Lecture 5: Stochastic HJB Equations, Kolmogorov Forward Equations

ECO 521: Advanced Macroeconomics I

Benjamin Moll

Princeton University

Fall 2012

Outline

(1) Hamilton-Jacobi-Bellman equations in stochastic settings (without derivation)

(2) Ito's Lemma

(3) Kolmogorov Forward Equations

(4) Application: Power laws (Gabaix, 2009)

Stochastic Optimal Control

- Generic problem:

$$V(x_0) = \max_{u(t)_{t=0}^{\infty}} \mathbb{E}_0 \int_0^{\infty} e^{-\rho t} h(x(t), u(t))\, dt$$

subject to the law of motion for the state

$$dx(t) = g(x(t), u(t))\, dt + \sigma(x(t)) dW(t) \text{ and } u(t) \in U$$

for $t \geq 0$, $x(0) = x_0$ given.

- Deterministic problem: special case $\sigma(x) \equiv 0$.
- In general $x \in \mathbb{R}^m$, $u \in \mathbb{R}^n$. For now do scalar case.

Stochastic HJB Equation: Scalar Case

- Claim: the HJB equation is

$$\rho V(x) = \max_{u \in U} \; h(x,u) + V'(x) g(x,u) + \frac{1}{2} V''(x) \sigma^2(x)$$

- Here: on purpose no derivation ("cookbook")
- In case you care, see any textbook, e.g. chapter 2 in Stokey (2008)
- Sidenote: can again write this in terms of the Hamiltonian

$$\rho V(x) = \max_{u \in U} \; \mathcal{H}(x, u, V'(x)) + \frac{1}{2} V''(x) \sigma^2(x)$$

Just for Completeness: Multivariate Case

- Let $x \in \mathbb{R}^m, u \in \mathbb{R}^n$.
- For fixed x, define the $m \times m$ covariance matrix

$$\sigma^2(x) = \sigma(x)\sigma(x)'$$

(this is a function $\sigma^2 : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$)

- The HJB equation is

$$\rho V(x) = \max_{u \in U} \; h(x,u) + \sum_{i=1}^{m} \frac{\partial V(x)}{\partial x_i} g_i(x,u) + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\partial^2 V(x)}{\partial x_i \partial x_j} \sigma_{ij}^2(x)$$

- In vector notation

$$\rho V(x) = \max_{u \in U} \; h(x,u) + \nabla_x V(x) \cdot g(x,u) + \frac{1}{2} \mathrm{tr}\left(\Delta_x V(x) \sigma^2(x)\right)$$

- $\nabla_x V(x)$: gradient of V (dimension $m \times 1$)
- $\Delta_x V(x)$: Hessian of V (dimension $m \times m$).

HJB Equation: Endogenous and Exogenous State

- Lots of problems have the form $x = (x_1, x_2)$
  - x_1: endogenous state
  - x_2: exogenous state

$$dx_1 = \tilde{g}(x_1, x_2, u)dt$$

$$dx_2 = \tilde{\mu}(x_2)dt \quad + \quad \tilde{\sigma}(x_2)dW$$

- Special case with

$$g(x) = \begin{bmatrix} \tilde{g}(x_1, x_2, u) \\ \tilde{\mu}(x_2) \end{bmatrix}, \quad \sigma(x) = \begin{bmatrix} 0 \\ \tilde{\sigma}(x_2) \end{bmatrix}$$

- Claim: the HJB equation is

$$\rho V(x_1, x_2) = \max_{u \in U} \; h(x_1, x_2, u) + V_1(x_1, x_2)\tilde{g}(x_1, x_2, u)$$
$$+ V_2(x_1, x_2)\tilde{\mu}(x_2) + \frac{1}{2} V_{22}(x_1, x_2)\tilde{\sigma}^2(x_2)$$

Example: Real Business Cycle Model

$$V(k_0, A_0) = \max_{c(t)_{t=0}^{\infty}} \mathbb{E}_0 \int_0^{\infty} e^{-\rho t} U(c(t)) dt$$

subject to

$$dk = [AF(k) - \delta k - c]dt$$
$$dA = \mu(A)dt + \sigma(A)dW$$

for $t \geq 0$, $k(0) = k_0$, $A(0) = A_0$ given.

