The best and the worst of VaR in a Basel III context

Jean-Paul Laurent, Univ. Paris 1 Panthéon – Sorbonne, PRISM & Labex Refi
Hassan Omidi Firouzi, Royal Bank of Canada & Labex Refi
Séminaire Compta Contrôle Finance Sorbonne
7 April 2016, updated 5 September 2016

Key messages for regulation

- Hidden impacts of risk modelling choices on financial stability and pro-cyclicality under Basel III FRTB
  - Even when considering simple exposures (S&P500)
  - And complexity (optional products, correlations) left aside
- Backtesting / Quantitative Impact Studies poorly discriminates among models under calm periods
  - Danielsson (2002)
- Questionable benchmarking on hypothetical portfolios
  - Highly unstable ranking of risk models
- Promote smart supervision, model risk validation and enhanced disclosure on risk methodologies

Messages for market risk managers

- Favour Volatility Weighted Historical Simulation (VWHS) over Historical Simulation (HS) for VaR and Expected Shortfall computations?
  - Standard backtesting procedures are of little help
- Historical Simulation works poorly in stressed periods
  - Hidden procyclicality: patterns of VaR exceptions under stress and fall-back to costly Standard Approach
- BUT large estimation errors when computing the decay factor in VWHS
  - Challenge the .94 golden risk number?
  - Consider smaller values of decay factor(s)?

The best and worse out of VaR in a Basel III context: outlook

- Market risks: regulatory outlook
- The rise of historical simulation
- Backtesting and VaR exceptions
- Pointwise volatility estimation: The conundrum
- Assessment of risk models under Basel III
  - Limited usefulness of econometric techniques
  - Hypothetical Portfolio Exercises useless?
  - Lower decay factors to mitigate disruptions in the computation of Risk Weighted Assets?
Market risks: regulatory outlook

- Market risks are not the main driver of banks’ risks
  - But are prominent for large dealer banks

- Computing market RWA (Risk Weighed Assets)
  - Basel amendment for market risks (1996)
  - Fixing Basel II after 2008 turmoil
    - Stressed VaR based on year 2008
    - Credit risk: IRC, CRM, VaR on CVA, ...
  - Minimum capital requirements for market risk (2016)
    - Implementation scheduled in 2019
    - Laurent (2016) for an overview of ongoing issues

- Basel III: Internal Models Approach (IMA) still applicable
- 97.5% Stressed Expected Shortfall (ES)
  - Liquidity horizons: 10 days or more
    - No scaling from 1D to 10D (Danielsson & Zigrand (2006))
- Backtesting based on 97.5% and 99% 1 day VaR
  - Not directly on ES as in Du & Escanciano (2016)
  - Number of VaR exceptions over past year
  - At trading desk level: Danciulescu (2010), Wied et al. (2015)
    - VaR exception if « loss » greater than VaR
  - BCBS QIS also requests reporting of 1D 97.5% ES + p – values

The rise of Historical Simulation (HS)

1% HS VaR (based on 250 rolling days) and S&P500 returns over past 10 years. Nominal = 1
The rise of historical simulation

- **Backtesting**: compare 1 day VaR with both hypothetical and actual daily Profit and Loss (P&L)
  - **Hypothetical P&L**
    - Banks holdings frozen over risk horizon
    - « Uncontaminated P&L »: not accounting for banks’ fees (Frésard et al. (2011)).
    - Computed according to all risk factors and pricing tools being used by Front Office (FO)
    - Full revaluation is implicit when computing hypothetical P&L

- **Use of risk-theoretical P&L to compute VaR**
  - Changes in P&L according to bank’s internal risk model (which includes risk representation and pricing tools)
    - Use of modellable risk factors within risk systems (FRTB/Basel 3) or risks in VaR when applicable
    - Subset of risk factors used in Front Office systems.
    - Delta/gamma approximations, PV grids or full revaluation might be used in repricing books
  - Rank daily P&L over past 250 trading days (1Y)
    - In between 2nd and 3rd worst loss provides 99% VaR

- **Huge literature to compare approaches to VaR/ES**
  - Historical, FHS, VWHS, EWMA, Parametric (multivariate Gaussian), GARCH family, EVT, CAViaR, ...
  - Focus on backtesting performance
    - Lack of implementation details, choice of backtest portfolios, historical periods make comparisons difficult
  - Dealing with operational issues is also of importance
    - Large dimensionality: several thousands of risk factors,
    - Costly to price optional products,
    - Data requirements.

