Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

ifo Working Papers

Survey-based indicators vs. hard data: What improves export forecasts in Europe?

Robert Lehmann

Ifo Working Paper No. 196

March 2015

Ifo Working Paper No. 196

Survey-based indicators vs. hard data: What improves export forecasts in Europe?

Abstract

In this study, we evaluate whether survey-based indicators produce lower forecast errors for export growth than indicators obtained from hard data such as price and cost competitiveness measures. Our pseudo out-of-sample analyses and forecast-encompassing tests reveal that survey-based indicators outperform the benchmark model as well as the indicators from hard data for most of the twenty European states focused on in our study and the aggregates EA-18 and EU-28. The most accurate forecasts are on average produced by the confidence indicator in the manufacturing sector, the economic sentiment indicator and the production expectations. However, large country differences in the forecast accuracy of survey-based indicators emerge. These differences are mainly explained by country-specific export compositions. A larger share in raw material or oil exports worsens the accuracy of soft indicators. The accuracy of soft indicators improves if countries have a larger share in exports of machinery goods. For hard indicators, we find only weak evidence for the export composition to explain differences in forecast accuracy.

JEL Code: F01, F10, F17.

Keywords: Export forecasting, European business and consumer survey, export expectations, price and cost competitiveness.

Robert Lehmann
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich
Dresden Branch
Einsteinstr. 3
01069 Dresden, Germany
Phone: +49(0)351/26476-21
lehmann@ifo.de

1. Motivation

When it comes to macroeconomic forecasting, the main figure recognized by the public is gross domestic product (GDP). However, from a practical point of view, economic forecasts are more than just the prediction of a single number. Most forecast suppliers, such as supra-national organizations, research institutes or banks, predict each single component of GDP (e.g. private consumption or exports) separately and merge them together to form a plausible and most likely forecast of total output. Such a disaggregated approach of forecasting GDP is also found to be preferable compared to a direct approach by the academic literature (see, among others, Angelini *et al.*, 2010; Drechsel and Scheufele, 2012). Thus, the forecast error for GDP can significantly be reduced by forecasting single components such as private consumption or exports. Academics have studied forecasts of private consumption in particular (see, among others, Vosen and Schmidt, 2011). The other components are more or less disregarded. In this paper, we focus on exports and ask whether export forecasts for a multitude of European states can be improved by either hard data, such as price and cost competitiveness measures, or by qualitative information gained from surveys.

From the demand-side calculation of GDP, exports are one of the major components. Considering that the share of exports of goods and services in total GDP rose from almost 30% in 1995 to 45% in 2013 for the EU-15, exports are one major source of the creation of business cycles, since they transfer international shocks into the domestic economy. Fiorito and Kollintzas (1994) find for the G7 that exports are procyclical and coincide with the business cycle of total output. So trade is an important pillar for the economic development of countries, as the empirical literature shows (see Frankel and Romer, 1999). Thus, especially unbiased export forecasts can, c.p., significantly reduce forecast errors of GDP.

Only a few studies exist that focus on the improvement of export forecasts. An early attempt has been made by Baghestani (1994). He finds that survey results obtained from professional forecasters improve predictions for US net exports. In the case of Portugal, Cardoso and Duarte (2006) find that business surveys improve the forecasts for export growth. For Taiwan, standard autoregressive integrated moving average (ARIMA) models are able to improve export forecasts compared to heuristic methods (Wang *et al.*, 2011). Additionally, two German studies exist. Jannsen and Richter (2012) use a capacity utilization weighted indicator obtained from major export partners to forecast German capital goods exports. Elstner *et al.* (2013) use hard data (e.g. foreign new orders in manufacturing) as well as indicators from the Ifo business survey (e.g. Ifo export expectations) to improve forecasts for German exports. Overall, survey indicators produce lower forecast errors than hard indicators do. Finally, Hanslin and Scheufele (2014) show that a weighted Purchasing Manager Index (PMI) from major trading partners improves Swiss exports more than other indicators.

Next to these country-specific studies, some contributions focus on country-aggregates. Keck *et al.* (2009) show that trade forecasts for the OECD25 can be improved by applying standard time series models in comparison to a 'naïve' prediction based on a deterministic trend. Economic theory names two major drivers of exports: relative prices and domestic demand of the importing trading partners. Thus, Ca'Zorzi and Schnatz (2010) use different measures of price and cost competitiveness to forecast extra Euro-area exports and find that for a recursive estimation approach the real effective exchange rate based on the export price index outperforms the other measures as well as a 'random walk' benchmark. For the Euro area, Frale *et al.* (2010) find that survey results play an important role for export forecasts. From a global perspective, Guichard and Rusticelli (2011) show that the industrial production (IP) and Purchasing Manager Indices are able to improve world trade forecasts.

We contribute to this existing literature by creating a forecasting competition between indicators gained from hard data and different survey-based indicators for a multitude of European countries. We do not focus solely on one indicator or state, but rather analyze sixteen indicators for twenty European states and the aggregates EA-18 and EU-28 in the period from 1996 to 2013. From a pseudo out-of-sample analysis and forecast-encompassing tests we can conclude that survey-based indicators produce the most accurate export forecasts and cannot be beaten by hard indicators.

In general, it is common knowledge that business and consumer surveys are powerful tools for macroeconomic forecasting. However, business surveys are not free of criticism. Croux *et al.* (2005) mention that surveys are very expensive and time-consuming for both the enterprise and the consumer. This expense, in terms both time and money, should result in any informative or even predictive character of the questions asked in the specific survey. The study by Croux *et al.* (2005) finds an improvement in industrial production forecasts through the usage of production expectations expressed by European firms. Despite the forecasting power of a survey indicator for European industrial production, the results for different macroeconomic aggregates are mixed. This leads to the conclusion by Claveria *et al.* (2007) that we actually have no definite idea why some qualitative indicators work for specific macroeconomic variables, whereas others do not. With this paper, we ask whether survey-based indicators are able to predict export growth for a multitude of European states. Additionally, our paper searches for the reasons of country differences in the forecasting performance of survey-based indicators. With standard regression techniques, we find that in particular the composition of exports plays a crucial role for the forecast accuracy of soft indicators. In countries with a high share in raw materials or oil exports, the forecast accuracy of survey-based indicators worsens. The opposite holds for countries with a high share in machinery and transport equipment exports. These results are underpinned by studying the impact of export diversity. It turns out that survey indicators produce, on average, lower forecast errors in countries with a higher degree of export diversification.

To evaluate the competition between soft and hard indicators to forecast export growth, the paper is organized as follows. In Section 2, we present the data and our empirical setup. Section 3 discusses our results in detail. Section 4 offers a conclusion.

2. Data and Empirical Setup

2.1. Data

2.1.1. Target Variable

Eurostat supplies comprehensive export data on a quarterly basis for all member states of the European Union plus Switzerland and Norway. These figures are comparable to each other, since they are based on consistent standards within national accounts. We use total exports, which is the sum of exports of goods and services.[1] These total export figures are measured in real terms and are seasonally adjusted by the CENSUS X-12-ARIMA procedure. Since we are interested in growth forecasts rather than levels, we transform the export figures into year-on-year (yoy) growth rates. Our forecast experiment relies on quarterly data from 1996Q1 to 2013Q4 for a large sample of European states. Due to some data restrictions (e.g. missing export data or survey results), we eliminate some countries, leaving us with the following 20 European states in the sample: Austria, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Italy, Latvia, Lithuania, Luxemburg, the Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden and the United Kingdom. Additionally, we test the indicators for the aggregates EA-18 and EU-28. Descriptive statistics are available upon request.

2.1.2. Indicators

The European Commission (EC) provides both survey indicators and hard data. The survey-based indicators are collected within the *Joint Harmonised EU Programme of Business and Consumer Surveys* on behalf of the European Commission. The survey is harmonized across all European states. The samples in each country are representative. For the business survey, the sample comprises firms from different sectors (industry, construction, retail trade and services).[2] We concentrate on the survey results obtained from the manufacturing sector for two reasons. First, the majority of exports are goods produced in the manufacturing sector. Second, the survey in the service sectors was first conducted in the mid-2000s, so the time series is too short for our purposes. The survey program in manufacturing is divided into monthly and quarterly questions. The most intuitive candidate to predict future export growth in a specific country is the following question, which we call export expectations (*EXEXP*): 'How do you expect your export orders to develop over the next three months?' The respondents can answer this question in three ways: (+) increase,

[1] The code of the corresponding time series is: *namq_exi_k*. All the data can be downloaded free of charge under *http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home*. Note that the figures used here are based on the European System of Accounts of 1995 (ESA 1995).

[2] The European Commission wants to keep the sample representative for each month. To ensure this, sample updates are necessary on occasion due to (for example) 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 Economic and Financial Affairs (2014).

(=) remain unchanged or (-) decrease. Since *EXEXP* is measured on a quarterly basis (as exports), no transformation is necessary. In line with the literature, we assess the forecasting power of "balances". These balances are expressed as differences between the weighted share of firms whose exports will increase and the weighted share of those that expect a decrease. The weights are based on the size of the firms (see European Commission Economic and Financial Affairs, 2014). All firms with a response "remain unchanged" are not considered. However, balances are not indisputable in the existing literature (for a critical discussion, see Croux *et al.*, 2005; Claveria *et al.*, 2007, and the references therein).

