Lecture notes on risk management, public policy, and the financial system

Assessing Value-at-Risk

Allan M. Malz

Columbia University
Backtesting of VaR

Critiques of VaR
Backtesting of VaR

Overview
Unconditional coverage test procedure
Examples of backtesting
Limitations of the unconditional coverage test

Critiques of VaR
Challenges in validating VaR

- How do we measure “poor performance” of VaR? → **model risk**
- VaR **backtesting**: type of **model validation**
- VaR not a point forecast, but statement about distribution of future outcomes
- VaR **exceedance, exception** or **excession**: event the portfolio loss exceeds the VaR
  - Loss over the VaR horizon is compared with VaR computed just prior
  - E.g. for daily VaR, compare VaR reported at close of trading with loss over subsequent trading day
- For single position, exceedance can be defined in terms of return: for each of $T$ observations,
  - Parametric: compare realized return with estimated volatility
  - Historical simulation: compare realized log or arithmetic return with quantile of historical sample
- Practical problem: portfolio is likely to be changing over time
  - Backtest comparison assume static portfolio
Assessing Value-at-Risk

Backtesting of VaR

Overview

Testable dimensions of VaR

**Unconditional coverage:** is proportion of exceedances in entire sample consistent with VaR confidence level?

**Independence:** frequency and timing of exceedances, e.g. absence of clustering

**Magnitude** of exceedances: somewhat larger or much larger than VaR?
Brief review of statistical hypothesis testing

- Formulate **statistical hypothesis** testable with available data
  - Framed as a **null hypothesis** $H_0$ about a distributional characteristic of the data
  - $H_0$ expressed through a **test statistic**, so **falsifiable** based on data
  - $H_0$ guides choice of test statistic; data determines its value
  - So **falsifiable** based on data
- $H_0$ guides choice of test statistic; data determines its value
  - So **falsifiable** based on data
Errors in statistical hypothesis testing

**Type I:** reject $\mathcal{H}_0$ even though $\mathcal{H}_0$ true

- Often referred to as “false positive”
  - Since rejection often taken as *confirmation* of a theory
  - When framed as “treatment has effect” or “factor has influence”
- **Significance level** of test: a prespecified, chosen probability of Type I error, e.g. 0.01
  - *p-value*: probability, if $\mathcal{H}_0$ true, of having a test statistic at least as unfavorable to $\mathcal{H}_0$ as that actually obtained

**Type II:** fail to reject $\mathcal{H}_0$ even though $\mathcal{H}_0$ false

- “False negative”
- **Power** of a test: probability of Type II error
Sample space of a statistical test

- **Sample space**: all the possible configurations of the data
- Identify in the sample space for a given significance level:
  - **Critical** or **rejection region** within which $H_0$ rejected
  - **Acceptance** or **non-rejection region** within which $H_0$ *not* rejected
    - is complement in sample space of critical region
- Sample $\in$ critical region leads to test statistic with $p$-value $< \text{significance level}$
Statistical framework for unconditional coverage test

- VaR associated with a confidence level $\alpha$
- VaR model accurate $\Rightarrow$ exceedances occur $\approx$ every $(1 - \alpha)^{-1}$ periods
  - For example, with daily VaR at 95 percent, expect $\approx 1$ per month
  - $\Rightarrow$ Null hypothesis $H_0$: exceedance frequency or fraction of exceedances $= 1 - \alpha$
- Backtest is a sequence of comparisons of current VaR estimate with P&L realized at the VaR forecast horizon
- Under $H_0$, comparisons are Bernoulli trials/random variables:
  
  $\begin{cases} 
  1 - \alpha \\
  \alpha 
  \end{cases}$

  result is $\begin{cases} 
  1 \ (\text{VaR exceedance}) \\
  0 \ (\text{VaR not exceeded}) 
  \end{cases}$

