Advanced Computational Methods in Finance

Julien Guyon

Ecole des Ponts ParisTech

Department of Mathematics
Baruch College, City University of New York

Master of Financial Engineering

The field of mathematics is very wide and it is not easy to predict what happens next, but I can tell you it is alive and well. Two general trends are obvious and will surely persist. In its pure aspect, the subject has changed, much for the better I think, by moving to more concrete problems. In both its pure and applied aspects, an equally beneficial shift to nonlinear problems can be seen. Most mathematical questions suggested by Nature are genuinely nonlinear, meaning very roughly that the result is not proportional to the cause, but varies with it as the square or the cube, or in some more complicated way. The study of such questions is still, after two or three hundred years, in its infancy. Only a few of the simplest examples are understood in any really satisfactory way. I believe this direction will be a principal theme in the future.

— **Henry P. McKean**, *Some Mathematical Coincidences (May 2003)*

Syllabus

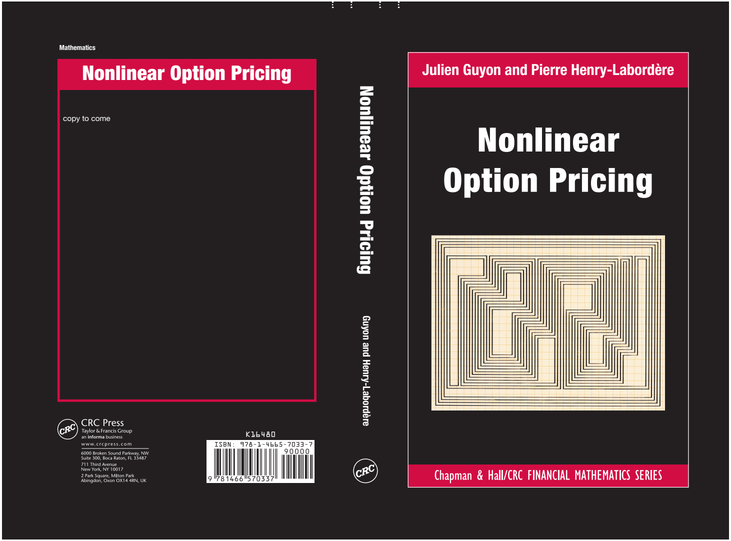
- The classical curriculum of mathematical finance programs generally covers the link between linear second order parabolic partial differential equations (PDEs) and stochastic differential equations (SDEs), resulting from Feynman-Kac's formula.
- However, many challenges faced by today's practitioners involve nonlinear PDEs.
- The aim of this course is to provide you with the mathematical tools and numerical methods required to tackle these issues, and illustrate the methods with practical case studies like:
  - American option pricing,
  - uncertain volatility, uncertain mortality, different rates for borrowing and lending,
  - calibration of models to market smiles,

Syllabus

- We will spend some time on the theory:
  - optimal stopping,
  - stochastic control,
  - backward stochastic differential equations (BSDEs),
  - McKean SDEs and the particle method.
- But the main focus will deliberately be on ideas and numerical examples, which we believe help a lot in understanding the tools and building intuition.
- PDE methods suffer from the curse of dimensionality. Since most quantitative finance problems are high-dimensional, we will mostly focus on simulation-based methods (a.k.a. Monte Carlo algorithms).
- This course exposes the students with a wide variety of Machine Learning techniques, old and new, including parametric regression, nonparametric regression, neural networks, etc. These techniques allow us to compute some quantities (such as conditional expectations) that are key ingredients of the nonlinear Monte Carlo algorithms.

Syllabus

- Homeworks will allow you to check your understanding of the course by solving exercises inspired by our experience as a quantitative analyst. You are expected to use the Python programming language to complete the 2 homeworks (60% of the final grade).
- A final exam will count towards 30% of the final grade. A written homework (10%) will help you train for the final.
- The main reference for this course is the monograph *Nonlinear Option Pricing* by Julien Guyon and Pierre Henry-Labordère.


Mathematics
Nonlinear Option Pricing
copy to come
CRC Press
Taylor & Francis Group
an informa business
www.crcpress.com
6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK
K16480
ISBN: 978-1-4665-7033-7
90000
9 781466 570337
Nonlinear Option Pricing
Guyon and Henry-Labordère
CRC
Julien Guyon and Pierre Henry-Labordère
Nonlinear
Option Pricing
Chapman & Hall/CRC FINANCIAL MATHEMATICS SERIES
K16480_Cover.indd 1
4/11/13 8:36 AM

Prerequisites

- We assume familiarity with classical stochastic analysis (Brownian motion, Itô formula, stochastic differential equations) and Black-Scholes option pricing.
- The students needing a reminder on these matters could consult:
  - Karatzas, I., Shreve, S.: Brownian Motion and Stochastic Calculus, Springer-Verlag, 1991.
  - Øksendal, B.: Stochastic Differential Equations: An Introduction with Applications, 5th ed., Springer-Verlag, 1998.
- We will briefly recap linear option pricing: self-financing strategies, arbitrage, equivalent martingale measures (risk-neutral measures), market completeness, super-replication, Black-Scholes PDE, Feynman-Kac theorem (link between parabolic second order linear PDEs and SDEs).
- Regarding the new tools introduced, the course will be self-contained.

Numerical tools: Finite difference schemes vs Monte Carlo methods

Problems described by a nonlinear PDE

$$\begin{aligned} \partial_t u(t,x) + H(t,x,u(t,x),Du(t,x),D^2u(t,x)) &= 0, \quad x \in \Omega, \quad t \in [0,T), \quad (1) \\ u(T,x) &= g(x) \end{aligned}$$

- Ω: open domain of $\mathbb{R}^n$
- Du: gradient vector of the solution u with respect to spatial variables x
- D^2u: Hessian matrix of the solution u with respect to spatial variables x

In one dimension:

$$\begin{aligned} \partial_t u(t,x) + H(t,x,u(t,x),\partial_x u(t,x),\partial_x^2 u(t,x)) &= 0, \quad x \in \Omega, \quad t \in [0,T), \\ u(T,x) &= g(x) \end{aligned}$$

No analytical solution. Need for numerical approximation:

- Deterministic methods: finite difference schemes, finite elements
- Probabilistic methods (based on simulation): Monte Carlo

Finite difference schemes

- Localize the problem on a bounded domain
- Discretize the solution u on a space–time grid $\mathcal{G}_h = \Delta t\{0, 1, \ldots, n_T\} \times \Delta x \mathbb{Z}^n$ where $h = (\Delta t, \Delta x)$ and $n_T \Delta t = T$:

$$u_i^k = u(t_k, x_i), \qquad \forall (t_k, x_i) \in \mathcal{G}_h$$

- Replace the differentiation operators ∂_t, D, and D^2 by their finite difference approximations. In one dimension, one can use

$$\begin{aligned}
\partial_t^{\mathrm{fwd}} u(t_k, x_i) &= \frac{u_i^{k+1} - u_i^k}{\Delta t} \\
D^{\mathrm{fwd}} u(t_k, x_i) &= \frac{u_{i+1}^k - u_i^k}{\Delta x} \\
D^{\mathrm{bwd}} u(t_k, x_i) &= \frac{u_i^k - u_{i-1}^k}{\Delta x} \\
D^{\mathrm{cen}} u(t_k, x_i) &= \frac{u_{i+1}^k - u_{i-1}^k}{2\Delta x} \\
(D^2)^{\mathrm{cen}} u(t_k, x_i) &= \frac{u_{i+1}^k + u_{i-1}^k - 2u_i^k}{\Delta x^2}
\end{aligned}$$

Finite difference schemes

- PDE (1) can then be approximated by a (nonlinear) algebraic equation with unknowns u_i^k.
- By using centered discrete differentiation operators, an explicit scheme that uses centered finite differences reads

$$\frac{u_i^{k+1}-u_i^k}{\Delta t}+H(t_{k+1},x_i,u_i^{k+1},D^{\mathrm{cen}}u(t_{k+1},x_i),(D^2)^{\mathrm{cen}}u(t_{k+1},x_i))=0, \qquad \forall (t_{k+1},x_i)\in\mathcal{G}_h$$

- The implicit scheme reads

$$\frac{u_i^{k+1}-u_i^k}{\Delta t}+H(t_k,x_i,u_i^k,D^{\mathrm{cen}}u(t_k,x_i),(D^2)^{\mathrm{cen}}u(t_k,x_i))=0, \qquad \forall (t_k,x_i)\in\mathcal{G}_h\backslash\{t=T\}$$

- A θ-scheme reads

$$\frac{u_i^{k+1}-u_i^k}{\Delta t}+\theta H(t_{k+1},x_i,u_i^{k+1},D^{\mathrm{cen}}u(t_{k+1},x_i),(D^2)^{\mathrm{cen}}u(t_{k+1},x_i)) + (1-\theta)H(t_k,x_i,u_i^k,D^{\mathrm{cen}}u(t_k,x_i),(D^2)^{\mathrm{cen}}u(t_k,x_i))=0$$

Monte Carlo methods

- Curse of dimensionality: Finite diff schemes only work in small dimension
- No more than 3 space variables (assets and path-dependent variables)
- Need to turn to simulation-based methods (Monte Carlo methods)

Problem	Numerical tool
American options	ML techniques, duality
Uncertain Lapse and Mortality	ML techniques, BSDEs
Uncertain Volatility	Parameterization, ML techniques, BSDEs
Different rates for borrowing/lending	ML techniques, BSDEs
Calibration of models to market smiles	Particle method, ML techniques, Malliavin

ML techniques will mostly be used to estimate conditional expectations $\mathbb{E}[Y|X=x]$. They include: nearest neighbors, kernel regression, parametric regression, neural networks, random forests, kernel trick... In this course we will apply them to simulated data, but they can be readily applied to data science problems.

Course agenda

1. Week 1: American options
2. Week 2: Optimal control
3. Week 3: Backward Stochastic Differential Equations (BSDEs)
4. Week 4: Uncertain Default Rate Model
5. Week 5: Uncertain Volatility Model
6. Week 6: Portfolio Optimization
7. Week 7: Nonlinear SDEs and calibration of models to market smiles

Models of financial markets

- A filtered probability space $\left(\Omega, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P}^{\mathrm{hist}}\right)$
- $\mathbb{P}^{\mathrm{hist}}$ is the historical or real probability measure under which we model our market.
- A market model is defined by an n-dimensional stochastic differential equation (SDE)

$$dX_t^i = b_i(t, X_t)\,dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t)\,dW_t^j, \qquad i \in \{1, \dots, n\} \tag{2}$$

and by another positive stochastic process B_t, called the money-market account, representing the value of cash, which satisfies

$$dB_t = r_t B_t\,dt, \qquad B_0 = 1$$

i.e.,

$$B_t = \exp\left(\int_0^t r_s\,ds\right)$$

- r_t is the short term interest rate.

Models of financial markets

- We set

$$D_{tu} := B_t B_u^{-1} = \exp\left(-\int_t^u r_s\, ds\right)$$

  which is the discount factor from date u to date t.
- Throughout the course, we will denote by $\tilde{Y}_t := D_{0t} Y_t$ the discounted value of any price process Y_t.
- Certain market components X^i may not be sold or bought in the market, such as the short term interest rate, or a stochastic volatility. Throughout this course, a market component X^i that can be sold and bought in the market is called an "asset."

Early Exercise Problems

Pricing and hedging of American options

If you want a happy ending, that depends, of course, on where you stop your story. — Orson Welles

European, American and Bermudan options

- A *European* option can be exercised only at the expiration date T,

$$V_t^E = \mathbb{E}^{\mathbb{Q}}\left[D_{t,T} F_T | \mathcal{F}_t\right]$$

- An *American* option may be exercised anytime on or before the expiration date T,

$$V_t^A = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau} F_\tau | \mathcal{F}_t\right]$$

  conditional on the option not being exercised at or before t, where $\mathcal{T}_{t,T}$ denotes the set of all stopping times τ such that $\tau \in [t, T]$.

- A *Bermudan* option can be exercised on certain pre-defined discrete set of dates $t_1 < \cdots < t_N = T$,

$$V_t^B = \sup_{\tau \in \mathcal{T}_D} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau} F_\tau | \mathcal{F}_t\right]$$

  conditional on the option not being exercised at or before t, where $\mathcal{T}_D$ denotes the set of all stopping times $\tau \in \{t_1, \cdots, t_N\} \bigcap [t, T]$.

$$V_t^A \geq V_t^B \geq V_t^E$$

American Call and Put - No dividend

Assume positive interest rates. Martingale condition yields

$$\mathbb{E}^{\mathbb{Q}}\left[D_{t,T}S_T|\mathcal{F}_t\right] = S_t$$

American Call

Early exercise of an American call option leads to

- Loss of the time value of the option (from volatility)
- Loss of the time value of strike (from discounting)

Never optimal to exercise an American call option early.

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[D_{t,T}\left(S_T-K\right)^+\middle|\mathcal{F}_t\right] &\geq \left(\mathbb{E}^{\mathbb{Q}}\left[D_{t,T}S_T-D_{t,T}K|\mathcal{F}_t\right]\right)^+ \\ &= \left(S_t-D_{t,T}K\right)^+ \geq \left(S_t-K\right)^+\end{aligned}$$

American Put

Early exercise of an American put option leads to

- Loss of the time value of the option (from volatility)
- Gain on the time value of strike (from discounting)

Early exercise could be beneficial when the option is deeply in the money (time value of the option is small compared to the gain of the time value of the strike)

American Call and Put - Dividend

Continuous Dividend Yield $\mathbb{E}^{\mathbb{Q}}\left[D_{t,T}S_T|\mathcal{F}_t\right] = e^{-q(T-t)}S_t, \quad q > 0$

- American Call

  Early exercise of an American call option could be beneficial if the dividend yield is large enough compared to the time value of the option (from volatility) and the time value of strike (from discounting) lost

- American Put

  Large dividend yield benefit continuation

Discrete Dividend $S_{t_d-} - S_{t_d+} = D > 0, \quad D = \text{dividend}, \quad t_d = \text{ex-date}$

- American Call

  The optimal exercise time of an American call should be either right before the ex-date or at the final maturity

- American Put

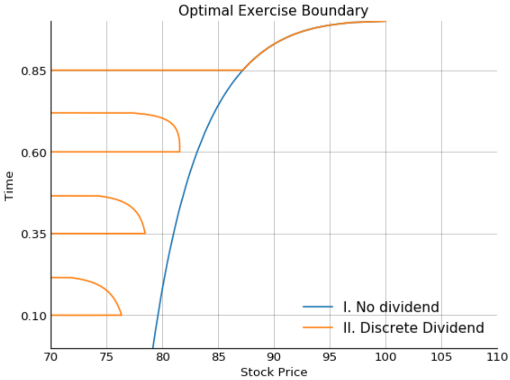
  It is never optimal to exercise an American put right before the ex-date

American Call and Put - Optimal Exercise Boundary

Example. American Put, $S = 100$, $K = 100$, $T = 1$, $\sigma = 25\%$, $r = 8\%$

I. no dividend

II. \$1 cash dividend paid at $t = 0.1, 0.35, 0.6, 0.85$



American Call and Put

- Put-Call parity does **not** hold for American options.
- Most listed options on individual stocks are American options with physical delivery
- Most listed options on stock indices are European options with cash settlement

Markovian Case - European options

Assume that there exists a (deterministic) function r such that $r_t = r(t, X_t)$ and g such that $F_T = g(X_T)$, then

$$u(t,x) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s,X_s)ds} g(X_T) \middle| X_t = x \right]$$

Black-Scholes pricing PDE

$$\partial_t u(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x) = 0, \quad u(T,x) = g(x).$$

where $\mathcal{L}$ is the Ito generator of X:

$$\mathcal{L} = \sum_{i=1}^{n} b_i(t,x)\partial_i + \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{d} \sigma_{i,k}(t,x)\sigma_{j,k}(t,x)\partial_{ij}$$

In dimension one $(n = d = 1)$,

$$\mathcal{L} = b(t,x)\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2}{\partial x^2}$$

Markovian Case - American options

Consider an American option that pays $g(X_t)$ if exercised at time t.

Let $u(t,x)$ be the value of the American option at time t if $X_t = x$ and the option has not been exercised yet. Then

$$u(t,x) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^\tau r(s,X_s)ds} g(X_\tau) \middle| X_t = x \right]$$

where $\mathcal{T}_{t,T}$ is the collection of all stopping times τ such that $t \leq \tau \leq T$.

Since $t \in \mathcal{T}_{t,T}$,

$$u(t,x) \geq g(x).$$

As long as $u(t,X_t) > g(X_t)$, it is not optimal to exercise, and the optimal stopping time τ^* is

$$\tau^* = \inf \left\{ s \geq t | u(s,X_s) = g(X_s) \right\}.$$

Markovian Case - American options - Variational Inequality

Variational Inequality

$$\max\left(\partial_t u(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x),\ g(x) - u(t,x)\right) = 0$$
$$u(T,x) = g(x)$$

Let $\mathcal{J} = \partial_t + \mathcal{L} - r$.

- The option is worth at least its exercise value:

$$g(x) \leq u(t,x)$$

- The supermartingale property of $e^{-\int_0^t r(s,X_s)ds}u(t,X_t)$:

$$\mathcal{J}u(t,x) \leq 0$$

- $\mathcal{J}u = 0$ in the continuation region and $g = u$ in the exercise region:

$$\mathcal{J}u(t,x)\left(g(x) - u(t,x)\right) = 0$$

Markovian Case - American options - Variational Inequality

Formal proof of variational inequality

Assume that u is a smooth solution of the variational inequality

$$\max\left(\partial_t u(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x),\ g(x) - u(t,x)\right) = 0$$

we shall show that u solves the optimal stopping problem, i.e.

$$u(t,X_t) = \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}g(X_t)|\mathcal{F}_t\right].$$

First, $\partial_t u + \mathcal{L}u - ru \leq 0$ implies that $D_{0,t}u(t,X_t)$ is a super-martingale so that for any $\tau \in \mathcal{T}_{t,T}$, by the optional sampling theorem and the fact $g(X_\tau) \leq u(\tau, X_\tau)$, we have

$$\mathbb{E}^{\mathbb{Q}}\left[D_{0,\tau}g\left(X_\tau\right)|\mathcal{F}_t\right] \leq \mathbb{E}^{\mathbb{Q}}\left[D_{0,\tau}u(t,X_\tau)|\mathcal{F}_t\right] \leq D_{0,t}u(t,X_t)$$

Taking the supremum over all $\tau \in \mathcal{T}_{t,T}$, we obtain

$$\sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}g(X_t)|\mathcal{F}_t\right] \leq u(t,X_t).$$

Markovian Case - American options - Variational Inequality

Formal proof of variational inequality (continued)

On the other hand, define $\tau^* = \inf\{s \geq t | u(s, X_s) = g(X_s)\}$. Then

$$D_{0,\tau^*} u(\tau^*, X_{\tau^*}) = D_{0,t} u(t, X_t)$$
$$+ \int_t^{\tau^*} D_{0,s} \left(\partial_t u(s, X_s) + \mathcal{L}u(s, X_s) - r(s, X_s) u(s, X_s)\right) ds + \int_t^{\tau^*} \cdots dW_s$$

The definition of τ^* guarantees that $u(\tau^*, X_{\tau^*}) = g(X_{\tau^*})$ and

$$\partial_t u(s, X_s) + \mathcal{L}u(s, X_s) - r(s, X_s) u(s, X_s) = 0 \quad \text{for all } s \in [t, \tau^*],$$

so that

$$u(t, X_t) = \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau^*} g(X_{\tau^*}) | \mathcal{F}_t\right]$$

i.e. the supremum over $\mathcal{T}_{t,T}$ is attained by the optimal stopping time τ^*.

Markovian Case - Bermudan options

Assume the option can only be exercised on a selected number of dates $t_1, \cdots, t_N$:

$$u(t,x) = \sup_{\tau \in \mathcal{T}_{\{t_1,\dots,t_N\} \cap [t,T]}} \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^\tau r(s,X_s)ds} g(X_\tau) \middle| X_t = x \right]$$

Between two dates t_{i+1} and t_i:

$$\partial_t u(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x) = 0$$

At date t_i (**Jump condition**):

$$u(t_i^-, x) = \max(u(t_i^+, x), g(x))$$

Optimal stopping time

$$\tau^* = \inf\{s \in \{t_1, \dots, t_N\} \cap [t,T] \mid u(s,X_s) = g(X_s)\}$$

Finite Difference Scheme

Finite difference scheme can be easily adapted to price options with early exercises features.

Consider a Bermudan option with exercise dates $t_1 < \cdots < t_N = T$ and payoff function $g(X_t)$

- Let $u(T, x) = g(x)$ and then proceed backward
- Between t_i and t_{i+1}, solve the PDE $\partial_t u + \mathcal{L}u - ru = 0$ by a classic finite difference scheme (fully explicit, fully implicit and Crank-Nicholson)
- Replace $u(t_i^+, x)$ by $u(t_i^-, x) = \max\left(u\left(t_i^+, x\right), g(x)\right)$ and continue the backward induction from t_i with the new terminal condition $u(t_i^-, x)$.

American options can be priced by applying the jump condition to every point in the time grid of the finite difference scheme. Usually only first-order in time.

Bermudan Options and Discrete Time Snell Envelope

Definition (Snell Envelope (Discrete Time))

Given a discrete-time stochastic process $\{Z_k\}_{1 \leq k \leq N}$, the process U_k defined by

$$U_k = \sup_{\tau \in \mathcal{T}_{k,N}} \mathbb{E}\left[Z_\tau | \mathcal{F}_k\right], \quad k = 1, \cdots, N$$

is called the *Snell envelope* of $\{Z_k\}$.

Consider a Bermudan option with exercise dates $t_1 < \cdots < t_N$ and payoffs $F_{t_1}, \cdots, F_{t_N}$. Let V_{t_k} be the value of the option at time t_k if it has not been exercised yet.

$$V_{t_k} = \sup_{\tau \in \mathcal{T}_{\{t_k, \cdots, t_N\}}} \mathbb{E}^{\mathbb{Q}}\left[D_{t_k,\tau} F_\tau | \mathcal{F}_{t_k}\right]$$

$$D_{0,t_k} V_{t_k} = \sup_{\tau \in \mathcal{T}_{\{t_k, \cdots, t_N\}}} \mathbb{E}^{\mathbb{Q}}\left[D_{0,\tau} F_\tau | \mathcal{F}_{t_k}\right].$$

The discounted value process $\{D_{0,t_k} V_{t_k}\}_{k=1}^N$ is the Snell envelope of the discounted payoff $\{D_{0,t_k} F_{t_k}\}_{k=1}^N$.

Recursive Scheme of Building Discrete Time Snell Envelope

The Snell envelope $\{U_k\}$ can be constructed explicitly by the following recursive scheme:

$$U_k = \sup_{\tau \in \mathcal{T}_{k,N}} \mathbb{E}\left[Z_\tau | \mathcal{F}_k\right], \quad k = 0, \cdots, N \tag{3}$$

if and only if

$$\begin{cases} U_N = Z_N \\ U_k = \max\{Z_k, \mathbb{E}\left[U_{k+1} | \mathcal{F}_k\right]\} \quad k = 0, \cdots, N-1 \end{cases} \tag{4}$$

Proof. First assume U_k is given by (3). Let $\tau^*_{k+1} = \text{argmax}_{\tau \in \mathcal{T}_{k+1,N}} \mathbb{E}\left[Z_\tau | \mathcal{F}_{k+1}\right]$ be the optimal stopping time. Then

$$\begin{aligned} \mathbb{E}\left[U_{k+1} | \mathcal{F}_k\right] &= \mathbb{E}\left[\mathbb{E}\left[Z_{\tau^*_{k+1}} \middle| \mathcal{F}_{k+1}\right] \middle| \mathcal{F}_k\right] \\ &= \mathbb{E}\left[Z_{\tau^*_{k+1}} \middle| \mathcal{F}_k\right] \leq \sup_{\tau \in \mathcal{T}_{k,N}} \mathbb{E}\left[Z_\tau | \mathcal{F}_k\right] = U_k \end{aligned}$$

The second-to-last inequality holds since $\tau^*_{k+1} \in \mathcal{T}_{k,N}$. Obviously, $Z_k \leq U_k$ since $k \in \mathcal{T}_{k,N}$. Hence $\max\{Z_k, \mathbb{E}\left[U_{k+1} | \mathcal{F}_k\right]\} \leq U_k$.

On the other hand, for any $\tau \in \mathcal{T}_{k,N}$, $Z_\tau = Z_k \mathbf{1}_{\{\tau = k\}} + Z_{\tau \vee (k+1)} \mathbf{1}_{\{\tau > k\}}$.

Recursive Scheme of Building Discrete Time Snell Envelope

$$
\begin{aligned}
\mathbb{E}\left[Z_\tau|\mathcal{F}_k\right] &= Z_k\mathbf{1}_{\{\tau=k\}} + \mathbb{E}\left[Z_{\tau\vee(k+1)}|\mathcal{F}_k\right]\mathbf{1}_{\{\tau>k\}} \\
&= Z_k\mathbf{1}_{\{\tau=k\}} + \mathbb{E}\left[\mathbb{E}\left[Z_{\tau\vee(k+1)}|\mathcal{F}_{k+1}\right]|\mathcal{F}_k\right]\mathbf{1}_{\{\tau>k\}} \\
&\leq Z_k\mathbf{1}_{\{\tau=k\}} + \mathbb{E}\left[U_{k+1}|\mathcal{F}_k\right]\mathbf{1}_{\{\tau>k\}} \\
&\leq \max\{Z_k, \mathbb{E}\left[U_{k+1}|\mathcal{F}_k\right]\}
\end{aligned}
$$

Thus

$$U_k = \sup_{\tau\in\mathcal{T}_{k,N}} \mathbb{E}\left[Z_\tau|\mathcal{F}_k\right] \leq \max\{Z_k, \mathbb{E}\left[U_{k+1}|\mathcal{F}_k\right]\}.$$

Next assume that U_k is given by the recursive scheme (4). Clearly U_k is a supermartingale that dominates Z_k. Let $\tau_k^* = \inf\{n \geq k; Z_n = U_n\}$. Then

$$U_{(n+1)\wedge\tau_k^*} - U_{n\wedge\tau_k^*} = 1_{n+1\leq\tau_k^*}\left(U_{n+1} - U_n\right) = 1_{n+1\leq\tau_k^*}\left(U_{n+1} - \mathbb{E}\left[U_{n+1}|\mathcal{F}_n\right]\right)$$

Note that $\{n+1 \leq \tau_k^*\} \in \mathcal{F}_n$. Hence $\mathbb{E}\left[U_{(n+1)\wedge\tau_k^*}\middle|\mathcal{F}_n\right] = U_{n\wedge\tau_k^*}$, i.e. $U_{n\wedge\tau_k^*}$ is a martingale so that $U_k = \mathbb{E}\left[U_{\tau_k^*}\middle|\mathcal{F}_k\right] = \mathbb{E}\left[Z_{\tau_k^*}\middle|\mathcal{F}_k\right]$. Furthermore, for $\forall\tau \in \mathcal{T}_{k,N}$,

$$\mathbb{E}\left[Z_\tau|\mathcal{F}_k\right] \leq \mathbb{E}\left[U_\tau|\mathcal{F}_k\right] \leq U_k.$$

$\square$

Discrete-Time Snell Envelope and Bermudan Options

Corollary. Given $\{Z_n\}_{n\leq N}$, the Snell envelope $\{U_k\}$ is the smallest supermartingale dominating $\{Z_n\}$. For $\forall k \leq N$, the optimal stopping time τ_k^* is given by

$$\tau_k^* = \inf\{n \geq k; Z_n = U_n\}$$

and the stopped process $\left\{U_{\tau_k^* \wedge n}\right\}_{n\geq k}$ is a martingale.

Bermudan Option

Let V_{t_k} be the price of a Bermudan option at time t_k provided the option has not been exercised before t_k. Then $\{D_{0,t_k} V_{t_k}\}$ is the smallest supermartingale that dominates the discounted payoff $\{D_{0,t_k} F_{t_k}\}$ and can be constructed by the recursive scheme:

$$\begin{cases} V_{t_N} & = & F_{t_N} \\ V_{t_i} & = & \max\left(F_{t_i}, \mathbb{E}^{\mathbb{Q}}[D_{t_i t_{i+1}} V_{t_{i+1}} | \mathcal{F}_{t_i}]\right) \end{cases}$$

If the option has not been exercised before t_k, the optimal stopping time τ_k^* is

$$\tau_k^* = \inf\{t_j \geq t_k \,|\, V_{t_j} = F_{t_j}\}$$

The stopped value process $\{D_{0,\tau_k^* \wedge t_j} V_{\tau_k^* \wedge t_j}\}_{j\geq k}$ is a martingale.

Monte Carlo Methods

Why do we need it?

Curse of dimensionality for finite difference methods when the number of variables exceeds three.

- Multi-asset options
- Path-dependent payoff
- Stochastic volatility/stochastic interest rates

Why is it difficult?

Need to estimate the continuation value $\mathbb{E}^{\mathbb{Q}}\left[D_{t_i,t_{i+1}} V_{i+1} \middle| \mathcal{F}_{t_i} \right]$. Naive nested Monte Carlo is explosive.

Backward Induction

First, we approximate American options by Bermudan options with exercise dates $t_1, t_2, \cdots, t_N = T$ and payoff $F_{t_1}, F_{t_2}, \cdots, F_{t_N}$.

Let V_{t_i} be the value of the option at time t_i.

At the final expiry $t_N = T$,

$$V_{t_N} = F_{t_N}$$

Then we proceed by backward induction

$$V_{t_i} = \max\left(\mathbb{E}^{\mathbb{Q}}\left[D_{t_i,t_{i+1}} V_{t_{i+1}} | \mathcal{F}_{t_i}\right], g(X_{t_i})\right), \quad i = 1, \cdots, N-1$$

The key is to determine the continuation value $C_{t_i} = \mathbb{E}^{Q}\left[D_{t_i,t_{i+1}} V_{t_{i+1}} | \mathcal{F}_{t_i}\right]$

Regression algorithms

Continuation value at t_i:

$$C_{t_i} = \mathbb{E}^{\mathbb{Q}}\left[D_{t_i,t_{i+1}} V_{t_{i+1}} | \mathcal{F}_{t_i}\right] \quad \text{(A)}$$

Let τ_{i+1} be the optimal stopping time at t_{i+1}. Then

$$V_{t_{i+1}} = \mathbb{E}^{\mathbb{Q}}\left[D_{t_{i+1},\tau_{i+1}} F_{\tau_{i+1}} | \mathcal{F}_{t_{i+1}}\right]$$

so that

$$C_{t_i} = \mathbb{E}^{\mathbb{Q}}\left[D_{t_i,\tau_{i+1}} F_{\tau_{i+1}} | \mathcal{F}_{t_i}\right] \quad \text{(B)}$$

The conditional expectation can be estimated from the cross-sectional information in the simulated paths.

- Tsitsiklis-Van Roy (2001) uses (A) and regress $\hat{V}(t_{i+1}, X_{t_{i+1}})$ on some functions of X_{t_i}
- Longstaff-Schwartz (2001) uses (B) and regress the sum of subsequent realized cash flows on some functions of X_{t_i} (more robust, more accurate)

The Tsitsiklis-Van Roy algorithm

Bermudan option with exercise dates $t_1 < \cdots < t_N$ and payoff $F_{t_1}, \cdots, F_{t_N}$. Let $\hat{V}_{t_i}(\omega)$ denote the approximation to $V_{t_i}(\omega)$ on each path.

Procedure

- Simulate paths until maturity $T = t_N$
- Let $\hat{V}_{t_N} = F_{t_N}$
- For $i = N-1, N-2, \ldots, 1$, estimate the continuation value

$$\hat{C}_{t_i} = \mathbb{E}^{\mathbb{Q}}[D_{t_i t_{i+1}} \hat{V}_{t_{i+1}} | \mathcal{F}_{t_i}]$$

  and then let

$$\hat{V}_{t_i} = \max(F_{t_i}, \hat{C}_{t_i})$$

- Estimate V_0 by

$$\hat{V}_0 = \mathbb{E}^{\mathbb{Q}}[D_{0 t_1} \hat{V}_{t_1}]$$

  where $\mathbb{E}^{\mathbb{Q}}$ is replaced by the average value over all simulated paths

The Longstaff-Schwartz algorithm

Procedure

- Simulate paths until maturity $T = t_N$
- Let $\hat{\tau}_N = t_N$
- For $i = N-1, \dots, 1$,

$$\hat{C}_{t_i} = \mathbb{E}^{\mathbb{Q}}[D_{t_i \hat{\tau}_{i+1}} F_{\hat{\tau}_{i+1}} | \mathcal{F}_{t_i}]$$

$$\hat{\tau}_i = \begin{cases} t_i & \text{if} \quad \hat{C}_{t_i} \leq F_{t_i} \\ \hat{\tau}_{i+1} & \text{otherwise} \end{cases}$$

- Eventually, estimate the optimal strategy by $\hat{\tau} = \hat{\tau}_1$ and the price V_0 by

$$\hat{V}_0 = \mathbb{E}^{\mathbb{Q}}[D_{0\hat{\tau}} F_{\hat{\tau}}]$$

where $\mathbb{E}^{\mathbb{Q}}$ is replaced by the average value over all simulated paths

Tsitsiklis-Van Roy v.s. Longstaff-Schwartz

Both Tsitsiklis-Van Roy (TVR) and Longstaff-Schwartz (LS) estimate the continuation value from the cross-sectional information in the simulated paths by regression.

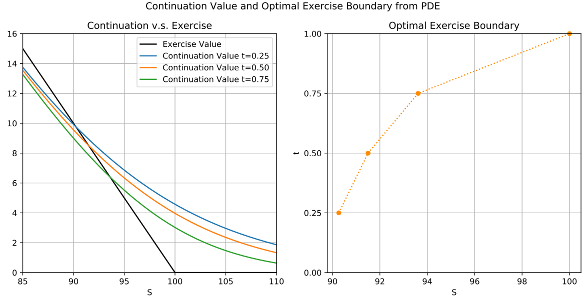
- TVR uses the fitted values from the previous step in the regression at current step whereas LS uses subsequent realized cash flows.
- The continuation estimate $\hat{C}_{t_i}$ is used directly in building value function V_{t_i} in TVR, but only used in making exercise decision in LS.
- The regression error from estimating $\hat{C}_{t_i}$ leads to error in $\hat{V}_{t_i}$ in TVR, but error in exercise decision in LS. In both algorithms the errors propagate through the backward induction. In general the error propagation is worse in TVR.
- In LS, one only needs to estimate the continuation value where option is in the money.
- In general, the LS algorithm is more accurate and robust than the TVR algorithm.

Tsitsiklis-Van Roy v.s. Longstaff-Schwartz - Example

Example: Bermudan put option, quarterly exercise, Black-Scholes model,

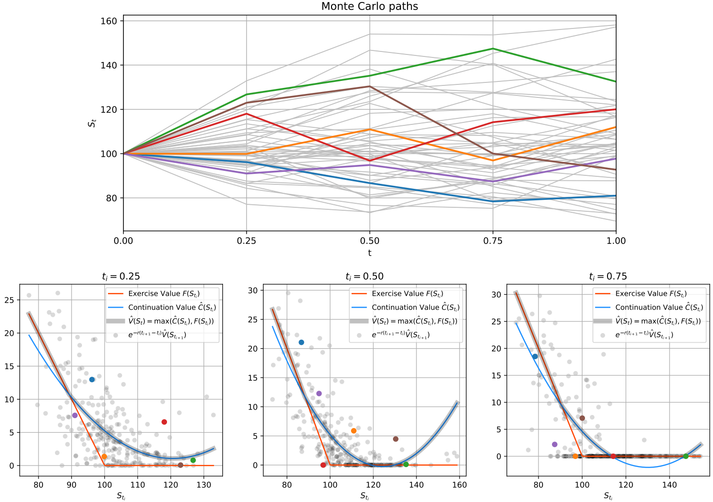
$$S_0 = 100, \sigma = 0.2, r = 0.1, q = 0.02, K = 100, T = 1. \quad F(S) = (K - S)^+.$$

PDE:

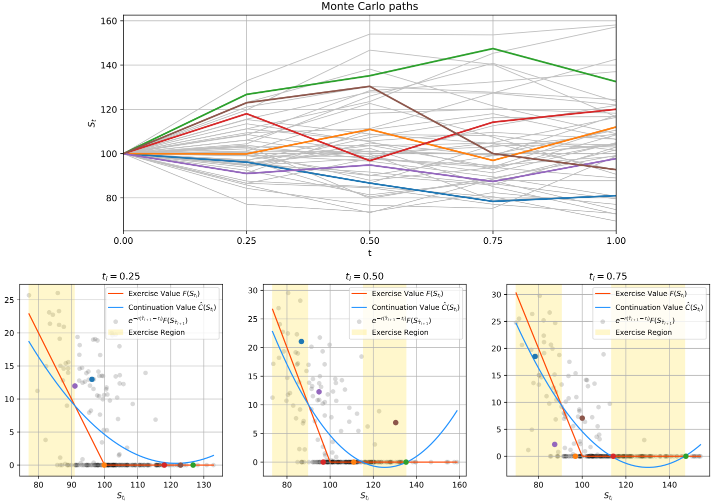


Monte Carlo: estimate of continuation value is obtained by quadratic polynomial fit

Tsitsiklis-Van Roy - Example

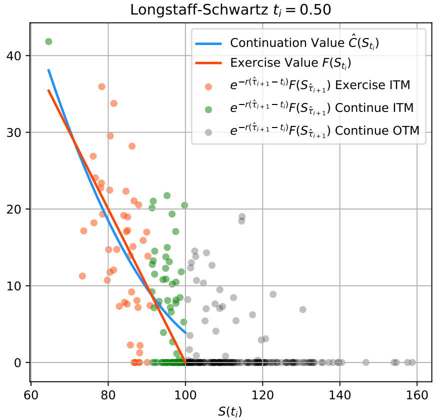


Longstaff-Schwartz - Example



Longstaff-Schwartz - Example

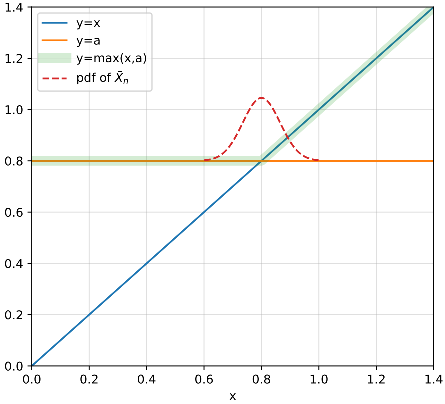
Continuation values can be estimated only where the option is in the money.



