9 Market Risk

9.1 Risk factors and mapping

9.1.1 The loss operator

- The key idea in this section is that of a loss operator for expressing the change in value of a portfolio in terms of risk-factor changes.
- Let the current time be t and assume the current value V_t os an asset portfolio is known, or can be computed with appropriate valuation models.
- We are interested in value changes or losses over a relatively short time period $[t, t+1]$, for example one day, two weeks or month.
- Scaling may be applied to derive capital requirements for longer periods.
- We assume there is no change to the composition of the portfolio over the time period.
- The future value V_{t+1} is modelled as a random variable.

- We want to determine the distribution of the loss distribution of $L_{t+1} = -(V_{t+1} - V_t)$.
- We map the value at time t using the formula

$$V_t = g(\tau_t, \boldsymbol{Z}_t)$$

where τ_t is time t expressed in units of valuation time.

The issue of time

- We will be quite precise about the modelling of time.
- The natural time unit for valuation of positions might be yearly; e.g. in Black-Scholes valuation, the volatility is expressed in annualized terms.
- On the other hand the risk modelling time horizon $[t, t+1]$ is typically shorter.
- Let Δt be the length of the time horizon in valuation time.
- For example, suppose that valuation time is yearly. Then a monthly time horizon would be $\Delta t = 1/12$ and a trading day $\Delta t = 1/250$.

- We set $\tau_t = t(\Delta t)$ for all t so that $\tau_{t+1} - \tau_t = \Delta t$.

From the mapping to the loss operator

- The risk factor changes over the time horizon are

$$\mathbf{X}_{t+1} = \mathbf{Z}_{t+1} - \mathbf{Z}_t.$$

- Typically, historical risk factor data are available as a time series $\mathbf{X}_{t-n}, \ldots, \mathbf{X}$ and these are used to model the behaviour of $\mathbf{X}_{t+1}$.
- We have

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) \\ &= -\left(g(\tau_{t+1}, \mathbf{Z}_{t+1}) - g(\tau_t, \mathbf{Z}_t)\right) \\ &= -(g(\tau_t + \Delta t, \mathbf{Z}_t + \mathbf{X}_{t+1}) - g(\tau_t, \mathbf{Z}_t)). \end{aligned} \tag{50}$$

- Since the risk factor values $\mathbf{Z}_t$ are known at time t, the loss L_{t+1} is determined by the risk factor changes $\mathbf{X}_{t+1}$.

- Given a realization $\mathbf{z}_t$ of $\mathbf{Z}_t$, the loss operator at time t is defined to be

$$l_{[t]}(\mathbf{x}) = -(g(\tau_t + \Delta t, \mathbf{z}_t + \mathbf{x}) - g(\tau_t, \mathbf{z}_t)), \tag{51}$$

so that

$$L_{t+1} = l_{[t]}(\mathbf{X}_{t+1}).$$

- The loss operator embodies the idea of full revaluation.
- From the perspective of time t the loss distribution of L_{t+1} is determined by the multivariate distribution of $\mathbf{X}_{t+1}$.

9.1.2 Delta and delta–gamma approximations

- If the mapping function g is differentiable and Δt is relatively small we can approximate g with a first-order Taylor series approximation

$$g(\tau_t + \Delta t, \boldsymbol{z}_t + \boldsymbol{x}) \approx g(\tau_t, \boldsymbol{z}_t) + g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \sum_{i=1}^{d} g_{z_i}(\tau_t, \boldsymbol{z}_t) x_i, \tag{52}$$

where the τ-subscript and z_i-subscript denote partial derivatives with respect to (valuation) time and the risk factors respectively.

- This allows us to approximate the loss operator in (51) by the linear loss operator at time t given by

$$l_{[t]}^{\Delta}(\boldsymbol{x}) := -\left(g_{\tau}(\tau_t, \boldsymbol{z}_t)\Delta t + \sum_{i=1}^{d} g_{z_i}(\tau_t, \boldsymbol{z}_t)x_i\right). \tag{53}$$

- Note that, when working with a short time horizon Δt, the term $g_{\tau}(\tau_t, \boldsymbol{z}_t)\Delta t$ is typically small and is sometimes omitted in practice.

Example 9.1 (European call option)

- Consider portfolio consisting of one standard European call on a non-dividend paying stock S with maturity T and exercise price K.
- The Black-Scholes value of this asset at time t is $C^{BS}(t, S_t, r, \sigma)$ where

$$C^{BS}(t, S; r, \sigma) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2),$$

  Φ is standard normal df, r represents risk-free interest rate, σ the volatility of underlying stock, and where

$$d_1 = \frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \text{ and } d_2 = d_1 - \sigma\sqrt{T-t}.$$

- While in the BS model, it is assumed that interest rates and volatilities are constant, in reality they tend to fluctuate over time; they should be added to our set of risk factors.

- The risk factors: $\mathbf{Z}_t = (\log S_t, r_t, \sigma_t)'$.
- The risk factor changes: $\mathbf{X}_t = (\log(S_t/S_{t-1}), r_t - r_{t-1}, \sigma_t - \sigma_{t-1})'$.
- The mapping:

$$V_t = C^{BS}(\tau_t, S_t; r_t, \sigma_t) = g(\tau_t, \boldsymbol{Z}_t)$$

- For derivative positions it is quite common to use the linear loss operator

$$L^{\Delta}_{t+1} = l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1}) = -\left(g_\tau(\tau_t, \mathbf{z}_t)\Delta t + \sum_{i=1}^{3} g_{z_i}(\tau_t, \mathbf{z}_t) X_{t+1,i}\right),$$

  where g_τ, g_{z_i} denote partial derivatives.
- Δt is the length of the time interval expressed in years since Black-Scholes parameters relate to units of one year.

- It is more common to write the linear loss operator as

$$l_{[t]}^{\Delta}(\boldsymbol{x}) = -\left(C_t^{BS} + C_S^{BS} S_t x_1 + C_r^{BS} x_2 + C_\sigma^{BS} x_3\right),$$

in terms of the derivatives of the BS formula or the Greeks.

  - C_S^{BS} is known as the delta of the option.
  - C_σ^{BS} is the vega.
  - C_r^{BS} is the rho.
  - C_t^{BS} is the theta.

  Note the appearance of S_t in the C_S^{BS} term. This is because the risk factor is $\ln S_t$ rather than S_t and $C_{\ln S}^{BS} = C_S^{BS} S_t$.

Quadratic loss operator

- Recall the first-order Taylor series approximation of mapping in (52).
- Let $\boldsymbol{\delta}(\tau_t, \boldsymbol{z}_t) = (g_{z_1}(\tau_t, \boldsymbol{z}_t), \ldots, g_{z_d}(\tau_t, \boldsymbol{z}_t))'$ be the first-order partial derivatives of the mapping with respect to the risk factors.

- Let $\boldsymbol{\omega}(\tau_t, \boldsymbol{z}_t) = (g_{z_1\tau}(\tau_t, \boldsymbol{z}_t), \ldots, g_{z_d\tau}(\tau_t, \boldsymbol{z}_t))'$ denote the mixed partial derivatives with respect to time and the risk factors.
- Let $\Gamma(\tau_t, \boldsymbol{z}_t)$ denote the matrix with (i, j)th element given by $g_{z_i z_j}(\tau_t, \boldsymbol{z}_t)$; this matrix contains gamma sensitivities to individual risk factors on the diagonal and cross gamma sensitivities to pairs of risk factors off the diagonal.
- The full second-order approximation of the mapping function is g is

$$\begin{aligned} g(\tau_t + \Delta t, \boldsymbol{z}_t + \boldsymbol{x}) \approx\ & g(\tau_t, \boldsymbol{z}_t) + g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \boldsymbol{\delta}(\tau_t, \boldsymbol{z}_t)'\boldsymbol{x} + \\ & \frac{1}{2}\big(g_{\tau\tau}(\tau_t, \boldsymbol{z}_t)(\Delta t)^2 + 2\boldsymbol{\omega}(\tau_t, \boldsymbol{z}_t)'\boldsymbol{x}\Delta t + \\ & \boldsymbol{x}'\Gamma(\tau_t, \boldsymbol{z}_t)\boldsymbol{x}\big)\,. \end{aligned}$$

- In practice, we would usually omit terms of order $o(\Delta_t)$ (terms that tend to zero faster than Δ_t). In standard continuous-time financial models like Black-Scholes the risk-factor changes $\boldsymbol{x}$ are of order $\sqrt{\Delta_t}$.