- And independently and identically distributed (i.i.d.)
- In reality, clustered exceedances are routine
- $H_0$ doesn't state returns are lognormal, just that VaR procedure accurate for confidence level $\alpha$
Test statistic of unconditional coverage test

- **Likelihood function** of $T$ i.i.d. observations of VaR forecast and subsequent realized loss:
  \[
  L(\alpha; x) = (1 - \alpha)^x \alpha^{T-x}
  \]

- $x$ is the number of exceedances out of $T$
- $L(\alpha)$: probability of $x$ in-sample exceedances if exceedance probability $1 - \alpha$

- **Maximum likelihood estimator** of $\alpha$ is $1 - \frac{x}{T}$
  - Likelihood function then takes on value
  \[
  L \left( \frac{x}{T}; x \right) = \left( \frac{x}{T} \right)^x \left( 1 - \frac{x}{T} \right)^{T-x}
  \]

- The test statistic is the **log likelihood ratio**
  \[
  2 \left\{ \ln \left[ L \left( \frac{x}{T}; x \right) \right] - \ln \left[ L(\alpha; x) \right] \right\} \\
  = 2 \left\{ \ln \left[ \left( \frac{x}{T} \right)^x \left( 1 - \frac{x}{T} \right)^{T-x} \right] - \ln \left[ (1 - \alpha)^x \alpha^{T-x} \right] \right\}
  \]
Distribution of unconditional coverage test statistic

- Test statistic measures distance between data and model prediction
  - Log of ratio of what we observe to what $H_0$ leads us to expect
- Follows a $\chi^2$ distribution (for large enough $T$) if $H_0$ is true
  - With one degree of freedom (df), for the one parameter $\alpha$
  - $\chi^2$ test a standard approach to assessing goodness of fit of a distributional hypothesis
  - In this case, exceedances i.i.d. Bernoulli trials with parameter $\alpha$
- $p$-value: probability, if $H_0$ true, of a test statistic greater than or equal to that actually obtained in the sample
  - I.e. 1 minus cumulative probability of a $\chi^2[1]$ variate with a value equal to the test statistic
- Independence requirement $\rightarrow$ non-overlapping observations if risk horizon $> \text{observation frequency}$
\( \chi^2[1] \) distribution

Cumulative distribution function of a \( \chi^2 \) variate with one degree of freedom.

<table>
<thead>
<tr>
<th>Significance level</th>
<th>0.95</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical value</td>
<td>3.8415</td>
<td>6.6349</td>
</tr>
</tbody>
</table>
Critical value and acceptance range

- Reject $H_0$ only if test statistic $> \text{critical value}$
  - Critical value is a quantile of $\chi^2[1]$, the $\chi^2$ distribution with 1 df
  - Quantile is chosen to correspond to significance level of backtest

- $\rightarrow$ **Acceptance range**: range of number of exceedances s.t. test statistic $< \text{critical value}$
  - If number of exceedances falls outside acceptance range, reject null hypothesis
  - Too many or too few exceedances $\rightarrow$ high value of test statistic
  - But caveat: $\chi^2$ nonetheless a one-tailed test

- **Example**: 1 year (252 daily observations), VaR confidence level 0.99

<table>
<thead>
<tr>
<th>No. of exceedances</th>
<th>0</th>
<th>3</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test statistic</td>
<td>5.0654</td>
<td>0.0870</td>
<td>12.8331</td>
</tr>
<tr>
<td>$\chi^2$ cumulative probability</td>
<td>0.9756</td>
<td>0.2320</td>
<td>0.9997</td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.0244</td>
<td>0.7680</td>
<td>0.0003</td>
</tr>
</tbody>
</table>

- Zero exceedances results in rejection of $H_0$ at a significance level of 0.95, but not 0.99
Significance and confidence levels in the test

