CAESar:
Conditional Autoregressive Expected Shortfall

Federico Gatta[1], Fabrizio Lillo[1,2], Piero Mazzarisi[3]

[1]Scuola Normale Superiore, [2]Università di Bologna,
[3]Università di Siena

federico.gatta@sns.it

SCUOLA
NORMALE
SUPERIORE

Summary

1. Expected Shortfall

2. CAESar

3. Experimental Stage

4. Conclusion

Expected Shortfall

Framework

Let's define:

- $\boldsymbol{y} = \{y_t\}_{t=1}^{T}$ the asset return time series;
- $F_{t-1}(\cdot) = F(\cdot|\mathcal{I}_{t-1})$ and $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{I}_{t-1}]$ the distribution function and the expectation conditional on the information set at time $t-1$, $\mathcal{I}_{t-1}$.

Let θ be the target level of confidence. The main measures used to summarize financial risk are:

Framework

Let's define:

- $\boldsymbol{y} = \{y_t\}_{t=1}^{T}$ the asset return time series;
- $F_{t-1}(\cdot) = F(\cdot|\mathcal{I}_{t-1})$ and $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{I}_{t-1}]$ the distribution function and the expectation conditional on the information set at time $t-1$, $\mathcal{I}_{t-1}$.

Let θ be the target level of confidence. The main measures used to summarize financial risk are:

- **Value at Risk (VaR)** - It is defined as the θ quantile of the conditional distribution:

$$VaR_t(\theta) = \inf\{F_{t-1}^{-1}([\theta, 1])\}$$

- **Expected Shortfall (ES)** - It is defined as the conditional tail expectation, that is:

$$ES_t(\theta) = \mathbb{E}[y_t|\mathcal{I}_{t-1}, y_t \leq VaR_t(\theta)]$$

Noticeable probability levels θ are $0.05, 0.025, 0.01$.

VaR and ES - Critical Points

From VaR to ES

So far, VaR has been the most popular approach. However, the regulator is moving towards the ES, as stated in the Basel III regulatory framework for banks[1]. Why?

- ES better captures negative events as it contains information on the whole tail.
- It has nicer mathematical properties. One for all, it is subadditive.

[1] https://www.bis.org/bcbs/basel3.htm

VaR and ES - Critical Points

From VaR to ES

So far, VaR has been the most popular approach. However, the regulator is moving towards the ES, as stated in the Basel III regulatory framework for banks[1]. Why?

- ES better captures negative events as it contains information on the whole tail.
- It has nicer mathematical properties. One for all, it is subadditive.

ES Drawback

The ES is not *elicitable* - Gneiting (2011). That is, there does not exist a loss function $\mathcal{L}$ such that the expected loss under a distribution takes its unique minimum at the functional value of this distribution.
On the contrary, the VaR is elicitable. A popular loss function is the *Quantile Loss*:

$$\mathcal{L}_q^{\theta}(\hat{\boldsymbol{q}}, \boldsymbol{y}) := \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{q}_t)\left(\theta - \mathbb{1}_{\{y_t < \hat{q}_t\}}\right) \tag{1}$$

where $\hat{\boldsymbol{q}} = \{\hat{q}_t\}_{t=1}^{T}$ is the $\boldsymbol{VaR}(\theta) = \{VaR_t(\theta)\}_{t=1}^{T}$ predictions time series.

[1] `https://www.bis.org/bcbs/basel3.htm`

VaR and ES Joint Elicitability

ES and VaR are jointly elicitable - Fissler and Ziegel (2016).

Noteworthy contributions are:

VaR and ES Joint Elicitability

ES and VaR are jointly elicitable - Fissler and Ziegel (2016).
Noteworthy contributions are:

Barrera et al. (2022)

The estimator $\hat{\boldsymbol{r}}$ of the time series difference $\boldsymbol{r} := \boldsymbol{ES} - \boldsymbol{VaR}$ is in the argmin (fixed the hypothesis class $\mathcal{R}$) of the following loss function:

$$\mathcal{L}_B^{\theta}(\hat{\boldsymbol{r}}, \boldsymbol{y}; \boldsymbol{VaR}(\theta)) := \frac{1}{T} \sum_{t=1}^{T} \left(\hat{r}_t + \frac{1}{\theta} \left(VaR_t(\theta) - y_t \right)^+ \right)^2 \tag{2}$$

Intuition: $\mathcal{L}_B^{\theta}$ is the mean square error between $\hat{\boldsymbol{r}}$ and the "excess loss" $\left(\boldsymbol{VaR} - \boldsymbol{y}\right)^+$.

