Assessing the quality of VaR forecasts

"All models are wrong but some might be useful." (G.E.P. Box)

W. Härdle, Z. Hlávka, G. Stahl

Institut für Statistik und Ökonometrie
Humboldt-Universität zu Berlin
email: hlavka@wiwi.hu-berlin.de
www.md-stat.com

Outline of the talk

1. introduction ✓
2. Value-at-Risk
3. define some simple models based on dimension reduction
4. DAX example
5. exceedances and quality of VaR forecasts
6. methods for comparing probability forecasts
7. conclusion: which method is the best one?

Notation

$\Lambda_t \in \mathbb{R}^p$ nominal volumes at time t

P_s the price vector

Π_t is the portfolio

The market value of Π_t at time s is given by

$$\nu_s(\Pi_t) = \Lambda_t^\top P_s.$$

We are interested in modelling the risk of price changes at time t given $P_t = p_t$:

$$\begin{aligned} \nu_{t+h}(\Lambda_t) - \nu_t(\Lambda_t) &= \Lambda_t^\top (P_{t+h} - p_t) \\ &= (\Lambda_t p_t)^\top \frac{(P_{t+h} - p_t)}{p_t} \\ &= w_t^\top R_{t+h}, \end{aligned}$$

where w_t is the exposure and R_{t+h} are the returns.

The conditional distribution

$$\mathcal{L}(L_{t+h}|\mathcal{H}_t)$$

of possible losses $L_{t+h} = -w_t^\top R_{t+h}$ is the object of interest.

$\mathcal{H}_t$, the information available at time t consists in the simplest case of moving window of past returns of the assets.

Value-at-Risk (VaR) is just the α quantile of $\mathcal{L}(L_{t+h}|\mathcal{H}_t)$.

Let $\mathrm{VaR}_i(\alpha, t)$ denote the Value-at-Risk of i-th asset and $\mathrm{VaR} = (\mathrm{VaR}_1, \ldots, \mathrm{VaR}_p)^\top$. Then the Value-at-Risk of the portfolio Π_t is:

$$\mathrm{VaR}(\Pi_t) = u_\alpha \sqrt{w_t^\top \Sigma_t w_t} = \sqrt{\mathrm{VaR}^\top C_t \, \mathrm{VaR}}.$$

Dimension Reduction—Mapping

Using a model based on the equality

$$\mathrm{VaR}(\Pi_t) = \sqrt{\mathrm{VaR}^\top \, C_t \, \mathrm{VaR}}$$

would require the estimation of all the $p(p+1)/2$ parameters. This becomes unfeasible for large number of large portfolios.

There is a need for procedures which would reduce the computation cost of such operations.

We search for a small vector of *risk factors*

$$\widetilde{R} = (\widetilde{R}_1, \ldots, \widetilde{R}_d)$$

such that

$$\mathrm{VaR}(\Pi_t) \approx \mathrm{VaR}(\widetilde{\Pi}_t).$$

This can be formally described as a mapping of the original risk factors onto a lower dimensional space:

$$\mathcal{M} : R \longrightarrow \mathcal{M}(R) = \widetilde{R}.$$

We consider the following simple models:

1. full model with all $p(p+1)/2$ parameters,
2. beta factor model,
3. mapping on principal components,
4. random effects model.

The main focus of the paper is to develop and present methods for comparing these simple methods.

Beta Factor Model

R_{it} is the returns i-th asset.

R_{mt} is the return of an index (e.g. DAX).

The dimension reduction is based on the regression of R_{it}'s on R_{mt}:

$$R_{it} = \beta_i R_{mt} + \epsilon_{it}.$$

It follows that

$$\begin{aligned} \mathrm{Var}(R_{it}) &= \sigma_{it}^2 = \beta_i^2 \sigma_{mt}^2 + \sigma_{\epsilon,it}^2, \\ \mathrm{Cov}(R_{it}, R_{jt}) &= \sigma_{ijt}^2 = \beta_i \beta_j \sigma_{mt}^2 \end{aligned}$$

which implies that

$$w_t^\top \Sigma_t w_t = \sigma_{mt}^2 w_t^\top \beta \beta^\top w_t + w_t^\top D_{\epsilon t} w_t.$$

$$w_t^\top \Sigma_t w_t \quad = \quad \sigma_{mt}^2 w_t^\top \beta\beta^\top w_t + w_t^\top D_{\epsilon t} w_t.$$

