Institute for Employment Research

The Research Institute of the Federal Employment Agency

IAB-Discussion Paper 30/2015

Articles on labour market issues

Forecasting employment in Europe

Are survey results helpful?

Robert Lehmann
Antje Weyh

ISSN 2195-2663

Forecasting employment in Europe

Are survey results helpful?

Robert Lehmann (ifo Institute, Branch Dresden)
Antje Weyh (IAB)

Contents

Abstract

In this paper we evaluate the forecasting performance of employment expectations for employment growth in 15 European states. Our data cover the period from the first quarter 1998 to the fourth quarter 2014. With in-sample analyses and pseudo out-of-sample exercises, we find that for most of the European states considered, the survey-based indicator model outperforms common benchmark models. It is therefore a powerful tool for generating more accurate employment forecasts. We observe the best results for one quarter ahead predictions that are primarily the aim of the survey question. However, employment expectations also work well for longer forecast horizons in some countries.

Zusammenfassung

In diesem Papier evaluieren wir für 15 europäische Staaten, inwiefern Beschäftigungserwartungen die Prognosegüte des Erwerbstätigenwachstums verbessert. Unser Beobachtungszeitraum beginnt mit dem ersten Quartal 1998 und endet mit dem vierten Quartal 2014. Mit In-Sample- und Out-of-Sample-Analysen, können wir zeigen, dass unser befragungsbasiertes Indikatormodell den üblichen Benchmark-Modellen überlegen ist. Deshalb liefern Befragungsdaten auch gute Indikatoren für die Erstellung genauerer Beschäftigungsprognosen. Die größten Verbesserungen konnten wir für Prognosen ein Quartal im Voraus beobachten, die genau Zielhorizont der Befragung sind. Die Beschäftigungserwartungen sind in einigen Ländern aber auch in der Lage das Erwerbstätigenwachstum in der längeren Frist vorherzusagen.

JEL classification: E27, J00, J49

Keywords: employment forecasting, European business survey, employment expectations, Granger causality

Acknowledgements: Our special thanks goes to Alfred Garloff and Marcel Thum for their careful reading of our manuscript. We are grateful to Wolfgang Nagl, Michael Weber and seminar participants at the Technische Universität Dresden for helpful comments and suggestions; and to Lisa Giani Contini for editing this text. We also thank the participants at the meeting of the Regional Research Network from the Institute for Employment Research in November 2013. Comments from the 1st CGDE Doctoral Workshop are also gratefully acknowledged. An earlier version of this paper has been published as *Ifo Working Paper No. 182* and is available free of charge under *https://ideas.repec.org/s/ces/ifowps.html*.

1 Introduction

Business and consumer surveys have become a widely accepted source in the field of macroeconomic forecasting. Because of their rapid availability, qualitative survey results are useful tools for the assessment of the economic constitution. Several groups, such as politicians, employers and researchers, are interested in early information about the course of an economy that is not available from secondary data bases. Additionally and in contrast to such quantitative data, survey results do not suffer from major revisions, which make them a powerful tool for economic forecasting.

There is a large body of literature dealing with the forecasting performance of survey-based indicators. Whereas most studies evaluate the predictive power of these indicators for fairly standard economic variables, such as gross domestic product (see, e.g., Hansson/Jansson/Löf, 2005; Abberger, 2007a), industrial production (see, e.g., Hanssens/Vanden Abeele, 1987; Fritsche/Stephan, 2002; Croux/Dekimpe/Lemmens, 2005) or inflation (see, e.g., Ang/Bekaert/Wei, 2007), analyses for labor market variables are scarce. We fill this gap with our paper. We use qualitative information from the *Joint Harmonised EU Programme of Business and Consumer Surveys*, i.e., employment expectations for three months ahead (*EEXP*), to forecast employment growth[1] on a quarterly basis for 15 European states in the period from 1998Q1 to 2014Q4. We test the forecasting performance of *EEXP* with Granger causality tests and pseudo out-of-sample exercises for every country considered in this study. The results show that for most of the countries, *EEXP* is an efficient indicator for forecasting employment growth in the short-term (one quarter ahead). Despite the fact that employment expectations can be seen as a short-term indicator, we also test the forecasting performance for longer horizons up to four quarters. As expected, the indicator loses its power with increasing forecast horizons. However, for some countries (for example Belgium, Estonia and France) a model including the indicator also significantly beats the benchmark in the long run. Especially for Bulgaria and Hungary, a model that includes *EEXP* has no higher forecast accuracy in comparison to our chosen benchmark models, no matter the forecast horizon considered.

We contribute to the existing literature in several ways. Firstly, we systematically analyze the forecasting performance of a survey-based leading indicator (employment expectations) for employment growth. Interestingly, only five studies have addressed the forecasting properties of similar survey-based qualitative indicators for employment growth of single states. For Canada, an early attempt is the study by Hartle (1958). He used data from the Employment Forecast Survey, where industrial establishments in Canada were asked to forecast their own future employment for the next three and six months and then studies whether it is possible to forecast employment for the Canadian industrial sector more accurately with these firm-specific forecasts. He concluded that these survey results are not able to provide reliable forecasts for employment change in the Canadian industry. The study of dos Santos (2003) examined the relationship between a large amount of qualitative indicators (among

[1] We have to mention that studies that evaluate survey results for the prediction of the unemployment rate exist (see, e.g., Claveria/Pons/Ramos, 2007; Österholm, 2010; Hutter/Weber, 2015; Martinsen/Ravazzolo/Wulfsberg, 2014). Survey results help to improve unemployment rate forecasts especially in the short-term.

them employment expectations) and several different macroeconomic variables for Portugal. Through a cross-correlation analysis, employment expectations were statistically associated with the annual growth rate of employment in some sectors (e.g., the industrial sector) with a lead of up to two quarters. More recent studies are those from Abberger (2007b) for Germany and Siliverstovs (2013) as well as Graff/Mannino/Siegenthaler (2012) for Switzerland. Abberger (2007b) analyzed whether employment expectations from the monthly Ifo business survey in Germany (Ifo Employment Barometer[2]) can serve as a leading indicator for annual employment changes. Applying three approaches (smoothing techniques, error correction models and probit estimates), he found that the survey-based indicator has a lead of two to four months and is able to date turning points in employment growth. For Switzerland, Siliverstovs (2013) used the KOF Employment Barometer[3] provided by the KOF Swiss Economic Institute to evaluate whether this survey-based indicator improves in-sample and out-of-sample forecast accuracy of Swiss employment. He found that the barometer has predictive power for nowcasts and one quarter ahead predictions. The study by Graff/Mannino/Siegenthaler (2012) confirmed these results by showing that the KOF Employment Barometer as well as a survey-based indicator obtained by the Federal Statistical Office of Switzerland are able to predict employment one quarter ahead. With the exception of Hartle (1958), all the other studies found an improvement in the accuracy of employment forecasts by using survey results.

The second contribution of our paper is the examination of forecast improvement by employment expectations for a multitude of European states. Most of the studies either analyzed the Euro area as an aggregate (see Claveria/Pons/Ramos, 2007) or just one single state (see Hansson/Jansson/Löf, 2005; Österholm, 2010; Martinsen/Ravazzolo/Wulfsberg, 2014: for Sweden or Norway). Only the study of Croux/Dekimpe/Lemmens (2005) analyzed the capability of production expectations in forecasting industrial production for 12 European states. We add to the existing literature by studying the predictive content of employment expectations for employment growth in 15 European countries separately. To the best of our knowledge, this has not been documented in the literature.

Our third contribution is that we do not focus on one single sector (e.g., industry) when forecasting employment. This paper uses the survey results from the industrial sector, construction and retail trade together to evaluate the forecasting power for total employment growth. Both in-sample (Granger causality) and out-of-sample properties (root mean squared forecast errors in comparison to benchmark models) are discussed in our analyses.

[2] The data are periodically updated and available at *http://www.cesifo-group.de/ifoHome/facts/Time-series-and-Diagrams/Zeitreihen/Reihen-Beschaeftigungsbarometer.html*.
[3] A description and new press releases can be found at *http://www.kof.ethz.ch/en/surveys*.

Fourthly, we add to the existing literature on survey results in giving a deeper understanding on how qualitative information work for macroeconomic forecasting or rather for one specific macroeconomic variable. As it was stated in Croux/Dekimpe/Lemmens (2005), business tendency surveys are expensive as well as time-consuming. In order to justify the different questions of this time-consuming and expensive survey, the results should have some predictive power for macroeconomic variables. Since the results for several macroeconomic aggregates are mixed, Claveria/Pons/Ramos (2007) concluded that it is not clear why some indicators are able to predict specific macroeconomic variables whereas others cannot. Our paper adds to this discussion by evaluating the forecast performance of employment expectations so that we are able to explain a piece of this apparent puzzle.

The paper is organized as follows. In Section 2, we present our data and the empirical setup along with some descriptive statistics as well as statistic properties of the data. By using in-sample approaches and out-of-sample methods, Section 3 presents our findings in detail, provides a robustness check and discusses the results. Section 4 concludes.

2 Data and empirical setup

2.1 Data

The European Commission collects monthly survey results within their *Joint Harmonised EU Programme of Business and Consumer Surveys* for a multitude of European states. This program is harmonized across the countries in terms of questions and methods. It comprises establishments from different sectors (industry, construction, retail trade and services).[4] We excluded the results from the service sector because the time series is too short for our purpose. In the end, we used qualitative information from the industrial sector, construction and retail trade for the period from January 1998 to December 2014. Due to some further data restrictions (e.g., missing employment data), we eliminated some countries so that the following European states remain in our sample: Austria, Belgium, Bulgaria, the Czech Republic, Estonia, Finland, France, Germany, Hungary, Italy, the Netherlands, Portugal, Slovakia, Sweden and the United Kingdom. These 15 states cover more than 82% of gross domestic product and 75% of EU-28-employment in 2014.

For our analysis, the question of interest is: "How do you expect your firm's employment to change over the next 3 months?" (i.e., employment expectations [*EEXP*]). The respondents have three possibilities to answer this question: (+) increase, (=) remain unchanged and (-) decrease. In line with the literature, we assessed the forecasting power of "balances" (for a critical discussion, see Croux/Dekimpe/Lemmens, 2005; Claveria/Pons/Ramos, 2007: and the references therein). These balances are expressed as differences between the weighted share of firms whose employment will increase and the weighted share of those who expect that their total employment level will decrease. The weights, therefore, are based on the

[4] The aim of the European Commission is to keep the sample representative for each month. To ensure this, sample updates are necessary on occasion due to, e.g., start-ups or bankruptcies. However, the samples for the business survey are very stable in each state. Additional details on the sample composition can be found in European Commission (2007).

size of the firms (see European Commission, 2007). All firms with a response "remain unchanged" are not considered.

Since our target variable (employment) is not available on a monthly basis, we had to transform the balances into quarterly data. To obtain quarterly survey results, we calculated a three-month average (*EEXP_av*). To verify our results, we additionally used the third month of each quarter (*EEXP_tm*) as one possible robustness check. All the survey results are provided with or without seasonal adjustment. In line with the literature, we chose seasonal adjusted data. In order to summarize the balances from the three different sectors (industry, construction and retail trade) to one single indicator, we applied time-varying weights obtained from quarterly employment figures for every sector and single state.

As already mentioned, our variable of interest is the development of employment in 15 EU countries. A source for comprehensive quarterly employment figures are national accounts of single states. Eurostat makes these data available for all member states of the European Union plus Norway. In addition to the total sum of employment, data for 10 branches of the economy are provided.[5] All the data are seasonally adjusted and transformed into quarter-on-quarter (qoq) growth rates. This transformation is very suitable because the firms were asked about their employment development within the next three months.

In addition to total employment (*EMP*), we also use employment in those sectors that are directly addressed by the survey results (*MEMP*). Hence, our second variable *MEMP* is the difference of total employment minus agriculture, forestry, fishing and advanced services.[6] The variable *MEMP* comprises almost 50% of the total employment for all states in this sample and therefore still a large part of the private sector economy. To summarize, we analyze the forecast accuracy of employment expectations for employment growth (either total or a sub-sample of sectors) on a quarterly basis from 1998Q1 to 2014Q4 for 15 European states.

