Blaskowitz, Oliver J.; Herwartz, Helmut

Working Paper

On economic evaluation of directional forecasts

SFB 649 Discussion Paper, No. 2009-052

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Blaskowitz, Oliver J.; Herwartz, Helmut (2009) : On economic evaluation of directional forecasts, SFB 649 Discussion Paper, No. 2009-052, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/39335

SFB 649 Discussion Paper 2009-052

On economic evaluation of directional forecasts

Oliver Blaskowitz*
Helmut Herwartz**

* Humboldt-Universität zu Berlin, Germany
** Christian-Albrechts-Universität Kiel, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

On economic evaluation of directional forecasts *

Oliver Blaskowitz
Humboldt–Universität zu Berlin
Institute of Statistics and Econometrics
Phone: +49 – 30 – 2093 – 5705
Email: blaskowitz@wiwi.hu–berlin.de

Helmut Herwartz
Christian–Albrechts–Universität zu Kiel
Institute of Statistics and Econometrics
Phone: +49 – 431 – 880 – 2417
Email: h.herwartz@stat–econ.uni–kiel.de

Abstract

It is commonly accepted that information is helpful if it can be exploited to improve a decision making process. In economics, decisions are often based on forecasts of up– or downward movements of the variable of interest. We point out that directional forecasts can provide a useful framework to assess the economic forecast value when loss functions (or success measures) are properly formulated to account for realized signs and realized magnitudes of directional movements. We discuss a general approach to evaluate (directional) forecasts which is simple to implement, robust to outlying or unreasonable forecasts and which provides an economically interpretable loss/success functional framework. As such, the measure of directional forecast value is a readily available alternative to the commonly used squared error loss criterion.

Keywords: Directional forecasts, directional forecast value, forecast evaluation, economic forecast value, mean squared forecast error, mean absolute forecast error.
JEL classification: C52, E17, E27, E37, E47, F17, F37, F47, G17.

*This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 'Economic Risk'. We thank two anonymous referees and the Associate Editor Michael Clements for helpful suggestions.

1 Introduction

Diebold and Mariano (1995) and Granger and Pesaran (2000a,b), among others, point out that in order to evaluate the usefulness of forecasts, measuring the realized economic value is more sensible than assessing a realized 'statistical value' in terms of mean squared or absolute forecast errors. Other loss functions based on forecast errors exist and find some support when evaluating the accuracy of various forecast methods across many series. These are, for example, the geometric mean of the relative absolute error, the mean absolute scaled error or the log mean squared error ratio (e.g. Thompson 1990, Armstrong and Collopy 1992, Hyndman and Koehler 2006). Generally, such forecast criteria suffer from lacking economic interpretability. Moreover, criteria based on forecast errors are not suitable whenever a forecast method is akin to produce unreasonable forecasts which are far away from the realizations of the variable of interest. Robustness to outliers is particularly relevant in applied research when numerous (econometric) forecast procedures have to be compared (e.g. Armstrong and Collopy 1992, Makridakis 1993).

It is commonly accepted that information is helpful if it can be exploited to improve a decision making process. Frequently, the available information set is used to produce forecasts. Hence, information is useful if the forecasts help to make decisions that reduce losses/costs or increase gains/utility. From this perspective, a forecast evaluation criterion should be related to decision making (see also Armstrong and Collopy 1992, Granger and Pesaran 2000a,b, Pesaran and Skouras 2002). In economics, decisions are often based on forecasts of directional up– or downward movements of the variable of interest. This paper focuses on some aspects of the economic evaluation of directional forecasts (DFs). We argue that commonly used approaches to evaluate DFs that rely on signs are mostly incomplete measures of the economic value. We point out that DFs can, nevertheless, provide a convenient framework to assess the economic forecast value. This is accomplished when loss functions (or success measures) are properly formulated to account for realized signs and realized magnitudes of directional movements. Accordingly, we suggest a success measure that is easy to implement and interpret, robust to outlying forecasts, and, thus, matches core requirements of comparative forecast performance analysis (Ahlburg 1992).

In the next section we review the evaluation of DFs when considering directional signs only. In Section 3 we sketch a general framework to assess the economic value of DFs and

provide an illustration in Section 4. Section 5 contains some concluding remarks.

2 Directional forecasts: signs only

In economic applications the forecast user is often interested in directional (up–/downward) movements of the variable of interest denoted by Y_t henceforth. A prominent macroeconomic example is given by a monetary authority who raises interest rates if inflation is predicted to rise. In finance, a speculator buys the stock if its price is expected to rise. Various other examples exist.

To formalize the forecast evaluation procedure we let h denote the forecast horizon. The forecast for Y_{t+h} using the information available in t is given by X_t^h. Using the indicator function $I(\bullet)$, the realized and predicted directions are given by $\widetilde{Y}_t = I(Y_{t+h} - Y_t > 0)$ and $\widetilde{X}_t = I(X_t^h - Y_t > 0)$. (In-)correct DFs are defined by the binary variable $\widetilde{Z}_t = I(\widetilde{X}_t = \widetilde{Y}_t)$. Directions can also be determined using a non–zero threshold. In principle, DFs need not necessarily be derived from forecast and current levels X_t^h and Y_t. Any other forecast method generating $\widetilde{X}_t$ is allowed. For example, DFs can be based on probability forecasts of changes in Y_t.

