returns of bonds

Fixed Income Products and Analysis

Bonds Yield rates

In this lecture. . .

prices

- names and features of the basic and most important fixed-income products

derivatives

- swaps
- the relationship between swaps and zero-coupon bonds
- an overview of fixed-income modeling

dr

- simple ways to analyze the market value of the instruments: yield, duration and convexity

Risk

- how to construct yield curves and forward rates

Bootstrapping

By the end of this lecture you will

- be able to decompose a swap into a portfolio of bonds

- understand the main approaches to interest-rate modeling

- be able to construct the forward curve from simple bonds

- understand the concepts of yield, duration and convexity

Introduction

Fixed Income.

This lecture is an introduction to some basic instruments and concepts in the world of fixed income, that is, the world of cash-flows that are in the simplest cases independent of any stocks, commodities etc.

We will see the most elementary of fixed-income instruments, the coupon-bearing bond, and how to determine various properties of such bonds to help in their analysis.

$100

CF.

Simple fixed-income contracts and features

The zero-coupon bond

Price $100

Par / Face / Nominal / Notional :

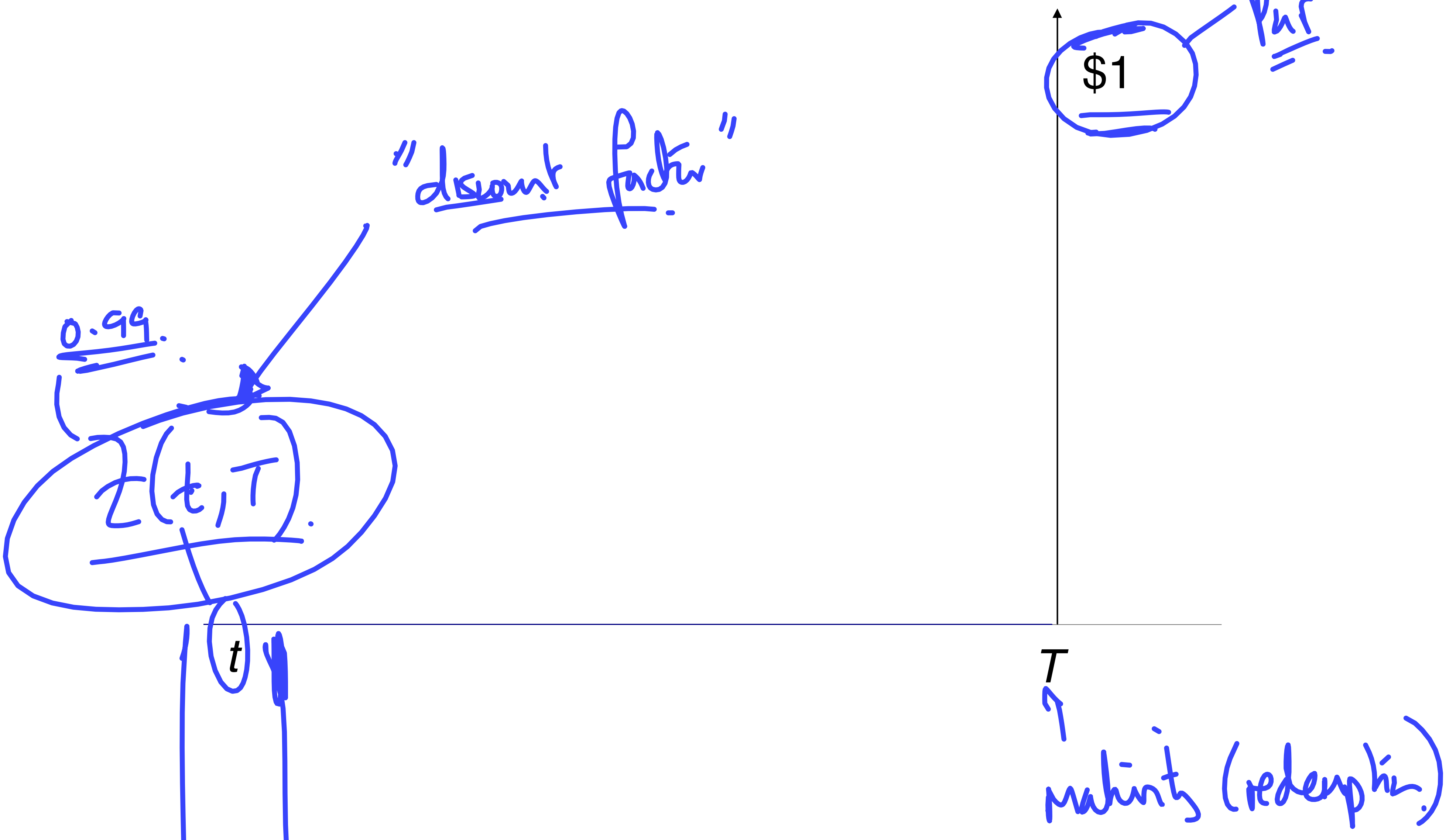
- The **zero-coupon bond** is a contract paying a known fixed amount, the **principal**, at some given date in the future, the **maturity** date T.

Size :

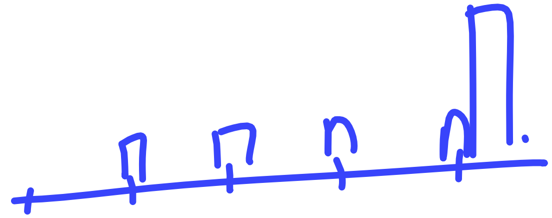
This promise of future wealth is worth something now: it cannot have zero or negative value.

Furthermore, the amount you pay initially will be smaller than the amount you receive at maturity.

ZCB



The coupon-bearing bond



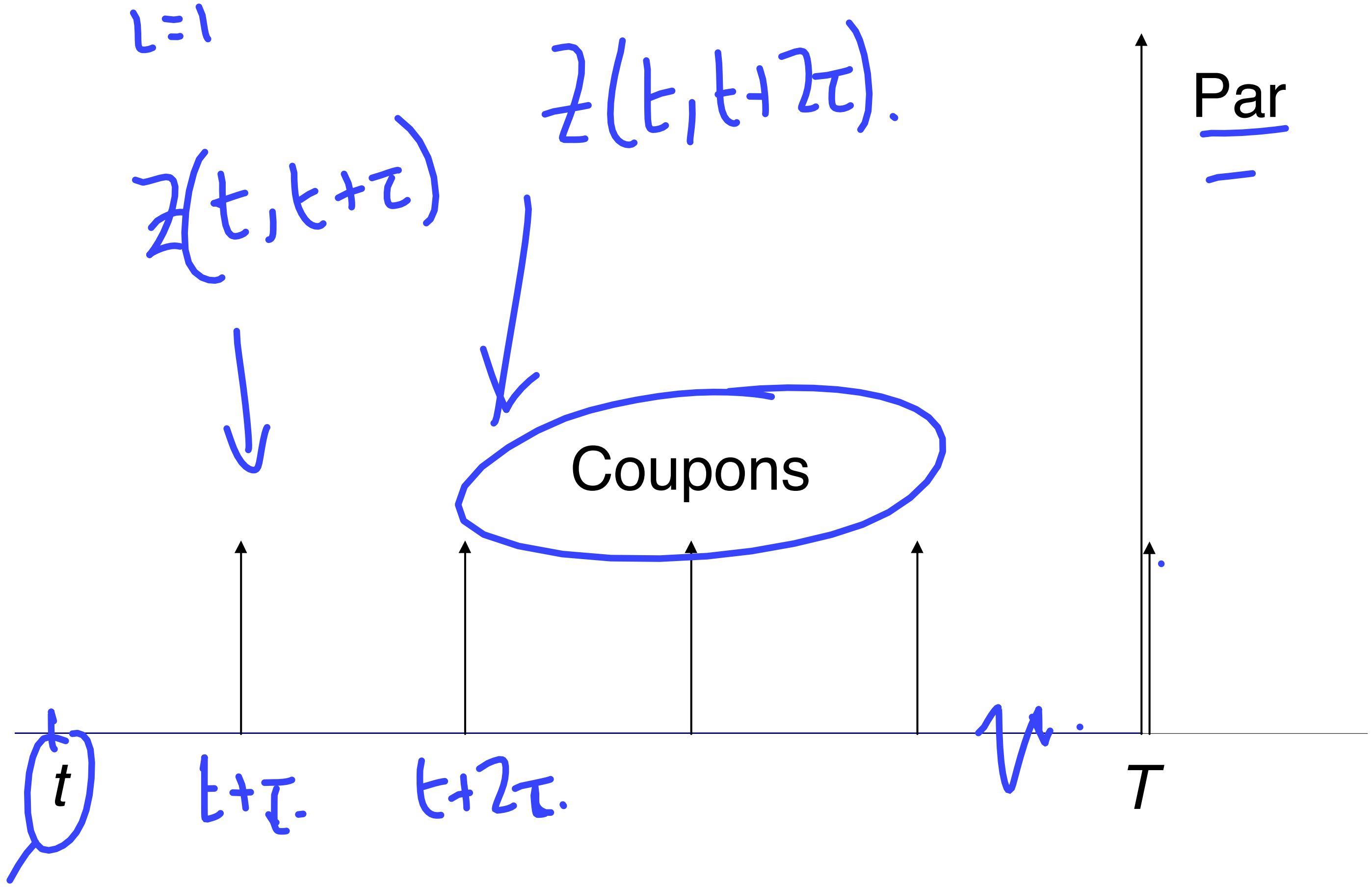
- A **coupon-bearing bond** is similar to the above except that as well as paying the principal at maturity, it pays smaller quantities, the coupons, at intervals up to and including the maturity date.

% of par. 4% bond.

These coupons are usually specified fractions of the principal. For example, the bond pays \$1 in 10 years and 2%, i.e. 2 cents, every six months.

This bond is clearly more valuable than the bond in the previous example because of the coupon payments.

$$\text{Price} = c\sum_{i=1}^{n} Z(t, t+i\tau) + Z(t,T).$$



We can think of the coupon-bearing bond as a portfolio of zero-coupon bearing bonds:

- one zero-coupon bearing bond for each coupon date with a principal being the same as the original bond's coupon, and then a final zero-coupon bond with the same maturity as the original.

A bank account

- Simply an account that accumulates interest compounded at a rate that varies from time to time.

The rate at which interest accumulates is usually a short-term and unpredictable rate.

In the sense that money held in a bank account will grow at an unpredictable rate, such an account is risky when compared with a one-year zero-coupon bond.

On the other hand, the bank account can be closed at any time but if the bond is sold before maturity there is no guarantee how much it will be worth at the time of the sale.

FRN

reference rate BLM + margin $r(t)$

Floating rate bonds

In its simplest form a **floating interest rate** is the amount that you get on your bank account. This amount varies from time to time, reflecting the state of the economy. This uncertainty is compensated by the flexibility of your deposit, it can be withdrawn at any time.

- The most common measure of interest is **London Interbank Offer Rate or LIBOR***. LIBOR comes in various maturities, one month, three month, six month etc., and is the rate of interest offered between Eurocurrency banks for fixed-term deposits.

UK → SONIA

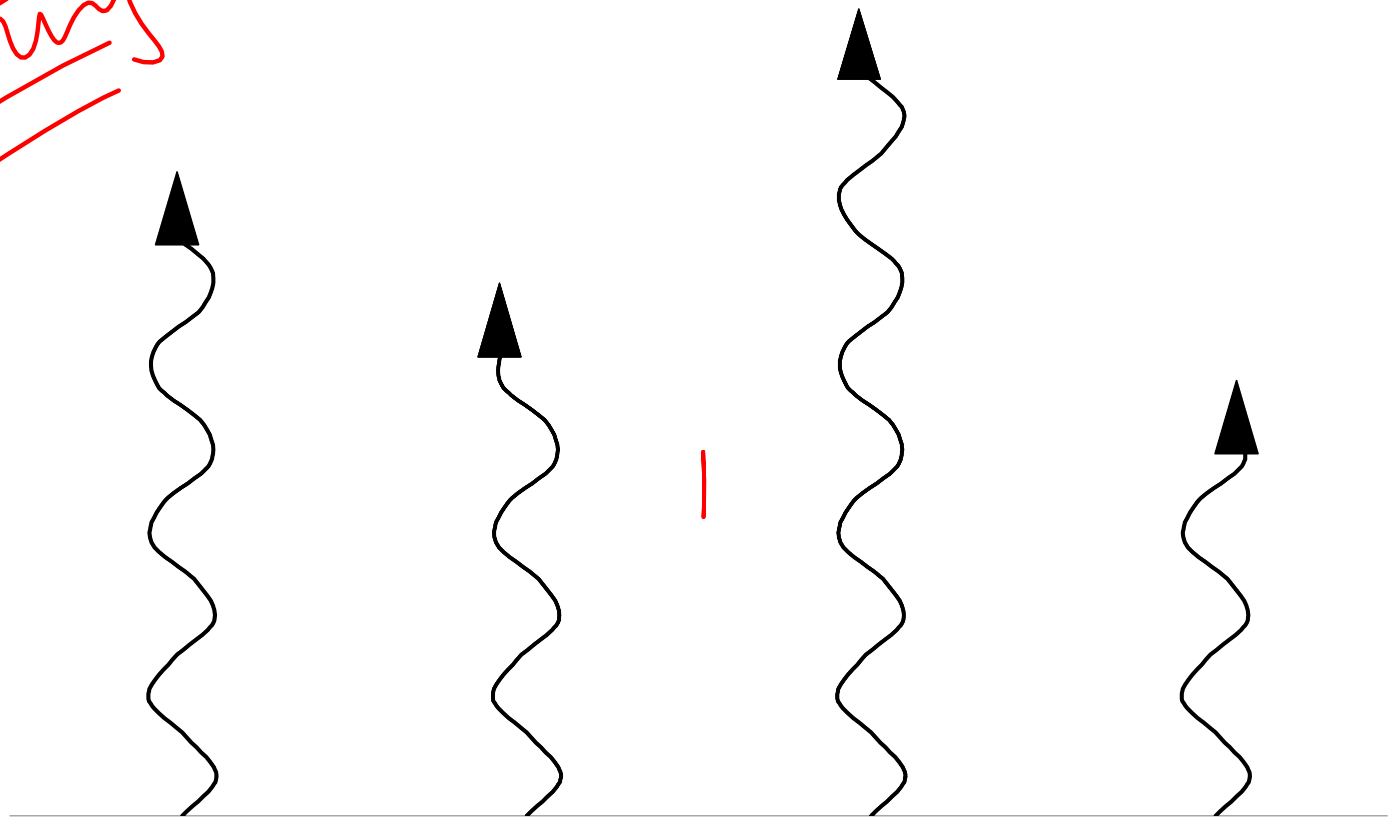
**Regulators are pushing to phase out LIBOR as a benchmark by the end of 2021. Alternative bechmark rates are based on overnight funding rates which are often used in practice as a proxy for risk-free institutional borrowing costs.*

https://www.bis.org/publ/qtrpdf/r_qt1903e.pdf

US → SOFR

Fixed

Floating



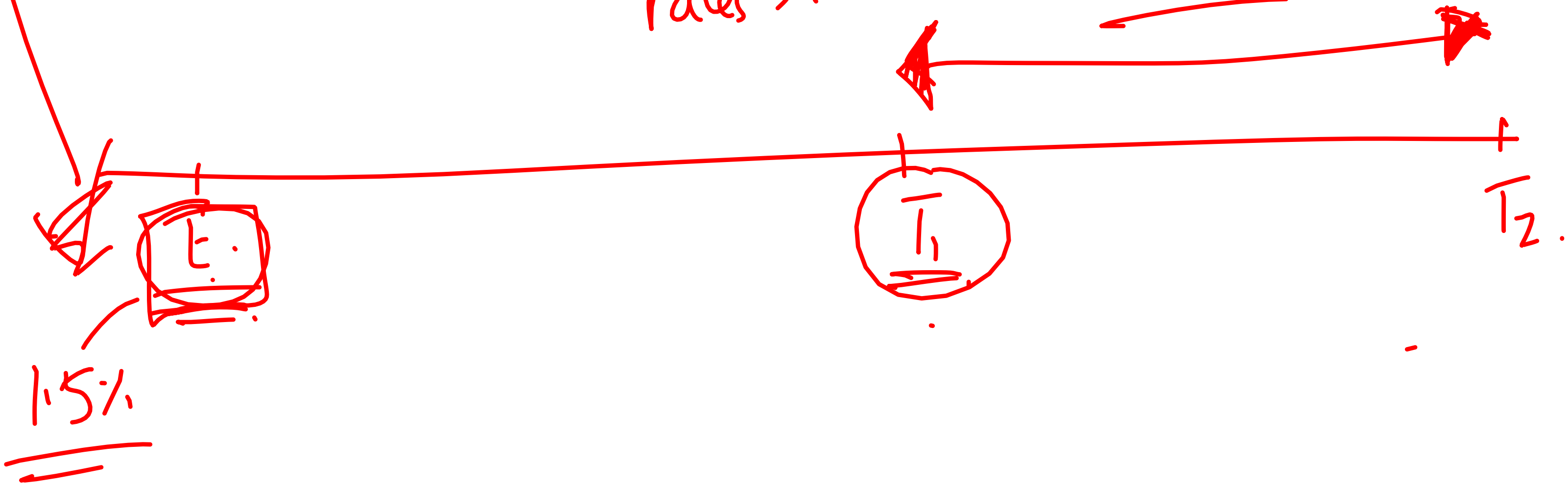
Derivative:

Forward rate agreements

LONG FRA ↑ rates

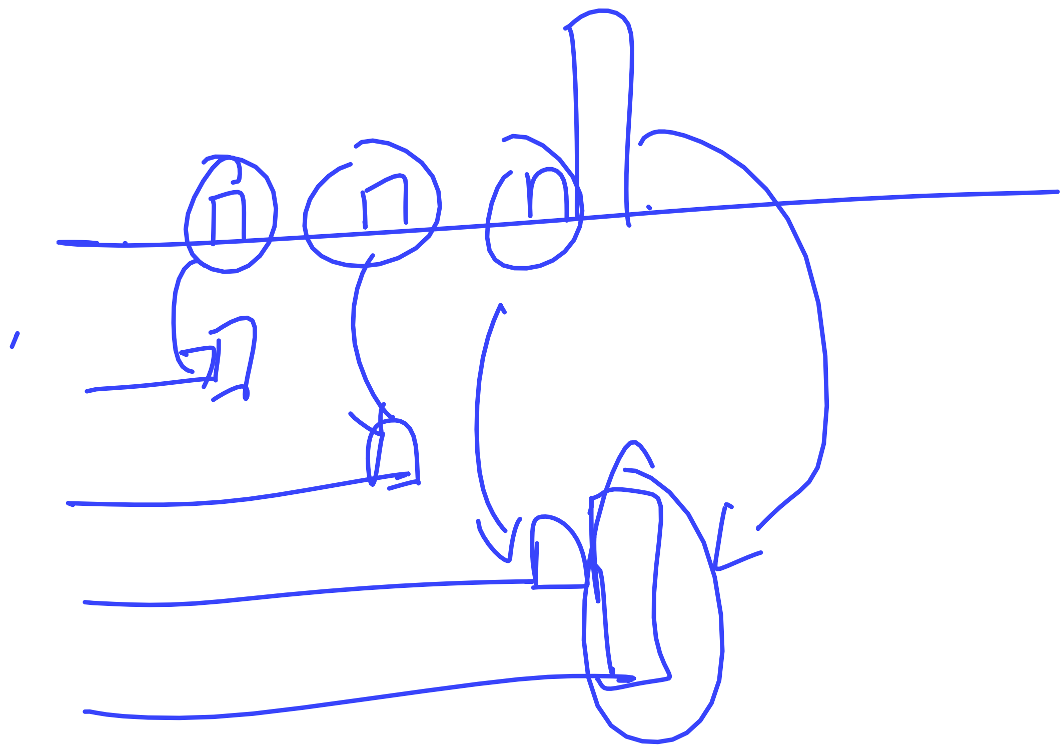
- A **Forward Rate Agreement** (**FRA**) is an agreement between two parties that a prescribed interest rate will apply over some specified period in the future.

rates > 1.5% ⇒ LONG WINS



STRIPS

STRIPS stands for 'Separate Trading of Registered Interest and Principal of Securities'. The coupons and principal of normal bonds are split up, creating artificial zero-coupon bonds of longer maturity than would otherwise be available.



Amortization

Securities

In all of the above products ~~we have~~ assumed that the principal remains fixed at its initial level.

repaying loan

- Sometimes this is not the case, the principal can **amortize** or decrease during the life of the contract. The principal is thus paid back gradually and interest is paid on the amount of the principal outstanding.

Such amortization is arranged at the initiation of the contract and may be fixed, so that the rate of decrease of the principal is known beforehand, or can depend on the level of some index, if the index is high the principal amortizes faster for example.

Call provision callable bond

Some bonds have a **call provision**. The issuer can call back the bond on certain dates or at certain periods for a prescribed, possibly time-dependent, amount. This lowers the value of the bond.

Callable Bond = Straight Bond − call option.

Swaps

- A **swap** is an agreement between two parties to exchange, or swap, future cashflows.

The size of these cashflows is determined by some formulæ, decided upon at the initiation of the contract. The swaps may be in a single currency or involve the exchange of cashflows in different currencies.

The total notional principal amount is,in US dollars,currently comfortably in *15* figures.

(https://www.bis.org/statistics/derstats.htm)

The vanilla interest rate swap

In the **interest rate swap** the two parties exchange cashflows that are represented by the interest on a notional principal. Typically,

KNOWN AT START.

- one side agrees to pay the other a fixed interest rate

PRICE

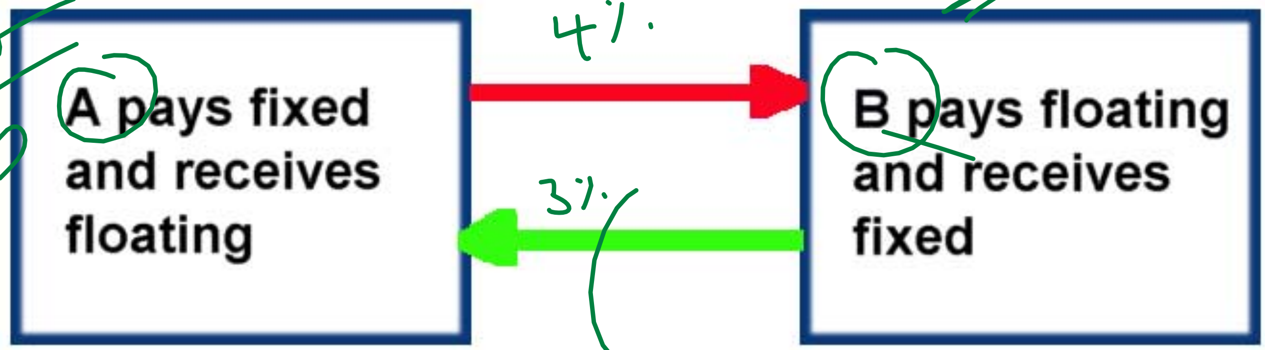
- the cashflow in the opposite direction is a **floating rate**.

One of the commonest floating rates used in a swap agreement is LIBOR*, London Interbank Offer Rate.

Pay Fixed

Swap

Receive Fixed



A wins when rates ↑

Periodically — netted off

The parties to a simple interest rate swap.

Commonly in a swap, the exchange of the fixed and floating interest payments occur every six months.

In this case the relevant LIBOR rate would be the six-month rate. At the maturity of the contract the principal is *not* exchanged.

Why are swaps so popular?

1. Comparative advantage

2. Hedging

3. Speculation

Swaps: Comparative advantage

Swaps were first created to exploit **comparative advantage**. This is when two companies who want to borrow money are quoted fixed and floating rates such that by exchanging payments between themselves they benefit, at the same time benefitting the intermediary who puts the deal together.

Here's an example.

Two companies A and B want to borrow \$10MM, to be paid back in two years. They each have a choice of a fixed- or floating-rate loan. They are quoted the interest rates for borrowing at fixed and floating rates shown here.

	Fixed	Floating
A	7%	six-month LIBOR + 30bps
B	8.2%	six-month LIBOR + 100bps

Note that both must pay a premium over LIBOR to cover risk of default, which is perceived to be greater for company B.

Ideally, company A wants to borrow at floating and B at fixed. This will possibly be because of the cashflows in their business.

If they each borrow directly then they pay the following in total:

six-month LIBOR + 30bps +8.2% = six-month LIBOR + 8.5%.

A	six-month LIBOR + 30bps (floating)
B	8.2% (fixed)

However, if A borrowed at fixed and B at floating they'd only be paying

$$\text{six-month LIBOR} + 100\text{bps} + 7\% = \text{six-month LIBOR} + 8\%.$$

A	7% (fixed)
B	six-month LIBOR + 100bps (floating)

That's a saving between them of 0.5%.

Let's suppose that A borrows fixed and B floating, even though that's not what they want. Their total interest payments are six-month LIBOR plus 8%.

Now let's see what happens if we throw a swap into the pot.

A is currently paying 7% and B six-month LIBOR plus 1%.

They enter into a swap in which A pays LIBOR to B and B pays 6.95% to A.

They have swapped interest payments.

Looked at from A's perspective they are paying 7% and LIBOR while receiving 6.95%, a net floating payment of LIBOR plus 5bps. Not only is this floating, as A originally wanted, but it is 25bps better than if they had borrowed directly at the floating rate.

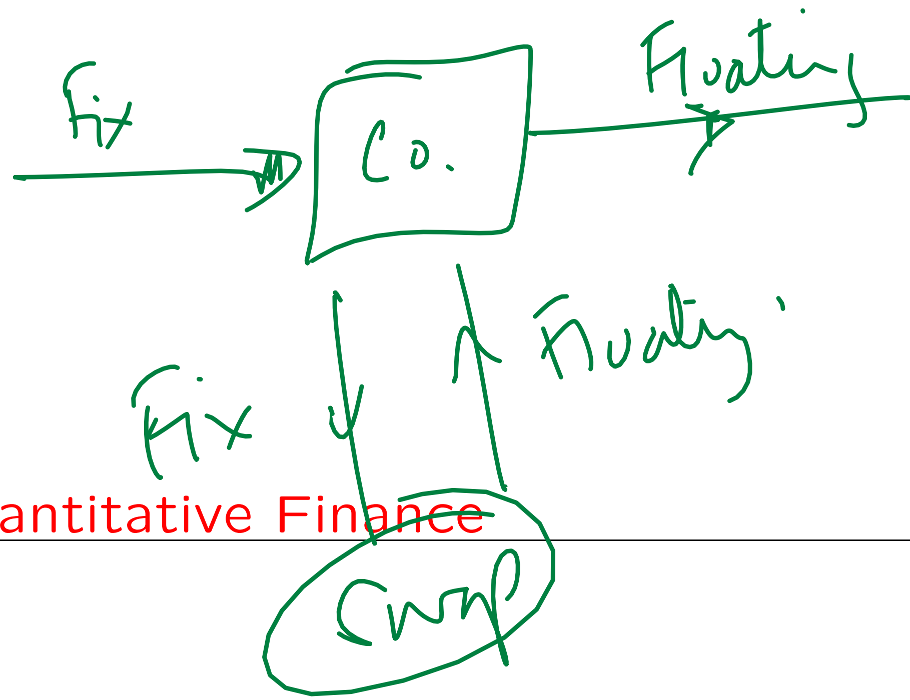
There's still another 25bps missing, and, of course, B gets this. B pays LIBOR plus 100bps and also 6.95% to A while receiving LIBOR from A. This nets out at 7.95%, which is fixed, as required, and 25bps less than the original deal.

Swaps: Hedging

Swaps can be used to balance cashflows.

Income might be at a fixed rate with outgoings at a floating rate.

A simple example would be the cashflows associated with a rented out property: Income would be fixed, the rent; Outgoings might vary, the mortgage.



Swaps: Speculation

Because the floating legs of the swap vary with the level of interest rates, swaps can be used to speculate on the future direction of these rates.

The par swap starts life with zero value but gives immediate exposure to interest rates. Once interest rates move the swap will have a non-zero value. This may be positive or negative depending on the direction in which the floating legs move.

The swap can then be closed out resulting in a profit or loss.

Let's see an example of how a swap works.

Example

Suppose that we enter into a five-year swap on 8th July 2018, with semi-annual interest payments.

Price

We will pay to the other party a rate of interest fixed at 6% on a notional principal of $100 million,the counterparty will pay us six-month LIBOR.

Annualised Rate.

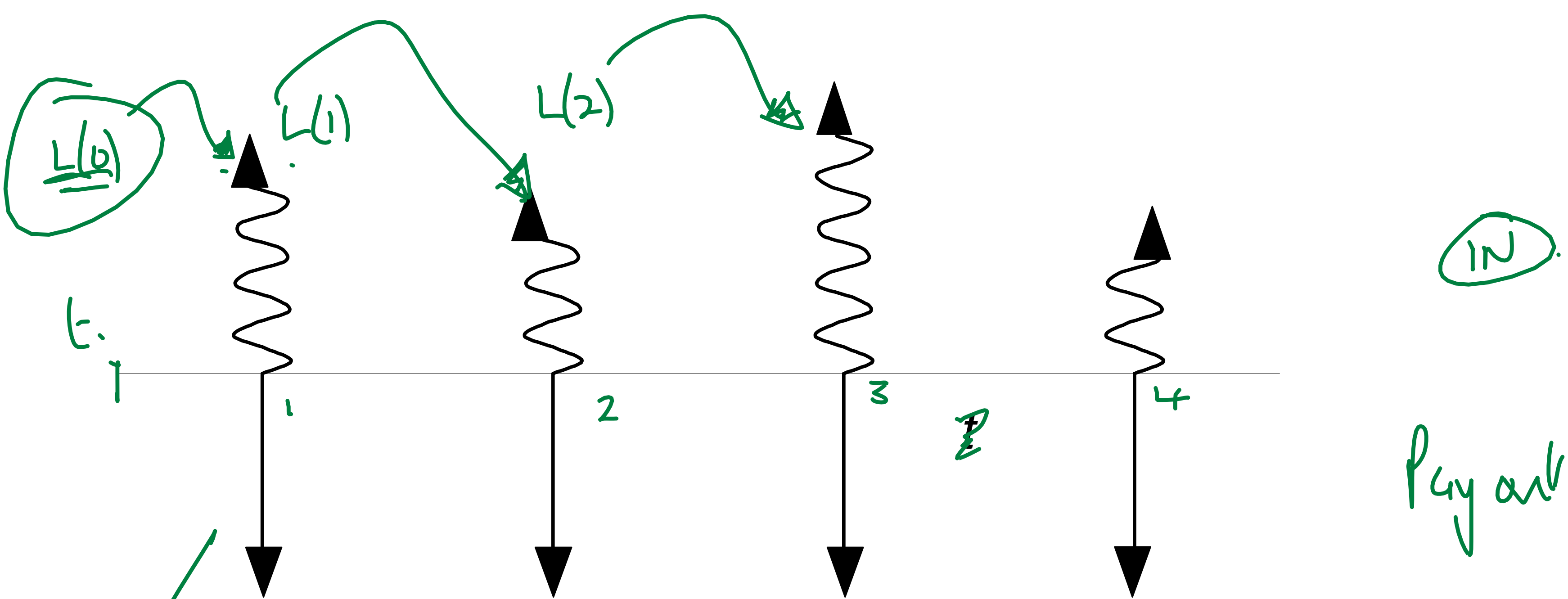
3% every 6m.

period

The cashflows in this contract are shown below.

known at outset → paid at end

The straight lines denote a fixed rate of interest and thus a known amount, the curly lines are floating rate payments.



$0.03 \times \$100m = \$3m$

The first exchange of payments is made on 8th January 2019, six months after the deal is signed. How much money changes hands on that first date?

We must pay $0.03 \times \$100{,}000{,}000 = \$3{,}000{,}000$.

The cashflow in the opposite direction will be at six-month LIBOR, *as quoted six months previously* i.e. at the initiation of the contract.

This is a very important point.

- The LIBOR rate is set six months before it is paid, so that in the first exchange of payments the floating side is known. This makes the first exchange special.

The second exchange takes place on 8th July 2019. Again we must pay $3,000,000, but now we receive LIBOR, as quoted on 8th January 2019.

Netted off.

- Every six months there is an exchange of such payments, with the fixed leg always being known and the floating leg being known six months before it is paid.

This continues until the last date, 8th July 2023

How did we price this swap? −6%

"Fair" payment: NPV = 0 at initiation.

Relationship between swaps and bonds

There are two sides to a swap, the fixed-rate side and the floating-rate side.

- The fixed interest payments, since they are all known in terms of actual dollar amount, can be seen as the sum of zero-coupon bonds.

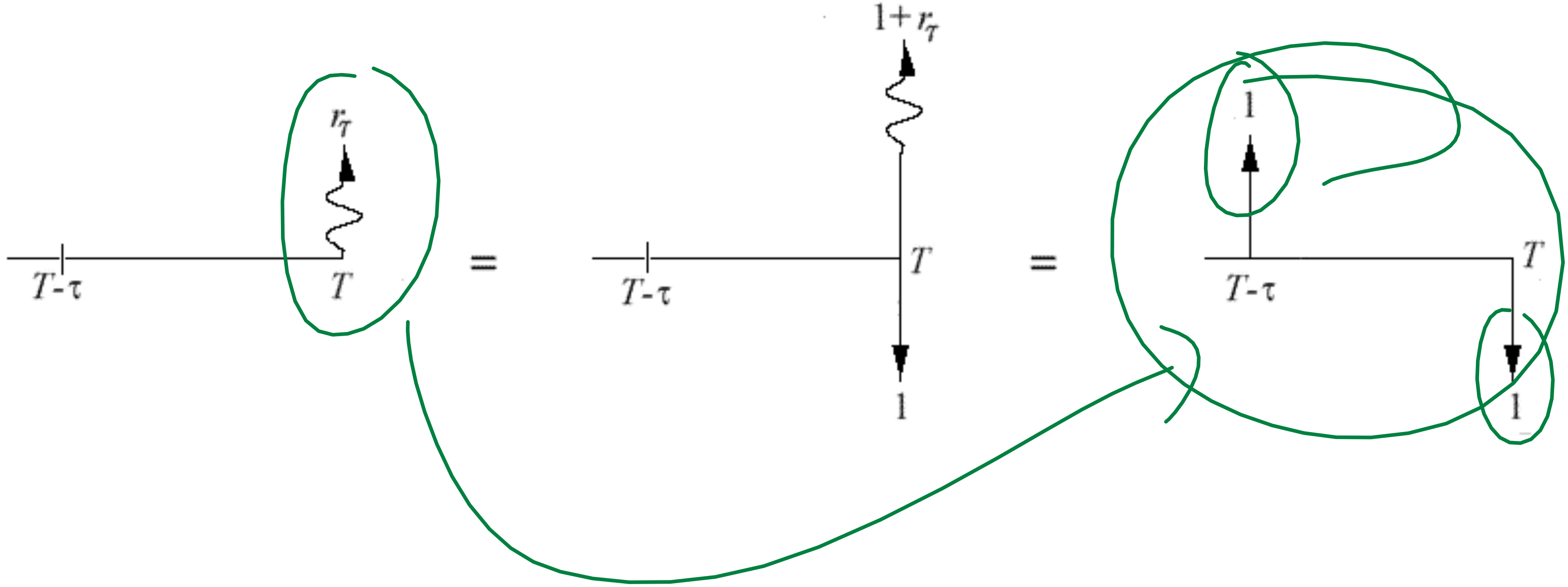
If the fixed rate of interest is r_s then the fixed payments add up to

$$r_s \tau \sum_{i=1}^{N} Z(t; T_i),$$

where τ is the time interval between payments measured in years.

(This assumes a principal of 1.)

To see the simple relationship between the floating leg and zero-coupon bonds we draw some schematic diagrams and compare the cashflows.



At time T_i there is payment of $r_\tau \tau$ of the notional principal, where r_τ is the period τ rate of LIBOR, set at time $T_i - \tau$.

Add and subtract \$1 at time T_i to get the second diagram. The first and the second diagrams obviously have the same present value.

Now recall the precise definition of LIBOR. It is the interest rate paid on a fixed-term deposit. Thus the $\$1 + r_\tau \tau$ at time T_i is the same as \$1 at time $T_i - \tau$. This gives the third diagram.

- It follows that the single floating rate payment is equivalent to two zero-coupon bonds.

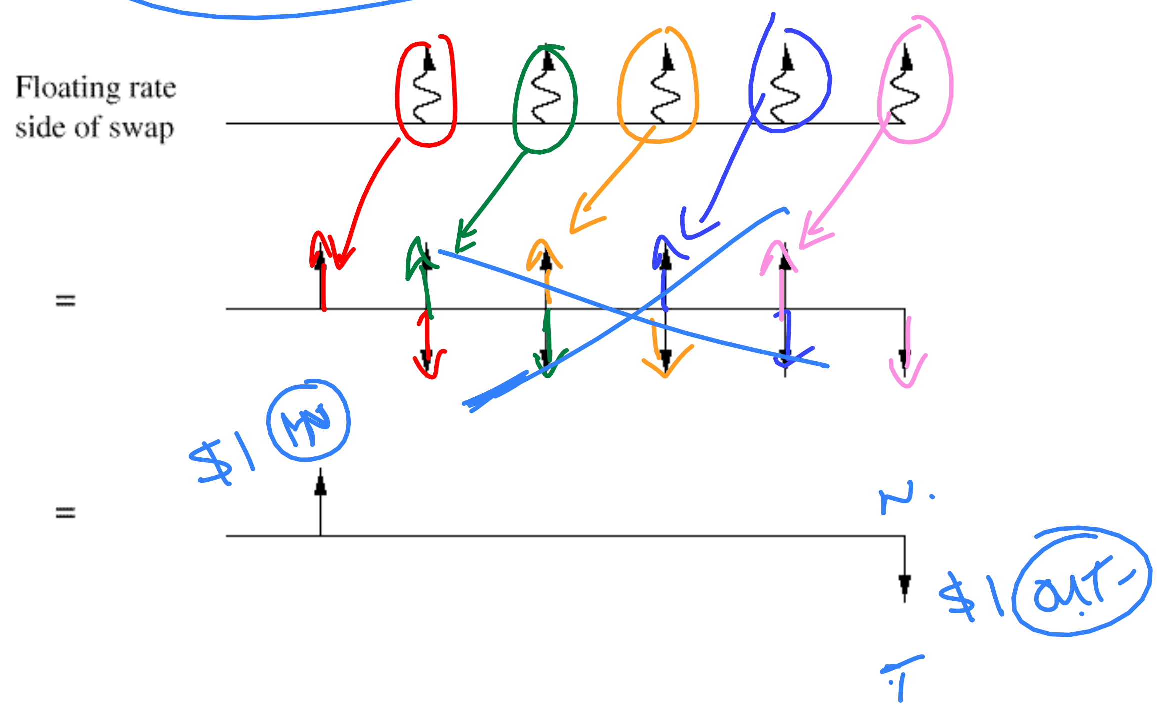
PV(floating).

A single floating leg of a swap at time T_i is *exactly* equal to a deposit of \$1 at time $T_i - \tau$ and a withdrawal of \$1 at time *Ti*.

Add up all the floating legs, note the cancelation of all cashflows except for the first and last. The floating side has value

$$1 - Z(t; T_N).$$

PV :



The floating legs are equivalent to two zero-coupon bonds.

Bring the fixed and floating sides together to find that the value of the swap, to the receiver of the fixed side, is

$$-1 + Z(t;T_N) + r_s\tau \sum_{i=1}^{N} Z(t;T_i).$$

- This result is *model independent*. This relationship is independent of any mathematical model for bonds or swaps.

The swap curve

- When the swap is first entered into it is usual for the deal to have no value to either party (i.e. zero value). This is done by a careful choice of the fixed rate of interest r_s.

  Such a swap is called a **par swap**.

Thus

Make sure you can do this

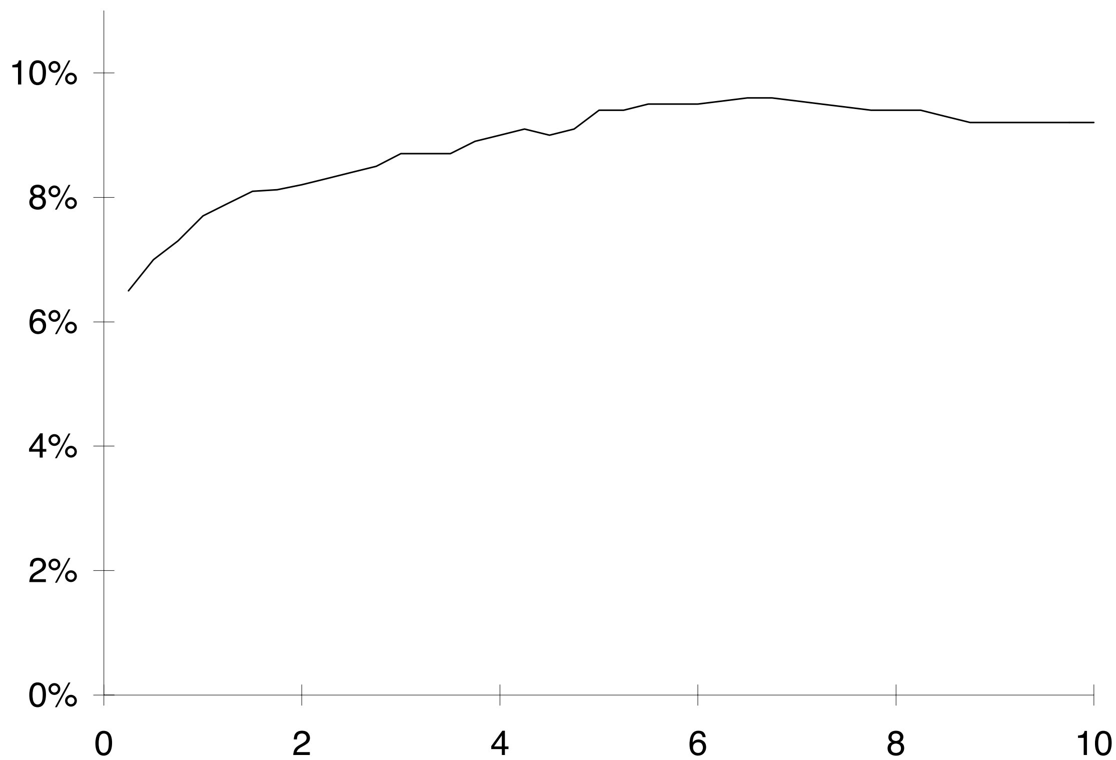
$$r_s = \frac{1 - Z(t;T_N)}{\tau \sum_{i=1}^{N} Z(t;T_i)}. \tag{1}$$

This is the quoted swap rate.

In other words, the present value, let us say, of the fixed side and the floating side both have the same value, netting out to zero.

The rates of interest in the fixed leg of a swap are quoted at various maturities. These rates make up the **swap curve**.

Swap fixed leg



Maturity

Swaps are now so liquid and exist for an enormous range of maturities that their prices determine the yield curve and not *vice versa*.

In practice one is given $r_s(T_i)$ for many maturities T_i and one uses (1) to calculate the prices of zero-coupon bonds and thus the yield curve.

Other features of swaps contracts

The above is a description of the vanilla interest rate swap. There any many features that can be added to the contract that make it more complicated, and most importantly, model dependent.

option to call in fixed (e.g. to close a pay fixed swap)

Callable and puttable swaps

A **callable** or **puttable swap** allows one side or the other to close out the swap at some time before its natural maturity. If you are receiving fixed and the floating rate rises more than you had expected you would want to close the position.

Mathematically we are in the early exercise world of American-style options.

right to pay out fixed leg

Extendible swaps

The holder of an **extendible swap** can extend the maturity of a vanilla swap at the original swap rate.

Index amortizing rate swaps

The principal in the vanilla swap is constant. In some swaps the principal declines with time according to a prescribed schedule. The index amortizing rate swap is more complicated still with the amortization depending on the level of some index, say LIBOR, at the time of the exchange of payments.

Currency swaps

A **currency swap** is an exchange of interest payments in one currency for payments in another currency. The interest rates can both be fixed, both floating or one of each. As well as the exchange of interest payments there is also an exchange of the principals (in two different currencies) at the beginning of the contract and at the end.

The key point about swaps is that they can be priced in terms of bonds (and vice versa).

Most financial instruments require a model for interest rates, however.

And if the instrument has some convexity in rates, then we need a model for the randomness in rates.

There now follows an overview of interest-rate modeling...

Overview

The different approaches to interest-rate modeling

short rate

$r(t)$

1. Deterministic

one path

Bonds - yields

2. Black '76

applied Black Scholes to $B \Rightarrow dBX$.

$\hookrightarrow$ r constant

3. Stochastic spot rate

77 Vasicek

mean. rates.

$\boxed{dr} = a(\bar{r} - r)\,dt + \_\_\, dr\, dX$

speed of mean reversion

$\sqrt{r}$.

Cox Ingersoll Ross

4. Multi factor

Ho-Lee

calibrated

function

$dr = \_\_\, dt + \_\_\, dX_1 \_\_\, dX_2$.

5. Heath, Jarrow & Morton

$\boxed{\text{HJM}}$

Model curve directly

6. LIBOR Market Model

model key rates + $\boxed{\text{correlation}}$

Deterministic rates

The history of interest-rate modeling begins with deterministic rates, and the ideas of yield to maturity, duration etc. We will see this later in this lecture.

Briefly, one assumes that there is a quantity called the **spot interest rate**, this being the interest paid over a very short period of time. This quantity could be constant in the simplest case (and we often assume this when pricing equity options), or a time-dependent function

- $r(t)$

How do we know what is the function $r(t)$?

This is the subject of **bootstrapping**.

The assumption of determinism is not at all satisfactory for pricing derivatives.

Black '76

In 1976 Fischer Black introduced the idea of treating bonds as underlying assets so as to use the Black–Scholes equity option formulas for fixed-income instruments.

- All you need to know is the volatility of the 'underlying' bond.

This is not entirely satisfactory since there can be contradictions in this approach. On one hand bond prices are random, yet on the other hand interest rates used for discounting from expiration to the present are assumed to be deterministic.

However, the main advantage of this approach is that for simple fixed-income contracts there are simple formulas for their prices.

For more complicated, less liquid instruments an internally consistent stochastic rates approach was needed.

Stochastic spot rate

The first step on the stochastic interest rate path used the very short-term interest rate, the spot rate, as the random factor driving the entire yield curve.

The mathematics of these spot-rate models is identical to that for equity models, and the fixed-income derivatives satisfied similar equations as equity derivatives.

- Diffusion equations governed the prices of derivatives, and derivatives prices could be interpreted as the risk-neutral expected value of the present value of all cashflows as well.

And so the solution methods of finite-difference methods for solving partial differential equations, trees and Monte Carlo simulation carried over.

Models of this type are **Vasicek**, **Cox, Ingersoll & Ross**.

The advantages of these models are

1. they are internally consistent

2. they are easy to solve numerically by many different methods

But there are several aspects to the downside:

1. the spot rate does not exist, it has to be approximated in some way

2. with only one source of randomness the yield curve is very constrained in how it can evolve, essentially parallel shifts

3. the yield curve that is output by the model will not match the market yield curve. (Model is *uncalibrated*)

Multi factor

Models were then designed to get around the second and third of these problems.

- A second random factor was introduced, sometimes representing the long-term interest rate (**Brennan & Schwartz**), and sometimes the volatility of the spot rate (**Fong & Vasicek**). This allowed for a richer structure for yield curves.

- And an arbitrary time-dependent parameter (or sometimes two or three such) was allowed in place of what had hitherto been constant(s). The time dependence allowed for the yield curve (and other desired quantities) to be instantaneously matched. Thus was born the idea of calibration, the **Ho & Lee** model.

Heath, Jarrow & Morton

The business of calibration in such models was rarely straightforward.

The next step in the development of models was by Heath, Jarrow & Morton (HJM) who modeled the evolution of the *entire* yield curve directly so that calibration simply became a matter of specifying an initial curve.

The model is easy to implement via simulation.

Because of the non-Markov nature of the general HJM model it is not possible to solve these via finite-difference solution of partial differential equations, the governing partial differential equation would generally be in an infinite number of variables, representing the infinite memory of the general HJM model.

Since the model is usually solved by simulation it is straightforward having any number of random factors and so a very, very rich structure for the behaviour of the yield curve.

The only downside with this model, as far as implementation is concerned, is that it assumes a continuous distribution of maturities and the existence of a spot rate.

LIBOR Market Model

The LIBOR Market Model (LMM) as proposed by Miltersen, Sandmann, Sondermann, Brace, Gatarek, Musiela and Jamshidian in various combinations and at various times, models *traded* forward rates of different maturities as correlated random walks.

- The key advantage over HJM is that only prices which exist in the market are modelled, the LIBOR rates.

Each traded forward rate is represented by a stochastic differential equation model with a drift rate and a volatility, as well as a correlation with each of the other forward rate models.

And now... deterministic interest rate modeling...

(The other, more interesting, models will be seen in other lectures.)

"short rate" r(t) consistent with bond yields

Excel : spreadsheet.

Come back at half past!

Break!

Measures of yield

There is such a variety of fixed-income products, with different coupon structures, amortization, fixed and/or floating rates, that it is necessary to be able to consistently compare different products.

One way to do this is through measures of how much each contract earns, there are several measures of this all coming under the name **yield**.

bonds' return

Yield to maturity (YTM) or internal rate of return (IRR)

Suppose that we have a zero-coupon bond maturing at time T when it pays one dollar.

- At time t it has a value $Z(t;T)$.

Compare the following two investments:

simple strategy

1. Buy one zero-coupon bond that matures at time T. So you spend $Z(t;T)$ now to get \$1 at T.

2. Put an amount (cash) of $Z(t;T)$ in a bank account earning a $\boxed{\text{fixed,}}$ continously compounded, interest rate of y. So you spend $Z(t;T)$ now to get $Z(t;T)e^{y(T-t)}$ at T.

Question: What y makes the two amounts at time T the same?

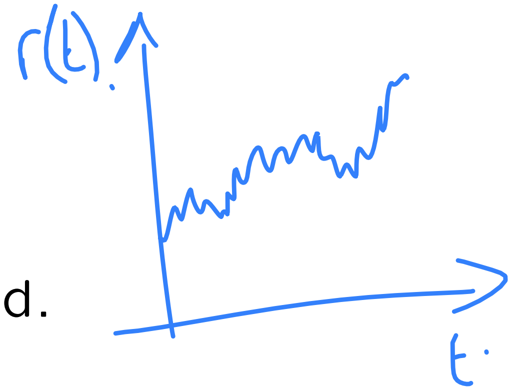
$$Z(t;T)e^{y(T-t)} = 1.$$

$$Z(t;T) = e^{-y(T-t)}.$$

It follows that

$$y = -\frac{\log Z}{T-t}.$$

This is the simple relationship between the value of a bond with a single cashflow and a continuously compounded rate of interest.



Let us generalize this.

Suppose that we have a coupon-bearing bond.

- Discount all coupons and the principal to the present by using some interest rate y. *And we use the same rate of interest for present valuing each cashflow.*

constant yield earned by all CFs of bond.

single "y" which makes PV(CF) = Price

If P is the principal, N the number of coupons, C_i the coupon paid on date t_i then the present value of the bond, at time t, is

Price

$$V = Pe^{-y(T-t)} + \sum_{i=1}^{N} C_i e^{-y(t_i - t)}. \tag{2}$$

Same y for cash flows

If the bond is a traded security then we know the price at which the bond can be bought.

If this is the case then we can calculate the **yield to maturity** or **internal rate of return** as the value y that we must put into Equation (2) to make V equal to the traded price of the bond.

This calculation must be performed by some trial and error/iterative procedure.

Example:

A five-year bond with principal of \$1 has twice-yearly payments of 3 cents.

It has a market value of 96 cents.

What is its yield to maturity?

We ask

- 'What is the rate of return we must use to give these cash flows a total present value of 96 cents?'

This value is the yield to maturity.

We must solve

$$\overset{V}{0.96} = \overset{Pw}{1} \times e^{-5y} + \sum_{i=1}^{10} 0.03 \times e^{-0.5\,\underline{i}\,y}$$

for y.

Time	**Coupon**	**Principal repayment**	**PV (discounting at 6.8406%)**
0			0
0.5	.03		0.0290
1.0	.03		0.0280
1.5	.03		0.0270
2.0	.03		0.0262
2.5	.03		0.0253
3.0	.03		0.0244
3.5	.03		0.0236
4.0	.03		0.0228
4.5	.03		0.0220
5.0	.03	1.00	0.7316
		Total	0.9600

An example of a coupon-bearing bond.

In the fourth column in this table is the present value (PV) of each of the cashflows using a rate of 6.8406%: since the sum of these present values is 96 cents the YTM or IRR is 6.8406%.

This yield to maturity is a valid measure of the return on a bond if we intend to hold it to maturity.

	A	B	C	D	E	F	G	H	I	J	K
1					Date	Coupon	Principal	PVs	Time	Time^2	
2					0				wtd	wtd	
3		YTM	4.95%		0.5	2%		0.0195	0.0098	0.0049	
4		Mkt price	0.921		1	2%		0.0190	0.0190	0.0190	
5		Th. Price	0.921		1.5	2%		0.0186	0.0279	0.0418	
6		Error	1.4E-08		2	2%		0.0181	0.0362	0.0725	
7		Duration	8.2544		2.5	2%		0.0177	0.0442	0.1104	
8		Convexity	76.8728		3	2%		0.0172	0.0517	0.1551	
9					3.5	2%		0.0168	0.0589	0.2060	
10					4	2%		0.0164	0.0656	0.2625	
11					4.5	2%		0.0160	0.0720	0.3241	
12					5	2%		0.0156	0.0781	0.3903	
13					5.5	2%		0.0152	0.0838	0.4607	
14					6	2%		0.0149	0.0892	0.5349	
15					6.5	2%		0.0145	0.0942	0.6124	
16					7	2%		0.0141	0.0990	0.6929	
17					7.5	2%		0.0138	0.1035	0.7760	
18					8	2%		0.0135	0.1077	0.8613	
19					8.5	2%		0.0131	0.1116	0.9485	
20					9	2%		0.0128	0.1153	1.0374	
21					9.5	2%		0.0125	0.1187	1.1276	
22					10	2%	1	0.6216	6.2161	62.1614	
23											
24											
25											
26											
27											
28											
29											
30											
31											
32											
33											
34											
35											
36											
37											

= C4-C5

= SUM(H3:H22)

= SUM(I3:I22)/C5

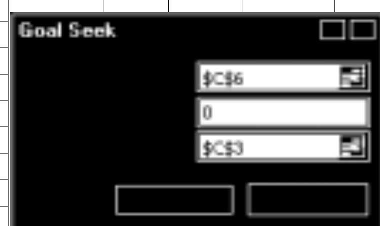
= SUM(J3:J22)/C5

= F20*EXP(-E20*C3)

= (G22+F22)*EXP(-E22*C3)

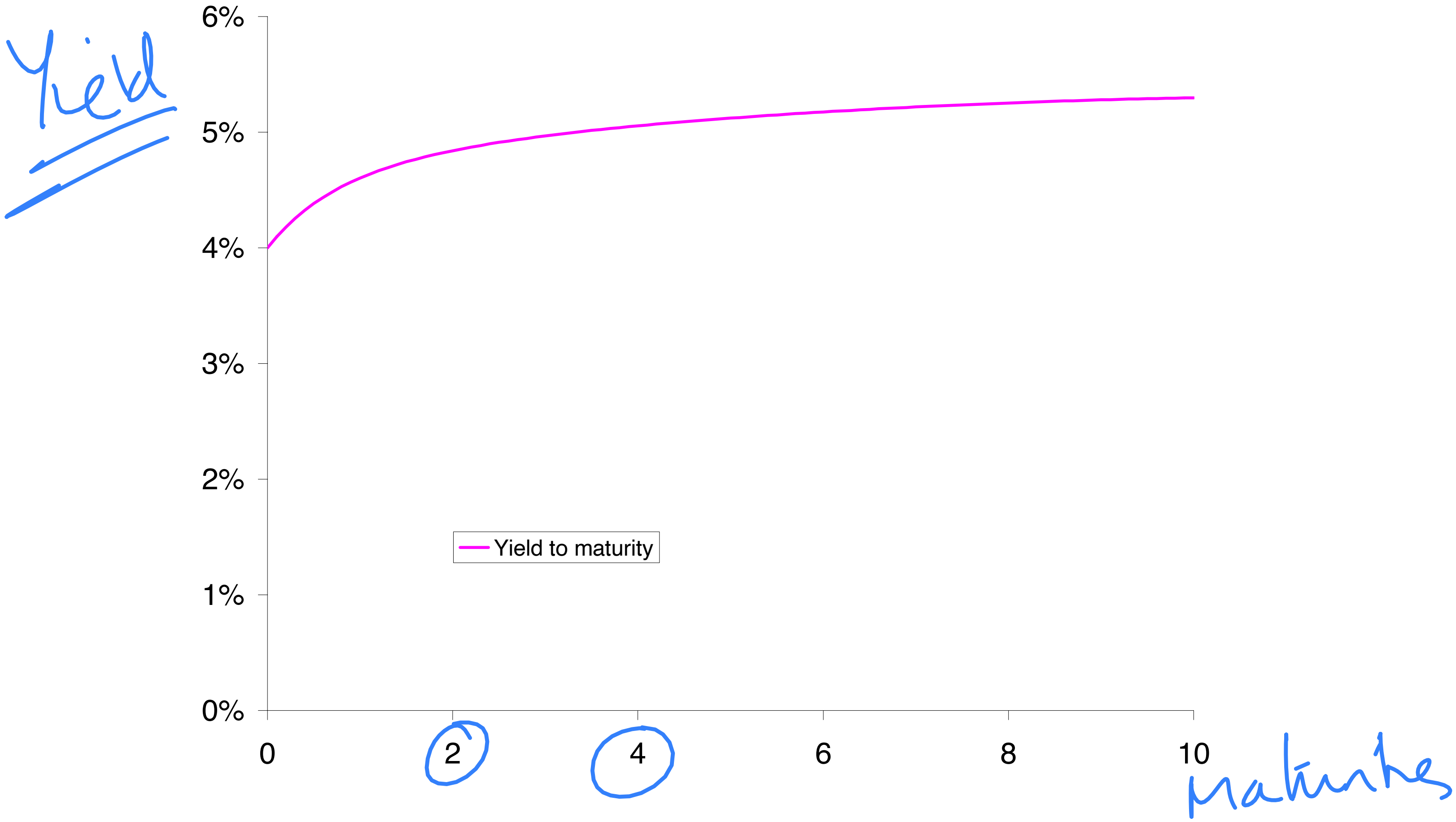
= E20*H20

= E20*I20



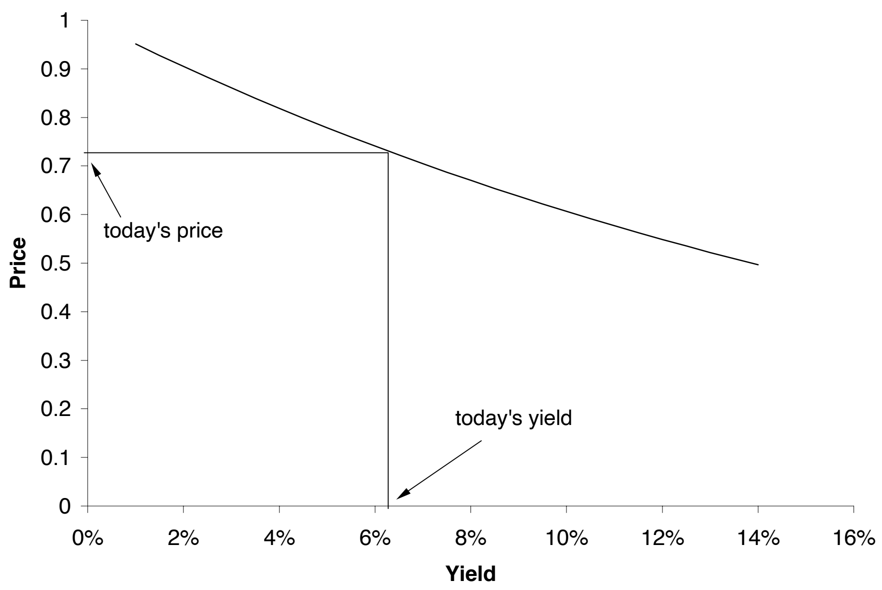
Every bond has an associated yield to maturity.

We can therefore plot yield versus maturity, the so called **yield curve**.



Price/yield relationship

From Equation (2) we can easily see that the relationship between the price of a bond and its yield is of the form shown below (assuming that all cash flows are positive).



Since we are often interested in the sensitivity of instruments to the movement of certain underlying factors it is natural to ask how does the price of a bond vary with the yield, or vice versa.

To a first approximation this variation can be quantified by a measure called the **duration**.

estimate it based on

$$\frac{dV}{dy}$$

measure of price risk.

$$\boxed{\%\Delta V \text{ for } 1\% \Delta y.}$$

Duration

$\frac{dV}{dy}$ dimension = $\frac{\$}{\text{rates}}$ ← 1/time = \$time

From Equation (2) we find that

$$\frac{dV}{dy} = -(T-t)Pe^{-y(T-t)} - \sum_{i=1}^{N} C_i(t_i - t)e^{-y(t_i-t)}.$$

This is the slope of the Price/Yield curve.

%ΔP for 1% Δy.

The quantity

%ΔV.

- $$-\frac{1}{V}\frac{dV}{dy}$$

\$

is called the **Macauley duration**.

Time

Average time! – w

In the expression for the duration the time of each coupon payment is weighted by its present value.

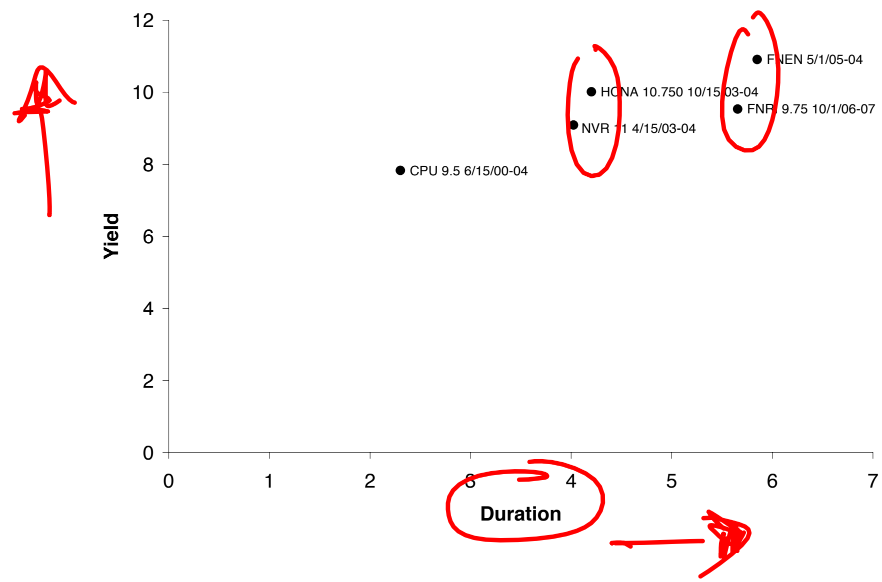
The higher the value of the present value of the coupon the more it contributes to the duration.

Also, since y is measured in units of inverse time, the units of the duration are time.

- The duration is a measure of the **average life of the bond**.

One of the most common uses of the duration is in plots of yield *versus* duration for a variety of instruments.

- We can group together instruments with the same or similar durations and make comparisons between their yields.



Convexity

→ based on 2nd deriv

The Taylor series expansion of V gives

$$dV = \frac{dV}{dy}\delta y + \frac{1}{2}\frac{d^2V}{dy^2}(\delta y)^2 + \cdots,$$

where δy is a change in yield.

For very small movements in the yield, the change in the price of a bond can be measured by the duration.

For larger movements we must take account of the curvature in the Price/Yield relationship.

The **dollar convexity** is defined as

$$\frac{d^2V}{dy^2} = (T-t)^2 P e^{-y(T-t)} + \sum_{i=1}^{N} C_i (t_i - t)^2 e^{-y(t_i - t)}.$$

and the **convexity** is

%ΔV

$\times \frac{1}{V}$.

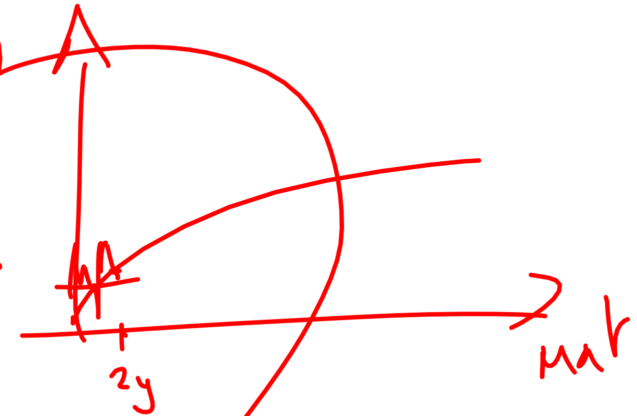
$$\frac{1}{V}\frac{d^2V}{dy^2}.$$

Average time squared

A problem. . .

Every bond has its own yield.

Bond prices



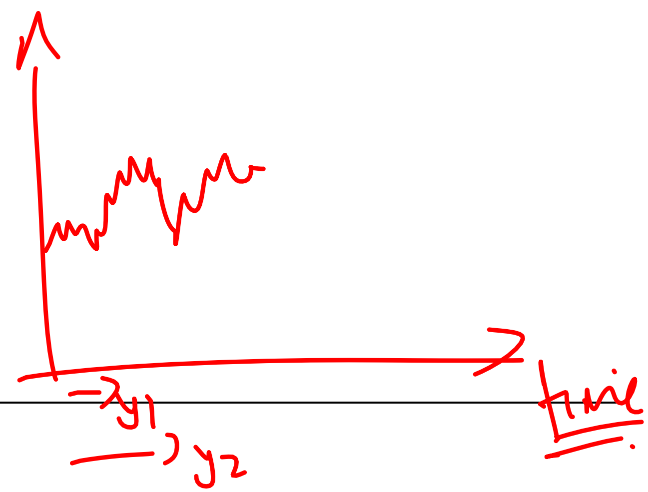
Is it possible to construct an interest rate model that is simultaneously consistent with *all* bonds?

Yes.

bootstrapping

r(t)

- The simplest way to do this is to introduce one time-dependent interest rate, representing the interest received over an infinitesimal period of time.



Time-dependent interest rate

Let's look at bond pricing when we have an interest rate that is a **known function of time**. The interest rate we consider will be what is called the **short-term interest rate** or **spot interest rate** $r(t)$.

- This means that the rate $r(t)$ is to apply at time t: interest is compounded at this rate at each moment in time.

We begin with a zero-coupon bond example.

The bond price is a function of time: $Z = Z(t;T)$, with maturity T being a parameter.

Because we receive \$1 at time $t = T$ we know that $Z(T;T) = 1$.

Let's derive an equation for the value of the bond at a time before maturity, $t < T$.

The change in the value of the bond in a time-step dt (from t to $t + dt$) is

$$dZ.$$

Arbitrage considerations—there is no risk or randomness in this model—again lead us to equate this with the return from a bank deposit receiving interest at a rate $r(t)$:

$$dZ = r(t)\ Z\ dt.$$

The solution of this equation is

- $$Z(t;T) = e^{-\int_t^T r(s)ds}.$$

So, given an $r(t)$ we can find prices of zero-coupon bonds.

Unfortunately, there is no one to tell us what $r(t)$ is!

Instead we are usually given the market prices of zero-coupon bonds. The question still remains, can we find *one* function $r(t)$ that is consistent with all bond prices? ...

$r(t)$

Forward rates and bootstrapping

OUT OF BONDS

Let us suppose that we are in a perfect world in which we have a continuous distribution of zero-coupon bonds with all maturities T. Call the prices of these at time t, $Z(t;T)$.

The **forward rate** is the curve of a time-dependent spot interest rate that is consistent with the market price of instruments.

If this rate is $r(t)$ at time t then it satisfies

$$\boxed{Z(t;T) = e^{-\int_t^T r(\tau)d\tau}.}$$

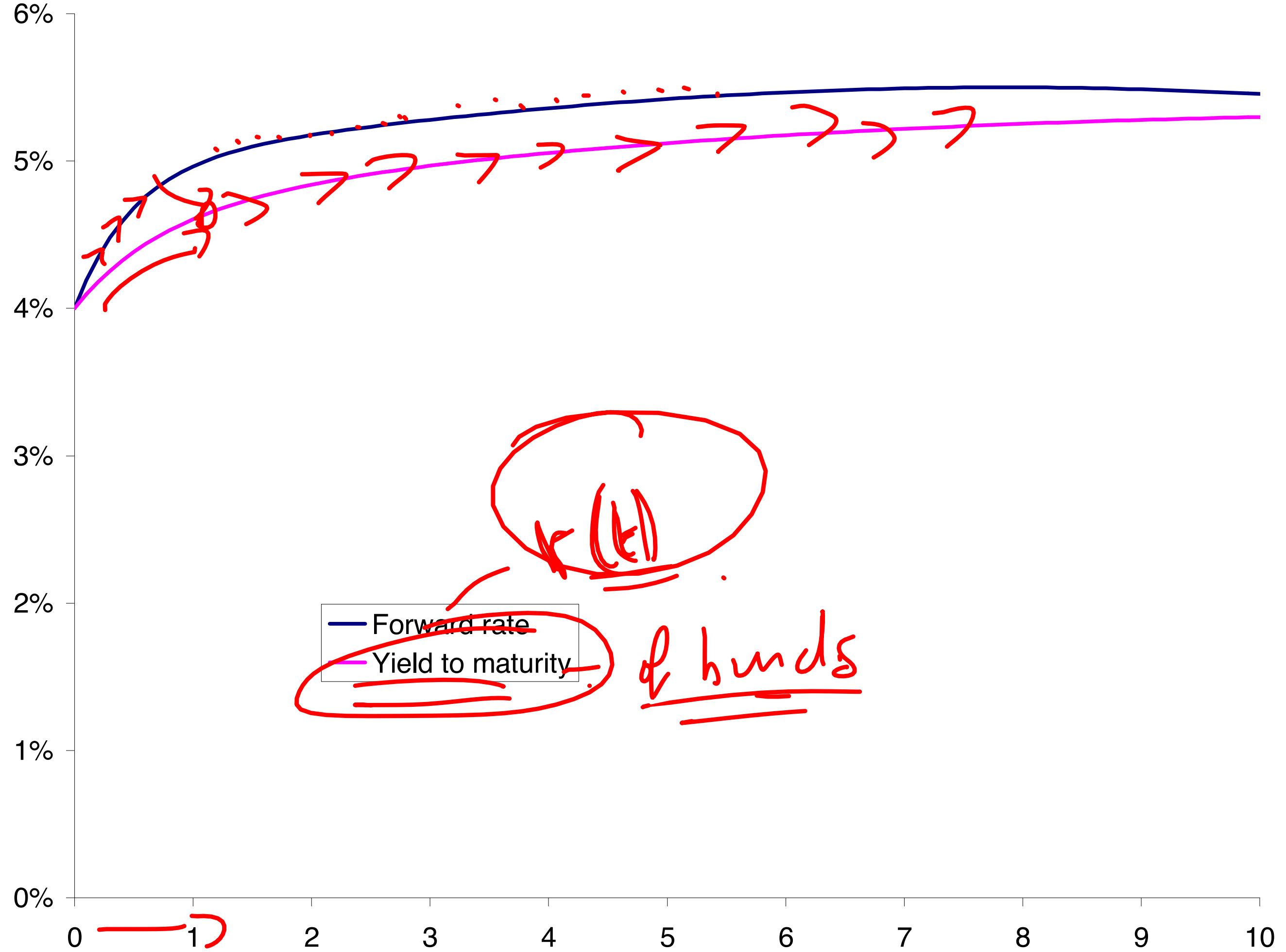
On rearranging and differentiating with respect to T this gives

$$r(T) = -\frac{\partial}{\partial T}(\log Z(t;T)).$$

Of course, this should really be quoted as a function of calendar time t, so we just change variables and write

$$r(t) = -\frac{\partial}{\partial T}(\log Z(t;T))|_{T=t}$$

- Forward rates are interest rates that are assumed to apply over given periods *in the future* for *all* instruments. This contrasts with yields which are assumed to apply up to maturity, with a different yield for each bond.



6%
5%
4%
3%
2%
1%
0%
0
1
2
3
4
5
6
7
8
9
10
Forward rate
Yield to maturity
r(t)
of bonds

Joining the dots

In the less-than-perfect real world we must do with only a discrete set of data points.

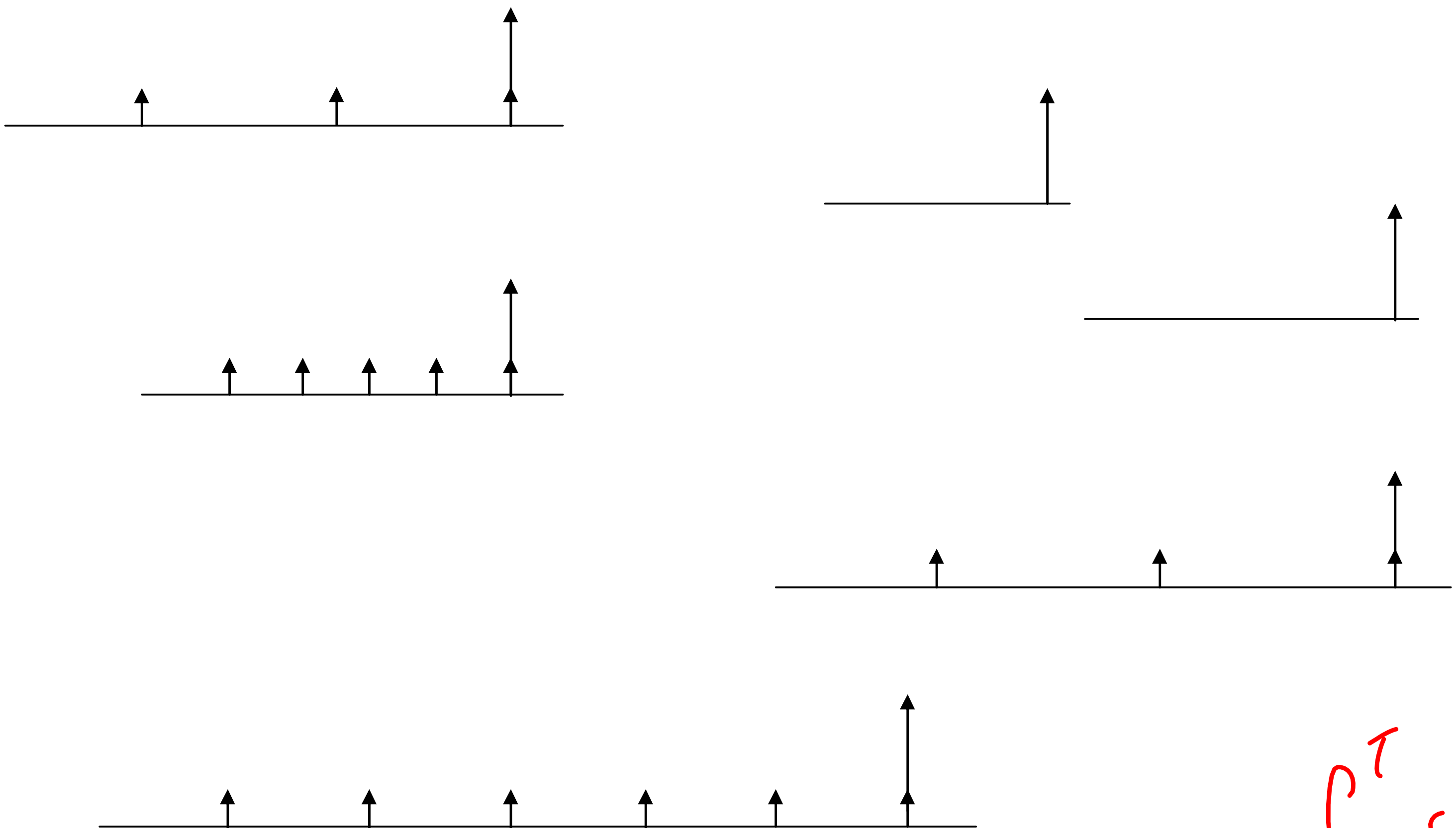
We continue to assume that we have zero-coupon bonds but now we will only have a discrete set of them.

We can still find an implied forward rate curve as follows.

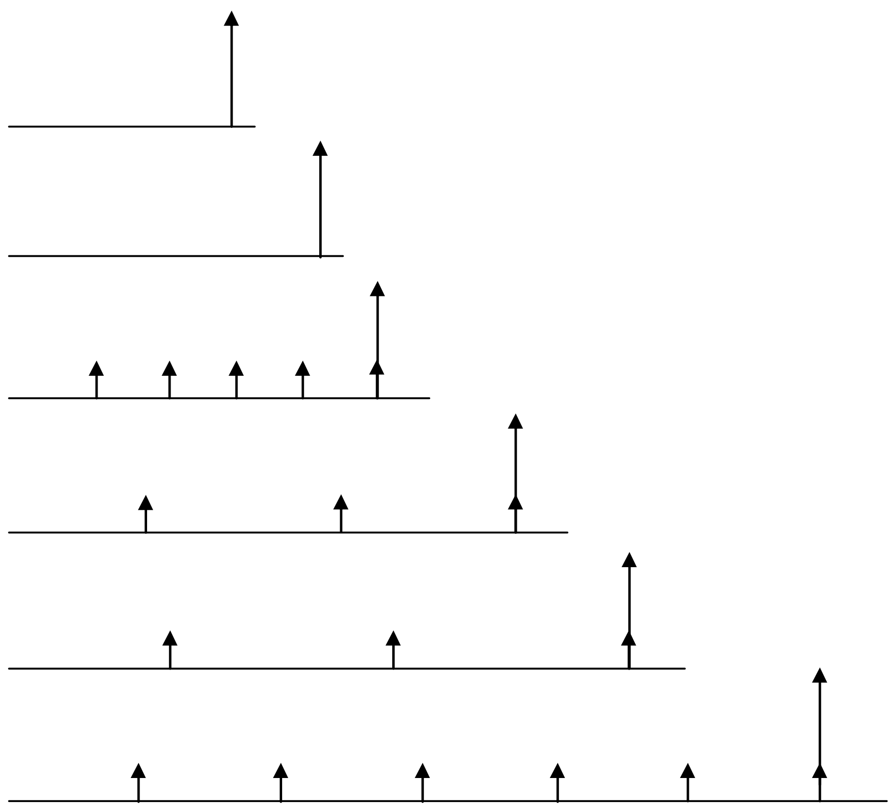
- Rank the bonds according to maturity, with the shortest maturity first. The market prices of the bonds will be denoted by Z_i^M where i is the position of the bond in the ranking.

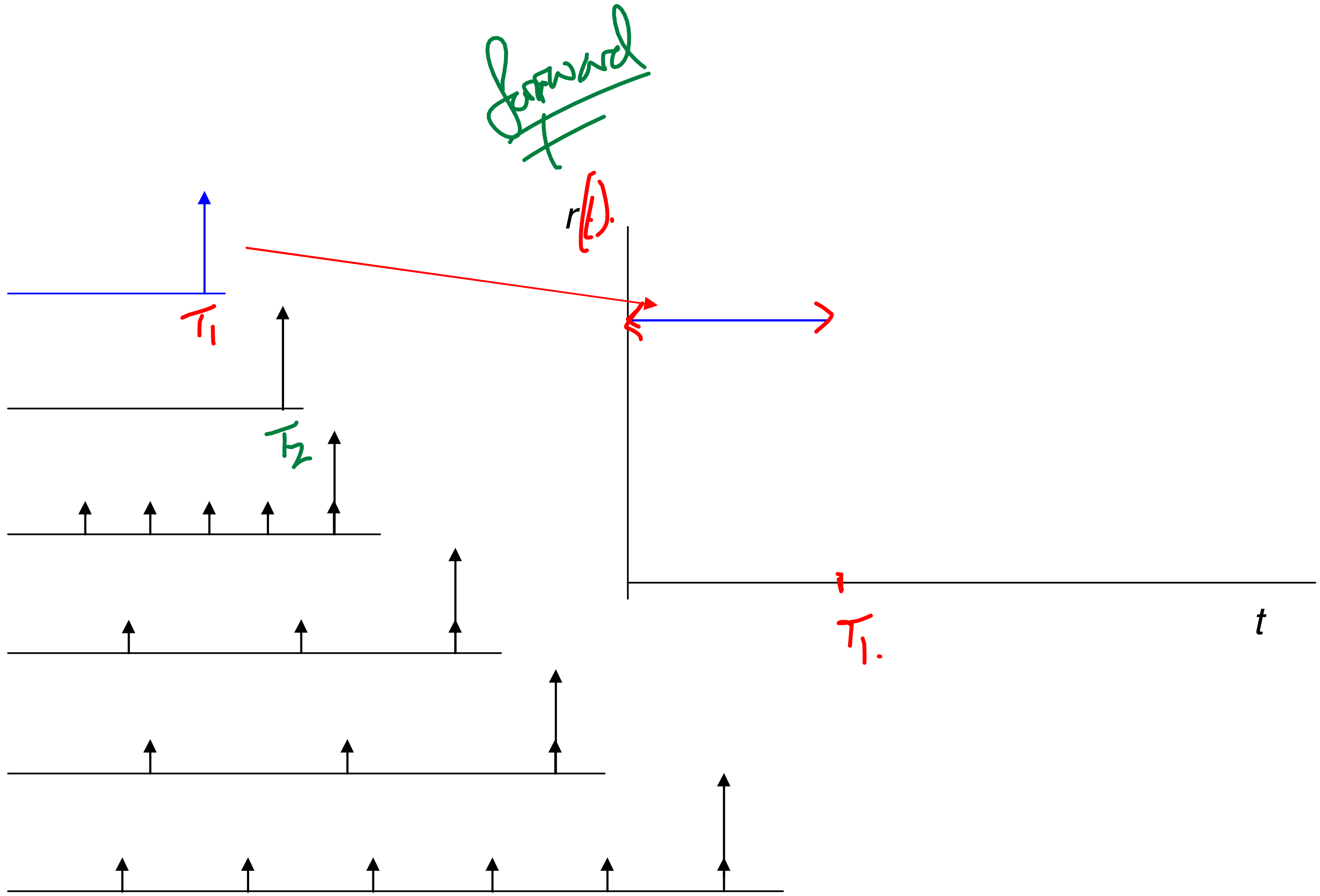
First, the idea . . .

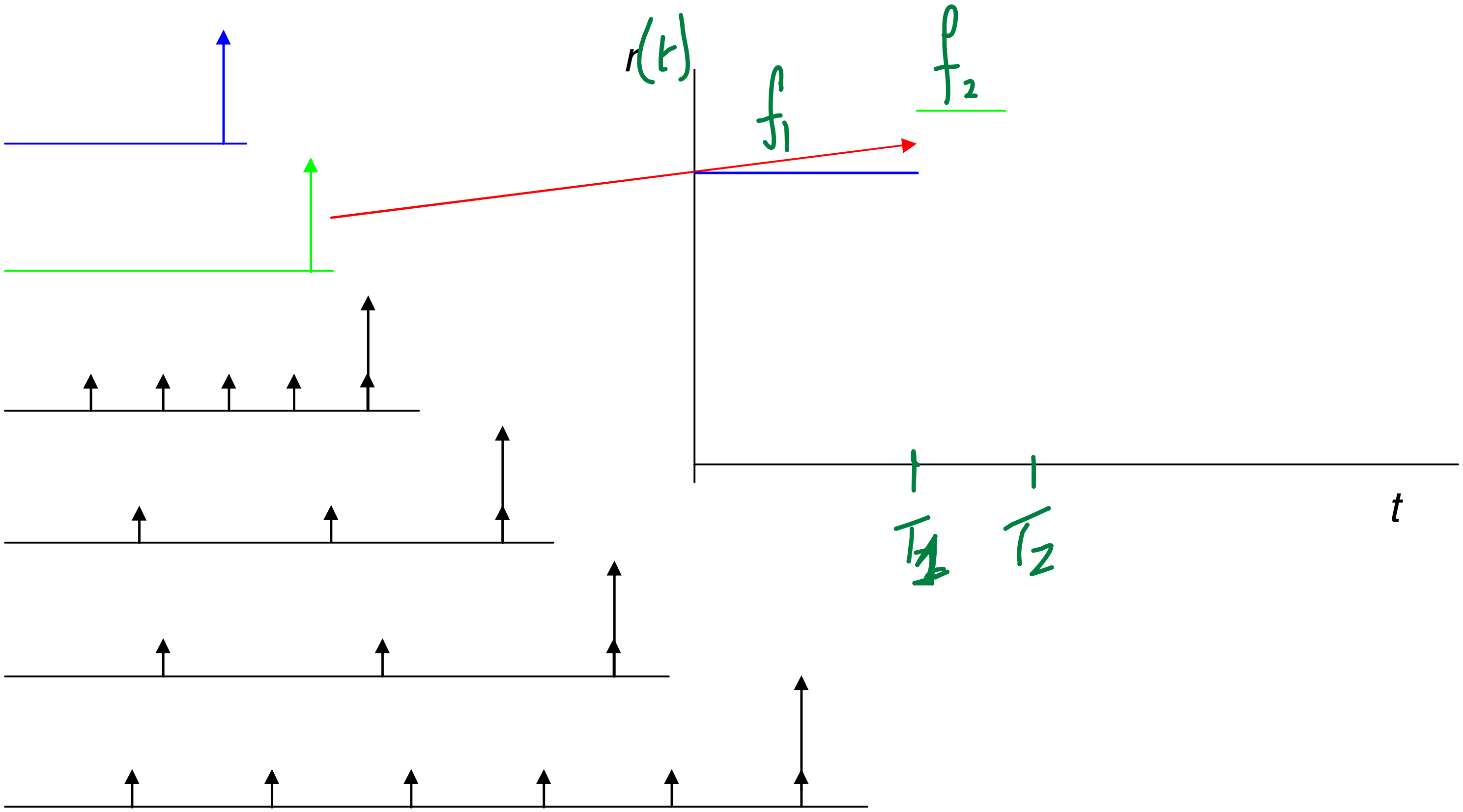
BONDS

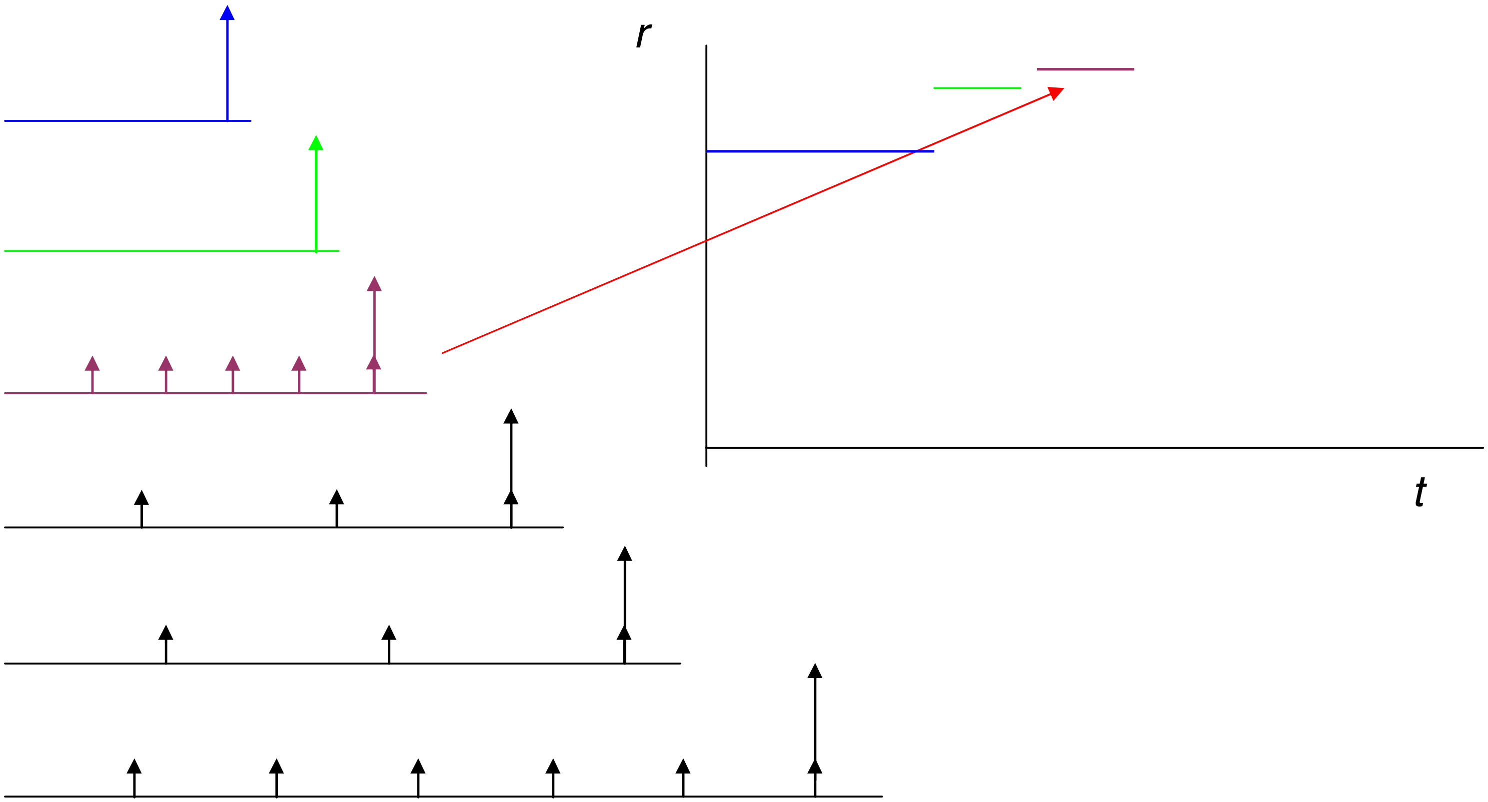


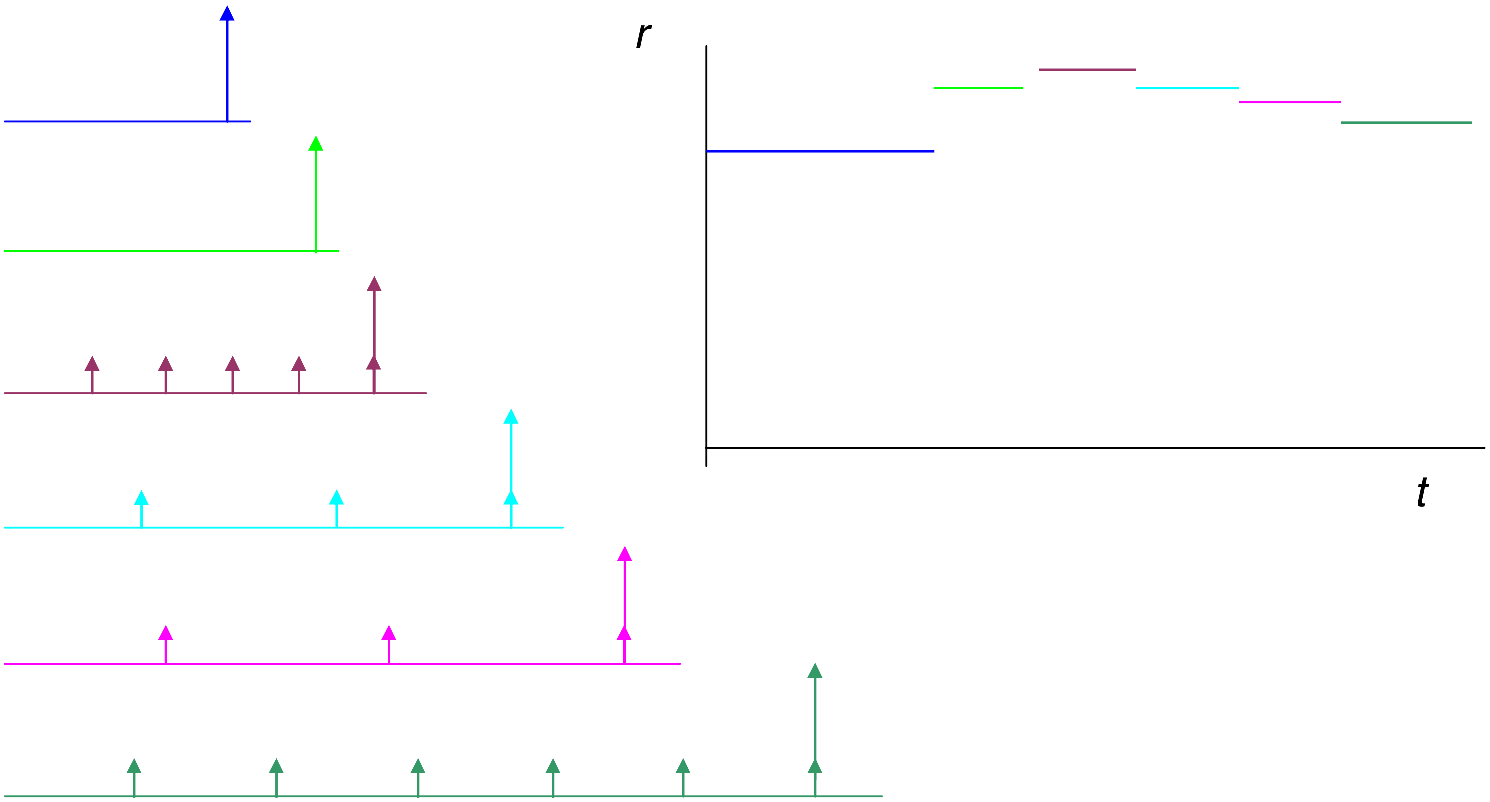
$$Z(t,T) = e^{-\int_t^T r(s)\,ds}$$




forward
$r(t)$
T_1
T_2
t
T_1.

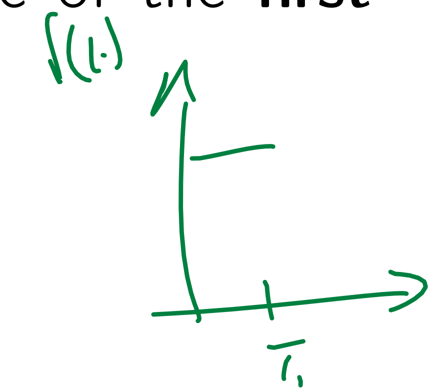

r(t)
f1
f2
t
T1
T2


r
t


r
t

Then the maths . . .

What interest rate is implied by the market price of the **first bond**? The answer is y_1, the solution of



$$Z_1^M = e^{-y_1(T_1-t)},$$

i.e.

$$y_1 = -\frac{\log(Z_1^M)}{T_1 - t}.$$

This rate will be the rate that we use for discounting between the present and the maturity date T_1 of the first bond.

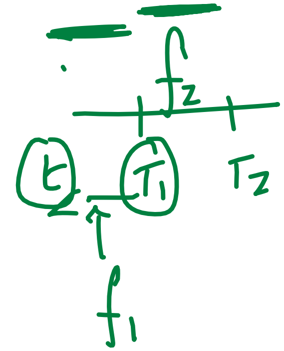
- It will be applied to *all* instruments whenever we discount over this period.

Z_1

Now move on to the **second bond**, having maturity date T_2.

We know the rate to apply between now and time T_1, but at what interest rate must we discount between dates T_1 and T_2 to match the theoretical and market prices of the second bond?

The answer is y_2 which solves the equation



$$Z_2^M = e^{-y_1(T_1-t)} e^{-y_2(T_2-T_1)},$$

i.e.

$$y_2 = -\frac{\log\left(Z_2^M / Z_1^M\right)}{T_2 - T_1}. \quad = \frac{y_2(T_2-t) - y_1(T_1-t)}{T_2-T_1}$$

- By this method of **bootstrapping** we build up the forward curve. Note how the forward rates are applied between two dates, for which period we have assumed they are constant.

$$= \frac{R_2 T_2 - R_1 T_1}{T_2 - T_1}$$

	A	B	C	D	E
1	**Time to**	**Market**	**Yield to**	**Forward**	
2	**maturity**	**price z-c b**	**maturity**	**rate**	
3	0.25	0.9809	7.71%	7.71%	
4	0.5	0.9612	7.91%	8.12%	
5	1	0.9194	8.40%	8.89%	
6	2	0.8436	8.50%	8.60%	
7	3	0.7772	8.40%	8.20%	
8	5	0.644	8.80%	9.40%	
9	7	0.5288	9.10%	9.85%	
10	10	0.3985	9.20%	9.43%	
11					
12					
13					
14					
15					
16					

= -LN(B10)/A10

= (C10*A10-C9*A9)/(A10-A9)

Interpolation

We have explicitly assumed in the previous section that the forward rates are piecewise constant, jumping from one value to the next across the maturity of each bond. Other methods of **interpolation** are also possible.

For example, the forward rate curve could be made continuous, with piecewise constant gradient. Some people like to use cubic splines.

Whatever interpolation method you use you would expect the resulting curve to have certain nice properties (such as being non negative, perhaps with continuity and smoothness).

STUARTJACKAMAN@GMAIL.COM

Summary

Please take away the following important ideas

- A vanilla swap can be decomposed exactly into a portfolio of bonds
- The main ideas behind interest-rate modeling . . . . .
- Yield, duration and convexity and important measures of interest rate and sensitivities
- The forward curve can be constructed from simple bonds and swaps

Bootstrapping

Greetings

Stochastic Interest Rate Modeling

Spot rate $r(t)$

r_t $\boxed{dt}$

In this lecture...

- stochastic models for interest rates
- how to derive the pricing equation for many fixed-income products

(*) Popular class of bond prices

MPoR

- the structure of many popular one-factor interest rate models
- the theoretical framework for multi-factor interest rate modeling

2 factor I.R model → B.P.E

- popular two-factor models

Certificate in Quantitative Finance

By the end of this lecture you will

- be able to derive the pricing equation for fixed-income instruments with one and two random factors / sources of randomness

- appreciate the meaning of the market price of interest rate risk

- know the names of many popular interest rate models

- Extend to 2 factors

Introduction

Whenever you price non linear derivatives products you require a model for the randomness in the underlying.

This is as true for fixed-income instruments as it is for equity derivatives.

In this lecture we see the ideas behind modeling interest rates using stochastic differential equations to model randomness.

We begin by having one source of randomness, the spot interest rate.

This is the subject of **one-factor interest rate modeling**.

$Z(t;T)$

- The model will allow the short-term interest rate, the spot rate, to follow a random walk.

SDE.

This model leads to a parabolic partial differential equation for the prices of bonds and other interest rate derivative products.

B.S.E

V – bond $V(r, t; T)$

var. time Maturity

Later we will consider modeling the fixed-income world using *two* sources of randomness.

The spot interest rate

The ‘spot rate’ that we will be modeling is a very loosely-defined quantity, meant to represent the yield on a bond of infinitesimal maturity.

In practice one should take this rate to be the yield on a liquid finite-maturity bond, say one of one month.

Bonds with one *day* to expiry do exist but their price is not necessarily a guide to other short-term rates.

Stochastic interest rates

Since we cannot realistically forecast the future course of an interest rate, it is natural to model it as a random variable.

- We are going to model the behavior of r, the interest rate received by the shortest possible deposit.

From this we will see the development of a model for all other rates.

The interest rate for the shortest possible deposit is commonly called the **spot interest rate**.

$t \to t + dt$

Let us suppose that the interest rate r is governed by a stochastic differential equation of the form

real world drift

- $$\boxed{dr = u(r,t)\,dt + w(r,t)\,dX.}$$

diffusion

The functional forms of $u(r,t)$ and $w(r,t)$ determine the behavior of the spot rate r.

For the present we will not specify any particular choices for these functions.

The pricing equation for the general model

When interest rates are stochastic a fixed-income instrument has a price of the form $V(r,t)$. $\longrightarrow V(r,t;T)$

- We are not modeling a *traded* asset; the traded asset (the bond, say) is a derivative of our independent variable r.

Pricing a fixed-income instrument presents new technical problems, and is in a sense harder than pricing an option since *there is no underlying asset with which to hedge*.

$\mathbb{P}$ $\quad \mu \longrightarrow r \quad$ $\mathbb{Q}$

- The only way to construct a hedged portfolio is by hedging one bond with a bond of a different maturity.

Spot rate $r(t)$ or r_t – non traded.

With equities $\boxed{\Pi = V - \Delta S}$ ← traded.

1 stock (packet of 100/1000 shares)

Hedging one bond with another

We set up a portfolio containing two bonds with different maturities T_1 and T_2.

The bond with maturity T_1 has price $V_1(r, t; T_1)$ and the bond with maturity T_2 has price $V_2(r, t; T_2)$.

We hold one of the former and a number $-\Delta$ of the latter.

We have

instead of S

$$\Pi = V_1 - \Delta V_2.$$

Δ is for hedging purposes to eliminate risk

Δ-hedge to make Π risk-free

$t \longrightarrow t + dt$

The change in this portfolio in a time dt is given by

$$d\Pi = \frac{\partial V_1}{\partial t}dt + \boxed{\frac{\partial V_1}{\partial r}} dr + \frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2}dt - \Delta\left(\frac{\partial V_2}{\partial t}dt + \boxed{\frac{\partial V_2}{\partial r}} dr + \frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}dt\right),$$

where we have applied Itô's lemma to functions of r and t.

Which of these terms are random?

$$\frac{\partial V_1}{\partial r} - \Delta \frac{\partial V_2}{\partial r} = 0$$

$$\Longrightarrow \Delta$$

Once you've identified them you'll see that the choice

$$\boxed{\Delta = \frac{\partial V_1}{\partial r} \Big/ \frac{\partial V_2}{\partial r}}$$

eliminates all randomness in $d\Pi$. This is because it makes the coefficient of dr zero.

Substituting Δ with $\frac{\partial V_1}{\partial r} \Big/ \frac{\partial V_2}{\partial r}$

We then have

$$d\Pi = \left(\frac{\partial V_1}{\partial t} + \frac{1}{2} w^2 \frac{\partial^2 V_1}{\partial r^2} - \underbrace{\left(\frac{\partial V_1}{\partial r} \Big/ \frac{\partial V_2}{\partial r} \right)}_{\Delta} \left(\frac{\partial V_2}{\partial t} + \frac{1}{2} w^2 \frac{\partial^2 V_2}{\partial r^2} \right) \right) dt$$

$$= r\Pi\, dt = r \left(V_1 - \overbrace{\left(\frac{\partial V_1}{\partial r} \Big/ \frac{\partial V_2}{\partial r} \right)}^{\Delta} V_2 \right) dt,$$

where we have used arbitrage arguments to set the return on the portfolio equal to the risk-free rate. This risk-free rate is just the spot rate.

Collecting all V_1 terms on the left-hand side and all V_2 terms on the right-hand side we find that

$$\frac{\frac{\partial V_1}{\partial t}+\frac{1}{2}w^2\frac{\partial^2 V_1}{\partial r^2}-rV_1}{\frac{\partial V_1}{\partial r}}=\frac{\frac{\partial V_2}{\partial t}+\frac{1}{2}w^2\frac{\partial^2 V_2}{\partial r^2}-rV_2}{\frac{\partial V_2}{\partial r}}.$$

Indep. of T_2 Indep of T_1

both depend on (r,t)

At this point the distinction between the equity and interest-rate worlds starts to become apparent.

- This is *one* equation in *two* unknowns.

Fortunately, the left-hand side is a function of T_1 but not T_2 and the right-hand side is a function of T_2 but not T_1.

The only way for this to be possible is for both sides to be independent of the maturity date.

Furthermore, neither side can have any dependence of the specific contract at all.

Both sides must be equal to a universal constant.

The previous statement is not strictly true. . .

Both sides can be functions of the 'variables' in the model, r and t, since these are common to all instruments.

Dropping the subscript from V, we have

$$\boxed{\frac{\frac{\partial V}{\partial t} + \frac{1}{2} w^2 \frac{\partial^2 V}{\partial r^2} - rV}{\frac{\partial V}{\partial r}} = a(r,t)}$$

indep of T_i

for some function $a(r,t)$.

We shall find it convenient to write

w.l.o.g

$$\boxed{a(r,t) = w(r,t)\lambda(r,t) - u(r,t);}$$

for a given $u(r,t)$ and non-zero $w(r,t)$ this is always possible.

The fixed-income pricing equation is therefore

B.P.E

new drift.

$$\boxed{\frac{\partial V}{\partial t}} + \frac{1}{2}w^2 \boxed{\frac{\partial^2 V}{\partial r^2}} + (u - \lambda w)\frac{\partial V}{\partial r} - rV = 0. \qquad (1)$$

Bond Pricing Eqn

Rearrange 1 so that

$$\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + u\frac{\partial V}{\partial r} = \boxed{\lambda w\frac{\partial V}{\partial r} + rV}$$

subst in

∂V

This is another parabolic partial differential equation.

Structurally, it is very similar to the Black–Scholes equation.

The only differences between this equation and the Black–Scholes equation are

- in the coefficients of the gamma and delta terms

- that the variable has another name, r instead of S

As with equity derivatives, we must tell the equation which contract we are solving by specifying a final condition, corresponding to the payoff at maturity, T.

For example, the final condition for a zero-coupon bond is

Normalised principal

$$\boxed{V(r,T)=1.}$$

$$\boxed{V(r,T;T)=1}$$

Redemption value

The market price of risk? $\lambda(r,t)$

Imagine that you hold an unhedged position in one bond with maturity date T. $V(r,t;T)$ do Itô on this

In a time step dt this bond changes in value by $t \to t+dt$

$$dV = \underbrace{w\frac{\partial V}{\partial r}}_{\text{diff}} dX + \underbrace{\left(\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + u\frac{\partial V}{\partial r}\right)}_{\text{drift.}} dt.$$

$$dV = \left[\lambda w \frac{\partial V}{\partial r} + rV\right] dt + w\frac{\partial V}{\partial r}\, dX$$

rVdt

This may be written as

u drift

w diff.

$$dV = w\frac{\partial V}{\partial r}dX + \left(w\lambda\frac{\partial V}{\partial r} + rV\right)dt,$$

t → t+dt

X → X+dX

or

risk

return on unhedged bond

$$dV - \underbrace{rV\,dt}_{\text{risk-free return}} = w\frac{\partial V}{\partial r}(dX + \underbrace{\lambda\,dt}).$$

This expression contains a deterministic term in dt and a random term in dX.

The deterministic term may be interpreted as the excess return above the risk-free rate for accepting a certain level of risk.

- In return for taking the risk the portfolio profits by $\lambda\,dt$ per unit of risk, dX. The function λ is called the **market price of risk**.

Interpreting the market price of risk, and risk neutrality

The bond pricing equation contains references to the functions $u - \lambda w$ and w. The former is the coefficient of the first-order derivative with respect to the spot rate, and the latter appears in the coefficient of the diffusive, second-order derivative.

The four terms in the equation represent, in order as written,

- time decay, $\frac{\partial V}{\partial t}$
- diffusion, $\frac{1}{2} w^2 \frac{\partial^2 V}{\partial r^2}$
- drift and $\boxed{(u - \lambda w)} \frac{\partial V}{\partial r}$
- discounting. $-rV$

$$\frac{dS}{S} = \mu dt + \sigma dX \qquad rS\frac{\partial V}{\partial S}$$

$$dr = u dt + w dX \qquad (u - \lambda w)\frac{\partial V}{\partial r}$$

risk-neutral drift.

$$dr = u dt + w dX$$

risk-neutral spot

$$dr = (u - \lambda w) dt + w dX$$

The equation is similar to the backward equation for a probability density function except for the final discounting term.

- As such we can interpret the solution of the bond pricing equation as the expected present value of all cashflows.

This idea should be familiar from equity derivatives.

As with equity options, this expectation is not with respect to the *real* random variable, but instead with respect to the *risk-neutral* variable.

- There is this difference because the drift term in the equation is not the drift of the real spot rate u, but the drift of another rate, called the **risk-neutral spot rate**. This rate has a drift of $u - \lambda w$.

$$dr = (u - \lambda w)\,dt + w\,dX$$

Recall that for equities we have a real drift rate denoted by μ. However, we price as if the asset grows with a rate r, the risk-free rate.

For fixed-income products the real growth of the spot interest rate may be $u(r,t)$ but we price as if it were $u(r,t) - \lambda(r,t)w(r,t)$.

The latter is the **risk-adjusted drift rate**.

When pricing interest rate derivatives it is important to model, and price, using the risk-neutral rate.

The risk-neutral spot rate evolves according to

$$\boxed{dr = (u - \lambda w)dt + w\, dX.}$$

We can use Monte Carlo simulations for pricing fixed-income products, but we ensure that we simulate the risk-neutral spot rate process.

The relationship between prices and expectations

In the equity world we can write

$$V(S,t) = e^{-r(T-t)} E^*[\text{Payoff}],$$

when interest rates are constant.

We can write

variable function of time

$$V(S,t) = e^{-\int_t^T r(\tau)\, d\tau} E^*[\text{Payoff}],$$

when rates are deterministic, $r(t)$,...

When we have a fixed-income product and rates are stochastic the present value term must go inside the expectation...

$* \equiv Q$

$$\boxed{V(S,t) = E^* \left[e^{-\int_t^T r(\tau)\, d\tau} \, \text{Payoff} \right]}$$

But the idea is the same...

The value of a derivative is the risk-neutral expectation of the present value of the payoff.

Rule of Thumb:

We need the new market-price-of-risk term because our modeled variable, r, is not traded.

- If a modeled quantity is traded then the risk-neutral growth rate is r

$\mu \longrightarrow r$

$\mathbb{P} \longrightarrow \mathbb{Q}$

- If a modeled quantity is not traded then the risk-neutral growth rate is

$$\text{real growth} - \text{market price of risk} \times \text{volatility}$$

$$u \quad - \quad \lambda w$$

Tractable models and solutions of the pricing equation

We have built up the pricing equation for an arbitrary model.

That is, we have not specified the risk-neutral drift, $u - \lambda w$, or the volatility, w.

How can we choose these functions to give us a good model?

There are two ways to proceed:

- Choose a model that matches reality as closely as possible

- Choose a model which is easy to solve

Let us examine some choices for the risk-neutral drift and volatility that lead to tractable models, that is, models for which the solution of the pricing equation for zero-coupon bonds can be found analytically.

We will discuss these models and see what properties we like or dislike.

Named models

There are many interest rate models, associated with the names of their inventors.

1 • Vasicek $dr = (\eta - \gamma r)\,dt + \sqrt{\beta}\,dX$

2 • Cox, Ingersoll & Ross $dr = (\eta - \gamma r)\,dt + \sqrt{\beta r}\,dX$

• Ho & Lee $dr = \eta(t)\,dt + c\,dX$

• Hull & White 1 & 2 with time dep parameters

and many more.

Vasicek

The Vasicek model (for the risk-neutral spot rate) takes the form

$$\beta$$

- $$dr = \overbrace{(\eta - \gamma r)}^{u - \lambda w} dt + \beta^{1/2} dX.$$

Thus the pricing equation is $\frac{\partial V}{\partial t} + \frac{1}{2} w^2 \frac{\partial^2 V}{\partial r^2} + (u - \lambda w) \frac{\partial V}{\partial r} - rV = 0$

$$w^2$$

$$\frac{\partial V}{\partial t} + \frac{1}{2}\beta \frac{\partial^2 V}{\partial r^2} + \underbrace{(\eta - \gamma r)}_{u - \lambda w} \frac{\partial V}{\partial r} - rV = 0.$$

BPE

And the final condition for a zero-coupon bond is

$$V(r, T) = 1.$$

$$\boxed{V(r, T; T) = 1}$$

We are very lucky that this can be solved exactly.

And the form of the solution is very simple.

Affine sol^n's $\quad V(r,t;T) = e^{A(t)-rB(t)}.$ $\qquad A = A(t;T)$

$B = B(t;T)$

Let's confirm this by substituting this expression into the partial differential equation

First of all we need some derivatives:

$$\frac{\partial}{\partial t}\left(e^{A(t)-rB(t)}\right) = \left(\dot{A}(t) - r\dot{B}(t)\right)e^{A(t)-rB(t)}, \quad = \left(\dot{A} - r\dot{B}\right) V$$

where $\dot{}$ means d/dt.

$$\frac{\partial}{\partial r}\left(e^{A(t)-rB(t)}\right) = -B(t)e^{A(t)-rB(t)}, \quad = -B\,V$$

and

$$\frac{\partial^2}{\partial r^2}\left(e^{A(t)-rB(t)}\right) = B(t)^2 e^{A(t)-rB(t)}. \quad = B^2\,V$$

Substituting these expressions into the pricing equation for the Vasicek model we get

$$\left(\dot{A}(t) - r\dot{B}(t)\right)\cancel{e^{A(t)-rB(t)}} + \tfrac{1}{2}\beta B(t)^2\cancel{e^{A(t)-rB(t)}}$$
$$-(\eta - \gamma r)B(t)\cancel{e^{A(t)-rB(t)}} - r\cancel{e^{A(t)-rB(t)}} = 0. + 0 \times r$$

There is a common factor of $e^{A(t)-rB(t)}$. Divide by that and what is left is linear in r:

$$\left(\dot{A}(t) + \tfrac{1}{2}\,\beta B(t)^2 - \eta B(t)\right) + r\left(-\dot{B}(t) + \gamma B(t) - 1\right) = 0 + 0 \times r$$

$$\frac{dB}{dt} = \gamma B - 1 \qquad (1)$$

$$\frac{dA}{dt} = \eta B - \frac{1}{2}\beta B^2 \qquad (2)$$

Solve (1) for B. Then sub. B in (2) & solve for A

Both of the expressions in parentheses must be zero.

We have two ordinary differential equations for $A(t)$ and $B(t)$.

In order for the final condition to be satisfied we need

$V(r,T;T)=1$ at $t=T$

$$\boxed{e^{A(T)-rB(T)} = 1} \Rightarrow A(T) - rB(T) = 0 \quad \forall r$$

and so

$$A(T) = B(T) = 0.$$

$A(T;T)=0$

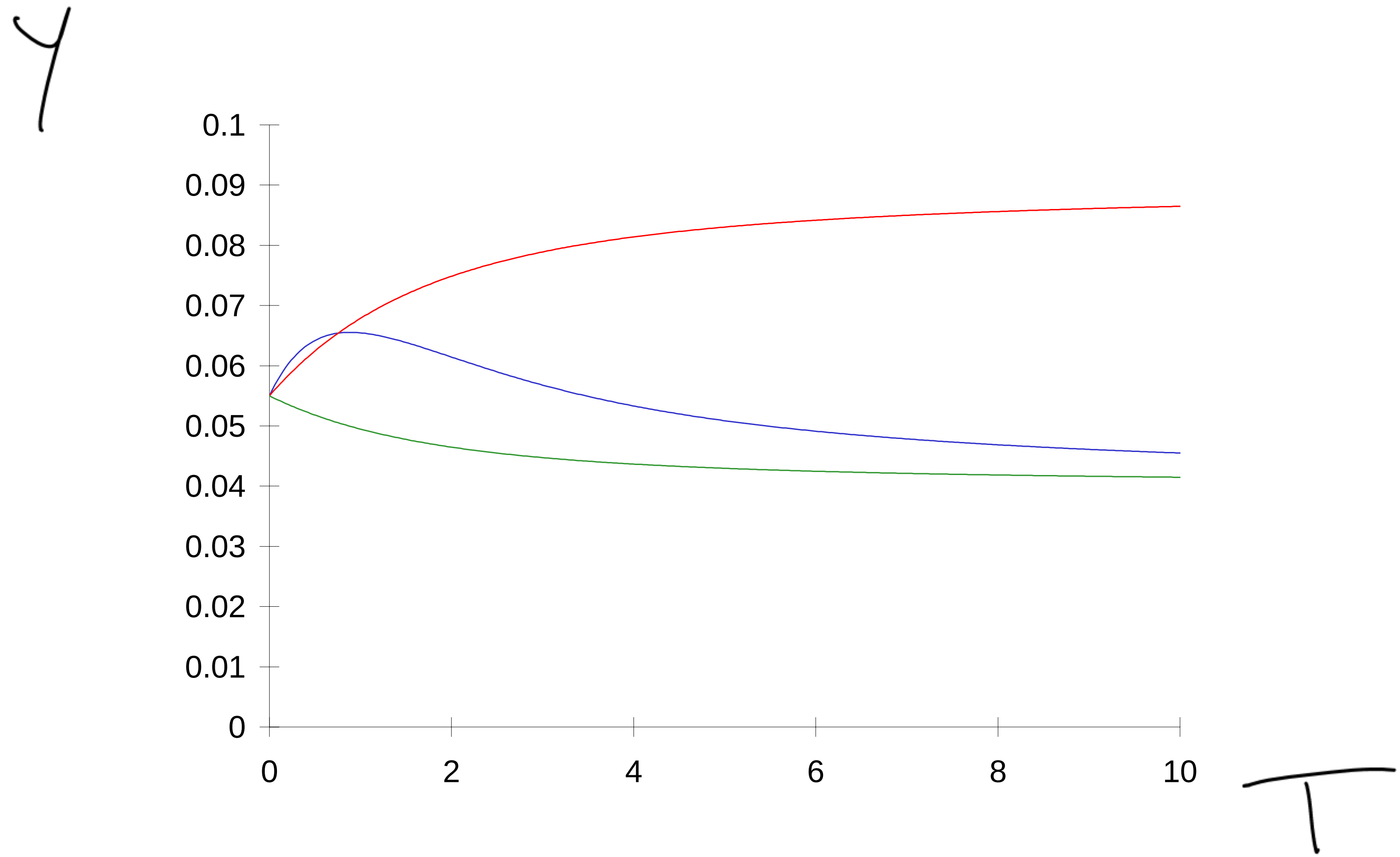
$B(T;T)=0$

The solution is

$$B = \frac{1}{\gamma}(1 - e^{-\gamma(T-t)})$$

and

$$A = \frac{1}{\gamma^2}(B(t;T) - T + t)(\eta\gamma - \tfrac{1}{2}\beta) - \frac{\beta B(t;T)^2}{4\gamma}.$$



Three types of yield curve given by the Vasicek model.

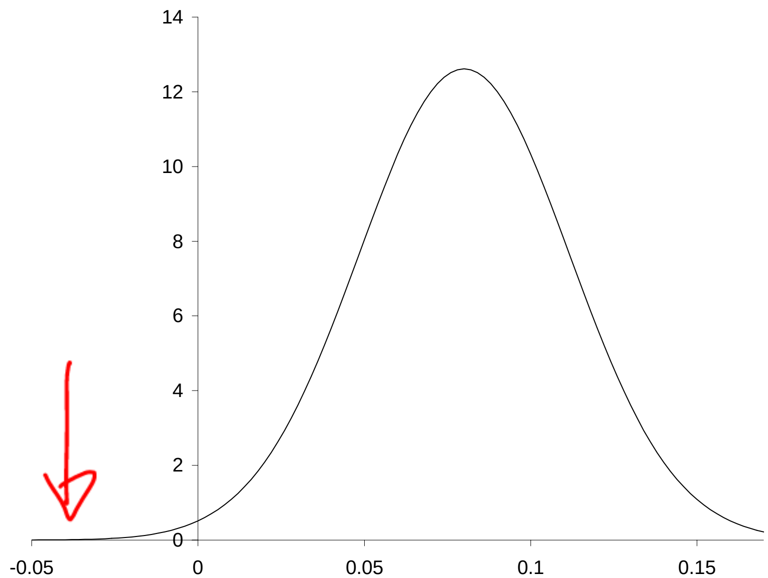
The Fokker–Planck equation can be used to find the probability distribution for the risk-neutral spot rate.

The steady-state probability density function for the Vasicek model is

Steady state distn

- $$P_\infty(r) = \sqrt{\frac{\gamma}{\beta\pi}}\, e^{-\frac{\gamma}{\beta}\left(r-\frac{\eta}{\gamma}\right)^2}.$$

Thus, in the long run, the spot rate is Normally distributed.



$r \to \pm\infty$

The spot rate can go negative!

Cox, Ingersoll & Ross

The CIR model takes the form

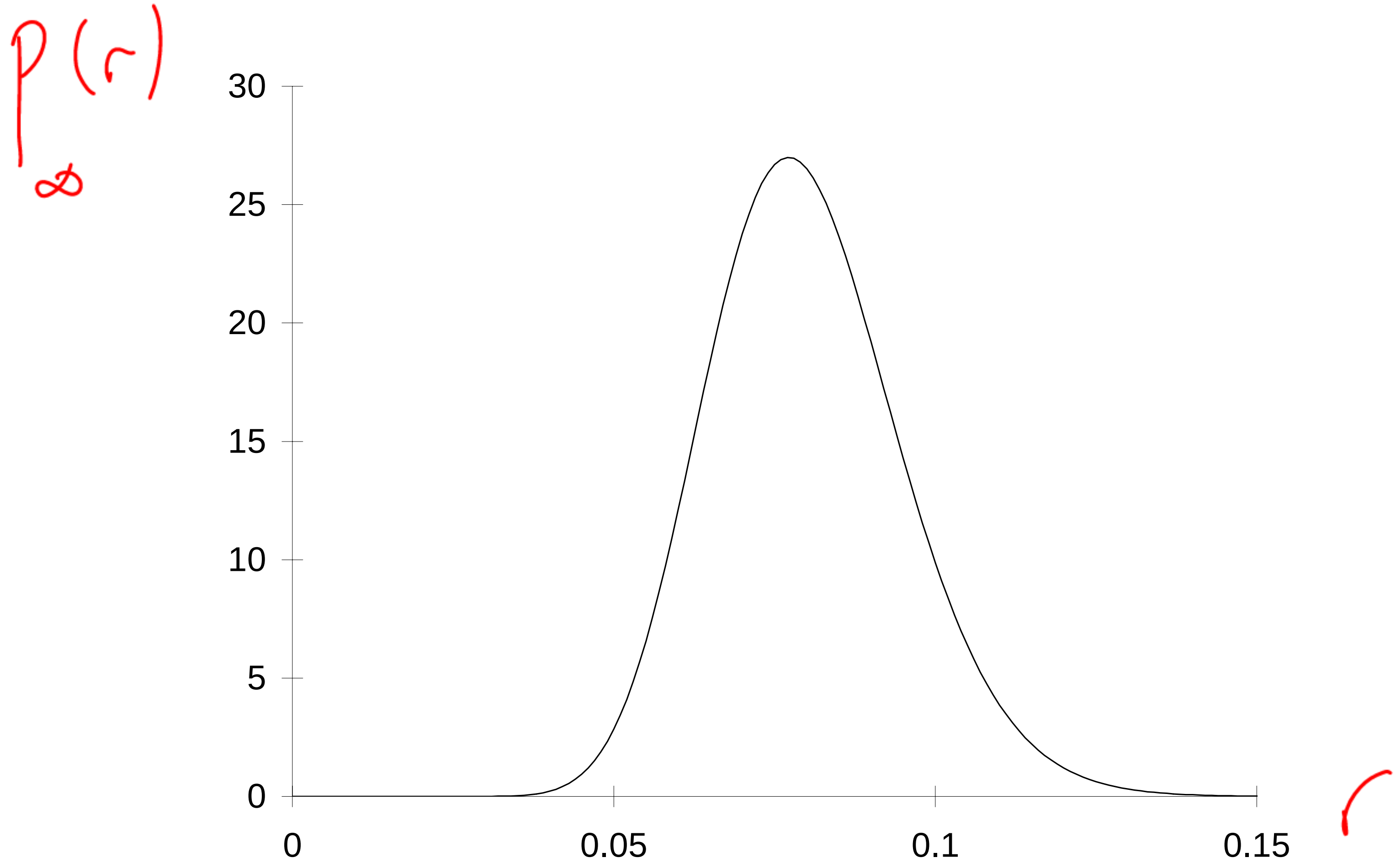
- $$dr = \overbrace{(\eta - \gamma r)}^{u-\lambda w}dt + \overbrace{\sqrt{\alpha r}}^{\text{diff.}}\, dX.$$

The spot rate is mean reverting and if $\eta > \alpha/2$ the spot rate stays positive.

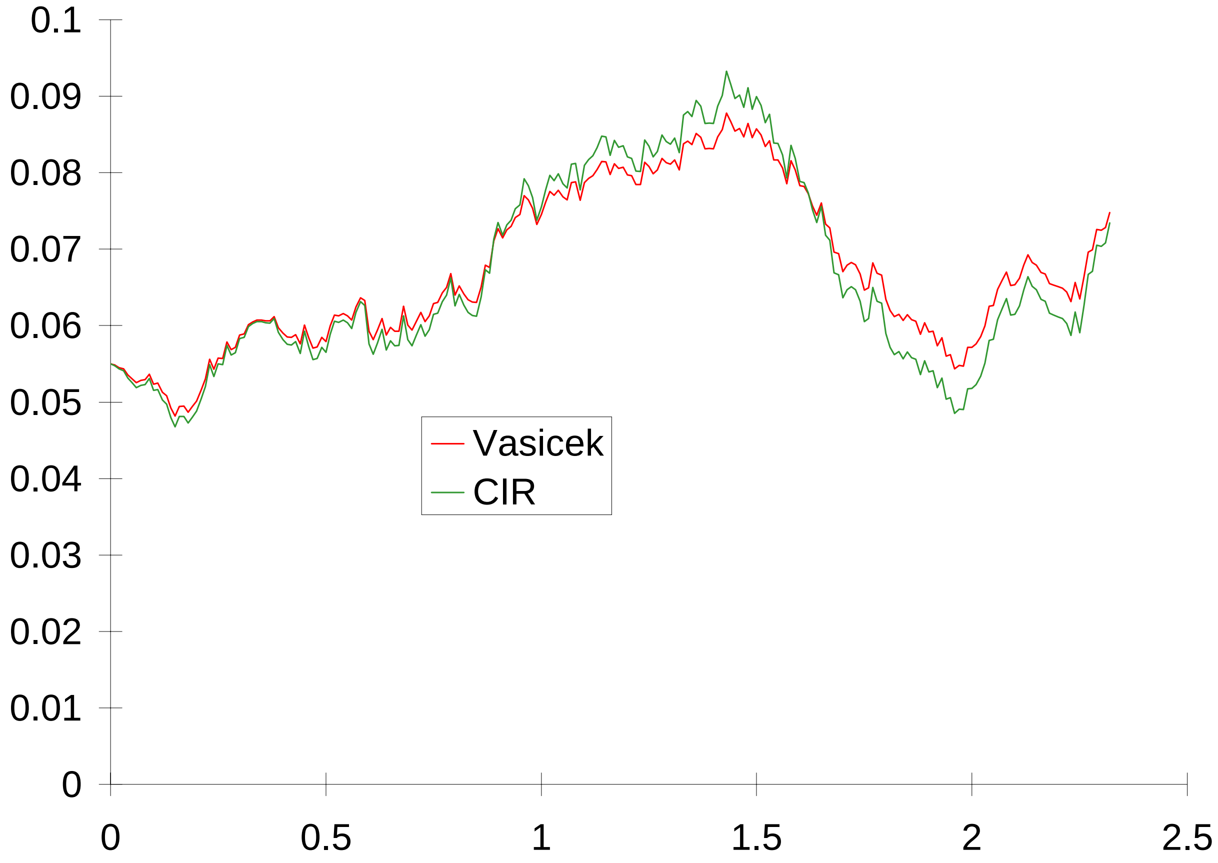
The value of a zero-coupon bond is again of the form

$$V(r,t;T) = e^{A(t)-rB(t)},$$

for different (and more complicated) functions A and B.



The steady-state probability density function for the risk-neutral spot rate in the CIR model.



A simulation of the Vasicek and CIR models using the same random numbers.

Hull & White I and II

Hull & White have extended both the Vasicek and the CIR models to incorporate time-dependent parameters:

$$1) \quad dr = (\eta(t) - \gamma(t)r)dt + \beta(t)^{1/2} dX \qquad \text{Vasicek}$$

$$2) \quad dr = (\eta(t) - \gamma(t)r)dt + \sqrt{\alpha(t)r}\, dX \qquad \text{CIR}$$

This time dependence again allows the yield curve (and even a volatility structure) to be fitted.

A more general model

Assume that $u - \lambda w$ and w take the form

Vasicek & CIR

$$u(r,t) - \lambda(r,t)w(r,t) = \boxed{\eta(t) - \gamma(t)r},$$

e^{A-rB}

$$\boxed{w(r,t) = \sqrt{\alpha(t)r + \beta(t)}.}$$

$\alpha = 0 \quad \beta \in \mathbb{R}$ Vasicek

$\beta = 0 \quad$ CIR.

We have chosen u and w in the stochastic differential equation for r to take special functional forms for a very special reason.

With these choices the solution for the zero-coupon bond is of the simple form

- $$V(r,t) = e^{A(t)-rB(t)}. \qquad (2)$$

We will see how this works in a moment but first let's look at the properties of this random walk.

By suitably restricting these time-dependent functions, we can ensure that the random walk for r has the following nice properties:

- Positive interest rates

- Mean reversion

Positive interest rates:

Except for a few pathological cases interest rates are positive.

With the above model the spot rate can be bounded below by a positive number if $\alpha(t) > 0$ and $\beta \leq 0$. The lower bound is $-\beta/\alpha$.

Note that r can still go to infinity, but with probability zero.

Mean reversion:

Examining the drift term, we see that for large r the (risk-neutral) interest rate will tend to decrease towards the mean, which may be a function of time.

When the rate is small it will move up on average.

We also want the lower bound to be non-attainable, we don't want the spot interest rate to get forever stuck at the lower bound or have to impose further conditions to say how fast the spot rate moves away from this value.

This requirement means that

$$\boxed{\eta(t) \geq -\beta(t)\gamma(t)/\alpha(t) + \alpha(t)/2.}$$

- The model with all of α, β, γ and η non-zero is the most general stochastic differential equation for r which leads to a solution of the form (2).

Let's see how this works.

$V = e^{A-rB}$ in B.P.

Substitute (2) into the pricing equation (1). This gives

$$\frac{\partial A}{\partial t} - r\frac{\partial B}{\partial t} + \tfrac{1}{2}w^2B^2 - (u - \lambda w)B - r = 0. \qquad (3)$$

general drift & diffusion

Some of these terms are functions of t and T (i.e. A and B) and others are functions of r and t (i.e. u and w).

Differentiating (3) with respect to r gives

$$\longrightarrow \quad -\cancel{\frac{\partial B}{\partial t}} + \tfrac{1}{2}B^2\frac{\partial}{\partial r}(w^2) - B\frac{\partial}{\partial r}(u - \lambda w) \cancel{-1} = 0.$$

Differentiate again with respect to r and divide through by B:

$$\tfrac{1}{2}B\underbrace{\frac{\partial^2}{\partial r^2}(w^2)} - \underbrace{\frac{\partial^2}{\partial r^2}(u - \lambda w)} = 0.$$

Set both indep equal to zero.

In this, only B can depend on the bond maturity T, therefore we must have

$$\frac{\partial^2}{\partial r^2}(w^2) = 0, \quad -\frac{\partial^2}{\partial r^2}(u - \lambda w) = 0.$$

Therefore $u - \lambda w$ and w must be linear in r as proposed.

$\int dr$: $\frac{\partial}{\partial r}(w^2) = \alpha(t)$

$\int dr$: $w^2 = \alpha(t)r + \beta(t)$

$\therefore$ $w = \sqrt{\alpha(t)r + \beta(t)}$

$-\frac{\partial^2}{\partial r^2}(u - \lambda w) = 0$

$\int dr$: $-\frac{\partial}{\partial r}(u - \lambda w) = \gamma(t)$

$\frac{\partial}{\partial r}(u - \lambda w) = -\gamma(t)$

$\int dr$: $(u - \lambda w) = -\gamma(t)r + \eta(t)$

$dr = (\eta(t) - \gamma(t)r)dt + \sqrt{\alpha(t)r + \beta(t)}\, dX$

The equations for A and B are

$$\frac{\partial A}{\partial t} = \eta(t)B - \tfrac{1}{2}\beta(t)B^2 \tag{4}$$

and

$$\frac{\partial B}{\partial t} = \tfrac{1}{2}\alpha(t)B^2 + \gamma(t)B - 1. \tag{5}$$

In order to satisfy the final data that $Z(r,T;T) = 1$ we must have

$$A(T;T) = 0 \quad \text{and} \quad B(T;T) = 0.$$

Redemption Value.

Solution for constant parameters

The solution for arbitrary α, β, γ and η is found by integrating the two ordinary differential equations (4) and (5).

(Generally speaking, though, when these parameters are time dependent this integration cannot be done explicitly.)

The simplest case is when α, β, γ and η are all constant. Then

$$\frac{\alpha}{2}A = a\psi_2 \log(a - B) + (\psi_2 + \tfrac{1}{2}\beta)b \log((B + b)/b) - \tfrac{1}{2}B\beta - a\psi_2 \log a,$$

and

$$\boxed{B(t;T) = \frac{2(e^{\psi_1(T-t)} - 1)}{(\gamma + \psi_1)(e^{\psi_1(T-t)} - 1) + 2\psi_1},}$$

$(T-t) \to \infty$

where

$$b, a = \frac{\pm\gamma + \sqrt{\gamma^2 + 2\alpha}}{\alpha},$$

and

$$\psi_1 = \sqrt{\gamma^2 + 2\alpha} \quad \text{and} \quad \psi_2 = \frac{\eta - a\beta/2}{a + b}.$$

When all four of the parameters are constant it is obvious that both A and B are functions of only the one variable $\tau = T - t$, and not t and T individually; this would not necessarily be the case if any of the parameters were time dependent.

A wide variety of yield curves can be predicted by the model. As $\tau \to \infty$,

$\tau = T - t$

$$B \to \frac{2}{\gamma + \psi_1}$$

and the yield curve Y has long term behavior given by

$$Y \to \frac{2}{(\gamma + \psi_1)^2} \left(\eta(\gamma + \psi_1) - \beta \right).$$

Thus for constant and fixed parameters the model leads to a fixed long-term interest rate, independent of the spot rate.

The probability density function, $P(r,t)$, for the risk-neutral spot rate satisfies

$$\boxed{\frac{\partial P}{\partial t} = \frac{1}{2}\frac{\partial^2}{\partial r^2}(w^2 P) - \frac{\partial}{\partial r}((u - \lambda w)P).}$$

In the long term this settles down to a distribution, $P_\infty(r)$, that is independent of the *initial* value of the rate.

This distribution satisfies the ordinary differential equation

- $$\boxed{\frac{1}{2}\frac{d^2}{dr^2}(w^2 P_\infty) = \frac{d}{dr}((u - \lambda w)P_\infty).}$$

The solution of this for the general model with constant parameters is

$$P_\infty(r) = \frac{\left(\frac{2\gamma}{\alpha}\right)^k}{\Gamma(k)} \left(r + \frac{\beta}{\alpha}\right)^{k-1} e^{-\frac{2\gamma}{\alpha}(r+\frac{\beta}{\alpha})}$$

where

$$k = \frac{2\eta}{\alpha} + \frac{2\beta\gamma}{\alpha^2}$$

and $\Gamma(\cdot)$ is the gamma function. The boundary $r = -\beta/\alpha$ is non-attainable if $k > 1$.

The mean of the steady-state distribution is

$$\frac{\alpha k}{2\gamma} - \frac{\beta}{\alpha}.$$

Multi-factor interest rate modeling

The simple one-factor stochastic spot interest rate models cannot hope to capture the rich yield-curve structure found in practice: from a given spot rate at a given time they will predict the whole yield curve.

Generally speaking, the one source of randomness, the spot rate, may be good at modeling the overall level of the yield curve but it will not necessarily model shifts in the yield curve that are substantially different at different maturities.

For some instruments this may not be important. For example, for instruments that depend on the *level* of the yield curve it may be sufficient to have one source of randomness, i.e. one factor.

More sophisticated products depend on the difference between yields of different maturities and for these products it is important to model the tilting of the yield curve.

One way to do this is to invoke a second factor, a second source of randomness.

Theoretical framework

$dX_1 \sim \phi_1 \sqrt{dt}$

$dX_2 \sim \phi_2 \sqrt{dt}$

$\mathbb{E}[\phi_1 \phi_2] = \rho$

Assume that simple interest rate depend on two variables r, the spot interest rate, and another independent variable l where

2 factor model

$$\boxed{dr = u\,dt + w\,dX_1}$$

dX_1

$\mathbb{E}[dX_1 dX_2] = \rho\, dt$

and

$$dl = p\,dt + q\,dX_2.$$

dX_2

Simple instruments will then have prices which are functions of r, l and t, $V(r, l, t)$.

$V = V(r, l, t; T)$

All of u, w, p and q are allowed to be functions of r, l and t. The correlation coefficient ρ between dX_1 and dX_2 may also depend on r, l and t.

Note that we have not said what l is.

It could be another interest rate, a long rate, say, or the yield curve slope at the short end, or the volatility of the spot rate, for example.

We set up the framework in general and look at specific models later.

Since we have two sources of randomness now, in pricing one bond we must hedge with *two* others to eliminate the risk:

$$\Pi = V(r,l,t;T) - \Delta_1 V_1(r,l,t;T_1) - \Delta_2 V_2(r,l,t;T_2).$$

What happens to Π as $t \to t + dt$?

$d\Pi = dV - \Delta_1 dV_1 - \Delta_2 dV_2$ Using Itô on $V(r,l,t;T)$

$$dV = \frac{\partial V}{\partial t}dt + \frac{\partial V}{\partial r}dr + \frac{\partial V}{\partial l}dl + \frac{1}{2}\frac{\partial^2 V}{\partial r^2}dr^2 + \frac{1}{2}\frac{\partial^2 V}{\partial l^2}dl^2 + \frac{\partial^2 V}{\partial r \partial l}dr\,dl$$

$$= \underbrace{\left(\frac{\partial V}{\partial t} + \frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + \frac{1}{2}q^2\frac{\partial^2 V}{\partial l^2} + \rho w q\frac{\partial^2 V}{\partial r \partial l}\right)dt}_{\mathcal{L}(V)\,dt} + \underline{\frac{\partial V}{\partial r}dr} + \underline{\frac{\partial V}{\partial l}dl}$$

$$\mathcal{L} \equiv \left(\frac{\partial}{\partial t} + \frac{1}{2}w^2\frac{\partial^2}{\partial r^2} + \frac{1}{2}q^2\frac{\partial^2}{\partial l^2} + \rho w q\frac{\partial^2}{\partial r \partial l}\right)$$

The change in the value of this portfolio is given by

$$d\Pi = \boxed{\left(\mathcal{L}(V) - \Delta_1 \mathcal{L}(V_1) - \Delta_2 \mathcal{L}(V_2)\right) dt} = r\Pi\, dt$$

$$+ \underbrace{\left(\frac{\partial V}{\partial r} - \Delta_1 \frac{\partial V_1}{\partial r} - \Delta_2 \frac{\partial V_2}{\partial r}\right)}_{=0} dr + \underbrace{\left(\frac{\partial V}{\partial l} - \Delta_1 \frac{\partial V_1}{\partial l} - \Delta_2 \frac{\partial V_2}{\partial l}\right)}_{=0} dl, \quad (6)$$

with the obvious notation for V, V_1 and V_2.

$d\Pi = r\Pi\, dt$

Here

$$\mathcal{L}(V) = \frac{\partial V}{\partial t} + \tfrac{1}{2} w^2 \frac{\partial^2 V}{\partial r^2} + \rho w q \frac{\partial^2 V}{\partial r \partial l} + \tfrac{1}{2} q^2 \frac{\partial^2 V}{\partial l^2}.$$

$$\left[\mathcal{L}(V) - \Delta_1 \mathcal{L}(V_1) - \Delta_2 \mathcal{L}(V_2)\right] dt = r\left(V - \Delta_1 V_1 - \Delta_2 V_2\right) dt$$

$\mathcal{L}(V) - rV$ Call this $\mathcal{L}'$

Now choose Δ_1 and Δ_2 to make the coefficients of dr and dl in (6) equal to zero. The corresponding portfolio is risk free and should earn the risk-free rate of interest, r.

