Certificate in Quantitative Finance
Final Project Brief

June 2014 Cohort (One Topic)

Instructions

Please note that Workshops will be held to discuss the project topics in detail. Before that, use the brief in order to select topics and earmark material for a revision.

This document outlines each available topic together with submission requirements for the project report. By-step instructions are provided to offer a structure; they are not exhaustive of what you can implement. The brief should be used together with preparatory reading and review of the relevant CQF lectures.

Exclusively for the use by current CQF delegates. No distribution.

1 Instructions

Assessment for CQF Module Six is carried out by the means of a programming project. The project is designed to test your understanding of mainstream models and numerical methods used in quantitative finance as well as your ability to implement and validate a model to the standard of a financial product.

To complete the project, you must implement **one** of the following extended topics and submit working code, together with a well-written report and originality declaration.

1. **Pricing Basket Credit Default Swap (sampling by copula)**
2. **Factors in Interest Rate Models (forward curve data)**
3. **Portfolio Construction using Black-Litterman Model (time series)**

The topics mostly rely on material up to and including Module Five on Credit Modeling. Therefore, it is possible and recommended to start working on projects and their implementation early. Introduction to advanced modelling techniques will be completed throughout Module Six.

1.1 Submission date is Monday, 12 January 2015

There is no extension to the Final Project. All late submissions will automatically be deferred to the next cohort.

- The report must be soft bound in thesis style. Normal length is 15-30 pages, ex. code.
- The software must be provided on a DVD in a plastic pocket attached to the inside back cover. A 'clip-it' USB flash drive is also acceptable when securely attached.

- One printed copy must be sent to Fitch Learning, 4 Chiswell Street, London EC1Y 4UP, United Kingdom. Fitch Learning will keep the submitted reports.
- The report must be postmarked or delivered by hand on the submission date at the latest. E-mailed submissions will not be accepted.

Projects without declaration or code are incomplete and will be returned.

All projects are checked for originality. The CQF faculty reserves an option to invite a delegate for *a viva voce* before the qualification can be awarded.

1.2 Project Code

- Traditionally, project code is written in VBA or C++. Depending on topics chosen, delegates also use Python with libraries, MATLAB with toolboxes or R's integration with Wolfram's Mathematica. Java and C# with .NET Framework are also acceptable.
  If you are using other coding environments please consult with a CQF Tutor in advance.

- Use of Excel spreadsheets or Mathematica is usually limited to data processing, visualisation as well as validation of interim and final results.

- The aim of the project is to enable you to code numerical methods and develop model prototypes in a production environment. Therefore, it is recommended that you minimise the use of Excel application functions. They are unstable and computationally inefficient.

- It is desirable that you re-code numerical methods that are central to your model. However, a balanced use of libraries is allowed. When to re-use ready code is ultimately your decision. When using ready functions, consider their limitations in the report.
  Note: submission of a project that is solely based on unchanged code that was provided in CQF lectures and textbooks would be a serious drawback.

- The code must be thoroughly tested and well-documented: each function must be described, and comments must be used. Provide instructions on how to run the code.

- It is up to delegates to develop their own test cases, sensibility checks, and validation. It is normal to observe irregularities when the model is implemented on real life data. If in doubt, reflect on this in your project report.

1.3 Project Report

It is a technical report with the main purpose to facilitate access to implementation of numerical methods (the code) and pricing results.

- The report must contain a sufficient description of the mathematical model, numerical methods and their properties. In-depth study is welcome but report should be relevant.

- Include notes on algorithms and code in your report, noting which numerical methods you have re-coded.

- Please give due attention and space for presentation and discussion of your pricing results. Use charts, test cases and comparison to research results where available.

- Mathematical sections of the report can be prepared using LaTeX, Microsoft Equation Editor (Word) or Mathematica.

The next sections outline requirements and implementation steps. A separate **Final Project Q&A** document discusses the most frequent issues encountered.

