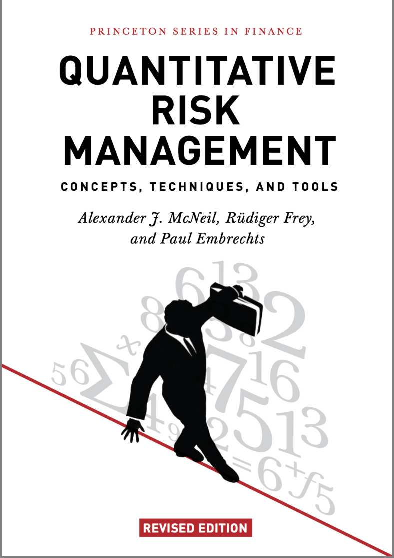
Backtesting of Financial Risk Models

First Banco de Santander Financial Engineering School

Alexander J. McNeil, University of York

2nd March 2021

QRM Textbook: McNeil et al. (2015)



Overview

1. Fundamentals of Financial Risk Modelling
   - Modelling Value and Value Change
   - Risk Measurement
   - Regulatory Background
   - Some Concepts in Time Series Analysis
2. Fundamentals of Backtesting
   - Introduction
   - Backtesting VaR and ES
   - Elicitability and Backtesting
3. New Backtesting Methods Using PIT Values
   - Tests Based on PIT Values
   - Spectral backtests
   - Conditional spectral tests

Overview

1 Fundamentals of Financial Risk Modelling

2 Fundamentals of Backtesting

3 New Backtesting Methods Using PIT Values

Overview

1. Fundamentals of Financial Risk Modelling
   - Modelling Value and Value Change
   - Risk Measurement
   - Regulatory Background
   - Some Concepts in Time Series Analysis
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values

Balance Sheet of a Bank

Bank XYZ (31st December 2012)			
Assets		Liabilities	
Cash (and central bank balance)	£10M	Customer deposits	£80M
Securities	£50M	Bonds issued	
- bonds		- senior bond issues	£25M
- stocks		- subordinated bond issues	£15M
- derivatives		Short-term borrowing	£30M
Loans and mortgages	£100M	Reserves (for losses on loans)	£20M
- corporates			
- retail and smaller clients		Debt (sum of above)	£170M
- government			
Other assets	£20M		
- property			
- investments in companies		Equity	£30M
Short-term lending	£20M		
Total	£200M	Total	£200M

The Trading Book

- Contains assets that are available to trade.
- Can be contrasted with the more traditional banking book which contains loans and other assets that are typically held to maturity and not traded.
- Trading book assets are supposed to be easy to trade, highly liquid and straightforward to value (mark-to-market) at any point in time.
- Examples: fixed income instruments; standardized derivatives.
- The trading book is often identified with market risk whereas the banking book is largely affected by credit risk.
- The Basel rules allow banks to use internal models to measure market risks in the trading book.
- The trading book was abused in the financial crisis of 2007–2009. Many securitized credit instruments (e.g. CDO tranches) were held in the trading book where they were subject to lower capital requirements.

Modelling Value and Value Change

- The risk factors at time t are denoted by the vector $\boldsymbol{Z}_t = (Z_{t,1}, \ldots, Z_{td,})$. These include, for example, equity prices, exchange rates, interest rates for different maturities and volatility parameters.
- The value of the trading book is given by formula of form

$$V_t = f(t, \boldsymbol{Z}_t) \tag{1}$$

  where f is the portfolio mapping which is assumed to be known.
- The risk factors $\boldsymbol{Z}_t$ are observable at time t and hence V_t is known at t.
- Assuming positions are held over the period $[t, t+1]$, the trading book loss is described by the random variable

$$\begin{aligned} L_{t+1} = -(V_{t+1} - V_t) &= -\left(f(t+1, \boldsymbol{Z}_{t+1}) - f(t, \boldsymbol{Z}_t)\right) \\ &= -\left(f(t+1, \boldsymbol{Z}_t + \boldsymbol{X}_{t+1}) - f(t, \boldsymbol{Z}_t)\right) \\ &= l_{[t]}(\boldsymbol{X}_{t+1}) \end{aligned}$$

  where $\boldsymbol{X}_{t+1} = \boldsymbol{Z}_{t+1} - \boldsymbol{Z}_t$ are the risk-factor changes and $l_{[t]}$ is a function we refer to as the loss operator.

The Essence of the Problem

- To estimate the conditional loss distribution

$$F_{t+1}(x) = P\left(l_{[t]}(\boldsymbol{X}_{t+1}) \leq x \mid \mathcal{F}_t\right)$$

  where $\mathcal{F}_t$ denotes the available information at time t.
- To estimate risk measures like VaR and expected shortfall (ES) that describe the tail of F_{t+1}.
- Note that some methods used in practice apply an unconditional approach, assuming stationarity of past risk-factor changes $(\boldsymbol{X}_s)_{s\leq t}$ and estimating the distribution of $l_{[t]}(\boldsymbol{X})$ under stationarity assumption.