- Confidence level of backtest is distinct from confidence level of VaR
  - *Confidence* level of VaR enters into test statistic (together with number of observations, number of exceedances)
  - *Significance* level of backtest determines $\chi^2$ quantile to compare (together with number of degrees of freedom)
- Acceptance range depends on significance level of backtest
  - Acceptance range is wider at a higher significance level
  - Greater departure from expected exceedance count required to reject null that VaR accurate
  - Any realization outside acceptance range has $p$-value below significance level of backtest
Test statistic and acceptance range

Points represent values for 1 year of daily VaR estimates; $T = 252$ and $\alpha = 0.99$ of test statistic $2 \left\{ \ln \left[ (\frac{x}{T})^x \left(1 - \frac{x}{T}\right)^{T-x} \right] - \ln \left[ (1 - \alpha)^x \alpha^{T-x} \right] \right\}$ for integer values of exceedances $x$ from 0 to 7. The acceptance range at a 95 percent confidence level is $x \in [1, 6]$. 

Quantile function of the $\chi^2[1]$ distribution

Test statistic

Cumulative probability

$\chi^2[1]$ value

Critical value: 0.95 quantile of $\chi^2[1]$ distribution

Acceptance range

No. of outliers

Test statistic

$T = 252$ and $\alpha = 0.99$ of test statistic $2 \left\{ \ln \left[ (\frac{x}{T})^x \left(1 - \frac{x}{T}\right)^{T-x} \right] - \ln \left[ (1 - \alpha)^x \alpha^{T-x} \right] \right\}$ for integer values of exceedances $x$ from 0 to 7. The acceptance range at a 95 percent confidence level is $x \in [1, 6]$. 

Quantile function of the $\chi^2[1]$ distribution

Test statistic

Cumulative probability

$\chi^2[1]$ value

Critical value: 0.95 quantile of $\chi^2[1]$ distribution

Acceptance range

No. of outliers
Setting up the examples

- Unconditional coverage test of daily VaR at 99 percent confidence level
  - Using 5 years of data 30Sep2014 to 30Sep2019
  - Use parametric VaR with EWMA volatility estimate
- Assume constant position size each day, backtest in return terms
- Backtest two single-position portfolios:
  - Long position in S&P 500
  - Short position in AUD against USD
    - AUD-USD exchange rate expressed as USD price of A$1
    - Short loss if exchange rate rises
S&P 500 and AUD-USD returns and excessions

Points denote daily returns, solid plot the 98 percent confidence level, expressed as a return and measured using a EWMA volatility estimate with a decay factor of $\lambda = 0.94$. Orange x’s denote excessions of the VaR. Left: long position in the S&P 500 index. Right: short position in AUD against USD.
Results for the examples

- Reject $\mathcal{H}_0$ for long position in S&P 500 at 0.95 and 0.99 significance levels
- Reject $\mathcal{H}_0$ for short position in AUD-USD at neither 0.95 nor 0.99 significance levels

<table>
<thead>
<tr>
<th></th>
<th>Long S&amp;P 500</th>
<th>Short AUD-USD</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. obs.</td>
<td>1258</td>
<td>1304</td>
</tr>
<tr>
<td>acceptance range (0.99 significance level)</td>
<td>7–20</td>
<td>7–20</td>
</tr>
<tr>
<td>no. exceptions</td>
<td>28</td>
<td>17</td>
</tr>
<tr>
<td>% exceptions</td>
<td>2.23</td>
<td>1.30</td>
</tr>
<tr>
<td>value of test statistic</td>
<td>14.157</td>
<td>1.109</td>
</tr>
</tbody>
</table>
Limitations of the unconditional coverage test

- Weak test: hard to reject $H_0$ unless number of observations $T$ very large
- Disregards size of exceedances (→expected shortfall)
- Disregards clustering of exceedances (→alternative tests, return models)
Backtesting of VaR

Critiques of VaR

Overview
Variability of VaR estimates
The coherence critique of VaR
Limitations of VaR