This motivates the following two approximations of Σ:

Beta Factor Model I

$$\Sigma_{\beta,1} = \sigma_m^2 \beta\beta^\top.$$

Beta Factor Model II

$$\Sigma_{\beta,2} = \sigma_m^2 \beta\beta^\top + D_\epsilon.$$

Plugging $\Sigma_{\beta,1}$ and $\Sigma_{\beta,2}$ into the formula for $\mathrm{Var}(\mathcal{M}(\Pi))$ leads to:

$$\begin{aligned} \mathrm{VaR}_{\beta,1}(\mathcal{M}(\Pi)) &= \sqrt{\mathrm{VaR}_{\mathcal{M}}^\top(\mathbf{1})\,\mathrm{VaR}_{\mathcal{M}}}, \\ \mathrm{VaR}_{\beta,2}(\mathcal{M}(\Pi)) &= \sqrt{\mathrm{VaR}_{\mathcal{M}}^\top(\mathbf{1} + B^{-1} D_\epsilon B^{-1})\,\mathrm{VaR}_{\mathcal{M}}}. \end{aligned}$$

Principal Components

This method is based on the orthogonalization of the variance matrix:

$$\Sigma = QDQ^\top.$$

The Principal Components transformation $Y = Q^\top R$ satisfies the following property:

$$\Sigma = \operatorname{Var} R = \operatorname{Var} QY = Q(\operatorname{Var} Y)Q^\top = QDQ^\top.$$

Keeping only small number k of principal components allows good approximation of Σ and of the Value-at-Risk:

$$\text{VaR} = \Phi^{-1}(\alpha)\sqrt{w_t^\top Q_k D Q_k^\top w_t}.$$

Calculating the PC's from $\text{Var}(w_t R_t)$ leads to different result (and possibly better approximation of $\text{VaR}(\Pi_t)$).

Random Effects

Assume that the following model holds:

$$R_{it}/\sigma_i = \alpha_t + \epsilon_{it}.$$

It follows that

$$Cov(R_{it}, R_{jt}) = \sigma_i \sigma_j \sigma_\alpha^2$$

which is equivalent to the assumption of constant correlation between all returns in the portfolio.

The dimension of the problem is reduced by imposing constraint of constant correlation on Σ.

