Forecasting VaR and CVaR based on a skewed exponential power mixture, in compliance with the new market risk regulation

Samir Saissi Hassani and Georges Dionne
Canada Research Chair in Risk Management and Finance Department, HEC Montréal

25 November 2022

Abstract

We demonstrate how a mixture of two SEP3 densities (skewed exponential power distribution of Fernández et al., 1995) can model the conditional forecasting of VaR and CVaR to efficiently cover market risk at regulatory levels of 1% and 2.5%, as well as at the additional 5% level. Our data consists of a sample of market asset returns, relating to a period of extreme market turmoil, showing typical leptokurtosis and skewness. The SEP3 mixture outcomes are benchmarked using various competing models, including the generalized Pareto distribution. Appropriate scoring functions quickly highlight valuable models, which undergo conventional backtests. As an additional backtest, we argue for and apply the CVaR part of the optimality test of Patton et al. (2019) to assess the conditional adequacy of CVaR. An additional aim of this paper is to present a collaborative framework that relies on both comparative and conventional backtesting tools, all in compliance with the recent Basel regulation for market-risk.

Keywords

Conditional forecasting, VaR, CVaR, Backtesting, Basel regulation for market risk, Heavy tailed distributions.

Highlights

- Conditional forecasting of VaR and CVaR
- Mixture of two skewed exponential power distributions (SEP3)
- Testing the conditional adequacy of CVaR
- In line with Basel, performing required and supplementary VaR and CVaR evaluations

Introduction

In 2016, the Basel Committee decided that banks' market risk capital should be calculated with the conditional value at risk (CVaR), also known as expected shortfall,[1] at the 97.5% confidence level, while maintaining model backtesting at the 99% value at risk (VaR) (BCBS, 2016, 2019). This shift toward CVaR is motivated by the inadequacy of the risk coverage calculated by VaR, which has been noted over time. Market risk is now jointly managed via CVaR and VaR, at two different probabilities: p = 2.5% and p = 1%, respectively.[2]

In recent decades, it has become increasingly important to assess financial risk accurately; selecting a suitable distribution has turned out to be a major challenge. The empirical evidence against normality is considerable (see, e.g., Haas et al., 2006). Accounting for heavy tails, autocorrelations, and volatility clustering, the stylized aspect of financial time series is more critical than ever for gauging market risk (McNeil et al., 2015).

On the one hand, for these reasons, various methods to transform the normal density, such as Gram-Charlier expansions, have been created to meet the desired features, for increased flexibility in fitting empirical distributions (see, e.g., Zoia et al., 2018; Molina-Muñoz et al., 2021). On the other hand, alternative distributions to replace Gaussian law have been studied, including stable distributions and several Student-*t* extensions (see, e.g., Lee and Lin, 2012; Iqbal et al., 2020). A great deal of effort has also been made based on exponential power (EP) distribution, which, with three parameters, can reproduce any kurtosis level. Fernández et al. (1995) introduced a four-parameter skewed EP, which will be abbreviated as SEP3 density here. Zhu and Zinde-Walsh (2009) presented a five-parameter asymmetric exponential power distribution (AEPD) to capture both the right- and left-tail thicknesses independently. Downside risk measurement and forecasting accuracy improve when the relevant tail is more adequately captured. Zhu and Galbraith (2011) and, more recently, Kim and Lee (2021) used AEPD to build a conditional heteroscedastic specification.

[1] CVaR is also called expected shortfall in the literature. Both measures are equivalent for continuous distributions without jumps (Rockafellar and Uryasev, 2002). See also Dionne (2019).

[2] In this paper, we use the letter p to refer to the probability that the VaR is exceeded, and 1-p for the corresponding confidence level. The *p*-value notation is for statistical tests.

Earlier, Rombouts and Bouaddi (2009) showed that a mixture of a skewed construction using an EP distribution worked properly to model standardized innovations in an asymmetric GARCH-type model. Indeed, a distribution mixture is an interesting alternative, offering a degree of freedom to represent the left and right tails relatively separately in terms of skewness and tail thickness. It is in this line that we link our work.

We propose to show that a mixture of two SEP3 works well for conditional forecasting with two stages, as in McNeil and Frey (2000) and as shown in Nolde and Ziegel (2017), the latter of which we refer to as NZ (2017) in the sequel. The first stage, based on an AR(1)-GARCH(1,1), will model the standardized innovations in two ways: a Gaussian and the Skew-*t* of Fernández and Steel (1998). The first stage standardized residuals will be estimated in the second stage using competing models starting with a simple normal law and moving on to a SEP3 mixture.

To measure risk, we focus on VaR, well-known and elicitable, and CVaR, a coherent and subadditive risk measure. Let *Y* represent the return on a given asset over a given horizon. Negative returns, which are subject to loss risk, are located at the left tail of the distribution of *Y*. Also, p-levels are small, close to 0. At a given p-level $\in (0,1)$, p-VaR is defined by

$$\mathrm{VaR}^{\mathrm{p}}(\mathrm{Y}) = -\inf\{\mathrm{y} \mid \mathrm{F}_{\mathrm{Y}}(\mathrm{y}) \geq \mathrm{p}\} \tag{1}$$

where F_{Y} is the cumulative distribution function of Y. p-VaR can also be defined as the negated value of *Y* p-quantile . CVaR is the negated value of the expected return *Y* conditional on *Y* being below the p-quantile of Y. Assuming F_{Y} strictly increases with finite mean, p-CVaR is defined by

$$\mathrm{CVaR}^{\mathrm{p}}(\mathrm{Y}) = -\frac{1}{\mathrm{p}} \int_0^p \mathrm{F}_{\mathrm{Y}}^{-1}(\mathrm{u})\mathrm{du}. \tag{2}$$

Due to our convention, VaR and CVaR are then typically positive; the greater the value, the riskier the situation being measured. Assuming normal returns, 2.5%-CVaR represents a risk similar to 1%-VaR.

In addition to the regulatory requirements for capital, the new Basel regulations strongly recommend supplementing statistical procedures to ensure the ex-post suitability of models (BCBS, 2016, page 82; BCBS, 2019, paragraph 32.13). For these reasons, models are evaluated through a collaborative approach using tools from both branches of backtesting: conventional (called standard) and comparative. Depending on level p = 1% or 2.5%, appropriate scoring functions guide the analysis to focus on worthwhile models, whereas models are validated individually using standard backtests and on the basis of the DM test (Diebold and Mariano, 1995).

Three standard VaR tests are applied: the uc test of Kupiec (1995), the cc of Christoffersen (1998), and the DQ of Engle and Manganelli (2004). For CVaR, we use the Z_{ES} test of Acerbi and Szekely (2019) and the RC test of Righi and Ceretta (2015). Moreover, we argue for and employ the CVaR part of the optimality test of Patton et al. (2019) to evaluate the CVaR's conditional adequacy. One aim of this paper is to orchestrate all these aspects into a validation process that complies with the regulations in force.

We utilize a sample of data from three assets: IBM, General Electric, and Walmart, whose returns all refute normality over the time period under examination, which encompasses the extreme price fluctuations of the last economic recession in the United States (NBER[3] December 2007 to June 2009) and the financial crisis of August 2007 to February 2009 (TED spread). Our competing models have a role to play in identifying and documenting specific aspects of the research. Due to their well-known performance at high quantiles, the generalized Pareto distribution (GP, McNeil et al. 2015) and the exponential generalized beta type 2 (EGB2, Kerman and McDonald, 2015) are also in the running to benchmark the findings.

We provide mathematical derivations of VaR and CVaR for our parametric models in the Online Appendices. Appendix A1 describes the model naming. Appendices A2 and A3 derive formulas to compute CVaR for distribution mixtures, while Appendices A4 to A10 derive the CVaR formula for individual parametric distributions. With the aforementioned assets, we design in Appendix A11 an optimal portfolio, which will serve as input data beginning in Section 2.

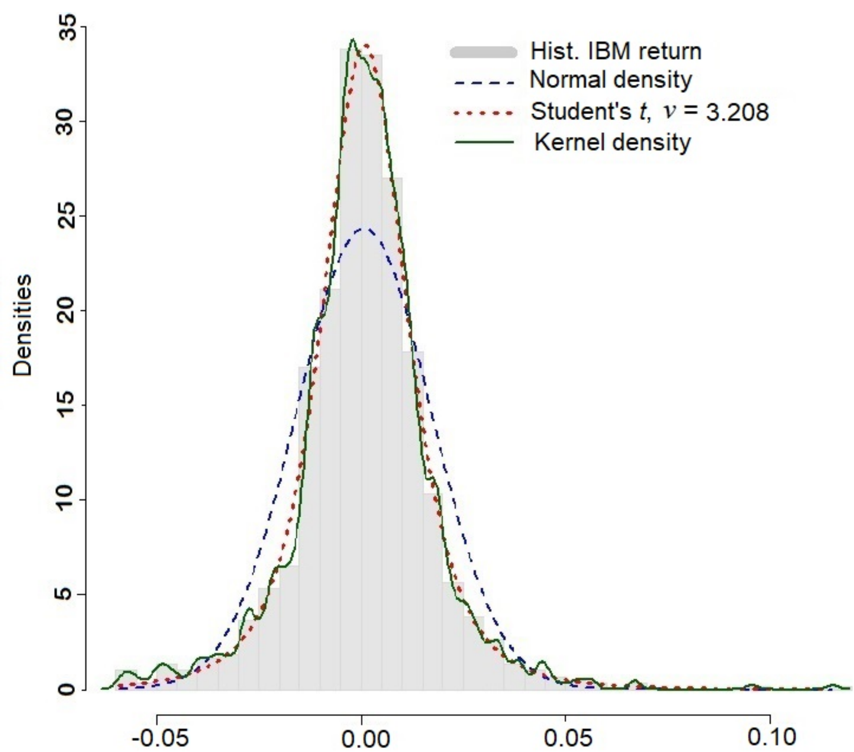
The first section, which follows, presents the data. Section 2 provides a preliminary analysis and establishes an initial in sample modeling. Section 3 presents the conditional forecasting and backtesting framework. Section 4 covers backtesting results. The final section discusses the results and concludes the paper.

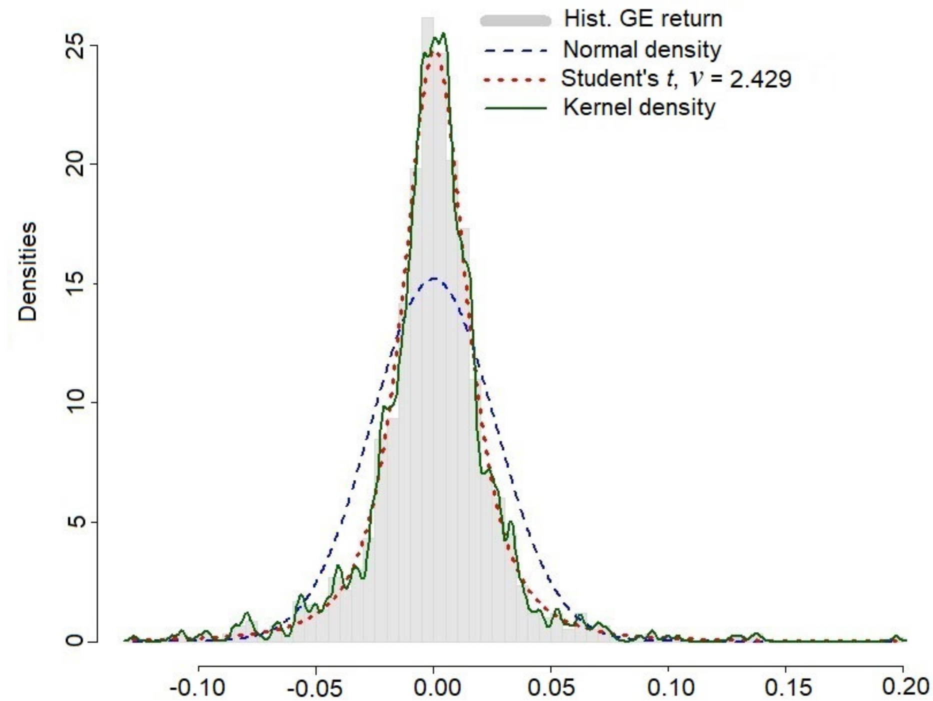
[3] https://www.nber.org/cycles.html

1. Data

Our data contains three risky stocks: IBM, General Electric, and Walmart. The period consists of 1,200 days, from June 18, 2007 to March 20, 2012. The actual daily price extraction period starts 251 days earlier, on June 16, 2006. In total, 1,451 daily prices are collected (=1,200+250+1 day), to provide 250-day rolling windows for the out-of-sample framework, plus one additional day to compute the first day's return. The return of asset *i* for time *t* is calculated using the formula: $\mathrm{Return_{i,t}} = (\mathrm{Price_{i,t}} + \mathrm{divid_{i,t}})/\mathrm{Price_{i,t-1}} - 1$ ($\mathrm{Price_{i,t}}$ and $\mathrm{divid_{i,t}}$ are the price of asset *i* and its dividend, if any, at t).

Figure 1 shows that, over the period in question, the asset return distributions are far from normal. By fitting a Student-*t* distribution to the returns, using the standard maximum likelihood approach, the estimated degrees of freedom are 3.2, 2.4, and 3.2, respectively, indicating the presence of a very fat tail.





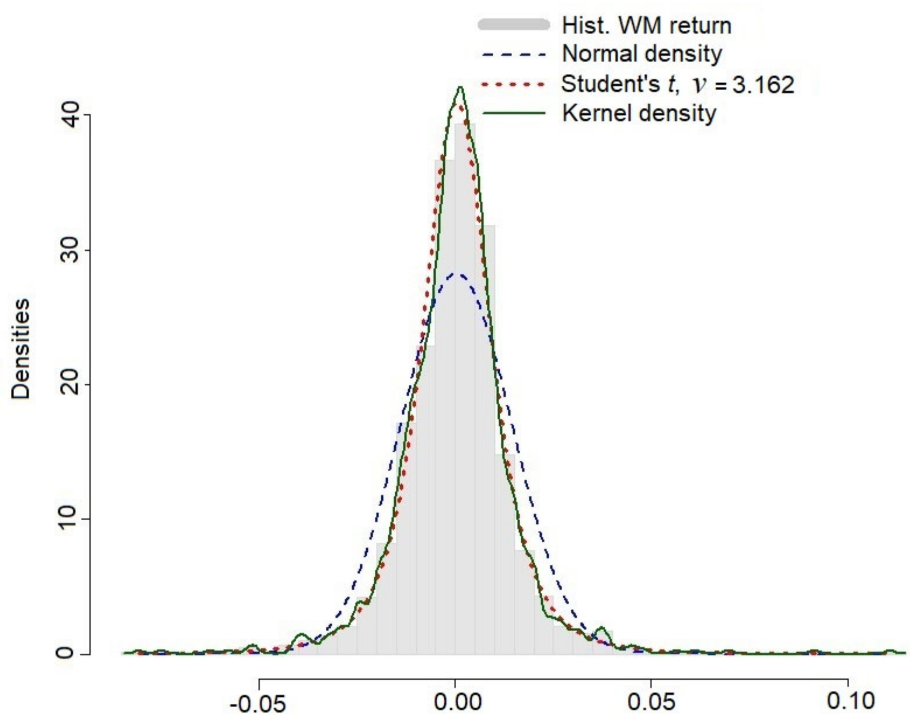


Figure 1: Histograms and densities of IBM, GE, and WM stocks

Table 1 displays descriptive statistics, including a correlation matrix. During this period of the financial crisis, the correlations are all positive and very strong, at around 50%.

Table 1: Descriptive statistics of individual assets

	IBM	General Electric	Walmart
Correlations			
IBM	1		
General Electric	0.56759	1	
Walmart	0.49146	0.43083	1
Covariances*			
IBM	0.02688	0.02443	0.01138
General Electric	0.02443	0.06894	0.01598
Walmart	0.01138	0.01598	0.01995
Mean	0.07580%	-0.00286%	0.03472%
Variance*	0.02688	0.06894	0.01995
Skewness	0.27190	0.35375	0.35429
Kurtosis	7.43415	9.95718	10.68244

*Variances and covariances are multiplied by 100 to show more decimal places.

On average, the daily returns are practically nil. The skewness coefficients are positive, which is surprising given that the extreme market turmoil should instead cause a shift toward the left tail of losses. The empirical kurtosis coefficients are very large, roughly three times higher than the known normal of 3. This confirms the presence of a large tail thickness.

Using the three assets, we construct an optimal portfolio Y_t^*. The optimal weight vector β is computed to minimize the relative VaR, defined as $\mathrm{VaR}_{\mathrm{portfolio}} + \mu_{\mathrm{portfolio}}$, assuming normal returns. The weights are $\beta = (0.38894, -0.04651, 0.65756)$. For detailed calculations, see Online Appendix A11. To simplify the notations, we will refer to this optimal portfolio as Y_t in the following. Y_t is assumed to remain optimal for all competing models, to ensure results comparability.

As demonstrated in Table 2, the portfolio skewness = 0.35887 > 0 and kurtosis = 9.81578 indicate that the same conclusion applies to the portfolio as it does to the individual assets.

Table 2 : Optimal portfolio (normal model)

	Portfolio					
	Weights	0.38894	-0.04651	0.65756		
Mean	Variance*	Std-deviation	Skewness	Kurtosis	Min	Max
0.05244%	0.01680	1.29631%	0.35887	9.81578	-7.15216%	10.54010%

*Variance is multiplied by 100 to show more decimal digits.

Figure 2 clearly shows that a Gaussian cannot fit. Student-*t* is not sharp enough and does not keep enough mass around the mode. This rather suggests a Laplace density.