Challenges include:

- The volatility and heavy tails of typical risk-factor changes.
- Dimensionality of problem. The number of risk factors d may be very large, although dimension reduction strategies (factor models) are used.
- The loss operator is typically highly non-linear.

Overview

1. Fundamentals of Financial Risk Modelling
   - Modelling Value and Value Change
   - Risk Measurement
   - Regulatory Background
   - Some Concepts in Time Series Analysis
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values

VaR and Expected Shortfall

Let L be a random variable with df F and let $0 < \alpha < 1$.

- Value at Risk is defined to be

$$\mathrm{VaR}_\alpha(L) = q_\alpha(F) = F^{\leftarrow}(\alpha)\,, \tag{2}$$

  where we use the notation $q_\alpha(F)$ for a quantile of the distribution of L and $F^{\leftarrow}$ for the (generalized) inverse of F.

- Provided $E(|L|) < \infty$, expected shortfall is defined to be

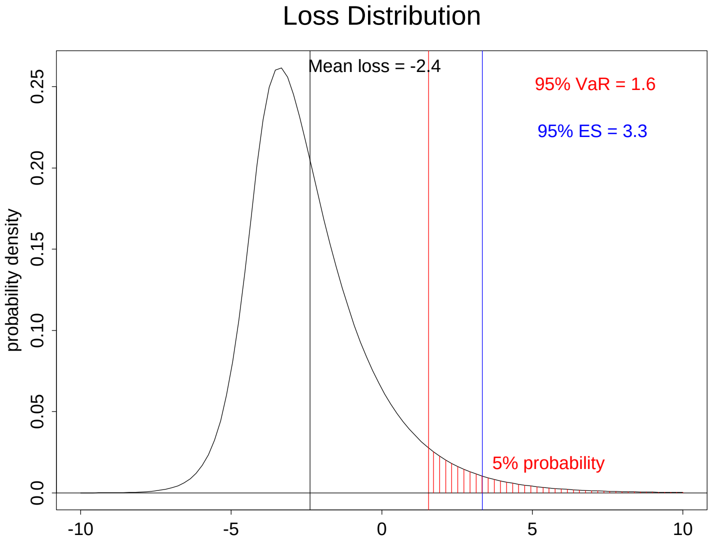
$$\mathrm{ES}_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F)du. \tag{3}$$

  If F is a continuous df then

$$\mathrm{ES}_\alpha(L) = E(L \mid L \geq \mathrm{VaR}_\alpha(L))\,.$$

These are the two most widely applied risk measures in practice.

Losses and Profits



Overview

1. Fundamentals of Financial Risk Modelling
   - Modelling Value and Value Change
   - Risk Measurement
   - Regulatory Background
   - Some Concepts in Time Series Analysis
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values

Fundamental Review of the Trading Book (FRTB)

- As a result of the FRTB, a new standard for capital for market risk in the trading book (Basel Committee on Banking Supervision, 2016) has been designed and adopted in Basel III.
- Banks still have a choice between a standardized approach and an internal-models approach, both of which have been revised.
- There is a more rigorous internal model approval process.
- From FRTB:
  - *The bank must conduct regular backtesting and P&L attribution programmes.*
  - *Backtesting requirements are based on comparing each desk's 1-day static value-at-risk measure (calibrated to the most recent 12 months' data) at both the 97.5th and 99th percentile.*
  - *If any given desk experiences either more than 12 exceptions at the 99th or 30 at the 97.5th percentile, all of its positions must be capitalised using the standardised approach.*

Ingredients in the Capital Calculation

$$\mathrm{ES}_{h_1}(P,j) = \boxed{h_1\text{-day 97.5\%-ES w.r.t. risk factors with liquidity horizon} \geq h_j}$$

$$\underbrace{\mathrm{ES}_{R,S}}_{\text{reduced risk-factor set; stressed calibration}} = \underbrace{\sqrt{\sum_{j=1}^{5}\left(\sqrt{\frac{h_j - h_{j-1}}{h_1}}\,\mathrm{ES}_{h_1}(P,j)\right)^2}}_{(h_1,h_2,h_3,h_4,h_5)=(10,20,40,60,120),\ h_0=0}$$

$$\mathrm{IMCC}(C) = \mathrm{ES}_{R,S} \times \underbrace{\frac{\mathrm{ES}_{F,C}}{\mathrm{ES}_{R,C}}}_{\frac{\text{full risk-factors; current calibration}}{\text{reduced risk-factors; current calibration}}} \qquad \underbrace{\mathrm{IMCC}(C_i) = \mathrm{ES}_{R,S,i} \times \frac{\mathrm{ES}_{F,C,i}}{\mathrm{ES}_{R,C,i}}}_{\text{standalone calc. for risk factor class } i \text{ (IR,FX,EQ,etc.)}}$$

