Certificate in Quantitative Finance

Model Implementation and Robust Estimation in Rates, Credit and Portfolio Construction

Richard Diamond

This Version: February 16, 2016

Instructions

These comprehensive notes address the issues encountered in final project implementation. Coding and model validation—that is, checking that a model provides sensible numbers—is part of independent project work.

Please print selected sections/pages of this document for your reference when working on the CQF Final Project.

Interest Rates

Factorisation and Projection

In the world of fixed income, the modelling is done for either the short rate $r(t)$ or forward rates $\bar{f}_j$, each is instantaneous and belonging a fixed tenor τ_j. SDEs model $dr(t)$ and $d\bar{f}_j$, correspondingly, diffusion is estimated from *changes* in interest rates. We measure changes as simple differences or log-differences – how they behave depends on volatility regimes often but not always conditioned on the level of interest rates (high/low). Empirical research findings are different for the US, UK and Japan's government debt.[1]

We can explore rate changes at selected 'important' tenors. Tenors are linked/most sensitive to PCA factors.[2] The histograms show: rate differences are an *iid* process with a high-peaked elliptic distribution.[3]

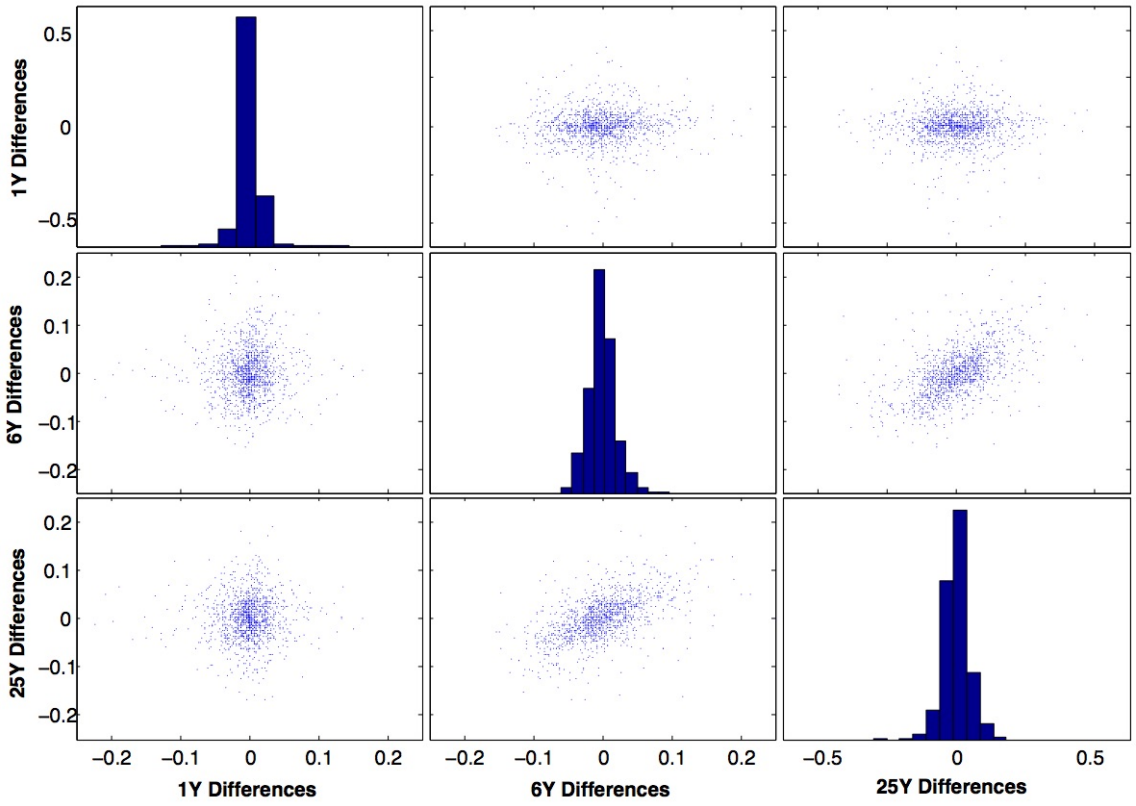

Figure 1: Observe how 1Y forward rate changes are naturally independent from changes at 6Y and 25Y tenors. The plot is not symmetric making 6Y vs. 25Y relationship to look different (illusion).

[1] If interested please see *Neither "Normal" nor "Lognormal": Modeling Interest Rates Across all Regimes* by Meucci and Loregian (2013).

[2] Identification of those important tenors was driven by PCA results on Pound Sterling data for Government Liability Curve, pre-2007. It is safe to propose these tenors are the most sensitive to PCA factors.

[3] To develop an intuition about PCA, consider these 2D 'clouds' as projections of a 3D location-dispersion ellipsoid in the form of a cigar. Eigenvectors provide the most appropriate orthogonal directions to characterise variance in this system.

Below is simple scheme for **statistical projection** (simulation) using relevant value from row j (corresponds to tenor) from principal component $\boldsymbol{e}^{(i)}$, an eigenvector that has 1 column and $j = 1..n$ tenors. Three factors were selected. Observe how it is a case of *linear factorisation*.

$$f_j(t+dt) = f_j(t) + \boldsymbol{e}_j^{(1)} X_1 + \boldsymbol{e}_j^{(2)} X_2 + \boldsymbol{e}_j^{(3)} X_3$$

For each factor, a Normal random variable X_k is scaled with standard deviation $SD\{X_k\} = \sqrt{\lambda_k}$

$$\mathbf{d\bar{f}_j(t,\tau_j)} = \cdots + \sum_{\mathbf{k=1}}^{\mathbf{3}} \sqrt{\lambda_\mathbf{k}} \mathbf{e^{(k)} dX_k} \quad \text{in the HJM SDE}$$

Forward rates can be simulated in two ways as follows (omitting index for tenor τ_j):

- HJM implements Gaussian factors – each volatility function is essentially a diffusion coefficient, which we multiply by a random Normal variable. It follows that the change in a forward rate $d\bar{f}$ is a scaled Normal variable. When simulating, we add

$$\bar{f}(t+dt) = \bar{f}(t) + d\bar{f}$$

- LMM simulation is done using Log-Normal factors, $d\ln\bar{f}(t) = \ldots dt + \sigma dX$

$$\bar{f}(t+dt) = \bar{f}(t) \times \exp\left[d\ln\bar{f}(t)\right]$$

On the first look at the HJM PCA spreadsheet a question is asked: why do we estimate from differences? We are essentially analysing shocks to interest rates (rate differences are an *iid* process) so it makes sense that a decomposed covariance matrix gives volatility functions $\bar{\nu} = \sqrt{\lambda_k}\mathbf{e}^{(k)}$ for tenors $j = 1..25$. It is volatility regimes that are of interest, not simply high rates/low rates.

Risk Premia To make the projected rates compatible with observed forward curves, **HJM model adds a drift**, also linearly. The drift represents a no arbitrage condition, the economic reasons for which are presented using the following terms (increasing for longer tenors):

1. Credit/default risk – modelled by adding a jump term to an SDE or via CVA adjustment to derivative value.
2. Term liquidity premium – early call and getting less yield as result is referred to as maturity risk; here, CVA is subject to position close-out by the counterparty.
3. Jump in inflation expectations – assets are expected to grow at the risk-free rate, which represents inflation. This factor can be related directly to $\boldsymbol{e}^{(1)}$ PCA factor (first principal component) that typically represents the level of interest rates.

Covariance Data and Attribution

Covariance matrix $\mathbf{\Sigma}$ is estimated on differences among interest rates.

- Covariance of daily differences has to be re-annualised by the factor of $\times 252$. Bank of England data has rates as percentages so, we adjust covariance by $\frac{1}{100^2}$. The total adjustment is $\times \frac{252}{10{,}000}$.
- A technical condition for eigenvalues being real and positive is for the matrix to be positive definite. Eigenvalues they represent variance for each eigendirection.
- Ensure that covariance is calculated from data without structural breaks. BoE data has missing values for long-term maturities which can be overlooked on a visual inspection – it is not uncommon for some longer term swaps/FRAs not being quoted on a particular date. A simple solution is to remove the row with missing forwards from your data.

Uncertain Attribution: if principal components have unclear attributions (i.e., they do not match their roles as parallel shift, skew, and 'twist' of the curve), then you might have to reduce the time period. For example, if you mix the data from the regime of the higher rates and their volatility (pre 2008) and low rates (2010- mid 2013) the input does not make for a good model and the eigenvectors would not behave well – there is no consistent data structure to reveal by PCA and the model becomes a 'garbage in, garbage out' story:

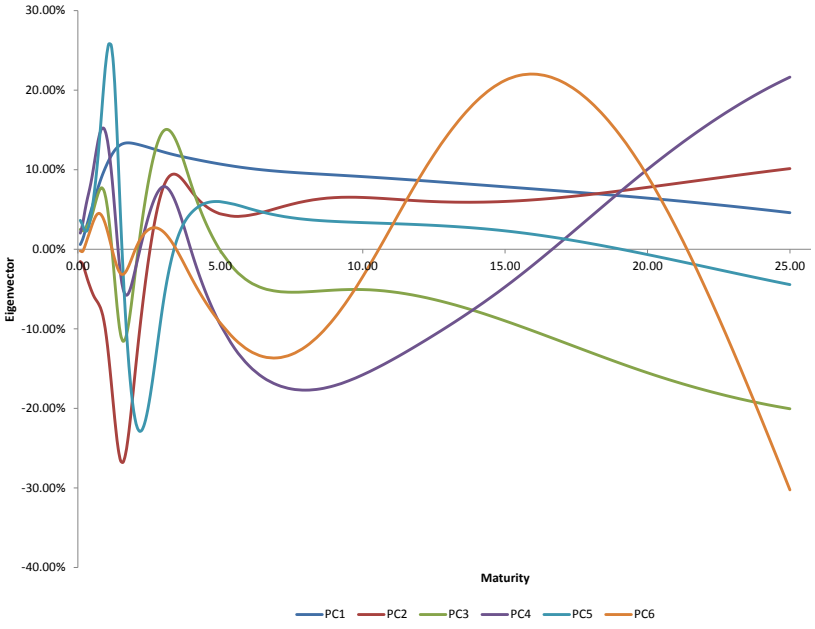

Figure 2: Principal Components with Uncertain Attribution

The illustration should help you to identify such situation: notice how the parallel shift that is typical for PC1 dominates through to the second eigendimension, while higher principal components replicate each other serving as a noise. Avoid the 'regime-switching' period of mid 2008–mid 2009 in your sample.