- Here: $x_1 = k, x_2 = A, u = c$
- $h(x, u) = U(u)$
- $g(x, u) = F(x) - \delta x - u$

Example: Real Business Cycle Model

- HJB equation is

$$\rho V(k,A) = \max_c \; U(c) + V_k(k,A)[AF(k) - \delta k - c] + V_A(k,A)\mu(A) + \frac{1}{2} V_{AA}(k,A)\sigma^2(A)$$

Example: Real Business Cycle Model

- Special Case 1: A is a geometric Brownian motion

$$dA = \mu A dt + \sigma A dW$$

$$\rho V(k,A) = \max_c \; U(c) + V_k(k,A)[AF(k) - \delta k - c] + V_A(k,A)\mu A + \frac{1}{2} V_{AA}(k,A)\sigma^2 A^2$$

See Merton (1975) for an analysis of this case.

- Special Case 2: A is a Feller square root process

$$dA = \theta(\bar{A} - A)dt + \sigma\sqrt{A}dW$$

$$\rho V(k,A) = \max_c \; U(c) + V_k(k,A)[AF(k) - \delta k - c] + V_A(k,A)\theta(\bar{A} - A) + \frac{1}{2} V_{AA}(k,A)\sigma^2 A$$

Special Case: Stochastic AK Model with log Utility

- Preferences: $U(c) = \log c$
- Technology: $AF(k) = Ak$
- A follows any diffusion

$$\rho V(k,A) = \max_c \log c + V_k(k,A)[Ak - \delta k - c] + V_A(k,A)\mu(A) + \frac{1}{2}V_{AA}(k,A)\sigma^2(A)$$

- **Claim:** Optimal consumption is $c = \rho k$ and hence capital follows

$$dk = [A - \rho - \delta]kdt$$

$$dA = \mu(A)dt + \sigma(A)dt$$

- Solution prop's? Simply simulate two SDEs forward in time.

Special Case: Stochastic AK Model with log Utility

- **Proof:** Guess and verify

$$V(k,A) = v(A) + \kappa \log k$$

- FOC:

$$U'(c) = V_k(k,A) \quad \Leftrightarrow \quad \frac{1}{c} = \frac{\kappa}{k} \Leftrightarrow \quad c = \frac{k}{\kappa}$$

- Substitute into HJB equation

$$\begin{aligned}\rho[v(A) + \kappa \log k] = & \log k - \log \kappa + \frac{\kappa}{k}\left[Ak - \delta k - k/\kappa\right] \\ & + v'(A)\mu(A) + \frac{1}{2}v''(A)\sigma^2(A)\end{aligned}$$

- Collect terms involving $\log k \Rightarrow \kappa = 1/\rho \Rightarrow c = \rho k.\square$
- Comment: log-utility $\Rightarrow$ offsetting income and substitution effects of future $A \Rightarrow$ constant savings rate ρ.

General Case: Numerical Solution with FD Method

- See `HJB_stochastic_reflecting.m`
- Solve on bounded grids $k_i, i = 1, ..., I$ and $A_j, j = 1, ..., J$
- Use short-hand notation $V_{i,j} = V(k_i, A_j)$. Approximate

$$V_k(k_i, A_j) \approx \frac{V_{i+1,j} - V_{i-1,j}}{2\Delta k}$$
$$V_A(k_i, A_j) \approx \frac{V_{i,j+1} - V_{i,j+1}}{2\Delta A}$$
$$V_{AA}(k_i, A_j) \approx \frac{V_{i,j+1} - 2V_{i,j} + V_{i,j-1}}{(\Delta A)^2}$$

- Discretized HJB

$$\begin{aligned}\rho V_{i,j} =& U(c_{i,j}) + V_k(k_i, A_j)[A_j F(k_i) - \delta k_i - c_{i,j}] \\ &+ V_A(k_i, A_j)\mu(A_j) + \frac{1}{2} V_{AA}(k_i, A_j)\sigma^2(A_j)\end{aligned}$$
$$c_{i,j} = (U')^{-1}[V_k(k_i, A_j)]$$