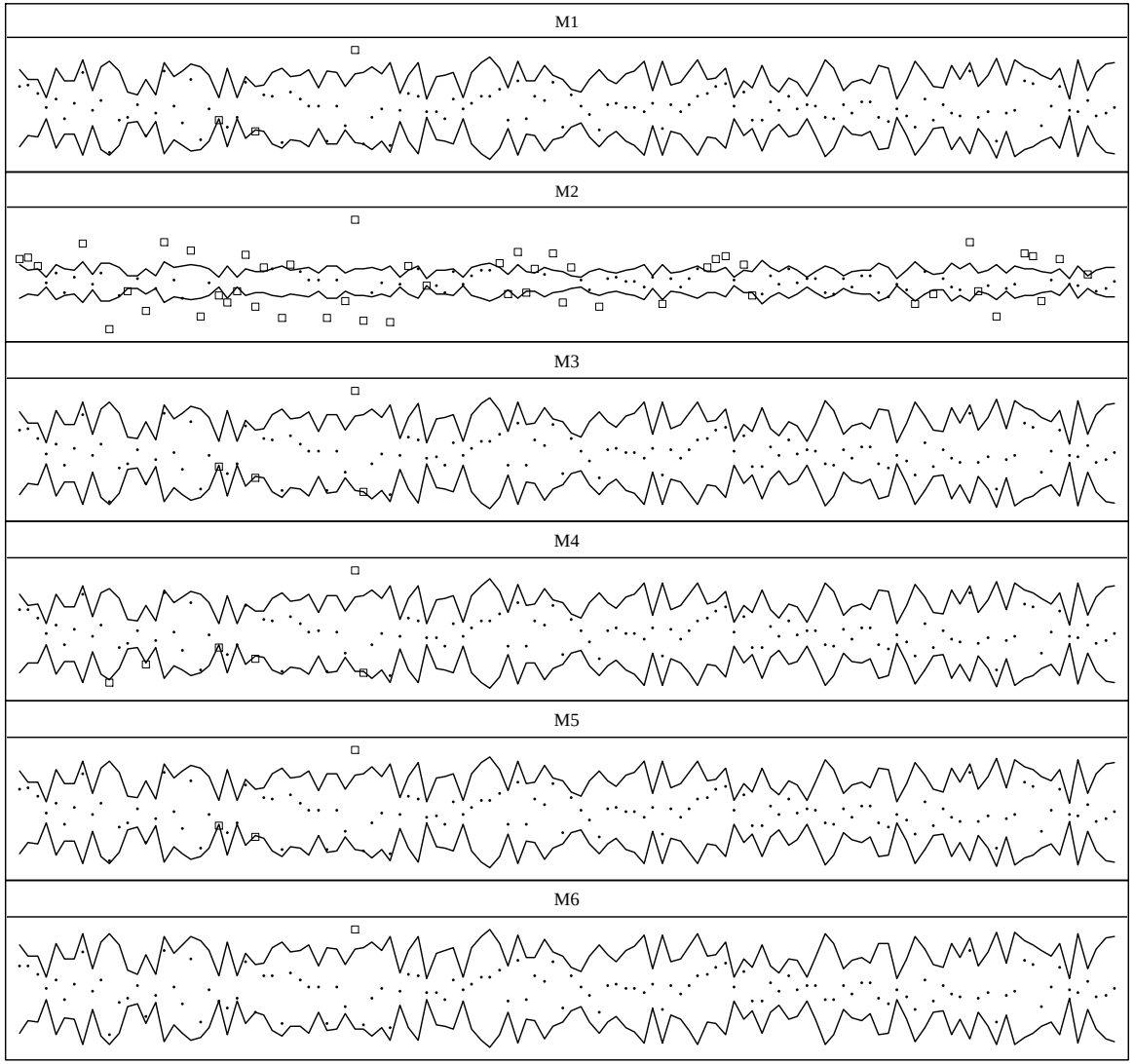
Simple Example

Time series of 18 German stocks + DAX from 1.1.1997–18.6.1999. For each day in 1998 and 1999, we calculated Value-at-Risk using history of 249 observations.

The computational scheme included random changes of the weighting scheme w_t over time.

The following methods were used:

1. saturated model using the covariance matrix of all returns,
2. classical beta-factor model,
3. improved beta-factor model,
4. principal components,
5. weighted principal components $(w_t R_t)$,
6. all off-diagonal elements of the correlation matrix are identical.

