Discretizing Stochastic Processes with Exact Conditional Moments

Leland E. Farmer   Alexis Akira Toda

University of California San Diego

May 4, 2015

Motivation

- Nonlinear dynamic economic models (DSGE, optimal portfolio, etc.) often imply a set of integral equations (*e.g.*, Euler equations) that do not admit explicit solutions.
- Finite-state Markov chain approximations of stochastic processes are a useful way of reducing the complexity of solving these models: $\int$ becomes $\sum$.
- Existing methods are not good at or not even capable of approximating high-dimensional and highly persistent processes, and only apply to VARs.

Contribution

1. Extend Tanaka and Toda "Discrete Approximations of Continuous Distributions by Maximum Entropy" (EL, 2013) to the approximation of stochastic processes.
2. Show that our method is computationally tractable and provides more accurate approximations than existing methods for VARs and stochastic volatility models.
3. Our method improves solution accuracy of simple asset pricing models by many orders of magnitude over existing methods.

Literature

Discrete approximations of VARs **Tauchen (1986)**, Tauchen and Hussey (1991), **Rouwenhorst (1995)**, Adda & Cooper (2003), Flodén (2008), Galindev & Lkhagvasuren (2010), Kopecky & Suen (2010), Terry & Knotek (2011), **Gospodinov & Lkhagvasuren (2014)**.

Maximum entropy Shannon (1948), Jaynes (1957), Shore & Johnson (1980), Borwein & Lewis (1991), Caticha and Giffin (2006), Tanaka & Toda (2013, 2014, 2015), & many many more.

Tanaka and Toda (2013)

- **Goal**: Approximate the probability density function f on $\mathbb{R}^K$ by probabilities $P = \{p_n\}_{n=1}^N$ on a finite discrete subset $D_N = \{x_{n,N}\}_{n=1}^N \subset \mathbb{R}^K$.
- Assume some moments $\bar{T} = \int_{\mathbb{R}^K} T(x)f(x)\mathrm{d}x$ are given, where $T : \mathbb{R}^K \to \mathbb{R}^L$ is a measurable function.
- **Example**: If first and second moments are given, we would have:

$$T(x) = (x_1, \dots, x_K, x_1^2, \dots, x_k x_l, \dots, x_K^2)$$

with $L = K + K + \frac{K(K-1)}{2}$ for the K expected values, K variances, and $\frac{K(K-1)}{2}$ covariances.

Tanaka and Toda (2013)

- To match the given moments with a discrete distribution, it suffices to find probabilities $\{p_n\}_{n=1}^N$ that satisfy:

$$\sum_{n=1}^{N} p_n T(x_{n,N}) = \bar{T}, \qquad (L \text{ equations})$$

$$\sum_{n=1}^{N} p_n = 1. \qquad (1 \text{ equation})$$

- **Problem**: Number of points in D_N, N, is in general much larger than number of equations, $L+1$.
  $\implies \{p_n\}_{n=1}^N$ underdetermined (ill-posed problem).
- **Solution**: Maximum entropy.

Primal problem

- Given a discrete subset $D_N \subset \mathbb{R}^K$, an initial approximation $Q = \{q_n\}_{n=1}^N$, a moment defining function $T : \mathbb{R}^K \to \mathbb{R}^L$, and moments $\bar{T} \in \mathbb{R}^L$, solve:

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

- That is, find the least informative posterior (in terms of Kullback-Leibler information) that matches the moments.

Dual problem

- Solution to the previous problem, $\{p_n\}_{n=1}^N$, is given by:

$$p_n = \frac{q_n \mathrm{e}^{\lambda_N' T(x_{n,N})}}{\sum_{n=1}^N q_n \mathrm{e}^{\lambda_N' T(x_{n,N})}},$$

  where λ_N is the Lagrange multiplier to the moment constraint.
- λ_N is a solution to the dual problem (Borwein & Lewis, 1991):

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

Tanaka & Toda (2015, R&R SIAM J. Num. Anal.)