Bearing in mind that other survey indicators may also deliver important information to forecast export growth, we evaluate the following monthly indicators as well: (i) the confidence indicator in manufacturing (COF), (ii) the assessment of export order-book levels (EOBL), (iii) the assessment of order-book levels (OBL), (iv) production expectations for the month ahead (PEXP), (v) the assessment of stocks of finished products (SFP), (vi) a self-constructed capacity-based indicator in the style of the Kiel Institute for the World Economy (IfW; see Jannsen and Richter, 2012) and (vii) the economic sentiment indicator (ESI) of the whole economy. In addition, we use the consumer confidence indicator (CCOF) as a possible predictor. Since the balances of these eight additional indicators are on a monthly basis, we transform these balances with a simple three-month average to obtain quarterly data. All survey results are seasonally adjusted by the provider via the procedure DAINTIES.[3] All in all we end up with nine survey-based indicators.

Since the purpose of the paper is to create a "horse race" between survey-based indicators and hard data, we have to specify which variables are found in the category of hard data. One major driver for exports is the price and cost competitiveness of a specific country. The Department of Economic and Financial Affairs (ECFIN) at the European Commission provides price and cost competitiveness measures based on different price weights. We choose the quarterly real effective exchange rate (REER) against 37 industrial countries for each specific state in our sample.[4] The ECFIN provides REER data based on five different price weights: (i) harmonized consumer price index (HCPI), (ii) nominal unit labor costs of the total economy (ULCTOT), (iii) nominal unit wage costs in manufacturing (UWCMAN), (iv) the GDP deflator (GDPDEF) and (v) the price deflator for exports of goods and services (EXPI).[5] The discussion in Ca'Zorzi and Schnatz (2010) reveals different advantages and shortcomings of each of these five indicators (see Table 1 for an overview). The EXPI, in particular, has some remarkable disadvantages, such as heavy data revisions. We test the forecasting performance of each indicator and evaluate which of them works best.

[3]We are aware of the fact that an intensive discussion about seasonal adjustment and the forecasting properties of survey indicators exists in the academic literature. However, this issue is beyond the scope of this paper.

[4]More information can be found at: http://ec.europa.eu/economy_finance/db_indicators/competitiveness/index_en.htm.

[5]However, there is no standard indicator that measures price and cost competitiveness best (see Ca'Zorzi and Schnatz, 2010).

Table 1: Advantages (+) and shortcomings (–) of different REER measures

Price Weights	(+)	(–)
HCPI	• homogeneity across countries	• non-tradable goods included • no capital or intermediate goods included • distortions through subsidies and taxes
ULCTOT	• whole economy considered	• non-tradable goods included • only a fraction of the firm's costs considered • measurement problems
UWCMAN	• focus on cost side • labor productivity included	• only manufacturing considered
GDPDEF	• services included	• no complete comparability across countries • distortions through subsidies and taxes
EXPI	• direct prices of exports	• endogenous to exchange rate changes • if measured in values per physical unit, then export composition unfortunately changes competitiveness • publication lags and heavy revisions • no complete comparability across countries

Source: Authors' illustration based on Ca'Zorzi and Schnatz (2010).

As for the soft indicators, we do not only test these price and cost competitiveness measures as hard indicators exclusively. Thus, we decide to add two additional indicators to the horse race: the specific national industrial production index (PIPROD) and the industrial production index of the United States (PIPRODUS).[6] It could be argued that the national production index partially reflects foreign demand and should therefore be a good predictor for national exports. Additionally, PIPROD is a widely accepted business cycle indicator with a high forecasting power. We choose PIPRODUS since the United States is one of the most important export partners for a multitude of European states.

2.2. Empirical Setting

2.2.1. Forecast Model

We generate our pseudo out-of-sample forecasts by employing the following autoregressive distributed lag (ADL) model:

$$y_{t+h} = \alpha + \sum_{i=1}^{p} \beta_i y_{t-i} + \sum_{j=m}^{q} \gamma_j x_{t+1-j} + \varepsilon_{t+h} \ , \qquad (1)$$

where y_{t+h} is the h-step-ahead forecast for export growth and x_t represents one of the single indicators. The forecast horizon h is defined in the range of $h \in \{1, 2\}$ quarters since survey-based indicators are usually applied for short-term forecasts (see, among others, Gayer, 2005). We allow a maximum of four lags for our target variable and each single indicator: $p, q \leq 4$. The optimal lag length is determined by the Bayesian Information Criterion (BIC). The initial estimation period for Equation (1) ranges from 1996Q1 to 2004Q3 ($T_E = 35$). The

[6]One important indicator in the existing literature is foreign new orders in the manufacturing sector (see Elstner *et al.*, 2013). Unfortunately, to the best of our knowledge, Eurostat stopped reporting this indicator in 2012, so we cannot use it as a hard indicator.

period is then expanded successively by one quarter with a new specification of the model; the first forecast for y_t is calculated for 2004Q4 and the last for 2013Q4. 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 ($T_F = 37$) for every forecast horizon h. More details on direct-step forecasting can be found in Robinzonov and Wohlrabe (2010). As the benchmark model we chose a common AR(p) process.

2.2.2. Forecast Evaluation

To evaluate the forecast accuracy of our different models, we calculate forecast errors. 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 AR(p) 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. For the h-step-ahead indicator-based forecast, the RMSFE is:

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

The RMSFE for the benchmark model is $RMSFE^{ARp}_h$. To decide whether one indicator performs, on average, better than the autoregressive process, we calculate the relative RMSFE between the indicator model and the benchmark:

$$rRMSFE_h = \frac{RMSFE_h}{RMSFE^{ARp}_h}\,. \tag{3}$$

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 the test proposed by Diebold and Mariano (1995). Under the null hypothesis, the test states that the expected difference in the MSFE equals zero. With our notation this gives:

$$H_0 : E\left[\left(FE^{ARp}_{t+h}\right)^2 - \left(FE_{t+h}\right)^2\right] = E\left[MSFE^{ARp}_{t+h} - MSFE_{t+h}\right] = 0\,. \tag{4}$$

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 each of our single indicators – 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 and 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 and Mariano (1995), lose their power. On this account, we follow the literature (see, among others, Weber and Zika, 2013; Lehmann and Weyh, 2014) and apply the adjusted test statistic by Clark and West (2007):

$$CW_h = \sqrt{\frac{1}{\widehat{V}(a_{t+h})T_F} \sum_{t=1}^{T_F} \left(\underbrace{MSFE_{t+h}^{ARp} - \left[MSFE_{t+h} - \left(FE_{t+h} - FE_{t+h}^{ARp} \right)^2 \right]}_{a_{t+h}} \right)} \;, \tag{5}$$

with $\widehat{V}(a_{t+h})$ as the sample variance of a_{t+h} and $\left(FE_{t+h} - FE_{t+h}^{ARp}\right)^2$ 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.

2.2.3. Forecast Encompassing Test

In order to give a formal statement whether survey-based indicators or hard data perform better, we apply a standard forecast encompassing test. To keep it simple, we separately averaged the forecast errors from all soft (FE_{t+h}^{soft}) and all hard (FE_{t+h}^{hard}) indicators. With a forecast encompassing test, we can easily answer the question of whether a group of indicators (here: soft indicators) has more information content to forecast a target variable in comparison to the other group (here: hard data). The encompassing test follows the idea of Granger and Newbold (1973), who state that it is insufficient to compare only mean squared forecast errors between competing models. Their suggestion deals with the optimality of a forecast. The preferred forecast does not necessarily comprise all available information and is thus not optimal. This principle is known as "conditional efficiency". The preferred forecast encompasses the competitor, if the competing forecast has no more additional information (see Clements and Hendry, 1993). In our export case, we examine whether soft indicators (FE_{t+h}^{soft}) contain additional information compared to hard data (FE_{t+h}^{hard}). This can simply be answered with the following regression:

$$FE_{t+h}^{hard} = \lambda \left(FE_{t+h}^{hard} - FE_{t+h}^{soft} \right) + \varepsilon_{t+h} \,. \tag{6}$$

We apply standard ordinary-least-squares (OLS) with corrected standard errors in the style of Newey and West (1987). We test the null hypothesis $H_0 : \lambda = 0$. Whenever the test rejects the null, soft indicators contain more information than their competitors based on hard data.

