Model uncertainty in FHS risk models

Pedro Gurrola-Perez & Michael Wood
Bank of England

Workshop on Central Clearing Counterparties, LSE
24 May 2019.

This is a work in progress. The views expressed in this paper are those of the authors and not necessarily those of the Bank of England

Table of Contents

General modelling steps:

- Determine a statistical model which describes P&L dynamics. Example: GARCH(1,1) parameters estimated using MLE over a given estimation window.
- Selected a risk measure θ (e.g. VaR), and estimate a θ-point forecast at a fixed horizon.
- Test the model: Create a time series of forecasts obtained using a rolling window and compare with realized P&L at each point (backtesting).
- Apply some statistical test to interpret the backtesting results (i.e. assess, with certain confidence level, whether the model captures the true risk).

In the case of CCP margin models, it may look like this:

	CCP's
Statistical model describing P&L dynamics	FHS with EWMA volatility estimates over a +5yr window.
Estimate a ϕ point forecast for a fixed horizon.	VaR of CVaR, 2 to 7 day MPOR, 99% or 99.5% confidence level
Backtesting	Count exceptions over >1yr window
Statistical test to interpret the backtesting results	Kupiec's and Christoffersen's tests.

Potential sources of model risk

	CCP's	Model risk
Statistical model describing P&L dynamics	FHS with EWMA volatility estimates over a +5yr window.	Wrong model, parameter estimation error
Estimate a ϕ point forecast for a fixed horizon.	VaR of CVaR, 2 to 7 day MPOR, 99% or 99.5% confidence level	Inadequate risk measure or MPOR
Backtesting	Count exceptions over 2 to 5yr window	Inadequate testing model
Statistical test to interpret the backtesting results	Kupiec's and Christoffersen's tests.	

Questions

	CCP's	Model risk	Questions
Statistical model describing P&L dynamics	FHS with EWMA volatility estimates over a +5yr window.	Wrong model, parameter estimation error	Can we quantify uncertainty around the choice of parameters?
Estimate a ϕ point forecast for a fixed horizon.	VaR of CVaR, 2 to 7 day MPOR, 99% or 99.5% confidence level	Inadequate risk measure or MPOR	
Backtesting	Count exceptions over 2 to 5yr window	Inadequate testing model	How powerful are the tests? Is backtesting the right approach?
Statistical test to interpret the backtesting results	Kupiec's and Christoffersen's tests.		

Table of Contents

Filtered samples

- Aim is to incorporate a volatility updating scheme to increase the sensitivity of historical simulation models to the arrival of new information.

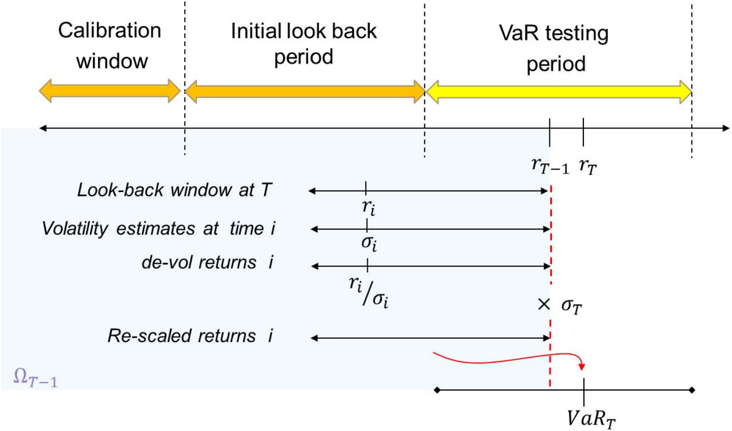
Filtered samples

- Aim is to incorporate a volatility updating scheme to increase the sensitivity of historical simulation models to the arrival of new information.
- Common approaches are variants of the Filtered Historical Simulation (FHS) methods suggested by John Hull and Allan White (1998) and Barone-Adesi, Bourgoin and Giannopoulos (1998).

