CQF Final Project Topics:
Time Series, Interest Rates

June 2016 Cohort – Workshop II

Implement ONE topic plus CVA Component.

Topic and Design Choice

- Start collecting data and planning how to complete the project.

- It is up to you to source and clean the suitable input data.

  We will review the data necessary for each topic, further description and links are in the Project Brief. Any suitable alternative data can be used, including pre-simulated data.

- Model validation (test cases, sensibility checks) is part of the task assigned.

  We will cover suggestions in these slides as well as Project Q&A.

  - In particular, set your option strikes and maturities, missing forward rates, CDS spreads.

Topic and Design Choice - Numerical Techniques

- Implementation of numerical techniques from the first principles is the purpose. Pricing of a derivative/credit instrument/robust allocation is the result.
  - In particular, make a choice of methods of curve fitting, numerical integration, and random numbers generation.
- Refer to the relevant CQF Lectures and do extra reading on pricing methodologies, quant finance models and numerical techniques.

Project Report

- A full **mathematical description** of the models employed as well as any numerical methods. Include mathematical measures of *accuracy and convergence*.
- Results presented using **a plenty of tables and figures**, together with sensible interpretation.
- **Pros and cons** of a particular model and its implementation, together with possible improvements.
- **Demonstrate 'the specials'** of your implementation: own research, re-coding of numerical methods, using the industrial-strength libraries of C++, Python or VBA + NAG.
- Instructions on how to use software if that is not obvious. The code must be thoroughly tested and well-documented.

Time Series Analysis and Backtesting

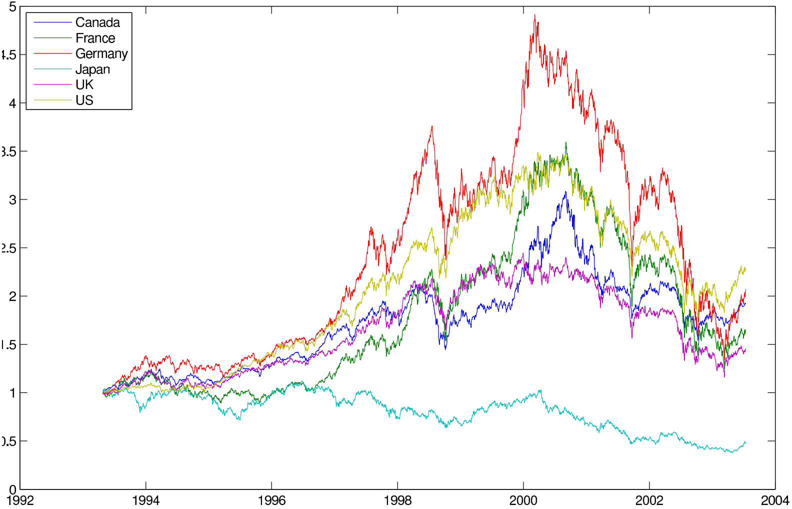
Introduction

We will use **returns** from six broad market indices to demonstrate how to implement *Vector Autoregression*.

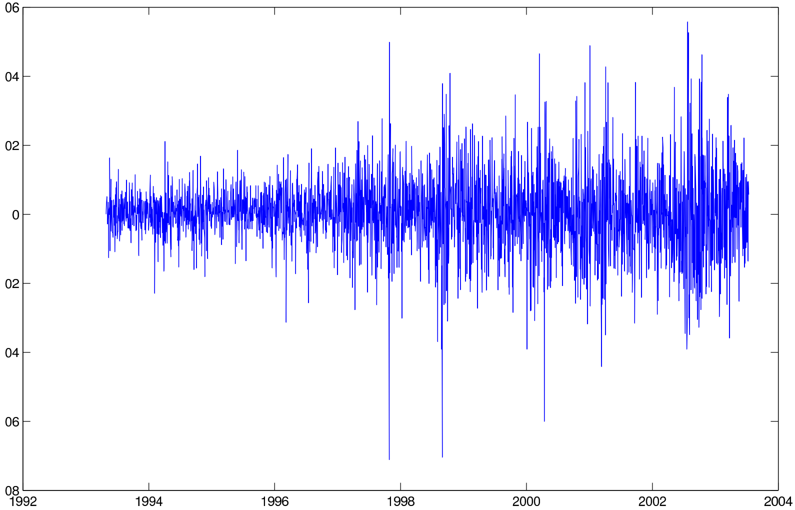
- Japan, Germany, France, UK, US, Canada

You can extend the regression with exogenous variables, such as rates, commodities, VIX, even derivatives but those have to be *returns* data.

Relative Equity Indices



US Daily Index Returns



Linear Model

For these stationary returns, we can set up a system of endogenous variables that depend only on their past (lagged) values.

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

Think about forecasting powers of this model-free set up.

Empirical observation: the vector autoregression is **not good** at predicting daily returns.

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

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

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

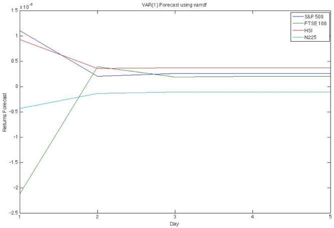
Forecasting Accuracy

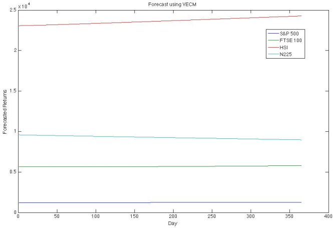
R command *accuracy()* returns the common measures of forecasting accuracy.

ME	RMSE	MAE	MPE	MAPE	MASE
0.00120	1.0514	0.8059	-Inf	Inf	0.79298

The mean error, root mean squared error, mean absolute error, mean percentage error, mean absolute percentage error, and mean absolute scaled error.

Forecasting with regression





The useful part of a forecast does not extend beyond 2-3 days.

$$\mathbb{E}^{VECM}[Y_{T+i}] = \mathbb{E}^{VAR}[Y_{T+i}] + [\Delta Y_T]$$

Adding a small correction, if cointegrated relationship exists, only produces a linear trend $\Delta Y_T =$ constant.