3. Results

3.1. Pseudo Out-of-sample Analysis

Do soft or hard indicators best improve export growth forecasts? The very simple answer is that survey-based indicators do. Table 2 shows the pseudo out-of-sample results for all twenty European states in our study and the aggregates EA-18 and EU-28. The target variables are the growth rates of total exports yoy; an expanding window is applied (see *yoy, expanding* in the caption of Table 2). The table is divided into the two forecasting horizons ($h = 1, 2$). For every country and forecast horizon, the performance of each soft and hard indicator is presented. Whenever a cell is shown in gray, the specific indicator significantly outperforms the autoregressive benchmark model, thus, the relative root mean squared forecast error ($rRMSFE$) is smaller than one. A white-colored cell shows that the specific indicator has no higher forecast accuracy than the AR(p) process. Whenever an indicator series was too short for our forecasting purposes, a dash ("–") appears in the specific cell. Detailed results can be found in Table 7 in Appendix A.[7]

To summarize the large amount of information from Table 2, we compare the results in two different ways. First, we discuss performance differences across indicators. In a second step, we discuss country differences. Survey-based indicators beat the benchmark model quite often compared to hard indicators, since more cells for soft indicators are shown in gray. Turning to the indicator comparison, it is favorable to work with simple ranks. Therefore, we first assign country-specific ranks for each indicator. Then, we calculate average ranks for each indicator over all countries. This has been done for the two forecast horizons separately. For $h = 1$ the best indicator is the confidence indicator for the manufacturing sector (COF), followed by the specific economic sentiment indicator (ESI) and the production expectations (PEXP). For the larger forecast horizon ($h = 2$), COF and ESI change their positions. Again, production expectations are ranked in the third place. But how large are the forecast improvements of these indicators? We only discuss the results for $h = 1$. The results for $h = 2$ can also be found in Table 7 in Appendix A. For the COF, the improvement over the benchmark model ranges from 40% for Spain to more than 3% for Poland. In the case of ESI, the range runs from 35% for the EA-18 to 4% in Italy. The PEXP indicator outperforms the benchmark model of almost 35% for the EA-18 and nearly 8% for the Netherlands. The overall performance of the export expectations (EXEXP) indicator is rather poor in comparison to the three best indicators. From sixteen possible indicators, EXEXP ranks sixth for $h = 1$ and ninth for $h = 2$. The improvement of EXEXP ranges from 32% in Denmark to more than 5% in Sweden for the shorter forecast horizon.

[7] The results table in the appendix presents the $rRMSFE$ for all soft and hard indicators plus three additional benchmark models. One exception is the number for the AR(p) process: here we present the forecast errors in percentage points. Asterisks denote significant differences between the forecast errors based on the outcome of the Clark-West test.

Table 2: Pseudo Out-of-sample results for export growth (yoy, expanding)

	h=1															
	Soft indicator									**Hard indicator**						
Country	EXEXP	COF	EOBL	OBL	PEXP	SFP	IfW	ESI	CCOF	HCPI	ULCTOT	UWCMAN	GDPDEF	EXPI	PIPROD	PIPRODUS
Austria																
Bulgaria									–						–	
Czech Republic																
Denmark																
Estonia																
Finland																
France																
Germany																
Italy																
Latvia															–	
Lithuania									–						–	
Luxemburg																
Netherlands																
Poland									–							
Portugal																
Slovakia									–							
Slovenia																
Spain																
Sweden																
United Kingdom																
EA-18																
EU-28																
	h=2															
	Soft indicator									**Hard indicator**						
Country	EXEXP	COF	EOBL	OBL	PEXP	SFP	IfW	ESI	CCOF	HCPI	ULCTOT	UWCMAN	GDPDEF	EXPI	PIPROD	PIPRODUS
Austria																
Bulgaria									–						–	
Czech Republic																
Denmark																
Estonia																
Finland																
France																
Germany																
Italy																
Latvia															–	
Lithuania									–						–	
Luxemburg																
Netherlands																
Poland									–							
Portugal																
Slovakia									–							
Slovenia																
Spain																
Sweden																
United Kingdom																
EA-18																
EU-28																

Note: A gray-colored box indicates that the indicator significantly improves the autoregressive benchmark model. A white-colored box denotes no improvement. – not available.

The worst hard indicator is the real effective exchange rate based on unit wage costs in the manufacturing sector (UWCMAN). This indicator is ranked in sixteenth place for the shorter forecast horizon and in fifteenth place for $h = 2$. We ascertain that the hard indicators have in general a poorer forecast performance than the soft ones. However, there is one exception: the US industrial production. For $h = 1$ and $h = 2$, the average rank for PIPRODUS is four. This result is clearly indicated by the gray-colored boxes in Table 2. By focusing only on the price and cost competitiveness measures, the most intuitive candidate, a real effective exchange rate (REER) based on export price indices (EXPI), is the "less worse" one.

Now we deal with observable country differences. Since we have argued before that US industrial production performs well, we base our country comparison on the performance between soft indicators and the different price and cost competitiveness measures. For this purpose, we can summarize the countries in four possible groups: (i) only soft indicators can beat the benchmark model; (ii) only real effective exchange rates (REER) are better than the autoregressive process; (iii) at least one indicator from both groups work; (iv) no indicator delivers better results at all. Most of the countries fall into the first group. In eleven countries (or country aggregates), only soft indicators beat the benchmark model (see here and subsequent Table 3). There is no case where only the price and cost competitiveness measures are better than the autoregressive process. The third group consists of eight countries. In this group soft indicators as well as price and cost competitiveness measures beat the benchmark. There are three countries (Bulgaria, Latvia and Lithuania) where almost no indicator works at all. Especially for these Eastern Europe countries we find no improvement, with the exception of the industrial production of the United States (PIPRODUS) and $h = 2$ for Latvia, through any of our considered indicators. We have to conclude that especially in those three countries, the AR(p) process is a hard-to-beat benchmark model.

Table 3: Country differences between soft and hard indicators

Group	Countries
(i): only soft	Austria, Czech Republic, Denmark, France, Italy, Netherlands, Poland, Slovenia, Spain, EA-18, EU-28
(ii): only REER	–
(iii): soft and REER	Estonia, Finland, Germany, Luxemburg, Portugal, Slovakia, Sweden, United Kingdom
(iv): no indicator	Bulgaria, Latvia, Lithuania

As the analysis of the ranks revealed, we observe a high heterogeneity in the forecasting performance of soft and hard indicators between countries. In Section 3.4 we apply standard regression techniques to explain these differences. We especially ask whether the country-specific export composition is able to give some deeper insights into why certain groups of indicators work, while others do not.

3.2. Encompassing Test

Before we present some robustness checks as well as a discussion of why the forecast performance of indicators varies between countries, we first show that soft indicators perform better than hard data. Table 4 shows the forecast encompassing test results from Equation (6) for the two forecast horizons $h = 1$ and $h = 2$. Asterisks (for the standard significance levels 1%, 5% and 10%) indicate that soft indicators have significantly more information to forecast export growth in comparison to their hard counterparts.

Table 4: Encompassing results (yoy, expanding)

Country	**h=1**	**h=2**
Austria	***	***
Bulgaria	***	
Czech Republic	***	***
Denmark	***	***
Estonia		**
Finland	***	***
France	***	***
Germany	***	***
Italy	**	***
Latvia		***
Lithuania		
Luxemburg	***	***
Netherlands	***	***
Poland	***	***
Portugal	***	***
Slovakia		
Slovenia	***	***
Spain	***	***
Sweden	**	***
United Kingdom		
EA-18	***	***
EU-28	***	***

Note: Estimation with robust standard errors. ***, **, * indicate a p-value below the 1%, 5% or 10% level.

The table clearly underlines that soft indicators produce lower forecast errors than hard data for almost all of the countries in the sample. However, we observe some exceptions from this clear pattern. For Lithuania, Slovakia and the United Kingdom we find no significant differences between soft and hard indicators. Table 3 shows that Lithuania falls into group (iv), where no indicator works, and Slovakia and the United Kingdom are in group (iii), where soft as well as hard indicators work. For these three countries, regardless of whether the indicators improve forecast accuracy or not, no information advance of soft indicators exist. This is also the case for Estonia and Latvia by looking at the shorter forecast horizon ($h = 1$)

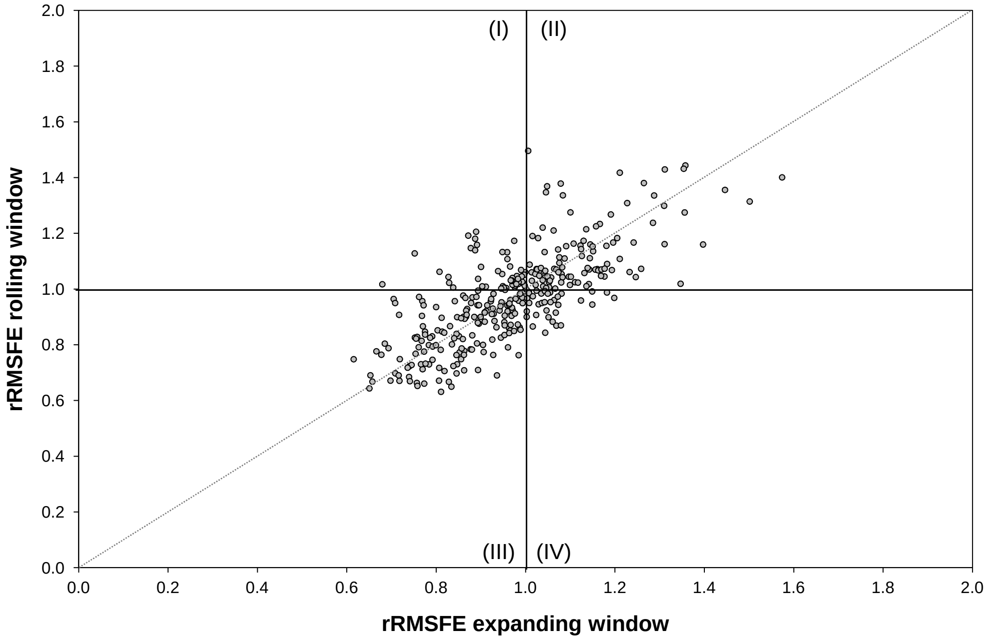
and for Bulgaria for $h = 2$. All in all, the results of the encompassing test strengthen the findings from the previous subsection. Whenever it comes to a practical application of export predictions, the forecaster should rely on soft indicators, especially the three mentioned above: the confidence indicator for the manufacturing sector (COF), the country-specific economic sentiment indicator (ESI) and the production expectations (PEXP).

