

# 9 Market Risk

9.1 Risk factors and mapping

9.2 Market risk measurement

9.3 Backtesting

# 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  $V_t$  of 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

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

where  $\tau_t$  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.
- 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  $\Delta t$  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  $\mathbf{X}_{t-n}, \dots, \mathbf{X}_{t-1}, \mathbf{X}_t$  and these are used to model the behaviour of  $\mathbf{X}_{t+1}$ .
- We have

$$\begin{aligned} L_{t+1} &= -(V_{t+1} - V_t) \\ &= -(g(\tau_{t+1}, \mathbf{Z}_{t+1}) - g(\tau_t, \mathbf{Z}_t)) \\ &= -(g(\tau_t + \Delta t, \mathbf{Z}_t + \mathbf{X}_{t+1}) - g(\tau_t, \mathbf{Z}_t)). \end{aligned} \quad (51)$$

- 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

$$l_{[t]}(\mathbf{x}) = -(g(\tau_t + \Delta t, \mathbf{z}_t + \mathbf{x}) - g(\tau_t, \mathbf{z}_t)), \quad (52)$$

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  $\mathbf{X}_{t+1}$ .
- Generally we consider the **conditional** distribution of  $L_{t+1}$  given history  $\mathcal{F}_t$  up to and including time  $t$ .
- Alternatively we can consider the **unconditional** distribution under assumption that  $(\mathbf{X}_t)$  form stationary time series.

## 9.1.2 Delta and delta–gamma approximations

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

$$g(\tau_t + \Delta t, \mathbf{z}_t + \mathbf{x}) \approx g(\tau_t, \mathbf{z}_t) + g_\tau(\tau_t, \mathbf{z}_t)\Delta t + \sum_{i=1}^d g_{z_i}(\tau_t, \mathbf{z}_t)x_i, \quad (53)$$

where the  $\tau$ -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 (52) by the linear loss operator at time  $t$  given by

$$l_{[t]}^\Delta(\mathbf{x}) := -\left(g_\tau(\tau_t, \mathbf{z}_t)\Delta t + \sum_{i=1}^d g_{z_i}(\tau_t, \mathbf{z}_t)x_i\right). \quad (54)$$

- Note that, when working with a short time horizon  $\Delta t$ , the term  $g_\tau(\tau_t, \mathbf{z}_t)\Delta t$  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

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

$\Phi$  is standard normal df,  $r$  represents risk-free interest rate,  $\sigma$  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:

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

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

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

where  $g_{\tau}$ ,  $g_{z_i}$  denote partial derivatives.

- $\Delta t$  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

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

in terms of the derivatives of the BS formula or [the Greeks](#).

- ▶  $C_S^{BS}$  is known as the [delta](#) of the option.
- ▶  $C_{\sigma}^{BS}$  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 (53).
- Let  $\delta(\tau_t, \mathbf{z}_t) = (g_{z_1}(\tau_t, \mathbf{z}_t), \dots, g_{z_d}(\tau_t, \mathbf{z}_t))'$  be the first-order partial derivatives of the mapping with respect to the risk factors.

- Let  $\omega(\tau_t, \mathbf{z}_t) = (g_{z_1\tau}(\tau_t, \mathbf{z}_t), \dots, g_{z_d\tau}(\tau_t, \mathbf{z}_t))'$  denote the mixed partial derivatives with respect to time and the risk factors.
- Let  $\Gamma(\tau_t, \mathbf{z}_t)$  denote the matrix with  $(i, j)$ th element given by  $g_{z_i z_j}(\tau_t, \mathbf{z}_t)$ ; 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

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

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

- This leaves us with the **quadratic loss operator**

$$l_{[t]}^{\Delta\Gamma}(\mathbf{x}) = -(g_{\tau}(\tau_t, \mathbf{z}_t)\Delta t + \boldsymbol{\delta}(\tau_t, \mathbf{z}_t)' \mathbf{x} + \frac{1}{2} \mathbf{x}' \Gamma(\tau_t, \mathbf{z}_t) \mathbf{x}) \quad (55)$$

which is more accurate than the linear loss operator (54).