$$\underbrace{\mathrm{IMCC}}_{\text{calculated daily}} = \rho \cdot \underbrace{\mathrm{IMCC}(C)}_{\text{diversified}} + (1-\rho)\underbrace{\sum_i \mathrm{IMCC}(C_i)}_{\text{undiversified}}, \qquad \rho = 0.5$$

$$C_A = \max\{\mathrm{IMCC}_{t-1} + \underbrace{\mathrm{SES}_{t-1}}_{\text{non-modellable risk factors}},\ \underbrace{m_c}_{\text{multiplier}} \cdot \underbrace{\overline{\mathrm{IMCC}} + \overline{\mathrm{SES}}}_{\text{running averages}}\}$$

$$\mathrm{ACC} = \underbrace{C_A}_{\text{approved desks}} + \underbrace{\mathrm{DRC}}_{\text{default risk charge}} + \underbrace{C_U}_{\text{unapproved desks}}$$

The Multiplier

From FRTB:

- *The multiplication factor m_c will be 1.5. Banks must add to this a plus directly related to the ex-post performance of the model.*
- *The plus will range from 0 to 0.5 based on the outcome of the backtesting of the bank's daily VaR at the 99th percentile based on the current observations of the full set of risk factors.*

The traffic light system:

$$
\begin{aligned}
N_{\text{exceptions}} \leq 4 \quad &\Longrightarrow \quad m_c = 1.5 \\
N_{\text{exceptions}} = 5, 6, 7, 8, 9 \quad &\Longrightarrow \quad m_c = 1.70, 1.76, 1.83, 1.88, 1.92 \\
N_{\text{exceptions}} \geq 10 \quad &\Longrightarrow \quad m_c = 2, \text{ regulatory intervention}
\end{aligned}
$$

Additional Model Validation Standards

Additional tests are required which may include, for instance:

- *Testing carried out for longer periods than required for the regular backtesting programme (eg three years); or*
- *Testing carried out using the entire forecasting distribution using the p-value of the desk's profit or loss on each day. For example the bank could be required to use in validation and make available to the supervisor the following information for each desk for each business day over the previous three years, with no more than a 60-day lag:*
  - (i) *Two daily VaR's for the desk calibrated to a one-tail 99.0 and 97.5 percent confidence level, and a daily ES calibrated to 97.5;*
  - (ii) *The daily P&L for the desk; and*
  - (iii) *The p-value for the P&L on each day for each desk (that is, the probability of observing a profit that is less than, or a loss that is greater than the amount reported according to the model used to calculate ES).*
- *Testing of portfolios must be done at both the trading desk and bank-wide level.*

Overview

1. Fundamentals of Financial Risk Modelling
   - Modelling Value and Value Change
   - Risk Measurement
   - Regulatory Background
   - Some Concepts in Time Series Analysis
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values

Stylized facts of Financial Time Series

- Consider discrete observations of a financial risk factor Z_t made at daily, weekly, or perhaps even monthly intervals. We are interested in risk-factor changes: $X_t = Z_t - Z_{t-1}$.
- Z_t might be a log asset price ($Z_t = \ln S_t$) in which case

$$X_t = (\ln S_t - \ln S_{t-1}) \approx (S_{t-1} - S_t) / S_{t-1}$$

  is often called a (log) return.
- A realistic model should reflect stylized facts of risk-factor change series:
  - Risk-factor changes not iid but correlation low
  - Absolute changes highly correlated
  - Volatility appears to change randomly with time
  - Risk-factor changes are often leptokurtic or heavy–tailed
  - Extremes appear in clusters
- GARCH-type models can be useful for modelling this behaviour.

ARCH and GARCH: Models for Conditional Variance

- Let $(Z_t)_{t\in\mathbb{Z}}$ follow a strict white noise process with mean zero and variance one. This just means iid.
- ARCH and GARCH processes $(X_t)_{t\in\mathbb{Z}}$ take the general form

$$X_t = \sigma_t Z_t, \quad t \in \mathbb{Z}, \tag{4}$$

  where σ_t, the volatility, is a function of the history up to time $t-1$ represented as usual by $\mathcal{F}_{t-1}$.
- The conditional variance is $\mathrm{var}(X_t \mid \mathcal{F}_{t-1}) = \sigma_t^2$.
- Thus volatility is the conditional standard deviation of the process.
- Depending on time horizon, a conditional mean term μ_t depending on $\mathcal{F}_{t-1}$ may be introduced so that $X_t = \mu_t + \sigma_t Z_t$.

ARCH and GARCH Processes

Definition (ARCH(p) process)

(X_t) follows an ARCH(p) process if, for all t,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2, \quad \text{with } \alpha_j > 0.$$

Intuition: volatility influenced by large observations in recent past.

Definition (GARCH(p,q))

(X_t) follows a GARCH(p,q) process (generalised ARCH) if, for all t,

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{k=1}^{q} \beta_k \sigma_{t-k}^2, \quad \text{with } \alpha_j, \beta_k > 0. \tag{5}$$

Intuition: more persistence is built into the volatility.