High vs. low rates regime: HJM lecture example has been done for high interest rate regime (pre-2008), under which it was fine to use PCA on rate differences, but since after 2008 we are in the low rates regime and it is best to conduct the PCA on log-differences in forward rates.

If one needs 10 factors to cover the reasonable $> 95\%$ movement of the curve, it is a signal that eigenvectors reproduce noise and PCA is non-informative. Here is a quick overview of recipes to deal with the situation:

- reduce your sample size to two years of stable curve regime;
- apply PCA to log differences in rates (research suggests this is suitable for the current low rates regime);
- work with weekly data (weekly changes would have less noise).

Fitting Volatility based on Principal Components

Musiela parameterisation HJM SDE volatility functions are calculated as $\overline{\nu}_i(t,\tau) = \sqrt{\lambda_i}\,\mathbf{e^{(i)}}$ and fitted by a cubic spline as implemented in *HJM_MC.xls* using Excel regression functions.

$$\overline{\nu}_i(t,\tau) = \beta_0 + \beta_1\tau + \beta_2\tau^2 + \beta_3\tau^3$$

Because volatility functions represent **factors**, the global fitting by a polynomial spline is more likely to remain robust (factors can change when regime changes but that is less frequent than yield curve fluctuations). The alternatives to polynomial fitting are as follows:

- Multiple splines: fitting parts of the curve with different splines (increases complexity).
- A logarithmic function $a + b\ln\tau$, particularly for the first component.
- An exponential function of the form $(a+b\tau)e^{c\tau}+d$, which is used in LMM to parametrise the 'skeleton term structure' of FRA volatility $\sigma_i(t)$ with $\tau = t_i - t$.

A general criterion is that a good fit crosses axe X about the same points as the fitted function.

Before adding more terms $O(\tau^4)$ to improve the fit please analyse if you can give a clear attribution to principal components (see above). Otherwise, including the higher powers will increase in-sample accuracy of fit but represent **an overfitting**.

HJM Model-Specific Issues

Overfitting and 'being stuck' at high or low levels of volatility (memory phenomenon) can be attributed to **the non-Markov nature of HJM**.

In a Markov process, only a present state determines the possible future (albeit random) state. We demonstrated that under the HJM framework, SDE for a spot rate dr has drift that depends on the history of stochastic increments dX (Itô integrals).[4] Econometricians would say that $r(t)$ is an integrated process with memory.[5] HJM model makes the evolution of the spot rates non-Markov, so we cannot obtain apply a method like the binomial tree to obtain a solution for $r(t) = f(t,t)$ or in fact, any $f(t,T)$.

HJM is Gaussian Model meaning that it can generate **negative interest rates**.

- If volatility is high then the rate evolution path can jump into negative territory from a relatively high level. This can happen for both, short and long term maturities (eg, 4Y and 10Y forwards were problematic in some Brasil rates simulations). To discriminate against unrealistic states of the curve, Monte Carlo methodology offers *rejection sampling*.
- If volatility is low, simulation paths get 'stuck' near zero and do not get out fast enough.

Apart from leaving negative rates 'as is' in a learning project, modelling options are very limited: adding a small positive floor, e.g., 0.001, as rates approach zero introduces a bias, while excluding undesired MC simulations presents continuity problem (affects convergence).

Validation – HJM is not a market model

Achieving price convergence (across multiple simulations) is sufficient.

- HJM-simulated bond prices can be compared to actual bonds, historic data on the spot curve and other models.
  As a mathematical exercise, price the same bonds using known solutions for Ho&Lee or Vasicek models – that would require calibrated values for the volatility of volatility $\eta(t)$ and mean-reversion speed $\gamma(t)$ best sourced from research papers.
- Caps and floors values can be converted into implied volatility using Black formula.

Because HJM is not a market model it cannot be 'validated' against prices in the strict sense. While it is possible to compare quantities, such as market caplets (in implied vol. terms) and HJM-priced caplets, the smarter approach would be to explosure whether the HJM model signals that the market is offering under- or over-valued claims.

[4] Slide 30 of HJM Lecture.

[5] An integrated process has unit root and allows 'an infinite' diffusion. For HJM model it means, typically generating higher volatilities than can be expected for interest rates. Using too many factors (for better fit) can amplify this problem.

Curve Interpolation. Instantaneous rates

1. Interpolation is required because usually the input curve data is provided in larger increments than necessary, for example, curve is in $6M$ increments while contracts to price (caplets) pay every $3M$.

2. Interpolating over the yield curve (forward curve) has its own specifics, such as avoiding negative forward rates, preserving convexity and concavity monotonically, and ideally, providing continuity by ensuring the derivative exists. Interpolating other curves, such as volatility term structures, might not have these requirements.

3. The third problem is about **the instantaneous forward rates** (IFR) $\bar{f}_j$ as an input to HJM calibration. BoE and ECB make effort to present forward curve data in terms of instantaneous quantities that belong only to given tenor j.

   The relationship between ZCB and IFR involves a continuous derivative *wrt* tenor time. Compared to simple forward rate observed at time t $F(t; T_i, T_{t+1})$, the instantaneous rate has $\Delta \to 0$

   $$\begin{aligned} f(T_i, T_i + \Delta) &= -\frac{1}{Z(T_i, T_i + \Delta)} \frac{\partial Z(T_i, T_i + \Delta)}{\partial T} \\ &= \frac{\partial}{\partial T} \log Z(T_i, T_i + \Delta) \end{aligned}$$

   When working with the IFR, an assumption must be made whether the instantaneous rate at T_i is made 'to belong' to the whole period $[T_i, T_{i+1}]$ or the half of it (see piecewise methods below). The smaller the period the less important the assumption is.

Linear Interpolation (here on forward rates)

For the purpose of pricing caps with 2Y-5Y maturity – via a set of forward-starting caplets – we are interested in relatively monotonic sections of the curve that, at times, could be almost linear.

For rate at any tenor τ located as $\tau_i < \tau < \tau_{i+1}$ (not just a midpoint), the linear interpolation as provided in Hagan & West (2005, page 7) is

$$\bar{f}(\tau) = \underbrace{\frac{\tau - \tau_i}{\tau_{i+1} - \tau_i}} \bar{f}(\tau_{i+1}) + \underbrace{\frac{\tau_{i+1} - \tau}{\tau_{i+1} - \tau_i}} \bar{f}(\tau_i)$$

$\bar{f}(\tau)$ refers to an instantaneous forward rate at tenor τ (bar comes from Musiela notation) under the necessary assumption of the rate being piecewise constant. The formula does a time-weighted averaging across rates; the seemingly inverse position of $\bar{f}(\tau_{i+1}), \bar{f}(\tau_i)$ terms ensures that the higher weight is given to 'the closest' rate in terms of a tenor (check computationally).

Some sources distinguish between piecewise constant and piecewise linear; the latter reduces linear relationship to an averaging across rates $\frac{\bar{f}(\tau_i) + \bar{f}(\tau_{i+1})}{2}$. That assumption is only good for interpolating the exact mid-point τ as the sum of weights is equal to one.

Method 3. Monotone convex spline [DRAFT]

The scheme is as follows:

$$r \quad \rightarrow \quad f_i^d \quad \rightarrow f_i$$

where f_i^d are discrete forwards obtained as $f_i^d = \dfrac{r_{i+1}\tau_{i+1} - r_i\tau_i}{\tau_{i+1} - \tau_i}$ and belonging to the entire interval $[\tau_{i+1}, \tau_i]$.

We calculate instantaneous forward rate f_i as being on the straight line that joins the adjacent midpoints f_i^d, f_{i+1}^d.

You can always create more yield curve points for any increment of τ (less then $d\tau = 0.5$ of a year). An input rate f_i^d belongs to the entire interval $[\tau_i, \tau_{i+1}]$ rather than point τ_i only.

$$\begin{aligned} f_i &= \frac{\tau_i - \tau_{i-1}}{\tau_{i+1} - \tau_{i-1}} f_{i+1}^d + \frac{\tau_{i+1} - \tau_i}{\tau_{i+1} - \tau_{i-1}} f_i^d \qquad \text{for } i = 1, 2, \ldots, n-1 \\ f_i^d &= \frac{1}{\tau_i - \tau_{i-1}} \int_{\tau_{i-1}}^{\tau_i} f(s)ds \end{aligned}$$

This means linearly interpolating for the mid-point interval. See Hagan and West, Section 7 (page 19) for method details and its development to suit the forward curve.

When interpolating over the curve or discount factors, the question often considered is "What does interpolation mean in terms of $r(t)$?" The answer to the setup of discrete forward rates input is given in Section 7.7 (p. 28) and is as follows:

$$\begin{aligned} r(\tau)\tau = \int_0^\tau \bar{f}(s)ds &= \int_0^{\tau_{i-1}} \bar{f}(s)ds + \int_{\tau_{i-1}}^{\tau} \bar{f}(s)ds \\ &= r_{i-1}\tau_{i-1} + f_i^d(\tau - \tau_{i-1}) + I_\tau \end{aligned}$$

where f_i^d is 'a simple average' forward rate over the period obtained by integration.

The main choice in interpolation methods is between the global approaches[6] and monotone-preserving spline in which the interpolatory values are determined by local behaviour.

You need to interpolate the forward curve by a suitable method to obtain Forward LIBOR for 3M caplets (future-starting).

[6]For HJM volatility functions, we used simple cubic spline as one global function giving the entire curve. Some local behaviour was invariably ignored by that function.

The Forward Curve: an analytical view

This section stems from reviews Lesniewski (2008) and Glasserman (2003), available online, identifying certain critical derviations and linkages between the formulae.

Let's start with a classic **zero-coupon bond** and its pricing under Monte Carlo from simulated evolutions of short rate $r(t)$[7]

$$Z(t,T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T r(s)ds}\right]$$

In terms of the forward curve, using the steps $r(t) = f(t,t) = \bar{f}(0,\tau)$ for today's curve $t = 0$

$$\begin{aligned} Z(0,T_i) &= \exp\left(-\int_0^{T_i} \bar{f}(0,\tau)d\tau\right) * \qquad \text{integrating over curve} \\ &\quad \text{discretised as} \\ &= \exp\left(-\sum_{j=0}^{i-1} \bar{f}(0,T_j)(T_{j+1}-T_j)\right) \qquad F(T_j) = f_j \text{ applies to } [T_j, T_{j+1}] \end{aligned}$$

Indexes i, j – both are applied to tenor time from 0 to T_i (curve data in row). When index i is used then the interval $\tau = T_{i+1} - T_i$ is large (eg, $6M$, $1Y$, $2Y$) and so, index j is used for rates inside of that interval $T_i < T_j < \ldots < T_{i+1}$.