Convexity of $x \mapsto \max(x, a)$ and Jensen's Bias

Example. Let $X_1, \ldots, X_n$ be independent samples of a random variable X and $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ be the sample mean. Assume that $\mathbb{E}[X] = a = 0.8$. Then

$$\mathbb{E}\left[\max\left(a, \bar{X}_n\right)\right] \geq \max(a, \mathbb{E}[X]) = a.$$



Bias and Lower Bound Pricing

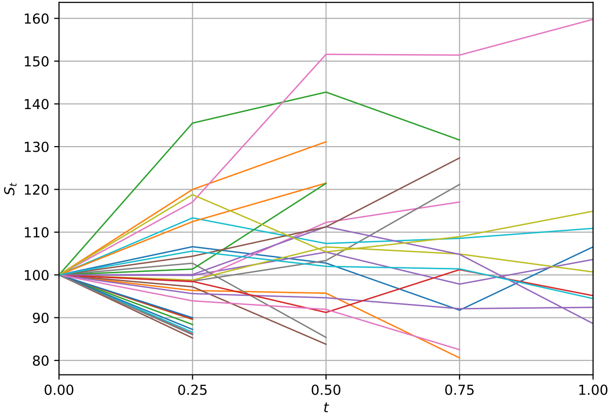
- There are both bias and variance from the estimation of the continuation value $\hat{C}_{t_i}$, which lead to bias in the estimation of the final price $\hat{V}_0$.
- The variance in $\hat{C}_{t_i}$ creates high bias in $\hat{V}_0$ due to Jensen's bias. The Jensen's bias diminishes as the number of paths goes to infinity.
- The bias in $\hat{C}_{t_i}$ typically leads to the bias of the same direction in $\hat{V}_0$ in the TVR algorithm.
- The bias in $\hat{C}_{t_i}$ leads to a sub-optimal strategy, which produces low bias in $\hat{V}_0$ in the Longstaff-Schwartz algorithm.
- Typically, both Tsitsiklis-Van Roy and Longstaff-Schwartz algorithms have undetermined bias.
- *Lower Bound (or Low-biased) pricing* requires a two-step procedure.
  1. Run TVR or Longstaff-Schwartz algorithm with p_1 paths and obtain an estimate of the optimal strategy $\hat{\tau}_1$:

$$\hat{\tau}_1 = \inf\{t_j | \hat{C}_{t_j} \leq F_{t_j}\}$$

  2. Simulate p_2 new independent paths that are stopped according to $\hat{\tau}_1$.

Lower Bound Pricing - Example

Example: Independent Monte Carlo simulation to get lower bound price with exercise strategy given by the Longstaff-Schwartz algorithm.



Conditional Expectation $\mathbb{E}[Y|X]$

1. (Best Predictor). If $Y \in L^2$, $f^*(X) \triangleq \mathbb{E}[Y|X]$ minimizes the mean squared error $\mathbb{E}\left[(Y - f(X))^2\right]$ among all f s.t. $f(X) \in L^2$. Therefore $f^*(X)$ can be interpreted as a projection of Y on the space of all functions of X and is the best predictor of Y (with the smallest mean squared error) given the value of X.

2. If both X and Y are continuous random variables, then

$$\mathbb{E}[Y|X = x] = \int y f_{Y|X}(y|x) dy$$

where the conditional density $f_{Y|X}$ is the ratio of the joint density $f_{X,Y}$ and marginal density f_X,

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Conditional Expectation Estimation - Parametric Regression

We approximate $\mathbb{E}[Y|X]$ by restricting the subspace of functions of X to a smaller subspace of functions of some parametric form $\{f(x;\alpha), \alpha \in \mathcal{A}\}$:

$$\alpha^* = \text{argmin}_{\alpha \in \mathcal{A}} \mathbb{E}\left[(Y - f(X;\alpha))^2\right] \quad \text{and} \quad \mathbb{E}[Y|X] \approx f(X;\alpha^*)$$

In particular, we may choose $f(x;\alpha) = \sum_{k=1}^{n} \alpha_k f_k(x)$ to be the linear sum of basis functions $\{f_k\}_{k=1}^{n}$, i.e.

$$\mathbb{E}[Y|X] \approx \sum_{k=1}^{n} \alpha_k f_k(X)$$

Given N observations $((x_1, y_1), \cdots, (x_N, y_N))$ of X and Y, finding $\mathbb{E}[Y|X]$ comes down to solving least square problem $\min_{\alpha \in \mathbb{R}^n} \|A\alpha - y\|_2$, where

$$A = \begin{bmatrix} f_1(x_1) & \cdots & f_n(x_1) \\ \vdots & \ddots & \vdots \\ f_1(x_N) & \cdots & f_n(x_N) \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}$$

$$\alpha^* = \left(A^T A\right)^{-1} A^T y.$$

Conditional Expectation Estimation - Parametric Regression

- Select simple regressors and avoid overfitting problem.
- Select regressors similar to the payoff function.
- Piecewise polynomials can be used as basis functions, e.g., piecewise-linear functions, regression splines.
- Use `numpy.polyfit(x, y)` for polynomial regression and `numpy.linalg.lstsq(A, b)` for general parametric regression.
- Avoid overfitting with regularization (Ridge Regression or Tikhonov regularization):

$$\min_{\alpha \in \mathbb{R}^n} \|A\alpha - y\|_2^2 + \lambda \|\alpha\|_2^2, \quad \lambda > 0$$

  - λ is a hyperparameter (remaining fixed during optimization/training). The optimal solution α^* is given by

$$\alpha^* = \left(A^T A + \lambda I\right)^{-1} A^T y.$$

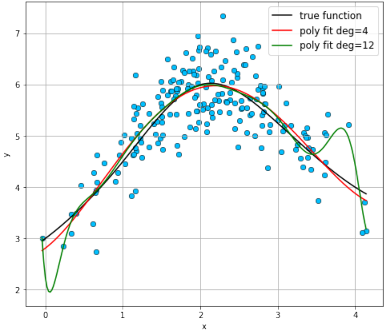
  - Normalize variables beforehand if necessary
  - Penalty can be applied to selected variables
  - Use `sklearn.linear_model.Ridge` for ridge regression

Conditional Expectation Estimation - Polynomial Regression

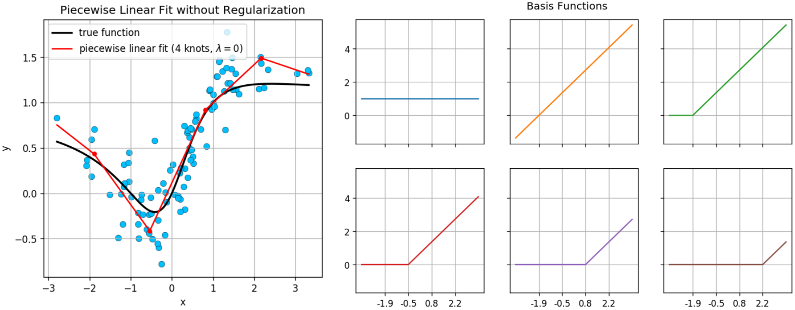
Polynomial basis:

$$f_1(x) = 1, \quad f_2(x) = x, \quad f_3(x) = x^2, \ \cdots, \ f_{n+1}(x) = x^n$$

Polynomial regression with higher order polynomials are more likely to overfit; its behavior near the boundaries tends to be erratic and extrapolation can be dangerous.



Conditional Expectation Estimation - Piecewise Linear Fit



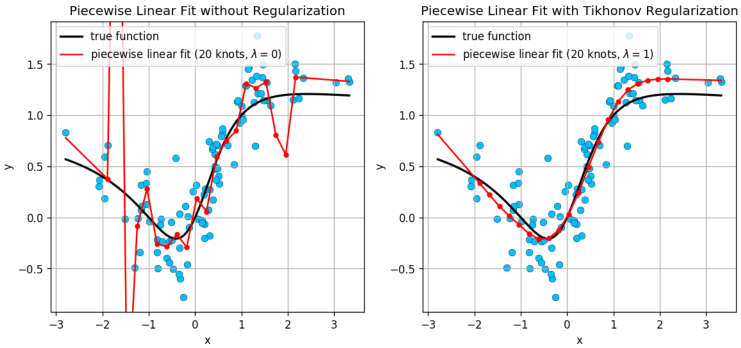
For given knots $x_1 < \cdots < x_n$, the basis functions are

$$f_1(x) = 1, \;\; f_2(x) = x - x_1$$

$$f_{2+i}(x) = (x - x_i)_+, \quad i = 1, \ldots, n$$

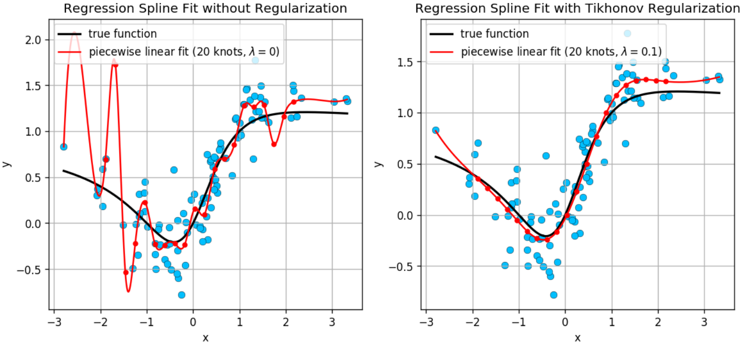
Conditional Expectation Estimation - Piecewise Linear Fit

Tikhonov Regularization



Penalties are applied to α_{2+i}, $i = 1, \ldots, n$.

Conditional Expectation Estimation - Regression Cubic Splines



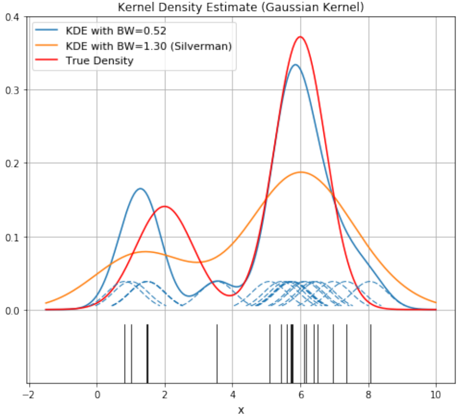
For given knots $x_1 < \cdots < x_n$, the basis functions are

$$f_1(x) = 1, \;\; f_2(x) = x - x_1, \;\; f_3(x) = (x - x_1)^2, \;\; f_4(x) = (x - x_1)^3$$

$$f_{4+i}(x) = (x - x_i)_+^3, \quad i = 1, \cdots, n$$

Penalties are applied to α_{4+i}, $i = 1, \cdots, n$.

Kernel Density Estimation (KDE)



$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad \text{with} \quad K \geq 0 \text{ and } \int K(x)dx = 1$$

Kernel Density Estimation (KDE)

Commonly used kernel functions:

- Epanechnikov $K(x) = \frac{3}{4}\left(1 - x^2\right)$ for $-1 \leq x \leq 1$ and 0 elsewhere;
- Quartic $K(x) = \frac{15}{16}(1+x)^2(1-x)^2$ for $-1 \leq x \leq 1$ and 0 elsewhere;
- Gaussian $K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ for all x.

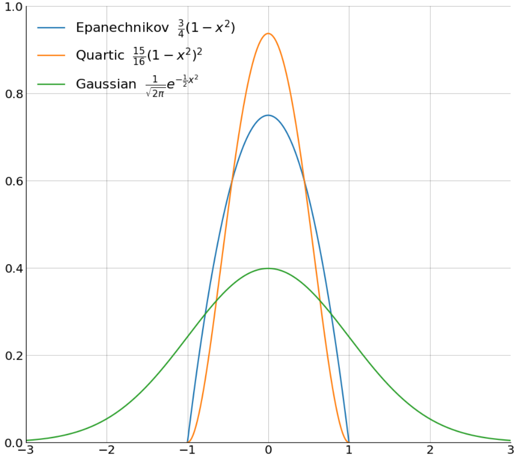
Bandwidth Selection:

Balance the bias-variance tradeoff. Large h gives lower variance but higher bias. Silverman's rule of thumb:

$$h = \left(\frac{4\hat{\sigma}^5}{3N}\right)^{\frac{1}{5}} \approx 1.06\hat{\sigma} N^{-\frac{1}{5}} \quad \text{where } \hat{\sigma}^2 \text{ is the sample variance}$$

Optimal (in integrated mean square error terms) if Gaussian kernel is used to estimate Gaussian density.

Kernel Density Estimation (KDE) - Commonly Used Kernels



Conditional Expectation Estimation - Nonparametric Regression

Nadaraya-Watson Kernel Regression (Locally weighted average)

$$\mathbb{E}[Y|X=x] \approx \frac{\sum_{i=1}^{N} K_h(x-x_i) y_i}{\sum_{i=1}^{N} K_h(x-x_i)}$$

where K is the kernel, $h>0$ is the bandwidth and $K_h(x) = K(x/h)/h$.

Derivation.

$$\mathbb{E}\left[Y|X=x\right] = \int y \frac{f_{X,Y}(x,y)}{f_X(x)} dy$$

Use kernel density estimation for f_X and $f_{X,Y}$ with kernel K_h:

$$f_X(x) \approx \frac{1}{n}\sum_{i=1}^{n} K_h(x-x_i), \quad f_{X,Y}(x,y) \approx \frac{1}{n}\sum_{i=1}^{n} K_h(x-x_i) K_h(y-y_i)$$

and (assuming that K is centered: $\int y K_h(y) dy = 0$) the fact that

$$\int y K_h(y-y_i) dy = y_i.$$

□

Conditional Expectation Estimation - Nonparametric Regression

Local Linear Regression

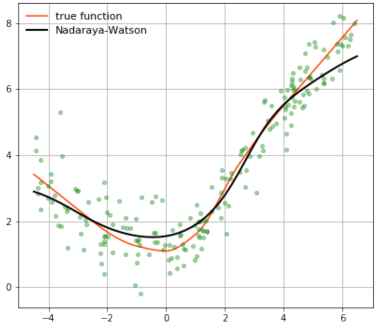
The locally weighted linear regression solves a separate weighted least squares problem at each target point x,

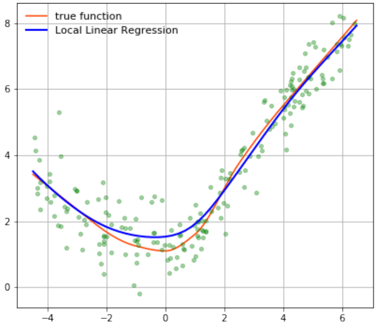
$$\min_{\alpha,\beta} \sum_{i=1}^{N} K_h(x - x_i) \left[y_i - \alpha - \beta x_i\right]^2$$

Then the estimate is $\hat{\alpha} + \hat{\beta}x$, where $\hat{\alpha}$ and $\hat{\beta}$ are dependent on x.

- The locally-weighted averages can be badly biased on the boundaries. This bias can be removed by local linear regression to first order.
- Cross validation can be used for bandwidth selection.

NW v.s. Local Linear Regression



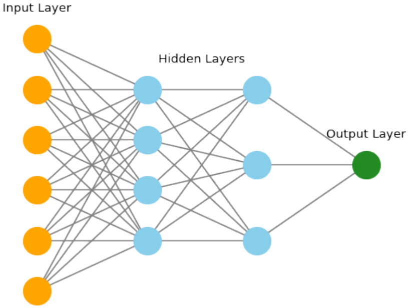


Artificial Neutral Networks

Feedforward Neutral Network

Inputs:	$a_0 = x \in \mathbb{R}^p$
Hidden layers	$a_i = f_i\left(W_i a_{i-1} + b_i\right), \quad i = 1, \cdots, N-1$
Outputs	$y = a_N = W_N a_{N-1} + b_N \in \mathbb{R}^q$

The nonlinear function f_i is called activation function. Some common choices are sigmoid, hyperbolic tangent, ReLu and ELU functions.



Universal Approximation Theorem

Theorem

Let ϕ be a non-constant bounded and monotone-increasing continuous function on $\mathbb{R}$. Given any continuous function f on $[0,1]^n$ and $\epsilon > 0$, there exists an integer m and real constants β_i, b_i and w_{ij}, with $i = 1, \ldots, m$ and $j = 1, \ldots, n$, such that the function F defined by

$$F(x_1, \ldots, x_n) = \sum_{i=1}^{m} \beta_i \phi \left(\sum_{j=1}^{n} w_{ij} x_j + b_i \right)$$

is a uniform ϵ-approximation to f:

$$|F(x_1, \ldots, x_n) - f(x_1, \ldots, x_n)| < \epsilon \quad \text{for all } x_1, \ldots, x_n.$$

- Any continuous function can be approximated to arbitrary accuracy (on a compact set) by a neural network with a single hidden layer, provided the hidden layer is sufficiently wide (m large enough).

Artificial Neutral Networks

- A neural network *learns* the basis functions instead of using fixed ones.
- Much of the power of neural networks comes from the repeated composition of simpler functions to represent a more complex function.
- A loss function needs to be specified to train the model. For regression tasks, a typical choice is the mean squared error.
- Gradient with respect to the network parameters can be computed efficiently by *backpropagation*. Then the network parameters can be updated by a gradient-based optimizer such as stochastic gradient descent.
- A neural network typically has a huge number of parameters. Regularization techniques are helpful to avoid overfitting.
- The network parameters are initialized randomly (with variances properly chosen) to break the symmetry and avoid the exploding/vanishing gradient problems.
- It is also important to fine-tune the hyperparameters.
- Popular deep learning libraries include TensorFlow, CNTK, Theano and PyTorch. Keras is a high-level API with multibackend support.

Primal and Dual Problems

Primal Problem (maximizing over stopping times)

$$\sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau} F_\tau | \mathcal{F}_t\right] \tag{5}$$

Dual Problem (minimizing over martingales) [Rogers 2002; Haugh & Kogan 2004]

$$\inf_{M \in \mathcal{M}_{t,0}} \mathbb{E}^{\mathbb{Q}}\left[\sup_{t \leq s \leq T} \left(D_{t,s} F_s - M_s\right) \middle| \mathcal{F}_t\right]. \tag{6}$$

where $\mathcal{M}_{t,0}$ denotes the set of all right-continuous martingales $(M_s)_{s\in[t,T]}$ with $M_t = 0$. Let $(U_s)_{s\in[t,T]}$ be the Snell envelope of discounted payoff $(D_{t,s}F_s)_{s\in[t,T]}$. Then the optimal martingale M^* is the martingale part of the Doob-Meyer decomposition of $(U_s)_{s\in[t,T]}$, minus its value at time t.

$$\boxed{\sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau} F_\tau | \mathcal{F}_t\right] = \inf_{M \in \mathcal{M}_{t,0}} \mathbb{E}^{\mathbb{Q}}\left[\sup_{t \leq s \leq T} \left(D_{t,s} F_s - M_s\right) \middle| \mathcal{F}_t\right]}$$

Primal and Dual Problems

Proof. For any $\tau \in \mathcal{T}_{t,T}$ and $M \in \mathcal{M}_{t,0}$, it follows from the optional sampling theorem that

$$\mathbb{E}^{\mathbb{Q}}\left[M_\tau | \mathcal{F}_t\right] = M_t = 0$$

Therefore

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau | \mathcal{F}_t\right] &= \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau - M_\tau | \mathcal{F}_t\right] + \mathbb{E}^{\mathbb{Q}}\left[M_\tau | \mathcal{F}_t\right] \\ &\leq \mathbb{E}^{\mathbb{Q}}\left[\sup_{t \leq s \leq T}\left(D_{t,s}F_s - M_s\right) \middle| \mathcal{F}_t\right]\end{aligned}$$

so that

$$\sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau | \mathcal{F}_t\right] \leq \inf_{M \in \mathcal{M}_{t,0}} \mathbb{E}^{\mathbb{Q}}\left[\sup_{t \leq s \leq T}\left(D_{t,s}F_s - M_s\right) \middle| \mathcal{F}_t\right]. \qquad (7)$$

On the other hand, for any fixed t, let $U_s = \sup_{\tau \in \mathcal{T}_{s,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau | \mathcal{F}_s\right]$, $s \in [t,T]$, be the Snell envelope of the discounted payoff $\left(D_{t,s}F_s\right)_{s \in [t,T]}$. Then $\left(U_s\right)_{s \in [t,T]}$ is the smallest supermartingale that dominates $\left(D_{t,s}F_s\right)_{s \in [t,T]}$. In particular, it has unique Doob-Meyer decomposition

$$U_s = M_s - A_s, \quad \forall s \in [t,T]$$

where $\left(M_s\right)_{s \in [t,T]}$ is a martingale with $M_t = U_t$ and $\left(A_s\right)_{s \in [t,T]}$ a predictable increasing process with $A_t = 0$.

Primal and Dual Problems

Let $M_s^* = M_s - M_t$, $s \in [t,T]$. Then

$$D_{t,s}F_s - M_s^* = D_{t,s}F_s - U_s - A_s + M_t \leq M_t = U_t.$$

Since this holds for all $s \in [t,T]$, we must have

$$\sup_{t \leq s \leq T} \left(D_{t,s}F_s - M_s^*\right) \leq U_t = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau \middle| \mathcal{F}_t\right].$$

Taking conditional expectation $\mathbb{E}^{\mathbb{Q}}\left[\cdot \middle| \mathcal{F}_t\right]$ and using inequality (7) we obtain

$$\mathbb{E}^{\mathbb{Q}}\left[\sup_{t \leq s \leq T} \left(D_{t,s}F_s - M_s^*\right) \middle| \mathcal{F}_t\right] = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau \middle| \mathcal{F}_t\right] \tag{8}$$

which shows that the infimum in (6) is achieved by M^*, thus the equivalence of the primal and dual problems. $\square$

Remark. It is easy to see that (8) holds even without expectation $\mathbb{E}^{\mathbb{Q}}\left[\cdot\right]$ on the left-hand side, i.e.

$$\sup_{t \leq s \leq T} \left(D_{t,s}F_s - M_s^*\right) = \sup_{\tau \in \mathcal{T}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{t,\tau}F_\tau \middle| \mathcal{F}_t\right] \quad \text{almost surely}$$

Lower and Upper bounds

Let u_0 be the price of an American option at time $t = 0$, i.e.,

$$u_0 = \sup_{\tau \in \mathcal{T}_{0,T}} \mathbb{E}^{\mathbb{Q}}\left[D_{0,\tau} F_\tau\right] = \inf_{M \in \mathcal{M}_{0,0}} \mathbb{E}^{\mathbb{Q}}\left[\sup_{0 \le t \le T} \left(D_{0,t} F_t - M_t\right)\right]$$

- Any stopping time (exercise strategy) $\tau \in \mathcal{T}_{0,T}$ gives a lower bound:

$$\mathbb{E}^{\mathbb{Q}}\left[D_{0,\tau} F_\tau\right] \le u_0$$

- Any martingale (self-financing hedge) M with $M_0 = 0$ gives an upper bound:

$$u_0 \le \mathbb{E}^{\mathbb{Q}}\left[\sup_{0 \le t \le T} \left(D_{0,t} F_t - M_t\right)\right].$$

$$\boxed{\text{Primal problem} \Longrightarrow \text{Lower bound,} \quad \text{Dual problem} \Longrightarrow \text{Upper bound}}$$

Upper bound

Suppose an agent sells an American option at time 0 and let M be the value of the self-financing hedge less its initial value, then

$$\mathbb{E}^{\mathbb{Q}}\left[\sup_{0\leq t\leq T}\left(D_{0,t}F_t - M_t\right)\right]$$

gives the maximum exposure from the short option position and its hedge on average.

Example 1. Let $M_t \equiv 0$ (no hedge), i.e.

$$u_0 \leq \mathbb{E}^{\mathbb{Q}}\left[\sup_{0\leq t\leq T}\left(D_{0,t}F_t\right)\right].$$

Example 2. In case of American put option, let M_t be the European put price (discounted to time 0) with the same final maturity and strike less the initial price.

Doob-Meyer Decomposition - Discrete Time

Theorem (Doob-Meyer Decomposition)

Let U_n be an adapted and integrable process. Then there exists a unique decomposition

$$U_n = M_n - A_n,$$

where M_n is a martingale and A_n is predictable process with $A_0 = 0$. In particular, U_n is a supermartingale if and only if A_n is increasing; U_n is a submartingale if and only if A_n is decreasing.

Proof.

Put $M_0 = U_0$ and $A_0 = 0$. For $n \geq 0$, let

$$\begin{aligned} M_{n+1} - M_n &= U_{n+1} - \mathbb{E}\left[U_{n+1}|\mathcal{F}_n\right] \\ A_{n+1} - A_n &= -\mathbb{E}\left[U_{n+1}|\mathcal{F}_n\right] + U_n. \end{aligned}$$

Then obviously M_n is a martingale and $A_n \in \mathcal{F}_{n-1}$. □

Dual Method: Broadie-Andersen algorithm

Consider a Bermudan option that pays F_{t_i} if exercised at t_i, $i = 1, \ldots, N$.

Step 1

First run a primal algorithm with n_1 paths (e.g. the Longstaff-Schwartz algorithm) to obtain a sequence of (nearly) optimal stopping times τ_i, $i = 1, \ldots, N$. Let V be the (discounted) value process stopped by τ_i, i.e.

$$V_{t_i} = \mathbb{E}^{\mathbb{Q}} \left[D_{0,\tau_i} F_{\tau_i} \,\middle|\, \mathcal{F}_{t_i} \right]$$

We shall use V_t as an approximation to the Snell envelope $\mathcal{S}_t$ of $D_{0,t} F_t$ and then extract the martingale part of V_t as an approximation to the optimal martingale M_t^*, which is the martingale part of $\mathcal{S}_t$.

Step 2

Simulate a new set of n_2 independent paths. For each of these paths and at each exercise date t_i, we need to estimate V_{t_i} and $\mathbb{E}^{\mathbb{Q}} \left[V_{t_{i+1}} | \mathcal{F}_{t_i} \right]$.

Dual Method: Broadie-Andersen algorithm

- If τ_i indicates continuation, i.e. $\tau_i > t_i$, then simulate n_c independent subpaths starting from (t_i, X_{t_i}) and stopped by τ_i

$$V_{t_i} = \mathbb{E}^{\mathbb{Q}}\left[V_{t_{i+1}}|\mathcal{F}_{t_i}\right] \approx \frac{1}{n_c}\sum_{j=1}^{n_c} D_{0,\tau_i}^{(j)} F_{\tau_i}^{(j)}$$

- If τ_i indicates exercise, i.e. $\tau_i = t_i$, then $V_{t_i} = D_{0,t_i} F_{t_i}$, but we still need to estimate $\mathbb{E}^{\mathbb{Q}}\left[V_{t_{i+1}}|\mathcal{F}_{t_i}\right]$. Simulate n_e independent subpaths starting from (t_i, X_{t_i}) and stopped by τ_{i+1},

$$\mathbb{E}^{\mathbb{Q}}\left[V_{t_{i+1}}\middle|\mathcal{F}_{t_i}\right] \approx \frac{1}{n_e}\sum_{j=1}^{n_e} D_{0,\tau_{i+1}}^{(j)} F_{\tau_{i+1}}^{(j)}$$

Step 3

Build the martingale M: $M_0 = 0$ and

$$M_{t_{i+1}} = M_{t_i} + V_{t_{i+1}} - \mathbb{E}^{\mathbb{Q}}\left[V_{t_{i+1}}|\mathcal{F}_{t_i}\right]$$

and then obtain the upper bound price

$$\mathbb{E}^{\mathbb{Q}}\left[\max_{1\le i\le N}\left(D_{0,t_i}F_{t_i} - M_{t_i}\right)\right]$$

Dual method: Broadie-Andersen algorithm

- The nested simulations in Step 2 are not recursive, i.e., no further simulations are run on each subpath. The number of simulated paths is at most
$$n_1 + n_2 \times (N-1) \times \max(n_c, n_e)$$
- For the martingale (M_{t_i}) generated from a nearly optimal strategy, $\max_{1\le i\le N}(D_{0,t_i}F_{t_i} - M_{t_i})$ has very small variance. Therefore, in most cases, even small values of n_2 give accurate results.
- A large duality gap is usually an indication of poor exercise rules from the primal algorithm (often caused by regression problems, such as bad selection of basis functions, overfitting, $\cdots$)
- The nested Monte Carlo generates independent sampling error with zero mean. However, applying the $\sup$ operator subsequently produces a high-biased estimator of the upper bound.

Option Pricing in a Nutshell

Option Pricing in a Nutshell

In this chapter:

- We recall classical results on stochastic representations of solutions of linear parabolic PDEs.
- This allows us to revisit key notions of option pricing and set our notations.
- Most of the material presented here is standard and can be found in various classical references. Yet we also highlight several important notions which are not often covered:
  - super-replication
  - pricing in incomplete models

1. The superreplication paradigm

Models of financial markets

- A filtered probability space $\left(\Omega, (\mathcal{F}_t)_{0\le t\le T}, \mathbb{P}^{\text{hist}}\right)$
- $\mathbb{P}^{\text{hist}}$ is the historical or real probability measure under which we model our market.
- A market model is defined by an n-dimensional stochastic differential equation (SDE)

$$dX_t^i = b_i(t, X_t)\,dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t)\,dW_t^j, \qquad i \in \{1,\dots,n\} \tag{9}$$

and by another positive stochastic process B_t, called the money-market account, representing the value of cash, which satisfies

$$dB_t = r_t B_t\,dt, \qquad B_0 = 1$$

i.e.,

$$B_t = \exp\left(\int_0^t r_s\,ds\right)$$

- r_t is the short term interest rate. It is adapted to $\mathcal{F}_t$, which is the (natural) filtration generated by the d-dimensional uncorrelated standard Brownian motion $\{W_t^j\}_{1\le j\le d}$.

Models of financial markets

- In order to ensure that SDE (9) admits a unique strong solution (see e.g., Karatzas), we assume that b and σ satisfy:

  Assum(SDE): The functions b and σ are Lipschitz-continuous in x uniformly in t, and satisfy a linear growth condition: there exists a positive constant C such that for all $t \geq 0$, $x, y \in \mathbb{R}^n$,

$$\begin{aligned} |b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| &\leq C|x-y| \\ |b(t,x)| + |\sigma(t,x)| &\leq C\,(1+|x|) \end{aligned}$$

- We set

$$D_{tu} := B_t B_u^{-1} = \exp\left(-\int_t^u r_s\,ds\right)$$

  which is the discount factor from date u to date t.
- Throughout the course, we will denote by $\tilde{Y}_t := D_{0t} Y_t$ the discounted value of any price process Y_t.
- Certain market components X^i may not be sold or bought in the market, such as the short term interest rate, or a stochastic volatility. Throughout this course, a market component X^i that can be sold and bought in the market is called an "asset."

Self-financing portfolios

- Let us assume that we have a portfolio consisting of m assets, say $X_t^1, \ldots, X_t^m$, and the money-market account B_t. It is convenient to use the notation X^0 for B.
- The portfolio at a time t is composed of Δ_t^i assets X_t^i and Δ_t^0 units of X_t^0 (cash).
- The Δ_t^i's must be $\mathcal{F}_t$-measurable, i.e., we cannot look into the future.
- The portfolio value π_t is

$$\pi_t := \sum_{i=0}^{m} \Delta_t^i X_t^i \tag{10}$$

- As time passes, we can readjust the allocations Δ_t^i, but no cash is ever injected into or removed from the portfolio: between t and $t + dt$, the variation in the portfolio value is only due to the variation of the values of the assets, i.e.,

$$d\pi_t = \sum_{i=0}^{m} \Delta_t^i \, dX_t^i \tag{11}$$

We then speak of a *self-financing portfolio*.

Self-financing portfolios

- In terms of discounted values, this reads

$$d\tilde{\pi}_t = \sum_{i=0}^{m} \Delta_t^i \, d\tilde{X}_t^i = \sum_{i=1}^{m} \Delta_t^i \, d\tilde{X}_t^i \tag{12}$$

because for any price process Y_t, $d\tilde{Y}_t = D_{0t}(dY_t - r_t Y_t \, dt)$, so

$$\tilde{\pi}_t = \pi_0 + \sum_{i=1}^{m} \int_0^t \Delta_s^i \, d\tilde{X}_s^i = \pi_0 + \int_0^t \Delta_s \cdot d\tilde{X}_s \tag{13}$$

where $\cdot$ denotes the usual scalar product in $\mathbb{R}^m$.[1]
- As a technical condition, we need to introduce:

Definition

$(\Delta_t, 0 \leq t \leq T)$ defines an **admissible** portfolio if $\tilde{\pi}_t$ is bounded from below for all t $\mathbb{P}^{\text{hist}}$-a.s., i.e., there exists $M \in \mathbb{R}$ such that

$$\mathbb{P}^{\text{hist}}\left(\forall t \in [0, T], \; \tilde{\pi}_t \geq M\right) = 1$$

[1] The stochastic integral is well-defined with the condition $\int_0^t (\Delta_s^i)^2 \, d\langle X \rangle_s < \infty$ $\mathbb{P}^{\text{hist}}$-a.s. $\tilde{\pi}_t$ is then a local martingale.

Arbitrage and arbitrage-free models

- An arbitrage is a self-financing strategy that is worth zero initially and yields a positive gain without any risk:

Definition

A self-financing admissible portfolio is called an **arbitrage** if the corresponding value process π_t satisfies $\pi_0 = 0$ and

$$\pi_T \geq 0 \quad \mathbb{P}^{\mathrm{hist}} - \mathrm{a.s} \qquad \text{and} \qquad \mathbb{P}^{\mathrm{hist}}(\pi_T > 0) > 0$$

- Arbitrageurs are a special kind of traders. Their role is precisely to detect and take full advantage of arbitrage opportunities as soon as they appear in the market. This impacts market prices: arbitrage opportunities tend to disappear as soon as they arise. Absence of arbitrage opportunities is therefore a natural modeling assumption.

Arbitrage and arbitrage-free models

- The next lemma gives a sufficient condition under which we exclude arbitrage opportunities in our market model:

Lemma (Sufficient condition excluding arbitrage)

Suppose there exists a measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ such that $\mathbb{Q} \sim \mathbb{P}^{\text{hist}}$ and such that, for all **asset** *X^i, the discounted price process $\{\tilde{X}_t^i\}_{t\in[0,T]}$ is a local martingale with respect to $\mathbb{Q}$. Then the market $\{X_t\}_{t\in[0,T]}$ has no arbitrage.*

- $\mathbb{P}$ is equivalent to $\mathbb{Q}$ (denoted by $\mathbb{P} \sim \mathbb{Q}$) if and only if $\forall A \in \mathcal{F}_T$, $\mathbb{P}(A) = 0 \Longleftrightarrow \mathbb{Q}(A) = 0$.
- Throughout this course, unless stated otherwise, martingales and local martingales are considered with respect to the filtration $(\mathcal{F}_t)_{0\le t\le T}$.
- Note that the assumption of this lemma bears only on *assets* X^i only, not on non-tradable components of X, such as instantaneous interest rates, instantaneous stochastic volatility, etc.

Arbitrage and arbitrage-free models

Proof.

Let us suppose that there exists an arbitrage. We have

$$\tilde{\pi}_t = \pi_0 + \sum_{i=1}^{m} \int_0^t \Delta_s^i \, d\tilde{X}_s^i = \pi_0 + \int_0^t \Delta_s \cdot d\tilde{X}_s$$

and, for all asset X^i, the discounted price process $\{\tilde{X}_t^i\}_{t\in[0,T]}$ is a local martingale with respect to $\mathbb{Q}$. Therefore $\tilde{\pi}_t$ is a local martingale w.r.t. $\mathbb{Q}$. Moreover, Δ defines an admissible portfolio, so $\tilde{\pi}_t$ is a lower bounded local martingale w.r.t. $\mathbb{Q}$. By a standard result, $\tilde{\pi}_t$ is thus a $\mathbb{Q}$-supermartingale and hence $\mathbb{E}^{\mathbb{Q}}[\tilde{\pi}_T] \leq \tilde{\pi}_0 = 0$. But since $\tilde{\pi}_T \geq 0$ $\mathbb{P}^{\text{hist}}$-a.s. and $\mathbb{Q} \sim \mathbb{P}^{\text{hist}}$, we have $\tilde{\pi}_T \geq 0$ $\mathbb{Q}$-a.s. Hence $\tilde{\pi}_T = 0$ $\mathbb{Q}$-a.s., so $\tilde{\pi}_T = 0$ $\mathbb{P}^{\text{hist}}$-a.s., which yields $\pi_T = 0$ $\mathbb{P}^{\text{hist}}$-a.s. This contradicts the fact that $\mathbb{P}^{\text{hist}}(\pi_T > 0) > 0$. $\square$

Arbitrage and arbitrage-free models

- This simple lemma invites one to introduce the notion of an equivalent local martingale measure (in short ELMM):

Definition (Equivalent local martingale measure)

Any measure $\mathbb{Q} \sim \mathbb{P}^{\text{hist}}$ on $(\Omega, \mathcal{F}_T)$ such that, for all *asset* X^i, the discounted price process $\{\tilde{X}^i_t\}_{t\in[0,T]}$ is a local martingale w.r.t. $\mathbb{Q}$ is called an equivalent local martingale measure (ELMM). An ELMM is also called a risk-neutral measure.

- The previous lemma now reads: if there exists an ELMM, then the market has no arbitrage opportunities.
- Throughout this course we will always assume that we are in a situation where there exists at least one ELMM.
- Let us consider an *asset* X^i. Under an ELMM $\mathbb{Q}$, $\{\tilde{X}^i_t\}_{t\in[0,T]}$ is an $(\mathcal{F}_t)$-local martingale, hence has zero drift. As a consequence, the drift of the *asset* X^i under an ELMM is $r_t X^i_t$:

$$dX^i_t = \exp\left(\int_0^t r_s\, ds\right)(d\tilde{X}^i_t + r_t \tilde{X}^i_t\, dt) = r_t X^i_t\, dt + (\cdots)\cdot dW_t$$

This is not the case for non-tradable components of the market.