A commonly used loss function for DFs is given by

$$L_t^{\mathrm{DA}}(X_t^h, Y_{t+h}, Y_t) = \begin{cases} a & \text{if } \widetilde{Z}_t = 1 \\ b & \text{if } \widetilde{Z}_t = 0, \end{cases}$$

where $(a, b) \neq (0, 0)$. Note that the abbreviation DA refers to directional accuracy. For notational convenience, we neglect the arguments the loss function depends on and write L_t^{DA} instead of $L_t^{\mathrm{DA}}(X_t^h, Y_{t+h}, Y_t)$. In this framework, a correct DF has a 'value' of a and an incorrectly predicted direction a 'value' of b. Frequently $(a, b) = (1, -1)$ or $(a, b) = (1, 0)$. Hence, it makes more sense to call L_t^{DA} a success function. Leitch and Tanner (1995), Greer (2005), Blaskowitz and Herwartz (2009b) employ $(a,\ b) = (1, -1)$. Other authors use $(a, b) = (1, 0)$, e.g. Swanson and White (1995, 1997a,b), Gradojevic and Yang (2006) and Diebold (2007). Note that $\mathbb{E}[L_t^{\mathrm{DA}}] = (a - b)\mathbb{P}[\widetilde{Z}_t = 1] + b$. Consequently, using this loss function amounts to considering the number of correct, respectively, incorrect DFs. While L_t^{DA} is robust to outlying forecasts X_t^h, it ignores the size of realized directional movements. Therefore, it does not measure the economic value to the forecast user whenever correctly

predicted small, respectively, large realized directional changes have distinct benefits/losses to the forecast user.

Merton's (1981) theory implies that DFs have no value if the forecast user's subjective probability function for $\widetilde{Y}_t$ given the forecast user's information set does not change when the user obtains a forecast $\widetilde{X}_t$. Within the framework of Merton (1981) it holds that DFs have no value if and only if

$$HM = \mathbb{P}[\widetilde{X}_t = 1|\widetilde{Y}_t = 1] + \mathbb{P}[\widetilde{X}_t = 0|\widetilde{Y}_t = 0] = 1 \ .$$

The notation HM is due to the follow up contribution by Henriksson and Merton (1981). Moreover, DFs have positive value if and only if

$$HM > 1 \ .$$

In this case, the subjective probability function of the forecast user changes such that she considers up–/downward movements more likely when the forecast is an up–/downward movement. For applications of the HM statistic, see Schnader and Stekler (1990), Mills and Pepper (1999) and Ashiya (2006), among others. Merton's framework is not equivalent to the loss functional approach described earlier as pointed out, for instance, in Merton (1981) and Blaskowitz and Herwartz (2008). Notably, it is easily verified that

$$HM - 1 = \frac{\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right)}{\mathbb{V}\left[\widetilde{Y}_t\right]} \ ,$$

where $\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right)$ and $\mathbb{V}\left[\widetilde{Y}_t\right]$ denote the covariance between realized and predicted directions respectively the variance of realized directions. Hence, DFs have no value if and only if $\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) = 0$. Equivalently, $\widetilde{X}_t$ and $\widetilde{Y}_t$ are independent in this case. DFs have positive value if and only if $\text{Cov}\left(\widetilde{X}_t, \widetilde{Y}_t\right) > 0$, i.e. $\widetilde{X}_t$ and $\widetilde{Y}_t$ are positively correlated. A prominent naive benchmark strategy for DFs is given by forecasting always an upward (or downward) movement. Such naive DFs have no value in the sense of Merton. Hence, HM measures the additional value of a DF when compared to naive predictions. Consequently, the HM measure is not only robust to outlying forecasts, it also has a sensible and intuitive economic interpretation. Yet, it considers only the sign and neglects the magnitude of changes in the movement of Y_t.

The nonparametric test of predictive performance presented in Pesaran and Timmermann (1992) tests the null hypothesis that predicted and realized signs $\widetilde{X}_t$ and $\widetilde{Y}_t$ are independent. The latter hypothesis is equivalent to the null hypothesis implied by the Merton framework. Applications include, for instance, Pesaran and Timmermann (1995), Pons (2001), Schneider and Spitzer (2005).

While the DA criterion (as well as criteria based on forecast errors) does not measure the economic value of (directional) forecasts, the HM and the Pesaran and Timmermann (1992) approaches provide an 'all–purpose' measure for an economic value of DFs in a rather restrictive sense. A more appropriate context–specific assessment of the economic value of DFs is explained in the next Section.

3 The economic value of directional forecasts

To formalize the economic evaluation of DFs we define

$$L_t^{\mathrm{DV}}(X_t^h, Y_{t+h}, Y_t) = \begin{cases} H_t^{UU} = H^{UU}(Y_{t+h}, Y_t) & \text{if correct upward prediction} \\ H_t^{DD} = H^{DD}(Y_{t+h}, Y_t) & \text{if correct downward prediction} \\ H_t^{UD} = H^{UD}(Y_{t+h}, Y_t) & \text{if incorrect upward prediction} \\ H_t^{DU} = H^{DU}(Y_{t+h}, Y_t) & \text{if incorrect downward prediction.} \end{cases} \tag{3.1}$$