LIBOR model gives the following relationship between LIBOR and ZCB. We will be interested in Forward LIBOR, so indexing changed from $Z(0,T_i)$ to $Z(T_i,T_{i+1})$.

$$\begin{aligned} Z(T_i,T_{i+1}) &= \frac{1}{1+\tau_i f_i} \\ (1+\tau_i f_i) &= \frac{1}{Z(T_i,T_{i+1})} \qquad \Longrightarrow \qquad f_i = \frac{1}{T_{i+1}-T_i}\left(\frac{1}{Z(T_i,T_{i+1})} - 1\right) \end{aligned}$$

$$f_i = \frac{1}{T_{i+1}-T_i}\left[\exp\left(\int_{T_i}^{T_{i+1}} \bar{f}(T_i,\tau)d\tau\right) - 1\right] \qquad \text{using } Z(T_i,T_{i+1}) = \exp\left(-\int_{T_i}^{T_{i+1}} \bar{f}(T_i,\tau)d\tau\right) **$$

$$\begin{aligned} &\text{often presented simply as} \\ &= \frac{1}{\tau}\left[\exp\int_t^T f(s)ds - 1\right] \qquad \text{where } t = T_i \text{ and } T = T_{i+1} \end{aligned}$$

Given that $\mathbb{E}^{\mathbb{Q}}_{i+1}[L(T_i,T_{i+1})] = f_i$ we found a formula for the **Forward LIBOR**. The numerarie for this rolling expectation is a bond $Z(0,T_{i+1})$.

f_i is defined as the simple annualised rate on a **Forward Rate Agreement** starting at T_i and ending at T_{i+1}. f_i is a known rate quoted for a FRA contract and used in LMM calibration.[8]

[7] In the HJM Framework, the first column (forward rate of the shortest tenor) is a proxy for $r(t)$ and can be used to price a bond. That removes a need to calibrate a one-factor model such as Hull-White, Vasicek from market-traded bonds (spot curve).

[8] FRAs are used to construct an instantaneous forward rate curve $\bar{f}(t,\tau)$. In HJM calibration, we start with a ready curve and that is why we need a Forward LIBOR formula to back out f_i.

Caplet Pricing (with HJM output). Note by Dr Richard Diamond, CQF.

The simple interest rate options are caps and floors. A cap can be treated as a sum of caplets, each **re-setting (expiring) at** T_i and **maturing (paid) at** T_{i+1}. Caplet rate is paid for the time $\tau_i = T_{i+1} - T_i$. The amount of interest accrued is equal to $(T_{i+1} - T_i)L(T_i, T_{i+1}) = \tau_i f_i$.

Interest rate options deliver a cashflow at maturity, ie, interest paid on the notional. The interest payment is unknown until the rate is fixed (re-set). Caplet cashflow is valued as

$$\mathbb{E}^{\mathbb{Q}}\left[\exp\left(-\int_0^{T_i} \bar{f}(t,\tau)d\tau\right) \times \text{Payoff}_{Cpl}\right]$$

Here, integration over the forward curve (row) at time t gives a discounting factor for the period $[0, T_i]$. Now, for the **forward-starting** caplet, we define the payoff as

$$\text{Payoff}_{Cpl} \quad = \quad \frac{\tau_i}{1+\tau_i f_i} \max\left[f_i - K, 0\right]$$

where $1/(1+\tau_i f_i)$ can be recognised as discount factor for the period $[T_i, T_{i+1}] \equiv \tau_i$.
$f_i \equiv L(T_i, T_{i+1})$ is Forward LIBOR, a simple annualised rate applied to that future period.

Price a caplet simply by

$$\boxed{\text{DF}_{\text{OIS}}(0, T_{i+1}) \times \max\left[L(T_i, T_{i+1}) - K, 0\right] \times \tau}$$

Notice that discounting factor covers both periods, $[0, T_i]$ and $[T_i, T_{i+1}]$. DF can be taken from the OIS spot curve. Proper pricing would require a stochastic DF obtained from the same rolling measure $\mathbb{E}^{\mathbb{Q}}_{i+1}$ as applied to the $f_i = L(T_i, T_{i+1})$. That would require calibrating a separate HJM model for OIS forwards – this is not for implementation.

Once decided on discounting, the question becomes how to obtain that **Forward LIBOR**?

1. FRA formula derived in the previous section gives direct result

$$\begin{aligned} L(T_i, T_{i+1}) \quad &= \quad \boxed{\frac{1}{T_{i+1} - T_i}\left[\exp\left(\int_{T_i}^{T_{i+1}} \bar{f}(T_i,\tau)d\tau\right) - 1\right]} \quad \ddagger \\ &= \quad \frac{1}{T_{i+1} - T_i}\left[\frac{1}{Z(T_i, T_{i+1})} - 1\right] \qquad \text{where} \quad Z(T_i, T_{i+1}) = \frac{Z(0, T_{i+1})}{Z(0, T_i)} \end{aligned}$$

   Choice 1. *Yield Curve.xlsm* implements this Forward LIBOR calculation **from the same curve 'today' because no future simulated curves are available.**

   Choice 2. Integrate over the simulated curves $\bar{f}(T_i, \tau)$ while decreasing the tenor. Consider cashflows in reverse: the first carries the longest credit risk. For example, $L[0.75, 1]$ is taken from the curve today $T_i = 0$ and reflects the credit risk of 1Y loan; $L[0.5, 0.75]$ from curve $T_i = 0.25, \ldots, L[0, 0.25]$ from $T_i = 0.75$.
   The method also gives LIBOR over longer terms, eg, $L[0, 0.75]$ can be obtained by integration over the curve today $T_i = 0$ from 0 to $\tau = 0.75$; $L[0, 0.5]$ from curve $T_i = 0.25$ integrating from 0 to $\tau = 0.5$. This choice combines the logic of taking credit risk as seen today with the idea that over time we work with shorter sections of the curve.

2. Alternatively, we can start with a yield on a future-starting AA-rated bond

$$L(T_i, T_{i+1}) = -\frac{\log Z(T_i, T_{i+1})}{T_{i+1} - T_i}$$

Log discount factors means being linear in rates, so operating on a forward curve

$$L(T_i, T_{i+1}) = \frac{1}{T_{i+1} - T_i} \int_{T_i}^{T_{i+1}} \bar{f}(T_i, \tau) d\tau \quad \Rightarrow \quad \boxed{\frac{1}{n} \sum_{j=1}^{n} \bar{f}_j} \quad \dagger$$

where n is a number of points between tenors $[T_i, T_{i+1}]$ and integration is over the curve(s).

Proper LIBOR is a simple annualised rate, which for short periods can be converted using

$$L' = m \left(e^{\frac{L}{m}} - 1 \right) \quad \dagger$$

where m is compounding frequency per year. *Examples:* 3M LIBOR compounded 4 times a year, 6M LIBOR - 2 times. Calculate † and convert to L'.

Choice 3. The notation $\bar{f}(T_i, \tau)$ implies picking each Forward LIBOR from the future curve as T_i increases. For example, $L(T_i, T_{i+1}) = L(4, 4.25)$ is picked from the curve simulated at time $T_i = 4$ (row 400 with $dt = 0.01$) and the column for tenor $\tau = 4$. Why? Even 3M rate for $[4, 4.25]$ has to carry the credit risk of a longer-term loan.

Using the forward curve simulated for $t = 4$ is awkward. The notation $\bar{f}(t = T_i, \tau)$ is best suited to **the constant maturity case: when we are not exposed to the long-term credit risk but only to 3M or 6M LIBOR paid over and over**. For example, for an IRS (swap) we can use the same column of HJM output to pick Forward LIBOR in $t = 0.5$ (row 50), in $t = 1$ (row 100) and so on.

Market risk factors that affect caplet price are the inputs into pricing formula. The terminology comes from the formal risk management, which regards any input as 'a risk factor'.

1. The standard risk factors are strike, maturity and 'bucket risks', ie, bumping rates at particular tenors and re-pricing. In a sense, HJM does that by introducing shocks $\sqrt{\lambda}\phi\delta t$.
2. Semi-annual vs. quarterly expiry. BOE forward curve is already smoothed for 6M increments – re-interpolation not advisable.
3. Choice of discounting factor: from the model vs. OIS vs. simulated OIS forwards to match the forward risk measure.

We can **validate** by comparing pricing results that utilize HJM output (eg, constant maturity or reverse cashflows method) to prices implied by the forward curve at $t = 0$, known with certainty.

Caplet-floorlet parity is given by

$$\mathbf{Cpl}(0; T_i, T_{i+1}, K) - \mathbf{Flr}(0; T_i, T_{i+1}, K) = Z(0, T_{i+1}) \times (L(T_i, T_{i+1}) - K) \times \tau$$

where $t = 0$ means that we are pricing from today's curve but the caplets (floorlets) are forward-starting at time T_i. Discount factor $Z(0, T_{i+1})$ can be substituted with one taken from OIS spot curve $\mathrm{DF}_{OIS}(0, T_{i+1})$ (or simulated OIS forwards to match the expectation).

Black Formula

The formula converts the discounted cashflow of caplet into implied volatility quotation and applies. Market volatility for 3M caplets $\mathbf{Cpl}^{LMM}$ requires bootstrapping from traded caps of longer maturity. As usual, $t = 0$ means that we are pricing from today's curve but the caplets (floorlets) are forward starting at time T_i.

$$\begin{aligned}\mathbf{Cpl}^{LMM}(0; T_i, T_{i+1}, K) &= \mathbf{Cpl}^{Bl}(0; T_i, T_{i+1}, K, \zeta_i) \\ &= Z(0, T_i) \times Bl(K, L(T_i, T_{i+1}), \zeta_i)\end{aligned}$$

- Complete notation $L(t; T_i, T_{i+1})$ is equivalent to $L_i(t)$ and $F_i(0)$ means Forward LIBOR fixed from today's curve $t = 0$. $F(T_i)$ notation also used, referring to the same quantity.
- Discount factor $Z(0, T_i)$ is separated from $Z(T_i, T_{i+1})$ (which is technically under the risk-neutral expectation).

$$\begin{aligned}Bl(K, L(T_i, T_{i+1}), \zeta_i) &= [L(T_i, T_{i+1})\,\Phi(d_1) - K\Phi(d_2)]\,\frac{\tau}{1 + \tau L(T_i, T_{i+1})} \\ &= [f_i\,\Phi(d_1) - K\Phi(d_2)]\,\frac{\tau}{1 + \tau f_i}\end{aligned}$$

$$\text{where} \quad d_{1,2} = \frac{\ln \frac{f_i}{K} \pm \frac{\zeta^2}{2}}{\zeta}$$

in these formulae $\zeta^2 = \zeta_i^2 T_i$ where ζ_i is an implied volatility of the caplet that expires at T_i and pays off at T_{i+1}. Once the rate is fixed at T_i, there is no more volatility!

