Discretizing nonlinear, non-Gaussian Markov processes with exact conditional moments

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

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

Approximating stochastic processes by finite-state Markov chains is useful for re-
ducing 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 en-
tropy. In contrast to existing methods, our approach is not limited to linear Gaus-
sian autoregressive processes. We apply our method to numerically solve asset
pricing models with various underlying stochastic processes for the fundamen-
tals, including a rare disasters model. Our method outperforms the solution accu-
racy 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, numeri-
cal methods, solution accuracy.

JEL classification. C63, C68, G12.

1. Introduction

Many nonlinear dynamic economic models such as dynamic stochastic general equi-
librium (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 ap-
proximations of stochastic processes are a useful way to reduce computational complex-
ity when solving and estimating such models because integration is replaced by summa-

tion.[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 that 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 to discretize. 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 in most situations—their method cannot be

---

[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, Kubler, and Schmedders, 2011), asset pricing (Zhang, 2005, Guvenen, 2009), DSGE models (Aruoba, Fernándes-Villaverde, and Rubio-Ramírez, 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.

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 so as 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) (autoregressive) processes. We further discuss the relation of our method to the existing literature in Section 3.3.

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] of-

---

[3]Several papers, such as Aruoba, Fernándes-Villaverde, and Rubio-Ramírez (2006) and Caldara et al. (2012), compare the accuracy of various solution techniques (log-linearization, value function iteration,

ten by several orders of magnitude, 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 to solve 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 let $g : \mathbb{R}^K \to \mathbb{R}$ be any bounded continuous function. The first step is to pick a quadrature formula

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

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. For example, if we let

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

---

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 the simpler notation x_n and w_n.

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

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

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 $\bar{T} = \int_{\mathbb{R}^K} T(x) f(x) \, dx$ 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

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

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

---

[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 see Owen (2001), Tsao (2004), Kitamura (2007), and Tsao and Wu (2013) for discussions on the computational aspects of empirical likelihood methods, which are mathematically related.

(P) to the dual problem

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

which is a *low dimensional* (*L* unknowns) *unconstrained* concave maximization problem and hence is 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. (i) *The primal problem* (P) *has a solution if and only if $\bar{T} \in \text{co}\, T(D_N)$. If a solution exists, it is unique.*

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

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

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

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

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

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

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

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

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

---

[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

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 $\bar{T} \in \text{int co } T(D_N)$ 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

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

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

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

where C is a constant explicitly given in their 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).

2.2 *Discretizing general Markov processes*

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

---

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.

2.2.1 *Description of method*   Consider the time-homogeneous first-order Markov process

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

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

$$\bar{T}_n = \mathrm{E}\big[T(x_t)|x_n\big] = \int T(x)\, dF(x, x_n),$$

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

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

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

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

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

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

holds. We summarize our procedure in Algorithm 2.2 below.

ALGORITHM 2.2 (Discretization of Markov processes).

Step 1. Select a discrete set of points $D_N = \{x_n\}_{n=1}^N$ and an initial approximation $Q = (q_{nn'})$.

Step 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.

Step 3. For each $n = 1, \ldots, N$, solve the 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 $\bar{T}_n = \mathrm{E}[T(x_t)|x_{t-1} = x_n]$, 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 in which 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*   So as 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 evenly 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 ex-

ponentially 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 evenly 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 one 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 $D_N = \{\bar{x}_1, \dots, \bar{x}_N\}$ 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 $\bar{T}_n \in \text{int co } T(D_N)$ may fail to hold near the boundary. In the examples below, we consider matching up to four conditional moments whenever we can.

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 we prove some theoretical properties.

3.1 VAR(1)

Suppose we want to discretize a VAR(1) process

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

---

[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.

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 so as 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 $y_t = (y_{1t}, \ldots, y_{Kt})$ 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 $j = 1, \ldots, J$ be an index of the state, corresponding to a particular combination of points $(\bar{y}_{1n}(j), \ldots, \bar{y}_{Kn}(j))$. 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).

Step 1. For each component of $y_t = (y_{1t}, \ldots, y_{Kt})$, select a discrete set of points $D_{k,N_k} = \{\bar{y}_{kn}\}_{n=1}^{N_k}$.

