Working Papers of the Priority Programme 1859
Experience and Expectation.
Historical Foundations of Economic Behaviour
Edited by Alexander Nützenadel und Jochen Streb

No 22 (2020, September)

Foltas, Alexander / Pierdzioch, Christian

Business-Cycle Reports and the Efficiency of Macroeconomic Forecasts for Germany

Arbeitspapiere des Schwerpunktprogramms 1859 der Deutschen Forschungsgemeinschaft
„Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns“ /
Working Papers of the German Research Foundation's Priority Programme 1859
"Experience and Expectation. Historical Foundations of Economic Behaviour"

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Published in co-operation with the documentation and publication service of the Humboldt University, Berlin (https://edoc.hu-berlin.de).

ISSN: 2510-053X

Redaktion: Alexander Nützenadel, Jochen Streb, Ingo Köhler

V.i.S.d.P.: Alexander Nützenadel, Jochen Streb

SPP 1859 "Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns"
Sitz der Geschäftsführung:
Humboldt-Universität
Friedrichstr. 191-193, 10117 Berlin
Tel: 0049-30-2093-70615, Fax: 0049-30-2093-70644
Web: https://www.experience-expectation.de
Koordinatoren: Alexander Nützenadel, Jochen Streb
Assistent der Koordinatoren: Ingo Köhler

Recommended citation:

Foltas, Alexander / Pierdzioch, Christian (2020): *Business-Cycle Reports and the Efficiency of Macroeconomic Forecasts for Germany*. Working Papers of the Priority Programme 1859 "Experience and Expectation. Historical Foundations of Economic Behaviour" No 22 (September), Berlin

© 2020 DFG-Schwerpunktprogramm 1859 „Erfahrung und Erwartung. Historische Grundlagen ökonomischen Handelns“

The opinions and conclusions set forth in the Working Papers of the Priority Programme 1859 *Experience and Expectation. Historical Foundations of Economic Behaviour* are those of the authors. Reprints and any other use for publication that goes beyond the usual quotations and references in academic research and teaching require the explicit approval of the editors and must state the authors and original source.

Business-Cycle Reports and the Efficiency of Macroeconomic Forecasts for Germany

Alexander Foltas[a*] and Christian Pierdzioch[a]

September 2020

Abstract

We study the efficiency of growth and inflation forecasts published by three leading German economic research institutes during a period of time ranging from 1970 to 2017. To this end, we examine whether the information used by the research institutes when they formed their forecasts helps to explain the ex-post realized forecast errors. We identify the information that the research institutes used to set up their quantitative forecasts by applying computational-linguistics techniques to decompose the business-cycle reports published by the research institutes into various topics. Our results show that several topics have predictive value for the forecast errors.

JEL classification: C53; E32; E37

Keywords: Growth forecasts; Inflation forecasts; Forecast efficiency; Business-cycle reports

Address:

[a] Department of Economics, Helmut Schmidt University, Holstenhofweg 85, P.O.B. 700822, 22008 Hamburg, Germany

[*] Corresponding author. E-mail: foltasa@hsu-hh.de.

Funding:
This research was supported by the German Science Foundation (Project: Exploring the experience-expectation nexus in macroeconomic forecasting using computational text analysis and machine learning; Project number: 275693836).

Declarations of interest: None.

1 Introduction

A classic topic in the extensive literature on business-cycle forecasting is whether macroeconomic forecasts are efficient. Forecast efficiency requires that the ex-post realized forecast error cannot be predicted by means of information that were available to a forecaster when a forecast was being published. Hence, once a researcher has decided on how to model the information set of a forecaster, it is straightforward to test forecast efficiency by estimating a regression equation that links the forecast error to the variables that a researcher thinks are good proxies of the historical information that were available to a forecaster (Mincer and Zarnowitz, 1969).

The hard part of this exercise is to model a forecaster's information set. A common approach is to use standard macroeconomic and financial variables to proxy a forecaster's information set. The array of macroeconomic and financial variables that a researcher can use to this end, however, is potentially large (see, for example Behrens, Pierdzioch, and Risse, 2018a), and researchers using different variables may draw different conclusions regarding the efficiency of the same series of forecasts (Stekler (2004)). For this reason, we depart from the common approach and use textual data to describe a forecaster's information set.

