Marginal-Copula-Scores for Multivariate Forecasting Evaluation

Florian Ziel
University of Duisburg-Essen

February 28, 2019

UNIVERSITÄT DUISBURG ESSEN
Open-Minded

Motivation: forecasting and evaluation

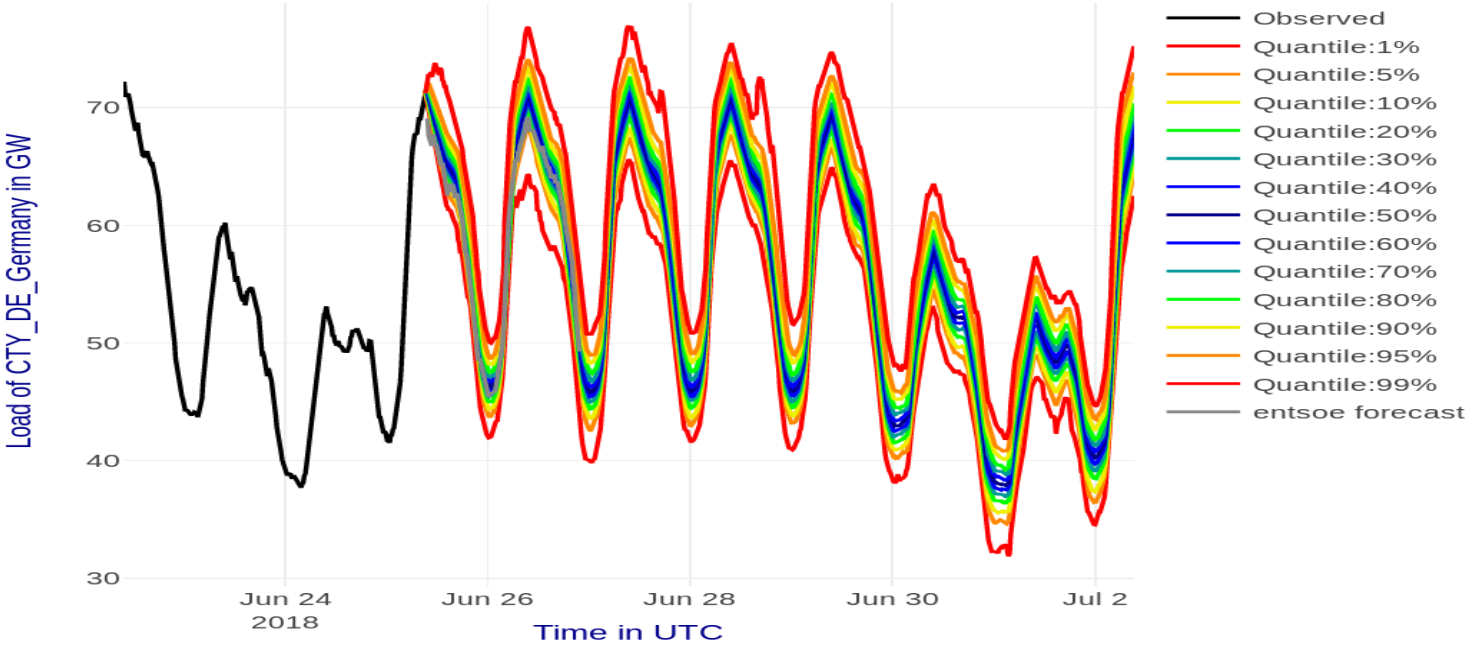
Standard setting:

- given historic data (and a statistical model) creating a forecast for a target of interest
- target of interest $\boldsymbol{Y} \sim \boldsymbol{F_Y}$, H-dimensional random variable
  e.g. maximum load of tomorrow and day after tomorrow (2-dim.)
- in practice we never know the true $\boldsymbol{F_Y}$ we just observe $\boldsymbol{y}$
- if we have a *forecast* we can only compare the performance by comparing it with $\boldsymbol{y}$
- evaluation relies on some repeatability of the forecasting experiment
- ways to report forecast
  - **point forecasting:**
    - $\widehat{\boldsymbol{X}}$ estimator for e.g. $\mathbb{E}(\boldsymbol{Y})$ or $\text{Med}(\boldsymbol{Y})$
    - evaluation based on forecasting error $\boldsymbol{Y} - \widehat{\boldsymbol{X}}$, resp. $\boldsymbol{y} - \widehat{\boldsymbol{X}}$
    - $\mathbb{E}$ can be strictly proper evaluated using MSE (mean square error)
    - Med can be strictly proper evaluated using MAE (mean absolute error)

Forecasting evaluation

probabilistic forecasting: (everything beyond point forecasting)

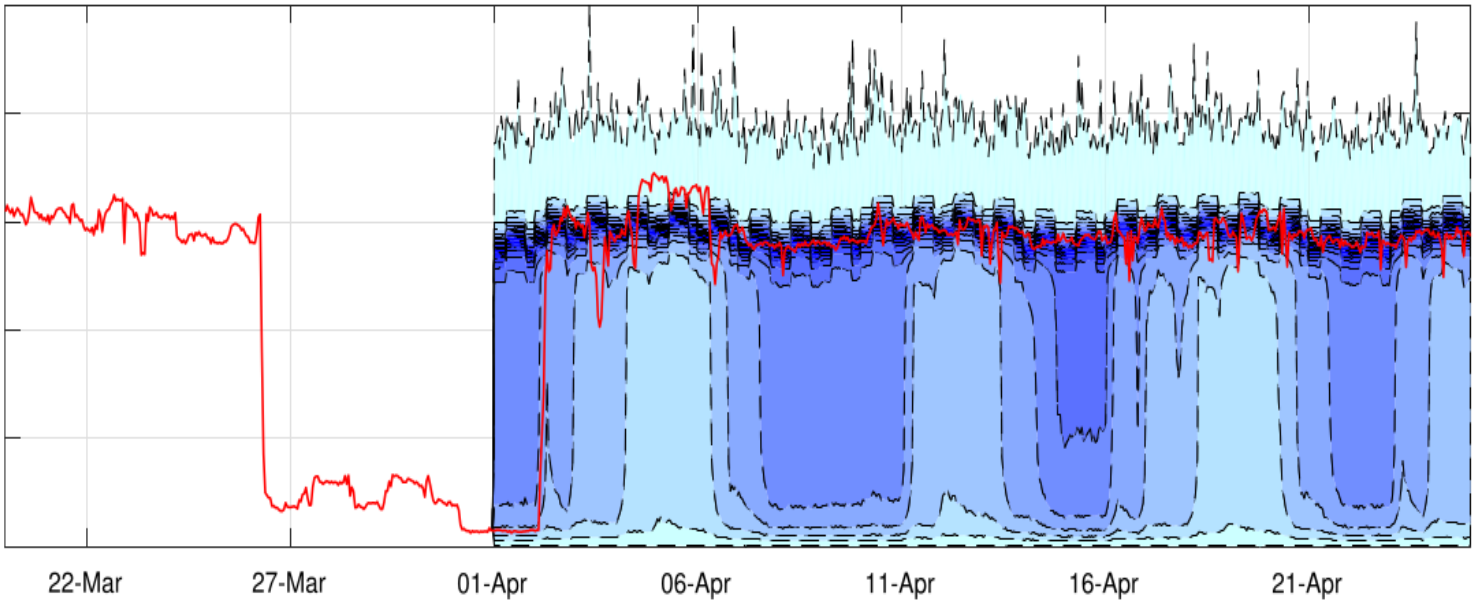
- characterises the uncertainty in the forecast, e.g.:



- usually forecast of marginal distributions F_{Y_h} of $\boldsymbol{Y} = (Y_1, \ldots, Y_H)'$, marginal densities f_{Y_h}, or quantiles $q_\alpha(Y_h)$ for $\alpha \in \mathcal{A}$
- strictly proper evaluation methods available, e.g. continuous rank probability score (**CRPS**) for F_{Y_h}
- used in e.g. Global Energy Forecasting Competitions (99%-tiles)

Probabilistic forecasting evaluation

- Problem with standard probabilistic methods:
  - forecasting only the marginals distributions
  - ignoring the dependency structure



(source: Berk, Hoffmann, Müller (2017) *International Journal of Forecasting*)

- require full forecast $\boldsymbol{F_X}$ for $\boldsymbol{F_Y}$ and strictly proper evaluation method

Evaluation measures for multivariate distributions

some measures available

- **Energy score**

$$\mathrm{ES}_\beta(\boldsymbol{F_X}, \boldsymbol{y}) = \mathbb{E}\left(\|\boldsymbol{X} - \boldsymbol{y}\|_2^\beta\right) - \frac{1}{2}\mathbb{E}\left(\|\boldsymbol{X} - \widetilde{\boldsymbol{X}}\|_2^\beta\right) \tag{1}$$

  - $\beta > 0$, $\boldsymbol{X}, \widetilde{\boldsymbol{X}} \overset{iid}{\sim} \boldsymbol{F_X}$
  - if $H = 1$ and $\beta = 1 \rightsquigarrow$ CRPS
  - strictly proper

- **Variogram score**

$$\mathrm{VS}_p(\boldsymbol{F_X}, \boldsymbol{y}; \boldsymbol{W}) = \sum_{i=1}^{H}\sum_{j=1}^{H} w_{i,j}(|y_i - y_j|^p - \mathbb{E}|X_i - X_j|^p)^2$$

  - with $p > 0$ and weight matrix $\boldsymbol{W} = (w_{i,j})_{i,j}$ (usually $w_{i,j} = c$)
  - not strictly proper *(forecasts with shifted mean have same score)*

Evaluation measures for multivariate distributions

- **Log-score**

$$\text{LogS}(\boldsymbol{F_X}, \boldsymbol{y}) = \log(\boldsymbol{f_X}(\boldsymbol{y})).$$

  - where $\boldsymbol{f_X}$ is density of $\boldsymbol{F_X}$
  - strictly proper
  - density forecast for $\boldsymbol{X}$ often not available (even if $\boldsymbol{X}$ is continuous)

- **Dawid-Sebastiani score**

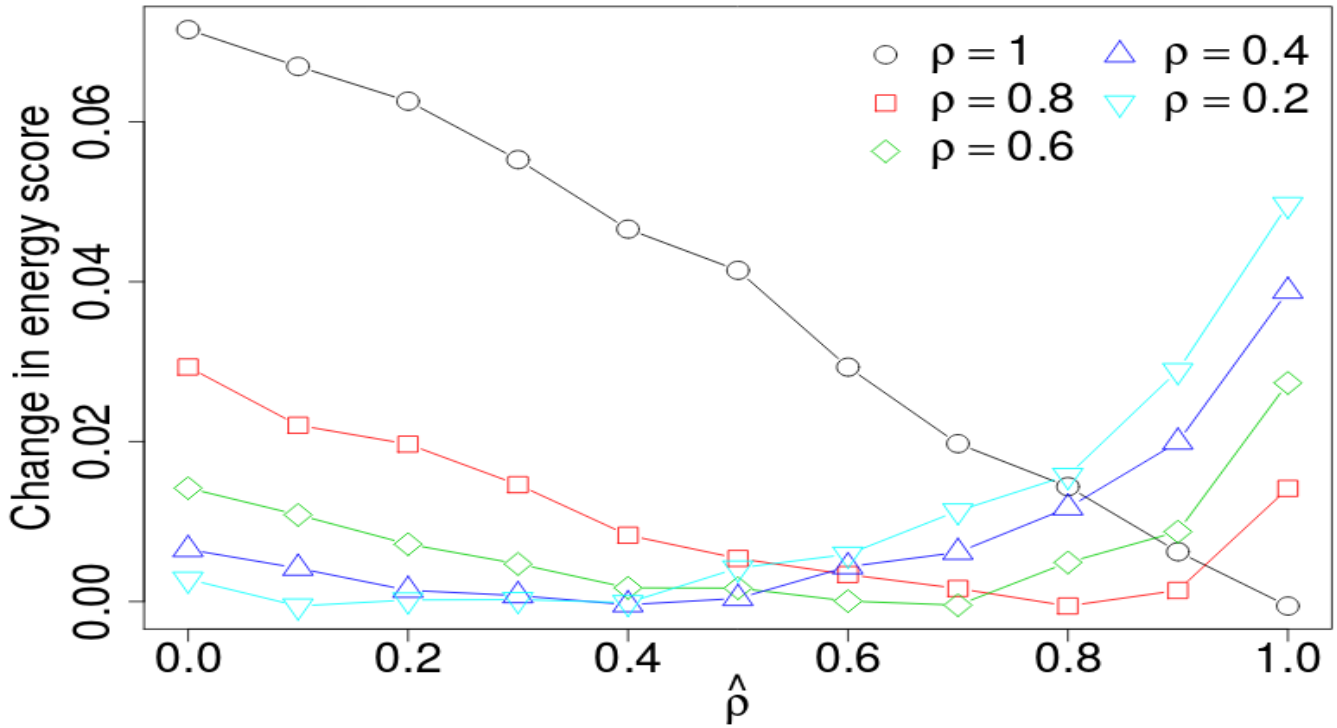
$$\text{DSS}(\boldsymbol{F_X}, \boldsymbol{y}) = \log(|\boldsymbol{\Sigma_X}|) + (\boldsymbol{y} - \boldsymbol{\mu_X})' \boldsymbol{\Sigma_X}^{-1} (\boldsymbol{y} - \boldsymbol{\mu_X})$$

  - with $\boldsymbol{\mu_X}$ and $\boldsymbol{\Sigma_X}$ as mean and covariance matrix of $\boldsymbol{X}$.
  - optimal if $\boldsymbol{Y}$ is normally distributed
  - not strictly proper

- Summary:
  - only energy score and log-score strictly proper
  - log-score not useful for practice as density forecast is required

Energy Score

- Energy score seems to be preferable, still it is hardly applied
- Pinson,Tastu(2013) state that **energy score is not sensitive in changes in dependency structure**, based on simulation results from a bivariate normal distribution



(source: Pinson, Tatsu (2013))

- Conclusion derived by looking at relative change in scores with respect to the true distribution

Marginal-Copula Scores

Idea:

- instead of full distribution $\boldsymbol{F_Y}$ evaluate
  - marginal distributions F_{Y_h}
  - copula $\boldsymbol{C_Y}$ of $\boldsymbol{Y}$
  - apply copula theory (Sklar's theorem)
  - hope: control somehow marginal and dependency measures
- **marginal score**: MS

$$\mathrm{MS}(\boldsymbol{a}) = \boldsymbol{a}'\mathbf{MS} = \sum_{h=1}^{H} a_h \mathrm{MS}_h.$$

  - where MS_h is a univariate scoring rule for Y_h (e.g. CRPS)
  - $\boldsymbol{a} = (a_1, \ldots, a_H)'$ a weight vector (usually $a_h = 1$)
- **copula score**: CS
  - CS is a multivariate score for the copula $\boldsymbol{C_X}$ of the copula $\boldsymbol{X}$

Marginal-Copula Scores

Problem:

- Combine marginal score MS and copula score CS to one score
- looking for $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ defines CES $= g(\text{MS}, \text{CS})$ such that strictly proper scoring rules can be achieved
- g must be strictly isotonic:

$$g(x_1, y_1) - g(x_1, y_2) - g(x_2, y_1) + g(x_2, y_2) > 0$$

for $x_1, x_2 \in \text{supp(MS)}$ and $y_1, y_2 \in \text{supp(CS)}$ with $x_1 < x_2, y_1 < y_2$

- possible options for g:
  a) $g(x, y) = w_1 x + w_2 y$ for weights $w_i > 0$ works
  b) $g(x, y) = wxy$ for weights $w > 0$ works on e.g. $(0, \infty) \times (0, \infty)$
- option a) is not intuitive (due to scaling/ units)

Consider forecast for $\boldsymbol{Y}$ and $c\boldsymbol{Y}$ with $c > 0$. For most marginal scores it follows that $\text{MS}(F_{\boldsymbol{Y}}, \boldsymbol{Y}; \boldsymbol{a}) \neq \text{MS}(F_{c\boldsymbol{Y}}, c\boldsymbol{Y}; \boldsymbol{a})$. For the popular CRPS we even have $\text{MS}(F_{\boldsymbol{Y}}, \boldsymbol{Y}; \boldsymbol{a}) = \frac{1}{c}\text{MS}(F_{c\boldsymbol{Y}}, c\boldsymbol{Y}; \boldsymbol{a})$, but for the copula it holds $\text{CS}(\boldsymbol{C}_{\boldsymbol{Y}}, \boldsymbol{U}_{\boldsymbol{Y}}) = \text{CS}(\boldsymbol{C}_{c\boldsymbol{Y}}, \boldsymbol{U}_{c\boldsymbol{Y}})$ with $\boldsymbol{U}_{\boldsymbol{Y}} \sim \boldsymbol{C}_{\boldsymbol{Y}}$ and $\boldsymbol{U}_{c\boldsymbol{Y}} \sim \boldsymbol{C}_{c\boldsymbol{Y}}$.

