Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 23 (2007) 695 – 705
international journal of forecasting
www.elsevier.com/locate/ijforecast

Predictive financial models of the euro area: A new evaluation test

Ekaterini Panopoulou *

National University of Ireland, Maynooth, Ireland
Department of Statistics and Insurance Science, University of Piraeus, Greece

Abstract

This paper investigates the predictive ability of financial variables for euro area growth. Our forecasts are built from univariate autoregressive and single equation models. Euro area aggregate forecasts are constructed both by employing aggregate variables and by aggregating country-specific forecasts. The forecast evaluation is based on a recently developed test for equal predictive ability between nested models. Employing a monthly dataset from the period between January 1988 and May 2005 and setting the out-of-sample period to be from 2001 onwards, we find that the single most powerful predictor on a country basis is the stock market returns, followed by money supply growth. However, for the euro area aggregate, the set of most powerful predictors includes interest rate variables as well. The forecasts from pooling individual country models outperform those from the aggregate itself for short run forecasts, while for longer horizons this pattern is reversed. Additional benefits are obtained when combining information from a range of variables or combining model forecasts.

JEL classification: E52; C14

Keywords: Forecasting accuracy; Financial variables; Output growth; Aggregation

1. Introduction

A vast body of literature in finance and macroeconomics is devoted to the forecasting ability of financial variables for real economic activity. The empirical evidence is mixed, and the results are not robust with respect to model specification, sample choice or forecast horizon (see Stock & Watson, 2003, for a review of the empirical literature).

A voluminous body of literature exists on the choice of candidate variables, but there is little consensus as to what the most appropriate variables. We choose a set of variables that are the ones most frequently used in the literature. We include forward-looking financial variables – stock market returns, short-term interest rates, interest rates spreads and the dollar exchange rate – that are thought to embody future economic expectations. Studies such as Barro (1990), Fama (1990), Lee (1992), Estrella and Mishkin (1998), Hassapis and Kalyvitis (2002), Hassapis (2003) and Panopoulou, Pittis, and Kalyvitis (2006), among others, find that stock market

* Department of Statistics and Insurance Science, University of Piraeus, Greece. Tel.: +30 210 4142728; fax: +30 210 4142340.
E-mail address: apano@unipi.gr.

0169-2070/$ - see front matter © 2007 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2007.04.001

returns improve forecasting ability. Interest rate measures have also enjoyed success in predicting output growth. Both short-term rates (see Bernanke & Blinder, 1992) and (more usually term) spreads are used (see Boulier & Stekler, 2000; Davis & Fagan, 1997; Estrella & Hardouvelis, 1991; Estrella & Mishkin, 1997, 1998; Harvey, 1988; Stock & Watson, 2003). These have mixed forecasting performance, and there is evidence that in the US that these variables' ability to predict output growth has fallen over the past two decades. We also investigate the forecasting performance of the domestic money supply, the dollar exchange rate as well as extraneous data such as US growth and oil prices. Money supply growth, exchange rates and oil prices have been employed by Stock and Watson (2003), among others, while the effect of US variables on their EU equivalents has been documented by Marcellino, Stock, and Watson (2003) and Banerjee, Masten, and Marcellino (2005).

With the exception of a few cases, the aforementioned studies have concentrated on and examined the predictive ability of financial variables for forecasting US future growth. Similar evidence for the euro area countries is quite scarce, and what there is is more recent. Studies such as Davis and Fagan (1997), Forni, Hallin, Lippi, and Reichlin (2003), Sensier, Artis, Osborn, and Birchenhall (2004), Marcellino et al. (2003), Moneta (2005), Duarte, Venetis, and Paya (2005), and Banerjee et al. (2005), among others, confirm the widely held belief that a variety of variables act as leading indicators for output growth, albeit in some counties more than others, and at different horizons.

We concentrate on single equation linear models, which are often found to outperform both non-linear alternatives (Banerjee & Marcellino, 2006) and multivariate models (Marcellino et al., 2003). Specifically, we examine a range of nested models, using the simple autoregressive model as a benchmark and augmenting it with a number of the aforementioned candidate variables. We initially assess the forecasting ability of the models by analysing their Mean Squared Forecast errors (MSFE). We then extend this approach by testing for statistical differences in forecasting accuracy, using the OOS-F statistic of equal predictive ability for nested models developed by McCracken (2004). The employment of this testing methodology gives us a clear comparison between the competing models, and thus provides an advance on other studies of economic forecasting within the euro area.

Our results from testing the forecasting accuracy of the variables at hand are in line with the consensus that for some countries and horizons, some variables contain useful information for predicting future growth. On a country basis, we find that in the vast majority of cases, financial variables add significant predictive content over and above that already contained in the autoregressive model, with the exception of the 3-month horizon where only marginal gains are observed. However, at the aggregate euro area level, our results are more promising, as a longer list of candidate variables proves to provide more accurate forecasts either at an aggregate level or when pooled forecasts from country-specific models are considered.

