

Non-linearity and instability in the Euro area ^{*}

Massimiliano Marcellino ^{**}

IEP-Università Bocconi, IGER and CEPR

March 2005

Abstract

In this chapter we evaluate the relative performance of linear, non-linear and time-varying models for about 500 macroeconomic variables for the countries in the Euro area, using a pseudo real-time forecasting methodology. It turns out that linear models work well for about 35% of the series under analysis, time-varying models for another 35% and non-linear models for the remaining 30% of the series. The gains in forecasting accuracy from the choice of the best model can be substantial, in particular for longer forecast horizons. These results emerge from a detailed disaggregated analysis, while they are hidden when an average loss function is used. In the latter case AR models are ranked first, which suggests that they provide a robust tool for automatic forecasting a large number and variety of series. To explore in more detail the issue of parameter instability, we then apply a battery of tests, detecting non-constancy in about 20-30% of the time series. For these variables the forecasting performance of the time-varying and non-linear models further improves, with larger gains for a larger fraction of the series. Finally, we evaluate whether non-linear models perform better for three key macroeconomic variables: industrial production, inflation and unemployment. It turns out that this is often the case. Hence, overall, our results indicate that there is a substantial amount of non-linearity and instability in the Euro area, and suggest that it can be worth going beyond linear models for several macroeconomic variables.

J.E.L. Classification: C2, C53, E30

Keywords: Non-linearity, Instability, Time-Varying models, Non-linear models, European Monetary Union

^{*} I am grateful to Jim Stock and Mark Watson for several conversations on the topics addressed in this paper, and for making available the GAUSS programs that form the basis of the code used in this paper. The Editors of this book and participants at the 2002 NBER/NSF EFWW Meeting and at a seminar at the ECB also provided useful comments. The usual disclaimers apply.

^{**} Correspondence to: IGER, Università Bocconi, Via Salasco 5, 20136, Milan, Italy.

E-mail: massimiliano.marcellino@uni-bocconi.it Phone: +39-02-5836-3327 Fax: +39-02-5836-3302

1. Introduction

The many social, economic and political changes that occurred in the European countries since the early '80s can be expected to make modeling macroeconomic variables with constant parameter linear models particularly difficult. In this context, time-varying and non-linear models should have a comparative advantage, even though their estimation is complicated by the short time span usually available. On the other hand, the use of rolling estimation combined with linear models (see e.g. Swanson (1998)) is not sufficient to address the issue of parameter non constancy in our context, due to the rather short sample available and the multiplicity of the possible structural breaks.

In this chapter we compare a variety of linear and non-linear models on the basis of their real time forecasting performance, using a very large dataset of about 500 macroeconomic variables for the countries in the Euro area. A similar analysis is performed by Stock and Watson (1999, SW henceforth) for the US.¹ Here we have a larger dataset, for several countries, but for a shorter sample. Moreover, we consider more forecasting models, and evaluate the role of instability in explaining the results.

We compare three main forecasting methods. The linear method, which includes autoregressive (AR) models, exponential smoothing and random walk models. The time-varying method, which includes time-varying AR models and smooth transition AR models. The non-linear method, which includes artificial neural network models. Within each method we consider several alternative specifications, for a total of 58 models. We focus on three forecast horizons: one, three and six month a-head. Longer horizons are not worth analyzing because of the rather short sample available.

The competing forecasts are compared on the basis of three measures, in increasing level of disaggregation. First, we compute the average value over all variables of several loss functions, including the common mean absolute and mean square forecast error (mae and msfe, respectively). The results are quite clear cut. The best model is a simple AR either with fixed lag length or with AIC selection, in line with what SW have found for the US. This finding indicates that AR models provide a robust method for automatic forecasting a large number and variety of short European macroeconomic time series. A

similar conclusion has been reached for aggregate Euro area series by Marcellino (2004a) and, in a different context, by Marcellino, Stock and Watson (2003) and Banerjee, Marcellino and Masten (2003). Another interesting finding is that time-varying models, in particular time-varying AR models, perform better than non-linear models, the best ranked neural network model is only 35th.

Second, we rank the models on the basis of the percentage of variables for which they are among the top-N models, for several values of N. The picture is now rather different. The best models for the largest proportion of series belong to the artificial neural network class, they lead to the lowest msfe for 27% of the 480 series when the forecast horizon is $h=1$, and for 32% when $h=6$. The smooth transition models can be ranked second, with 28% of the lowest msfe for $h=3$. The AR models rank third, with values of 23% for $h=1$ and 6, and of 20% for $h=3$. With respect to the findings in SW for the US, compare their tables 3 and 4, it emerges that AR and time-varying models perform much better for Europe. This is likely due both to the shorter sample size, that favours simple models, and to the amount of parameter instability, that favours time-varying models. It is also interesting to mention that SW find quite a good performance for pooling methods, while this is not the case for Europe, see Marcellino (2004b) for details.

Finally, for each forecasting model we compute the empirical distribution function (over variables) of its mean square forecast error, relative to a benchmark AR model, and report selected percentiles of this distribution. In line with the previous results, it turns out that the AR model has a much more stable performance over variables and forecast horizons than time-varying and non-linear models, but the latter can yield substantial gains for some variables and forecast horizons. This indicates that care and expertise is required to select the proper non-linear model for a specific variable, but the exercise is worth undertaking.

To explore whether the results summarized so far are driven by extensive parameter instability, we then apply a battery of stability tests, including statistics for constancy of the parameters versus the alternative of random walk or stationary random coefficients,

¹ Other interesting articles evaluating the relative forecasting performance of linear and non-linear models are Montgomery, Zarnowitz, Tsay and Tiao (1998), Clements, Franses, Smith and Van Dijk (2003) and the

and recursive F-tests for parameter changes at unknown dates. We detect non-constancy in about 20-40% of the time-series, though the percentage decreases to 10-20% when using bootstrapped critical values for the test statistics. The forecasting performance of the time-varying and non-linear models for the unstable variables further improves, with larger gains for a larger fraction of the series.

Finally, we evaluate whether non-linear models perform better for three key macroeconomic variables: industrial production, inflation and unemployment. It turns out that this is often the case, an important result because of the key role for policy making of accurate forecasts of these variables. Joint models for these variables also confirm the relevance of nonlinearities, see e.g. Krolzig, Marcellino and Mizon (2002) and Milas and Rothman (2003).

The structure of the chapter is the following. The dataset is illustrated in Section 2. Section 3 describes the forecasting models and the evaluation criteria. The results of the forecast comparison are presented in Section 4. The issue of instability is analyzed in Section 5. Section 6 presents detailed results for industrial production, unemployment and inflation. Section 7 summarizes and offers some concluding remarks.

2. The data

The dataset we use is taken from Marcellino, Stock and Watson (2003), to whom we refer for additional information. It includes the OECD main economic indicators, monthly, for the period 1982:1-1997:8, for the 11 countries originally in the EMU in the year 2000. The dataset and the sample range are chosen in order to have rather homogenous variables over countries, for a long enough comparable time span. Overall, there are 480 series.

In particular, for each country there are output variables (industrial production and sales, disaggregated by main sectors); labour market variables (employment, unemployment, wages and unit labour costs); prices (consumer and producer, disaggregated by type of goods); monetary aggregates, interest rates (different maturities), stock prices; exchange rates (effective and nominal); imports, exports and net

references therein.

trade; and other miscellaneous series. A complete list of the variables is available upon request.

All the (OECD) series under analysis are seasonally adjusted. This can have implications for the forecasting performance of nonlinear models, see e.g. Ghysels, Granger and Syklos (1996) and Lee and Syklos (1997), an interesting issue for additional research in this field.

3 Forecasting methods

The formulation of a generic forecasting model is

$$y_{t+h}^h = f(Z_t; \theta_h) + \varepsilon_{t+h}, \quad (1)$$

where y_t is the variable being forecast, h indicates the forecast horizon, Z_t is a vector of predictor variables, ε_t is an error term, and θ_h is a vector of parameters, possibly evolving over time. We introduce a distinction between forecasting methods and forecasting models. Forecasting methods differ for the choice of the functional form of the relationship between y_{t+h}^h and Z_t , f . Within each method, different models are determined by the choice of the regressors Z_t and the stationarity transformation applied to y_t .