VaR and ES Joint Elicitability

ES and VaR are jointly elicitable - Fissler and Ziegel (2016).
Noteworthy contributions are:

Barrera et al. (2022)

The estimator $\hat{\boldsymbol{r}}$ of the time series difference $\boldsymbol{r} := \boldsymbol{ES} - \boldsymbol{VaR}$ is in the argmin (fixed the hypothesis class $\mathcal{R}$) of the following loss function:

$$\mathcal{L}_B^{\theta}(\hat{\boldsymbol{r}}, \boldsymbol{y}; \boldsymbol{VaR}(\theta)) := \frac{1}{T} \sum_{t=1}^{T} \left(\hat{r}_t + \frac{1}{\theta} \left(VaR_t(\theta) - y_t \right)^+ \right)^2 \tag{2}$$

Intuition: $\mathcal{L}_B^{\theta}$ is the mean square error between $\hat{\boldsymbol{r}}$ and the "excess loss" $\left(\boldsymbol{VaR} - \boldsymbol{y}\right)^+$.

Patton, Ziegel, and Chen (2019)

The estimators $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{e}} = \{\hat{e}_t\}_{t=1}^{T}$, for the time series $\boldsymbol{VaR}$ and $\boldsymbol{ES}$, are in the argmin of the following:

$$\mathcal{L}_P^{\theta}(\hat{\boldsymbol{e}}, \hat{\boldsymbol{q}}, \boldsymbol{y}) := \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\hat{q}_t}{\hat{e}_t} - \frac{\hat{q}_t - y_t}{\theta \hat{e}_t} \mathbb{1}_{\{y_t \leq \hat{q}_t\}} + \log\left(-\hat{e}_t\right) \right) \tag{3}$$

Intuition: $\mathcal{L}_P^{\theta}$ is a special case of the jointly elicitable losses discussed in Fissler and Ziegel (2016).

CAESar

Model Description

CAViaR Regression - Engle and Manganelli (2004)

CAViaR is an autoregressive approach for VaR forecasting. Its most popular specification is the Asymmetric Slope (AS):

$$\hat{q}_t = \beta_0 + \beta_1 (y_{t-1})^+ + \beta_2 (y_{t-1})^- + \beta_3 \hat{q}_{t-1} \tag{4}$$

where $(y_{t-1})^+$ and $(y_{t-1})^-$ are the positive and negative parts of y_{t-1}.

CAESar Estimator

Our proposal is to generalize the CAViaR regression to handle ES. Specifically, the quantile estimator becomes:

$$\hat{q}_t = \beta_0 + \beta_1 (y_{t-1})^+ + \beta_2 (y_{t-1})^- + \beta_3 \hat{q}_{t-1} + \beta_4 \hat{e}_{t-1} \tag{5}$$

As for the Expected Shortfall:

$$\hat{e}_t = \gamma_0 + \gamma_1 (y_{t-1})^+ + \gamma_2 (y_{t-1})^- + \gamma_3 \hat{q}_{t-1} + \gamma_4 \hat{e}_{t-1} \tag{6}$$

Motivation

Autoregressive approaches for ES estimators have already been proposed:

- Generalized Autoregressive Score (GAS) models in Patton, Ziegel, and Chen (2019). Two formulations are proposed: with one (**GAS1**) and two latent factors (**GAS2**).
- Asymmetric Laplace-based models in Taylor (2019): the VaR estimator has the same functional form as the CAViaR. The loss function is derived from the Asymmetric Laplace log-likelihood. Two formulations are proposed: ES multiple of VaR (**AL-Multi**); ES-VaR modeled as an AR process (**AL-AR**).

Why should CAESar overcome these models?

We expect the CAESar formulation to be more general and less limiting:

- **GAS1** and **AL-Multi assume the ratio** $\frac{Var_t}{ES_t}$ **is constant**. However, the ratio can change based on the evolution of the returns' conditional distribution shape, leading to potential model misspecification errors.
- **GAS2** and **AL-AR** address this issue, but they **only considers part of the information set**. Indeed, information on the magnitude of y_t is used only when it breaks the quantile. Specifically, our model proves to be faster in adjusting forecasts within regime-switch frameworks.
- As stated by the authors, GAS model optimization is **sensitive to starting coefficients**.