Fair Spread for Basket CDS

Summary

Price a fair spread for a Basket CDS with at least 5 reference names using the sampling from a copula algorithms. The input of the appropriate default correlation, preferably estimated from the historical CDS data, is required. The fair spread is calculated as an expectation over the joint distribution of default times. In order to convert simulated uniform random variables to default times, you will need their marginal distributions, defined empirically by hazard rates. The hazard rates are bootstrapped from traded credit spreads (CDS) under the assumption of being piecewise constant.

A successful project will implement sampling from both, Gaussian and Student's t copulae, pricing for 1st to 5th to default basket swaps as well as elaborate sensitivity analysis for the prices of these contracts. Key numerical aspect for the topic is the use of low discrepancy sequences (e.g., Halton, Sobol).

Data Requirements

- A snapshot on a particular day of credit curves to bootstrap hazard rates (separate term structure for each name). This is used to find default time, $u_i \to \tau_i$.
- *Historical* credit spreads data (liquid tenor, 5Y) to estimate inferred default correlation matrix. Convergence and robustness of spread pricing by t copula will depend on this historical data due to explicit MLE estimation. Data quality is defined by the degree with which estimated correlation reflects default correlation, as opposed to daily market fluctuations in credit spreads.

These are two **separate** datasets. Discounting curve data is also necessary for each; in this project, discount factors can be approximate.

Corporate credit spreads are not in the open access, you will need obtain them from a data provider like Bloomberg or Thomson Reuters via your firm or a colleague. Sovereign default probabilities (including historical series) have been made available by the DB research
`http://www.dbresearch.com/servlet/reweb2.ReWEB?rwnode=DBR_INTERNET_EN-PROD\$EM\&rwobj=CDS.calias\&rwsite=DBR_INTERNET_EN-PROD`

Step-by-Step Instructions

1. For each reference name, bootstrap implied default probabilities from quoted CDS and convert them to hazard rates.
2. Calibrate appropriate parameters for 'sampling from copula': correlation matrix and *df*.
3. For each simulation, repeat the following routine:

(a) Generate a vector of correlated uniform random variables – you will need to implement sampling from both Gaussian and Student's t copulæ separately.

(b) For each reference name, use its term structure of hazard rates in order to determine the year of default t_m and the year fraction δt (exact default time or quarterly/semi-annual accrual method).

(c) Based on τ_k for kth-to-default calculate the discounted values of premium and default legs.

4. Average premium and default legs across simulations separately. Calculate the fair spread.

Model Validation

- The fair spread for k-th to default Basket CDS should be less than k-1 to default. Why?
- Project Report on this topic should have a section on **Risk and Sensitivity Analysis** of the fair spread *w.r.t.*
  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.
- The most important parameter is correlation. Ensure that you explain the following:
  1. historical sampling of default correlation matrix, and
  2. choice of the stress-testing levels of correlation.

Make sure you discuss and compare sensitivities for all five instruments.

Reading List

- The starting source on *sampling from copula* algorithm is *Monte Carlo Methods in Finance* textbook by Peter Jaekel (2002). Please see Chapter 5, particularly pages 46-50.
- Most likely, you will need to re-visit *CDO Lecture* material, particularly about slides 48-52 that illustrate Elliptical copula densities and discuss factorisation.
- Rank correlation coefficients are introduced *Correlation Sensitivity and State Dependence* and Jaekel (2002) as above. Any good statistical textbook can serve as a reference.
- Bootstrapping of survival probabilities is covered in *Credit Default Swaps Lecture.*

Introduction to Copula

MathWorks have tutorials on copula fitting and applications – I can recommend one from their older CD (dated 2005). Please search for *Market Risk using GARCH, EVT and Copula.*

Factors in Interest Rate Models

Summary

Focus on factorisation of the yield curve and numerical techniques of Principal Component Analysis. Identify and attribute factors that determine evolution of the forward curve. Choose a sampling and differencing approach that contributes to the robustness of covariance matrix of interest rate changes (differences). The sampling and factorisation efforts will ensure robust calibration of volatility functions and acceptable results from the HJM SDE simulations.

A successful project will provide **a.** calibration of volatility functions from your original sample and **b.** convergent pricing of bonds, caps and floors supported by charts and analysis of Monte-Carlo efficiency. Option pricing examples must offer a view across a range of strikes and tenors. Report on this topic should concentrate on numerical techniques and volatility fitting rather than generic HJM/LMM models.