### Example 9.2 (European call option)

The quadratic loss operator is

$$l_{[t]}^{\Delta\Gamma}(\mathbf{x}) = l_{[t]}^{\Delta}(\mathbf{x}) - 0.5 \left( C_{SS}^{BS} S_t^2 x_1^2 + C_{rr}^{BS} x_2^2 + C_{\sigma\sigma}^{BS} x_3^2 \right) \\ - \left( C_{Sr}^{BS} S_t x_1 x_2 + C_{S\sigma}^{BS} S_t x_1 x_3 + C_{r\sigma}^{BS} x_2 x_3 \right) .$$

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  $T \rightarrow p(t, T)$  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  $T \rightarrow p(t, T)$  is known at time  $t$ .
- However the future term structure  $T \rightarrow p(t + x, T)$  for  $x > 0$  is not known at time  $t$  and must be modelled stochastically.

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

$$y(t, T) = -\frac{\ln p(t, T)}{T - t}. \quad (56)$$

- 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  $T \rightarrow y(t, T)$  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  $i = 1, \dots, d$ . Assume  $p(T_i, T_i) = 1$  for all  $i$ .

- By  $\lambda_i$  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

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

- We set

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

where  $\Delta t$  is risk management time horizon in years.

- We obtain a mapping of the form

$$V_t = V(\tau_t) = g(\tau_t, \mathbf{Z}_t) = \sum_{i=1}^d \lambda_i \exp(-(T_i - \tau_t)Z_{t,i}). \quad (57)$$

## The loss operator and its approximations

- The portfolio loss is

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

- Reverting to standard bond pricing notation the loss operator is

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

where  $x_i$  represents the change in yield of the  $i$ th bond.

- The first derivatives of the mapping function (57) are

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

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

- Inserting these in (54) and reverting to standard bond pricing notation we obtain

$$l_{[t]}^{\Delta}(\mathbf{x}) = - \sum_{i=1}^d \lambda_i p(\tau_t, T_i) \left( y(\tau_t, T_i) \Delta t - (T_i - \tau_t) x_i \right), \quad (58)$$

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

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

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



- The quadratic loss operator (55) is

$$l_{[t]}^{\Delta\Gamma}(\boldsymbol{x}) = - \sum_{i=1}^d \lambda_i p(\tau_t, T_i) \left( y(\tau_t, T_i) \Delta t - (T_i - \tau_t) x_i + \frac{1}{2} (T_i - \tau_t)^2 x_i^2 \right). \quad (59)$$

## 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

$$l_{[t]}^{\Delta}(x) = -V_t \left( A_t \Delta t - D_t x \right), \quad (60)$$

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  $l_{[t]}(x) \approx V_t D_t x$ .
- **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 (59) becomes

$$l_{[t]}^{\Delta\Gamma}(x) = -V_t \left( A_t \Delta t - D_t x + \frac{1}{2} C_t x^2 \right), \quad (61)$$

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  $\Delta t$ , 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  $X_{t+1}$  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

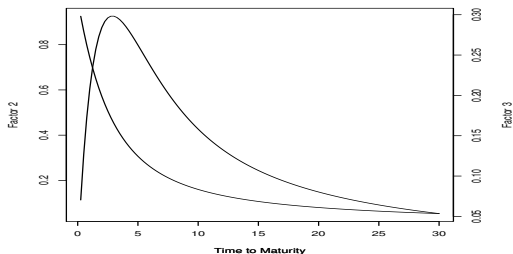
- We assume that at time  $t$  the yield curve can be modelled by

$$y(\tau_t, T) = Z_{t,1} + k_2(T - \tau_t, \eta_t)Z_{t,2} + k_3(T - \tau_t, \eta_t)Z_{t,3}, \quad (62)$$

where the functions  $k_2$  and  $k_3$  are given by

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

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



- $\eta$  is an extra tuning parameter to improve fit.
- There are other simple factor models including the [Svensson model](#).
- Clearly  $\lim_{s \rightarrow \infty} k_2(s, \eta) = \lim_{s \rightarrow \infty} k_3(s, \eta) = 0$  while  $\lim_{s \rightarrow 0} k_2(s, \eta) = 1$  and  $\lim_{s \rightarrow 0} k_3(s, \eta) = 0$ .
- It follows that

$$\lim_{T \rightarrow \infty} y(\tau_t, T) = Z_{t,1},$$

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

$$\lim_{T \rightarrow \tau_t} y(\tau_t, T) - \lim_{T \rightarrow \infty} y(\tau_t, T) = Z_{t,2}.$$

