A simple benchmark for forecasts of growth and inflation *

Massimiliano Marcellino **
IEP-Università Bocconi, IGIER and CEPR

This version: October 2006

Abstract

A theoretical model for growth or inflation should be able to reproduce the empirical features of these variables better than competing alternatives. Therefore, it is common practice in the literature, whenever a new model is suggested, to compare its performance with that of a benchmark model. However, while the theoretical models become more and more sophisticated, the benchmark typically remains a simple linear time series model. Recent examples are provided, e.g., by articles in the real business cycle literature or by new-keynesian studies on inflation persistence. While a time series model can provide a reasonable benchmark to evaluate the value added of economic theory relative to the pure explanatory power of the past behavior of the variable, recent developments in time series analysis suggest that more sophisticated time series models could provide more serious benchmarks for economic models. In this paper we evaluate whether these complicated time series models can really outperform standard linear models for GDP growth and inflation, and should therefore substitute them as benchmarks for economic theory based models. Since a complicated model specification can over-fit in sample, i.e. the model can spuriously perform very well compared to simpler alternatives, we conduct the model comparison based on the out of sample forecasting performance. We consider a large variety of models and evaluation criteria, using real time data and a sophisticated bootstrap algorithm to evaluate the statistical significance of our results. Our main conclusion is that in general linear time series models can be hardly beaten if they are carefully specified, and therefore still provide a good benchmark for theoretical models of growth and inflation. However, we also identify some important cases where the adoption of a more complicated benchmark can alter the conclusions of economic analyses about the driving forces of GDP growth and inflation. Therefore, comparing theoretical models also with more sophisticated time series benchmarks can guarantee more robust conclusions.

J.E.L. Classification: *C2, C53, E30*
Keywords: Growth, Inflation, Time-Varying models, Non-linear models

---

* I am grateful to Todd Clark, Frank Diebold, Toni Espasa, Niels Haldrup, George Kapetanios, Lutz Killian, Hashem Pesaran, Mark Watson and seminar participants at Bocconi, Venezia, Humboldt, Carlos III, Aarhus, Bundesbank 8th Spring Conference and NBER 2006 EFWW Meeting for useful comments on a previous draft. Marco Aiolfi and Andrea Carriero provided excellent research assistance. The usual disclaimers apply.
** Correspondence to: IGIER, 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

A theoretical model for growth or inflation should be able to reproduce the empirical features of these variables better than competing alternatives. Therefore, it is common practice in the literature, whenever a new model is suggested, to compare its performance with that of a benchmark model. However, while the theoretical models become more and more sophisticated, the benchmark typically remains a simple linear time series model. Recent examples are provided in the literature on DSGE models, which incorporate most of the recent theoretical advancements in macroeconomic theory, but are usually compared with a linear VAR for a subset of the variables under analysis, see, e.g., Del Negro, Schorfheide, Smets, and Wouters (2004).

While a time series model can provide a reasonable benchmark to evaluate the value added of economic theory relative to the pure explanatory power of the past behavior of the variable, the many social, economic and political changes that occurred in the US after World War II can be expected to make modeling macroeconomic variables with constant parameter linear models particularly difficult. In this context, time-varying and non-linear models could have a comparative advantage over linear specifications, and there is an ever growing literature on this topic. On the other hand, there is no consensus on the magnitude and relevance of the structural breaks, and therefore on the need of going beyond linear models for the conditional expectation of growth or inflation, see, e.g., the debate between Cogley and Sargent (2001) and Sims (2001).

In this paper we conduct a detailed analysis of univariate time series models for US GDP growth and inflation. Our main goal is to establish whether simple autoregressive (AR) models can still be used as a sound benchmark for economic theory based models, or whether they should be substituted for more sophisticated specifications. We focus on univariate benchmark models because the theory based model should be at least capable of improving upon models that only rely on the past behavior of the variable of interest. Of course more sophisticated multivariate time series models or even structural models could be used as tougher competitors, but this would entail a substantial increase of complexity which few macroeconomists would be willing to undertake for evaluating their economic based models. Moreover, univariate linear models are often more robust than their multivariate (VAR) counterparts, see e.g. Banerjee and Marcellino (2005).

Therefore, beating a serious univariate time series model for growth and inflation can be considered as a good testing ground for an economic theory based model.

