THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 23_13

Wolfgang Polasek
Institute for Advanced Studies, Austria
The Rimini Centre for Economic Analysis (RCEA), Italy

FORECAST EVALUATIONS FOR MULTIPLE TIME SERIES: A GENERALIZED THEIL DECOMPOSITION

The Rimini Centre for Economic Analysis
Legal address: Via Angherà, 22 – Head office: Via Patara, 3 - 47900 Rimini (RN) – Italy
www.rcfea.org - secretary@rcfea.org

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

Wolfgang Polasek

Institute of Advanced Studies
Stumpergasse 56, 1060 Wien, Austria
Email: Polasek@ihs.ac.at

May 1, 2013
2. Draft

Abstract

The mean square error (MSE) compares point forecasts or a location parameter of the forecasting distribution with actual observations by the quadratic loss criterion. This paper shows how the $Theil$ decomposition of the MSE error into a *bias*, *variance* and *noise* component which was proposed for univariate time series can be used to evaluate and compare multiple time series forecasts. Thus, for multivariate time series the ordinary and the alternative $Theil$ decomposition is applied to decompose the MSE matrix. As an alternative we propose the average predictive ordinate criterion ($APOC$) which evaluates the ordinates of the predictive distribution for comparing forecasts of volatile time series. The multivariate $Theil$ decomposition for the MSE and $APOC$ criterion is used to compare and evaluate 3-dimensional VAR-GARCH-M time series forecasts for stock indices and exchange rates.

`Keywords`: Forecast comparisons, average predictive ordinate criterion $APOC$, MSE matrix and multivariate predictions, multivariate and alternative $Theil$ decomposition.

1 Introduction

Time series models are often evaluated by their forecasting performance and financial time series models evaluate forecasts in the presence of a volatility model. Are the point forecasts and evaluation criteria like the MSE criterion appropriate in such cases? Recently, Christofferson (1998) and Diebold et al. (1998) have proposed new ways how to evaluate interval and density forecasts for financial time series applications. Christofferson (1998) suggests a likelihood ratio test of conditional coverage for interval forecasts as an alternative to point forecasts. While such an approach lies in the tradition of predictive model testing we want to explore in the present paper a direct approach to quantify the predictability potential of volatile time series. We suggest an average predictive ordinate criterion ($APOC$) which is in the spirit of the predictive inference approach: a non-time series introduction into this field can be found in Geisser (1993). Evaluations by predictive ordinates have as advantage to generalize easily to multivariate time series models. Thus, we will compare the $APOC$ approach with the mean squared error (MSE) for multiple time series.

Therefore we will first describe the MSE approach for forecasts: Let $\hat{x}_{T+1}$ be the prediction and x_{T+1} the actual observation, then the MSE for period $T+1$ is defined by

$$MSE(\hat{x}_{T+1}) = (x_{T+1} - \hat{x}_{T+1})^2. \tag{1}$$

In case of unbiased forecasts we require $E\hat{x}_{T+1} = x_{T+1}$. If the forecast is the mean of the time series then $\hat{x}_{T+1} = \bar{x}$ and the MSE reduces to the variance σ^2 and the relative MSE, $RMSE = MSE(\hat{x}_{T+1})/\sigma^2$, can be used to compare time series forecasts with different measurement scales. 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 cannot 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 propose a forecast evaluation criterion which is based on the whole forecasting (or predictive) distribution. At the time T forecast origin, we forecast the next observation 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 ordinate of the probability density for the new observation. These ordinates can be used for discriminating between competing models. Choosing between two models, we prefer the model where the new observation falls into a more likely density region of the predictive distribution. Such a measure is provided by the ordinate of the predictive distribution.

This evaluations of forecasts is called predictive ordinate (PO) and we call the principle of choosing a model with the highest predictive ordinates the predictive ordinate criterion (POC). A new observation which realizes just at the mode of the (unimodal) predictive distribution attains the maximum ordinate. If the

observation realizes in the tails of the predictive distribution then the predictive ordinates will be small. Over a horizon of H future observations we suggest to take the average of the predictive ordinates ($APOC$) as a new evaluation criterion for forecasts.
What is the difference to the MSE criterion? The MSE evaluation of forecasts does not differentiate if the forecasts were coming from a 'normal' distribution or a fat tailed distribution. Thus, the evaluation of forecasts by the $APOC$ criteria also implies a judgement of the quality of the predictive distribution. E.g., a $GARCH$ model could produce the same forecast with a small and a large conditional variance. The $APOC$ criteria assigns the forecast with the small conditional variance a larger weight. Also, the forecasting process can be evaluated over H time periods. Using the $APOC$ criterion we will prefer the model which produces higher predictive ordinates on average over H periods. This will be the case for forecasts with narrower predictive densities and particularly if the observation realizes in the center of the predictive distributions. We can distinguish between two types of forecasting situations:

1. Keep the forecast origin T fixed and consider evolving lead times up to horizon H. We define as average ordinate criterion ($APOC$)

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

   where $d_T(h), h = 1, \ldots, H$ is the ordinate of the predictive density $p(x_{T+h}|I_T)$, given the information set I_T and 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 $T + h - 1$.

In particular we argue that the $APOC$ criterion is better suited to evaluate forecasts of volatility or $GARCH$ models. First of all, it takes into account different distributional assumptions of the time series models since this is reflected in the shape of the predictive distribution. Secondly, the $APOC$ criterion generalizes in a straightforward way to multivariate time series since predictive ordinates are simple one-dimensional summary measures of multiple time series forecasts. We will compare the $APOC$ criterion with the $Theil$ decomposition of the MSE matrix in the multivariate case. Since a MSE matrix is no longer a one dimensional summary measure we will explore the possibility of interpreting the properties of the MSE matrix by an eigenvalue decomposition.
The plan of the paper is as follows: In section **??** we introduce the $APOC$ and the relative predictive ordinate criterion $RPOC$ for comparing the $APOC$ of the current model with the $APOC$ of the no-change forecasts. Section **??**

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 **??**. Moreover we show how the $APOC$ criterion can be extended to the multivariate case. Furthermore, we discuss the merits of a $Theil$ decomposition of the $APOC$ criterion. This is in principle possible but the interpretation of the components of the $Theil$ decomposition need some adjustments. In the last section we evaluate the returns of a 3-dimensional model of international stock indices and three exchange rates for the VAR-GARCH model. In the conclusion we summarize our experience to evaluate multiple time series forecasts by generalized $Theil$ decompositions.