Filtered samples

- Aim is to incorporate a volatility updating scheme to increase the sensitivity of historical simulation models to the arrival of new information.
- Common approaches are variants of the Filtered Historical Simulation (FHS) methods suggested by John Hull and Allan White (1998) and Barone-Adesi, Bourgoin and Giannopoulos (1998).
- Examples: initial margin methodologies for interest rate products used by LCH Swapclear, CME and Eurex (Gregory, 2014).

FHS


Calibration window
Initial look back period
VaR testing period
r_{T-1} r_T
Look-back window at T
r_i
Volatility estimates at time i
σ_i
de-vol returns i
r_i/σ_i
$\times$ σ_T
Re-scaled returns i
Ω_{T-1}
VaR_T

EWMA volatility estimates

- The conditional volatility estimates derived from an EWMA volatility updating scheme or from a GARCH process.
- EWMA recursive formula:

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)r_t^2 \tag{1}$$

The decay factor, $\lambda \in [0,1]$, determines the responsiveness of the process to the arrival of new information.

Is backtesting (+ Kupiec + Christoffersen) the correct tool?

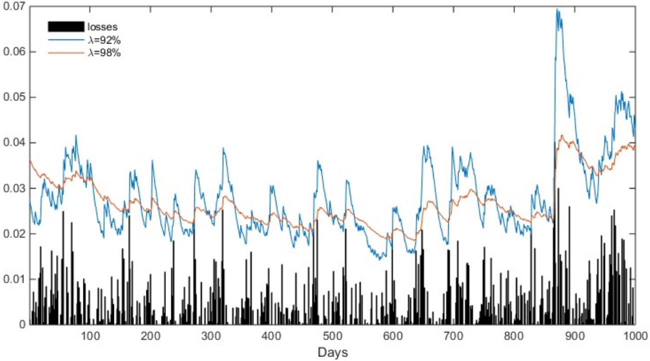


Figure : Backtesting of FHS VaR estimates for SPX using decay factors $\lambda = 0.92$ and $\lambda = 0.98$. Both cases produce 5 exceptions (which is exactly what the model predicts at 99.5% confidence). Both models pass the Kupiec's and Christoffersen's tests.

Analysis has shown the impact of FHS on higher moments (Gurrola and Murphy, 2015), and the need of more sophisticated testing tools (Gurrola, 2018). In particular,

- Hypothesis tests are largely insensitive to the dynamics resulting from different decay factors.
- They tend to favor overreacting calibrations, especially at high coverage levels, which is an undesirable outcome in terms of the procyclicality.
- Asymmetric piece-wise linear (APL) score functions improve performance, which is in line with Gneiting (2012) (applying scoring functions which are consistent for the α-quantile functional).
- Calibration and validation of the model cannot only rely on backtesting, even when accounting for the size and duration of the exceptions, and additional criteria needs to be considered.

Table of Contents

- *The problem:* Quantify the impact of parameter uncertainty on the output of VaR models that rely on EWMA volatility estimates, including its sensitivity to the time period considered.
- *Why it is important:* Quantifying the uncertainty of those outputs can help the risk manager take more informed and transparent decisions about the amount of initial margin required.
- *What we do:* We apply a Bayesian approach to quantify parameter uncertainty in EWMA-based VaR models.

Why Bayesian?

- Classical (frequentist) approach:
  - In general not well-suited to answer questions of parameter uncertainty because the only uncertainty they deal with is in sampling.
  - One typically identifies point estimates, such the MLE, of certain model parameters and, in doing so, overlooks the stochastic nature of the estimation of these parameters.
- Bayesian estimation:
  - Point estimates for parameters are substituted by probability distributions that describe the uncertainty surrounding the estimation process.
  - Allows the modeller to incorporate their prior knowledge.
  - The outcome is a joint posterior distribution of the model parameters and the projected portfolio outcomes.