The h -step forecast is

$$\hat{y}_{t+h}^h = f(Z_t; \hat{\theta}_h), \quad (2)$$

with associated forecast error

$$e_{t+h} = y_{t+h}^h - \hat{y}_{t+h}^h. \quad (3)$$

When y_t is treated as stationary, it is $y_{t+h}^h = y_{t+h}$, while if y_t is I(1) then $y_{t+h}^h = y_{t+h} - y_t$. We present results for both cases. Moreover, we also consider a pre-test forecast where the decision on the stationarity of y_t is based on a unit root test, which often improves the forecasting performance, see e.g. Diebold and Kilian (2000). In particular, we use the Elliott, Rothenberg and Stock (1996) DF-GLS statistics, which performed best in the simulation experiments in Stock (1996). Note that $e_{t+h} = y_{t+h} - \hat{y}_{t+h}$, independently of

whether y_t is treated as stationary or not, so that forecast errors from the three different cases (stationary, I(1) and pre-test) are directly comparable.

Because of the short sample period available, the forecast horizons we consider are 1, 3 and 6 months. When h is larger than one, the " h -step ahead projection" approach in (1), also called dynamic estimation (e.g. Clements and Hendry (1996)), differs from the standard approach of estimating a one-step ahead model, then iterating that model forward to obtain h -step ahead predictions. The h -step ahead projection approach has two main advantages in this context. First, the potential impact of specification error in the one-step ahead model can be reduced by using the same horizon for estimation as for forecasting. Second, we need not resort to simulation methods to obtain forecasts from non-linear models. The resulting forecasts could be slightly less efficient, see e.g. Granger and Terasvirta (1993, Ch.8), but the computational savings in our real time exercise with many series are substantial.

A few forecast errors from the non-linear and time-varying methods are very large. This is due to problems in the estimation of these models, because of multiple local optima and the short estimation sample available (1982:1-1993:12). In order not to bias the comparison against these methods, we automatically trim the forecasts. In particular, when the absolute value of a forecasted change is larger than any previously observed change, a no change forecast is used.

Let us now list the methods and models we compare, and briefly discuss their main characteristics and estimation issues. More details can be found in Stock and Watson (1996, 1999).

Linear methods

Autoregression (AR). Box and Jenkins (1970) popularized the use of these models for forecasting economic variables, and they have performed rather well in forecast comparison exercises, see e.g. Meese and Geweke (1984), or Marcellino, Stock and Watson (2003) for the Euro area. The f function in (1) is linear, and Z_t includes lags of the y variable and a deterministic component. The latter can be either a constant or also a linear trend. The lag length is either fixed at 4, or it is chosen by AIC or BIC with a

maximum of 6 lags. Recalling that the y_t variable can be treated as stationary, I(1), or pre-tested for unit roots, overall we have 18 models in this class.

Exponential smoothing (ES). Exponential smoothing is a rather simple forecasting technique that can be shown to be optimal in the mean square forecast error sense only when the underlying process follows a particular ARMA structure, see e.g. Granger and Newbold (1986, Ch.5). Yet, Makridakis et al. (1982) found this method to perform rather well in practice in a forecast comparison exercise. We consider both single and double exponential smoothing, which are usually adopted for, respectively, stationary and trending series. Estimation of the parameters is conducted by means of (recursive) non-linear least squares (see e.g. Tiao and Xu (1993)). The third model in this class is given by a combination of the single and double models, based on the outcome of the unit root test.

No change. This simple forecast is based on a random walk model, so that it is $\hat{y}_{t+h} = y_t$. Notwithstanding its simplicity, in a few cases it was found to outperform even forecasts from large-scale structural models, see e.g. Artis and Marcellino (2001).

Time-varying methods

Time-varying autoregression (ARTV). In this case the parameters of the AR models evolve according to the following multivariate random walk model (see e.g. Nyblom (1989)):

$$\theta_{ht} = \theta_{h,t-1} + u_{ht}, \quad u_{ht} \sim iid(0, \lambda^2 \sigma^2 Q), \quad (4)$$

where σ^2 is the variance of the error term ε in (1), $Q = (E(Z_t Z_t'))^{-1}$, and we inspect several values of λ : 0 (no evolution), 0.0025, 0.005, 0.0075, 0.01, 0.015, or 0.020. We consider first a specification with a constant, 3 lags and $\lambda = 0.005$, and then we allow for selection of the number of lags (1,3,6) jointly with the value of λ by either AIC or BIC. In each case, y_t can be either stationary, or I(1) or pre-tested, so that we have a total of 9 ARTV models. The models are estimated by the Kalman filter.

Logistic smooth transition autoregression (LSTAR). The generic model can be written as

$$y_{t+h}^h = \alpha' \zeta_t + d_t \beta' \zeta_t + \varepsilon_{t+h}, \quad (5)$$

where $d_t = 1/(1 + \exp(\gamma_0 + \gamma_1 v_t))$, and $\zeta_t = (1, y_t, y_{t-1}, \dots, y_{t-p+1})$ if y_t is treated as stationary or $\zeta_t = (1, \Delta y_t, \Delta y_{t-1}, \dots, \Delta y_{t-p+1})$ if y_t is I(1). The smoothing parameter γ_1 regulates the shape of parameter change over time. When $\gamma_1 = 0$ the model becomes linear, while for large values of γ_1 the model tends to a self-exciting threshold model, see e.g. Granger and Terasvirta (1993), Terasvirta (1998) for details. For models specified in levels we consider the following choices for the threshold variable in d_t : $v_t = y_t$, $v_t = y_{t-2}$, $v_t = y_{t-5}$, $v_t = y_t - y_{t-6}$, $v_t = y_t - y_{t-12}$. For differenced variables, it can be $v_t = \Delta y_t$, $v_t = \Delta y_{t-2}$, $v_t = \Delta y_{t-5}$, $v_t = y_t - y_{t-6}$, $v_t = y_t - y_{t-12}$. In each case the lag length of the model was either 1 or 3 or 6. We report results for the following models: 3 lags and $v_t = y_t$ (or $v_t = \Delta y_t$ for the I(1) case); 3 lags and $v_t = y_t - y_{t-6}$; AIC or BIC selection of both the number of lags and the specification of v_t . In each case, y_t can be either stationary, or I(1) or pre-tested, so that overall there are 12 LSTAR models. Estimation is carried out by (recursive) non-linear least squares, using an optimizer developed by Stock and Watson (1999).

Non-linear methods

Artificial neural network (ANN). Artificial neural networks can provide a valid approximation to the generating mechanism of a vast class of non-linear processes, see e.g. Hornik, Stinchcombe and White (1989), and Swanson and White (1997) for their use as forecasting devices. The so called single layer feedforward neural network model with n_1 hidden units (and a linear component) is specified as:

$$y_{t+h}^h = \beta_0' \zeta_t + \sum_{i=1}^{n_1} \gamma_{1i} g(\beta_{1i}' \zeta_t) + \varepsilon_{t+h}, \quad (6)$$

where $g(x)$ is the logistic function, $g(x) = 1/(1 + e^{-x})$. Note that when $n_1=1$ the model can be interpreted as a logistic smooth transition autoregression, with the parameter evolution being determined by the linear combination of variables $\beta_{11}' \zeta_t$. A more complex model is the double layer feedforward neural network with n_1 and n_2 hidden units:

$$y_{t+h}^h = \beta_0' \zeta_t + \sum_{j=1}^{n_2} \gamma_{2j} g\left(\sum_{i=1}^{n_1} \beta_{2ji} g(\beta_{1i}' \zeta_t)\right) + \varepsilon_{t+h}. \quad (7)$$

We report results for the following specifications: $n_1=2, n_2=0, p=3$ (recall that p is number of lags in ζ_t); $n_1=2, n_2=1, p=3$; $n_1=2, n_2=2, p=3$; AIC or BIC selection with $n_1=(1,2,3), n_2=(1,2 \text{ with } n_1=2), p=(1,3)$. For each case y_t can be either stationary, or I(1) or pre-tested, which yields a total of 15 ANN models. The models are estimated by (recursive) non-linear least squares, using an algorithm developed by Stock and Watson (1999).