The implied variance of forward rate f_i (also referred to as FRA rate) is an integrated instantaneous variance $\zeta_i^2 = \frac{1}{T_i}\int_0^{T_i} \sigma^2(s)ds$.

LMM SDE

Discretised LMM SDE to implement looks confusing. **Notation t_{j-k-1} just means that we refer to the previous time step $k - 1$.**

We can express the SDE in simpler terms by changing $f(t_{k+1}) \to f(t + 1) \to f(t + \delta t)$ and $f(t_k) \to f(t) \to f$. There is duplicate index i in $\sigma_i(t_{i-k-1})$, while $i - k$ relates to how we bootstrapped the instantaneous volatility of f_i from caplets, given that after time t_i the forward rate is fixed and there is no volatility. Lets write $\sigma_i(t_{i-k-1}) \to \sigma_{i,t-1} \to \sigma_i$ (can also say, we take sigmas 'from the previous row'). LMM SDE must look simpler,

$$f_i(t + \delta t) = f_i \exp\left[\left(\sigma_i \sum_{j=0}^{i} \frac{\tau_j f_j\,\sigma_j \rho_{ij}}{1 + \tau_j f_j} - \frac{1}{2}\sigma_i^2\right)\delta t\right] + \sigma_i \phi_i \sqrt{\delta t} \tag{1}$$

Note that this is a single-factor SDE. With the discrete market model dt can be large.

Understanding HJM Output. Note by Dr Richard Diamond, CQF.

There is **a multiplicity of ways** to extract rates and discounting factor information from the HJM model's simulated output:

1. Use the first column – this reduces the HJM to a short rate model for $r(t) = f(t,t)$.
2. Average across the evolution of rate $\bar{f}(\tau)$ (data in column) – numerically, this is to the same effect as the next way. This gives the forward rate of constant maturity per period.
3. Pick values from the relevant cell **under Monte-Carlo.** For example, you can pick $L(0.5, 1)$ as in HJM MC caplet pricing example, or $L(4, 4.25)$ as above and average.
4. Integrate over the curve from 0 to τ (data in row) – this assumes we are not simulating future curves. It is technically possible to calculate $L(T_i, T_{i+1})$ and $DF(0, T_{i+1})$ from today's curve as done in *Yield Curve.xlsm* but that disregards simulated curves.

Caplet section above shows the actual pricing method is more complex than these ways. It utilises several curves $\bar{f}(T_i; \tau)$ taken at rate reset times T_i and pick LIBOR values a step-like fashion under Monte-Carlo. Therefore, it relies on two ways in order to correspond to the appropriate forward risk expectation and include the credit risk of the rates for longer term loans (money markets) with committed notional.

Evolution of the forward rate at each tenor with the timestep $dt = 0.01$ offers paths ready for precise integration. It is theoretically consistent to calculate LIBOR-per-period as an average of the instantaneous rate over that period (column for τ_j from T_i to T_{i+1}). Unfortunately, this approach is neither affine nor simple. If we start averaging over columns the samples will be of different length T_i. To pick values from cells under Monte-Carlo remains the best way.

Expiry fallacy: Consider a cap on 3M LIBOR (paying quarterly) with the final rate re-set in 9M time (expiry).

- One can choose the curves at $T_i = 0, 0.25, 0.5, 0.75$ 9M pick the rate from period $[0, 0.25]$ from each curve $\bar{f}(T_i, \tau = 0.25)$. One treats the 6M-9M section of the today's forward curve as 'an equivalent' to the 0M-3M section of the future simulated curve from $T_i = 0.5$.

We are reduced to the case of constant-maturity forward rates. The fallacy here is as follows: the credit risk of capital committed never factors into the pricing. For caplet pricing, constant maturity case is likely to produce a relatively flat term structure of implied volatility – that would be inconsistent with the idea of interest rate risk (eg, 3M LIBOR is uncertain after today's fix) and its components of credit premium and term liquidity premium.

Plot simulated curves for each timestep $dt = 0.01$ on a single 3D surface to observe the dynamics and make own conclusions.

Relationships across Yield Curve (with statistical modelling)

The HJM model evolves the full term structure (yield curve) $\bar{\boldsymbol{f}}(\boldsymbol{T};\boldsymbol{\tau})$. As a by-product we have evolutions of the forward rates for each tenor τ. For historical data, we evaluate the relationship between time series $\bar{\boldsymbol{f}}(\tau_j)$, $\bar{\boldsymbol{f}}(\tau_k)$ using either cointegration analysis or correlation coefficients.

Correlation is always estimated between changes (differences) $\Delta\bar{\boldsymbol{f}}$, which themselves are stationary. At the start of the document, there are plots of changes in forward rates at $1Y$ vs. $6Y$ vs. $25Y$ tenors as well as discussion about common types of premium for interest rates. The plots demonstrate independence (in daily movements) of the short rates up to and around $\tau = 1$ year. This is expected because the short rates have no significant credit risk premium component (factor) that can be shared by the rates of various tenors.

[An open study task is to build cross-correlation as well as autocorrelation profile of the yield curve (rates at different tenors) in order to capture dynamics of those correlations.]

The next logical step is **linking rates at different tenors via constant spreads**:

$$\bar{f}(\tau_k) = \bar{f}(\tau_j) + \text{Credit Premium} + \text{Term Premium} \quad \dagger$$

It is market practice for IR Derivatives to strip spreads that capture these premiums.

- One can make empirical estimations of credit risk and term liquidity risk for different tenors, then add to the short rate r_T to construct a full curve. This can be done for the stochastic process of $r(t)$ in order to obtain the full yield curve at each simulation step.
- One can see an actual yield curve as a collection of flat curves for rates of different tenors. This naturally creates a collection of tenor spreads, from which OIS and other relevant spreads (currency) can be subtracted.

  $$\text{L}_{3M} = \text{OIS}_{3M} + \text{Basis Spread} \qquad \text{vs.} \qquad \text{L}_{3M} = \text{OIS}_{ON} + \text{3M to OIS Spread}$$

  '3M to OIS Spread' is an example of basis swap contract traded or stripped from the market. Therefore, you can take 3M USD LIBOR and build up the yield curve in the currency of choice (as farther as the basis swap spread curve allows). You can also take 6M USD LIBOR and build a different yield curve. Use the curve of a matching frequency (ie, 3M vs. 6M) to price and strip the traded instruments.

The simple proposition † 'works' because there is a **cointegrated relationship** between rates at neighbouring tenors (and across the curve): each tenor is stochastic but *on average*, two tenors keep a constant distance and the error correction coefficient $(1-\alpha)$ comes up as significant in

$$\Delta\bar{f}_{k-j} = -(1-\alpha)e_t = -(1-\alpha)(\bar{f}_j - \beta_C\bar{f}_k - \text{Premium})$$

Naively, $\Delta\bar{f}_{k-j}$ = constant but the concept of long run equilibrium suggests that it will be sufficient to have significant term e_t, a stationary residual. This kind of analysis detailing a cointegrated relationship between spot rates of 10Y and 25Y tenors is given in Cointegration Case B (CQF Lecture). The premium can be extracted from the deterministic trend of cointegration.

Fair Spread for A Basket CDS

Spread Calculation

Premium Accrual Analytical expression for spread computation with discrete-time accruals becomes complicated for 2nd-to-default and above. Remember that while multiple defaults are possible, only one is protected from.

- Continue with 2nd-to-default example and assuming that 1st default occurs before Year 2 end $\tau_1 < 2$, and 2nd default occurs in Year 3 $\tau_2 < 3$, then accrual can look like

$$s = \frac{(1-R)Z(0,\tau_2) \times \frac{1}{5}}{[Z(0,t_1)\Delta t_1 + Z(0,\tau_1)\delta t_{k=1}] \times \frac{5}{5} + [Z(0,t_2)(t_2-\tau_1) + Z(0,\tau_2)\delta t_{k=2}] \times \frac{4}{5}}$$

This and workshop implementations assume there is a default. If there is no default, keep accruing the premium until the 5Y end (expiry). This will only be one simulated value of PL.

Implied curve fitting The generalised coding solution is: create a running variable that accumulates PL at each time step $dt = 0.01$. For that you will need discount factors for each time step (a fitted discounting curve). This approximates continuous-time accrual – if we calculate the PL to default time exactly then, we do not need to adjust DL for accrued premium.

The alternative to near-continuous accruals are quarterly, semi-annual, annual in arrears (end of the year), or upfront payment of the premium. Each of these choices will modify PL (the denominator), often leading to substantial simplification. **Code any suitable choice.**

Averaging PL and DL **Remember** to calculate the averages of DL and PL across simulations separately. Then calculate the spread once as

$$s = \frac{< DL >}{< PL >}$$

You are likely to find out that *a few early defaults* (e.g., from one day to 6M) are likely to bring PL average down and over-estimate the spread (introduce a bias). Therefore, you might impose a restriction that the first default can only occur (be settled) after first quarter, $\min \tau_k = 0.25$ or implement premium accrual for the whole year (paid in arrears). Here, you can also try variance reduction and importance sampling techniques, such as giving higher weight to simulations that have defaults in later years. Usual MC weights each simulation as $\frac{1}{T}$.

It is not advised to mix data for 1st, 2nd-, 3rd-to-default pricing for clarity reason.

These recommendations aim to improve the convergence of fair spread.

However, it has been reported up to $600,000$ rounds of Monte-Carlo were required for the spread to stabilise. Apart from code bugs, several inputs and factors impact computational stability and MC covergence in copula methods: **a.** averaging spread (rather than DL, PL) typically over-estimates the spread, **b.** a few early defaults introduce bias, **c.** using low spreads (i.e., 10-30 *bps*) is on the border of perturbations/computational instability, and **d.** correlation matrix is problematic (i.e., correlations are very high or homogeneous) and has no determinant.

