Forecast evaluations for volatile time series: A generalized *Theil* decomposition

Wolfgang Polasek

Institute of Statistics and Econometrics
University of Basel
Holbeinstrasse 12, 4051 Basel, Switzerland
Email: Wolfgang@iso.iso.unibas.ch

May 27, 1999
1. Draft

Abstract

Standard forecasting criteria like the mean square error (MSE) compare point forecasts or a location parameter of the forecasting distribution with actual observations. Such criteria are less suited to comparing forecasts of volatile time series. Therefore we use the average predictive ordinate ($APOC$) criterion which evaluates the ordinate of the predictive distribution. Using the comparison to a no-change forecasting rule, we suggest taking the $RPOC$, the ratio of predictive ordinate criteria. We also suggest comparing two volatile forecasts by a decomposition of the squared distances of ordinates into a *bias*, *variance* and *noise* component. The new criteria are demonstrated for stock indices and exchange rates forecasts.

`Keywords`: Forecasting comparison, predictive ordinates, $APOC$, volatile time series, *Theil* decomposition, MSE matrix.

1 Introduction

Forecast evaluations are important for applied time series analysis. Successful time series models are often measured in terms of their forecasting performance, like the mean squared error (MSE) or the root MSE ($RMSE$). If $x_T(h)$ denotes the forecast of a time series at point T for a h-step ahead observation y_{T+h}, then the MSE is defined as

$$MSE = \frac{1}{H}\sum_{h=1}^{H}(x_T(h) - x_{T+h})^2,$$

where H is the time series horizon for which the forecasts are evaluated. For better interpretation, the square root of the MSE, $RMSE = \sqrt{MSE}$ can be used. There are two possibilities for calculating the MSE:
1) Keep the forecast origin T fixed and evaluate the forecasts for evolving lead times, e.g. $x_T(1), \ldots, x_T(H)$ up to the time horizon H.
2) Keep the lead time fixed, e.g. one-step ahead forecasts, but evaluate the forecasts for evolving time origins, like

$$x_T(1), x_{T+1}(1), \ldots, x_{T+H-1}(1).$$

In this case the information set for the forecasts is expanding, but the forecast errors will be measured for a constant step size into the future. We will denote this measure by

$$MSE_1 = \frac{1}{H}\sum_{h=0}^{H-1}(x_{T+h}(1) - x_{T+h+1})^2$$

if the lead time is one (equi-distant) time step. This concept can be extended for $MSE(k)$ if k-step ahead forecasts have to be evaluated for an expanding information set.
One disadvantage of the MSE is the focus on point forecasts, but the advantage is the simple mathematical tractability and the connection to conditional expectations. Also, the MSE can't discriminate between forecasts coming from homoskedastic or heteroskedastic time series models. In order to evaluate forecasts from a volatile time series model, we have to take into account the whole forecasting distribution.
In this paper we will propose a forecast evaluation criterion which is based on

the whole forecasting (predictive) distribution. At the time T forecast origin, we can make a forecast for the next step by a predictive distribution which means that we quantify the uncertainty of the future observation by a probability distribution. When the new observation realizes, we can calculate the probability density of the new observation. Choosing between two predictive distributions, we prefer the predictive distribution where the new observation falls into a more likely density region. Such a measure is provided by the ordinate of the predictive distribution. A new observation which realizes just at the mode of the predictive distribution attains the maximum ordinate. If the observation realizes in the tails of the predictive distribution then the predictive ordinates will be small. Thus, we suggest taking the average of the predictive ordinates as a new evaluation criterion for forecasts. As before, we can distinguish between two cases:
1) Keep the forecast origin T fixed and consider evolving lead times up to horizon H. Then we get as average ordinate criteria ($APOC$)

$$APOC = \frac{1}{H}(d_T(1) + \ldots + d_T(H)),$$

where $d_T(h)$ is the ordinate of the predictive density $p(x_{T+h}|I_T)$, given the information set I_T, evaluated at the actual observation x_{T+h}.
2) Keep the lead time fixed and evaluate the predictive density for evolving time origins up to horizon H. This leads to the criterion

$$APOC_1 = \frac{1}{H}(d_T(1) + \ldots + d_{T+H-1}(1)),$$

where $d_{T+h}(1)$ is the ordinate of the predictive density $p(x_{T+h}|I_{T+h-1})$ and I_{T+h-1} is the information set up to time $T+h-1$.
The plan of the paper is as follows: In section 3 we introduce the $APOC$ and the relative predictive ordinate criterion for comparing the $APOC$ of the current model with the $APOC$ of the no-change forecasts. Section 2 reviews the mean square error (MSE) criterion and discusses the extension to the multivariate case. Also, the $Theil$ decomposition of the MSE of the point forecasts and for the predictive ordinates are discussed in Section 3. Also we show how the $APOC$ criterion can be used in extended to the multivariate case. In the last section we discuss an example involving the VAR-GARCH model of international stock and exchange rate markets. We conclude with some final remarks.

2 The relative mean square error

We denote by $\hat{x}_{T+1}, \ldots, \hat{x}_{T+H}$ the point forecasts of the current time series model at time T for the horizon H. Then the MSE (mean square error) for horizon H is given by

$$MSE_1 = \frac{1}{H} \sum_{h=1}^{H} (\hat{x}_{T+h} - x_{T+h})^2, \tag{1}$$

where $\hat{x}_{T+h}$ denotes the successive one step ahead forecasts over the horizon H. The no-change mean square error (MSE_1^N) for one step ahead predictions is given by

$$MSE_1^N = \frac{1}{H} \sum_{h=1}^{H} (x_{T+h} - x_{T+h-1})^2.$$

The relative improvement of the root MSE ($RMSE$) over the no-change forecasts is given by

$$RMSE_1 = \sqrt{\frac{MSE_1}{MSE_1^N}}. \tag{2}$$

This criterion can be evaluated over different time horizons H.