2.2 Descriptive results

To illustrate the structure and development, Figure 1 shows the qoq growth rate of *EMP* as well as employment expectations *EEXP_av* for each of the 15 European states.[7] On the left y-axis, *EMP* (gray bars) is displayed, while the right y-axis illustrates *EEXP_av*, shown as a black and continuous line. The x-axis represents the sample period (1998Q1 to 2014Q4). Employment expectations seem to serve as an indicator for predicting employment growth, but a high heterogeneity exists between the considered European states. *EEXP_av* seems to be a good predictor for the Scandinavian states (Finland and Sweden) and large European

[5] The code of the corresponding time series is: *namq_10_a10_e*. All the data can be downloaded free of charge under *http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home*. The data used in this paper were downloaded on May 15, 2015. To keep our analysis as up-to-date as possible, we use employment figures based on the new European System of National and Regional Accounts 2010 (ESA 2010).

[6] Advanced services comprise the sectors information and communication, financial services, real estate, scientific and administrative services, public administration as well as arts and other service activities. For more details on the specific sectors, see Eurostat (2008).

[7] Table A.1 in the Appendix shows the typical descriptive results for all considered series.

economies, such as France and Germany; for Bulgaria or Hungary one can see completely different movements in *EEXP_av* and *EMP*.

To underpin our idea that in most cases *EEXP_av* could serve as a predictor for *EMP*, we first examined standard cross-correlations between the two variables. We calculated simple correlation coefficients for all European states in our sample by holding *EMP* fixed and applied a lag or lead to *EEXP_av* by four quarters (see Figure 2). To compare the results across all states, the pictures have identical scales for the y-axis (correlation coefficient). In addition, we highlight the correlation coefficient observed by lag one since the question of *EEXP_av* aims at a leading character of the indicator by one quarter. Only in the case of Finland we observe the highest correlation coefficient by lagging employment expectations by one quarter. Mostly, the highest correlation coefficients between *EEXP_av* and *EMP*, with the exception of Austria, Finland, France, Germany, Italy and Portugal, were observed for the contemporaneous values of the two series. One possible explanation could be the aggregation from monthly to quarterly data. Moreover, this is not a problem at all since it shows that the indicator could also be used for nowcasts as well. Additionally, the still large correlation coefficients for higher lags than one quarter suggest that the indicator can also be an adequate predictor for larger forecasting horizons in comparison to the benchmark. Altogether, the *EEXP_av* series has some leading characteristics for *EMP*.[8] We observe the strongest linear relationship between *EMP* and the first lag of *EEXP_av* for the Netherlands, followed by Belgium and the Czech Republic. The weakest relation is found for Hungary, Bulgaria and Slovakia. One would have expected this from Figure 1. To sum up, the correlation analyses also show that *EEXP_av* could serve as a potential indicator for predicting *EMP* in most of the countries and possibly for forecasting horizons longer than one quarter.

2.3 Empirical setting

Next to the simple analyses of correlation coefficients, potential leading characteristics of the single employment expectation series has to be ensured with more elaborate methods. As a first step, we applied standard Granger causality tests (in-sample analysis) to check whether *EEXP* is basically helpful to describe employment growth. To check the forecasting performance of *EEXP*, we present pseudo out-of-sample exercises in a second step.

[8] The same holds for the series *EEXP_tm*.

2.3.1 In-sample analyses

A necessary condition to test Granger causality is stationarity of the time series considered. To give a broad and reliable picture of stationarity, we applied two different tests: the Ng-Perron (NP) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. Whereas the NP test states a unit root under the null, the KPSS test is applied against stationarity. Whenever a series has no unit root, the NP test should reject the null. Since the KPSS test is a test on stationarity, it should not reject the null hypothesis. We chose these two different tests for three reasons: Firstly, the widely used Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test have different properties within finite samples, thus they can produce misleading results. Secondly, the ADF test and the PP test also have low power against $I(0)$ alternatives when the series is close to $I(1)$. In such cases the NP test performs much better (Ng/Perron, 2001). In other words, the NP test is much more accurate when time series are nearly integrated of order one. Thirdly, the KPSS test is applied since it is a stationarity test instead of a unit root test. It proposes a stationary time series under the null hypothesis and is therefore a complement against the NP test. With the KPSS and NP tests, we can distinguish between series that are stationary and series that contain a unit root.

In the first step, the NP test and the KPSS test are applied to the levels of the series (qoq growth rates or balances). We performed the tests in two ways: (i) only with a constant and (ii) with a constant and a linear trend. Whenever a series is not stationary in levels, we tested the first differences in a second step.

Table 1: Results of the unit root tests for *EMP* and *EEXP_av*

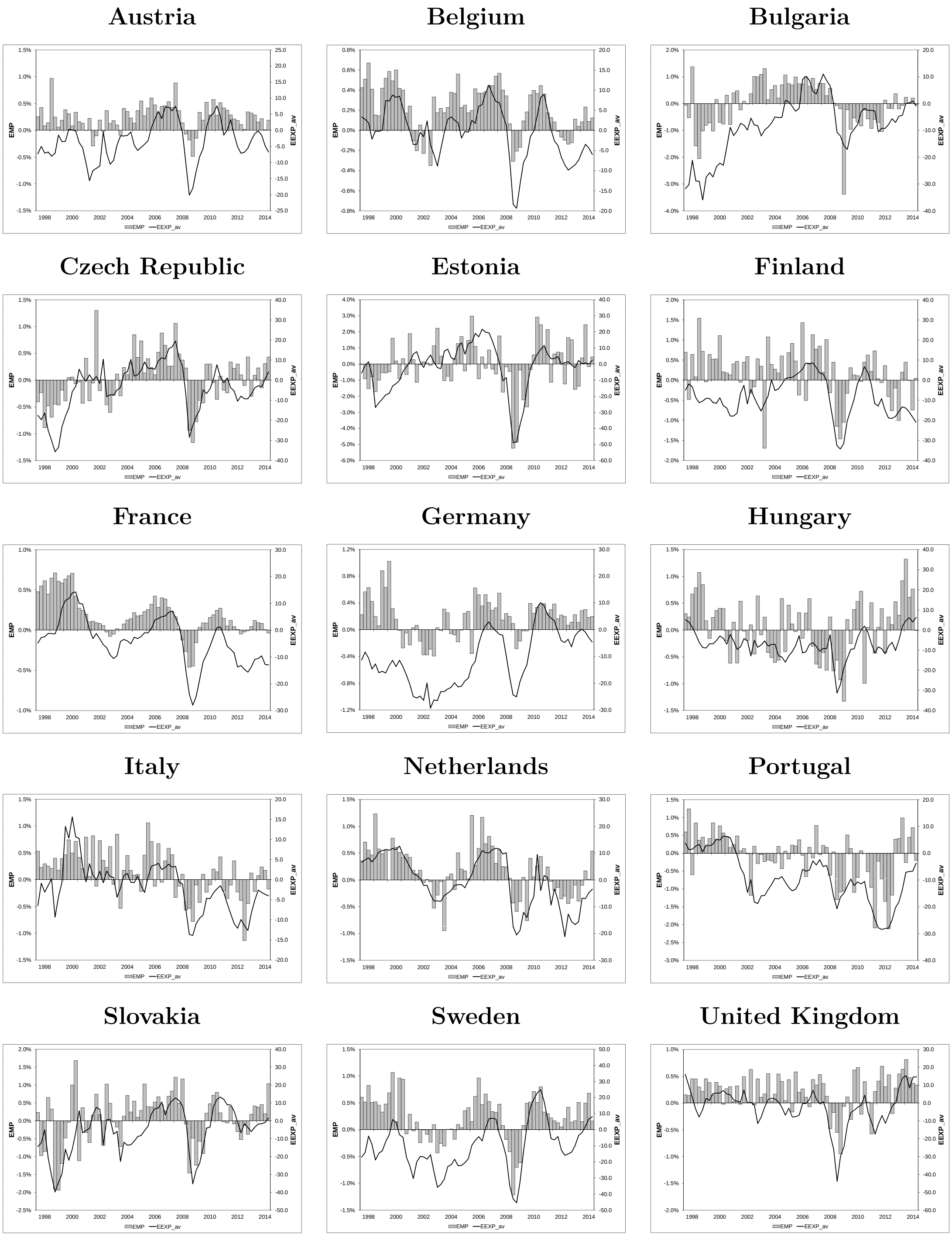
Country	**EMP**					**EEXP_av**				
	KPSS		NP		Δ	KPSS		NP		Δ
	Const.	Trend	Const.	Trend		Const.	Trend	Const.	Trend	
Austria			***	***				***	***	
Belgium			***	***				***	***	
Bulgaria		**	*		X	**	**			X
Czech Republic		*	***	***			**			X
Estonia			***	***				***	*	
Finland	*		***	***				***	**	
France	**		*	*	X	**		***	***	
Germany			***	**		**		***	***	
Hungary		**	***	*			**	*		X
Italy	***		**	***		**		**		X
Netherlands			***	***		**		*		X
Portugal	**		***	***		**		*	***	
Slovakia			***	***		*		**		X
Sweden			***	***				***	***	
United Kingdom		*	***	***			*			X

Note: Calculations are based on the whole sample (1998Q1–2014Q4). The Ng-Perron (NP) test states a unit root under the null hypothesis. The Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) has stationarity under the null hypothesis. For the KPSS and the NP tests, only test statistics and critical values are available. We therefore use asterisks to show whether the null can be rejected or not. In all cases we first tested stationarity in levels (qoq qrowth rate or balances). If the levels turned out to be non-stationary, we then tested first differences of the variables. The column Δ presents the decision on the transformation of the variables. An **X** indicates that first differences are applied. ***, ** and * indicate the rejection of the null hypothesis at the 1%, 5% and 10% significance level.

Table 1 presents the unit root test results for *EMP* and *EEXP_av*.[9] and shows that most of the series are stationary in levels, although the results of the tests are not always consistent

[9] The results of the unit root tests for *MEMP* and *EEXP_tm* can be found in Table A.2 in the Appendix. For these two variables, the stationarity tests yield relatively similar results as for *EMP* and *EEXP_av*.

Figure 1: Development of *EMP* and *EEXP_av* for each European state



Source: European Commission, Eurostat, author's calculations and illustrations.

Figure 2: Cross-correlations between *EMP* and *EEXP_av* for each European state

Note: Calculations are based on the whole sample (1998Q1–2014Q4).

Source: European Commission, Eurostat, author's calculations and illustrations.

with each other. In such cases, we decided whether the series is stationary or not by all means with the NP test, because the KPSS test could suffer from a finite sample bias. Table 1 also shows which series have to be transformed. This is indicated by an X in column Δ. After the transformation into first differences, all series are stationary.

Granger causality is commonly used to show whether an indicator has some leading characteristics for a specific target variable.[10] It is also possible to check whether feedback effects between the two series are present (Granger, 1969). This is the case whenever an indicator variable explains the target variable and vice versa. In the worst case, the target variable has a leading character for the indicator and not the other way around (reverse Granger causality). To test for (reverse) Granger causality, we estimated the following two equations:

$$y_t = \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j x_{t-j} + \varepsilon_{1,t} \ , \tag{1}$$

$$x_t = \sum_{j=1}^{q} \gamma_j x_{t-j} + \sum_{i=1}^{p} \delta_i y_{t-i} + \varepsilon_{2,t} \ . \tag{2}$$

The qoq growth rate of either *EMP* or *MEMP* is denoted with y_t. Employment expectations, either as the three-month average (*EEXP_av*) or the last value of the quarter (*EEXP_tm*), are defined as x_t. We allow a maximum of four lags for p and q in Equations (1) and (2).