From Perignon & Smith (2010) based on 2005 data

Mehta et al. (2012)
The rise of historical simulation

- Volatility Weighted Historical Simulation (VWHS)
  - Hull & White (1998), Barone-Adesi et al. (1999), not to be confused with Boudoukh et al. (1998)
  - Volatility not constant over VaR estimation period
  - Rescale returns by ratio of current volatility to past volatility
    - $\sigma_t$ volatility at time $t$, $r_{t-h}$ return at $t-h$
    - Rescaled past returns $\frac{\sigma_{t-h}}{\sigma_{t-h}} \times r_{t-h}$
    - VWHS: empirical quantile of rescaled returns

The rise of historical simulation

- (Location) scale models: $r_t = \sigma_t \times \varepsilon_t$
  - GARCH: $\varepsilon_t$ has a given stationary distribution
    - Such as $t$ ($\nu$): parametric approach to $\varepsilon_t$
  - VaR: $q_\alpha (r_t) = \sigma_t \times q_\alpha (\varepsilon_t)$
    - EVT could be used to assess $q_\alpha (\varepsilon_t)$, McNeil & Frey (2000), Diebold et al. (2000), Jalal & Rockinger (2008)
  - VWHS: same approach to VaR
    - BUT $q_\alpha (\varepsilon_t)$ empirical quantile of standardised returns $r_t / \sigma_t$
    - Above decomposition shows two sources of model risk: volatility estimation $\sigma_t$, tails of standardized returns $\varepsilon_t$

The rise of historical simulation

- Issues with previous approaches
  - Standardised returns $\varepsilon_t = r_t / \sigma_t$ not directly observed
    - Since $\varepsilon_t$ depends on volatility estimates $\sigma_t$
    - Use of Diebold & Mariano (2002) to compare predictive accuracy questionable.
  - Large uncertainty when deriving $\sigma_t$?
    - See page 29 when using EWMA
  - Issues with GARCH(1,1) modelling: Pritsker (2006)
    - Misspecification of $\varepsilon_t$ distribution?
    - Tail dynamics only driven by volatility $\sigma_t$

(Var1%/VaR2.5%)/ (Φ⁻¹(99%)/Φ⁻¹ (97.5%))
EWMA volatility estimates, decay factor = .8

Descriptive statistics of standardised returns $\varepsilon_{t+1}$

For Gaussian $\varepsilon_t$ and well-specified decay factor, ratio should be equal to one
Ratio higher than 1 means fat tails
(Var1%/VaR2.5%) / (Φ⁻¹(99%)/Φ⁻¹(97.5%))
EWMA volatility estimates, decay factor = .8

$\varepsilon_t = \frac{\tau_t}{\sigma_t}$ shows some left tail dynamics.

Descriptive statistics of standardised returns $\varepsilon_t$

Daily 97.5% ES (black) vs 99% VaR (red), $\lambda = .97$

Expected Shortfall computations:

$ES_\alpha(r_t) = \sigma_t \times ES_\alpha(\varepsilon_t)$

Over past 10 years, patterns are similar, but ES is less stable than VaR due to outliers

Daily Expected Shortfall of Standardised returns

Ratio of 97.5% ES to 99% VaR ($\lambda = .94$)

$ES_{97.5}(\varepsilon_t)$ is unstable over past 10 years
Median (3.1), 1st decile (2.5), 9th decile (4.1) with peaks up to 10

Daily ES instability confirmed by considering ratio of ES to VaR
Backtesting and VaR exceptions

- Basel III regulatory reporting
  - 10 days Expected Shortfall (capital requirement)
  - Computed over different subsets of risk factors (partial ES), scaled-up to various time horizons
  - Computed over stressed period, averaged and submitted to multiplier (in between 1.5 and 2)
- 1 day 99% and 97.5% VaR (backtesting)
  - $q_{99}(\epsilon_t) = \sigma_t \times q_{99}(\epsilon_t)$
  - $q_{97.5}(\epsilon_t) = \sigma_t \times q_{97.5}(\epsilon_t)$

Volatility Weighted Historical Simulation outperforms Historical Simulation

- Number of VaR exceptions over past 10 years (S&P 500)

<table>
<thead>
<tr>
<th></th>
<th>1% VaR</th>
<th>2.5% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical Simulation</td>
<td>40</td>
<td>89</td>
</tr>
<tr>
<td>Volatility Weighted Historical Simulation (RiskMetrics)</td>
<td>26</td>
<td>68</td>
</tr>
<tr>
<td>Expected</td>
<td>25</td>
<td>63</td>
</tr>
</tbody>
</table>