Super-replication

- Let us assume that, at time t, we buy and delta-hedge a European option written on m assets, say $X_t^1, \ldots, X_t^m$, with maturity T and payoff F_T, at the price z.
- In general, the payoff F_T is a function of the paths $(X_t^i, 0 \leq t \leq T)$ followed by the prices of the m assets between times 0 and T.
- The final value of the buyer's portfolio, discounted at time 0, is

$$\begin{aligned}
\tilde{\pi}_T^B &= -D_{0t} z + \sum_{i=1}^{m} \int_t^T \Delta_s^i \, d\tilde{X}_s^i + D_{0T} F_T \\
&= -D_{0t} z + \int_t^T \Delta_s \cdot d\tilde{X}_s + D_{0T} F_T
\end{aligned}$$

We then define the buyer's super-replication price at time t as the greatest price z s.t. the value of the buyer's portfolio $\tilde{\pi}_T^B$ is $\mathbb{P}^{\text{hist}}$-a.s. nonnegative:

Super-replication

Definition (Buyer's price)

$$\mathcal{B}_t(F_T) = \sup \Big\{ z \in \mathcal{F}_t \;\Big|\; \text{there exists an admissible portfolio } \Delta \text{ such that}$$
$$\tilde{\pi}_T^B := -D_{0t} z + \int_t^T \Delta_s \cdot d\tilde{X}_s + D_{0T} F_T \geq 0 \;\; \mathbb{P}^{\text{hist}} - a.s. \Big\} \quad (14)$$

- The price z must be $\mathcal{F}_t$-measurable, denoted by $z \in \mathcal{F}_t$, i.e., we cannot look into the future.
- Similarly, we can define the seller's super-replication price as:

Definition (Seller's price)

$$\mathcal{S}_t(F_T) = \inf \Big\{ z \in \mathcal{F}_t \;\Big|\; \text{there exists an admissible portfolio } \Delta \text{ such that}$$
$$\tilde{\pi}_T^S := D_{0t} z + \int_t^T \Delta_s \cdot d\tilde{X}_s - D_{0T} F_T \geq 0 \;\; \mathbb{P}^{\text{hist}} - a.s. \Big\} \quad (15)$$

Super-replication

Theorem (Arbitrage-free bounds)

Assume that there exists an ELMM $\mathbb{Q} \sim \mathbb{P}^{\text{hist}}$*. Then*

$$\mathcal{B}_t(F_T) \leq \mathbb{E}^{\mathbb{Q}}[D_{tT} F_T | \mathcal{F}_t] \leq \mathcal{S}_t(F_T)$$

Proof.

From the definition of $\mathcal{B}_t(F_T)$, we assume that there exists an admissible portfolio Δ such that

$$-D_{0t} z + \int_t^T \Delta_s \cdot d\tilde{X}_s + D_{0T} F_T \geq 0 \qquad \mathbb{P}^{\text{hist}} - a.s.$$

$\int_0^t \Delta_s \cdot d\tilde{X}_s$ is a lower bounded $\mathbb{Q}$-local martingale, hence a $\mathbb{Q}$-supermartingale, so $\mathbb{E}^{\mathbb{Q}}[\int_t^T \Delta_s \cdot d\tilde{X}_s | \mathcal{F}_t] \leq 0$ and we deduce that

$$D_{0t} z \leq \mathbb{E}^{\mathbb{Q}}[D_{0T} F_T | \mathcal{F}_t] \qquad \text{i.e.,} \qquad z \leq \mathbb{E}^{\mathbb{Q}}[D_{tT} F_T | \mathcal{F}_t]$$

Taking the supremum over z, we get $\mathcal{B}_t(F_T) \leq \mathbb{E}^{\mathbb{Q}}[D_{tT} F_T | \mathcal{F}_t]$. The inequality $\mathbb{E}^{\mathbb{Q}}[D_{tT} F_T | \mathcal{F}_t] \leq \mathcal{S}_t(F_T)$ can be derived similarly. □

Attainable payoff

- These bounds can be sharpened by assuming that the market is complete.
- In order to define what it means for a market to be complete, we need to introduce the notion of attainable payoff:

Definition (Attainable payoff)

A payoff F_T is said to be attainable (at time 0) if there exists an admissible portfolio Δ and a real number z such that $z + \int_0^T \Delta_s \cdot d\tilde{X}_s - D_{0T} F_T = 0$ $\mathbb{P}^{\text{hist}}$-a.s. and $\int_0^t \Delta_s \cdot d\tilde{X}_s$ is a (true) $\mathbb{Q}$-martingale with $\mathbb{Q} \sim \mathbb{P}^{\text{hist}}$.

- Stated otherwise, a payoff F_T is said to be attainable if we can generate the wealth F_T at time T from an initial wealth z by trading only in the underlying assets and the cash in a self-financing way.
- The portfolio Δ and the real number z are unique because if $z' + \int_0^T \Delta'_s \cdot d\tilde{X}_s - D_{0T} F_T = 0$, then

$$\int_0^T (\Delta'_s - \Delta_s) \cdot d\tilde{X}_s = z - z'$$

and since $\tilde{X}$ is a (nontrivial) $\mathbb{Q}$-local martingale, this yields $\Delta'_s = \Delta_s$, hence $z' = z$.

Complete markets

- As a consequence, there is a unique fair price at $t = 0$ for an attainable payoff, and this price is z.
- The assumption that $\int_0^t \Delta_s \cdot d\tilde{X}_s$ is a (true) $\mathbb{Q}$-martingale guarantees that $z = \mathbb{E}^{\mathbb{Q}}[D_{0T} F_T]$, so we have a formula to compute the price z from the payoff F_T.

Definition (Complete market)

A market is said to be complete when every payoff is attainable at time 0.

Complete markets

Theorem

For a complete market, for any ELMM $\mathbb{Q}$ and any payoff F_T, we have

$$\mathcal{B}_t(F_T) = \mathcal{S}_t(F_T) = \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t]$$

Proof.

By completeness, we can find a real number z and an admissible portfolio Δ such that $z + \int_0^T \Delta_s \cdot d\tilde{X}_s - D_{0T}F_T = 0$ $\mathbb{P}^{\text{hist}}$-a.s., hence $\mathbb{Q}$-a.s. As a consequence,

$$D_{0t}z_t + \int_t^T \Delta_s \cdot d\tilde{X}_s - D_{0T}F_T = 0 \qquad \mathbb{Q}\text{−a.s.}$$

with $z_t = D_{0t}^{-1}\left(z + \int_0^t \Delta_s \cdot d\tilde{X}_s\right)$ being $\mathcal{F}_t$-measurable. Since $\int_0^t \Delta_s \cdot d\tilde{X}_s$ is a $\mathbb{Q}$-martingale, we have $D_{0t}z_t = \mathbb{E}^{\mathbb{Q}}[D_{0T}F_T|\mathcal{F}_t]$, i.e., $z_t = \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t]$, whence $\mathcal{B}_t(F_T) \geq \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t]$. Combined with the reverse inequality, we get $\mathcal{B}_t(F_T) = \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t]$. We proceed similarly for the other equality. □

Super-replication in incomplete markets

- With incomplete markets or markets with friction, perfect replication is no longer possible.
- In the theorem below, we state that the buyer's super-replication price is given by the infimum of expectations of the discounted payoff over the set ELMM of equivalent local martingale measures:

Theorem (Super-replication price (El Karoui-Quenez, Kramkov))

The buyer's super-replication price and the seller's super-replication price are given by

$$\begin{aligned}\mathcal{B}_t(F_T) &= \inf_{\mathbb{Q}\in\mathrm{ELMM}} \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t] \\ \mathcal{S}_t(F_T) &= \sup_{\mathbb{Q}\in\mathrm{ELMM}} \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t]\end{aligned}$$

Characterization of complete markets

Corollary (Complete market)

A market is complete if and only if there exists a unique ELMM.

Proof.

Assume the market is complete. Let $\mathbb{Q}^0$ be an ELMM. For any payoff F_T,

$$\mathcal{B}_0(F_T) = \mathcal{S}_0(F_T) = \mathbb{E}^{\mathbb{Q}^0}[D_{0T}F_T] = \inf_{\mathbb{Q}\in\mathrm{ELMM}} \mathbb{E}^{\mathbb{Q}}[D_{0T}F_T] = \sup_{\mathbb{Q}\in\mathrm{ELMM}} \mathbb{E}^{\mathbb{Q}}[D_{0T}F_T].$$

Hence, for all F_T, $\mathbb{E}^{\mathbb{Q}}[D_{0T}F_T] = \mathbb{E}^{\mathbb{Q}^0}[D_{0T}F_T]$ for all $\mathbb{Q} \in \mathrm{ELMM}$, so that $\mathrm{ELMM} = \{\mathbb{Q}^0\}$. Conversely, if $\mathrm{ELMM} = \{\mathbb{Q}^0\}$, then $\mathcal{B}_0(F_T) = \mathcal{S}_0(F_T) = \mathbb{E}^{\mathbb{Q}^0}[D_{0T}F_T]$. As a consequence, there exist two admissible portfolios Δ^B and Δ^S (one for the buyer and one for the seller) such that $\mathbb{P}^{\mathrm{hist}}$-a.s., i.e., $\mathbb{Q}^0$-a.s.

$$\begin{aligned} -\mathbb{E}^{\mathbb{Q}^0}[D_{0T}F_T] + \int_0^T \Delta_t^B \cdot d\tilde{X}_t + D_{0T}F_T &\geq 0 \\ \mathbb{E}^{\mathbb{Q}^0}[D_{0T}F_T] + \int_0^T \Delta_t^S \cdot d\tilde{X}_t - D_{0T}F_T &\geq 0 \end{aligned}$$

Summing those inequalities, this means that $\mathbb{Q}^0$-a.s. $\int_0^T (\Delta_t^B + \Delta_t^S) \cdot d\tilde{X}_t \geq 0$, and, because the discounted prices of all *assets* $(\tilde{X}_t^i)$ are $\mathbb{Q}^0$-local martingales, this yields $\Delta_t^B = -\Delta_t^S$ $\mathbb{Q}^0$-a.s. Eventually, we have that $\mathbb{E}^{\mathbb{Q}^0}[D_{0T}F_T] + \int_0^T \Delta_t^S \cdot d\tilde{X}_t - D_{0T}F_T = 0$ $\mathbb{Q}^0$-a.s., hence $\mathbb{P}^{\mathrm{hist}}$-a.s., and the market is complete. $\square$

Complete models versus incomplete models

- From the proof of the super-replication price theorem, the market model (9) is complete if and only if there exists a unique $\mathcal{F}_t$-adapted vector $\lambda_t \in \mathbb{R}^d$ s.t.

$$\forall i \in \text{asset}, \qquad b_i(t, X_t) - r_t X_t^i = \sum_{j=1}^{d} \sigma_{i,j}(t, X_t) \lambda_t^j \tag{16}$$

- The unique ELMM $\mathbb{Q}$ is then given by

$$\frac{d\mathbb{Q}}{d\mathbb{P}^{\text{hist}}}_{|\mathcal{F}_T} = \prod_{j=1}^{d} e^{-\int_0^T \lambda_t^j \, dW_t^j - \frac{1}{2} \int_0^T \left(\lambda_t^j\right)^2 dt}$$

- The unique arbitrage-free price is

$$\mathcal{B}_t(F_T) = \mathcal{S}_t(F_T) = \mathbb{E}^{\mathbb{Q}}[D_{tT} F_T | \mathcal{F}_t]$$

- An inspection of (16) reveals that if the market is complete, then the rank of $\sigma(t, X_t)$ is equal to d a.s. (which implies that $\#\text{assets} \geq d$). In the case where $\#\text{assets} < d$, the market cannot be complete.
- Moreover, if the number of assets coincides with the number of Brownian motions, i.e., $\#\text{assets} = d$, and $(\sigma_{i,j}(t, X_t))_{i \in \text{asset}, 1 \leq j \leq d}$ is invertible, then the market is complete.

Complete models versus incomplete models

- Examples of complete models that are commonly used by practitioners include
  - Dupire's local volatility model (see Dupire)
  - Libor market models with local volatilities, e.g., BGM with deterministic volatilities (see Brace-Gatarek-Musiela)
  - and Markov functional models (see Hunt-Kennedy-Pelsser).
- Common examples of incomplete models are stochastic volatility models (in short SVMs). Here $\#\text{assets} < d$.
- An example of stochastic volatility model is the Heston model. The dynamics of the underlying, denoted by X_t, reads under a risk-neutral measure $\mathbb{Q}^0 \sim \mathbb{P}^{\text{hist}}$ as

$$\frac{dX_t}{X_t} = r_t\,dt + \sqrt{V_t}\left(\rho\,dW_t^2 + \sqrt{1-\rho^2}\,dW_t^1\right) \tag{17}$$

$$dV_t = -k(V_t - m)\,dt + \sigma\sqrt{V_t}\,dW_t^2 \tag{18}$$

with W_t^1, W_t^2, two uncorrelated standard $\mathbb{Q}^0$-Brownian motions.

Complete models versus incomplete models

- V_t is the instantaneous variance, it is *not* a tradable instrument. The only asset is X_t. Its drift under any ELMM is $r_t X_t$.
- The incompleteness of such a model can be detected as the drift term in V_t can be arbitrarily modified with a change of measure from $\mathbb{Q}^0$ to $\mathbb{Q}^1 \sim \mathbb{P}^{\text{hist}}$:

$$\frac{d\mathbb{Q}^1}{d\mathbb{Q}^0}_{|\mathcal{F}_T} = \prod_{j=1}^{2} e^{-\int_0^T \lambda_t^j \, dW_t^j - \frac{1}{2}\int_0^T \left(\lambda_t^j\right)^2 dt}$$

such that

$$\rho\lambda_t^2 + \sqrt{1-\rho^2}\lambda_t^1 = 0 \tag{19}$$

- From Girsanov's theorem, Equation (19) guarantees that the drift of X_t under $\mathbb{Q}^1$ is still $r_t X_t$, i.e., that $\mathbb{Q}^1$ is still an ELMM.

Complete models versus incomplete models

- We can show that the seller's super-replication (undiscounted) price at $t = 0$ of a European vanilla payoff $F_T := g(X_T)$ in such an SVM is $g^{\text{conc}}(X_0)$ with g^{conc} the concave envelope of g, i.e., the smallest concave function that is greater than or equal to g.
- Similarly, the buyer's super-replication (undiscounted) price at $t = 0$ of a European vanilla payoff $g(X_T)$ in such an SVM is $g^{\text{conv}}(X_0)$ with g^{conv} the convex envelope of g, i.e., the largest convex function that is smaller than or equal to g.
- For example, for a call option $g(x) = (x - K)^+$, $g^{\text{conc}}(x) = x$ and $g^{\text{conv}}(x) = (x - K)^+$.
- This simple example illustrates the main drawback of the super-replication framework: the buyer's and seller's super-replication prices are not close and therefore cannot be used in practice.[2]

[2] This drawback can be circumvented by adding extra market instruments, such as vanilla options, in the superhedging strategy. This leads to tight bounds and a nice generalization of the optimal transport theory (see Beiglböck-Henry-Labordère-Penkner, Galichon-Henry-Labordère-Touzi) where now the transport is performed along a martingale measure.

Pricing in practice

- In practice, the seller's price at time t is computed by picking out a particular ELMM $\mathbb{Q}$:

$$u_t := \mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t] \tag{20}$$

  Under $\mathbb{Q}$, the drift for an asset X^i is fixed to $b_i(t, X_t) = r_t X_t^i$.
- In an incomplete market, $\mathbb{Q}$ does not necessarily achieve the supremum in the super-replication price theorem, and we lose the superhedging strategy paradigm. Selling options becomes a risky business.
- In practice, picking a particular ELMM simplifies a lot the pricing problem: it becomes a *linear* problem: the price of the (European) payoff $F_T^1 + F_T^2$ equals the sum of the prices of the (European) payoffs F_T^1 and F_T^2.

2. Stochastic representation of solutions of linear PDEs

The Cauchy problem

- We will always assume that there exists a (deterministic) function r such that $r_t = r(t, X_t)$. Then $D_{t_1 t_2} = \exp\left(-\int_{t_1}^{t_2} r(s, X_s)\, ds\right)$
- We denote by $\mathcal{L}$ the Itô generator of X defined under a risk-neutral measure $\mathbb{Q}$:

$$\mathcal{L} = \sum_{i=1}^{n} b_i(t,x)\partial_i + \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{d}\sigma_{i,k}(t,x)\sigma_{j,k}(t,x)\partial_{ij} \tag{21}$$

  If X^i is an asset, $b_i(t,x) = r(t,x)x_i$.
- If there exists g such that $F_T = g(X_T)$, we speak of a vanilla option. In such a case, by the Markov property of X,

$$u_t = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s, X_s)\, ds\right) g(X_T) \middle| \mathcal{F}_t\right] =: u(t, X_t)$$

  is a function u of (t, X_t); the discounted price process

$$D_{0t}u_t = D_{0t}\mathbb{E}^{\mathbb{Q}}[D_{tT}F_T|\mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[D_{0T}F_T|\mathcal{F}_t]$$

  is a $\mathbb{Q}$-martingale. Then a straightforward application of Itô's lemma to $D_{0t}u(t,X_t) = \exp\left(-\int_0^t r(s,X_s)\, ds\right) u(t, X_t)$ shows that the pricing function u, if smooth enough, satisfies the second order parabolic linear PDE:

$$\partial_t u(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x) = 0 \tag{22}$$

  with the terminal condition $u(T,x) = g(x)$.

The Cauchy problem

Conversely, a solution u to (22) admits the stochastic representation (20) as stated by the Feynman-Kac theorem:

Theorem (Feynman-Kac, see e.g., Karatzas)

Let b, σ satisfy **Assum(SDE)**, *r be uniformly bounded from below and f have quadratic growth in x uniformly in t. Let $u \in C^{1,2}([0,T) \times \mathbb{R}^d) \cap C^0([0,T] \times \mathbb{R}^d)$ with quadratic growth in x uniformly in t and solution to the parabolic PDE:*

$$\partial_t u(t,x) + \mathcal{L}u(t,x) - r(t,x)u(t,x) + f(t,x) = 0$$

with terminal condition $u(T,x) = g(x)$. Then u admits the stochastic representation

$$u(t,x) = \mathbb{E}^{\mathbb{Q}}\left[\int_t^T e^{-\int_t^s r(u,X_u)du} f(s,X_s)ds + e^{-\int_t^T r(s,X_s)ds} g(X_T) \middle| X_t = x\right]$$

Think of f as intermediate payoffs.

Path-dependent options

- Path-dependent payoffs are those payoffs for which one cannot write $F_T = g(X_T)$, i.e., whose value depends on the asset values not only at maturity T, but also at previous dates $t < T$.
- For instance, with one asset $X_t = X_t^1$:
  - Asian options: $F_T = g\left(\frac{1}{T}\int_0^T X_t\,dt\right)$
  - Barrier and lookback options: $F_T = g\left(\min_{t\in[0,T]} X_t, \max_{t\in[0,T]} X_t, X_T\right)$
  - Cliquet options: $F_T = g\left(X_{T_i}/X_{T_{i-1}}, 1 \le i \le N\right)$
- In some cases, one can write SDEs for the path-dependent variables and add them to the SDEs describing the market X so that one can write $F_T = g(X'_T)$ for an enlarged market X', and the option price satisfies a PDE in the variables x'. For instance, the price of an Asian option satisfies a PDE in the variables (x, a) where $A_t = \int_0^t X_s\,ds$ and $dA_t = X_t\,dt$.

Path-dependent options

- For a general path-dependent payoff $F_T = g(X_{t_1}, \ldots, X_{t_n})$, depending on the spot values at the observation dates $t_1 < \cdots < t_n = T$, the PDE, depending on the past spot values $Z^1, \ldots, Z^n$, can be written as

$$\partial_t u_i(t, x, Z^1, \ldots, Z^{i-1}) + \mathcal{L} u_i(t, x, Z^1, \ldots, Z^{i-1}) = 0, \qquad \forall t \in [t_{i-1}, t_i)$$

with the matching conditions at the dates $t_1, \ldots, t_n$,

$$\begin{aligned} u_n(t_n, x, Z^1, \ldots, Z^{n-1}) &= g(Z^1, \ldots, Z^{n-1}, x) \\ u_i(t_i^-, x, Z^1, \ldots, Z^{i-1}) &= u_{i+1}(t_i^+, x, Z^1, \ldots, Z^{i-1}, x) \end{aligned}$$

- The final price is $u_1(0, X_0)$.

Path-dependent options - An example

Example. Consider a path-dependent payoff at time t_n written as

$$g\left(\frac{X_{t_1} + \cdots + X_{t_n}}{n}\right)$$

Introduce an auxiliary variable I such that the pricing function $u(t,x,I)$ satisfies

$$\partial_t u_i(t,x,I) + \mathcal{L}u(t,x,I) = 0, \qquad t_{i-1} < t < t_i, \quad i = 2,\ldots,n$$

with terminal condition at t_n,

$$u(t_n,x,I) = g\left(I + \frac{x}{n}\right)$$

and jump condition at t_i, $i = 1,\ldots,n-1$,

$$u(t_i^-,x,I) = u\left(t_i^+,x,I + \frac{x}{n}\right).$$

Discretize I-values into m different values, and solve m different one-dimensional PDEs backward in time and exchange information with each other at each t_i by the jump condition. Note that in the first time interval $0 < t < t_1$, the auxillary variable I is not needed and the pricing function u simply satisfies

$$\partial_t u(t,x) + \mathcal{L}u(t,x) = 0, \qquad 0 < t < t_1.$$

Finally the price at time 0 is $u(0,X_0)$.

PDE with default time

- Let us now derive the pricing PDE for an option that delivers $g(\tau, X_\tau)$ at time τ if $\tau < T$, or $g(T, X_T)$ at time T if $T \leq \tau$, where τ is the first time of jump of a Poisson process with deterministic intensity $\lambda(t) \in \mathbb{R}_+$, independent of the filtration $(\mathcal{F}_t)$. Think of τ as a default time.
- Assume that $\tau > t$. Then the price of the option at time t is

$$\begin{aligned}
u(t,x) &= \mathbb{E}\left[\mathbf{1}_{\tau \geq T} e^{-\int_t^T r(s,X_s)ds} g(T,X_T) + \mathbf{1}_{\tau < T} e^{-\int_t^\tau r(s,X_s)ds} g(\tau, X_\tau) \middle| X_t = x\right] \\
&= \mathbb{E}\Big[g(T,X_T) e^{-\int_t^T (r(s,X_s)+\lambda(s))ds} \\
&\qquad + \int_t^T \lambda(s) g(s,X_s) e^{-\int_t^s (r(a,X_a)+\lambda(a))da}\, ds \Big| X_t = x\Big]
\end{aligned}$$

- Feynman-Kac's theorem states that if v is a solution to the PDE

$$\partial_t v + \mathcal{L}v + \lambda(g - v) - rv = 0 \tag{23}$$

with the terminal condition $v(T,x) = g(T,x)$, then $v = u$. Under the usual assumptions guaranteeing the existence and uniqueness of a solution to the above PDE, we deduce that (23) is the pricing PDE for such an option.

Selected references

- Beiglböck, M., Henry-Labordère, P., Penkner, F.: *Model-independent bounds for option prices: A mass transport approach*, Finance and Stoch., 17(3):477–501, 2013.
- Bick, A.: *Quadratic-variation-based dynamic strategies*, Management Science, 41:722–732, 1995.
- Brace, A., Gatarek, D., Musiela, M.: *The market model of interest rate dynamics*, Mathematical Finance, 7(2):127–154, 1997.
- Dupire, B.: *Pricing with a smile*, Risk magazine, 7:18–20, 1994.

Selected references

- El Karoui, N., Quenez, M.-C.: *Dynamic programming and pricing of contingent claims in an incomplete market*, SIAM Journal on Control and Optimization archive, 33(1), 1995.
- Galichon, A., Henry-Labordère, P., Touzi, N.: *A stochastic control approach to no arbitrage bounds given marginals, with an application to lookback options*, to appear in Ann. Appl. Probab., available at http://ssrn.com/abstract=1912477, 2011.
- Hunt, P., Kennedy, J., Pelsser, A.: *Markov-functional interest rate models*, Finance and Stoch., 4(4):391–408, 2000.
- Kramkov, D.: *Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets*, Probab. Theory Related Fields, 105:459–479, 1996.

Stochastic optimal control and the Hamilton-Jacobi-Bellman PDE

Optimal control problems and HJB PDE: summary

Controlled Diffusion:

$$dX_t^{\alpha,i} = b_i(t, X_t^\alpha, \alpha_t)\, dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t^{\alpha,i}, \alpha_t)\, dW_t^j, \qquad i \in \{1, \dots, n\}$$

$$u(t,x) = \sup_{(\alpha_s, s\in[t,T]) \text{ adapted},\ \alpha_s \in \mathcal{A}} \mathbb{E}^{\mathbb{Q}}\left[\int_t^T f(s, X_s^\alpha, \alpha_s) ds + g(X_T^\alpha) \middle| X_t^\alpha = x \right]$$

- For a fixed control α:

$$\begin{aligned} \partial_t u(t,x) + \mathcal{L}^\alpha u(t,x) + f(t,x,\alpha) &= 0 \\ u(T,x) &= g(x) \end{aligned}$$

- Hamilton-Jacobi-Bellman PDE:

$$\begin{aligned} \partial_t u(t,x) + \sup_{\alpha\in\mathcal{A}} \{ f(t,x,\alpha) + \mathcal{L}^\alpha u(t,x) \} &= 0 \\ u(T,x) &= g(x) \end{aligned}$$

- Many nonlinear problems that have recently arisen in quantitative finance are described by HJB PDEs.

Stochastic optimal control problems

- $(\Omega, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{Q})$ denotes a probability space equipped with a d-dimensional Brownian motion W; $(\mathcal{F}_t)$ denotes the natural filtration of W.
- We introduce the following SDE

$$dX_t^{\alpha,i} = b_i(t, X_t^\alpha, \alpha_t)\, dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t^\alpha, \alpha_t)\, dW_t^j, \qquad i \in \{1, \ldots, n\} \quad (24)$$

  where the drift b and the volatility σ depend on t, X_t^α and a control parameter α_t, which is an $\mathcal{F}_t$-adapted process valued in a domain $\mathcal{A}$, not necessarily compact.
- We denote

$$\mathcal{A}_{t,T} = \{(\alpha_s)_{t\leq s\leq T} \text{ adapted} \mid \forall s \in [t, T],\ \alpha_s \in \mathcal{A}\}$$

Stochastic optimal control problems

Controlled diffusion process

$$dX_t^{\alpha,i} = b_i(t, X_t^\alpha, \alpha_t)\,dt + \sum_{j=1}^{d} \sigma_{i,j}(t, X_t^\alpha, \alpha_t)\,dW_t^j, \qquad i \in \{1, \ldots, n\}$$

with Ito generator

$$\mathcal{L}^a = \sum_{i=1}^{n} b_i(t,x,a)\partial_i + \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{d} \sigma_{i,k}(t,x,a)\sigma_{j,k}(t,x,a)\partial_{ij}$$

- In order to ensure that SDE (25) admits a strong solution, we assume that b, σ satisfy:

  Assum(SDE_α**)**: b and σ are Lipschitz-continuous functions in x uniformly in t and α, and satisfy a linear growth condition:

$$\begin{aligned} |b(t,x,\alpha) - b(t,y,\alpha)| + |\sigma(t,x,\alpha) - \sigma(t,y,\alpha)| &\leq C|x-y| \\ |b(t,x,\alpha)| + |\sigma(t,x,\alpha)| &\leq C\,(1+|x|) \end{aligned}$$

  for every $t \geq 0$, $x, y \in \mathbb{R}^n$ and $\alpha \in \mathcal{A}$ where C is a positive constant.
- This condition ensures that

$$\mathbb{E}^{\mathbb{Q}}\left[\sup_{0\leq s\leq T} |X_s^\alpha|^2\right] < \infty \qquad (25)$$

Stochastic optimal control problems

- A standard stochastic optimal control problem consists of maximizing a reward function (or minimizing a cost function) J given by

$$J(t,x,\alpha) := \mathbb{E}^{\mathbb{Q}}\left[\int_t^T f(s,X_s^\alpha,\alpha_s)\,ds + g(X_T^\alpha)\middle|X_t^\alpha = x\right]$$

with respect to the control $\alpha \in \mathcal{A}_{t,T}$:

$$u(t,x) := \sup_{\alpha\in\mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{Q}}\left[\int_t^T f(s,X_s^\alpha,\alpha_s)\,ds + g(X_T^\alpha)\middle|X_t^\alpha = x\right] \qquad (26)$$

where the initial condition is $X_t^\alpha = x$.

- In order to ensure that the reward function is finite, we assume that the supremum is taken over the subset $\mathcal{A}_{t,T}^*$ of $\mathcal{A}_{t,T}$ satisfying

$$\mathbb{E}^{\mathbb{Q}}\left[\int_t^T |f(s,X_s^\alpha,\alpha_s)|\,ds + |g(X_T^\alpha)|\right] < \infty \qquad (27)$$

Stochastic optimal control problems: standard form ($f = 0$)

- By augmenting the state variables X_t^α with the path-dependent continuous variable

$$dZ_t = f(t, X_t^\alpha, \alpha_t)\, dt, \qquad Z_0 = 0$$

our stochastic control problem can be written without loss of generality as

$$u(t, \tilde{x}) = \sup_{\alpha \in \mathcal{A}^*_{t,T}} \mathbb{E}^{\mathbb{Q}}[\tilde{g}(\tilde{X}_T^\alpha) | \tilde{X}_t^\alpha = (x, z)] - z$$

with $\tilde{X}^\alpha = (X^\alpha, Z)$ and $\tilde{g}(\tilde{x}) = g(x) + z$.

- We shall appeal to this standard form (i.e., $f = 0$)

$$u(t, x) := \sup_{\alpha \in \mathcal{A}^*_{t,T}} \mathbb{E}^{\mathbb{Q}}[g(X_T^\alpha) | X_t^\alpha = x] \tag{28}$$

to review results on stochastic control.

- **Assum(g)**: g has quadratic growth, i.e., there exists $C \geq 0$ such that for all $x \in \mathbb{R}^n$, $|g(x)| \leq C\left(1 + |x|^2\right)$.
- From the integrability result (25), this assumption on g implies that $\mathcal{A}^*_{t,T} = \mathcal{A}_{t,T}$.

Bellman's principle

Bellman's principle, also known as dynamic programming principle (DPP):

$$u(t,x) = \sup_{\alpha \in \mathcal{A}_{t,t+h}} \mathbb{E}^{\mathbb{Q}}[u(t+h, X^\alpha_{t+h}) | X^\alpha_t = x] \tag{29}$$

Bellman's principle also holds for any stopping time τ with $t \leq \tau \leq T$, $\mathbb{Q}-$a.s.

$$u(t,x) = \sup_{\alpha \in \mathcal{A}_{t,\tau}} \mathbb{E}^{\mathbb{Q}}[u(\tau, X^\alpha_\tau) | X^\alpha_t = x]$$

Proof.

(i). Using iterated conditional expectation, one gets

$$u(t,x) = \sup_{\alpha \in \mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{Q}} \left[\mathbb{E}^{\mathbb{Q}} \left[g\left(X^\alpha_T\right) \middle| \mathcal{F}_{t+h} \right] \middle| X^\alpha_t = x \right]$$

By the Markov property of X^α and the definition of u, one has

$$\mathbb{E}^{\mathbb{Q}} \left[g\left(X^\alpha_T\right) \middle| \mathcal{F}_{t+h} \right] = \mathbb{E}^{\mathbb{Q}} \left[g\left(X^\alpha_T\right) \middle| X^\alpha_{t+h} \right] \leq u\left(t+h, X^\alpha_{t+h}\right), \quad \forall \alpha \in \mathcal{A}_{t+h,T}.$$

Hence

$$u(t,x) \leq \sup_{\alpha \in \mathcal{A}_{t,t+h}} \mathbb{E}^{\mathbb{Q}} \left[u\left(t+h, X^\alpha_{t+h}\right) \middle| X^\alpha_t = x \right].$$

Bellman's principle

(ii). For any $\alpha \in \mathcal{A}_{t,t+h}$, there is an $\alpha^\epsilon \in \mathcal{A}_{t+h,T}$ such that

$$u\left(t+h, X^\alpha_{t+h}\right) \leq \mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\alpha^\epsilon}_T\right) \middle| X^\alpha_{t+h}\right] + \epsilon.$$

We glue together α and α^ϵ:

$$\tilde{\alpha}^\epsilon_s = \begin{cases} \alpha_s & \text{if } s \in [t, t+h) \\ \alpha^\epsilon_s & \text{if } s \in [t+h, T] \end{cases}.$$

Then $\tilde{\alpha}^\epsilon_s \in \mathcal{A}_{t,T}$ and

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}\left[u\left(t+h, X^\alpha_{t+h}\right) \middle| X^\alpha_t = x\right] &\leq \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\alpha^\epsilon}_T\right) \middle| X^\alpha_{t+h}\right] \middle| X^\alpha_t = x\right] + \epsilon \\
&= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\alpha^\epsilon}_T\right) \middle| \mathcal{F}_{t+h}\right] \middle| X^\alpha_t = x\right] + \epsilon \\
&= \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\tilde{\alpha}^\epsilon}_T\right) \middle| \mathcal{F}_{t+h}\right] \middle| X^\alpha_t = x\right] + \epsilon \\
&= \mathbb{E}^{\mathbb{Q}}\left[g\left(X^{\tilde{\alpha}^\epsilon}_T\right) \middle| X^\alpha_t = x\right] + \epsilon \leq u(t,x) + \epsilon.
\end{aligned}$$

Letting $\epsilon \to 0$, we obtain the reverse inequality.

Formal derivation of the HJB PDE

- Let us take an arbitrary constant control $\alpha_s = a$ with $a \in \mathcal{A}$ during the interval $[t, t+h]$. From the Bellman principle, we get

$$u(t,x) \geq \mathbb{E}^{\mathbb{Q}}[u(t+h, X^a_{t+h}) | X^a_t = x]$$

- By applying Itô's lemma to $u(t+h, X^\alpha_{t+h})$ (u should be smooth enough, $u \in C^{1,2}([0,T) \times \mathbb{R}^n)$), we get

$$u(t,x) \geq \mathbb{E}^{\mathbb{Q}}\left[u(t,x) + \int_t^{t+h} (\partial_s + \mathcal{L}^a)\, u(s, X^a_s)\, ds + M \middle| X^a_t = x\right]$$

  where $M := \int_t^{t+h} Du(s, X^a_s)\sigma(s, X^a_s, a)\, dW_s$ and $\mathcal{L}^a$ is the Itô generator associated to X_t (25) controlled by the constant $a \in \mathcal{A}$.

- By assuming that M is not only a local martingale but a true martingale, we obtain

$$\mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} (\partial_s + \mathcal{L}^a)\, u(s, X_s)\, ds \middle| X_t = x\right] \leq 0$$

  Dividing by $h > 0$, letting $h \to 0$, and permuting the limit and the expectation using the dominated convergence theorem, we finally get

$$\partial_t u(t,x) + \mathcal{L}^a u(t,x) \leq 0, \qquad \forall a \in \mathcal{A} \tag{30}$$

Formal derivation of the HJB PDE

- Then we take the optimal control α^* that realizes the supremum in Equation (29). By following a similar route, we get

$$\partial_t u(t,x) + \mathcal{L}^{\alpha_t^*} u(t,x) = 0 \tag{31}$$

- By combining Equations (30) and (31), we get that u, as given by Equation (28), is a solution to the following parabolic second order fully nonlinear PDE, the so-called HJB equation:

$$\partial_t u(t,x) + \sup_{a\in\mathcal{A}} \mathcal{L}^a u(t,x) = 0, \qquad u(T,x) = g(x) \tag{32}$$

- The interpretation is clear: For a deterministic constant "control" a, the linear PDE reads

$$-\partial_t u(t,x) = \mathcal{L}^a u(t,x), \qquad u(T,x) = g(x)$$

  At each time, we actually have the possibility to choose a in $\mathcal{A}$ so as to maximize $u(0,\cdot)$. Since u is known at the final date, $a \in \mathcal{A}$ must be chosen so that $-\partial_t u$ is maximized. This yields

$$-\partial_t u(t,x) = \sup_{a\in\mathcal{A}} \mathcal{L}^a u(t,x), \qquad u(T,x) = g(x)$$

  which is Equation (32).

HJB PDE with source term f and discount factor

- For our initial stochastic control problem with a source term (26), the HJB PDE reads

$$\partial_t u(t,x) + \sup_{a\in\mathcal{A}}\{\mathcal{L}^a u(t,x) + f(t,x,a)\} = 0, \qquad u(T,x) = g(x) \tag{33}$$

- When we include a discount factor:

$$u(t,x) = \sup_{\alpha\in\mathcal{A}_{t,T}} \mathbb{E}^{\mathbb{Q}}\Big[g(X_T^\alpha)e^{-\int_t^T r(s,X_s^\alpha,\alpha_s)\,ds} + \int_t^T e^{-\int_t^s r(u,X_u^\alpha,\alpha_u)du} f(s,X_s^\alpha,\alpha_s)\,ds \Big| X_t = x\Big] \tag{34}$$

the HJB PDE reads

$$\begin{aligned} \partial_t u(t,x) + \sup_{a\in\mathcal{A}}\{\mathcal{L}^a u(t,x) + f(t,x,a) - r(t,x,a)u(t,x)\} &= 0 \qquad (35)\\ u(T,x) &= g(x) \end{aligned}$$

- Note that if the control can only be chosen at discrete dates, the HJB equation is not applicable because it assumes that the control can be chosen continuously in time. In these cases, we must rely on the (discrete) Bellman's dynamic programming equation.

Derivation of HJB PDE directly with source term f

- Bellman's dynamic programming principle:

$$u(t,x) = \sup_{\alpha_{s\in[t,t+h]}\in\mathcal{A}} \mathbb{E}^{\mathbb{P}}\left[\int_t^{t+h} f(s,X_s^\alpha,\alpha_s)\,ds + u(t+h,X_{t+h}^\alpha)\middle| X_t^\alpha = x\right]$$

- For any constant control a on $[t,t+h]$,

$$u(t,X_t^a) \geq \mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} f(s,X_s^a,a)\,ds + u(t+h,X_{t+h}^a)\middle| X_t^a = x\right]$$

$$= \mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} f(s,X_s^a,a)\,ds + u(t,X_t^a) + \int_t^{t+h} (\partial_s + \mathcal{L}^a)\,u(s,X_s^a)\,ds + M\middle| X_t^a = x\right]$$

$$0 \geq \mathbb{E}^{\mathbb{Q}}\left[\int_t^{t+h} f(s,X_s^a,a)\,ds + \int_t^{t+h} (\partial_s + \mathcal{L}^a)\,u(s,X_s^a)\,ds\middle| X_t^a = x\right]$$

Letting $h \to 0$, one gets

$$f(t,x,a) + (\partial_t + \mathcal{L}^a)\,u(t,x) \leq 0.$$

- For the optimal control α^*, the inequality becomes equality:

$$f(t,x,\alpha_t^*) + \left(\partial_t + \mathcal{L}^{\alpha_t^*}\right) u(t,x) = 0$$

$$\partial_t u(t,x) + \sup_{\alpha\in\mathcal{A}} \left\{f(t,x,\alpha) + \mathcal{L}^\alpha u(t,x)\right\} = 0$$

Verification

Theorem (Verification Theorem)

Let $U \in C^{1,2}([0,T) \times \mathbb{R}^n) \cap C^0([0,T] \times \mathbb{R}^n)$ be a function with quadratic growth in x, uniformly in t, and for all $(t,x) \in [0,T) \times \mathbb{R}^n$, there exists $\alpha^*(t,x) \in \mathcal{A}$ such that

$$-\partial_t U(t,x) - \sup_{\alpha \in \mathcal{A}} \{\mathcal{L}^\alpha U(t,x)\} = -\partial_t U(t,x) - \mathcal{L}^{\alpha^*} U(t,x) = 0$$

and that the SDE with the control α^* admits a strong solution $X_t^{\alpha^*}$. We also assume that $U(T,\cdot) = g$. Then $U = u$ in $[0,T] \times \mathbb{R}^n$.