- $Z_{t,3}$  has an interpretation as a [curvature factor](#).



- Using (62), the portfolio mapping (57) becomes

$$V_t = g(\tau_t, \mathbf{Z}_t) = \sum_{i=1}^d \lambda_i \exp \left( - (T_i - \tau_t) \sum_{j=1}^3 k_j (T_i - \tau_t, \eta_t) Z_{t,j} \right),$$

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

- It is then straightforward to derive the loss operator  $l_{[t]}(\mathbf{x})$  or its linear version  $l_{[t]}^{\Delta}(\mathbf{x})$  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  $z_t$  for the risk factors  $\mathbf{Z}_t$ . We have to overcome the fact that **the Nelson-Siegel factors  $\mathbf{Z}_t$  are not directly observed** at time  $t$ . Instead **they have to be estimated** from observable yield curve data.
- Let  $\mathbf{Y}_t = (y(\tau_t, \tau_t + s_1), \dots, y(\tau_t, \tau_t + s_m))'$  denote the data vector at time  $t$ , containing the yields for  $m$  different times to maturity,  $s_1, \dots, s_m$ , where  $m$  is large.

- This is assumed to follow the factor model

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

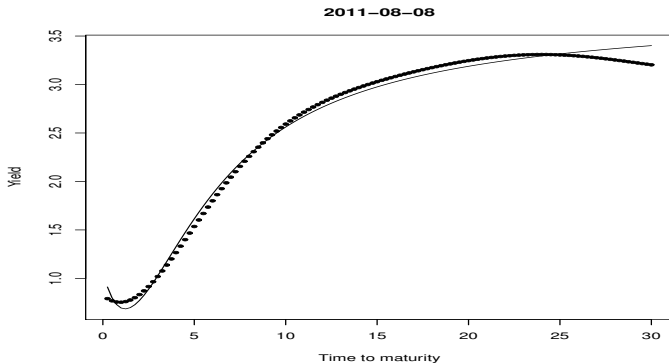
where  $B_t \in \mathbb{R}^{m \times 3}$  is the matrix with  $i$ th row  $(1, k_2(s_i, \eta_t), k_3(s_i, \eta_t))$  and  $\boldsymbol{\varepsilon}_t \in \mathbb{R}^m$  is an error vector.

- For a given value of  $\eta_t$  the estimation of  $\mathbf{Z}_t$  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  $\eta_t$  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 values 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**  $\mathbf{X}_{t+1}$  rather than the risk factors  $\mathbf{Z}_t$ .

- We recall that PCA can be used to construct factor models of the form

$$\mathbf{X}_{t+1} = \boldsymbol{\mu} + \Gamma_1 \mathbf{F}_{t+1} + \boldsymbol{\varepsilon}, \quad (63)$$

where  $\mathbf{F}_{t+1}$  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  $\mathbf{X}_{t+1}$  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  $\mathbf{X}_{t+1} \approx \Gamma_1 \mathbf{F}_{t+1}$ .
- In the case of the **linear loss operator for the bond portfolio** in (58) we

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

$$l_{[t]}^{\Delta}(\mathbf{F}_{t+1}) = - \sum_{i=1}^d \lambda_i p(\tau_t, T_i) (y(\tau_t, T_i) \Delta t - (T_i - \tau_t)(\Gamma_1 \mathbf{F}_{t+1})_i), \quad (64)$$