What else do we do?

- Econometric studies for stationary data of weekly or monthly changes are **a.** forecasting, **b.** impulse response analysis, and **c.** Granger causality analysis.

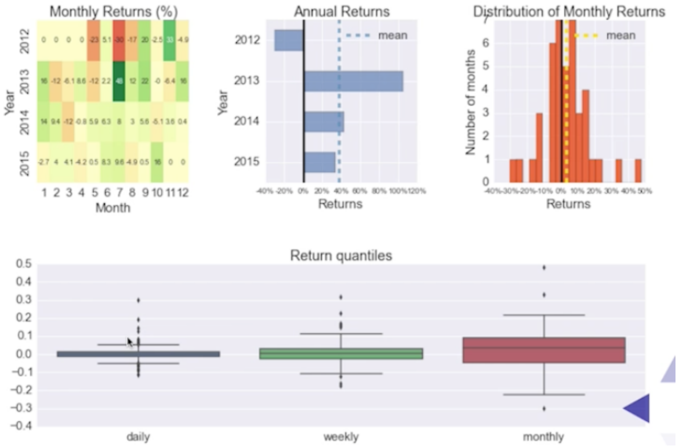
- **Correlation studies** – see experiments with rank vs. linear correlation in Credit Topic.

  Often, before using Cholesky and decomposition a fix is required by estimating the nearest correlation matrix.

- **Covariance matrix issues** – collinearity is common and prevents matrix inversion for portfolio allocation purposes.

  Dimension reduction is often required for large covariance matrices but they are noisy and PCA not as effective as we observed for interest rate changes (HJM Calibration).

Performance Explorations



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

Vector Autoregression

It is a multivariate regression with past values.
VAR(p) is the simplest way of structural equation modelling.

It models a system of endogenous variables that depend only on their past (lagged) values.

$$Y_t = C + A_1\, Y_{t-1} + \cdots + A_{t-p}\, Y_{t-p} + \epsilon_t \qquad (1)$$

$Y_t = (y_{1,t}, \cdots, y_{n,t})'$ is a column vector $Nvar \times 1$.
A_p is a $n \times n$ matrix of coefficients for lagged variables Y_{t-1} ... Y_{t-p}

$$A_p = \begin{bmatrix} a^p_{11} & \cdots & a^p_{1n} \\ \vdots & \ddots & \vdots \\ a^p_{n1} & \cdots & a^p_{nn} \end{bmatrix}$$

Vector Autoregression: Estimation

Although VAR(p) can be exceedingly large, it is a system of **seemingly unrelated regressions** that can be estimated separately (line by line) using *ordinary least squares* (OLS).

Matrix manipulation is available in many packages (Matlab, R, Python) allowing to specify a concise form and estimate all lines of Vector Autoregression **in one go**.

Dependent Matrix

First, we need to form the matrices as follows, with $T = N_{obs}$:

1. *Dependent data matrix* has *observations* for the first p lags removed (here, observations are in rows from time $p+1$ to most recent observation at T)

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

$$\begin{bmatrix} \cancel{y_{1,t=1}} & \cancel{y_{1,\dots}} & \cancel{y_{1,p}} & y_{1,p+1} & y_{1,p+2} \cdots & y_{1,t=T} \end{bmatrix}$$

refers to all historic observations of the variable y_1.

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

$$Z = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{y}_p & \mathbf{y}_{p+1} & \cdots & \mathbf{y}_{T-1} \\ \mathbf{y}_{p-1} & \mathbf{y}_p & \cdots & \mathbf{y}_{T-2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{y}_{n,1} & \mathbf{y}_{n,2} & \cdots & \mathbf{y}_{n,T-p} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_{1,p} & y_{1,p+1} & \cdots & y_{1,T-1} \\ y_{2,p} & y_{2,p+1} & \cdots & y_{2,T-1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n=Nvar,p} & y_{n,p+1} & \cdots & y_{n,T-1} \\ & & & \\ y_{1,p-1} & y_{1,p} & \cdots & y_{1,T-2} \\ y_{2,p-1} & y_{2,p} & \cdots & y_{2,T-2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n=Nvar,p-1} & y_{n,p} & \cdots & y_{n,T-2} \\ & & & \\ y_{1,1} & y_{1,2} & \cdots & y_{1,T-p} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,T-p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n=Nvar,1} & y_{n,2} & \cdots & y_{n,T-p} \end{bmatrix}$$

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

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

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

Residuals

4. Disturbance matrix (innovations, residuals)

$$\epsilon = \begin{bmatrix} \epsilon_{p+1} & \epsilon_{p+2} & \cdots & \epsilon_T \end{bmatrix} = \begin{bmatrix} e_{1,p+1} & e_{1,p+2} & \cdots & e_{1,T} \\ e_{2,p+1} & e_{2,p+2} & \cdots & e_{2,T} \\ \vdots & \ldots & \ddots & \vdots \\ e_{n,p+1} & e_{n,p+2} & \cdots & e_{n,T} \end{bmatrix}$$

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

5. *Coefficient matrix* includes the intercept C

$$B = \begin{bmatrix} C & A_1 & A_2 & \cdots & A_p \end{bmatrix}$$

Calculating VAR(p) Estimates

Given our matrix specifications, VAR(p) system can be written as

$$Y = BZ + \epsilon$$

- Calculate the multivariate OLS estimator for *regression coefficients matrix* B as

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

  This estimator is consistent and asymptotically efficient.

- Back out regression residuals

  $$\hat{\epsilon} = Y - \hat{B}Z$$

A bit of MLE

Consider the multivariate Normal Likelihood function and its log

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

To maximise the Log-Likelihood *by varying* β

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

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

Residuals and Parameters Significance

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

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

2. Covariance matrix of regression coefficients

$$Cov\left(Vec(\hat{B})\right) = (ZZ')^{-1} \otimes \hat{\Sigma} \quad = \quad I^{-1}$$

Vec denotes *vectorization* and $\otimes$ is the *Kronecker product*.