Step 2. For $j = 1, \ldots, J$, perform the following steps:

(a) For $k = 1, \ldots, K$ (note that we can treat each component k separately because the variance–covariance matrix D is diagonal),

---

[8]Clearly there are infinitely many such decompositions. Experience tells us that the quality of discretization is best when each component of the y_t process in (3.2) has the same unconditional variance. We can construct such a decomposition as follows. 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 $\mathrm{Var}[\varepsilon_t] = U'IU = I$. The unconditional variance of the y process is then $\Sigma = U'\tilde{\Sigma}U$. Since $\mathrm{tr}\,\Sigma = \mathrm{tr}\,\tilde{\Sigma}$, the diagonal elements of Σ become equal if $\Sigma_{kk} = (U'\tilde{\Sigma}U)_{kk} = \frac{1}{K}\mathrm{tr}\,\tilde{\Sigma}$. We can make this equation (approximately) true by solving the optimization problem

$$\text{minimize} \quad \sum_{k=1}^{K}\left((U'\tilde{\Sigma}U)_{kk} - \frac{1}{K}\mathrm{tr}\,\tilde{\Sigma}\right)^2$$

$$\text{subject to} \quad U'U = I.$$

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

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

(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'.

Step 3. Collect the conditional probabilities $\{p_{jj'}\}_{j'=1}^{J}$ into a matrix $P = (p_{jj'})$.

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] 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 one-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* $\bar{T}_n \in \text{int co } T(D_N)$ *holds, and hence our method matches the conditional*

---

[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 there are N_k discrete points for y_{kt}, we divide the real line $\mathbb{R}$ into N_k intervals using the nth 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 $N(\mu_k(j), \sigma_k^2(j))$.

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.

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

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

where $0 \le \rho < 1$. 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*

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

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

When the grid $\{\bar{x}_n\}$ is evenly 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 evenly spaced, let $\sigma = 1/\sqrt{1 - \rho^2}$ be the unconditional standard deviation, and let $M = \max_n \bar{x}_n$. Suppose one of the following alternatives hold:*

(i) *We have $\rho \leq 1 - \frac{2}{N-1}$ and $\sigma < M \leq \sqrt{2}\sigma\sqrt{N-1}$.*

(ii) *We have $\rho > 1 - \frac{2}{N-1}$ 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]

3.2 AR(1) *with stochastic volatility*

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

$$y_t = \lambda y_{t-1} + u_t, \quad u_t \sim N(0, e^{x_t}), \tag{3.4a}$$

$$x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \tag{3.4b}$$

where x_t is the unobserved log variance process and y_t is the observable, for example, 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

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

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

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

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

---

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

With some more algebra, we can show that

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

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 evenly 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

---

[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."

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 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 evenly 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 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. More important questions are 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? 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 coefficient of relative risk aversion (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 let P_t be its price. When log consumption and dividend growth

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

follow a VAR(1) process with independent and identically distributed (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 Appendix, available in a supplementary file on the journal website, http://qeconomics.org/supp/737/supplement.pdf, for details.

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

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

Dividing (4.1) by D_t, we obtain

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

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

$$v = \beta P \operatorname{diag}\left(\mathrm{e}^{X\alpha}\right)(v+1) \quad \Longleftrightarrow \quad v = \left(I - \beta P \operatorname{diag}\left(\mathrm{e}^{X\alpha}\right)\right)^{-1} \beta P \mathrm{e}^{X\alpha}. \tag{4.4}$$

---

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

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

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

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.

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

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

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 and $\mathcal{M}$ are homogeneous of degree 1 (which is true for almost all applications) and that the underlying

---

[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.

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

stochastic process is discretized. Dividing (4.6) by C_t, we can solve for the S nonlinear equations in S unknowns,

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

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 the Federal Reserve Economic Data (FRED), and 12-month moving sums of dividends paid on the Standard and Poors (S&P) 500 obtained from the spreadsheet in Welch and Goyal (2008).[18] For the two univariate specifications, we assume that $C_t = D_t$, that is, $x_{1,t} = x_{2,t} = x_t$, 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 $\frac{1-\gamma}{1-\rho} \le s \le 1 - \gamma$ (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, to make a difference for asset pricing, we need to either (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 Appendix, we evaluate the accuracy by the $\log_{10}$ relative errors

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

---

[17]The idea of using discretization to solve asset pricing models is not particularly new: see, for example, Mehra and Prescott (1985), Cecchetti, Lam, and Mark (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/.

