Combination of Forecast Methods Using Encompassing Tests: An Algorithm-Based Procedure ; For the revised version of this paper, see Working Paper 240, Economics Series, June 2009, which includes some changes. The most important change regards the reference of Kisinbay (2007), which was not reported in the previous version. The hierarchical procedure proposed in the paper is based on the approach of Kisinbay (2007), but some modifications of that approach are provided.

Mauro Costantini
Carmine Pappalardo

INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

Impressum

Author(s):
Mauro Costantini, Carmine Pappalardo

Title:
Combination of Forecast Methods Using Encompassing Tests: An Algorithm-Based Procedure ; For the revised version of this paper, see Working Paper 240, Economics Series, June 2009, which includes some changes. The most important change regards the reference of Kisinbay (2007), which was not reported in the previous version. The hierarchical procedure proposed in the paper is based on the approach of Kisinbay (2007), but some modifications of that approach are provided.

ISSN: Unspecified

2008 Institut für Höhere Studien - Institute for Advanced Studies (IHS)
Josefstädter Straße 39, A-1080 Wien
E-Mail: office@ihs.ac.at
Web: www.ihs.ac.at

228

Reihe Ökonomie

Economics Series

Combination of Forecast Methods Using Encompassing Tests:

An Algorithm-Based Procedure

Mauro Costantini, Carmine Pappalardo

"For the revised version of this paper, see Working Paper 240, Economics Series, June 2009, which includes some changes. The most important change regards the reference of Kisinbay (2007), which was not reported in the previous version. The hierarchical procedure proposed in the paper is based on the approach of Kisinbay (2007), but some modifications of that approach are provided."

228

Reihe Ökonomie

Economics Series

Combination of Forecast Methods Using Encompassing Tests:

An Algorithm-Based Procedure

Mauro Costantini, Carmine Pappalardo

October 2008

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Mauro Costantini
Department of Economics
University of Vienna BWZ
Bruenner Str. 72
1210 Vienna, Austria
☎: +43/1/4277-37478
fax: +43/1/4277-37498
email: mauro.costantini@univie.ac.at

Carmine Pappalardo
ISAE, Institute for Studies and Economic Analysis
Rome

Abstract

This paper proposes a strategy to increase the efficiency of forecast combining methods. Given the availability of a wide range of forecasting models for the same variable of interest, our goal is to apply combining methods to a restricted set of models. To this aim, an algorithm procedure based on a widely used encompassing test (Harvey, Leybourne, Newbold, 1998) is developed. First, forecasting models are ranked according to a measure of predictive accuracy (RMSFE) and, in a consecutive step, each prediction is chosen for combining only if it is not encompassed by the competing models. To assess the robustness of this procedure, an empirical application to Italian monthly industrial production using ISAE short-term forecasting models is provided.

Keywords

Combining forecasts, econometric models, evaluating forecasts, models selection, time series

JEL Classification

C32, C53

Comments

The first draft of this paper was mostly prepared while Mauro Costantini was working as a post-doc at ISAE (Institute for Studies and Economic Analyses). The authors would like to thank Michael P. Clements for his very useful suggestions. The authors would also like to thank the participants at the 28th annual International Symposium on Forecasting, Nice 2008. None of them is responsible for any remaining error. The opinions expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of ISAE or its staff.

Contents

1 Introduction

Forecast combination is often used to improve forecast accuracy. A linear combination of two or more predictions may often yield more accurate forecasts than using a single prediction to the extent that the component forecasts contain useful and independent information (Newbold and Harvey, 2002). To generate independent forecasts two ways could be followed. One is to examine different data, and the other is to use different forecasting methods. On one hand, the use of several sources of the data can add useful information and can also adjust for biases. On the other hand, forecasting combining methods can reduce errors arising from faulty assumption, bias, or incorrect data. In this paper, a new algorithm-based procedure to increase the efficiency of forecasting combining methods is provided. The algorithm relies on a widely used encompassing test (Harvey, Leybourne, Newbold, (HLN) 1998). Rather than for evaluating forecasts, the above statistics is implemented to select a subset of forecasts to be combined. According to this procedure, overall forecasting models are first ranked using RMSFE measure and, in a consecutive step, each prediction is chosen for combining only if it is not encompassed by the competing forecasting models. A multiple encompassing test (Harvey and Newbold, 2000) is used to assess the robustness of the models selecting procedure. An empirical application to Italian monthly industrial production is provided. We exploit several short-term forecasting models currently used at ISAE to obtain forecasts up to 6 steps ahead, both in a recursive and rolling regression framework. In a majority of cases, forecasts deriving from the algorithm procedure outperform those obtained by combining overall models in terms of RMSFE. The paper is organized as follows. Section 2 presents the seven models under evaluation. In section 3, the linkages between forecast encompassing test and RMSFE are presented. In section 4, the algorithm procedure and the methods used for combining are described. Section 5 presents empirical results. Section 6 concludes.

2 Forecasting models

In this section, seven time series models for forecasting the industrial production (*IPI*) are briefly described: four single-equation models, a dynamic factor model, a VAR model and an ARIMA model. The econometric specifications selected for the empirical exercise are based on a collection of coincident and leading indicators of the short-term pattern of manufacturing activity. This set of variables consists of both *hard* and *soft* data. The former includes time series related to energy and intermediate inputs as the quantity of raw materials transported by rails ($TONN$), the volume of natural gas demanded by the industrial sector (*snam*), the supply of electric energy as a whole (GW). *Soft* data relies on businesses' production expectations (PP) deriving from ISAE business surveys and on the purchasing managers' index *PMI*. Other than *PMI*, all the variables are not seasonally adjusted and, also with the exception of *PP*, observations at the end of sample are revised by further data releases. IPI and related indicators are considered at their latest available updates.