- Let $E^Q_{g,N} = \left\| \int f(x)g(x)\mathrm{d}x - \sum_{n=1}^N q_n g(x_n) \right\|$ be the integration error of initial distribution $Q = \{q_n\}$, and define $E^P_{g,N}$ similarly.
- Let $g(x) \approx b_{g,T}(x) = \sum_{l=1}^L b_l T_l(x)$ be the approximation of the integrand using the basis functions $T = \{T_l\}_{l=1}^L$, and $r_{g,T} = \frac{g-b_{g,T}}{\left\|g-b_{g,T}\right\|_\infty}$ be normalized residual.
- Obtain the error estimate

$$E^P_{g,N} \le \|g - b_{g,T}\|_\infty \left(E^Q_{r_{g,T},N} + \frac{2}{\sqrt{C}} E^Q_{T,N} \right),$$

  so the error improves by the factor $\|g - b_{g,T}\|_\infty$.
- In particular, $\{p_n\}_{n=1}^N$ weakly converges to f as $N \to \infty$ when $\{q_n\}_{n=1}^N$ does.

Some notes

- If $\bar{T}$ is in the interior of the convex hull of $T(D_N)$, then the objective function of dual problem is continuous, strictly convex, and a unique solution λ_N exists.
- Our original, high-dimensional constrained optimization problem (N unknowns, $L+1$ equality constraints, N inequality constraints) reduces to a low-dimensional unconstrained convex minimization problem (L unknowns, no constraints).
- **Example**: In 3 dimensions, 10 grid points in each dimension, and 2 moments (mean & variance) to match, primal problem has 1000 unknowns and 1010 constraints, while dual has 9 unknowns and no constraints.
- Complexity of primal problem is exponential in dimension, but dual is polynomial.

  $\Longrightarrow$ Computationally tractable in multivariate case.

Application to VAR processes

- Assume we have a general VAR(1) of the form
$$x_t = b + Bx_{t-1} + \eta_t, \ \eta_t \sim N(0, \Psi).$$
- Define the unconditional mean $\mu = (I - B)^{-1}b$, regular matrix C and diagonal D such that $CDC' = \Psi$ (*e.g.*, Cholesky decomposition).
- Finding a discretization of x_t is equivalent to finding a discretization of
$$y_t = Ay_{t-1} + \epsilon_t, \ \epsilon_t \sim N(0, D),$$
where $y_t = C^{-1}(x_t - \mu)$, $A = C^{-1}BC$, and $\epsilon_t = C^{-1}\eta_t$.
- Hence suffices to discretize y_t and use the inverse transformation $x_t = \mu + Cy_t$ to recover a discretization of x_t.

Application to VAR processes

- Problem reduces to discretizing the VAR(1)

$$y_t = Ay_{t-1} + \epsilon_t, \ \epsilon_t \sim N(0, D),$$

  so VAR has zero mean and cross-uncorrelated shocks.
- More precisely, we wish to define a finite-state Markov chain with $(S \times S)$ transition matrix P that approximates the dynamics of our VAR.
- Row s of P corresponds to the conditional probability measure of moving from state s to any other state in the chain. If $P = (p_{ss'})$, then $p_{ss'}$ is probability of $s_t = s'$ conditional on $s_{t-1} = s$.

Outline of procedure

1. Define a set $D_S = \{\bar{y}_s\}_{s=1}^{S} \subset \mathbb{R}^K$, which is the support of the Markov chain. (More on how to pick D_S later.)
2. Conditional on being in state s at time $t-1$, the distribution of y_t is $N(A\bar{y}_s, D)$.
3. Since D diagonal, can discretize coordinate-by-coordinate to discretize vector y_t.

Note:

- Procedure matches conditional moments by construction.
- Since minimizing KL information puts positive probability, transition probability matrix P is positive.
  $\implies$ Markov chain is stationary and ergodic by construction. (*c.f.*, Perron-Frobenius theorem.)

Practical concerns

- Picking the support of the Markov chain:
  - Tensor grids: Even-spaced (Tauchen, 1986), quadrature-based (Tauchen & Hussey, 1991), quantiles (Adda & Cooper, 2003).
  - Non-rectangular grids: Epsilon-distinguishable sets (Maliar & Maliar, 2014).
- Picking how many moments to match: **2**? 3? 4? more?
- Picking initial approximation: proportional, quadrature-based, uniform.