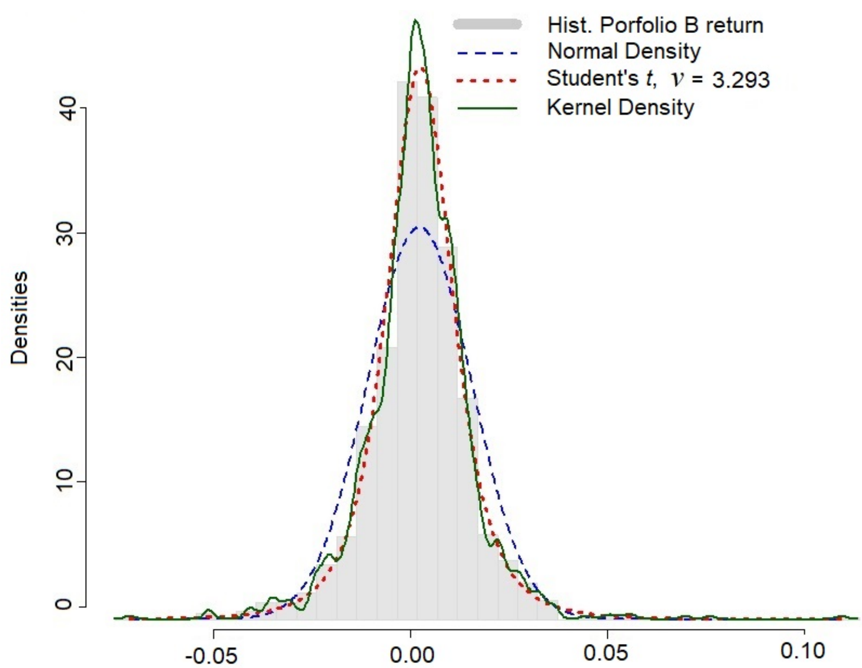


Figure 2: Histogram and densities of the optimal portfolio

2. Preliminary analysis

In this section, models M1 through M8 are directly fitted to $\{Y_t\}_{t=1}^{1200}$ using the maximum likelihood approach in static modeling. The goal is to deliver a quick overview of model traits before going deeper into them in Sections 3 and 4. The eight competing models for the data modeling are listed in Table 3.

Table 3 : Model Symbol Definitions

Model	Symbol	Description of the model
M1	1:NO	Normal distribution (see Appendix A4)
M2	1:T	Student-*t* distribution (see Appendix A5)
M3	1:EGB2	Exponential GB2 distribution (Exponential Generalized Beta 2; Appendix A6)
M4	2:NO	Mixture of 2 normal distributions (Appendices A2, A3, A4)
M5	2:T	Mixture of 2 Student-*t* distributions (Appendices A2, A3, and A5)
M6	3:NO	Mixture of 3 normal distributions (Appendices A2, A3, A4)
M7	2:SN2	Mixture of 2 SN2 distributions (Skewed Normal of Fernández et al., 1995, Appendix A7)
M8	2:SEP3	Mixture of 2 SEP3 distributions (Skewed Exponential Power type 3 of Fernández et al., 1995; Appendices A3, A8, and A9)

The results for p=5% are shown in Tables 4 and 5 alongside usual validation criteria: AIC, BIC, and the KS goodness-of-fit test. Also, model-derived kurtosis and asymmetry coefficients can be compared to empirical ones (Table 2). For mixtures, c_j identifies the weight of the j[th] individual density in the mixture.

Obviously, M1=1:NO is not suitable for data. M2=1:T, assuming a single Student-*t*, has infinite kurtosis (degree of freedom $\nu = 3.288 < 4$). M3=1:EGB2 using a single EGB2 density could capture both thickness and skewness using parameters ν and τ. Values $\nu = 0.165$ and $\tau = 0.158$ tend to 0. Therefore, EGB2 tends toward a Laplace (Lemma 2 of Caivano and Harvey, 2014). This confirms Section 1 remarks.

Mixture models M4=2:NO, M5=2:T and M6=3:NO improve AIC and attain higher kurtosis. To capture the skewness, M7=2:SN2 injects two parameters. M8=2:SEP3 adds two tail thickness parameters. AIC and BIC indicate a better fit to the data. Based on thickness and skewness parameters $\tau_1 = 0.959 \approx 1$, $\nu_1 = 1.031 \approx 1$, $\tau_2 = 2.108 \approx 2$, and $\nu_2 = 0.613 < 1$, the first SEP3 is a Laplace degenerate, whereas the second SEP3 is practically an asymmetric normal, which is an SN2. These facts confirm Section 1 findings and illustrate relationships between our models.

For further analysis, Tables 4 and 5 provide exhaustive information and Appendix A11 (Online Appendices) presents detailed discussion. Also, Table A.3 (Online Appendices) illustrates the behavior of the eight models at p = 2.5% and 1%, which is similar to p=5%.

Table 4
Model estimation – Panel A

	1:NO	1:T	1:EGB2	2:NO	2:T
μ_1	0.0005244	0.0006974**	0.0008884*	−0.0004845	0.0012920**
	(0.0003741)	(0.0002977)	(0.0004982)	(0.0015691)	(0.0005553)
σ_1	0.0129631***	0.0085310***	0.0014108**	0.0226636***	0.0066854***
	(0.0002645)	(0.0003410)	(0.0006812)	(0.0018632)	(0.0009171)
ν_1		3.2887197***	0.1587161**		23,642.31***
		(0.3809600)	(0.0796200)		(0.0000001)
τ_1			0.1652522*		
			(0.0851634)		
μ_2				0.0008151**	−0.0004740
				(0.0003448)	(0.0008931)
σ_2				0.0082545***	0.0140598***
				(0.0005136)	(0.0025828)
ν_2					6.4162601**
					(2.5707612)
τ_2					
c_1				0.2231962***	0.5158049***
				(0.0497856)	(0.1538992)
No of params	2	3	4	5	7
LogLik	3,512.55	3,627.57	3,625.58	3,619.04	3,628.64
AIC	−7,021.10	−7,249.14	−7,243.16	−7,228.09	−7,243.28
BIC	−7,010.92	−7,233.87	−7,222.80	−7,202.63	−7,207.65
KS (*p*-value)	(0.0015)	(0.1285)	(0.3490)	(0.2180)	(0.1100)
5%-VaR	2.079%	1.868%	2.006%	1.953%	2.029%
5%-CvaR	2.621%	3.012%	2.895%	3.113%	3.041%
Skewness	0	0	-0.081	-0.138	-0.154
Kurtosis	3	-	5.807	6.678	8.399

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$; *p*-values in parentheses; number of observations = 1,200

Table 5
Model estimation – Panel B

	3:NO	2:SN2	2:SEP3
μ_1	−0.0004753	-0.0173572***	−0.0007520***
	(0.0009649)	(0.0059701)	(0.0001560)
σ_1	0.0150441***	0.0235020***	0.0045291***
	(0.0022041)	(0.0020111)	(0.0014071)
ν_1		1.4398353***	1.0315089***
		(0.1888699)	(0.0376383)
τ_1			0.9598700***
			(0.1180946)
μ_2	0.0043390	-0.0001414	0.0075456**
	(0.0098212)	(0.0007373)	(0.0032033)
σ_2	0.0376531***	0.0089036***	0.0065018**
	(0.0101797)	(0.0004863)	(0.0025539)
ν_2		1.1003833***	0.6137048***
		(0.0657146)	(0.2182171)
τ_2			2.1083901*
			(1.1436395)
μ_3	0.0011752***		
	(0.0004491)		
σ_3	0.0065771***		
	(0.0008483)		
c_1	0.4433715***	0.1378343***	0.7389303***
	(0.1089861)	(0.0366570)	(0.1181466)
c_2	0.0334707		
	(0.0303812)		
No. of params	8	7	9
LogLik	3,630.06	3,619.48	3,631.96
AIC	−7,244.13	-7,224.96	−7,245.93
BIC	−7,203.41	-7,189.33	−7,200.12
KS (*p*-value)	(0.2280)	(0.1099)	(0.3040)
5%-VaR	2.038%	1.878%	1.992%
5%-CvaR	3.004%	2.968%	2.973%
Skewness	0.122	0.169	0.005
Kurtosis	9.432	7.529	7.175

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$; *p*-values in parentheses; number of observations = 1,200

3. Out-of-sample forecasting and backtesting

3.1 Conditional modeling

We are now interested in VaR and CVaR conditional forecasting given $\Im_{t-1}$ being the information set available up to $t-1$. The framework is based on an out-of-sample rolling window of 250-day. The purpose is to capture time dynamics to account for autocorrelations and conditional volatility, including volatility clustering, which are typically observed in financial series of market risk (McNeil et al., 2015). The return series Y_t given $\Im_{t-1}$ follows a conditional distribution F_t: $Y_t|\Im_{t-1} \sim F_t$. We assume F_t continuous, strictly increasing with finite mean. Conditional risk measures at p-level are defined by

$$\mathrm{VaR}_t^p(Y_t) = -F_t^{-1}(p) \quad \text{and} \quad \mathrm{CVaR}_t^p(Y_t) = -E\left[Y_t | Y_t < -\mathrm{VaR}_t^p(Y_t), \Im_{t-1}\right].$$

Following NZ (2017) with adjustment due to the left tail, assume that Y_t can be written as

$$Y_t = -\mu_t + \sigma_t Z_t \tag{3}$$

where Z_t is an i.i.d. random standardized variable independent of $\Im_{t-1}$. $\mu_t = -E[Y_t|\Im_{t-1}]$ and $\sigma_t^2 = \mathrm{var}[Y_t|\Im_{t-1}]$ are the negated conditional expectation and conditional variance of Y_t at *t* given $\Im_{t-1}$. The conditional risk measures of $\{Y_t\}$ for $t+1$ can be obtained from those of $\{Z_t\}$ using (3) and with one-step ahead predicted μ_{t+1} and σ_{t+1} (see Online Appendix A12.1 for detailed derivation), such as

$$\mathrm{VaR}_{t+1}^p(Y_{t+1}) = \mu_{t+1} + \sigma_{t+1}\mathrm{VaR}_{t+1}^p(Z_{t+1}) \text{ and } \mathrm{CVaR}_{t+1}^p(Y_{t+1}) = \mu_{t+1} + \sigma_{t+1}\mathrm{CVaR}_{t+1}^p(Z_{t+1}) \tag{4}$$

An ARMA-GARCH specification allows for such a design. We will use an $\mathrm{AR}(1)-\mathrm{GARCH}(1,1)$ to model μ_t and σ_t. The $\mathrm{AR}(1)-\mathrm{GARCH}(1,1)$ system is written as

$$y_t = \xi_0 + \xi_1 y_{t-1} + \varepsilon_t \tag{5}$$

$$\sigma_t^2 = \omega + \delta_1 \varepsilon_{t-1}^2 + \delta_2 \sigma_{t-1}^2 \tag{6}$$

$$z_t = \varepsilon_t / \sigma_t \, ; \quad Z_t \sim G(\theta) \tag{7}$$

where $(\xi_0, \xi_1, \omega, \delta_1, \delta_2, \theta)$ are the $AR(1) - GARCH(1,1)$ parameters. Once standardized, the error term ε_t produces the residuals Z_t, such as $z_t = \varepsilon_t / \sigma_t$. Z_t follows a distribution *G*, whose parameters θ ensure zero mean and unit variance. If *G* is known, $AR(1) - GARCH(1,1)$ can be estimated by the maximum likelihood method. We then derive

$$\mu_{t+1} = -E[Y_{t+1} | \Im_t] = -\xi_0 - \xi_1 y_t \tag{8}$$

$$\sigma_{t+1}^2 = \operatorname{var}\left[Y_{t+1} | \Im_t \right] = \omega + \delta_1 \left(\sigma_t z_t \right)^2 + \delta_2 \sigma_t^2 \tag{9}$$

and finally calculate risk measures related to *G*. In the real world, however, *G* is unknown. So, we proceed in two stages: In the first, identified as prefiltering, we impose a distribution on *G* (usually a normal, a Student-*t* or a Skew-*t* distribution). We then estimate $(\xi_0, \xi_1, \omega, \delta_1, \delta_2, \theta_t)$ using the 250-day rolling window $\{Y_s\}_{s=t-249}^{s=t}$. Meanwhile, the prefiltering stage suffers from a misspecification introduced by imposing an arbitrary distribution on unknown *G*.

Our competing models come into play in the second stage, which we henceforth call fitting. The models can more or less effectively reduce the impact of the aforementioned misspecification. Models fitted to $\{Z_t\}$ series produce $VaR_{t+1}^p(Z_{t+1})$ and $CVaR_{t+1}^p(Z_{t+1})$, which are converted back to $VaR_{t+1}^p(Y_{t+1})$ and $CVaR_{t+1}^p(Y_{t+1})$ by applying the forecasted μ_{t+1} and σ_{t+1} using (8) and (9) into equalities (4).

There will be two prefilterings. The first, denoted n-prefiltering, forces *G* to follow a Gaussian, whereas the second, called e-prefiltering, imposes the Skew-*t* of Fernández and Steel (1998).[4]

This Skew-*t* distribution, named ST3, has four parameters, say, μ, σ, ν, τ, and its density function is written for $x \in R$, such that

[4] Nolde and Ziegel (2017) do prefiltering in three ways: Gaussian, Student-*t*, and Skew-*t* of Fernández and Steel (1998).

$$f_{ST3}\left(x|\mu,\sigma,\nu,\tau\right)=\frac{c}{\sigma}\left\{\left[1+\frac{\nu^2u^2}{\tau}\right]^{-(\tau+1)/2}1_{y<\mu}+\left[1+\frac{u^2}{\nu^2\tau}\right]^{-(\tau+1)/2}1_{y\geq\mu}\right\} \tag{10}$$

where $u=(x-\mu)/\sigma;\quad \mu\in R;\quad \sigma,\nu,\tau>0,$ and $c=2\nu\left[\left(1+\nu^2\right)B\left(1/2,\tau/2\right)\tau^{1/2}\right]^{-1}$.

We add a prefix *n* or *e* to properly specify the conditional models. Thus, the 3:NO basic model (Section 2) will correspond to two conditional models: n|3:NO (n-prefiltered) and e|3:NO (e-prefiltered). Likewise, e|1:T is based on a single Student-*t* distribution fitted to Z_t series obtained from e-prefiltering.

Three new basic models are required. The first, $M9=1:ST3,$ (single Skew-*t* density) is related to the conditional model $e|1:ST3$, which will benchmark e-prefiltered models because it has ST3 in both stages. The predictive power of the resulting conditional model is expected to be average or poor when the same density is arbitrarily forced in both stages (see NZ, 2017). $M10=1:SEP3$ is the second basic model (single SEP3 density). The conditional model $e|1:SEP3\ (n|1:SEP3)$ will serve to identify the mixture effect on $e|2:SEP3\ (n|2:SEP3)$ performance (see discussion later).

We define a final model, M11, which estimates the well-known generalized Pareto distribution (GP) in the fitting stage. GP is frequently needed to challenge the ability of studied models to account for extreme values in the data. Standardized residuals series Z_t will be momentarily negated to fit GP. For a threshold u and $z\geq u$, the cumulative function of GP is written as

$$F_{GP}(z)=1-\left(1+\xi\frac{z-u}{\beta}\right)^{-\frac{1}{\xi}},\qquad \xi\neq 0$$

where β and ξ are the scale and shape parameters of GP. The threshold *u* must be determined. We use pragmatic method of McNeil and Frey (2000) and, like NZ (2017), we fit GP to the 30 largest values from the 250-day rolling window; u_t is the 31st largest value (30 = 12% of 250). Finally, the formulas are as follows:

$$\mathrm{VaR}_t^p(Z_t) = u_t + \frac{\beta_t}{\xi_t}\left(\left(\frac{p}{N_u/N_w}\right)^{-\xi_t} - 1\right) \quad \text{and} \quad \mathrm{CVaR}_t^p(Z_t) = \frac{\mathrm{VaR}_t^p(Z_t)}{1-\xi_t} + \frac{\beta_t - u_t\xi_t}{1-\xi_t} \tag{11}$$

where (u_t, β_t, ξ_t) parameters correspond to day *t* fitted GP; $N_u = 30$, $N_w = 250$. $\mathrm{VaR}_{t+1}(Z_{t+1})$ and $\mathrm{CVaR}_{t+1}(Z_{t+1})$ computed following (11) are converted to $\mathrm{VaR}_{t+1}(Y_{t+1})$ and $\mathrm{CVaR}_{t+1}(Y_{t+1})$ using (8) and (9) into equalities (4).

In total, we have 22 conditional models (11 for each prefiltering).

3.2. Standard backtesting

VaR testing

3.2.1. Unconditional coverage test

Unconditional coverage testing is based on the fact that counting independent p-VaR violations follows a binomial distribution. With *T* and N_1 denoting time length and number of violations, respectively, Kupiec's (1995) test compares $p^* = N_1 / T$ (failure rate) to p (theoretical rate) with a likelihood ratio test. The null hypothesis is $H_0: p^* = p$ against $H_1: p^* \neq p$. The test statistic LR_{uc} is

$$LR_{uc} = -2\log\left[\frac{(1-p)^{T-N_1}p^{N_1}}{(1-N_1/T)^{T-N_1}(N_1/T)^{N_1}}\right] \sim \chi^2_{df=1}, \tag{12}$$

which asymptotically follows a chi-square distribution with one degree of freedom.

3.2.2. Independence test

Violations of p-VaR should be independent and well dispersed in time. The occurrence of violation clustering may result in the inability to renew coverage capital in time. Christoffersen (1998) considers the violation process I_t as first-order Markovian with two states. Its transition matrix, Π, is

$$\Pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}, \tag{13}$$

where π_{ij} is the transition probability from state *i* to state *j*. The process likelihood is

$$L(\Pi) = (1-\pi_{01})^{N_{00}} \pi_{01}^{N_{01}} (1-\pi_{11})^{N_{10}} \pi_{11}^{N_{11}} \tag{14}$$

where N_{ij} is the number of observations with a state *j* following a state *i*. There is violation independence if $H_0 : \pi_{01} = \pi_{11}$. $L(\Pi)$ can be estimated empirically by $L(\hat{\Pi})$, where π_{01} and π_{11} are replaced by $\hat{\pi}_{01} = N_{01}/(N_{00} + N_{01})$ and $\hat{\pi}_{11} = N_{11}/(N_{10} + N_{11})$. Thus, under H_0, the likelihood (14) becomes

$$L(p^*) = (1-p^*)^{N_{00}} (p^*)^{N_{01}} (1-p^*)^{N_{10}} (p^*)^{N_{11}} = L(1-p^*)^{T-N_1} (p^*)^{N_1} \tag{15}$$

Note that $N_{00} + N_{10} = T - N_1$ and $N_{01} + N_{11} = N_1$. Finally, the LR_{ind} statistic is

$$LR_{ind} = -2\log\left[\frac{L(p^*)}{L(\hat{\Pi})}\right] \sim \chi^2_{df=1}. \tag{16}$$

3.2.3. Conditional coverage tests

The goal is to evaluate the adequacy of coverage, conditional on the independence of p-VaR violations. The first backtest is the Christoffersen (1998) cc. Its null hypothesis is $H_0 : \pi_{01} = \pi_{11} = p$. Thus, its statistic can be derived directly from (16), where p replaces p^*:

$$LR_{cc} = -2\log\left[\frac{L(p)}{L(\hat{\Pi})}\right] \sim \chi^2_{df=2} \tag{17}$$

Moreover, $L(p^*)$ in (15) is exactly the denominator of the LR_{uc} statistic. This implies that $LR_{cc} = LR_{uc} + LR_{ind}$. Under H_0, LR_{cc} follows a χ^2 with 2 degrees of freedom.