2.1 The MSE matrix

In the multivariate case we have H forecasts $\{\hat{\mathbf{x}}_{T+1}, \ldots, \hat{\mathbf{x}}_{T+H}\}$ of dimension M and the forecasting properties can be summarized in a MSE matrix, i.e.

$$MSE_1 = \frac{1}{H} \sum_{h=1}^{H} (\hat{\mathbf{x}}_{T+h} - \mathbf{x}_{T+h})(\hat{\mathbf{x}}_{T+h} - \mathbf{x}_{T+h})', \tag{3}$$

where the $\hat{\mathbf{x}}_{T+h}$ are successive one step ahead forecasts over the horizon H. The no-change MSE matrix for one step ahead prediction is given by

$$MSE_1^N = \frac{1}{H} \sum_{h=1}^{H} (\hat{\mathbf{x}}_{T+h} - \mathbf{x}_{T+h-1})(\hat{\mathbf{x}}_{T+h} - \mathbf{x}_{T+h-1})'.$$

A simple generalization of the univariate ratio of root mean square error in the multivariate case is given by the ratio of the root of the determinants:

$$RMSE_1 = \sqrt{\frac{det(MSE_1)}{det(MSE_1^N)}}. \tag{4}$$

Note: An alternative is to calculate the relative gain in eigenvalues, i.e.

$$\frac{\lambda_1}{\lambda_1^N}, \ldots, \frac{\lambda_p}{\lambda_p^N}$$

with $\lambda_1, \ldots, \lambda_p$ being the eigenvalues of MSE_1 and $\lambda_1^N, \ldots, \lambda_p^N$ from those of MSE_1^N.

2.2 The *Theil* decomposition of the mean square error

The following decomposition of the MSE is explained, e.g. in Pindyck and Rubinfeld (1998) and relates to the inequality coefficients developed in Theil (1961).

Theorem 2.1 The ordinary *Theil* decomposition of the MSE
Let x_t be the actual observations and y_t the forecasts from a model. Then the mean square error decomposition for the horizon H is

$$\begin{aligned} \Delta d^2 &= \frac{1}{H} \sum (x_t - y_t)^2 \\ &= (\bar{x}_t - \bar{y}_t)^2 + (\sigma_x - \sigma_y)^2 + 2(1-\rho)\sigma_x \sigma_y \end{aligned} \tag{5}$$

or

$$\Delta d^2 = bias^2 + variance + noise,$$

where the relative proportions of the decomposition are given as

$$D_{bias} = (\bar{x} - \bar{y})^2 / \Delta d^2,$$

$$D_{var} = (\sigma_x - \sigma_y)^2 / \Delta d^2,$$

$$D_{noise} = (1 - \rho_{xy}) \sigma_x \sigma_y / \Delta d^2 \tag{6}$$

with

$$\sigma_x^2 = \frac{1}{H} \sum (x_t - \bar{x})^2, \quad and \quad \sigma_y^2 = \frac{1}{H} \sum (y_t - \bar{y})^2,$$

$$\bar{x} = \frac{1}{H} \sum x_t, \quad and \quad \bar{y} = \frac{1}{H} \sum y_t,$$

and

$$\rho = \frac{1}{H}\frac{1}{\sigma_x\sigma_y}\sum(x_t - \bar{x})(y_t - \bar{y}).$$

Proof: Follows from standard manipulations.

The *bias* proportion D_{bias} measures the discrepancy of the average forecasts over the forecast horizon. For '*unbiased*' forecasts we expect this proportion to be close to zero.
Any large deviation from zero will cast doubt on the forecasting model.
The *variance* proportion D_{var} shows if the variability of the forecasts is in agreement with the variability of the observations. If the *variance* term is not close to zero then we are faced with a mismatch in the volatility of the observed and predicted series. This could be because the forecasts are over- or underestimating the standard deviation of the observations. To justify forecasts of volatility models, this component should indicate if the volatility component of a model should be changed. Finally, the *noise* proportion D_{noise} measures the unexplained proportion of the MSE. It is called *noise* component because it measures the uncorrelatedness of the current forecasts with comparable observations. If ρ_{xy} is zero, then D_{noise} is close to 1 and we have found satisfactory forecasts. If the *variance* and the *bias* term is close to zero, then the *noise* term will be close to 1. Thus, good forecasting models in the the mean square sense are found if D_{noise} can be made as large as possible.

Theorem 2.2 The alternative $Theil$ decomposition of the MSE
Let y_t be the actual values and x_t the predicted values of the time series. Then the MSE given by

$$\Delta d^2 = \frac{1}{H}\sum(x_t - y_t)^2$$

can be alternatively decomposed as

$$\begin{aligned} \Delta d^2 &= (\bar{x} - \bar{y})^2 + (\sigma_x - \rho\sigma_y)^2 + (1-\rho^2)\sigma_y^2 \\ &= bias^2 + Adj.variance + mmse, \end{aligned} \tag{7}$$

where $MMSE$ stands for minimum mean squared error.
The relative decomposition is given by

$$D_{bias} + D_{adjvar} + D_{mmse} = 1, \tag{8}$$

$$D_{bias} = (\bar{x} - \bar{y})^2 / \Delta d^2,$$
$$D_{adjvar} = (\sigma_x - \rho\sigma_y)^2 / \Delta d^2,$$
$$D_{mmse} = (1 - \rho)\sigma_y^2 / \Delta d^2,$$

with

$$\rho = \frac{1}{H} \frac{1}{\sigma_x \sigma_y} \sum (x_t - \bar{x})(y_t - \bar{y}).$$

Proof: Squaring the nominators of the last two terms in (7) gives

$$\sigma_x^2 + \rho^2\sigma_y^2 - 2\rho\sigma_x\sigma_y + \sigma_y^2 - \rho^2\sigma_y^2,$$

where as squaring the last two terms in (5) gives

$$\sigma_x^2 - 2\sigma_x\sigma_y + \sigma_y^2 + 2\sigma_x\sigma_y + 2\rho\sigma_x\sigma_y.$$

Note that these expressions are equal.
If the MSE is minimized then the minimum MSE is given by the remainder term in (7):

$$MMSE = \underset{x_t}{\text{Min}} \Delta d^2 = (1 - \rho^2)\sigma_y^2. \tag{9}$$

As Clements and Hendry (1998, P. 64) point out, the characteristics of the minimum MSE forecasts are given by the solutions of the first order conditions:

$$\frac{\partial \Delta d^2}{\partial \bar{x}} = -2(\bar{x} - \bar{y}) = 0,$$
$$\frac{\partial \Delta d^2}{\partial \sigma_x} = -2(\sigma_x - \rho\sigma_y) = 0, \tag{10}$$
$$\frac{\partial \Delta d^2}{\partial \rho} = -2\sigma_x\sigma_y = 0.$$

Thus (7) will be minimized if $\bar{x} = \bar{y}$ and $\sigma_x = \rho\sigma_y$. The minimum in (9) is obtained by substitution. The ratio $\Delta d^2/MMSE$ can be interpreted as a scale independent potential MSE which can be reduced.
The difference between the two $Theil$ decompositions lies in the D_{var} and the D_{adjvar} terms. If the correlation coefficients between actual and forecasted values is low, then condition (10) implies that the MSE is minimized if the variance of the forecasts is sufficiently small. The $MMSE$ prefers smooth predictions over erratic ones.