Overall, there are 58 models in the forecast comparison exercise, 22 belong to the linear class, 21 are time-varying, and 15 are non-linear. They are summarized in Table 1. To mimic real time situations, for each variable, method and model the unit-root tests, estimation and model selection are repeated each month over the forecasting period, which is 1994:1-1997:8.

4. Forecast Evaluation

We now have to evaluate the relative forecasting performance of the $M=58$ models for the $N=480$ variables in the dataset. The starting point is the choice of a loss function.

For variable n and forecasting method m , we define the loss function as

$$Loss_{n,m}^h = \frac{1}{T-h} \sum_{t=1}^{T-h} |e_{t+h,n,m}|^\rho, \quad (8)$$

where e_{t+h} is the h -step ahead forecast error, and ρ can be equal to 1, 1.5, 2, 2.5 or 3. The values $\rho = 1$ and $\rho = 2$ correspond to the familiar choices of, respectively, the mean absolute and the mean square forecast error as the loss function.

In order to compare the loss over the whole set of variables, we adopt the following loss function for method m :

$$Loss_m^h = \frac{1}{N} \sum_{n=1}^N \frac{Loss_{n,m}^h}{Loss_{n,1}^h}, \quad (9)$$

namely, a weighted average of the loss for each variable, with weights given by the inverse of the loss of a benchmark forecast, which makes the magnitude of the losses

comparable across variables. As a benchmark, we adopt throughout an AR model with 4 lags and a constant, specified in levels.²

In Table 2 we report the ranking of the models, for different values of ρ . The results are quite clear cut, and three main comments are in order. First, the best model for any value of ρ is an AR, with 4 lags when $h=1$ or 3 (our benchmark), and with AIC lag selection for $h=6$. The AR(4) ranks second for $h=6$, and the AR(AIC) ranks second for $h=1,3$. In both cases, it is better not to impose the presence of a unit root and not to include a linear trend in the model. AR models with unit roots imposed or pre-tested for appear in the ranking at, respectively, the third and fourth place for some values of ρ and h .

Second, time-varying models perform better than non-linear models, and the best time-varying models are the TVARs. In particular, when $\rho = 2$, the TVAR with 3 lags, constant and no unit root imposed ranks 4th for $h=6$, 10th for $h=3$ and 17th for $h=1$. Several other TVAR models are ranked 15th or lower, while the only LSTAR model in the top-15 has 3 lags, no unit root, and the transition variable is $\zeta_t = y_t - y_{t-6}$.

Third, the performance of the neural network models is very poor, the best ranked model is only 35th. To make sure that this result is not driven by estimation problems due to the small sample size, we have repeated the estimation and forecasting exercises several times with different starting values and a very large number of iterations in the optimization routines, but the figures did not change. Moreover, Stock and Watson (1999, SW) got similar results with longer series for the US.³

The ranking in Table 2 is based on the loss function in (9), which is an average over all the variables, so that the TV and non-linear models could still be valuable for some

² Notice that testing for the significance of the loss function (9) across models is complicated by the fact that we are averaging over a large number of time series. In principal, the bootstrap could be used to provide standard errors around the estimated loss functions, see e.g. Marcellino, Stock and Watson (2004) for an application concerning the evaluation of dynamic estimation for forecasting with linear models. Yet, in our context with an even larger data set and many nonlinear models, the exercise becomes computationally cumbersome.

³ Note that since we are comparing the average loss over about 500 variables, it is virtually impossible to provide standard errors for the measure in equation (9), and hence tests for whether the average loss between two models is statistically different from zero. Yet, the ranking in Table 2 definitely favors linear models.

series. To evaluate whether this is the case, we adopt the loss function in (8) with $\rho = 2$, i.e. the mean square forecast error (msfe), and we find the best forecasting model for each variable. In Table 3a we report the fraction of series for which a given method is the best (results for each model are available upon request).

The picture is now rather different. The best methods for the largest proportion of series are ANN, which leads to the lowest msfe for 27% of the 480 series when $h=1$ and for 32% when $h=6$, and LSTAR, with 28% of the lowest msfe for $h=3$. The AR method ranks third, with values of 23% for $h=1$ and 6, and of 20% for $h=3$. Even though Stock and Watson (1999) do not explicitly mention it, a similar pattern emerges also for the US, see their Tables 3 and 4.

The second panel of Table 3a shows that if we aggregate the models into linear, time-varying and non-linear methods, there is a substantial equivalence of the three approaches in terms of the fraction of best forecasting models, with only slightly lower values for the non-linear approach.

In Table 3b we provide additional results for the models that perform best for the highest fraction of series in each class. In particular, we compute the fraction of series for which these models are the best or among the top 5, 10, 15 and 20. The resulting figures are of comparable magnitude across models, with slightly lower values for the no-change forecast.

As mentioned in the Introduction, these results differ from what SW have found for the US because of the stronger importance of AR and time-varying models, likely due to the shorter and more unstable sample.

To explain the mismatching ranking of Tables 2 and 3 we have to take a more disaggregate approach. First, for each variable we compute the relative msfe (rmsfe) of each forecasting model with respect to the benchmark AR(4), so that an rmsfe higher than one indicates that the method under analysis is worse than the benchmark. In formulae, the rmsfe of model j for variable m is:

$$rmsfe^h_{j-AR4,m} = \left(\sum_{t=1}^{T-h} e_{j,t+h,m}^2 \right) / \left(\sum_{t=1}^{T-h} e_{AR4,t+h,m}^2 \right). \quad (10)$$

Then, for each model, we calculate the empirical distribution of the rmsfe over the variables. In Table 4 we report the mean of the distribution and some percentiles for

selected models (the best in Table 2 and those in Table 3b, results for all models are available upon request).

The ranking in Table 2 for $\rho = 2$ is based on the mean of this distribution, while that in Table 3 in practice is based on the lower percentiles. Hence, we expect the ARFC0a (and the benchmark) to have a lower average rmsfe and a more concentrated distribution, while the ANN models, and more generally the models in Table 3b, should present higher dispersion and good performance in the lower tail of the distribution.

The figures in Table 4 confirm our expectations. In particular, values of the rmsfe in the lower 10% tail of the distribution are substantially smaller than one for ANN, but also for LSTAR models, while some values in the upper 10% tail are very large. This could suggest using the median rather than the mean of the distribution to construct the ranking in Table 2. Yet, from the 6th column of Table 4 the median is also in general higher for non-linear than for linear models. It is important to take into consideration the whole distribution.

It is also worth noting that the AR in levels with a constant and AIC lag length selection (ARFC0a) has a highly concentrated distribution over variables, most rmsfe are in the range 0.85 to 1.19 for $h=1$. The corresponding values for the ANN model ANF0b are 0.66 to 2.15, and 0.68 to 2.04 for the LSTAR model LS0063 (these are the best models in their class from Table 3b). Moreover, for these models the dispersion of the distribution increases much more rapidly with the forecast horizon, h , than for the AR model.

These characteristics imply that the AR model has a much more stable performance over variables and forecast horizons than time-varying and non-linear models, but the latter can yield substantial gains for some variables and forecast horizons. In Section 6 we will evaluate whether this is the case for some key macroeconomic variables, while in the next Section we analyze in more detail the issue of instability.

5. Measuring the extent of instability

Time-varying models work better than linear specifications for about one third of the series under analysis. Moreover, the success of the non-linear models can be also partly

due to instability. In this section we try to measure the extent of instability in the Monetary Union by applying a battery of parameter constancy tests to all the 480 macroeconomic variables. We then repeat the forecasting exercise for the subset of unstable series, and verify whether the time-varying and non-linear models do perform better in this case.

5.1 Instability Tests

Following Stock and Watson (1996), who present a detailed analysis of instability for the US, we consider three different types of statistics.