Optimization Approach

CAESar parameters optimization consists of three steps:

- Step 1: CAViaR for VaR Estimate
- Parameters Rough Approximation via Barrera Approach
- Ultimate Approximation via Patton Approach

As for the first step, we need a reliable VaR estimate. CAViaR is used. The parameters $\beta_0, \cdots, \beta_3$ minimize the $\mathcal{L}_q^\theta$ (Pinball) loss function. Specifically, 100 random parameter vectors are sampled, and out of them, the three corresponding to the lowest loss are used as starting points for the optimization. After optimization, the most promising parameters vector is used as the $\beta_0, \cdots, \beta_3$ estimate.

Optimization Approach

CAESar approach consists of three steps:

- CAViaR for VaR Estimate
- Step 2: Parameters Rough Approximation via Barrera Approach
- Ultimate Approximation via Patton Approach

As for the first approximation, we use $\hat{e}_t = \hat{q}_t + \hat{r}_t$ with:

$$\hat{r}_t = \tilde{\gamma_0} + \tilde{\gamma_1}(y_{t-1})^+ + \tilde{\gamma_2}(y_{t-1})^- + \tilde{\gamma_3}\hat{q}_{t-1} + \tilde{\gamma_4}\hat{r}_{t-1} \tag{7}$$

The parameters $\tilde{\gamma_0}, \cdots, \tilde{\gamma_4}$ minimize the following loss function:

$$\mathcal{L}_r^\theta(\hat{\boldsymbol{r}}, \boldsymbol{y}; \boldsymbol{VaR}(\theta)) := \mathcal{L}_B^\theta(\hat{\boldsymbol{r}}, \boldsymbol{y}; \boldsymbol{VaR}(\theta)) + \lambda_r \sum_{t=1}^{T} (r_t)^+ \tag{8}$$

where a soft threshold has been added to the loss in Barrera to ensure the *monotonicity condition*, i.e., $\hat{e}_t < \hat{q}_t$. λ_r has been set equal to 10. Finally, the $\gamma_0, \cdots, \gamma_4$ coefficients can easily be recovered as:

$$\gamma_j = \tilde{\gamma}_j + \beta_j \; j = 0, 1, 2; \qquad \gamma_3 = \tilde{\gamma_3} + \beta_3 - \tilde{\gamma_4}; \qquad \gamma_4 = \tilde{\gamma_4} \tag{9}$$

Optimization Approach

CAESar approach consists of three steps:

- CAViaR for VaR Estimate
- Parameters Rough Approximation via Barrera Approach
- Step 3: Ultimate Approximation via Patton Approach

The ultimate step is the joint optimization of $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{e}}$. Namely, the two estimators are:

$$\begin{bmatrix} \hat{q}_t \\ \hat{e}_t \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} (y_{t-1})^+ \\ (y_{t-1})^- \end{bmatrix} + \begin{bmatrix} \beta_3 & \beta_4 \\ \gamma_3 & \gamma_4 \end{bmatrix} \begin{bmatrix} \hat{q}_{t-1} \\ \hat{e}_{t-1} \end{bmatrix} \tag{10}$$

As the starting point, we use the coefficients estimated in the previous steps. As before, soft constraints are added to ensure the monotonicity condition:

$$\mathcal{L}^{\theta}_{q,e}(\hat{\boldsymbol{e}}, \hat{\boldsymbol{q}}, \boldsymbol{y}) := \mathcal{L}^{\theta}_{P}(\hat{\boldsymbol{e}}, \hat{\boldsymbol{q}}, \boldsymbol{y}) + \lambda_e \sum_{t=1}^{T} (\hat{e}_t - \hat{q}_t)^+ + \lambda_q \sum_{t=1}^{T} (\hat{q}_t)^+ \tag{11}$$

Experimental Stage

The Experimental Framework

The Synthetic Dataset

GARCH data-generating process; separate experiments with Gaussian and Student's t innovations. For each innovation distribution, 3 sets of coefficients. For each set, 20 time series, 7 years long.

The Experimental Framework

The Synthetic Dataset

GARCH data-generating process; separate experiments with Gaussian and Student's t innovations. For each innovation distribution, 3 sets of coefficients. For each set, 20 time series, 7 years long.

The Index Dataset

A real-world dataset of daily observations from 10 stock indices has been used (SPX, FTSE, DAX, CAC, NIKKEI, IBEX, MXX, BOVESPA, KOSPI, HSI).
The period covered is from 1993-07-01 to 2023-06-30. *Block cross-validation* has been used: 6 years for train and validation, 1 for test. That is, 24 folds.

The Experimental Framework

The Synthetic Dataset

GARCH data-generating process; separate experiments with Gaussian and Student's t innovations. For each innovation distribution, 3 sets of coefficients. For each set, 20 time series, 7 years long.