Overview

1. Fundamentals of Financial Risk Modelling
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values

Overview

1 Fundamentals of Financial Risk Modelling

2 Fundamentals of Backtesting
- Introduction
- Backtesting VaR and ES
- Elicitability and Backtesting

3 New Backtesting Methods Using PIT Values

Essence of Backtesting

- Backtesting is the practice of evaluating risk measurement procedures by comparing out-of-sample estimates of risk measures with actual realized losses and gains.
- Backtesting allows us to evaluate the question of whether a given estimation procedure produces credible risk measure estimates.
- Suppose that a model has been used to estimate risk measures for the distribution of losses in the next period.
- At the end of the next period we have the opportunity to compare the risk measure estimates with the actual realized loss.
- When this procedure is repeated over many time periods we can monitor the performance of methods and compare their relative performance.

Set-Up for Backtesting Discussion

- Recall that, for $t \in \mathbb{N}$, the trading book loss over $[t-1, t]$ is the negative P&L given by an expression of form

$$L_t = l_{[t-1]}(\boldsymbol{X}_t)$$

  where $l_{[t-1]}$ is a $\mathcal{F}_{t-1}$-measurable function and $\boldsymbol{X}_t$ is a random vector of changes in fundamental risk factors over $[t-1, t]$.
- The true (conditional) loss distribution is

$$F_t(x) = P\left(L_t \leq x \mid \mathcal{F}_{t-1}\right). \tag{6}$$

- We denote the α-VaR and α-ES of this distribution by $\mathrm{VaR}_{\alpha,t}$ and $\mathrm{ES}_{\alpha,t}$.
- For $t \in \mathbb{N}$ the risk modelling group forms an estimate $\widehat{F}_t$ of F_t based on information up to time $t-1$.
- Models may be parametric or non-parametric (e.g. HS).
- The estimate $\widehat{F}_t$ is used to compute estimates $\widehat{\mathrm{VaR}}_{\alpha,t}$ and $\widehat{\mathrm{ES}}_{\alpha,t}$.

Overview

1 Fundamentals of Financial Risk Modelling

2 Fundamentals of Backtesting
- Introduction
- Backtesting VaR and ES
- Elicitability and Backtesting

3 New Backtesting Methods Using PIT Values

Backtesting VaR

- We refer to the event $\{L_t > \mathrm{VaR}_{\alpha,t}\}$ as a VaR violation or exception.
- Denote the indicator of this event by $I_t = I_{\{L_t > \mathrm{VaR}_{\alpha,t}\}}$.
- Assuming a continuous loss distribution, we have, by definition of the quantile, that

$$E(I_t \mid \mathcal{F}_{t-1}) = P(L_t > \mathrm{VaR}_{\alpha,t} \mid \mathcal{F}_{t-1}) = 1 - \alpha\,, \tag{7}$$

which may also be written as

$$E\left(h_\alpha(\mathrm{VaR}_{\alpha,t}, L_t) \mid \mathcal{F}_{t-1}\right) = 0$$

where h_α is a so-called identification function given by

$$h_\alpha(q, l) = I_{\{l > q\}} - (1 - \alpha). \tag{8}$$

- It may be shown that (7) holds if and only if the sequence of VaR violation indicators (I_t) forms a Bernoulli trials process, i.e. a process of iid Bernoulli random variables with event probability $1 - \alpha$.

Properties of a Bernoulli Trials Process

1. If we sum the violation indicators over a number of different times, we obtain binomially distributed random variables. For example,

$$M = \sum_{t=1}^{m} I_t \sim B(m, 1-\alpha).$$

2. Suppose that the violations occur at the random times $1 \leq T_1 < \cdots < T_M \leq m$. If we set $T_0 = 0$, then the spacings $S_j = T_j - T_{j-1}$ will be independent geometrically distributed random variables 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 can be tested in empirical data.

Note that, for small event probability $1-\alpha$, the Bernoulli Trials Process may be well approximated by a Poisson process and the geometric distribution may be approximated by an exponential distribution.

Testing for a Bernoulli Trials Process

- Define the empirical indicator variables by $\widehat{I}_t = I_{\{L_t > \widehat{\mathrm{VaR}}_{\alpha,t}\}}$.
- If $P(L_t > \widehat{\mathrm{VaR}}_{\alpha,t} \mid \mathcal{F}_{t-1}) = 1 - \alpha$ at each time t, then the sequence of indicator variables $(\widehat{I}_t)_{1 \leq t \leq m}$ should behave like a realization from a Bernoulli trials process with event probability $(1 - \alpha)$.
- There are a number of classical tests for a binomial distribution including a likelihood ratio test (LRT), score test and Wald test.
- The Basel internal model approval rules and traffic lights are essentially based on the binomial LRT, although binomial score test is better sized.
- Exponential spacings can be tested numerically or with a Q–Qplot.
- A number of tests have been proposed to test for binomial behaviour and independence, including the well-known test of Christoffersen (1998). These are often called tests of conditional coverage and distinguished from tests of unconditional coverage.