Generalization to arbitrary stochastic processes

- Unlike other procedures previously proposed in the literature, our method is neither limited to the approximation of VARs nor the use of tensor grids.
- Consider a general nonlinear stochastic process given by

$$x_t = \phi(x_{t-1}, \varepsilon_t), \ \varepsilon_t \sim F_\varepsilon.$$

- Assuming that x_t is a stationary, ergodic process, and that certain moments of x_t conditional on x_{t-1} can be computed (often the case in economic models, *e.g.*, when we know $f(x_t|x_{t-1})$), we can apply our method.
- Difficulty becomes how to pick support of the Markov chain: most promising method is probably epsilon-distinguishable sets.

AR(1)

- Consider AR(1) process

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

  with unconditional variance $\sigma^2 = \frac{1}{1-\rho^2}$.
- Persistence: $\rho \in \{0.5, 0.9, 0.99, 0.999, 0.9999\}$, number of discrete points: $N \in \{9, 15, 21\}$.
- Construct 3 different Markov chain approximations:
  1. optimized Tauchen (1986),
  2. Rouwenhorst (1995), and
  3. our method (matching two conditional moments).

Simulation exercise

- Sample size: $T = 2,000,000$, discard first 200,000 observations as burn-in. Number of Monte Carlo replications: 1,000.
- For each Monte Carlo sample, estimate persistence $\hat{\rho}$ and unconditional variance $\hat{\sigma}^2$ by OLS.
- For each parameter, compute the root mean-squared error (RMSE), bias, and standard deviation (SD) relative to their true values. Example:

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{\theta}_m - \theta)^2} \Big/ \theta \, ,$$

where $m = 1, \dots, M = 1,000$ are Monte Carlo replications and $\theta = 1 - \rho, \sigma^2$ is parameter of interest.

Root mean-squared error, scaled by 10^{-3}

N	ρ	Tauchen (optimized)		Rouwenhorst		FTT (2 moments)	
		$1-\hat{\rho}$	$\hat{\sigma}^2$	$1-\hat{\rho}$	$\hat{\sigma}^2$	$1-\hat{\rho}$	$\hat{\sigma}^2$
9	0.5	14.785	1.155	1.244	1.209	1.231	0.897
	0.9	94.069	2.328	3.054	2.801	3.017	2.187
	0.99	41.396	6.349	9.976	9.390	10.151	7.271
	0.999	N/A	N/A	30.228	28.892	32.141	18.892
	0.9999	N/A	N/A	102.599	101.321	104.471	81.096
15	0.5	6.532	1.227	1.228	1.270	1.243	1.086
	0.9	43.638	2.647	3.113	3.034	3.103	2.579
	0.99	226.847	6.454	9.756	9.570	10.007	8.521
	0.999	N/A	N/A	30.592	29.564	32.230	18.518
	0.9999	N/A	N/A	93.009	94.525	99.412	77.916
21	0.5	3.767	1.212	1.269	1.264	1.220	1.139
	0.9	25.788	2.715	2.995	2.912	3.116	2.658
	0.99	165.361	6.798	9.736	9.512	10.207	8.562
	0.999	163.418	162.371	31.399	30.575	32.317	19.887
	0.9999	N/A	N/A	99.173	101.920	98.734	82.396

Bias, scaled by 10^{-3}

N	ρ	Tauchen (optimized)		Rouwenhorst		FTT (2 moments)	
		$1-\hat{\rho}$	$\hat{\sigma}^2$	$1-\hat{\rho}$	$\hat{\sigma}^2$	$1-\hat{\rho}$	$\hat{\sigma}^2$
9	0.5	-14.736	-0.007	0.057	-0.022	-0.032	-0.011
	0.9	-94.042	-0.039	-0.075	0.058	0.003	-0.029
	0.99	-40.812	0.070	0.037	0.059	-0.118	-0.027
	0.999	N/A	N/A	-0.777	1.792	-0.901	1.856
	0.9999	N/A	N/A	-6.341	16.928	-1.685	14.527
15	0.5	-6.415	0.042	0.063	-0.014	-0.028	0.040
	0.9	-43.557	-0.005	-0.104	0.134	0.081	-0.073
	0.99	-226.788	0.216	-0.318	0.301	0.234	0.028
	0.999	N/A	N/A	-0.191	0.948	-0.288	0.811
	0.9999	N/A	N/A	-11.422	20.950	-3.076	16.806
21	0.5	-3.573	-0.004	-0.022	-0.011	0.041	0.033
	0.9	-25.644	-0.009	0.011	-0.072	-0.035	0.140
	0.99	-165.259	0.436	0.066	0.079	0.518	-0.240
	0.999	161.562	160.713	0.490	0.907	-0.588	1.602
	0.9999	N/A	N/A	-11.229	20.659	0.688	15.841

