Ostfalia
Hochschule für angewandte Wissenschaften

Wolfsburg
Fakultät Wirtschaft

WWP

Wolfsburg Working Papers No. 11-05

On Assessing Economic Forecasts

An evaluation of predictions on GDP, industrial production and private consumer spending trends in twelve industrial nations

Markus Spiwoks, Johannes Scheier and Oliver Hein, 2011

On Assessing Economic Forecasts

An evaluation of predictions on GDP, industrial production and private consumer spending trends in twelve industrial nations

Prof. Dr. Markus Spiwoks (corresponding author)

Ostfalia University of Applied Sciences, Faculty of Business Administration, Chair in Finance, Sieg-fried-Ehlers-Straße 1, D-38440 Wolfsburg, Germany

Georg August University Göttingen, Faculty of Economic Sciences, Chair in Economic Policy and SME Research, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany

E-mail: m.spiwoks@ostfalia.de; markus.spiwoks@wiwi.uni-goettingen.de

Phone: +49–5361–8922–25100; Fax: +49–5361–8922–25004; Mobil: +49–173–6515835

Dipl.-Kfm. (FH) Johannes Scheier

Ostfalia University of Applied Sciences, Faculty of Business Administration, Chair in Finance, Sieg-fried-Ehlers-Straße 1, D-38440 Wolfsburg, Germany

Georg August University Göttingen, Faculty of Economic Sciences, Chair in Economic Policy and SME Research, Platz der Göttinger Sieben 3, D-37073 Göttingen, Germany

E-mail: jo.scheier@ostfalia.de

Prof. Dr. Oliver Hein

Giessen Friedberg University of Applied Sciences, Department of Mathematics, Natural Sciences and Computer Science, Chair in Information Systems, Wilhelm-Leuschner-Str. 13, D-61169 Friedberg, Germany

E-mail: oliver.hein@mnd.fh-friedberg.de

Abstract

This study focuses on economic forecasts and their assessment. We evaluate forecasts on the growth of GDP, industrial production and private consumer spending in twelve major industrial nations. In addition to the unbiasedness test and comparisons with naive forecasts, we apply the information growth test and the TOTA coefficient. Economic forecasts are generally revealed to be quite successful. It is furthermore shown that the unbiasedness test is unsuited when it comes to distinguishing success from failure in terms of forecasting.

Key words: Macro-economic forecasts, forecast accuracy, unbiasedness test, information growth test, TOTA coefficient.

JEL classification: C19, E21, E23, E27, E32, E37

1. Introduction

Economic forecasts are indispensable. Government action and numerous company decisions heavily depend on the prediction of economic trends.

Estimated inland revenue and social spending, and hence the entire financial policy of a country, relies primarily on anticipated economic developments. Should a government feel the need to intervene in these developments, it is likewise dependent on economic forecasts. Even central banks compile elaborate economic forecasts in order, for example, to ward off the imminent overheating of economic growth at an early stage and thus the danger of inflation.

Companies in the manufacturing sector take their lead from economic predictions when it comes to planning future investments and human resources or the timely securing of access to raw materials. Economic growth predictions also play a major role in the banking sector. Forecasts allow banks to assess the probability of default in the case of current and future borrowers, and are essential if they are to make provision for an appropriate risk premium in the interest margin. Economic forecasts are also of major significance for the investment sector. Furthermore, economic trends exert considerable influence on the price of shares and bonds. Financial market predictions are therefore normally based on economic forecasts.

Economic forecasts are as controversial as they are vital. This is mostly due to forecast success falling short of high expectations. Predicting economic trends is a difficult task. The economy is a complex entity in a process of permanent change. Evolutionary objects are particularly difficult to fathom. This was brought home to the business community once again in 2008 in the aftermath of the financial market crisis and ensuing economic crisis.

The limitation to generating reliable economic forecasts is in stark contrast to the often utterly exaggerated demands of forecast users. With his rational expectation hypothesis, Muth (1961) may well have contributed decisively to the unrealistic appraisal of economic forecast possibilities. The rational expectation hypothesis assumes that if all the relevant information is taken into account, a largely accurate economic prediction can be made. Only unexpected random influences (so-called white noise) cause forecasters to occasionally produce inaccurate forecasts. The unbiasedness test is usually applied to identify whether forecast errors are solely the result of unsystematic white noise. If forecasts prove to be biased, they are not considered rational.

Forecast users demand that experts make predictions that are both reasonable and beneficial. It is sometimes difficult to convey that reasonable and beneficial forecasts are not necessarily *rational* in the sense of Muth (1961). Muth assumes that the relevant economic model is well-known and that therefore the workings of the economy have to a large extent been explored. This assumption, however, is obviously incorrect. Only the basic features of the conditions for economic growth are commonplace. Several influencing factors have hitherto not been taken into consideration sufficiently or correctly, or indeed at all. In a world of imperfect knowledge about the internal relations and workings of the economy, forecasters must be judged by other means than the unbiasedness test.

The present study pursues four aims:

1. The study provides a brief overview of the most significant methods of measuring the forecasting quality of macro-economic forecasts.
2. Two new methods of evaluating macro-economic forecasts will be introduced: the information growth test and the TOTA coefficient. We consider these to be particularly suitable approaches, as they offer forecasters realistic challenges.
3. It will be shown that the unbiasedness test does not allow for adequate distinction of forecast successes and failures.
4. The study gives an overview of forecasting success in twelve industrial nations. Specifically this refers to consensus forecasts on the development of GDP, industrial production and private consumer spending in the USA, Japan, Germany, France, the UK, Italy, Spain, Canada, the Netherlands, Switzerland, Sweden and Norway.

The next chapter details the forecast data used in the study. This is followed by a chapter on methodology. The subsequent chapter presents empirical results, while the closing chapter summarizes the most important outcomes of the study.

2. Data

The present study focuses on the assessment of economic forecasts. In particular it is concerned with the growth of GDP, industrial production and private consumer spending. Forecasts refer to the respective percentage changes (deviation) from the previous calendar year. They are taken from the magazine *Consensus Forecasts*, which has been publishing forecasts of various macro-economic magnitude on a monthly basis since October 1989. *Consensus Forecasts* interviews analysts and researchers from banks, insurance companies, investment companies, research institutes and associations. It publishes all of the expert predictions and compiles them into consensus forecasts.

Forecasts refer to percentage changes in the current year compared to the year before, as well as to those of the coming year in comparison with the current year. Experts submit their assessments at the beginning of each month with forecasts referring to the end of the year, when the percentage deviation from the previous year can be calculated. Since forecasts are submitted monthly, each forecast item has twenty-four different forecast horizons. A forecast generated at the beginning of January 2011 for the following year (effective date 31.12.2012), for example, has a forecast horizon of twenty-four months. A forecast for the current year (effective date: 31.12.2011) is likewise generated, producing a forecast horizon of twelve months. Forecasts submitted at the beginning of February for the following and the current year have shortened forecast horizons of twenty-three and eleven months respectively. The two forecasts generated at the beginning of December consequently occasion thirteen-month and one-month forecast horizons. This means that each forecast item and year leads to twenty-four consensus forecasts with horizons ranging from twenty-four months to one month. Hence two years before the end of a specific calendar year, financial experts begin to generate their forecasts. Each month they produce new forecasts until these finally materialize after twenty-four months. This particular pattern of forecast data allows for stimulating methods of analysis.