General Case: Numerical Solution with FD Method

- As boundary conditions, use

$$V_A(k, A_1) = 0 \quad \text{all } k \quad \Rightarrow \quad V_{i,0} = V_{i,2}$$

$$V_A(k, A_J) = 0 \quad \text{all } k \quad \Rightarrow \quad V_{i,J+1} = V_{i,J-1}$$

- These correspond to "reflecting barriers" at lower and upper bounds for productivity, A_1 and A_J (Dixit, 1993).
- In theory also need boundary condition for k (possibility: reflecting barrier at k_I)
- Instead, use "dirty fix": backward and forward rather than central differences at boundaries

$$V_k(k_1, A) = \frac{V_{2,j} - V_{1,j}}{\Delta k}, \quad V_k(k_I, A) = \frac{V_{I,j} - V_{I-1,j}}{\Delta k}$$

General Case: Numerical Solution with FD Method

- Iterate using same explicit method as in deterministic case.
- Guess, V^0, update using:

$$\frac{V_{i,j}^{n+1} - V_{i,j}^n}{\Delta} + \rho V_{i,j}^n = U(c_{i,j}^n) + V_k^n(k_i, A_j)[A_j F(k_i) - \delta k_i - c_{i,j}^n] + V_A^n(k_i, A_j)\mu(A_j) + \frac{1}{2} V_{AA}^n(k_i, A_j)\sigma^2(A_j)$$

- See `HJB_stochastic_reflecting.m`
- **Extremely** inefficient: need 112,140 iterations.
- Implicit Method?

Ito's Lemma

- Let x be a scalar diffusion

$$dx = \mu(x)dt + \sigma(x)dW$$

- We are interested in the evolution of $y(t) = f(x(t))$ where f is **any** twice differentiable function.
- **Lemma**: $y(t) = f(x(t))$ follows

$$df(x) = \left(\mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)\right)dt + \sigma(x)f'(x)dW$$

- Extremely powerful because it says that **any** (twice differentiable) function of a diffusion is also a diffusion.
- Can also be extended to vectors.
- FYI: this is also where the $V'(x)\mu(x) + \frac{1}{2}V''(x)\sigma^2(x)$ term in the HJB equation comes from (it's $\frac{\mathbb{E}[dV(x)]}{dt}$).

Application: Brownian vs. Geometric Brownian Motion

- Let x be a geometric Brownian motion

$$dx = \mu x dt + \sigma x dW$$

- Claim: $y = \log x$ is a Brownian motion with drift $\mu - \sigma^2/2$ and variance σ^2.
- Derivation: $f(x) = \log x$, $f'(x) = 1/x, f''(x) = -1/x^2$

  By Ito's Lemma

$$\begin{aligned} dy = df(x) &= \left(\mu x(1/x) + \frac{1}{2}\sigma^2 x^2(-1/x^2)\right) dt + \sigma x(1/x) dW \\ &= \left(\mu - \sigma^2/2\right) dt + \sigma dW \end{aligned}$$

- Note: naive derivation would have used $dy = dx/x$ and hence

$$dy = \mu dt + \sigma dW \quad \text{wrong unless } \sigma = 0!$$

Just for Completeness: Multivariate Case

- Let $x \in \mathbb{R}^m$. For fixed x, define the $m \times m$ covariance matrix

$$\sigma^2(x) = \sigma(x)\sigma(x)'$$

- **Ito's Lemma:**

$$df(x) = \left(\sum_{i=1}^{n} \mu_i(x) \frac{\partial f(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma_{ij}^2(x) \frac{\partial^2 f(x)}{\partial x_i x_j} \right) dt + \sum_{i=1}^{m} \sigma_i(x) \frac{\partial f(x)}{\partial x_i} dW_i$$

- In vector notation

$$df(x) = \left(\nabla_x f(x) \cdot \mu(x) + \frac{1}{2} \mathrm{tr} \left(\Delta_x f(x) \sigma^2(x) \right) \right) dt + \nabla_x f(x) \cdot \sigma(x) dW$$

- $\nabla_x f(x)$: gradient of f (dimension $m \times 1$)
- $\Delta_x f(x)$: Hessian of f (dimension $m \times m$).