First, tests for constant versus randomly time-varying coefficients. This set includes Nyblom's (1989, NY) locally most powerful test against the alternative of random walk coefficients ($\lambda=0$ versus $\lambda >0$ in equation (4)), and a Breusch and Pagan (1979, BP) Lagrange multiplier test against the alternative of iid random coefficients with constant mean and variance.

Second, tests based on functions of the cumulative sum of OLS residuals from equation (1), see Ploberger and Kramer (1992). We consider the supremum of the cumulative sum (KP1), and its mean square (KP2).

Third, F-tests for constancy of the parameters against the alternative of a single break at an unknown date. The tests are computed recursively for a range of dates, say $[t_0, \dots, t_1]$, where t_0 and t_1 are selected in order to discard the first and last 15% of the sample. Three functions of the resulting sequence of statistics are considered: the supremum (Quandt (1960, QLR)); the mean (Hansen (1992), Andrews and Ploberger (1994), MLR); and the so called average exponential (Andrews and Ploberger (1994), ALR).

Stationarity transformations, i.e. logarithms and differencing, are applied to all series when needed (a detailed list is available upon request), and all series are represented as an AR process in levels, with 3 lags and a constant. In Table 5a we report results for the stability tests, using different significance levels and asymptotic critical values from the papers listed above. More precisely, we report the percentage of series for which the null hypothesis of stability is rejected.

Using a 10% significance level, the figures are in the range 20-40%, with lower values from the KP1 and KP2 tests, and higher values from the recursive F-tests. There is no clear-cut pattern of rejection of stability among series and countries (detailed results for each variable are available upon request, see also the Data Appendix and Section 6), even though price series and monetary aggregates appear often to be unstable.

In Table 5b we investigate whether the detected amount of instability is robust to three extensions. First, instead of using a fixed lag length for all the series, we allow for BIC selection for each variable. Second, because of the short sample, the finite sample distribution of the tests could be different from its asymptotic counterpart. To address this issue, we have generated finite sample critical values by bootstrapping, following a procedure suggested by Stock and Watson (1996). Third, the performance of the Nyblom's statistic and of the F-tests could be affected by the presence of heteroskedasticity in the errors. Hence, we consider a robust version of these tests, where the relevant covariance matrices are estimated by the White's (1980) method.

From Table 5b, BIC selection does not affect the outcome of the tests. Instead, the use of finite sample critical values substantially decreases the fraction of unstable series according to Nyblom's statistic and the F-tests, values are now in the range 10-25%. The KP1, KP2 and BP tests are robust to the change of critical values, and still reject stability for about 20% of the series. The robust estimation of the covariance matrix further decreases the number of unstable series according to the F-tests, to values around 10%.

Overall, we can conclude that tests for parameter stability reject this hypothesis for about 20-30% of the series.

5.2 Forecast evaluation for unstable series

We now evaluate whether the time-varying and non-linear models under analysis forecast better than linear specifications for the unstable series. In particular, we select those series for which the Nyblom's test rejects at the 10% level using finite sample critical values, so that TVAR models should be particularly favored, see the Data Appendix for a list of these 134 variables.

Table 6 reports the ranking of the top-10 models, according to the loss function in equation (9). An AR in levels, with a constant, and either a fixed or AIC selected number

of lags is still the best forecasting model on average. The main difference with respect to Table 2 is that now the TVAR3 ranks second when $h=1$ and $\rho = 2$. More generally, TV models appear more frequently in the top-10, actually they are the only models in the top-10 not in the linear class, but still most of the top-10 models are linear. The ranking of STAR or ANN models does not improve significantly. This implies that, on average, linear models forecast comparatively well also in the presence of changing parameters.

Yet, if we increase the level of disaggregation by focusing on the percentage of series for which a model forecasts best in msfe terms, the outcome is rather different. From Table 7a, the improved performance of time-varying models for unstable series is evident, in particular for $h=3$ and 6. When $h=3$ the TV models forecast best for 51% of the unstable variables, versus 38% in Table 3a, and for 37% when $h=6$, versus 29% in Table 3a. The performance of the linear models deteriorates at all forecast horizons, in particular for $h=1,3$, while that of the non-linear models improves slightly for $h=1$ and 6 but deteriorates for $h=3$.

Within the TV class, the performance of the TV-AR improves for all forecast horizons, that for the STAR models for $h=3,6$. This pattern is also reflected in Table 7b, the fraction of series for which the TV models are among the top-N, with $N=5, 10, 15, 20$, increases for all values of N with respect to Table 3b.

We can further increase the level of disaggregation, by considering the performance of each model for each variable. In table 8 we report, for the best model in Table 6 and those in Table 7b, the mean of the empirical distribution of the msfe relative to the benchmark ARFC04, and some percentiles. If we focus on the mean, the ranking in Table 6 (for $\rho = 2$) is obviously confirmed. But if we look at the percentiles, it is evident that the time-varying and non-linear models in Table 7b can perform substantially better than the benchmark and the ARFC0a for a consistent fraction of series. Moreover, the whole empirical distribution for the TV and non-linear models is in general shifted to the left with respect to the case where all the variables are analyzed, see Table 4, which is consistent with the fact that these models should perform better for unstable variables.

In summary, time-varying models forecast better for a larger fraction of unstable series, even though on average linear models are still the best.

6. Forecasting industrial production, unemployment and inflation

In this section we focus on three key macroeconomic variables, namely, industrial production (IP), unemployment (UNEM) and inflation (INFL), for all the 11 countries originally in the EMU. More precisely, we focus on IP growth, the change in the unemployment rate, and the growth in consumer prices.

In Table 9 we report the outcome of the stability tests for these variables. If we label unstable a variable when at least one test rejects, then inflation is the most unstable with 10 rejections out of 11, followed by unemployment with 5 out of 10 (unemployment is not available for Portugal over the whole sample), and IP ranks third with 5 rejections out of 11. Overall, there appears to be enough instability for the time-varying and non-linear models to be potentially useful.

In Table 10 we then compare the ranking of the competing models using the average loss function in equation (9) for several values of ρ , only the top-2 models are reported for each variable to save space. In the case of IP growth, the best models are linear, but for inflation an ANN model ranks second for $h=1$ and a TV model for $h=3$, while for unemployment time-varying models rank first (in particular, a TVAR for $h=1$ and STAR models for $h=3,6$). Hence, for this variable even the average msfe can be reduced by forecasting with a model outside the linear class.

Table 11 presents the ranking of the methods based on the fraction of series for which they msfe dominate. In this case time-varying methods perform rather well also for IP growth, in particular models in the STAR class and when $h=1,3$. The ANN models rank first in about 50% of the cases for unemployment, and they achieve good results also for inflation. Overall, linear models work well only in about 25% of the cases.

7. Conclusions

In this chapter we have provided an accurate analysis of the relative forecasting performance of linear, time-varying and non-linear models, using a very large dataset of 480 macroeconomic variables for the countries in the European Monetary Union. The

main finding is that models outside the linear class are useful for a substantial fraction of the series, though on average over all variables AR models perform best.

Three final comments are in order to interpret this result. First, as shown for example in Clements and Hendry (1998, 1999), the best forecasting model does not necessarily provide the closest approximation to the generating mechanism of the process. Yet, the fact that time-varying and non-linear models work well for a substantial fraction of the series, combined with the outcome of the instability tests and with the many social, economic and institutional changes that affected European countries over the past 20 years, provide enough evidence against the untested use of linear models.

Second, unfortunately, the sample available is rather short, so that estimation and forecast evaluation have to be conducted on a limited number of observations. We have compensated for this effect by increasing substantially the number of series under analysis. Moreover, a longer sample would have likely improved the estimation of non-linear and time-varying models, so that the results could point even more in their favor.

Third, other non-linear or time-varying models could perform even better, and computation of the forecasts using simulation methods could further increase the forecast gains.

Thus, this chapter would like to encourage applied researchers to go beyond linearity, and economists to investigate in more detail the economic rationale underlying the good performance of non-linear models.