EWMA as an IGARCH

EWMA process is a particular case of an integrated GARCH process (IGARCH), as defined by Engle and Bollerslev (1986). In general, an IGARCH(1,1) process has the following specification:

$$\begin{aligned} \sigma_t^2 &= \omega_0 + \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2 \\ r_t &\sim \mathcal{G}(0,\sigma_t) \end{aligned} \tag{2}$$

where $\mathcal{G}(0,\sigma_t)$ is a standardized distribution and ω_0 is the drift parameter. In our analysis, we will assume the drift is zero, so that the conditional volatility in (2) is an EWMA process.[1]

[1] Although a zero-drift IGARCH process has the undesirable property of converging almost surely to zero (Nelson, 1990), this should not be a problem when working at short term horizons.

We expand the IGARCH specification with zero-drift by converting the exogenous parameters λ and σ_0 into internal parameters that are modelled as random variables

$$
\begin{aligned}
(\lambda, \sigma_0) &\sim p(\lambda, \sigma_0) \\
(\sigma_t^2 | \lambda, \sigma_0) &= \lambda^t \sigma_0^2 + (1-\lambda) \sum_{i=1}^{t} r_{i-1}^2 \lambda^{t-i} \qquad (3) \\
(r_t | \sigma_t^2) &\sim \mathcal{G}(0, \sigma_t)
\end{aligned}
$$

Let $\hat{r}_i$ be a set of observed returns.We apply Bayes' rule

$$p(\lambda, \sigma_0 | \hat{r}_1, ..., \hat{r}_T) = \frac{p(\lambda, \sigma_0) \times p(\hat{r}_1, ..., \hat{r}_T | \lambda, \sigma_0)}{p(\hat{r}_1, ..., \hat{r}_T)} \tag{4}$$

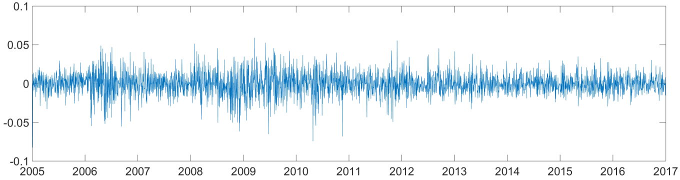
This posterior distribution can then be used to forecast the next unknown observable, r_{T+1}. These forecasts are the posterior predictive distributions and are expressed as a weighted average of the model's conditional predictions weighted by the posterior:

$$p(r_{T+1} | \hat{r}_T, ..., \hat{r}_1) = \int_0^\infty \int_0^1 p(r_{T+1} | \lambda, \sigma_0) p(\lambda, \sigma_0 | \hat{r}_T, ..., \hat{r}_1) d\lambda d\sigma_0 \tag{5}$$

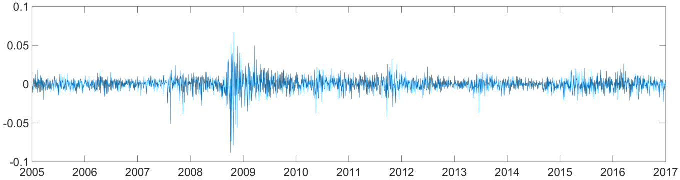
We will assess:

- The magnitude of the parameter uncertainty around λ for some typical market risk factors,
- The resulting uncertainty in the model's forecasts and on its accuracy,
- The impact of market data sample size on the level of uncertainty around λ.

Data: daily returns covering a 12-year period, from 3 January 2005 to 30 December 2016.

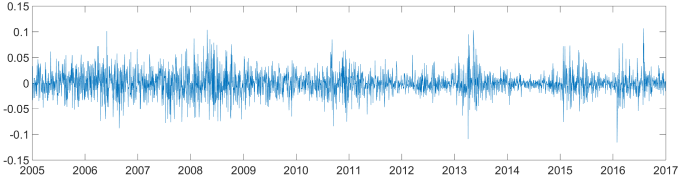


(a) Aluminiun 3 Month forward relative returns

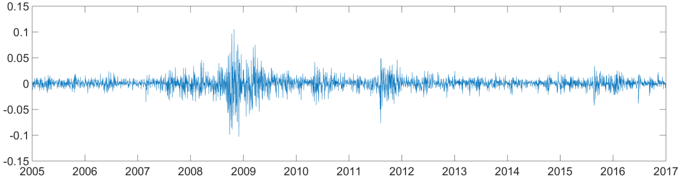


(b) AUS/USD FX spot rate relative returns

Data: daily returns covering a 12-year period, from 3 January 2005 to 30 December 2016.