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 1001 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 evenly 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 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

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

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 evenly spaced grid with two or four 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 1 shows the graphs of $\log_{10}$ relative errors for the VAR(1) model. Table 1 shows the mean and maximum $\log_{10}$ relative errors over the entire grid.

For all choices of N, the Gaussian quadrature based methods, ME-Quad and TH, perform the best, with ME-Quad being always about 2 orders of magnitude more accurate than TH. For evenly 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 1, the conclusions drawn from Figure 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 nonnormality in financial data. One might prefer to specify a parametric distribution with fatter tails or to refrain from parametric specifications altogether. For this reason, we consider an AR(1) with i.i.d., but non-Gaussian shocks:
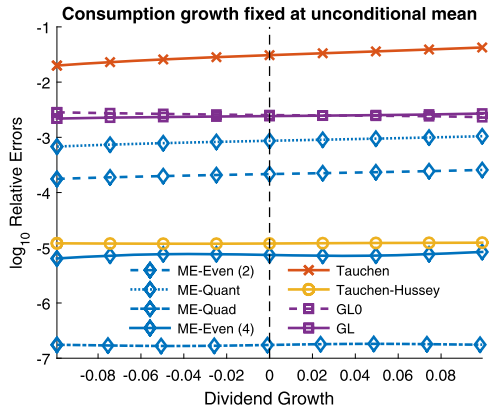
$$x_t = (1 - \rho)\mu + \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim F.$$
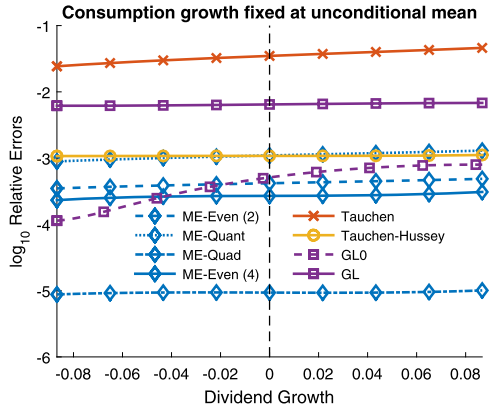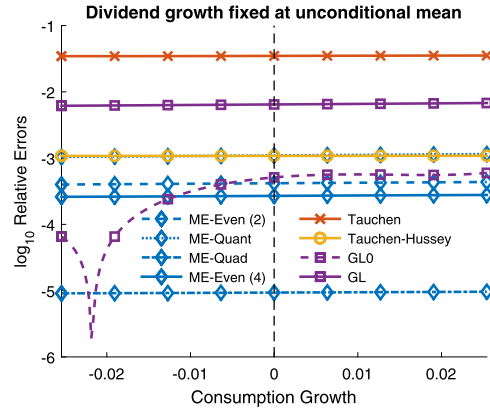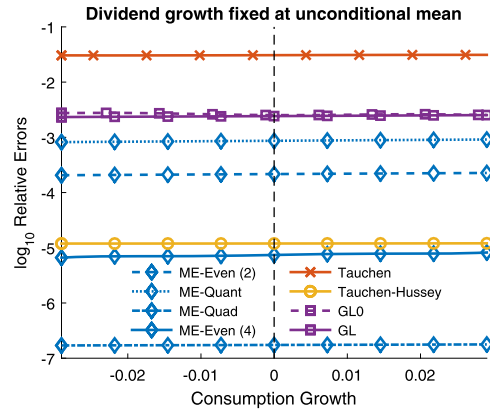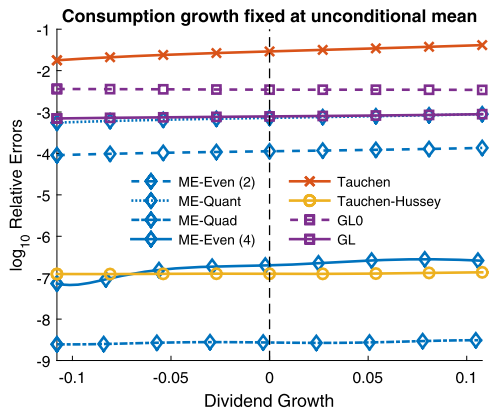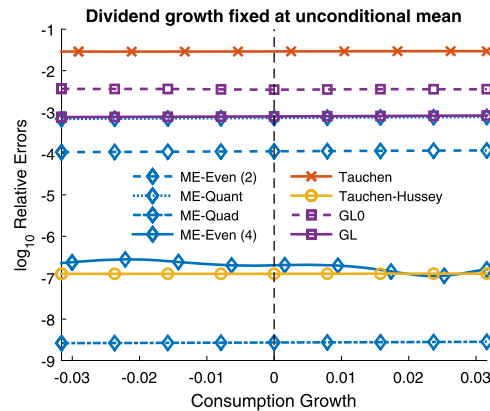
FIGURE 1. The $\log_{10}$ relative errors of the price–dividend ratio with various discretization methods and number of points for the VAR(1) model. Each row corresponds to a certain number of grid points ($N = 5, 7, 9$). The left panels show the accuracy along the dividend growth dimension, fixing consumption 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.