The second backtest is the DQ of Engle and Manganelli (2004). The null hypothesis can be rewritten as $H_0 : E_{t-1}[h_t] = 0$, where $h_t = I_t - p$. Hence, H_0 can be tested with the following linear regression model:

$$h_t = \alpha_0 + \alpha_1 h_{t-1} + \ldots + \alpha_n h_{t-n} + \beta_1 VaR^p_{t-1} + \beta_2 VaR^p_{t-2} + \ldots + \beta_n VaR^p_{t-n} + u_t,$$

where all the coefficients, $\alpha_0, \alpha_k, \beta_k, k=1,...,n$, must be simultaneously nil. This can be verified using Wald test, whose statistic follows a χ^2 with $2n+1$ degrees of freedom.

CVaR testing

3.2.4. Z_{ES} test of Acerbi and Szekely (2019)

The backtest of Acerbi and Szekely (2019) exists in two versions: absolute or relative. We will consider the absolute version, whose statistic is

$$Z_{ES}(Y_t) = \frac{1}{T}\sum_{t=1}^{T}\left[\mathrm{CVaR}_t^p - \mathrm{VaR}_t^p + \frac{(Y_t + \mathrm{VaR}_t^p)}{p} I_t\right] \tag{18}$$

where $I_t = 1_{(Y_t < -\mathrm{VaR}_t^p)}$ (indicator function).

The null hypothesis is $H_0 : E\left[Z_{ES}(Y_t)\right] = 0$ against $H_1 : E\left[Z_{ES}(Y_t)\right] \neq 0$. To evaluate H_0, we draw N random vectors, $\left\{X_t^j\right\}_{t=1,...,T}^{j=1,...,N}$ depending on model parameters relating to each day *t*. Applying (18) to $\left\{X_t^j\right\}$ vectors generates the series $\left\{Z_{ES}\left(X_t^j\right)\right\}^{j=1,..,N}$. The two-sided *p*-value is then determined by

$$2 \times \min\left\{P\left(Z_{ES}\left(X_t^j\right) < Z_{ES}(Y_t)\right), P\left(Z_{ES}\left(X_t^j\right) > Z_{ES}(Y_t)\right)\right\}.$$

3.2.5. RC test of Righi and Ceretta (2015)

The second backtest, denoted RC, is proposed in Righi and Ceretta (2015). Its statistic is defined by the expression

$$\mathrm{RC}(Y_t) = \frac{1}{T}\sum_{t=1}^{T}\frac{(Y_t - \mathrm{CVaR}_t^p)}{\mathrm{SD}_{p,t}} I_t, \tag{19}$$

where $I_t = 1_{(Y_t < -\mathrm{VaR}_t^p)}$ (indicator function) and $\mathrm{SD}_{p,t} = \sqrt{\mathrm{variance}[Y_t \times I_t]}$. The test *p*-value is obtained by a construction similar to Z_{ES}.

3.3. Comparative backtesting

The comparative backtesting is based on scoring functions. A scoring function assigns an asymmetric penalty depending on the deviation between the realized return and its corresponding ex-ante risk coverage. Being a penalty, it is thus meant to be minimized.

3.3.1. Scoring functions for VaR

Let y_t be the realized return and v_t a forecast of the p-VaR of day *t*. Gneiting and Raftery (2007, eq. (40)) show a general form of strictly consistent scoring functions for VaR:

$$S_v(v_t, y_t) = (I_t - p) \times G(-v_t) - I_t \times G(y_t) + \eta(y_t), \tag{20}$$

where $I_t = 1_{(y_t < -v_t)}$ and G and η are functions. G is continuously differentiable and strictly increasing. The functions η and G must allow $E[S_v(.,.)] < \infty$ and v_t is supposed to be $\mathfrak{I}_{t-1}$-mesurable (up to t-1 only) . We then have

$$\mathrm{VaR}_t^p = \arg\min_{v_t} E[S_v(v_t, y_t) | \mathfrak{I}_{t-1}].$$

The value $S_v(v_t, y_t)$ is the score of day *t*. The expectation $E[S_v(v_t, y_t)]$ is the model score. We fix $G(x) = -\log(-x)$ and $\eta(x) = -1_{(x<-1)} \times \log(-x)$ (see Online Appendix A12.2 for detailed discussion and derivations). The scoring function $S_{\log}$ thus obtained has the expression

$$S_{\log}(v_t, y_t) = \begin{cases} (p-1)\log(v_t) + \log(-y_t)\,, & \text{if } y_t < -v_t \\ p\log(v_t)\,, & \text{if } y_t \geq -v_t \end{cases} \tag{21}$$

$S_{\log}$ is now defined for $\forall\, y_t \in R$ and $v_t > 0$, which is the case for all p-VaR forecasts.

The choice of $S_{\log}$ is motivated in NZ (2017, eq.(2.20)). Indeed, $S_{\log}$ allows loss differences between competing forecasts to be 0-homogenous (homogeneous of degree 0). This property is essential to have optimal power of the DM test (Diebold and Mariano, 1995).

3.3.2. Scoring functions for CVaR

Let y_t, v_t, and e_t be, respectively, the realized return and the values of p-VaR and p-CVaR forecasts for day *t*. Fissler et al. (2016) propose a general form of strictly consistent scoring functions for CVaR:

$$S_{v,e}(v_t, e_t, y_t) = (I_t - p)\big(G_1(-v_t) - G_1(y_t)\big) - \frac{1}{p} G_2(-e_t) I_t (v_t + y_t) \\ + G_2(-e_t)(-e_t + v_t) - g_2(-e_t) + \eta(y_t) \tag{22}$$

where $I_t = 1_{(y_t < -v_t)}$. η is a function and $E[\eta(Y_t)]$ must exist. G_1 is an increasing function[5] twice continuously differentiable and $E[G_1(Y_t)]$ must exist. G_2 is the derivative of $g_2\left(g_2^{'} = G_2\right)$, with g_2 strictly increasing and strictly convex. v_t, e_t must be $\mathfrak{I}_{t-1}$-mesurable (up to t-1). We then have, for optimal $VaR/CVaR$:

$$\left(VaR_t^p, CVaR_t^p\right) = \arg\min_{(v_t, e_t)} E\left[S_{v,e}\left(v_t, e_t, y_t\right) \mid \mathfrak{I}_{t-1}\right].$$

We choose $G_1(x) = 0$, $G_2(x) = -1/x \ (x \neq 0)$ and $\eta(x) = 0$. Therefore, $g_2(x) = -\log(-x)$. The scoring function S_{FZ} thus obtained is

$$S_{FZ}(v_t, e_t, y_t) = \frac{1}{e_t}\left(v_t - \frac{I_t}{p}(y_t + v_t)\right) + \log(e_t) - 1. \tag{23}$$

S_{FZ} is defined for $\forall\, y_t \in R$ and $e_t > 0$, which is the case for all p-CVaR forecasts. S_{FZ} is chosen because it allows loss differences between competing forecasts to be 0-homogeneous (NZ, 2017, eq.(2.24), and Patton et al., 2019, eq.(6)).

The variable v_t is present in (23) and (22) and can't be eliminated. More generally, Gneiting (2011) shows that there is no strictly consistent scoring function where e_t appears alone without v_t. This is why CVaR is not elicitable by itself, but the pair $VaR/CVaR$ is.

3.3.3. DES test (goodness-of-fit test of Patton et al., 2019)

From equation (23) defining S_{FZ}, Patton et al. (2019) show that, under the assumption of a correct model specification for VaR and CVaR, the variables $\lambda_{v,t}^s = I_t - p$ and $\lambda_{e,t}^s = I_t y_t/(p e_t) - 1$ are conditionally mean zero[6], such that

[5] This is a sufficient condition on G_1. If not satisfied, Theorem 5.2 in Fissler and Ziegel (2016) shows that the function defined by $f(v_t) = -v_t G_2(-e) + G_1(-v_t)$ must be strictly increasing with respect to v_t.

[6] To understand intuitively this result, consider the first-order condition derivatives of $E[S_{FZ}(v_t, e_t, y_t)|\mathfrak{I}_{t-1}]$ with respect to v_t and e_t.

$$E\left[\lambda_{v,t}^{s}|\mathfrak{I}_{t-1}\right] = E[I_t|\mathfrak{I}_{t-1}] - p = 0 \tag{24}$$

$$E\left[\lambda_{e,t}^{s}|\mathfrak{I}_{t-1}\right] = E\left[\frac{I_t y_t}{p\times e_t}|\mathfrak{I}_{t-1}\right] - 1 = 0. \tag{25}$$

Equality (24) is, interestingly, the same as for the DQ null hypothesis. So, to test (25), Patton et al. (2019) employ the same technique as Engle and Manganelli (2004) and rely on the linear regression:

$$\lambda_{e,t}^{s} = b_0 + b_1\lambda_{e,t-1}^{s} + b_2 e_t + u_t, \tag{26}$$

where u_t is an error term, and (b_0, b_1, b_2) are the regression coefficients, which must be simultaneously zero under the null hypothesis (against the two-sided alternative). Compared to the Patton et al. (2019) approach, these authors state on page 406 that "similar conditional calibration tests are presented in Nolde and Ziegel (2017)."

Based on the aforementioned facts, we now know that DQ does more than evaluate conditional coverage: it also assesses the models' ability to produce an optimal VaR while minimizing S_{log}. Using the preceding arguments, and the parallel analogy between VaR and CVaR, we assert that the CVaR part of the Patton et al. (2019) construction (eq.(25)) is valid for testing the CVaR conditional adequacy. We refer to it as the DES test henceforth. This illustrates the key advantages of exploiting the close relationship between traditional and comparative backtesting techniques and employing them concurrently.

3.3.4. DM test of Diebold and Mariano (1995)

The purpose of the test is to compare forecasts between two models, M_i and M_j, using an appropriate scoring function, *S*. We construct the loss-difference vector $d_{ij} = \left\{S(M_i) - S(M_j)\right\}_{t=1}^{T}$. The null hypothesis is $H_0 : E\left[S(M_i) - S(M_j)\right] = 0$. The DM statistic is the mean of d_{ij}, normalized by its standard deviation, estimated with a heteroskedasticity and autocorrelation consistent (HAC) correction. We apply the Harvey et al. (1997) correction with a lag = 5 to account for the strong autocorrelation persistence. Under H_0, the DM statistic follows a Student-*t* of $T-1$ degrees of freedom.

3.4. Approach to conducting backtesting

Let us recall that our primary goal is to orchestrate all the procedures and actions to specifically meet the Basel regulations' requirements (BCBS, 2016, p.77; BCBS, 2019, §32.5). We must also fulfill the Basel recommendation to foresee additional complementary procedures and statistical tests, with varying degrees of confidence, to support the model's accuracy and robustness (BCBS, 2016, p.82; BCBS, 2019, §32.13).

The components involved are as follows: 6 backtests $(uc, cc, DQ, Z_{ES}, RC, DES)$, 3 p-levels $(1\%,\ 2.5\%,\ 5\%)$, and 2 scoring functions (S_{log} and S_{FZ}), for a total of 22 competing models. The calculations yield 22×6 backtests $\times$ 3 p-levels $= 396$ backtests to investigate, and $22 \times 2 \times 3 = 132$ scores to analyze pairwise using $2 \times 3 = 6$ DM test matrices. These are symmetrical non-diagonal matrices, each of which has $22 \times (22-1)/2 = 231$ pairs of statistics/*p*-values. Therefore, $6 \times 231 = 1{,}386$ DM statistics should be examined for signs and significance. To summarize, $396 + 1386 = 1{,}782$ different statistical tests would have to be conducted and reviewed!

We are obviously not going to do that. To achieve the backtesting process much more efficiently, we propose the following collaborative approach. The study begins with p=1%, which corresponds to Basel's explicit requirement of backtesting on VaR. The guiding idea is to select from the top-10 competing models on the basis of their ranking by S_{log}. Each selected model first undergoes an evaluation via the six backtests. The forecasts of successful models are then compared in pairs by DM matrices.

The resulting subset of models undergoes a similar circuit for $p = 2.5\%$, the regulatory coverage by CVaR level. Finally, the remaining models are assessed for $p = 5\%$, as an additional complement. The number of models thus diminishes progressively. Therefore, scoring functions concentrate on worthwhile models, whereas conventional backtests perform the final individual validation along with the DM test.

4. Backtesting results

We begin by focusing on 1%-VaR backtest. For this purpose, our criteria (1a) demand that all three VaR tests, uc, cc, and DQ, be carried out with no rejection at the 10%-threshold. We also concentrate on CVaR, as per Basel recommendations, for additional tests. We deploy three tests: Z_{ES}, RC, and the DES test. To meet criteria (1b), at least two of the three tests must not reject the model at the 5%-threshold. The supplement (1b) ensures that enough tail thickness is captured.

Table 6 displays the results. Columns (1) to (4) show the three VaR tests; columns (5) and (6), the Z_{ES} and RC standard tests; column (11) presents DES results on the conditional adequacy of CVaR. Scores and ranks of S_{log} are in columns (7) and (8). Columns (9) and (10) contain the corresponding elements for S_{FZ}. Rankings are based on the entire 22 models, including both prefilterings.

For p=1%, Panel A presents the 12 highest-rated models using S_{log}. Two of them, however, are problematic. The first is e|1:T, which, despite ranking fourth out of 22, is nonetheless strongly rejected by uc, contradicting the VaR scoring ranking. Furthermore, cc and DQ reject this model (respective *p*-values as low as 0.005, 0.012, and 0.001). Additionally, all CVaR tests reject e|1:T with *p*-values of 0.001, 0.006, and 0.015.

This example shows that a favorable ranking by a consistent scoring function does not prevent adverse results in conventional backtests. On the other hand, it is interesting to note that six models from Panel A have the exact same number of violations of 1%-VaR, 17, and therefore, identical uc results. So, while uc results alone cannot be used to differentiate among them, the VaR scoring function can. This combination of observations highlights not only the complementarity of conventional and comparative backtesting components, but also the need to integrate the capabilities they each offer.

We have a second case of rejection. Model e|1:ST3, 7th out of 22, fails uc as well as DQ, Z_{ES}, and RC (respective *p*-values of 0.061, 0.054 < 10%; 0.001 and 0.026 < 5%). In fact, e|1:ST3 is expected to deliver average to poor results, since it has an ST3 density arbitrarily imposed on both stages. Meanwhile, being ranked 7th, e|1:ST3 is therefore believed to be equivalent to or better than almost three-quarters of the 22 models. This provides a first indication of how inefficient the n-prefiltering stage might be. At the same time, it also shows how difficult it is to model the 1%-level properly.

The point to note now is that n|2:SEP3 is the only model resulting from Gaussian prefiltering. This shows, yet again, how weak this n-prefiltering is. Interestingly, despite the problematic n-prefiltering, the n|2:SEP3 model ranked 5th and 3rd overall for VaR and CVaR, respectively (columns (8) and (11)), revealing one aspect of the superiority of the 2:SEP3 mixture.

Let us now highlight that e|2:SEP3 takes first place in Panel A and passes all six tests with very comfortable *p*-values. In addition to e|2:SEP3, n|2:SEP3 is the only model in Panel A that meets criteria (1a) and surpasses (1b), since it passes all CVaR tests. The last four models of Panel A are removed for failing two or more CVaR tests. We are left with six potential models: e|2:SEP3, e|1:EGB2, e|1:SEP3, n|2:SEP3, e|1:GP, and e|2:SN2. We need the DM test to discern their loss differences. The e|2:SEP3 score of -3.2707 clearly stands out from the others, which are all rather close to each other.

The DM statistics of Table 7 are calculated by subtracting the forecasts of the row models from those of the column models. Thus, a negative statistic favors the line model, and vice versa. The improvement is confirmed when the statistic is significant at 5%. The e|2:SEP3 model shows all negative and significant statistics at 5% or better. The predictive power of this model is much higher than the others. The remaining five models, according to DM test, cannot be discriminated at 5%. That said, keep in mind that n|2SEP3 has an indisputable advantage over the other four models in meeting criteria (1a) and exceeding (1b).

Panel B of Table 6 displays the six models selected from Panel A, now sorted according to S_{FZ} scores (column (9)). Recall that this is the level of regulatory coverage calculation with 2.5%-CVaR. So, the CVaR criteria should be tighter. Criteria (2a) require no rejection by any of Z_{ES}, RC, or DES at the 10%-threshold. Criteria (2b) for VaR also require no rejection by any of the three VaR tests at 10%, which can be quite stringent due to the joint elicitability of VaR and CVaR. A proper VaR assessment is essential for a successful CVaR validation.

Once again, e|2:SEP3 takes first place. This model easily passes all conditions (2a) and (2b) with comfortable *p*-values. Model e|2:SN2 also succeeds in all criteria, with decent *p*-values (see also the SN2's performance in Dionne and Saissi Hassani, 2017). Note that e|1:GP is ranked third, but fails (2a) criteria (RC *p*-value = 0.089 < 10%). Yet again, an excellent scoring does not always imply success in backtests. The next model to consider is n|2:SEP3, which is ranked 4th and almost as easily meets (2a) and (2b) criteria as e|2:SEP3.

Notice how close the S_{FZ} scores for e|2:SN2 and e|1:GP are. Given the excellent performance of the first and the rejection of the latter, there is inconsistency. We want to see if the DM test can detect it. We also need confirmation that n|2:SEP3 can be considered the third choice in Panel 2.5% (after removing e|1GP).

The pairwise DM test matrix of the six Panel B models is shown in Table 8. There are two key facts to point out. The first is that there is not enough evidence to identify the defaulter, e|1:GP, from the others: e|2:SEP3, e|2:SN2, and n|2:SEP3. In fact, e|1:GP would even have been considered satisfactory as Panel B's third choice if we hadn't done the backtests. The complementary nature of backtesting tools proves quite useful once again. The second point is that, despite small advantages between models e|2:SEP3, e|2:SN2, and n|2:SEP3, there is no evidence to differentiate them (DM statistics are negative but not significant). As a result, the three models that emerge through the criteria of Panels A and B are e|2:SEP3, n|2:SEP3, and e|2:SN2.

