Journal of Risk Model Validation Volume 6/Number 4, Winter 2012/13 (3–16)

Backtesting value-at-risk: a comparison between filtered bootstrap and historical simulation

Dario Brandolini
Symphonia SGR, Via Gramsci 7, 10121 Torino, Italy;
email: dario.brandolini@symphonia.it

Stefano Colucci
Symphonia SGR, Via Gramsci 7, 10121 Torino Italy;
email: stefano.colucci@symphonia.it

The purpose of this paper is to compare ex ante *value-at-risk (VaR) estimation produced by two risk models: historical simulation and Monte Carlo filtered bootstrap. We perform three tests: unconditional coverage, independence and conditional coverage. We present results on both* $\text{VaR}_{1\%}$ *and* $\text{VaR}_{5\%}$ *on a one-day horizon for the following indexes: S&P 500, Topix, Dax, MSCI United Kingdom, MSCI France, Italy Comit Globale, MSCI Canada, MSCI Emerging Markets and RJ/CRB. Our results show that the Monte Carlo filtered bootstrap approach satisfies conditional coverage for all tested indexes, while historical simulation has many rejection cases. We also test the two models in a regulatory framework (rolling window of 250 daily observations) and discuss the advantages of using a conditional coverage methodology to validate risk models.*

1 INTRODUCTION

In the last ten years financial markets have suffered many periods of turbulence, such as the dot-com bubble (2001–2), emerging markets fall (2004), the subprime financial crisis with the defaults of large investment banks (2008) and the sovereign debt crisis (2010–11), when losses reached values well above the Gaussian hypotheses in terms of frequencies and amounts. Many models can be used to estimate market risk but they must satisfy two conditions: unconditional coverage and independence. For backtesting purposes, the regulator (the Committee of European Securities Regulators (CESR), now the European Securities and Markets Authority) will accept no

The authors thank Loriana Pelizzon of Venice University for her useful comments during the XIII Workshop on Quantitative Finance. The views expressed in this paper represent those of the authors and not necessarily those of Symphonia SGR.

more than seven failures in a 250-day rolling window when $\mathrm{VaR}_{1\%}$ is estimated on a one-day horizon. This is a coverage condition, while no tests are required regarding independence.

The aim of this paper is to discuss and compare two different models (filtered bootstrap (FB) and historical simulation (HS)) with no Gaussian hypothesis over a relatively long period of time (from January 3, 2000 to August 22, 2011), and to then analyze the test results. We perform both coverage (unconditional and conditional) and independence tests to validate the risk models and check their compliance with regulatory rules.

The outline of the paper is as follows. Data is presented in Section 2. In Section 3 we briefly describe how value-at-risk (VaR) is estimated in the two models. In Sections 4 and 5, unconditional coverage independence and conditional coverage tests are discussed, together with the regulatory "hit function" for both models.

2 DATA PRESENTATION

We perform both $\mathrm{VaR}_{1\%}$ and $\mathrm{VaR}_{5\%}$ estimations on a one-day horizon for the FB approach (Barone-Adesi *et al* (1999)) and the HS approach for the following indexes: S&P 500, Topix (TPX), Dax (DAX), MSCI United Kingdom (MXGB), MSCI France (MXFR), Italy Comit Globale (COMIT), MSCI Canada (MXCA), MSCI Emerging Markets (MXEF), RJ/CRB (CRYTR) in US dollars and also in local currency (TPX LC, DAX LC, MXGB LC, MXFR LC, COMIT LC, MXCA LC) if the US dollar is not the original currency. Value-at-risk estimation will cover the period from January 3, 2000 to August 22, 2011, but data sets start 500 days earlier in order to initialize the estimation. All indexes follow the US financial calendar so there are no missing values. If one market is closed, the previous value is replicated and the return for that day is zero. In that way we have the same data set for all markets and it is possible to evaluate the whole sample in joint VaR estimation.

3 VAR ESTIMATION: FILTERED BOOTSTRAP AND HISTORICAL SIMULATION

A simple HS model specification is applied. For each index and each working day, we calculate the 1% and 5% quantiles of the distribution of equally weighted realized returns of the previous 500 working days. Because we analyze each index independently, this one is the risk factor that we want to estimate. Thus, it is sufficient to calculate the historical percentile (1 and 5 in our case) of the index realized distribution on 500 working days to obtain the VaR estimate for next day. There are many methods of computing quantile Q_p. We use the default quantile MATLAB estimator

$Q_p = \frac{1}{2}(x_{\lceil h-1/2 \rceil} + x_{\lfloor h+1/2 \rfloor})$, where $h = Np + \frac{1}{2}$, p is the probability and N is the sample size. This estimator is slightly more conservative than the Excel function.

