Different Methods of Backtesting VaR and ES

Hsiao Yen Lok

Heriot Watt University

May 17, 2015

Overview

1. Value at Risk
   - Interpretation
   - Violation Based Test
   - Independence Based Test
   - Score Based Test
2. Expected Shortfall
   - Interpretation
   - Zero Mean Test
3. Extensions
   - Other Methods for Value at Risk

Value at Risk

Also known as generalised quantile. Given some confidence level $\alpha \in (0,1)$, the VaR_α of some loss distribution F_L, assuming F_L is right continuous, is the generalized inverse $F_L^{\leftarrow}$, given by

$$\text{VaR}_\alpha = F_L^{\leftarrow}(\alpha) = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}.$$

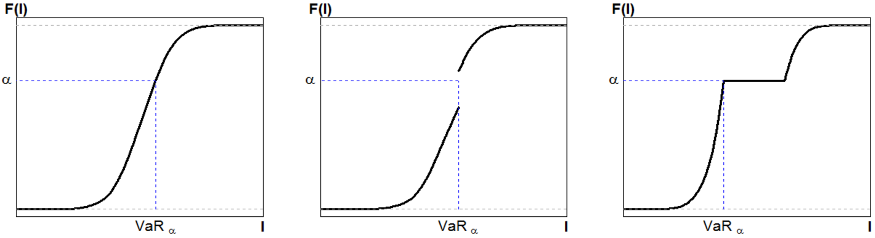

Figure: VaR_α of various types of distribution functions.

Backtesting VaR: Violation Based Test

- The probability that the loss L exceeds VaR_α, $P(L \geq \text{VaR}_\alpha) \leq 1 - \alpha$. We obtain equality in the equation by assuming that the loss distribution F_L is continuous.
- This leads naturally to the Binomial test. Suppose we have a series of observed loses L_t, $t = 1, \ldots, n$, and the corresponding VaR_α^t forecast estimated using our assumed distribution G_L. We will refer to the event $\{L_t > \text{VaR}_\alpha^t\}$ as a violation, with the corresponding indicator function $1_{\{L_t > \text{VaR}_\alpha^t\}}$.
- Each of the $1_{\{L_t > \text{VaR}_\alpha^t\}}$ should behave as Bernoulli distributed r.v.'s with success probability $(1 - \alpha)$, and the sum $S_n = \sum_{t=1}^n 1_{\{L_t > \text{VaR}_\alpha^t\}}$ should be Binomial distributed, with

$$S_n \sim B(n, 1 - \alpha)$$

Backtesting VaR: Violation Based Test

- Kupiec test:
  - Kupiec (1995) have proposed to used the likelihood ratio statistic LR_{uc} to test the violation rate.
  - Under the null hypothesis that the observed violation rate $\frac{S_n}{n}$ is statistically equal to the expected violation rate $p = 1 - \alpha$.
  - Making use of the result $2(l(\hat{\theta}) - l(\theta_0)) \sim \chi^2_1$, we get

$$LR_{uc} = -2 \ln \left(\frac{(1-p)^{n-S_n} p^{S_n}}{\left(1 - \frac{S_n}{n}\right)^{n-S_n} \left(\frac{S_n}{n}\right)^{S_n}} \right) \sim \chi^2_1$$

Backtesting VaR: Independence Based Test

It is well known that financial data exhibit some form of volatility clustering. A good model should be able to capture this characteristic.

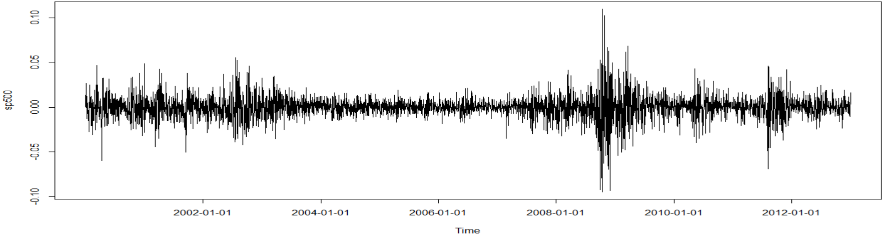

Figure: Log returns of S&P 500 from year 2000 to year 2010.