- **Accuracy:**
  - Inadequate treatment of frequency and size of tail risk ⇒ generally poor performance during crises
  - But even when no recent financial crisis, low power, i.e., hard to reject null
- **VaR doesn’t tell risk manager how large loss might be if VaR exceeded**
  - In VaR limit system, may incentivize traders to take more risk
    - Trades may increase return, as well as probability of tail losses much larger than VaR, while increasing VaR much less
  - Can be addressed through use of (*→*) **expected shortfall**
- **Even if the distribution model were right:** nonlinear risks, options
- **The devil in the details:** subtle and not-so-subtle differences in how VaR is computed ⇒ large differences in results
- **VaR is not coherent** because it is not subadditive: a portfolio may have a VaR larger than the sum of the individual positions’ VaR
- **Procyclicality:** widespread use of similar VaR models in setting trading limits can amplify price fluctuations
Getting whatever answer you want from VaR

- Compute 10-day (2-week) VaR four different ways
  1. Parametric: assume log returns normally distributed
     1.a Using 10-day volatility, computed via exponentially weighted moving average (EWMA) using non-overlapping observations
     1.b Using 1-day volatility times $\sqrt{10}$
  2. Historical simulation using non-overlapping observations
     2.a Using 2 years of data
     2.b Using 5 years of data
- Express results as a return (easy to turn into a dollar amount)
- Results: large differences among approaches

<table>
<thead>
<tr>
<th>Technique</th>
<th>12Mar2003</th>
<th>26Nov2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parametric: 10-day volatility</td>
<td>9.90</td>
<td>14.43</td>
</tr>
<tr>
<td>Parametric: 1-day volatility $\times \sqrt{10}$</td>
<td>9.03</td>
<td>28.75</td>
</tr>
<tr>
<td>Historical simulation: 2 years of data</td>
<td>8.15</td>
<td>24.60</td>
</tr>
<tr>
<td>Historical simulation: 5 years of data</td>
<td>9.66</td>
<td>20.15</td>
</tr>
</tbody>
</table>
Backtesting the four models

Backtesting VaR, 99 percent confidence level. With $T = 513$ and $\alpha = 0.99$, the acceptance range is $[2, 10]$. Points denote returns, blue plot the VaR, expressed as a return, red $\times$'s denote excursions.
Variability and model risk

- **Model risk**: Risk of losses due to errors in models and how applied
- Choice of VaR model can lead to over- or underestimate of risk *ex post*
- →Subject to manipulation
  - Choice of computational technique, historical lookback period
  - Distributional hypothesis, pricing models in simulations
  - Choice of risk factors, e.g. mapping resi subprime to AAA corporate
  - Mapping position and hedge to same risk factor: voil‘a, no basis risk
Coherence of risk measures

- **Coherence** is a set of standards for risk measures to ensure they do not lead to perverse or counterintuitive rankings of strategies.
- Defined mathematically, but implement these intuitions:
  - **Monotonicity:** if one portfolio’s return is always greater than that of another, its measured risk must be smaller.
  - **Homogeneity of degree one:** doubling every position in a portfolio should exactly double its measured risk.
  - **Subadditivity:** the risk of a portfolio should be no greater than the sum of the risks of its constituents.
  - **Translation invariance:** adding a riskless asset to a portfolio should reduce its measured risk by that same amount.
- VaR doesn’t satisfy the subadditivity condition.
Examples of failure of subadditivity of VaR

- Examples are easy to generate: require
  - Positions susceptible to large loss, but with low probability, i.e. below $1 - \alpha$, with $\alpha$ the VaR confidence level
  - Each position has zero or negative VaR
  - Positions are independent, or have low correlation, or low probability of joint event of loss
  - Loss probabilities and correlations are such that probability of loss on at least one position exceeds $\alpha$

- **Examples** of positive-VaR portfolios at the 99 percent confidence level consisting of zero- or negative-VaR positions
  - Market-risk VaR: two option positions, short a far out-of-the-money (OTM) call and OTM put, each with probability of exercise just less than 1 percent
  - Credit-risk VaR: two loans, each with a default probability just less than 1 percent and low default correlation