We thus have the three equations

$$\frac{\partial V}{\partial r} - \Delta_1 \frac{\partial V_1}{\partial r} - \Delta_2 \frac{\partial V_2}{\partial r} = 0,$$ Coeff. of dr

$$\frac{\partial V}{\partial l} - \Delta_1 \frac{\partial V_1}{\partial l} - \Delta_2 \frac{\partial V_2}{\partial l} = 0$$ Coeff. of dl

and

$$\mathcal{L}'(V) - \Delta_1 \mathcal{L}'(V_1) - \Delta_2 \mathcal{L}'(V_2) = 0$$

where

$$\mathcal{L}'(V) = \mathcal{L}(V) - rV.$$

3 eq$\hat{=}$s in 2 unknowns

These are three simultaneous equations for Δ_1 and Δ_2. As such, this system is over-prescribed and for the equations to be consistent we require

$$\boxed{\det(\mathbf{M}) = 0}$$

where

$$\mathbf{M} = \begin{pmatrix} \mathcal{L}'(V) & \mathcal{L}'(V_1) & \mathcal{L}'(V_2) \\ \partial V/\partial r & \partial V_1/\partial r & \partial V_2/\partial r \\ \partial V/\partial l & \partial V_1/\partial l & \partial V_2/\partial l \end{pmatrix}.$$

$$\begin{bmatrix} \mathcal{L}'(V) & - & - & - \\ \partial V/\partial r & - & - & - \\ \partial V/\partial \lambda & - & - & - \end{bmatrix} \begin{bmatrix} 1 \\ -\Delta_1 \\ -\Delta_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$a(r,t) = \lambda w - u$

The first row of the matrix **M** is a linear combination of the second and third rows.

$$\mathcal{L}'(V) = \alpha_1 \frac{\partial V}{\partial r} + \alpha_2 \frac{\partial V}{\partial l}$$

We can therefore write

$$\mathcal{L}'(V) = (\lambda_r w - u)\frac{\partial V}{\partial r} + (\lambda_l q - p)\frac{\partial V}{\partial l}$$

where the two functions $\lambda_r(r,l,t)$ and $\lambda_l(r,l,t)$ are the market prices of risk for r and l respectively, and are again independent of the maturity of any bond.

$$dr = (u - \lambda_r w)\,dt + w\,dX_1$$

$$dl = (p - \lambda_l q)\,dt + q\,dX_2$$

In full, we have $V(r, \lambda, t; T)$

$$\frac{\partial V}{\partial t}+\frac{1}{2}w^2\frac{\partial^2 V}{\partial r^2}+\rho w q\frac{\partial^2 V}{\partial r\partial l}+\frac{1}{2}q^2\frac{\partial^2 V}{\partial l^2}+(u-\lambda_r w)\frac{\partial V}{\partial r}+(p-\lambda_l q)\frac{\partial V}{\partial l}-rV=0. \tag{7}$$

The model for interest rate derivatives is defined by the choices of w, q, ρ, and the risk-adjusted drift rates $u-\lambda_r w$ and $p-\lambda_l q$.

This is yet another parabolic partial differential equation.

It has more variables than other partial differential equations we have seen.

Popular models

In this section we see some popular models.

Most of these models are popular because the pricing equations (7) for these models have explicit solutions.

In these models sometimes the second factor is the long rate and sometimes it is some other, usually unobservable, variable.

Brennan and Schwartz (1982)

In the Brennan and Schwartz model the risk-adjusted spot rate satisfies

$$dr = (a_1 + b_1(l - r))dt + \sigma_1 r\, dX_1$$

and the long rate satisfies

random

$$dl = l(a_2 - b_2 r + c_2 l)dt + \sigma_2 l\, dX_2.$$

Brennan and Schwartz choose the parameters statistically.

Because of the relatively complicated functional forms of the terms in these equations there are no simple solutions of the bond pricing equation.

The random terms in these two stochastic differential equations are of the lognormal form, but the drift terms are more complicated than that, having some mean reversion character.

The main problem with the Brennnan and Schwartz model is that it can blow up in a finite time, meaning that rates can go to infinity.

Fong and Vasicek (1991) stochastic vol model

Fong & Vasicek consider the following model for risk-adjusted variables:

vol.

$$dr = a(\bar{r} - r)dt + \sqrt{\xi}\,dX_1$$

and

c is vol. of vol.

$$d\xi = b(\bar{\xi} - \xi)dt + c\sqrt{\xi}\,dX_2.$$

Thus they model r, the risk-adjusted spot rate, and ξ the square root of the volatility of the spot rate.

The latter cannot be observed, and this is an obvious weakness of the model. But it also makes it harder to show that the model is wrong.

The simple linear mean reversion and the square roots in these equations results in explicit equations for simple interest rate products.

Longstaff and Schwartz (1992)

Longstaff & Schwartz consider the following model for risk-adjusted variables:

$$dx = a(\bar{x} - x)dt + \sqrt{x}\, dX_1$$

and

$$dy = b(\bar{y} - y)dt + \sqrt{y}\, dX_2,$$

where the spot interest rate is given by

$$\boxed{r = cx + dy.}$$

Again, the simple nature of the terms in these equations results in explicit equations for simple interest rate products.

General affine model

$$dr = (\eta(t) - \gamma(t) r)\,dt + \sqrt{\alpha(t) r + \beta(t)}\,dX$$

If r and l satisfy the following:

- the risk-adjusted drifts of both r and l are linear in r and l (but can have an arbitrary time dependence)
- the random terms for both r and l are both square roots of functions linear in r and l (but can have an arbitrary time dependence)
- the stochastic processes for r and l are uncorrelated $\rho = 0$

then. . .

. . . the two-factor pricing equation (7) has a solution for a zero-coupon bond of the form

$$V(r,\lambda,t;T) = e^{\boxed{A(t)-B(t)r}-C(t)l}.$$

The ordinary differential equations for A, B and C must in general be solved numerically.

Summary

Please take away the following important ideas

- interest rates can be modeled as stochastic variables

λ

- whenever a modeled quantity is not traded the pricing equation contains a market price of risk term

- the pricing equation is another partial differential equation, similar in form to the Black–Scholes equation B.P.E

- Extended 1 factor model to 2 factors

Backtesting and Essentials of Statistical Arbitrage (Pairs Trading)

Robust Portfolio Construction

by Dr Richard Diamond

CQF FINAL PROJECT 2020

Fundamentals + Techniques

Choices and Preparation

- Start collecting data and planning the project for you.
- It is up to you to source and clean the suitable data.

  Webex session on Equities Data. If you can't get hold of some data: make reasonable assumptions, even generate the data by Monte-Carlo.

  Set your option strikes and maturities, clean rows with missing rates in BOE data, assume reasonable CDS spreads, etc.

- A. You can adopt code for specific partial tasks, not model as a whole, amending it for your purpose (not copy/paste). B. Use ready libraries with expertise – quadratic optimisation, kernels, etc – methods have to be suited to the task. C. You are welcome to implement complex numerical methods vs. ready solution if able to.

Numerical Techniques

Implement numerical techniques from the first principles as necessary. Pricing a spread/CVA from simulated curve/optimal allocation is the result.

What to code: pricing formulae, Black-Litterman calculation, SDE simulation, matrix form regression, as Engle-Granger, interpolation, numerical integration, Cholesky, t-copula formula, CDS bootstrap, features computation...

Ledoit-Wolf (py, Matlab code)

Use ready solutions for: covariance shrinkage, nearest correlation, ML numerical methods (eg, decision trees, neural nets), low latency RNs, kernel density (cdf estimation), QR-decomposition (PCA), EGARCH estimation, Johansen Procedure...

The lists are not exhaustive.

Project Report

- A full **mathematical description** of the models employed as well as numerical methods. Remember *accuracy and convergence*!
- Results presented using **a plenty of tables and figures**, which must be interpreted not just thrown at the reader.
- **Pros and cons** of a model and its implementation, together with possible improvements.
- **Demonstrate 'the specials'** of your implementation: own research, own coding of complex methods, use of the industrial-strength libraries of C++, Python or VBA + NAG.
- Instructions on how to use software if not obvious.
  The code must be thoroughly tested and well-documented.

CQF Electives

See Brief for Full Table

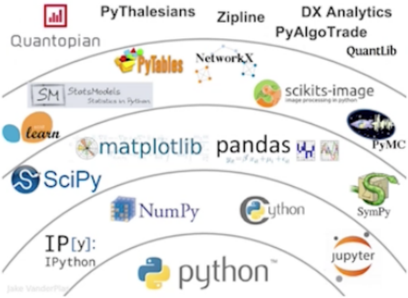
Counterparty Credit Risk – CR, IR	CDS, survival probabilities and hazard rates reviewed. Three key numerical methods for quant finance pricing (Monte-Carlo, Binomial Trees, Finite Difference). Monte Carlo for simple LMM. Review of Module Five on Credit with a touch on the copula method. **Outcome:** covers CVA Computation clearly and reviews of credit spread pricing techniques.
Risk Budgeting – PC primary choice	Reviews the nuance of Modern Portfolio Theory, ties in VaR and Risk Decomposition with through derivations and expectation algebra. Gives simple examples of figures you need to compute and then combine with portfolio optimisation. Risk-budgeting portfolio from Video Part 10.
Advanced Volatility Modeling – LV primary	Considers the main kinds of PDEs (stochastic volatility and jump diffusion) and their analytical solution: the main approach to solve stochastic volatility (Heston model) is via Fourier Transform. In-depth on integration **Outcome:** Local Volatility topic offers a classic pricing PDE, which can be solved by techniques from this elective.
C++ – Dev	Consider this a revised version of C++ Primer/ initial certificate course.

Final Day is 12th July 2020

Don’t Extend Your Luck!

The Quant Finance PyData Stack

Source: [Jake VanderPlas: State of the Tools](https://www.youtube.com/watch?v=5GlNDD7qbP4)



Electives on **Data Analytics** and **Algotrading** and *Python*. Source: QI 2015 Quantopian talk on *Portfolio and Risk Analytics with PyFolio*.

Key Libraries

1. **C++** has the famous but lesser documented QuantLib. ← CR, IR
   `www.quantlib.org/docs.shtml`
   Ported option pricing and rates modelling to R quantlib.

2. **R** packages to explore: portfolioSim, PerformanceAnalytics, QuantStrat.

3. **Quantopian** backtesting system can solve the issues with (time series) orders
   - rule-based trading strategies
   - access to clean data to backtest against
   - producing professional output

   your P&L

   Factors

   Quantopian 'playback' illustrates recurrent trading (reallocation), not for static or infrequent allocation.

The coding language is special Python, in fact based on open-source package *zipline*.

Code **HAS TO BE** executed on Quantopian's own platform (IDE window on their website).
That backtesting result can be pulled to your Python notebook by 'tearsheet reference' and *pyfolio* package (Webex Session to come).

```
import numpy as np
import statsmodels.api as sm
import pandas as pd

import quantopian.optimize as opt
import quantopian.algorithm as algo
```

Pairs Trading

`www.quantopian.com/lectures/example-pairs-trading-algorithm`

Portfolio Construction

`www.quantopian.com/tutorials/getting-started#lesson7`

```python
def initialize(context):
    # Quantopian backtester specific variables
    set_slippage(slippage.FixedSlippage(spread=0))
    set_commission(commission.PerTrade(cost=1))
    set_symbol_lookup_date('2014-01-01')

    context.stock_pairs = [(symbol('ABGB'), symbol('FSLR')),
                           (symbol('CSUN'), symbol('ASTI'))]

    context.stocks = symbols('ABGB', 'FSLR', 'CSUN', 'ASTI')

    context.num_pairs = len(context.stock_pairs)
    # strategy specific variables
    context.lookback = 20 # used for regression
    context.z_window = 20 # used for zscore calculation, must be <= loo

    context.target_weights = pd.Series(index=context.stocks, data=0.25)

    context.spread = np.ndarray((context.num_pairs, 0))
    context.inLong = [False] * context.num_pairs
    context.inShort = [False] * context.num_pairs
```

Rolling beta

1. We will look at **how to relate P&L** to the market and factors, to understand what drives P&L, what you make money on.
2. Then, we will talk about **evaluating P&L** with drawdown control and VaR.

   Rolling SR $SR \propto 1/\sqrt{\tau}$

3. You can look for suitable models for algorithmic **order flow** and liquidity impact. [Optional]

Please refer to Algotrading Elective and Quantopian material

`https://www.quantopian.com/lectures/fundamental-factor-models`

Alpha and Beta

Beta is the strategy's market exposure, for which you should not pay much as it is easy to gain by buying an ETF or index futures contract.

Alpha is the excess return after subtracting return due to market movements.

Rets from SP500

Returns from Strategy

$$R_t^S = \alpha + \beta R_t^M + \epsilon_t$$

$$\mathbb{E}[R_t^S - \beta R_t^M] = \alpha$$

$R_t^S \quad + \; R_t^M \quad + \epsilon_t$

$R_t^M = R_t - r_f$ is the time series of returns representing **the market factor**.

3M US Treasuries

Risk-Reward Ratios

Information Ratio (IR) focuses on risk-adjusted *abnormal* return, the risk-adjusted alpha!

$$\frac{\alpha}{\sigma(\epsilon)}$$

of residuals

$$\frac{\mu_\pi - r_f}{\sigma(\text{rets})}$$

(That doesn't tell us how much dollar alpha is there. It can be eaten by transaction costs.)

Sharpe Ratio measures return per unit of risk. Familiar form:

$$\frac{\mathbb{E}(R_t - r_f)}{\sigma(R_t - r_f)}$$

Factors

Evaluating performance **against factors** is the central part of the backtesting.

We saw the separation of alpha and beta in regression *wrt* one market factor

$$R_t^S = \alpha + \beta R_t^M + \epsilon_t$$

We see that a factor is a time series of changes, similar to the series of asset returns.

L Long / Short portfolio

L Time series of returns

Named Factors

- **Up Minus Down** (UMD) or **momentum** factor would leverage on stocks that are going up. The recent month's returns are excluded from calculation to avoid a spurious signal.
- **Small Minus Big** (SMB) factor shorts large cap stocks, so β^{SMB} measures the tilt towards small stocks.
- Long-short **High Minus Low** (HML) or **value** factor: buy top 30% long of companies with the high book-to-market value and sell the bottom 30% (expensive stocks). short

1) Except for HML, the impact/presence of other factors questionable.

2) Since 2015, Fama-French moved to 5-factor model that include profitability RMW and investment CMA but ignore the proper 3) Momentum factor and 4) Low Volatility (Betting Against Beta) factors. BAB

Factors Backtesting

$R_t^S = \alpha + \beta^M R_t^M + \epsilon_t$

0.8

So how do we check against those factors?

Set up a regression!

≈ 0.6 ≈ 0.2

$$R_t^S = \alpha + \beta^M R_t^M + \beta^{HML} R_t^{HML} + \epsilon_t$$

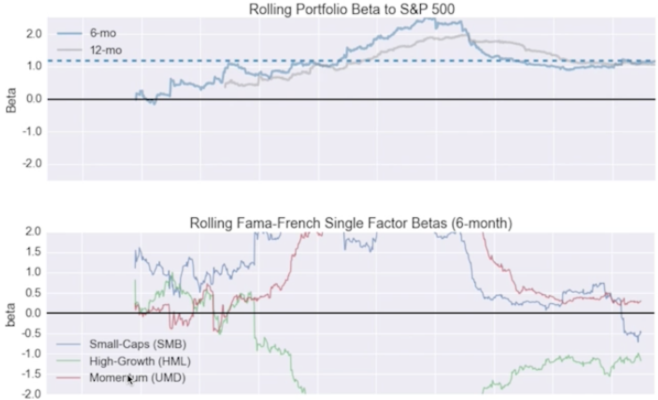
Value Factor

where R_t^{HML} is return series from the long-short HML factor.

- We can **add factors** to this regression.
- We can have **rolling estimates** of these betas for each day/week.

- Scale returns to have the same volatility as the benchmark – put on the same plot for correct comparison.
- Rolling Sharpe Ratio – changes **not** desirable).
- Rolling market factor beta – $\beta > 1$ **not** desirable.
- Rolling betas *wrt* to UMD (momentum), SMB, and industry sectors.

1) pyfolio on own Python Notebook
2) Quantopian IDE (full backtest)
3) zipline (no data)



From: *Portfolio and Risk Analytics with PyFolio*, T. Wiecki, QI 2015

Drawdowns

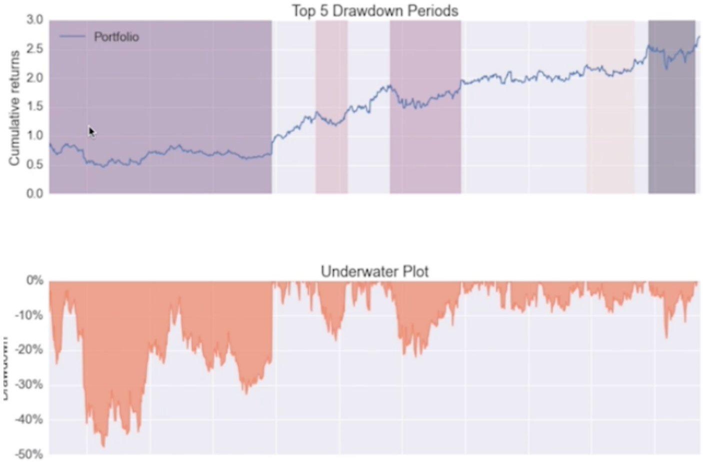
The drawdown is the cumulative percentage loss, given the loss in the initial timestep.

Let's define the highest past peak performance as High Water Mark

$$\mathrm{DD}_t = \frac{\mathrm{HWM}_t - P_t}{\mathrm{HWM}_t}$$

where P_t is the cumulative return (or portfolio value Π_t).

It makes sense to evaluate a maximum drawdown over past period $\max_{t \leq T} \mathrm{DD}_t$.



From: Quant Insights, Oct 2015, *Portfolio and Risk Analytics with PyFolio*, Thomas Wiecki (Quantopian)

Drawdown Control

The strategy must be able to survive without running into a close-out.

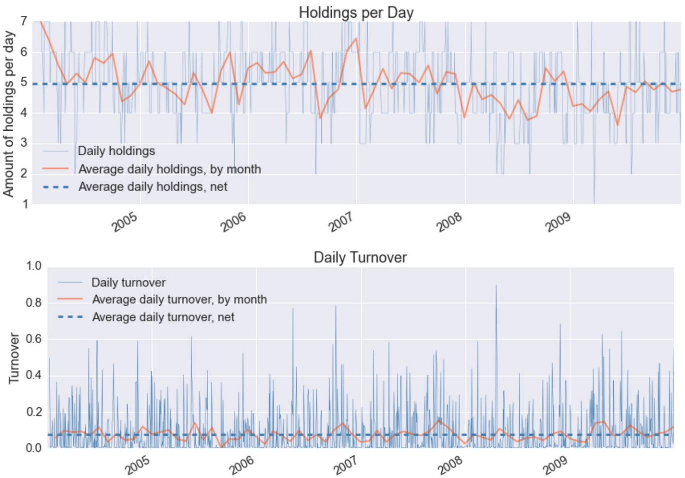
It makes sense to pre-define Maximum Acceptable Drawdown (MADD) and trace

$$\mathrm{VaR}_t \leq \mathrm{MADD} - \mathrm{DD}_t$$

where VaR_t is today's VaR and DD_t is current drawdown.

Backtesting Risk and Liquidity - Summary

1. Does cumulative P&L behave as expected (eg, for a cointegration trade)? Behaviour of risk measures (volatility/VaR/Drawdown)?
2. Is P&L coming from a few large trades or many smaller trades? Does all profit come from a particular period. Concentration in assets and its attribution – as intended?
3. Turnover (good or bad?), impact of transaction costs (slippage). Plot the P&L value (or its alpha) vs. number of transactions.



From: Quant Insights, Oct 2015, *Portfolio and Risk Analytics with PyFolio*, Thomas Wiecki (Quantopian)

Algorithmic Flow

3. Optionally, introduce liquidity and algorithmic flow considerations (a model of order flow). How would you be entering and accumulating the position? What impact *your transactions* will make on the market order book?

4. Related issue is the possible leverage for the strategy. While the maximum leverage is $1/\text{Margin}$, the more adequate solution is a maximally leveraged market-neutral gain or alpha-to-margin ratio

$$\text{AM} = \frac{\alpha}{\text{Margin}}.$$

Quantopian Platform: Strat Backtesting & Risk

1. Overview of New Backtesting platform (2018)
   `www.quantopian.com/posts/improved-backtest-analysis`

2. Quantopian Risk Model for portfolio construction
   `www.quantopian.com/posts/`
   `new-tool-for-quants-the-quantopian-risk-model`

   White paper at `https://www.quantopian.com/papers/risk`

3. The previous version and backtesting capabilities at
   `github.com/quantopian/pyfolio`

Quick Algo Checker

```
Results: 2016-05-02 to 2018-05-14
Score                        0.0792  Constraints met 8/9
Returns                        7.9%  PASS: Positive
Positions                 4.88|5.17  PASS: Max position concentration 4.88% <= 5.0%
Leverage       0.95|0.97|1.03|1.06  PASS: Leverage range 0.97x-1.03x between 0.8x-1.1x
Turnover           3.9|4.3|8.3|8.6  FAIL: 2nd percentile turnover 4.3% < 5.0x
Net exposure                1.7|2.1  PASS: Net exposure (absolute value) 1.7% <= 10.0%
Beta-to-SPY               0.24|0.28  PASS: Beta 0.24 between +/-0.30
Sectors                   0.08|0.08  PASS: All sector exposures between +/-0.20
Style                     0.28|0.29  PASS: All style exposures between +/-0.40
Tradable                     96|100  PASS: Investment in QTradableStocksUS >= 95.0%
```

From: https://www.quantopian.com/posts/contest-constraint-check-notebook-with-compact-output

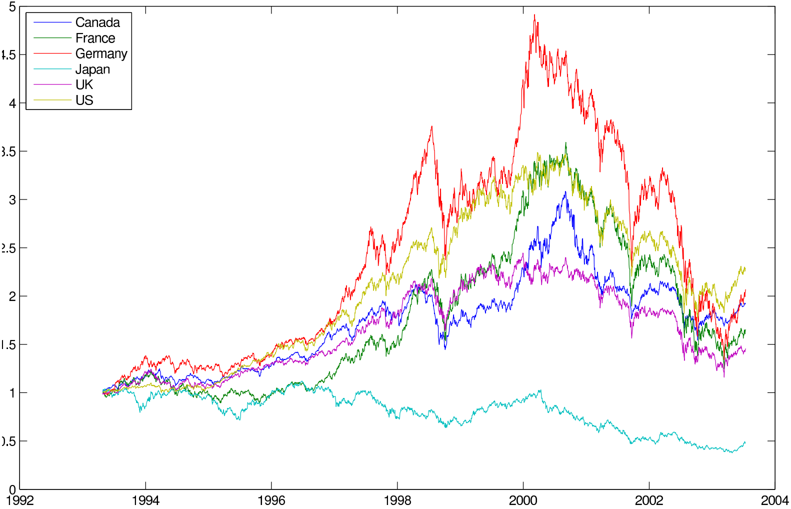
Developing a Trading Business

There are libraries for anything: data download, regression and ML, backtesting and tear sheets/trading analytics.

Building an Energy Trading Business from Scratch, Teodora Baeva (BTG Pactual), QI 2015,

Techniques from Time Series

Relative Equity Indices

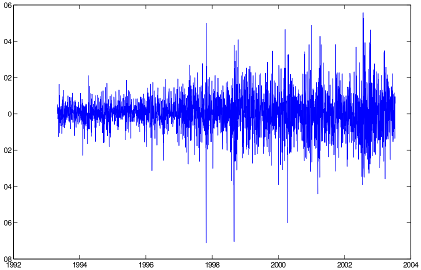


US Daily Index Returns

We will use index returns to demo *Vector Autoregression*.

- Canada, France, Germany, Japan, UK, US

Below is typical plot for the market daily returns (US). Observe the regimes of low, then high volatility.



Linear Model on Returns

For stationary returns, we set up a model-free endogenous system: variables depend on their past (lagged) values.

$N_{lags} = 3$

AIC , BIC

$$
\begin{aligned}
R_t^{CA=1} &= \beta_{1,0} + \beta_{11} R_{t-1}^{CA} + \beta_{12} R_{t-1}^{FR} + \dots \beta_{1n} R_{t-1}^{US} + \dots_{t-2} \dots + \epsilon_{1,t} \\
R_t^{FR=2} &= \beta_{2,0} + \beta_{21} R_{t-1}^{CA} + \beta_{22} R_{t-1}^{FR} + \dots \beta_{2n} R_{t-1}^{US} + \dots_{t-2} \dots + \epsilon_{2,t} \\
\dots & \quad \dots \\
R_t^{US=n} &= \beta_{n,0} + \beta_{n1} R_{t-1}^{CA} + \beta_{nn} R_{t-1}^{FR} + \dots \beta_{nn} R_{t-1}^{US} + \dots_{t-2} \dots + \epsilon_{n,t}
\end{aligned}
$$

CA

Asset n

lag 1

A_1

Consider forecasting powers of this model-free set up.

Empirical Forecasting

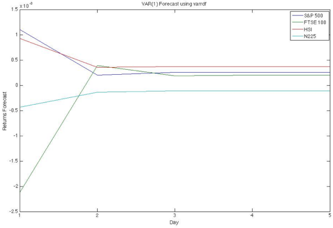
Vector Autoregression **FAILS** at forecasting daily returns (2011 data).

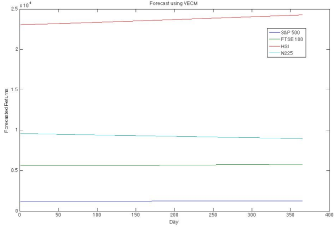
	S&P 500	FTSE 100	HSE	N225
MSE	0.0001	0.0001	0.0001	0.0001
MAPE	1.0175	1.3973	2.5325	1.0111

Table: Forecasting Accuracy: Market Index Returns (next day)

MAPE results suggest a deviation $O(100\%)$ to $O(200\%)$ per cent. Granted, daily returns for a broad market are a very small, close to negligible, quantity.

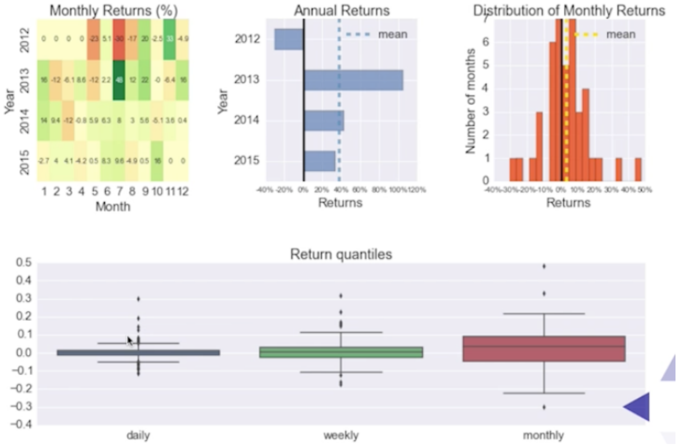
For properly stationary variables, Econometrics offers forecasting, impulse-response (IRF), and Granger causality analyses. **NONE** applicable to financial time series.





Without updating (recomputing) the regression, the forecast is a straight line.

Investment Performance



From: Quantopian Backtesting, T. Wiecki at QI2015

Vector Autoregression

VAR(p) is the structural equation model of *seemingly unrelated regressors*:

$$y_{1,t} = \beta_{1,0} + \beta_{11} y_{1,t-1} + \beta_{12} y_{2,t-1} + \dots \beta_{1n} y_{n,t-1} + \dots_{t-2} \dots + \epsilon_{1,t}$$

by OLS

$$y_{2,t} = \beta_{2,0} + \beta_{21} y_{1,t-1} + \beta_{22} y_{2,t-1} + \dots \beta_{2n} y_{n,t-1} + \dots_{t-2} \dots + \epsilon_{2,t}$$

by OLS

$$\dots \qquad \dots$$

$$y_{n,t} = \beta_{n,0} + \beta_{n1} y_{1,t-1} + \beta_{nn} y_{2,t-1} + \dots \beta_{nn} y_{n,t-1} + \dots_{t-2} \dots + \epsilon_{n,t}$$

Instead of estimating by OLS line-by-line, all beta coefficients can be computed in concise form, **in one go**.

Dependent Matrix

$p=3$

1. *Dependent matrix* has *observations* for p lags removed. Time series in rows, $p+1$ to the most recent observation at T.

row-wise $\longrightarrow$

$$Y = [\mathbf{y}_{p+1}\ \mathbf{y}_{p+2} \cdots \mathbf{y}_T] = \begin{array}{c} \text{CA Rets} \\ \text{FR Rets} \\ \\ \\ \end{array} \begin{pmatrix} y_{1,p+1} & y_{1,p+2} & \cdots & y_{1,T} \\ y_{2,p+1} & y_{2,p+2} & \cdots & y_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ y_{n,p+1} & y_{n,p+2} & \cdots & y_{n,T} \end{pmatrix}$$

$$\left[\cancel{y_{1,t=1}} \quad \cancel{y_{1,\ldots}} \quad \cancel{y_{1,p}} \quad y_{1,p+1} \quad y_{1,p+2} \cdots \quad y_{1,t=T}\right]$$ ROW-WISE

For **lag** $p=3$, we use the first three values to predict y_{p+1} (and so on).

$$n = Nvar \qquad\qquad T = Nobs$$

3. *Explanatory data matrix (assume p=3)*

$$
Z = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{y}_p & \mathbf{y}_{p+1} & \cdots & \mathbf{y}_{T-1} \\ \mathbf{y}_{p-1} & \mathbf{y}_p & \cdots & \mathbf{y}_{T-2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_{n,1} & \mathbf{y}_{n,2} & \cdots & \mathbf{y}_{n,T-p} \end{bmatrix}
=
\begin{array}{r}
\\ \text{CA ret} \rightarrow \\ \text{FR ret} \rightarrow \\ \\ \\ \\ \text{CA ret} \rightarrow \\ \text{FR ret} \rightarrow \\ \\ \\ \\ \text{CA rets} \rightarrow \\ \text{FR rets} \rightarrow \\ \\ \\
\end{array}
\begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_{1,p=3} & y_{1,p+1} & \cdots & y_{1,T-1} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n=Nvar,p} & y_{n,p+1} & \cdots & y_{n,T-1} \\ \\ y_{1,p-1=2} & y_{1,p} & \cdots & y_{1,T-2} \\ y_{2,p-1} & y_{2,p} & \cdots & y_{2,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n=Nvar,p-1} & y_{n,p} & \cdots & y_{n,T-2} \\ \\ \overset{1}{y_{1,1}} & y_{1,2} & \cdots & \overset{T-3}{y_{1,T-p}} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,T-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n=Nvar,1} & y_{n,2} & \cdots & y_{n,T-p} \end{bmatrix}
$$

Coded in *Matlab*, the algorithm forms the matrix from the top,

```
ymat = y(nlag+1:end,:)'; % Forming dependent matrix Y

zmat = [ones(1,nobs-nlag)]; % Forming explanatory matrix Z
for i=1:nlag
    zmat = [zmat; y(nlag-i+1:end-i,:)'];
end;
```

$$n = Nvar \qquad\qquad T = Nobs$$

Residuals

4. Disturbance matrix (innovations, residuals)

Row-wise

$$\epsilon = \begin{bmatrix} \epsilon_{p+1} & \epsilon_{p+2} & \cdots & \epsilon_T \end{bmatrix} = \begin{array}{r} \text{CA} \rightarrow \\ \text{FR} \rightarrow \\ \\ \\ \end{array} \begin{bmatrix} e_{1,p+1} & e_{1,p+2} & \cdots & e_{1,T} \\ e_{2,p+1} & e_{2,p+2} & \cdots & e_{2,T} \\ \vdots & \cdots & \ddots & \vdots \\ e_{n,p+1} & e_{n,p+2} & \cdots & e_{n,T} \end{bmatrix}$$

Each row of residuals matches variables $y_1, y_2, \dots, y_{n=Nvar}$ respectively. The most recent observation is at T.

Residuals are computed once we estimated $\hat{B}$

$$\hat{\epsilon} = \overset{\text{dep.}}{Y} - \underset{?}{\hat{B}} Z \leftarrow \text{expl.}$$

Calculating VAR(p) Estimates

- Calculate the multivariate OLS estimator for *the coefficients*

$$\boxed{\hat{B} = YZ'(ZZ')^{-1}}$$

This estimator is consistent and asymptotically efficient.

- For the simple case of variables x and y, regression coefficients estimated with

$$\beta_1 = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2} \quad \text{and} \quad \beta_0 = \bar{y} - \beta_1 \bar{x}$$

A bit of MLE

Consider the Log-likelihood function for multivariate Normal

$$
\begin{aligned}
L &= \underbrace{\prod_t^T N(y_t, x_t, \beta, \sigma^2)} = (2\pi\sigma^2)^{-T/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) \\
\ln L &= -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \left(\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right)
\end{aligned}
$$

To maximise the Log-Likelihood *by varying* β

$$
\begin{aligned}
\frac{\partial \ln L}{\partial \beta} &= \frac{1}{\sigma^2}(Y - X\beta)'X = 0 \\
\hat{\beta} &= \underline{YX'(XX')^{-1}}.
\end{aligned}
$$

This is how $\hat{B} = \underline{YZ'(ZZ')}^{-1}$ result was obtained.

1. Estimator of the *residual covariance matrix* with $T \equiv N_{obs}$

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t'$$

row-wise

2. Standard errors of beta coefficients will be inside the inverse of Information Matrix (on the diagnoal)

$$\mathbb{C}ov\left[Vec(\hat{B})\right] = (ZZ')^{-1} \otimes \hat{\Sigma} \quad = \quad I^{-1}$$

Explanatory

Residual Covariance

$\otimes$ is the *Kronecker product*.

$$I = \frac{\partial^2 L}{\partial \beta \, \partial \beta}$$

Hessian
(matrix form derivative)
2nd order

VAR(1) Estimation

$N_{vars} = 6$

$d.f. = 6 - 1 = 5$

$t\ stat = \frac{\beta}{\sigma_\beta}$

t dist. tables → C.V. at 2.5%

p-value

		Const	Canada(-1)	France(-1)	Germany(-1)	Japan(-1)	UK(-1)	US(-1)
Canada	Estimates	0.0002	0.0489	0.0164	-0.0343	-0.0165	-0.0017	0.1113
	Std err	0.0002	0.0273	0.0234	0.0198	0.0136	0.0276	0.0240
	t-stats	1.0954	1.7939	0.7020	-1.7339	-1.2158	-0.0600	4.6467
France	Estimates	0.0000	0.0434	-0.0899	0.0235	-0.0424	-0.0960	0.4545
	Std err	0.0003	0.0390	0.0335	0.0283	0.0194	0.0395	0.0343
	t-stats	0.1781	1.1128	-2.6859	0.8313	-2.1817	-2.4331	13.2627
Germany	Estimates	0.0002	0.0256	0.0826	-0.1930	-0.0632	-0.0091	0.4392
	Std err	0.0003	0.0422	0.0362	0.0306	0.0210	0.0427	0.0371
	t-stats	0.5438	0.6059	2.2809	-6.3110	-3.0094	-0.2133	11.8475
Japan	Estimates	-0.0004	0.0556	0.0921	0.0140	-0.0888	0.0535	0.3079
	Std err	0.0003	0.0378	0.0325	0.0274	0.0188	0.0383	0.0333
	t-stats	-1.7341	1.4690	2.8349	0.5091	-4.7149	1.3974	9.2589
UK	Estimates	0.0000	0.0146	-0.0427	-0.0069	-0.0477	-0.0779	0.3774
	Std err	0.0002	0.0301	0.0259	0.0218	0.0150	0.0305	0.0265
	t-stats	0.1620	0.4853	-1.6524	-0.3155	-3.1786	-2.5537	14.2523
US	Estimates	0.0003	-0.0098	0.0217	-0.0010	-0.0246	0.0024	0.0068
	Std err	0.0002	0.0315	0.0270	0.0229	0.0157	0.0319	0.0277
	t-stats	1.4256	-0.3105	0.8013	-0.0446	-1.5690	0.0766	0.2472

Residual Covariance Matrix

	Canada	France	Germany	Japan	UK	US
Canada	100%	42%	46%	14%	42%	69%
France	42%	100%	75%	15%	75%	46%
Germany	46%	75%	100%	16%	67%	51%
Japan	14%	15%	16%	100%	17%	10%
UK	42%	75%	67%	17%	100%	45%
US	69%	46%	51%	10%	45%	100%

- since our residuals $\sim N(0, \sigma^2)$ this is also correlation.
- notice the correlation for US/Canada and UK/France, UK/Germany pairs. That hints at **collinearity**, a difficulty to separate.

Optimal Lag Selection

Optimal Lag p is determined by the lowest values of AIC, BIC statistics, constructed using the penalised likelihood principle.

- *Akaike Information Criterion*

det. of cov. matrix

$$AIC = \log|\widehat{\Sigma}| + \frac{2k'}{T}$$

$p=1$

$p=2$

- *Bayesian Information Criterion* (also Schwarz Criterion)

$$SC = \log|\widehat{\Sigma}| + \frac{k'}{T}\log(T)$$

$k' = n \times (n \times p + 1)$ is the total number of coefficients in VAR(p)

$|\widehat{\Sigma}|$ is the determinant of the residual covariance matrix

Example: Optimal Lag Selection

Lag	AIC	SC (BIC)
1	-38.9814	-38.8886
2	-38.9727	-38.8003
3	-38.9736	-38.7217
4	-38.954	-38.6225
5	-38.9434	-38.5324
6	-38.9173	-38.4266
7	-38.8996	-38.3294
8	-38.8817	-38.2319
9	-38.8577	-38.1284
10	-38.8364	-38.0275

largest by modulus

Stability Condition

It requires for the eigenvalues of each relationship matrix A_p to be inside the unit circle (< 1).

Eigenvalue	Modulus < 1
-0.22	0.22
-0.17	0.17
-0.01-0.11i	0.11
-0.01+0.11i	0.11
0.04	0.04
-0.01	0.01

β coeffs for lag 1

A $\quad n \times n$

This VAR system satisfies stability condition $|\lambda\mathbf{I} - \mathbf{A}| = 0$.

If $p > 1$, coefficient matrix for each lag A_p to be checked separately.

Multivariate Cointegration Analysis

Johansen Procedure can be used as a powerful **pre-screening** tool.

Consider our cointegration cases:

- Global market indicies, such as FTSE vs DAX – cointegration transpires over the 15-20 year period – even as daily Prices used.

- Cointegration lecture focuses on different sections of the yield curve, such as r_{10Y} and r_{25Y}, May 2013-May 2015

- Present within the segments of the commodities market, eg, heating oil vs. gas.

T_1, T_2, T_3

May June

Moving onto **cointegration in equities** (daily Prices)

- Learning Cointegration article (Appendix B) has APPL vs GOOG and AMZN vs EBAY for Feb 2012 - Feb 2013. AMZN vs EBAY had a reverting spread, up to 8-10 trades.

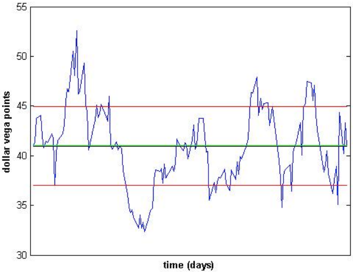
- Webex illustration (forthcoming) shows Ford vs. GM for 2011-2015 (full years). Cointegrated but spread ‘smoothed’, trades taking to 6M time.

- Marriott vs IHG was **a special situation** Jan 2014 - Jan 2017, when Marrott was looking for an acquisition. More in *Cointegration Lecture*.

Cointegrated System

Prices move together in long term = stationary spread.

$$e_t = \underbrace{\beta'_{\boldsymbol{Coint}} \boldsymbol{P_t}} = P_t^A + \beta_B P_t^B + \cdots + \beta_G P_t^G$$



Cointegration reduces several observed Prices to one common factor: spread e_t. There is **long/short** position with weights $[1, \pm\beta_B, \dots, \pm\beta_G]$.

Vector Autoregression reincarnated

Vector Autoregression is a model for **Returns**. Can't run on Prices except for special cointegrated case:

$\boldsymbol{\Delta P_t}$ have small but SIGNIFICANT correction to the long run spread.

$$
\begin{aligned}
\boldsymbol{\Delta P_t} &= \boldsymbol{\Pi\, P_{t-1}} + \boldsymbol{\Gamma_1 \Delta P_{t-1}} + \boldsymbol{\epsilon_t} \\
&= \alpha\beta'_{\boldsymbol{C}}\, \boldsymbol{P_{t-1}} + \boldsymbol{\Gamma_1 \Delta P_{t-1}} + \boldsymbol{\epsilon_t} \\
&= \alpha(\underbrace{\beta'_{\boldsymbol{C}}\, \boldsymbol{P_{t-1}}} + \mu_{\boldsymbol{e}}) + \boldsymbol{\Gamma_1 \Delta P_{t-1}} + \boldsymbol{\epsilon_t}
\end{aligned}
$$

$\boldsymbol{\Pi}$ is matrix of regression coefficients (similar to $\hat{\boldsymbol{B}}$ for returns).

Cointegration Rank

The matrix $\mathbf{\Pi}$ **must have a reduced rank**, otherwise the stationary differences $\mathbf{\Delta P}_t$ equate to non-stationary price levels $\boldsymbol{P}_t$ on *rhs*.

$$\mathbf{\Delta P}_t = \mathbf{\Pi}\, \boldsymbol{P}_{t-1} + \mathbf{\Gamma}_1 \mathbf{\Delta P}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\begin{aligned} \mathbf{\Pi} &= \alpha\beta' \\ (n \times n) &= (n \times r) \times (r \times \underbrace{[r + (n - r)]}) \end{aligned}$$

- r columns of β are cointegrating vectors
- $n - r$ columns are linearly dependent on r. They just represent the common stochastic trends of the system.

Cointegrating Vector Estimators β'_{Coint}

	1	2	3	4	5	6	7
Canada	6.78395	-1.96320	-9.07554	7.03629	2.56142	6.25519	-2.08045
France	4.86921	4.86043	-2.08623	-7.28739	2.28808	-1.59825	-1.60875
Germany	-15.76001	-5.94947	0.12170	3.34469	-0.01972	-4.04040	4.24522
Japan	-1.22250	5.52024	-0.70856	1.03285	-0.17938	-0.08242	1.76463
UK	27.19903	-13.06796	-0.55980	-0.36245	-1.03954	-1.76308	0.23821
US	-10.25644	13.17254	7.00734	-0.56186	-5.15207	2.16214	-2.37646
Const	-117.01015	-5.47002	59.45116	-32.77753	5.05186	-8.11528	-7.19582

Take the first column and standardise it.

$$\begin{bmatrix} 1 & 0.7178 & -2.3231 & -0.1802 & 4.0093 & -1.5119 & -17.2481 \end{bmatrix}$$

The allocations $\hat{\beta}'_{Coint}$ provide a mean-reverting spread.

Sequential Testing for Cointegration Rank

Trace Statistic and Maximum Eigenvalue tests rely on eigenvalues of $\mathbf{\Pi}$.

r	lambda	1-lambda	ln(1-lambda)	Trace	CV trace	MaxEig	CV MaxEig
0	0.0167	0.9833	-0.0168	105.7518	103.8473	44.8038	40.9568
1	0.0094	0.9906	-0.0094	60.9479	76.9728	25.1283	34.8059
2	0.0046	0.9954	-0.0046	35.8197	54.0790	12.3440	28.5881
3	0.0038	0.9962	-0.0038	23.4757	35.1928	10.2469	22.2996
4	0.0031	0.9969	-0.0031	13.2287	20.2618	8.3510	15.8921
5	0.0018	0.9982	-0.0018	4.8777	9.1645	4.8777	9.1645

- Trace statistic $H_0 : r = r^*$, and $H_1 : r > r^*$. **Table above $r^* = 1$**

$$LR_{r^*} = -T \sum_{i=r^*+1}^{n} \ln(1 - \lambda_i)$$

- Maximum eigenvalue statistic $H_0 : r = r^*$, and $H_1 : r = r^* + 1$

$$LR_{r^*} = -T \ln(1 - \lambda_{r^*+1})$$

Cointegration Analaysis – Johansen Procedure can be used as a powerful **pre-screening** tool.

The workhorse is *ca.jo()* function from the **R package** *urca*.

3 relationships

```
          test 10pct  5pct  1pct
r <= 6 |  4.67  7.52  9.24 12.97
r <= 5 |  5.87 13.75 15.67 20.20
r <= 4 |  9.78 19.77 22.00 26.81
r <= 3 | 24.98 25.56 28.14 33.24
r <= 2 | 44.91 31.66 34.40 39.79
r <= 1 | 46.88 37.45 40.30 46.82
r = 0  |101.10 43.25 46.45 51.91
```

c.v.

cajorls() presents the output as a set of familiar OLS equations with EC term, separate line for each price.

4x futures

Prices

Estimating Cointegration - Pairwise

Pairwise Estimation: select two prices likely to have a stationary spread: gas vs. heating oil futures, two pharmas or hotel chains where one interested in a merger.

- **Step 1.** Regress one price P_t^A on another P_t^B, and test the fitted residual by ADF with lag=1. If stationary, proceed.
- **Step 2.** Confirm significance of correction term in the eqns for ΔP_t^A, ΔP_t^B.
- Step 3. Fit the stationary spread to OU SDE solution (by autoregression) to evaluate mean-reversion: μ_e, θ, σ_{eq}.

There will be *CQF Lecture on Cointegration*, but R code and Case Study available now in Additional Files.

Engle-Granger Procedure: One Slide

Step 1. Obtain the fitted residual and ADF-test for stationarity.

$$\widehat{e}_t = P_t^A - \widehat{\beta}_C P_t^B - \widehat{\mu_e}$$

Price A Price B

- Cointegrating vector $\boldsymbol{\beta}'_{Coint} = [1, -\widehat{\beta}_C]$ and equilibrium level is $\mathbb{E}[\widehat{e}_t] = \mu_e$
- **If the residual non-stationary** then no long-run relationship exists and regression is spurious.

Step 2. Plug the residual from Step 1 into the **error correction** equation in order to confirm the statistical significance of coefficients

$$\Delta P_t^A = \phi \Delta P_t^B - (1 - \alpha) \underbrace{\left(P_{t-1}^A - \beta_C P_{t-1}^B - \mu_e\right)}$$

- It is required **to confirm the significance for** $(1 - \alpha)$ coefficient.

Relevant Econometric Advances

In addition to *urca* library, explore

`cran.r-project.org/web/views/Econometrics.html`

1. Estimation of regression adaptively (via a state-space model known as Kalman filter) removes the need for rolling parameters

   `www.thealgoengineer.com/2014/online_linear_regression_kalman_filter/`

   Explore if it makes sense (a) to estimate cointegrating eqn with Kalman filter recursively vs. (b) to improve price or return prediction by applying Kalman filter directly to VAR or equation for price itself as an OU process.

2. Overview of robust statistical methods in R

   `cran.r-project.org/web/views/Robust.html`

Statistical Arbitrage

- Signal Generation from Cointegrated Spread
- Fitting to OU Process: how good the mean-reversion is
- Trade Bounds Optimisation and Backtesting. EXTRA

Statistical Arbitrage

Cointegrated prices have a mean-reverting spread $e_t = \beta'_{Coint}\boldsymbol{P}_t$, when it goes *significantly* above/below μ_e, it gives a signal.

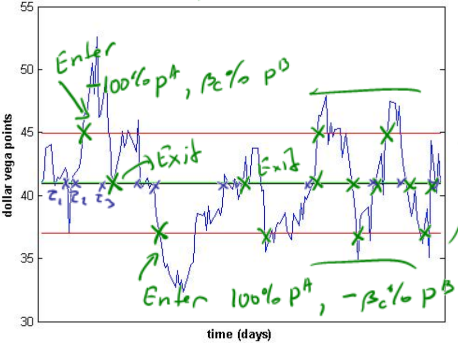
1. **How to generate P&L?** Trade design and algorithmic considerations.

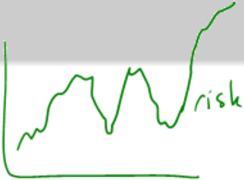
2. **How to evaluate P&L?** Drawdown control and backtesting.

An example of signal generation from e_t crossing $\mu_e + \sigma_{eq}$ or e_t crossing $\mu_e - \sigma_{eq}$

Signals from Mean-Reverting Spread

$e_t = -P_t^A + \beta_c P_t^B$





$\langle \tau \rangle$ is half-life

$\mu_e + Z^* \sigma_{eq}$

$\mu_e - Z^* \sigma_{eq}$

DF test

$Z^*_{optimal} = 0.7$?

Signal generation and positions for assets A and B.

$e_t \gg \mu_e$ enter with $[-100\%\, P^A, +\beta_C\%\, P^B]$

$e_t \ll \mu_e$ enter with $[100\%\, P^A, -\beta_C\%\, P^B]$

0.7, 0.8, 0.9

To make the trading systematic and controlled, you will need:

- **Loadings** β_{Coint} give positions, the spread is coint residual

$$e_t = P_t^A + \beta_B P_t^B + \cdots + \beta_G P_t^G$$

- **Bounds** $\mu_e \pm Z\,\sigma_{eq}$ give **entry** signal, while **exit** at $e_t \approx \mu_e$

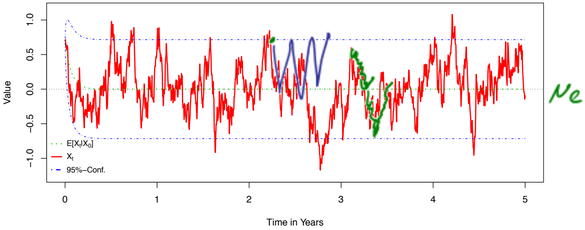
  Instead of assuming $Z = 1$ you can vary in the range $[0.7, 1.3]$ or as fitting to your spread.

- **Half-life** between the crossings $e_t = \mu_e$.

$$\left[\ \tilde{\tau} \propto \ln 2/\theta\ \right]$$

  Average time between exists that fix positive P&L.

OU Process simulated



μ_e

We consider the process because it generates **mean-reversion**.
Your empirical spread e_t might/might not be as good as this. σ

$$\left[de_t = -\theta\underbrace{(e_t - \mu_e)}_{\text{reversion}}\, dt + \sigma\, dX_t \right] \tag{1}$$

θ - speed of mean reversion

BM

Solved. diffusion

- $\boxed{\theta \ll 0}$ is the speed of reversion to the equilibrium μ_e σ_{eq}
- σ is the scatter of BM diffusion (not of reversion σ_{eq}).

Fitting to OU Process

SDE Solved

$$e_{t+\tau} = \underbrace{\boxed{\left(1 - e^{-\theta\tau}\right)\mu_e}}_{C} + \underbrace{\boxed{e^{-\theta\tau}}}_{B} e_t + \epsilon_{t,\tau}$$

Two terms of SDE solution: reversion and autoregression

$e_t = -P^A_t - \beta P^B_t$ ← residual   AR(1)   shift by lag 1

$$e_t = \boxed{C} + \boxed{B} e_{t-1} + \epsilon_{t,\tau}$$ run a regression

$$e^{-\theta\tau} = B \quad \Rightarrow \quad \boxed{\theta = -\frac{\ln B}{\tau}} \tag{2}$$

$$\left(1 - e^{-\theta\tau}\right)\mu_e = C \quad \Rightarrow \quad \boxed{\mu_e = \frac{C}{1 - B}} \tag{3}$$

Signal-generating Bounds

$e_t = C + B e_{t-1} + \epsilon$

$\mu_e \pm 0.7\sigma_{eq}$

$$\left[\ \sigma_{eq} = \sqrt{\frac{\cancel{\Sigma_\tau}\ RSE}{1 - e^{-2\theta\tau}}}\ \right] \qquad (4)$$

$z^* = \{0.7,\ 1,\ 1.2\}$

SSE is sum of squared residuals of your AR(1) regression for e_t. In one-variable case SSE represents residuals covariance matrix Σ_τ.

σ_{OU} is parameter of the SDE, Brownian Motion diffusion *over each small dt*. Not needed for trading *per se*.

$$\sigma_{OU} = \sigma_{eq}\sqrt{2\theta}$$

$$= \sqrt{\frac{2\theta\,\Sigma_\tau}{1 - e^{-2\theta\tau}}}.$$

$\sigma_{OU} = \sigma$ into SDE

OU Fit – Model Risk

IN PRACTICE we want to trade with tight bounds $Z < 1$ of the higher frequency spread.

$$\mu_e \pm Z\,\sigma_{eq}$$

For the largest profit per trade, typically $Z > 1.5$, the strategy is prone to the breakouts (partitioning of the coint relationship).

Ex ante testing for regime-change is of little help. Adaptive estimation with Kalman or other filtration means unwanted rebalancing, however.

You are constructing the model (cointegration) as much as you are testing for it. There are a number of ways where model not suitable, typically, (a) spread too tight, below bid/ask spread, and (b) OU process might not fit well.

Before we conclude, the words of wisdom from Fischer Black:

1. “In the real world of research, conventional tests of [statistical] significance seem almost worthless.”

2. “It is better to estimate a model than to test it. Best of all, though, is to explore a model.”

On model risk in time series from American Statistical Society:

1. “Running multiple tests on the same data set at the same stage of an analysis increases the chance of obtaining at least one invalid result.”

2. “Selecting one ‘significant’ result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion.’

Robust Portfolio Construction

Non-Robust Allocation

*Plugging historical mean and variance into a mean-variance optimizer and implementing its portfolio advice is a **terrible guide** to investing. Practically anything does better. $1/N$ does better.*

$\tau\Sigma$ error of estimation of μ

John Cochrane

$\operatorname{argmin} \mu w' - \frac{1}{2} w \Sigma w$

Sample covariance matrix contains estimation error of the kind most likely to perturb a mean-variance optimizer.

Ledoit & Wolf, 2003

μ

In MV, accurate mean estimates are about 20 times more important than covariance estimates.

Bill Ziemba (2016)

Taking a closer look at optimisation expression

^ means "sample estimate"

$$\underset{\boldsymbol{w}}{\operatorname{argmax}} \left\{ \boldsymbol{w}' \cancel{\hat{\boldsymbol{\mu}}} - \lambda \boldsymbol{w}' \boldsymbol{\Sigma} \boldsymbol{w} \right\}$$

μ_{BL}

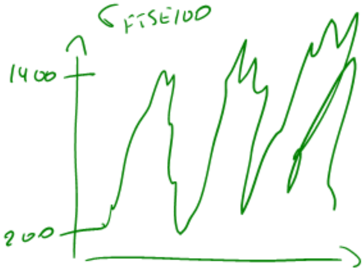
we can understand Black-Litterman insights,

- Expected returns are changed to equilibrium returns $\hat{\mu} \rightarrow \boldsymbol{\pi}$, which requires knowledge of index weightings.
- Equilibrium returns are updated with the views, posterior is computed from prior, $\boldsymbol{\pi} \rightarrow \boldsymbol{\mu}_{BL}$.
- $\tau\,\boldsymbol{\Sigma}$ gives standard error of the estimate, its uncertainty.
- Updated eq returns (the posterior) overshadow covariance input.

Historic variance plagued by **heteroskedasticity**: $\hat{\boldsymbol{S}}$ varies each sample to sample. FTSE100 variance varies $\times 7$ fold.

Empirical covariance have a signature of a random matrix.

Corner solution. Expected return dominates MV optimisation: invest 100% into the high risk/high return asset.



a. Apply shrinkage towards true estimates.

b. Devise BL views using ML classifiers.

c. Apply risk budgeting.

Covariance Shrinkage: 2003-2014 linear

$$\boldsymbol{\Sigma}_{Shrink} = (1 - \delta)\,\hat{\boldsymbol{S}} + \delta\,\boldsymbol{F}$$

For the structured estimator (skeleton variance) $\boldsymbol{F}$, a diagonal matrix of covariances (each asset vs S&500) is simple choice.

Min Covariance Determinant is covariance computed with Mahalanobis distance instead of squared difference, suitable for $\boldsymbol{F}$.

Instead of sample covariance (formula is MLE result from the joint Normal density), *sklearn.covariance.empirical_covariance()* is an alternative $\hat{\boldsymbol{S}}$.

Back to Black-Litterman:

Updating returns input into optimisation to be posterior $\pi \rightarrow \mu_{BL}$

Black-Litterman Model was a path-breaking invention. Allowed to blend views with the market-driven allocation.

- Source benchmarks and construct the prior, define input views, compute the posterior (expected returns). Perform optimisation.
- The prior distribution of *excess returns* $f_{\boldsymbol{\mu}}(\boldsymbol{\mu})$

Prior

$$\boldsymbol{\mu} \sim N(\underset{\text{eq. rets}}{\boldsymbol{\pi}}, \tau\underset{\text{error}^2}{\boldsymbol{\Sigma}}) \tag{5}$$

$\tau = 1/252 \approx 0.004$ gives low uncertainty on estimates $\tau\boldsymbol{\Sigma}$, and so we choose up to $\tau = 0.4$ to impose more uncertainty.

We start with ‘an oversimplified example’ from

The Black Litterman Approach guide by Attilio Meucci (2010):

Six market indicies of Italy, Spain, Switzerland, Canada, US and Germany with similar annualised historic volatilities

$$\boldsymbol{\sigma} = (21\%, 24\%, 24\%, 25\%, 29\%, 31\%)$$

with and correlations in the range $0.4 \leq \rho \leq 0.8$.

This is more informative than the covariance matrix, recovered as

$$\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\sigma})\,\mathbf{Corr}\,\mathrm{diag}(\boldsymbol{\sigma})$$

BL Inputs: expected returns

We take **market allocations $\tilde{\mathbf{w}}$** from some global benchmark

$$\tilde{w}' = (4\%,\ 4\%,\ 5\%,\ 8\%,\ 71\%,\ 8\%)$$

and calculate the equilibrium returns with risk-aversion $\lambda = 1.12$

$$\boldsymbol{\pi}' = (6\%,\ 7\%,\ 9\%,\ 8\%,\ 17\%,\ 10\%)$$

π is **the mean** of the reference distribution (**the prior**).

$$\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma})$$

How did we calculate $\tilde{\boldsymbol{w}} \to \pi$?

$$\frac{\partial}{\partial w}\left[w'\pi - \lambda w'\Sigma w\right]$$

$$\pi - 2\lambda \Sigma w = 0$$

$$\hookrightarrow \pi = \ldots$$

$$\underset{\boldsymbol{w}}{\operatorname{argmax}} \{\boldsymbol{w}'\pi - \lambda \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}\}$$

Solve **reverse optimisation** problem to obtain equilibrium returns π from market index allocations $\tilde{\boldsymbol{w}}$

$$\left[\pi = 2\lambda\boldsymbol{\Sigma}\tilde{\boldsymbol{w}}\right]$$

Direct optimisation gives solution for allocations $\boldsymbol{w}^*$,

$$\hookrightarrow w = \ldots$$

$$\boldsymbol{w}^* = \frac{1}{2\lambda}\boldsymbol{\Sigma}^{-1}\pi$$

Choice of the prior

How do we choose a prior? Say, we invest in Emerging Markets.

- **Do not** naively optimise the mean-variance over estimates calculated from a sample.
- Form equilibrium allocations π using the weights of the relevant MSCI capped index $\tilde{\boldsymbol{w}}$.
- Calculating weights by market cap is possible but gives a bias. If not benchmarked, one can start with allocations $\tilde{\boldsymbol{w}} = 1/N$.

Devise a suitable prior by referring to weights in specialised indicies (ready multi-asset indicies are rare).

- MSCI, S&P Dow Jones, FTSE for equities. Barclays Capital Aggregate for bonds. Markit iTraxx, SovX and CDX North America for credit. Sources $\tilde{w}$

Equilibrium returns π anchor the expectations and determine a vast majority of optimal portfolio allocations. This is an empirical finding.

Overweight and underweight the market

Findings and Analysis section of the project report must give a comparison as follows (see Table 6 in BL Guide by T. Idzorek):

- equilibrium allocations for the posterior vs. the prior ($\boldsymbol{\mu}_{BL} - \boldsymbol{\pi}$)
- devised optimal allocations and market index weights ($\mathbf{w}^* - \tilde{\mathbf{w}}$)

[Active Risk] You should only hold something different than market weights if you are identifiably different than the market average investor.

John Cochrane

Views Example

Too many views will pull the prior in multiple ways.

- the market in Spain will rise 12% (*absolute view*)
- the spread US-Germany will drop 10% (*relative view*)
  Germany outperforms the US

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

pick matrix

$$\boldsymbol{v} = \begin{pmatrix} 12\% \\ -10\% \end{pmatrix} \quad \textit{annualised}$$

views

The common notation is K views on N assets so, the pick matrix $\boldsymbol{P}$ has $K \times N$ dimensions.

Views Distribution

Definition: a view is a statement on the market (returns) that **potentially clashes** with the reference market distribution.

Views are expressed on expected returns $\boldsymbol{P\mu}$.

$$\boldsymbol{P\mu} \sim N(\boldsymbol{v}, \boldsymbol{\Omega}) \quad \leftarrow \text{dist of Views} \tag{6}$$

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \qquad \begin{pmatrix} \mu_{IT} & \mu_{ES} & \mu_{CH} & \mu_{CA} & \mu_{US} & \mu_{DE} \end{pmatrix}$$

$$\boldsymbol{\Omega} = \operatorname{diag}\left(\boldsymbol{P}(\tau\boldsymbol{\Sigma})\boldsymbol{P}'\right) \tag{7}$$

with $\tau \in [0.004, 0.4]$, the **uncertainty on views $\boldsymbol{\Omega}$** is less than full covariance $\boldsymbol{\Sigma}$.

Updating Multivariate Density

Asset Returns

Using multivariate Normal density $\boldsymbol{\mu} \sim N(\boldsymbol{\pi}, \tau\boldsymbol{\Sigma})$ (matrix notation)

$$f_{\boldsymbol{\mu}}(\boldsymbol{\mu}) = \frac{|(\tau\boldsymbol{\Sigma})|^{-\frac{1}{2}}}{(2\pi)^{\frac{N}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}-\boldsymbol{\pi})'(\tau\boldsymbol{\Sigma})^{-1}(\boldsymbol{\mu}-\boldsymbol{\pi})\right)$$

pdf

$(x-\mu)$ $\quad \frac{1}{\tau\Sigma}$ $\quad (x-\mu)$

$e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

univariate Normal pdf

The views can be written $\boldsymbol{v} \overset{D}{=} \boldsymbol{P}\boldsymbol{\mu} + \epsilon$ where $\epsilon \sim N(0, \boldsymbol{\Omega})$.

$$\boldsymbol{V}|\boldsymbol{\mu} \sim N(\boldsymbol{P}\boldsymbol{\mu}, \boldsymbol{\Omega})$$

$$f_{\boldsymbol{V}|\boldsymbol{\mu}}(\boldsymbol{v}) = \frac{|\boldsymbol{\Omega}|^{-\frac{1}{2}}}{(2\pi)^{\frac{K}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{P}\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{v}-\boldsymbol{P}\boldsymbol{\mu})\right)$$

pdf

f_μ , $f_{V|\mu}$

We are interested in **the posterior** distribution of returns updated with the views $\mu|\boldsymbol{V}$

$$f_{V|\mu} = \frac{f_{\mu,V}}{f_\mu} \qquad f_{\mu,V} = f_{V|\mu} \times f_\mu$$

joint

$$f_{\mu|\boldsymbol{V}}(\boldsymbol{\mu}) = \frac{f_{\mu,\boldsymbol{V}}(\boldsymbol{\mu},\boldsymbol{V})}{f_{\boldsymbol{V}}(\boldsymbol{V})} = \frac{f_{\boldsymbol{V}|\mu}(\boldsymbol{V})\, f_\mu(\boldsymbol{\mu})}{\int f_{\boldsymbol{V}|\mu}(\boldsymbol{V})\, f_\mu(\boldsymbol{\mu})d\boldsymbol{\mu}} \tag{8}$$

assets conditioned on Views

By substituting the respective Normal *pdfs* into the Bayes formula and much tedious working, we will have a solution for **the posterior**.

$$f_{\mu|\boldsymbol{V}}(\boldsymbol{\mu}) \sim N(\boldsymbol{\mu}_{BL}, \boldsymbol{\Sigma}^{\mu}_{BL})$$

BL Solution for computation

Attillio Meucci (2010) suggested **computationally stable** formulation (less matrix inversions) than original BL.

$f_{\mu|v}$

$$\underline{\boldsymbol{\mu}_{BL}} = \boldsymbol{\pi} + \tau \boldsymbol{\Sigma} \boldsymbol{P}' \left(\tau \boldsymbol{P} \boldsymbol{\Sigma} \boldsymbol{P}' + \boldsymbol{\Omega}\right)^{-1} (\boldsymbol{v} - \boldsymbol{P}\boldsymbol{\pi}) \tag{9}$$

$p_v \ldots$

$$\underline{\cancel{\boldsymbol{\Sigma}}_{BL}} = \underbrace{(1+\tau)\boldsymbol{\Sigma}} - \tau^2 \boldsymbol{\Sigma} \boldsymbol{P}' \left(\tau \boldsymbol{P} \boldsymbol{\Sigma} \boldsymbol{P}' + \boldsymbol{\Omega}\right)^{-1} \boldsymbol{P} \boldsymbol{\Sigma} \tag{10}$$

$(1+\tau)\boldsymbol{\Sigma}$ is the most distorted covariance of the posterior when views are not informative.

Total covariance is also $\boldsymbol{\Sigma}_{BL} = \boxed{\boldsymbol{\Sigma}} + \boldsymbol{\Sigma}^{\mu}_{BL}$.

comes from shrinkage

Mean-Variance Optimisation

Obtain the allocations $\boldsymbol{w}^*$ by optimisation that uses the posterior.

Usually, the optimisation disregards $\boldsymbol{\Sigma}_{BL} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}^{\mu}_{BL}$. Why?

$$\underset{\boldsymbol{w}}{\operatorname{argmax}} \left\{ \boldsymbol{w}' \mu_{BL} - \lambda \boldsymbol{w}' \boldsymbol{\Sigma} \boldsymbol{w} \right\}$$

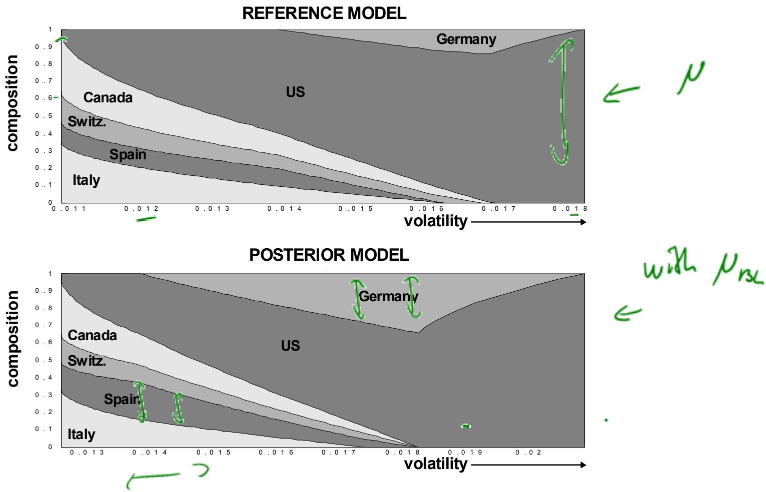
The analytical solution, no constraints is

robust shrunk Σ

$$\boldsymbol{w}^* = \frac{1}{2\lambda} \hat{\boldsymbol{\Sigma}}^{-1} \mu_{BL}$$

Unconstrained mean-variance has a convenient analytical solution. Constraints improve robustness but require numerical optimisation, *quadprog()*, Solver.

BL Outputs



From: *The Black Litterman Approach* guide by Attilio Meucci (2010).

Efficient Frontier

The above way of looking at the Efficient Frontier, via changing allocations, allows us to evaluate their robustness.

We aim for views to reduce ‘a corner solution’, investing all in the highest return/highest risk asset (the US index).

Compare allocations, the prior vs. posterior:

- allocation to Spain increases (up to certain volatility level),
- allocation to Germany increases (since US is expected to grow 10% less than Germany)

Risk Aversion

Risk aversion λ is a constant but **changing it affects allocations**. Study how allocations change for

λ_{kelly} + Σ robust

- $\lambda = 0.01/2$ for to a near-Kelly investor (many concentrated bets)
- $\lambda = 2.24/2$ as suggested by BL (average investor, market)
- $\lambda = 6/2$ for a risk-averse investor (trustee).

Why $/2$? The unmodified values are for optimisation over $\frac{1}{2}\boldsymbol{\lambda w'\Sigma w}$.

In these slides, $\boldsymbol{w}^* = \frac{1}{2\lambda}\,\hat{\boldsymbol{\Sigma}}^{-1}\,\mu_{BL}$ so use the modified values.

Alternative Optimisations: Max SR

1. Maximum Sharpe Ratio (Tangency Portfolio) explicitly operates with excess returns and a budget constraint

$$\underset{\boldsymbol{w}}{\text{argmax}} \qquad \frac{\boldsymbol{w}'\boldsymbol{\mu_{BL}} - r_f}{\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}}} = \frac{\mu_\Pi - r_f}{\sigma_\Pi}$$

$$\text{s.t.} \;\; \boldsymbol{w}'\mathbf{1} = 1$$

$$\boldsymbol{w}^* = \frac{\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1})}{\mathbf{1}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - r_f\mathbf{1})}$$

Max SR and Global Min Var are independent of risk aversion, but an implied quantity can be computed as $\lambda = \frac{\mu_{Pi}}{2\sigma^2_{Pi}}$.

Alternative Optimisations: Min VaR/ES

2. Minimum VaR, which is not quadratic programming

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} \left\{ -\boldsymbol{w}'\boldsymbol{\mu_{BL}} + \text{Factor} \times \sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}} \right\}$$

   Value at Risk factor is $\Phi^{-1}(1-c)$

   Expected Shortfall factor is $\frac{1}{1-c}\int_0^{1-c} \Phi^{-1}(\gamma)d\gamma$

3. Risk Budgeting offers superior way of incorporating VaR and other risk limits

$$\mathcal{RC}_i = w_i \left(-\mu_i + \text{Factor} \times \frac{(\boldsymbol{\Sigma}\boldsymbol{w})_i}{\sqrt{\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}}} \right)$$

$$\forall\, w_i \geq 0, \quad \sum w_i = 1$$

See Cases & Examples (Tab 3 - 2.19, Ex 11) *Risk Budgeting Elective*.

Markowitz MPT Model Risk

Unconstrained mean-variance optimisation is of little value.

There is also **estimation risk**: the use of naive estimators of expected risk and return (i.e., historical sample mean) in optimization generates the Efficient Frontier that is far from robust.

Known Market

Let's generate a sample of $N = 10$ assets with ranked volatilities

$$\boldsymbol{\sigma} = (5\%, 8.9\%, 12.8\%, 16.7\%, 20.6\%, 24.4\%, 28.3\%, 32.2\%, 36.1\%, 40\%)$$

Set correlation $\rho = 0.3$ for all assets, and introduce obtain expected returns as a proportion to each asset's volatility.

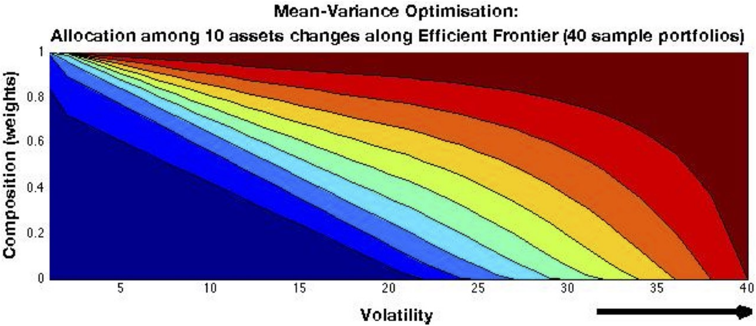
$$\boldsymbol{\mu} = (0.9\%, 1.6\%, 2.4\%, 3.3\%, 4.2\%, 5.2\%, 6.2\%, 7.25\%, 8.4\%, 9.55\%)$$

This is our **knowledge set** of the market. No estimation error.

Σ = nil

Unconstrained Mean-Variance

True



This is how allocations behave as acceptable portfolio risk increases: without constraints we end up investing in the most risky asset alone.

Sampling from mulativariate Normal

1. Using the **knowledge set** of our market (expected returns, covariance), we generate the multivariate Normal random values for time series with $T = 256$ observations

   This is simulated history of returns for correlated assets.

2. Then, estimate **sample covariance and mean** and optimise using these (instead of true parameters)

$$\boldsymbol{\Sigma} \rightarrow \hat{\boldsymbol{\Sigma}} \qquad \mu \rightarrow \hat{\mu}$$

$$\begin{array}{cc} A & B \\ \end{array}$$
$$\begin{pmatrix} 1 & 1 \\ \vdots & \vdots \\ 256 & 256 \end{pmatrix}$$

Estimation error looks very small, less than half of a per cent.

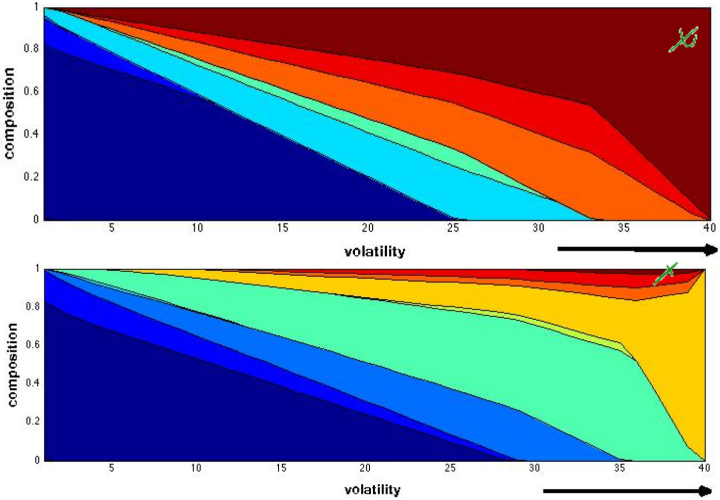
$$\tau = \frac{1}{T} = \frac{1}{256} = 0.0039 \approx 0.4\%$$

Mean-Variance with naive estimators

Both frontiers come form **the same knowledge** about the market.

$\hat{\Sigma}_1, \hat{\mu}_1$

$\hat{\Sigma}_2 \quad \hat{\mu}_2$



$$\arg\min_{w} = \mu w' - \frac{1}{2}\lambda w'\Sigma w - \text{skew} - \text{kurdosis}$$

The words of wisdom from Fischer Black:

"It is better to estimate a model than to test it. Best of all, though, is to explore a model."

Making Allocations Robust

- Covariance Shrinkage – practical techniques
- Quantitative Approach to BL Views

sklearn.covariance.LedoitWolf() performs Basic Shrinkage:

$$\boldsymbol{\Sigma}_{Shrink} = (1-\delta)\,\hat{\boldsymbol{S}} + \delta\,\frac{\mathrm{Tr}(\,\hat{\boldsymbol{S}}\,)}{n}\,\mathbf{1}$$

shrinkage towards the trace (sum of diagonal elements) of empirical covariance $\hat{\boldsymbol{S}}$ itself.

- Mathematically that reduces the ratio between the smallest and the largest eigenvalues.
- Improved Estimation of Covariance (LW 2003) implements this. **Code:** *L1_covMarket.m*, or *L1_ledoit_and_wolf_2001.py.*

`scikit-learn.org/stable/modules/covariance.html`

`http://mathworld.wolfram.com/MatrixTrace.html`

Covariance Shrinkage: 2017 non-linear, direct kernel

The updated shrinkage recipe is **non-linear**, with **direct kernel** for distribution of eigenvalues of a large matrix $n \times n$, **Goldilocks** for large n_{Assets} compared to N_{obs}.

- Advantages for portfolio construction: (1) large-dimensional covariance matrices, (2) implemented for Maximum Sharpe Ratio kind of optimisation.
- Winger's semicircle kernel chosen from 48 functions, has a closed-form Hilbert transform. Rotation occurs on the complex plane.

Suitable reading: *Nonlinear Shrinkage of the Covariance Matrix: Markowitz Meets Goldilocks* (LW 2017)

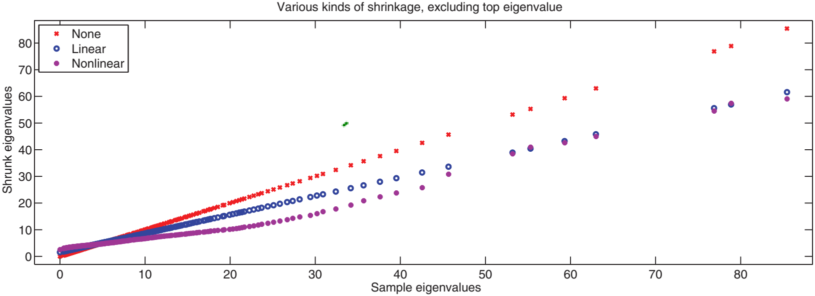
Understand Matlab code *direct_kernel.m* with Technical WP – each line of code has eqn reference. Python code *direct_kernel.py*.

Implementation Notes - Shrinkage

2001

```
>>> import numpy as np
>>> from sklearn.covariance import LedoitWolf
>>> real_cov = np.array([[.4, .2],
...                      [.2, .8]])
>>> np.random.seed(0)
>>> X = np.random.multivariate_normal(mean=[0, 0],
...                                   cov=real_cov,
...                                   size=50)
>>> cov = LedoitWolf().fit(X)
>>> cov.covariance_
array([[0.4406..., 0.1616...],
       [0.1616..., 0.8022...]])
>>> cov.location_
array([ 0.0595... , -0.0075...])
```

```
>>> import numpy as np
>>> from sklearn.covariance import MinCovDet
>>> from sklearn.datasets import make_gaussian_quantiles
>>> real_cov = np.array([[.8, .3],
...                      [.3, .4]])
>>> np.random.seed(0)
>>> X = np.random.multivariate_normal(mean=[0, 0],
...                                   cov=real_cov,
...                                   size=500)
>>> cov = MinCovDet(random_state=0).fit(X)
>>> cov.covariance_
array([[0.7411..., 0.2535...],
       [0.2535..., 0.3053...]])
>>> cov.location_
array([0.0813... , 0.0427...])
```



GARCH + Correlation

Alternative simple recipe to re-construct the covariance

$$\boldsymbol{\Sigma}_{ij} = \rho_{ij}\,\sigma_i^{GARCH}\,\sigma_j^{GARCH}$$

EGARCH equation estimated for each asset **individually**, from **log-returns**. The output is smoothed annualised vol.

$$\ln \sigma_t^2 = \alpha_0 + \frac{\alpha_1 u_{t-1} + \gamma_1 |u_{t-1}|}{\sigma_{t-1}} + \beta_1 \ln \sigma_{t-1}^2$$

- Note the feature $\gamma_1 |u_{t-1}|$ in the EGARCH above.
- Prediction much dependent on the long run average variance, GARCH result $\overline{\sigma}^2 = \frac{\omega}{1-\alpha-\beta}$.
- Consider using **EWMA** alternative with high δ. Why?

$$\sigma_t^2 = (1-\delta)u_{t-1}^2 + \delta\sigma_{t-1}^2$$

Implementation Notes - GARCH

- Create separate spreadsheet for each asset, like *ftse gjr.xlsx* from ARCH Lecture. Estimate by Solver.

- Matlab MFE Toolbox gives you control over econometric routines `https://www.kevinsheppard.com/MFE_Toolbox`. Plus teaching guides and exercises from Oxford MFE.

  Python version **for your simple ARCH/EWMA needs**, examples at `https://arch.readthedocs.io/en/latest/univariate/univariate_volatility_modeling.html`

- To understand GARCH you might choose to re-implement its MLE:

```python
# To fit the model first we define the objective function we seek to maximise
def log_likelihood(returns, mu, omega, alpha, beta):
    # Calculate excess returns
    hs, zs, sv = apply_garch(returns, mu, omega, alpha, beta)

    T = returns.size
    return -0.5*(T*np.log(2*np.pi) + np.sum(np.log(hs) + zs*zs))
```

Back to Black-Litterman:

Steps to quantify the views and make predictions

η_k of $\times\{-2,-1,0,1,2\}$ are five labels for ‘very bearish’, ‘bearish’, ‘neutral’, ‘bullish’, or ‘very bullish’ states Y for the view k.

Run multinomial logistic classifier (also NaiveBayes, SVM) on some past statistical data, which you might need to generate, especially dependent variable Y.

	prices	returns	ret_1	ret_2	ret_3	ret_4	ret_5	ols_pred	ols_returns	log_pred
Date										
2010-01-12	127.35	-0.022977	-0.024335	0.026717	-0.017160	-0.018282	0.005883	1.0	-0.022977	1.0
2010-01-13	129.11	0.013726	-0.022977	-0.024335	0.026717	-0.017160	-0.018282	1.0	0.013726	1.0
2010-01-14	127.35	-0.013726	0.013726	-0.022977	-0.024335	0.026717	-0.017160	-1.0	0.013726	1.0
2010-01-15	127.14	-0.001650	-0.013726	0.013726	-0.022977	-0.024335	0.026717	1.0	-0.001650	1.0
2010-01-19	127.61	0.003690	-0.001650	-0.013726	0.013726	-0.022977	-0.024335	1.0	0.003690	1.0

The above output is from CQF ML Elective

gist.github.com/yhilpisch/648565d3d5d70663b7dc418db1b81676

The scheme is for both, training and prediction:

$$\text{CI} \quad = \quad \text{Equilibrium Rtn} \quad \pm \quad \eta \times \text{Volatility Estimator}$$

1. Estimate past volatility on rolling basis and the asset move $\eta_t = \dfrac{r_{t,\text{Period}}}{\hat{\sigma}_t}$. That will give a series of past η_t for training.

2. Estimate recent robust volatility (forecast), to compute the view **once**.

Calibration and Data Analysis

In this lecture. . .

Part I

- the theoretical yield curve and the market yield curve, should they be the same?

market

v.s.

Theoretical

- how calibration works, the pros and cons

- how to analyze short-term interest rates to determine the best model

Part II

- how to analyze the slope of the yield curve to get information about the market price of interest rate risk

By the end of this lecture you will

- know the meaning of ‘calibration’

- appreciate the pros and cons of calibration

- be able to analyze data to find a good model

Paul's favourite

Introduction

We have seen the theory behind one-factor (and multi-factor) interest rate models.

These models will have as an output a 'theoretical' yield curve.

This theoretical yield curve will not be the same as the yield curve seen in the market.

Is this good or bad?!

1. Affine sol^n, e^{A-rB} (closed form)
2. PDE
3. M.C.

1 **It is bad if. . .** your job is to price exotic, structured products which must be hedged with simple instruments. (How can you be expected to price exotics 'correctly' if you can't even match simple instruments?!)

2 **It is good if. . .** you believe your theoretical model and you are looking for arbitrage opportunities among simple instruments. (If your model output matched market prices then you'll never find any arbitrage!)

So in using any model we have to decide how to choose the parameters.

Should the parameters be chosen to match

- the market yield curve? Or Calibration
- historical interest rate data? Data analysis

The former is **calibration** to a snapshot of the market at one instant in time. t^* today

The latter is fitting to time series data.

Let's start with calibration.

Calibration

1 Because of this need to correctly price liquid instruments, the idea of **yield curve fitting** or **calibration** has become popular.

When stochastic models are used in practice they are almost always fitted.

To match a theoretical yield curve to a market yield curve requires a model with enough degrees of freedom. (You are matching a curve, i.e. an 'infinite' number of points, so you need infinite degrees of freedom!)

2 This is done by making one or more 'parameters' time dependent.

3
- This functional dependence on time is then carefully chosen to make an output of the model, the price of zero-coupon bonds, exactly match the market prices for these instruments.

Ho & Lee

The Ho & Lee spot interest rate model is the *simplest that can be used to fit the yield curve*.

Now we don't necessarily say this model or idea is great, but we will go through the mathematics to see how it can be done.

Recap: In the Ho & Lee model the process for the *risk-neutral* spot rate is

$\eta \rightarrow \eta(t)$

$\eta_i = \eta(t_i)$

$$dr = \eta(t)dt + c\,dX.$$

const. vol.

The standard deviation of the spot rate process, c, is constant, the drift rate η is time dependent.

As we've seen in an earlier CQF lecture, for this model the solution of the bond pricing partial differential equation for a zero-coupon bond is simply

$$Z(r,t;T) = e^{A(t;T)-r(T-t)},$$

Affine soln

Take from market.

where

$$A(t;T) = -\int_t^T \eta(s)(T-s)\,ds + \frac{1}{6}c^2(T-t)^3.$$

(Note that the variables are r and t, but we are also explicitly referring to the parameter T, the bond maturity.)

Working forwards: If we know $\eta(t)$ then the above gives us the theoretical value of zero-coupon bonds of all maturities. I.e. start with model ($\eta(t)$) and find answer (Z).

An inverse problem: But what if we know Z from the market, but don't know the unobservable η? Turn this relationship around and ask the question

in terms of market bond price

$\eta(t) = - - - - - -$

- 'What functional form must we choose for $\eta(t)$ to make the theoretical value of the discount rates for all maturities equal to the market values?'

That is calibration!

historical time series / rates

(What about the parameter c?)

$t^* = 26/5/2020$ look at yield curve

Suppose we want to calibrate our model today, time t^*. Today's spot interest rate is r^* and the discount factors *in the market* are $Z_M(t^*; T)$.

Call the special, calibrated, choice for η, $\eta^*(t)$.

t^* — today

r^* — spot rate on date t^*

$\eta^*(t)$ — η using above data.

Assume cts dist^n of maturities.

To match the market and theoretical bond prices, we must solve

Market price of ZCB

$$Z_M(t^*;T) = \underbrace{e^{A(t^*;T)-r^*(T-t^*)}.}_{\text{Theoretical price.}}$$

Taking logarithms of this and rearranging slightly we get

$$\int_{t^*}^{T} \eta^*(s)(T-s)ds = -\log(Z_M(t^*;T)) - r^*(T-t^*) + \frac{1}{6}c^2(T-t^*)^3. \quad (1)$$

We know everything on the right-hand side. So this is an **integral equation** for $\eta^*(t)$.

(Luckily for us, it is quite easy to solve!)

Diff under the integral sign - Leibnitz Rule

Observe what happens if we differentiate the integral term with respect to T.

First differentiate once with respect to T

$$\frac{d}{dT}\int_{t^*}^{T}\eta^*(s)(T-s)ds = \int_{t^*}^{T}\eta^*(s)ds.$$

Differentiate again

$$\frac{d^2}{dT^2}\int_{t^*}^{T}\eta^*(s)(T-s)ds = \eta^*(T).$$

So, differentiating (1) twice with respect to T we get

$$\boxed{\eta^*(t) = c^2(t - t^*) - \frac{\partial^2}{\partial t^2}\log(Z_M(t^*; t)).}$$

The solution!

With this choice for the time-dependent parameter $\eta(t)$ the theoretical and actual market prices of zero-coupon bonds are the same.

Notes:

Complex

- Now that we know $\eta(t)$ we can price other fixed income instruments.

yield curve fitting

- We say that our prices are **consistent with the yield curve**.

- The same idea can be applied to other spot interest rate models.

- This is an inverse problem, and will typically be sensitive to input data (the Z).

Another calibrated model: The extended Vasicek model of Hull & White

Most one-factor models have the potential for fitting, the more tractable the model the easier the fitting. If the model is not at all tractable then we can always resort to numerical methods.

The next easiest model to fit is the Vasicek model. The Vasicek model has the following stochastic differential equation for the risk-neutral spot rate

$$dr = (\eta - \gamma r)dt + c\,dX.$$

Hull & White extend this to include a time-dependent parameter

$$dr = (\eta(t) - \gamma r)dt + c\,dX.$$

Assuming that γ and c have been estimated statistically, say, we choose $\eta = \eta^*(t)$ at time t^* so that our theoretical and the market prices of bonds coincide.

keep $\gamma, c \in \mathbb{R}$

$\eta \longrightarrow \eta(t)$

Ex: Solve the B.P.E for this model

Just $V = e^{A - rB}$

Again, as covered in an earlier CQF lecture, under this risk-neutral process the value of a zero-coupon bond is

$$Z(r,t;T) = e^{A(t;T) - rB(t;T)},$$

where

$$A(t;T) = -\int_t^T \eta(s)B(s;T)ds$$

$$+\frac{c^2}{2\gamma^2}\left(T - t + \frac{2}{\gamma}e^{-\gamma(T-t)} - \frac{1}{2\gamma}e^{-2\gamma(T-t)} - \frac{3}{2\gamma}\right).$$

and

$$B(t;T) = \frac{1}{\gamma}\left(1 - e^{-\gamma(T-t)}\right).$$

As before, to fit the yield curve at time t^* we must make $\eta^*(t)$ satisfy

Use Leibnitz

$$-\int_{t^*}^{T} \eta^*(s)B(s;T)ds$$
$$+\frac{c^2}{2\gamma^2}\left(T-t^*+\frac{2}{\gamma}e^{-\gamma(T-t^*)}-\frac{1}{2\gamma}e^{-2\gamma(T-t^*)}-\frac{3}{2\gamma}\right)$$
$$=\log(Z_M(t^*;T))+r^*B(t^*,T). \quad (2)$$

This is an integral equation for $\eta^*(t)$ if we are given all of the other parameters and functions, such as the market prices of bonds, $Z_M(t^*;T)$.

Equation (2) is easy to solve by differentiating the equation twice with respect to T. This gives

$$\eta^*(t) = -\frac{\partial^2}{\partial t^2}\log(Z_M(t^*;t)) - \gamma\frac{\partial}{\partial t}\log(Z_M(t^*;t)) + \frac{c^2}{2\gamma}\left(1 - e^{-2\gamma(t-t^*)}\right). \tag{3}$$

Note:

(Please don't get the idea from this that all models are easy to calibrate or all integral equations are easy to solve!)

Back to Ho & Lee:
Calibration in practice

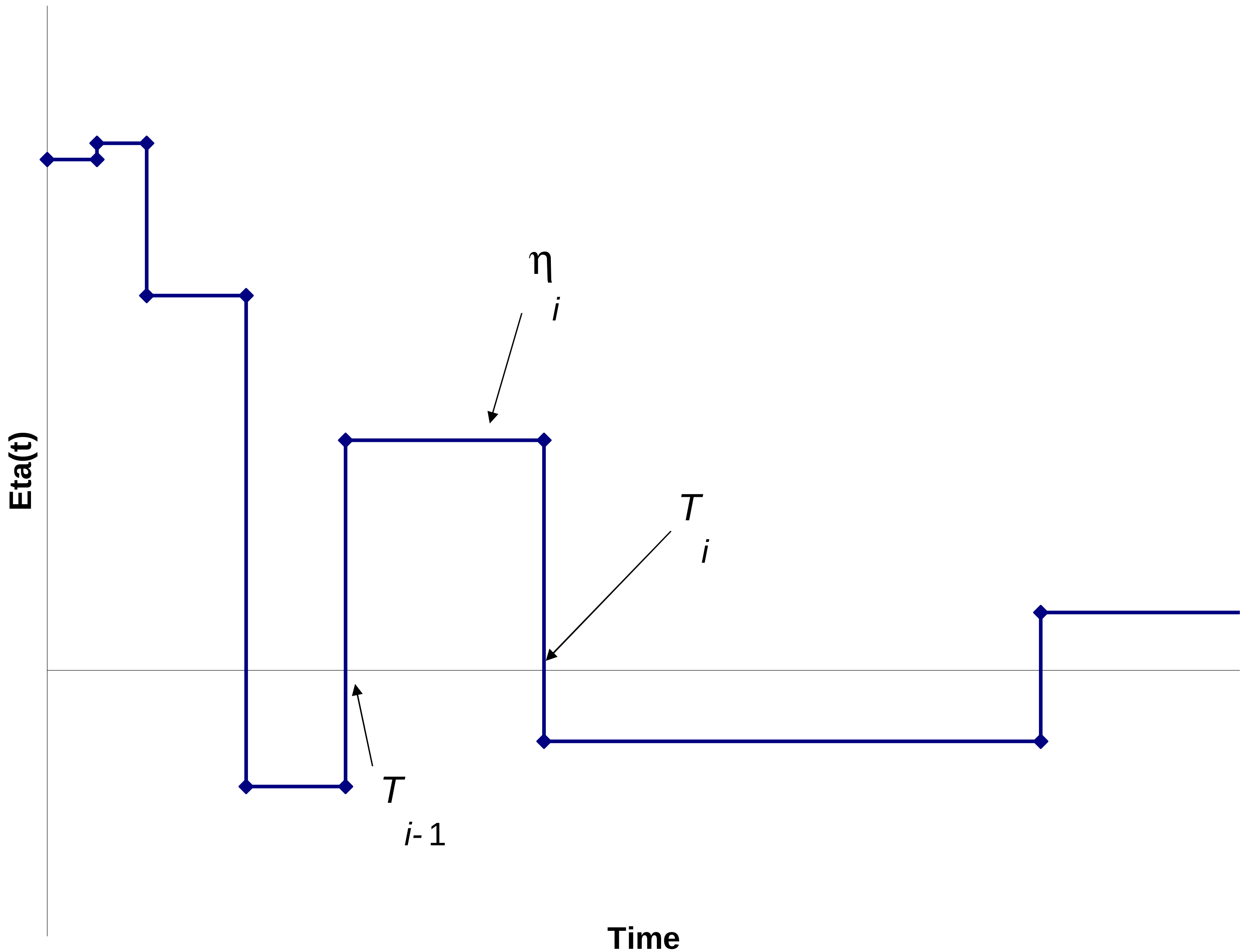
In practice we do not have a differentiable yield curve, instead we have a finite number of bond prices. Just some dots, one per maturity.

To get a unique 'solution' for $\eta(t)$ we have to make some assumptions about its structure.

Examples: $\eta(t)$ is piecewise constant; $\eta(t)$ is piecewise linear and continuous; etc.

Let's assume that $\eta(t)$ is piecewise constant. It's the easiest to do the maths for!

$\eta(t)$



Put $t^* = 0$

$$\int_0^T \eta^*(s)(T-s)ds = \underbrace{-\log(Z_M(0;T)) - r^*T + \tfrac{1}{6}c^2T^3}_{\text{Known}}.$$

First η_1:

1st maturity

$$\int_0^{T_1} \eta_1(T_1 - s)ds = \tfrac{1}{2}T_1^2\eta_1 = -\log(Z_M(0;T_1)) - r^*T_1 + \tfrac{1}{6}c^2T_1^3.$$

Therefore

rearranging

$$\boxed{\eta_1 = \frac{2}{T_1^2}\left(-\log(Z_M(0;T_1)) - r^*T_1 + \tfrac{1}{6}c^2T_1^3\right).}$$

Next η_2: $$\int_0^{T_2} = \int_0^{T_1} + \int_{T_1}^{T_2}$$

$$\int_0^{T_1} \eta_1(T_2 - s)ds + \int_{T_1}^{T_2} \eta_2(T_2 - s)ds = \underbrace{-\log(Z_M(0;T_2)) - r^* T_2 + \tfrac{1}{6}c^2 T_2^3.}$$

$$= \tfrac{1}{2}\eta_1 \left(T_2^2 - (T_2 - T_1)^2\right) + \tfrac{1}{2}\eta_2 (T_2 - T_1)^2.$$

Therefore

$$\eta_2 = \frac{2}{(T_2 - T_1)^2}\left(-\log(Z_M(0;T_2)) - r^* T_2 + \tfrac{1}{6}c^2 T_2^3 \right.$$
$$\left. -\tfrac{1}{2}\eta_1 \left(T_2^2 - (T_2 - T_1)^2\right)\right).$$

Generally, the left-hand side of equation (1) becomes

$$\sum_{i=1}^{j} \eta_i \int_{T_{i-1}}^{T_i} \left(T_j - s\right)\, ds$$

$$= \frac{1}{2} \sum_{i=1}^{j} \eta_i \left(\left(T_j - T_{i-1}\right)^2 - \left(T_j - T_i\right)^2 \right).$$

$$= \frac{1}{2}\eta_j \left(T_j - T_{j-1}\right)^2 + \frac{1}{2} \sum_{i=1}^{j-1} \eta_i \left(\left(T_j - T_{i-1}\right)^2 - \left(T_j - T_i\right)^2 \right).$$

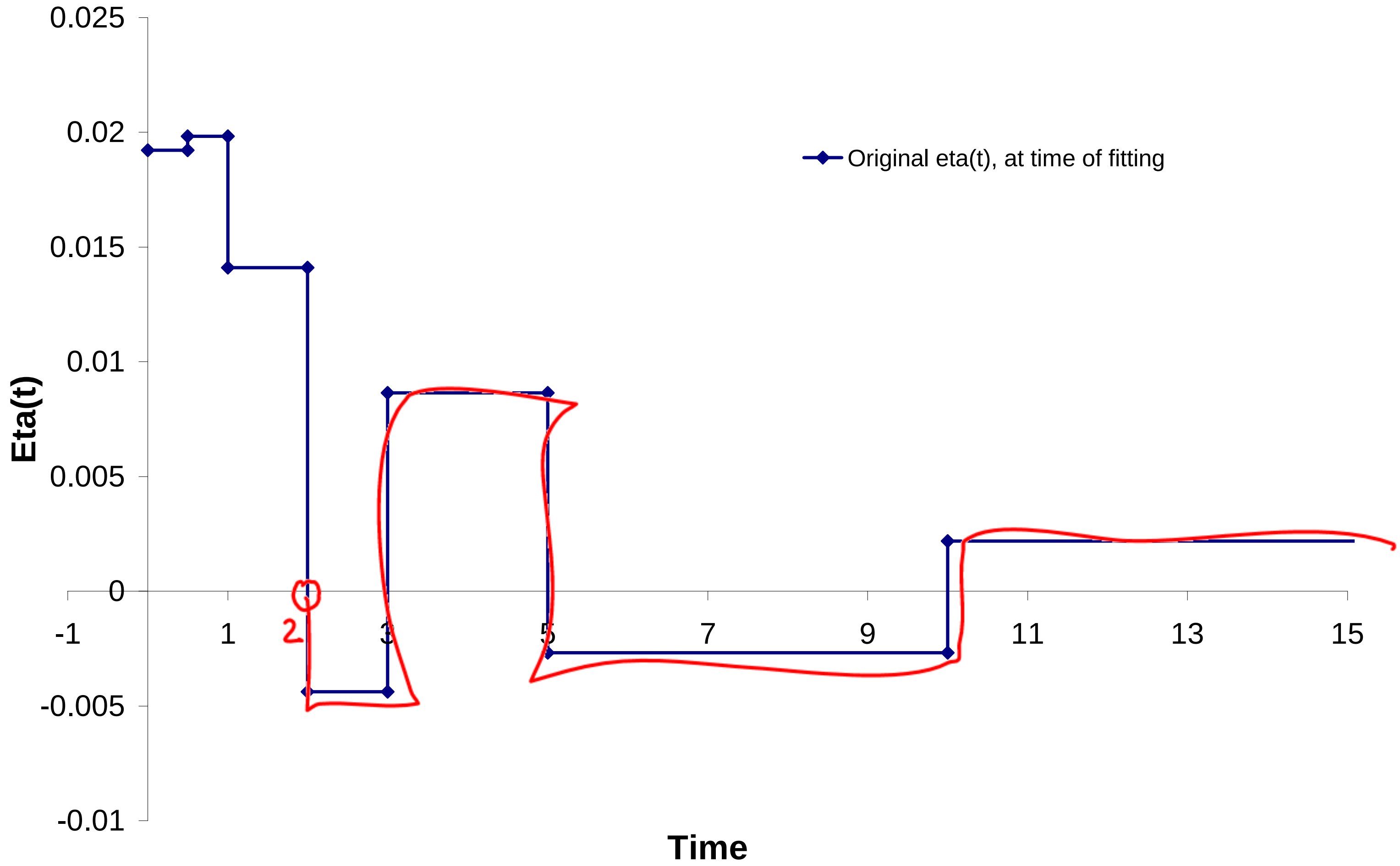
And so... for a general η

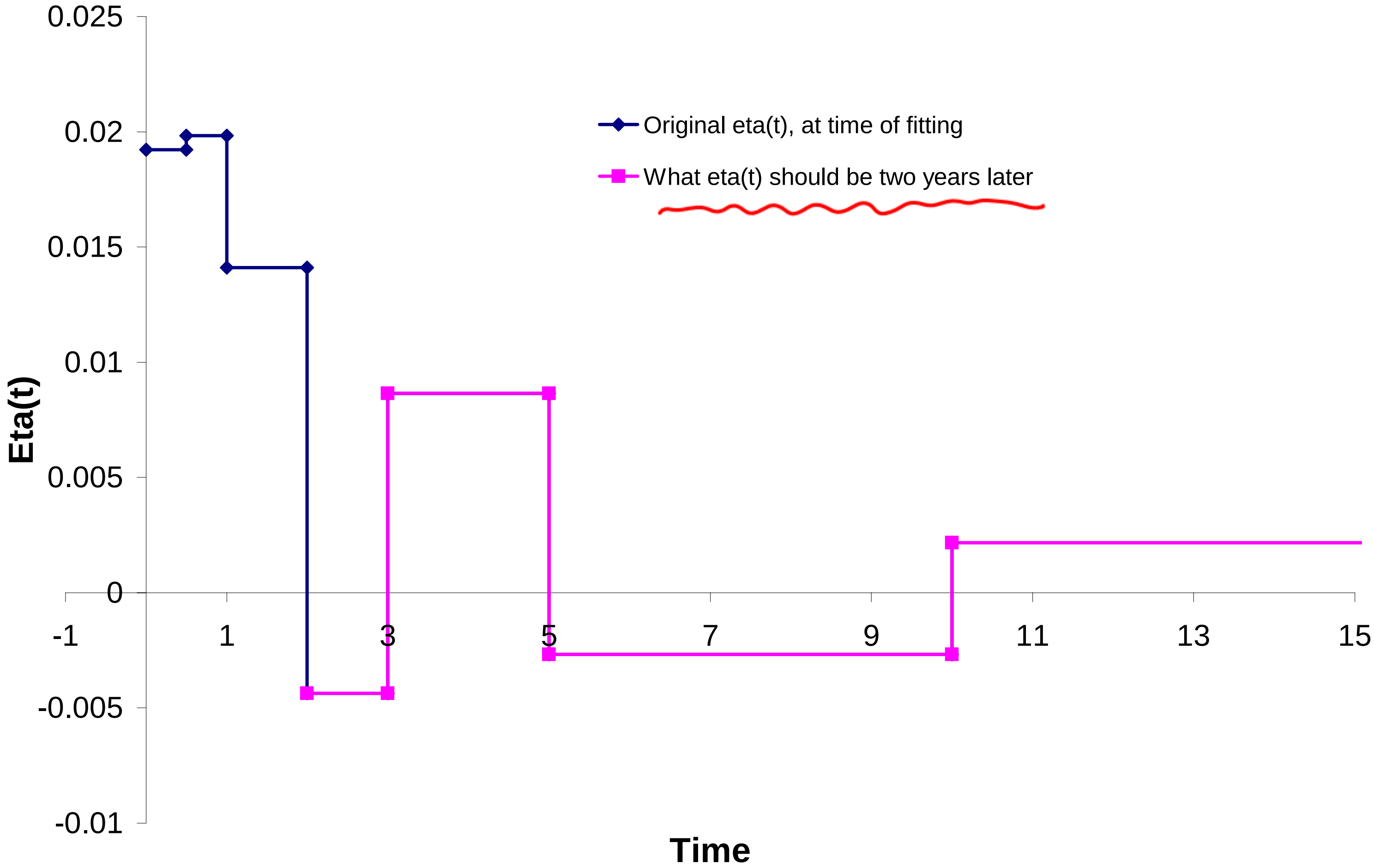
$$\eta_j = \frac{2}{\left(T_j - T_{j-1}\right)^2}\left(-\log\left(Z_M(0;T_j)\right) - r^* T_j + \frac{1}{6}c^2 T_j^3 \right.$$

$$\left. -\frac{1}{2}\sum_{i=1}^{j-1}\eta_i\left(\left(T_j - T_{i-1}\right)^2 - \left(T_j - T_i\right)^2\right)\right)$$

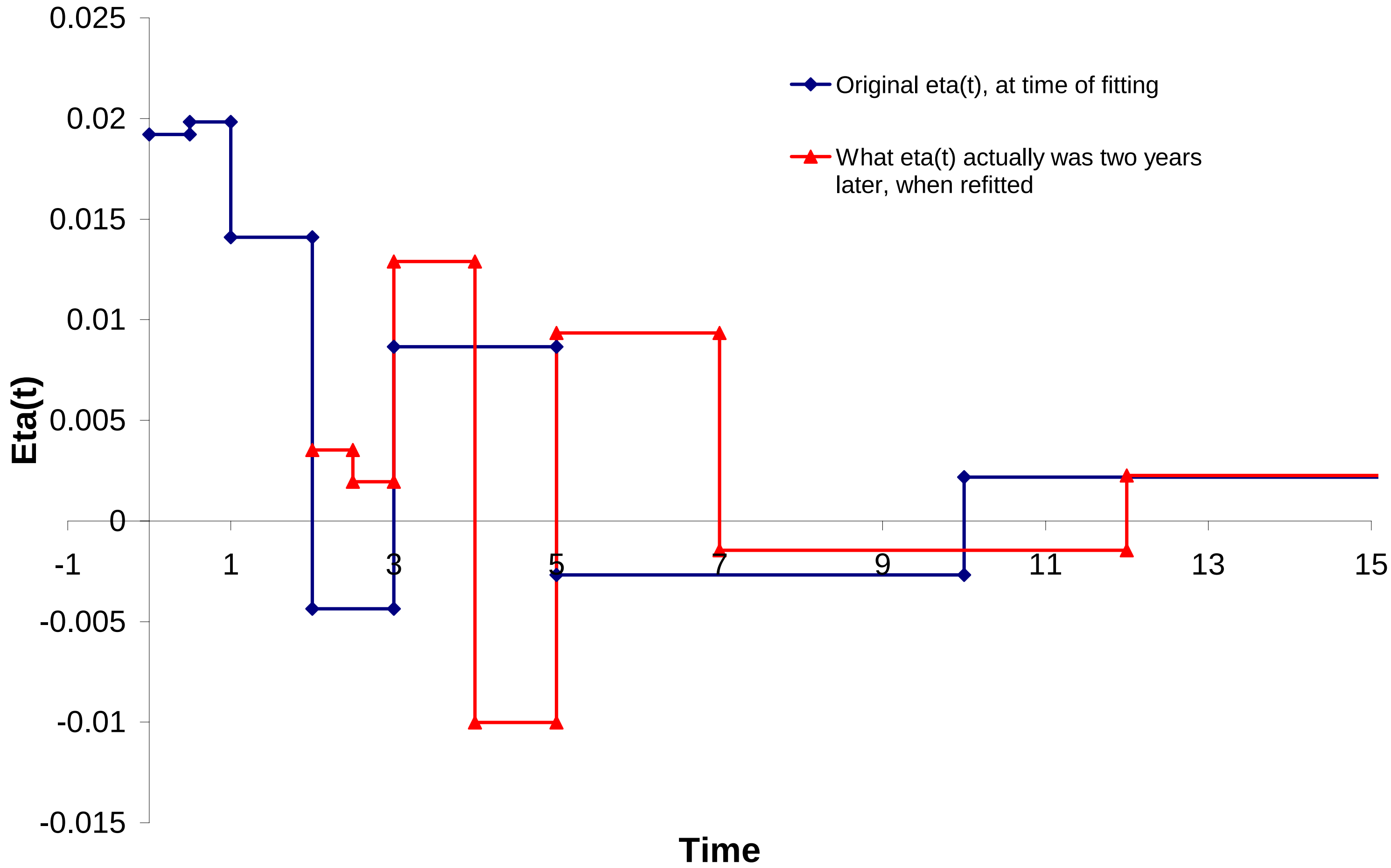
between j and $j-1$

And this is what we find...


0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
Eta(t)
-1
1
2
3
5
7
9
11
13
15
Time
Original eta(t), at time of fitting



0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
Eta(t)
-1
1
3
5
7
9
11
13
15
Time
Original eta(t), at time of fitting
What eta(t) should be two years later



Eta(t)
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-1
1
3
5
7
9
11
13
15
Time
Original eta(t), at time of fitting
What eta(t) actually was two years later, when refitted

Yield-curve fitting: For and against

For

- The building blocks of the bond pricing equation are delta hedging and no arbitrage. If we are to use a stochastic model correctly then we must abide by the delta hedging assumptions. We must buy and sell instruments to remain delta neutral. The buying and selling of instruments must be done at the market prices. We *cannot* buy and sell at a theoretical price.

- Perhaps by hedging with other instruments the dependence of the model on its parameters and assumptions is reduced anyway.

Against

If the market prices of simple bonds were correctly given by a model calibrated at time t^* then, when we come back a week later, t^* + one week, say, to refit the function $\eta^*(t)$, we would find that this function *had not changed* in the meantime.

This *never* happens in practice. We find that the function η^* has changed.

This means that the original model was incorrect.

One problem with calibration is that because it matches an instantaneous 'snapshot' of the market it is difficult to see how wrong it is!

And now for the opposite approach, analyze historical data. . . no reliance on a single 'snapshot.'

Part II

Data analysis to find the 'best' model

e^{A-rB}

The one (or more)-factor models for the spot interest rate that we have seen were all chosen for their nice properties; for most of them we were able to find simple closed-form solutions of the bond-pricing equation.

Clearly, this means that the models are not necessarily a good description of reality.

Let's recap these models quickly. . .

Popular one-factor spot-rate models

The real spot rate r satisfies the stochastic differential equation

Real

$$dr = u(r,t)dt + w(r,t)dX. \tag{4}$$

RN

Model	$u(r,t) - \lambda(r,t)w(r,t)$	$w(r,t)$
Vasicek	$a - br$ linear	c const.
CIR	$a - br$	$cr^{1/2}$
Ho & Lee	$a(t)$	c
Hull & White I	$a(t) - b(t)r$	$c(t)$
Hull & White II	$a(t) - b(t)r$	$c(t)r^{1/2}$
General affine	$a(t) - b(t)r$	$(c(t)r - d(t))^{1/2}$

Here $\lambda(r,t)$ denotes the market price of risk. The function $u - \lambda w$ is the risk-adjusted drift.

For all of these models the zero-coupon bond value is of the form $Z(r,t;T) = e^{A(t,T) - rB(t,T)}$.

The time-dependent coefficients in all of these models allow for the fitting of the yield curve and other interest-rate instruments.

Here ignore all previous models.

From now on in this lecture we will see how to deduce a model for the spot rate from data; it is therefore unlikely to be nice and tractable!

The method

The method that we use assumes that

- the model is time homogeneous drop time dep.

$dr = u(r)\,dt + w(r)\,dX$ (t)

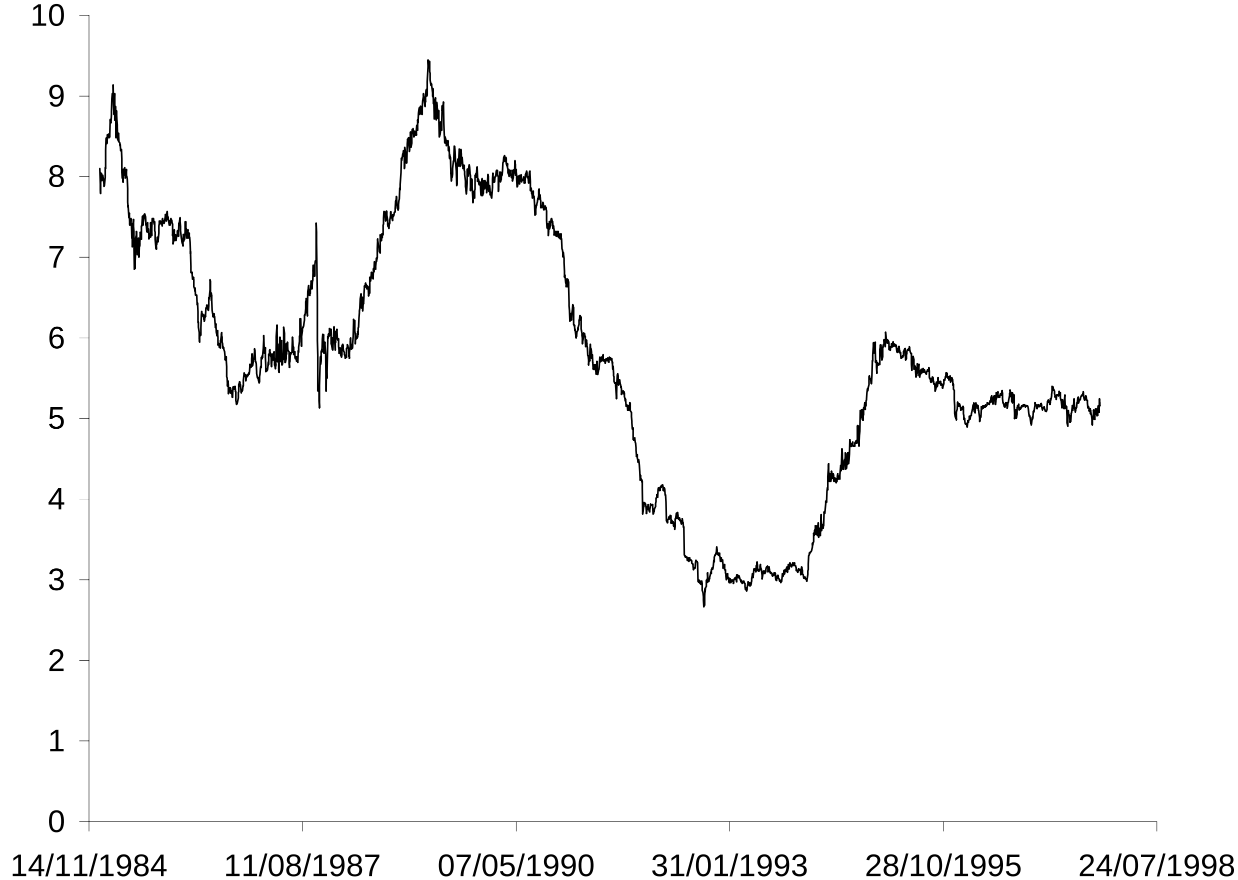
- the spot rate is well behaved (i.e. doesn't wander too far away, go to zero or infinity for example)

The downside to the resulting model is that we cannot find closed-form solutions for contract values, the risk-neutral drift and the volatility don't have a sufficiently nice structure.

In the figure are shown the US one-month LIBOR rates, daily, for the period 1985–1997, and is the data that we use in our analysis.

The ideas that we introduce can be applied to any currency, but here we use US data for illustration.

Aside: This method isn't specific to interest rates, it has also been used to model the gold price, equity and index volatility, and the rate of inflation.



Extra Lecture, "Fear & Greed In F.I Makets"

$u(r)$, $w(r)$ $\lambda(r)$

There are three key stages:

1. By differencing spot rate time series data we determine the volatility dependence on the spot rate $w(r)$. diffusion / vol

2. By examining the steady-state probability density function for the spot rate we determine the functional form of the drift rate $u(r)$.

3. We examine the slope of the yield curve to determine the market price of risk $\lambda(r)$.

The volatility structure

Our first observation is that many popular models take the form

power law in vol.

SDE of choice

$$dr = \overbrace{u(r)}dt + \nu r^{\beta} dX. \qquad (5)$$

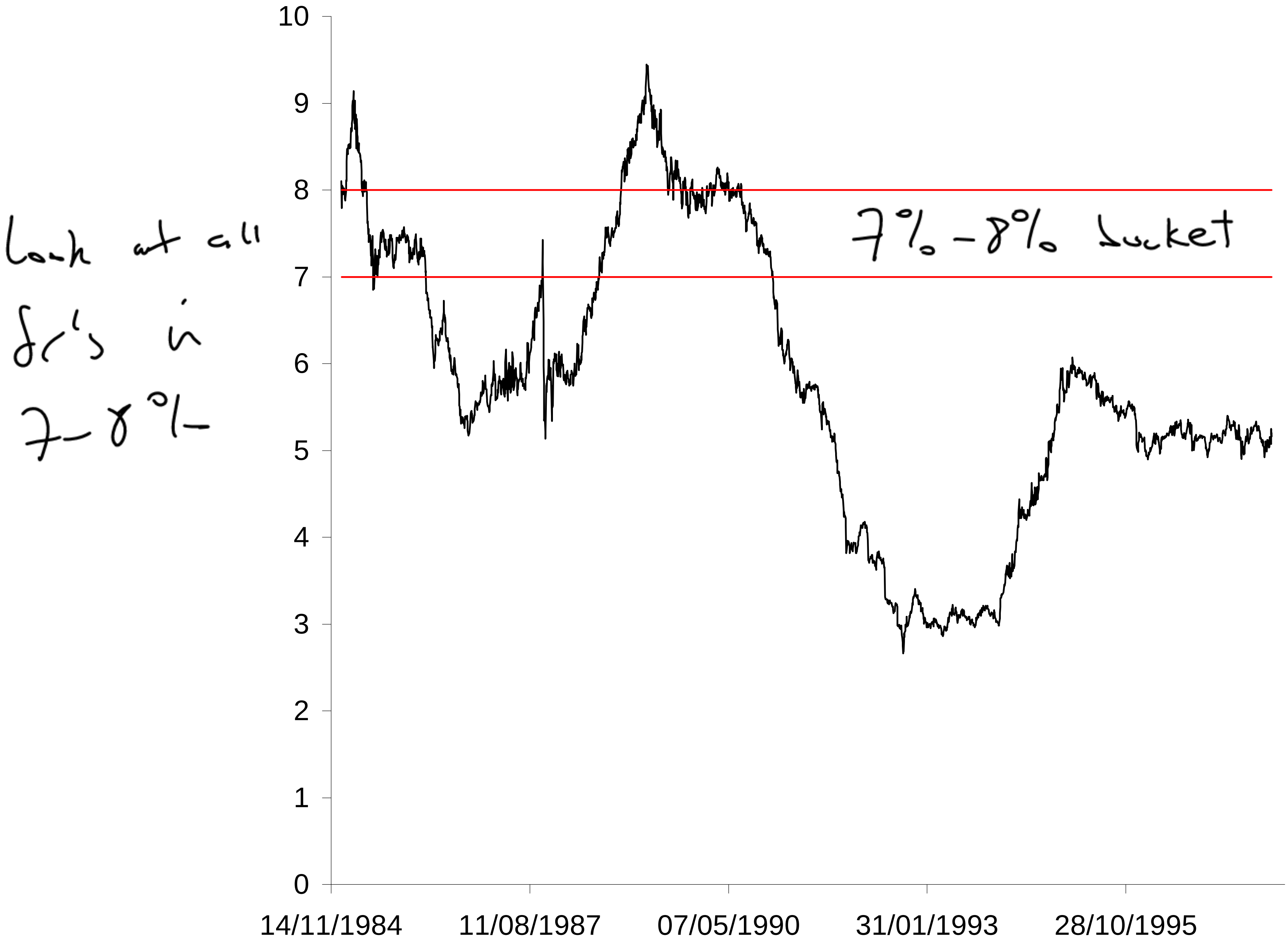
Examples of such models are the Ho & Lee ($\beta = 0$), Vasicek ($\beta = 0$) and Cox, Ingersoll & Ross ($\beta = 1/2$) models.

$\nu \quad \beta \longrightarrow u(r)$

$$E\left[dr^2\right] = \nu^2 r^{2\beta} dt \Rightarrow \sqrt{\frac{E\left[\delta r^2\right]}{\delta t}} = \nu r^{\beta}$$

Using our US spot rate data we can estimate the best value for β using a very simple bucketing technique.

From the time-series data divide the changes in the interest rate, δr, into buckets covering a range of r values.


10
9
8
7
6
5
4
3
2
1
0
14/11/1984
11/08/1987
07/05/1990
31/01/1993
28/10/1995
7% - 8% bucket
Look at all fr's in 7-8%

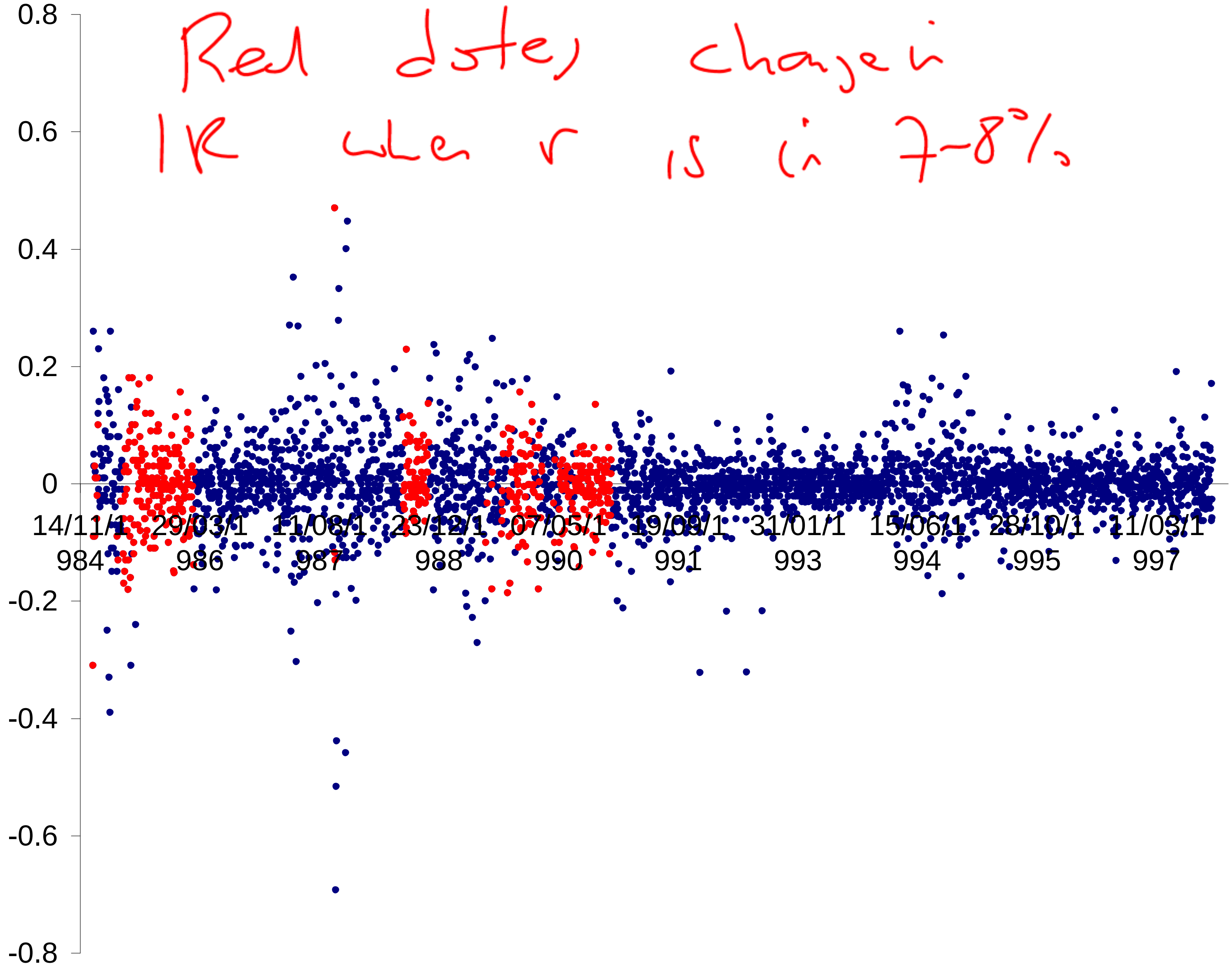
Just examine the δrs associated with each bucket.

Move on to next bucket & repeat

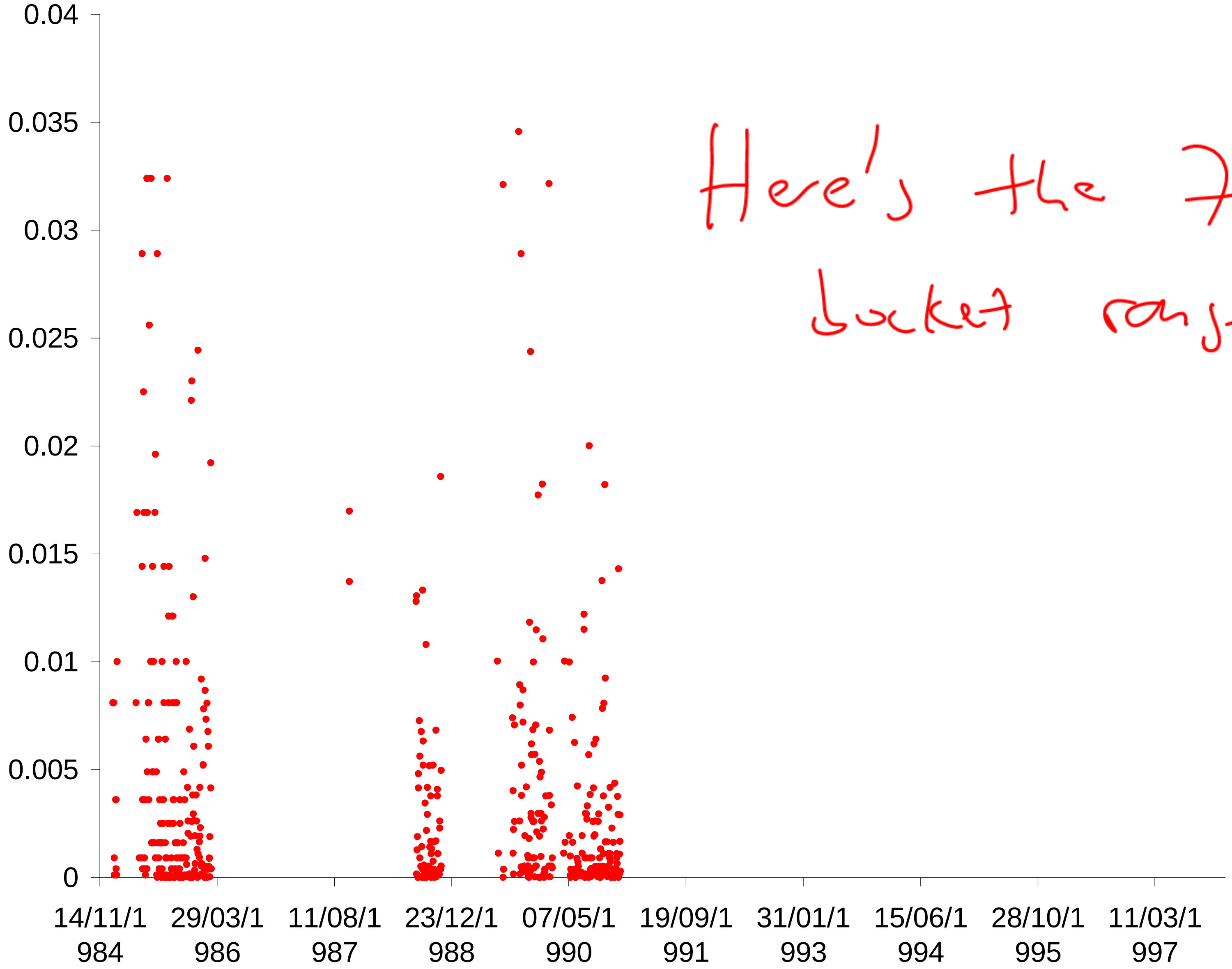
i.e 8—9%, 9–10%

Calc δr's in each bucket/range.

Red dates, change in IR when r is in 7–8%



Then calculate the average value of $(\delta r)^2$, for each bucket.



Here's the 7-8% bucket range.

Certificate in Quantitative Finance

If the model (5) is correct we would expect

Average of δr

$$\boxed{E[(\delta r)^2] = \nu^2 r^{2\beta}\delta t}$$

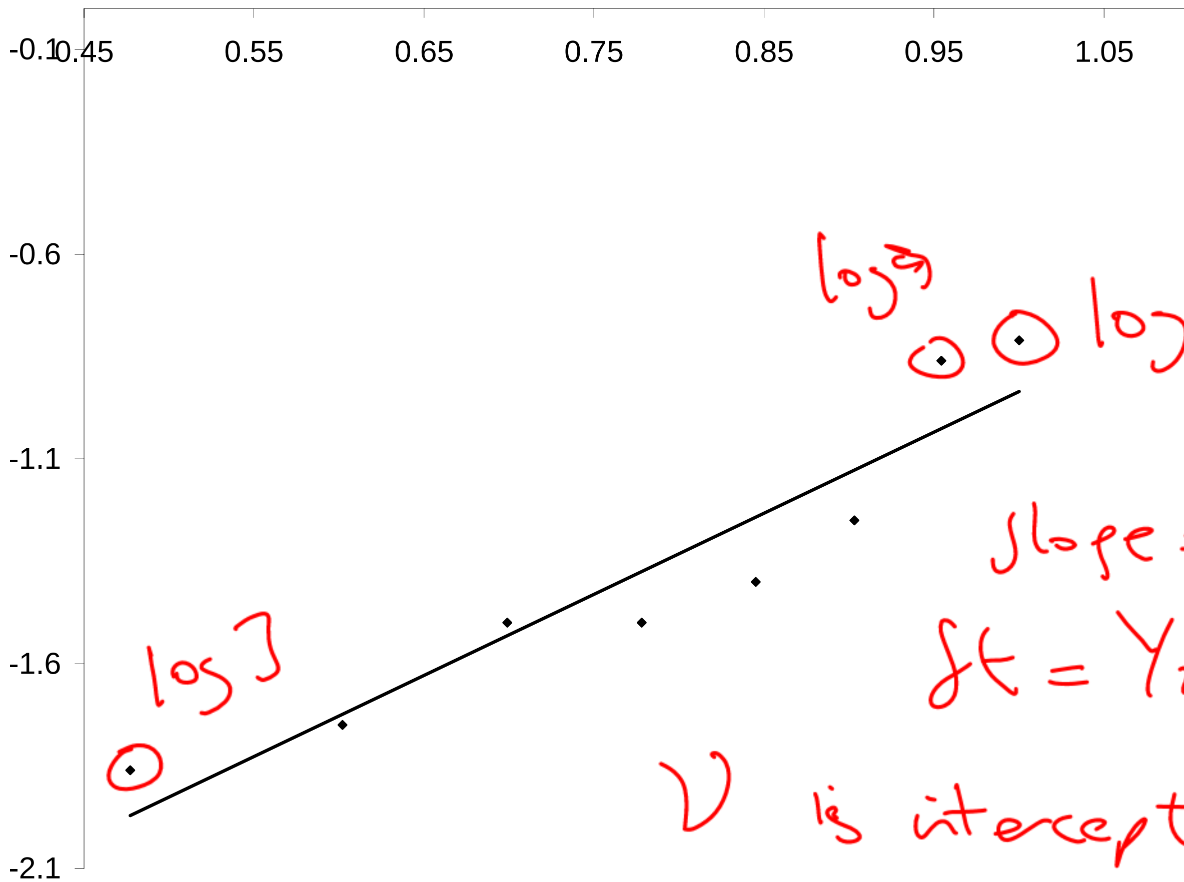
to leading order in the time step δt, which for our data is one day. LHS is from data. RHS assumed model

Now plot $\log(E[(\delta r)^2])$ against $\log r$ using the data.

The slope of this 'line' gives an estimate for 2β and where the line crosses the vertical axis can be used to find ν.

log r (bucket values)

$\log(\mathbb{E}[dr^2])$



slope $= 2\beta \rightarrow \beta$

$\delta t = 1/252$

ν is intercept with y-axis

We have $\beta \rightarrow$ slope

Estimation of β

$\nu \rightarrow$ intercept.

We can see that the line is very straight.

From this calculation it is estimated that

$$\boxed{\beta = 1.13} \quad \text{and} \quad \boxed{\nu = 0.126.}$$

This confirms that the spot rate randomness increases as the spot rate increases, approximately linearly.

$dS = \mu S\,dt + \sigma S\,dX$

(And this rules out Vasicek, Ho & Lee, etc. etc.)

The drift structure

If model is

$dr = a\,dt + b\,dX$

It is statistically harder to estimate the drift term from the data; this term is smaller than the volatility term and thus subject to larger relative errors.

$\mathbb{E}[dr] \longrightarrow a\,dt$

Our approach to finding the drift function is via the empirical and analytical determination of the steady-state probability density function for r.

If r satisfies the s.d.e. (5) then the probability density function $p(r,t)$ for r satisfies the Fokker–Planck equation

$t \to \infty$

$p \equiv p_\infty(r)$

$$\frac{\partial p}{\partial t} = \tfrac{1}{2}\nu^2 \frac{\partial^2}{\partial r^2}(r^{2\beta}p) - \frac{\partial}{\partial r}(u(r)p). \qquad (6)$$

The steady state $p_\infty(r)$ will satisfy the time-independent version of (6):

$$\boxed{\tfrac{1}{2}\nu^2 \frac{d^2}{dr^2}(r^{2\beta}p_\infty) - \frac{d}{dr}(u(r)p_\infty) = 0.} \qquad (7)$$

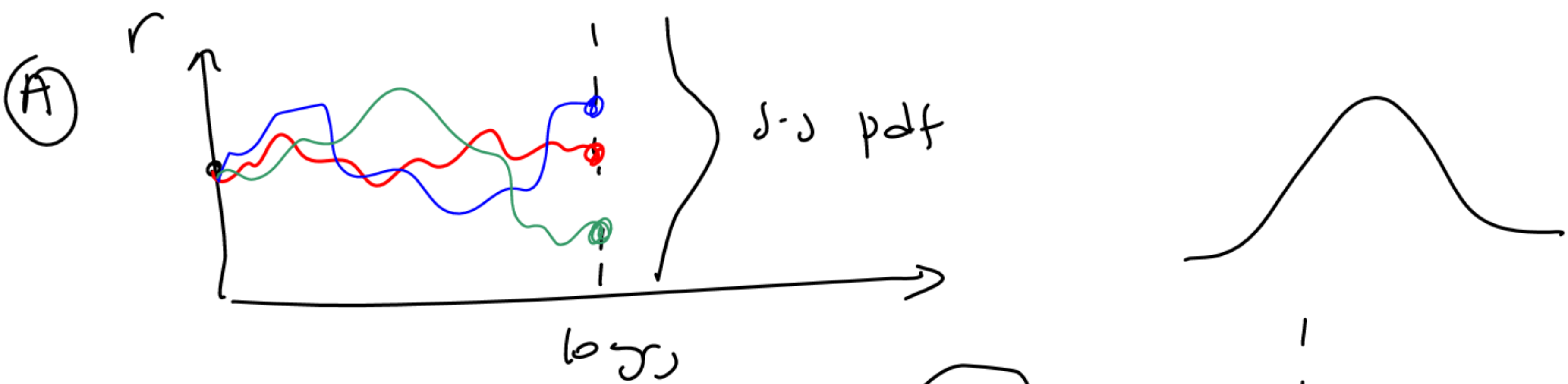
If $u(r)$ known, solve (7) for $P_\infty(r)$

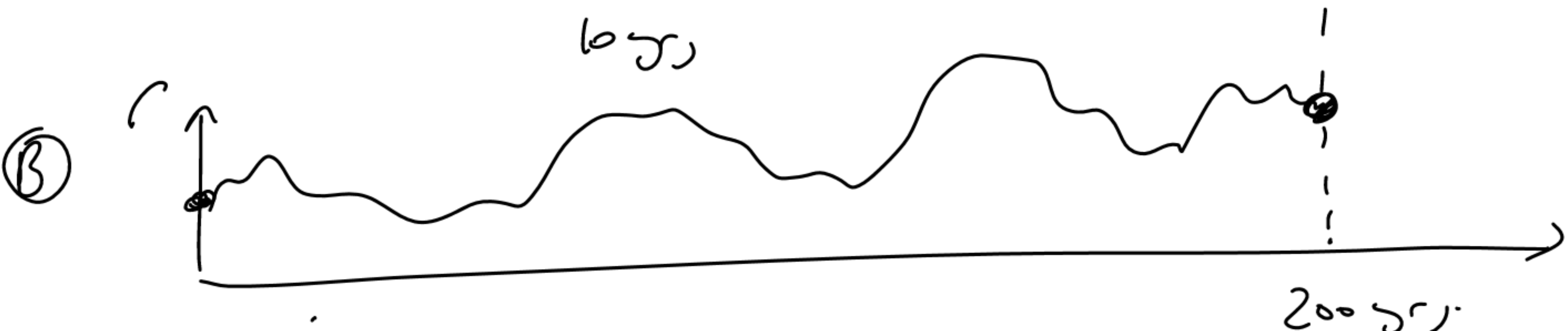
By integrating (7) we find the relationship between the steady-state probability density function and the drift function:

$$u(r) = \nu^2 \beta r^{2\beta-1} + \nu^2 \frac{1}{2} r^{2\beta} \frac{d}{dr}(\log p_\infty). \qquad 7a$$

If we know one we can find the other.

Do we know $p_\infty(r)$? Yes ∵ Ergodic Th^{m}



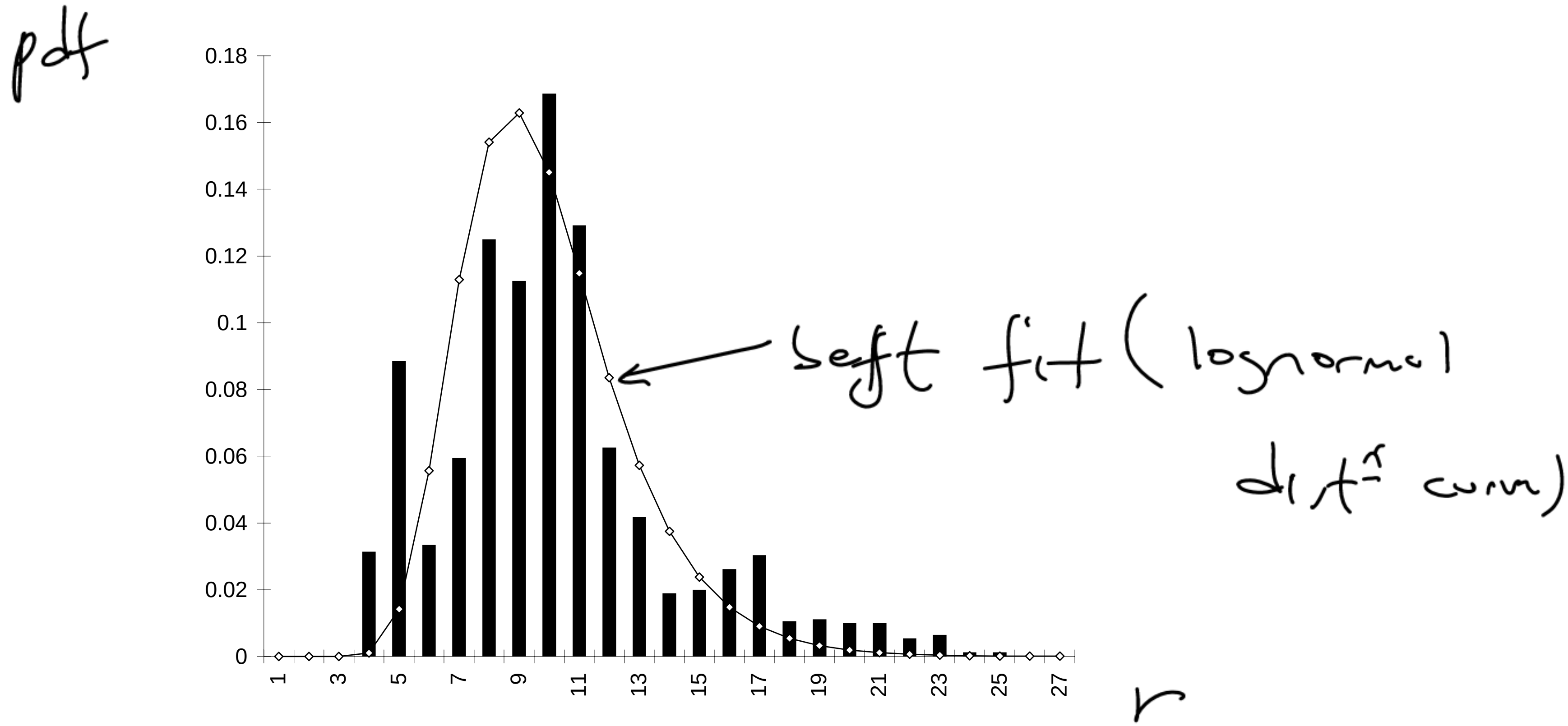


Ergodic Th^{m}: ① If SDE time homog.; ② steady state pdf exists

THEN:

A) dist^{n} at end of many paths ≡ B) dist^{n} at every pt. in time for one very long path.

We can determine a plausible functional form for $p_\infty(r)$ from one-month US LIBOR rates.



This figure shows empirical data, the bars, and a fitted function, the line.

Our choice for $p_\infty(r)$ is

$$p_\infty(r) = \frac{1}{ar\sqrt{2\pi}} \exp\left(-\frac{1}{2a^2}(\log(r/\bar{r}))^2\right)$$

where $a = 0.4$ and $\bar{r} = 0.08$. From this we find that for the US market

$$u(r) = \nu^2 r^{2\beta-1}\left(\beta - \frac{1}{2} - \frac{1}{2a^2}\log(r/\bar{r})\right).$$

People assume mean reversion to look like $(\eta - \gamma r)$ or $-\gamma(r - \bar{r})$

Mean reversion

$r \gg 1$ $(r > \bar{r})$ log is +ve so −ve trend in u

$r \ll 1$ $r < \bar{r}$ log is −ve so +ve trend in u

Advantages of working with the probability density function to find the drift function:

- more stable than other methods

- easy to see whether the probability density function ‘makes sense’

- spot rate cannot go to zero or infinity if probability density function zero there

Finally $\lambda(r)$

The slope of the yield curve and the market price of risk

Now we have found $w(r)$ and $u(r)$, it only remains for us to find $\lambda(r)$.

We shall again allow λ to have a spot-rate dependence, but not a time dependence.

Note: There is no information about the market price of risk in the spot-rate process!

Such information is contained within instruments of finite (not infinitesimal) maturity.

We will examine the short end of the curve for this information.

Slope is $\frac{1}{2}(u - \lambda w)$

Power

Let us expand $Z(r,t;T)$ in a Taylor series about $t=T$, this is the short end of the yield curve.