The Index Dataset

A real-world dataset of daily observations from 10 stock indices has been used (SPX, FTSE, DAX, CAC, NIKKEI, IBEX, MXX, BOVESPA, KOSPI, HSI).
The period covered is from 1993-07-01 to 2023-06-30. *Block cross-validation* has been used: 6 years for train and validation, 1 for test. That is, 24 folds.

The Stock Dataset

The dataset is made up of daily observations of 8 stocks in the US banking sector (BAC, BK, C, GS, JPM, MS, STT, WFC).
The period covered is from 2000-01-01 to 2023-12-31. *Block cross-validation* as before; 18 folds.

The Experimental Framework

The Synthetic Dataset

GARCH data-generating process; separate experiments with Gaussian and Student's t innovations. For each innovation distribution, 3 sets of coefficients. For each set, 20 time series, 7 years long.

The Index Dataset

A real-world dataset of daily observations from 10 stock indices has been used (SPX, FTSE, DAX, CAC, NIKKEI, IBEX, MXX, BOVESPA, KOSPI, HSI).
The period covered is from 1993-07-01 to 2023-06-30. *Block cross-validation* has been used: 6 years for train and validation, 1 for test. That is, 24 folds.

The Stock Dataset

The dataset is made up of daily observations of 8 stocks in the US banking sector (BAC, BK, C, GS, JPM, MS, STT, WFC).
The period covered is from 2000-01-01 to 2023-12-31. *Block cross-validation* as before; 18 folds.

Experiments have been carried out with probability levels $\theta = 0.05$, $\theta = 0.025$, and $\theta = 0.01$.

The Experimental Framework

The Competing Models

- **BCGNS** The neural network developed in Barrera et al. (2022).

The Experimental Framework

The Competing Models

- **BCGNS** The neural network developed in Barrera et al. (2022).
- **GAS1** The Generalized Autoregressive Score model with one latent factor - Patton, Ziegel, and Chen (2019).
- **GAS2** The GAS model with two latent factors.

The Experimental Framework

The Competing Models

- **BCGNS** The neural network developed in Barrera et al. (2022).
- **GAS1** The Generalized Autoregressive Score model with one latent factor - Patton, Ziegel, and Chen (2019).
- **GAS2** The GAS model with two latent factors.
- **AL-Multi** The AL-based approach with ES multiple of VaR - Taylor (2019).
- **AL-AR** AL-based approach with ES-VaR modeled as an AR process.

The Experimental Framework

The Competing Models

- **BCGNS** The neural network developed in Barrera et al. (2022).
- **GAS1** The Generalized Autoregressive Score model with one latent factor - Patton, Ziegel, and Chen (2019).
- **GAS2** The GAS model with two latent factors.
- **AL-Multi** The AL-based approach with ES multiple of VaR - Taylor (2019).
- **AL-AR** AL-based approach with ES-VaR modeled as an AR process.
- **K-CAViaR** Following Emmer, Kratz, and Tasche (2013), we construct an $ES(\theta)$ estimator by averaging the CAViaR predictions for $n = 10$ equi-spaced quantiles $VaR(\theta_j)$ $j = 1, \cdots, n$ with $0 < \theta_1 < \cdots < \theta_n = \theta$.
- **K-QRNN** The approach in Emmer, Kratz, and Tasche (2013), where the quantile prediction is provided by a Quantile Regression Neural Network (QRNN). One single network is trained, with n outputs, one for each quantile $VaR(\theta_j)$ $j = 1, \cdots, n$.

Results Synthetic Dataset - Root Mean Squared Error

$\theta=0.05$	N - I	N - II	N - III	t - I	t - II	t - III
CAESar	0.002	**0.002**	0.002	0.342	0.028	0.008
K-CAViaR	**0.001**	0.002	**0.001**	0.132	0.034	0.01
BCGNS	0.03	0.033	0.036	0.829	0.145	0.054
K-QRNN	0.008	0.009	0.007	0.937	0.177	0.037
GAS1	0.089	0.029	0.081	0.35	0.068	0.022
GAS2	0.02	0.247	0.004	0.129	0.072	0.039
AL-Multi	0.063	0.213	0.072	**0.108**	**0.019**	**0.006**
AL-AR	0.025	0.021	0.058	0.228	0.038	0.019