- This leaves us with the quadratic loss operator

$$l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) = -\left(g_\tau(\tau_t, \boldsymbol{z}_t)\Delta t + \boldsymbol{\delta}(\tau_t, \boldsymbol{z}_t)'\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}'\Gamma(\tau_t, \boldsymbol{z}_t)\boldsymbol{x}\right) \tag{54}$$

which is more accurate than the linear loss operator (53).

Example 9.2 (European call option)

The quadratic loss operator is

$$\begin{aligned} l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) =& l_{[t]}^{\Delta}(\boldsymbol{x}) - 0.5\left(C_{SS}^{BS}S_t^2x_1^2 + C_{rr}^{BS}x_2^2 + C_{\sigma\sigma}^{BS}x_3^2\right) \\ &- \left(C_{Sr}^{BS}S_tx_1x_2 + C_{S\sigma}^{BS}S_tx_1x_3 + C_{r\sigma}^{BS}x_2x_3\right) . \end{aligned}$$

The names of the second-order Greeks (with the exception of gamma) are rather obscure. Here are some of them:

- C_{SS}^{BS} is known as the gamma of the option;
- $C_{\sigma\sigma}^{BS}$ is the vomma;
- $C_{S\sigma}^{BS}$ is the vanna.

9.1.3 Mapping bond portfolios

Basic definitions for bond pricing

- Let $p(t,T)$ denote the price at time t of a default-free zero-coupon bond paying one at time T (also called a discount factor).
- Time is measured in years.
- Many other fixed-income instruments such as coupon bonds or standard swaps can be viewed as portfolios of zero-coupon bonds.
- The mapping $T \to p(t,T)$ for different maturities is one way of describing the so-called term structure of interest rates at time t. An alternative description is based on yields.
- The term structure $T \to p(t,T)$ is known at time t.
- However the future term structure $T \to p(t+x,T)$ for $x > 0$ is not known at time t and must be modelled stochastically.

- The continuously compounded yield of a zero-coupon bond is

$$y(t,T) = -\frac{\ln p(t,T)}{T-t}. \tag{55}$$

- We have the relation

$$p(t,T) = \exp(-(T-t)y(t,T)).$$

- The yield is the constant, annualized rate implied by the price $p(t,T)$. Also known as spot rate.
- The mapping $T \to y(t,T)$ is referred to as the continuously compounded yield curve at time t.
- Yields are comparable across different times to maturity.

Detailed mapping of a bond portfolio

- Consider a portfolio of d default-free zero-coupon bonds with maturities T_i and prices $p(t,T_i)$ for $i = 1,\ldots,d$. Assume $p(T_i,T_i) = 1$ for all i.

- By λ_i we denote the number of bonds with maturity T_i in the portfolio.
- The portfolio value at time t is given by

$$V(t) := \sum_{i=1}^{d} \lambda_i p(t, T_i) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - t) y(t, T_i)).$$

- In a detailed analysis of the change in value one takes all yields $y(t, T_i)$, $1 \leq i \leq d$, as risk factors.
- We want to put this in the general discrete-time framework of the mapping

$$V_t = g(\tau_t, \boldsymbol{Z}_t).$$

- We set

$$\tau_t = t(\Delta t), \quad V_t = V(\tau_t), \quad Z_{t,i} = y(\tau_t, T_i)$$

where Δt is risk management time horizon in years.

- We obtain a mapping of the form

$$V_t = V(\tau_t) = g(\tau_t, \boldsymbol{Z}_t) = \sum_{i=1}^{d} \lambda_i \exp(-(T_i - \tau_t) Z_{t,i}). \tag{56}$$

The loss operator and its approximations

- The portfolio loss is

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) \\ &= -\sum_{i=1}^{d} \lambda_i e^{-(T_i - \tau_t) Z_{t,i}} \Big(\exp\left(Z_{t,i} \Delta t - (T_i - \tau_{t+1}) X_{t+1,i} \right) - 1 \Big) . \end{aligned}$$

- Reverting to standard bond pricing notation the loss operator is

$$l_{[t]}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) \Big(\exp\left(y(\tau_t, T_i) \Delta t - (T_i - \tau_{t+1}) x_i \right) - 1 \Big) ,$$

where x_i represents the change in yield of the ith bond.

- The first derivatives of the mapping function (56) are

$$g_\tau(\tau_t, \boldsymbol{z}_t) = \sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) z_{t,i}$$

$$g_{z_i}(\tau_t, \boldsymbol{z}_t) = -\lambda_i (T_i - \tau_t) \exp(-(T_i - \tau_t) z_{t,i}).$$

- Inserting these in (53) and reverting to standard bond pricing notation we obtain

$$l_{[t]}^{\Delta}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) \Big(y(\tau_t, T_i)\Delta t - (T_i - \tau_t) x_i \Big), \tag{57}$$

- For the second-order approximation we need the second derivatives with respect to yields which are

$$g_{z_i z_i}(\tau_t, \boldsymbol{z}_t) = \lambda_i (T_i - \tau_t)^2 \exp(-(T_i - \tau_t) z_{t,i})$$

and $g_{z_i z_j}(\tau_t, \boldsymbol{z}_t) = 0$ for $i \neq j$.

- The quadratic loss operator (54) is

$$l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i) \left(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)x_i + \frac{1}{2}(T_i - \tau_t)^2 x_i^2 \right). \tag{58}$$

Relationship of linear operator to duration

- Consider a very simple model for the yield curve at time t in which

$$y(\tau_{t+1}, T_i) = y(\tau_t, T_i) + x$$

  for all maturities T_i.
- In our mapping notation

$$Z_{t+1,i} = Z_{t,i} + X_{t+1}, \quad \forall i.$$

- In this model we assume that a parallel shift in level takes place along the entire yield curve.

- This is unrealistic but frequently assumed in practice.
- In this model the loss operator and its linear and quadratic approximations are functions of a scalar variable x, the change in level.
- Under the parallel shift model we can write

$$l^{\Delta}_{[t]}(x) = -V_t\Big(A_t\Delta t - D_t x\Big), \tag{59}$$

where

$$D_t := \sum_{i=1}^{d} \frac{\lambda_i p(\tau_t, T_i)}{v_t}(T_i - \tau_t), \quad A_t := \sum_{i=1}^{d} \frac{\lambda_i p(\tau_t, T_i)}{V_t} y(\tau_t, T_i).$$

- D_t is usually called the (Macaulay) duration of the bond portfolio.
- It is a weighted sum of the times to maturity of the different cash flows in the portfolio, the weights being proportional to the discounted values of the cash flows.

Interpreting duration

- Over short time intervals losses of value in the bond portfolio will be determined by $l^{\Delta}_{[t]}(x) \approx V_t D_t x$.
- Increases in level of yields lead to losses; decreases lead to gains.
- The duration D_t is the bond pricing analogue of the delta of an option.
- Any two bond portfolios with equal value and duration will be subject to similar losses when there is a small parallel shift of the yield curve.
- Duration is an important tool in traditional bond-portfolio or asset-liability management.
- An asset manager, who invests in various bonds to cover promised cash flows in the future, invests in such a way that the duration of the overall portfolio of bonds and liability cash flows is equal to zero.
- Portfolios are immunized against small parallel shifts in yield curve, but not changes of slope and curvature.

