Journal of Forecasting, *J. Forecast.* (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/for.2449

Do Media Data Help to Predict German Industrial Production?

DIRK ULBRICHT,[1]* KONSTANTIN A. KHOLODILIN[2] AND TOBIAS THOMAS[3]

[1] *IFF Hamburg, Germany*
[2] *DIW Berlin, Germany*
[3] *DICE, Düsseldorf, Germany*

ABSTRACT

In an uncertain world, decisions by market participants are based on expectations. Therefore, sentiment indicators reflecting expectations have a proven track record at predicting economic variables. However, survey respondents largely perceive the world through media reports. Here, we want to make use of that. We employ a rich dataset provided by Media Tenor International, based on sentiment analysis of opinion-leading media in Germany from 2001 to 2014, transformed into several monthly indices. German industrial production is predicted in a real-time out-of-sample forecasting experiment and media indices are compared to a huge set of alternative indicators. Media data turn out to be valuable for 10- to 12-month horizon forecasts, which is in line with the lag between monetary policy announcements and their effect on industrial production. This holds in the period during and after the Great Recession when many models fail. Copyright © 2016 John Wiley & Sons, Ltd.

KEY WORDS media data; German industrial production; forecast breakdown; real-time experiment; model confidence set

INTRODUCTION

Typically, data on gross domestic product (GDP) are available on a quarterly basis. In addition, they are published half a quarter following the end of the reference quarter. Therefore, in order to gain quick insight into the current economic situation, a monthly series of industrial production is used. It is considered to be a key monthly indicator for business activity. This is especially true in the case of Germany. Although the share of industrial production has been shrinking since the 1980s, it remains relatively high when compared to other OECD and, especially, other EU member countries.[1] Furthermore, the European Commission plans to raise the contribution of industry to GDP to as much as 20% by 2020 (European Commission, 2014) in order to increase the competitiveness of the EU. Moreover, industrial production contributes substantially to the business cycle dynamics.

Consequently, there have been many attempts to improve the forecast accuracy of this variable.[2] Most of these studies employ 'hard economic' indicators, such as interest rates or manufacturing orders. There are also several studies using 'soft data', such as business surveys, including the ifo and ZEW indicators (see, for example, Abberger and Wohlrabe, 2006, or Hüfner and Schröder, 2002). It is demonstrated that, owing to their forward-looking nature as well as their timely availability, the soft data are well suited for forecasting industrial production. The underlying idea of this approach is to employ a measure of the intentions or the expectations of the managers or analysts, respectively. The main advantages of these indicators are their high frequency, timeliness and the fact that they are subject only to minor revisions,[3] unlike many other statistical indicators.

Alternatively, media data could be used to improve the forecast accuracy. The use of media data for such an analysis in an uncertain environment is rather straightforward. While in classical economics the *Homo oeconomicus* is omniscient and decides independently, with his decisions leading to efficient outcomes at the market level, Keynes (1937) underlines the role of uncertainty concerning decisions and behavior as well as the related (suboptimal) outcomes at the macro level, just as von Hayek (1989) points to the pretense of knowledge. Similarly, both Simon (1957) and Kahneman and Tversky (1979) show that actual human behavior clearly deviates from the behavior predicted by standard economic models. Due to their limited information processing capacity, individuals use subjective models for the perception of reality. If these models are shared because of common cultural background and

*Correspondence to: Dirk Ulbricht, IFF Hamburg, Germany E-mail: dirk.ulbricht@iff-hamburg.de

[1] According to the *OECD Factbook 2011: Economic, Environmental and Social Statistics*, in 2010, the percentage of total value added in industry (including energy) was 24% in Germany, 19% in the EU and 21% in the OECD countries.

[2] See, for example, Kholodilin and Siliverstovs (2006), Robinzonov and Wohlrabe (2010) or Drechsel and Scheufele (2012).

[3] As pointed out by an anonymous referee, surveys undergo revisions but of much smaller magnitude than, for example, the indicators of the national accounts. As a rule, these small revisions are related to the fact that not all firms manage to respond in timely manner to the questionnaires sent to them. Thus incorporation of their responses leads to revisions in survey-based indicators. Another source of revisions is an updating of the seasonal factors, in case a survey indicator needs to be seasonally adjusted.

experience, in accordance with Denzau and North (1994), one can speak of shared mental models. In media societies, media reporting forms relevant parts of those shared mental models not only because investors, consumers, politicians and voters receive lots of information via the media, but because additional information perceived directly is interpreted on the basis of the frame determined by the media reporting. Therefore, what is on the agenda ('agenda setting') and what is not ('agenda cutting') becomes highly relevant, as well as the way in which these things are framed in the media, such as with a positive, negative or neutral tone. At least in part, individuals decide and behave based on the information they receive from the media. This is also important in the context of business surveys, as respondents interpret their own economic situation and build their expectations within the frame set by the media.

A growing literature employs media data to explain economic sentiment. For instance, Goidel and Langley (1995) as well as Doms and Morin (2004) show an impact of media reporting on consumer climate. For Nadeau *et al.* (2000) and Soroka (2006) the assessment of the state of the economy depends at least in part on media reports. In their comprehensive contribution Lamla and Maag (2012) analyze the role of media reporting for inflation forecasts of households and professional forecasters.

The literature can be split into two main streams. The first simply counts the number of times a single word or a group of words that can be associated with a certain event occur in the media. The second strand of literature captures content expressed in the media.

Most economic analyses using media focus on the USA. Using word counts, *The Economist* newspaper introduced the R-word index, which is intended to be a proxy for the US business cycle. It counts how many articles in *The Washington Post* and *The New York Times* use the word 'recession' in a quarter. This simple indicator was expanded by Doms and Morin (2004), who count the number of articles in 30 American newspapers that contain nine keywords or expressions in the title or the first paragraph of an article and use this statistic to forecast US private consumption.

Beyond simple word counts, content analysis focuses on the underlying sentiment expressed in media reports using both automated methods and human analysts to evaluate the news. Tetlock (2007) assesses the sentiment of *Wall Street Journal* articles, while Uhl (2011, 2012) uses sentiment data of newspaper and TV news, provided by Media Tenor International, to forecast US private consumption.

Bordino *et al.* (2012) take advantage of the number of queries of listed companies in the US search engine Yahoo! as a predictor for stock market volumes. Using the number of queries in Google, Kholodilin *et al.* (2010) try to improve forecasts of US private consumption. Bollen *et al.* (2011) employ the OpinionFinder software to analyze Twitter tweeds automatically with the aim of forecasting stock prices.

For Germany, the R-word index was adopted by HypoVereinsbank, which counted the word '*Rezession*' ('recession' in German) in articles published in the *Frankfurter Allgemeine Zeitung*, *Handelsblatt* and *WirtschaftsWoche*; but publication of the index was quickly dropped.[4] Grossarth-Maticek and Mayr (2008) revived the index for their study, but at that time the Great Recession period had not yet started and thus could not be considered. Their study compares two media-based indices—the R-word index for Germany and a Media Tenor International index—by examining their ability to forecast the year-on-year growth rate of the German real GDP and to predict the business cycle turning points. The results are rather mixed for both indices. Other media studies include Iselin and Siliverstovs (2013, 2015), who use the R-word index to make one-step-ahead forecasts of the growth rates of real GDP in Germany and Switzerland. While media indicators are helpful for Switzerland, it is not the case for Germany. Ammann *et al.* (2014) compute the number of mentions of a lexicon of 236 words in the online archive of *Handelsblatt* with the aim of predicting yields of the German stock market DAX index. They show that newspaper content is a valuable predictor of future DAX returns.

Here, we test the predictive content of media indicators performing a pseudo-out-of-sample forecast experiment. Since we want to compare the performance of the media indicators to that of a set of rival indicators that is as complete as possible, we collect and update the database of Drechsel and Scheufele (2012), which set a standard in this respect.