On the other hand, e|1:EGB2 and e|1:SEP3 are rejected by both RC and DES. Incidentally, the single-density models, e|1:EGB2, e|1:SEP3, and e|1:GP, appear to behave similarly at the 2.5% and 1% levels, all passed p=1% decently, but are now rejected for the 2.5% level.

In Panel C of Table 6, we assess our three models' adequacy for the traditional p = 5% level. The criteria we set here are the same as before, namely, (3a), no rejection by any CVaR tests at 10%, and (3b), no rejection by any VaR tests at 10%, either. Everything is fine for e|2:SEP3 and e|2:SN2. However, n|2:SEP3 does show a weakness on both VaR and CVaR (DQ *p*-value = 0.023; DES *p*-value = 0.060). This could be due to residual autocorrelations or volatility clustering left by the n-prefilter. Despite this, given its strong performance at the regulatory and extremely difficult levels, we do not believe this to be a problem for n|2:SEP3.

Finally, to provide a further analysis of individual model evaluations not discussed here, tables A.4 and A.5 of the Online appendix A13 contain complete information about all backtesting outcomes for n-prefiltering and e-prefiltering, respectively. For instance, we can see in Table A.4, regarding n-prefiltered models, a significant decline in the CVaR sufficiency, as evidenced by the three CVaR tests, compared to e-prefiltered models in Table A.5. In addition, DQ and DES reveal serious difficulties for n-prefiltered models to deal with residual autocorrelations and volatility clustering that the Gaussian prefilter fails to account for.

Discussion and conclusion

We can draw some conclusions from our examination of the facts outlined above. First and foremost, the superiority of the SEP3 mixture has been demonstrated for our extremely difficult data in VaR and CVaR conditional forecasting at the regulatory levels of 1% and 2.5%, as well as the complementary 5% level. With an effective Skew-*t* prefilter, e|2:SEP3 behaved strongly. It is far more interesting, however, to learn that n|2:SEP3 even overcame the poor Gaussian prefiltering, whereas no other n-prefiltered model even appeared in the top-10 regulatory 1% panel. The lower efficacy of the single-density e|1:SEP3, as compared to e|2:SEP3, again illustrated the strength of the 2:SEP3 mixture in conditional modeling. Our findings support and expand those of Haas (2009), Broda and Paolella (2011), Miao et al. (2016), Rombouts and Bouaddi (2009), Zhu and Galbraith (2011), and Kim and Lee (2021).

When we compare GP to the 2:SEP3 mixture, we see that, unlike n|2:SEP3, the Gaussian prefiltered n|1:GP ranked rather near the end: 15th in the 1% panel (see Table A.4). On the other hand, the e-prefiltered e|1:GP was not well ranked in Panel A (6th), well behind e|2:SEP3 (1st) as well as behind n|2:SEP3 (5th). Furthermore, in the 2.5% panel, even e|1:GP was rejected. Finally, GP had no chance against the 2:SEP3 mixture.

A possible shortcoming of this work is that we could also have done the calculations according to a third prefilter built on Student-*t*. The results would have been more complete, but this would have overloaded the paper. Also, a second point we wanted to look into was which models could fill in the substantial gaps left by poor Gaussian prefiltering. n|2:SEP3 was the only one.

Two-density mixtures (made up of two densities) appear to be an optimal combination. They can improve risk modeling by more precisely capturing the properties of the relevant tail for downside risk. Each density can represent the tail fatness and skewness of each side of the data distribution separately. To approach this fact differently, despite being Skew-*t* prefiltered, no single-density models were retained in the final selection, including the well-known GP and EGB2—not even the single-SEP3.

The next point to consider is how crucial it is to model both the prefiltering and fitting stages properly. While the inefficiency of n-prefiltering is obvious, e-prefiltering might still have left some residual thickness and skewness unaccounted for. Interestingly, we can see this in the severe rejection of e|3:NO and e|2:NO by two CVaR tests, while models like e|1:GP were able to capture residual tail thickness. The presence of residual skewness in e-prefiltering can be proved, since e|2:SN2 managed to capture it using its two supplementary parameters, whereas neither e|2:NO nor e|3:NO could. As a result, even a successful prefiltering, such as the one built with Skew-*t*, could still be incomplete by itself. The quality of the fitting stage is unquestionably equally critical.

We applied what we called a collaborative approach to enhance model-evaluation efficiency, using mixed backtesting tools. The effort can concentrate on relevant and worthwhile competitors pinpointed among numerous models by appropriate scoring functions. On the other hand, a prominent example to illustrate the need to combine the two branches of backtesting is e|1:T, which was severely rejected by all standard backtests for p=1%, while it was rather well rated by $S_{\log}$. This was also the case, albeit less intensely, for e|1:ST3 at 1% and e|1:GP at the 2.5% level.

Regarding the abovementioned issue of models being rejected by backtests despite having excellent scoring, the exact opposite situation was stated in NZ(2017). Indeed, their study revealed models that rated poorly but that were not rejected by traditional backtesting, while they should have been (data simulated with misspecification). They advise checking models that pass traditional backtests alongside an appropriate scoring function. We support this proposition while also advocating for the opposite outcome. Excellent scoring alone, or passing conventional backtests alone, should no longer be considered enough to approve a model.

By analogy with DQ, we employed DES to backtest the conditional adequacy of CVaR, whereas its original purpose was to assess model forecasting optimality. The mechanism is straightforward to document. Low-rated models would not minimize the scoring function. First-order conditions cannot be met. Thus, DQ or DES (or equivalent conditional tests) should disqualify them. This further illustrates how interesting it can be to take advantage of the close relationship between both backtesting branches. Backtesting can thus achieve significant theoretical and practical advancements.

In addition to fully supporting the concurrent use of all backtesting approaches, we believe that regulators will soon consider incorporating the best of them into a future standardized "minimum backtesting requirement" procedure.

Table 6
Backtesting results

Models	p	%	Standard Backtests						--- Comparative Backtests ---				
			uc		cc	DQ	Z_{ES}	RC	S_{log}		S_{FZ}		DES
	%	Viol.	Stat.	p-val	p-val	p-val	p-val	p-val	Scores	#R	Scores	#R	p-val
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A:													
e|2:SEP3	1	0.92	0.0001	0.769	0.865	0.802	0.579	0.979	-3.2707	1	-3.2513	1	0.701
e|1:EGB2	1	1.42	0.0016	0.172	0.308	0.183	0.092	0.019	-3.2178	2	-3.1857	2	0.185
e|1:SEP3	1	1.50	0.0022	0.105	0.204	0.098	0.170	0.022	-3.2153	3	-3.1720	4	0.126
e|1:T	1	1.92	0.0067	0.005	0.012	0.001	0.001	0.006	-3.2116	4	-3.1407	12	0.015
n|2:SEP3	1	1.33	0.0010	0.269	0.438	0.186	0.076	0.051	-3.2112	5	-3.1846	3	0.219
e|1:GP	1	1.42	0.0016	0.172	0.308	0.160	0.094	0.014	-3.2094	6	-3.1677	7	0.143
e|1:ST3	1	1.58	0.0029	0.061	0.128	0.054	0.001	0.026	-3.2089	7	-3.1416	11	0.078
e|2:SN2	1	1.42	0.0016	0.172	0.308	0.385	0.054	0.006	-3.2069	8	-3.1696	5	0.260
e|2:T	1	1.42	0.0016	0.172	0.308	0.135	0.016	0.009	-3.2030	9	-3.1691	6	0.156
e|3:NO	1	1.42	0.0016	0.172	0.308	0.128	0.034	0.008	-3.2027	10	-3.1584	8	0.156
e|2:NO	1	1.42	0.0016	0.172	0.308	0.123	0.038	0.003	-3.1934	11	-3.1571	9	0.126
e|1:NO	1	2.17	0.0103	0.000	0.001	0.000	0.017	0.000	-3.1853	12	-3.1110	18	0.007
Panel B:													
e|2:SEP3	2.5	2.50	0.0000	1.000	0.463	0.373	0.747	0.809	-8.7904	1	-3.4860	1	0.320
e|2:SN2	2.5	3.00	0.0010	0.282	0.559	0.438	0.148	0.120	-8.7860	3	-3.4705	2	0.260
e|1:GP	2.5	3.25	0.0021	0.111	0.273	0.252	0.217	0.089	-8.7893	2	-3.4702	3	0.168
n|2:SEP3	2.5	2.92	0.0007	0.368	0.666	0.227	0.301	0.408	-8.7720	4	-3.4680	4	0.185
e|1:EGB2	2.5	3.25	0.0021	0.111	0.273	0.097	0.104	0.082	-8.7565	7	-3.4628	6	0.097
e|1:SEP3	2.5	3.25	0.0021	0.111	0.273	0.081	0.202	0.041	-8.7542	8	-3.4579	9	0.079
Panel C:													
e|2:SEP3	5	5.00	0.0000	1.000	1.000	0.432	0.406	0.550	-18.5713	2	-3.6757	1	0.300
n|2:SEP3	5	5.92	0.0017	0.156	0.076	0.023	0.280	0.422	-18.4889	16	-3.6548	12	0.060
e|2:SN2	5	5.92	0.0017	0.156	0.150	0.107	0.140	0.163	-18.5322	7	-3.6626	2	0.121

Scores of S_{log} multiplied by 100 to show more digits; Panel A: Models sorted by S_{log} (1%-VaR)
Panel B: Selected models, sorted by S_{FZ} (2.5%-CVaR); Panel C: Selected models only. #R columns show model ranking according to their respective scoring functions.

Table 7
DM test matrix for Panel A (S_{log} scoring 1%-VaR)

			--------------------------- Competing Models				-------------------------	
Models	Scores	#R	e|2:SEP3 Stat (*p*-value)	e|1:EGB2 Stat (*p*-value)	e|l:SEP3 Stat (*p*-value)	n|2:SEP3 Stat (*p*-value)	e|l:GP Stat (*p*-value)	e|2:SN2 Stat (*p*-value)
e|2:SEP3	−3.2707	1		**−2.428** ** (0.015)	**−2.330** ** (0.020)	**−3.060** *** (0.002)	**−2.480** ** (0.013)	**−2.609** *** (0.009)
e|1:EGB2	-3.2178	2	**2.428** ** (0.015)		−0.562 (0.574)	−0.835 (0.404)	−1.334 (0.182)	−1.812 (0.070)
e|1:SEP3	-3.2153	3	**2.3302** ** (0.020)	0.562 (0.574)		−0.401 (0.688)	−0.719 (0.472)	−1.026 (0.305)
n|2:SEP3	-3.2112	4	**3.060** *** (0.002)	0.835 (0.404)	0.401 (0.688)		−0.208 (0.835)	−0.464 (0.642)
e|1:GP	-3.2094	5	**2.480** ** (0.013)	1.334 (0.182)	0.719 (0.472)	0.208 (0.835)		−0.624 (0.533)
e|2:SN2	-3.2069	6	**2.609** *** (0.009)	1.812 (0.070)	1.026 (0.305)	0.464 (0.642)	0.624 (0.533)	

***p < 0.01, **p < 0.05; statistics highlighted in bold are significant at least at 5%; *p*-values in parentheses below the statistics

Table 8
DM test matrix for Panel B (S_{FZ} scoring 2.5%-VaR and CVaR)

			------------------------ Competing Models				--------------------------	
Models	Scores	#R	e|2:SEP3 Stat (*p*-value)	e|2:SN2 Stat (*p*-value)	e|l:GP Stat (*p*-value)	n|2:SEP3 Stat (*p*-value)	e|l:EGB2 Stat (*p*-value)	e|l:SEP3 Stat (*p*-value)
e|2:SEP3	-3.4860	1		-1.311 (0.190)	-1.279 (0.201)	-1.755 (0.079)	**-2.013** ** (0.044)	**-2.318** ** (0.021)
e|2:SN2	-3.4705	2	1.311 (0.190)		-0.086 (0.931)	-0.477 (0.633)	**2.111** ** (0.035)	**-2.776** *** (0.006)
e|l:GP	-3.4702	3	1.279 (0.201)	0.086 (0.931)		-0.413 (0.679)	-1.732 (0.084)	**-2.497** ** (0.013)
n|2:SEP3	-3.4680	4	1.755 (0.079)	0.477 (0.633)	0.413 (0.679)		-0.961 (0.337)	-1.628 (0.104)
e|l:EGB2	-3.4628	5	**2.013** ** (0.044)	**2.111** ** (0.035)	1.732 (0.084)	0.961 (0.337)		**-2.697** *** (0.007)
e|l:SEP3	-3.4579	6	**2.318** ** (0.021)	**2.776** *** (0.006)	**2.497** ** (0.013)	1.628 (0.104)	**2.697** *** (0.007)	

***p < 0.01, **p < 0.05; statistics highlighted in bold are significant at least at 5%; *p*-values in parentheses below the statistics

Acknowledgements and declaration of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper. They thank Claire Boisvert for her collaboration in the preparation of the manuscript.

References

Acerbi, C. and Szekely, B. (2019). The minimally biased backtest for es. *Risk net* **29**: 1-6.

Basel Committee on Banking Supervision (BCBS) (2016). Minimum capital requirements for market risk, publication no 352. *Bank For International Settlements (BIS)*:1-92.

Basel Committee on Banking Supervision (BCBS) (2019). Minimum capital requirements for market risk, publication no 457. *Bank For International Settlements (BIS)*:1-136.

Broda, S.A. and Paolella, M.S. (2011). Expected shortfall for distributions in finance. In: *Statistical Tools for Finance and Insurance*, p. 57-99. Springer. https://doi.org/10.1007/978-3-642-18062-0_2

Caivano, M. and Harvey, A. (2014). Time-series models with an EGB2 conditional distribution. *Journal of Time Series Analysis* **35**(6):558-571. https://doi.org/10.1007/978-3-642-18062-0_2

Christoffersen, P.F. (1998). Evaluating interval forecasts. *International Economic Review* **39**(4):841-862. https://doi.org/10.2307/2527341

Cummins, J.D., Dionne, G., McDonald, J.B. and Pritchett, B.M. (1990). Applications of the GB2 family of distributions in modeling insurance loss processes. *Insurance: Mathematics and Economics* **9**(4):257-272. https://doi.org/10.2307/2527341

Diebold, F.X. and Mariano, R.S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**(3):253-263. https://doi.org/10.1080/07350015.1995.10524599

Dionne, G. (2019). Corporate risk management: Theories and applications. John Wiley, 384 pages.

Dionne, G. and Saissi Hassani, S. (2017). Hidden Markov regimes in operational loss data: Application to the recent financial crisis. *Journal of Operational Risk* **12**(1):23-51. https://doi.org/10.21314/JOP.2017.188

Engle, R.F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* **22**(4):367-381. https://doi.org/10.1198/073500104000000370

Fernández, C. and Steel, M.F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* **93**(441):359-371. https://doi.org/10.1080/01621459.1998.10474117

Fernández, C., Osiewalski, J. and Steel, M.F. (1995). Modeling and inference with *v*-spherical distributions. *Journal of the American Statistical Association* **90**(432):1331-1340. https://doi.org/10.1080/01621459.1995.10476637

Fissler, T. and Ziegel, J.F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics* **44**(4):1680-1707. https://doi.org/10.1214/16-AOS1439

Fissler, T., Ziegel, J.F., and Gneiting, T. (2016). Expected shortfall is jointly elicitable with value at risk implications for backtesting. *Risk Magazine* arXiv:1507.00244.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association* **102**(477):359-378. https://doi.org/10.1198/016214506000001437

Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association* **102**(477):359-378. https://doi.org/10.1198/016214506000001437

Haas, M. (2009). Modelling skewness and kurtosis with the skewed Gauss-Laplace sum distribution. *Applied Economics Letters* **16**(12):1277-1283. https://doi.org/10.1080/17446540802400441

Haas, M., Mittnik, S., and Paolella, M.S. (2006). Modelling and predicting market risk with Laplace-gaussian mixture distributions. *Applied Financial Economics* **16**(15):1145-1162. https://doi.org/10.1080/09603100500438817

Iqbal, R., Sorwar, G., Baker, R., and Choudhry, T. (2020). Multiday expected shortfall under generalized *t* distributions: Evidence from global stock market. *Review of Quantitative Finance and Accounting* **55**(3):803-825. https://doi.org/10.1007/s11156-019-00860-1

Kerman, S.C. and McDonald, J.B. (2015). Skewness-kurtosis bounds for egb1, egb2, and special cases. *Communications in Statistics-Theory and Methods* **44**(18):3857-3864. https://doi.org/10.1080/03610926.2013.844255

Kim, M. and Lee, S. (2021). Risk measurement for conditionally heteroscedastic location-scale time series models with astd and aepd innovations. *Journal of Statistical Computation and Simulation*, pages 1-23. https://doi.org/10.1080/00949655.2021.2005062

Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives* **3**(2):73-84. https://doi.org/10.3905/jod.1995.407942

Lee, S.C. and Lin, X.S. (2012). Modeling dependent risks with multivariate Erlang mixtures. *ASTIN* Bulletin: The Journal of the IAA **42**(1):153-180.

McDonald, J.B. and Michelfelder, R.A. (2016). Partially adaptive and robust estimation of asset models: Accommodating skewness and kurtosis in returns. *Journal of Mathematical Finance* **7**(1):219-237. https://doi.org/10.4236/jmf.2017.71012

McNeil, A.J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance* **7**(3-4):271-300. https://doi.org/10.1016/S0927-5398(00)00012-8

McNeil, A.J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-Revised Edition*. Princetown University Press.

Miao, D.W.C., Lee, H.C. and Chen, H. (2016). A standardized normal-Laplace mixture distribution fitted to symmetric implied volatility smiles. *Communications in Statistics-Simulation and Computation* **45**(4):1249-1267. https://doi.org/10.1080/03610918.2013.816559

Molina-Munoz, E., Mora-Valencia, A., and Perote, J. (2021). Backtesting expected shortfall for world stock index etfs with extreme value theory and gram-charlier mixtures. *International Journal of Finance and Economics* **26**(3):4163-4189. https://doi.org/10.1002/ijfe.2009

Nolde, N. and Ziegel, J.F. (2017). Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics* **11**(4):1833-1874. https://doi.org/10.1214/17-AOAS1041

Patton, A.J., Ziegel, J.F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics* **211**(2):388-413. https://doi.org/10.1016/j.jeconom.2018.10.008

R CORE TEAM (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at http://www.R-project.org/

Rigby, B., Stasinopoulos, M., Heller, G. and Voudouris, V. (2014). The distribution toolbox of GAMLSS. *gamlss.org*.