We know that zero-coupon bonds satisfy

$$\frac{\partial Z}{\partial t}+\frac{1}{2}w(r,t)^2\frac{\partial^2 Z}{\partial r^2}+(u(r,t)-\lambda(r,t)w(r,t))\frac{\partial Z}{\partial r}-rZ=0.$$

Look for a solution for small times to maturity of the form

$$Z \sim A(r)+a(r)(T-t)+b(r)(T-t)^2+\cdots=\sum_{n=0}^{\infty} f_n(r)(T-t)^n$$

$$Z \sim 1+a(r)(T-t)+b(r)(T-t)^2+\cdots.$$

We know that at $t=T$ $Z=1$ $\therefore$ $\boxed{A(r)=1}$

Now subst Z is $\frac{\partial Z}{\partial t}+\frac{1}{2}w^2\frac{\partial^2 Z}{\partial r^2}+(u-\lambda w)\frac{\partial Z}{\partial r}=rZ$

and equate powers of $(T-t)$

Put this form into the bond pricing equation and equate powers of $(T-t)$ and you will find that

$$Z(r,t;T) \sim 1 - r(T-t) + \frac{1}{2}(T-t)^2(r^2 - u + \lambda w) + \ldots \quad \text{as } t \to T.$$

This is just a simple Taylor series approximation to the solution for a zero-coupon bond, *for any one-factor model*!

Take ln of Z then ÷ thro' by $-(T-t)$

From this we can find the shape of the yield curve near the short end:

intercept

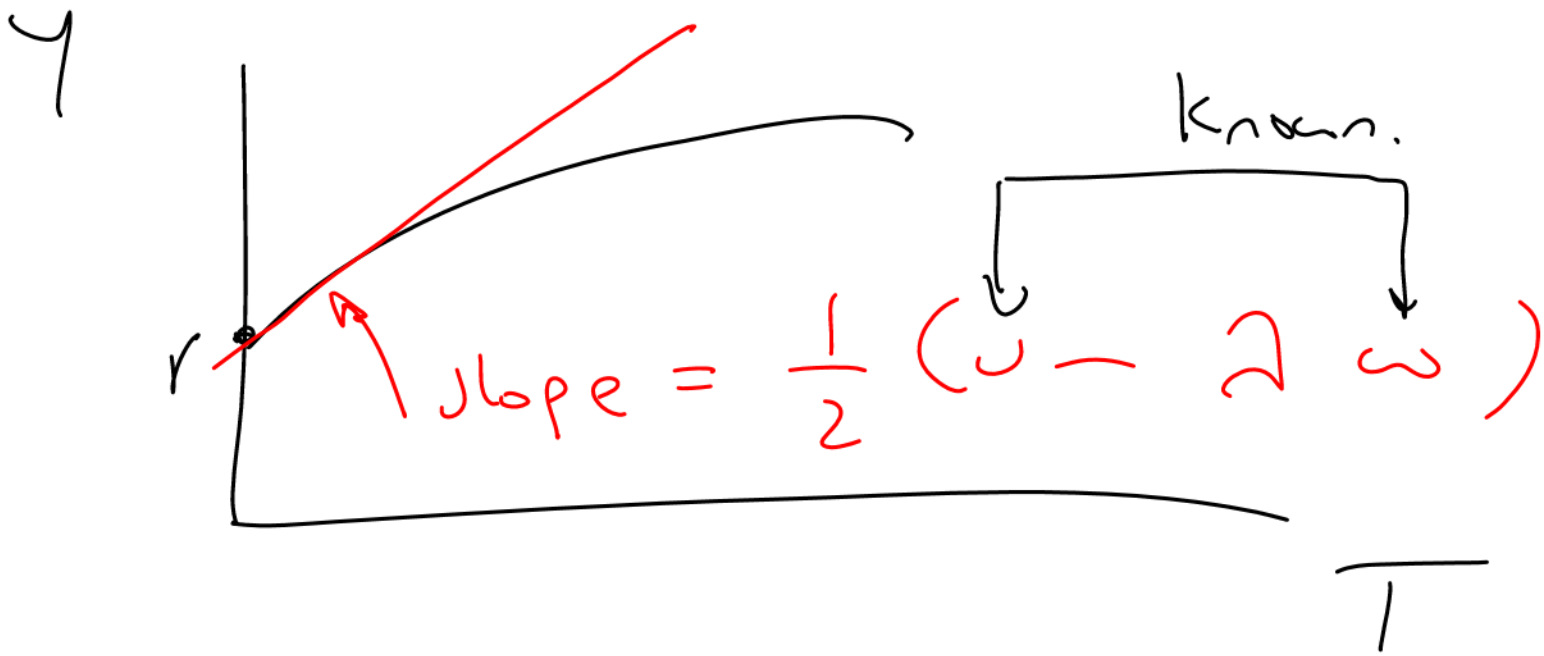
unknown

$$y \qquad -\frac{\ln Z}{T-t} \sim r + \underbrace{\tfrac{1}{2}(u-\lambda w)}_{\text{gradient}}(T-t) + \dots \quad \text{as } t \to T. \tag{8}$$

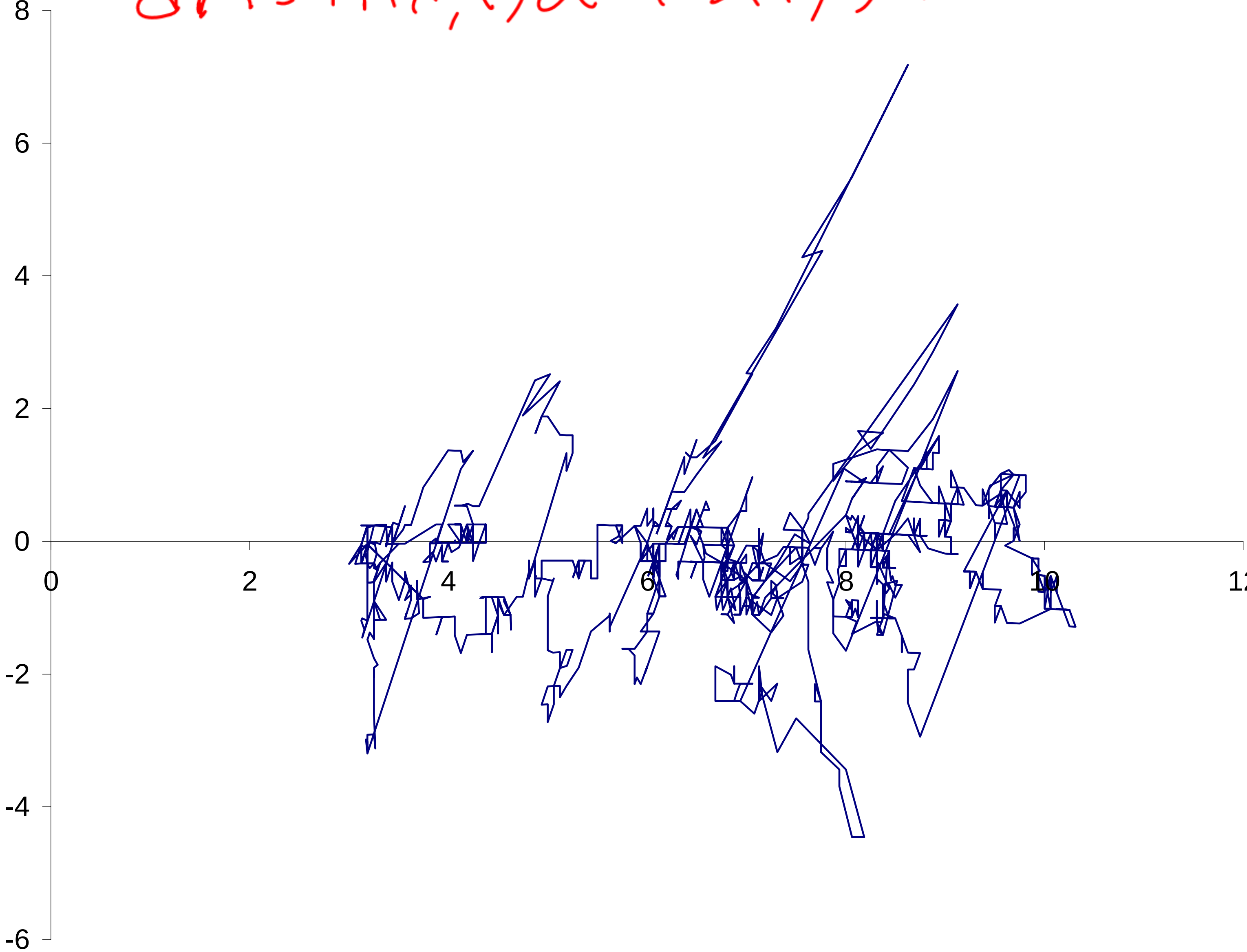
The first term says that the short end of the yield curve is r (obvious!), and the second term says that the slope of the yield curve at the short end in this one-factor model is simply $(u - \lambda w)/2$.

We can use this result together with time-series data to determine the form for $u - \lambda w$ empirically. . . and since we have functional forms for $u(r)$ and $w(r)$ that means we can find $\lambda(r)$.

And so, the parameter λ as a function of r is . . .

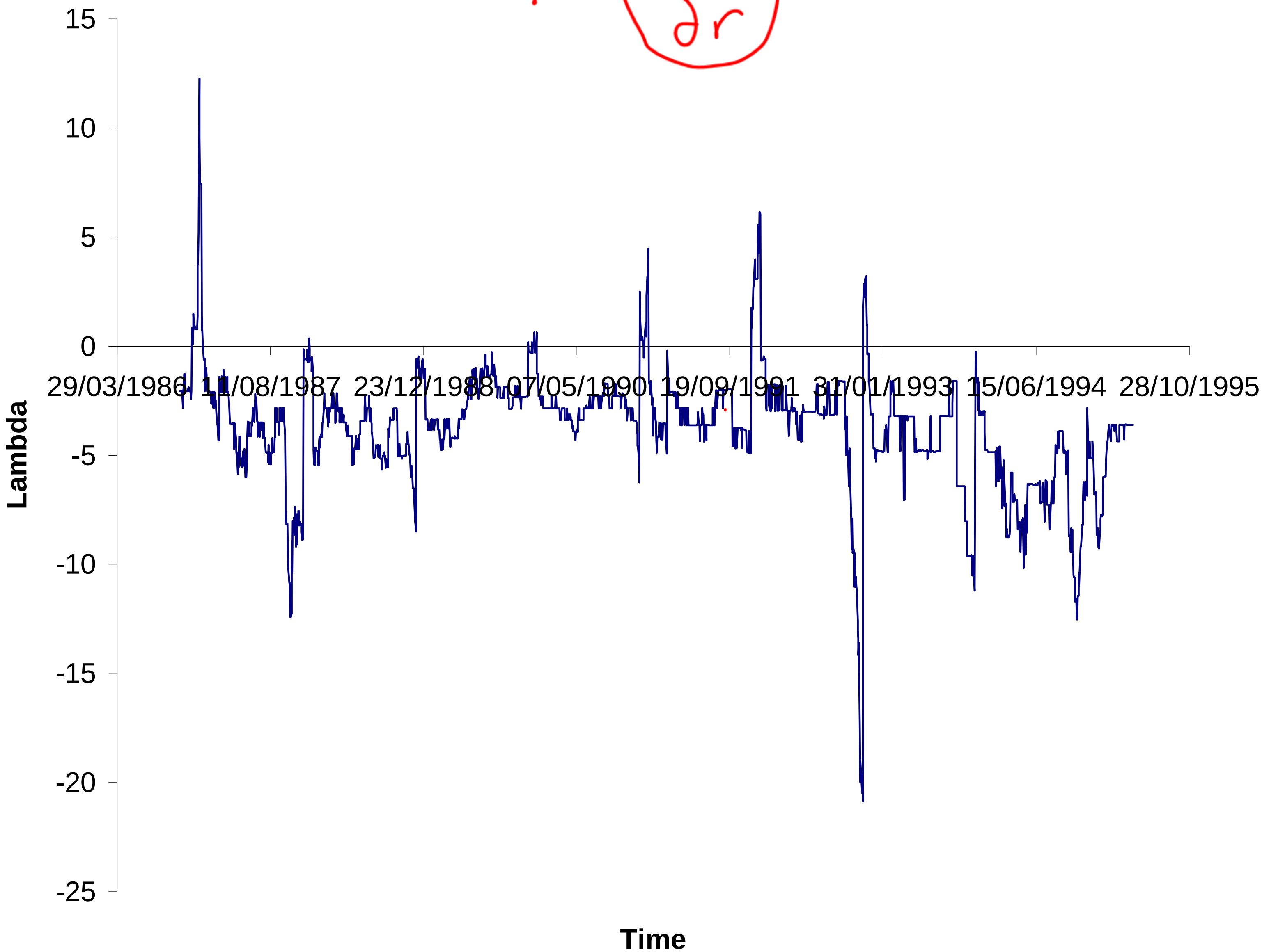


$$d\lambda = A(\lambda,t)\,dt + B(\lambda,t)\,dX$$



A mess!

$\lambda \frac{\partial V}{\partial r}$



Possible conclusion from this:

- it is 'easy' to model the spot interest rate!

- but difficult to model the market price of interest rate risk!!

Summary

Please take away the following important ideas

- Spot interest rate models are usually calibrated to match market data, in particular the forward curve

- This calibration is in practice always inconsistent

Ilder

- There are simple methods for examining interest rate data to find good models

$dr = (u - \lambda w)\,dt + w\,dX$

- Thanks for your interest
- This is my last lecture

Fixed Income and Credit – Lecture 5

Martingales and Fixed Income Valuation

May 2020

CQF

Martingales and Fixed Income Valuation

Introduction

Introduction to martingale and fixed income!

Financial mathematics useful.

Short-term rates and risk free zero-coupon bonds do not exists in practice.

The lecture is useful for financial mathematics modelling, presents the building block of the technical tool, but is not a faithful description of the actual market.

This lecture

Apply probabilistic and martingale methods to the pricing of bonds and interest rate derivatives using short-term rate and bond models. We will see:

- the pricing of interest rate products in a probabilistic setting; the equivalent martingale measures;
- the fundamental asset pricing formula for bonds;
- the dynamics of bond prices; the forward measure;
- the fundamental asset pricing formula for derivatives on bonds;

Martingales and Fixed Income Valuation

A Model for the Short-Term Rate

In the world of equity, a single class of models (the Geometric Brownian Motion and its extensions) dominates the landscape. The situation is different in the world of interest rates, where several classes of models coexist:

1. Single factor short-term rate models, such as the Merton, Vasicek, Cox-Ingersoll-Ross, Ho-Lee or Hull-White models;
2. Multifactor models such as the Brennan-Schwartz, Longstaff-Schwartz and Fong-Vasicek, G2++ (Hull-White 2-factor) models;
3. Forward instantaneous rate models such as the Heath-Jarrow-Morton model
4. Forward market rate model like Brace, Gatarek and Musiela;

Modelled quantities can be (theoretical) short-term rate, (theoretical) forward rates, discount factors or market rates.

A Model for the Short-Term Rate

Underlying probability space: $(\Omega, \mathcal{F}, \mathbb{P})$. The filtration $\mathcal{F}$ is the $\mathbb{P}$-completed version of the filtration generated by the underlying d-dimensional standard Brownian motion X. Time interval: $[0, \bar{T}]$.

This lecture starting point: short-term rate models whose dynamics under the physical measure $\mathbb{P}$ is of the form

$$\underline{dr(t) = \mu(t, r(t))\, dt + \sigma(t, r(t)) \cdot dX(t)}$$

where $r(t)$ is the short-term rate at time t and $X(t)$ is a Brownian motion.

To simplify the notation, we will write this equation as

$$\underline{dr(t) = \mu_t dt + \sigma_t \cdot dX(t)}$$

This specification is general enough to cover both

1. Equilibrium models such as the Vasicek or Cox-Ingersoll-Ross models;
2. No arbitrage models such as Ho-Lee and Hull-White;

A Model for the Short-Term Rate

Equipped with a short-term interest rate model, we define a *cash account*, *money market account* or *money-in-the-bank* process $A(t)$ as

$$A(t) = \exp\left(\int_0^t r(s)ds\right).$$

The starting value is $A(0) = 1$ and the account grows at the instantaneous risk free rate r:

$$dA(t) = r(t)A(t)dt.$$

Martingales and Fixed Income Valuation

The Zero-Coupon Bond Market

A zero-coupon bond $B(t, T)$ is a bond which pays 1 at maturity time T. We define the zero-coupon bond market as the set of all the zero-coupon bonds $B(t, T)$ for all $t \leq T \leq \bar{T}$.

Riskless zero-coupon bonds do not exists in practice. This is an idealised quantity that is useful for intuition but need to be adapted to the market, e.g. OIS discounting.

We cannot expect to find a zero-coupon bond maturing for all times $t \leq T \leq \bar{T}$. Only a discrete number of maturities exists, at most one for each good business day T.

This representation is useful for financial mathematics modelling, as the building block of the technical tool, but not as a faithful description of the market.

Arbitrage-free family of bonds

Definition (Arbitrage-free family of bonds)

A family $B(t,T)$ $(t \leq T \leq \bar{T})$ of adapted processes is called an arbitrage-free family of bond prices relative to short-term interest rate process r if

1. $B(T,T) = 1$ *for every* $T \in [0, \bar{T}]$.
2. *there exists a probability measure* $\mathbb{Q}$ *on* $(\Omega, \mathcal{F}_{\bar{T}})$ *equivalent to* $\mathbb{P}$ *and such that for any maturity* $T \in [0, \bar{T}]$ *the relative bond price*

$$\boxed{Z(t,T) = \frac{B(t,T)}{A(t)}}$$

is a $\boxed{\textit{martingale}}$ *under* $\mathbb{Q}$.

Any such probability measure $\mathbb{Q}$ is called a (spot) martingale measure for the family $B(t,T)$ relative to r.

Equivalent Measure or Equivalent Measures?

First, Definition is consistent with the way we defined the equivalent martingale measure for a stock. But the key difference in the zero-coupon bond case is that the equivalent martingale measure does not apply to a single security, but to a continuum of securities.

As shown in a previous lecture, the existence of (at least) one equivalent martingale measure implies an absence of arbitrage opportunity.

Any probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ is given by the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \mathcal{E}_{\bar{T}}\left(\int_0^{\cdot} -\theta_s \cdot dX_s\right), \quad \mathbb{P} \text{ a.s.}$$

for some predictable $\mathbb{R}^d$-valued process θ.

Equivalent Martingale Measure and No-Arbitrage

The process $\theta(t)$ satisfies the Novikov condition

$$\mathrm{E}^{\mathbb{P}}\left[\exp\left(\frac{1}{2}\int_0^{\bar{T}} |\theta(s)|^2\, ds\right)\right] < \infty$$

We suppose the existence of an equivalent martingale measure $\mathbb{Q}$. The conditional expectation of the derivative is

$$\frac{d\mathbb{Q}}{d\mathbb{P}}_{|\mathcal{F}_t} = \eta(t) = \exp\left(-\int_0^t \theta(s)\cdot dX(s) - \frac{1}{2}\int_0^t |\theta(s)|^2 ds\right).$$

Note that η satisfies

$$\underline{d\eta(t) = -\eta(t)\,\theta(t)\cdot dX(t).}$$

Equivalent Martingale Measure and No-Arbitrage

By Girsanov's theorem, the process X^θ defined as

$$X^\theta(t) = X(t) + \int_0^t \theta(s)\, ds, \quad t \in [0, \bar{T}]$$

is a standard Brownian Motion on $(\Omega, \mathcal{F}, \mathbb{Q})$.

As a result, the dynamics of $r(t)$ under $\mathbb{Q}$ is

$$dr(t) = (\mu_t - \sigma_t \cdot \theta(t))\, dt + \sigma_t \cdot dX^\theta(t).$$

Equivalent Martingale Measure and No-Arbitrage

Applying the martingale property of the rescaled bond, we get

$$B(t, T) = A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{B(T,T)}{A(T)}\right|\mathcal{F}_t\right].$$

Since $B(T,T) = 1$ and $A(T) = \exp\left(\int_0^T r(s)\,ds\right)$ then the formula simplifies to

$$B(t, T) = \mathrm{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)\,ds\right)\middle|\,\mathcal{F}_t\right].$$

Bond Pricing in Practice: Analytical Solutions

Some form of analytical solutions exist for the most popular models such as the Vacicek model or the CIR model. One way to derive them is to apply Feynman-Kac to the fundamental asset pricing formula to obtain a PDE, and then solve that PDE.

$$\frac{\partial B}{\partial t}(t,r) + (\mu_t - \sigma_t \theta(t))\frac{\partial B}{\partial r}(t,r) + \frac{1}{2}\sigma_t^2 \frac{\partial^2 B}{\partial^2 r}(t,r) - r(t)B(t,r) = 0 \quad B(T,r) = 1$$

This lecture is related to martingale approach.

Bond Pricing in Practice: Numerical Solutions

Given a short-rate model with dynamics

$$dr(t) = (\mu_t - \sigma_t \cdot \theta(t))\, dt + \sigma_t \cdot dX^{\theta}(t).$$

under the equivalent martingale measure $\mathbb{Q}$ and the fundamental asset pricing formula applied to zero-coupon bonds

$$B(t, T) = \mathrm{E}^{\mathbb{Q}} \left[\exp\left(-\int_t^T r(s)\, ds \right) \middle| \mathcal{F}_t \right].$$

it is not difficult to obtain a numerical solution using Monte Carlo methods.

Bond Pricing in Practice: Numerical Solutions

A Monte-Carlo algorithm on the short-term rate (dimension 1) looks like

- $r(0) = r$
- Simulations $j = 1, \dots, N$
  - Time steps t_i, $i = 1, \dots M$
    - Simulate $N(0,1)$ random variable $Z_{j,i}$
    - Simulate short rate

$$r_j(t_i) = r_j(t_{i-1}) + (\mu(t_{i-1}, r_j(t_{i-1})) - \sigma(t_{i-1}, r_j(t_{i-1}))\theta(t_{i-1}))\,(t_i - t_{i-1}) + \sigma(t_{i-1}, r_j(t_{i-1}))Z_{j,i}\sqrt{t_i - t_{i-1}}$$

    - Compute integral $\int_0^T r(s)ds$: $\text{Int} = \text{Int} + 0.5 * (r_j(t_i) + r_j(t_{i-1}))(t_i - t_{i-1})$.
  - Zero-bond under simulation $B_j = \exp(-\text{Int})$
- Value of the bond:

$$B = \frac{1}{M}\sum_{j=1}^{M} B_j.$$

Monte Carlo on r not most efficient numerically. Do it on B! (see end of section on *Pricing bond derivatives*)

Martingales and Fixed Income Valuation

Dynamics of the Zero-Coupon Bond Price

We will consider the pricing of derivatives struck on a zero-coupon bond. To price such derivatives, we will need to understand not only the dynamics of the short-term rate but also the dynamics of the zero-coupon bond prices.

Our starting point is to look at the process

$$M(t) = Z(t, T)\eta(t) = \frac{B(t, T)}{A(t)}\eta(t)$$

where $\eta(t)$ is the Radon Nikodym derivative.

The reason for considering this process M(t) is that:

If a process $Y(t)$ is a martingale under $\mathbb{Q}$ and $\eta(t) = \frac{d\mathbb{Q}}{d\mathbb{P}}_{|\mathcal{F}_t}$ then the process $M(t) = Y(t)\eta(t)$ is a martingale under $\mathbb{P}$.

Dynamics of the Zero-Coupon Bond Price

Since $Z(t,T)$ is a martingale under $\mathbb{Q}$ (by definition of $\mathbb{Q}$), $M(t) = Z(t,T)\eta(t)$ is a martingale under $\mathbb{P}$. By the *Martingale Representation Theorem*, there exists a stochastic process γ such that $M(t)$ can be represented as

$$\boxed{M(t) \quad = \quad M(0) + \int_0^t \gamma_s \cdot dX(s)}$$

$$= \quad Z(0,T) + \int_0^t \gamma_s \cdot dX(s)$$

or equivalently

$$\underline{dM(t) = \gamma_t \cdot dX(t).}$$

We express $Z(t,T)$ in term of $M(t)$ as $Z(t,T) = M(t)\eta^{-1}(t)$ and

$$dZ(t,T) = d(M(t)\eta^{-1}(t))$$

Dynamics of the Zero-Coupon Bond Price

The differential of η is

$$d\eta(t) = -\eta(t)\,\theta(t)\cdot dX_t = -\eta(t)\,\theta(t)\cdot\left(dX^{\theta}(t) - \theta(t)dt\right)$$

and thus $\frac{1}{\eta}$ $\quad -\frac{1}{\eta^2}(-\eta\theta\, dX) + \frac{\cancel{2}}{\eta^3}\frac{1}{\cancel{2}}\eta^2\theta^2 dt$

$$d\eta^{-1}(t) = \eta^{-1}(t)\left(\theta(t)\cdot dX_t + |\theta(t)|^2 dt\right) = \eta^{-1}(t)\left(\theta(t)\cdot dX^{\theta}(t)\right)$$

Using the Itô product rule we have

$$\begin{aligned}
dZ(t,T) &= d(M(t)\eta^{-1}(t)) = dM(t)\eta^{-1}(t) + M(t)d\eta^{-1}(t) + \eta^{-1}(t)\gamma(t)\cdot\theta(t)dt \\
&= \eta^{-1}(t)\gamma(t)\cdot(dX^{\theta}(t) - \cancel{\theta(t)dt}) + M(t)\eta^{-1}(t)\theta(t)\cdot dX^{\theta}(t) + \cancel{\eta^{-1}(t)\gamma(t)\cdot\theta(t)dt} \\
&= \eta^{-1}(t)(\gamma(t) + M(t)\theta(t))\cdot dX^{\theta}(t).
\end{aligned}$$

Dynamics of the Zero-Coupon Bond Price

Recalling that $Z(t,T) = B(t,T)/\cancel{A(t)}$, we get the dynamic of $B(t,T)$ under $\mathbb{Q}$

$$
\begin{aligned}
dB(t,T) &= d(Z(t,T)A(t)) \\
&= Z(t,T)dA(t) + dZ(t,T)A(t) \\
&= r(t)\underline{Z(t,T)A(t)}dt + A(t)(\eta^{-1}(t)\gamma(t) + \underline{Z(t,T)}\theta(t)) \cdot dX^{\theta}(t).
\end{aligned}
$$

So

$$
\begin{aligned}
dB(t,T) &= r(t)B(t,T)dt + (\eta^{-1}(t)\underline{A(t)}\gamma(t) + B(t,T)\theta(t)) \cdot dX^{\theta}(t) \\
&= r(t)B(t,T)dt + \left(B(t,T)\frac{\gamma(t)}{Z(t,T)\eta(t)} + B(t,T)\theta(t)\right) \cdot dX^{\theta}(t) \\
&= r(t)B(t,T)dt + B(t,T)\left(\frac{\gamma(t)}{M(t)} + \theta(t)\right) \cdot dX^{\theta}(t).
\end{aligned}
$$

Dynamics of the Zero-Coupon Bond Price

$$
\begin{aligned}
dB(t,T) &= r(t)B(t,T)dt + b^{\theta}(t,T)B(t,T)\cdot dX^{\theta}(t) \\
&= B(t,T)\left(r(t)dt + \underline{b^{\theta}(t,T)}\cdot dX^{\theta}(t)\right)
\end{aligned}
$$

where

$$
\boxed{b^{\theta}(t,T) = \frac{\gamma(t)}{M(t)} + \theta(t).}
$$

The process $b^{\theta}(t,T)$ is called the *volatility* of the zero-coupon bond of maturity T.

Dynamics of the Zero-Coupon Bond Price

The dynamics of a zero-coupon bond $B(t, T)$ under the equivalent martingale measure $\mathbb{Q}$ is given by

$$\frac{dB(t,T)}{B(t,T)} = r(t)dt + b^{\theta}(t,T)\cdot dX^{\theta}(t) \quad B(T,T) = 1$$

where b^{θ} is the volatility of the zero-coupon bond. Therefore

$$\begin{aligned} \frac{B(t,T)}{A(t)} &= B(0,T)\exp\left(\int_0^t b^{\theta}(s,T)\cdot dX^{\theta}(s) - \frac{1}{2}\int_0^t |b^{\theta}(s,T)|^2 ds\right) \\ &\boxed{= B(0,T)\mathcal{E}_t\left(\int_0^{\cdot} b^{\theta}(s,T)\cdot dX^{\theta}(s)\right)}. \end{aligned}$$

An interest rate models family will describe the form of b.

Martingales and Fixed Income Valuation

Market price of risk

Under an equivalent martingale measure $\mathbb{Q}$, the dynamics of a zero-coupon bond is given by

$$\frac{dB(t,T)}{B(t,T)} = r(t)dt + b^{\theta}(t,T)\cdot dX^{\theta}(t) \quad B(T,T) = 1.$$

The dynamic under the actual probability measure $\mathbb{P}$ is

$$\frac{dB(t,T)}{B(t,T)} = \underbrace{\boxed{\Big(r(t) + b^{\theta}(t,T)\cdot\theta(t)\Big)}}_{\mu} dt + b^{\theta}(t,T)\cdot dX(t) \quad B(T,T) = 1.$$

Market price of risk

The process $\theta(t)$ through which we defined the relation between the physical measure $\mathbb{P}$ and the equivalent martingale measure $\mathbb{Q}$ has an economic meaning: it is the risk premium or market price of interest rate risk.

The market price of risk represents the compensation paid by the market to an investor per unit of risk. In our framework, the risk is represented by the volatility of the zero-coupon bond, $b^{\theta}(t, T)$, and the total compensation for risk that an investor requires is equal to $b^{\theta}(t, T) \cdot \theta(t)$.

This observation is consistent with the equity case. In the stock option case, the market price of risk was the process $\theta = (\mu - r)/\sigma$.

The relationship between the physical measure $\mathbb{P}$ and an equivalent martingale $\mathbb{Q}$ measure is established by the market price of risk which acts as the change of measure process. In complete markets the equivalent martingale measure is unique and so is the market price of risk. In incomplete markets, we may have several equivalent martingale measures, each with its own market price of risk.

Martingales and Fixed Income Valuation

~~Starting back at~~ 7:15 ~~BST~~

Pricing bond derivatives

Fixed income markets present a wide diversity of instruments: bonds, of course, but also forwards, options, caps and floors, numerous swaps and swaptions, structure notes... and this is without mentioning the interconnection between fixed income markets and credit market or between fixed income products and inflation-linked products.

The pricing zero-coupon bond is simply the starting point of any attempt to price fixed income products. In this section, we start to expand our horizons by considering the pricing of derivatives on zero-coupon bonds.

Pricing bond derivatives

In this section, we denote by $X^{\mathbb{Q}}(t)$ a $\mathbb{Q}$-standard Brownian motion and express the bond price dynamics as

$$\frac{dB(t,T)}{B(t,T)} = r(t)dt + b(t,T)\cdot dX^{\mathbb{Q}}(t) \quad B(T,T) = 1.$$

This means

$$\frac{B(t,T)}{A(t)} = B(0,T)\exp\left(\int_0^t b(s,T)\cdot dX^{\mathbb{Q}}(s) - \frac{1}{2}\int_0^t |b(s,T)|^2 ds\right).$$

Applying the Fundamental Asset Pricing Formula

The fundamental asset pricing formula tells us that the time t price of a contingent claim paying some (random) amount Y at time T is given by

$$\pi_t(Y) = A(t)\,\mathrm{E}^{\mathbb{Q}}\left[A^{-1}(T)Y\middle|\,\mathcal{F}_t\right].$$

In particular, the value of a zero-coupon bond maturing at time T is given by

$$B(t,T) = A(t)\,\mathrm{E}^{\mathbb{Q}}\left[A^{-1}(T)\middle|\,\mathcal{F}_t\right].$$

Now, what would happen if we wanted to price a call option $C(t)$ on a zero-coupon bond maturing at time U? The call option has strike K and expiry $T < U$.
The payoff in T is

$$\boxed{(B(T,U) - K)^+.}$$

Applying the Fundamental Asset Pricing Formula

$$
\begin{aligned}
C(t) &= A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{(B(T,U)-K)^{+}}{A(T)}\right|\mathcal{F}_t\right] \\
&= A(t)\left(\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{B(T,U)}{A(T)}\mathbb{1}(B(T,U)>K)\right|\mathcal{F}_t\right] - K\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{1}{A(T)}\mathbb{1}(B(T,U)>K)\right|\mathcal{F}_t\right]\right) \\
&= A(t)\left(\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{A(T)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{1}{A(U)}\right|\mathcal{F}_T\right]}{A(T)}\mathbb{1}(B(T,U)>K)\right|\mathcal{F}_t\right] - K\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{\mathbb{1}(B(T,U)>K)}{A(T)}\right|\mathcal{F}_t\right.\right. \\
&= A(t)\left(\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{\mathbb{1}(B(T,U)>K)}{A(U)}\right|\mathcal{F}_t\right] - K\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{\mathbb{1}(B(T,U)>K)}{A(T)}\right|\mathcal{F}_t\right]\right)
\end{aligned}
$$

Applying the Fundamental Asset Pricing Formula

And that's it. We cannot go any further.
To go any further, we would need to know at time t the joint distribution of $B(T,U)$, $A(U)$ and $A(T)$. This is unlikely, unless we make very explicit model.

One way out of this situation would be to look for a measure $\mathbb{P}^T$ such that the expectation in the fundamental asset pricing formula would be a sole function of the derivative payoff $(B(T,U)-K)^+$.

This idea implies that rather than having the "classic" formula

$$C(t) = A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{(B(T,U)-K)^+}{A(T)}\right|\mathcal{F}_t\right]$$

we would start with the "modified" formula

$$C(t) = B(t,T)\,\mathrm{E}^{\mathbb{P}^T}\left[(B(T,U)-K)^+\middle|\mathcal{F}_t\right].$$

Applying the Fundamental Asset Pricing Formula

To be in a position to use this "modified" formula, we must answer 2 questions:

1. We do not know what $\mathbb{P}_T$ is. In fact, we do not even know if $\mathbb{P}_T$ exists.
2. Given information up to time t, what would $B(T, U)$ be equal to?

Let's start with the second, and easiest, question. If we are at time t and we would like to know the price at some future time T of a bond maturing at time U, we would use the forward price for that bond.

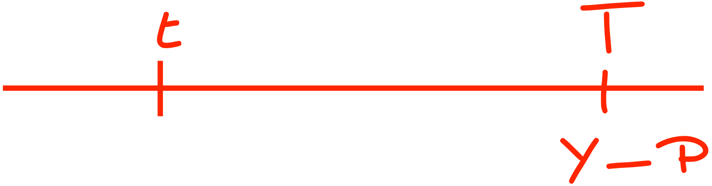
This leads us to a possible answer to our first question. To define $\mathbb{P}_T$, we could look for a *forward martingale measure*, that is, an equivalent martingale measure defined with respect to forward prices.

Hence, to use the *modified* fundamental asset pricing formula, we need to know a little bit about forwards.

Forward Contracts and Forward Prices

Forward contracts are OTC derivatives securities in which the long party has the obligation to buy an agreed upon quantity of an underlying asset (securities, commodities or others) at an agreed upon time and at an agreed upon price called the forward price.

Forward contracts are symmetrical contracts. Therefore, the obligations of the short party mirror those of the long party. The contract is settled at maturity and typically no cash flow is exchanged in the meantime. As they are OTC derivatives, forward contracts are subject to counterparty risk.



Forward Contracts and Forward Prices

Let's say that we want to enter into a (long) forward contract on a financial instrument (stock, bond, currency...) whose value at time t is Y(t). The forward matures at time T. Based on our definition of forward contracts, the payoff $G(T, Y_T)$ of the contract is

$$G(T, Y_T) = Y(T) - F_Y(t, T)$$

where $F_Y(t, T)$ is the forward price of Y determined at time t for delivery at time T.

Forward Contracts and Forward Prices

Plugging this into the fundamental asset pricing formula, we see that the time $t \leq \tau \leq T$ value $\pi_\tau(G)$ of a forward contract entered into at time t is equal to

$$\pi_\tau(G) = A(\tau)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{Y(T) - F_Y(t,T)}{A(T)}\right|\mathcal{F}_\tau\right]$$

This formula can be simplified by noting that $F_Y(t,T)$ is a $\mathcal{F}_t$ measurable

$$\pi_\tau(G) = A(\tau)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{Y(T)}{A(T)}\right|\mathcal{F}_\tau\right] - A(\tau)F_Y(t,T)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{1}{A(T)}\right|\mathcal{F}_\tau\right]$$

and now we know the value of a forward contract for any time $t \leq \tau \leq T$.

Forward Contracts and Forward Prices

The forward price $F_Y(t,T)$ was originally set at time t so that the value of the forward contract at time t is 0. Hence,

$$\cancel{\pi_t(G) =} A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{Y(T)}{A(T)}\right|\mathcal{F}_t\right] = A(t)F_Y(t,T)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{1}{A(T)}\right|\mathcal{F}_t\right] \cancel{= 0}$$

Rearranging,

F

$$\boxed{F_Y(t,T) = \frac{A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{Y(T)}{A(T)}\right|\mathcal{F}_t\right]}{A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{1}{A(T)}\right|\mathcal{F}_t\right]} = \frac{\pi_t(Y)}{B(t,T)}.}$$

where

$$\pi_t(Y) = Y(t) = A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\left.\frac{Y(T)}{A(T)}\right|\mathcal{F}_t\right].$$

that is $Y(t)$ is the value of a claim paying $Y(T)$ at time T.

The forward price is the number that discounted gives us today's value.

Forward Contracts and Forward Prices

If the underlying asset is a zero-coupon bond of maturity $U > T$, the forward price becomes

$$\boxed{F_{B(.,U)}(t,T) = \frac{B(t,U)}{B(t,T)}.}$$

We use also the notation $F_{B(.,U)}(t,T) = F_B(t,T,U)$.

The Forward Martingale Measure

We will define the T-forward martingale measure, or simply *forward measure*, via the Radon-Nikodym derivative λ_t defined as

$$\boxed{\lambda_T = \frac{d\mathbb{P}_T}{d\mathbb{Q}} = \frac{A(\cancel{0})}{A(T)} \frac{B(\cancel{T,} T)}{B(0,T)}}$$

Thus

$$\lambda_t = \frac{d\mathbb{P}_T}{d\mathbb{Q}}_{|\mathcal{F}_t} = \frac{1}{B(0,T)} \mathrm{E}^{\mathbb{Q}} \left[\frac{1}{A(T)} \middle| \mathcal{F}_t \right] = \frac{1}{A(t)B(0,T)} \left(A(t)\, \mathrm{E}^{\mathbb{Q}} \left[\frac{1}{A(T)} \middle| \mathcal{F}_t \right] \right)$$
$$= \boxed{\frac{A(0)}{A(t)} \frac{B(t,T)}{B(0,T)}}$$

Note that λ_t defined through a conditional expectation is a $\mathbb{Q}$-martingale and $\mathrm{E}^{\mathbb{Q}}[\lambda_T] = \lambda_0 = 1$.

The Forward Martingale Measure

We know that the time t value of a zero-coupon bond is given by

$$\frac{B(t,T)}{A(t)} = B(0,T)\exp\left(\int_0^t b(s,T)\cdot dX^{\mathbb{Q}}(s) - \frac{1}{2}\int_0^t |b(s,T)|^2 ds\right)$$

with

$$A(t) = \exp\left(\int_0^t r(s)ds\right).$$

The Forward Martingale Measure

The derivative is such that

$$\begin{aligned}\lambda_t &= \frac{1}{A(t)}\frac{B(t,T)}{B(0,T)} \\ &= \exp\left(\int_0^t b(s,T)\cdot dX^{\mathbb{Q}}(s) - \frac{1}{2}\int_0^t |b(s,T)|^2 ds\right)\end{aligned}$$

It is an exponential martingale with a known exponent.
By Girsanov's theorem, the process X^T defined as

$$X_t^T = X_t^{\mathbb{Q}} - \int_0^t b(s,T)ds$$

is a standard Brownian Motion under the forward measure $\mathbb{P}_T$.
The random process X^T is called the *T-forward Brownian motion*.

Pricing a Derivative Under the Forward Measure

We now have a measure $\mathbb{P}_T$. But before we can use the forward asset pricing formula

$$V(t) = B(t, T)\, \mathrm{E}^{\mathbb{P}_T}\left[\frac{Y}{B(T, T)} \middle| \mathcal{F}_t \right]$$

we need to make sure that it will give the same result as the "classic" fundamental asset pricing formula

$$V(t) = A(t)\, \mathrm{E}^{\mathbb{Q}}\left[\frac{Y}{A(T)} \middle| \mathcal{F}_t \right].$$

Pricing a Derivative Under the Forward Measure

The price of a contingent claim on Y is

$$
\begin{aligned}
\pi_t(Y) &= A(t)\,\mathrm{E}^{\mathbb{Q}}\left[\frac{Y}{A(T)}\middle|\,\mathcal{F}_t\right] \\
&= A(t)\,\mathrm{E}^{\mathbb{Q}}\left[A(\cancel{T})B(0,T)\lambda_T\frac{Y}{A(\cancel{T})}\middle|\,\mathcal{F}_t\right] \\
&= A(t)B(0,T)\,\mathrm{E}^{\mathbb{Q}}\left[\underline{Y\lambda_T}\,|\,\mathcal{F}_t\right]
\end{aligned}
$$

Pricing a Derivative Under the Forward Measure

By extension of Bayes' formula (see for example Musiela, Rutkowsky, *Martingale Methods in Financial Modelling*, Lemma A.0.4)

$$\mathrm{E}^{\mathbb{P}_T}\left[Y \middle| \mathcal{F}_t\right] = \frac{\mathrm{E}^{\mathbb{Q}}\left[Y\lambda_T \middle| \mathcal{F}_t\right]}{\mathrm{E}^{\mathbb{Q}}\left[\lambda_T \middle| \mathcal{F}_t\right]}.$$

Therefore

$$\begin{aligned} V(t) &= A(t)B(0,T)\,\mathrm{E}^{\mathbb{Q}}\left[\lambda_T \middle| \mathcal{F}_t\right]\mathrm{E}^{\mathbb{P}_T}\left[Y \middle| \mathcal{F}_T\right] \\ &= A(t)B(0,T)\lambda_t\,\mathrm{E}^{\mathbb{P}_T}\left[Y \middle| \mathcal{F}_T\right] \\ &= A(t)B(0,T)\frac{A(0)B(t,T)}{A(t)B(0,T)}\,\mathrm{E}^{\mathbb{P}_T}\left[Y \middle| \mathcal{F}_T\right] \\ &= B(t,T)\,\mathrm{E}^{\mathbb{P}^T}\left[Y \middle| \mathcal{F}_T\right] \end{aligned}$$

Pricing a Derivative Under the Forward Measure

The forward (martingale) measure $\mathbb{P}_T$ is defined in terms of the equivalent martingale measure $\mathbb{Q}$ via the Radon-Nikodym derivative

$$\boxed{\frac{d\mathbb{P}_T}{d\mathbb{Q}} = \frac{A(0)}{A(T)} \frac{B(T,T)}{B(0,T)}.}$$

Key Fact (European derivative)

The price of a European derivative expiring at time T with payoff Y is

$$\pi_t(Y) = \boxed{B(t,T)\,\mathrm{E}^{\mathbb{P}_T}\left[Y \middle| \mathcal{F}_t\right]}$$

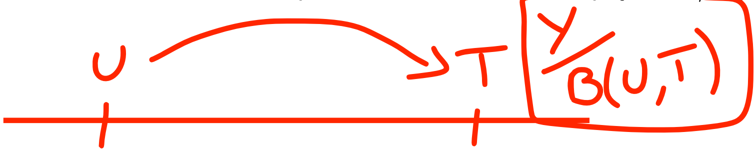
Pricing a Derivative Under the Forward Measure

Key Fact (European derivative)

The price of a European derivative expiring at time $U < T$ with payoff Y is

$$\boxed{\pi_t(Y) = B(t,T)\,\mathrm{E}^{\mathbb{P}_T}\left[\left.\frac{Y}{B(U,T)}\right|\mathcal{F}_t\right]}$$

Strategy: Invest from U into zero-coupon bond to obtain the payoff $Y/B(U,T)$ in T.



Pricing a Call on a Zero-Coupon Bond

We express the call payoff at time T, not in terms of the zero-coupon bond, but in terms of a forward on the zero-coupon bond as

$$(B(T,U)-K)^+ = (\boxed{F_B(T,T,U)} - K)^+$$

where $F_B(t,T,U)$ is the forward price at time t for settlement at time T of a zero-coupon bond maturing at time $U > T$. Note that $F_B(T,T,U)$ is the instantaneous forward price at time T, which is equal to the spot price $B(T,U)$.

Pricing a Call on a Zero-Coupon Bond

Applying the forward pricing formula, we deduce that the zero-coupon forward price $F_B(t, T, U)$ is given by

$$\boxed{F_B(t, T, U) = \frac{B(t, U)}{B(t, T)}}$$

Pricing a Call on a Zero-Coupon Bond

As a result, the $\mathbb{Q}$-dynamics of the forward price is given by

$$\begin{aligned} F_B(t,T,U) &= \frac{B(0,U)}{B(0,T)} \exp\left(\int_0^t (b(s,U)-b(s,T)) \cdot dX^{\mathbb{Q}}(s) - \frac{1}{2}\int_0^t |b(s,U)|^2 - |b(s,T)|^2 ds\right) \\ &= F_B(0,T,U) \exp\left(-\int_0^t (b(s,U)-b(s,T)) \cdot b(s,T) ds\right) \\ &\quad \exp\left(\int_0^t (b(s,U)-b(s,T)) \cdot dX^{\mathbb{Q}}(s) - \frac{1}{2}\int_0^t |b(s,U)-b(s,T)|^2 ds\right) \end{aligned}$$

or

$$\frac{dF_B(t,T,U)}{F_B(t,T,U)} = (b(t,U)-b(t,T)) \cdot dX^{\mathbb{Q}}(t) - (b(t,U)-b(t,T)) \cdot b(t,T)dt.$$

Pricing a Call on a Zero-Coupon Bond

Deriving the dynamics of $F_B(t, T, U)$ under the $\mathbb{Q}$-measure is a promising start. As we are going to price the call option using the forward asset pricing formula, we need to know the dynamics of the forward price $F_B(t, T, U)$ under the forward measure $\mathbb{P}_T$. Recalling that

$$X_t^T = X_t^{\mathbb{Q}} - \int_0^t b(s, T)ds$$

is a standard Brownian Motion under the forward measure $\mathbb{P}^T$, we immediately get

$$\frac{dF_B(t, T, U)}{F_B(t, T, U)} = (b(t, U) - b(t, T)) \cdot dX^T(t).$$

and

$$\boxed{F_B(t, T, U) = F_B(0, T, U)\mathcal{E}_t\left(\int_0^{\cdot} (b(s, U) - b(s, T)) \cdot dX^T(s)\right)}$$

which implies that the forward price is a martingale under the forward measure.

Pricing a Call on a Zero-Coupon Bond

We now have all we need to solve the Call option pricing problem using the forward asset pricing formula

$$C(t) = \boxed{B(t,T)\,\mathrm{E}^{\mathbb{P}_T}\left[(F_B(T,T,U)-K)^+\middle|\,\mathcal{F}_t\right].}$$

In the case of a call on a zero-coupon bond and deterministic volatility, a Black-Scholes-type formula exists.

Pricing a Call on a Zero-Coupon Bond

Key Fact (European Call)

When the volatility $b(., V)$ is deterministic (for all $V \leq \bar{T}$), the time t price of a European Call expiring at time T and with strike K, written on a zero-coupon bond maturing at time $U > T$ is given by the following Black-Scholes type of formula

$$C(t) = B(t, U)N(d_+(B(t, U), t, T)) - KB(t, T)N(d_-(B(t, U), t, T))$$

where

$$d_\pm(b, t, T) = \left(\ln\left(\frac{b}{KB(t, T)}\right) \pm \frac{1}{2}\sigma_U^2(t, T)\right)\frac{1}{\sigma_U(t, T)}$$

and

$$\sigma_U^2(t, T) = \int_t^T |b(s, U) - b(s, T)|^2 ds.$$

Call Pricing in Practice: Numerical Solutions 2

Monte Carlo on the zero-coupon bonds B (deterministic b), e.g. Hull-White one-factor model.

The random variable $\int_0^t b(s, T) \cdot dX^{\mathbb{Q}}(s)$ normally distributed. When b is deterministic, its variance is explicitly known: $\alpha^2(t, T) = \int_0^t |b^\theta(s, T)|^2 ds$.

When possible, significantly more efficient numerically than working on r. More regularity: $B = \int r$, one level of regularity more!

The quantities $B(0, T)$ provided by the market. The quantity $r(0)$ undefined in practice! It is the limit for $t \to 0$ of a quantity known only discreetly (at best daily maturities).

Call Pricing in Practice: Numerical Solutions 2

A Monte-Carlo algorithm on the bond looks like (with $Z(t,T) = B(t,T)/A(t)$)

- $Z(0,T) = B(0,T)$ and $Z(0,U) = B(0,U)$
- Simulations $j = 1, \ldots, N$
  - Time steps not needed
  - Simulate $N(0,1)$ random variables W_j
  - Simulate bonds $(X = T, U)$

$$Z_j(T,X) = B(0,X) \exp\left(-\alpha(T,X)W_j - 0.5 * \alpha^2(T,X)\right)$$

- Value of the bond option:

$$C = \frac{1}{M} \sum_{j=1}^{M} (Z_j(T,U) - KZ_j(T,T))^+.$$

Advantage with respect to Monte-Carlo on r: $B(0,T)$ given by the market, no time discretisation, no differentiation, no integration.

Martingales and Fixed Income Valuation

Why this approach?

If we want the bond price to have a lognormal-type behaviour in a short rate model, the bond volatility function $b(t, T)$ may look like a "fudge function". This will however motivate us to turn our models around and specify a bond dynamics first, and then deduce a dynamics of interest rates. This approach forms the base of forward rate models such as the HJM class of models.

Forward measure. The existence of the forward measure and the critical role played by forwards in the pricing of interest rate derivatives also provides a powerful motivation for looking at the term structure of forward (as opposed to spot) rates (see the HJM class of models).

As long as the bond price follows a geometric dynamics, irrespective of the specific interest rate model we chose, the value of a bond derivative will always be of the same form. Natural question: if we assume a geometric dynamics for the bond price, how many interest rate models do we have access to? What is the most general interest rate model we can find such that the bond price follows a geometric dynamics?

What is next? Forward Rate Model

The answer to this question, and next chapter in the development of interest rate models, is the derivation of models of the forward rate dynamics. This critical step was achieved by Heath, Jarrow and Morton (1992) and then further developed by Brace, Gatarek and Musiela (1997).

The key attraction of forward rate models is they start from a (nice) geometric dynamics for the zero-coupon bond price and then deduce the behaviour of the term structure of forward rates; they are a “meta”-model which encompasses all existing interest rate models; as such, you can use them to price or manage the risk of anything, from vanilla derivatives to complex fixed income portfolios (which are heavily dependent on an accurate modelling of the term structure).

Forward rate models are not necessarily different from short rate models. The same practical model can have several representation: short rate, forward instantaneous rate, market rate, discount factors.

What is next? Forward Rate Model

The key problems related to forward rate models: mathematically sophisticated, sometimes too sophisticated for the applications at hand (such as pricing vanilla derivatives); potentially non-Markov. A good part of the mathematics we have are based on Markov models. Hence, we need to choose our parameters carefully and make assumptions to ensure that the forward rate models we work with are indeed Markov.

Meta-model: no clear indication of which form to use and when to use it.

What is next? Forward Rate Model

In this lecture, we have seen...

- the pricing of interest rate products in a probabilistic setting; the equivalent (spot) martingale measures;
- the fundamental asset pricing formula for bonds;
- the dynamics of bond prices;
- the forward measure;
- the fundamental asset pricing formula for derivatives on bonds;

Credit Spread Pricing
Caplet Volatility Stripping

by Dr Richard Diamond

CQF FINAL PROJECT 2020

Fundamentals + Techniques

Preparation

- Start collecting data and planning the project for you.
- It is up to you to source and clean the suitable data.

  Webex session on Equities Data. If you can't get hold of some data: **make reasonable assumptions**, even generate the data.

  Set your option strikes and maturities, clean rows with missing rates in BOE data, assume reasonable CDS spreads, etc.

- A. You can adopt code for specific partial tasks, not model as a whole, amending it for your purpose (not copy/paste). B. Use ready libraries with expertise – quadratic optimisation, kernels, etc – methods have to be suited to the task. C. You are welcome to implement complex numerical methods vs. ready solution if able to.

Numerical Techniques

Implement numerical techniques from the first principles as necessary. Pricing a spread/CVA from simulated curve/optimal allocation is the result.

What to code: pricing formulae, Black-Litterman calculation, SDE simulation, matrix form regression, Engle-Granger, interpolation, numerical integration, Cholesky, t-copula formula, CDS bootstrap, features computation...

Use ready solutions for: covariance shrinkage, nearest correlation, ML numerical methods (eg, decision trees, neural nets), low latency RNs, kernel density (cdf estimation), QR-decomposition (PCA), EGARCH estimation, Johansen Procedure...

The lists are not exhaustive.

Project Report

- A full **mathematical description** of the models employed as well as numerical methods. Remember *accuracy and convergence*!
- Results presented using **a plenty of tables and figures**, which must be interpreted not just thrown at the reader.
- **Pros and cons** of a model and its implementation, together with possible improvements.
- **Demonstrate 'the specials'** of your implementation: own research, own coding of complex methods, use of the industrial-strength libraries of C++, Python or VBA + NAG.
- Instructions on how to use software if not obvious.
  The code must be thoroughly tested and well-documented.

CQF Electives

See Project Brief for the full current table.

Counterparty Credit Risk – CR, IR	CDS, survival probabilities and hazard rates reviewed. Three key numerical methods for quant finance pricing (Monte-Carlo, Binomial Trees, Finite Difference). Monte Carlo for simple LMM. Review of Module Five on Credit with a touch on the copula method. **Outcome:** covers CVA Computation clearly and reviews of credit spread pricing techniques.
Risk Budgeting – PC primary choice	Reviews the nuance of Modern Portfolio Theory, ties in VaR and Risk Decomposition with through derivations and expectation algebra. Gives simple examples of figures you need to compute and then combine with portfolio optimisation. Risk-budgeting portfolio from Video Part 10.
Advanced Volatility Modeling – LV primary	Considers the main kinds of PDEs (stochastic volatility and jump diffusion) and their analytical solution: the main approach to solve stochastic volatility (Heston model) is via Fourier Transform. In-depth on integration **Outcome:** Local Volatility topic offers a classic pricing PDE, which can be solved by techniques from this elective.
C++ – Dev	Consider this a revised version of C++ Primer/ initial certificate course.

Final Day is ~~12th~~ 13th July 2020

Don't Extend Your Luck!

1. Introduction to CQF Final Project

2. Pricing a credit product: kth-to-default Basket CDS

3. Aspects of LMM Calibration

4. Interest Rate Swap: Exposure Profile, OIS Discounting

Pricing for Credit Products

Case Study

In 2008 AIG Financial Products Corp had notional CDS exposure to highly rated CDSs of roughly **$450-500 billion** concentrated on banks, with about $60 billion exposed to subprime mortgages.

Viewing CDS as a leveraged purchase of bond, consider the implications.

- Outright purchase of $450 billion worth of corporate bonds (especially subordinated bank debt) would have attracted everyone's attention: C-level management, counterparties, regulators. **CDS positions did not**.

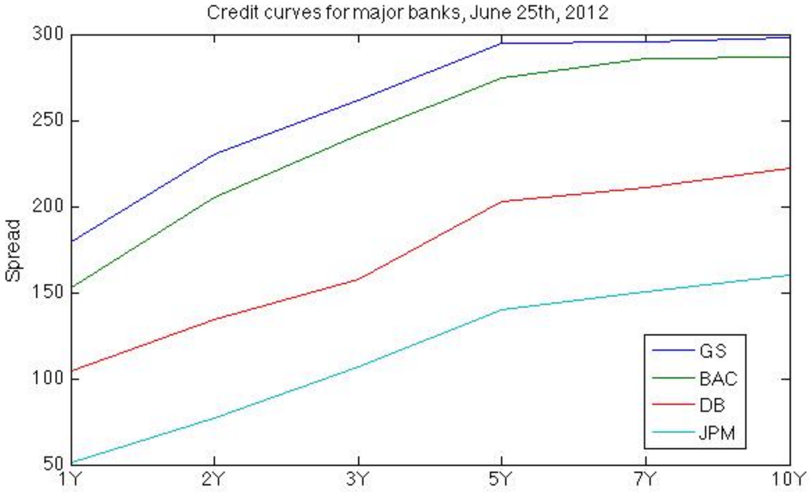
- Consider the exposure as if you were a risk manager: CDSs typically have 5Y maturities, and rates were about 5.5% in 2008.

  A five-year par bond with a rate of 5.5% has a sensitivity to credit spreads, or credit DV01, of about $435 per **one** basis point for $1 million notional.

  $450 billion notional will have exposure of $200 million per **one** basis point. A move of 50 bps would generate roughly $10 billion in losses. Mid-2007 through early 2008, spreads on five-year AAA financial issues rose from about 50 bps to about 150 bps...

Re-worked from: *A Practical Guide to Risk Management*, Coleman (2011)

Single-name CDS: credit curves



What is a **Basket CDS**?

Structured credit product, an OTC credit exotic. Product characteristics and pricing methodology give a practical insight in how to price tranched products (pools of securities), with correlation.

If the *k* number of defaults occur

- the contract terminates, the protection buyer receives

$$\text{LGD} = (1 - R) \times \frac{1}{5} \times N$$

Protection seller receives $PL = s \times 10$ million per year. Paid periodically in arrears (vs. upfront fee CDS). Five reference names, 2m notional each, with maturity T = 5 years.

Potential Pricing

1st to default 'Equity'	22 bps Gaussian	t copula
2nd to default	10 bps	
5th to default	1.6 bps	

$s = \frac{DL}{PL}$ is computed separately for each k-th to default. The spread is an expectation over the joint loss distribution (with simplification).

Credit Spreads

There are spreads on standardised CDS indices, such as CDX North America and iTraxx Europe (5Y maturity) baskets of 50 top credit quality corporate names.

Why interest in synthetic credit? **Selling protection** (CDS) means the same exposure as **buying a bond** – leveraged,eg, borrowing the initial purchase price of the bond. Receive premiums until the maturity or default, and pay out the principal in case of default.

Price a fair spread for a basket of five reference names by sampling default times from both, Gaussian and Student's t copulae.

Hazard Rates Data (five snapshots)

What does default time τ_k depend on?

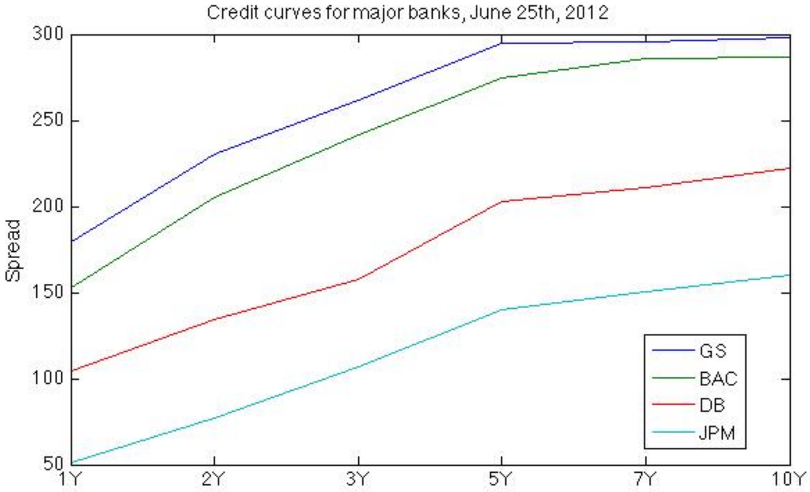
$$\boldsymbol{\tau} \sim Exp(\lambda_{1Y}, \ldots, \lambda_{5Y})$$

For each reference name, we have to bootstrap hazard rates from the credit curve 'today'.

- 5 single-name par spread CDS with maturities up to 5Y.
- Matching discounting curve for 1Y,..., 5Y points.
- assume recovery rate $R = 40\%$ (there are LGD models)

CDS Lecture illustrates how to bootstrap implied survival probabilities and thus, hazard rates.

Credit Curves (term structure), bps



Survival Probabilities

The cumulative survival probability relates to hazard rate function:

$$\log P(0, t_m) = -\int_0^{t_m} \lambda_s ds = -\sum_{i=j}^{m} \lambda_j \Delta t_j \tag{1}$$

- $P(0, t_m)$ is survival probability up to time t_m
- λ_j is hazard rate between $j - 1$ to j
- Δt_j is gap between each period, likely to be 1 year

$P(0, t_m)$ is 'like' a discounting factor.

Hazard Rate for Each Tenor

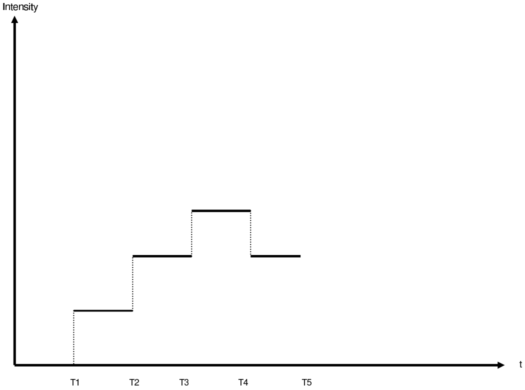
Assuming hazard rate function as piecewise constant, we bootstrap iteratively for each tenor Δt_j (year)

$$\begin{aligned}\lambda_1 &= -\frac{1}{\Delta t}\log P(0, t_1) \\ \lambda_m &= -\frac{1}{\Delta t}\log P(0, t_m) - \sum_{j=1}^{m-1}\lambda_j\end{aligned} \tag{2}$$

We can express the intensity as a ratio of survival probabilities:

$$\begin{aligned}\lambda_m &= -\frac{1}{\Delta t}\log P(0, t_m) + \frac{1}{\Delta t}\log P(0, t_{m-1}) \\ \lambda_m &= -\frac{1}{\Delta t}\log \frac{P(0, t_m)}{P(0, t_{m-1})}\end{aligned} \tag{3}$$

IHP calibrated on single-name CDS



For each name, we have a term structure $\widehat{\lambda}_{1Y}, \ldots, \widehat{\lambda}_{5Y}$, a calibration to the Inhomogenous Poisson process (IHP).

Exponential inter-arrival times

- Suppose a simulation gives us correlated $(u_1, \ldots, u_5)$.
- Our task is to convert $u_i \to \tau_i$, done individually for **each** reference name by using (five) hazard rate from the curve

$$
\begin{aligned}
\tau_{\text{Name 1}} &\sim Exp(\lambda_{1Y}, \ldots, \lambda_{5Y}) \\
\tau_{\text{Name 2}} &\sim Exp(\lambda_{1Y}, \ldots, \lambda_{5Y})
\end{aligned}
$$

Exponential CDF is $u = 1 - e^{-\lambda\tau}$ so,

$$\log(1 - u) = -\lambda_\tau \tau$$

There is input u but **two unknowns** λ_τ and τ.

Marginal default time (for each name)

$$\boxed{\tau = t_{m-1} + \delta t}$$

1. First, we find the year of default,
   i.e., determine that default occurs between t_{m-1} and t_m.
2. Second, we estimate the year fraction δt or use accruals.

Year of default

- Iterate adding up hazard rates λ_j

$$\tau = \inf\left\{ t > 0 \,:\, \log(1-u) \geq -\sum^{t} \lambda_m \right\}$$

where default occurs if inequality holds and

$t_{m-1} \leq \tau \leq t_m$

Comparison is done on negative scale because $\log(1-u) < 0$.

- If the inequality **holds** after adding λ_m then default occurs.

Validating example

- Using absolute values to compare on positive scale

$$|\log(1-u)| \le \sum \lambda$$

- We construct a validation table, where small u implies a default

u	$\lvert log(1-u)\rvert$
0.90	2.3
0.50	0.69
0.25	0.2877
0.10	0.1054
0.05	0.0513

Exact default time

Exact default time $\tau = t_{m-1} + \delta t$ requires year fraction δt

$$1 - u = \exp\left(-\int_0^{t_{m-1}+\delta t} \lambda_s ds\right) = P(0, t_{m-1}) \exp\left(-\int_{t_{m-1}}^{t_{m-1}+\delta t} \lambda_s ds\right)$$

$$\log\left(\frac{1-u}{P(0, t_{m-1})}\right) = -\delta t \lambda_m$$

$$\delta t = -\frac{1}{\lambda_m} \log\left(\frac{1-u}{P(0, t_{m-1})}\right) \tag{4}$$

What matters in practice is how default event is determined and settled. Assume $\delta t = 0.5$ is called accruals.

$$\boxed{\tau = t_{m-1} + \delta t}$$

Copula. Correlation of Default Events

Marginal Distributions

Question

What is the distribution of default time is for each reference name?

Answer

It is an **Exponential Distribution** parametrised empirically by a set of five hazard rates (piecewise constant, per year).

$$
\begin{aligned}
\tau_{\text{Name 1}} &\sim Exp(\lambda_{1Y}, \dots, \lambda_{5Y}) \\
\tau_{\text{Name 2}} &\sim Exp(\lambda_{1Y}, \dots, \lambda_{5Y})
\end{aligned}
$$

Copula Method

A great deal of flexibility by separating:

- **Marginal distributions** for default times τ_i
- **Dependence structure** (correlation: linear, rank, calibrated based on MLE by an optimiser)

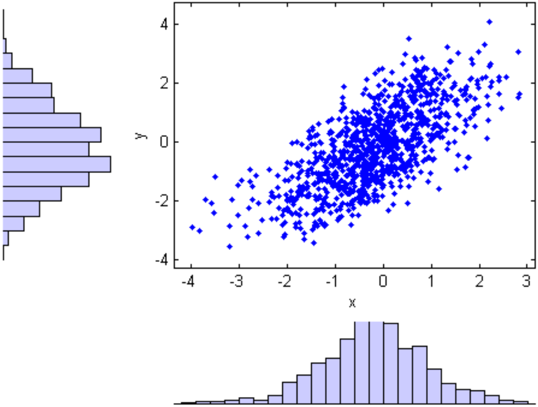
The joint distribution for k-th to default time across all reference names $\tau_k \sim F_k(t_1, t_2, \dots, t_n)$ has **no closed form**. However,

$$F(x_1, x_2, \dots, x_n) \equiv C(u_1, u_2, \dots, u_n)$$

Let's review the concept of **copula**.

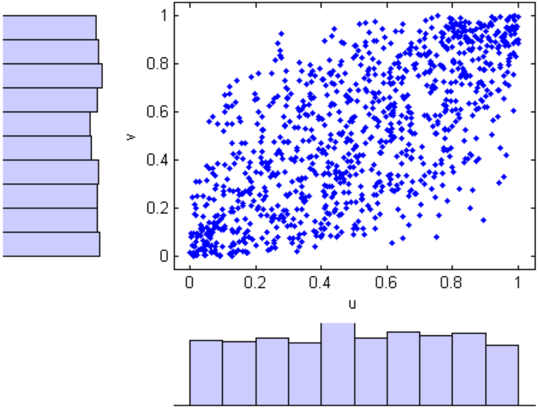
Joint Student's t Distribution

Joint distributions are cumbersome to work with and might have no analytical solution for CDF and ICDF.



Student's t Copula

Applying Student's t **CDF** to marginals means re-scaling to a uniform [0, 1] projection. This should look familiar:



Sampling from copula

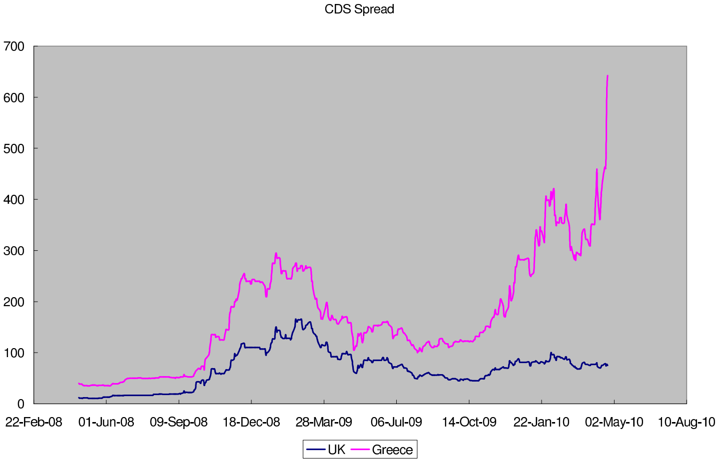
Copula method is a way to structure Monte Carlo: generate a correlated set $(u_1, u_2, u_3, u_4, u_5)$.

The approach is non-parametric which gives it a flexibility. The joint distribution of default time τ_k **unknown** for each case of *k-th to default*.

$$\tau_k \sim F_k(t_1, t_2, \dots, t_n)$$

- Implementation boils down to Cholesky decomposition $\widehat{\boldsymbol{\Sigma}}_\rho = \mathbf{AA}'$
- We generate independent Normal RNs, and impose correlation by $\boldsymbol{X}^{Sim} = \boldsymbol{AZ}$, then convert back $\boldsymbol{U}^{Sim} = \Phi(\boldsymbol{X}^{\boldsymbol{Sim}})$

Historic Credit Spreads, bps



Correlating *levels* of credit spreads is spurious (unit root variables!).

Data for Default Correlation

Estimate correlation from **historical PD** data, as implied by credit spreads.

- 5Y tenor is a good reference point;
- daily or weekly changes; one-two year period.

Getting daily historical quotes for CDS and discounting a year or two back can be a challenge. Common substitutes:

- historical returns data (equity, bond yields) OR *base correlation* available from traded correlation instruments.

Distribution Fitting:

To apply correlation/transform to uniform easily we need data close to Normal histograms

- Obtain the time series for 5Y tenor point (CDS, PD, prices) and plot histograms. Non-stationary variables produce bi-modal histograms.

- Convert variables to *changes* ΔCDS, ΔPD or returns, subtracting means if necessary, and generate histograms.

  Hazard rate are already log-differences of the survival probabilities $\propto -\log \frac{P(0,t_m)}{P(0,t_{m-1})}$.

This is experimentation: demo can be provided but coding not to be shared.

Pseudo-samples: Normalising

Maximum Likelihood requires sample data transformed into uniform pseudo-samples $\mathbf{U}_t^{Hist}$. A simple recipe:

- Take data and convert into standardised changes (returns) **Z**.

$$\Delta\text{CDS},\ \Delta\text{PD} \rightarrow \mathbf{Z}^{\mathbf{Hist}}$$

- Under all assumptions (Normal distribution, volatility known)

$$\mathbf{U}_t^{Hist} = \Phi(\mathbf{Z}^{\mathbf{Hist}})$$

The drawback is that histograms of $\mathbf{U}_t^{Hist}$ might not be that uniform! Interferes with correlation.

Pseudo-samples: Kernel Smoothing (for pdf)

The proper recipe is $\mathbf{U}_t^{Hist} = \widehat{F}(\mathbf{PD})$ – you are not limited to PD data. Also computable $\mathbf{Z} = \Phi^{-1}(\mathbf{U})$.

This is numerically involved and involves at least two steps,

- fitting a pdf to a kernel function
- numerical integration over kernel to compute Empirical CDF $\widehat{F}()$

Matlab *ksdensity()* gives ready implementation. Use 'ecdf' and vary bandwidth 'bw' to obtain good-looking uniform histograms (each reference name). If things not uniform try more amenable data, eg, returns.

Dependence Fitting:

Plot 2D scatter plots (one reference name vs. another)

- Original data, variables such as CDS_{5Y}, prices – is this a valid move?
- Changes ΔCDS if using those.
- Pseudo-samples **U** – what does this plot represent?

The plots helps to visualise dependence, now moving on to formal estimation of 5×5 correlation matrix.

Correlation Matrix

- Once we got Normal $\mathbf{Z}$ from pseudo-samples $\mathbf{U}_t^{Hist}$ it is straightforward to calculate linear correlation matrix $\mathbf{\Sigma} = \rho(\mathbf{Z})$. Good enough for Gaussian copula.

- But t copula sampling requires **rank correlation**.

  Spearman's rho is estimated on pseudo-samples $\mathbf{\Sigma}_S = \rho(\mathbf{U})$. A separate formula $\mathbf{\Sigma}_\tau = \rho_\tau(\mathbf{X})$ is defined for Kendall's tau.

  Use $\rho = 2\sin\left(\frac{\pi}{6}\rho_S\right)$ and $\rho = \sin\left(\frac{\pi}{2}\rho_\tau\right)$ 'to linearlise' rank correlation.

The inferred linear correlation matrix is not guaranteed to be positive definite as required for Cholesky – so the nearest correlation matrix is obtained.

t copula

Fitting t copula gives an explicit exercise in **Maximum Likelihood**.

Two parameters to calibrate: correlation matrix $\widehat{\boldsymbol{\Sigma}}$ and ν, degrees of freedom. This is what ready software does and comes up with unrealistic correlations by optimisaiton.

$$\underset{\nu}{\operatorname{argmax}} \left\{ \sum_{t=1}^{T} \log c(\mathbf{u}_t^{Hist}; \nu, \widehat{\boldsymbol{\Sigma}}) \right\}$$

$\mathbf{u}'_{Hist}$ is a 1×5 row vector of observations for five reference names .

Log-likelihood of t copula

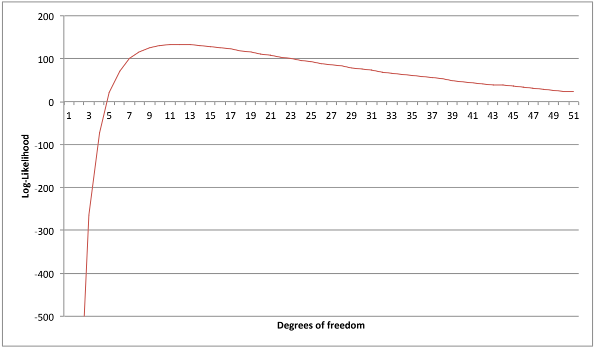


Figure: Use rank correlation matrix input, repeat calculation of the sum of log-likelihoods *for each value* of $\nu = 1 \dots 25$ and plot.

t copula density

$$c(\mathbf{u};\nu,\widehat{\boldsymbol{\Sigma}}) = \frac{1}{\sqrt{|\Sigma|}}\frac{\Gamma\left(\frac{v+n}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}\left(\frac{\Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+1}{2}\right)}\right)^{n}\frac{\left(1+\frac{T_v^{-1}(\boldsymbol{u}')\boldsymbol{\Sigma}^{-1}T_v^{-1}(\boldsymbol{u})}{v}\right)^{-\frac{v+n}{2}}}{\prod_{i=1}^{n}\left(1+\frac{T_v^{-1}(u_i)^2}{v}\right)^{-\frac{v+1}{2}}}$$

- $\mathbf{u}'$ is a row vector of 1×5 that represents an observation of spreads (scaled) for five reference names on a given day.
- the denominator is calculated element-wise by drawing u_i from $\mathbf{u}'$.
- the nominator of the last term produces a scalar $1\times 5\ \times 5\times 5\ \times 5\times 1$.
- in VBA use *EXP(GAMMALN())*. **ICDF** for Student's t T_v^{-1} code in Q&A, apply elementwise.

Sampling from Gaussian copula

1. Compute decomposition of correlation matrix $\widehat{\mathbf{\Sigma}} = \mathbf{A}\,\mathbf{A}'$.
   Use the simplest option of Cholesky decomposition if the matrix is positive definite.
2. Draw an n-dimensional vector of independent standard Normal variables $\mathbf{Z} = (z_1, \cdots, z_n)'$.
3. Compute a vector of correlated variables by $\mathbf{X} = \mathbf{AZ}$.
4. Use Normal CDF to map to a uniform vector $\mathbf{U} = \Phi(\mathbf{X})$.

Convert each uniform variable to default time $u_i \to \tau_i$ using hazard rates structure for each name.

Inefficiency: quasi RN generators generate uniform variables, which we convert to Normal plane to impose correlation, and then convert back to uniform.

Sampling from t copula

Differences are **rank correlation** and **chi-squared RN**.

1. Compute decomposition of correlation matrix $\widehat{\boldsymbol{\Sigma}} = \mathbf{A}\,\mathbf{A}'$.
2. Draw an n-dimensional vector of independent standard Normal variables $\mathbf{Z} = (z_1, \cdots, z_n)'$.
3. Draw an independent chi-squared random variable $s \sim \chi^2_\nu$. Compute n-dimensional Student's t vector $\mathbf{Y} = \mathbf{Z}/\sqrt{\frac{s}{\nu}}$.
4. Impose correlation by $\mathbf{X} = \mathbf{AY}$.
5. Map to a correlated uniform vector by $\mathbf{U} = T_v(\mathbf{X})$ using t CDF.

Why t copula?

1. If your calibrated (or reasonably assumed) d.f. $\nu = 7$.
2. The chi-squared random variable $s \sim \chi^2_\nu$ is obtained by drawing ν squared Normal random variables, separately

$$s = Z_1^2 + Z_2^2 + ... + Z_7^2$$

t copula means stronger co-movement. If simulated $u_1 = 0.1$, then $u_2 \approx 0.1$ is likely.

Think of the impact on the spread: multiple defaults together raise the spread for k-th to default, compared to Gaussian.

Where is the copula?

Question

Usually at this stage, a question is asked: “Where in these algorithms is the copula?”

Answer

What we do with imposing correlation by Cholesky result $\boldsymbol{A}$ is equivalent to **factorisation** of the copula into a set of linear equations.

Check CDO lecture for two-dimensional Cholesky solution for $\boldsymbol{A}$ (p.52)

$$\begin{pmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{pmatrix}$$

Review [Reference]

We have covered how to:

- estimate appropriate correlation matrix
- sample a correlated random $U = (u_1, \ldots, u_n)$ from copula
- convert each random variable into default time $u_i \rightarrow \tau_i$

In our simulation of correlated default events, their marginal distributions $\tau_i \sim Exp(\widehat{\lambda})$ are kept separately from dependence structure, a linear correlation matrix $\widehat{\boldsymbol{\Sigma}}$.

The joint distribution for k*th*-to-default time across all reference names $\tau_k \sim F_k(t_1, t_2, \ldots, t_n) \equiv C(u_1, u_2, \ldots, u_n)$ has been represented by a factorised copula (Cholesky linear system).

Spread Computation

kth-to-default spread

Par spread of k*th*-to-default swap is derived by equating DL $=$ PL.

$$s = \frac{< DL >}{< PL_{\$} >} = \frac{(1-R)\sum_{i=1}^{m} Z(0,t_i)\,\left(F_k(t_i) - F_k(t_{i-1})\right)}{\Delta t \sum_{i=1}^{m} Z(0,t_i)\,\left(1 - F_k(t_i)\right)}$$

Assume, we simulated default times $(\tau_{N1}, \tau_{N2}, \tau_{N3}, \tau_{N4}, \tau_{N5})_{1..10{,}000}$

The joint distribution $\tau_k \sim F_k(t_1, t_2, \dots, t_n)$ remains **unknown**. It is likely to be a different distribution for each kth-to-default instrument.

Total Expected Loss

Structured credit pricing is simplified by introducing the **Total Expected Loss**, an expectation over the joint distribution,

$$\mathbb{E}[F_k(t)] = L_k \qquad L_i - L_{i-1} = \frac{1}{5} \times \text{Notional}$$

$$\mathbb{E}[s] = \frac{(1-R)\sum_{i=1}^{m} Z(0,t_i)\,(L_i - L_{i-1})}{\Delta t \sum_{i=1}^{m} Z(0,t_i)\,(NP - L_i)}$$

To satisfy the expectation, the fair spread is calculated using Monte-Carlo.

Upon k-th to default, the notional payment is made by protection seller for the defaulted entity.

- s is fair spread of the contract paid $\frac{1}{dt}$ times per annum until τ_k **or** maturity. $\Delta t \approx t_i - t_{i-1}$ is an accrual factor.
- summation in the spread is over m years, with $\tau = t_{m-1} + \delta t$.
- $Z(t, T)$ is a risk-free zero coupon bond price as discount factor.
- R is recovery rate, $LGD = 1 - R$.
- $NP = 1$ is notional principal. We invest $\frac{1}{5}$ of notional in each name.

Spread Computation: 1st to default, τ_1

Loss Function per time period $L_i - L_{i-1} = \frac{1}{5} \times NP$. This simplifies computation.

$$s = \frac{(1-R)Z(0,\tau_1) \times \frac{1}{5}}{Z(0,\tau_1)\,\tau_1 \times \frac{5}{5}}$$

For $\tau_1 < 5$ years, $Z(0,\tau_1)$ in the DL numerator coincides with one in PL denominator. We keep it in both expressions because we compute average PL and DL separately.

Each kth-to-default basket is priced as **a separate instrument**. Simulations can be saved and re-used but N_{sim} likely to be different. Price by both, Gaussian and t copulæ.

Spread Computation: 1st to default, table

If default time $\tau_k \geq 5$ years then $DL = 0$ but the paid premium has to be discounted. Assuming annual payment, discretisation goes

$$Z(0,1) \times 1 + Z(0,2) \times 1 + ... + Z(0,5) \times 1$$

0	$(1-R)/5$	0	0	0	$(1-R)/5$	$\dots$
$DF \times 5$	τ_1	$DF \times 5$	$DF \times 5$	$DF \times 5$	τ_1	$\dots$

Very small default times τ_k lead to large spreads and interfere with convergence. Can introduce a floor $\tau_k = \max(\hat{\tau}_k, 0.25)$.

Average DL and PL across simulations **separately**, and calculate the spread one time. Done to improve convergence.

Spread Computation: 2nd to default

2nd-to-default also protects from the loss in **single name**

$$s = \frac{(1-R)Z(0,\tau_2)\times\frac{1}{5}}{Z(0,\tau_1)(\tau_1-0)\times\frac{5}{5}+Z(0,\tau_2)(\tau_2-\tau_1)\times\frac{4}{5}}$$

Poses problem of 1) more than two defaults within the same increment of time (year) and 2) removal of referenced entity (non-removal has the simpler PL $= Z(0,\tau_2)\tau_2$).

Reference entities that have defaulted *before* k-th default are removed from the portfolio, reducing its value by $\frac{1}{5}\times NP$ each.

The rule is consistent with the Removal of Defaulted Reference Entity provisions of ISDA Terms.

Model Validation

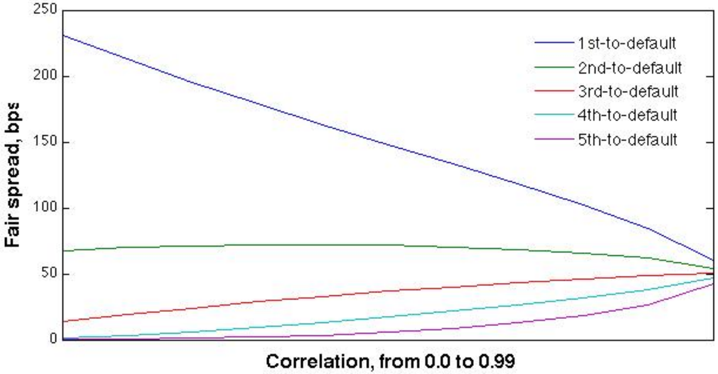
- The fair spread for k-th to-default Basket CDS should be less than k-th -1. Why?

- **Risk and Sensitivity Analysis** of the spread is important
  1. default correlation among reference names: either stress-test by constant high/low correlation or $\pm$ percentage change in correlation from the actual estimated levels.
  2. credit quality of each individual name (change in credit spread, credit delta) as well as recovery rate.

- Correlation matrix is key input, so make sure to explain:
  1. historical sampling of default correlation matrix, and
  2. choice of the stress-testing levels of correlation, i.e., what kind of event they represent

Sensitivity to constant correlation

As default correlation increases to very high levels, spreads for different kth-to-default instruments lapse. Why?



Which default correlation levels have you obtained from linear and rank correlation measures?

Data Requirements - Reference

1. *A snapshot* of credit spreads on a given day is used in estimation of hazard rates:
   - For each reference name, the term structure of hazard rates for 1Y, 2Y,...5Y (non-cumulative) parametrises the distribution of default time τ.
2. *Historical* credit spreads data is needed for estimation of the (inferred) linear correlation matrix of PD.
   - Alternative estimation of default correlations is possible. Please see below and consult with the Q&A.
3. Discounting curve data is necessary for both, hazard rates bootstrapping and basket spread s calculation. Approximate.

Basket CDS Implementation Step-by-Step

1. For each reference name, bootstrap implied default probabilities from quoted CDS and convert them to hazard rates.
2. Estimate the appropriate inputs for 'sampling from copula', i.e., correlation matrix and degrees of freedom.
3. For each simulation, repeat the following routine:
    1. Sample a vector of correlated uniform random variables – you will need to implement sampling from both Gaussian and Student's t copula separately.
    2. Use hazard rates of each reference name to convert the corresponding uniform variable of u_i into exact default time τ_i.
    3. Based on τ_k calculate the discounted values of premium and default legs.
4. Average premium and default legs across simulations separately. Calculate the fair spread s.

1. Introduction to CQF Final Project
2. Pricing a credit product: kth-to-default Basket CDS
3. Aspects of LMM Calibration
4. Interest Rate Swap: Exposure Profile, OIS Discounting

LMM Calibration: Caplet Volatility Stripping

Caplet

A caplet is an interest rate option that pays **a cashflow** based on the value of LIBOR at a re-set time T_i.

$$\mathrm{DF}_{\mathrm{OIS}}(0, T_{i+1}) \times \max(L(T_i, T_{i+1}) - K, 0) \times \tau \times N \tag{5}$$

- $L(T_i, T_{i+1})$ is the forward LIBOR. Assume $L(T_i, T_{i+1}) = f_i$
- τ is year fraction that converts an annualised rate
- N is the notional that can be scaled as $N = 1$

The cashflow is paid for the period $\tau = [T_i, T_{i+1}]$ in arrears.

Payoff and parity

Buying a caplet gives protection from an increase in LIBOR rate:

$$L - (L - K)^{+} = \min(L, K)$$

Alternatively for a floorlet:

$$\max(L, K)$$

Put-call parity for caplet and floorlet becomes:

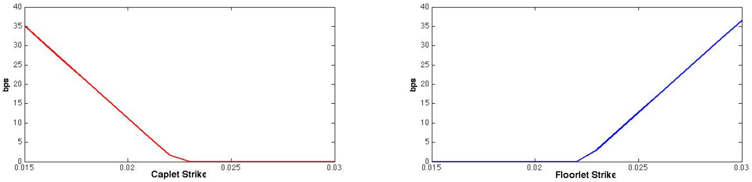
Buying a caplet and selling a floorlet with the same strike gives a payoff equal to a FRA contract. Consider

$$(L - K)^{+} - (K - L)^{+} = (L - K) \qquad \text{always}$$

where FRA fixed rate is equal to the strike, so $(f - K)$.

Pricing Skew

Caplet/floorlet cashflow **in basis points** for a range of strikes:



ITM options just have the larger cashflows – the relationship almost linear. But this is not in implied volatility terms.

Volatility skew present in cap/floor markets. Derivatives models, such as LMM, are calibrated for each strike separately.

Caplet cashflow is computed in HJM MC.xlsm, however you need to run Solver manually on fixed inputs (spreadsheet updating off).

To find implied volatility σ^{cap}_{imp}, use the root-finding on Black (1976) formula

$$\begin{aligned}\textbf{Cap cash} &= Z(0, T_i)\,[f_i\,N(d_1) - K\,N(d_2)]\,\frac{\tau_i}{1 + f_i\tau_i}\\ d_{1,2} &= \frac{\ln(f_i/K) \pm 0.5\sigma^2 T}{\sigma\sqrt{T}}\end{aligned}$$

$f_i = F(t, T_i, T_{i+1})$ forward LIBOR at the caplet expiry T_i, paid over $[T_i, T_{i+1}]$.

$Z(T_i, T_{i+1}) = 1/(1 + f_i\tau_i)$ is discounting factor for LMM. For OIS discounting do $f_i^* = f_i - \text{LOIS}$.

To calibrate the LMM = To strip caplet volatility.

The **market-quoted caps** $\sigma^{cap}(T_{i-1}, T_i)$ trade with expiries $1Y, 2Y, 3Y$, etc.

Stripped **3M caplet sequence** $\sigma^{cpl}(T_{i-1}, T_i)$ gives an approximation to time-dependent volatility function $\sigma^{inst}(t)$.

Calibration to swaptions (Rebonato Method) is appropriate if you have those instruments.

5×5 swaption matures in 5 years, after which IRS will be alive for a further 5 years. The implied volatility denoted $V_{5,10}$.

Market Cap Quotes

Tenor T_i	Date	Discount factor $B(0, T_i)$	Cap volatility $\sigma^{cap}(T_0, T_i)$
$t = 0$	21-01-2005	1.0000000	N/A
T_0	25-01-2005	0.9997685	N/A
T_{SN}	26-01-2005	0.9997107	N/A
T_{SW}	01-02-2005	0.9993636	N/A
T_{2W}	08-02-2005	0.9989588	N/A
T_{1M}	25-02-2005	0.9979767	N/A
T_{2M}	25-03-2005	0.9963442	N/A
T_{3M}	25-04-2005	0.9945224	N/A
T_{6M}	25-07-2005	0.9890361	N/A
T_{9M}	25-10-2005	0.9832707	N/A
T_{1Y}	25-01-2006	0.9772395	0.1641
T_{2Y}	25-01-2007	0.9507588	0.2137
T_{3Y}	25-01-2008	0.9217704	0.2235

From: *LIBOR Market Model in Practice* by Gatarek, et al. (2006), Ch 7.

Calibrating LMM on caplets

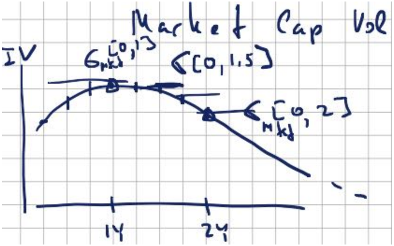
We make sense of Algorithm 7.1 in Gatarek, et al. (2006, page 76).

1. ATM strikes for caplets are equal to the forward-starting swap rates, obtained directly from the forward curve *today*.

$$S(t, T_i, T_{i+1}) \quad \text{or} \quad S(T_i; T_0^*, T_{3M}^*) \quad \text{gives} \quad K^{cpl}$$

   Textbook recites the common formula 7.4 (page 72). These strikes are used when converting implied volatilities to cash prices.

2. Extrapolate market volatilities $[0, T]$, assuming flat volatility for the first period $\sigma^{cap}_{Mkt}(t, T_{6M}) = \sigma^{cap}_{Mkt}(t, T_{1Y})$



$$\cancel{\sigma^{cap}(t, T_{3M})}, \; \sigma^{cap}(t, T_{6M}), \; \sigma^{cap}(t, T_{9M}), \; \sigma^{cap}_{Mkt}(t, T_{12M}), \; \sigma^{cap}(t, T_{18M}), \ldots$$

3. Convert into cash prices $\text{cap}_{Mkt}(t, T_{9M})$, $\text{cap}_{Mkt}(t, T_{12M})$, ... using Black formula.

4. Actual caplet stripping

$$\text{cpl}(T_{6M}, T_{9M}) = \text{cap}_{Mkt}(t, T_{9M}) - \text{cpl}(T_{3M}, T_{6M})$$

$$\text{cpl}(T_{9M}, T_{12M}) = \text{cap}_{Mkt}(t, T_{1Y}) - \text{cpl}(T_{6M}, T_{9M}) - \text{cpl}(T_{3M}, T_{6M})$$

The expression relies on the model-free fact that **caplet cashflows add up to the cap cashflow**: $\text{cap}_T = \sum_i^T \text{cpl}_i$

$$\text{cap}_{Mkt}(t, T_{9M}) = \cancel{\text{cpl}(t, T_{3M})} + \text{cpl}(T_{3M}, T_{6M}) + \text{cpl}(T_{6M}, T_{9M})$$

$$\text{cap}_{Mkt}(t, T_{12M}) = 0 + \text{cpl}(T_{3M}, T_{6M}) + \text{cpl}(T_{6M}, T_{9M}) + \text{cpl}(T_{9M}, T_{12M})$$

5. Use the root-finding on Black formula to convert caplet cashflows into volatilities $\sigma(0.25, 0.5)$, $\sigma(0.5, 0.75)$, $\sigma(0.75, 1)$

$$\text{cpl}(T_{i-1}, T_i) \quad \Leftrightarrow \quad \sigma^{cap}(T_{i-1}, T_i).$$

Even if the annualised implied volatility σ is taken the same for the initial period, say 16%

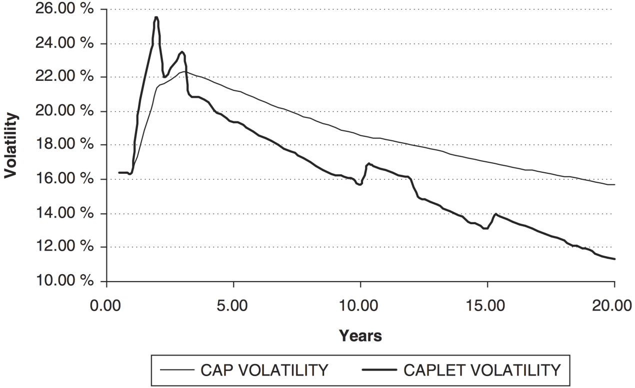
$$\text{cpl}(T_{3M}, T_{6M}) \quad < \quad \text{cap}(t, T_{6M})$$

$$\text{cpl}(T_{3M}, T_{6M}) \quad \ll \quad \text{cap}(t, T_{1Y})$$

$$\textbf{Cpl cash} \quad = \quad Z(0, T_i)\ [f_i\, N(d_1) - K\, N(d_2)]\ \frac{\tau_i}{1 + f_i \tau_i}$$

Caplets

Stripped caplet volatility on a given day (Gatarek, et al., 2006)



Consistent with actual vs implied (*Understanding Volatility Lecture*).

Instantaneous Volatility

After volatility stripping, we moved **from 1.0 year to 0.25** increment. But, the implied volatility remains an average over the actual, instantaneous volatility (via integration).

$$\sigma^{cap}(t, T_{i-1}, T_i) = \sqrt{\frac{1}{T_{i-1} - t} \int_t^{T_{i-1}} \sigma^{inst}(\tau)^2 \, d\tau}$$

- Fitting to a, b, c, d function.
- Piecewise constant instantaneous volatility assignment – Section 7.4 in Gatarek et al. (2006).

$$\boldsymbol{\Sigma}_{cpl} \Rightarrow \boldsymbol{\Sigma}_{inst}$$

Volatility Fitting

We would like the term structure of volatility to be time-homogeneous,

$$\sigma^{inst}(t) = \phi_i \left[(a + b(T_{i-1} - t)) \times e^{-c(T_{i-1}-t)} + d \right] \times \mathbf{1}_{\{t<T_{i-1}\}}$$

a, b, c, d are **the same** for all tenors! Obtained by setting up the optimisation task: assume $\phi_i = 1$, row with $\sigma_{Stripped}$, guess a, b, c, d and compute initial σ_{Fitted} in another row

$$\text{argmin} \sum \left(\sigma_{Stripped} - \sigma_{Fitted} \right)^2$$

for each cell compute the squared difference and run Solver that varies a, b, c, d to minimise the sum of squared differences.

Optimisation can be enhanced by modifier $0.9 < \phi_i < 1.1$ to create a near-perfect fit.

Parametrised Instantaneous Volatility

$$\int_t^{T_{i-1}} \sigma^{inst}(\tau)^2 \, d\tau \quad = \quad \frac{1}{4c^3}\left(4ac^2 d[e^{2c(t-T_{i-1})}] + ...\right)$$

The so called FRA/FRA covariance matrix for instantaneous volatility, our $\mathbf{\Sigma}_{inst}$, also has a, b, c, d-parametrised, closed-form solution – these 'integrated covariances' can be used in LMM SDE **if pre-multiplied** by DF and year fraction.

$$\int \rho_{ij}\,\sigma_i(\tau)\,\sigma_j(\tau)\,d\tau = e^{-\beta|t_i-t_j|}\phi_i\phi_j\,\frac{1}{4c^3}\left(4ac^2 d[e^{c(t-T_i)} + e^{2c(t-T_j)}] + ...\right)$$

The complete parametric solutions to be found in CQF Lecture on the LMM and Peter Jaekel's textbook.

Parametric Correlation

The simplest parametric fit for correlations with $\beta \approx 0.1$ has merits for longer tenors

$$\rho_{ij} = e^{-\beta(t_i - t_j)}$$

The two-factor parametric form of Schoenmakers and Coffey (2003):

$$\rho_{ij} = \exp\left(-\frac{|i-j|}{m-1}\left[-\ln\beta_1 + \beta_2 \ldots\right]\right)$$

works for situations that are different from the stylised empirical observations.

Empirical Correlation

First, changes (in forward rates) at the neighbouring tenors tend to correlate stronger

$$\mathrm{Corr}[\Delta f_{i-1}, \Delta f_i] > \mathrm{Corr}[\Delta f_{i-3}, \Delta f_i]$$

Second, correlation is higher towards the long end of the curve.

$$\mathrm{Corr}[\Delta f_{i-1}, \Delta f_i] < \mathrm{Corr}[\Delta f_{j-1}, \Delta f_j] \qquad \text{for } j \gg i$$

At the short end the rates tend to behave more independently from one another. This is due to being most sensitive to the principal component/primary risk factor of rising the level in the risk-free rate and the entire curve. Further, for 3M, 6M and 1Y tenors there is own dynamics because of how specific market instruments are traded.

LIBOR Market Model SDE

LMM Notation

The LIBOR Market Model was designed to operate with forward rates and denotes them as f_i, where

$$f_i = F(t; t_i, t_{i+1})$$

The forward rate re-sets at time t_i and matures at time t_{i+1}.

Discount factor is represented in the LIBOR model as

$$Z(t; T_{i+1}) \equiv \frac{1}{1 + \tau_i f_i}$$

This is discount factor over *the forward period* $\tau_i = t_{i+1} - t_i$! We need 'one step back' in LMM SDE.

Rolling-forward risk-neutral world

LMM SDE is defined under the measure $\mathbb{Q}^{m(t)}$, known as **the rolling forward risk-neutral world**.

- We just keep discounting the drift.

If you would like to see how LMM SDE is derived, please review *CQF Lifelong Lecture on LMM* by Tim Mills.

$Z(t; T_i)$ is a function of forward rates $F_j(0 \geq j \geq i-1)$. It is **not** a function of forward rates F_i

$$
\begin{aligned}
Z(t; T_i) &= Z(t; T_j) \times Z(T_j, T_i) = \frac{Z(t; T_j)}{1+\tau_j F_j} \quad \text{where } j = i-1 \\
&\quad \text{Taylor series means } dZ_i \propto dt,\ dF_j,\ dF_j dF_k,\ dF_j dt, \ldots \\
dF_j dF_k &= F_j F_k \sigma_j \sigma_k \rho_{jk} dt
\end{aligned}
$$

LMM SDE

Using a discretely rebalanced money market account as Numeraire, the forward rate f_i follows the log-normal process

$$\frac{df_i}{f_i} = \sum_{j=m(t)}^{i} \frac{\tau_j f_j}{1+\tau_j f_j}\,\sigma_i\sigma_j\rho_{ij}dt + \sigma_i\,dW_i^{\mathbb{Q}^{m(t)}} \tag{6}$$

$m(t)$ is an index for the next re-set time. This means that $m(t)$ is the smallest integer such that $t^* \leq t_{m(t)}$.

LMM SDE discretised (single-factor)

The SDE (6) is for the **log-normal** dynamics of f_i given by the instantaneous FRAs. It is solved into a discretised version as follows:

$$
\begin{aligned}
f_i(t_{k+1}) \quad = \quad & f_i(t_k) \exp\left[\left(\sigma_i(t_{i-k-1}) \sum_{j=k+1}^{i} \frac{\tau_j f_j(t_k)\, \sigma_j(t_{j-k-1})\rho_{ij}}{1+\tau_j f_j(t_k)}\right.\right. \\
& \left.\left. -\frac{1}{2}\sigma_i^2(t_{i-k-1})\right)\tau_k + \sigma_i(t_{i-k-1})\phi_i\sqrt{\tau_k}\right] \qquad (7)
\end{aligned}
$$

where $f_j(t_k) = f_j$ and $\sigma_j(t_k) = \sigma_j(t)$ for $t_k < t < t_{k+1}$.

Notation t_{j-k-1} means we refer to the previous time step $k-1$.

This discretisation is optimal HOWEVER we might fall back to the original SDE – if we prefer to use integrated covariances and/or more affine computation.

By column arrangement reveals the logic all rates being under the same measure.

$L_1(0)$			
$L_2(0)$	$L_2(3\text{M})$		
$L_3(0)$	$L_3(3\text{M})$	$L_3(6\text{M})$	
$L_4(0)$	$L_4(3M)$	$L_4(6\text{M})$	$L_4(9\text{M})$

LMM model output with credit to Numerical Methods book and *CCP Elective* by Dr Alonso Pena.

The result, simulated curve, will be on the diagonal.

Consider rate $L_4(9M) = f_i(t_{k+1})$, is the last simulated tenor, it will have **no drift** because it is 'under the terminal measure'.

- Rate $L_4(6M) = f_i(t_{k+1})$ has only one integrated covariance.
- Rate $L_4(3M) = f_i(t_{k+1})$ will have the largest summation in the drift, that encapsulates $[3M, 6M]$ and $[6M, 9M]$ caplets.

$$L_2(3M) = L_2(0) \times \exp\left[\sigma_2(t_0) \times \sigma_2(t_0) \times 1 \ldots\right]$$

$$L_3(3M) = L_3(0) \times \exp\left[\sigma_3(t_0) \times (\sigma_2(t_0) \times \rho_{3,2} + \sigma_3(t_0) \times 1) \ldots\right]$$

because $j = k + 1 = 2$, assume $\sigma_1(t_0)$ does not exist as corresponds to $L_1(0)$ which has no caplet.

Column $L_i(0)$ represents fixed and known LIBOR spot today (here, in 0.25 increment).

SDE Simulation

- LMM evolves **log-normal** dynamics of f_i

$$f(t + dt) = f(t) \exp(df)$$

The curve evolved in discrete tenor chunks, arranged in **column**.

- HJM evolves only the Normal increment df_i (Gaussian model)

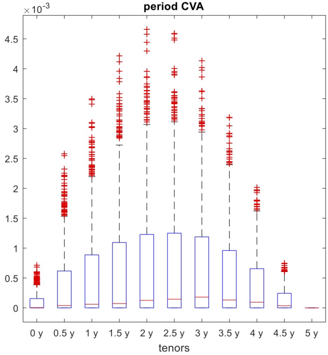
$$\bar{f}(t + dt, \tau) = \bar{f}(t, \tau) + d\bar{f}$$

$d\bar{f}$ are Normally distributed and the curve **in row** assumes evolution in continuous time dt (not $dt = 0.25$).

For CVA Calculation

Interest Rate Swap MtM exposure over tenor time

(a) MtM simulations of exposure, and (b) Expected Exposure **EE** as the worst case, (c) Potential Future Exposure **PFE**.



An example of exposure analytics that encompasses, both EE and multiple PFE. From: Fernando R. Liorente, CQF Delegate

EE and PFE consider **positive exposure only**, when $L_{6M} > K$ and the payer swap cashflow is positive.

- MtM values are cashflows (each reset point, discounted) form the full curve $\mapsto [0, 5Y]$,

  then the curve $\mapsto [0, 4.5]$ simulated at $t = 0.5$,

  then $\mapsto [0, 4]$ simulated at $t = 1$,
  etc.

HJM Model MC IRS.xlsm provides one kind of implementation – pricing is simplified and based on one-off curve – your implementation might vary, for example in how you approach discounting, and use full simulated curves.

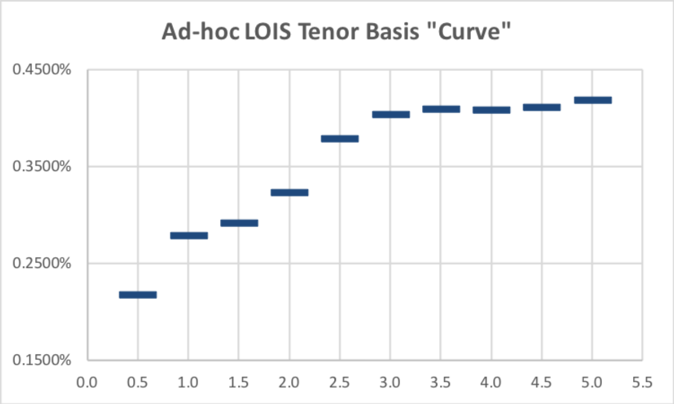
LOIS Spread: constant or curve

LOIS scan be constant spread or tenor-based 'curve'.

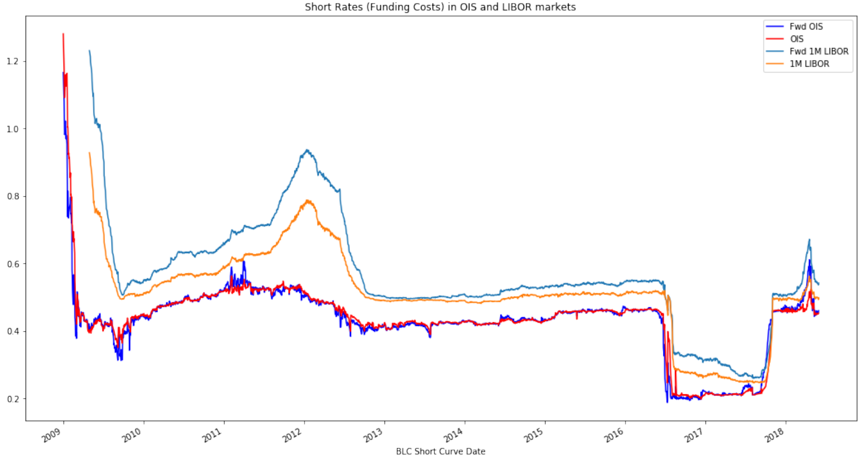
		0.5	1.0	1.5	2.0
LIBOR-OIS (spot)	61.10	**0.2285%**	0.3381%	0.4552%	0.5833%
LIBOR - Fwd OIS	35.78	**0.2179%**	0.2798%	0.2935%	0.3259%
Fwd Inst -Fwd OIS	35.36	**0.2168%**	0.2783%	0.2912%	0.3227%
	Average, bps. Evaluate the range as well				
LIBOR-Fwd OIS	spread	35.78	bps		

$$L_{i,6M} - \text{6M to OIS spread} \qquad \forall i$$

Fwd LIBOR		0.6617%	0.9422%	1.2346%	1.5090%
However, no new OIS curve available (or the new OIS data is stale)!					
Implied Fwd OIS		0.3039%	0.5843%	0.8768%	1.1512%



Source: *Yield Cuve v3.xls* by Richard Diamond



Fwd 1M LIBOR can be seen as risk-adjusted 1M LIBOR (spot)

Source: BOE/Bloomberg data processed by Richard Diamond

The Heath, Jarrow and Morton Model

Forward Curve. Factorisation with Principal Component Analysis

In this lecture. . .

- The short rate process, bond price and forward rates
- Evolving the forward curve with HJM model – in fact, a system of SDEs for each term forward rate
- Derivatives pricing by Monte-Carlo (caplet, floorlet, spread option)
- HJM is your first multi-factor model. Calibration done with PCA
- Yield Curve Data Analysis: revealing the internal structure

By the end of this lecture you will:

- understand forward rates and their bootstrapping
- get introduced to quant modelling of a yield curve
- be able to analyse the yield curve changes data to calibrate fwd rate volatility
- understand the HJM framework and its calibration issues
- be able to price simple interest rate derivatives by Monte-Carlo

Introduction

The Heath, Jarrow & Morton approach was a major breakthrough that improved risk management in fixed income.

It models the yield curve as a whole.

This is a good starting point to learn about forward rates and their volatility/SDE for curve simulation.

Then you can move onto term forward rates (IBORs), and LIBOR Market Model (LMM).

One-factor models, Vasicek, CIR, Ho & Lee, Hull & White, evolve the short rate $dr(t)$.

- $r(t) = f(t, t)$ represents **one point on the curve**.

  Curve is evolved with $r + dr$, a series of steps. Can price a bond numerically by integrating over the evolution.

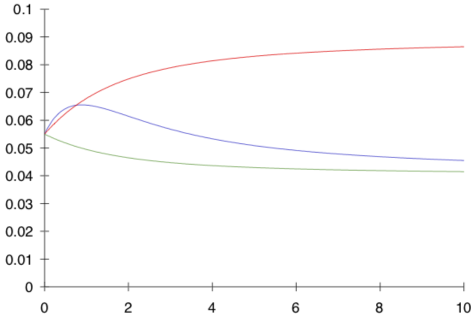
- Calibrated to short-term $Z_M(t^*, T)$ market bonds ONLY.

  There are bond futures, IRS, FRAs written directly on expected LIBOR at tenors t^* $(f_{t^*,6M} - k)$.

  Offer analytical ‘bond pricing’ $Z(t; T) = e^{A(T) - r\,B(T)}$.

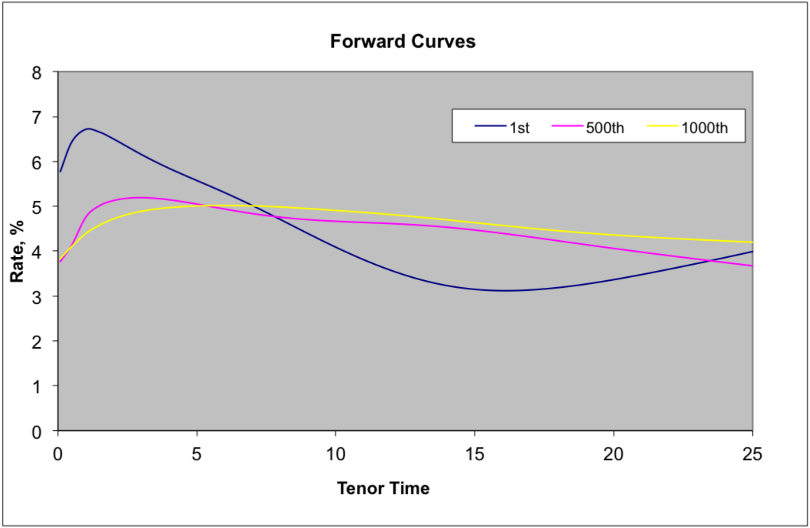
One-factor simulated curves

Types of yield curve given by the Vasicek model $dr = (\eta - \gamma r)\, dt + \beta^{1/2}\, dX$.



From: CQF Lecture on Stochastic Interest Rates, Slide 43.

Market Curves (forward rates, pound sterling)



- 'One factor' translates to one kind of movement: parallel up/down.

$$Corr[\Delta r_t, \Delta f_j] \approx 1$$

- Empirical fact is that **short-term** funding rates move rather independently from long-term rates.

$$Corr[\Delta r_t, \Delta f_j] \approx 0$$

Both cannot be true.

<u>Outcome:</u> One-factor models are affine, easy to simulate but produce simplistic moves for a whole curve. Good as a quick fix for risk calculations: CVA, IRS pricing examples.

It makes sense to model a sequence of forward rates $f(0; t_i, t_{i+1})$,

rather than overlapping discount factors $Z(0, T_1), Z(0, T_2), Z(0, T_3)\ldots$

$$
\begin{aligned}
Z(0, T_1) \times Z(0; T_1, T_2) &= Z(0, T_2) \\
Z(0; T_1, T_2) &= \exp\left[-f(0; t_1, t_2)(t_2 - t_1)\right] \\
f_2 &= -\frac{\ln Z_2 - \ln Z_1}{t_2 - t_1}
\end{aligned}
$$

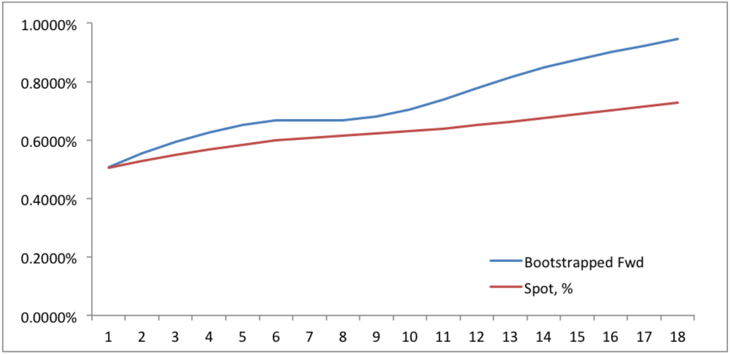
Computing forward rates

A scheme for each discrete tenor j follows from the instantaneous forward rate maths (1)

$$f_0 = -\frac{\ln Z_1 - \ln 1}{T_1 - 0} \qquad f_1 = -\frac{\ln(Z_2/Z_1)}{T_2 - T_1}$$

Tenor, T	**0.08**	**0.17**	**0.25**	**0.33**
Spot	0.5052%	0.5295%	0.5500%	0.5682%
Z(0, T)	0.9996	0.9991	0.9986	0.9981
Forward	0.5063%	0.5552%	0.5927%	0.6244%

Forward vs. Spot Curve



Data: Bank Liability Curve as of 30 January 2015 (Bank of England)

An **instantaneous forward rate** $F(t, T, T + \Delta t)$, where rate expires at time T and applies over an instant $\Delta t \to 0$.

$$\begin{aligned} f(t, T) &= -\lim_{\Delta t \to 0} \frac{lnZ(0; T + \Delta t) - \ln Z(0; T)}{\Delta t} \\ &= -\frac{\partial}{\partial T} \ln Z(0; T) \end{aligned} \tag{1}$$

Now we use instantaneous notation $f(t, T) \equiv F(t, T, T + \Delta)$.

LIBOR, a simple forward rate

Market-quoted and traded LIBOR is a simple forward rate $L(t, T_{i-1}, T_i)$ or $L(T_{i-1}, T_i)$ or $L_i(t)$

$$\text{LIBOR} = m\left(e^{\text{inst fwd}/m} - 1\right) \tag{2}$$

where m is compounding frequency for a year, eg 3M LIBOR compounded $m = 4$ times.

Interest Rate Swaps, Forward Rate Agreements reference to $L(T_{i-1}, T_i)$, a traded quantity.

$$\text{FRA}(T_{i-1}, T_i) - L(T_{i-1}, T_i) = 0 \tag{3}$$

At reset time, FRA rate and LIBOR fixing will coincide.

Modelling bond price

The log-normal SDE must be a familiar model:

$$\frac{dZ}{Z} = \mu(t, T)\, dt + \sigma(t, T)\, dX \tag{4}$$

where bond price evolves with t but the maturity date T is fixed.

Empirical observation: if we construct a constant maturity bond price, the changes in the yield give us **the model invariant** (an *iid* process)

$$\Delta f \quad \propto \quad \ln Z(t_{i+1}; T) - \ln Z(t_i; T)$$

$$\Delta f \quad \sim \quad Normal(\mu, \sigma^2 \tau)$$

The model is Normal for forward rates $f(t, T)$ and Log-Normal for bond prices $Z(t; T)$.

HJM Calibration

Data and Computational Preview

1. Take the data of *changes* in forward rates about constant tenors Δf_j and compute $j \times j$ covariance matrix $\boldsymbol{\Sigma}$.

$$\boldsymbol{\Sigma} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}'$$

2. Conduct PCA on the covariance matrix $\boldsymbol{\Sigma}$. Each component is a linear contribution to change in the yield curve.

3. Fit volatility functions from components (eigenvectors). This gives both, diffusion and drift for HJM SDE.

4. Pricing of interest rate derivatives is done by Monte-Carlo.

Start with the forward rate equation $f(t, T) = -\frac{\partial}{\partial T} \ln Z(t, T)$ and use the bond price 'GBM' SDE to derive the evolution of forward rates:

$$df(t, T) = \frac{\partial}{\partial T} \left[\frac{1}{2}\sigma^2(t, T) - \mu(t, T) \right] dt - \frac{\partial}{\partial T}\sigma(t, T)dX \qquad (5)$$

The SDE carries the drift of the bond price $\mu(t, T)$.

When we come to pricing, such drift terms are replaced by the risk-free interest rate $r(t)$.

Risk to risk-neutral

In the real world, our model for he evolution of the forward curve (5) was derived as

$$df(t,T) = \frac{\partial}{\partial T}\left[\frac{1}{2}\sigma^2(t,T) - \underline{\mu(t,T)}\right]dt - \frac{\partial}{\partial T}\sigma(t,T)dX$$

In the risk-neutral world, under measure $\mathbb{Q}$, the model becomes

$$df(t,T) = \frac{\partial}{\partial T}\left[\frac{1}{2}\sigma^2(t,T) - \underline{r(t)}\right]dt - \frac{\partial}{\partial T}\sigma(t,T)dX^{\mathbb{Q}} \qquad (6)$$

But $r(t)$ is not a function of T, so $\frac{\partial}{\partial T}r(t) = 0$

Change of measure

Replacing $\mu(t,T)$ with $r(t)$ is a change of measure. Girsanov theorem gives

$$X_t^{\mathbb{Q}} = X_t + \int_t^T \theta_s ds \qquad \frac{d\mathbb{Q}}{d\mathbb{P}} = exp\left\{-\frac{1}{2}\theta^2 T - \theta X_T\right\}$$

The result for market price of risk is familiar $\theta = \frac{\mu - r}{\sigma}$.

$$\frac{dZ}{Z} = \mu(t,T)dt + \sigma dX_t$$

$$\frac{dZ}{Z} = r(t)dt + \sigma dX_t^{\mathbb{Q}}$$

In the risk-neutral economy, the expected return on any traded investment (a bond) is simply $r(t)$.

- By setting up a hedged portfolio $\Pi = Z(t, T_1) - \Delta Z(t, T_2)$ we found that to cancel the drift we require

$$\frac{\mu(t, T_1) - r(t)}{\sigma(t, T_1)} = \frac{\mu(t, T_2) - r(t)}{\sigma(t, T_2)}$$

- Only possible if both sides are equal to some parameter, which is independent of maturity dates T_1, T_2

$$\mu(t, T) = r(t) + \lambda(r, t)\sigma(t, T)$$

  where $\lambda(r, t)$ is the market price of risk (MPOR), a unifying global parameter, not a constant!

More in the CQF Extra *The Market Price of Risk: Fear and Greed...*

Drift derivation

We expressed the risk-neutral drift as a function of volatility – of bond prices for now. Chain rule was used for differentiation *wrt* ∂T

$$df(t,T) = \underbrace{\sigma(t,T)\frac{\partial}{\partial T}\sigma(t,T)dt}_{} - \frac{\partial}{\partial T}\sigma(t,T)dX^{\mathbb{Q}} \tag{7}$$

Equivalent to,

$$df(t,T) = -\sigma(t,T)\nu(t,T)dt + \nu(t,T)dX^{\mathbb{Q}}$$

Forward rate volatility

$$\nu(t,T) = -\frac{\partial}{\partial T}\sigma(t,T) \qquad \sigma(t,T) = -\int_t^T \nu(t,s)ds$$

this is a conversion into **volatility of forward rates** $\nu(t,T)$.

$$df(t,T) = \nu(t,T)\underbrace{\left(-\sigma(t,T)\right)}\,dt + \nu(t,T)dX^{\mathbb{Q}}$$

$$df(t,T) = \left[\nu(t,T)\int_t^T \nu(t,s)ds\right]dt + \nu(t,T)dX^{\mathbb{Q}}$$

From now on, we operate in the risk-neutral world, so we drop $\mathbb{Q}$.

Risk-neutral HJM dynamics

The drift is a function of volatility, **a no arbitrage condition**.

$$m(t, T) = \nu(t, T) \int_t^T \nu(t, s) ds \tag{8}$$

HJM SDE expressed deceiptively simply

$$\boldsymbol{f}(t, T) = m(t, T)dt + \nu(t, T)dX \tag{9}$$

Notation in bold, $\boldsymbol{f}(t, T)$ refers to evolution of the whole curve – that is at any point of tenor time.

We cannot model ***barf***(t,τ) at an infinite number of tenors, so we have to discretise into *a system* of SDEs – each is evolved in **a separate column** of *HJM MC.xlsm*.

$$
\begin{aligned}
d\bar{f}(t,\tau_1) &= \bar{m}(t,\tau_1)dt + \sum_{i=1}^{k} \bar{\nu}_i(t,\tau_1)dX_i \\
\vdots \quad & \quad\quad \vdots \quad\quad\quad\quad \vdots \\
d\bar{f}(t,\tau_j) &= \bar{m}(t,\tau_j)dt + \sum_{i=1}^{k} \bar{\nu}_i(t,\tau_j)dX_i
\end{aligned}
$$

Forward rate at each tenor point $\tau_j = T_j - t$ is its own stochastic variable. j is a counter for tenor 0.5Y, 1Y, 1.5Y, etc.

We have introduced something else: instead of one volatility function $\bar{\nu}_i(t,\tau)$, we have several! That summation means we have **a multi-factor** stochastic model.

$$\sum_{i=1}^{k} \bar{\nu}_i(t,\tau_j)\, dX_i$$

Curve stress-testing: diffusion has several independent sources of randomness dX_i. Each represents uncertainty about curve movement, from a different factor.

Each volatility function is a Principal Component

$$\bar{\nu}_i(t,\tau) = \sqrt{\lambda_i}\, \boldsymbol{e}_{\boldsymbol{\tau}}^{(\boldsymbol{i})}$$

Through **calibration by PCA** and **fitting by cubic spline** we have the volatility functions $\bar{\nu}_i(t,\tau)$

$$\bar{m}(t,\tau) = \sum_{i=1}^{k} \bar{\nu}_i(t,\tau) \int_0^{\tau} \bar{\nu}_i(t,s)ds + ...$$

Drift computation requires numerical integration $\int_0^{\tau} \bar{\nu}_i(t,s)ds$ implemented by Trapezium Method inside *m()* function in VBA.

Pricing by Monte-Carlo

Simulated Output of HJM

1. Simulation Simulate an evolution of the whole risk-neutral curve for the necessary length of time, from today t^* to T^*.

- For 'a risk-neutral forward curve' we use GLC data (UK Gilts).
- Realizations of the curve $\boldsymbol{f}(t, T)$ over time steps $dt = 0.01$ are **in rows**.
- The paths of forward rates for discretised tenors τ_j are **in columns**.

2. Discounting factors Obtain ZCB values for all required tenors up to T^*. However, discounting factors can come from the outside (e.g., OIS curve) creating a problem of how to match expectations.

Risk-neutral discount factor

Pricing a Zero Coupon Bond with any stochastic short rate $r(t)$

$$Z(t,T) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)ds\right)\right]$$

Here is a case when we do not need the entire simulated curve.

Looking at the HJM output, we already know that we only an evolution of the first point on the yield curve $f(t,t) = r(t)$.

$$\begin{aligned} Z(t,T) &= \exp\left(-\sum r_t\,\Delta t\right) \qquad \text{under MC} \qquad (10)\\ &= \exp\left(-\text{SUM(COLUMN)} \times 0.01\right) \end{aligned}$$

Monte-Carlo simulation is always discretised over time step dt. Therefore, integration becomes summation.

3. Calculate the value of cashflows of interest and apply discounting. Example: consider $(L - K)^+$ payoff and discount

(Discounting is assumed within the same risk-neutral expectation.)

Return to Step 1 to perform another round of simulation.

- Keep track of the running average of simulated prices, i.e., 1st, 1st + 2nd, 1st + 2nd + 3rd, etc. .
- It must demonstrate convergence/reduction in its variance.

Let's review the pricing.

The analytical solution (PDE approach) is not feasible under the HJM. That leaves us with two choices:

1. **Monte-Carlo method** – estimation of an expectation by simulating the evolution of forward rates.

   Implemented under risk-neutral HJM dynamics and measure $\mathbb{Q}$.

2. The other is to build up **a tree** structure and formalise it into a Finite Difference grid.

EXTRA. HJM and the short rate r(t)

Given by the forward rate for a maturity equal to the current date, i.e.

$$r(t) = f(t, t)$$

If the forward curve today is $f(t^*, T)$ then the short rate for *any* time t in the future is

$$r(t) = f(t^*, t) + \int_{t^*}^{t} df(s, t)$$

In terms of HJM output, $df(s, ...)$ means evolving a rate **in column**.

Check with $f(t^*, t) + f(t, t) - f(t^*, t) = f(t, t)$.

The HJM SDE in terms of spot rate is jammed into

$$dr(t) = \left[\frac{\partial f(t^*, t)}{\partial t} - \left. \frac{\partial \mu(t, s)}{\partial s} \right|_{s=t} \right.$$
$$+ \int_{t^*}^{t} \left(\sigma(s, t) \frac{\partial^2 \sigma(s, t)}{\partial t^2} + \left(\frac{\partial \sigma(s, t)}{\partial t} \right)^2 - \frac{\partial^2 \mu(s, t)}{\partial t^2} \right) ds$$
$$\left. - \underline{\int_{t^*}^{t} \frac{\partial^2 \sigma(s, t)}{\partial t^2} dX(s)} \right] dt - \left. \frac{\partial \sigma(t, s)}{\partial s} \right|_{s=t} dX$$

This SDE is odd because **the drift depends on the history of σ** from the date t^* to the future date t and the stochastic increments dX.

Building a tree

If we attempt to use an evolution path for a forward rate $f(t, t) = r(t)$ to build a tree

- Then we'll find ourselves with an unfortunate result: an up move followed by a down move will **not** end up in the same state.
- Our tree structure becomes 'bushy'. The number of branches grows *exponentially* with the addition of new time steps.

This is a feature of a non-Markov model. Equivalence of paths is what makes the pricing by Binomial Method so efficient.

Non-Markovian nature of HJM

The highly path-dependent drift of dr(t) makes the movement of the short rate **non-Markov**.

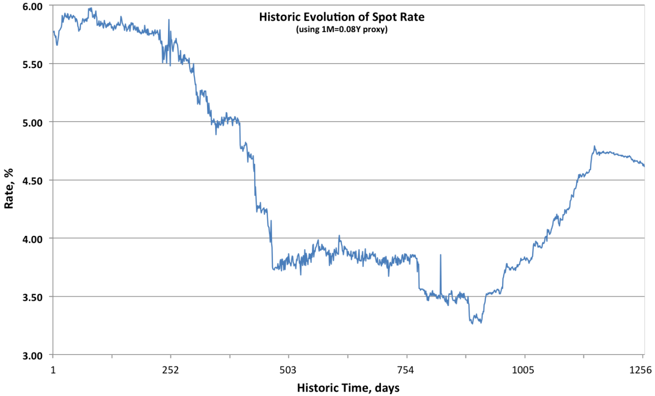
$$r + dr$$

is not recombinable.

In a **Markov chain** only the present state of a variable determines the possible future (albeit random) state.

Markov process is a stochastic process without a **memory**. But let's look at the empirical $r(t)$.

Memory in r(t)



The **long memory** for stationary time series means that decay in autocorrelation is slower than exponential $\text{Corr}[r_t, r_s] = \beta^{t-s}$.

Working in tenor time τ

Modelling in tenor time τ (change of variable) is called **Musiela Parametrisation** of the HJM SDE, which we have derived as $df(t, T)$

- as t changes however, the time distance $\tau = T - t$ changes too. Would be modelling a different tenor rate!!

$$T_{6M-0.00}, T_{6M-0.01}, T_{6M-0.02}, \dots, T_{6M-0.42}, \dots$$

The volatility function keeps its form:

$$\nu(t, T) = -\frac{\partial}{\partial T}\sigma(t, T) \quad \Rightarrow \quad -\frac{\partial}{\partial \tau}\bar{\sigma}(t, \tau)\frac{\cancel{\partial \tau}}{\cancel{\partial T}} = \bar{\nu}(t, \tau)$$

$$\frac{\partial}{\partial \tau}\frac{\partial \tau}{\partial T} \equiv \frac{\partial}{\partial T}\frac{\partial \tau}{\partial \tau}$$

$$\bar{f}\,(t, T-t) \quad \Leftrightarrow \quad \bar{f}(t,\tau)$$

$$T(t) = t + \tau \qquad \text{and} \qquad t(T) = T - \tau$$

We want to evolve an SDE in dt (lhs), while 'keeping τ constant'

$$\frac{d\bar{f}(t,\tau)}{dt} \equiv \frac{\partial}{\cancel{\partial T}}\frac{\cancel{\partial T}}{\partial t} + \frac{\partial}{\cancel{\partial t}}\frac{\cancel{\partial t}}{\partial T}$$

$$\frac{d\bar{f}(\tau)}{dt} \equiv \frac{\partial}{\partial t}f(t, t+\tau) + \frac{\partial}{\partial T}f(T-\tau, T) \qquad \text{or} \qquad \frac{d\bar{f}}{dt} \equiv \frac{df}{dt} + \frac{\partial f}{\partial T}$$

$$d\bar{f} = df + \frac{\partial f}{\partial T}dt \quad \text{or} \quad \boxed{d\bar{f} = df + \frac{\partial \bar{f}}{\partial \tau}dt}$$

$$\begin{aligned}
d\bar{f}(t,\tau) &= \underline{df(t,T)} + \frac{\partial f(t,T)}{\partial T}dt \\
&= \underline{\left(\nu(t,T)\int_t^T \nu(t,s)ds\right)dt + \nu(t,T)dX} + \frac{\partial f(t,T)}{\partial T}dt \\
&= \left(\bar{\nu}(t,\tau)\int_0^\tau \bar{\nu}(t,s)ds + \frac{\partial \bar{f}(t,\tau)}{\partial \tau}\right)dt + \bar{\nu}(t,\tau)dX
\end{aligned}$$

The drift gains an extra term $\frac{\partial \bar{f}}{\partial \tau}$, which is a slope of the yield curve.

When we come to simulate in *HJM MC.xlsm* the complete multi-factor SDE (in drift and in diffusion) is simulated column-by-column.

$$d\bar{f}(t,\tau) = \left(\sum_{i=1}^{k} \bar{\nu}_i(t,\tau) \int_0^{\tau} \bar{\nu}_i(t,s)ds \right) dt + \sum_{i=1}^{k} \bar{\nu}_i(t,\tau)dX_i + \frac{\partial \bar{f}}{\partial \tau}dt \quad (11)$$

$$d\bar{f}(t,\tau) = \bar{m}(\tau)dt + \sum_{i=1}^{k=3} \bar{\nu}_i(t,\tau)dX_i + \frac{\partial \bar{f}}{\partial \tau}dt \quad (12)$$

We simulate,

$$\bar{f}_{t+dt} = \bar{f}_t + d\bar{f}$$

In cells, computation $d\bar{f} = \bar{m}(\tau)dt + SUM(Vol_i * \phi_i)\sqrt{dt} + \frac{d\bar{f}}{d\tau}dt$

```
B$2*dt +

(B$3*$BC13+B$4*$BD13+B$5*$BE13)*SQRT(dt)+

(C12-B12)/(C$11-B$11)*dt
```

$k = 3$ independent factors feature in SDE for each term rate (each column).

Volatility functions $\sqrt{\lambda_i}\,\boldsymbol{e}_{\boldsymbol{\tau}}^{(\boldsymbol{i})}$ were fitted to cubic spline by $LINEST()$

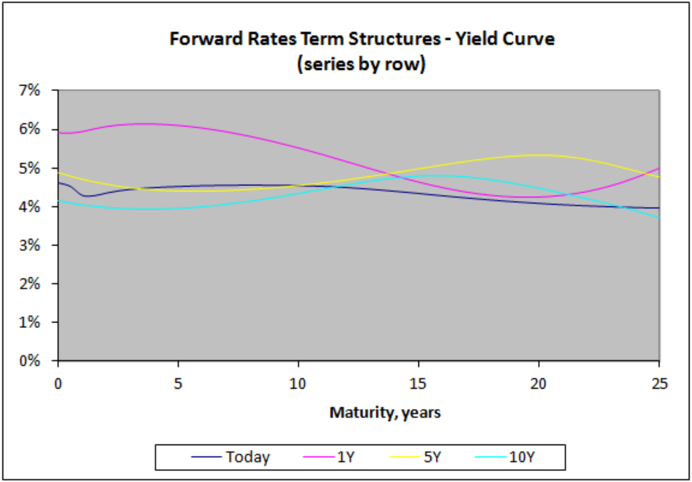
```
Vol_i=b0+b1*tau+b2*tau^2+b3*tau^3
```

Numerical Methods (reference)

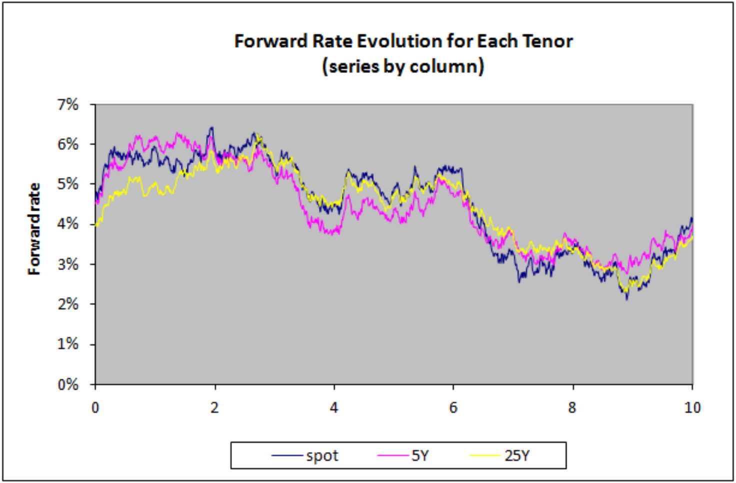
- Drift $\bar{m}(t,\tau)$ computed using *numerical integration* over the fitted volatility functions. **Trapezium rule** is used.
- Eigenvectors and eigenvalues obtained by rotation $\boldsymbol{\Sigma}' = \boldsymbol{P_{p,q}}^{\boldsymbol{T}} \boldsymbol{\Sigma} \boldsymbol{P_{p,q}}$ to eliminate the largest off-diagonal element p, q. Rotation repeated. [See Solutions]
- To initialise, the first row is a static curve from data.
- The forward derivative $\frac{\partial \bar{f}}{\partial \tau}$ is calculated using the row above.

We simulate a realisation of the entire curve at each time step. The simulation is Gaussian HJM and negative rates are possible.

Simulated Forward Curves



Simulated Instantaneous Forward Rates



ZCB Pricing

There are **two approaches** for pricing a bond under HJM framework.

- Integrating over a current forward curve $\bar{f}(t^*, \tau_j)$ – in a row

$$Z(t, T) = \exp\left(-\int_0^{T=\tau} \bar{f}(t^*, \tau)d\tau\right)$$

  This requires a no-arbitrage interpolation of the curve, followed by numerical integration. Things can get very technical!

- Using a simulated path of $r(t) = f(t, t)$ - -from the first column

$$Z(t, T) = \mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)ds\right)\right]$$

Risk-neutral discount factor

Using $r(t)$, the calculation becomes a summation over the first column of simulated data (HJM Model - MC Excel)

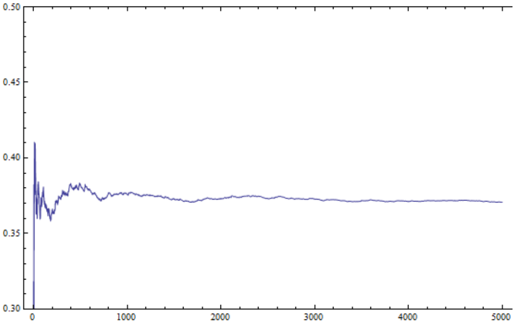
$$
\begin{aligned}
Z(t,T) &= \exp\left(-\sum r_t\,\Delta t\right) \qquad \text{under MC} \qquad (13)\\
&= \exp\left(-\mathrm{SUM(COLUMN)}\times 0.01\right)
\end{aligned}
$$

Solution: to price a half-year bond starting today Z(0; 0.5), we will carry out summation over 50 rows of the first column, $\Delta t = 0.01$.

Convenient!

To satisfy the risk-neutral expectation $\mathbb{E}^{\mathbb{Q}}$ we have to conduct the Monte-Carlo.

ZCB price convergence, $T > 10Y$



Monte-Carlo pricing means that we produce **a running average** of simulated prices, e.g., 1st, 1st + 2nd, 1st + 2nd + 3rd. The running average must demonstrate convergence/reduction in its variance.

Please take away the following important ideas:

- with the HJM model, we evolve the entire forward curve
- calibration means linear factorisation of forward rate volatility
- Principal Component Analysis reveals the internal data structure:

  the key factors of curve movement are level, steepness/flatness and curvature
- pricing under the HJM is done using the Monte-Carlo

Slides next offer **PCA Tutorial**

for **Data Analytics on Yield Curve**.

We start with a case for multi-factor modelling of the yield curve, and review the PCA side of implementation.

Case for a multi-factor model

A single-factor model for the short rate $r(t)$ can't hope to capture the richness of yield curve movements.

Consider *a spread option*. Its payoff is the difference between rates at two different tenors, e.g., $(L_{6M} - L_{3M})$.

- **If movements of two rates are not correlated (going up/down in sync), there is an extra source of risk**, ie, another factor.

  Instrument is sensitive to more than one factor.

  Cannot hedge with a single bond.

Bucket risks vs. Curve movement

Consider two ‘natural’ bucket risks, some short rate $f_{0.08Y}$ and a long-term rate f_{7Y},

the common risk methodology will study the CVA or derivative price *wrt* change at a single bucket.

However, if rates at $0.08, 7$ tenors move in the opposite direction, they represent another kind of curve movement, one systematic factor:

- steepening or flattening of the curve.

$$Corr[\Delta f_{3Y}, \Delta f_{5Y}], Corr[\Delta f_{3Y}, \Delta f_{10Y}], ... \gg 0$$

Interest rate *changes* are well-correlated across distant tenors (except *wrt* the short end).

But the covariance of changes $\boldsymbol{\Sigma(\Delta f_j, \Delta f_{j+h})}$ can be explained with a few independent factors.

There follows a possibility to represent the change about tenor τ_j as a linear decomposition of 'orthogonal' (independent) changes:

$$\Delta f_j = \mathrm{PC}_1 + \mathrm{PC}_2 + ... + \mathrm{PC}_k$$

The linear components come from the **Principal Component Analysis**.

- Systematic factors that describe movement of a curve as a whole.
- Factor attribution is well-established for yield curve analysis.

Note that for the HJM SDE (11), each **volatility function** is equal to the scaled principal component *i*.

$$\bar{\nu}(\tau) = \text{Std Dev} \times \text{Eigenvector}_{\tau} \qquad \text{or} \qquad \sqrt{\lambda_i}\, \boldsymbol{e}_{\boldsymbol{\tau}}^{(\boldsymbol{i})}$$

volatility structure matches the data and calibration is fully numerical.

Data Preparation

If we have time series of each rate going back a few years, we can calculate covariances between **changes** in the rates.

- Inst. forward rates from BOE Yield Curve Statistics (use BLC)
- $\tau = 0.5$ increment for $0.08Y$ ... $25Y$ gives 50 columns.
- Jan 2002 – Jan 2007 regime. Consider regimes since then.

Government Liability Curve (GLC) bootstrapped from repo agreements, spot bonds (Gilts) and bond futures.

Bank Liability Curve built from short sterling futures, and FRAs. It is more suitable for pricing IR derivatives.

Covariance Matrix – Changes Percentage Rates

To estimate the covariance matrix $\boldsymbol{\Sigma}$,

1. Compute **daily differences** in fwd rate at each tenor, columnwise. Subtract the mean, if not a small quantity.

   ```
   B3-B2 - AVERAGE(B), B4-B3 - AVERAGE(B), ...
   5.7680 - 5.7733, 5.7757 - 5.7680
   ```

2. $\boldsymbol{\Sigma} = \frac{1}{N}\boldsymbol{X}\boldsymbol{X}'$ where $\boldsymbol{X}$ relates to the dataset of Differences, see the tab in *HJM PCA.xls*.

3. We annualise covariance, and correct for the fact that we used percentages, eg, 5.768 not 0.05768 when computing differences,

$$\times \frac{252}{100 \times 100}.$$

Matrix Decomposition

$\mathbf{\Sigma}$ is the covariance matrix of fwd rate **changes**. Such symmetric matrix can be decomposed according to *the spectral theorem*:

$$\mathbf{\Sigma} = \mathbf{V\Lambda V'}$$

- $\mathbf{\Lambda}$ is a diagonal matrix with eigenvalues $\lambda_1 > \cdots > \lambda_n > 0$ positive and usually ranked in software output (Matlab, R).

$$\mathbf{\Lambda} = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}$$

- $\mathbf{V}$ is a vectorised matrix of eigenvectors $vec(\mathbf{e}^{(1)}\ \mathbf{e}^{(2)}\ \ldots\ \mathbf{e}^{(n)})$.

$$\Delta f(\tau_j) = \sqrt{\lambda_1}\boldsymbol{e}^{(1)}_{\tau_j} + \sqrt{\lambda_2}\boldsymbol{e}^{(2)}_{\tau_j} + \sqrt{\lambda_3}\boldsymbol{e}^{(3)}_{\tau_j} + \ldots \quad \text{in rows}$$

Using PCA output, we express a system of HJM SDEs in matrix form,

$$d\boldsymbol{f}(t,T) = \boldsymbol{M}(t,T)dt + \boldsymbol{V\Lambda}^{\frac{1}{2}}\, d\boldsymbol{X} \tag{14}$$

where $d\boldsymbol{X}$ is a multi-dimensional Brownian Motion representing k independent factors.