Analogous to L_t^{DA}, we alleviate notation by using the abbreviation L_t^{DV} for $L_t^{\mathrm{DV}}(X_t^h, Y_{t+h}, Y_t)$. In (3.1) H_t^{UU} resp. H_t^{DD} denote the benefit/gain/value to the forecast user when she believes in a directional up– resp. downward forecast and an up– resp. downward movement realizes. Similarly, H_t^{UD} resp. H_t^{DU} denote the cost/loss/value to the forecast user in case of an incorrect directional prediction. As L_t^{DV} depends only on the DF $\widetilde{X}_t$ and not on the exact value of X_t^h it is robust to forecasts which are far apart from Y_{t+h}. Testing hypothesis about $\mathbb{E}\left[L_t^{\mathrm{DV}}\right]$ is readily accomplished within the framework of Diebold and Mariano (1995), as long as L_t^{DV} is stationary. Moreover, testing equality in prediction accuracy of alternative methods, such as naive DFs, can be implemented easily. Notably, for the special case $H_t^{UU} = H_t^{DD} = a$ and $H_t^{UD} = H_t^{DU} = b$, $L_t^{\mathrm{DV}} = L_t^{\mathrm{DA}}$.

The framework implied by (3.1) can also be interpreted as a particular decision environment. More precisely, suppose based on the DF the forecast user decides to take a particular action or not. Depending on the realized direction, the decision/action undertaken implies a

cost or gain defined by (3.1), see below for more specific examples. Notably, L_t^{DV} relates to the payoff matrices in two–state two–action decision environments discussed in Granger and Pesaran (2000a,b), Skouras (2001b), Pesaran and Skouras (2002) or Elliot and Lieli (2009). However, the examples discussed in these papers consider only decision environments in which the payoffs do not depend on the size of the realized movements. Skouras (2001b) considers in an example a similar payoff matrix but focuses on estimating the sign of a mean regression and not on forecast evaluation. Elliot and Lieli (2009) concentrate on constructing predictions for binary outcomes taking a decision–theoretic approach. While they allow the payoff matrix to depend on G observable variables $W_{tg}, g = 1, ..., G$, they do not account for the size of realized movements of Y_t. Furthermore, issues of comparative forecast evaluation are not addressed.

The measure in (3.1) directly targets at the evaluation of the realized economic value of DFs (resp. of the decisions derived from the DFs) and ignores how the decisions are determined. To be more precise consider a reformulation of (3.1)

$$L_t^{\mathrm{DV}} = H_t^{UU}\widetilde{X}_t\widetilde{Y}_t + H_t^{DD}(1-\widetilde{X}_t)(1-\widetilde{Y}_t) + H_t^{UD}\widetilde{X}_t(1-\widetilde{Y}_t) + H_t^{DU}(1-\widetilde{X}_t)\widetilde{Y}_t \ .$$

Granger and Pesaran (2000a,b) and Pesaran and Skouras (2002) measure the realized economic value analogously. But the values of $\widetilde{X}_t$ and $(1-\widetilde{X}_t)$ are derived from the optimal decisions which are determined by comparing the expected costs/gains of taking action and not taking action, $\widetilde{X}_t = 1$, say, if it turns out optimal to take an action. This framework requires the specification of the decision environment of individual agents and distributional assumptions about the underlying DGP. Furthermore, in practical applications for most decision problems complex numerical optimizations are necessary. Pesaran and Skouras (2002) note that: "A widespread application of the decision–based approach in economics is likely to take decades rather than years before becoming a reality." In contrast, L_t^{DV} allows to evaluate/compare forecast methods in a decision environment even if decisions (based on DFs) are not optimal in a decision–theoretic framework.

We illustrate the flexibility of L_t^{DV} by means of some examples. Let $H_t^{UU} = H_t^{DD} = |Y_{t+h} - Y_t|$ and $H_t^{UD} = H_t^{DU} = -|Y_{t+h} - Y_t|$. Then L_t^{DV} captures the ability to forecast the sign and the magnitude of realized changes. See Blaskowitz and Herwartz (2009b) for an application. Such a property is particularly relevant in a decision making context, for instance, when Y_t is a stock price and the DFs are used to make buy/sell decisions. Suppose

the investor obtains a forecast that Y_t will rise within the next h days, i.e. $\widetilde{X}_t = 1$, and she decides to buy one share of the stock. Then, h periods later, she realizes a cash flow of $Y_{t+h} - Y_t$, which is positive if $Y_{t+h} > Y_t$ (correct DF) and negative in case of an incorrect DF. For a downward movement forecast a similar reasoning applies. Hence in this simple decision environment L_t^{DV} is the realized cash flow from the position set up based on the DFs. In the framework of Skouras (2001a) a risk–neutral artificial technical analyst chooses from a set of competitive directional forecasting methods the one which maximizes expected utility. The latter is accomplished by maximizing expected cash flows. Note also that numerous loss functions are scaled in arbitrary units. As a particular merit, the scale of L_t^{DV} is in the units of the forecast variable allowing an immediate interpretation of the forecast value.

An obvious modification measuring realized returns derived from DFs is given by

$$L_t^{\mathrm{DV}} = \begin{cases} |(Y_{t+h} - Y_t)/Y_t| & \text{if } \widetilde{Z}_t = 1 \\ -|(Y_{t+h} - Y_t)/Y_t| & \text{if } \widetilde{Z}_t = 0 \ , \end{cases}$$

where we assume that $Y_t > 0$ (see Gencay (1998) or Anatolyev and Gerko (2005) for an application). The decision making context is the same as the one for L_t^{DV} described above. Another context is provided in Granger and Pesaran (2000a) who derive optimal decisions of switching between stocks and bonds. While they determine optimal decisions based on a payoff matrix without accounting for the size of movements, they assess a trading strategy in terms of realized economic returns. Note that in this case L_t^{DV} is unit–free which is particularly useful when comparing forecast methods for various series with different scale (Armstrong and Collopy 1992). The excess profitability test of Anatolyev and Gerko (2005) can be viewed as a test of the null hypothesis that $\mathbb{E}\left[L_t^{\mathrm{DV}}\right]$ is greater than the expected profits from an artificial benchmark strategy. While the buy/sell signal frequencies of the benchmark and the trading strategy under investigation are equal, the artificial strategy generates buy/sell signals randomly.