The FB model[1] step by step procedure is as follows.

(1) Fit the best autoregressive moving average generalized autoregressive conditional heteroskedasticity (ARMA-GARCH) model.

(2) Standardize residuals: divide residuals by estimated sigmas.

(3) Bootstrap standardized residuals.

(4) Pass the bootstrapped residuals in a forward simulation using the ARMA-GARCH estimated model.

(5) Collect estimated returns.

(6) Calculate VaR on the distribution of returns.

Each day, the best ARMA(p, q)-exponential general autoregressive conditional heteroskedastic (EGARCH)(p, q) model is fitted to the data according to the following hierarchy.

- ARMA(1,0)-EGARCH(1,1): if the AR coefficient is statistically significant.
- ARMA(0,1)-EGARCH(1,1): if the MA coefficient is statistically significant.
- ARMA(0,0)-EGARCH(1,1): if the leverage coefficient is statistically significant.
- ARMA(1,0)-GARCH(1,1): if the AR coefficient is statistically significant.
- ARMA(0,1)-GARCH(1,1): if the MA coefficient is statistically significant.
- ARMA(0,0)-GARCH(1,1).

Using the best model estimate, the simulation process then generates 1000 scenarios and calculates 1% and 5% quantiles of the distribution of returns; again, we use the same estimator as in HS.

[1] For details on the filtered bootstrap approach, refer to Barone Adesi *et al* (1999), Brandolini *et al* (2000) and Pallotta and Zenti (2000).

4 TEST AND RESULTS

In this section we apply the benchmark test proposed by Christoffersen and Pelletier (2004). Consider a time series of daily *ex post* portfolio returns, R_t, $t \in 1, \dots, T$, and a corresponding time series of *ex ante* value-at-risk forecasts, $\mathrm{VaR}_t(p)$, $t \in 1, \dots, T$, with expected coverage rate p, such that ideally:

$$\Pr_{t-1}(R_t < \mathrm{VaR}_t(p)) = p$$

Define the hit sequence of VaR_t violations as:

$$I_t = \begin{cases} 1 & \text{if } R_t < \mathrm{VaR}_t(p) \\ 0 & \text{otherwise} \end{cases}$$

Note that the hit sequence discards the information regarding the size of violations.

Christoffersen (1998) tests the null hypothesis that:

$$I_t \sim \text{iid Bernoulli}(p)$$

against the alternative that:

$$I_t \sim \text{iid Bernoulli}(\pi)$$

and refers to this as the test of correct unconditional coverage (UC):

$$H_{0,\mathrm{uc}} : \pi = p$$

thus, testing that on average the coverage is correct. In fact, the above test implicitly assumes that the hits are independent: an assumption that has to be tested explicitly. In order to test this hypothesis, an alternative test is proposed where the hit sequence follows a first-order Markov sequence with switching probability matrix:

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}$$

where π_{ij} is the probability of an event i on day $t - 1$ being followed by an event j on day t.

The test of independence (ind) is then:

$$H_{0,\mathrm{ind}} : \pi_{01} = \pi_{11}$$

Finally, the two tests can be combined in a test of conditional coverage (CC):

$$H_{0,\mathrm{CC}} : \pi_{01} = \pi_{11} = p$$

The idea behind the Markov alternative is that clustered violations might be a signal of risk model misspecification. Violation clustering is important as it suggests repeated severe losses, which together could result in bankruptcy.

The likelihood function for a sample $I_1, \dots, I_T$ of iid observations from a Bernoulli variable with known probability p is written as:

$$L(I, p) = p^{T_1}(1 - p)^{T - T_1}$$

where $T_1 = \sum_{t=1}^{T} I_t$ is the number of 1s in the sample. The likelihood function for an iid Bernoulli variable with unknown probability parameter, π_1, to be estimated is:

$$L(I, \pi_1) = \pi_1^{T_1}(1 - \pi_1)^{T - T_1}$$

The maximum-likelihood (ML) estimate of π_1 is:

$$\hat{\pi}_1 = T_1/T$$

and we can thus write a likelihood ratio test of unconditional coverage as:

$$\text{LR}_{\text{UC}} = 2[\ln L(I, \hat{\pi}_1) - \ln L(I, p)]$$

For the independence test, the likelihood under the alternative hypothesis is:

$$L(I, \pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_0} - T_{01}\pi_{01}^{T_{01}}(1 - \pi_{11})^{T_1} - T_{11}\pi_{11}^{T_{11}}$$

where T_{ij} denotes the number of observations when a j follows an i. The ML estimates are:

$$\hat{\pi}_{01} = \frac{T_{01}}{T_0} \quad \text{and} \quad \hat{\pi}_{11} = \frac{T_{11}}{T_1}$$

and the independence test statistic is:

$$\text{LR}_{\text{ind}} = 2[\ln L(I, \hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(I, \hat{\pi}_1)]$$

Finally, the test of conditional coverage is written as:

$$\text{LR}_{\text{CC}} = 2[\ln L(I, \hat{\pi}_{01}, \hat{\pi}_{11}) - \ln L(I, p)]$$

We note that all the tests are carried out conditionally on the first observation. The tests are asymptotically chi-square distributed with one degree of freedom for the UC and ind tests and two for the CC test.

TABLE 1 Unconditional coverage test on filtered bootstrap VaR estimation (in percent).

Index	π when $p = 1\%$	p-value	π when $p = 5\%$	p-value
S&P 500	1.47	1.70	5.64	11.98
TPX	1.23	22.68	5.26	51.71
TPX LC	1.37	5.87	5.50	21.95
DAX	1.44	2.63	6.12	**0.73**
DAX LC	1.47	1.70	6.05	1.16
MXGB	1.67	**0.08**	5.88	51.71
MXGB LC	1.40	3.97	5.88	3.37
MXFR	1.57	**0.41**	6.56	**0.00**
MXFR LC	1.33	8.51	5.60	14.05
COMIT	1.47	1.70	6.25	**0.27**
COMIT LC	1.47	1.70	5.78	5.99
MXCA	1.47	1.70	5.91	2.75
MXCA LC	1.64	**0.14**	6.63	**0.00**
MXEF	1.33	8.51	5.98	1.80
CRYTR	1.44	2.63	5.23	57.26

Finally, as a practical matter, if the sample has $T_{11} = 0$, as can easily happen in small samples and with small coverage rates, then the first-order Markov likelihood is computed as:

$$L(I, \pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_0 - T_{01}} \pi_{01}^{T_{01}}$$

and the tests are carried out as above.

Following this procedure, as in Christoffersen and Pelletier (2004), we perform all three tests (UC, ind and CC) for the two methods. A confidence interval of 99% is chosen to make sure that we reject the null hypothesis only when the estimate is far from the expected value. Table 1 and Table 2 on the facing page display the results of the UC tests for $\text{VaR}_{1\%}$ and $\text{VaR}_{5\%}$, as estimated using FB and HS. Hypotheses rejected al 99% confidence are given in bold. For FB, there are 3 rejections in the $\text{VaR}_{1\%}$ case, and 4 in the $\text{VaR}_{5\%}$ case. For HS, the numbers of rejections are 8 and 3, so FB seems to be superior in terms of unconditional coverage.

Table 3 on the facing page and Table 4 on page 10 display the corresponding results of independence tests. In these, FB performs much better (1 and 2 rejections at 99% confidence) than HS (8 and 11 rejections), the latter being vulnerable to highly dependent VaR estimations, indicated by clustering of ones in the hit function.

Table 5 on page 10 and Table 6 on page 11 display the corresponding results of the conditional coverage test. Again, FB performs much better (5 and 3 rejections at 99% confidence) than HS (13 and 11 rejections).

TABLE 2 Unconditional coverage test on historical simulation VaR estimation (in percent).

Index	π when $p = 1\%$	p-value	π when $p = 5\%$	p-value
S&P 500	1.88	**0.00**	5.64	11.98
TPX	1.16	39.06	5.50	21.95
TPX LC	1.50	1.08	5.13	75.46
DAX	1.67	**0.08**	5.50	5.13
DAX LC	1.61	**0.24**	6.08	**0.92**
MXGB	1.78	**0.01**	6.02	1.45
MXGB LC	1.67	**0.08**	5.67	10.16
MXFR	1.71	**0.05**	6.39	**0.05**
MXFR LC	1.50	1.08	6.36	**0.12**
COMIT	1.40	3.97	5.84	4.11
COMIT LC	1.50	1.08	5.81	4.97
MXCA	1.47	1.70	5.74	7.19
MXCA LC	1.57	**0.41**	5.43	28.78
MXEF	1.30	12.06	5.57	16.39
CRYTR	1.67	**0.08**	5.43	28.78

TABLE 3 Independence test on filtered bootstrap VaR estimation.