Relationship of quadratic operator to convexity

- It is possible to get more accurate approximations for the loss in a bond portfolio by considering second-order effects.
- The analogue of the gamma of an option is convexity. Under the parallel shift model, the quadratic loss operator (58) becomes

$$l_{[t]}^{\Delta\Gamma}(x) = -V_t\left(A_t\Delta t - D_t x + \frac{1}{2}C_t x^2\right), \tag{60}$$

  where

$$C_t := \sum_{i=1}^{d} \frac{\lambda_i p(\tau_t, T_i)}{V_t}(T_i - \tau_t)^2$$

  is the convexity of the bond portfolio.
- The convexity is a weighted average of the squared times to maturity and is (minus) the derivative of the duration with respect to yield.

Interpreting convexity

- Consider two portfolios (1) and (2) with identical durations $D_t^{(1)} = D_t^{(2)}$ but differing convexities satisfying $C_t^{(1)} > C_t^{(2)}$.
- Ignoring terms in Δt, the difference in loss operators satisfies

$$l_{[t]}^{\Delta\Gamma,1}(x) - l_{[t]}^{\Delta\Gamma,2}(x) \approx -\frac{1}{2} V_t (C_t^{(1)} - C_t^{(2)}) x^2 < 0.$$

- Since $l_{[t]}^{\Delta\Gamma,1}(x) < l_{[t]}^{\Delta\Gamma,2}(x)$ an increase in the level of yields ($x > 0$) will lead to smaller losses for portfolio (1)
- Since $-l_{[t]}^{\Delta\Gamma,1}(x) > -l_{[t]}^{\Delta\Gamma,2}(x)$ a decrease in the level of yields ($x < 0$) will lead to larger gains.
- For this reason higher convexity is considered a desirable attribute of a bond portfolio in risk management.

9.1.4 Factor models for bond portfolios

The need for factor models

- The parallel shift model is unrealistic in practice.
- For large portfolios of fixed-income instruments, such as the overall fixed-income position of a major bank, modelling changes in the yield for every cash flow maturity date becomes impractical.
- Moreover, the statistical task of estimating a distribution for $\boldsymbol{X}_{t+1}$ is difficult because the yields are highly dependent for different times to maturity.
- A pragmatic approach is therefore to build a factor model for yields that captures the main features of the yield curve.
- Three-factor models of the yield curve in which the factors typically represent level, slope and curvature are often used in practice.

The approach based on the Nelson and Siegel (1987) model

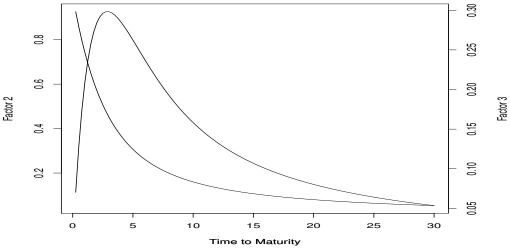
- We assume that at time t the yield curve can be modelled by

$$y(\tau_t, T) \approx Z_{t,1} + k_2(T - \tau_t, \eta_t) Z_{t,2} + k_3(T - \tau_t, \eta_t) Z_{t,3}, \tag{61}$$

  where the functions k_2 and k_3 are given by

$$k_2(s, \eta) = \frac{1 - \exp(-\eta s)}{\eta s}, \ k_3(s, \eta) = k_2(s, \eta) - \exp(-\eta s) \, .$$

- Nelson-Siegel functions $k_2(s, \eta)$ and $k_3(s, \eta)$ for an η value of 0.623:



- η is an extra tuning parameter to improve fit.
- There are other simple factor models including the Svensson model.
- Clearly $\lim_{s\to\infty} k_2(s,\eta) = \lim_{s\to\infty} k_3(s,\eta) = 0$ while $\lim_{s\to 0} k_2(s,\eta) = 1$ and $\lim_{s\to 0} k_3(s,\eta) = 0$.
- It follows that
$$\lim_{T\to\infty} y(\tau_t, T) = Z_{t,1},$$
so that the first factor is usually interpreted as a long-term level factor.
- $Z_{t,2}$ is interpreted as a slope factor because the difference in short-term and long-term yields satisfies
$$\lim_{T\to\tau_t} y(\tau_t, T) - \lim_{T\to\infty} y(\tau_t, T) = Z_{t,2}.$$
- $Z_{t,3}$ has an interpretation as a curvature factor.
- Using (61), the portfolio mapping (56) becomes
$$V_t = g(\tau_t, \boldsymbol{Z}_t) = \sum_{i=1}^{d} \lambda_i \exp\left(-(T_i - \tau_t) \sum_{j=1}^{3} k_j(T_i - \tau_t, \eta_t) Z_{t,j} \right),$$

where $k_1(s,\eta) = 1$.

- It is then straightforward to derive the loss operator $l_{[t]}(\boldsymbol{x})$ or its linear version $l^{\Delta}_{[t]}(\boldsymbol{x})$ which are functions on $\mathbb{R}^3$ rather than $\mathbb{R}^d$.
- To use this method to evaluate the loss operator at time t we require realized values $\boldsymbol{z}_t$ for the risk factors $\boldsymbol{Z}_t$. We have to overcome the fact that the Nelson-Siegel factors $\boldsymbol{Z}_t$ are not directly observed at time t. Instead they have to be estimated from observable yield curve data.
- Let $\boldsymbol{Y}_t = (y(\tau_t, \tau_t + s_1), \ldots, y(\tau_t, \tau_t + s_m))'$ denote the data vector at time t, containing the yields for m different times to maturity, $s_1, \ldots, s_m$, where m is large.
- This is assumed to follow the factor model

$$\boldsymbol{Y}_t = B_t \boldsymbol{Z}_t + \boldsymbol{\varepsilon}_t,$$

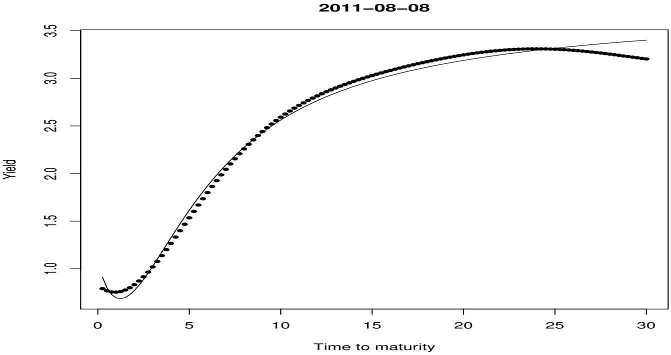
where $B_t \in \mathbb{R}^{m\times 3}$ is the matrix with ith row $(1, k_2(s_i,\eta_t), k_3(s_i,\eta_t))$ and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^m$ is an error vector.

- For a given value of η_t the estimation of $\boldsymbol{Z}_t$ can be carried out as a cross-sectional regression using weighted least squares. It is a fundamental factor model where the loading matrix B_t is known.
- To estimate η_t a more complicated optimization is carried out.

Example 9.3

- The data are daily Canadian zero-coupon bond yields for 120 different quarterly maturities ranging from 0.25 years to 30 years.
- They have been generated using pricing data for Government of Canada bonds and treasury bills.