Independence is achieved by decomposition of covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{V\Lambda}^{\frac{1}{2}} \left(\mathbf{V\Lambda}^{\frac{1}{2}}\right)' = \boldsymbol{AA}' \quad \text{Cholesky decomposition}$$

The covariance matrix is estimated from **changes in forward rates**

$$\boldsymbol{\Sigma} = \mathbb{C}ov[\Delta f(\tau_j),\, \Delta f(\tau_{j+h})].$$

Eigenvectors

Volatility functions of τ

- For each *column* eigenvector $\mathbf{e}^{(\mathbf{i})}$, the first entry is the movement of one-month rate ($\tau = 0.08$), the second entry is of the six-month rate ($\tau = 0.5$) and so on.

$$\overline{\nu}_i(t^*, \tau) = \sqrt{\lambda_i}\, \boldsymbol{e}_{\boldsymbol{\tau}}^{(\boldsymbol{i})} \tag{15}$$

To obtain a volatility function, it is naturally convenient to fit a column eigenvector to tenor τ. Eigenvector $\boldsymbol{e}_\tau^{(i)}$ has values at

$$\tau = 0.08Y,\ 0.5Y,\ 1Y,\ \dots,\ 25Y$$

- Instead of picking numbers from the matrix of eigenvectors $\mathbf{V}$, **we use the fitted volatility functions**.

Polynomial Fitting

- The fitting is done by a single **cubic spline** *wrt* tenor τ

$$\overline{\nu}(t,\tau) = \beta_0 + \beta_1\tau + \beta_2\tau^2 + \beta_3\tau^3 \quad \forall\, \tau_j$$

A spline is a piecewise-defined smooth polynomial function.

In general, we can fit exactly using a piecewise polynomial – here, an improvement can be made by enquiring into fitting and methods behind functions like *polyfit()* in Matlab and *nls()* in R.

Fitting recipes are domain-specific for yield curve, implied vol. under LMM/LVM/SABR. Principal components are polynomials, eg, the best PC4 fit requires τ^4.

Despite using *LINEST()* to calculate β, here we are **not** conducting any regression analysis.

In our PCA application to Pound Sterling curve, three factors explain 93.33% of movement (variation) in the yield curve.

Tenor	λ	Cum. R^2
1Y	0.002027	71.31%
25Y	0.000463	87.58%
6Y	0.000164	93.33%

But how did we choose these $k = 3$ eigenvectors to be our volatility functions?

$$\boldsymbol{e}^{(1Y)}, \boldsymbol{e}^{(6Y)}, \boldsymbol{e}^{(25Y)}$$

By the largest corresponding eigenvalue.

Factor Significance

Eigenvalue λ_i **is variance** of the movements of a curve in each eigendirection. For example, the first factor explains

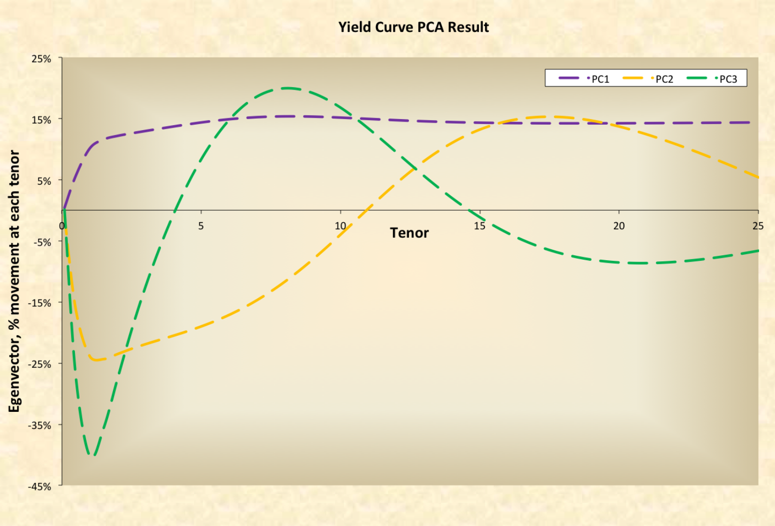
$$\frac{\lambda_1}{\lambda_1 + \lambda_2 + \cdots + \lambda_N}$$

The cumulative goodness of fit statistic for the k-factor model is

$$\text{Cum. } R^2 = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i}$$

By choosing the largest-impact factors we reduce an N-dimensional model to the three-factor model. Each factor represents systemic movement by the curve.

PCA Result: Three Largest Factors



Factor Attribution

- **Parallel shift in overall level of rates** is the largest principal component of forward curve movement, common to all tenors.
- **Steepening/flattening of the curve** is the second important component (i.e., change of *skew* across the term structure)

  Inverted curve (backwardation for commodities term structure) would have a different shape for the PC2.
- **Bending about specific maturity points** is the third component to curve movement that mostly affects *curvature* (convexity).

Disclaimer. These are commonly accepted attributions but please read next.

Factors and Causality in Curve Data Analytics with PCA

Changes in the rate at certain tenor 1Y, 6Y, 25Y can be particularly sensitive to a systematic factor **BUT** causality is not proven, and eigenvectors tend to rotate particularly for PC2 vs. PC3 and above.

Short end of the curve has low correlation of changes with the longer tenors. Short end is sensitive to and often represents PC1.

In the current regime of low interest rates (to 2017) and flattened curve, PC2, PC3 components might have no attribution and rotate often. PCA is a limited tool for periods of rapid shifts in interest rates.

END OF TUTORIAL

THE LIBOR MARKET MODEL

*OTC Analytics

Contents

I. The market view of the yield curve

Financial markets are created by the activities of agents dealing in tradeable securities.

The *yield curve* is given by the distillation of many different market-observable prices into one condensed representation. The distillation procedure is called *yield curve stripping*. This usually involves market quotes from several considerably distinct market segments:

- Interbank cash deposit rates
- Government bond repo rates (this is how the bank of England sets interest rates)
- Interest rate futures. These are the exchange-traded equivalent to *Forward Rate Agreements* (FRAs).
- Swap rates (usually quoted by brokerage services)

Some of these market quotes are subject to credit risk, and for others it can practically be ignored. Some of these quotes are subject to convexity adjustments, and others are not.

In the financial market, there is no such thing as traded yield curve, but there are many interest rate related tradeable contracts.

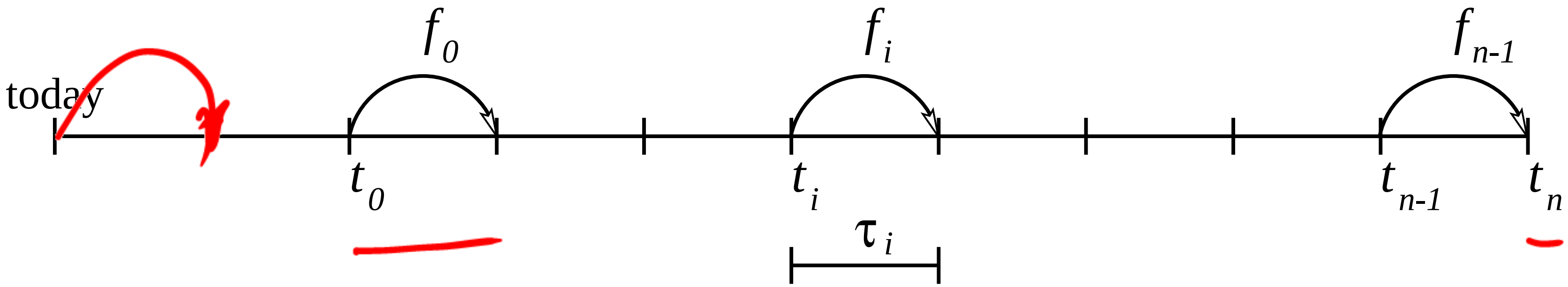
Since most of these contracts are sufficiently distinct, their prices are only very weakly linked by strict no-arbitrage requirements.

The yield curve is a theoretical construct. What matters financially is the trading of securities and contingent claims.

II. Yield curve discretisation

The concept of a *market model* is to describe directly the dynamics of observable market quotes of financially tradeable contracts, rather than to fall back to a hidden process driving the entirety of the fixed income market.

A *Libor market model* is based on the discretisation of the yield curve into *discrete spanning forward rates*.



Each forward rate immediately represents the (modelled) market quote for an associated Forward Rate Agreement (FRA).

A forward rate agreement quote f_i for period $t_i \rightarrow t_{i+1}$ with accrual factor $\tau_i \simeq t_i - t_{i+1}$ means:

- Upon deposit of a notional at time t_i, at the later time t_{i+1} the notional plus interest amounting to $f_i \cdot \tau_i$ times the notional is returned to the depositor.
- For a borrower of money, the effetive funding discount factor over the (forward) interval $t_i \rightarrow t_{i+1}$ is given by $1/\left(1 + f_i\tau_i\right)$.

The fair value of a Forward Rate Agreement on rate f_i struck at K is $P(t, t_{i+1}) \cdot (f_i - K)\,\tau_i$ where $P(t, t_{i+1})$ is the value at time t of a zero coupon bond paying one domestic currency unit at time t_{i+1}.

At time t_i, the value becomes $(f_i - K)\,\tau_i/\left(1 + f_i\tau_i\right)$.

Cash settlement

III. Standard Libor market model dynamics

In the standard Libor market model for discretely compounded interest rates, we assume that each of n spanning forward rates f_i evolves according to the stochastic differential equation

$df_i = f_i \mu_i dt + f_i \sigma_i d\tilde{W}_i$

$$\frac{\mathrm{d}f_i}{f_i} = \mu_i(\boldsymbol{f}, t)\mathrm{d}t + \sigma_i(t)\mathrm{d}\widetilde{W}_i \ . \tag{1}$$

This ensures that all interest rates remain positive at all times. The drift terms are yet to be determined.

Correlation is incorporated by the fact that the n standard Wiener processes in equation (1) satisfy

$$\mathsf{E}\left[\mathrm{d}\widetilde{W}_i\,\mathrm{d}\widetilde{W}_j\right] = \varrho_{ij}\mathrm{d}t \ . \tag{2}$$

The elements of the instantaneous covariance matrix $C(t)$ of the n forward rates are thus

$$c_{ij}(t) = \sigma_i(t)\sigma_j(t)\varrho_{ij} \ . \tag{3}$$

Using a decomposition of $C(t)$ into a pseudo-square root $\tilde{A}$ such that

$$C = \tilde{A}\tilde{A}^\top \,, \tag{4}$$

we can transform equation (1) to

$$\frac{\mathrm{d}f_i}{f_i} = \mu_i \mathrm{d}t + \sum_j \tilde{a}_{ij}\mathrm{d}W_j \tag{5}$$

with $\mathrm{d}W_j$ being n independent standard Wiener processes where dependence on time has been omitted for clarity.

The matrix A may sometimes be referred to as the *driver* or as the *dispersion*[1] matrix.

[1] Karatzas and Shreve [KS91], page 284.

IV. Numéraire and measure

A fundamental principle of financial mathematics is that of *relative* or *numéraire denominated valuation*.

$\left(\frac{Q}{N}\right)$

- Select a tradeable security whose value $N(t)$ can be readily determined at any point in time of interest for the valuation of the given contingent claim at hand. This security will be called the *numéraire*.
- Establish an equivalent martingale measure in which the values of *all* tradeable securities *relative to the numéraire* are martingales.
- Carry out all calculations as values *relative to the numéraire*.
- Multiply the numéraire-denominated result by today's value of the numéraire asset.

See also en.wikipedia.org/wiki/Numéraire, the 1995 article by Geman, Karoui, and Rochet [GKR95], or [BBW01] for a less formal exposition.

In other words, the value $v(t)$ of a contingent claim that pays a sequence of conditional cashflows $c_j(t_j)$ is to be computed as

$$E_P\left[\frac{c(T)}{N(T)}\right] = \left(\frac{v_T(0)}{N(0)}\right)$$

$$v(t) \;=\; N(0)\cdot \mathsf{E}_{\mathbb{Q}_N}\left[\sum_j \frac{c_j(t_j)}{N(t_j)}\right] . \tag{6}$$

- *Discounting is but one very special case of relative valuation induced by choosing a cash account as the numéraire.*

- *The choice of numéraire is arbitrary and does not change absolute values.*

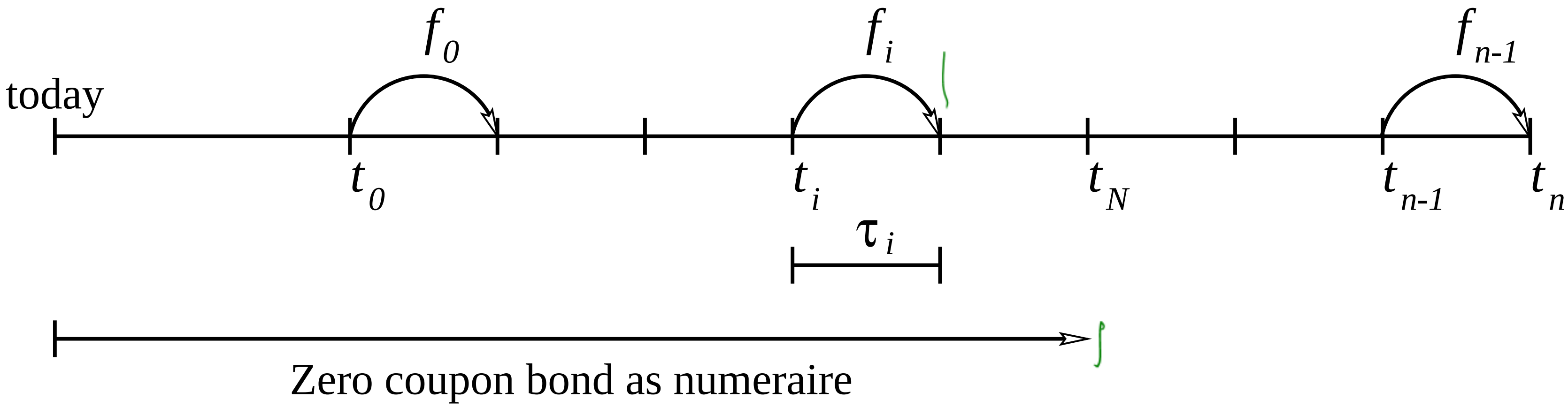
- A good choice of numéraire can make life a lot easier.
  - To value a caplet, a zero coupon bond makes a good numéraire.
  - For a swaption, the associated annuity is the natural choice.
  - For an option to exchange one asset for another [Mar78], select one of the two assets as numéraire.

$(f_i - K) \cdot \tau_i \cdot P_{i+1}$

V. The drift

$f_i \cdot \tau_i \cdot P_{i+1}$

The common choice of numéraire is some zero coupon bond that pays one currency unit at t_N.



The drifts μ_i in equations (1) and (5) can then be calculated with the following Ansatz:

$$\mathsf{E}_{t_N}\left[\mathrm{d}\left(\frac{f_i P_{i+1}}{P_N}\right)\right] = 0, \tag{7}$$

$= f_i \cdot X_i$

where P_i, $\forall\, i = 0, \ldots, n$ are the t_i discount bonds and $\mathsf{E}_{t_N}[\cdot]$ the expectation operator under the equivalent martingale measure induced by the choice of the discount bond P_N as numéraire.

Under the latter assumption, by the fundamental theorem of asset pricing, for the market to be free of arbitrage, all ratios of tradeable assets divided by the numéraire value have to form martingales [HP81], i.e. we also require

$$\mathsf{E}_{t_N}\left[\mathrm{d}\left(\frac{P_i}{P_N}\right)\right] = 0, \qquad \forall\, i = 0, \ldots, n, \tag{8}$$

since the discount bonds are assumed to be traded assets.

Now, introducing the bond ratio

$$X_i := P_{i+1}/P_N \tag{9}$$

and invoking Itô's formula on equation (7) yields

$= 0$

$$\mathsf{E}_{t_N}[X_i \mathrm{d}f_i + f_i \mathrm{d}X_i + \mathrm{d}f_i \mathrm{d}X_i] = 0\;. \tag{10}$$

Since $\mathrm{d}X$ is drift-free, this reduces to $X_i(t)$

$$\mathsf{E}_{t_N}[X_i\,\mathrm{d}f_i + \mathrm{d}X_i\,\mathrm{d}f_i] \;=\; \mathsf{E}_{t_N}[X_i\mu_i f_i \mathrm{d}t] + \mathsf{E}_{t_N}[\mathrm{d}X_i\,\mathrm{d}f_i] \;=\; 0\;. \tag{11}$$

$X_i(\mu_i f_i\, dt + \sigma_i f_i\, dW_i)$

In the following, we will use the instantaneous relative covariance brackets $[a,b]$ defined[2] by the instantaneous drift of the product of the infinitesimal *relative* increments of any two stochastic processes a and b, i.e.

$$[a,b] \;:=\; \mathsf{E}\left[\frac{\mathrm{d}a}{a}\frac{\mathrm{d}b}{b}\right] \Big/ \mathrm{d}t \,. \tag{12}$$

The definition (12) immediately gives us

$$[a,bc] = [a,b] + [a,c] \qquad \text{and} \qquad [a,b/c] = [a,b] - [a,c] \,. \tag{13}$$

Using this notation, we return to the derivation of the drift of the discrete forward rates. From equation (11), we obtain

$$\mu_i = -\left[f_i, \frac{P_{i+1}}{P_N}\right] \tag{14}$$

[2] Rebonato [Reb98] refers to the bracket $[a,b]$ as the *Vaillant bracket* quoting a company-internal report of Barclays BZW by N. Vaillant in 1995.

whose computation is aided by the bracket notation:

$$\left[f_i, \frac{P_{i+1}}{P_N}\right] = [f_i, P_{i+1}] - [f_i, P_N]$$
$$= \left[f_i, \Pi_{k=0}^{i}(1+f_k\tau_k)^{-1}\right] - \left[f_i, \Pi_{k=0}^{N-1}(1+f_k\tau_k)^{-1}\right]$$
$$= -\sum_{k=0}^{i}[f_i, 1+f_k\tau_k] + \sum_{k=0}^{N-1}[f_i, 1+f_k\tau_k]\ . \tag{15}$$

μ_i $= -\sum_{k=i+1}^{N-1}[f_i, 1+f_k\tau_k]$

By the definition of the bracket (12) and the dynamics of the individual forward rates (1), each of the terms in the sums on the right hand side of equation (15) is easily computed:

$$[f_i, 1+f_k\tau_k] = \frac{f_k\tau_k}{1+f_k\tau_k}\sigma_i\sigma_k\varrho_{ik} \tag{16}$$

Finally, cancellation of summation terms leads to the drift formulæ

$$\mu_i(\boldsymbol{f}(t),t) = \begin{cases} -\sigma_i \sum_{k=i+1}^{N-1} \frac{f_k(t)\tau_k}{1+f_k(t)\tau_k} \sigma_k \varrho_{ik} & \text{for} \quad i < N-1 \\ 0 & \text{for} \quad i = N-1 \\ \sigma_i \sum_{k=N}^{i} \frac{f_k(t)\tau_k}{1+f_k(t)\tau_k} \sigma_k \varrho_{ik} & \text{for} \quad i \geq N \end{cases} \tag{17}$$

Note that this means that the drift of all of the forward rates but one are indirectly stochastic, i.e. it is stochastic due to its explicit dependence on the stochastic forward rates themselves.

When $i = N-1$, i.e. for a drift-free forward rate f_i, we call the numéraire associated with the pricing measure the *natural numéraire* of the forward rate f_i.

VI. Factor reduction

It is possible to drive the evolution of the n forward rates with fewer underlying independent standard Wiener processes than there are forward rates, say only m of them.

$\vec{\tilde{w}}(t) \not\to \vec{f}(t)$

In this case, the coefficient matrix $\tilde{A} \in \mathbb{R}^{n\times n}$ is to be replaced by $A \in \mathbb{R}^{n\times m}$ which must satisfy

$$\sum_{j=1}^{m} a_{ij}^2 = c_{ii} \tag{18}$$

in order to retain the calibration of the options on the FRAs, i.e. the caplets. In practice, this can be done very easily by calculating the decomposition as in equation (4) as before and rescaling according to

$C = \tilde{A} \cdot \tilde{A}^T$

$$\underline{a_{ij}} = \tilde{a}_{ij}\sqrt{\frac{c_{ii}}{\sum_{k=1}^{m}\tilde{a}_{ik}^2}} \,. \tag{19}$$

The effect of this procedure is that the individual variances of each of the rates are still correct, even if we have reduced the number of driving factors to one, but the effective covariances will differ.

Using fewer factors than discrete forward rates means

- a destruction of
  - either the term structure of instantaneous volatility of FRAs
  - or the correlation structure of the FRAs
  - or both
- that simultaneous calibration to market instruments of different nature such as caplets and swaptions becomes practically impossible
- the model loses its *market* feature and becomes a *factor* model
- virtually no speed gain unless you have significantly fewer than $n/4$ factors

VII. Parametrisation of volatility and correlation

Stable calibration of any market model relies on the specification of a robust yet flexible a *reference volatility structure*. We call a specification of instantaneous volatility *time-homogeneous* or *stationary* if the volatility of any forward rate f_T that will fix at time T depends on calendar time t only in terms of $T-t$, i.e.

$$\sigma_T(t) = \sigma(T-t) \; . \tag{20}$$

One cannot fit many market prices with this strict assumption.

In fact, there are frequently good economic reasons why time-homogeneity may not be given for the term structure of instantaneous volatility of forward rates.

In practice, we may want to use an initial parametrised fit in order to find the values a, b, c, and d, such that (only) the caplet implied volatilities resulting from the instantaneous FRA volatility

$$\sigma_i(t) = k_i \left[(a + b \cdot (t_i - t)) \cdot \mathrm{e}^{-c \cdot (t_i - t)} + d \right] \cdot \mathbf{1}_{\{t < t_i\}} \tag{21}$$

are perfectly matched to the caplet volatility entries in the swaption matrix with all of the adjustment coefficients k_i being as near to 1 as possible. Note that this is only to obtain a reasonable skeleton for the term strucure of FRA volatility. The so obtained parameters a, b, c, and d then determine the *reference* or *skeleton* term structure of instantaneous volatility

$$\sigma_T^{\text{reference}}(t) = \Big[(a + b \cdot (T - t)) \cdot \mathrm{e}^{-c\cdot(T-t)} + d\Big] \cdot \mathbf{1}_{\{t<T\}} \,. \tag{22}$$

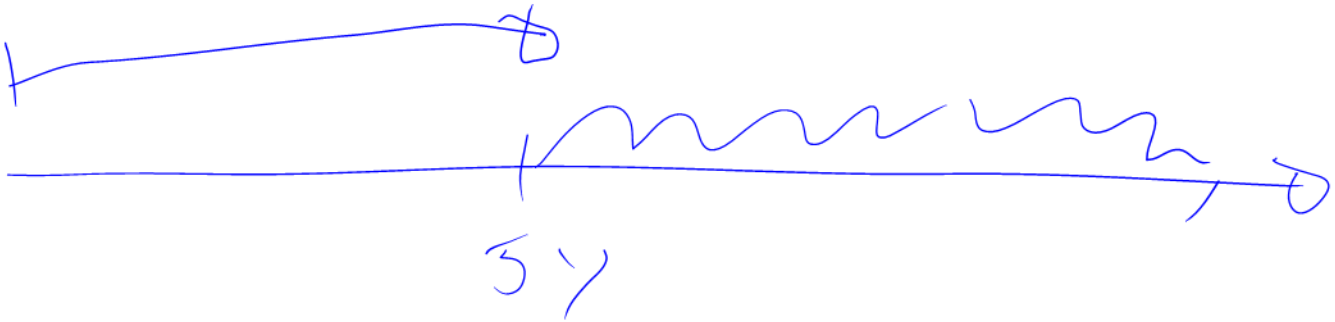
As for the instantaneous correlation between forward rates, a parametrisation that is economically, econometrically, and analytically appealing is the functional form

$$\varrho_{ij} \;=\; \mathrm{e}^{-\beta\cdot(t_i-t_j)} \tag{23}$$

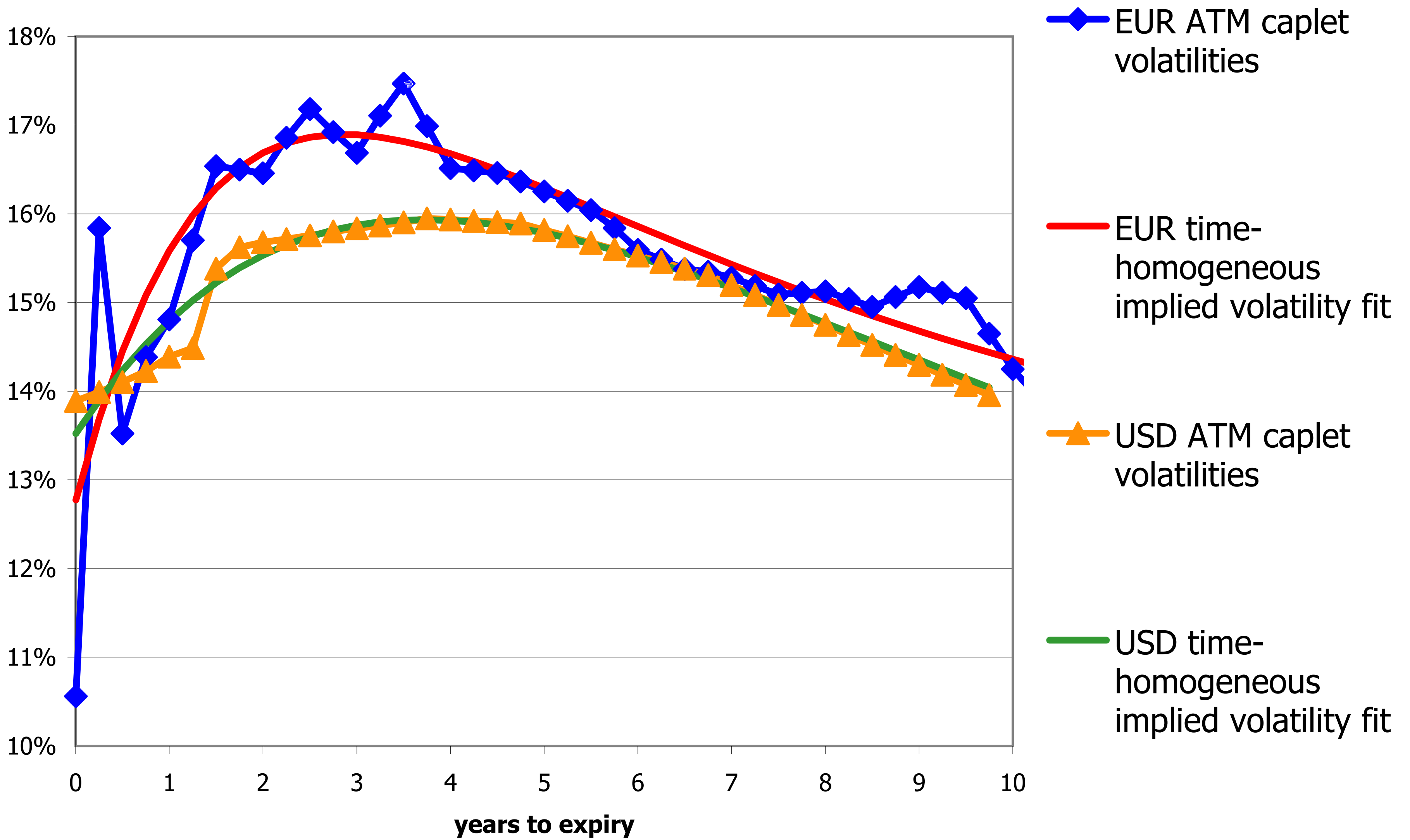
with t_i and t_j, as before, being the expiry times of caplets #i and #j. A good value for the overall correlation coefficient is $\beta \approx 0.1$.

Since the instantaneous correlation function doesn't actually depend on calendar time t, integrated FRA/FRA covariances can be computed analytically:

$$
\begin{aligned}
\int \varrho_{ij}\sigma_i(t)\sigma_j(t)\,\mathrm{d}t \quad = \quad & \mathrm{e}^{-\beta\left|t_i - t_j\right|} \cdot k_i k_j \cdot \frac{1}{4c^3} \cdot \\
& \cdot \Bigg(4ac^2 d \left[\mathrm{e}^{c(t-t_j)} + \mathrm{e}^{c(t-t_i)} \right] + 4c^3 d^2 t \\
& \quad - 4bcd\mathrm{e}^{c(t-t_i)} \left[c(t-t_i) - 1 \right] - 4bcd\mathrm{e}^{c(t-t_j)} \left[c(t-t_j) - 1 \right] \\
& \quad + \mathrm{e}^{c(2t-t_i-t_j)} \Big(2a^2c^2 + 2abc \left[1 + c(t_i + t_j - 2t) \right] \\
& \qquad\qquad + b^2 \left[1 + 2c^2(t-t_i)(t-t_j) + c(t_i + t_j - 2t) \right] \Big) \Bigg)
\end{aligned}
\tag{24}
$$



The quality of a reference fit of the implied volatilities consistent with equation (21) for a typical yield curve and caplet market both in EUR and USD is usually very good.

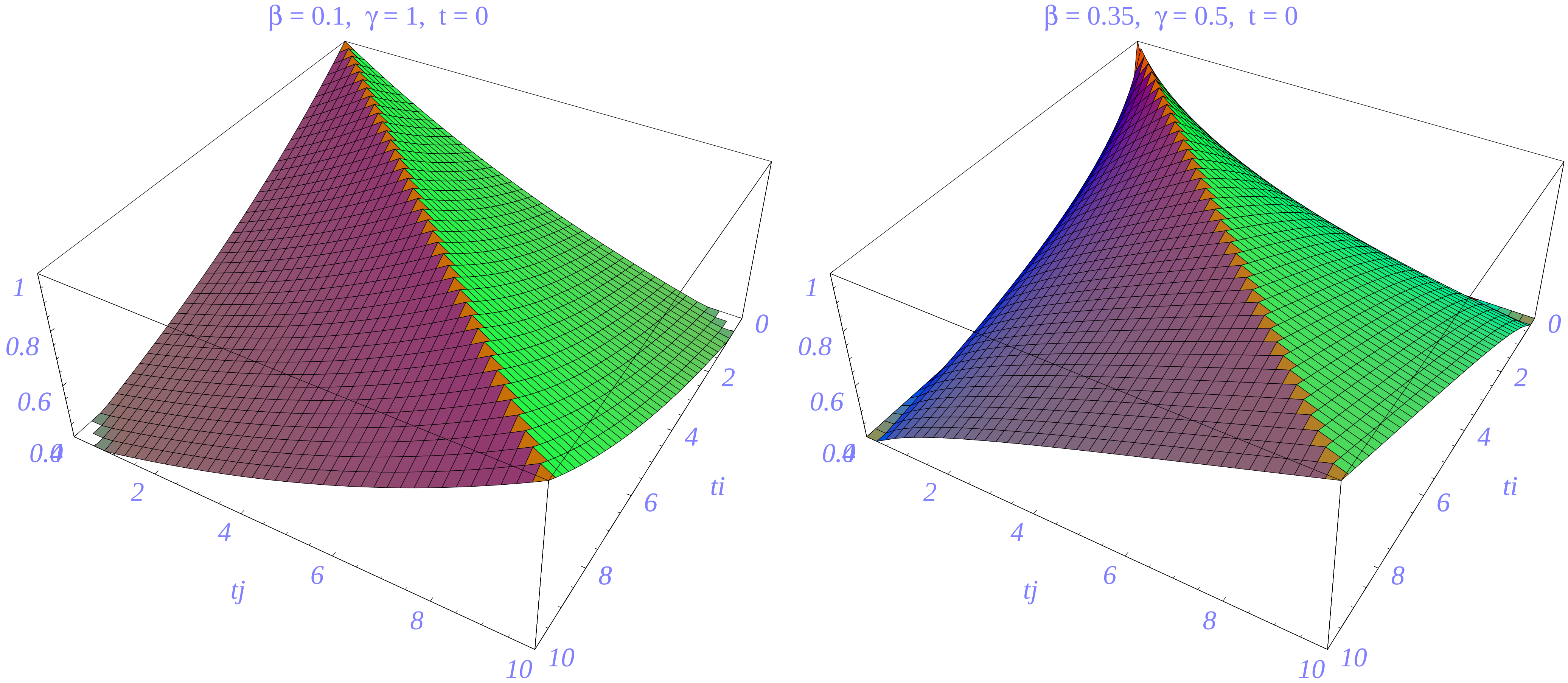


An alternative to the time independent correlation function (23) is

$e^{-\beta\cdot(t_i - t_j)}$

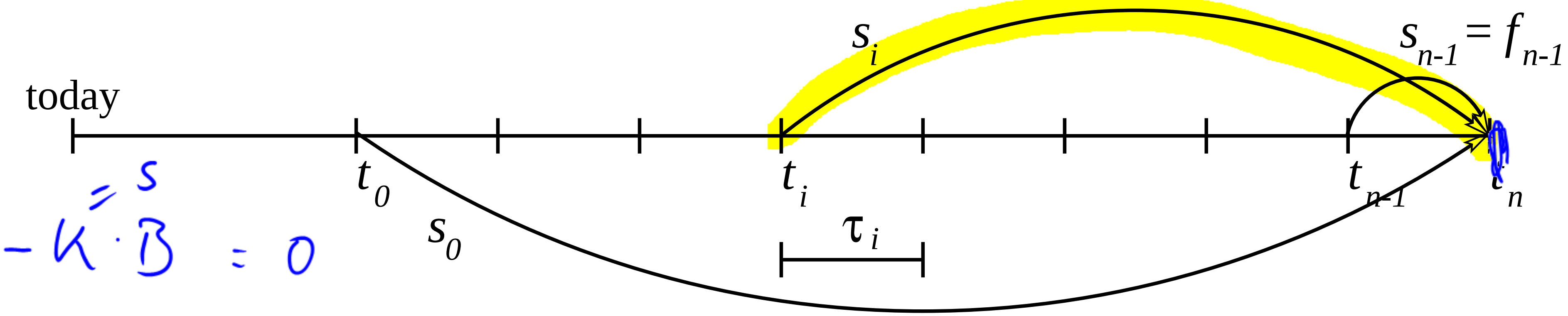
$$\varrho_{ij}(t) \quad = \quad (1-\eta)\cdot \mathrm{e}^{-\beta\left|(t_i-t)^\gamma-(t_j-t)^\gamma\right|} \; + \; \eta \tag{25}$$

with $\eta \in [-1, 1]$. Clearly, for $\gamma = 1$ and $\eta = 0$ this functional form is identical to (23). For the functional form (25), suitable parameters are $\gamma \approx 0.5$, $\beta \approx 0.35$, and $\eta \approx 0$.



VIII. Calibration to European swaptions

A forward swap rate s_i



$A - \overset{=s}{K} \cdot B = 0$

(starting with the reset time of the forward rate f_i) can be written as the ratio

$$s_i = \frac{A_i}{B_i} \tag{26}$$

of the floating leg value

$$A_i = \sum_{j=i}^{n-1} P_{j+1} f_j \tau_j N_j \quad \text{for} \quad i = 0 \ldots n-1 \tag{27}$$

and the annuity

$$B_i = \sum_{j=i}^{n-1} P_{j+1} \tau_j N_j \quad \text{for} \quad i = 0 \ldots n-1 \,. \tag{28}$$

fixed leg npv = $K \cdot B$

N_j is the notional associated with accrual period τ_j.

Since the market convention of price quotation for European swaptions uses the concept of implied Black volatilities for the forward swap rate, it seems appropriate to think of the swap rates' covariance matrix in relative terms just as much as for the forward rates themselves.

For a set of coterminal swaps all ending with a final payment at t_n, the elements of the swap rate covariance matrix C^s can therefore be written as

$$\begin{aligned} C^s_{ij} &= \left\langle \frac{\mathrm{d}s_i}{s_i} \cdot \frac{\mathrm{d}s_j}{s_j} \right\rangle \Big/ \mathrm{d}t \\ &= \sum_{k=0,l=0}^{n-1,n-1} \frac{\frac{\partial s_i}{\partial f_k} \cdot \frac{\partial s_j}{\partial f_l}}{s_i \cdot s_j} \cdot f_k f_l \cdot \left\langle \frac{\mathrm{d}f_k}{f_k} \frac{\mathrm{d}f_l}{f_l} \right\rangle \Big/ \mathrm{d}t \\ &= \sum_{k=0,l=0}^{n-1,n-1} \frac{\partial s_i}{\partial f_k} \frac{f_k}{s_i} \cdot C^f_{kl} \cdot \frac{f_l}{s_j} \frac{\partial s_j}{\partial f_l} \,. \end{aligned} \tag{29}$$

Defining the elements of the matrix $Z^{f\to s}$ by

$$Z_{ik}^{f\to s} = \frac{\partial s_i}{\partial f_k}\frac{f_k}{s_i}\ , \tag{30}$$

the mapping from the FRA covariance matrix C^{FRA} to the swap rate covariance matrix C^s can be seen as a matrix multiplication:

$$C^s = Z^{f\to s}\cdot C^f \cdot {Z^{f\to s}}^\top \qquad . \tag{31}$$

Equations (30) and (31) are the basis of all fast constructive calibration algorithms.

When the floating and fixed payments of a swap occur simultaneously with the same frequency, it is possible to find a simple formula for the swap rate coefficients $Z_{ik}^{f\to s}$. Using

$$\frac{\partial P_{i+1}}{\partial f_k} = -P_{i+1}\frac{\tau_k}{1+f_k\tau_k}\cdot \mathbf{1}_{\{k\geq i\}}\ , \tag{32}$$

where $\mathbf{1}_{\{k\geq i\}}$ is one if $k \geq i$ and zero otherwise, and equations (27), (28), and (26), we have

$$\frac{\partial s_i}{\partial f_k} = \left\{ \frac{P_{k+1}\tau_k N_k}{B_i} - \frac{\tau_k}{1+f_k\tau_k} \cdot \frac{A_k}{B_i} + \frac{\tau_k}{1+f_k\tau_k} \cdot \frac{A_i B_k}{B_i^2} \right\} \cdot \mathbf{1}_{\{k\geq i\}} \ . \tag{33}$$

This enables us to calculate the elements of the forward rate to swap rate covariance transformation matrix $Z^{f\rightarrow s}$ to obtain the expression

$$Z_{ik}^{f\rightarrow s} = \left[\underbrace{\frac{P_{k+1}N_k f_k \tau_k}{A_i}}_{\text{constant weights approximation}} + \underbrace{\frac{(A_i B_k - A_k B_i) f_k \tau_k}{A_i B_i (1+f_k\tau_k)}}_{\text{shape correction}} \right] \cdot \mathbf{1}_{\{k\geq i\}} \ . \tag{34}$$

The second term inside the square brackets of equation (34) is a *shape correction*. Rewriting it as

$$\frac{(A_i B_k - A_k B_i) f_k \tau_k}{A_i B_i (1 + f_k \tau_k)} = \frac{f_k \tau_k}{A_i B_i (1 + f_k \tau_k)} \cdot \sum_{l=i}^{k-1} \sum_{m=k}^{n-1} P_{l+1} P_{m+1} N_l N_m \tau_l \tau_m (f_l - f_m) \quad (35)$$

highlights that it is a weighted average over inhomogeneities of the yield curve.

In fact, for a flat yield curve, all of the terms $(f_l - f_m)$ are identically zero and the mapping matrix $Z^{f \to s}$ is equivalent to the so-called constant-weights approximation.

In practice the yield curve is never entirely flat which makes it necessary to compute the swap rate coefficients via the full derivative calculation (30).

When floating and fixed schedules differ, we have to compute the partial dependencies of the swap rates' floating payments, floating payment discount factors, and fixed payment discount factors individually.

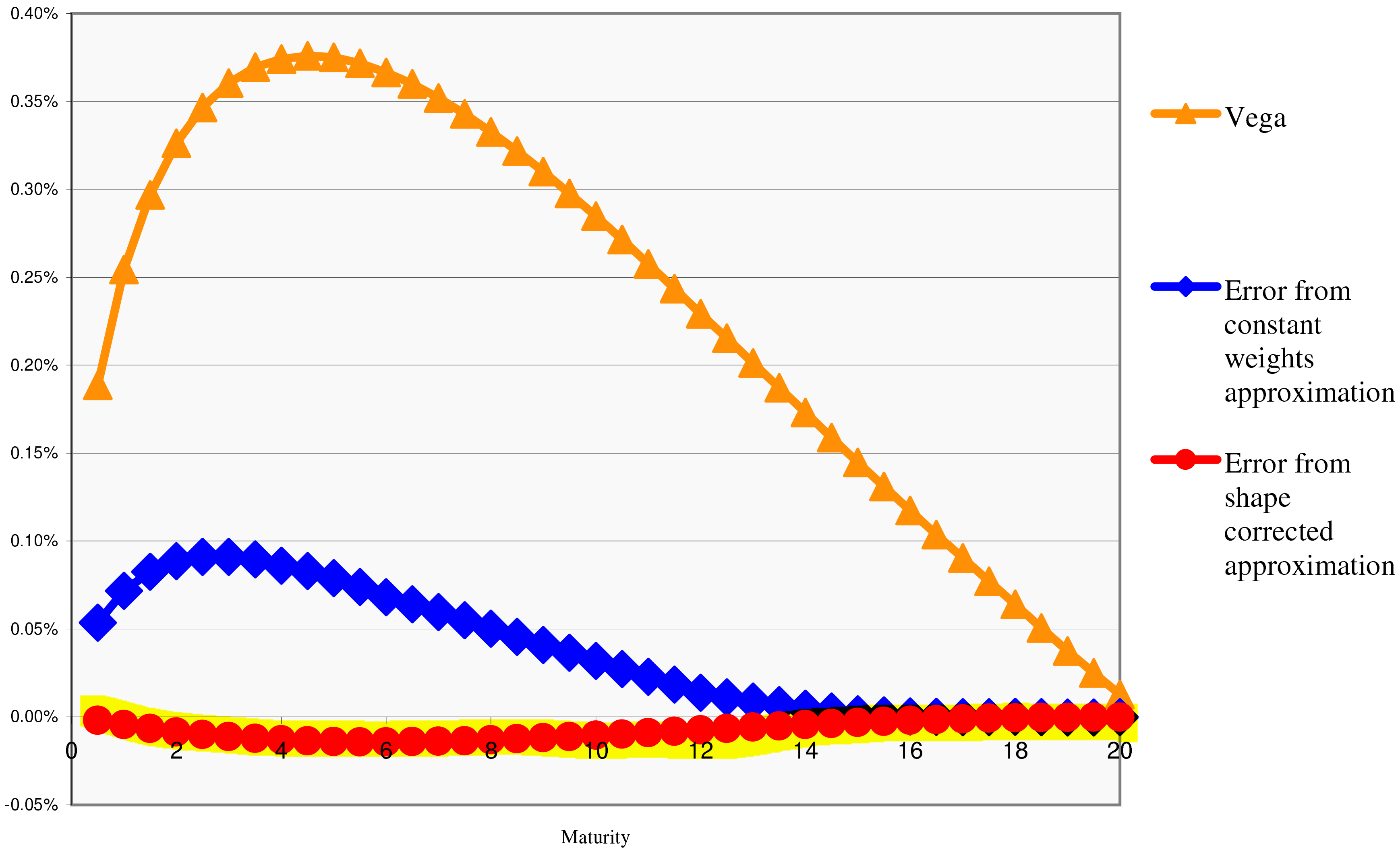
As things stand at this point, we have a map between the instantaneous FRA/FRA covariance matrix and the instantaneous swap/swap covariance matrix.

Unfortunately, though, the map involves the state of the yield curve at any one given point in time via the matrix Z.

The price of a European swaption, though, does not just depend on one single realised state or even path of instantaneous volatility. It is much more appropriate to think about some kind of *path integral average volatility*. Using arguments of factor decomposition and equal probability of up and down moves (in log space), it can be shown [JR00, Kaw02, Kaw03] that the specific structure of the map allows us to approximate the effective implied swaption volatilities by simply using today's state of the yield curve for the calculation of the mapping matrix Z:

$$\hat{\sigma}_{s_i}(t,T) = \sqrt{\sum_{k=i,l=i}^{n-1} Z_{ik}^{f\to s}(0) \cdot \frac{\int_t^T \sigma_k(t')\sigma_l(t')\varrho_{kl}\mathrm{d}t'}{T-t} \cdot Z_{il}^{f\to s}(0)} \tag{36}$$

This approximate equivalent implied volatility can now be substituted into the Black swaption formula to produce a price *without the need for a single simulation*. In practice, the formula (36) works remarkably well.



We can now design a non-iterative calibration procedure that connects the stepwise covariance matrices of the logarithms of the realisations of the forward rates directly to the calibration volatilities of a set of European swaptions (including caplets). For any given time step from t to T, populate the time-unscaled FRA/FRA covariance matrix

$$C^f_{kl} = \frac{\int_{t'=t}^{T} \sigma_k(t')\sigma_l(t')\varrho_{kl}(t')\mathrm{d}t'}{T-t} \ . \tag{37}$$

Next, map this matrix into a time-unscaled swap/swap covariance matrix using the Z matrix calculated from the initial state of the yield curve

$$C^s = Z \cdot C^f \cdot Z^\top \ . \tag{38}$$

Note that this swap rate/swap rate covariance matrix is associated with forward swap rates that expire at times equal to or later than T. Its diagonal elements are the mean square volatilities of the n swap rates over the time step $t \to T$. For $t = 0$ and $T = t_1$, we notice that the diagonal element C^s_{11} represents the square of the FRA-covariance matrix implied Black volatility of the first swaption, which, if the model was already calibrated, should equate the market implied volatility of the swaption expiring at time t_1 denoted by $\sigma_{s_1}^{\text{market}}$. Since variances are additive, we have

$$C_{ij}^{s}(0,t_k)\cdot t_k = C_{ij}^{s}(0,t_{k-1})\cdot t_{k-1} + C_{ij}^{s}(t_{k-1},t_k)\cdot(t_k - t_{k-1}) \quad \text{for} \quad k \geq \max(i,j)\ . \tag{39}$$

In other words, we can compute the time-integrated (smaller) covariance matrix for a set of swaptions expiring at a later date by adding a subset of the (larger) time-integrated covariance matrix to an earlier date and the time-integrated covariance matrix from that earlier date to the later date.

This additive feature of covariances means that we can accomplish calibration of each swaption individually by rescaling the whole swap rate covariance matrix such that the diagonal elements, when averaged to the expiry date of any individual swaption, match the square of the respective market given implied volatility.

For this purpose, define the diagonal matrix Ξ by

$$\Xi_{gh} = \frac{\hat{\sigma}_{s_h}^{\text{market}}}{\hat{\sigma}_{s_h}(0, t_h)} \cdot \delta_{gh} \tag{40}$$

with δ_{gh} being the Kronecker symbol (which is zero unless $g = h$ when it is one) and $\hat{\sigma}_{s_h}(0, t_h)$ calculated from the FRA instantaneous volatility parametrisation through equation (36). The calibrated swap rate/swap rate covariance matrix for any time step $t \to T$ is thus given by

$$C^{s}_{\text{calibrated}} = \Xi \cdot Z \cdot C^{f} \cdot Z^{\top} \cdot \Xi \,. \tag{41}$$

When Z is invertible, we can therefore define the calibration matrix

$$M := Z^{-1} \cdot \Xi \cdot Z \tag{42}$$

and express the entire calibration procedure as the simple operation

$$C^{f}_{\text{calibrated}} = M \cdot C^{f}_{\text{parametric}} \cdot M^{\top} \,. \tag{43}$$

In order to use the matrix $C^f_{\text{calibrated}}$ for the evolution of the yield curve over the time step $t \longrightarrow T$ from a set of standard normal variates, we now simply need to compute a pseudo-square root $A^f_{\text{calibrated}}$ such that

$$C^f_{\text{calibrated}} = A^f_{\text{calibrated}} \, {A^f_{\text{calibrated}}}^\top \tag{44}$$

just as we would have done without calibration to swaptions.

In practice, a user may wish to specify not exactly as many swaptions as there are forward rates to calibrate to. Instead, it may be desirable to specify fewer, or even more than n swaption volatilities. In this case, the swap rate coefficient matrix Z may be over- or under-determined. Either way, it is still possible to find a matrix M that can be used in equation (43). To find it, let us first consider the singular value decomposition [PTVF92] of the transpose of Z :

$$Z^\top = U \cdot W \cdot V^\top \tag{45}$$

In the underdetermined case, the diagonal matrix W will have some zero entries on the diagonal. Let us define W' as the diagonal matrix whose diagonal elements are the inverse of the corresponding elements in W where they are nonzero, and zero otherwise. The matrix product $(W'W)$ then has unit elements wherever W has nonzero entries, and formally constitutes a *projection* matrix by virtue of the fact that its repeated application to any target vector has the same result as a single multiplication, i.e.

$$(W'W)^k \cdot X = (W'W) \cdot X \quad \forall\, k \geq 1 \text{ and } \forall\, X \ . \tag{46}$$

Now, the calibration procedure in the present framework amounts to the identification of $A^f_{\text{calibrated}}$ that satisfies

$$Z \cdot A^f_{\text{calibrated}} = \Xi \cdot Z \cdot A^f_{\text{parametric}} \tag{47}$$

but remains as close to the original $A^f_{\text{parametric}}$ as possible, i.e. to find the matrix $A^f_{\text{calibrated}}$ that meets equation (47) and simultaneously minimises

$$\left|\left|A^f_{\text{calibrated}} - A^f_{\text{parametric}}\right|\right| \tag{48}$$

for some suitable matrix norm. Now, denote the Moore-Penrose inverse [Alb72] of Z as $\widetilde{Z^{-1}}$, and write the product of $\widetilde{Z^{-1}}$ times Z itself as Q :

$$Q := {U^\top}^{-1} W'V^\top \cdot VWU^\top \tag{49}$$

By the aid of the orthogonality conditions satisfied by the constituents U and V of the singular value decomposition of Z, and by the fact that $(W'W)$ is a projection, both Q and the matrix

$$P := \mathbf{1} - Q \tag{50}$$

are also projection operators. In fact, P is the projection onto the kernel of Z. The desired matrix $A^f_{\text{calibrated}}$ can thus be found by adding the projection of $A^f_{\text{parametric}}$ onto the kernel of Z and the Moore-Penrose solution to equation (47), i.e.

$$A^f_{\text{calibrated}} = P \cdot A^f_{\text{parametric}} + \widetilde{Z^{-1}} \cdot \Xi \cdot Z \cdot A^f_{\text{parametric}} \,. \tag{51}$$

Since $A^f_{\text{parametric}}$ appears as the last multiplicand in both of the summands on right hand side, we can rewrite this as

$$\begin{aligned} A^f_{\text{calibrated}} &= (\mathbf{1} - UW'WU^\top + UW'V^\top \Xi Z) \cdot A^f_{\text{parametric}} \\ &= \left(\mathbf{1} - UW'\left(WU^\top - V\top\Xi Z\right)\right) \cdot A^f_{\text{parametric}} \\ &= \left(\mathbf{1} - UW'\left(1 - V^\top \Xi V\right) WU^\top\right) \cdot A^f_{\text{parametric}} \,. \end{aligned} \tag{52}$$

This means, the sought calibration matrix M is given by

$$M = \mathbf{1} - U \cdot W' \cdot \left(1 - V^\top \cdot \Xi \cdot V\right) \cdot W \cdot U^\top \,. \tag{53}$$

The key to this calibration procedure in the underdetermined case is that a *minimal* solution to the raw calibration problem (47) is combined with as much of the original covariance information as possible that has no effect on the calibration problem. In more formal terms, we combine the minimal solution to the calibration problem with the projection of the desired covariance structure onto the calibration kernel.

When Z is overdetermined, the correction matrix M cannot achieve calibration to all the desired market prices. Instead, the calibration procedure based on the linear algebraic operations above will result in a least squares fit in some suitable norm by virtue of the use of the singular value decomposition of Z.

Within the limits of the approximation (36), the operation given in equation (43) will provide calibration to European swaption prices whilst retaining as much calibration to the caplets as is possible without violating the overall FRA/FRA correlation structure too much.

IX. Long time steps using the predictor-corrector scheme

In order to price an exotic interest rate derivative in a Monte Carlo framework, we need to evolve the set of forward rates $\boldsymbol{f}$ from its present values into the future according to the stochastic differential equation

$$\mathrm{d}f_i(t) = f_i(t) \cdot \mu_i(\boldsymbol{f}(t), t)\mathrm{d}t + f_i(t) \cdot \sum_{j=1}^{m} a_{ij}\mathrm{d}W_j \tag{54}$$

driven by an m-dimensional standard Wiener process $\boldsymbol{W}$.

The drift terms given by equation (17) are clearly state-dependent and thus indirectly stochastic which forces us to use a numerical scheme to solve equation (54) along any one path.

A simple explicit Euler scheme would be

$$f_i^{\text{Euler}}(\boldsymbol{f}(t), t + \Delta t) = f_i(t) + f_i(t) \cdot \mu_i(\boldsymbol{f}(t), t)\Delta t + f_i(t) \cdot \sum_{j=1}^{m} a_{ij}(t) z_j \sqrt{\Delta t} \tag{55}$$

with z_j being m independent normal variates. This would imply that we approximate the drift as constant over the time step $t \rightarrow t + \Delta t$.

Moreover, this scheme effectively means that we are using a normal distribution for the evolution of the forward rates over this time step.

Whilst we may agree to the approximation of a piecewise constant (in time) drift coefficient μ_i, the normal distribution may be undesirable, especially if we envisage to use large time steps Δt for reasons of computational efficiency.

However, when we assume piecewise constant drift, we might as well carry out the integration over the time step Δt analytically and use the scheme

$c_{ii} = \int_t^{t+\Delta t} \sigma_i^2(t)\,dt$

$$f_i^{\text{Constant drift}}(\boldsymbol{f}(t), t + \Delta t) = f_i(t) \cdot \mathrm{e}^{\mu_i(\boldsymbol{f}(t),t)\Delta t - \frac{1}{2}c_{ii} + \sum\limits_{j=1}^{m} a_{ij} z_j} \tag{56}$$

whereby the time step scaling by $\sqrt{\Delta t}$ for A and by Δt for C has been absorbed into the respective matrices. In other words, we have set $A' := A \cdot \sqrt{\Delta t}$ and $C' := C \cdot \Delta t$ and dropped the primes. Equation (56) can also be viewed as the Euler scheme in logarithmic coordinates.

The above procedure works very well as long as the time steps Δt are not too long and is widely used and also referred to in publications [And00, GZ99].

Since the drift term appearing in the exponential function in equation (56) is in some sense a stochastic quantity itself, we will begin to notice that we are ignoring Jensen's inequality when the term $\mu_i \Delta t$ becomes large enough.

This happens when we choose a big step Δt, or the forward rates themselves or their volatility are large. Therefore, we should use a *hybrid predictor-corrector* method which models *only the drift* as indirectly stochastic.

A method that works very well in practice is as follows.

1. Given a current evolution of the yield curve denoted by $\boldsymbol{f}(t)$, we calculate the predicted solution $\boldsymbol{f}^{\text{Constant drift}}(\boldsymbol{f}(t), t + \Delta t)$ using one m-dimensional normal variate draw $\boldsymbol{z}$ following equation (56).

   $\vec{f}(t+\Delta t)$

2. We recalculate the drift using this evolved yield curve. The predictor-corrector approximation $\tilde{\mu}_i$ for the drift is then given by the average of

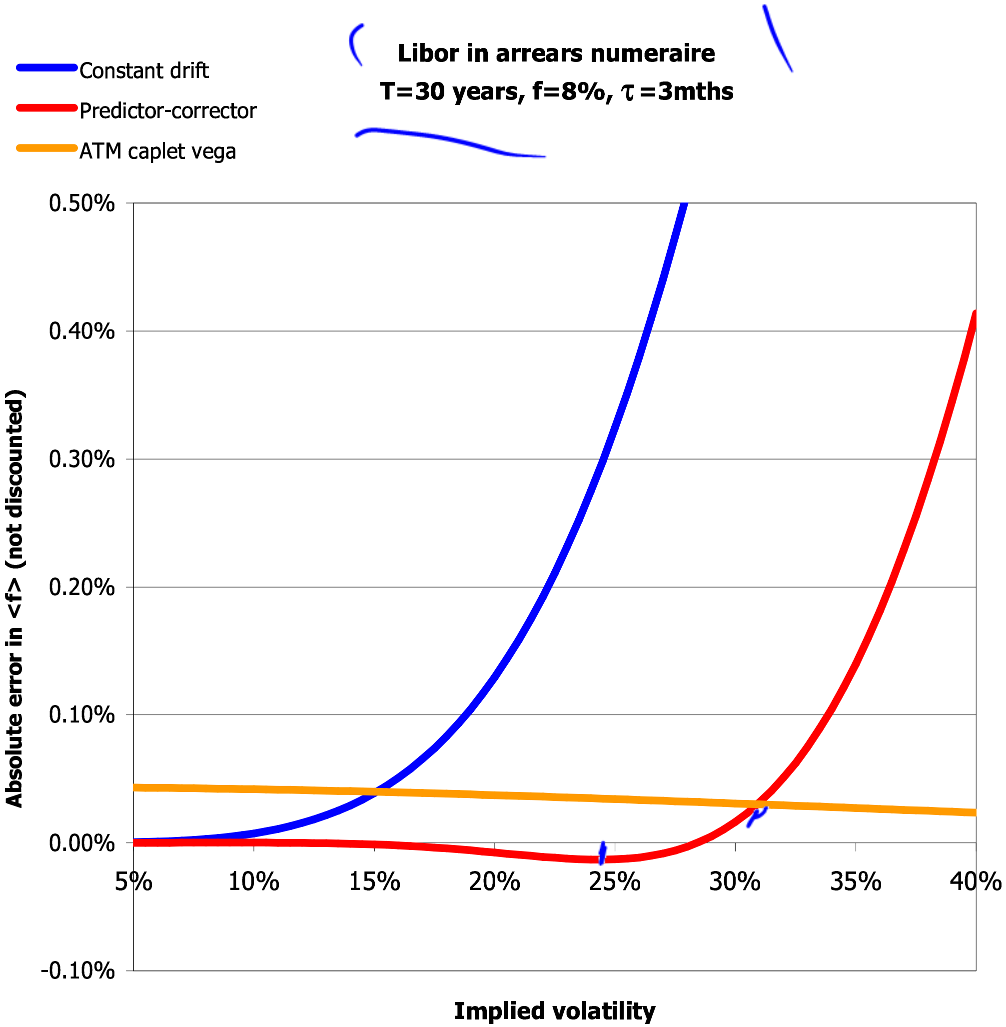
these two calculated drifts, i.e.

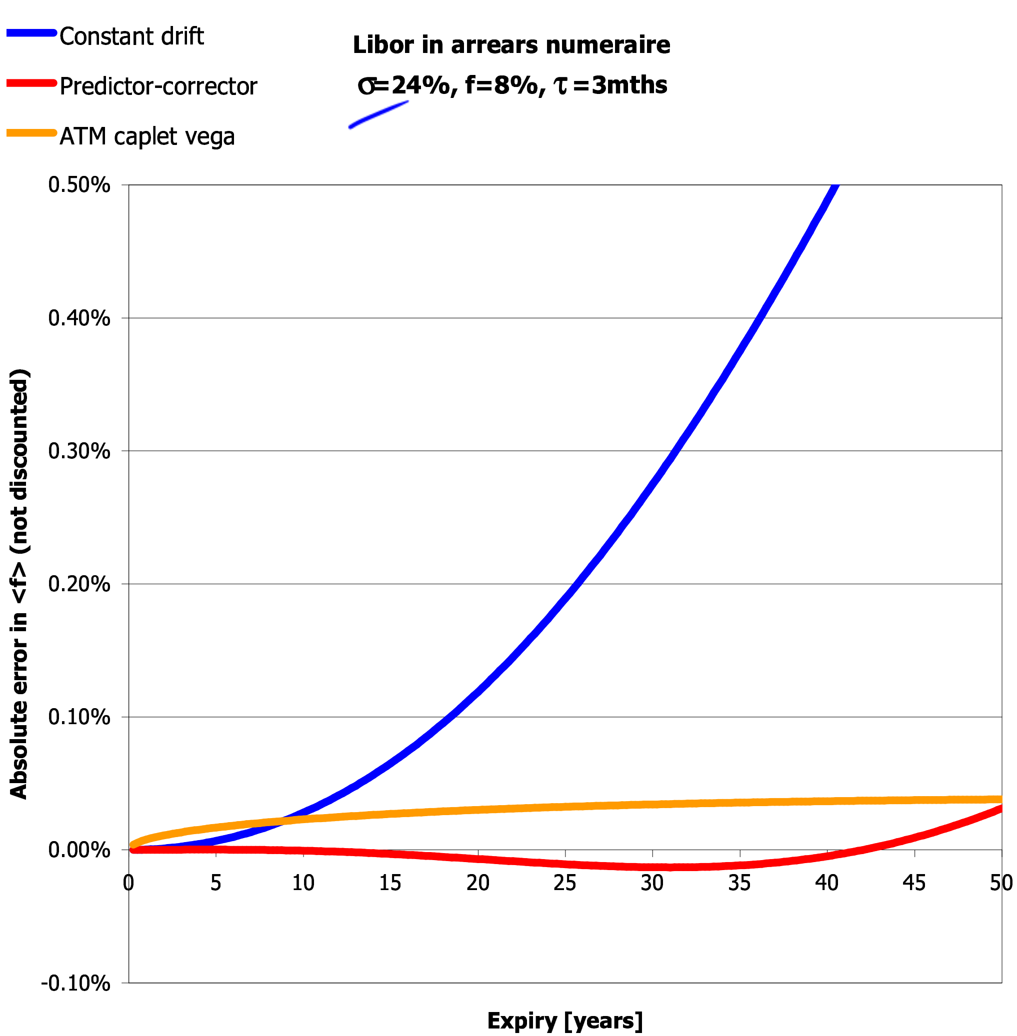
$$\tilde{\mu}_i(\boldsymbol{f}(t), t \;\to\; t + \Delta t) = {}^1\!/_2 \Big\{ \mu_i(\boldsymbol{f}(t), t) + \mu_i(\boldsymbol{f}^{\text{Constant drift}}(\boldsymbol{f}(t), t + \Delta t), t) \Big\} \;. \tag{57}$$

3. The predictor-corrector evolution is given by

$$f_i^{\text{Predictor-corrector}}(\boldsymbol{f}(t), t + \Delta t) = f_i(t) \cdot \mathrm{e}^{\tilde{\mu}_i(\boldsymbol{f}(t), t \to t + \Delta t)\Delta t - \frac{1}{2}c_{ii} + \sum_{j=1}^{m} a_{ij} z_j} \tag{58}$$

wherein we re-use the same normal variate draw $\boldsymbol{z}$, i.e. we only correct the drift of the predicted solution.





The stability of the predictor-corrector drift method as a function of volatility level (left) and time to expiry (right) for the Libor-in-arrears convexity.

⇒ The predictor-corrector drift approximation is highly accurate!

X. Bermudan swaptions

A Bermudan swaption contract denoted by 'X-non-call-Y' gives the holder the right to enter into a swap at a prespecified strike rate K on a number of exercise opportunities.

The first exercise opportunity in this case would be Y years after inception. The swap that can be entered into has always the same terminal maturity date, namely X, independent on when exercise takes place.

A Bermudan swaption that entitles the holder to enter into a swap in which he pays the fixed rate is known as *payer's*, otherwise as *receiver's*.

For the owner of a payer's Bermudan swaption, the present value of exercising at time t_j is given by the intrinsic value $I(t_j)$ of the swap to be entered into at that time

$$I(t_j) = \sum_{k=j}^{n-1} \left[P_{k+1}(t_j) \cdot (f_k(t_j) - K)\, \tau_k \right] \; . \tag{59}$$

In order to decide optimally about early exercise at time t_j, the holder compares the present intrinsic value with the expected profit to be made by not exercising at that time. Thus, the t_j-value of the Bermudan swaption $V(t_j)$ is given by

$$V(t_j) = \begin{cases} \max\left\{I(t_j), \mathsf{E}_{t_j}[V(t_{j+1})]\right\} & \text{for} \quad j = 1\ldots n-2 \\ \max\{I(t_j), 0\} & \text{for} \quad j = n-1 \end{cases} \quad . \tag{60}$$

In the marketplace, many variations are common such as differing payment frequencies between fixed and floating leg, margins on top of the floating payment, varying notionals (*roller coaster* or *amortizing* swaptions are not uncommon), time-varying strike of the swap to enter into, cross-currency payoff (quanto), and many more.

XI. The Bermudan swaption exercise domain

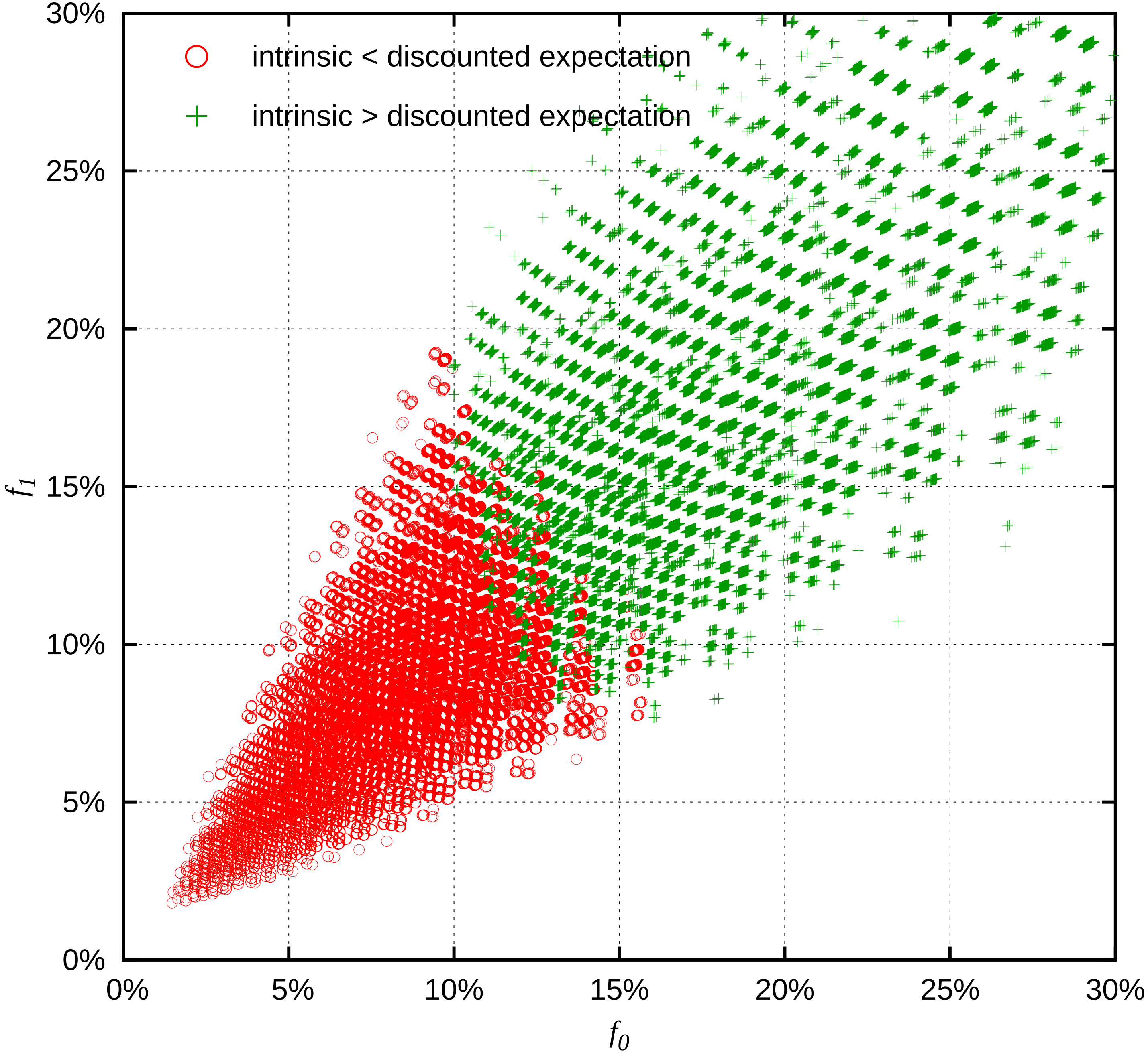
There are many ways to describe a given discretised yield curve unambiguously:

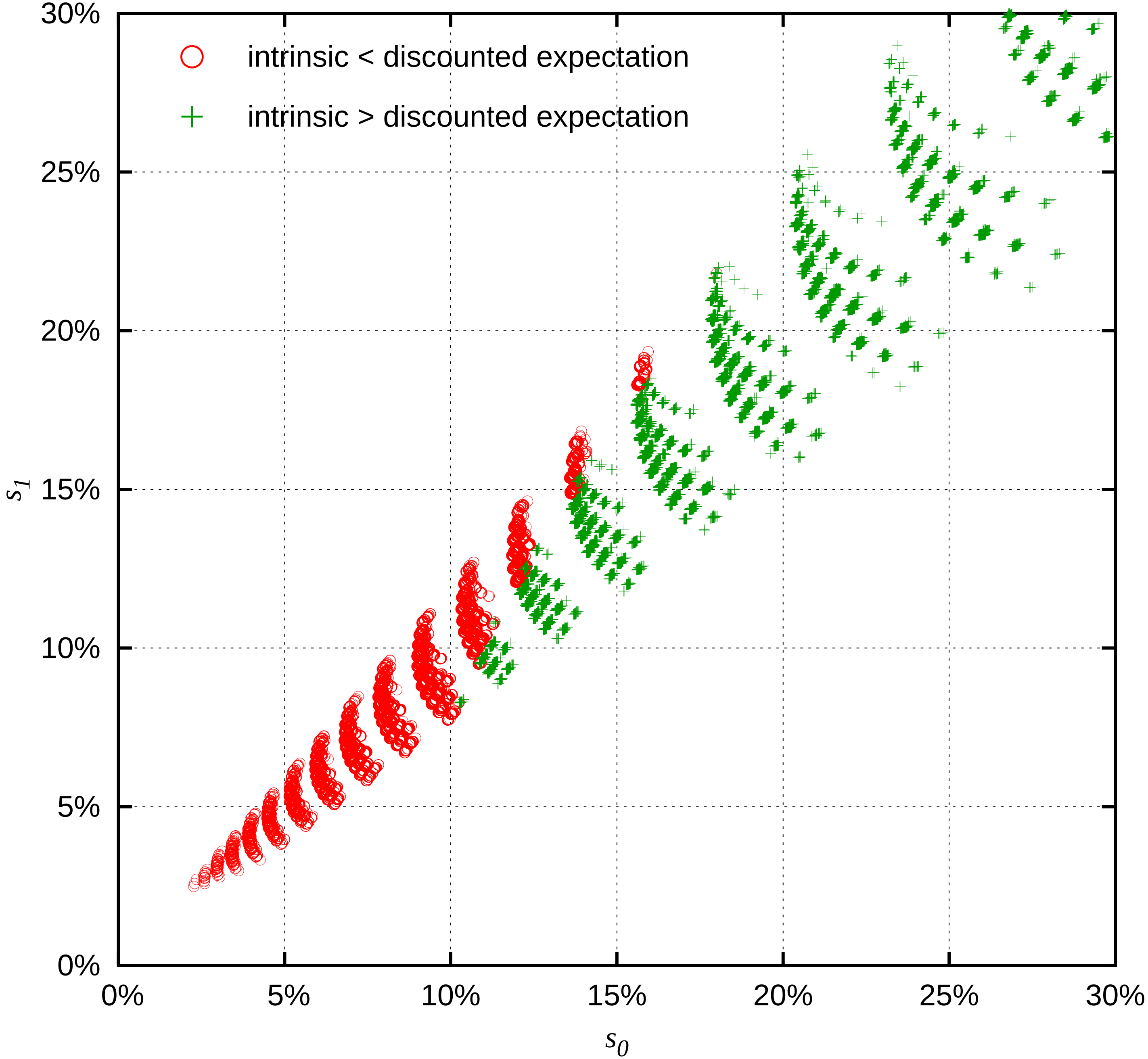
- a complete set of discrete spanning forward rates $\{f_i\}$;
- a complete set of discrete coterminal (or, indeed, coinitial) swap rates $\{s_i\}$;
- a complete set of discount factors $\{P_i\}$;

and more.

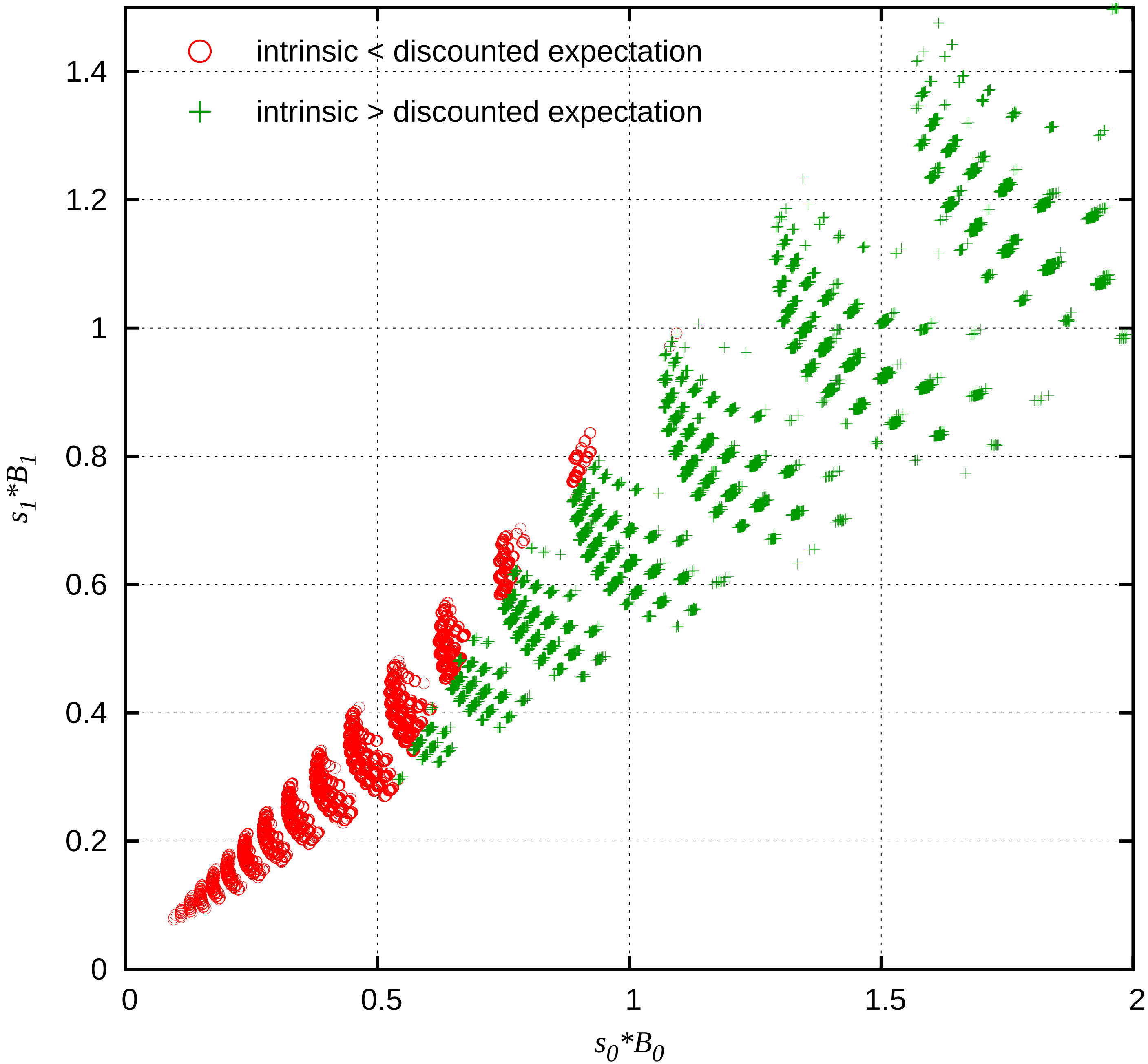
Given any complete specification of the discretised yield curve, we can project points in yield curve space onto any alternative coordinate choice we please.

It depends on our choice of projection to what extent the domain of optimal exercise and the domain of optimal continuation appear to overlap.

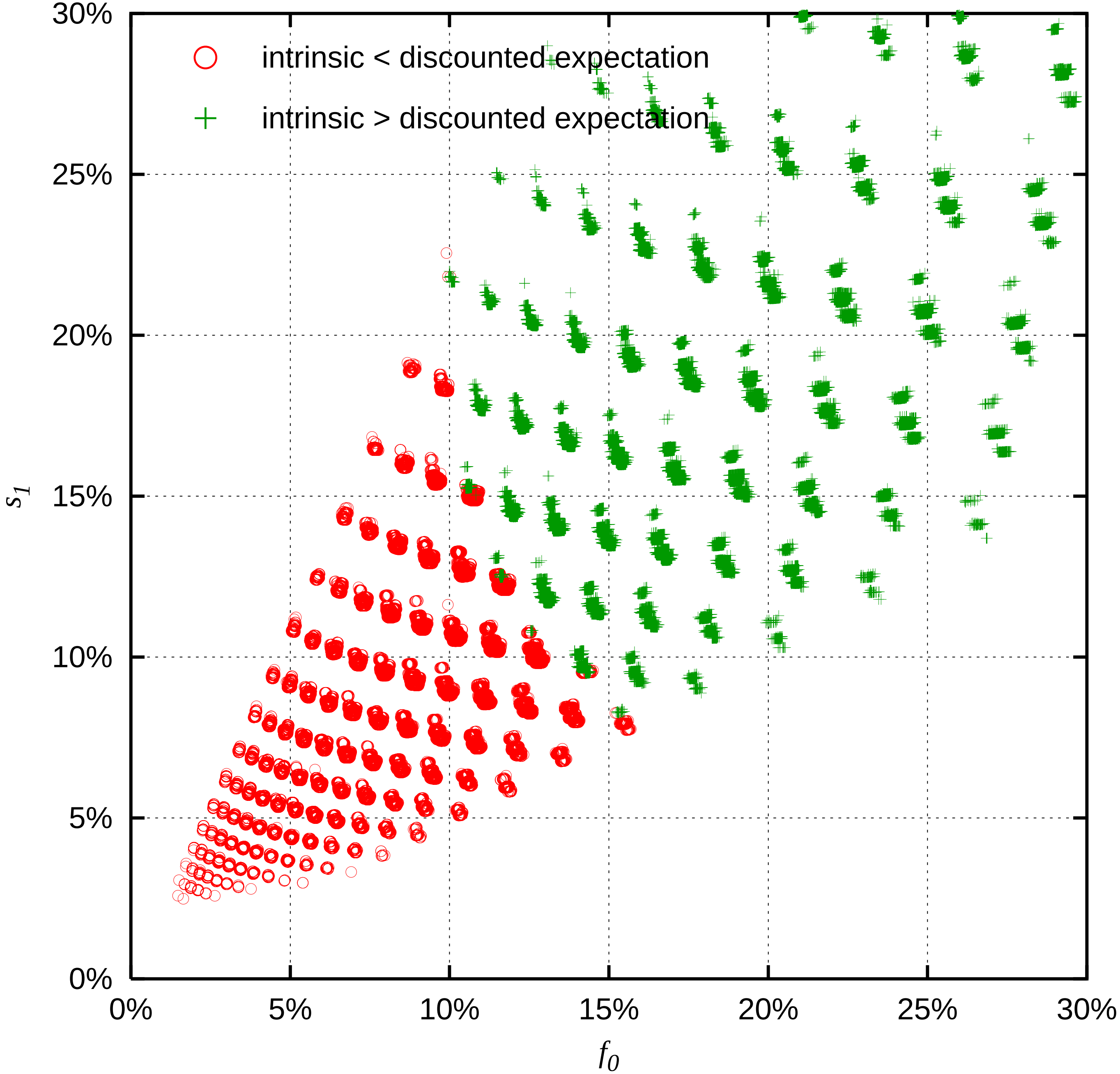

intrinsic < discounted expectation
intrinsic > discounted expectation
30%
25%
20%
15%
10%
5%
0%
0%
5%
10%
15%
20%
25%
30%
f_1
f_0



intrinsic < discounted expectation
intrinsic > discounted expectation
30%
25%
20%
15%
10%
5%
0%
0%
5%
10%
15%
20%
25%
30%
s_1
s_0



intrinsic < discounted expectation
intrinsic > discounted expectation
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
1.5
2
$s_1 * B_1$
$s_0 * B_0$


intrinsic < discounted expectation
intrinsic > discounted expectation
30%
25%
20%
15%
10%
5%
0%
0%
5%
10%
15%
20%
25%
30%
s_1
f_0

XII. Bermudan swaption exercise boundary parametrisation

Taking into account all of the heuristic observations about the shape of the exercise boundary in various projections for many different shapes of the yield curve and volatility structures, the following function can be chosen as the basis for the subsequent exercise decision strategy in the Monte Carlo simulation:

$$\mathcal{E}_i(\boldsymbol{f}(t_i)) = \phi_\pm \cdot \left(f_i(t_i) - \left[p_{i1} \cdot \frac{s_{i+1}(0)}{s_{i+1}(t_i) + p_{i2}} + p_{i3} \right] \right) \tag{61}$$

with

$$\phi_\pm = \left\{ \begin{array}{ccc} +1 & \text{for} & \text{payer's swaptions} \\ -1 & \text{for} & \text{receiver's swaptions} \end{array} \right\} \tag{62}$$

This function is hyperbolic in s_{i+1} and depends on three coefficients, the initial (i.e. at the calendar time of evaluation or inception of the derivative contract) value of $f_i(0)$ and $s_{i+1}(0)$, and their respective evolved values as given by the simulation procedure. For a more general discussion, see [JA10].

Since we have to make an exercise decision at each exercise opportunity time t_i, we allow for a new set of exercise function coefficients for each such time slice.

For non-standard Bermudan swaptions that have payments in between exercise dates, we use the shortest swap rate from t_i to the next exercise time instead of f_i. The parametric exercise decision given an evolved yield curve is then simply to exercise if $\mathcal{E}_i > 0$.

At the very last exercise opportunity at time t_{n-1} we have exact knowledge if exercise is optimal, namely when the residual swap is in the money. This easily integrates into the parametric description given by equation (61) by setting $p_{(n-1)\ 1}$ and $p_{(n-1)\ 2}$ to zero and $p_{(n-1)\ 3}$ to the strike:

$$\begin{aligned} p_{(n-1)\ 1} &= 0 \\ p_{(n-1)\ 2} &= 0 \\ p_{(n-1)\ 3} &= K \end{aligned} \tag{63}$$

XIII. The Bermudan Monte Carlo algorithm

- For a *training set* $\mathbb{P}^{\text{Training}}$ of N_{Training} evolutions of the yield curve into the future out to the last exercise time t_{n-1}, precalculate:

$$\mathbb{P}^{\text{Training}} = \{\boldsymbol{f}_{jk}\} \;\;,\; j = 1 \ldots N_{\text{Training}} \,, k = 0 \ldots n-1 \tag{64}$$

- For each evolution of the yield curve, precalculate and store the residual intrinsic value I_{jk} in the chosen numéraire as seen at time t_k.

- Carry out $n-1$ optimisations, one for each exercise opportunity t_i apart from the last one[3] in order to determine the best values to use for the coefficients p_{ij}.

The optimisations are to be done in reverse order, starting with the penultimate exercise time t_{n-2}.

[3]On the very last exercise opportunity, the optimal exercise parameters are given by equation (63) whence no optimisation is required for them.

Prior to each optimisation, we assign a path-value $v_j, j = 1 \ldots N_{\text{Training}}$ to each evolution path in the training set $\mathbb{P}^{\text{Training}}$ which represents the value of the Bermudan swaption on this path if no exercise occurs up until and including t_i.

The path-value vector $\boldsymbol{v}$ is initialised to be zero in all its elements before we enter the following loop which counts down in the time index variable i from $(n-2)$ to 0:

1. For each path $\boldsymbol{f}_{j\,(\cdot)}$ in $\mathbb{P}^{\text{Training}}$,

   if $(\mathcal{E}_{i+1}(\boldsymbol{f}_{ji}) > 0)$ **and** $(I_{j\;(i+1)} > 0)$,

      re-assign $v_j := I_{j\;(i+1)}$,

   else

      leave v_j unchanged.

2. Optimise the average of the exercise-decision dependent value

   $$U_i(\boldsymbol{p}_i) = \frac{1}{N_{\text{Training}}} \sum_{j=1}^{N_{\text{Training}}} \left\{ \begin{array}{ccc} I_{ji} & \textbf{if} & (\mathcal{E}_i(\boldsymbol{f}_{ji}\,;\,\boldsymbol{p}_i) > 0) \\ v_j & \textbf{else} & \end{array} \right\} \tag{65}$$

   over the three parameters p_{i1}, p_{i2}, and p_{i3}. Specifically, one can use the Broyden-Fletcher-Goldfarb-Shanno multi-dimensional variable metric method for this optimisation [PTVF92].

   It is worth noting that, since *absolutely all values are precalculated and stored*, the function to be optimised given by equation (65) requires merely N_{Training} evaluations of the exercise decision function $\mathcal{E}_i(\boldsymbol{f}_{ji}\,;\,\boldsymbol{p}_i)$ and the same number of additions and is thus *linear in the number of training paths and independent on the dimensionality or maturity of the problem*.

3. Decrement i by 1

4. **if** $(i \geq 0)$ continue with step 1.

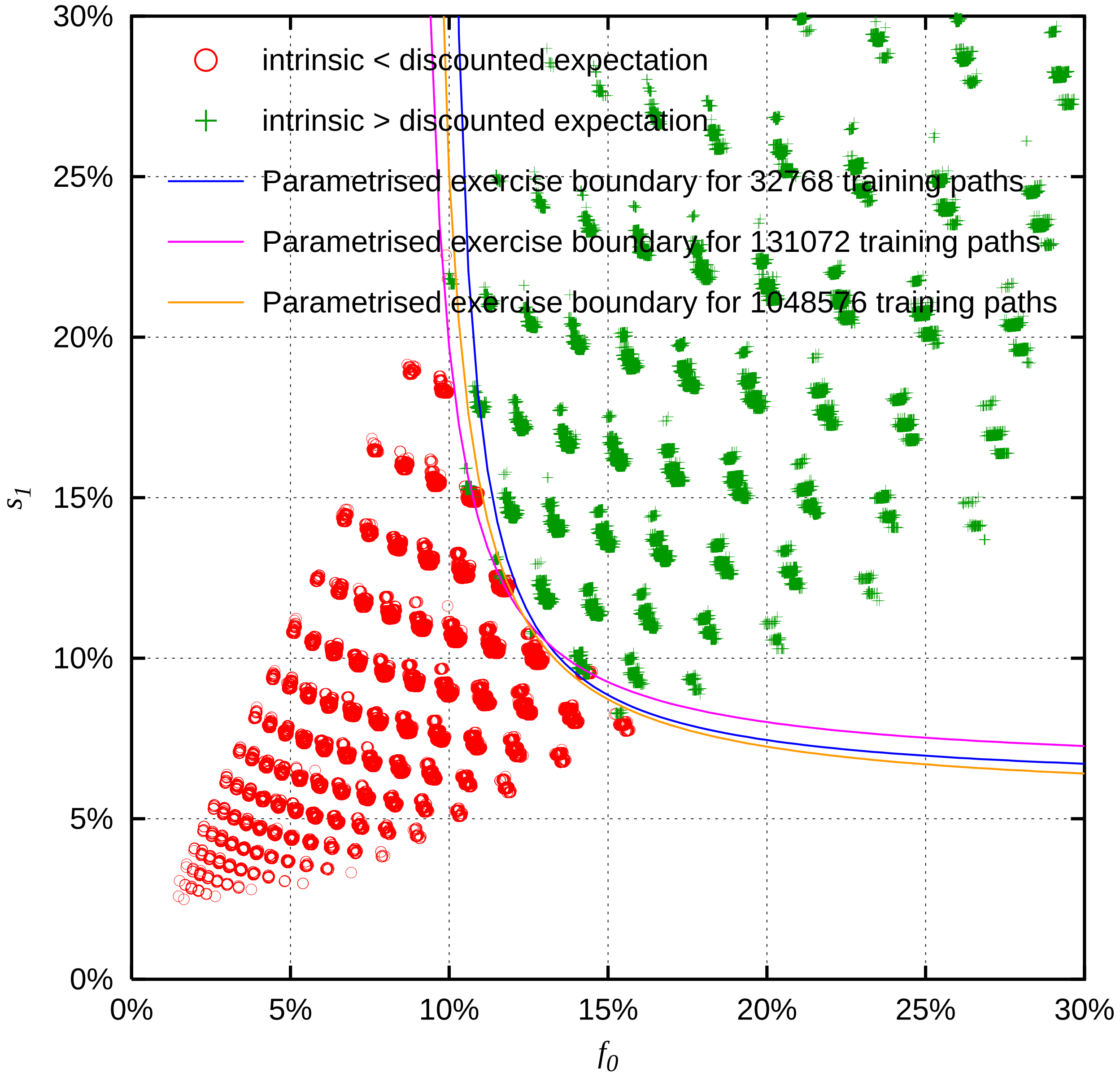
The final value $U_0(\boldsymbol{p}_0)$ gives then an estimate of the value of the Bermudan swaption with a slight upward bias.

Therefore, we finally re-run the simulation with a new set of N_{Sampling} yield curve evolutions using the established exercise strategy parametrisation given by the set of n exercise decision functions $\mathcal{E}_i$.

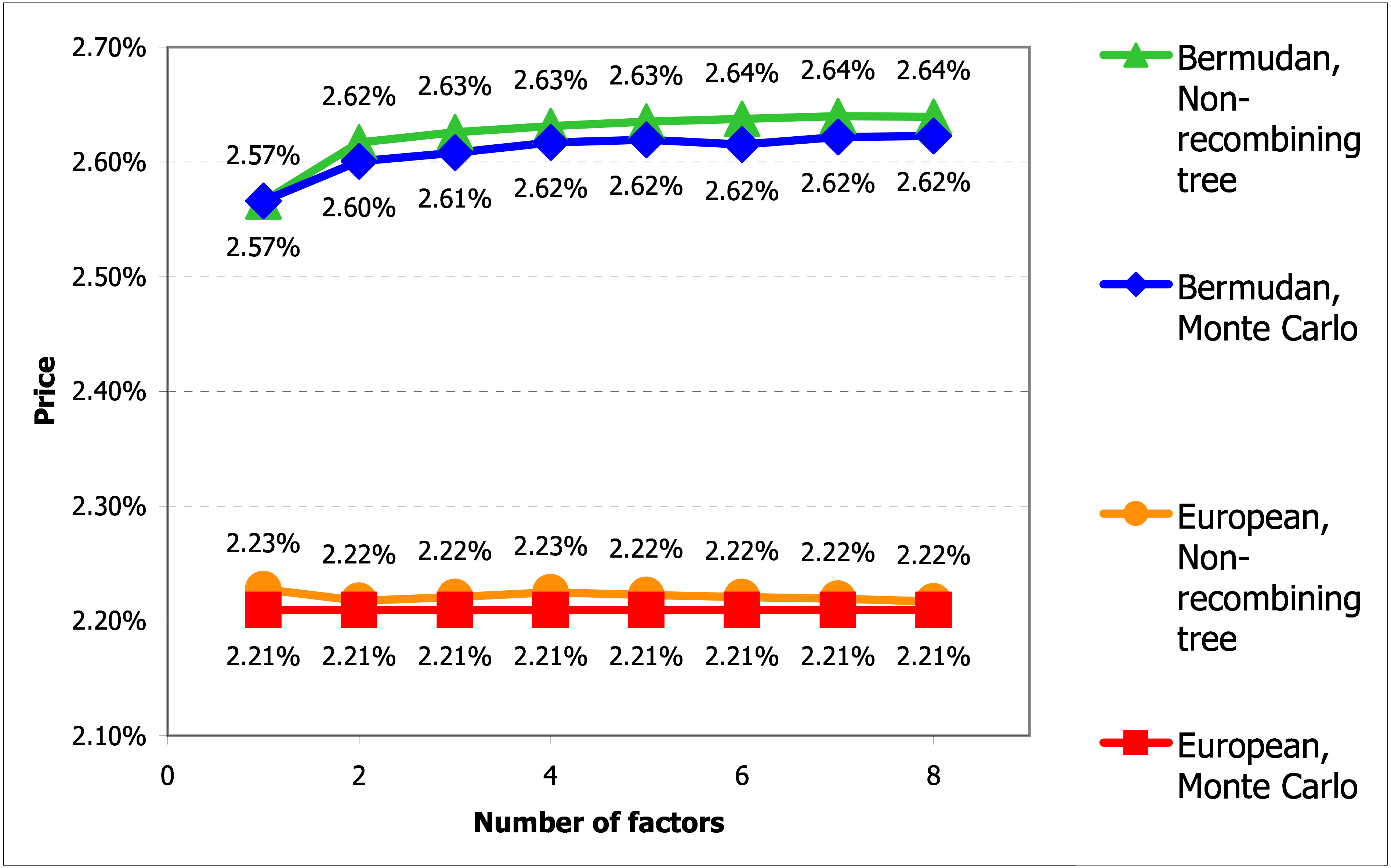
$N_{\text{Sampling}} \simeq 2N_{\text{Training}}$ is typically well sufficient, especially when the driving number generator method was a low-discrepancy sequence.

In practice, we may use a specific choice of coordinates for the specification of the exercise decision function, e.g. (f_i, s_{i+1}). Naturally, since we then have $\mathcal{E}_i = \mathcal{E}_i(f_i, s_{i+1})$, we only store (f_i, s_{i+1}) for each path and each exercise horizon.

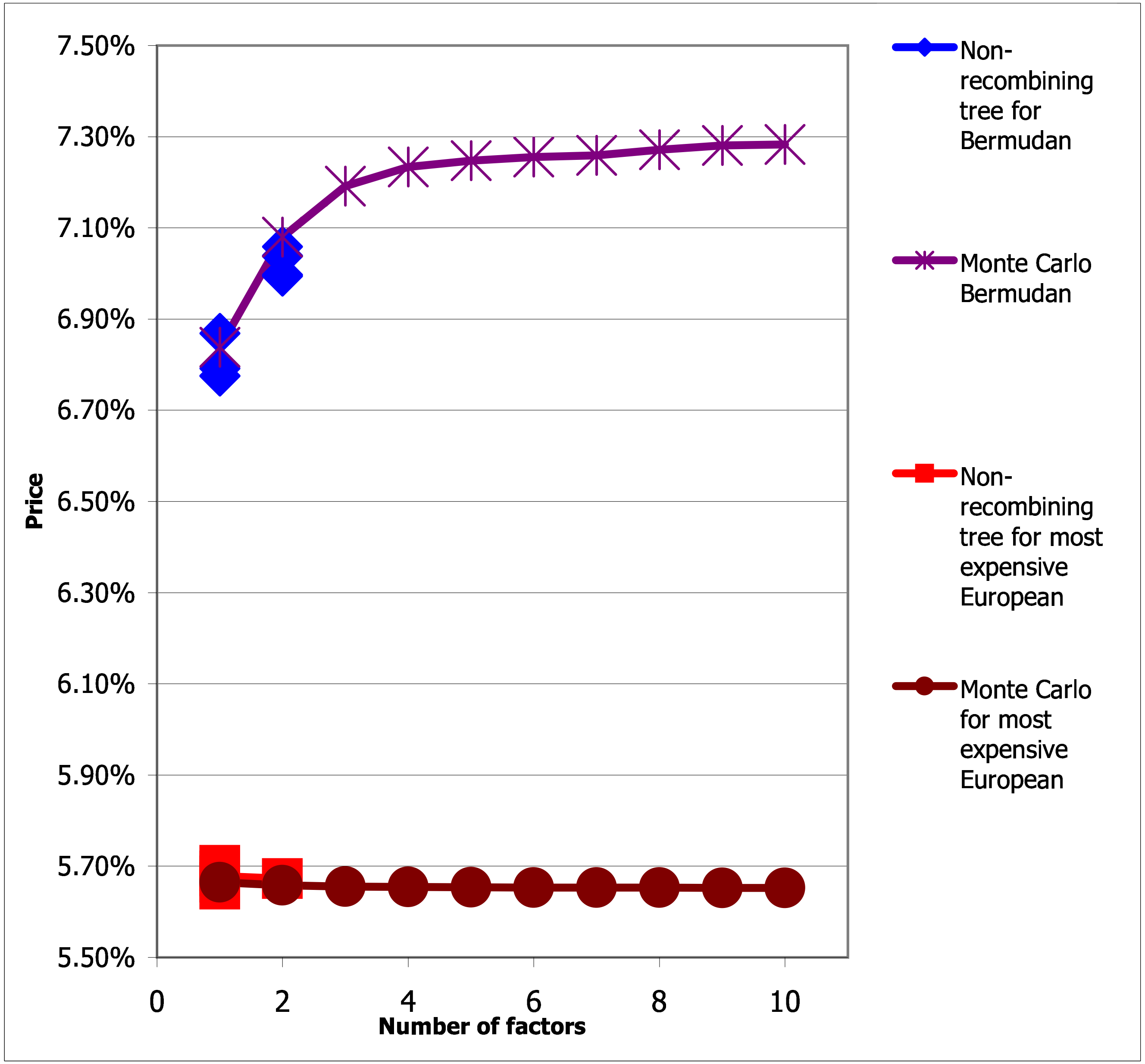
Only storing what is needed for the decision $(\mathcal{E}_i(\boldsymbol{f}_{ji}\,;\,\boldsymbol{p}_i) > 0)$ for each path at each exercise horizon greatly reduces the memory requirements and increases the speed of the algorithm!



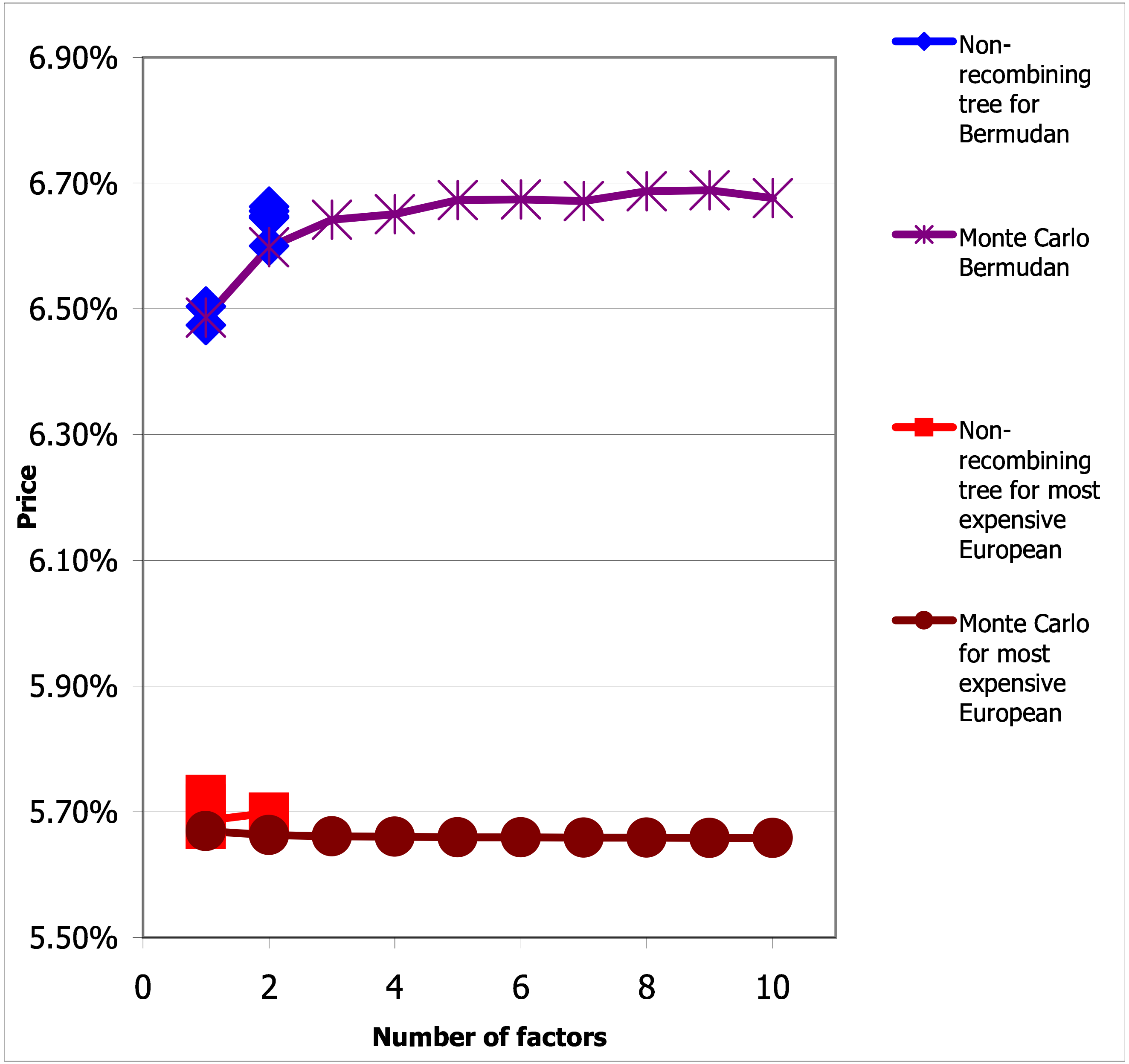
The exercise domain in the f_i-s_{i+1} projection of the evolved yield curve at $t_i = 2$, together with the parametrised exercise boundary resulting from training with different sizes of the training set.



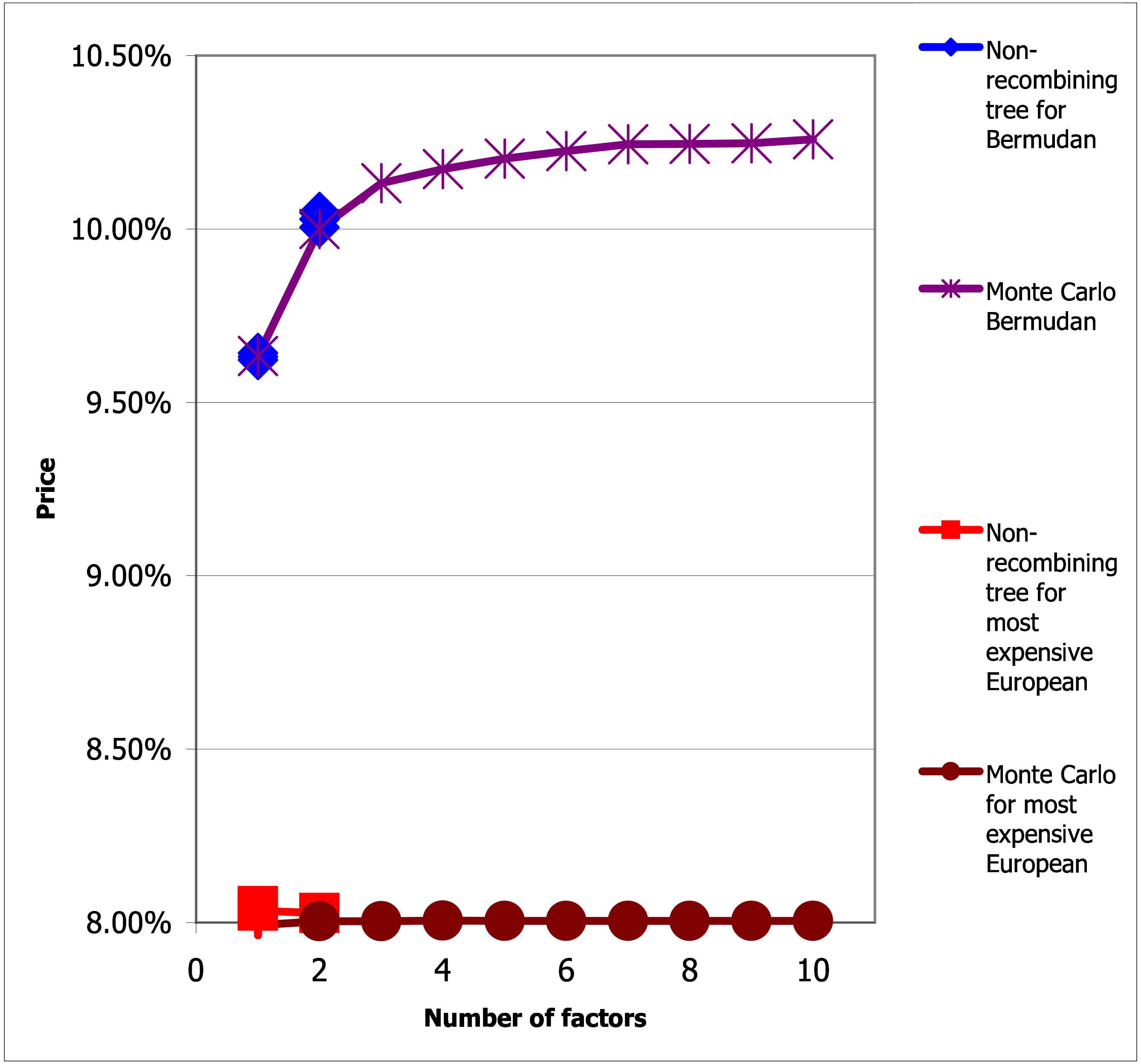
Bermudan swaption prices from the Monte Carlo model in comparison to those obtained from a non-recombining tree model for a 6-non-call-2 semi-annual payer's swaption.



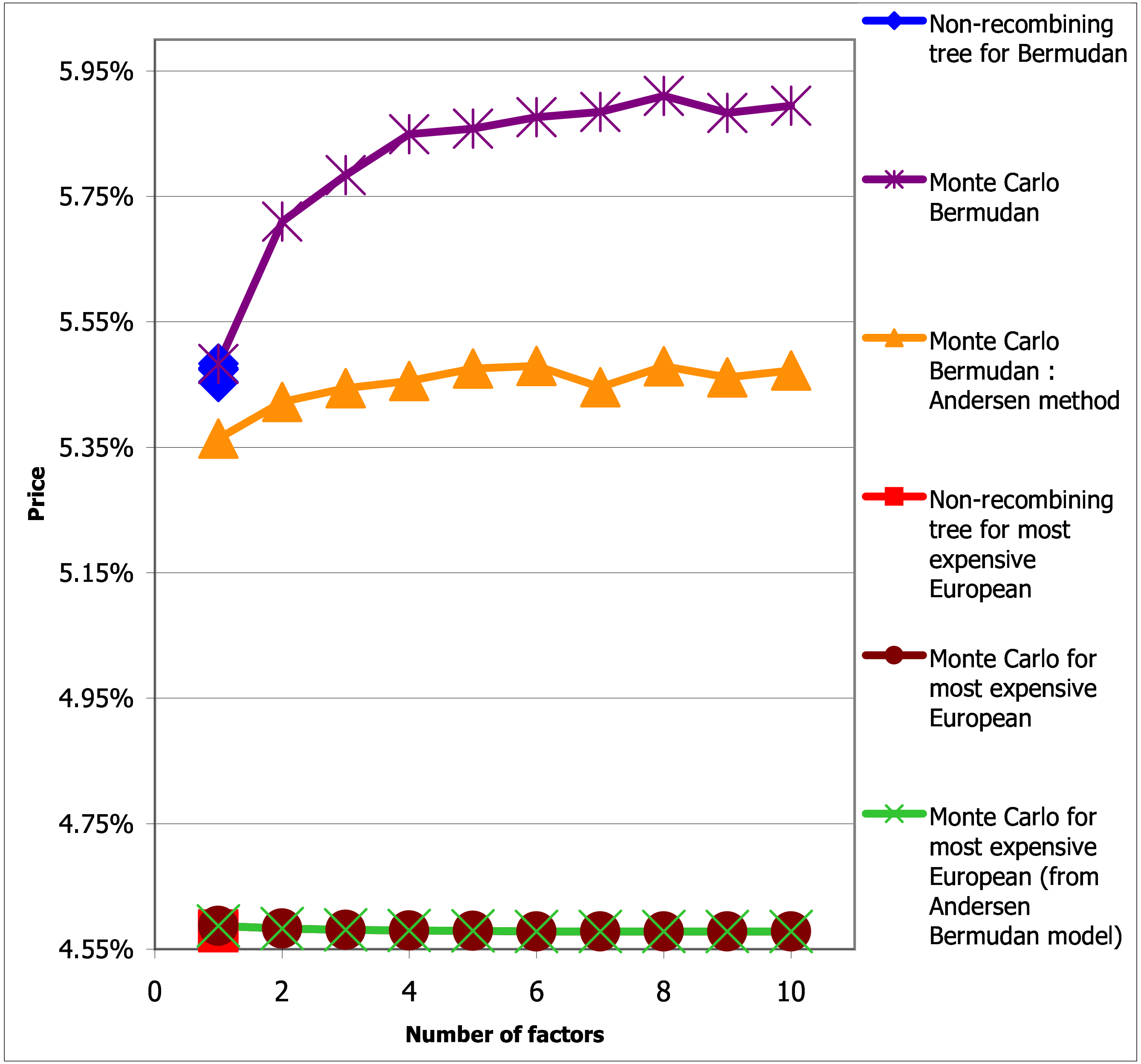
15-non-call-5 annual payer's



15-non-call-5 annual receiver's



15-non-call-5 annual payer's for steeply upwards sloping yield curve



20-non-call-10 semi-annual payer's in comparison to Andersen's method I [And00]

References

[AA00] L. Andersen and J. Andreasen. Volatility Skews and Extensions of the Libor Market Model. *Applied Mathematical Finance*, 7(1):1–32, March 2000.

[Alb72] A. Albert. *Regression and the Moore-Penrose Pseudo-Inverse*. Academic Press, 1972.

[And00] L. Andersen. A simple approach to the pricing of Bermudan swaptions in the multifactor LIBOR market model. *Journal of Computational Finance*, 3(2):5–32, Winter 1999/2000.

[BBW01] S. Benninga, T. Bjork, and Z. Wiener. On the Use of Numeraires in Option Pricing. *SSRN eLibrary*, 2001. ssrn.com/paper=297056.

[BGM97] A. Brace, D. Gatarek, and M. Musiela. The Market Model of Interest Rate Dynamics. *Mathematical Finance*, 7(2):127–155, 1997.

[BR96] M. Baxter and A. Rennie. *Financial Calculus*. Cambridge University Press, 1996.

[Con10] R. Cont, editor. *Encyclopedia of Quantitative Finance*. John Wiley and Sons, 2010. www.wiley.com/go/eqf.

[GKR95] H. Geman, N. Karoui, and J-C. Rochet. Changes of numéraire, changes of probability measure, and option pricing. *Journal of Applied Probability*, 32:443–458, 1995.

[GZ99] P. Glasserman and X. Zhao. Fast greeks by simulation in forward Libor models. *Journal of Computational Finance*, 3(1):5–39, Fall 1999.

[HJM92] D. Heath, R. Jarrow, and A. Morton. Bond pricing and the term structure of interest rates. *Econometrica*, 61(1):77–105, 1992.

[HP81] J. M. Harrison and S. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic processes and their applications*, 11:215–260, 1981.

[JA10] P. Jäckel and L. Andersen. Exercise boundary optimization methods. In Cont [Con10]. www.jaeckel.org/ExerciseBoundaryOptimization.pdf.

[Jäc02] P. Jäckel. *Monte Carlo methods in finance*. John Wiley and Sons, February 2002.

[JR00] P. Jäckel and R. Rebonato. The link between caplet and swaption volatilities in a Brace-Gatarek-Musiela/Jamshidian framework: approximate solutions and empirical evidence. *Journal of Computational Finance*, 6(4):41–59, 2003 (submitted in 2000). www.jaeckel.org/LinkingCapletAndSwaptionVolatilities.pdf.

[Kaw02] A. Kawai. Analytical and Monte Carlo swaption pricing under the forward swap measure. *Journal of Computational Finance*, 6(1):101–111, 2002. www.maths.unsw.edu.au/statistics/files/preprint-2002-12.pdf.

[Kaw03] A. Kawai. A new approximate swaption formula in the LIBOR market model: an asymptotic expansion approach. *Applied Mathematical Finance*, 10(1):49–74, March 2003. www.maths.unsw.edu.au/statistics/files/preprint-2002-13.pdf.

[KS91] I. Karatzas and S. E. Shreve. *Brownian motion and Stochastic Calculus*. Springer, 1991.

[LS98] F. A. Longstaff and E. S. Schwartz. Valuing American options by simulation: A simple least squares approach. Working paper, The Anderson school, UCLA, 1998.

[Mar78] W. Margrabe. The value of an option to exchange one asset for another. *Journal of Finance*, 33:177–186, 1978.

[MSS97] K. R. Miltersen, K. Sandmann, and D. Sondermann. Closed-form solutions for term structure derivatives with lognormal interest rates. *Journal of Finance*, 52(1):409–430, 1997.

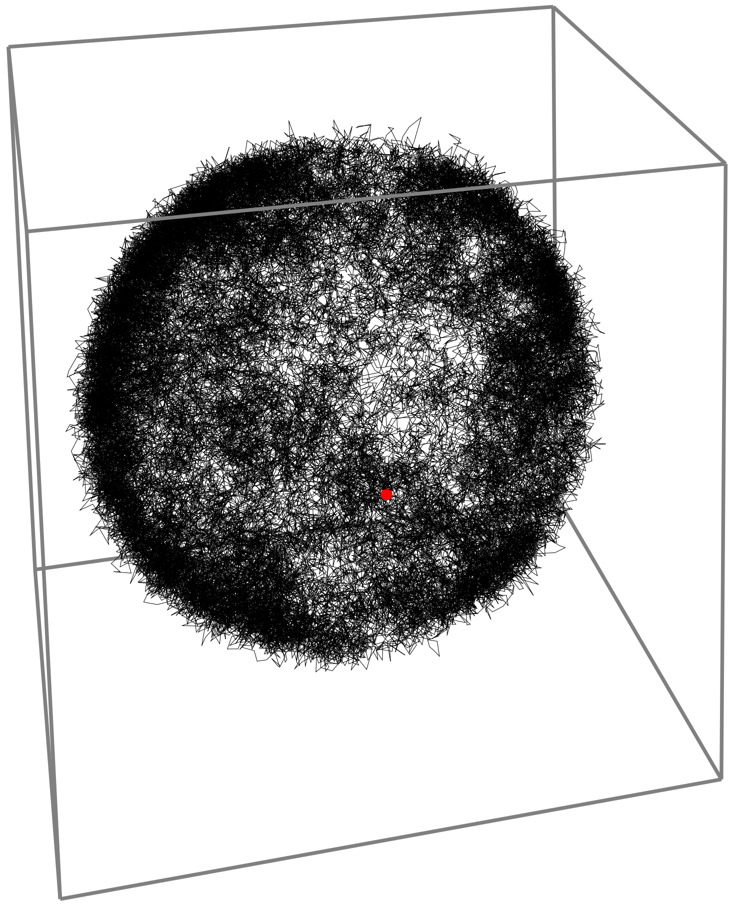
[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992. www.nrbook.com/a/bookcpdf.php.

[Reb98] R. Rebonato. *Interest Rate Option Models*. John Wiley and Sons, 1998.

Certificate in Quantitative Finance

Peter Jäckel*

MONTE CARLO METHODS



*OTC Analytics

Contents

I. Basics

A *Monte Carlo* method is any method that involves a *numerical* approximation by means of *sampling a domain*.

Against popular belief, the term *Monte Carlo method* does not, in any way, imply that any form of randomness is involved.

The most common calculation evaluated by Monte Carlo methods is

$$\int_{\mathcal{D}} g(x)\,\mathrm{d}x \tag{1}$$

over some domain $\mathcal{D}$ for some integrand $g(x)$. Almost always, the integrand $g(x)$ is comprised of two parts: the value function $f(x)$ and the *measure* $\psi(x)$ by

$$g(x) = f(x) \cdot \psi(x)\ . \tag{2}$$

II. The connection to statistics

- A *random experiment* is a process or action subject to non-deterministic uncertainty.

- We call the outcome of a random experiment a *draw* or a *variate* from a distribution.

- A *distribution density* or *probability density function* is a generalised function that assigns *likelihood* or *probability density* to all possible results of a random experiment.

- For our purposes, a *generalised function* can be an ordinary function, or a linear combination of an ordinary function and any finite number of *Dirac densities* $\delta(x - x_0)$.

- The *Dirac density* is the equivalent of assigning a finite probability to a single number on a continuous interval. This means, the Dirac density $\delta(x - x_0)$ is zero everywhere where it is defined, and strictly speaking undefined at x_0. However, its integral is given by the *Heaviside function* $h(x)$, i.e.

$$h(x - x_0) = \int_{x'=-\infty}^{x} \delta(x' - x_0)\,\mathrm{d}x' \;,$$

which is zero for $x < x_0$ and one for $x > x_0$.

- We call the set of all attainable outcomes X of a random experiment the *domain* $\mathcal{D}(X)$ *of the random experiment*. Whenever $\mathcal{D}(X)$ is an ordered set, i.e. when we can decide whether any one element is less than any of the other elements of $\mathcal{D}(X)$, we define the *cumulative probability function* as

$$\Psi(x) \;=\; \int_{x'=\inf(\mathcal{D})}^{x} \psi(x')\,\mathrm{d}x' \;=\; \text{Pr}\left[X < x\right] \;. \tag{3}$$

- All distribution densities are, by definition, normalised, i.e.

$$\int_{\mathcal{D}} \psi(x)\,\mathrm{d}x = 1\ . \tag{4}$$

- The *expected value* of a quantity subject to uncertainty is the probability weighted average. Our notation for the expected value of a quantity f with respect to a probability density ψ is

$$\mathsf{E}_\psi[f] \quad : \quad \text{expected value of } f \text{ with respect to } \psi. \tag{5}$$

- An alternative notation frequently encountered for $\mathsf{E}_\psi[f]$ is

$$\langle f \rangle\ , \tag{6}$$

- The *variance* of a quantity subject to uncertainty is defined as

$$\mathsf{E}_\psi\left[f^2\right] - \left(\mathsf{E}_\psi[f]\right)^2 \;=\; \left\langle f^2 \right\rangle - \langle f \rangle^2\ . \tag{7}$$

- Very often, there is no ambiguity about the relevant distribution. In this case we may just write $\mathsf{E}[f]$. Alternatively, there may be a parametric description of the distribution of f. This means that f is actually a function of some other uncertain variable x. Given the distribution density of x, say $\psi(x)$, we would denote the expectation of f as $\mathsf{E}_{\psi(x)}[f(x)]$. This is just to say

$$\mathsf{E}_{\psi(x)}[f(x)] = \int_{-\infty}^{\infty} f(x)\psi(x)\,\mathrm{d}x\ . \tag{8}$$

- The connection between Monte Carlo methods and the statistical quantity known as the *expectation* is simply the fact that the target of a Monte Carlo calculation is to compute

$$\int f(x)\psi(x)\,\mathrm{d}x\ . \tag{9}$$

which, in a statistical sense, is, of course, equal to $\mathsf{E}_{\psi(x)}[f(x)]$.

III. The Feynman-Kac theorem

R. Feynman and M. Kac [Fey48, Kac51]:

Given the set of stochastic processes

$$\mathrm{d}X_i = b_i \mathrm{d}t + \sum_{j=1}^{n} a_{ij} \mathrm{d}W_j \quad \text{for} \quad i = 1..n \; , \tag{10}$$

with formal solution

$$X_i(T) = X_i(t) + \int_t^T b_i \mathrm{d}t + \int_t^T \sum_{j=1}^{n} a_{ij} \mathrm{d}W_j \; , \tag{11}$$

any function $V(t, \boldsymbol{X})$ with boundary conditions

$$V(T, \boldsymbol{X}) = f(\boldsymbol{X}) \tag{12}$$

that satisfies the partial differential equation

$$\frac{\partial V}{\partial t} + \sum_{i=1}^{n} b_i \frac{\partial V}{\partial X_i} + \frac{1}{2} \sum_{i,j=1}^{n} c_{ij} \frac{\partial^2 V}{\partial X_i \partial X_j} + g = rV \tag{13}$$

with

$$c_{ij} := \sum_{k=1}^{n} a_{ik} a_{jk} \tag{14}$$

can be represented as the expectation

$$V(t, \boldsymbol{X}) = \mathsf{E}\left[f(\boldsymbol{X}_T) \mathrm{e}^{-\int_t^T r \,\mathrm{d}u} + \int_t^T g\, \mathrm{e}^{-\int_t^s r \,\mathrm{d}u} \mathrm{d}s \right] . \tag{15}$$

Hereby, all of the coefficients a_{ij}, b_i, r, and g can be functions both of time t and the state vector $\boldsymbol{X}(t)$.

As with most mathematical theorems, there is a whole host of additional conditions for good behaviour of all the coefficients and functions involved and the reader is referred to, e.g., Karatzas and Shreve [KS91] (page 366).

IV. The basic Monte Carlo algorithm

The easiest form of *Monte Carlo integration* of integrals like equation (9) can be summarised as follows.

- Establish a procedure of drawing variates $\boldsymbol{x}$ from the target distribution density $\psi(\boldsymbol{x})$.

- Set up a running sum variable

```
double runingSum=0;
```

  and a counter variable

```
unsigned long i=0;
```

- Draw a variate vector $\boldsymbol{x}_i$ and evaluate $f_i := f(\boldsymbol{x}_i)$.
- Add the computed function value to `runningSum`, i.e.

```
runningSum += f;
```

- Increment i, i.e.

```
++i;
```

- Set the running average to

```
const double runningAverage = runningSum/i;
```

  This gives us the *Monte Carlo estimator*

$$\hat{v}_N \; := \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}_i) \quad . \tag{16}$$

- Keep iterating until either the required number of iterations has been carried out, or a specific error estimate has decreased below a predetermined threshold.

A practitioner's list of key points to remember:-

- Do not trust any single Monte Carlo simulation result unless you feel you have good reasons to believe that it is sufficiently converged.

- Do not trust the fact that an interval of one standard error around your result is still within an acceptable range for your purposes - one in three simulations should be outside one standard error interval!

- Do not repeat a simulation using a different “seed” - you have no idea how much overlap you may have!

- If you wish to repeat a simulation, do so using a different number generation algorithm.

- The most intuitive control over convergence is given by a convergence diagram - I recommend plotting the running average as a function of the number of iterations N with plot points for N being powers of 2.

V. The standard error

A statistical error estimate for a quantity subject to probabilistic uncertainty is the *standard deviation*

$$\sigma_a = \left\langle a^2 \right\rangle - \left\langle a \right\rangle^2 \ . \tag{17}$$

Note: the standard deviation is strictly only applicable when the distribution of a itself is normal. It can still be used as a vague measure of uncertainty, **if and only if**

- the distribution of a is unimodal, i.e. its probability density function has a single local maximum which must be its global maximum;
- the probability density function of a is concave only in a singly connected domain near its maximum and non-concave everywhere else.

In general, we don't actually know the variance σ_a^2 of the random variate a whose expectation we are trying to estimate. However, by virtue of the combination of the central limit and the continuous mapping theorem, we can use the variance of the simulation instead as an estimate for σ_a^2:

$$\hat{\sigma}_a^{(N)} = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N} a_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} a_i\right)^2} \tag{18}$$

This leads us to the definition of the *standard error*:

$$\epsilon_a^{(N)} = \frac{\hat{\sigma}_a^{(N)}}{\sqrt{N}} \tag{19}$$

Important:

- Every single Monte Carlo simulation result for a normally distributed variate a has a probability of approximately **one third** of being more than one standard error away from the unknown true expectation.
- In practice, almost nothing we are ever interested in is ultimately normally distributed.
- The less similarities the distribution of the target quantity has with the normal (Gaussian) distribution, the less meaning has the standard error figure.

Examples where the target distribution is distinctly non-normal:

– Correlation estimator

$$\hat{\rho}_{ab} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} a_i b_i\right) - \left(\frac{1}{N}\sum_{i=1}^{N} a_i\right)\left(\frac{1}{N}\sum_{i=1}^{N} b_i\right)}{\sqrt{\left[\left(\frac{1}{N}\sum_{i=1}^{N} a_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} a_i\right)^2\right]\left[\left(\frac{1}{N}\sum_{i=1}^{N} b_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} b_i\right)^2\right]}} \tag{20}$$

– Option payoff

Despite all these shortcomings, without further knowledge of the task at hand, the *standard error* is still the best generic estimator of convergence available, irrespective of whether the simulation algorithm was based on statistical methods or not.

VI. Uniform variates

Standard uniform variates are supposed to be equally distributed on $[0,1]$, $[0,1)$, or $(0,1)$ (depending on the algorithm used). For applications in finance, *without exception*, we need to ensure that the number generator is confined to $(0,1)$. There are two main variations of number generation algorithms:-

- Pseudo-random number generators. Can be used as *number pipelines*.

  These generators have no built-in mechanism for sampling efficiency. They typically lead to *clusters* and *gaps* which may or may not be desirable.

  $\Rightarrow$ emulate randomness.

- Low-discrepancy number generators (sometimes also, *misleadingly*, referred to as quasi-random number generators).

  These algorithms are designed to take advantage of our knowledge of the dimensionality and ordering of importance of the given simulation problem at hand.

  $\Rightarrow$ holistic sampling scheme aiming for as little randomness as possible.

Pseudo-random numbers

- Uniform coverage is ensured by virtue of *serial decorrelation*.

- The most common simple algorithms are *congruential generators*

$$m_{n+1} = (a\,m_n + c) \mod M\ . \tag{21}$$

for the integer m_n. Real-valued variates on $(0,1)$ are then produced from m_n by setting

$$x_n = (m_n + 1)/(M + 1)\ . \tag{22}$$

- There are many number generators: see the reviews in [PTVF92, SB02, L'E01]. To name but one reliable one: the Mersenne Twister [MN97].

- Number generators are highly sensitive to tampering - don't do it! Any changes to an existing number generator without a deep understanding of the involved number theory is practically guaranteed to destroy the desired features!

Low-discrepancy numbers

- Uniform coverage is ensured by *selective placement* within one dimension and *incommensurate sampling* across dimensions.

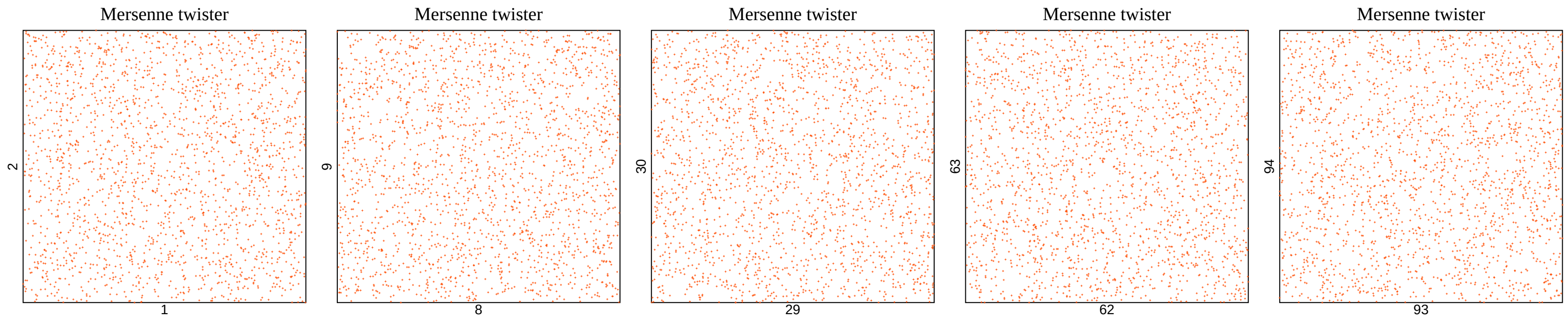
- The mathematics can be very involved, though the resulting run-time algorithm is usually extremely efficient.

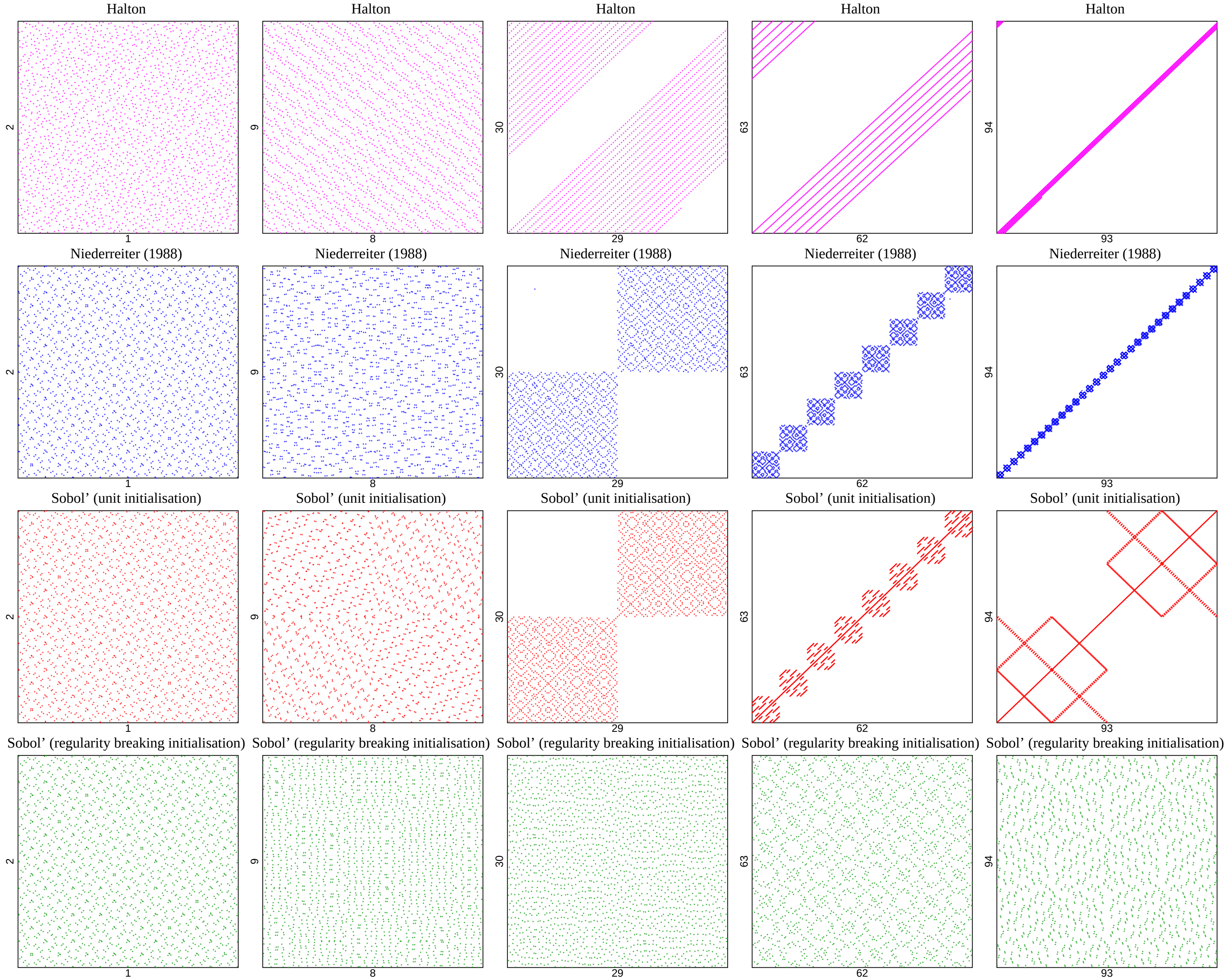
- Monte Carlo simulations using low-discrepancy numbers are, in theory, expected to converge as
$$\sim c(d)\frac{(\ln N)^d}{N}$$
whereas the convergence using random numbers is, expected to be
$$\sim \frac{1}{\sqrt{N}}\ .$$

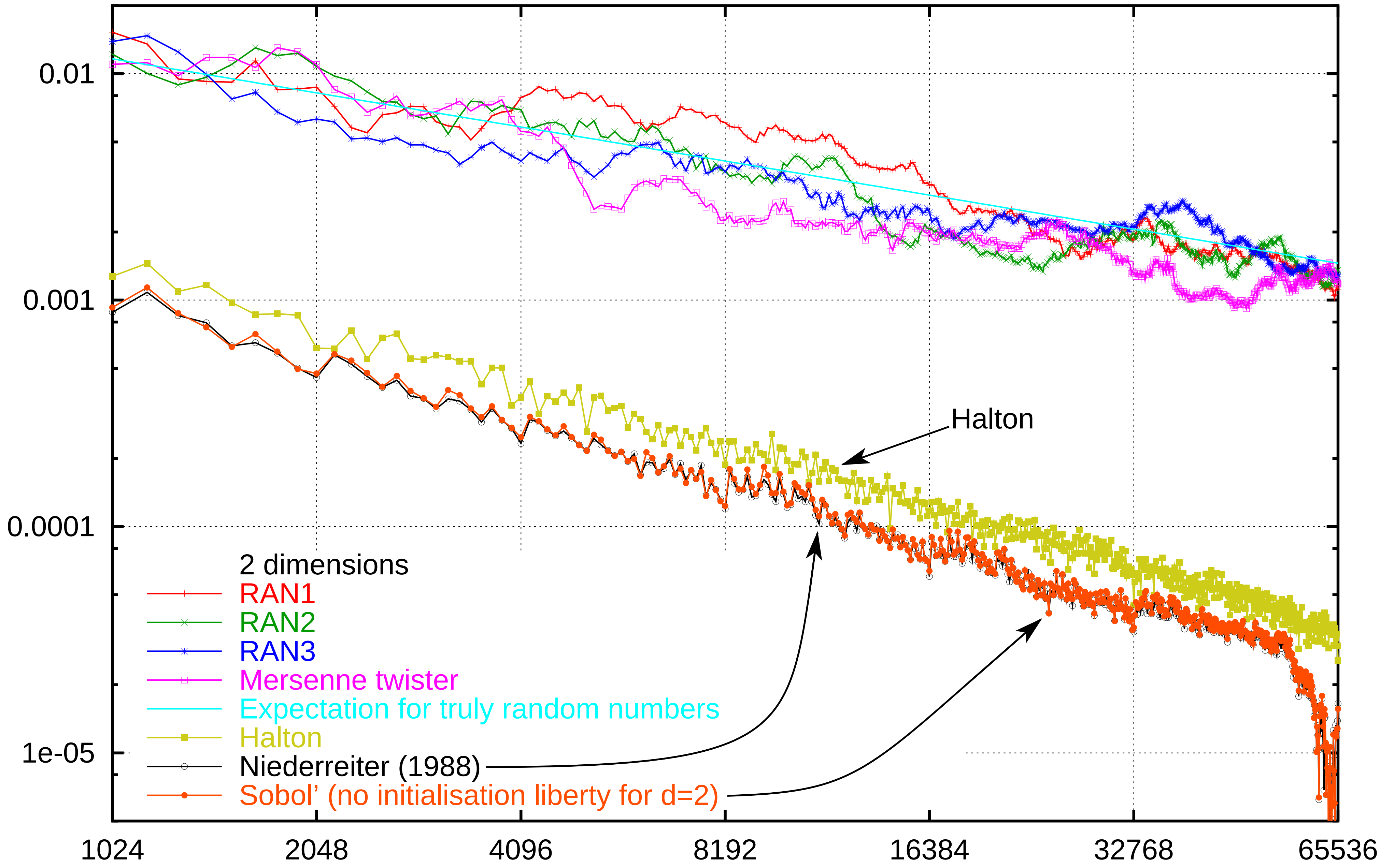
- The problem with low-discrepancy numbers is that $c(d)$ can explode geometrically with d.

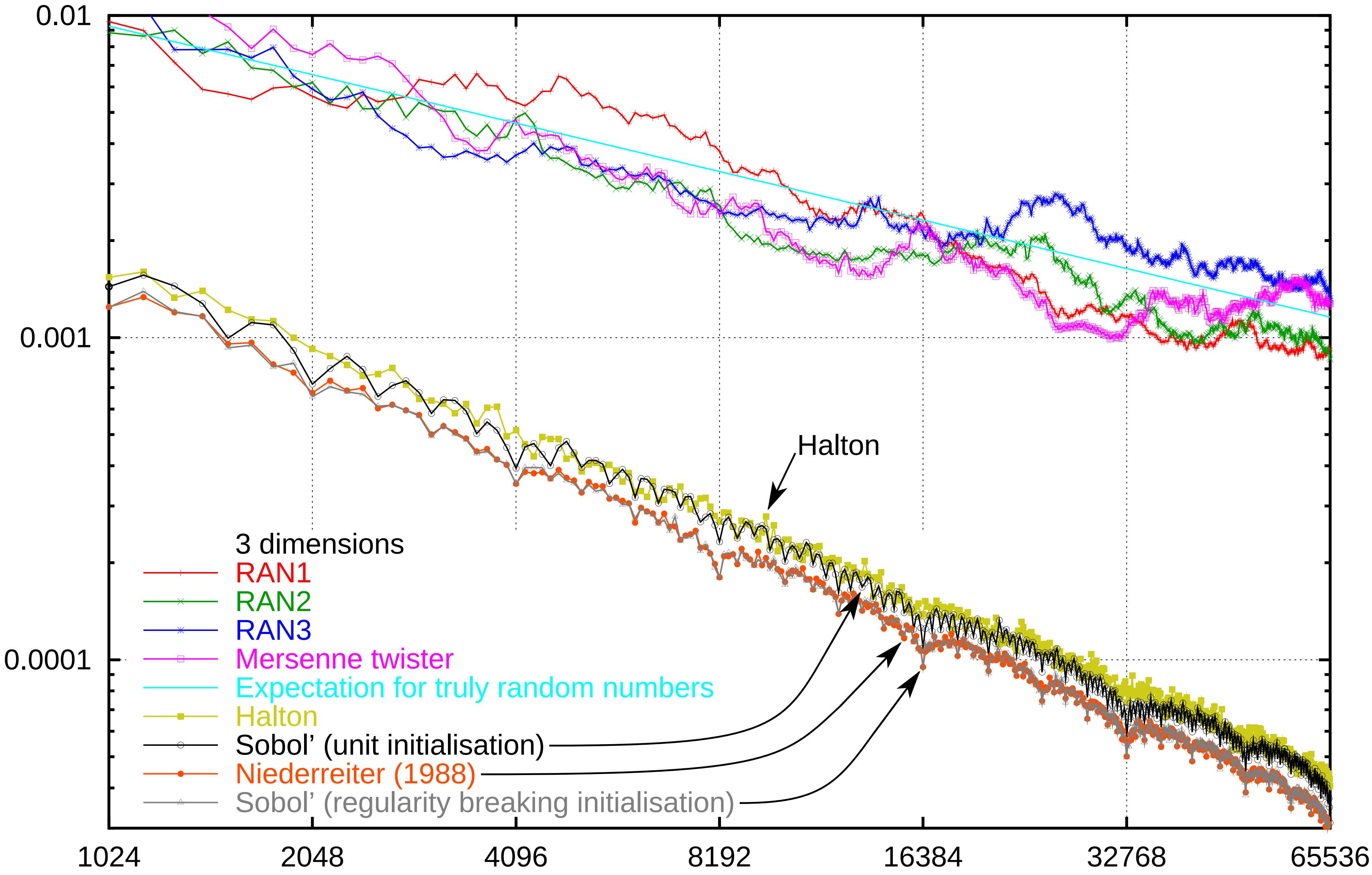
- However, it has been shown that there are low-discrepancy algorithms that do not suffer the dimensionality breakdown, namely Sobol' numbers with appropriate initialisation [Sob76], and Niederreiter-Xing numbers [NX95, NX96].