2 Forecast evaluation by the MSE

Forecast evaluations by the mean squared error (MSE), the root MSE or $\sqrt{MSE}$ and the relative root MSE ($RRMSE$ or briefly as $RMSE$) are important for applied time series analysis. Successful time series models are often measured in terms of their forecasting performance. 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. 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, but evaluate the forecasts for evolving time origins up to the time horizon H, e.g. for the one-step ahead forecasts the sequence $x_T(1), x_{T+1}(1), \ldots, x_{T+H-1}(1)$. 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 (equidistant) time step. This concept can be extended to the MSE_k if k-step ahead forecasts have to be evaluated for an expanding information set.

2.1 The relative root mean square error ($RRMSE$)

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

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

where the $\hat{x}_{T+h}, h = 1, \ldots, H$ denotes the successive one-step ahead forecasts over the horizon H. The no-change mean square error (MSE_1^{no}) for one-step ahead predictions is given by

$$MSE_1^{no} = \frac{1}{H}\sum_{h=1}^{H}(x_{T+h-1} - x_{T+h})^2.$$

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

$$RRMSE_1^{no} = \sqrt{\frac{MSE_1}{MSE_1^{no}}}. \tag{3}$$

This criterion can be computed for different time horizons H.

2.2 The MSE matrix

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

$$MSE_1 = \frac{1}{H}\sum_{h=1}^{H}(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})(\mathbf{y}_{T+h} - \hat{\mathbf{y}}_{T+h})', \tag{4}$$

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

$$MSE_1^{no} = \frac{1}{H}\sum_{h=1}^{H}(\mathbf{y}_{T+h-1} - \mathbf{y}_{T+h})(\mathbf{y}_{T+h-1} - \mathbf{y}_{T+h})'.$$

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:

$$RMSE_1^{det} = \sqrt{\frac{det(MSE_1)}{det(MSE_1^{no})}}. \tag{5}$$

A further possibility to compare MSE matrices is to calculate the relative gain in eigenvalues of the MSE matrix, i.e.

$$\frac{\lambda_1}{\lambda_1^{no}}, \ldots, \frac{\lambda_p}{\lambda_p^{no}} \tag{6}$$

with $\lambda_1, \ldots, \lambda_p$ being the eigenvalues of MSE_1 and $\lambda_1^{no}, \ldots, \lambda_p^{no}$ from those of MSE_1^{no}. The disadvantage is the rather technical nature of the interpretation of eigenvalues: The relative gains of eigenvalues in (**??**) indicate if the current forecasts are improving the predictive information over independent dimensions by the same rate.

2.3 The univariate *Theil* decomposition of the *MSE*

The following decomposition of the *MSE* can be found in Pindyck and Rubinfeld (1998) and Clements and Hendry (1998) and relates to the inequality coefficients developed in Theil (1961).
Theorem ??.1 The ordinary *Theil* decomposition of the *MSE*
Let y_t be the actual observations and x_t the forecasts from a model. Then the mean square error decomposition for the horizon H is

$$\begin{aligned} \Delta d^2 &= \frac{1}{H}\sum_{t=1}^{H}(x_t - y_t)^2 \\ &= (\bar{x} - \bar{y})^2 + (\sigma_x - \sigma_y)^2 + 2(1-\rho)\sigma_x\sigma_y \end{aligned} \tag{7}$$

or

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

The relative *Theil* decomposition is given by

$$1 = D_{bias} + D_{var} + D_{noise} \tag{8}$$

where the proportions of the decomposition are given as

$$\begin{aligned} D_{bias} &= (\bar{x} - \bar{y})^2/\Delta d^2, \\ D_{var} &= (\sigma_x - \sigma_y)^2/\Delta d^2, \\ D_{noise} &= (1 - \rho_{xy})\sigma_x\sigma_y/\Delta d^2. \end{aligned} \tag{9}$$

The descriptive first and second moments are calculated for the forecast horizon:

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

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

and the correlation coefficient $\rho = \rho_{xy}$ is given by

$$\rho = \frac{1}{H}\frac{1}{\sigma_x\sigma_y}\sum(x_t - \bar{x})(y_t - \bar{y}). \tag{10}$$

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. For the evaluation of volatility model forecasts, the D_{var} component indicates 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 the actual 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 mean square sense are detected by the relative $Theil$ decomposition if D_{noise} can be made as large as possible.
Theorem ??.2 The alternative $Theil$ decomposition of the MSE
As before, let y_t be the actual values and x_t the predicted values of the time series. Then the MSE for the period H is

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

and can be alternatively decomposed as

$$\begin{aligned} \Delta d^2 &= (\bar{x} - \bar{y})^2 + (\sigma_x - \rho\sigma_y)^2 + (1 - \rho^2)\sigma_y^2 \\ &= bias^2 + adjvar + mmse, \end{aligned} \qquad (11)$$

where $adjvar$ stands for adjusted variance and $mmse$ stands for minimum mean squared error. The relative decomposition is given in analogy to (**??**) by

$$D_{bias} + D_{adjvar} + D_{mmse} = 1, \qquad (12)$$

As in the direct $Theil$ decomposition, D_{bias} is a measure for the quadratic distance between the observations and the forecasts. With the relative components defined as

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

and the first two moments are given as in theorem 2.1.1 and the correlation coefficient is computed as in (**??**).

Proof: Squaring the last two terms in (**??**) yields

$$(\sigma_x - \rho\sigma_y)^2 + (1 - \rho^2)\sigma_y^2 = \sigma_x^2 + \rho^2\sigma_y^2 - 2\rho\sigma_x\sigma_y + \sigma_y^2 - \rho^2\sigma_y^2,$$

and squaring the last two terms in (**??**) yields the same value ($\sigma_x^2+\sigma_y^2+2\rho\sigma_x\sigma_y$):

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

Clements and Hendry (1998, p.64) show that the minimum MSE forecasts are found by the solutions of the first order conditions:

$$\begin{aligned} \frac{\partial \Delta d^2}{\partial \bar{x}} &= -2(\bar{x}-\bar{y}) = 0, \\ \frac{\partial \Delta d^2}{\partial \sigma_x} &= -2(\sigma_x - \rho\sigma_y) = 0, \\ \frac{\partial \Delta d^2}{\partial \rho} &= -2\sigma_x\sigma_y = 0. \end{aligned} \quad (13)$$

