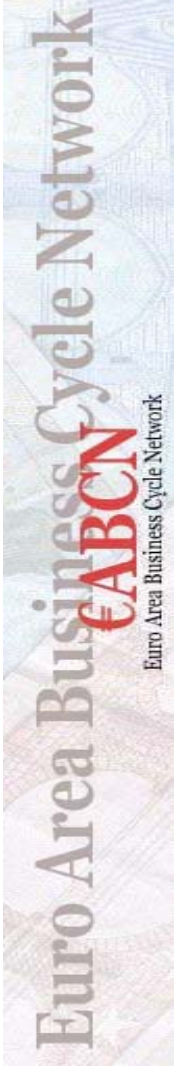




Centre for
Economic
Policy
Research



Centre for
Financial
Analysis &
Policy



CAMBRIDGE
Judge Business School

Euro Area Business Cycle Network (EABCN) Workshop Using Euro Area Data: Issues and Consequences for Economic Analysis

Cambridge, 26-28 March 2008

Hosted by University of Cambridge

Forecasting Euro Area Real GDP: Optimal Pooling of Information

Johannes Mayr

We are grateful to the following institutions for their financial and organizational support: University of Cambridge and the Euro Area Business Cycle Network

The views expressed in this paper are those of the author(s) and not those of the funding organization(s) or of CEPR, which takes no institutional policy positions.

Forecasting Euro Area Real GDP: Optimal Pooling of Information

Oliver Hülsewig* Johannes Mayr* Timo Wollmershäuser†

March 18, 2008

Preliminary Version

Abstract

This paper proposes a new method to pool the information contained in national indicator series to a single euro-area-wide indicator. Instead of using economic weights or dynamic factor models we use weights that minimize the variance of the out-of-sample forecast error for euro area real GDP growth. In an out-of-sample forecast experiment we find that our method generally outperforms alternative weighting and forecasting methods in terms of forecast accuracy.

JEL classifications: C13, C51, C53, C82, E37

Key words: Forecasting, Aggregation, Model Averaging, Real Time Experiment

*Ifo Institute for Economic Research, Poschingerstr. 5, 81679 München, Germany.
Email: <Hulsewig@ifo.de> and <Mayr@ifo.de>

†CESifo and Ifo Institute for Economic Research, Poschingerstr. 5, 81679 München, Germany.
Email: <Wollmershaeuser@ifo.de>

1 Introduction

An prompt assessment of the current state of the economy is a crucial ingredient for policy making. Since statistical agencies publish the first official release of real GDP with a delay of several weeks after the end of each quarter, information about the current economic situation can be extracted from timely available business-cycle indicators — such as industrial production, retail sales, composite indicators and confidence surveys — by using so-called bridge models.

In the euro area business-cycle indicators are typically collected at a national level by national statistical agencies or national survey institutes. In such a data-rich environment professional forecasters who aim at predicting real GDP growth rates for the euro area as a whole, can choose between two forecast strategies: pooling of forecasts and pooling of information. Pooling of forecasts uses national indicator series as predictors in the bridge model equations. One strategy is to produce a large number of forecasts of euro area real GDP growth rates and to combine them to a single forecasts of the area-wide target variable. The optimal weight of each model results from a minimization of the ex-post forecast error and hence takes the correlations of past forecast errors of each model into account (Granger and Ramanathan, 1984). Alternatively, real GDP growth rates of each euro area member country can be forecasted separately and then be aggregated to a single euro area real GDP growth rate by using the relative economic weight of each member country (Marcellino, Stock, and Watson, 2003). Both strategies, however, entail a trade-off between the informational gain of using disaggregate national data and the risk of making specification errors due to the need to estimate a large number of regression models.

Pooling of information generates a projection of euro area real GDP growth rates by using area-wide indicators as predictors that combine all national information. It thereby reduces the number of regressions to one. The simplest strategy is to employ the area-wide indicators which are provided by Eurostat (the European statistical agency) or other institutions (the European Commission or the OECD) and which are economically weighted averages of national indicators. Alternatively, professional forecasters may individually combine the set of national information by extracting common dynamic factors or principal

components (Forni, Hallin, Lippi, and Reichlin, 2000, and Stock and Watson 2002).

This paper proposes a new method to pool the information contained in national indicator series to a single area-wide indicator. Instead of using economic weights or dynamic factor models we use weights that minimize the variance of the out-of-sample forecast error for euro area real GDP growth. By allowing a pre-aggregation of individual information to national indicator series, the optimal pooling of information problem is reduced to a manageable number of variables, which avoids the construction of a “super model” (Timmermann, 2005) whose computation is often deemed to be prohibitively costly or even impossible.

We focus on three euro area business-cycle indicators, which are all available at both, the national and the area-wide level: the CESifo World Economic Survey (WES) indicator for the euro area, the Economic Sentiment Indicator (ESI) of the European Commission and the Composite Leading Indicator (ECI) of the OECD. The forecast models are specified as Autoregressive Distributed Lag (ADL) models, which are estimated employing a model averaging strategy in order to avoid misspecification due to lag length selection. In a first step, we compute the optimized weights and compare the forecast performance of the optimally pooled area-wide indicators with the officially published area-wide indicators under the assumption of full information. Our main result is that optimally pooled area-wide indicators significantly reduce the out-of-sample forecast root mean squared errors (RMSE) for euro area quarterly real GDP growth rates by 20% on average.

A major challenge of the optimal pooling of information approach is that it computes weights on the basis of out-of-sample forecast errors and hence relies on ex-ante information, which is not available at the time the projections are computed. As a consequence, the application of optimally pooled information in practice requires at least some stability of the weights. Thus, in a second step we evaluate the applicability of the optimal pooling of information approach in real-time by employing a pseudo out-of-sample forecast experiment, in which optimally pooled area-wide indicators are computed using only information that would have been available in real-time. The optimized weights are derived from a short training sample, which is then excluded from the forecast evaluation process, and are fixed for each of the recursive forecast steps. The performance of the op-

timal pooling of information approach is then compared with several traditional methods, which include pooling of forecast strategies (optimal combination of area-wide GDP forecasts and aggregation of national GDP forecasts) as well as pooling of information strategies (economic weights, OLS weights, principal components analysis, dynamic factor models). We find that the optimal pooling of information approach outperforms the traditional forecast methods in 22 out of 24 cases in terms of forecast accuracy as measured by the out-of-sample forecast RMSE. Furthermore, we find that the optimal weights derived from a short training sample are almost identical to those derived from the entire forecast sample. These results indicate a certain stability of the optimal weights and support the application of our approach in real-time.

The remainder of the paper is structured as follows. Section 2 reviews the traditional forecast strategies. In Section 3 we introduce the optimal pooling of information approach. The empirical part of the paper (Section 4) describes the forecast models and the data, presents the forecast exercise and summarizes our empirical results. Section 5 concludes.