2.3 Forecasting volatilities

Consider the following AR-GARCH model

$$y_t = \beta_0 + \sum_{i=1}^{k} y_{t-i}\beta_i + \varepsilon_t, \quad t = 1, \ldots, T, \quad \varepsilon_t \sim N[0, h_t], \tag{11}$$

$$h_t \;\; = \;\; \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{t-i} + \sum_{i=1}^{q} \phi_i \varepsilon_{t-i}^2. \tag{12}$$

We can extend the MSE decomposition for the predictions of the volatilities in an AR-GARCH model with

$$\hat{y}_t = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{y}_{t-i}\hat{\beta}_i,$$

then the observed volatility is $h_t = \hat{\varepsilon}_t^2 = (y_t - \hat{y}_t)^2$. The predicted volatilities are given by

$$\hat{h}_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \hat{h}_{t-i} + \sum_{i=1}^{q} \phi_i \hat{\varepsilon}_{t-i}^2.$$

If $\hat{h}_t = x_t$ then the following decomposition holds:

$$\begin{aligned} \Delta d^2 &= \frac{1}{H}\sum (h_t - x_t)^2 \\ &= (\bar{h} - \bar{x})^2 + (\sigma_h - \sigma_x)^2 + 2(1-\rho)\sigma_h\sigma_x \end{aligned}$$

or in relative term

$$D_{bias} + D_{var} + D_{noise} = 1 \tag{13}$$

with

$$D_{bias} = (\bar{h} - \bar{x})^2 / \Delta d^2,$$

$$D_{var} = (\sigma_h - \sigma_x)^2 / \Delta d^2,$$

$$D_{noise} = (1 - \rho_{hx})\sigma_h \sigma_x / \Delta d^2$$

with

$$\sigma_h^2 = \frac{1}{H}\sum (h_t - \bar{h})^2, \quad and \quad \sigma_x^2 = \frac{1}{H}\sum (x_t - \bar{x})^2,$$

$$\bar{h} = \frac{1}{H}\sum h_t, \quad and \quad \bar{x} = \frac{1}{H}\sum x_t,$$

and

$$\rho = \frac{1}{H}\frac{1}{\sigma_h \sigma_x}\sum (h_t - \bar{h})(x_t - \bar{x}).$$

The one step ahead forecasts of the AR-GARCH model are given by

$$y_t = x_t^{'}\beta + \varepsilon_t, \quad var(\varepsilon_t) = h_t,$$

or

$$y_t \sim N[x_t^{'}\beta, h_t = z_t^{'}\gamma], \quad t = 1, \ldots, T,$$

where the conditional variance h_t is parameterized as in (12).

2.4 The multivariate MSE decomposition

This section shows how the $Theil$ decomposition of the MSE can be generalized for multivariate time series. First, we define the MSE matrix and then we proceed with the $Theil$ decomposition.

Theorem 2.3: Multivariate MSE decomposition
We consider two multivariate models which produce the forecasts $\mathbf{x}_1, \ldots, \mathbf{x}_H$ and $\mathbf{y}_1, \ldots, \mathbf{y}_H$. The MSE matrix for the horizon H defined as

$$\mathbf{D} = \frac{1}{H}\sum_t (\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)'$$

can be decomposed as

$$\mathbf{D} = \mathbf{M} + (\mathbf{\Sigma}_{\mathbf{xx}}^{1/2} - \mathbf{\Sigma}_{\mathbf{yy}}^{1/2})^2 + 2(\mathbf{\Sigma}_{\mathbf{xx}}^{1/2}\mathbf{\Sigma}_{\mathbf{yy}}^{1/2} - \mathbf{\Sigma}_{\mathbf{xy}}) \tag{14}$$

with

$$\mathbf{M} = (\bar{\mathbf{x}} - \bar{\mathbf{y}})(\bar{\mathbf{x}} - \bar{\mathbf{y}})',$$

$$\mathbf{\Sigma}_{\mathbf{xx}} = \frac{1}{H}\sum_t (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})',$$

$$\mathbf{\Sigma}_{\mathbf{yy}} = \frac{1}{H}\sum_t (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})',$$

$$\mathbf{\Sigma_{xy}} = \frac{1}{H}\sum_t (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{y}_t - \bar{\mathbf{y}})',$$

and $\mathbf{\Sigma}^{-1/2}$ is defined as $\mathbf{\Sigma}^{-1/2} = \mathbf{U} diag(\lambda_1^{1/2}, \ldots, \lambda_M^{1/2})\mathbf{U}^T$, where $EV(\mathbf{\Sigma}) = (\lambda_1, \ldots, \lambda_M)$ are the eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{U}$ is associated matrix of eigenvectors. Now we can define two multivariate MSE decompositions:

a) In terms of the determinants the generalized MSE decomposition is given by

$$D_{bias} = \frac{|\mathbf{M}|}{|\mathbf{D}|},$$

$$D_{var} = \frac{|\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{\Sigma_{yy}}^{1/2}|^2}{|\mathbf{D}|},$$

$$D_{noise} = 1 - D_{bias} - D_{var}.$$

b) In terms of the eigenvalues of $\mathbf{D}$, we use the $1 \times M$ vector of eigenvalues by $EV(\mathbf{D}) = (\lambda_1^D, \ldots, \lambda_M^D)$ for the M-dimensional relative MSE decomposition defined as

$$D^{\lambda}_{mean} = EV(\mathbf{M}) : / : EV(\mathbf{D}),$$

$$D^{\lambda}_{var} = EV(\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{\Sigma_{yy}}^{1/2})^2 : / : EV(\mathbf{D}),$$

$$D^{\lambda}_{noise} = 1 - D^{\lambda}_{mean} - D^{\lambda}_{var},$$

where : / : denotes elementwise division.