The layout of this paper is as follows: Section 2 outlines the methodology for testing the out-of-sample predictability of financial variables for growth. Section 3 presents and comments on the empirical results for the euro area countries and the euro area as a whole, and Section 4 summarizes the main findings of the paper.

2. Construction of out-of-sample forecasts and evaluation

In this section, we briefly review the forecasting methodology, which is fairly standard (see, inter alia, Marcellino et al., 2003; Stock & Watson, 2003). Specifically, we estimate several univariate models for each series to be forecast, and focus on forecast horizons (h) of 1, 3, 6 and 12 months. Contrary to the textbook approach of estimating a one-step ahead model and then iterating it forward to get the h-step predictions, we set the h-step ahead variable to be forecast, y^h_{t+h}, to be equal to $\sum_{s=t+1}^{t+h} y_s$ In our case, the variable of interest is the output growth, y^h_{t+h}, which represents the growth of output over the next h periods. The models considered are all nested within the following class of Autoregressive Distributed Lag (ADL) models:

$$y^h_{t+h} = c + \alpha(L)y_t + B(L)'Z_t + \varepsilon^h_{t+h}, \tag{1}$$

where c is a constant, $\alpha(L)$ is a scalar lag polynomial, $B(L)$ is a vector lag polynomial, and Z_t is a vector of financial (predictor) variables. Our specification of Z_t differentiates between the models. The number of lags for both y_t and Z_t is selected by the Schwartz Bayesian information criterion (SIC), setting the maximum lag length at 12 to avoid estimating any models with low degrees of freedom.

Not including financial variables in (1), i.e. setting $B(L)$ equal to zero, provides us with the simple autoregressive model (AR) which will be used as a benchmark when evaluating the forecasts of the various models. The remaining models include either one of the elements of Z_t at a time, or all of them simultaneously.

Evaluating the forecasting accuracy of the candidate models is just important as constructing the forecasts. In this respect, the estimation procedure is designed to allow us to implement formal statistical tests for the comparison of the forecasts provided by the various models. Specifically, we first estimate an AR model for each country by setting $B(L)$ equal to zero. Our simulated out-of-sample forecasting experiment proceeds recursively in the following manner. For each date of the out-of-sample forecasting period, the AR model is re-estimated by keeping the lag-order fixed, providing us with a sequence of forecasts. We then estimate alternative models, adding Z_t to our model. We keep the order of $\alpha(L)$ fixed, and once more use the SIC to select the order of $B(L)$. Consequently, the AR model is always nested within the alternative models. The lag structure of the models, however, is allowed to vary across countries.

The forecasting performance of the various models is assessed using the simulated out-of-sample mean squared forecast error (MSFE), relative to the MSFE of the benchmark AR model. A value of this ratio which is lower than one suggests superiority of the respective model over a simple AR model, and indicates that the candidate variable is a useful predictor of output growth. However, a ratio lower than one does not necessarily mean that the alternative model generates better forecasts than the benchmark, as this lower MSFE may be due to sample variation. In order to establish the statistical significance of this ratio, one has to test the hypothesis that the population relative MSFE is equal to 1, against the alternative of a ratio less than one. Techniques for comparing the forecasting performance of two nested models, which we need because the AR model is always nested within the remaining models considered, have only recently been developed. In this study, we use the following F-statistic proposed by McCracken (2004):

$$OOS - F = \frac{\sum_{t=1}^{P} [\epsilon_{1,t}^2 - \epsilon_{2,t}^2]}{P^{-1} \sum_{t=1}^{P} \epsilon_{2,t}^2} \tag{2}$$

where $\epsilon_{i,t}$, $i=1, 2$ are the forecast errors of the restricted and unrestricted models, respectively, and P is the number of out-of-sample observations. Under the null hypothesis, the two models have equal MSFEs, while under the alternative, the MSFE of the unrestricted model is less than that of the restricted one. The limiting distribution of the aforementioned test-statistic is non-standard, and numerical estimates of the asymptotic critical values for valid inference are provided by McCracken (2004). These values depend on the ratio of in sample to out-of-sample observations, and the number of parameter restrictions.

While the asymptotic critical values of the aforementioned test are valid for one-step ahead horizons, these values cannot be employed for horizons greater than one, as the null distribution of the OOS-F test statistic is non-standard and non-pivotal for horizons greater than one, and also for nested models. For these cases, Clark and McCracken (2005) recommend basing inference on a bootstrap procedure along the lines of Kilian (1999). Following this recommendation, we base our inferences on this bootstrap procedure.[1]

3. Model forecasts and evaluation

In this section we report and discuss the results of applying the techniques outlined in the previous section to examining the empirical relationship between growth and financial variables in the euro area.[2]

[1] We do not describe this procedure for the sake of brevity. A detailed description of this procedure can be found in Clark and McCracken (2005).

[2] All the reported results were obtained by programs written in E-views 4.1, with the exception of the bootstrap procedure, and are available from the author upon request. The critical values for the h-step ahead forecasts were obtained using programs written in Gauss, and are available from David Rapach's website: http://pages.slu.edu/faculty/rapachde/Research.htm.