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

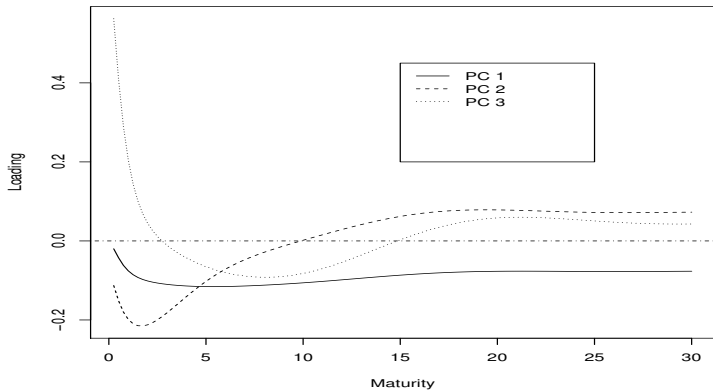
- To calibrate this function, **we require an estimate for the matrix  $\Gamma_1$** . This can be obtained from historical time-series data on yield changes by estimating sample principle components.

### Example 9.4

- To estimate the  $\Gamma_1$  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  $\mathbf{Y}_t = (y(\tau_t, \tau_t + s_1), \dots, y(\tau_t, \tau_t + s_d))$  of yields for different maturities.

- For simplicity assume that the times-to-maturity  $T_1 - \tau_t, \dots, T_d - \tau_t$  of the bonds in the portfolio correspond exactly to the times to maturity  $s_1, \dots, s_d$  available in the historical dataset.
- Assume also that the risk management horizon  $\Delta t$  is one day.
- We analyse the first differences of the data  $\mathbf{X}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$  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  $\Gamma_1$  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.





## 9.2 Market risk measurement

The goal in this section is to estimate the distribution of

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

or a linear or quadratic approximation thereof, where

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

## 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  $F_{\mathbf{X}_{t+1}|\mathcal{F}_t}$ , 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  $F_{\mathbf{X}_{t+1}|\mathcal{F}_t}$ , that is, the distribution with distribution function

$$F_{L_{t+1}|\mathcal{F}_t} = P(l_{[t]}(\mathbf{X}_{t+1}) \leq l \mid \mathcal{F}_t).$$

- In the unconditional approach we assume that the process of risk-factor changes  $(\mathbf{X}_s)_{s \leq t}$  forms a stationary time series, at least in the recent past.

- In this case we can estimate the stationary distribution  $F_{\mathbf{X}}$  and then evaluate the unconditional loss distribution of  $l_{[t]}(\mathbf{X})$  where  $\mathbf{X}$  represents a generic random vector in  $\mathbb{R}^d$  with distribution function  $F_{\mathbf{X}}$ .
- The unconditional loss distribution is thus the distribution of the loss operator  $l_{[t]}(\cdot)$  under  $F_{\mathbf{X}}$ .
- 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  $F_{\mathbf{X}_{t+1}|\mathcal{F}_t} = F_{\mathbf{X}}$ , 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**  $F_{\mathbf{X}_{t+1}|\mathcal{F}_t}$  **is a multivariate normal distribution.**
- In other words, we assume that  $\mathbf{X}_{t+1} \mid \mathcal{F}_t \sim N_d(\boldsymbol{\mu}_{t+1}, \Sigma_{t+1})$ .
- The estimation of  $F_{\mathbf{X}_{t+1}|\mathcal{F}_t}$  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  $\Sigma_{t+1}$ .
  - ▶ Alternatively use the **exponentially weighted moving-average (EWMA)** procedure;  $\Sigma_{t+1}$  estimated recursively by

$$\hat{\Sigma}_{t+1} = \theta \mathbf{X}_t \mathbf{X}_t' + (1 - \theta) \hat{\Sigma}_t$$

where  $\theta$  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

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

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

- We have seen examples, including derivative and bond portfolios.
- We infer that, conditional on  $\mathcal{F}_t$ ,

$$L_{t+1}^{\Delta} = l_{[t]}^{\Delta}(\mathbf{X}_{t+1}) \sim N(-c_t - \mathbf{b}_t' \boldsymbol{\mu}_{t+1}, \mathbf{b}_t' \Sigma_{t+1} \mathbf{b}_t). \quad (65)$$