Theorem

If the MS_h is a strictly proper score for Y_h and CS is a strictly proper score for the copula $\boldsymbol{C_Y}$ of $\boldsymbol{Y} = (Y_1, \ldots, Y_H)'$ then the marginal-copula score

$$MCS(\boldsymbol{F_X}, \boldsymbol{y}; \boldsymbol{a}) = MS((F_{X_1}, \ldots, F_{X_H})', \boldsymbol{y}; \boldsymbol{a}) CS(\boldsymbol{C_X}, \boldsymbol{u_y})$$

$$= CS(\boldsymbol{C_X}, \boldsymbol{u_y}) \sum_{h=1}^{H} a_h MS_h(F_{X_h}, y_h)$$

with $\boldsymbol{F_X}$ as cumulative distribution function, with continuous marginals $F_{X_1}, \ldots, F_{X_H}$ and copula $\boldsymbol{C_X}$, of $\boldsymbol{X} = (X_1, \ldots, X_H)'$ which forecasts $\boldsymbol{Y}$, observation vector $\boldsymbol{y}$ and copula observations $\boldsymbol{u_Y} = (u_{\boldsymbol{Y},1}, \ldots, u_{\boldsymbol{Y},H})' = (F_{Y_1}(y_1), \ldots, F_{Y_H}(y_H))'$ and $\boldsymbol{a} = (a_1, \ldots, a_H)'$ with $a_h > 0$ is a strictly proper scoring rule.

Marginal-Copula Scores

- Possible choices for marginal scores F_{Y_i} of $\boldsymbol{Y} = (Y_1, \ldots, Y_H)'$
  - **CRPS** (univariate energy score)
  - Log score
  - Dawid-Sebastiani score
  - pinball score / quantile loss on a dense grid on $(0, 1)$

  many properties known
- Possible choice of the copula score for copula $\boldsymbol{C_Y}$ of $\boldsymbol{Y}$
  - **Energy score**
  - **Variogram score**
  - Log score
  - **Dawid-Sebastiani score**
- (originally) proposed score:
  - MS: CRPS
  - CS: Energy score
- Notation:
  - $\boldsymbol{U_Y} = (U_{\boldsymbol{Y},1}, \ldots, U_{\boldsymbol{Y},H})' = (F_{Y_1}(Y_1), \ldots, F_{Y_H}(Y_H))' = \boldsymbol{F_Y}(\boldsymbol{Y})$
  - $\boldsymbol{u_Y} = (u_{\boldsymbol{Y},1}, \ldots, u_{\boldsymbol{Y},H})' = (F_{Y_1}(y_1), \ldots, F_{Y_H}(y_H))' = \boldsymbol{F_Y}(\boldsymbol{y})$.

Copula Energy Scores (CES)

CES: energy score of the copula minus its lower bound scaled by $H^{-\frac{1}{2}}$

$$\begin{aligned}\mathrm{CES}(\boldsymbol{C_X}, \boldsymbol{u_Y}) &= \frac{1}{\sqrt{H}} \left(\mathrm{ES}(\boldsymbol{C_X}, \boldsymbol{u_Y}) - \mathrm{lb}_{\mathrm{CES}}\right) \\ &= \frac{1}{\sqrt{H}} \left(\mathbb{E}\left(\|\boldsymbol{U_X} - \boldsymbol{u_Y}\|_2\right) - \frac{1}{2}\mathbb{E}\left(\|\boldsymbol{U_X} - \widetilde{\boldsymbol{U}}_{\boldsymbol{X}}\|_2\right) - \mathrm{lb}_{\mathrm{CES}} \right)\end{aligned}$$

where $\mathrm{lb}_{\mathrm{CES}} = \frac{1}{4} - \frac{1}{2}\frac{1}{\sqrt{6}}$ due to

Lemma

$$\frac{\sqrt{H}}{4} \leq \mathbb{E}\,\|\boldsymbol{U_X} - \boldsymbol{u_Y}\|_2 \leq \frac{\sqrt{H}}{\sqrt{3}} \text{ and } \frac{\sqrt{H}}{3} \leq \mathbb{E}\left(\|\boldsymbol{U_X} - \widetilde{\boldsymbol{U}}_{\boldsymbol{X}}\|_2\right) \leq \frac{\sqrt{H}}{\sqrt{6}}$$

- not a strict bound (ongoing research)

Copula Variogram Score (CVS)

$$\mathrm{CVS}_p(\boldsymbol{C_X}, \boldsymbol{u_Y}; \boldsymbol{W}) = \frac{1}{\mathbf{1}'\boldsymbol{W}\mathbf{1}} \mathrm{CS}_p(\boldsymbol{C_X}, \boldsymbol{u_Y}; \boldsymbol{W}) \tag{2}$$

$$= \sum_{i=1}^{H}\sum_{j=1}^{H} w_{i,j}\left(|u_{\boldsymbol{Y},i} - u_{\boldsymbol{Y},j}|^p - \mathbb{E}|U_{\boldsymbol{X},i} - U_{\boldsymbol{X},j}|^p\right)^2$$

- with $\boldsymbol{U_X} = (U_{\boldsymbol{X},1}, \ldots, U_{\boldsymbol{X},H})' \sim \boldsymbol{C_X}$, $p > 0$ and weight matrix $\boldsymbol{W} = (w_{i,j})_{i,j}$.
- upper bound:

$$\mathrm{VS}_p(\boldsymbol{C_X}, \boldsymbol{u_Y}; \boldsymbol{W}) = \sum_{i=1}^{H}\sum_{j=1}^{H} w_{i,j}\left(|u_{\boldsymbol{Y},i} - u_{\boldsymbol{Y},j}|^p - \mathbb{E}|U_{\boldsymbol{X},i} - U_{\boldsymbol{X},j}|^p\right)^2 \tag{3}$$

$$\leq \sum_{i=1}^{H}\sum_{j=1}^{H} w_{i,j}\left(1^p - 0\right)^2 = \mathbf{1}'\boldsymbol{W}\mathbf{1}, \tag{4}$$

  which justifies the scaling constant.
- lower bound is zero, as for is holds $\mathrm{VS}_p(\boldsymbol{M}_H, \boldsymbol{U_Y}; \boldsymbol{W}) = 0$ if $\boldsymbol{U_Y} \sim \boldsymbol{M}_H$.

Copula Dawid-Sebastiani Score (CDSS)

$$\begin{aligned}\text{CDSS}(\boldsymbol{C_X}, \boldsymbol{u_y}) &= \text{DSS}(\boldsymbol{C_X}, \boldsymbol{u_y}) \\ &= \log(\det(\boldsymbol{\Sigma_{U_X}})) + (\boldsymbol{u_y} - \boldsymbol{\mu_{U_X}})' \boldsymbol{\Sigma_{U_X}^{-1}} (\boldsymbol{u_y} - \boldsymbol{\mu_{U_X}}) \quad (5)\end{aligned}$$

- $\mu_{\boldsymbol{U_X}}$ and $\boldsymbol{\Sigma_{U_X}}$ as mean and covariance matrix of $\boldsymbol{U_X} \sim \boldsymbol{C_X}$.
- as $\boldsymbol{\mu_{U_X}} = \frac{1}{2}\mathbf{1}$ and $\boldsymbol{\Sigma_{U_X}} = \boldsymbol{S R_{U_X} S}$ where $\boldsymbol{S} = \frac{1}{\sqrt{12}}\boldsymbol{I}$ and correlation matrix $\boldsymbol{R_{U_X}}$ it holds

$$\text{CDSS}(\boldsymbol{C_X}, \boldsymbol{u_y}) = -H \log(12 \det(\boldsymbol{R_{U_X}})) + \frac{1}{12} \left(\boldsymbol{u_y} - \frac{1}{2}\mathbf{1}\right)' \boldsymbol{R_{U_X}^{-1}} \left(\boldsymbol{u_y} - \frac{1}{2}\mathbf{1}\right)$$

- $\Rightarrow$ only measures dependency in correlation of the copula
- CDSS is unbounded:
  for $\boldsymbol{R_{U_X}}(\delta) = (1-\delta)\boldsymbol{I} + \delta \mathbf{1}\mathbf{1}'$ then it holds $\lim_{\delta \to 1} \det(\boldsymbol{R_{U_X}}(\delta)) = 0$.
  $\Rightarrow$ not applicable for multiplicative marginal-copula score

Reporting multivariate forecasts

- for sophisticated problems forecast distribution $\boldsymbol{F_X}$ (or density $\boldsymbol{f_X}$, or characteristic function $\boldsymbol{\varphi_X}$) is not explicitly available.
- reporting forecast as a large ensemble $\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(M)}$ for forecasting $\boldsymbol{Y}$
- repeat N (similar) forecasting experiments in a rolling window forecasting study: forecasts $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$ for $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$
- realised ensemble forecasts $\mathcal{X}_i = (\boldsymbol{x}_i^{(1)}, \ldots, \boldsymbol{x}_i^{(M)})'$ of the forecasting distribution $\boldsymbol{X}_i$ for $\boldsymbol{Y}_i$

Illustration rolling window forecasting study

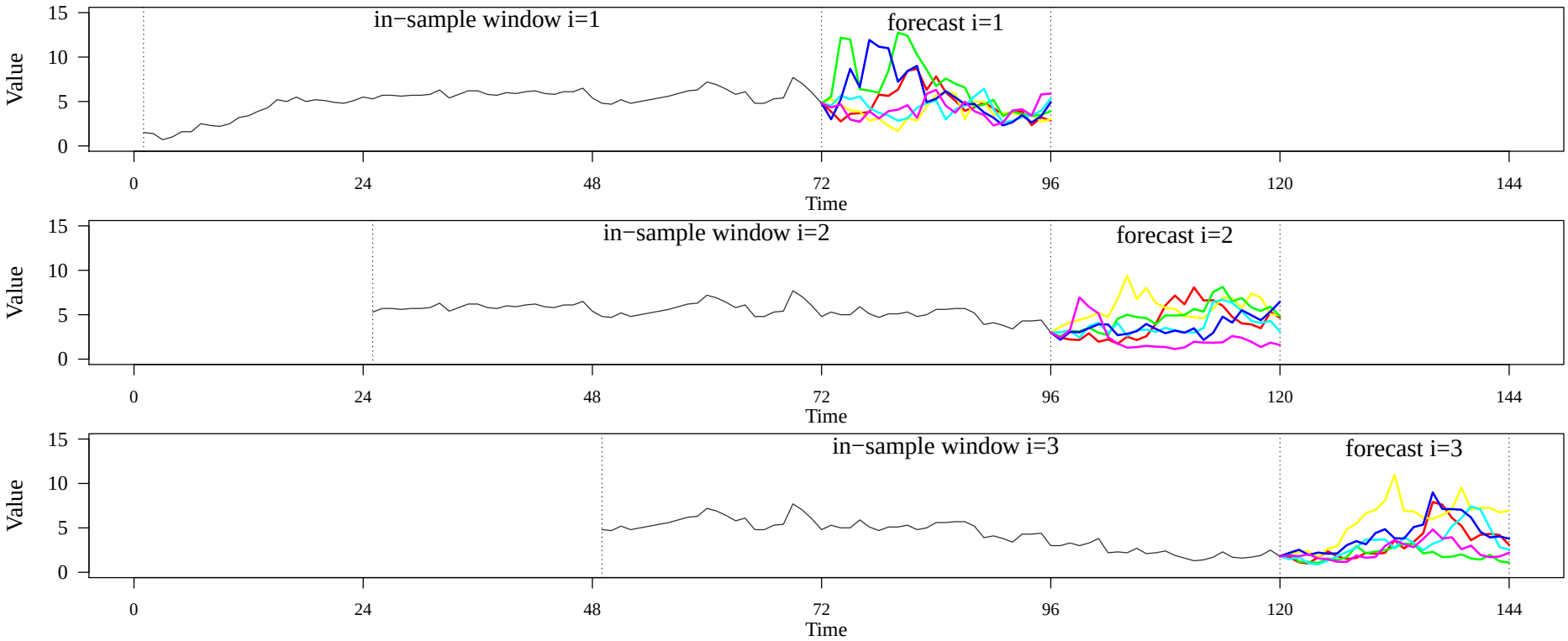


Figure: Illustration of a rolling window forecasting study with non-overlapping windows ($s_i = H(i-1)$) for $i = 1, \ldots, 3$ windows and $M = 6$ forecast samples $\boldsymbol{x}_{T,i}^{(1)}, \ldots, \boldsymbol{x}_{T,i}^{(M)}$ for each window i.