This paper makes five contributions. First, unlike Grossarth-Maticek and Mayr (2008), who use a single aggregate Media Tenor International business conditions index, we employ 16 more indicators that differ in their time perspective (present, future and climate), their underlying topic (government budget, monetary policy, labor market, business cycle and taxation) and the region it reports (Germany or the rest of the world, ROW). Second, in contrast to Grossarth-Maticek and Mayr (2008), we use monthly instead of quarterly data. Third, in contrast to Drechsel and Scheufele (2012) and Grossarth-Maticek and Mayr (2008), we take advantage of real-time data provided by the Deutsche Bundesbank for the industrial production as well as for other crucial variables such as inflation. This is particularly important in this context, as one of the great benefits of media data is that they do not need to be revised ex post. Fourth, we test media-based models against *data snooping* using the model confidence set (MCS) approach (Hansen *et al.*, 2011). Fifth, we separately test the media indicators' relative performance over a period when forecasting becomes particularly difficult. To the best of our knowledge, we are the first to identify such a period using

[4] The German R-word index was reported in the mass media several times in the early 2000s but no instances were identified after 5 July 2001.

the concept of forecast breakdowns introduced by Giacomini and Rossi (2009). Finally, we employ the same concept to compare the reliability of different indicators.

To summarize, we find that media indicators based on news that are related to future events are very useful for predicting 10–12 months ahead. This is in line with the time lag between monetary policy announcements in the media and subsequent changes in industrial production.[5] This is also true during periods when forecasting becomes harder.

This paper is structured as follows. The next section presents the empirical approach, while the third section describes the forecast accuracy tests employed. In the fourth section the data used in the analysis are presented. The forecast accuracy testing results are reported in the fifth section and discussed in the sixth section . The final section draws conclusions.

EMPIRICAL APPROACH

We follow the approach of existing studies that concentrate on the comparison of single models including one different alternative indicator at a time in a horse race with respect to forecast accuracy.

Due to the different release lags of indicators such as macroeconomic or survey indicators, data unbalancedness often emerges at the end of multivariate samples. This phenomenon is sometimes referred to as the ragged edge of the data. In order to account for this we adapt the basic empirical set-up for individual models from Drechsel and Scheufele (2012), which explicitly addresses this issue. The estimation equation for the individual models is given as

$$y_{t+h}^{h} = \alpha + \sum_{p=\underline{p}}^{P} \beta_p y_{t-p} + \sum_{q=\underline{q}}^{Q} \gamma_q x_{t-q} + \epsilon_{t+h}^{h} \tag{1}$$

where y_{t+h}^{h} is the annualized growth rate of industrial production at time t over the next h months, $y_{t+h} = \frac{1200}{h} \times \ln(\mathrm{IP}_t/\mathrm{IP}_{t-h})$. The growth rate of industrial production in levels, IP_t, is defined as $y_t = \Delta \ln \mathrm{IP}_t$; x_t is a candidate predictor; ϵ_{t+h}^{h} is an error term; and α, β, and γ are regression coefficients. The timely availability of indicator variable can be taken account of by $\underline{p}$ and $\underline{q}$. German industrial production is available only with a time lag of about 6 weeks, such that $\underline{p} = 2$ months. The publication lags of the candidate regressors vary from $\underline{q} = 0$, mostly for financial and media data, up to $\underline{q} = 2$ months. The lag length is optimized using the Akaike information criterion. First, P is estimated. Then, holding P fixed, Q is estimated. The maximum lag length tested is 12 months.

Forecasts for 1- to 12-month horizons are computed in a pseudo-out-of-sample real-time set-up:

- The first forecast is made based on the information set as it was available in 2005:12.
- Each iteration, the information set is extended by 1 month; i.e. in the second iteration data are used as they have been available in 2006:01, in the third iteration data available in 2006:02 are used, and so on.
- For each horizon, 171 individual models are estimated and forecasts are made. The models differ as different exogenous variables and different transformations are used (see Table IV).
- Each iteration, the lag lengths are optimized and the model is updated.
- All real indicators are obtained by deflating the nominal ones using consumer price index as it has been available at the point in time the forecast is made.
- The estimations are based on a rolling window of 60 months. Employing rolling windows is a method of dealing with slowly moving non-stationarities (see, for example, Pesaran and Timmermann (2004). Furthermore, it is required for the MCS test (see Hansen *et al.*, 2011).
- The last forecast is based on the information set available in 2014:11 for $h = 1$ and, respectively, 2013:11 for $h = 12$.

After having completed the estimation and forecasting exercise, the forecasts are used to compute the forecast errors based on the data as they have been available 2015:02 (the last-vintage data available at the time of investigation).

FORECAST ACCURACY TESTS

To compare the usefulness of media indicators relative to rival indicators for predicting German industrial production, we need to evaluate the forecast performance of competing models. Thereby, we resort to simple forecast accuracy measures, pairwise and multiple statistical tests, and a test for forecast breakdowns.

[5] The Bank of England, for example, states that 'official interest rate decisions have their fullest effect on output with a lag of around one year'; p. 3 in George *et al.* (1999).

Simple measures

In the following, we concentrate on one of the standard loss functions adopted in the forecasting literature: the squared forecast error, $L_{it} = e_{it}^2$. Let $e_{it} = y_t - \hat{y}_{it}$ be the forecast error of model i in period t, where y_t is the realization of the target variable and $\hat{y}_{it}$ is the point forecast by model i. Usually, the forecast performance of alternative models is compared using the mean squared forecast error (MSFE), defined as

$$\text{MSFE}_i = \frac{\sum_{t=T_\text{E}+1}^{T_\text{E}+T_\text{F}} e_{it}^2}{T_\text{F}} \tag{2}$$

where T_E is the length of estimation period, and T_F is the number of forecasts. However, mean squared forecast errors (MSFE_i) tend to overly emphasize differences between models. Thus we will concentrate on the root mean squared forecast error, RMSE_i. Moreover, for pairwise comparisons we employ the RMSE ratio, defined as

$$\text{RR}_i = \frac{\text{RMSE}_i}{\text{RMSE}_0} \tag{3}$$

measuring the relative performance of model i, RMSE_i, with respect to that of a benchmark model, RMSE_0. A value exceeding one implies that the alternative model i is less accurate than the benchmark, whereas a value lower than one implies that it is more accurate.

Tests

Pairwise tests

Our aim is to test whether media data contain any valuable information for forecasting purposes. A minimum requirement is that media data improve upon the forecasts obtained using a simple autoregressive (AR) model. This involves pairwise comparisons of the AR model and alternative models consisting of the AR model augmented with one additional explanatory variable at a time. Let the loss difference be $d_{it} = L_{0t} - L_{it}$. The starting point for pairwise forecast comparisons of a benchmark model, 0, and an alternative model i is the Diebold and Mariano (1995) test statistic, defined as

$$\text{DM} = \frac{\bar{d}_i}{\hat{V}(\bar{d}_i)} \tag{4}$$

where $\bar{d}_i$ and $\hat{V}(\bar{d}_i)$ are the estimated mean and long-run variance of d_{it}, respectively. The null hypothesis of the test is

$$H_0\colon E(d_{it}) = 0 \tag{5}$$

meaning that the two models perform equally well.

By construction, the AR is nested in the augmented model. Under the null hypothesis, the additional variables are useless and their regression coefficients are zero. However, estimating additional variables introduces noise into the forecasts of the alternative model. Consequently, under the null, the forecast accuracy of the smaller benchmark is higher than that of the larger alternative model. Thus, in the context of nested models, the DM statistic has lower power and size. Therefore, we apply the modified test statistic of Clark and West (2007), which involves an adjustment term to improve upon the DM statistic when nested models are compared:

$$\text{CW} = \frac{\bar{d}_i - \bar{a}_i}{\hat{V}(\bar{d}_i - \bar{a}_i)} \tag{6}$$

where $\bar{a}_i = \frac{1}{T_\text{F}} \sum_{t=T_\text{E}+1}^{T_E+T_\text{F}} (\hat{y}_{0t} - \hat{y}_{it})^2$.