	$p = 1\%$				$p = 5\%$			
Index	π_{01}	π_{11}	$\mathbf{LR_{ind}}$	p-**value**	π_{01}	π_{11}	$\mathbf{LR_{ind}}$	p-**value**
S&P 500	1.39%	6.98%	4.88	2.72%	5.69%	4.85%	0.21	64.32%
TPX	1.14%	8.33%	6.74	**0.94**%	5.23%	5.84%	0.11	74.40%
TPX LC	1.35%	2.50%	0.31	57.68%	5.46%	6.21%	0.16	69.03%
DAX	1.46%	0.00%	1.22	26.87%	6.08%	6.70%	0.11	73.89%
DAX LC	1.46%	2.33%	0.19	66.43%	6.29%	2.26%	6.09	1.36%
MXGB	1.60%	6.12%	3.62	5.70%	5.88%	5.81%	0.00	97.04%
MXGB LC	1.39%	2.44%	0.27	60.57%	5.81%	6.98%	0.38	53.84%
MXFR	1.49%	6.52%	4.22	4.00%	6.51%	7.29%	NaN	**0.00%**
MXFR LC	1.32%	2.56%	0.36	54.83%	5.81%	6.71%	0.38	53.87%
COMIT	1.49%	0.00%	1.28	25.74%	6.23%	6.56%	0.03	86.21%
COMIT LC	1.49%	0.00%	1.28	25.74%	5.84%	4.73%	0.38	53.83%
MXCA	1.39%	6.98%	4.88	2.72%	5.78%	8.09%	1.43	23.22%
MXCA LC	1.60%	4.17%	1.37	24.24%	6.52%	8.25%	NaN	**0.00%**
MXEF	1.35%	0.00%	1.05	30.47%	5.96%	6.29%	0.03	86.18%
CRYTR	1.42%	2.38%	0.23	63.48%	5.30%	3.92%	0.60	43.69%

NaN (“not a number”) denotes test failed.

TABLE 4 Independence test on historical simulation VaR estimation.

Index	$p = 1\%$				$p = 5\%$			
	π_{01}	π_{11}	LR_{ind}	p-value	π_{01}	π_{11}	LR_{ind}	p-value
S&P 500	1.81%	5.45%	2.61	10.65%	5.40%	9.70%	4.56	3.28%
TPX	1.04%	11.76%	12.12	**0.05%**	5.21%	10.56%	6.79	**0.92%**
TPX LC	1.42%	6.82%	4.65	3.10%	5.31%	7.55%	1.32	25.07%
DAX	1.49%	12.24%	14.68	**0.01%**	4.32%	20.00%	44.64	**0.00%**
DAX LC	1.46%	10.64%	11.23	**0.08%**	5.53%	14.61%	18.13	**0.00%**
MXGB	1.60%	11.54%	13.35	**0.03%**	5.27%	17.61%	31.21	**0.00%**
MXGB LC	1.49%	12.24%	14.68	**0.01%**	5.11%	15.06%	20.95	**0.00%**
MXFR	1.53%	12.00%	14.22	**0.02%**	5.91%	13.37%	12.87	**0.03%**
MXFR LC	1.39%	9.09%	8.25	**0.41%**	6.06%	10.75%	5.48	1.93%
COMIT	1.28%	9.76%	9.28	**0.23%**	5.48%	11.70%	9.05	**0.26%**
COMIT LC	1.42%	6.82%	4.65	3.10%	5.30%	14.12%	17.02	**0.00%**
MXCA	1.42%	4.65%	1.96	16.13%	5.08%	16.67%	27.53	**0.00%**
MXCA LC	1.49%	6.52%	4.22	4.00%	4.95%	13.84%	16.92	**0.00%**
MXEF	1.28%	2.63%	0.41	52.03%	4.63%	21.47%	52.13	**0.00%**
CRYTR	1.63%	4.08%	1.26	26.10%	5.31%	7.55%	1.32	25.07%

TABLE 5 Conditional coverage test on filtered bootstrap VaR estimation.