We first tested whether all four lags of the indicator (x_t) have a significant effect on the target variable y_t. Under the null hypothesis "employment expectations (x_t) do NOT Granger cause employment growth (y_t)". If the null is rejected, then *EEXP* is able to explain our variable of interest (*EMP*, *MEMP*). Then, in a second step, the reverse way is tested with the null hypothesis "employment growth does NOT Granger cause employment expectations". If this hypothesis is rejected, then *EMP* or *MEMP* can explain *EEXP*. From the Granger causality tests, four different cases emerge: (i) *EEXP* only Granger causes employment growth, (ii) there are feedback effects between the two series, (iii) *EMP* or *MEMP* only Granger causes *EEXP* and (iv) there is no relationship. As already mentioned, the third case is the worst one. Whenever case (iii) occurs, employment expectations are probably not a suitable predictor for employment growth. The same holds for the fourth event. In case (i) and (ii), *EEXP* can probably be used as an indicator for forecasting employment growth, i.e., y can be better forecasted with the additional information of x. It is well known that the Granger concept has some weaknesses (see, e.g., Lütkepohl, 2005). It is often argued that data transformation, such as first differences or the elimination of a trend, go along with a loss of information. This loss causes the Granger concept not to be able to distinguish between long-term and short-term relationships between variables. However, we are not interested in long-term or short-term movements between series but rather to check the survey-based indicator's ability to forecast employment. Another weakness of Granger causality originates from the specification of Equations (1) and (2): The results of the Granger causality tests may be sensitive due to the maximum lag length of p and q. We tested different specifications of p and q with fairly robust results. We also checked the

[10] So it is by no means a test on causality between two variables or on the exogeneity of a series.

necessary assumptions (e.g., homoscedasticity or no autocorrelation) to estimate the models in Equations (1) and (2) and found that these assumptions are predominantly fulfilled. As our focus is not on short-term or long-term relationships and as we debilitate the second main criticism, the Granger concept seems to be an adequate approach for our purpose.

2.3.2 Out-of-sample examination

Forecast model

To generate our pseudo out-of-sample forecasts, we employ an autoregressive distributed lag (ADL) model

$$y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=1}^{q} \gamma_j x_{t+1-j} + \varepsilon_t \, , \tag{3}$$

where y_{t+h} is the h-step ahead forecast of either *EMP* or *MEMP* and x_t represents the employment expectations *EEXP*. The forecast horizon h is defined in the range of $h \in \{1,$ 2, 3, 4$\}$ quarters. We allow, as in the in-sample analyses, a maximum of four lags for our target variable (p) and the employment expectations (q). The optimal lag length is determined by the Bayesian Information Criterion. Robinzonov/Wohlrabe (2010) showed for Germany that choosing either a recursive approach or a rolling window can lead to different forecasting results. Thus, we generated our forecast in both ways. The initial estimation period for Equation (3) ranges from 1998Q1 to 2004Q4 ($T_E = 28$). The period is then successively moved forward by one quarter with a new specification of the model in each step. The expanding window approach serves as a robustness check for our results obtained from the rolling approach.[11] It uses an initial window, which is successively enlarged by one quarter in each step; the first forecast for y_t is calculated for 2005Q1 and the last for 2014Q4. To avoid a prediction of the indicator x_t or the dependent variable y_t itself, we implement the ADL model in a direct-step fashion. This means that y_{t+h} is directly explained with lagged values of the dependent variable and the indicator. This results in the same number of forecasts ($T_F = 40$) for every forecast horizon h. More details on direct-step forecasting can be found in Robinzonov/Wohlrabe (2010). As the benchmark model, we chose a common AR(p) process.

Forecast evaluation

To evaluate the forecast accuracy of our different models, we first have to calculate forecast errors from our exercises. Let $\widehat{y}_{t+h}$ denote the h-step ahead forecast produced at time t, then the resulting forecast error is defined as $FE_{t+h} = y_{t+h} - \widehat{y}_{t+h}$. The corresponding forecast error of our benchmark model is FE^{ARp}_{t+h}. To assess the performance of an indicator-based model, we calculate the root mean squared forecast error (RMSFE) as the loss function.

[11] Whenever breaks in the time series are present, the rolling window approach is preferable. An expanding window is suitable when there are no breaks in the series or the whole cyclicality of the series should be captured. The recursive approach then leads to more precise estimates of the parameters (Weber/Zika, 2013).

For the h-step ahead indicator-based forecast, the RMSFE is

$$RMSFE_h = \sqrt{\frac{1}{T_F}\sum_{n=1}^{T_F}\left(FE_{t+h,n}\right)^2} \,. \tag{4}$$

The RMSFE for the benchmark model is $RMSFE_h^{ARp}$. To decide whether employment expectations perform, on average, better than the autoregressive process, we calculate the relative RMSFE between the indicator model and the benchmark

$$rRMSFE_h = \frac{RMSFE_h}{RMSFE_h^{ARp}} \,. \tag{5}$$

Whenever this ratio is smaller than one, the indicator-based model performs better than the benchmark. Otherwise, the AR(p) process is preferable. Nonetheless, calculating this ratio does not clarify whether the forecast errors of the indicator-based model and the benchmark are statistically different from each other. To check this, we apply a test such as the one proposed by Diebold/Mariano (1995). Under the null hypothesis, the test states that the expected difference in the MSFE equals zero. With our notation, this gives

$$H_0 : E\left[\left(FE_{t+h}^{ARp}\right)^2 - \left(FE_{t+h}\right)^2\right] = E\left[MSFE_{t+h}^{ARp} - MSFE_{t+h}\right] = 0 \,. \tag{6}$$

In other words, the null hypothesis states that the AR(p) is the data generating process. Adding an indicator to this process can then cause a typical problem of nested models. The larger model – with our survey-based indicator – therefore introduces a bias through estimating model parameters that are zero within the population. Thus, the AR(p) process nests the indicator model by setting the parameters of the indicator to zero. As stated by Clark/West (2007), this causes the MSFE of the larger model to be biased upwards since redundant parameters have to be estimated. As a result, standard tests, such as the one proposed by Diebold/Mariano (1995), loose their power. On this account, we follow the literature (e.g., Weber/Zika, 2013) and apply the adjusted test statistic by Clark/West (2007)

$$CW_h = \sqrt{\frac{1}{\widehat{V}(a_{t+h})T_F}} \sum_{t=1}^{T_F}\left(\underbrace{MSFE_{t+h}^{ARp} - \left[MSFE_{t+h} - \left(\widehat{y}_{t+h} - \widehat{y}_{t+h}^{ARp}\right)^2\right]}_{a_{t+h}}\right) \,, \tag{7}$$

with $\widehat{V}(a_{t+h})$ as the sample variance of a_{t+h} and $\left(\widehat{y}_{t+h} - \widehat{y}_{t+h}^{ARp}\right)^2$ as the adjustment term. After this adjustment, standard critical values from the Student's t-distribution with $T_F - 1$ degrees of freedom can be used to decide whether forecast errors are statistically significant from each other.

3 Results

The figures and cross-correlations in Section 2.2 have shown that employment expectations in most of the countries could serve as a potential indicator for predicting employment. Only for a few of the observed countries (e.g., Bulgaria or Hungary) was a leading character of *EEXP_av* unlikely. In order to analyze the forecasting performance of employment expectations, the following two subsections present the results of our in-sample and out-of-sample analyses. We discuss the forecasting performance of *EEXP_av* for *EMP* and *MEMP*. Tables with results for *EEXP_tm* can be found in the Appendix.

3.1 Results of the in-sample analyses

Do employment expectations deliver some additional information to forecast employment growth? In most of the cases, they do. The results for *EMP* are shown in the upper part of Table 2. The lower part of the table comprises the Granger causality results for *MEMP*. Column two shows the test results for Granger causality from employment expectations to employment growth (*EEXP_av* $\rightarrow$ *EMP* or *MEMP*). Column three presents this information in reverse (*MEMP* or *EMP* $\rightarrow$ *EEXP_av*). All numbers represent p-values. The last column shows whether the indicator has leading characteristics (+), whether feedback effects between the two series are present (FB) or whether the indicator has no predictive content or there is no relationship at all (X).

In most of the countries considered, employment expectations serve as a leading indicator. For *EMP*, *EEXP_av* probably does not deliver additional information in Bulgaria, Germany, Hungary and Italy. For Germany, Hungary and Italy, no relationship between the two series seems to exists, whereas for Bulgaria, employment growth serves as an indicator for *EEXP_-av*. Feedback effects are evident for four countries (Austria, the Czech Republic, France and the Netherlands).

For *MEMP*, the results are superior to those for *EMP*. In more cases compared to *EMP*, employment expectations have additional information to forecast a sub-sample of total employment. Overall, the results changed for Germany, Italy and Slovakia. There, employment expectations provide additional information and improve the forecasting for the sub-sample of total employment growth (*MEMP*). For Bulgaria and Hungary, the results for *MEMP* are in line with those for *EMP*. In contrast, feedback effects are now present for Estonia.[12]

To summarize, the in-sample analyses revealed potential forecasting information from employment expectations for most of the countries in our sample. However, we found differences between our two potential target variables *EMP* and *MEMP*. We suggest that it is important which sectors are asked as possibly our indicators is less able to reproduce the cyclical movement of total employment. Adding employment expectations of the service sector to our indicator may solve this puzzle. However, as argued in Section 2.1, the times series for the service sector is too short for our analysis. To examine whether *EEXP_av* produces lower forecast errors than a benchmark model and to underpin the statements from our in-sample analyses, we conducted pseudo out-of sample exercises in the following section.

[12] Turning to *EEXP_tm*, the third month of each quarter as representative serves as a leading indicator as well. Especially for Italy, the third month has explanatory power for *EMP* as well as for *MEMP*. Detailed in-sample results for *EEXP_tm* can be found in Table A.3 in the Appendix.

Table 2: Granger causality results for *EMP*, *MEMP* and *EEXP_av*

Country	EMP		
	EEXP_av → EMP	EMP → EEXP_av	Result
Austria	0.001	0.021	FB
Belgium	0.049	0.253	+
Bulgaria	0.106	0.095	X
Czech Republic	0.068	0.009	FB
Estonia	0.007	0.163	+
Finland	0.000	0.976	+
France	0.001	0.087	FB
Germany	0.298	0.263	X
Hungary	0.435	0.890	X
Italy	0.209	0.386	X
Netherlands	0.018	0.096	FB
Portugal	0.000	0.223	+
Slovakia	0.039	0.120	+
Sweden	0.015	0.886	+
United Kingdom	0.036	0.624	+
Country	**MEMP**		
	EEXP_av → MEMP	MEMP → EEXP_av	Result
Austria	0.000	0.082	FB
Belgium	0.002	0.467	+
Bulgaria	0.170	0.139	X
Czech Republic	0.028	0.027	FB
Estonia	0.024	0.099	FB
Finland	0.001	0.675	+
France	0.021	0.033	FB
Germany	0.007	0.285	+
Hungary	0.768	0.872	X
Italy	0.062	0.314	+
Netherlands	0.001	0.189	+
Portugal	0.000	0.151	+
Slovakia	0.104	0.182	X
Sweden	0.002	0.657	+
United Kingdom	0.008	0.364	+

Note: Calculations are based on the whole sample (1998Q1–2014Q4). The table presents p-values from the Granger causality test. *Acronyms:* +, *EEXP* only Granger causes employment growth (case [i]); FB, feedback effects are present (case [ii]); X, employment growth only Granger causes *EEXP* (case [iii]) or no relationship (case [iv]).

3.2 Results of the pseudo out-of-sample analyses

Are employment expectations able to produce lower forecast errors in comparison to a common benchmark model? The simple answer is yes, for most of the countries in our sample. Table 3 presents the pseudo out-of-sample results for all 15 European states, produced with a rolling window.[13] We divided the table into two parts *EMP* and *MEMP*. Each column represents the forecasting outcome for a specific forecast horizon, ranging from one to four quarters for the both target variables. For each single country, we added the forecasting performance of six different models: (i) AR(p) is the chosen benchmark, (ii) an AR(1) process, (iii) the in-sample mean (ISM), (iv) a Random Walk (RW)[14] and finally (v) and (vi) the outcomes from employment expectations (*EEXP_av*, *EEXP_tm*). Each entry in the rows *AR(p) in %* illustrates one single $RMSFE_h^{ARp}$ of the benchmark model in percentage points. These figures are separated from the other results for each country with

[13] The results from the expanding window are presented in Table A.4 in the Appendix.

[14] The ISM is defined as $y_{t+h} = \overline{y}$, representing the sample average of the estimation window. The Random Walk prediction is simply the last known value of the target variable $y_{t+h} = y_{t-1}$.

dashed lines. The other numbers in Tables 3 and A.4 are the model-specific $rRMSFE_h$, i.e., the RMSFE of each model compared to the benchmark. Asterisks typically denote significant differences between the forecast errors based on the outcome of the Clark-West test.