The general specification of single−equation multivariate models is:

$$\Delta_{12}y_t = \alpha + \gamma\Delta_{12}y_{t-h} + \sum_{j=h}^{p}\beta_j x_{t-j}^{m_h} + \delta d_t + \varepsilon_t^{m_h} \tag{1}$$

where m denotes the models for each forecasting step (h=1,...,6), $\Delta_{12} = (1 - L^{12})$, d_t are the deterministic components and ε_t is the idiosyncratic error term. Specifically:

1. the *SE* model includes *TONN* and the composite index *PMI* as regressors;

2. in the *GW* model, the lagged endogenous variable, *PMI* and the electric indicator behave as regressors;

3. in the third model GW_c, the electricity demand, *PMI* (both lagged by 1 period) and the variable $\tilde{\mathrm{C}}_{q,t}$ are included. The last one, which is defined as the deviation of $\mathrm{C}_{q,t}$ (the current temperature in period q and year t) from its average level observed in the same month over the latest five years ($\mathrm{C}_{q,t-1}$,..., $\mathrm{C}_{q,t-5}$), is considered since some of the electricity components could be significantly affected by temperature patterns (the electricity demand of households and the service sector);

4. the *Gas* model is based on *snam* and *PMI* indicators.

As a common feature, the deterministic component of the above models consists of a sequential specification (up to 1 lag) of the month-on-month trading days variation. The model GW_c is supplemented with a set of seasonal dummies, taking value equal to $\tilde{\mathrm{C}}_{q,t}$ in the reference month, zero otherwise. The reduced form of all single-equation models has been achieved applying the following criteria: *i*) each general unrestricted model (*GUM*) is estimated over the period 1997:1-2005:9 (except for *GAS* model) using up to the 12th lag of the independent and dependent variables; *ii*) the General-to-Specific approach for model reduction is performed running Pc−Gets (Hendry and Krolzig, 1999; Krolzig and Hendry, 2001).
To get forecasts for more than one step ahead (h=2,..., 6), each *GUM* is constructed by lagging the available information. The general unrestricted models up to h=2 still include contemporary values of the explanatory variables. The explanatory variables are lagged by 1 period if h=3, by 4 months if h=6. This allows multi-step forecasting in a single-equation context not involving any prediction of the selected indicators (Rünstler and Sédillot, 2003; Rathjens and Robins, 1993).
In addition, two other models, based on different functional forms, are considered. Following Stock and Watson (1998, 2002), a dynamic factor model (*Factor*) is estimated:

$$\Delta_{12}y_t^{m_h} = \beta_0 + \sum_{i=1}^{4} B_i\hat{F}_{i,t-h} + \gamma\Delta_{12}y_{t-h} + \hat{\varepsilon}_t^{m_h} \tag{2}$$

where m denotes the models for each step (h=1,...,6) and $i = 1, ..., 4$ are the number of estimated factors ($\hat{F}_{it}$). Lagged values of the dependent variables also appear as predictors since the error term can be serially correlated. The factors are extracted from a large data-set of ISAE business surveys (current assessments on demand, production and inventories, short-term prospects for orders, production and prices). The number of factors are computed using the IC(3) criterion proposed in Bai and Ng (2002) and their estimates are obtained using the Principal Component method.

In the second model, *VECM* (Bruno and Lupi, 2004), the indicators are re-parameterized in seasonal differences, since this proves useful to obtain quasi-orthogonal regressors. The starting unrestricted model takes the form:

$$\Delta\Delta_{12}y_t = \alpha\Delta_{12}y_{t-1} + \sum_{j=1}^{13}\beta_j\Delta\Delta_{12}y_{t-j} + \phi d_t + \varepsilon_t \tag{3}$$

where $y_t = (IPI_t, TONN_t, PP_t)$, $\Delta = (1-L)$, $\Delta_{12} = (1-L^{12})$, d_t are the deterministic components which include, other than the usual specification of trading days effect, two dichotomous variables that take value 1 in both August and December if production prospects in the corresponding previous months (July and November) are positive, -1 if negative. Finally, we get an ARIMA time series model as a benchmark model. It involves double differencing, both at regular and seasonal frequencies. According to the Schwarz information criterion for lag length selection, the final specification consists of an ARMA(2,3) polynomial for the regular part, $\text{MA}(1)_{12}$ for the seasonal frequencies.

3 Forecast encompassing and RMSFE

The aim of this section is to define the theoretical linkages between two most used criteria for forecast evaluation, the forecast encompassing test and the minimizing the RMSFE. Using RMSFE and encompassing tests as complementary rather than competing forecast criteria stems from a theoretical contribution of Ericsson (1992). To improve the detection of the predictive ability across non-nested models, the author shows that the forecast encompassing test is a *sufficient* condition for RMSFE dominance, i.e. for minimizing RMSFE of a given model. The starting point is to consider two alternative non-nested linear models for the same dependent variable y_t, estimated over the sample period $[1,\ T]$:

$$M_1 : y_t = \delta_1' z_{1t} + \nu_{1t} \tag{4}$$

$$M_2 : y_t = \delta_2' z_{2t} + \nu_{2t} \tag{5}$$

where z_{1t} and z_{2t} do not have regressors in common and are linked by the relation $z_{1t} = \Pi z_{2t} + \varepsilon_{1t}$. Substituting into (4) yields the following restrictions for equation (5)