Backtesting and VaR exceptions

- VaR exception: whenever loss exceeds VaR
  - For 250 trading days and 1% VaR, average number of VaR exceptions = 2.5
  - For well-specified VaR model, number of VaR exceptions follows a Binomial distribution
    - So-called « unconditional coverage ratios » or traffic light approach (Kupiec, 1995, Basel III, 2016)
  - Regulatory thresholds at bank’s level: green zone, up to 4 exceptions, yellow zone, in between 5 and 9 exceptions, red zone, 10 or above
  - At desk level: 12 exceptions at 1%, 30 at 2.5%

Volatility estimation: the conundrum

- EWMA (Exponentially Weighted Moving Average)
  - $\sigma_t^2 = \lambda \times \sigma_{t-1}^2 + (1 - \lambda) \times r_t^2$
  - $\lambda$: decay factor, 1 $-$ $\lambda$ speed at which new returns are taken into account for pointwise volatility estimation
    - RiskMetrics (1996), $\lambda = 0.94$ « Golden number »
    - Single parameter model
  - EWMA is a special case of GARCH(1,1)
    - With no mean reversion of volatility.
    - $\sigma_t^2$ is not floored and become quite close to zero in calm periods (Murphy et al. (2014))
Volatility estimation: the conundrum

Pattern of estimated volatility: EWMA with decay factor = .94

Numerous techniques to estimate decay factor $\lambda$
- RiskMetrics (1996): minimizing the average squared error on variance estimation
  \[
  \hat{\lambda} = \arg\min_{\lambda \in (0, 1)} \frac{1}{T} \sum_{t=1}^{T} [\sigma_t^2(\lambda) - r_t^2]^2
  \]

Other approaches:
- Guermat & Harris (2002) to cope with non-Gaussian returns
- Minimization of check-loss function: González-Rivera et al. (2007)

For S&P500, Estimates of decay factor are highly unstable and could range from 0.8 to 0.98 wild around the 0.94 RiskMetrics « golden number »
- Note that $\lambda = 1$ corresponds to plain HS

Building volatility filters is even more intricate when considering different risk factors (Dave & Stahl (1998))
Volatility estimation: the conundrum

Ratios of daily volatility estimates over past 10Y with decay factor 0.94 and 0.8 are highly volatile.

Assessment of VaR (risk) models

VaR1%/VaR1% for decay factors .8 and .94 respectively: shaky volatility estimates leads to large VaR estimation uncertainty and huge time instability.

Assessment of risk models

- Number of VaR Exceptions over past 10 years (S&P 500)

<table>
<thead>
<tr>
<th></th>
<th>1% VaR</th>
<th>2.5% VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>VWHS ( \lambda = 0.8 )</td>
<td>28</td>
<td>68</td>
</tr>
<tr>
<td>VWHS ( \lambda = 0.94 ) (RiskMetrics)</td>
<td>26</td>
<td>68</td>
</tr>
<tr>
<td>Expected</td>
<td>25</td>
<td>63</td>
</tr>
</tbody>
</table>

Note: Almost same results for tests based on number of VaR exceptions (unconditional coverage)

Assessment of risk models

- Smaller decay factors imply **prompter VaR increases** when volatility rises and slightly better behaviour during stressed periods.

<table>
<thead>
<tr>
<th>( \lambda )</th>
<th>Number of Exceptions for 99% VaR over period January 2008 – January 2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda = 0.8 )</td>
<td>5</td>
</tr>
<tr>
<td>( \lambda = 0.94 )</td>
<td>8</td>
</tr>
<tr>
<td>( \lambda = 0.97 )</td>
<td>11</td>
</tr>
</tbody>
</table>

Note: Stressed period based on high levels of VaR and of VIX.

Similar results in Boucher et al. (2014), where plain HS \( (\lambda = 1) \) provides poor results under stress. See also O'Brien & Szerszen (2014).
Assessment of risk models

- PIT (Probability Integral Transform) adequacy tests
  - Crnkovic and Drachman (1995), Diebold et al. (1997), Berkowitz (2001)
  - Regulators: Fed, ongoing BCBS QIS
    - Check whether the loss distribution (instead of a single quantile) is well predicted.
    - If $F_t$ is the well-specified (predicted) conditional loss distribution, $F_t(\tau_{t+1}) \sim U[0,1]$
    - $F_t(\tau_{t+1})$: p-values

Good news: risk models are not a vacuum!