Standard errors of regression coefficients will be along the diagonal. Useful to calculate t-statistic.

Example: VAR(1) Estimation

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

Example: Residual Covariance Matrix

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

As our market index data was standardised to start at 1, the covariance matrix is the same as correlation matrix.

Optimal Lag Selection

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

- *Akaike Information Criterion*

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

- *Bayesian Information Criterion* (also Schwarz Criterion)

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

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

Example: Optimal Lag Selection

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

Stability Condition

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

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

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

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

Augmented DF Test for Unit Root

To improve the Dickey-Fuller procedure, lagged differences Δy_{t-k} 'augment' the test, improving robustness *wrt* serial correlation

$$\Delta y_t = \phi y_{t-1} + \sum_{k=1}^{p} \phi_i \Delta y_{t-k} + \epsilon_t$$

- Insignificant ϕ means unit root for series y_t.

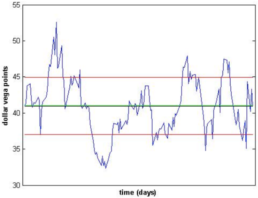
$$\phi = \beta - 1 = 0 \qquad \Rightarrow \qquad \beta = 1$$

- The critical value is taken from the empirically tabulated Dickey-Fuller distribution.

Cointegration Analysis and Estimation

Cointegrated System

A linear combination $\beta'_{Coint} \boldsymbol{Y_t} = e_t$ must generate cointegrated residual (spread) as below $e_t \sim I(0)$



To obtain the spread, allocation is done in β_{Coint} as weights.

There is a cancellation of <u>a common stochastic process</u> in each of $y_{i,t}$

$$\begin{aligned} e_t &= y_{1,t} \pm \beta_2 y_{2,t} \pm \cdots \pm \beta_n y_{n,t} \\ \beta_{Coint} &= [1, \pm\beta_2, \ldots, \pm\beta_n] \end{aligned}$$

Common Factor (Cont)

Assume we have two bond prices that are cointegrated, so

$$Z(t;\tau_1) - \beta_C Z(t;\tau_2) = e_t \qquad \text{stationary } I(0)$$

- **Cointegrated factor** e_t is different from **stochastic factor** dX. It works with the slow speed of correction $(1-\alpha)$

$$\Delta Z_t = \phi \Delta Z_{t-1} + (1-\alpha)\left[Z(\tau_1) - \beta_C Z(\tau_2) - \mu_e\right]_{t-1}$$

- The level of equilibrium $\mathbb{E}[e_{t-1}] = \mu_e$ gives a risk factor: a parallel shift of the yield curve.

Estimating Cointegration - Pairwise

- **Pairwise Estimation**: select two candidate time series and apply ADF test for stationarity to the joint residual. Use the estimated residual to continue with the Engle-Granger procedure.

See CQF Lecture on Cointegration and Cointegration Case B with the case study in R that re-implements ECM explicitly.

Engle-Granger Procedure

Step 1. Obtain the fitted residual $\widehat{e}_t = y_t - \widehat{b}x_t - \widehat{a}$ and test for unit root.

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

Step 2. Plug the residual from Step 1 into the ECM equation and estimate parameters ϕ, α

$$\Delta y_t = \phi \Delta x_t - (1 - \alpha)\widehat{e}_{t-1}.$$

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

Estimating Cointegration - Multivariate

To understand **multivariate cointegration** you need to know:

- Principle of reduced rank of a matrix
- Johansen trace and max eigenvalue tests
- Choice of deterministic trends in Johansen tests

- Johansen MLE Procedure estimates $\mathbf{\Pi} = \alpha\beta'$ and provides inference about the number of cointegrating relationships r (rank of cointegration).

Multivariate Equilibrium Correction

For non-stationary variables, such as prices, $\mathbf{Y_t}$ we specify equilibrium correction equations

$$
\begin{aligned}
\Delta \boldsymbol{Y}_t &= \boldsymbol{\Pi}\, \boldsymbol{Y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{Y}_{t-1} + \epsilon_t \\
\Delta \boldsymbol{Y}_t &= \alpha\beta'\, \boldsymbol{Y}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{Y}_{t-1} + \epsilon_t \\
\Delta \boldsymbol{Y}_t &= \alpha(\beta'\, \boldsymbol{Y}_{t-1} + \mu_e) + \boldsymbol{\Gamma}_1 \Delta \boldsymbol{Y}_{t-1} + \epsilon_t
\end{aligned}
$$

The last case is a restricted constant (deterministic trend in $\Delta \boldsymbol{Y}_t$).

Beware! Multivariate ECM notation is different from Engle-Granger.

Multivariate Equilibrium Correction

The matrix $\mathbf{\Pi}$ **must have a reduced rank**, otherwise the stationary $\Delta\mathbf{Y}_t$ will be 'equal' to non-stationary $\mathbf{\Pi Y}_{t-1}$. It's factorised as

$$
\begin{aligned}
\mathbf{\Pi} &= \boldsymbol{\alpha}\boldsymbol{\beta}' \\
(n \times n) &= (n \times r) \times (r \times n)
\end{aligned}
$$

r columns of $\boldsymbol{\beta}$ are cointegrating vectors, and $n - r$ columns are common stochastic trends (unit roots) of the system.

Next, we test for the rank of $\mathbf{\Pi}$ that determines the number of cointegrating vectors.

Johansen Test for Cointegration Rank

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

Comparing Trace Statistic to its Critical Value, we reject $r = 0$ but can't reject $r \leq 1$. That suggests **only one** cointegrated relationship.

Cointegrating Vector Estimators β'_{Coint}

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

Take the first column and standardise it.

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

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

Planners vs. Hedgers

'The social planner' approach, implicit in regression forecasting, is dangerous with markets – what would you do if the next price step is not according to a forecast?