$\theta=0.025$	N - I	N - II	N - III	t - I	t - II	t - III
CAESar	**0.002**	**0.002**	**0.002**	0.229	0.038	0.011
K-CAViaR	0.002	0.002	0.002	0.193	0.044	0.013
BCGNS	0.018	0.018	0.015	0.876	0.147	0.044
K-QRNN	0.01	0.01	0.008	1.065	0.225	0.048
GAS1	0.013	0.049	0.012	**0.108**	0.088	0.036
GAS2	0.077	0.196	0.229	0.384	0.214	0.096
AL-Multi	0.088	0.142	0.2	0.203	**0.026**	**0.009**
AL-AR	0.017	0.049	0.014	0.162	0.053	0.023

$\theta=0.01$	N - I	N - II	N - III	t - I	t - II	t - III
CAESar	**0.003**	**0.004**	**0.003**	0.286	0.062	0.017
K-CAViaR	0.003	0.004	0.003	0.334	0.067	0.024
BCGNS	0.012	0.013	0.010	1.018	0.148	0.04
K-QRNN	0.013	0.013	0.010	1.33	0.257	0.055
GAS1	0.026	0.058	0.020	**0.131**	0.191	0.058
GAS2	0.458	0.333	0.198	0.498	0.369	0.266
AL-Multi	0.017	0.143	0.013	0.223	**0.041**	**0.014**
AL-AR	0.076	0.053	0.141	0.303	0.057	0.049

Results Index Dataset - Patton Loss Function

$\theta=0.05$	SPX	MXX	BOVESPA	CAC	KOSPI	IBEX	NIKKEI	HSI	DAX	FTSE
CAESar	0.847	**0.901**	**1.25**	0.985	**1.01**	1.074	**1.061**	**1.05**	**0.998**	0.797
K-CAViaR	**0.844**	0.908	1.286	**0.978**	1.054	**1.039**	1.092	1.061	1.005	**0.792**
BCGNS	2.44	2.571	8.451	5.984	2.204	2.168	4.249	3.126	15.583	2.516
K-QRNN	1.152	1.071	1.427	1.303	1.218	1.26	1.23	1.282	1.339	1.104
GAS1	1.216	0.985	1.31	1.72	1.154	1.296	1.166	1.108	1.071	0.944
GAS2	1.313	1.718	1.656	1.672	1.775	2.065	1.568	2.484	1.957	1.484
AL-Multi	1.171	1.573	1.472	1.522	1.353	1.556	1.601	1.232	1.215	1.789
AL-AR	1.308	1.084	1.581	0.979	1.14	2.153	1.4	1.7	2.229	2.404

$\theta=0.025$	SPX	MXX	BOVESPA	CAC	KOSPI	IBEX	NIKKEI	HSI	DAX	FTSE
CAESar	1.064	**1.096**	1.475	**1.148**	1.275	**1.191**	**1.243**	**1.255**	**1.169**	0.965
K-CAViaR	**1.06**	1.119	1.487	1.151	**1.27**	1.191	1.268	1.261	1.172	**0.96**
BCGNS	5.263	2.218	2.536	2.023	2.523	1.97	4.342	1.84	3.0	2.807
K-QRNN	1.466	1.294	1.693	1.518	1.492	1.492	1.467	1.517	1.507	1.379
GAS1	1.25	1.328	1.886	1.415	1.401	1.376	1.505	1.443	1.636	1.386
GAS2	2.65	2.288	1.696	1.975	2.451	1.551	1.887	2.303	1.834	2.187
AL-Multi	1.591	1.277	**1.429**	1.424	1.596	1.366	1.486	1.598	1.546	2.192
AL-AR	2.702	1.381	1.561	1.467	1.791	1.468	1.878	2.176	1.683	1.886

$\theta=0.01$	SPX	MXX	BOVESPA	CAC	KOSPI	IBEX	NIKKEI	HSI	DAX	FTSE
CAESar	**1.427**	**1.376**	**1.652**	1.394	**1.493**	**1.389**	**1.482**	1.476	1.461	1.191
K-CAViaR	1.554	1.412	1.756	**1.387**	1.541	1.394	1.499	**1.433**	**1.441**	**1.172**
BCGNS	3.48	2.262	2.335	1.993	2.539	2.161	2.603	2.438	2.222	2.131
K-QRNN	1.815	1.508	2.049	1.77	1.782	1.735	1.752	1.811	1.717	1.659
GAS1	1.443	1.78	2.046	2.103	1.774	1.82	2.203	2.246	1.612	1.555
GAS2	2.312	1.84	2.026	1.903	2.297	1.766	2.235	1.962	1.9	2.153
AL-Multi	2.32	1.682	1.888	1.646	1.604	1.549	1.887	1.487	1.791	1.291
AL-AR	4.079	1.545	1.983	2.195	2.133	2.831	1.563	1.489	2.608	2.287