Note that all components of the decomposition add up to 1.

Proof: First note that as in the univariate case we find

$$\begin{aligned}
\mathbf{D} &= \frac{1}{H}\Sigma(\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)' \\
&= \frac{1}{H}\Sigma(\bar{\mathbf{x}} - \bar{\mathbf{y}})(\bar{\mathbf{x}} - \bar{\mathbf{y}})' + \frac{1}{H}\Sigma(\mathbf{x}_t - \bar{\mathbf{x}} + \bar{\mathbf{y}} - \mathbf{y}_t)(\mathbf{x}_t - \bar{\mathbf{x}} + \bar{\mathbf{y}} - \mathbf{y}_t)' \\
&= \mathbf{M} + \mathbf{R}.
\end{aligned}$$

Furthermore, the last term can be decomposed into

$$\begin{aligned}
\mathbf{R} &= \frac{1}{H}\Sigma(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})' + \frac{1}{H}\Sigma(\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})' - \frac{2}{H}\Sigma(\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{y}_t - \bar{\mathbf{y}})' \\
&= \mathbf{\Sigma_{xx}} + \mathbf{\Sigma_{yy}} - 2\mathbf{\Sigma_{xy}}.
\end{aligned}$$

For numerical calculations we use

$$(\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{\Sigma_{yy}}^{1/2})^2 = \mathbf{\Sigma_{xx}} + \mathbf{\Sigma_{yy}} - 2\mathbf{\Sigma_{xx}}^{1/2}\mathbf{\Sigma_{yy}}^{1/2}.$$

Substituting all these terms gives (14).
Note: A further multivariate decomposition of the MSE matrix is given by the trace of MSE matrix $\mathbf{D}$. This is equivalent to a summation over the M univariate components:

$$tr\mathbf{D} = \sum_{m=1}^{M} [(\bar{x}_m - \bar{y}_m)^2 + (\sigma_{x,m} - \sigma_{y,m})^2 - 2(1 - \rho_{xy,m}(\sigma_{x,m}\sigma_{y,m})]$$

with

$$\rho_{xy,m} = \frac{1}{\sigma_{x,m}\sigma_{y,m}} \Sigma(x_{t,m} - \bar{x}_m)(y_{t,m} - \bar{y}_m).$$

Since this trace measure depends on the scale of the M components it is of little practical use. It could serve only as a reasonable summary measure for multivariate observations measured on the same scale.
As in the univariate context, the multivariate proportional contributions can be interpreted as before: Good forecasts are those where D_{bias} and D_{var} are close to zero and D_{noise} is close to 1.

3 The Average Predictive Ordinate Criterion ($APOC$)

Consider a time series $\{x_1, \ldots, x_T\}$ for which at time T we calculate H predictive densities for future observations $x_{T+1} \ldots x_{T+H}$, $i.e. f_{T+1}(x), \ldots, f_{T+H}(x)$. The predictive ordinate criterion (POC) is the average predictive ordinate, defined for a horizon of length H as

$$APOC = \frac{1}{H} \sum_{h=1}^{H} f_{T+h}(x_{T+h}|\hat{\theta}_T), \tag{15}$$

where $\hat{\theta}_T$ is the estimated parameter of the time series model for time T. For one-step ahead forecasts we define

$$APOC_1 = \frac{1}{H} \sum_{h=1}^{H} f_{T+h}(x_{T+h}|\hat{\theta}_{T+h-1})$$

e.g. for the one-step ahead normal distribution this is

$$APON_1 = \frac{1}{H} \sum_{h=1}^{H} N[x_{T+h}|\hat{\mu}_{T+h-1}, \hat{\sigma}^2_{T+h-1}].$$

$\hat{\mu}_{T+h}$ and $\hat{\sigma}^2_{T+h}$ are the mean and the variance of the one-step predictive densities at time h of the time horizon H starting at time T.
The relative improvement to no-change forecasts is given by the relative predictive ordinate criterion ($RPOC$) for one-step ahead predictions over the horizon H:

$$RPOC_1 = \frac{APOC_1}{APON_1}, \tag{16}$$

where the $APOC$ of the simpler (or no-change normal distribution) prediction is given by

$$APON_1 = \frac{1}{H} \sum_{h=1}^{H} N[x_{T+h} | x_{T+h-1}, \hat{\sigma}^2]$$

with

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=1}^{T-1} (x_{t+1} - x_t)^2.$$

3.1 The decomposition of the $APOC$

In analogy to the $Theil$ decomposition of the MSE we can decompose the squared distance of the $APOC$ criterion for any predictive ordinate of a model or the no-change forecasts. Let d_t be the predictive ordinate for the ARCH model and let d_t^N be the predictive ordinate for the no-change forecasts assuming a normal distribution

$$\Delta d^2 = \frac{1}{H} \sum (d_t - d_t^N)^2 = (\bar{d} - \bar{d}^N)^2 + (\sigma - \sigma_N)^2 + 2(1-\rho)\sigma\sigma_N, \tag{17}$$

or

$$\Delta d^2 = bias^2 + variance + noise,$$

where $\rho = cov(d_t, d_t^N)/\sigma\sigma_N$ is the correlation between the predictive ordinates of the two models.
In terms of relative proportions the 3 components of the mean squared predictive ordinate ($MSPO$) criterion is

a) $MSPO\ bias$:

$$D_{bias} = (\bar{d} - \bar{d}^N)^2 / \Delta d^2.$$

b) *MSPO variance*:

$$D_{var} = (\sigma - \sigma_N)^2 / \Delta d^2.$$

c) *MSPO noise*:

$$D_{noise} = 1 - D_{bias} - D_{var},$$

with

$$\sigma^2 = \frac{1}{H} \sum (d_t - \bar{d})^2, \quad \bar{d} = \frac{1}{H} \sum d_t,$$

and

$$\sigma_N^2 = \frac{1}{H} \sum (d_t^N - \bar{d}^N)^2, \quad \bar{d}^N = \frac{1}{H} \sum d_t^N.$$