Table 1: Data taken from *Consensus Forecasts* magazine

Country	Publication begin	First completed forecast year	Last completed forecast year	No. of forecasts analysed
US	Oct. 1989	1991	2009	1,368
Japan	Oct. 1989	1991	2009	1,368
Germany	Oct. 1989	1991	2009	1,368
France	Oct. 1989	1991	2009	1,368
UK	Oct. 1989	1991	2009	1,368
Italy	Oct. 1989	1991	2009	1,368
Spain	Dec. 1994	1996	2009	1,008
Canada	Oct. 1989	1991	2009	1,368
Netherlands	Dec. 1994	1996	2009	1,008
Switzerland	June 1998	2000	2009	720
Sweden	Dec. 1994	1996	2009	1,008
Norway	June 1998	2000	2009	720
Total				14,040

Since October 1989, *Consensus Forecasts* has published forecasts for the USA, Japan, Germany, France, the UK, Italy and Canada. December 1994 saw the addition of Spain, Sweden and the Netherlands, followed in June 1998 by Norway and Switzerland. Our study only takes years into account for which all twenty-four forecast horizons are available. Analysis of the first group of countries mentioned above covers the years 1991 to 2009, the second group 1996 to 2009, and the third, 2000 to 2009 (see Table 1).

Twenty-four individual forecasts for each of the nineteen years (1991 – 2009) observed are available for the first group of countries, i.e., a total of 456 predictions. Since we look at three forecast items (GDP, industrial production, private consumer spending) for each country, we reach a sum total of 1,368 predictions per country in the first group. Altogether we evaluate 14,040 consensus forecasts.

3. Methodology

At best, forecasts should be unbiased (Diebold, 2007). In other words the expected value deviation between forecasts and actual events should amount to null. The residual time series should furthermore contain no systematic components but rather be characterized by random distribution.

The most common version of the unbiasedness test is the Mincer-Zarnowitz regression (Ball, 1962, Mincer and Zarnowitz, 1969). A_t represents the actual event at the moment in time t, P_t represents the prediction of this event, and u_t a residual at the moment in time t.

(1) $$A_t = \alpha + \beta\, P_t + u_t$$

According to this equation, forecasts are unbiased if α does not significantly differ from 0, β does not significantly differ from 1, and the error term u is not autocorrelated.

We, however, take a very critical stance on the unbiasedness test, which is rooted in the theory of rational expectations. In order that forecasts be considered unbiased, several criteria must be fulfilled: 1. Events in the real economy should have stable structures. 2. The model must be accurate and unambiguous. 3. Financial forecasters must constantly absorb the relevant information to a satisfactory degree.

The initial euphoria of neo-classical theory may have prompted some willingness to acknowledge at least the approximate accuracy of such assumptions. On the other hand, the bank and financial market crises in 2008, the subsequent economic crisis in 2009, the sovereign debt crisis in 2010, and the imminent currency crisis have forced economic scientists to question the current model.[1] Behavioral economics has produced numerous documents to corroborate that even highly gifted subjects are not in a position to evaluate all of the relevant information to an appropriate degree. Transaction cost theory illustrates that – even if it were possible – it is not always expedient to do so. Evolutionary economics has given rise to deep-seated doubts as to whether long-term, stable structures are in fact a feature of this economy model. Against this backdrop we regard demands on financial analysts to generate unbiased forecasts as highly unjustified.

Our study nevertheless includes the Mincer-Zarnowitz version of the unbiasedness test. We intend to show that this widely used approach leads to quite different results than the methods we suggest for forecast assessment.

The Holden and Peel (1990) efficiency test is based on the Mincer-Zarnowitz regression.

$$A_t = \alpha + \beta\, P_t + \gamma\, X_t + u_t \tag{2}$$

X_t is an optional piece of available information. According to equation 2, forecasts are efficient if α and γ do not significantly differ from 0, β does not significantly differ from 1, and the error term u is not autocorrelated.

If γ differs significantly from null, information X has not been adequately incorporated into the forecasts. Criticism of this approach is largely akin to that of the unbiasedness test.

Another common version of the efficiency test examines whether the actual values prior to forecast submission have a systematic influence on forecast errors (Simon, 1989). Should this be the case, actual events contain information not taken into account in the forecasts. A_t represents the actual event at the moment in time t, P_t represents the forecast of this event, h the forecast horizon and u_t a residual at the moment in time t.

$$A_t - P_t = \alpha + \sum_{i=1}^{I} \beta_i\, A_{t-h-i} + u_t \tag{3}$$

If the available information has been used efficiently, experts' forecast errors should not be correlated with the lags. Whether an existing correlation between the forecast errors and the lag variables can be viewed as significant is determined with the aid of the F test.

[1] Very few economists at the annual meeting of the American Economic Association (AEA) 2010 in Atlanta, like Thomas Sargent of New York University and Robert Barro of Harvard University, declared their unswerving loyalty to the basics of neo-classical economics, e.g., the theory of rational expectations (Storbeck, 2010).

We see this approach in a more positive light. Forecasters must acknowledge actual events of the recent past prior to forecast submission. If a systematic correlation can be established beween past and future events, financial analysts must be expected to make use of this possibility.

The sign accuracy test (Merton, 1981; Henriksson and Merton, 1981) is a commonly used tool for forecast assessment. It does not focus on the extent of forecast deviation but merely verifies the accuracy of the forecast trend. Forecasts are then incorporated in a 2 x 2 contingency table. A distinction is made, on the one hand, between predicted economic growth acceleration or slowdown and, on the other hand, whether growth acceleration or an economic slowdown has in fact taken place. The principal diagonal in the 2 x 2 contingency table indicates the accurate forecast trend. The inaccurate forecast trend is conveyed by the secondary diagonal. A χ^2 test is now applied to examine whether the distribution frequency of the four fields is significantly different from a random walk forecast (cf. Diebold and Lopez, 1996; Joutz and Stekler, 2000). If this is the case, it should be determined whether the forecasts under review were significantly better or significantly worse than a random walk forecast.

We consider this a suitable approach. It examines whether forecasters are ill-informed and merely guessing or have at least gained a rough understanding of the forecast subject matter. No exaggerated demands are placed on the experts.