- We model the yield curve on the 8th August 2011.
- The estimated value are $z_{t,1} = 3.82$, $z_{t,2} = -2.75$, $z_{t,3} = -5.22$ and $\hat{\eta}_t = 0.623$. Thus the curves $k_2(s, \eta)$ and $k_3(s, \eta)$ are as shown earlier.
- The fitted Nelson-Siegel curve and the data are shown below:



The approach based on PCA

- The key difference to the Nelson-Siegel approach is that here the dimension reduction via factor modelling is applied at the level of the risk factor changes $\boldsymbol{X}_{t+1}$ rather than the risk factors $\boldsymbol{Z}_t$.
- We recall that PCA can be used to construct factor models of the form

$$\boldsymbol{X}_{t+1} = \boldsymbol{\mu} + \Gamma_1 \boldsymbol{F}_{t+1} + \boldsymbol{\varepsilon}\,, \tag{62}$$

  where $\boldsymbol{F}_{t+1}$ is a p-dimensional vector of principal component factors $(p < d)$, $\Gamma_1 \in \mathbb{R}_{d \times p}$ contains the corresponding loading matrix, $\boldsymbol{\mu}$ is the mean vector of $\boldsymbol{X}_{t+1}$ and $\boldsymbol{\varepsilon}$ is an error vector.
- Typically, the error term is neglected and $\boldsymbol{\mu} \approx \mathbf{0}$, so that we make the approximation $\boldsymbol{X}_{t+1} \approx \Gamma_1 \boldsymbol{F}_{t+1}$.
- In the case of the linear loss operator for the bond portfolio in (57) we

basically replace $l^{\Delta}_{[t]}(\boldsymbol{X}_{t+1})$ by

$$l^{\Delta}_{[t]}(\boldsymbol{F}_{t+1}) = -\sum_{i=1}^{d} \lambda_i p(\tau_t, T_i)(y(\tau_t, T_i)\Delta t - (T_i - \tau_t)(\Gamma_1 \boldsymbol{F}_{t+1})_i), \quad (63)$$

so that a p-dimensional function replaces a d-dimensional function.

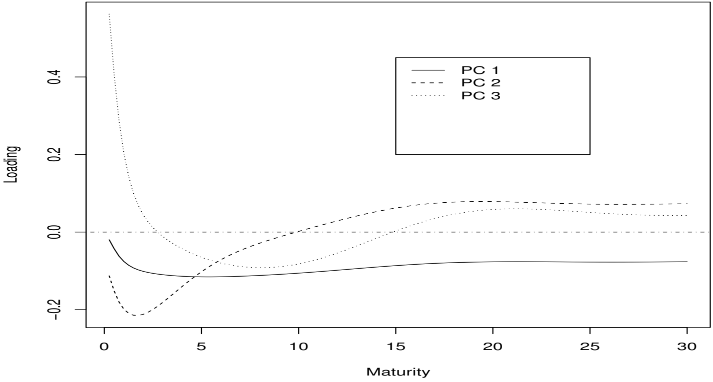
- To calibrate this function, we require an estimate for the matrix Γ_1. This can be obtained from historical time-series data on yield changes by estimating sample principle components.

Example 9.4

- To estimate the Γ_1 matrix of principal component loadings we require longitudinal (time-series) data rather than the cross-sectional data.
- We again analyse Canadian data. Recall that we have data vectors $\boldsymbol{Y}_t = (y(\tau_t, \tau_t + s_1), \ldots, y(\tau_t, \tau_t + s_d))$ of yields for different maturities.

- For simplicity assume that the times-to-maturity $T_1 - \tau_t, \ldots, T_d - \tau_t$ of the bonds in the portfolio correspond exactly to the times to maturity $s_1, \ldots, s_d$ available in the historical dataset.
- Assume also that the risk management horizon Δt is one day.
- We analyse the first differences of the data $\boldsymbol{X}_t = \boldsymbol{Y}_t - \boldsymbol{Y}_{t-1}$ using PCA under an assumption of stationarity.
- In the Canadian dataset we have 2488 days of data from 2 January 2002 to 30 December 2011.
- (Note that a small error is incurred by analysing daily returns of yields with fixed times-to-maturity rather than fixed maturity date.)
- The first principle component explains 87.0% of the variance of the data, the first two components explain 95.9% and the first three components explain 97.5%.

- We choose to work with the first three principal components, meaning that we set $p = 3$ and set the columns of Γ_1 equal to the first three principal component loading vectors.
- These vectors are shown graphically below and lend themselves to a standard interpretation.
- The first principal component has negative loadings for all maturities; the second has negative loadings up to 10 years and positive loadings thereafter; the third has positive loadings for very short maturities (less than 2.5 years) and very long maturities (greater than 15 years) but negative loadings otherwise.
- This suggests that the first principal component can be thought of as inducing a change in the level of all yields, the second induces a change of slope and the third a change in the curvature of the yield curve.


PC 1
PC 2
PC 3
Loading
Maturity
-0.2
0.0
0.2
0.4
0
5
10
15
20
25
30

9.2 Market risk measurement

The goal in this section is to estimate the distribution of

$$L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$$

or a linear or quadratic approximation thereof, where

- $\boldsymbol{X}_{t+1}$ is the vector of risk-factor changes from time t to time $t+1$;
- $l_{[t]}$ is the known loss operator function at time t.

The problem comprises two tasks:

1) on the one hand we have the statistical problem of estimating the distribution of $\boldsymbol{X}_{t+1}$;
2) on the other hand we have the computational or numerical problem of evaluating the distribution of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$.

9.2.1 Conditional and unconditional loss distributions

- Generally, we want to compute conditional measures of risk based on the most recent information about financial markets.
- In this case, the task is to estimate $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, the conditional distribution of risk-factor changes, given $\mathcal{F}_t$, the sigma field representing the available information at time t.
- The conditional loss distribution is the distribution of the loss operator $l_{[t]}(\cdot)$ under $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$, i.e. the distribution with df

$$F_{L_{t+1}|\mathcal{F}_t}(l) = \mathbb{P}(l_{[t]}(\boldsymbol{X}_{t+1}) \le l \,|\, \mathcal{F}_t).$$

- In the unconditional approach we assume that $(\boldsymbol{X}_s)_{s\le t}$ forms a stationary time series, at least in the recent past.
- In this case we can estimate the stationary distribution $F_{\boldsymbol{X}}$ and then evaluate the unconditional loss distribution of $l_{[t]}(\boldsymbol{X})$ where $\boldsymbol{X} \sim F_{\boldsymbol{X}}$. The unconditional loss distribution is thus $F_{L_{t+1}}(l) = \mathbb{P}(l_{[t]}(\boldsymbol{X}) \le l)$.

- The unconditional approach may be appropriate for longer time intervals, or for stress testing during quieter periods.
- If the risk-factor changes form an iid series, we obviously have $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t} = F_{\boldsymbol{X}}$, so that the conditional and unconditional approaches coincide.

9.2.2 Variance-covariance method

- The variance–covariance method is an analytical method in which strong assumptions of (conditional) normality and linearity are made.
- We assume that the conditional distribution of risk-factor changes $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ is a multivariate normal distribution.
- In other words, we assume that $\boldsymbol{X}_{t+1} \,|\, \mathcal{F}_t \sim \mathrm{N}_d(\boldsymbol{\mu}_{t+1}, \Sigma_{t+1})$.
- The estimation of $F_{\boldsymbol{X}_{t+1}|\mathcal{F}_t}$ can be carried out in a number of ways:
  - Fit multivariate ARMA-GARCH model with multivariate normal innovations; use model to derive estimates of $\boldsymbol{\mu}_{t+1}$ and Σ_{t+1}.

- Alternatively use the exponentially weighted moving-average (EWMA) procedure; Σ_{t+1} estimated recursively by $\hat{\Sigma}_{t+1} = \theta \boldsymbol{X}_t \boldsymbol{X}_t' + (1-\theta)\hat{\Sigma}_t$ where θ is a small positive number (typically $\theta \approx 0.04$).