Thus, (**??**) will be minimized if $\bar{x} = \bar{y}$ and $\sigma_x = \rho\sigma_y$ and the minimum MSE is given by the remainder term in (**??**):

$$MMSE = \tau \min \Delta d^2 = (1-\rho^2)\sigma_y^2. \quad (14)$$

The ratio $\Delta d^2/MMSE$ can be interpreted as a scale independent potential MSE which can be reduced by the forecasting model.
The difference between the simple and the alternative $Theil$ decomposition lies in the D_{var} and the D_{adjvar} terms. If the correlation coefficient between the actual and forecasted values is low, then condition (**??**) implies that the MSE is minimized if the variance of the forecasts is sufficiently small. The $MMSE$ decomposition prefers smooth predictions over erratic ones.
The correlation coefficient ρ is a measure for the difficulty to forecasts the time series, in short a measure for 'forcastability'. A small ρ implies through the variance term of the alternative $Theil$ decomposition that the variance of the forecasts should be made smaller. In the extreme case of zero correlation the forecasts have to be set to a constant (i.e. zero variance) to qualify as good forecasts according to the alternative $Theil$ decomposition.

2.4 Forecasting volatilities

For financial time series the prediction of conditional variances is sometimes more important than the predictions of returns. Therefore we discuss the possibility to apply the previous discussed MSE approach to evaluate volatility forecasts. Taking the AR-GARCH model as example we show in this section how the MSE criterion can be used to evaluate the forecasts of conditional variances (volatilities).
Consider the following AR-GARCH model with mean equation

$$y_t = \beta_0 + \sum_{i=1}^{k} y_{t-i}\beta_i + \varepsilon_t, \quad \varepsilon_t \sim N[0, h_t], \quad (15)$$

and variance equation

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{t-i} + \sum_{i=1}^{q} \phi_i \varepsilon_{t-i}^2, \quad t = 1, \ldots, T.$$

and we pursue the question, if the observed and predicted conditional variances are in agreement. First we have to define what we understand by 'observed' volatilities. Via the fitted values of the AR-GARCH model

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

we compute the observed volatility by $\nu_t = \hat{\varepsilon}_t^2 = (y_t - \hat{y}_t)^2$. The predicted volatilities are given by the estimated $GARCH$ equation

$$\hat{h}_t = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i \hat{h}_{t-i} + \sum_{i=1}^{q} \hat{\phi}_i \hat{\varepsilon}_{t-i}^2. \tag{16}$$

In analogy to Theorem 2.1.1 we define the MSE decomposition for volatilities:

$$\begin{aligned} \Delta d^2 &= \frac{1}{H} \sum_{t=1}^{H} (\hat{h}_t - \nu_t)^2 \\ &= (\bar{h} - \bar{\nu})^2 + (\sigma_h - \sigma_\nu)^2 + 2(1-\rho)\sigma_h \sigma_\nu \end{aligned}$$

or in relative terms the $Theil$ decomposition for volatilities is given by

$$D_{bias} + D_{var} + D_{noise} = 1 \tag{17}$$

where

$$\begin{aligned} D_{bias} &= (\bar{h} - \bar{\nu})^2 / \Delta d^2, \\ D_{var} &= (\sigma_h - \sigma_\nu)^2 / \Delta d^2, \\ D_{noise} &= (1 - \rho_{h\nu}) \sigma_h \sigma_\nu / \Delta d^2. \end{aligned}$$

The first and second moments over the forecast horizon H are calculated as

$$\sigma_h^2 = \frac{1}{H} \sum_{t=1}^{H} (\hat{h}_{T+t} - \bar{h})^2, \quad \text{and} \quad \sigma_\nu^2 = \frac{1}{H} \sum_{t=1}^{H} (\nu_{T+t} - \bar{\nu})^2,$$

$$\bar{h} = \frac{1}{H} \sum_{t=1}^{H} \hat{h}_{T+t}, \quad \text{and} \quad \bar{\nu} = \frac{1}{H} \sum_{t=1}^{H} \nu_{T+t},$$

and the correlation coefficient is

$$\rho = \frac{1}{H} \frac{1}{\sigma_h \sigma_\nu} \sum_{t=1}^{H} (\hat{h}_{T+t} - \bar{h})(\nu_{T+t} - \bar{\nu}).$$

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

$$\hat{y}_{T+h} = \hat{\beta}_0 + \sum_{i=1}^{k} \mathrm{E_T y_{T+h-i}} \hat{\beta}_\mathrm{i}$$

where the conditional expectation is given by $\mathrm{E_T y_{T+h}} = \mathrm{y_{T+h}}$ for $h \leq 0$ and $\mathrm{E_T y_{T+h}} = \mathrm{\hat{y}_{T+h}}$ for $h > 0$. At time T the h-step ahead forecasts are given by

$$\hat{h}_{T+h} = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i \hat{h}_{T+h-i} + \sum_{j=1}^{q} \hat{\phi}_j \mathrm{E_T} \hat{\varepsilon}^2_{\mathrm{T+h-j}}$$

with $\mathrm{E_T}\hat{\varepsilon}^2_{\mathrm{T+h}} = \hat{\varepsilon}^2_{\mathrm{T+h}}$ for $h \leq 0$ and $\mathrm{E_T}\hat{\varepsilon}^2_{\mathrm{T+h}} = 0$ else.

2.5 The multivariate *Theil* decomposition

This section discusses the *Theil* decomposition for multivariate time series can be extended in different ways. In theorem 2.1.3 we show in analogy to the univariate decomposition that the *MSE* matrix as defined in (**??**) can be decomposed in matrix components. In theorem 2.1.4 we suggest to decompose certain summary measures of the *MSE* matrix by the *Theil* decomposition.
Theorem ??.3: The matrix *MSE* decomposition
We consider a multivariate time series $\mathbf{y}_1, \ldots, \mathbf{y}_T$ which produce the forecasts $\mathbf{x}_1, \ldots, \mathbf{x}_H$ of a fitted model where the observed values are $\mathbf{y}_1, \ldots, \mathbf{y}_H$. The *MSE* matrix for the horizon H is defined as

$$\mathbf{D} = \frac{1}{H} \sum_{t=1}^{H} (\mathbf{x}_t - \mathbf{y}_t)(\mathbf{x}_t - \mathbf{y}_t)'$$