The golden parameter for measuring forecast quality remains the comparison with naive forecasts. Let us assume that a black box generates a quantifiable event in regular time intervals. We can observe the time series of these events, but we have no insight whatsoever into the processes occurring inside the black box or how the visible results were generated. Let us also assume that despite our complete ignorance we have to make a forecast on the future tendency of the time series. As we have no information on the genesis of events, the future increasing and decreasing course of the time series are equally probable. Thus it seems sensible to assume an unchanged situation in the future (naive forecast). This idea goes back to the French mathematician Pierre Simon Laplace (1814), who introduced it into the literature as the "principle of insufficient reason". The naive forecast has been appraised since then as the rock bottom of forecast quality. Even if nothing is known about the forecast subject, naive forecast quality is easy to achieve. If an expert at least roughly understands the processes to be forecast, his forecasts should be of better quality than naive forecasts (Spiwoks, Bedke and Hein, 2009, p. 8).

There are several ways of making comparisons with naive forecasts. First of all, simple forecast accuracy measurements are easily calculated for both expert and naive forecasts. The most common forecast accuracy measurements are mean error (*ME*), root mean squared error (*RMSE*), mean absolute error (*MAE*), median absolute error (*MdAE*), mean absolute percentage error (*MAPE*), median absolute percentage error (*MdAPE*), symmetric mean absolute percentage error (*sMAPE*), symmetric median absolute percentage error (*sMdAPE*), mean relative absolute error (*MRAE*), median relative absolute error (*MdRAE*) and geometric mean relative absolute error (*GMRAE*) (De Gooijer and Hyndman, 2006; Mathews and Diamentopoulos, 1994).

Comparison with naive forecasts is more manageable if it is implicit in the forecast accuracy measurement. Theil's *U* is a well-known example (Theil, 1971). If a disproportionate emphasis on large forecast errors is to be avoided, the mean absolute error relative to naive forecasts (*MAERNF*) should be used.

(4)
$$MAERNF = \frac{\sum_{t=1}^{T} |P_t - A_t|}{\sum_{t=1}^{T} |A_{t-h} - A_t|}$$

A_t represents the actual event at the moment in time t, P_t represents the forecast of this event und h the forecast horizon.

Fair and Shiller (1990) have shown that in certain constellations these approaches can lead to controversial results. For this reason the Diebold Mariano test for forecast encompassing (Diebold and Mariano, 1995) is more often applied. The initial premise here is that a forecasted situation P_m is described by two competing forecast models i and j:

(5)
$$P_m = (1-\lambda)\, P_{i,m} + \lambda\, P_{j,m}$$

where $0 \le \lambda \le 1$. If $\lambda = 0$, then the forecasts generated by model i are said to encompass the forecasts generated by model j, as model j does not contribute any useful information – apart from that already contained in model i – to the formation of an optimal composite forecast. Harvey, Leybourne and Newbold (1998) develop a statistic to test the null hypothesis that H_0: $\lambda = 0$ against the alternative that H_1: $\lambda > 0$. If the null hypothesis is rejected, then the forecasts contain distinct predictive information that is useful in forming the optimal forecast P_m. Taking naive forecasts as model i and expert forecasts as model j allows for determination of whether the information content of expert forecasts significantly exceeds that of naive forecasts.

The comparison with naive forecasts is useful. Prediction methods that fail to produce better results than the corresponding naive forecasts are undoubtedly ill-advised. The Diebold Mariano test for forecast encompassing offers a highly elegant method of conducting a naive forecast comparison. We doubt the sense, however, of only regarding statistically significant better results as successful, and that at every turn. No one at the Olympics, for instance, would think of asking whether the 100 metres gold medalist could run statistically significant faster than the silver medalist.[2] When it comes to avoiding exaggerated demands on expert forecasters, comparisons on the basis of simple forecast accuracy measurements are still a good option. We apply the mean absolute error relative to naive forecasts (*MAERNF*) because it constitutes a straightforward measure of forecast accuracy, the interpretation of which is both easy and conclusive.

It is not advisable, however, to ascertain the measure of forecast accuracy for all twenty-four forecast horizons. Actual events occur annually, not monthly. Hence there is only one naive forecast per year that can serve as a measure of comparison in terms of the *MAERNF*. We evaluate forecasts with a twenty-four and a twelve-month forecast horizon, since forecasts submitted in January can be suitably compared with naive forecasts (actual event of the previous year). It could be argued here that data pertaining to actual economic growth in the previous year is not yet available at the beginning of January. As a rule it takes up to March for the first largely reliable figures to appear, which explains why we also conduct a *MAERNF* evaluation of forecasts submitted at the beginning of April. This applies to forecasts with forecast horizons of twenty-one and nine months. Since we are keen to avoid exorbitant demands on forecasters, we deem the examination of predictions submitted at the beginning of April productive. After all the low-cost alternative of constant reliance on naive forecasts is not available until the beginning of April.

The further (nearer) non-deterministic events lie in the future, the more difficult (easy) they are to predict. This is because events observed over time lead to a more accurate assessment of the subject matter (Poulizac, Weale and Young, 1996, Ackert and Hunter, 1994, pp. 390-391, Döpke and Fritsche, 2006, p. 789). It is easier to predict the final outcome of a soccer match in the eightieth minute than at the kick-off. Predicting the losing team in a soccer match is not difficult when two players from one team have been shown red cards and sent off by the

[2] It is certainly possible to include statistical significance tests in the assessment of sporting achievements, albeit sprinters would have to compete against each other thirty times over. Average times and the scattering of individual results would allow for calculation of whether the winner runs significantly faster than the runner-up. Most sport enthusiasts, however, would consider this an absurd approach.

eightieth minute and the other team is winning by six goals to nil. Forecasting the outcome of a hundred soccer matches at the kick off would therefore generate many more errors than if predictions were made in the eightieth minute of the game. The precondition is a rudimentary knowledge of soccer and the acknowledgement of the relevant events (e.g., goals, send-offs, cautions) leading up to the eightieth minute.

Similar could be expected of economic experts. They should have basic knowledge of economic processes[3] and at least reflect on the most salient of the pertinent events. Once these two criteria have been fulfilled, their predictions should on average be all the more accurate, the shorter the forecast horizon.

It could be assumed that economic forecasters generally meet these demands without any great effort or difficulty. In reality, however, financial market analysts – as one example – by no means meet the requirements profile consistently. Spiwoks (2009) illustrates that the forecast error in profit trend predictions for US companies does not decrease with the lessening of the forecast horizon. Forecasts with a forecast horizon of only one month are on average less accurate than those with a twenty-four-month forecast horizon. This is particularly surprising, since these companies publish their profits on a quarterly basis. In the case of short forecast horizons, therefore, key elements of the events to be predicted are already known. Analysts are apparently unwilling to include such interim results in their predictions.

We therefore suggest applying an information growth test. This new and at the same time simple approach examines whether forecasters are willing and in a position to take relevant events during a gradually diminishing forecast horizon into account. The test determines from a statistical perspective whether there is a significant decline in forecast errors as the forecast horizon diminishes.