More general functions of Y_{t+h} and Y_t can be accommodated within this framework. Consider a swap trading example. A receiver (payer) swap is an agreement between two counterparties. One receives (pays) a fixed amount of money that is determined by the fixed rate R of the agreement on an annual basis, for instance. The other makes (obtains), say, semiannual payments that depend on a floating leg such as the 6 month EURIBOR rate. In a receiver (payer) swap agreement the investor receives (pays) the fixed leg. The fair value

swap rate is defined to be the fixed rate which makes the swap agreement to have a value of zero to both counterparties. Let Y_t denote the fair value swap rate at time t. Furthermore, let $RSW(Y_t, R, \tau)$ be the value of a receiver swap agreement with fixed rate R and termination date τ when the current fair value swap rate is Y_t. Similarly $PSV(Y_t, R, \tau)$ denotes the value of a payer swap. For simplicity, we neglect the dependence of the swap value on other variables (see e.g. Miron and Swannell 1991). The current value of a payer swap with fixed rate $R = Y_t$ is zero, $PSV(Y_t, Y_t, \tau - h) = 0$. If $R < Y_t$ then $PSV(Y_t, R, \tau) > 0$. In such a swap agreement the payer–counterparty pays only the fixed rate R which is less than the current fair value swap rate Y_t at which the swap would have a value of zero. Thus, in swap trading, a speculator decides to enter a payer swap agreement if she expects the fair value swap rate to rise. On the other hand, if the fair value swap rate is expected to fall, a receiver swap agreement is entered. Consequently, a success measure is given by

$$L_t^{\mathrm{DV}} = \begin{cases} PSV(Y_{t+h}, Y_t, \tau - h) & \text{if } PSV(X_t^h, Y_t, \tau - h) > 0 \text{ and } PSV(Y_{t+h}, Y_t, \tau - h) > 0 \\ RSV(Y_{t+h}, Y_t, \tau - h) & \text{if } RSV(X_t^h, Y_t, \tau - h) > 0 \text{ and } RSV(Y_{t+h}, Y_t, \tau - h) > 0 \\ PSV(Y_{t+h}, Y_t, \tau - h) & \text{if } PSV(X_t^h, Y_t, \tau - h) > 0 \text{ and } PSV(Y_{t+h}, Y_t, \tau - h) < 0 \\ RSV(Y_{t+h}, Y_t, \tau - h) & \text{if } RSV(X_t^h, Y_t, \tau - h) > 0 \text{ and } RSV(Y_{t+h}, Y_t, \tau - h) < 0 \ . \end{cases}$$

Notably, $PSV(X_t^h, Y_t, \tau - h)$ and $RSV(X_t^h, Y_t, \tau - h)$ can be any signal that indicates rising or falling values of swap agreements. Moreover, $PSV(Y_{t+h}, Y_t, \tau - h)$ and $RSV(Y_{t+h}, Y_t, \tau - h)$ can be theoretical or observed market prices (see also Blaskowitz and Herwartz (2009a) for an application).

The measure defined in (3.1) can deal with numerous other specifications. For example, instead of assessing the value of directional swap rate forecasts any financial derivative such as stock options can easily be analyzed. H_t^{ij} could also be determined by a utility function such as the negative exponential utility function as in West, Edison and Cho (1993). Furthermore, the framework of DF evaluation is not restricted to financial applications. Business applications include decisions of a company whether to increase production by, say, 3% or not, conditional on predicted changes of macroeconomic aggregates as, for instance, GDP. The DF value could be determined by incremental sales or revenues. In macroeconomics, monetary authorities who have to decide whether to increase or decrease interest rates by 25 basis points given DFs for inflation could use a social welfare/cost function to measure the economic value of DFs. Öller and Barot (2000) investigate the directional accuracy of

European growth and inflation forecasts. Their discussion suggests further macroeconomic applications for the DF measure (3.1). L_t^{DV} also accommodates situations in which directional costs/benefits are asymmetric. For example, consider a strategy to short put options until maturity when the market is predicted to go up or to invest in the cash market when it is expected to go down. In this case, an incorrect upward prediction might be more expensive than an incorrect downward movement, $H_t^{UD} < H_t^{DU}$.

4 Empirical illustration

To highlight the issues discussed above we provide an empirical example. We consider $h = 5$ day ahead forecasts for the 2yr EURIBOR swap rate determined by means of the principal components analysis (PCA) based approach analyzed in Blaskowitz and Herwartz (2009a). They estimate K principal components (or factors) from ω observations for the EURIBOR swap term structure defined by the 3 and 6 month EURIBOR rates, and the 1yr (year), 2yr, 3yr, 5yr, 7yr, 10yr 12yr, 15yr swap rates. Factor forecasts are computed using a vector autoregressive (VAR) model with p lags. Overall, Blaskowitz and Herwartz (2009a) consider 100 different models by combining five estimation windows $\omega \in \{42, 63, 126, 189, 252\}$, five factor choices $K \in \{1, 2, 3, 4, 5\}$ and four lag orders $p \in \{0, 1, 2, 3\}$. For illustrative purposes we focus on the model specification defined by an estimation window of $\omega = 252$ observations, $K = 4$ factors and $p = 1$ autoregressive lag (we abbreviate the model by 252/4/1). Altogether 80 forecasts are produced for the period September 3, 2001 to December 21, 2001.