Backtesting Expected Shortfall

- Consider the identification function

$$h(q, e, l) = \left(\frac{l - e}{e}\right) I_{\{l>q\}}$$

  and note that $E(h(\mathrm{VaR}_{\alpha,t}, \mathrm{ES}_{\alpha,t}, L_t) \mid \mathcal{F}_{t-1}) = 0$.
- We can define a conditional mean-zero process (K_t) by $K_t = h(\mathrm{VaR}_{\alpha,t}, \mathrm{ES}_{\alpha,t}, L_t)$ for all t.
- Under the stronger assumption that (L_t) follows a model of the form $L_t = \sigma_t Z_t$, where σ_t is $\mathcal{F}_{t-1}$-measurable and the (Z_t) are $\mathrm{SWN}(0, 1)$ innovations, (K_t) is iid mean-zero process.
- This suggests we form violation residuals of the form $\widehat{K}_t = h(\widehat{\mathrm{VaR}}_{\alpha,t}, \widehat{\mathrm{ES}}_{\alpha,t}, L_t)$.
- We can test for mean-zero behaviour using a bootstrap test (or t-test) on the non-zero violation residuals (McNeil and Frey, 2000).

Overview

1 Fundamentals of Financial Risk Modelling

2 Fundamentals of Backtesting
- Introduction
- Backtesting VaR and ES
- Elicitability and Backtesting

3 New Backtesting Methods Using PIT Values

Elicitability and Scoring Functions

- The elicitability concept has been introduced into the backtesting literature by Gneiting (2011); see also important papers by Bellini and Bignozzi (2015) and Ziegel (2016).
- 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.
- Law-invariant risk measures ϱ can be considered as functionals T of the distribution of the loss F.
- Suppose that for some class of loss distribution functions $\mathcal{X}$ a real-valued statistical functional T satisfies

$$T(F) = \arg\min_{y\in\mathbb{R}} \int_{\mathbb{R}} S(y, l)\mathrm{d}F(l) = \arg\min_{y\in\mathbb{R}} E(S(y, L)) \tag{9}$$

for a scoring function S and any loss distribution $F \in \mathcal{X}$.

Elicitability and Scoring Functions II

- Suppose moreover that $T(F)$ is the unique minimizing value in (9).
- The scoring function S is said to be strictly consistent for T.
- The functional $T(F)$ (or corresponding risk measure) is said to be elicitable.
- Note that (9) implies that

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

  where h is the derivative of the scoring function.
- Thus elicitability theory also indicates how we may derive so-called identification functions for hypothesis tests involving $T(F)$.

Elicitability: Examples

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

$$S_\alpha^q(y, l) = |1_{\{l \le y\}} - \alpha||l - y| \tag{10}$$

  is strictly consistent for T.

- If we take the negative of the derivative of this function with respect to y we get the identification function $h_\alpha(q, l)$ in (8).

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

$$S_\alpha^e(y, l) = |1_{\{l \le y\}} - \alpha|(l - y)^2. \tag{11}$$

  This risk measure is elicitable by definition.

- Bellini and Bignozzi (2015) and Ziegel (2016) 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.

Comparative Backtesting

- We define so-called VaR scores $\{S^q_\alpha(\widehat{\mathrm{VaR}}_{\alpha,t}, L_t) : t = 1,\ldots,m\}$ where S^q_α is the scoring function in (10).
- The statistic
$$Q_0 = \frac{1}{m}\sum_{t=1}^{m} S^q_\alpha(\widehat{\mathrm{VaR}}_{\alpha,t}, L_t)$$
can be used as a measure of relative model performance.
- If two models A and B deliver VaR estimates $\{\widehat{\mathrm{VaR}}^{(A)}_{\alpha,t}, t = 1,\ldots,m\}$ and $\{\widehat{\mathrm{VaR}}^{(B)}_{\alpha,t}, t = 1,\ldots,m\}$ 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.

Overview

1. Fundamentals of Financial Risk Modelling
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values

Overview

1 Fundamentals of Financial Risk Modelling

2 Fundamentals of Backtesting

3 New Backtesting Methods Using PIT Values
- Tests Based on PIT Values
- Spectral backtests
- Conditional spectral tests

Probability integral transform

- Increasingly, regulators observe more than just VaR exceedances.
- Consider the PIT process given by $P_t = \widehat{F}_t(L_t)$.
- Reported PIT values contain information about VaR exceedances at every level α.

$$P_t \geq \alpha \iff L_t \geq \widehat{\mathrm{VaR}}_{\alpha,t}$$

- The ideal forecaster. If the $(\widehat{F}_t)$ coincide with the true (F_t), then the process (P_t) is iid $U[0,1]$ (Rosenblatt, 1952).
- In the US, banks on the Internal Models Approach for the trading book have been required to report PIT values to regulators since 2013.
- What is the best way to exploit this additional information?