- For any admissible control α, Itô's lemma gives
  $$U(T, X_T^\alpha) = U(t,x) + \int_t^T (\partial_s U + \mathcal{L}^\alpha U(s, X_s^\alpha))\, ds + \text{martingale}.$$
  Taking expectation $\mathbb{E}[\cdot | X_t^\alpha = x]$ yields $U(t,x) \geq \sup_\alpha \mathbb{E}\left[g(X_T^\alpha) \middle| X_t^\alpha = x\right]$.
  Setting $\alpha = \alpha^*$ one has $U(t,x) = \mathbb{E}\left[g(X_T^{\alpha^*}) \middle| X_t^{\alpha^*} = x\right]$. Thus $U = u$.
- For non-smooth solutions, consider viscosity solutions.

Example 1: The uncertain volatility model

- Here the nonlinearity arises as a result of modeling choices.
- We model the asset X_t $(n = d = 1)$ by a positive local Itô $(\mathcal{F}_t, \mathbb{Q})$-martingale (zero interest rates, repos, and dividends for simplicity):

$$dX_t = \sigma_t X_t \, dW_t$$

- The volatility σ_t is unspecified for the moment.
- In the UVM introduced in Avellaneda-Levy-Paras and Lyons (1995), the volatility is uncertain. As a minimal modeling hypothesis, we only assume that:
    - the volatility is valued in a compact interval $[\underline{\sigma}, \overline{\sigma}]$, and
    - it cannot look into the future (i.e., it is an $\mathcal{F}_t$-adapted process).

The uncertain volatility model

- Consider an option delivering a payoff F_T at maturity T which is a function of the asset path $(X_t, 0 \leq t \leq T)$.
- Then the time-t value u_t of the option is (worst case pricing):

$$u_t = \sup_{[t,T]} \mathbb{E}^{\mathbb{Q}}[F_T | \mathcal{F}_t] \tag{36}$$

  Here, $\sup_{[t,T]}$ means that the supremum is taken over all $(\mathcal{F}_t)$-adapted processes $(\sigma_s)_{t \leq s \leq T}$ such that for all $s \in [t, T]$, σ_s belongs to the domain $[\underline{\sigma}, \overline{\sigma}]$.
- We *cannot* control the volatility of the underlying. However, the pricing formulation is equivalent to that of the value function in a stochastic control problem.

The uncertain volatility model: vanilla options

- For vanilla payoffs $F_T = g(X_T)$, the price $u(t,x)$ is solution to a nonlinear PDE:

$$\partial_t u(t,x) + \frac{1}{2} \sup_{\sigma \in [\underline{\sigma}, \overline{\sigma}]} \sigma^2 x^2 \partial_x^2 u(t,x) = 0, \qquad u(T,x) = g(x)$$

- Taking the supremum over σ, we get a fully nonlinear PDE called the Black-Scholes-Barenblatt equation (in short BSB)

$$\begin{aligned} \partial_t u(t,x) + \frac{1}{2} x^2 \Sigma\left(\partial_x^2 u(t,x)\right)^2 \partial_x^2 u(t,x) &= 0, \quad (t,x) \in [0,T) \times \mathbb{R}_+^* \\ u(T,x) &= g(x), \quad x \in \mathbb{R}_+^* \end{aligned}$$

  with $\Sigma(\Gamma) = \underline{\sigma} \mathbf{1}_{\Gamma < 0} + \overline{\sigma} \mathbf{1}_{\Gamma \geq 0}$.
- Fully nonlinear PDE: the nonlinearity affects the 2nd order derivative $\partial_x^2 u$.
- If we consider a convex payoff, u coincides with the Black-Scholes price with the upper volatility $\overline{\sigma}$.

Example 2: The uncertain mortality model for reinsurance deals

- Examples of reinsurance deals: GMxB deals.
- GMxB stands for GMIB (Guaranteed Minimum Investment Benefit) or GMDB (Guaranteed Minimum Death Benefit).
- Let us ignore the usual lapse feature of GMxB deals. Then those deals consist of only two payoffs:
    - at maturity T, the seller of the option pays the payoff $g(X_T)$ with X_T the value of an asset;
    - in the case where the insurance subscriber dies before maturity, the seller pays at time t the payoff $g_{\mathrm{D}}(X_t)$ to the beneficiaries of the insurance policy.

  The index D stands for default.

The uncertain mortality model for reinsurance deals

- Like in credit modeling, we model the time to default τ (here, the time of death) by the first time of default of a Poisson process with intensity λ_t (the mortality rate).
- In the uncertain mortality model, this rate λ_t is assumed to be uncertain and to range into the interval $[\underline{\lambda}(t), \overline{\lambda}(t)]$. Also, the rate cannot look into the future, i.e., it must be adapted.
- The fair value of this option is then given by (worst case pricing):

$$u(t,x) = \sup_{\lambda \in \Lambda_{t,T}} \mathbb{E}^{\mathbb{Q}}[g(X_T)\mathbf{1}_{\tau \geq T} + g_{\mathrm{D}}(X_\tau)\mathbf{1}_{\tau < T} | X_t = x]$$

where $\Lambda_{t,T}$ is the set of all adapted processes λ such that for all $s \in [t,T]$, $\lambda_s \in [\underline{\lambda}(s), \overline{\lambda}(s)]$.

The uncertain mortality model for reinsurance deals

- u is solution to a nonlinear PDE:

$$\partial_t u(t,x) + \mathcal{L}u(t,x) + \sup_{\lambda \in [\underline{\lambda}(t), \overline{\lambda}(t)]} \lambda \left(g_{\mathrm{D}}(x) - u(t,x) \right) = 0$$

  which is equivalent to

$$\partial_t u(t,x) + \mathcal{L}u(t,x) + \Lambda \left(t, g_{\mathrm{D}}(x) - u \right) = 0$$

  with the terminal condition $u(T,x) = g(x)$ and Λ defined by

$$\Lambda(t,y) = \begin{cases} \overline{\lambda}(t)y & \text{if } y \geq 0 \\ \underline{\lambda}(t)y & \text{if } y < 0 \end{cases}$$

- Semilinear PDE: the nonlinearity does not affect the 2nd order derivative $\partial_x^2 u$. Here the nonlinearity affects only u.
- No analytical solution. How to numerically compute u? PDE solver useful only when X has dimension 1 or 2. We will discuss Monte Carlo methods (Least Squares Monte Carlo, BSDEs).

Example 3: Portfolio optimization

- Merton problem (a genuine stochastic control problem): consider an investor who can invest over the horizon $[0, T]$ in two assets: a risk-free asset B_t with rate of return $r > 0$ and a risky asset S_t ("stock"):

$$\begin{aligned} dB_t &= rB_t\,dt \\ dS_t &= \mu S_t\,dt + \sigma S_t\,dW_t \end{aligned}$$

- Let X_t^B denote the investor's wealth in the bank at time $t \geq 0$, X_t^S the wealth in stock, $X_t = X_t^B + X_t^S$ the investor's total wealth, and $\pi_t = X_t^S / X_t$ the proportion of wealth invested in stock.
- The investor has some initial capital $X_0 = x > 0$ to invest and there are no further cash injections.
- The investor can decide at each time $t \geq 0$ how much money to hold in stock: by choosing, at time t, the value π_t the investor controls the evolution of his/her wealth in the future.
- The investor's goal is to maximize expected total utility at a final time T:

$$J(t, x, \pi) = \mathbb{E}\left[g(X_T) | X_t = x\right], \qquad u(t, x) = \sup_{\pi \in \mathcal{A}(t,x)} J(t, x, \pi)$$

- u is solution to a nonlinear PDE.

Portfolio optimization: Exercise

1. What are the *quantities* Δ_t^B and Δ_t^S of risk-free asset and risky asset held by the investor at time t? Show that, given a portfolio process π, the wealth of the investor evolves according to the SDE

$$dX_t^{x,\pi} = \left(r + \pi_t(\mu - r)\right) X_t^{x,\pi} dt + \sigma \pi_t X_t^{x,\pi}\, dW_t \tag{37}$$

2. Write down the HJB PDE for this problem.
3. Show that it reads

$$\begin{aligned} \partial_t v + rx\partial_x v + \pi^*(\mu - r)x\partial_x v + \frac{1}{2}\left(\pi^*\right)^2 \sigma^2 x^2 \partial_x^2 v &= 0 \\ v(T,x) &= g(x) \end{aligned}$$

with

$$\pi^*(t,x) = -\frac{(\mu - r)x\partial_x v(t,x)}{\sigma^2 x^2 \partial_x^2 v(t,x)}$$

Backward Stochastic Differential Equations

I find it hard to focus looking forward. So I look backward.
— Iggy Pop

Classical Feynman-Kac Formula

Linear Parabolic PDE:

$$\partial_t u(t,x) + \mathcal{L}u(t,x) + a(t,x)u(t,x) + c(t,x) = 0, \quad u(T,x) = g(x)$$

where $(t,x) \in [0,T] \times \mathbb{R}^d$ and

$$\mathcal{L} = \sum_{i=1}^{d} b_i(t,x)\partial_i + \frac{1}{2}\sum_{i,j=1}^{d}\sum_{k=1}^{r} \sigma_{i,k}\sigma_{j,k}\partial_{i,j}$$

We consider a d-dimensional Itô process defined by the (forward) SDE

$$dX_t = b(t,X_t)\,dt + \sigma(t,X_t)\,dW_t$$

where $b \in \mathbb{R}^d$, $\sigma \in \mathbb{R}^{d\times r}$ and W is an r-dimensional standard Brownian motion.

Let $Y_t = u(t, X_t)$. Applying Itô's lemma gives

$$\begin{aligned} dY_t &= (\partial_t u + \mathcal{L}u)\, dt + \sigma(t, X_t)^T D_x u(t, x) \cdot dW_t \\ &= (-a(t, X_t)u(t, X_t)dt - c(t, X_t))\, dt + \sigma(t, X_t)^T D_x u(t, X_t) \cdot dW_t. \end{aligned}$$

and therefore

$$e^{\int_0^t a(s,X_s)ds} Y_t + \int_0^t c(s, X_s) e^{\int_0^s a(v,X_v)dv} ds \text{ is a martingale.}$$

Classical Feynman-Kac

$$u(t, x) = \mathbb{E}\left[g(X_T) e^{\int_t^T a(r,X_r)dr} + \int_t^T c(r, X_r) e^{\int_t^r a(s,X_s)ds}\, dr \,\middle|\, X_t = x \right]$$

Provides probabilistic representation of solutions of linear parabolic PDEs.

First-Order BSDE - Motivation

Semilinear Parabolic PDE:

$$\partial_t u(t,x) + \mathcal{L}u(t,x) + f\left(t, x, u(t,x), \sigma(t,x)^T D_x u(t,x)\right) = 0, \quad u(T,x) = g(x)$$

$$\begin{aligned} dY_t &= (\partial_t u + \mathcal{L}u)\, dt + \sigma(t, X_t)^T D_x u(t, X_t) \cdot dW_t \\ &= -f\left(t, X_t, Y_t, \sigma(t, X_t)^T D_x u(t, X_t)\right) dt + \sigma(t, X_t)^T D_x u(t, X_t) \cdot dW_t. \end{aligned}$$

Let $Z_t = \sigma(t, X_t)^T D_x u(t, X_t)$,

$$\boxed{dY_t = -f\left(t, X_t, Y_t, Z_t\right) dt + Z_t \cdot dW_t}$$

with the terminal condition

$$\boxed{Y_T = g\left(X_T\right)}$$

First-Order BSDE - Definition

Definition (First-Order BSDE)

A solution to a (Markovian) first-order BSDE is a couple (Y, Z) of $(\mathcal{F}_t)$-*adapted* Itô processes taking values in $\mathbb{R} \times \mathbb{R}^r$ and satisfying

$$dY_t = -f(t, X_t, Y_t, Z_t)\, dt + Z_t \cdot dW_t$$

with the *terminal* condition $Y_T = g(X_T)$. The above equation means that

$$Y_t = g(X_T) + \int_t^T f(s, X_s, Y_s, Z_s)\, ds - \int_t^T Z_s \cdot dW_s$$

The deterministic function f is called the *driver* or *generator*. We emphasize that the diffusion term Z is a part of the solution.

They provide a probabilistic representation of solutions of semi-linear parabolic PDEs, generalizing the Feynman-Kac formula.

Interpretation: payoff $g(X_T)$, price at time t, Y_t, delta strategy Z_t, source term $f(t, X_t, Y_t, Z_t)$.

First-Order BSDE - Existence and Uniqueness

Theorem (Pardoux-Peng)

Assume that $f : [0,T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^r \to \mathbb{R}$ *is a deterministic function satisfying the Lipschitz condition*

$$|f(t,x,y_1,z_1) - f(t,x,y_2,z_2)| \leq K\left(|y_1 - y_2| + |z_1 - z_2|\right)$$

for some constant K *independent of* $(y_1, y_2) \in \mathbb{R}$ *and* $(z_1, z_2) \in \mathbb{R}^r$*, that*

$$\mathbb{E}\left[\int_0^T |f(t,X_t,0,0)|^2 \, dt\right] < \infty$$

and that $g \in \mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{R}^d)$*. Then there is a unique adapted solution* (Y,Z) *to the 1-BSDE satisfying*

$$\sup_{0 \leq t \leq T} \mathbb{E}\left[Y_t^2\right] + \mathbb{E}\left[\int_0^T |Z_t|^2 \, dt\right] < \infty$$

Martingale Representation Theorem

Example (Martingale Representation Theorem)

We consider the case where f does not depend on Y and Z,

$$dY_t = -f(t, X_t)dt + Z_t \cdot dW_t.$$

Then $Y_t + \int_0^t f(s, X_s)ds$ is an $(\mathcal{F}_t)$-martingale with terminal value $g(X_T) + \int_0^T f(s, X_s)\, ds$ at $t = T$.

$$Y_t + \int_0^t f(s, X_s)\, ds = \mathbb{E}\left[g(X_T) + \int_0^T f(s, X_s)\, ds \middle| \mathcal{F}_t \right] =: M_t$$

According to the martingale representation theorem, there exists a unique Z s.t.

$$M_t = M_0 + \int_0^t Z_s \cdot dW_s.$$

Define $Y_t := M_t - \int_0^t f(s, X_s)\, ds = \mathbb{E}\left[g(X_T) + \int_t^T f(s, X_s)\, ds \middle| \mathcal{F}_t \right]$. Then Y is adapted and $Y_T = g(X_T)$, $dY_t = Z_t \cdot dW_t - f(t, X_t)dt$. The pair (Y, Z) gives the solution to the BSDE.

Linear First-Order BSDE

Example (Linear 1-BSDE)

$$dY_t = -\left(a_t Y_t + b_t \cdot Z_t + c_t\right) dt + Z_t \cdot dW_t, \quad Y_T = g\left(X_T\right)$$

where a_t, b_t, and c_t are deterministic functions of time. Let

$$d\Gamma_t = \Gamma_t(a_t\, dt + b_t \cdot dW_t), \qquad \Gamma_0 = 1.$$

$$\Gamma_t = \exp\left(\int_0^t b_r \cdot dW_r - \frac{1}{2}\int_0^t |b_r|^2\, dr + \int_0^t a_r\, dr\right)$$

Applying Itô's lemma to $\Gamma_t Y_t$, we obtain

$$d\left(\Gamma_t Y_t\right) = -c_t \Gamma_t dt + \Gamma_t\left(Z_t + b_t Y_t\right) dW_t.$$

Then $\Gamma_t Y_t + \int_0^t c_r \Gamma_r\, dr$ is a $\mathcal{F}_t$-martingale with terminal value $g(X_T)\Gamma_T + \int_0^T c_r \Gamma_r dr$ at $t = T$. Define $M_t := \mathbb{E}\left[g(X_T)\Gamma_T + \int_0^T c_r \Gamma_r dr \middle| \mathcal{F}_t\right]$ and Y_t by

$$Y_t := \Gamma_t^{-1}\left(M_t - \int_0^t c_r \Gamma_r dr\right) = \Gamma_t^{-1}\mathbb{E}\left[g(X_T)\Gamma_T + \int_t^T c_r \Gamma_r\, dr \middle| \mathcal{F}_t\right].$$

The existence and uniqueness of Z_t is guaranteed from the martingale representation theorem applied to M_t.

First-Order BSDE - Numerical Simulation

- Divide $(0,T)$ into subintervals (t_{i-1},t_i), $1 \leq i \leq n$, and set $\Delta t_i = t_i - t_{i-1}$, $\Delta W_{t_i} = W_{t_i} - W_{t_{i-1}}$, and $\Delta = \max_i \Delta t_i$
- Simulating the forward process X gives us values of $X_T = X_{t_n}$ and $Y_T = Y_{t_n} = g(X_{t_n})$. Then we proceed backward

$$Y_{t_i} - Y_{t_{i-1}} = -\int_{t_{i-1}}^{t_i} f(s, X_s, Y_s, Z_s)\, ds + \int_{t_{i-1}}^{t_i} Z_s \cdot dW_s$$

Discretize the integral

$$Y_{t_i} - Y_{t_{i-1}} \approx -f\left(t_{i-1}, X_{t_{i-1}}, Y_{t_{i-1}}, Z_{t_{i-1}}\right) \Delta t_i + Z_{t_{i-1}} \cdot \Delta W_{t_i}$$

Taking conditional expectation $\mathbb{E}_{i-1} := \mathbb{E}\left[\cdot | \mathcal{F}_{t_{i-1}}\right]$

$$\begin{aligned} Y_{t_{i-1}} &= \mathbb{E}_{i-1}[Y_{t_i}] + f(t_{i-1}, X_{t_{i-1}}, Y_{t_{i-1}}, Z_{t_{i-1}})\Delta t_i \\ Z_{t_{i-1}} &= \frac{1}{\Delta t_i}\mathbb{E}_{i-1}[Y_{t_i} \Delta W_{t_i}] \end{aligned}$$

First-Order BSDE - Numerical Simulation

- *Implicit* scheme: Use regression to estimate the conditional expectations $\mathbb{E}_{i-1}\left[Y_{t_i}\right]$ and $\mathbb{E}_{i-1}\left[Y_{t_i}\Delta W_{t_i}\right]$, and then solve, say, by a root-finding routine, for $Y_{t_{i-1}}$.
- *Explicit* schemes:

$$Y_{t_{i-1}} = \mathbb{E}_{i-1}[Y_{t_i} + f(t_i, X_{t_i}, Y_{t_i}, Z_{t_i})\Delta t_i]$$

or

$$Y_{t_{i-1}} = \mathbb{E}_{i-1}[Y_{t_i}] + f(t_{i-1}, X_{t_{i-1}}, \mathbb{E}_{i-1}[Y_{t_i}], Z_{t_{i-1}})\Delta t_i$$

- θ-schemes

$$Y_{t_{i-1}} = \mathbb{E}_{i-1}\left[Y_{t_i}\right] + \theta\mathbb{E}_{i-1}\left[f\left(t_i, X_{t_i}, Y_{t_i}, Z_{t_i}\right)\right]\Delta t_i + (1-\theta) f\left(t_{i-1}, X_{t_{i-1}}, Y_{t_{i-1}}, Z_{t_{i-1}}\right)\Delta t_i$$

or

$$Y_{t_{i-1}} = \mathbb{E}_{i-1}\left[Y_{t_i}\right] + \theta f\left(t_i, X_{t_i}, \mathbb{E}_{i-1}\left[Y_{t_i}\right], Z_{t_i}\right)\Delta t_i + (1-\theta) f\left(t_{i-1}, X_{t_{i-1}}, Y_{t_{i-1}}, Z_{t_{i-1}}\right)\Delta t_i$$

First-Order BSDE - Numerical Simulation - Example

Example (CVA pricing)

The semilinear PDE arising in the pricing of counterparty risk in case of risky close-out is

$$\partial_t u + \mathcal{L}u + \beta(u^+ - u) = 0, \qquad u(T,x) = g(x)$$

The corresponding BSDE is

$$dY_t = -\beta(Y_t^+ - Y_t)\,dt + Z_t \cdot dW_t, \qquad Y_T = g(X_T)$$

This can be solved numerically using either of the following schemes:

$$\begin{aligned} & Y_{t_{i-1}} &=& \frac{1}{1+\beta\Delta t_i \mathbf{1}_{\mathbb{E}_{i-1}[Y_{t_i}]\le 0}}\mathbb{E}_{i-1}[Y_{t_i}] \\ \text{or} \quad & Y_{t_{i-1}} &=& \mathbb{E}_{i-1}[Y_{t_i} + \beta\Delta t_i(Y_{t_i}^+ - Y_{t_i})] \\ \text{or} \quad & Y_{t_{i-1}} &=& \mathbb{E}_{i-1}[Y_{t_i}] + \beta\Delta t_i(\mathbb{E}_{i-1}[Y_{t_i}]^+ - \mathbb{E}_{i-1}[Y_{t_i}]) \end{aligned}$$

or θ-schemes.

Second Order BSDE - Motivation

Consider the fully nonlinear parabolic PDE

$$\partial_t u(t,x) + f(t,x,u,D_x u, D_x^2 u) = 0, \quad u(T,x) = g(x)$$

where $(t,x) \in [0,T) \times \mathbb{R}^d$. Assume that

$$dX_t = \sigma(t,X_t)dW_t$$

Let $Y_t = u(t,X_t)$, $Z_t = D_x u(t,X_t)$, $\Gamma_t = D_x^2 u(t,X_t)$, and apply Itô's formula

$$dY_t = \left(-f(t,X_t,Y_t,Z_t,\Gamma_t) + \frac{1}{2}\mathrm{tr}\left(\sigma(t,X_t)\sigma(t,X_t)^T\Gamma_t\right)\right)dt + Z_t \cdot \sigma(t,X_t)\,dW_t$$

$$dZ_t = (\partial_t + \mathcal{L})D_x u(t,X_t)\,dt + \Gamma_t \sigma(t,X_t)dW_t$$

Second Order BSDE - Definition

Definition (Second order BSDE)

Let $(Y_t, Z_t, \Gamma_t, \alpha_t)_{t\in[0,T]}$ be a quadruple of $(\mathcal{F}_t)$-adapted processes taking values in $\mathbb{R}$, $\mathbb{R}^d$, $\mathcal{S}^d$, and $\mathbb{R}^d$ respectively.[a] We call (Y, Z, Γ, α) a solution to a (Markovian) 2-BSDE if

$$
\begin{aligned}
dY_t &= -f(t, X_t, Y_t, Z_t, \Gamma_t)\, dt + Z_t \circ \sigma(t, X_t)\, dW_t \\
dZ_t &= \alpha_t\, dt + \Gamma_t \sigma(t, X_t)\, dW_t
\end{aligned}
$$

where $Y_T = g(X_T)$. The use of the Stratonovich product $\circ$ is only for convenience and can be replaced by an Itô integral

$$
dY_t = \left(-f(t, X_t, Y_t, Z_t, \Gamma_t) + \frac{1}{2}\operatorname{tr}\left(\sigma(t, X_t)\sigma(t, X_t)^T \Gamma_t \right) \right) dt + Z_t \cdot \sigma(t, X_t)\, dW_t
$$

[a] $\mathcal{S}^d$ stands for the space of d-dimensional symmetric matrices.

- They provide a probabilistic representation of solutions of fully non-linear parabolic PDEs, generalizing the Feynman-Kac formula.

Second Order BSDE - Numerical Simulation

Following the same discretization scheme as in 1-BSDE:

$$
\begin{aligned}
Y_{t_n} &= g(X^{\Delta}_{t_n}) \\
Z_{t_n} &= Dg(X_{t_n}) \\
Y_{t_{i-1}} &= \mathbb{E}_{i-1}[Y_{t_i}] + \Big(f(t_{i-1}, X_{t_{i-1}}, Y_{t_{i-1}}, Z_{t_{i-1}}, \Gamma_{t_{i-1}}) \\
&\quad - \frac{1}{2}\mathrm{tr}[\sigma(t_{i-1}, X_{t_{i-1}})\sigma(t_{i-1}, X_{t_{i-1}})^T \Gamma_{t_{i-1}}] \Big) \Delta t_i \\
Z_{t_{i-1}} &= \frac{1}{\Delta t_i} \sigma(t_{i-1}, X_{t_{i-1}})^{T^{-1}} \mathbb{E}_{i-1}[Y_{t_i} \Delta W_{t_i}] \\
\Gamma_{t_{i-1}} &= \frac{1}{\Delta t_i} \mathbb{E}_{i-1}[Z_{t_i} \Delta W^T_{t_i}] \sigma(t_{i-1}, X_{t_{i-1}})^{-1}
\end{aligned}
$$

Uncertain Lapse and Mortality Model

Yes, man is mortal, but that would be only half the trouble. The worst of it is that he's sometimes unexpectedly mortal — there's the trick!
— Mikhail Bulgakov

Reinsurance Deals

Reinsurance deals are sold by banks to insurance companies that need to hedge their book of variable annuities. A typical reinsurance deal consists of three payoffs:

- At maturity, in the case where the insurance subscriber is still alive, the issuer delivers a put on the net asset value (NAV) X of the underlying fund, whose strike is denoted K_{mat}.

$$u^{\mathrm{mat}}(x) = (K_{\mathrm{mat}} - x)_+$$

- In the case where the insurance subscriber dies before maturity, the issuer delivers a put on X whose strike is denoted K_D.

$$u^D(t, x) = (K_D - x)_+$$

- Each month, the issuer receives a constant fee.

All three payoffs are canceled as soon as the insurance subscriber lapses, which he or she can do whenever he or she wants:

$$u^L(t, x) = 0$$

Deterministic lapse and mortality model

Assumptions

- We model death (D) and lapse (L) as independent default events, jointly independent of the path $(X_t, 0 \le t \le T)$ of the NAV of the underlying fund.
- They are supposed to be first jump times of two independent Poisson processes, τ^D and τ^L, with respective deterministic intensities λ_t^D and λ_t^L.
- Once one of these defaults occurs, the product is canceled, and the NAV is unchanged.
- The payoffs $u^{\text{mat}}(\cdot)$, $u^D(t,\cdot)$ and $u^L(t,\cdot)$ are functions of the underlying asset price at the time (Markov payoff).
- For simplicity, we assume the underlying asset has constant volatility σ, and interest and repo rates are zero.

Deterministic lapse and mortality model

The fair value of this contract is

$$u(t,x) = \mathbb{E}^{\mathbb{Q}}_{t,x}\Big[u^{\mathrm{mat}}(X_T)\mathbf{1}_{\tau^D \geq T}\mathbf{1}_{\tau^L \geq T} - \int_t^T \alpha \mathbf{1}_{\tau^D \geq s}\mathbf{1}_{\tau^L \geq s}\,ds$$
$$+ u^D(\tau^D, X_{\tau^D})\mathbf{1}_{\tau^D < T}\mathbf{1}_{\tau^D \leq \tau^L} + u^L(\tau^L, X_{\tau^L})\mathbf{1}_{\tau^L < T}\mathbf{1}_{\tau^L < \tau^D}\Big]$$

$$\mathbb{E}^{\mathbb{Q}}_{t,x}\left[u^{\mathrm{mat}}(X_T)\mathbf{1}_{\tau^D \geq T}\mathbf{1}_{\tau^L \geq T}\right] = e^{-\int_t^T \lambda^D_s\,ds - \int_t^T \lambda^L_s\,ds}\mathbb{E}^{\mathbb{Q}}_{t,x}\left[u^{\mathrm{mat}}(X_T)\right]$$

$$\mathbb{E}^{\mathbb{Q}}_{t,x}\left[-\int_t^T \alpha \mathbf{1}_{\tau^D \geq s}\mathbf{1}_{\tau^L \geq s}\,ds\right] = -\alpha\int_t^T e^{-\int_t^s \lambda^D_v\,dv}e^{-\int_t^s \lambda^L_v\,dv}ds$$

$$\mathbb{E}^{\mathbb{Q}}_{t,x}\left[u^D(\tau^D, X_{\tau^D})\mathbf{1}_{\tau^D < T}\mathbf{1}_{\tau^D \leq \tau^L}\right] = \mathbb{E}^{\mathbb{Q}}_{t,x}\left[\int_t^T u^D(s, X_s)\lambda^D_s e^{-\int_t^s \lambda^D_v\,dv}e^{-\int_t^s \lambda^L_v\,dv}ds\right]$$

$$\mathbb{E}^{\mathbb{Q}}_{t,x}\left[u^L(\tau^L, X_{\tau^L})\mathbf{1}_{\tau^L < T}\mathbf{1}_{\tau^L < \tau^D}\right] = \mathbb{E}^{\mathbb{Q}}_{t,x}\left[\int_t^T u^L(s, X_s)\lambda^L_s e^{-\int_t^s \lambda^L_v\,dv}e^{-\int_t^s \lambda^D_v\,dv}ds\right]$$

where $\mathbb{E}_{t,x}\left[\cdot\right] = \mathbb{E}\left[\cdot\middle|\, X_t = x, \tau^D > t, \tau^L > t\right]$.

Deterministic lapse and mortality model

Summing up, we get

$$u(t,x) = \mathbb{E}^{\mathbb{Q}}_{t,x}\left[e^{-\int_t^T \left(\lambda_s^D + \lambda_s^L\right) ds} u^{\mathrm{mat}}\left(X_T\right)\right] +$$
$$\mathbb{E}^{\mathbb{Q}}_{t,x}\left[\int_t^T e^{-\int_t^s \left(\lambda_v^D + \lambda_v^L\right) dv} \left(-\alpha + u^D\left(s, X_s\right)\lambda_s^D + u^L\left(s, X_s\right)\lambda_s^L\right) ds\right]$$

By the classical Feynman-Kac formula, we have the linear PDE

$$\boxed{\partial_t u + \frac{1}{2}\sigma^2 x^2 \partial_x^2 u - \alpha + \lambda_t^D\left(u^D - u\right) + \lambda_t^L\left(u^L - u\right) = 0}$$

with the terminal condition $u(T,x) = u^{\mathrm{mat}}(x)$.

Derivatives with Default Risk - Recall

We model the time of default, τ, to be the first jump time of a Poisson process, with deterministic intensity λ_t. That is, for any $t \geq 0$, we have

$$\mathbb{P}\left[\tau \in (t, t+dt) | \tau > t\right] = \lambda_t dt \tag{38}$$

Let $\Phi(t) = \mathbb{P}\left[\tau \leq t\right]$ be the CDF of τ. Then it follows from (38) that

$$\frac{\Phi'(t)dt}{1 - \Phi(t)} = \lambda_t dt$$

Solving this ODE gives us the following:

$$\begin{aligned}
\mathbb{P}\left[\tau \leq t\right] &= 1 - e^{-\int_0^t \lambda_s ds} \\
\mathbb{P}\left[\tau > t\right] &= e^{-\int_0^t \lambda_s ds} \\
\mathbb{P}\left[\tau \in (t, t+dt)\right] &= \lambda_t e^{-\int_0^t \lambda_s ds} dt \\
\mathbb{P}\left[\tau > T | \tau > t\right] &= e^{-\int_t^T \lambda_s ds}, \quad T > t \\
\mathbb{P}\left[\tau \in (s, s+ds) | \tau > t\right] &= \lambda_s e^{-\int_t^s \lambda_a da} ds, \quad s > t
\end{aligned}$$

Derivatives with Default Risk - Recall

Consider European payoff $g(X_T)$ at time T where the underlying asset X_t is given by

$$\frac{dX_t}{X_t} = \left[r\left(t, X_t\right) - q\left(t, X_t\right)\right] dt + \sigma\left(t, X_t\right) dW_t.$$

If the counter-party defaults at time τ before T, the contract pays $u_D(\tau, X_\tau)$. The default time τ is assumed to be independent from X. The fair value of this contract at time t, provided that the counter-party has not defaulted yet at time t, is

$$\begin{aligned}
u(t,x) &= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s,X_s)ds} g(X_T) 1_{\tau \geq T} + e^{-\int_t^\tau r(s,X_s)ds} u_D(\tau, X_\tau) 1_{\tau < T} \middle| X_t = x, \tau > t \right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s,X_s)ds} g\left(X_T\right) e^{-\int_t^T \lambda_s ds} \middle| X_t = x \right] + \\
&\quad \mathbb{E}^{\mathbb{Q}}\left[\mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^\tau r(s,X_s)ds} u_D\left(\tau, X_\tau\right) 1_{\tau < T} \middle| \left(X_s\right)_{t \leq s \leq T}, \tau > t \right] \middle| X_t = x, \tau > t \right] \\
&= \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T (r(s,X_s)+\lambda_s)ds} g\left(X_T\right) + \int_t^T e^{-\int_t^s (r(\nu,X_\nu)+\lambda_\nu)d\nu} \lambda_s u_D(s, X_s) ds \middle| X_t = x \right]
\end{aligned}$$

From the Feynman-Kac formula, u satisfies the PDE

$$\frac{\partial u}{\partial t} + \left(r(t,x) - q(t,x)\right) x \frac{\partial u}{\partial x} + \frac{1}{2}\sigma^2(t,x)x^2 \frac{\partial^2 u}{\partial x^2} + \lambda_t \left(u_D(t,x) - u(t,x)\right) - r(t,x)u = 0$$

Deterministic lapse and mortality model

The P&L of the issuer's delta-hedged position incurred between t and $t+dt$, averaged over mortality and lapse events

$$
\begin{aligned}
&\mathbb{E}^{\mathbb{Q}}\left[d\mathrm{P\&L}_t \,|\, X_s, 0 \le s \le T, \text{ and no mortality or lapse has occured before } t\right] \\
&= \left(1-\lambda_t^D dt - \lambda_t^L dt\right)\left(-\partial_t u(t,X_t)dt - \frac{1}{2}\sigma^2 X_t^2 \partial_x^2 u(t,X_t)dt + \left(\Delta - \partial_x u(t,X_t)\right) dX_t\right) \\
&\quad + \alpha dt \\
&\quad - \lambda_t^D dt \left(u^D(t,X_t) - u(t,X_t)\right) \\
&\quad - \lambda_t^L dt \left(u^L(t,X_t) - u(t,X_t)\right) \\
&\quad + O\left(dt^2\right)
\end{aligned}
$$

- Choose $\Delta = \partial_x u(t,X_t)$ to hedge the risk from the underlying asset
- Mortality and lapse risks are hedged in average. The total P&L per contract vanishes in the limit by the strong law of large numbers when the number of contracts tends to infinity.

Deterministic lapse and mortality model - Monte Carlo

I. Direct simulation of default times

For simplicity, assume λ_t^D and λ_t^L are constants λ^D and λ^L respectively.

1. Simulate N random future market paths, denote by X_t^p, $0 \le t \le T$, the underlying asset price at time t on path p;
2. For each path p, simulate exponential random variables $\tau^{D,p}$ and $\tau^{L,p}$ with rates λ^D and λ^L respectively;
3. For each path p, compare $\tau^{D,p}$, $\tau^{L,p}$ and T, and then compute the payoff as follows:
   - if T is the smallest, payoff is $u^{\text{mat}}\left(T, X_T^p\right) - \alpha T$;
   - if τ^D is the smallest, payoff is $u^D\left(\tau^{D,p}, X^p_{\tau^{D,p}}\right) - \alpha\tau^{D,p}$;
   - if τ^L is the smallest, payoff is $u^L\left(\tau^{L,p}, X^p_{\tau^{L,p}}\right) - \alpha\tau^{L,p}$;
4. Averaging payoffs across all N paths gives an estimate of the fair value of the contract at time zero.