We consider a large variety of models, including: AR models with different deterministic components, stationarity assumptions, and lag length specification. Time-varying AR models whose parameter evolution can capture the small but frequent changes in model structure that emerge, for example, from the analysis of Stock and Watson (1996). Smooth transition AR models, which allow for more general patterns of parameter variation, ranging from a smooth evolution across regimes to close to abrupt changes, as in Markov switching models à la Hamilton (1989) or SETAR models, and therefore can accommodate larger structural changes. Artificial neural networks, which are hardly interpretable from an economic point of view but provide a powerful tool to approximate virtually any kind of non-linear behavior, see, e.g., Hornik, Stinchcombe and White (1989). Overall, we consider a total of 55 alternative specifications, described in details in Section 2, providing a close to exhaustive analysis of the role of nonlinearity in modeling GDP growth and inflation using univariate models. While macroeconomists might consider some of these models too complex, we can anticipate that we find a very limited role for nonlinearity for both GDP growth and inflation, while a careful specification of the linear models is very important.

The model comparison exercise can be conducted in-sample or out-of-sample. Within an in-sample framework, the models can be evaluated on the basis of their goodness of fit or of their capacity of replicating some features of the data, such as persistence or unconditional moments. As an alternative, more formal statistical procedures can be applied, such as tests for model selection, information criteria or encompassing evaluation, see e.g. Pesaran and Deaton (1976), Mizon and Richard (1986), Vuong (1989) or Marcellino and Mizon (2006) for an overview.

A comparison of the in sample goodness of fit of linear and non-linear models would be likely biased in favor of the latter because of their extensive parameterization, see, e.g., Van Dick and Franses (2003), and this could also affect the other evaluation criteria. Moreover, the in-sample data have been already used for estimation or calibration of the models under analysis, so that fresh data are better suited for an unbiased evaluation. These comments suggest that the relative performance of linear and non-linear models for GDP growth and inflation is better evaluated out-of-sample, on the basis of their forecasting ability.

The forecasting ability of the models can be assessed using alternative criteria, such as forecast encompassing tests, where the issue is whether the forecasts from one model can explain the forecast errors from another model, see e.g. Chong and Hendry (1986); pooling regressions, where the best model should have a weight of one in the pooled forecast and the competing model a weight of zero, see e.g. Diebold (1989); or the relative size of the loss function associated with the forecast errors generated by the models, see e.g. West (1996). We prefer the last approach, and compare the forecasts from the competing models on the basis of several loss functions, including the common mean absolute and mean square forecast error (mae and mse, respectively). Notice that if a model does not dominate the others on the basis of the mse, it also does not forecast encompass them and it does not have a weight of one in a pooling regression, see e.g. Marcellino (2000).

Out-of-sample evaluation is not immune to criticisms. In particular, it is typically less informative than in-sample comparisons since the latter are often based on the full parameter set. Moreover, a valid economic-theory based model can produce worse forecasts than a time-series model, see e.g. Clements and Hendry (1998) for a theoretical discussion or Favero and Marcellino (2005) for an empirical example and simulation evidence. However, if the goal of the analysis is forecasting gdp growth or inflation, the out-of-sample forecasting performance appears as the natural metric for ranking the competing models.

Our forecast evaluation covers a rather long time span, 1980-2004, in order to include a few business cycles and periods with a rather different behavior of inflation. We also consider separately the ‘80s and ‘90s, and either the periods of expansions and recessions for GDP growth or those of reserve and inflation targeting for inflation.

As a first step, we follow standard practice and conduct a pseudo real time forecasting exercise, where model specification, estimation and forecasting are repeated for each month or quarter in the evaluation period, using the latest available vintage of data. Detailed results are presented in Section 3 for GDP growth and Section 4 for CPI inflation.

Overall, as anticipated, we find that the quantitative gains from using the time-varying or non-linear models rather than linear specifications are either limited or totally non-existent. The loss reduction using the best non-linear specification is in the range 5-15% for inflation and even lower for GDP growth, *if* the linear models are carefully specified, e.g., by pre-testing for the presence of a unit root, carefully choosing the deterministic component, or using information criteria for lag length selection. It is worth recalling that we focus here on point forecasts, while it could be that the nonlinear models perform better for density or interval forecasts, see, e.g., Pesaran and Potter (1997) in the case of GDP growth.