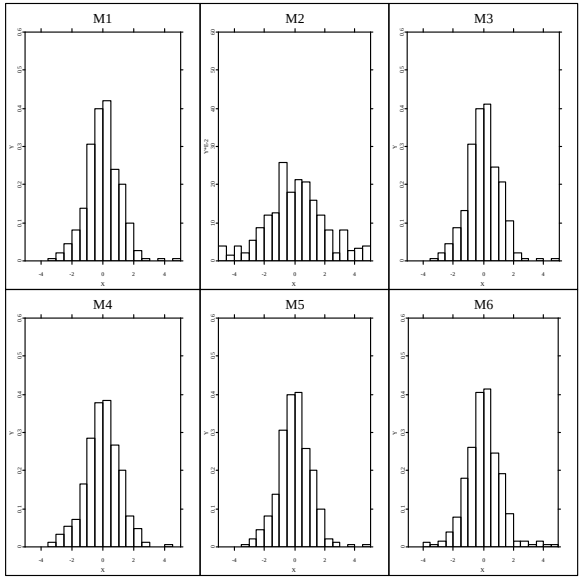

M1
M2
M3
M4
M5
M6

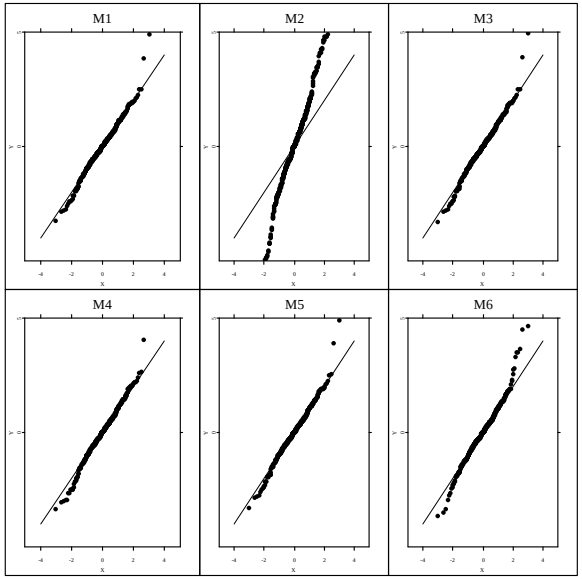
Checking the Quality of the Model

Under the assumptions of normality, the variable

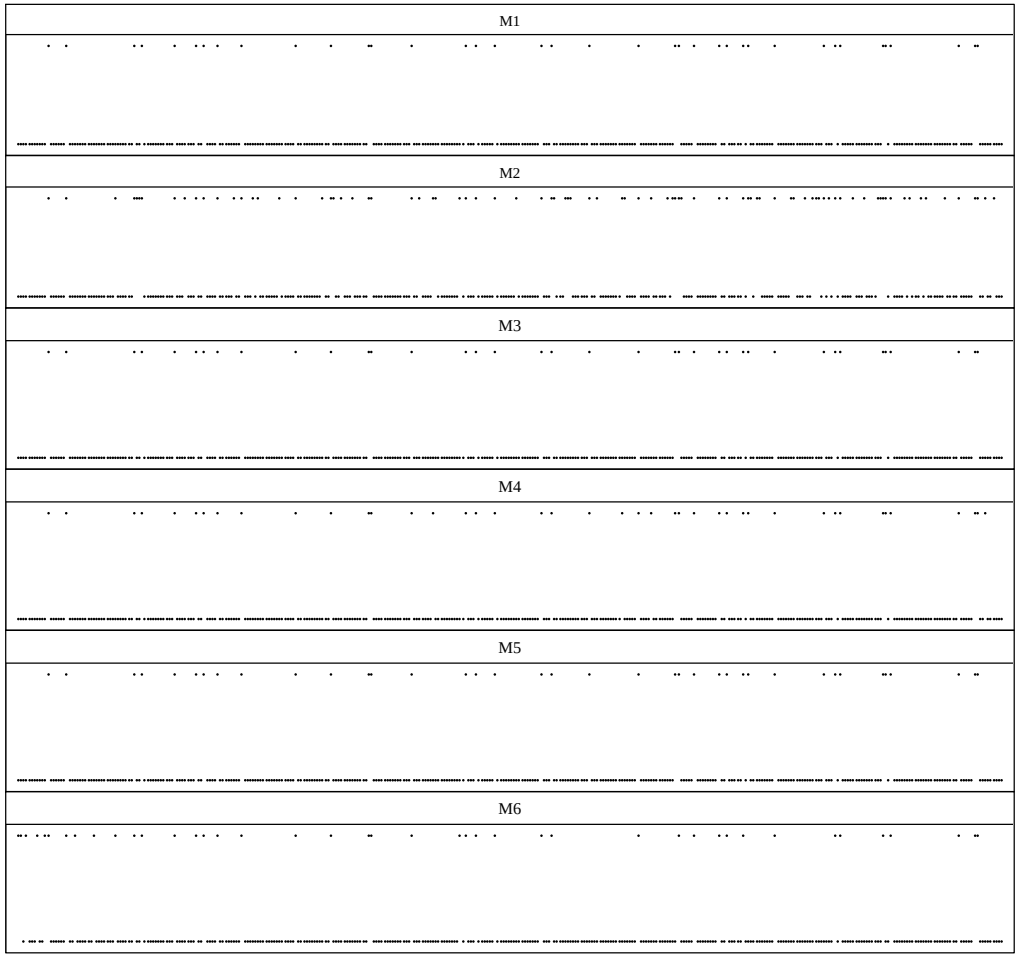
$$L_t / \mathrm{VaR}_{t-1}$$

should have approximately Normal distribution $N(0, 2.33^{-2})$. The assumption of normality and of the variance can be verified using histograms and quantile-quantile plots:





Exceedances of 80% Value-at-Risk



Relative frequencies of exceedances of VaR at level $\alpha = 0.8$:

	M1	M2	M3	M4	M5	M6
1.1998	0.14	0.22	0.14	0.14	0.14	0.14
2.1998	0.17	0.37	0.17	0.17	0.17	0.17
3.1998	0.26	0.52	0.26	0.24	0.26	0.26
4.1998	0.14	0.44	0.14	0.14	0.14	0.14
1.1999	0.38	0.70	0.38	0.38	0.38	0.38
2.1999	0.17	0.63	0.17	0.17	0.17	0.17

Probabilistic Nature of VaR Forecasts

Above presented methods concentrate mainly on the probability of exceedances of the predicted Value-at-Risk.

The VaR estimation is more complex and there is need for other methods of evaluating the quality of the forecasts. VaR model allows to estimate the probability that the loss of portfolio will fall into certain "regions of interest".

In each step, we calculate the probability that the loss will fall into the specified region.

After running the simulations, we investigate the observed joint distribution of the forecasts and the true results of the experiment.

Let us consider, e.g., the interval $I = (50.000, 100.000)$

P_t estimate of the probability that your loss next day will fall in I

$E_t = \boldsymbol{I}$(the loss falls into the interval I)

Brier score

$$BS = \frac{1}{N} \sum_{i=1}^{N} (P_t - E_t)$$

measures the accuracy of the 6 methods:

	M1	M2	M3	M4	M5	M6
BS	0.1541	0.2060	0.1542	0.1552	0.1541	0.1558

The joint probability function of P and E can be decomposed in the following two ways:

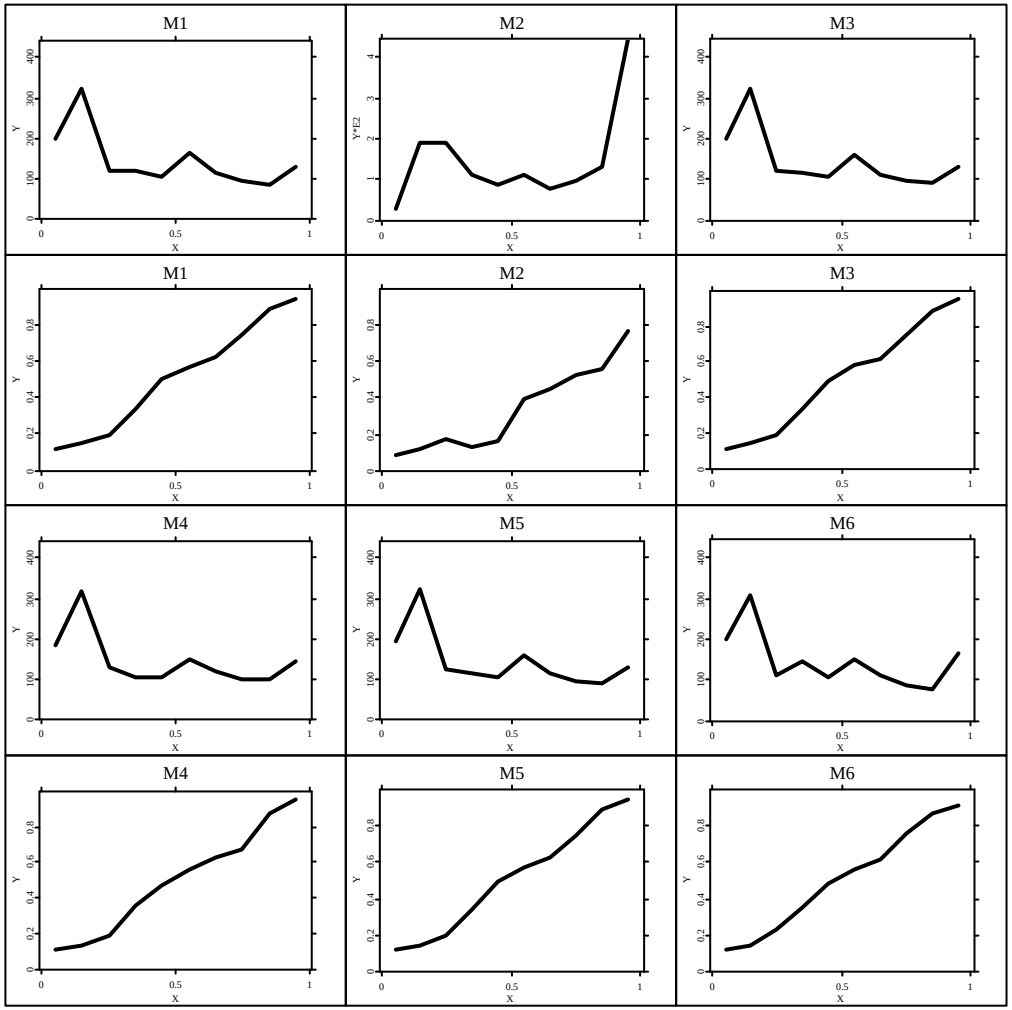
$$\begin{aligned} h(p,e) &= h(e|p)\,h(p) \\ &= h(p|e)\,h(e) \end{aligned}$$