To be more specific, we study the efficiency of growth and inflation forecasts published by three leading German economic research institutes. The research institutes regularly publish business-cycle reports that accompany their quantitative macroeconomic forecasts. The business-cycle reports contain detailed information as to how the research institutes assess several important facets of macroeconomic and policy developments. This information comes in the form of textual data. For this reason, we must quantity the economic information embedded in the business-cycle reports before we can use this information to test forecast efficiency.

To this end, we apply techniques developed in the computational-linguistics literature that render it possible to extract "topics" from the business-cycle reports (Foltas, 2020). The topics represent various key aspects of macroeconomic and financial developments. For example, there is an investment topic, a labor-market topic and a topic that represents taxation and social security

issues. The relative importance of the topics transforms the textual data into quantitative data that we can then use to represent the information set of the research institutes at the time they published their macroeconomic forecasts.

In addition to presenting a novel way of testing for forecast efficiency, our research contributes to recent research on macroeconomic forecasts for Germany. For example, Behrens, Pierdzioch, and Risse (2018a) analyze the multivariate efficiency of growth and inflation forecasts for Germany, while Behrens, Pierdzioch, and Risse (2018b) study forecast efficiency under flexible loss. Heilemann and Stekler (2013) study the time-varying accuracy of forecasts, Döpke and Fritsche (2006) use panel-data methods to study forecast efficiency, and Kirchgässner and Müller (2006) shed light on costly forecast revisions.

2 Modeling Framework

Assuming that the loss function is of the classic mean-squared error form, a classic forecast-efficiency regression equation is of the following format (Holden and Peel, 1990):

$$fe_{t+h} = \alpha + \gamma x_t + \varepsilon_t, \tag{1}$$

where fe_{t+h} denotes the forecast error (in our empirical analysis defined as forecast minus actual value) at forecast horizon h, x_t denotes a variable that represents a forecaster's information set at the time a forecast was formed, and ε_t denotes a disturbance term.

An intercept coefficient, α, that is significantly different from zero implies that forecasts are biased. A slope coefficient, γ, that systematically differ from zero implies that a variable, x_t, contains information about the subsequently realized forecast error. In case the right-hand side variable, x_t, represents the lagged forecast error and we have $\alpha = \gamma = 0$, then we cannot reject the hypothesis that forecasts weakly efficient. In case the right-hand side variable, x_t, represents any other variable a forecaster knew when a forecast was being formed (but not a topic), then fore-

casts are said to be strongly efficient (see Timmermann (2007)).[1] We call this form of forecast efficiency the classic form of strong forecast efficiency.

A problem with Equation (1) is that the right-hand-side variable, x_t, can represent a potentially large number of macroeconomic and financial variables, and it is unlikely that researchers fully agree on which variables to include in a classic forecast-efficiency regression equation.[2] It is, therefore, not surprising that no consensus has emerged in the forecasting literature as to which variables should be used to test forecast efficiency. For this reason, we propose a modified forecast-efficiency regression equation. Our modified forecast-efficiency regression equation accounts for the information contained in the business-cycle reports of the research institutes and, thus, approximates their information set by using data published by the research institutes rather than extraneous macroeconomic or financial data published by statistical agencies. Our modified forecast-efficiency regression equation is given by

$$fe_{t+h} = \alpha + \beta T_t + \varepsilon_t, \tag{2}$$

where T_t denotes a topic extracted from the business-cycle reports of the research institutes. We use Equation (2) to test what we call the textual form of strong forecast efficiency. As in the case of the classic form of strong forecast efficiency, the textual form of strong forecast efficiency requires $\alpha = \beta = 0$, a hypothesis that can be tested by means of an F-test.

One can further study whether forecasts satisfy the requirement of "global" strong forecast efficiency by including x_t as another right-hand-side variable of Equation (2) and testing the hypothesis $\alpha = \beta = \gamma = 0$. If so, we cannot reject the hypothesis that forecasts are unbiased and both the classic and the textual form of strong forecast efficiency apply.

An assumption underlying the forecast-efficiency regression equations given in Equations (1) and (2) is that the loss function is of the mean-squared error type. In order to account for the

[1] Some researchers use the terms weak and strong forecast rationality rather than weak and strong forecast forecast-efficiency. See Stekler (2004).