[**Optional**] The advanced option for delegates with interest rates experience would be pricing vanilla and Bermudan swaptions using *a market model*.[1] While HJM calibration means fitting volatility from the historic data of interest rate changes, LMM utilizes market data of implied volatility and therefore is deemed to be forward-looking. Both models rest on the no-arbitrage argument: the drift cannot be random and depends on volatility of forward rates.

Data Monte-Carlo simulation within the HJM framework relies on consistent forward curve representation by the instantaneous forward rates data. A discretised market model such as the LMM requires caplet data for simple volatility bootstrapping or swaption data for Rebonato Method (calibration by optimisation), which is usually limited to trading desks and interdealer brokers. However, snapshots of caplet data can be found in research papers or obtained from market data providers.

Discounting is an intimate part of IR derivatives pricing because they are decomposable into cashflows. Usage of OIS or SONIA rates to construct discount factors is compatible with the market model that relies on shifting numerarie, i.e., OIS is an equivalent of money market/bank account. Discounting with unsecured overnight borrowing rates generates a higher cost of funding than is practised on collateralised assets and repo transactions.

Other approaches to discounting can be categorised into (a) risk-free with a model-generated spot rate $r(t)$ and (b) funding premium with the customised curve built for discounting of specific cashflows (e.g., debt or multi-currency swaps). Construction of customized yield curves for discounting would require an independent study of interpolation methods (see Reading List).

[1] The Brief provides guidance for the LMM however, the arbitrage-free SABR can also be used; it offers explicit analytical solutions for volatility fitting.

Data Requirements

- (HJM) Forward rates data for Pound Sterling Commercial Bank Liability Curve (the equivalent *short-term* data are LIBOR rates) is readily available from the Bank of England `http://www.bankofengland.co.uk/statistics/Pages/yieldcurve/archive.aspx`

  ECB provides comparative instantaneous forward rate data for the Euro area yield curve `http://www.ecb.europa.eu/stats/money/yc/html/index.en.html` (Government Liability)

  Sampling Covariance matrix is estimated on historical rates data that includes the short and long ends of the curve. Daily frequency over two-year period is typical choice but you can experiment with other frequencies (weekly) and periods. Statistically, estimating covariance matrix on *iid rate differences* is equivalent to working with covariance matrix of residuals–the modelling focuses on innovations (shocks) to interest rates.

- (LMM) Caplet or swaption data is usually maintained by trading desks and interdealer brokers. Data for certain markets is available from Thomson Reuters and Bloomberg.

- Discounting that reflects the cost of funding is preferable but requires an extra source of data for discounting curve (i.e., OIS) or curve construction. The simplified alternative is to calculate discount factors from the simulated risk-neutral spot rate process $r(t)$ as done for the HJM model.

It useful to look up how 'instantaneous forward rates' are derived and refresh rates notation maths. Methodology for stripping of the yield curves from the underlying traded instruments and bootstrapping forward curves is covered by BIS and BOE documentation.

Building Curves. OIS Discounting

Below is the list of CQF Extras on IR derivatives and discounting curves. 'OIS Discounting' term refers to the curve use *and* construction. Though 'construction' part is under-explained in vendor presentations – fall back to BIS documents and Hagan&West on interpolation methods.

- *Valuation framework for interest rate derivatives in today's Libor world* - Wojciech Slusarski **with Excel examples**

- *OIS Discounting: General Background* - Sol Steinberg

- *OIS and its Impact on Modelling, Calibration and Funding of OTC Derivatives* - Satyam Kancharla

Bloomberg Curves Toolkit focuses on stripping data and fitting the curves. A tutorial available through the platform's Excel Add-in **XCTK** with links to white papers.
One can also explore Matlab resources and case studies
`http://www.mathworks.co.uk/discovery/yield-curve.html`

Step-by-Step Instructions - HJM Model (Yield Curve Data)

Part I: Data and Volatility Estimation

1. Obtain time series data on evolution of instantaneous forward rates.
   - Include both, short and long ends of a yield curve.
   - Even if you model a market other than sterling, check out data format on the Bank of England website (Yield Curve Data section).
2. Convert interest rates data into differences, group daily observations by tenors and calculate a covariance matrix.
   - This is 'a quest for invariance' as we effectively analyse the covariance matrix of innovation process.