As expected from descriptive statistics and the in-sample analyses, employment expectations seem to be able to predict employment growth more accurately than a simple benchmark model. Compared to the other three possible benchmarks (AR(1), ISM, and RW) this conclusion holds as well. The best results can be found for short-term forecasts with a forecasting horizon of one quarter ahead ($h = 1$). This is straightforward because the survey-based indicators used here are short-term indicators by construction. Firms were asked to provide a statement about their expected employment development for the next three months. However, some noteworthy exceptions exist. For Belgium, Estonia, Finland and France, the indicator model produced significantly lower forecast errors for longer forecast horizons, i.e., the ADL model is also significantly better than the benchmark for $h > 1$. The best relative performance of employment expectations in terms of $rRMSFE$ in the short term can be found for Finland and Sweden.

Table 3: Out-of-sample results (rolling) for *EMP* and *MEMP*

Model	**EMP**				**MEMP**			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
				Austria				
AR(p) in %	0.228	0.274	0.324	0.392	0.228	0.378	0.434	0.448
AR(1)	0.962	0.936	0.845*	0.762	1.000	0.990	0.887	0.939
ISM	1.132	0.966	0.833*	0.695*	1.112	0.993	0.880*	0.865*
RW	1.152	1.177	1.076	1.005	1.219	1.116	1.063	1.246
EEXP_av	0.977**	1.022	1.390	1.522	0.950***	1.012	1.050	1.261
EEXP_tm	1.021	0.957**	1.313	1.330	1.010	0.978***	1.142	1.225
				Belgium				
AR(p) in %	0.163	0.214	0.285	0.351	0.163	0.301	0.384	0.624
AR(1)	0.863**	0.956	0.880*	0.854*	0.895***	0.949**	0.896**	0.615*
ISM	1.376	1.082	0.832*	0.685**	1.419	1.046	0.839**	0.524*
RW	1.233	1.231	1.096	0.954*	1.282	1.205	1.131	0.757*
EEXP_av	0.941*	0.899*	1.014	1.289	0.860**	1.104	1.076	1.275
EEXP_tm	0.858**	0.958*	0.977*	1.056	0.809***	0.975***	0.935***	1.050
				Bulgaria				
AR(p) in %	0.826	0.801	1.176	1.193	0.826	1.910	1.862	3.900
AR(1)	0.951	1.017	0.866	0.801	0.750	0.933	0.813*	0.367
ISM	0.985	0.995	0.681	0.675	0.789	0.810	0.845*	0.410
RW	1.287	1.332	1.068	0.865*	0.843	0.873	0.830*	0.410
EEXP_av	1.017	1.264	1.191	1.064	0.943	0.966	1.020	0.946
EEXP_tm	1.128	1.196	1.094	1.083	0.880	0.970	1.019	1.031
				Czech Republic				
AR(p) in %	0.432	0.502	0.585	0.669	0.432	0.659	0.718	0.726
AR(1)	0.967	0.947*	0.881**	0.936*	0.958	0.928	0.898**	0.989
ISM	1.174	1.036	0.904*	0.803**	1.126	0.946	0.878*	0.872**
RW	1.201	1.143	1.131	1.014	1.331	1.225	1.201	1.238
EEXP_av	0.969*	1.198	0.970	1.217	0.953*	1.078	0.956	1.006
EEXP_tm	1.109	1.144	1.051	1.042	1.043	1.115	1.182	1.007

Continued on next page...

Table 3: Out-of-sample results (rolling) for *EMP* and *MEMP* – continued

Model	EMP				MEMP			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Estonia								
AR(p) in %	1.772	2.032	1.948	2.181	1.772	3.579	3.528	3.718
AR(1)	0.989*	0.936*	0.997	0.969	0.969	1.002	0.979*	0.909**
ISM	1.052	0.932**	0.979*	0.886**	0.975	0.936	0.947**	0.900**
RW	1.336	1.063	1.200	1.200	1.423	1.258	1.175	1.165
EEXP_av	1.010	0.965**	1.362	1.286	0.934*	1.034	1.215	1.297
EEXP_tm	0.928**	0.959**	1.167	1.156	0.884**	0.990*	1.202	1.083
Finland								
AR(p) in %	0.682	0.663	0.695	0.720	0.682	1.338	1.260	1.227
AR(1)	0.985	0.990	0.935**	0.948*	0.993	0.851	0.938*	0.985*
ISM	0.926*	0.963***	0.928**	0.902**	0.978**	0.830	0.887*	0.915**
RW	1.094	1.253	1.343	1.384	1.259	1.087	1.179	1.179
EEXP_av	0.750**	0.997**	1.252	1.033	0.822**	0.848**	1.525	2.013
EEXP_tm	0.769**	0.994**	1.237	1.035	0.939**	0.963**	1.553	1.638
France								
AR(p) in %	0.107	0.124	0.120	0.130	0.107	0.203	0.261	0.352
AR(1)	0.877	0.855	0.929*	0.847*	1.120	1.001	0.899	0.696
ISM	0.908*	0.795	0.824	0.762*	2.100	1.167	0.931*	0.703*
RW	1.201	1.067	1.192	1.097	1.815	1.233	1.100	0.896
EEXP_av	0.943**	0.942*	1.169	1.160	1.016	1.348	1.700	1.578
EEXP_tm	0.948*	0.925*	1.166	0.991*	0.895***	1.189	1.821	1.295
Germany								
AR(p) in %	0.226	0.238	0.262	0.292	0.226	0.387	0.405	0.438
AR(1)	0.992	1.021	0.991	0.989	0.976*	0.985	1.007	0.998
ISM	1.088	1.055	0.974	0.892*	1.106	1.048	1.019	0.959*
RW	1.169	1.113	1.161	1.135	1.127	1.120	1.145	1.219
EEXP_av	1.086	1.033	1.157	1.231	0.901***	0.856***	1.057	1.126
EEXP_tm	1.002	1.029	0.914**	1.122	0.895***	0.834***	0.932***	0.789***
Hungary								
AR(p) in %	0.599	0.586	0.617	0.644	0.599	1.466	1.405	1.311
AR(1)	0.982	1.000	0.988	1.003	0.977*	0.931	0.987*	1.000
ISM	1.001	1.038	1.000	0.972*	0.962**	0.895	0.938	0.998
RW	1.120	1.196	1.284	1.178	1.265	1.356	0.953**	1.318
EEXP_av	1.075	1.029	1.074	1.147	1.037	0.928	1.023	1.059
EEXP_tm	1.089	1.115	1.215	1.324	1.060	0.996	1.076	1.286
Italy								
AR(p) in %	0.442	0.453	0.512	0.536	0.442	0.579	0.580	0.600
AR(1)	0.998	0.985	0.879	0.83*	1.002	0.975	1.013	0.977
ISM	0.980*	0.971*	0.867	0.835*	1.057	0.997	1.008	0.982
RW	1.194	1.238	1.079	1.082	1.214	1.236	1.167	1.160
EEXP_av	0.979*	1.113	0.977**	1.060	1.022	1.005	1.005	0.932***
EEXP_tm	0.931	1.089	1.063	1.128	0.972	1.086	1.100	1.262
Netherlands								
AR(p) in %	0.408	0.397	0.450	0.576	0.408	0.687	0.716	0.656
AR(1)	0.956***	0.942**	0.955**	0.894**	1.003	0.882	0.970	0.987
ISM	1.085	1.134	1.022	0.817*	1.178	0.870	0.841	0.917
RW	0.990**	1.129	1.274	1.019	1.454	1.260	1.138	1.334
EEXP_av	0.988**	1.081	1.023	0.937**	1.068	1.462	1.299	1.135
EEXP_tm	1.278	1.311	1.249	1.565	1.056	0.958***	1.015	1.031

Continued on next page...

Table 3: Out-of-sample results (rolling) for *EMP* and *MEMP* – continued

Model	**EMP**				**MEMP**			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
				Portugal				
AR(p) in %	0.725	0.817	0.855	0.833	0.725	1.260	1.255	1.054
AR(1)	0.973*	0.958	0.969	0.993	0.943	0.892*	0.982	1.139
ISM	1.019	0.918**	0.885*	0.910	1.024	0.875**	0.890**	1.066
RW	1.273	1.216	1.070	1.036	1.146	1.092	1.089	1.319
EEXP_av	1.060	0.909*	1.062	1.173	0.947**	0.771**	1.024	1.476
EEXP_tm	0.944*	0.964	1.105	1.254	0.918***	0.741**	1.130	1.224
				Slovakia				
AR(p) in %	0.536	0.672	0.813	1.237	0.536	0.945	1.008	1.186
AR(1)	1.023	0.996	0.869	0.614	0.983*	0.992	1.013	0.831
ISM	1.198	0.979	0.825*	0.547	1.147	1.005	0.960	0.821
RW	1.163	1.149	1.078	0.745	1.129	1.364	1.367	1.180
EEXP_av	1.081	0.959*	1.207	1.188	0.955*	0.993	1.548	1.423
EEXP_tm	1.413	1.240	1.312	1.228	1.395	1.386	1.866	1.785
				Sweden				
AR(p) in %	0.328	0.397	0.519	0.541	0.328	0.679	0.822	0.835
AR(1)	0.939***	0.978***	0.883**	0.905**	0.943**	0.981	0.881**	0.946**
ISM	1.259	1.067	0.833***	0.810***	1.041	0.980**	0.822**	0.815**
RW	1.235	1.245	1.164	1.203	1.148	1.270	1.139	1.215
EEXP_av	0.854**	1.053	1.254	1.533	0.925**	1.016	1.240	1.926
EEXP_tm	0.810**	1.091	1.096	1.517	0.944**	1.180	1.079	1.982
				United Kingdom				
AR(p) in %	0.439	0.551	0.556	0.472	0.439	0.640	0.775	0.871
AR(1)	0.949*	0.768**	0.777**	0.974	1.034	0.861**	0.714*	0.717
ISM	0.930**	0.750**	0.752**	0.896*	1.052	0.856**	0.714*	0.643*
RW	1.113	0.891**	1.053	1.240	1.239	0.993**	0.972*	0.897*
EEXP_av	0.946**	1.057	0.990	1.036	1.161	1.012	1.089	1.030
EEXP_tm	0.988*	1.020	0.977	1.060	1.197	1.075	1.026	0.940*

Note: Calculations are based on the whole sample (1998Q1–2014Q4). The table presents the relative root mean squared forecast errors (*rRMSFE*) of the different models and the benchmark. The row AR(p) shows the *RMSFE* (in %) for the benchmark model. *ISM*, in-sample mean; *RW*, Random Walk. Asterisks show significant differences between forecast errors due to the Clark-West test. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

What about the countries for which the in-sample analyses suggest that *EEXP* does not serve as an indicator to predict employment?[15] For *EMP*, the indicator model has a lower forecast accuracy than the benchmark in Bulgaria, Germany and Hungary. For these countries we find for almost all forecast horizons that $rRMSFE$ is larger than one. The opposite holds for Italy, where a forecasting model with employment expectations is able to produce lower forecasts errors than the benchmark. For the most part, the in-sample analyses have indicated correct out-of-sample performance. From Table 3 we additionally can conclude that there are differences between the forecasting performance of *EEXP_av* and *EEXP_tm*

[15] One would argue that adding an indicator and therefore getting a better in-sample fit for the data has to result in a better out-of-sample performance. This may not be the case (see Chatfield, 1995). Overfitting the model or parameter instabilities (see Rossi/Sekhposyan, 2011) are some explanations why in-sample and out-of-sample performance may differ.

for Austria, the Czech Republic, Estonia, the Netherlands and Portugal. On the one hand it is possible that the firms in those countries mainly decide on their future employment figures at the end of each quarter, thus, resulting in a better forecasting performance of *EEXP_tm*. On the other hand noisy signals could be introduced in *EEXP_av*, because of using a three month average instead of an appointed date. However, this trade-off has to be investigated in future research activities in detail.