2 Review of Traditional Forecast Strategies

In order to present the different forecast strategies that can be found in the literature with respect to the aggregation problem, we make the following notations. Suppose that we forecast the aggregate target variable Y_t (i.e. area-wide GDP growth) using a broad set of disaggregate information variables, denoted by $X_{i,t}$, with t being time and i referring to the disaggregate unit (i.e. member countries in the currency area). The number of disaggregate units (i.e. member countries) is given by K . The data sample that is available for the evaluation of the forecast model ranges from $t = 1, \dots, \Theta_2$. The forecast model is estimated over $t = 1, \dots, T$, with $T = \Theta_1, \dots, \Theta_2 - 1$. The out-of-sample forecasts of the area-wide target variable, which are denoted by $\hat{Y}_{T+1|T+1}$, are computed for $T + 1$ using the national information indicators already available at $T + 1$.¹ Thus, the number of out-of-sample forecasts is given by $\Theta_2 - \Theta_1$. Note that at this stage of the paper

¹Since in our set-up the current quarter is estimated, the literature often uses the notion nowcast instead of forecast Domenico, Reichlin, and Small (2006).

we assume a static structure of the forecasting model to keep the review as simple as possible; below in the empirical part, we will follow the literature and allow for more dynamics.

2.1 Pooling of Forecasts

Pooling of forecasts, or forecast combination, describes the combination of two or more individual forecasts to generate one single, pooled forecast. The idea of improving the accuracy of predictions regarding a certain target variable by combining the forecasts of different models was developed by Bates and Granger (1969) and Granger and Ramanathan (1984) and mainly follows the ideas of portfolio optimization and diversification gains. A large number of theoretical and empirical studies, e.g. Timmermann (2005) and Stock and Watson (2004), have shown the superiority of combined model based predictions. In the context of forecasting euro area real GDP, this multiple equation set-up allows for two alternative pooling of forecast strategies. In both cases, the single forecasts to combine are derived from national indicator series. While in the first approach a large number of euro area real GDP forecasts are pooled to a single area-wide forecast, in the second approach national real GDP series are forecasted and then aggregated to euro area real GDP. The crucial issue in both strategies is the determination of an adequate weighting scheme.

2.1.1 Optimal Combination of Area-Wide GDP Forecasts

For each of the K national indicators the following model is estimated over the period $t = 1, \dots, T$

$$Y_t = \delta + c_i X_{i,t} + \varepsilon_t^i. \quad (1)$$

The K forecasts resulting from each model are then linearly combined to a single forecast for the area-wide target variable according to

$$\widehat{Y}_{T+1|T+1} = \sum_{i=1}^K \omega_i \widehat{Y}_{T+1|T+1}^i, \quad (2)$$

where the superscript i attached to $\widehat{Y}_{T+1|T+1}$ denotes the forecast of the area-wide target variable obtained from the model using the national indicator $X_{i,t}$ (not to

confuse with $\widehat{Y}_{i,T+1|T+1}$ which denotes a forecast of the national target variable, see next Section).

The weighting of the national information ω_i , and hence the weights attributed to each model, depends on the model's out-of-sample performance. Under the assumption that the forecasts are unconditionally unbiased, the $\Theta_2 - \Theta_1$ out-of-sample forecast errors of model i , $\widehat{\varepsilon}_{T+1|T+1}^i = Y_{T+1} - \widehat{Y}_{T+1|T+1}^i$, are normally distributed around zero with variance σ_i^2 and covariance $\rho_{ij}\sigma_i\sigma_j$ for $j = 1, \dots, K$. Defining $\boldsymbol{\omega}$ as the $K \times 1$ vector containing the weights of each model and $\boldsymbol{\Sigma}_{\widehat{\varepsilon}}$ the $K \times K$ variance-covariance matrix of the out-of-sample forecast errors, the optimal weights are obtained from minimizing the variance of the combined out-of-sample forecast error:

$$\boldsymbol{\omega}^{opt} = \arg \min_{\boldsymbol{\omega}} \left[\boldsymbol{\omega}' \boldsymbol{\Sigma}_{\widehat{\varepsilon}} \boldsymbol{\omega} \right] \text{ s.t. } \sum_{i=1}^K \omega_i = 1. \quad (3)$$

As time moves on, the vector of forecast errors used for calculation of the optimal weights changes, and thus the weights themselves vary over time. Furthermore, the weights also depend on the number of out-of-sample forecasts N obtained from each model. The major challenge of the approach is the estimation of the covariance matrix of the forecast error series of the single models. If the number of models K increases, the computation of optimal weights becomes very complex. As a back door, Newbold and Granger (1974) propose to simply neglect the covariances between forecast errors and calculate the weights solely on basis of the variances of the forecast errors (i.e. the forecast RMSE). A further simplification of the weighting scheme is the use of equal weights. Timmermann (2005) derives conditions, under which the simple average of a number of forecasts outperforms single model based forecasts as well as more elaborated weighting schemes. Among others, Stock and Watson (2004) confirm the superiority of the equal weighting scheme in a broad empirical application, i.e. the so-called "forecast combination puzzle" holds in their analysis.

The combination of forecasts approach benefits from the possibility of including a large number of candidate regressors in forecasting a certain target series without of running into the problem of overparametrization or overfitting. However, as the data generating process is typically unknown, the need to specify

a large number of parsimonious regression models leads to specification errors (Lütkepohl, 1987).

- discuss problem of ex-post forecast error, real-time experiment, information availability

2.1.2 Aggregation of National GDP Forecasts

In line with Marcellino, Stock, and Watson (2003), the following model is estimated for each disaggregate unit (i.e. member country of the monetary union) $i = 1, \dots, K$ over the period $t = 1, \dots, T$

$$Y_{i,t} = \delta_i + c_i X_{i,t} + \varepsilon_{i,t}. \quad (4)$$

Building on these equations, forecasts of euro area real GDP are generated by computing weighted averages of the national predictions,

$$\widehat{Y}_{T+1|T+1} = \sum_{i=1}^K \omega_i \widehat{Y}_{i,T+1|T+1} = \sum_{i=1}^K \omega_i \widehat{\delta}_i + \sum_{i=1}^K \omega_i \widehat{c}_i X_{i,T+1|T+1}, \quad (5)$$

where ω_i are economic weights (e.g. GDP shares), reflecting the relative importance of country i in the monetary union. As equation 5 shows, the weighting of national information $X_{i,t}$ is thus influenced by both, the in-sample fit of the disaggregate model for country i and the economic weight of country i . Again, due to the specification of a number of parsimonious models, the approach hardly suffers from overfitting, but incorporates the drawback of larger specification errors when the data generating process is unknown.