If equilibrium condition changes $\mu_e^{Old} \rightarrow \mu_e^{New}$, cointegrated models give correction in the precisely opposite direction

$$\Delta \mathbf{Y_t} = \mu_{0,\tau} + \alpha \left[\beta' \mathbf{Y}_{t-1} - \mu_{\mathbf{e}} \right] + \epsilon_{t,\tau}$$

- Cointegration analysis helps to understand how the common factor drives many prices. However, decoupling a set of cointegrated prices into forecasting equations does not work.
- Instead, one would devise a trade around the mean-reverting spread.

Statistical Arbitrage with Cointegration

Statistical Arbitrage

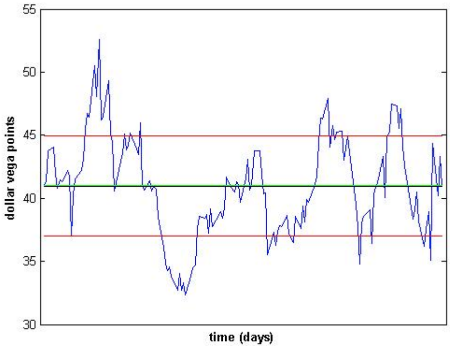
Cointegrated prices generate a mean-reverting spread. It is possible to enter systematic trades that generate P&L.

1. **How to generate P&L?** – Design a trade and evaluate profitability.
2. **How to evaluate P&L?** – Drawdown control and backtesting.

For systematic trading, you will need specifics:

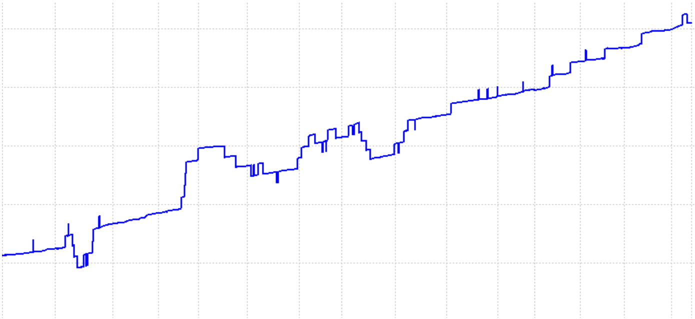
- loadings β_C give positions $\beta_C' \boldsymbol{P_t}$,
- (optimised) bounds give entry/exit, and
- speed of reversion gives idea of profitability over time.

Mean-Reverting Spread



- Using a cointegrated relationship among Volatility Futures, we obtained a mean-reverting spread $\theta \gg 0$.
- The bounds are calculated by fitting to **the OU process**.

P&L from a spread trade



The P&L was achieved by trading around the spread $e_t = \beta'_C \boldsymbol{P_t}$. Positions were taken as

$$\beta_{1,C}\, P_{1,t} + \beta_{2,C}\, P_{2,t} + ... + \beta_{n,C}\, P_{n,t}$$

Pairs Trading P&L

For the **price levels** P of two assets, stock S and market index M, we have a cointegrated residual

$$e_t = P_t^S - \beta_C P_t^M - \mu_e \qquad \mathbb{E}[e_t] = 0$$

e_t is a mean-reverting time series with the mean μ_e.

The dollar MtM P&L $\Delta e_t = e_t - e_{t-1}$ does not depend on the level μ_e

$$\begin{aligned}
\Delta e_t &= (P_t^S - \beta_C P_t^M - \mu_e) - (P_{t-1}^S - \beta_C P_{t-1}^M - \mu_e) \\
&= (P_t^S - P_{t-1}^S) - \beta_C (P_t^M - P_{t-1}^M) \\
&= \Delta P_t^S - \beta_C \Delta P_t^M
\end{aligned}$$

Pairs Trading P&L as Factor

$$\frac{\Delta e_t}{e_{t-1}} = \frac{\Delta P_t^S - \beta_C \Delta P_t^M}{P_{t-1}^S - \beta_C P_{t-1}^M} = R_t^C$$

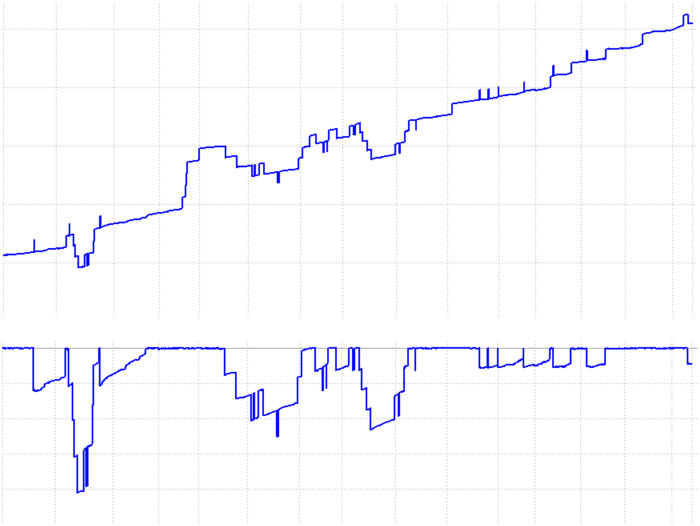
The series of returns R_t^C is a technical presentation of **the factor**.

$$\text{Corr}[R_t^C, R_t^M]$$

Experiment: Correlate returns from a strategy with another factor.

For example, returns from a pairs trading P&L vs. market index returns. What impact does it make to have the strategy returns of frequency higher than daily (1Min, 10Min)?