3.1. Data and models

Our dataset for the 12 euro area countries is monthly, and covers the period from January 1988 to May 2005. As a measure of growth, we employ the growth in the industrial production index for the 12 euro area countries. The financial variables we consider are the term spread, the real stock market returns, real money supply growth, exchange rate returns and short-term interest rates. We also include two non-financial variables, the oil price and US growth.[3]

Our simulated out-of-sample forecasting experiment is conducted using the semi-recursive methodology outlined in the previous section. The out-of-sample forecast period is 2001:4 to 2005:5 (50 observations), covering the more recent period of the monetary union and generating a ratio of out-of-sample (P) to in-sample observations (R) equal to approximately 0.3 ($P/R=0.3$). For the period 2001:4 onwards, we reestimate all the candidate models by adding one observation at a time. The h-step ahead forecasts are generated for the periods of 1, 3, 6 and 12 months, and the corresponding MSFEs are calculated.

The models estimated in the forecasting experiment are the following:

Model (1) The benchmark AR model, i.e. Z_t, is excluded from (1).

Models (2)–(8) The AR specification is combined with lags of Z_t, which contains only one element of the available predictors.

Model (9) The AR specification is combined with lags of all of the candidate variables simultaneously. This model combines information (CI) in order to generate an ultimate forecast.

We also consider the issue of combining the forecasts from models (2)–(8) in order to generate a country-specific combination forecast (CF). One could apply the theory of optimal linear combination forecasts (see Bates & Granger, 1969; Granger & Ramanathan, 1984), of which suggests that combination forecasts are weighted averages of individual forecasts with weights obtained as regression coefficients of the true future value on the various forecasts. Given that optimal linear combination forecasts are often found to be inferior to simpler ones, such as means or medians, we construct the respective combination forecasts by simply averaging the forecasts of the individual models (see Stock & Watson, 2004).

The aforementioned models are estimated for the 12 euro area countries and the euro area aggregate. The relevant aggregated series are constructed as the weighted average of the (transformed) country level data for all 12 countries. A fixed-weighting scheme is employed, using each country's GDP share in the euro area aggregate in PPP exchange rates averaged over 2005.[4]

3.2. 1-step ahead forecasts

The results for the 1-month forecast horizon are reported in Table 1. Specifically, the second row reports the MSFE of the benchmark AR model in decimal values, while rows 4 to 11 tabulate the ratio of the MSFE of models (2) to (9) relative to the AR benchmark. The last row of Table 1 reports the MSFE ratio of the average of the forecasts of models (2) to (8). In addition to forecasting the euro area aggregates directly using the respective aggregated series as independent variables, we also consider pooling country-specific forecasts in order to construct the euro area forecast. These pooled forecasts are constructed by employing the same fixed-weighting scheme using GDP weights, and the last column reports the respective figures for euro area pooled forecasts. If the country-specific models are time invariant, correctly specified, and have parameters that differ across countries, pooling country-specific forecasts would give more accurate forecasts than the ones based on aggregated series (see Lutkepohl, 1987).

To evaluate our forecasts, we calculate the OOS-F statistic from Eq. (2), which tests for the statistical significance of the ratio of the forecasts. The value of the OOS-F statistic is then compared to the corresponding tabulated values from McCracken (2004), taking into account the number of parameter restrictions and the ratio of out-of-sample to in-sample

[3] A list of the variables, along with details about data transformations and sources, is given in the Appendix.

[4] Source: Statistics Pocket Book, April 2006, European Central Bank.

Table 1
Out-of-sample forecasts: 1 month forecast horizon

Out-of-sample MSFE	Austria	Belgium	Finland	France	Germany	Greece	Ireland	Italy	Luxembourg	Netherlands	Portugal	Spain	Euro area	Pooled euro area
(1) AR[1]	2.790	1.550	4.280	0.709	1.430	4.500	30.870	0.456	7.460	2.980	4.260	1.510	0.438	0.495
MSFE relative to AR														
(2) AR+spread	***0.912***	1.004	1.071	**0.974**	***0.886***	1.036	1.007	**0.966**	0.990	1.019	1.048	1.014	**0.975**	***0.923***
(3) AR+stock market	**0.986**	0.986	***0.767***	***0.797***	***0.886***	1.239	0.993	**0.944**	***0.901***	0.999	1.009	**0.980**	***0.765***	***0.805***
(4) AR+money supply	0.997	**0.973**	***0.905***	1.072	1.089	0.999	**0.975**	1.082	**0.975**	1.000	1.011	1.022	1.990	1.069
(5) AR+exchange rates	1.000	1.015	1.076	1.006	***0.928***	***0.910***	0.997	***0.872***	1.090	1.119	**0.964**	**0.954**	***0.936***	***0.923***
(6) AR+oil	***0.907***	1.180	**0.974**	**0.958**	***0.939***	1.117	**0.978**	***0.883***	0.987	1.116	***0.933***	1.066	0.989	0.995
(7) AR+short-term interest rates	0.999	1.051	1.012	**0.963**	**0.960**	1.004	1.003	0.997	1.001	**0.979**	1.007	1.005	**0.978**	**0.961**
(8) AR+US growth	***0.886***	1.344	1.113	***0.876***	0.985	1.072	1.084	**0.971**	0.997	***0.917***	**0.970**	1.073	1.075	***0.893***
(9) AR+all	***0.836***	***0.886***	***0.696***	***0.655***	***0.705***	1.119	***0.828***	***0.626***	***0.953***	1.122	***0.632***	***0.948***	***0.547***	***0.590***
(10) Pooled model of (2)–(8)	***0.936***	0.993	***0.928***	***0.908***	***0.917***	1.015	**0.961**	***0.899***	**0.972**	0.986	***0.939***	**0.981**	***0.875***	***0.907***