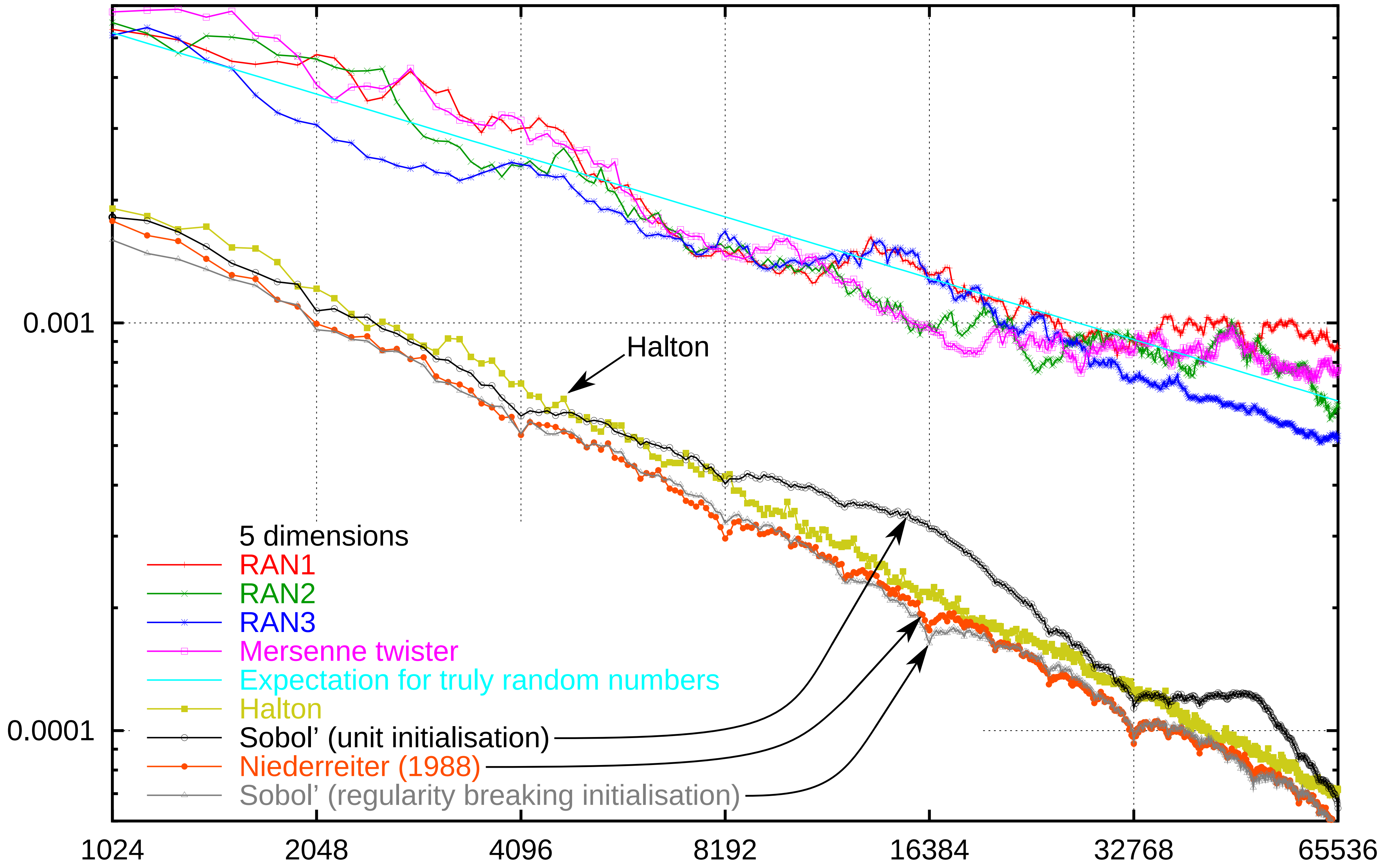
- Two methods are commonly employed to test low-discrepancy numbers: numerical evaluation of a measure for non-uniformity (i.e. discrepancy), and pairwise projections:-

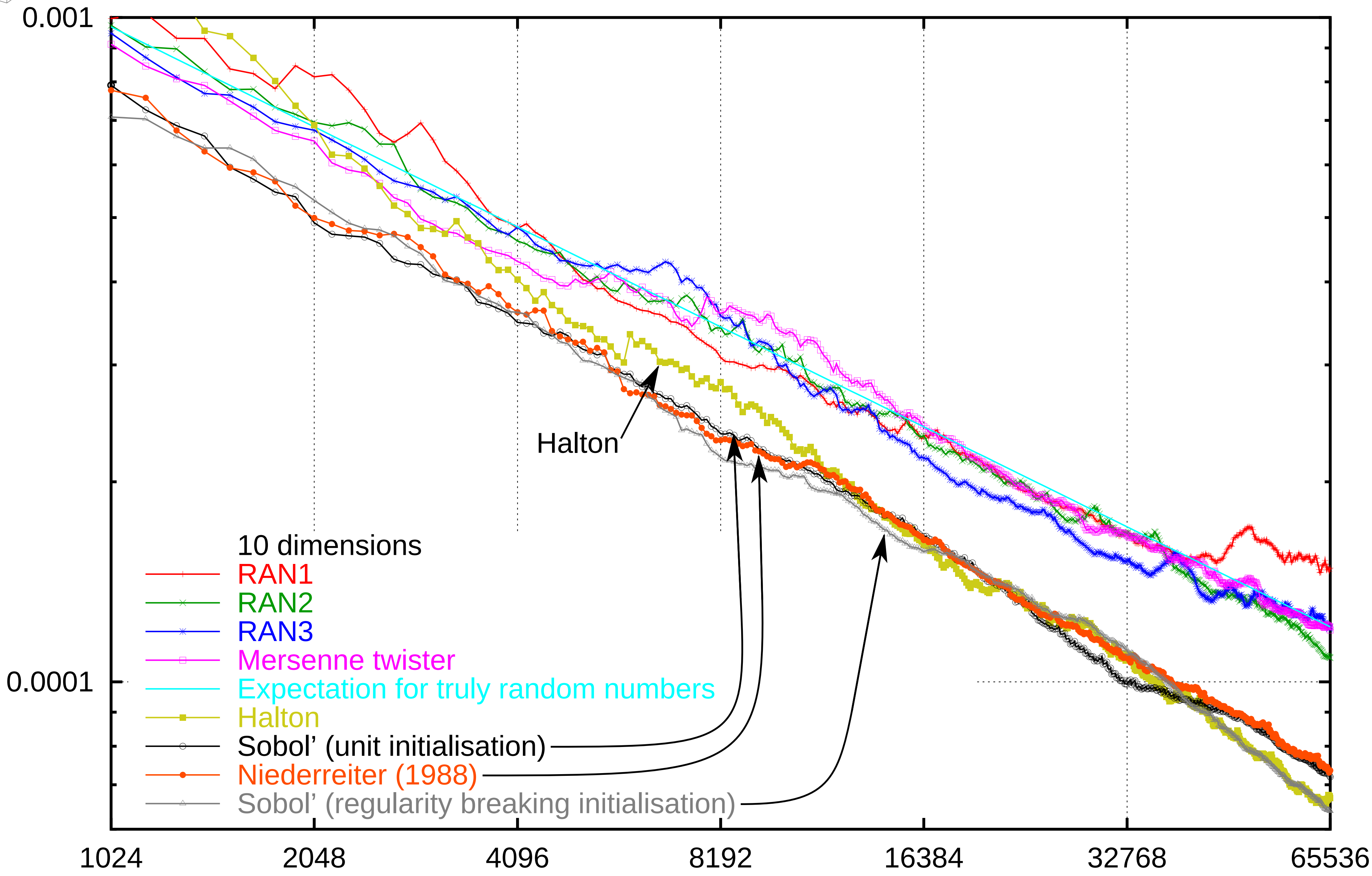


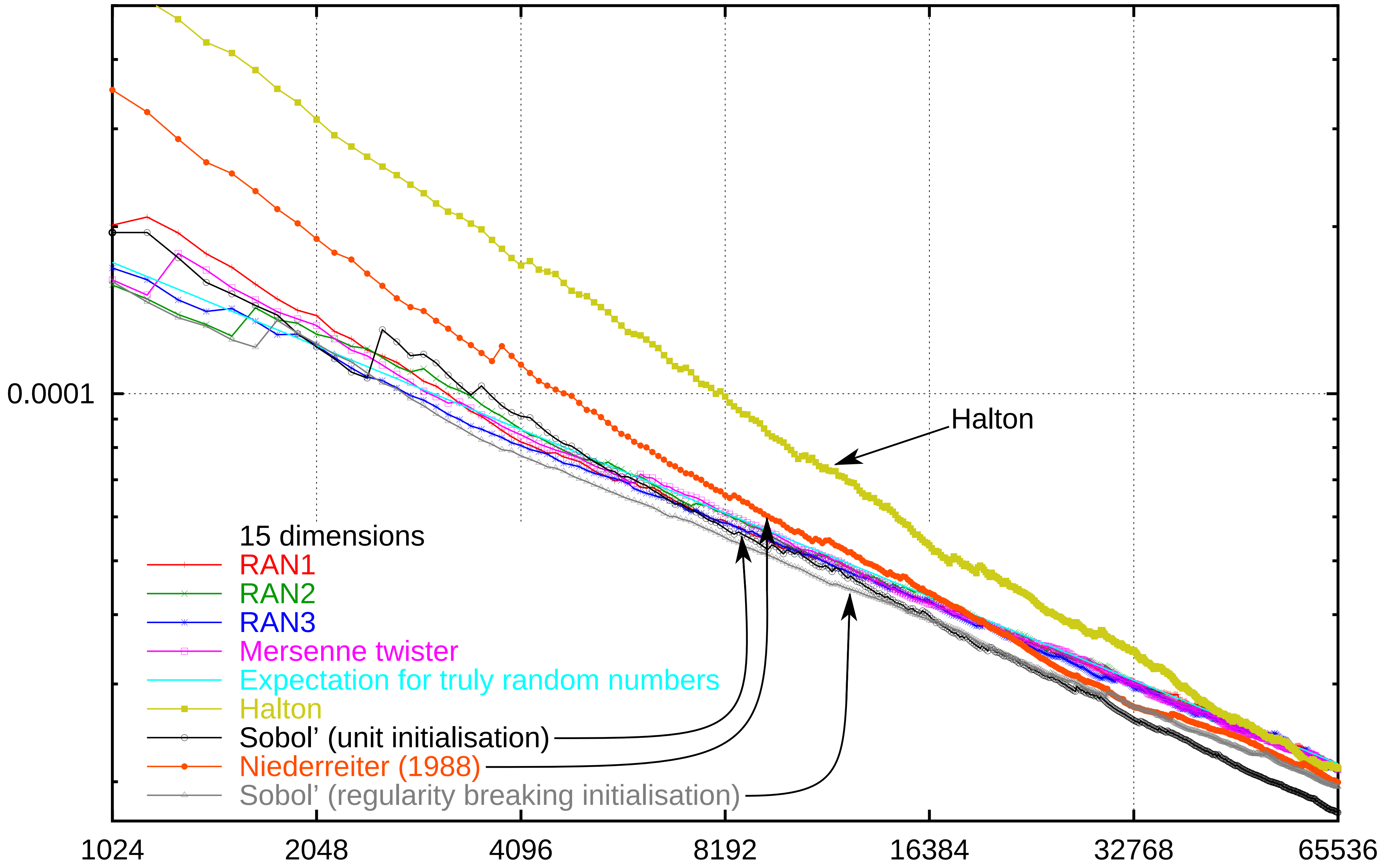

Halton
Halton
Halton
Halton
Halton
2
1
9
8
30
29
63
62
94
93
Niederreiter (1988)
Niederreiter (1988)
Niederreiter (1988)
Niederreiter (1988)
Niederreiter (1988)
2
1
9
8
30
29
63
62
94
93
Sobol' (unit initialisation)
Sobol' (unit initialisation)
Sobol' (unit initialisation)
Sobol' (unit initialisation)
Sobol' (unit initialisation)
2
1
9
8
30
29
63
62
94
93
Sobol' (regularity breaking initialisation)
Sobol' (regularity breaking initialisation)
Sobol' (regularity breaking initialisation)
Sobol' (regularity breaking initialisation)
Sobol' (regularity breaking initialisation)
2
1
9
8
30
29
63
62
94
93

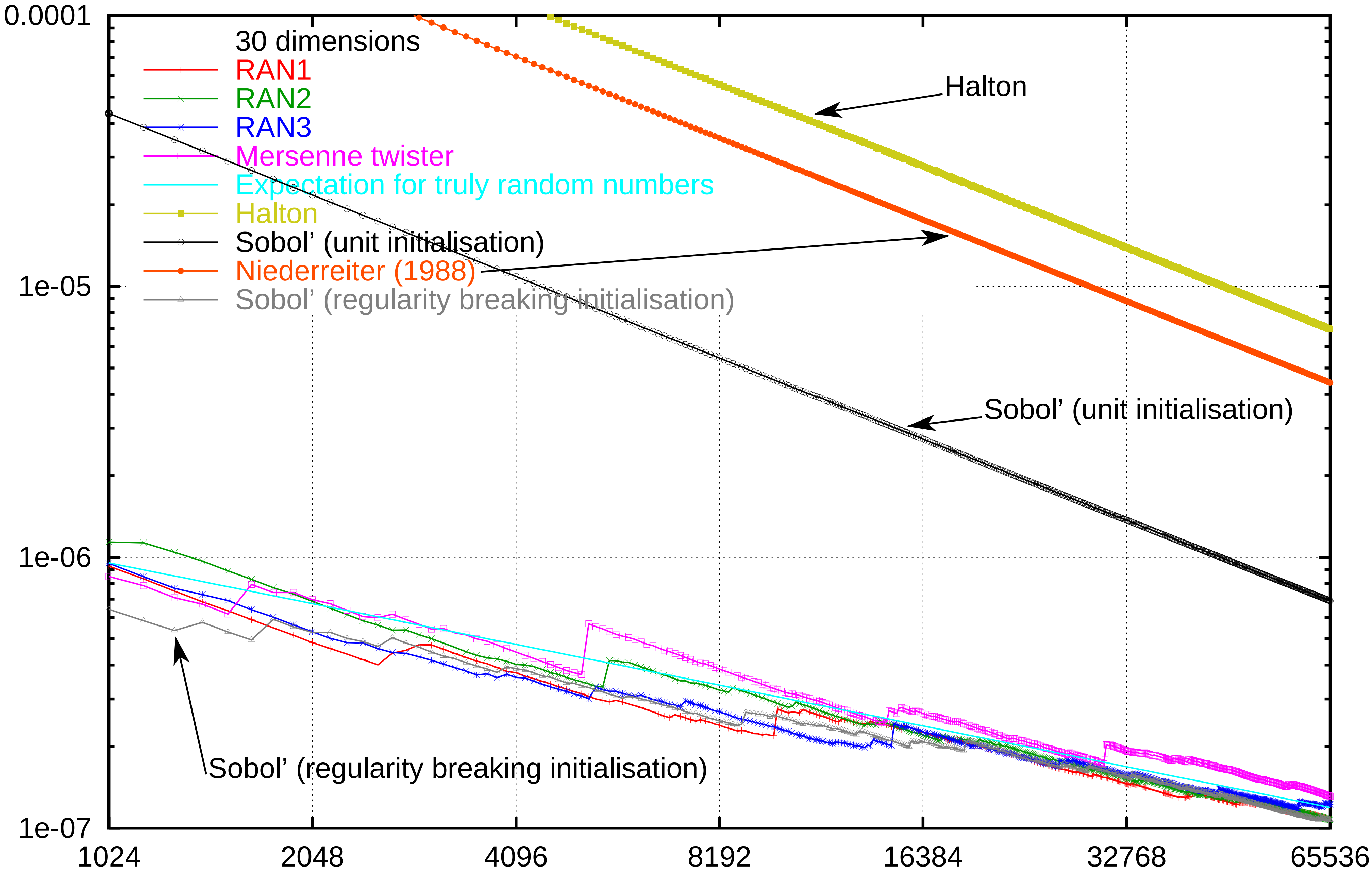

0.01
0.001
0.0001
1e-05
1024
2048
4096
8192
16384
32768
65536
Halton
2 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Niederreiter (1988)
Sobol' (no initialisation liberty for d=2)

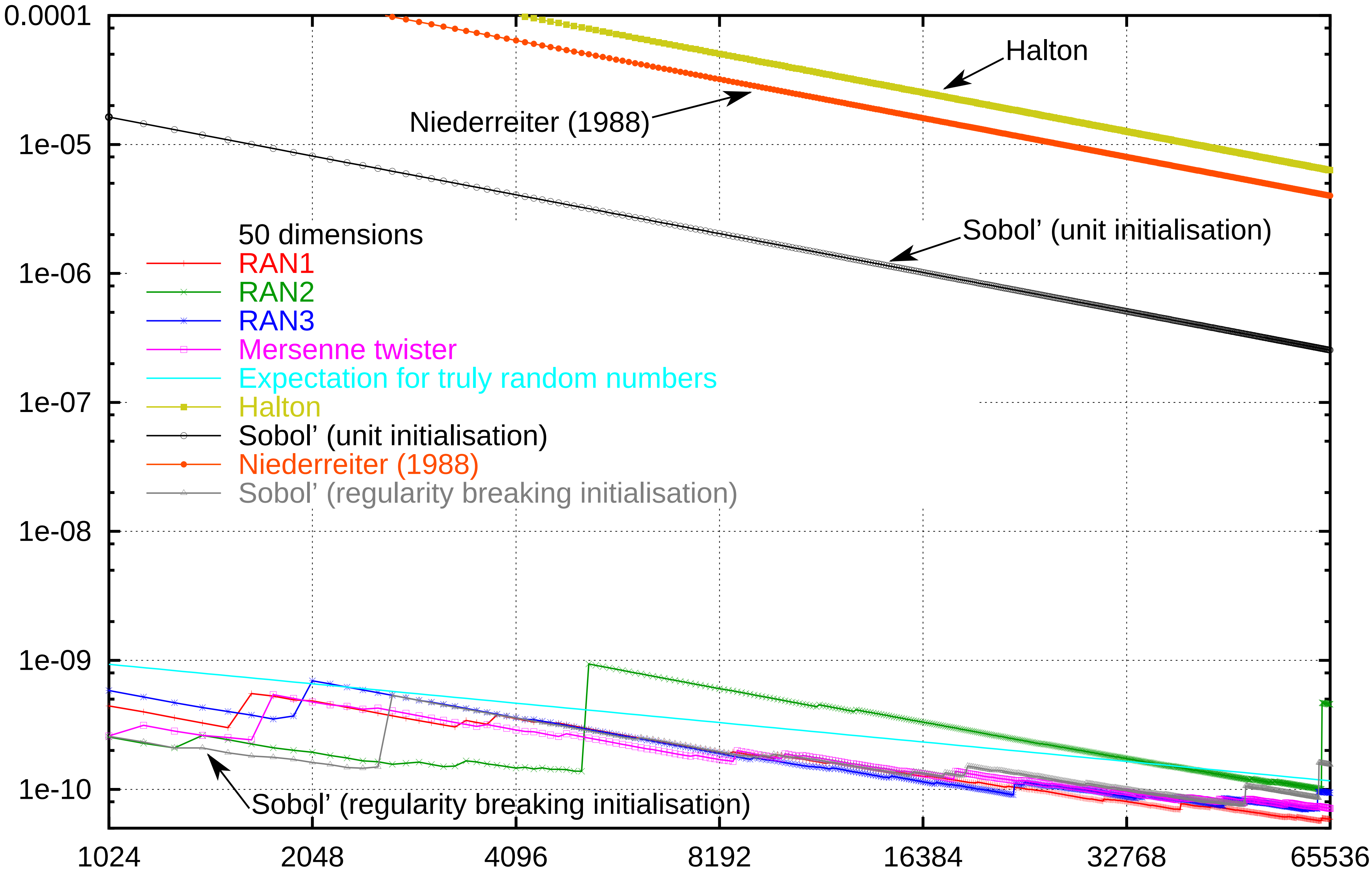

0.01
0.001
0.0001
1024
2048
4096
8192
16384
32768
65536
Halton
3 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)

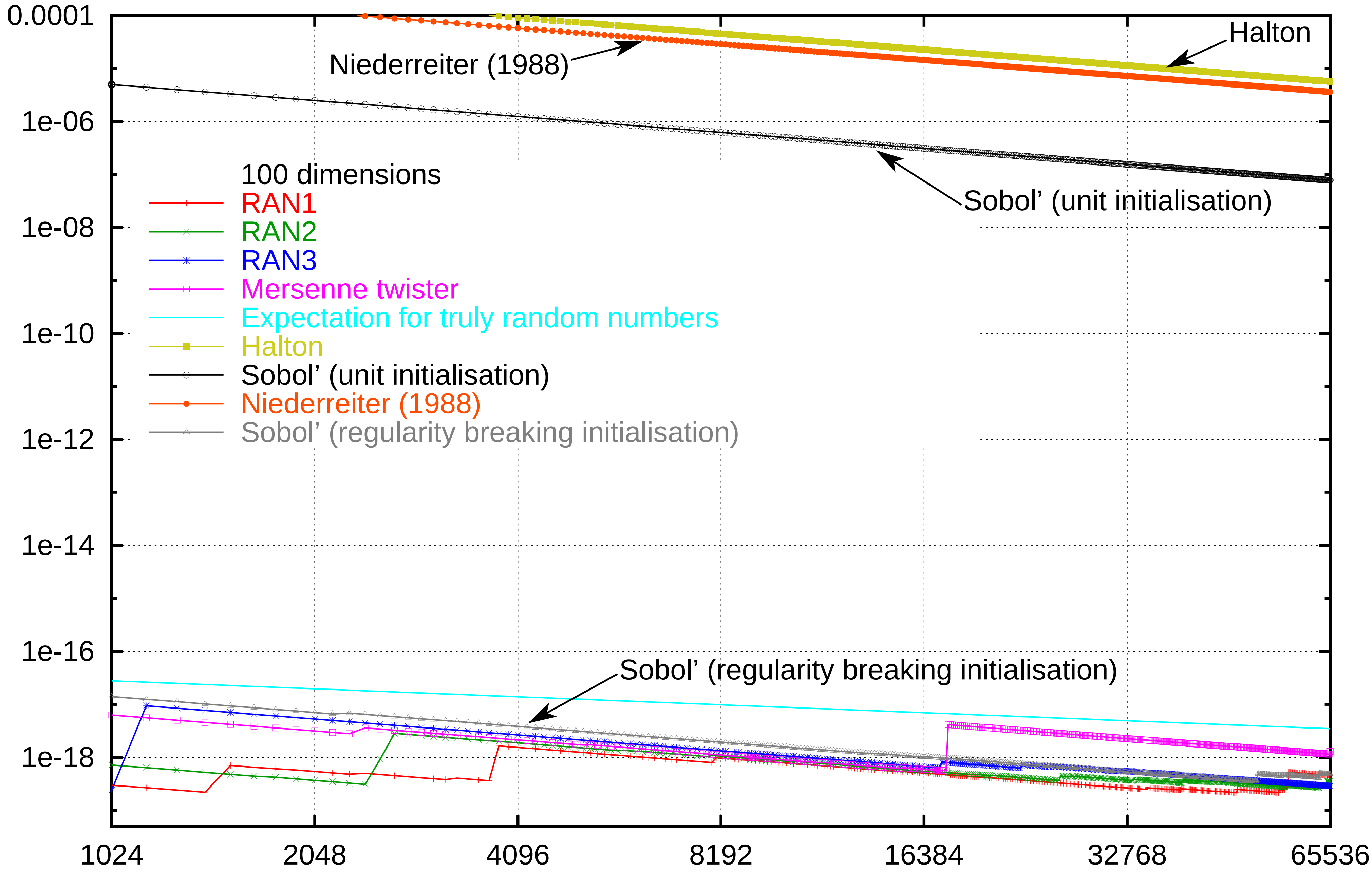

0.001
0.0001
1024
2048
4096
8192
16384
32768
65536
Halton
5 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)


0.001
0.0001
1024
2048
4096
8192
16384
32768
65536
Halton
10 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)


0.0001
1024
2048
4096
8192
16384
32768
65536
Halton
15 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)


0.0001
1e-05
1e-06
1e-07
1024
2048
4096
8192
16384
32768
65536
30 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)
Halton
Sobol' (unit initialisation)
Sobol' (regularity breaking initialisation)


0.0001
1e-05
1e-06
1e-07
1e-08
1e-09
1e-10
1024
2048
4096
8192
16384
32768
65536
Halton
Niederreiter (1988)
Sobol' (unit initialisation)
50 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)
Sobol' (regularity breaking initialisation)


0.0001
1e-06
1e-08
1e-10
1e-12
1e-14
1e-16
1e-18
1024
2048
4096
8192
16384
32768
65536
Halton
Niederreiter (1988)
Sobol' (unit initialisation)
Sobol' (regularity breaking initialisation)
100 dimensions
RAN1
RAN2
RAN3
Mersenne twister
Expectation for truly random numbers
Halton
Sobol' (unit initialisation)
Niederreiter (1988)
Sobol' (regularity breaking initialisation)

VII. Non-uniform variates

Inverting the cumulative distribution function

Given a uniform variate $u \in (0,1)$, the distribution density $\psi(x)$, its integral $\Psi(x)$ (i.e. the cumulative distribution function), and the inverse of the cumulative distribution function $\Psi^{-1}(u)$, the variate $x := \Psi^{-1}(u)$ is distributed $\sim \psi(x)$.

Whenever the cumulative distribution function can be accurately inverted at moderate computational cost, this is the preferred method of choice.

When an efficient and accurate approximation for the inverse of $\Psi(x)$ is not available but $\Psi(x)$ is, an efficient inverse lookup table can be set up as follows:-

1. Choose the number of interpolation points to be used.

2. Populate a vector of values x_i such that the whole domain of x is reasonably equally covered.

3. Compute the associated cumulative probability values $\Psi(x_i)$ and store as elements u_i in a second vector.

4. Build an interpolation object of at least second order interpolation accuracy using the vector $\boldsymbol{u}$ as abscissa, and the vector $\boldsymbol{x}$ as ordinate. Simple polynomials of third or higher order are usually already sufficient but the more accurate the better [Kva00]. Ensure that your interpolation object selects the interpolation bracket given any lookup variate u using a binary nesting algorithm.

5. The resulting interpolation object now represents a numerical instantiation of the inverse cumulative function $\Psi^{-1}(u)$

The use of such an interpolation object for the transformation of uniform $(0, 1)$ variates to a target distribution $\psi(x)$ is highly efficient and completely generic.

Ensure that $u \in (0, 1)$, i.e. that $u \neq 0$ and $u \neq 1$ since at least one of the two end points would for most distributions of interest result in `NaN` (also known as `#Num`).

Using a sampler density

Assume that the target density, $\psi(x)$, is much harder to match directly than a very similar density, $\widetilde{\psi}(x)$, and that $\widetilde{\psi}(x)$ is non-zero wherever $\psi(x)$ is non-zero.

Since

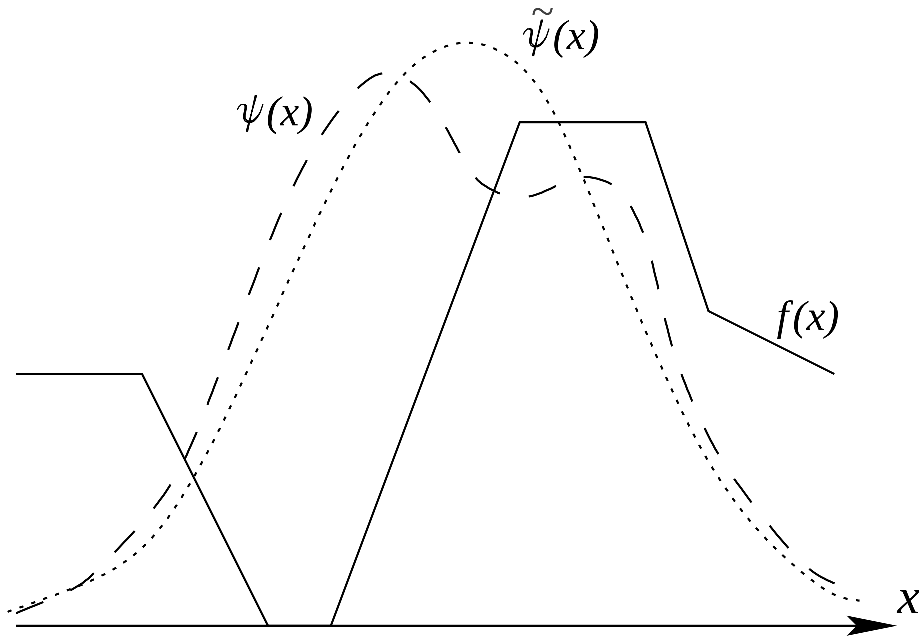
$$\int f(x) \cdot \psi(x)\,\mathrm{d}x \quad = \quad \int f(x) \frac{\psi(x)}{\widetilde{\psi}(x)} \cdot \widetilde{\psi}(x)\,\mathrm{d}x \tag{23}$$

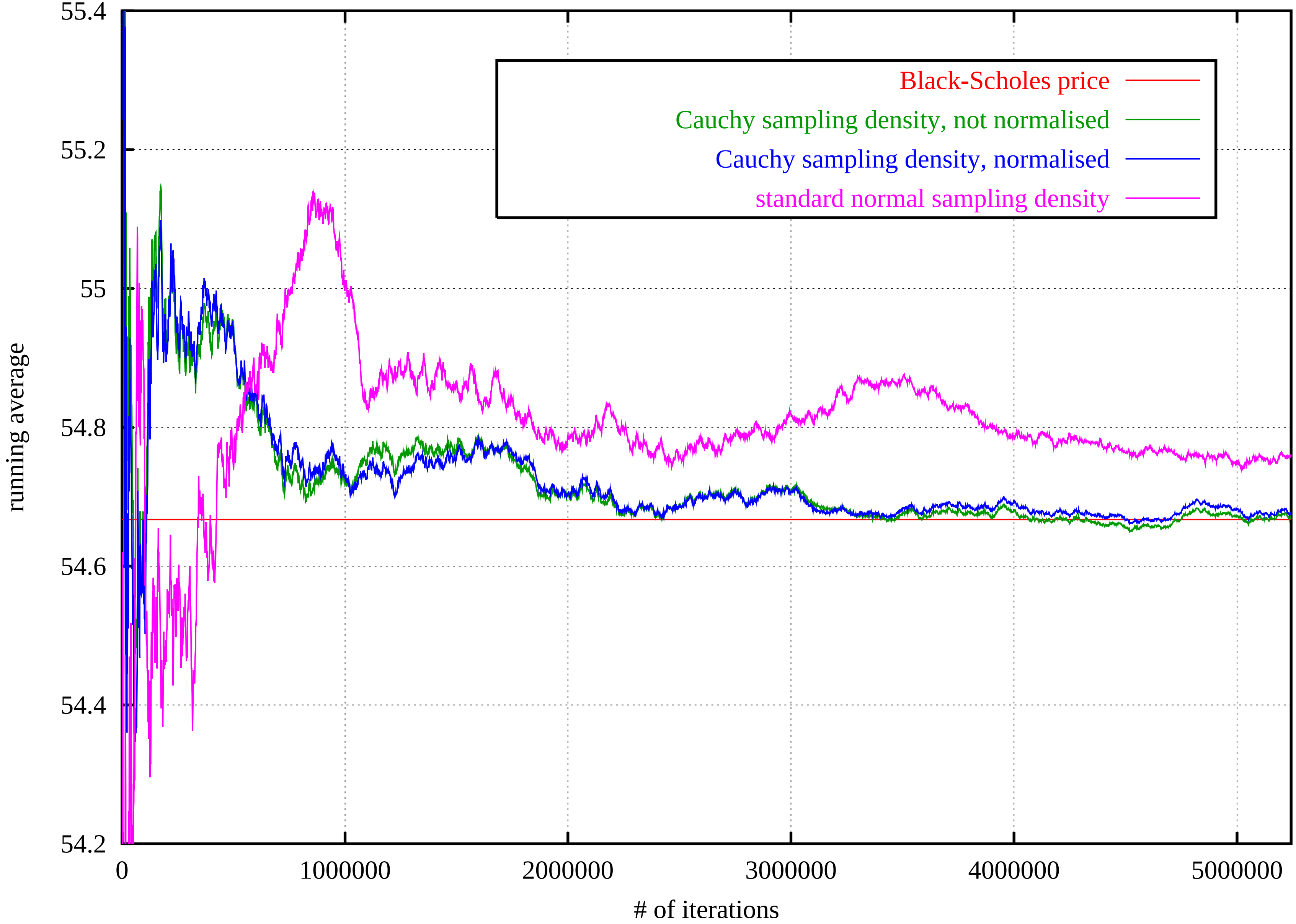
we can, instead of (16), use the *sampler density Monte Carlo estimator*:

$$\hat{v}_N \quad := \quad \frac{1}{N} \sum_{i=1}^{N} f(x_i) \left(\frac{\psi(x_i)}{\widetilde{\psi}(x_i)} \right) \quad . \tag{24}$$

When $f(x)$ is nearly constant, it may be preferable to use the *normalised sampler density Monte Carlo estimator*:

$$\hat{\upsilon}_N \quad := \quad \frac{\sum_{i=1}^{N} f(x_i)\left(\frac{\psi(x_i)}{\widetilde{\psi}(x_i)}\right)}{\sum_{i=1}^{N}\left(\frac{\psi(x_i)}{\widetilde{\psi}(x_i)}\right)} \quad . \tag{25}$$





Pricing an out-of-the-money call option using a Cauchy sampling density $\widetilde{\psi}(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}$ with $\widetilde{\Psi}^{-1}(u) = tan\left(\pi \cdot (u - {}^1\!/2)\right)$.

Importance sampling

is but a suitable choice of a sampler density such that the Monte Carlo simulation to compute

$$\mathsf{E}_{\widetilde{\psi}(x)}\left[f(x) \left(\frac{\psi(x_i)}{\widetilde{\psi}(x_i)} \right) \right]$$

converges more rapidly than

$$\mathsf{E}_{\psi(x)}[f(x)] \ ,$$

even though their convergence levels are identical.

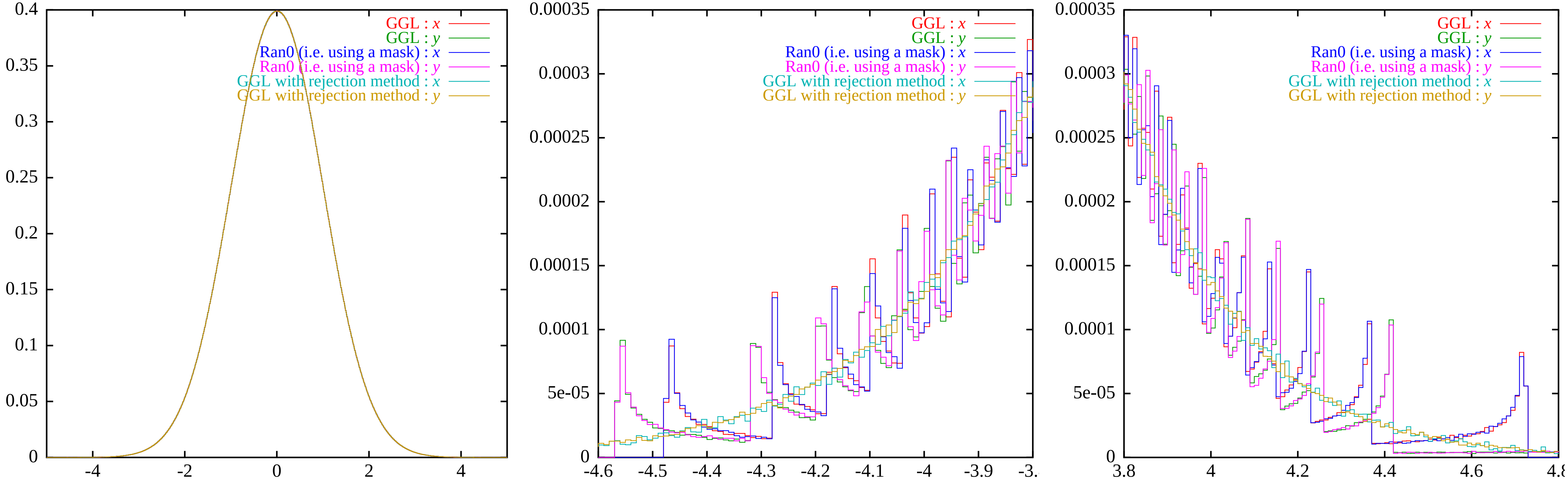
Normal variates

- The preferred method should always be Peter Acklam's direct inverse cumulative normal [Ack00].
- *Never* use Excel's `Application.NormSInv` for any simulation.
- Box-Muller is deprecated for three reasons:

1. It requires a pairwise pipelined generator.
2. It is less efficient than the direct inverse cumulative normal.
3. The Neave effect and similar entrainments.

The Neave effect

On a global scale, all combinations look pretty much the same



The dramatic Neave [Nea73] effect above only arises for the direct trigonometric version of the Box-Muller algorithm [BM58].

However, there are indications that other undesirable effects [Tez95] can arise due to the interaction of the nonlinear coupling of the transformation scheme and the number generation algorithm itself. In the physical and mathematical sciences of nonlinear dynamics, this phenomenon is known as *entrainment*. It is practically impossible to predict for what number generators it can arise. It is much safer to stay away from transformation methods that introduce unintended coupling.

VIII. Efficiency ratio and yield

Assume that we are using a variate generation algorithm based on the sampler density $\chi(x)$, and that the target density is $\psi(x)$, with the domain where ψ is non-zero being a subset of the domain where χ is non-zero. This means that for any x, we have

$$\psi(x) > 0 \quad \Rightarrow \quad \chi(x) > 0 \tag{26}$$

$$\chi(x) \quad \leq \quad \psi(x) \tag{27}$$

The standard error of a simulation based on χ whose objective is to compute $\mathsf{E}_{\psi(x)}[f(x)]$ is governed by the variance

$$V_\chi[f] \;=\; \int \left(f(x)\frac{\psi(x)}{\chi(x)} \right)^2 \chi(x)\,\mathrm{d}x \;-\; \left(\int f(x)\frac{\psi(x)}{\chi(x)}\chi(x)\,\mathrm{d}x \right)^2 . \tag{28}$$

We define the *efficiency ratio* of the Monte Carlo simulation based on the sampler density $\chi(x)$ as

$$\eta_\chi[f] \;:=\; \frac{V_\psi[f]}{V_\chi[f]}\;. \tag{29}$$

Note that the efficiency ratio is a functional of the function f whose expectation we intend to compute.

Next, we define the *yield* of the Monte Carlo simulation based on the sampler density $\chi(x)$ as

$$y_\chi[f] \;:=\; \frac{\int f^2(x)\psi(x)\,\mathrm{d}x}{\int f^2(x)\frac{\psi(x)}{\chi(x)}\psi(x)\,\mathrm{d}x}\;. \tag{30}$$

Without any knowledge of f, we can use the *unbiased yield*

$$y_\chi \;:=\; \left(\int \frac{\psi(x)}{\chi(x)}\psi(x)\,\mathrm{d}x\right)^{-1} \tag{31}$$

to assess the computational burden of a sampler density algorithm.

Note that, whilst the functional yield $y_\chi[f]$ can be greater than 100% (importance sampling methods), the unbiased yield y_χ cannot.

IX. Codependence

Given the joint distribution density $\psi(x, y)$ of the two variables x and y, the *marginal* distribution density function of x is defined as

$$\psi_x(x) = \int \psi(x, y)\, \mathrm{d}y \ , \tag{32}$$

and analogously,

$$\psi_y(y) = \int \psi(x, y)\, \mathrm{d}x \ . \tag{33}$$

The marginal distribution density of any one of the two variables is nothing other than the probability density disregarding the value of the second variable.

Two variates x and y are considered *independent* if their joint distribution density function separates into the product of their individual distribution density functions, i.e.

$$\psi(x, y) = \psi_x(x)\psi_y(y) \ . \tag{34}$$

There are several ways to measure codependence.

Linear correlation

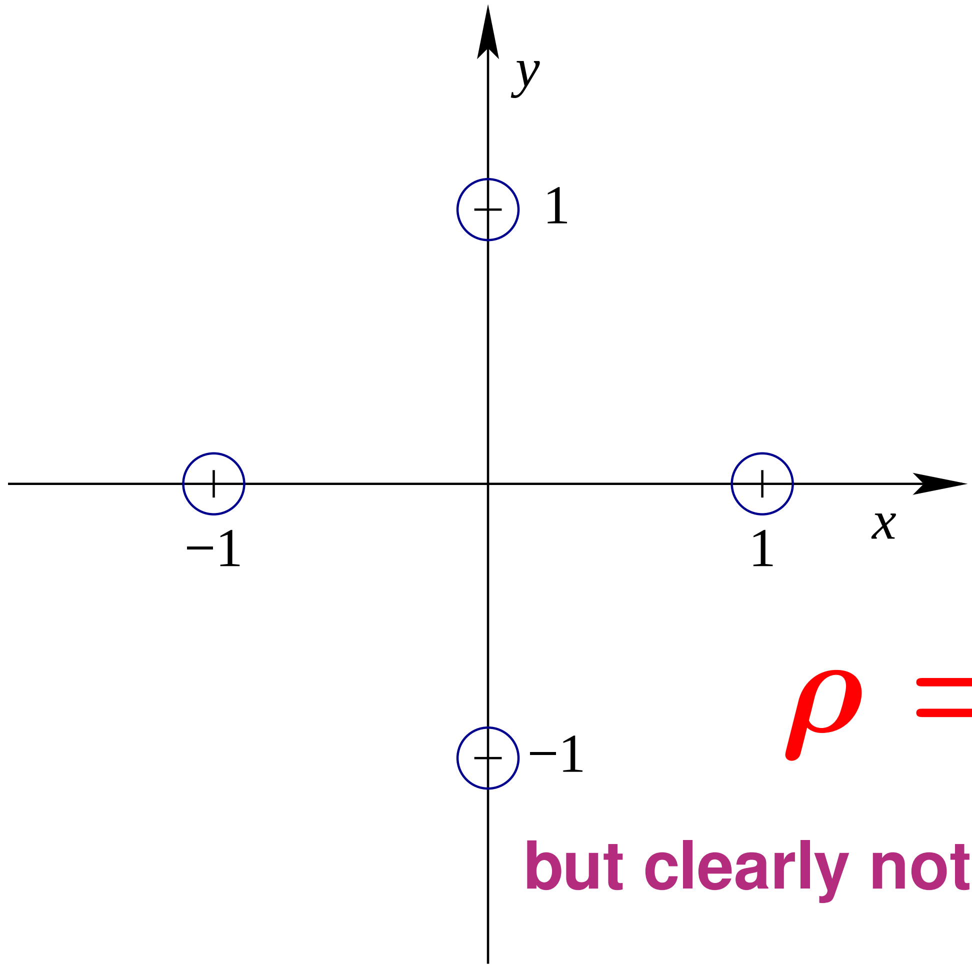
The *linear correlation* $\rho_{xy} := \mathsf{Corr}[x,y]$ of two variates x and y is defined as

$$\rho_{xy} = \frac{\mathsf{Cov}[x,y]}{\sqrt{\mathsf{V}[x]\,\mathsf{V}[y]}} = \frac{\int xy\psi(x,y)\mathrm{d}x\mathrm{d}y - \int x\psi_x(x)\mathrm{d}x \int y\psi_y(y)\mathrm{d}y}{\sqrt{\int x^2\psi_x(x)\mathrm{d}x - \left[\int x\psi_x(x)\mathrm{d}x\right]^2}\sqrt{\int y^2\psi_y(y)\mathrm{d}y - \left[\int y\psi_y(y)\mathrm{d}y\right]^2}} \,. \tag{35}$$

Linear correlation is a good measure for the co-dependence of normal variates.

For strongly non-normal distributions, it can be highly misleading.

Consider the discrete distribution where the variate pair (x, y) can take on the possible combinations $\{(0,1),(0,-1),(1,0),(-1,0)\}$ with equal probability:



$$\rho = 0,$$

but clearly not independent!

Correlation is a term that makes sense only for some types of codependence.

Spearman's rho

Spearman's rho is defined as the linear correlation coefficient of the probability-transformed variates, i.e. of the variates transformed by their own cumulative marginal distribution functions:

$$\rho_{\mathsf{S}} := \frac{\iint \Psi_x(x)\Psi_y(y)\psi(x,y)\mathrm{d}x\mathrm{d}y - \int \Psi_x(x)\psi_x(x)\mathrm{d}x \int \Psi_y(y)\psi_y(y)\mathrm{d}y}{\sqrt{\int \Psi_x(x)^2\psi_x(x)\mathrm{d}x - \left[\int \Psi_x(x)\psi(x)\mathrm{d}x\right]^2}\sqrt{\int \Psi_y(y)^2\psi_y(y)\mathrm{d}y - \left[\int \Psi_y(y)\psi(y)\mathrm{d}y\right]^2}} \qquad (36)$$

Spearman's rho can be expressed as

$$\rho_{\mathsf{S}} := 12\iint \Psi_x(x)\Psi_y(y)\psi(x,y)\,\mathrm{d}x\mathrm{d}y - 3\ . \qquad (37)$$

Spearman's rho is independent with respect to variable transformations, whether linear or not.

Kendall's tau

For continuous distributions, *Kendall's tau* is given by

$$\tau_{\mathsf{K}} = 4 \iint \Psi(x,y)\psi(x,y)\,\mathrm{d}x\mathrm{d}y - 1 \ . \tag{38}$$

Since Kendall's tau is defined on the joint cumulative probability, it is also invariant with respect to transformations.

Kendall's tau and Spearman's rho belong to the category of *rank correlations*.

Rank correlations have the nice property that for any two marginal distribution densities $\psi_x(x)$ and $\psi_y(y)$, there always exists a joint distribution density $\psi(x,y)$ for every possible value in $[-1,1]$ of the rank correlation.

This is not necessarily guaranteed for linear correlation!

Gaussian variates

Given the covariance matrix C, and *any* pseudo-square root A of C such that $C = A \cdot A^\top$, a vector $\boldsymbol{z}$ of independent Gaussian variates can be transformed into a vector of correlated Gaussian variates $\boldsymbol{x}$ by setting

$$\boldsymbol{x} = A \cdot \boldsymbol{z} \ . \tag{39}$$

The covariances of the entries of the vector x are then

$$\begin{aligned}
\langle \boldsymbol{x} \cdot \boldsymbol{x}^\top \rangle &= \langle A \cdot \boldsymbol{z} \cdot \boldsymbol{z}^\top \cdot A^\top \rangle \\
&= A \cdot \langle \boldsymbol{z} \cdot \boldsymbol{z}^\top \rangle \cdot A^\top \\
&= A \cdot \mathbf{1} \cdot A^\top \\
&= A \cdot A^\top \\
&= C
\end{aligned}$$

as desired.

For specific covariance matrices, individual ways to compute a pseudo-square root are available. For instance, the covariance matrix of the path realisations of a Wiener process can be split into the *Brownian bridge*.

The two most popular methods to split any *generic* covariance matrix C into $A \cdot A^\top$ are:-

- Cholesky decomposition.

  Advantages: the effort in computing $\boldsymbol{x} = A \cdot \boldsymbol{z}$ for $\boldsymbol{z} \in \mathbb{R}^n$ is only $n \cdot \frac{n+1}{2}$.

  Disadvantages:

  – Telescopic dependence of the last variates on the first makes this method susceptible to roundoff error accumulation.
  – The standard Cholesky method as found in textbooks on numerical analysis fails for rank-deficient or ill conditioned covariance matrices. However, with some extra care, this problem can be remedied since the Cholesky algorithm can be readily amended to handle rank deficiencies.

- Spectral decomposition. Take the eigendecomposition

$$C = S \cdot \Lambda \cdot S^\top \tag{40}$$

with Λ being a diagonal matrix of non-negative entries and set

$$A := S \cdot \sqrt{\Lambda}\ . \tag{41}$$

Disadvantages: the effort in computing $\boldsymbol{x} = A \cdot \boldsymbol{z}$ for $\boldsymbol{z} \in \mathbb{R}^n$ is always n^2.

Advantages:

– Rank deficiencies are handled gracefully, especially if a stable eigensystem calculation algorithm is used.

– The resulting matrix A can be designed to associate the modes responsible for the majority of the variance in C with any particular entry of $\boldsymbol{z}$ at will.

Copulæ

A *copula* of two variables x and y is a cumulative probability function defined directly as a function of the marginal cumulative probabilities of x and y.

A copula of n variables is a function $C : [0,1]^n \to [0,1]$.

$$\Psi(x,y) = C\left(\Psi_x(x), \Psi_y(y)\right) \ . \tag{42}$$

For strictly increasing cumulative marginals $\Psi_x(x)$ and $\Psi_y(y)$, we can also write

$$C(u,v) = \Psi\left(\Psi_x^{-1}(u), \Psi_y^{-1}(v)\right) \ . \tag{43}$$

The copula of independent variables, not surprisingly, is given by

$$C_{\text{independent}}(u,v) = u \cdot v. \tag{44}$$

By virtue of the definition on the cumulative marginal distribution functions, the copula of a set of variables (x,y) is invariant with respect to a set of strictly increasing transformations $(f(x), g(y))$.

The Gaussian Copula

Assume that we are given a symmetric, positive-semidefinite matrix R of codependence coefficients that are to be used to generate vectors of codependent uniform $(0, 1)$ variates:

- Find a suitable pseudo-square root A of R such that $R = A \cdot A^\top$.
- Draw a vector $\boldsymbol{z} \in \mathbb{R}^n$ of uncorrelated standard normal variates.
- Compute $\tilde{\boldsymbol{z}} := A \cdot \boldsymbol{z}$.
- Map $\tilde{\boldsymbol{z}}$ back to a vector of uniform variates $\boldsymbol{v} \in [0, 1]^n$ by setting $v_i = \mathsf{N}(\tilde{z}_i)$.

It can be shown [LMS01, Kau01] that Kendall's tau of two variables connected by a Gaussian copula with codependence coefficient ρ is given by

$$\tau_{\mathsf{K}} = \frac{2}{\pi} \arcsin \rho \; . \tag{45}$$

Example: *Two uniform variates under the Gaussian copula*

The linear correlation η between the two dependent uniform variates under the Gaussian copula can be calculated:

$$\eta(\rho) = 12 \iint \mathsf{N}(z_1)\mathsf{N}(\rho z_1 + \sqrt{1-\rho^2} z_2)\varphi(z_1)\varphi(z_2)\,\mathrm{d}z_1\mathrm{d}z_2 - 3\;. \tag{46}$$

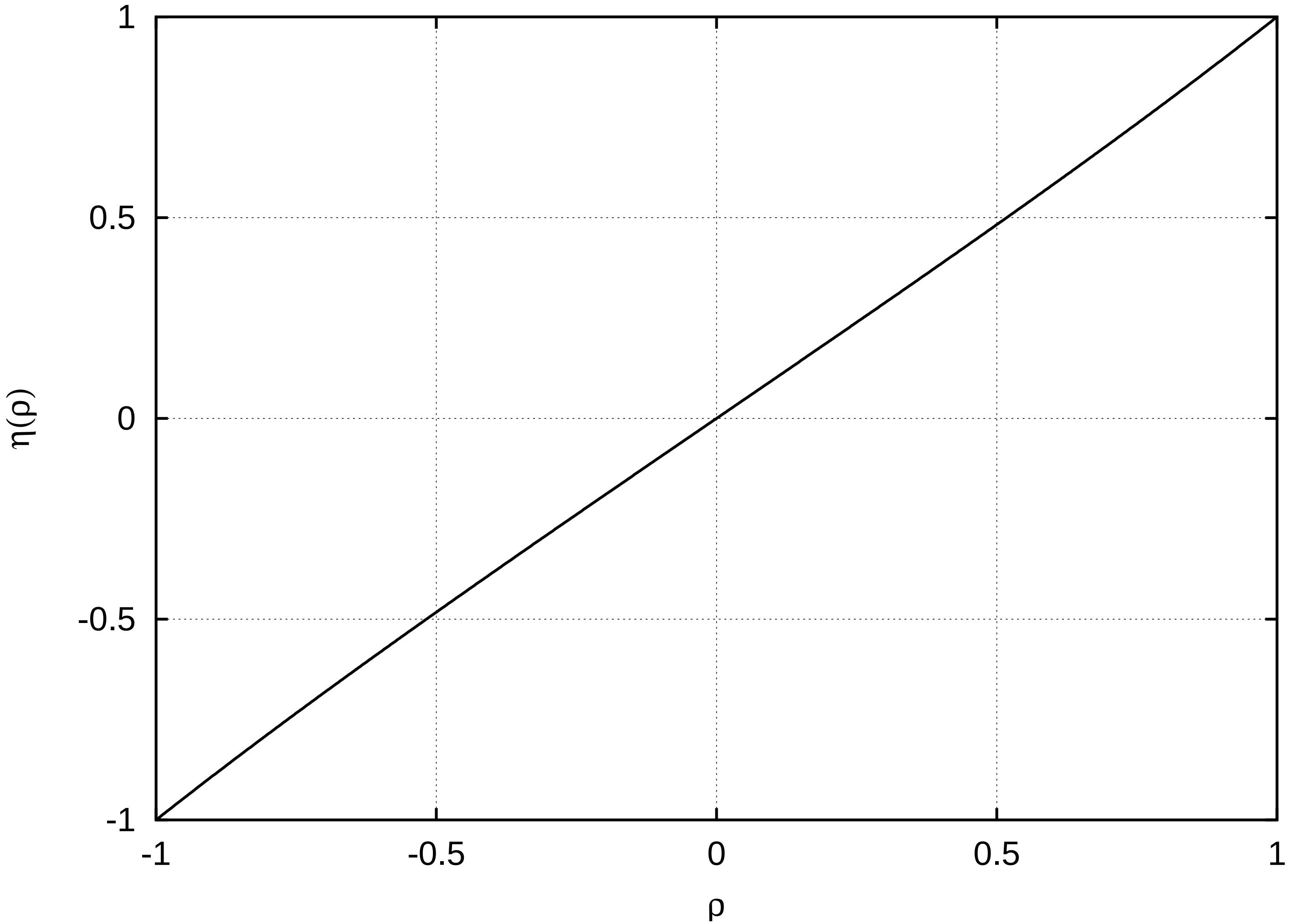
Using equation (26.3.19) in [AS84], namely

$$\mathsf{N}(0,0,\rho) \;=\; \rho_{\mathsf{S}} \;=\; \frac{1}{4} + \frac{1}{2\pi}\arcsin\rho\;,$$

we have the exact relationship

$$\eta(\rho) \;=\; \rho_{\mathsf{S}} \;=\; 2\cdot\frac{3}{\pi}\arcsin\frac{\rho}{2}\;. \tag{47}$$

Near the origin, this means: $\eta(\rho) \;=\; \rho_{\mathsf{S}} \approx \frac{3}{\pi}\rho$ for $|\rho| \ll 1$.



Don't be misled by the apparently straight line: there is a little bit of curvature in there, although a straight line would certainly be a good approximation for it.

Example: *Two exponential variates under the Gaussian copula*

$$\tau \sim \lambda \mathrm{e}^{-\lambda\tau} \ . \tag{48}$$

Random draws for τ can be generated from a uniform variate u by setting

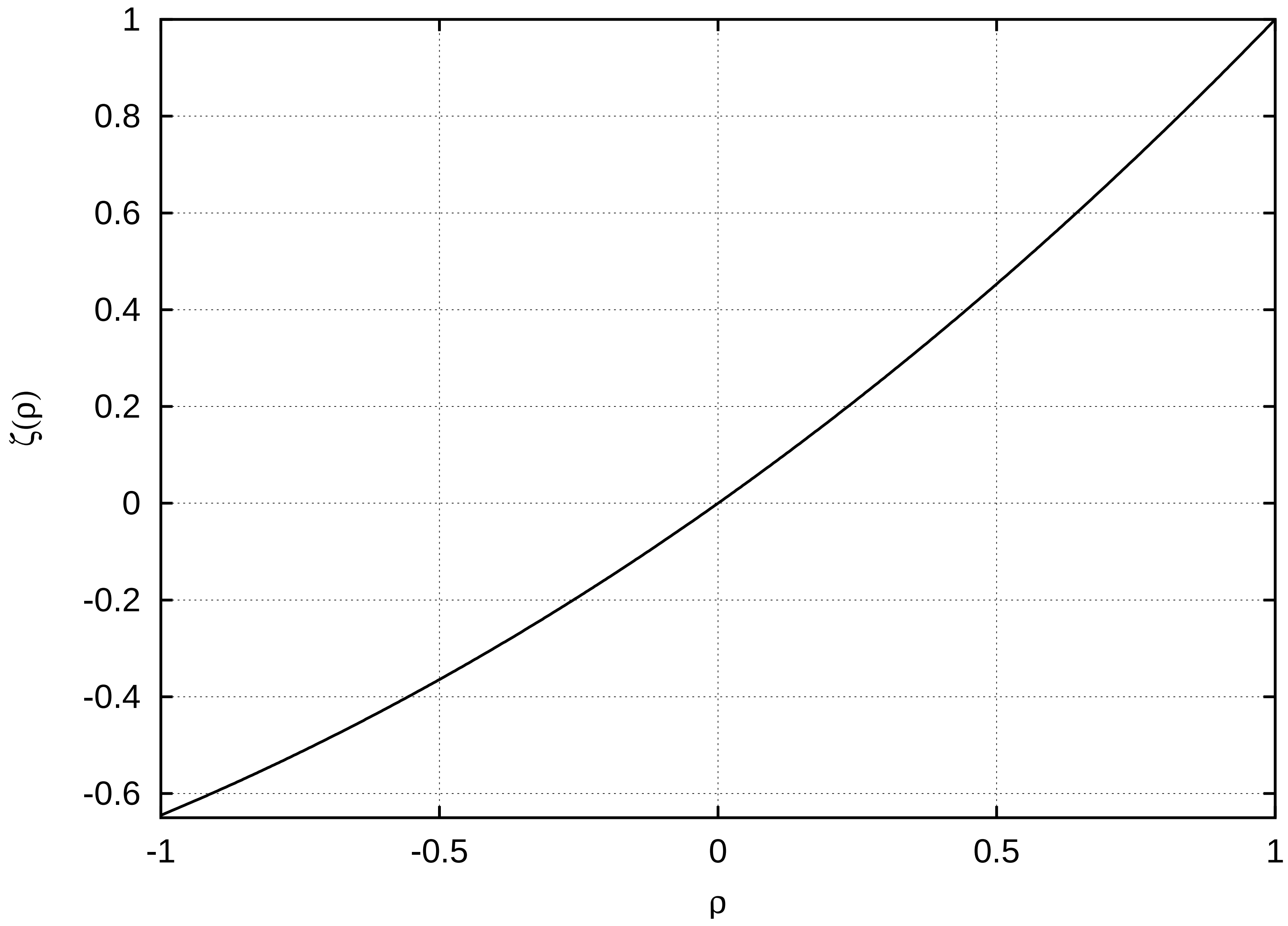
$$\tau = -\frac{\ln(1-u)}{\lambda} \ . \tag{49}$$

The linear correlation of two events τ_a and τ_b using a Gaussian copula with codependence coefficient ρ is given by

$$\zeta(\rho) = \iint \ln(1-\mathsf{N}(z_a))\ln(1-\mathsf{N}(\rho z_a + \sqrt{1-\rho^2} z_b))\varphi(z_a)\varphi(z_b)\,\mathrm{d}z_a\mathrm{d}z_b - 1 \ . \tag{50}$$

We can evaluate analytically what interval $\zeta(\rho)$ is confined to:

$$\zeta \in [1-\frac{\pi^2}{6}, 1] \quad \text{for} \quad \rho \in [-1,1] \ , \quad \text{where} \quad 1-\frac{\pi^2}{6} \approx -0.6449341.$$


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
ζ(ρ)
-1
-0.5
0
0.5
1
ρ

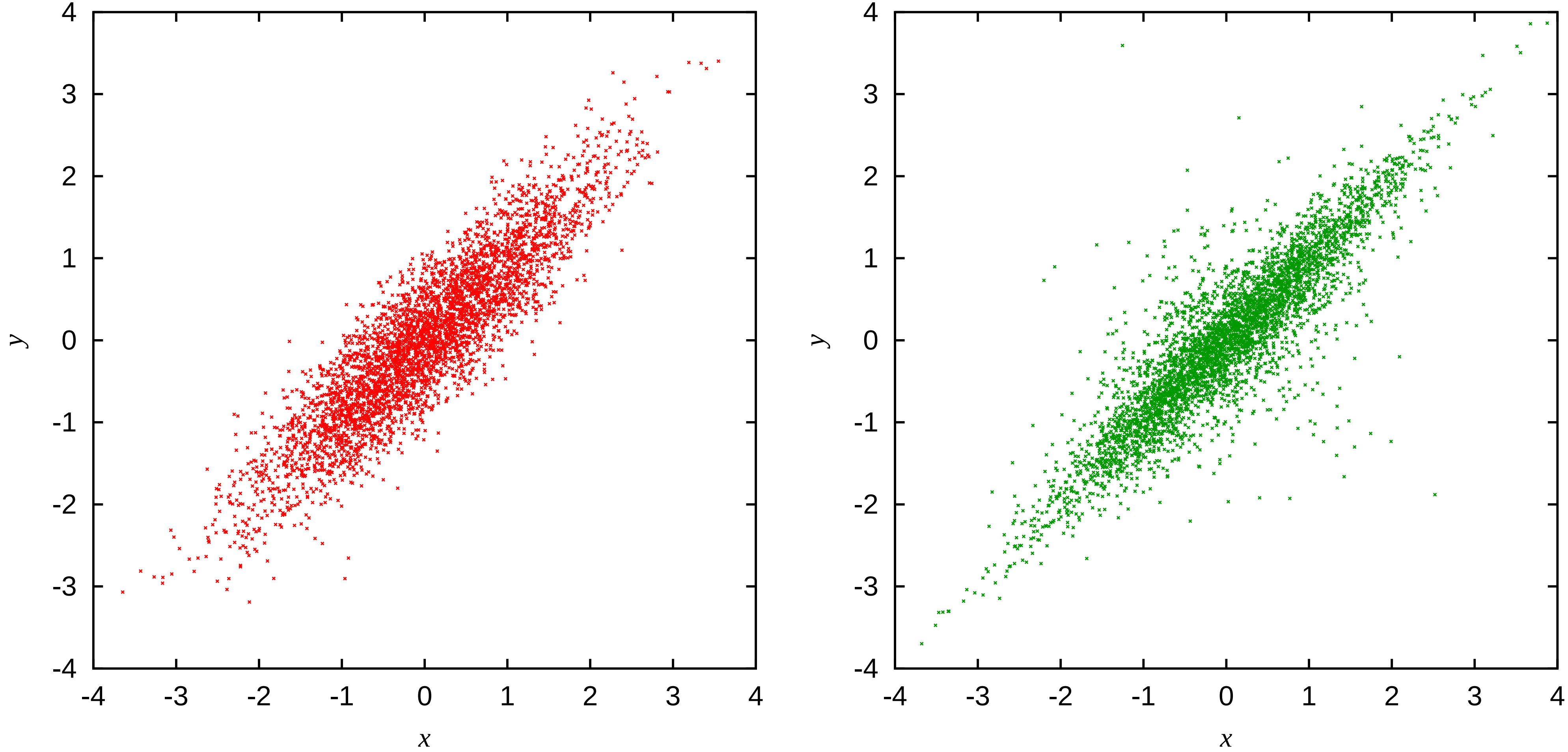
Generic elliptic copulæ

- Select a standard correlation coefficient matrix R and find a pseudo-square root A of R such that $R = A \cdot A^\top$.
- Draw a vector $\boldsymbol{z} \in \mathbb{R}^n$ of uncorrelated standard normal variates.
- Compute $\tilde{\boldsymbol{z}} := A \cdot \boldsymbol{z}$.
- Draw an independent, non-negative, variate q, from some distribution χ.
- Set $\boldsymbol{x} := q \cdot \tilde{\boldsymbol{z}}$.
- Map $\boldsymbol{x}$ back to a vector of uniform variates $\boldsymbol{v} \in [0,1]^n$ using the cumulative probability function of $q \cdot z$ with $q \sim \chi(q)$ and $z \sim \varphi(z)$ that has to be computed according to the choice of χ.

Kendall's tau for any pair of variates connected by a generic elliptic copula is also given by equation (45).

The t-copula

is an elliptic copula with the specific choice that $q := \sqrt{\frac{\nu}{s}}$ with $s \sim \chi^2_\nu(s)$, i.e. s is a chi-square variate with ν degrees of freedom. The distribution of $q \cdot z$ is that of a Student-t variate with ν degrees of freedom.



Both figures show data that are marginally normally distributed. The figure on the left depicts ordinary correlated normal variates with $\rho = 0.9$, and the figure on the right was created from data under a t_2-copula, also with $\rho = 0.9$.

There are many other copulæ. In particular, there is the family of *Archimedean copulæ*. All members of this class of copulæ have in common that they are generated by a strictly decreasing convex function $\phi(u)$ which maps the interval $(0,1]$ onto $[0,\infty)$ such that $\lim_{\epsilon\to 0}\phi(\epsilon)=\infty$ and $\phi(1)=0$. An Archimedean copula is generated from a given function $\phi(u)$ by

$$C(u,v)=\phi^{-1}\left(\phi(u)+\phi(v)\right)\ . \tag{51}$$

Two uniform variates u and v under any Archimedean copula can be produced by the following algorithm:

- Draw two independent uniform variates s and q.
- Solve

$$q=t-\frac{\phi(t)}{\phi'(t)} \tag{52}$$

  for t.

- Set

$$u := \phi^{-1}\left(s\phi(t)\right) \tag{53}$$

$$v := \phi^{-1}\left((1-s)\phi(t)\right) \ . \tag{54}$$

For further details see [ELM01].

The main drawback of these copulæ is that it is difficult to extend them to higher dimensions whilst allowing for one separate codependence coefficient for every possible pair.

X. Wiener path construction

In many applications, we need to construct a simulated discretised path of a standard Wiener process over a set $\{t_i\}, i = 1..n,$ points in time.

A discretised Wiener path is given by the associated values $w_i := W(t_i)$.

We can view the values $w_i := W(t_i)$ of the Wiener process at those points in time as a vector of Gaussian variates.

The elements of their covariance matrix C, are given by

$$c_{ij} = \text{Cov}[W(t_i), W(t_j)] = \min(t_i, t_j) \ . \tag{55}$$

The special structure of the covariance matrix gives us immediately three choices for the construction of paths.

Incremental

$$w_{i+1} = w_i + \sqrt{\Delta t_{i+1}} \cdot z_i, \quad \text{with} \quad z_i \sim \mathcal{N}(0,1) \ . \tag{56}$$

The construction matrix of the incremental method is given by the Cholesky decomposition of the covariance matrix:

$$A_{\text{incremental}} = \begin{pmatrix} \sqrt{\Delta t_1} & 0 & 0 & 0 & \cdots & 0 \\ \sqrt{\Delta t_1} & \sqrt{\Delta t_2} & 0 & 0 & \cdots & 0 \\ \sqrt{\Delta t_1} & \sqrt{\Delta t_2} & \sqrt{\Delta t_3} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sqrt{\Delta t_1} & \sqrt{\Delta t_2} & \sqrt{\Delta t_3} & \cdots & \cdots & \sqrt{\Delta t_n} \end{pmatrix} \tag{57}$$

The advantage of this method is that it is fast. For each path over n points in time, we need a total of n multiplications, and $n-1$ additions.

Spectral

The most efficient and reliable way to construct Wiener paths from a spectral decomposition of the pathwise covariance is to use the generic spectral decomposition approach.

Whilst certain analytical approximations are possible in the general case, and whilst an exact decomposition is known when all time steps are exactly equal, all in all it is hardly ever worth considering anything other than an efficient numerical eigensystem decomposition as described, for instance, in [PTVF92].

The Brownian bridge

Similar to the spectral path construction method, the *Brownian bridge* is a way to construct a discretised Wiener process path by using the first Gaussian variates in a vector draw $\boldsymbol{z}$ to shape the overall features of the path, and then add more and more of the fine structure.

The very first variate z_1 is used to determine the realisation of the Wiener path at the final time t_n of our n-point discretisation of the path by setting $W_{t_n} = \sqrt{t_n} z_1$.

The next variate is then used to determine the value of the Wiener process as it was realised at an intermediate timestep t_j conditional on the realisation at t_n (and at $t_0 = 0$ which is, of course, zero).

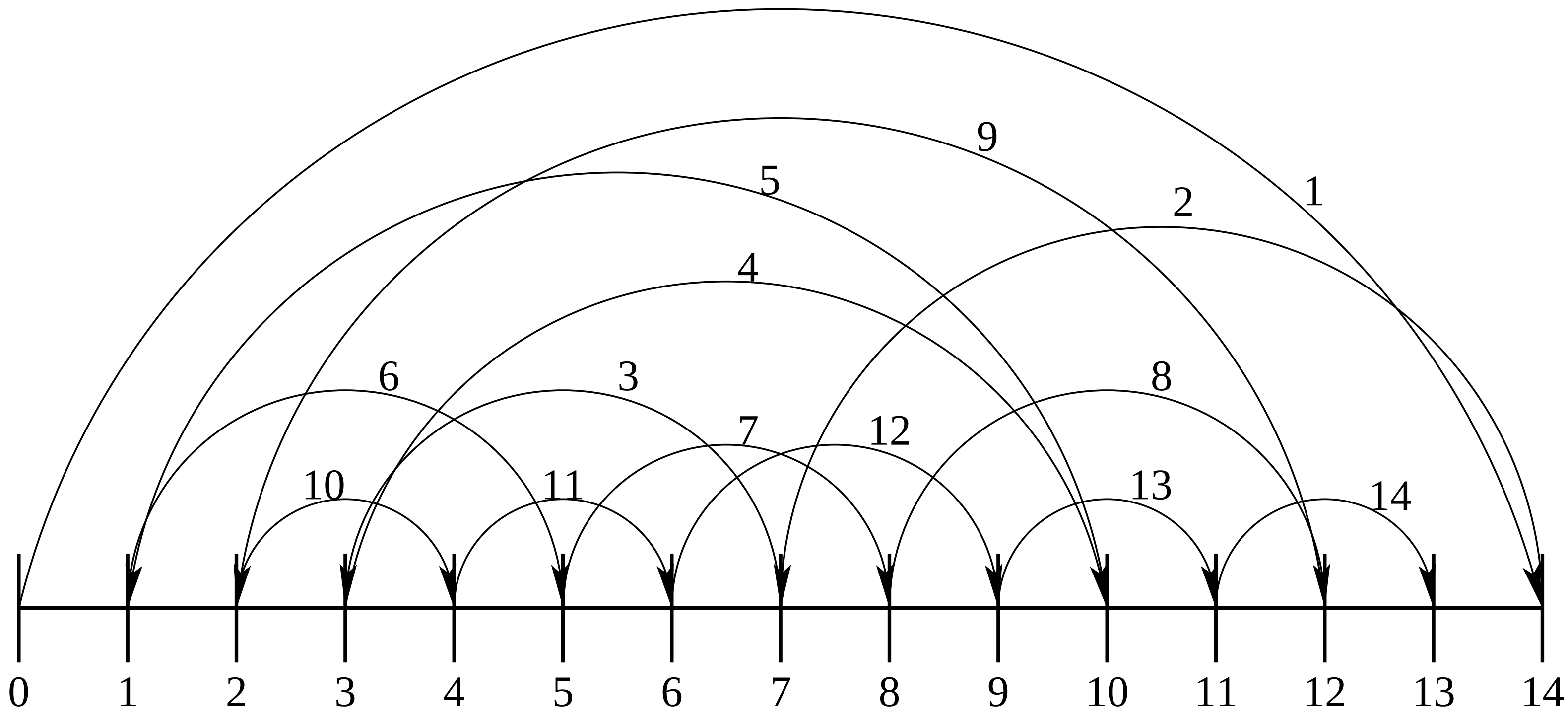
The procedure is then repeated to gradually fill in all of the realisations of the Wiener process at all intermediate points, in an ever refining algorithm.

In each step of the refinement procedure to determine W_{t_j} given that we have already established W_{t_i} and W_{t_k} with $t_i < t_j < t_k$, we make use of the fact that the conditional distribution of W_{t_j} is Gaussian with mean

$$\mathsf{E}\left[W_{t_j}\right] = \left(\frac{t_k - t_j}{t_k - t_i}\right) W_{t_i} + \left(\frac{t_j - t_i}{t_k - t_i}\right) W_{t_k} \tag{58}$$

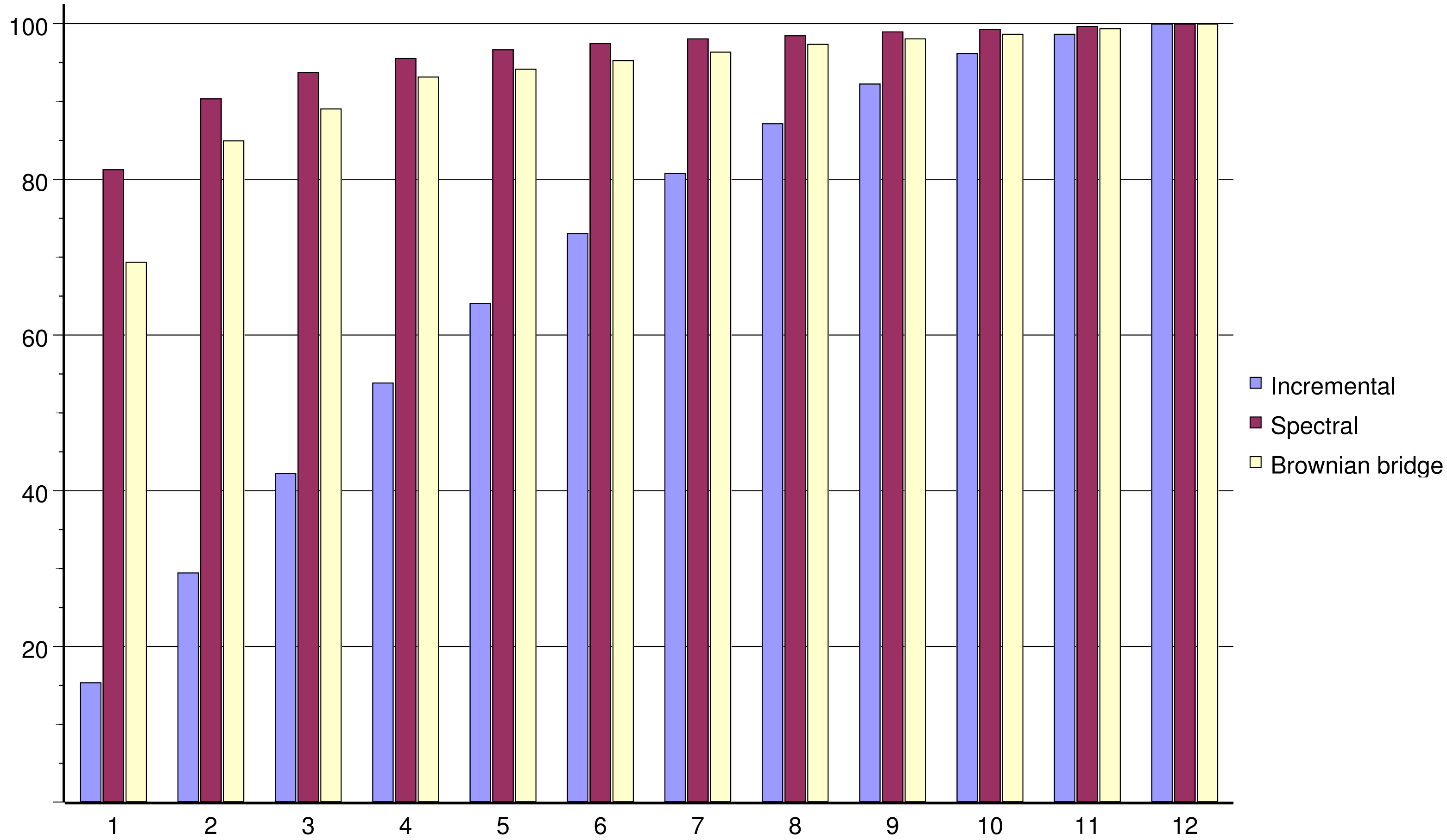
and variance

$$\mathsf{V}\left[W_{t_j}\right] = \frac{(t_j - t_i)\,(t_k - t_j)}{(t_k - t_i)}\ . \tag{59}$$



Comparison of path construction methods

The *variability explained* is the variance that remains if only the first m column vectors in a complete path construction matrix A are used. It is given by the sum of all of the squares of the elements of the first m column vectors of A.

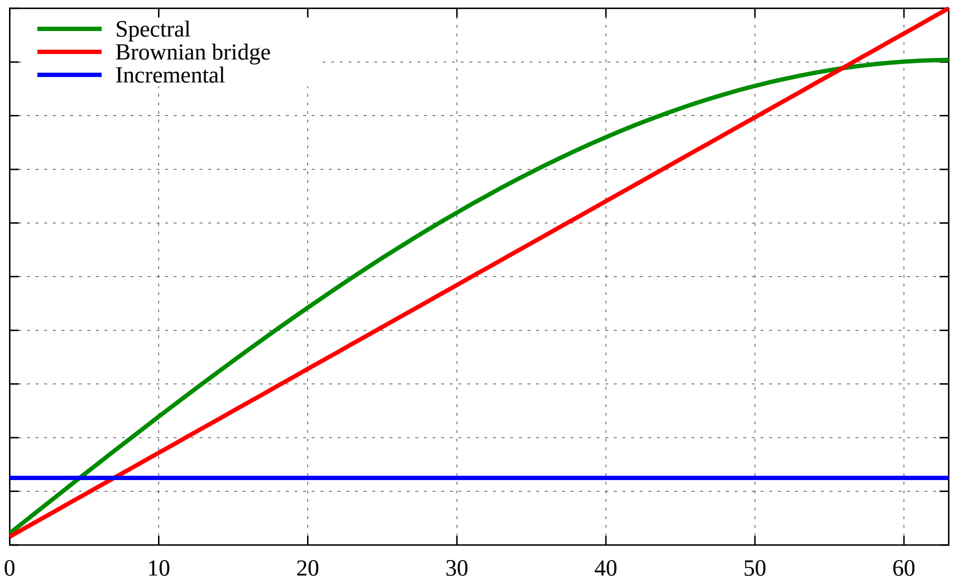


The cumulative variability explained (in %) by an increasing number of dimensions for three path construction methods over 12 equal time steps.

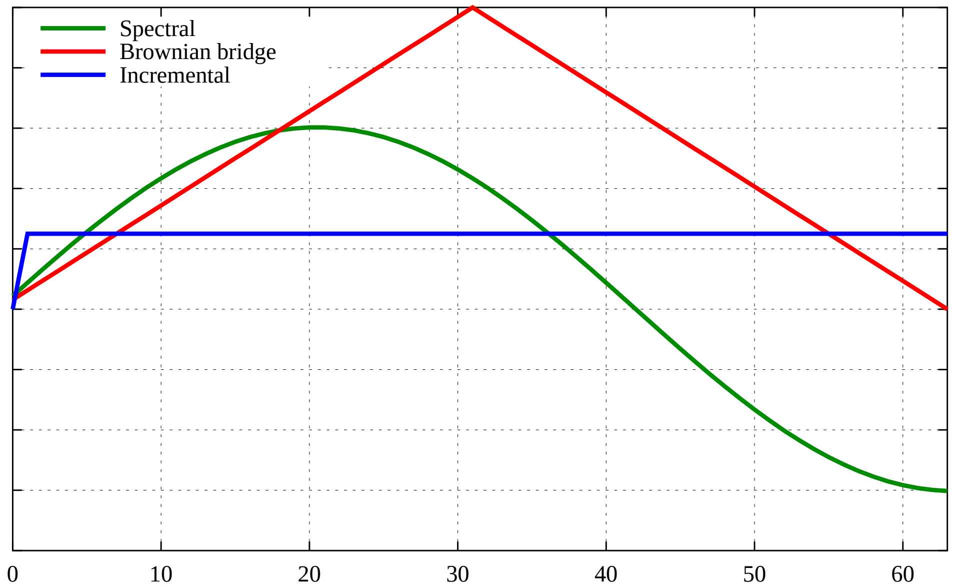
The Brownian bridge construction captures variability similarly to the spectral method whilst being almost as efficient as incremental path construction.

The construction modes of the Brownian bridge are akin to a piecewise linear approximation to the construction modes of the spectral method.

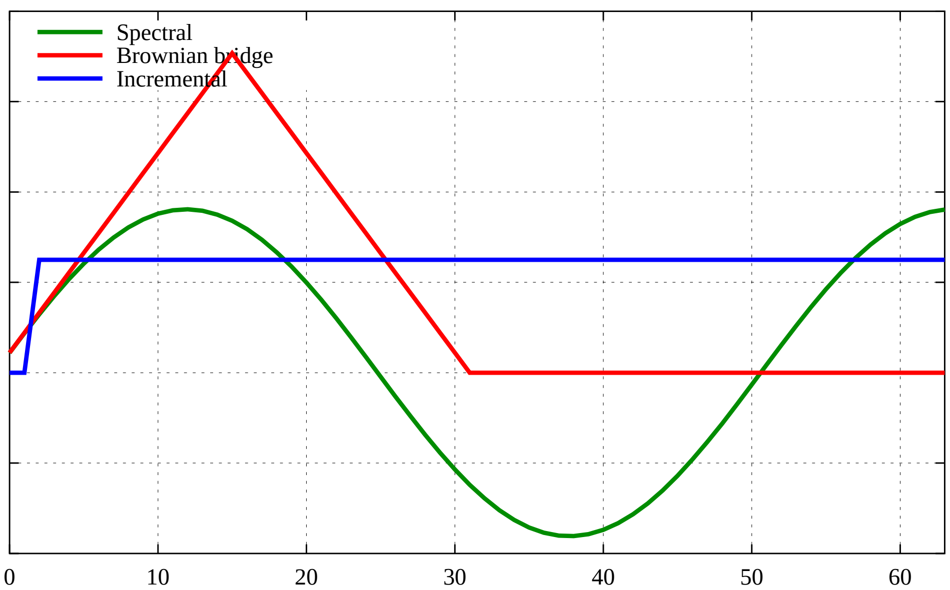
First column vectors of the construction matrices



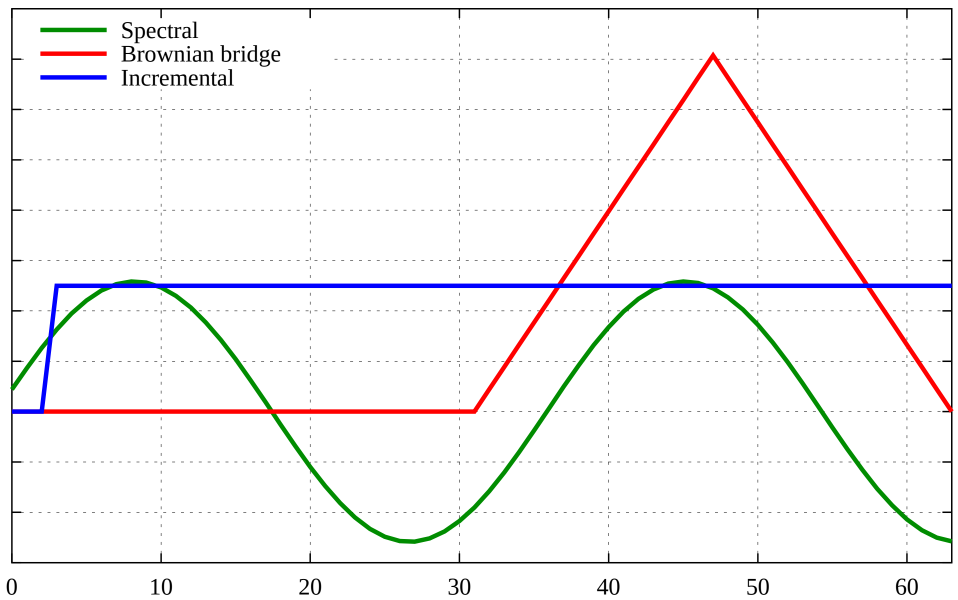
Second column vectors of the construction matrices



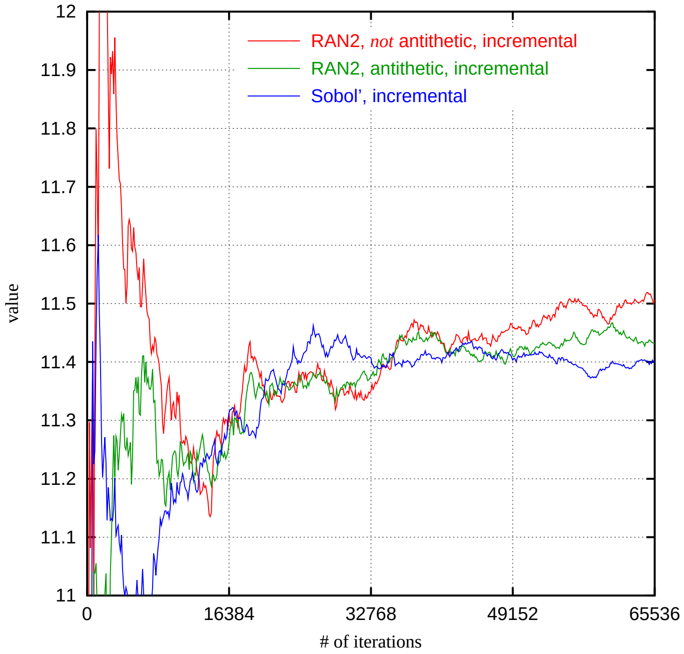
Third column vectors of the construction matrices

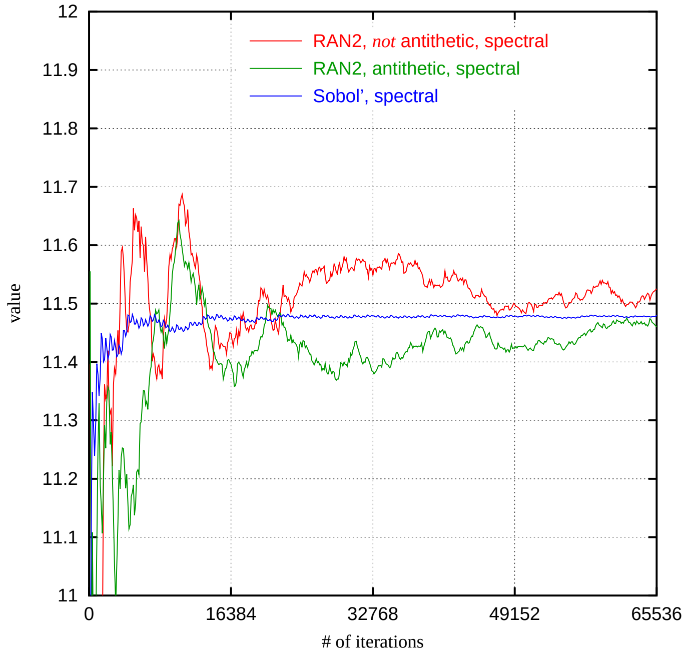


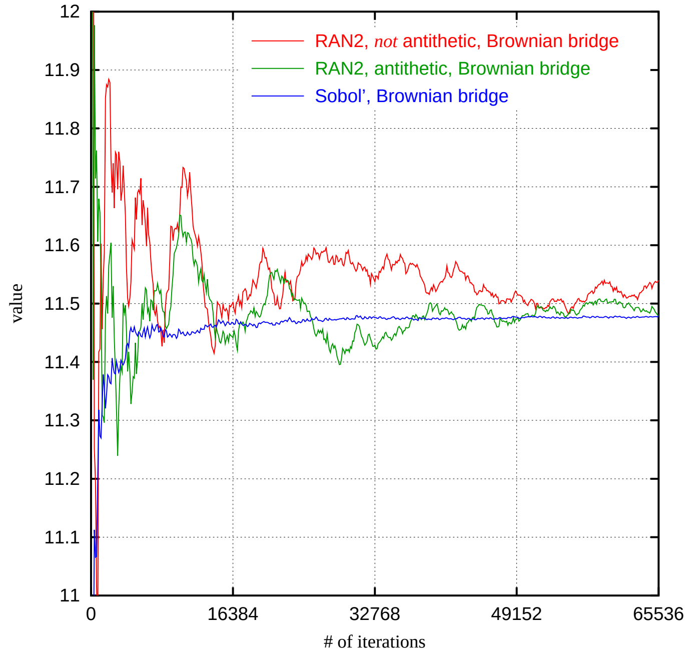
Fourth column vectors of the construction matrices



This explains the convergence behaviour we see for the different path construction methods in comparative Monte Carlo simulations.

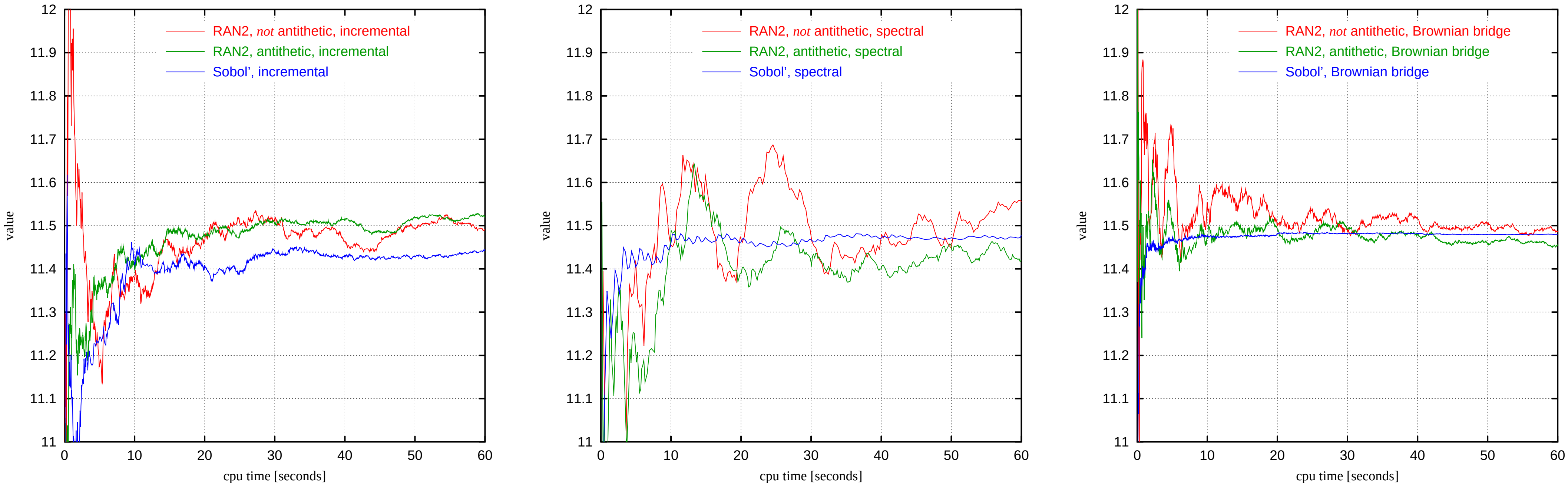






The example was a standard Asian option with one year to maturity and 252 monitoring points.

If we take into account the time spent in the path construction, the Brownian bridge immediately becomes the method of choice, irrespective of the specifics of the calculation.



XI. Poisson path construction

A Poisson process of intensity λ is given by the increments $\mathrm{d}N$ that over any infinitesimal timer interval $\mathrm{d}t$ are:-

$$\mathrm{d}N = \begin{cases} 1 & \text{with probability} \quad \lambda \mathrm{d}t \\ 0 & \text{with probability} \quad 1 - \lambda \mathrm{d}t \end{cases} . \tag{60}$$

A Poisson path is fully described by the set of event times. There are (at least) three ways to construct such a Poisson path:-

1. Draw Bernoulli variates over a time discretisation of interval size Δt. *Very bad*.

2. Draw independent exponential variates $\tau \sim \lambda \mathrm{e}^{-\lambda\tau}$ that represent the time between events. *Good and easy*.

3. Construct a *Poisson bridge* [Fox96]. *Excellent but somewhat more involved*.

A sketch of the Poisson bridge from 0 to T:

- Draw the number of events n to occur for this path in $(0, T)$ from the Poisson distribution

$$\Pr[n = k] = \mathrm{e}^{-\lambda T} \frac{(\lambda T)^k}{k!} \ . \tag{61}$$

- *Conditional on there being* n *events on* $(0, T)$, they are uniform on $(0, T)$.

- Set j to be the largest integer less than or equal to $n/2$.

- Draw the u_j from a beta distribution $\beta(j, n + 1 - j)$.

- Set the event time t_j to be $u_j \cdot T$.

- Now we need to fill $j - 1$ events into the interval $(0, t_j)$ and $n - j - 1$ events into the interval (t_j, T), which is to be done using the same algorithm after rescaling.

The generalization of a Poisson process leads to a *Gamma* process which can be constructed with a *Gamma bridge* [ALT03].

XII. Numerical integration of stochastic differential equations

Anyone ever considering numerical integration of SDEs should buy [KP99].

Let us assume that we have a stochastic differential equation of the form

$$\mathrm{d}X = a\,\mathrm{d}t + b\,\mathrm{d}W\ . \tag{62}$$

Note that both a and b can be functions of the process variable X and time. In the multi-dimensional case of m state variables X_i driven by d independent Wiener processes, we have

$$\mathrm{d}X_i = a_i(t, \boldsymbol{X})\,\mathrm{d}t + \sum_{j=1}^{d} b_{ij}(t, \boldsymbol{X})\,\mathrm{d}W_j\ . \tag{63}$$

The Euler scheme

Denote the numerical approximation to the solution of (62) for a scheme over equal steps of size Δt at time $n \cdot \Delta t$ as $Y(t_n)$. The explicit Euler scheme is then given by

$$Y(t_{n+1}) = Y(t_n) + a(t_n, Y(t_n))\, \Delta t + b(t_n, Y(t_n))\, \Delta W \ . \tag{64}$$

Not surprisingly, the implicit Euler scheme is given by

$$Y(t_{n+1}) = Y(t_n) + a(t_n, Y(t_{n+1}))\, \Delta t + b(t_n, Y(t_{n+1}))\, \Delta W \ . \tag{65}$$

The Milstein scheme

The Milstein scheme involves the addition of the next order terms of the Itô-Taylor expansion of equation (62). This gives

$$Y(t_{n+1}) = Y(t_n) + a(t_n, Y(t_n))\,\Delta t + b(t_n, Y(t_n))\,\Delta W + \tfrac{1}{2} b b' \left[\Delta W^2 - \Delta t\right] \ . \tag{66}$$

with

$$b' = \frac{\partial b(t, X)}{\partial X} \ . \tag{67}$$

The Milstein scheme is definitely manageable in the one-dimensional case.

However, its general multi-dimensional extension is not as straightforward as one may expect since it requires not only the drawing of standard normal variates for the simulation of the standard Wiener process increments ΔW for each dimension, but also additional ones to account for the Itô integrals involving the mixing terms $\sum_{j=1}^{d} b_{ij}(t, \boldsymbol{X})\,\mathrm{d}W_j$.

Predictor-Corrector

First take an explicit Euler step to arrive at the *predictor*

$$\bar{Y}_i(t_{n+1}) \quad = \quad Y_i(t_n) + a_i(t_n, \boldsymbol{Y}(t_n))\,\Delta t + \sum_{j=1}^{d} b_{ij}(t_n, \boldsymbol{Y}(t_n))\,\Delta W_j\ . \tag{68}$$

Next, select two weighting coefficients α and η in the interval $[0,1]$, usually near $^1\!/_2$, and calculate the *corrector*

$$\begin{aligned} Y_i(t_{n+1}) \quad = \quad & Y_i(t_n) + \left\{\alpha\,\bar{a}_i(t_{n+1}, \bar{\boldsymbol{Y}}(t_{n+1}); \eta) + (1-\alpha)\bar{a}_i(t_n, \boldsymbol{Y}(t_n); \eta)\right\}\Delta t \\ & + \sum_{j=1}^{m} \left\{\eta\, b_{ij}(t_{n+1}, \bar{\boldsymbol{Y}}(t_{n+1})) + (1-\eta) b_{ij}(t_n, \boldsymbol{Y}(t_n))\right\}\sqrt{\Delta t}\, z_j \end{aligned} \tag{69}$$

with

$$\bar{a}_i(t, \boldsymbol{Y}; \eta) := a_i(t, \boldsymbol{Y}) - \eta \sum_{j=1}^{m} \sum_{k=1}^{d} b_{kj}(t, \boldsymbol{Y}) \partial_{Y_k} b_{ij}(t, \boldsymbol{Y})\ . \tag{70}$$

The predictor-corrector scheme is very easy to implement, in particular for the special case that the coefficients b_{ij} don't depend on the state variables.

Practitioner's hint: if at all possible, transform your equations to shift all dependence on the state variables from the volatility term over to the drift term!

The biggest problems arise when none of the above schemes are guaranteed to remain in the domain of the process for the stochastic differential equation at hand.

In that case, we may need to select a boundary condition for the numerical scheme.

Alternatively, we can try to find a numerical approximation to Doss's approach of *pathwise solutions* (see [Dos77] or [KS91] (pages 295–296)).

Numerical integration by approximation of pathwise solutions

If there is a strong solution to

$$\mathrm{d}X = a(X)\mathrm{d}t + b(X)\mathrm{d}W \tag{71}$$

then it can be written as a function of the standard Wiener process W (with $W(0) = 0$) and a second process Y, i.e.

$$X = u(W, Y) \quad \text{with boundary condition} \quad u(0, Y) = Y \tag{72}$$

whereby the dynamics of the process Y are governed by an *ordinary differential equation*

$$\mathrm{d}Y = f(W, Y)\mathrm{d}t \tag{73}$$

for some function $f(W, Y)$ to be determined [Dos77]. Itô's lemma for X gives us

$$\mathrm{d}X = \frac{\partial u(W,Y)}{\partial W}\mathrm{d}W + \frac{\partial u(W,Y)}{\partial Y}\mathrm{d}Y + \frac{1}{2}\frac{\partial^2 u(W,Y)}{\partial W^2}\mathrm{d}t \tag{74}$$

which means that u must satisfy

$$\frac{\partial u(W,Y)}{\partial W} = b(u) \qquad \text{with} \qquad u(0,Y) = Y\ . \tag{75}$$

Differentiating with respect to Y gives us

$$\frac{\partial^2 u(W,Y)}{\partial W\,\partial Y} = b'(u)\frac{\partial u(W,Y)}{\partial Y}\ . \tag{76}$$

This is equivalent to

$$\frac{\partial \frac{\partial u(W,Y)}{\partial Y}}{\partial W} = b'(u)\frac{\partial u(W,Y)}{\partial Y} \tag{77}$$

which has the formal solution

$$\frac{\partial u(W,Y)}{\partial Y} = \mathrm{e}^{\int_0^W b'(u(z,Y))\mathrm{d}z} \tag{78}$$

Matching the terms proportional to $\mathrm{d}t$ in equations (71) and (74) gives

$$\frac{\partial u(W,Y)}{\partial Y}f(W,Y)+\frac{1}{2}b(u)b'(u)=a(u)\ . \tag{79}$$

Therefore, we have from (75),

$$f(W,Y) \;=\; \left[a\left(u(W,Y)\right)-\frac{1}{2}\cdot b\left(u(W,Y)\right)\cdot b'\left(u(W,Y)\right)\right]\cdot \mathrm{e}^{-\int\limits_0^W b'(u(z,Y))\,\mathrm{d}z} \tag{80}$$

aid the numerical scheme which is now solely governed by

$$\mathrm{d}Y \;=\; f(W,Y)\,\mathrm{d}t \tag{81}$$

and the state variable X at any point in time is given by

$$X(t) \;=\; u\left(W(t),Y(t)\right)\ . \tag{82}$$

Naturally, we must have $X(0) \;=\; u\left(W(0),Y(0)\right) \;=\; u\left(0,Y(0)\right) \;=\; Y(0)$.

Example:

$$\mathrm{d}X = \kappa(\theta - X)\mathrm{d}t + \sigma X \mathrm{d}W \tag{83}$$

Assuming

$$\kappa \geq 0\ , \quad \theta \geq 0\ , \quad \sigma \geq 0\ , \quad \text{and} \quad X(0) \geq 0\ ,$$

we must have

$$X(t) \geq 0 \qquad \text{for all} \qquad t > 0\ .$$

Now for equation (75):

$$\frac{\partial u(W,Y)}{\partial W} = \sigma u \quad \text{and thus} \quad u(W,Y) = Y\mathrm{e}^{\sigma W}\ . \tag{84}$$

This means

$$\mathrm{d}Y \ = \ \left[\kappa\theta\mathrm{e}^{-\sigma W} - \left(\kappa + \tfrac{1}{2}\sigma^2\right)Y\right]\mathrm{d}t \tag{85}$$

We cannot solve this equation directly.

Also, a directly applied explicit Euler scheme would permit Y to cross over to the negative half of the real axis and thus $X = u(W,Y) = Y\mathrm{e}^{\sigma W}$ would leave the domain of (83).

An explicit Euler scheme applied to equation (85) would mean that, within the scheme, we interpret the $W(t)$ as a piecewise constant function.

We can do better than that!

Recall that, for the given time discretisation, we explicitly construct the Wiener process values $W(t_i)$ and thus, for the purpose of numerical integration of equation (83), they are known along any one given path.

If we now approximate $W(t)$ as a piecewise linear function in between the known values at t_n and t_{n+1}, i.e.

$$W(t) \;\simeq\; \alpha_n + \beta_n t \quad \text{for} \quad t \in [t_n, t_{n+1}] \tag{86}$$

with

$$\alpha_n \;=\; W(t_n) - \beta_n t_n \qquad \text{and} \qquad \beta_n \;=\; \frac{W(t_n) - W(t_{n+1})}{t_n - t_{n+1}} \;,$$

then we have the approximating ordinary differential equation

$$\mathrm{d}\hat{Y} \;=\; \left[\kappa\theta \mathrm{e}^{-\sigma(\alpha_n+\beta_n t)} - \left(\kappa + \tfrac{1}{2}\sigma^2\right)\hat{Y}\right]\mathrm{d}t \;. \tag{87}$$

Using the abbreviations

$$\delta_n := \kappa + \tfrac{1}{2}\sigma^2 - \sigma\beta_n \;, \qquad \Delta t_n := t_{n+1} - t_n \;, \qquad \text{and} \qquad W_{n+1} := W(t_{n+1})$$

we can write the solution to equation (87) as

$$\hat{Y}_{n+1} = \hat{Y}_n \mathrm{e}^{-\left(\kappa+\frac{1}{2}\sigma^2\right)\Delta t_n} + \kappa\theta \cdot \mathrm{e}^{-\sigma W_{n+1}} \cdot \left(\frac{1-\mathrm{e}^{-\delta_n \Delta t_n}}{\delta_n}\right) \;. \tag{88}$$

This scheme is unconditionally stable!

Note: The ordinary differential equation of Doss's pathwise solution can also be obtained by looking for the transformation $Y = Y(W, X)$ such that the stochastic differential equation for Y resulting from Itô's lemma degenerates into an ordinary differential equation.

For the example: apply Itô's lemma to $Y = X\mathrm{e}^{-\sigma W}$:

$$
\begin{aligned}
\mathrm{d}Y &= \frac{\partial Y}{\partial X}\mathrm{d}X + \frac{\partial Y}{\partial W}\mathrm{d}W \\
&\quad + \frac{1}{2}\frac{\partial^2 Y}{\partial X^2}\sigma^2 X^2 \mathrm{d}t + \frac{1}{2}\frac{\partial^2 Y}{\partial W^2}\mathrm{d}t + \frac{\partial^2 Y}{\partial W \partial X}\sigma X \mathrm{d}t \\
&= \mathrm{e}^{-\sigma W} \cdot [\kappa(\theta - X)\mathrm{d}t + \sigma X\,\mathrm{d}W] - \sigma Y\,\mathrm{d}W \\
&\quad + \frac{1}{2}\sigma^2 Y\,\mathrm{d}t - \sigma \mathrm{e}^{-\sigma W}\sigma X \mathrm{d}t \\
&= \left[\kappa\theta\mathrm{e}^{-\sigma W} - \left(\kappa + \tfrac{1}{2}\sigma^2\right) Y\right]\mathrm{d}t
\end{aligned}
$$

Alternative stable schemes for the numerical solution of equation (85):-

Implicit.

$$
\begin{aligned}
Y_{n+1} - Y_n &= \left[\kappa\theta \mathrm{e}^{-\sigma W_{n+1}} - \left(\kappa + \tfrac{1}{2}\sigma^2\right) Y_{n+1}\right] \Delta t_n \\
Y_{n+1} &= \frac{Y_n + \kappa\theta \mathrm{e}^{-\sigma W_{n+1}} \Delta t_n}{1 + \left(\kappa + \frac{1}{2}\sigma^2\right) \Delta t_n}
\end{aligned}
$$

Mixed.

$$
\begin{aligned}
Y_{n+1} - Y_n &= \left[\kappa\theta \mathrm{e}^{-\sigma \frac{1}{2}(W_n + W_{n+1})} - \left(\kappa + \tfrac{1}{2}\sigma^2\right) Y_{n+1}\right] \Delta t_n \\
Y_{n+1} &= \frac{Y_n + \kappa\theta \mathrm{e}^{-\sigma \frac{1}{2}(W_n + W_{n+1})} \Delta t_n}{1 + \left(\kappa + \frac{1}{2}\sigma^2\right) \Delta t_n}
\end{aligned}
$$

XIII. Variance reduction techniques

Reducing the variance of the Monte Carlo estimator (which is a variate itself), improves the Monte Carlo convergence behaviour.

This is why Monte Carlo convergence enhancements are referred to as *variance reduction techniques*.

- A good method to construct paths weighting the dimensions of relevance by their importance is one of the most important variance reduction methods.

- Importance sampling, naturally arising from a fortunate choice of sampler density, is another powerful variance reduction method.

Antithetic sampling for Gaussian variates

For each drawn Gaussian variate vector $\boldsymbol{z}$, reevaluate the functional $v(\boldsymbol{z})$ with $-\boldsymbol{z}$, i.e. compute $v(-\boldsymbol{z})$.

Since $\boldsymbol{z}$ and $-\boldsymbol{z}$ have equal probability, this method automatically matches the first moment.

Only count the pairwise average $\bar{v}_i = \frac{1}{2}(v(\boldsymbol{z}_i) + v(-\boldsymbol{z}_i))$ as an individual sample, because the pairwise averages $\bar{v}_i$ are independent!

This method improves convergence whenever $v(\boldsymbol{z})$ is monotonic in $\boldsymbol{z}$. When v is symmetric in $\boldsymbol{z}$, it decreases performance.

Antithetic sampling is good for:

- one sided payoff functionals

Antithetic sampling is bad for:

- double sided payoff functionals (double knock-outs, range accruals)
- payoff functionals with pronounced discretely symmetric features

Note: Sobol' numbers, whenever the number of iterations is a Mersenne number, i.e. $2^n - 1$ for some integer n, unlike pseudo-random numbers, have the antithetic feature (approximately) built into them.

$\Rightarrow$ **Do not use this method with low-discrepancy numbers!**

Matching the second moment

Moment matching used to be a very popular method before efficient and reliable low-discrepancy numbers became available. This method does usually give more accurate results for calculations that use pseudo-random numbers.

Matching the second moment is not guaranteed to improve convergence.

Assume a Monte Carlo simulation is to be carried out using a total of N variate vectors $\boldsymbol{v}$ of dimensionality d of a known joint distribution density $\psi(\boldsymbol{v})$.

Then, we can calculate the moments actually realised by the drawn variate vector set $V := \{v_{ij}\}$ with $i = 1..N$ and $j = 1..d$. The first moment for dimension j is given by

$$\langle \boldsymbol{v} \rangle_j = \frac{1}{N} \sum_{i=1}^{N} v_{ij} \ , \quad j = 1..d \ . \tag{89}$$

Using (89), we can construct a set of first-moment-corrected variates $\tilde{V}$ by subtraction of the average in each dimension, i.e.

$$\tilde{v}_{ij} = v_{ij} - \langle \boldsymbol{v} \rangle_j \ . \tag{90}$$

The realised covariance of the mean-corrected variate set can be concisely represented as

$$\tilde{C} = \tilde{V}^\top \tilde{V} \tag{91}$$

if we view $\tilde{V}$ as a matrix whose rows comprise the individual d-dimensional mean-corrected vector draws.

We can construct a new matrix $\hat{V}$ whose entries will meet the desired covariance C of the target distribution density ψ exactly.

Define the elements of the desired covariance matrix C as

$$c_{jk} = \int v_j v_k \psi(\boldsymbol{v}) \mathrm{d}v_j \mathrm{d}v_k \ . \tag{92}$$

Also, define the pseudo-square roots of both C and $\tilde{C}$ by

$$\tilde{C} = \tilde{A} \cdot \tilde{A}^\top \quad \text{and} \quad C = A \cdot A^\top \ . \tag{93}$$

The correction matrix K that transforms $\tilde{V}$ to $\hat{V}$ can be computed by solving the linear system

$$\tilde{A}^\top \cdot K = A^\top \ , \quad \text{i.e.} \quad K = \tilde{A}^{\top\,-1} \cdot A^\top \ . \tag{94}$$

The moment-corrected variate matrix is thus

$$\hat{V} = K \cdot \tilde{V} \ . \tag{95}$$

Since $\tilde{C}$ may (very rightfully) be rank-deficient, you must always use a failsafe method for the solution of the linear system (94) such as the Moore-Penrose pseudo-inverse.

Notes:

- When using this method to correct the first and the second moment of a set of drawn variates it should be applied to the variates *after* having transformed them from the uniform $(0, 1)$ distribution to whatever distribution is actually used, e.g. a joint normal distribution.

- For anything but comparatively low-dimensional methods, the calculation of the covariance matrix and the subsequent correction of each drawn vector variate can very easily dominate the overall computational effort *by several orders of magnitude*!

- Sobol' numbers (approximately) match the second moment (as well as the first moment), when the number of iterations is a Mersenne number.

- **Do not use this method with low-discrepancy numbers!**

Control variates

Many Monte Carlo calculations are carried out for problems that we can almost solve analytically.

The idea behind *control variates* is as follows:

Let's assume that we wish to calculate the expectation $\mathsf{E}[v]$ of a function $v(\boldsymbol{u})$ for some underlying vector draw $\boldsymbol{u}$, and that there is a related function $g(\boldsymbol{u})$ whose expectation $g^* := \mathsf{E}[g]$ we know exactly. Then, we have

$$\mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} v(\boldsymbol{u}_i)\right] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} v(\boldsymbol{u}_i) + \beta\left(g^* - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{u}_i)\right)\right] \tag{96}$$

for any given $\beta \in \mathbb{R}$ and thus we can replace the ordinary Monte Carlo estimator

$$\hat{v} = \frac{1}{n}\sum_{i=1}^{n} v(\boldsymbol{u}_i) \tag{97}$$

by

$$\hat{v}_{\mathsf{CV}} = \frac{1}{n}\sum_{i=1}^{n}\left[v(\boldsymbol{u}_i) + \beta\left(g^* - g(\boldsymbol{u}_i)\right)\right] \ . \tag{98}$$

The optimal choice of β is

$$\beta^* = \frac{\mathsf{Cov}[v,g]}{\mathsf{V}[g]} \tag{99}$$

which minimises the variance of of $\hat{v}_{\mathsf{CV}}$.

Note that the function $g(\boldsymbol{u}_i)$ does not have to be the payoff of an analytically known option. It could also be the profit from a self-financing dynamic hedging strategy, i.e. a strategy that starts with zero investment capital. For risk-neutral measures, the expected profit from any such strategy is zero which means that the control variate is simply the payoff from the dynamic hedging strategy along any one path.

An intuitive understanding of the control variate method is to consider the case when v and g are positively correlated. For any draw $v(\boldsymbol{u}_i)$ that overestimates the result, $g(\boldsymbol{u}_i)$ is likely to overestimate g^*. As a result, the term multiplied by β in equation (98) is likely to correct the result by subtracting the aberration.

The precise value of β^* is, of course, not known but can be estimated from the same simulation that is used to calculate $\hat{v}_{\text{CV}}$. As in all the situations when the parameters determining the result are calculated from the same simulation, this can introduce a bias that is difficult to estimate. In the limit of very large numbers of iterations, this bias vanishes, but the whole point of variance reduction techniques is to require *fewer* simulations and thus shorter run time. A remedy to the problem of bias due to a correlated estimate of the control parameter β is to use an initial simulation, possibly with fewer iterates than the main run, to estimate β^* in isolation.

Key point: **Choose a good control variate specific to each problem. A (very) bad control variate can make matters worse!**

XIV. Sensitivity calculations

A Monte Carlo estimator, invariably, is not just a function of the payoff-specific parameters, but also of the initial values of the state variables and the model parameters

$$\hat{v} = \hat{v}(\boldsymbol{x}(0), \boldsymbol{\lambda}) \ . \tag{100}$$

The partial derivatives with respect to the elements of $\boldsymbol{x}(0)$ and $\boldsymbol{\lambda}$ are known as *Greeks* and are one of the biggest problems of Monte Carlo simulations.

Finite differencing with path recycling

We can always re-do the entire valuation with varied inputs reflecting the potential change in the underlying asset, and use an explicit finite-differencing approach to compute the Greek we are after.

Forward differencing for Delta:

$$\text{Delta} = \frac{\partial v}{\partial S_0} \approx \frac{v(S_0 + \Delta S_0) - v(S_0)}{\Delta S_0} \ . \tag{101}$$

Centre differencing for Delta:

$$\text{Delta} = \frac{\partial v}{\partial S_0} \approx \frac{v(S_0 + \Delta S_0) - v(S_0 - \Delta S_0)}{2\Delta S_0} \,. \tag{102}$$

Using the centre-differencing approach in equation (102) has the added advantage that we can then directly approximate Gamma as

$$\text{Gamma} = \frac{\partial^2 v}{\partial S_0^2} \approx \frac{v(S_0 + \Delta S_0) - 2v(S_0) + v(S_0 - \Delta S_0)}{\Delta S_0^2} \,. \tag{103}$$

Note: $\frac{\Delta S_0}{S_0}$ should scale like the fourth root of your machine precision!

For Gamma of a plain vanilla option, this means we are trying to compute the value of an approximation to the Dirac-spike as a payoff!

Monte Carlo Greeks with conventional finite differencing can be prohibitively noisy!

Importance sampling can help!

Pathwise differentiation

All occurrences of the true price v in equation (102) are numerically evaluated as a Monte Carlo approximation to it. Thus,

$$\widehat{\text{Delta}} = \frac{\partial \hat{v}}{\partial S_0} = \frac{\partial}{\partial S_0}\left[\frac{1}{m}\sum_{j=1}^{m}\pi\Big(\boldsymbol{S}(\boldsymbol{z}^j;S_0)\Big)\right] \quad . \tag{104}$$

An infinitesimal change of the initial spot level S_0 can only give rise to infinitesimal changes of the spot level at any of the monitoring dates.

For Lipschitz-continuous payoff functions, the order of differentiation and expectation can be interchanged. For such payoff functions, it is perfectly consistent to assign a Delta of zero to all paths that terminate out of the money, and a Delta equal to $\sum_i \frac{\partial \pi\left(\boldsymbol{S}(\boldsymbol{z}^j;S_0)\right)}{\partial S_i} \cdot \frac{\partial S_i}{\partial S_0}$.

This method is called *pathwise differentiation* and can easily be transferred to other Greeks such as Vega. However, for the calculation of Gamma, we still have to implement a finite differencing scheme for two individually calculated

Deltas for an up and a down shift, and both of these can individually be computed using pathwise differentiation.

Alas, this method does not apply (easily) to discontinuous payoff functionals, of which there are so many.

The likelihood ratio method

The option pricing problem by Monte Carlo is a numerical approximation to an integration:

$$v = \int \pi(S)\ \psi(S)\,\mathrm{d}S \tag{105}$$

Numerically, we construct evolutions of the underlying assets represented by S given a risk-neutral distribution $\psi(S)$.

We hereby typically construct the paths by the aid of a set of standard normal variates which corresponds to

$$v = \int \pi(S(z;\alpha))\ \varphi(z)\,\mathrm{d}z\ , \tag{106}$$

and all dependence on further pricing parameters (herein represented by α) such as the spot level at inception, volatility, time to maturity, etc., is absorbed into the path construction $S(z;\alpha)$.

Any derivative with respect to any of the parameters will thus suffer from any discontinuities of π in S:

$$\frac{\partial v}{\partial \alpha} = \int \frac{\partial}{\partial \alpha}\pi(S(z;\alpha))\ \varphi(z)\,\mathrm{d}z\ . \tag{107}$$

The key insight behind the *likelihood ratio* method is to shift the dependence on any of the parameters over into the density function. In other words, a transformation of the density is required to look at the pricing problem in the form of equation (105). This way, the Greek evaluation problem becomes

$$\frac{\partial v}{\partial \alpha} = \int \pi(S)\ \frac{\partial}{\partial \alpha}\psi(S;\alpha)\,\mathrm{d}S = \int \pi(S)\ \frac{\frac{\partial \psi(S;\alpha)}{\partial \alpha}}{\psi(S;\alpha)}\ \psi(S;\alpha)\,\mathrm{d}S\ . \tag{108}$$

The calculation of the desired Greek now looks exactly like the original pricing problem, only with a new payoff function

$$\chi(S;\alpha) := \pi(S) \cdot \omega(S;\alpha) \tag{109}$$

with

$$\omega(S;\alpha) := \frac{\frac{\partial\psi(S;\alpha)}{\partial\alpha}}{\psi(S;\alpha)} . \tag{110}$$

The term $\omega(S;\alpha)$ may be interpreted as a *likelihood ratio* since it is the quotient of two density functions, whence the name of the method. Using this definition, the Greek calculation becomes

$$\frac{\partial v}{\partial \alpha} = \int \chi(S;\alpha)\,\psi(S;\alpha)\,\mathrm{d}S . \tag{111}$$

The beauty of this idea is that for the probability density functions that we typically use such as the one corresponding to geometric Brownian motion, the function $\chi(S;\alpha)$ is $\in \mathcal{C}^\infty$ in the parameter α and thus doesn't cause the trouble that we have when we approach the Greek calculation problem in the form of equation (106).

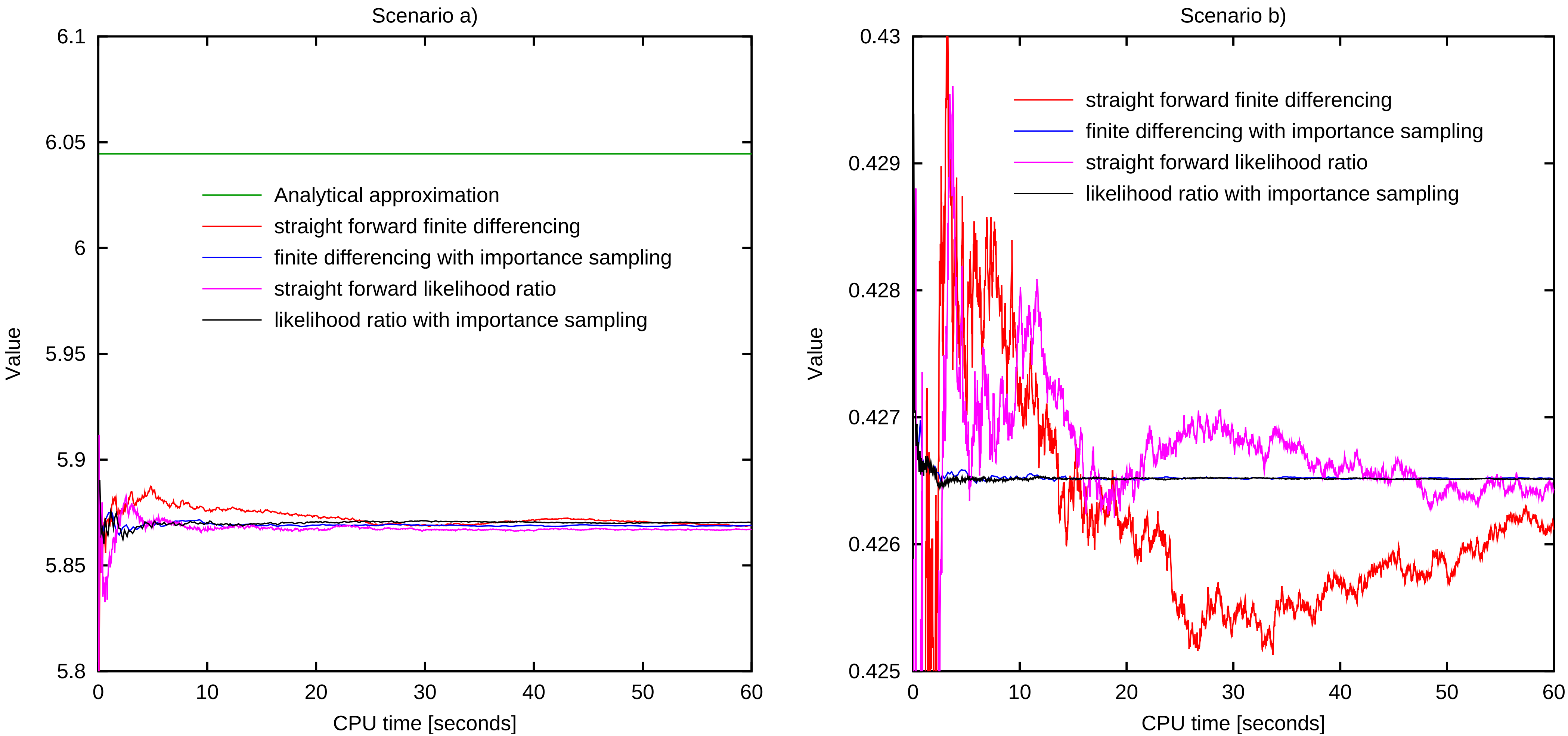
The application is now straightforward. Alongside the calculation of the option price, for each constructed path, apart from calculating the payoff $\pi(S)$, also calculate the likelihood ratio $\omega(S;\alpha)$. The approximation for Delta, for instance, thus becomes

$$\widehat{\text{Delta}} = \frac{1}{m}\sum_{j=1}^{m} \pi\big(\boldsymbol{S}^j; S_0)\big)\ \omega\left(\boldsymbol{S}^j; S_0)\right) \ . \tag{112}$$

For geometric Brownian motion:

$$\omega_{\widehat{\text{Delta}}} = \frac{z_1}{S_0\sigma\sqrt{\Delta t_1}} \tag{113}$$

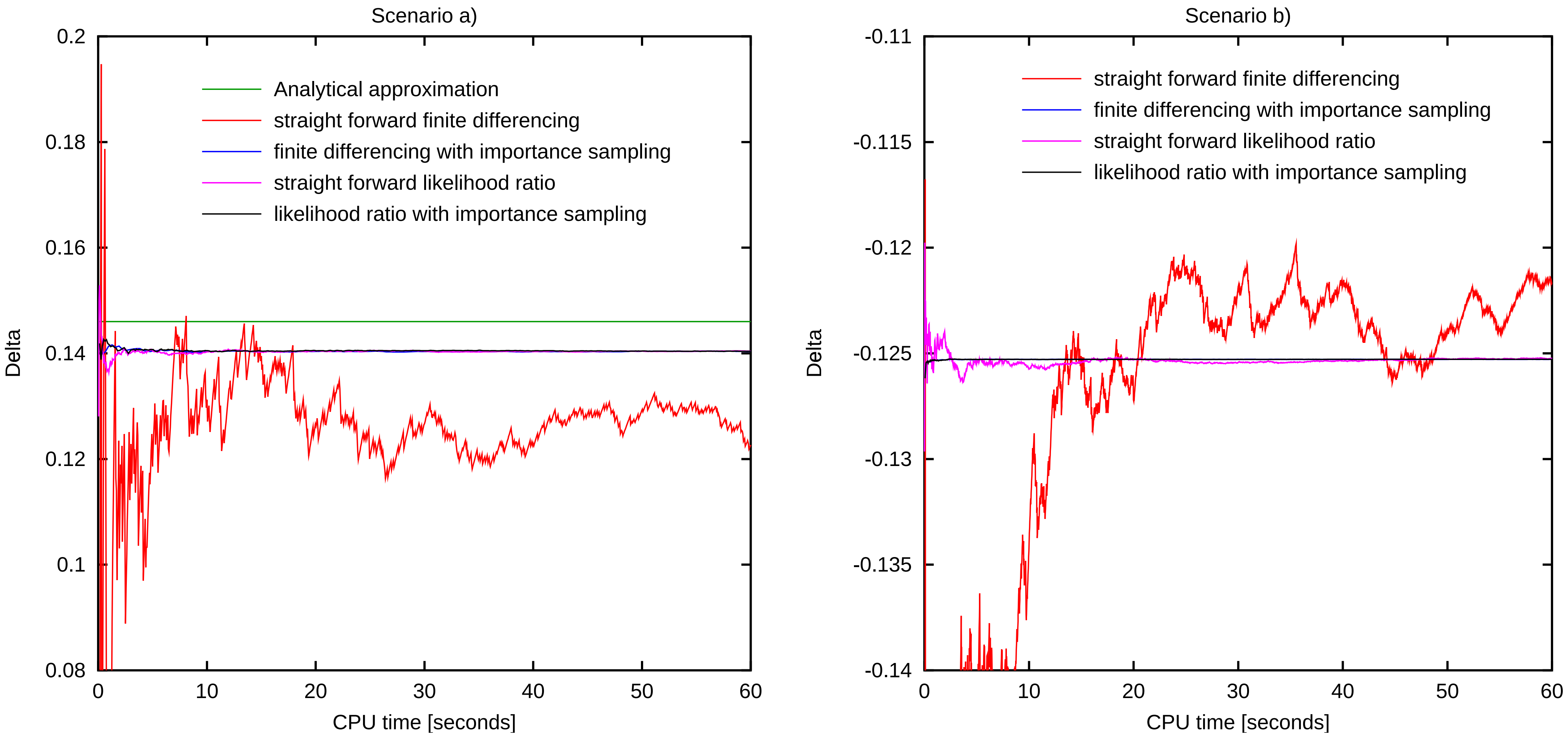
$$\omega_{\widehat{\text{Gamma}}} = \frac{z_1^2 - z_1\sigma\sqrt{\Delta t_1} - 1}{S_0^2\sigma^2\Delta t_1} \tag{114}$$



The value of two Up-Out-Call options.

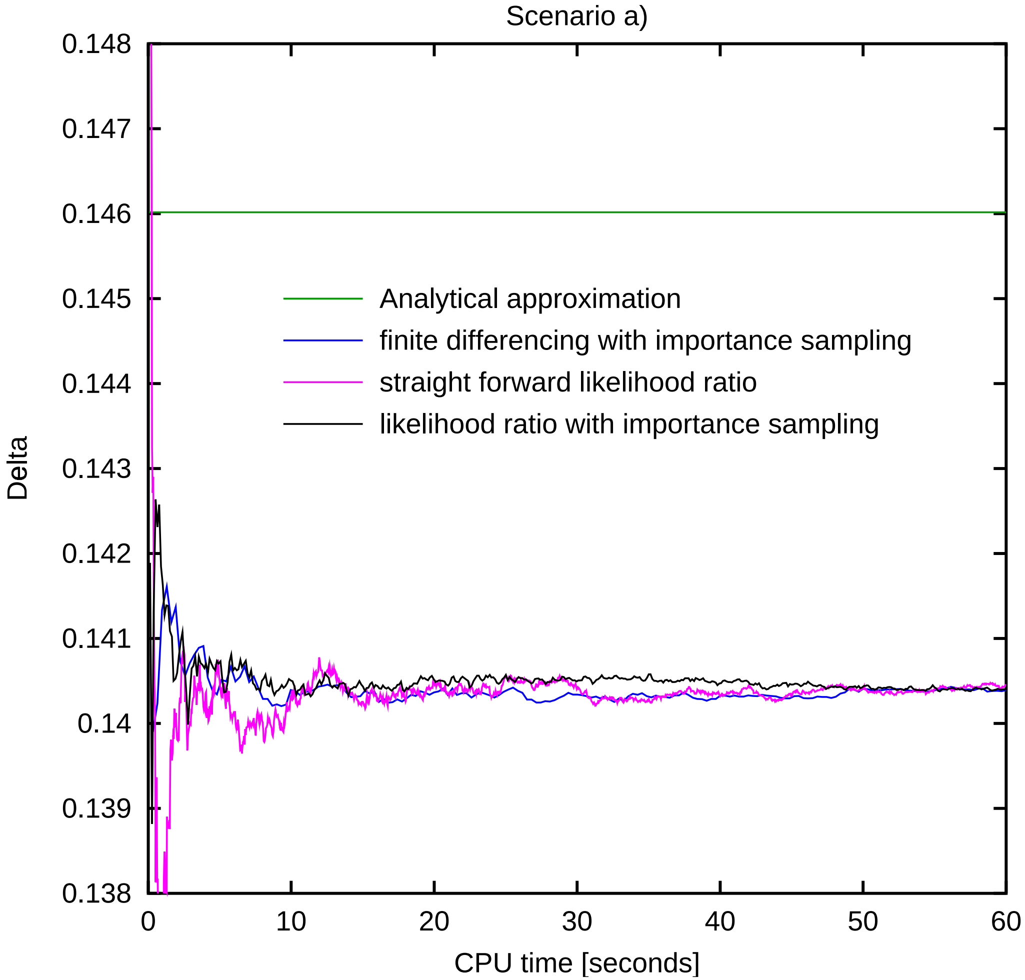
Scenario a): $T = 1$, $S = 100$, $K = 100$, $H = 150$, $\sigma = 30\%$, $t_i = \frac{i}{12}$.

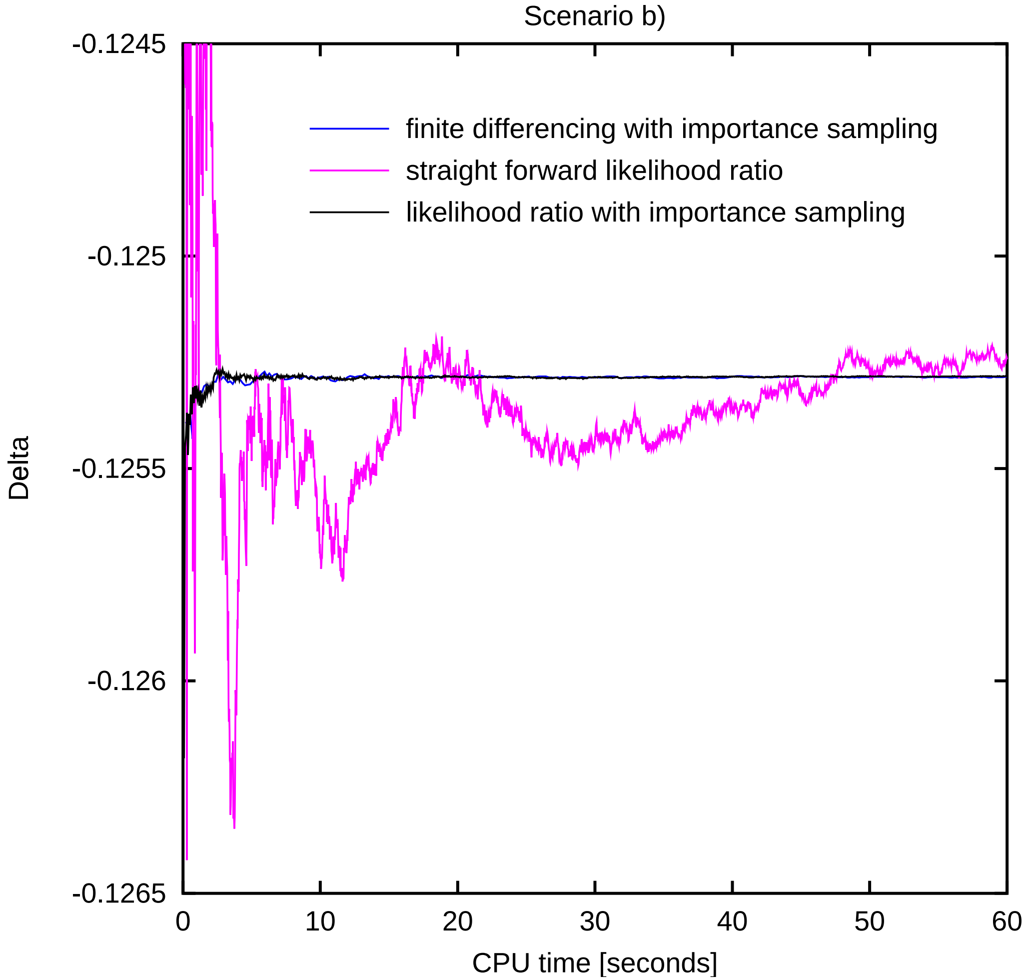
Scenario b): $T = 0.52$, $S = 160$, $K = 100$, $H = 150$, $\sigma = 30\%$, $t_1 = \frac{5}{250}$, $t_i = \frac{i}{12} + \frac{5}{250}$ for $i = 2, 3, 4, 5, 6, 7$.



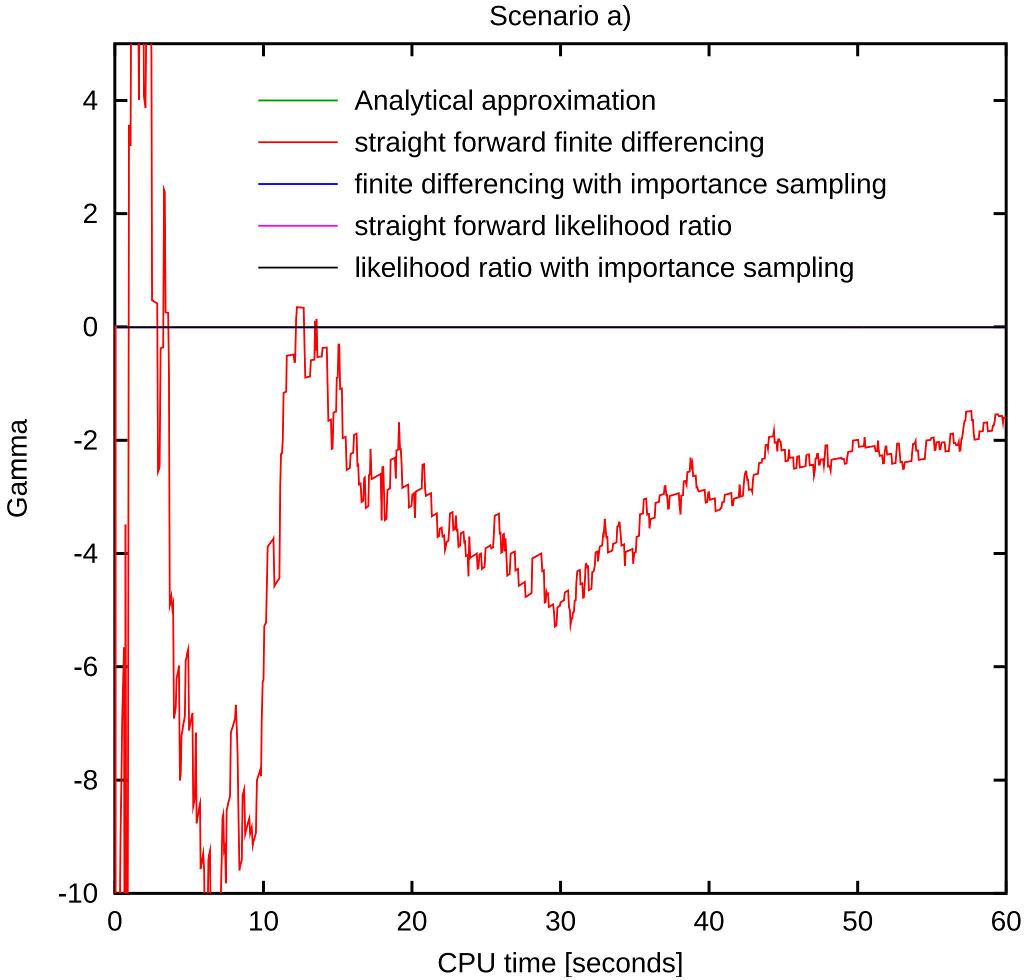
Delta of two Up-Out-Call option.

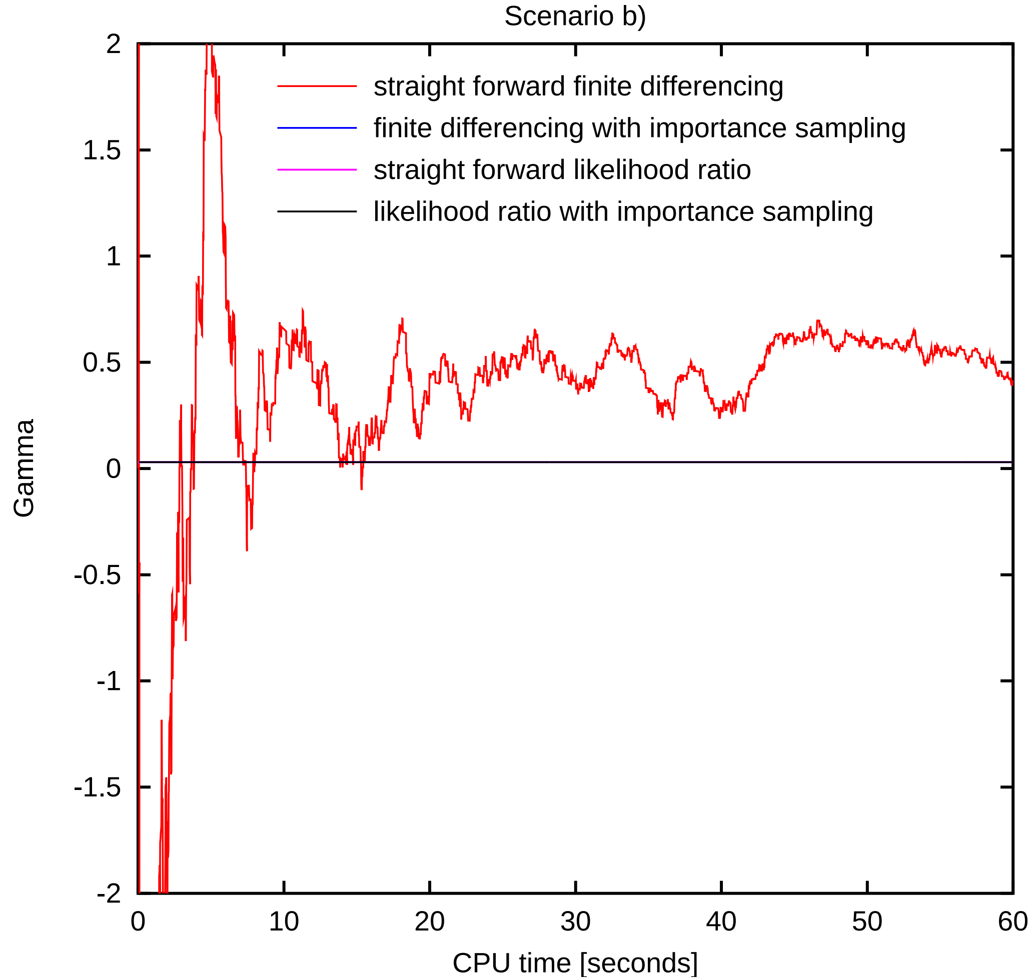
Hardware: AMD K6-III processor running at 400MHz.



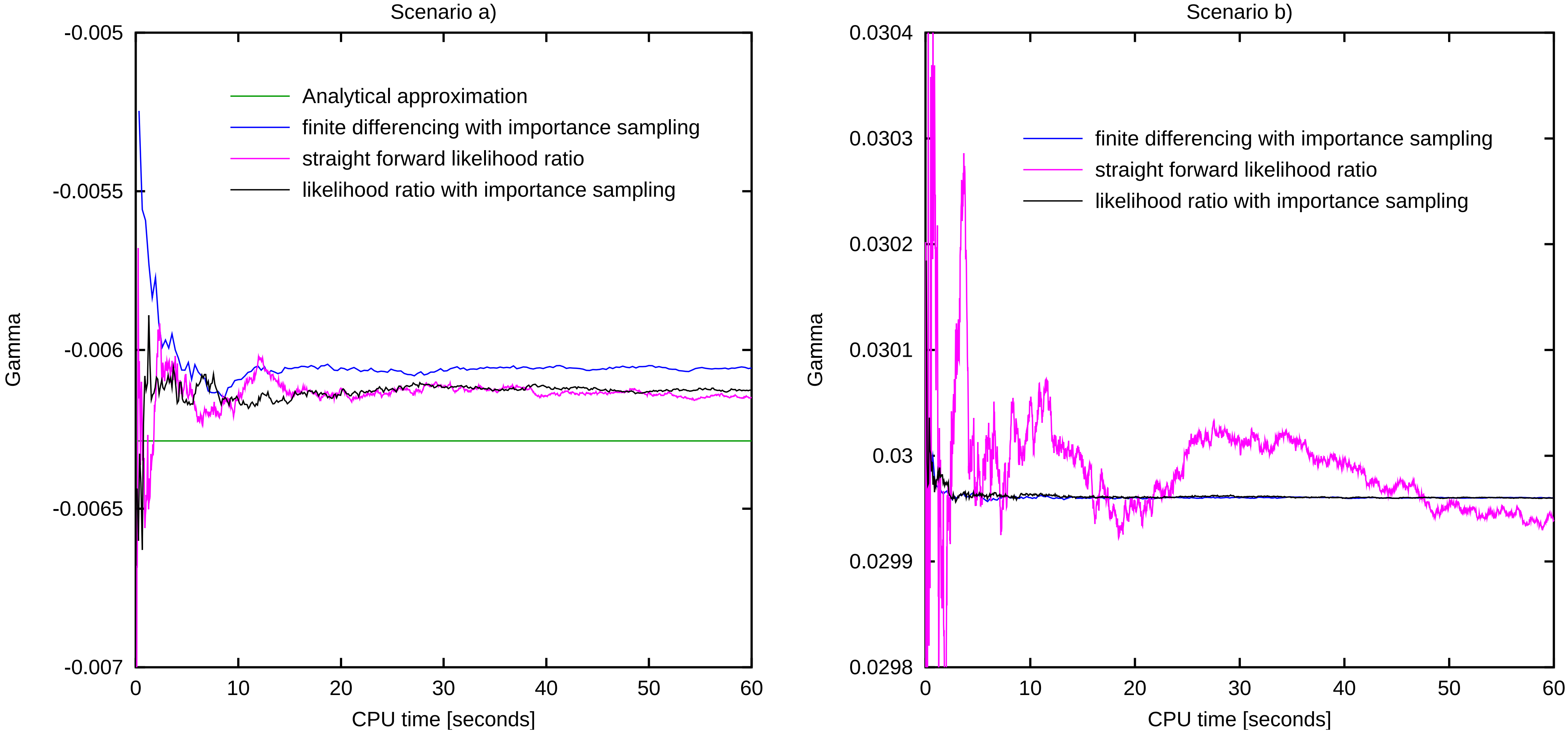


Enlargement for Delta.