Forecast errors are calculated as absolute errors (*AE*) or absolute percentage errors (*APE*) for each forecast horizon. Subsequently the correlation coefficient between the forecast horizon and the forecast error is calculated. If the coefficient indicates a statistically significant positive sign, forecast errors decrease discernibly with the shortening of the forecast horizon. The minimum requirement for forecasters can be seen as fulfilled in this case.

The information growth test belongs to the efficiency test category, since it ascertains whether forecasts have considered the relevant information adequately into account.

[3] In the light of current economic science, more is probably not possible.

When forecasts at their date of issue show a stronger correlation to actual trends than they do at their date of validity, we speak of topically oriented trend adjustment (TOTA). Such an adjustment is present when forecasts describe the progression of naive forecast time series rather than the actual future progression of the forecast object. The TOTA coefficient serves to detect possible topically oriented trend adjustments.

Prior to calculating the TOTA coefficient (see Andres and Spiwoks, 1999, Bofinger and Schmidt, 2003), the coefficient of determination of the forecast data and actual events is worked out (R^2_A). This is followed by calculation of the coefficient of determination of the forecast data and actual events from the forecast date of issue (R^2_B).

(6) $$\text{TOTA coefficient} = \frac{R^2_{\text{forecasts (validity date); actual}}}{R^2_{\text{forecasts (issue date); actual}}} = \frac{R^2_A}{R^2_B}$$

If the value of the TOTA coefficient is < 1, a topically oriented trend adjustment must be assumed. In this case the forecast time series transferred back to the issue date shows a higher correspondence with actual values than it did at the date of validity. Hence for a TOTA coefficient < 1, the forecast time series reflects the present more strongly than the future.

Topically oriented trend adjustments occur regularly in financial market forecasts. Spiwoks, Gubaydullina and Hein (2011) examine around 1,200 interest rate forecast time series with approximately 160,000 individual predictions. It was found that 98.5% of all forecast time series are characterized by topically oriented trend adjustment. Financial market analysts are evidently unable or refuse to break away from the prevailing market situation. Thus for the most part they "forecast" the present, rarely the future. Schuh (2001, p. 42), Mehra (2002, pp. 21-22), Mankiw, Reis and Wolfers (2003, p. 212), Bowles et al. (2007, p. 18), Andolfatto, Hendry and Moran (2008, p. 407) and Dovern, Fritsche and Slacalek (2009, pp. 38-43) give first indications that topically oriented trend adjustments can also occur in macro-economic forecasts.

The TOTA coefficient has not been applied up to now in the analysis of macro-economic forecasts. This is most likely because the TOTA coefficient is only suitable for analysis of forecast time series with a consistent forecast horizon. If individual horizons are selected from the available twenty-four forecast horizons, the TOTA coefficient can also be applied to macro-economic forecast time series. We use it on forecasts with horizons of twelve and nine months, both of which seem particularly suited to the purpose. Since forecasters will still have the past year in mind in January, there may be an inclination towards topically oriented trend adjustment. Fairly reliable data on economic growth in the previous year is usually available by the beginning of April. It is perhaps this data that prompts topically oriented trend adjustment.

We analyse economic forecasts with four of the methods presented. We consider the first three, i.e., the mean absolute error relative to naive forecasts (*MAERNF*), the information growth test and the TOTA coefficient as particularly suitable tools. Finally we use the unbiasedness test to illustrate how serious miscalculations are produced when an unsuitable approach is adopted for forecast assessment.

4. Results

The comparison with forecasts based on the mean absolute error relative to naive forecasts (*MAERNF*) shows that with few exceptions economic forecasters are successful (Table 2). In the case of forecast horizons of twenty-four and twelve months, predictions on thirty-three of the thirty-six forecast items (91.7%) were more accurate than those of the naive forecast alternative.[4] With a forecast horizon of twenty-one months, expert predictions showed less forecast errors than the corresponding naive forecasts in thirty-four of the thirty-six cases (94.4%). With regard to the shortest forecast horizon under review (nine months), forecasts for all thirty-six forecast items were superior to naive forecasts.

Naive forecasting constitutes the rock bottom of forecast quality (Fildes and Stekler, 2002, p. 439). In order to justify their activity, economic experts must generate forecasts that are significantly better than the corresponding naive forecasts. Free of charge, the latter represent a convenient alternative. Hence there seems to be no cause for jubilation when economic forecasters deliver a higher performance than naive forecasting in 136 of 144 cases (94.4%).

In order to show worthy appreciation of economic experts, we should call to mind the success quota of financial market forecasters. Numerous studies show that between 80 and 100% of financial market forecast time series produce results that are worse than the corresponding naive forecasts (cf. for share forecasts: Lakonishok, 1980, Fraser and MacDonald, 1993, Spiwoks, 2004; cf. for interest forecasts: Brooks and Gray, 2004, Mose, 2005, Spiwoks,

[4] Osterloh (2008) uses Theil's *U* as a benchmark and reaches a less favourable outcome for German GDP forecasts with twenty-four-month forecast horizons. The variation in results could be due to the choice of benchmark. It might also be the result of the specific observation period. Osterloh evaluates forecast data from 1995 to 2005.

Bedke and Hein, 2008, Spiwoks, Bedke and Hein, 2009, Spiwoks, Bedke and Hein, 2010; cf. for exchange rate forecasts: Manzur, 1988, Chinn and Frankel, 1994, Bofinger and Schmidt, 2003). Seen in this light the economic forecasts under review appear quite successful.

The information growth test likewise shows a positive image (Table 3). A glance at the correlation coefficients between forecast horizons and absolute errors (*AE*) reveals a positive sign in thirty-five of thirty-six forecast items (97.2%). This outcome is shown to be statistically significant in thirty-three of these thirty-five cases (94.3%) with an error probability of 1%. The remaining two cases show statistical significance with an error probability of 5%. As the sole exception, Norwegian forecasts on manufacturing production show a negative sign. Statistically, however, this is insignificant.

The correlation appears somewhat less compelling when correlation coefficients between forecast horizons and absolute percentage errors (APE) are considered (Table 3). Although all thirty-six correlation coefficients show a positive sign, only twenty-seven (75%) can be seen as statistically significant with an error probability of 1%. Thirty-two of the thirty-six correlation coefficients (88.9%) prove to be statistically significant despite a 5% error probability. In the case of a 10% error probability, this applies to thirty-three of thirty-six correlation coefficients (91.7%).

Thus the information growth test shows an absence of statistical significance in only four out of seventy-two cases (5.6%). Sixty-eight of the seventy-two cases (94.4%), on the other hand, give an indication that the shorter the forecast horizon, the smaller the number of forecast errors. The overwhelming majority of economic forecasters is therefore willing and in a position to absorb the relevant information over time and incorporate it into their predictions.