Standard deviation, scaled by 10^{-3}

N	ρ	Tauchen (optimized)		Rouwenhorst		FTT (2 moments)	
		$1-\hat{\rho}$	$\hat{\sigma}^2$	$1-\hat{\rho}$	$\hat{\sigma}^2$	$1-\hat{\rho}$	$\hat{\sigma}^2$
9	0.5	1.204	1.156	1.243	1.209	1.231	0.898
	0.9	2.243	2.328	3.054	2.802	3.018	2.188
	0.99	6.891	6.352	9.981	9.394	10.156	7.275
	0.999	N/A	N/A	30.233	28.851	32.145	18.810
	0.9999	N/A	N/A	102.454	99.947	104.509	79.824
15	0.5	1.234	1.226	1.227	1.270	1.243	1.085
	0.9	2.660	2.648	3.113	3.032	3.103	2.579
	0.99	5.181	6.454	9.755	9.570	10.010	8.525
	0.999	N/A	N/A	30.606	29.564	32.244	18.509
	0.9999	N/A	N/A	92.351	92.220	99.414	76.120
21	0.5	1.195	1.213	1.270	1.265	1.220	1.139
	0.9	2.724	2.716	2.996	2.912	3.117	2.656
	0.99	5.816	6.787	9.740	9.516	10.199	8.562
	0.999	24.573	23.156	31.411	30.577	32.328	19.832
	0.9999	N/A	N/A	98.584	99.855	98.781	80.900

VAR(1)

- Consider VAR(1) process used in Gospodinov & Lkhagvasuren (2014) (henceforth GL):

$$y_t = Ay_{t-1} + \epsilon_t, \qquad \epsilon_t \sim N(0, \Psi),$$
$$A = \begin{bmatrix} 0.9809 & 0.0028 \\ 0.0410 & 0.9648 \end{bmatrix}, \qquad \Psi = \begin{bmatrix} 0.0087^2 & 0 \\ 0 & 0.0262^2 \end{bmatrix}.$$

- Number of points in each dimension: $N \in \{9, 15, 21\}$.
- Construct 3 different Markov chain approximations:
  1. GL0 (generalization of Rouwenhorst to VAR),
  2. GL (modify GL by targeting 2 conditional moments in primal problem, so computationally infeasible in high dimension), and
  3. our method (matching 2 conditional moments).

Simulation exercise

- Sample size: $T = 2,000,000$, discard first 200,000 observations as burn-in. Number of Monte Carlo replications: 1,000.
- For each Monte Carlo simulation, compute the unconditional variances, covariance and the eigenvalues of $\hat{A}$.
- Compute the root mean-squared error, bias, and standard deviation of all the estimates relative to their true values.