Table 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

Table 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, $j = 1, \ldots, J$. Growths μ and ρ are estimated by ordinary leastsquares (OLS); $\sigma = \sqrt{\mathrm{Var}[\varepsilon_t]}$ 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).

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 2 shows the parameter estimates.

Figure 2 plots the probability density functions (PDFs) of ε_t fitted to the dividend growth data under the assumptions of normal and Gaussian mixture shocks, as well as the nonparametric kernel density estimate. The Gaussian mixture with three components appears to capture the skewness and kurtosis lacking in the normal specification by placing more weight on large negative realizations of the shock as well as ones close to zero.

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 evenly spaced (ME-Even) or Gauss–Hermite quadrature grid (ME-GH), each with two or four 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
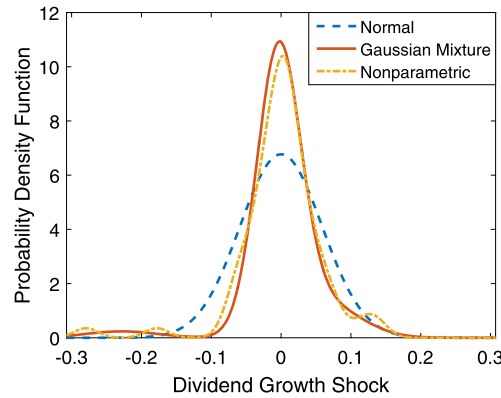
FIGURE 2. Densities fitted to the AR(1) OLS residuals.

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 $w'_n = w_n \frac{f(x_n)}{\phi(x_n)}$ to discretize $f(x)$. Figure 3 plots the $\log_{10}$ relative errors of the AR(1) model with Gaussian mixture shocks. Table 3 shows the mean and maximum $\log_{10}$ relative errors.

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 four moments instead of two increases the solution accuracy by about 1–2 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 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 four moments have 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

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

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

(c) $N = 9$, $\gamma = 5$.
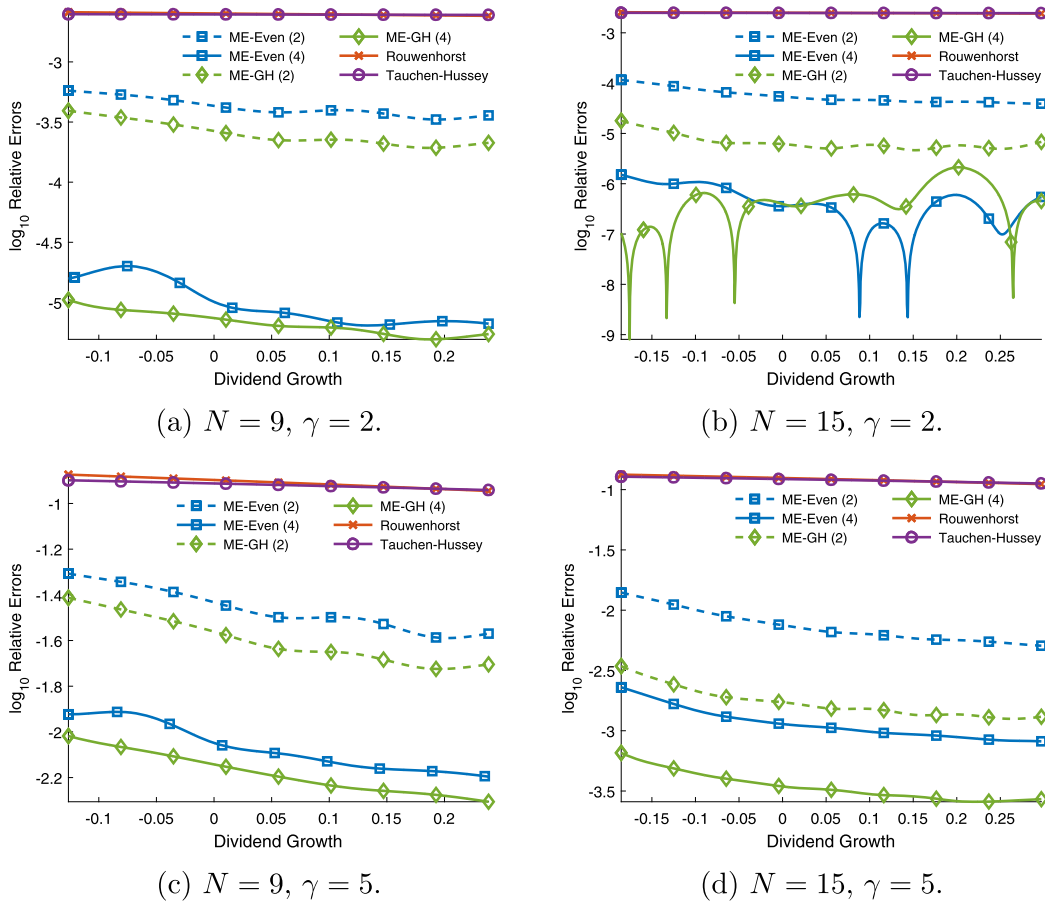
(d) $N = 15$, $\gamma = 5$.