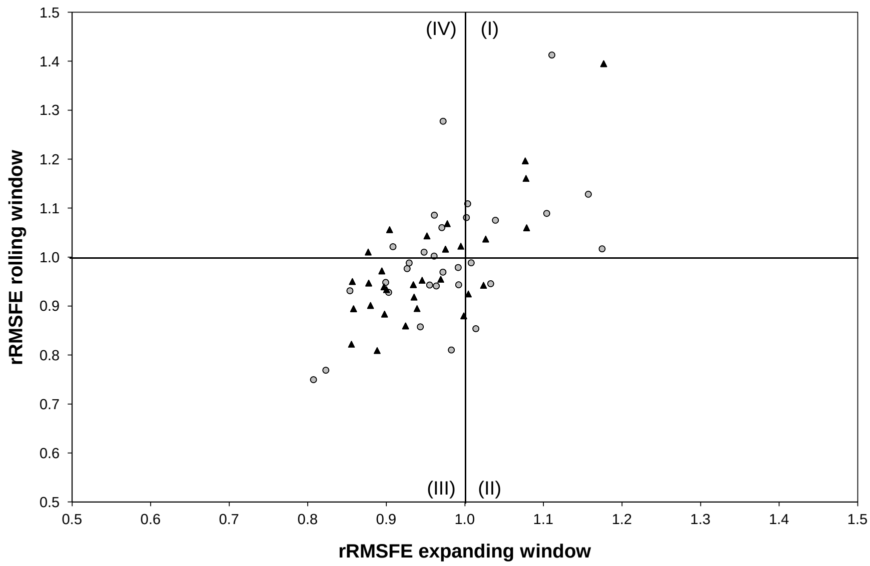
Turning to *MEMP*, we find tremendous changes in the forecasting performance of employment expectations for six countries. For Germany and Portugal, employment expectations are able to produce lower forecast errors than the benchmark if we forecast *MEMP* instead of *EMP*. Especially in the German case, the relative performance of employment expectations increases significantly between the two target variables. Employment expectations lose their relative performance partly or completely for France, Italy, the Netherlands and the United Kingdom. For those countries, our indicator produces lower forecast errors than the benchmark by forecasting total employment instead of *MEMP*. The performance between *EMP* and *MEMP* is very similar for the remaining countries, which is promising since the rapid availability of a firm's employment expectations provides a rich source to forecast either total employment or the sector aggregate.

3.3 Robustness check

Instead of using a rolling window, we applied an expanding window approach to verify the results and conducted a serious robustness check (see Table A.4 in the Appendix for detailed results). Figure 3 shows a comparison of the relative forecasting performance ($rRMSFE$) between the expanding and the rolling window approach. Since employment expectations mainly operate as short-term indicators, we only present the robustness check for the shortest forecast horizon ($h = 1$).[16] The *rRMSFE* for *EEXP_av* and *EEXP_tm* from the rolling window approach are drawn on the y-axis. The corresponding *rRMSFE* from the expanding window approach can be found on the x-axis. Since we have two target variables, the results for *EMP* are shown in gray circles and those for *MEMP* are displayed by black triangles. Each dot (triangle) represents an x-y-pair of either *EEXP_av* or *EEXP_tm* for a specific country. To facilitate interpretation, we add a horizontal and a vertical line that cross the value of the *rRMSFE* of one, thus, indicating whether an indicator performs better or worse than the benchmark model. The interpretations for the resulting quadrants (I) and (III) are straightforward. A dot (triangle) showing up in quadrant (I) indicates that employment expectations produce higher forecast errors compared to the benchmark in the rolling window, as well as the expanding window approach. The opposite holds for quadrant (III), thus producing lower forecast errors than the autoregressive process. Whenever a dot (triangle) lies in quadrant (II), the forecasting performance of employment expectations becomes worse in an expanding window compared to the rolling window approach. For quadrant (IV) the indicator beats the benchmark in an expanding window setup, but not in a rolling one.

[16] The results for longer forecast horizons are available upon request.

Figure 3: Relative forecast errors in expanding vs. rolling windows ($h = 1$)



Note: Calculations are based on the whole sample (1998Q1–2014Q4). Gray circles indicate the relative forecast errors for total employment (*EMP*). The relative forecasting performance for *MEMP* is represented by black triangles.

As shown by Figure 3, the results for most countries remain fairly robust in both the rolling and the expanding window approach. However, 28% of all forecasts, either for *EMP* or for *MEMP*, change their relative performance, thus lying in quadrant (II) or mostly in (IV). But a large share of those forecasts improve if an expanding window is applied.

We also observe some country differences. For *EMP* the relative performance of employment expectations changes for Germany and the United Kingdom. In the German case, employment expectations beat the benchmark model under an expanding window approach, whereas they are not able to do so under a rolling window approach. The opposite holds for the United Kingdom, where the relative performance of employment expectations deteriorates if an expanding window is applied compared to the rolling case. For *MEMP* the most remarkable differences can be found for the Netherlands. Here, *EEXP* beats the benchmark model under an expanding window, but does not improve on average by applying a rolling window. One possible explanation is that the rolling window is not able to capture all the cyclicality in the target series for these countries, meaning that the expanding window approach produces superior parameter estimates. To summarize, employment expectations serve as an indicator for predicting short-term employment for most of the European states in our sample. The only exceptions are Bulgaria, Germany and Hungary. In the German case, however, the results change tremendously if an expanding window is applied.

3.4 Discussion of the results

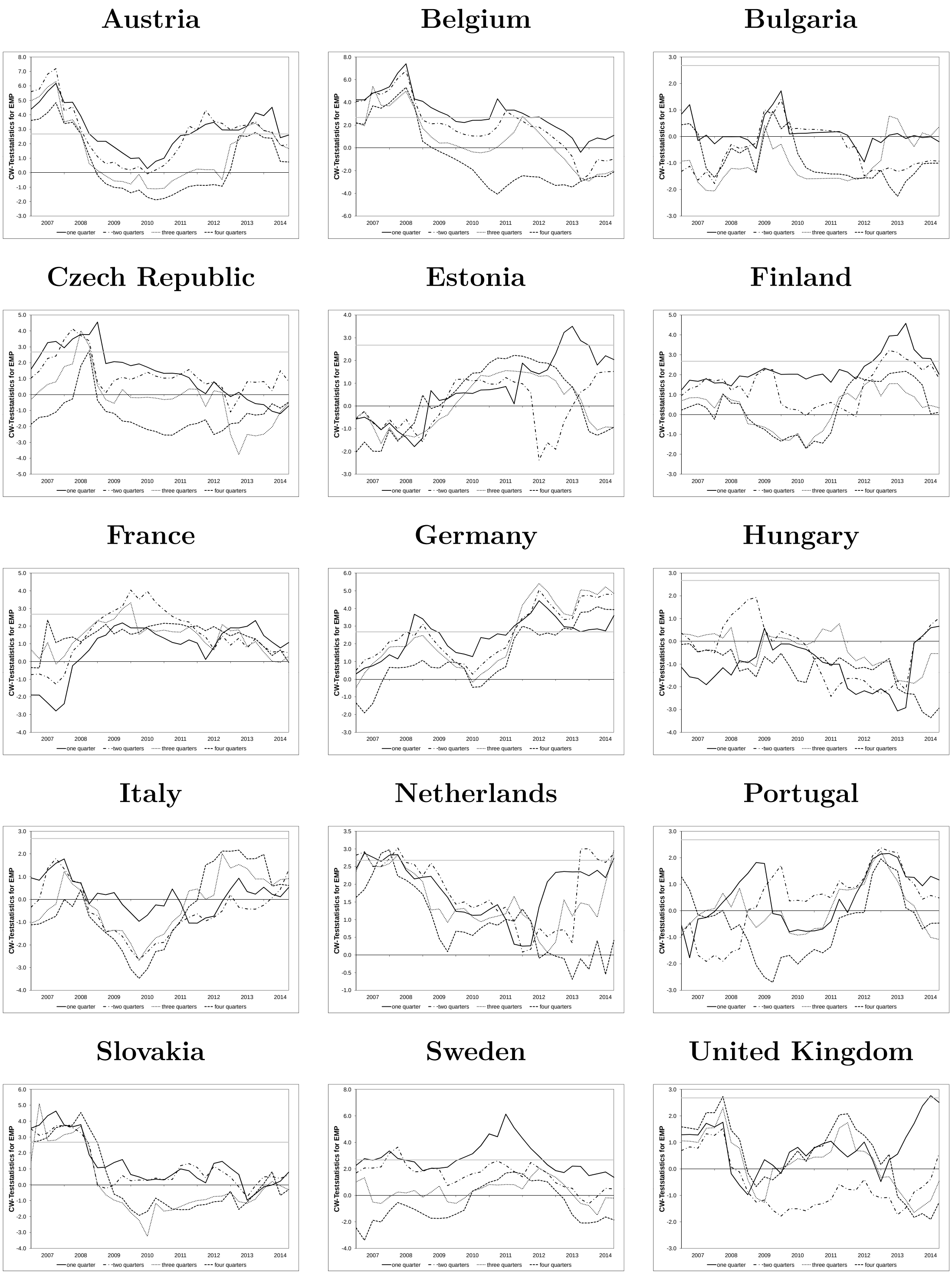
One main criticism that can be made of our results is the fact that our sample comprises the global crisis of the years 2008/2009. Thus, the forecasting performance of employment expectations can change over time. Anticipating this criticism, we applied the so called Fluctuation test proposed by Giacomini/Rossi (2010). The main idea of this test is to check whether the indicator model is preferable over time. With standard tests such as those conducted by Clark/West (2007), we are only able to assess the forecasting performance of an indicator over the entire period under investigation. The Fluctuation test can assess the performance over different sub-samples instead. To conduct the Fluctuation test we need the Clark-West test statistics for different sub-samples and an adequate rolling window that meets the requirements of Giacomini/Rossi (2010). The statistical power of the Fluctuation test crucially depends on the length of the estimation period (T_E) and on the ratio $\mu = m/T_F$, with m as the size of the rolling window mentioned above and T_F as the length of our out-of-sample window. In our case, we operate with $T_E = 28$, $T_F = 40$, $m = 10$ and $\mu = 0.3$. As proposed by Giacomini/Rossi (2010), this combination ensures that the Fluctuation test has meaningful statistical power. Since we decided to apply a rolling window of 10 quarters, we are able to show the Fluctuation test for the period from 2007 to 2014. Figure 4 shows the outcome of the Fluctuation test for all of the European countries in the sample. Each panel contains four different lines, each representing the rolling Clark-West statistics for one specific forecast horizon. We also highlight the zero line, thus indicating whether employment expectations produce lower forecast errors compared to the benchmark model. A value higher than zero denotes that the indicator model is better than the autoregressive benchmark. The opposite holds for values smaller than zero. However, showing the rolling Clark-West statistics over time does not tell us much about significant differences between the competing models. We therefore include the gray horizontal lines in each panel, indicating the critical value for the one-sided Fluctuation test to the 10% significance level. Whenever a line crosses this critical value, the indicator produces lower forecast errors than the benchmark model.[17]

The outcome of the Fluctuation test is very interesting, since there are phases where employment expectations are significantly better than the benchmark, followed by phases where no major differences are observable. Let us turn to those countries where we find no forecast improvement through our indicator, namely: Bulgaria, Germany, and Hungary. For Bulgaria and Hungary we clearly see that there is no period between 2007 and 2014 where employment expectations are significantly better than the autoregressive benchmark. However, for Germany the picture is different. Whereas employment expectations tend to perform rather poorly in the overall evaluation, the Fluctuation test clearly reveals that there are periods where the indicator works quite well in forecasting total employment. This last finding lets us conclude again, while bearing the other results of the Fluctuation test from Figure 4 in mind, that *EEXP* is an effective instrument for forecasting short-term employment.

Another question is the role played by an European aggregate of *EEXP_av* in terms of forecasting national employment. It is possible that the employment expectations of the European Union may contain more information than national indicators, thus boosting their forecasting performance. We therefore draw a picture similar to Figure 3. Instead of showing the relative performance of the indicators in an expanding and rolling setup, Figure 5 compares the *rRMSFE* of national *EEXP_av* and a similar indicator for the EU-28 (*EEXP-EU-28_av*). The relative performance of *EEXP_av* can be found on the x-axis, whereas the *rRMSFEs* for *EEXP-EU-28_av* are displayed on the y-axis. As before, we

[17] In order to keep the number of results clear and to save space, we only present the Fluctuation test results for *EMP* and *EEXP_av*. All the other results and pictures are available upon request.