$$\delta_2' = (\delta_1'\Pi) \tag{6}$$

$$\nu_{2,t} = \nu_{1,t} + \delta_1'\varepsilon_{1,t}. \tag{7}$$

Assuming that the forecasts from the models (4) and (5) are $\hat{y}_{1j} = \delta' z_{1j}$ and $\hat{y}_{2j} = \delta' z_{2j}$, ($j$ = T+1,..., T+n), restriction (6) (forecast-model encompassing) implies that z_{2j} has no power in explaining the forecast error given z_{1j}. This is equivalent to testing for $\gamma = 0$ in the equation $y_j = \delta_1' z_{1j} + \gamma z_{2j} + \nu_{1j}$. From restriction (7), we get

$$\mathrm{E}(y_j - \hat{y}_{2j})^2 = \mathrm{E}(y_j - \hat{y}_{1j})^2 + \delta_1' \Omega \delta_1 \tag{8}$$

where $\mathrm{E}(y_j - \hat{y}_{1j})^2$ is the RMSFE of model 1, $\mathrm{E}(y_j - \hat{y}_{2j})^2$ is the RMSFE of model 2 and $\Omega = \mathrm{E}(\varepsilon_{1j}\varepsilon_{1j}')$. Testing this hypothesis is equivalent to testing for $\alpha = 0$ (forecast encompassing) in the equation

$$y_j = \delta_t' z_{1j} + \alpha \hat{y}_{2j} + \nu_{1j}. \tag{9}$$

As shown in Ericsson (1992), the sufficient condition for this is that $\gamma = 0$. Further, this implies that Ω is a positive definite matrix, so that $RMSFE_1 < RMSFE_2$ (RMSFE dominance). From the above discussion, with not-nested forecasting models, the sufficient condition to minimize the RMSFE of a given model is to verify that it encompasses all the other competing models. This implies to perform the encompassing test only in one direction (model with the lowest RMSFE against the model with higher RMSFE).

4 Forecast encompassing and combining methods

The issue of complementarity between RMSFE and encompassing test is used to develop an algorithm for the efficient selection of non-nested forecasts to combine. The algorithm considers pseudo out-of-sample forecast as inputs. The basic idea is to compare all forecasting models with each other using HLN (1998) encompassing test, to eliminate the encompassed models, and to use several forecast combining methods for combining the remaining forecasts. The encompassing algorithm is described as follows:

Step 1. Calculate the RMSFE of the out-of-sample forecast for each model using out-of-sample forecasts and realized values. Rank the models according to their past performance based on RMSFE;

Step 2. Pick the best model (i.e, model with the lowest RMSFE), and test sequentially whether the best forecasting model encompasses other models, using the HLN test. If the best model encompasses the alternative model at some significance level α, delete the alternative model from the list;

Step 3. Repeat step 2, with the second best model. The list of model includes only those that are not encompassed by the best model, and the best model;

Step 4. Continue with the third best model, and so on, until no encompassed model remains in the list;

Last Step Obtain the algorithm combining forecast (ACF) by using several forecast combining methods with all models previously selected.

As regards to empirical application, several issues should be addressed.
First, an initial set of out-of-sample forecast of 24 observation is considered for applying the HLN (1998) test. Harvey et al. (1998) developed a test of the null hypothesis of forecast encompassing using the methodology of the test for equal accuracy discussed in Diebold and Mariano (1995) and Harvey et al. (1997).
Second, in our empirical application, we consider several significance levels α of the HLN test: 0.01,0.05,0.10, 0.15,0.20,0.25.
Third, a multiple encompassing F-test (Harvey and Newbold, 2000) is applied to verify the robustness of our models selecting procedure based on the HLN encompassing test. The F-test would confirm whether, at each step of the sequence procedure, the best model encompasses or not all the competitors.
Fourth, several methods of forecasts combinations are used. All of the combining methods take the form of a linear combination of the individual forecast:

$$\hat{y}^h_{c,t+h|t} = w_{0,t} + \sum_{i=1}^{n} w_{i,t}\hat{y}^h_{i,t+h|t}, \tag{10}$$

where $\hat{y}^h_{c,t+h|t}$ is a given combination forecast whose weights, $\{w_{i,t}\}^n_{i=0}$, are computed using the individual out-of-sample forecast, n is the number of the models and h is the forecast horizon. In particular, we consider:
a) three simple combining methods: the mean, median, and trimmed mean. In the mean case, we set $\mathrm{w_{0,t}} = 0$ and $\mathrm{w_{i,t}} = \frac{1}{\mathrm{n}}$ for i=1,...,n in the equation (10); for the median, we use the sample median of $\{\hat{y}^h_{i,t+h|t}\}^n_{i=1}$; the Trimmed mean uses $\mathrm{w_{0,t}} = 0$ and $\mathrm{w_{i,t}} = 0$ for the individual models that generate the smallest and largest forecasts at time t, while $\mathrm{w_{i,t}} = \frac{1}{(\mathrm{n}-2)}$;
b) the unrestricted OLS combining method (see Granger and Ramanathan, 1984). The combining weights are calculated using OLS regression;
c) the WLS combining method proposed by Diebold and Pauly (1987). We apply the "t-lambda" method. It consists of a combining method with weights calculated by WLS estimator. Diebold and Pauly (1987) suggested to use the weighting matrix $\mathbf{\Psi} = diag[\Psi_{tt}] = [\kappa t^{\gamma}]$, where $\kappa, \gamma > 0, t = 1, ..., T$ and T is the number of observations used in the WLS regression. In our empirical application, we use $\gamma = 1$ (weights that decrease at constant rate) and $\gamma = 3$ (weights that decrease at increasing rate).
d) the DMSFE (Discount Mean Square Forecast Errors) combining methods. Following Stock and Watson (2004), the weights in equation (10) depend inversely on the historical forecasting performance of the individual models:

$$w_{i,t} = \frac{\lambda^{-1}_{it}}{\sum_{j=1}^{n} \lambda^{-1}_{jt}}, \tag{11}$$

where

$$\lambda_{i,t} = \sum_{s=T+1}^{T+K} \delta^{T+K-s}(y^h_s - \hat{y}^h_{i,s|s-h})^2, \tag{12}$$

$w_{0,t} = 0$, and δ is a discount factor. When $\delta = 1$, there is no discounting; when $\delta < 1$, greater importance is attributed to the recent forecast performance of the individual models. We use $\delta = 0.9, 1.0$.
Finally, to assess the robustness of the algorithm-based procedure we compare the models by evaluating the relative RMSFE. For each combining method we

calculate the RMSFE for the algorithm forecast ($\mathrm{RMSFE_{ACF}}$) and the RMSFE from combining all models ($\mathrm{RMSFE_{ALL}}$). The relative RMSFE, $\frac{RMSFE_{ACF}}{RMSFE_{ALL}}$, provides a scale-free metric, where the ratio less than one denotes that the algorithm forecast outperforms the combining forecast from all models.

5 Empirical results

All models presented in section 2 are estimated over a common sample (except for GAS model, which estimation sample begins in 2002:1, the remaining models are estimated over the common time interval 1997:7-2005:9) and usual diagnostics to evaluate in-sample correct specification have been performed (results are not reported for space reasons and are available upon request). The forecasting exercise is carried out using both the recursive and rolling schemes. This latter framework is generally used when there are concerns about turning points and biases from the use of older information: it plays the role of a sensitivity analysis with respect to the results of the combination obtained through the encompassing algorithm. The dimension of the rolling window is different for each model to account for the different time span over which the indicators are available (starting in 1979 for TONN, in 1991 for IPI and PP, in 1997 for PMI, in 2001 for snam).

In Table 1 the RMSFE of the out-sample forecasts for each model is reported. On the basis of these results, the models are ranked from the best to the worst for each forecast horizon (see Table 2). Different results in terms of the ranked models are found for recursive and rolling estimation respectively. Looking at the first-step ahead horizon, SE model is firstly ranked in both estimation frameworks but its forecasting accuracy worsens for successive steps ahead. When the recursive scheme is considered, the GW_c model shows higher rankings at several prediction steps. Its performance is slightly better than the rolling estimates for h=1, 2. When 6 steps ahead are considered, the VAR model is characterized by significantly higher performances and outperforms all the other models, as it's generally expected. Its predictive ability slightly improves in the rolling schemes, due to the cutting down of the older observations, in presence of the long lag structure of model equations.

On the basis of the rank classification, the HLN test is applied to eliminate models that are encompassed by others. Findings are reported in Tables 3-4. Given the number of steps ahead and the estimation scheme, the number of models selected for combination depends on the significance level. The lower is the significance level α, the stronger is the selection between competing models through the encompassing filtering. As α rises, a larger number of forecasts is selected for combination. Looking at the first-step ahead horizon, at 25% significance level, four models are selected both in the recursive $(\mathrm{SE}, \mathrm{GW}, \mathrm{GW_c}, \mathrm{Gas})$ and rolling $(\mathrm{SE}, \mathrm{VAR}, \mathrm{Gas}, \mathrm{GW_c})$ scheme. Lowering α, the selected models reduces to three (the fourth model is ruled out in each scheme, respectively) and, for $\alpha = 0.01$, only the SE model is selected in both estimation frameworks. In the recursive scheme, the selected models are $\mathrm{GW_c}$ and SE for h=2 (irrespective of the significance levels), $\mathrm{GW_c}$ and Gas for h=3 (VAR is only chosen when $\alpha = 0.25$). As regards to rolling estimates, four models enter the combination for h=2 and $\alpha \in [0.05, 0.25]$ $(\mathrm{GW}, \mathrm{GW_c}, \mathrm{Gas}, \mathrm{Factor})$. Models based on the electricity indicator are taken for when h=3 and trim down to GW as $\alpha \leq 0.15$.