2.2 Pooling of Information

Given that pooling a number of different models or predictions improves forecast accuracy, it should theoretically be possible to construct one single model that makes optimal use of the different kinds of information used by the component predictions in the combination. Compared to the multi-equation approaches reported above, constructing one so-called “super-model” and thus employing a single equation approach yields the advantage of smaller specification errors, especially when the data generating process is unknown. Diebold and Pauly (1990)

remark that while pooling of forecasts is suboptimal relative to pooling of information sets, it must be recognized that in many forecasting situations, particularly in real time, pooling of information sets is either impossible or prohibitively costly. The crucial issue within the pooling of information framework is again the determination of an adequate weighting scheme. When forecasting euro area real GDP, the question is which weighting scheme can be used in order to aggregate national indicators to an area-wide indicator.

2.2.1 Economic Weights

The most straightforward forecasting strategy is to use the area-wide indicators that are officially published as regressors of the forecasting model

$$Y_t = \delta + cX_t + \varepsilon_t. \quad (6)$$

The area wide indicator X_t is produced as a weighted average of national information variables

$$X_t = \sum_{i=1}^K \omega_i X_{i,t}, \quad (7)$$

where the ω_i 's typically reflect country i 's relative economic weight in the currency area.

Employing economic weights to construct one aggregate indicator series implies that these weights are exogenously given. Thus, any correlation between the national indicator series is ignored. Furthermore, the approach does not take any correlations between the resulting indicator series and the area-wide target into account.

2.2.2 OLS Weights

Employing OLS weights to generate an area-wide indicator series takes the correlation structure between the national indicator series explicitly into account. If the forecast model

$$Y_t = \delta + \sum_{i=1}^K c_i X_{i,t} + \varepsilon_t \quad (8)$$

is estimated over the period $t = 1, \dots, T$, the weighting of national information is given by the point estimates for c_i , which are derived from the minimization

of the in-sample regression errors. Thus, the in-sample fit of this approach with respect to the aggregate target variable must be superior to the multiple equation approach of Section 2.1.2. The problem of this approach is, however, that with an increasing number of disaggregate information variables K , the regression model more likely suffers from overfitting. As overparametrization leads to higher estimation uncertainty in finite samples, the out-of-sample performance of the OLS weighting approach worsens.

2.2.3 Factor Models

The use of factor models tries to mitigate the problem of parameter proliferation. While the forecasting model has the same structure as in Section 2.2.1, it starts with a principal component analysis that pools disaggregate information to a common factor X_t , before the target variable Y_t is forecasted. The common factor corresponds to the first principal component, which accounts for as much of the variability in the disaggregate indicators as possible. The weights ω_i are the squared elements of the eigenvector, which is associated with the first principal component. If the resulting common factor explains a large part of the variance of $X_{i,t}$, then $X_{i,t}$ is attributed a high weight.

The advantage of factor models is that information of a possibly large set of indicators is pooled by taking into account the covariances between the candidate regressors. In the context of business cycle analysis a useful extension of the static version of the factor model is the generalized dynamic factor model of Forni, Hallin, Lippi, and Reichlin (2000), which takes into account phase differences between disaggregate indicator time series by appropriately weighting leading and lagging variables.

The main drawback of the factor model is that the construction of the common factor ignores any correlation between the common factor X_t and the aggregate target variable Y_t . Thus, the weighting of national information only reflects in-sample correlation patterns between disaggregate indicators and is independent of the forecasting model.

3 Optimal Pooling of Information

Within the optimal pooling of information approach, the weighting of the national indicators results from minimizing the variance of the resulting out-of-sample errors of predictions of euro area real GDP from an area-wide model which builds on the newly constructed optimal indicator series. The procedure involves a non-linear numerical optimization routine which accounts for correlations between both, the disaggregate indicator time series and the aggregate target series. The determination of the optimal weights involves the following steps. We begin with an initial estimate for the weights $\boldsymbol{\omega} = (\omega_1, \dots, \omega_K)'$. We then compute the area-wide indicator X_t according to equation (7) and estimate equation (6) over the period $t = 1, \dots, T$. Finally, we compute the out-of-sample forecasts:

$$\widehat{Y}_{T+1|T+1}(\boldsymbol{\omega}) = \widehat{\delta} + \widehat{c} \sum_{i=1}^K \omega_i X_{i,T+1|T+1} = \widehat{\delta} + \widehat{c} \mathbf{X}'_{T+1|T+1} \boldsymbol{\omega}, \quad (9)$$

with $\mathbf{X}_{T+1|T+1}$ being a $K \times 1$ vector containing the national indicators at time $T + 1$, and the related out-of-sample forecast error:

$$\widehat{\varepsilon}_{T+1|T+1}(\boldsymbol{\omega}) = Y_{T+1} - \widehat{Y}_{T+1|T+1}(\boldsymbol{\omega}). \quad (10)$$

The optimal weights then result from the minimization of the variance of the out-of-sample forecast error:

$$\boldsymbol{\omega}^{opt} = \arg \min_{\boldsymbol{\omega}} \left[(\widehat{\varepsilon}_{T+1|T+1}(\boldsymbol{\omega}))^2 \right] \text{ s.t. } \sum_{i=1}^K \omega_i = 1. \quad (11)$$

By optimizing the weights on basis of the ex-ante forecast performance of the resulting area-wide model, we ensure the construction of a “super” indicator in terms of forecasting. The main drawback of this procedure is the reliance on ex-ante information, i.e. information stemming from the forecast sample is explicitly exploited. As a consequence, the application of the approach in practice requires at least some stability of the weights.

4 Empirical Results

4.1 Forecast Model Selection

Following Banerjee, Marcellino, and Masten (2003) and Marcellino, Stock, and Watson (2003), within each of the forecast frameworks we obtain short-term projections of real GDP for the euro area and its member states employing autoregressive distributed lag (ADL) models, so-called bridge models of the form

$$A(L)Y_t = \delta + B(L)X_t + \varepsilon_t, \quad (12)$$

where Y_t denotes real GDP expressed in quarterly growth rates, δ is a constant term, X_t describes the quarterly values of a business-cycle indicator, $A(L)$ and $B(L)$ are lag polynomials and ε_t denotes the error terms. Quarterly projections of real GDP growth are derived by exploiting the timely information contained in the contemporaneous business-cycle indicator in addition to the information provided by past realizations.

An important issue in specifying a time series model for forecasting purposes is the choice of the number of lags of the endogenous and exogenous included as regressors. Two approaches stand vis-à-vis: In-sample selection criteria and out-of-sample selection criteria. The in-sample selection of the optimal lag length is based on information criteria that build on the likelihood value adjusted by a penalty function. However, it is well known that these goodness-of-fit measures still lead to overfitting which results in a poor forecast performance (see e.g. Clark and McCracken (2005)). Alternatively, out-of-sample criteria might be employed that are directly based on the forecast errors of the resulting model. Granger (1993) recommends to choose those model specifications that provide the lowest value of the pseudo out-of-sample forecast RMSE as a criterion of model selection, since in-sample selection measures – such as the standard information criteria – frequently fail to provide strong implications for the out-of-sample performance. These so called performance-based approaches tend to ensure a parsimonious model specification which usually yields a comparatively proper forecast accuracy. An argument in favor of performance-based selection criteria often found in the literature is their higher robustness in case of structural changes.