(1) The MSFE is calculated for real growth in decimal values.
(2) The pooled euro area refers to the combination of forecasts using a fixed-weighting GDP scheme based on GDP shares calculated in 2005 PPP exchange rates.
(3) Bold denotes significance at the 10% level based on the McCracken (2004) critical values.
(4) Italics denote significance at the 10% level based on robust for data-mining bootstrap critical values.

observations. Given that McCracken (2004) does not tabulate critical values for P/R equal to 0.3, we base our inference on interpolated critical values of P/R equal to 0.2 and 0.4. An issue arises with respect to the selection of appropriate values for testing the significance of the forecasting accuracy of the combined and euro area pooled forecasts. To be on the safe side, the combined forecasts are evaluated on the basis of the respective critical values of model (9), which includes all variables simultaneously. With respect to the pooled euro area model, inference is based on critical values of the respective euro area aggregate models. To save space and increase the readability of the paper, we do not report the values of the respective statistics. Instead, we denote rejection of the null hypothesis of equal forecasting ability at the 10% significance level by bold MSFE ratios, and only comment on these outcomes.

The information content in Table 1 may be summarised as follows:

(i) On a country/variable basis, at least one financial variable is helpful in predicting the next month's growth (with a significant ratio relative to the benchmark AR model). The most informative of the financial variables appears to be the stock market returns, which provides additional information for 7 of the 12 euro area countries, followed by developments in exchange rates, which improves forecasts for 5 out of the 12 countries. Quite interestingly, the non-financial variables employed, namely the US growth and oil prices, lead to an improvement in the forecasts in 5 and 7 countries, respectively.

(ii) Combining information leads to a significant improvement in the growth forecasts for all of the countries at hand except for Greece and the Netherlands, while combining forecasts from the individual models on a country basis results in slightly reduced gains in terms of forecasting ability. In detail, the countries that combining models does not cause to improve upon the benchmark are Belgium, Greece and the Netherlands.

(iii) Turning to the euro area aggregate (column 14), all of the financial variables, with the exception of money supply, provide additional information on growth either employed as single predictors or combined in the same model. The non-financial variables do not improve the forecasting accuracy for the euro area aggregate. The same picture emerges when the forecasts of the euro area growth are generated by taking the GDP-weighted average forecast (column 15), with the exception of US growth. Moreover, the absolute euro area benchmark MSFE turns out to be one of the lowest when compared to the individual countries, suggesting that the forecasting accuracy improves when models are employed for the euro area as a whole.

(iv) All of the methods of combining information or forecasts for the euro area turn out to be successful, as is demonstrated by the significance of the MSFE ratios. However, the lowest relative ratios are attained by combining information either at the aggregate euro area level or at the pooled level.

3.3. h-step ahead forecasts

We next turn our attention to the accuracy of growth forecasts at longer horizons, namely 3, 6 and 12 months ahead. Tables 2–4 report the MSFEs of the benchmark specification, along with the relative MSFEs of the other specifications considered. As previously, bold denotes rejection of the null hypothesis of equal forecasting ability at the 10% significance level. However, for these multi-step direct forecasts, the significance of the calculated OOS-F statistics is based on bootstrapped critical values.[5]

As expected, the benchmark AR forecasts become less accurate as we increase the time horizon; however, instances of improvement in forecast accuracy exhibit a U-shaped pattern. In particular, the forecasting performance of the competing models deteriorates at the 3 month horizon, and then improves at the 6 and 12 month horizons. In more detail, at the 3 month horizon, the most powerful predictor is the stock market return, as it improves the forecast accuracy in 7 out of the 12 countries, followed by the US growth, while the remaining variables prove significant in only 2–3 countries. At the longer horizons, the candidate variables improve forecasts in at least 4 countries, with

[5] Detailed tables of the bootstrapped critical values at the 1%, 5% and 10% critical levels for each specification and horizon considered are available upon request from the author.