At six-step horizon, the VAR model outperforms all the other models.
The multiple encompassing F-test would seem to confirm HLN test findings (similar results are found using MS^* test. They are available upon request). Selection models results are reported in Table 5. The F-test is applied to each best model at each step of the sequence algorithm and is performed against all other competing models. For each significance level, the rejection of the null indicates that the rival models (at least, some of them) are not encompassed by the best model. Looking at the first-step horizon for recursive scheme, the multiple encompassing F-test supports the HLN tests results at each algorithm procedure step since the null hypothesis that the best model encompasses all the competitors is always rejected at the same significance level. In this sense, the multiple encompassing test reveals the robustness of our algorithm procedure in selecting models that are not encompassed. Similar results are found for rolling scheme. As regards the second step-horizon for the recursive scheme, the multiple encompassing F-test confirms results based on HLN test. The only difference is found in the first step of the algorithm procedure. The multiple encompassing F-test reject the null at 5% level instead of 1% as in the HLN test. In the rolling scheme, for h=1,2, the F-test results show higher probability than that one of HLN test. Thus there is greater tendency to accept the null hypothesis of encompassing **since the number of observations remains constant** in the use of rolling estimation window. For h=3,6, multiple encompassing F-test confirms the HLN test results.
To assess the robustness of the algorithm-based procedure we compare the models by evaluating the relative RMSFE (Table 6). Results are statistically significant in many cases at 5% level and, only in few cases, the ratio is larger than one. These findings confirm the goodness of the algorithm procedure. Irrespective of the combination methods and of the number of steps ahead, the best results in terms of relative RMSFE are obtained for higher significance levels, hence allowing for the averaging of a large number of models. At the low significance levels, very few forecasts are considered for combination and the overall forecast benefits less from the advantages of combining. A significant improvement in this framework is that the relative RMSFE$_h$ (h=1,...,6) remains roughly constant below unity almost for the majority of combination methods and significance levels: this result is a key feature of the filtering algorithm. Moreover, the several combination methods can be ranked in terms of relative RMSFE. In the recursive scheme, the basic linear pooling methodologies (Mean, Median and Tmean) have performed remarkably better than other combination methods. This result is fully consistent with the prevailing evidence in the empirical literature. The relative RMSFE in the case of simple averaging (Mean) ranges in 0.908-0.924 for h=1, in 0.937-0.942 when h=3. Apart from simple linear averaging methods, relative RMSFE is minimized through discounted combination algorithms (DMSFE), in which the weights are estimated so to be mostly affected by recent past model performance (Newbold and Harvey, 2002). The worst results are obtained applying OLS and WLS procedures, the only ones for which the relative RMSFE is greater than 1. Recent literature has stressed the lower performance of these combining methods (Newbold, Zumwalt and Kannan, 1987).

6 Conclusions

In this paper, a new algorithm-based procedure to increase the efficiency of forecasting combining methods is provided. The algorithm considers pseudo out-of-sample forecast as inputs. Results from Ericsson (1992), who shows that the forecast encompassing of a given model versus the other non-nested models is a sufficient condition for minimizing RMSFE, are used in the algorithm procedure. The basic idea is to compare all forecasting models with each other using the Harvey et al. (1998) encompassing test, to eliminate those are encompassed by others, and to use several forecast combining methods for combining the remaining forecasts. To assess the robustness of this procedure, an empirical application to Italian monthly industrial production using seven ISAE short-term forecasting models is provided. Results confirm the goodness of the algorithm we proposed.

References

Bai, J. & Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70, 191-221.

Bruno, G. & Lupi, C. (2004). Forecasting industrial production and the early detection of turning points. *Empirical Economics*, 29, 647-671.

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

Diebold, F.X. & Pauly, P. (1987). Structural Change and the Combination of Forecasts. *Journal of Forecasting*, 6, 21-40.

Ericsson, N. R. (1992). Parameter Constancy, Mean Square Forecast Errors, and Measuring Forecast Performance: An Exposition, Extensions, and Illustration. *Journal of Policy Modelling*, 4, 465-495.

Harvey, D., Leybourne, S. & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281-291.

Harvey, D. I., Leybourne, S. & Newbold, P. (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics*, 16, 254-259.

Harvey, D. & Newbold, P. (2000). Tests for multiple forecast encompassing, *J*ournal of Applied Econometrics, 15, 471-482.

Hendry, D. F. & Krolzig, H.M. (1999). Improving on Data mining reconsidered by K.D. Hoover and S.J. Perez. *Econometrics Journal*, 2, 202-219.

Granger, C. W. J. & Ramanathan, R. (1984). Improved Methods of Combining Forecasts. *Journal of Forecasting*, 3, 197-204.

Krolzig, H. M. & Hendry, D. F. (2001). Computer automation of general-to-specific model selection procedures. *Journal of Economic Dynamics and Control*, 25, 831-866.

Newbold, P. & Harvey, I. H. (2002) Forecast Combination and Encompassing. In: M.P. Clements & D.F. Hendry *A Companion to Economic Forecasting.* Blackwell Publishers, pp. 268-283.

Newbold, P., Zumwalt, J. K. & Kannan, S. (1987). Combining forecasts to improve earnings per share prediction: and examination of electric utilities. *International Journal of Forecasting*, 3, 229-238.

Rathjens, P. & Robins, R. P. (1993). Forecasting Quarterly Data Using Monthly Information. *Journal of Forecasting*, 12, 321-330.

Rünstler G. & Sédillot, F. (2003). Short-term Estimates of Euro Area Real GDP by Means of Monthly Data, ECB Working Paper, n. 276.

Stock, J. H. & Watson, M. W. (1998). Diffusion indexes. Working Paper 6702, *NBER*, 1-65.

Stock, J. H. & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indices. *Journal of Business and Economic Statistics*, 20, 147-162.

Stock J. H. & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23, 405-430.