Deterministic lapse and mortality model - Monte Carlo

II. Averaging over default events

1. Simulate N random future market paths at dates $0 = t_0, \ldots, t_n = T$. On each path p, denote by $X^p_{t_k}$ the underlying asset price at time t_k, $C^p_{t_k}$ (or more precisely, our estimate of) the sum of all future cashflows from t_k onwards, averaging over death and lapse events, provided that no death or lapse has occurred up to time t_k.
2. At t_n, $C^p_{t_n} = u^{\text{mat}}\left(t_n, X^p_{t_n}\right)$
3. At t_k with $0 \leq k < n$,

$$
\begin{aligned}
C^p_{t_k} = & \exp\left(-\lambda^D_{t_{k+1}} \Delta t_{k+1}\right) \exp\left(-\lambda^L_{t_{k+1}} \Delta t_{k+1}\right) C^p_{t_{k+1}} \\
& - \alpha \Delta t_{k+1} \\
& + \left(1 - \exp\left(-\lambda^L_{t_{k+1}} \Delta t_{k+1}\right)\right) \exp\left(-\lambda^D_{t_{k+1}} \Delta t_{k+1}\right) u^L\left(t_{k+1}, X_{t_{k+1}}\right) \\
& + \left(1 - \exp\left(-\lambda^D_{t_{k+1}} \Delta t_{k+1}\right)\right) u^D\left(t_{k+1}, X_{t_{k+1}}\right)
\end{aligned}
$$

4. Averaging $C^p_{t_0}$ across all N paths gives an estimate of the fair value of the contract at time zero.

Uncertain Lapse and Mortality Model (ULMM)

Assume that λ_t^D and λ_t^L are unknown but adapted and belong to a moving corridor, i.e. $\underline{\lambda}^D(t) \leq \lambda_t^D \leq \overline{\lambda}^D(t)$ and $\underline{\lambda}^L(t) \leq \lambda_t^L \leq \overline{\lambda}^L(t)$.

$$u(t,x) = \sup_{\lambda^D,\lambda^L} \mathbb{E}^{\mathbb{Q}}_{t,x} \Big[e^{-\int_t^T (\lambda_s^D + \lambda_s^L) ds} u^{\mathrm{mat}}(X_T) + \int_t^T e^{-\int_t^s (\lambda_v^D + \lambda_v^L) dv} \left(-\alpha + u^D(s, X_s)\lambda_s^D + u^L(s, X_s)\lambda_s^L \right) ds \Big]$$

HJB equation:

$$\partial_t u + \frac{1}{2}\sigma^2 x^2 \partial_x^2 u - \alpha + \Lambda^D\left(t, u^D - u\right)\left(u^D - u\right) + \Lambda^L\left(t, u^L - u\right)\left(u^L - u\right) = 0$$

with terminal condition $u(T,x) = u^{\mathrm{mat}}(x)$, where

$$\Lambda^D(t,y) = \begin{cases} \overline{\lambda}^D(t) & \text{if } y \geq 0 \\ \underline{\lambda}^D(t) & \text{otherwise} \end{cases}$$

$$\Lambda^L(t,y) = \begin{cases} \overline{\lambda}^L(t) & \text{if } y \geq 0 \\ \underline{\lambda}^L(t) & \text{otherwise} \end{cases}$$

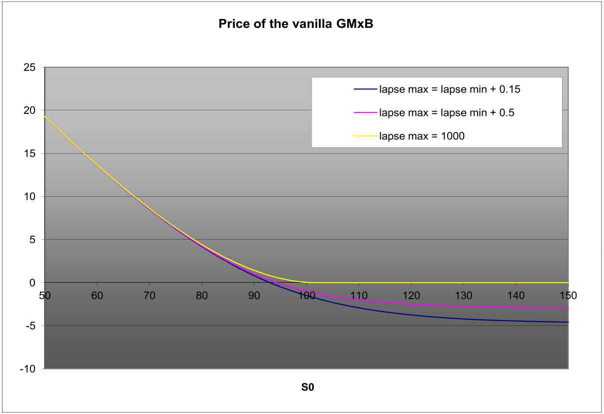
Uncertain Lapse and Mortality Model (ULMM)

The insurance subscriber can always lapse. Assume that the fee α matches exactly the managing fee that the subscriber pays to the insurance company. A rational policyholder should lapse as soon as $u^L > u$. A GMxB deal is therefore an American option. Why not price it as such? It is always possible to price reinsurance deals as American options, but the resulting price is often far from the market quotes. In fact, it has been observed that, so far, policyholders do not exercise optimally, which is taken into account through the $\overline{\lambda}^L$ function. The $\overline{\lambda}^L$ function accounts for the fact that policyholders may not lapse when they should. The $\underline{\lambda}^L$ function accounts for the fact that policyholders may lapse when they should not. Choosing $\overline{\lambda}^L = +\infty$ and $\underline{\lambda}^L = 0$ boils down to pricing the GMxB as an American option:

$$\max\left(\partial_t u + \frac{1}{2}\sigma^2 x^2 \partial_x^2 u - \alpha + \Lambda^D\left(t, u^D - u\right)\left(u^D - u\right), u^L - u\right) = 0$$

with terminal condition $u(T,x) = u^{\text{mat}}(x)$.

Numerical results



Longstaff-Schwartz-like method (G., 2007)

1. Simulate N_1 random future market paths at dates $t_0, \ldots, t_n = T$. On each path p, denote by
   - $X_{t_k}^p$ the underlying asset price at time t_k
   - $\lambda_{t_k}^{L,p}$ and $\lambda_{t_k}^{D,p}$ our estimates of the optimal lapse and death rates at time t_k
   - $C_{t_k}^p$ our estimate of the sum of all future cashflows from t_k onwards, averaging over death and lapse events with lapse and death rates $\lambda_{t_j}^{L,p}$ and $\lambda_{t_j}^{D,p}$ for $t_j > t_k$, provided that no death or lapse has occurred up to time t_k.
2. At date $t_n = T$,

$$
\begin{aligned}
C_{t_n}^p &= u^{\text{mat}}\left(t_n, X_{t_n}^p\right) \\
\lambda_{t_n}^{L,p} &= \Lambda^L\left(t_n, u^L\left(t_n, X_{t_n}^p\right) - C_{t_n}^p\right) \\
\lambda_{t_n}^{D,p} &= \Lambda^D\left(t_n, u^D\left(t_n, X_{t_n}^p\right) - C_{t_n}^p\right)
\end{aligned}
$$

Longstaff-Schwartz-like method (G., 2007)

3. At date t_k with $0 = k < n$,

$$
\begin{aligned}
C^p_{t_k} = {} & \exp\left(-\lambda^{L,p}_{t_{k+1}} \Delta t_{k+1}\right) \exp\left(-\lambda^{D,p}_{t_{k+1}} \Delta t_{k+1}\right) C^p_{t_{k+1}} \\
& - \alpha \Delta t_{k+1} \\
& + \left(1 - \exp\left(-\lambda^{L,p}_{t_{k+1}} \Delta t_{k+1}\right)\right) \exp\left(-\lambda^{D,p}_{t_{k+1}} \Delta t_{k+1}\right) u^L\left(t_{k+1}, X^p_{t_{k+1}}\right) \\
& + \left(1 - \exp\left(-\lambda^{D,p}_{t_{k+1}} \Delta t_{k+1}\right)\right) u^D\left(t_{k+1}, X^p_{t_{k+1}}\right)
\end{aligned}
$$

Longstaff-Schwartz-like method (G., 2007)

4. Find a functional estimate $\hat{u}\left(t_k, X^p_{t_k}\right)$ of $\mathbb{E}^{\mathbb{Q}}\left[C^p_{t_k} \middle| X^p_{t_k}\right]$ by parametric or non-parametric regression.

   Update the mortality and lapse rates at t_k by

$$\lambda^{L,p}_{t_k} = \begin{cases} \overline{\lambda}^L(t_k) & \text{if } u^L\left(t_k, X^p_{t_k}\right) - \hat{u}\left(t_k, X^p_{t_k}\right) \geq 0 \\ \underline{\lambda}^L(t_k) & \text{otherwise} \end{cases}$$

$$\lambda^{D,p}_{t_k} = \begin{cases} \overline{\lambda}^D(t_k) & \text{if } u^D\left(t_k, X^p_{t_k}\right) - \hat{u}\left(t_k, X^p_{t_k}\right) \geq 0 \\ \underline{\lambda}^D(t_k) & \text{otherwise} \end{cases}$$

   With these definitions of $\lambda^{L,p}_{t_k}$ and $\lambda^{D,p}_{t_k}$, we can now estimate the $C^p_{t_{k-1}}$ at t_{k-1}, etc. This is how the backward induction works.

Longstaff-Schwartz-like method

5. Simulate N_2 new independent paths to obtain a lower-bound price.

   At date $t_n = T$,

$$C^p_{t_n} = u^{\mathrm{mat}}\left(t_n, X^p_{t_n}\right)$$

   At date t_k, $k < n$,

$$\begin{aligned} C^p_{t_k} = &\exp\left(-\lambda^{L,p}_{t_{k+1}} \Delta t_{k+1}\right) \exp\left(-\lambda^{D,p}_{t_{k+1}} \Delta t_{k+1}\right) C^p_{t_{k+1}} \\ &- \alpha \Delta t_{k+1} \\ &+ \left(1 - \exp\left(-\lambda^{L,p}_{t_{k+1}} \Delta t_{k+1}\right)\right) \exp\left(-\lambda^{D,p}_{t_{k+1}} \Delta t_{k+1}\right) u^L\left(t_k, X^p_{t_{k+1}}\right) \\ &+ \left(1 - \exp\left(-\lambda^{D,p}_{t_{k+1}} \Delta t_{k+1}\right)\right) u^D\left(t_{k+1}, X^p_{t_{k+1}}\right) \end{aligned}$$

   where $\lambda^{L,p}_{t_k}$ and $\lambda^{D,p}_{t_k}$ $(1 \leq p \leq N_2)$ are determined by comparing $u^L\left(t_k, X^p_{t_k}\right)$ and $u^D\left(t_k, X^p_{t_k}\right)$ with the functional estimate $\hat{u}(t_k, X^p_{t_k})$ where $\hat{u}$ was computed during step 4.

   Averaging $C^p_{t_0}$ across all N_2 paths to get a lower-bound estimate.

BSDE for ULMM

$$\partial_t u + \frac{1}{2}\sigma^2 x^2 \partial_x^2 u - \alpha + \Lambda^D\left(t, u^D - u\right)\left(u^D - u\right) + \Lambda^L\left(t, u^L - u\right)\left(u^L - u\right) = 0$$

The nonlinear part of f involves u but not $\partial_x u, \partial_x^2 u$, we will use 1-BSDEs. $u(0,x)$ can be represented by the solution Y_0^x to the 1-BSDE

$$dY_t = -f(t, X_t, Y_t, Z_t)\,dt + Z_t\,dW_t$$

with the terminal condition $Y_T = u^{\mathrm{mat}}(X_T)$, where

$$dX_t = \sigma X_t\,dW_t, \qquad X_0 = x$$

and

$$f(t,x,y,z) = -\alpha + \Lambda^D\left(t, u^D(t,x) - y\right)\left(u^D(t,x) - y\right) + \Lambda^L\left(t, u^L(t,x) - y\right)\left(u^L(t,x) - y\right)$$

BSDE for ULMM

$$Y_{t_n}^{\Delta} = u^{\mathrm{mat}}(X_{t_n}^{\Delta})$$

Implicit Scheme

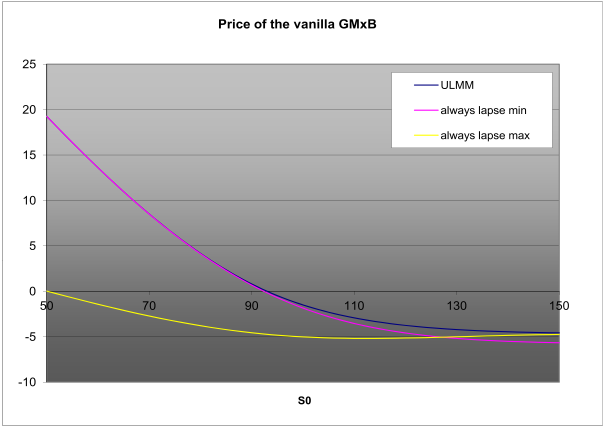
$$\begin{aligned} Y_{t_{i-1}}^{\Delta} &= \mathbb{E}_{i-1}^{\mathbb{Q}}[Y_{t_i}^{\Delta}] - \alpha \Delta t_i \\ &\quad + \Lambda^D\left(t_{i-1}, u^D(t_{i-1}, X_{t_{i-1}}^{\Delta}) - Y_{t_{i-1}}^{\Delta}\right)\left(u^D(t_{i-1}, X_{t_{i-1}}^{\Delta}) - Y_{t_{i-1}}^{\Delta}\right)\Delta t_i \\ &\quad + \Lambda^L\left(t_{i-1}, u^L(t_{i-1}, X_{t_{i-1}}^{\Delta}) - Y_{t_{i-1}}^{\Delta}\right)\left(u^L(t_{i-1}, X_{t_{i-1}}^{\Delta}) - Y_{t_{i-1}}^{\Delta}\right)\Delta t_i \end{aligned}$$

Explicit Scheme

$$\begin{aligned} Y_{t_{i-1}}^{\Delta} &= \mathbb{E}_{i-1}^{\mathbb{Q}}\Big[Y_{t_i}^{\Delta} - \alpha \Delta t_i \\ &\quad + \Lambda^D\left(t_i, u^D(t_i, X_{t_i}^{\Delta}) - Y_{t_i}^{\Delta}\right)\left(u^D(t_i, X_{t_i}^{\Delta}) - Y_{t_i}^{\Delta}\right)\Delta t_i \\ &\quad + \Lambda^L\left(t_i, u^L(t_i, X_{t_i}^{\Delta}) - Y_{t_i}^{\Delta}\right)\left(u^L(t_i, X_{t_i}^{\Delta}) - Y_{t_i}^{\Delta}\right)\Delta t_i\Big] \end{aligned}$$

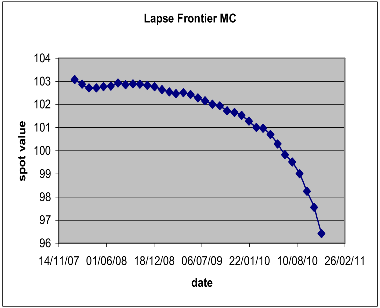
Note: No need to compute $Z_{t_{i-1}}^{\Delta}$ as f does not depend on z.

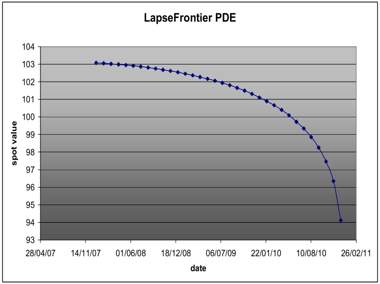
Numerical results



Numerical results

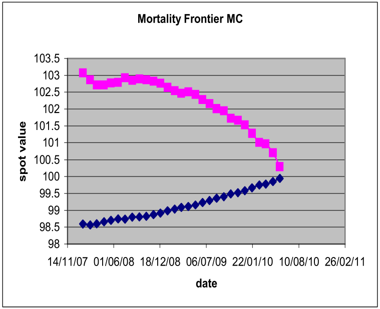
Optimal lapse frontier in the time-spot plane. Left: using MC. Right: using PDE

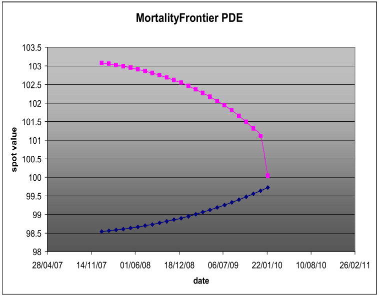




Numerical results

Optimal mortality frontier in the time-spot plane. Left: using MC. Right: using PDE





Uncertain Volatility Model

Il n'est pas certain que tout soit incertain.[3]
— Blaise Pascal

[3] "It is not certain that everything is uncertain."

The 1D Uncertain Volatility Model (UVM)

- Avellaneda-al (1995), Lyons (1995)

$$dX_t = \sigma_t X_t dW_t, \quad \sigma_t \in [\underline{\sigma}, \overline{\sigma}], \quad V_t = \sup_{[t,T]} \mathbb{E}[g(X_T)|\mathcal{F}_t]$$

where the supremum is taken over all $(\mathcal{F}_s)$-adapted processes $(\sigma_s)_{t \leq s \leq T}$ such that for all $s \in [t,T]$, $\sigma_s \in [\underline{\sigma}, \overline{\sigma}]$.

- Worst-case scenario for the sell-side
- HJB: $V_t = u(t, X_t)$ with

$$\partial_t u(t,x) + \frac{1}{2} \sup_{\sigma} \left(x^2 \sigma^2 \partial_x^2 u(t,x) \right) = 0$$

Black-Scholes-Barenblatt (BSB) PDE:

$$\partial_t u(t,x) + \frac{1}{2} x^2 \Sigma \left(\partial_x^2 u \right)^2 \partial_x^2 u(t,x) = 0, \quad (t,x) \in [0,T) \times \mathbb{R}_+^* \qquad (39)$$

with $u(T,x) = g(x)$ and $\Sigma(\Gamma)^2 = \overline{\sigma}^2 1_{\Gamma \geq 0} + \underline{\sigma}^2 1_{\Gamma < 0}$.

Analogies

- **Transaction cost** [Leland]: Same equation with

$$\Sigma(\Gamma)^2 = \sigma^2 + \sqrt{\frac{2}{\pi}}\frac{k\sigma}{\sqrt{\delta t}}\mathrm{sign}(\Gamma)$$

- **Pricing in illiquid market** [Willmott-al]: Same equation with

$$\Sigma(\Gamma)^2 = \frac{\sigma^2}{(1-\epsilon S\Gamma)^2}$$

Leland's Approach to Transaction Costs

Assume we sold an option that we delta-hedge. When re-hedging from t to $t+\delta t$, the delta position needs to be re-adjusted from Δ_t to $\Delta_{t+\delta t}$. The cost associated to either buying or selling is assumed to be $kS_t|\delta\Delta_t|$, where k is a constant depending on the liquidity of the underlying.

Note $\pi_t = \Delta_t S_t + \varphi_t B_t$ the self-financing portfolio. The hedging error between t and $t+\delta t$, assuming zero rates and dividends, is given by $\varepsilon_t = \pi_t - u(t, S_t)$. That gives the P&L between t and $t+\delta t$ to be

$$\begin{aligned}
\delta\varepsilon_t &= \delta\pi_t - \delta u(t, S_t) \\
&= (\Delta_t - \partial_s u(t, S_t))\delta S_t - kS_t|\delta\Delta_t| \\
&\quad - \left(u(t, S_t)\partial_t u(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 \partial_s u(t, S_t)\right)\delta t + O(\delta t)
\end{aligned}$$

Leland's Approach to Transaction Costs

The (conditional at time t) average hedging cost is:

$$\mathbb{E}_t(\delta\varepsilon_t) = -\left(u(t,S_t)\partial_t u(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 \partial_{ss} u(t,S_t)\right)\delta t - kS_t\mathbb{E}_t(|\delta\Delta_t|)$$

We need to evaluate the term

$$\mathbb{E}_t(|\delta\Delta_t|) = \mathbb{E}_t\left(\left|\partial_t\Delta_t\delta t + \frac{1}{2}\sigma^2 S_t^2\partial_s^2\Delta_t\delta t + \partial_s\Delta_s\sigma S_t\delta W_t\right|\right)$$

Recall that $\mathbb{E}(|\alpha+\beta Z|) = \alpha(2\Phi(\frac{\alpha}{\beta})-1) + |\beta|\sqrt{\frac{2}{\pi}}e^{-\frac{1}{2}\left(\frac{\alpha}{\beta}\right)^2}$ and as δt is negligible before $\sqrt{\delta t}$, we get

$$\mathbb{E}_t(|\delta\Delta_t|) = \sigma S_t|\partial_s^2 u(t,S_t)|\sqrt{\frac{2}{\pi}}\sqrt{\delta t} + O(\sqrt{\delta}t)$$

Leland's Approach to Transaction Costs

Considering the natural constraint of having a zero P&L at order δt leads to

$$\partial_t u(t,S_t) + \frac{1}{2} S_t^2 \partial_s^2 u(t,S_t) \left(\sigma^2 + \sigma \mathrm{sign}\left(\partial_s^2 u(t,S_t)\right) 2 \frac{k\sigma}{\sqrt{\delta t}} S_t^2 \sqrt{\frac{2}{\pi}} \right) = 0$$

We note the term $\frac{k\sigma}{\sqrt{\delta t}}$ would got to $+\infty$ as $\delta t \to 0$ and seems problematic. Many justifications are given in the literature, ranging from "let's heuristically just do that, the math works out" to something more plausible such as "the constant k in fact depends on the re-hedging frequency and would tend to be smaller making the ratio $\frac{k}{\sqrt{\delta t}}$ rather constant".

Pricing in Illiquid Market

Let's assume that when observing the market, the underlying evolves as $dX_t = \mu dt + \sigma X_t dW_t$. As we follow a delta-hedging strategy $\Delta_t = \Delta(t, X_t)$, when re-balancing our portfolio we need to buy a quantity $d\Delta_t$ of shares. If this quantity if positive, the stock price will go slightly up, and it will go (slightly) otherwise. This leads us to consider the following dynamics for the stock price

$$dX_t = \mu X_t dt + \sigma X_t dW_t + \varepsilon\varphi(X_t) d\Delta_t$$

We note $\Gamma_t = \Gamma(t, X_t) = \partial_x \Delta(t, X_t)$. Using Ito's formula, we formally get

$$\begin{aligned} dX_t &= \mu X_t dt + \sigma X_t dW_t + \varepsilon\varphi(X_t)(\partial_t \Delta_t dt + \frac{1}{2}\partial_x^2 \Delta_t d\langle X\rangle_t + \partial_x \Delta_t dX_t) \\ &\Rightarrow (1 - \varepsilon\varphi(X_t)\Gamma_t) dX_t = \tilde{\mu} X_t dt + \sigma X_t dW_t \\ &\Rightarrow dX_t = \frac{\tilde{\mu}}{1 - \varepsilon\varphi(X_t)\Gamma_t} X_t dt + \frac{\sigma X_t}{1 - \varepsilon\varphi(X_t)\Gamma_t} dW_t \end{aligned}$$

Pricing in Illiquid Market

In particular

$$d\langle X\rangle_t = \left(\frac{\sigma X_t}{(1-\varepsilon\varphi(X_t)\Gamma_t)}\right)^2 dt$$

In a complete market setting with zero rates and dividends with $\varphi(x) = x$ (market impact are fully proportional to the share price and quantity of shares transacted), the pricing PDE then becomes

$$\partial_t u(t,x) + \frac{1}{2}\left(\frac{\sigma X_t}{1-\varepsilon x\partial_x^2 u(t,x)}\right)^2 \partial_x^2 u(t,x) = 0$$

Solving the pricing equation

1D Black-Scholes-Barenblatt (BSB) PDE:

$$\partial_t u(t,x) + \frac{1}{2} x^2 \Sigma(\partial_x^2 u)^2 \partial_x^2 u(t,x) = 0, \quad (t,x) \in [0,T) \times \mathbb{R}_+^*$$

with $u(T,x) = g(x)$ and $\Sigma(\Gamma)^2 = \overline{\sigma}^2 1_{\Gamma \geq 0} + \underline{\sigma}^2 1_{\Gamma < 0}$

- No analytical solution
- Finite-difference schemes
- Monte Carlo compulsory when the number of variables (underlyings and path-dep variables) is 3 or more

Robust super-replication

- The buyer's and seller's super-replication prices have been defined w.r.t. the historical measure $\mathbb{P}^{\text{hist}}$ (the definition involves the space of ELMM $\sim \mathbb{P}^{\text{hist}}$), meaning that the volatility process of the asset is fixed.
- In practice, the volatility process is unknown. In the uncertain volatility framework, we consider instead the set $\mathcal{P}$ of all measures $\mathbb{Q}$ under which X_t is a local martingale such that

$$\underline{\sigma}^2 \leq \frac{d\langle \ln X \rangle_t}{dt} \equiv \sigma_t^2 \leq \overline{\sigma}^2$$

- The seller's super-replication price in the UVM is defined by

$$\mathcal{S}_t(F_T) = \inf \left\{ z \in \mathcal{F}_t \,\middle|\, \text{there exists an admissible portfolio } \Delta \text{ such that} \right.$$
$$\left. \pi_T^S \equiv z + \int_t^T \Delta_s \, dX_s - F_T \geq 0 \quad \mathbb{Q}\text{−a.s} \quad \forall \mathbb{Q} \in \mathcal{P} \right\} \tag{40}$$

Robust super-replication

$$\mathcal{S}_t(F_T) = \inf \Big\{ z \in \mathcal{F}_t \,\Big|\, \text{there exists an admissible portfolio } \Delta \text{ such that} \\ \pi_T^S \equiv z + \int_t^T \Delta_s \, dX_s - F_T \geq 0 \quad \mathbb{Q}\text{−a.s} \quad \forall \mathbb{Q} \in \mathcal{P} \Big\}$$

Theorem

$$\mathcal{S}_t(F_T) \quad = \quad \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[F_T | \mathcal{F}_t] \quad = \quad \sup_{[t,T]} \mathbb{E}[F_T | \mathcal{F}_t]$$

This theorem justifies our definition of the UVM price.

Robust super-replication: proof

Proof. We assume that $F_T = g(X_T)$ with $g \in C_b^3(\mathbb{R}_+)$ and $\overline{\sigma} < \infty$.

(i) We set $z = u(t, X_t) = \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}^{\mathbb{Q}}[g(X_T)|\mathcal{F}_t]$ and $\Delta_s = \partial_x u(s, X_s)$ with u the solution of PDE (39). As $u \in C^{1,2}([0,T) \times \mathbb{R}_+)$, Itô's lemma gives

$$z + \int_t^T \Delta_s \, dX_s - g(X_T) = \int_t^T \Big(-\partial_s u(s, X_s) - \frac{1}{2} X_s^2 \sigma_s^2 \partial_x^2 u(s, X_s) \Big) \, ds$$

Using the PDE (39), we get

$$z + \int_t^T \Delta_s \, dX_s - g(X_T) = \frac{1}{2} \int_t^T X_s^2 \partial_x^2 u(s, X_s) (\Sigma(\partial_x^2 u(s, X_s))^2 - \sigma_s^2) \, ds \geq 0$$

$\mathbb{Q}$-a.s. for all $\mathbb{Q} \in \mathcal{P}$ as $\left(\Sigma\left(\partial_x^2 u(s, X_s)\right)^2 - \sigma_s^2\right) \partial_x^2 u(s, X_s) \geq 0$ for all adapted σ_s such that for all $s \in [t, T]$, $\sigma_s \in [\underline{\sigma}, \overline{\sigma}]$. This implies that $\mathcal{S}_t(g(X_T)) \leq u(t, X_t) \equiv \sup_{\mathbb{Q}\in\mathcal{P}} \mathbb{E}^{\mathbb{Q}}[g(X_T)|\mathcal{F}_t]$ as we have obtained that the delta-hedging strategy $(z = u(t, X_t), \Delta_s = \partial_x u(s, X_s))$ super-replicates the payoff.

Robust super-replication: proof

(ii) (Weak duality) Let z be such that there exists an admissible portfolio Δ such that

$$z + \int_t^T \Delta_s \, dX_s - g(X_T) \geq 0 \quad \mathbb{Q}\text{−a.s} \quad \forall \mathbb{Q} \in \mathcal{P}$$

By taking the conditional expectation we obtain $z \geq \mathbb{E}^{\mathbb{Q}}[g(X_T)|\mathcal{F}_t]$ for all $\mathbb{Q} \in \mathcal{P}$. We have used that the local martingale $\int_t^T \Delta_s \, dX_s$ bounded from below is a supermartingale. This implies that

$$\mathcal{S}_t(g(X_T)) \geq \sup_{\mathbb{Q} \in \mathcal{P}} \mathbb{E}^{\mathbb{Q}}[g(X_T)|\mathcal{F}_t]$$

The multidim uncertain volatility model

$$V_t = \sup_{[t,T]} \mathbb{E}[F_T|\mathcal{F}_t], \quad dX_t^\alpha = \sigma_t^\alpha X_t^\alpha dW_t^\alpha, \quad dW_t^\alpha dW_t^\beta = \rho_t^{\alpha\beta} dt$$

- Supremum taken over all $(\mathcal{F}_s)$-adapted processes $(\xi_s)_{t\le s\le T} \equiv ((\sigma_s^\alpha, \rho_s^{\alpha\beta})_{1\le\alpha<\beta\le d})_{t\le s\le T}$ such that for all $s \in [t,T]$, ξ_s belongs to some compact domain D
- We will consider domains D of the form $D = [\underline{\sigma}, \overline{\sigma}]$ when $d = 1$, and $D = [\underline{\sigma}^1, \overline{\sigma}^1] \times [\underline{\sigma}^2, \overline{\sigma}^2] \times [\underline{\rho}, \overline{\rho}]$ when $d = 2$

Pricing vanilla options

Consider vanilla payoff $F_T = g(X_T)$. We have $u_t = u(t, X_t)$ where $u(\cdot,\cdot)$ is the unique (viscosity) solution with quadratic growth of the nonlinear PDE

$$\partial_t u(t,x) + H(x, D_x^2 u(t,x)) = 0, \qquad (t,x) \in [0,T) \times (\mathbb{R}_+^*)^d \tag{41}$$

with the terminal condition $u(T,x) = g(x)$ and the Hamiltonian

$$H(X,\Gamma) = \frac{1}{2} \max_{(\sigma^\alpha, \rho^{\alpha\beta})_{1\le\alpha<\beta\le d} \in D} \sum_{\alpha,\beta=1}^{d} \rho^{\alpha\beta}\sigma^\alpha\sigma^\beta X^\alpha X^\beta \Gamma^{\alpha\beta} \tag{42}$$

If $g \in C^3((\mathbb{R}_+^*)^d)$ and $g, D_x g$ have quadratic growth, then $u \in C^{1,2}([0,T] \times (\mathbb{R}_+^*)^d)$.

Pricing vanilla options - Computation of the Hamiltonian

When $d=2$, the Hamiltonian reads

$$H(X,\Gamma)=\max_{(\sigma^1,\sigma^2,\rho)\in D}\left[\frac{1}{2}\left(\sigma^1\right)^2(X^1)^2\Gamma^{11}+\frac{1}{2}\left(\sigma^2\right)^2\left(X^2\right)^2\Gamma^{22}+\rho\sigma^1\sigma^2X^1X^2\Gamma^{12}\right]$$

where $D=\left[\underline{\sigma}^1,\overline{\sigma}^1\right]\times\left[\underline{\sigma}^2,\overline{\sigma}^2\right]\times\left[\underline{\rho},\overline{\rho}\right]$. The optimal correlation is bang-bang:

$$R(\Gamma^{12})=\underline{\rho}1_{\Gamma^{12}<0}+\overline{\rho}1_{\Gamma^{12}\geq 0}.$$

Then the problem becomes a two-dimensional quadratic form maximization under a double inequality constraint, which can be solved by first freezing σ^2 and maximize in σ^1 and then compute the maximum over σ^2.

When $d>2$, maximize in ρ_{ij} first and then solve the quadratic programming problem.

Pricing path-dependent options

- We consider the price of an option that depending on path-dependent variables whose values can change continuously.
- The Hamiltonian may not involve only the gammas
- Example: if $u(t, x, v)$ depends on the continuously compounded realised variance v,

$$H(x, \partial_x^2 u, \partial_v u) = \max_{\underline{\sigma} \leq \sigma \leq \overline{\sigma}} \sigma^2 \left(\frac{1}{2} x^2 \partial_x^2 u + \partial_v u \right) = \Sigma^2 \left(\frac{1}{2} x^2 \partial_x^2 u + \partial_v u \right)$$

The BSDE approach: Probabilistic representation of PDEs

- The classical link between Monte Carlo and finite-difference methods as stated by the Feynman-Kac formula is only valid for linear second-order parabolic PDEs
- Cheridito-Touzi-Soner-Victoir provide a stochastic representation for solutions of fully nonlinear parabolic PDEs by introducing a new class of BSDEs, the so-called second-order BSDEs

2-BSDE [Cheridito-al]

- Let $(Y_t, Z_t, \Gamma_t, \alpha_t)_{t\in[0,T]}$ a quadruple of $\mathcal{F}_t$-adapted processes taking values in $\mathbb{R}$, $\mathbb{R}^d$, S^d (space of d-dimensional symmetric matrices) and $\mathbb{R}^d$ respectively. Then, we call (Y, Z, Γ, α) a solution of a 2-BSDE if

$$
\begin{aligned}
dX_t &= \sigma(t, X_t)dW_t \\
dY_t &= -f(t, X_t, Y_t, Z_t, \Gamma_t)dt + Z_t \circ \sigma(t, X_t)dW_t \\
dZ_t &= \alpha_t dt + \Gamma_t \sigma(t, X_t)dW_t \\
Y_T &= g(X_T)
\end{aligned} \tag{43}
$$

  where $\circ$ is the Stratonovich integral

- The use of the Stratonovich product is only for convenience and can be replaced by an Itô integral

$$
dY_t = \left(-f(t, X_t, Y_t, Z_t, \Gamma_t) + \frac{1}{2}\mathrm{tr}\left(\sigma(t, X_t)\sigma(t, X_t)'\Gamma_t\right)\right) dt + Z_t.\sigma(t, X_t)\, dW_t
$$

Generalization of the Feynman-Kac formula

Let us consider the fully nonlinear parabolic PDE

$$\begin{aligned} \partial_t u(t,x) + f(t,x,u,\partial_x u,\partial_x^2 u) &= 0 \\ u(T,x) &= g(x) \end{aligned} \tag{44}$$

We recall

Theorem

Let u be a function smooth (enough to apply Itô's formula) satisfying the above PDE. Then $(Y_t = u(t,X_t), Z_t = \partial_x u(t,X_t), \Gamma_t = \partial_x^2 u(t,X_t), \alpha_t = (\partial_t + \mathcal{L})\,\partial_x u(t,X_t))$ is a solution of the 2-BSDE (43).

Y = price, Z = delta, Γ = gamma

Discretization scheme for 2-BSDEs

- Numerical simulation of 2-BSDEs = a way to build approximations of solutions to PDE (44)
- Proceeding by analogy with 1-BSDEs, we obtain the following (implicit) discretization scheme for 2-BSDEs (Cheridito *et al.*):
  Scheme 2-BSDE:

$$
\begin{aligned}
Y_{t_n}^{\Delta} &= g(X_{t_n}^{\Delta}) \\
Z_{t_n}^{\Delta} &= Dg(X_{t_n}^{\Delta}) \\
Y_{t_{i-1}}^{\Delta} &= \mathbb{E}_{i-1}[Y_{t_i}^{\Delta}] + \Big(f(t_{i-1}, X_{t_{i-1}}^{\Delta}, Y_{t_{i-1}}^{\Delta}, Z_{t_{i-1}}^{\Delta}, \Gamma_{t_{i-1}}^{\Delta}) \\
&\quad - \frac{1}{2} \mathrm{tr}[\sigma(t_{i-1}, X_{t_{i-1}}^{\Delta}) \sigma(t_{i-1}, X_{t_{i-1}}^{\Delta})' \Gamma_{t_{i-1}}^{\Delta}] \Big) \Delta t_i \\
Z_{t_{i-1}}^{\Delta} &= \frac{1}{\Delta t_i} \sigma(t_{i-1}, X_{t_{i-1}}^{\Delta})'^{-1} \mathbb{E}_{i-1}[Y_{t_i}^{\Delta} \Delta W_{t_i}] \\
\Gamma_{t_{i-1}}^{\Delta} &= \frac{1}{\Delta t_i} \mathbb{E}_{i-1}[Z_{t_i}^{\Delta} \Delta W_{t_i}'] \sigma(t_{i-1}, X_{t_{i-1}}^{\Delta})^{-1}
\end{aligned}
\tag{45}
$$

Discretization scheme for 2-BSDEs

One can also consider explicit schemes like

$$Y^{\Delta}_{t_{i-1}} = \mathbb{E}_{i-1}\Big[Y^{\Delta}_{t_i} + \Big(f(t_i, X^{\Delta}_{t_i}, Y^{\Delta}_{t_i}, Z^{\Delta}_{t_i}, \Gamma^{\Delta}_{t_i}) - \frac{1}{2}\mathrm{tr}[\sigma(t_i, X^{\Delta}_{t_i})\sigma(t_i, X^{\Delta}_{t_i})'\Gamma^{\Delta}_{t_i}]\Big)\Delta t_i\Big] \quad (46)$$

or

$$Y^{\Delta}_{t_{i-1}} = \mathbb{E}_{i-1}[Y^{\Delta}_{t_i}] + \Big(f(t_{i-1}, X^{\Delta}_{t_{i-1}}, \mathbb{E}_{i-1}[Y^{\Delta}_{t_i}], Z^{\Delta}_{t_{i-1}}, \Gamma^{\Delta}_{t_{i-1}}) - \frac{1}{2}\mathrm{tr}[\sigma(t_{i-1}, X^{\Delta}_{t_{i-1}})\sigma(t_{i-1}, X^{\Delta}_{t_{i-1}})'\Gamma^{\Delta}_{t_{i-1}}]\Big)\Delta t_i \quad (47)$$

or linear combinations of those (θ-schemes).

Discretization scheme for 2-BSDEs

- The quantity $\mathrm{P\&L} \times \Delta t_i$ where

$$\mathrm{P\&L} \equiv f(t_{i-1}, X^{\Delta}_{t_{i-1}}, Y^{\Delta}_{t_{i-1}}, Z^{\Delta}_{t_{i-1}}, \Gamma^{\Delta}_{t_{i-1}}) - \frac{1}{2}\mathrm{tr}[\sigma(t_{i-1}, X^{\Delta}_{t_{i-1}})\sigma(t_{i-1}, X^{\Delta}_{t_{i-1}})'\Gamma^{\Delta}_{t_{i-1}}] \quad (48)$$

  is the so-called gamma-theta P&L between t_{i-1} and t_i.
- Note that PDE (44) does not depend on the function σ (volatility of forward process X), which can be chosen arbitrarily.
- This scheme requires a final condition for $Z_{t_n} = Dg(X_{t_n})$. Since the payoffs usually considered in finance are not smooth, this scheme may perform poorly. We now suggest a new scheme in the case of the UVM.
- Using an induction argument, one can show that the random variables $Y^{\Delta}_{t_i}$, $Z^{\Delta}_{t_i}$, and $\Gamma^{\Delta}_{t_i}$ are deterministic functions of $X^{\Delta}_{t_i}$. Hence the conditional expectations above can be replaced by

$$\begin{aligned} \mathbb{E}_{i-1}[Y^{\Delta}_{t_i}\Delta W_{t_i}] &= \mathbb{E}[Y^{\Delta}_{t_i}\Delta W_{t_i}|X^{\Delta}_{t_{i-1}}] \\ \mathbb{E}_{i-1}[Z^{\Delta}_{t_i}\Delta W'_{t_i}] &= \mathbb{E}[Z^{\Delta}_{t_i}\Delta W'_{t_i}|X^{\Delta}_{t_{i-1}}] \end{aligned}$$

Discretization scheme for the BSB 2-BSDE

Scheme 2-BSDE for the UVM (dropping superscript Δ for clarity):

$$
\begin{aligned}
X^\alpha_{t_i} &= X^\alpha_0 e^{-(\hat\sigma^\alpha)^2 \frac{t_i}{2} + \hat\sigma^\alpha W^\alpha_{t_i}}, \qquad \mathbb{E}[\Delta W^\alpha_{t_i} \Delta W^\beta_{t_i}] = \hat\rho^{\alpha\beta} \Delta t_i \\
Y_{t_n} &= g(X_{t_n}) \qquad (49) \\
Y_{t_{i-1}} &= \mathbb{E}[Y_{t_i} | X_{t_{i-1}}] + \left(f(X_{t_{i-1}}, \Gamma_{t_{i-1}}) - \frac{1}{2} \sum_{\alpha,\beta=1}^{d} \hat\rho^{\alpha\beta} \hat\sigma^\alpha \hat\sigma^\beta X^\alpha_{t_{i-1}} X^\beta_{t_{i-1}} \Gamma^{\alpha\beta}_{t_{i-1}} \right) \Delta t_i \\
(\Delta t_i)^2 \hat\sigma^\alpha \hat\sigma^\beta X^\alpha_{t_{i-1}} X^\beta_{t_{i-1}} \Gamma^{\alpha\beta}_{t_{i-1}} &= \mathbb{E}\left[Y_{t_i} \left(U^\alpha_{t_i} U^\beta_{t_i} - \Delta t_i \hat\rho^{-1}_{\alpha\beta} - \Delta t_i \hat\sigma^\alpha U^\alpha_{t_i} \delta_{\alpha\beta} \right) \middle| X_{t_{i-1}} \right]
\end{aligned}
$$

with $U^\alpha_{t_i} \equiv \sum_{\beta=1}^d \hat\rho^{-1}_{\alpha\beta} \Delta W^\beta_{t_i}$

- We have introduced explicitly the Malliavin weight for a log-normal diffusion with volatility $\hat\sigma$ and correlation $\hat\rho$
- Analogy with American options: $Y_{t_{i-1}} = \max\left(\mathbb{E}^{\mathbb{Q}}_{i-1}[Y_{t_i}], g(X_{t_{i-1}})\right)$

Discretization Scheme for the BSB 2-BSDE in 1D

Scheme 2-BSDE for the UVM in 1D:

$$
\begin{aligned}
X_{t_i} &= X_0 e^{-\hat{\sigma}^2 \frac{t_i}{2} + \hat{\sigma} W_{t_i}} \\
Y_{t_n} &= g(X_{t_n}) \\
Y_{t_{i-1}} &= \mathbb{E}[Y_{t_i} | X_{t_{i-1}}] + \left(f(X_{t_{i-1}}, \Gamma_{t_{i-1}}) - \frac{1}{2} \hat{\sigma}^2 X_{t_{i-1}}^2 \Gamma_{t_{i-1}} \right) \Delta t_i \\
&= \mathbb{E}[Y_{t_i} | X_{t_{i-1}}] + \frac{1}{2} (\Sigma(\Gamma_{t_{i-1}})^2 - \hat{\sigma}^2) X_{t_{i-1}}^2 \Gamma_{t_{i-1}} \Delta t_i \\
(\Delta t_i)^2 \hat{\sigma}^2 X_{t_{i-1}}^2 \Gamma_{t_{i-1}} &= \mathbb{E}\left[Y_{t_i} \left(\Delta W_{t_i}^2 - \Delta t_i (1 + \hat{\sigma} \Delta W_{t_i}) \right) | X_{t_{i-1}} \right]
\end{aligned}
\tag{50}
$$

Malliavin Weights for Black-Scholes Delta and Gamma

Let $h(x) = \mathbb{E}\left[g\left(xe^{\sigma W_T - \frac{1}{2}\sigma^2 T}\right)\right]$. Then

$$h'(x) = \frac{1}{\sigma x T}\mathbb{E}\left[g\left(xe^{\sigma W_T - \frac{1}{2}\sigma^2 T}\right) W_T\right]$$

$$h''(x) = \frac{1}{\sigma^2 x^2 T^2}\mathbb{E}\left[g\left(xe^{\sigma W_T - \frac{1}{2}\sigma^2 T}\right)\left(W_T^2 - T\left(1 + \sigma W_T\right)\right)\right]$$

Proof.