Backtesting VaR: Independence Based Test

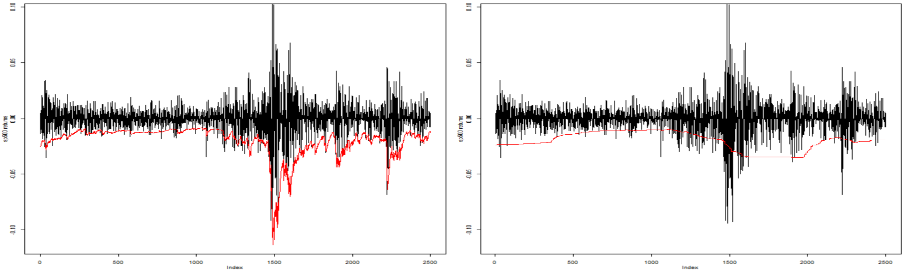

Figure: Comparison of fitted 1-step ahead $VaR_{0.95}$ forecast of 2 models, with a 500 days rolling window size. Left: GARCH-EVT model, Right: HS model.

- First Model: GARCH(1,1) dynamics with Extreme Value Theory (EVT) applied to the residuals.
  Result: 113 violations out of 2500 days (4.5%), p value of 0.263.
- Second Model: Standard rolling historical simulation.
  Result: 132 violations out of 2500 days (5.3%), p value of 0.524.

Backtesting VaR: Independence Based Test

- Christofferssen test:
  - Christofferssen (1998) have proposed to used the likelihood ratio statistic LR_{uc} to test whether the violation indicator $1^t = 1_{\{L_t > \mathrm{VaR}_\alpha^t\}}$.
  - The null hypothesis is that the violation indicator 1^t does not exhibits a first order Markov property, i.e.

$$P(1^t = 0 | 1^{t-1} = 0) = P(1^t = 0 | 1^{t-1} = 1) = 1 - \alpha.$$

  - Making use of the result $2(l(\hat{\theta}) - l(\theta_0)) \sim \chi_1^2$, and defining n_{ij} to be the number of observation with value i followed by j, and $\pi_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$, we obtain the test statistic

$$LR_{ind} = -2 \ln \left(\frac{\alpha^{n_{00}+n_{10}} (1-\alpha)^{n_{01}+n_{11}}}{(1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}}} \right) \sim \chi_1^2$$

Backtesting VaR: Independence Based Test

- Weibull test (Christoffersen and Pelletier 2004):
  - Ideally, the duration between two VaR violation should be i.i.d.
  - We consider the exponential distribution since it is the only memoryless continuous distribution. We want to find a distribution which have exponential distribution as a special case.
  - The Weibull distribution has the density function

$$f_W(x, a, b) = a^b b x^{b-1} e^{-(ax)^b},$$

    where the exponential distribution is the special case when $b = 1$.
  - We fit the Weibull distribution to the duration data, and test the null hypothesis

$$H_0 : b = 1 \quad \text{(duration is exponential distributed.)}$$

Backtesting VaR: Independence Based Test

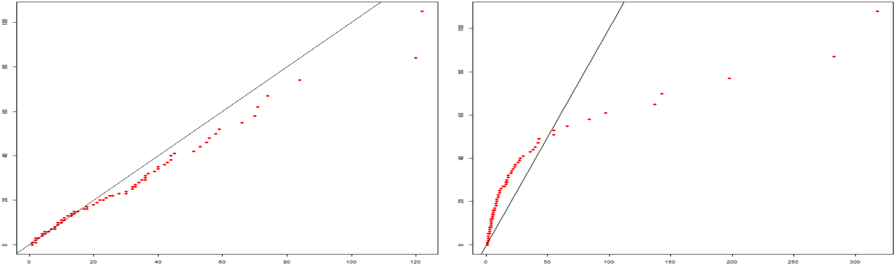

Figure: QQ-Exponential Plot Comparison. Left: GARCH-EVT model, Right: HS model.