Note that the correlation between the predictive ordinates could be estimated by

$$\rho = \frac{1}{H} \frac{1}{\sigma \sigma_N} \sum (d_t - \bar{d})(d_t^N - \bar{d}^N).$$

In the case of the *Theil* decomposition of the *MSPO* we need a slightly different interpretation of the 3 components. Since we compare no-change forecasts with forecasts from another model, we generally expect that elaborate models (e.g. a GARCH model) forecast better than no-change models. This implies that the predictive ordinates of the elaborate model are larger than the ordinates of the no-change model. Thus, the *bias* term will only be close to zero if the current model is equally good (or bad) as the no-change model. If the current model beats the no-change model, we expect the *bias* term to be large as well.

Interestingly, the *variance* term in the *MSPO* decomposition doesn't imply anything about the volatility in the model. The *variances* of the predictive ordinates are measuring how the actual observations are scattered in terms of the ordinates under the predictive distribution of both models. Thus, the *variance* term should be close to zero. If this is not the case, then the predictive ordinates are differently scattered for both models under consideration. In this case the predictive distributions should be checked as well as the choice of the horizon period. This might be the case because a period of unusual observations could have been observed. The *noise* term can be interpreted like before as the remainder term in the decomposition. Now a good predictive model is not only characterized by a *noise* component close to 1. Since the *bias* term can be large, the *MSPO* can be split into the *bias* and the *noise* component.

3.2 A Multivariate Predictive Ordinate Criterion

Consider a time series $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ for which we want to predict H one-step ahead forecasts at time T :

$$f_{T+1}(\mathbf{x}), \ldots, f_{T+H}(\mathbf{x}).$$

For one-step ahead forecasts at time T and for horizon H we define the average predictive ordinate criterion

$$APOC_1 = \frac{1}{H} \sum_{h=1}^{H} f_{T+h}(\mathbf{x}_{T+h} | \hat{\theta}_{T+h-1}), \tag{18}$$

e.g. for one-step ahead forecasts assuming a multivariate normal distribution, the $APOC$ is

$$APOC_1 = \frac{1}{H} \sum_{h=1}^{H} N[\mathbf{x}_{T+h} | \hat{\mathbf{x}}_{T+h-1}, \hat{\mathbf{\Sigma}}].$$

For no-change prediction we calculate a ratio of average predictive ordinates assuming a multivariate normal distribution

$$RPOC_1 = \frac{APOC_1}{APON_1}, \tag{19}$$

with

$$APON_1 = \frac{1}{H} \sum_{h=1}^{H} N[\mathbf{x}_{T+h} | \mathbf{x}_{T+h-1}, \hat{\mathbf{\Sigma}}],$$

where $\hat{\mathbf{\Sigma}}$ is calculated as the covariance matrix of the first differences of the time series which is defined as

$$\hat{\mathbf{\Sigma}} = \frac{1}{T-1} \sum_{t=1}^{T-1} (\mathbf{x}_{t+1} - \mathbf{x}_t)(\mathbf{x}_{t+1} - \mathbf{x}_t)'.$$

To demonstrate the application of the above formulas we consider the following

Example: Consider the AR(1)-ARCH(1,1) model with $x_t = (1, y_{t-1})$ and $z_t = (1, h_{t-1}, \varepsilon_{t-1}^2)$.

For $h = 1$ the one step ahead predictions are

$$\mu_{t+1} = E(y_{t+1}) = \beta_0 + \beta_1 y_t,$$
$$h_{t+1} = Var(y_{t+1}) = \gamma_0 + \gamma_1 h_t + \gamma_2 \varepsilon_t^2,$$

with the residual

$$\begin{aligned} \varepsilon_t &= y_t - \mu_t \\ &= y_t - \beta_0 - \beta_1 y_{t-1}. \end{aligned}$$

For $h = 2$ the forecasting equations are

$$\mu_{t+2} = E(y_{t+2}) = \beta_0 + \beta_1 y_{t+1},$$
$$h_{t+2} = Var(y_{t+2}) = \gamma_0 + \gamma_1 h_{t+1} + \gamma_2 \varepsilon_{t+1}^2,$$
$$\varepsilon_{t+1} = y_{t+1} - \beta_0 - \beta_1 y_t,$$

and similarly for higher order predictions.

4 Forecasting exchange rates and stock returns

Financial time series like stock returns or returns on exchange rates have been modeled in recent times by a wide variety of ARCH models and this has led to the discussion of how to compare forecast performances of volatile time series.

Therefore we consider the forecasts of daily exchange rates of the Deutsche Mark and the Japanese Yen against the US dollar and the DM/Yen exchange rates and stock returns in the US, Germany and Japan.

Exchange rates exhibit volatile behavior; we have therefore estimated univariate AR-GARCH(1,1) models for the returns of the stock market and the exchange rates for Japan, German and the US in Tables 2 and 3. Table 2 shows the $RPOC$ for daily one-step ahead forecasts starting on June 12, 1998 for 10 days. The smallest improvement over the no-change prediction can be seen for the one-step forecasts for exchange rates and for stock indices. The picture changes for large time horizons, starting from a horizon length of two days, where the improvement in ordinate length is at least 50%. The