P&L vs Drawdowns



To set up an arbitrage trade, one requires the following items of information:

1. **Weights β'_{Coint}** to obtain the spread as

$$\beta_{1,C}P_{1,t} + \beta_{2,C}P_{2,t} + \ldots + \beta_{n,C}P_{n,t}$$

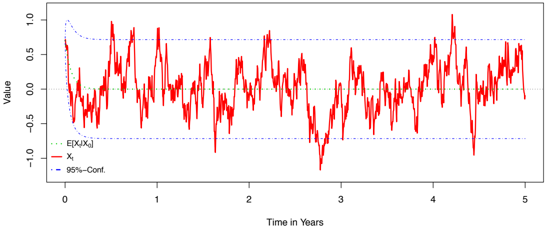
2. **Speed of mean-reversion** in the spread θ, which can be converted into half life (expected position holding time) as

$$\widetilde{\tau} \propto \ln 2/\theta$$

3. **Entry and exit signals** defined by σ_{eq}. Optimisation involved.

$$e_t \qquad \text{crosses} \qquad \mu_e + \sigma_{eq}$$

OU Process



We consider the process because it generates **mean-reversion.**

$$dY_t = -\theta(Y_t - \mu_e)\,dt + \sigma\,dX_t \qquad (2)$$

- θ is the speed of reversion
- μ_e is equilibrium level
- σ is the scatter of diffusion

Evaluating mean-reversion

OU SDE solution for $e_{t+\tau}$ has mean-reverting and autoregressive terms

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

Estimate a simple regression as follows:

$$e_t = C + Be_{t-1} + \epsilon_{t,\tau}$$

$$e^{-\theta\tau} = B \quad \Rightarrow \quad \boxed{\theta = -\frac{\ln B}{\tau}} \tag{3}$$

$$\left(1 - e^{-\theta\tau}\right)\mu_e = C \quad \Rightarrow \quad \boxed{\mu_e = \frac{C}{1 - B}} \tag{4}$$

Bounds of reversion

The scatter of the OU process relates to the total variance of cointegrating residual e_t (where τ is data frequency)

$$\sigma_{OU} = \sqrt{\frac{2\,\theta}{1 - e^{-2\theta\tau}}\,\mathbb{V}ar[e_{t,\tau}]} \tag{5}$$

σ_{OU} is diffusion *over small time scale* (volatility coming from small ups and downs of BM). **But**, we are interested in reversion from/to μ_e.

To plot trading bounds we use

$$\sigma_{eq} \approx \sigma_{OU}/\sqrt{2\theta}.$$

for potential entry/exit signals $\mu_e \pm \sigma_{eq}$.

Model Risk

There are formal statistical tests and procedures to estimate cointegraiton (eg, Engle-Granger, Johansen). But, the **presupposition** is that you take those tools and uncover some existential equilibrium.

The **reality** is that you are constructing and imposing such relationship as much as you are testing for it.

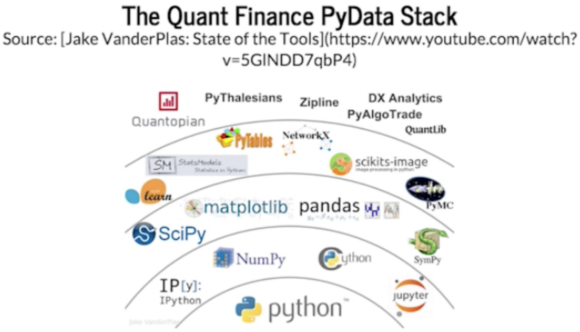
- Take *Cointegration Case B* among spot rates. We tested for cointegration between short end and long end, recent data and data collected over the long-term.

All project designs (whether learning-level or in-depth) should include backtesting of a strategy. Stat arb strategy is realised by using cointegrating coefficients β_C as allocations $\boldsymbol{w}$.

That creates a long-short portfolio that generates a mean-reverting spread (cointegrated residual) e_t.

The portfolio also explicates a factor.

Backtesting



`https://github.com/quantopian/pyfolio`

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

Preview

1. We will look at **how to relate P&L** to the market and factors (to understand what drives it, what you make money on).
2. Then, we will talk about **evaluating P&L** (through Drawdown Control).
3. You can look for suitable models for algorithmic **order flow** and liquidity impact. [Optional]

Please refer to the Algotrading Elective and QI Quantopian talk for illustrations and cases.

Alpha and Beta

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

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

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

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

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

Risk-Reward Ratios

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

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

(But this does not tell us how much dollar alpha is there. It can be eaten by transaction costs.)

Sharpe Ratio must be familiar $\frac{\mathbb{E}(R_t - r_f)}{\sigma(R_t - r_f)}$. It measures return per unit of risk.

Factors

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

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

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

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

Named Factors

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

Factors Backtesting

So how do we check against those factors?

We can set up a regression!

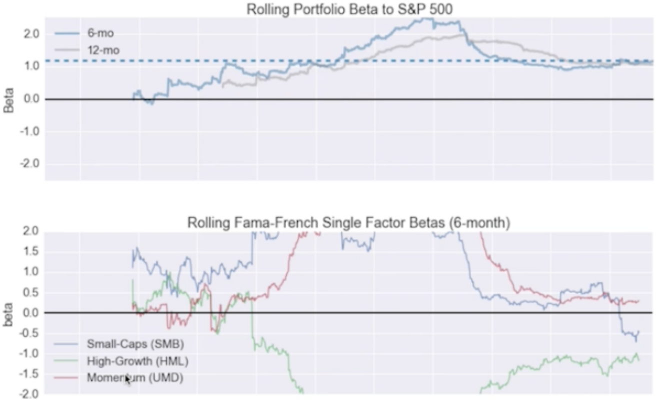
$$R_t^S = \alpha + \beta^M R_t^M + \beta^{HML} R_t^{HML} + \epsilon_t$$

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

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

Factors Backtesting (Advanced)

- Scale returns to have the same volatility as the benchmark (put on the same plot for correct comparison).
- Rolling Sharpe Ratio, 12M data (changes **not** desirable).
- Rolling market factor beta 6M, 12M data ($\beta > 1$ and changes **are not** desirable).
- Rolling betas *wrt* to HML, SMB, UMD (value factor, small business factor, and momentum factor).