Bad news: PIT does not discriminate among risk models! (lack of conditionality)

Expected values: 25 exceptions at 1% level, 38 in between 1% and 2.5%; good fit with VWHS

Focusing on tails: VWHS vs plain HS

QQ plot for p-values for VWHS with lambda=.8

QQ plot for p-values for VWHS with lambda=.94

Histogram of p-values for VWHS and $\lambda=.94$
Focusing on tails: VWHS vs plain HS

Histogram of p-values for plain HS, $\lambda=1$

Expected values: 25 exceptions at 1% level, 38 in between 1% and 2.5%: bad fit with HS

Assessment of risk models

- Clustering of VaR exceptions, i.e. several blows in a row might knock-out bank’s capital
- Are VaR exceptions clustered during stressed periods?
  - “We are seeing things that were 25-standard deviation moves, several days in a row”
    - Quoted from David Viniar, Goldman Sachs CFO, August 2007 in the Financial Times
  - Crotty (2009), Danielsson (2008), Dowd (2009), Dowd et al. (2011)
- Tests based on duration between VaR exceptions
  - Christoffersen & Pelletier (2004), Haas (2005), Candelon et al. (2010)

Overshoots for VaR exceptions using VWHS and lambda=.8 at 1% confidence level

Not too much clustering with lower values of decay factor

Assessment of risk models

- Conditional coverage tests
  - $I_t = 1, 0$ depending on occurrence of an exception
  - $E_t[|I_{t+1}|] = 1\%$, 2.5%
    - $E_t$ conditional expectation
  - Conditional probability of VaR exception consistent with confidence level
- Instrumental variables: past VaR exceptions and current + past level of the VIX volatility index
  - Leads to GMM type approach
Assessment of risk models

\[ I_t = \alpha_0 + \sum_{i=1}^{I} \alpha_i I_{t-i} + \sum_{j=0}^{K} \beta_j VIX_{t-j} + u_t \]

- VaR model is well-specified if \( \alpha_0 = 1\% \), 2.5\% and \( \beta_j = 0, \alpha_i = 0, i \geq 1 \)

- We rather follow the logistic regression approach
  - Berkowitz et al. (2008)
- Choosing number of lags \( I, K \) is uneasy
  - Number of lags depend on confidence level
  - And considered portfolio/trading desk
  - Bayesian Information Criteria (BIC), backward model selection, partial autocorrelation function (PACF) are not discriminant

Assessment of risk models

- Preliminary results suggests that \( \lambda \leq 0.9 \)
  - Would reject \( \lambda = 0.94 \) (Riskmetrics standard)

| Parameters (two regressors, \( I_{-3}, I_{-4} \)) | Estimate | Std. Error | \( z \) value | \( Pr(> | z |) \) |
|--------------------------------------------------|----------|------------|---------------|----------------|
| \( \alpha_0 \)                                  | -4.0561  | 0.5043     | -8.043        | 8.77e-16***    |
| \( \alpha_3 \)                                  | 2.4467   | 1.2060     | 2.029         | 0.0425*        |
| \( \alpha_4 \)                                  | 2.4467   | 1.2060     | 2.029         | 0.0425*        |

- But results of statistical tests are difficult to interpret (depend on the chosen lags)
- Rejection for lags (3,4) acceptance for lag 3 only

| Parameters (one regressor, \( I_{-3} \)) | Estimate | Std. Error | \( z \) value | \( Pr(> | z |) \) |
|------------------------------------------|----------|------------|---------------|----------------|
| \( \alpha_0 \)                          | -3.8544  | 0.4519     | -8.529        | < 2e-16***     |
| \( \alpha_3 \)                          | 2.2450   | 1.1850     | 1.894         | 0.0582         |

Assessment of risk models

- Results for S&P500 2.5\% confidence level
  - Red cells are acceptable: no lag for VIX, but lags 2,3,4 or (3,4) for \( I_{t-i} \) could be considered

|               | GMM model | (1|0) | (1|1) | (1|2) | (2|0) | (2|1) | (2|2) |
|---------------|-----------|-----|-----|-----|-----|-----|-----|-----|
| BIC           | 67.18     | 72.25| 69.70| 65.07| 70.21| 67.80|
| GMM model     | (3|0) | (3|1) | (3|2) | (4|0) | (4|1) | (4|2) |
| BIC           | 45.07     | 70.13| 67.71| 65.07| 70.14| 66.55|
| \( \lambda = 0.22 \) | (1|2|1) | (1|2|2) | (2|3|0) | (2|3|1) | (2|3|2) |
| BIC           | 70.33     | 75.44| 73.02| 67.82| 73.08| 70.66|
| GMM model     | (3|4|0) | (3|4|1) | (3|4|2) | (3|4|3) | (4|3|1) | (4|3|2) |
| BIC           | 67.86     | 73.03| 70.42| 69.97| 75.05| 72.63|