Estimating the scores

- for standard (multivariate) scores estimation straight forward, e.g.

$$\mathsf{ES}_{i,\beta}(\boldsymbol{F}_{\boldsymbol{X}_i}, \boldsymbol{y}_i) = \mathbb{E}\left(\|\boldsymbol{X}_i - \boldsymbol{y}_i\|_2^\beta\right) - \frac{1}{2}\mathbb{E}\left(\|\boldsymbol{X}_i - \widetilde{\boldsymbol{X}}_i\|_2^\beta\right) \tag{6}$$

$$= \mathsf{ED}_{\beta,i}(\boldsymbol{X}_i, \boldsymbol{y}_i) - \frac{1}{2}\mathsf{EI}_{\beta,i}(\boldsymbol{X}_i, \boldsymbol{y}_i) \tag{7}$$

- estimated by

$$\widehat{\mathsf{ED}}_{i,\beta} = \frac{1}{M}\sum_{j=1}^{M}\left\|\boldsymbol{X}_i^{(j)} - \boldsymbol{y}_i\right\|_2^\beta.$$

and

$$\widehat{\mathsf{EI}}_{i,\beta}^{K\text{band}} = \frac{1}{M}\sum_{j=1}^{M}\sum_{k=j}^{K}\left\|\boldsymbol{X}_i^{(j)} - \boldsymbol{X}_i^{(j+k)}\right\|_2^\beta$$

  - $K = M$ computationally expensive - but optimal
  - $K = 1$ fast

Estimating the copula scores

- $\boldsymbol{U}_{\boldsymbol{Y}_i} = (U_{\boldsymbol{Y}_i,1}, \ldots, U_{\boldsymbol{Y}_i,H})' = (F_{Y_{i,1}}(Y_{i,1}), \ldots, F_{Y_{i,H}}(Y_{i,H}))'$ with copula observations
  $\boldsymbol{u}_{\boldsymbol{Y}_i} = (u_{\boldsymbol{y}_i,1}, \ldots, u_{\boldsymbol{y}_i,H})' = (F_{Y_{i,1}}(y_{i,1}), \ldots, F_{Y_{i,H}}(y_{i,H}))'$ depend on true marginals $F_{Y_{i,h}}$.
- estimate $F_{Y_{i,h}}$ empirical distribution function (ecdf), e.g.
  $\widehat{F}_{Y_{i,h}}(z) = \widehat{F}_{Y_{i,h}}(z; \mathcal{X}_i) = \frac{1}{M}\sum_{j=1}^{M} \mathbb{1}\{\boldsymbol{x}_i^{(j)} \leq z\}$,
  (or the mid-point rule $\widehat{F}^{\text{mid}}_{Y_{i,h}}(z) = \frac{1}{2M}\sum_{j=1}^{M} \mathbb{1}\{\boldsymbol{x}_i^{(j)} \leq z\} + \mathbb{1}\{\boldsymbol{x}_i^{(j)} < z\}$)

$$\widehat{\boldsymbol{u}}_{\boldsymbol{Y}_i} = \widehat{\boldsymbol{u}}_{\boldsymbol{Y}_i}(\mathcal{X}_i) = (\widehat{F}_{Y_{i,1}}(y_{i,1}; \mathcal{X}_i), \ldots, \widehat{F}_{Y_{i,H}}(y_{i,H}; \mathcal{X}_i))'.$$

Estimating the copula scores

- $\boldsymbol{U}_{\boldsymbol{Y}_i} = (U_{\boldsymbol{Y}_i,1}, \dots, U_{\boldsymbol{Y}_i,H})' = (F_{Y_{i,1}}(Y_{i,1}), \dots, F_{Y_{i,H}}(Y_{i,H}))'$ with copula observations
  $\boldsymbol{u}_{\boldsymbol{Y}_i} = (u_{\boldsymbol{y}_i,1}, \dots, u_{\boldsymbol{y}_i,H})' = (F_{Y_{i,1}}(y_{i,1}), \dots, F_{Y_{i,H}}(y_{i,H}))'$ depend on true marginals $F_{Y_{i,h}}$.
- estimate $F_{Y_{i,h}}$ empirical distribution function (ecdf), e.g.
  $\widehat{F}_{Y_{i,h}}(z) = \widehat{F}_{Y_{i,h}}(z; \mathcal{X}_i) = \frac{1}{M} \sum_{j=1}^{M} \mathbb{1}\{\boldsymbol{x}_i^{(j)} \leq z\}$,
  (or the mid-point rule $\widehat{F}^{\text{mid}}_{Y_{i,h}}(z) = \frac{1}{2M} \sum_{j=1}^{M} \mathbb{1}\{\boldsymbol{x}_i^{(j)} \leq z\} + \mathbb{1}\{\boldsymbol{x}_i^{(j)} < z\}$)

$$\widehat{\boldsymbol{u}}_{\boldsymbol{Y}_i} = \widehat{\boldsymbol{u}}_{\boldsymbol{Y}_i}(\mathcal{X}_i) = (\widehat{F}_{Y_{i,1}}(y_{i,1}; \mathcal{X}_i), \dots, \widehat{F}_{Y_{i,H}}(y_{i,H}; \mathcal{X}_i))'.$$

- Problem:
  misspecified marginal can lead to estimated copula lower scores than the true model

Estimating the copula scores

- $\boldsymbol{U}_{\boldsymbol{Y}_i} = (U_{\boldsymbol{Y}_i,1}, \ldots, U_{\boldsymbol{Y}_i,H})' = (F_{Y_{i,1}}(Y_{i,1}), \ldots, F_{Y_{i,H}}(Y_{i,H}))'$ with copula observations
  $\boldsymbol{u}_{\boldsymbol{Y}_i} = (u_{\boldsymbol{y}_i,1}, \ldots, u_{\boldsymbol{y}_i,H})' = (F_{Y_{i,1}}(y_{i,1}), \ldots, F_{Y_{i,H}}(y_{i,H}))'$ depend on true marginals $F_{Y_{i,h}}$.
- estimate $F_{Y_{i,h}}$ empirical distribution function (ecdf), e.g.
  $\widehat{F}_{Y_{i,h}}(z) = \widehat{F}_{Y_{i,h}}(z; \mathcal{X}_i) = \frac{1}{M}\sum_{j=1}^{M} \mathbb{1}\{\boldsymbol{x}_i^{(j)} \leq z\}$,
  (or the mid-point rule $\widehat{F}^{\text{mid}}_{Y_{i,h}}(z) = \frac{1}{2M}\sum_{j=1}^{M} \mathbb{1}\{\boldsymbol{x}_i^{(j)} \leq z\} + \mathbb{1}\{\boldsymbol{x}_i^{(j)} < z\}$)

$$\widehat{\boldsymbol{u}}_{\boldsymbol{Y}_i} = \widehat{\boldsymbol{u}}_{\boldsymbol{Y}_i}(\mathcal{X}_i) = (\widehat{F}_{Y_{i,1}}(y_{i,1}; \mathcal{X}_i), \ldots, \widehat{F}_{Y_{i,H}}(y_{i,H}; \mathcal{X}_i))'.$$

- Problem:
  misspecified marginal can lead to estimated copula lower scores than the true model
- Solution:
  Force the marginals to be uniform, while preserving the dependency structure

Estimating the copula scores

- $R_{i,h}$ the rank of $\widehat{u}_{\boldsymbol{Y}_i,h}$ within $\widehat{u}_{\boldsymbol{Y}_1,h},\ldots,\widehat{u}_{\boldsymbol{Y}_N,h}$.
- define the *adjusted estimated copula observations* by

$$\widehat{u}^*_{\boldsymbol{Y}_i,h} = \frac{2R_{i,h}-1}{2N}.$$

  taking values on $\frac{1}{2N},\ldots,\frac{2N-1}{2N}$
- resulting ecdf has minimal Komogorov-Smirnov (KS) distance to the uniform distribution. (no other justification)

Estimating the copula scores

- $R_{i,h}$ the rank of $\widehat{u}_{\boldsymbol{Y}_i,h}$ within $\widehat{u}_{\boldsymbol{Y}_1,h},\ldots,\widehat{u}_{\boldsymbol{Y}_N,h}$.
- define the *adjusted estimated copula observations* by

$$\widehat{u}^*_{\boldsymbol{Y}_i,h} = \frac{2R_{i,h}-1}{2N}.$$

  taking values on $\frac{1}{2N},\ldots,\frac{2N-1}{2N}$
- resulting ecdf has minimal Komogorov-Smirnov (KS) distance to the uniform distribution. (no other justification)
- estimate $\boldsymbol{C}_{\boldsymbol{X}_i}$ by the empirical copula, we suggest

$$\widehat{\boldsymbol{C}}_{\boldsymbol{X}_i}(u_1,\ldots,u_H) = \frac{1}{M}\sum_{j=1}^{M}\mathbb{1}\{\widetilde{R}_{i,j,1}/M \leq u_1,\ldots,\widetilde{R}_{i,j,H}/M \leq u_H\}$$

  with the ranks

$$\widetilde{R}_{i,j,h} = \frac{1}{2}\sum_{k=1}^{M}\mathbb{1}\{x^{(k)}_{i,h} \leq x^{(j)}_{i,h}\} + \mathbb{1}\{x^{(k)}_{i,h} < x^{(j)}_{i,h}\}$$

  which break ties by the mid-point rule.

Estimating the copula scores

- Further problem: We know

$$\mathbb{E}[\mathrm{MS} \cdot \mathrm{CS}] = \mathbb{E}[\mathrm{MS}]\mathbb{E}[\mathrm{CS}] + \mathbb{C}\mathrm{ov}[\mathrm{MS}, \mathrm{CS}]$$

- Thus we estimate

$$\widehat{\mathrm{MS}(\boldsymbol{a})\text{-}\mathrm{CS}}_i = \widehat{\mathrm{MS}}_i(\boldsymbol{a})\widehat{\mathrm{CS}}_i - \widehat{\sigma}_{\mathrm{MS,CS}}$$

with $\widehat{\sigma}_{\mathrm{MS,CS}}$ as estimator for $\mathbb{C}\mathrm{ov}[\mathrm{MS}, \mathrm{CS}]$.

Application in simulation studies

9 scores:

i) Energy score (ES)

ii) Variogram score (VS)

iii) Dawid-Sebastiani score (DSS)

iv) CRPS-copula energy score

v) CRPS-copula variogram score

vi) CRPS

vii) Copula energy score

viii) Copula variogram score

ix) Copula DSS score

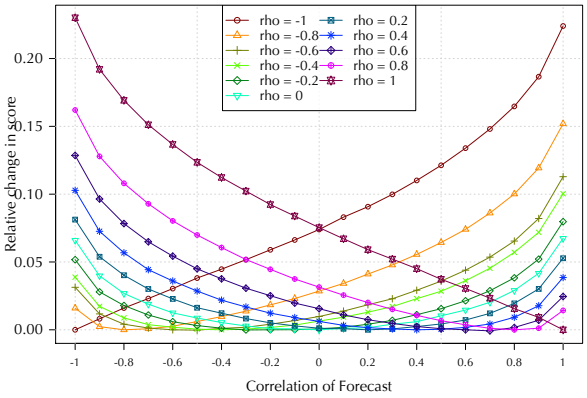
$\rightarrow$ evaluation for score SC using two criteria

I) relative change in score with respect to best: $\text{RelCh}(\text{SC}) = \frac{\overline{\text{SC}} - \overline{\text{SC}}^*}{\overline{\text{SC}}^*}$

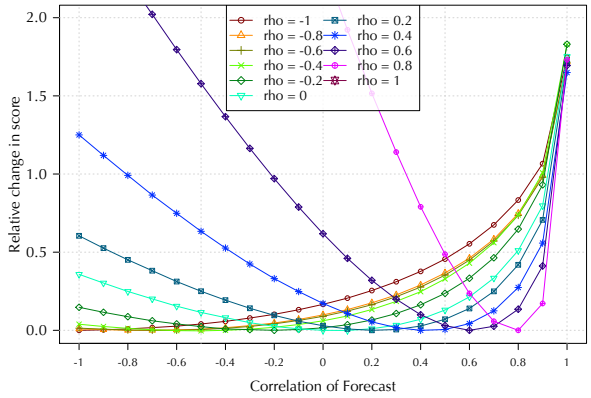
II) DM-test statistics with respect to the best

first study of Pinson, Tatsu (2013) [change in correlation of bivariate normal] using

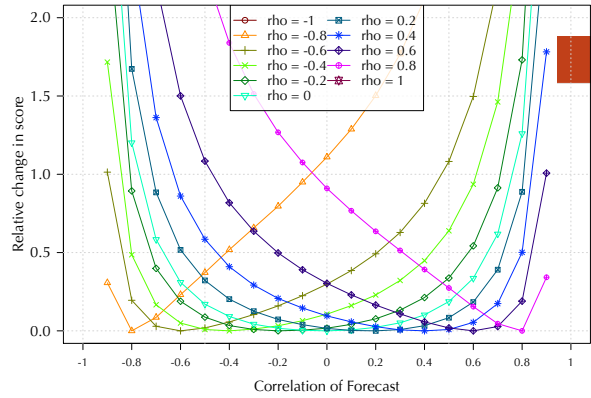
- $N = 2^9 = 512$ (window length), $M = 2^{14} = 16384$ (ensemble size)