$h(e|p)$ is the distribution of the events (probability of falling into I) conditioned on the value of the probability forecast.

For reasonable VaR forecasts, the relative frequencies of $\{L_t \in I | P_t = p\}$ should lie close to p for each p.

Forecasts with this property are **calibrated**.

Calibration



Calibration

The following table shows:

1. measure of miscalibration,
2. regression based measure of miscalibration.

	M1	M2	M3	M4	M5	M6
$E_p(\mu_{e\|p} - p)^2$	0.00127	0.03718	0.00124	0.00146	0.00124	0.00124
$[\rho_{pe} - (\sigma_p/\sigma_e)]^2$	0.00004	0.01119	0.00005	0.00047	0.00007	0.00005

The information can be displayed also on the **attributes diagram**.

$h(p|e)$ is the distribution of the forecasts conditioned on the loss falling into interval I.

For good probability forecasts, there should be a large difference between the conditional distributions

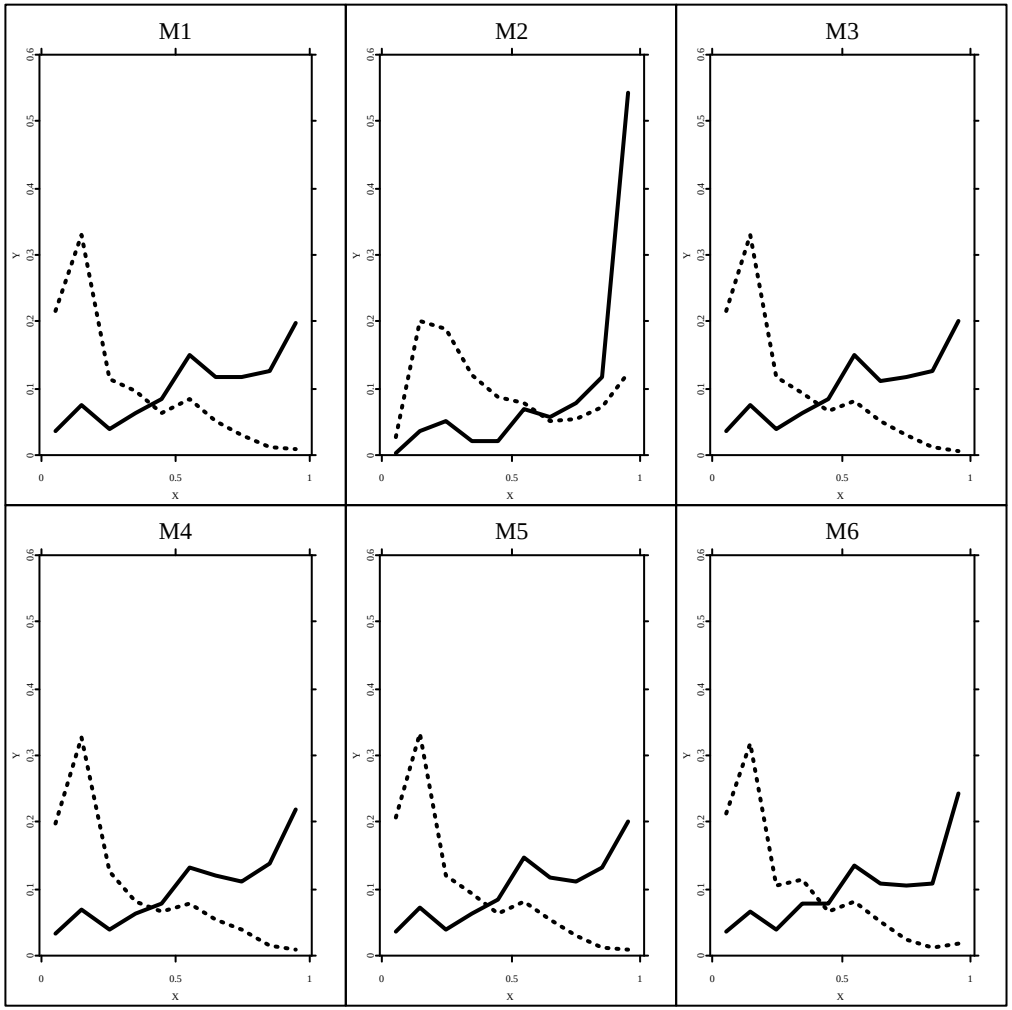
$$h(p|e=0) \quad \text{and} \quad h(p|e=1).$$

The 6 methods can now be compared, e.g., using difference in the conditional means or difference in conditional variances:

	M1	M2	M3	M4	M5	M6
$\mu_{p\|e=1} - \mu_{p\|e=0}$	0.3532	0.3332	0.3536	0.3549	0.3529	0.3516
$\sigma^2_{p\|e=1}$	0.0716	0.0577	0.0718	0.0719	0.0719	0.0716