Table 1: Forecast errors measures. Recursive and Rolling estimation

Models (recursive)	RMSFE(1)	RMSFE(2)	RMSFE(3)	RMSFE(6)
GW	0.0192	0.0211	0.0388	0.0423
Gas	0.0205	0.0203	0.0214	0.0270
GW_c	0.0201	0.0168	0.0202	0.0222
SE	0.0180	0.0182	0.0399	0.0379
VAR	0.0236	0.0240	0.0243	0.0206
ARIMA	0.0297	0.0265	0.0391	0.0413
Factor	0.0279	0.0210	0.0372	0.0368
Models (rolling)	RMSFE(1)	RMSFE(2)	RMSFE(3)	RMSFE(6)
GW	0.0375	0.0194	0.0202	0.0248
Gas	0.0342	0.0207	0.0215	0.0254
GW_c	0.0344	0.0206	0.0207	0.0230
SE	0.0178	0.0422	0.0412	0.0352
VAR	0.0314	0.0315	0.0303	0.0213
ARIMA	0.0367	0.0353	0.0353	0.0354
Factor	0.0388	0.0272	0.0400	0.0367

Table 2: Rank Classification. Recursive and Rolling estimation

	Recursive				Rolling			
rank	h=1	h=2	h=3	h=6	h=1	h=2	h=3	h=6
1	SE	GW_c	GW_c	VAR	SE	GW	GW	VAR
2	GW	SE	Gas	GW_c	VAR	GW_c	GW_c	GW_c
3	GW_c	Gas	VAR	Gas	Gas	Gas	Gas	GW
4	Gas	Factor	Factor	Factor	GW_c	Factor	VAR	Gas
5	VAR	GW	GW	SE	ARIMA	VAR	ARIMA	SE
6	Factor	VAR	SE	ARIMA	GW	ARIMA	Factor	ARIMA
7	ARIMA	ARIMA	ARIMA	GW	Factor	SE	SE	Factor

Table 3: Encompassing (HLN) test results. Recursive estimation.

Recursive	h=1	h=2	h=3	h=6
1° step	Best Model: SE	Best Model: GW_c	Best Model: GW_c	Best Model: VAR
$\mathrm{p-values}$ (Models)	0.150 (SE/GW)	0.004 (GW_c/GW)	0.220 (Gw_c/GW)	0.278 (VAR/GW)
	0.112 (SE/Gas)	0.073 (GW_c/Gas)	0.005 (GW_c/Gas)	0.422 (VAR/Gas)
	0.103 (SE/GW_c)	0.061 (GW_c/SE)	0.230 (GW_c/SE)	0.286 (VAR/Gw_c)
	0.183 (SE/VAR)	0.561 (GW_c/VAR)	0.375 (GW_c/VAR)	0.336 (VAR/SE)
	0.426 (SE/ARIMA)	0.794 (GW_c/ARIMA)	0.009 (GW_c/ARIMA)	0.381 (VAR/ARIMA)
	0.052 (SE/Factor)	0.880 (GW_c/Factor)	0.465 (GW_c/Factor)	0.261 (VAR/Factor)
2° step	Best Model: GW	Best Model: SE	Best Model: Gas	Best Model: -
$\mathrm{p-values}$ (Models)				
$\alpha = 0.25$	0.167 (GW/Gas)	0.508 (SE/GW)	0.194 (Gas/GW)	
	0.033 (GW/GW_c)	0.409 (SE/Gas)	0.420 (Gas/ARIMA)	
	0.031 (GW/VAR)		0.190 (Gas/SE)	
	0.075 (GW/Factor)			
$\alpha = 0.20$	0.167 (GW/Gas)	0.508 (SE/GW)	0.420 (Gas/ARIMA)	
	0.033 (GW/GW_c)	0.409 (SE/Gas)		
	0.031 (GW/VAR)			
	0.075 (GW/Factor)			
$\alpha = 0.15$	0.167 (GW/Gas)	0.508 (SE/GW)	0.420 (Gas/ARIMA)	
	0.033 (GW/GW_c)	0.409 (SE/Gas)		
	0.075 (GW/Factor)			
$\alpha = 0.10$	0.075 (GW/Factor)	0.508 (SE/GW)	0.420 (Gas/ARIMA)	
		0.490 (SE/Gas)		
$\alpha = 0.05, 0.01$	-	0.508 (SE/GW)	0.420 (Gas/ARIMA)	
3° step	Best Model: GW_c	Best Model: -	Best Model: VAR	Best Model: -
$\mathrm{p-values}$ (Models)				
$\alpha = 0.25$	0.220 (GW_c/Gas)		0.373 (VAR/GW)	
	0.322 (GW_c/VAR)		0.279 (VAR/SE)	
	0.151 (GW_c/Factor)			
$\alpha = 0.20$	0.220 (GW_c/Gas)			
	0.322 (GW_c/VAR)			
	0.151 (GW_c/Factor)			
$\alpha = 0.15, 0.10$	0.151 (GW_c/Factor)			
4° step	Best Model: Gas	Best Model: -	Best Model: -	Best Model: -
$\mathrm{p-values}$ (Models)				
$\alpha = 0.25, 0.20$	0.273 (Gas/Factor)			

Table 4: Encompassing (HLN) test results. Rolling estimation.