- The second critical assumption in the variance–covariance method is that the linear loss operator is sufficiently accurate. The linear loss operator is a function of the form

$$l_{[t]}^{\Delta}(\boldsymbol{x}) = -(c_t + \boldsymbol{b}_t'\boldsymbol{x})$$

for some constant c_t and constant vector $\boldsymbol{b}_t$, known at time t.

- We infer that, conditional on $\mathcal{F}_t$,

$$L_{t+1}^{\Delta} = l_{[t]}^{\Delta}(\boldsymbol{X}_{t+1}) \sim \mathrm{N}(-c_t - \boldsymbol{b}_t'\boldsymbol{\mu}_{t+1},\ \boldsymbol{b}_t'\Sigma_{t+1}\boldsymbol{b}_t).$$

- Under normality, VaR_α and ES_α may be easily calculated:
  - $\widehat{\mathrm{VaR}}_\alpha = -c_t - \boldsymbol{b}_t'\widehat{\boldsymbol{\mu}}_{t+1} + \sqrt{\boldsymbol{b}_t'\widehat{\Sigma}_{t+1}\boldsymbol{b}_t}\,\Phi^{-1}(\alpha).$
  - $\widehat{\mathrm{ES}}_\alpha = -c_t - \boldsymbol{b}_t'\widehat{\boldsymbol{\mu}}_{t+1} + \sqrt{\boldsymbol{b}_t'\widehat{\Sigma}_{t+1}\boldsymbol{b}_t}\,\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$

Pros and cons, extensions

Pros: In contrast to the methods that follow, variance-covariance offers analytical solution with no simulation.

Cons:
- Assumption of multivariate normality may seriously underestimate the tail of the loss distribution.
- Linearization may be a crude approximation.

Extensions: Instead of assuming normal risk factors, the method could be easily adapted to use multivariate Student t or multivariate hyperbolic risk-factor changes without sacrificing tractability (the method works for all elliptical distributions but linearization is crucial here).

9.2.3 Historical simulation

- Historical simulation is by far the most popular method used by banks for the trading book.
- Instead of estimating the distribution of $l_{[t]}(\boldsymbol{X}_{t+1})$ under an explicit parametric model for $\boldsymbol{X}_{t+1}$, the historical simulation method can be thought of as estimating the distribution of the loss operator under the *empirical distribution* of historical data $\boldsymbol{X}_{t-n+1}, \ldots, \boldsymbol{X}_t$.
- Construct the historically simulated losses (under the current portfolio):

$$\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) : s = t-n+1, \ldots, t\}.$$

- One may apply the linear/quadratic loss operator (if that was already used; avoids revaluation).
- $\tilde{L}_s$ shows what would happen to the current portfolio if the risk-factor change on day s were to recur.
- Use $(\tilde{L}_s)$ to make inferences about the loss distribution and risk measures.

- **Inference about the loss distribution**
  - One could use empirical quantile estimation to estimate VaR_α.
    But: What about precision (sample size; confidence intervals)?
  - Or fit a parametric distribution to the historical losses $L_{t-n+1}, \dots, L_t$ and calculate risk measures from this distribution.
    But: Which distribution to fit (body or tail)?
  - One could use extreme value theory to estimate the tail of the loss distribution and related risk measures based on the historical losses $L_{t-n+1}, \dots, L_t$.

Theoretical justification

If $\boldsymbol{X}_{t-n+1}, \dots, \boldsymbol{X}_t$ are iid or, more generally, stationary, convergence of the empirical distribution to the true distribution is ensured by a suitable version of the Law of Large Numbers (e.g. Glivenko–Cantelli theorem).

Pros and Cons

Pros:
- Easy to implement.
- No statistical estimation of the distribution of $\boldsymbol{X}$ necessary (the empirical df of $\boldsymbol{X}$ is used implicitly).

Cons:
- It may be difficult to collect sufficient quantities of relevant, synchronized data for all risk factors.
- Historical data may not contain examples of extreme scenarios ("driving a car by only looking in the back mirror").

Note:
- The dependence here is given by the empirical df of $\boldsymbol{X}$.
- "Historical simulation method" is a bit of a misnomer; there is no simulation in the sense of random number generation.

In its standard form HS is an unconditional method. There are a number of ways of extending historical simulation to take account of volatility dynamics (filtered HS).

9.2.4 Dynamic Historical Simulation

A univariate approach:

- Assume that $\{\tilde{L}_s = l_{[t]}(\boldsymbol{X}_s) : s = t-n+1, \dots, t\}$ are realizations from a stationary process $(\tilde{L}_s)$ of the form $\tilde{L}_s = \mu_s + \sigma_s Z_s$, where
  - μ_s and σ_s are $\mathcal{F}_{s-1}$-measurable;
  - (Z_s) are $\mathrm{SWN}(0,1)$ innovations with distribution function F_Z.

  Example: ARMA-GARCH model.
- We can easily calculate that for the next loss $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$ ahead

$$F_{L_{t+1}|\mathcal{F}_t}(l) = \mathbb{P}(\mu_{t+1} + \sigma_{t+1} Z_{t+1} \le l \mid \mathcal{F}_t) = F_Z((l - \mu_{t+1})/\sigma_{t+1}).$$

- Writing VaR_α^t for $F^{\leftarrow}_{L_{t+1}|\mathcal{F}_t}(\alpha)$ and ES_α^t for ES, we obtain

$$\mathrm{VaR}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{VaR}_\alpha(Z),$$
$$\mathrm{ES}_\alpha^t = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z),$$

  where Z is a random variable with distribution function F_Z.

- **Estimation**
  - Formal parametric time series modelling to estimate μ_{t+1}, σ_{t+1}, $\mathrm{VaR}_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$.
  - Often $\mu_{t+1} \approx 0$ and can be neglected. We can use EWMA to estimate $\sigma_{t-n+1}, \dots, \sigma_t, \sigma_{t+1}$ and use the standardized residuals $\{\widehat{Z}_s = \tilde{L}_s/\hat{\sigma}_s,\ s = t-n+1, \dots, t\}$ to estimate $\mathrm{VaR}_\alpha(Z)$ and $\mathrm{ES}_\alpha(Z)$.

A multivariate approach:

- We (implictly) assume risk-factor change data $\boldsymbol{X}_{t-n+1}, \dots, \boldsymbol{X}_t$ are realizations from process $(\boldsymbol{X}_s)$ which satisfies

$$\boldsymbol{X}_s = \boldsymbol{\mu}_s + \Delta_s \boldsymbol{Z}_s, \quad \Delta_s = \mathrm{diag}(\sigma_{s,1}, \dots, \sigma_{s,d}),$$

  where $(\boldsymbol{\mu}_s)$ is a process of vectors and (Δ_s) a process of diagonal matrices (all assumed $\mathcal{F}_{s-1}$-measurable) and $(\boldsymbol{Z}_s) \sim \mathrm{SWN}(\mathbf{0}, P)$ for some correlation matrix P.
- The vector $\boldsymbol{\mu}_s$ contains the conditional means and the matrix Δ_s contains the volatilities of the component series at time s.
- An example of a model that fits into this framework is the CCC-GARCH (constant conditional correlation) process.
- The key idea of the method is to apply historical simulation to the unobserved innovations $(\boldsymbol{Z}_s)$.
- The first step is to compute estimates $\{\hat{\boldsymbol{\mu}}_s : s = t-n+1, \dots, t\}$ and $\{\hat{\Delta}_s : s = t-n+1, \dots, t\}$.