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

Drawdowns

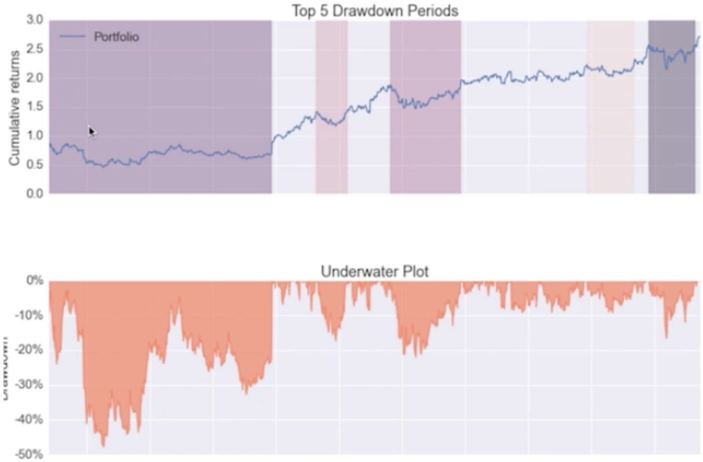
The drawdown is the cumulative percentage loss, given the loss in the initial timestep.

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

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

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

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



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

Drawdown Control

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

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

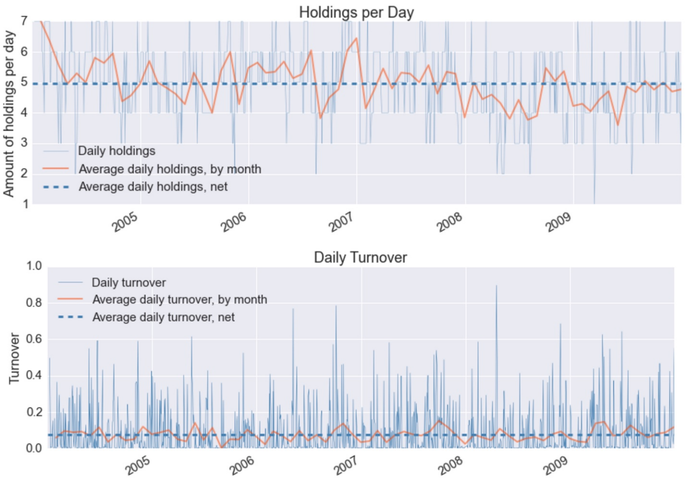
$$\mathrm{VaR}_t \leq \mathrm{MADD} - \mathrm{DD}_t$$

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

Backtesting Risk and Liquidity - Summary

1. Does cumulative P&L behave as expected (eg, for a cointegration trade)? Is P&L coming from a few or lot of trades/time period? What are the SR/Maximum Drawdown? Behaviour of risk measures (volatility/VaR)? Concentration in assets and attribution?

2. Impact of transaction costs (plot the expected P&L value (alpha) vs. number of transactions). Is there any P&L from the spread left after bid-ask spread? What is an impact of transaction costs (*aka* slippage assumption)? That would depend on turnover.



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

Algorithmic Flow

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

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

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

Developing a Trading Business

You can find the full suite of libraries for data prep, modelling, analytical/backtesting reports generation and testing.

From: For Python Quants, Nov 2015, *Building an Energy Trading Business from Scratch*, Teodora Baeva (BTG Pactual)

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

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

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

The large part of model risk for time series analysis (statistical tests) is formalised as the **multiple testing problem**. Here are the guidelines from the American Statistical Society,

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

"Selecting one 'significant' result from a multiplicity of parallel tests poses a grave risk of an incorrect conclusion."

Cap Pricing, Volatility and LMM Calibration

Caplet

A caplet is an interest rate option that pays **a cashflow** based on the value of LIBOR at a re-set time T_i.

The cashflow is paid for the period $\tau = [T_i, T_{i+1}]$ in arrears.

$$\mathrm{DF}_{\mathrm{OIS}}(0, T_{i+1}) \times \max(L(T_i, T_{i+1}) - K, 0) \times \tau \times N \tag{6}$$

- $L(T_i, T_{i+1})$ is the forward LIBOR. Assume $L(T_i, T_{i+1}) = f_i$
- τ is year fraction that converts an annualised rate
- N is the notional that can be scaled as $N = 1$

Payoff and parity

Buying a caplet gives protection from an increase in LIBOR rate:

$$L - (L - K)^{+} = \min(L, K)$$

Alternatively for a floorlet:

$$\max(L, K)$$

Put-call parity for caplet and floorlet becomes:

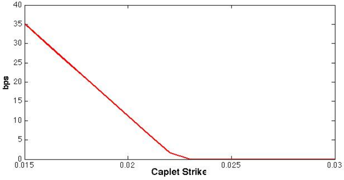
Buying a caplet and selling a floorlet with the same strike gives a payoff equal to a FRA contract. Consider

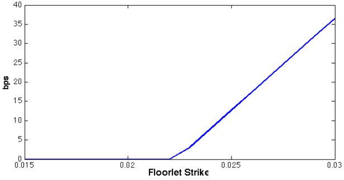
$$(L - K)^{+} - (K - L)^{+} = (L - K) \qquad \text{always}$$

where FRA fixed rate is equal to the strike, so $(f - K)$.

Pricing Skew

Plotting caplets/floorlets in *bps* for a range of strikes gives ‘a skew’.





The skew is due to a simple fact: the more an option is in the money the larger cashflow it delivers. The relationship is almost linear.

Richer volatility surfaces are possible.

Let's look at caplet pricing using curve data simulated by the HJM Model (and gain more understanding about simulation).

Black (1976) formula is used to quote implied volatility σ_i^{cap}.

$$\begin{aligned}
\textbf{Cap} &= Z(0,T_i)\,[f_i\,N(d_1) - K\,N(d_2)]\,\frac{\tau_i}{1+f_i\tau_i} \\
d_{1,2} &= \frac{\ln(f_i/K) \pm 0.5\sigma^2 T}{\sigma\sqrt{T}}
\end{aligned}$$

where $f_i = F(t, T_i, T_{i+1})$ becomes set at the caplet expiry T_i and paid over $[T_i, T_{i+1}]$.

Volatility Stripping and Fitting

Volatility

To calibrate the LIBOR model = To strip caplet volatility

We will consider stripping caplet volatility from market-quoted caps. This has elements of bootstrapping time-dependent volatility

$$\sigma^{BS}(T_{i-1}, T_i) \qquad or \qquad \sigma_i(t)$$

For the forward curve, caplets are in sequence with 3M reset. Each caplet is on Forward LIBOR $L(0, T_{i-3M}, T_i)$, in Gatarek notation.