Results are reported for the mean squared forecast error (MSFE, multiplied by 10^6), for the mean absolute forecast error (MAD, multiplied by 10^3) and for the mean DF value (MDV, multiplied by 100). The latter is defined by average cash flows derived from a swap trading strategy. In each time point t a DF for the 2yr swap rate is derived from the factor model. As outlined in Section 3, an investor decides to enter a 2yr payer (receiver) swap agreement if an increase (decrease) in the 2yr swap rate is predicted. Five days later the economic value of this swap position is determined by means of the comparison swap valuation technique (Miron and Swannell 1991) and the realized 2yr swap rate. We assume that the swap value translates into a hypothetical cash flow if the position were closed in the market. Note that to determine cash flows the swap rate forecast is not needed making the evaluation measure insensitive to outlying forecasts. The economic interpretability of the

MDV measure as average cash flows is obvious as opposed to the economic content of the MSFE resp. MAD criteria. Moreover, note that MSFE/MAD and MDV do not only differ in their economic interpretability. In fact, they assess different forecast properties. While, the criteria MSFE/MAD ignore the direction of the movement and measure (only) the squared/absolute distance of the predicted from the realized swap rate, the MDV accounts for the direction and the size of the movement of the swap rate. Different model rankings for both measures are likely. A model that closely forecasts the 2yr swap rate, may always incorrectly predict the directional movement. On the other hand, a model of which the forecasts are always far away from the outcome may still correctly predict the direction offering a higher economic value to the forecast user in a swap trading application.

The factor model specification 252/4/1 implies a MSFE of 3.80, a MAD of 1.32 and a MDV of 3.27. From Table 1, left panel, it can be seen that in the set of 100 considered forecasting models it is the 65th, 60th and 47th best model in terms of MSFE, MAD and MDV. Inspection of the time series plot of forecasts and actuals for the above model, given in Figure 1, reveals that the 14th forecast is somewhat unreasonably far away from both other forecasts and actual realizations. In order to separate the impact of the outlying forecast from the comparison, we delete it from all models. Then, the model 252/4/1 has a MSFE of 2.59, a MAD of 1.21 and a MDV of 3.01, see the right panel of Table 1. With respect to the MSFE and MAD criteria it is now 4th resp. 14th best model and remains 47th in terms of MDV. Removing the outlier leads to a 30% resp. 8.3% reduction in MSFE resp. MAD and a substantial improvement in the model ranking relative to the remaining 99 specifications, while the MDV comparison remains unaffected.

Percentile	with outlier			outlier removed		
	MSFE$*10^6$	MAD$*10^3$	MDV$*100$	MSFE$*10^6$	MAD$*10^3$	MDV$*100$
1st	2.53	1.14	8.77	2.56	1.15	8.63
10th	2.61	1.19	8.23	2.63	1.20	8.07
30th	2.75	1.23	4.94	2.78	1.23	4.75
40th	2.84	1.24	4.06	2.84	1.25	3.85
60th	3.34	1.32	2.51	3.36	1.33	2.28
70th	3.88	1.46	2.04	3.93	1.48	1.81

Table 1. Percentiles for MSFE$*10^6$, MAD$*10^3$ and MDV$*100$ out–of–sample forecast performance of 5 day–ahead forecasts of 2yr swap rates for the period of September 3, 2001 to December 21, 2001 of 100 PCA–VAR models.

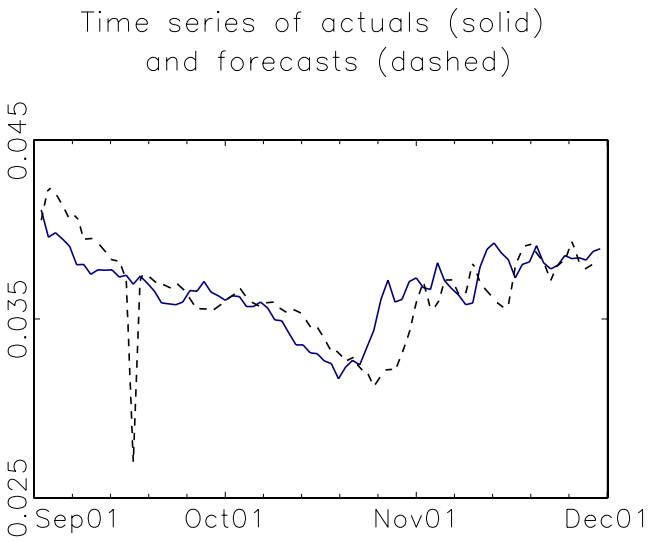


Figure 1. Time series of actuals and out–of–sample 5 day–ahead forecasts for the 2yr swap rate for the period of September 3, 2001 to December 21, 2001 of the model specification 252/4/1.