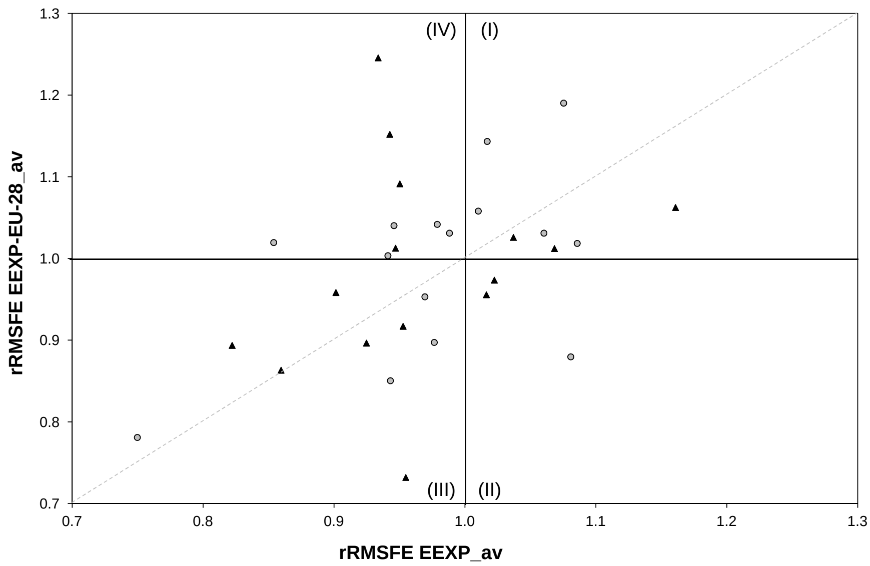
Figure 4: Fluctuation test statistic for *EMP*, *EEXP_av* and each European state



Note: Calculations based on the forecasting period (2005Q1–2014Q4) and a rolling window of ten quarters.

only present the results for the shortest forecast horizon $h = 1$. We also enclose the results for *EMP* in gray circles and use black triangles for *MEMP*. Dots or triangles lying in quadrant (I) indicate that *EEXP_av* as well as *EEXP-EU-28_av* are not able to beat the benchmark model. The opposite holds for quadrant (III), where both indicators perform better than the autoregressive process. However, the main quadrants of interest here are (II) and (IV). A dot (triangle) lying in quadrant (II) indicates that the European indicator can beat the benchmark model, whereas the national cannot. Thus, by using the European pendant of employment expectations, we are able to predict national employment growth. All dots or triangles found in quadrant (IV) document that there is no improvement with the European employment indicator, whereas its national counterpart is able to predict employment growth better than the benchmark model. In addition to Figure 3, we add a 45° line in Figure 5. All *rRMSFE* combinations below this line indicate that the European indicator produces a lower forecast error than national employment expectations. x-y-pairs above the 45° line indicate lower *rRMSFEs* of *EEXP_av* compared to *EEXP-EU-28_av*.

Figure 5: Relative forecast errors for *EEXP_av* vs. *EEXP-EU-28_av* ($h = 1$)



Note: Calculations are based on the whole sample (1998Q1–2014Q4). Gray circles indicate the relative forecast errors for total employment (*EMP*). The relative forecasting performance for *MEMP* is represented by black triangles.

All in all, national employment expectations often do a very good job of predicting national employment growth. There are, however, some improvements to be obtained by using the European indicator. We find one country (Slovakia) for *EMP* and two countries (France and Italy) for *MEMP* for which the European employment expectations are clearly better than the national indicator. There are also some combinations in quadrant (III) where *EEXP-EU-28_av* produces lower relative forecasting errors than the national indicator. For total employment we find improvements for Austria, the Czech Republic and France. Looking at *MEMP*, the relative performance of the European indicator is better for the Czech Republic, Slovakia and Sweden. For some countries the European indicator is therefore a tough competitor compared to national employment expectations.

The last issue raised by this paper is: why does the survey-based indicator not work similarly for all countries? The explanation is manifold and a conclusive answer is beyond the scope of this paper. It does, however, discuss some preliminary explanations. Firstly, in some countries the survey may suffer from non-responsive firms, leading to a significant bias that deteriorates the accuracy of survey-based indicators (see Seiler, 2014). Since we were not able to analyze firm-level data for each European state in such detail, it could be the case that employment expectations are biased in Bulgaria and Hungary due to non-responses, which would explain why the indicator loses its power for forecasting employment growth in these countries. Secondly, the two countries where a forecasting model with employment expectations does not beat the benchmark (Bulgaria and Hungary) are still seen as transition economies (see EBRD, 2013). A lack of experience or the false prediction of future developments may give rise to wrong answers from respondents in these states. A third, and very simple explanation is that labor markets naturally vary between states. A high degree of heterogeneity can be seen between country-specific labor market institutions and matching processes between firms and the unemployed. This heterogeneity can lead to spreads between employment expectations and observable employment in an economy. Firms may want to fill a vacancy, but are not able to get suitable candidates due to their low matching efficiency, for instance. Fourthly, a discussion of the aggregation of firm's responses, e.g., balances, exists in the literature (Croux/Dekimpe/Lemmens, 2005; Claveria/Pons/Ramos, 2007). Alongside this discussion of the sheer aggregation of the raw data, several methods exist to transform qualitative indicators into quantitative information. Rather than using the direction of employment change, a quantitative measure extracted from employment expectations may have predictive power for the magnitude of employment change (Claveria/Pons/Ramos, 2007). However, our paper gives initial insights into the forecasting performance of survey-based indicators for predicting labor market variables. Follow-up studies may concentrate on these issues.

We close our discussion by comparing our results with those of existing studies and formulate some words of caution that point to further avenues for future research. The two studies our results can compared to are those by dos Santos (2003) and Abberger (2007b). Whereas our results for Portugal are fairly in line with dos Santos (2003), they partially differ from those presented by Abberger (2007b). There are three possible reasons why the results for Germany differ between our study and that of Abberger. The first, and probably the most important reason, is that Abberger uses year-on-year growth rates instead of quarter-on-quarter growth. As shown by Lehmann (2015) for European export forecasts, the performance of survey-based indicators increases by applying them to year-on-year growth rates instead of quarter-on-quarter growth. This could also be the case for employment data. Secondly, Abberger uses a monthly sample from January 1993 to May 2004, whereas we use quarterly data from 1998 to 2014. Either the aggregation to quarterly information or the different time span can cause the results to differ between the studies. Thirdly, Abberger identifies the leading properties of employment expectations with in-sample techniques, whereas we additionally provide a pseudo out-of-sample examination. This study comparison, however, is important for the practical relevance of employment expectations and perfectly fits into the discussion brought forward by Gayer (2005). His recommendation is to clarify which survey indicator refers to which target variable and their possible transformations. This could be one reason why we observe that our indicator of employment expectations is able to reduce forecast errors in countries, whereas it is not able to do so in other countries. Such a similar discussion can be conducted for the differences between total employment (*EMP*) and a sub-sample of sectors (*MEMP*). Since the service sector is not part of our indicator yet, including it could improve the forecasting performance of employment expectations. All of these points can be investigated by future research activities.

4 Summary and conclusion

Survey-based indicators serve as powerful tools for forecasting different macroeconomic variables. Since the survey results used here are immediately available at the end of each month and do not suffer from major revisions, the outcome of specific questions can easily be used to analyze the recent state of the economy. Most of the existing studies in the field of economic forecasting try to evaluate the forecasting performance of survey-based indicators for fairly standard macroeconomic variables, such as gross domestic product; articles for labor market variables are scarce. Our paper fills this gap in the literature on this topic.

We used the results from the *Joint Harmonised EU Programme of Business and Consumer Surveys* to forecast employment growth in Europe. We specifically focused on the question of employment expectations. Our sample consisted of 15 European states, which covered almost 75% of total employment in the EU-28, for the period 1998Q1 to 2014Q4. To evaluate the forecasting performance of employment expectations, we applied both in-sample and out-of-sample techniques. Some descriptive statistics, as well as Granger causality tests, suggest that an indicator based on employment expectations can be used to forecast employment growth for most countries in our sample. The out-of-sample examination based on an autoregressive distributed lag model showed that our indicator produces significantly lower forecast errors than several benchmark models. Whereas the best relative performance of the indicator model can be found for Finland and Sweden, employment expectations clearly have no better predictive value than the benchmark model for Bulgaria and Hungary. To verify our findings, we ran a robustness check and find that our results are fairly robust irrespective of whether a rolling or an expanding window approach is adopted. Only 28% of all forecasts change their relative performance. The discussion of our results revealed, among other things, two major insights. Firstly, the Fluctuation test by Giacomini/Rossi (2010) indicated that there are phases in which employment expectations are significantly better than the benchmark, followed by phases in which no statistically significant differences can be observed. Secondly, in addition to national employment expectations, a similar indicator for the EU-28 aggregate shows a forecasting improvement for France, Italy and Slovakia. For all the other countries, the national indicator is doing a better job.

Our contribution to survey-based forecasting is manifold: we focus on a very important part of any economy, the labor market and add to the discussion by Croux/Dekimpe/Lemmens (2005) that different survey results should have some predictive power for different macroeconomic variables. Here, we contribute by analyzing employment expectations and employment growth. Moreover, we examined the forecasting performance of a survey-based indicator not for the Euro area as a whole but rather for a large number of single states. This gives a broader picture of how survey results work as indicators in different states. As our results highlight, employment expectations are an indicator for forecasting employment growth. However, for some countries the indicator fails to improve forecasts in comparison to a simple benchmark model. We also provide a discussion of forecasting stability over time that is rarely included in previous literature on this topic and gives some preliminary explanations as to why the results from our study emerge. An in-depth analysis of these ideas offers scope for future research.

References

Abberger, K. (2007a): Forecasting Quarter-on-Quarter Changes of German GDP with Monthly Business Tendency Survey Results. Ifo Working Paper No. 40.

Abberger, K. (2007b): Qualitative business surveys and the assessment of employment – A case study for Germany. In: International Journal of Forecasting, Vol. 23, No. 2, p. 249–258.

Ang, A.; Bekaert, G.; Wei, M. (2007): Do macro variables, asset markets, or surveys forecast inflation better? In: Journal of Monetary Economics, Vol. 54, No. 4, p. 1163–1212.

Chatfield, C. (1995): Model Uncertainty, Data Mining and Statistical Inference. In: Journal of the Royal Statistical Society, Series A, Vol. 158, No. 3, p. 419–466.

Clark, T. E.; West, K. D. (2007): Approximately normal tests for equal predicitve accuracy in nested models. In: Journal of Econometrics, Vol. 138, No. 1, p. 291–311.

Claveria, O.; Pons, E.; Ramos, R. (2007): Business and consumer expectations and macroeconomic forecasts. In: International Journal of Forecasting, Vol. 23, No. 1, p. 47–69.

Croux, C.; Dekimpe, M. G.; Lemmens, A. (2005): On the predictive content of production surveys: A pan-European study. In: International Journal of Forecasting, Vol. 21, No. 2, p. 363–375.

Diebold, F. X.; Mariano, R. S. (1995): Comparing Predictive Accuracy. In: Journal of Business and Economic Statistics, Vol. 13, No. 3, p. 253–263.

dos Santos, R. H. (2003): The use of qualitative data for short term analysis. Banco de Portugal Economic Bulletin September 2003.

EBRD (2013): Transition Report 2013. European Bank for Reconstruction and Development, ISBN 9781898802402.

European Commission (2007): The Joint Harmonised EU Programme of Business and Consumer Surveys – User Guide. Directorate General for Economic and Financial Affairs, http://ec.europa.eu/economy_finance/db_indicators/surveys/documents/userguide_en.pdf.

Eurostat (2008): NACE Rev. 2 - Statistical classification of economic activities in the European Community. Eurostat Methodologies and Working Papers, ISSN 1977-0375.

Fritsche, U.; Stephan, S. (2002): Leading Indicators of German Business Cycles. In: Journal of Economics and Statistics, Vol. 222, No. 3, p. 289–315.