Numerical results: VAR(1), scaled by 10^{-3}

	$N=9$			$N=15$			$N=21$		
	GL0	**GL**	**FTT**	**GL0**	**GL**	**FTT**	**GL0**	**GL**	**FTT**
Root mean squared error									
$\hat{\sigma}_z^2$	90.557	8.855	5.873	75.016	7.792	6.356	63.251	7.634	6.993
$\hat{\sigma}_g^2$	121.753	9.172	4.641	102.118	6.154	5.117	86.865	6.483	5.604
$\hat{\rho}_{zg}$	23.606	10.666	8.635	11.711	8.707	8.579	10.598	8.684	8.884
$1-\hat{\zeta}_1$	14.372	10.561	8.568	8.623	8.536	8.356	8.503	8.321	8.599
$1-\hat{\zeta}_2$	6.203	5.252	4.901	4.935	4.984	4.878	5.079	4.964	5.216
Bias									
$\hat{\sigma}_z^2$	-90.312	4.910	-0.129	-74.641	0.504	0.118	-62.810	0.376	0.252
$\hat{\sigma}_g^2$	-121.645	6.860	-0.058	-101.958	0.353	0.191	-86.677	0.334	0.104
$\hat{\rho}_{zg}$	21.773	6.421	-0.056	7.608	0.561	0.028	5.778	0.582	0.254
$1-\hat{\zeta}_1$	-12.007	-6.872	0.273	-0.578	-0.517	-0.103	0.172	-0.345	-0.158
$1-\hat{\zeta}_2$	-3.302	-2.012	-0.009	-0.521	0.052	-0.277	0.131	-0.027	0.042
Standard deviation									
$\hat{\sigma}_z^2$	6.663	7.373	5.875	7.501	7.779	6.358	7.462	7.629	6.992
$\hat{\sigma}_g^2$	5.111	6.092	4.643	5.724	6.147	5.116	5.707	6.477	5.606
$\hat{\rho}_{zg}$	9.125	8.521	8.639	8.907	8.694	8.583	8.889	8.669	8.884
$1-\hat{\zeta}_1$	7.902	8.023	8.568	8.608	8.524	8.360	8.505	8.318	8.602
$1-\hat{\zeta}_2$	5.253	4.854	4.903	4.910	4.986	4.873	5.080	4.966	5.218

Stochastic volatility model

- Consider the stochastic volatility model

$$\begin{aligned} y_t &= \lambda y_{t-1} + \exp(x_t/2)u_t, & u_t &\sim N(0,1) \\ x_t &= \mu(1-\rho) + \rho x_{t-1} + \epsilon_t, & \epsilon_t &\sim N(0,\sigma^2) \end{aligned}$$

  where x_t: log variance, y_t: observable (*e.g.*, stock returns).
- Since unconditional variance is

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

  can choose an even-spaced 3σ grid for y_t.
- Discretize x_t as before (Rouwenhorst or our method). Discretize $y_t|x_{t-1}, y_{t-1}$ using Tauchen or our method.

Simulation exercise

- Parameter values taken from Caldara, Fernándes-Villaverde, Rubio-Ramírez, & Yao (RED, 2012):

Parameter	λ	μ	ρ	σ
Value	0.95	-9.3332	0.9	0.06

- Generate 1,000 samples of length 100,000 each, run OLS, and compute $\hat{\lambda}$, $\hat{\sigma}_y$ (standard deviation of y_t), and $\hat{\kappa}_y$ (kurtosis of y_t).
- Report RMSE relative to exact value.

Relative RMSE of stochastic volatility discretization

N_x	N_y	λ	Tauchen-Rouwenhorst			FTT		
			$\hat{\lambda}$	$\hat{\sigma}_y$	$\hat{\kappa}_y$	$\hat{\lambda}$	$\hat{\sigma}_y$	$\hat{\kappa}_y$
9	9	0	0.0032	0.0178	0.0398	0.0031	0.0024	0.0331
		0.5	0.0059	0.0246	0.0413	0.0056	0.0030	0.0294
		0.9	0.0023	0.1010	0.0635	0.0016	0.0068	0.0246
		0.95	0.0017	0.1690	0.0844	0.0010	0.0097	0.0532
		0.99	0.0087	0.2795	0.1322	0.0006	0.0278	0.1144
15	15	0	0.0030	0.0028	0.0422	0.0032	0.0023	0.0480
		0.5	0.0060	0.0045	0.0425	0.0054	0.0030	0.0421
		0.9	0.0021	0.0286	0.0533	0.0016	0.0070	0.0190
		0.95	0.0015	0.0597	0.0644	0.0010	0.0096	0.0207
		0.99	0.0018	0.1979	0.1149	0.0005	0.0224	0.0455
21	21	0	0.0032	0.0032	0.0433	0.0033	0.0023	0.0545
		0.5	0.0057	0.0033	0.0431	0.0056	0.0030	0.0481
		0.9	0.0020	0.0110	0.0514	0.0016	0.0071	0.0211
		0.95	0.0014	0.0253	0.0570	0.0011	0.0101	0.0214

Solving asset pricing models

- Consider a standard representative-agent asset pricing model with CRRA utility
$$\sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\gamma}}{1-\gamma},$$
where $\beta = 0.95$ and $\gamma = 2$.
- Single asset (stock) with dividend D_t. Consider three models:
  1. $C_t = D_t$, and consumption growth $\log C_t/C_{t-1}$ is AR(1).
  2. $C_t \neq D_t$, stock is in zero net supply, and consumption growth and dividend growth $(\log C_t/C_{t-1}, \log D_t/D_{t-1})$ is VAR(1).
  3. $C_t = D_t$, and consumption growth obeys the stochastic volatility model.