- GARCH-EVT p-value: Weibull (0.389), Christofferssen (0.405)
- HS p-value: Weibull (0.000), Christofferssen (0.029)

Backtesting VaR: Elicitability Theory

- VaR is known to be elicitable. This means that there exist some scoring function $S^q_\alpha(y, L), y \in \mathbb{R}$, that is consistent for the VaR_α.
- This scoring function induces a accuracy rewarding property in the sense that predictive distribution that produces VaR estimates that are closer to the "true" VaR will give a lower expected score.
- One example of such a scoring function for VaR_α is

$$S^q_\alpha(y, l) = |1_{\{l \leq y\}} - \alpha||l - y|.$$

- A simple rejection scheme:
  - For a realization L_i, $i = 1, \ldots, n$, choose a benchmark model G_B, and compute the benchmark score $S_{G_B} = \frac{1}{n}\sum_{i=1}^{n} S^q_\alpha(\text{VaR}^{G_{B,i}}_\alpha, L_i)$.
  - For a set of $\text{VaR}^{G_j}_\alpha$ that comes from unknown predictive distribution G_j, compute the associated score S_{G_j}, accept the $\text{VaR}^{G_j}_\alpha$ if $S_{G_j} \leq S_{G_B}$, reject otherwise.

Backtesting VaR: An Experiment

We conduct an experiment to test the power of the mentioned backtest methods:

1. Generate a sample data path of length 3000 using a GARCH(1,1) model with student-t innovations.
2. Fit the following model to the data to obtain the respective $\text{VaR}_{0.95}$:
   - GARCH(1,1) model with student-t innovations.
   - GARCH(1,1) model with standard normal innovations.
   - ARCH(1) model with student-t innovations.
   - ARCH(1) model with standard normal innovations.
   - Historical simulation method.
3. Backtest the obtained $\text{VaR}_{0.95}$ using previously discussed methods (for the score based method, we will use the dynamic historical simulation method as the benchmark model).
4. Repeat step one to step three 500 times to estimate the rejection rate of each test.

Backtesting VaR: An Experiment

Model	Binomial	Weibull	Score Based
GARCH t	2.8%	10.4%	1.6%
GARCH normal	13.4%	19.2%	6.7%
ARCH t	36.4%	88.0%	95.4%
ARCH normal	75.4%	99.0%	96.0%
Historical Simulation	42.2%	99.4%	97.2%

Table: Rejection rate of fitted models using different backtest methods.

Expected Shortfall

For a continuous loss distribution, the expected shortfall is given by the expression

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha} E[L; L > \mathrm{VaR}_\alpha] = E[L | L > \mathrm{VaR}_\alpha],$$

which is the expected loss given violation occurred. This is also known as the Tail Value at Risk (TVaR).
For a discontinuous loss distribution F_L, the formula for the expected shortfall becomes slightly more complicated, given by

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha} \left(E[L; L > \mathrm{VaR}_\alpha] + \mathrm{VaR}_\alpha (1 - \alpha - P(L \geq \mathrm{VaR}_\alpha)) \right).$$

In the second equation we have an extra continuity correction term.

Expected Shortfall: Graphical Representation

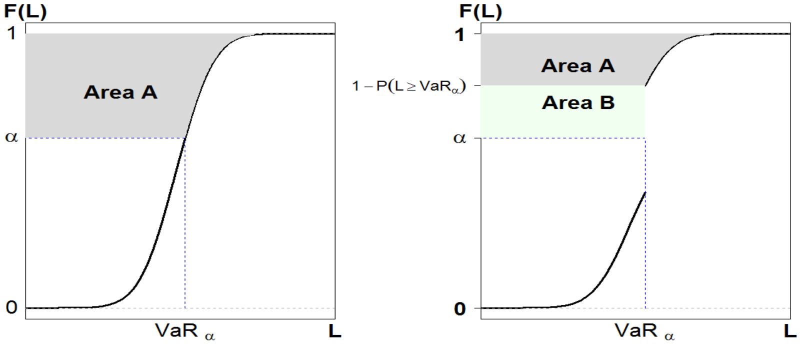

Figure: Left: F_L is continuous, Right: F_L is discontinuous

$$\begin{aligned} \text{Area A} &= E[L; L > \text{VaR}_\alpha] \\ \text{Area B} &= \text{VaR}_\alpha(1 - \alpha - P(L \geq \text{VaR}_\alpha)) \end{aligned}$$

Backtesting ES: Zero Mean Test

- McNeil et. al. (2005) have proposed to backtest ES using the zero mean test. We observe that ES_α can be written as

$$\text{ES}_\alpha = \text{VaR}_\alpha + \underbrace{(\text{ES}_\alpha - \text{VaR}_\alpha)}_{\text{Excess Loss}}.$$

- The VaR component can be backtested using previously mentioned methods.
- Given that the VaR estimate passes the test, the excess loss component can be backtested using the test statistic

$$S = (L - \text{ES}_\alpha)1_{\{L > \text{VaR}_\alpha\}},$$

  where S should have mean of zero.
- To test for zero mean, we can either use a bootstrap test similar to those discussed in Efron and Tibshirani (1994), which requires no assumption on the distribution of S, or we can use a standard one sample t test, with the assumption that S is i.i.d. distributed.