Gayer, C. (2005): Forecast Evaluation of European Commission Survey Indicators. In: Journal of Business Cycle Measurement and Analysis, Vol. 2005, No. 2, p. 157–183.

Giacomini, R.; Rossi, B. (2010): Forecast comparisons in unstable environments. In: Journal of Applied Econometrics, Vol. 25, No. 4, p. 595–620.

Graff, M.; Mannino, M.; Siegenthaler, M. (2012): A real time Evaluation of Employment Forecasts in Switzerland. KOF Working Papers No. 320.

Granger, C. W. J. (1969): Investigating Causal Relations by Econometric Models and Cross-spectral Methods. In: Econometrica, Vol. 37, No. 3, p. 424–438.

Hanssens, M.; Vanden Abeele, P. M. (1987): A Time-Series Study of the Formation and Predictive Performance of EEC Production Survey Expectations. In: Journal of Business & Economic Statistics, Vol. 5, No. 4, p. 507–519.

Hansson, J.; Jansson, P.; Löf, M. (2005): Business survey data: Do they help in forecasting GDP growth? In: International Journal of Forecasting, Vol. 21, No. 2, p. 377–389.

Hartle, D. (1958): Predictions Derived from the Employment Forecast Survey. In: The Canadian Journal of Economics and Political Science, Vol. 24, No. 3, p. 373–390.

Hutter, C.; Weber, E. (2015): Constructing a new leading indicator for unemployment from a survey among German employment agencies. In: Applied Economics, Vol. 47, No. 3, p. 3540–3558.

Lehmann, R. (2015): Survey-based indicators vs. hard data: What improves export forecasts in Europe? Ifo Working Paper No. 196.

Lütkepohl, H. (2005): New Introduction to Multiple Time Series Analysis. Springer-Verlag Berlin Heidelberg.

Martinsen, K.; Ravazzolo, F.; Wulfsberg, F. (2014): Forecasting macroeconomic variables using disaggregate survey data. In: International Journal of Forecasting, Vol. 30, No. 1, p. 65–77.

Ng, S.; Perron, P. (2001): Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power. In: Econometrica, Vol. 69, No. 6, p. 1519–1554.

Robinzonov, N.; Wohlrabe, K. (2010): Freedom of Choice in Macroeconomic Forecasting. In: CESifo Economic Studies, Vol. 56, No. 2, p. 192–220.

Rossi, B.; Sekhposyan, T. (2011): Understanding models' forecasting performance. In: Journal of Econometrics, Vol. 164, No. 1, p. 158–172.

Seiler, C. (2014): On the robustness of balance statistics with respect to nonresponse. In: OECD Journal: Journal of Business Cycle Measurement and Analysis, Vol. 2014, No. 2, p. 45–62.

Siliverstovs, B. (2013): Do Business Tendency Surveys Help in Forecasting Employment? A Real-Time Evidence for Switzerland. In: OECD Journal: Journal of Business Cycle Measurement and Analysis, Vol. 2013, No. 2, p. 1–20.

Österholm, P. (2010): Improving Unemployment Rate Forecasts Using Survey Data. In: Finnish Economic Papers, Vol. 23, No. 1, p. 16–26.

Weber, E.; Zika, G. (2013): Labour market forecasting – Is disaggregation useful? IAB-Discussion Paper 14/2013.

A Appendix

Table A.1: Descriptive statistics

Variable	Mean	Std. Dev.	Min.	Max.
Austria				
EMP (in %)	0.246	0.241	-0.487	0.967
MEMP (in %)	0.118	0.326	-0.989	0.890
EEXP_av (balances)	-3.390	6.094	-20.138	7.588
EEXP_tm (balances)	-3.573	6.184	-21.249	8.217
Belgium				
EMP (in %)	0.223	0.241	-0.351	0.671
MEMP (in %)	-0.062	0.342	-0.936	0.698
EEXP_av (balances)	-0.418	6.544	-19.369	11.274
EEXP_tm (balances)	-0.422	6.649	-20.299	10.629
Bulgaria				
EMP (in %)	-0.009	0.834	-3.383	1.374
MEMP (in %)	0.025	1.416	-5.668	4.206
EEXP_av (balances)	-8.277	10.940	-35.920	11.005
EEXP_tm (balances)	-8.104	10.789	-36.219	9.813
Czech Republic				
EMP (in %)	0.019	0.470	-1.163	1.301
MEMP (in %)	-0.054	0.583	-1.674	1.268
EEXP_av (balances)	-3.739	12.025	-35.598	19.458
EEXP_tm (balances)	-3.997	12.392	-36.956	20.287
Estonia				
EMP (in %)	-0.021	1.530	-5.238	2.984
MEMP (in %)	-0.042	2.933	-6.780	7.113
EEXP_av (balances)	-1.842	14.796	-48.961	21.658
EEXP_tm (balances)	-1.866	15.469	-55.467	23.368
Finland				
EMP (in %)	0.194	0.631	-1.695	1.541
MEMP (in %)	-0.004	1.035	-3.756	2.026
EEXP_av (balances)	-8.492	9.723	-34.292	8.604
EEXP_tm (balances)	-8.753	10.335	-35.757	9.987
France				
EMP (in %)	0.181	0.251	-0.461	0.712
MEMP (in %)	0.077	0.295	-0.644	0.986
EEXP_av (balances)	-3.786	8.930	-27.998	14.263
EEXP_tm (balances)	-3.857	9.370	-28.552	15.597
Germany				
EMP (in %)	0.175	0.292	-0.394	1.021
MEMP (in %)	0.004	0.398	-0.821	0.749
EEXP_av (balances)	-11.041	10.392	-29.305	10.090
EEXP_tm (balances)	-11.085	10.888	-32.623	12.023
Hungary				
EMP (in %)	0.057	0.544	-1.330	1.323
MEMP (in %)	-0.043	1.155	-1.872	5.515
EEXP_av (balances)	-7.097	6.706	-31.337	6.289
EEXP_tm (balances)	-6.925	6.978	-35.283	6.032

Continued on next page...

Table A.1: Descriptive statistics – continued

Variable	Mean	Std. Dev.	Min.	Max.
Italy				
EMP (in %)	0.141	0.424	-1.136	1.061
MEMP (in %)	0.023	0.528	-1.596	0.967
EEXP_av (balances)	-1.747	6.191	-13.865	15.605
EEXP_tm (balances)	-1.489	6.211	-15.637	15.590
Netherlands				
EMP (in %)	0.175	0.442	-0.946	1.233
MEMP (in %)	0.018	0.510	-1.053	1.568
EEXP_av (balances)	-0.649	9.615	-21.148	12.646
EEXP_tm (balances)	-0.716	9.975	-23.127	13.215
Portugal				
EMP (in %)	-0.080	0.681	-2.114	1.242
MEMP (in %)	-0.261	1.003	-2.568	1.736
EEXP_av (balances)	-9.219	9.280	-28.431	6.398
EEXP_tm (balances)	-9.410	9.320	-29.471	6.516
Slovakia				
EMP (in %)	0.076	0.735	-1.944	1.687
MEMP (in %)	0.074	0.984	-2.958	1.624
EEXP_av (balances)	-5.369	12.908	-39.794	13.302
EEXP_tm (balances)	-5.152	13.744	-43.788	14.507
Sweden				
EMP (in %)	0.240	0.416	-1.216	1.067
MEMP (in %)	0.138	0.638	-1.495	1.454
EEXP_av (balances)	-10.281	14.978	-45.467	24.870
EEXP_tm (balances)	-10.180	15.030	-44.729	26.757
United Kingdom				
MEMP (in %)	-0.047	0.460	-1.442	0.883
EMP (in %)	0.219	0.329	-0.950	0.812
EEXP_av (balances)	-2.028	10.494	-43.837	16.036
EEXP_tm (balances)	-2.101	10.688	-42.351	17.526

Note: Calculations are based on the whole sample (1998Q1–2014Q4).
Source: European Commission, Eurostat, author's calculations.

Table A.2: Results of the unit root tests for *MEMP* and *EEXP_tm*

Country	**MEMP**					**EEXP_tm**				
	KPSS		NP		Δ	KPSS		NP		Δ
	Const.	Trend	Const.	Trend		Const.	Trend	Const.	Trend	
Austria			***	***				**		
Belgium			***	***				***	***	
Bulgaria		**	***	***		**	**			X
Czech Republic		*	***	***			**			X
Estonia			***	***				***	*	
Finland	*		***	***				***	**	
France	**		**	***		*		***	***	
Germany			***	***		**		**		
Hungary			***	***			**	**		
Italy	***		***	***		**		***	**	
Netherlands			**		X	**		**	*	
Portugal	**		**	***		**			***	X
Slovakia		*	***	***				**	*	
Sweden			**	*				***	***	
United Kingdom		*	***	***			*	*		X

Note: Calculations are based on the whole sample (1998Q1–2014Q4). The Ng-Perron (NP) test states a unit root under the null hypothesis. The Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test has stationarity under the null hypothesis. For the KPSS and the NP tests, only test statistics and critical values are available. We therefore use asterisks to show whether the null can be rejected or not. In all cases we first tested stationarity in levels (qoq qrowth rate or balances). If the levels turned out to be non-stationary, we then tested first differences of the variables. The column Δ presents the decision on the transformation of the variables. An **X** indicates that first differences are applied. ***, ** and * indicate the rejection of the null hypothesis at the 1%, 5% and 10% significance level.

Table A.3: Granger causality results for *EMP*, *MEMP* and *EEXP_tm*

Country	**EMP**		
	EEXP_tm → EMP	EMP → EEXP_tm	Result
Austria	0.002	0.027	FB
Belgium	0.006	0.735	+
Bulgaria	0.128	0.531	X
Czech Republic	0.006	0.114	+
Estonia	0.002	0.180	+
Finland	0.001	0.997	+
France	0.000	0.482	+
Germany	0.125	0.052	X
Hungary	0.553	0.122	X
Italy	0.008	0.107	+
Netherlands	0.046	0.012	FB
Portugal	0.002	0.260	+
Slovakia	0.001	0.313	+
Sweden	0.002	0.678	+
United Kingdom	0.008	0.283	+
Country	**MEMP**		
	EEXP_tm → MEMP	MEMP → EEXP_tm	Result
Austria	0.000	0.168	+
Belgium	0.000	0.429	+
Bulgaria	0.116	0.712	X
Czech Republic	0.007	0.060	FB
Estonia	0.029	0.032	FB
Finland	0.005	0.668	+
France	0.001	0.049	FB
Germany	0.006	0.022	FB
Hungary	0.189	0.552	X
Italy	0.026	0.153	+
Netherlands	0.000	0.213	+
Portugal	0.001	0.150	+
Slovakia	0.003	0.634	+
Sweden	0.000	0.535	+
United Kingdom	0.017	0.084	FB

Note: Calculations are based on the whole sample (1998Q1–2014Q4). The table presents p-values from the Granger causality test. *Acronyms:* +, *EEXP* only Granger causes employment growth (case [i]); FB, feedback effects are present (case [ii]); X, employment growth only Granger causes *EEXP* (case [iii]) or no relationship (case [iv]).