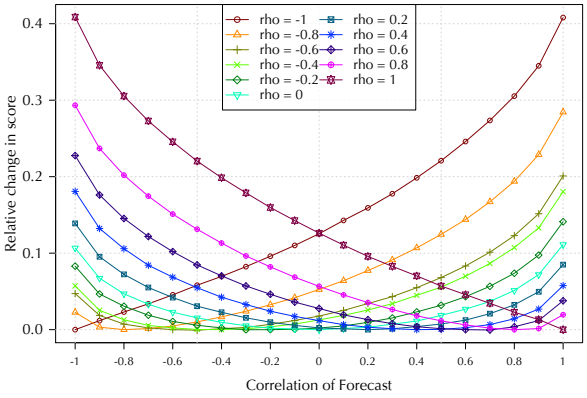
ES



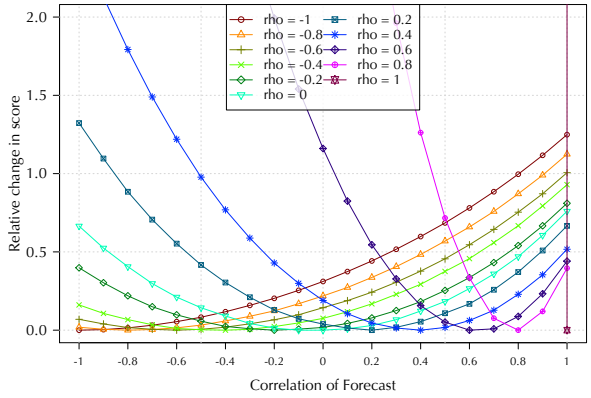
VS



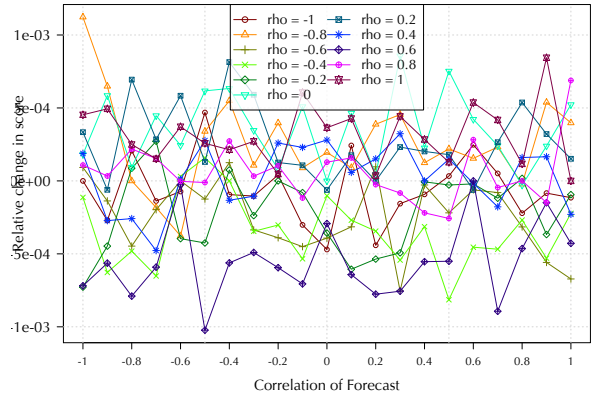
DSS



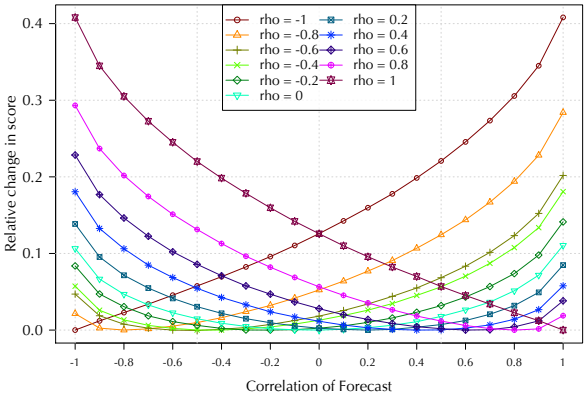
CRPS-CES



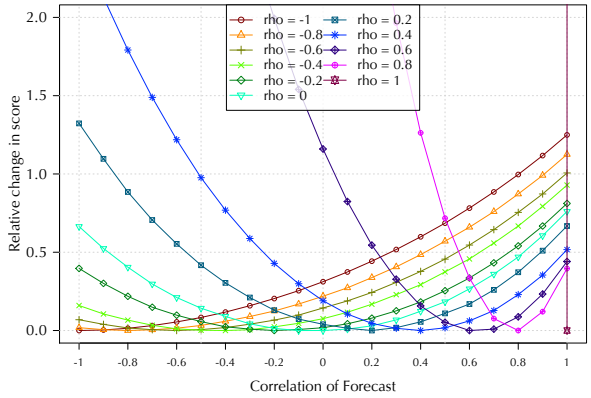
CRPS-CVS



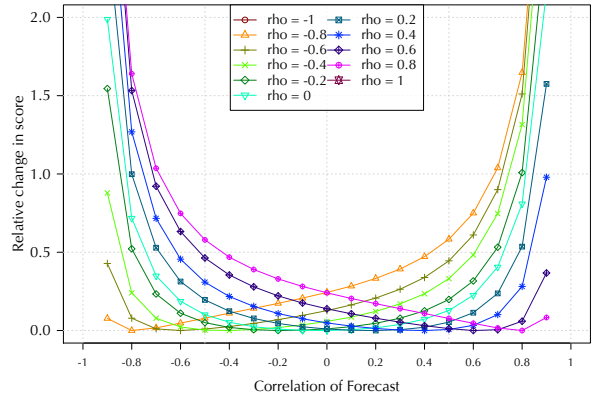
CRPS



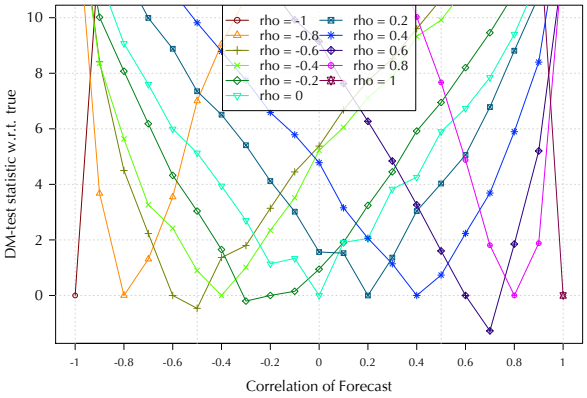
CES



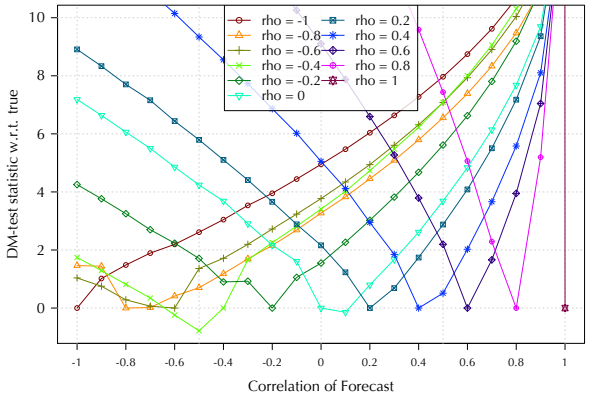
CVS



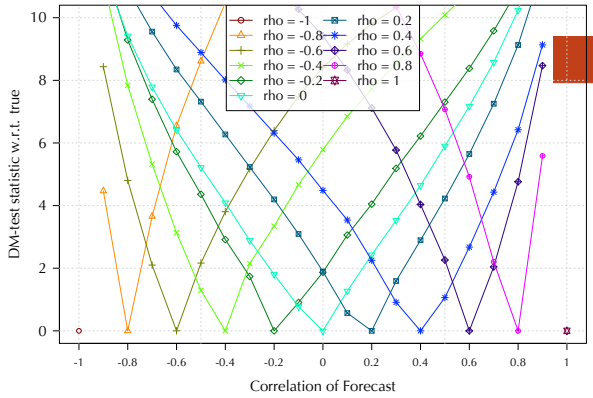
CDSS



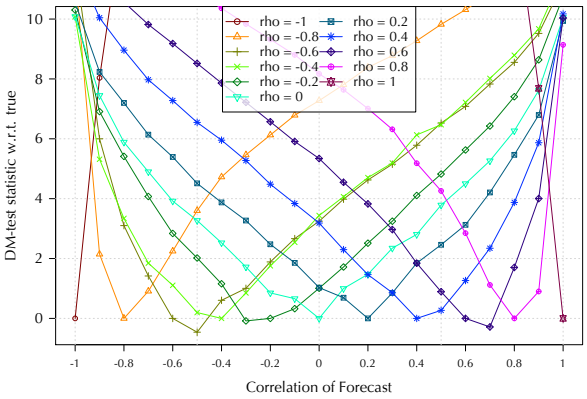
ES



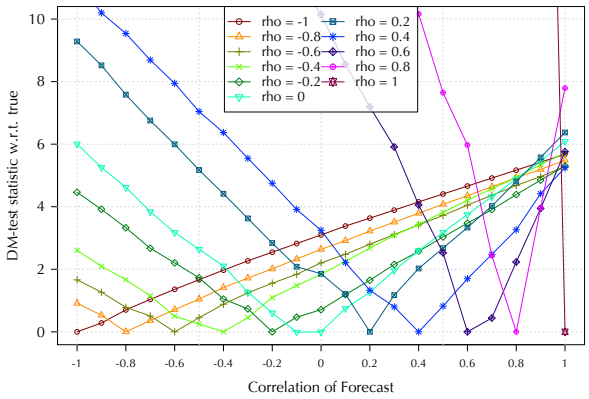
VS



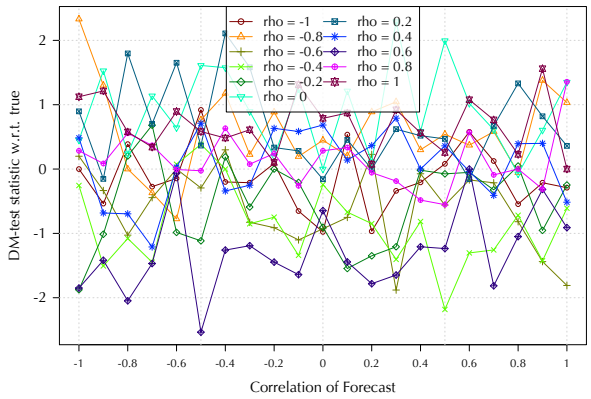
DSS



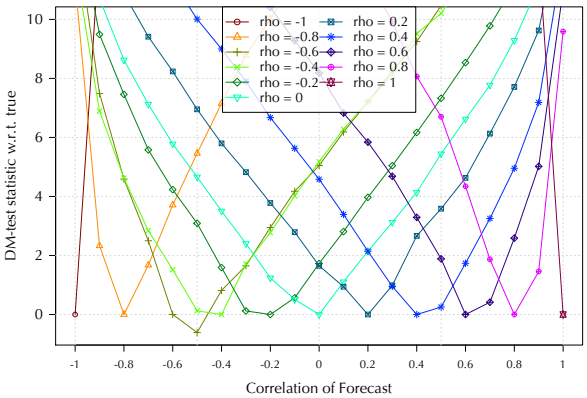
CRPS-CES



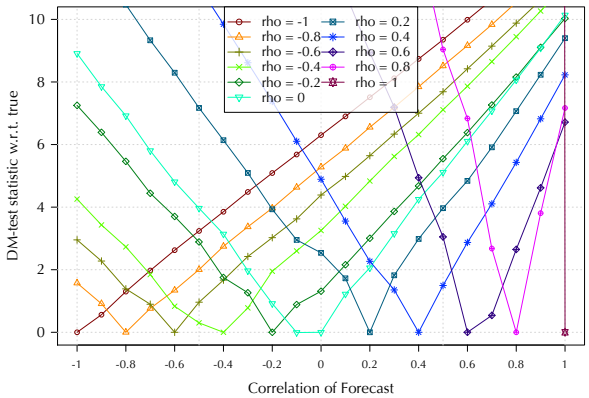
CRPS-CVS



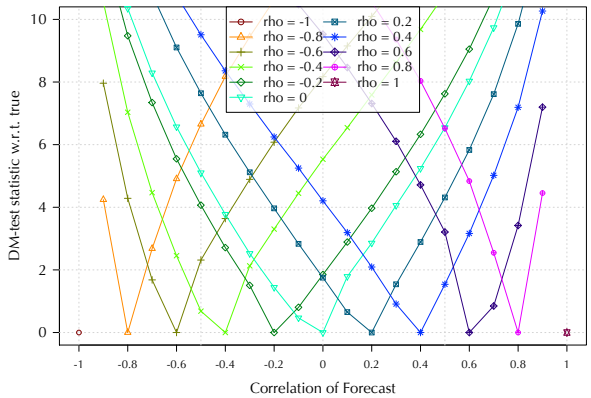
CRPS



CES



CVS



CDSS

2nd Experiment: on bivariate normal distribution

- With $\boldsymbol{\mu} = (0,0)'$ and $\boldsymbol{\Sigma}(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
  - i) (true setting): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\rho = \sqrt{2}/2$
  - ii) (symmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + a_1\mathbf{1}, \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iii) (asymmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + (a_2, -a_2)', \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iv) (smaller variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_3\boldsymbol{\Sigma}(\rho))$ with $a_3 < 1$ and $\rho = \sqrt{2}/2$
  - v) (larger variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_4\boldsymbol{\Sigma}(\rho))$ with $a_4 > 1$ and $\rho = \sqrt{2}/2$
  - vi) (smaller correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_5))$ with $a_5 < \rho$
  - vii) (larger correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_6))$ with $a_6 > \rho$

2nd Experiment: on bivariate normal distribution

- With $\boldsymbol{\mu} = (0,0)'$ and $\boldsymbol{\Sigma}(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
  - i) (true setting): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\rho = \sqrt{2}/2$
  - ii) (symmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + a_1\mathbf{1}, \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iii) (asymmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + (a_2, -a_2)', \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iv) (smaller variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_3\boldsymbol{\Sigma}(\rho))$ with $a_3 < 1$ and $\rho = \sqrt{2}/2$
  - v) (larger variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_4\boldsymbol{\Sigma}(\rho))$ with $a_4 > 1$ and $\rho = \sqrt{2}/2$
  - vi) (smaller correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_5))$ with $a_5 < \rho$
  - vii) (larger correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_6))$ with $a_6 > \rho$
- change in biased model so that change theoretical likelihood is the same for all settings ii) - vii)

2nd Experiment: on bivariate normal distribution

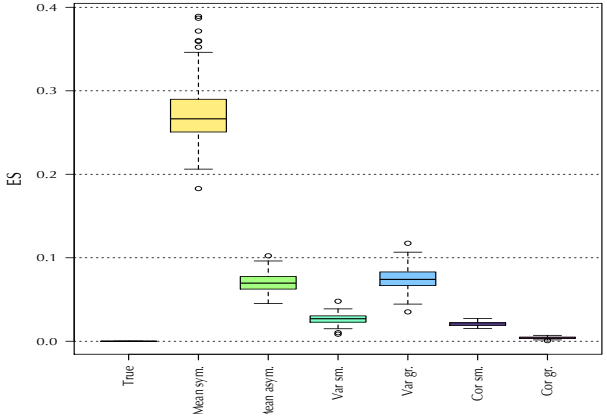
- With $\boldsymbol{\mu} = (0,0)'$ and $\boldsymbol{\Sigma}(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
  - i) (true setting): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\rho = \sqrt{2}/2$
  - ii) (symmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + a_1 \mathbf{1}, \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iii) (asymmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + (a_2, -a_2)', \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iv) (smaller variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_3 \boldsymbol{\Sigma}(\rho))$ with $a_3 < 1$ and $\rho = \sqrt{2}/2$
  - v) (larger variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_4 \boldsymbol{\Sigma}(\rho))$ with $a_4 > 1$ and $\rho = \sqrt{2}/2$
  - vi) (smaller correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_5))$ with $a_5 < \rho$
  - vii) (larger correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_6))$ with $a_6 > \rho$
- change in biased model so that change theoretical likelihood is the same for all settings ii) - vii)
- consider $\rho = \sqrt{2}/2 \approx 0.707$, $a_5 = 0$ (1 degree of freedom) $\Rightarrow$ likelihood reduction $\delta = \frac{1}{2}\log(2)$

$(a_1 = \sqrt{\frac{\delta}{2-\sqrt{2}}}, a_2 = \sqrt{\frac{\delta}{2+\sqrt{2}}}, a_3 \approx 0.48124, a_4 \approx 2.62729$ and $a_6 \approx 0.89032)$

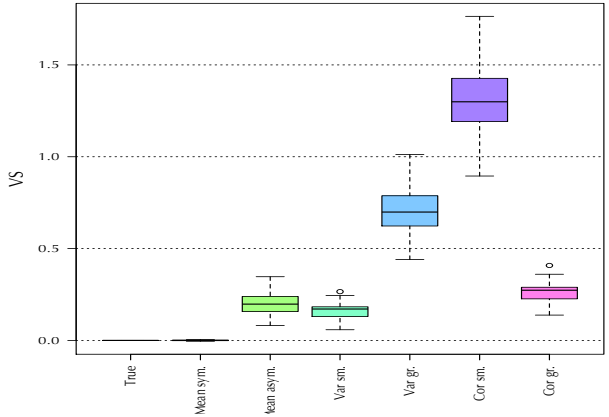
2nd Experiment: on bivariate normal distribution

- With $\boldsymbol{\mu} = (0,0)'$ and $\boldsymbol{\Sigma}(\rho) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$
  - i) (true setting): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\rho = \sqrt{2}/2$
  - ii) (symmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + a_1 \mathbf{1}, \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iii) (asymmetric mean bias): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu} + (a_2, -a_2)', \boldsymbol{\Sigma}(\rho))$ with $\rho = \sqrt{2}/2$
  - iv) (smaller variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_3 \boldsymbol{\Sigma}(\rho))$ with $a_3 < 1$ and $\rho = \sqrt{2}/2$
  - v) (larger variance): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, a_4 \boldsymbol{\Sigma}(\rho))$ with $a_4 > 1$ and $\rho = \sqrt{2}/2$
  - vi) (smaller correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_5))$ with $a_5 < \rho$
  - vii) (larger correlation): $\boldsymbol{X} \sim \mathcal{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma}(a_6))$ with $a_6 > \rho$
- change in biased model so that change theoretical likelihood is the same for all settings ii) - vii)
- consider $\rho = \sqrt{2}/2 \approx 0.707$, $a_5 = 0$ (1 degree of freedom) $\Rightarrow$ likelihood reduction $\delta = \frac{1}{2}\log(2)$