[2] Another problem is that, when a researcher uses macroeconomic variables to represent the right-hand side variable, x_t, it is important to account for ex-post data revisions and possibly time-varying publication lags.

possibility that the loss function is not of this specific type, we use the quantile test proposed by Patton and Timmermann (2007). We implement the quantile test by estimating the following regression equation:

$$I_{t+h} = \alpha + \beta T_t + \varepsilon_t, \tag{3}$$

where $I_{t+h} = 1$ if $fe_{t+h} \leq 0$ and $I_{t+h} = 0$ otherwise. Equation (3) can be estimated as a linear probability model or by means of a logit model, where the latter better captures the dichotomous nature of the left-hand-side variable. Textual forecast efficiency requires that the hypothesis $\beta = 0$ cannot be rejected.

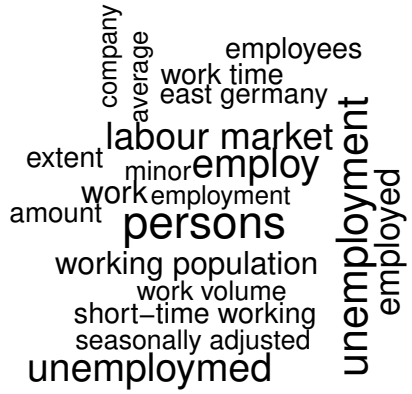
3 The Data

We study the annual growth forecasts of three leading German economic research institutes for the sample period 1970−2017.[3] While the publication frequency of forecasts differs across the research institutes and also varies over time, the most common forecasts are one-year-ahead (denoted q_4 forecasts) annual forecasts that the institute4s publish at the turn of the year, and six-month-ahead (denoted q_2 forecasts) annual forecasts, which are published mid-year. We subtract the realized growth and inflation rate (measured using first-release data) from the forecasts in order to compute the forecast errors, where we take into account for each institute the impact of German reunification (for further details, see Behrens, Pierdzioch, and Risse, 2018a). In order to quantify the information embedded in the business-cycle reports that the research institutes publish, we process the business-cycle reports with a combination of word embeddings and the latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan, 2003) topic model.[4] In this way, we

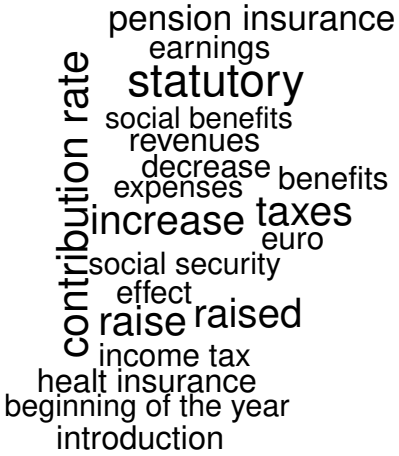
[3]The research institutes are (in alphabetical order): Deutsches Institut für Wirtschaftsforschung, Ifo Institut, and Institut für Weltwirtschaft.

[4]The basic idea is to map each word of the business-cycle reports via the skip-gram method in vector space and allows for a mathematical representation of the word sense. An LDA algorithm then extracts topics out of the obtained co-occurrence matrix. For an extensive explanation of this approach, see Panigrahi, Simhadri, and Bhattacharyya (2019) and Foltas (2020).

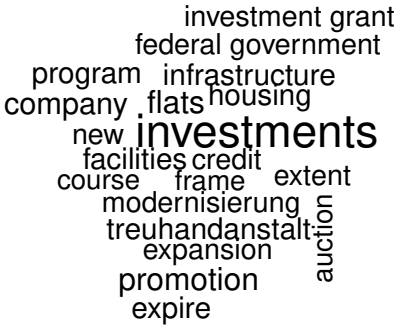
Figure 1: Wordclouds



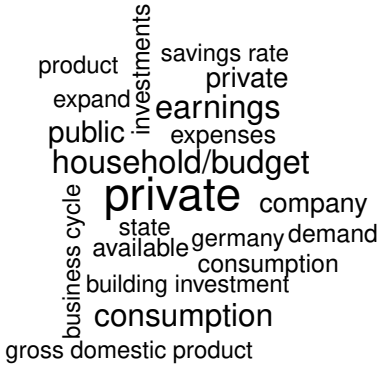
(a) Labour-market topic



(b) Taxation/ social-security topic



(c) Investment topic



(d) Public/private-expenses topic

obtain 18 topics that are related to specific economic subjects[5]. The topic proportions inform about the share of particular economic themes in the business-cycle reports and, thereby, represent the research institutes' information set. We plot four word clouds containing the twenty most prevalent words of the respective topics in Figure 1: one topic captures information regarding the labour market, one topic represents taxes and social-security issues, then there is one of two investment topics (in this case, there is a particular emphasis on the connection between investments and infrastructure), and finally, we plot a topic that combines discussions regarding public and private expenses. For an overview of all topics, we refer the reader to Foltas (2020).