Backtesting ES: Zero Mean Test

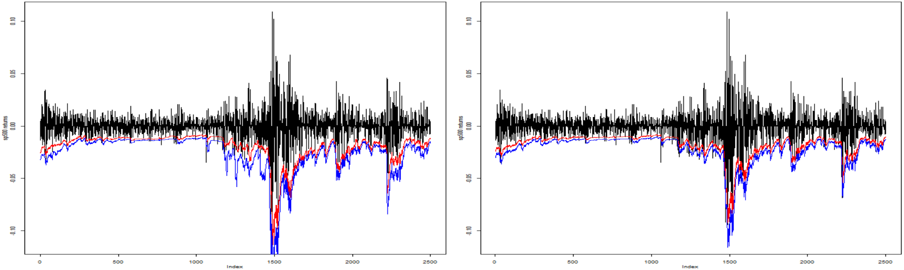

Figure: Comparison of fitted 1-step ahead $\text{VaR}_{0.95}$ (Red) and $\text{ES}_{0.95}$ (Blue) forecast of 2 models, with a 500 days rolling window size. Left: GARCH-EVT model, Right: GARCH-normal model.

- First Model: GARCH(1,1) dynamics with EVT innovations.
  Result: Binomial p-value of 0.263. Weibull p-value is 0.348.
- Second Model: GARCH(1,1) dynamics with N(0,1) innovations.
  Result: Binomial p-value of 0.176. Weibull p-value is 0.674.

Backtesting ES: Zero Mean Test

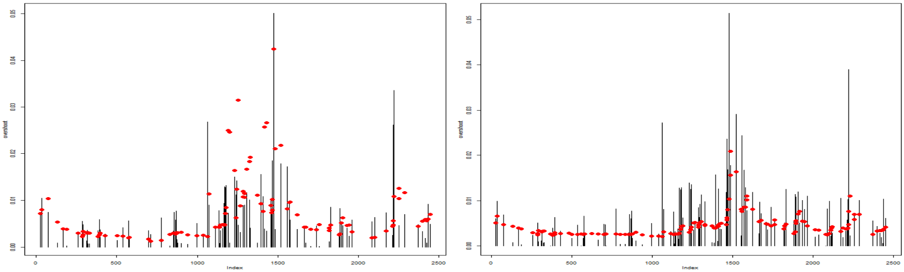

Figure: Observed excess loss and fitted excess loss (Red). Left: GARCH-EVT model, Right: GARCH-normal model.

Backtesting ES: Zero Mean Test

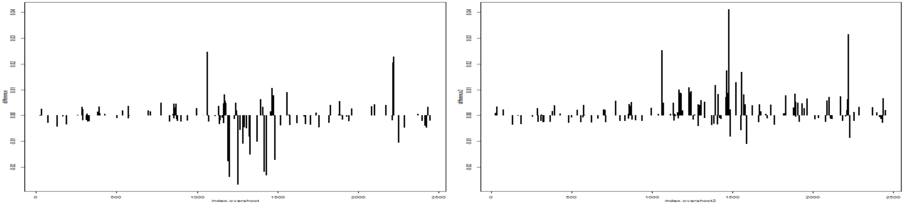

Figure: Difference of the observed excess loss and fitted excess loss. Left: GARCH-EVT model, Right: GARCH-normal model.

- First Model: GARCH(1,1) dynamics with EVT innovations.
  Result: Mean difference is -0.0008, t-test p-value is 0.867.
- Second Model: GARCH(1,1) dynamics with N(0,1) innovations.
  Result: Mean difference is 0.0026, t-test p-value is 0.
- Conclusion: Even though the GARCH normal model is okay for estimating the $\text{VaR}_{0.95}$ for this set of data, it is inadequate for estimating the $\text{ES}_{0.95}$.