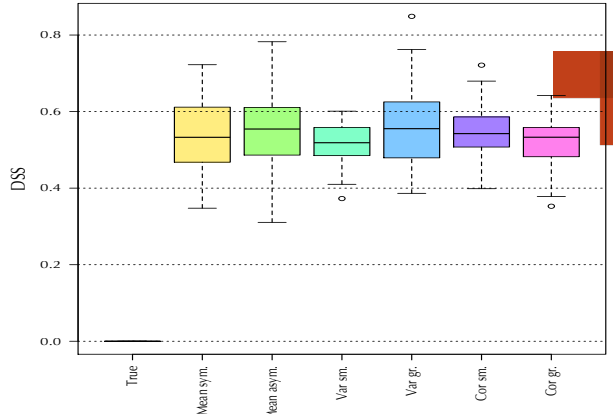
  ($a_1 = \sqrt{\frac{\delta}{2-\sqrt{2}}}$, $a_2 = \sqrt{\frac{\delta}{2+\sqrt{2}}}$, $a_3 \approx 0.48124$, $a_4 \approx 2.62729$ and $a_6 \approx 0.89032$)
- $M = 2^{13} = 8192$ (ensemble sample size), $N = 2^8 = 256$ (rolling window length), $L = 2^6 = 64$ (replications)



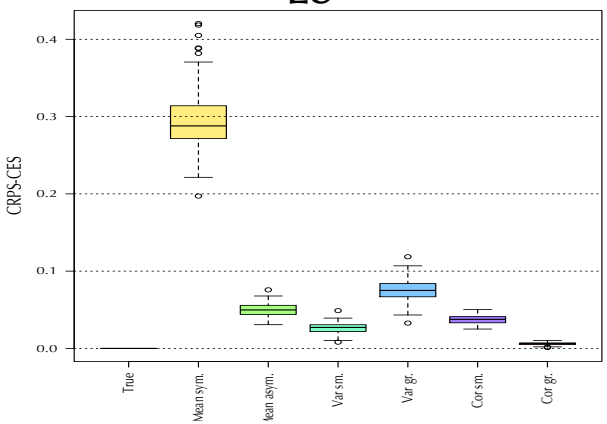
ES



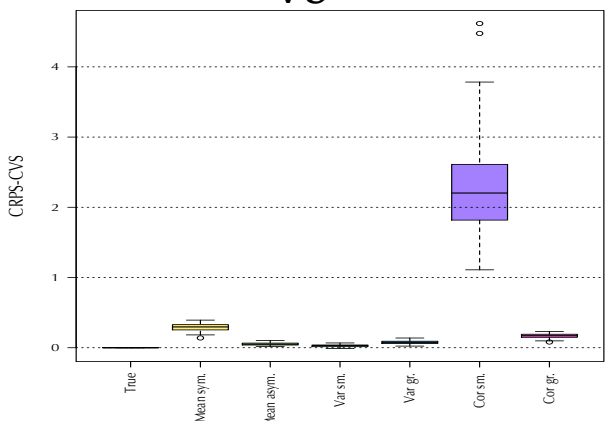
VS



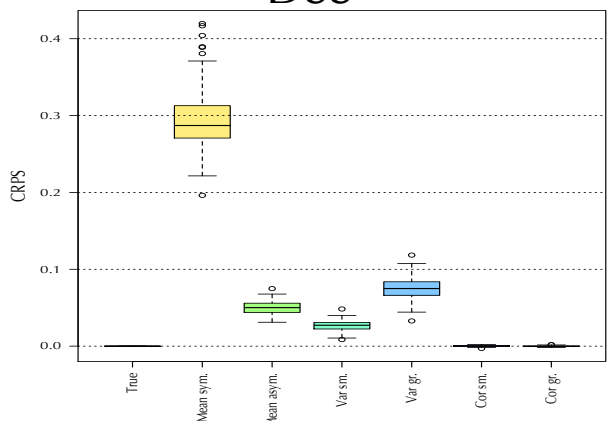
DSS



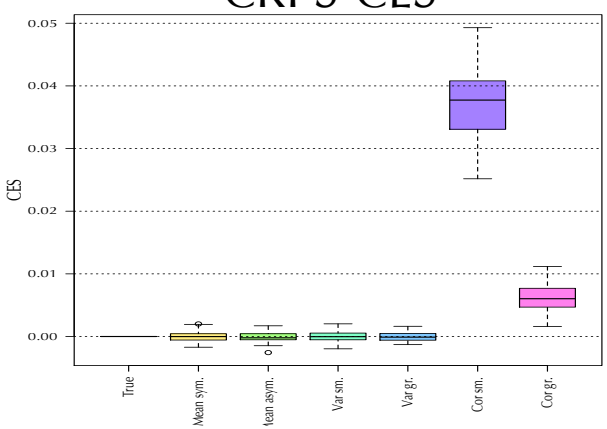
CRPS-CES



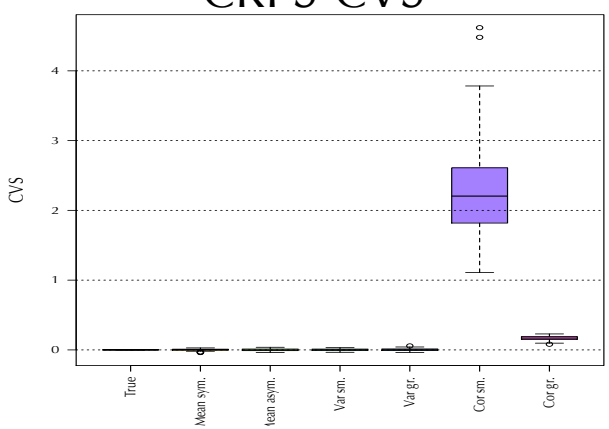
CRPS-CVS



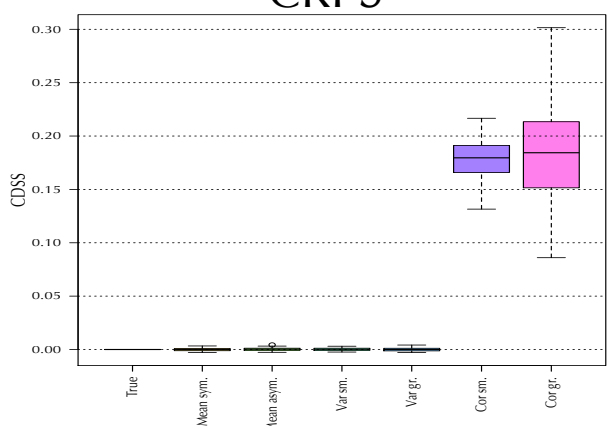
CRPS



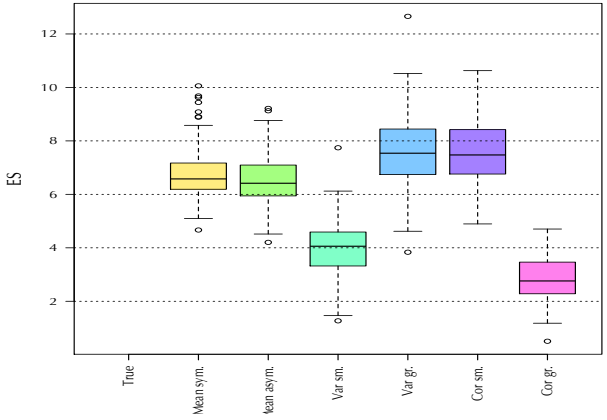
CES



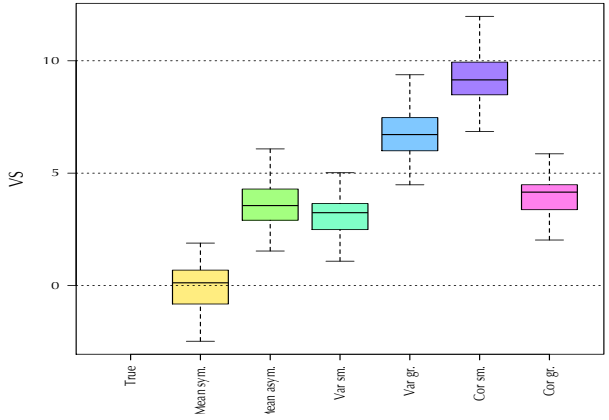
CVS



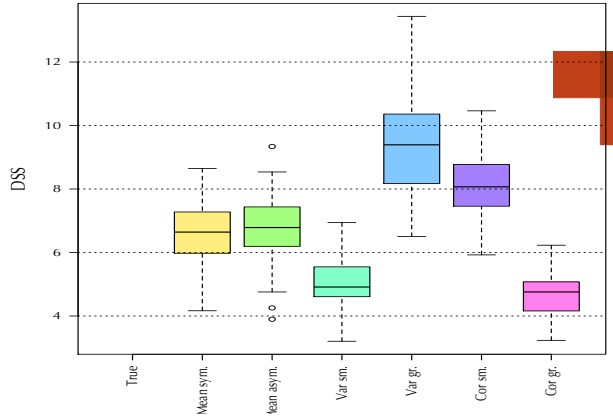
CDSS



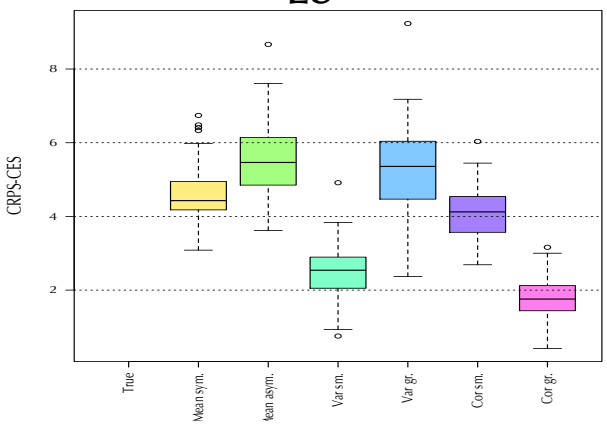
ES



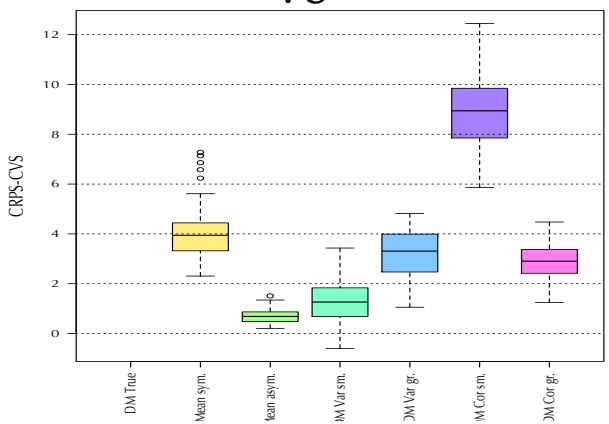
VS



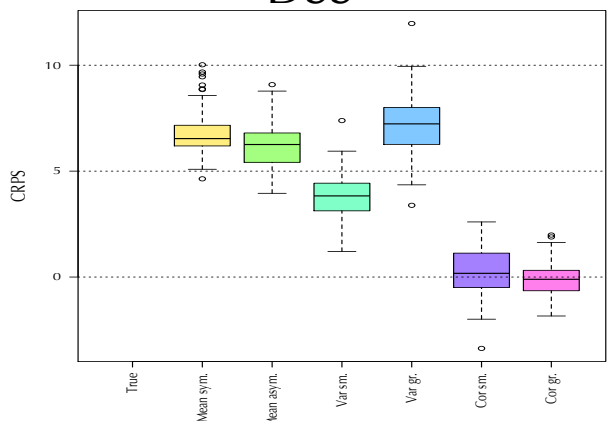
DSS



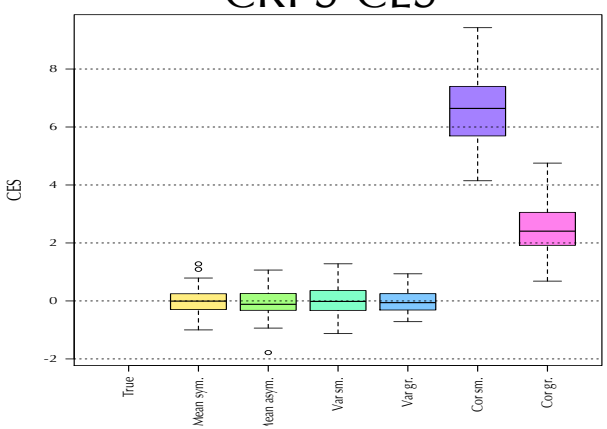
CRPS-CES



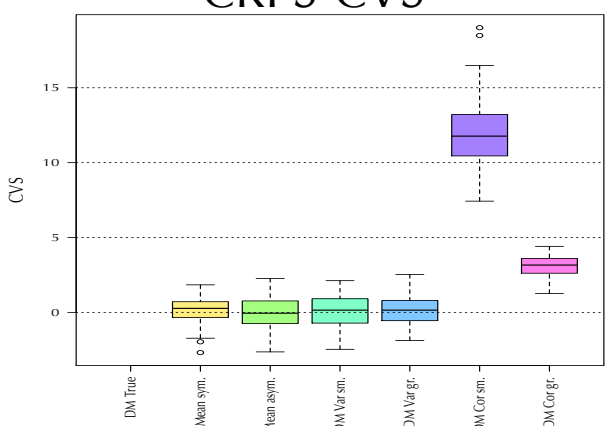
CRPS-CVS



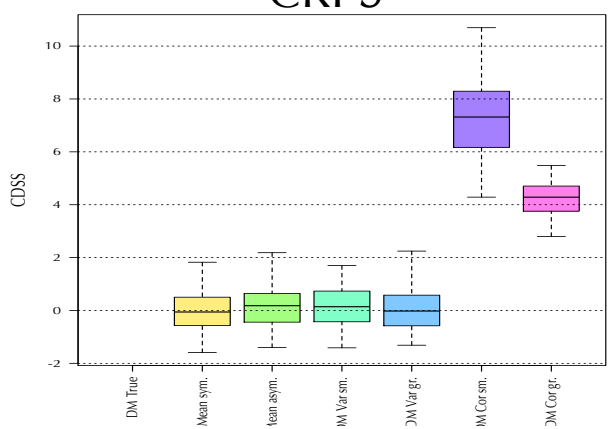
CRPS



CES



CVS



CDSS

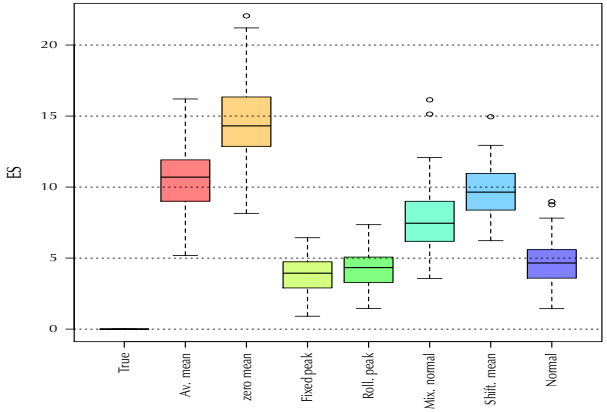
3rd Experiment: Random peak study

- i) (true) $\boldsymbol{X}_i = \boldsymbol{Y}_i + Q\boldsymbol{Z}_i$ with $\boldsymbol{Y}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{Z}_i \overset{\text{iid}}{\sim} \mathcal{U}(\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_H\})$
- ii) (average mean) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\mu\boldsymbol{1}, \boldsymbol{I})$ with $\mu = \frac{Q}{H}$
- iii) (zero mean) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{0}, \boldsymbol{I})$
- iv) (fixed peak) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{\mu}, \boldsymbol{I})$ with $\boldsymbol{\mu} = (Q, 0, \dots, 0)'$
- v) (rolling peak) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{\mu}_i, \boldsymbol{I})$ with $\boldsymbol{\mu}_i = \boldsymbol{e}_{1+(i-1)\text{mod}H}$ with as unit vector for the i'th coordinate.
- vi) (mixture normal with same marginals) $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,H})'$ with $X_{i,j} \overset{\text{iid}}{\sim} \begin{cases} \mathcal{N}_1(0,1) & U_j \leq \frac{H-1}{H} \\ \mathcal{N}_1(Q,1) & U_j > \frac{H-1}{H} \end{cases}$ where $U_j \overset{\text{iid}}{\sim} \mathcal{U}([0,1])$.
- vii) (shifted mean) $\boldsymbol{X}_i = \boldsymbol{Y}_i + Q\boldsymbol{Z}_i$ with $\boldsymbol{Y}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(Q/H\boldsymbol{1}, \boldsymbol{I})$ and $\boldsymbol{Z}_i \overset{\text{iid}}{\sim} \mathcal{U}(\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_H\})$
- viii) (normal with true mean and covariance) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \frac{Q}{H}\boldsymbol{1}$ and $\boldsymbol{\Sigma} = (H + Q^2)/H\boldsymbol{I} - Q^2/H^2\boldsymbol{1}\boldsymbol{1}'$