Calculating Exact Default Time τ_i from simulated u_i

Exact default time $\tau = t_{m-1} + \delta t$ is estimated by first, identifying the year by an iterative procedure (comparing $\log(1-u)$ to the running sum of hazard rate) and second, calculating the year fraction δt.

Extending derivation for $\tau \overset{D}{=} F^{-1}(u)$ as the inverse of Exponential Distribution **CDF**

$$\tau \overset{D}{=} -\frac{\log(1-u)}{\lambda_\tau} \quad \Rightarrow \quad \log(1-u) = -\lambda_\tau \tau = -\int_0^\tau \lambda_s ds = \log P(0,\tau)$$

If $1-u = P(0,\tau)$ is the exact probability of survival, and default occurs as $t_{m-1} \leq \tau \leq t_m$

$$P(0,t_m) \;\leq\; P(0,\tau) \;\leq\; P(0,t_{m-1})$$

because survival probability is cumulative and decreasing over time (see CDS Bootstrapping). Re-specified in terms of intensities $\log P(0,t_m) = -\sum_{j=1}^{m} \lambda_j \Delta t_j$, the inequality becomes:

$$-\sum_{j=1}^{m} \lambda_j \Delta t_j \;\leq\; \log(1-u) \;\leq\; -\sum_{j=1}^{m-1} \lambda_j \Delta t_j \tag{2}$$

First, the inequality is used in the iterative procedure to determine the year of default.

We can continue to re-arrange around u

$$\begin{aligned} P(0,t_m) \;\leq\; 1-u \;&\leq\; P(0,t_{m-1}) \\ -P(0,t_{m-1}) \;\leq\; u-1 \;&\leq\; -P(0,t_m) \\ 1-P(0,t_{m-1}) \;\leq\; u \;&\leq\; 1-P(0,t_m) \\ \mathrm{PD}_{m-1} \;\leq\; u \;&\leq\; \mathrm{PD}_m \end{aligned}$$

Since we compare the threshold u to the cumulative probability of default, then if u is large (more than cumulative PD for year 5) it has a meaning of default occurring after year 5.

Second, calculating the year fraction also relies on $1-u = P(0,\tau)$, where $\tau = t_{m-1} + \delta t$

$$\delta t = -\frac{1}{\lambda_m} \log\left(\frac{1-u}{P(0,t_{m-1})}\right) = -\frac{1}{\lambda_m} \log\left(\frac{P(0,t_{m-1}+\delta t)}{P(0,t_{m-1})}\right) \tag{3}$$

This essentially refers to calculation of hazard rate as a log-ratio of survival probabilities.

Copula Fitting. Note by Dr Richard Diamond, CQF.

Copula estimation operates with pseudo-samples U rather than original data X, consists of five columns of historical data, used for estimation of correlation matrix $\mathbf{\Sigma}_{5\times5}$. Once chosen how to convert the historical data into uniform the **sampling from copula** algorithm implementation is a straightforward by-step. Choices are **converting to Normal variable by *differencing* $\Delta\mathbf{X} \to \mathbf{Z}$** or **applying probability integral transform and *kernel smoothing* $\mathbf{X} \to \mathbf{U} \to \mathbf{Z}$**.

For Gaussian copula, we estimate linear correlation ρ on Normally distributed $\mathbf{Z}$ so $\mathbf{\Sigma} = \rho(\mathbf{Z})$, for t copula we estimate correlation on the ranks of $\mathbf{X}$ so $\mathbf{\Sigma}_S = \rho(\mathbf{U})$ for Spearman's rho, while separate formula $\mathbf{\Sigma}_\tau = \rho_\tau(\mathbf{X})$ is defined for Kendall's tau. To convert into linear correlation $\rho = 2\sin\left(\frac{\pi}{6}\rho_S\right)$ and $\rho = \sin\left(\frac{\pi}{2}\rho_\tau\right)$ elementwise. This converted matrix is not guaranteed to be positive definite as required for Cholesky – so the nearest correlation matrix is obtained.

Historical sample data $\mathbf{X}^{\text{Hist}}$ (five columns credit spreads/default probabilities/hazard rates) is converted to pseudo-samples $\mathbf{U}^{\text{Hist}} = \widehat{F}(\mathbf{X}^{\text{Hist}})$. That is achieved by special transformation of data by its own *Empirical CDF* and involves *kernel density estimation* in order to guarantee uniformity.[9] Estimation done without making assumption about distribution of marginals is non-parametric and called Canonical Maximum Likelihood.[10]

Let's consider notation as we go from copula fitting (ie, calibration) to simulation,

- $\mathbf{Z} = \Phi^{-1}(\mathbf{U})$ obtained from pseudo-samples, so can be expressed as $\mathbf{Z}^{\text{Hist}}$. Use $\mathbf{\Sigma} = \rho(\mathbf{Z}^{\text{Hist}})$ The shortcut which avoids kernel smoothing is first, take differences $\mathbf{X} = \Delta\mathbf{X}^{\text{Hist}}$ and second, standartise $\mathbf{Z}_t^{(j)} = \frac{\mathbf{X}_t^{(j)} - \mu_j}{\sigma_j}$ for each row (observation) t of column j.
- For calculation of copula density, $\mathbf{U}_t \equiv \mathbf{U}_{t,1\times5}$ refers to **a row** of values for five reference names as observed at time t.
- $\mathbf{Z_{t+}}^{\text{Sim}}$ or simply $\mathbf{Z_{t+}}$ is a vector of simulated 1×5 Standard Normal random variables, and so $\mathbf{U_{t+}} = \Phi(\mathbf{Z_{t+}})$ for Gaussian or $\mathbf{U_{t+}} = T_\nu(\mathbf{Z_{t+}})$ for t copula.

  For the simulated 1×5 $\mathbf{U_{t+}}$, each value is converted to default time $u \to \tau$ using *its own* term structure of hazard rates

$$\tau \sim Exp(\hat{\lambda}_{1Y}, \ldots, \hat{\lambda}_{5Y}).$$

Elliptical copulae might fail to fit dependence structure of empirical data (eg, higher density of tail observations, low density of the middle high-peaked observations). That is a model risk the copula method. A quick recipe is to **check bivariate scatters between the columns of U** – the scatter should have the familiar pattern of Elliptical copula density.

[9] There is no analytical formula for Empirical CDF function. It is obtained via a set of algorithms.

[10] Each column of $\mathbf{X}^{\text{Hist}}$ is 'a marginal' with its own univariate distribution that is usually bi-modal for raw credit spreads. Therefore, we have to work with *changes* in spreads $\Delta\mathbf{X}$ (daily or weekly).

Kernel Smoothing: An Introduction

The term refers to the estimation (fitting) of analytical probability density function $\widehat{f}()$ to the data. Most software-implemented kernel smoothers fit probability density function (PDF), from which additional steps have to be taken to obtain CDF $\widehat{F}$ – those are numerical integration over kernel PDF and interpolation. Altogether, the set of algorithms is known as Probability Integral Transform:

$$\mathbf{U} = \widehat{F}(\mathbf{X})$$

Performing MLE on pseudo-samples $\mathbf{U}$ instead of the original data $\mathbf{X}$ is a superior approach. For example, applying the familiar linear correlation formula on ranks $\mathbf{U}$ immediately delivers Spearman's rho, a rank correlation measure.

$$\mathbf{\Sigma}_S = \rho(\mathbf{U})$$

- $\mathbf{X} \to \mathbf{U} \to \mathbf{Z}$. Kernel smoothing on $\mathbf{X}^{\text{Hist}}$ by Empirical CDF algorithm (as implemented in Matlab/R/NAG functions), where implementation guarantees the uniformity of $\mathbf{U}^{\text{Hist}}$.
- $\mathbf{X} \to \mathbf{Z} \ldots \mathbf{U}$. Hidden assumption that original data $\mathbf{X}^{\text{Hist}}$ is converted to near-Normal, for example, by differencing $\mathbf{X} = \Delta\mathbf{X}^{\text{Hist}}$. Next steps are standardization $\mathbf{Z}_t^{(j)} = \frac{\mathbf{X}_t^{(j)} - \mu_j}{\sigma_j}$ and inferring pseudo-samples $\mathbf{U} = \Phi(\mathbf{Z})$. However, empirical pseudo-samples obtained this way (without kernel smoothing) might be insufficiently uniform.[11]

Where possible *use the ready implementation of kernel smoothing that gives Empirical CDF*, such as Matlab *ksdensity(), and check the uniformity of the output* $\mathbf{U}$ by plotting a histogram for each column (reference name).

When using kernel smoothing **the choice of bandwidth is very important**! Think of it as a bucket of observations for cumulative probability step (standard deviation on uniform scale). MATLAB's *ksdensity()* calibrates some optimal bandwidth, however you might be able to get better smoothing result in terms of uniformity of output $\mathbf{U}$ by interactively experimenting with the 'bw' setting from default down to circa 0.0001.

NAG kernel density estimation (PDF only) does require bandwidth as a ready input, so interactive experiment is necessary. Setting the bandwidth (window width) too high results in the data being represented as fully Normal and therefore, *oversmoothed* and highly correlated across names. Setting the bandwidth too low represents data very close to original (*undersmoothed*) and results in u_i that are zero or close, creating a problem with TINV calculation $T_\nu^{-1}(\mathbf{U})$.

- Each data column of $\mathbf{X}^{\text{Hist}}$ might require calibration of its own bandwidth setting. Here, for kernel PDF data are *changes* in credit spreads/default probabilities/hazard rates.
- Credit monitors rely on weekly changes and drop $1-3\%$ of extreme observations. [12]

[11] $\mathbf{Z} \ldots \mathbf{U}$ by $\mathbf{U} = \Phi(\mathbf{Z})$ is actually **the wrong way** but we are trying to see what $\mathbf{U}$ is implied.

[12] In utmost generality, one can look at changes between 5Y hazard rates averaged per period.

Kernel smoothing in various contexts

Let's begin with the idea of Empirical CDF $\widehat{F}()$. CDF is a function that assigns the lower weight on common observations and the higher weight on tail observations in order to transform a bell-shaped histogram to uniform histogram. Consider two simple cases of Empirical CDF.

- Case 1: bucket of one, each observation is converted as $u_i = \frac{1}{N}$. Then the uniformity is perfect but correlating the set of the same u_i with another alike set is not useful.
- Case 2: create buckets and calculate an average of each bucket. By the Central Limit Theorem the distribution of averages is guaranteed to be Normal.

Task: Fit the uniform density (ie, shown as a histogram) from the Normal data $\mathbf{X}$ (or close). The center of the density is not the target for conditioning.