Simulated example of a backtest dataset

Days	VaR	Loss	Exceed?	PIT
1	2.492	0.278	0	0.602
2	2.968	0.716	0	0.713
3	3.336	-0.759	0	0.298
4	3.018	-0.451	0	0.364
5	2.654	2.955	1	0.995
6	3.335	-1.697	0	0.118
7	3.137	0.184	0	0.554
8	2.641	1.091	0	0.832

Priorities for model performance

- Diebold et al. (1998) develop forecast density tests based on testing PIT values for iid $U[0, 1]$.
- We will explain the testing framework for PIT values developed in Gordy and McNeil (2020).
- In a risk-management context, some quantiles of the forecast distribution are more important than others.
- Accuracy in “good tail” of high profits (low P_t) is generally much less important than accuracy in the “bad tail” of large losses (high P_t).
- We study a class of backtests for forecast distributions in which the test statistic weights exceedances by a function of the probability level α.
- The kernel function makes explicit the priorities for model performance.

Overview

1. Fundamentals of Financial Risk Modelling
2. Fundamentals of Backtesting
3. New Backtesting Methods Using PIT Values
   - Tests Based on PIT Values
   - Spectral backtests
   - Conditional spectral tests

Spectral transformations

- Tests are based on transformations of the indicator function for PIT exceedances and are termed “spectral” in the integral transform sense.
- The transformations take the form

$$W_t = \int_0^1 I_{\{P_t \geq u\}} \mathrm{d}\nu(u) = \nu([0, P_t])$$

  where ν is a Lebesgue-Stieltjes measure on $[0, 1]$.
- W_t increases in P_t and can be thought of as a weighted PIT.
- ν is chosen to apply weight to different levels in the unit interval, typically in the region of the VaR level $\alpha = 0.99$.
- We refer to ν as the kernel measure for the transform.
- The support of the measure describes subsets of $[0, 1]$ that are weighted.

Spectral backtests

- Univariate spectral backtests are backtests based on $W_1, \ldots, W_n$.
- Multivariate tests based on $\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n$ where $\boldsymbol{W}_t = (W_{t,1}, \ldots, W_{t,J})'$ and $W_{t,j} = \nu_j([0, P_t])$ for distinct measures $\nu_1, \ldots, \nu_J$.
- Null hypothesis. Let F_W^0 denote df of $\boldsymbol{W}_t$ when P_t is uniform.

$$H_0 : \quad \boldsymbol{W}_t \sim F_W^0 \text{ and } \boldsymbol{W}_t \perp\!\!\!\perp \mathcal{F}_{t-1}, \ \forall t. \tag{12}$$

- Within the class of spectral backtests, we have tests of unconditional and conditional coverage.
- Unconditional coverage: test for correct distribution F_W^0;
- Conditional coverage: correct distribution and independence from $\mathcal{F}_{t-1}$.

Z-tests

- Univariate Z-tests are based on the asymptotic normality under H_0 of $\overline{W}_n = n^{-1} \sum_{t=1}^{n} W_t$.
- Solve for $\mu_W = \mathbb{E}(W_t)$ and $\sigma_W^2 = \mathrm{var}(W_t)$ in the null model F_W^0.
- Then it trivially follows from CLT that, under H_0,

$$Z_n = \frac{\sqrt{n}(\overline{W}_n - \mu_W)}{\sigma_W} \xrightarrow[n\to\infty]{d} N(0,1).$$

- Multivariate Z-tests are based on

$$T_n = n \left(\overline{\boldsymbol{W}}_n - \boldsymbol{\mu}_W\right)' \Sigma_W^{-1} \left(\overline{\boldsymbol{W}}_n - \boldsymbol{\mu}_W\right) \xrightarrow[n\to\infty]{d} \chi_J^2.$$

Discrete kernels

- Dirac case
  - A Dirac kernal $\nu = \delta_\alpha$ yields $W_t = I_{\{P_t \geq \alpha\}}$, the α-VaR exceedance indicator.
  - The (W_t) are iid Bernoulli$(1-\alpha)$ under H_0.
  - The Z-test is the binomial score test promoted in Kratz et al. (2018).
- Univariate
  - A general discrete kernel $\nu = \sum_{i=1}^{m} k_i \delta_{\alpha_i}$ yields $W_t = \sum_{i=1}^{m} k_i I_{\{P_t \geq \alpha_i\}}$.
  - W_t satisfies

$$P(W_t = q_i) = \alpha_{i+1} - \alpha_j = i, \quad i = 0, \ldots, m \tag{13}$$

    where $q_i = \sum_{j=1}^{i} k_j,\ q_0 = 0,\ \alpha_0 = 1$ and $\alpha_{m+1} = 1$.
  - The Z-test is a new test which allows user to vary the weights k_i.
- Multivariate
  - A set of m distinct Dirac kernels $\nu_1 = \delta_{\alpha_1}, \ldots, \nu_m = \delta_{\alpha_m}$ yields multivariate tests based on $\boldsymbol{W}_t = (I_{\{P_t \geq \alpha_1\}}, \ldots, I_{\{P_t \geq \alpha_m\}})'$.
  - The Z-test is identical to Pearson's celebrated chi-squared test which has been proposed by Campbell (2007) for backtesting.