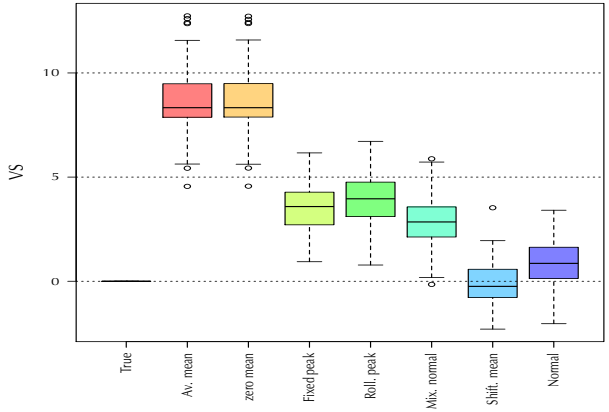
3rd Experiment: Random peak study

- i) (true) $\boldsymbol{X}_i = \boldsymbol{Y}_i + Q\boldsymbol{Z}_i$ with $\boldsymbol{Y}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{Z}_i \overset{\text{iid}}{\sim} \mathcal{U}(\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_H\})$
- ii) (average mean) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\mu\boldsymbol{1}, \boldsymbol{I})$ with $\mu = \frac{Q}{H}$
- iii) (zero mean) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{0}, \boldsymbol{I})$
- iv) (fixed peak) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{\mu}, \boldsymbol{I})$ with $\boldsymbol{\mu} = (Q, 0, \dots, 0)'$
- v) (rolling peak) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{\mu}_i, \boldsymbol{I})$ with $\boldsymbol{\mu}_i = \boldsymbol{e}_{1+(i-1)\text{mod}H}$ with as unit vector for the i'th coordinate.
- vi) (mixture normal with same marginals) $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,H})'$ with $X_{i,j} \overset{\text{iid}}{\sim} \begin{cases} \mathcal{N}_1(0,1) & U_j \leq \frac{H-1}{H} \\ \mathcal{N}_1(Q,1) & U_j > \frac{H-1}{H} \end{cases}$ where $U_j \overset{\text{iid}}{\sim} \mathcal{U}([0,1])$.
- vii) (shifted mean) $\boldsymbol{X}_i = \boldsymbol{Y}_i + Q\boldsymbol{Z}_i$ with $\boldsymbol{Y}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(Q/H\boldsymbol{1}, \boldsymbol{I})$ and $\boldsymbol{Z}_i \overset{\text{iid}}{\sim} \mathcal{U}(\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_H\})$
- viii) (normal with true mean and covariance) $\boldsymbol{X}_i \overset{\text{iid}}{\sim} \mathcal{N}_H(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu} = \frac{Q}{H}\boldsymbol{1}$ and $\boldsymbol{\Sigma} = (H + Q^2)/H\boldsymbol{I} - Q^2/H^2\boldsymbol{1}\boldsymbol{1}'$

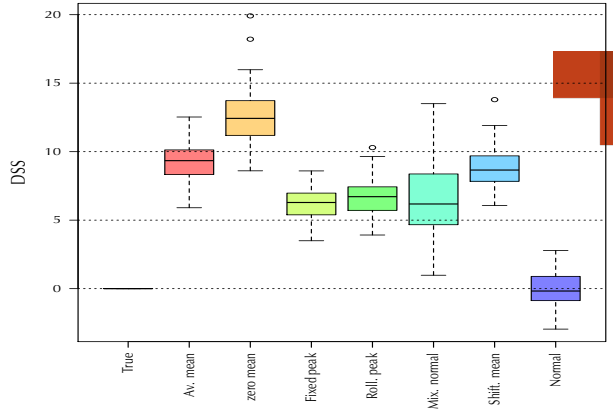
- First: $H = 3$-dim. case with a peak size of $Q = 5$
- $M = 2^{14} = 16384$ (ensemble sample size), $N = 2^5 = 32$ (rolling window length), $L = 2^6 = 64$ (replications), only DM-test



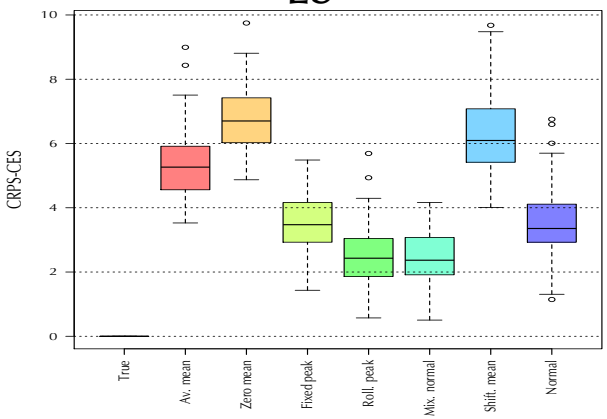
ES



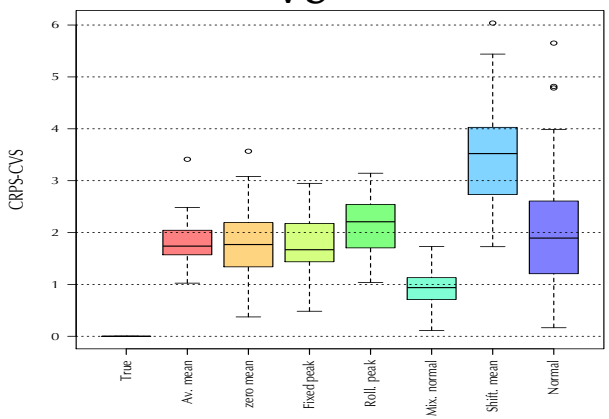
VS



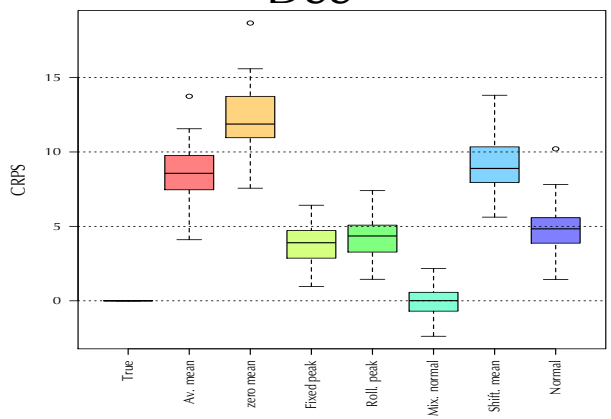
DSS



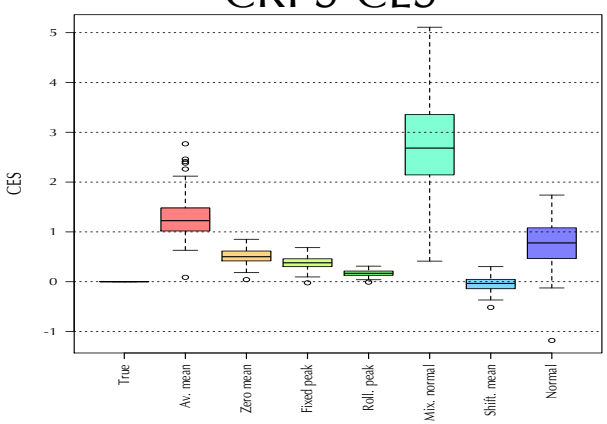
CRPS-CES



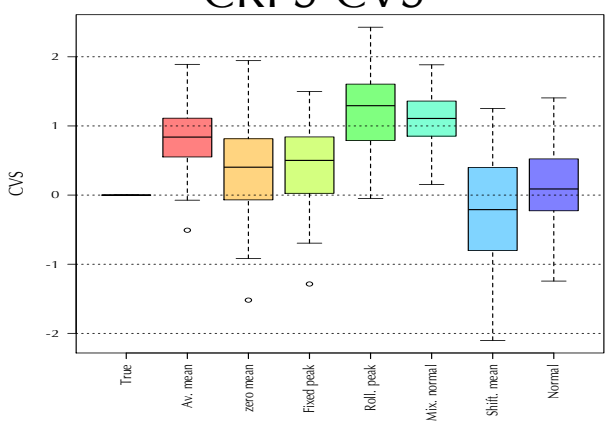
CRPS-CVS



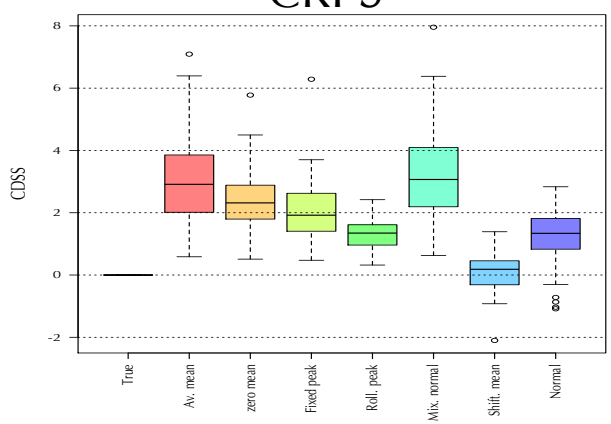
CRPS



CES



CVS



CDSS

Variants of 3rd Experiment: Random peak study

a) Effect of ensemble sample size M

$H = 3$-dim. case with a peak size of $Q = 5$

$M \in \mathcal{M} = \{2^i | i \in \{4, \dots, 14\}\} = \{2^4, 2^5, \dots, 2^{14}\} = \{16, 32, \dots, 16384\}$

Variants of 3rd Experiment: Random peak study

a) Effect of ensemble sample size M

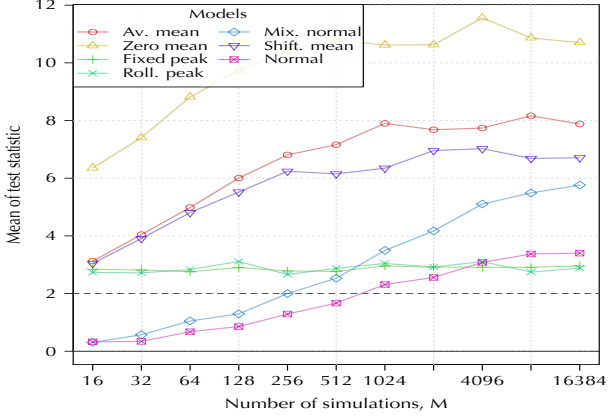
$H = 3$-dim. case with a peak size of $Q = 5$

$M \in \mathcal{M} = \{2^i | i \in \{4, \dots, 14\}\} = \{2^4, 2^5, \dots, 2^{14}\} = \{16, 32, \dots, 16384\}$

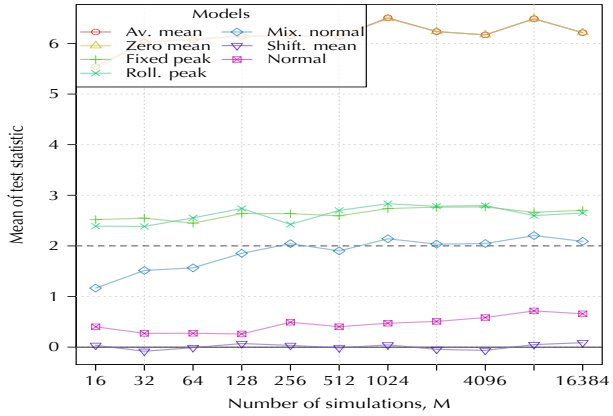
b) Effect of ensemble sample dimension H

$H = 9$-dim. case with a peak size of $Q = 5$

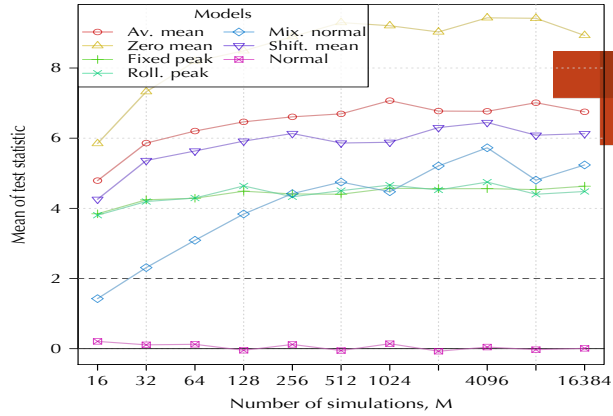
$M \in \mathcal{M} = \{2^i | i \in \{4, \dots, 14\}\} = \{2^4, 2^5, \dots, 2^{14}\} = \{16, 32, \dots, 16384\}$