Let $n(y) = \frac{1}{\sqrt{2\pi T}}e^{-\frac{y^2}{2T}}$ be the probability density function of W_T. Then

$$h(x) = \int_{\mathbb{R}} g\left(xe^{\sigma y - \frac{1}{2}\sigma^2 T}\right) n(y)dy.$$

$$h'(x) = \int_{\mathbb{R}} g'\left(xe^{\sigma y - \frac{1}{2}\sigma^2 T}\right) e^{\sigma y - \frac{1}{2}\sigma^2 T} n(y)dy = \int_{\mathbb{R}} \frac{d}{dy}\left(g\left(xe^{\sigma y - \frac{1}{2}\sigma^2 T}\right)\right)\frac{1}{\sigma x} n(y)dy$$

$$= -\frac{1}{\sigma x}\int_{\mathbb{R}} g\left(xe^{\sigma y - \frac{1}{2}\sigma^2 T}\right) n'(y)dy = \frac{1}{\sigma x}\int_{\mathbb{R}} g\left(xe^{\sigma y - \frac{1}{2}\sigma^2 T}\right)\frac{y}{T} n(y)dy$$

$$= \frac{1}{\sigma x T}\mathbb{E}\left[g\left(xe^{\sigma W_T - \frac{1}{2}\sigma^2 T}\right) W_T\right]$$

The expression for h'' can be proved similarly. □

Interpretation

- Consider the case where $i = n$: since we simulated lognormal X's, the price $Y_{t_{n-1}}$ is the sum of the Black-Scholes price $\mathbb{E}^{\mathbb{Q}}[Y_{t_n} | X_{t_{n-1}}]$ and the gamma-theta P&L correction

$$\left(f(X_{t_{n-1}}, \Gamma_{t_{n-1}}) - \frac{1}{2} \sum_{\alpha,\beta=1}^{d} \hat{\rho}^{\alpha\beta} \hat{\sigma}^{\alpha} \hat{\sigma}^{\beta} X^{\alpha}_{t_{n-1}} X^{\beta}_{t_{n-1}} \Gamma^{\alpha\beta}_{t_{n-1}} \right) \Delta t_n \qquad (51)$$

- This last term requires that we estimate the gamma at time t_{n-1}. We use the Black-Scholes gamma, as given by the last equation in (49), which uses the Malliavin weight for the lognormal diffusion.
- In dimension 1, the gamma-theta P&L correction reads

$$\left(f(X_{t_{n-1}}, \Gamma_{t_{n-1}}) - \frac{1}{2} \hat{\sigma}^2 X^2_{t_{n-1}} \Gamma_{t_{n-1}} \right) \Delta t_n = \frac{1}{2} X^2_{t_{n-1}} (\Sigma(\Gamma_{t_{n-1}})^2 - \hat{\sigma}^2) \Gamma_{t_{n-1}} \Delta t_n$$

with $\Sigma(\Gamma)^2 = \overline{\sigma}^2 1_{\Gamma \geq 0} + \underline{\sigma}^2 1_{\Gamma < 0}$

The algorithm (G. and H.-L., 2008)

- Simulate N_1 replications of X with a log-normal diffusion.
- Apply the backward algorithm using a regression approximation.
- Simulate N_2 independent replications of X using the gamma functions $\Gamma_{t_i} = \phi(t_i, X_{t_i})$ which are the result of the regression at the previous step.

Numerical experiment: Call on one underlying $(X_T - 100)^+$

At-the money call option valued using the BSDE approach with $\hat{\sigma} = 0.15$. The true price is $u = 7.97$

Δ	M_1	12	13	14	15	16	17
$1/2$	Price	8.04	8.06	8.00	8.00	8.00	8.00
$1/4$	Price	8.53	8.29	8.00	7.87	7.89	7.86
$1/8$	Price	9.28	8.72	8.02	7.79	7.78	7.68

At-the money call option valued using the BSDE approach with $\hat{\sigma} = 0.15$ and an independent second MC run. The true price is $u = 7.97$

Δ	M_1	12	13	14	15	16	17
$1/2$	Price	7.93	7.94	7.95	7.95	**7.96**	**7.96**
$1/4$	Price	7.88	7.90	7.92	7.93	7.95	7.94
$1/8$	Price	7.53	7.93	7.58	7.60	**7.96**	**7.96**

Numerical experiments

- $T = 1$, $t_i = i/n$, $\Delta = 1/n$
- For each asset α, $X_0^\alpha = 100$, $\underline{\sigma}^\alpha = 0.1$, $\overline{\sigma}^\alpha = 0.2$ and $\hat{\sigma}^\alpha = 0.15$
- We compare with a parametric approach (not presented in this course)
- **Param**: In the gamma calibration stage, we pick $N_1 = 2^{M_1}$ with $M_1 = 12$, and the N_1 replications of X use a time step $t_{k+1} - t_k = 1/100$
- **BSDE**: We allow M_1 to vary from 12 to 17
- Pricing stage: the $N_2 = 2^{15}$ replications of X use a time step $\Delta_2 = 1/400$

Call spread: $(X_T - 90)^+ - (X_T - 110)^+$

- The true price (PDE) is $u_{\text{PDE}} = 11.20$
- **Param**: We pick $\theta_i \in \mathbb{R}$ and $\lambda_{t_i}(X;\theta_i) = \overline{\sigma}$ if $\theta_i - \ln(X/X_0) \geq 0$, $= \underline{\sigma}$ otherwise
- **BSDE**: we use non-parametric regressions

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	**11.19**	Price	11.08	11.07	11.06	11.06	11.06	11.06
$1/4$	**11.19**	Price	11.01	**11.12**	11.06	11.07	11.11	11.11
$1/8$	11.18	Price	10.74	10.55	10.73	11.01	11.04	11.11

Algo/$\hat{\sigma}$	2%	5%	10%	15%	20%	30%	50%
Param	**11.19**	**11.19**	**11.19**	**11.19**	**11.19**	**11.19**	11.14
BSDE	9.36	10.72	11.01	**11.12**	11.07	11.00	10.81
Black-Scholes	10.00	9.97	9.76	9.52	9.30	8.87	8.06

Digital option: $100 \times 1_{X_T \geq K}$

- The true price (PDE) is $u_{\mathrm{PDE}} = 63.33$
- **Param**: same parameterization as for the call spread
- **BSDE**: we use non-parametric regressions

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	63.13	Price	**62.83**	62.83	62.74	62.75	62.75	62.74
$1/4$	**63.14**	Price	62.53	62.86	62.77	62.35	62.45	62.43
$1/8$	62.68	Price	60.06	59.16	60.56	60.59	60.94	60.53

Call Sharpe: $(X_T - 100)^+/\sqrt{V_T}$

- $V_T = \frac{1}{T}\sum_{l=1}^{12}\left(\ln \frac{X_{t_l}}{X_{t_{l-1}}}\right)^2$ is the monthly realized volatility.

$$A_t^1 = \sum_{\{l|t_l \leq t\}}\left(\ln \frac{X_{t_l}}{X_{t_{l-1}}}\right)^2, \quad A_t^2 = X_{\sup_{\{l|t_l\leq t\}} t_l}, \quad u(t, X_t, A_t^1, A_t^2)$$

- The true price (PDE) is $u_{\mathrm{PDE}} = 58.4$
- **Param**: We take $\theta_i = (\theta_i^1, \theta_i^2) \in \mathbb{R}^2$ and pick $\gamma_{t_i}(X, A; \theta_i) = \theta_i^1\sqrt{A^1} + \theta_i^2 - \ln(X/X_0)$
- **BSDE**: we assume as a first approximation that $\mathbb{E}_{i-1}[\cdot] \equiv \mathbb{E}[\cdot|X_{t_{i-1}}, A^1_{t_{i-1}}, A^2_{t_{i-1}}] \simeq \mathbb{E}[\cdot|X_{t_{i-1}}]$. The latter is computed using a one dimensional non-parametric regression

Δ	Param	M_1	12	13	14	15	16	17	18
1/2	54.98	Price	47.73	47.18	48.82	48.09	48.10	48.01	48.09
1/4	**55.55**	Price	46.93	47.34	48.01	48.92	48.67	49.38	49.44
1/12	54.32	Price	48.03	49.26	49.78	50.87	51.11	51.66	**52.12**

Outperformer option: $(X_T^1 - X_T^2)^+$

- Two underlyings and no uncertainty on correlation
- Outperformer option with 2 uncorrelated assets valued using $\hat{\sigma}^i$, $i = 1, 2$ and an independent MC. The true price is $u = 11.25$
- **Param:** We choose $\theta_i = (\theta_i^1, \theta_i^2) \in \mathbb{R}^2$, $\sigma_{t_i}^1(X; \theta_i) = \overline{\sigma}^1$ if $\theta_i^1 - \ln(X^1/X_0^1) \geq 0$, $= \underline{\sigma}^1$ otherwise, and $\sigma_{t_i}^2(X; \theta_i) = \overline{\sigma}^2$ if $\theta_i^2 - \ln(X^2/X_0^2) \geq 0$, $= \underline{\sigma}^2$ otherwise

BSDE: Basis functions $= \{1, X_1, X_2\}$:

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	**11.26**	Price	**11.24**	**11.24**	**11.24**	**11.24**	**11.24**	**11.24**
$1/4$	**11.26**	Price	9.65	10.11	10.04	10.16	10.28	9.69
$1/8$	**11.26**	Price	9.17	9.45	9.14	9.26	9.40	9.47
$1/12$	**11.26**	Price	9.17	9.67	9.47	9.38	9.32	9.84

Outperformer option: $(X_T^1 - X_T^2)^+$

BSDE: Basis functions $= \{1, X_1, X_2, X_1^2, X_2^2, X_1 X_2\}$:

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	**11.26**	Price	11.09	11.15	11.15	11.15	11.15	11.15
$1/4$	**11.26**	Price	10.95	11.07	11.10	11.13	11.16	**11.19**
$1/8$	**11.26**	Price	10.35	10.71	10.88	10.94	11.07	11.14
$1/12$	**11.26**	Price	10.39	10.68	10.74	10.91	11.02	11.13

Outperformer option: $(X_T^1 - X_T^2)^+$

- $\rho = -0.5$. The true price is $u = 13.75$
- BSDE: Basis functions $= \{1, X_1, X_2, X_1^2, X_2^2, X_1 X_2\}$

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	**13.77**	Price	13.58	13.66	**13.68**	13.66	13.66	13.65
$1/4$	13.74	Price	13.13	13.48	13.50	13.58	13.64	13.68
$1/8$	13.74	Price	12.45	12.87	13.12	13.20	13.48	13.63
$1/12$	13.75	Price	12.52	12.82	12.98	13.27	13.44	13.59

Outperformer spread option: $(X_T^2 - 0.9X_T^1)^+ - (X_T^2 - 1.1X_T^1)^+$

- $\rho = -0.5$. The true price is $u_{\mathrm{PDE}} = 11.41$
- **Param:** We choose $\theta_i = (\theta_i^1, \theta_i^2) \in \mathbb{R}^2$, $\sigma_{t_i}^1(X;\theta_i) = \overline{\sigma}^1$ if $\theta_i^1 - \ln(X^2/X^1) \geq 0$, $= \underline{\sigma}^1$ otherwise, and $\sigma_{t_i}^2(X;\theta_i) = \overline{\sigma}^2$ if $\theta_i^2 - \ln(X^2/X^1) \geq 0$, $= \underline{\sigma}^2$ otherwise

BSDE: Basis functions $= \{1, X_1, X_2, X_1^2, X_2^2, X_1X_2\}$:

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	**11.37**	Price	11.07	11.07	11.10	**11.11**	11.09	11.10

BSDE: Basis functions $= \{X_1, X_2, \frac{X_2^2}{X_1}, \frac{X_2^3}{X_1^2}\}$:

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	**11.37**	Price	11.24	11.25	11.26	11.26	11.27	11.27
$1/4$	**11.37**	Price	10.85	11.12	11.20	11.27	**11.29**	11.28

Options with two underlyings and uncertainty on correlation

- Outperformer spread option
- $\underline{\rho} = -0.5$, $\overline{\rho} = 0.5$ and $\hat{\rho} = 0$. The (2d) PDE price is 12.83
- **Param:** We choose $\theta_i = (\theta_i^1, \theta_i^2, \theta_i^3) \in \mathbb{R}^3$, $\sigma^1_{t_i}(X; \theta_i) = \overline{\sigma}^1$ if $\theta_i^1 - \ln(X^2/X^1) \geq 0$, $= \underline{\sigma}^1$ otherwise, $\sigma^2_{t_i}(X; \theta_i) = \overline{\sigma}^2$ if $\theta_i^2 - \ln(X^2/X^1) \geq 0$, $= \underline{\sigma}^2$ otherwise, and $\rho_{t_i}(X; \theta_i) = \overline{\rho}$ if $-\theta_i^3 + \ln(X^2/X^1) \geq 0$, $= \underline{\rho}$ otherwise
- **BSDE:** We use the basis functions $= \{1, X^1, X^2, \frac{(X^2)^2}{X^1}, \frac{(X^2)^3}{(X^1)^2}\}$

Δ	Param	M_1	12	13	14	15	16	17
$1/2$	12.50	Price	12.37	12.40	12.41	12.41	12.39	12.38
$1/4$	**12.67**	Price	11.58	11.79	12.38	**12.44**	12.44	12.41

Comments

- Results depend greatly on the choice of regressors which may require numerical experimentations and good understanding of the financial derivatives under consideration. This is a common feature in the pricing of American options
- Recent advances based on deep neural networks: Deep BSDE method. For semilinear PDEs, approximate the delta Z_t as a function of X_t using neural nets, then simulate Y_t in a purely forward fashion based on this parameterization of Z, and use stochastic gradient descent to minimize $\mathbb{E}\left[(Y_T - g(X_T))^2\right]$. See Han, Jentzen, E:*Solving high-dimensional partial differential equations using deep learning*, PNAS 115(34):8505–8510, 2018.
- Combination of the BSDE and parametrical approaches
- The payoffs g that we will use in our numerical experiments below do not satisfy the regularity condition $g \in \mathcal{C}^3$ under which existence and uniqueness the BSB BSDE hold. However, even for these non-smooth payoffs, our discretization scheme seems numerically convergent
- Remark about the generation of the first N_1 paths

Selected references

- Andersen, L. (1999): *A Simple Approach to the pricing of Bermudan Swaptions in the Multi-Factor Libor Market Model* : Journal of Computational Finance, 3:5–32.
- Avellaneda, M., Levy, A., Paras, A. (1995): *Pricing and hedging derivative securities in markets with uncertain volatilities* : Journal of Applied Finance, Vol 1.
- Bouchard, B., Touzi, N. (2004): *Discrete-time approximation and Monte Carlo simulation of backward stochastic differential equations*, Stochastic Processes and their Applications, 111, 175–206.
- Cheridito, P., Soner, M., Touzi, N., Victoir, N. (2007): *Second Order Backward Stochastic Differential Equations and Fully Non-Linear Parabolic PDEs*, Communications on Pure and Applied Mathematics, Volume 60, Issue 7, 1081–1110.

Selected references

- El Karoui, N., Peng, S., Quenez, M.C. (1997): *Backward Stochastic Differential Equations in Finance*, Mathematical Finance, Vol. 47, No. 1, 1–77.
- Guyon, J., Henry-Labordère, P. (2011) : *Uncertain Volatility Model: A Monte Carlo approach*, Journal of Computational Finance, 14(3):37–71.
- Lyons, T.J. (1995): *Uncertain volatility and the risk-free synthesis of derivatives*, Applied Mathematical Finance, 2(2):117–133.
- Pardoux, E., Peng, S. (1992): *Backward Stochastic Differential Equations and Quasilinear Parabolic Partial Differential Equations*, Lecture Notes in CIS, vol. 176. Springer-Verlag, 200–217.

Nonlinear Stochastic Differential Equations in the sense of McKean

Traditional SDE's

- Traditional Stochastic Differential Equation:

$$dX_t = b(t, X_t)\ dt + \sigma(t, X_t) \cdot dW_t, \tag{52}$$

- The pdf $p(t, \cdot)$ of X_t is solution to the Fokker-Planck PDE:

$$-\partial_t p - \sum_{i=1}^{n} \partial_i \left(b^i(t,x) p(t,x) \right) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} \left(\sum_{k=1}^{d} \sigma_k^i(t,x) \sigma_k^j(t,x) p(t,x) \right) = 0$$

$$\lim_{t \to 0} p(t,x) = \delta(x - x_0)$$

- The PDE for $p(t, .)$ is **linear**, so we speak of **linear** SDEs
- Uniqueness and existence proved if drift and volatility are Lipschitz or locally Lipschitz

Calibration of SLV models to market smiles

SLV model (deterministic rate and div yield):

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \sigma(t, S_t)a_t dW_t$$

The model exactly calibrated to market smile iff for all t, K

$$\sigma(t, K)^2 \mathbb{E}[a_t^2 | S_t = K] = \sigma_{\text{loc}}(t, K)^2$$

Nonlinear SDE in the sense of McKean:

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \frac{\sigma_{\text{loc}}(t, S_t)}{\sqrt{\mathbb{E}\left[a_t^2 | S_t\right]}} a_t dW_t \qquad (*)$$

The local volatility function depends on the joint pdf $p(t, S, a)$ of (S_t, a_t):

$$\sigma(t, S, p) = \sigma_{\text{loc}}(t, S)\sqrt{\frac{\int p(t, S, a')da'}{\int a'^2 p(t, S, a')da'}}$$

Equation $(*)$ is an example of **McKean SDE**.

McKean SDEs (Henry McKean, 1966)

- McKean SDE: $X_t \in \mathbb{R}^n$

$$dX_t = b(t, X_t, \mathbb{P}_t)\; dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \mathrm{Law}\,(X_t) \qquad (53)$$

- One-factor SLV: $n = 2$, $X_t = (S_t, a_t)$
- The pdf $p(t, \cdot)$ of X_t is solution to the Fokker-Planck PDE:

$$-\partial_t p - \sum_{i=1}^{n} \partial_i \left(b^i(t, x, \mathbb{P}_t) p(t, x)\right) + \frac{1}{2} \sum_{i,j=1}^{n} \partial_{ij} \left(\sum_{k=1}^{d} \sigma_k^i(t, x, \mathbb{P}_t) \sigma_k^j(t, x, \mathbb{P}_t) p(t, x) \right) = 0$$

$$\lim_{t \to 0} p(t, x) = \delta(x - X_0)$$

- It is **nonlinear** because $b^i(t, x, \mathbb{P}_t)$ and $\sigma_k^i(t, x, \mathbb{P}_t)$ depend on the unknown $p \Rightarrow$ We speak of **nonlinear** SDEs
- Uniqueness and existence can be proved if drift and volatility coefficients are Lipschitz-continuous functions of $\mathbb{P}_t$, with respect to the Wasserstein metric (see Sznitman, Méléard, Villani)

McKean SDEs (Henry McKean, 1966)

- McKean SDE: $X_t \in \mathbb{R}^n$

$$dX_t = b(t, X_t, \mathbb{P}_t)\ dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \mathrm{Law}\,(X_t)$$

- One-factor SLV: $n = 2$, $X_t = (S_t, a_t)$
- The pdf $p(t, \cdot)$ of X_t is solution to the Fokker-Planck PDE:

$$-\partial_t p - \sum_{i=1}^{n} \partial_i \left(b^i(t, x, p(t,x))p(t,x)\right) + \frac{1}{2}\sum_{i,j=1}^{n} \partial_{ij}\left(\sum_{k=1}^{d} \sigma_k^i(t, x, p(t,x))\sigma_k^j(t, x, p(t,x))p(t,x)\right) = 0 \tag{54}$$

$$\lim_{t\to 0} p(t, x) = \delta(x - X_0)$$

- It is **nonlinear** because $b^i(t, x, \mathbb{P}_t)$ and $\sigma_k^i(t, x, \mathbb{P}_t)$ depend on the unknown p ⇒ We speak of **nonlinear** SDEs
- Uniqueness and existence can be proved if drift and volatility coefficients are Lipschitz-continuous functions of $\mathbb{P}_t$, with respect to the Wasserstein metric (see Sznitman, Méléard, Villani)

Wasserstein Metric

The Monge-Kantorovich Distance (or Wasserstein metric) between two probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ is defined as:

$$d_{\mathrm{MK}}(\mathbb{P}_1, \mathbb{P}_2)^p = \inf_{\tau \in \mathcal{P}(\mathbb{R}^n, \mathbb{R}^n) \text{ with marginals } \mathbb{P}_1 \text{ and } \mathbb{P}_2} \left(\int_{\mathbb{R}^n \times \mathbb{R}^n} d(x,y)^p \, \tau(dx, dy) \right)$$

Optimal transport interpretation

Financial interpretation ($n = 1$):

- Consider a European option on two assets $X = S_T^1$ and $Y = S_T^2$ paying $d(X,Y)^p$
- A model τ is calibrated to the vanilla options on S_T^1 and S_T^2 if and only if the marginal distributions of τ are the risk-neutral distributions $\mathbb{P}_1$ and $\mathbb{P}_2$
- The pricing model means choosing the joint distribution τ that minimizes the fair value of the option:

$$\mathbb{E}^{\tau}\left[d(X,Y)^p\right]$$

- This is precisely the Monge-Kantorovich distance between $\mathbb{P}_1$ and $\mathbb{P}_2$

Wasserstein Metric

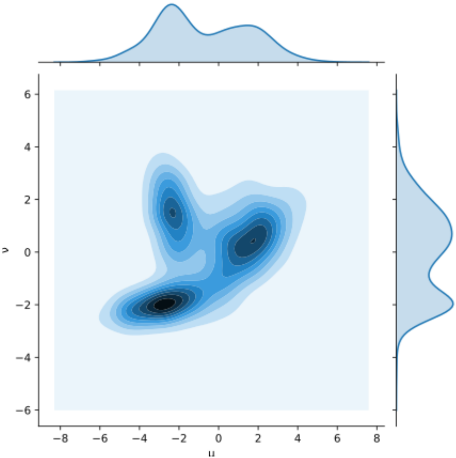


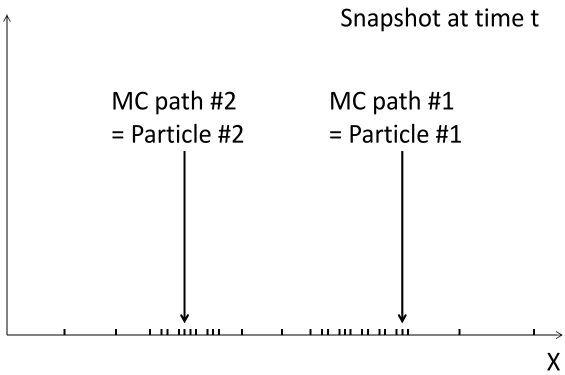
Figure: Example of a transport plan. Source: Wikipedia

How to simulate a McKean SDE?

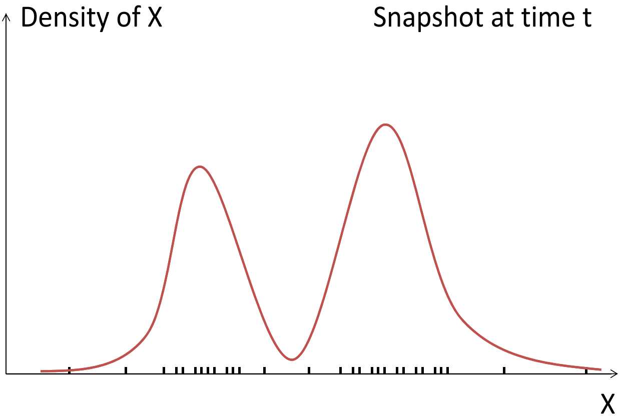
$$dX_t = b(t, X_t, \mathbb{P}_t)\ dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \text{Law}\,(X_t)$$

How to simulate a McKean SDE?

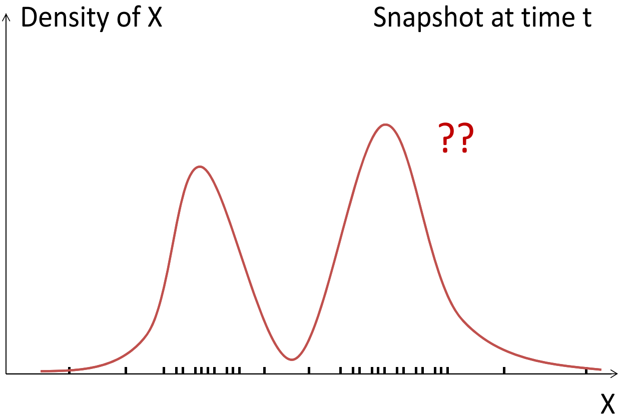
$$dX_t = b(t, X_t, \mathbb{P}_t)\ dt + \sigma(t, X_t, \mathbb{P}_t) \cdot dW_t, \qquad \mathbb{P}_t = \mathrm{Law}\,(X_t)$$



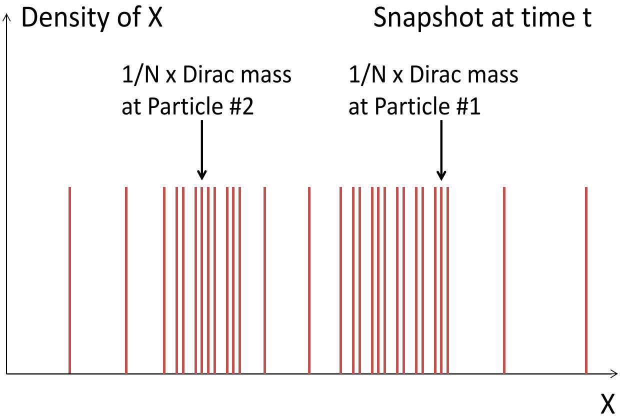
How to simulate a McKean SDE?



How to simulate a McKean SDE?



How to simulate a McKean SDE?



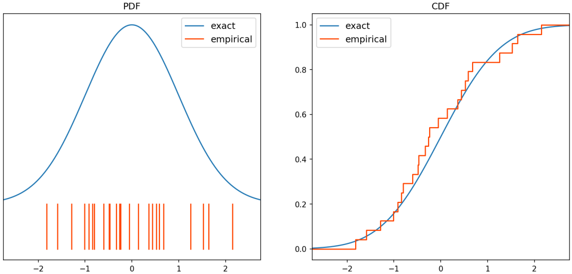
Empirical Measure

Definition (Empirical measure)

Let $X_1, \ldots, X_N$ be i.i.d. random variables with law μ. The empirical measure associated to the configuration $(X_1, \ldots, X_N)$ is

$$\hat{\mu}^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_i}, \quad \text{or} \quad \hat{\mu}^N(A) = \frac{1}{N}\sum_{i=1}^{N} I_A(X_i)$$

Note that the empirical measure $\hat{\mu}^N$ is a random probability measure with expectation $\mathbb{E}_\mu[\hat{\mu}^N] = \mu$, meaning that for all events A, $\mathbb{E}_\mu[\hat{\mu}^N(A)] = \mu(A)$.



Particle Method

- Principle: replace the law $\mathbb{P}_t$ by the empirical distribution

$$\mathbb{P}_t^N = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_t^{i,N}}$$

where the $X_t^{i,N}$ are solution to the $(\mathbb{R}^n)^N$-dimensional **classical** SDE

$$dX_t^{i,N} = b\left(t, X_t^{i,N}, \mathbb{P}_t^N\right) dt + \sigma\left(t, X_t^{i,N}, \mathbb{P}_t^N\right) \cdot dW_t^i, \qquad \text{Law}\left(X_0^{i,N}\right) = \mathbb{P}_0$$

- $\{X_t^i\}_{1\le i\le N}$ = system of N **interacting particles**. The interaction with the other $N-1$ particles comes from $\mathbb{P}_t^N$
- Propagation of chaos (Sznitman, Méléard, Villani) implies the convergence of the particle method:

$$\frac{1}{N}\sum_{i=1}^{N} f(X_t^{i,N}) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int_{\mathbb{R}^d} f(x)p(t,x)dx$$

- In the large N limit, the $(\mathbb{R}^n)^N$-dimensional **linear** Fokker-Planck PDE approximates the **nonlinear** low-dimensional (n-dimensional) Fokker-Planck PDE

Particle Method

$$D_n = \sum_{i \leq n} D(i) + \binom{n}{1}^{-1} \sum_{i<j\leq n} D(ij) + \binom{n}{2}^{-1} \sum_{i<j<k\leq n} D(ijk) + \ldots + D(12\ldots n).$$

[A *backward* operator D is the dual of a *forward* infinitesimal operator D^* as described in §1.] $D(1)$ is an (additive) 1-molecule backward infinitesimal operator acting on the first coordinate of Q^n. $D(i)$ is a copy of $D(1)$ acting on the ith coordinate ($i \leq n$). $D(12) = D(21)$ is a 2-molecule operator acting on the pair of coordinates 12. $D(ij)$ is a copy of $D(12)$ acting on the pair ij ($i < j \leq n$), etc. $D(1)$ governs the motion of molecule number 1 in isolation, $D(12)$ governs the interaction (double collisions) between molecules 1 and 2, $D(123)$ the (triple) collisions between molecules 1, 2, 3, etc. $D_n{}^*$ commutes with permutations of the coordinates so that if the initial n-molecule distribution f^n is symmetrical, then so is $v = e(tD_n{}^*)f^n$ at any later time.

Given a 1-molecule distribution function f, let f^n be the outer product $f \otimes \ldots \otimes f$ so that the molecules are independent at time $t = 0$ with common distribution f. A formal computation indicates that as $n \uparrow \infty$, $v = e(tD_n{}^*)f^n$ tends to the infinite outer product $u^\infty = u \otimes u \otimes$, etc., of the 1-molecule distribution function $u = \lim_{n\uparrow\infty} P_{f^n}[x_1(t) \in db]$. Kac,[4] describing the first known instance of this phenomenon, called it *the propagation of chaos*. u is the (formal) solution of $\partial u/\partial t =$

An archetypal example: the McKean-Vlasov SDE

- For $1 \leq k \leq n$ and $1 \leq l \leq d$

$$\begin{aligned} b_k\left(t, x, \mathbb{P}_t\right) = \int b_k(t, x, y)\mathbb{P}_t(dy) &= \mathbb{E}\left[b_k(t, x, X_t)\right] \\ \sigma_{kl}\left(t, x, \mathbb{P}_t\right) = \int \sigma_{kl}(t, x, y)\mathbb{P}_t(dy) &= \mathbb{E}\left[\sigma_{kl}(t, x, X_t)\right] \end{aligned} \tag{55}$$

- Particle method:

$$dX_t^{i,N} = \left(\int b(t, X_t^{i,N}, y)d\mathbb{P}_t^N(y)\right) dt + \left(\int \sigma(t, X_t^{i,N}, y)d\mathbb{P}_t^N(y)\right) \cdot dW_t^i$$

- This is equivalent to

$$dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^{N} b\left(t, X_t^{i,N}, X_t^{j,N}\right) dt + \frac{1}{N}\sum_{j=1}^{N} \sigma\left(t, X_t^{i,N}, X_t^{j,N}\right) \cdot dW_t^i \tag{56}$$

The Particle Method

- Convergence of the particle method is guaranteed when the propagation of chaos property holds:
- **Propagation of chaos**: If at $t = 0$, the $X_0^{i,N}$ are independent particles, then as $N \to \infty$, for any fixed $t > 0$, the $X_t^{i,N}$ are asymptotically independent and their empirical measure $\mathbb{P}_t^N$ converges in distribution toward the true measure $\mathbb{P}_t$.
- This means that, in the space of probabilities over the space of probabilities, the distribution of the random measure $\mathbb{P}_t^N$ converges toward a Dirac mass at the deterministic measure $\mathbb{P}_t$. Practically, it means that for all functions $\varphi \in C_b(\mathbb{R}^n)$

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int_{\mathbb{R}^n}\varphi(x)p(t,x)\,dx$$

where $p(t,\cdot)$ is the fundamental solution to the nonlinear Fokker-Planck PDE (54). Hence, if propagation of chaos holds, the particle method is convergent.

μ-chaotic distributions

Definition (μ-chaotic distribution)

Let $\{\mu^N\}_{N\in\mathbb{N}}$ be a sequence of *symmetric* probabilities on $(\mathbb{R}^n)^N$. Let μ be a probability measure on $\mathbb{R}^n$. We say that μ^N is μ-chaotic if for each integer $k \geq 1$ and for all test functions $\varphi_1, \ldots, \varphi_k \in C_b(\mathbb{R}^n)$ (i.e., for all bounded continuous functions), we have

$$\int \varphi_1(x_1) \cdots \varphi_k(x_k)\, d\mu^N(x_1, \ldots, x_N) \underset{N\to\infty}{\longrightarrow} \int \varphi_1\, d\mu \cdots \int \varphi_k\, d\mu$$

Stated otherwise, k particles (within N) are asymptotically independent and identically distributed as $N \to \infty$ (k being fixed). We say that μ^N is chaotic if there exists μ such that μ^N is μ-chaotic.

Definition (Propagation of chaos)

Let us consider an N-dimensional SDE flow that associates to an initial probability measure μ_0^N a probability μ_t^N at time t. We say that this flow propagates the chaos if, for any initial chaotic measure μ_0^N and any $t > 0$, μ_t^N is chaotic.

Intuition

What the propagation of chaos essentially says is that as the number of particles becomes very large:

- A **fixed subset of particles** becomes independent as the sample increases
- The empirical measure converges to the law of a single particle
- It is related to the idea that in the large N-limit a linear Fokker-Planck PDE of high dimension approximates a non-linear low dimensional Fokker-Planck PDE

Some Theory

Theorem

Let $\{\mu^N\}_{N\in\mathbb{N}}$ be a sequence of symmetric probabilities on $(\mathbb{R}^n)^N$, and μ be a probability measure on $\mathbb{R}^n$. The following four properties are equivalent:

(i) $\{\mu^N\}_{N\in\mathbb{N}}$ *is μ-chaotic.*

(ii) *For all test functions $\varphi_1, \varphi_2 \in C_b(\mathbb{R}^n)$:*

$$\int \varphi_1(x_1)\varphi_2(x_2)\, d\mu^N(x_1, \dots x_N) \underset{N\to\infty}{\longrightarrow} \int \varphi_1\, d\mu \int \varphi_2\, d\mu$$

(iii) *Let $X_1, \dots, X_N$ be random variables such that $\mathrm{Law}(X_1, \dots, X_N) = \mu^N$. Then, for all $\varphi \in C_b(\mathbb{R}^n)$,*

$$\frac{1}{N}\sum_{i=1}^{N} \varphi(X_i) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int \varphi\, d\mu$$

(iv) *Let $\hat{\mu}^N$ be the empirical measure associated to μ^N. Then*

$$\mathbb{E}_{\mu^N}\left[\left|\int \varphi\, d\hat{\mu}^N - \int \varphi\, d\mu\right|\right] \underset{N\to\infty}{\longrightarrow} 0$$

for all $\varphi \in C_b(\mathbb{R}^n)$.

Some Theory

(i) $\Longrightarrow$ (ii) follows from the definition with $k=2$.

(ii) $\Longrightarrow$ (iii). We have

$$
\begin{aligned}
& \mathbb{E}_{\mu^N}\left[\left(\frac{1}{N}\sum_{i=1}^N \varphi(X_i) - \int \varphi\, d\mu\right)^2\right] \\
= \;& \mathbb{E}_{\mu^N}\left[\frac{1}{N^2}\sum_{i,j=1}^N \varphi(X_i)\varphi(X_j) - 2\int \varphi\, d\mu \frac{1}{N}\sum_{i=1}^N \varphi(X_i) + \left(\int \varphi\, d\mu\right)^2\right] \\
= \;& \frac{1}{N}\mathbb{E}_{\mu^N}[\varphi(X_1)^2] + \frac{N-1}{N}\mathbb{E}_{\mu^N}[\varphi(X_1)\varphi(X_2)] \\
& -2\int \varphi\, d\mu \mathbb{E}_{\mu^N}[\varphi(X_1)] + \left(\int \varphi\, d\mu\right)^2 \\
& \underset{N\to\infty}{\longrightarrow} 0 + \left(\int \varphi\, d\mu\right)^2 - 2\left(\int \varphi\, d\mu\right)^2 + \left(\int \varphi\, d\mu\right)^2 = 0
\end{aligned}
$$

as from (ii) $\mathbb{E}_{\mu^N}[\varphi(X_1)\varphi(X_2)] \underset{N\to\infty}{\longrightarrow} \left(\int \varphi\, d\mu\right)^2$ and $\mathbb{E}_{\mu^N}[\varphi(X_1)] \underset{N\to\infty}{\longrightarrow} \int \varphi\, d\mu$.