4 Empirical Results

Table 1: Descriptive

Institute	Forecast error	Mean	N	SD	t-test	AR(1)	F-test
Institute I	GDP_{q2}	-0.18	37.00	0.89	0.23	0.03	0.55
	GDP_{q4}	0.05	48.00	1.45	0.82	0.01	0.74
	CPI_{q2}	-0.00	36.00	0.39	0.97	0.00	0.21
	CPI_{q4}	-0.10	40.00	0.74	0.40	0.15	0.01
Institute II	GDP_{q2}	-0.07	42.00	0.73	0.51	0.04	0.22
	GDP_{q4}	0.13	45.00	1.25	0.50	-0.26	0.41
	CPI_{q2}	0.01	39.00	0.41	0.88	0.09	0.44
	CPI_{q4}	0.07	41.00	0.57	0.44	0.01	0.20
Institute III	GDP_{q2}	-0.12	38.00	0.86	0.38	-0.13	0.04
	GDP_{q4}	-0.05	42.00	1.09	0.79	0.00	0.14
	CPI_{q2}	0.01	38.00	0.45	0.94	0.42*	0.00
	CPI_{q4}	-0.00	43.00	0.78	0.98	0.11	0.11

Note: N denotes the number of observations. SD denotes the standard deviation of forecast errors. The t-test tests the hypothesis that the mean forecast error is zero. AR(1) tests the hypothesis that the coefficient of first-order autocorrelation is zero. The F-test tests the hypothesis of classic strong forecast efficiency, where the control variables are a short term interest rate, the returns of the oil price (West Texas Intermediate), the returns of the real effective exchange rate, and the growth rate of industrial production.

[5]Our model extracts 24 topics, though six are not used in our analysis, as they describe methodical approaches or consist of self-references and words without economic meaning.

Table 1 summarizes some descriptive statistics of the data and depicts the results of tests of unbiasedness of forecasts, weak forecast efficiency, and the classic form of strong forecast efficiency. Several results emerge. We have more data on the q_4 than on the q_2 forecasts, and, as expected, the standard deviation of the q_4 forecast errors is larger than for the q_2 forecast errors. The mean forecast error for the CPI_{q2} forecasts is close to zero. The tests for unbiasedness of forecasts (two sided t-test) and first-order autocorrelation of forecast errors (based on the Pearson correlation coefficient) are not significant. Hence, we cannot reject the hypothesis of weakly efficient forecasts. Further, we cannot reject the hypothesis of the classic form of forecast efficiency for Institutes I and II.[6] For Institute III, we reject the classic form of forecast efficiency at the 5% level of significance for the short-term GDP_{q2} and CPI_{q2} forecasts. Table 2 summarizes

Table 2: Textual form of strong forecast efficiency

Forecasts	Institute	Topic	Intercept	Topic	F-test
			p-values		
GDP_{q2}	Institute II	monetary policy	0.25	0.02	0.02
GDP_{q2}	Institute II	investments	0.01	0.00	0.00
GDP_{q2}	Institute III	investments	0.22	0.03	0.03
CPI_{q2}	Institute II	government spending	0.07	0.03	0.03
CPI_{q4}	Institute I	current account	0.52	0.05	0.05
CPI_{q4}	Institute I	public/private-expenses	0.01	0.01	0.01

Note: This table summarizes results for those cases in which a topic has significant predictive power at least at the 5% level of significance. The F-test tests the hypothesis of textual strong forecast efficiency, while a t-test tests the significance of the intercept and the topic coefficient.

the results of tests for the textual version of strong forecast efficiency, where we focus on those models that feature an estimated coefficient of a topic is significantly that is different from zero at a marginal significance level of at least 5%. We reject the textual version of strong forecast

[6]We approximate the research institutes information sets in terms of a short term interest rate, the returns of the oil price (West Texas Intermediate), the returns of the real effective exchange rate, and the growth rate of industrial production (see Döpke and Fritsche, 2006). Like Behrens, Pierdzioch, and Risse (2018a,b), we take into account a forecast formation lag (that is, we assume that the research institutes use macroeconomic data for the month preceding the month in which a forecast is formed) and publication lags.