FIGURE 3. The $\log_{10}$ relative errors with various discretization methods and number of points for the Gaussian mixture model. 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 $\gamma = 5$.

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

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

TABLE 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*) denotes an evenly spaced grid with *L* moments; GH (*L*) denotes a Gauss–Hermite quadrature grid with *L* moments; R denotes the Rouwenhorst (1995) method; TH denotes the Tauchen and Hussey (1991) method.

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 quantity, which he calls resilience,

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

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, and F_t individually, Gabaix directly specifies the dynamics of $H_t = H_* + \widehat{H}_t$ as

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

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 equation (13) in his paper).
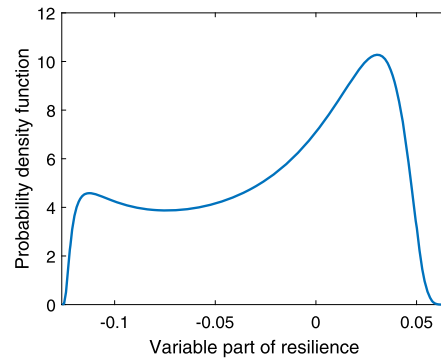
FIGURE 4. Ergodic distribution of the variable part of resilience $\widehat{H}$.

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 $\{\widehat{H}_t\}$ 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 in this way, it is straightforward to discretize the Markov process using our method. See the Appendix for the details on discretization and the computation of the numerical solution. Although there are no accepted standard ways to solve 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 $B = 0.66$, and the speed of mean reversion $\phi_H = 0.13$. The implied value for the constant H_* in (5.2) is 0.09. Figure 4 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.

For our method, we consider the evenly 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 is $N = 5, 11, 21, 41, 81$. For the perturbation method in Levintal (2014), we consider up to the fifth-order approximation (the maximum allowed). So as to apply the perturbation method, we need to supply the unconditional standard deviation of the innovation in resilience, ε_{t+1}^H. We compute this number using the ergodic distribution in Figure 4, 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 shows the $\log_{10}$ relative errors of the price–dividend ratio. Table 4 shows the mean and maximum $\log_{10}$ relative errors over the entire grid.
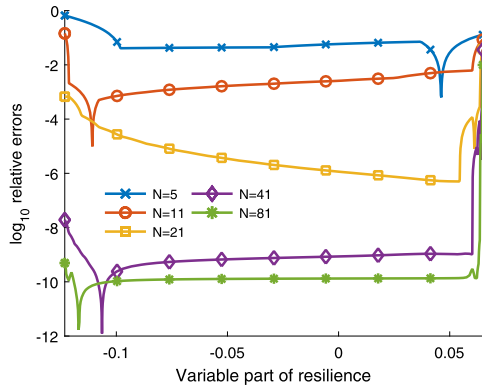