References

- Andrews, D.W.K. and Ploberger, W. (1994), "Optimal tests when a nuisance parameter is present only under the alternative", *Econometrica*, 62, 1383-1414.
- Artis, M. and Marcellino, M. (2001), "Fiscal forecasting: the track record of IMF, OECD and EC", *Econometrics Journal*, 4, s20-s36.
- Banerjee, A., Marcellino, M. and Masten, I. (2003), "Leading indicators for Euro area inflation and GDP growth", CEPR WP 3893.
- Box, G.E.P. and Jenkins, G.M. (1970), *Time series analysis, forecasting and control*, San Francisco: Holden Day.
- Breusch, T.S. and Pagan, A.R. (1979), "A simple test for heteroschedasticity and random coefficient variation", *Econometrica*, 47, 1287-1294.

- Clements, M.P. and Hendry, D.F. (1996), "Multi-step estimation for forecasting", *Oxford Bulletin of Economics and Statistics*, 58, 657-684.
- Clements, M.P. and Hendry, D.F. (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.
- Clements, M.P. and Hendry, D.F. (1999), *Forecasting Non-Stationary Economic Time Series*, Cambridge (MA): MIT Press.
- Clements, M.P., P.H. Franses, J. Smith, and D. van Dijk (2003), "On SETAR Non-Linearity and Forecasting", *Journal of Forecasting*, 22, 359-375.
- Diebold, F.X. and Kilian, L. (2000), "Unit Root Tests are Useful for Selecting Forecasting Models," *Journal of Business and Economic Statistics*, 18, 265-273.
- Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996), "Efficient tests for an autoregressive unit root", *Econometrica*, 64, 813-36.
- Ghysels, E., Granger, C.W.J. and Siklos P.L. (1996), "Is seasonal adjustment a linear or nonlinear data-filtering process?", *Journal of Business and Economic Statistics* 14, 374-386.
- Granger, C.W.J. and Newbold, P. (1986), *Forecasting economic time series*, San Diego: Academic Press.
- Granger, C.W.J. and Terasvirta, T. (1993), *Modelling nonlinear economic relationships*, Oxford: Oxford University Press.
- Hansen, B. (1992), "Tests for parameter instability in regressions with I(1) processes," *Journal of Business and Economic Statistics*, 10, 321-336.
- Hornik, K., Stinchcombe, M. and White, H. (1989), "Multilayer feedforward networks are universal approximators", *Neural Networks*, 2, 359-66.
- Krolzig, H.-M., Marcellino, M., and Mizon G.E. (2002), "Markov-switching vector equilibrium correction model of the UK labour market", *Empirical Economics* 27, 233- 254.
- Lee, H.S. and Siklos, P. (1997), "The role of seasonality in economic time series-reinterpreting money-output causality in US data", *International Journal of Forecasting* 13, 381-391.
- Makridakis, S. Anderson, A., Carbonne, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E., Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition", *Journal of Forecasting*, 1, 111-153.
- Marcellino, M. (2004a) "Forecasting EMU macroeconomic variables", *International Journal of Forecasting*, 20, 359-72.
- Marcellino, M. (2004b) "Forecast pooling for short time series of macroeconomic variables", *Oxford Bulletin of Economics and Statistics*, 66, 91-112.

- Marcellino, M., Stock, J.H. and Watson, M.W. (2003), "Macroeconomic forecasting in the Euro area: country specific versus Euro wide information", *European Economic Review*, 47, 1-18.
- Marcellino, M., Stock, J.H. and Watson, M.W. (2004), "A Comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series", mimeo.
- Meese, R. and Geweke, J. (1984), "A comparison of autoregressive univariate forecasting procedures for macroeconomic time series", *Journal of Business and Economic Statistics*, 2, 191-200.
- Milas, C. and Rothman P. (2003), "Multivariate STAR unemployment rate forecasts", East Carolina University Discussion Paper in Economics 03/07. <http://www.ecu.edu/econ/wp/03/ecu0307.pdf>
- Montgomery, A.L., V. Zarnowitz, R.S. Tsay, and G.C. Tiao (1998), "Forecasting the U.S. Unemployment Rate", *Journal of the American Statistical Association*, 93, 478–493.
- Nyblom, J. (1989), "Testing for constancy of parameters over time", *Journal of the American Statistical Association*, 84, 223-230.
- Ploberger, W. and Kramer, W. (1992), "The CUSUM test with OLS residuals", *Econometrica*, 60, 271-286.
- Quandt, R.E. (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes", *Journal of the American Statistical Association*, 55, 324-330.
- Swanson, N.R. and White, H. (1997), "A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks", *Review of Economics and Statistics*, 79, 540-550.
- Terasvirta, T. (1998), "Modelling economic relationships with smooth transition regressions" in Ullah, A. and Giles, D.E.A. (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, 507-552.
- Tiao, G.C. and Xu, D. (1993), "Robustness of maximum likelihood estimates for multi-step predictions: the exponential smoothing case", *Biometrika*, 80, 623-641.
- Stock, J.H. (1996), "VAR, error correction and pretest forecasts at long horizons", *Oxford Bulletin of Economics and Statistics*, 58, 685-701.
- Stock, J.H. and Watson, M.W. (1996), "Evidence on structural instability in macroeconomic time series relations", *Journal of Business and Economic Statistics*, 14, 11-30.
- Stock, J.H. and Watson, M.W. (1999), "A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series", in Engle, R. and White, R. (eds), *Cointegration, causality, and forecasting: A festschrift in honor of Clive W.J. Granger*, Oxford: Oxford University Press, 1-44.
- Swanson, N.R. (1998), "Money and Output Viewed Through a Rolling Window", *Journal of Monetary Economics* 41, 455–473.
- White, H. (1980), "A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48, 817-830.

Table 1 – Forecasting models

Linear methods

ARF(X,Y,Z)	<i>Autoregressive models</i> (18 models) X = C (const.) or T (trend) Y = 0 (stationary), 1 (I(1)), P (pre-test) Z = 4 (4 lags), a (AIC), b (BIC)
EX(X)	<i>Exponential smoothing</i> (3 models) X = 1 (single), 2 (double), P (pre-test)
NOCHANGE	<i>No change forecast</i> (1 model)

Time-varying methods

ARTVF(X,Y,Z)	<i>Time-varying AR models</i> (9 models) X = C (const.) Y = 0 (stationary), 1 (I(1)), P (pre-test) Z = 3 (3 lags), a (AIC), b (BIC)
LS(X,Y,Z)	<i>Logistic smooth transition</i> (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) Y = transition variable, 10 ($\zeta_t = y_t$), 06 ($\zeta_t = y_t - y_{t-6}$) Z = 3 (p, lag length)
LSF(X,W)	<i>Logistic smooth transition</i> (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) W = a (AIC on transition variable and p), b (BIC)

Non-linear methods

AN(X,Y,Z,W)	<i>Artificial neural network models</i> (9 models) X = 0 (stationary), 1 (I(1)), P (pre-test) Y = 2 (n_1) Z = 0, 1, 2 (n_2) W = 3 (p, lag length)
ANF(X,S)	<i>Artificial neural network models</i> (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) S = a (AIC on n_1, n_2, p), b (BIC)