Table 2
Out-of-sample forecasts: 3 month forecast horizon

Out-of-sample MSFE	Austria	Belgium	Finland	France	Germany	Greece	Ireland	Italy	Luxembourg	Netherlands	Portugal	Spain	Euro area	Pooled euro area
(1) AR[(1)]	4.280	3.510	8.170	1.170	1.780	8.540	41.770	1.310	11.070	5.360	4.380	2.080	0.577	0.810
MSFE relative to AR														
(2) AR + spread	**0.970**	1.018	1.119	1.049	1.079	1.004	1.001	1.123	0.991	1.134	1.125	1.031	1.416	1.094
(3) AR + stock market	**0.971**	1.381	0.990	***0.842***	***0.897***	**0.966**	1.019	**0.946**	1.091	**0.975**	1.063	***0.911***	***0.728***	***0.696***
(4) AR + money supply	1.002	***0.913***	1.128	1.059	**0.944**	0.991	**0.972**	1.079	1.029	1.011	0.990	1.073	1.238	1.027
(5) AR + exchange rates	1.033	1.190	1.058	**0.948**	1.164	1.000	1.003	0.997	**0.973**	1.058	1.000	1.121	1.056	1.005
(6) AR + oil	**0.974**	1.028	**0.978**	1.064	1.064	0.999	1.013	1.168	1.001	1.998	1.076	0.990	**0.945**	0.959
(7) AR + short-term interest rates	1.003	**0.980**	1.008	1.004	0.977	1.016	0.996	**0.983**	0.997	1.093	**0.988**	1.079	1.097	1.005
(8) AR + US growth	0.997	**0.945**	**0.979**	**0.956**	***0.923***	1.172	1.112	**0.957**	1.028	1.206	1.054	**0.979**	1.078	***0.845***
(9) AR + all	0.992	1.771	1.287	1.341	1.193	1.011	1.101	1.379	1.029	1.225	1.503	1.296	1.307	1.046
(10) Pooled model of (2)–(8)	0.985	0.988	0.970	**0.953**	**0.943**	0.987	1.005	0.978	1.004	1.043	0.993	0.992	0.957	***0.899***

(1) MSFE calculated for real growth in decimal values.
(2) Pooled euro area refers to combination of forecasts using a fixed-weighting GDP scheme based on GDP shares calculated in 2005 PPP exchange rates.
(3) Bold denotes significance at the 10% level based on bootstrapped critical values.
(4) Italics denotes significance at the 10% level based on robust for data-mining bootstrap critical values.

Table 3

Out-of-sample forecasts: 6 month forecast horizon

Out-of-sample MSFE	Austria	Belgium	Finland	France	Germany	Greece	Ireland	Italy	Luxembourg	Netherlands	Portugal	Spain	Euro area	Pooled euro area
(1) AR$^{(1)}$	6.390	5.920	12.510	1.680	3.830	6.780	61.220	3.160	14.020	7.460	5.840	2.480	2.350	2.750
MSFE relative to AR														
(2) AR+spread	**0.910**	***0.799***	1.142	1.095	***0.826***	1.141	1.024	1.028	***0.911***	0.924	1.285	***0.896***	0.908	**0.894**
(3) AR+stock market	0.990	1.031	***0.691***	***0.816***	***0.866***	1.136	**0.933**	**0.953**	0.980	***0.899***	1.004	**0.958**	***0.828***	**0.937**
(4) AR+money supply	**0.964**	***0.805***	***0.724***	1.109	**0.940**	**0.977**	***0.881***	1.036	1.008	1.007	**0.950**	0.991	***0.869***	**0.902**
(5) AR+exchange rates	1.074	1.097	1.061	1.145	1.272	1.010	**0.981**	**0.951**	***0.983***	**0.938**	***0.806***	1.237	1.076	1.212
(6) AR+oil	1.028	**0.977**	0.993	**0.966**	1.013	1.097	***0.877***	**0.955**	**0.951**	1.015	***0.806***	1.048	**0.959**	1.026
(7) AR+short-term interest rates	0.995	0.994	**0.980**	0.983	0.986	**0.987**	1.000	***0.844***	0.996	1.001	**0.961**	1.014	**0.936**	***0.815***
(8) AR+US growth	**0.959**	***0.896***	1.258	***0.870***	0.970	0.990	1.014	1.032	0.987	1.020	**0.971**	***0.847***	1.079	1.089
(9) AR+all	1.026	***0.716***	***0.368***	***0.534***	***0.717***	1.339	***0.635***	***0.549***	***0.840***	***0.776***	***0.562***	**0.931**	***0.424***	0.962
(10) Pooled model of (2)–(8)	0.946	***0.866***	***0.830***	***0.868***	***0.873***	0.979	***0.931***	***0.895***	0.958	0.943	***0.883***	***0.856***	***0.835***	0.919

Notes: see Table 2.