3.3. Robustness Checks

To check the validity of our results, we present two types of robustness checks. We decided to check for robustness in two ways. First, we use a rolling window instead of applying an expanding window approach. This means that the initial estimation window for Equation (1) is not successively enlarged by one quarter but is rather fixed and moved forward by one quarter in each single step. Especially if breaks are present in the time series of export growth, the rolling window approach is more suitable. The advantage of the expanding window approach is its ability to capture the whole cyclicality of the underlying time series. In our second robustness check, we apply a different transformation of the target variable. Instead of using year-on-year growth rates, we calculate quarter-on-quarter (qoq) growth rates. Such a transformation captures the cyclical movement of the target variable during the year. In practice, forecasts of macroeconomic aggregates are usually based on the qoq transformation. Thus, we use this transformation as the second robustness check.

Let us first stick to the rolling window approach. Figure 1 shows a comparison of the relative root mean squared forecast errors ($rRMSFE$) for the short forecast horizon ($h = 1$); the target variables are yoy growth rates (yoy in the caption of Figure 1). Detailed results are available upon request. The rRMSFEs from the rolling window approach are drawn on the y-axis. The rRMSFEs from the expanding window approach can be found on the x-axis. Each dot represents an x-y-pair of an indicator for a specific country (e.g., performance EXEXP for Germany). To ease interpretation of the figure, we add the 45° line as well as a horizontal and vertical line, which both cross the value of the rRMSFE of one, thus, indicating whether an indicator performs better or worse compared to the specific benchmark model. Each dot below the 45° line means that the rRMSFE of the rolling window approach is lower than the one from an expanding window. The opposite holds for values above the 45° line. The horizontal and vertical lines divide the figure into four quadrants. The interpretations for quadrant (II) and (III) are straightforward. A dot in quadrant (II) stands for an indicator that produces a higher root mean squared forecast error (RMSFE) in comparison to the benchmark within the expanding as well as rolling window approach. The opposite holds for an indicator lying in quadrant (III), thus, producing a lower RMSFE in both approaches. Whenever an indicator enters quadrant (I) its performance becomes worse in a rolling window approach compared to an expanding window. For quadrant (IV) the indicator beats the benchmark in a rolling setup, whereas it fails to do so in the expanding approach.

Figure 1: Relative forecast errors in expanding vs. rolling window (yoy, h=1)



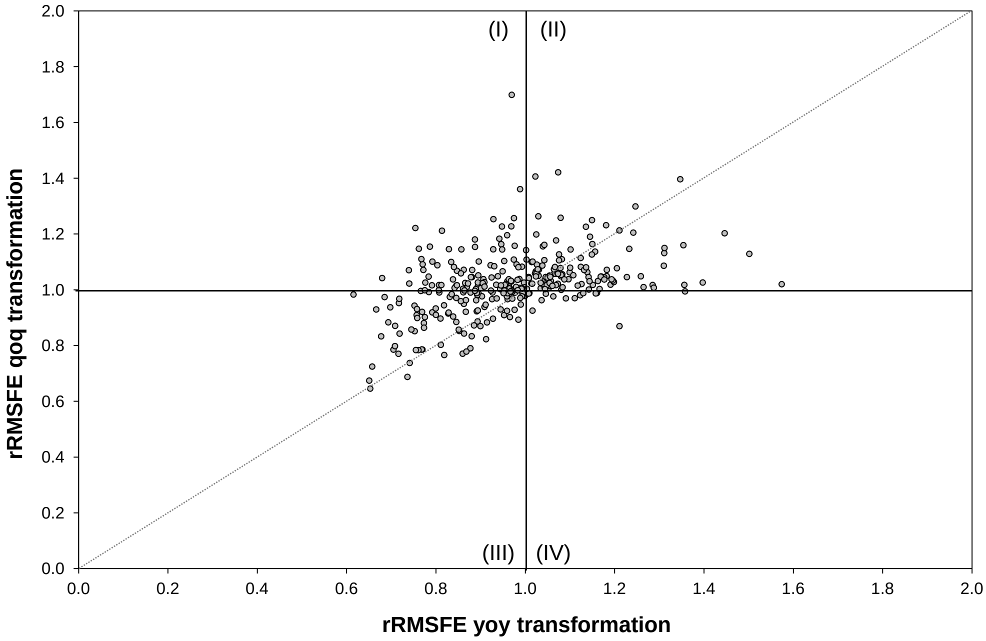
The results would be perfectly robust to the applied window if all dots lay on the 45° line. Figure 1 reveals that this is not the case. The results do not vary much between the two approaches, however, since the dots are located close to the 45° line. Only 24% of all indicators either become better or worse with the rolling window approach compared to the expanding window. However, most of these differences are not statistically significant. The remaining 76% remain either in quadrant (II) or (III). We conclude that the results are fairly robust for the shorter forecast horizon. The figure for the larger forecast horizon ($h = 2$) can be found in the Appendix (see Figure 3 in Section B). In that case, 30% of all results lie in either quadrant (I) or (IV). Still, 70% of all indicators stay robust in their relative performance. This is a confirmation of the results from the expanding window approach in Section 3.1.

The second robustness check is based on an alternative transformation of our target variable: qoq growth rates.[8] As for the rolling window, we present a similar figure as for the alternative transformation. For $h = 1$, Figure 2 compares the relative performance of the indicators in both transformations; the expanding window approach is applied (*expanding* in the caption of Figure 2). The results are not as robust as for the rolling window. 32% of all indicators change their relative performance for $h = 1$ by applying qoq instead of yoy growth rates. The bulk of these indicators are located in quadrant (I), thus, the relative performance worsens. For the larger forecast horizon ($h = 2$) even more indicators can be found in quadrant (I) or (IV). Nearly 42% change their relative performance between the two transformations (see Figure 4 in Appendix B). Almost 68% ($h = 1$) and 58% ($h = 2$) of

[8]All numerical results are available upon request.

all indicators keep their relative performance; thus, most of the findings remain the same. Additionally, qoq growth rates show a higher volatility compared to their yoy counterparts and are thus not that persistent. This fact makes them harder to predict. Gayer (2005) recommends clarifying to which reference series different survey indicators refer. This statement is directly transferable to our question. Do our indicators refer to yoy or qoq export growth rates? From the previous findings we suggest that most of the indicators clearly refer to yoy export growth rates. Whenever it comes to predicting exports of goods and services, the forecaster should rely on yoy instead of qoq growth rates.

Figure 2: Relative forecast errors yoy vs. qoq transformation (expanding, h=1)



3.4. Discussion of the Results

In the final step, we try to find some explanations for the high heterogeneity in performance between countries. Why do soft indicators work better in country A compared to country B? A similar question can be raised for hard indicators. To answer these questions, we run the following regression:

$$\overline{rRMSFE}_i^k = c^k + \beta_1 East_i + \beta_2 Service_i + \sum_{j=1}^{7} \beta_j SITC_i + \beta_8 HHI_i + \varepsilon_i^k \,. \tag{7}$$

First, we calculate the average $\overline{rRMSFE}$ of all soft (hard) indicators, here abbreviated with $k \in \{soft, hard\}$, for each country (i). Second, we ask the question of which variables may explain the differences in relative forecast errors. Since the sample is not too large, we end up with the composition of total exports. Therefore we use the average share of service exports in total exports (*Service*) between 2005 and 2013. Additionally, we add the shares of different product groups. For instance, Germany exports more cars, whereas the United Kingdom has a higher share in oil exports. Maybe it is easier for firm A to expect what their car exports will be instead of the highly uncertain or more volatile exports of oil from firm B. Thus, maybe the performance of soft and hard indicators depends crucially on the composition of exports and therefore the possibility of a firm to correctly anticipate future developments in foreign markets. In the end, we add average shares of seven different product groups based on the Standard International Trade Classification (*SITC*) between 2005 and 2013. The codes as well as the corresponding product groups can be found in Table 9 in the Appendix. Instead of using each single product group in the regression, we calculate a standard Hirschman-Herfindahl-Index (*HHI*) to measure the diversification of exports. At last, we add a Dummy for Eastern Europe countries in the sample (*East*). This dummy accounts for the observed differences in forecast performance between Eastern and Western Europe countries. We focus on the short forecast horizon $h = 1$, yoy export growth rates and an expanding window (see the caption of the following tables). Equation (7) is estimated with OLS and robust standard errors based on the Huber-White-Sandwich-Estimator.