(iii) $\Longrightarrow$ (iv) follows from the definition of convergence in mean.

Some Theory

(iv) $\Longrightarrow$ (i). Let $k \geq 1$ and $\varphi_1, \ldots, \varphi_k \in C_b(\mathbb{R}^n)$. We have (from the triangle inequality):

$$
\begin{aligned}
&\left|\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)] - \int \varphi_1\, d\mu \cdots \int \varphi_k\, d\mu\right| \\
&\quad \leq \left|\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)] - \mathbb{E}_{\mu^N}\left[\int \varphi_1\, d\hat{\mu}^N \cdots \int \varphi_k\, d\hat{\mu}^N\right]\right| \\
&\qquad + \left|\mathbb{E}_{\mu^N}\left[\int \varphi_1\, d\hat{\mu}^N \cdots \int \varphi_k\, d\hat{\mu}^N\right] - \int \varphi_1\, d\mu \cdots \int \varphi_k\, d\mu\right|
\end{aligned}
$$

The second term on the r.h.s. converges to zero thanks to hypothesis (iv). The first term on the r.h.s. reads

$$
\left|\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)] - \frac{1}{N^k}\sum_{i_1,\ldots,i_k=1}^{N} \mathbb{E}_{\mu^N}[\varphi_1(X_{i_1})\cdots\varphi_k(X_{i_k})]\right|
$$

In the above sum, $\frac{N!}{(N-k)!}$ terms are such that the indices $i_1, \ldots, i_k$ are all different. By symmetry, they are equal to $\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)]$ and can be added to the first term. The other terms can be bounded by M^k with $M = \sum_j ||\varphi_j||_\infty$, so the first term is bounded by

$$
\begin{aligned}
&\mathbb{E}_{\mu^N}[\varphi_1(X_1)\cdots\varphi_k(X_k)]\left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right) + \left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right)M^k \\
&\qquad \leq 2M^k\left(1 - \frac{1}{N^k}\frac{N!}{(N-k)!}\right) \underset{N\to\infty}{\longrightarrow} 0
\end{aligned}
$$

Propagation of Chaos - McKean-Vlasov SDE

Theorem

The propagation of chaos property holds for the McKean-Vlasov SDE (56)

The intuition of the proof relies on a sort of "coupling method", which consists of introducing a process Y_t^m, $1 \leq m \leq N$, defined as:

$$dY_t^m = b(Y_t^m, \mathbb{P}_t)\,dt + \sigma(Y_t^m, \mathbb{P}_t)\,dW_t^m, \qquad Y_0^m = X_0$$

with $b(y, \mathbb{P}_t) := \int b(y,z)\mathbb{P}_t(dz)$, $\sigma(y, \mathbb{P}_t) := \int \sigma(y,z)\mathbb{P}_t(dz)$, and $\mathbb{P}_t := \mathrm{Law}(X_t)$. These are standard SDEs which admit a strong solution as b and σ are Lipschitz-continuous functions. The density $q_m(t,x)$ of Y_t^m satisfies the *linear* Fokker-Planck equation

$$-\partial_t q_m(t,x) - \sum_{i=1}^n \partial_i \left(b^i(t,x,\mathbb{P}_t) q_m(t,x)\right) + \frac{1}{2}\sum_{i,j=1}^n \partial_{ij}\left(\sum_{k=1}^d \sigma_k^i(t,x,\mathbb{P}_t)\sigma_k^j(t,x,\mathbb{P}_t) q_m(t,x)\right) = 0$$

with Dirac initial condition. From (54), the density $p(t,x)$ of X_t is also a solution. By uniqueness, $\mathbb{P}_t = \mathrm{Law}(Y_t^m)$.

Propagation of Chaos - McKean-Vlasov SDE

Proposition

Let $(X^{i,N})_{1\leq i\leq N}$ be defined by the particle method. Then

$$\mathbb{E}[|X_t^{1,N} - Y_t^1|] \leq \frac{C(t)}{\sqrt{N}}$$

where $C(t)$ is a smooth function of time independent of N.

Proof of Chaos Propagation Theorem for McKean-Vlasov SDE
Let us denote by μ_t the law of the solution X_t of the McKean SDE (55). Then from Theorem 29 (iii), it is enough to show that for all $\varphi \in C_b(\mathbb{R}^n)$,

$$\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) \xrightarrow[N\to\infty]{\mathrm{L}^1} \int \varphi\, d\mu_t \tag{57}$$

where the $X_t^{i,N}$ are defined by (56).

Propagation of Chaos - McKean-Vlasov SDE

$$
\mathbb{E}_{\mu^N}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\varphi(X_t^{i,N}) - \int \varphi\, d\mu_t\right|\right]
$$

$$
\leq \mathbb{E}_{\mu^N}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\left(\varphi(X_t^{i,N}) - \varphi(Y_t^i)\right)\right|\right] + \mathbb{E}_{\mu^N}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\varphi(Y_t^i) - \int \varphi\, d\mu_t\right|\right]
$$

As by construction the processes $\{Y_t^i\}_{1\leq i\leq N}$ are i.i.d. with law μ_t, the second term above goes to zero as $N \to \infty$ from the law of large numbers.
For all $\epsilon > 0$, there exists a Lipschitz-continuous function φ_ϵ such that $|\varphi - \varphi_\epsilon| \leq \epsilon$. The first term is then bounded by

$$
2\epsilon + \mathbb{E}_{\mu^N}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\left(\varphi_\epsilon(X_t^{i,N}) - \varphi_\epsilon(Y_t^i)\right)\right|\right]
$$

$$
\leq 2\epsilon + ||\varphi_\epsilon||_{\mathrm{Lip}}\mathbb{E}[|X_t^{1,N} - Y_t^1|] \leq 2\epsilon + ||\varphi_\epsilon||_{\mathrm{Lip}}\frac{C(t)}{\sqrt{N}}
$$

from the proposition above.

Calibration of models to market smiles

The tools and concepts that will be used: A bit of history

- **Concept of propagation of chaos**: Kac, 1960's
- **Nonlinear stochastic differential equations**: McKean, 1966
  A Class of Markov Processes Associated with Nonlinear Parabolic Equations (Proc Natl Acad Sci U S A. 1966 December; 56(6): 1907–1911)
- **Particle method and propagation of chaos**: Sznitman, Méléard, Chorin, Puckett, Talay, Bossy, Jourdain, etc., 1980's and 1990's
- **Application to various smile calibration problems in quantitative finance**: G. and Henry-Labordère, 2011
  Being Particular About Calibration (Risk, January 2012)

Local Volatility Model (LVM)

- Local Volatility Model (deterministic rate and div yield):

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma_{\text{loc}}(t, S_t)\,dW_t$$

where r_t and q_t are deterministic, and $\sigma_{\text{loc}}(t, S)$ is a function of t and S.

- Local volatility function from call option prices (*Dupire's Formula*)

$$\sigma^2_{\text{loc}}(T, K) = \frac{\frac{\partial C}{\partial T} + (r_t - q_t)\,K\,\frac{\partial C}{\partial K} + q_t C}{\frac{1}{2} K^2 \frac{\partial^2 C}{\partial K^2}}.$$

where

$$C(T, K) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_t\,dt}\,(S_T - K)^+\right]$$

is the price of the call option on S with maturity T and strike K.

$$\sigma^2_{\text{loc}}(T, K) := \frac{\frac{\partial C^{\text{mkt}}}{\partial T} + (r_t - q_t)\,K\,\frac{\partial C^{\text{mkt}}}{\partial K} + q_t C^{\text{mkt}}}{\frac{1}{2} K^2 \frac{\partial^2 C^{\text{mkt}}}{\partial K^2}}.$$

Local Volatility Model (LVM)

- From Implied Volatility to Local Volatility:

$$\sigma^2_{\text{loc}}(t,x) = \frac{\frac{\partial w}{\partial T}}{1 - \frac{x}{w}\frac{\partial w}{\partial x} + \frac{1}{4}\left(-\frac{1}{4} - \frac{1}{w} + \frac{x^2}{w^2}\right)\left(\frac{\partial w}{\partial x}\right)^2 + \frac{1}{2}\frac{\partial^2 w}{\partial x^2}}$$

where log-moneyness $x = \log(K/F_T)$, $F_T = S_0 e^{\int_0^T (r_t - q_t)\,dt}$ is the forward with maturity T, and $w(T,x) = \sigma^2_{\text{imp}}(K,T)T$.

- Stochastic Volatility model

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + a_t\,dW_t$$

(where a_t is stochastic) is calibrated to the market smile if and only if

$$\mathbb{E}\left[a_t^2 | S_t = K\right] = \sigma^2_{\text{loc}}(t,K).$$

(Proof will be given later)

Stochastic Local Volatility (SLV) models

- Exotic options bear exotic risks such as volatility-of-volatility risk, forward smile risk, or spot/volatility correlation risk.
- To price and hedge these risks, one should not use the Black-Scholes model, or the local volatility model, because such models give no control on them.
- One would rather use a Stochastic Volatility (SV) model

$$\frac{dS_t}{S_t} = (r_t - q_t)\, dt + a_t\, dW_t \tag{58}$$

with a_t the stochastic volatility, a random process (e.g., Heston, SABR, double-lognormal) which allows us to handle exotic risks through parameters such as volatility-of-volatility, spot/volatility correlation, or mean-reversion.

Heston Stochastic Volatility Model - Example of SV Model

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sqrt{V_t}\,dW_t^S$$
$$dV_t = \kappa\,(\theta - V_t)\,dt + \omega\sqrt{V_t}\,dW_t^V$$

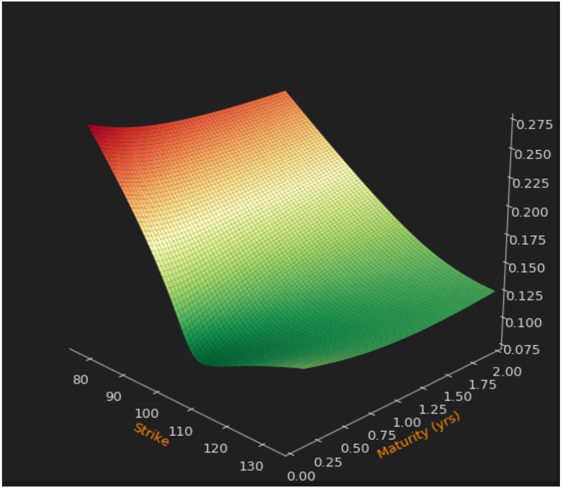
with

$$d\langle W^S, W^V\rangle_t = \rho\,dt.$$

Model Parameters

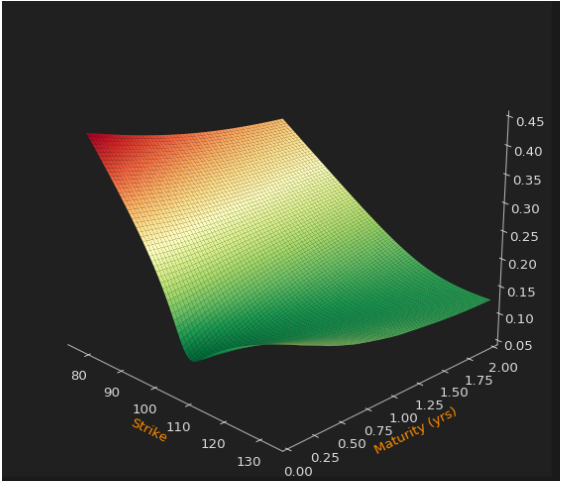
- κ: mean reversion rate
- V_0: initial variance
- θ: Long term variance
- ρ: spot-vol correlation
- ω: volatility of variance

Heston Stochastic Volatility Model - Implied Volatility Surface



$$V_0 = 0.0165, \kappa = 1.3093, \theta = 0.0541, \omega = 0.6344, \rho = -0.744, S_0 = 100$$

Heston Stochastic Volatility Model - Local Volatility Surface



$$V_0 = 0.0165, \kappa = 1.3093, \theta = 0.0541, \omega = 0.6344, \rho = -0.744, S_0 = 100$$

Stochastic Local Volatility (SLV) models

- SV models produce a smile of implied volatilities (see the seminal papers by Hull-White, Renault-Touzi; see also Bergomi-Guyon for an analysis of the smile in general multi-factor second-generation SV models).
- Since they have a finite number of parameters (typically, 3 to 10), SV models cannot be perfectly calibrated to full (inter- and extrapolated) market smiles, indexed by all strikes and maturities (until some final maturity T).
- Now, vanilla options often provide a good hedge of exotic options. Dynamic or static trading of vanilla options often results in reducing the variance of the final P&L. In those cases, it is important that the SV model incorporates the correct initial prices of the hedging instruments—the vanilla options.

Stochastic Local Volatility (SLV) models

- How do we build an SV model that calibrates exactly to a full surface of implied volatilities?
- Due to the double infinity of constraints, indexed by strikes and maturities, one needs to introduce a double infinity of parameters.
- A natural way to do so, and a common practice in the foreign exchange market, is to embed a local volatility $\sigma(t, S)$ (also called "leverage function") into the SV model:

$$\frac{dS_t}{S_t} = (r_t - q_t)\, dt + a_t \sigma(t, S_t)\, dW_t \tag{59}$$

  We speak of stochastic local volatility (SLV) models.
- The main issue we address in these lecture notes is how to build the local volatility $\sigma(t, S)$ to ensure that the market smile is exactly calibrated.
- Note that this local volatility function differs from Dupire's local volatility. For instance, if the smile produced by the "naked" SV model (58) is close to the market smile, one expects the calibrated local volatility $\sigma(t, S)$ to be uniformly close to 1.

Calibration of SLV models to market smiles

- SLV model (deterministic rate and div yield):

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \sigma(t, S_t)a_t dW_t$$

- Model exactly calibrated to market smile iff for all t, K

$$\sigma(t, K)^2 \mathbb{E}[a_t^2 | S_t = K] = \sigma_{\text{loc}}(t, K)^2$$

- $\Rightarrow$ **Nonlinear** SDE:

$$\frac{dS_t}{S_t} = (r_t - q_t)dt + \frac{\sigma_{\text{loc}}(t, S_t)}{\sqrt{\mathbb{E}[a_t^2 | S_t]}} a_t dW_t \qquad (60)$$

- The local volatility function depends on the joint pdf $p\,(t, S, a)$ of (S_t, a_t):

$$\sigma(t, S, p) = \sigma_{\text{loc}}(t, S)\sqrt{\frac{\int p(t, S, a')da'}{\int a'^2 p(t, S, a')da'}}$$

- Equation (60) is an example of **McKean SDE**

The calibration condition

Proposition. Let us consider the following dynamics for an asset S, where the volatility α_t, the interest rate r_t, and the repo q_t, inclusive of the dividend yield, are all stochastic processes:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \alpha_t\,dW_t \tag{61}$$

Model (61) is exactly calibrated to the market smile of S if and only if

$$\begin{aligned}\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]} &= \sigma_{\text{loc}}(t,K)^2 \\ &- \frac{\mathbb{E}\left[D_{0t}\left(r_t - q_t - (r_t^0 - q_t^0)\right)\mathbf{1}_{S_t>K}\right]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t,K)} + \frac{\mathbb{E}\left[D_{0t}\left(q_t - q_t^0\right)(S_t - K)^+\right]}{\frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)}\end{aligned} \tag{62}$$

for all (t,K), where $D_{0t} = \exp\left(-\int_0^t r_s ds\right)$ is the discount factor, r_t^0 and q_t^0 are deterministic rates and repos, and

$$\sigma_{\text{loc}}(t,K)^2 = \frac{\partial_t\mathcal{C}(t,K) + (r_t^0 - q_t^0)K\partial_K\mathcal{C}(t,K) + q_t^0\mathcal{C}(t,K)}{\frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)} \tag{63}$$

with $\mathcal{C}(t,K)$ the market price of the call option on S with strike K and maturity t.

Case where rates and repo are deterministic

Proposition. In the case where rates r_t and repo/dividend yield q_t are deterministic, Model (61) is exactly calibrated to the market smile of S if and only if

$$\mathbb{E}[\alpha_t^2 | S_t = K] = \sigma_{\mathrm{loc}}(t, K)^2$$

for all (t, K), where

$$\sigma_{\mathrm{loc}}(t, K)^2 = \frac{\partial_t \mathcal{C}(t, K) + (r_t - q_t) K \partial_K \mathcal{C}(t, K) + q_t \mathcal{C}(t, K)}{\frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t, K)} \tag{64}$$

with $\mathcal{C}(t, K)$ the market price of the call option on S with strike K and maturity t.

When $\alpha_t = \sigma(t, S_t) a_t$ the calibration condition reads

$$\sigma(t, K)^2 \mathbb{E}[a_t^2 | S_t = K] = \sigma_{\mathrm{loc}}(t, K)^2$$

Proof of the general proposition

By applying Itô-Tanaka's formula on a discounted vanilla call payoff with maturity t and strike K, $\mathcal{P}_t \equiv D_{0t}(S_t - K)^+$, we have:

$$\begin{aligned} d\mathcal{P}_t &= -D_{0t}(S_t - K)^+ r_t\, dt + D_{0t}\mathbf{1}_{S_t > K} S_t\left((r_t - q_t)\, dt + \alpha_t dW_t\right) \\ &\quad + \frac{1}{2} S_t^2 \alpha_t^2 D_{0t}\delta(S_t - K)\, dt \\ &= D_{0t}\mathbf{1}_{S_t > K}(r_t - q_t)K\, dt - D_{0t} q_t (S_t - K)^+\, dt + D_{0t}\mathbf{1}_{S_t > K}\alpha_t S_t dW_t \\ &\quad + \frac{1}{2} K^2 \alpha_t^2 D_{0t}\delta(S_t - K)\, dt \end{aligned}$$

By taking the expectation $\mathbb{E}[\cdot]$ on both sides of the above equation and by assuming that $M_t = \int_0^t D_{0s}\mathbf{1}_{S_s > K}\alpha_s S_s dW_s$ is a true martingale, we get

$$\begin{aligned} \partial_t \mathcal{C}_{\mathrm{m}}(t, K) = K\mathbb{E}[D_{0t}(r_t - q_t)\mathbf{1}_{S_t > K}] - \mathbb{E}[D_{0t} q_t (S_t - K)^+] \\ + \frac{1}{2} K^2 \mathbb{E}[D_{0t}\alpha_t^2 \delta(S_t - K)] \end{aligned}$$

where $\mathcal{C}_{\mathrm{m}}(t, K) = \mathbb{E}[\mathcal{P}_t]$ denotes the price of the call option in the model.

Proof of the general proposition (continued)

Then, by using that $\partial_K \mathcal{C}_{\mathrm{m}}(t,K) = -\mathbb{E}[D_{0t}\mathbf{1}_{S_t>K}]$ and $\partial_K^2 \mathcal{C}_{\mathrm{m}}(t,K) = \mathbb{E}[D_{0t}\,\delta(S_t-K)]$, we deduce that

$$\partial_t \mathcal{C}_{\mathrm{m}}(t,K) = K\mathbb{E}[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))\mathbf{1}_{S_t>K}] - (r_t^0 - q_t^0)K\partial_K \mathcal{C}_{\mathrm{m}}(t,K)$$
$$-\mathbb{E}[D_{0t}(q_t - q_t^0)(S_t - K)^+] - q_t^0 \mathcal{C}_{\mathrm{m}}(t,K) + \frac{1}{2}K^2\partial_K^2 \mathcal{C}_{\mathrm{m}}(t,K)\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t=K]}{\mathbb{E}[D_{0t}|S_t=K]}$$

with the initial condition $\mathcal{C}_{\mathrm{m}}(0,K) = (S_0 - K)^+$ so by uniqueness of the solution to this PDE the model is calibrated to the market smile of S if and only if

$$\partial_t \mathcal{C}(t,K) = K\mathbb{E}[D_{0t}(r_t - q_t - (r_t^0 - q_t^0))\mathbf{1}_{S_t>K}] - (r_t^0 - q_t^0)K\partial_K \mathcal{C}(t,K)$$
$$-\mathbb{E}[D_{0t}(q_t - q_t^0)(S_t - K)^+] - q_t^0 \mathcal{C}(t,K) + \frac{1}{2}K^2\partial_K^2 \mathcal{C}(t,K)\frac{\mathbb{E}[D_{0t}\alpha_t^2|S_t=K]}{\mathbb{E}[D_{0t}|S_t=K]}$$

From the definition of $\sigma_{\mathrm{loc}}(t,K)$, this is equivalent to (62), which completes the proof.

Back to the calibration of SLV models

- In the SLV model, the approximated conditional expectation is

$$\mathbb{E}^{\mathbb{P}_t^N}\left[a_t^2|S_t = S\right] = \frac{\int a'^2 p_N(t,S,a')da'}{\int p_N(t,S,a')da'} = \frac{\sum_{i=1}^N (a_t^{i,N})^2 \delta(S_t^{i,N} - S)}{\sum_{i=1}^N \delta(S_t^{i,N} - S)}$$

- Instead of the Dirac delta function $\delta(\cdot)$, we use a kernel $\delta_{t,N}(\cdot)$ and approximate $\mathbb{E}^{\mathbb{P}_t}[a_t^2|S_t = S]$ by (simple nonparametric regression!)

$$\bar{a}(t,S)^2 = \frac{\sum_{i=1}^N \left(a_t^{i,N}\right)^2 \delta_{t,N}\left(S_t^{i,N} - S\right)}{\sum_{i=1}^N \delta_{t,N}\left(S_t^{i,N} - S\right)} \quad (65)$$

- Then we define

$$\sigma_N(t,S) = \frac{\sigma_{\mathrm{loc}}(t,S)}{\bar{a}(t,S)}$$

and simulate

$$\frac{dS_t^{i,N}}{S_t^{i,N}} = (r_t - q_t)dt + \sigma_N(t, S_t^{i,N}) a_t^{i,N} dW_t^i \quad (66)$$

- $O(N^2)$ operations at each discretization date $\Rightarrow$ it is crucial to design acceleration techniques

How to make the particle method efficient in practice

- Computing $\sigma_N(t, S_t^{i,N})$ for all i is useless $\Rightarrow$ Compute $\sigma_N(t,S)$ for only a grid of values $G_{S,t}$ of S, of a size much smaller than N, and then inter- and extrapolate
- In the sums in (65), a large number of terms make a negligible contribution $\Rightarrow$ Disregard $S_t^{i,N}$ when it is far from S, say, when $\delta_{t,N}(S_t^{i,N} - S) < \eta$ $\Rightarrow$ Sort particles according to spot value. Cost of sorting, $O(N \ln N)$, is more than compensated by the acceleration in the evaluations of (65)
- Alternative methods for estimating such conditional expectations include B-spline techniques (Corlay, 2013), neural networks, random forests, nearest neighbors, etc.
- Calibration and pricing can be achieved in the course of the same Monte-Carlo simulation
- If we need more paths for pricing, we can reuse the calibration paths, and simulate new paths using the already calibrated local volatility function

Choice of kernel and bandwidth

- It is natural to take $\delta_{t,N}(x) = \frac{1}{h_{t,N}} K\left(\frac{x}{h_{t,N}}\right)$ where K is a fixed, symmetric kernel with a bandwidth $h_{t,N}$ that tends to zero as N grows to infinity.
- Typical examples:

$$\begin{aligned}
\text{Exponential kernel:} \qquad & K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right) \\
\text{Epanechnikov kernel:} \qquad & K(x) = \frac{3}{4}\left(1 - x^2\right) \mathbf{1}_{\{|x| \leq 1\}} \\
\text{Quartic kernel:} \qquad & K(x) = \frac{15}{16}\left(1 - x^2\right)^2 \mathbf{1}_{\{|x| \leq 1\}}
\end{aligned}$$

The last two are faster to compute than the exponential.

Choice of kernel and bandwidth

- We take
$$h_{t,N} = \kappa S_0 \sigma_{\mathrm{VS},t} \sqrt{\max(t, t_{\min})} N^{-\frac{1}{5}}$$
with $\sigma_{\mathrm{VS},t}$ the variance swap volatility at maturity t.
- The factor $N^{-\frac{1}{5}}$ comes from the minimization of the asymptotic mean integrated squared error of the Nadaraya-Watson estimator, which is the sum of two terms: bias and variance. The smaller the bandwidth, the smaller the bias, but the larger the variance. The critical bandwidth that minimizes the sum of bias and variance decreases as $N^{-\frac{1}{5}}$ for large N.
- Following Silverman's rule of thumb, the prefactor $\kappa S_0 \sigma_{\mathrm{VS},t} \sqrt{\max(t, t_{\min})}$ is on the order of the standard deviation of the regressor S_t, if κ is of the order of magnitude of 1. In practice, we take $\kappa \in [1, 3]$, $t_{\min} = 0.1$.

Particle algorithm (G. and Henry-Labordère, 2011)

1. Initialize $k = 1$ and set $\sigma_N(t,S) = \frac{\sigma_{\text{loc}}(0,S)}{\alpha}$ for all $t \in [t_0 = 0, t_1]$
2. Simulate the N processes $\{S_t^{i,N}, a_t^{i,N}\}_{1 \le i \le N}$ from t_{k-1} to t_k using a discretization scheme for (66) - say a log-Euler scheme
3. Sort particles according to spot value. For $S \in G_{S,t_k}$, find the smallest index $\underline{i}(S)$ and the largest index $\overline{i}(S)$ for which $\delta_{t_k,N}\left(S_{t_k}^{i,N} - S\right) > \eta$, and compute the local volatility

$$\sigma_N(t_k,S) = \sigma_{\text{loc}}(t_k,S)\sqrt{\frac{\sum_{i=\underline{i}(S)}^{\overline{i}(S)} \delta_{t_k,N}\left(S_{t_k}^{i,N} - S\right)}{\sum_{i=\underline{i}(S)}^{\overline{i}(S)} \left(a_{t_k}^{i,N}\right)^2 \delta_{t_k,N}\left(S_{t_k}^{i,N} - S\right)}}$$

   Interpolate the square root using cubic splines, and extrapolate (e.g., flat). Set $\sigma_N(t,S) \equiv \sigma_N(t_k,S)$ for all $t \in [t_k, t_{k+1}]$
4. Set $k := k+1$. Iterate steps 2 and 3 up to the maturity date T

Adding stochastic rates

- SIR-SLV model:
$$\frac{dS_t}{S_t} = r_t dt + \sigma(t, S_t) a_t dW_t$$
- Model exactly calibrated to market smile iff for all t, K
$$\sigma(t,K)^2 \frac{\mathbb{E}[D_{0t} a_t^2 | S_t = K]}{\mathbb{E}[D_{0t} | S_t = K]} = \sigma_{\text{loc}}(t,K)^2 - \frac{\mathbb{E}\left[D_{0t}\left(r_t - r_t^0\right) 1_{S_t > K}\right]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t,K)} \quad (67)$$
where r_t^0 is a deterministic rate and
$$\sigma_{\text{loc}}(t,K)^2 = \frac{\partial_t \mathcal{C}(t,K) + r_t^0 K \partial_K \mathcal{C}(t,K)}{\frac{1}{2} K^2 \partial_K^2 \mathcal{C}(t,K)}$$
- Equivalently, using the t-forward measure $\mathbb{P}^t$,
$$\sigma(t,K)^2 \mathbb{E}^{\mathbb{P}^t}\left[a_t^2 | S_t = K\right] = \sigma_{\text{loc}}(t,K)^2 - P_{0t} \frac{\mathbb{E}^{\mathbb{P}^t}\left[\left(r_t - r_t^0\right) 1_{S_t > K}\right]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t,K)}$$
- Usually $r_t^0 = -\partial_t \ln P_{0t} = \mathbb{E}^{\mathbb{P}^t}[r_t]$
- Does the r.h.s. of (67) stay nonnegative?

Adding stochastic rates

- $\Rightarrow$ The calibrated model follows the nonlinear SDE

$$\frac{dS_t}{S_t} = r_t\,dt + \sigma\left(t, S_t, \mathbb{P}_t\right)\, a_t\,dW_t$$

where

$$\sigma\left(t, K, \mathbb{P}_t\right)^2 = \left(\sigma_{\mathrm{loc}}(t,K)^2 - \frac{\mathbb{E}[D_{0t}\left(r_t - r_t^0\right) 1_{S_t > K}]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t,K)}\right) \frac{\mathbb{E}^{\mathbb{P}}[D_{0t} | S_t = K]}{\mathbb{E}^{\mathbb{P}}[D_{0t} a_t^2 | S_t = K]} \quad (68)$$

- Equivalently:

$$\frac{df_t}{f_t} = \sigma\left(t, P_{tT} f_t, \mathbb{P}_t^T\right)\, a_t\,dW_t^T - \sigma_{\mathrm{P}}^T(t).dB_t^T$$

where

$$\sigma\left(t, K, \mathbb{P}_t^T\right)^2 = \left(\sigma_{\mathrm{loc}}(t,K)^2 - P_{0T} \frac{\mathbb{E}^{\mathbb{P}^T}[P_{tT}^{-1}\left(r_t - r_t^0\right) 1_{S_t > K}]}{\frac{1}{2} K \partial_K^2 \mathcal{C}(t,K)}\right) \frac{\mathbb{E}^{\mathbb{P}^T}[P_{tT}^{-1} | S_t = K]}{\mathbb{E}^{\mathbb{P}^T}[P_{tT}^{-1} a_t^2 | S_t = K]} \quad (69)$$

Particle algorithm (G. and Henry-Labordère) - Case I

If we use (68) we define

$$\sigma_N(t,S)^2 = \left(\sigma_{\mathrm{loc}}(t,S)^2 - \frac{\frac{1}{N}\sum_{i=1}^N D_{0t}^{i,N}\left(r_t^{i,N} - r_t^0\right)1_{S_t^{i,N}>S}}{\frac{1}{2}S\partial_K^2\mathcal{C}(t,S)}\right) \times \frac{\sum_{i=1}^N D_{0t}^{i,N}\delta_{t,N}\left(S_t^{i,N}-S\right)}{\sum_{i=1}^N D_{0t}^{i,N}\left(a_t^{i,N}\right)^2\delta_{t,N}\left(S_t^{i,N}-S\right)}$$

and simulate

$$\frac{dS_t^{i,N}}{S_t^{i,N}} = r_t^{i,N}dt + \sigma_N\left(t, S_t^{i,N}\right)a_t^{i,N}dW_t^i$$

Particle algorithm (G. and Henry-Labordère) - Case II

In many commonly used short rate models, P_{tT} bond has a closed form formula, we can use (69), define

$$\sigma_N(t,S)^2 = \left(\sigma_{\text{loc}}(t,S)^2 - P_{0T}\frac{\frac{1}{N}\sum_{i=1}^N \left(P_{tT}^{i,N}\right)^{-1}\left(r_t^{i,N} - r_t^0\right) 1_{S_t^{i,N}>S}}{\frac{1}{2}S\partial_K^2\mathcal{C}(t,S)}\right) \times \frac{\sum_{i=1}^N \left(P_{tT}^{i,N}\right)^{-1}\delta_{t,N}\left(S_t^{i,N} - S\right)}{\sum_{i=1}^N \left(P_{tT}^{i,N}\right)^{-1}\left(a_t^{i,N}\right)^2\delta_{t,N}\left(S_t^{i,N} - S\right)}$$

and simulate

$$df_t^{i,N} = f_t^{i,N}\sigma_N\left(t, f_t^{i,N}P_{tT}^{i,N}\right)a_t^{i,N}dW_t^i - f_t^{i,N}\sigma_{\mathrm{P}}^{T,i,N}(t).dB_t^i$$

where W^i and B^i are $\mathbb{P}^T$-Brownian motions.

Adding stochastic repo/dividend yield (G., 2013)

- SIR-SDY-SLV model:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma(t, S_t)a_t\,dW_t$$

- Model exactly calibrated to market smile iff for all (t, K)

$$\sigma(t,K)^2 \frac{\mathbb{E}[D_{0t}a_t^2|S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]} = \sigma_{\mathrm{loc}}(t,K)^2 - \frac{\mathbb{E}\left[D_{0t}\left(r_t - q_t - (r_t^0 - q_t^0)\right)1_{S_t > K}\right]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t,K)} + \frac{\mathbb{E}\left[D_{0t}\left(q_t - q_t^0\right)(S_t - K)^+\right]}{\frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)}$$

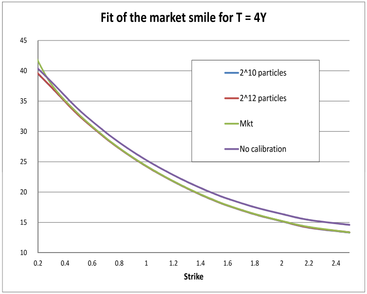
where r_t^0 and q_t^0 are deterministic rates and repos and

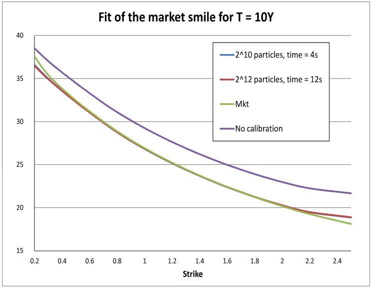
$$\sigma_{\mathrm{loc}}(t,K)^2 = \frac{\partial_t\mathcal{C}(t,K) + (r_t^0 - q_t^0)K\partial_K\mathcal{C}(t,K) + q_t^0\mathcal{C}(t,K)}{\frac{1}{2}K^2\partial_K^2\mathcal{C}(t,K)}$$

- The calibrated model follows a McKean SDE
- Does $\sigma(t,K)^2$ stay nonnegative?

LV + Ho-Lee

- DAX market smiles (30-May-11)
- Ho-Lee parameters: $\sigma_{\mathrm{r}} = 1\%$, $\rho = 40\%$
- Particle algorithm: $\Delta = 1/100$, $N = 2^{10}$ particles. MC pricing: $N = 2^{15}$ paths





LV + Ho-Lee

Strike	0.5	0.7	0.8	0.9	1	1.1	1.2	1.3	1.5	1.8
with Malliavin	14	10	10	9	8	7	6	5	3	1
without Malliavin	16	8	7	4	1	1	1	3	3	5

Table: DAX (30-May-11) Implied volatilities $T = 10$Y. Errors in bps using the particle method with $N = 2^{10}$ particles

LV + Bergomi

$$
\begin{aligned}
df_t &= f_t \sigma(t, f_t) \sqrt{\xi_t^t} dW_t \\
\xi_t^T &= \xi_0^T f^T(t, x_t^T) \\
f^T(t, x) &= \exp(2\nu x - 2\nu^2 h(t, T)) \\
x_t^T &= \alpha_\theta \left((1-\theta) e^{-k_X (T-t)} X_t + \theta e^{-k_Y (T-t)} Y_t \right) \\
\alpha_\theta &= \left((1-\theta)^2 + \theta^2 + 2\rho_{XY} \theta (1-\theta) \right)^{-1/2} \\
dX_t &= -k_X X_t dt + dW_t^X \\
dY_t &= -k_Y Y_t dt + dW_t^Y
\end{aligned}
$$

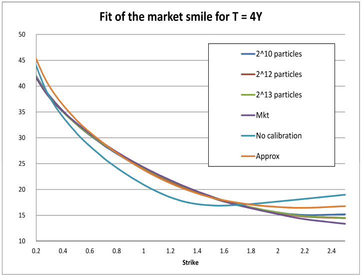
where

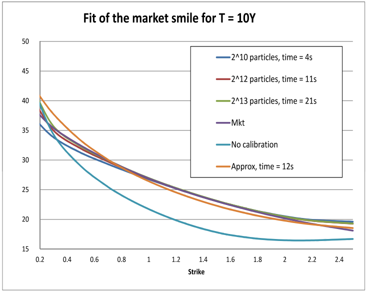
$$
\begin{aligned}
h(t, T) &= \mathrm{Var}(x_t^T) = \alpha_\theta^2 \Big((1-\theta)^2 e^{-2k_X(T-t)} \mathbb{E}\left[X_t^2\right] + \theta^2 e^{-2k_Y(T-t)} \mathbb{E}\left[Y_t^2\right] \\
&\quad + 2\theta(1-\theta) e^{-(k_X+k_Y)(T-t)} \mathbb{E}\left[X_t Y_t\right] \Big)
\end{aligned}
$$

$$
\mathbb{E}\left[X_t^2\right] = \frac{1-e^{-2k_X t}}{2k_X}, \quad \mathbb{E}\left[Y_t^2\right] = \frac{1-e^{-2k_Y t}}{2k_Y}, \quad \mathbb{E}\left[X_t Y_t\right] = \rho_{XY} \frac{1-e^{-(k_X+k_Y)t}}{k_X+k_Y}
$$

LV + Bergomi

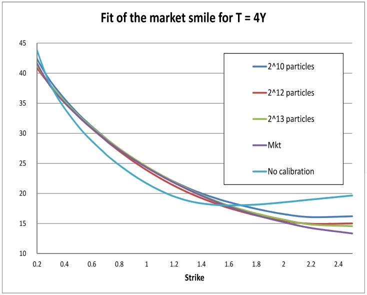
- Parameters of Bergomi model: $\nu = 200\%$, $\theta = 22.65\%$, $k_X = 4$, $k_Y = 12.5\%$, $\rho = 30\%$, $\rho_{\mathrm{SX}} = -50\%$, $\rho_{\mathrm{SY}} = -50\%$

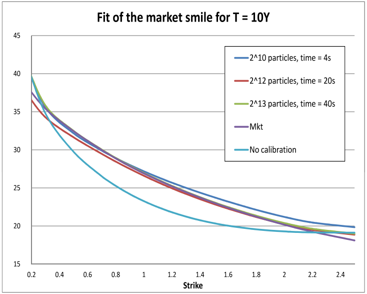




LV + Bergomi + Ho-Lee

- Ho-Lee parameters: $\sigma_{\rm r} = 1\%$, $\rho_{\rm Sr} = 40\%$
- Parameters of Bergomi model: $\nu = 200\%$, $\theta = 22.65\%$, $k_X = 4$, $k_Y = 12.5\%$, $\rho = 30\%$, $\rho_{\rm SX} = -50\%$, $\rho_{\rm SY} = -50\%$. Correlation Vol/Rate $= 0$
- Requires solving a 4d-PDE!