- Value-at-Risk and expected shortfall may be easily calculated for a normal loss distribution:
  - ▶  $\widehat{\text{VaR}}_{\alpha} = -c_t - \mathbf{b}_t' \hat{\boldsymbol{\mu}}_{t+1} + \sqrt{\mathbf{b}_t' \hat{\Sigma}_{t+1} \mathbf{b}_t} \Phi^{-1}(\alpha).$
  - ▶  $\widehat{\text{ES}}_{\alpha} = -c_t - \mathbf{b}_t' \hat{\boldsymbol{\mu}}_{t+1} + \sqrt{\mathbf{b}_t' \hat{\Sigma}_{t+1} \mathbf{b}_t} \frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}.$

## Pros and cons, extensions

- Pros:** In **contrast** to the methods that follow, variance-covariance offers **analytical solution** with no simulation.
- Cons:**
- ▶ **Linearization** may be a crude approximation.
  - ▶ Assumption of **multivariate normality** may seriously underestimate the **tail** of the loss distribution.
- 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.

- **The Idea**

Instead of estimating the distribution of  $l_{[t]}(\mathbf{X}_{t+1})$  under an explicit parametric model for  $\mathbf{X}_{t+1}$ , the historical-simulation method can be thought of as estimating the distribution of the loss operator under the *empirical distribution* of historical data  $\mathbf{X}_{t-n+1}, \dots, \mathbf{X}_t$ .

- We construct a univariate dataset by applying the loss operator to historical observations of the risk-factor change vector to get a set of historically simulated losses:

$$\{\tilde{L}_s = l_{[t]}(\mathbf{X}_s) : s = t - n + 1, \dots, t\}. \quad (66)$$

- To avoid full revaluation, we may apply linear/quadratic loss operator.

- The values  $\tilde{L}_s$  show what would happen to the current portfolio if the risk-factor changes on day  $s$  were to recur.
- We make inferences about the loss distribution and risk measures using these historically simulated loss data.
- **Inference about the loss distribution**
  - ▶ Use empirical quantile estimation to estimate the VaR directly from the simulated data.  
But: What about precision (sample size; confidence intervals)?
  - ▶ 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)?
  - ▶ Powerful: Use techniques of 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  $X_{t-n+1}, \dots, X_t$  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  $X$  necessary (the **empirical df** of  $X$  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  $X$ .

- ▶ “Historical simulation method” is actually a misnomer; there is no simulation in the sense of random number generation.

**Extensions:** 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 the historical simulation data  $\{\tilde{L}_s = l_{[t]}(\mathbf{X}_s) : s = t - n + 1, \dots, t\}$  are realizations from a stationary process  $(\tilde{L}_s)$ .
- To predict  $L_{t+1} = l_{[t]}(\mathbf{X}_{t+1})$ , the next random variable in this process, we assume  $(\tilde{L}_s)$  follows a model of the form  $\tilde{L}_s = \mu_s + \sigma_s Z_s$ , where
  - ▶  $\mu_s$  and  $\sigma_s$  are  $\mathcal{F}_{s-1}$ -measurable;
  - ▶  $(Z_s)$  are SWN(0, 1) innovations with distribution function  $F_Z$ .

- An example would be a **GARCH model with ARMA mean structure**.
- Writing  $\text{VaR}_\alpha^t$  for the  $\alpha$ -quantile of  $F_{L_{t+1}|\mathcal{F}_t}$ ,  $\text{ES}_\alpha^t$  for the corresponding expected shortfall, we can obtain formulas:

$$\begin{aligned}\text{VaR}_\alpha^t &= \mu_{t+1} + \sigma_{t+1} \text{VaR}_\alpha(Z), \\ \text{ES}_\alpha^t &= \mu_{t+1} + \sigma_{t+1} \text{ES}_\alpha(Z),\end{aligned}$$

where  $Z$  is a random variable with distribution function  $F_Z$ .