It should be noted that the criterion of model selection on the basis of the pseudo out-of-sample forecast RMSE implies that information coming from the forecast sample is explicitly exploited, which means that the selected specifications might not have been found in real time. However, when the aim is at comparing the forecast potential of the candidate indicators the out-of-sample model selection criterion allows to tap the full forecast information of each indicator.

Alternatively, in order to make the out-of-sample criteria feasible in practice, the model evaluation must be based on data not used in the specification search. Ashley (1980) recommend, that the data must be strictly divided into in-sample and out-of-sample portions, with the latter reserved for forecast comparison. First using all of the data in the model search and then using an in-sample portion to re-estimate the chosen model and an out-of-sample portion to generate forecasts is of relatively little help in avoiding overfitting.

In order to avoid the drawbacks of the two lag selection approaches and to account for model selection uncertainty, we do not restrict the forecasting models to a certain lag length but employ the framework of pooling of forecasts described in chapter 2.1.1 in detail. We thus follow the idea that it is a priori impossible to discard a certain lag order from the forecast exercise. Although theoretical optimal weighting schemes can be derived by making use of the variances and covariances of the forecast errors of the single models, a growing number of empirical studies finds that simple combination schemes, i.e. attributing equal weights, which do not make use of the historical forecast performance outperform those approaches. This phenomenon is known as the “forecast combination puzzle”. Building on these ideas, we derive forecasts from an business-cycle indicator within an ADL-model by allowing for a certain maximum number of lags of the exogenous and the endogenous. The different model specifications are then built by permutating the candidate variables, given that the contemporaneous value of the indicator forms part of each model. One-step ahead forecasts of each model specification are derived and combined using equal weights.

4.2 Data Set

Our data for the euro area comprises real GDP, which is collected area-wide and for the member states, and business cycle indicators for the sample period from 1990Q1 to 2007Q2. Real GDP is taken from the Main Economic Indicators (MEI) Original Release Data and Revisions Database of the OECD that comprises vintage data, which is published each month from February 1999 onwards. In order to ensure consistency, we use the vintage series starting in 1995Q1 and extend the time series backwards by using the data of the Area Wide Model provided by the EACBN business cycle network of the European Central Bank in order to obtain the missing values for the period from 1990Q1 to 1994Q4. Real GDP is seasonally adjusted and converted into quarterly growth rates to satisfy stationarity conditions.

Since the availability of real GDP for the euro area member states is limited in the sample period under consideration, we focus on a subset of member states – as summarized in Table 1 – that cover almost 95% of area-wide economic activity.

Table 1: Selected euro area member states

Rank	Country	Share of GDP in %
1.	Germany	27.4
2.	France	21.1
3.	Italy	17.4
4.	Spain	11.3
5.	Netherlands	6.3
6.	Belgium	3.7
7.	Austria	3.0
8.	Finland	2.0
9.	Portugal	1.8

Notes: Euro area member states considered. Share of country GDP in overall economic activity evaluated in the year 2006.

Our selection of the candidate business-cycle indicators is bounded to those satisfying the following restrictions: (1) The business cycle indicator is published area-wide as well as on a national level. (2) It is constructed to describe the

current or future economic situation and is thus adequate to generate forecasts of real GDP. (3) The indicator is published quarterly or at a higher frequency. (4) It covers a sufficient time span starting at least in 1990. (5) The indicator is either not revised or vintage data is available covering the respective time span. Keeping these guidelines in mind, we end up with three business-cycle indicators, namely the CESifo World Economic Survey (WES) for the euro area, the Economic Sentiment Indicator (ESI) of the European Commission and the Composite Leading Indicator (ECI) published by the OECD, which all aim at tracking overall economic activity.

The WES summarizes the judgement of economic experts about the current economic situation by revealing their appraisals and expectations. It is exclusively based on qualitative information and is timely released within the survey quarter on a quarterly basis.² The WES is collected for each member state of the euro area, whereby the aggregate area-wide index is calculated as a weighted average of the individual country indices. The weighting scheme adopted refers to the share of a single country in total world trade (Stangl, 2007).³

The ESI combines the weighted information contained in confidence indicators of different sectors – namely industry, services, construction, retail trade and consumers – that are in turn constructed from survey data. Since the ESI is published on a monthly basis, the quarterly value is computed as an average of the monthly releases within the survey quarter. The indicator is built in two steps. In a first step, the area-wide confidence indicators of each sector are derived by aggregating the individual country sector confidence indicators. The weights are the shares of each of the member states in an area-wide reference series – here GDP growth – and are smoothed by calculating a two year moving average. In a second step, the area-wide confidence indicators are combined by using survey weights, which are based on two criteria: (i) the importance of the corresponding sector in the overall economy, and (ii) the ability of tracking the movements of the reference series (European Commission, 2007).

The ECI is constructed for the euro area on the basis of the individual indi-

²The WES is usually released in the middle of the second month of each quarter.

³The calculation of the national trade volumes is based on the foreign trade statistic published by the United Nations. The weighting scheme is readjusted once a year.

cators for the member states using weights that reflect the relative importance of economic activity as measured by GDP originating in industry that is raised by means of the purchasing power parity of 1995. Industrial production is the main reference series covering all industries excluding construction, which is used because of its cyclical sensitivity and monthly availability. The indicator is intended to display a reasonably consistent leading relationship with the reference series for the macroeconomic cycle.⁴ The ECI for each member state is compiled on the basis of a number of component series that include business surveys on production, stocks, order books and new orders, monetary aggregates, interest rates, share prices, construction indicators and terms of trade. These component series are selected according to a number of criteria such as economic significance, cyclical behavior, data quality and availability, the absence of structural breaks and freedom of excessive revisions. The ECI is calculated by aggregating the component series that are smoothed and normalized in advance employing a dynamic factor framework. It is released on a monthly basis so that the quarterly value is derived from the monthly average. The indicator is subject to data revisions due to revisions of the component series. Vintage data is provided by the OECD from 1985 onwards.⁵

4.3 Simulated Out-of-Sample Forecast Experiment

4.3.1 The Experiment

We generate quarterly forecasts of euro area real GDP by estimating ADL models for each business-cycle indicator recursively over the forecast sample from 1999Q4 to 2006Q2. The projections are derived as current quarter nowcasts for every quarter following the starting sample from 1990Q1 to 1999Q3 that is stepwise augmented by including an additional quarter.

⁴Since the ECI is not really intended for quantitative forecasts, the indicator does not match exactly our selection guidelines. This notwithstanding, the ECI is frequently used for economic forecasts in practice.

⁵It should be noted that the WES and ESI are by construction stationary, while the ECI has to be detrended to become stationary.