Table A.4: Out-of-sample results (expanding) for *EMP* and *MEMP*

Model	**EMP**				**MEMP**			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
				Austria				
AR(p) in %	0.218	0.269	0.283	0.278	0.218	0.382	0.384	0.394
AR(1)	1.004	0.932	0.921^{**}	0.974^{*}	0.997	0.955	0.944^{***}	0.974^{**}
ISM	1.146	0.944	0.907^{*}	0.926	1.127	0.961^{*}	0.964	0.945^{*}
RW	1.206	1.201	1.233	1.415	1.255	1.106	1.201	1.414
EEXP_av	0.926^{**}	0.892^{***}	0.983^{**}	1.094	0.857^{***}	0.887^{***}	0.955^{**}	0.993
EEXP_tm	0.908^{**}	0.932^{**}	0.942^{***}	1.094	0.877^{**}	0.828^{**}	0.929^{**}	1.055
				Belgium				
AR(p) in %	0.139	0.188	0.225	0.243	0.139	0.248	0.311	0.328
AR(1)	0.977^{**}	0.992^{*}	0.994	0.993	0.982^{**}	1.030	0.988	0.987
ISM	1.631	1.223	1.035	0.965	1.689	1.260	1.020	0.974^{**}
RW	1.451	1.400	1.388	1.380	1.513	1.458	1.396	1.442
EEXP_av	0.963^{**}	0.947^{***}	1.079	1.200	0.924^{**}	1.044	1.111	1.639
EEXP_tm	0.943^{**}	1.001	1.045	1.137	0.888^{**}	1.022	1.013	1.324
				Bulgaria				
AR(p) in %	0.706	0.799	0.817	0.796	0.706	1.437	1.411	1.384
AR(1)	0.977	1.007	0.991^{*}	0.997	0.999	0.999	1.007	0.997
ISM	1.141	0.996	0.976	1.006	1.076	1.007	1.036	1.066
RW	1.505	1.335	1.537	1.295	1.219	1.161	1.096	1.154
EEXP_av	1.175	1.032	1.007	0.998	1.024	1.009	0.994	1.063
EEXP_tm	1.157	1.024	1.016	0.996	0.998	1.024	0.995	1.037
				Czech Republic				
AR(p) in %	0.417	0.467	0.485	0.537	0.417	0.593	0.610	0.664
AR(1)	0.990	0.991	1.013	1.009	0.963^{*}	1.013	1.020	0.971
ISM	1.213	1.103	1.076	0.984	1.139	1.060	1.041	0.961
RW	1.245	1.229	1.364	1.264	1.328	1.363	1.414	1.354
EEXP_av	0.972^{*}	0.964	1.027	1.017	0.945^{**}	0.959^{*}	1.010	0.985
EEXP_tm	1.003	0.946	0.981	1.035	0.952^{**}	0.932^{*}	1.036	0.907^{**}
				Estonia				
AR(p) in %	1.716	1.884	1.832	1.934	1.716	3.456	3.337	3.356
AR(1)	0.997	0.966^{*}	0.989	0.987	0.975	0.991	1.002	0.976
ISM	1.052	0.967^{*}	0.998	0.951^{*}	0.986	0.948	0.979^{**}	0.974
RW	1.380	1.147	1.276	1.353	1.463	1.303	1.242	1.291
EEXP_av	0.948^{**}	0.901^{**}	0.983^{**}	0.932^{*}	0.900^{**}	0.943	0.934^{**}	1.008
EEXP_tm	0.903^{**}	0.893^{**}	0.981^{**}	0.974	0.898^{**}	0.923^{*}	0.982^{**}	1.089
				Finland				
AR(p) in %	0.677	0.671	0.660	0.688	0.677	1.197	1.146	1.152
AR(1)	1.006	0.987^{*}	1.005	0.993	1.001	0.967	1.001	1.002
ISM	0.976	0.991	1.014	0.977	1.005	0.955^{*}	1.002	1.000
RW	1.103	1.239	1.414	1.448	1.251	1.215	1.297	1.256
EEXP_av	0.807^{**}	0.900^{***}	1.079	1.145	0.855^{***}	0.974^{*}	0.995	1.014
EEXP_tm	0.823^{**}	0.896^{***}	1.038	1.074	0.897^{**}	1.015	1.035	1.034
				France				
AR(p) in %	0.092	0.105	0.102	0.102	0.092	0.203	0.241	0.263
AR(1)	1.000	0.959^{*}	0.993	0.998	1.118	0.984^{*}	1.007	1.014
ISM	1.049	0.925^{*}	0.954	0.962	2.836	1.527	1.311	1.221
RW	1.397	1.259	1.401	1.408	1.846	1.229	1.190	1.201
EEXP_av	0.955^{**}	0.883^{**}	0.874^{*}	0.905^{**}	0.975^{**}	0.951^{**}	1.044	1.043
EEXP_tm	0.899^{**}	0.851^{*}	0.837^{*}	0.908^{**}	0.939^{***}	0.909^{**}	1.021	0.996^{*}

Continued on next page...

Table A.4: Out-of-sample results (expanding) for *EMP* and *MEMP* – continued

Model	EMP				MEMP			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
Germany								
AR(p) in %	0.213	0.225	0.226	0.252	0.213	0.363	0.387	0.426
AR(1)	0.993	0.989*	0.994	0.955*	0.985**	0.996	1.006	0.977**
ISM	1.044	0.999	1.004	0.904**	1.172	1.108	1.051	0.964
RW	1.242	1.178	1.350	1.313	1.191	1.192	1.198	1.253
EEXP_av	0.961***	1.001	1.016	1.026	0.880***	0.899***	0.827***	0.891***
EEXP_tm	0.961***	1.002	0.977***	1.006	0.858***	0.815***	0.796***	0.773***
Hungary								
AR(p) in %	0.591	0.600	0.631	0.649	0.591	1.339	1.296	1.267
AR(1)	0.959***	0.941*	0.932*	0.942	0.991	0.993	1.032	1.002
ISM	0.997	0.989	0.948	0.928*	0.976**	0.987	1.024	1.044
RW	1.135	1.168	1.255	1.168	1.278	1.485	1.034	1.363
EEXP_av	1.039	1.073	0.998*	1.082	1.026	1.001	1.029	1.010
EEXP_tm	1.104	1.029	1.028	1.079	1.078	0.997	1.021	1.005
Italy								
AR(p) in %	0.479	0.490	0.552	0.567	0.479	0.604	0.646	0.648
AR(1)	0.991*	0.937	0.883	0.855*	0.999	0.981	0.990	0.988
ISM	1.008	0.997	0.892	0.876*	1.103	1.055	0.996	1.000
RW	1.099	1.143	1.001	1.022	1.142	1.185	1.048	1.075
EEXP_av	0.991	1.027	1.001	0.978*	0.995	1.002	0.936**	0.927**
EEXP_tm	0.854**	0.981*	0.984*	0.966	0.894**	0.955**	0.923**	0.967**
Netherlands								
AR(p) in %	0.360	0.366	0.422	0.562	0.360	0.498	0.544	0.602
AR(1)	1.083	1.004	0.997	0.856	1.074	0.966***	1.008	0.901
ISM	1.254	1.246	1.095	0.833	1.041	1.017	0.940***	0.856**
RW	1.121	1.227	1.361	1.044	1.134	1.256	1.219	1.197
EEXP_av	0.929**	1.216	1.156	0.999**	0.977**	1.138	1.065	1.010
EEXP_tm	0.972**	1.311	1.174	1.047	0.904***	1.066	1.144	1.202
Portugal								
AR(p) in %	0.696	0.857	0.838	0.809	0.696	1.130	1.159	1.132
AR(1)	1.010	0.890	0.946	0.962	0.982*	0.942	1.000	0.998
ISM	1.123	0.920	0.947	0.984	1.135	1.017	1.001	1.030
RW	1.327	1.159	1.091	1.066	1.208	1.218	1.179	1.228
EEXP_av	0.971**	0.866**	0.955	1.009	0.877***	0.837**	1.021	1.122
EEXP_tm	0.992	0.844**	0.935**	0.948**	0.935**	0.897**	0.939**	0.998
Slovakia								
AR(p) in %	0.522	0.635	0.649	0.652	0.522	0.871	0.938	0.948
AR(1)	1.015	1.017	0.994	1.026	1.001	1.005	0.989	0.980
ISM	1.214	1.013	1.004	1.008	1.268	1.049	0.988	0.980
RW	1.195	1.215	1.350	1.415	1.285	1.478	1.470	1.477
EEXP_av	1.002	0.889**	0.907**	0.980**	0.969*	1.036	1.171	0.973*
EEXP_tm	1.111	0.937**	0.928**	1.024	1.177	1.083	1.103	1.071
Sweden								
AR(p) in %	0.320	0.378	0.444	0.469	0.320	0.629	0.668	0.691
AR(1)	0.943***	0.990*	0.997	0.998	0.994	0.993	0.992	0.986
ISM	1.303	1.119	0.965	0.922*	1.092	1.029	0.977*	0.949***
RW	1.267	1.308	1.360	1.388	1.227	1.370	1.400	1.469
EEXP_av	1.014	0.985**	1.050	1.194	1.004	0.941**	1.123	1.054
EEXP_tm	0.983**	1.034	1.021	1.158	0.934***	0.977**	1.048	1.076

Continued on next page...

Table A.4: Out-of-sample results (expanding) for *EMP* and *MEMP* – continued

Model	**EMP**				**MEMP**			
	h=1	h=2	h=3	h=4	h=1	h=2	h=3	h=4
				United Kingdom				
AR(p) in %	0.417	0.423	0.417	0.410	0.417	0.548	0.556	0.569
AR(1)	0.978	0.966	1.002	1.000	0.933***	0.977	0.982	0.984
ISM	0.949**	0.942**	0.960*	0.983*	0.969*	0.977	0.969*	0.952
RW	1.171	1.161	1.404	1.429	1.158	1.160	1.357	1.372
EEXP_av	1.033	0.963*	1.159	0.971*	1.078	1.004	1.086	0.938**
EEXP_tm	1.008	0.966**	1.105	0.971*	1.077	1.056	1.067	0.954**

Note: Calculations are based on the whole sample (1998Q1–2014Q4). The table presents the relative root mean squared forecast errors (*rRMSFE*) of the different models and the benchmark. The row AR(p) shows the *RMSFE* (in %) for the benchmark model. *ISM*, in-sample mean; *RW*, Random Walk. Asterisks show significant differences between forecast errors due to the Clark-West test. ***, ** and * indicate statistical significance at the 1%, 5% and 10% level, respectively.

Recently published

No.	Author(s)	Title	Date
15/2015	Zapf, I.	Individual and workplace-specific determinants of paid and unpaid overtime work in Germany	4/15
16/2015	Horbach, J. Janser, M.	The role of innovation and agglomeration for employment growth in the environmental sector	6/15
17/2015	Dorner, M. Fryges, H. Schopen, K.	Wages in high-tech start-ups – do academic spin-offs pay a wage premium?	6/15
18/2015	Möller, J.	Verheißung oder Bedrohung? Die Arbeitsmarktwirkungen einer vierten industriellen Revolution	6/15
19/2015	Hecht, V.	Location choice of German multinationals in the Czech Republic: The importance of agglomeration economies	7/15
20/2015	Wiemers, J.	Endogenizing take-up of social assistance in a microsimulation model: A case study for Germany	7/15
21/2015	Wanger, S. Weigand, R. Zapf, I.	Measuring hours worked in Germany: Contents, data and methodological essentials of the IAB working time measurement concept	8/15
22/2015	Weigand, R. Wanger, S. Zapf, I.	Factor structural time series models for official statistics with an application to hours worked in Germany	8/15
23/2015	Zapf, I.	Who profits from working-time accounts? Empirical evidence on the determinants of working-time accounts on the employers‘ and employees‘ side	8/15
24/2015	Dietrich, H.	Jugendarbeitslosigkeit aus einer europäischen Perspektive: Theoretische Ansätze, empirische Konzepte und ausgewählte Befunde	9/15
25/2015	Christoph, B.	Empirische Maße zur Erfassung von Armut und materiellen Lebensbedingungen: Ansätze und Konzepte im Überblick	9/15
26/2015	Bauer, A.	Reallocation patterns across occupations	9/15
27/2015	Dauth, W. Fuchs, M. Otto, A.	Long-run processes of geographical concentration and dispersion: Evidence from Germany	9/15
28/2015	Klinger, S. Weber, E.	Detecting unemployment hysteresis: A simultaneous unobserved components model with Markov switching	10/15
29/2015	Eckman, S. Kreuter; F.	Misreporting to looping questions in surveys: Recall, motivation and burden	10/15

As per: 2015-11-02

For a full list, consult the IAB website
http://www.iab.de/de/publikationen/discussionpaper.aspx

Imprint

IAB-Discussion Paper 30/2015

Editorial address
Institute for Employment Research
of the Federal Employment Agency
Regensburger Str. 104
D-90478 Nuremberg

Editorial staff
Regina Stoll, Jutta Palm-Nowak

Technical completion
Gertrud Steele

Website
http://www.iab.de

Download of this Discussion Paper
http://doku.iab.de/discussionpapers/2015/dp3015.pdf

ISSN 2195-2663

For further inquiries contact the author:

Antje Weyh
Phone +49.371 9118 642
E-mail antje.weyh@iab.de