Rolling	h=1	h=2	h=3	h=6
1° step	Best Model	Best Model	Best Model	Best Model
	SE	GW	GW	VAR
p − values (Models)	0.154 (SE/GW)	0.874 (GW/Gas)	0.657 (GW/Gas)	0.269 (VAR/GW)
	0.070 (SE/Gas)	0.279 (GW/GW_c)	0.290 (GW/GW_c)	0.546 (VAR/Gas)
	0.243 (SE/GW_c)	0.256 (GW/SE)	0.274 (GW/SE)	0.256 (VAR/Gw_c)
	0.065 (SE/VAR)	0.306 (GW/VAR)	0.331 (GW/VAR)	0.597 (VAR/SE)
	0.595 (SE/ARIMA)	0.502 (GW/ARIMA)	0.214 (GW/ARIMA)	0.304 (VAR/ARIMA)
	0.206 (SE/Factor)	0.019 (GW/Factor)	0.197 (GW/Factor)	0.359 (VAR/Factor)
2° step	Best Model: VAR	Best Model: GW_c	Best Mode: GW_c	Best Model: -
p − values (Models)				
$\alpha = 0.25$	0.086 (VAR/GW)	0.007 (GW_c/Factor)	0.707 (GW_c/ARIMA)	
	0.049 (VAR/Gas)		0.467 (GW_c/Factor)	
	0.085 (VAR/GW_c)			
	0.029 (VAR/Factor)			
$\alpha = 0.20$	0.086 (VAR/GW)	0.007 (GW_c/Factor)	0.467 (GW_c/Factor)	
	0.049 (VAR/Gas)			
$\alpha = 0.15, 0.10$	0.049 (VAR/Gas)	0.007 (GW_c/Factor)		
$\alpha = 0.05$	-	0.007 (GW_c/Factor)		
$\alpha = 0.01$	-	-		
3° step	Best Model: Gas	Best Model: Gas	Best Model: -	Best Model: -
p − values (Models)				
$\alpha = 0.25$	0.776 (Gas/GW)	0.144 (Gas/Factor)		
	0.243 (Gas/GW_c)			
	0.075 (Gas/Factor)			
$\alpha = 0.20, 0.15, 0.10$	0.776 (Gas/GW)	0.144 (Gas/Factor)		
$\alpha = 0.05$		0.144 (Gas/Factor)		
4° step	Best Model: GW_c	Best Model -	Best Model -	Best Model -
p − values (Models)				
$\alpha = 0.25$	0.600 (GW_c/Factor)			

Table 5: Encompassing (HN) test results. Recursive and Rolling estimation.

Recursive	h=1	h=2	h=3	h=6
1° step	Best Model: SE	Best Model: GW_c	Best Model: GW_c	Best Model: VAR
F − test (p−values)	3.16 (0.069)	3.96 (0.041)	5.89 (0.018)	0.94 (0.380)
2° step	Best Model: GW	Best Model: SE	Best Model: Gas	Best Model: -
F − test (p−values)	4.12 (0.034)	0.63 (0.543)	2.65 (0.183)	
3° step	Best Model: GW_c	Best Model: −	Best Model: VAR	Best Model: -
F − test (p−values)	2.12 (0.182)		0.87 (0.452)	
4° step	Best Model: Gas	Best Model: −	Best Model: −	Best Model: -
F − test (p−values)	1.45 (0.281)			
Rolling	**h=1**	**h=2**	**h=3**	**h=6**
1° step	Best Model: SE	Best Model: GW	Best Model: GW	Best Model: VAR
F − test (p−values)	1.96 (0.064)	3.87 (0.044)	1.85 (0.210)	0.76 (0.517)
2° step	Best Model: VAR	Best Model: GW_c	Best Model: GW_c	Best Model: -
F − test (p−values)	2.80 (0.092)	2.14 (0.180)	0.35 (0.643)	
3° step	Best Model: Gas	Best Model: Gas	Best Model: VAR	Best Model: -
F − test (p−values)	1.80 (0.99)	1.31 (0.298)		
4° step	Best Model: GW_c	Best Model: −	Best Model: −	Best Model: -
F − test (p−values)	0.26 (0.763)			

Table 6: Relative RMSFE results. Recursive and Rolling estimation.

	Recursive				Rolling			
Combining method	h=1	h=2	h=3	h=6	h=1	h=2	h=3	h=6
Mean								
$\alpha = 0.25$	0.908^{***}	0.933^{**}	0.937^{**}	-	0.924^{**}	0.931^{**}	0.937^{**}	-
$\alpha = 0.20$	0.908^{***}	0.933^{**}	0.937^{**}	-	0.928^{**}	0.931^{**}	0.397^{**}	-
$\alpha = 0.15$	0.924^{**}	0.933^{**}	0.942^{**}	-	0.928^{**}	0.931^{**}	-	-
$\alpha = 0.10$	0.924^{**}	0.933^{**}	0.942^{**}	-	0.928^{**}	0.931^{**}	-	-
$\alpha = 0.05$	-	0.933^{**}	0.942^{**}	-	-	0.931^{**}	-	-
$\alpha = 0.01$	-	0.933^{**}	0.942^{**}	-	-	-	-	-
Median								
$\alpha = 0.25$	0.930^{**}	0.942^{**}	0.938^{**}	-	0.940^{**}	0.943^{**}	0.944^{**}	-
$\alpha = 0.20$	0.930^{**}	0.942^{**}	0.938^{**}	-	0.949^{**}	0.943^{**}	0.944^{**}	-
$\alpha = 0.15$	0.934^{**}	0.942^{**}	0.943^{**}	-	0.949^{**}	0.943^{**}	-	-
$\alpha = 0.10$	0.934^{**}	0.942^{**}	0.943^{**}	-	0.949^{**}	0.943^{**}	-	-
$\alpha = 0.05$	-	0.942^{**}	0.943^{**}	-	-	0.943^{**}	-	-
$\alpha = 0.01$	-	0.942^{**}	0.943^{**}	-	-	-	-	-
Tmean								
$\alpha = 0.25$	0.931^{**}	0.941^{**}	0.942^{**}	-	0.944^{**}	0.941^{**}	0.941^{*}	-
$\alpha = 0.20$	0.931^{**}	0.941^{**}	0.942^{**}	-	0.954^{**}	0.941^{**}	0.941^{*}	-
$\alpha = 0.15$	0.939^{**}	0.941^{**}	0.947^{**}	-	0.954^{**}	0.941^{*}	-	-
$\alpha = 0.10$	0.939^{**}	0.941^{**}	0.947^{**}	-	0.954^{**}	0.941^{*}	-	-
$\alpha = 0.05$	-	0.941^{**}	0.947^{**}	-	-	0.941^{**}	-	-
$\alpha = 0.01$	-	0.941^{**}	0.947^{**}	-	-	-	-	-
OLS								
$\alpha = 0.25$	0.979^{**}	0.990^{**}	0.985^{**}	-	0.954^{**}	0.985^{**}	0.987^{**}	-
$\alpha = 0.20$	0.979^{**}	0.990^{**}	0.985^{**}	-	0.967^{**}	0.985^{**}	0.987^{**}	-
$\alpha = 0.15$	0.997^{*}	0.990^{**}	0.990^{**}	-	0.967^{**}	0.985^{**}	-	-
$\alpha = 0.10$	0.997^{*}	0.990^{**}	0.990^{**}	-	0.967^{**}	0.985^{**}	-	-
$\alpha = 0.05$	-	0.990^{**}	0.990^{**}	-	-	0.985^{**}	-	-
$\alpha = 0.01$	-	0.990^{**}	0.990^{**}	-	-	-	-	-