Other Methods for Value at Risk

- We can model the violation series $1_t = 1_{\{L_t > \text{VaR}_\alpha^t\}}$ as a Bernoulli sequence $\text{Be}(p_t)$, of which under the null hypothesis, $p_t = p = 1 - \alpha$.
- Possible specification for p_t are:
  - CaViaR model, where $p_t = g(\theta_t)$, $\theta_t = \mu + \beta 1_t + \gamma \text{VaR}_\alpha^t$, where g is the link function, and possible g includes the probit function ($g = \Phi$) and logit function ($g(x) = (1 + \exp(-x))^{-1}$). (Berkowitz et al. (2011))
  - DQDB model, where $p_t = g(\theta_t)$, $\theta_t = \mu + \delta\theta_{t-1} + \beta 1_t + \gamma \text{VaR}_\alpha^t$. (Dumitrescu et al. (2011))

Other Methods for Value at Risk

Model	Binomial	Weibull	CaViaR	DQDB
GARCH t	2.8%	10.4%	6.6%	7.4%
GARCH normal	13.4%	19.2%	11.8%	12.2%
ARCH t	36.4%	88.0%	47.6%	96.8%
ARCH normal	75.4%	99.0%	98.4%	100.0%
HS	42.2%	99.4%	92.6%	100.0%

Table: Rejection rate of fitted models using different backtest methods, for $\alpha = 0.95$.

Other Methods for Expected Shortfall: Multinomial Test

- A good ES_α estimator should provide reliable VaR_u estimates for all $0 \leq \alpha \leq u \leq 1$.
- Hence, one way to backtest ES is to simultaneously backtest multiple VaR estimates computed using the same model used to compute the ES estimate.
- One way to do this is to simply count the number of realized loss that falls between the sets of VaR forecast, and test for the correct proportion using the goodness of fit test.

Backtesting ES: Multinomial Test

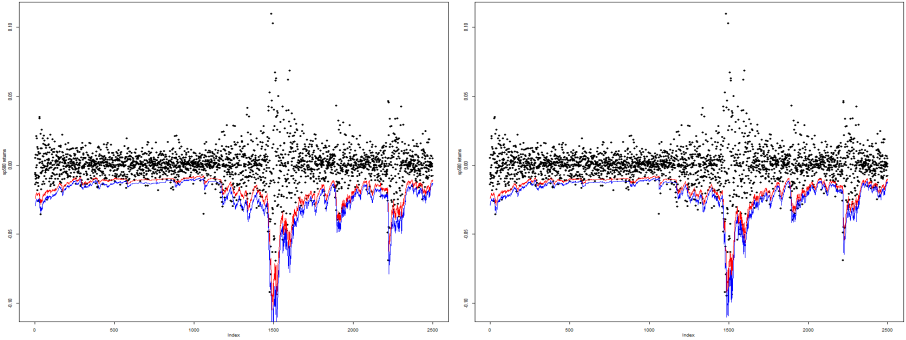

Figure: Proportion of loss that lies within each VaR interval, with $VaR_{0.95}$ (red) and $VaR_{0.975}$ (blue). Left: GARCH-HS model, Right: GARCH-normal model.

- GARCH-HS proportion: 94.76%, 2.32%, 2.92%, p-value=0.349.
- GARCH-normal proportion: 94.4%, 1.8%, 3.8%, p-value=0.000.

Summary

- We have reviewed the popular methods to backtest VaR and ES.
- For VaR, we have reviewed backtest methods based on violation rate, such as the Binomial test. We have also reviewed backtest methods based on duration between violation, such as the Weibull test.
- We have also seen how we can make use of the elicitability theory to reject models.
- For ES, we have the zero mean test for the excess loss provided that it first passes that test for VaR.

References

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D. (1999). Coherent measures of risk. Mathematical Finance 9(3), 203228.

Christofferssen, P. (1998). Evaluating Interval Forecasts. International Economic Review, 39, 841-862.

Christoffersen, Peter F. and Denis Pelletier (2004). Backtesting Value-at-Risk: A Duration-Based Approach. Journal of Financial Econometrics, 2, 84-108.

Efron, B. and Tibshirani, R.J. (1994). An Introduction to the Bootstrap, Chapman & Hall, New York.

Kupiec, P.H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. Journal of Derivatives, 3, 73-84.

McNeil, A.J., Frey, R., Embrechts, P. (2005). Quantitative Risk Management. Princeton University Press, Princeton, NJ.

The End