Cross-dependency between $\hat{q}_t$ and $\hat{e}_t$

Remember the CAESar specification:

$$\begin{bmatrix} \hat{q}_t \\ \hat{e}_t \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} (y_{t-1})^+ \\ (y_{t-1})^- \end{bmatrix} + \begin{bmatrix} \beta_3 & \beta_4 \\ \gamma_3 & \gamma_4 \end{bmatrix} \begin{bmatrix} \hat{q}_{t-1} \\ \hat{e}_{t-1} \end{bmatrix}$$

we assess the importance of the cross terms β_4 and γ_3 via an ablation study:

Cross-dependency between $\hat{q}_t$ and $\hat{e}_t$

Remember the CAESar specification:

$$\begin{bmatrix} \hat{q}_t \\ \hat{e}_t \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} (y_{t-1})^+ \\ (y_{t-1})^- \end{bmatrix} + \begin{bmatrix} \beta_3 & \beta_4 \\ \gamma_3 & \gamma_4 \end{bmatrix} \begin{bmatrix} \hat{q}_{t-1} \\ \hat{e}_{t-1} \end{bmatrix}$$

we assess the importance of the cross terms β_4 and γ_3 via an ablation study:

$$\begin{bmatrix} \hat{q}_t^{NC} \\ \hat{e}_t^{NC} \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} (y_{t-1})^+ \\ (y_{t-1})^- \end{bmatrix} + \begin{bmatrix} \beta_3 & 0 \\ 0 & \gamma_4 \end{bmatrix} \begin{bmatrix} \hat{q}_{t-1}^{NC} \\ \hat{e}_{t-1}^{NC} \end{bmatrix}$$

Cross-dependency between $\hat{q}_t$ and $\hat{e}_t$

Remember the CAESar specification:

$$\begin{bmatrix} \hat{q}_t \\ \hat{e}_t \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} (y_{t-1})^+ \\ (y_{t-1})^- \end{bmatrix} + \begin{bmatrix} \beta_3 & \beta_4 \\ \gamma_3 & \gamma_4 \end{bmatrix} \begin{bmatrix} \hat{q}_{t-1} \\ \hat{e}_{t-1} \end{bmatrix}$$

we assess the importance of the cross terms β_4 and γ_3 via an ablation study:

$$\begin{bmatrix} \hat{q}^{NC}_t \\ \hat{e}^{NC}_t \end{bmatrix} = \begin{bmatrix} \beta_0 \\ \gamma_0 \end{bmatrix} + \begin{bmatrix} \beta_1 & \beta_2 \\ \gamma_1 & \gamma_2 \end{bmatrix} \begin{bmatrix} (y_{t-1})^+ \\ (y_{t-1})^- \end{bmatrix} + \begin{bmatrix} \beta_3 & 0 \\ 0 & \gamma_4 \end{bmatrix} \begin{bmatrix} \hat{q}^{NC}_{t-1} \\ \hat{e}^{NC}_{t-1} \end{bmatrix}$$

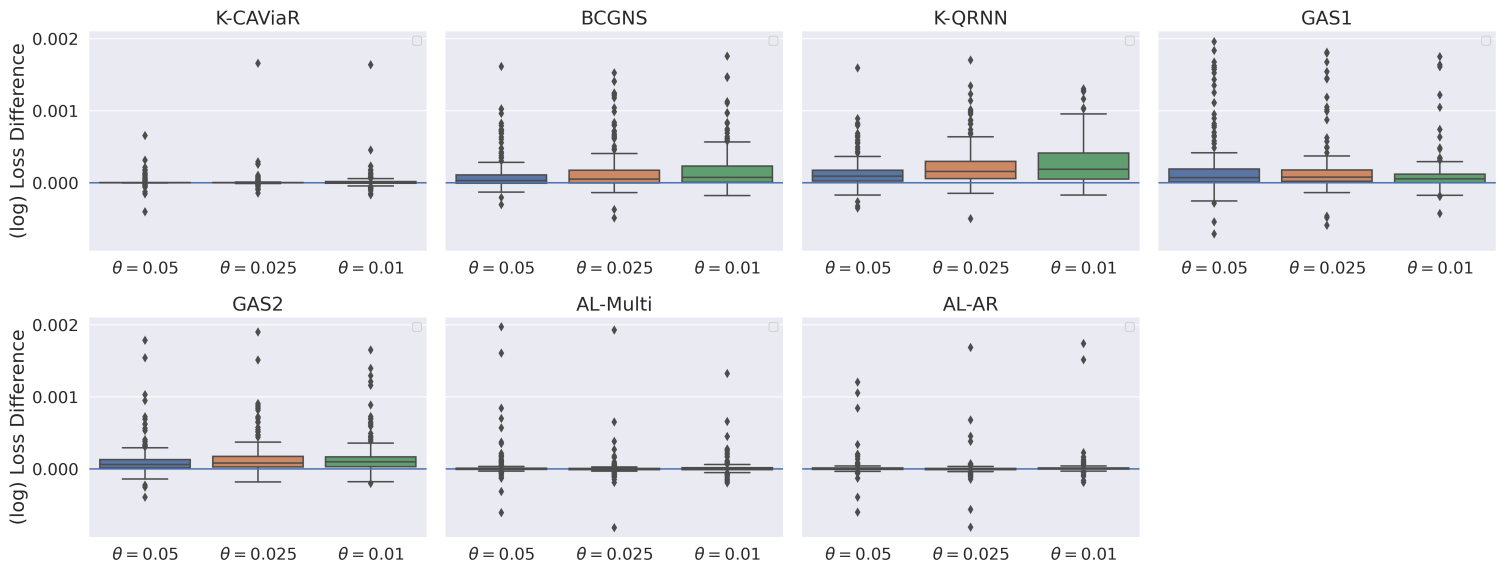
The loss comparison shows that the ”full” specification roughly outperforms the other. This suggests that the intrinsic relationship between VaR and ES is not merely related to elicitability, suggesting a kind of Granger causality between these variables.