Multiple tests

White (2000) points out that, even when no exploitable forecasting relation exists, looking hard enough at a given set of data will often reveal one or more forecasting models that appear to be good, but are, in fact, useless. In particular, sequential testing a number of models by comparing two of them at a time invalidates standard critical values and might result in *data snooping*. In much of economic time series analysis this is aggravated by the fact that typically there is only a limited number of observations available. To overcome this problem White (2000) proposes a reality check test (RC-test) for data snooping. It compares the whole set of m rival models at a time to the benchmark. The null hypothesis is that none of the rival models is inferior to the benchmark:

$$H_0\colon E(d_{it}) \leq 0 \quad \forall i = 1, \ldots, m \tag{7}$$

It is rejected when at least one of the rivals yields significantly better forecasts. The expected loss differential can be consistently estimated using the sample mean, $\bar{d}_i$. White proposes the sample mean statistic

$$\text{RC} = \max_{k=1,\ldots,m} T_F^{1/2} \bar{d}_k \tag{8}$$

This approach is refined by Hansen (2005). He shows that the test statistic proposed by White is very conservative, if the set of rival models contains very poorly performing models, and proposes a refinement consisting of a Studentized version of the RC-test that is known as the test for superior predictive ability (SPA).

However, selecting the benchmark independently of the data introduces the problem of *multiple comparisons with control*. To overcome this Hansen *et al.* (2011) propose the model confidence set (MCS) approach. This approach involves looking for a set of best models, $\mathcal{M}^*$, such that, given the set of all forecasting models, $\mathcal{M}_0$, the MCS identifies the set of forecasting models that cannot be rejected as statistically inferior at a certain level of confidence:

$$\mathcal{M}^* = \left\{ i \in \mathcal{M}_0 \colon \mu_{ij} \leq 0 \text{ for all } j \in \mathcal{M}_0 \right\} \tag{9}$$

where $\mu_{ij} = E(d_{ij})$ is the expected loss differential $d_{ij} = L_{it} - L_{jt}$ based on one-by-one comparisons of all models.[6] The MCS is implemented using the following steps, where initially $\mathcal{M}$ is set $\mathcal{M} = \mathcal{M}_0$:

1. The null of equal predictive accuracy (EPA), $H_{0,\mathcal{M}}\colon \mu_{ij} \leq 0\, \forall i, j$ is tested at significance level α.
2. If the null is rejected, the worst-performing model is eliminated from $\mathcal{M}$.
3. The procedure is repeated until the null cannot be rejected. The set $\hat{\mathcal{M}}^*_{1-\alpha}$ with the remaining models is defined as the MCS.

In order to test the null hypothesis in step 1, we apply the $T_{\max,\mathcal{M}}$ statistic.[7] Let $\bar{d}_{ij} \equiv T_{\text{F}}^{-1} \sum_{t=T_{\text{E}}+1}^{T_{\text{E}}+T_{\text{F}}} d_{ij}$ be the relative sample loss between the ith and the jth models. Then, let $\bar{d}_{i\cdot}$ be the loss of the ith model relative to the average across models in $\mathcal{M}$, $\bar{d}_{i\cdot} \equiv m^{-1} \sum_{j \in \mathcal{M}} \bar{d}_{ij}$, where the models in $\mathcal{M}$ are again indexed by $i = 1, \ldots, m$. Then, the t-statistic is defined as

$$t_{i\cdot} = \frac{\bar{d}_{i\cdot}}{\sqrt{\hat{V}(\bar{d}_{i\cdot})}} \tag{10}$$

where $\hat{V}(\bar{d}_{i\cdot})$ is an estimate of $V(\bar{d}_{i\cdot})$. The t-statistic can be associated with the null $H_{i\cdot}\colon \mu_{i\cdot} = 0$, where $\mu_{i\cdot} = \text{E}(\bar{d}_{i\cdot})$. Hansen *et al.* (2011) show that the hypothesis $H_{0,\mathcal{M}}$ is equivalent to $\{H_{i\cdot}$ for all $i \in \mathcal{M}\}$ such that $\{\mu_{i\cdot} \leq 0$ for all $i \in \mathcal{M}\}$. Thus the null hypothesis $H_{0,\mathcal{M}}$ can be tested using the statistic

$$T_{\max,\mathcal{M}} = \max_{i \in \mathcal{M}} t_{i\cdot} \tag{11}$$

The asymptotic distribution of $T_{\max,\mathcal{M}}$ is non-standard as it depends on nuisance parameters. However, it can be estimated using bootstrap methods.

In the best of all cases, the MCS contains only one model. However, if the data are uninformative, the MCS contains many or even all models in $\mathcal{M}_0$. As a useful feature, the MCS procedure yields p-values for all models under consideration, where a small p-value indicates that the corresponding model is unlikely to enter the set of superior models, $\mathcal{M}^*$ (for details, see Hansen *et al.*, 2011).

Forecast breakdowns

Reliability of models is crucial for forecasters. However, models may become unstable over time due to instabilities in the economy. Thus forecasting performance depends on the success of the model at adapting to changes. Giacomini and Rossi (2009) introduce an approach to test whether the model's future performance is consistent with what is expected on the basis of its past performance. They define a forecast breakdown as a situation, in which the out-of-sample performance of a forecast model, judged by some loss function, is significantly worse than its in-sample performance. Their test is based on the intuition that, in the absence of a forecast breakdown, the difference between expected out-of-sample and in-sample performances should be close to zero.

The approach has two major advantages. First, in contrast to the literature concentrating on structural breaks, such as Bai and Perron (2003), the forecast breakdown test allows for model misspecification. Second, the test is applicable in cases of instabilities in the data-generating process of unknown form.

Let y_{t+h} be a sequence of h-step-ahead forecasts from an out-of-sample forecast experiment, which involves dividing the sample of size T into an in-sample window of size T_{E} and an out-of-sample window of size $T_{\text{F}}^h =$

[6] Here again, we choose the squared forecast error as loss.
[7] Hansen *et al.* (2011) propose an alternative test statistic, $T_{\text{R},\mathcal{M}}$. Here, we only report the results for the $T_{\max,\mathcal{M}}$ statistic as it yielded the most conservative results, i.e. the smallest confidence sets.

Table I. Analyzed media set

TV program/newspaper	Name	Number of news items
TV newscasts	ARD Tagesschau	11,472
	ARD Tagesthemen	14,933
	ZDF heute	10,158
	ZDF heute journal	15,415
	RTL Aktuell	6,167
Weekly magazines	Spiegel	4,833
	Focus	7,111
Daily newspaper	Bild	10,586
Total		80,675

$T - T_{\mathrm{E}} - h + 1$. A surprise loss at time $t + h$ is defined as the difference between the out-of-sample loss at time $t + h$ and the average in-sample loss:

$$\mathrm{SL}_{t+h} = L_{t+h} - \bar{L}_t, \quad \text{for } t = T_{\mathrm{E}}, \ldots, T - h \tag{12}$$

where $\bar{L}_t$ is the average in-sample loss computed over the in-sample window. If a forecast is reliable, this mean should be close to zero. Thus the null hypothesis is defined as

$$H_0\colon\ E\left(T_{\mathrm{F}}^{-1} \sum_{t=T_{\mathrm{E}}}^{T-h} \mathrm{SL}_{t+h}\right) = 0 \tag{13}$$

where $T_{\mathrm{F}}^{-1} \sum_{t=T_{\mathrm{E}}}^{T-h} \mathrm{SL}_{t+h}$ is the out-of-sample mean of the surprise losses. The test statistic is then defined as

$$t_{T_{\mathrm{E}},T_{\mathrm{F}},h} = T_{\mathrm{F}}^{1/2} \frac{\overline{\mathrm{SL}}_{T_{\mathrm{E}},T_{\mathrm{F}}}}{\hat{\sigma}_{T_{\mathrm{E}},T_{\mathrm{F}}}} \tag{14}$$

where $\hat{\sigma}^2_{T_{\mathrm{E}},T_{\mathrm{F}}}$ is an asymptotic variance estimator. The test statistic follows asymptotically the standard normal distribution.