4.3.2 Optimized Area–Wide Indicators

Our analysis starts with generating pseudo out–of–sample nowcasts of euro area real GDP from the newly constructed area–wide indicators based on optimized weights. The weights are derived by focusing on the entire forecast horizon from 1999Q4 to 2006Q2. We use equal weights to initialize the optimization algorithm.⁶ Table 2 compares the forecast performance of the newly generated optimized area–wide indicator series to that of the corresponding economically weighted indicator series. The results indicate that forecast accuracy in terms of the out–of–sample RMSE calculated over the entire forecasting period is on average improved by 20%. The HLN (1997) test of forecast accuracy clearly confirms the significance of the improvement.

Table 2: Forecast performance of optimized weighted indicators vs. economic weighted indicators

		RMSE ratio	HLN–Test
CESifo Economic Climate	WES	0.79	–2.06
Economic sentiment	ESI	0.77	–2.44
OECD Leading indicator	ECI	0.81	–1.45

Notes: The RMSE ratios summarize the RMSE of area–wide models employing optimized weights relative to the RMSE of area–wide models based on economic weights. For the HLN (1997) test the corresponding critical value is ± 1.32 for the 5% level with 26 degrees of freedom. A value of the HLN statistic below -1.32 implies an improvement, while a value above $+1.32$ implies a worsening of the forecast compared to the benchmark predictions.

For an insight into the composition of the optimized area–wide indicators, Table 3 depicts the weights that the optimization algorithm attributes to the single national indicator series. The target function that underlies the algorithm is the RMSE over the 27 data points spanning the out–of–sample forecast period. For each of the three indicators under consideration only a small number of national indicator series are selected. As regards the WES the algorithm attributes large weights to Germany, which also poses the dominant member state in terms of

⁶We begin with an initial estimate for the weights of $\omega_i = \frac{1}{9}$ for $i = 1, \dots, 9$ for the WES and the ECI and of $\omega_i = \frac{1}{8}$ for $i = 1, \dots, 8$ for the ESI.

economic weights, as well as to a subset of indicators of smaller countries. The same holds for the ESI except for the difference that the number of selected national indicators is more extensive. In contrast, the ECI attributes a very large weight to Belgium, whereas the German indicator is completely neglected. Surprisingly, the Portuguese indicators contribute within each of the optimized area-wide indicators with weights that lie far beyond the Portuguese economic share. Even more surprisingly, despite the eminent role of their economies within the euro area, the French and the Italian indicators appear to be almost irrelevant.

Table 3: Optimal Weighting Schemes

National Indicator Series	WES	ESI	ECI
Austria	0.00	–	0.00
Belgium	0.00	0.17	0.51
Finland	0.00	0.07	0.00
France	0.00	0.00	0.00
Germany	0.50	0.38	0.00
Italy	0.00	0.00	0.19
Netherlands	0.25	0.21	0.00
Portugal	0.25	0.12	0.31
Spain	0.00	0.04	0.00

Notes: The weights are derived by minimizing the resulting out-of-sample RMSE that follows from 27 one-step ahead forecasts.

In order to analyze why certain national indicators enter the optimized area-wide indicators, Table 4 reports the out-of-sample RMSE of area-wide models solely based on each of the national indicator series relative to an area-wide model based on the area-wide indicators series as published. For all indicators at least one national series outperforms the area-wide economic weighted indicator in terms of out-of-sample RMSE. Consequently, the area-wide indicators based on economic weights are suboptimal in terms of forecasting. Comparing these figures with Table 3 shows, that our optimization algorithm attributes the highest weights to those national indicators series that exhibit a superior forecast performance.

A closer look at the correlations of the forecast errors resulting from the area-wide models solely based on the national indicator series explains why a subset of

Table 4: Area-wide ADL-models based on national indicators

National Indicator Series	WES	ESI	ECI
	RMSE ratio		
Austria	1.14	–	1.00
Belgium	1.15	1.19	0.88
Finland	1.32	1.16	1.09
France	1.20	1.24	1.07
Germany	0.90	0.77	1.06
Italy	1.34	1.27	1.03
Netherlands	1.10	0.91	1.04
Portugal	1.23	1.25	0.98
Spain	1.22	1.50	1.34

Notes: The RMSE are calculated relative to the RMSE of area-wide models employing economic weighting schemes.

assumed non-relevant national series enters the optimal indicators, additionally to the best performing indicator series. As shown in Table 5 the optimization algorithm attributes a weight larger than zero to those national series whose forecast errors show a low correlation to the ones of the dominant series identified in Table 4. Although the Portuguese indicators perform relative poorly in the area-wide approach for the WES and the ESI and worse than the Austrian series for the ECI, it is attributed a higher weight as its forecast errors from the area-wide models show a low correlation to the ones of the respective dominant series.

4.4 Stability Analysis within the Optimal Pooling of Information approach

The critical point of our optimization framework is the use of the resulting out-of-sample RMSE as target function of the optimization algorithm. By explicitly using information stemming from the forecast period, the approach somehow seems unfeasible within a real-time forecasting experiment. However, this critic can be considered baseless if the resulting weights reported in table 3 remain stable over time. Stability thereby means that within the optimal pooling of information approach each national indicator series is attributed a similar weight,

Table 5: Correlations of forecast errors of the Area-wide models based on national indicators

CESifo Economic Climate (WES)									
	Aus	Bel	Fin	Fra	Ger	Ita	Ned	Por	Spa
Aus	1.00	0.92	0.92	0.91	0.71	0.92	0.90	0.89	0.88
Bel	0.92	1.00	0.87	0.93	0.74	0.88	0.81	0.85	0.88
Fin	0.92	0.87	1.00	0.94	0.63	0.97	0.96	0.91	0.93
Fra	0.91	0.93	0.94	1.00	0.76	0.96	0.90	0.88	0.94
Ger	0.71	0.74	0.63	0.76	1.00	0.66	0.62	0.55	0.66
Ita	0.92	0.88	0.97	0.96	0.66	1.00	0.94	0.91	0.95
Ned	0.90	0.81	0.96	0.90	0.62	0.94	1.00	0.86	0.87
Por	0.89	0.85	0.91	0.88	0.55	0.91	0.86	1.00	0.86
Spa	0.88	0.88	0.93	0.94	0.66	0.95	0.87	0.86	1.00

Economic sentiment indicator (ESI)									
	Aus	Bel	Fin	Fra	Ger	Ita	Ned	Por	Spa
Aus	1.00								
Bel		1.00	0.65	0.78	0.70	0.70	0.77	0.65	0.56
Fin		0.65	1.00	0.87	0.65	0.84	0.67	0.72	0.84
Fra		0.78	0.87	1.00	0.66	0.88	0.69	0.78	0.88
Ger		0.70	0.65	0.66	1.00	0.69	0.53	0.59	0.62
Ita		0.70	0.84	0.88	0.69	1.00	0.74	0.79	0.89
Ned		0.77	0.67	0.69	0.53	0.74	1.00	0.71	0.65
Por		0.65	0.72	0.78	0.59	0.79	0.71	1.00	0.81
Spa		0.56	0.84	0.88	0.62	0.89	0.65	0.81	1.00