Experimental design

- Estimate each process from 1947–2014 US data.
- For AR(1) and VAR models, a closed-form solution exists (Burnside, JEDC 1998).
- We solve each model using Chebyshev collocation for each discretization method and compare solution accuracy.

Results for AR(1) model

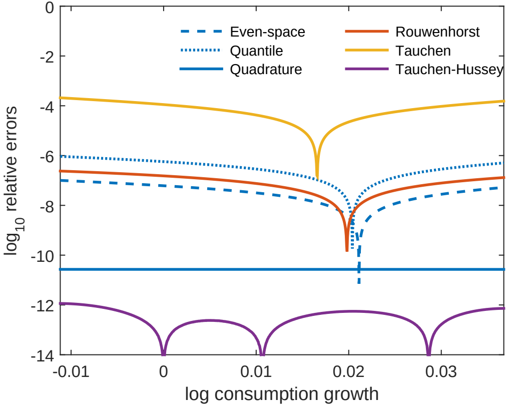


Figure Baseline model, $\rho = 0.27$

Results for AR(1) model

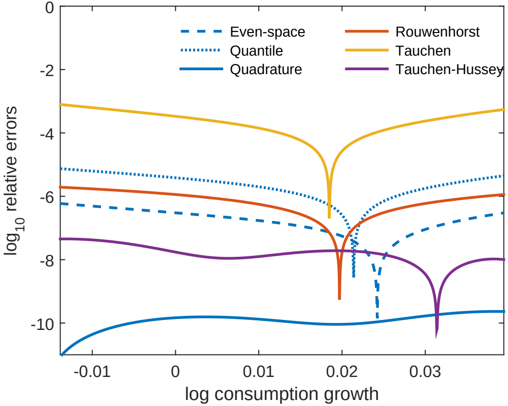


Figure $\rho = 0.50$

Results for AR(1) model

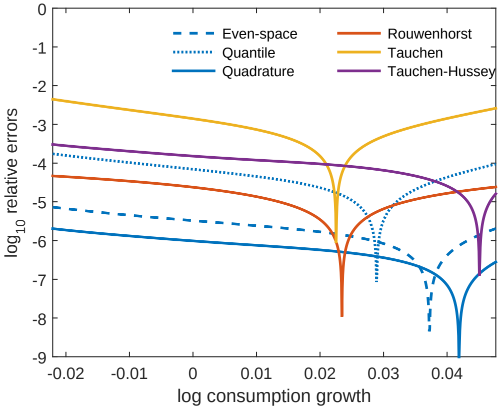
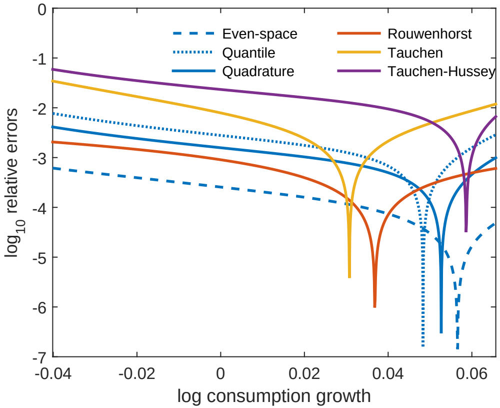