THE DATA

Our analysis is based on the assessment by Media Tenor International (MTI), the Swiss-based media analysis institute, of the content of opinion-leading media in Germany, including five television news programs, two weekly magazines and one daily tabloid newspaper. News items only referring to the state of the economy in the media set are analyzed over the period 1 January 2000 to 31 March 2014. Hence the analyzed dataset can be seen as a subset of a much bigger dataset covering news items on all possible protagonists, including individuals (politicians, entrepreneurs, managers, celebrities, etc.) and institutions (political parties, companies, football clubs, etc.). Each of these news items is analyzed with regard to the topic mentioned (unemployment, inflation, etc.), the region of reference (e.g. Germany, EU, USA, UK, BRIC, whole world), the time reference (past, present and future), the source of information (journalist, politician, expert, etc.), as well as with regard to the tone of the information (negative, positive or neutral).[8] Overall, the analysis includes 80,675 news items about the state of the economy. For a description of the analyzed media set see Table I. Of all the topics only government budget, monetary policy, labor market, business cycle and taxation contain enough observations to compute sub-indices with a complete set of monthly observations. Table II is a contingency table of these topics. It reports the number of observations of the topics in temporal (present, past and future) and spatial dimensions (Germany vs. the rest of the world). In particular, it can be seen that a slight majority of media reports concerning past and present are about Germany, while twice as many reports concerning future events deal with Germany. Moreover, 61% of media reports on economic issues are related to the present and 34% to the future. Thus a very large part of the reports refer to the future events.

[8] MTI employs professional coders to carry out media analysis. Only coders who achieve a minimum reliability of 0.85 are cleared for coding. That means that the coding of these coders deviates at most by 0.15 from the trainers' master versions. The reliability of the coding is checked on an ongoing basis, both with quarterly standard tests and random spot checks. For each month and coder, three analyzed reports are selected at random and checked. Coders scoring lower than 0.80 are removed from the coding process. The mean deviation among all coders is below 0.15 in all months.

Table II. Media data: contingency table

Time reference:	Present		Past		Future		All times		
Region:	Germany	ROW	Germany	ROW	Germany	ROW	Germany	ROW	Total
Government budget	4,286	6,470	232	441	2,864	3778	7,382	10,689	18,071
Monetary policy	1,145	1,115	115	136	382	449	1,642	1,700	3,342
Labor market	8,865	2,140	519	126	4,296	361	13,680	2,627	16,307
Business cycle	4,227	6,257	253	510	2,719	1649	7,199	8,416	15,615
Taxation	3,167	694	123	21	3,420	295	6,710	1,010	7,720
Others	6,905	4,343	549	591	4,235	2997	11,689	7,931	19,620
Total	28,595	21,019	1791	1825	17,916	9529	48,302	32,373	80,675

Note: ROW, rest of the world.

We computed 17 MTI indices using different (sub)samples of the news data. We distinguish between two types of indices depending on the way they are constructed.

The first type is the difference between the percentage share of the positive ratings and that of the negative ratings:

$$B_{i,j,t} = 100 \times \frac{A^{+}_{i,j,t} - A^{-}_{i,j,t}}{A^{+}_{i,j,t} + A^{-}_{i,j,t} + A^{0}_{i,j,t}} \tag{15}$$

where $A^{+}_{i,j,t}$ is the number of positive ratings of media reports about events happening in time i in country j, published in the period t; $A^{-}_{i,j,t}$ is the number of negative ratings; and $A^{0}_{i,j,t}$ is the number of neutral ratings. The index varies between -100 (all reports are negatively rated) and 100 (all reports are positively rated).

The second type uses the indices of the present and the future sentiment to construct a so-called *media climate index* analogous to the ifo business climate index:

$$\mathrm{MCI} = \sqrt{(B^{\mathrm{P}}_{j,t} + 100)(B^{\mathrm{F}}_{j,t} + 100)} \tag{16}$$

where $B^{\mathrm{P}}_{j,t}$ is the present sentiment index and $B^{\mathrm{F}}_{j,t}$ is the future sentiment index. By construction, the MCI can take values between 0, indicating extremely bad media climate, and 200, pointing to an excellent media climate.

Based on these two types, we distinguish between four groups of MTI indicators:

1. The first set of indicators differs only in the time reference of the respective databases. MT.all uses all the data with all time references (past, present and future). MT.present and MT.future are based on the data with the time references present and future, correspondingly. Finally, MT.climate is computed based on the items with present and future reference as in equation (16).
2. The second group is the same as the first. However, it exclusively uses data referencing Germany. The indices are labeled with 'de'.
3. The second group of indicators only uses data related to specific topics: the government budget (MT.budget), monetary issues (MT.monetary), the labor market (MT.labor), the business cycle (MT.cycle) and taxation (MT.taxation). The selection of subsets is restricted by data availability. The more specialized the topic, the fewer observations are available.
4. The fourth group is the same as the third; however, it exclusively uses data with a reference to Germany. The indices are again labeled with 'de'.[9]

Table III shows the mean and the median of the media indices. Except for MT.climate and MT.de.climate, all indices have a negative mean. This is in line with the literature (see, for example, Brettschneider, 2000) that finds that negative news dominates the media. Moreover, the standard deviations of some sub-indices, particularly MT.cycle and MT.de.cycle, are very high when compared to MT.all, which is based on the whole dataset.

In addition to the media indicators, we collect variables from the database used in Drechsel and Scheufele (2012). This is a large dataset comprising financial indicators, survey indices, prices and wages, indicators of the real economy and composite indicators. For details see Table IV and Drechsel and Scheufele (2012).[10] The different transformations

[9] As the number of observations for MT.de.monetary would have been too small, it is not constructed.

[10] Some of their variables are missing in our dataset. We could not construct five spread series involving corporate rates computed by Merryl Lynch and the early bird indicator of the Commerzbank, as the data are not publicly available. Furthermore, we only obtained very short series for the purchasing manager index of Markit. Finally, we did not include monetary aggregates as, since the introduction of the euro, no country-specific monetary aggregate can be identified. Since the introduction of the euro in Germany, there is only an indicator for money supply of the entire euro area available.