Table 5 presents the regression results for the soft indicators. It should be noticed that we use the average values of the rRMSFE for each country so that we end up with twenty observations, one for each country, in the regression. All these results thus should be interpreted with caution since the number of observations is rather small. In the end, we estimate the model with only one SITC variable, in order not to not stress the few degrees of freedom. Therefore, the output tables contain eight columns, each for one single SITC group plus the HHI. For soft indicators, we find that the average rRMSFE is higher in Eastern Europe states than in non-Eastern Europe countries. We find no statistically significant correlation between the share of service exports and the relative performance of soft indicators. Hence, we expect that the performance of soft indicators is almost independent of the target variable. It seems to make no difference whether we forecast exports of goods, exports of services or the sum of both.[9]

Now let us turn to the SITC variables. Obviously the share of three product groups correlate with the relative performance of our soft indicators. These are: SITC24 – raw materials etc., SITC3 – mineral fuels etc. and SITC7 – machinery and transport equipment. Whenever a country has a higher export share in raw materials (0.674) or, for example, oil (0.791), the relative forecast performance of soft indicators worsens. Thus, it seems either harder for the firms to really anticipate future developments of exports or confidence indicators are not able, from a time series perspective, to grab export growth in a meaningful way. On the other hand, a higher share of machinery goods leads to a significant improvement in the forecasting performance of survey-based indicators. These three results are underpinned by

[9] We run our forecasting exercise for the two components of total exports as well. On average, we find no large difference, which explains the insignificant *Service* coefficient. Some performance differences do exist, but we do not want to discuss these results in detail; these results are available upon request.

the significant negative coefficient for the HHI. Since the HHI is coded in a way that a larger number represents a lower degree of diversification, the negative coefficient is interpreted as follows: the more diversified the exports of a country are, the better the performance of soft indicators.

Table 5: Composition of exports and performance of soft indicators (yoy, expanding, $h = 1$)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
East	0.094**	0.078*	0.080**	0.132**	0.095**	0.076**	0.100	0.108**
	(0.038)	(0.040)	(0.036)	(0.051)	(0.044)	(0.029)	(0.058)	(0.038)
Service	0.048	0.035	0.077	0.118	0.050	0.001	0.059	0.078
	(0.102)	(0.096)	(0.084)	(0.115)	(0.145)	(0.061)	(0.125)	(0.077)
SITC01	0.430							
	(0.407)							
SITC24		0.674*						
		(0.354)						
SITC3			0.791***					
			(0.212)					
SITC5				0.732				
				(0.494)				
SITC68					0.047			
					(0.365)			
SITC7						-0.471***		
						(0.092)		
SITC9							0.133	
							(2.273)	
HHI								-0.630**
								(0.228)
c	0.832***	0.846***	0.805***	0.756***	0.853***	1.057***	0.861***	1.028***
	(0.049)	(0.042)	(0.046)	(0.086)	(0.099)	(0.048)	(0.070)	(0.060)
R^2	0.302	0.313	0.467	0.330	0.258	0.527	0.258	0.411
Obs.	20	20	20	20	20	20	20	20

Note: Robust standard errors are in parentheses.
***, **, * indicate statistical significance at the 1%, 5%, 10% level.

The same exercise can be done for the average performance of our hard indicators; Table 6 shows the corresponding results. We find no significant difference for Eastern Europe countries and no impact of the share in service exports. The composition of goods exports seems to matter only in a minor way. Only a higher share of products in the group SITC01 (food, beverages and tobacco) seems to worsen the relative performance of hard indicators.

All in all the composition of exports seem to matter for the relative performance of indicators. However, we suspect that firm characteristics in particular explain these observed country differences, i.e., firm samples of each country over time would offer a rich source of variation. With this information, future research activities could either analyze the number of exporting firms or their corresponding characteristics could explain our observed differences in forecasting performance. To the best of our knowledge, no European study exists that links firm-level information to the macroeconomic forecasting performance of survey-based indicators. However, there is some literature which links so called non-responses of firms to the accuracy of survey-based indicators (for Germany see Seiler, 2014).

Table 6: Composition of exports and performance of hard indicators (yoy, expanding, $h = 1$)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
East	-0.033	-0.014	-0.024	0.027	-0.016	-0.033	-0.070	-0.018
	(0.048)	(0.056)	(0.052)	(0.047)	(0.044)	(0.051)	(0.072)	(0.051)
Service	-0.156	-0.120	-0.132	-0.045	-0.104	-0.156	-0.142	-0.122
	(0.119)	(0.102)	(0.112)	(0.124)	(0.107)	(0.108)	(0.103)	(0.119)
SITC01	1.002**							
	(0.426)							
SITC24		-0.353						
		(0.508)						
SITC3			-0.018					
			(0.349)					
SITC5				1.072				
				(0.754)				
SITC68					-0.162			
					(0.387)			
SITC7						-0.193		
						(0.212)		
SITC9							-2.370	
							(1.394)	
HHI								-0.322
								(0.321)
c	1.035***	1.120***	1.112***	0.952***	1.150***	1.190***	1.170***	1.195***
	(0.052)	(0.061)	(0.071)	(0.103)	(0.127)	(0.094)	(0.079)	(0.097)
R^2	0.246	0.050	0.037	0.172	0.046	0.076	0.102	0.072
Obs.	20	20	20	20	20	20	20	20

Note: Robust standard errors are in parentheses.
***, **, * indicate statistical significance at the 1%, 5%, 10% level.

Another explanation for the country differences could lie in the aggregation of firm responses. In this paper, we use a rather standard measure: the balances of positive and negative responses. However, there is a broad discussion on the usefulness of balances (see Croux *et al.*, 2005; Claveria *et al.*, 2007). Future research activities could focus on a sensitivity analysis with respect to different aggregation methods.

Finally, we return to the discussion brought forward by Gayer (2005). He asks which survey indicator refers to which specific reference series. The European Commission also provides survey indicators for different sub-sectors in manufacturing. Since the discussion before reveals the fact that the export composition matters for the relative performance of soft indicators, maybe sectoral results are more closely linked to total export growth. However, we leave all these issues for follow-up studies.

4. Summary and Conclusion

Macroeconomic forecasts consist of more than the prediction of a single number, namely gross domestic product (GDP). In practice it is standard to forecast each single component (e.g. exports) of total output. Disaggregated GDP forecasts are also in the academic literature seen as more accurate than direct predictions. Thus, better forecasts on each single component lead, c.p., to lower forecast errors for GDP. In this paper we concentrate on one major aggregate in total output: exports of goods and services. In conclusion, do soft or hard indicators have better predictive power for export growth? This paper evaluates this question with pseudo out-of-sample techniques and forecast-encompassing tests for twenty single European states and the aggregates EA-18 and EU-28. Our period of investigation runs from the first quarter 1996 to the fourth quarter of 2013 and therefore covers more than one business cycle. For most of our countries we find a significant improvement in forecast accuracy through survey-based indicators. Hard indicators such as price and cost competitiveness measures are only in a few cases able to beat the benchmark model. One exception of a hard indicator is US industrial production, which is a tough competitor compared to the soft indicators. Two robustness checks confirm our results.

All in all, we find remarkable differences in forecast accuracy between the countries in the sample. We therefore ask: what are the reasons for these country differences? It turns out that export composition in particular has an impact on the forecast accuracy of survey-based indicators. The relative performance of soft indicators worsens the higher the export shares of raw materials or oil become. The opposite holds for a higher share in machinery exports. For hard indicators, we find only weak results for the export composition.

This paper expands the discussion on export forecasts in several ways. First, we use a multitude of indicators for the forecasting exercise and employ a competition between soft and hard data. Second, we analyze this competition for a multitude of European states, thus broadening the picture of the usefulness of indicators for export forecasts. Third, we implicitly stick to the discussion by Claveria *et al.* (2007) by searching for the reasons for observed country differences. We find that the accuracy of soft indicators depends on export composition. However, further investigation of this result is needed. Finally, this paper gives some suggestions for future research activities to develop a broader understanding of the forecasting power of survey results for exports in particular and different macroeconomic variables in general.

Acknowledgments: Special thanks go to Steffen R. Henzel, Tobias Lohse, Marcel Thum, Michael Weber and Klaus Wohlrabe for their helpful comments and suggestions. We are also grateful to seminar participants at the Technische Universität Dresden and the ifo/CES Christmas Conference 2014.

References

Angelini, E., Banbura, M. and Rünstler, G. (2010). Estimating and forecasting the euro area monthly national accounts from a dynamic factor model. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, **2010** (1), 5–26.

Baghestani, H. (1994). Evaluating multiperiod survey forecasts of real net exports. *Economics Letters*, **44** (3), 267–272.

Cardoso, F. and Duarte, C. (2006). *The use of qualitative information for forecasting exports*. Banco de Portugal Economic Bulletin Winter 2006.

Ca'Zorzi, M. and Schnatz, B. (2010). Explaining and forecasting euro area exports: which competitiveness indicator performs best? In P. de Grauwe (ed.), *Dimensions of Competitiveness*, CESifo Seminar Series September 2010, MIT Press, Cambridge, pp. 121–148.

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predicitve accuracy in nested models. *Journal of Econometrics*, **138** (1), 291–311.

Claveria, O., Pons, E. and Ramos, R. (2007). Business and consumer expectations and macroeconomic forecasts. *International Journal of Forecasting*, **23** (1), 47–69.

Clements, H. P. and Hendry, D. F. (1993). On the limitations of comparing mean squared forecast errors. *Journal of Forecasting*, **12** (8), 617–637.