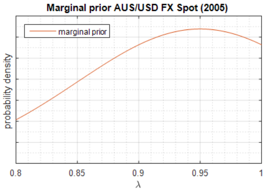
(c) Japan 10Y Bond yield absolute returns



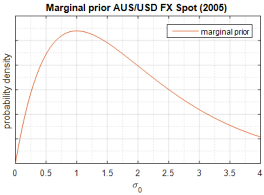
(d) S&P 500 Index relative returns

Prior distributions (chosen to match our prior information):

- $p(\lambda)$ is defined as a truncated normal distribution, with mean 0.95 and standard deviation 0.1, which is truncated outside the range $0.8 \leq \lambda \leq 1$ (and re-normalised);
- $p(\sigma_0)$ is defined as a gamma distribution, with shape parameter 2 and scale parameter 1, so that the peak value is at 1% and the standard deviation is $\sqrt{2}$.

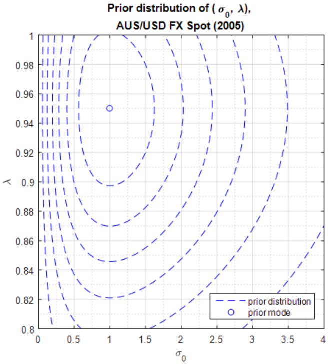


(e) The marginal prior $p(\lambda)$

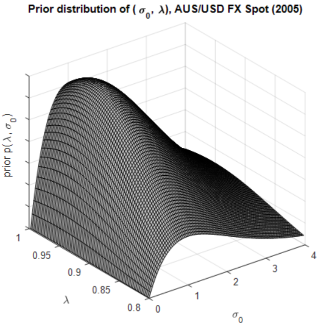


(f) The marginal prior $p(\sigma_0)$

Prior joint distribution:



(g) Contour plot



(h) Joint distribution

Likelihood function $p(\hat{r}_0, ..., \hat{r}_T | \lambda, \sigma_0)$:

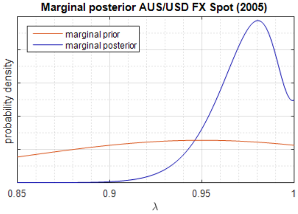
Estimation based on the approach proposed by Nakatsuma (1998): For each parameter pair (λ, σ_0) we express the joint distribution as a product of densities for each individual return, where each density is conditioned on the preceding returns:

$$
\begin{aligned}
p(\hat{r}_0, ..., \hat{r}_T | \lambda, \sigma_0) &= p(\hat{r}_0 | \lambda, \sigma_0) \times p(\hat{r}_1 | \hat{r}_0, \lambda, \sigma_0) \times \dots \\
\dots &\times p(\hat{r}_T | \hat{r}_0, ..., \hat{r}_{T-1}, \lambda, \sigma_0)
\end{aligned} \tag{6}
$$

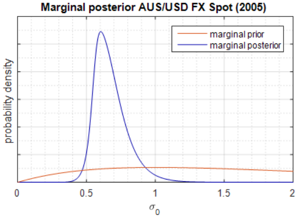
Posterior distribution of the parameters $p(\lambda, \sigma_0 | \hat{r}_0, \ldots, \hat{r}_T)$

- In more complex models with larger parameter spaces this is typically achieved by Markov-Chain Monte Carlo (MCMC) sampling.
- In our model, the parameter space is small enough to allow direct computation of the posterior on a grid of values in parameter space.
- Grid consisting of 2,001 uniformly-spaced points between $\lambda = 0.8$ and 1 (inclusive), and 1,501 uniformly-spaced points between $\sigma_0 = 0$ (exclusive) and an upper value that is 4% where relative returns are used, or 8 bps where absolute returns are used.
- The posterior is calculated directly at each grid point in parameter space and then normalising over the entire space.