OECD Composite leading indicator (ECI)									
	Aus	Bel	Fin	Fra	Ger	Ita	Ned	Por	Spa
Aus	1.00	0.80	0.87	0.90	0.94	0.85	0.85	0.72	0.84
Bel	0.80	1.00	0.73	0.89	0.85	0.69	0.90	0.55	0.58
Fin	0.87	0.73	1.00	0.86	0.80	0.94	0.83	0.90	0.90
Fra	0.90	0.89	0.86	1.00	0.90	0.82	0.89	0.73	0.79
Ger	0.94	0.85	0.80	0.90	1.00	0.76	0.86	0.61	0.79
Ita	0.85	0.69	0.94	0.82	0.76	1.00	0.77	0.86	0.91
Net	0.85	0.90	0.83	0.89	0.86	0.77	1.00	0.68	0.74
Por	0.72	0.55	0.90	0.73	0.61	0.86	0.68	1.00	0.86
Spa	0.84	0.58	0.90	0.79	0.79	0.91	0.74	0.86	1.00

independent of the point on the timeline where the indicator is constructed.

In order to analyze whether the optimal pooling of information approach is applicable in real time, we first evaluate whether the optimal weights shown in table 3 can be derived using only a smaller number of out-of-sample forecast errors as target function. This setting can thus also be considered as a real-time experiment, assuming that one uses the weights derived only in the following forecasting steps. In a first iteration step, we optimize the weights of the national series using only the first one-step ahead forecast of the aggregate model, i.e. the forecast of the 4th quarter 1999. We then expand the window for calculation of the RMSE and thus the corresponding weighting scheme by one additional forecast step. At each iteration step, a new weighting scheme is derived which bases on a growing number on out-of-sample forecast errors for calculation of the RMSEs. At each step, we compare the forecast performance of the new indicator relative to the performance of the published indicators assuming that one uses only ex-post information for calculation of the weights. The evaluation window is thus scaled down as the window for calculation of the RMSE based weights increases. As table 2 reports, the forecast accuracy in terms of the out-of-sample RMSE calculated over the whole forecasting period regarding euro area real GDP is on average improved by 20 % using the optimal pooling scheme. Figure 1 shows the development of the relative out-of-sample RMSE of the above described experiment. As the graph shows, the optimal pooling of information approach gains accuracy with each ex-post forecast error added within the calculation of the weighting scheme. The relative RMSE reaches the minimum value of 0.8, which corresponds to an 20% gain in forecast accuracy towards the published indicators at a window size of the ex-post errors used of about 13 forecast steps or quarters. As with each new forecast period added for the calculation of the weighting scheme, the number of remaining datapoints for the evaluation of the forecast performance of the resulting indicator decreases, we stop this iterative procedure here at a minimum number of 14 datapoints for accuracy evaluation in order to ensure significance of the results. Additionally, the graphs show the relative RMSE based on the whole 27 datapoints spanning forecast window, ending in the minimum value given in table 2.