Deleting outliers from the forecast evaluation is not necessarily the best choice for several reasons. First, it is a delicate matter to define outliers. It might be that large observed forecast errors belong to the tail of the forecast error distribution in which case a removal boils down to truncating this distribution. Second, deleting predictions from all models leads to a loss of information. This is particularly relevant when relatively few forecasts are available as in numerous macroeconomic applications. In addition, taking the evolution of actuals and forecasts into account an applied analyst would doubt the exact value of the outlier(s) but she would probably admit that a further directional movement is not unreasonable. For example, in the case of the 14th forecast as shown in Figure 1 an analyst might believe in a further downward movement. The directional prediction content of the 14th forecast may still be of value. Moreover, visual inspection of the corresponding plots for the 100 models reveals that there are further outliers from time to time. Accounting for the widespread use of PCA–VAR approaches, especially in term structure modelling, it would be inappropriate to discuss the suitability of the forecast method itself. Given the large number of models, a manual outlier removal is time consuming and subjective. Applying an 'insanity filter' based on ad–hoc rules to define and delete outliers reduces the workload but still remains subjective, see, for instance, Elliot and Timmermann (2008). Summarizing, in the presence

of outliers a forecast comparison in terms of MSFE and MAD is subjected to some risk of getting distorted. The robust DF measure represents a meaningful tool for forecast evaluations as it is readily interpretable in economic terms and circumvents numerous problems invoked with outlying forecasts.

5 Conclusion

We discuss a general approach to evaluate (directional) forecasts which is simple to implement, robust to outlying or unreasonable forecasts and which provides an economically interpretable loss/success functional framework. As such, the measure of directional forecast value presented here, is a readily available alternative to the commonly used squared error loss criterion.

Christoffersen and Diebold (1996, 1997), Granger and Pesaran (2000a,b) and Skouras (2007), among others, argue in favor of an integrated approach to allow for general loss functions in modelling, estimation, model selection, prediction and forecast evaluation. By focusing only on the evaluation of forecasts, we account for the fact that frequently only the predictions are available without knowing the method used to produce the latter (e.g. survey/analysts/judgemental forecasts). The underlying rationale is that even if such forecasts are not produced optimally within the above integrated framework, they may contain valuable information with respect to a distinct loss function.

Armstrong and Collopy (1992) argue that a forecast evaluation criterion should be related to decision making. The framework we investigate is related to decision making as it provides the economic value of DFs in a very simple decision problem (buy/sell stocks, increase interest rates or not, etc.). Even if it does not encompass all possible decision problems, it can be seen as a compromise between an individualized decision–theoretic framework and a generalized loss functional approach in a decision making environment.

References

Ahlburg, D. A., 'A commentary on error measures', *International Journal of Forecasting*, **8**, 99–100 (1992).

Armstrong, J. S. and F. Collopy, 'Error measures for generalizing about forecasting methods', *International Journal of Forecasting*, **8**, 69–80 (1992).

Ashiya, M., 'Are 16–month–ahead forecasts useful?: A directional analysis of japanese GDP forecasts', *Journal of Forecasting*, **25**, 201–207 (2006).

Blaskowitz, O. and H. Herwartz, 'PCA based ex–ante forecasting of swap term structures', *International Journal of Theoretical and Applied Finance*, **12**, 465–489 (2009a).

Blaskowitz, O. and H. Herwartz, 'Adaptive forecasting of the EURIBOR swap term structure', *Journal of Forecasting*, **28**, 575–594 (2009b).

Blaskowitz, O. and H. Herwartz, 'Testing for the economic value of directional forecasts in the presence of serial correlation', *SFB 649, Discussion Paper 2008–73* (2008).

Christoffersen P. F. and F. X. Diebold, 'Further results on forecasting and model selection under asymmetric loss', *Journal of Applied Econometrics*, **11**, 561–571 (1996).

Christoffersen P. F. and F. X. Diebold, 'Optimal prediction under asymmetric loss', *Econometric Theory*, **13**, 808–817 (1997).

Diebold, F.X., *Elements of forecasting*, Cincinnati, South-Western College Publishing (2007).

Diebold, F.X. and R. Mariano, 'Comparing predictive accuracy', *Journal of Business & Economic Statistics*, **13**, 253–263, (1995).

Elliot, G. and R.P. Lieli, 'Predicting Binary Outcomes', *Downloaded from http://www.eco.utexas.edu/ rpl227/fbin.pdf*, (2009).

Elliot, G. and A. Timmermann, 'Economic Forecasting', *Journal of Economic Literature*, **46**, 3–56, (2008).

Gençay, R., 'Optimization of technical trading strategies and the profitability in security markets', *Economics Letters*, **59**, 249–254 (1998).

Gradojevic, N. and J. Yang, 'Non–linear, non–fundamental exchange rate forecasting', *Journal of Forecasting*, **25**, 227–245 (2006).

Granger, C.W.J. and M.H. Pesaran, 'A decision–theoretic approach to forecast evaluation', *In Chan, W.S., Li, W.K. and Tong, H. (eds.), Statistics and Finance: An Interface*, London, Imperial College Press (2000a).

Granger, C.W.J. and M.H. Pesaran, 'Economic and statistical measures of forecast accuracy', *Journal of Forecasting*, **19**, 537–560 (2000b).

Greer, M.R., 'Combination forecasting for directional accuracy: An application to survey interest rate forecasts', *Journal of Applied Statistics*, **32**, 607–615 (2005).

Henriksson, R.D. and R.C. Merton, 'On market timing and investment performance II: Statistical procedures for evaluating forecasting skills', *Journal of Business*, **54**, 513–533 (1981).