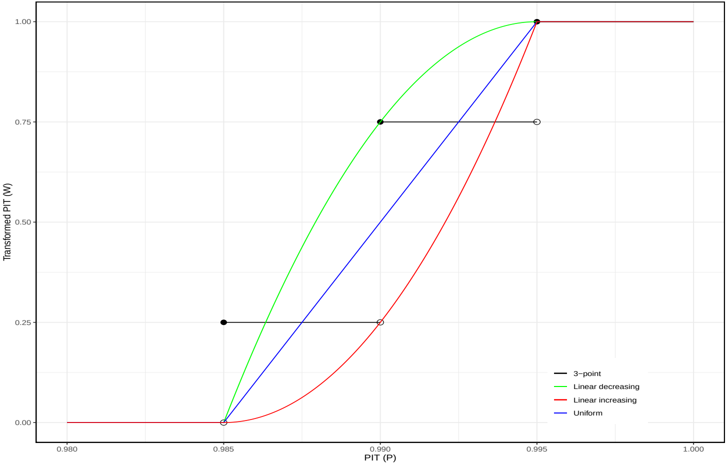
Continuous kernels

- Univariate
  - A continuous kernel measure has density $\mathrm{d}\nu(u) = g(u)\mathrm{d}u$ for some non-negative function g on $[0, 1]$.
  - We study measures with support given by a window $[\alpha_1, \alpha_2] \subset [0, 1]$.
  - Univariate Z-tests are considered by Costanzino and Curran (2015) and Du and Escanciano (2017) who use uniform kernel $g(u) = I_{\{\alpha_1 \leq u \leq \alpha_2\}}$.
- Bivariate
  - Bispectral Z-tests using two measures with common support can be effective.
  - While univariate tests can detect systematic under or overestimation of quantiles in kernel window, bivariate tests can detect more complex misspecifications (crossing).
- Mixed kernels
  - We also consider one test (probitnormal score test) with a kernel that corresponds to a mixed measure (discrete and continuous parts).

Kernel list for examples

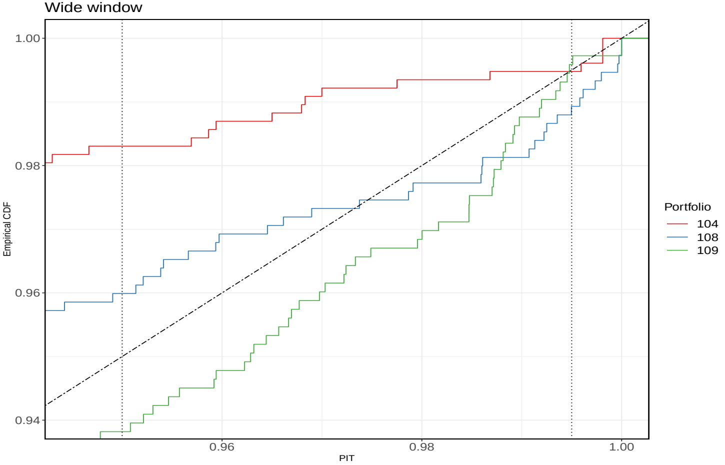
BIN	Binomial score test (Dirac case)
Z3p	Univariate discrete with weight on 3 points
Pearson3	Multivariate discrete on 3 points (Pearson chi-squared test)
ZU	Univariate continuous test with uniform kernel
ZLp	Univariate continuous test with linear increasing kernel
ZLL	Bivariate continuous test with two linear kernels.
PNS	Bivariate mixed kernel test.

Illustration of selected kernels



In this case the kernel window is $[\alpha_1, \alpha_2] = [0.985, 0.995]$.

What the tests detect



Empirical distribution of PITs should be close to diagonal. Tests pick up deviations within the kernel window - under/overestimation, crossing.

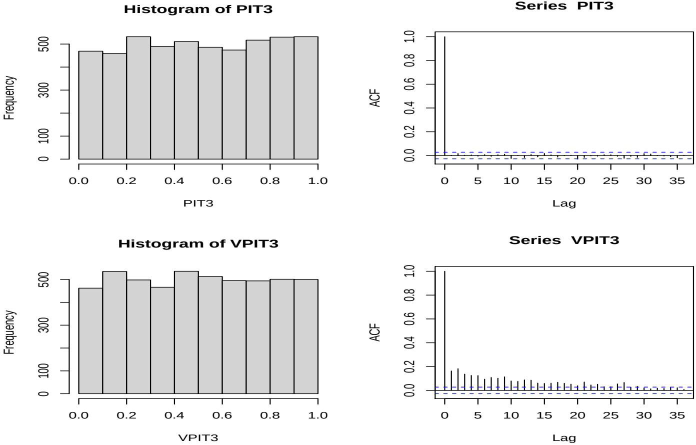
Overview

1 Fundamentals of Financial Risk Modelling

2 Fundamentals of Backtesting

3 New Backtesting Methods Using PIT Values
- Tests Based on PIT Values
- Spectral backtests
- Conditional spectral tests

Unmodelled volatility in PIT values



Histograms and ACF plots of PIT-values (P_t) and transformed PIT-values ($|2P_t - 1|$). Volatility of returns has not been captured.

