Portfolio tail risk validation

George Dharma

Of four daily risk models, filtered historical simulation produced the fewest 99% value-at-risk breaches and the lowest error score. It still saw 64 breaches over 5,031 sessions, versus 50.3 expected. I would use it for monitoring with a separate stress limit.

The reference portfolio holds 60% in the US equity market and 40% in Treasury bills, rebalanced daily. A breach occurs when the realised loss exceeds the forecast value-at-risk (VaR). Expected shortfall (ES) estimates the average loss in the tail beyond that threshold.[1]

99% model	Breaches	Breach rate	Loss score (bp)	Joint test p
Historical	80	1.59%	2.97	<0.001
Normal	137	2.72%	3.25	<0.001
EWMA normal	118	2.35%	2.51	<0.001
Filtered historical	**64**	**1.27%**	**2.40**	**0.029**

Filtered simulation scales past shocks by the volatility prevailing when each shock occurred. It then applies the observed tail of those shocks to the current volatility estimate. That gives extreme past moves a role without forcing today’s raw volatility to match theirs.[2]

The 99% joint coverage and independence test still rejects the filtered model at the 5% level ($p = 0.029$). Its lower error score identifies the best of the four tested forecasts for this sample. The remaining clustering matters when the model influences daily trading limits.

1 Kenneth R. French, US daily factors and Treasury-bill returns

2 John Hull and Alan White, Incorporating volatility updating into historical simulation

How each day is forecast

The forecast for any session uses only earlier observations. This matters most at the start of a market shock. The first bad day must count as a forecast error even if it immediately makes tomorrow's risk estimate jump.

Model	Volatility estimate	Where the tail comes from
Historical	No separate scaling	Rank the previous 250 portfolio losses.
Normal	Standard deviation over 250 prior sessions	A normal distribution gives a 99% cut-off near 2.33 standard deviations.
EWMA normal	Exponential smoothing with 94% weight on the prior variance	A normal tail is applied to the changing volatility estimate.
Filtered historical	Same 94% exponential volatility estimate	The previous 250 standardised losses determine the tail factor.

With 250 observations, 99% ES weights the two largest losses plus half the third, for 2.5 observations. The 99% historical threshold is the 248th observation sorted from low to high. Even independent continuous losses would exceed that rank about 3 in 251 times, or 1.20%. I report the nominal 1% test and interpret the short sample alongside it.[3]

Tail severity in dollars

At a constant $10m portfolio value, the filtered model's average 99% VaR was about $193,000 and its average ES was $237,000. The worst actual day lost $719,600. Across filtered-model breaches, the average realised loss was 1.60%, against the model's average ES of 1.53% on those same dates. That shortfall measures the severity of the misses.

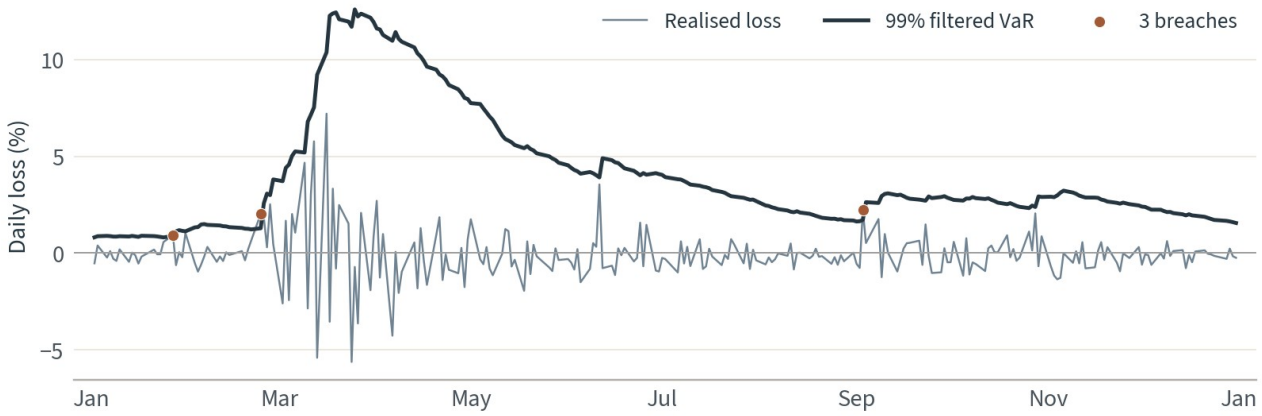
3 George Dharma, Portfolio tail risk validation, Excel workbook, Forecast tape and Model tests.

What changes across time and confidence levels

The filtered 99% model was weaker in the first decade. It recorded 37 breaches in 2006–2015 and 27 in 2016–2025. A model that passes a test in the later period still needs a way to handle conditions resembling the earlier one.

Filtered model	Period	Breaches	Breach rate	Joint test p
99.0%	2006–2015	37	1.47%	0.026
99.0%	2016–2025	27	1.07%	0.541
97.5%	2006–2015	74	2.94%	0.201
97.5%	2016–2025	70	2.78%	0.275

Daily losses and the 99% forecast during 2020



The orange dots mark the 3 sessions in 2020 when the realised loss exceeded the forecast. Each forecast uses data available before that session. Negative losses are gains.[4]

At 97.5% confidence the filtered model recorded 144 breaches over the whole sample, compared with about 126 expected. Its joint test p-value was 0.059. The EWMA normal model had a marginally lower quantile loss score at this confidence level, despite producing 196 breaches. A single score does not establish reliable coverage.[5]

For daily monitoring I would show both the filtered forecast and the largest recent realised losses. I would track exceedance counts in a rolling 250-session window and review clusters as well as single large misses. The workbook lets the reader change method, confidence level and portfolio value and inspect the days behind each result.

4 George Dharma, Portfolio tail risk validation, Excel workbook, Forecast tape.

5 George Dharma, Portfolio tail risk validation, Excel workbook, Model tests.