### ■ Estimation Options:

- ▶ Formal **parametric time series modelling** to estimate  $\mu_{t+1}$ ,  $\sigma_{t+1}$ ,  $\text{VaR}_\alpha(Z)$  and  $\text{ES}_\alpha(Z)$ .
- ▶ 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 carry out separate analysis of the **residuals**  $\{\hat{Z}_s = \tilde{L}_s / \hat{\sigma}_s, s = t - n + 1, \dots, t\}$  to estimate  $\text{VaR}_\alpha(Z)$  and  $\text{ES}_\alpha(Z)$ .

## A multivariate approach:

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

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

where  $(\boldsymbol{\mu}_s)$  is a process of vectors and  $(\Delta_s)$  a process of diagonal matrices (all assumed  $\mathcal{F}_{s-1}$ -measurable) and  $(\mathbf{Z}_s) \sim \text{SWN}(\mathbf{0}, P)$  for some correlation matrix  $P$ .

- The vector  $\boldsymbol{\mu}_s$  contains the conditional means and the matrix  $\Delta_s$  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  $(\mathbf{Z}_s)$ .

- 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

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

and treat these as “observations” of the unobserved innovations.

- We then construct the dataset

$$\{\tilde{L}_s = l_{[t]}(\hat{\boldsymbol{\mu}}_{t+1} + \hat{\Delta}_{t+1}\hat{\mathbf{Z}}_s) : s = t - n + 1, \dots, t\} \quad (67)$$

and treat these as observations of  $L_{t+1} = l_{[t]}(\mathbf{X}_{t+1})$ .

- To estimate VaR (or expected shortfall) we can apply simple empirical estimators directly to these data.

## 9.2.5 Monte Carlo

### Idea

- We estimate the distribution of  $L = \ell_{[t]}(\mathbf{X}_{t+1})$  under some **explicit parametric model** for  $\mathbf{X}_{t+1}$ .
- 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 **Monte Carlo** methods, which involves **simulating** new risk factor data.

### The method

- 1) Based on the historical risk-factor data  $\mathbf{X}_{t-n+1}, \dots, \mathbf{X}_t$ , **estimate** a suitable statistical **model** for the **risk-factor changes**.
- 2) Simulate  $N$  new data vectors  $\mathbf{X}_{t+1}^{(1)}, \dots, \mathbf{X}_{t+1}^{(N)}$  from this model.

- 3) Construct the Monte Carlo simulated losses  $L_k = \ell_{[t]}(\mathbf{X}_{t+1}^{(k)})$ ,  $k \in \{1, \dots, N\}$ .
- 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:** ▶ Any distribution for  $\mathbf{X}_{t+1}$  can be taken  $\Rightarrow$  general

**Cons:** ▶ Can be time consuming if loss operator is difficult to evaluate, which depends on size and complexity of the portfolio.

▶ Note that MC approach does not address the problem of determining the distribution of  $\mathbf{X}_{t+1}$ .

## 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  $L_1, \dots, L_n$  from an underlying loss distribution  $F_L$  and the aim is to estimate  $\text{VaR}_\alpha = q_\alpha(F_L) = F_L^{\leftarrow}(\alpha)$  or  $\text{ES}_\alpha = (1 - \alpha)^{-1} \int_\alpha^1 q_\theta(F_L) d\theta$ . 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  $L_{(1)} \leq \dots \leq L_{(n)}$ .
- **GPD-based estimators.** These are semi-parametric estimators based on GPD approximations described in EVT chapter.