Righi, M. and Ceretta, P.S. (2015). A comparison of expected shortfall estimation models. *Journal of Economics and Business* **78**:14-47. https://doi.org/10.1016/j.jeconbus.2014.11.002

Rockafellar, R.T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26**(7):1443-1471. https://doi.org/10.1016/S0378-4266(02)00271-6

Rombouts, J.V. and Bouaddi, M. (2009). Mixed exponential power asymmetric conditional heteroskedasticity. *Studies in Nonlinear Dynamics & Econometrics* **13**(3):1-32. https://doi.org/10.2202/1558-3708.1645

Taylor, J.W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business and Economic Statistics* **37**(1):121-133. https://doi.org/10.1080/07350015.2017.1281815

Zhu, D. and Galbraith, J.W. (2011). Modeling and forecasting expected shortfall with the generalized asymmetric student-t and asymmetric exponential power distributions. *Journal of Empirical Finance* **18**(4):765-778. https://doi.org/10.1016/j.jempfin.2011.05.006

Zhu, D. and Zinde-Walsh, V. (2009). Properties and estimation of asymmetric exponential power distribution. *Journal of Econometrics* **148**(1):86-99. https://doi.org/10.1016/j.jeconom.2008.09.038

Zoia, M. G., Biffi, P., and Nicolussi, F. (2018). Value at risk and expected shortfall based on gram-charlier-like expansions. *Journal of Banking and Finance* **93**:92-104. https://doi.org/10.1016/j.jbankfin.2018.06.001

Forecasting VaR and CVaR based on Skewed Exponential Power mixture in compliance with the new market risk regulation

25 November 2022

Online Appendices

The appendices present mathematical developments regarding VaR and CVaR formulas of competing models. Table A.1 shows the symbols of different models. Appendix A2 presents the general expression of CVaR. Appendix A3 shows how to compute CVaR for a mixture of distributions. Appendices A4 to A10 develop the CVaR formulas for different statistical distributions. Appendix A11 derives an optimal portfolio, which serves as input data beginning in Section 2. Appendix A12 contains complementary derivations. Complementary tables are presented in Appendix A13.

A1. Estimated models

Table A.1
Model Symbol Definitions

Model	Symbol	Description of the model
M1	1:NO	Normal distribution
M2	1:T	Student-*t* distribution
M3	1:EGB2	Exponential GB2 distribution
M4	2:NO	Mixture of 2 normal distributions
M5	2:T	Mixture of 2 Student-*t* distributions
M6	3:NO	Mixture of 3 normal distributions
M7	2:SN2	Mixture of 2 SN2 distributions
M8	2:SEP3	Mixture of 2 SEP3 distributions
M9	1:ST3	Skew-*t* distribution (type 3)
M10	1:SEP3	SEP3 distribution
M11	1:GP	Generalized Pareto distribution

A2. General expression of CVaR

We are interested in the family of location-scale parametric distributions $\overline{\mathrm{F}}$. For a distribution $\mathrm{F} \in \overline{\mathrm{F}}$ and a r.v. $\mathrm{Y} \sim \mathrm{F}$, the reduced variable $\mathrm{Z} = (\mathrm{Y} - \mu)/\sigma$, where μ and σ denote location and scale parameters, follows the distribution F^0 defined by:

$$\mathrm{F}(\mathrm{y}) = \mathrm{F}^0(\mathrm{z})$$

where *y* and *z* are generic variables for Y and Z random variables, respectively.

F^0 is the reduced cumulative function of F. The reduced density $\mathrm{f}^0(\cdot)$ is related to the density $\mathrm{f}(\cdot)$ by:

$$\mathrm{f}(\mathrm{y}) = \frac{\mathrm{f}^0(\mathrm{z})}{\sigma} \tag{A0}$$

In the following, downside risk is located in the left tail (negative returns). The risk measures are negated, so they are positive. The probability level p corresponds to the degree of confidence $(1-\mathrm{p})$. The p-quantile $\mathrm{q} = \text{Quantile}\,(\mathrm{Y},\ \text{level} = \mathrm{p})$, so $\mathrm{q} < 0$ (left tail). In the other hand, we note $\mathrm{e} = \mathrm{E}\left[\mathrm{Y} \middle| \mathrm{Y} \leq \mathrm{q}\right]$, which means $\mathrm{e} < 0$ (left tail). Then we compute $\mathrm{p-VaR} = -\mathrm{q}$ and $\mathrm{p-CVaR} = -\mathrm{e}$.

According to Broda and Paolella (2011), the tail quantity of a density $\mathrm{f}(\cdot)$ at point *x* is defined by: $\mathrm{tail}_{\mathrm{f}}(\mathrm{x}) \overset{\mathrm{def}}{=} \int_{-\infty}^{\mathrm{x}} \mathrm{t}\,\mathrm{f}(\mathrm{t})\,\mathrm{dt}$. We develop the expression of CVaR using this definition:

$$\mathrm{e} = \mathrm{E}\left[\mathrm{Y} \middle| \mathrm{Y} \leq \mathrm{q}\right] = \frac{1}{\mathrm{F}(\mathrm{q})} \int_{-\infty}^{\mathrm{q}} \mathrm{y}\,\mathrm{f}(\mathrm{y})\,\mathrm{dy} \tag{A1}$$

$$= \frac{1}{\mathrm{p}} \left\{ \int_{-\infty}^{\frac{\mathrm{q}-\mu}{\sigma}} (\mu + \sigma \mathrm{z}) \frac{\mathrm{f}^0(\mathrm{z})}{\sigma} \sigma \mathrm{d}(\mathrm{z}) \right\} \tag{A2}$$

$$= \frac{1}{\mathrm{p}} \left\{ \int_{-\infty}^{\frac{\mathrm{q}-\mu}{\sigma}} \mu\, \mathrm{f}^0(\mathrm{z})\,\mathrm{dz} + \sigma \int_{-\infty}^{\frac{\mathrm{q}-\mu}{\sigma}} \mathrm{z}\, \mathrm{f}^0(\mathrm{z})\,\mathrm{d}(\mathrm{z}) \right\} \tag{A3}$$

$$= \frac{1}{\mathrm{p}} \left\{ \mu\, \mathrm{F}^0\left(\frac{\mathrm{q}-\mu}{\sigma}\right) + \sigma \mathrm{Tail}_{\mathrm{f}^0}\left(\frac{\mathrm{q}-\mu}{\sigma}\right) \right\} \tag{A4a}$$

Equation (A2) is obtained by using (A0) after a change of variable $z=(y-\mu)/\sigma$, or $y=\mu+\sigma z$, so, $dy=\sigma dz$. Note that there are two parts in formula (A4a): the first one is μ/p times the centered reduced cumulative $F^0(\cdot)$ evaluated at the centered reduced quantity $(q-\mu)/\sigma$. The second one is σ/p times the tail of f^0, also evaluated at $(q-\mu)/\sigma$. CVaR is finally:

$$\mathrm{CVaR}_f=-e=-\frac{1}{p}\left\{\mu\, F^0\left(\frac{q-\mu}{\sigma}\right)+\sigma \mathrm{Tail}_{f^0}\left(\frac{q-\mu}{\sigma}\right)\right\} \tag{A4}$$

An important remark is that we have of course $F^0((q-\mu)/\sigma)=p$, which would simplify the expression (A4). Even so, we will leave the expression as it is so that it will be of the same form as for mixtures of distributions where there will be several cumulatives $F_i^0(\cdot)$, for which $F_i^0((q-\mu)/\sigma)\neq p$.

A3. CVaR of a mixture of distributions

Let $m(\cdot)$ be a mixture of *n* densities $f_i(\cdot), i=1,...,n$. Each density $f_i\in\overline{F}$ has a parameter of location μ_i and scale σ_i. The mixture density $f_m(\cdot)$ and its distribution $F_m(\cdot)$ are written as:

$$f_m(y)=\sum_1^n c_i\, f_i(y), \qquad F_m(y)=\sum_1^n c_i\, F_i(y)$$

where c_i is a probability, to be estimated, regarding the weight of density $f_i(\cdot)$. The sum of the c_i is equal to 1.

Let q_m be the p-quantile of the mixture. We denote $f_i^0(\cdot)$ and $F_i^0(\cdot)$ as the reduced density and the reduced cumulative of the *i*th density. The calculations are such as:

$$
\begin{aligned}
e &= E\left[Y \middle| Y \le q_m\right] = \frac{1}{p}\int_{-\infty}^{q_m} y\, f_m(y)\, dy \\
&= \frac{1}{p}\sum_{1}^{n} c_i \int_{-\infty}^{q_m} y\, f_i(y)\, d(y) \\
&= \frac{1}{p}\sum_{1}^{n} c_i \int_{-\infty}^{\frac{q_m-\mu_i}{\sigma_i}} (\mu_i + \sigma_i z_i)\frac{f_i^0(z_i)}{\sigma_i}\sigma_i d(z_i) \\
&= \frac{1}{p}\sum_{1}^{n} c_i \left\{ \mu_i F_i^0\left(\frac{q_m-\mu_i}{\sigma_i}\right) + \sigma_i \mathrm{Tail}_{f_i^0}\left(\frac{q_m-\mu_i}{\sigma_i}\right)\right\}
\end{aligned}
$$

or, in vector form:

$$
\mathrm{CVaR}_m = -e = -\frac{1}{p}\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}^T \times \left[\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} \times \begin{pmatrix} F_1^0\left(\frac{q_m-\mu_1}{\sigma_1}\right) \\ \vdots \\ F_n^0\left(\frac{q_m-\mu_n}{\sigma_n}\right) \end{pmatrix} + \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{pmatrix} \times \begin{pmatrix} \mathrm{Tail}_{f_1}^0\left(\frac{q_m-\mu_1}{\sigma_1}\right) \\ \vdots \\ \mathrm{Tail}_{f_n}^0\left(\frac{q_m-\mu_n}{\sigma_n}\right) \end{pmatrix} \right].
$$

Generally, q_m is found numerically as a solution to the equation $F_m(q_m) - p = 0$. Note that for a distribution *i*, $F_i^0\left((q_m - \mu_i)/\sigma_i\right) \ne p$.

A4. CVaR of a normal distribution

The density $\phi_{\mu,\sigma}(\cdot)$ of a normal distribution $N(\mu,\sigma)$ is:

$$
\phi_{\mu,\sigma}(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right).
$$

We denote both $\phi_0(\cdot)$ and $\Phi_0(\cdot)$ as the density and the cumulative of the standard normal distribution $N(0,1)$. It is easy to show that:

$$
\frac{\partial}{\partial x}\phi_0(x) = -x\phi_0(x). \tag{A5a}
$$

For $Y \sim N(\mu,\sigma)$, q being the p-quantile of Y is found by solving $P(Y \le q) = P\left(Z \le \frac{q-\mu}{\sigma}\right) = p$, hence:

$$q = \mu + \sigma\Phi_0^{-1}(p).$$

Further, with the definition of tail and with the help of equation (A5a) , we find:

$$\mathrm{Tail}_{\phi_0}(x) = \int_{-\infty}^{x} z\,\phi_0(z)\,dz = -\phi_0(x). \tag{A5}$$

We apply (A4) and (A5) to obtain:

$$\mathrm{CVaR}_{\phi,\mu,\sigma} = -\frac{1}{p}\left\{\mu\Phi_0\left(\frac{q-\mu}{\sigma}\right) - \sigma\phi_0\left(\frac{q-\mu}{\sigma}\right)\right\}. \tag{A6}$$

A5. CVaR of a Student-*t* distribution

The density $f_{T,\mu,\sigma,\nu}(\cdot)$ of the Student-*t* of parameters μ (location), σ (scale) and ν (degrees of freedom) is written as:

$$f_{T,\mu,\sigma,\nu}(y) = \frac{A}{\sigma}\left(1 + \left(\frac{y-\mu}{\sigma}\right)^2 \frac{1}{\nu}\right)^{-b}$$

where $A = \left[\sqrt{\nu} \times B(1/2, \nu/2)\right]^{-1}$ and $b = (\nu+1)/2$. $B(\cdot)$ is the beta function.[1] The reduced functions are noted $f_{T,\nu}^{0}(\cdot)$ and $F_{T,\nu}^{0}(\cdot)$. We determine *q* from the VaR of $Y \sim t(\mu,\sigma,\nu)$ to the degree of confidence $(1-p)$:

$$P(Y \le q) = P(\mu + \sigma Z \le q) \Rightarrow F_{T,\nu}^{0}\left(\frac{q-\mu}{\sigma}\right) = p$$

$$q = \mu + \sigma F_{T,\nu}^{0-1}(p)$$

where $F_{T,\nu}^{0-1}(\cdot)$ is the quantile (or inverse) function of $F_{T,\nu}^{0}(\cdot)$. The tail at point *x* is by definition:

$$\mathrm{Tail}_{f_{T,\nu}^{0}}(x) = \int_{-\infty}^{x} z\, f_{T,\nu}^{0}(z)\,dz = A\int_{-\infty}^{x} z\left(1 + z^2/\nu\right)^{-b} dz. \tag{A7}$$

[1] There is another way to write the constant A with the gamma function $\Gamma(\cdot)$ instead of the beta function.

We change the variable $u = z^2/\nu$, hence $zdz = \nu du/2$. The integral of equation (A7) becomes:

$$\mathrm{Tail}_{f^0_{T,\nu}}(x) = A\frac{\nu}{2}\int_{-\infty}^{\frac{x^2}{\nu}}(1+u)^{-b}\,du$$

$$= \frac{A}{2}\frac{\nu}{-b+1}\Big[(1+u)^{-b+1}\Big]_{-\infty}^{\frac{x^2}{\nu}} = \frac{\nu}{2(1-b)}\left(1+\frac{x^2}{\nu}\right)\times A\left(1+\frac{x^2}{\nu}\right)^{-b} \quad \text{(A8)}$$

$$= -\frac{\nu+x^2}{\nu-1}\times f^0_{T,\nu}(x). \quad \text{(A9)}$$

In equation (A8), we replace b with its value $(\nu+1)/2$. The final expression of the tail is simplified in (A9). In order to be valid we need to have $\nu > 1$. We now apply (A9) in (A4) to find:

$$\mathrm{CVaR}_{f_{T,\mu,\sigma,\nu}} = -\frac{1}{p}\left\{\mu F^0_{T,\nu}\left(\frac{q-\mu}{\sigma}\right) - \sigma\frac{\nu+\left(\frac{q-\mu}{\sigma}\right)^2}{\nu-1}f^0_{T,\nu}\left(\frac{q-\mu}{\sigma}\right)\right\}.$$

Important: In Excel, the functions related to Student-*t* distribution consider the degree of freedom ν to be an integer. Therefore, calculations cannot be made in standard form, and an additional module is required. The XRealStats.xlam module is used. It must be downloaded from the website[2], placed in the C:/TP5 directory and activated to use the functions that allow calculations with ν∈R. The cumulative and density functions are called by T_DIST. The inverse of the cumulative function is T_INV.

A6. Exponential Generalized Beta type 2 distribution: EGB2

The EGB2 (Exponential Generalized Beta type 2) density has four parameters and is written, according to Kerman and McDonald (2015), for $y \in R$:

$$f(y|\mu,\sigma,\nu,\tau) = \frac{e^{\nu z}}{|\sigma|\times B(\nu,\tau)\left(1+e^z\right)^{\nu+\tau}}$$

[2] http://www.real-statistics.com/free-download/

where $z = (y-\mu)/\sigma$, $\mu, \sigma \in R$, $\nu, \tau > 0$. $B(\cdot)$ is the standard beta function.

The parameters ν and τ characterize both tail thickness and the asymmetry of the distribution. The distribution has a negative or positive asymmetry, or is symmetrical when $\nu < \tau$, $\nu > \tau$ or $\nu = \tau$ respectively. As for the tail thickness, the smaller the ν, the thicker the left tail (all other parameters being equal).

The EGB2 includes many parametric distributions as special cases. Specifically, when $\nu \approx \tau \to +\infty$, the distribution converges to the normal. In practice, this convergence can be considered to have been reached when $\nu \approx \tau > 15$. When $\nu = \tau = 1$, EGB2 becomes a logistic distribution. Further, lemma 2 of Caivano and Harvey (2014) shows that EGB2 tends toward a Laplace density when $\nu \approx \tau \approx 0$. Other interesting special cases of EGB2 and GB2 are presented by Kerman and McDonald (2015).

Cummins, Dionne, McDonald, and Pritchett (1990) applies the GB2 to compute reinsurance premiums and quantiles for the distribution of total insurance losses. EGB2 is increasingly used in finance, as the studies by Caivano and Harvey (2014), McDonald and Michelfelder (2016), and in operational risk management (Dionne and Saissi Hassani, 2017).

A7. CVaR of a Skew-Normal distribution: SN2

The definition of the density of Skew-Normal (SN2) of Fernandez et al. (1995) for $y \in R$ can be written as:

$$f_{SN2,\mu,\sigma,\nu}(y) = \frac{2\nu}{\sigma\sqrt{2\pi}\left(1+\nu^2\right)}\left\{\exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 \nu^2\right) I_{(y<\mu)} + \exp\left(-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 \frac{1}{\nu^2}\right) I_{(y\geq\mu)}\right\} \quad \text{(A10)}$$

where $\mu \in R$, $\sigma > 0$, $\nu > 0$. SN2 refers to the classification proposed by Rigby et al. (2014).

If $\nu < 1$, asymmetry is to the left (negative returns); if $\nu > 1$, asymmetry is positive. When $\nu = 1$, we return to a normal (symmetrical) distribution. This density is also useful to compute capital in operational risk management, as in the study by Dionne and Saissi Hassani (2017). A random variable $Y \sim F_{SN2,\mu,\sigma,\nu} \Rightarrow Z = (Y-\mu)/\sigma \sim F^0_{SN2,\nu}$. For $z < 0$, only the left side of the equation (A10) is non-zero. The reduced density is then written as:

$$f^0_{SN2,\nu}(z) = \frac{2\nu}{1+\nu^2}\,\phi_0(z \times \nu).$$

The cumulative at point $z < 0$ is written as:

$$F^0_{SN2,\nu}(z) = \int_{-\infty}^{z} f^0_{SN2,\nu}(t)\,dt = \frac{2}{1+\nu^2}\,\Phi_0(z \times \nu).$$

The functions $\phi_0(\cdot)$ and $\Phi_0(\cdot)$ designate the cumulative and the reduced centered normal density $N(0,1)$. The previous equation allows us to find the expression of the VaR at the confidence level $(1-p)$:

$$P(Y \le q) = P\left(Z \le \frac{q-\mu}{\sigma}\right) \Rightarrow F^0_{SN2,\nu}\left(\frac{q-\mu}{\sigma}\right) = p$$

$$\frac{2}{1+\nu^2}\Phi_0\left(\frac{q-\mu}{\sigma} \times \nu\right) = p$$

$$q = \mu + \sigma\frac{1}{\nu}\Phi_0^{-1}\left(p\,\frac{1+\nu^2}{2}\right). \tag{A11}$$

The expression (A11) is valid only if $p(1+\nu^2)/2 \le 1$, otherwise $\Phi^{-1}(\cdot)$ would not be defined. This requires that $\nu \le \sqrt{2/p-1}$.