can be decomposed as

$$\begin{aligned} \mathbf{D} &= \mathbf{M} + \mathbf{\Sigma_{xx}} + \mathbf{\Sigma_{yy}} - \mathbf{\Sigma_{xy}} - \mathbf{\Sigma_{yx}} \\ &= \mathbf{M} + (\mathbf{\Sigma}^{1/2}_{\mathbf{xx}} - \mathbf{\Sigma}^{1/2}_{\mathbf{yy}})^2 + 2\mathbf{\Sigma}^{1/2}_{\mathbf{xx}}\mathbf{\Sigma}^{1/2}_{\mathbf{yy}} - \mathbf{\Sigma_{xy}} - \mathbf{\Sigma_{yx}} \end{aligned} \quad (18)$$

with the mean difference matrix given as

$$\mathbf{M} = (\bar{\mathbf{x}} - \bar{\mathbf{y}})(\bar{\mathbf{x}} - \bar{\mathbf{y}})' \quad (19)$$

and the covariance matrices are given by

$$\begin{aligned} \mathbf{\Sigma_{xx}} &= \frac{1}{H} \sum_{t=1}^{H} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{x}_t - \bar{\mathbf{x}})', \\ \mathbf{\Sigma_{yy}} &= \frac{1}{H} \sum_{t=1}^{H} (\mathbf{y}_t - \bar{\mathbf{y}})(\mathbf{y}_t - \bar{\mathbf{y}})', \\ \mathbf{\Sigma_{xy}} &= \frac{1}{H} \sum_{t=1}^{H} (\mathbf{x}_t - \bar{\mathbf{x}})(\mathbf{y}_t - \bar{\mathbf{y}})' \end{aligned}$$

and the mixed moment matrix has the property

$$\mathbf{\Sigma_{xy}} = \mathbf{\Sigma'_{yx}}.$$

Proof: See Appendix.
Note that $\mathbf{\Sigma}^{1/2}$ is the "square root" of a positive definite matrix defined through the eigenvalue decomposition as $\mathbf{\Sigma}^{1/2} = \mathbf{U} diag(\lambda_1^{1/2}, \ldots, \lambda_M^{1/2})\mathbf{U}^T$, where we denote with $\lambda(\mathbf{\Sigma}) = (\lambda_1, \ldots, \lambda_M)$ the eigenvalues of $\mathbf{\Sigma}$ and $\mathbf{U}$ is the associated matrix of eigenvectors. $\mathbf{\Sigma_{xx}}$ and $\mathbf{\Sigma_{yy}}$ are jointly diagonalized.

In the next theorem we suggest to extend the *Theil* decomposition by special summary measures of the *MSE* matrix.
Theorem ??.4 Special multivariate *Theil* decompositions
Based on theorem 2.1.3 we can define three *Theil* decompositions for functions of the multivariate *MSE* matrix:

1. The *Theil* decomposition for determinants;
   In terms of the determinants the generalized *MSE* decomposition is given by

$$\begin{aligned} D_{bias} &= \frac{|\mathbf{M}|}{|\mathbf{D}|}, \\ D_{var} &= \frac{|\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{\Sigma_{yy}}^{1/2}|^2}{|\mathbf{D}|}, \\ D_{noise} &= 1 - D_{bias} - D_{var}. \end{aligned} \tag{20}$$

   Since the determinant of a covariance matrix $|\mathbf{D}|$ is also called 'generalized variance' we can call the determinant of the multivariate *MSE* matrix the 'generalized *MSE*'.

2. The *Theil* decomposition for eigenvalues;
   Collecting the eigenvalues of $\mathbf{D}$, we define the $1 \times M$ vector of eigenvalues by $\lambda(\mathbf{D}) = (\lambda_1^D, \ldots, \lambda_M^D)$ and a M-dimensional relative *MSE* decomposition can be computed for all eigenvalues individually:

$$\begin{aligned} D^\lambda_{bias,i} &= \lambda_i(\mathbf{M})/\lambda_i(\mathbf{D}), \\ D^\lambda_{var,i} &= \lambda_i(\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{\Sigma_{yy}}^{1/2})^2/\lambda_i(\mathbf{D}), \end{aligned} \tag{21}$$

   and the relative decomposition exists if the noise term is positive

$$D^\lambda_{noise,i} = 1 - D^\lambda_{mean,i} - D^\lambda_{var,i}, \quad i = 1, ..., M.$$

3. The *Theil* decomposition for the trace;
   The relative components are defined as

$$\begin{aligned} D^{tr}_{bias} &= \frac{tr\mathbf{M}}{tr\mathbf{D}}, \\ D^{tr}_{var} &= \sum_{m=1}^{M} (\sigma_{x,m} - \sigma_{y,m})^2/tr\mathbf{D}, \\ D^{tr}_{noise} &= 1 - D^{tr}_{bias} - D^{tr}_{var}. \end{aligned}$$

The trace of the MSE matrix $\mathbf{D}$ is equivalent to a summation over the M separate univariate components:

$$tr\mathbf{D} = \sum_{m=1}^{M} [(\bar{x}_m - \bar{y}_m)^2 + (\sigma_{x,m} - \sigma_{y,m})^2 - 2(1 - \rho_{xy,m})(\sigma_{x,m}\sigma_{y,m})] \qquad (22)$$

with the correlation coefficient of the m^{th} component is defined as

$$\rho_{xy,m} = \frac{1}{\sigma_{x,m}\sigma_{y,m}} \sum_{t=1}^{T} (x_{t,m} - \bar{x}_m)(y_{t,m} - \bar{y}_m). \qquad (23)$$

Use Theorem 2.1.3 and the multivariate matrix calculus. Note that the eigenvalues of a sum of p.d. matrices is not the sum of the eigenvalues of each matrix (see Schott 1997). %**Remarks:**

1. Since the trace of a matrix depends on the scale of the M components it is of limited practical use. We recommend the trace as a reasonable summary measure for multivariate observations if the time series are measured on the same scale.

2. As in the univariate context, the multivariate decompositions can be interpreted as before: Good forecasts have the property that D_{bias} and D_{var} are close to zero and D_{noise} is close to 1.

2.6 The alternative $Theil$ decomposition

In analogy to the univariate case in theorem 2.1.2. we define alternative $Theil$ decompositions for the multivariate MSE matrix. The next theorem generalizes the alternative $Theil$ decompositions for the whole MSE matrix and 3 special MSE matrix functions.