maximum improvement is attained with $RPOC = 2.2$ for horizon length $H = 5$ for the US$/DM exchange rate. For stock indices the improvement in the $RPOC$ is in general a little bit higher. The best improvement can be found for the Nikkei index for horizon length $H = 7$ ($RPOC = 2.8$). It is interesting to note that the $RPOC$ increases only for the Nikkei index while $RPOC$ decreases for the DAX and the Dow Jones indices after the horizon length $H = 2$.
The picture is different if we look at the relative mean square error ($RMSE$) in Table 3. The $RMSE$ seems to stabilize for large horizon length H and gives improvements of about 50% for the DM/Yen, 35% for the US$/Yen and 25% for the US$/DM.
The $RMSE$ behavior is again different for stock indices. The best $RMSE$ can be found in long horizons H for the Nikkei index, followed by the DAX and the Dow Jones indices. This result is similar to the analysis for the $RPOC$ criterion.
Table 4 shows the $MSPO$ decomposition for AR(1)-GARCH(1,1) models for exchange rates and stock returns individually. The *bias* proportion for exchange rates lies between 11 and 16% and the rest of the decomposition lies in the *bias* component. This result makes sense, since we see that the average predictive ordinate for the AR-GARCH model is larger than the one for the no-change model, but the variances are about the same. The picture is different for the returns of the stock indices. Only the Nikkei index shows the same behavior as the exchange rates. For the DAX and the Dow Jones indices the *bias* proportion lies between 4 and 6% while the *variance* proportion is large: 24% for the DAX and 32% for the Dow Jones returns. Also we see from the first line in Table 4 that the $MSPO$ is two to three times larger than the $MSPO$ of the other time series. This shows that we have serious mismatch between the observed values and the predictive distributions for the returns in the evaluation period June 12 to June 22, 1998. Tables 10 and 11 show the multivariate gain in predictive ordinates ($RPOC$) for one-step ahead forecasts of the VAR(1) model of exchange rates and the univariate and multivariate $RMSE$. Table 10 shows the $RPOC$ for the 3-dimensional forecasting distribution and we can see that the $RPOC$ increases over a horizon period up to six days; after day six the predictive power of the model declines. Not surprising is the low performance of the point forecasts for the relative gain in the root mean square error. The improvement lies between 27 and 52% below the no-change root MSE. Also, the picture is confirmed for

the 3-dimensional model, where we compare the improvement with regard to the root determinant. The last column of Table 10 shows that the improvements lie between 40 and 48% percent. Table 11 contains the above analysis for the stock returns. The first column shows that the $RPOC$ for the VAR(1) model increases only up to lag three and then starts decreasing slowly. The $RMSE$s for the three stock returns are decreasing to a smaller degree than for the exchange rates. The $RMSE$ of the 3-dimensional system lies between 28 and 41%. The $APOC$ decomposition for the 3-dimensional VAR(1) model is displayed in Table 12. Now the squared distances for exchange rates and stock returns are about the same size and the *bias* proportion is about 1/3, leaving 2/3 for the *noise* term. This shows that substantial predictive improvements can be found in a multivariate model.

5 Conclusions

The paper proposes an alternative for the usual evaluations of forecasts by the MSE or the $RMSE$. The new average predictive ordinate criterion ($APOC$) evaluates the actual observation given a parametric conditional predictive density. This concept allows the comparison of forecasts of volatile time series, since it takes into account the quality of the predictions if the variances change over time. This $APOC$ criterion can be compared with a no-change forecast for a normal distribution and leads to the relative predictive ordinate criterion ($RPOC$) which is a measure of increase or decrease of the predictability between two models. The $APOC$ and $RPOC$ are also extended to the multivariate case for joint predictions.

References

[1] Christoffersen, P.F. and Diebold, F.X. (1996), Further results on forecasting and model selection under asymmetric loss, Journal of Applied Economics, 11, 561-572.

[2] Clements, M.P. and Hendry, D.F. (1998), Forecasting Economic Time Series, Cambridge University Press.

[3] Diebold, F.X. and Mariano, R. (1995), Comparing Predictive Accuracy, JBES 13, 253-265.

[4] Granger, C.W.J. (1998), Comments on the Evaluation of Economic Models and of Forecasts, UCSD, San Diego, mimeo.

[5] Granger, C.W.J. and Newbold, P. (1996), Forecasting Economic Time Series, (2nd ed.) Academic Press, San Diego.

[6] Pindyck, R.S. and Rubinfeld, D.C. (1998), Econometric Models and Economic Forecasts, (4th ed) McGraw-Hill, Boston.

[7] Polasek, W. and Ren, L. (1998), Volatility analysis during the Asia crisis: A multivariate GARCH-M model for stock returns in the US, Germany and Japan, University of Basel, *http://www.unibas.ch/iso*.

[8] Theil, H. (1961), Economic Forecasts and Policy, North Holland, Amsterdam.

k	r	p	q	Marginal likelihood					
				US$/DM	US$/Yen	DM/Yen	Nikkei	DAX	Dow Jones
1	1	1	1	-674.4645	-609.9565	-629.8232	-607.3409	-654.2450	-637.5071
1	2	1	1	-636.3213	-550.2960	-572.8846	-673.2638	-639.5340	-597.2565
1	1	2	1	-696.6606	-520.3694	-640.3129	-661.6067	-625.3239	-630.8174
1	1	1	2	-673.4059	-492.0668*	-578.4710	-677.2201	-663.6978	-584.5854
1	1	2	2	-644.8745	-596.9241	-611.4474	-494.4379*	-610.9003	-635.0603
2	1	2	2	-629.5650	-601.5632	-552.8091	-546.2119	-587.0933*	-651.9961
2	1	1	1	-613.5182*	-588.4128	-591.0394	-685.4334	-646.1929	-571.2153
2	1	2	1	-639.3854	-571.6018	-545.5851*	-550.5207	-684.1608	-561.1770*
2	2	2	2	-669.0917	-565.1104	-508.5802	-590.0872	-629.6237*	-653.3686

Table 1: The marginal likelihood for the AR(k)-GARCH(p,q)-M(r) model ($US\$/DM$,US$/Yen, DM/Yen, Nikkei, DAX and Dow Jones) from June, 21, 1996 to June, 12, 1998

H	US$/DM	US$/Yen	DM/Yen	Nikkei	DAX	Dow Jones
1	1.1250	1.2260	1.1384	1.3028	1.6392	1.3777
2	1.4389	1.2647	1.3313	1.5965	1.7129	1.3343
3	1.4394	1.4315	1.6757	1.5747	1.7245	1.3144
4	1.4517	1.4338	1.5869	1.5558	1.7787	1.3019
5	1.4424	1.4407	1.5720	1.5333	1.7068	1.2812
6	1.4398	1.4754	1.5677	1.5242	1.6975	1.2718
7	1.4220	1.4256	1.5552	1.4708	1.6828	1.2663
8	1.3605	1.4234	1.4984	1.4600	1.6373	1.2585
9	1.2878	1.3589	1.4810	1.4203	1.6283	1.2354
10	1.2726	1.2641	1.4798	1.4125	1.6189	1.2235