Kolmogorov Forward Equations

- Let x be a scalar diffusion

$$dx = \mu(x)dt + \sigma(x)dW, \quad x(0) = x_0$$

- Suppose we're interested in the evolution of the **distribution** of x, $f(x,t)$, and in particular in the limit $\lim_{t\to\infty} f(x,t)$.
- Natural thing to care about especially in heterogenous agent models
- Example 1: $x =$ wealth
  - $\mu(x)$ determined by savings behavior and return to investments
  - $\sigma(x)$ by return risk.
  - microfound later
- Example 2: $x =$ city size, will cover momentarily

Kolmogorov Forward Equations

- **Fact:** Given an initial distribution $f(x,0) = f_0(x)$, $f(x,t)$ satisfies the PDE

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[\mu(x)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x)f(x,t)]$$

- This PDE is called the "Kolmogorov Forward Equation"
- Note: in math this often called "Fokker-Planck Equation"
- Can be extended to case where x is a vector as well.
- **Corollary**: if a stationary distribution, $\lim_{t\to\infty} f(x,t) = f(x)$ exists, it satisfies the ODE

$$0 = -\frac{d}{dx}[\mu(x)f(x)] + \frac{1}{2}\frac{d^2}{dx^2}[\sigma^2(x)f(x)]$$

Just for Completeness: Multivariate Case

- Let $x \in \mathbb{R}^m$.
- As before, define the $m \times m$ covariance matrix

$$\sigma^2(x) = \sigma(x)\sigma(x)'$$

- The Kolmogorov Forward Equation is

$$\frac{\partial f(x,t)}{\partial t} = -\sum_{i=1}^{m} \frac{\partial}{\partial x_i}[\mu_i(x) f(x,t)] + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \frac{\partial^2}{\partial x^2}[\sigma_{ij}^2(x) f(x,t)]$$

Application: Stationary Distribution of RBC Model

- Recall RBC Model

$$\rho V(k,A) = \max_c \; U(c) + V_k(k,A)[AF(k) - \delta k - c] + V_A(k,A)\mu(A) + \frac{1}{2} V_{AA}(k,A)\sigma^2(A)$$

- Denote the optimal policy function by

$$\dot{k}(k,A) = AF(k) - \delta k - c(k,A)$$

- Then $f(k,A,t)$ solves

$$\frac{\partial f(k,A,t)}{\partial t} = -\frac{\partial}{\partial k}[\dot{k}(k,A)f(k,A,t)] - \frac{\partial}{\partial A}[\mu(A)f(k,A,t)] + \frac{1}{2}\frac{\partial^2}{\partial A^2}[\sigma^2(A)f(k,A,t)]$$

- Can discretize using FD method, run forward, see if it converges to stationary distribution.

Application: Power Laws

- See Gabaix (2009), "Power Laws in Economics and Finance," very nice, very accessible!
- Pareto (1896!!!): upper-tail distribution of number of people with an income or wealth S greater than a large x is proportional to $1/x^{\zeta}$ for some $\zeta > 0$

$$\Pr(S > x) = kx^{-\zeta}$$

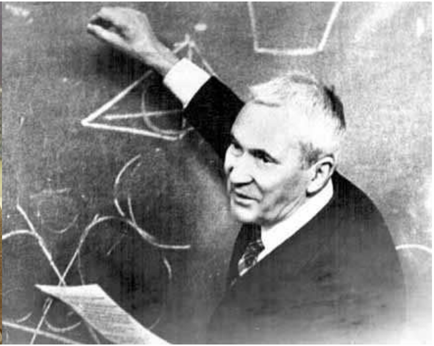
- **Definition:** We say that a variable, x, follows a power law (PL) if there exist $k > 0$ and $\zeta > 0$ such that

$$\Pr(S > x) = kx^{-\zeta}, \quad \text{all } x$$

- x follows a PL $\Leftrightarrow$ x has a Pareto distribution
- Holds for surprisingly many variables.