Summarizing this experiment, one can conclude that our information pooling

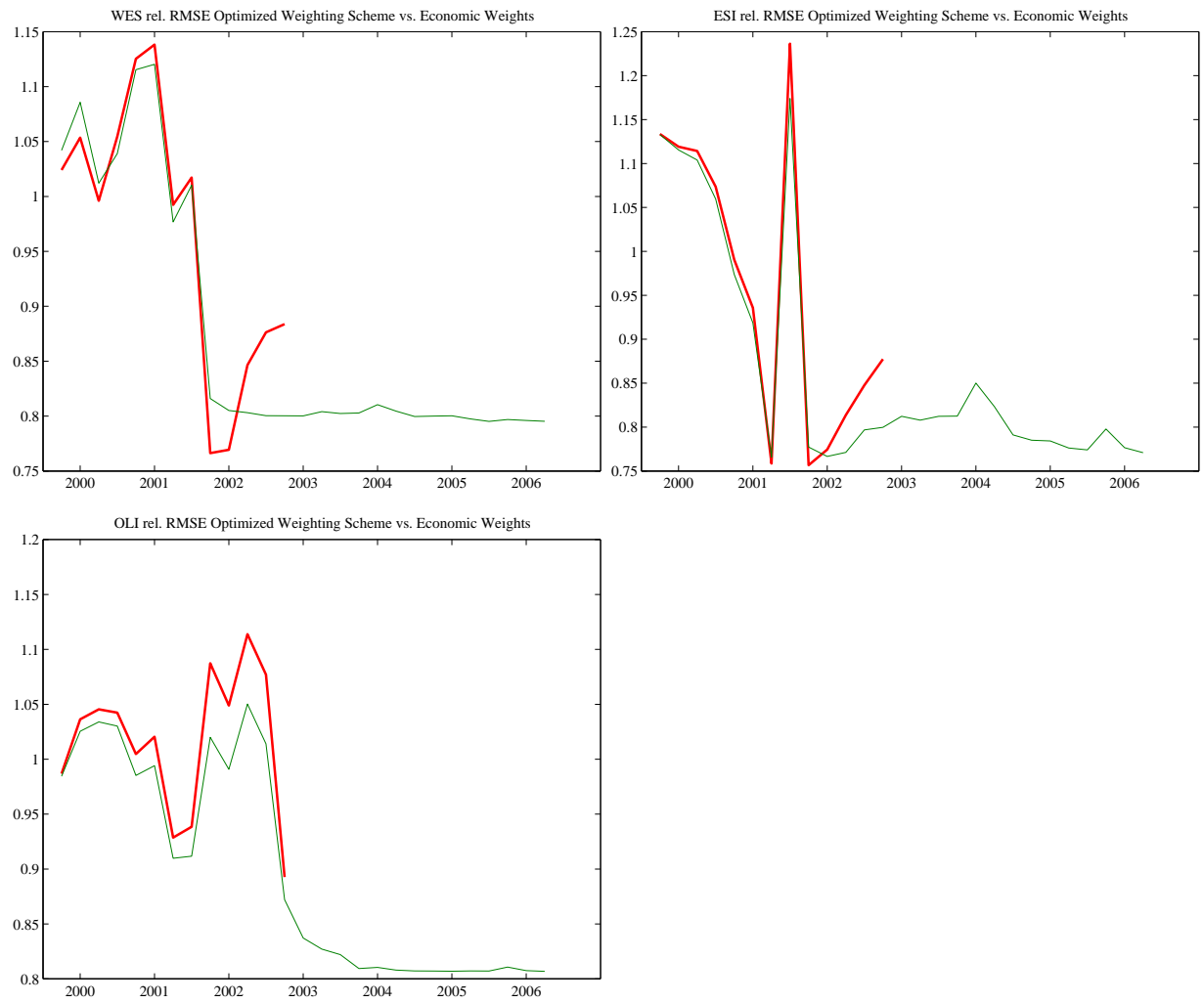


Figure 1: Recursive evaluation of optimized weights

procedure works well in an real time experiment, using only ex–post information for calculation of the weights attributed to each national indicator series. Considering the stability condition, it furthermore seems interesting whether the weights derived from a 13 forecast error window are independent of the point on the timeline. We therefore compare the weights of the different indicators derived from a 13 datapoints covering window moving along the total 27 datapoints spanning forecast window. Due to the shortness of the time series, we end up with 14 of these windows which can be used to derive the weights. Figure 2 shows the resulting weights for the different indicators. Although some variation can be seen especially for the ESI, it can be concluded that for each of the windows the setting of indicators entering the weighting scheme remains mainly unchanged. In order to evaluate the forecast accuracy of these new indicators, one can use a framework similar to the cross–validation approach. Neglecting the chronological structure, we use the newly generated weighting scheme to forecast those periods that have not been used within the construction of the weights.

Although the economic weights pose the most popular aggregation scheme, a number of alternative benchmark models are at disposal. In order to take our optimal pooling approach to a tougher test we compare its forecast accuracy in the following to a number of competing econometric weighting schemes and prediction approaches within the real–time experiment. We thereby use the weights, minimizing the sum of the first 13 out–of–sample squared forecast errors for calculation of our optimized indicator. We then forecast the remaining out–of–sample periods within a recursive approach, but keeping the weights constant. Table 6 reports the forecast RMSE of the alternative approaches described in chapter 2 relative to the one of our optimized pooling of information approach.

As the results indicate, our optimal pooling of information approach outperforms the economic weighting schemes for each of the indicators as well as almost all competing econometric approaches at disposal. Only for the ECI, the dynamic factor model and the optimal pooling of forecasts perform slightly better. However, an important feature of our optimal pooling of information approach is that we do not update the weights of the national indicators as time moves on but employ a fixed weighting scheme. This procedure enables us to present a new area–wide indicator which is not revised and thus can simply be updated as

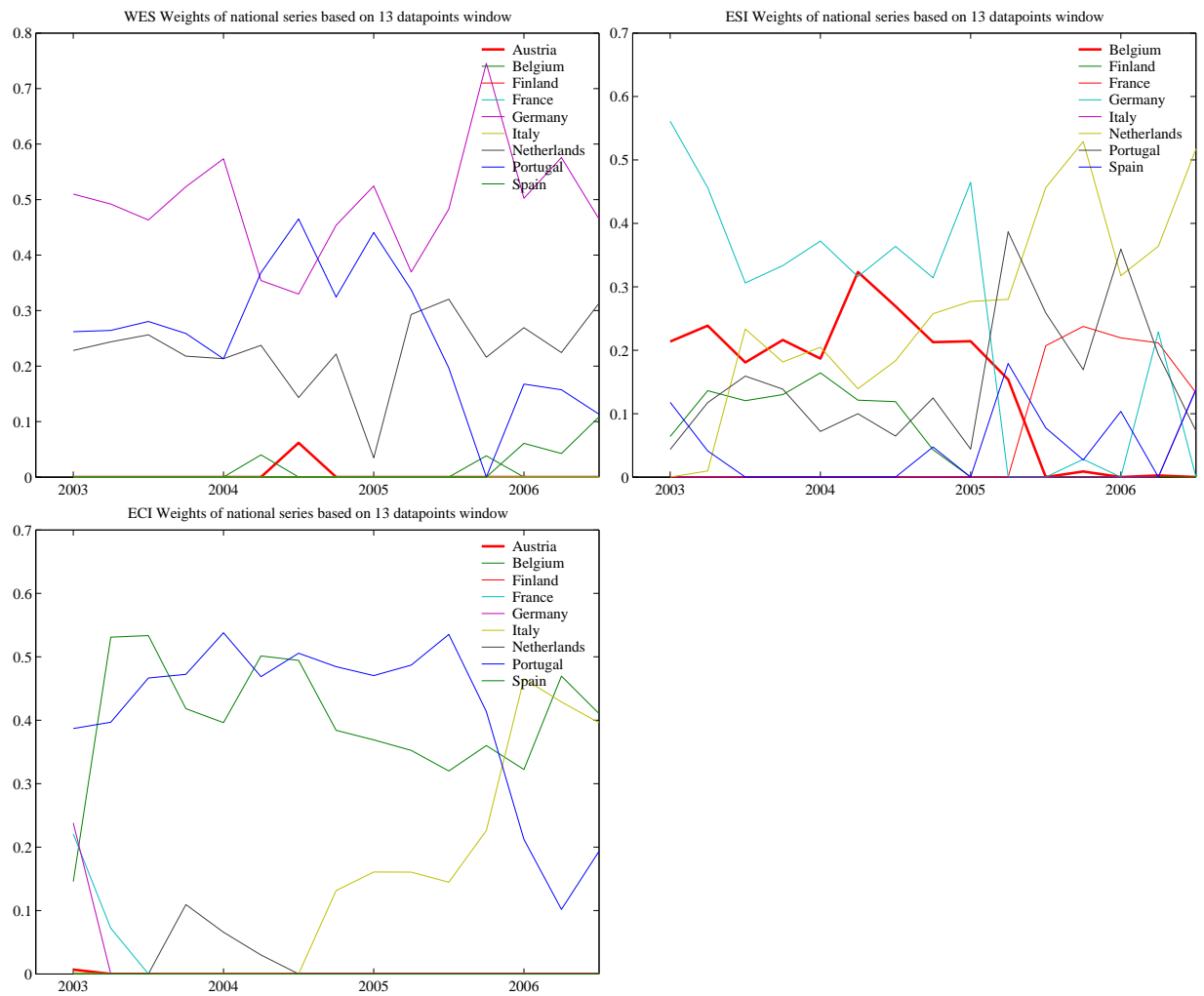


Figure 2: Rolling evaluation of optimized weights - Rolling window with 13 forecast errors