Marginal posterior (AUS/USD)

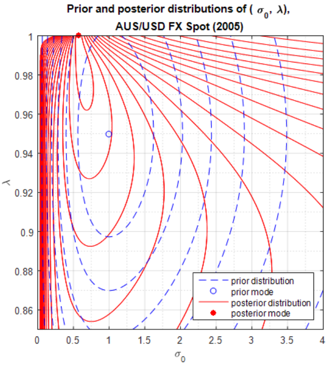


(i) λ

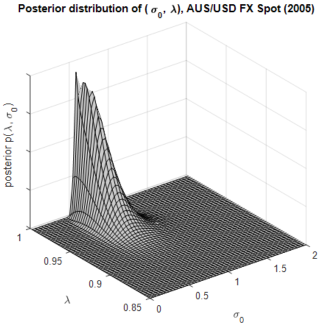


(j) σ_0

Posterior $p(\lambda, \sigma_0)$

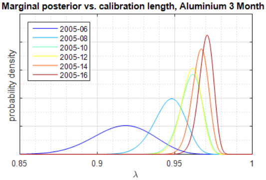


(k) Contour

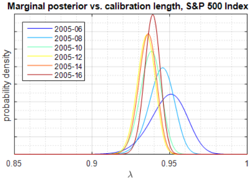


(l) Joint

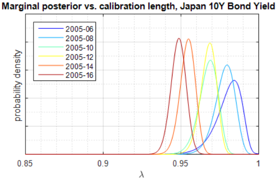
Evolution over time of the mean and standard deviations of marginal posteriors.



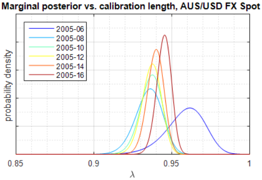
(m) Aluminium 3M



(n) S&P 500

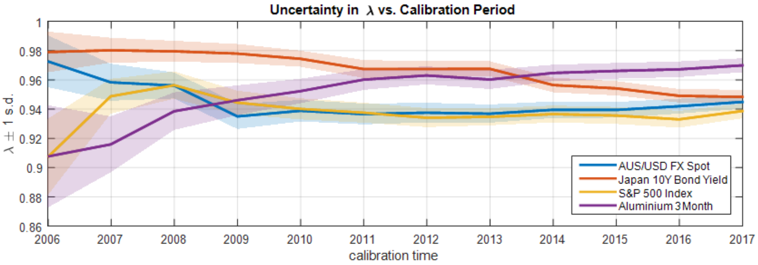


(o) Japan 10Y Bond

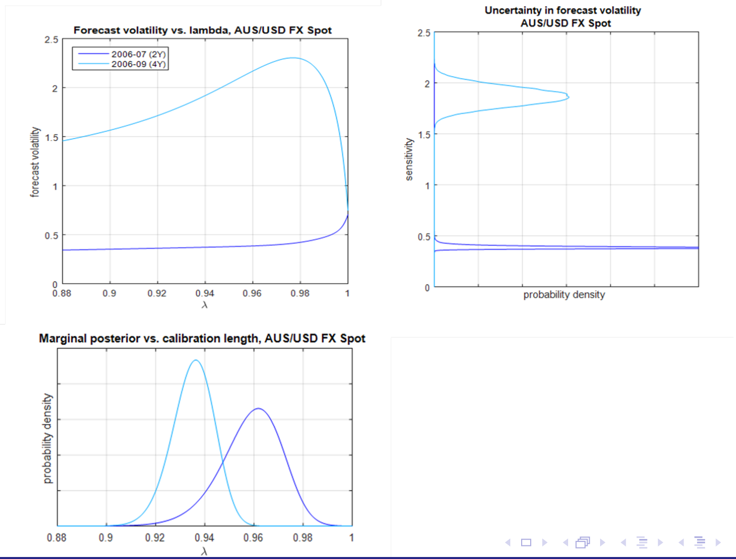


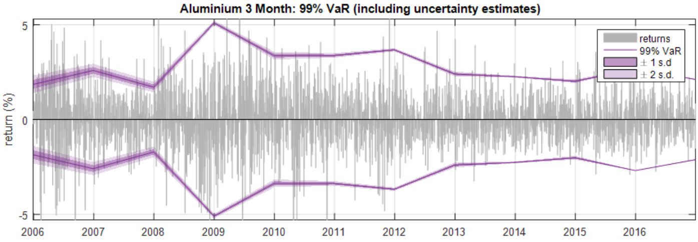
(p) AUS/USD FX spot rate

Evolution of the mean and standard deviation of the marginal posterior distributions $p(\lambda|\text{data})$ for six selected time periods.