History Interlude

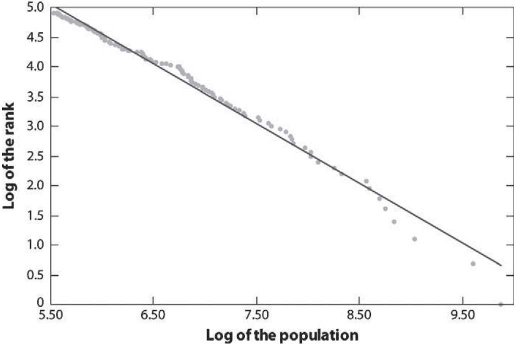




Vilfredo Pareto   Kiyoshi Ito   Andrei Kolmogorov

City Size

- Order cities in US by size (NY as first, LA as second, etc)
- Graph ln Rank ($\ln \text{Rank}_{NY} = \ln 1, \ln \text{Rank}_{LA} = \ln 2$) vs. ln Size
- Basically plot log quantiles $\ln \Pr(S > x)$ against $\ln x$



City Size

- Surprise 1: straight line, i.e. city size follows a PL

$$\Pr(S > x) = kx^{-\zeta}$$

- Surprise 2: slope of line ≈ -1, regression:

$$\ln \text{Rank} = 10.53 - 1.005 \ln \text{Size}$$

i.e. city size follows a PL with exponent $\zeta \approx 1$

$$\Pr(S > x) = kx^{-1}.$$

- A power law with exponent $\zeta = 1$ is called "Zipf's law"
- Two natural questions:
  (1) Why does city size follow a power law?
  (2) Why on earth is $\zeta \approx 1$ rather than any other number?

Where Do Power Laws Come from?

- Gabaix's answer: random growth
- Economy with continuum of cities.
- S_t^i: size of city i at time t

$$S_{t+1}^i = \gamma_{t+1}^i S_t^i, \quad \gamma_{t+1}^i \sim f(\gamma) \tag{RG}$$

- S_t^i follows random growth process $\Leftrightarrow$ $\log S_t^i$ follows random walk.
- Gabaix shows: (RG) + friction (e.g. minimum size) $\Rightarrow$ power law. Use "Champernowne's equation"
- Easier: continuous time approach.

Random Growth Process in Continuous Time

- Consider random growth process over time intervals of length Δt

$$S^i_{t+\Delta t} = \gamma^i_{t+\Delta t} S^i_t$$

- Assume in addition that $\gamma^i_{t+\Delta t}$ takes the particular form

$$\gamma^i_{t+\Delta t} = 1 + g\Delta t + v\varepsilon^i_t\sqrt{\Delta t}, \quad \varepsilon^i_t \sim N(0,1)$$

- Substituting in

$$S^i_{t+\Delta t} - S^i_t = (g\Delta t + v\varepsilon^i_t\sqrt{\Delta t})S^i_t$$

- Or as $\Delta t \to 0$

$$dS^i_t = gS^i_t dt + vS^i_t dW^i_t$$

i.e. a geometric Brownian motion!

Stationary Distribution

- Assumption: city size follows random growth process

$$dS_t^i = gS_t^i dt + vS_t^i dW_t^i$$

- Does this have a stationary distribution? No! In fact

$$\log S_t^i \sim N((g - v^2/2)t, v^2 t)$$

$\Rightarrow$ distribution explodes.

- Gabaix insight: random growth process + friction does have a stationary distribution and that's a PL
- Simplest possible friction: minimum size $S_{\min}$. If process goes below $S_{\min}$ it is brought back to $S_{\min}$ ("reflecting barrier")

Stationary Distribution

- Use Kolmogorov Forward Equation.
- Recall: stationary distribution satisfies

$$0 = -\frac{d}{dx}[\mu(x)f(x)] + \frac{1}{2}\frac{d^2}{dx^2}[\sigma^2(x)f(x)]$$

- Here geometric Brownian motion: $\mu(x) = gx, \sigma^2(x) = v^2x^2$

$$0 = -\frac{d}{dx}[gxf(x)] + \frac{1}{2}\frac{d^2}{dx^2}[v^2x^2f(x)]$$

Stationary Distribution

- **Claim:** solution is a Pareto distribution, $f(x) = S_{\min}^{\zeta} x^{-\zeta-1}$
- **Proof:** Guess $f(x) = Cx^{-\zeta-1}$ and verify

$$
\begin{aligned}
0 &= -\frac{d}{dx}[gxCx^{-\zeta-1}] + \frac{1}{2}\frac{d^2}{dx^2}[v^2x^2Cx^{-\zeta-1}] \\
&= Cx^{-\zeta-1}\left[g\zeta + \frac{v^2}{2}(\zeta-1)\zeta\right]
\end{aligned}
$$