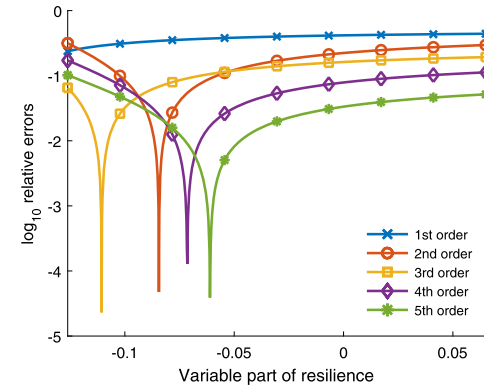[19]https://sites.google.com/site/orenlevintal/5th-order-perturbation.
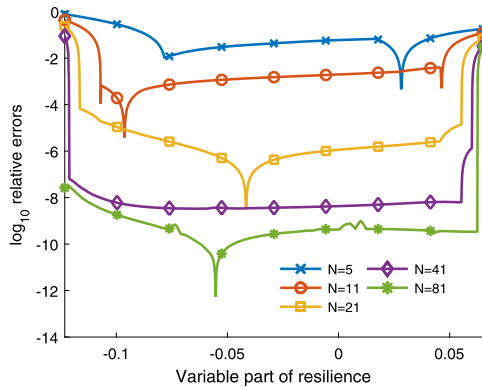
(a) Even-spaced grid.
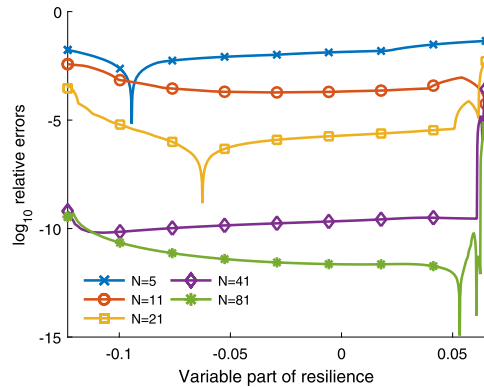
(b) Gauss-Legendre quadrature grid.

(c) Clenshaw-Curtis quadrature grid.

(d) Perturbation method.

FIGURE 5. The $\log_{10}$ relative errors of the price–dividend ratio with various methods and number of points or order of approximation for the variable rare disaster model.

TABLE 4. Mean and maximum $\log_{10}$ relative errors for the variable disaster model.

	ME Methods			Perturbation	
	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

Because the resilience process (5.2) is highly nonlinear, we need many grid points to obtain an accurate solution. Overall using the Gauss–Legendre quadrature grid (Figure 5(b)) 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(c)) is similar to Gauss–Legendre, as documented in Trefethen (2008). The performance of the evenly spaced grid (Figure 5(a)) 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(d)) is not so accurate, with about 10% error with third-order approximation and 2.6% error with fifth-order. Even the 5-point Gauss–Legendre discretization is more accurate than the fifth-order perturbation in terms of both mean and maximum $\log_{10}$ errors.

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 shows the results. As expected from Figure 5 and Table 4, 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 first-order approximation, the stock return is 4 percentage points higher than the true value; the third-order approximation is off by 10–20%, and the fifth-order approximation is off by about 10% for the standard deviation.

Based on the numerical results in the last two sections, we provide some recommendations 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 evenly spaced or quadrature grid) seems best. Numerical results in the Appendix suggest that for univariate Gaussian AR(1) processes, 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, Schmedders, and Wilms (2015) suggest that to solve 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.

TABLE 5. 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.

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 that 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, Ozkan, and Song, 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

Adda, J. and R. W. Cooper (2003), *Dynamic Economics: Quantitative Methods and Applications*. MIT Press, Cambridge, MA. [659, 662, 665]

Aguirregabiria, V. and P. Mira (2007), "Sequential estimation of dynamic discrete games." *Econometrica*, 75 (1), 1–53. [652]

Aiyagari, S. R. (1994), "Uninsured idiosyncratic risk and aggregate saving." *Quarterly Journal of Economics*, 109 (3), 659–684. [652]

Aiyagari, S. R. (1995), "Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting." *Journal of Political Economy*, 103 (6), 1158–1175. [652]

Aruoba, S. B., J. Fernándes-Villaverde, and J. F. Rubio-Ramírez (2006), "Comparing solution methods for dynamic equilibrium economies." *Journal of Economic Dynamics and Control*, 30 (12), 2477–2508. [652, 653]