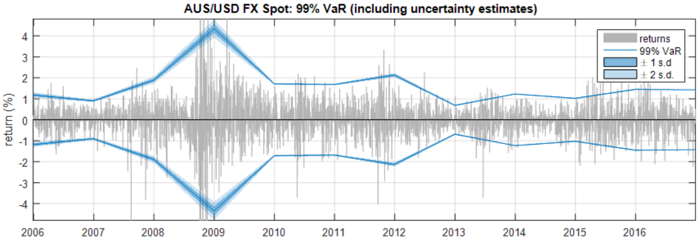


The propagation of uncertainty into the model's outcome

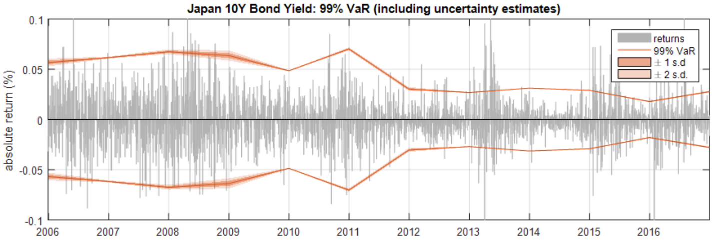




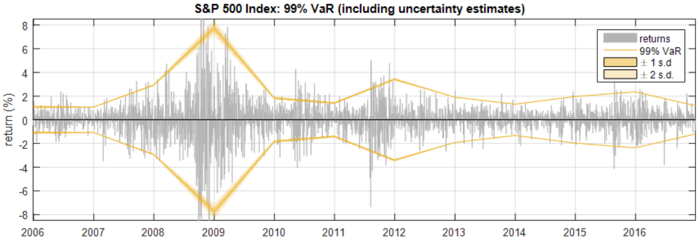
(q) Aluminium 3 Month 1-day relative returns



(r) AUS/USD FX spot 1-day relative returns



(s) Japan 10Y Bond yield 1-day absolute returns



(t) S&P 500 Index 1-day relative returns

Dispersion of the marginal posterior distributions, using the standard deviation of this distribution as a metric for parameter uncertainty.

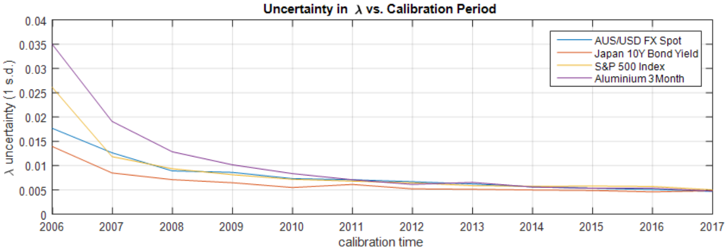


Figure : Plots of the standard deviations of the posterior distributions $p(\lambda|\hat{r}_0, ..., \hat{r}_T)$ for all six time periods and all four data series. The points have been linearly joined only for illustration purposes.

Table of Contents

- Bayesian inference provides a useful framework for modelling parameter uncertainty in EWMA estimates.
- The model specification appears unstable, which reduces confidence in using the EWMA-VaR approach to accurately estimate quantile measures of risk.
- Propagation method causes more variation in prediction uncertainty than changes in parameter uncertainty, over time.

- Understanding and monitoring such uncertainty may help improving risk management practices in various ways. In the case of CCPs, for example by
    - Considering richer processes for the evolution of returns.
    - Articulating a model risk tolerance by specifying what could be the maximum acceptable amount of uncertainty around the model outputs.
    - Monitoring uncertainty around outputs and use it as key indicator of potential model failure.
- The benefits of model accuracy should be balanced against other priorities as, for example, the economic cost of calling additional resources, or the importance of calibrating models in a way that they are not prone to overreacting.