- This is a quadratic equation with two roots $\zeta = 0$ and

$$\zeta = 1 - \frac{2g}{v^2}$$

- For mean to exist, need $\zeta > 1 \Rightarrow$ impose $g < 0$.
- Remains to pin down C. We need

$$1 = \int_{S_{\min}}^{\infty} f(x)dx = \int_{S_{\min}}^{\infty} Cx^{-\zeta-1}dx \quad \Rightarrow \quad C = S_{\min}^{\zeta}.\square$$

Zipf's Law

- Why would Zipf's Law ($\zeta = 1$) hold? We have that

$$\bar{S} = \int_{S_{\min}}^{\infty} x f(x) dx = \frac{\zeta}{\zeta - 1} S_{\min}$$

$$\Rightarrow \quad \zeta = \frac{1}{1 - S_{\min}/\bar{S}} \to 1 \quad \text{as} \quad S_{\min}/\bar{S} \to 0.$$

- Zip's law obtains as friction becomes small.

Alternative Friction: Death

- No minimum size.
- Instead: die at Poisson rate δ, get reborn at S_*.
- Can show: correct way of extending KFE (for $x \neq S_*$) is

$$\frac{\partial f(x,t)}{\partial t} = -\delta f(x,t) - \frac{\partial}{\partial x}[\mu(x) f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(x) f(x,t)\right]$$

- Stationary $f(x)$ satisfies (recall $\mu(x) = gx, \sigma^2(x) = v^2x^2$)

$$0 = -\delta f(x) - \frac{d}{dx}[gxf(x,t)] + \frac{1}{2}\frac{d^2}{dx^2}\left[\sigma^2 x^2 f(x)\right] \qquad \text{(KFE')}$$

Alternative Friction: Death

- To solve (KFE'), guess $f(x) = Cx^{-\zeta-1}$

$$0 = -\delta + \zeta g + \frac{v^2}{2}\zeta(\zeta - 1)$$

- Two roots: $\zeta_+ > 0$ and $\zeta_- < 0$. General solution to (KFE'):

$$\Rightarrow \quad f(x) = C_- x^{-\zeta_- -1} + C_+ x^{-\zeta_+ -1} \quad \text{for } x \neq S_*$$

- Need solution to be integrable

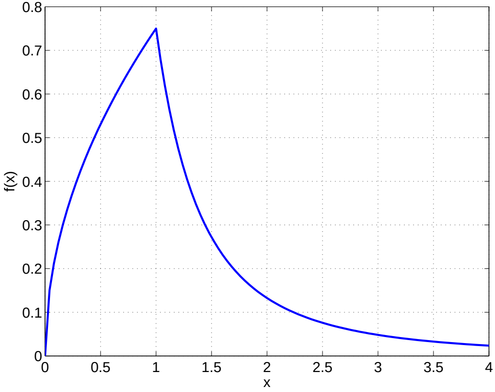
$$\int_0^\infty f(x)dx = f(S_*) + \int_0^{S_*} f(x)dx + \int_{S_*}^\infty f(x)dx < \infty$$

- Hence $C_- = 0$ for $x > S_*$, otherwise $f(x)$ explodes as $x \to \infty$.
- And $C_+ = 0$ for $x < S_*$, otherwise $f(x)$ explodes as $x \to 0$.

Alternative Friction: Death

- Solution is a **Double Pareto** distribution:

$$f(x) = \begin{cases} C(x/S_*)^{-\zeta_- -1} & \text{for } x < S_* \\ C(x/S_*)^{-\zeta_+ -1} & \text{for } x > S_* \end{cases}$$



Alternative Friction: Death

- Again, Zipf's Law ($\zeta = 1$) obtains as friction gets small.

  Here: $\delta \to 0$.

- Other cases in Gabaix's paper:
  (1) Extension to jump processes
  (2) Approximate power laws with generalized growth process

$$\frac{dS_t}{S_t} = g(S_t)dt + v(S_t)dt$$