Results - Quantile Loss Function

It is interesting to understand what happens to the quantile estimate when we update the $\hat{q}_t$ parameters in the third step of the CAESar optimization procedure. Specifically, a degradation in the VaR forecast capability could undermine CAESar's reliability. Actually, we find out this is not the case. To answer this question, we compare the Pinball loss from CAESar and competing approaches.

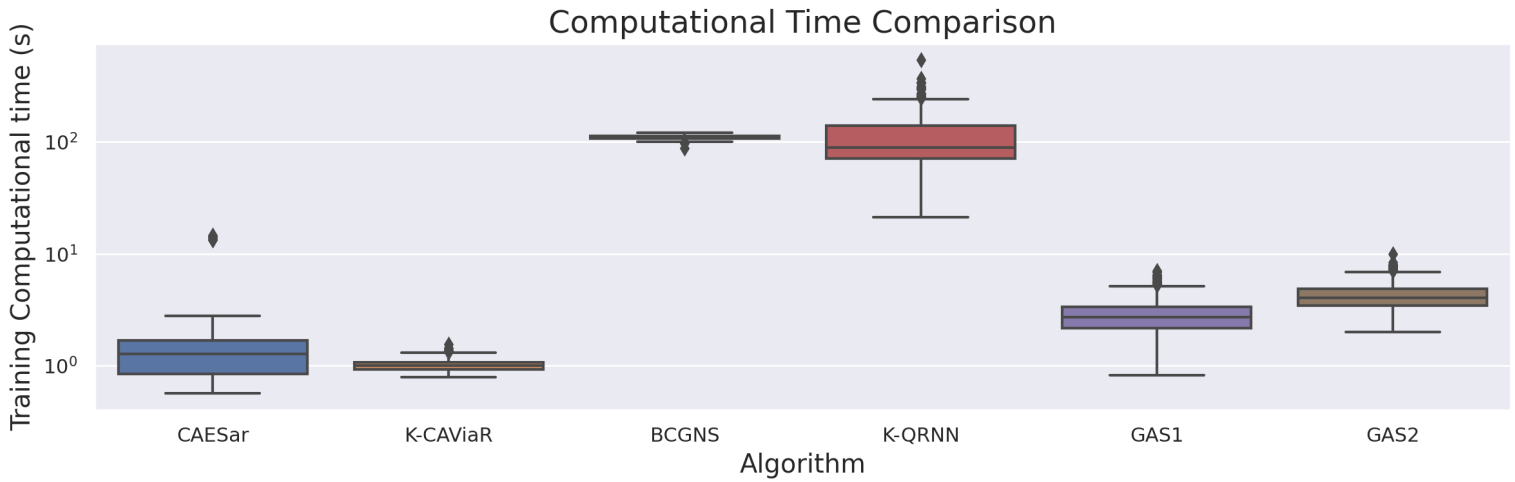
Results - Quantile Loss Function

It is interesting to understand what happens to the quantile estimate when we update the $\hat{q}_t$ parameters in the third step of the CAESar optimization procedure. Specifically, a degradation in the VaR forecast capability could undermine CAESar's reliability. Actually, we find out this is not the case. To answer this question, we compare the Pinball loss from CAESar and competing approaches.