Assessment of risk models

- Vast litterature on model risk due to parameter uncertainty, choice of estimation method.
  - Our focus is more narrow: concentrate on a key parameter left in the shadow, i.e. decay factor, and implications for risk management under Basel III
    - Recall that Historical Simulation, EWMA/Riskmetrics and FHS/VWHS are quite different
Tackling RWA (Risk Weighted Assets) variability

- VaR models with strikingly different outputs would not fail backtests
  - Not new! But what to do with this?
- This can feed suspicion on internal models
  - Hidden model complexity, tweaked RWAs?
  - Standardized Basel III risk models
  - Floors based on Hypothetical Portfolios Exercises

Floors based on Hypothetical Portfolio Exercises (HPE)?

- Basel 2013 RCAP (Regulatory Consistency Assessment Programme) BCBS240, BCBS267 & EBA (2013) show large variations across banks regarding VaR outputs for hypothetical portfolios
  - Partly related to discrepancies under various jurisdictions
  - Partly due to modelling choices
    - Length of data sample to estimate VaR, relative weights on dates in filtered historical simulation
    - And as shown in our study HS vs VWHS

Floors based on Hypothetical Portfolio Exercises (HPE)?

- Our controlled experiment shows that ranking of models varies dramatically through time
  - Model A can much more conservative than model B one day, the converse could be observed next day
  - Though in average models A and B provide the same VaRs
- This is problematic regarding the interpretation of HPE and RWA variability
  - Above approach would favour the use of the same possibly misspecified 0.94 golden number...

Tweaking internal models?

- Strategic/opportunistic choice of decay factor?
  - Sticky choice of decay factor: supervisory process
  - Does not change average capital requirements
  - Could change the pattern of VaR dynamics
    - Higher decay factor leads to smoother patterns and ease management (risk limits)
    - Regulatory capital requirements are based on stressed period only and on averages over past 60 days
    - No procyclicality issue with using smaller decay factors
Undue internal model complexity?

- **Haldane and Madouros** (2012), **Dowd** (2016) tackle undue model complexity
- Our approach is simple and widely documented
  - No correlation modelling or pricing models of exotic products is involved
  - No sophisticated econometric methods
  - However, HS can be fine tuned
- Making things simpler (Standard Approaches, output floors based on SA, leverage ratio) might reduce risk sensitivity

Traps in market risk capital requirements

- Avoiding the procyclical trap
  - Using lower values of decay factor for prompter updates in volatility prediction
  - Smaller number of VaR exceptions in volatile periods
  - Resilience of internal models against market tantrum
  - Managing reputation (see above Goldman’s case study)
- Lowering decay factor should not increase capital requirements
  - No bias in average variance estimates
  - ES computed on a stressed period only + averaging

**Traps in market risk capital requirements**

- Procyclical trap when using today’s risk models
  - Ratio of IMA to SA quite large in a number of cases
    - Plain historical simulation or use Riskmetrics decay factor results in large number of VaR exceptions under stress and fallback to SA
    - If a IMA desk is disqualified, huge increase in capital requirements
    - Issue not foreseen: QIS are related to a calm period
    - Use of outfloors based on a percentage of SA would not solve above issue

**Traps in market risk capital requirements**

- Avoiding the FRTB procyclical trap?
  - Banks are currently faced with other top priorities regarding desk eligibility to IMA
    - Data management to reduce NMRF scope
    - PnL attribution tests: reconciliation of risk and front office risk representations and pricing tools, dealing with reserves and fair value adjustments
    - Threshold number of VaR exceptions at desk level is high.
  - BUT large number of desks (100?) and local or global market tantrums might be devastating
    - Forget about unfrequent recalibration of risk models!
Conclusion

- Focus on decay factor impacts for risk measurement in the new Basel III setting
  - Desk-level validation and back-testing
- Beware of plain historical simulation methods and challenge the .94 golden number
  - Further research with internal bank data might prove useful
  - Lower decay factors for dedicated trading desks
- Challenge the outcomes of Hypothetical Portfolio Exercises on RWA variability

References


References

- BCBS, 2013, *Regulatory consistency assessment program (RCAP) – Analysis of risk-weighted assets for market risk*.

References

References


References

References