Table 6: Forecast RMSE of Optimal Pooling of Information approach (fixed weights) relative to Forecast RMSE of alternative benchmark approaches - Real time experiment

	rel. RMSE opt	HLN-Test
<i>WES</i>		
Univariate Approach	0.67	1.79
Pooling of Information -		
Economic Weights	0.88	0.96
OLS Weights	0.99	0.06
Principal Component Analysis	0.79	1.78
Dynamic Factor Model	0.82	1.39
Pooling of Forecasts -		
Optimal Combination of Area-wide Forecasts	0.99	0.08
Equal Weighting of Area-wide Forecasts	0.76	1.63
Aggregation of National Forecasts	0.72	2.02
<i>ESI</i>		
Univariate Approach	0.73	1.51
Pooling of Information -		
Economic Weights	0.88	1.60
OLS Weights	0.90	0.66
Principal Component Analysis	0.90	1.29
Dynamic Factor Model	0.99	0.11
Pooling of Forecasts -		
Optimal Combination of Area-wide Forecasts	0.92	0.86
Equal Weighting of Area-wide Forecasts	0.99	0.13
Aggregation of National Forecasts	0.90	1.03
<i>ECI</i>		
Univariate Approach	0.95	1.19
Pooling of Information -		
Economic Weights	0.89	1.15
OLS Weights	0.84	0.77
Principal Component Analysis	0.97	0.56
Dynamic Factor Model	1.03	-0.41
Pooling of Forecasts -		
Optimal Combination of Area-wide Forecasts	1.06	-0.90
Equal Weighting of Area-wide Forecasts	0.95	0.59
Aggregation of National Forecasts	0.88	1.30

Notes: The first column reports the Forecast RMSE of the Optimal Pooling of Information Approach relative to the Forecast RMSE of the respective benchmark approach. The second column reports modified Diebold–Mariano test statistics which follow a t -distribution with $n - 1$ degrees of freedom. Here, $n = 14$ is the number of out-of-sample forecasts within the real-time experiment. The critical value to reject H_0 (equal forecast performance) is ± 1.35 . A value of the test statistic above 1.35 implies an improvement of the Optimal Pooling of Information Approach towards the corresponding benchmark approach, while a value below -1.35 implies a worsening of the forecast compared to the benchmark predictions.

soon as new information on the national sub-indices is published. In contrast, all competing approaches that employ econometric weighting schemes benefit from updated weights.

A further improvement of forecast performance can be achieved by offsetting the restrictions on the weights of the national indicators. The weights are then allowed to take infinite and negative values. This enables a further improvement of forecast accuracy between 4 and 12%. However, the resulting weights lose their stability and are thus not applicable within a real-time experiment.

Table 7: Optimized weighted indicator without constraints relative to optimized weighted indicator

		relative RMSE	HLN-Test
CESifo Economic Climate	WES	0.88	-0.73
Economic sentiment	ESI	0.94	-0.73
OECD Composite Leading indicator	ECI	0.96	-1.28

Notes: The RMSE are calculated relative to the RMSE of area-wide models employing economic weighting schemes. For the HLN (1997) test the corresponding critical value is ± 1.32 for the 5% level with 26 degrees of freedom. A value of the HLN statistic below -1.32 implies an improvement, while a value above +1.32 implies a worsening of the forecast compared to the benchmark predictions.

5 Conclusion

This paper utilizes short-term forecasts of euro area real GDP – so-called nowcasts – to propose a new method to construct area-wide indicators by pooling the information contained in country-specific indicator series. We construct area-wide indicators by adopting optimized weights instead of economic weights. We focus on three business-cycle indicators, namely the CESifo World Economic Survey (WES) indicator for the euro area, the economic sentiment indicator (ESI) of the European Commission and the composite leading indicator (ECI) of the OECD, which are all available at the area-wide and country-specific level. The official area-wide indicators are constructed by means of economic weights. We construct the new area-wide indicators from country-specific indicator series by

using optimized weights that result from minimizing the variance of the resulting out-of-sample forecast errors of predictions of the area-wide target variable based on the new area-wide indicators series. Comparing the forecast performance of the area-wide indicators based on optimal weights to those based on economic weights yields that projections generated from indicators based on optimized weights perform on average 20% better in terms of the out-of-sample forecast root mean squared errors (RMSE) than those derived from indicators with economic weights.

In order to evaluate the adaptability of our forecasting method in real-time, we employ a pseudo out-of-sample forecast experiment. The area-wide indicators are thereby constructed using only information that would have been available in real-time. We find that our method of forecasting improves upon the bulk of alternative methods in terms of forecast accuracy as measured by the out-of-sample forecast RMSE. Furthermore, we find that the optimal weights solely based on the starting period used are almost identical to those that are derived for the entire forecast sample. These results strongly promote a certain stability of the optimal weighting scheme and the practicability of our approach in real-time.

References

- BANERJEE, A., M. MARCELLINO, AND I. MASTEN (2003): “Leading Indicators for Euro-Area Inflation and GDP Growth,” Discussion Paper 3893, CEPR.
- BATES, J., AND C. W. J. GRANGER (1969): “The combination of forecasts,” *Operations Research Quarterly*, 20, 319–25.
- CLARK, T., AND M. MCCracken (2005): “The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence,” *Journal of Money, Credit, and Banking*.
- DIEBOLD, F. X., AND P. PAULY (1990): “The use of prior information in forecast combination,” *International Journal of Forecasting*, 6, 503–508.

- DOMENICO, G., L. REICHLIN, AND D. SMALL (2006): “Nowcasting GDP and inflation: the realtime informational content of macroeconomic data releases,” Discussion paper, Working Paper Series from European Central Bank 633.
- EUROPEAN COMMISSION (2007): “The Joint Harmonised EU Programme of Business and Consumer Surveys: User Guide,” Economic studies and research, European Commission.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): “The Generalized Dynamic-Factor Model: Identification and Estimation,” *The Review of Economics and Statistics*, 82(4), 540–554.
- GRANGER, C. W. J. (1993): “On the limitations of comparing mean square forecast errors: Comment,” *Journal of Forecasting*, 12(8), 651–652.
- GRANGER, C. W. J., AND R. RAMANATHAN (1984): “Improved methods for combining forecasts,” *Journal of Forecasting*, 3, 197–204.
- LÜTKEPOHL, H. (1987): *Forecasting Aggregated Vector ARMA Processes*. Springer Verlag.
- MARCELLINO, M., J. STOCK, AND M. WATSON (2003): “Macroeconomic forecasting in the euro area: Country specific versus area-wide information,” *European Economic Review*, 47(1), 1–18.
- NEWBOLD, P., AND C. GRANGER (1974): “Experience with forecasting univariate time series and the combination of forecasts,” *Journal of Royal Statistical Society A*, 137, 131–146.
- STANGL, A. (2007): “World Economic Survey,” in *Handbook Of Survey-Based Business Cycle Analysis*, ed. by G. Goldrian, pp. 57–67. Ifo Economic Policy Series, Edward Elgar.
- STOCK, J., AND M. WATSON (2002): “Macroeconomic Forecasting Using Diffusion Indexes,” *Journal of Business and Economic Statistics*, 20(2), 147–62.
- STOCK, J. H., AND M. WATSON (2004): “Combination Forecasts of Output Growth in a Seven-country Data Set,” *Journal of Forecasting*, 23, 405–430.

TIMMERMANN, A. (2005): "Forecast Combinations," in *Handbook of Economic Forecasting*, ed. by G. C. W. J. Elliott, G., and A. Timmermann. North Holland.