Index	$p = 1\%$				$p = 5\%$			
	π_{01}	π_{11}	LR_{CC}	p-value	π_{01}	π_{11}	LR_{CC}	p-value
S&P 500	1.39%	6.98%	10.57	**0.51%**	5.69%	4.85%	2.64	26.78%
TPX	1.14%	8.33%	8.20	1.65%	5.23%	5.84%	0.53	76.86%
TPX LC	1.35%	2.50%	3.88	14.34%	5.46%	6.21%	1.67	43.46%
DAX	1.46%	0.00%	6.16	4.59%	6.08%	6.70%	7.32	2.58%
DAX LC	1.46%	2.33%	5.88	5.28%	6.29%	2.26%	12.46	**0.20%**
MXGB	1.60%	6.12%	14.80	**0.06%**	5.88%	5.81%	4.51	10.48%
MXGB LC	1.39%	2.44%	4.50	10.56%	5.81%	6.98%	4.89	8.68%
MXFR	1.49%	6.52%	12.46	**0.20%**	6.51%	7.29%	NaN	**0.00%**
MXFR LC	1.32%	2.56%	3.33	18.96%	5.81%	6.71%	2.55	27.93%
COMIT	1.49%	0.00%	6.98	3.06%	6.23%	6.56%	9.04	1.09%
COMIT LC	1.49%	0.00%	6.98	3.06%	5.84%	4.73%	3.92	14.10%
MXCA	1.39%	6.98%	10.57	**0.51%**	5.78%	8.09%	6.29	4.32%
MXCA LC	1.35%	4.17%	4.02	13.41%	6.52%	8.25%	5.62	6.01%
MXEF	1.42%	0.00%	11.53	**0.31%**	5.96%	6.29%	NaN	**0.00%**
CRYTR	1.42%	2.38%	5.16	7.56%	5.30%	3.92%	0.92	63.04%

NaN (“not a number”) denotes test failed.

TABLE 6 Conditional coverage test on historical simulation VaR estimation.

	$p = 1\%$				$p = 5\%$			
Index	π_{01}	π_{11}	**LR**$_{CC}$	p**-value**	π_{01}	π_{11}	**LR**$_{CC}$	p**-value**
S&P 500	1.81%	5.45%	20.78	**0.00%**	5.40%	9.70%	6.98	3.05%
TPX	1.04%	11.76%	12.85	**0.16%**	5.21%	10.56%	8.30	1.58%
TPX LC	1.42%	6.82%	11.15	**0.38%**	5.31%	7.55%	2.45	29.38%
DAX	1.49%	12.24%	25.86	**0.00%**	4.32%	20.00%	44.74	**0.00%**
DAX LC	1.46%	10.64%	20.40	**0.00%**	5.53%	14.61%	24.92	**0.00%**
MXGB	1.60%	11.54%	27.85	**0.00%**	5.27%	17.61%	37.18	**0.00%**
MXGB LC	1.49%	12.24%	25.86	**0.00%**	5.11%	15.06%	23.63	**0.00%**
MXFR	1.53%	12.00%	26.47	**0.00%**	5.91%	13.37%	23.86	**0.00%**
MXFR LC	1.39%	9.09%	14.75	**0.06%**	6.06%	10.75%	15.96	**0.03%**
COMIT	1.28%	9.76%	13.51	**0.12%**	5.48%	11.70%	13.22	**0.13%**
COMIT LC	1.42%	6.82%	11.15	**0.38%**	5.30%	14.12%	20.87	**0.00%**
MXCA	1.42%	4.65%	7.66	2.18%	5.08%	16.67%	30.77	**0.00%**
MXCA LC	1.49%	6.52%	12.46	**0.20%**	4.95%	13.84%	18.05	**0.01%**
MXEF	1.28%	2.63%	2.82	24.37%	4.63%	21.47%	54.07	**0.00%**
CRYTR	1.63%	4.08%	12.45	**0.20%**	5.31%	7.55%	2.45	29.38%

We conclude that FB seems to be superior to HS in terms of UC, independence and CC. In particular, FB is less vulnerable to periods of clustered large losses.