Bansal, R. and A. Yaron (2004), "Risks for the long run: A potential resolution of asset pricing puzzles." *Journal of Finance*, 59 (4), 1481–1509. [678]

Barro, R. J. (2006), "Rare disasters and asset markets in the twentieth century." *Quarterly Journal of Economics*, 121 (3), 823–866. [674]

Bonomo, M., R. Garcia, N. Meddahi, and R. Tédongap (2011), "Generalized disappointment aversion, long-run volatility risk, and asset prices." *Review of Financial Studies*, 24 (1), 82–122. [669]

Borwein, J. M. and A. S. Lewis (1991), "Duality relationships for entropy-like minimization problems." *SIAM Journal on Control and Optimization*, 29 (2), 325–338. [655]

Burnside, C. (1998), "Solving asset pricing models with Gaussian shocks." *Journal of Economic Dynamics and Control*, 22 (3), 329–340. [666, 667]

Caldara, D., J. Fernándes-Villaverde, J. F Rubio-Ramírez, and W. Yao (2012), "Computing DSGE models with recursive preferences and stochastic volatility." *Review of Economic Dynamics*, 15 (2), 188–206. [652, 653]

Cecchetti, S. G., P. Lam, and N. C. Mark (1993), "The equity premium and the risk-free rate: Matching moments." *Journal of Monetary Economics*, 31 (1), 21–45. [669]

Collard, F. and M. Juillard (2001), "Accuracy of stochastic perturbation methods: The case of asset pricing models." *Journal of Economic Dynamics and Control*, 25 (6-7), 979–999. [667]

Dávila, J., J. H. Hong, P. Krusell, and J. Ríos-Rull (2012), "Constrained efficiency in the neoclassical growth model with uninsurable idiosyncratic shocks." *Econometrica*, 80 (6), 2431–2467. [652]

Davis, P. J. and P. Rabinowitz (1984), *Methods of Numerical Integration*, second edition. Academic Press, Orlando, FL. [659]

Farmer, L. E. (2016), "The discretization filter: A simple way to estimate nonlinear state space models." Available at https://ssrn.com/abstract=2780166. [654, 680]

Flodén, M. (2008), "A note on the accuracy of Markov-chain approximations to highly persistent AR(1) processes." *Economics Letters*, 99 (3), 516–520. [665]

Gabaix, X. (2012), "Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance." *Quarterly Journal of Economics*, 127 (2), 645–700. [653, 673, 674, 675, 676, 679]

Galindev, R. and D. Lkhagvasuren (2010), "Discretization of highly persistent correlated AR(1) shocks." *Journal of Economic Dynamics and Control*, 34 (7), 1260–1276. [665]

Gospodinov, N. and D. Lkhagvasuren (2014), "A moment-matching method for approximating vector autoregressive processes by finite-state Markov chains." *Journal of Applied Econometrics*, 29 (5), 843–859. [652, 653, 665, 666, 670]

Gourio, F. (2012), "Disaster risk and business cycles." *American Economic Review*, 102 (6), 2734–2766. [674]

Guvenen, F. (2009), "A parsimonious macroeconomic model for asset pricing." *Econometrica*, 77 (6), 1711–1750. [652]

Guvenen, F., S. Ozkan, and J. Song (2014), "The nature of countercyclical income risk." *Journal of Political Economy*, 122 (3), 621–660. [680]

Haliassos, M. and A. Michaelides (2003), "Portfolio choice and liquidity constraints." *International Economic Review*, 44 (1), 143–177. [652]

Hamilton, J. D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica*, 57 (2), 357–384. [654]

Heaton, J. and D. J. Lucas (1996), "Evaluating the effects of incomplete markets on risk sharing and asset pricing." *Journal of Political Economy*, 104 (3), 443–487. [652]

Heiss, F. and V. Winschel (2008), "Likelihood approximation by numerical integration on grids." *Journal of Econometrics*, 144 (1), 62–80. [660]

Judd, K. L. (1992), "Projection methods for solving aggregate growth models." *Journal of Economic Theory*, 58 (2), 410–452. [654, 668]

Judd, K. L., F. Kubler, and K. Schmedders (2011), "Bond ladders and optimal portfolios." *Review of Financial Studies*, 24 (12), 4123–4166. [652]

Kitamura, Y. (2007), "Empirical likelihood methods in econometrics: Theory and practice." In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress* (R. Blundell, W. Newey, and T. Persson, eds.), Econometric Society Monographs, Vol. 3, 174–237, Cambridge University Press, New York. [655]