Gamma of two Up-Out-Call options.



Enlargements for Gamma of the two Up-Out-Call options.

XV. Weighted Monte Carlo

These methods are also known as under the names *equivalent entropy projection methods* and *perturbation of measures* and are due to Avellaneda and Gamba [AG02].

A standard Monte Carlo estimator over N paths is given by

$$\langle v \rangle_N = \sum_{i=1}^{N} \frac{1}{N} v_i \ . \tag{115}$$

where v_i is the net present value of all cashflows (numéraire dominated payoffs).

The concept of *Weighted Monte Carlo* is to generalise this to

$$\widetilde{\langle v \rangle}_N = \sum_{i=1}^{N} p_i v_i = \boldsymbol{p}^\top \cdot \boldsymbol{v} \ . \tag{116}$$

Define the value of the M calibration instruments as g_j^* for $j = 1..M$, and the vector of these values as $\boldsymbol{g}^*$.

Define the aggregated net present value of all cashflows occuring in path # i for the calibration instrument # j as $g_{ij} = (G)_{ij}$.

In order to have stable hedge ratios, we demand a perfect match of the market values of the calibration instruments:

$$\boldsymbol{g}^{*\top} \;=\; \boldsymbol{p}^\top \cdot G \tag{117}$$

Another constraint is

$$\sum_{i=1}^{N} p_i \;=\; 1\;. \tag{118}$$

<u>Since $M \ll N$, this alone does not define all of the p_i uniquely.</u>

One can remedy the ambiguity by the aid of a strictly convex objective function of $\boldsymbol{p}$ in combination with (117) and (118) using Lagrange multipliers.

Avellaneda and Gamba suggest

$$\sum_{i=1}^{N}\Psi(p_i)-\sum_{j=1}^{M}\lambda_j\cdot\left[\sum_{i=1}^{N}p_i g_{ij}-g_j^*\right]-\mu\cdot\left[\sum_{i=1}^{N}p_i-1\right] \tag{119}$$

with $\Psi(x)$ a convex function. Specifically, we will consider the following two choices:-

1. The Euclidean distance from equiprobability given by the quadratic entropy function
$$\Psi^{\text{(QE)}}(p)=(p-1/N)^2\ . \tag{120}$$

2. The Shannon entropy [Sha48] function (which is identical to the Kullback-Leibler relative entropy function against a uniform distribution [KL51])
$$\Psi^{\text{(KL)}}(p)\ =\ p\ln p\ . \tag{121}$$

Differentiation of (119),

$$\nabla_{\boldsymbol{p}} \cdot \left[\sum_{i=1}^{N} \Psi(p_i) - \left(\boldsymbol{p}^\top \cdot G - \boldsymbol{g}^{*\top} \right) \cdot \boldsymbol{\lambda} - \mu \cdot \left(\sum_{i=1}^{N} p_i - 1 \right) \right] \;=\; 0 \qquad (122)$$

yields that the minimum must satisfy

$$p_i \;=\; \phi\left(\sum_{j=1}^{M} g_{ij}\lambda_j + \mu \right) \qquad \text{with} \qquad \phi(x) \;:=\; \Psi'^{\,-1}(x) \qquad (123)$$

which, for the specific choices of (120) and (121), means

$$p_i^{(\mathrm{QE})} \;=\; \frac{1}{N} + \frac{1}{2}\left(\sum_{j=1}^{M} g_{ij}\lambda_j^{(\mathrm{QE})} + \mu \right) \qquad (124)$$

and

$$p_i^{(\mathrm{KL})} \;=\; \mathrm{e}^{\sum_{j=1}^{M} g_{ij}\lambda_j^{(\mathrm{KL})} + \mu - 1} \;. \qquad (125)$$

Combining this with the condition $\sum_{i=1}^{N} p_i \;=\; 1$, we obtain

$$p_i^{\text{(QE)}} \;=\; \frac{1}{N} + \frac{1}{2}\sum_{j=1}^{M}\Big(g_{ij} - \langle g_j \rangle_N\Big) \cdot \lambda_j^{\text{(QE)}} \tag{126}$$

and

$$p_i^{\text{(KL)}} \;=\; \frac{\mathrm{e}^{\sum_{j=1}^{M} g_{ij}\lambda_j^{\text{(KL)}}}}{Z(\boldsymbol{\lambda}^{\text{(KL)}})} \tag{127}$$

with

$$\langle g_j \rangle_N \;:=\; \frac{1}{N}\sum_{i=1}^{N} g_{ij} \qquad \text{and} \qquad Z(\boldsymbol{\lambda}^{\text{(KL)}}) \;:=\; \sum_{i=1}^{N} \mathrm{e}^{\sum_{j=1}^{M} g_{ij}\lambda_j^{\text{(KL)}}} \tag{128}$$

The key is now to find the values for $\boldsymbol{\lambda}$ such that (117) and (118), i.e.

$$g_j^* \;=\; \widetilde{\langle g_j \rangle}_N \;=\; \textstyle\sum_{i=1}^{N} p_i g_{ij} \qquad \text{and} \qquad \textstyle\sum_{i=1}^{N} p_i \;=\; 1\;,$$

are satisfied.

For the quadratic penalty function (120), this means we have to solve the linear system

$$\underbrace{\left(\langle \boldsymbol{g} \cdot \boldsymbol{g}^\top \rangle_N - \langle \boldsymbol{g} \rangle_N \cdot \langle \boldsymbol{g} \rangle_N^\top\right)}_{\text{sample covariance}} \cdot \boldsymbol{\lambda}^{\text{(QE)}} = \frac{2}{N} \left(\boldsymbol{g}^* - \langle \boldsymbol{g} \rangle_N\right) . \tag{129}$$

Substituting this back into (126) and (116) yields

$$p_i^{\text{(QE)}} = \frac{1}{N} \cdot \left(1 + \textstyle\sum_{j=1}^{M} \left(g_{ij} - \langle g_j \rangle_N\right) \cdot \gamma_j\right) \quad \text{with} \quad \boldsymbol{\gamma} := \langle \boldsymbol{g}, \boldsymbol{g}^\top \rangle_N^{-1} \cdot \left(\boldsymbol{g}^* - \langle \boldsymbol{g} \rangle_N\right) \tag{130}$$

and

$$\widetilde{\langle v \rangle}_N^{\text{(QE)}} = \langle v \rangle_N + \langle v, \boldsymbol{g}^\top \rangle_N \cdot \langle \boldsymbol{g}, \boldsymbol{g}^\top \rangle_N^{-1} \cdot \left(\boldsymbol{g}^* - \langle \boldsymbol{g} \rangle_N\right) \tag{131}$$

where $\langle \cdot, \cdot \rangle_N$ stands for the sample covariance.

The choice of a quadratic entropy function
centered at the equiprobability level
is equivalent to the use of the calibration instruments
as conventional control variates.

For the Kullback-Leibler entropy function (121), Avellaneda and Gamba noticed that

$$\sum_{i=1}^{N} p_i^{(\mathsf{KL})} g_{ij} \;=\; \frac{\partial \ln Z(\boldsymbol{\lambda}^{(\mathsf{KL})})}{\partial \lambda_j^{(\mathsf{KL})}} \;. \tag{132}$$

Thus, instead of solving (117) directly, i.e.

$$\boldsymbol{g}^* \;=\; \boldsymbol{p}^{(\mathsf{KL})\,\top} \cdot G \;=\; \nabla_{\boldsymbol{\lambda}^{(\mathsf{KL})}} \cdot \ln Z(\boldsymbol{\lambda}^{(\mathsf{KL})}) \;, \tag{133}$$

one may minimise

$$\ln Z(\boldsymbol{\lambda}^{(\mathsf{KL})}) - \sum_{j=1}^{M} g_j^* \lambda_j^{(\mathsf{KL})} \;. \tag{134}$$

This can be done, for instance, using the Broydn-Fletcher-Goldfarb-Shanno algorithm [PTVF92]. A second order expansion in $\boldsymbol{\lambda}$ for $\ln Z(\boldsymbol{\lambda}^{(\mathsf{KL})})$ gives

$$\boldsymbol{\lambda}^{(\mathsf{KL})} \;\simeq\; N/2 \cdot \boldsymbol{\lambda}^{(\mathsf{QE})} \tag{135}$$

which can be used as the initial guess for the iteration.

The connection between the Monte Carlo estimator of the exotic derivative $\widetilde{\langle v \rangle}_N$ and the calibration instruments $\boldsymbol{g}^*$ is given by the fact that the discrete distribution encoded in the probability vector $\boldsymbol{p}$ is a function of the parameter vector $\boldsymbol{\lambda}$, i.e.

$$\boldsymbol{p} \;=\; \boldsymbol{p}(\boldsymbol{\lambda}) \qquad \text{with} \qquad \boldsymbol{p} \in \mathbb{R}^N, \boldsymbol{\lambda} \in \mathbb{R}^M \;, \tag{136}$$

and that the vector $\boldsymbol{\lambda}$ is chosen to match the calibration constraints (117), i.e.

$$\boldsymbol{\lambda} \;=\; \boldsymbol{\lambda}(\boldsymbol{g}^*) \qquad \text{with} \qquad \boldsymbol{g}^* \in \mathbb{R}^M \;. \tag{137}$$

By virtue of parametric adjustments of the sample distribution $\boldsymbol{p}(\boldsymbol{\lambda})$, we can adjust for changes in the given calibration instruments.

Model calibration, in general, is given by the process of parametric adjustments of a model's probability measures, i.e. implied market variable distributions, to match the prices of calibration instruments.

This means, we can estimate the effect of changes in the values of calibration instruments on $\widetilde{\langle v \rangle}_N$ by calculating the changes in $\widetilde{\langle v \rangle}_N = \boldsymbol{p}^\top \cdot \boldsymbol{v}$ in response to parametric changes of the sample distribution:

$$\nabla_{\boldsymbol{g}^*} \cdot \widetilde{\langle v \rangle}_N \;=\; J \cdot \nabla_{\boldsymbol{\lambda}} \cdot \boldsymbol{p}^\top \cdot \boldsymbol{v} \quad \text{with} \quad J \;:=\; \left(\frac{(\partial \boldsymbol{\lambda})}{(\partial \boldsymbol{g}^*)} \right) \quad \text{such that} \quad J_{kl} = \partial_{g_k^*} \lambda_l \tag{138}$$

From (123), we have

$$\partial_{\lambda_j} p_k \;=\; \phi'_k \cdot \left(g_{kj} + \partial_{\lambda_j} \mu \right) \tag{139}$$

where we have set

$$\phi'_k := \phi' \left(\textstyle\sum_{j=1}^M g_{kj} \lambda_j + \mu \right) \;.$$

Differentiating the probability normalisation condition (118) with respect to λ_j gives us

$$0 \;=\; \sum_{k=1}^N \partial_{\lambda_j} p_k \;=\; \sum_{k=1}^N \phi'_k \cdot \left(g_{kj} + \partial_{\lambda_j} \mu \right) \;, \tag{140}$$

i.e.

$$\partial_{\lambda_j}\mu \;=\; -\frac{\sum_{k=1}^{N}\phi'_k g_{kj}}{\sum_{k=1}^{N}\phi'_k} \;=\; -\widehat{\langle g_j\rangle}_N \; . \tag{141}$$

where we have defined

$$s := \sum_{k=1}^{N}\phi'_k \;, \quad w_i := \phi'_i/s \;, \quad \text{and} \quad \widehat{\langle v\rangle}_N := \sum_{i=1}^{N} w_i v_i \; . \tag{142}$$

This enables us to write concisely

$$\partial_{\lambda_j}\widetilde{\langle v\rangle}_N \;=\; s\cdot\sum_{k=1}^{N}\Big(w_k g_{kj} v_k - w_k\widehat{\langle g_j\rangle}_N v_k\Big) \tag{143}$$

and thus

$$\nabla_{\boldsymbol{\lambda}}\cdot\widetilde{\langle v\rangle}_N \;=\; s\cdot\widehat{\langle \boldsymbol{g}, v\rangle}_N \tag{144}$$

with

$$\widehat{\langle a,b\rangle}_N := \widehat{\langle ab\rangle}_N - \widehat{\langle a\rangle}_N\widehat{\langle b\rangle}_N \; . \tag{145}$$

Since this holds for any derivative contract priced in this measure, it also holds for the calibration instruments:

$$\underbrace{\nabla_{\boldsymbol{g}^*} \cdot \widetilde{\langle \boldsymbol{g} \rangle}_N^\top}_{\text{identity matrix}} = J \cdot \nabla_{\boldsymbol{\lambda}} \cdot \widetilde{\langle \boldsymbol{g} \rangle}_N^\top = J \cdot s \cdot \widehat{\langle \boldsymbol{g}, \boldsymbol{g}^\top \rangle}_N \tag{146}$$

whence

$$J = \frac{1}{s} \widehat{\langle \boldsymbol{g}, \boldsymbol{g}^\top \rangle}_N^{-1} . \tag{147}$$

We finally arrive at the formula for the hedge positions in the calibration instruments:

$$\underline{\underline{\nabla_{\boldsymbol{g}^*} \cdot \widetilde{\langle v \rangle}_N = \widehat{\langle \boldsymbol{g}, \boldsymbol{g}^\top \rangle}_N^{-1} \cdot \widehat{\langle \boldsymbol{g}, v \rangle}_N}} \tag{148}$$

The vector of price-sensitivities is equal to the vector of regression coefficients (under the calibration-adjusted measure $\widehat{\langle \cdot \rangle}$) of the payoff of the target contract on the linear space generated by the payoff-vectors of the calibration instruments [AG02].

In the case of the quadratic entropy function, we have:

$$
\begin{aligned}
\Psi^{(\mathrm{QE})}(p) &= \left(p-\frac{1}{N}\right)^2 \\
\Psi'^{(\mathrm{QE})}(p) &= 2\left(p-\frac{1}{N}\right) \\
\phi^{(\mathrm{QE})}(y) &= \frac{y}{2}+\frac{1}{N} \\
\phi'^{(\mathrm{QE})}(y) &= \frac{1}{2} \\
s^{(\mathrm{QE})} &= \frac{N}{2} \\
w_k^{(\mathrm{QE})} &= \frac{1}{N}
\end{aligned}
$$

which means

$$\underline{\underline{\widehat{\langle\cdot\rangle}_N^{(\mathrm{QE})} = \langle\cdot\rangle_N}} \tag{149}$$

For the Shannon/Kullback-Leibler entropy:

$$\begin{aligned}\Psi^{(\mathsf{KL})}(p) &= p\ln p\\ \Psi'^{(\mathsf{KL})}(p) &= 1+\ln p\\ \phi^{(\mathsf{KL})}(y) &= \mathrm{e}^{y-1}\\ \phi'^{(\mathsf{KL})}(y) &= \phi^{(\mathsf{KL})}(y)\end{aligned}$$

Using equation (123), this means

$$\phi'^{(\mathsf{KL})}_k \;=\; \phi^{(\mathsf{KL})}\big(\textstyle\sum_{j=1}^{M} g_{kj}\lambda_j^{(\mathsf{KL})}+\mu^{(\mathsf{KL})}\big) \;=\; p_k^{(\mathsf{KL})}$$

and therefore

$$\begin{aligned}s^{(\mathsf{KL})} &= 1\\ w_k^{(\mathsf{KL})} &= p_k^{(\mathsf{KL})}\;,\end{aligned}$$

which ultimately results in

$$\widehat{\langle\cdot\rangle}_N^{(\mathrm{KL})} = \widetilde{\langle\cdot\rangle}_N^{(\mathrm{KL})} . \tag{150}$$

$\Rightarrow$

$$\nabla_{\boldsymbol{g}^*} \cdot \widetilde{\langle v\rangle}_N^{(\mathrm{KL})} = \widetilde{\langle \boldsymbol{g}, \boldsymbol{g}^\top\rangle}_N^{(\mathrm{KL})\,-1} \cdot \widetilde{\langle \boldsymbol{g}, v\rangle}_N^{(\mathrm{KL})} \tag{151}$$

Under the Shannon/Kullback-Leibler entropy, the hedge ratios are equal to the regression coefficients of the payoff of the target contract onto the payoff-vector of the calibration instruments under the pricing measure [AG02].

Note: reverse projection onto price changes with respect to observable market variables (such as spot levels) can be accomplished if the sensitivities of the calibration instruments with respect to the market variables are available by independent, e.g. analytic, means:

$$\partial_{S_j} \cdot = \left(\partial_{S_j} \boldsymbol{g}^{*\top}\right) \cdot \nabla_{\boldsymbol{g}^*} \cdot \tag{152}$$

The benefits of Weighted Monte Carlo schemes are:-

- Exact matching of calibration instruments
- Hedge ratios are computed as a *free by-product*.

The potential downsides are:-

- Negative probablities $\Rightarrow$ destruction of equivalent martingale measure property.
- Numerical fitting after N iterations $\Rightarrow$ convergence diagrams change meaning.

Question: does the system of SDEs

$$
\begin{aligned}
\mathrm{d}x &= -\kappa \cdot \frac{x}{s} \cdot (s-l)\mathrm{d}t + \sigma \mathrm{d}W_x \\
\mathrm{d}y &= -\kappa \cdot \frac{y}{s} \cdot (s-l)\mathrm{d}t + \sigma \mathrm{d}W_y \\
\mathrm{d}z &= -\kappa \cdot \frac{z}{s} \cdot (s-l)\mathrm{d}t + \sigma \mathrm{d}W_z
\end{aligned}
$$

with $s = \sqrt{x^2+y^2+z^2}$ and $l = 1$, $\sigma = 1$, and $\kappa = 350$, have a stationary distribution?

What shape is it?

References

[Ack00] P.J. Acklam. An algorithm for computing the inverse normal cumulative distribution function. home.online.no/~pjacklam/notes/invnorm/index.html, June 2000. University of Oslo, Statistics Division.

[AG02] M. Avellaneda and R. Gamba. Conquering the Greeks in Monte Carlo: Efficient Calculation of the Market Sensitivities and Hedge-Ratios of Financial Assets by Direct Numerical Simulation. In Geman et al. [GMPV02], pages 93–109. www.math.nyu.edu/faculty/avellane/ConqueringTheGreeks.pdf.

[ALT03] A.N. Avramidis, P. L'ecuyer, and P.-A. Tremblay. Efficient simulation of gamma and variance-gamma processes. In Chick et al. [CSF+03], pages 319–326. www.informs-cs.org/wsc03papers/039.pdf.

[And07] L. B. Andersen. Efficient Simulation of the Heston Stochastic Volatility Model. *SSRN eLibrary*, 2007. ssrn.com/paper=946405.

[AS84] M. Abramowitz and I.A. Stegun. *Pocketbook of Mathematical Functions*. Harri Deutsch, 1984. ISBN 3-87144-818-4.

[BK04] M. Broadie and Ö. Kaya. Exact Simulation of Stochastic Volatility and other Affine Jump Diffusion Processes. Working paper, Columbia University, New York, 2004. www.orie.cornell.edu/~aberndt/FEseminar/papers04/exact_sim_200409.pdf.

[BM58] G. E. P. Box and M. E. Muller. A note on the generation of random normal deviates. *Annals Math. Statist.*, 29:610–611, 1958.

[CSF+03] S. Chick, P.J. Snchez, D. Ferrin, D.J. Morrice, and A.F. Seila, editors. *Proceedings of the 1998 Winter Simulation Conference*, 2003.

[Dos77] H. Doss. Liens entre équations différentielles stochastiques ordinaires. *Annales de l'Institut Henrie Poincaré. Probabilités et Statistiques*, 13:99–125, 1977.

[ELM01] P. Embrechts, F. Lindskog, and A. McNeil. Modelling Dependence with Copulas and Applications to Risk Management. Working paper, Department of Mathematics, ETHZ CH-8092, Zürich, Switzerland, 2001. www.risklab.ch/ftp/papers/DependenceWithCopulas.pdf.

[Fey48] R. Feynman. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.*, 20:367–387, 1948.

[Fox96] B. L. Fox. Generating Poisson processes by quasi-Monte Carlo. Technical report, SIM-OPT Consulting, 872 Timber Lane, Boulder, CO 80304, 1996. citeseer.nj.nec.com/cache/papers/cs/4246/http:zSzzSzwww.math.hkbu.edu.hkzSzqmczSzgpp.pdf/fox96generating.pdf.

[GMPV02] H. Geman, D. Madan, S. R. Pliska, and T. Vorst, editors. *Mathematical Finance - Bachelier Congress 2000*, 2002.

[Jäc02] P. Jäckel. *Monte Carlo methods in finance*. John Wiley and Sons, February 2002.

[Kac51] M. Kac. On some connections between probability theory and differential and integral equations. In *Proc. 2nd Berkeley Symposium on Math. Stat. & Probability*, pages 189–215, 1951.

[Kau01] R. Kaufmann. Copulas as an Integrated Risk Management Tool. Risk 2001 europe conference, april 10-11, paris, RiskLab, Departement Mathematik, ETH Zürich, Switzerland, 2001. www.math.ethz.ch/~kaufmann/Copulas/Paris2001.pdf.

[KL51] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79–86, 1951.

[KP99] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992, 1995, 1999.

[KS91] I. Karatzas and S. E. Shreve. *Brownian motion and Stochastic Calculus*. Springer, 1991.

[Kva00] B. Kvasov. *Methods of Shape-Preserving Spline Approximation*. World Scientific Publishing, 2000. ISBN 9810240104.

[L'E01] P. L'Ecuyer. Software for Uniform Random Number Generation: Distinguishing the Good and the Bad. In Peters et al. [PMR01]. www.iro.umontreal.ca/~lecuyer/myftp/papers/wsc01rng.pdf.

[LMS01] F. Lindskog, A. McNeil, and U. Schmock. Kendall's Tau for Elliptical Distributions. Working paper, RiskLab, Departement Mathematik, ETH Zürich, Switzerland, 2001. www.risklab.ch/ftp/papers/KendallsTau.pdf.

[MN97] M. Matsumoto and T. Nishimura. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. www.math.keio.ac.jp/~matumoto/emt.html, 1997.

[Nea73] H. R. Neave. On using the Box-Muller transformation with multiplicative congruential pseudo-random number generators. *Applied Statistics*, 22:92–97, 1973.

[NX95] H. Niederreiter and C. P. Xing. Low-discrepancy sequences obtained from algebraic function fields over finite fields. *Acta Arithmetica*, 72:281–298, 1995.

[NX96] H. Niederreiter and C. P. Xing. Low-discrepancy sequences and global function fields with many rational places. *Finite Fields and Their Applications*, 2:241–273, 1996. www.dismat.oeaw.ac.at/pirs/niedxing.html.

[PMR01] B. A. Peters, J. S. Smithand D. J. Medeiros, and M. W. Rohrer, editors. *Proceedings of the 2001 Winter Simulation Conference*, Dec 2001.

[PTVF92] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1992. www.library.cornell.edu/nr/cbookcpdf.html.

[SB02] N. J. Smelser and P. B. Baltes, editors. *International Encyclopedia of the Social and Behavioral Sciences*, chapter Random Numbers (P. L'Ecuyer), pages 12735–12738. Pergamon, 2002. www.iro.umontreal.ca/~lecuyer/myftp/papers/ensbs.pdf.

[Sha48] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948.

[Sob76] I. M. Sobol'. Uniformly Distributed Sequences with an Additional Uniform Property. *USSR Computational Mathematics and Mathematical Physics*, 16(5):236–242, 1976.

[Tez95] S. Tezuka. *Uniform Random Numbers: Theory and Practice*. Kluwer Academic Publishers, 1995.

Modelling Long Run Relationships in Time Series

In this lecture. . .

- Financial time series and cointegration

- Integrated series. Testing for stationarity

- How the long-run relationship works: equilibrium correction

- Case Study: cointegration among spot rates (market data)

- Testing for multivariate cointegration *extra*

By the end of this lecture you will be able to . . .

- know why you cannot regress on prices
- understand integrated time series and DF test for unit root
- understand error correction formulation and Engle–Granger procedure
- confirm the long-run relationship between a pair of prices

Introduction

Cointegration analysis is a powerful tool for investigating *common factors* – sources of randomness – among price time series.

$$\text{Price}_A - \beta\,\text{Price}_B$$

is not a random walk! This spread is forecastable.

We have tried to predict individual asset price direction using ML Classifiers.

The proper statistical model 'to predict' one price from another (forecast the spread) is cointegration.

How do we work with empirical time series *in levels*, such as asset prices, CDS levels, or interest rates?

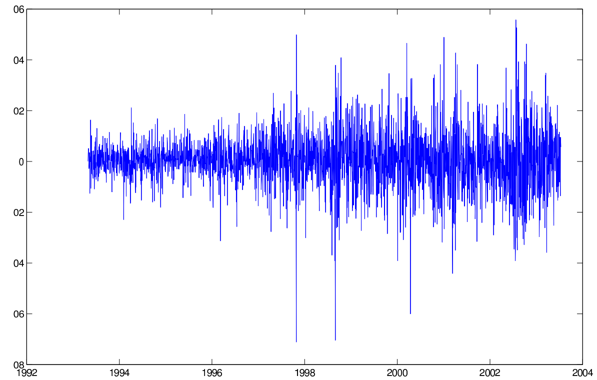
- The price levels are non-stationary.

- Regressing one price series on another is *spurious*.
  R^2 improves with N_{obs}

- Unlike with differences or returns, **we do not correlate**.

What is the underlying model of correlation?

Asset Returns



$$
\begin{aligned}
R_t^{CA=1} &= \beta_{1,0} + \beta_{11} R_{t-1}^{CA} + \beta_{12} R_{t-1}^{FR} + \ldots \beta_{1n} R_{t-1}^{US} + \ldots_{t-2} \ldots + \epsilon_{1,t} \\
R_t^{FR=2} &= \beta_{2,0} + \beta_{21} R_{t-1}^{CA} + \beta_{22} R_{t-1}^{FR} + \ldots \beta_{2n} R_{t-1}^{US} + \ldots_{t-2} \ldots + \epsilon_{2,t} \\
\ldots & \quad \ldots \\
R_t^{US=n} &= \beta_{n,0} + \beta_{n1} R_{t-1}^{CA} + \beta_{nn} R_{t-1}^{FR} + \ldots \beta_{nn} R_{t-1}^{US} + \ldots_{t-2} \ldots + \epsilon_{n,t}
\end{aligned}
$$

For returns... VAR is appropriate but **forecast is poor**.

Forecasting Market Returns

Vector Autoregression **fails** at forecasting daily returns.

For the 2011 data the full model, AIC BIC-tested for optimal lag and stability-checked, forecasting accuracy as follows:

	S&P 500	FTSE 100	HSE	N225
MSE	0.0001	0.0001	0.0001	0.0001
MAPE	1.0175	1.3973	2.5325	1.0111

MAPE results suggest a deviation $O(100\%)$ to $O(200\%)$.

Daily returns for a broad market are a very small quantity.

Static Equilibrium Between Returns

$$\mathbb{E}[r_A] \;=\; \beta\left(\mathbb{E}[r_M] - r_{rf}) + r_{rf}\right)$$

$$\mathbb{E}[r_A - r_{rf}] \;=\; \beta\,\mathbb{E}\left[r_M - r_{rf}\right]$$

CAPM assumes the existence of true and constant β.

Aside: betas are estimated by OLS wrt factors (eg, market or market segment returns series) in form of Linear Factor Model:

$$R_t^A \;=\; \alpha + \beta\,R_t^M + \beta_j F^j + \epsilon_t$$

Cointegration in Prices

The static equilibrium in changes in two asset prices – the steady-state means existence of the true and constant β_g, a growth rate.

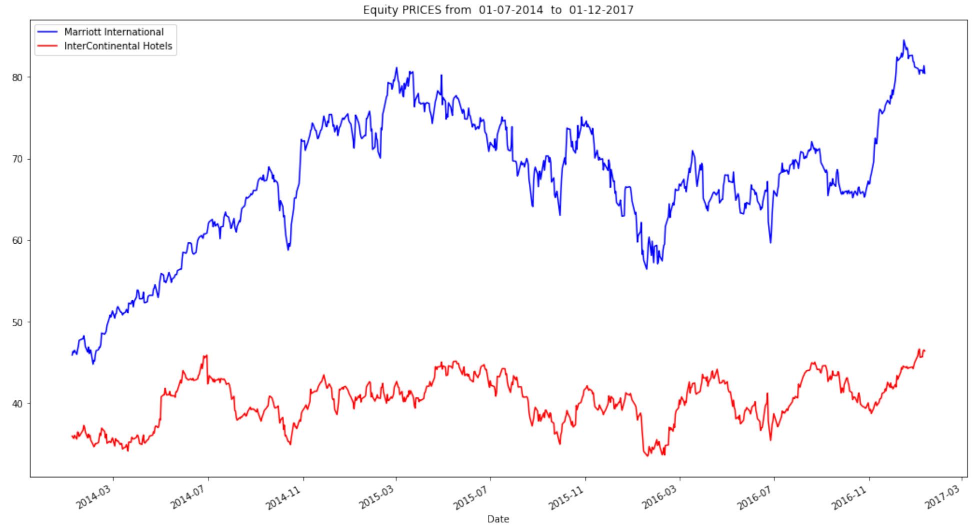
$$\Delta P_t^A = \beta_g \, \Delta P_t^B$$

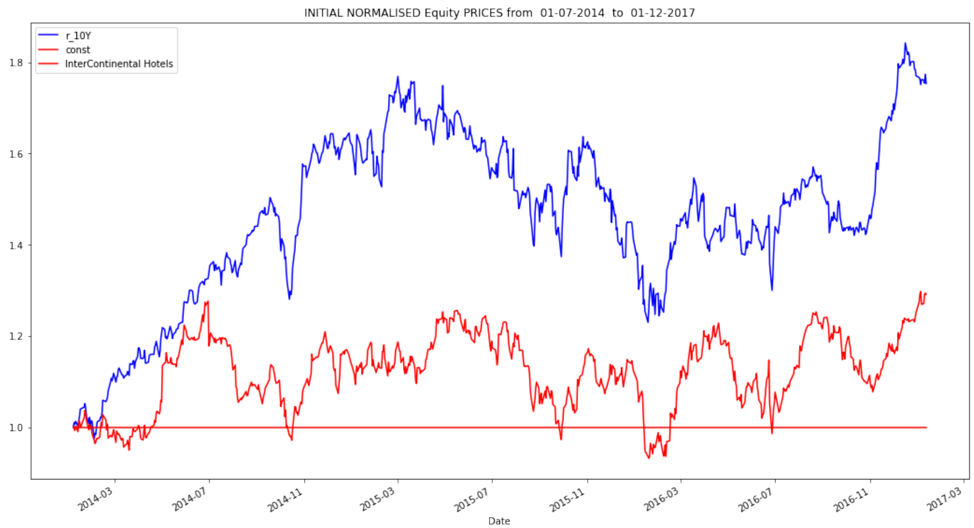
What about **the long run**?

$$\begin{aligned} \Delta P_t^{A=1} &= \beta_{11} \, \Delta P_t^B + \beta_{12} \, \text{Coint\_Factor} \\ \Delta P_t^{B=2} &= \beta_{21} \, \Delta P_t^A + \beta_{22} \, \text{Coint\_Factor} \end{aligned}$$

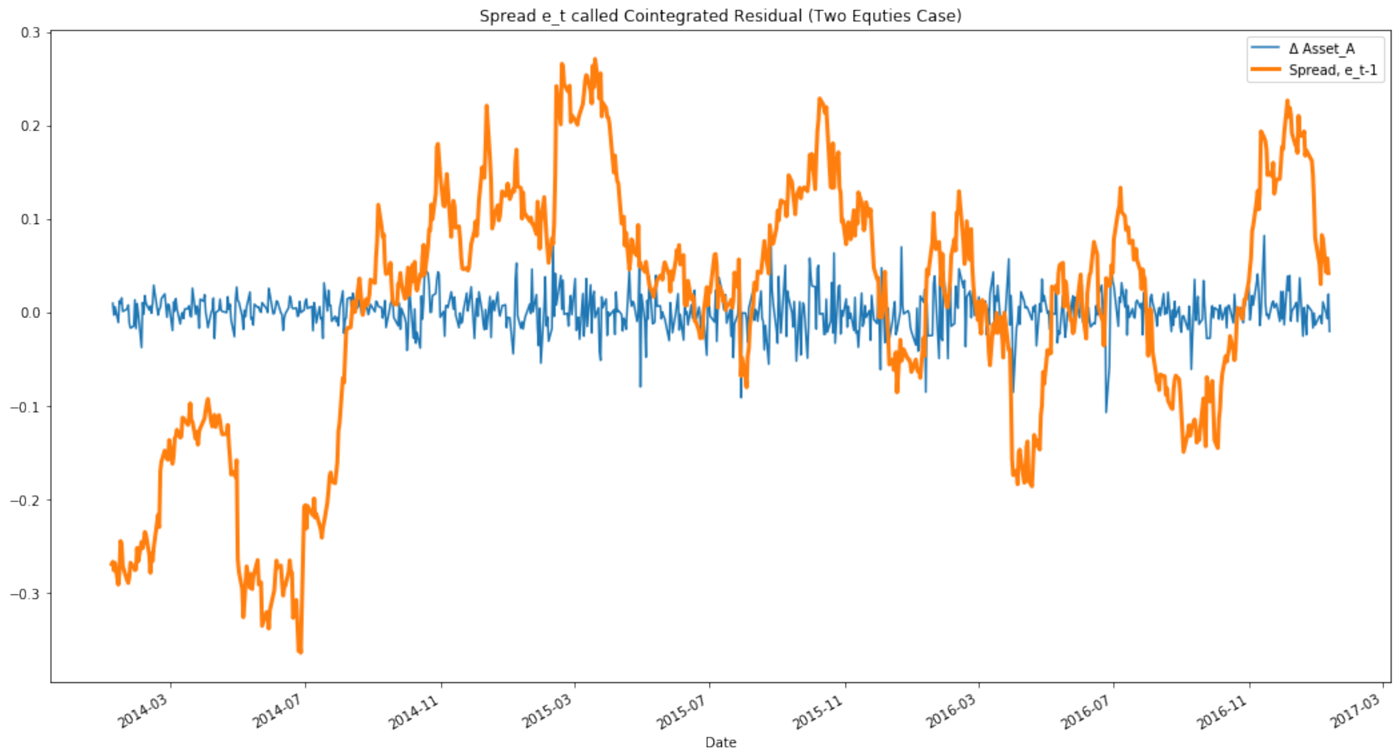
For equities/futures/rate levels... this correction model is appropriate, not naive regression of P_t^A on P_t^B .

Cointegration in Equities





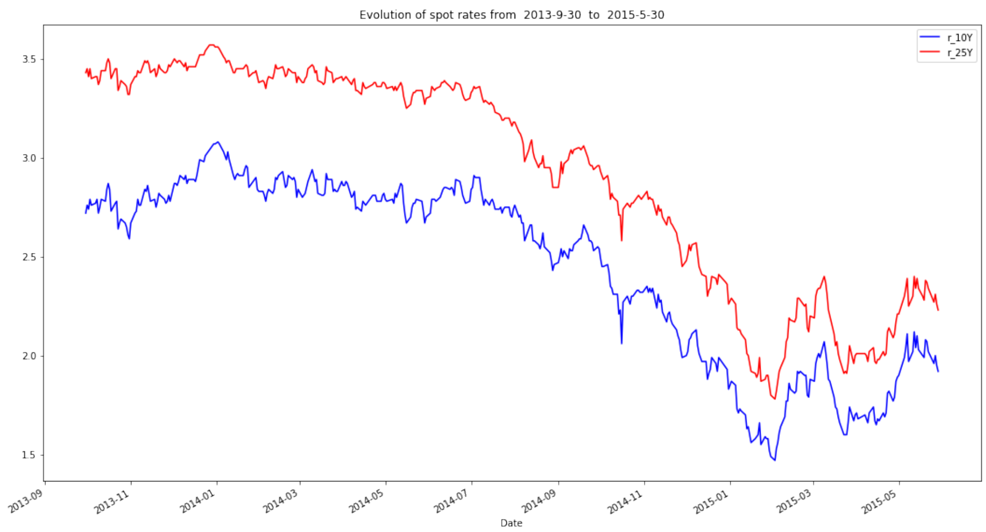
Cointegrated Residual – equities

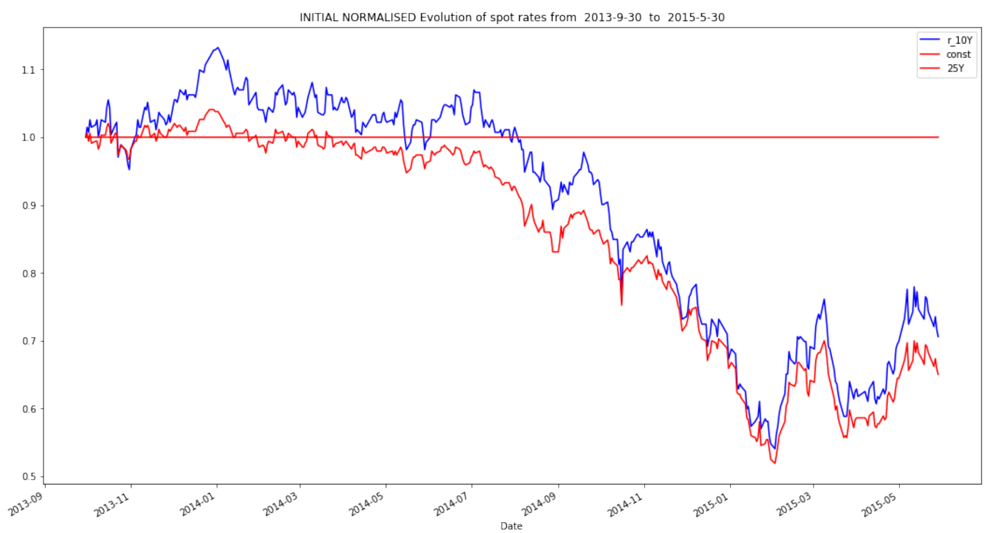


Those two equities proven to be cointegrated over the period – carefully chosen by examination over different time 'windows'.

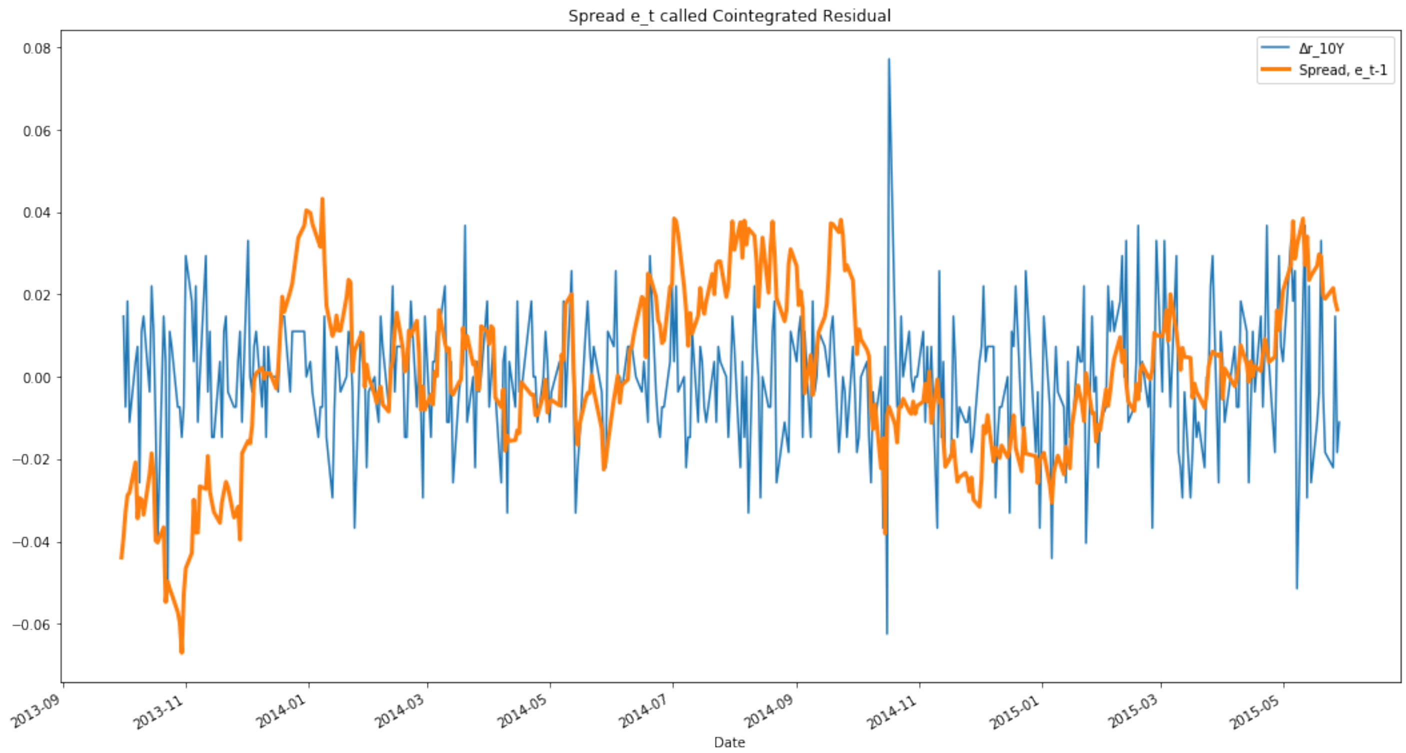
Half-life $< 3 - 4$ months, it is remarkable but these arb opportunities do exist.

Cointegration in Rates





Cointegrated residual – rates



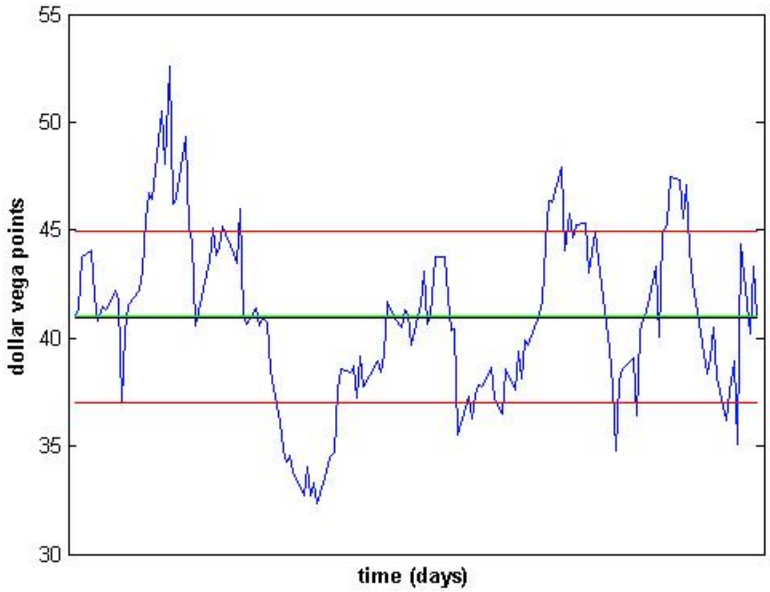
Those two rates r_{10Y} and r_{25Y} proven to be cointegrated over given period.

The residual is $r_{10Y} - \beta_C\, r_{25Y}$ confirmed to be 1) stationary and 2) common factor in Error Correction.

Dynamic Equilibrium

This mean-reverting spread produced by two co-moving series – is a dynamic equilibrium.

As opposed to static equilibrium which is CAPM β.



Spread example from a hedged basket of VIX Futures. Cointegrated residual $e_t = P_t^{Fut1} - \beta_2 P_t^{Fut2} - \ldots - \beta_n P_t^{Fut\,N}$ is fitted to OU process $(\theta, \mu, \sigma_{OU}, \sigma_{eq})$.

Diamond, R (2013) *Learning and Trusting Cointegration*, WILMOTT.

Uses of cointegration

Cointegration is useful in hedging applications: particularly when hedges are from a different asset class (eg, oil price vs oil-extracting country FX).

- **Equities**: the company and target of its acquisition (particularly when stock swap involved)
- **Commodities**: US heating oil vs. natural gas (competing fuels). Agricultural futures.
- **Rates**: funding rates at short end, long-term rates affecting middle of yield curve.
- **Term Structure**: VIX futures, a very liquid market

For very large samples $> 10Y$, it is likely to find co integration in commodities, global equity indicies.

Granger-Johansen Representation

The stationarity of spread e_t was **a discovery** in statistics.

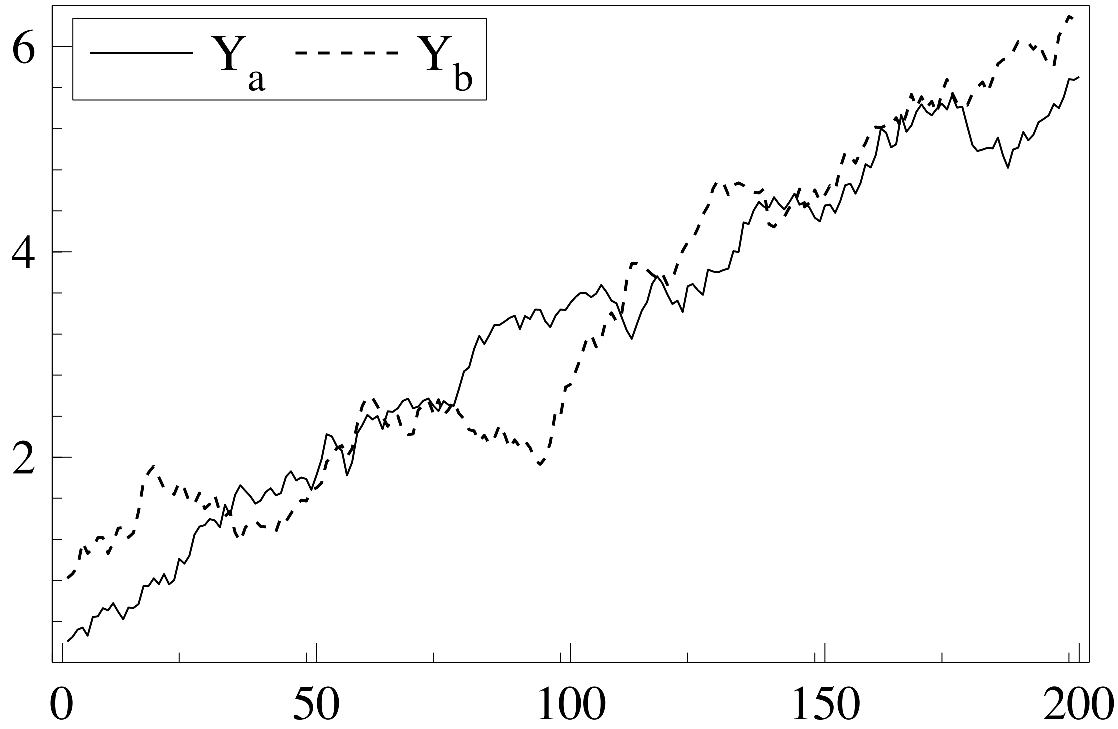
Implies there is a stochastic process (e.g. random walk) that is so similar in two series that it gets removed by linear differencing,

$$\begin{aligned} P_t^A &= P_0^A + \underbrace{\sum \epsilon_s^A} && \text{integrated process} \\ P_t^B &= \beta P_{t-1}^A + \epsilon_t^B && \text{regress and assume } \beta \approx 1 \\ P_t^B &= P_0^A - \epsilon_t^A + \epsilon_t^B + \underbrace{\sum \epsilon_s^A} \end{aligned}$$

"There are fewer feedbacks than variables."

Explaining Example 1

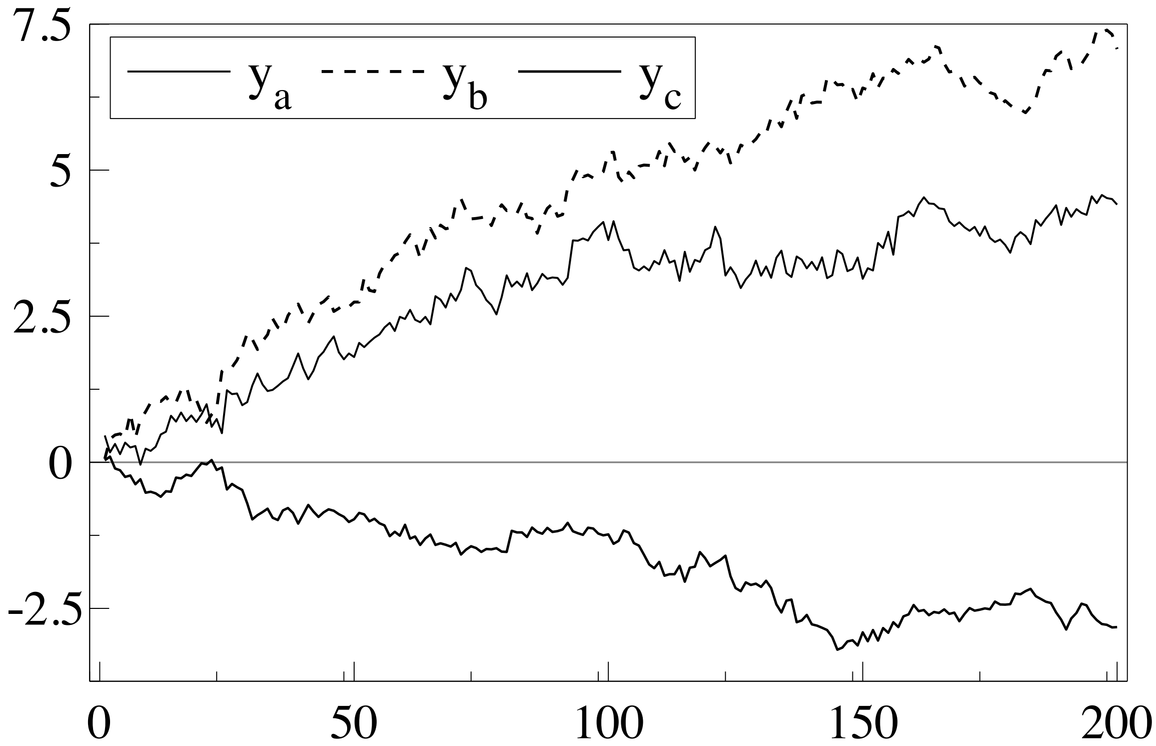
Two series below move together and end up in the similar state, but **not** cointegrated.



From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Explaining Example 2

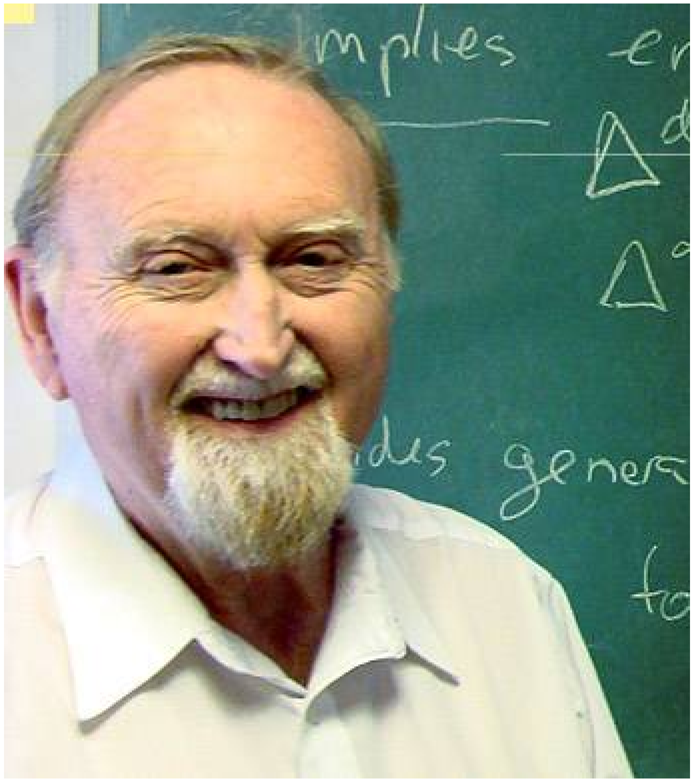
Three series are **cointegrated** and driven by the same common factor $e_t = \boldsymbol{\beta}'_{Coint}\boldsymbol{Y_t}$.



From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Phillips in an interview to *Econometric Theory*: "Clive set out to prove that such linear combinations of integrated variables would in fact remain integrated... In the process, he instead established the conditions under which cointegration could occur:

– when a dynamic system with a reduced-rank feedback matrix must generate integrated data." (Photo: Clive Granger)



Integrated Random Process, Unit Root and Stationarity Testing

Integrated process (unit root)

Take Vector Autoregression model, lag 1 , so VAR(1).

- We start with $\boldsymbol{Y_t = \beta Y_{t-1} + \epsilon_t}$
- $\boldsymbol{Y_{t-1}}$ depends on $\boldsymbol{Y_{t-2}}$, and so, $\boldsymbol{Y_t = \beta\left(\beta Y_{t-2} + \epsilon_{t-1}\right) + \epsilon_t}$.
- Next insertion gives

$$\boldsymbol{Y_t = \beta\left(\beta\left(\beta Y_{t-3} + \epsilon_{t-2}\right) + \epsilon_{t-1}\right) + \epsilon_t}$$

By induction,

$$\boldsymbol{Y_t = \beta^n\, Y_{t-n} + \sum_{n=1}^{t}\left(\beta^{n-1}\,\epsilon_{t-(n-1)}\right)}$$

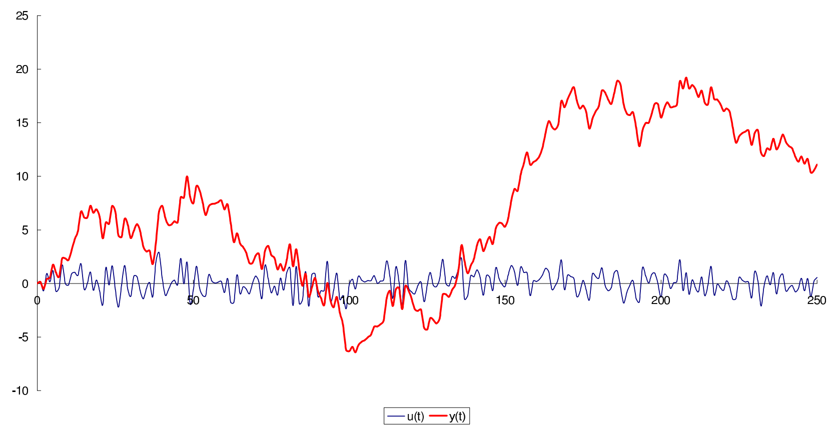
Think about the perfect **unit root** case $\boldsymbol{\beta = 1}$,

- then $\boldsymbol{Y_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2} + \ldots + Y_0 = \sum \epsilon_s + Y_0}$

In continuous time summation becomes an integration

$$\epsilon_{t,\tau} \stackrel{D}{=} \int_t^{t+\tau} \boldsymbol{\sigma} dW_s.$$

We say **the process is integrated** of order one, $Y_t \sim I(1)$.



Simulated random walk (BM) process adds up increments,

$$Y_t = \sum \sigma dW_s = \sum \epsilon_s$$

Testing BROWNIAN MOTION for a unit root

```
Null Hypothesis: Y_T has a unit root
Exogenous: None
Lag Length: 0 (Fixed)
============================================================
                                        t-Statistic  Prob.*
============================================================
Augmented Dickey-Fuller test statistic   -0.432663   0.5261
Test critical values1% level             -2.574245
                   5% level              -1.942099
                  10% level              -1.615852
============================================================
```

DF relies on the higher than t critical values, making it difficult to reject a unit root hypothesis.

Dickey-Fuller Test

$$\begin{aligned} Y_t &= \beta Y_{t-1} + \epsilon_t \\ Y_t - Y_{t-1} &= (\beta - 1) Y_{t-1} + \epsilon_t \\ \Delta Y_t &= \phi Y_{t-1} + \epsilon_t \qquad \textbf{test equation} \end{aligned}$$

The alternative hypothesis $H_1 : \phi \neq 0$ (phi significant).

The null hypothesis $H_0 : \phi = 0$ (insignificant, true value is 0).

Custom statistical tests work like this:

1. **Test statistic** for significance of ϕ calculated as usual.

$$\frac{\phi}{\text{std error}}$$

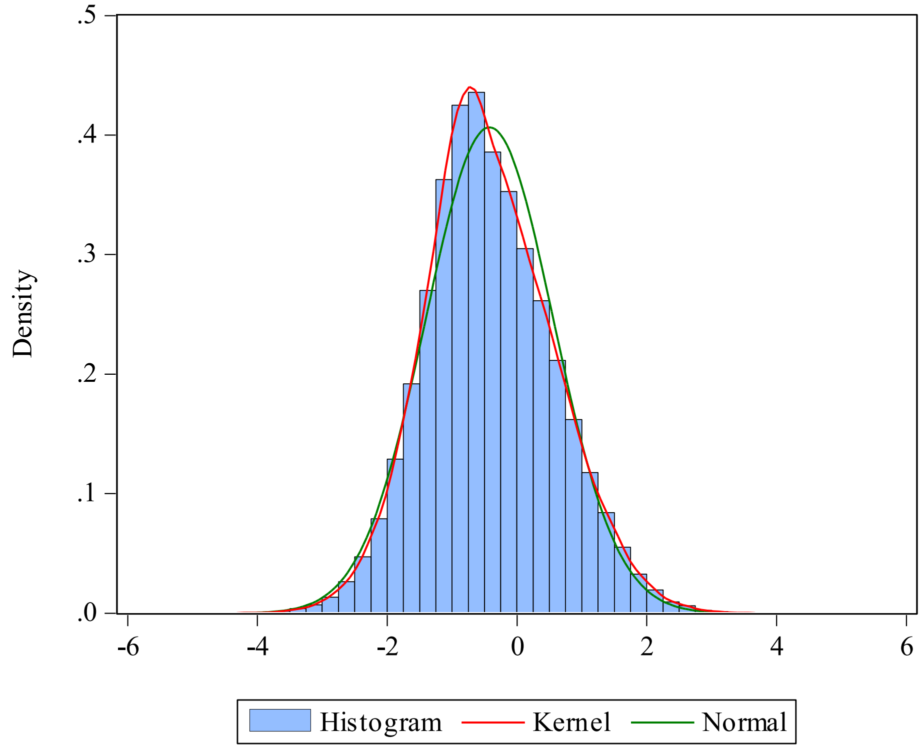
2. Standard error for ϕ is under-estimated, and so we use 'the right distribution for the wrong test statistic.'

   Conventional critical values (t distrib) give to over-rejection of H_0 when it is true.

$$H_0 : \phi = \beta - 1 = 0 \qquad Y_t = \beta^{=1} Y_{t-1} + \epsilon_t \qquad \Delta Y_t = \epsilon_t$$

   Critical value says 't-Statistic' but taken from the Dickey-Fuller as tabulated by MacKinnon (2010 update).

Bootstrapped Dickey-Fuller distribution



DF distribution is bootstrapped (form of Monte Carlo as used in econometrics) by generating *iid* residuals ϵ_t, where $Y_t \sim I(1)$

$$H_0 : \Delta Y_t = (1 - \beta)Y_{t-1} + \epsilon_t$$

Augmented Dickey-Fuller

Lagged differences Δy_{t-k} improve robustness if there is noticeable serial correlation

$$\Delta y_t = \phi y_{t-1} + \sum_{k=1}^{p} \phi_i \Delta y_{t-k} + \epsilon_t$$

Insignificant $\phi = \beta - 1 = 0$ means unit root for series y_t. DF critical values re-tabulated for each number of lags k.

Ready statistical tests (R, Matlab, etc) offer to add constant '**drift**' or time-dependent '**trend**'.

$$\Delta y_t = \phi y_{t-1} + \sum_{k=1}^{p} \phi_i \Delta y_{t-k} + \underbrace{\text{const} + \beta_t t} + \epsilon_t$$

These modifications are your false friends because they create temporary dependency and give **overfitted results**.

Statistical tests implemented in R usually present the underlying regression equation.

So you are able to identify parameters and understand whether an excessive specification was used.

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

Test regression trend

lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

Equilibrium Correction Model (ECM)

How Cointegration Works

Estimating Cointegration - Pairwise

Pairwise Estimation: select two likely candidates to have a stationary spread: gas vs. heating oil futures, two pharmas or hotel chains where one interested in merger.

- **Step 1.** Regress one price P_t^A on another P_t^B, and test the fitted residual by ADF with lag=1. If stationary, proceed.

- **Step 2.** Confirm significance of correction term in the eqns for ΔP_t^A, ΔP_t^B.

- Step 3. Fit the stationary spread to OU SDE solution (by autoregression) to evaluate mean-reversion: μ_e, θ, σ_{eq}.

Engle-Granger Procedure - Fitted Residual

Step 1. Obtain the fitted residual and test it for stationarity

$$P_t^A = \beta_0 + \beta_1 P_t^B + \epsilon_t$$

$$\widehat{e}_t = P_t^A - \widehat{\beta}_C P_t^B - const \qquad \beta_1 = \widehat{\beta}_C$$

- **If the residual non-stationary** then no long-run relationship exists and regression is spurious. Use MacKinnon (2010) for $\tau_c, N = 2$. 'No-constant' assumption in testing eqn is attractive but highly unrealistic.

- Use Augmented Dickey Fuller test with lag=1.

Engle-Granger Procedure - Error Correction

Step 2. Plug the stationary fitted residual $\widehat{e}_{t-1}$ from Step 1 as shifted into the error correction linear regression eqn and confirm statistical significance of its coefficient.

$$\Delta P_t^A = \phi \Delta P_t^B - (1-\alpha)\, \widehat{e}_{t-1}$$

$$\Delta P_t^A = \phi \Delta P_t^B - (1-\alpha) \left(P_{t-1}^A - \beta_C P_{t-1}^B - \mu_e \right)$$

- It is required **to confirm the significance for** $(1-\alpha)$ coefficient.

- Correction to equilibrium comes in very small moves as $(1-\alpha) \ll 1$.

$$\Delta P_t^A = \phi_{shortrun}(P_t^B - P_{t-1}^B) + \phi_{longrun}(P_{t-1}^A - \beta_C P_{t-1}^B)$$

Planners vs. Hedgers

EC model has Vector Autogregression representation

$$\Delta P_t^{A=1} = \beta_{11}\,(e_{t-1} - \mu_e) + \beta_{12}\,\Delta P_t^{B} + \beta_{13}\,\Delta P_{t-1}^{A}$$

$$\Delta P_t^{B=2} = \beta_{21}\,\mathsf{CointFactor} + \beta_{22}\,\mathsf{StaticEq} + \beta_{23}\,\mathsf{Augment}$$

$$\boldsymbol{\Delta P_t} = \boldsymbol{\Pi}\,\boldsymbol{P_{t-1}} + \sum_i \boldsymbol{\Gamma_i}\,\boldsymbol{\Delta P_{t-i}} + \boldsymbol{\mu_0}$$

1. Notation in bold is known as matrix form VECM (constant μ_0, no trend).
2. Coint_Factor has non-unique $[1, -\beta_C]$ cointegrating weights (uncertainty, see Hedging Puzzle choices in Case Study)
3. The challenge to robustness is 'sudden' shift $\boldsymbol{\mu}_e^{Old} \to \boldsymbol{\mu}_e^{New}$.

This is not forecasting!

Implementation in Python

```
from statsmodels.tsa.stattools import coint
coint(PriceA, PriceB)

import statsmodels.tsa.stattools as ts
ts.adfuller(CointResidual)
```

ts.adfuller() gives a rudimentary output for DF Test for stationarity.

Critical values were fixed to MacKinnon(2010) after been wrong for about 2011-2017!

Requirement (TS Topic): implement Engle-Granger procedure from the first principles. Enclosed R code gives a complete example.

VECM in Python

```
import statsmodels.tsa.vector_ar.vecm as cajo
johansen_test = cajo.coint_johansen(Prices, 0, 2)
```

Python routines output will be very similar to VECM output from R routines (package *urca*), such as *cajorls()*.

- please see Cointegration Case extra slides (about p.58).

VECM seem to be included from statsmodels v0.11.0. However, to install dev version, use *git()* instead of *pip*. Refer to the source code comments to understand inputs and outputs.

https://www.statsmodels.org/dev/generated/statsmodels.tsa.vector_ar.vecm.VECM.html

Implementation in R: Multivariate Cointegration

Johansen Procedure tests for the number (how many) of cointegrated relationships.

It is a powerful **screening tool**: to identify dependencies among and within sections of the yield curve, segments of the market, special M&A situations.

The workhorse is *ca.jo()* function from the **R package** *urca*.

```
           test 10pct  5pct  1pct
r <= 6 |   4.67  7.52  9.24 12.97
r <= 5 |   5.87 13.75 15.67 20.20
r <= 4 |   9.78 19.77 22.00 26.81
r <= 3 |  24.98 25.56 28.14 33.24
r <= 2 |  44.91 31.66 34.40 39.79
r <= 1 |  46.88 37.45 40.30 46.82
r = 0  | 101.10 43.25 46.45 51.91
```

cajorls() presents the output as a set of familiar OLS equations with EC term, separate line for each price.

Vector Autoregression to Multivariate Cointegration [EXTRA]

Remember, we learned Vector Autoregression, an endogenous system of equations for **Returns**. We can't use VAR on **Prices** – that would be a spurious model.

Prices can be tied up with special error correction equations:

Changes in prices ΔP_t can have a special correction, that makes prices to move together over the long term.

$$\Delta P_t \;=\; \Pi\, P_{t-1} + \Gamma_1 \Delta P_{t-1} + \epsilon_t$$

$\Longrightarrow$ Π must have **a reduced rank**, otherwise *rhs* will not balance *lhs*. Differences ΔP_t will not equate to non-stationary, random prices P_t on *rhs*.

Now, to make this look alike Engle-Granger, we decompose coefficients $\mathbf{\Pi} = \boldsymbol{\alpha}\,\boldsymbol{\beta}'_C$

$$\begin{aligned}
\boldsymbol{\Delta P_t} &= \mathbf{\Pi}\,\boldsymbol{P_{t-1}} + \mathbf{\Gamma}\boldsymbol{\Delta P_{t-1}} + \boldsymbol{\epsilon_t} \\
\boldsymbol{\Delta P_t} &= \boldsymbol{\alpha}\,(\underbrace{\boldsymbol{\beta}'_C\,\boldsymbol{P_{t-1}} + \boldsymbol{\mu_e})} + \mathbf{\Gamma_1}\boldsymbol{\Delta P_{t-1}} + \boldsymbol{\epsilon_t}
\end{aligned}$$

μ_e is called 'a restricted constant' or deterministic trend.

$$(n \times n) \;=\; (n \times r) \times (r \times n)$$

Above are dimensions for $\mathbf{\Pi} = \boldsymbol{\alpha}\,\boldsymbol{\beta}'_C$. r columns of vectorised $\boldsymbol{\beta}'$ are linearly independent – cointegrating vectors.

Eigenvalues of $\mathbf{\Pi}$ are utilised to compute both, Trace Statistic and Max Eigenvalue Statistic.

Sequential Testing for Cointegration Rank

r	lambda	1-lambda	ln(1-lambda)	Trace	CV trace	MaxEig	CV MaxEig
0	0.0167	0.9833	-0.0168	105.7518	103.8473	44.8038	40.9568
1	0.0094	0.9906	-0.0094	60.9479	76.9728	25.1283	34.8059
2	0.0046	0.9954	-0.0046	35.8197	54.0790	12.3440	28.5881
3	0.0038	0.9962	-0.0038	23.4757	35.1928	10.2469	22.2996
4	0.0031	0.9969	-0.0031	13.2287	20.2618	8.3510	15.8921
5	0.0018	0.9982	-0.0018	4.8777	9.1645	4.8777	9.1645

- Trace statistic $H_0 : r = r^*$, and $H_1 : r > r^*$. **Stop at $r^* = 1$**

$$LR_{r^*} = -T \sum_{i=r^*+1}^{n} \ln(1 - \lambda_i)$$

- Maximum eigenvalue statistic $H_0 : r = r^*$, and $H_1 : r = r^* + 1$

$$LR_{r^*} = -T \ln(1 - \lambda_{r^*+1})$$

Cointegrating Vector Estimators β'_{Coint}

	1	2	3	4	5	6	7
Canada	6.78395	-1.96320	-9.07554	7.03629	2.56142	6.25519	-2.08045
France	4.86921	4.86043	-2.08623	-7.28739	2.28808	-1.59825	-1.60875
Germany	-15.76001	-5.94947	0.12170	3.34469	-0.01972	-4.04040	4.24522
Japan	-1.22250	5.52024	-0.70856	1.03285	-0.17938	-0.08242	1.76463
UK	27.19903	-13.06796	-0.55980	-0.36245	-1.03954	-1.76308	0.23821
US	-10.25644	13.17254	7.00734	-0.56186	-5.15207	2.16214	-2.37646
Const	-117.01015	-5.47002	59.45116	-32.77753	5.05186	-8.11528	-7.19582

- $n-1$ columns are linearly dependent on the 1st column.

- $r=1$ columns of $\boldsymbol{\beta}$ are cointegrating vectors, take the first column and standardise it (row echelon form).

$$\begin{bmatrix} 1 & 0.7178 & -2.3231 & -0.1802 & 4.0093 & -1.5119 & -17.2481 \end{bmatrix}$$

The allocations $\hat{\beta}'_{Coint}$ provide a mean-reverting spread.

Cointegration Estimation

1. **Engle-Granger Procedure** for a pair of time series.

   - Cause/effect (leading variable) can be established and removes uncertainty about non-unique cointegrating weights $[1, \beta_C]$.

   - Estimation of various basis, eg, tenor basis between rates..

2. **Johansen Procedure** for a set of cointegrating relationships in a multivariate setting.

   - Relies on theorem for a reduced-rank matrix with r linearly independent rows.

     First, $\boldsymbol{\beta}'_C$ estimated, second $\boldsymbol{\alpha}$ are inferred making this a calibration.

Summary

Please take away the following ideas...

- Financial time series are non-stationary and so, naïve linear correlation and regression are spurious models with $R^2 \approx 1$.
- Large samples and cross-validation only amplify the issue.
- **Cointegration**: linear combination of time series produces a stationary spread. Long-term relationship relies on the common stochastic process.
- Engle-Granger procedure is a stepping stone which assumes that process (naïve coint. residual). Multivariate cointegration properly checks the reduced rank.
- To construct a trade: apply the special statistical arb techniques, such as fitting to OU process, and backtest.

Case Study B: Pound Sterling Spot Rates

- Traditionally, cointegration is tested in the very long run
- We had Case Study A testing for an equilibrium between US T-Bills and Treasuries over the horizon of 1960-2010.

  HOWEVER

- As quants we have to look for co-movement in the current, frequent market data.

We will use this opportunity to get introduced to R.

Spot Curve

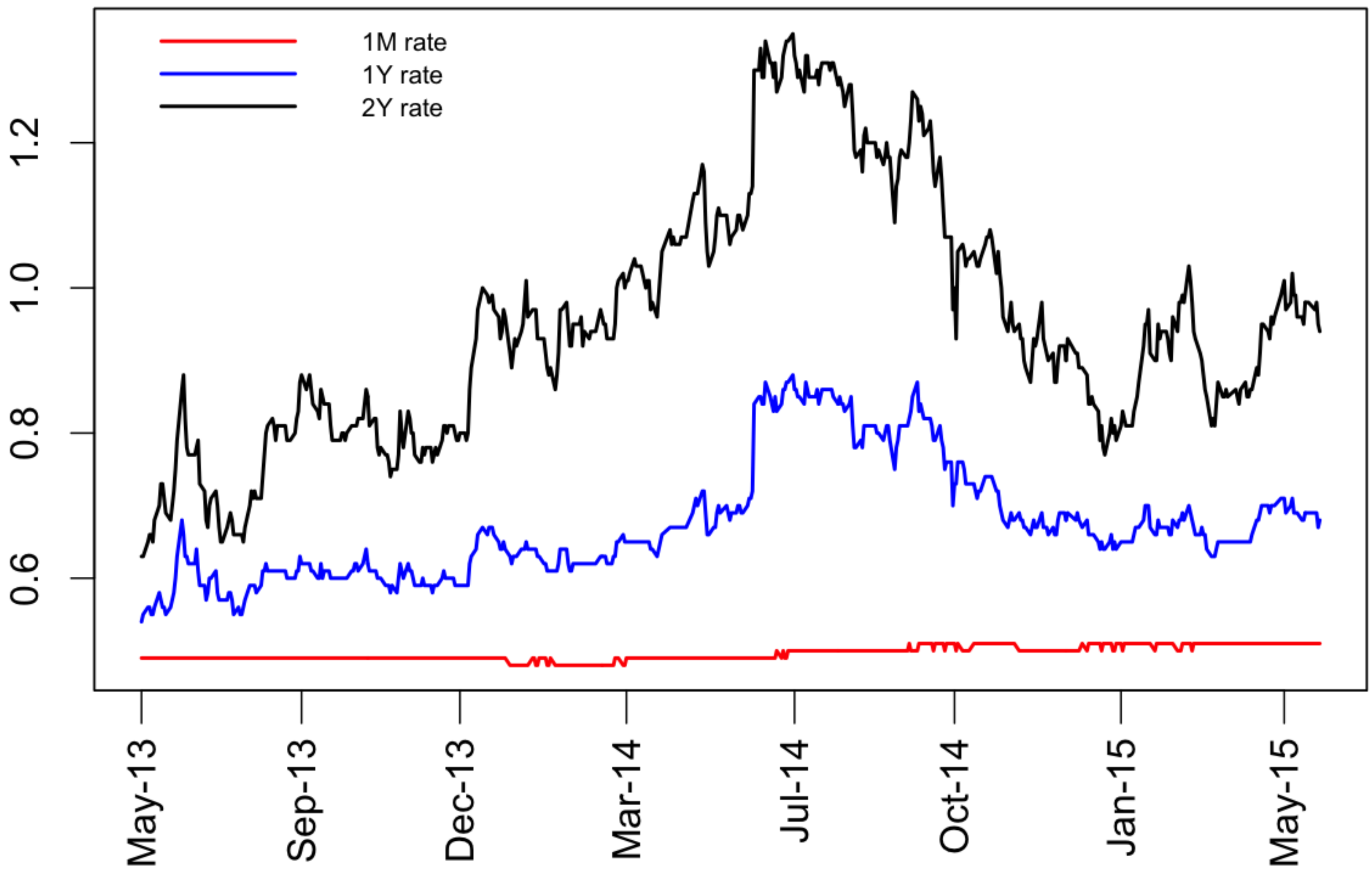
The Bank of England provides the daily yield curve data. It makes sense to consider smaller windows of the long timeframe:

- two-year window May 2013 – May 2015 (charts below) vs.

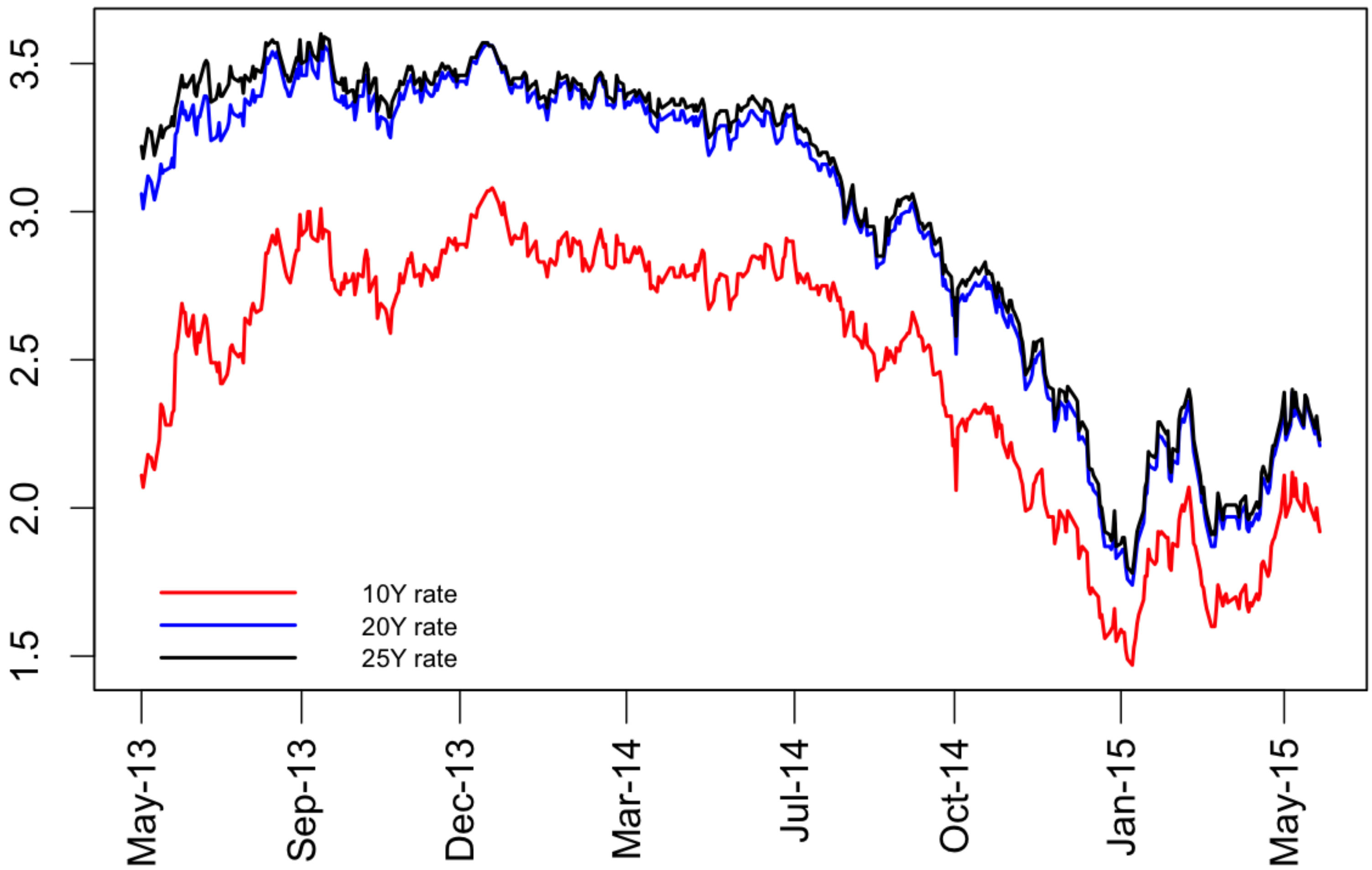
- all data from from Jan 2005 to May 2015.

We have to learn the equilibrium-correction mechanics (**ECM**) but it's worthwhile to have a peek from the multivariate test for cointegration.

Spot Rates at Short End



Spot Rates at Long End



Problems with curve data

1. r_t at the short end (0.8Y, 1Y, 2Y) and y_t at the long end (7Y, 10Y, 20Y) **do not** come as cointegrated in samples of two-three year period.

   There is simply not enough horizon for a cointegrated relationship to transpire.

2. Let's play a game: **Which long-end rates are co-integrated?** Choose pairs among 10Y, 20Y, 25Y.

   Can't decouple that easily.

   Similar pattern comes up for the short end, if all data included in the testing.

   Parallel to that, short rates have independent co-movement.

Engle-Granger preview

Let's choose a model with **10Y and 25Y tenors** because of their importance as benchmarks.

- We set up a naive cointegrating equation

$$r_{10Y} = \beta\, r_{25Y} + e_t \qquad \Rightarrow \qquad \hat{e}_t = r_{10Y} - \beta\, r_{25Y}$$

- We test this estimated residual $\hat{e}_t$ for stationarity by CADF.

  If the residual is stationary, it means that r_{10Y} and r_{25Y} have a unit root **in common**, removable by differencing

Long-run relationship r_{10Y} on r_{25Y}

```
lm(formula = curve2.this$X10 ~ curve2.this$X25)

                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.15878    0.03132    5.07  5.6e-07 ***
curve2.this$X25  0.76980    0.01023   75.28  < 2e-16 ***

Residual standard error: 0.1231 on 504 degrees of freedom
Multiple R-squared:  0.9183,Adjusted R-squared:  0.9182

Residuals:
     Min       1Q   Median       3Q      Max
-0.53675 -0.03449  0.01926  0.07920  0.18461
```

As usual, regressing one non-stationary series on another gives *extremely* significant coefficients. Large N_{obs} makes $R^2 \to 1$.

Long-run relationship if cointegrated

$$\widehat{r}_{10Y} = 0.159 + 0.77\, r_{25Y} + \widehat{e}_t$$

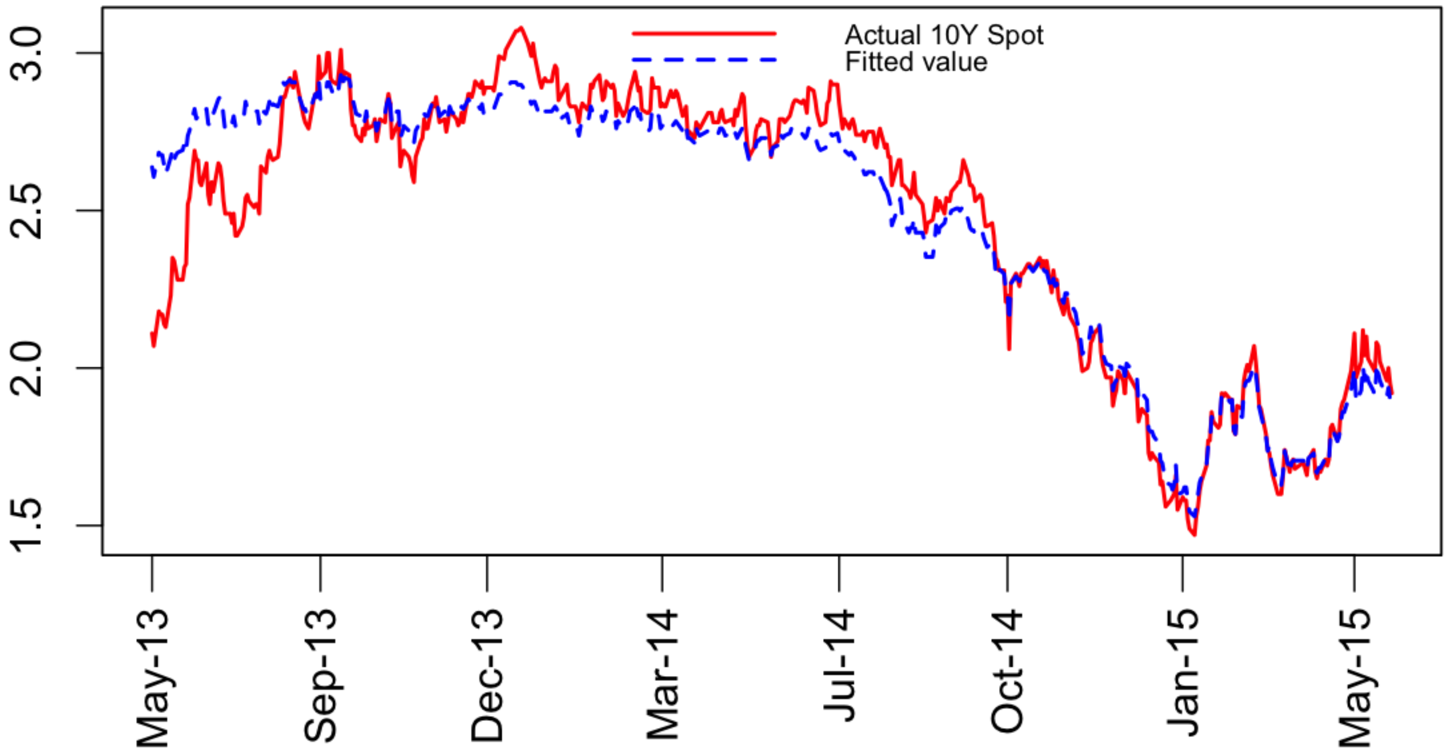
This model is valid only if it produces stationary $\widehat{e}_t$, so there is co-integration between r_{10Y} and r_{25Y}

It only works in the context of the equilibrium correction over the long-run, producing stationary and mean-reverting residual:

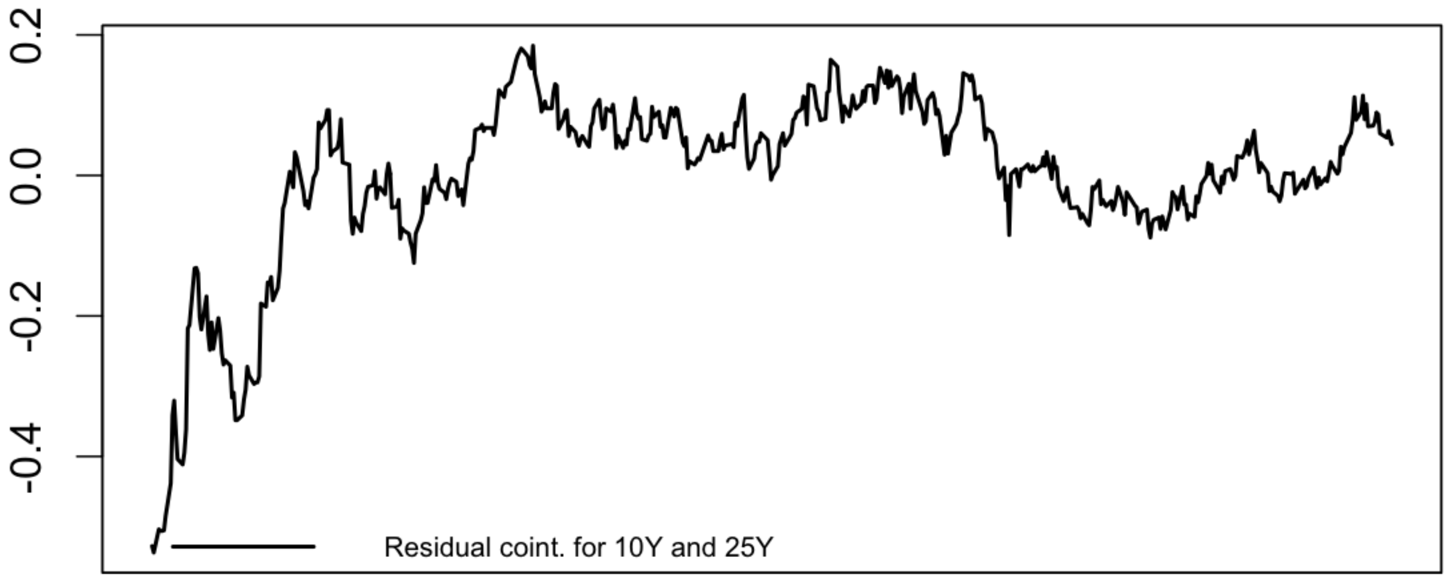
$$\widehat{e}_t = r_{10Y} - (0.159 + 0.77\, r_{25Y}).$$

Linear regression FIT for r_{10Y}

Our linear model aims to obtain $\widehat{e}_t$ so we would be differencing actual r_{10Y} with fitted $\widehat{r}_{10Y}$.



Stationary cointegrating residual $\widehat{e}_t$



We will confirm the stationarity of residual, and proceed with forming error-correction equations.

Stationarity test for $\widehat{e}_t$

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.038559   0.008548  -4.511 8.06e-06 ***
z.diff.lag -0.042376   0.043711  -0.969    0.333

[DF test-statistic is -4.5107, for which critical values]
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

Residual standard error: 0.02318 on 502 degrees of freedom
Multiple R-squared:  0.04071,Adjusted R-squared:  0.03689
```

Dickey-Fuller Test reminder

Null Hypothesis: time series has a unit root

We assume a linear trend, so ΔY_t will have a constant

$$\Delta Y_t = \text{Const} + \phi Y_{t-1} + \phi_1 \Delta Y_{t-1}$$

If ϕ is insignificant the time series has a unit root.

We can augment the test equation with more lags in $\phi_k \Delta Y_{t-k}$ or time-dependence $\phi_t t$ where ϕ_t is the drift.

That is likely to increase significance. However, beware you might be innocently introducing time dependence (growth/decrease) where there is none.

Long-run relationship (cointegrated)

ECM estimation [R code provided for your exploration] gives

- **the calbirated parameter** of interest is the speed of correction towards the equilibrium $(1-\alpha)$

  It is inevitably small but **must be** significant for cointegration to exist.

- We have quite good correlation between differences Δr_{10Y} and Δr_{25Y}. There is co-movement on the short timescale.

  For the lower frequency samples, you might find that Δr_t (for the short rate) and Δy_t (for some long-term rate) are cointegrated but correlated weakly negatively.

Equilibrium Correction Model: two-way, two residuals

$$\Delta r_{10Y} = 1.086\Delta r_{25Y} - 0.02716\, e^{10Y}_{t-1} + \epsilon_t$$

```
             Estimate Std. Error t value Pr(>|t|)
tenorX.diff  1.085090   0.022986  47.206  < 2e-16 ***
ec_term.lag -0.027164   0.007202  -3.772 0.000181 ***

Residual standard error: 0.01981   Multiple R-squared: 0.8202
```

$$\Delta r_{25Y} = 0.752\Delta r_{10Y} - 0.01206 e^{25Y}_{t-1} + \epsilon_t$$

```
             Estimate Std. Error t value Pr(>|t|)
tenorY.diff  0.751627   0.015910  47.243   <2e-16 ***
ec_term1.lag -0.012059   0.004851  -2.486   0.0132 *

Residual standard error: 0.01649   Multiple R-squared: 0.8175
```

Summary

Please take away the following ideas...

- this case of evolution of spot rates at different tenors is a case of a basis relationship,
- so imposing a long-run relationship and using Engle-Granger procedure has more statistical power,
- r_{10Y} and r_{25Y} series each have a unit root,
- it turns out that by differencing these time series, the unit root got cancelled and a stationary residual obtained,
- that means the time series are co-integrated.

Case Extra Slides

- Restricted VECM from Johansen Procedure
- Engle-Granger Procedure for r_{25Y} on r_{10Y} (other way)
- Linear regression on differences Δr_{25Y}, Δr_{10Y}
- Hedging ratio puzzle

Restricted VECM for Δr_{10Y} and Δr_{25Y}

```
cajorls(johansen.test)

lm(formula = substitute(form1), data = data.mat)

          X10.d     X25.d
ect1      -0.05842  -0.02647
X10.dl1   -0.13888  -0.09543
X25.dl1    0.07943   0.06495

[Cointegrating Equation (EC term)]
               ect1
X10.l2    1.0000000
X25.l2   -0.7870489
constant -0.1435463
```

Long-run relationship r_{25Y} on r_{10Y} (other way)

The linear model $r_{25Y} = \beta\, r_{10Y} + \epsilon_t$ only aims to obtain $\widehat{e}_t$.

```
lm(formula = curve2.this$X25 ~ curve2.this$X10)

                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.05686    0.03989   1.425    0.155
curve2.this$X10  1.19295    0.01585  75.285   <2e-16 ***

Residual standard error: 0.1532 on 504 degrees of freedom
Multiple R-squared:  0.9183,Adjusted R-squared:  0.9182
F-statistic:  5668 on 1 and 504 DF,  p-value: < 2.2e-16

Residuals:
     Min       1Q    Median       3Q      Max
-0.18591 -0.08516 -0.03819  0.02177  0.65373
```

Stationarity test for $\hat{e}_t$ (other way)

```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################

lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.033920   0.007759  -4.372  1.5e-05 ***
z.diff.lag -0.038024   0.043779  -0.869    0.386

[DF test-statistic is -4.3718, for which critical values]
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62

Residual standard error: 0.02619 on 502 degrees of freedom
Multiple R-squared:  0.03792,Adjusted R-squared:  0.03409
```

Comparison to linear regression

OLS on simple differences Δr_{25Y} and Δr_{10Y} gives min variance relationship – cointegration plays a completely separate role.

```
lm(formula = diff(curve2.this$X25) ~ diff(curve2.this$X10) + 0)

                      Estimate Std. Error t value Pr(>|t|)
diff(curve2.this$X10)  0.74570    0.01581   47.16   <2e-16 ***

Residual standard error: 0.01657 on 504 degrees of freedom
Multiple R-squared:  0.8153, Adjusted R-squared:  0.8149

Residuals:
      Min        1Q     Median        3Q       Max
-0.081683 -0.010172 -0.002371  0.007629  0.050172