Croux, C., Dekimpe, M. G. and Lemmens, A. (2005). On the predictive content of production surveys: A pan-European study. *International Journal of Forecasting*, **21** (2), 363–375.

Department of Economic and Social Affairs of the United Nations (2006). *Standard International Trade Classification, Revision 4*. Statistical Papers Series M No. 34/Rev. 4.

Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, **13** (3), 253–263.

Drechsel, K. and Scheufele, R. (2012). *Bottom-up or direct? Forecasting German GDP in a data-rich environment*. Swiss national bank working papers 2012-16.

Elstner, S., Grimme, C. and Haskamp, U. (2013). Das ifo Exportklima – ein Fruehindikator fuer die deutsche Exportprognose. *ifo Schnelldienst*, **66** (4), 36–43.

European Commission Economic and Financial Affairs (2014). *The Joint Harmonised EU Programme of Business and Consumer Surveys – User Guide*. Directorate General for Economic and Financial Affairs.

Fiorito, R. and Kollintzas, T. (1994). Stylized facts of business cycles in the G7 from a real business cycles perspective. *European Economic Review*, **38** (2), 235–269.

Frale, C., Marcellino, M., Mazzi, G. L. and Proietti, T. (2010). Survey Data as Coincident or Leading Indicators. *Journal of Forecasting*, **29** (1-2), 109–131.

Frankel, J. A. and Romer, D. (1999). Does Trade Cause Growth? *The American Economic Review*, **89** (3), 379–399.

Gayer, C. (2005). Forecast Evaluation of European Commission Survey Indicators. *Journal of Business Cycle Measurement and Analysis*, **2005** (2), 157–183.

Granger, C. W. J. and Newbold, P. (1973). Some comments on the evaluation of economic forecasts. *Applied Economics*, **5** (1), 35–47.

Guichard, S. and Rusticelli, E. (2011). *A Dynamic Factor Model for World Trade Growth*. OECD Economics Department Working Papers No. 874.

Hanslin, S. and Scheufele, R. (2014). *Foreign PMIs: A reliable indicator for Swiss exports*. Paper presented at the 2014 IWH Workshop Nowcasting and Forecasting, mimeo.

Jannsen, N. and Richter, J. (2012). Kapazitätsauslastung im Ausland als Indikator für die deutschen Investitionsgüterexporte. *Wirtschaftsdienst*, **92** (12), 833–837.

Keck, A., Raubold, A. and Truppia, A. (2009). Forecasting International Trade: A Time Series Approach. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, **2009** (2), 157–176.

Lehmann, R. and Weyh, A. (2014). *Forecasting employment in Europe: Are survey results helpful?* Ifo Working Paper No. 182.

Newey, W. K. and West, K. D. (1987). A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55** (3), 703–708.

Robinzonov, N. and Wohlrabe, K. (2010). Freedom of Choice in Macroeconomic Forecasting. *CESifo Economic Studies*, **56** (2), 192–220.

Seiler, C. (2014). On the Robustness of the Balance Statistics with respect to Nonresponse. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, forthcoming.

Vosen, S. and Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting*, **30** (6), 565–578.

Wang, C., Hsu, Y. and Liou, C. (2011). A comparison of ARIMA forecasting and heuristic modelling. *Applied Financial Economics*, **21** (15), 1095–1102.

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

A. Out-of-sample Results Expanding Window

Table 7: Detailed out-of-sample results for export growth (yoy, expanding)

Model	h=1	h=2	Model	h=1	h=2
	Austria			**Bulgaria**	
AR(p) in %	3.899	6.001	AR(p) in %	14.621	14.879
AR(1)	1.692	1.421	AR(1)	0.875	1.067
ISM	2.053	1.351	ISM	0.871	0.871
RW	1.560	1.414	RW	1.053	1.321
EXEXP	0.925*	0.990**	EXEXP	1.160	1.142
COF	0.852**	0.863**	COF	1.009	1.587
EOBL	0.960**	1.070	EOBL	1.073	1.113
OBL	0.984**	1.108	OBL	1.034	1.079
PEXP	0.769**	0.824**	PEXP	1.080	1.159
SFP	0.908**	1.044	SFP	1.006	1.192
IfW	1.211	1.395	IfW	1.242	1.289
ESI	0.851**	0.888**	ESI	1.264	1.656
CCOF	1.016	1.041	CCOF	–	–
HCPI	1.051	1.052	HCPI	1.151	1.075
ULCTOT	1.064	1.052	ULCTOT	1.166	1.178
UWCMAN	1.054	1.061	UWCMAN	1.357	1.442
GDPDEF	1.044	1.053	GDPDEF	1.157	1.215
EXPI	1.067	1.087	EXPI	1.191	1.264
PIPROD	0.999	1.011	PIPROD	–	–
PIPRODUS	0.997	1.017	PIPRODUS	1.198	1.374
	Czech Republic			**Denmark**	
AR(p) in %	8.793	9.322	AR(p) in %	5.363	6.279
AR(1)	1.010	1.088	AR(1)	1.015	1.023
ISM	1.122	1.070	ISM	1.229	1.064
RW	1.031	1.276	RW	1.038	1.153
EXEXP	0.847*	0.976**	EXEXP	0.677*	0.654*
COF	0.863*	0.963**	COF	0.845*	0.841*
EOBL	0.986	1.067	EOBL	0.800*	0.763*
OBL	0.982*	1.112	OBL	0.773*	0.751*
PEXP	0.791*	0.871**	PEXP	0.708*	0.685*
SFP	1.002	1.075	SFP	1.052	1.078
IfW	1.022	1.153	IfW	1.044	1.120
ESI	0.693*	0.818**	ESI	0.862**	0.802**
CCOF	0.960*	1.069	CCOF	0.976	0.908**
HCPI	1.007	1.153	HCPI	1.082	1.002
ULCTOT	1.123	1.113	ULCTOT	1.064	1.029
UWCMAN	1.142	1.092	UWCMAN	1.156	1.245
GDPDEF	1.111	1.093	GDPDEF	1.124	1.161
EXPI	1.130	1.069	EXPI	1.066	1.236
PIPROD	0.740**	0.998*	PIPROD	0.976	0.949**
PIPRODUS	0.959*	1.068	PIPRODUS	0.757*	0.742*
	Estonia			**Finland**	
AR(p) in %	10.628	14.532	AR(p) in %	10.461	11.457
AR(1)	1.285	1.165	AR(1)	1.053	1.059
ISM	1.311	1.155	ISM	1.130	1.041
RW	1.310	1.299	RW	1.122	1.290
EXEXP	0.968	1.024	EXEXP	0.867*	0.937**
COF	0.942	0.835**	COF	0.698**	0.760**
EOBL	0.807***	0.913**	EOBL	0.806*	0.820*
OBL	0.948**	0.991	OBL	0.807**	0.841*
PEXP	1.001	1.030	PEXP	0.784**	0.903*

Continued on next page...

Table 7: Out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	Model	h=1	h=2
SFP	0.911*	0.935**	SFP	0.903**	1.114
IfW	0.856**	0.966**	IfW	0.783	0.981
ESI	0.881*	0.824**	ESI	0.766**	0.783**
CCOF	1.024	0.956*	CCOF	0.775*	0.774**
HCPI	0.984**	0.824***	HCPI	1.030	1.057
ULCTOT	0.980	1.069	ULCTOT	0.938*	1.001
UWCMAN	0.970	0.947*	UWCMAN	0.812**	0.928**
GDPDEF	1.052	1.035	GDPDEF	1.053	1.023
EXPI	0.895***	0.940**	EXPI	1.057	1.085
PIPROD	1.039	0.890***	PIPROD	0.906*	0.956**
PIPRODUS	0.834*	0.883*	PIPRODUS	0.792*	0.920*
	France			**Germany**	
AR(p) in %	4.100	5.335	AR(p) in %	7.180	8.137
AR(1)	1.256	1.187	AR(1)	1.095	1.144
ISM	1.523	1.184	ISM	1.200	1.070
RW	1.278	1.319	RW	1.100	1.283
EXEXP	0.926*	1.097	EXEXP	0.718**	0.917**
COF	0.705**	0.745*	COF	0.716**	0.805**
EOBL	0.945**	0.947*	EOBL	0.773**	0.848*
OBL	0.810**	0.850*	OBL	0.752**	0.844*
PEXP	0.768**	0.749*	PEXP	0.811**	0.802**
SFP	0.769**	0.864**	SFP	0.745**	0.862**
IfW	1.347	1.319	IfW	0.857*	1.009
ESI	0.708**	0.693*	ESI	0.845**	0.919*
CCOF	0.827**	0.817*	CCOF	0.963	0.983*
HCPI	1.285	1.180	HCPI	1.083	1.077
ULCTOT	1.356	1.264	ULCTOT	1.035	1.052
UWCMAN	1.574	1.429	UWCMAN	0.924**	0.998*
GDPDEF	1.287	1.206	GDPDEF	1.097	1.095
EXPI	1.083	1.216	EXPI	1.141	1.109
PIPROD	1.311	1.048	PIPROD	1.149	1.140
PIPRODUS	0.862**	1.003	PIPRODUS	0.776**	0.916*
	Italy			**Latvia**	
AR(p) in %	5.923	7.323	AR(p) in %	6.894	8.660
AR(1)	1.254	1.235	AR(1)	1.191	1.136
ISM	1.463	1.197	ISM	1.416	1.142
RW	1.272	1.384	RW	1.246	1.305
EXEXP	0.994	1.052	EXEXP	1.124	1.035
COF	0.899*	0.925*	COF	0.990	0.957
EOBL	0.891*	1.026	EOBL	0.935	0.868**
OBL	0.952*	0.982*	OBL	1.091	1.044
PEXP	0.914*	0.953*	PEXP	1.069	1.029
SFP	0.971	0.966	SFP	1.046	0.994*
IfW	1.149	1.380	IfW	1.013	0.964**
ESI	0.959*	0.926*	ESI	1.007	1.002
CCOF	0.992	0.990	CCOF	1.100	1.036
HCPI	1.080	1.060	HCPI	1.023	1.031
ULCTOT	1.055	1.029	ULCTOT	1.022	1.039
UWCMAN	1.069	1.039	UWCMAN	1.049	1.047
GDPDEF	1.080	1.039	GDPDEF	0.997	1.009
EXPI	1.136	1.144	EXPI	0.998	0.998
PIPROD	0.974*	0.915*	PIPROD	–	–
PIPRODUS	0.953*	0.987	PIPRODUS	0.946*	0.963**

Continued on next page...