Testing the martingale difference property

- Let $\tilde{W}_t = W_t - \mu_W$ for all t. Under H_0 the martingale difference (MD) property hold: $\mathbb{E}(\tilde{W}_t \mid \mathcal{F}_{t-1}) = 0$.
- For any $\mathcal{F}_{t-1}$-measurable $\boldsymbol{h}_{t-1}$ vector this implies $\mathbb{E}(\boldsymbol{h}_{t-1}\tilde{W}_t) = \mathbf{0}$.
- We consider $\boldsymbol{h}_{t-1} = (1, h(P_{t-1}), \ldots, h(P_{t-k}))'$ for some choice of h.
- Let $\boldsymbol{Y}_t = \boldsymbol{h}_{t-1}\tilde{W}_t$ for $t = k+1, \ldots, n$. Let $\overline{\boldsymbol{Y}} = (n-k)^{-1}\sum_{t=k+1}^{n} \boldsymbol{Y}_t$ and let $\hat{\Sigma}_Y$ denote a consistent estimator of $\Sigma_Y := \operatorname{cov}(\boldsymbol{Y}_t)$.
- Giacomini and White (2006) show that under very weak assumptions, for large enough n and fixed k,

$$(n-k)\, \overline{\boldsymbol{Y}}'\, \hat{\Sigma}_Y^{-1}\, \overline{\boldsymbol{Y}} \sim \chi^2_{k+1}.$$

Conditional spectral tests

- This framework generalizes the dynamic quantile test of Engle and Manganelli (2004) which corresponds to $\nu = \delta_\alpha$ and $h(p) = I\{p \geq \alpha\}$.
- Case $k = 0$ is ordinary spectral Z-test.
- Case $k = 1$ is an analog of Markov chain LR-test of Christoffersen (1998).
- Martingale-difference extensions of multispectral tests are also available.
- We choose $h(p) = |2p - 1|$ to target unmodelled stochastic volatility.
- Results show that unconditional spectral tests are not sensitive to violations of the MD hypothesis caused by serial dependence in the PITs, provided their distribution is close to uniform.
- The conditional spectral tests can pick up the serial dependence.

Concluding Remarks

- Under FRTB model validation and backtesting requirements have become more stringent and now extend to desk level.
- Backtesting exceptions may lead to a higher multiplier applied to a firm's capital requirement and may also lead to withdrawal of internal model status for particular desks or the whole trading book.
- To comply with the new requirements, and anticipate developments, banks can implement tests based on realized *p*-values that go beyond simple binomial exception tests.
- Realized *p*-values give an overarching framework for discussing many backtesting approaches that have been proposed.
- Multinomial exception tests at several levels are an easy extension to binomial tests that are much more powerful, particularly at exposing systematic underestimation of risk.
- Tests that can identify serial dependence in PIT values, caused by failure to adequately model the dynamics of trading book risk, can also be constructed.

For Further Reading

Basel Committee on Banking Supervision (2016). Minimum capital requirements for market risk. Publication No. 352, Bank of International Settlements.

Bellini, F. and Bignozzi, V. (2015). On elicitable risk measures. *Quantitative Finance*, 15(5):725–733.

Campbell, S. (2007). A review of backtesting and backtesting procedures. *Journal of Risk*, 9(2):1–18.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4).

Costanzino, N. and Curran, M. (2015). Backtesting general spectral risk measures with application to expected shortfall. *The Journal of Risk Model Validation*, 9(1):21–31.

Diebold, F., Gunther, T., and Tay, A. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883.

For Further Reading (cont.)

Du, Z. and Escanciano, J. (2017). Backtesting expected shortfall: accounting for tail risk. *Management Science*, 63(4):940–958.

Engle, R. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Giacomini, R. and White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74:1545–1578.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gordy, M. and McNeil, A. (2020). Spectral backtests of forecast distributions with applications to risk management. *Journal of Banking & Finance*, 116.

Kratz, M., Lok, Y., and McNeil, A. (2018). Multinomial VaR backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking & Finance*, 88:393–407.

For Further Reading (cont.)

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.

McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton, 2nd edition.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, 23:470–472.

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

Ziegel, J. F. (2016). Coherence and elicitability. *Mathematical Finance*, 26(4):901–918.