Table III. Descriptive statistics of media indices

	Mean	SD
MT.all	−29.89	16.04
MT.de	−20.85	19.07
MT.present	−35.14	16.96
MT.future	−20.13	18.91
MT.climate	71.58	16.26
MT.monetary	−26.77	34.24
MT.taxation	−26.33	15.17
MT.cycle	−22.37	35.43
MT.labor	−28.06	19.74
MT.budget	−42.83	26.92
MT.de.present	−24.92	22.27
MT.de.future	−14.41	18.94
MT.de.climate	79.57	18.62
MT.de.taxation	−26.67	15.61
MT.de.cycle	−1.57	47.91
MT.de.labor	−23.35	22.00
MT.de.budget	−29.07	33.48

Table IV. Data: definitions, transformations and sources

Block	Name	Label	L	D	Dln	D2ln	Lag	Source
Dependent variable	Industrial production	IP	—	—	1	—	2	Buba RTDB
Financial	Money market rate (monthly average)	IS-M	1	1	—	—	0	Buba
	Discount rate/short-term repo rate (monthly average)	IS-D	1	1	—	—	0	Buba
	3-month money market rate (monthly average)	IS-3M	1	1	—	—	0	Buba
	Yields on debt securities outstanding (maturity 3–5 years)	IL-3	1	1	—	—	0	Buba
	Yields on debt securities outstanding (maturity 5–8 years)	IL-5	1	1	—	—	0	Buba
	Long-term government bond yield—9–10 years	IL-10	1	1	—	—	0	Buba
	Term spread (10 years—money market rate)	SPR-10Y-M	1	—	—	—	0	Buba
	Term spread (10 years—discount rate)	SPR-10Y-D	1	—	—	—	0	Buba
	Term spread (10 years—3-month money market rate)	SPR-10Y-3M	1	—	—	—	0	Buba
	Term spread (discount rate—money market rate)	SPR-1D-M	1	—	—	—	0	Buba
	Corporate bond—government bonds	SPR-C-G	1	—	—	—	0	Buba
	Nominal effective exchange rate	EX	—	—	1	—	1	Buba
	Real effective exchange rate	EXR	—	—	1	—	1	Buba
	DAX	DAX	—	—	1	—	0	Buba
	DAX volatility new	VOLA1	1	1	—	—	0	Buba
	DAX volatility old	VOLA2	1	1	—	—	0	Buba
	HWWA index of world market prices of raw materials	HWWA	—	—	1	1	1	datastream
	HWWA index, real	HWWAR	—	—	1	1	—	datastream
	HWWA index, energy	HWWA-E	—	—	1	1	1	Buba
	HWWA index, energy real	HWWA-ER	—	—	1	1	—	Buba
	HWWA index, excl. energy	HWWA-EX	—	—	1	1	1	Buba
	HWWA index, excl. energy real	HWWA-EXR	—	—	1	1	—	Buba
	Oil prices (euros per barrel)	OIL	—	—	1	1	0	ECB
	Oil prices (euros per barrel), real	OILR	—	—	1	1	—	ECB
Surveys	Ifo index climate	IFO-C	1	1	—	—	0	ifo
	Ifo expectations climate	IFO-EXP	1	1	—	—	0	ifo
	Ifo index manufacturing	IFOM-C	1	1	—	—	0	ifo
	Ifo expectations manufacturing	IFOM-EXP	1	1	—	—	0	ifo
	Ifo index capital goods	IFOMI-C	1	1	—	—	0	ifo
	Ifo expectations capital goods	IFOMI-EXP	1	1	—	—	0	ifo
	Ifo index intermediate goods	IFOMV-C	1	1	—	—	0	ifo
	Ifo expectations intermediate goods	IFOMV-EXP	1	1	—	—	0	ifo

Table IV. Continued

Block	Name	Label	L	D	Dln	D2ln	Lag	Source
	Ifo index wholesale	IFOWH-C	1	1	—	—	0	ifo
	Ifo expectations wholesale	IFOWH-EXP	1	1	—	—	0	ifo
	GFK consumer climate survey—business cycle expectations	GFK-EXP	1	1	—	—	0	datastream
	ZEW economic sentiment	ZEW	1	1	—	—	0	datastream
	Assessment of order-book levels	ECBS2	1	1	—	—	0	EC
	Assessment of export order-book levels	ECBS3	1	1	—	—	0	EC
	Assessment of stocks of finished products	ECBS4	1	1	—	—	0	EC
	Production expectations for the months ahead	ECBS5	1	1	—	—	0	EC
	Selling price expectations for the months ahead	ECBS6	1	1	—	—	0	EC
	Employment expectations for the months ahead	ECBS7	1	1	—	—	0	EC
	Industrial confidence indicator (40%)	ESI-INDU	1	1	—	—	0	EC
	Services confidence indicator (30%)	ESI-SERV	1	1	—	—	0	EC
	Consumer confidence indicator (20%)	ESI-C	1	1	—	—	0	EC
	Retail trade confidence indicator (5%)	ESI-TRADE	1	1	—	—	0	EC
	Construction confidence indicator (5%)	ESI-CTR	1	1	—	—	0	EC
	Economic sentiment indicator (average)	ESI	1	1	—	—	0	EC
	Confidence Indicator (Q2 + Q4 − Q7 + Q11) / 4	ECCS99	1	1	—	—	0	EC
	Financial situation over last 12 months	ECCS1	1	1	—	—	0	EC
	Financial situation over next 12 months	ECCS2	1	1	—	—	0	EC
	General economic situation over last 12 months	ECCS3	1	1	—	—	0	EC
	General economic situation over next 12 months	ECCS4	1	1	—	—	0	EC
	Price trends over last 12 months	ECCS5	1	1	—	—	0	EC
	Price trends over next 12 months	ECCS6	1	1	—	—	0	EC
	Unemployment expectations over next 12 months	ECCS7	1	1	—	—	0	EC
	Major purchases at present	ECCS8	1	1	—	—	0	EC
	Major purchases over next 12 months	ECCS9	1	1	—	—	0	EC
	Savings at present	ECCS10	1	1	—	—	0	EC
	Savings over next 12 months	ECCS11	1	1	—	—	0	EC
	Statement on financial situation of household	ECCS12	1	1	—	—	0	EC
Prices and wages	Consumer price index	CPI	—	—	1	1	0	Buba RTDB
	Core consumer price index	CPI-EX	—	—	1	1	0	Buba RTDB
	Negotiated wage and salary level	TARIF	—	—	1	1	0	Buba RTDB
Real economy	Intermediate goods production	IP-VORL	—	—	1	—	0	Buba RTDB
	Manufacturing orders—consumer goods	ORD-C	—	—	1	—	0	Buba RTDB
	Manufacturing orders—capital goods	ORD-I	—	—	1	—	0	Buba RTDB
	Employed persons (workplace concept)	EW	—	—	1	—	0	Buba RTDB
	1 + unemployment (% civilian labour)	ALQ	—	1	—	—	1	datastream
	Vacancies	VAC	—	—	1	—	1	datastream
	Capacity utilization	CAPA	1	1	—	—	0	datastream
	Hours worked	WHOUR	1	1	—	—	0	Buba RTDB
Composite indicators	FAZ indicator	FAZ	—	—	1	—	1	datastream
	Composite leading indicator (amplitude restored)	OECDL1	1	1	—	—	2	OECD
	Composite leading indicator (trend restored)	OECDL2	—	1	—	—	2	OECD
	Composite leading indicator (normalized)	OECDL3	1	1	—	—	2	OECD
Media indicators	All observations	MT.all	1	1	—	—	0	MTI
	All observations related to the future	MT.future	1	1	—	—	0	MTI
	All observations related to the present	MT.present	1	1	—	—	0	MTI
	all observations related to the future and present	MT.climate	1	1	—	—	0	MTI
	All observations related to Germany	MT.de	1	1	—	—	0	MTI
	All observations related to Germany and the future	MT.de.future	1	1	—	—	0	MTI
	All observations related to Germany and present	MT.de.present	1	1	—	—	0	MTI
	All observations related to Germany, and future and present	MT.de.climate	1	1	—	—	0	MTI
	All observations related to government budget	MT.budget	1	1	—	—	0	MTI
	All observations related to monetary issues	MT.monetary	1	1	—	—	0	MTI
	All observations related to the labor market	MT.labor	1	1	—	—	0	MTI
	All observations related to the business cycle	MT.cycle	1	1	—	—	0	MTI
	All observations related to taxation	MT.taxation	1	1	—	—	0	MTI
	All observations related to the German government budget	MT.de.budget	1	1	—	—	0	MTI

Table IV. Continued

Block	Name	Label	L	D	Dln	D2ln	Lag	Source
	All observations related to the German labor market	MT.de.labor	1	1	—	—	0	MTI
	All observations related to the German business cycle	MT.de.cycle	1	1	—	—	0	MTI
	All observations related to German taxation	MT.de.taxation	1	1	—	—	0	MTI

Note: The different transformations of the raw data are either level (L), differences (D), differences of natural logarithms (Dln) or second differences of natural logarithms (D2ln). The publication lag (Lag) ranges from 0 to 3 months. The sources are Datastream, Deutsche Bundesbank (Buba), Deutsche Bundesbank Realtime Database (Buba RTDB), European Commission (EC), European Central Bank (ECB), Ifo institute for economic research (ifo), Organization for Economic Co-operation and Development (OECD) and Media Tenor International (MTI). As deflators real-time CPI series of Deutsche Bundesbank have been used.

of the raw data are: level (L), differences (D), differences of natural logarithms (Dln), or second differences of natural logarithms (D2ln). For the reference month the data are available with a publication lag $\underline{q}$ that varies from 0 to $\overline{2}$.