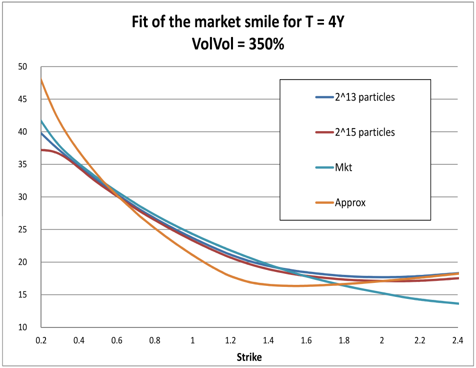


No optimization at all!

- Classical calibration procedures involve optimization over parameters: minimize the mean squared error between prices in the model and market prices.
- Time consuming, there could be local minima...
- And calibration to whole smile is imperfect (due to the finite number of parameters)
- It is remarkable that the particle method involves **no optimization at all**! And it gives perfect calibration.

Existence?

- Parameters of Bergomi model: $\nu = 350\%$, $\theta = 22.65\%$, $k_X = 4$, $k_Y = 12.5\%$, $\rho = 30\%$, $\rho_{\mathrm{SX}} = -50\%$, $\rho_{\mathrm{SY}} = -50\%$.



Existence?

$$\sigma(t,S,p)=\sigma_{\mathrm{loc}}(t,S)\sqrt{\frac{\int p(t,S,a)da}{\int a^2p(t,S,a)da}}$$

- Lipschitz condition w.r.t. $\mathbb{P}$ not satisfied $\Longrightarrow$ **uniqueness and existence results for calibrated SLV models are not at all obvious**
- Partial result by Abergel and Tachet (2010): the calibration problem for a SLV model is proved to be well posed (a) until some maturity T^*, (b) if the vol-of-vol is small enough, and (c) for suitably regularized initial conditions - not for a Dirac mass!
- Partial result by Jourdain and Zhou (2016): Regime-switching LV model = when the SV (a_t) is a pure jump process taking finitely many values, with jump intensities depending on the spot level. Range of (a_t^2) not too large $\Longrightarrow$ existence of a weak solution to the McKean SDE.
- Our numerical experiments with SLV show that the calibration fails for large vol-of-vol, whatever the algorithm used: PDE, particle method. Numerical issue? Non-existence?

2. Calibration of local correlation models to market smiles

Another example: calibration of local correlation (LC) models

- LC model for a triangle of FX rates (deterministic rates):

$$\begin{aligned} dS_t^1 &= (r_t^d - r_t^1)S_t^1\,dt + \sigma_1(t, S_t^1)S_t^1\,dW_t^1 \\ dS_t^2 &= (r_t^d - r_t^2)S_t^2\,dt + \sigma_2(t, S_t^2)S_t^2\,dW_t^2 \\ d\langle W^1, W^2\rangle_t &= \rho(t, S_t^1, S_t^2)\,dt \end{aligned}$$

- Example: $S^1 = \mathrm{EUR/USD}$, $S^2 = \mathrm{GBP/USD}$ and $S^{12} = \mathrm{EUR/GBP}$
- Model calibrated to market smile of cross rate $S^{12} := S^1/S^2$ iff for all t

$$\mathbb{E}_\rho^{\mathbb{Q}^f}\left[\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) - 2\rho(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)\,\middle|\,\frac{S_t^1}{S_t^2}\right] = \sigma_{12}^2\left(t, \frac{S_t^1}{S_t^2}\right) \quad (70)$$

- $\mathbb{Q}^f$ = risk-neutral measure associated to the foreign currency in S^2 (GBP):

$$\frac{d\mathbb{Q}^f}{d\mathbb{Q}} = \frac{S_T^2}{S_0^2}\exp\left(\int_0^T \left(r_t^2 - r_t^d\right)dt\right)$$

$$\frac{\mathbb{E}_\rho^{\mathbb{Q}}\left[S_t^2\left(\sigma_1^2(t, S_t^1) + \sigma_2^2(t, S_t^2) - 2\rho(t, S_t^1, S_t^2)\sigma_1(t, S_t^1)\sigma_2(t, S_t^2)\right)\,\middle|\,\frac{S_t^1}{S_t^2}\right]}{\mathbb{E}_\rho^{\mathbb{Q}}\left[S_t^2\,\middle|\,\frac{S_t^1}{S_t^2}\right]} = \sigma_{12}^2\left(t, \frac{S_t^1}{S_t^2}\right)$$

Local in cross correlation

- Assume ρ is local in cross: $\rho\left(t, \frac{S^1}{S^2}\right)$ (Reghai, G. and Henry-Labordère, 2011)
- Calibration condition reads

$$\mathbb{E}^{\mathbb{Q}^f}_{\rho}\left[\sigma_1^2(t,S_t^1)+\sigma_2^2(t,S_t^2)\,\middle|\,\frac{S_t^1}{S_t^2}\right]-2\rho\left(t,\frac{S_t^1}{S_t^2}\right)\mathbb{E}^{\mathbb{Q}^f}_{\rho}\left[\sigma_1(t,S_t^1)\sigma_2(t,S_t^2)\,\middle|\,\frac{S_t^1}{S_t^2}\right]=\sigma_{12}^2\left(t,\frac{S_t^1}{S_t^2}\right)$$

- Calibrated model follows the **McKean SDE**:

$$\begin{aligned}
dS_t^1 &= (r_t^d-r_t^1)S_t^1\,dt+\sigma_1(t,S_t^1)S_t^1\,dW_t^1\\
dS_t^2 &= (r_t^d-r_t^2)S_t^2\,dt+\sigma_2(t,S_t^2)S_t^2\,dW_t^2\\
d\langle W^1,W^2\rangle_t &= \frac{\mathbb{E}^{\mathbb{Q}^f}\left[\sigma_1^2(t,S_t^1)+\sigma_2^2(t,S_t^2)\,\middle|\,\frac{S_t^1}{S_t^2}\right]-\sigma_{12}^2\left(t,\frac{S_t^1}{S_t^2}\right)}{2\mathbb{E}^{\mathbb{Q}^f}\left[\sigma_1(t,S_t^1)\sigma_2(t,S_t^2)\,\middle|\,\frac{S_t^1}{S_t^2}\right]}dt\\
&= \frac{\mathbb{E}^{\mathbb{Q}}\left[S_t^2\left(\sigma_1^2(t,S_t^1)+\sigma_2^2(t,S_t^2)\right)\,\middle|\,\frac{S_t^1}{S_t^2}\right]-\sigma_{12}^2\left(t,\frac{S_t^1}{S_t^2}\right)\mathbb{E}^{\mathbb{Q}}\left[S_t^2\,\middle|\,\frac{S_t^1}{S_t^2}\right]}{2\mathbb{E}^{\mathbb{Q}}\left[S_t^2\sigma_1(t,S_t^1)\sigma_2(t,S_t^2)\,\middle|\,\frac{S_t^1}{S_t^2}\right]}dt
\end{aligned}$$

- Does the r.h.s. of last equation stay in $[-1,1]$?

Particle method for local in cross correl (G. and Henry-Labordère, 2011)

1. Set $k = 1$ and $\rho(t, S^1, S^2) = \frac{\sigma_1^2(0,S^1)+\sigma_2^2(0,S^2)-\sigma_{12}^2(0,\frac{S^1}{S^2})}{2\sigma_1(0,S^1)\sigma_2(0,S^2)}$ for $t \in [t_0, t_1]$
2. Simulate $(S_t^{1,i}, S_t^{2,i})_{1\leq i\leq N}$ from t_{k-1} to t_k using a discretization scheme
3. For all S^{12} in a grid G_{t_k} of cross rate values, compute non-parametric kernel estimates $E_{t_k}^{\mathrm{num}}(S^{12})$ and $E_{t_k}^{\mathrm{den}}(S^{12})$ of $\mathbb{E}_\rho^{\mathbb{Q}^f}\left[\sigma_1^2 + \sigma_2^2 \,\middle|\, \frac{S_t^1}{S_t^2}\right]$ and $\mathbb{E}_\rho^{\mathbb{Q}^f}\left[\sigma_1\sigma_2 \,\middle|\, \frac{S_t^1}{S_t^2}\right]$ at date t_k, define

$$\rho(t_k, S^{12}) = \frac{E_{t_k}^{\mathrm{num}}(S^{12}) - \sigma_{12}^2\left(t_k, S^{12}\right)}{2E_{t_k}^{\mathrm{den}}(S^{12})}$$

   interpolate $\rho(t_k,\cdot)$, e.g., using cubic splines, extrapolate, and, for all $t \in [t_k, t_{k+1}]$, set $\rho(t, S^1, S^2) = \rho(t_k, \frac{S^1}{S^2})$
4. Set $k := k + 1$. Iterate steps 2 and 3 up to the maturity date T

A new family of local correlation models (G., 2013)

- ρ local in cross/index = a particular modeling choice only guided by computational convenience
- "**Admissible correlation**" (FX) = any $\rho : [0,T] \times \mathbb{R}_+^* \times \mathbb{R}_+^* \to [-1,1]$ satisfying the calibration condition (70). How to build **all** admissible ρ?
- Let ρ be admissible. Pick two functions $a(t,S^1,S^2)$ and $b(t,S^1,S^2)$ such that b does not vanish and $a + b\rho$ is local in cross:

$$a(t,S^1,S^2) + b(t,S^1,S^2)\rho(t,S^1,S^2) \equiv f\left(t, \frac{S^1}{S^2}\right)$$

- We can always do so, e.g., $b \equiv 1$, $a(t,S^1,S^2) = f\left(t, \frac{S^1}{S^2}\right) - \rho(t,S^1,S^2)$
- Local in cross correl: $a \equiv 0$ and $b \equiv 1$ (Reghai, G. and Henry-Labordère)
- Local in cross volatility of the cross: $a = \sigma_1^2 + \sigma_2^2$ and $b = -2\sigma_1\sigma_2$ (Langnau, Kovrizhkin): $\rho^* = \frac{\sigma_1^2+\sigma_2^2-\sigma_{12}^2}{2\sigma_1\sigma_2}$

$$\sigma_{12}^2\left(t, \frac{S_t^1}{S_t^2}\right) = \mathbb{E}_\rho^{\mathbb{Q}^f}\left[\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2 \,\middle|\, \frac{S_t^1}{S_t^2}\right]$$
$$= \mathbb{E}_\rho^{\mathbb{Q}^f}\left[\sigma_1^2 + \sigma_2^2 + 2\frac{a}{b}\sigma_1\sigma_2 \,\middle|\, \frac{S_t^1}{S_t^2}\right] - 2(a+b\rho)\left(t, \frac{S_t^1}{S_t^2}\right)\mathbb{E}_\rho^{\mathbb{Q}^f}\left[\frac{\sigma_1\sigma_2}{b} \,\middle|\, \frac{S_t^1}{S_t^2}\right]$$

A new family of local correlation models (G., 2013)

$$\sigma_{12}^2\left(t,\frac{S_t^1}{S_t^2}\right)=\mathbb{E}_\rho^{\mathbb{Q}^f}\left[\sigma_1^2+\sigma_2^2+2\frac{a}{b}\sigma_1\sigma_2\,\middle|\,\frac{S_t^1}{S_t^2}\right]-2\,(a+b\rho)\left(t,\frac{S_t^1}{S_t^2}\right)\mathbb{E}_\rho^{\mathbb{Q}^f}\left[\frac{\sigma_1\sigma_2}{b}\,\middle|\,\frac{S_t^1}{S_t^2}\right]$$

$$\Longrightarrow\rho_{(a,b)}=\frac{1}{b}\left(\frac{\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f}\left[\sigma_1^2+\sigma_2^2+2\frac{a}{b}\sigma_1\sigma_2\,\middle|\,\frac{S_t^1}{S_t^2}\right]-\sigma_{12}^2}{2\mathbb{E}_{\rho_{(a,b)}}^{\mathbb{Q}^f}\left[\frac{\sigma_1\sigma_2}{b}\,\middle|\,\frac{S_t^1}{S_t^2}\right]}-a\right)$$

- Calibrated model follows the **McKean SDE**:

$$\begin{aligned}
dS_t^1 &= (r_t^d-r_t^1)S_t^1\,dt+\sigma_1(t,S_t^1)S_t^1\,dW_t^1\\
dS_t^2 &= (r_t^d-r_t^2)S_t^2\,dt+\sigma_2(t,S_t^2)S_t^2\,dW_t^2\\
d\langle W^1,W^2\rangle_t &= \left(\frac{\mathbb{E}^{\mathbb{Q}}\left[S_t^2\left(\sigma_1^2+\sigma_2^2+2\frac{a}{b}\sigma_1\sigma_2\right)\,\middle|\,\frac{S_t^1}{S_t^2}\right]-\sigma_{12}^2\left(t,\frac{S_t^1}{S_t^2}\right)\mathbb{E}^{\mathbb{Q}}\left[S_t^2\,\middle|\,\frac{S_t^1}{S_t^2}\right]}{2\mathbb{E}^{\mathbb{Q}}\left[S_t^2\frac{\sigma_1\sigma_2}{b}\,\middle|\,\frac{S_t^1}{S_t^2}\right]}\right.\\
&\qquad\left.-a(t,S_t^1,S_t^2)\vphantom{\frac{a}{b}}\right)dt/b(t,S_t^1,S_t^2)
\end{aligned}$$

Local in index correl for stock indices (G. and Henry-Labordère, 2011)

- Consider an index $I_t = \sum_{i=1}^{N} \alpha_i S_t^i$ made of N weighted stocks, each modeled using a local volatility (rates are deterministic):

$$dS_t^i = r_t S_t^i dt + S_t^i \sigma_i(t, S_t^i) dW_t^i, \qquad d\langle W^i, W^j \rangle_t = \rho_{ij}(t, S_t^1, \ldots, S_t^N) dt$$

- Model calibrated to basket smile iff for all t, I

$$I^2 \sigma_{\mathrm{loc}}^I(t, I)^2 = \sum_{i,j=1}^{N} \alpha_i \alpha_j \mathbb{E}[\rho_{ij}(t, S_t^1, \ldots, S_t^N) \sigma_i(t, S_t^i) \sigma_j(t, S_t^j) S_t^i S_t^j | I_t = I]$$

- Assume that correl is local in index (e.g., $\rho^0 = \rho^{\mathrm{hist}}$, $\rho^1 = \mathbf{1}$):

$$\rho(t, I) = (1 - \lambda(t, I))\rho^0 + \lambda(t, I)\rho^1, \qquad \lambda(t, I) \in [0, 1]$$

- The calibrated model then follows the **McKean SDE**

$$\begin{aligned} dS_t^i &= r_t S_t^i dt + S_t^i \sigma_i(t, S_t^i) dW_t^i, \qquad d\langle W^i, W^j \rangle_t = \rho_{ij}(t, I_t) dt \\ \lambda(t, I) &= \frac{I^2 \sigma_{\mathrm{loc}}^I(t, I)^2 - \sum_{i,j=1}^{N} \alpha_i \alpha_j \rho_{ij}^0 \mathbb{E}[\sigma_i(t, S_t^i) \sigma_j(t, S_t^j) S_t^i S_t^j | I_t = I]}{\sum_{i,j=1}^{N} \alpha_i \alpha_j (\rho_{ij}^1 - \rho_{ij}^0) \mathbb{E}[\sigma_i(t, S_t^i) \sigma_j(t, S_t^j) S_t^i S_t^j | I_t = I]} \end{aligned}$$

- Does λ stay in $[0, 1]$?

Extensions (G., 2013)

- Same trick works for building a family of local correlations that calibrate to the market smile of a stock index, by assuming that

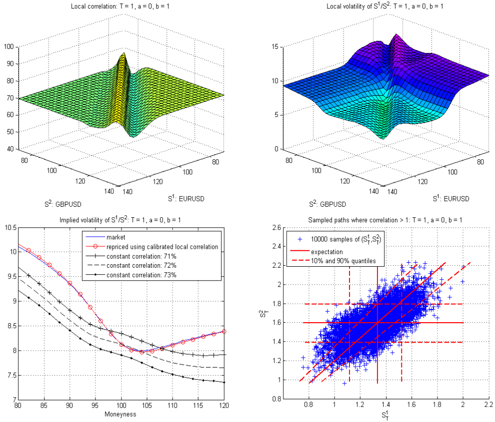
$$a(t, S^1, \dots, S^N) + b(t, S^1, \dots, S^N)\lambda(t, S^1, \dots, S^N) \equiv f(t, I_t)$$

  where $\rho = (1-\lambda)\rho^0 + \lambda\rho^1$, $\lambda \in [0, 1]$
- Another possible application: calibration to options on interest rate spreads
- The particle method and the construction of the new class of local correlations have been generalized (see G., 2013) to handle
  - LV+LC+stoch rates
  - LV+stoch vol+LC
  - LV+stoch vol+LC+stoch rates
  - LV+stoch vol+LC+stoch rates+stoch dividend yield

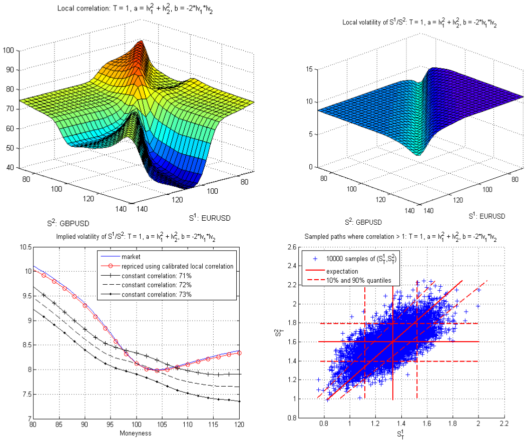
Numerical examples

- $S^1 = \mathrm{EURUSD}$, $S^2 = \mathrm{GBPUSD}$, $S^{12} = S^1/S^2 = \mathrm{EURGBP}$ (March 2012)
- $T = 1$
- $N = 10,000$ particles
- $\Delta t = \frac{1}{80}$
- $K(x) = (1 - x^2)^2 1_{\{|x| \leq 1\}}$
- Bandwidth $h = \kappa \bar{\sigma}^{12} S_0^{12} \sqrt{\max(t, t_{\min})} N^{-\frac{1}{5}}$, $\bar{\sigma}^{12} = 10\%$, $t_{\min} = 0.25$ and $\kappa \approx 3$
- The constant correlation that fits ATM implied volatility of cross rate at maturity $= 72\%$

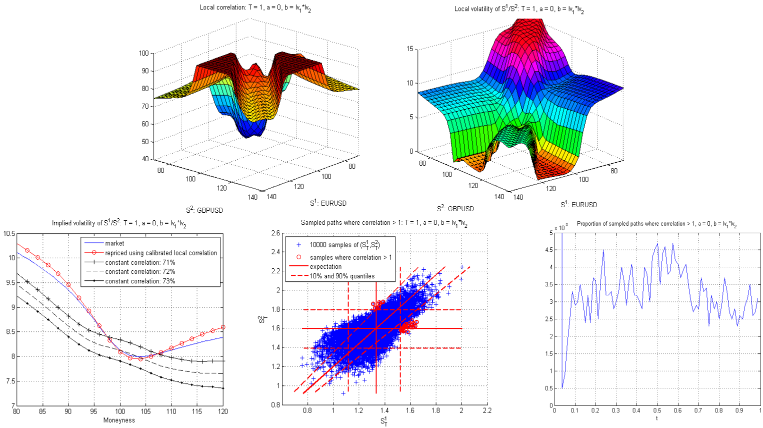
$a = 0,\ b = 1$ (local in cross correlation)



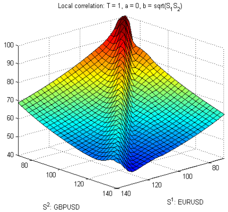
$a = \sigma_1^2 + \sigma_2^2$, $b = -2\sigma_1\sigma_2$ (local in cross volatility of the cross)

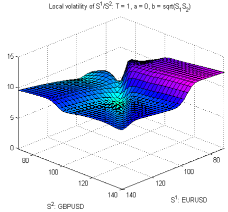


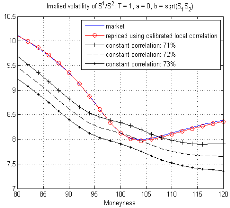
$a = 0,\ b = \sigma_1\sigma_2$ (local in cross covariance)

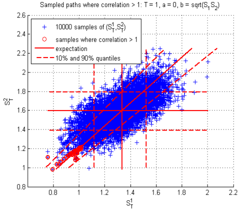


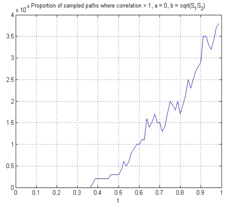
$a = 0,\ b = \sqrt{S^1 S^2}$



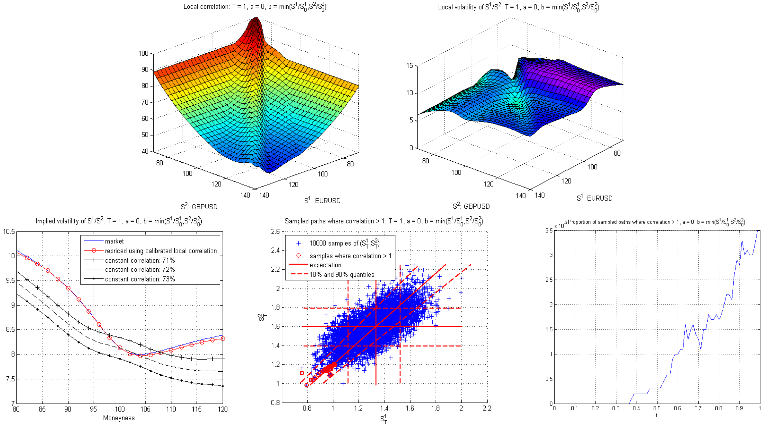




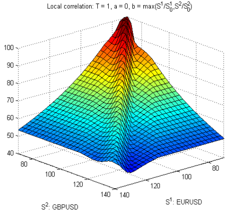


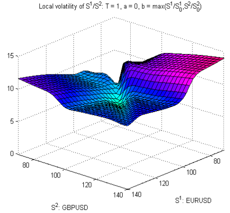


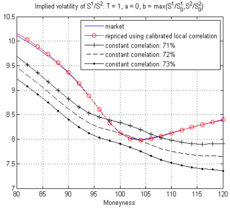
$a = 0,\ b = \min(S^1, S^2)$

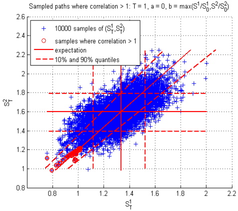

Local correlation: T = 1, a = 0, b = min(S1/S10,S2/S20)
S2: GBPUSD
S1: EURUSD
Local volatility of S1/S2: T = 1, a = 0, b = min(S1/S10,S2/S20)
Implied volatility of S1/S2: T = 1, a = 0, b = min(S1/S10,S2/S20)
market
repriced using calibrated local correlation
constant correlation: 71%
constant correlation: 72%
constant correlation: 73%
Moneyness
Sampled paths where correlation > 1: T = 1, a = 0, b = min(S1/S10,S2/S20)
10000 samples of (S1T,S2T)
samples where correlation > 1
expectation
10% and 90% quantiles
Proportion of sampled paths where correlation > 1, a = 0, b = min(S1/S10,S2/S20)
t

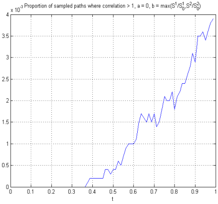
$a = 0,\ b = \max(S^1, S^2)$


Local correlation: T = 1, a = 0, b = max(S1/S10,S2/S20)
S2: GBPUSD
S1: EURUSD

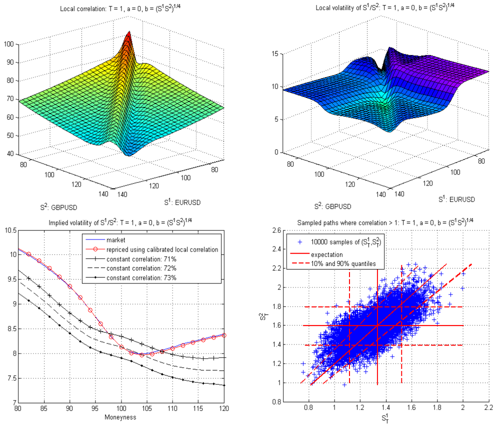

Local volatility of S1/S2: T = 1, a = 0, b = max(S1/S10,S2/S20)
S2: GBPUSD
S1: EURUSD


Implied volatility of S1/S2: T = 1, a = 0, b = max(S1/S10,S2/S20)
market
repriced using calibrated local correlation
constant correlation: 71%
constant correlation: 72%
constant correlation: 73%
Moneyness


Sampled paths where correlation > 1: T = 1, a = 0, b = max(S1/S10,S2/S20)
10000 samples of (S1T,S2T)
samples where correlation > 1
expectation
10% and 90% quantiles
S1T
S2T

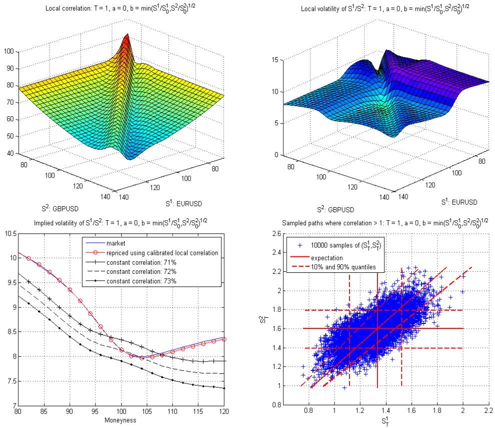

x 10-3 Proportion of sampled paths where correlation > 1, a = 0, b = max(S1/S10,S2/S20)
t

$a = 0,\ b = (S^1 S^2)^{1/4}$

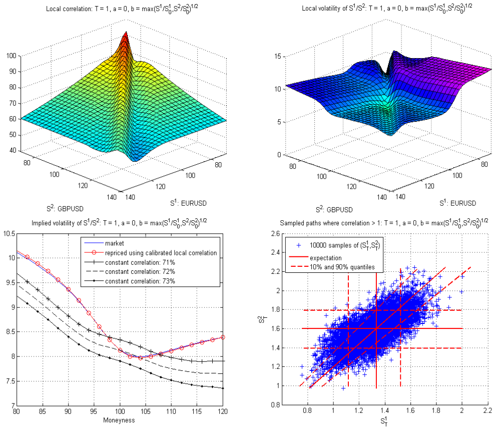


Local correlation: T = 1, a = 0, b = (S1S2)1/4
S2: GBPUSD
S1: EURUSD
Local volatility of S1/S2: T = 1, a = 0, b = (S1S2)1/4
S2: GBPUSD
S1: EURUSD
Implied volatility of S1/S2: T = 1, a = 0, b = (S1S2)1/4
market
repriced using calibrated local correlation
constant correlation: 71%
constant correlation: 72%
constant correlation: 73%
Moneyness
Sampled paths where correlation > 1: T = 1, a = 0, b = (S1S2)1/4
10000 samples of (S1T,S2T)
expectation
10% and 90% quantiles
S1T
S2T

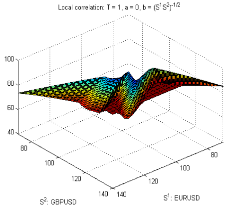
$a = 0,\ b = \sqrt{\min(S^1, S^2)}$

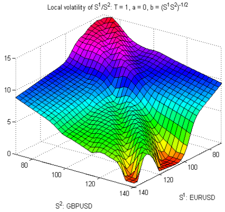


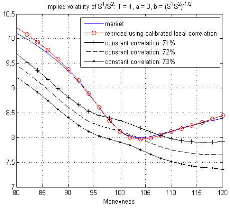
$a = 0,\ b = \sqrt{\max(S^1, S^2)}$

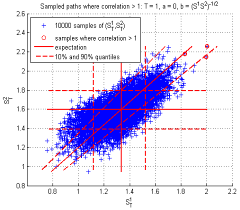


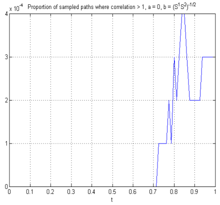
$a = 0,\ b = \frac{1}{\sqrt{S^1 S^2}}$



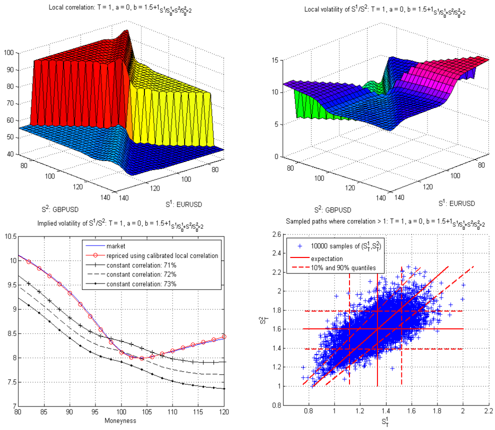




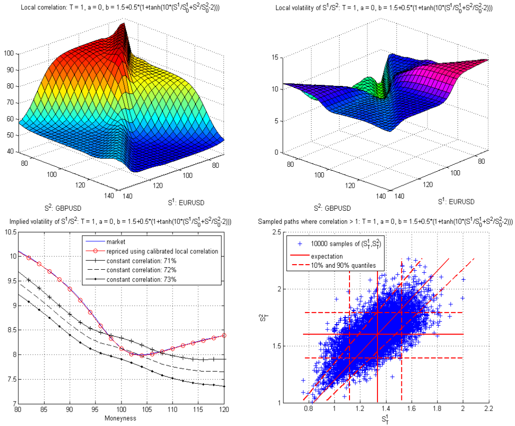




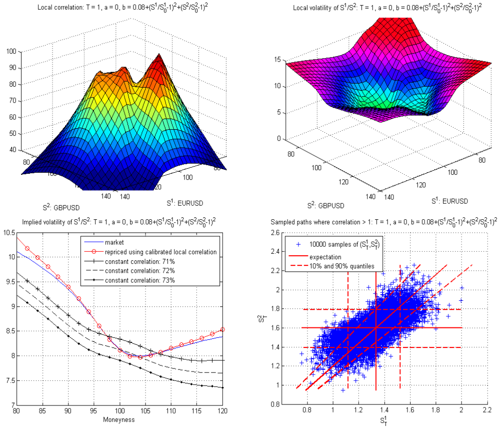
$a = 0,\ b = 1.5 + 1_{\left\{\frac{S^1}{S^1_0} + \frac{S^2}{S^2_0} > 2\right\}}$



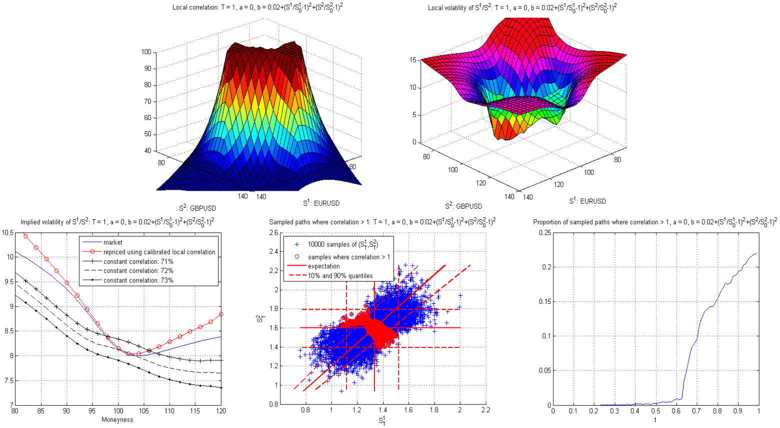
$$a = 0,\ b = 1.5 + \frac{1}{2}\left(1 + \tanh\left(10\left(\frac{S^1}{S^1_0} + \frac{S^2}{S^2_0} - 2\right)\right)\right)$$



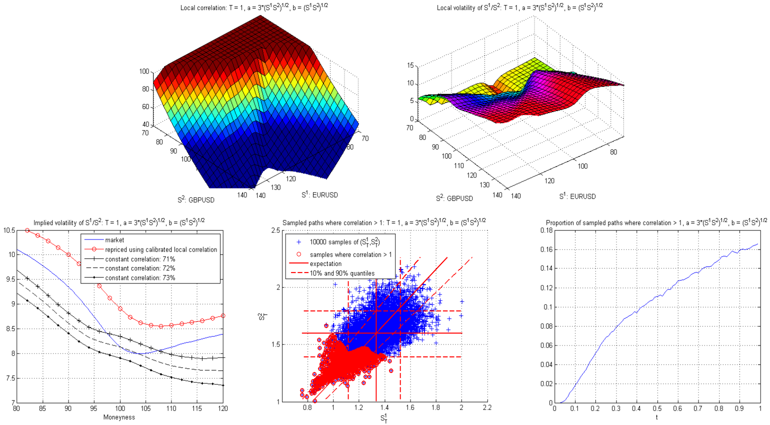
$a = 0,\ b = 0.08 + \left(\frac{S^1}{S^1_0} - 1\right)^2 + \left(\frac{S^2}{S^2_0} - 1\right)^2$



$a = 0,\ b = 0.02 + \left(\frac{S^1}{S^1_0} - 1\right)^2 + \left(\frac{S^2}{S^2_0} - 1\right)^2$ (bad choice of (a, b))



$a = 3\sqrt{S^1 S^2}$, $b = \sqrt{S^1 S^2}$ (bad choice of (a, b))



Path-dependent volatility (G., 2013)

- The $a+b\rho$ (or $a+b\lambda$) trick also works for path-dep correlation models:

$$a(t,S^1,S^2,X)+b(t,S^1,S^2,X)\rho(t,S^1,S^2,X)\equiv f\left(t,\frac{S^1}{S^2}\right)$$

- X can be any path-dep variable: running/moving averages, maximums/mimimums, realized correlations/variances (cf GARCH), etc.
- What works for correl works for vol! Path-dep vol model:

$$\frac{dS_t}{S_t}=(r_t-q_t)\,dt+\sigma(t,S_t,X_t)\,dW_t$$

- Determ. rates and div yield: model calibrated to the smile iff for all t

$$\mathbb{E}[\sigma(t,S_t,X_t)^2|S_t]=\sigma_{\mathrm{loc}}^2(t,S_t)$$

- All calibrated path-dep vol models can be built by picking a particular $\sigma(t,S,X)$ and computing $l(t,S)$ such that

$$\mathbb{E}[\sigma(t,S_t,X_t)^2 l(t,S_t)^2|S_t]=\sigma_{\mathrm{loc}}^2(t,S_t)$$

using the particle method, i.e.,

$$l(t,S)=\frac{\sigma_{\mathrm{loc}}(t,S)}{\sqrt{\mathbb{E}[\sigma(t,S_t,X_t)^2|S_t=S]}}$$

Path-dependent volatility (G., 2013)

- Calibrated model follows the **McKean SDE**

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma(t, S_t, X_t)\frac{\sigma_{\text{loc}}(t, S_t)}{\sqrt{\mathbb{E}[\sigma(t, S_t, X_t)^2 | S_t]}}\,dW_t$$

- Complete model: prices are uniquely defined. Joint spot-vol dynamics richer than in the local vol and stochastic vol models. Captures joint historical behaviour of vol and spot returns
- Extension to stochastic vol, stochastic rates, stochastic div yield is easy:

$$\frac{dS_t}{S_t} = (r_t - q_t)\,dt + \sigma(t, S_t, X_t)a_t\,dW_t$$

Model calibrated to the smile iff for all t, K

$$\frac{\mathbb{E}[D_{0t}\sigma^2(t, S_t, X_t)a_t^2 | S_t = K]}{\mathbb{E}[D_{0t}|S_t = K]} = \sigma^2_{\text{loc}}(t, K)$$
$$-\frac{\mathbb{E}\left[D_{0t}\left(r_t - q_t - (r_t^0 - q_t^0)\right)1_{S_t > K}\right]}{\frac{1}{2}K\partial_K^2\mathcal{C}(t, K)} + \frac{\mathbb{E}\left[D_{0t}\left(q_t - q_t^0\right)(S_t - K)^+\right]}{\frac{1}{2}K^2\partial_K^2\mathcal{C}(t, K)}$$

where σ_{loc} is the Dupire local volatility computed using r_t^0 and q_t^0

Conclusion

- The particle method is a very powerful tool for smile calibration problems
- **Single asset**: We calibrate to smile virtually any diffusive model with a LV component, possibly including any combination of stoch vol, path-dep vol, stoch rates, and stoch dividend yield
- **Multi-asset**: We calibrate to basket smile virtually any diffusive model with a LC component, possibly including any combination of LV, stoch vol, path-dep correl, stoch rates, and stoch dividend yield
- The particle method is fast: as fast as a standard MC
  - For low-dimensional models, it outperforms usual PDE implementations
  - For high-dimensional models, it is the first exact calibration method
- It is easy to implement
- It is robust: unlike PDE methods, it is insensitive to dimension
- + We can deal exactly with dividends (cash+yield) + Calibration and pricing can be achieved within the same MC procedure
- + There are very nice mathematics behind the scenes (McKean SDEs, propagation of chaos)

A few selected references

- Abergel, F., Tachet, R. : *A nonlinear partial integrodifferential equation from mathematical finance*, Discrete Cont. Dynamical Systems, Serie A, vol. 27, 3, p. 907-917, 2010.
- Bergomi, L., Guyon, J.: *Stochastic Volatility's Orderly Smiles*, Risk Magazine, May 2012.
- Corlay, S., *B-spline techniques for volatility modeling*, available at http://hal.archives-ouvertes.fr/hal-00830378, 2013.
- Dupire, B.: *Pricing with a smile*, Risk, January, 1994.
- Guyon, J. and Henry-Labordère, P., *Being Particular About Calibration*, Risk, Jan. 2012. Long version in Post-Crisis Quant Finance, Risk Books, 2013, and Quant of the Year 2000-2014, Risk Books, 2014.
- Guyon, J. and Henry-Labordère, P., *Nonlinear Option Pricing*, Chapman & Hall/CRC Financial Mathematics Series, Dec. 2013.
- Guyon, J. *A New Class of Local Correlation Models*, Risk, February 2014.

A few selected references

- Guyon, J. *Path-Dependent Volatility*, Risk, October 2014.
- Hull, J., White, A.: *The pricing of options on assets with stochastic volatilities*, The Journal of Finance, 42(2):281–300, 1987.
- Méléard, S. : *Probabilistic Models for Nonlinear Partial Differential Equations*, Series: Lecture Notes in Mathematics, Vol. 1627, Springer.
- Renault, E., Touzi, N.: *Option pricing and implied volatilities in a stochastic volatility model*, Mathematical Finance, 6:279–302, 1996.
- Sznitman, A.S. : *Topics in propagation of chaos*, Ecole d'été de probabilités de Saint-Flour XIX - 1989, volume 1464 of Lect. Notes in Math. Springer-Verlag (1991).