The expression of the tail is developed as follows:

$$\begin{aligned}
\mathrm{Tail}_{SN2,\nu}(x) &= \int_{-\infty}^{x} z\,f^0(z)\,dz = \int_{-\infty}^{x} \frac{2\nu}{1+\nu^2} z\,\phi_0(z \times \nu)\,dz \\
&= \frac{2\nu}{1+\nu^2}\int_{-\infty}^{x\times\nu} \frac{u}{\nu}\phi_0(u)\frac{du}{\nu} = \frac{2}{\nu(1+\nu^2)}\left[-\phi_0(u)\right]_{-\infty}^{x\times\nu} \\
&= -\frac{2}{\nu(1+\nu^2)}\,\phi_0(x \times \nu). \qquad \text{(A12)}
\end{aligned}$$

Equation (A12) is obtained by changing the variable $u = z \times \nu$ and using equation (A5a). Equations (A11) and (A12) in (A4) give the expression of CVaR:

$$\mathrm{CVaR}_{SN2,\mu,\sigma,\nu} = -\frac{1}{p}\left\{\frac{2}{1+\nu^2}\left[\mu\Phi_0\left(\frac{q-\mu}{\sigma}\nu\right)-\sigma\frac{1}{\nu}\phi_0\left(\frac{q-\mu}{\sigma}\nu\right)\right]\right\}.$$

Again, when $\nu = 1$ we find the CVaR of $N(\mu,\sigma)$.

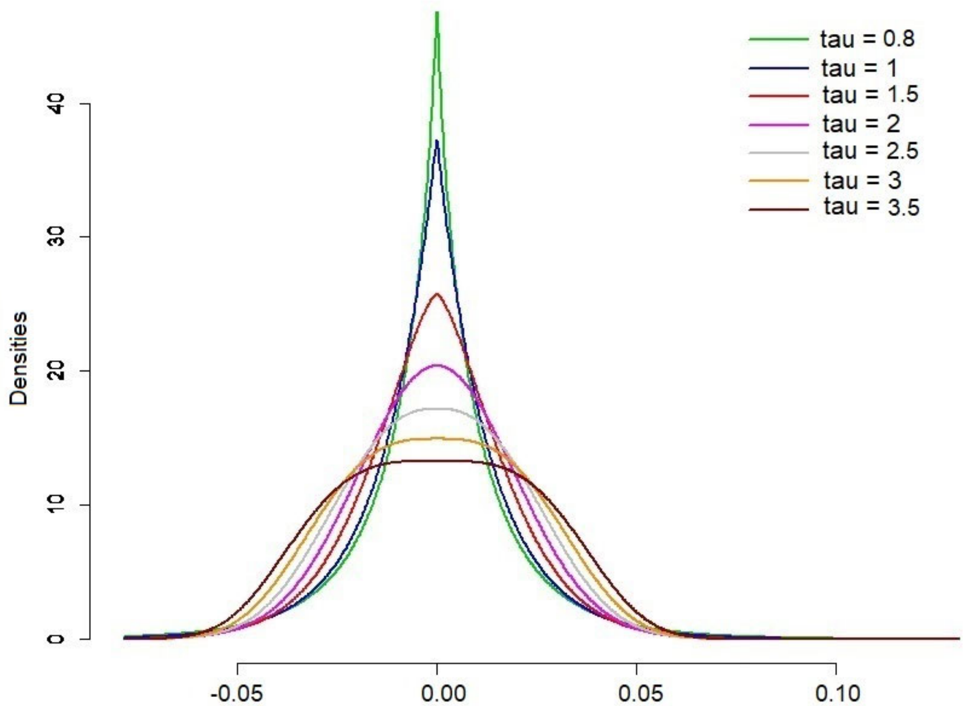
A8. Skewed Exponential Power type 3 Distribution: SEP3

Fernandez et al. (1995) defined this distribution. SEP3 refers to the classification proposed by Rigby et al. (2014). The density of SEP3 is written as:

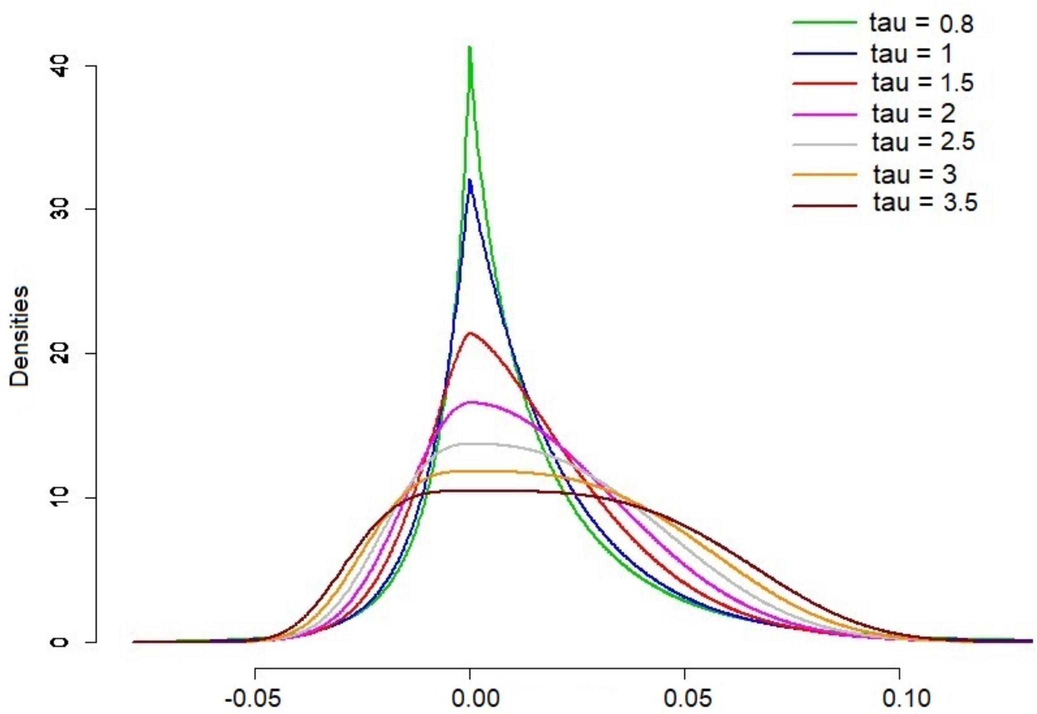
$$f_{SEP3,\mu,\sigma,\nu,\tau}(y) = \frac{c}{\sigma}\left\{\exp\left(-\frac{1}{2}\left|\frac{y-\mu}{\sigma}\nu\right|^{\tau}\right)I_{(y<\mu)} + \exp\left(-\frac{1}{2}\left|\frac{y-\mu}{\sigma}\frac{1}{\nu}\right|^{\tau}\right)I_{(y\geq\mu)}\right\}$$

where $c = \nu \times \tau \times \left[\left(1+\nu^2\right)2^{1/\tau}\Gamma\left(1/\tau\right)\right]^{-1}$ and where $\mu \in R, \sigma > 0, \nu \in R, \tau > 0$. They are respectively the parameters of location, scale, asymmetry, and tail thickness. SEP3 has as special cases the SN2 when $\tau = 2$ and a Laplace distribution (asymmetric version) when $\tau = 1$. Note that other names exist in the literature to designate distributions comparable to SEP3, such as AP (Asymmetric Power) and AEP (Asymmetric Exponential Power).

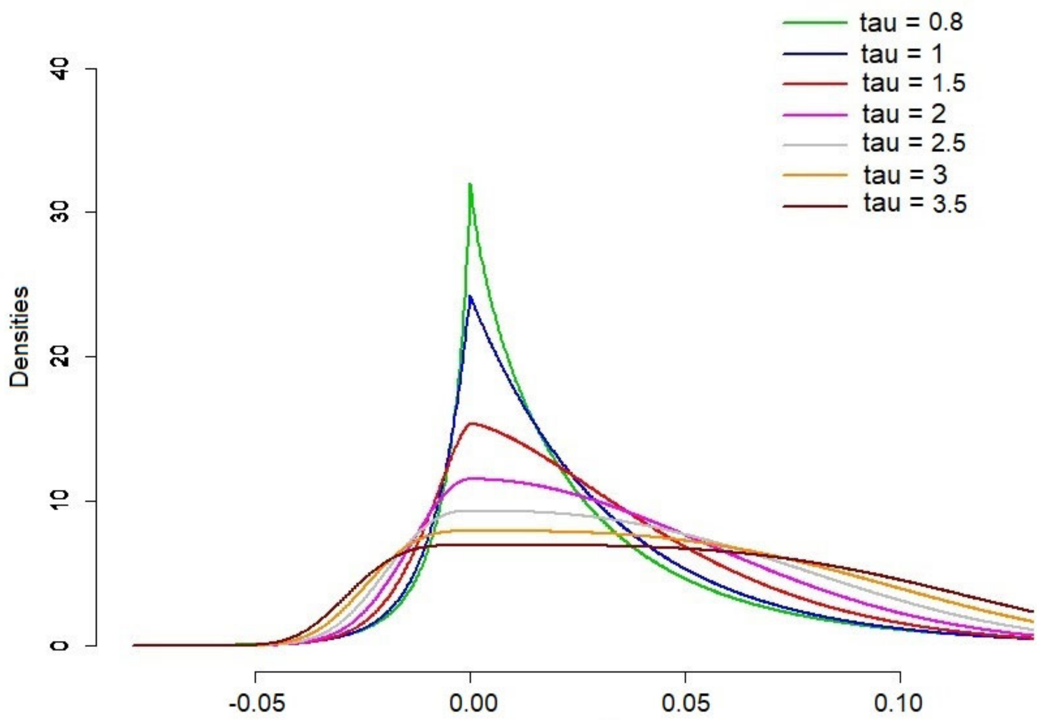
SEP3 can be leptokurtic when $\tau < 2$ or platykurtic when $\tau > 2$ (see Figure A1). VaR and CVaR calculations use gamma functions and the gamma distribution, as shown in the next section.


tau = 0.8
tau = 1
tau = 1.5
tau = 2
tau = 2.5
tau = 3
tau = 3.5
Densities
0
10
20
30
40
-0.05
0.00
0.05
0.10

$\nu = 1$


tau = 0.8
tau = 1
tau = 1.5
tau = 2
tau = 2.5
tau = 3
tau = 3.5
Densities
0
10
20
30
40
-0.05
0.00
0.05
0.10

$\nu = 1.5$



$\nu = 2$

Figure A1: Plots of SEP3 with different values of τ and ν

A9. VaR and CVaR of SEP3

As we did for SN2, we develop the expression of reduced cumulative of SEP3 for $z<0$ (left tail) by writing:

$$F^{0}_{SEP3,\nu,\tau}(z) = \int_{-\infty}^{z} \frac{\tau\nu}{\left(1+\nu^2\right)2^{1/\tau}\Gamma\left(1/\tau\right)} \times \exp\left(-\frac{1}{2}\left|w\nu\right|^{\tau}\right) dw$$

$$= \frac{2^{1/\tau}}{\nu\tau} \frac{\tau\nu}{\left(1+\nu^2\right)2^{1/\tau}\Gamma\left(1/\tau\right)} \int_{(z\nu)^{\tau}/2}^{+\infty} u^{1/\tau-1} e^{-u}\, du \tag{A13}$$

$$= \frac{1}{\left(1+\nu^2\right)\Gamma\left(1/\tau\right)} \int_{(z\nu)^{\tau}/2}^{+\infty} u^{1/\tau-1} e^{-u}\, du. \tag{A14}$$

Equation (A13) is immediate after the change of variable $u = \left(-w\nu\right)^{\tau}/2$ and by positing $s = -z > 0.$ Note that the inside of the integral $u^{1/\tau-1}e^{-u}\, du$ is reminiscent of the gamma function.

We need the complete gamma function $\Gamma(\cdot)$ and its incomplete version $\gamma(\cdot,\cdot)$, which are defined by:

$$\gamma(a,r) = \int_0^r t^{a-1}e^{-t}dt \qquad a > 0, r > 0$$

$$\Gamma(a) = \int_0^{+\infty} t^{a-1}e^{-t}dt \qquad a > 0 .$$

Parameter *a* is for the shape of these functions. It is easy to see that $\Gamma(a) = \gamma(a,+\infty)$. We also have a distribution that bears the same name, i.e. gamma,[3] whose cumulative parameter shape = a (and scale $= 1$ because it is standardized) evaluated at the point $x > 0$. It is written as $G_a(x) = [\Gamma(a)]^{-1}\gamma(a,x)$. The calculation of $F^0_{SEP3,\nu,\tau}(z)$ can be obtained from equality (A14):

$$F^0_{SEP3,\nu,\tau}(z) = \frac{1}{(1+\nu^2)\Gamma(1/\tau)}\int_{(z\nu)^\tau/2}^{+\infty} u^{1/\tau-1}e^{-u}\,du$$

$$= \frac{1}{(1+\nu^2)\Gamma(1/\tau)}\left\{\int_0^{+\infty} - \int_0^{(z\nu)^\tau/2}\right\} = \frac{1}{(1+\nu^2)}\frac{\Gamma(1/\tau) - \gamma\left(1/\tau,(z\nu)^\tau/2\right)}{\Gamma(1/\tau)} \qquad \text{(A15)}$$

$$= \frac{1}{1+\nu^2}\left(1 - G_{1/\tau}\left(\frac{|z\nu|^\tau}{2}\right)\right) \qquad \text{(A16)}$$

Equality (A15) is a cut-off of the integral's bounds that allows finding the gamma functions. To save space, we have not inserted the complete mathematical expressions of the two integrals in (A15), which are the same as in the previous equation. The expression is simplified by using the cumulative $G_{1/\tau}(\text{shape} = 1/\tau \text{ and scale} = 1)$. By inverting (A16), the quantile of VaR at the degree of confidence $(1-p)$ is immediate:

[3] Under the same "gamma" designation, three entities can be distinguished: the function $\Gamma(.)$ (complete from 0 to $+\infty$) and the incomplete function (its integral stops at a point $r < +\infty$). The third entity is the gamma distribution with two parameters: shape and scale.

$$F^0_{SEP3,\nu,\tau}\left(\frac{q-\mu}{\sigma}\right)=p$$

$$q=\mu+\sigma\times\frac{\left[2\times G^{-1}_{1/\tau}\left(1-p\left(1+\nu^2\right)\right)\right]^{1/\tau}}{\nu}$$

The calculation of the tail of SEP3 is similar to that done for the cumulative, but with a shape $2/\tau$ parameter for $x<0$:

$$\begin{aligned}\mathrm{Tail}_{SEP3,\nu,\tau}(x)&=\int_{-\infty}^{x}z f^0(z)\,dz=\int_{-\infty}^{x}c\times z\times\exp\left(-\frac{1}{2}|z\nu|^{\tau}\right)dz\\&=\frac{-2^{1/\tau}}{\nu\left(1+\nu^2\right)\Gamma(1/\tau)}\int_{(-x\nu)^{\tau}/2}^{+\infty}u^{2/\tau-1}e^{-u}\,du\\&=\frac{-2^{1/\tau}}{\nu\left(1+\nu^2\right)\Gamma(1/\tau)}\Gamma(2/\tau)\left(1-G_{2/\tau}\left(\frac{|x\nu|^{\tau}}{2}\right)\right).\end{aligned}\qquad\text{(A17)}$$

Finally, by putting (A16) and (A17) in (A4) we find:

$$\mathrm{CVaR}_{SEP3,\mu,\sigma,\nu,\tau}=-\frac{1}{p}\left\{\frac{1}{1+\nu^2}\left[\mu\times\left(1-G_{1/\tau}\left(\frac{\left|\frac{q-\mu}{\sigma}\nu\right|^{\tau}}{2}\right)\right)-\sigma\times\frac{2^{1/\tau}}{\nu}\frac{\Gamma(2/\tau)}{\Gamma(1/\tau)}\left(1-G_{2/\tau}\left(\frac{\left|\frac{q-\mu}{\sigma}\nu\right|^{\tau}}{2}\right)\right)\right]\right\}.$$

Remember that $G_{n/\tau}(x)$ is the cumulative gamma distribution of shape $=n/\tau$ and scale = 1 evaluated at point *x*. When $\tau=2$, we return to SN2. If $\tau=2$, and $\nu=1$, we get a normal distribution. The gamma distribution and the complete gamma function exist in standard Excel.

A10. Skew-*t* distribution: ST3

The density function of Skew-*t* of Fernandez et al. (1995), can be written for $x\in R$ as:

$$f_{ST3}\left(x\,|\,\mu,\sigma,\nu,\tau\right)=\frac{c}{\sigma}\left\{\left[1+\frac{\nu^2u^2}{\tau}\right]^{-(\tau+1)/2}1_{y<\mu}+\left[1+\frac{u^2}{\nu^2\tau}\right]^{-(\tau+1)/2}1_{y\geq\mu}\right\}\qquad\text{(A18)}$$

where $u=(x-\mu)/\sigma$; $\mu\in R$; $\sigma,\nu,\tau>0$, and $c=2\nu\left[\left(1+\nu^2\right)B(1/2,\tau/2)\tau^{1/2}\right]^{-1}$.