Table 7: Out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	Model	h=1	h=2
Lithuania			**Luxemburg**		
AR(p) in %	7.705	9.590	AR(p) in %	6.891	8.154
AR(1)	1.248	1.163	AR(1)	1.067	1.050
ISM	1.338	1.084	ISM	1.230	1.051
RW	1.384	1.489	RW	1.086	1.189
EXEXP	1.030	1.111	EXEXP	0.869**	0.936**
COF	1.145	1.241	COF	0.860**	0.945**
EOBL	1.181	1.285	EOBL	0.845***	0.868***
OBL	1.150	1.179	OBL	0.861***	0.928**
PEXP	1.075	1.214	PEXP	0.893**	1.022
SFP	1.075	1.094	SFP	0.865***	0.918**
IfW	1.003	0.991	IfW	0.813	1.001
ESI	1.072	1.229	ESI	0.791**	0.918**
CCOF	–	–	ESI	0.791**	0.918**
HCPI	1.015	1.036	HCPI	1.047	1.014
ULCTOT	1.024	1.033	ULCTOT	0.900**	0.924*
UWCMAN	1.045	1.024	UWCMAN	1.205	1.226
GDPDEF	1.014	1.039	GDPDEF	0.965*	0.965
EXPI	1.042	1.076	EXPI	0.959**	0.979
PIPROD	–	–	PIPROD	0.863*	0.920*
PIPRODUS	1.232	1.167	PIPRODUS	0.922	0.973
Netherlands			**Poland**		
AR(p) in %	4.024	5.202	AR(p) in %	7.307	7.803
AR(1)	1.299	1.198	AR(1)	0.957*	0.967***
ISM	1.443	1.128	ISM	1.055	0.999*
RW	1.351	1.418	RW	1.011	1.244
EXEXP	0.944***	0.978**	EXEXP	0.754***	0.881**
COF	0.912**	0.990*	COF	0.966*	1.132
EOBL	0.965**	0.990*	EOBL	0.963*	1.069
OBL	0.892**	1.041	OBL	0.960**	1.069
PEXP	0.923**	1.014	PEXP	0.772***	0.891**
SFP	0.893**	0.904*	SFP	0.761**	0.943**
IfW	1.135	1.266	IfW	0.969	1.381
ESI	0.775**	0.837**	ESI	1.029	1.167
CCOF	0.841***	0.929***	CCOF	–	–
HCPI	1.144	1.068	HCPI	1.038	1.175
ULCTOT	1.126	1.084	ULCTOT	1.211	1.208
UWCMAN	1.196	1.043	UWCMAN	1.397	1.438
GDPDEF	1.310	1.153	GDPDEF	1.227	1.174
EXPI	1.062	1.026	EXPI	1.048	1.154
PIPROD	1.043	0.979	PIPROD	0.818*	1.012
PIPRODUS	0.896**	1.039	PIPRODUS	1.023	1.113
Portugal			**Slovakia**		
AR(p) in %	6.592	7.129	AR(p) in %	11.158	12.949
AR(1)	1.042	1.102	AR(1)	0.986	0.995
ISM	1.130	1.056	ISM	1.105	0.965
RW	1.090	1.310	RW	1.010	1.126
EXEXP	0.887*	1.054	EXEXP	0.948**	0.979
COF	0.799*	0.926	COF	0.767**	0.901**
EOBL	0.739**	0.890*	EOBL	1.038	1.095
OBL	0.786**	0.893*	OBL	1.026	1.045
PEXP	0.831*	1.042	PEXP	0.685**	0.812**
SFP	0.963	1.052	SFP	0.872	0.886
IfW	1.073	1.191	IfW	0.887*	1.236
ESI	0.867*	1.003	ESI	0.930***	0.940**

Continued on next page...

Table 7: Out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	Model	h=1	h=2
CCOF	0.992	1.155	CCOF	–	–
HCPI	0.993	1.077	HCPI	0.877	0.898*
ULCTOT	0.966	1.047	ULCTOT	0.889	0.908
UWCMAN	1.054	1.061	UWCMAN	1.078	1.150
GDPDEF	0.964	1.053	GDPDEF	0.891	0.898
EXPI	0.953	1.043	EXPI	0.887	0.905
PIPROD	0.908	1.030	PIPROD	0.894	0.911
PIPRODUS	0.928*	1.045	PIPRODUS	0.829**	0.837**
	Slovenia			**Spain**	
AR(p) in %	7.247	9.321	AR(p) in %	6.044	7.267
AR(1)	1.197	1.124	AR(1)	1.022	1.027
ISM	1.387	1.092	ISM	1.192	1.003
RW	1.156	1.203	RW	1.044	1.114
EXEXP	0.984*	1.041	EXEXP	0.894*	0.928*
COF	0.828**	0.822***	COF	0.615**	0.549**
EOBL	0.893***	0.887***	EOBL	0.717**	0.659**
OBL	0.758***	0.885***	OBL	0.752**	0.699**
PEXP	0.847**	0.973**	PEXP	0.770**	0.682**
SFP	0.989**	0.968**	SFP	0.666**	0.619**
IfW	1.079	1.187	IfW	0.928**	0.962*
ESI	0.936***	0.923***	ESI	0.679**	0.568**
CCOF	1.036	1.090	CCOF	0.838**	0.695*
HCPI	1.050	1.047	HCPI	1.099	0.987*
ULCTOT	1.002	1.004	ULCTOT	1.182	1.001
UWCMAN	1.073	1.065	UWCMAN	1.182	1.022
GDPDEF	1.056	1.027	GDPDEF	1.139	1.033
EXPI	1.008	1.036	EXPI	1.131	1.045
PIPROD	1.101	1.207	PIPROD	0.988*	0.936*
PIPRODUS	0.974	1.026	PIPRODUS	0.803**	0.928*
	Sweden			**United Kingdom**	
AR(p) in %	4.665	6.407	AR(p) in %	8.428	7.967
AR(1)	1.355	1.268	AR(1)	0.953*	1.064
ISM	1.667	1.229	ISM	0.944*	1.006
RW	1.307	1.305	RW	1.042	1.301
EXEXP	0.949***	0.933**	EXEXP	0.988	1.065
COF	0.905**	0.959**	COF	1.034	1.050
EOBL	1.060	1.079	EOBL	1.107	1.063
OBL	0.757**	0.872**	OBL	1.087	1.079
PEXP	0.953	0.951*	PEXP	0.991	1.032
SFP	1.016	1.018	SFP	1.044	1.058
IfW	0.976*	1.033	IfW	1.028	1.195
ESI	0.835**	0.838**	ESI	0.979	0.997
CCOF	0.881**	0.870**	CCOF	0.890**	0.865*
HCPI	0.969*	1.010	HCPI	0.970*	0.978
ULCTOT	0.957*	1.010	ULCTOT	0.969**	0.998
UWCMAN	0.984	1.029	UWCMAN	0.951**	0.978
GDPDEF	0.968*	1.009	GDPDEF	0.967**	0.991
EXPI	0.954**	1.007	EXPI	0.983*	0.991
PIPROD	0.978*	0.995	PIPROD	0.945*	1.010
PIPRODUS	0.878**	0.907**	PIPRODUS	0.908*	0.950
	EA-18			**EU-28**	
AR(p) in %	4.892	6.317	AR(p) in %	4.953	6.160
AR(1)	1.325	1.239	AR(1)	1.274	1.223
ISM	1.488	1.166	ISM	1.427	1.160
RW	1.309	1.368	RW	1.267	1.366

Continued on next page...