Results - Computational Time

The average computational times of the different methods. The times are computed on a Ubuntu notebook equipped with a *13th Gen Intel Core™ i7-13620H x 16* processor and a *NVIDIA GeForce RTX 4060* video card.



Statistical Tests - Overview

Statistical tests for ES can be divided into two main categories:

Model Comparison

Tests used for comparing two ES estimators. Usually, they follow the structure:

- The null hypothesis is that the estimators have equal predictive accuracy.
- The alternative hypothesis is that one of them statistically outperforms the other. Thus, this kind of test is quite conservative.
- These tests are target-agnostic, as they only take into account the loss function.

Noticeable examples: Diebold-Mariano; Encompassing; Corrected resampled t-test.

Statistical Tests - Overview

Statistical tests for ES can be divided into two main categories:

Model Comparison

Tests used for comparing two ES estimators. Usually, they follow the structure:

- The null hypothesis is that the estimators have equal predictive accuracy.
- The alternative hypothesis is that one of them statistically outperforms the other. Thus, this kind of test is quite conservative.
- These tests are target-agnostic, as they only take into account the loss function.

Noticeable examples: Diebold-Mariano; Encompassing; Corrected resampled t-test.

Direct Approximation

Tests that directly evaluate the ES approximation. That is, if the dynamic provided by the estimator is coherent with the ES definition. The structure followed is:

- The null hypothesis is that the estimator correctly fits the ES.
- The alternative hypothesis is that the estimator cannot replicate the ES dynamic.
- These tests are loss-agnostic, as they are obtained by rearranging the ES definition.

Noticeable examples: McNeil and Frey; Acerbi and Szekely Z statistics.

Conclusion

Findings Summary

We propose CAESar, a three-step approach for ES estimation in a dynamic context. The idea is to use an autoregressive formulation together with the joint elicitability VaR-ES.

After the experimental stage, some patterns emerge:

- Conditional Autoregressive approaches (CAESar and K-CAViaR) show the best performances. Specifically, **CEASar outperforms** K-CAViaR.

Findings Summary

We propose CAESar, a three-step approach for ES estimation in a dynamic context. The idea is to use an autoregressive formulation together with the joint elicitability VaR-ES.

After the experimental stage, some patterns emerge:

- Conditional Autoregressive approaches (CAESar and K-CAViaR) show the best performances. Specifically, **CEASar outperforms** K-CAViaR.

- The **goodness** of the fit seems to **decrease as** θ. The most extreme and unpredictable events in the deep tail are mitigated by few values in the shallow tail.

Findings Summary

We propose CAESar, a three-step approach for ES estimation in a dynamic context. The idea is to use an autoregressive formulation together with the joint elicitability VaR-ES.

After the experimental stage, some patterns emerge:

- Conditional Autoregressive approaches (CAESar and K-CAViaR) show the best performances. Specifically, **CEASar outperforms** K-CAViaR.

- The **goodness** of the fit seems to **decrease as** θ. The most extreme and unpredictable events in the deep tail are mitigated by few values in the shallow tail.

- The effectiveness of the model design has been verified through **ablation** studies:
  - Model specification.
  - Number of lags.
  - Optimization scheme.
  - Soft-constraints in the loss function for monotonicity $\hat{e}_t \leq \hat{q}_t$.

References I

Barrera, D et al. (2022). “Learning value-at-risk and expected shortfall”. In: *arXiv preprint arXiv:2209.06476*.

Emmer, Susanne, Marie Kratz, and Dirk Tasche (2013). “What is the best risk measure in practice? A comparison of standard measures”. In: *arXiv preprint arXiv:1312.1645*.

Engle, Robert F and Simone Manganelli (2004). “CAViaR: Conditional autoregressive value at risk by regression quantiles”. In: *Journal of business & economic statistics* 22.4, pp. 367–381.

Fissler, Tobias and Johanna F Ziegel (2016). “Higher order elicitability and Osband’s principle”. In.

Gneiting, Tilmann (2011). “Making and evaluating point forecasts”. In: *Journal of the American Statistical Association* 106.494, pp. 746–762.

Patton, Andrew J, Johanna F Ziegel, and Rui Chen (2019). “Dynamic semiparametric models for expected shortfall (and value-at-risk)”. In: *Journal of econometrics* 211.2, pp. 388–413.

References II

Taylor, James W (2019). "Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution". In: *Journal of Business & Economic Statistics* 37.1, pp. 121–133.

Thanks for your attention!