Theorem 2.1.5 Alternative multivariate $Theil$ decompositions

First we extend the we multivariate $Theil$ decomposition for the matrix case in theorem 2.1.3. The alternative matrix decomposition needs the correlation coefficients which are defined in (**??**) to create the diagonal matrix

$$\mathbf{D}_\rho = diag(\rho_{xy,1}, \ldots, \rho_{xy,M}). \qquad (24)$$

The alternative matrix MSE decomposition now is

$$\mathbf{D} = \mathbf{M} + (\mathbf{\Sigma}_{\mathbf{xx}}^{1/2} - \mathbf{D}_\rho \mathbf{\Sigma}_{\mathbf{yy}}^{1/2})^2 + \mathbf{Q} \qquad (25)$$

where $\mathbf{Q}$ is given by

$$\mathbf{Q} = 2\mathbf{\Sigma}_{\mathbf{xx}} + \mathbf{\Sigma}_{\mathbf{yy}} - \mathbf{\Sigma}_{\mathbf{xy}} - \mathbf{\Sigma}_{\mathbf{yx}} - (\mathbf{\Sigma}_{\mathbf{xx}}^{1/2} - \mathbf{D}_\rho \mathbf{\Sigma}_{\mathbf{yy}}^{1/2})^2. \qquad (26)$$

Based on univariate decomposition in theorem 2.1.4, we define 3 further special $Theil$ decompositions for the multivariate MSE matrix.

1. The alternative $Theil$ decomposition by determinants;
   The bias term is given as in theorem 2.1.4 and the adjusted variance term is defined as

$$\begin{aligned} D_{adjvar} &= \frac{|\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{D}_{\rho}\mathbf{\Sigma_{yy}}^{1/2}|^2}{|\mathbf{D}|}, \\ D_{mmse} &= 1 - D_{bias} - D_{adjvar}. \end{aligned} \tag{27}$$

   The minimum mean square error term is calculated simply as remainder term of the $Theil$ decomposition.

2. The alternative $Theil$ decomposition by eigenvalues;
   The bias term is given as before by $D^{\lambda}_{bias,i} = \lambda_i(\mathbf{M})/\lambda_i(\mathbf{D})$, and the adjusted variance term is calculated as

$$D^{\lambda}_{adjvar,i} = \lambda_i(\mathbf{\Sigma_{xx}}^{1/2} - \mathbf{D}_{\rho}\mathbf{\Sigma_{yy}}^{1/2})^2/\lambda_i(\mathbf{D}), \tag{28}$$

   with the remainder terms

$$D^{\lambda}_{mmse,i} = 1 - D^{\lambda}_{mean,i} - D^{\lambda}_{adjvar,i}, \quad i = 1, ..., M. \tag{29}$$

3. The alternative $Theil$ decomposition by the trace;
   This decomposition can be viewed as an additive extension of theorem 2.1.2. The bias term is $D^{tr}_{bias} = tr\mathbf{M}/tr\mathbf{D}$, and the other terms are

$$\begin{aligned} D^{tr}_{adjvar} &= \sum_{m=1}^{M} (\sigma_{x,m} - \rho_{xy,m}\sigma_{y,m})^2/tr\mathbf{D}, \\ D^{tr}_{mmse} &= 1 - D^{tr}_{bias} - D^{tr}_{adjvar}. \end{aligned}$$

Proof: Insert the term $(\mathbf{\Sigma}^{1/2}_{\mathbf{xx}} - \mathbf{D}_{\rho}\mathbf{\Sigma}^{1/2}_{\mathbf{yy}})^2$ into the decomposition of theorem 2.1.3. The trace decomposition of the MSE matrix $\mathbf{D}$ is equivalent to a summation over the M univariate components:

$$tr\mathbf{D} = \sum_{m=1}^{M} [(\bar{x}_m - \bar{y}_m)^2 + (\sigma_{x,m} - \rho_{xy,m}\sigma_{y,m})^2 - 2(1 - \rho^2_{xy,m})\sigma^2_{y,m}] \tag{30}$$

where the correlation coefficient of the m^{th} component is defined as before. The components add to 1: $D^{tr}_{bias} + D^{tr}_{adjvar} + D^{tr}_{mmse} = 1$.

3 The Average Predictive Ordinate Criterion ($APOC$)

Consider a time series $\{x_1, \ldots, x_T\}$ being a sample from a stochastic process with density $\{f_t(x)\}_{t=1,\ldots,T}$ 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

$$APOC = \frac{1}{H} \sum_{h=1}^{H} f_{T+h}(x_{T+h}|\hat{\theta}_T), \tag{31}$$

where $\hat{\theta}_T$ is the estimated parameter of the time series model using information up to time T. For repeated one-step ahead forecasts which are updated by the parameter estimator $\hat{\theta}_T$ we define as average predictive density

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

This definition is general enough to cover time series predictions for different distributional assumptions for the error term (e.g. t and stable distributions, hyperbolic, *log*-normal distributions etc.) For the one-step ahead normal distribution we use the notation "$APON_1$ criterion" since it is calculated from the one-step ahead predictors with normal distribution:

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

The parameters of the normal distribution $\hat{\mu}_{T+h-1}$ and $\hat{\sigma}^2_{T+h-1}$ are the mean and the variance of the one-step predictive densities at time $T+h-1$ for 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:

$$RPOC_1 = \frac{APOC_1}{APON_1}, \tag{32}$$

we propose to calculate the $APOC$ of the denominator by the no-change normal distribution prediction which is given by

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

The parameters of the normal distribution for this no-change prediction are $\hat{\mu}_{T+h-1} = x_{T+h-1}$ (that is the best observation) and

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

is the variance of the no-change forecasts (i.e. first differences) up to time $T+h-1$.

3.1 The *APOC* Criterion for Multiple Time Series

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

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

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

$$APOC_1 = \frac{1}{H} \sum_{h=1}^{H} f_{T+h}(\mathbf{y}_{T+h} | \hat{\theta}_{T+h-1}), \tag{33}$$

where $\hat{\theta}_t$ denote estimated parameters. For one-step ahead forecasts assuming a multivariate normal distribution, the $APOC$ is