Table 2: Relative predictive ordinate criterion $RPOC_1$ of AR-GARCH-M model for daily exchange rates and stock indices from June, 21, 1996 to June, 12, 1998

H	US$/DM	US$/Yen	DM/Yen	Nikkei	DAX	Dow Jones
1	0.6467	0.8153	0.7310	0.5303	0.5204	0.8352
2	0.7565	0.9176	0.7552	0.6115	0.5581	0.7338
3	0.7687	0.7419	0.6464	0.6293	0.6249	0.6835
4	0.7755	0.7371	0.6312	0.6387	0.6324	0.6450
5	0.7937	0.7125	0.7247	0.6584	0.6508	0.6244
6	0.8076	0.9742	0.8164	0.6655	0.6475	0.5843
7	0.8427	0.8364	0.7522	0.7656	0.6630	0.5642
8	0.9117	0.8302	0.7497	0.7902	0.7730	0.5476
9	0.9321	0.8148	0.7250	0.8799	0.7200	0.5107
10	0.9173	0.7837	0.7176	0.7887	0.7168	0.4915

Table 3: The relative mean square error of AR-GARCH-M model for daily exchange rates and stock indices from June, 21, 1996 to June, 12, 1998

	US$/DM	US$/Yen	DM/Yen	Nikkei	DAX	Dow Jones
Δd^2	0.8165	1.1185	0.8861	0.8375	2.0506	2.8296
D_{bias}	0.1611	0.1153	0.1120	0.1820	0.0622	0.0403
D_{var}	0.0023	0.0060	0.0050	0.0414	0.2265	0.3041
D_{noise}	0.8376	0.8787	0.8830	0.7766	0.7113	0.6556
σ^2	0.7430	1.1365	0.7722	1.1414	1.6400	2.2462
$\bar{d}$	4.6861	4.4841	4.5003	3.6980	3.8174	4.1001
σ_N^2	0.6839	1.0305	0.7046	0.9421	2.3215	3.2329
$\bar{d}^N$	4.3201	4.1251	4.1756	3.3022	3.4516	3.7622
ρ	0.5222	0.5440	0.4719	0.6763	0.5239	0.5128

Table 4: The decomposition of *APOC* for AR-GARCH-M model for daily exchange rates and stock indices from June, 21, 1996 to June, 12, 1998

	US$/DM	US$/Yen	DM/Yen	Nikkei	DAX	Dow Jones
Δd^2	0.000033	0.000075	0.000099	0.001213	0.000450	0.000227
D_{bias}	0.032622	0.003052	0.093545	0.013030	0.037227	0.161651
D_{var}	0.078032	0.008306	0.009920	0.008668	0.075611	0.042444
D_{noise}	0.889346	0.988641	0.896535	0.978302	0.887163	0.795904
D_{adjvar}	0.019886	0.348702	0.135186	0.076497	0.231242	0.212792
D_{mmse}	0.947492	0.648246	0.771269	0.910473	0.000288	0.625557
σ_x^2	0.000006	0.000039	0.000029	0.000357	0.000124	0.000081
$\bar{x}$	0.000532	-0.002886	-0.000933	0.001058	0.001014	0.003367
σ_y^2	0.000017	0.000050	0.000041	0.000490	0.000289	0.000146
$\bar{y}$	0.001567	-0.002408	-0.003971	0.005034	0.003081	0.002691
ρ	0.397616	0.159952	0.269685	0.418399	0.731531	0.169651

Table 5: The returns MSE decomposition for AR-GARCH-M forecasts of daily exchange rates and stock indices from June, 21, 1996 to June, 12, 1998

	US$/DM	US$/Yen	DM/Yen	Nikkei	DAX	Dow Jones
Δd^2	0.0085	0.0078	0.0027	0.0045	0.0036	0.0032
D_{bias}	0.1122	0.0097	0.3959	0.0422	0.0001	0.1155
D_{var}	0.0159	0.0202	0.0061	0.0010	0.0199	0.0024
D_{noise}	0.8719	0.9702	0.5980	0.9569	0.9799	0.8821
D_{adjvar}	0.0812	0.0582	0.1505	0.0487	0.4869	0.2952
D_{mmse}	0.8066	0.9322	0.4535	0.9091	0.5130	0.5893
$\sigma_{\bar{h}}^2$	0.0022	0.0021	0.0011	0.0014	0.0020	0.0013
$\bar{h}$	0.0507	0.0597	0.0394	0.0721	0.0584	0.0577
σ_x^2	0.0034	0.0034	0.0014	0.0013	0.0013	0.0011
$\bar{x}$	0.0815	0.0510	0.0721	0.0583	0.0577	0.0770
ρ	0.3522	0.4194	0.3528	0.6386	0.0817	0.1642

Table 6: The forecasts of volatilities MSE decomposition for AR-GARCH-M model for daily exchange rates and stock indices from June, 21, 1996 to June, 12, 1998

k	r	p	q	Marginal likelihood	
1	1	1	1	-4188.5511	-4124.1339
2	1	1	1	-3902.6256	-4567.3613
1	2	1	1	-3783.0488	-4487.5251
1	1	2	1	-4425.1437	-3764.4254
1	1	1	2	-3526.0746	-3959.6954
1	1	2	2	-3271.3936*	-3169.1208*
2	1	2	2	-3702.6883	-3432.3842
2	1	1	1	-3902.6256	-4567.3613
2	1	2	1	-4246.0742	-4347.4885
2	2	2	2	-4372.8123	-4687.9028

Table 7: The marginal likelihood for the VAR(k)-GARCH(p,q)-M(r) model (for y_t^1 = US$/DM, y_t^2 = US$/Yen, y_t^3 = DM/Yen and y_t^1 = Nikkei, y_t^2 = DAX, y_t^3 = Dow Jones)