However, unlike Drechsel and Scheufele (2012), we employ real-time data. Revisions of measures of real economic activity, such as employment, sales and, in particular, industrial production, are sometimes large, and may occur years after official figures are first released. The information set that has been available at different points in time—the different vintages of the data—frequently conveys a different picture of the same period of time. Thus the use of current-vintage data, i.e. the data as they are available when the experiment is conducted, can lead an analyst to include variables in his forecasting model that, in real time, have little marginal predictive power (see, for example, Orphanides, 2001, or Koenig *et al.*, 2003). Furthermore, if we assume that revisions lead to an improvement of the indicators, this puts indicators that are unrevised, such as financial data or the media data that are analyzed here, in a disadvantageous position.

RESULTS

We analyze the whole set of out-of-sample predictions for each horizon. Furthermore, we are interested in the models' reliability during challenging times from a forecaster's perspective. Drechsel and Scheufele (2012) implement this by comparing the forecast performance before and during the Great Recession, as identified by the Euro Area Business Cycle Dating Committee at the Centre for Economic Policy Research (CEPR). However, we find that applying this chronology to the German case is inappropriate. CEPR analyzes recession dates for the whole euro area but individual countries reacted differently to the crisis. In particular, in Germany the recession was much milder and shorter than in other euro area countries, especially Greece and Spain. Thus it would be more appropriate to consider instead Germany-specific recession dates, such as the dates provided by the Economic Cycle Research Institute (ECRI). While according to CEPR the recession covered the period from 2008:01 to 2009:06, according to ECRI it started 2008:02 and lasted until 2009:02. We use ECRI chronology as a reference. However, we are aware of the fact that the official recession dating and the times, when the forecast performance of many models substantially deteriorates, need not coincide. We focus on those periods when considerably more models than usual suffer a forecast breakdown, as defined by Giacomini and Rossi (2009). We define an unstable period as a time when more than 20% of all models across different forecast horizons suffer a breakdown. The period of model instability starts in 2008:12 and ends in 2010:06. Thus most of the forecast breakdowns of single models occur when industrial production is predicted at the end of the recession and long after it is over.

This is reflected in the reaction of industrial production to the recession. Figure 1 shows German industrial production over the forecast subsample. On top of the figure, the recession period as defined by ECRI (empty circles) and periods of the model instability (filled circles) are shown. Large changes of the dependent variable occur at the end of the official recession dates and many months thereafter. At the beginning of the official start of the recession in 2008:05, industrial production dropped only by <5%. It took six more months until the end of the ECRI recession in 2009:01 for it to go down to about 80% of its pre-crisis level. However, the trough was not reached until 2009:04. Only in mid 2011 did industrial production return to its pre-crisis level.

Table V reports the RMSE, Theil's U, the rank according to RMSE ratio and the MCS p-value of the best media models for each horizon over all periods. The last column presents the models with the lowest RMSE ratio of all models, not only the media models. In line with the literature, our benchmark model is the autoregressive model. It is defined as in equation (1) under the restriction that $\gamma_q = 0, \forall q$, which means that only own lags of the dependent variable are included. Two models—MT.de.future and MT.de.climate—stand out as particularly useful for horizons from 10 to 12 months. The former has the lowest RMSE ratio for 10- and 12-month horizons, with values of 0.83 and 0.77, respectively. Furthermore, for both horizons it forms part of the $\mathcal{M}^*_{75\%}$ as the best model,[11] having an MCS

[11] Hansen *et al.* (2011) employ $\mathcal{M}^*_{75\%}$ and $\mathcal{M}^*_{90\%}$. We use the former, as it is more restrictive in the sense that it selects fewer models into the MCS.

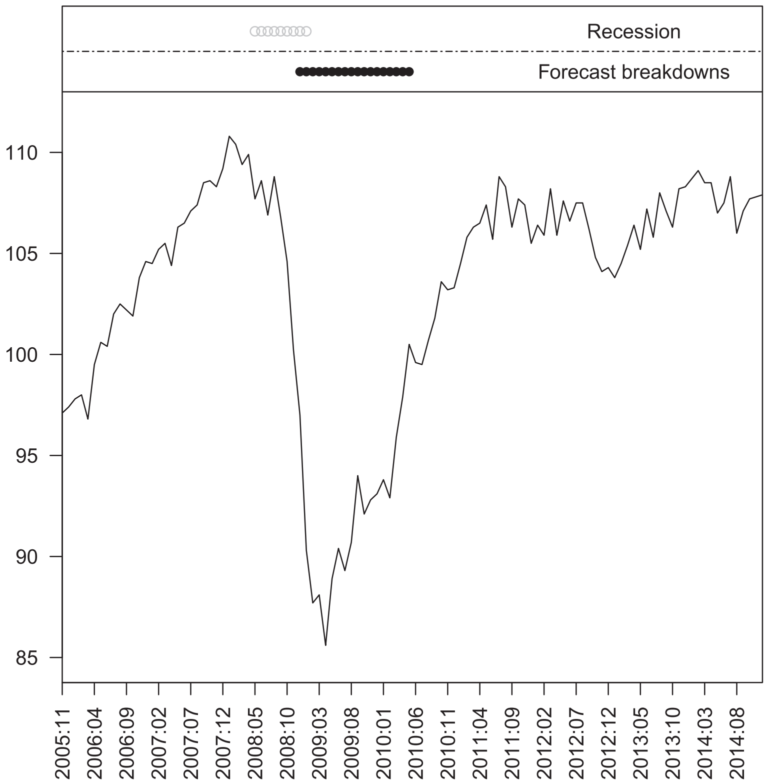

Figure 1. Industrial production, recession and forecast breakdowns

Table V. Best media models, all periods (2005:12–2014:04)

h	Best media model	RMSE	RMSE ratio	Rank RMSE ratio	MCS p-value	Best model each horizon
1	MT.de.present	28.37	0.91	93	0.00	ECBS5
2	MT.de.present	32.73	0.89	94	0.00	IFOMI.C
3	MT.present	33.78	0.87	90	0.00	DOECDL2
4	MT.present	31.08	0.84	71	0.00	IFOMI.C
5	MT.de.present	23.27	0.74	44	0.13	DOECDL2
6	MT.de.present	17.50	0.66	33	0.25*	IFOMI.C
7	MT.present	13.33	0.68	27	0.24	DOECDL2
8	MT.present	10.53	0.83*	13	0.40*	DIFO.C
9	MT.de.future	9.17	0.85**	5	0.43*	DIFO.C
10	MT.de.future	8.43	0.83**	1	1.00*	MT.de.future
11	MT.de.climate	8.69	0.77**	1	1.00*	MT.de.climate
12	MT.de.future	8.46	0.77**	1	1.00*	MT.de.future

Note: * (**) for RMSE ratio denotes significant outperformance of the AR model to the 5 (1)% level using the CW-statistic, whereas * for MCS p-values indicates that the respective model is in the $\mathcal{M}^*_{75\%}$. The best model refers to the model having the lowest RMSE ratio. The forecast horizon in months is denoted by h.

p-value of 1. Moreover, it significantly outperforms the AR model at the 1% level for both horizons. For the 11-month horizon MT.de.climate is the best model, both with respect to the RMSE ratio as well as the MCS p-value. It also significantly outperforms the AR at the 1% level. For horizons up to 9 months the models with the lowest RMSE ratio are survey-based indicators that directly measure the production decisions in the industry once they are taken.