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

where $\hat{\mathbf{y}}_\mathbf{t}$ and $\hat{\mathbf{\Sigma}}_t$ are the parameters (the first two moments) estimated with the information set up to time t. For no-change prediction we calculate a ratio of average predictive ordinates assuming a multivariate normal distribution

$$RPOC_1 = \frac{APOC_1}{APON_1}, \tag{34}$$

with

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

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

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

To demonstrate the application of the above formulas we consider the following
Example: One and two-step ARCH predictions
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

$$\hat{\mu}_{t+1} = \mathrm{E_t(y_{t+1})} = \beta_0 + \beta_1 \mathrm{y_t},$$
$$\hat{\sigma}_{t+1}^2 = \mathrm{Var_t(y_{t+1})} = \gamma_0 + \gamma_1 \mathrm{h_t} + \gamma_2 \varepsilon_\mathrm{t}^2, \tag{35}$$

where $\mathrm{E_t}$ and $\mathrm{Var_t}$ are the conditional mean and variance given the information up to time t and the residuals for the variance prediction in (**??**) calculated as

$$\varepsilon_t = y_t - \mu_t = y_t - \beta_0 - \beta_1 y_{t-1}.$$

For $h = 2$ the forecasting parameters are given by

$$\hat{\mu}_{t+2} = \mathrm{E_t(y_{t+2})} = \beta_0 + \beta_1 \mathrm{y_{t+1}},$$

$$\hat{\sigma}^2_{t+2} = \mathrm{Var_t(y_{t+2})} = \gamma_0 + \gamma_1 \mathrm{h_{t+1}} + \gamma_2 \varepsilon^2_{\mathrm{t+1}},$$

with the residual

$$\varepsilon_{t+1} = y_{t+1} - \beta_0 - \beta_1 y_t,$$

and similarly for higher order predictions.
Remark: The $Theil$ decomposition of the $APOC$

In analogy to the $Theil$ decomposition of the MSE we can decompose the squared distance of the $APOC$ criterion between the predictive ordinate of a model and 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

$$\Delta d^2 = \frac{1}{H} \sum (d_t - d_t^N)^2 = (\bar{d} - \bar{d}^N)^2 + (\sigma - \sigma_N)^2 + 2(1 - \rho)\sigma\sigma_N, \qquad (36)$$

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.
The practical applications in section 4.2 will show that this approach is less suited for forecast evaluations of volatile time series, mainly because the interpretation of the decomposition is rather abstract.

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. In this section we will evaluate the time series forecasts for univariate AR-GARCH and multivariate VAR-GARCH models for stock returns and exchange rates. Details of the modeling and the estimation process can be found in Polasek and Ren (2000).

4.1 Univariate forecast evaluations

Daily observations of stock returns and exchange rates exhibit a volatile behaviour. Therefore we have estimated univariate $AR(k)$-$GARCH(p,q)$-$M(r)$ models for the returns of the stock market indices in Japan, Germany and the US and the exchange rates between these countries.
Table 1 summarizes the order selection for univariate AR(k)-GARCH(p,q)-M(r) models by the marginal likelihood criterion, a Bayesian model selection criterion like BIC (see Polasek and Ren 2000). We conclude from the estimated orders

that there is a rich interaction pattern for volatilities and the orders are different for the returns of the stock indices and the exchange rates. (The returns are defined as the differences of the logarithms of the time series).
First of all, we try to find out if the $APOC$ or the MSE criterion is better to describe the forecasting properties of stock returns and the exchange rates returns for Japan, German and the US. Tables **??** and **??** list the numerical values and Figure 1 and 2 show the graphs for these tables. We see that the graphs seem to tell different stories: while Figure 1 shows a rather similar picture for the 6 forecasts using $APOC$ criterion, Figure 2 gives a quite heterogeneous answer if the MSE criterion is used. Therefore a more careful analysis of the results is necessary.
Table **??** shows the $RPOC$ over 10 days for daily one-step ahead forecasts starting on June 12, 1998. The smallest improvement over the no-change prediction can be seen for the first forecasts of all the 3 exchange rates and the 3 stock returns. Except for the Dow Jones returns all forecast improve in the first 3 days and stabilize if the next 5 days are included in the forecast evaluation. Including all 10 days decreases the $APOC$ for all 6 series, but leaves the relative ordering unchanged: The maximum improvement over the no-change forecasts is found for the DAX returns with $RPOC = 1.78$ if the first 4 days are considered. It is interesting to note that the $RPOC$ decreases only for the Dow Jones index and give the lowest values if 3 to 10 days are considered, improving the no-change forecasts only by $22 - 30\%$.
The forecast comparison is different if we look at the relative mean square error ($RMSE$) in Table **??**. Over the 10 days, the relative MSE seems to be decreasing only for the Dow Jones returns and gives improvements of about 0.4915 or 50%. The worst improvement can be found for the US$/DM exchange rate (0.93) while all other time series lie between those two values. We conclude that summarizing the forecasts by the $RPOC$ criterion can be quite different from a relative MSE analysis. It seems that point forecast evaluated by second moments produce more erratic forecast evaluations than forecasts evaluated by the ordinates relative to the normal distribution.

4.2 Direct and alternative $Theil$ decompositions

Table **??** shows the $Theil$ decomposition for the $APOC$ or mean squared predictive ordinate ($MSPO$) 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 is absorbed by the *noise* component since the variance part is almost zero. 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 **??** 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.
Table 5 contains also the alternative $Theil$ decomposition, defined in (12), for the 6 time series. Note that the numbers add to 1 and involve only a recalculation of the last two terms, the adjusted variance term D_{adjar} and the minimum MSE term D_{mmse}. The comparison between D_{noise} and D_{mmse} indicates where the most forecasting potential is to be expected. Interestingly the numbers indicate that the largest potential seems to be present for the USD/Yen exchange rate and the Dow Jones returns.
Table 6 gives the MSE decompositions for volatilities as defined in (17). The bias component is the largest for the DM/Yen relationship which is a sign that this model should not be used for volatility forecasting. (Note that the means for the true and predicted volatilities, $\bar{x}$ and $\bar{h}$, are quite different.) Interestingly, all the variance components are small, but not so the adjusted variance components. This indicates that possibly better volatility forecast models can be found for the DAX returns because of the gap between the *noise* and the *mmse* component. Overall, we see that the evaluation of the volatility forecasts by the $Theil$-MSE decomposition have to be interpreted with care.

4.3 Multivariate VAR-GARCH-M forecast evaluations

Table 7 summarizes the order selection procedure by marginal likelihoods for the 3-dimensional VAR-GARCH-M models for the returns of the stock indices and the exchange rates separately. Interestingly, the same order is picked for the multivariate model between exchange rates and stock returns. The multivariate MSE decomposition of section 2.4 can be found for the exchange rates and stock returns in Table 13. The *bias* term is very small and is also smaller for stock returns than for exchange rate returns. The *noise* term for the multivariate model is larger for stock returns than for exchange rates returns. This shows that stock indices are more difficult to predict in a VAR-GARCH-M model than exchange rates. Note that the multivariate *noise* term is generally smaller for the univariate series which have been discussed in Table 5.
The first row of Table 14 lists the eigenvalues for the MSE matrix for exchange rates and stock returns. The $Theil$ decomposition for the eigenvalues for the MSE matrix according to formula (**??**) can be found in Table 14. We see that the *bias* term of the eigenvalues is practically zero while the *noise* term - being very high - lies between 82% and 89%. This means that in terms of the eigenvalue decomposition the multivariate prediction for exchange rates and stock returns is quite accurate and performs quite well according to the $Theil$ decomposition.