Table 7: Out-of-sample results for export growth (yoy, expanding) – continued

Model	h=1	h=2	**Model**	h=1	h=2
EXEXP	0.718**	0.983*	EXEXP	0.838**	0.920**
COF	0.736**	0.759**	COF	0.818*	0.800*
EOBL	0.877*	0.937*	EOBL	0.927*	0.992
OBL	0.859**	0.839*	OBL	0.880*	0.882*
PEXP	0.653**	0.682**	PEXP	0.741*	0.717*
SFP	0.868**	0.870**	SFP	0.885*	0.885*
IfW	1.246	1.438	IfW	1.042	1.174
ESI	0.650**	0.627**	ESI	0.657**	0.632**
CCOF	0.761**	0.750**	CCOF	0.755**	0.715**
HCPI	1.176	1.095	HCPI	1.354	1.341
ULCTOT	1.162	1.084	ULCTOT	1.258	1.125
UWCMAN	1.117	1.050	UWCMAN	1.193	1.110
GDPDEF	1.170	1.102	GDPDEF	1.311	1.406
EXPI	1.169	1.111	EXPI	1.177	1.105
PIPROD	1.501	1.315	PIPROD	1.446	1.233
PIPRODUS	0.856**	0.954*	PIPRODUS	0.840**	0.943**

Note: The table presents the relative root mean squared forecast errors (*rRMSFE*) of the different models and the benchmark. The row AR(p) in % shows the *RMSFE* 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.

B. Further Results and Additional Material

Table 8: Encompassing results (yoy, rolling)

Country	h=1	h=2
Austria	* * *	* * *
Bulgaria	**	
Czech Republic	* * *	* * *
Denmark	**	**
Estonia		* * *
Finland	* * *	* * *
France	* * *	* * *
Germany	* * *	* * *
Italy		* * *
Latvia		* * *
Lithuania	*	
Luxemburg	* * *	* * *
Netherlands	* * *	* * *
Poland	* * *	**
Portugal	* * *	*
Slovakia	* * *	* * *
Slovenia	**	* * *
Spain	* * *	* * *
Sweden	* * *	* * *
United Kingdom		
EA-18	* * *	* * *
EU-28	* * *	* * *

Note: Estimation with robust standard errors. ***, **, * indicate a p-value below the 1%, 5% or 10% level.

Figure 3: Relative forecast errors in expanding vs. rolling window (yoy, h=2)

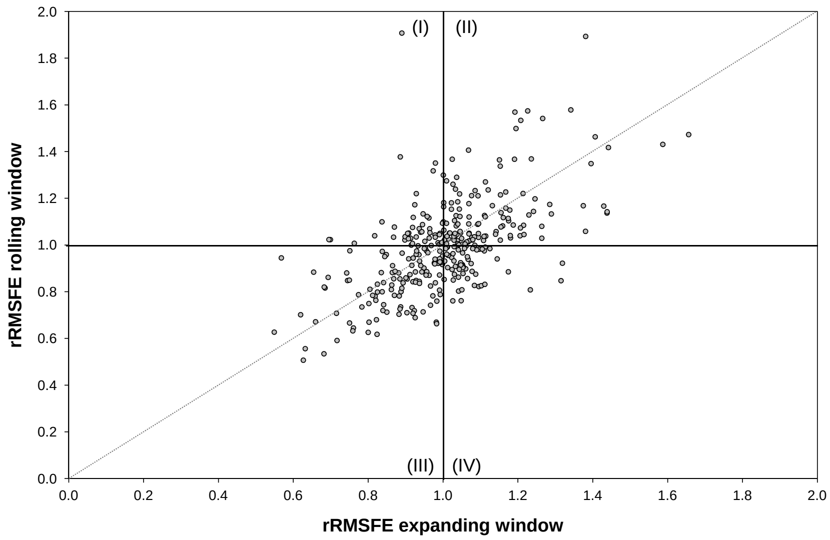


Figure 4: Relative forecast errors yoy vs. qoq transformation (expanding, h=2)

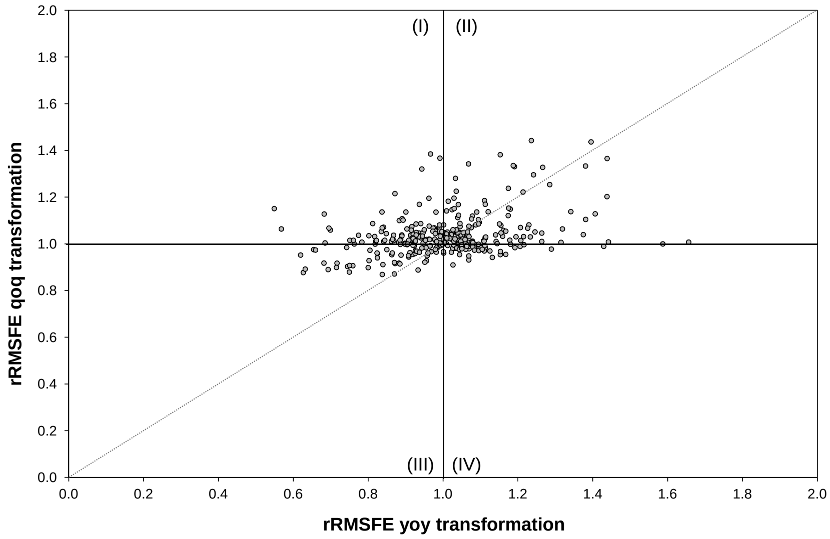


Table 9: SITC codes and product groups

Code	**Product group**
SITC01	Food and live animals, beverages and tobacco
SITC24	Crude materials, inedible, except fuels, animal and vegetable oils, fats and waxes
SITC3	Mineral fuels, lubricants and related materials
SITC5	Chemicals and related products, n.e.s.
SITC68	Manufactured goods classified chiefly by material, miscellaneous manufactured articles
SITC7	Machinery and transport equipment
SITC9	Commodities and transactions not classified elsewhere in the SITC

Source: Department of Economic and Social Affairs of the United Nations (2006).

Ifo Working Papers

No. 195 Fabritz, N., ICT as an Enabler of Innovation: Evidence from German Microdata, January 2015.

No. 194 Kauder, B. and N. Potrafke, Just hire your spouse! Evidence from a political scandal in Bavaria, December 2014.

No. 193 Seiler, C., Mode Preferences in Business Surveys: Evidence from Germany, November 2014.

No. 192 Kleemann, M. and M. Wiegand, Are Real Effects of Credit Supply Overestimated? Bias from Firms' Current Situation and Future Expectations, November 2014.

No. 191 Kauder, B, Spatial Administrative Structure and Intra-Metropolitan Tax Competition, October 2014.

No. 190 Auer, W. and N. Danzer, Fixed-Term Employment and Fertility: Evidence from German Micro Data, October 2014.

No. 189 Rösel, F., Co-Partisan Buddies or Partisan Bullies? Why State Supervision of Local Government Borrowing Fails, October 2014.

No. 188 Kauder, B., Incorporation of Municipalities and Population Growth – A Propensity Score Matching Approach, October 2014.

No. 187 Méango, R., Financing Student Migration: Evidence for a Commitment Problem, September 2014.

No. 186 Nagl, W. and M. Weber, Unemployment compensation and unemployment duration before and after the German Hartz IV reform, September 2014.

No. 185 Potrafke, N. and M. Reischmann, Explosive Target balances of the German Bundesbank, July 2014.

No. 184 Eppinger, P.S. and G.J. Felbermayr, Bilateral Trade and Similarity of Income Distributions: The Role of Second Moments, July 2014.

No. 183 Wohlrabe, K., Das FAZ-Ökonomenranking 2013: Eine kritische Betrachtung, Juli 2014.

No. 182 Lehmann, R. and A. Weyh, Forecasting employment in Europe: Are survey result helpful?, June 2014.

No. 181 Schinke, C., Government Ideology, Globalization, and Top Income Shares in OECD Countries, June 2014.

No. 180 Benz, S., M. Larch and M. Zimmer, The Structure of the German Economy, May 2014.

No. 179 Meier, V. and H. Rainer, Pigou Meets Ramsey: Gender-Based Taxation with Non-Cooperative Couples, May 2014.

No. 178 Kugler, F., G. Schwerdt und L. Wößmann, Ökonometrische Methoden zur Evaluierung kausaler Effeke der Wirtschaftspolitik, April 2014.

No. 177 Angerer, S., D. Glätzle-Rützler, P. Lergetporer and M. Sutter, Donations, risk attitudes and time preferences: A study on altruism in primary school children, March 2014.

No. 176 Breuer, C., On the Rationality of Medium-Term Tax Revenue Forecasts: Evidence from Germany, March 2014.

No. 175 Reischmann, M., Staatsverschuldung in Extrahaushalten: Historischer Überblick und Implikationen für die Schuldenbremse in Deutschland, März 2014.

No. 174 Eberl, J. and C. Weber, ECB Collateral Criteria: A Narrative Database 2001–2013, February 2014.

No. 173 Benz, S., M. Larch and M. Zimmer, Trade in Ideas: Outsourcing and Knowledge Spillovers, February 2014.

No. 172 Kauder, B., B. Larin und N. Potrafke, Was bringt uns die große Koalition? Perspektiven der Wirtschaftspolitik, Januar 2014.

No. 171 Lehmann, R. and K. Wohlrabe, Forecasting gross value-added at the regional level: Are sectoral disaggregated predictions superior to direct ones?, December 2013.

No. 170 Meier, V. and I. Schiopu, Optimal higher education enrollment and productivity externalities in a two-sector-model, November 2013.