cor(diff(curve2.this$X25), diff(curve2.this$X10))
[1] 0.903719
```

Hedging ratio puzzle

What would you use as a hedging ratio for assets r_{10Y} and r_{10Y} in presence of cointegration between them?

Multiple Choice:

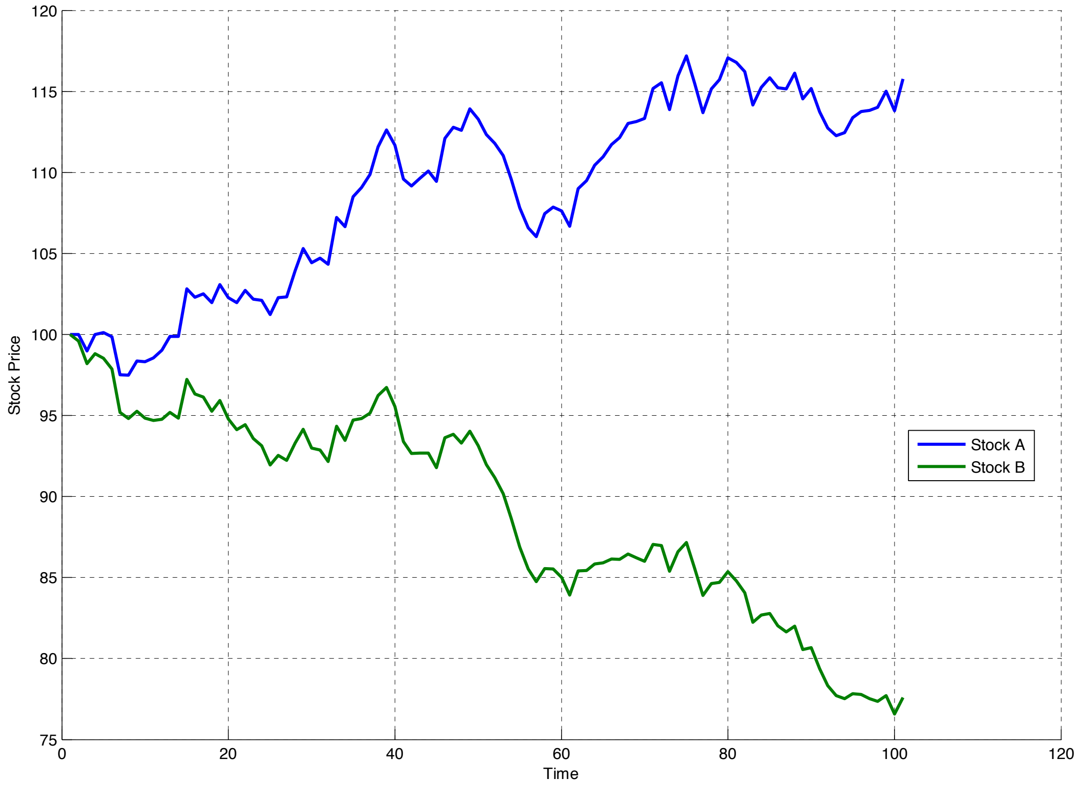
- 0.7698 from linear regression of r_{10Y} on r_{25Y}
- 0.7457 from linear regression on differences Δr_{25Y} on Δr_{10Y}
- 0.7516 from correction equation of Δr_{25Y}
- 0.7870 from a coint regression VECM output.

Time Series Cointegration

Extra Slides

Correlated Series

These time series are highly correlated but not cointegrated. Their spread possibly has an exponential fit.



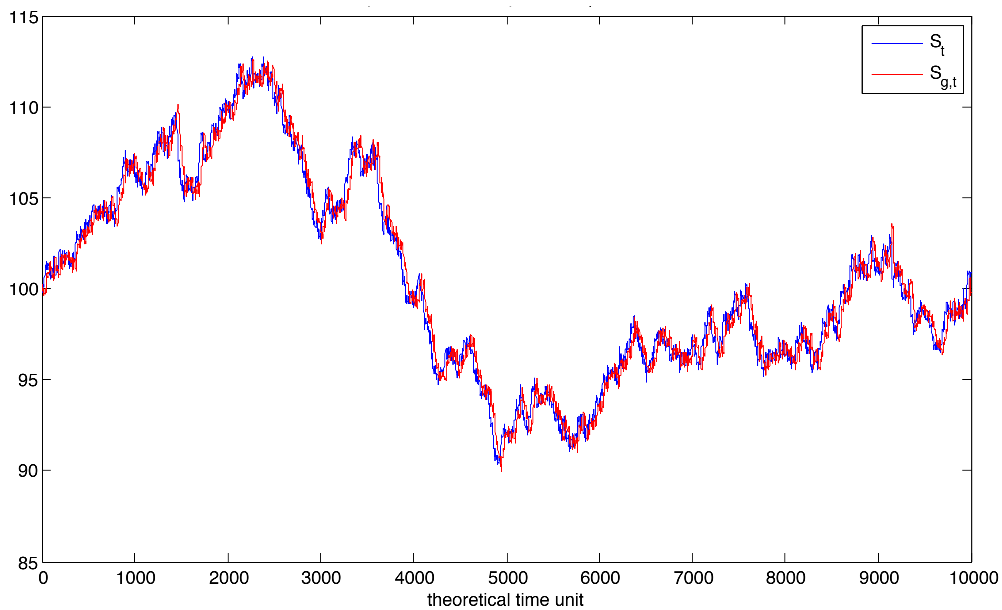
From *Correlation Sensitivity* CQF Lecture.

Cointelation

These series have been generated from the following processes:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t$$

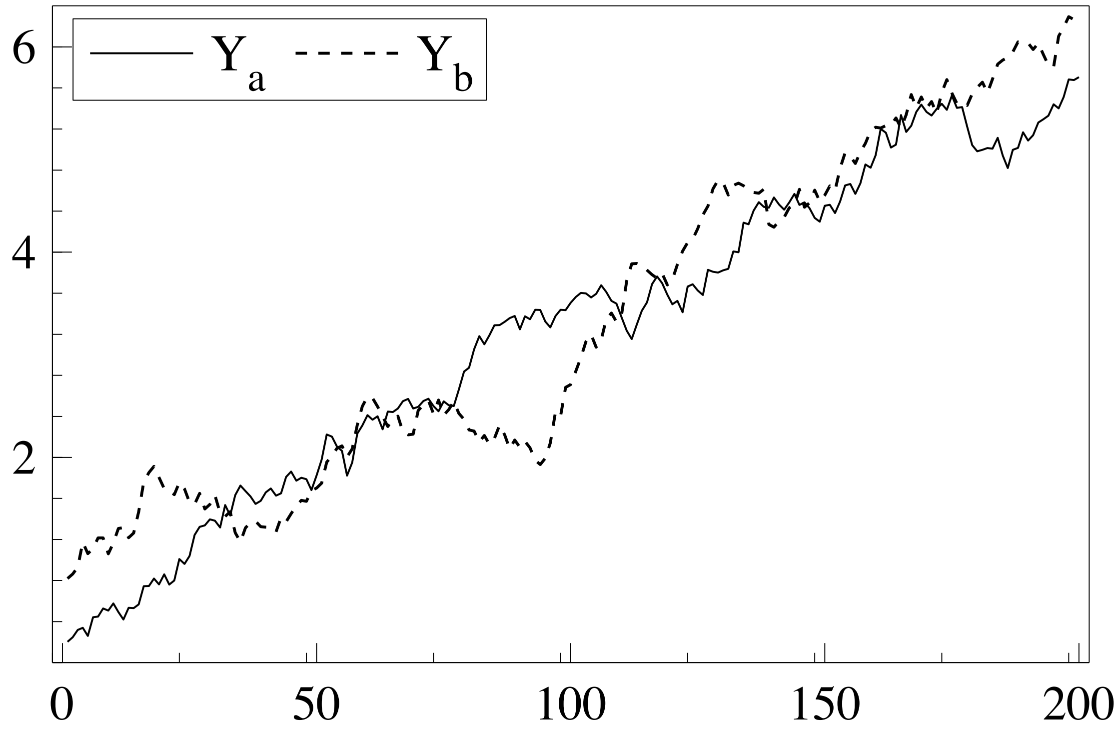
$$dS_{l,t} = -\theta(S_{l,t} - S_t)dt + \sigma S_{l,t}\,(\rho dW_t + \sqrt{1-\rho^2}dW_t^{\perp})$$



From Damghani (2014). *Introduction to the Cointelation Model*. CQF Extra

No linear equilibrium

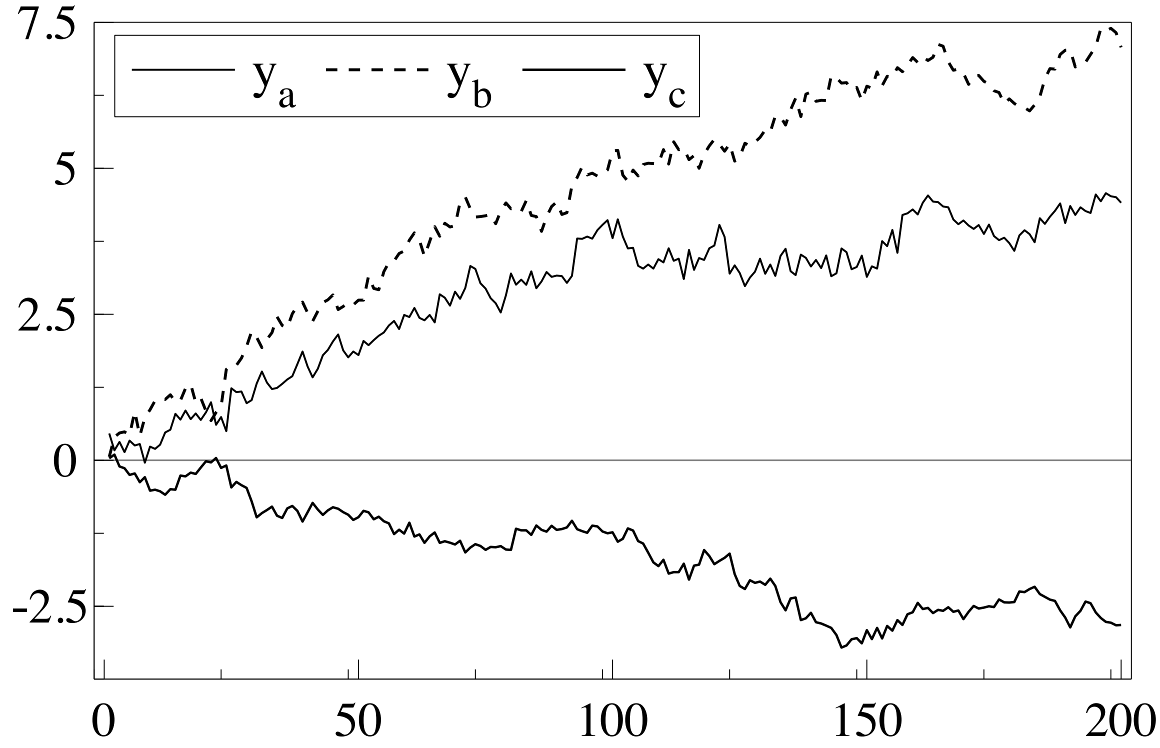
These series are **not** cointegrated. In fact, their spread contains a unit root. Multicollinearity is potentially irresolvable.



From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Cointegrated Series

These series are **cointegrated**. Their linear combination produces a mean-reverting spread (common factor) e_t.



From: Hendry & Juselius (2000). *Explaining Cointegration Analysis: Part II*

Cointegrated system

"There are fewer feedbacks than variables."

In a cointegrated system, **the common stochastic trend(s) drive all the related variables in the long-run**.

We are interested to trade **cointegrating residual** e_t which is

- Stationary (has no unit roots) $I(0)$
- Autoregressive $AR(1)$, **not** decomposable as $MA(\infty)$ series
- Mean-reverting $\theta \gg 0$

Common Factor: rates example

The linear combination $\boldsymbol{\beta}'_{Coint}\mathbf{Y}_t$ exposes a shared unit root, called *'a stochastic process in common'*.

Think of exposure to the common factor and what it could be.

- Cointegrating vector $[1, -\beta]$ gives hedging ratios for bonds.

- $Z(t;\tau_1) - \beta Z(t;\tau_2) = e_t$ is stationary $I(0)$

- Risk Factor: parallel shift of the yield curve.

A Multivariate Linear Combination

If a linear combination with some *special weights* $\boldsymbol{\beta}'_C$ produces **a stationary spread**:

$$\begin{aligned} e_t &= \boldsymbol{\beta}'_C \mathbf{Y}_t \qquad e_t \sim I(0) \\ &= \pm\beta_1 y_{1,t} \pm \beta_2 y_{2,t} \pm \cdots \pm \beta_n y_{n,t} \end{aligned} \tag{1}$$

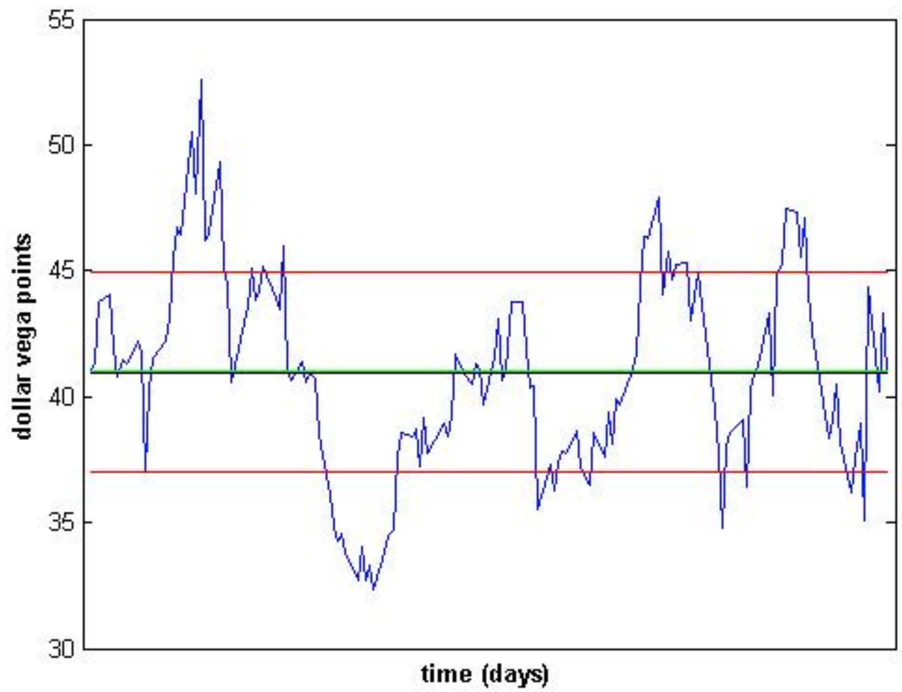
then we can explore mispricing that occurs when asset prices $y_{i,t}$ produce **a disequilibrium** $e_t \neq \mu_e$.

- The cointegration is *alike differencing* among time series.
- Left after the differencing is a **cointegrating residual** e_t. It is stationary $I(0)$ and mean-reverting $\theta > 0$.

Mean-reverting spread

The linear cointegrating combination $\boldsymbol{\beta}_C' \boldsymbol{Y}_t = e_t$ produces a stationary and mean-reverting spread:

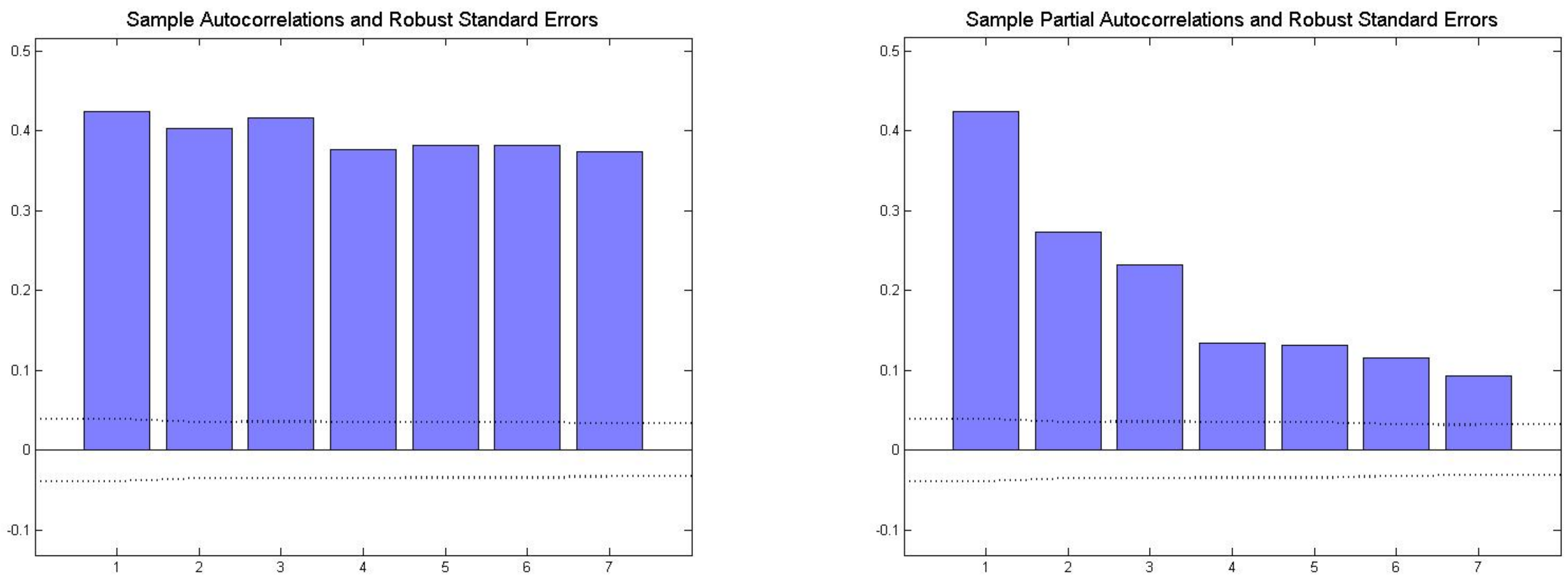
- Reversion speed $\theta \approx 44$ and bounds are calculated as $\sigma_{OU}/\sqrt{2\theta}$



From: Diamond (2013). *Learning and Trusting Cointegration*

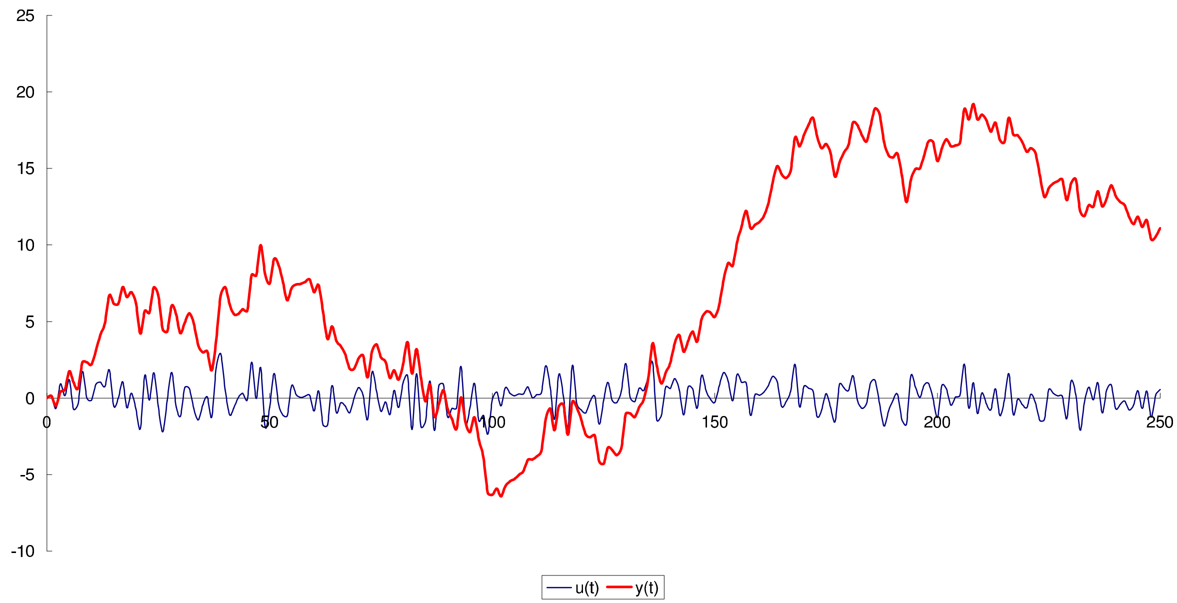
ACF and PACF for a high frequency spread e_t

Here, serial autocorreation $AR()$ noticeable for e_t.



The process is stationary but ACF has no exponential decay in autocorrelation Corr$[Y_t, Y_s]$.

ASIDE. Brownian Motion as Integrated Process

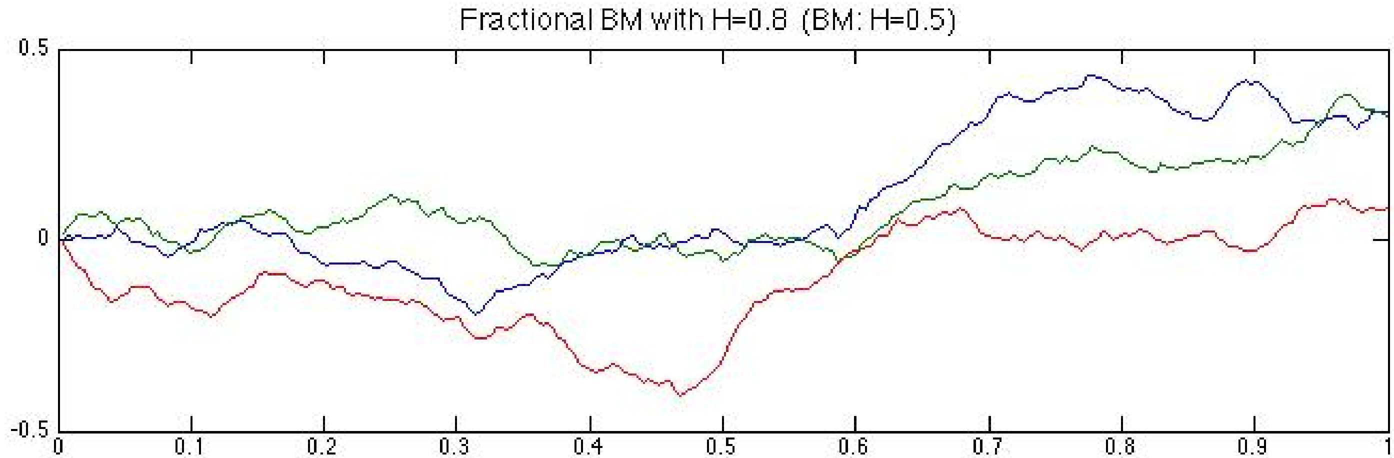


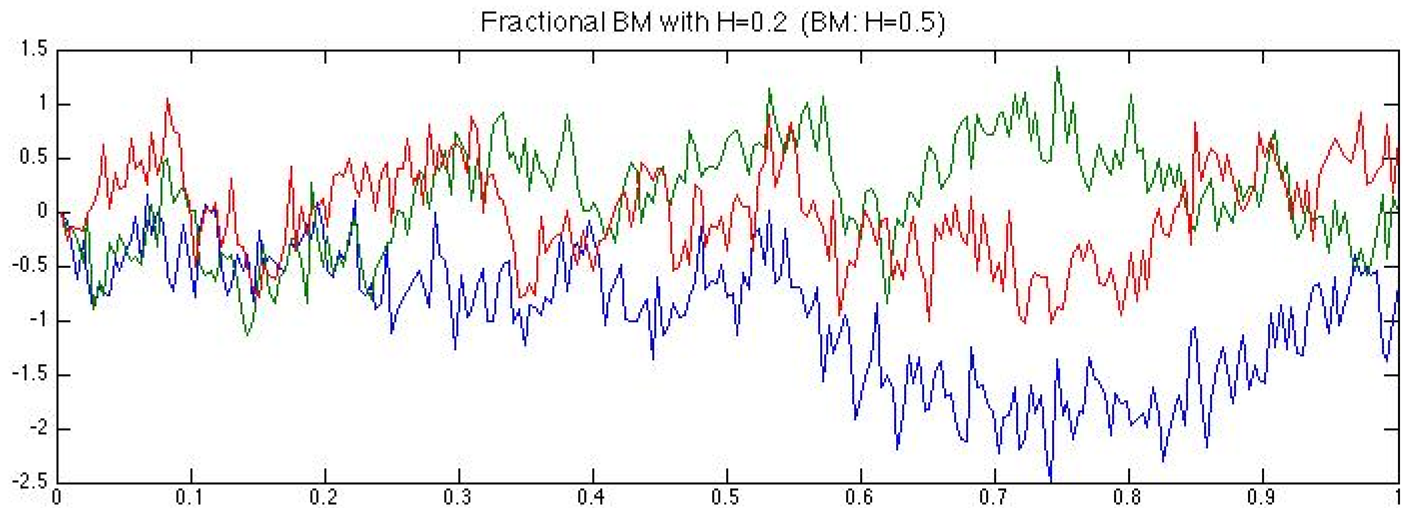
Regression residual $\epsilon_{t,\tau}$ is an increment of Brownian Motion and so, $Y_t = \sum \epsilon_s$ is the Brownian Motion under the unit root condition that, $\boldsymbol{\beta} = \mathbf{1}$.

$$\epsilon_{t,\tau} \stackrel{D}{=} \int_t^{t+\tau} \boldsymbol{\sigma} dW_s$$

It is integrated, stationarity test will confirm the unit root.

What if our common factor is **Fractional Brownian Motion**?





From: Algorithm credit to Yingchun Zhou and Stilian Stoev (2005)

Long memory

Fractional BM decays according to **the power law** τ^{2d-1} which is slower and delivers **long memory** for $H \gg 0.5$.

$$H = d + \frac{1}{2}$$

Hurst exponent $H > 0.5$ opens modelling of integrated series with long memory (think evolution of r_t and 1M LIBOR).

$H = 0.5$ recovers the Brownian Motion $\sim I(1)$.

Stationary-like series (simulated) have low values of Hurst exponent $H < 0.2$.

Stationary process $\beta < 1$ has exponentially decaying autocorrelations Corr $\approx e^{\tau \ln \beta}$. OU process has $e^{-\theta\tau}$.

END OF ASIDE

Dynamic Equilibrium in Econometrics

(for low-frequency, macro data)

Static Equilibrium Model

The familiar linear regression is **the** equilibrium model!

$$y = \beta_0 + \beta_1 x$$

In this *static*, stationary y_t and x_t produce **constant** b_g for

$$\Delta y = b_g \Delta x$$

The steady-state of equilibrium transpires through this constant growth rate β_g.

Static Equilibrium Model

CAPM a case of static equilibrium model! linear factor model. It relies on constant beta.

$$\mathbb{E}[r_I] = \beta\,(\mathbb{E}[r_M] - r_{rf}) + r_{rf}$$

$$\mathbb{E}[r_I - r_{rf}] = \beta\,\mathbb{E}[r_M - r_{rf}]$$

Since regression is involved, CAPM is also a Linear Factor Model.

Asset returns are regressed on Factors $\beta_j F_j$. The factors are linearly independent among themselves.

Equilibrium in STOCHASTIC Models

Assume that y_t and x_t are non-stationary time series **in levels** (e.g., prices/CDS/rates).

The static equilibrium model gives a short run relationship.

$$\Delta y = \beta_g \Delta x$$

Correlation is estimated among the differences...

What about the relationship in the long run?

If there is a common factor, it must affect *the changes* in y_t.

The same principle as with portfolio factor models (HML, SMB):

we regress returns (differences) on the common factor

$$\Delta y = \beta_g \Delta x + \underbrace{\text{Factor Term}} + \ldots + \epsilon_t$$

It turns out that the common factor is

$$\hat{e}_t \;=\; y - \hat{b}x - \hat{a}$$

$$\Delta y \approx \Delta x \qquad\qquad \text{and} \qquad \Delta y \approx (y - \hat{b}x)$$

s.t. $\hat{e}_t$ being stationary so that $[1, b]$ is a co-integrating vector.

Equilibrium Correction Model

The model addresses both, the short-run correlation-like $\beta_1 \Delta x_t$, and equilibrium correction working (slowly!) over the long-run

$$\Delta y_t \;=\; \beta_1 \Delta x_t - (1-\alpha)\left(y_{t-1} - b_e x_{t-1} - a_e\right) + \epsilon_t$$

where $e_{t-1} = y_{t-1} - b_e x_{t-1} - a_e$ and $\mathbf{E}[e_{t-1}] = a_e$

The disequilibrium $e_{t-1} \neq a_e$ is corrected over the long-run.

The speed of correction $-(1-\alpha)$ is inevitably small, but must be significant for cointegration to exist.

Modelling problems

$$\Delta y_t = \beta_1 \Delta x_t - (1-\alpha)e_{t-1} + \epsilon_t$$

- The assumption of x_t **being leading/exogenous/causing** variable.

- Equilibrium-correction mechanism is **linear**: if the 'error' e_{t-1} above μ_e the model suggests a small correction downwards (and vice versa).

- Non-unique cointegrating a, b are empirically possible so the speed of correction becomes **a calibrated parameter**

Estimating Cointegration - Pairwise

- **Pairwise Estimation**: select two candidate time series and apply ADF test for stationarity to the joint residual.

  Use the estimated residual to continue with the Engle-Granger procedure.

  Perform the Engle-Granger procedure in both ways,

  $$\Delta y_t \quad \text{on} \quad \Delta x_t$$

  $$\Delta x_t \quad \text{on} \quad \Delta y_t$$

Cointegration Case B offers R code that re-implements the ECM estimation explicitly. Then, VECM estimation routine is used to analyse further.

Engle-Granger procedure

Step 1. Obtain the fitted residual $\widehat{e}_t = y_t - \widehat{b}x_t - \widehat{a}$ and test for unit root.

- That *assumes* cointegrating vector $\boldsymbol{\beta}'_{Coint} = [1, -\widehat{b}]$ and equilibrium level $\mathbb{E}[\widehat{e}_t] = \widehat{a} = \mu_e$.

- **If the residual non-stationary** then no long-run relationship exists and regression is spurious.

Step 2. Plug the residual from Step 1 into the ECM equation and estimate parameters β_1, α (with linear regression)

$$\Delta y_t = \beta_1 \Delta x_t - (1-\alpha)\widehat{e}_{t-1}.$$

- It is required **to confirm the significance for** $(1-\alpha)$ coefficient.

Equilibrium Correction Mode (review)

The model addresses both, the short-run correlation-like $\beta_1 \Delta x_t$, and equilibrium correction working (slowly!) over the long-run

$$\Delta y_t \;=\; \beta_1 \Delta x_t - (1-\alpha)\,(y_{t-1} - b_e x_{t-1} - a_e) + \epsilon_t$$

$$\text{where } e_{t-1} = y_{t-1} - b_e x_{t-1} - a_e \quad \text{and} \quad \mathbf{E}[e_{t-1}] = a_e$$

The disequilibrium $e_{t-1} \neq a_e$ is corrected over the long-run.

Equilibrium-correction mechanism is **linear**: if the 'error' e_{t-1} above μ_e the model suggests a small correction downwards (and vice versa).

The speed of correction $-(1-\alpha)$ is inevitably small, but must be significant for cointegration to exist.

Non-unique cointegrating a, b are empirically possible so the speed of correction becomes **a calibrated parameter**

Stationarity Tests: reference

There are several ways of testing for unit root (the process being non-stationary).

- We focused on **Dickey Fuller** test that keeps familiar AR(p) model and is popular because of its simplicity and generality.
- Non-parametric *Phillips-Perron* test transforms t-statistic to further account for autocorrelation and heteroscedasticity.
- *Cointegration Regression Durbin-Watson* replaces Engle-Granger procedure (presented next). It contains of one simplistic regression (one price on another) and Durbin-Watson test on the residuals.

Statistical Arbitrage

A Few Fundamentals

Statistical arbitrage fundamentals

Makes two claims that **a.** relative mispricing persists and **b.** pricing inefficiencies are identifiable with statistical models.

The product of hedging is a hedging error, and the manageable error behaves like **the cointegrated residual** $e(t)$.

- Otherwise, the hedging error behaves like a random walk (unbounded) due to the unit root in $\boldsymbol{Y_t}$ when $\beta \approx 1$, common to all financial time series in levels.

Quality of mean-reversion

Q: How do we evaluate the quality of mean-reversion and find out entry/exit trade points?

A: We fit the spread to the Ornstein-Uhlenbeck process.

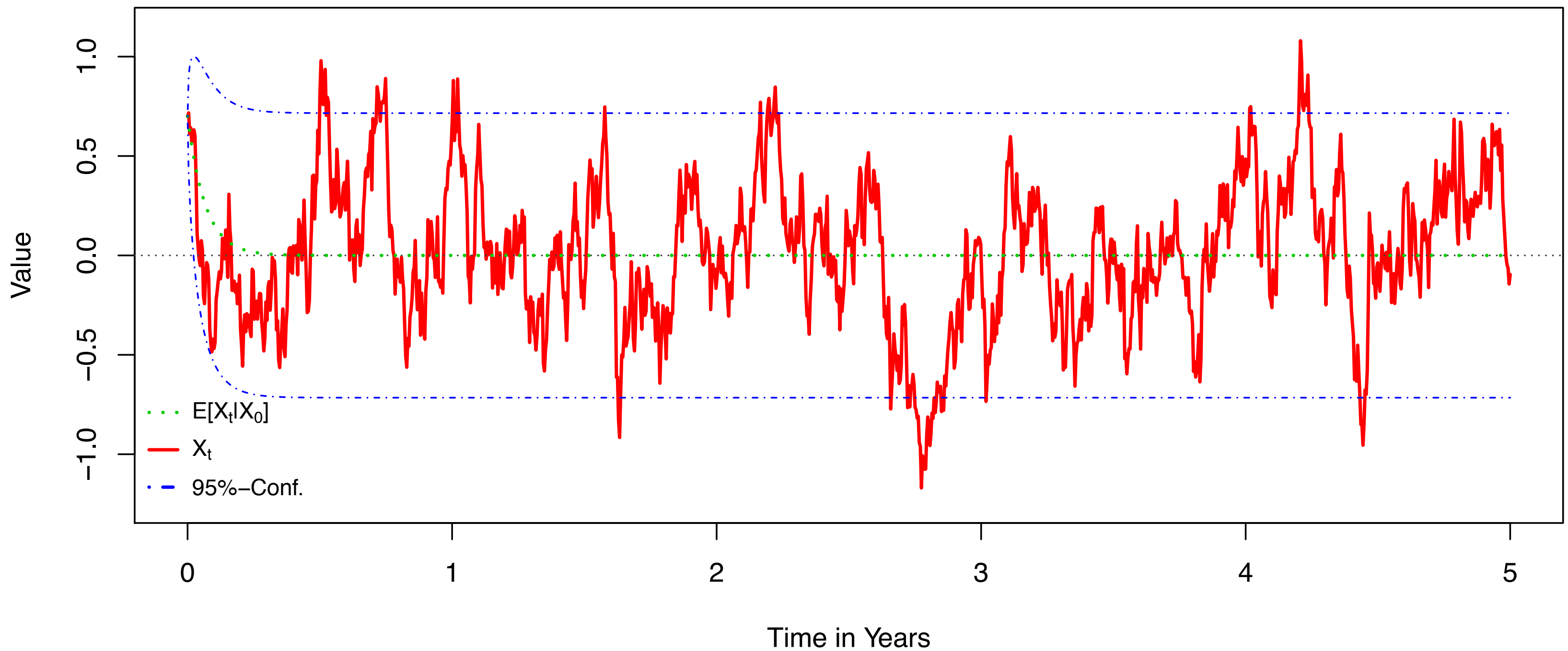
- Quality of mean-reversion is associated with the higher critical values equilibrium-correcting term (empirically).

- Cointegration is a filter on data: mean-reversion is of lower frequency than the data.

E.g., 10 Min data can generate half-life counted in weeks.

Simulated Ornstein-Uhlenbeck process

Here is how the simulated OU process looks like (sample path)



From: Harlacher (2012). *Cointegration Based Statistical Arbitrage*

Mean-reverting but has a different kind of stationarity than an AR(1) process. Why? There is diffusion.

Designing a trade

In order to design an arbitrage trade, one requires the following items of information:

1. **Weights** β'_{Coint} for a set of instruments to obtain the spread.

2. **Speed of mean-reversion** in the spread. For explanation purposes, the speed can be presented as **half-life**, the time between the equilibrium situations, when spread $e_t = \mu_e$.

$$\theta \to \tau$$

3. **Entry and exit levels** defined by σ_{eq}. Optimisation involved.

The inputs allow to backtest P&L and estimate drawdowns.

CQF: Certificate in Quantitative Finance

Credit Default Swaps

17th June 2020

Dr Jon Gregory
jon@solum-financial.com

Content

- CDS and credit spreads
- Definition of a CDS
- Pricing CDS
- Counterparty risk
- CDS indices
- Proxy CDS curves

CDS and Credit Spreads

FitchLearning

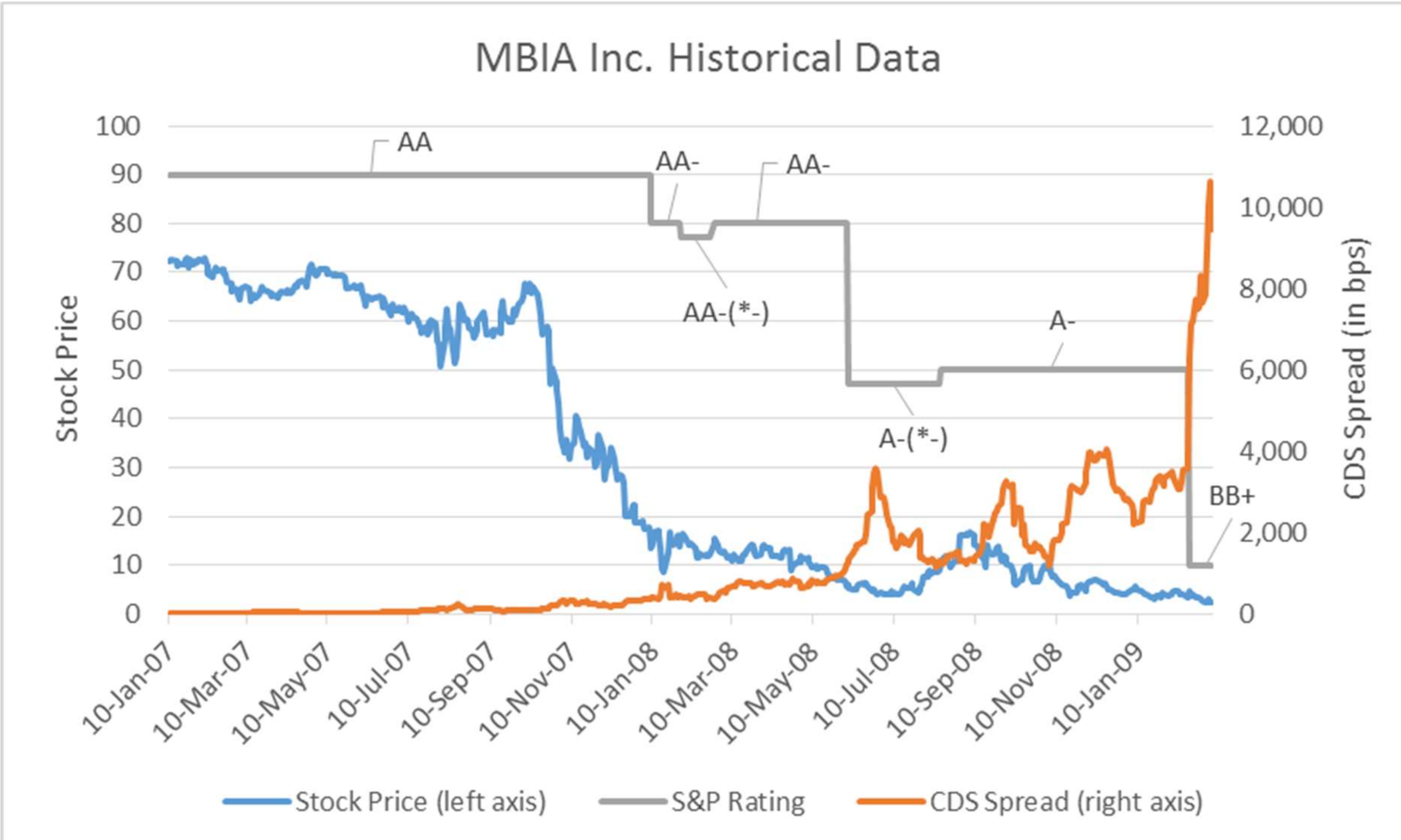
CDS Spreads

- Reflect the cost of credit risk
- Compared to ratings, may be more forward-looking
- These quotes are in basis points (bps) per annum

CNBC MARKETS BUSINESS INVESTING TECH POLITIC

*DUBAI CDS 5YR	203.254
*EGY CDS 5YR	557.52
*FIN CDS 5YR	18.416
*FRA CDS 5YR	30.01
*GER CDS 5YR	18.571
*GRE CDS 5YR	189.683
*HUN CDS 5YR	76.755
*INA CDS 5YR	136.619
*IRE CDS 5YR	34.593
*ITA CDS 5YR	191.137
*JPN CDS 5YR	22.298
*KOR CDS 5YR	25.267
*NED CDS 5YR	16.747
*PAN CDS 5YR	97.366
*POR CDS 5YR	76.566
*SVK CDS 5YR	52.303
*ESP CDS 5YR	81.863
*SWE CDS 5YR	13.884
*UK CDS 5YR	29.928

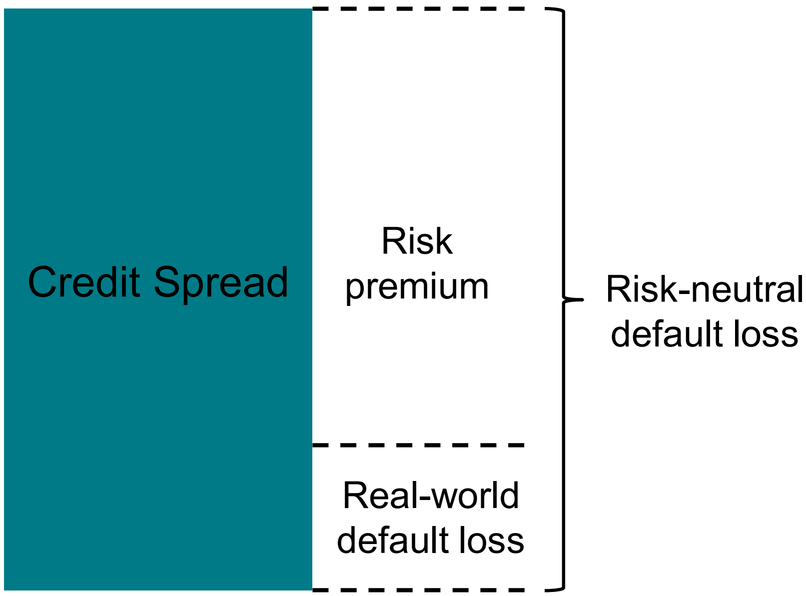
Example – Stock Price, Rating and Credit Spread



Market Implied or Real World?

Credit spreads compared to historical default probabilities

- ‘Credit risk premium’ is substantial



	Real world loss (bps)	Risk neutral loss (bps)	Ratio
Aaa	4		
Aa	6		
A	13		
Baa	47		
Ba	240		
B	749		
Caa	1690		

Source: Hull, J., M. Predescu and A. White, 2004

Credit Default Swap: Aim

- A product to buy or sell protection against a 'credit event' such as a default

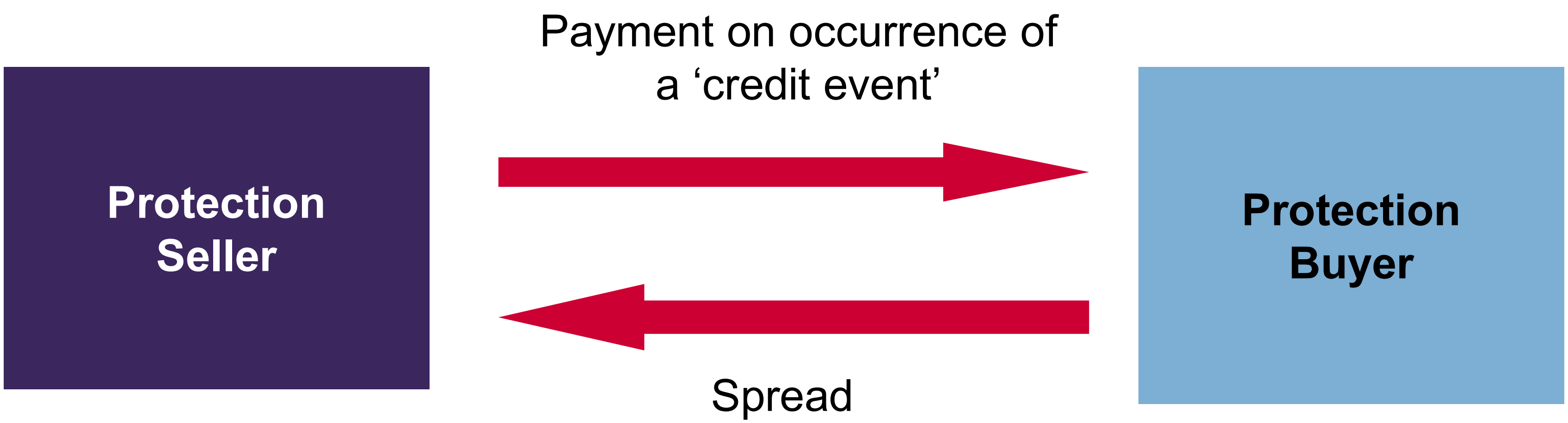
Uses:

- Buy protection to hedge credit risk on existing credit exposure:
  - Loans
  - Bonds
  - Other credit exposure (e.g. derivatives)
- Sell or buy protection to create a synthetic long or short position in particular counterparty/group of counterparties (portfolio management)
- Note that naked CDS positions have seen regulatory scrutiny
  - For example, in the EU buying 'naked' CDS protection on sovereign names is banned

Definition of a CDS

FitchLearning

CDS Structure



- Spread is usually paid quarterly and is quoted in basis points (bps) per annum
- Usually CDS trade with a fixed spread (100 bps) and upfront payments
  - This makes them more standardised
- The market still quotes on an all running spread basis

CDS: Credit Event Definitions

- Bankruptcy
- Failure to pay
- Obligation cross default (or acceleration)
- Repudiation/moratorium
- Restructuring

Recovery Rates and CDS Payoff

The loss given default (LGD) is the actual % loss rate experienced by a lender on a credit exposure if the borrower defaults

It is given by one minus the **recovery rate (RR)** and can take any value between 0% and 100%. Formally:

$$LGD = 1 - RR$$

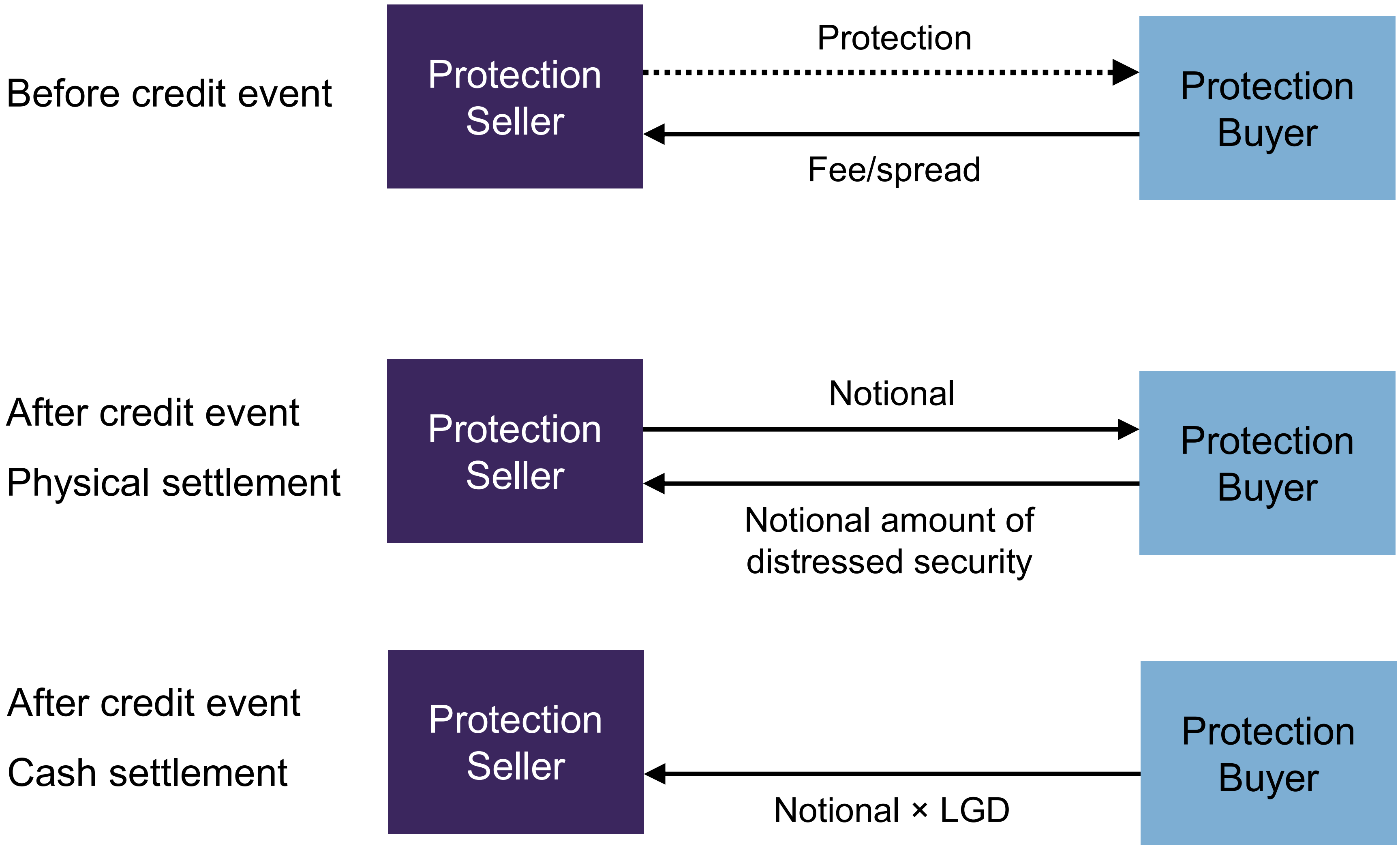
The LGD is never known when a new loan is issued, but is seen in the secondary market after the default (e.g. Lehman Brothers bonds traded at around 10% after default)

It therefore makes sense for the payoff of a CDS contract to be equal to the LGD

For example, buy a bond and hedge by buying CDS protection:

	Scenario 1	Scenario 2	Scenario 3
Bond	40	80	10
CDS	60	20	90
Total	100	100	100

Physically or Cash Settled CDS Structures

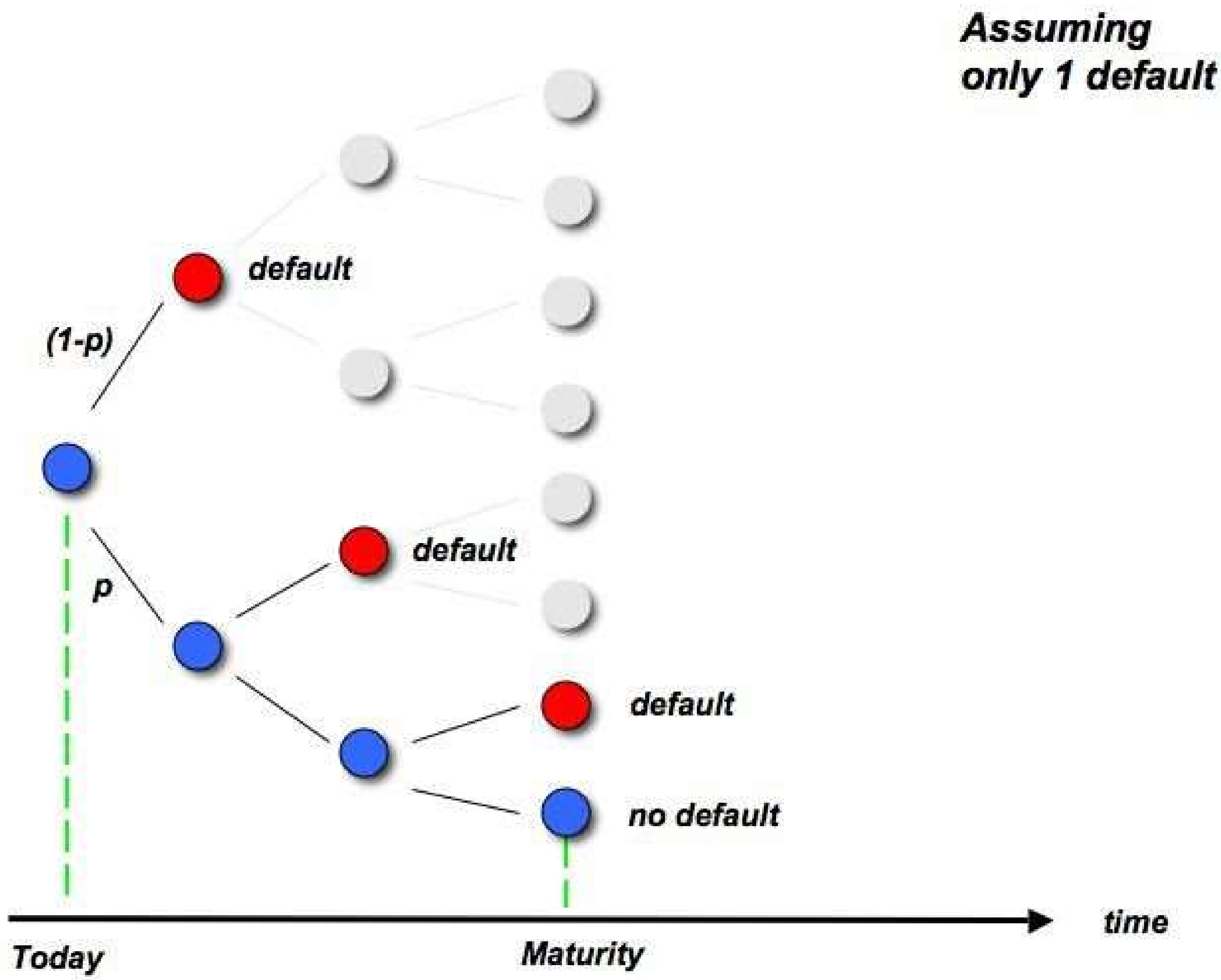


CDS Settlement

- Settlement process has a number of potential problems
  - Cheapest-to-deliver optionality (Conseco Corp, Railtrack restructurings 2001)
  - Delivery squeeze (Parmalat 2003, Delphi 2005), Lehman had $400 billion of (gross) CDS notional and $155 billion of debt
- Most large credit events are now settled by an Auction

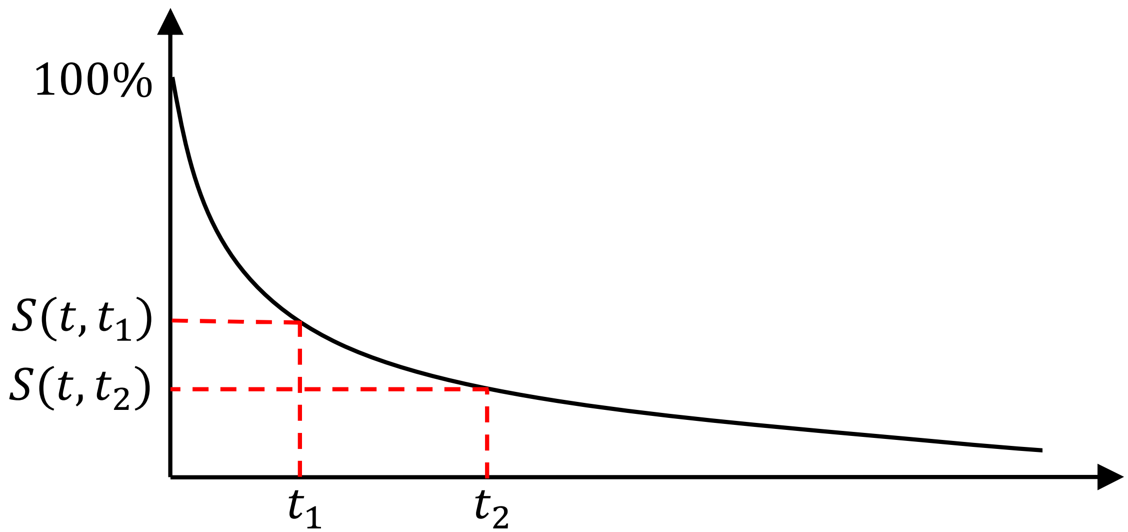
CDS Pricing

FitchLearning


Assuming
only 1 default
default
(1-p)
default
p
default
no default
time
Today
Maturity

CDS Valuation (I)

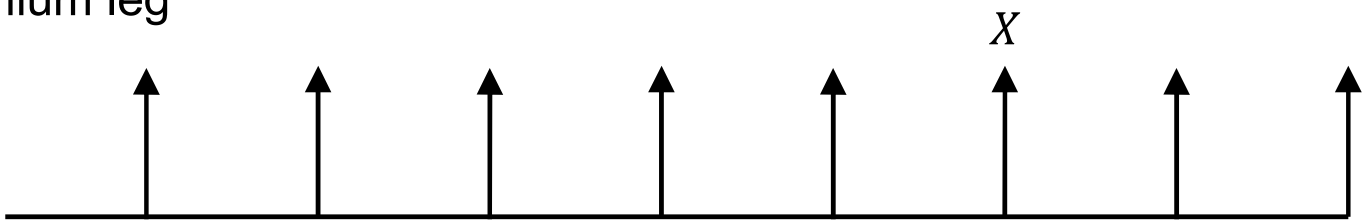
- A CDS can be split into two legs
  - Premium payments that occur contingent on no prior default (aka credit event)
  - A default payment that occurs contingent on a credit event
  - (there is also an accrued premium payment that occurs when there is a default payment)
- Generally it is assumed that interest rates, defaults and LGDs are independent
- Define the survival probability between time t and time u as $S(t,u)$



- Discount factors $B(t,u)$

CDS Valuation (II)

- Premium leg



$$PV^{X}_{premium}(t) = X \sum_{i=1}^{n} B(t, t_i) \times S(t, t_i)$$

- Default leg

$$PV_{default}(t) = LGD \int_{t}^{T} B(t, u) dS(t, u) \approx LGD \sum_{j=1}^{m} B(t, t_j)\left[S\left(t, t_j\right) - S\left(t, t_{j-1}\right)\right]$$

Mathematical Model for Default

- Default can be modelled as a Poisson process
- The process is driven by an intensity or hazard rate h
- Probability of default in an infinitively small period conditional on survival is hdt
- Probability of no default is given by $S(t,u) = \exp[-h(u-t)]$
- In order to fit a term structure of CDS quotes then we need a (deterministic) functionality form for the hazard rate

$$S(t,u) = \exp\left[-\int_t^u h(x)dx\right]$$

Simple Formulas

- Assuming a single deterministic hazard rate and interest rate:

$$PV^1_{premium} = \sum_{i=1}^{n} B(t,t_i)S(t,t_i) \approx \int_t^T \underbrace{\exp\big(-(r+h)(u-t)\big)}_{\text{Risky discount factor}} du = \frac{1-\exp(-(r+h)(T-t))}{(r+h)(T-t)}$$

$$PV_{default} = LGD \int_t^T B(t,u)dS(t,u) = -\,LGD \int_t^T h\exp\big(-(r+h)(T-t)\big)$$

$$= LGD.h.\frac{1-\exp(-(r+h)(T-t))}{(r+h)(T-t)}$$

- Leads to the approximate relationship between CDS spread and hazard rate

$$s_{CDS} = \frac{PV_{default}}{PV^1_{premium}} \approx h \times LGD \qquad h \approx \frac{s_{CDS}}{LGD}$$

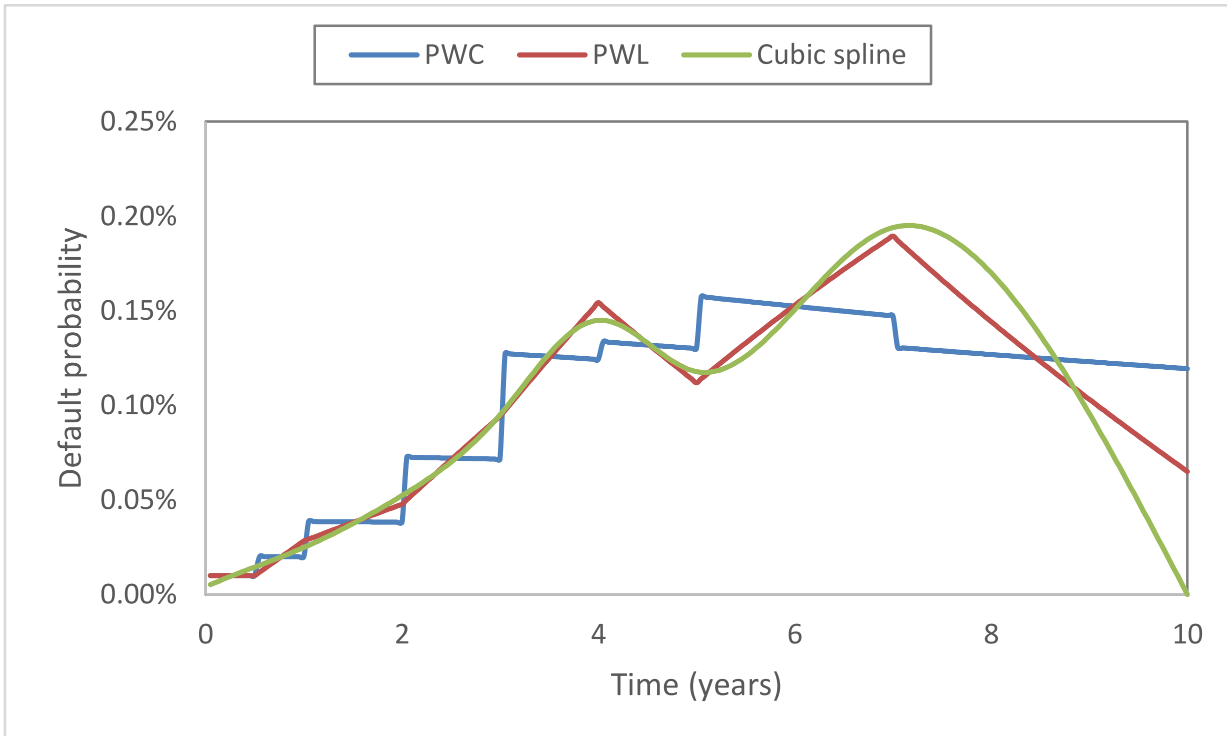
Calibration to CDS Curve (Excel example)

- Approximate method assuming a constant hazard rate (credit spread curve)

$$S(t,u) = \exp[-s_u^{CDS}(u-t)/LGD]$$

- Accurate by fitting a hazard rate function
  - Piecewise constant
  - Piecewise linear
  - Cubic spline

	CDS
6M	0.12%
1Y	0.18%
2Y	0.32%
3Y	0.50%
4Y	0.75%
5Y	0.92%
7Y	1.20%
10Y	1.35%

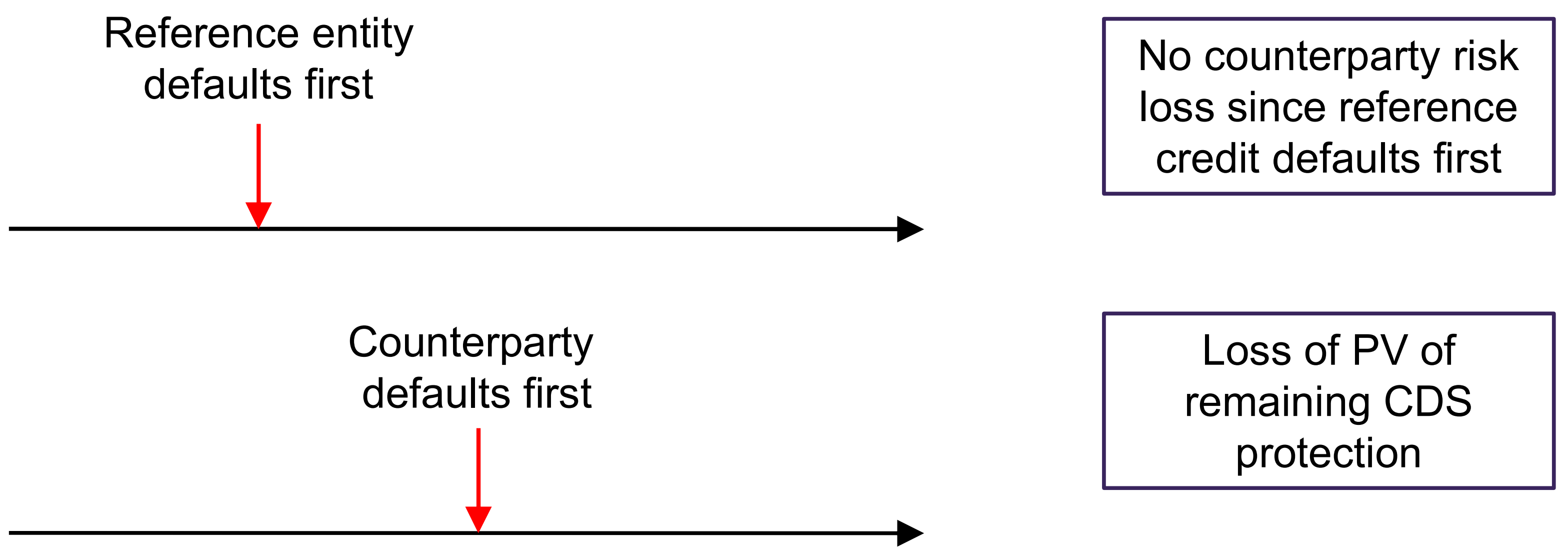


Counterparty Risk

FitchLearning

Counterparty Risk in CDS

- Classic case of wrong-way risk when buying credit protection
  - Payoff of CDS contract is determined by default of reference entity
  - But counterparty must not be in default
  - Dependency (e.g. default corrrelation) is very important



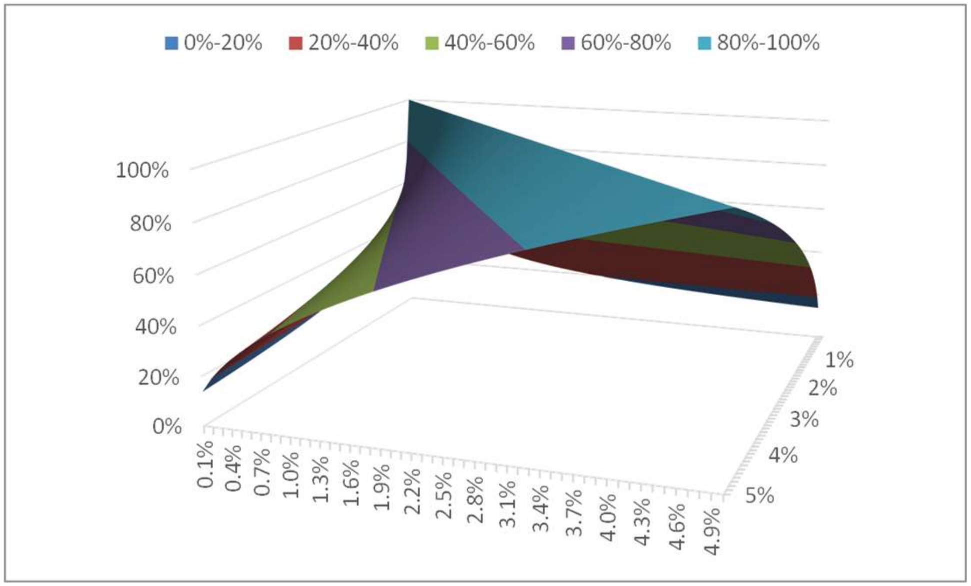
Default Correlation (I)

- Joint default probability p_{12}
- Clearly need to model default correlation (counterparty and reference entity)
- Default correlation definition

$$\rho_{12} = \frac{p_{12} - p_1 p_2}{\sqrt{p_1(1-p_1)p_2(1-p_2)}}$$

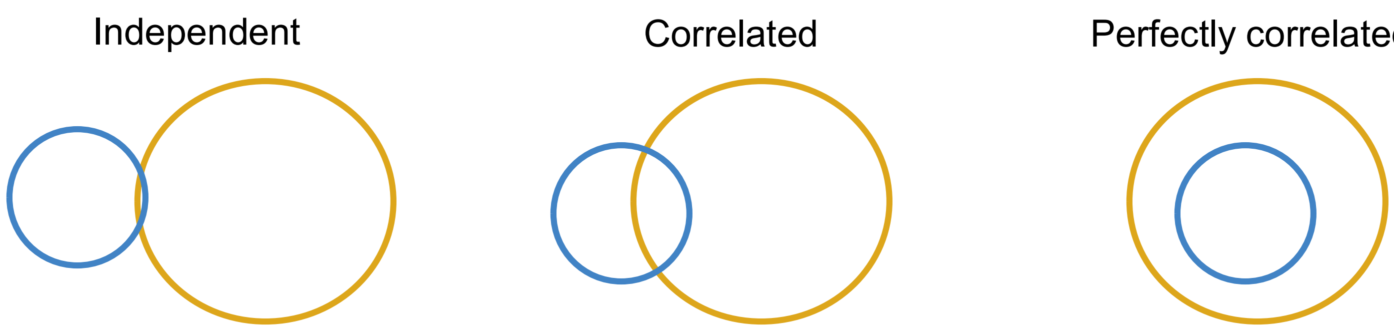
- Since $p_{12} \le p_1$

$$\rho_{12} \le \sqrt{\frac{p_1(1-p_2)}{p_2(1-p_1)}}$$



Default Correlation (II)

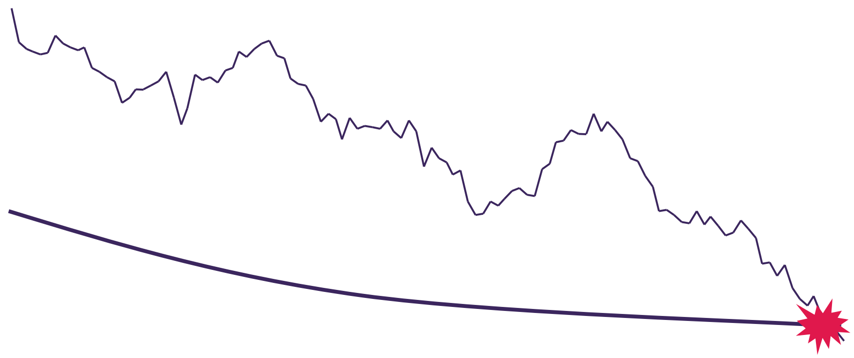
- For non-equal default probabilities, there is a implicit maximum default correlation



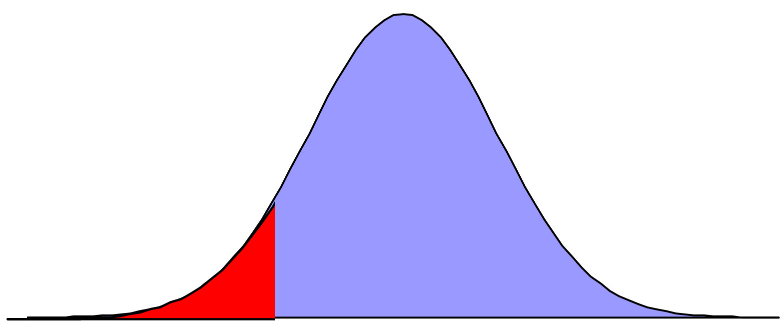
- Concept of default correlation is hard to deal with
- Calibration of parameters
  - Sparse data
  - Would be limited to some classification (e.g. by rating/sector/region)
- Problem to build a reasonable model
  - No economic structure
  - Specification of joint distribution (probability of many defaults)

The Merton Model

- Merton [1974], value of the firm follows a geometric Brownian motion



- Default occurs if this value falls below a certain level (value of assets < liabilities)
- Now defaults are driven by a Gaussian variable



Gaussian Approach to Default Modelling

- Default of a single name is determined by Gaussian variable V_1

$$V_1 < \Phi^{-1}(p_1)$$

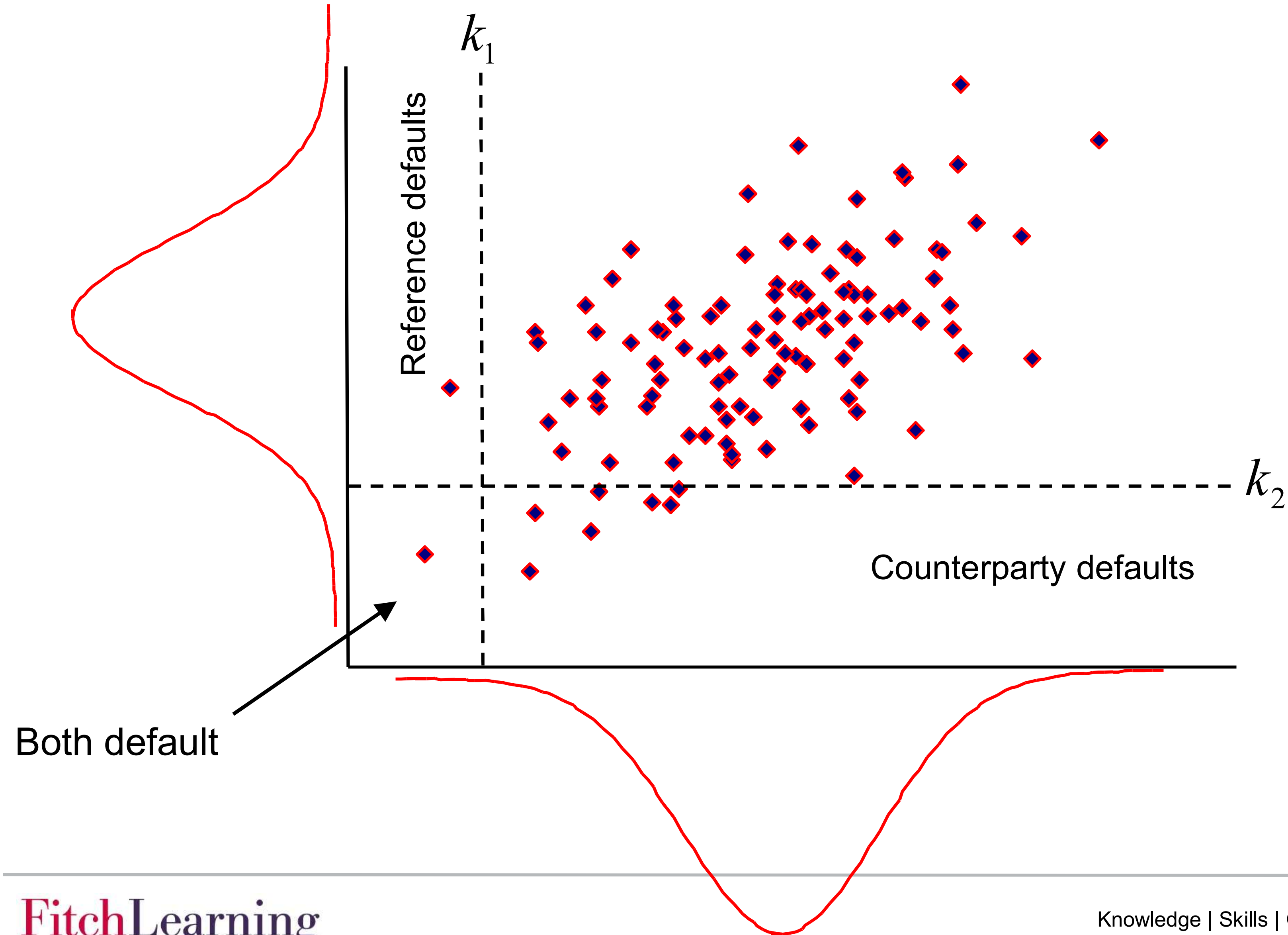
- Default of second correlated name

$$V_2 = \rho_{12} V_1 + \sqrt{1 - \rho_{12}^2}\tilde{V} < \Phi^{-1}(p_2)$$

- Joint default is given by bivariate gaussian distribution

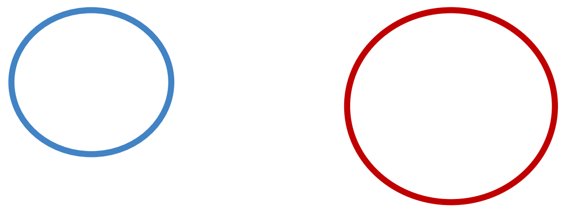
$$\Phi_{2d}(\Phi^{-1}(p_1), \Phi^{-1}(p_2); \rho_{12})$$

Joint Defaults

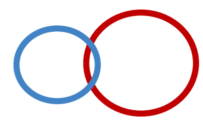


Gaussian Approach to Default Modelling - Example

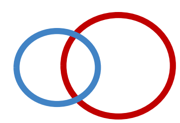
$p_1 = 1\%$ $\quad$ $p_2 = 1\%$



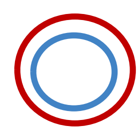
$\rho_{12} = 0\%$ $\quad$ $p_{12} = 0.02\%$



$\rho_{12} = 50\%$ $\quad$ $p_{12} = 0.21\%$

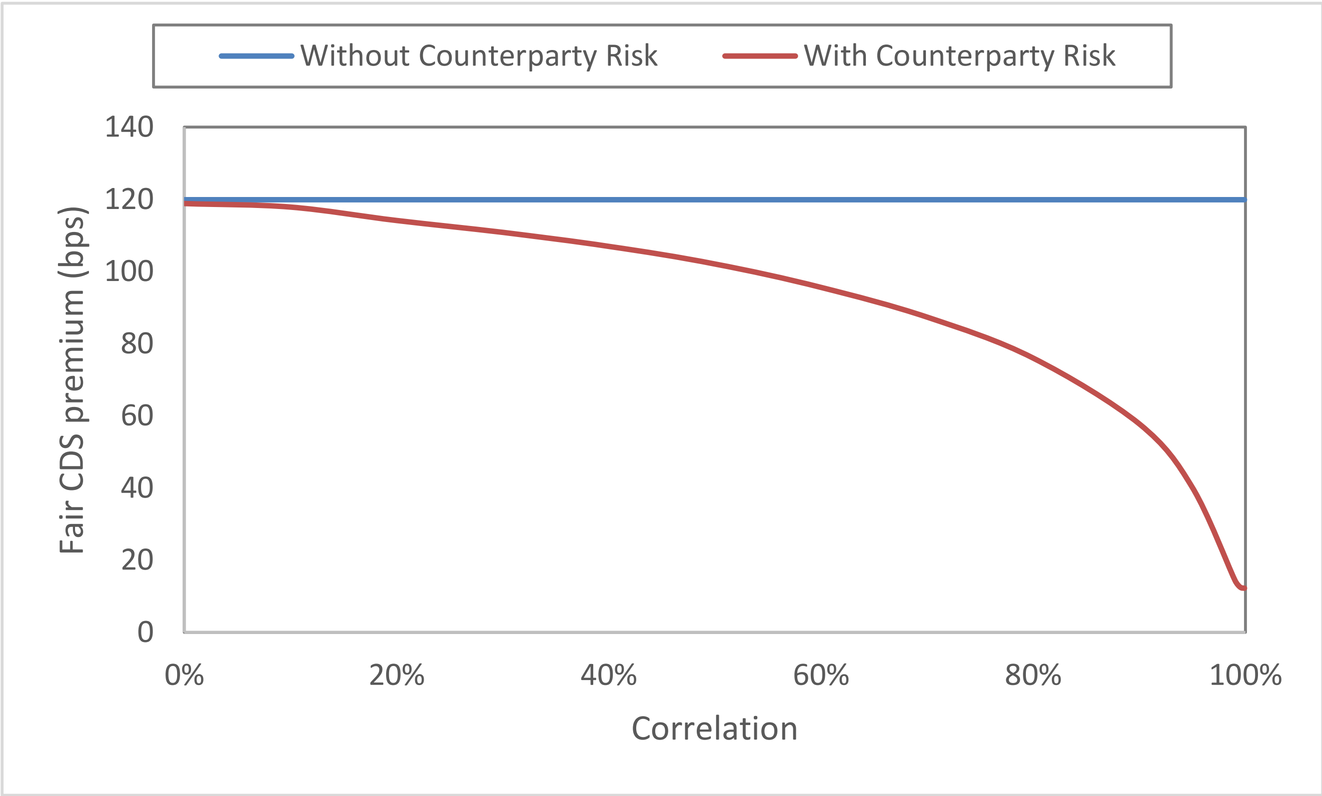


$\rho_{12} = 100\%$ $\quad$ $p_{12} = 1\% = p_1$



CDS Counterparty Risk – Example 1

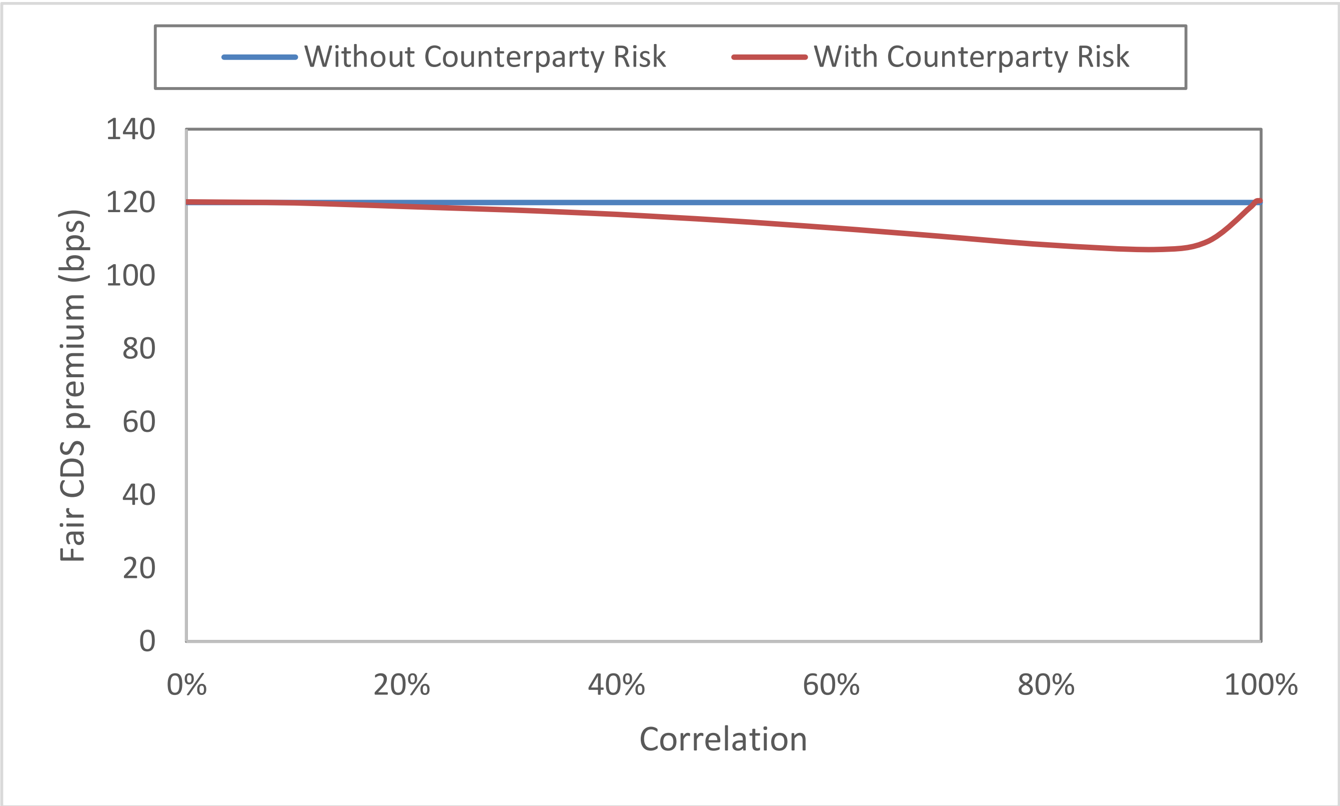
- What would you pay to buy CDS on a reference entity with a spread of 1.2% from a counterparty with a spread of 2.4% (recovery rate at 10%)



Gregory J., 2011, "Counterparty risk in credit derivative contracts", The Oxford Handbook of Credit Derivatives, A. Lipton and A. Rennie (Eds), Oxford University Press.

CDS Counterparty Risk – Example 2

- What would you pay to buy CDS on a reference entity with a spread of 1.2% from a counterparty with a spread of 0.5% (recovery rate at 10%)



CDS Counterparty Risk

'Monoline' insurers at risk of losing AAA credit rating

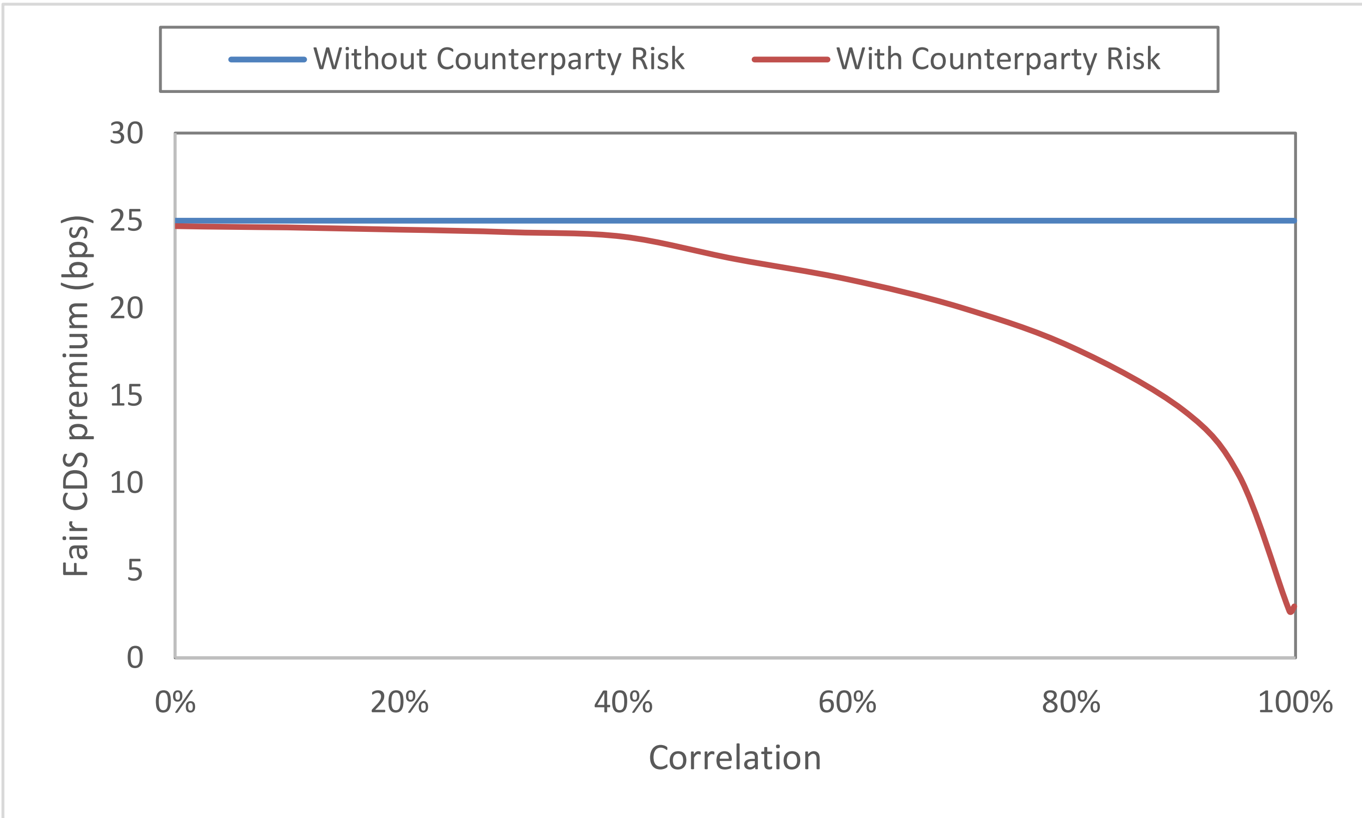
- Has caused many problems around the time of the global financial crisis

	CVA Charges						CVA Charges	Mortgage
(Amounts in $ billions)	**2007**	**2008**	**2009**	**2010**	**Total**	**2006 Equity**	**/ Equity**	**Losses***
Bank of America Merrill Lynch	$3.3	$11.3	$0.9	$0.0	$15.5	$174.3	8.9%	$56.2
Bank of America	0.2	0.9	0.9	0.0	2.0	135.3	1.5	16.4
Merrill Lynch	3.1	10.4	0.0	0.0	13.5	39.0	34.6	39.8
Barclays	0.0	0.6	0.8	0.0	1.4	72.3	2.0	13.3
Citigroup	1.0	5.7	1.3	(0.5)	7.5	119.8	6.2	42.6
Credit Suisse	(0.0)	0.6	0.1	(0.1)	0.5	35.7	1.5	11.1
Deutsche Bank	0.1	1.8	0.0	0.2	2.1	43.3	4.9	6.9
Goldman Sachs	0.0	0.0	0.0	0.0	0.0	31.0	0.0	4.6
J.P. Morgan	0.0	0.0	0.0	0.0	0.0	115.8	0.0	2.0
Morgan Stanley	0.0	1.9	0.2	0.9	3.0	35.4	8.4	3.2
Royal Bank of Scotland	1.7	6.6	4.0	(0.3)	12.0	103.9	11.6	19.8
Societe General	1.3	1.8	0.6	0.0	3.7	44.0	8.4	11.1
UBS	0.7	7.0	0.7	(0.7)	7.8	40.7	19.1	33.1
Total	**$11.4**	**$48.7**	**$9.5**	**($0.5)**	**$53.6**	**$816.2**	**6.6%**	**$203.9**

- One of the reasons why all CDS contracts are collateralised and are also being centrally cleared by central counterparties (CCPs)
  - When buying protection, how much does collateral held (clue: Lehman Brothers defaulted from a CDS spread of about 7%)
  - Should a CCP clear a trade where a bank sells protection on their own sovereign?

CDS Counterparty Risk – Example 3

- What would you pay to buy CDS on a reference entity with a spread of 0.25% from a counterparty with a spread of 0.5% (recovery rate at 10%)



CDS Indices

FitchLearning

CDS Indices

- A credit index is a combination of single name CDS
  - Usually equally weighted
  - 125 names in the most liquid iTraxx Europe and CDX North America indices
- Generally much more liquid than the single name equivalents

Counterparty	Rating	Index
Corporates	BBB and better	iTraxx / CDX Non-Financials
	BBB and below	iTraxx / CDX Crossover / HY
Financials		iTraxx /CDX Financials
Sovereigns		SovX

CDS Indices – Uses

- Gain exposure to corporate credit
  - Participants can easily gain exposure to a diversified portfolio of credits with much smaller transaction and operational costs than entering the single name market
- Hedge credit risk
  - Indices are an efficient and liquid method to hedge a portfolio's overall macro credit risk via a transparent and standardised portfolio
- Provide a benchmark for development of other credit products
  - Index options, index tranches
- Enhance liquidity in the single name market
  - The liquidity of the index flows into the single name CDS market

CDS Indices – Arbitrage

- Buying protection on a CDS index is very similar to buying 1/n protection on all of the constituent names
- The difference is
  - In the index case we pay a fixed premium for all CDS names
  - In the single name case we pay different premiums for each CDS name (typically pay more for names more likely to default)
  - This means the index spread is slightly lower than the average CDS spread
- Bid-offer
  - Bid-offer on single names can easily be several basis points
  - Liquidity on the index means bid-offers of ½ or ¼ of a basis point

Proxy CDS Curves

Rationale – CVA Capital Requirements

Not all counterparties have traded credit spreads. However, the FRTB-CVA framework must capitalise CVA risk arising from dealing with all counterparties, including ones that are not actively traded in credit markets ("illiquid counterparties"). Therefore, in order to use the FRTB-CVA framework, a bank is required to have a methodology for approximating the credit spreads of illiquid counterparties (see Section B.1(f) of the draft Accord text).

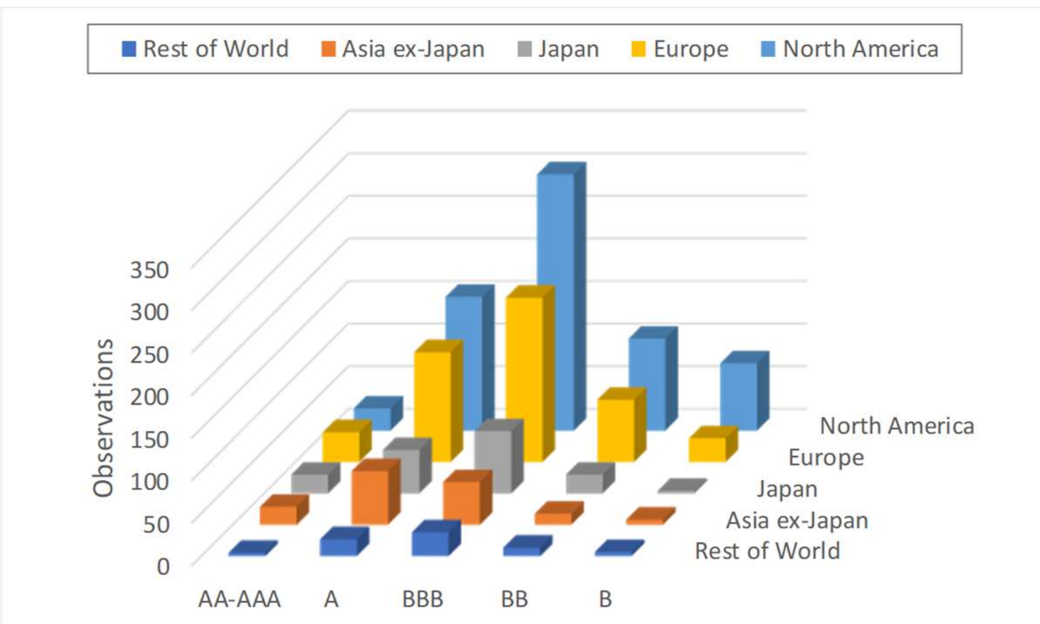
Banks normally develop the capability of calculating CVA sensitivities in order to manage their CVA risk. Typically, CVA risk management is performed by a dedicated function, such as the CVA desk. CVA sensitivities calculated by a bank without any internal function to use them would not be deemed reliable. Thus, the existence of a dedicated CVA risk management function will be a requirement.

Estimation of CDS Proxy Curves

- Two potential approaches
  - Bucketing approach (cannot use very granular buckets due to lack of data)
  - Regression approach

$$Log\ Spread_i = X_{global} + X_{sector(i)} + X_{region(i)} + X_{rating(i)}$$

Financial CDS population



Region
Asia
Eastern Europe
Europe
Japan
Middle East
Latin America
North America
Oceania

Industry
Basic materials
Consumer goods
Consumer services
Energy
Financials
Government
Healthcare
Industrials
Technology
Telecommunications
Utilities

Rating
Aaa
Aa
A
Baa
Ba
B
Caa

CDS Proxy Curves – Example

Rating	Coeff
Aaa	0.0000
Aa	0.4672
A	0.7611
Baa	1.2448
Ba	1.9351
B	2.5882
Caa	3.9599

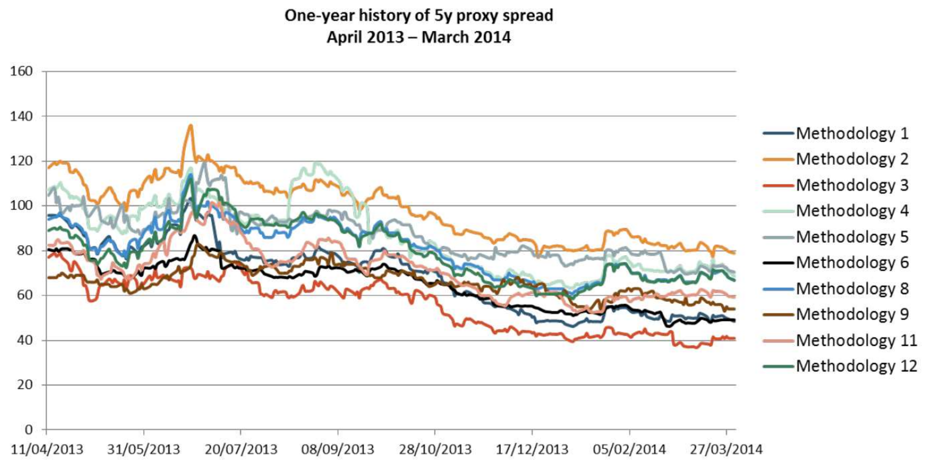
Region	Coeff
Asia	-0.1820
Eastern Europe	-0.0444
Europe	-0.4152
Japan	-0.6069
Middle East	0.1509
Latin America	0.0000
North America	-0.2574
Oceania	-0.2106

Industry	Coeff
Basic materials	-0.0025
Consumer goods	-0.0168
Consumer services	-0.0093
Energy	0.2062
Financials	0.1814
Government	-0.1635
Healthcare	-0.2806
Industrials	-0.0650
Muni	0.3385
Technology	0.0000
Telecommunications	-0.1149
Utilities	-0.1191

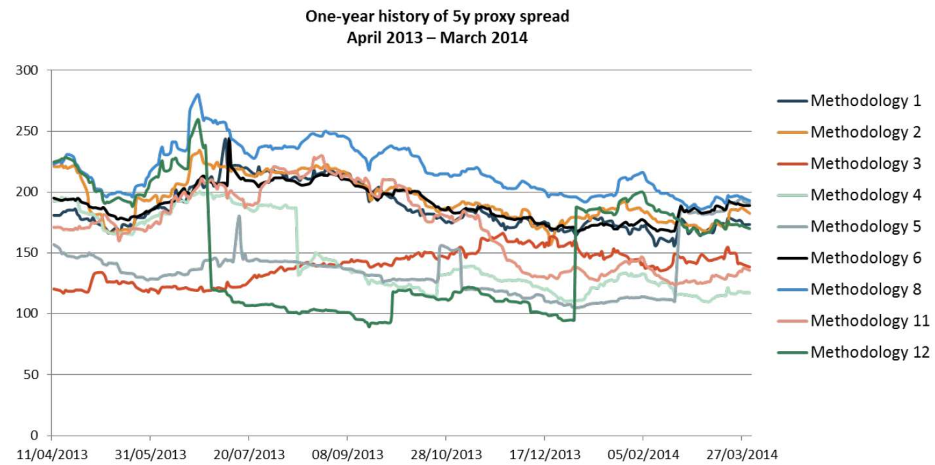
Rating	Region	Industry	Intercept	Rating	Region	Industry	Spread
Aaa	N America	Financial	-5.9364	0.0000	-0.2574	0.1814	0.245%
Aaa	Europe	Financial	-5.9364	0.0000	-0.4152	0.1814	0.209%
A	N America	Healthcare	-5.9364	0.7611	-0.2574	-0.2806	0.330%
A	N America	Muni	-5.9364	0.7611	-0.2574	0.3385	0.613%
A	Middle East	Financial	-5.9364	0.7611	0.1509	0.1814	0.788%
BB	Europe	Energy	-5.9364	1.9351	-0.4152	0.2062	1.484%
CCC	Europe	Healthcare	-5.9364	3.9599	-0.0444	-0.2806	6.910%
CCC	E Europe	Energy	-5.9364	3.9599	-0.0444	0.2062	16.290%

CDS Proxy Curves – Comparison

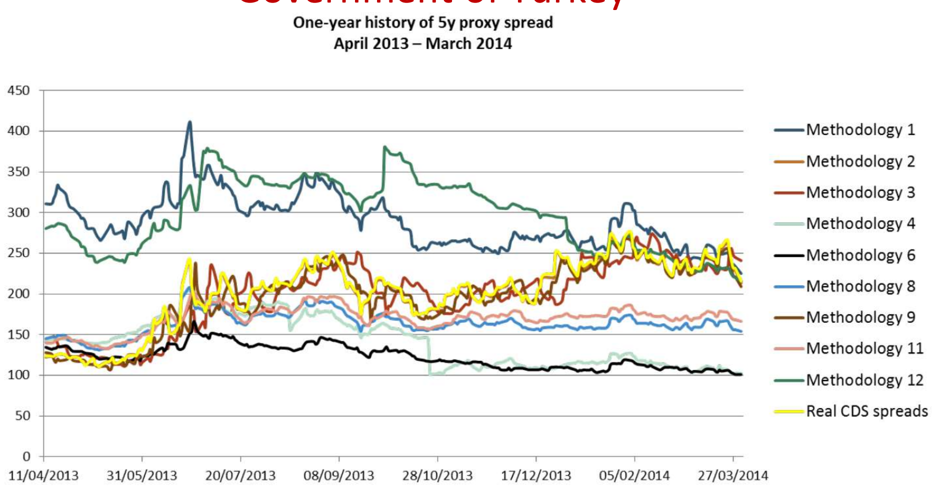
UK AA-rated insurance company



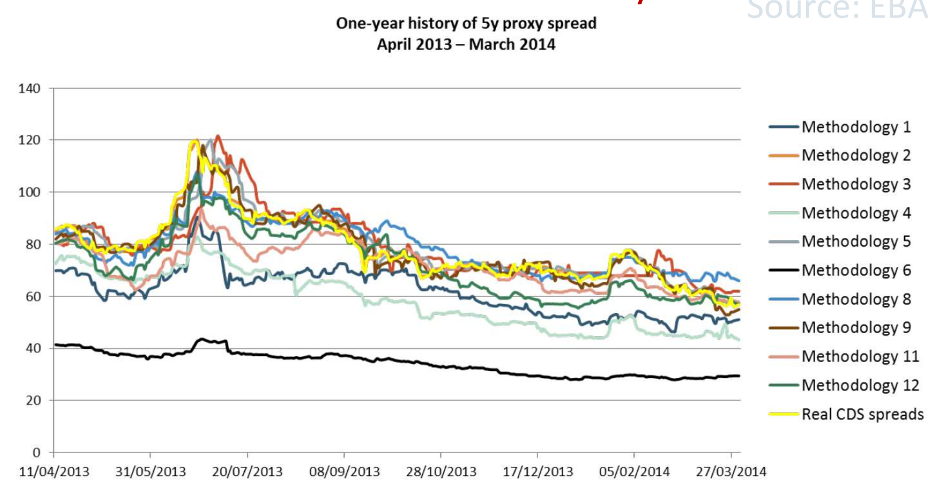
Japanese BB-rated airline



Government of Turkey



Berkshire Hathaway



Source: EBA

Summary

- Single-name and index CDS are fundamental building blocks of credit derivatives
  - Portfolio credit derivatives are more complex
- CDS pricing is straightforward assuming independence between default, recovery and interest rates
  - Counterparty risk is a more difficult and leads to the use of structural models
- Most credits do not have liquid observable CDS quotes
  - Proxy methodologies are important

CQF: Certificate in Quantitative Finance

Credit Derivatives and Structural Models

18th June 2020

Dr Jon Gregory
jon@solum-financial.com

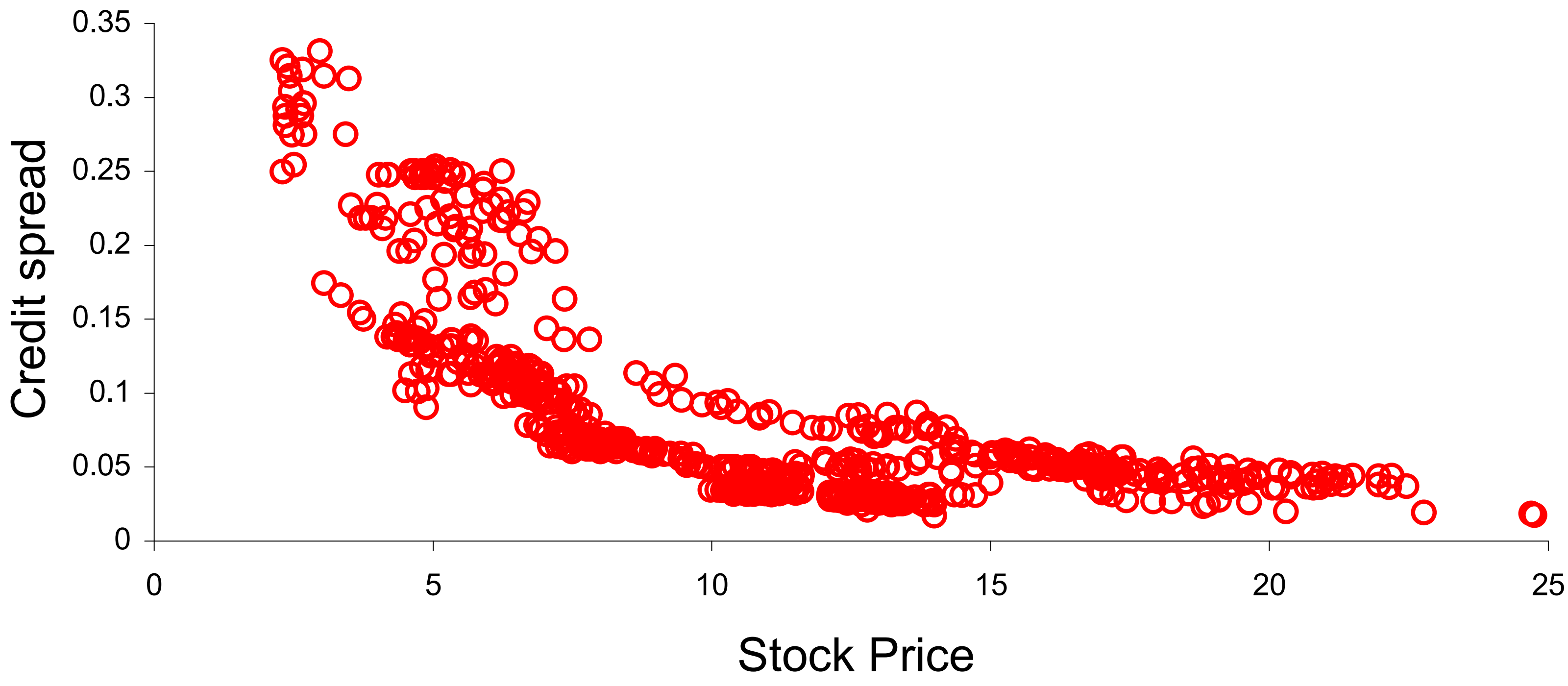
Content

- The Merton Model
- Structural Models and Default Prediction
- Portfolio Credit Derivatives
- Structural Models, Basket Credit Derivatives and CDOs
- CDOs and the Financial Crisis
- Does Securitisation have any Economic Value?

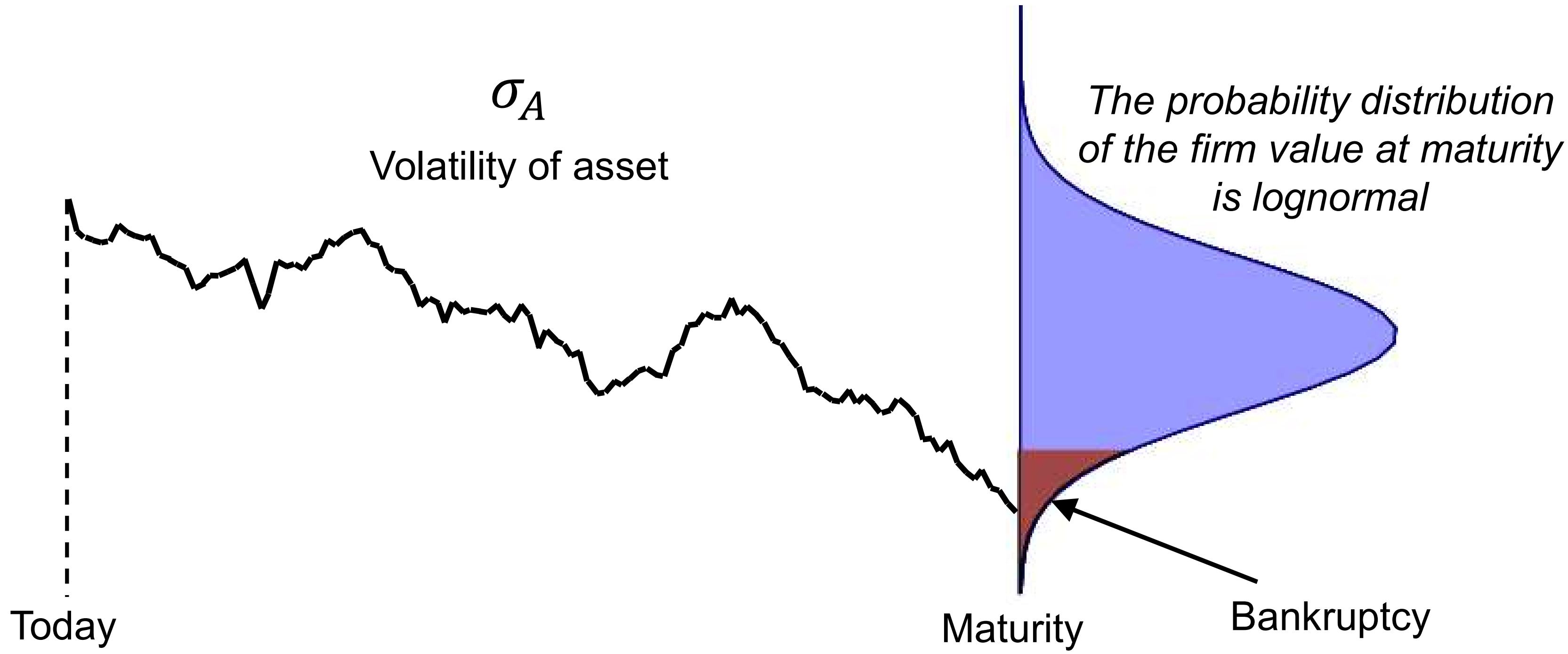
The Merton Model

FitchLearning

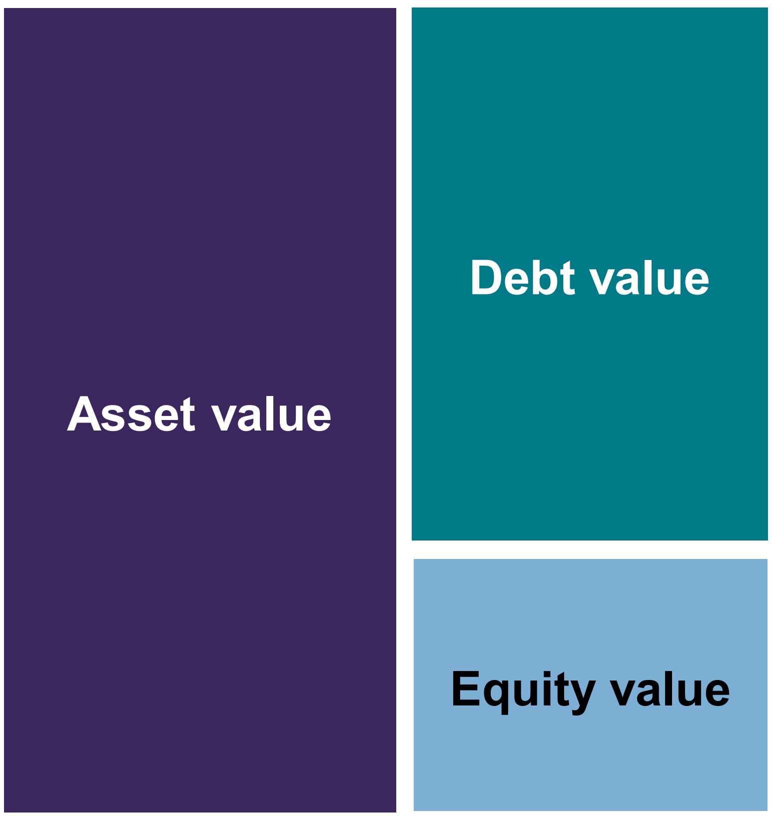
Structural Models – Empirical Data



Merton Model Assumptions

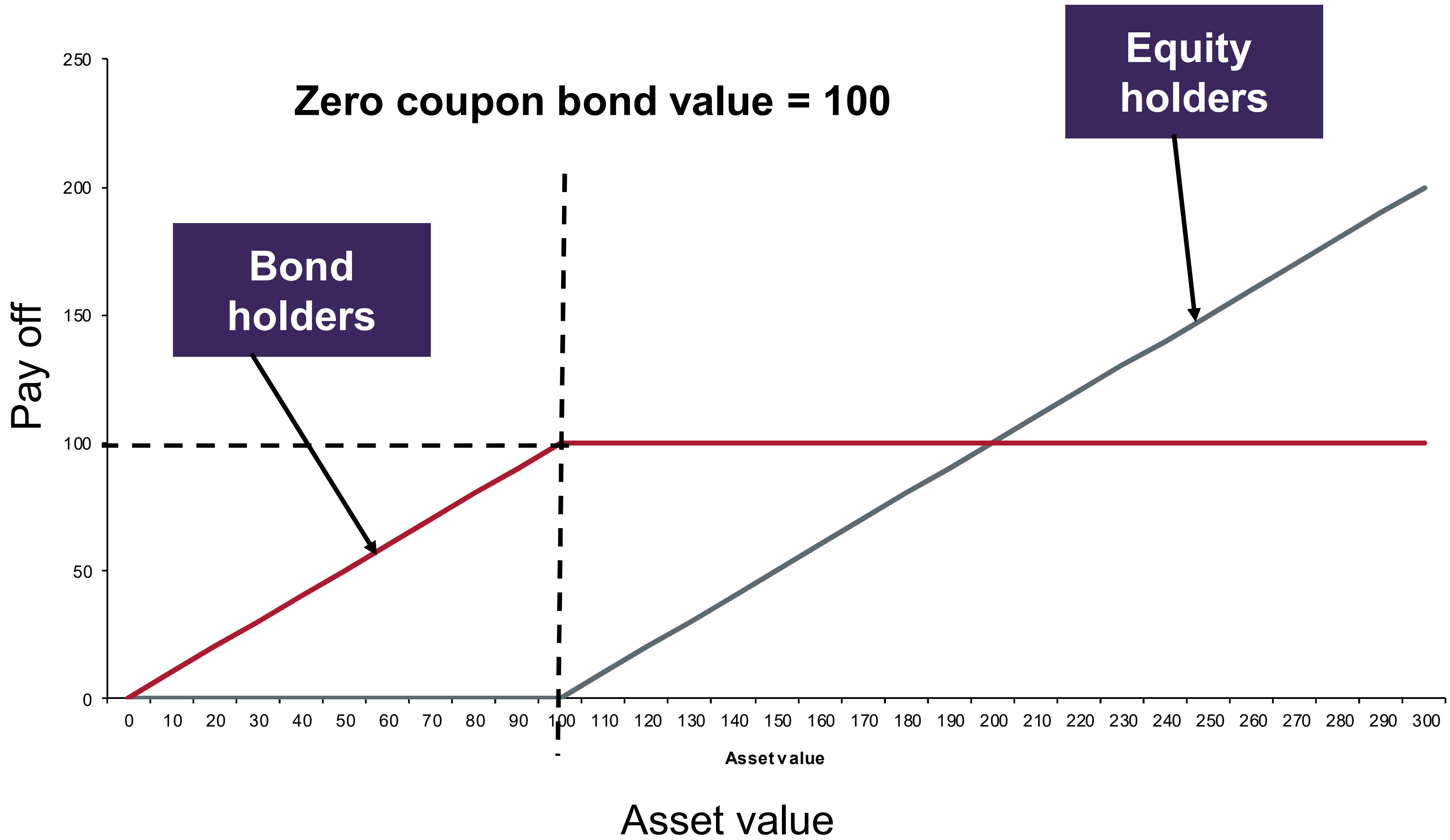


Deriving PDs from Market Prices: Merton Model



- Firm defaults when its asset value falls below its debt obligation
- Debts are contingent claims on firms assets
  - Pay-off is lower of firms asset value or face value of debt
  - Value of debt is ZCB of face value plus short put option on firm’s assets
- Survival probability is the probability of the asset value staying above the default point of the firm
- Default probability = 1- survival probability

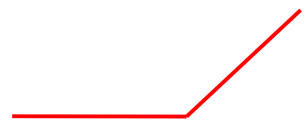
Default in the Merton Model: Pay Off at Maturity



The Merton Model in Summary

- Definitions
  - Firm value is V
  - Equity value is E
  - Face value of debt is F
- Value of equity is option on value of firm
  - Equity is residual value after bondholders have been paid

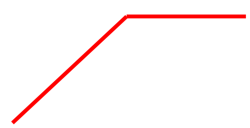
$$E = \max(V - F, 0)$$



$$E = V\Phi(d_1) - e^{-rT}F\Phi(d_2)$$

- Value of debt is option on value of firm
  - Bondholders have first claim on assets of company

$$D = \min(F, V)$$



$$D = V\Phi(-d_1) + e^{-rT}F\Phi(d_2)$$

$$PD = \Phi(-d_2)$$

$$d_1 = \frac{\ln(V/F) + (r + \sigma_A^2/2)T}{\sigma_A\sqrt{T}} = d_2 + \sigma_A\sqrt{T}$$

Merton Model Example

$$F = 100$$

$$r = 5\%$$

$$T = 1 \text{ year}$$

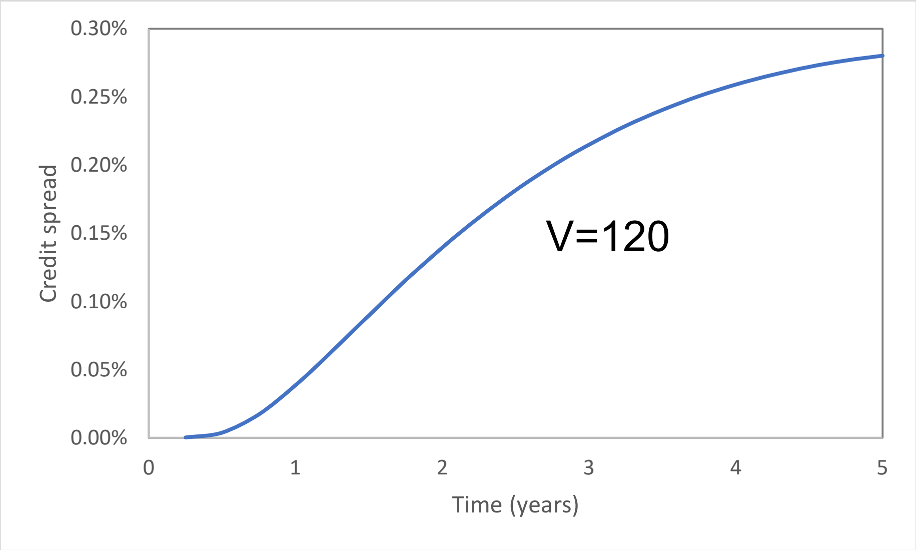
$$\sigma = 20\%$$

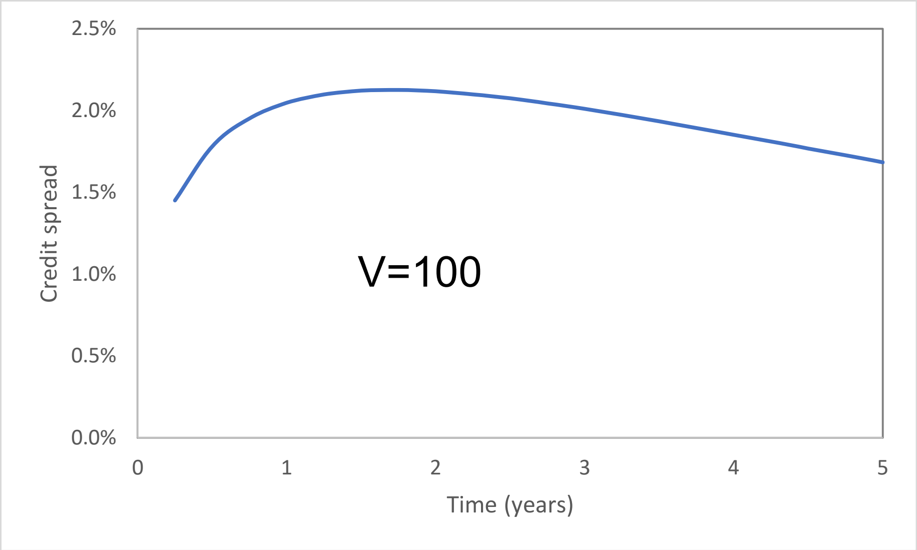
$$V = 80 \quad \Rightarrow D = 78.14, \quad E = 1.86 \quad PD = 83.2\%$$

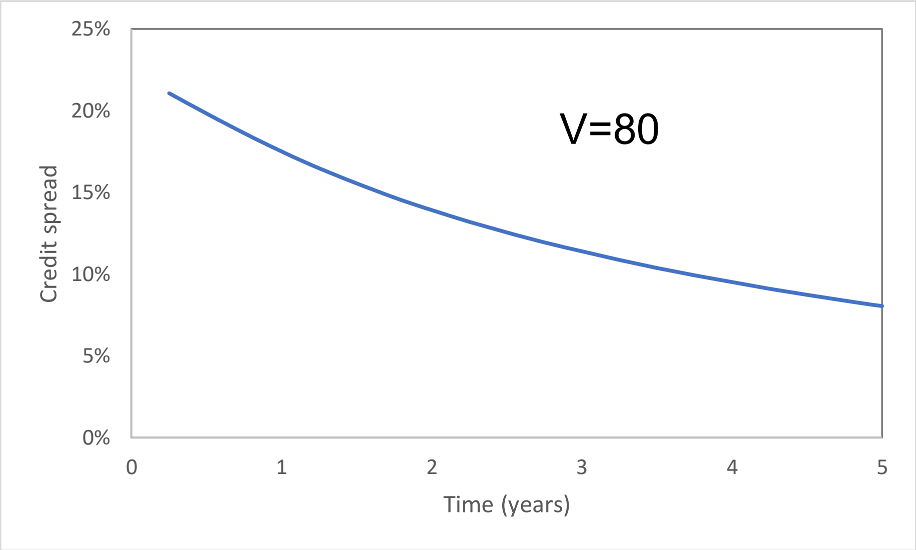
$$V = 100 \quad \Rightarrow D = 89.55, \quad E = 10.45 \quad PD = 44.0\%$$

$$V = 120 \quad \Rightarrow D = 93.83, \quad E = 26.17 \quad PD = 14.4\%$$

Merton Model Example – Credit Spread Curves







Extensions to the Merton Model

- Black and Cox (1976)
  - Default occurs at the first time that the firm's asset value drops below a certain time dependent barrier
- Leland and Toft (1994, 1996)
  - Bankruptcy costs and tax on coupons to consider optimal capital structure
- KMV (1990s)
  - Default prediction with Merton model
  - Acquired by Moody's in 2002
- Capital structure arbitrage (1990s onwards)
  - Proprietary trading desks trading different securities of a company
  - Bonds, CDS, equity and equity puts

Structural Models and Default Prediction

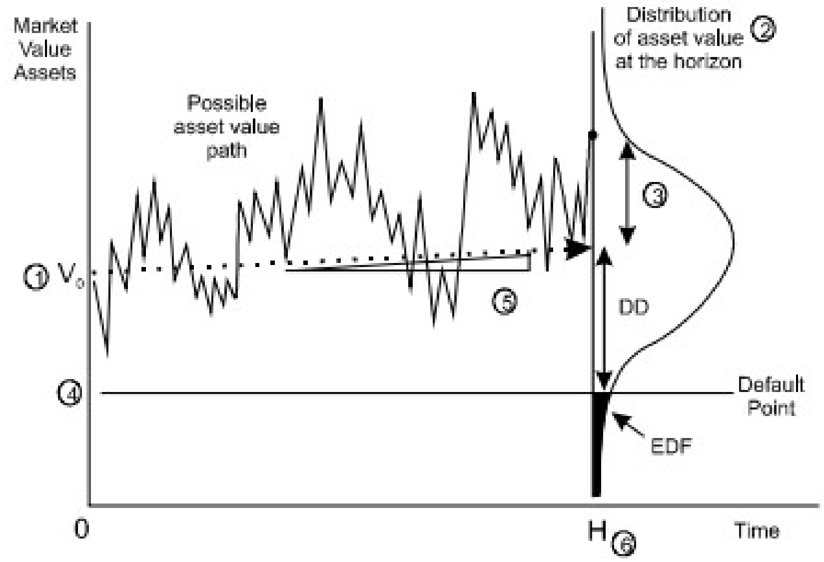
FitchLearning

Structural Models in Practice: Moody's KMV

Moody's | K·M·V

MOODY'S KMV RISKCALC™ v3.1 MODEL

NEXT-GENERATION TECHNOLOGY FOR PREDICTING PRIVATE FIRM CREDIT RISK

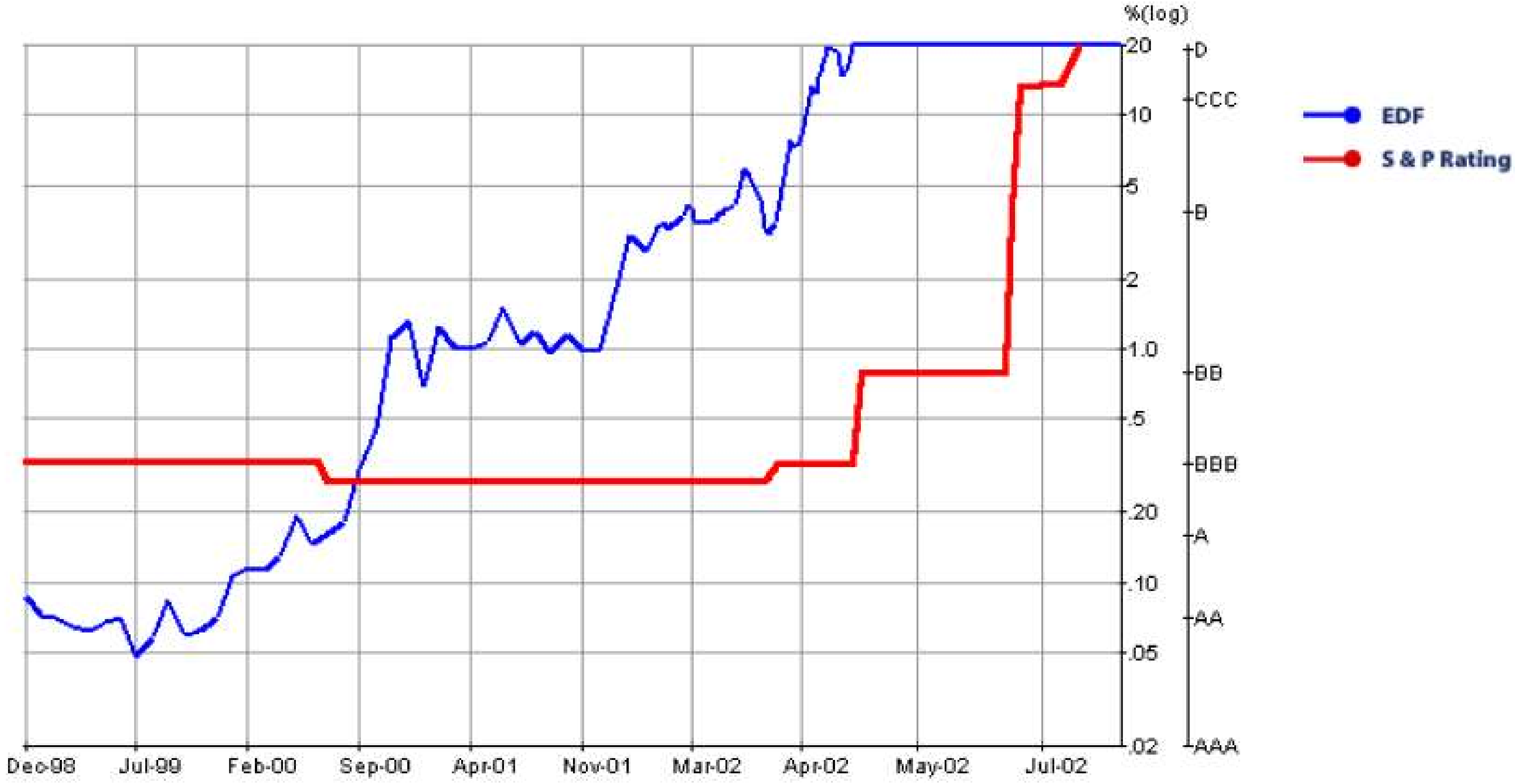


3.2 Calculate the Distance-to-default

There are six variables that determine the default probability of a firm over some horizon, from now until time *H* (see Figure 8):

1. The current asset value.
2. The distribution of the asset value at time H.
3. The volatility of the future assets value at time H.
4. The level of the default point, the book value of the liabilities.
5. The expected rate of growth in the asset value over the horizon.
6. The length of the horizon, H.

EDF and Credit Ratings



Source: Moody's KMV Case Study: Worldcom

Issues With Applying the Merton Model

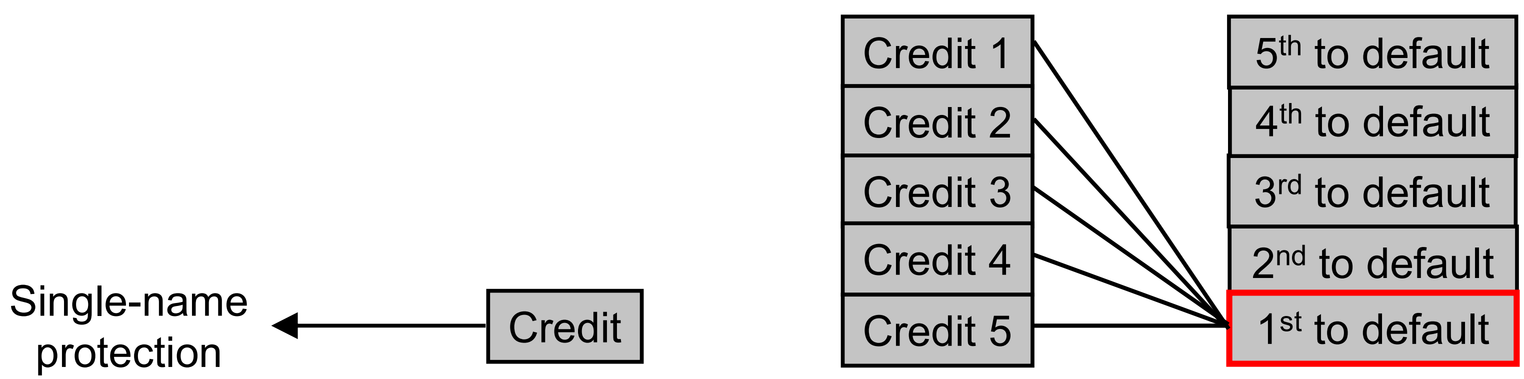
- Relies on a static model of the firm's capital structure
- Higher stock price does not necessarily reflect a lower PD
  - E.g. if the firm's projects are potentially valuable but high risk
- Cannot explain the magnitude of credit spreads
- Difficult to use with firms with more complex liability structure
  - Generally not applicable to financial institutions
- Not applicable to sovereign exposures
- Much of KMV's model is proprietary
  - e.g. Calculation the volatility of the default put

Portfolio Credit Derivatives

FitchLearning

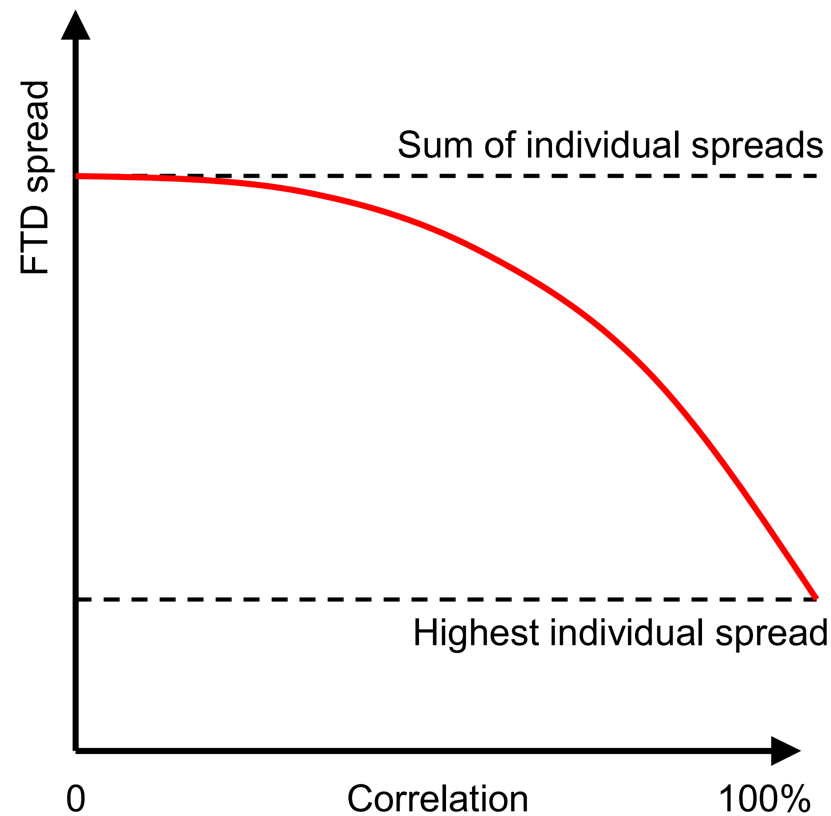
Basket Credit Default Swaps

- Credit default swap linked to a variety of underlying reference assets ('basket')
- The protection seller provides protection against a specified reference entity suffering a credit event, usually the 'first to default' (FTD)
- The protection seller is therefore assuming the risk of any default in the basket
- Can generalise to an 'nth to default' payoff
- Value depends crucially on the dependency between the different credits



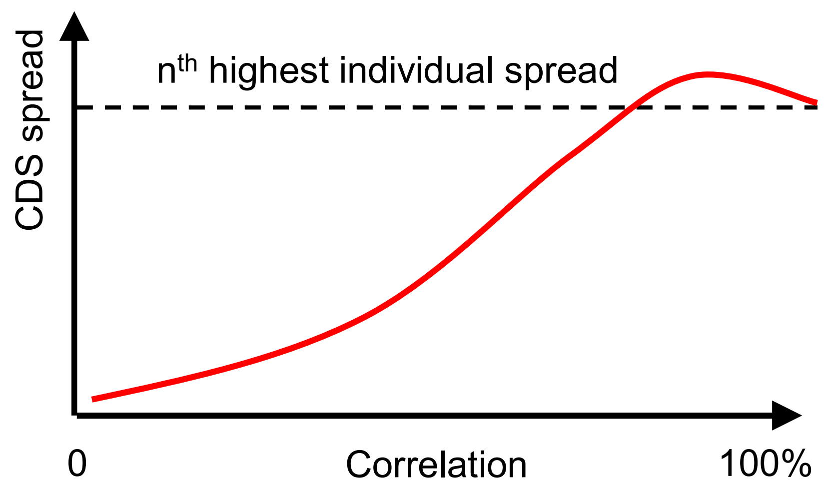
Behaviour – First to Default

- As the dependency becomes maximum (e.g. correlation 100%) the basket premium tends towards the largest single name default premium
  - Highest premium represents the greatest probability of default
  - In practice 100% correlation is a strange concept if the credits have unequal default probabilities

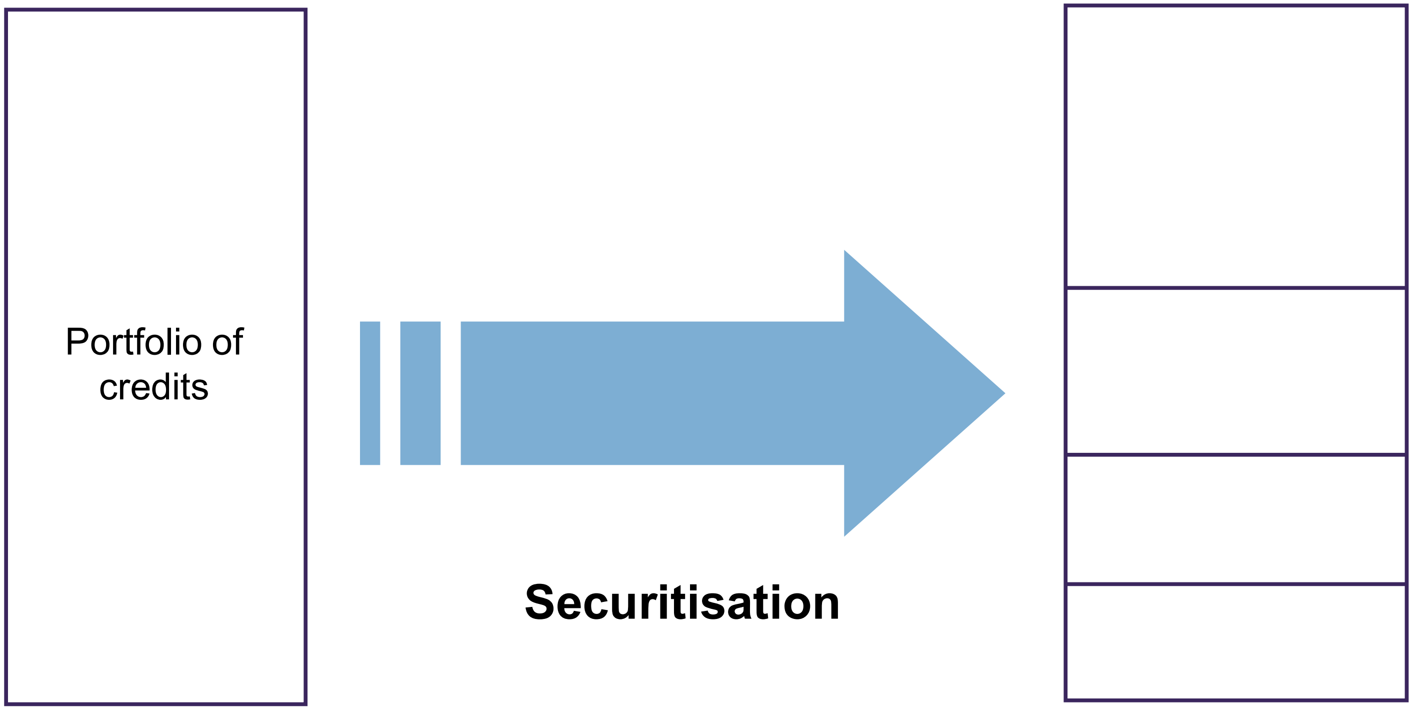


Behaviour – Nth to Default

- For low dependency, the probability of greater than average number of the underlyings defaulting becomes rather small
  - Most nth baskets (n>1) correspond to this situation
- Basket premium will therefore typically increase with correlation
- In the limit, the premium will tend towards the nth highest spread
  - Behaviour may not be monotonic



Securitisation

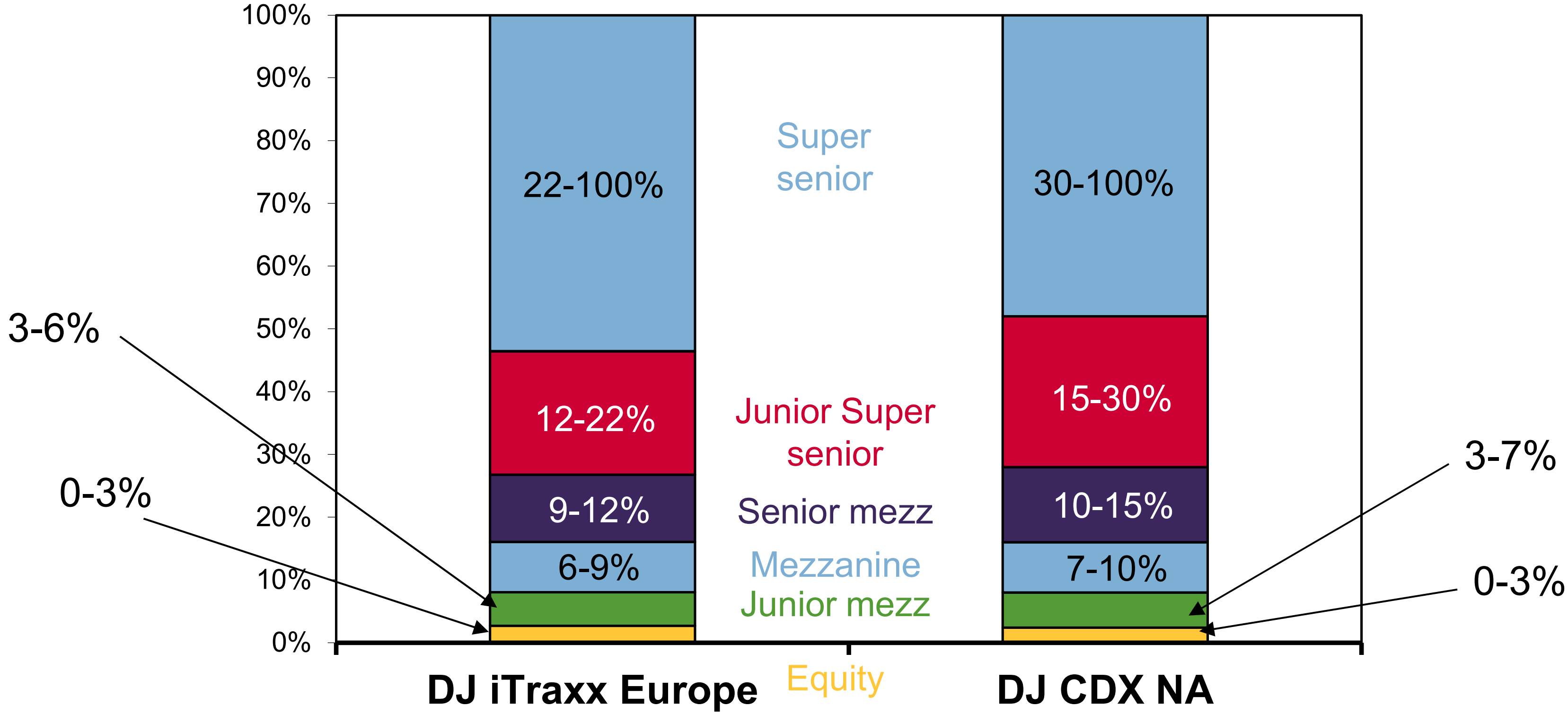


	Size	Tranching	Rating
Super senior	850	[15-100%]	NR
Class A	50	[10-15%]	Aaa/AAA
Class B	30	[7-10%]	Aa2/AA
Class C	30	[4-7%]	Baa2/BBB
Equity	40	[0-4%]	NR

Credit Ratings: Agency Definitions

Description	Fitch and S&P		Moody's		Short term	Explanation
Highest credit quality	AAA		Aaa		F1	Exceptionally strong capacity for timely payment of financial commitments which is highly unlikely to be adversely affected by foreseeable events.
Very high credit quality	AA	AA+ AA AA-	Aa	Aa1 Aa2 Aa3		Very strong capacity for timely payment of financial commitments which is not significantly vulnerable to foreseeable events.
High credit quality	A	A+ A A-	A	A 1 A 2 A 3	F2	Strong capacity for timely payment of financial commitments which may be more vulnerable to changes in circumstances/ economic conditions.
Good credit quality	BBB	BBB+ BBB BBB-	Baa	Baa1 Baa2 Baa3	F3	Adequate capacity for timely payment of financial commitments but adverse changes in circumstances/ economic conditions are more likely to impair this capacity.
Speculative	BB	BB+ BB BB-	Ba	Ba1 Ba2 Ba3		Possibility of credit risk developing, particularly due to adverse economic change over time. Business/financial alternatives may be available to allow financial commitments to be met.
Highly speculative	B	B+ B B-	B	B1 B2 B3	B	Significant credit risk with a limited margin of safety. Financial commitments currently being met; however, continued payment is contingent upon a sustained, favourable business and economic environment.
High default risk	CCC		Caa		C	Default is a real possibility. Capacity for meeting financial commitments is solely reliant upon sustained, favourable business or economic developments.
Probable default	CC		Ca			Default of some kind appears probable.
Likely default	C		C			Default imminent.

Index Tranche CDS: Attachment and Detachment Points



Structural Models, Basket Credit Derivatives and CDOs

FitchLearning

Pricing an Index Tranche or CDO

- To price tranche protection in the range $[X, Y]$, need to know the expected tranche loss over time (Gregory and Laurent [2003])

Maturity of tranche

$$E\left[\int_t^T B(t,s)\,dM(t,s)\right]$$

Risk-free discount factor

Tranche loss process

- This, in turn, is calculated from the loss distribution itself since the tranche loss is a simple call spread payoff

$$M(t,s) = (L(t,s), Y)_+ - \max(L(t,s), X)_+$$

- Need a model for the portfolio loss distribution $L(t,s)$
- A structural model is the obvious choice (hazard rate models do not produce strong enough dependency)

Merton Model for Portfolio Defaults

- Recall that default of a single name is determined by variable V_1

$$V_1 < \Phi^{-1}(p_1)$$

- Correlation structure (e.g. Gaussian) between variables $V_1, V_2, \dots \dots, V_n$

- A factor model is useful: $V_i = \sqrt{\rho} Z + \sqrt{1-\rho}\, Y_i$

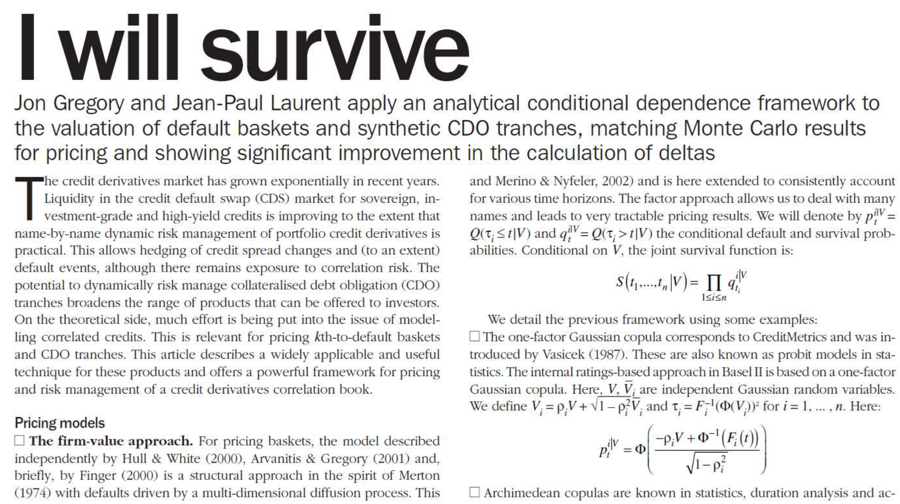
Idiosyncratic variable

Correlation

Systematic variable

- Conditional independence:

$$p_i = \Pr\left(V_i < \Phi^{-1}(p_i)\right)$$
$$= \Phi\left(\frac{\Phi^{-1}(p_i) - \sqrt{\rho} Z}{\sqrt{1-\rho}}\right)$$



I will survive

Jon Gregory and Jean-Paul Laurent apply an analytical conditional dependence framework to the valuation of default baskets and synthetic CDO tranches, matching Monte Carlo results for pricing and showing significant improvement in the calculation of deltas

The credit derivatives market has grown exponentially in recent years. Liquidity in the credit default swap (CDS) market for sovereign, investment-grade and high-yield credits is improving to the extent that name-by-name dynamic risk management of portfolio credit derivatives is practical. This allows hedging of credit spread changes and (to an extent) default events, although there remains exposure to correlation risk. The potential to dynamically risk manage collateralised debt obligation (CDO) tranches broadens the range of products that can be offered to investors. On the theoretical side, much effort is being put into the issue of modelling correlated credits. This is relevant for pricing *k*th-to-default baskets and CDO tranches. This article describes a widely applicable and useful technique for these products and offers a powerful framework for pricing and risk management of a credit derivatives correlation book.

Pricing models

☐ **The firm-value approach.** For pricing baskets, the model described independently by Hull & White (2000), Arvanitis & Gregory (2001) and, briefly, by Finger (2000) is a structural approach in the spirit of Merton (1974) with defaults driven by a multi-dimensional diffusion process. This

and Merino & Nyfeler, 2002) and is here extended to consistently account for various time horizons. The factor approach allows us to deal with many names and leads to very tractable pricing results. We will denote by $p_t^{i|V} = Q(\tau_i \le t|V)$ and $q_t^{i|V} = Q(\tau_i > t|V)$ the conditional default and survival probabilities. Conditional on V, the joint survival function is:

$$S(t_1,\dots,t_n|V) = \prod_{1\le i\le n} q_{t_i}^{i|V}$$

We detail the previous framework using some examples:

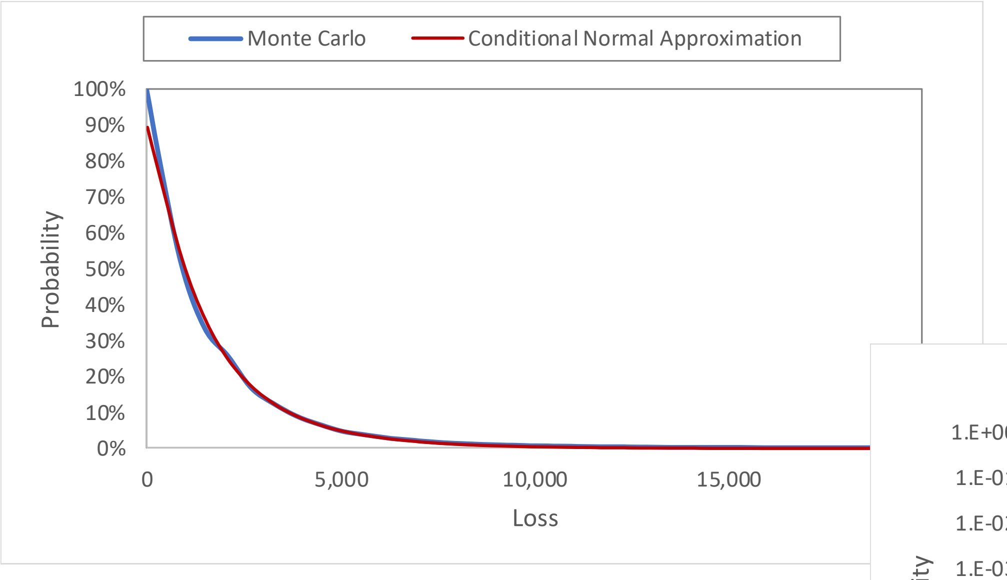
☐ The one-factor Gaussian copula corresponds to CreditMetrics and was introduced by Vasicek (1987). These are also known as probit models in statistics. The internal ratings-based approach in Basel II is based on a one-factor Gaussian copula. Here, $V, \bar{V}_i$ are independent Gaussian random variables. We define $V_i = \rho_i V + \sqrt{1-\rho_i^2}\bar{V}_i$ and $\tau_i = F_i^{-1}(\Phi(V_i))^2$ for $i = 1, \dots, n$. Here:

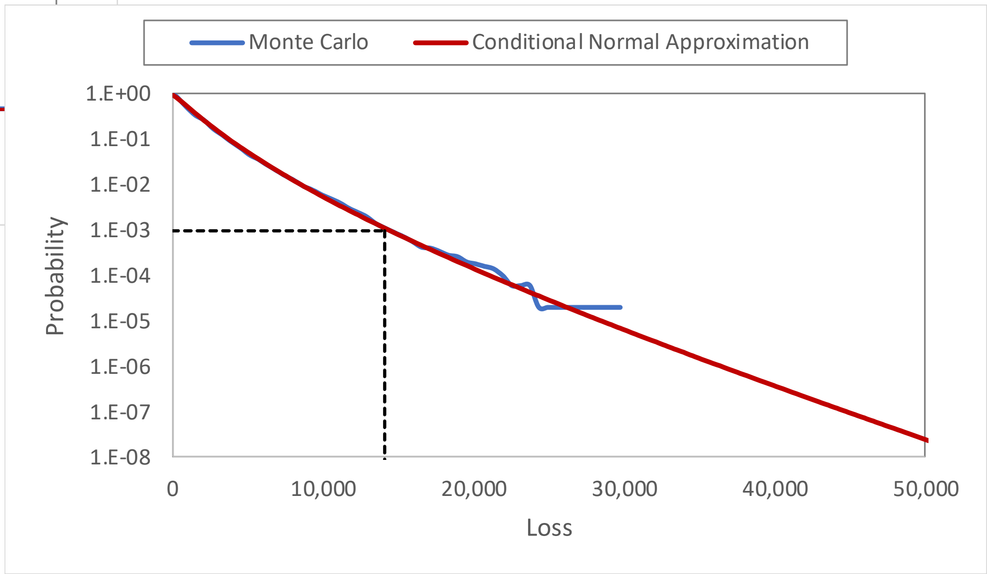
$$p_t^{i|V} = \Phi\left(\frac{-\rho_i V + \Phi^{-1}(F_i(t))}{\sqrt{1-\rho_i^2}}\right)$$

☐ Archimedean copulas are known in statistics, duration analysis and ac-

Analytical Factor Models

- With conditional independence can calculate loss distribution accurately without Monte Carlo simulation





Model Risk : Choice of Copula

- How sensitive is our choice of a Gaussian copula?

	Gaussian	Student-t	Clayton	Marshall-Olkin
1st to default	723	723	723	723
2nd to default	277	278	274	160
3rd to default	122	122	123	53
4th to default	55	55	56	37
5th to default	24	24	25	36
6^{tht} to default	11	10	11	36
7th to default	3.6	3.5	4.3	36
8th to default	1.2	1.1	1.5	36
9th to default	0.28	2.5	0.39	36
10th to default	0.04	0.04	0.06	36

10 names, spreads from 60 bps to 150 bps, recovery = 40%, maturity = 5 years, Gaussian correlation = 30%

The Failure of the Gaussian Copula Model

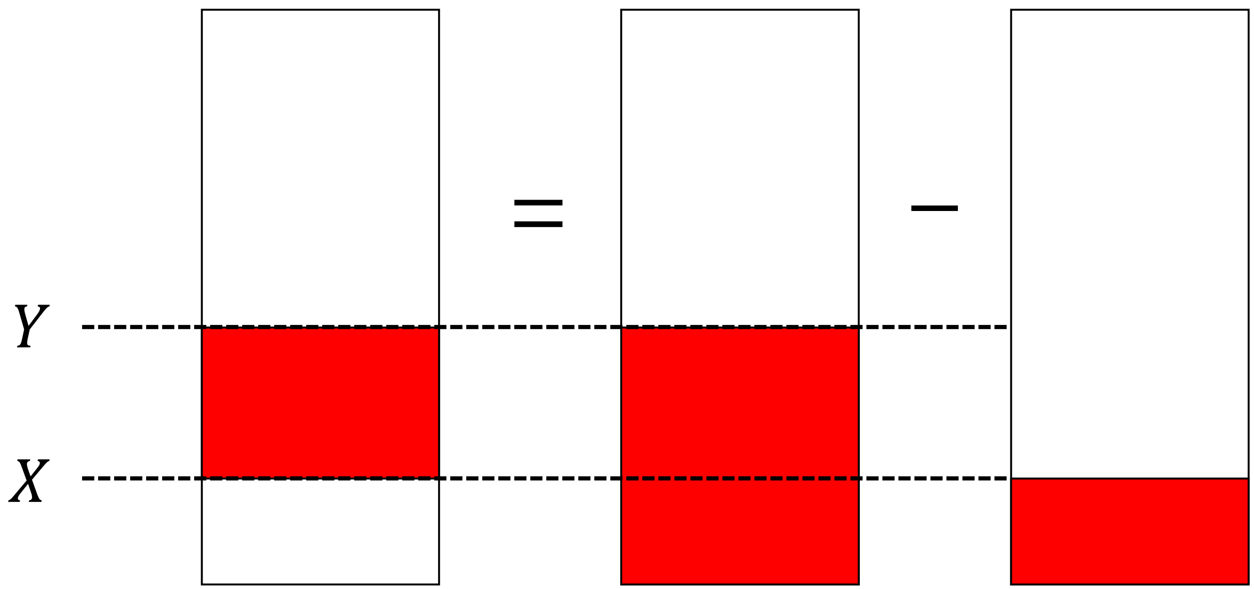
Tranche	Market (bp pa)	Gaussian Copula Model
Super Senior 22-100%	14.9	0.38
12-22%	34.3	24.3
9-12%	60.5	87.1
6-9%	84.9	180.0
3-6%	156.3	417.7
Equity 0-3%	25.05%	13.9%

Upfront premium

Base Correlation

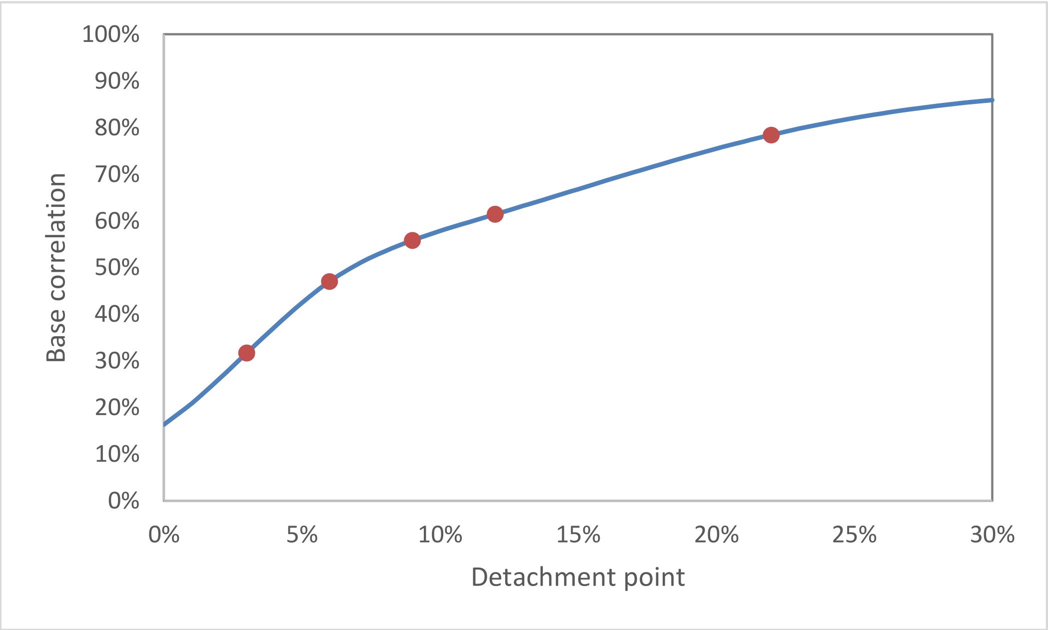
- A CDO tranche is a call spread payoff
- Split into two option type 'base tranches'

$$CDO[X,Y] \equiv CDO[0,Y] - CDO[0,X]$$



Base Correlation Example

	Market	Correlation	Model
[0-3%]	25.05%	31.7%	
[3-6%]	156.3	47.0%	
[6-9%]	84.9	55.8%	
[9-12%]	60.5	61.4%	
[12-22%]	34.3	78.4%	
[22-100%]			14.9
[8-15%]			57.3



CDOs and the Financial Crisis

The answer to how we got in this mess... if you can do the 'math'

WITH impeccable timing comes Standard & Poor's Handbook of Structured Finance, 785 pages with an incomprehensible equation (like this one below) on nearly all of them.

This tome, it says on the back, "provides a comprehensive overview of quantitative techniques needed to measure and manage risks..."

It also promises to "employ risk-measurement techniques such as ratings, and 'The Greeks' in structured deals". Perhaps they meant to say "Geeks", since neither word is in the index.

$$C_j(0) = D(T)\,EL_j(T) + \int_0^T EL_j(t)\,dD(t)$$
$$= D(T)\,EL_j(T) + \int_0^T EL_j(t)\,D(t)\,f(t)\,dt$$

S&P, of course, is the rating agency that OK'd all that sub-prime mortgage junk which is poisoning the world's financial system, and if you wondered how all those clever people could screw things up so royally, this book tells you.

As these instruments got more and more complicated, so ever-more sophisticated models were needed to justify their splendid credit ratings and thus the high prices their creators had demanded from the mugs who were buying.

So arcane was the "math" (it's mostly American, after all) that there was no room for anything like common sense. How the sow's ear of a subprime mortgage can somehow be transmuted into the silk purse of a top-notch credit is swamped under the bell curves, analytical models and (of course) management fees.

Arnaud de Servigny, one of the book's authors, was head of S&P's quantitative analytics, whatever they are. He's now doing something similar at Barclays Wealth, a business that sounds somewhat oxymoronic today.

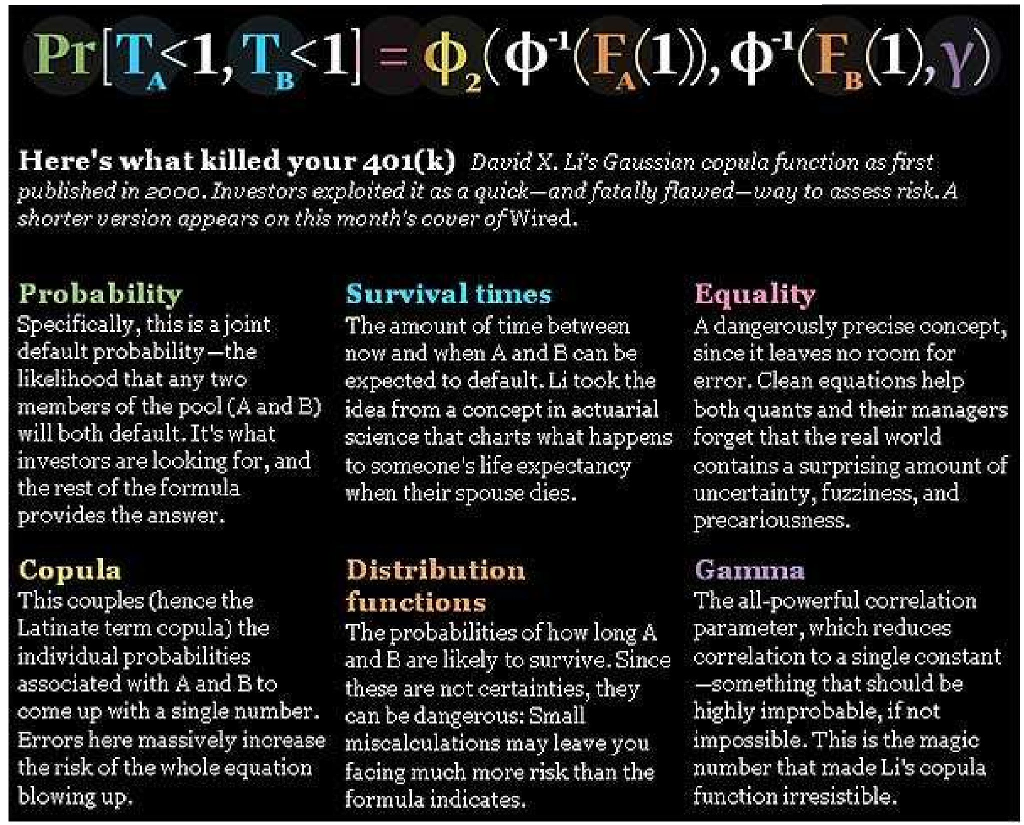
I wish I could recommend his handbook but you'd need a PhD in higher maths to understand it, and besides, after the summer meltdown, you can't afford the 54 quid S&P want for it.

Source : Evening Standard August 2007

On Default Correlation: A Copula Function Approach

David X. Li

This draft: April 2000
First draft: September 1999



$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1), \gamma)$$

Here's what killed your 401(k) *David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of* Wired.

Probability
Specifically, this is a joint default probability—the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Survival times
The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Equality
A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

Copula
This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Distribution functions
The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

Gamma
The all-powerful correlation parameter, which reduces correlation to a single constant—something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

Last month, Fitch paid $27m to settle the *Mid-Coast Council v Fitch Ratings* case. The settlement resulted in the recovery of about 95% of the losses incurred by claimants in their purchase of synthetic collateralised debt obligations (SCDOs), which were rated by Fitch.

BUSINESS NEWS AUGUST 10, 2018 / 4:15 AM / 2 YEARS AGO

S&P settles landmark derivatives-rating lawsuit in Australia

3 MIN READ

SYDNEY (Reuters) - Standard & Poor's said on Friday it has settled a lawsuit in Australia over claims by pension funds and local governments that the credit rating firm had overlooked risks when awarding high ratings to opaque investments that imploded in the global financial crisis.

Australia's Federal Court approved the settlement on Thursday, S&P said in an emailed statement which did not disclose the settlement sum or terms.

"S&P Global is pleased to reach a settlement on the class action lawsuit, the last of the significant litigation pertaining to our previous ratings actions on collateralised debt obligations," it said.

The U.S.-based ratings agency was sued for at least A$190 million ($140 million) by two local governments and two pension funds in Australia, which lost money on

Does Securitisation have any Economic Value?

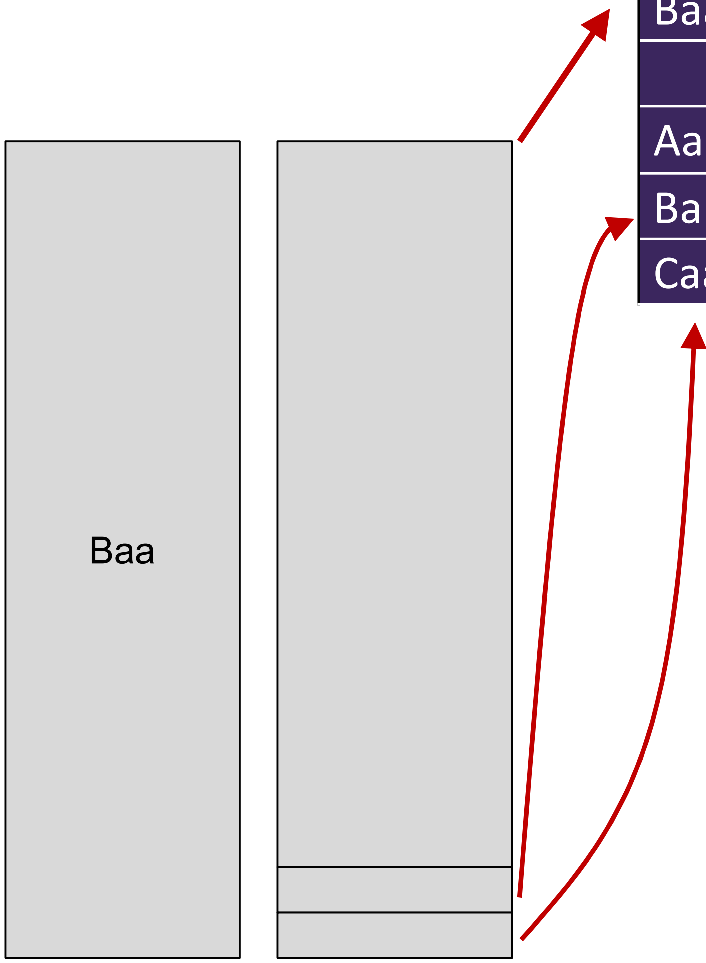
Do CDOs Work?

FitchLearning

Overview

- In the period 1998 to 2007, CDOs increased exponentially in both volume and diversity
  - Prior to 2007, the CDO was seen as a successful financial innovation
- However, the global financial crisis was partly catalysed by an implosion in the CDO market and caused massive losses for:
  - Issuers (banks) through investments held, litigation, failed hedges, reputation
  - Investors, both in terms of default losses and those from forced liquidation
  - Third parties (e.g. rating agencies through loss of fees, reputation issues and litigation)
- An obvious question is therefore:
  - Is there something fundamentally wrong with the concept of a CDO – and more broadly – the concept of securitisation?
  - Does it have economic value?

Simple CDO Economics



	Tranching	Rating	Coupon
Baa	[0-100%]	Baa	250
Aaa	[8-100%]	Aaa	100
Ba	[4-8%]	Ba	600
Caa	[0-4%]	Caa	1500

$$
\begin{aligned}
&\textit{Excess Spread} \\
&= (250 \\
&-100 \times 92\% \\
&-600 \times 4\% \\
&-1500 \times 4\%) \\
&\underline{= 74\ bps}
\end{aligned}
$$

Due to the short duration of the equity tranche, the excess spread is a conservative measure of the profitability of a CDO

Economics of a CDO

- Suppose there is a continuum of underlying tranches (full capital structure)
- Consider expected loss (EL) as the main quantitative characteristic of the tranche
  - Expected loss must be conserved across the structure

$$EL_P = \sum_i m_i EL_i \qquad \sum_i m_i = 1$$

Expected loss for unit tranche (under physical measure)

Tranche size

- Investors will demand a premium for the losses they take
  - Represented via a multiplier α which represents the risk aversion for a given seniority and will be implicitly determined by the coupon demanded by investors
  - The CDO will "work" if

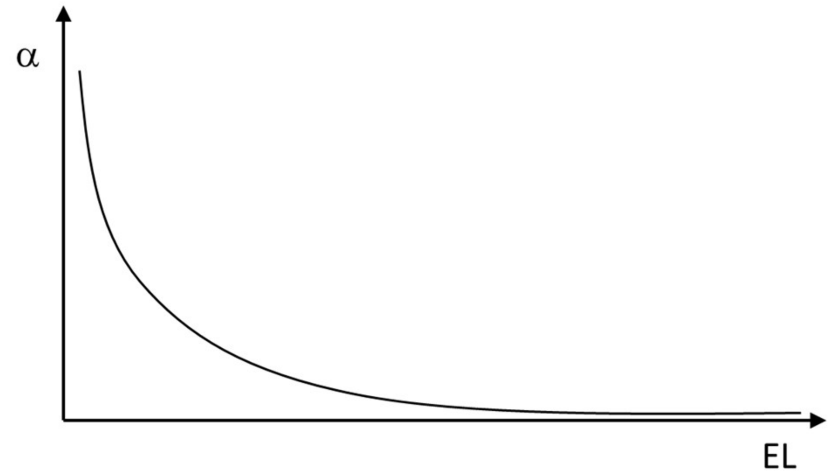
$$\alpha_p EL_P > \sum_i \alpha_i m_i EL_i$$

  - This basically requires that it is possible to buy protection cheaper via the CDO tranches than it is on the underlying portfolio

Risk Aversion and Seniority

- How do we represent α?
  - The primary consideration of investors is the rating of the underlying tranche
  - In turn, the fundamental driver of ratings would be the expected loss of a tranche
  - Hence we assume

$$\alpha_j = \left(\frac{a}{EL_j}\right)^b$$



- Properties
  - Risk-neutral investors, $b = 0$
  - Risk aversion for $a, b > 0$
  - More relative risk aversion for small expected losses

Requirement for CDO to Work

- What parameters are required for a CDO to work?
  - We require:

$$\alpha_p EL_P > \sum_i \alpha_i m_i EL_i \qquad\qquad \alpha_j = \left(\frac{a}{EL_j}\right)^b$$

  - Which becomes:

$$\left(\frac{a}{EL_p}\right)^b EL_P > \sum_i \left(\frac{a}{EL_i}\right)^b m_i EL_i$$

  - Simplifying to:

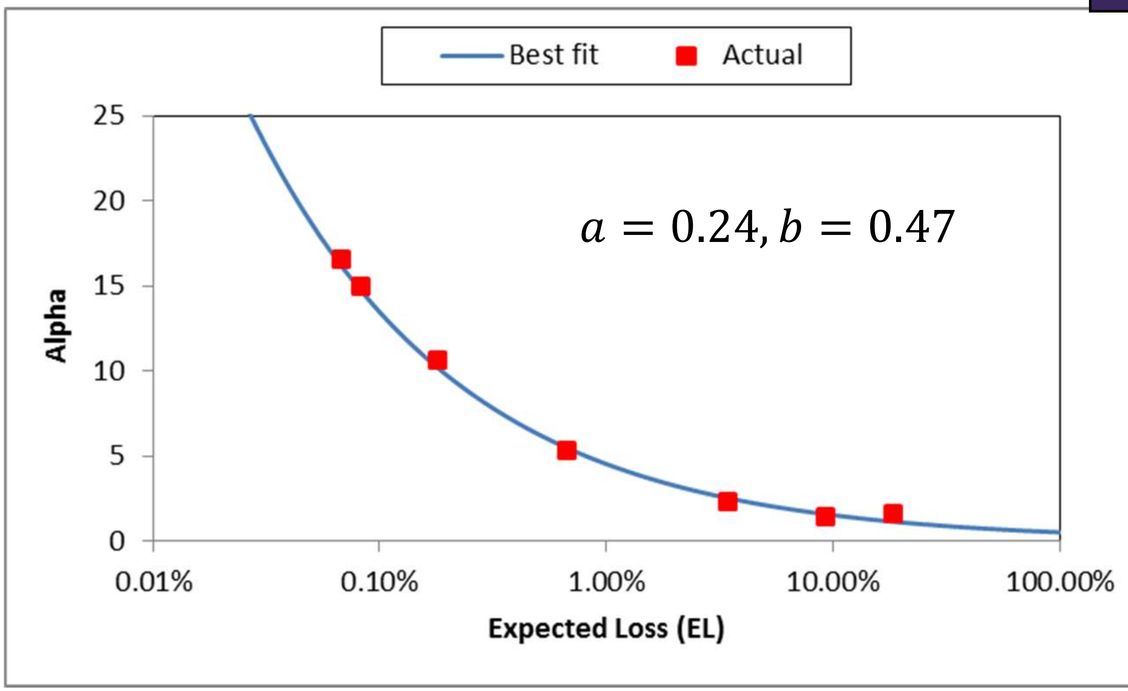
$$EL_P^{1-b} > \sum_i m_i EL_i^{1-b}$$

  - Which is satisfied when $b < 1$

Example Calibration

- Hull, Predescu and White (2005)
  - Time period, December 1996 to July 2004
  - Merrill Lynch bond indices and Moody's data

	Real world loss (bps)	Risk neutral loss (bps)	Ratio
Aaa	4	67	16.8
Aa	6	78	13.0
A	13	128	9.8
Baa	47	238	5.1
Ba	240	507	2.1
B	749	902	1.2
Caa	1690	2130	1.3



Conclusions

- A CDO works due to
  - The risk preferences of investors
  - The expected loss methodology used in the ratings process
- What did go wrong then?
- Lack of proper assessment of counterparty risk in the structuring process
  - The more senior the tranche, the more counterparty risk (relatively) and monoline insurers (Gregory 2008)
- Lack of appreciation of the systemic risk in senior tranches
  - Were investors sufficiently compensated for this?
  - Gibson, M., 2004, "Understanding the risk of synthetic CDOs", Finance and Economics Discussion Paper, 2004–36, Federal Reserve Board, Washington DC / Coval et al, 2009, "Economic catastrophe bonds," American Economic Review, 99(3), 628—66.

Reduced Form Models

Dr. Siyi Zhou

CQF Fitch Learning

2020

Main Topics

In this lecture we will

- Model default event using Poisson Process.
- Derive risky bond pricing PDE assuming stochastic interest rate and / or stochastic default intensity.
- Review common recovery assumptions that can be adopted in risky bond pricing.
- Illustrate Fundamental Pricing Formula for general contingent claims subject to default risk.
- Introduce basic theory of affine intensity models.
- Show an example of two-factor Vasicek intensity model.

Take Away

By the end of this lecture you will be able to

- Apply the basic reduced form model to price contingent claim subject to default risk.
- Explain pros and cons of intensity based model relative to structural models.
- Derive risky bond pricing equations assuming stochastic interest rate and / or default intensity.
- Calibrate default probability on bond prices.
- Solve simple affine intensity based models analytically.

Intensity Based (Reduced Form) Model

- In previous lecture we have studied structural approach to the modeling default risk. In this lecture we will introduce a different approach to model default risk in a bond or in a general contingent claim. These models are named as "reduced-form" or "intensity based" models, in which default is treated as an unpredictable event governed by an exogenous intensity process.
- The intensity is linked to (actually determines) the likelihood of default. In addition, the SDEs employed to model intensity are similar to the ones that are used to model short interest rate. Therefore many term structure models that are developed for short interest rate can also be used to model default risk.

The intensity based models, therefore, are one of the most popular credit risk models.

Model Default Risk using Poisson Process

Before embark on valuation of a risky bond, we need to understand the basic concept of intensity and how it can be used to model default events.

The Poisson process is one of the most important stochastic processes in probability theory. It is widely used to model random points in time and space. The process has a beautiful mathematical structure, and is used as a foundation for building a number of more complicated stochastic processes. Like many other discrete and countable events, such as the number of buses will arrive in the next 10 min, default events can be modeled in a Poisson process.

Several important probability distributions arise naturally from the Poisson processthe Poisson distribution, the exponential distribution, and the gamma distribution

Definition of Poisson Process I

A Poisson Process with intensity λ is a stochastic process

$$N_t : t \geq 0$$

taking values in $S = \{0, 1, 2, \cdots\}$ such that

1. $N_0 = 0$
2. if $s < t$, then $N_s \leq N_t$
3. if $s < t$, then the increment $N_t - N_s$ is independent of what happened during $[0, s]$

Definition of Poisson Process II

4. let $h \to 0^+$

$$Pr(N_{t+h} = n + m | N_t = n) = \begin{cases} \lambda h + o(h), & m = 1 \\ o(h), & m > 1 \\ 1 - \lambda h + o(h), & m = 0 \end{cases}$$

which means within an infinitesimal time interval h, maximum only one event can happen.

Distribution function of N_t

N_t has Poisson distribution with parameter λt

$$Pr(N_t = i) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

where

$$i = 0, 1, 2, \cdots$$

N_t is the number of events occurred by time t. It is also called "counting process". We can use Poisson Process as a starting point to model any default event.

Arrival Time in Poisson Process

Define T_n to be the arrival time of the nth event, i.e.,

$$T_n = \inf\{t : N_t = n\}, \quad T_0 = 0,$$

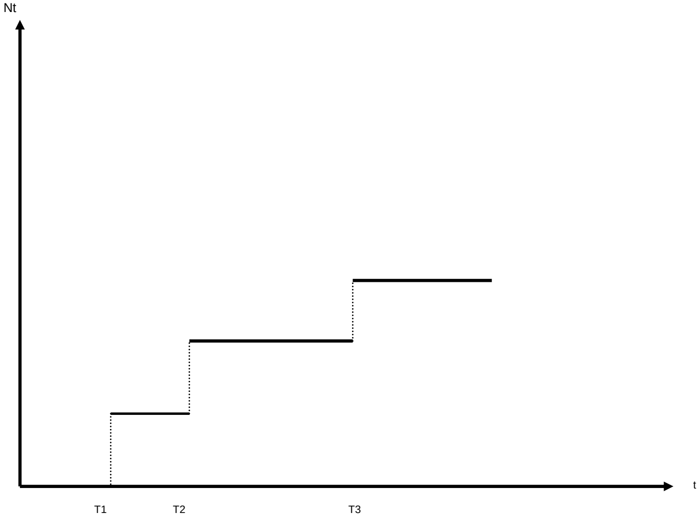
and τ_n to be an inter-arrival time, which is given by:

$$\tau_n = T_n - T_{n-1}.$$

So T_n can be expressed as the sum of all inter-arrive time:

$$T_n = \sum_{i=1}^{n} \tau_i$$

Sample Path of Poisson Process



Distribution of Inter-Arrival Time τ

- Since Poisson Process has independent increments, the associated inter-arrival times are also independent, i.e., $\forall i \neq j$, τ_i and τ_j are independent.
- Further more, every inter-arrival time has an exponential distribution with intensity λ. To summarize τ_i are i.i.d. $\exp(\lambda)$.
- To prove the above, we need to revert back to the 4th property of the Poisson Process which ultimately determines the timing of each jump.

Definition of Intensity

Denote $F(\tau)$ the Cumulative Distribution Function (CDF) of τ_1(the very first event), and define the survival function as:

$$S(\tau) = 1 - F(\tau).$$

According to the definition of Poisson process (point 4), when $h \to 0^+$

$$Pr(t < \tau \leq t + h | \tau > t) = \lambda h + o(h).$$

So

$$\lambda = \lim_{h \to 0^+} \frac{Pr(t < \tau \leq t + h | \tau > t)}{h}$$

Distribution of τ

$$\begin{aligned}\lambda &= \lim_{h\to 0^+}\frac{Pr(t<\tau\le t+h)}{hPr(\tau>t)}\\ &= \lim_{h\to 0^+}\frac{S(t)-S(t+h)}{hS(t)}\\ &= -\frac{d\log S(t)}{dt}\end{aligned}$$

Solve the above ODE for $S(t)$ with initial condition $S(0)=1$, we get

$$S(t)=e^{-\lambda t},$$

which implies

$$S(t)\sim \exp(\lambda).$$

Simulating Poisson Process

So far we have derived distribution for the 1st inter-arrival time, what about the rest of them? By the third property of Poisson Process we know

$$S_2(t) = Pr(\tau_2 > t|\tau_1) = Pr(\tau_2 > t).$$

Then follow the same argument for τ_1 we have

$$\tau_2 \sim \exp(\lambda),$$

and so on for τ_i.

As a result, simulating a Poisson process is equivalent to simulating consecutive i.i.d. exponential random variables. We will see how to do this in subsequent M5 lecture.

Inhomogeneous Poisson process

What if the intensity isn't constant but deterministic? Then the counting process is called inhomogenous Poisson Process. The analysis are almost the same as before. The the survival function in this case becomes

$$S(t) = \exp\left(-\int_0^t \lambda_s \, ds\right).$$

Cox Process

Later on we will derive pricing PDE in which the intensity is not deterministic but a stochastic process. The pure probabilistic approach is slightly more complex and we don't pursue it in more details here. The general idea is to use conditional expectation, i.e., conditional on the filtration to which the path of the intensity is adapted, then the survival function is known. The survival probability in this case is equal to

$$S(t) = \mathbb{E}\left[\exp\left(-\int_0^t \lambda_s \, ds\right)\right].$$

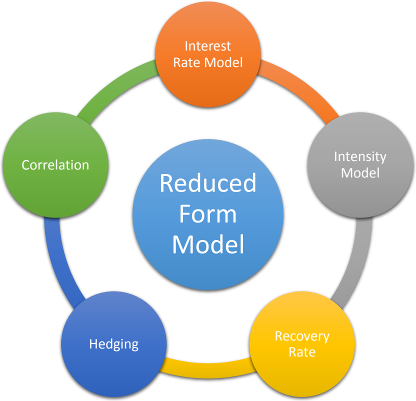
Cox Process: Basic Concept

Later on we will derive risk bond pricing PDEs where the default intensity is stochastic rather than deterministic. **Cox Process** is an inhomogeneous Poisson Process driven by a stochastic intensity. In more rigorous mathematical language we assume

- The stochastic intensity process λ_t which drives the probability of default is adapted to a filtration $\mathcal{G}_t$. This filtration can also contain other market observable underlying variables such as rates, FX and so on.
- Jumps (N_t) in the Cox process is adapted to another filtration $\mathcal{H}_t$.
- So the full filtration is obtained by $\mathcal{F}_t = \mathcal{G}_t \cup H_t$.

The process Nt is called a Cox Process with stochastic intensity h_t, however, conditional on the background filtration $\mathcal{G}_t$, Nt becomes an inhomogeneous Poisson Process with deterministic intensity.

Input for Reduced Form model



BPE Plan

In the next a few slides, We will derive 3 BPEs for a risky bond based on different assumptions made on default intensity, hedging strategy and recovery rate.

In the order of increasing complexity the assumptions made are

1. Constant intensity and zero recovery without the hedging of default risk.
2. Stochastic intensity and zero recovery with the hedging of default risk.
3. Stochastic intensity and positive recovery with hedging of default risk.

We always assume stochastic interest rate and interest rate risk are delta hedged.

Model Assumptions

- Suppose a corporate's default follows a homogenous Poisson process with intensity p, and a ZCB with maturity T is issued by this company, the value of the bond is denoted by $V(t, r; p)$.
- Like usual suppose $Z(t, r)$ is the value of a riskless ZCB with exactly the same maturity where the short interest rate dynamics follows a diffusion process

$$dr = u(r, t)\, dt + w(r, t) dX.$$

- For simplicity we will assume that there is no correlation between the diffusive change in the short interest rate and the Poisson process.

Hedging Portfolio

Now following our convention when pricing fixed income product, we construct a ‘hedged’ portfolio:

$$\Pi = V(r, t; p) - \Delta Z(r, t).$$

Note here only the interest rate risk is hedged.

Case A: Without default

There is a probability of $(1 - p\,dt)$ that the bond does not default. Then the change in the value of the portfolio during an infinitesimal time step dt is

$$d\Pi = \left(\frac{\partial V}{\partial t} + \tfrac{1}{2}w^2\frac{\partial^2 V}{\partial r^2}\right)dt + \frac{\partial V}{\partial r}dr$$
$$-\Delta\left(\left(\frac{\partial Z}{\partial t} + \tfrac{1}{2}w^2\frac{\partial^2 Z}{\partial r^2}\right)dt + \frac{\partial Z}{\partial r}dr\right).$$

- Choose Δ to eliminate the risky dr term.

Case B: With Default

On the other hand, if the bond defaults, with a probability of $p\,dt$, then the change in the value of the portfolio is

$$d\Pi = -V + O(dt^{1/2}).$$

This is due to the default loss of the risky bond, the second term represents the changes in the riskless bond.

Systematic and Idiosyncratic Risk

- In Robert Merton's Jump Diffusion Model, jump components are assumed to be idiosyncratic risk. And the risk associated with the jumps is diversifiable since the jumps in the individual assets is uncorrelated with the market as a whole.
- Similar assumption can be made here, where default of a bond is similar to a jump in a share price.
- Therefore, the beta of the hedge portfolio is zero, the expected return of the zero-beta portfolio should be equal to the risk-free rate.

Risky Bond Pricing Equation

Taking expectation and using the bond-pricing equation of the riskless bond, we find that the value of the risky bond satisfies the following BPE

$$\frac{\partial V}{\partial t} + \tfrac{1}{2} w^2 \frac{\partial^2 V}{\partial r^2} + (u - \lambda w) \frac{\partial V}{\partial r} - (r + p) V = 0.$$

Feynman Kac:1st call

$$V(t,T) = \mathbb{E}\left(e^{-\int_t^T (r_s+p)\,ds} | \mathcal{F}_t\right)$$

Very similar analysis can be carried out with deterministic default intensity, the result will be almost identical apart from changing p to p_t, that is

$$V(t,T) = \mathbb{E}\left(e^{-\int_t^T (r_s+p_s)\,ds} | \mathcal{F}_t\right)$$

Yield Spread

The yield to maturity on this bond is now given by

$$y = -\frac{\log(Z(t,T)S_t(T))}{T-t} = y_f + \frac{1}{T-t}\int_t^T p_s ds,$$

where y_f is the yield to maturity of a risk free bond with the same maturity as the risky bond.

Thus the effect of the risk of default on the yield is to add a spread on riskless yield. In this simple model, the spread will be the average of the hazard rate from t to T.

Forward Rate Spread

If one calculates the forward rate implied by the risky bond

$$-\frac{\partial}{\partial T}\log(V(t,T)) = f(t,T) + p_T.$$

The spread is simply equal to default intensity.

Implied Default Probability: Zero Recovery

Given term structure of risk free bond and risky bond, one can extract implied default probability by using

$$S_t(T) = \frac{V(t, T)}{Z(t, T)} = \exp(-(T - t)(y - y_f)).$$

One can also calculate implied hazard rate by using forward spread.

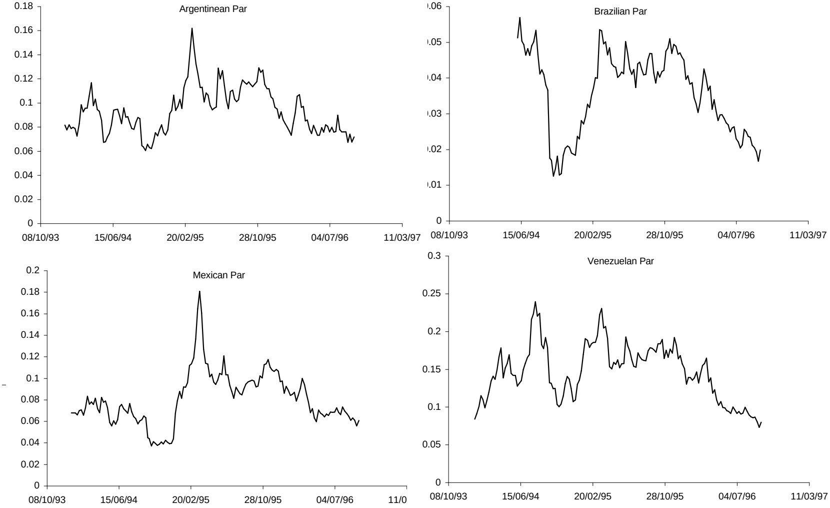
PD Calibration with Zero Recovery Rate

Year	Riskfree zero rate	Risky bond zero rate	Cummulative PD	Marginal PD
1	5%	5.25%	0.2497%	0.2497%
2	5%	5.50%	0.9950%	0.7453%
3	5%	5.70%	2.0781%	1.0831%
4	5%	5.85%	3.3428%	1.2647%
5	5%	5.95%	4.6390%	1.2961%

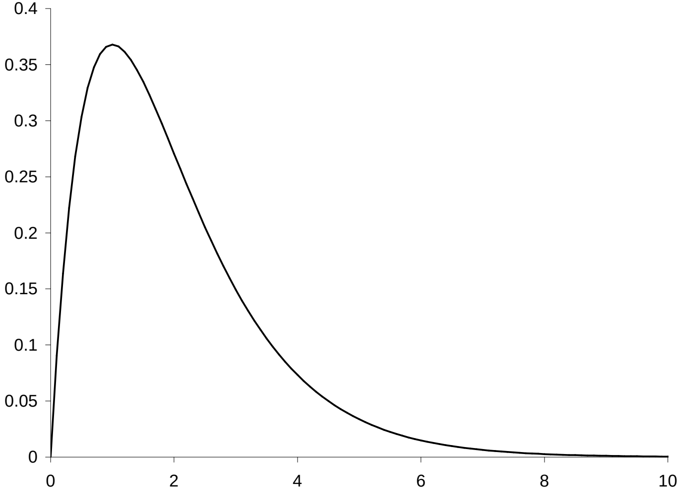
Risky Bond Valuation

1. Find the risk-free yield for the maturity of each cashflow in the risky bond;
2. Add a constant spread, p, to each of these yields;
3. Use this new yield to calculate the present value of each cashflow;
4. Sum all the present values.

Implied PD for 4 Latin American Brady Bonds



Implied Time Dependent Intensity



Stochastic Default Intensity

Now consider a model in which the default intensity is itself random:

$$dp = \gamma(r, p, t)dt + \delta(r, p, t)dX_1,$$

with the interest rate still given by

$$dr = u(r, t)dt + w(r, t)dX_2,$$

where

$$dX_1\, dX_2 = \rho\, dt.$$

Hedging Default Risk

- In the previous model we used riskless bonds to hedge the random movements in the spot interest rate.
- Can we introduce another risky bond or bonds into the portfolio to help with the hedging of the default risk?
- To do this we must assume that default in one bond automatically triggers default in the other bond issued by the same counterparty.

Hedged portfolio

To value our risky zero-coupon bond we construct a portfolio with one of the risky bond, with value $V(r, p, t)$, and delta hedged by shorting Δ unit of a riskless bond, with value $Z(r, t)$, and Δ_1 shorting another risky bond issued by the same company with different maturity, with value $V_1(r, p, t)$:

$$\Pi = V(r, p, t) - \Delta Z(r, t) - \Delta_1 V_1(r, p, t).$$

Case A: Without Default I

Suppose that the bond does not default, the change in the value of the portfolio during an infinitesimal time step is

$$d\Pi = dV - \Delta\, dZ - \Delta_1 dV_1.$$

By using Itô's lemma, above can be written as

$$d\Pi = \left(\mathcal{L}'(V) - \Delta\mathcal{L}(Z) - \Delta_1\mathcal{L}'(V_1)\right) dt$$

$$+ \left(\frac{\partial V}{\partial r} - \Delta\frac{\partial Z}{\partial r} - \Delta_1\frac{\partial V_1}{\partial r} r\right) dr$$

$$+ \left(\frac{\partial V}{\partial p} - \Delta_1\frac{\partial V_1}{\partial p}\right) dp$$

Case A: Without Default II

where

$$\mathcal{L}'(V) = \frac{\partial V}{\partial t} + \tfrac{1}{2} w^2 \frac{\partial^2 V}{\partial r^2} + \rho w \delta \frac{\partial^2 V}{\partial r \partial p} + \tfrac{1}{2} \delta^2 \frac{\partial^2 V}{\partial p^2}$$

$$\mathcal{L}(Z) = \frac{\partial Z}{\partial t} + \tfrac{1}{2} w^2 \frac{\partial^2 Z}{\partial r^2}$$

and ρ is the correlation between dX_1 and dX_2.

Case A: Without Default III

Choose Δ to eliminate the risky terms.