It may at first make little impression that economic experts only submit forecasts that have taken account of relevant facts over time. In the case of financial market forecasters, however, it has been shown that this cannot be taken for granted.[5]

[5] Cf., for example, Spiwoks (2009).

Table 2: Results of the comparison with naive forecasts based on the mean absolute error relative to naive forecasts (*MAERNF*) for selected forecast horizons

Country	Forecast item	*MAERNF* 24 months	*MAERNF* 21 months	*MAERNF* 12 months	*MAERNF* 9 months
US	GDP	0.972	0.941	0.793	0.639
	Industrial production	0.989	0.966	0.808	0.645
	Personal consumption	0.914	0.827	1.012	0.741
Japan	GDP	0.825	0.780	0.736	0.551
	Industrial production	0.741	0.726	0.643	0.432
	Private consumption	0.932	0.836	0.878	0.633
Germany	GDP	0.813	0.772	0.670	0.519
	Industrial production	0.750	0.739	0.620	0.415
	Private consumption	0.810	0.871	0.952	0.885
France	GDP	0.785	0.779	0.751	0.502
	Manufactur. production	0.879	0.859	0.825	0.623
	Household consumption	0.836	0.860	1.074	0.914
UK	GDP	0.798	0.769	0.809	0.640
	Manufactur. production	0.953	0.927	0.708	0.485
	Household consumption	0.879	0.820	0.800	0.595
Italy	GDP	0.891	0.879	0.728	0.561
	Industrial production	0.951	0.968	0.867	0.612
	Household consumption	0.943	0.892	0.888	0.746
Spain	GDP	0.914	0.838	0.931	0.822
	Industrial production	0.846	0.831	0.852	0.555
	Household consumption	0.744	0.726	0.983	0.813
Canada	GDP	0.789	0.794	0.774	0.535
	Industrial production	1.087	1.084	1.217	0.877
	Personal expenditure	0.737	0.749	0.800	0.669
Netherlands	GDP	0.875	0.853	0.883	0.730
	Manufactur. production	0.788	0.797	0.767	0.568
	Private consumption	1.010	0.978	0.970	0.932
Switzerland	GDP	0.773	0.773	0.713	0.653
	Industrial production	0.647	0.653	0.625	0.561
	Private consumption	0.702	0.685	0.815	0.884
Sweden	GDP	0.823	0.827	0.855	0.673
	Manufactur. production	0.754	0.781	0.690	0.511
	Household consumption	0.854	0.871	0.859	0.735
Norway	GDP	0.836	0.842	0.756	0.714
	Manufactur. production	1.218	1.182	0.497	0.525
	Private consumption	0.809	0.767	0.742	0.689

MAERNF = mean absolute error relative to naive forecasts

Table 3: Results of the information growth test for selected forecast accuracy measures and TOTA coefficients (TOTA) for selected forecast horizons

Country	Forecast item	Correlation fh - *AE*	Correlation fh - *APE*	TOTA 12 months	TOTA 9 months
US	GDP	0.408***	0.215***	1.562	2.102
	Industrial production	0.342***	0.233***	1.606	2.072
	Personal consumption	0.375***	0.177***	1.029	1.432
Japan	GDP	0.312***	0.182***	0.627	1.585
	Industrial production	0.368***	0.272***	1.168	4.904
	Private consumption	0.233***	0.201***	0.507	1.011
Germany	GDP	0.386***	0.127***	3.032	6.248
	Industrial production	0.371***	0.083***	3.099	6.329
	Private consumption	0.187***	0.227***	1.439	1.957
France	GDP	0.415***	0.122***	1.178	2.225
	Manufactur. production	0.335***	0.276***	1.643	2.617
	Household consumption	0.283***	0.164***	0.760	1.592
UK	GDP	0.212***	0.089**	1.704	1.860
	Manufactur. production	0.390***	0.234***	2.680	2.765
	Household consumption	0.287***	0.205***	1.624	1.900
Italy	GDP	0.417***	0.155***	1.413	1.692
	Industrial production	0.370***	0.251***	1.598	2.969
	Household consumption	0.280***	0.223***	0.836	1.285
Spain	GDP	0.281***	0.208***	1.309	1.405
	Industrial production	0.352***	0.257***	0.775	1.104
	Household consumption	0.250***	0.144**	1.166	1.273
Canada	GDP	0.425***	0.314***	2.394	2.888
	Industrial production	0.334***	0.156***	3.004	3.403
	Personal expenditure	0.283***	0.147***	1.347	1.882
Netherlands	GDP	0.330***	0.171***	1.352	2.033
	Manufactur. production	0.285***	0.154***	2.749	3.323
	Private consumption	0.420***	0.218***	1.409	1.295
Switzerland	GDP	0.500***	0.267***	4.501	9.678
	Industrial production	0.366***	0.163**	2.476	2.858
	Private consumption	0.160**	0.054	0.816	0.643
Sweden	GDP	0.310***	0.133**	0.840	1.229
	Manufactur. production	0.242***	0.163***	1.573	2.634
	Household consumption	0.325***	0.077	0.268	0.909
Norway	GDP	0.141**	0.102*	1.004	2.524
	Manufactur. production	–0.043	0.111	1.739	3.372
	Private consumption	0.391***	0.192**	1.138	1.438

Correlation = coefficient of correlation; fh = forecast horizon; *AE* = absolute error; *APE* = absolute percentage error; TOTA = TOTA coefficient; level of significance: 1%***, 5%** and 10%*.

Table 4: Results of the unbiasedness test in its Mincer-Zarnowitz version

Country	Forecast item	α	β	F test p value	DW p value
	GDP	0.264	0.890	**0.001** $^{\eta}$	0.000
US	Industrial production	-0.402	0.937	0.000	0.000
	Personal consumption	0.272	1.029	0.000 $^{\eta}$	0.000
	GDP	-0.059	0.650	0.000	0.000
Japan	Industrial production	-1.627	0.803	0.000	0.000
	Private consumption	0.347	0.484	0.000 $^{\eta}$	0.000
	GDP	-0.064	0.879	0.000	0.000
Germany	Industrial production	-1.101	0.973	0.000	0.000
	Private consumption	0.453	0.717	0.000 $^{\eta}$	0.000
	GDP	-0.230	0.915	0.000	0.000
France	Manufactur. production	-1.759	0.977	0.000	0.000
	Household consumption	0.010	0.914	0.001 $^{\eta}$	0.000
	GDP	-0.259	1.143	**0.000** $^{\eta}$	0.000
UK	Manufactur. production	-1.151	0.907	0.000	0.000
	Household consumption	-0.382	1.249	0.000 $^{\eta}$	0.000
	GDP	-0.671	1.012	0.000	0.000
Italy	Industrial production	-1.765	1.047	0.000	0.000
	Household consumption	-0.197	0.832	0.000	0.000
	GDP	-0.475	1.300	0.000 $^{\eta}$	0.000
Spain	Industrial production	-2.044	1.028	0.000 $^{\eta}$	0.000
	Household consumption	-0.984	1.410	0.000 $^{\eta}$	0.000
	GDP	-0.699	1.172	0.000	0.000
Canada	Industrial production	-0.648	0.654	0.000 $^{\eta}$	0.000
	Personal expenditure	0.143	1.010	**0.002** $^{\eta}$	0.000
	GDP	0.089	1.045	0.081	0.000
Netherlands	Manufactur. production	-0.443	0.940	0.000	0.000
	Private consumption	0.178	0.984	**0.179** $^{\eta}$	0.000
	GDP	0.133	1.014	0.191	0.000
Switzerland	Industrial production	-0.816	1.065	0.062	0.000
	Private consumption	0.693	0.581	0.000	0.000
	GDP	-0.387	1.153	0.125	0.000
Sweden	Manufactur. production	-3.188	1.228	0.000 $^{\eta}$	0.000
	Household consumption	0.801	0.688	0.000 $^{\eta}$	0.000
	GDP	0.702	0.419	0.000 $^{\eta}$	0.000
Norway	Manufactur. production	-1.535	0.396	0.000	0.000
	Private consumption	0.465	1.005	0.000	0.000