Thank you!

Barone-Adesi, G., Bourgoin,F., Giannopoulos,K. (1998), Don't look back. *Risk*, August.

Berkowitz, J., Christoffersen, P.F., and Pelletier, D. (2011), Evaluating Value -at-Risk Models with Desk-Level Data.Management Science, vol. 57, no. 12, December 2011, 2213-2227.

Blanco, C. and Ihle, G. (1998) , "How good is your VaR? Using backtesting to assess system performance", *Financial Engineering News*, August, 1-2.

Campbell, S. (2005), A Review of Backtesting and Backtesting Procedure, Finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C.(2005)

Christoffersen, P.F.(1998). Evaluating Interval Forecasts. *International Economic Re- view*, 39, 841-862.

Christoffersen, P.F., and Pelletier, D. (2004), Backtesting Value-at-Risk: A Duration-Based

Colletaz,G., Hurlin C., and Perignon C. (2011). The Risk Map: A New Tool for Backtesting Risk Models. Working paper, HEC Paris.

Davis, M. (2014). Consistency of risk measure estimates, Working Paper, Imperial College, 2014

Diebold, F.X., Schiermann, T., and Stroughair J., (1998). Pitfalls and opportinities in the use of Extreme Value Theory in riks management, in A.-P. N. Refenes, A.N. Burgess and J.D. Moody (eds.), Decision Technologies for Computational Finance, 3-12. Amsterdam: Kluwer Academic Publishers.

Dowd, K.(2005). Measuring market risk, 2nd Ed. Wiley Finance, London.

Engle, R.F. and T. Bollerslev (1986), Modeling the Persistence of Conditional Variances, *Econometric Reviews*, 5, 1-50.

Engle, R., Manganelli (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. Journal of Business and Economics Statistic, 22, 367-381.

Gneiting, T. (2012). Making and evaluating point forecasts, *Journal of the American Statistical Association*, 106:494, 746-762, DOI: 10.1198/jasa.2011.r10138.

Gregory, J. (2014), Central Counterparties. Mandatory clearing and bilateral margin requirements for OTC derivatives. Wiley.

Gurrola, P.(2018). Backtesting filtered historical simulation. *Journal of Model Risk Validation*, 2018.

Gurrola, P., Murphy, D.(2015). Filtered historical simulation Value-at-Risk models and their competitors . *Bank of England, Working Paper Series*, No. 525.

Haas, M. (2001), New Methods in Backtesting, Financial Engineering, Research Center Caesar, Bonn.

Hull J., and White A. (1998). Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk. Journal of Risk (Fall); Vol. 1, No. 1; Pages: 5-19.

Hurlin C., and Tokpavi, S.(2008). Une Evaluation des Procedures de Backtesting: Tout va pour le Mieux dans le Meilleur des Mondes, Finance, vol. 29, no. 1, 2008, 53-80.

Karimalis, E., Alexander, P., Cerezetti, F. (2016). Testing the test: How reliable are risk model backtesting results? *Bank of England, Bank Underground Blog*, 15th January 2016.

Kupiec, P. (1994). The performance of S&P 500 futures margins under the SPAN margining system, *The Journal of Futures Markets*, Vol 14, No 7, 789-811.

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

Lopez, J. (1998), Methods for Evaluating Value-at-Risk Estimates, *Economic Policy Review*, October 1998, 119-64.

Lopez, J. (1999), Regulatory Evaluation of Value-at-Risk Models, *Journal of Risk* 1, 37-64.

Nakatsuma, T. (1998). A Markov-Chain Sampling Algorithm for GARCH Models. *Studies in Nonlinear Dynamics and Econometrics*, July 1998, 3(2): 107-117.

Nelson, D.B. (1990), Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory*, 6, 318-334.

Zangari, P.,(1996), Estimation and Forecast, RiskMetrics Technical Document, 4th Edition, J.P. Morgan, New York.