- This can be achieved by fitting univariate time series models of ARMA-GARCH type to each of the component series in turn; alternatively we can use the univariate EWMA approach for each series.
- In the second step we construct residuals

$$\{\hat{\boldsymbol{Z}}_s = \hat{\Delta}_s^{-1}(\boldsymbol{X}_s - \hat{\boldsymbol{\mu}}_s) : s = t-n+1, \dots, t\}$$

  and treat these as "observations" of the unobserved innovations.
- We then construct the dataset

$$\{\tilde{L}_s = l_{[t]}(\hat{\boldsymbol{\mu}}_{t+1} + \hat{\Delta}_{t+1}\hat{\boldsymbol{Z}}_s) : s = t-n+1, \dots, t\} \tag{64}$$

  and treat these as observations of $L_{t+1} = l_{[t]}(\boldsymbol{X}_{t+1})$.
- To estimate VaR (or expected shortfall) we can apply simple empirical estimators directly to these data.

9.2.5 Monte Carlo

- Estimate the distribution of $L = \ell_{[t]}(\boldsymbol{X}_{t+1})$ under some explicit parametric model for $\boldsymbol{X}_{t+1}$.
- In contrast to the variance-covariance approach we do not necessarily make the problem analytically tractable by linearizing the loss and making an assumption of normality for the risk factors.
- Instead, make inference about L using simulated risk factor data.

The method

1) Based on the historical risk-factor data $\boldsymbol{X}_{t-n+1}, \dots, \boldsymbol{X}_t$, estimate a suitable statistical model for the risk-factor changes.
2) Simulate N new risk-factor changes $\boldsymbol{X}_{t+1}^{(1)}, \dots, \boldsymbol{X}_{t+1}^{(N)}$ from this model.
3) Construct the simulated losses $L_k = \ell_{[t]}(\boldsymbol{X}_{t+1}^{(k)})$, $k \in \{1, \dots, N\}$.

4) Make inference about the loss distribution F_L and risk measures using L_k, $k \in \{1, \dots, N\}$ (similar possibilities as for the historical simulation method: non-parametric/parametric/EVT).

Pros and Cons

Pros:
- General. Any distribution for $\boldsymbol{X}_{t+1}$ can be taken.

Cons:
- Can be time consuming if loss operator is difficult to evaluate (depends on size and complexity of the portfolio).
- Note that MC approach does not address the problem of determining the distribution of $\boldsymbol{X}_{t+1}$.

9.2.6 Estimating risk measures

Aim: In both the historical simulation and Monte Carlo methods we estimate risk measures using simulated loss data. Let us suppose that we have data $L_1, \ldots, L_n$ from an underlying loss distribution F_L and the aim is to estimate $\mathrm{VaR}_\alpha = q_\alpha(F_L) = F_L^{\leftarrow}(\alpha)$ or $\mathrm{ES}_\alpha = (1-\alpha)^{-1} \int_\alpha^1 q_\theta(F_L)\,\mathrm{d}\theta$. In the book we consider two possibilities:

- L-estimators. These are linear combinations of sample order statistics. Easiest to use notation for lower order statistics $L_{(1)} \leq \cdots \leq L_{(n)}$.
- GPD-based estimators. These are semi-parametric estimators based on GPD approximations described in EVT chapter.

L-estimators:

VaR: $\mathrm{VaR}_\alpha(L) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}$. Replacing F_L by $\hat{F}_L$ we obtain an L-estimator.

$$\begin{aligned}\widehat{\mathrm{VaR}}_\alpha(L) &= \inf\{x \in \mathbb{R} : \hat{F}_L(x) \ge \alpha\} \\ &= \inf\Big\{x \in \mathbb{R} : \sum_{i=1}^n I_{\{L_i \le x\}} \ge \lceil n\alpha \rceil\Big\} \\ &= \inf\Big\{x \in \mathbb{R} : \sum_{i=1}^n I_{\{L_{(i)} \le x\}} \ge \lceil n\alpha \rceil\Big\} = L_{(\lceil n\alpha \rceil)}.\end{aligned}$$

In practice, most software uses an average of two order statistics.

ES: Assume F_L is continuous so that

$$\mathrm{ES}_\alpha(L) = \frac{\mathbb{E}(LI_{\{L > F_L^{\leftarrow}(\alpha)\}})}{1-\alpha} = \frac{\mathbb{E}(LI_{\{L > F_L^{\leftarrow}(\alpha)\}})}{\mathbb{E}(I_{\{L > F_L^{\leftarrow}(\alpha)\}})}.$$

Replacing F_L by $\hat{F}_L$ leads to the canonical estimator

$$\widehat{\mathrm{ES}}_\alpha(L) = \frac{\sum_{i=1}^n L_i I_{\{L_i > \widehat{\mathrm{VaR}}_\alpha(L)\}}}{\sum_{i=1}^n I_{\{L_i > \widehat{\mathrm{VaR}}_\alpha(L)\}}}.$$

GPD-based estimators:

We set a high threshold $u = L_{(n-k)}$ at an order statistic and fit a GPD distribution to the k excess losses over u to obtain maximum likelihood estimates $\hat{\xi}$ and $\hat{\beta}$.

For $k/n > 1 - \alpha$ we can form the risk measure estimates:

$$\widehat{\mathrm{VaR}}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}}\left(\left(\frac{1-\alpha}{k/n}\right)^{-\hat{\xi}} - 1\right)$$

$$\widehat{\mathrm{ES}}_\alpha = \frac{\widehat{\mathrm{VaR}}_\alpha}{1-\hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1-\hat{\xi}}.$$

9.2.7 Losses over several periods and scaling

- **Goal:** Go from single-period risk measure (e.g. one day/one week VaR/ES) to multi-period risk measure using simple formula.
- **Idea:** The loss between today and h periods ahead is

$$\begin{aligned} L_{t+h}^{(h)} &= -(V_{t+h} - V_t) = -(g(\tau_{t+h}, \boldsymbol{Z}_{t+h}) - g(\tau_t, \boldsymbol{Z}_t)) \\ &= -(g(\tau_{t+h}, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1} + \cdots + \boldsymbol{X}_{t+h}) - g(\tau_t, \boldsymbol{Z}_t)) \\ &= L\Big(\sum_{i=1}^{h} \boldsymbol{X}_{t+i}\Big). \end{aligned}$$

- **Question:** How do risk measures scale with h?
- There is no general answer.
- If $\boldsymbol{X}_{t+i} \overset{\text{ind.}}{\sim} \mathrm{N}(\boldsymbol{\mu}, \Sigma)$, then $\boldsymbol{Y} = \sum_{i=1}^{h} \boldsymbol{X}_{t+i} \sim \mathrm{N}(h\boldsymbol{\mu}, h\Sigma)$. Then

$$L_{t+h}^{(h)\Delta} = -g_\tau(\tau_t, \boldsymbol{Z}_t) - \sum_{j=1}^{d} g_{z_j}(\tau_t, \boldsymbol{Z}_t)\Big(\sum_{i=1}^{h} X_{t+i,j}\Big) = -(c_t + \boldsymbol{b}_t'\boldsymbol{Y}).$$

- We infer that $L_{t+h}^{(h)\Delta} \sim \mathrm{N}(-c_t - h\boldsymbol{b}_t'\boldsymbol{\mu},\ h\boldsymbol{b}_t'\Sigma\boldsymbol{b}_t)$.
- If we assume $c_t \approx 0$, $\boldsymbol{\mu} \approx \mathbf{0}$ (typical for daily data) we obtain square-root-of-time scaling formulas for VaR and ES.
- $\mathrm{VaR}_\alpha(L_{t+h}^{(h)\Delta}) = 0 + \sqrt{h\boldsymbol{b}_t'\Sigma\boldsymbol{b}_t}\,\Phi^{-1}(\alpha) = \sqrt{h}\,\mathrm{VaR}_\alpha(L_{t+1}^{\Delta})$.
- $\mathrm{ES}_\alpha(L_{t+h}^{(h)\Delta}) = 0 + \sqrt{h\boldsymbol{b}_t'\Sigma\boldsymbol{b}_t}\,\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha} = \sqrt{h}\,\mathrm{ES}_\alpha(L_{t+1}^{\Delta})$.
- Note the many underlying assumptions:
  - $\boldsymbol{X}_{t+i}$ independent;
  - $\boldsymbol{X}_{t+i}$ multivariate normal;
  - The linearized loss provides a sufficiently good approximation to the true loss distribution.
- Note also that we have only considered the scaling of unconditional risk measures.