Table VI shows the results for the unstable period. They remain robust. MT.de.future has an RMSE ratio of 0.82 and 0.77 for the 10- and 12-month horizon forecasts and significantly outperforms the AR model. However, it is slightly worse, according to the model ratings, since it attains only the third and second rank for the 10- and 11-month horizons. While still forming part of $\mathcal{M}^*_{75\%}$ with high MCS p-values of 0.80 and 0.89 for the 10- and 12-month horizons, it is not the best model in the MCS. MT.de.climate significantly outperforms the AR, ranking second for the 11-month forecast horizon. As with MT.de.future, it is not the best model in the $\mathcal{M}^*_{75\%}$, having an MCS p-value of 0.76. Here, for all horizons survey-based indicators outperform the alternatives with respect to RMSE ratio. The reason might be that in unstable periods volatile decisions can best be measured directly.

In the first two rows of Table VII, the percentage of times MT.de.future and MT.de.climate failed over all periods and the unstable period for horizons 10–12 months is reported. The next rows show the results of the indicators that

Table VI. Best media models, unstable period (2008:12–2010:06)

h	Best media model	RMSE	RMSE ratio	Rank RMSE ratio	MCS p-value	Best model each horizon
1	MT.de.present	63.73	0.88	92	0.00	ECBS5
2	MT.de.present	74.88	0.88	94	0.00	IFOMI.C
3	MT.present	77.27	0.85	89	0.00	IFOMI.C
4	MT.present	70.76	0.82	71	0.00	DOECDL2
5	MT.de.present	51.98	0.72	46	0.00	DOECDL2
6	MT.de.present	37.96	0.63	35	0.02	IFOMI.C
7	MT.de.present	26.60	0.62	26	0.00	DOECDL2
8	MT.present	19.94	0.77	12	0.00	ECCS5
9	MT.de.future	18.09	0.86**	14	0.00	ECCS5
10	MT.de.future	15.90	0.82**	3	0.80*	ECCS5
11	MT.de.climate	14.08	0.77**	2	0.76*	ECCS5
12	MT.de.future	13.98	0.77**	2	0.89*	ECCS5

Note: * (**) for RMSE ratio denotes significant outperformance of the AR model to the 5% (1%) level using the CW-statistic, whereas * for MCS p-values indicates that the respective model is in the $\mathcal{M}^*_{75\%}$. The best model refers to the model having the lowest RMSE ratio. The forecast horizon in months is denoted by h.

Table VII. Percentage of forecast breakdowns

	$h=10$		$h=11$		$h=12$	
Model	All periods	Unstable	All periods	Unstable	All periods	Unstable
MT.de.future	0.19	0.74	0.21	0.58	0.20	0.63
MT.de.climate	0.18	0.58	0.20	0.58	0.23	0.68
Lowest value h=10 (DESI.INDU)	0	0	0.2	0.74	0.11	0.32
Lowest value h=11,12 (ZEW)	0.14	0.42	0	0	0.02	0.11
mean all models	0.19	0.56	0.20	0.51	0.21	0.51

performed best for horizons 10–12 months and the last row shows the respective averages over all models. A value of 0 implies that the model never failed, while a value of 1 means that the model always failed. For all periods, the average of all models rises slightly between $h=10$ and $h=12$ from 19% to 21%. Both media models' percentages are roughly at the same level. However, for the unstable periods, they are markedly less reliable than the average. On average all models fail 56%, 51%, and 51% of times for horizons 10, 11 and 12 respectively. MT.de.future fails 74%, 58% and 63% of times, whereas MT.de.climate fails 58%, 58%, and 68% of times for horizons 10, 11 and 12, correspondingly. Notably, with respect to 10-month horizon the ESI industrial indicator (ESI.INDU) and, with respect to 11-month and 12-month horizon, the ZEW indicator (nearly), never fail. However, this comes at a considerable cost in terms of forecast accuracy. The RMSE ratio for ESI.INDU for the 10-month horizon is 0.90, and for the ZEW for the 11-month and 12-month horizon the RMSE ratios are 0.89 and 0.90.

DISCUSSION

The results appear to be at odds with the existing literature. Grossarth-Maticek and Mayr (2008) find that the R-word index is suitable for detecting turning points in the business cycle in real time. Iselin and Siliverstovs (2015) find that media indicators are helpful for one-quarter-ahead forecasts. Partly, they are even better than other indicators. In contrast, we show that media indicators display better forecast accuracy at 10- to 12-month forecast horizons. The question is whether our results are so different from those two studies and, if yes, what the differences are in our set-up that can explain such a deviation in our results.

The papers cited above have several common points. First, both studies use quarterly GDP series or GDP-based recession dates as a dependent variable. We use a monthly series of industrial production. Although relatively strongly correlated, GDP and industrial production display different time profiles and possess different time-series properties. Second, both studies consider only short-term horizons. Grossarth-Maticek and Mayr (2008) focus on the first two lags (1 and 2 quarters), whereas Iselin and Siliverstovs (2015) concentrate only on the one-step-ahead forecast. In contrast, we examine different horizons varying between 1 and 12 months ahead. Therefore, we simply do not know what would be the outcomes of the two alternative studies had they looked at the higher forecast horizons. It is possible that the forecast accuracy of their media-based forecast would be even higher at longer horizons and, thus, similar to our results. To some extent this is supported by the fact that Grossarth-Maticek and Mayr (2008) find (see their Table II) that the out-of-sample RMSFE for R-Wort(-1) is smaller than for R-Wort, which points to a better forecast accuracy at longer lags.

Furthermore, it seems implausible at first glance that survey indicators (such as ifo) perform better in the short run than the media indicators. As media influence the expectation formation process immediately, it should also affect survey indicators. One explanation is that business survey indicators depend besides media on the order books of the firms. Therefore, the expectations of businesses have an important sticky component, for the order books have been filled over the past periods.

CONCLUSION

In this paper we examine the usefulness of media indicators for predicting the monthly series of German industrial production. We use MTI indices that are based on human analysis of reports in opinion-leading media in Germany. The forecast performance is evaluated through a real-time forecast experiment covering the period December 2005 to March 2014. In addition, we identify a period when many models fail with respect to their previous performance. This period begins at the end of the recession as dated by ECRI and ends more than a year afterwards. In doing this we evaluate the stability of models based on the media indices and determine whether they are useful during large economic fluctuations. The performance of media indices is compared to that of a large set of alternative indicators.

The forecast performance is evaluated using several criteria. First, we use two measures of forecast accuracy, namely the root mean squared forecast error and the RMSE ratio with respect to the simple autoregressive process. The performance of individual models is tested against that of the autoregressive model using the test proposed by Clark and West (2007). Moreover, the best-performing models are identified using the model confidence set procedure of Hansen *et al.* (2011). Finally, as a measure of reliability, we employ the test for forecast breakdowns proposed by Giacomini and Rossi (2009).

The results clearly show that models using media data outperform models without media data for relatively long forecast horizons, ranging from 10 to 12 months. Media data that contain economic information on all issues for only Germany and are constructed along the lines of the ifo climate and expectation indices rank first, according to RMSE, RMSE ratio and the p-value of the model confidence set, and outperform the benchmark autoregressive model individually.