4.4 Diagnostics: Multivariate VAR forecast evaluations

This section has a diagnostic purpose and evaluates the forecasts of a VAR(1) model with the *APOC* and the *Theil* decomposition. We are interested in the question: How does the volatility component show up in the proposed evaluation measures? Tables **??** and **??** show the multivariate gain in relative predictive ordinates (*RPOC*) for one-step ahead forecasts of the VAR(1) model of exchange rates and the univariate and multivariate *RMSE*.

Table **??** 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 surprisingly we see a low performance of the point forecasts for the relative gains in the root mean square error. The improvement lies between 27 and 52% below the no-change root *MSE*. This picture is confirmed for the 3-dimensional model, where we compare the improvement by the root determinant. The last column of Table **??** shows that the improvements in *MSE* around the same magnitude as the determinant *MSE* for the VAR-GARCH-M model.

Table **??** contains the above analysis for the stock returns. The first column shows that the *RPOC* for the VAR(1) model is much less than the *RPOC* for the VAR-GARCH-M model in Table 9. The *RMSE*s for the three stock returns show again a very heterogeneous behavior over the 10 days. The *RMSE* of the 3-dimensional VAR system is slightly worse than for the VAR-GARCH-M model in Table 9. The *APOC* decomposition for the 3-dimensional VAR(1) model is displayed in Table **??**. The relative *Theil* decomposition splits the *APOC* for exchange rates and stock returns in about the same size: the *bias* proportion is about 1/3, leaving 2/3 for the *noise* term. This shows that the VAR model is 33% '*APOC*-better' than a multivariate no-change model. A similar result is obtained by VAR-GARCH-M forecasts and shows that the *APO*C can uncover substantial forecasting potential in a multivariate model. Note that the *MSE* decomposition in Table 13 turns out to be less informative for an evaluation: The relative variance term is about the double size for stock indices than for exchange rates, but there is no obvious interpretation for this phenomenon.

Thus, we conclude that the forecast improvements through the multivariate VARCH component is much better captured by the *APOC* criterion or the relative *Theil* decomposition. The *MSE* criterion has not a satisfactory ability to discriminate between good and bad volatility models based their forecasts.

5 Conclusions

The paper proposes some alternative evaluation measures for multivariate forecasts based on the multivariate *MSE* criterion. We have shown that the multivariate *MSE* matrix can be decomposed by a matrix or multivariate *Theil* decomposition and we analysed 3 special summary measures: the determinant, the eigenvalues or the trace of the *MSE* matrix. From the application to financial time series we see that the new average predictive ordinate criterion

($APOC$) seems to be a useful alternative to the traditional evaluation by the MSE criterion. We argue that the advantage of the $APOC$ lies in the ability that it can be used in the same way for univariate and multivariate forecast evaluations. On first sight the $APOC$ criterion is more difficult to interpret since it evaluates the probability density of the actual observation given a parametric conditional predictive density. But we have seen that this concept is particularly useful for the comparison of forecasts of volatile time series, since it takes into account the quality of the predictions if the variance changes over time. While the MSE criterion only uses second moments, the $APOC$ criterion is based on distributional assumptions. It allows a comparison with no-change forecasts if a particular error distribution is assumed to be the 'base line' distribution and in many cases it will be the normal distribution. This leads to the relative predictive ordinate criterion ($RPOC$) which is a measure of increase or decrease of the predictability between two models based on a simple density forecast.
Since the forecasts of volatile time series will certainly affect the shape of the predictive distribution, the $APOC$ criterion is a better choice for the forecast evaluation of volatile time series than the MSE criterion. In the example involving stock returns and exchange rates shows clearly how much better the $APOC$ criterion can discriminate between VAR and VAR-VARCH-M forecasts than the MSE criterion. The improvements in the relative MSE are smaller in size and more difficult to interpret than the improvements by the APOC criterion. We conclude that while the MSE criterion is easier to define than the $APOC$ criterion, the forecast evaluation of the $APOC$ criterion is more informative than by the MSE criterion. Applying the relative $Theil$ decomposition for exchange rates and stock returns we find that stock returns of 3 regional MSCI indices are more difficult to predict than exchange rates.

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] Christofferson, P.F. (1998), Evaluating Interval Forecasts, Enternational Economic Review, 39/4, 841-862

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

[4] Diebold, F.X., Gunter T.A. and A.S. Tay (1998) Evaluating Density Forecasts with Applications to Financial Risk Management, Int. Economic Review 39, 863-883.

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

[6] Geisser S. (1993), Predictive Inference: An Introduction, Chapman and Hall, NY.

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

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

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

[10] Polasek, W. and Ren, L. (2000), Volatility analysis during the Asia crisis: A multivariate GARCH-M model for stock returns in the US, Germany and Japan, to appear in Applied Stochastic Models in Bus. and Ind. 2001, *http://www.unibas.ch/iso*.

[11] Schott, J.R. (1997), Matrix Analysis for Statistics, John Wiley, NY.