Part II: Calibration (Factorisation)

3. The required numerical techniques *each can be improved* as compared to Excel demonstrations for Lecture 4.4 on HJM.
   (a) Conduct Principal Component Analysis on the covariance matrix. Use the sum of eigenvalues to evaluate contribution of components (factors) and make a decision about their number.
   (b) Convert principal components to volatility functions using $\sqrt{\lambda_i}$ and apply curve-fitting techniques, such as a power series regression, to each volatility function.
      - *Avoid over-fitting* with more factors than necessary.
   (c) Calculation of the drift requires numerical integration over a fitted curve, comprised of volatility functions.

Part III: Pricing by Monte-Carlo

4. Implementation of PCA, curve-fitting, and numerical integration will enable you to conduct Monte-Carlo. To achieve price convergence it might be necessary to conduct $> 10,000$ simulations.
   - For each round of simulation, generate Normal random vectors to be used as factors.
   - An improvement can be made by using quasi-random (low latency) number generators.
5. Price interest rate derivatives, also *on each simulation*
   - ZCBs, caps/floors with strikes and maturities of your choice.
   - Pricing examples must offer a view on the volatility simile (across a range of strikes and/or maturities).
6. The fair prices for derivatives are obtained by averaging the results across MC simulations. The cumulative average should demonstrate price convergence.

Step-by-Step Instructions - LIBOR Market Model (Rate Options Data)

As you will notice in the *market* model, we start with interest rates derivatives and data on implied volatility. This part is an advanced and **optional choice**, not a requirement.

Part I: Data

1. You will need market prices data of cap and floor options (simply referred to as 'caps'). The data can be two columns (series) of cap prices and corresponding discount factors. Caps are quoted in terms of implied volatility σ^{cap} (notation is $\sigma_i^{cap}\sqrt{T}$ or ζ_i for a caplet).
   (a) Black formula is conventional means of converting the cashflow of a cap (caplet) into an implied volatility figure.
2. The second set of data to which model fitting can be done is swaptions, for which the deliverable asset is an interest rate swap.

Part II: Calibration (Volatility Stripping)

3. Stripping caplet volatilities $\sigma_i^{cap}\sqrt{T}$ or ζ_i as LMM requires more discretisation than cap prices traded in one-year increment.
   Algorithm involves calibration of strikes as forward swap rates $S(t; T_{i-1}, T_i)$.
4. Equally, volatilities can be calibrated from vanilla swaptions as they are options and give another source of implied volatility υ_i (Rebonato method makes Black(1976) suitable).
5. Fitting the *abcd* instantaneous volatility function $\sigma(\tau)$ defined for each tenor as $\sigma_i(t)$. Coefficients *abcd* are estimated by optimisation that can be joint *wrt* caplet implied volatilities σ_i^{cap} and swaption implied volatilities υ_i. Goal is to minimise the squared differences between two implied volatilities (for the same tenor enumerated i).
   (a) Correlation structure ρ_{ij} is defined parametrically.

In a sense, pricing of swaptions has already been done in the process of calibration (stripping) of caplet volatilities because forward swap rates $S(t; T_{i-1}, T_i)$ has been calculated. Pricing of path-dependent options, such as Bermudans that give exercise flexibility on some or all payment dates $T_i, T_{i+1}, ..., T_m$, would require the modified Least Squares Monte-Carlo simulation.

Reading List

- Start with a methodology paper from Bank of England that explains forward rates: *Inferring market interest rate expectations from money market rates.*
- Sources covering curve construction: *Interpolation Methods for Curve Construction* by Pat Hagan and Graeme West and *Zero-coupon Yield Curves: Technical Documentation*, BIS.
- The source to start on the LMM is *The LIBOR Market Model in Practice* by Gatarek, et al. (2006) available from Wiley Online Library. (Textbook not supplied with core CQF.)