η = Since heteroskedasticity cannot be excluded, the p value was calculated with robust standard errors; p values that altered with robust standard error estimates are emphasized in bold. DW = Durbin-Watson test.

Table 5: Comparison of forecast successes (+) and forecast failures (–) with selected benchmarks

Country	Forecast item	*MAERNF* 9 months	Correlation fh - *AE*	TOTA 9 months	Unbiasedness test
US	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Personal consumption	+	+	+	–
Japan	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Private consumption	+	+	+	–
Germany	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Private consumption	+	+	+	–
France	GDP	+	+	+	–
	Manufactur. production	+	+	+	–
	Household consumption	+	+	+	–
UK	GDP	+	+	+	–
	Manufactur. production	+	+	+	–
	Household consumption	+	+	+	–
Italy	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Household consumption	+	+	+	–
Spain	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Household consumption	+	+	+	–
Canada	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Personal expenditure	+	+	+	–
Netherlands	GDP	+	+	+	–
	Manufactur. production	+	+	+	–
	Private consumption	+	+	+	–
Switzerland	GDP	+	+	+	–
	Industrial production	+	+	+	–
	Private consumption	+	+	–	–
Sweden	GDP	+	+	+	–
	Manufactur. production	+	+	+	–
	Household consumption	+	+	–	–
Norway	GDP	+	+	+	–
	Manufactur. production	+	–	+	–
	Private consumption	+	+	+	–

+ = forecast success; – = forecast failure; fh = forecast horizon; *AE* = absolute error; *MAERNF* = mean absolute error relative to naive forecasts; TOTA = TOTA coefficient.

Similar to naive forecast comparison and the information growth test, the TOTA coefficient likewise conveys a fairly positive image of economic expert achievements (Table 3). Twenty-eight of the thirty-six forecast time series with a twelve-month forecast horizon (77.8%) and thirty-four out of thirty-six forecast time series with a nine-month forecast horizon show no indication of topically oriented trend adjustment. In other words the overwhelming majority of forecast time series reflects the future to a greater extent than the present. Only ten out of seventy-two forecast time series (13.9%) display a TOTA coefficient > 1. These time series correspond more closely to actual events at the time of forecast submission than to those at the time of forecast validity and are therefore characterized by topically oriented trend adjustment.

Forecasts should be geared to the future. Those that tend to reflect the present rather than the future are in fact superfluous. Hence topically oriented trend adjustment in the domain of economic forecasting should be the exception to the rule. In financial market forecasts, however, topically oriented trend adjustments represent the rule rather than the exception. Countless studies have illustrated that more than 98% of all financial market forecast time series feature topically oriented trend adjustments (Bofinger and Schmidt, 2003, Spiwoks, 2004, Spiwoks, Bedke and Hein, 2008, Spiwoks, Bedke and Hein, 2009, Spiwoks, Bedke and Hein, 2010, Spiwoks, Gubaydullina and Hein, 2011). Against this backdrop it should therefore be regarded as an economic forecasting accomplishment that only a small number of forecast time series are characterized by topically oriented trend adjustment.

The unbiasedness test produces a completely different assessment of forecasting success (Table 4). It gives the impression that forecaster efforts have failed utterly. In thirty-one of the thirty-six forecast items observed (86.1%), the F test with an error probability of 1% indicates that $\alpha \neq 0$ and/or $\beta \neq 1$. With an error probability of 10%, this even applies to thirty-three of the thirty-six forecast items (91.7%).[6]

The situation appears even more unfavourable when forecast errors are examined for systematic elements with the help of the Durbin-Watson test. With an error probability of well under 1%, all thirty-six cases show evidence of autocorrelation.

[6] The Breusch-Pagan test (Breusch and Pagan, 1979) demonstrates that heteroscedasticity must be assumed in sixteen of the thirty-six data items. We applied the F test with robust standard errors to these sixteen forecast items (marked η). Under consideration of three decimal places this led in four cases (bold emphasis) only to an alteration of the p value.

Hence the unbiasedness test earmarks forecasts on all thirty-six forecast items as biased.[7] Inasmuch as this can be considered an appropriate benchmark, it indicates that forecast endeavours have failed dismally. Keane and Runkle (1990) as well as Bonham and Cohen (2001) point out that consensus forecasts may appear unbiased although the individual predictions they contain are indeed biased. Even this factor did not lead to a more favourable outcome in the present data analysis.

With the exception of the unbiasedness test, many of the observed forecast items satisfy all eight benchmarks (Tables 2 and 3):

Forecasts for GDP growth in the USA, Germany, France, Italy, Spain, Canada, the Netherlands and Switzerland (a) prove to be superior for all four forecast horizons (24, 21, 12 and 9 months) to those of naive forecasting; (b) indicate information growth over time by means of absolute error and absolute percentage error with an error probability of 1%; (c) show no evidence in the twelve- and nine-month forecast horizons of topically oriented trend adjustment.

Forecasts on industrial production growth for the USA, Japan, Germany, France, the UK, Italy and the Netherlands (a) are shown for all four forecast horizons (24, 21, 12 and 9 months) to be superior to naive forecasts; (b) show evidence of information growth over time by means of absolute error and absolute percentage error with an error probability of 1%; (c) forecast horizons of both twelve and nine months show no evidence of topically oriented trend adjustment.

Forecasts on growth of private consumer spending for Germany, the UK and Canada (a) are shown for all four forecast horizons (24, 21, 12 and 9 months) to be superior to naive forecasts; (b) signal information growth over time by means of absolute error and absolute percentage error with an error probability of 1%; (c) show no evidence in the case of twelve- and nine-month forecast horizons of topically oriented trend adjustment.