Table 4

Out-of-sample forecasts: 12 month forecast horizon

Out-of-sample MSFE	Austria	Belgium	Finland	France	Germany	Greece	Ireland	Italy	Luxembourg	Netherlands	Portugal	Spain	Euro area	Pooled area euro
(1) AR$^{(1)}$	11.930	8.200	31.260	4.830	9.560	10.230	104.140	8.050	18.280	10.190	14.210	5.500	5.820	4.730
MSFE relative to AR														
(2) AR+spread	**0.875**	***0.826***	1.373	1.159	**0.641**	1.331	1.022	0.943	***0.836***	0.965	1.459	0.920	**0.856**	1.336
(3) AR+stock market	**0.983**	1.071	**0.915**	***0.836***	1.020	1.131	**0.947**	**0.959**	1.034	**0.914**	***0.731***	**0.961**	**0.943**	1.281
(4) AR+money supply	***0.677***	***0.857***	***0.678***	1.040	**0.859**	**0.938**	***0.850***	0.993	**0.951**	1.008	**0.895**	1.172	***0.750***	**0.966**
(5) AR+exchange rates	**0.967**	1.084	1.172	1.003	1.171	**0.933**	***0.847***	**0.891**	1.016	**0.969**	***0.727***	1.639	1.140	1.501
(6) AR+oil	0.986	**0.945**	**0.858**	**0.970**	**0.952**	1.166	***0.890***	**0.894**	0.993	**0.908**	***0.856***	0.993	**0.898**	1.091
(7) AR+short-term interest rates	1.034	**0.888**	***0.601***	***0.783***	**0.958**	1.009	0.998	***0.734***	0.993	0.985	**0.961**	**0.925**	**0.896**	**0.847**
(8) AR+US growth	**0.972**	**0.927**	***0.817***	***0.587***	**0.880**	1.560	**0.953**	***0.767***	0.974	0.992	0.982	1.121	1.004	1.420
(9) AR+all	***0.626***	***0.568***	***0.327***	***0.324***	***0.628***	1.604	***0.475***	***0.266***	***0.880***	***0.732***	***0.501***	1.126	***0.352***	1.847
(10) Pooled model of (2)–(8)	**0.872**	***0.840***	***0.703***	***0.789***	***0.835***	0.979	***0.891***	***0.798***	0.942	0.944	***0.828***	0.929	***0.769***	1.132

Notes: see Table 2.

the most valuable predictors being the stock market returns and the money supply growth, which improves predictions in 7 and 8 countries, respectively, for the two horizons under consideration. For the most part, our results are consistent with the existing literature on country-specific forecasts. For example, Stock and Watson (2003) find evidence against the predictive ability of the term spread for France, Germany and Italy for their more recent sample and the 4-quarter horizon, while our results for this horizon indicate that the term spread is a useful predictor for Germany. Similarly, the authors find that stock market returns are useful only for France, while our findings indicate that stock market returns are useful for Italy as well.

Turning to the euro area forecasts, a similar picture emerges. In particular, at a 3 month horizon, only stock market returns and oil prices help forecast the euro area aggregate, with the addition of US growth when the pooled forecasts are considered. Moving to longer horizons, the set of valuable predictors is enriched with interest rate variables, namely the term spread and short-term interest rates, and monetary variables, i.e. the money supply growth. For these horizons, neither the dollar exchange rate nor the US growth contain any useful information for euro area growth. These variables were identified as leading indicators by Banerjee et al. (2005), albeit in quite a different forecasting set-up. Specifically, the authors found that short-term interest rates and commodity prices do better than a simple AR model in more than 7 of 16 evaluation periods, while the success of exchange rates, money supply growth, and the term spread is confined to 3–6 of the evaluation periods.

Our finding that aggregate euro area forecasts are more accurate than pooled forecasts is in contrast with Marcellino et al. (2003), who found that forecasts constructed by aggregating the country-specific models are more accurate than forecasts constructed using the aggregate data.

Addressing the issue of combining information or combining forecasts, we have to note that neither of these methods leads to significant gains, as combining models only leads to significant MSFE ratios for the 3 month horizon in the cases of Germany, France and the pooled euro area. This picture is reversed for longer horizons, since we find that both methods improve forecasts substantially. In particular, combining information forecasts is marginally superior to combining models forecasts, with the former improving forecasts in 11 of 14 cases and the latter in 9 of 14 cases. This finding is consistent with Stock and Watson (2003), who find that combination forecasts improve upon the benchmark AR specification, sometimes by a substantial amount.

Evaluating the euro area forecasts generated by the various methods, our results indicate that at a 3 month horizon, significant gains emerge only from pooled combined forecasts, while at longer horizons both CI and CF methods for the euro area aggregate work well. In particular, the more accurate of the two is CI, yielding MSFE ratios as low as 0.424 and 0.352 for the 6 and 12 month horizons, respectively. In an extensive analytical and Monte Carlo study, Huang and Lee (2007) demonstrate that in the majority of the cases considered. CF tends to outperform CI, especially in small samples where the parameter uncertainty is greater. Our results, however, do not support such a conjecture, with the exception of the 3 month horizon.

3.4. Overall assessment and further issues

When we were attempted to rank the performance of the financial variables, the leading predictor financial variable turned out to be the returns of the stock market, as it improved the forecasting accuracy in 65% of the cases considered, followed by money supply growth, which succeeded in 45% of the cases. Interestingly, the figures for the non-financial variables employed, namely the oil price and US growth, hover around 47%. These percentages, however, fluctuate with the horizon considered.