### L-estimators:

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



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

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

**ES:** Assume  $F_L$  is continuous so that

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

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

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

### 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:

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

## 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

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

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

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

- We infer that  $L_{t+h}^{(h)\Delta} \sim N(-c_t - h\mathbf{b}'_t\boldsymbol{\mu}, h\mathbf{b}'_t\Sigma\mathbf{b}_t)$ .
- 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.
- $\text{VaR}_\alpha(L_{t+h}^{(h)\Delta}) = 0 + \sqrt{h\mathbf{b}'_t\Sigma\mathbf{b}_t}\Phi^{-1}(\alpha) = \sqrt{h}\text{VaR}_\alpha(L_{t+1}^\Delta)$ .
- $\text{ES}_\alpha(L_{t+h}^{(h)\Delta}) = 0 + \sqrt{h\mathbf{b}'_t\Sigma\mathbf{b}_t}\frac{\varphi(\Phi^{-1}(\alpha))}{1-\alpha} = \sqrt{h}\text{ES}_\alpha(L_{t+1}^\Delta)$ .
- Note the many underlying assumptions:
  - ▶  $\mathbf{X}_{t+i}$  independent;
  - ▶  $\mathbf{X}_{t+i}$  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  $\text{VaR}_\alpha^t$  denote the  $\alpha$ -quantile of the conditional loss distribution  $F_{L_{t+1}|\mathcal{F}_t}$  and consider the event indicator variable  $I_{t+1} = I_{\{L_{t+1} > \text{VaR}_\alpha^t\}}$ .
- The event  $\{L_{t+1} > \text{VaR}_\alpha^t\}$  is a VaR *violation* or *exception*.
- Assuming a continuous loss distribution, we have, by definition of the quantile,

$$E(I_{t+1} \mid \mathcal{F}_t) = P(L_{t+1} > \text{VaR}_\alpha^t \mid \mathcal{F}_t) = 1 - \alpha, \quad (68)$$

- $I_{t+1}$  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  $1 \leq T_1 < \dots < T_M \leq m$  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

$$P(S_j = k) = \alpha^{k-1}(1 - \alpha), \quad k \in \mathbb{N}.$$

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

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

- The sequence  $(\hat{I}_t)_{1 \leq t \leq m}$  **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-Qplot.

## 9.3.2 Violation-based tests of expected shortfall

- Let  $ES_{\alpha}^t$  denote the one-period expected shortfall and  $\widehat{ES}_{\alpha}^t$  its estimate.
- Assume  $(L_t)$  follows a model of the form  $L_t = \sigma_t Z_t$ , where  $\sigma_t$  is a function of  $\mathcal{F}_{t-1}$  and the  $(Z_t)$  are  $\text{SWN}(0, 1)$  innovations.
- Then we can define a process  $(K_t)$  by

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

and note that it is a **zero-mean iid sequence**.

- This suggests we form **violation residuals** of the form

$$\widehat{K}_{t+1} = \frac{(L_{t+1} - \widehat{ES}_{\alpha}^t)}{\widehat{ES}_{\alpha}^t} \widehat{I}_{t+1}. \quad (69)$$

- 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 elicibility 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

$$T(F_L) = \arg \min_{y \in \mathbb{R}} \int_{\mathbb{R}} S(y, l) dF_L(l) = \arg \min_{y \in \mathbb{R}} E(S(y, L)) \quad (70)$$

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 (70) implies that

$$\begin{aligned} \left. \frac{d}{dy} E(S(y, L)) \right|_{y=T(F_L)} &= \int_{\mathbb{R}} \frac{d}{dy} S(y, l) dF_L(l) \Big|_{y=T(F_L)} \\ &= E(h(T(F_L), L)) = 0 \end{aligned}$$

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

$$S_{\alpha}^q(y, l) = |1_{\{l \leq y\}} - \alpha| |l - y| \quad (71)$$

is strictly consistent for  $T$ .

- The  $\alpha$ -expectile of  $L$  is defined to be the risk measure that minimizes  $E(S_\alpha^e(y, L))$  where the scoring function is

$$S_\alpha^e(y, l) = |1_{\{l \leq y\}} - \alpha|(l - y)^2. \quad (72)$$

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  $\alpha$ -expectile risk measure for  $\alpha \geq 0.5$ ; see also Weber (2006). [Expected shortfall is not elicitable](#).
- $\text{VaR}_\alpha^t$  minimizes

$$E\left(S_\alpha^q(\text{VaR}_\alpha^t, L_{t+1}) \mid \mathcal{F}_t\right)$$

for the scoring function in (71). We refer to  $S_\alpha^q(\text{VaR}_\alpha^t, L_{t+1})$  as a (theoretical) [VaR score](#).