Method: Implement state-conditioning using exponents for each tail. Exponential kernel gives flexible probabilities p_t, which will be applied to the original data as $u_i \propto p_i \, X_i$

$$p_t|z^* \propto e^{-\frac{|z_t - z^*|}{h}}$$

where bandwidth h decreasing as N_{obs} increases and $z^*_{1,2}$ 'cut' each of the tails.

Extra Smoothing: Given the data has already been converted to uniform $\mathbf{U}_t = \widehat{F}(\mathbf{X}_t)$, in this context kernel smoothing means fitting a continuous PDF from discrete observations $\mathbf{U}_t$.

$$\hat{f}(u, h) = \frac{1}{hN} \sum_{t=1}^{N} K\left(\frac{u - U_t}{h}\right)$$

u is a continuously smoothed variable, and bandwidth h decreases as N increases. The formula applies for any density. The choice of a PDF-like function for K is not important, for computational efficiency 'Gaussian kernel' (similar to exponential) is used

$$K(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \qquad \text{where} \quad \int_{-\infty}^{+\infty} K(t)dt = 1$$

For the uniform density the recommended kernel is $K = \frac{\Gamma(2)}{2\,\Gamma(1)\Gamma(1)}\, 1_{|u|<1}$. However, since the fraction is equals to $1/2$ it is easy to see that this expression is over-fitting and results in the same u_i (Case 1).

Histogram: is simply defined for any random x (as opposed to just u) as

$$\hat{f}(x, h) = \frac{1}{hN} N_j$$

where N_j is the number of observations falling in the interval $a + (j-1)h < x < a + jh$. For example, take sample of $N = 100$ and bandwidth of $h = 0.1$ ($N_s = 10$ observations per bucket expected if data is uniform) so, if $N_j = 8$ observations then bar height is 0.8, if $N_j = 15$ observations bar height is 1.5.

Copula Fitting: Rank vs. Linear Correlation

The model behind linear correlation is a linear regression, ie, the ability to draw a straight line to explain relationship between changes in two variables. Due to this shortcoming, the low linear correlation does not imply low dependence because the relationship can be strong but non-linear.

The idea behind rank correlation is rather simple, **each value can be converted to rank, then ranks are correlated**. Estimating from ranks gives scale-invariant measures; converting into ranks requires either formula or algorithm but the method itself is non-parametric.[13] Rank correlation measures are also invariant to any non-linear transformation in variable as long as it is monotonic. These properties are very suitable for copula fitting because **a.** we deal with credit spreads/PDs/hazard rates data and **b.** we desire robust estimation of dependence structure. Rank correlation matrix is unlikely to vary as we transform data, particularly to pseudo-samples $\mathbf{U}^{\text{Hist}}$.

- **Linear correlation** is a good measure of co-dependence for Normal variables. If the Normality assumption is violated, linear correlation can be very misleading. Linear correlation is not preserved under any general transformation. There is problem with non-elliptical distributions and non-linear transformation of a random variable (factor). For example, transforming from Normal to log-Normal variable by simply taking a log creates the problem of *attainable correlation* – if correlation measured between log-Normal variables, then its transformation into linear by something like $e^{\rho \cdots}$ will not give the full range $[-1, 1]$.

  The linear correlation of two Normal random variables is

  $$\rho(X_1, X_2) = \frac{\mathbb{C}\text{ov}(X_1, X_2)}{\sqrt{\mathbb{V}\text{ar}(X_1)\mathbb{V}\text{ar}(X_2)}}$$

- **Spearman's rho** is formally defined as a linear correlation of associated CDFs. The ranks of the data points are usually obtained by applying Empirical CDF to the data $\mathbf{U} = \widehat{F}(\mathbf{X})$ (up to a multiplicative factor).

  $$\rho_S = \rho(U_1, U_2) = \frac{\mathbb{C}\text{ov}(F(X_1),\, F(X_2))}{\sqrt{\mathbb{V}\text{ar}(F(X_1))\mathbb{V}\text{ar}(F(X_2))}}$$

  Spearman's rho estimation: for N pairs of observations $(X_{1,i}, X_{2,i})$, generate ranking in the ascending order and calculate

  $$\rho_S = 1 - \frac{6\sum D_i^2}{N(N^2-1)}, \qquad \text{where } D_i = (U_{1,i} - U_{2,i})$$

[13] Remember how applying CDF to the values is a special transformation [insert formula] that gives grades, and copula is referred to as a distribution of grades.

- **Kendall's tau** is another rank correlation measure, formally defined as the difference between probability of concordance and dis-concordance. It measures the degree to which the large values of one variable associate with the large values of another variable.

$$\begin{aligned}\rho_\tau &= \Pr\left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0\right] - \Pr\left[(X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0\right] \\ &= \mathbb{E}\left[\text{sign}\left((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2)\right)\right]\end{aligned}$$

A positive sign refers to the increasing pair (line goes upwards and downwards but its slope is positive), while a negative sign gives at least one decreasing case.

Pairs (X_1, X_2) and 'an independent copy' $(\tilde{X}_1, \tilde{X}_2)$ are drawn from the identical joint distributions $F(X_1, X_2) \equiv F(\tilde{X}_1, \tilde{X}_2)$ are concordant if $X_1 > \tilde{X}_1$ and $X_2 > \tilde{X}_2$.

Kendall's tau estimation: for N pairs of observations $(X_{1,i}, X_{2,i})$, we use a sign function $sign(x) = -1$ if $x < 0$ and $sign(x) = 1$ if $x > 0$.

$$A_{ij} = \text{sign}\left((X_{1,i} - X_{1,j})(X_{2,i} - X_{2,j})\right)$$

and estimate using

$$\rho_\tau = \mathbb{E}[\mathbf{A}] = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i}^{N} A_{ij} = \frac{N_c - N_d}{N(N-1)/2}$$

where N_c is the number of concordant pairs and N_d is the number of discordant pairs.

The linear correlation is a required input to both, Gaussian and t copulae. If using a rank correlation measure the result has to be converted as $\rho = 2\sin\left(\frac{\pi}{6}\rho_S\right)$ and $\rho = \sin\left(\frac{\pi}{2}\rho_\tau\right)$.

Linear Correlation from Returns Data $\mathbf{X} \to \mathbf{Z} \ldots \mathbf{U}$ estimation can be streamlined by relying on a reasonable assumption of Normal distribution for asset returns. This works better over the long-term. So, for estimating default correlation matrix using equity returns it makes sense to use a 3-year historical time period (longer period for weekly data).

1. Convert price data into linear returns $\mathbf{R} = .\dfrac{X_{t+1} - X_t}{X_t}$ where X_t is equity price at time t.
2. Normalise returns $\mathbf{Z} = \dfrac{\mathbf{R}_t^j - \widehat{\mu}}{\widehat{\sigma}}$ where $\widehat{\mu}$ is sample mean and $\widehat{\sigma}$ is sample std. deviation.
3. Calculate linear correlation $\widehat{\mathbf{\Sigma}} = \rho(\mathbf{Z})$ (between each two columns of data).

The *implied* pseudo-samples can be obtained $\mathbf{U} = \Phi(\mathbf{Z})$ and checked for uniformity (by histogram). On the other hand, one can revert to the preferred CML method by applying kernel smoothing to returns $\mathbf{U} = \widehat{F}(\mathbf{R})$ and obtaining $\mathbf{Z} = \Phi^{-1}(\mathbf{U})$ – in this case, the kernel smoothing of Empirical CDF $\widehat{F}(\mathbf{R})$ is a conventional task of smoothing over the Normal-like density of $\mathbf{R}$.

Important Disclaimer. Imposing assumption about distribution of changes the procedure from *Canonical Maximum Likelihood* to *Inference Function for Margins*. CML gives MLE-optimal copula fit without parametric assumptions about marginal distributions.

Where rank correlation fails

Remember that correlating *probabilities of default* is the preferred method to obtain default correlation matrix (while estimation from equity returns is just an acceptable choice). If one starts examining time series of credit spreads/PD/hazard rates empirically (see DB Research data for Sovereign CDS_{5Y} and ready-bootstrapped PD_{5Y}) the quick findings will be as follows:

- credit spread/PD time series are non-stationary (ie, have unit root); and
- exhibit bimodal distribution (histogram)

Rank correlation measures were developed for any random variable but their application to the levels of non-stationary variables remains problematic. The important aspect is that **even rank correlation measures require co-monotonic (linear, exponential, etc) dependence.** Correlation measures fail when there is a break in co-monotonicity (eg, zig-zag or cross shapes). Combining two variables with a bimodal distribution each seems to create non-monotonic dependence.

t Copula and Degrees of Freedom ν (MLE procedure)

Degree of freedom ν is an additional parameter required for t copula that regulates tail dependence – that is, increased likelihood of large(small) values in simulated $\mathbf{U}_{t+}$ appear together.

In our flexible Canonical Maximum Likelihood approach, we first obtain the default correlation matrix and then use the same underlying data (pseudo-samples $\mathbf{U}^{Hist}$), in order to calculate the log-likelihood function (of ν) as sum of densities.[14]

$$\underset{\nu}{\text{argmax}} \left\{ \sum_{t=1}^{T} \log c_t \left(\mathbf{U}_t^{Hist}; \nu, \widehat{\mathbf{\Sigma}} \right) \right\}$$

To calculate a contribution to copula density (joint density) we input $\mathbf{U}_{t,1\times 5}$ values for five reference names as observed at time t. Then, we calculate density contribution at time $t+1$ and up to T, **a row at a time**.[15]

For each value of ν from 1 to 25, we repeat the calculation of total log-likelihood and produce a plot that must be similar to Figure 3 below.

A shortcut to degrees of freedom estimation is to make an informed choice between 4 and 13 (circa 25/2). Do not hesitate using this shortcut at least initially. An explanation for the shortcut comes from the approach to degrees of freedom in a structural equation system.

[14] Inferring linear correlation explicitly first and then 'fixing' it is a restriction compatible with Canonical MLE.

[15] Unlike probability density, copula density $c()$ is calculated 'up to a multiple'. The result can be above 1 and scale depends on the original transformation into $\mathbf{U}$. Canonically, the sum of densities under log will be negative $O(-2000)$ to $O(-800)$.

1. Within a multivariate regression

$$d.f. = N_{total} - N_{var} - 1$$

For the 5×5 correlation matrix which is symmetric with ones on the diagonal,

$$d.f. = (5 \times 5 - 5)/2 - 1 = 9$$

2. If two variables are strongly correlated then we effectively have 4 independent variables,

$$d.f. = (4 \times 4 - 4)/2 - 1 = 5$$

d.f. parameter reduced to from 9 to 5 with increased correlation!