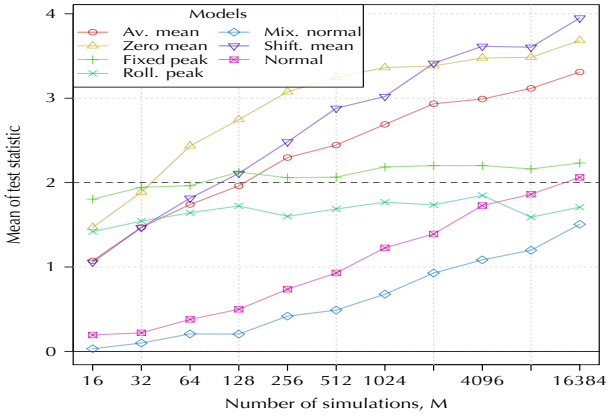
ES



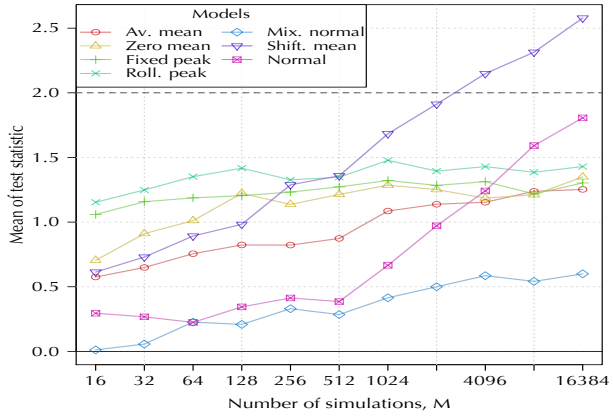
VS



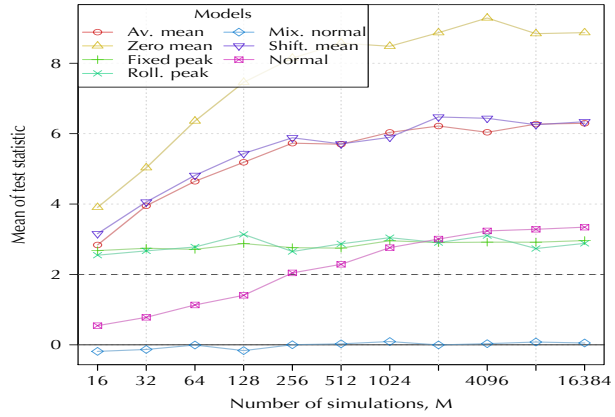
DSS



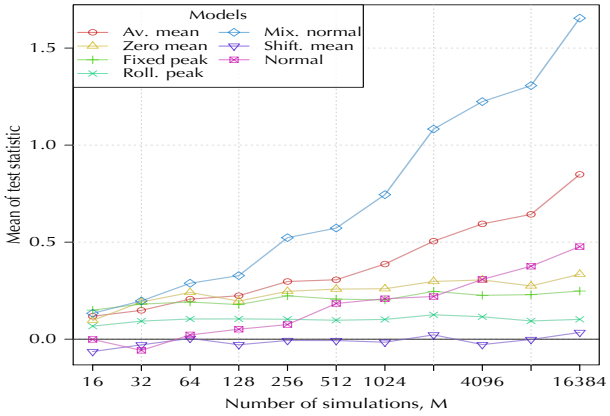
CRPS-CES



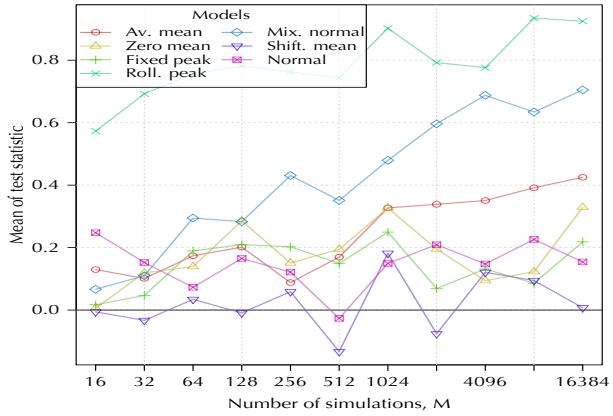
CRPS-CVS



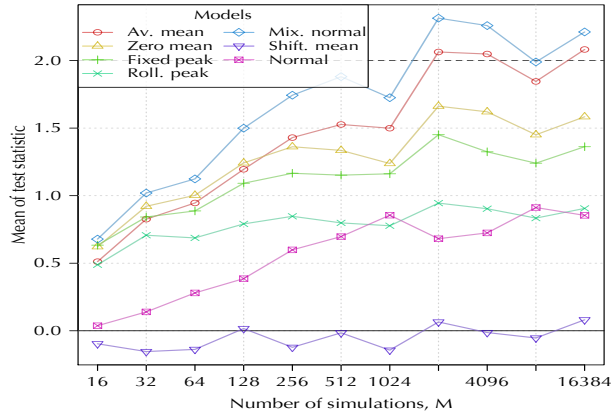
CRPS



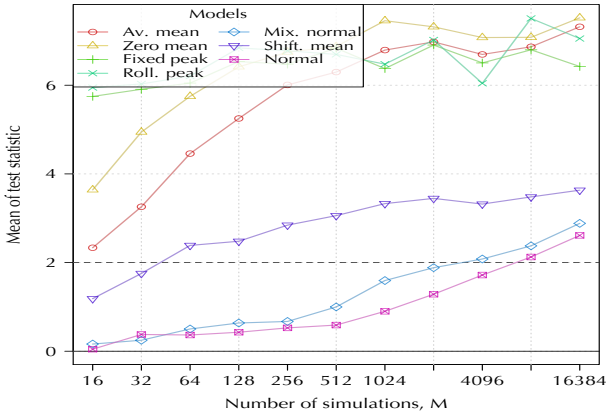
CES



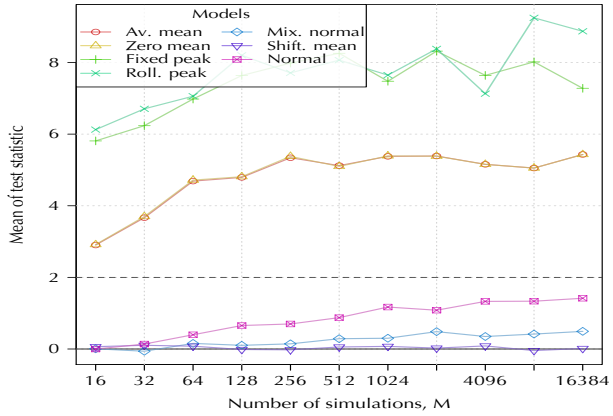
CVS



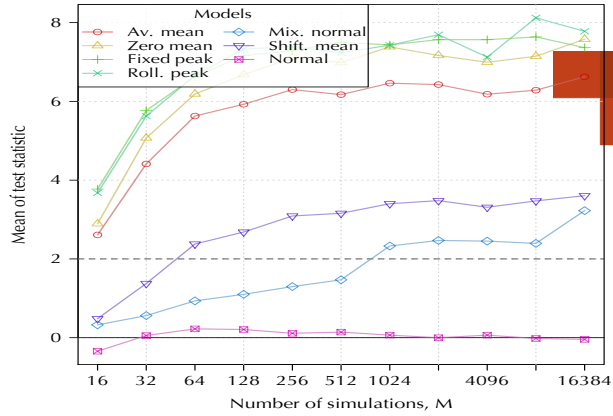
CDSS



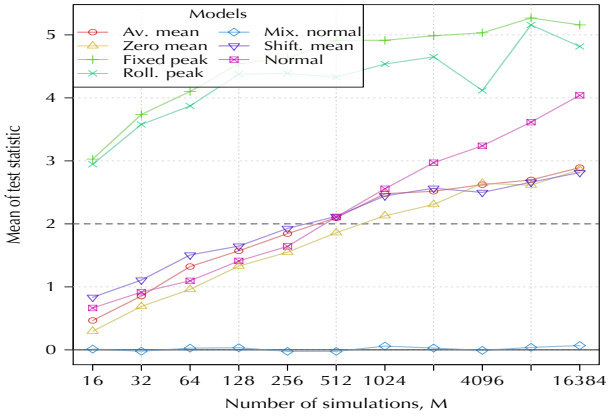
ES



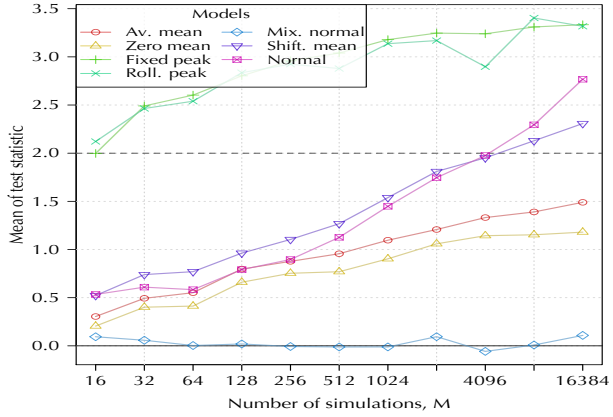
VS



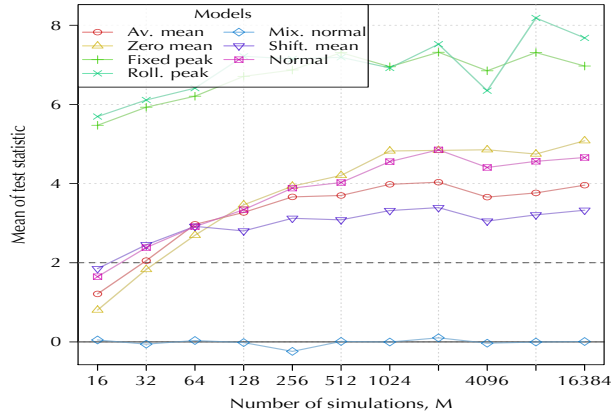
DSS



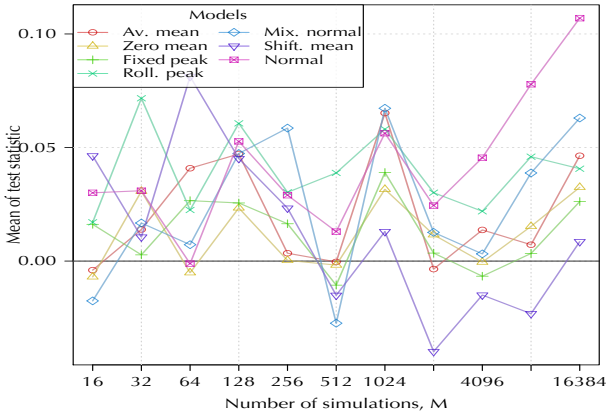
CRPS-CES



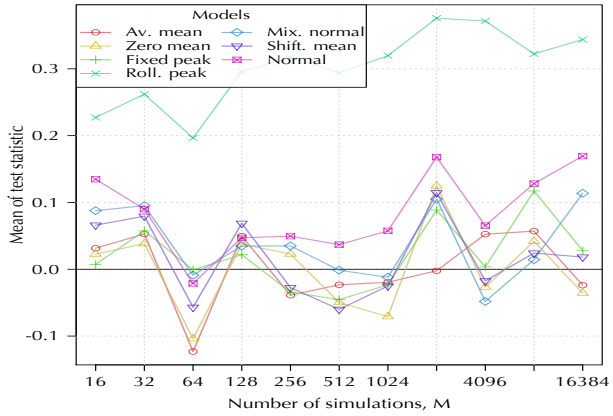
CRPS-CVS



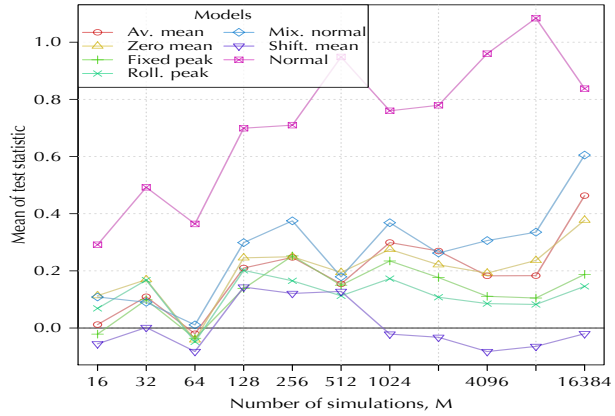
CRPS



CES



CVS



CDSS

Conclusions from the simulation studies

- Energy score is the only considered score which seems to be suitable
- In some cases other scores are slightly better in identifying special features

But:

? Why is the energy score so powerful?

Conclusions from the simulation studies

- Energy score is the only considered score which seems to be suitable
- In some cases other scores are slightly better in identifying special features

But:

- ? Why is the energy score so powerful?
- Reason: Structure of the energy distance

$$d_E(\boldsymbol{X}, \boldsymbol{Y}) = \mathbb{E}\|\boldsymbol{X} - \boldsymbol{Y}\|_2^\beta - \frac{1}{2}\mathbb{E}\|\boldsymbol{X} - \widetilde{\boldsymbol{X}}\|_2^\beta - \frac{1}{2}\mathbb{E}\|\boldsymbol{Y} - \widetilde{\boldsymbol{Y}}\|_2^\beta$$

- d_E yields energy score for observed $\boldsymbol{Y}(\omega) = \boldsymbol{y}$

Properties of the energy distance

- d_E is zero if and only if $\boldsymbol{X} \overset{d}{=} \boldsymbol{Y}$
- special weighted L^2-distance between characteristic functions:

$$d_E(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\pi^{\frac{H+1}{2}}}{\Gamma(\frac{H+1}{2})} \int_{\mathbb{R}^H} \frac{|\varphi_{\boldsymbol{X}}(\boldsymbol{z}) - \varphi_{\boldsymbol{Y}}(\boldsymbol{z})|^2}{\|\boldsymbol{z}\|_2^{H+\beta}} \, d\boldsymbol{z} \tag{8}$$

  for characteristic functions $\varphi_{\boldsymbol{X}}(\boldsymbol{z}) = \mathbb{E}(e^{i\boldsymbol{z}'\boldsymbol{X}})$ and $\varphi_{\boldsymbol{Y}}(\boldsymbol{z}) = \mathbb{E}(e^{i\boldsymbol{z}'\boldsymbol{Y}})$.
- If considering the weighted L^2-distance between $\varphi_{\boldsymbol{X}}$ and $\varphi_{\boldsymbol{Y}}$:

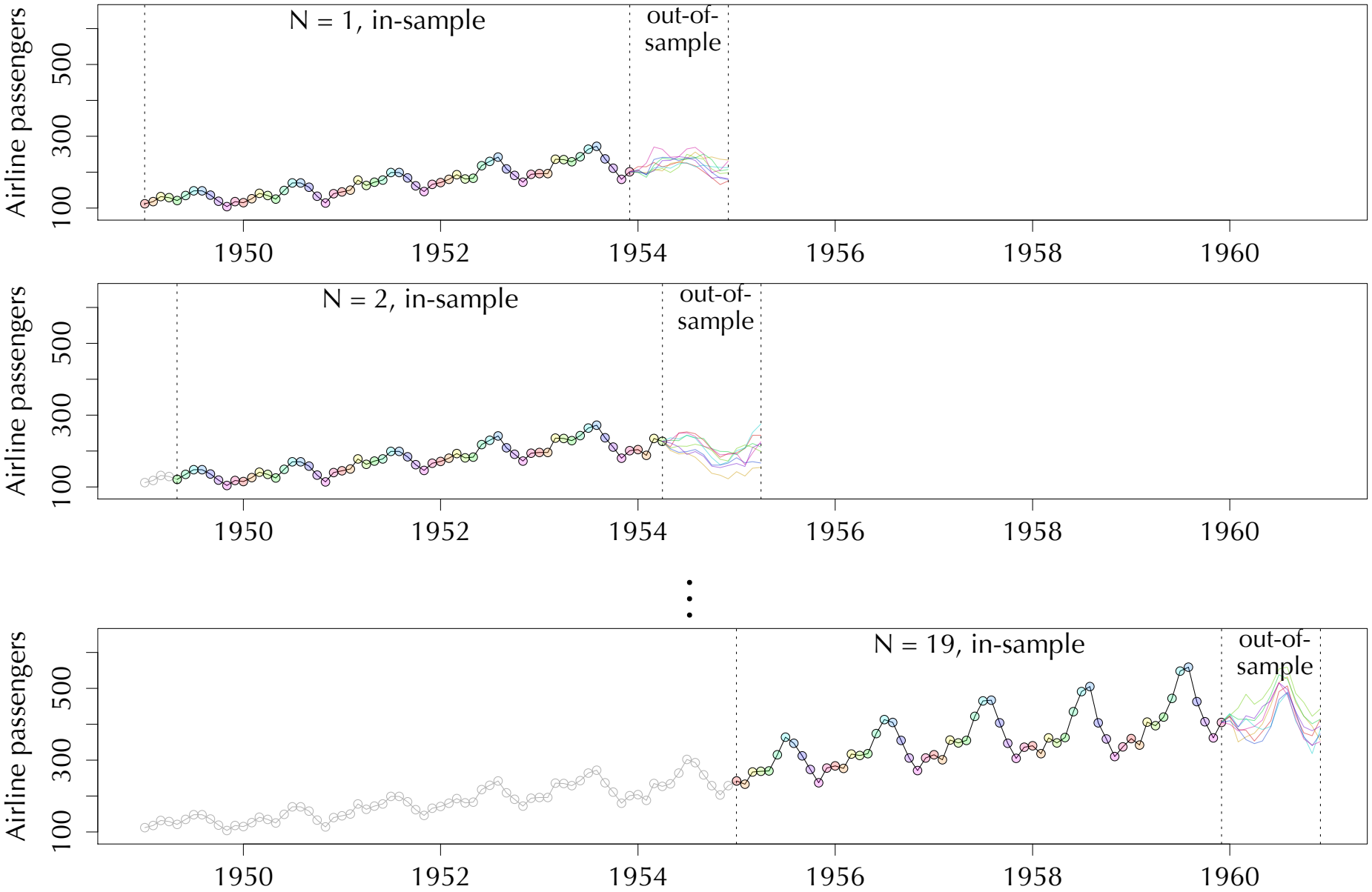
$$C \int_{\mathbb{R}^d} \xi(\boldsymbol{z}) |\varphi_{\boldsymbol{X}}(\boldsymbol{z}) - \varphi_{\boldsymbol{Y}}(\boldsymbol{z})|^2 \, d\boldsymbol{z}$$

  then $\xi(\boldsymbol{z}) = \|\boldsymbol{z}\|_2^{H+\beta}$ is the only choice such that the distance is scale equivariant and rotationally invariant
- d_E measures distances between $\boldsymbol{X}$ and $\boldsymbol{Y}$ in $\mathbb{C} \cong \mathbb{R}^2$ not $\mathbb{R}^H$ $\Rightarrow$ should be efficient for $H > 2$.

Properties of the energy distance

- Allows arbitrary 2-sample test
- Test for multivariate normality (more powerful than standard tests)
- Allows construction of 'β-distance-covariance':
  $\text{dstCov}_\beta(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{C} \int_{\mathbb{R}^H} \int_{\mathbb{R}^H} \frac{|\varphi_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{z},\boldsymbol{v}) - \varphi_{\boldsymbol{X}}(\boldsymbol{z})\varphi_{\boldsymbol{Y}}(\boldsymbol{v})|^2}{\|\boldsymbol{z}\|_2^{\beta+H} \|\boldsymbol{v}\|_2^{\beta+H}} d\,\boldsymbol{v} d\,\boldsymbol{z}$
- β-distance-covariance allows tests for (multivariate) independence(!)

Real data example



Real data example: 9 models

1a) AR(12): $Y_t = \phi_0 + \sum_{k=1}^{12} \phi_k Y_{t-k} + \varepsilon_t$ with ε_t iid and $\mathbb{E}(\varepsilon_t) = 0$.

2a) AR(13): $Y_t = \phi_0 + \sum_{k=1}^{13} \phi_k Y_{t-k} + \varepsilon_t$ with ε_t iid and $\mathbb{E}(\varepsilon_t) = 0$.

3a) AR(p): $Y_t = \phi_0 + \sum_{k=1}^{p} \phi_k Y_{t-k} + \varepsilon_t$ with ε_t iid, $\mathbb{E}(\varepsilon_t) = 0$ and $p \in \{1, \dots, T/2\}$ such that the corresponding Akaike information criterion (AIC) is minimized.

1b) AR(12) as in 1a) but with comonotone residuals (i.e. $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+1})$ have the copula $\boldsymbol{M}_2$)

2b) AR(13) as in 2a) but with comonotone residuals (i.e. $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+1})$ have the copula $\boldsymbol{M}_2$)

3b) AR(p) as in 3a) but with comonotone residuals (i.e. $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+1})$ have the copula $\boldsymbol{M}_2$)

1c) AR(12) as in 1a) but with countermonotone residuals (i.e. $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+1})$ have the copula $\boldsymbol{W}_2$)

2c) AR(13) as in 2a) but with countermonotone residuals (i.e. $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+1})$ have the copula $\boldsymbol{W}_2$)

3c) AR(p) as in 3a) but with countermonotone residuals (i.e. $(\widehat{\varepsilon}_t, \widehat{\varepsilon}_{t+1})$ have the copula $\boldsymbol{W}_2$)

Real data example

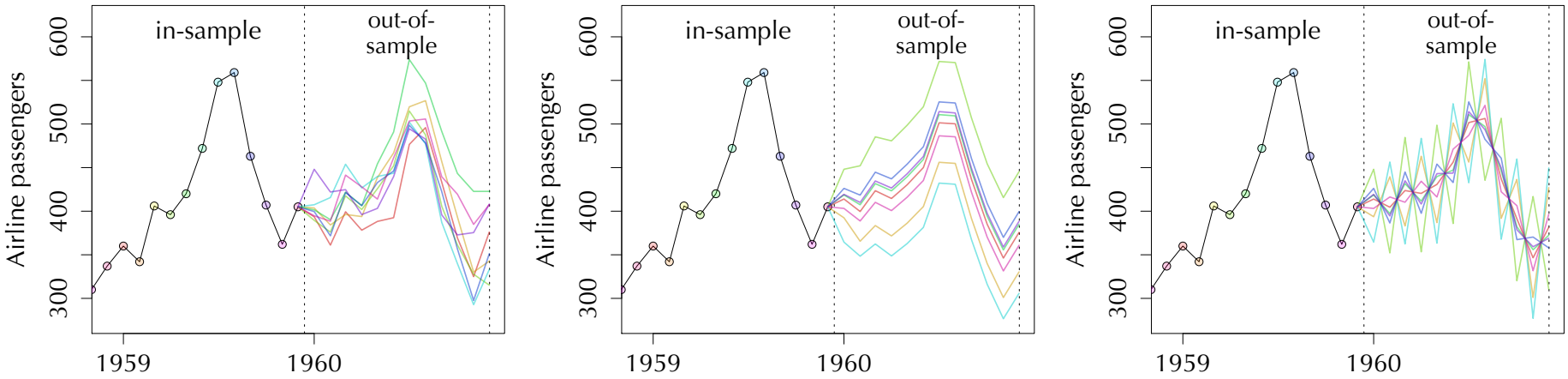


Figure: Illustration of standard (left), comonotone (center) and countermonotone (right) model simulations for the AR(13) with $M = 8$ paths for the last experiment ($N = 19$).

- ensemble size: $M = 2^{16} = 65536$

Real data example: results

Score\Model	AR(12)	AR(13)	AR(p)	AR(12)-M	AR(13)-M	AR(p)-M	AR(12)-W	AR(13)-W	AR(p)-W
ES	120.8	137.5	134.9	197.3	196.2	190.9	203.2	206.6	201.0
VS	158011	120657	111823	205513	133903	122607	175112	123631	114357
DSS	95.92	90.99	91.78	683635	372563	364726	995252	838370	762531
CRPS-CES	3.228	3.887	3.828	8.617	10.240	10.045	8.818	10.537	10.365
CRPS-CVS	0.3346	0.4638	0.4166	0.6357	0.7712	0.7248	0.7989	0.9954	0.9464
CRPS	28.28	33.94	33.36	28.32	33.90	33.36	28.29	33.92	33.35
CES	0.1142	0.1144	0.1146	0.3044	0.3020	0.3010	0.3118	0.3105	0.3106
CVS	0.01184	0.01367	0.01247	0.02247	0.02276	0.02171	0.02826	0.02935	0.02836
CDSS	-26.06	-24.54	-23.78	-	-	-	-	-	-

Table: Score averages $\overline{\text{SC}}$ across the $N = 19$ out-of-sample windows for the considered scores and models. -M indecates models with comonotone residuals, -W for countermonotone residuals.

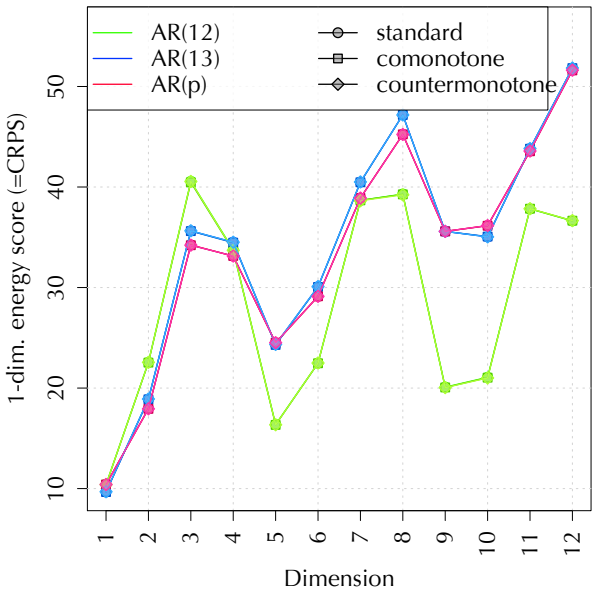
- AR(12) seems to be best concerning, but not uniformly $\Rightarrow$ improvements possible

Real data example: results

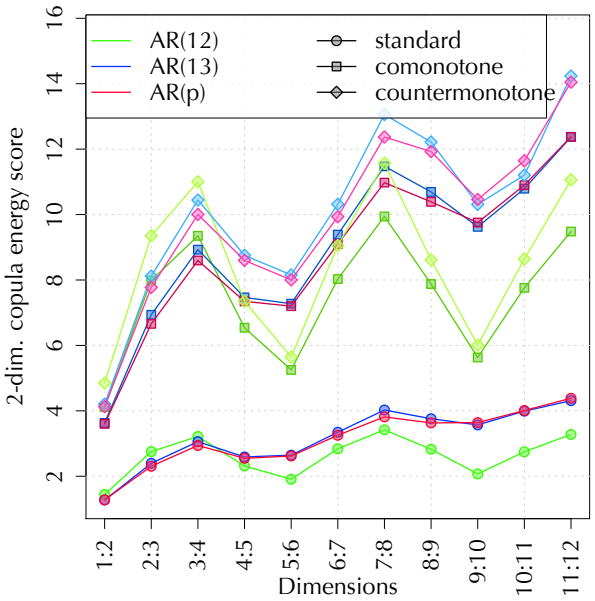
	AR(12)	AR(13)	AR(p)	AR(12)-M	AR(13)-M	AR(p)-M	AR(12)-W	AR(13)-W	AR(p)-W
AR(12)		−4.04 [<0.001]	−2.57 [0.005]	−26.53 [<0.001]	−19.57 [<0.001]	−13.39 [<0.001]	−30.02 [<0.001]	−17.90 [<0.001]	−13.02 [<0.001]
AR(13)	4.04 [>0.999]		0.73 [0.766]	−11.15 [<0.001]	−31.58 [<0.001]	−13.67 [<0.001]	−13.57 [<0.001]	−35.14 [<0.001]	−14.51 [<0.001]
AR(p)	2.57 [0.995]	−0.73 [0.234]		−9.12 [<0.001]	−14.06 [<0.001]	−29.95 [<0.001]	−10.74 [<0.001]	−16.87 [<0.001]	−31.22 [<0.001]
AR(12)-M	26.53 [>0.999]	11.15 [>0.999]	9.12 [>0.999]		0.26 [0.602]	1.08 [0.860]	−7.18 [<0.001]	−1.86 [0.031]	−0.55 [0.292]
AR(13)-M	19.57 [>0.999]	31.58 [>0.999]	14.06 [>0.999]	−0.26 [0.398]		1.39 [0.918]	−1.98 [0.024]	−9.32 [<0.001]	−1.09 [0.138]
AR(p)-M	13.39 [>0.999]	13.67 [>0.999]	29.95 [>0.999]	−1.08 [0.140]	−1.39 [0.082]		−2.28 [0.011]	−4.18 [<0.001]	−9.51 [<0.001]
AR(12)-W	30.02 [>0.999]	13.57 [>0.999]	10.74 [>0.999]	7.18 [>0.999]	1.98 [0.976]	2.28 [0.989]		−0.78 [0.219]	0.36 [0.642]
AR(13)-W	17.90 [>0.999]	35.14 [>0.999]	16.87 [>0.999]	1.86 [0.969]	9.32 [>0.999]	4.18 [>0.999]	0.78 [0.781]		1.39 [0.918]
AR(p)-W	13.02 [>0.999]	14.51 [>0.999]	31.22 [>0.999]	0.55 [0.708]	1.09 [0.862]	9.51 [>0.999]	−0.36 [0.358]	−1.39 [0.082]	

Table: DM-test statistics with corresponding p-value given in squared brackets for the energy score (ES).

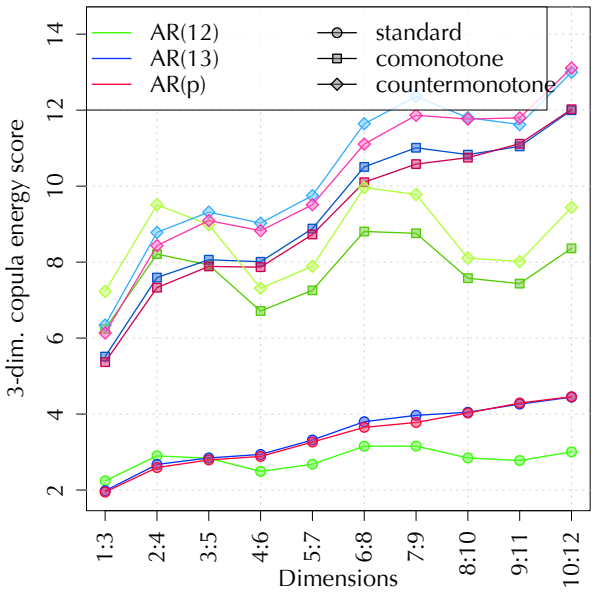
Real data example: results



CRPS



2-dim. CES



3-dim. CES

- CRPS and 2-dim. CES across horizon very useful to detect failures in performance

Summary

- Energy score is suitable distance for multivariate evaluation (in combination with significance tests)
- Ensemble sample size should be as large as computationally feasible
- Additionally consider
  - CRPS for checking individual marginals across the forecasting horizon
  - copula energy score for evaluation 2-way dependencies across the forecasting horizon

Summary

- Energy score is suitable distance for multivariate evaluation (in combination with significance tests)
- Ensemble sample size should be as large as computationally feasible
- Additionally consider
  - CRPS for checking individual marginals across the forecasting horizon
  - copula energy score for evaluation 2-way dependencies across the forecasting horizon

Thank you for your attention.