5 REGULATORS' BACKTESTING PROCEDURE

The CESR's guidelines[2] state that backtesting results must be in line with the selected $\text{VaR}_{1\%}$ confidence interval, ie, in the last 250 rolling days the hit function must present at most seven failures (hits or overshootings) at 99% confidence or six failures at 95% confidence. We test $\text{VaR}_{1\%}$ at the 95% confidence level to have an early warning on model performances, which can be translated in terms of frequency as $\pi \in [0\%, 2.40\%]$ (see Table 7 on the next page).

If the hit ratio is higher than 2.40%, then some kind of measure has to be taken in order to reduce the risk model misspecification, while, if there are no failures, nothing

[2] "Where the backtesting results give rise to consistently inaccurate estimates and an unacceptable number of 'overshootings' (that is to say, that the number of 'overshootings' is not in line with the confidence interval selected for the calculation of the VaR), competent authorities reserve the right to take measures and eg, apply stricter criteria to the use of VaR or, if need be, to disallow the use of the model for the purpose of measuring global exposure. The competent authorities may, for example, also require that results of the calculation of the UCITS VaR to be scaled up by a multiplication factor." (See CESR (2010).)

TABLE 7 Number of hits, corresponding frequency and associated p-value.

Number of hits	Frequency (%)	p-value (%)
1	0.4	27.8
2	0.8	74.2
3	1.2	75.8
4	1.6	38.0
5	2.0	16.2
6	2.4	5.9
7	2.8	1.9

has to be done, even if this is statistically incorrect. In the following paragraphs we will examine the empirical results applying this backtesting procedure to $\text{VaR}_{1\%}$ forecasts on a one-day horizon with both FB and HS. Figure 1 on the facing page shows, for nine indexes, the π time series for the two models (FB and HS) together with the frequency level consistent with $\text{VaR}_{1\%}$ forecasts (2.4%). Typically, the FB model is compliant more time than the HS model, the latter suffering from overshooting dependence.

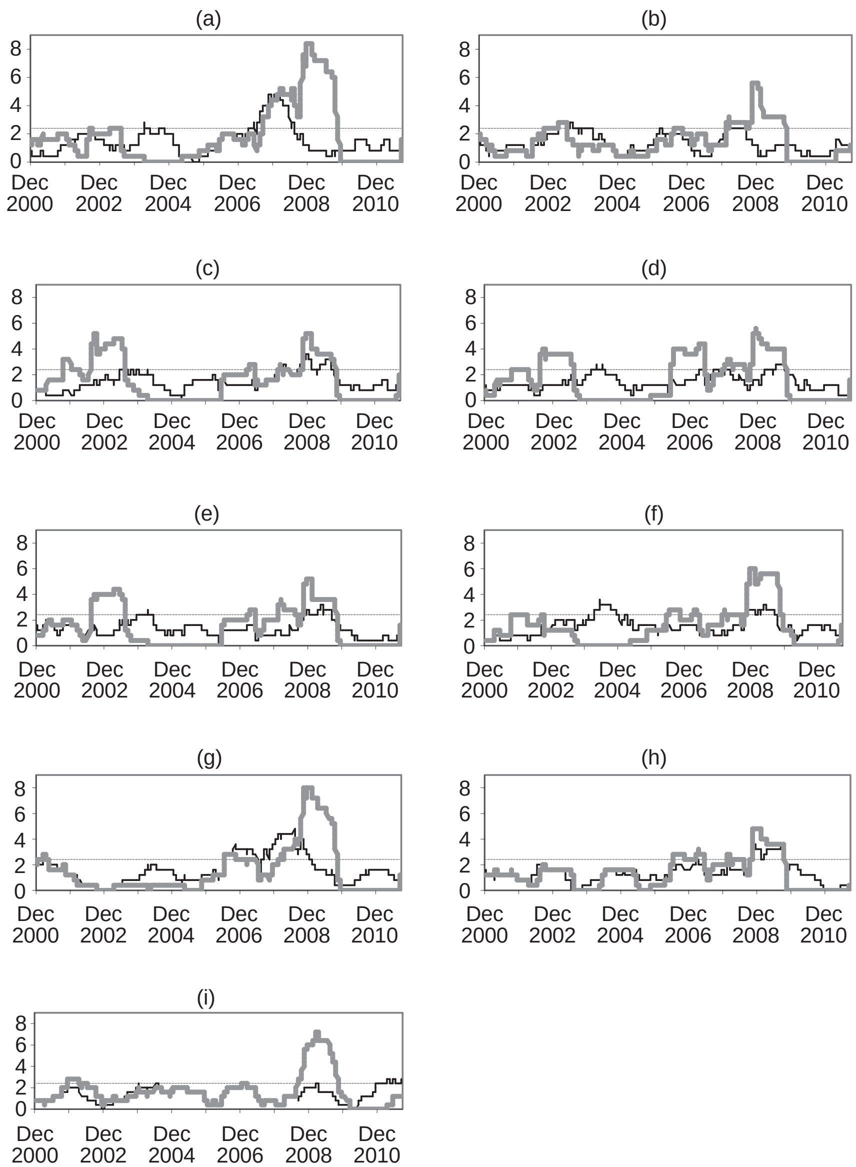
In Table 8 on page 14 and Table 9 on page 15 we summarize the results of the extensive backtesting procedure performed for filtered bootstrap and historical simulation models on a sample of equity indexes (in local currency and in US$) according to regulatory rules.