Table 2 - Ranking of competing models with different loss functions

Rank	Horizon	$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=2.5$	$\rho=3$
1	h=1	ARFC04	ARFC04	ARFC04	ARFC04	ARFC04
	h=3	ARFC04	ARFC04	ARFC04	ARFC04	ARFC04
	h=6	ARFC0a	ARFC0a	ARFC0a	ARFC0a	ARFC0a
2	h=1	ARFC0a	ARFC0a	ARFC0a	ARFC0a	ARFC0a
	h=3	ARFC0a	ARFC0a	ARFC0a	ARFC0a	ARFC0a
	h=6	ARFC04	ARFC04	ARFC04	ARFC04	ARFC04
3	h=1	ARFT1a	ARFT1a	ARFT1a	ARFC1a	ARFC1a
	h=3	ARFC0b	ARFC0b	ARFC0b	ARFC0b	ARFC0b
	h=6	ARFC0b	ARFC0b	ARFC0b	ARFC0b	ARFC0b
4	h=1	ARFC0b	ARFC0b	ARFC1a	ARFTPa	ARFTPa
	h=3	ARFC1a	ARFC1a	ARFC1a	ARFC1a	ARFC1a
	h=6	ARFC1a	ARFC1a	ARTVFC03	ARTVFC03	ARTVFC03
5	h=1	ARFT1b	ARFC1a	ARFTPa	ARFT1a	ARFCPa
	h=3	ARFTPa	ARFTPa	ARFTPa	ARFTPa	ARFTPa
	h=6	ARFCPa	ARFCPa	ARFC1a	ARFT1a	ARTVFC0b
6	h=1	ARFC1a	ARFTPa	ARFC0b	ARFC0b	ARFC0b
	h=3	ARFCPa	ARFCPa	ARFCPa	ARFCPa	ARFCPa
	h=6	ARFTPa	ARFTPa	ARFT1a	ARFC1a	ARFT1a
7	h=1	ARFT14	ARFT14	ARFCPa	ARFCPa	ARFT1a
	h=3	ARFC1b	ARFC1b	ARTVFC1b	ARTVFC03	ARTVFC03
	h=6	ARFT1a	ARFT1a	ARFCPa	ARFCPa	ARTVFC0a
8	h=1	ARFTPa	ARFT1b	ARFT14	ARTVFC1b	ARTVFC1b
	h=3	ARFTPb	ARTVFC1b	ARTVFCPb	ARTVFC1b	ARTVFC1b
	h=6	ARFC1b	ARTVFC03	ARFTPa	ARFTPa	ARFC1a
9	h=1	ARTVFC03	ARFCPa	ARTVFC1b	ARTVFC1a	ARTVFC1a
	h=3	ARFT1a	ARFT1a	ARFC1b	ARTVFCPb	ARTVFCPb
	h=6	ARTVFC1a	ARFT14	ARFT14	ARTVFC0b	ARFCPa
10	h=1	ARFCPa	ARTVFC1b	ARTVFC1a	ARFT14	ARTVFCP3
	h=3	ARTVFC1b	ARFTPb	ARTVFC03	ARTVFC1a	ARTVFCPa
	h=6	ARFCPb	ARFC1b	ARFT1b	ARFT1b	ARFT1b

Notes:

See Table1 for definition of models

The loss function is $Loss_m^h = \frac{1}{N} \sum_{n=1}^N \frac{Loss_{n,m}^h}{Loss_{n,1}^h}$, $Loss_{n,m}^h = \frac{1}{T-h} \sum_{t=1}^{T-h} |e_{t+h,n,m}|^\rho$, where the

benchmark model is ARFC04 and e_{t+h} is the h-step ahead forecast error

Table 3a – Fraction of series for which a forecasting method has lowest msfe

Method	AR	ES	NoChange	ARTV	LSTAR	ANN
h=1	0.23	0.11	0.02	0.13	0.25	0.27
h=3	0.20	0.09	0.02	0.10	0.28	0.27
h=6	0.23	0.11	0.03	0.06	0.23	0.32

Method	Linear	Time-varying	Non-linear
h=1	0.36	0.38	0.27
h=3	0.31	0.38	0.27
h=6	0.37	0.29	0.32

Notes:

Linear method includes AR,ES, No Change.

Time-varying method includes ARTV, LSTAR.

Non-linear method includes ANN.

Figures do not sum up to one because of rounding errors.

Table 3b – Fraction of series for which a forecasting model is in the top-N

	N=1	N=5	N=10	N=15	N=20
ARFT0b	0.03	0.12	0.22	0.32	0.39
	0.03	0.14	0.21	0.31	0.37
	0.04	0.14	0.19	0.26	0.32
EX1	0.05	0.16	0.25	0.30	0.36
	0.05	0.15	0.24	0.30	0.35
	0.05	0.16	0.25	0.29	0.36
NOCHANGE	0.02	0.08	0.12	0.16	0.21
	0.02	0.09	0.15	0.19	0.24
	0.03	0.13	0.19	0.23	0.27
ARTVFC03	0.04	0.14	0.24	0.35	0.44
	0.03	0.16	0.28	0.37	0.41
	0.02	0.07	0.21	0.31	0.38
LS0063	0.04	0.12	0.18	0.26	0.34
	0.04	0.13	0.22	0.30	0.37
	0.04	0.11	0.21	0.30	0.37
ANF0b	0.05	0.13	0.20	0.26	0.30
	0.04	0.14	0.22	0.28	0.33
	0.05	0.14	0.22	0.27	0.32

Notes: See Table 1 for definition of models

The figures report the fraction of series for which a model is among the N models with the lowest msfe.

The reported models are the best performers in each class for N=1.

For each model, the three rows report result for, respectively, h=1,3,6.

Table 4 – Mean and percentiles of relative msfe for selected forecasting models

Forecast	Mean	0.02	0.10	0.25	0.50	0.75	0.90	0.98
ARFC04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ARFC0a	1.01	0.85	0.94	0.99	1.00	1.03	1.08	1.19
	1.01	0.77	0.92	0.98	1.00	1.04	1.08	1.21
	1.00	0.66	0.84	0.94	1.00	1.03	1.14	1.38
ARFT0b	1.08	0.67	0.87	0.95	1.02	1.11	1.30	1.93
	1.22	0.45	0.74	0.88	1.04	1.26	1.79	3.44
	1.84	0.19	0.49	0.75	1.08	1.84	2.99	9.79
EX1	1.50	0.65	0.85	0.94	1.05	1.26	2.02	5.00
	1.81	0.43	0.76	0.91	1.06	1.40	3.08	9.75
	3.55	0.16	0.44	0.76	1.07	1.78	5.24	20.32
NOCHANGE	1.67	0.64	0.88	1.02	1.24	1.56	2.24	5.00
	1.90	0.42	0.77	0.95	1.22	1.63	3.04	9.75
	3.65	0.15	0.44	0.81	1.27	2.18	4.56	18.28
ARTVFC03	1.03	0.72	0.90	0.96	1.01	1.07	1.17	1.52
	1.07	0.59	0.81	0.92	1.01	1.12	1.34	2.14
	1.24	0.37	0.62	0.86	1.04	1.29	1.96	3.94
LS0063	1.14	0.68	0.88	0.97	1.05	1.16	1.36	2.04
	1.15	0.55	0.81	0.95	1.05	1.21	1.60	2.52
	1.36	0.27	0.60	0.84	1.03	1.35	2.23	4.93
ANF0b	1.31	0.66	0.88	0.98	1.08	1.24	1.53	2.15
	1.42	0.41	0.78	0.95	1.14	1.47	2.09	4.76
	2.75	0.16	0.48	0.87	1.30	2.15	4.13	12.17

Notes:

The models are the best in Table 2 and those in Table 3b

The benchmark model is ARFC04

For each forecast the three rows correspond to, respectively, h=1, 3, 6

See Table 1 for the definition of the models

Table 5a –Stability Tests

Percentage of series significant at

	NY	KP1	KP2	BP	QLR	MLR	ALR
10% level	27.9	22.5	20.6	31.5	40.8	34.6	41.7
5% level	21.0	16.0	16.9	26.7	35.2	28.5	35.2
1% level	8.3	5.2	9.6	17.3	24.6	17.5	23.3

Notes:

The model is an AR3, except for the second line where the lag length is selected by BIC

NY is Nyblom's (1989) test

KP1 and KP2 are the Ploerberger and Kramer's (1992) supremum and mean square tests

BP is Breusch and Pagan's (1979) Lagrange multiplier test

QLR is Quandt's (1960) supremum F-test

MLR is Andrews and Ploerberger's (1994) mean F-test

ALR is Andrews and Ploerberger's (1994) average exponential F-test

Table 5b –Stability Tests, sensitivity analysis

Percentage of series significant at the 10%

Lag L	Cval	H-R	NY	KP1	KP2	BP	QLR	MLR	ALR
Fixed	Asy.	No	27.9	22.5	20.6	31.5	40.8	34.6	41.7
BIC	Asy.	No	25.4	21.9	20.4	33.5	41.7	33.8	39.6
Fixed	MC	No	10.6	24.8	21.3	20.0	25.6	16.5	23.8
Fixed	MC	Yes	8.8	--	--	--	10.8	10.6	11.3