With respect to the performance of the OOS-F test statistic, we have to note that the employment of the test makes a difference. Specifically, 63% of the competing models considered turned out to improve forecasts, as indicated by a ratio of MSFEs which is lower than unity. This percentage increases to 70% when the 3 month horizon is excluded, the horizon that is marked with the worst performance. Quite interestingly, 75% of the lower than unity ratios proved significant. This result suggests that when we compare forecasts on the basis of the MSFE ratio only, we reach a false conclusion 25% of time.

Next, we consider whether the significant results presented so far on the basis of the OOS-F test statistic are due to data mining across a larger dataset of financial variables.[6] Although the conventional wisdom suggests that out-of-sample tests of predictability guard against data-mining, Inoue and Kilian (2004) challenged this notion by arguing that both in-sample and out-of-sample tests are equally susceptible to data mining. To address this issue we compute data-mining-robust bootstrap critical values based on a version of the data mining bootstrap procedure developed by Inoue and Kilian (2004), a detailed description of which can be found in Rapach and Wohar (2004, 2006). Italics in Tables 1–4 denote the rejection of the null hypothesis of equal forecasting ability at the 10% significance level based on these bootstrapped critical values. Given that these critical values are computed from the empirical distribution of the maximal OOS-F statistic across candidate variables, a deterioration in the forecasting accuracy of the predictors considered is expected.[7] Interestingly though, a sufficient percentage of the significant MSFE ratios survive the robust data-mining values, with percentages varying across horizons. For the 1-month horizon, 67% of the significant values remained significant, though this percentage falls to 60% and 51% for the 6 and 12 month horizons, respectively. The performance of the predictors for the 3 month horizon is quite disappointing, with successes accounting for only 28%. However, for every horizon considered, our results for the euro area aggregate or the pooled euro area forecasts remain unaffected by this stricter testing setting, especially when CI or CF methods are employed.

4. Conclusions

We compare forecasting models of economic growth in the context of the euro area, which now formulates economic policy to accommodate 12 countries. These countries differ greatly in terms of their economic and financial development, and this diversity makes the forecaster's problem even more difficult. We focus on single linear equations which have been shown to perform relatively well in times of economic change. Specifically, we focus on a range of nested models, using a simple AR model as our benchmark. We augment this with a number of financial variables; and by employing a recently developed test of equal forecasting ability for nested models, we test whether they add predictive content over and above that contained in the benchmark.

Our main findings can be summarised in two parts. Firstly, at the country level, none of the financial variables systematically outperform the benchmark. Admittedly, most of the variables manage to improve the forecast precision for some country and at some horizon, but it is not possible to identify patterns that would allow a forecaster to be confident that a particular variable adds predictive value across countries. The most prominent variable, though, is the stock market returns, which improves the forecasts in around two thirds of the cases considered. However, the safer strategy seems to be that of combining information from a range of variables in to one model or combining forecasts from a variety of models, since such a method succeeds in improving the forecast accuracy in the majority of the euro area countries.

Secondly, at the aggregate level, our results are more promising, since our selected financial variables also deliver more consistent performances over differing forecast horizons. In particular, all of the financial variables, with the exception of exchange rates, provide more accurate forecasts at least three of the four horizons considered. Our findings with respect to combining information or combining forecasts are relevant to the euro area as well. From a policy perspective, the safest way to conduct growth forecasts for the euro area is to rely on euro area aggregate data and combine information from a range of financial variables. Alternatively, combination forecast methods based on a GDP-weighting scheme can provide quite accurate forecasts.

Acknowledgements

Financial support from the EU under the "PYTHAGORAS: Funding of research groups in the University of Piraeus" through the Greek Ministry of National Education and Religious Affairs is greatly

[6] We thank an anonymous referee for bringing this issue to our attention.

[7] For the sake of brevity, we do not describe this procedure. The data-mining-robust bootstrap critical values for *h*-step ahead forecasts were obtained by programs written in Gauss and are available from David Rapach's website: http://pages.slu.edu/faculty/rapachde/Research.htm. Detailed tables of the bootstrapped critical values at the 1%, 5% and 10% critical levels for each horizon considered are available upon request from the author.

acknowledged. The author is grateful to the Editor Michael Clements, an anonymous Associate Editor, two anonymous referees of the journal, S. Kalyvitis, N. Pittis, D. Thomakos, F. Moneta, and participants at the Advances in Financial Forecasting section of the ICMMSE 2005 conference for helpful comments and suggestions. The usual disclaimer applies.

Appendix A. Data appendix

The variables we consider, along with data transformations and sources, are:

- the term spread (the difference between a long-term bond yield, mainly a 10-year one, and a short-term interest rate, mainly a three-month Treasury Bill obtained from the IMF, International Financial Statistics, Source: Ecowin);
- real stock market returns (first difference in the log-levels of Datastream-calculated composite indices deflated by the Consumer Price Index, CPI);
- real money supply growth (first difference in the log-levels of CPI deflated M3 money supply, Source: Datastream);
- exchange rate returns (first difference in the log-levels of the exchange rates against the US dollar, Source: Ecowin);
- oil price (first difference in the log-levels of the oil price, Source: FRED database);
- short-term interest rates (first difference in the levels of (mainly) a three-month Treasury Bill obtained from the IMF, International Financial Statistics, Source: Ecowin);
- US growth (first difference in the log-levels of the industrial production index, Source: FRED database);
- Euro area countries, growth (first difference in the log-levels of the industrial production index, Source: OECD Economic Indicators, Datastream).