Portfolio Construction with Time Series Extensions

Summary

Construct a portfolio of diverse assets using large historic time series and estimate allocations for multiple levels of risk aversion together with the key risk measures such as VaR, Expected Shortfall, Sharpe or other information ratios. The measures often used to quantify 'diversification'. The key point of the Black-Litterman model is introduction of the analysts' recommendations into portfolio construction. You will have to identify sensible recommendations and formalise them in form of constraints suitable for BL estimation.

The effective implementation will (a) improve on the sample covariance matrix estimator and (b) carefully diagnose and treat the optimisation problem, choosing from mean-variance, mean-TE, or mean-ES optimisation modes.[2] It is also necessary to study the robustness of allocations (i.e., what makes the weights to change) under different assumptions and constraints *ceteris paribus*; this study needs to be detailed and well-illustrated.

A successful project will have matrix form calculations and other numerical techniques coded (rather than spreadsheet calculations), robust optimisation results for different levels of risk aversion, and extensive graphic presentation. **Note:** a naive mean-variance optimisation on sample mean and covariance is of little value. Recommended to attempt time series analysis, such as variance estimation with GARCH models, returns projection, or cointegration analysis.

Data Requirements (Portfolio Design)

The objective is to come up with either (a) a multi-asset diversified portfolio or (b) a 'specialised' portfolio that focuses on an industry, emerging market(s), credit, etc; another example would be a managed futures portfolio. The first kind of portfolio can have under 15 diverse assets and sub-assets, while the specialised portfolio usually includes more.[3]

- Multi-asset includes equity, fixed income, credit and volatility. Commodities, real estate and other assets are optional. ETFs can be used to represent the asset classes for which the data is hard to obtain (e.g., bonds, futures).
- Replication of broad equity indices is convenient.[4] However, the multi-asset approach comes from the need to account for key factors that drive performance of names selected. It is possible to include a factor into optimisation as if it were a traded asset.
- Mean-variance optimisation is specified for *excess linear returns*. The technical challenges, such as the risk-free rate changing over time can be dealt using simplified assumptions.

[2] TE stands for Tracking Error and ES stands for Expected Shortfall mathematically known as CVaR.

[3] If you follow portfolio compiled in others' study please provide a reference.

[4] Because index weights can be converted into equilibrium weights, alternatively market cap approach can be used.

The minimum historical time period is 2-3 years (for daily returns) though you might use the shorter periods of 1-3 months for variance estimation; that would require robust estimation and exclusion of stressed periods. Portfolios that are tradeable strategies themselves might require the higher frequency data (1-10min). A starting source for historical daily close prices of US equities and ETFs is Yahoo!Finance.

Step-by-Step Instructions

Part I: The Black-Litterman Model

1. Construct the prior (reference distribution): equilibrium returns can come from a benchmark index, while covariance is estimated from historical data. Improve robustness.

2. Define input views of both kinds, relative and absolute.

3. Estimate the posterior distribution of excess returns using the Black-Litterman formulae.

Part II: Robust Allocation

5. Choose at least one-two more optimisation designs in addition to variance minimisation. Formulate reasonable optimisation constraints (if necessary in addition to views), for example 'no short positions in bonds'.

6. Obtain allocations for three levels of risk aversion.

7. Study robustness of allocations (explore optimisation behaviour) and check for common pitfalls such as 'corner solutions'. Can you pinpoint a few factors that drive allocations?

Part III: Time Series Analysis [Two suggestions]

8. Stabilise covariance matrix by GARCH model before conducting BL optimisation. Alternatively, you can test the performance and risk of optimised allocations on GARCH-simulated returns (forward projection).

9. Cointegration analysis can be used to identify candidates for the long-short and managed futures portfolios. An alternative is to focus on pair trade design, quality of mean-reversion (OU process fit), and properties of P&L.

Reading List

- CQF Lecture on *Fundamentals of Optimization and Application to Portfolio Selection*

- *A Step-by-step Guide to The Black-Litterman Model* (Incorporating user-specified confidence levels). Thomas Idzorek, 2002

- *The Black-Litterman Approach: Original Model and Extensions* Attilio Meucci, 2010. `http://ssrn.com/abstract=1117574`

- `http://blacklitterman.org/`