- Assume  $\text{VaR}_\alpha^t$  is replaced by an estimate at each time point and consider the VaR scores  $\{S_\alpha^q(\widehat{\text{VaR}}_\alpha^t, L_{t+1}) : t = 1, \dots, m\}$
- These can be used to address questions of [relative model performance](#).

- The statistic

$$Q_0 = \frac{1}{m} \sum_{t=1}^m S_{\alpha}^q(\widehat{\text{VaR}}_{\alpha}^t, L_{t+1})$$

can be used as a measure of relative model performance.

- If two models A and B deliver VaR estimates  $\{\widehat{\text{VaR}}_{\alpha}^{tA}, t = 1, \dots, m\}$  and  $\{\widehat{\text{VaR}}_{\alpha}^{tB}, t = 1, \dots, m\}$  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.

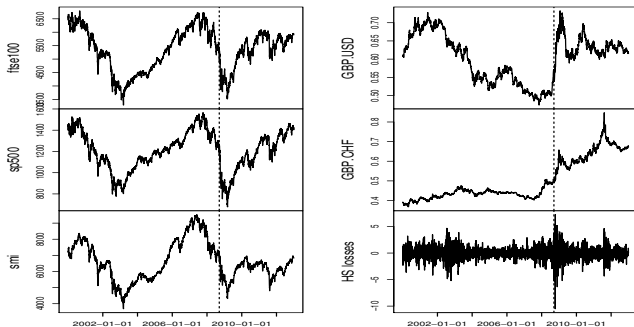
### 9.3.4 Empirical comparison of methods using backtesting concepts

- 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:

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

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

- $x_1$ ,  $x_2$  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 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 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

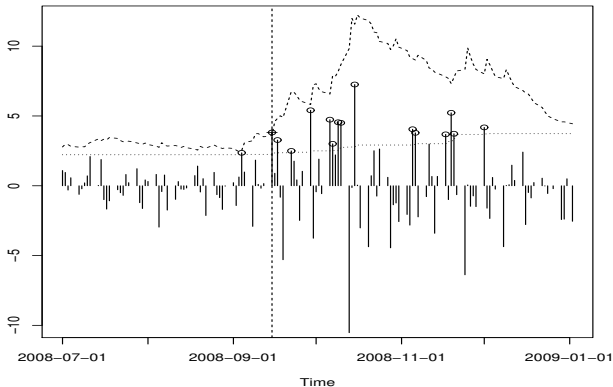
Expected no. of violations	13	13	13	13	13	13	13	13	103
VC	8	16	17	19	13	15	14	14	116
HS	0	6	28	49	19	6	10	1	119
HS-GARCH	9	13	22	22	13	14	9	15	117
HS-GARCH- $t$	9	14	23	22	14	15	10	15	122
HS-MGARCH	5	14	21	19	12	9	11	12	103

### 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

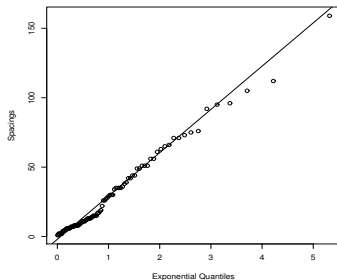


- The HS method **does not react to changing volatility:**



- 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.

- QQ plot of spacings between exceptions:



	Violation residual test			
	95% ES	( <i>n</i> )	99% ES	( <i>n</i> )
VC	0.00	116	0.05	43
HS	0.02	119	0.25	36
HS-GARCH	0.00	117	0.05	43
HS-GARCH- <i>t</i>	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  $\hat{U}_{t+1} = \hat{F}_{L_{t+1}|\mathcal{F}_t}(L_{t+1})$  on day  $t + 1$ .
- Suppose we estimate the predictive distribution on days  $t = 0, \dots, n - 1$  and form backtesting data  $\hat{U}_1, \dots, \hat{U}_n$ ; 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.