3. With 25-30 d.f. t distribution begins to closely approximate the Normal distribution.

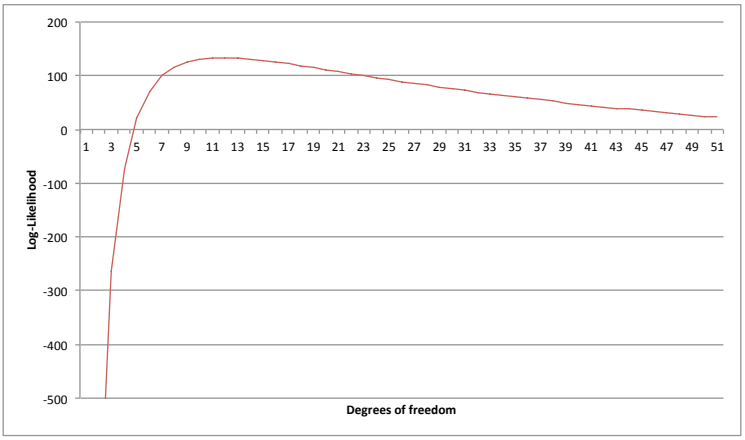

Figure 3: Log-likelihood for t copula as a function of degrees of freedom ν.

Tail dependence index means that with large (small) values of U_1 we expect to observe the large (small) values of U_2 (in probability terms). This applies applies to t copula and is the reason for sampling from it.[16]

$$\lambda_l = \lambda_u = 2t_{\nu+1}\left(-\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}\right)$$

where $\lambda_l = \lambda_u$ is coefficient of tail dependence, ν is degrees of freedom parameter and ρ is linear correlation between U_1 and U_2.

Inverse CDF Working with the t copula requires the CDF $T_v(\mathbf{X})$ (when sampling from) and its inverse $T_v^{-1}(\boldsymbol{U'})$ (when fitting – calculating density $c()$ to obtain likelihood). Student's t CDF formula (2.52) is on page 14 of *Monte-Carlo Methods in Finance* textbook by Peter Jaekel.

[16]Correlated values sampled from Gaussian copula are asymptotically independent in both tails. Even with high correlation input, there is no tail dependence (increase in likelihood of co-movement) from a Gaussian copula.

The formula uses Beta function which is not available in VBA but things can be done using the *BetaDist()* as below. To understand link between *Beta function* and *BetaDist()* compare (2.52) to (2.38) and make necessary derivations.

This works for $x > 0$ (to understand the modification for $x \leq 0$ refer to the formulae)

```
'x = Z-score,  n = degrees of freedom
Public Function MyTDIST(x As Double, n As Single) As Double
MyTDIST = 1 - Application.BetaDist(n / (n + x ^ 2), n / 2, 1 / 2) / 2
End Function
```

To calculate the inverse of a Student's t CDF, the following code can be used

```
Public Function MyTINV(x As Double, n As Single) As Double
Dim tmp As Double
tmp = Application.BetaInv(2 * Application.Min(x, 1 - x), n / 2, 1 / 2)
tmp = 1 / tmp
MyTINV = (Sgn(x - 0.5) * (n * (tmp - 1)) ^ 0.5)
End Function
```

It is up to delegates to confirm validated code.

Summary Estimation of both, correlation and degrees of freedom, is known as **copula fitting**. As an illustration please check out how copula fitting is done in MATLAB by using only a few ready functions. `http://www.mathworks.co.uk/help/stats/copulafit.html`

Sensitivity Analysis (required part of report)

Sensible hedging strategies for a credit portfolio (basket credit swap) rely on the notion of Default Basket Delta that measures sensitivity *w.r.t.* changes in credit spreads of reference names as

$$\Delta_i = \frac{\partial V_{Basket}}{\partial S_i} \Big/ \frac{\partial V_i}{\partial S_i}$$

or Change in Basket Spread / Change in Credit Spread (for a reference name).

If you can, come up with a different term structure of hazard rates to be used for pricing – that would be important aspect of testing the model risk. Sensitivity analysis *wrt* Recovery Rate and Discount Factor (when pricing) is optional. Generally one can explore the limitations of the Credit Triangle $S = \lambda(1 - R)$.

Correlation Sensitivity. Other key input for sensitivity analysis is default correlation matrix $\mathbf{\Sigma}$. *Imposing absolute levels of base default correlation*, ie, same for all names low 0.05 to high 0.95, is a simple stress-testing technique on the fair spread. With t copula, keep the estimated d.f. parameter fixed.

A smarter technique would be *to shift the values of correlation matrix by a percentage*, eg, 10% at a time. Fisher transform can be used to convert correlation to a variable such as $Z \in [-\infty, \infty]$ – this is useful in stress-testing

$$Z = \frac{1}{2} \log \frac{1+\rho}{1-\rho}$$

the inverse transform is

$$\rho = \frac{e^{2Z} - 1}{e^{2Z} + 1}.$$

The key model risk is about copula fitting from the empirical data, particularly if data are credit spreads/default probabilities. We have discussed the importance of checking bi-variate plots (visually) for both $\mathbf{X}^{\text{Hist}}$ and $\mathbf{U}^{\text{Hist}}$. This check is for tricky non-monotonic relationship (eg, zig-zag, cross) between correlated variables which can't be picked up by even rank measures.

Histograms for each column (reference name) of $\mathbf{X}^{\text{Hist}}$ as well as kernel-smoothed $\mathbf{U}^{\text{Hist}}$ will show the empirical marginal distributions and effectiveness of kernel smoothing – each histogram of $\mathbf{U}^{\text{Hist}}$ must have uniform bars of similar height. Remember that any histogram presents buckets of observations subject to bandwidth h.

Monte Carlo Methodology (All Topics)

Improvement for Credit Spread Pricing

Please see the section on Averaging PL and DL. The number of simulations required for convergence (with naive Excel random numbers) can reach into 100,000s, particularly for the higher kth-to-default. Check sensibility of your pricing result (is it as expected, is its variance reducing) even if it *appears* to converge after 10,000s simulations.

The key reference for simulation design and sampling strategies, particularly when sampling from copula required, is a textbook on *Monte-Carlo Methods in Finance* by Peter Jaekel.

Improvement for HJM Simulations

With $k = 3 \ldots 5$ factors (volatility functions) used to simulate the curve, the dimensionality of Monte Carlo simulation is sufficiently low. Potentially, fractal patterns in random numbers across the depth of multiple dimensions can lead to pricing patterns that can be arbitraged.

There is no requirement to simulate the curve for up to 25 years into the 'future time' (with $\Delta t = 0.01$ that would be 2500 rows on the HJM Spreadsheet). You can start with a 5-year future period – that will give enough data for most caplet pricing illustrations.

However, for each simulated 'table' of forward rates one can observe that the same kind of simulated curve propagates into the future time. That raises an issue about importance of re-sampling when pricing by Monte Carlo.

[Insert 3D plots here]

Random Numbers Generation (for faster convergence)

There are following RN generation methods that you can adopt from ready implementations (eg, NAG Library) to give professional quality to the Monte Carlo within your project. Low discrepancy generators provide statistically dependent numbers with improved evenness in the multidimensional space.

1. Mersenne Twister (Pseudo RN)
2. Halton Numbers (low discrepancy, Quasi RN)
3. Sobol Numbers (low discrepancy, Quasi RN)

The convergence provided by low discrepancy RN generators is the order of $c(d)\frac{(\ln N)^d}{N}$ vs. $\frac{1}{\sqrt{N}}$ for quasi RN generators such as Excel's *RAND()* and Mersenne Twister. N refers to the number of simulations, d is number of dimensions (eg, $d = 3$ factors in HJM SDE and $d = 5$ reference names in Basket CDS), and $c(d)$ is some scaling function.

Example for HJM implementation: while Excel-generated random numbers might not give convergence even after $2,000$ simulations, Monte Carlo with proper low discrepancy RNs gives a good (low variance) estimate right after 200 simulations and satisfactory convergence after about 600 simulations. The second problem brought by this example is that **the use of Excel's pseudo random numbers over-estimated the price of a derivative.**

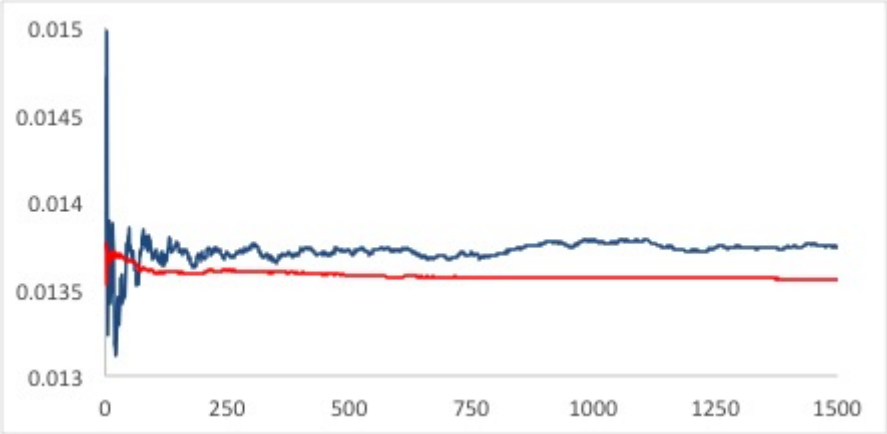

Figure 4: Dark blue line is a derivative price using *RAND()*, while red line result is for the same derivative obtained using low-discrepancy RNs.

Variance Reduction

Monte-Carlo simulation is evaluated by the speed (computational criterion) as well as variance of the estimator. That is, if you are estimating bond price or caplet price by Monte Carlo, you might would like to plot running standard deviation (after each 100 of simulations added).

The most important trick to reduce variance is to identify which simulated inputs increase the standard deviation of the estimate. For example in Basket CDS projects very small u imply default times τ_i which are less than one quarter. Because each default time is a stand-alone output (inter-arrival times are conditionally independent) it possible to remove those early defaults without breaking the continuity of Monte-Carlo. On the other hand, you will not be able to do the same thing when simulating a log-normal asset price (i.e., removing price levels at will).

Between Excel RAND() and Quasi RN, the variance of the estimator can vary threefold!

The note is an excerpt from the Q&A document on *Model Implementation and Robust Estimation* by Dr Richard Diamond, CQF.