ST3 refers to the classification proposed by Rigby et al. (2014). For z<0, its reduced density and cumulative are:

$$f^0_{ST3,\tau}(z) = \frac{2\nu}{1+\nu^2} f^0_{T,\tau}(z\nu) \quad \text{and} \quad F^0_{ST3,\tau}(z) = \frac{2}{1+\nu^2} F^0_{T,\tau}(z\nu)$$

where $f^0_{T,\tau}(.)$ and $F^0_{T,\tau}(.)$ are the reduced density and cumulative of a Student-*t*. The tail at point x is:

$$\begin{aligned}
\mathrm{Tail}_{f^0_{ST3}}(x) &= \int_{-\infty}^{x} z\, f^0_{ST3}(z)\,dz = \frac{2\nu}{1+\nu^2}\int_{-\infty}^{x} z\, f^0_{T,\tau}(z\nu)\,dz \\
&= \frac{2}{\nu(1+\nu^2)}\int_{-\infty}^{x\nu} u\, f^0_{T,\tau}(u)\,du \\
&= \frac{2}{\nu(1+\nu^2)}\mathrm{Tail}_{f^0_{T,\tau}}(x\nu)
\end{aligned}$$

Hence, $F^0_{ST3,\tau}(q_0) = p \quad \Rightarrow \quad \mathrm{VaR}_{ST3} = -q = -\left\{\mu + \sigma\frac{1}{\nu}F^{0\;-1}_{T,\tau}\left(p\frac{1+\nu^2}{2}\right)\right\}$ (A19)

Note the similarity with SN2 expressions. We now apply (A4) and (A9) to find

$$\begin{aligned}
\mathrm{CVaR}_f &= -\frac{1}{p}\left\{\mu\, F^0_{ST3}\left(\frac{q-\mu}{\sigma}\right) + \sigma \mathrm{Tail}_{f^0}\left(\frac{q-\mu}{\sigma}\right)\right\} \\
&= -\frac{2}{p\left(1+\nu^2\right)}\left\{\mu F^0_{T,\tau}\left(\frac{q-\mu}{\sigma}\nu\right) - \sigma\frac{1}{\nu}\frac{\tau + \left(\frac{q-\mu}{\sigma}\nu\right)^2}{\tau-1} f^0_{T,\tau}\left(\frac{q-\mu}{\sigma}\nu\right)\right\}
\end{aligned} \tag{A20}$$

A11. Calculations on optimal portfolio

A11.1. Determining optimal portfolio

Let Y_t be the return on the given asset over a time horizon *t*. Y_t follows a distribution F, assumed strictly increasing with finite mean. At a given p-level, we have

$$\mathrm{VaR}^{p} = -F^{-1}(p) \tag{A21}$$

$$\mathrm{CVaR}^{p} = -E\left[Y_t \middle| Y_t < -\mathrm{VaR}^{p}\right] \tag{A22}$$

We now introduce the relative VaR and CVaR. Relative VaR, denoted VaRr, is the distance between a given quantile of return distribution and its mean. Relative CVaR, or CVaRr, can be introduced as CVaR coverage including the expected return. Their expressions are:

$$\mathrm{VaRr}^{p} = \mathrm{VaR}^{p} + E\left[Y_t\right] \text{ and } \mathrm{CVaRr}^{p} = \mathrm{CVaR}^{p} + E\left[Y_t\right].$$

We begin by demonstrating that optimal portfolio weights are identical minimizing VaRr or CVaRr at $p = 5\%$. Assuming normal returns for now, the portfolio measures are:

$$\mathrm{VaR} = -\mu_{\text{portfolio}} - \sigma_{\text{portfolio}} \times \Phi_0^{-1}(p) \quad \text{and} \quad \mathrm{VaRr} = -\sigma_{\text{portfolio}} \times \Phi_0^{-1}(p) \tag{A23}$$

$$\mathrm{CVaR} = -\mu_{\text{portfolio}} + \sigma_{\text{portfolio}} \frac{\phi_0\left(\Phi_0^{-1}(p)\right)}{p} \quad \text{and} \quad \mathrm{CVaRr} = \sigma_{\text{portfolio}} \frac{\phi_0\left(\Phi_0^{-1}(p)\right)}{p} \tag{A24}$$

where $\sigma_{\text{portfolio}} = \sqrt{\beta^T \Sigma \beta}$, β is security weights vector, Σ is the return variance-covariance matrix, $\Phi_0^{-1}(\cdot)$ and $\phi_0^{-1}(\cdot)$ are the quantile and densty functions of $N(0,1)$. Since $\Phi_0^{-1}(p)$ and $\phi_0\left(\Phi_0^{-1}(p)\right)/p$ are constant, both VaRr and CVaRr are minimized on $\sqrt{\beta^T \Sigma \beta}$ alone. Hence, the optimal portfolio is the same. Moreover, optimal weights are independent of p.

Table A.2a shows the results[4]. The portfolio mean and standard deviation are reported next to those of the three assets. We focus on VaR line in order to comment weighted sum column. The portfolio $\mathrm{VaR} = 2.079\%$. $R = (2.621\%, 4.321\%, 2.288\%)$ represents the VaR vector of individual assets (normal returns, equation (A23)).

[4] See also the Excel file available on the Canada Research Chair website at https://chairegestiondesrisques.hec.ca/en/seminars-and-publications/book-wiley/

Table A.2a : Optimal portfolio (normal model)

		IBM	GE	Walmart
	Weight	0.3889444	-0.0465131	0.6575686
	Portfolio			
Mean	0.05244%	0.07580%	-0.00286%	0.03472%
Variance*	0.01680			
Standard deviation	1.29631%	1.63961%	2.62559%	1.41254%
Skewness	0.35887			
Kurtosis	9.81578			
p = 5%				
VaR	2.07980%	2.62112%	4.32157%	2.28871%
VaRr	2.13224%	2.69692%	4.31871%	2.32343%
CVaR	2.62147%	2.62112%	4.32157%	2.28871%
CvaRr	2.67392%	2.69692%	4.31871%	2.32343%

*Variance is multiplied by 100 to show more decimal digits.

For p=5%, we calculate the empirical $\mathrm{VaR_{np}}$ as the negated value of the 61st smallest value of $\mathrm{Y_t}$. The subscript np indicates nonparametric empirical estimates throughout the sequel.

$$\mathrm{VaR_{np}} = -\mathrm{Quantile}(\mathrm{T}, \mathrm{p}) \tag{A25}$$

$$\mathrm{CVaR_{np}} = -\frac{1}{\mathrm{N_T}}\sum_{t=1}^{T} \mathrm{Y_t} \times 1_{(\mathrm{Y_t}<-\mathrm{VaR_{np}})} \quad \text{where} \quad \mathrm{N_T} = \sum_{t=1}^{T} 1_{(\mathrm{Y_t}<-\mathrm{VaR_{np}})} \tag{A26}$$

where $\mathbf{1_d}$ is the indicator function equal to 1 if $\boldsymbol{d}$ is true else 0.

Table A.2b displays the optimal portfolio's computed $\mathbf{VaR_{np}}$ and $\mathbf{CVaR_{np}}$ alongside those of the individual assets (Portfolio $\mathbf{VaR_{np}} = \mathbf{2.0473}\%$ and $\mathbf{CVaR_{np}} = \mathbf{2.9779}\%$).

Table A.2b
Empirical measurements $\mathrm{VaR_{np}}$ and $\mathrm{CVaR_{np}}$

		IBM	GE	Walmart
Weight		0.38894	-0.04651	0.65756
p = 5%				
Order: 61	Portfolio			
$\mathrm{VaR_{np}}$	2.04736%	2.69619%	4.12926%	2.02620%
$\mathrm{CVaR_{np}}$	2.97795%	3.82711%	6.43375%	3.21706%

Figure A2 clearly shows that a Gaussian cannot fit. Student-*t* is not sharp enough and does not keep enough mass around the mode. This rather suggests a Laplace density.

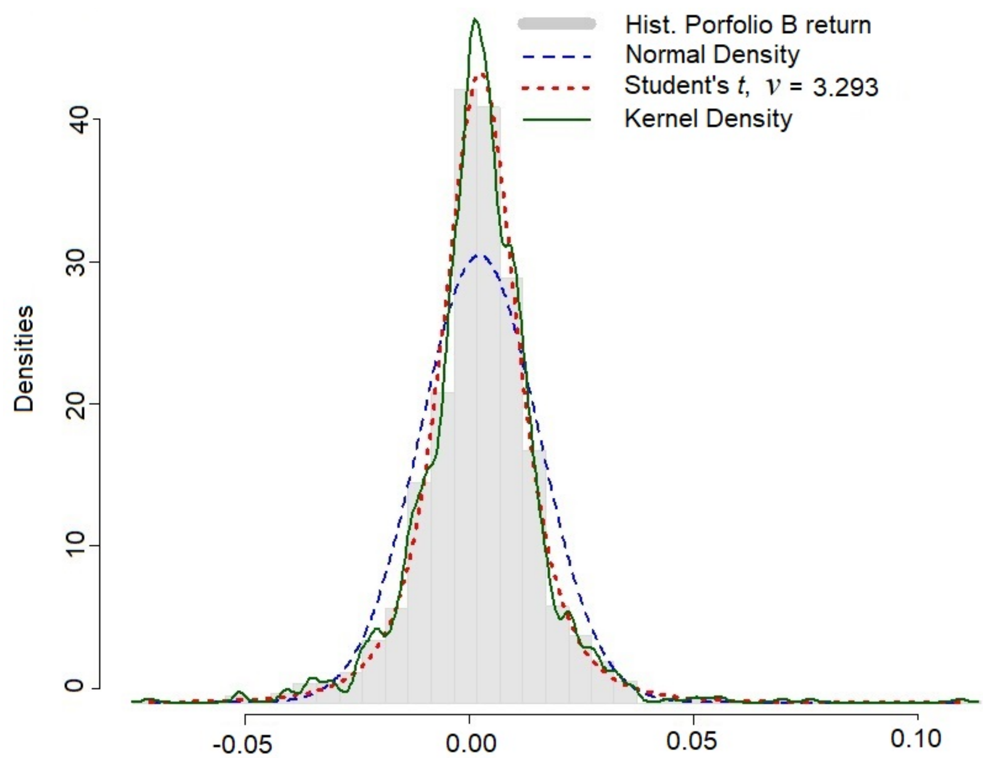


Figure A2: Histogram and densities of the optimal portfolio

A11.2. Preliminary estimations of competing models

The return portfolio Y_t is assumed to remain optimal for all competing models, to ensure, result comparability. Models M1 to M8 are directly fitted to $\{Y_t\}_{t=1}^{1200}$ using the maximum likelihood approach in static modeling.

Model validation consists of the usual criteria: AIC, BIC, and the KS goodness-of-fit test. Also, model-derived kurtosis and asymmetry coefficients are compared to empirical ones. Tables 3 and 4 depict results for p=5%. For mixtures, c_j identifies the weight of the j^{th} individual density in the mixture. Empirical $\text{VaR}_{\text{np}} = 2.0473\%$ and $\text{CVaR}_{\text{np}} = 2.9779\%$, are from Table A.2b.

Our first model, M1=1:NO, consists of a single Gaussian (Appendix A4). Obviously, M1 is far from allowing the targeted empirical skewness=0.35 and kurtosis=9.81. The next model, M2, assumes a single Student-*t* (Appendix A5). The estimated degree of freedom $\nu = 3.288 < 4$ confirms a high tail thickness. KS does not reject M2. AIC and BIC improve compared to M1, but asymmetry cannot be captured. M3=1:EGB2 has a single EGB2 density (exponential generalized beta 2; Appendix A6). Both thickness and skewness can be captured by ν and τ. Values $\nu = 0.165$ and $\tau = 0.158$ tend to 0. Therefore, EGB2 tends toward a Laplace (Lemma 2 of Caivano and Harvey, 2014). This confirms Section 1 remarks.

We now explore density mixtures. M4=2:NO is made of two normals (Appendices A2, A3, A4). $\text{Kurtosis} = 6.7$ improves. KS *p*-value$=$ $0.218 > 10\%$. However, $\text{CVaR} = 3.113\%$ is the farthest away from CVaR_{np}. Next is M5 =2:T, a two Student-*t* mixture (Appendices A2, A3, and A5). Kurtosis of 8.4 is close to 9.8. BIC=-7,207.6 deteriorates. $\text{VaR} \approx \text{VaR}_{\text{np}}$ but $\text{CVaR} = 3.041\% > \text{CVaR}_{\text{np}}$. A word about M6=3:NO (three normals), which almost reaches empirical kurtosis: $9.4 \approx 9.8$. $\text{VaR} \approx \text{VaR}_{\text{np}}$ and CVaR=3.004% is close to CVaR_{np}. However, it cannot capture data skewness.

To capture data skewness, we build M7=2:SN2, which is a two SN2 mixture (skewed normal type 2 of Fernández et al., 1995, Appendix A7). Both skewness parameters, $\nu_1 \approx 1.44 > 1$ and $\nu_2 \approx 1.10 > 1$, comply with the data. Although kurtosis and AIC deteriorate, the good news is that $\text{CVaR} = 2.968\%$ falls close to CVaR_{np}.

We now incorporate tail thickness, using two SEP3 distributions (skewed exponential power type 3 of Fernández et al., 1995). M8=2:SEP3 has nine parameters (Appendices A3, A8, and A9). AIC and BIC indicate a better fit to the data. $\text{VaR} = 1.992\% \le \text{VaR}_{\text{np}}$, but, interestingly, its $\text{CVaR} = 2.973\%$ falls close to CVaR_{np}.

Based on thickness and skewness parameters $\tau_1 = 0.959 \approx 1$, $\nu_1 = 1.031 \approx 1$, $\tau_2 = 2.108 \approx 2$, and $\nu_2 = 0.613 < 1$, the first SEP3 is a Laplace degenerate, whereas the second SEP3 is practically an asymmetric normal, which is an SN2. These facts illustrate relationships between our models. The results corroborate similar findings in the recent market-risk literature highlighting the Laplace-Gaussian mixture (Haas, 2009; Broda and Paolella, 2011; Miao et al., 2016).

For further analysis, Tables 4 and 5 (in main text) provide exhaustive information. Also, Table A.13 (Online Appendix A13) illustrates the behavior of the eight models at p = 2.5% and 1%, which is similar to p=5%.

To conclude this overview section, Figure A.3 depicts VaR and CVaR behaviors at 5% in the left tail for 2:SEP3 along with 1:NO and 1:T, including their densities. Risk measures are reported with negative values, so they can be shown with negative returns.

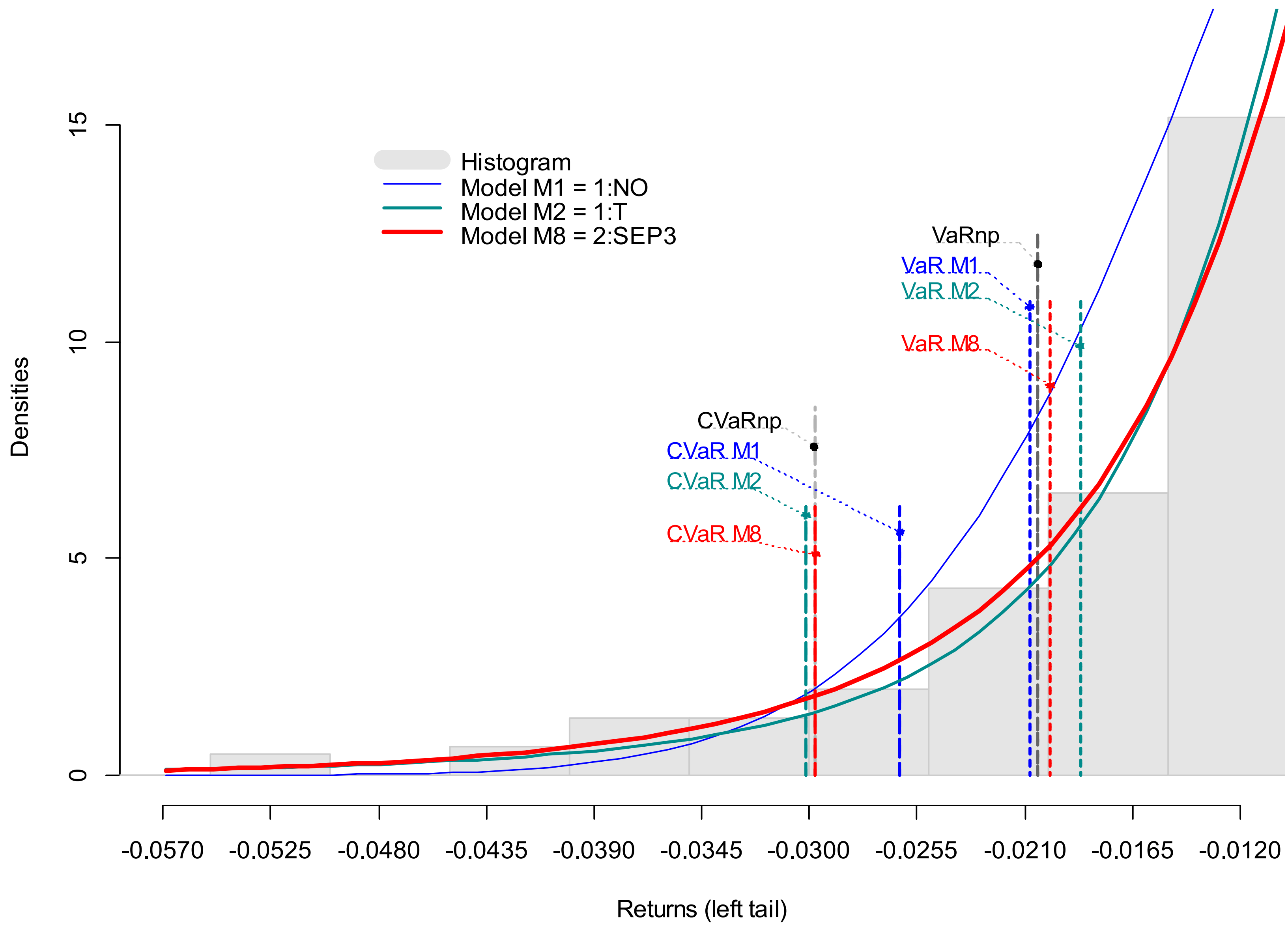


Figure A.3: VaR and CVaR plots of selected models in the left tail of returns

A12. Complementary derivations

A12.1. Converting between conditional risk measures of $\mathbf{Y_t}$ and $\mathbf{Z_t}$ given $\mathbf{\mathfrak{I}_{t-1}}$

The return series Y_t given $\mathfrak{I}_{t-1}$ follows a conditional distribution F_t: $Y_t|\mathfrak{I}_{t-1} \sim F_t$. We assume F_t continuous, strictly increasing with finite mean. Conditional risk measures at p-level are defined by

$$\mathrm{VaR}_t^p(Y_t) = -F_t^{-1}(p) \quad \text{and} \quad \mathrm{CVaR}_t^p(Y_t) = -E\left[Y_t|Y_t < -\mathrm{VaR}_t^p(Y_t), \mathfrak{I}_{t-1}\right].$$

Following NZ (2017) with adaption due to the left tail, assume that Y_t can be written as

$$Y_t = -\mu_t + \sigma_t Z_t \tag{A27}$$

where Z_t is an i.i.d. random standardized variable independent of $\mathfrak{I}_{t-1}$. $\mu_t = -E[Y_t|\mathfrak{I}_{t-1}]$ and $\sigma_t^2 = \mathrm{var}[Y_t|\mathfrak{I}_{t-1}]$.