Hyndman, R. J. and A. B. Koehler, 'Another look at measures of forecast accuracy', *International Journal of Forecasting*, **22**, 679–688 (2006).

Leitch G. and J.E. Tanner, 'Professional economic forecasts: Are they worth their costs?', *Journal of Forecasting*, **14**, 143–157 (1995).

Makridakis, S., 'Accuracy measures: theoretical and practical concerns', *International Journal of Forecasting*, **9**, 527–529 (1993).

Merton, R.C., 'On Market Timing and investment performance I: An equilibrium theory of value for market forecasts', *Journal of Business*, **54**, 363–406 (1981).

Mills, C.T. and G.T. Pepper, 'Assessing the forecasters: An analysis of the forecasting records of the treasury, the london business school and the national institute', *International Journal of Forecasting*, **15**, 247–257 (1999).

Miron, P. and P. Swannell, *Pricing and hedging swaps*. Euromoney Publications PLC, London (1991).

Öller, L.E. and B. Barot, 'The Accuracy of European Growth and Inflation Forecasts', *International Journal of Forecasting*, **16**, 293-315 (2000).

Pesaran, M.H. and S. Skouras, 'Decision–based methods for forecast evaluation', *In Clements, M.P., and D.F. Hendry (eds.), A companion to economic forecasting*, Oxford, Blackwell Publishing (2002).

Pesaran, M.H. and A.G. Timmermann, 'A simple nonparametric test of predictive performance', *Journal of Business & Economic Statistics*, **10**, 461–465 (1992).

Pesaran, M.H. and A.G. Timmermann, 'Predictability of stock returns: Robustness and economic significance', *Journal of Finance*, **50**, 1201–1228 (1995).

Pons, J., 'The rationality of price forecasts: A directional analysis', *Applied Financial Economics*, **11**, 287–290 (2001).

Schnader, M.H. and H.O. Stekler, 'Evaluating predictions of change', *Journal of Business*, **63**, 99–107 (1990).

Schneider, M. and M. Spitzer, 'Forecasting austrian GDP using the generalized dynamic factor model', *Oesterreichische Nationalbank (Austrian Central Bank), Working Paper No. 89* (2004).

Skouras, S., 'Financial returns and efficiency as seen by an artificial technical analyst', *Journal of Economic Dynamics & Control*, **25**, 213–244 (2001a).

Skouras, S., 'Sign of a mean regression: characterisation, estimation and applications', *available at http://ssrn.com/abstract=291922 or DOI: 10.2139/ssrn.291922* (2001b).

Skouras, S., 'Decisionmetrics: A decision–based approach to econometric modelling', *Journal of Econometrics*, **137**, 414–440 (2007).

Swanson, N.R. and H. White, 'A model selection approach to assessing the information in the term structure using linear models and artificial neural networks', *Journal of Business & Economic Statistics*, **13**, 265–275, (1995).

Swanson, N.R. and H. White, 'A model selection approach to real–time macroeconomic forecasting using linear models and artificial neural networks', *Review of Economics and Business Statistics*, **79**, 540–550, (1997a).

Swanson, N.R. and H. White, 'Forecasting economic time series using flexible versus fixed specification and linear versus non–linear econometric models', *International Journal of Forecasting*, **13**, 439–461, (1997b).

Thompson, P. A., 'An mse statistic for comparing forecast accuracy across series', *International Journal of Forecasting*, **6**, 219–227 (1990).

West K. D., H. J. Edison and D. Cho, 'A utility based comparison of some models of exchange rate volatility', *Journal of International Economics*, **35**, 23–45 (1993)