Portfolio Construction. Robust Estimation

When designing optimisation, use different risk aversion and implement comparison to variance minimisation, using at least one alternative index of satisfaction (e.g., TR, SR, VaR, CVaR).

Sigma or BL Sigma for Optimisation?

Some sources (e.g., Idzorek) seem to use $\boldsymbol{\mu}_{BL}$ while discarding $\boldsymbol{\Sigma}^{\mu}_{BL}$ result. They proceed using the asset covariance matrix $\boldsymbol{\Sigma}$ in optimisation and risk metrics/index of satisfaction calculations (Sharpe Ratio). This is done to make these results compatible with other unconditional estimations and match dimensions as $\boldsymbol{\Sigma}^{\mu}_{BL} \propto \tau\boldsymbol{\Sigma}$ is the order of standard error $O(\tau)$.

However, the recommended computationally stable solution (Meucci 2010) calculates

$$\boldsymbol{\Sigma_{BL}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma^{\mu}_{BL}}$$

which is of the same order as $\boldsymbol{\Sigma}$. Note the difference to the original BL model.

This larger covariance $\boldsymbol{\Sigma_{BL}} \propto (1+\tau)\boldsymbol{\Sigma}$ reflects more uncertainty and results in increased risk metrics (decreased Sharpe Ratio). The behaviour is expected but the results are conditioned on particular views – that makes it difficult to compare to other, typically unconditional risk measures and allocations. By using the sample covariance matrix $\boldsymbol{\Sigma}$ together with $\boldsymbol{\mu}_{BL}$ in optimisation, the model implementation 'pretends' to obtain unconditional allocations.

Another problem with sample covariance matrices is reduced rank of the matrix – multicollinearity of two or more columns being linearly dependent. In that way, the returns of multiple assets are driven by one factor and there is likely to be cointegration. The practical implication for portfolio construction is that such matrices are non-invertible, and we require covariance matrix inverse to calculate allocations.

Tracking Error

Portfolio relative **tracking error** is difference in return (over each period) as compared to the benchmark, usually an MSCI index for a relative market. The error can also be estimated as a quadratic mean over the period.

The alternative name for the tracking error is **annualized tracking risk**. So it is measured as the standard deviation of **active returns**

$$a_i = r_i - b_i$$

which is the difference between the portfolio *monthly* returns r_i and benchmark monthly returns b_i, measured over the past 120 months.

$$\mathrm{TR}_j(nY) = \sqrt{\frac{12}{12n-1}\sum_{i=1}^{12n}(a_{i,j} - \langle a_{i,j}\rangle)^2}$$

where $\langle a_{i,j}\rangle$ is the *inner product* (scalar product) represented by *SUMPRODUCT()* function in Excel and $j = 12n, \ldots, N$. The expression sets up sampling of annualised tracking risk from the set of observations for 120 months.

Optimal allocations can be obtained to minimise the tracking error.

Historical Simulation (HistSim)

You can enhance estimation of risk measures by implementing Historically Simulated VaR (or other risk measure),

1. Bootstrap GARCH-adjusted historic residuals Z_t^* for a particularly volatile period/regime change period. Suggested scenarios: 2008-2009, two-year credit crunch period; 2011 - Feb 2012 for European debt crisis.
   Sampled from a volatile time period, the VaR estimate will be a stressed (SVaR).

2. Sample from this dataset (in a chosen random way, with resampling) to generate *a projection* – a series of future returns.

Applications are twofold, (1) use the series of future returns to calculate volatility and VaR and (2) take optimal allocations $\boldsymbol{w}^*$ and calculate simulated performance $\boldsymbol{w^*\widehat{\mu}}$ as well as portfolio risk $\sqrt{\boldsymbol{w^{*\prime}\widehat{\Sigma}w^*}}$. Also, compare $\widehat{\boldsymbol{\mu}}$ to equilibrium returns $\boldsymbol{\pi}$. Note: $\widehat{\boldsymbol{\mu}}, \widehat{\boldsymbol{\Sigma}}$ are sample estimates from simulated (projected) returns data.

Boostrapping standardised residuals (returns) using GARCH

For a number of applications, we can standardise empirical residuals (returns) using GARCH-filtered standard deviation.

Standardised with *historic residuals* Z_t can be obtained as below from $u_{t,Hist}$, empirical returns or excess returns.

$$Z_t = \frac{u_{t,Hist}}{\sqrt{\sigma^2_{t,GARCH}}}.$$

$$u_t^* = \sqrt{\sigma^2_{t,GARCH}}\, Z_t^*$$

$$\sigma^2_{t,GARCH} = \omega + \alpha u^2_{t-1} + \beta\sigma^2_{t-1}$$

Also notice, that we can simulate Normal random variables Z_t^* and use them to obtain fully simulated residuals.

ROBUST COVARIANCE AND VOLATILITY

The Workshop suggests improving the robustness of covariance by 'reconstruction' for the matrix using GARCH-filtered volatilities and correlations from **a.** a long-term sample and/or **b.** standardised residuals Z_t^* (divide each return figure by GARCH volatility as done above)

$$\mathbf{\Sigma}_{ij} = \sigma_i \sigma_j \rho_{ij}$$

Individual variance and correlation estimates should have the same scale (remember, sample size $\neq$ timescale).

In addition,

- There are numerous research papers and studies from equity quant teams that demonstrate **a strong linkage between the flow of news and volatility of a financial asset** (equity, bond). The news flow dominates a short-term volatility forecast, with disclaimer that some equities 'are exempt'. The models implementing this are EGARCH with variables for positive/negative news volume, etc.

As an alternative,

- Use **range-based methods to estimate volatility**, also known as rescaled range: Parkinson's adjustment, Yang & Zhang (2000) drift-independent volatility. The methods use a high-low range observed for a price during each period (ie, daily) and the volatility is inferred. The methods take into account what happens inbetween closing prices and are more efficient than the traditional standard deviation of periodic (daily) returns.

Factorisation is always a tool at hand,

- One might fold covariance estimation into a Multivariate GARCH that, in turn, can be simplified with independent factors coming from the Principal Component Analysis (PCA GARCH). Zero correlation between factors allows to calculate allocations under various regimes of correlation .

Correlation estimation is subject to a compounding effect, that leads to higher correlation for large samples. The larger your sample, the more chance there will be correlated pairs. Also, the larger your *sample period*, the more chance that things will be 'on average' similar and therefore, highly correlated. Example: Monte-Carlo experiments show that correlation calculated using every 100th observation is higher than using every 10th (same sample size for each group).

Financial Time Series Analysis

These notes give early discussion of causality analysis for Vector Autoregression and Equilibrium Correction. All the **new content**, including Backtesting, Cointegration Case in R (spot rates market data), as well as *Learning and Trusting Cointegration in Statistical Arbitrage* to be presented in the relevant lectures and the Project Workshop. The updated Vector Autoregression and Johansen MLE Procedure notes can be found in *TS Workings* document that comes with the Cointegration Lecture.

Granger Causality

In addition to a decomposition of feedback relationship between two variables, the test is useful to determine whether a variable can be treated as exogenous, and therefore, dropped from VAR(p) system. Here is the original paper by Granger (1969) – the proofs, illustrations and discussion cover the case of two variables. The useful feature of his proof is the decomposition of feedback relationship between two variables in arms (A to B and B to A).
`http://www.sonoma.edu/users/c/cuellar/econ411/Granger.pdf`

Algorithm

Actual implementation of Granger Causality in software packages (MATLAB open-source econometrics toolboxes) forms the full VAR(1) matrix, calculate the unrestricted regression, then drop a variable and calculate the restricted regression. For example, for one unrestricted regression for row 1 (Y1,t) there are four restricted regressions (removing Y1,t-1, Y2,t-1, Y3,t-1, Y4,t-1). Repeat for row 2 (Y2,t).

Y1,t-1 Y2,t-1 Y3,t-1 Y4, t-1
Y1,t
Y2,t
Y3,t
Y4,t

Conditional Statement

For a case of more than two variables, the system would require a modified proof and statement about joint impact and conditioning. But econometrics literature says that keeping other variables in restricted and unrestricted regression approximates the bi-variate test.

I suggest that when Granger causality was introduced, the understanding of statistical models (a regression) as conditional statements was yet in development. In my opinion, using additional variables increases the degree of conditionality and limits the test. In other words the effect of a variable is evaluated in presence of other variables this sounds as a positive quality but, a simple VAR system does not represent the dynamics between financial time series well. Therefore, increase in conditionality actually limits the model which could be useful in simpler formulation and times when there are clear trends in the market.

Examples

There is an example of the use of Granger causality by Economist at the link below. Controversies apart, the statistical testing and results presentation were carried out with care see Charts 3a and 3b. `http://www.economist.com/node/21554185`

Adding lags to Granger causality testing increases statistical significance of links already identified with VAR(1) but could also introduce weaker links that appear comparable to stronger links in their significance. If partial autocorrelations do not suggest an autoregressive dependence beyond lag 1 then there would be a little point testing further. Information Criteria tests for lag p usually indicate the same as autocorrelations.

High correlation across time series could also be a problem: Granger causality testing would show that two highly correlated variables X, Y 'cause a third variable without identifying whether X or Y has the stronger impact. Not only such construction can fall apart but if we consider residuals: removing X will not reduce residuals much (Y would still be a good predictor) and so we conclude that X does not matter; same goes for Y and we are in Catch 22.

Pairs Trading [Earlier Notes]

Instead of solely relying on confirming significance of equilibrium correction (the small term $(1 - alpha)e_{t-1}$, the trade design should explore the properties of mean-reversion in the spread by the fitting to the OU process. Beware of the regime changes in non-stationary series: a series can shift in response to shocks leading to the change in equilibrium level or dismantling of a cointegrating relationship altogether.

Spread trading applications of cointegration work well in the presence of the a common factor or term structure (i.e., inflation and rates).

1. Once a cointegrated relationship in the pair is identified, establish the level of equilibrium (intercept) and plot the mean-reverting spread.

2. It will be prudent to study the joint distribution of *the common stochastic process.* It should consist of *i.i.d.* random variables, however if they produce a skew it is a dangerous sign. If it does not follow the Normal distribution, other problems arise: high-peaked and thin-tailed distribution indicates difficulty to find entry point, whereas fat-tailed distribution produces more entry points but potentially low P&L.

3. For trade design, long waiting times mean cost of carry (financing of the leverage) and increase the risk of regime change in the co-integrating relationship. Estimate half-life time of a co-integrating relationship returning to its equilibrium and explore trading partial half life.