The following requirements are simultaneously fulfilled for thirty-three of thirty-six forecast horizons (91.7%): (a) in the case of the nine-month forecast horizon, naive forecasts are sur-

---

[7] Ager, Kappler and Osterloh (2007) explore among other factors GDP forecasts from 1996 to 2006 for most of the countries under review here. In the majority of cases they come to the conclusion that the forecasts are unbiased (p. 13). Dovern and Weisser (2009) examine GDP predictions for the G7 countries from 1991 to 2005. They too reach the conclusion that most forecasts – particularly in periods with no serious structural changes – are unbiased (p. 20). These deviations are probably due to shorter observation periods, the varying number of forecasts or other versions of the unbiasedness test. Batchelor (2007) investigated GDP forecasts for the period 1990 to 2005. In the case of Japan, Italy, Germany and France, the forecasts proved to be biased (p. 20).

passed; (b) a close look at the absolute error reveals a statistically significant information growth; (c) there is no evidence of topically oriented trend adjustment in the case of nine-month forecast horizons (Table 5). As sole exceptions, growth forecasts on private consumer spending in Switzerland and Sweden, and forecasts on industrial production in Norway fail to fulfill these three criteria to the fullest extent.

Table 5 also indicates a strong discrepancy between our preferred benchmarks, on the one hand, and the unbiasedness test, on the other.

We conclude that economic forecasters are far more successful than financial market analysts. This sharp contrast, however, is levelled and obscured by the unbiasedness test. For this reason, we consider the unbiasedness test an unsuitable tool for the assessment of economic forecasts.

5. Conclusion

The present study addresses the assessment of economic forecasts. In addition to the elaboration of conventional benchmarks, two new evaluation methods are introduced: the information growth test and the TOTA coefficient.

The information growth test determines whether forecasts become more accurate with the shortening of forecast horizons. It is analysed whether forecasters are willing and in a position to absorb new information in the course of the forecast horizon, and to take it into account when generating forecasts. The TOTA coefficient identifies whether forecasts tend to reflect the future or the present.

We investigate consensus forecasts on the development of GDP, industrial production and private consumer spending in the USA, Japan, Germany, France, the UK, Italy, Spain, Canada, the Netherlands, Switzerland, Sweden and Norway.

Research is based on four selected assessment methods: 1. comparison with naive forecasts using the mean absolute error relative to naive forecasts (*MAERNF*), 2. the information growth test, 3. the TOTA coefficient and 4. the unbiasedness test.

The mean absolute error relative to naive forecasts (*MAERNF*) shows that economic forecasts in 136 of the 144 cases observed (94.4%) are more accurate than the corresponding naive forecasts (Table 2).

The information growth test also leads to a fairly positive evaluation of economic forecasters. Sixty-eight of seventy-two cases observed (94.4%) display a statistically signifcant reduction of forecast errors with the gradual shortening of the forecast horizon (Table 3). In the overwhelming majority of cases, therefore, economic experts are willing and in a position to take relevant events into account in their forecasts as forecast horizons gradually diminish.

Analysis of economic forecasts using the TOTA coefficient likewise leads to a favourable assessment. Sixty-two of the seventy-two forecast time series observed (86.1%) show no evidence of topically oriented trend adjustment (Table 3). Hence an inordinate number of forecasts reflect future trends rather than those of the present.

The following requirements are simultaneously fulfilled in thirty-three of the thirty-six forecast items (91.7%): (a) naive forecasts are surpassed when the forecast horizon is nine months; (b) a close look at the absolute error shows a statistically significant information growth; (c) forecasts with a nine-month horizon display no indication of topically oriented trend adjustment. This outcome is evidence of the sophisticated forecast abilities of economic experts.

The unbiasedness test, on the other hand, identifies the forecasts under consideration as failures without exception (Tables 4 and 5). We argue that the discrepancy between the unbiasedness test and the three other benchmarks is due to the unrealistic demands the unbiasedness test makes on forecasters.

Economic forecasts are far more accurate than financial market forecasts. This distinct edge is levelled and obscured when the unbiasedness test is applied. We therefore advocate that this test be given less attention in the future than has been the case up to now.

References

Ackert, L. F. and Hunter, W. C. (1994), Rational Expectations and the Dynamic Adjustment of Security Analysts' Forecasts to New Information, in: *The Journal of Financial Research*, Vol. 17, pp. 387-401.

Ager, P., Kappler, M. and Osterloh, S. (2007), The Accuracy and Efficiency of the Consensus Forecasts: A Further Application and Extension of the Pooled Approach, Centre for European Economic Research Discussion Paper, No. 07-058, Mannheim.

Andolfatto, D., Hendry S. and Moran, K. (2008), Are inflation expectations rational?, in: *Journal of Monetary Economics*, Vol. 55, pp. 406-422.

Andres, P. and Spiwoks, M. (1999), Forecast accuracy matrix, A Methodological Survey of Judging Forecast Quality of Capital Market Forecasts, in: *Journal of Economics and Statistics*, Vol. 219, pp. 513-542.

Ball, R. J. (1962), The Prediction of Wage-Rate Changes in the United Kingdom Economy 1957-60, in: *The Economic Journal*, Vol. 72, pp. 27-44.

Batchelor, R. (2007), Forecaster Behaviour and Bias in Macroeconomic Forecasts, ifo Working Papers, No. 39, ifo Institute of Economic Research at the University of Munich, Munich.

Bofinger, P. and Schmidt, R. (2003), On the reliability of professional exchange rate forecasts: An empirical analysis for the €/US-$ rate, in: *Financial Markets and Portfolio Management*, Vol. 17, pp. 437-449.

Bonham, C. S. and Cohen, R. H. (2001), To Aggregate, Pool, or Neither: Testing the Rational-Expectations Hypothesis Using Survey Data, in: *Journal of Business & Economic Statistics*, Vol. 19, pp. 278-291.

Bowles, C., Friz, R., Genre, V., Kenny, G., Meyler, A. and Rautanen, T. (2007), The ECB Survey of Professional Forecasters (SPF) – A Review After Eight Years' Experience, European Central Bank Occasional Paper Series, No. 59, April 2007, Frankfurt am Main.

Breusch, T. S. and Pagan, A. R. (1979), Simple test for heteroscedasticity and random coefficient variation, in: *Econometrica*, Vol. 47, pp.1287–1294.

Brooks, R. and Gray, B. (2004), History of forecasters: Naive forecasts are better than the consensus WSJ forecast, in: *Journal of Portfolio Management*, Vol. 31, pp. 113-117.

Chinn, M. and Frankel, J. (1994), Patterns in Exchange Rate Forecasts for Twenty-five Currencies, in: *Journal of Money, Credit, and Banking*, Vol. 26, pp. 759-770.

De Gooijer, J. G. and Hyndman, R. J. (2006), 25 years of time series forecasting, in: *International Journal of Forecasting*, Vol. 22, pp. 443-473.