9.3 Backtesting

- Backtesting is the practice of evaluating risk measurement procedures by comparing ex ante estimates/forecasts of risk measures with ex post realized losses and gains.
- It allows us to evaluate whether a model and estimation procedure produce credible risk measure estimates.

9.3.1 Violation-based tests for VaR

- Let VaR_α^t denote the α-quantile of the conditional loss distribution $F_{L_{t+1}|\mathcal{F}_t}$ and consider the event indicator variable $I_{t+1} = I_{\{L_{t+1} > \mathrm{VaR}_\alpha^t\}}$.
- The event $\{L_{t+1} > \mathrm{VaR}_\alpha^t\}$ is a VaR violation or exception.
- Assuming a continuous loss distribution, we have, by definition of the quantile,

$$\mathbb{E}(I_{t+1} \mid \mathcal{F}_t) = \mathbb{P}(L_{t+1} > \mathrm{VaR}_\alpha^t \mid \mathcal{F}_t) = 1 - \alpha\,, \tag{65}$$

- I_{t+1} is a Bernoulli variable with event probability $(1-\alpha)$.
- Moreover, the sequence of VaR exception indicators (I_t) is an iid sequence.
- The sum of exception indicators is binomially distributed:
$$M = \sum_{t=1}^{m} I_t \sim B(m, 1-\alpha).$$
- Assume exceptions occur at times $1 \leq T_1 < \cdots < T_M \leq m$ and set $T_0 = 0$. The spacings $S_j = T_j - T_{j-1}$ will be independent geometrically distributed rvs with mean $1/(1-\alpha)$, so that
$$\mathbb{P}(S_j = k) = \alpha^{k-1}(1-\alpha), \quad k \in \mathrm{N}\,.$$
- Both of these properties are testable in empirical data.
- For small event probability $1-\alpha$, the Bernoulli Trials Process may be well approximated by a Poisson process.
- Also for small $1-\alpha$ the geometric distribution may be approximated by an exponential distribution.

- Suppose we estimate VaR_α^t at time point t by $\widehat{\mathrm{VaR}}_\alpha^t$.
- In a backtest we consider empirical indicator variables

$$\widehat{I}_{t+1} = I_{\{L_{t+1} > \widehat{\mathrm{VaR}}_\alpha^t\}}.$$

- The sequence $(\widehat{I}_t)_{1 \le t \le m}$ should behave like a realization from a Bernoulli trials process with event probability $(1-\alpha)$.
- To test binomial behaviour for number of violations we compute a score test statistic

$$Z_m = \frac{(\sum_{t=1}^m \hat{I}_t) - m(1-\alpha)}{\sqrt{m\alpha(1-\alpha)}}$$

and reject Bernoulli hypothesis at 5% level if $Z_m > \Phi^{-1}(0.95)$.
- Exponential spacings can be tested numerically or with a Q-Q plot.

9.3.2 Violation-based tests of expected shortfall

- Let ES_α^t denote the one-period expected shortfall and $\widehat{\mathrm{ES}}_\alpha^t$ its estimate.

- Assume (L_t) follows a model of the form $L_t = \sigma_t Z_t$, where σ_t is a function of $\mathcal{F}_{t-1}$ and the (Z_t) are $\mathrm{SWN}(0,1)$ innovations.
- Then we can define a process (K_t) by

$$K_{t+1} = \frac{(L_{t+1} - \mathrm{ES}_\alpha^t)}{\mathrm{ES}_\alpha^t} I_{\{L_{t+1} > \mathrm{VaR}_\alpha^t\}} = \frac{Z_{t+1} - \mathrm{ES}_\alpha(Z)}{\mathrm{ES}_\alpha(Z)} I_{\{Z_{t+1} > q_\alpha(Z)\}},$$

  and note that it is a zero-mean iid sequence.
- This suggests we form violation residuals of the form

$$\widehat{K}_{t+1} = \frac{(L_{t+1} - \widehat{\mathrm{ES}}_\alpha^t)}{\widehat{\mathrm{ES}}_\alpha^t} \widehat{I}_{t+1}. \tag{66}$$

- We test for mean-zero behaviour using a bootstrap test on the non-zero violation residuals (McNeil and Frey (2000)).

9.3.3 Elicitability and comparison of risk measure estimates

- The elicitability concept has been introduced into the backtesting literature by Gneiting (2011); see also important papers by Bellini and

Bignozzi (2013) and Ziegel (2014).

- A key concept is that of a scoring function $S(y,l)$ which measures the discrepancy between a forecast y and a realized loss l.
- Forecasts are made by applying real-valued statistical functionals T (such as mean, median or other quantile) to the distribution of the loss F_L to obtain the forecast $y = T(F_L)$.
- Suppose that for some class of loss distribution functions a real-valued statistical functional T satisfies

$$T(F_L) = \arg\min_{y\in\mathbb{R}} \int_{\mathbb{R}} S(y,l)\,\mathrm{d}F_L(l) = \arg\min_{y\in\mathbb{R}} \mathbb{E}(S(y,L)) \qquad (67)$$

for a scoring function S and any loss distribution F_L in that class.

- Suppose moreover that $T(F_L)$ is the unique minimizing value.
- The scoring function S is said to be strictly consistent for T.
- The functional $T(F_L)$ is said to be elicitable.
- Note that (67) implies that

$$\frac{\mathrm{d}}{\mathrm{d}y}\mathbb{E}(S(y,L))\Big|_{y=T(F_L)} = \int_{\mathbb{R}} \frac{\mathrm{d}}{\mathrm{d}y} S(y,l)\mathrm{d}F_L(l)\Big|_{y=T(F_L)} = \mathbb{E}(h(T(F_L),L)) = 0$$

where h is the derivative of the scoring function.

- The VaR risk measure corresponds to $T(F_L) = F_L^{\leftarrow}(\alpha)$. For any $0 < \alpha < 1$ this functional is elicitable for strictly increasing distribution functions. The scoring function

$$S_\alpha^q(y,l) = |1_{\{l \leq y\}} - \alpha||l - y| \tag{68}$$

is strictly consistent for T.

- The α-expectile of L is defined to be the risk measure that minimizes $\mathbb{E}\left(S^e_\alpha(y,L)\right)$ where the scoring function is
$$S^e_\alpha(y,l) = |1_{\{l\le y\}} - \alpha|(l-y)^2. \tag{69}$$
This risk measure is elicitable by definition.
- Bellini and Bignozzi (2013) and Ziegel (2014) show that a risk measure is coherent and elicitable if and only if it is the α-expectile risk measure for $\alpha \ge 0.5$; see also Weber (2006). Expected shortfall is not elicitable.
- VaR^t_α minimizes
$$\mathbb{E}\left(S^q_\alpha(\mathrm{VaR}^t_\alpha, L_{t+1}) \mid \mathcal{F}_t\right)$$
for the scoring function in (68). We refer to $S^q_\alpha(\mathrm{VaR}^t_\alpha, L_{t+1})$ as a (theoretical) VaR score.
- Assume VaR^t_α is replaced by an estimate at each time point and consider the VaR scores $\{S^q_\alpha(\widehat{\mathrm{VaR}}^t_\alpha, L_{t+1}) : t = 1,\dots,m\}$
- These can be used to address questions of relative model performance.