When FB overshoots the backtesting hit ratio, the number of failures is always lower than in the HS case, meaning that the FB model is wrong for a shorter period of time (see the first two columns of Table 8 on page 14). This is particularly clear during 2002 and 2008, when market volatility was very high, and FB was better able than HS to catch risk jumps in equity markets.

Thanks to the independence property, the FB model is compliant with the maximum failures allowed by regulators in every market for a longer period of time than HS (see the last two columns of Table 9 on page 15). On average, FB has 3.6 violations but with a standard deviation of 2, and HS has 3.8 violations with a standard deviation of 4.

The ability of the FB model to quickly adjust the risk forecast to market conditional volatility is also evident in the opposite way, ie, when volatility is going down. Therefore, the FB model, having a greater ability to catch the ups and downs of portfolio risk estimates, is more suitable not only to comply with regulatory rules, but also to be used for a portfolio risk policy that dynamically controls the portfolio risk budget utilization.

FIGURE 1 π time series in FB and HS (in percent).



(a) S&P 500 VaR backtest regulator framework. (b) Topix VaR backtest regulator framework. (c) DAX VaR backtest regulator framework. (d) MSCI GB VaR backtest regulator framework. (e) MSCI FRANCE VaR backtest regulator framework. (f) COMIT VaR backtest regulator framework. (g) MSCI CANADA VaR backtest regulator framework. (h) MSCI EMERGING VaR backtest regulator framework. (i) CRYTR VaR backtest regulator framework. Thick black line: FB. Gray line: HS. Dashed black line: upper bound $\mathrm{VaR}_{1\%}$. All values are in local currency.

TABLE 8 Summary statistics of two models with respect to regulatory rules.

Index	FB Number of days not compliant	HS Number of days not compliant	FB Number of days not compliant in 2002 and 2008	HS Number of days not compliant in 2002 and 2008	Max π FB	Max π HS
S&P 500	292 (10.9%)	545 (20.4%)	135 (27.0%)	253 (50.6%)	4.80%	8.40%
TOPIX	207 (7.7%)	80 (3.0%)	0 (0.0%)	58 (11.6%)	3.20%	4.00%
TOPIX LC	111 (4.1%)	454 (17.0%)	24 (4.8%)	214 (42.8%)	2.80%	5.60%
DAX	229 (8.6%)	828 (30.9%)	13 (2.6%)	269 (53.8%)	3.20%	6.80%
DAX LC	213 (8.0%)	629 (23.5%)	100 (20.0%)	198 (39.6%)	3.60%	5.20%
MSCI UK	311 (11.6%)	853 (31.9%)	33 (6.6%)	316 (63.2%)	3.60%	6.40%
MSCI UK LC	82 (3.1%)	894 (33.4%)	0 (0.0%)	318 (63.6%)	2.80%	5.60%
MSCI France	426 (15.9%)	905 (33.8%)	23 (4.6%)	341 (68.2%)	3.60%	6.40%
MSCI France LC	119 (4.4%)	664 (24.8%)	23 (4.6%)	305 (61.0%)	3.20%	5.20%
COMIT	181 (6.8%)	474 (17.7%)	0 (0.0%)	67 (13.4%)	4.00%	6.00%
COMIT LC	367 (13.7%)	387 (14.5%)	62 (12.4%)	75 (15.0%)	3.60%	6.00%
MSCI Canada	111 (4.1%)	413 (15.4%)	111 (22.2%)	221 (44.2%)	2.80%	7.60%
MSCI Canada LC	567 (21.2%)	522 (19.5%)	235 (47.0%)	221 (44.2%)	4.80%	8.00%
MSCI Emerging Markets	225 (8.4%)	400 (14.9%)	62 (12.4%)	85 (17.0%)	3.60%	4.80%
RJ/CRB	70 (2.6%)	363 (13.6%)	0 (0.0%)	139 (27.8%)	2.80%	7.20%

TABLE 9 Statistics of violation of two models and time span on zero and over six hits.