Kopecky, K. A. and R. M. H. Suen (2010), "Finite state Markov-chain approximations to highly persistent processes." *Review of Economic Dynamics*, 13 (3), 701–714. [654, 664, 665, 666]

Krueger, D. and F. Kubler (2004), "Computing equilibrium in OLG models with stochastic production." *Journal of Economic Dynamics and Control*, 28 (7), 1411–1436. [660]

Kullback, S. and R. A. Leibler (1951), "On information and sufficiency." *Annals of Mathematical Statistics*, 22 (1), 79–86. [655]

Levintal, O. (2014), "Fifth order perturbation solution to DSGE models." Available at http://ssrn.com/abstract=2364989. [676]

Maliar, L. and S. Maliar (2015), "Merging simulation and projection approaches to solve high-dimensional problems with an application to a new Keynesian model." *Quantitative Economics*, 6 (1), 1–47. [660]

Mehra, R. and E. C. Prescott (1985), "The equity premium: A puzzle." *Journal of Monetary Economics*, 15 (2), 145–161. [669]

Owen, A. B. (2001), *Empirical Likelihood*. Monographs on Statistics and Applied Probability, Vol. 92. Chapman & Hall/CRC. [655]

Pohl, W., K. Schmedders, and O. Wilms (2015), "Higher-order dynamics in asset-pricing models with recursive preferences." Available at https://ssrn.com/abstract=2540586. [678]

Rietz, T. A. (1988), "The equity risk premium: A solution." *Journal of Monetary Economics*, 22 (1), 117–131. [674]

Rouwenhorst, K. G. (1995), "Asset pricing implications of equilibrium business cycle models." In *Frontiers of Business Cycle Research* (T. F. Cooley, ed.), 294–330, Princeton University Press. [653, 664, 665, 666, 672, 675]

Schmitt-Grohé, S. and M. Uribe (2004), "Solving dynamic general equilibrium models using a second-order approximation to the policy function." *Journal of Economic Dynamics and Control*, 28 (4), 755–775. [667]

Tanaka, K. and A. A. Toda (2013), "Discrete approximations of continuous distributions by maximum entropy." *Economics Letters*, 118 (3), 445–450. [652, 654, 655]

Tanaka, K. and A. A. Toda (2015), "Discretizing distributions with exact moments: Error estimate and convergence analysis." *SIAM Journal on Numerical Analysis*, 53 (5), 2158–2177. [652, 653, 654, 657, 659]

Tauchen, G. (1986), "Finite state Markov-chain approximations to univariate and vector autoregressions." *Economics Letters*, 20 (2), 177–181. [665, 670]

Tauchen, G. and R. Hussey (1991), "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models." *Econometrica*, 59 (2), 371–396. [659, 662, 665, 670, 672, 675]

Terry, S. J. and E. S. Knotek II (2011), "Markov-chain approximations of vector autoregressions: Application of general multivariate-normal integration techniques." *Economics Letters*, 110 (1), 4–6. [665]

Trefethen, L. N. (2008), "Is Gauss quadrature better than Clenshaw-Curtis?" *SIAM Review*, 50 (1), 67–87. [678]

Tsao, M. (2004), "Bounds on coverage probabilities of the empirical likelihood ratio confidence regions." *Annals of Statistics*, 32 (3), 1215–1221. [655]

Tsao, M. and F. Wu (2013), "Empirical likelihood on the full parameter space." *Annals of Statistics*, 41 (4), 2176–2196. [655]

Tsionas, E. G. (2003), "Exact solution of asset pricing models with arbitrary shock distributions." *Journal of Economic Dynamics and Control*, 27 (5), 843–851. [667, 669]

Vavra, J. (2014), "Inflation dynamics and time-varying volatility: New evidence and an *Ss* interpretation." *Quarterly Journal of Economics*, 129 (1), 215–258. [652]

Wachter, J. A. (2013), "Can time-varying risk of rare disasters explain aggregate stock market volatility?" *Journal of Finance*, 68 (3), 987–1035. [674]

Welch, I. and A. Goyal (2008), "A compehensive look at the empirical performance of equity premium prediction." *Review of Financial Studies*, 21 (4), 1455–1508. [669]

Zhang, L. (2005), "The value premium." *Journal of Finance*, 60 (1), 67–103. [652, 665]