efficiency for the GDP_{q2} forecasts of Institutes II and III, where we find that the monetary policy and investment topics have significant explanatory power for the forecast error. As for the CPI_{q2} forecasts, only the forecasts of Institute II violate the criterion of strong textual forecast efficiency, with government spending being a significant topic. For the CPI_{q4} forecasts, we reject strong textual forecast efficiency for Institute I, where the current-account and household-income topics have predictive power for the forecast error. Only two of the estimated intercept terms are significant at the 5% level, providing weak evidence against weak forecast efficiency. In contrast, all F-tests yield significant results, implying that we can reject the strong textual form of forecast efficiency for all but the GDP_{q4} forecasts.

Table 3: Global strong forecast efficiency

Series	Institute	Topic	Intercept	Topic	F-Test
			p-values		
GDP_{q2}	Institute I	investments	0.08	0.04	0.12
GDP_{q2}	Institute I	investments	0.14	0.03	0.12
GDP_{q2}	Institute I	public/private-expenses	0.07	0.04	0.13
GDP_{q2}	Institute II	monetary policy	0.65	0.03	0.03
GDP_{q2}	Institute II	investments)	0.04	0.00	0.00
GDP_{q4}	Institute III	employment	0.13	0.04	0.03
CPI_{q2}	Institute II	government spending	0.07	0.04	0.11
CPI_{q2}	Institute II	wages	0.10	0.02	0.08
CPI_{q2}	Institute III	GDP	0.00	0.01	0.00
CPI_{q2}	Institute III	taxes/social insurances	0.00	0.01	0.00
CPI_{q4}	Institute I	public/private-expenses	0.10	0.01	0.01
CPI_{q4}	Institute III	recession	0.02	0.04	0.03
CPI_{q4}	Institute III	GDP	0.01	0.02	0.02

Note: This table summarizes results for those cases in which a topic has significant predictive power at least at the 5% level of significance. The F-test tests the hypothesis of global strong forecast efficiency, while a t-test tests the significance of the intercept and the topic coefficient. The control variables are a short term interest rate, the returns of the oil price (West Texas Intermediate), the returns of the real effective exchange rate, and the growth rate of industrial production.

Table 3 summarizes the results of a test for global strong forecast efficiency. Again, we exclusively present results for those cases in which we find that a topic has significant predictive power

at least at the 5% level of significance. We reject global strong forecast efficiency for all four categories of forecasts, but there are differences across the research institutes. For Institute I, we reject the textual form of strong forecast efficiency of the GDP_{q2} forecasts (the investment and household-income topics are significant) but not the global form of strong forecast efficiency, where the evidence against unbiasedness of the forecasts is weak. For Institute II, in contrast, we reject global strong forecast efficiency of the GDP_{q2} forecasts (the monetary-policy and the investment topics are significant). For the GDP_{q4} forecasts, we reject the textual but not the global form of strong efficiency in case of Institute I (the statistical topic is significant), and we reject textual (the employment topic is significant) and global strong forecast efficiency for Institute III (but the forecasts of this institute are unbiased). Next, we reject the textual but not the strong form of strong forecast efficiency of the CPI_{q2} forecasts of Institute II (the government-spending and wages topics are significant). The CPI_{q2} forecasts of Institute III, in turn, violate both the textual and the strong form of forecast efficiency (and forecast unbiasedness), where the GDP and taxes/social insurance topics are significant. We reject the textual and global forecast efficiency of the CPI_{q4} forecasts of Institutes I and III (the household-income, recession, GDP topics are significant). Finally, we report estimates of Equation (3) in Table 4.[7] We reject textual forecast efficiency of the GDP_{q2} forecasts of Institute I and Institute III, where the current-account topic has significant predictive value for the forecast error in case of the former and the inflation topic is significant in case of the latter. As for the CPI_{q2} forecasts, we find evidence against textual forecast efficiency for all three research institutes. While for Institute I, the sectoral and GDP topics are significant, the wage and taxes/social-insurance topics have predictive value for the forecast errors made by Institute II and Institute III, respectively. Finally, we reject the textual form of efficiency of the CPI_{q4} forecasts of Institute III (the government-spending and employment topics are significant).