Diebold, F. X. (2007), Elements of Forecasting, 4th International Edition, Thomson Learning.

Diebold, F. X. and Lopez, J. A. (1996), Forecast Evaluation and Combination, in: Maddala, G. S. and Rao, C. R. (eds), Handbook of Statistics. 14: pp. 241-268, Amsterdam.

Diebold, F. X. and Mariano, R. S. (1995), Comparing Predictive Accuracy, in: *Journal of Business and Economic Statistics*, Vol. 13, pp. 253-263.

Döpke, J. and Fritsche, U. (2006), Growth and inflation forecasts for Germany – A panel-based assessment of accuracy and efficiency, in: *Empirical Economics*, Vol. 31, pp. 777-798.

Dovern, J., Fritsche, U. and Slacalek, J. (2009), Disagreement Among Forecasters in G7 Countries, European Central Bank Working Paper Series, No. 1082, Frankfurt am Main.

Dovern, J. and Weisser, J. (2009), Accuracy, Unbiasedness and Efficiency of Professional Macroeconomic Forecasts: An Empirical Comparison for the G7, Jena Economic Research Papers, No. 2009-091, Friedrich Schiller University, Jena.

Fair, R. C. and Shiller, R. J. (1990), Comparing Information in Forecasts from Econometric Models, in: *The American Economic Review*, Vol. 80, pp. 375-389.

Fildes, R. and Stekler, H. (2002), The state of macroeconomic forecasting, in: *Journal of Macroeconomics*, Vol. 24, pp. 435-468.

Fraser, P. and MacDonald, R. (1993), The Efficiency of CAC Stock Price Forecasts, in: *Revue économique*, Vol. 44, pp. 991-1000.

Harvey, D. I., Leybourne, S. J. and Newbold, P. (1998), Tests for Forecast Encompassing, in: *Journal of Business & Economic Statistics*, Vol. 16, pp. 254-259.

Henriksson, R. D. and Merton, R. C. (1981), On Market Timing and Investment Performance, II. Statistical Procedures for Evaluating Forecasting Skills, in: *Journal of Business*, Vol. 54, pp. 513-533.

Joutz, F. and Stekler, H. O. (2000), An evaluation of the predictions of the Federal Reserve, in: *International Journal of Forecasting*, Vol. 16, pp. 17-38.

Keane, M. P. and Runkle, D. E. (1990), Testing the Rationality of Price Forecasts: New Evidence from Panel Data, in: *The American Economic Review*, Vol. 80, pp. 714-735.

Lakonishok, J. (1980), Stock Market Return Expectations: Some General Properties, in : *The Journal of Finance*, Vol. 35, pp. 921-931.

Laplace, P. S. (1814), Essai philosophique sur les probabilités, Paris.

Mankiw, N. G., Reis, R. and Wolfers, J. (2003), Disagreement About Inflation Expectations, in: *NBER Macroeconomics Annual*, Vol. 18, pp. 209-248.

Manzur, M. (1988), How Much Are Exchange Rate Forecasts Worth? in: *Australian Journal of Management*, Vol. 13, pp. 93-113.

Mathews, B. P. and Diamentopoulos, A. (1994), Towards a Taxonomy of Forecast Error Measures, A Factor-comparative Investigation of Forecast Error Dimensions, in: *Journal of Forecasting*, Vol. 13, pp. 409-416.

Mehra, Y. P. (2002), Survey Measures of Expected Inflation: Revisiting the Issues of Predictive Content and Rationality, in: *Federal Reserve Bank of Richmond Economic Quarterly*, Vol. 88/3, pp. 17-36.

Merton, R. C. (1981), On Market Timing and Investment Performance, I. An Equilibrium Theory of Value for Market Forecasts, in: *Journal of Business*, Vol. 54, pp. 363-406.

Mincer, J. and Zarnowitz, V. (1969), The Evaluation of Economic Forecasts, in: Mincer, J. (Ed.), Economic Forecasts and Expectation, New York, Columbia University Press, pp. 3-46.

Mose, J. S. (2005), Expert Forecasts of Bond Yields and Exchange Rates, in: *Danmarks Nationalbank Monetary Review*, Vol. 4th Quarter 2005, pp. 91-95.

Muth, J. F. (1961), Rational Expectations and the Theory of Price Movements, in: *Econometrica*, Vol. 29, pp. 315-334.

Osterloh, S. (2008), Accuracy and Properties of German Business Cycle Forecasts, *in: Applied Economics Quarterly*, Vol. 54, pp. 27-57.

Poulizac, D., Weale, M. and Young, G. (1996), The Performance of National Institute Economic Forecasts, in: *National Institute Economic Review*, No. 156, pp. 55-62.

Schuh, S. (2001), An Evaluation of Recent Macroeconomic Forecast Errors, in: *New England Economic Review*, Vol. Jan./Feb. 2001, pp. 35-56.

Simon, D. P. (1989), The Rationality of Federal Funds Rate Expectations: Evidence from a Survey, in: *Journal of Money, Credit, and Banking*, Vol. 21, pp. 388-393.

Spiwoks, M. (2004), The Usefulness of ZEW Stock Market Forecasts for Active Portfolio Management Strategies, in: *Journal of Economics and Statistics*, Vol. 224, 2004, pp. 557-578.

Spiwoks, M. (2009), Deceptive Unity: Surprising Characteristics of Pre-Tax Corporate Profit Forecasts, in: *International Research Journal of Finance and Economics*, Vol. 26, pp. 170-177.

Spiwoks, M., Bedke, N. and Hein, O. (2008), Forecasting the past: The case of US interest rate forecasts, in: *Financial Markets and Portfolio Management*, Vol. 22, pp. 357-379.

Spiwoks, M., Bedke, N. und Hein, O. (2009), The Pessimism of Swiss Bond Market Analysts and the Limits of the Sign Accuracy Test – An Empirical Investigation of Their Forecasting Success Between 1998 and 2007, in: *International Bulletin of Business Administration*, Vol. 4, pp. 6-19.

Spiwoks, M., Bedke, N. and Hein, O. (2010), Topically Orientated Trend Adjustment and Autocorrelation of the Residuals - An Empirical Investigation of the Forecasting Behavior of Bond Market Analysts in Germany, in: *Journal of Money, Investment, and Banking*, Vol. 14, pp. 16-35.

Spiwoks, M., Gubaydullina, Z. and Hein, O. (2011), Trapped in the Here and Now – New Insights into Financial Market Analyst Behavior, Wolfsburg Working Papers, No. 11-01, Ostfalia University of Applied Sciences, Wolfsburg.

Storbeck, O. (2010), Wie die Finanzkrise die VWL auf den Kopf stellt, in: *Handelsblatt*, January 11th 2010, pp. 20-21.

Theil, H. (1971), Principles of Econometrics, New York.