Notes:

Lag L indicates whether the lag length is fixed at 3 or BIC determined

Cval indicates whether the critical values are asymptotic or for finite sample (MC)

H-R indicates robust estimation of the relevant covariance matrices

See the Notes to Table 5a for a description of the tests

Table 6 - Unstable series , ranking of competing models with different loss functions

Rank		$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=2.5$	$\rho=3$
1	h=1	ARFT1b	ARFC04	ARFC04	ARFC04	ARFC04
	h=3	ARFT1a	ARFC0a	ARFC0a	ARFC0a	ARFC04
	h=6	ARFC0a	ARFC0a	ARFC0a	ARFC0a	ARFC0a
2	h=1	ARFT1a	ARTVFC03	ARTVFC03	ARFC0a	ARFC0a
	h=3	ARFC0a	ARFT1a	ARFC04	ARFC04	ARFC0a
	h=6	ARFC04	ARFC04	ARFC04	ARFC04	ARFC04
3	h=1	ARFC04	ARFT1a	ARFT1a	ARFC0b	ARFC0b
	h=3	ARFT1b	ARFC04	ARFC0b	ARFC0b	ARFC0b
	h=6	ARFC0b	ARFC0b	ARFC0b	ARFC0b	ARFC0b
4	h=1	ARTVFC03	ARFT1b	ARFC0a	ARFT14	ARFT14
	h=3	ARFT14	ARFC0b	ARFT1a	ARFT1a	ARFT1a
	h=6	ARFT1a	ARFT1a	ARFT1a	ARFT1a	ARFT1a
5	h=1	ARFT14	ARFT14	ARFT14	ARFT1a	ARFT1a
	h=3	EXP	ARFT1b	ARFT14	ARTVFC03	ARTVFC03
	h=6	ARFT14	ARFT14	ARFT14	ARFT14	ARFT14
6	h=1	ARFC0b	ARFC0a	ARFT1b	ARTVFC03	ARFT1b
	h=3	ARFC0b	ARFT14	ARFT1b	ARFT14	ARFT14
	h=6	ARFT1b	ARFT1b	ARFT1b	ARFT1b	ARFT1b
7	h=1	ARFC0a	ARFC0b	ARFC0b	ARFT1b	ARTVFC03
	h=3	ARFC04	ARTVFC03	ARTVFC03	ARFT1b	ARFT1b
	h=6	ARTVFC03	ARTVFC03	ARTVFC03	ARTVFC03	ARTVFC03
8	h=1	EXP	EXP	EXP	ARTVFC1b	ARTVFC1b
	h=3	ARTVFC03	EXP	EXP	ARTVFC0b	ARTVFC0b
	h=6	EXP	ARTVFC0b	ARTVFC0b	ARTVFC0b	ARTVFC0b
9	h=1	EX2	EX2	ARTVFC1b	ARTVFC1a	ARTVFCpb
	h=3	EX2	ARTVFC0b	ARTVFC0b	ARTVFC0a	ARTVFC0a
	h=6	EX2	ARTVFC0a	ARTVFC0a	ARTVFC0a	ARTVFC0a
10	h=1	ARTVFC0a	ARTVFC1b	ARTVFC1a	ARTVFCpb	ARTVFC1a
	h=3	ARTVFC0b	ARTVFC0a	ARTVFC0a	EXP	EXP
	h=6	ARFC1a	EXP	EXP	EXP	EXP

Notes:

See Table1 for definition of models

The loss function is $Loss^h_m = \frac{1}{N} \sum_{n=1}^N \frac{Loss^h_{n,m}}{Loss^h_{n,1}}$, $Loss^h_{n,m} = \frac{1}{T-h} \sum_{t=1}^{T-h} |e_{t+h,n,m}|^\rho$, where the

benchmark model is ARFC04 and e_{t+h} is the h-step ahead forecast error

Table 7a – Fraction of unstable series for which a forecasting method has lowest msfe

Method	AR	ES	NoChange	ARTV	LSTAR	ANN
h=1	0.17	0.04	0.04	0.17	0.21	0.31
h=3	0.15	0.1	0.01	0.16	0.35	0.20
h=6	0.22	0.08	0.05	0.08	0.29	0.28

Method	Linear	Time-varying	Non-linear
h=1	0.25	0.38	0.31
h=3	0.17	0.51	0.20
h=6	0.35	0.37	0.28

Notes:

Linear method includes AR,ES, No Change.

Time-varying method includes ARTV, LSTAR.

Non-linear method includes ANN.

Figures do not sum up to one because of rounding errors.

Table 7b – Fraction of unstable series for which a forecasting model is in the top-N

	N=1	N=5	N=10	N=15	N=20
ARFT1b	0.02	0.12	0.30	0.39	0.54
	0.04	0.14	0.28	0.35	0.48
	0.04	0.17	0.25	0.34	0.43
EX1	0.01	0.13	0.20	0.25	0.34
	0.06	0.13	0.22	0.25	0.30
	0.04	0.14	0.21	0.24	0.31
NOCHANGE	0.04	0.13	0.18	0.24	0.29
	0.01	0.11	0.16	0.20	0.26
	0.05	0.13	0.21	0.24	0.28
ARTVFC03	0.09	0.23	0.34	0.49	0.57
	0.06	0.19	0.33	0.46	0.51
	0.03	0.07	0.23	0.34	0.40
LS0103	0.03	0.13	0.25	0.34	0.40
	0.07	0.25	0.33	0.39	0.46
	0.07	0.21	0.31	0.36	0.41
ANF0b	0.09	0.21	0.29	0.37	0.39
	0.04	0.18	0.27	0.34	0.40
	0.05	0.18	0.29	0.36	0.41

Notes: See Table 1 for definition of models

The figures report the fraction of series for which a model is among the N models with the lowest msfe.

The reported models are the best performers in each class for N=1.

Table 8 – Unstable series, mean and percentiles of relative msfe for selected forecasting models

Forecast	Mean	0.02	0.10	0.25	0.50	0.75	0.90	0.98
ARFC04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ARFC0a	1.01	0.87	0.97	0.99	1.00	1.03	1.07	1.20
	0.99	0.74	0.89	0.97	1.00	1.04	1.08	1.20
	0.98	0.62	0.81	0.91	0.98	1.03	1.15	1.37
ARFT1b	1.01	0.69	0.84	0.95	1.01	1.06	1.18	1.45
	1.03	0.34	0.62	0.87	1.03	1.17	1.34	2.06
	1.29	0.14	0.35	0.68	1.01	1.29	2.74	5.19
EX1	1.59	0.70	0.86	0.95	1.09	1.48	2.14	4.57
	1.91	0.39	0.74	0.92	1.15	1.86	3.84	7.42
	2.83	0.15	0.40	0.79	1.36	2.63	6.23	18.26
NOCHANGE	1.67	0.64	0.85	1.00	1.14	1.63	2.36	4.57
	1.94	0.36	0.73	0.94	1.18	1.96	3.84	7.42
	2.86	0.13	0.39	0.79	1.38	2.64	6.23	18.26
ARTVFC03	1.01	0.68	0.85	0.92	0.99	1.05	1.17	1.53
	1.03	0.56	0.75	0.87	0.97	1.10	1.29	2.57
	1.33	0.28	0.50	0.83	1.03	1.47	2.05	3.94
LS0103	1.15	0.68	0.88	0.96	1.06	1.18	1.45	1.89
	1.19	0.46	0.73	0.87	1.05	1.35	1.73	2.89
	1.67	0.12	0.40	0.70	1.09	1.68	3.46	8.21
ANF0b	1.13	0.67	0.82	0.94	1.05	1.21	1.41	2.15
	1.40	0.33	0.73	0.90	1.13	1.41	2.07	4.54
	2.57	0.08	0.33	0.66	1.15	2.63	6.28	14.97

Notes:

The benchmark model is ARFC04

The models are the best from Table 6 and those from Table 7b

For each forecast the three rows correspond to, respectively, h=1, 3, 6

See Table 1 for the definition of the models

Table 9 –Stability Tests for Unemployment, IP and Inflation

Series	NY	KP1	KP2	BP	QLR	MLR	ALR		
Ger-UNEM	0.86	1.08	0.25	5.98	11.76	5.38	3.33		
Ger-IP	0.78	1.14	0.19	4.85	12.11	4.63	3.87		
Ger-INFL	1.03	1.13	0.3	8.48	** 13.6	7.92	** 4.89	*	
Ita-UNEM	0.47	0.87	0.17	10.29	**	8.71	3.78	2.41	
Ita-IP	0.57	0.58	0.07	2.15		8.37	3.3	2.13	
Ita-INFL	0.78	0.73	0.21	7.54	*	11.23	5.3	3.08	
Spa-UNEM	0.83	0.79	0.16	1.34		19.65	** 6.89	* 6.81	**
Spa-IP	1.34	** 0.99	0.14	5.76		16.21	** 8.21	** 5.53	**
Spa-INFL	1.74	*** 1.51	** 1.07	*** 9.52	**	39.79	*** 15.01	*** 15.76	***
Fra-UNEM	0.34	0.66	0.08	3.44		6.89	2.06	1.5	
Fra-IP	0.52	0.64	0.09	7.43	*	6.69	2.95	1.82	
Fra-INFL	1.3	** 1.04	0.44	* 3.18		30.39	*** 15.45	*** 11.48	***
Aus-UNEM	0.4	0.69	0.06	9.16	**	9	2.4	1.91	
Aus-IP	0.8	0.64	0.07	0.78		8.85	4.59	2.79	
Aus-INFL	0.39	0.94	0.26	0.78		9.34	2.71	2.33	
Lux-UNEM	0.92	1.23	* 0.38	* 7.59	*	9.96	5.33	3.22	
Lux-IP	0.58	0.98	0.24	5.1		9.52	3.23	2.38	
Lux-INFL	0.77	1.08	0.19	5.32		19.47	** 4.89	5.7	**
Net-UNEM	1.99	*** 0.75	0.13	9.26	**	23.88	*** 11.03	*** 7.4	***
Net-IP	0.49	0.48	0.02	5.54		7.16	3.49	2.17	
Net-INFL	1.28	** 1.26	* 0.34	* 4.17		20.85	*** 10.5	*** 7.33	***
Fin-UNEM	1.92	*** 0.85	0.14	8.65	**	135.35	*** 54.7	*** 63.69	***
Fin-IP	1.27	** 1.31	* 0.31	3.45		17.09	** 8.71	** 5.66	**
Fin-INFL	1.41	** 1.5	** 0.78	*** 3.06		20.2	*** 12.47	*** 7.6	***
Por-IP	0.46	1.29	* 0.23	1.59		8.79	2.66	2.15	
Por-INFL	1.23	** 1.15	0.57	** 27.29	***	38.04	*** 11.05	*** 14.85	***
Bel-UNEM	0.66	0.98	0.18	4.65		8.93	4.47	2.76	
Bel-IP	0.32	0.72	0.08	11.21	**	5.44	1.48	0.89	
Bel-INFL	1.66	*** 1.48	** 0.5	** 12.35	***	35.86	*** 14.38	*** 13.41	***
Ire-UNEM	0.63	0.83	0.23	1.58		11.06	4.55	3.02	
Ire-IP	0.69	0.98	0.28	1.9		7.77	4.13	2.34	
Ire-INFL	1.4	** 0.93	0.2	0.98		16.13	** 8.49	** 6.14	**

Notes:

The model is an AR3 for the change in unemployment, and growth in IP and CPI

NY is Nyblom's (1989) test

KP1 and KP2 are the Ploebeger and Kramer's (1992) supremum and mean square tests

BP is Breusch and Pagan's (1979) Lagrange multiplier test

QLR is Quandt's (1960) supremum F-test

MLR is Andrews and Ploebeger's (1994) mean F-test

ALR is Andrews and Ploebeger's (1994) average exponential F-test

*, **, and *** indicate significance at, respectively, 10%, 5% and 1% level

Table 10 - Ranking of competing models with different loss functions, selected series

IP growth

Rank		$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=2.5$	$\rho=3$
1	h=1	ARFTP4	ARFTP4	ARFTP4	ARFTP4	ARFCP4
	h=3	ARFTP4	ARFCP4	ARFTP4	ARFCP4	ARFCP4
	h=6	ARFCPa	ARFCP4	ARFTP4	ARFTP4	ARTVFC13
2	h=1	ARFC14	ARFC14	ARFC14	ARFC14	ARFC14
	h=3	ARFC14	ARFC14	ARFC14	ARFC14	ARFC14
	h=6	ARFTPa	ARFC14	ARFC14	ARFC14	ARTVFCP3

Unemployment (change)

Rank		$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=2.5$	$\rho=3$
1	h=1	NOCHANGE	ARTVFCP3	ARTVFCP3	ARTVFC13	ARTVFC13
	h=3	LSP063	LS1063	LS1063	LSP063	LSP063
	h=6	ANP213	ANP213	LSF1a	ANF1b	ANFPb
2	h=1	ARFTPb	ARTVFC13	ARTVFC13	ARTVFCP3	ARTVFCP3
	h=3	LS1063	LSP063	LSP063	LS1063	LS1063
	h=6	AN1213	AN1213	LSFPa	ANFPb	ANF1b

CPI inflation

Rank		$\rho=1$	$\rho=1.5$	$\rho=2$	$\rho=2.5$	$\rho=3$
1	h=1	ARFC04	ARFC04	ARFC04	ARFC04	ARFC04
	h=3	LS0103	ARFC04	ARFC04	ARFC04	ARFC04
	h=6	AN0223	ARFC04	ARFC04	ARFC0a	ARFC0a
2	h=1	ANF0b	ANF0b	ANF0b	ANF0b	ANF0b
	h=3	ARFC04	LS0103	LS0103	LS0103	LS0103
	h=6	ARFC04	AN0223	ARFC0a	ARFC04	ARFC04

Notes:

See Table1 for definition of models

The loss function is $Loss^h_m = \frac{1}{N} \sum_{n=1}^N \frac{Loss^h_{n,m}}{Loss^h_{n,1}}$, $Loss^h_{n,m} = \frac{1}{T-h} \sum_{t=1}^{T-h} |e_{t+h,n,m}|^\rho$, where the

benchmark model is ARFC04 and e_{t+h} is the h-step ahead forecast error

Unemployment for Portugal is not available

Table 11 - Fraction of series for which a forecasting method has lowest msfe

IP growth

Method	AR	ES	NoChange	ARTV	LSTAR	ANN
h=1	2/11	-	-	1/11	7/11	1/11
h=3	1/11	1/11	-	2/11	6/11	1/11
h=6	4/11	-	-	1/11	-	6/11

Method	Linear	Time-varying	Non-linear
h=1	2/11	8/11	1/11
h=3	2/11	8/11	1/11
h=6	4/11	1/11	6/11

Unemployment (change)

Method	AR	ES	NoChange	ARTV	LSTAR	ANN
h=1	3/10	1/10	-	2/10	-	4/10
h=3	2/10	-	-	-	3/10	5/10
h=6	1/10	1/10	-	-	3/10	5/10

Method	Linear	Time-varying	Non-linear
h=1	4/10	2/10	4/10
h=3	2/10	3/10	5/10
h=6	2/10	3/10	5/10

CPI inflation

Method	AR	ES	NoChange	ARTV	LSTAR	ANN
h=1	3/11	1/11	-	1/11	-	6/11
h=3	2/11	-	-	2/11	5/11	2/11
h=6	1/11	-	1/11	1/11	3/11	5/11

Method	Linear	Time-varying	Non-linear
h=1	4/11	1/11	6/11
h=3	2/11	7/11	2/11
h=6	2/11	4/11	5/11

Notes:

Unemployment for Portugal is not available

Linear method includes AR,ES, No Change.

Time-varying method includes ARTV, LSTAR.

Non-linear method includes ANN.