Converting from $\mathrm{VaR}_{t+1}^p(Z_{t+1})$ to $\mathrm{VaR}_{t+1}^p(Y_{t+1})$:

Let $Y'_{t+1} = -\mathrm{VaR}_{t+1}^p(Y_{t+1})$ and $Z'_{t+1} = -\mathrm{VaR}_{t+1}^p(Z_{t+1})$. Using (A27), we can write:

$$Z'_{t+1} = (Y'_{t+1} + \mu_{t+1})/\sigma_{t+1}$$

$$-\mathrm{VaR}_{t+1}^p(Z_{t+1}) = (-\mathrm{VaR}_{t+1}^p(Y_{t+1}) + \mu_{t+1})/\sigma_{t+1}$$

$$\Rightarrow \mathrm{VaR}_{t+1}^p(Y_{t+1}) = \mu_{t+1} + \sigma_{t+1}\mathrm{VaR}_{t+1}^p(Z_{t+1}) \tag{A28}$$

Converting from $\mathrm{CVaR}_{t+1}^p(Z_{t+1})$ to $\mathrm{CVaR}_{t+1}^p(Y_{t+1})$:

$$\mathrm{CVaR}_{t+1}^p(Y_{t+1}) = -E[\,Y_{t+1}|Y_{t+1} < -\mathrm{VaR}_{t+1}^p\,(Y_{t+1}), \mathfrak{I}_t] \tag{A29}$$

$$= -E[\,-\mu_{t+1} + \sigma_{t+1}Z_{t+1}|-\mu_{t+1} + \sigma_{t+1}Z_{t+1} < -\mu_{t+1} - \sigma_{t+1}\mathrm{VaR}_{t+1}^p(Z_{t+1}), \mathfrak{I}_t] \tag{A30}$$

$$= \mu_{t+1} - \sigma_{t+1}E[Z_{t+1}|Z_{t+1} < -\mathrm{VaR}_{t+1}(Z_{t+1}), \mathfrak{I}_t]$$

$$\Rightarrow \mathrm{CVaR}_{t+1}^p(Y_{t+1}) = \mu_{t+1} + \sigma_{t+1}\mathrm{CVaR}_{t+1}(Z_{t+1}) \tag{A31}$$

Equation (A30) is the result of (A27) at t+1 and (A28) applied into (A29).

A12.2. Defining S_{log} scoring function of VaR

According to Gneiting and Raftery (2007, eq. (40)), a general form of strictly consistent scoring functions for VaR can be written as:

$$S_v(v_t, y_t) = (I_t - p) \times G(-v_t) - I_t \times G(y_t) + \eta(y_t), \tag{A32}$$

where $I_t = 1_{(y_t < -v_t)}$, G and η are functions. $E[\eta(Y_t)]$ must exist, and G continuously differentiable and strictly increasing.

We fix $G(x) = -\log(-x)$. In this case, the corresponding scoring function would be

$$S_v(v_t, y_t) = (p - I_t) \times \log(v_t) + I_t \times \log(-y_t) + \eta(y_t). \tag{A33}$$

If $\eta(.)$ is ignored in (A33), $S_v(v_t, y_t)$ would be not defined for $y_t > 0$. To fix this, notice $\min(y_t) = -7.15\% > -100\%$ (see Table 2). As a result, $1_{y_t < -1} = 0$ $\forall t$ $(t = 1, .., 1200)$. Therefore, for $v_t = 1$, we have

$$\begin{aligned} S_v(1, y_t) &= \left(p - 1_{y_t < -1}\right) \log(1) + 1_{y_t < -1} \log(-y_t) + \eta(y_t) \\ &= 1_{y_t < -1} \times \log(-y_t) + \eta(y_t) \end{aligned}$$

The quantity $1_{y_{t<-1}}$ is null but the expression $S_v(1, y_t)$ is not defined for $y_t \geq 0$ unless we choose η such that

$$1_{y_t < -1} \times \log(-y_t) + \eta(y_t) = \rho(y_t)$$

where ρ is a function defined $\forall y_t \in R$ and $E[\rho(y_t)]$ exists. The function $\rho(x) = 0$ does the job. Thus,

$$\eta(x) = -1_{x < -1} \times \log(-x)$$

has the ability to make the function S_v of equation (A33) be defined for all y_t. Indeed, the scoring function thus obtained is:

$$\begin{aligned} S_{log}(v_t, y_t) &= (p - I_t) \log(v_t) + I_t \log(-y_t) + 1_{y_t \geq -1} \times 0 + 1_{y_t < -1} \times -\log\left(-y_t\right) \\ &= (p - I_t) \log(v_t) + I_t \log(-y_t) - 1_{y_{t<-1}} \times \log(-y_t) \end{aligned} \tag{A34}$$

For $y_t < -v_t \implies S_{log}(v_t, y_t) = (p-1)\log(v_t) + \log(-y_t)$.

For $y_t \geq -v_t \implies S_{log}(v_t, y_t) = p\log(v_t) + 0 \times \log(-y_t) - 0 \times \log(-y_t)$

$$S_{log}(v_t, y_t) = p\log(v_t)$$

To summarize:

$$S_{log}(v_t, y_t) = \begin{cases} (p-1)\log(v_t) + \log(-y_t)\,, & \text{if } y_t < -v_t \\ p\log(v_t)\,, & \text{if } y_t \geq -v_t \end{cases} \tag{A35}$$

A13. Complementary tables

Table A.3
Calculation and comparison of risk measures and moments

P	Densities Mixtures	VaR (in %)	CvaR in%)	Mean (in%)	Variance (× 100)	Skewness	Kurtosis
0.050	non-param	2.04736	2.97795	0.0524	0.0168	0.3589	9.8158
0.050	1:NO	2.07980	2.62147	0.0524	0.0168	0.0000	3.0000
0.050	1:T	1.86805	3.01294	0.0697	0.0186	0.0000	
0.050	1:EGB2	2.00674	2.89562	0.0525	0.0157	-0.0813	5.8076
0.050	2:NO	1.95397	3.11363	0.0525	0.0168	-0.1386	6.6789
0.050	2:T	2.02945	3.04197	0.0437	0.0163	-0.1544	8.3993
0.050	3:NO	2.03846	3.00451	0.0549	0.0172	0.1224	9.4321
0.050	2:SN2	1.87846	2.96880	0.0586	0.0163	0.1697	7.5293
0.050	2:SEP3	1.99293	2.97395	0.0544	0.0163	0.0051	7.1752
0.025	non-param	2.54290	3.66665				
0.025	1:NO	2.48828	2.97808				
0.025	1:T	2.51522	3.87890				
0.025	1:EGB2	2.62287	3.51175				
0.025	2:NO	2.81354	3.90424				
0.025	2:T	2.71654	3.74976				
0.025	3:NO	2.66598	3.68928				
0.025	2:SN2	2.71156	3.70573				
0.025	2:SEP3	2.66110	3.66159				
0.010	non-param	3.59575	4.51902				
0.010	1:NO	2.96323	3.40250				
0.010	1:T	3.54473	5.29712				
0.010	1:EGB2	3.43734	4.32622				
0.010	2:NO	3.89559	4.82632				
0.010	2:T	3.62577	4.72258				
0.010	3:NO	3.47885	4.71115				
0.010	2:SN2	3.72513	4.51543				
0.010	2:SEP3	3.57259	4.58396				

Table A.4
Backtests of n-prefiltered models (Gaussian)

			Standard Backtests						**Comparative Backtests**				
Models	**p**	**%**	**uc**		**cc**	**DQ**	**Z_{ES}**	**RC**	**S_{log}**		**S_{FZ}**		**DES**
	%	Viol.	Stat.	p-val	p-val	p-val	p-val	p-val	Scores	#R	Scores	#R	p-val
			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
n|1:NO	1	2.17	0.0103	0.000	0.001	0.000	0.000	0.000	-3.1255	22	-3.0380	22	0.003
n|1:T	1	2.00	0.0078	0.002	0.006	0.000	0.009	0.020	-3.1623	19	-3.0907	21	0.006
n|1:EGB2	1	1.50	0.0022	0.105	0.204	0.068	0.053	0.003	-3.1775	13	-3.1431	10	0.102
n|2:NO	1	1.42	0.0016	0.172	0.308	0.087	0.001	0.000	-3.1528	21	-3.1028	19	0.090
n|2:T	1	1.42	0.0016	0.172	0.308	0.100	0.043	0.015	-3.1614	20	-3.1131	17	0.140
n|3:NO	1	1.50	0.0022	0.105	0.204	0.036	0.003	0.002	-3.1651	17	-3.1305	14	0.131
n|2:SN2	1	1.42	0.0016	0.172	0.308	0.116	0.006	0.000	-3.1710	16	-3.1251	16	0.106
n|2:SEP3	1	1.33	0.0010	0.269	0.438	0.186	0.076	0.051	-3.2112	5	-3.1846	3	0.219
n|1:ST3	1	1.58	0.0029	0.061	0.128	0.042	0.029	0.009	-3.1643	18	-3.0937	20	0.062
n|1:SEP3	1	1.50	0.0022	0.105	0.204	0.070	0.025	0.002	-3.1775	14	-3.1311	13	0.095
n|1:GP	1	1.42	0.0016	0.172	0.308	0.120	0.048	0.003	-3.1763	15	-3.1258	15	0.109
n|1:NO	2.5	3.58	0.0043	0.024	0.073	0.006	0.000	0.002	-8.6578	22	-3.3954	22	0.008
n|1:T	2.5	3.58	0.0043	0.024	0.073	0.007	0.014	0.015	-8.6747	20	-3.4132	21	0.012
n|1:EGB2	2.5	3.42	0.0031	0.054	0.146	0.029	0.065	0.066	-8.7022	16	-3.4374	15	0.039
n|2:NO	2.5	3.25	0.0021	0.111	0.273	0.070	0.032	0.095	-8.7110	15	-3.4329	17	0.056
n|2:T	2.5	3.33	0.0026	0.078	0.202	0.084	0.058	0.044	-8.6989	19	-3.4300	19	0.089
n|3:NO	2.5	3.17	0.0017	0.155	0.357	0.034	0.010	0.005	-8.7011	17	-3.4376	14	0.051
n|2:SN2	2.5	3.17	0.0017	0.155	0.357	0.117	0.061	0.163	-8.7458	10	-3.4501	10	0.085
n|2:SEP3	2.5	2.92	0.0007	0.368	0.666	0.227	0.301	0.408	-8.7720	4	-3.4680	4	0.185
n|1:ST3	2.5	3.42	0.0031	0.054	0.138	0.014	0.026	0.010	-8.6742	21	-3.4141	20	0.020
n|1:SEP3	2.5	3.42	0.0031	0.054	0.138	0.013	0.047	0.032	-8.7002	18	-3.4323	18	0.026
n|1:GP	2.5	3.33	0.0026	0.078	0.202	0.063	0.129	0.045	-8.7401	12	-3.4476	11	0.060
n|1:NO	5	5.92	0.0017	0.156	0.338	0.182	0.002	0.008	-18.5184	10	-3.6403	20	0.039
n|1:T	5	6.00	0.0020	0.123	0.218	0.103	0.038	0.024	-18.5059	13	-3.6454	17	0.048
n|1:EGB2	5	5.92	0.0017	0.156	0.248	0.139	0.060	0.078	-18.4738	19	-3.6457	16	0.080
n|2:NO	5	5.83	0.0014	0.196	0.390	0.227	0.069	0.120	-18.4829	17	-3.6445	19	0.090
n|2:T	5	5.83	0.0014	0.196	0.277	0.109	0.046	0.043	-18.4475	22	-3.6386	21	0.076
n|3:NO	5	5.67	0.0009	0.299	0.326	0.199	0.025	0.025	-18.4929	15	-3.6491	14	0.132
n|2:SN2	5	5.75	0.0011	0.244	0.441	0.243	0.086	0.137	-18.4967	14	-3.6514	13	0.127
n|2:SEP3	5	5.92	0.0017	0.156	0.076	0.023	0.280	0.422	-18.4889	16	-3.6548	12	0.060
n|1:ST3	5	5.92	0.0017	0.156	0.248	0.134	0.032	0.022	-18.4584	21	-3.6373	22	0.061
n|1:SEP3	5	5.83	0.0014	0.196	0.277	0.163	0.080	0.036	-18.4803	18	-3.6448	18	0.081
n|1:GP	5	5.83	0.0014	0.196	0.390	0.228	0.135	0.070	-18.4657	20	-3.6463	15	0.118

Scores of S_{log} multiplied by 100 to show more digits.

Table A.5
Backtests of e-prefiltered models (Skew-t)

Models	p %	% Viol.	Standard Backtests: uc Stat. (1)	uc p-val (2)	cc p-val (3)	DQ p-val (4)	Z_{ES} p-val (5)	RC p-val (6)	Comparative Backtests: S_{log} Scores (7)	S_{log} #R (8)	S_{FZ} Scores (9)	S_{FZ} #R (10)	DES p-val (11)
e|1:NO	1	2.17	0.0103	0.000	0.001	0.000	0.017	0.000	-3.1853	12	-3.1110	18	0.007
e|1:T	1	1.92	0.0067	0.005	0.012	0.001	0.001	0.006	-3.2116	4	-3.1407	12	0.015
e|1:EGB2	1	1.42	0.0016	0.172	0.308	0.183	0.092	0.019	-3.2178	2	-3.1857	2	0.185
e|2:NO	1	1.42	0.0016	0.172	0.308	0.123	0.038	0.003	-3.1934	11	-3.1571	9	0.126
e|2:T	1	1.42	0.0016	0.172	0.308	0.135	0.016	0.009	-3.2030	9	-3.1691	6	0.156
e|3:NO	1	1.42	0.0016	0.172	0.308	0.128	0.034	0.008	-3.2027	10	-3.1584	8	0.156
e|2:SN2	1	1.42	0.0016	0.172	0.308	0.385	0.054	0.006	-3.2069	8	-3.1696	5	0.260
e|2:SEP3	1	0.92	0.0001	0.769	0.865	0.802	0.579	0.979	-3.2707	1	-3.2513	1	0.701
e|1:ST3	1	1.58	0.0029	0.061	0.128	0.054	0.001	0.026	-3.2089	7	-3.1416	11	0.078
e|1:SEP3	1	1.50	0.0022	0.105	0.204	0.098	0.170	0.022	-3.2153	3	-3.1720	4	0.126
e|1:GP	1	1.42	0.0016	0.172	0.308	0.160	0.094	0.014	-3.2094	6	-3.1677	7	0.143
e|1:NO	2.5	3.58	0.0043	0.024	0.043	0.006	0.009	0.001	-8.7292	14	-3.4333	16	0.015
e|1:T	2.5	3.42	0.0031	0.054	0.138	0.030	0.002	0.018	-8.7432	11	-3.4435	12	0.030
e|1:EGB2	2.5	3.25	0.0021	0.111	0.273	0.097	0.104	0.082	-8.7565	7	-3.4628	6	0.097
e|2:NO	2.5	3.17	0.0017	0.155	0.357	0.139	0.106	0.154	-8.7571	6	-3.4592	7	0.115
e|2:T	2.5	3.25	0.0021	0.111	0.273	0.105	0.068	0.192	-8.7484	9	-3.4590	8	0.110
e|3:NO	2.5	3.00	0.0010	0.282	0.559	0.155	0.138	0.143	-8.7684	5	-3.4648	5	0.117
e|2:SN2	2.5	3.00	0.0010	0.282	0.559	0.438	0.148	0.120	-8.7860	3	-3.4705	2	0.260
e|2:SEP3	2.5	2.50	0.0000	1.000	0.463	0.373	0.747	0.809	-8.7904	1	-3.4860	1	0.320
e|1:ST3	2.5	3.33	0.0026	0.078	0.182	0.053	0.001	0.012	-8.7369	13	-3.4429	13	0.056
e|1:SEP3	2.5	3.25	0.0021	0.111	0.273	0.081	0.202	0.041	-8.7542	8	-3.4579	9	0.079
e|1:GP	2.5	3.25	0.0021	0.111	0.273	0.252	0.217	0.089	-8.7893	2	-3.4702	3	0.168
e|1:NO	5	6.08	0.0023	0.095	0.121	0.063	0.014	0.002	-18.5810	1	-3.6606	7	0.030
e|1:T	5	6.08	0.0023	0.095	0.121	0.061	0.006	0.084	-18.5654	3	-3.6624	3	0.048
e|1:EGB2	5	6.25	0.0031	0.055	0.052	0.027	0.084	0.121	-18.5328	6	-3.6621	5	0.046
e|2:NO	5	6.00	0.0020	0.123	0.071	0.036	0.144	0.203	-18.5333	5	-3.6609	6	0.064
e|2:T	5	6.08	0.0023	0.095	0.065	0.033	0.089	0.234	-18.5083	12	-3.6584	10	0.065
e|3:NO	5	5.83	0.0014	0.196	0.277	0.176	0.124	0.189	-18.5174	11	-3.6603	9	0.135
e|2:SN2	5	5.92	0.0017	0.156	0.150	0.107	0.140	0.163	-18.5322	7	-3.6626	2	0.121
e|2:SEP3	5	5.00	0.0000	1.000	1.000	0.432	0.406	0.550	-18.5713	2	-3.6757	1	0.300
e|1:ST3	5	6.00	0.0020	0.123	0.218	0.163	0.001	0.048	-18.5207	8	-3.6551	11	0.083
e|1:SEP3	5	5.92	0.0017	0.156	0.248	0.150	0.230	0.060	-18.5452	4	-3.6623	4	0.087
e|1:GP	5	6.00	0.0020	0.123	0.136	0.101	0.187	0.120	-18.5187	9	-3.6606	8	0.108

Scores of S_{log} multiplied by 100 to show more digits.