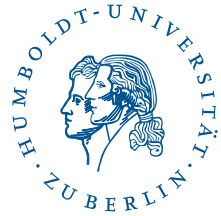
SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Implied Market Price of Weather Risk" by Wolfgang Härdle and Brenda López Cabrera, January 2009.
002 "On the Systemic Nature of Weather Risk" by Guenther Filler, Martin Odening, Ostap Okhrin and Wei Xu, January 2009.
003 "Localized Realized Volatility Modelling" by Ying Chen, Wolfgang Karl Härdle and Uta Pigorsch, January 2009.
004 "New recipes for estimating default intensities" by Alexander Baranovski, Carsten von Lieres and André Wilch, January 2009.
005 "Panel Cointegration Testing in the Presence of a Time Trend" by Bernd Droge and Deniz Dilan Karaman Örsal, January 2009.
006 "Regulatory Risk under Optimal Incentive Regulation" by Roland Strausz, January 2009.
007 "Combination of multivariate volatility forecasts" by Alessandra Amendola and Giuseppe Storti, January 2009.
008 "Mortality modeling: Lee-Carter and the macroeconomy" by Katja Hanewald, January 2009.
009 "Stochastic Population Forecast for Germany and its Consequence for the German Pension System" by Wolfgang Härdle and Alena Mysickova, February 2009.
010 "A Microeconomic Explanation of the EPK Paradox" by Wolfgang Härdle, Volker Krätschmer and Rouslan Moro, February 2009.
011 "Defending Against Speculative Attacks" by Tijmen Daniëls, Henk Jager and Franc Klaassen, February 2009.
012 "On the Existence of the Moments of the Asymptotic Trace Statistic" by Deniz Dilan Karaman Örsal and Bernd Droge, February 2009.
013 "CDO Pricing with Copulae" by Barbara Choros, Wolfgang Härdle and Ostap Okhrin, March 2009.
014 "Properties of Hierarchical Archimedean Copulas" by Ostap Okhrin, Yarema Okhrin and Wolfgang Schmid, March 2009.
015 "Stochastic Mortality, Macroeconomic Risks, and Life Insurer Solvency" by Katja Hanewald, Thomas Post and Helmut Gründl, March 2009.
016 "Men, Women, and the Ballot Woman Suffrage in the United States" by Sebastian Braun and Michael Kvasnicka, March 2009.
017 "The Importance of Two-Sided Heterogeneity for the Cyclicality of Labour Market Dynamics" by Ronald Bachmann and Peggy David, March 2009.
018 "Transparency through Financial Claims with Fingerprints – A Free Market Mechanism for Preventing Mortgage Securitization Induced Financial Crises" by Helmut Gründl and Thomas Post, March 2009.
019 "A Joint Analysis of the KOSPI 200 Option and ODAX Option Markets Dynamics" by Ji Cao, Wolfgang Härdle and Julius Mungo, March 2009.
020 "Putting Up a Good Fight: The Galí-Monacelli Model versus 'The Six Major Puzzles in International Macroeconomics'", by Stefan Ried, April 2009.
021 "Spectral estimation of the fractional order of a Lévy process" by Denis Belomestny, April 2009.
022 "Individual Welfare Gains from Deferred Life-Annuities under Stochastic Lee-Carter Mortality" by Thomas Post, April 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Pricing Bermudan options using regression: optimal rates of convergence for lower estimates" by Denis Belomestny, April 2009.
024 "Incorporating the Dynamics of Leverage into Default Prediction" by Gunter Löffler and Alina Maurer, April 2009.
025 "Measuring the effects of geographical distance on stock market correlation" by Stefanie Eckel, Gunter Löffler, Alina Maurer and Volker Schmidt, April 2009.
026 "Regression methods for stochastic control problems and their convergence analysis" by Denis Belomestny, Anastasia Kolodko and John Schoenmakers, May 2009.
027 "Unionisation Structures, Productivity, and Firm Performance" by Sebastian Braun, May 2009.
028 "Optimal Smoothing for a Computationally and Statistically Efficient Single Index Estimator" by Yingcun Xia, Wolfgang Härdle and Oliver Linton, May 2009.
029 "Controllability and Persistence of Money Market Rates along the Yield Curve: Evidence from the Euro Area" by Ulrike Busch and Dieter Nautz, May 2009.
030 "Non-constant Hazard Function and Inflation Dynamics" by Fang Yao, May 2009.
031 "De copulis non est disputandum - Copulae: An Overview" by Wolfgang Härdle and Ostap Okhrin, May 2009.
032 "Weather-based estimation of wildfire risk" by Joanne Ho and Martin Odening, June 2009.
033 "TFP Growth in Old and New Europe" by Michael C. Burda and Battista Severgnini, June 2009.
034 "How does entry regulation influence entry into self-employment and occupational mobility?" by Susanne Prantl and Alexandra Spitz-Oener, June 2009.
035 "Trade-Off Between Consumption Growth and Inequality: Theory and Evidence for Germany" by Runli Xie, June 2009.
036 "Inflation and Growth: New Evidence From a Dynamic Panel Threshold Analysis" by Stephanie Kremer, Alexander Bick and Dieter Nautz, July 2009.
037 "The Impact of the European Monetary Union on Inflation Persistence in the Euro Area" by Barbara Meller and Dieter Nautz, July 2009.
038 "CDO and HAC" by Barbara Choroś, Wolfgang Härdle and Ostap Okhrin, July 2009.
039 "Regulation and Investment in Network Industries: Evidence from European Telecoms" by Michał Grajek and Lars-Hendrik Röller, July 2009.
040 "The Political Economy of Regulatory Risk" by Roland Strausz, August 2009.
041 "Shape invariant modelling pricing kernels and risk aversion" by Maria Grith, Wolfgang Härdle and Juhyun Park, August 2009.
042 "The Cost of Tractability and the Calvo Pricing Assumption" by Fang Yao, September 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

043 "Evidence on Unemployment, Market Work and Household Production" by Michael C. Burda and Daniel S. Hamermesh, September 2009.
044 "Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, September 2009.
045 "Quantifizierbarkeit von Risiken auf Finanzmärkten" by Wolfgang Karl Härdle and Christian Wolfgang Friedrich Kirchner, October 2009.
046 "Pricing of Asian temperature risk" by Fred Benth, Wolfgang Karl Härdle and Brenda López Cabrera, October 2009.
047 "MM-Stat – MultiMedia-Statistik: Statistische Datenanalyse – webbasiert, interaktiv und multimedial" by Sigbert Klinke, Dina Kuhlee, Christian Theel, Cornelia Wagner and Christian Westermeier, October 2009.
048 "Migration of the Highly Skilled: Can Europe catch up with the US?" by Lydia Mechtenberg and Roland Strausz, October 2009.
049 "A blocking and regularization approach to high dimensional realized covariance estimation" by Nikolaus Hautsch, Lada M. Kyj and Roel C.A. Oomen, October 2009.
050 "Generalized single-index models: The EFM approach" by Xia Cui, Wolfgang Karl Härdle and Lixing Zhu, October 2009.
051 "The Market Impact of a Limit Order" by Nikolaus Hautsch and Ruihong Huang, October 2009.
052 "On economic evaluation of directional forecasts" by Oliver Blaskowitz and Helmut Herwartz, October 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".