$\sigma^2_{p\|e=0}$	0.0459	0.0808	0.0459	0.0499	0.0461	0.0458

More informative than the difference in mean or variance is the graphical representation of the two conditional distributions.

Discrimination



There are other useful measures of

refinement degree to which probabilities are close to 0 or 1: σ_p^2

M1	M2	M3	M4	M5	M6
0.0875	0.0981	0.0877	0.0902	0.0877	0.0872

resolution difference between the conditional distributions

$E_p(\mu_{e|p} - \mu_e)^2$

M1	M2	M3	M4	M5	M6
0.0867	0.0722	0.0868	0.0851	0.0864	0.0862

bias $ME = \mu_e - \mu_p$, regression based $[(\mu_e - \mu_p)/\sigma_e]^2$

M1	M2	M3	M4	M5	M6
−0.0102	0.1657	−0.0097	0.0033	−0.0083	−0.0122
0.0004	0.1120	0.0004	0.0001	0.0003	0.0006

skill relative accuracy, Skill Score (SS), measure of potential skill of perfectly calibrated forecasts ρ^2_{fx}

M1	M2	M3	M4	M5	M6
0.3488	0.1543	0.3490	0.3417	0.3477	0.3467
0.3493	0.2775	0.3495	0.3423	0.3480	0.3474

References

Davé, R.D. & Stahl, G. (1998). On the accuracy of VaR estimates based on the variance-covariance approach, in: Bol, Nakhaeizadeh, Vollmer (Eds.), *Risk Measurement, Econometrics and Neural Networks*, Heidelberg: Physica

Murphy, A.H. & Winkler, R.L. (1992). Diagnostic verification of probability forecasts, *International Journal of Forecasting* 7: 435–455