H	$RPOC_1$	$RMSE_1^1$ =US$/DM	$RMSE_1^2$ =US$/Yen	$RMSE_1^3$ =DM/Yen
1	15.3543	0.6884	0.6682	0.6316
2	19.6335	0.6524	0.6246	0.6291
3	22.0070	0.6239	0.6166	0.6053
4	24.3516	0.5801	0.5837	0.5977
5	25.7497	0.5283	0.5647	0.5111
6	30.5997	0.5133	0.5039	0.4661
7	18.1786	0.5553	0.5443	0.5350
8	17.5643	0.6489	0.6778	0.5479
9	16.8179	0.7507	0.6888	0.5538

Table 8: The multivariate relative predictive ordinate criterion $RPOC_1$ and the relative mean square error of VAR-GARCH-M(1,1,2,2) model for daily exchange rates for $\mathbf{x}_t^1$ =US$/DM, $\mathbf{x}_t^2$ =US$/Yen and $\mathbf{x}_t^3$ =DM/Yen from June, 21, 1996 to June, 12, 1998

H	$RPOC_1$	$RMSE_1^1$ =Nikkei	$RMSE_1^2$ =DAX	$RMSE_1^3$ =Dow Jones
1	5.7065	0.6318	0.9064	0.7814
2	6.5556	0.8098	0.8403	0.6490
3	7.2495	0.6017	0.8264	0.5893
4	5.4773	0.8805	0.9089	0.6821
5	4.8234	0.7404	0.8609	0.7518
6	4.0215	0.7773	0.9291	0.8383
7	3.2283	0.7929	0.9524	0.6838
8	2.6568	0.7842	0.9772	0.7469
9	2.5859	0.7782	0.9101	0.6923

Table 9: The multivariate relative predictive ordinate criterion $RPOC_1$ and the relative mean square error of VAR-GARCH-M(1,1,2,2) model for daily Stock indices of Japan, German and USA ($\mathbf{x}_t^1$ =Nikkei, $\mathbf{x}_t^2$ =DAX and $\mathbf{x}_t^3$ =Dow Jones) from June, 21, 1996 to June, 12, 1998

H	$RPOC_1$	$RMSE_1^1$ US$/DM	$RMSE_1^2$ US$/Yen	$RMSE_1^3$ DM/Yen	$RMSE$
1	1.3288	0.5220	0.6227	0.6231	0.5875
2	1.3545	0.6197	0.5764	0.5818	0.5719
3	1.4582	0.5305	0.5889	0.5158	0.5638
4	1.5617	0.6985	0.5592	0.4953	0.5402
5	1.5928	0.6475	0.5749	0.4832	0.5357
6	1.9192	0.4840	0.5171	0.4673	0.5297
7	1.6533	0.6623	0.7300	0.5299	0.5889
8	1.4416	0.6561	0.6695	0.5342	0.6077
9	1.3243	0.7462	0.6754	0.5400	0.6147

Table 10: The multivariate relative predictive ordinate criterion $RPOC_1$ and the relative mean square error of the VAR(1) model for daily exchange rates for $\mathbf{x}_t^1$ =US$/DM, $\mathbf{x}_t^2$ =US$/Yen and $\mathbf{x}_t^3$ =DM/Yen from June, 21, 1996 to June, 12, 1998

H	$RPOC_1$	$RMSE_1^1$ Nikkei	$RMSE_1^2$ DAX	$RMSE_1^3$ Dow Jones	$RMSE$
1	1.5290	0.6780	0.7475	0.7773	0.7236
2	1.6170	0.8004	0.7347	0.6492	0.6100
3	1.9890	0.6026	0.7022	0.5927	0.5774
4	1.7786	0.8861	0.9328	0.6863	0.5969
5	1.6224	0.7452	0.9870	0.7545	0.6112
6	1.5535	0.7837	0.9571	0.8408	0.6132
7	1.5383	0.8002	0.9677	0.6857	0.6253
8	1.4828	0.7917	0.9888	0.7479	0.6333
9	1.4570	0.8789	0.9220	0.6933	0.7039

Table 11: The multivariate relative predictive ordinate criterion $RPOC_1$ and the relative mean square error of the VAR(1) model for daily Stock indices of Japan, German and USA ($\mathbf{x}_t^1$ =Nikkei, $\mathbf{x}_t^2$ =DAX and $\mathbf{x}_t^3$ =Dow Jones) from June, 21, 1996 to June, 22, 1998

	Exchange rates	Stock indices
Δd^2	1.0875	1.0701
D_{bias}	0.3227	0.3483
D_{var}	0.0258	0.0461
D_{noise}	0.6514	0.6056
σ^2	1.3354	1.5465
$\bar{d}$	9.3134	7.5807
σ_N^2	1.1678	1.3244
$\bar{d}^N$	8.7210	6.9702
ρ	0.7074	0.7594

Table 12: The *APOC* decomposition for the VAR(1) model for daily exchange rates (US$/DM, US$/Yen and DM/Yen) and stock indices (Nikkei, DAX and Dow Jones) from June, 21, 1996 to June, 12, 1998

	Exchange rates	Stock indices
D	0.00017	0.00471
D_{bias}	0.02954	0.00983
D_{var}	0.10199	0.22520
D_{noise}	0.86847	0.76498

Table 13: The multivariate *MSE* decomposition of the VAR-GARCH-M model for daily exchange rates (US$/DM, US$/Yen and DM/Yen) and stock indices (Nikkei, DAX and Dow Jones) from June, 21, 1996 to June, 12, 1998

Exchange rates	Stock indices
0.0006942	0.0208560
0.0003514	0.0121810
0.0001715	0.0024779

Table 14: $EV(\mathbf{D})$ for daily exchange rates (US$/DM, US$/Yen and DM/Yen) and stock indices (Nikkei, DAX and Dow Jones) from June, 21, 1996 to June, 12, 1998

	Exchange rates			Stock indices		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
D_{bias}	0.00747	0.00000	0.00000	0.00465	0.00000	0.00000
D_{var}	0.12700	0.11306	0.17348	0.10782	0.12212	0.16972
D_{noise}	0.86553	0.88694	0.82652	0.88753	0.87788	0.83028

Table 15: The relative MSE decomposition for daily exchange rates (US$/DM, US$/Yen and DM/Yen) and stock indices (Nikkei, DAX and Dow Jones) from June, 21, 1996 to June, 12, 1998