The time lag corresponds to the time lag of the effect of monetary policy announcements on production (see George *et al.*, 1999) and thus is hardly surprising. In both cases the lag can be explained by the fact that media reports affect the expectations of economic agents about the future course of the economy. It takes some time before the change in expectation leads to new orders. In addition, on the side of producers, for technical reasons it takes some time before expectations materialize in the form of production plans. Afterwards, firms need time to hire or fire workers and to adjust their capacities. Thus the lag between the publication of the news and change in industrial production can be rather lengthy.

Nonetheless, during the unstable period at the end of and almost 1 year after the most recent recession, survey-based indicators are more reliable in the sense that they fail less frequently. However, this comes at a considerable cost in terms of forecast accuracy.

Taking all results into account, with respect to the prediction of industrial production, the use of the media sentiment indicators can be recommended for longer forecast horizons, since it significantly improves the performance of the forecast models.

REFERENCES

Abberger K, Wohlrabe K. 2006. Einige Prognoseeigenschaften des ifo Geschäaftsklimas: Ein Überblick über die neuere wissenschaftliche Literatur. *IFO Schnelldienst* **59**(22): 19–26.

Ammann M, Frey R, Verhofen M. 2014. Do newspaper articles predict aggregate stock returns? *Journal of Behavioral Finance* **15**(3): 195–213.

Bai J, Perron P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**(1): 1–22.

Bollen J, Mao H, Zeng X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* **2**: 1–8.

Bordino I, Battiston S, Caldarelli G, Cristelli M, Ukkonen A, Weber I. 2012. Web search queries can predict stock market volumes. *PLoS ONE* **7**(7): 1–17.

Brettschneider F. 2000. Reality bytes: Wie die Medienberichterstattung die Wahrnehmung der Wirtschaftslage beeinflußt. In *Wirklich ein Volk? Die politischen Orientierungen von Ost- und Westdeutschen im Vergleich*, Falter JW, Gabriel OW, Rattinger H (eds). Leske + Budrich: Opladen, Germany; 539–569.

Clark TE, West KD. 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* **138**(1): 291–311.

Denzau AT, North DC. 1994. Shared mental models: ideologies and institutions. *Kyklos* **47**(1): 3–31.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**(3): 253–263.

Doms M, Morin N. 2004. *Consumer sentiment, the economy, and the news media. Finance and Economics*, Discussion Series 2004-51, Board of Governors of the Federal Reserve System, Washington.

Drechsel K, Scheufele R. 2012. The performance of short-term forecasts of the German economy before and during the 2008/2009 recession. *International Journal of Forecasting* **28**(2): 428–445.

European Commission. 2014. For a European industrial renaissance. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions COM 14/2.
George E, King M, Clementi D, Budd A, Buiter W, Goodhart C, Julius D, Plenderleith I, Vickers J. 1999. *The Transmission Mechanism of Monetary Policy*. Bank of England: London.
Giacomini R, Rossi B. 2009. Detecting and predicting forecast breakdowns. *Review of Economic Studies* **76**(2): 669–705.
Goidel RK, Langley RE. 1995. Media coverage of the economy and aggregate economic evaluations: uncovering evidence of indirect media effects. *Political Research Quarterly* **48**: 313–328.
Grossarth-Maticek J, Mayr J. 2008. Medienberichte als Konjunkturindikator. *IFO Schnelldienst* **61**: 17–29.
Hansen PR. 2005. A test for superior predictive ability. *Journal of Business and Economic Statistics* **23**: 365–380.
Hansen PR, Lunde A, Nason JM. 2011. The model confidence set. *Econometrica* **79**(2): 453–497.
Hüfner F, Schröder M. 2002. Prognosegehalt von ifo-Geschäftserwartungen und ZEW- Konjunkturerwartungen: Ein ökonometrischer Vergleich. *Jahrbücher für Nationalökonomie und Statistik* **222**(3): 316–336.
Iselin D, Siliverstovs B. 2013. The R-word index for Switzerland. *Applied Economics Letters* **20**(11): 1032–1035.
Iselin D, Siliverstovs B. 2015. Using newspapers for tracking the business cycle: a comparative study for Germany and Switzerland. *Applied Economics* **48**(12): 1103–1118.
Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* **47**(2): 263–291.
Keynes JM. 1937. The general theory of employment. *Quarterly Journal of Economics* **51**(2): 209–223.
Kholodilin KA, Siliverstovs B. 2006. On the forecasting properties of the alternative leading indicators for the German GDP: Recent evidence. *Jahrbücher für Nationalökonomie und Statistik* **226**(3): 234–259.
Kholodilin KA, Podstawski M, Siliverstovs B. 2010. *Do Google searches help in nowcasting private consumption? A real-time evidence for the US*, Discussion Papers of DIW Berlin 997, German Institute for Economic Research.
Koenig EF, Dolmas S, Piger J. 2003. The use and abuse of real-time data in economic forecasting. *Review of Economics and Statistics* **85**(3): 618–628.
Lamla MJ, Maag T. 2012. The role of media for inflation forecast disagreement of households and professional forecasters. *Journal of Money, Credit and Banking* **44**(7): 1325–1350.
Nadeau R, Niemi RG, Amato T. 2000. Elite economic forecasts, economic news, mass economic expectations, and voting intentions in Great Britain. *European Journal of Political Research* **38**: 135–170.
Orphanides A. 2001. Monetary policy rules based on real-time data. *American Economic Review* **91**(4): 964–985.
Pesaran MH, Timmermann A. 2004. How costly is it to ignore breaks when forecasting the direction of a time series *International Journal of Forecasting* **20**(3): 411–425.
Robinzonov N, Wohlrabe K. 2010. Freedom of choice in macroeconomic forecasting. *CESifo Economic Studies* **56**(2): 192–220.
Simon HA. 1957. *Models of Man: Social and Rational—Mathematical Essays on Rational Human Behavior in a Social Setting*. Wiley: New York.
Soroka SN. 2006. Good news and bad news: Asymmetric responses to economic information. *Journal of Politics* **68**(2): 372–385.
Tetlock PC. 2007. Giving content to investor sentiment: the role of media in the stock market. *Journal of Finance* **62**(3): 1139–1168.
Uhl MW. 2011. Explaining US consumer behavior with news sentiment. *ACM Transactions on Management Information Systems* **2**(2): 9.
Uhl MW. 2012. And action: TV sentiment and the US consumer. *Applied Economics Letters* **19**(11): 1029–1034.
von Hayek FA. 1989. The pretence of knowledge. *American Economic Review* **79**(6): 3–7.
White H. 2000. A reality check for data snooping. *Econometrica* **68**(5): 1097–1126.

Authors' biographies:

Dirk Ulbricht is director of the institute for financial services (iff), Hamburg. His research interests are forecasting, housing, and overindebtedness.

Konstantin A. Kholodilin is a senior researcher at the DIW Berlin. His research interests include spatial econometrics, non-linear modelling, time-series forecasting, and analysis of the economic convergence.

Tobias Thomas is lecturer and research affiliate at Düsseldorf Institute for Competition Economics (DICE) and Head Research of media research institute Media Tenor International. In addition he is founding member and chairman of ECONWATCH - Society for Policy Analysis, and research fellow at the Institute of Economics at Helmut Schmidt University Hamburg, Germany.

Authors' addresses:

Dirk Ulbricht, iff, Rödingsmarkt 31/33, 20459 Hamburg, Germany.

Konstantin A. Kholodilin, DIW, Mohrenstraße 58, 10117, Berlin, Germany.

Tobias Thomas, Düsseldorf Institute for Competition Economics (DICE), Heinrich-Heine- Universität, Universitätsstraße 1, 40225 Düsseldorf, Germany.