$$\Delta_1 = \frac{\partial V}{\partial p} \Big/ \frac{\partial V_1}{\partial p}$$

and

$$\Delta = \frac{\frac{\partial V}{\partial r} - \Delta_1 \frac{\partial V_1}{\partial r}}{\frac{\partial Z}{\partial r}}$$

Case B: With Default

If the bond defaults then the change in the value of the portfolio is

$$d\Pi = -V + \Delta_1 V_1 + O(dt^{1/2}).$$

Taking expectations and using the bond-pricing equation for the riskless bond, we find that the value of the risky bond satisfies

$$\frac{\partial V}{\partial t} + \tfrac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + \rho w \delta \frac{\partial^2 V}{\partial r \partial p} + \tfrac{1}{2}\delta^2\frac{\partial^2 V}{\partial p^2} +$$

$$(u - \lambda w)\frac{\partial V}{\partial r} + (\gamma - \lambda' \delta)\frac{\partial V}{\partial p} - (r + p)V = 0.$$

Feynman Kac: 2nd call

Similar to the interest rate risk, λ' is called the market price of default risk. So the fundamental pricing formula for the risky bond under risk neutral measure is

$$V(t, T) = \mathbb{E}\left(e^{-\int_t^T (r_s + p_s)ds} | \mathcal{F}_t\right).$$

Recovery Rate Historic Analysis

In default there is usually *some* payment, not all of the money is lost. In the table are shown the mean and standard deviations for recovery according to the seniority of the debt. This emphasizes the fact that the rate of recovery is itself very uncertain.

Class	Mean (%)	Std Dev. (%)
Senior secured	53.80	26.86
Senior unsecured	51.13	25.45
Senior subordinated	38.52	23.81
Subordinated	32.74	20.18
Junior subordinated	17.09	10.90

There is also a statistical relationship between rate of recovery and default rates. (Years with low default rates have higher recovery when there is default.)

Recovery of Market Value

Suppose that on default we know that we will loss l percent of pre-default value. This will change the partial differential equation.

Upon default we have

$$d\Pi = -lV + l\Delta_1 V_1 + O(dt^{1/2});$$

we suffer a loss from the first bond but gain $1 - l$ from the second bond. The pricing equation becomes

$$\frac{\partial V}{\partial t} + \tfrac{1}{2}w^2\frac{\partial^2 V}{\partial r^2} + \rho w\delta\frac{\partial^2 V}{\partial r\partial p} + \tfrac{1}{2}\delta^2\frac{\partial^2 V}{\partial p^2}$$

$$+(u - \lambda w)\frac{\partial V}{\partial r} + (\gamma - \lambda'\delta)\frac{\partial V}{\partial p} - (r + lp)V = 0.$$

Recovery of Treasury Value

- Although the assumption of recovery on market value is convenient for the purpose of mathematical modeling and makes economic sense since it measures the loss in value associated with default, it is impossible to give immediate expression for implied default probability.
- Recovery on treasury assumes that, if a corporate bond defaults, its value will be replaced by a treasury bond with the same maturity. Under this assumption and with the independence of interest rate and hazard rate, the bond price will be

$$V(0,t) = (1 - F(t))\, Z(0,t) + F(t)\theta Z(0,t).$$

- Implied default probability can be easily extracted from the above relationship.

PD Calibration with recovery of Treasury Value

Year	Riskfree zero rate	Risky bond zero rate	Cummulative PD	Marginal PD
1	5%	5.25%	0.4161%	0.4161%
2	5%	5.50%	1.6584%	1.2422%
3	5%	5.70%	3.4635%	1.8051%
4	5%	5.85%	5.5714%	2.1079%
5	5%	5.95%	7.7316%	2.1602%
	Recovery	40%		

Feynman Kac: 3rd Call

$$V(t,T) = \mathbb{E}\left(e^{-\int_t^T (r_s + lp_s)ds} | \mathcal{F}_t\right).$$

It can be rewritten as

$$V(t,T) = \mathbb{E}\left(e^{-\int_t^T R_s ds} | \mathcal{F}_t\right).$$

Where

$$R_t = r_t + l\, p_t,$$

is called the risk adjusted discount rate.

Introduction

- In M4, we have seem affine short rate models lead to explicit solutions for bond prices, e.g. Vasicek, CIR and Ho& Lee etc.
- Similarly, affine intensity based models lead to analytical solutions for risk bond prices.

Pricing for a General Contingent Claim

For a general contingent claim $g(X_T)$ at time T, where

$$X_t = (x_{1t}, x_{2t}, \cdots, x_{nt})$$

is a vector of state variable, The fundamental pricing formula is

$$V(t,T) = \mathbb{E}\left(e^{-\int_t^T R(X_s)ds} g(X_T) | \mathcal{F}_t\right).$$

Solution of General Affine Model

If the model is affine, i.e.,

- X_t: affine process
- $R(X_t)$: affine in X_t
- $g(X_t)$: affine is X_t

then the general solution is

General Affine Model Solution

$$V(t, T) = e^{\alpha(t,T)+\beta(t,T)X_t}.$$

Affine Conditions

In previous section while we deriving risky bond pricing equation, we have the stochastic interest rate

$$dr = u(r,t)dt + w(r,t)dX_1,$$

and the stochastic intensity

$$dp = \gamma(p,t)dt + \delta(p,t)dX_2.$$

To be affine intensity model

- we must choose the functions $u - \lambda w$, w, $\gamma - \lambda' \delta$, δ and ρ carefully.
- We must choose $u - \lambda w$ and w^2 to be linear in state variables, same for $\gamma - \lambda' \delta$ and δ^2.
- The form of the correlation coefficient is assumed to be constant, i.e., $dX_1 dX_2 = \rho \, dt$.

Solution for Risky ZCB

Suppose recovery rate is a constant $\theta = 1 - l$, and $s_t = (1 - \theta)p_t$.
With appropriate choices of the functions in the two stochastic differential equations we find that the solution with final condition $V(r, s, T) = 1$ is

$$V = \exp\{A(t, T) - B(t, T)r - C(t, T)s\}$$

where A, B and C satisfy non-linear first-order ordinary differential equations.

Calibration

- In M4 fixed income, if we allow the spot interest rate model to have some simple time dependence then we have the freedom to fit the initial yield curve.
- Similarly, if there is time dependence in the model for the intensity, and the model is sufficiently tractable, then you can also fit risky bond term structure.

Standard Two-Factor Vasicek Model

Suppose there are two state variables X and Y whose dynamics can be written as

$$\begin{aligned} dX &= (a_1 - b_{11}X - b_{12}Y)dt + \sigma_1 dW_1 \\ dY &= (a_2 - b_{21}X - b_{22}Y)dt + \sigma_2 dW_2 \end{aligned}$$

where

$$dW_1 dW_2 = \rho dt$$

The risk adjusted discount rate is

$$R = g_0 + g_1 X + g_2 Y.$$

Modified Canonical Form

To simplify the parameterizations(standard form doesn't usually have unique solution) we work with demeaned canonical form

$$\begin{aligned} dX &= -a\,X dt + \sigma dW_1 \\ dY &= -b\,Y dt + \eta dW_2 \end{aligned}$$

and

$$R(t) = \phi(t) + X(t) + Y(t).$$

Note in order to calibrate on the risky bond yield we employ a time dependent parameter $\phi(t)$ into risk adjusted rate. Here we correspond the state variable X to the short rate r and Y to the spread $s = (1-\theta)p$.

Two-factor Vasicek risky BPE

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial x^2} + \frac{1}{2}\eta^2 \frac{\partial^2 V}{\partial y^2} + \sigma\eta\rho \frac{\partial^2 V}{\partial x \partial y} - ax\frac{\partial V}{\partial x} - by\frac{\partial V}{\partial y} - RV = 0$$

Risky Bond Solution I

Plug the general affine solution into the BPE, come up with 3 ODEs w.r.t A, B and C respectively, solve them one by one to obtain the price of risky bond.

$$\begin{aligned} V(t,T) &= \exp\left\{-\int_t^T \phi(s)ds - \frac{1-e^{-a(T-t)}}{a}X(t) - \right. \\ &\qquad \left. \frac{1-e^{-b(T-t)}}{b}Y(t) + \tfrac{1}{2}M(t,T)\right\} \end{aligned}$$

Risky Bond Solution II

where

$$
\begin{aligned}
M(t,T) \quad &= \quad \frac{\sigma^2}{a^2}\left[T - t + \frac{2}{a}e^{-a(T-t)} - \frac{1}{2a}e^{-2a(T-t)} - \frac{3}{2a}\right] \\
&+ \quad \frac{\eta^2}{b^2}\left[T - t + \frac{2}{b}e^{-b(T-t)} - \frac{1}{2b}e^{-2b(T-t)} - \frac{3}{2b}\right] \\
&+ \quad 2\rho\frac{\sigma\eta}{ab}\left[T - t - \frac{1 - e^{-a(T-t)}}{a} - \frac{1 - e^{-b(T-t)}}{b}\right. \\
&- \quad \left.\frac{e^{-(a+b)(T-t)-1}}{a+b}\right]
\end{aligned}
$$

Structural Model Vs Reduced Form Model

1. Structural models assume that the modeler has the same information set as the firm's manager-complete knowledge of all the firm's assets and liabilities. In contrast, reduced form models assume that the modeler has the same information set as the market-incomplete knowledge of the firm's condition.
2. Structural models use a firm's asset and debt values to determine the time of default, thus defaults are endogenously generated within the model. In contrast, the time of default in intensity models is determined by the first jump of an jump process whose hazard rate is given by exogenous stochastic process.

Take Away

Please Take Away the following important ideas

- Poisson process assumes constant intensity and it can be used to model default.
- Risky bond is discounted by risk-adjusted interest rate.
- stochastic default intensity increases dimension of bond pricing partial differential equation.
- Risky bond pricing partial differential equationis consistent with fundamental pricing formula through Feynman Kac.
- Reduced-form models are tractable when it satisfies affine structure.

CDO, Copula and Correlation

Dr. Siyi Zhou

CQF FitchLearning

2020

1. Introduction to CDO
   - What is CDO
   - CDO market pricing and risk management
   - Derive CDO Pricing Equation
2. Introduction to Copula Function
   - Motivation from Loss distribution
   - What is Copula function
   - Classification of copula functions
   - Simulating via Gaussian copula
3. Gaussian Copula Factor Model
   - Introduction
   - One factor model
   - Implementation
4. Analytical Solution for LHP
5. Conclusion

Main Topics

In this lecture we will:

1. Briefly review Collateralized Debt Obligation(CDO) from the perspective of pricing problem
2. Derive pricing formula for a synthetic Collateralized Debt Obligation
3. Introduce the basic concept and mathematical properties of copula models
4. Simulate default times using Gaussian copula
5. Introduce one factor model for credit risk modelling
6. Derive analytic solution for large homogenous portfolio based on one factor model
7. Analyse the impact of default correlation under asymptotic one factor model

Take Away

By the end of this lecture you will be able to

- Understand basic concept and mathematics of copula model
- How to implement Gaussian and / or t copula model to generate correlated default times
- Price Collateralized Debt Obligationand general credit derivatives using copula models
- Understand what roles default correlation play in credit risk models
- Apply factor models to price Collateralized Debt Obligationand general credit derives
- Be able to derive analytical solution for credit derives based on asymptotic one factor model

What is CDO?

CDO is a type of asset backed security, it is an investment on a pool of diversified assets (bonds, loans, CDS etc) in the form of tranched securities. Instead of directly selling the asset pool, the sponsor of CDO (banks, non-financial institutions, asset management companies) creates an independent legal entity called Special Purpose Vehicle(SPV) who repacks the asset pool and slices it into tranches according to the underlying asset pool's credit risk.

Typical CDO tranche

The tranches of a typical CDO are

- Senior tranche(AAA)
- Mezzanine tranche(AA to BB)
- Equity tranche(unrated)

The position of each tranche within the CDO capital structure is determined by its attachment and detachment points.

How CDO Works

- At good times, tranche Investors will receive premium (returns from asset pool) in turn periodically
- At bad times, when the total pool loss reaches the attachment point, investors in the tranche start to lose their capital, and when the total pool loss reaches the detachment point, the investors in the tranche lose all their capital and no further loss can occur to them
- Loss will be applied in reverse order of seniority
- The senior tranche is protected by the Mezzanine tranche and equity tranche; while mezzanine tranche is protected only by equity tranche

Motivation

First issued in the late 1980s, CDO emerged a decade later as the fastest growing sector in financial market.Banks that are actively trading CDO are motivated by two reasons, they attempt to

- explore arbitrage opportunities. The majority of CDO are arbitrage motivated
- off load credit risk from their loan book, and hence to reduce their regulatory capital requirement and improve return on risk capital without termination of their customer relationship

Cash CDO

CDO is a broad term that can refer to several different types of products, it can be categorized in several ways. From pricing point of view we are interested in distinguish CDO with respect of its funding,i.e., cash CDO or synthetic CDO.

Cash CDO involves a portfolio of cash assets, such as corporate bonds, loans, etc. Ownership of the assets is transferred to a SPV who is the issuer of CDO.

Synthetic CDO

If the SPV of a CDO does not own the physical asset pool, instead obtaining the credit risk exposure by selling CDSs on the reference portfolio, the CDO is referred to as a synthetic CDO.

Synthetic CDO can be unfunded, which means investors only pay when their tranches are affected by defaults. In this case, counterparty default risk must be taken into account by risk managers.

CDO Market

Overall, the CDO market consists of

- an illiquid segment, for example cash CDO "buy-to-hold" investors whose investment decisions are mainly based on ratings and yields
- an actively traded segment in which the underlying credit portfolio is based on the standardized CDS index (CDX) such as the iTraxx (European) or CDX (North American) index. This market are mostly traded by the correlation desks of hedge funds and banks

Because the net cash flows of index tranches are the same as synthetic CDO tranches, these tranches can be priced the same way as a synthetic CDO.

CDO Valuation and its Risk Management I

In the illiquid segment, CDO pricing and risk measurement are difficult.

- Due to lack of liquidity, hedging and market-to-market are generally unavailable
- A cash CDO often has complex waterfall structure so that the cashflow of its tranches are highly path-dependent, and it is generally actively managed, so structural and cashflow analysis are complicated. Plus key parameters like default correlation is difficult to estimate, so market-to-model is very weak
- Traditionally investors rely on rating agencies for their valuation and risk assessment, but since credit crunch 2007, rating agencies' credibility has been severely undermined

CDO Valuation and its Risk management II

However in the synthetic CDO market

- the cashflows of synthetic CDO tranches are very simple and not path-dependent;
- market participants can easily take long or short positions in the CDX and index tranches, and the underlying single name CDS
- because of strong arbitrage relationships among them, the basis between indices, tranches or single names CDS in the synthetic market tends to stay within a reasonable range

Therefore the valuation and risk management in this market are more feasible. In this lecture in order to help us focus on credit risk modelling we will price synthetic CDO, in particular we will treat CDO as a derivative of a portfolio of CDS contracts.

Notation

Let's get familiar with some notations before deriving the pricing equation

- Survival time for reference name i: τ_i
- Loss given default for reference name i: LGD_i
- exposure at default for reference name i: EAD_i
- Tranche: $[D, U]$
- Settlement date: t_j
- Payment frequency: $\Delta = t_{j+1} - t_j$
- Maturity date: T_M
- discount factor: $Z(t, T)$

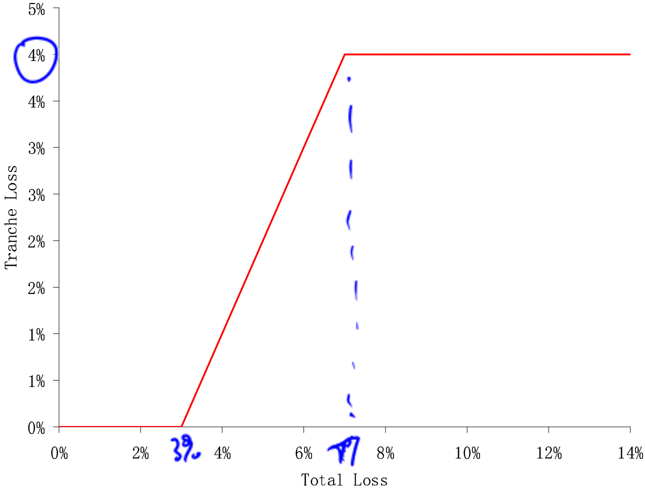
Loss Function

1. the loss for reference name i by time t is $L_i(t)$,
2. the total reference pool loss by time t is $L(t)$,
3. the loss for tranche $[d, u]$ by time t is $L(t; d, u)$.

By definition

$$
\begin{aligned}
L_i(t) &= LGD_i * EAD_i * I\{\tau_i < t\} \\
L(t) &= \sum_{i=1}^{N} L_i(t) \Big/ \sum_{i=1}^{n} EAD_i \\
\underbrace{L(t; u, d)}_{\varphi} &= \underbrace{\max\left[\min\left(L(t), u\right) - d, 0\right]}_{\varphi}
\end{aligned}
$$

Mezzanine Tranche Payoff



Synthetic CDO Pricing I

The present value of protection leg (assuming paid in arrear) is

marginal loss for t_j

$$\sum_{j=1}^{M} Z(0,t_j)\left[L(t_j;d,u) - L(t_{j-1};d,u)\right], \quad (1)$$

Assuming the fair tranche spread is s, so the present value of premium leg is

$$s\Delta \sum_{j=1}^{M} Z(0,t_j)\left[(u-d) - L(t_j;d,u)\right], \quad (2)$$

s : spr

E(1) = E(2)
Solve for S

Synthetic CDO Pricing II

Same principle as pricing interest rate swap or credit default swap, present value of both legs must be equal at inception of the contract.

Now take expectation of present value of both legs and equate them, the fair spread s for tranche (d, u) is

$$\boxed{s = \frac{\mathbb{E}\left\{\sum_{j=1}^{M} Z(0, t_j)\left[L(t_j; d, u) - L(t_{j-1}; d, u)\right]\right\}}{\Delta \mathbb{E}\left\{\sum_{j=1}^{M} Z(0, t_j)\left[(u - d) - L(t_j; d, u)\right]\right\}}} \tag{3}$$

? $L(t; d, u)$

Portfolio Loss Distribution

From equation (**??**), one can see that the key input to price a CDO tranche is the expected tranche loss which in turn depends on the distribution of entire reference portfolio. To derive this loss distribution, we need know the following:

1. Distribution of joint default
2. Specification of default risk parameters for each obligor, i.e. EAD and LGD

Suppose that it is not too difficult to make assumptions on loss give default and exposure at default for each name, so the main task to is to figure out what the joint default distribution of the asset pool is.

Marginal and Joint Distribution

By definition marginal distribution of a random variable X is

$$\boxed{F(x) = \Pr(X \leq x),}$$

r.v.

$X \leq x$

arbitrage number.

The joint distribution function of two random variables X and Y is

$$F(x, y) = \Pr(X \leq x, Y \leq y).$$

We can model default risk of a credit portfolio if we know joint default distribution function

$$\boxed{F(t_1, t_2, \cdots, t_n) = \Pr(\tau_1 \leq t_1, \tau_2 \leq t_2, \cdots, \tau_n \leq t_n).}$$

Problem with Joint Distribution

Directly work on joint distribution is inconvenient, because

- Marginal distribution are different, conventional joint distribution only accepts homogenous marginal distribution
- Extension to higher dimension maybe difficult
- Measures of dependence may appear in marginal distribution

Copula Approach

Instead of directly working on joint distribution which must be horrendous, the better way is to use copula function. Unlike joint distribution function, copula function can separate marginal distribution and their association completely, as a result, by copula, one can conveniently mix marginal distributions together with certain dependence structure to become a joint distribution.

Let's see how one can do that.

Definition of Copula

Definition [Copula function]: For k uniform random variables $(U_1, U_2, \cdots, U_k)$, the joint distribution function

$$C(u_1, u_2, \cdots, u_k; \rho)$$

is called a copula function. This definition shows that C is a multivariate distribution function with uniformly distributed marginals.

Copula Links to Joint Distribution

Since the distribution function of a random variable is uniformly distributed, so copula function can be used to link marginal distribution with a joint distribution.

Theorem: Suppose there are random variables $X_1, X_2, \cdots, X_k$, with distributions $F_1, F_2, \cdots, F_k$, then

$$C(F_1(x_1), F_2(x_2), \cdots, F_k(x_k)) = F(x_1, x_2, \cdots, x_k)$$

.

Sklar Theorem

Sklar proved converse version of above, that is any joint distribution can be written in the form of a copula function, and if the joint distribution is continuous then the copula function is unique.

Sklar's theorem shows that copula function can be used to model dependence structure. For any joint distribution function, the marginal distribution and the dependence structure can be isolated, with the latter completely described by copula.

Copula Density Function

By Sklar's theorem, one can derive the density function of a multivariate copula function.

$$
\begin{aligned}
f(x_1, \cdots, x_n) &= \frac{\partial^n \left[C(F_1(x_1), \cdots, F_n(x_n))\right]}{\partial F_1(x_1), \cdots, \partial F_n(x_n)} \prod_{i=1}^{n} f_i(x_i) \\
&= c\left(F_1(x_1), \cdots, F_n(x_n)\right) \prod_{i=1}^{n} f_i(x_i)
\end{aligned}
$$

So the density function of copula is

$$
c\left(F_1(x_1), \cdots, F_n(x_n)\right) = \frac{f(x_1, \cdots, x_n)}{\prod_{i=1}^{n} f_i(x_i)}
$$

With the copula density function it is possible to calibrate its parameters to market data via maximum likelihood estimation.

Classification of Copula

There are several families of copula functions, those are often used in quantitative finance due to their tractability are

- Elliptical copulae
  - Gaussian copula
  - t copula
- Archimedean copulae
  - Gumbel Copula
  - Clayton Copula
  - Frank copula

In the application of CDO, we will pay great attentions to Elliptical family, particularly for Gaussian copula.

Multivariate Gaussian Copula

Definition [Gaussian copula] Let Φ_n be the normalized multivariate standard normal distribution function and Φ be univariate standard normal distribution function, the multivariate Gaussian copula function is

$$C(u_1, u_2, \cdots, u_n) = \Phi_n\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \cdots, \Phi^{-1}(u_n); \Sigma\right)$$

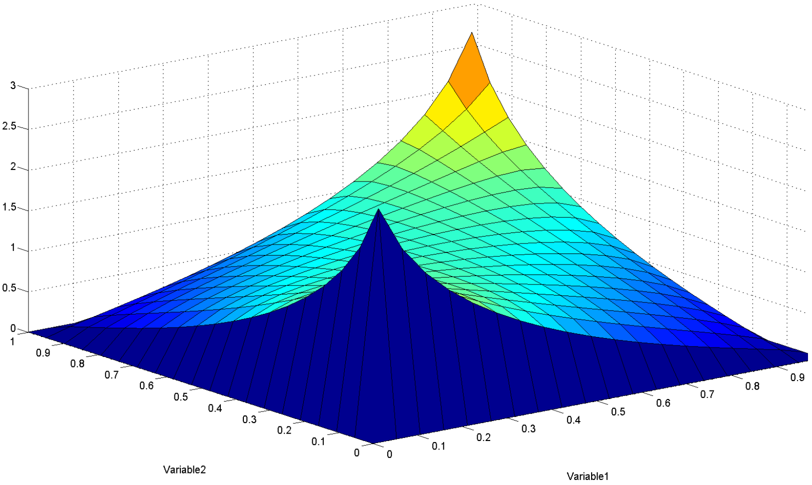
Its density function is

$$c\left(\phi(x_1), \cdots, \phi(x_n)\right) = \frac{\frac{1}{\sqrt{2\pi^n|\Sigma|}} \exp\left(-\frac{1}{2}X'\Sigma^{-1}X\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right)}$$

or in terms of marginal

$$c(u_1, \cdots, u_2) = \frac{1}{\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}\Phi^{-1}(U')(\Sigma^{-1} - I)\Phi^{-1}(U)\right)$$

Bivariate Gaussian Copula Density ($\rho = 0.5$)



Multivariate Student's t Copula

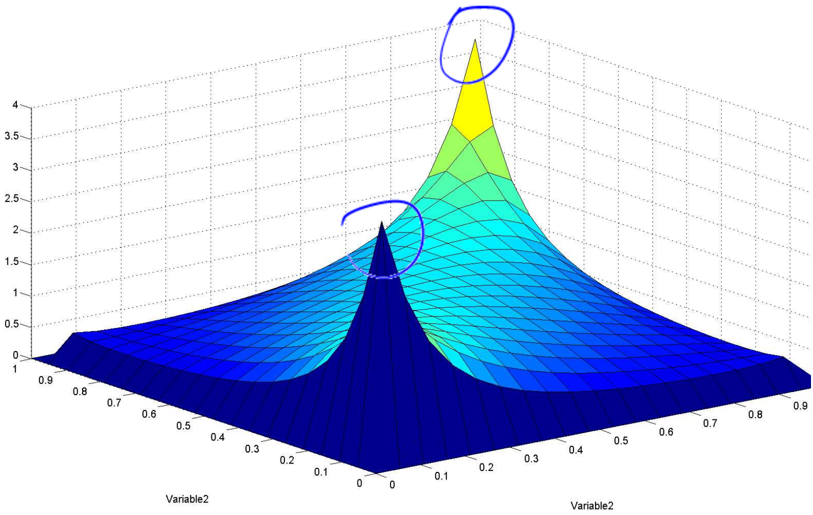
Definition [Student's t Copula]: Let T_v be the normalized multivariate Student's t distribution function with v degrees of freedom, and t_v is the normalized univariate t distribution function also with v degrees of freedom, then the multivariate Student's t Copula is

$$C(u_1, u_2, \cdots, u_n) = T_v\left(t_v^{-1}(u_1), t_v^{-1}(u_2), \cdots, t_v^{-1}(u_n); \Sigma\right)$$

The Student's t copula density function is

$$c(u_1, \cdots, u_2) = \frac{1}{\sqrt{|\Sigma|}} \frac{\Gamma\left(\frac{v+n}{2}\right)}{\Gamma\left(\frac{v}{2}\right)} \left(\frac{\Gamma\left(\frac{v}{2}\right)}{\Gamma\left(\frac{v+1}{2}\right)}\right)^n \frac{\left(1 + \frac{T_v^{-1}(U')\Sigma^{-1} T_v^{-1}(U)}{v}\right)^{-\frac{v+n}{2}}}{\prod_{i=1}^n \left(1 + \frac{T_v^{-1}(u_i)^2}{v}\right)^{-\frac{v+1}{2}}}$$

Bivariate Student's t copula density $\rho = 0.5, v = 3$



Gaussian Copula - Simulate Correlated Default Times

In our example of n assets portfolio, we assume, for each name i, the marginal distribution is $F_i(\tau_i)$, and suppose the correlation matrix is successfully estimated, so the normal copula for the joint survival time is

$$C\left(F_1(\tau_i), \cdots, F_n(\tau_n)\right) = \Phi_n\left(\Phi^{-1}(F_1(\tau_1)), \cdots, \Phi^{-1}(F_n(\tau_n)); \Sigma\right),$$

in which we define

$$x_i = \Phi^{-1}(F_i(\tau_i)).$$

We need to find out what τ_i is, $\forall i = 1, 2, \cdots, n$ through normal copula function. The first step is to generate correlated random variables x_is.

How to Simulate Correlated Multivariate Normal Distribution

Short answer is, instead of generating correlated multivariate normal variables directly, one can generate independent normal variables and then convert them into correlated ones according to predetermined correlation matrix.

Let's see how to do it mathematically.

Notation

Let's denote

$$\mathbf{Z}^T = (z_1, z_2, \cdots, z_d)$$

be an independent d-dimensional standard normal vector.

The most effective method to create correlated normal vector is just by linearly combining independent normal vector. So introduce a $n \times d$ matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nd} \end{pmatrix},$$

where each row represents a vector of weights allocated to elements in $\mathbf{Z}$.

Correlated Normal

Define a new vector

$$\mathbf{X}^T = (x_1, \cdots, x_n),$$

such that

$$\mathbf{X} = \mathbf{A}\,\mathbf{Z},$$

where

$$x_i = \sum_{j=1}^{d} a_{ij} z_j$$

Vector $\mathbf{X}$ is then a linear combination of independent normal vector $\mathbf{Z}$. Note the dimension is changed from d to n.

Covariance Matrix

Define the covariance matrix of $\mathbf{X}$ be $\mathbf{\Sigma}$, then by definition it is,

$$\mathbf{\Sigma} = E\left[\begin{pmatrix} x_1^2 & x_1x_2 & \dots & x_1x_n \\ \vdots & \dots & \dots & \vdots \\ x_nx_1 & x_nx_2 & \cdots & x_n^2 \end{pmatrix}\right] = E\left[\mathbf{X}\mathbf{X}^T\right]$$

Plug $\mathbf{X} = \mathbf{A}\,\mathbf{Z}$ into above equation come up with

$$\mathbf{\Sigma} = E[\mathbf{A}\mathbf{Z}\mathbf{Z}^T\mathbf{A}^T] = \mathbf{A}\mathbf{A}^T$$

So given Covariance matrix, if one can find $\mathbf{A}$, then by multiply independent vector $\mathbf{Z}$ by $\mathbf{A}$ end up with correlated vector $\mathbf{X}$.

Matrix Factorization

The act of Finding matrix $\mathbf{A}$ is called Matrix Factorization or Decomposition.

We all know that non-negative numbers have real square root, whereas negative number doesn't.

Similar result holds for matrices. Any symmetric at least semi-positive definite matrix, like $\mathbf{\Sigma}$ can be factorized. But the solution is not unique.

Methods to decompose covariance matrix

We are going to introduce two popular methods of decomposing correlation matrix,

- Cholesky Factorization
- Spectral Decomposition

Cholesky Factorization

The basic ideal of Cholesky Factorization is very easy, it claims that any symmetric positive definite matrix can be factorized in the form of triangular matrices.

Two Dimension Example I

The best way to see this via looking example. Let's suppose that $\mathbf{\Sigma}$ is a two-dimensional matrix:

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

Cholesky Factorization takes the form:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix} \times \begin{pmatrix} A_{11} & A_{21} \\ 0 & A_{22} \end{pmatrix}$$

2D Example II

$$\begin{pmatrix} A_{11}^2 & A_{11}A_{21} \\ A_{21}A_{11} & A_{21}^2 + A_{22}^2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

It will end up with 3 equations for 3 unknowns like this:

$$\begin{cases} A_{11}^2 = \sigma_{11} \\ A_{21}A_{11} = \sigma_{12} \\ A_{21}^2 + A_{22}^2 = \sigma_{22} \end{cases}$$

One can solve for A_{ij} sequentially, the answer is.

$$\begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}$$

General Algorithm for Multi Dimension

For the case of a d-dimension covariance matrix $\mathbf{\Sigma}$, we need to solve

$$\begin{pmatrix} A_{11} & & & \\ A_{21} & A_{22} & & \\ \vdots & \vdots & \ddots & \\ A_{d1} & A_{d2} & \cdots & A_{dd} \end{pmatrix} \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{d1} \\ & A_{22} & \cdots & A_{d2} \\ & & \ddots & \vdots \\ & & & A_{dd} \end{pmatrix} =$$

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ & \sigma_{22} & \cdots & \sigma_{2d} \\ & & \ddots & \vdots \\ & & & \sigma_{dd} \end{pmatrix}$$

General algorithm for multi dimension continue 2

Traversing the σ_{ij} by looping over i and then j produces,

$$\begin{aligned}
A_{11}^2 &= \sigma_{11} \\
A_{11}A_{21} &= \sigma_{12} \\
&\vdots \\
A_{11}A_{d1} &= \sigma_{1d} \\
A_{21}^2 + A_{22}^2 &= \sigma_{22} \\
&\vdots \\
A_{21}A_{d1} + A_{22}A_{d2} &= \sigma_{2d}
\end{aligned}$$

Exactly one new entry of the A matrix appears in each equation, making it possible to solve for the individual entries sequentially.

General Algorithm for Multi Dimension Continue III

More compactly from,

$$\sigma_{ij} = \sum_{k=1}^{i} A_{ik} A_{jk} \quad j \geq i,$$

We get have basic identity

$$A_{ji} = \left(\sigma_{ij} - \sum_{k=1}^{i-1} A_{ik} A_{jk} \right) / A_{ii} \quad j \geq i,$$

and

$$A_{ii} = \sqrt{\sigma_{ii} - \sum_{k=1}^{i-1} A_{ik}^2} \quad j = i$$

This formulae make a simply recursion to find Cholesky factor.

Spectral Decomposition I

spectral decomposition mainly relies on the fact that eigenvector of a symmetric matrix is orthogonal to each other.

Based on linear algebra, the spectral decomposition of of a symmetric matrix takes form

$$\mathbf{\Sigma} = \mathbf{V\Lambda V}^T$$

$\mathbf{V}$ is a matrix which collects eigenvectors in its column, and $\mathbf{\Lambda}$ is a diagonal matrix with its diagonal elements are eigenvalues of $\mathbf{\Sigma}$.

Spectral Decomposition II

If $\mathbf{\Sigma}$ is positive semi-definite, it can be expressed as

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{V}' = \left(\mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}\right)\left(\mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}\right)'$$

Thus

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}^{\frac{1}{2}}$$

Pros and Cons

Cholesky Factorization has particular structure providing a computational advantage. Spectral Decomposition doesn't have it and hence isn't faster than Cholesky.

In addition to Cholesky's inability to deal with semi-definite matrix, Spectral Decomposition do however have a statistical interpretation that is occasionally useful, that is related to Principal Component Analysis (PCA).

Gaussian Copula Simulation Procedure

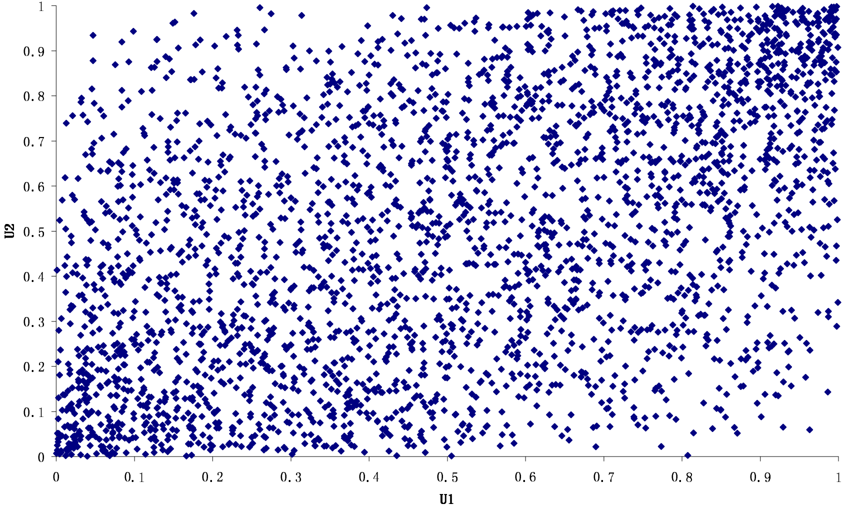
The procedure for generating random default times from normal copula with correlation maxtrix Σ proceeds as follows

1. Find a suitable (e.g. Cholesky) decomposition A from Σ, such that $\Sigma = AA'$
2. Draw a N-dimensional independent standard normal vector $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_n)'$
3. Let $X = AZ$ to obtain correlated normal vector
4. Calculate default time use the following equation

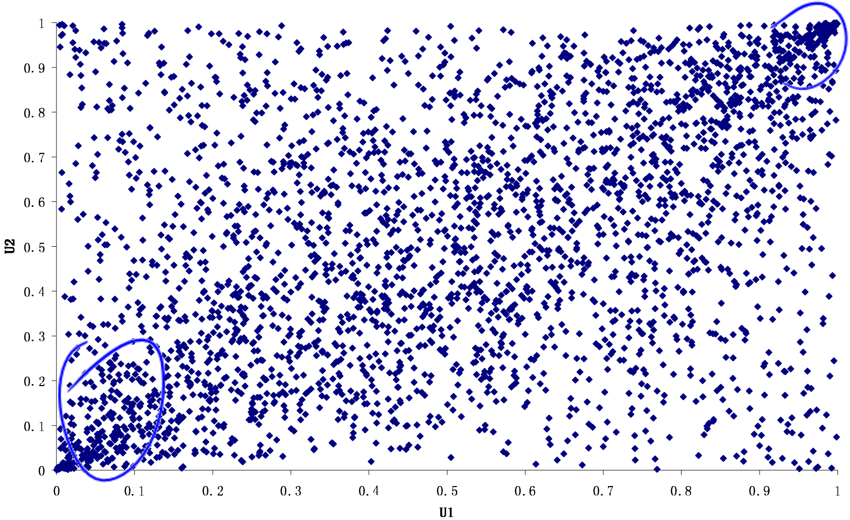
$$\tau_i = F^{-1}\left(\Phi(x_i)\right) \tag{4}$$

5. repeat 2 to 4 many times.

Simulated Bivariate Gaussian Copula ($\rho = 0.5$)



Simulated Bivariate Student's t copula ($\rho = 0.5, v = 2$)



Factor Model

For the time being we have seen copula models which are explicitly based on copula function. Actually there exits another type of copula model which has factor representation, these models have its advantages and are widely used by financial industry.

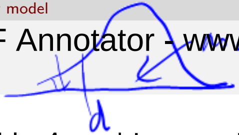
Factor model does the same job like explicit copula model to obtain joint distribution by mixing marginal distributions with measure of dependence. However, it is more easy to understand and less computation effort.

Asset Value Approach

One way to obtain loss distribution is asset value approach which was initially developed by CreditMetrics in 1990s.

The heart of this method is the assumption that, for each obligor, there exists an latent variable which determines the occurrence of default event. Since the the latent variable is commonly viewed as proxy to asset value, this approach is therefore in line with structure model.

Latent Variable



For each obligor i, there exists a latent variable A_i and its associated threshold d_i such that

$$\begin{aligned} \text{Obligor i default} \quad &\iff \quad A_i \leq d_i \\ \text{Obligor i not default} \quad &\iff \quad A_i > d_i \end{aligned} \tag{5}$$

Where

$$A_i = w_i Z + \sqrt{1 - w_i^2}\varepsilon_i \tag{6}$$

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j; \quad \text{cov}(Z, \varepsilon_i) = 0, \forall i$$

Where Z and ε_i are standard normal variables. So by construction A_i is also standard normal. w_i is called factor loading or sensitivity which ultimately links to correlation.

Default Correlation

Instead of directly imposing correlation structure on default themselves, the asset value approach represents it by imposing structure on latent variables.

$$\rho_{ij} = \text{cov}(A_i, A_j) = w_i\, w_j \tag{7}$$

Above can be easily shown by following:

$$\begin{aligned}
\text{cov}(A_i, A_j) &= \text{cov}(w_i Z + \cancel{\sqrt{1-w_i^2}\varepsilon_i},\ w_j Z + \cancel{\sqrt{1-w_j^2}\varepsilon_j}) \\
&= \text{cov}(w_i Z,\ w_j Z) = w_i w_j \text{var}(Z) \\
&= w_i w_j
\end{aligned}$$

Implementation:Monte Carlo simulation

One can easily implement asset value approach to obtain loss distribution by Monte Carlo simulation

1. Using (**??**), randomly draw latent variable for each obligor
2. For each obligor check if it defaulted according to (**??**). If yes, determine individual loss
3. Aggregate individual losses into portfolio loss
4. Repeat steps 1 to 3 many times to arrive at portfolio loss distribution

Homogenous portfolio

A portfolio is homogenous if each name in the portfolio shares the same default probability, recovery rate, correlation and notional principal.

The loss distribution for homogenous portfolio has closed form solution.

Default probability

Recall the one factor normal copula model, the asset value of a firm i is driven by

$$\boxed{A_i = w_i Z + \sqrt{1 - w_i^2}\,\varepsilon_i.} \tag{8}$$

We can drop subscript off since the reference portfolio is homogenous. The default probability for any i is

$$\boxed{F(t) = \Pr[\tau < t]} = \boxed{\Pr[A < d(t)]},$$

where $d(t)$ is default threshold.
Because of the assumption of normality across all i

$$F(t) = \Phi\left(d(t)\right),$$

from where the default threshold can be derived

$$\boxed{d(t)} = \Phi^{-1}(F(t)).$$

Conditional default probability

Conditional on common factor Z, the default probability will be

$$\begin{aligned} F(t|Z) &= \Pr\left[wZ + \sqrt{1-w^2}\,\varepsilon < d(t) \mid Z\right] \\ &= \Pr\left[\varepsilon < \frac{d(t) - w\,Z}{\sqrt{1-w^2}} \mid Z\right] \\ &= \Phi\left(\frac{d(t) - w\,Z}{\sqrt{1-w^2}}\right). \end{aligned}$$

$\rho = w^2.$

Define $\rho = w^2$, then

$$F(t|Z) = \Phi\left(\frac{d(t) - \sqrt{\rho}\,Z}{\sqrt{1-\rho}}\right). \tag{9}$$

Conditional independence

Conditioning on Z, the default of an obligor will be independent of any other obligor. As a result, the number of default conditional on Z follows binomial distribution.

Let N the the size of the reference portfolio, and K be the number of default occurred before time t, so

$$K \sim \textbf{Binomial}\left(N, F(t|Z)\right),$$

hence

$$\textbf{Pr}[K = k|Z] = \binom{N}{k} F(t|Z)^k \left(1 - F(t|Z)\right)^{N-k}. \tag{10}$$

$$\binom{N}{k} p^k (1-p)^{n-k}$$

Unconditional Distribution

The unconditional distribution function of K can be derived by iterated expectation,

$$\begin{aligned} \mathbf{Pr}[K = k] &= \mathbb{E}\left(\mathbf{Pr}[K = k|Z]\right) \\ &= \int_{-\infty}^{\infty} \binom{N}{k} F(t|Z)^k \left(1 - F(t|Z)\right)^{N-k} \, d\,\Phi(Z). \end{aligned} \tag{11}$$

Note, this integral must be solved numerically.

Loss distribution

Equation (**??**) is the distribution function for the number of default, we need to translate it to portfolio loss distribution.

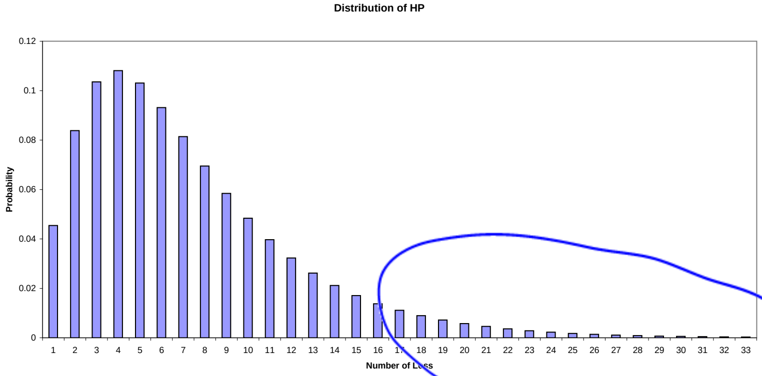
Denote $L(t)$ the percentage of loss for the reference portfolio by time t, and θ is the recovery rate. So when there is K defaults in the portfolio, the loss will be (assume unit NP):

$$\boxed{L(t)} = \frac{K(1-\theta)}{N}.$$

Thus the loss distribution function is

$$\mathbf{Pr}\left[L(t) = \frac{k(1-\theta)}{N}\right] = \mathbf{Pr}[K = k].$$

Skewed by tail.



Large Homogenous Portfolio

If the number of reference names N of the underlying portfolio becomes reasonably large (as it effectively is for a typical CDO), the distribution function for the portfolio loss can be further simplified. This allows for computation of the portfolio loss distribution without resorting to either Monte Carlo simulationnor numerical schemes to solve integral (**??**).

Fraction of default

Define the fraction of default as

$$Y = \frac{K}{N}.$$

$E(K) = Np.$

$Var(K) = Np(1-p)$

According to CLT, conditional on Z, Y is approximately normal with

$$\mathbb{E}[Y|Z] = \frac{\mathbb{E}(K|Z)}{N} = F(t|Z)$$

lim

$$\mathbf{Var}(Y|Z) = \frac{\mathbf{Var}(K|Z)}{N^2} = \frac{F(t|Z)(1-F(t|Z))}{N}. \to 0.$$

So

$$\boxed{\lim_{N\to\infty} Y = F(t|Z).}$$

Loss distribution for LHP

Define the cumulative distribution function of $Y(z)$ is G, then

$$
\begin{aligned}
G(y) &= \mathbf{Pr}[Y(z) \le y] = \mathbf{Pr}[F(t|Z) \le y] \\
&= \mathbf{Pr}\left[\Phi\left(\frac{d(t) - \sqrt{\rho}\,Z}{\sqrt{1-\rho}}\right) \le y\right] \\
&= \mathbf{Pr}\left[Z \le \frac{\sqrt{1-\rho}\,\Phi^{-1}(y) - d(t)}{\sqrt{\rho}}\right] \\
&= \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}(y) - d(t)}{\sqrt{\rho}}\right).
\end{aligned}
$$

Loss distribution for LHP

To derive ultimate loss distribution use the fact that percentage of loss equals to fraction of loss multiply by loss given default, i.e.,

$$L(t) = Y(1-\theta),$$

$Pr(L < \ell)$
$= Pr(Y(1-\theta) \le \ell)$
$= Pr(Y \le \frac{\ell}{1-\theta})$
$= G(\frac{\ell}{1-\theta})$

So the distribution function of $L(t)$ will be

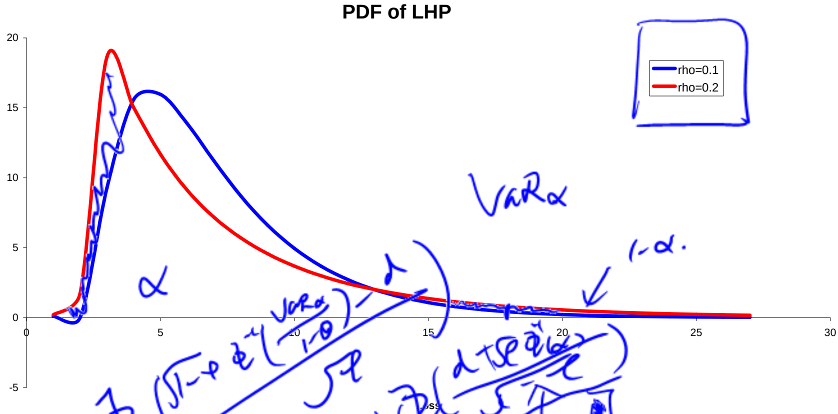
$$\mathbf{Pr}[L(t) \le l] = G\left(\frac{l}{1-\theta}\right).$$

The PDF of the loss distribution of LHP will be

$$f(y) = \sqrt{\frac{1-\rho}{\rho}} \exp\left(-\frac{1}{2\rho}\left(\sqrt{1-\rho}\Phi^{-1}(y) - d(t)\right)^2 + \frac{1}{2}\left(\Phi^{-1}(y)\right)^2\right).$$

$f_l = G'(\frac{l}{1-\theta}) = G'(y) / (1-\theta).$

$$Cap = VaR_\alpha - EL.$$



$$\alpha = \Phi\left(\frac{\sqrt{1-\rho}\,\Phi^{-1}\left(\frac{VaR_\alpha}{1-\theta}\right) - d}{\sqrt{\rho}}\right)$$

$$VaR_\alpha = (1-\theta)\,\Phi\left(\frac{d + \sqrt{\rho}\,\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)$$

Expected base tranche Loss

By using analytic distribution for LHP, one can use to calculate the expected base tranche loss(tranche with no subordination), for example the equity tranche loss function $L(t; 0, l)$.

$$\begin{aligned}
\mathbb{E}\left[L(t;0,l)\right] &= \mathbb{E}\left[L(t)\mathbf{I}\{L(t) < l\} + l\,\mathbf{I}\{L(t) \geq l\}\right] \\
&= \mathbb{E}\left[(1-\theta)F(t|Z)\mathbf{I}\{L(t) < l\}\right] + l\,\Phi(-a) \\
&= (1-\theta)\mathbb{E}\left[F(t|Z)\mathbf{I}\{Z > -a\}\right] + l\,\Phi(-a) \\
&= (1-\theta)\Phi_2\left(d(t), a, -\sqrt{\rho}\right) + l\,\Phi(-a),
\end{aligned}$$

where

$$a = \frac{\sqrt{1-\rho}\,\Phi^{-1}\left(\frac{l}{1-\theta}\right) - d(t)}{\sqrt{\rho}},$$

and $\Phi_2(x, y; r)$ is bivariate normal distribution with correlation parameter r.

Expected Loss for general tranche

$l_2 > l_1$

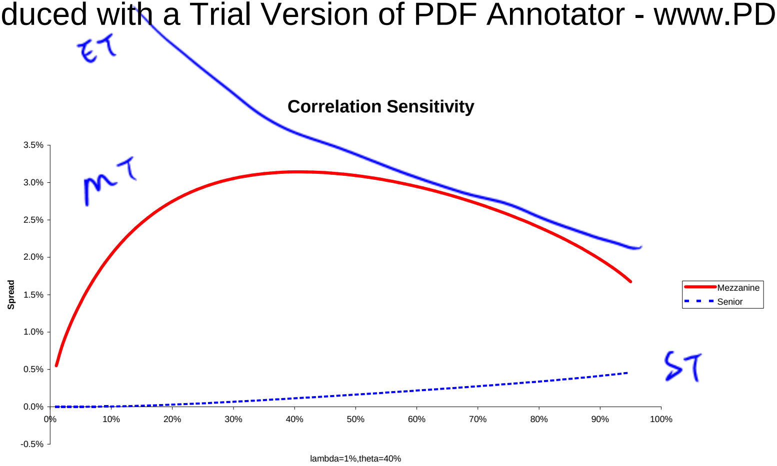
One can use the expected base tranche loss to calculate general tranche loss like this

$$\mathbb{E}\left[L(t; l_1, l_2)\right] = \mathbb{E}\left[L(t; 0, l_2)\right] - \mathbb{E}\left[L(t; 0, l_1)\right].$$

Therefore the closed for solution for tranche spread can be obtained by using this formula

$$s = \frac{\sum_{j=1}^{M} P(0, t_j)\left[\mathbb{E}[L(t_j; d, u)] - \mathbb{E}[L(t_{j-1}; d, u)]\right]}{\Delta \sum_{j=1}^{M} P(0, t_j)\left[1 - \mathbb{E}[L(t_j; d, u)]\right]}. \tag{12}$$

Note: $L(t_j; u, d)$ is redefined to be percentage of tranche loss.


ET
Correlation Sensitivity
MT
ST
Spread
3.5%
3.0%
2.5%
2.0%
1.5%
1.0%
0.5%
0.0%
-0.5%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Mezzanine
Senior
lambda=1%,theta=40%

Please take away the following important ideas

1. Joint default distribution is a key input for modeling default risk in a portfolio
2. Copula is a powerful method to model joint default
3. One factor model is an intuitively appealing and easy approach to portfolio default risk embedded
4. Both copula and factor model can be employed to price general multi-name credit derivatives
5. If it is LHP, one factor model may lead to analytical solution
6. Portfolio default risk is drive by default correlation among other key risk parameters, which would impact the valuation and risk assessment for credit derivatives

CQF: Certificate in Quantitative Finance

Valuation Adjustments ('xVAs')

Dr Jon Gregory
jon@solum-financial.com

Content

- Historical background and xVA overview
- From discounting to valuation adjustments
- CVA and DVA
- Exposure simulation
- Wrong-way risk
- FVA, MVA and KVA

The Birth of xVA

- Derivatives pricing was previously seen as pricing cashflows
- Now it is seen as being also related to
  - Credit risk
  - Funding
  - Collateral
  - Capital
  - Initial margin

$$V_{actual} = V_{base} + xVA$$

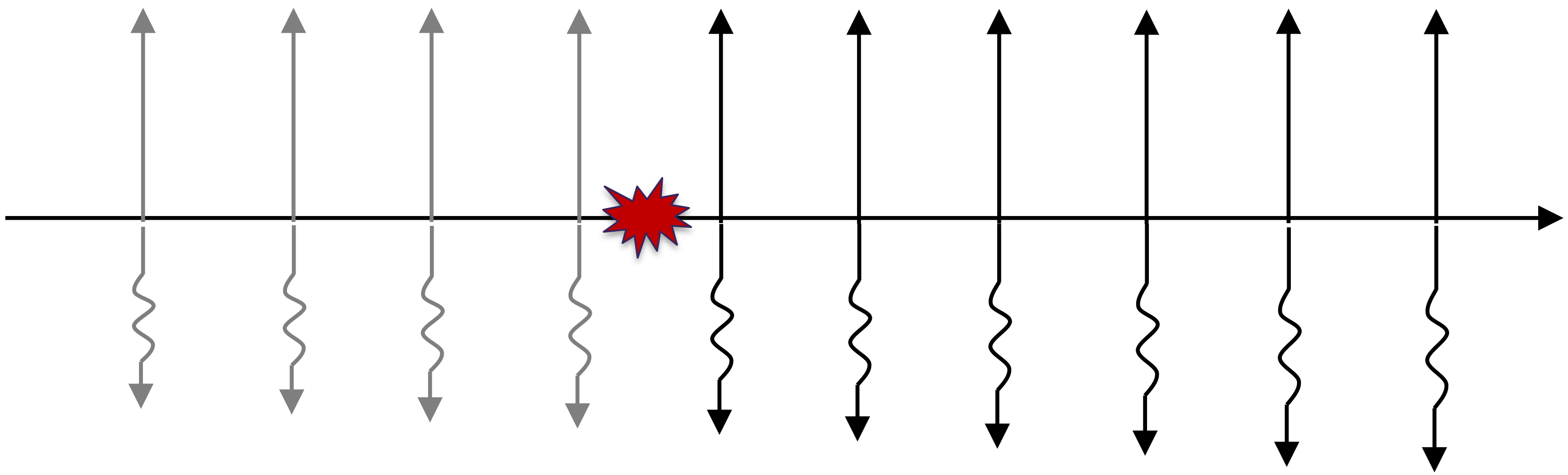
- These aspects are not mutually exclusive and often require portfolio level calculations
  - The has led to the birth of the "xVA desk" or "central resource desk"
  - This desk typically deals with most of the complexity in derivatives pricing

Content

- Historical background and xVA overview
- From discounting to valuation adjustments
- CVA and DVA
- Exposure simulation
- Wrong-way risk
- FVA, MVA and KVA

Expected Future Value (EFV)

- What is the (discounted) expected value of a transaction at a future date?

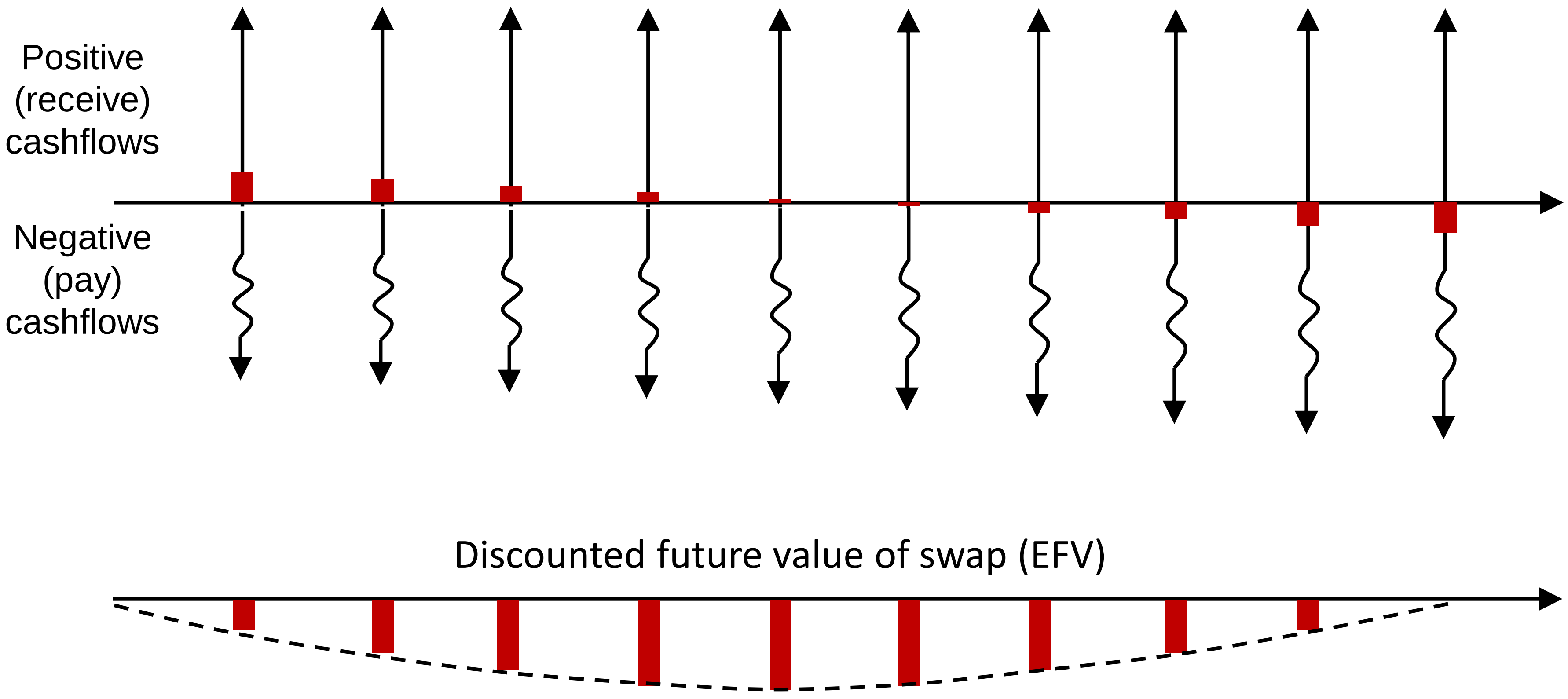


- Present value of all cashflows after that date
- This is the value of a forward starting transaction

$$EFV(t_i) = \sum_{i+1} CF_{t_i} DF_{t_i} \qquad CF_{t_i} DF_{t_i} = EFV_{t_{i-1}} - EFV_{t_i}$$

EFV Example

- Receive fixed swap, assuming the yield curve is upwards sloping



xVA and Discounting (I)

- Suppose we define xVA as the difference between discounting at two different rates (the difference between the rates is a spread s)

$$xVA = \sum_i CF_{t_i} DF_{t_i} \underbrace{\exp\left[-s_{t_i} \times t_i\right]}_{\textit{Additional discounting term}} - \sum_i CF_{t_i} DF_{t_i}$$

$$= \sum_i CF_{t_i} DF_{t_i} \left\{\exp\left[-s_{t_i} \times t_i\right] - 1\right\}$$

- Also define the EFV as the sum of all the future cashflows after a given date:

$$EFV(t_i) = \sum_{i+1} CF_{t_i} DF_{t_i}$$

$$CF_{t_i} DF_{t_i} = EFV_{t_{i-1}} - EFV_{t_i}$$

xVA and Discounting (II)

- xVA can then be written as:

$$xVA = \sum_i \left[EFV_{t_{i-1}} - EFV_{t_i}\right]\left\{\exp\left[-s_{t_i} \times t_i\right] - 1\right\}$$

- Which gives an xVA-like formula

$$= -\sum_i EFV_{t_{i-1}} \underbrace{\left\{\exp\left[-s_{t_{i-1}} \times t_{i-1}\right] - \exp\left[-s_{t_i} \times t_i\right]\right\}}_{\textit{Forward spread}}$$

- Conclusion
  - The difference is discounting assumptions can be written as an xVA formula (although this may be unnecessarily complex and the reverse is not true)
  - Situations this is relevant
    - ➢ Collateral discounting (discounting in collateral currency)
    - ➢ FVA (discounting at own cost of funding)

Discounting and ColVA - Example

- ColVA formula based on expected collateral balance (ECB = EFV)
- This example is simple due to the symmetry (default is not symmetric)

$$ColVA = -\sum_i ECB_{t-1}\left\{\exp\left[-s_{t_{i-1}} \times t_{i-1}\right] - \exp\left[-s_{t_i} \times t_i\right]\right\}$$

	Pay fixed	Receive fixed
Swap value discounted at base rate	-23,968	23,968
Swap value discounted at alternate rate	-33,401	33,401
Difference (using spread s_t)	-9,433	9,433
ColVA formula	-9,433	9,433

What's the point of a ColVA formula?

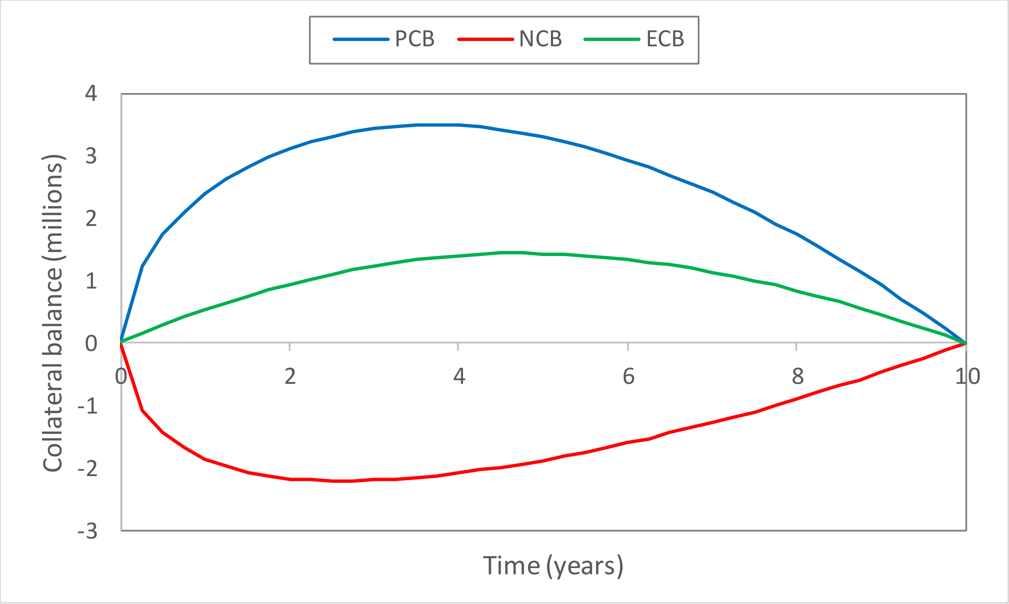
$ENE \approx NCB$

- Expand the ColVA formula into cost and benefit terms (*ECB* = *PCB* + *NCB*)

$EPE \approx PCB$

Collateral received adjustment $$ColRA = -\sum_i PCB_{t-1}\{\exp[-s_{t_{i-1}} \times t_{i-1}] - \exp[-s_{t_i} \times t_i]\}$$

Collateral posted adjustment $$ColPA = -\sum_i NCB_{t-1}\{\exp[-s_{t_{i-1}} \times t_{i-1}] - \exp[-s_{t_i} \times t_i]\}$$



	Pay fixed
Difference (as before)	-9,433
ColRA	-24,226
ColPA	14,799
ColRA + ColPA	-9,427

Content

- Historical background and xVA overview
- From discounting to valuation adjustments
- CVA and DVA
- Exposure simulation
- Wrong-way risk
- FVA, MVA and KVA

Deriving the CVA Formula (I)

- Time of default is denoted by τ
- Discounted base value (no default risk) of derivatives portfolio at time t with final maturity T: $V(\tau, T)$
- Default payoff is:

$$\delta V(\tau, T)_+ + V(\tau, T)_-$$

Recovery rate (%)

- Risky value:

$$\tilde{V}(t,T) = E_t\left[\begin{array}{c} \overbrace{I(\tau > T)}^{\text{No default before } T} V(t,T) + \\ I(\tau \le T) V(t,\tau) + \\ \underbrace{I(\tau \le T)(\delta V(\tau,T)_+ + V(\tau,T)_-)}_{\text{Default payoff}} \end{array}\right]$$

Full value if no default

Cashflows before default

Deriving the CVA Formula (II)

- Extract the risk-free value

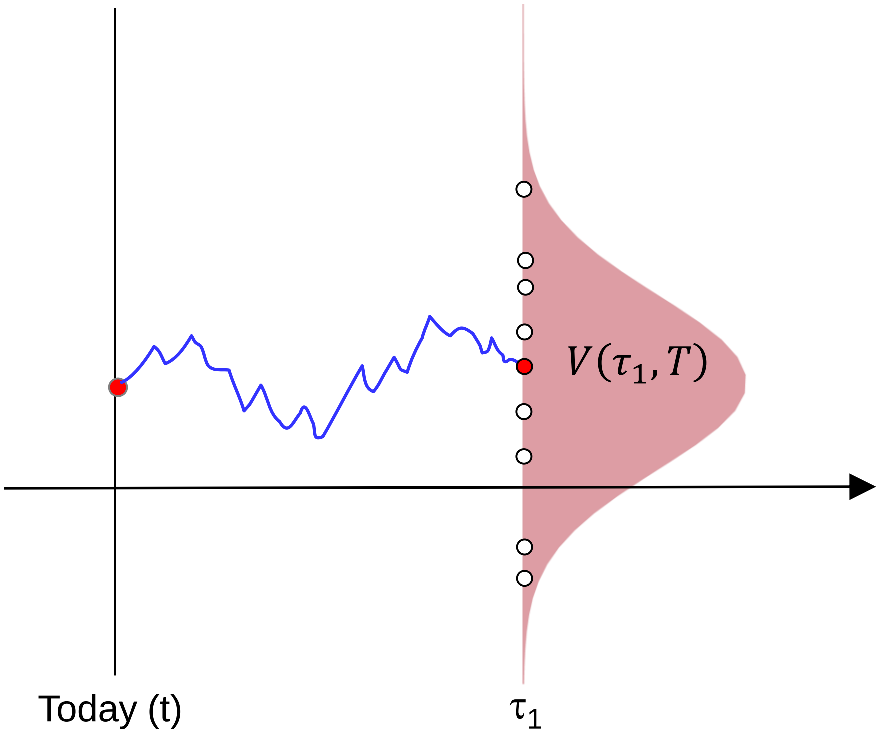
$$\tilde{V}(t,T) = E_t\begin{bmatrix} I(\tau > T)V(t,T) + \\ I(\tau \leq T)V(t,\tau) + \\ I(\tau \leq T)(\delta V(\tau,T)_+ + V(\tau,T)_-) \end{bmatrix}$$

$$\tilde{V}(t,T) = E_t\begin{bmatrix} I(\tau > T)V(t,T) + \\ I(\tau \leq T)V(t,\tau) + \\ I(\tau \leq T)(\delta V(\tau,T)_+ + V(\tau,T) - V(\tau,T)_+) \end{bmatrix}$$

$$\tilde{V}(t,T) = V(t,T) \underbrace{- E_t[I(\tau \leq T)V(\tau,T)_+(1-\delta)]}_{\text{CVA}}$$

Direct CVA Formula

$$CVA = -E_t[\underbrace{I(\tau \leq T)}_{\text{Default time } (\tau) \text{ is prior to final maturity } (T)} \times \underbrace{V(\tau, T)_+}_{\text{Exposure at default (discounted)}} \times \underbrace{(1-\delta)}_{\text{Loss given default \% (LGD)}}]$$



Possible scheme

1) Simulate time of default
2) Calculate discounted exposure at default $V(\tau, T)_+$
3) Multiply by LGD = $(1 - \delta)$
4) Repeat and average

Traditional CVA Formula

$$CVA = -E_t[\underbrace{I(\tau \leq T)}_{\text{Default}}\underbrace{(1-\delta)}_{\text{Loss given default (LGD)}}\underbrace{V(t,T)_+}_{\text{Positive exposure}}]$$

- If we assume independence (no wrong-way risk)

$$CVA = -LGD \int_t^T \underbrace{EPE(u)}_{\text{Discounted expected positive exposure}}\underbrace{dPD_C(u)}_{\text{Default probability}}$$

$$EPE(t) = E_t[V(t,T)_+]$$

Credit spread

$$\approx -E[LGD]\sum_i^T EPE(t_i)\left[\exp\left(-\left(\frac{s_{i-1}t_{i-1}}{E[LGD]}\right)-\exp\left(-\frac{s_i t_i}{E[LGD]}\right)\right)\right]$$

Example

The Firm estimates derivatives CVA using a scenario analysis to estimate the expected credit exposure across all of the Firm's positions with each counterparty, and then estimates losses as a result of a counterparty credit event. The key inputs to this methodology are (i) the expected positive exposure to each counterparty based on a simulation that assumes the current population of existing derivatives with each counterparty remains unchanged and considers contractual factors designed to mitigate the Firm's credit exposure, such as collateral and legal rights of offset; (ii) the probability of a default event occurring for each counterparty, as derived from observed or estimated CDS spreads; and (iii) estimated recovery rates implied by CDS, adjusted to consider the differences in recovery rates as a derivative creditor relative to those reflected in CDS spreads, which generally reflect senior unsecured creditor risk. As such, the Firm estimates derivatives CVA relative to the relevant benchmark interest rate.

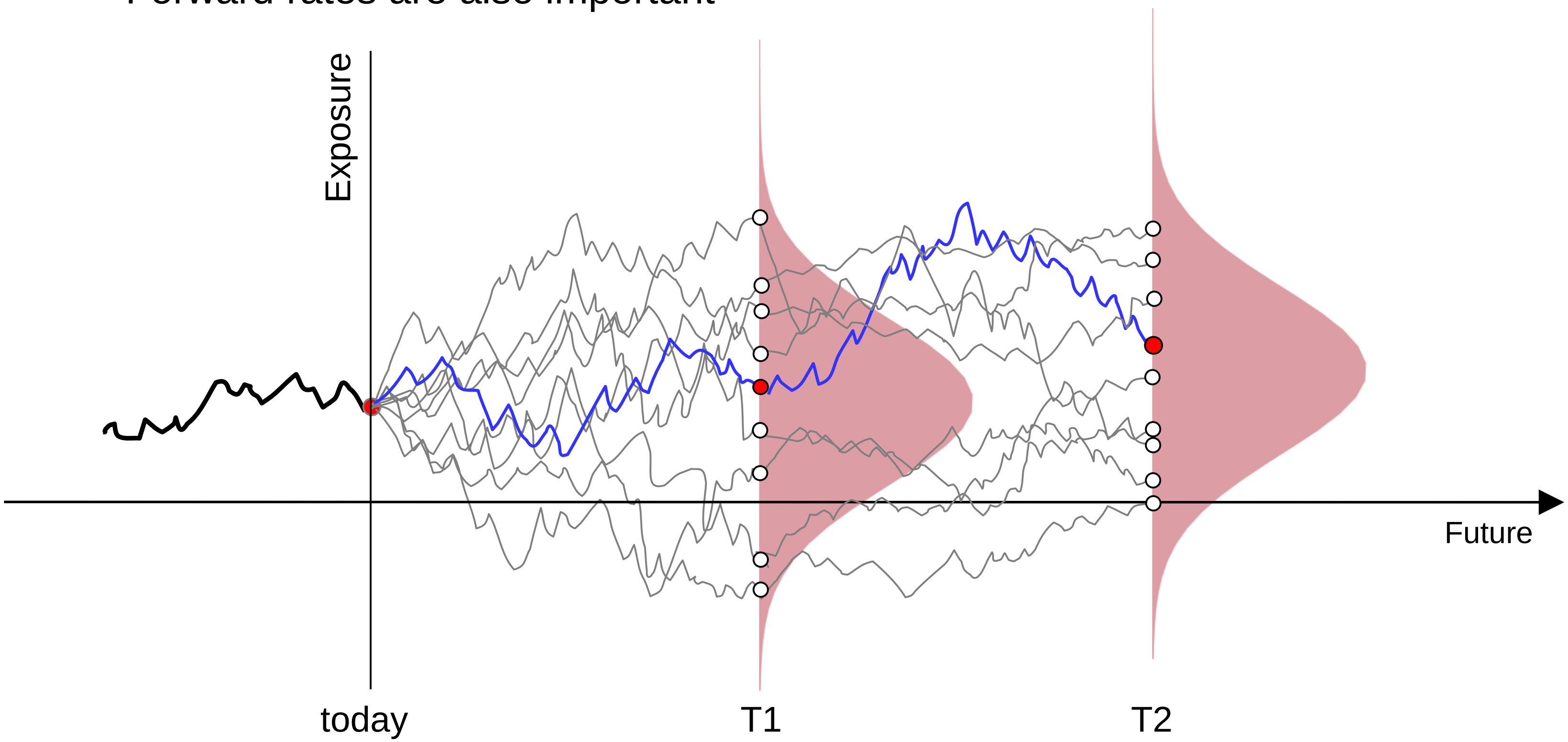
JP Morgan 2015 Annual Report

Content

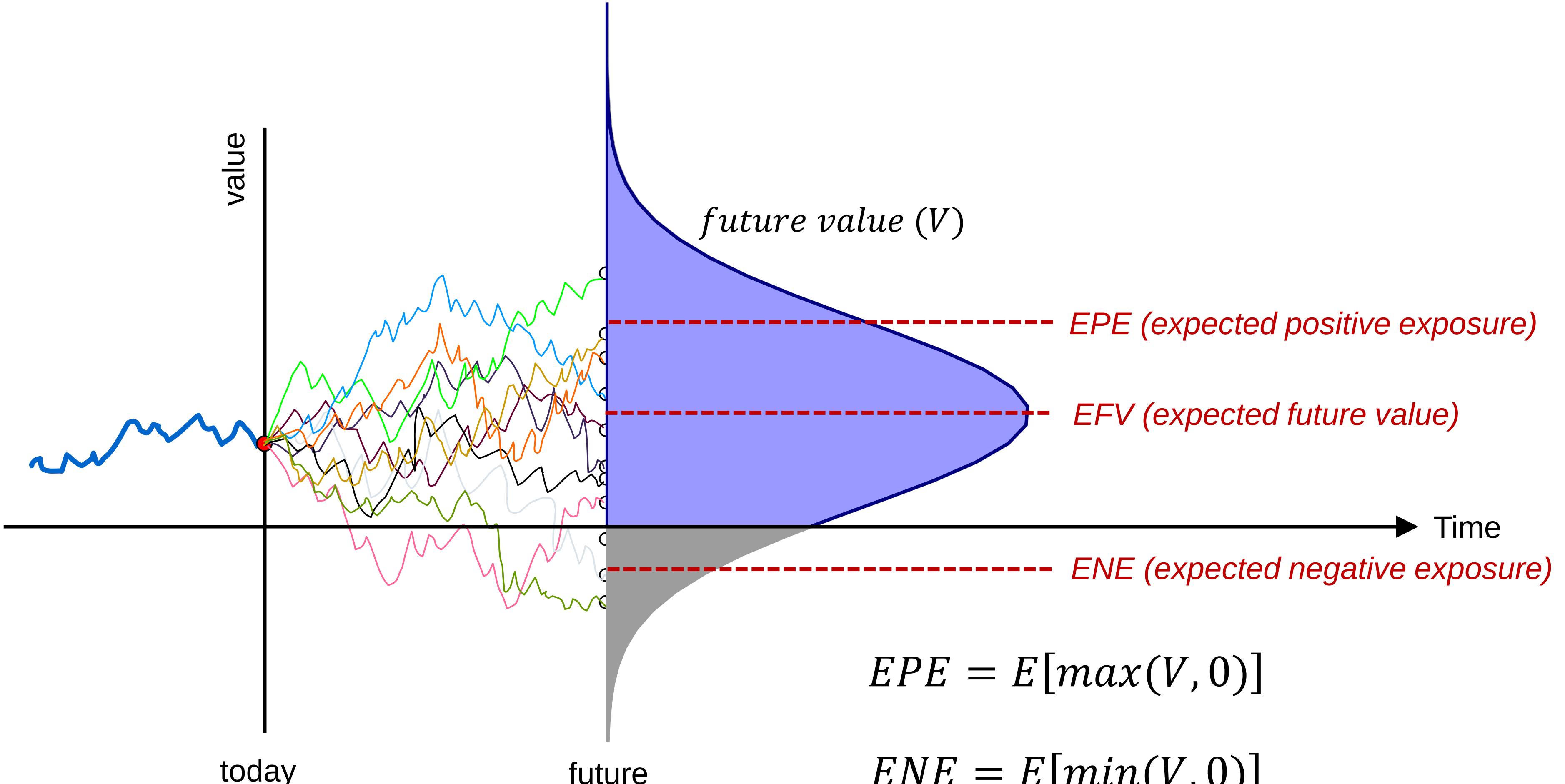
- Historical background and xVA overview
- From discounting to valuation adjustments
- CVA and DVA
- Exposure simulation
- Wrong-way risk
- FVA, MVA and KVA

Multi-Step Exposure Simulation

- Usually used to calculate quantities needed for valuation adjustments
  - Ageing is important (e.g. a 10-year swap becomes a 9-year swap in 1-year and options may get exercised)
  - Forward rates are also important



Metrics for xVA



$$EPE = E[max(V, 0)]$$

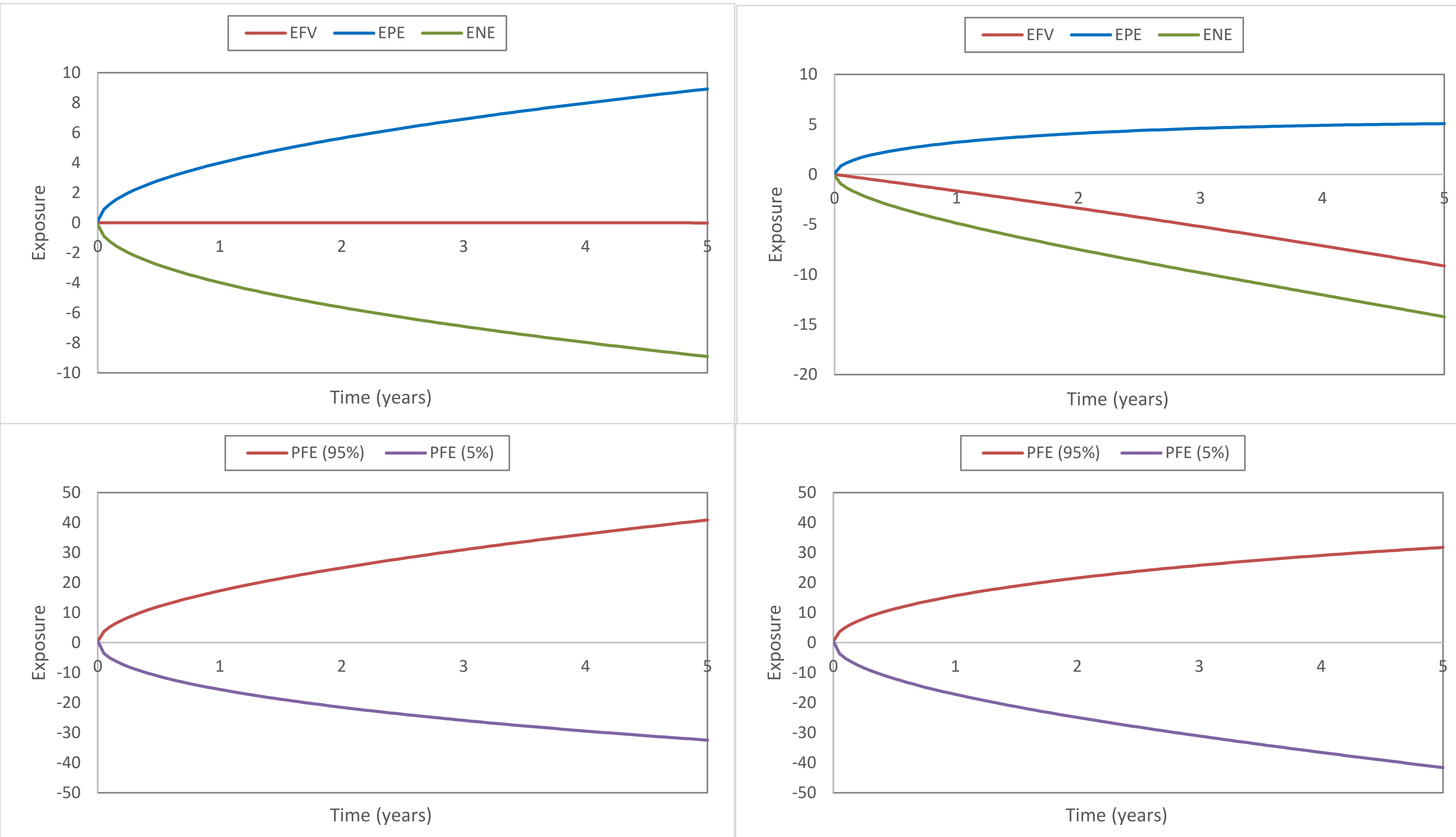
$$ENE = E[min(V, 0)]$$

$$EFV = E[V] = EPE + ENE$$

Example – FX Forward

Risk-neutral drift

Zero drift



Interest Rate Simulation – Model Choice

- Simplest model is a short-rate Gaussian model (e.g. one-factor Hull-White) which has the following benefits
  - Markovian (no memory)
  - Closed-form calculations of discount factors
  - Closed-form pricing of caps, floors and swaptions
- More advanced choice could be a ‘market model’ potentially with SABR
  - Can calibrate to entire volatility surface (with skew in the case of SABR)
  - Much less analytically tractable
- Tenor basis is also an important consideration

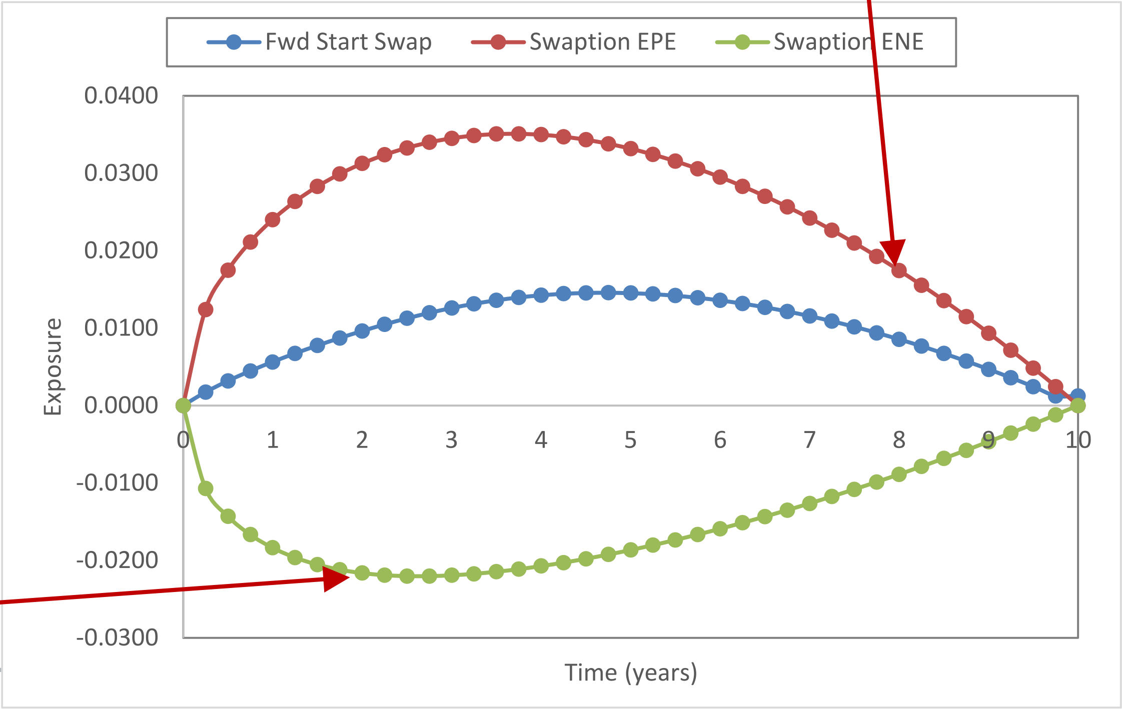
Thoughts on Volatility Calibration

- Sorenson and Bollier (1994)
  - Exposure (EPE) for a swap can be constructed from a series of European swaptions
- Intuition
  - Counterparty has the option to “cancel” the trade when they default

$$Exposure = max(value, 0)$$

Value of 8-year maturity European payer swaption on a 2-year swap

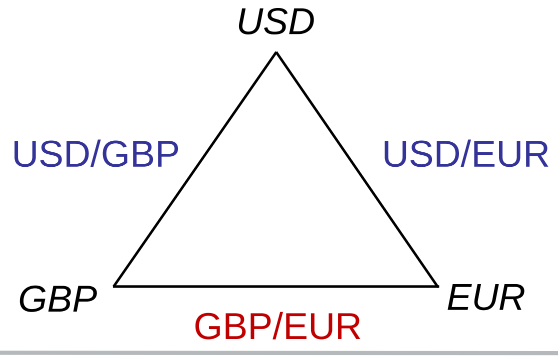
- Would imply a calibration to the ‘co-terminal swaptions’
  - 2x8, 4x6, 5x5, 6x4, 8x2 etc…..
  - However there are several problems with this?



Value of 2-year maturity European receiver swaption on a 8-year swap

FX Simulation

- A simple lognormal (Black-Scholes) model is common
- Must calibrate to FX forwards
  - Risk-free forwards from interest rate differential
  - Risky forwards from cross-currency basis swaps
- Volatility calibration
  - Piecewise constant calibrated to ATM FX options (within stochastic interest rate model setting)
  - Cannot calibrate to third pair of currency triangle or ‘cross-rates’ (changing the correlation between FX rates is one possibility)
  - For some banks, FX rates not involving their base currency may not be material



Exposure Allocation

- Two obvious ways to allocate exposure (and xVA)
- Incremental
  - Useful for pricing
  - Each trade is added sequentially

$$EPE_{inc,i} = EPE_{p+i} - EPE_p$$

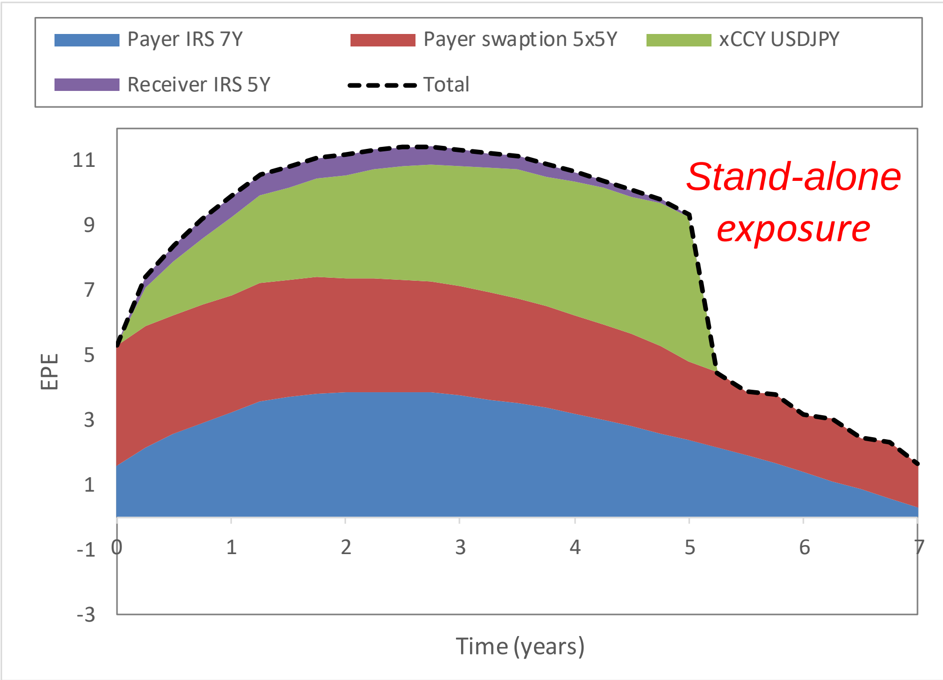
- Marginal
  - Useful for accounting / risk analysis
  - Euler allocation

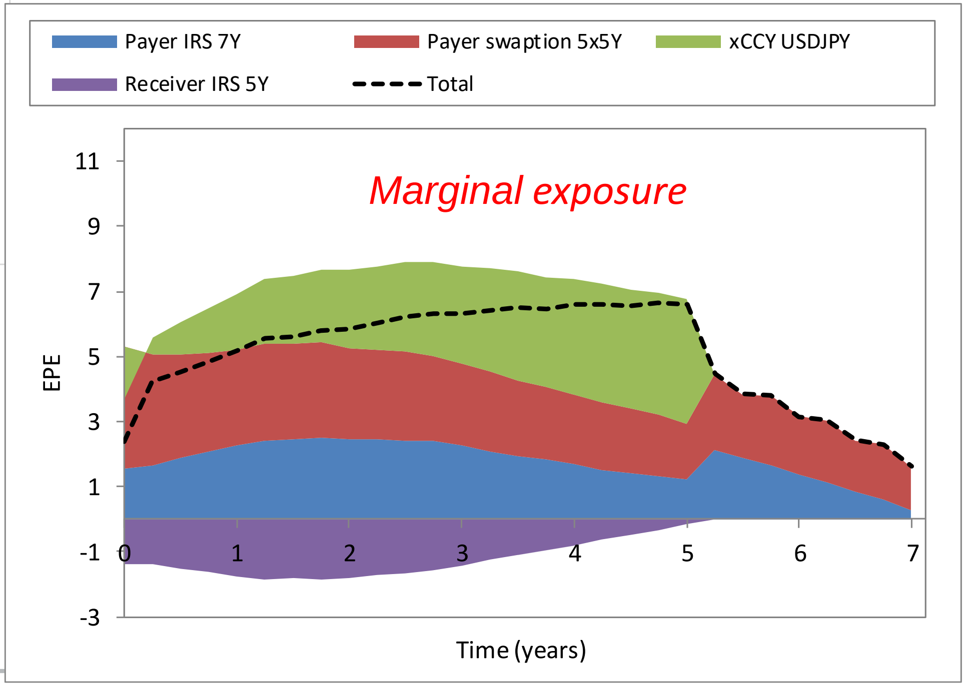
$$\underbrace{EPE_i'(s)}_{\text{Marginal exposure for trade } i} = E[V_i | V_{NS} > 0]$$

Future value of trade *i*

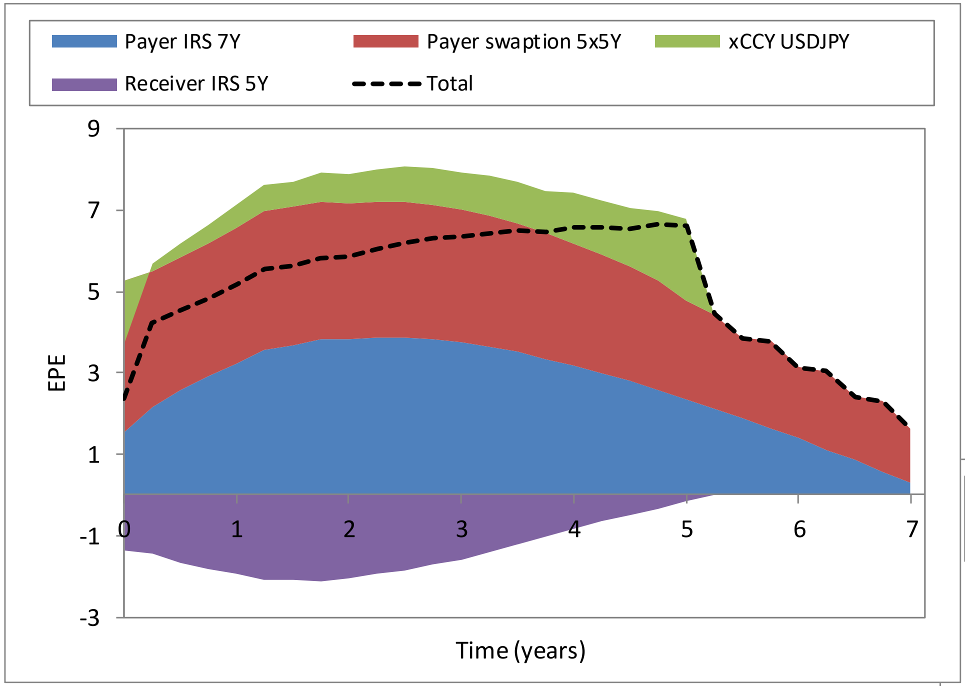
Future value of netting set

Stand-alone vs. Marginal Exposure



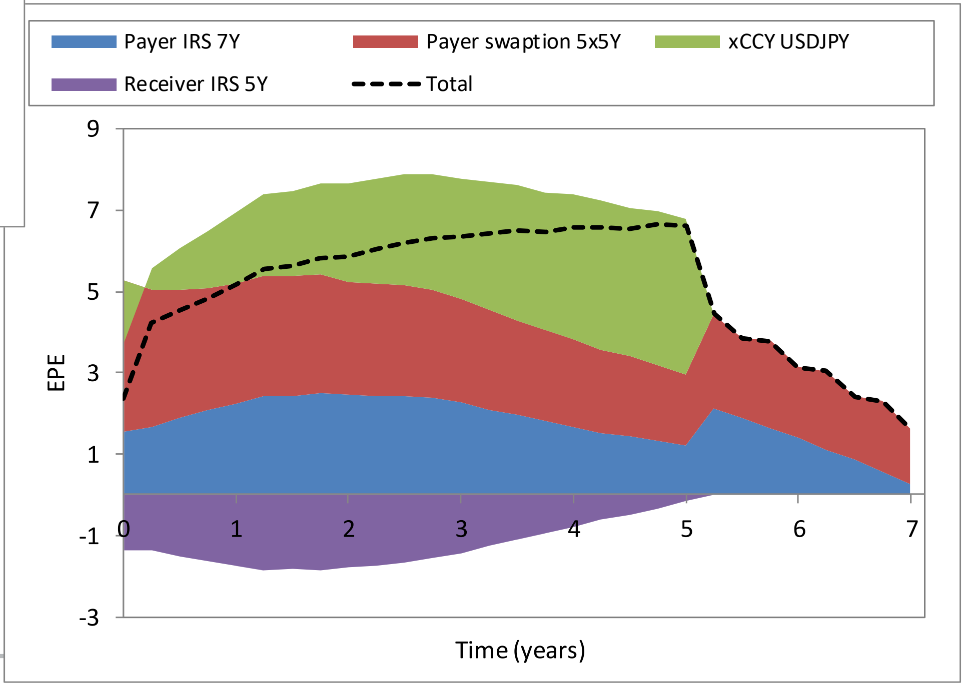


Marginal vs. Incremental Exposure



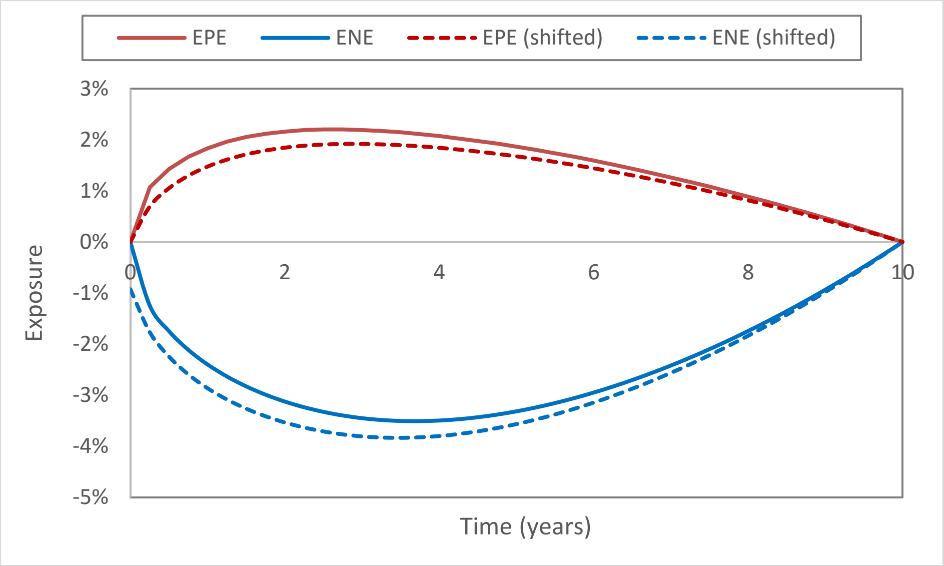
Incremental exposure (1-2-3-4)

Marginal exposure



CVA Hedging

- Interest rate swap (receive fixed) interest rate risk
  - Upwards move in rates (+10 bps)
  - Value of swap goes down (negative sensitivity to rates)
  - CVA goes down also (positive sensitivity to rates)



	6M	1Y	2Y	3Y	5Y	10Y	Total
Base value	(50)	(172)	(385)	(927)	(3,546)	(77,306)	(46,655)
CVA	27	(88)	94	(618)	(1,040)	5,613	3,988
CVA (credit spreads wider)	32	(107)	104	(754)	(1,265)	6,840	4,850

Bilateral CVA Formula

- Considering an bank's own default

Do we include survival probabilities (grey terms)?

$$BCVA = CVA + DVA =$$

$$-LGD_C \sum_j \underbrace{EPE(t_j)}_{\text{Expected positive exposure}} S_I(t_j) \underbrace{\Delta PD_C(t_{j-1}, t_j)}_{\text{Probability counterparty defaults}} \quad \text{CVA}$$

$$-LGD_I \sum_j \underbrace{ENE(t_j)}_{\text{Expected negative exposure}} S_C(t_j) \underbrace{\Delta PD_I(t_{j-1}, t_j)}_{\text{Probability } \textbf{we} \text{ default}} \quad \text{DVA}$$

The Debate Around DVA

Quant Congress USA: Ban DVA, counterparty risk quant says

Author: Laurie Carver

Source: Risk magazine | 16 Jul 2010

Categories: Credit Risk

Banks' profits boosted by DVA rule

The profits of British banks could be inflated by as much as £4bn due to a bizarre accounting rule that allows them to book a gain on the fall in the value of their debt.

Being two-faced over counterparty credit risk

*A recent trend in quantifying counterparty credit risk for over-the-counter derivatives has involved taking into account the bilateral nature of the risk so that an institution would consider their counterparty risk to be reduced in line with their own default probability. This can cause a derivatives portfolio with counterparty risk to be more valuable than the equivalent risk-free positions. In this article, **Jon Gregory** discusses the bilateral pricing of counterparty risk and presents an approach that accounts for default of both parties. He derives pricing formulas and then argues that the full implications of pricing bilateral counterparty risk must be carefully considered before it is naively applied for risk quantification and pricing purposes*

Counterparty credit risk is the risk that a counterparty in a financial contract will default prior to the expiry of the contract and fail to make future payments. Counterparty risk is taken by each party in an over-the-counter derivatives contract and is present in all asset classes, including interest rates, foreign exchange, equity derivatives, commodities and credit derivatives. Given the recent decline in credit quality and heterogeneous concentration of credit exposure, the high-profile defaults of Enron, Parmalat, Bear Stearns and Lehman Brothers, and writedowns associated with insurance purchased from monoline insurance companies, the topic of counterparty risk management remains ever-important.

A typical financial institution, while making use of risk mitigants such as collateralisation and netting, will still take a significant amount of counterparty risk, which needs to be priced and risk-managed appropriately. Over the past decade, some financial institutions have built up their capabilities for handling counterparty risk and active hedging has also become common, largely in the form of buying credit default swap (CDS) protection to mitigate large exposures (or future exposures). Some financial institutions have a dedicated unit that charges a premium to each business line and in return takes on the counterparty risk of each new trade, taking advantage of portfolio-level risk mitigants such as netting and collateralisation. Such units might operate partly on an actuarial basis, utilising the diversification benefits of the exposures, and partly on a risk-neutral basis, hedging key risks such as default and forex volatility.

A typical counterparty risk business line will have significant reserves held against some proportion of expected and unexpected losses, taking into account hedges. The recent significant increases in credit spreads, especially in the financial markets, will have increased such reserves and/or future hedging costs associated with counterparty risk. It is perhaps not surprising that many institutions, notably banks, are increasingly considering the two-sided or bilateral nature when quantifying counterparty risk. A clear advantage of doing this is that it will dampen the impact of credit spread increases by offsetting mark-to-market losses arising, for example, from increases in required reserves. However, it requires an institution to attach economic value to its own default, just as it may expect to make an economic loss when one of its counterparties defaults. While it is true a corporation does 'gain' from its own default, it might seem strange to take this into account from a pricing perspective. In this article, we will make a quantitative analysis of the pricing of counterparty risk and use this to draw conclusions about the validity of bilateral pricing.

Unilateral counterparty risk

The reader is referred to Pykhtin & Zhu (2006) for an excellent overview of measuring counterparty risk. We denote by $V(s, T)$ the value at time s of a derivatives position with a final maturity date of T. The value of the position is known with certainty at the current time $t (< s \leq T)$. We note that the analysis is general in the sense that $V(s, T)$ could indicate the value of a single derivatives position or a portfolio of netted positions[1], and could also incorporate effects such as collateralisation. In the event of default, an institution must consider the following two situations:

- $V(s, T) > 0$. In this case, since the netted trades are in the institution's favour (positive present value), it will close out the position but retrieve only a recovery value, $V(s, T)\delta_C$, with δ_C a percentage recovery fraction.
- $V(s, T) \leq 0$. In this case, since the netted trades are valued against the institution, it is still obliged to settle the outstanding amount (it does not gain from the counterparty defaulting).

[1] We note that since exposures within netted portfolios are linear then this case is suitably general

Risk February 2009

Using debt value adjustment to inflate profits

Financial results in large banks have been inflated in the third quarter due to an accounting rule called "debt value adjustment" (DVA). DVA states that banks are allowed to mark their debt to market. In other words, if their debt decreases in price on the market, this is interpreted as a decrease in liabilities and is reported as profit. In the third quarter, this rule created £10 billion in profits in the biggest U.K. banks and $12 billion in profits in the biggest U.S. banks.

Overlap Between DVA and FVA

- In January 2014, JP Morgan reported FVA for the first time
  - $1.5 billion pre-tax loss (delta around -$25 million per bp assuming a funding spread of 60 bps)

"The adjustment this quarter is largely related to uncollateralized derivatives receivables, as

- Collateralized derivatives already reflect the cost or benefit of collateral posted in valuations

- Existing DVA for liabilities already reflects credit spreads, which are a significant component of funding spreads that drive FVA"

- DVA sensitivity?
  - Q4 loss of $536 million on DVA (JPM CDS spread had tightened from 93 bps to 70 bps)
  - Delta around +$23.3 million per bp?

"P&L volatility of combined FVA/DVA going forward is expected to be lower than in the past."

- What JP Morgan calls FVA partially offsets their DVA results

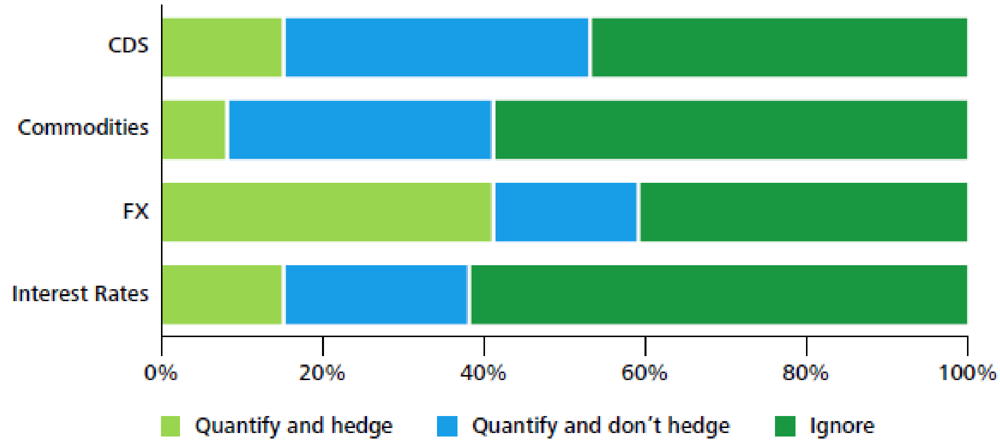
Content

- Historical background and xVA overview
- From discounting to valuation adjustments
- CVA and DVA
- Exposure simulation
- Wrong-way risk
- FVA, MVA and KVA

Wrong-Way Risk

- Standard CVA calculation assumes independence between
  - Default probability of counterparty
  - Exposure at default
- But in reality this is often wrong
  - Credit spreads and interest rates may be negatively correlated
  - FX rates may jump on major default in region
  - Credit derivatives

Figure 36. Treatment of wrong way risk



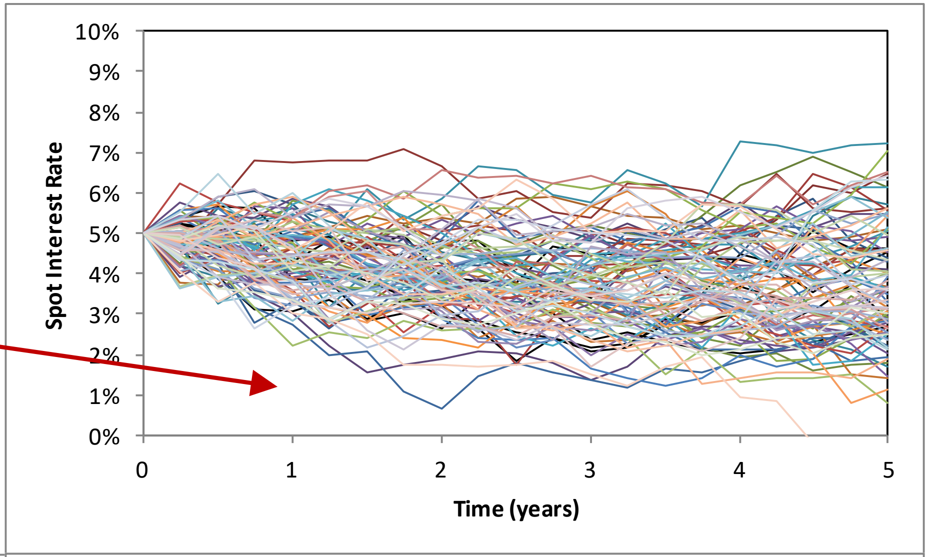
Source: Deloitte / Solum CVA Survey 2013

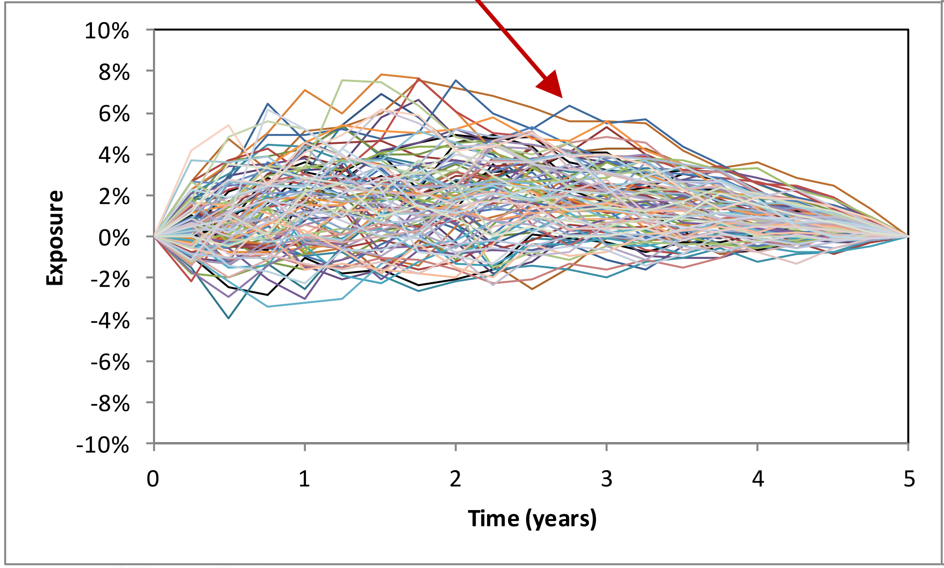
FRTB-CVA Text

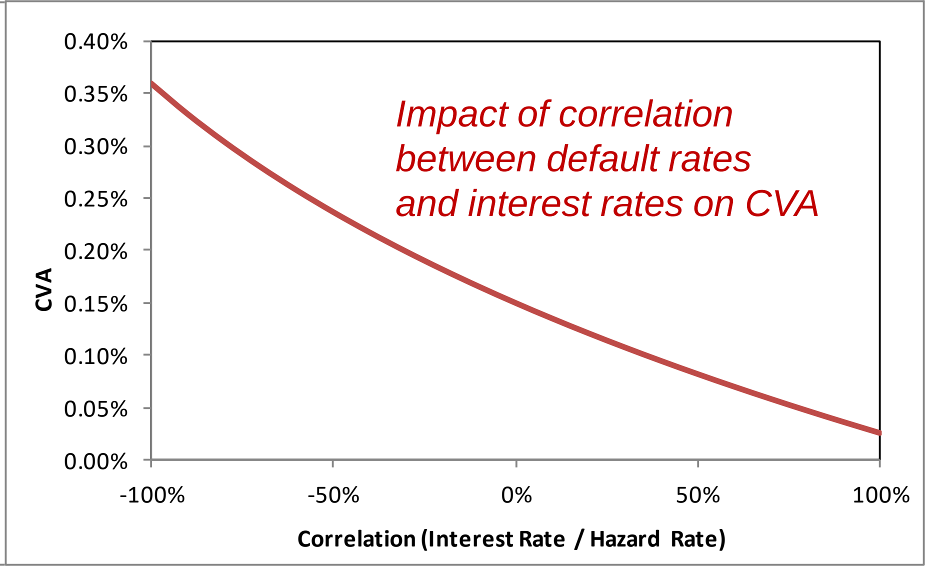
measured ES via a conservative multiplier. The proposed default level of the multiplier is [1.5]. The value of the multiplier can be increased from its default value by a bank's supervisory authority if a bank fails to capture the dependence between counterparty credit quality and exposure in its CVA calculations, or if it determines that a bank's CVA model risk is higher than its peer's.

Interest Rate Swap – Soft Approach

- Negative correlation between default rate and IR
  - Conditionally on default interest rates paths tend to decrease
  - Receiver swap exposure is higher and vice versa

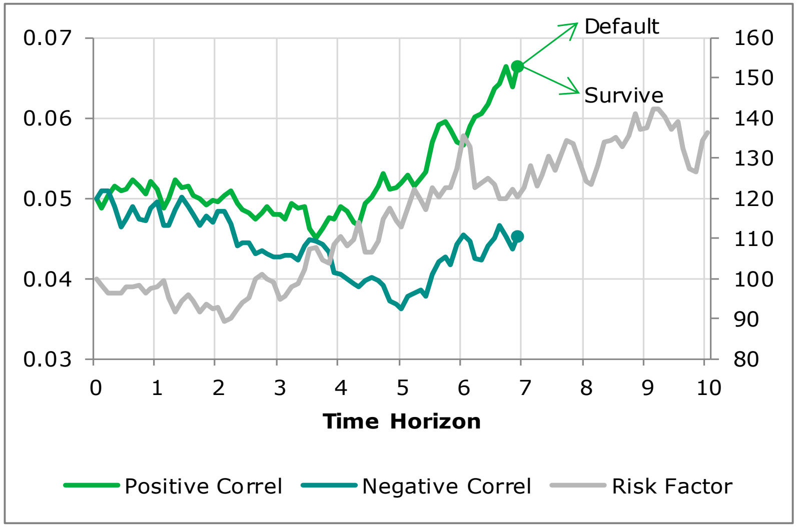




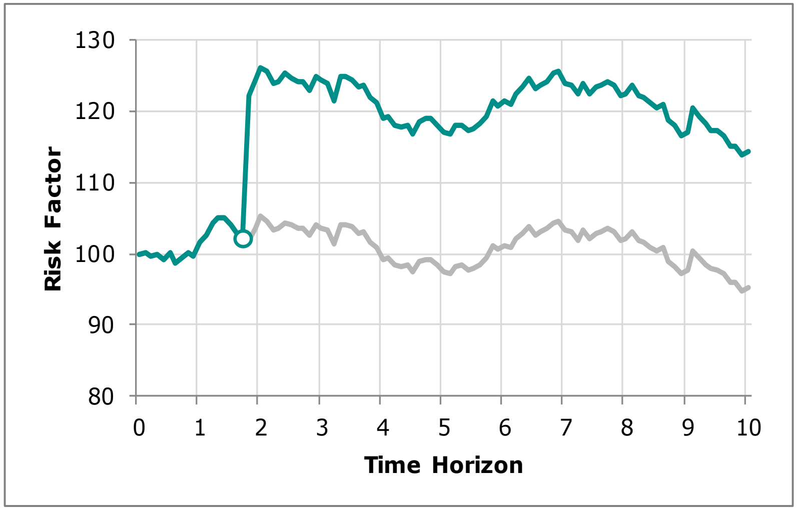


Wrong-Way Risk Case Study (II)

- Model 1
  - Soft WWR model correlating credit spread (~hazard rate) with FX process
  - Correlation estimated historically



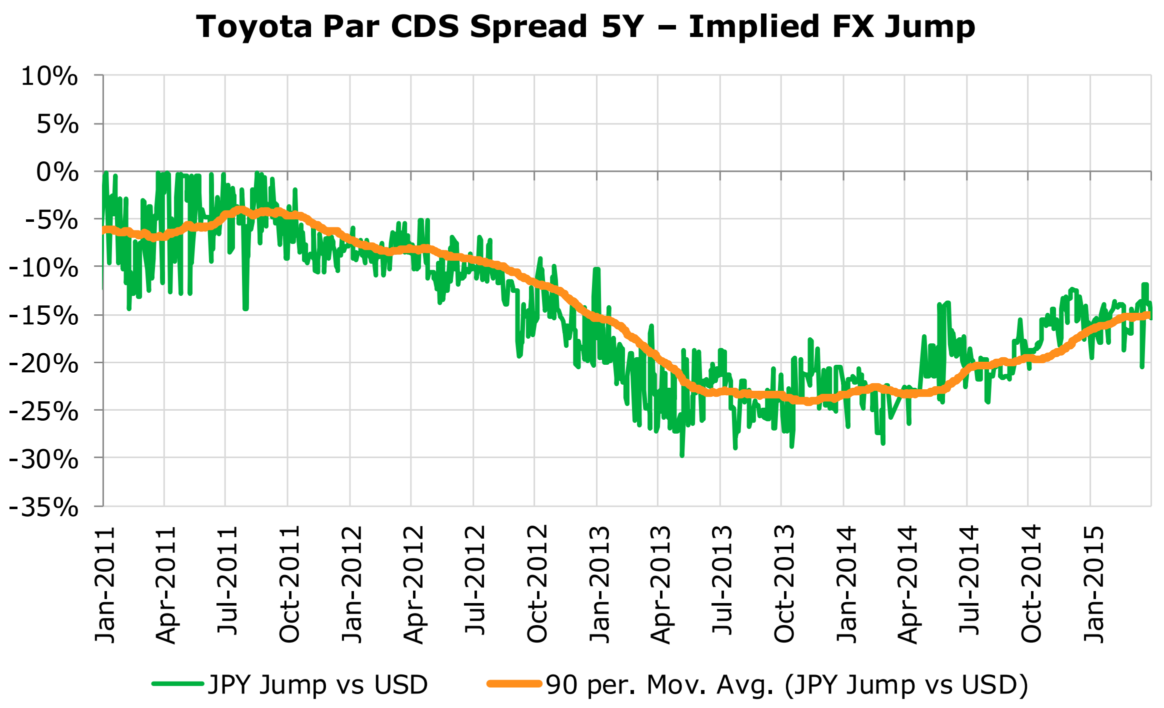
- Model 2
  - 'Hard' causal WWR model where FX rate jumps when the counterparty defaults
  - Correlation calibrated from CDS market



Source: Chung and Gregory [2019]

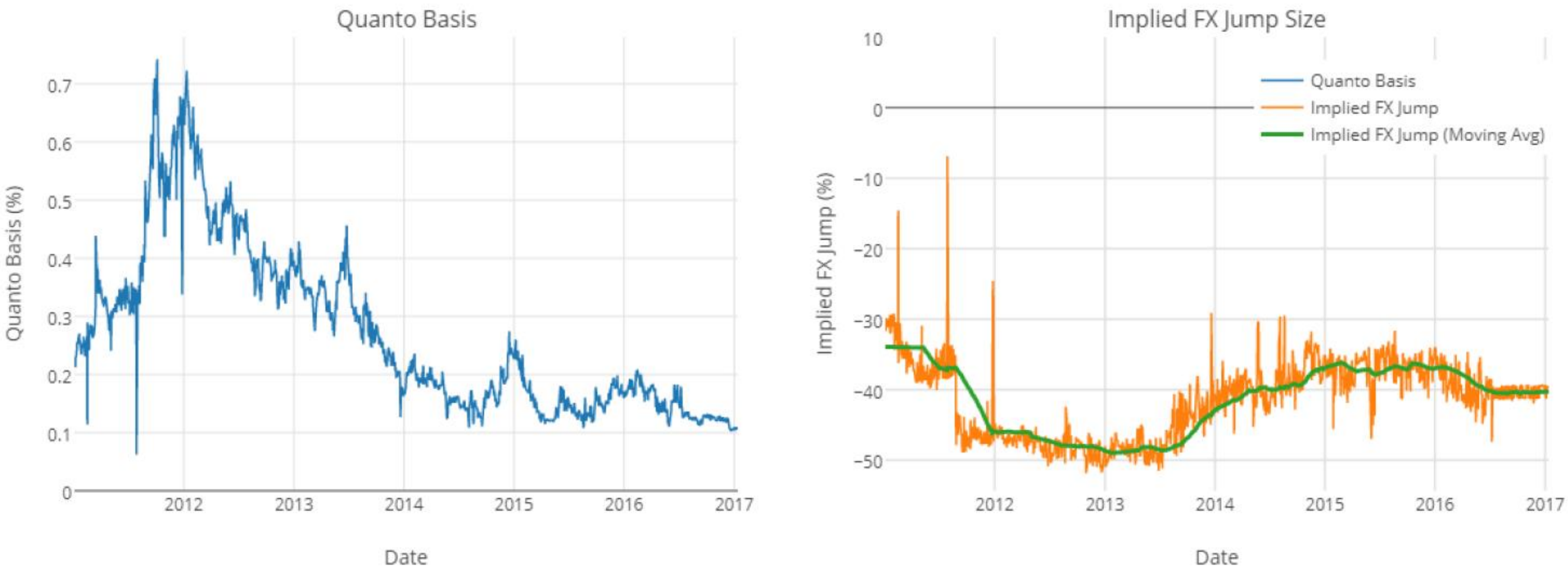
Wrong-Way Risk Case Study (III)

- Model 2 calibration
  - Implied jump can be calibrated from CDS in local current and USD
  - Similar jump size can be calibrated from the FX market



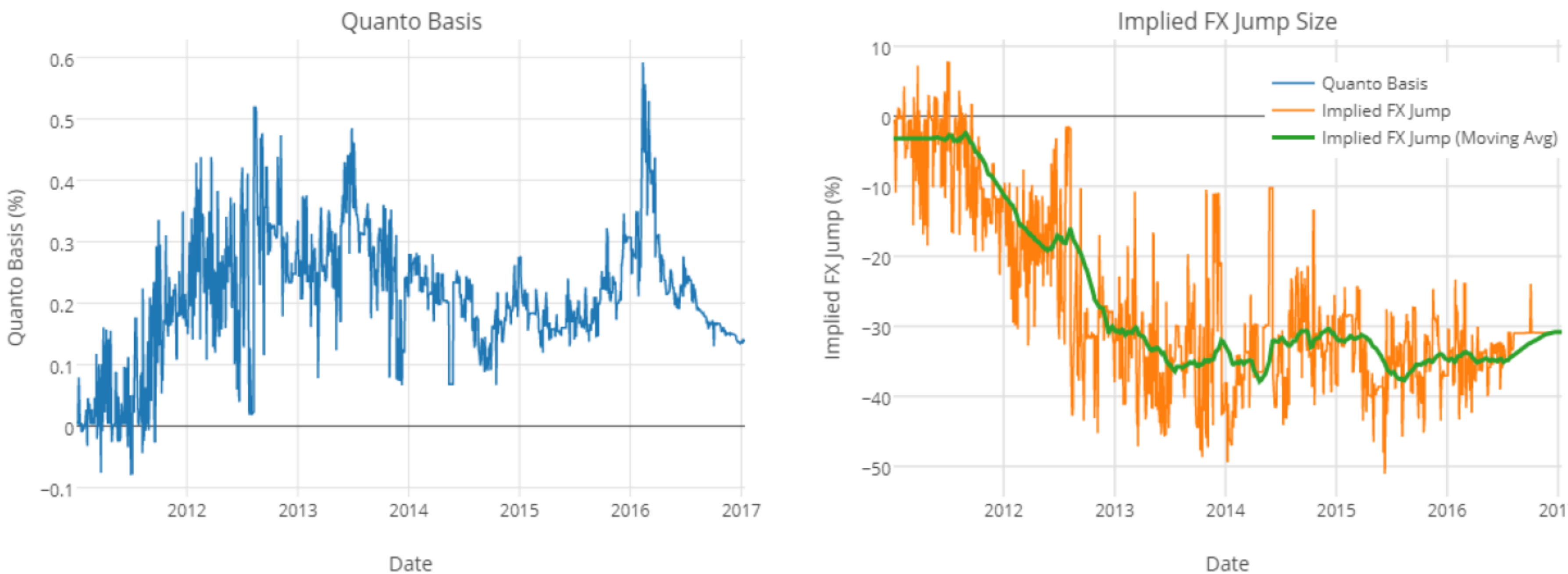
$$\mathit{Implied\ Jump} \approx \frac{\mathit{CDS\ JPY}}{\mathit{CDS\ USD}} - 1$$

Jump Calibration (I)



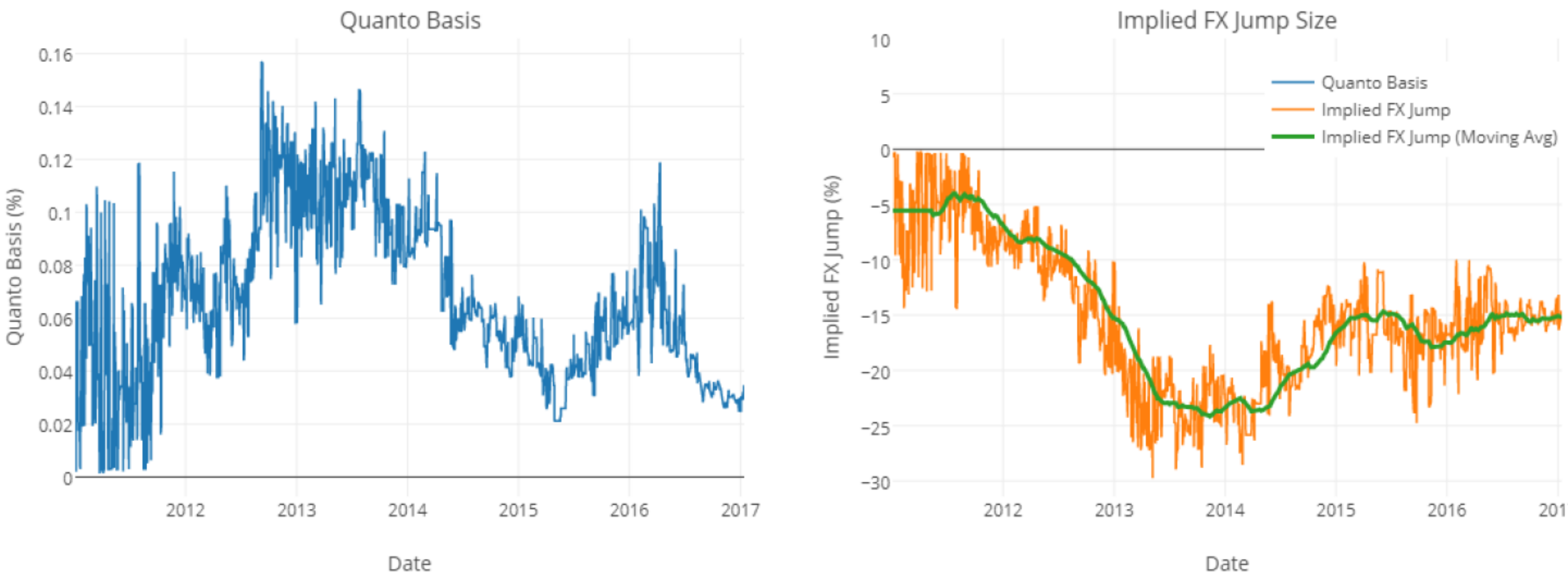
(a) Japan sovereign.

Jump Calibration (II)



(b) Bank of Tokyo Mitsubishi UFJ Ltd.

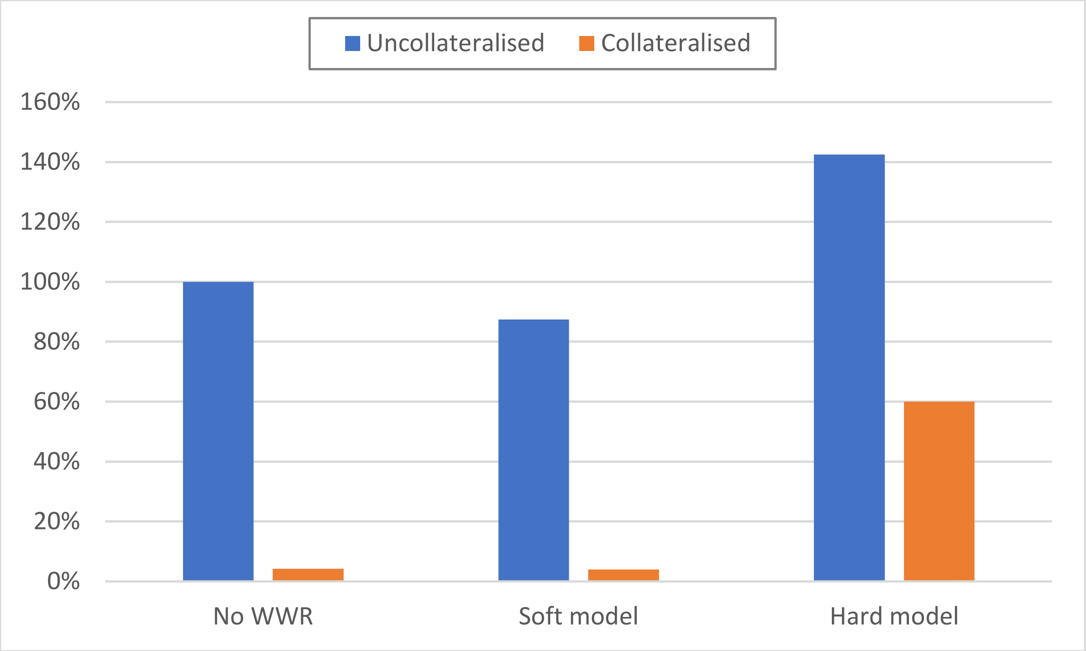
Jump Calibration (III)



(c) Toyota Motor Corp.

Results

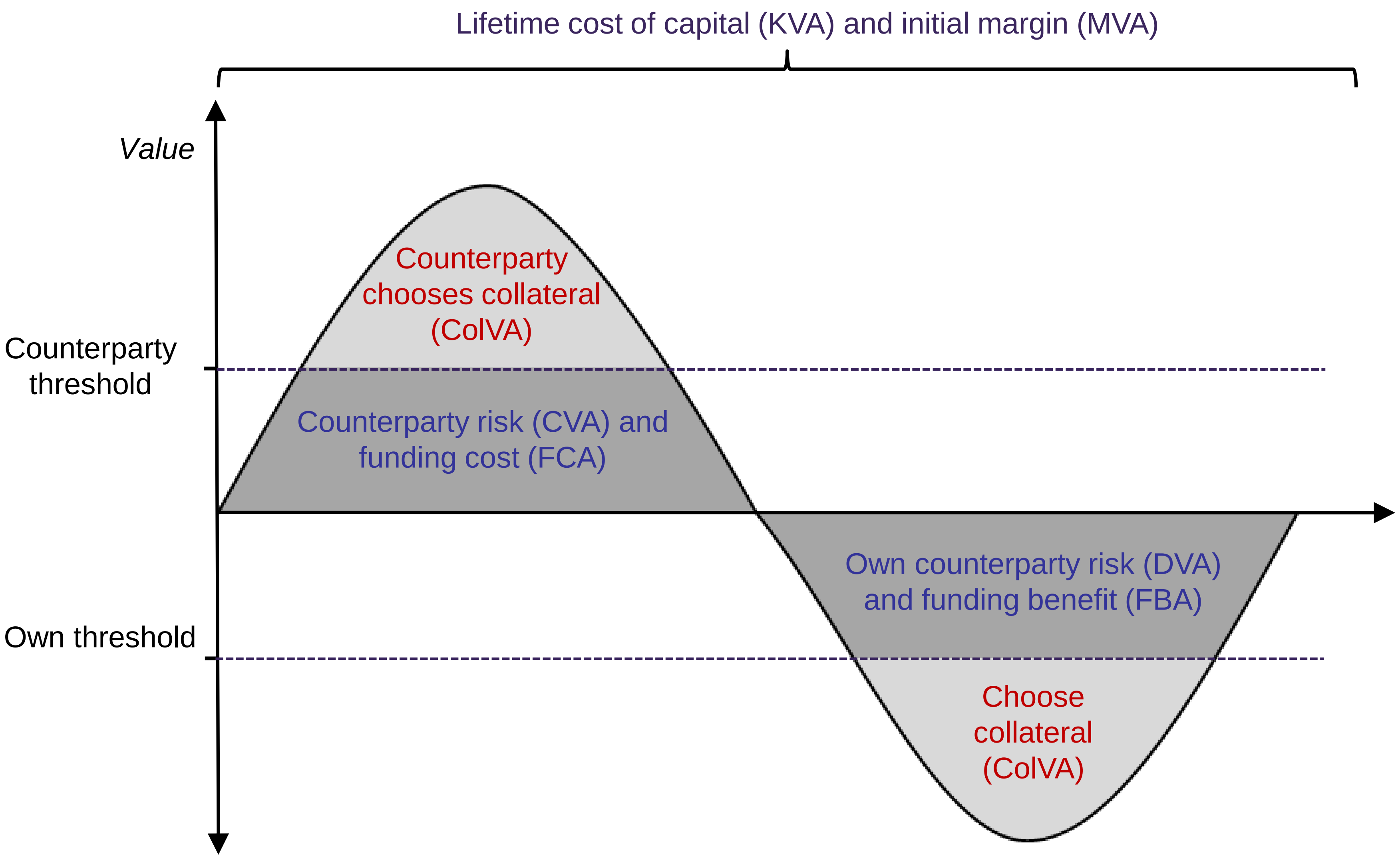
- Comparison of models for a directional portfolio
  - Soft WWR model gives lower CVA since correlation implies a weakening of JPY will be beneficial for the corporate
  - Hard WWR model gives much higher CVA since default of corporate implies devaluation of JPY
  - Soft WWR model cannot reproduce market prices



Content

- Historical background and xVA overview
- From discounting to valuation adjustments
- CVA and DVA
- Exposure simulation
- Wrong-way risk
- FVA, MVA and KVA

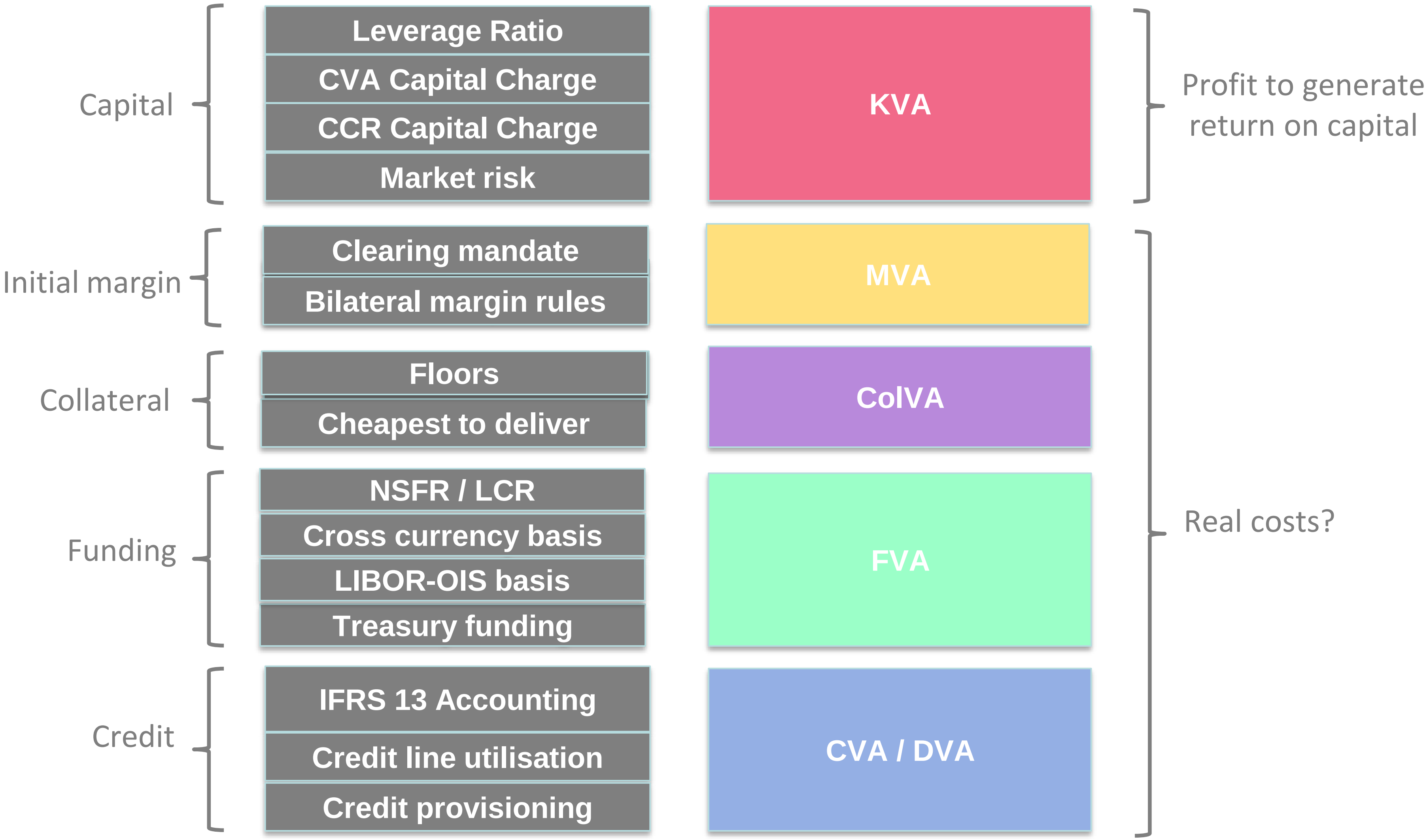
Economic Costs of Holding a Derivative Transaction



Beyond CVA - xVA

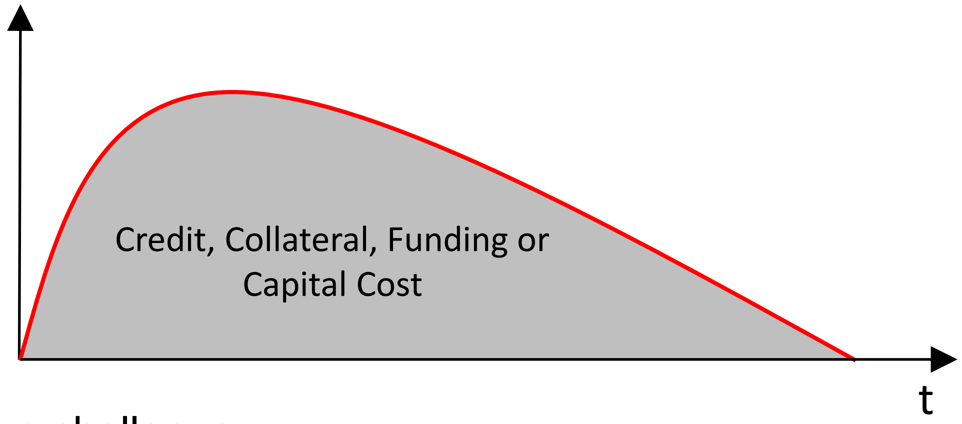
- CVA and DVA
  - Counterparty risk
- FVA (funding value adjustment)
  - Cost of funding derivatives (generally undercollateralisation)
  - Funding benefit is DVA
- ColVA
  - Adjustment for collateral effects (e.g. cheapest-to-deliver, remuneration, floors)
- MVA (margin value adjustment)
  - Cost of posting initial margin (generally overcollateralisation)
- KVA (capital value adjustment)
  - Cost of holding capital against transaction

The xVA Hierarchy



The xVA Calculation – General Comments

$$xVA = \int_0^\infty C(t)e^{-\int_0^t \beta(u)du} E[X(t)]dt$$



- xVA computation involves
  - Determination of curves, $\mathrm{C(t)}$ – qualitative challenge
  - Calculation of underlying profile, $\mathrm{X(t)}$ – quantitative challenge
- In some special cases, we are only really pricing forward contracts
  - xVA can be implemented by the correct choice of discount factor
- Discounting choices are not easy
  - Relates to questions such as close-out assumptions (e.g. do we close-out at an uncollateralised or collateralised price?)

Summary

- Virtually all derivatives have significant xVA considerations
  - Bilateral OTC uncollateralised – CVA, FVA, KVA
  - Bilateral OTC collateralised – CVA, KVA, ColVA
  - Central cleared – MVA, KVA
  - Exchange-traded – MVA, MVA
- xVA to some extent occupies the ground previously taken by exotic derivatives
- xVA represents a portfolio problem and terms are not mutually exclusive
- Represents a key area for quantitative finance going forward, for example:
  - Proxy credit spreads
  - Numerical methods for exposure simulation
  - Optimisation of xVA (e.g. initial margin compression)
  - Advanced xVA models (volatility skew, stochastic basis)
- xVA is often the core component in derivatives valuation