Figure $\rho = 0.75$

Results for AR(1) model



Figure$\rho = 0.90$

Results for VAR model

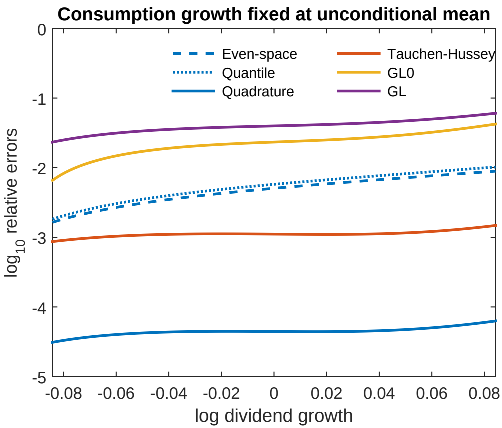


Figure $N = 5$ points

Results for VAR model

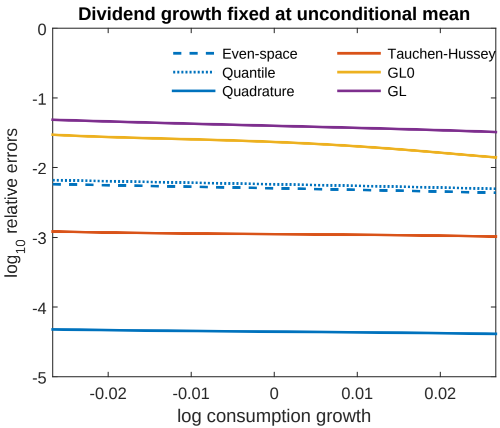


Figure $N = 5$ points

Results for VAR model

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

	Our method			Existing methods		
N	Even-space	Quantile	Quadrature	TH	GL0	GL
Mean $\log_{10}$ *errors*						
5	-2.333	-2.277	-4.351	-2.951	-1.651	-1.409
9	-2.669	-2.494	-7.984	-6.775	-2.047	-2.398
13	-2.975	-2.367	-7.702	-7.704	-2.047	-3.163
Maximum $\log_{10}$ *errors*						
5	-1.999	-1.939	-4.126	-2.679	-0.611	-0.580
9	-2.251	-2.056	-7.541	-6.273	-0.724	-0.693
13	-1.998	0.012	-7.363	-6.926	-1.373	-1.337

Results for stochastic volatility model

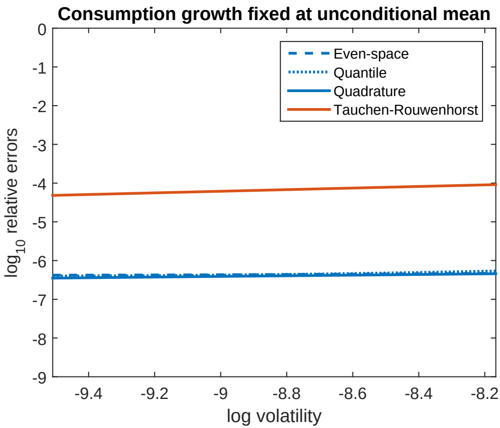
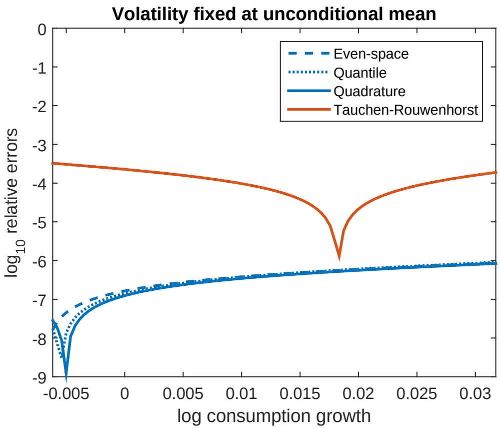


Figure $N = 5$ points

Results for stochastic volatility model



Figure$N = 5$ points

Conclusion

- New method for finite-state Markov chain approximation of stochastic processes using maximum entropy:
  1. Matches an arbitrary number of conditional moments given a fine enough grid.
  2. Outperforms existing methods along most dimensions in terms of accuracy for linear stochastic processes.
  3. Our method applies to any stochastic process. Parametric model not even necessary (*e.g.*, can estimate transition density nonparametrically and discretize).
- Useful for solving dynamic models with complicated dynamics (stochastic volatility, etc.) with high accuracy.

Future work

- Generalize to nonlinear stochastic processes.
- More interesting applications (*e.g.*, numerical option pricing).
- Use for efficiently estimating nonlinear state space models (as opposed to linear models, *e.g.*, Kalman filter). (Leland has nice preliminary work.)