References

Banerjee, A., & Marcellino, M. (2006). Are there any reliable leading indicators for US inflation and GDP growth? *International Journal of Forecasting*, *22*, 137−151.

Banerjee, A., Masten, I., & Marcellino, M. (2005). Leading indicators for euro-area inflation and GDP growth. *Oxford Bulletin of Economics and Statistics*, *67*, 785−813.

Barro, R. (1990). The stock market and investment. *Review of Financial Studies*, *3*, 115−131.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operations Research Quarterly*, *20*, 451−468.

Bernanke, B. S., & Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review*, *82*(4), 901−921.

Boulier, B. L., & Stekler, H. O. (2000). The term spread as a monthly cyclical indicator: an evaluation. *Economics Letters*, *66*, 79−83.

Clark, T., & McCracken, M. (2005). Evaluating direct multistep forecasts. *Econometric Reviews*, *24*(4), 369−404.

Davis, E. P., & Fagan, G. (1997). Are financial spreads useful indicators of future inflation and output growth in EU countries? *Journal of Applied Econometrics*, *12*, 701−714.

Duarte, A., Venetis, I. A., & Paya, I. (2005). Predicting real growth and the probability of recession in the Euro area using the yield spread. *International Journal of Forecasting*, *21*, 261−277.

Estrella, A., & Hardouvelis, G. (1991). The term structure as a predictor of real economic activity. *Journal of Finance*, *46*, 555−576.

Estrella, A., & Mishkin, F. (1997). The predictive power of the term structure of interest rates in Europe and the United States: Implications for the European Central Bank. *European Economic Review*, *41*, 1375−1401.

Estrella, A., & Mishkin, F. (1998). Predicting U.S. recessions: Financial variables as leading indicators. *Review of Economics and Statistics*, *80*, 45−61.

Fama, E. (1990). Stock returns, expected returns and real activity. *Journal of Finance*, *45*, 1089−1108.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the euro area? *Journal of Monetary Economics*, *50*, 1243−1255.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*, 197−204.

Harvey, C. (1988). The real term structure and consumption growth. *Journal of Financial Economics*, *22*, 305−333.

Hassapis, C. (2003). Financial variables and real activity in Canada. *Canadian Journal of Economics*, *36*, 421−442.

Hassapis, C., & Kalyvitis, S. (2002). Investigating the links between growth and stock price changes with empirical evidence from the G7 economies. *Quarterly Review of Economics and Finance*, *42*, 543−575.

Huang, H., & Lee, T. H. (2007). *To combine forecasts or to combine information?* University of California, Riverside Working Paper.

Inoue, A., & Kilian, L. (2004). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, *23*, 371−402.

Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics*, *14*(5), 491−510.

Lee, B-Soo (1992). Casual relationships among stock returns, interest rates, real activity, and inflation. *Journal of Finance*, *4*, 1591−1603.

Lutkepohl, H. (1987). *Forecasting aggregated vector ARMA processes.* Berlin Springer.

Marcellino, M., Stock, J., & Watson, M. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review*, *47*, 1−18.

McCracken, M. W. (2004). *Asymptotics for out-of-sample tests of Granger causality.* University of Missouri-Columbia Manuscript.

Moneta, F. (2005). Does the yield spread predict recessions in the euro area? *International Finance*, *8*(2), 263−301.

Panopoulou, E., Pittis, N., & Kalyvitis, S. (2006). Looking far in the past: re-visiting the growth-returns nexus with non-parametric tests. Downloadable from http://www.aueb.gr/users/kalyvitis/.

Rapach, D. E., & Wohar, M. E. (2004). Financial variables and the simulated out-of-sample forecastability of US output growth since 1985: An encompassing approach. *Economic Inquiry*, *42* (4), 717−738.

Rapach, D. E., & Wohar, M. E. (2006). In-sample vs out-of-sample tests of stock return predictability in the context of data mining. *Journal of Empirical Finance*, *13*, 231−247.

Sensier, M., Artis, M., Osborn, D. R., & Birchenhall, C. (2004). Domestic and international influences on business cycle regimes in Europe. *International Journal of Forecasting*, *20*, 343−357.

Stock, J. H., & Watson, M. W. (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature*, *41*, 788−829.

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

Ekaterini Panopoulou is currently a lecturer at the Department of Statistics and Insurance Science, University of Piraeus, Greece, and a Research Associate at the Institute for International Studies (IIIS), at Trinity College Dublin. Her current research interests cover a range of topics relevant to both theoretical and applied time series econometrics. Panopoulou has recent publications in the *Journal of International Money and Finance, Econometrics Journal, Journal of Applied Econometrics, Journal of Emerging Markets Finance* and *Mathematical and Computer Modelling*.