Index	FB violations mean	HS violations mean	FB violations std	HS violations std	FB time span on 0 hit	HS time span on 0 hit	FB time span in 1–6 hits	HS time span in 1–6 hits
S&P 500	3.8	4.6	2.6	5.4	2.1%	25.9%	87.0%	53.7%
TOPIX	2.8	2.4	1.9	2.3	13.3%	23.9%	79.0%	73.1%
TOPIX LC	3.3	3.5	1.7	2.9	0.0%	13.0%	95.9%	70.1%
DAX	3.7	4.3	1.9	4.7	1.4%	40.4%	90.0%	28.7%
DAX LC	3.7	3.9	1.8	3.9	0.2%	36.5%	91.9%	40.0%
MSCI UK	4.2	4.4	1.9	4.4	0.0%	30.2%	88.4%	37.9%
MSCI UK LC	3.6	4.1	1.5	4.1	0.0%	35.8%	96.9%	30.8%
MSCI France	3.8	4.1	2.4	4.2	11.3%	36.0%	72.8%	30.2%
MSCI France LC	3.3	3.7	1.6	3.8	0.7%	36.6%	94.8%	38.6%
COMIT	3.8	3.4	2.0	4.1	2.2%	32.4%	91.1%	49.9%
COMIT LC	3.8	3.7	1.8	3.9	0.0%	26.9%	86.3%	58.7%
MSCI Canada	3.6	3.6	1.7	4.5	1.8%	21.8%	94.1%	62.7%
MSCI Canada LC	4.1	3.8	2.8	4.9	4.9%	26.0%	73.9%	54.6%
MSCI Emerging Markets	3.4	3.4	1.9	3.0	7.8%	26.9%	83.8%	58.1%
RJ/CRB	3.4	4.2	1.7	3.9	3.2%	9.9%	94.2%	76.6%

The backtesting analysis also suggests some remarks on regulatory rules. First of all, overly conservative models always stay at the boundary (or even never producing any VaR violations at all) will be judged compliant, although they are highly questionable statistically (Kupiec (1995)). In particular, a model producing no VaR violations ($\pi = 0\%$) cannot pass coverage and independence tests (see columns 5 and 6 of Table 9 on the preceding page).

In addition, regulatory rules are not concerned with the distribution of violations: it does not matter much whether a model fails with seven overshootings in a row, or with seven violations distributed over 250 days (although the model will have been compliant for most of the time). The latter feature seems much less worrying than the former, but according to the backtesting hit ratio this does not make any difference in evaluating risk models.

REFERENCES

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* **9**(3), 203–228.

Barone-Adesi, G. K., Giannopoulos, K., and Vosper, L. (1999). VaR without correlations for portfolios of derivative securities. *Journal of Futures Markets* **19**(5), 583–602.

Barone-Adesi, G. K., Giannopoulos, K., and Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management* **8**(1), 31–58.

Basel Committee on Banking Supervision (1996). Amendment to the capital accord to incorporate market risks. Bank for International Settlements (January). URL: www.bis.org/publ/bcbs24.pdf.

Brandolini, D., Pallotta, M., and Zenti, R. (2000). Risk management in an asset management company: a practical case. Working Paper.

CESR (2010). CESR's guidelines on risk measurement and the calculation of global exposure and counterparty risk for UCITS. Report, CESR/10-788.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review* **39**, 841–862.

Christoffersen, P. (2003). *Elements of Financial Risk Management*. Academic Press, San Diego, CA.

Christoffersen, P., and Pelletier, D. (2004). Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics* **2**(1), 84–108.

Christoffersen, P., Hahn, J., and Inoue, A. (2001). Testing and comparing value-at-risk measures. *Journal of Empirical Finance* **8**, 325–342.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **3**, 73–84.

Pallotta, M., and Zenti, R. (2000). Risk analysis for asset managers: historical simulation, the bootstrap approach and value at risk calculation. Working Paper.