Caps are traded in $1Y, 2Y, 3Y, ...$ increments.

Market Cap Quotes

Tenor T_i	Date	Discount factor $B(0, T_i)$	Cap volatility $\sigma^{cap}(T_0, T_i)$
$t = 0$	21-01-2005	1.0000000	N/A
T_0	25-01-2005	0.9997685	N/A
T_{SN}	26-01-2005	0.9997107	N/A
T_{SW}	01-02-2005	0.9993636	N/A
T_{2W}	08-02-2005	0.9989588	N/A
T_{1M}	25-02-2005	0.9979767	N/A
T_{2M}	25-03-2005	0.9963442	N/A
T_{3M}	25-04-2005	0.9945224	N/A
T_{6M}	25-07-2005	0.9890361	N/A
T_{9M}	25-10-2005	0.9832707	N/A
T_{1Y}	25-01-2006	0.9772395	0.1641
T_{2Y}	25-01-2007	0.9507588	0.2137
T_{3Y}	25-01-2008	0.9217704	0.2235

From: *LIBOR Market Model in Practice* by Gatarek, et al. (2006).

Calibrating LMM on caplets

Caplet volatility stripping is nuanced. We are going to cover the inputs and principles.

1. First, we need ATM strikes for the caplets, which are equal to the forward-starting swap rates

$$S(t, T_i, T_{i+1}) \quad \text{or} \quad S(T_i; T_0^*, T_{3M}^*) \quad \text{gives} \quad K$$

   obtained directly from the forward curve *today*.

   Please see Yield Curve.xlsm on how to calculate ZCB prices *implied* by the forward curve.

Calibrating LMM on caplets (Cont.)

We use the assumption of flat volatility and extrapolation.

2. Prepare interpolated market volatilities

$\cancel{\sigma^{cap}(t, T_{3M})}$, $\sigma^{cap}(t, T_{6M})$, $\sigma^{cap}(t, T_{9M})$, $\sigma^{cap}_{Mkt}(t, T_{12M})$, $\sigma^{cap}(t, T_{18M})$, ...

Initial flat volatility gives

$$\sigma^{cap}(t, T_{3M}, T_{6M}) = \sigma^{cap}_{Mkt}(t, T_{6M}) = \sigma^{cap}_{Mkt}(t, T_{1Y})$$

use Black formula to obtain cashflow

$\mathrm{cpl}(T_{3M}, T_{6M})$.

Volatility Stripping

Instead of bootstrapping via variance, keep using cashflows

$$\text{cap}_T = \sum_i^T \text{cpl}_i$$

$$\text{cpl}(T_{6M}, T_{9M}) = \text{cap}_{Mkt}(t, T_{9M}) - \text{cpl}(T_{3M}, T_{6M}) - \cancel{\text{cpl}(t, T_{3M})}$$

$$\text{cpl}(T_{9M}, T_{12M}) = \text{cap}_{Mkt}(t, T_{1Y}) - \text{cpl}(T_{6M}, T_{9M}) - \text{cpl}(T_{3M}, T_{6M}) - \cancel{\text{cpl}(t, T_{3M})}$$

etc.

Use the Black formula to convert back

$$\text{cpl}(T_{i-1}, T_i) \quad \Leftrightarrow \quad \sigma^{cap}(T_{i-1}, T_i).$$

Volatility Stripping (Cont.)

With the Black formula, even though you use the same constant volatility $\sigma^{cap}(t, T_{3M}, T_{6M}) = \sigma^{cap}(t, T_{6M}) = \sigma^{cap}(t, T_{1Y})$

$$\text{cpl}(T_{3M}, T_{6M}) \quad \neq \quad \text{cpl}(T_{6M}, T_{9M}).$$

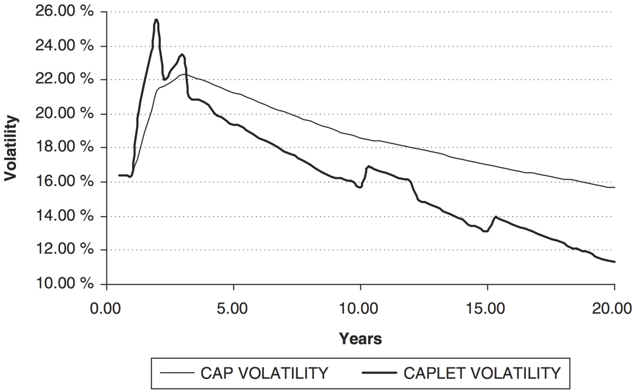
Because Forward LIBOR $L(T_i, T_{i+1}) = f_i$ and Forward Swap Rates (caplet strikes) $S(t, T_i, T_{i+1}) = K$ will be different for each 3M consecutive period .

Also, even if implied volatility σ is the same (annualised)

$$\begin{aligned} \text{cpl}(T_{3M}, T_{6M}) \quad &< \quad \text{cap}(t, T_{6M}) \\ \text{cpl}(T_{3M}, T_{6M}) \quad &\ll \quad \text{cap}(t, T_{1Y}). \end{aligned}$$

Market Cap Quotes

The result is caplet volatility on a given day (Gatarek, et al., 2006)



The solution is consistent with the bootstrapping we did for local time-dependent volatility (as actual) from implied volatility (*CQF UV Lecture*).

Instantaneous Volatility

With volatility stripping, we moved from 12M to 3M increment.

But, the caplet implied volatility is still an average (via integration) over the actual, instantaneous volatility, on time to maturity.

$$\sigma^{cap}(t, T_{i-1}, T_i) = \sqrt{\frac{1}{T_{i-1} - t} \int_t^{T_{i-1}} \sigma^{inst}(\tau)^2 \, d\tau}$$

Therefore, we have two different covariance matrices with an algorithm to match them

$$\mathbf{\Sigma}_{cpl} \Rightarrow \mathbf{\Sigma}_{inst}$$

The fitting done next applies to instantaneous volatility.