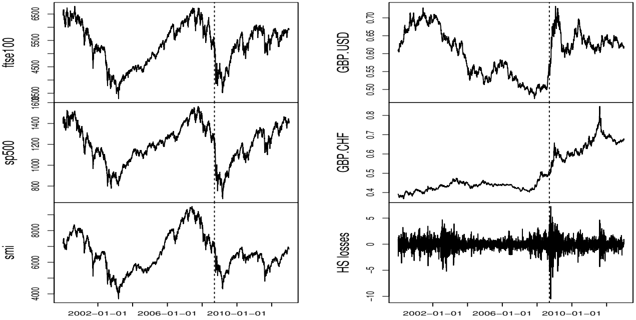
- The statistic
$$Q_0 = \frac{1}{m}\sum_{t=1}^{m} S_\alpha^q(\widehat{\mathrm{VaR}}_\alpha^t, L_{t+1})$$
can be used as a measure of relative model performance.
- If two models A and B deliver VaR estimates $\{\widehat{\mathrm{VaR}}_\alpha^{t\,A}, t = 1, \ldots, m\}$ and $\{\widehat{\mathrm{VaR}}_\alpha^{t\,B}, t = 1, \ldots, m\}$ with corresponding average scores Q_0^A and Q_0^B, then we expect the better model to give estimates closer to the true VaR numbers and thus a value of Q_0 that is lower.
- Of course, the power to discriminate between good models and inferior models will depend on the length of the backtest.

9.3.4 Empirical comparison of methods using backtesting concepts

- We apply various VaR estimation methods to the portfolio of a hypothetical investor in international equity indexes.
- The investor is assumed to have domestic currency sterling (GBP) and to invest in the Financial Times 100 Shares Index (FTSE 100), the Standard & Poor's 500 (S&P 500) and the Swiss Market Index (SMI).
- The portfolio is influenced by five risk factors.
- On any day t we standardize the total portfolio value V_t in sterling to be one and assume portfolio weights are 30%, 40% and 30%, respectively.
- The loss operator and linear loss operator are:

$$l_{[t]}(\boldsymbol{x}) = 1 - (0.3e^{x_1} + 0.4e^{x_2+x_4} + 0.3e^{x_3+x_5})$$
$$l_{[t]}^{\Delta}(\boldsymbol{x}) = -(0.3x_1 + 0.4(x_2 + x_4) + 0.3(x_3 + x_5))$$

- x_1, x_2 and x_3 represent log-returns on the three indexes and x_4 and x_5 are log-returns on the GBP/USD and GBP/CHF exchange rates.



- The final picture shows the corresponding historical simulation data. The vertical dashed line is Lehman Brothers bankruptcy.

Estimation methods:

VC. The variance–covariance method assuming multivariate Gaussian risk-factor changes and using the multivariate EWMA method to estimate the conditional covariance matrix of risk-factor changes.

HS. The standard unconditional historical simulation method.

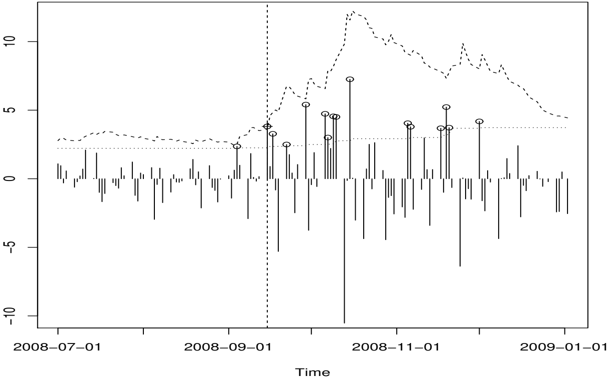
HS-GARCH. The univariate dynamic approach to historical simulation in which a $\mathrm{GARCH}(1,1)$ model with a constant conditional mean term and Gaussian innovations is fitted to the historically simulated losses to estimate the volatility of the next day's loss.

HS-GARCH-t. A similar method to HS-GARCH but Student t innovations are assumed in the GARCH model.

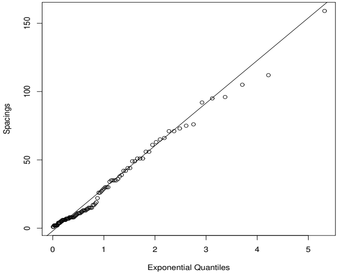
HS-MGARCH. The multivariate dynamic approach to historical simulation in which $\mathrm{GARCH}(1,1)$ models with constant conditional mean terms are fitted to each time series of risk-factor changes to estimate volatilities.

Year	2005	2006	2007	2008	2009	2010	2011	2012	All
Trading days	258	257	258	259	258	259	258	258	2065
Results for 95% VaR									
Expected no. of violations	13	13	13	13	13	13	13	13	103
VC	8	16	17	19	13	15	14	14	116
HS	0	6	28	49	19	6	10	1	119
HS-GARCH	9	13	22	22	13	14	9	15	117
HS-GARCH-t	9	14	23	22	14	15	10	15	122
HS-MGARCH	5	14	21	19	12	9	11	12	103
Results for 99% VaR									
Expected no. of violations	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	21
VC	2	8	8	8	2	4	5	6	43
HS	0	0	10	22	2	0	2	0	36
HS-GARCH	2	8	8	10	5	4	3	3	43
HS-GARCH-t	2	8	6	8	1	4	2	1	32
HS-MGARCH	0	4	4	5	0	1	2	1	17

- The HS method does not react to changing volatility:



- Dotted line is HS; dashed line is HS-MGARCH; vertical line is Lehmann.
- Circle is VaR violation for HS; cross is VaR violation for HS-MGARCH.
- Q-Q plot of spacings between exceptions (for HS-MGARCH):



- Violation residual test for ES (n: number of VaR violations):

	$\mathrm{ES}_{0.95}$	n	$\mathrm{ES}_{0.99}$	n
VC	0.00	116	0.05	43
HS	0.02	119	0.25	36
HS-GARCH	0.00	117	0.05	43
HS-GARCH-t	0.12	122	0.68	32
HS-MGARCH	0.99	103	0.55	17

9.3.5 Backtesting the predictive distribution

- As well as backtesting VaR and expected shortfall we can also devise tests that assess the overall quality of the estimated conditional loss distribution, or its tail.
- If L_{t+1} is a random variable with (continuous) distribution function $F_{L_{t+1}|\mathcal{F}_t}$, then $U_{t+1} = F_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$ is uniform (probability transform).
- In actual applications we estimate $F_{L_{t+1}|\mathcal{F}_t}$ from data up to time t and we backtest our estimates by forming $\widehat{U}_{t+1} = \widehat{F}_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$ on day $t+1$.
- Suppose we estimate the predictive distribution on days $t = 0, \ldots, n-1$ and form backtesting data $\widehat{U}_1, \ldots, \widehat{U}_n$; we expect these to behave like a sample of iid uniform data.
- The distributional assumption can be assessed by standard goodness-of-fit tests like the Kolmogorov–Smirnov test.

References

Bellini, F. and Bignozzi, V. (2013), Elicitable risk measures, Working paper, available at SSRN: http://ssrn.com/abstract=2334746.

Gneiting, T. (2011), Making and Evaluating Point Forecasts, *Journal of the American Statistical Association*, 106(494), 746–762.

McNeil, A. J. and Frey, R. (2000), Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach, *Journal of Empirical Finance*, 7, 271–300.

Nelson, C. R. and Siegel, A. F. (1987), Parsimonious modeling of yield curves, *Journal of Business*, 60, 473–489.

Weber, S. (2006), Distributions-invariant risk measures, information, and dynamic consistency, *Mathematical Finance*, 16(2), 419–441.

Ziegel, J. F. (2014), Coherence and elicitability, *Mathematical Finance*, to appear.