Continued overleaf

Table 6: Continued.

Combining method	Recursive h=1	Recursive h=2	Recursive h=3	Recursive h=6	Rolling h=1	Rolling h=2	Rolling h=3	Rolling h=6
WLS ($\gamma = 1$)								
$\alpha = 0.25$	0.981^{**}	0.998^{*}	0.971^{*}	-	1.003	0.996^{*}	0.998^{*}	-
$\alpha = 0.20$	0.981^{**}	0.998^{**}	0.971^{*}	-	1.002	0.996^{*}	0.998^{*}	-
$\alpha = 0.15$	0.986^{**}	0.998^{**}	0.976^{*}	-	1.002	0.996^{*}	-	-
$\alpha = 0.10$	0.986^{**}	0.998^{**}	0.976^{*}	-	1.002	0.996^{*}	-	-
$\alpha = 0.05$	-	0.998^{**}	0.976^{*}	-	-	0.996^{*}	-	-
$\alpha = 0.01$	-	0.998^{**}	0.976^{*}	-	-	-	-	-
WLS ($\gamma = 3$)								
$\alpha = 0.25$	0.982	0.993^{*}	1.035	-	1.007	1.017	1.010	-
$\alpha = 0.20$	0.982	0.993^{*}	1.035	-	1.013	1.017	1.010	-
$\alpha = 0.15$	0.995	0.993^{*}	1.041	-	1.013	1.017	-	-
$\alpha = 0.10$	0.995	0.993^{*}	1.041	-	1.013	1.017	-	-
$\alpha = 0.05$	-	0.993^{*}	1.041	-	-	1.017	-	-
$\alpha = 0.01$	-	0.993^{*}	1.041	-	-	-	-	-
DMSFE ($\delta = 1$)								
$\alpha = 0.25$	0.948^{**}	0.959^{**}	0.960^{**}	-	0.963^{**}	0.953^{**}	0.958^{**}	-
$\alpha = 0.20$	0.948^{**}	0.959^{**}	0.960^{**}	-	0.974^{**}	0.953^{**}	0.958^{**}	-
$\alpha = 0.15$	0.969^{**}	0.959^{**}	0.965^{**}	-	0.974^{**}	0.953^{**}	-	-
$\alpha = 0.10$	0.969^{**}	0.959^{**}	0.965^{**}	-	0.974^{**}	0.953^{**}	-	-
$\alpha = 0.05$	-	0.959^{**}	0.965^{**}	-	-	0.953^{*}	-	-
$\alpha = 0.01$	-	0.959^{**}	0.965^{**}	-	-	-	-	-
DMSFE ($\delta = 0.9$)								
$\alpha = 0.25$	0.948^{**}	0.956^{**}	0.959^{**}	-	0.954^{**}	0.939^{**}	0.944^{**}	-
$\alpha = 0.20$	0.948^{**}	0.956^{**}	0.959^{**}	-	0.967^{**}	0.939^{**}	0.944^{**}	-
$\alpha = 0.15$	0.995^{**}	0.956^{**}	0.964^{**}	-	0.967^{**}	0.939^{**}	-	-
$\alpha = 0.10$	0.995^{**}	0.956^{**}	0.964^{**}	-	0.967^{**}	0.939^{**}	-	-
$\alpha = 0.05$	-	0.956^{**}	0.964^{**}	-	-	0.939^{**}	-	-
$\alpha = 0.01$	-	0.956^{**}	0.964^{**}	-	-	-	-	-

Notes: *, **, *** Indicates rejection of the null hypothesis of equal forecasting accuracy at 10%, 5%, 1%.

Author: Mauro Costantini, Carmine Pappalardo

Title: Combination of Forecast Methods Using Encompassing Tests: An Algorithm-Based Procedure

Reihe Ökonomie / Economics Series 228

Editor: Robert M. Kunst (Econometrics)
Associate Editors: Walter Fisher (Macroeconomics), Klaus Ritzberger (Microeconomics)

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996