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

6 Appendix

Proof of theorem 2.3: First note that as in the univariate case we decompose in the multivariate case the MSE matrix in the following way:

$$\begin{aligned}\mathbf{D} &= \frac{1}{H}\sum_{t=1}^{H}(\mathbf{x}_t-\mathbf{y}_t)(\mathbf{x}_t-\mathbf{y}_t)' \\ &= \frac{1}{H}\sum_{t=1}^{H}(\bar{\mathbf{x}}-\bar{\mathbf{y}})(\bar{\mathbf{x}}-\bar{\mathbf{y}})' + \frac{1}{H}\sum_{t=1}^{H}(\mathbf{x}_t-\bar{\mathbf{x}}+\bar{\mathbf{y}}-\mathbf{y}_t)(\mathbf{x}_t-\bar{\mathbf{x}}+\bar{\mathbf{y}}-\mathbf{y}_t)' \\ &= \mathbf{M}+\mathbf{R}.\end{aligned}$$

Furthermore, the $\mathbf{R}$ matrix can be split into 4 components

$$\begin{aligned}\mathbf{R} &= \frac{1}{H}\sum_{t=1}^{H}(\mathbf{x}_t-\bar{\mathbf{x}})(\mathbf{x}_t-\bar{\mathbf{x}})' + \frac{1}{H}\sum_{t=1}^{H}(\mathbf{y}_t-\bar{\mathbf{y}})(\mathbf{y}_t-\bar{\mathbf{y}})' \\ &\quad -\frac{1}{H}\sum_{t=1}^{H}(\mathbf{x}_t-\bar{\mathbf{x}})(\mathbf{y}_t-\bar{\mathbf{y}})' - \frac{1}{H}\sum_{t=1}^{H}(\mathbf{y}_t-\bar{\mathbf{y}})(\mathbf{x}_t-\bar{\mathbf{x}})' \\ &= \mathbf{\Sigma_{xx}}+\mathbf{\Sigma_{yy}}-\mathbf{\Sigma_{xy}}-\mathbf{\Sigma_{yx}}.\end{aligned}$$

For numerical calculations we use the formula

$$(\mathbf{\Sigma_{xx}}^{1/2}-\mathbf{\Sigma_{yy}}^{1/2})^2 = \mathbf{\Sigma_{xx}}+\mathbf{\Sigma_{yy}}-2\mathbf{\Sigma_{xx}}^{1/2}\mathbf{\Sigma_{yy}}^{1/2}. \tag{37}$$

To calculate the mixed term a joint diagonalisation of $\mathbf{\Sigma_{xx}}$ and $\mathbf{\Sigma_{yy}}$ is necessary. Substituting all these terms gives

$$\mathbf{D} = \mathbf{M} + (\mathbf{\Sigma}_{\mathbf{xx}}^{1/2} - \mathbf{\Sigma}_{\mathbf{yy}}^{1/2})^2 + 2\mathbf{\Sigma}_{\mathbf{xx}}^{1/2}\mathbf{\Sigma}_{\mathbf{yy}}^{1/2} - \mathbf{\Sigma_{xy}} - \mathbf{\Sigma_{yx}} \tag{38}$$

which is exactly (**??**).

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 univariate AR(k)-GARCH(p,q)-M(r) models: $US\$/DM$,US$/Yen, DM/Yen, Nikkei, DAX and Dow Jones from June, 21, 1996 to June, 12, 1998

Figure 1: $APOC$ for daily exchange rates and stock returns

Figure 2: MSE for daily exchange rates and stock returns

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 models 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 $RMSE$ of the univariate AR-GARCH-M models of Table 1 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 the $APOC$ as in (7) for univariate AR-GARCH-M models 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}	0.119886	0.348702	0.135186	0.076497	0.231242	0.212792
D_{mmse}	0.847492	0.648246	0.771269	0.910473	0.731514	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 MSE decomposition for the returns of the 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
σ_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 MSE decomposition for volatilities as in (17) of the AR-GARCH-M forecasts for daily exchange rates and stock indices from June, 21, 1996 to June, 12, 1998.

				exchange rates	stock returns
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 two VAR(k)-GARCH(p,q)-M(r) models 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), separately estimated.

H	$RPOC_1$	$RMSE_1$ US$/DM	$RMSE_1$ US$/Yen	$RMSE_1$ DM/Yen	$RMSE_1^{det}$
1	15.3543	0.6884	0.6682	0.6316	0.5698
2	19.6335	0.6524	0.6246	0.6291	0.5986
3	22.0070	0.6239	0.6166	0.6053	0.5346
4	24.3516	0.5801	0.5837	0.5977	0.5373
5	25.7497	0.5283	0.5647	0.5111	0.5253
6	30.5997	0.5133	0.5039	0.4661	0.5741
7	18.1786	0.5553	0.5443	0.5350	0.6136
8	17.5643	0.6489	0.6778	0.5479	0.5925
9	16.8179	0.7507	0.6888	0.5538	0.6221

Table 8: Multivariate forecast evaluation for exchange rates: The $RPOC_1$, the individual $RMSE$ and the determinant ($RMSE_1^{det}$) for the 3-dimensional VAR(1)-GARCH(1,2)-M(2) model defined in (5).)

H	$RPOC_1$	$RMSE_1$ Nikkei	$RMSE_1$ DAX	$RMSE_1$ Dow Jones	$RMSE_1^{det}$
1	5.7065	0.6318	0.9064	0.7814	0.5894
2	6.5556	0.8098	0.8403	0.6490	0.6579
3	7.2495	0.6017	0.8264	0.5893	0.6537
4	5.4773	0.8805	0.9089	0.6821	0.9303
5	4.8234	0.7404	0.8609	0.7518	0.8406
6	4.0215	0.7773	0.9291	0.8383	0.9994
7	3.2283	0.7929	0.9524	0.6838	0.8140
8	2.6568	0.7842	0.9772	0.7469	0.8399
9	2.5859	0.7782	0.9101	0.6923	0.8335

Table 9: Multivariate forecast evaluation for stock returns: The $RPOC_1$, the individual $RMSE$ and the determinant ($RMSE_1^{det}$) for the 3-dimensional VAR(1)-GARCH(1,2)-M(2) model defined in (5).

H	$RPOC_1$	$RMSE_1$ US$/DM	$RMSE_1$ US$/Yen	$RMSE_1$ DM/Yen	$RMSE_1^{det}$
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: Diagnostics: The multivariate relative predictive ordinate criterion $RPOC_1$ given in (**??**) and the relative mean square error $RMSE$ of the 3-dimensional 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. The last column is the determinant criterion $RMSE_1^{det}$

.

H	$RPOC_1$	$RMSE_1$ Nikkei	$RMSE_1$ DAX	$RMSE_1$ Dow Jones	$RMSE_1^{det}$
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: Diagnostics: The multivariate relative predictive ordinate criterion $RPOC_1$, the relative mean square error of the 3-dimensional VAR(1) model for daily stock indices of Japan, Germany and USA from June, 21, 1996 to June, 22, 1998. The last column is the determinant criterion $RMSE_1^{det}$

.

	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		
	λ_1	λ_2	λ_3	λ_1	λ_2	λ_3
D	0.0006942	0.0003514	0.0001715	0.0208560	0.0121810	0.0024779
$\%D_{bias}$	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 14: 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 in terms of the eigenvalues.