Volatility Fitting

We would like the term structure of volatility to be time-homogeneous,

$$\sigma^{inst}(t) = \phi_i \left[(a + b(T_{i-1} - t)) \times e^{-c(T_{i-1} - t)} + d \right]$$

a, b, c, d are **the same** for all tenors! One set of numbers. This is regardless your optimization method.

What varies for each tenor is $0.9 < \phi_i < 1.1$ to make for near-perfect fit, such that

$$\text{argmin} \left(\sigma^{inst}_{Est} - \sigma^{inst}_{Fit} \right)^2$$

The fitted σ^{inst}_{Fit} agrees with the stripped one with only a minimal squared error.

Parametrised Instantaneous Volatility

$$\int_t^{T_{i-1}} \sigma^{inst}(\tau)^2 \, d\tau \quad = \quad \frac{1}{4c^3} \left(4ac^2 d[e^{2c(t-T_{i-1})}] + ... \right)$$

The so called FRA/FRA covariance matrix for instantaneous volatility, our $\mathbf{\Sigma}_{inst}$, also has a, b, c, d-parametrised, closed-form solution

$$\int \rho_{ij} \, \sigma_i(\tau) \, \sigma_j(\tau) \, d\tau = e^{-\beta|t_i - t_j|} \phi_i \phi_j \, \frac{1}{4c^3} \left(4ac^2 d[e^{c(t-T_i)} + e^{2c(t-T_j)}] + ... \right)$$

The complete parametric solutions to be found in CQF Lecture on the LMM and Peter Jaekel's textbook.

Parametric Correlation

The simplest parametric fit for correlations with $\beta \approx 0.1$ has merits for longer tenors

$$\rho_{ij} = e^{-\beta(t_i - t_j)}$$

The two-factor parametric form of Schoenmakers and Coffey (2003):

$$\rho_{ij} = \exp\left(-\frac{|i-j|}{m-1}\left[-\ln\beta_1 + \beta_2 \ldots\right]\right)$$

works for situations that are different from the stylised empirical observations.

Empirical Correlation

First, changes (in forward rates) at the neighbouring tenors tend to correlate stronger

$$\mathrm{Corr}[\Delta f_{i-1}, \Delta f_i] > \mathrm{Corr}[\Delta f_{i-3}, \Delta f_i]$$

Second, correlation is higher towards the long end of the curve.

$$\mathrm{Corr}[\Delta f_{i-1}, \Delta f_i] < \mathrm{Corr}[\Delta f_{j-1}, \Delta f_j] \qquad \text{for } j \gg i$$

At the short end the rates tend to behave more independently from one another. This is due to being most sensitive to the principal component/primary risk factor of rising the level in the risk-free rate and the entire curve. Further, for 3M, 6M and 1Y tenors there is own dynamics because of how specific market instruments are traded.

LIBOR Market Model SDE

LMM Notation

The LIBOR Market Model was designed to operate with forward rates and denotes them as f_i, where

$$f_i = F(t; t_i, t_{i+1})$$

The forward rate re-sets at time t_i and matures at time t_{i+1}.

Discount factor is represented in the LIBOR model as

$$Z(t; T_{i+1}) \equiv \frac{1}{1 + \tau_i f_i}$$

This is discount factor over *the forward period* $\tau_i = t_{i+1} - t_i$! We need 'one step back' in LMM SDE.

LMM Result

Using a discretely rebalanced money market account as Numeraire, the forward rate f_i follows the log-normal process

$$\frac{df_i}{f_i} = \sum_{j=m(t)}^{i} \frac{\tau_j f_j}{1+\tau_j f_j}\,\sigma_i \sigma_j \rho_{ij}\,dt + \sigma_i\, dW_i^{\mathbb{Q}^{m(t)}} \tag{7}$$

- $m(t)$ is an index for the next re-set time. This means that $m(t)$ is the smallest integer such that $t^* \leq t_{m(t)}$.

Preview: terms come from

$$df_j df_k = f_j f_k \sigma_j \sigma_k \rho_{jk} dt$$

in an SDE for dZ_i.

Rolling-forward risk-neutral world

The SDE is defined under the Spot LIBOR Measure $\mathbb{Q}^{m(t)}$, known as **the rolling forward risk-neutral world**.

- We just keep discounting the drift.

If you would like to see how LMM SDE is derived starting from an SDE for dZ_i and using the Ito lemma, please review *CQF Extra on LMM* by Tim Mills.

LMM SDE discretised (single-factor)

The SDE (7) is for the **log-normal** dynamics of f_i given by the instantaneous FRAs. It is solved into a discretised version as follows:

$$
\begin{aligned}
f_i(t_{k+1}) \quad = \quad & f_i(t_k)\exp\left[\left(\sigma_i(t_{i-k-1})\sum_{j=k+1}^{i}\frac{\tau_j f_j(t_k)\,\sigma_j(t_{j-k-1})\rho_{ij}}{1+\tau_j f_j(t_k)}\right.\right. \\
& \left.\left. -\frac{1}{2}\sigma_i^2(t_{i-k-1})\right)\tau_k + \sigma_i(t_{i-k-1})\phi_i\sqrt{\tau_k}\right] \qquad (8)
\end{aligned}
$$

where $f_j(t_k) = f_j$ and $\sigma_j(t_k) = \sigma_j(t)$ for $t_k < t < t_{k+1}$.

Notation t_{j-k-1} means we refer to the previous time step $k-1$.

SDE Simulation

- HJM is a Normal (Gaussian) model as we simulated

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

$d\bar{f}$ are Normally distributed and the curve **in row** is evolved in continuous time at fixed tenors.

- LMM is a Log-Normal model

$$f(t + dt) = f(t) \exp (df)$$

The curve evolved in discrete tenor chunks, arranged in **column**. For 6M LIBOR, the second column will lose $\tau = 0.5$ tenor and the third column will lose $\tau = 0.5, 1$ tenors – due to expiry. See Alonso et al. (2010).