[7] Estimating Equation (3) as a logit model yields qualitatively similar results (available upon request).

Table 4: Results for a flexible loss function

Series	Institute	Topic	Intercept	Topic	F-Test
			p-values		
GDP_{q2}	Institute I	current account	0.00	0.03	0.03
GDP_{q2}	Institute III	inflation	0.00	0.04	0.04
CPI_{q2}	Institute I	sectoral	0.00	0.00	0.00
CPI_{q2}	Institute I	GDP	0.00	0.02	0.02
CPI_{q2}	Institute II	wages	0.00	0.03	0.03
CPI_{q2}	Institute III	taxes/social insurances	0.06	0.03	0.03
CPI_{q4}	Institute III	government spending	0.00	0.04	0.04
CPI_{q4}	Institute III	employment	0.01	0.01	0.01

Note: This table summarizes results for those cases in which a topic has significant predictive power at least at the 5% level of significance. The F-test tests the hypothesis of textual strong forecast efficiency, while a t-test tests the significance of the intercept and the topic coefficient. The dependent variable is defined as $I_{t+h} = 1$ if $fe_{t+h} \leq 0$ and $I_{t+h} = 0$.

5 Concluding Remarks

We have used topics extracted by means of computatinal-linguistics techniques from the business-cycle reports published by three leading German research institutes to study the efficiency of macroeconomic forecasts for Germany. To this end, we have proposed an extension of the classic forecast-efficiency regression model in a way that makes it possible to test for unbiasedness, weak efficiency, the classic form of strong forecast efficiency, the textual form of strong forecast efficiency, and global strong forecast efficiency. We have found evidence against the textual and the global form of forecast efficiency, where we have documented differences between the (short-term and longer-term) growth and inflation forecasts and between the research institutes. In future research, it is interesting to apply our approach to testing for forecast efficiency to study the efficiency of other research institutes and international organizations.

References

Behrens, C. / *Pierdzioch,* C. / *Risse,* M. (2018a): "A test of the joint efficiency of macroeconomic forecasts using multivariate random forests", Journal of Forecasting 37(5), 560–572.

– (2018b): "Testing the optimality of inflation forecasts Under flexible Loss with random forests", Economic Modelling 72, 270–277.

Blei, D. M. / *Ng,* A. Y. / *Jordan,* M. I. (2003): "Latent Dirichlet Allocation", Journal of Machine Learning Research 3, 993–1022.

Döpke, J. / *Fritsche,* U. (2006): "Growth and inflation forecasts for Germany - A panel-based assessment of accuracy and efficiency", Empirical Economics 31(3), 777–798.

Foltas, A. (2020): "Testing investment forecast efficiency with textual data", Working Papers of the Priority Programme 1859 "Experience and Expectation. Historical Foundations of Economic Behaviour" (19), `https://edoc.hu-berlin.de/handle/18452/22538`.

Heilemann, U. / *Stekler,* H. O. (2013): "Has the accuracy of German macroeconomic forecasts improved?", German Economic Review 14(2), 235–253.

Holden, K. / *Peel,* D. A. (1990): "On testing for unbiasedness and efficiency of forecasts", The Manchester School 58(2), 120–127.

Kirchgässner, G. / *Müller,* U. K. (2006): "Are forecasters reluctant to revise their predictions? Some German evidence", Journal of Forecasting 25(6), 401–423.

Mincer, J. A. / *Zarnowitz,* V. (1969): "The evaluation of economic forecasts", in: Economic forecasts and expectations: Analysis of forecasting behavior and performance, ed. by Jacob A. Mincer, 3–46, National Bureau of Economic Research, New York.

Panigrahi, A. / *Simhadri,* H. V. / *Bhattacharyya,* C. (2019): "Word2Sense: Sparse Interpretable Word Embeddings", Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 5692–5705.

Patton, A. J. / *Timmermann,* A. (2007): "Testing forecasting optimality under unknown loss", Journal of the American Statistical Association 102(480), 1172–1184.

Stekler, H. O. (2004): "The rationality and Efficiency of individuals' forecasts: 10", in: A Companion to Economic Forecasting, ed. by Clements, M. P. / Hendry, D. H., Blackwell Publishing Ltd.

Timmermann, A. (2007): "An evaluation of the World Economic Outlook forecasts", IMF Staff Papers 54(1), 1–33.