Even though the forecasting gains from the more sophisticated time series models are quantitatively limited, they could be significant from a statistical and/or economic point of view. To evaluate whether the differences of each model with respect to the benchmark are statistically significant, we then bootstrap 200 series of GDP growth and inflation, compute the empirical distribution function of the mse for each model, and use the latter to obtain a 90% confidence interval for the mse of each model relative to the linear benchmark. The exercise is complicated by the fact that we do not want to bootstrap the series assuming that a specific model generates the data. Therefore, we use a sophisticated non-parametric bootstrap algorithm, see Politis and White (2004). The results, summarized in Section 5, indicate that the confidence interval for the relative mse are rather large but, in particular for inflation, there are some non-linear models whose associated relative mse falls outside the confidence interval, signaling a statistically significant improvement with respect to the linear benchmark.

To evaluate the economic significance of the forecasting gains from the more sophisticated time series models or improved specification methods for the linear models, we consider whether their size can be sufficient to question the conclusions of important studies on the determinants of inflation and growth. For example, the exhaustive and influential analysis of inflation models in Stock and Watson (1999a), led them to conclude that (from the abstract) "Inflation forecasts produced by the Phillips curve generally have been more accurate than forecasts based on other macroeconomic variables, including interest rates, money and commodity prices. These forecasts can however be improved upon using a generalized Phillips curve based on measures of real aggregate activity other than unemployment, especially a new index of aggregate activity based on 61 real economic indicator". However, if their benchmark linear Phillips curve for CPI inflation is substituted by an AR with time varying parameters, it turns out that the latter can be hardly beaten by *any* of the multivariate specifications considered by Stock and Watson (1999a). In other words, a careful modeling of parameter evolution, which seems to be required in this case since Stock and Watson (1999a) detected some instability in their estimated Phillips curves, matches the forecasting gains obtained from modeling a very large set of macroeconomic explanatory variables. A similar result emerges for the factor based inflation forecasts in Stock and Watson (2002). Of course, it may be possible to do even better by using non-linear multivariate models, but that would introduce additional substantial complications, in particular in a large data set context.

As another example, Stock and Watson (2003) conduct a careful examination of the role of financial variables and other macroeconomic time series for forecasting GDP growth during the periods 1971-84 and 1985-99. In both cases we find that there are no major advantages by adopting a nonlinear specification, but adding a linear trend into their benchmark AR model for 1985-99 makes it good enough to outperform even pooled forecasts using macroeconomic or financial indicators. Using a better benchmark is also very important, e.g., for the analysis in Ang, Piazzesi and Wei (2004) on the role of the yield curve for forecasting GDP growth. They conclude that the former is very important over the period 1990-2001, using an AR(1) as a benchmark in their forecasting exercise. However, a model with a linear trend and four lags would provide a much more serious competitor, its mse relative to the AR(1) for a three year ahead forecast is about 0.54, much lower than any of the models considered by Ang et al. (2004). Of course the linear trend in the model for GDP growth might just be approximating the behavior of the economic variables that Stock and Watson or Ang et al, consider: the point is that it seems to work better than these variables. It might be possible to do even better by combining the trend and the economic variables or by considering additional economic variables, but this is exactly the reason for comparing the economic theory based model with a tough benchmark and not with a simple one.

Another interesting by-product of the analysis of GDP growth over subperiods is that there is a marked decrease in the mean square error over the more recent subsample that, combined with the good performance of the linear specifications in both subsamples and the stable performance of the nonlinear models, suggests that the size of the shocks matters more than time-variation in the parameters in order to explain the reduction in the volatility of GDP growth observed in the more recent period.

This finding is in line with results in, e.g., Blanchard and Simon (2001), Kim and Nelson (1999), McConnel and Perez-Quiros (2000) and Sims and Zha (2004). Instead, it is in contrast with, e.g., DeJong, Liesenfeld and Richard (2003a), who find that a sophisticated model with time-varying parameters produces residuals with stable variability. However, the forecasting performance of this model, evaluated in DeJong, Liesenfeld and Richard (2003b), is unclear because of the choice of a simple benchmark, a random walk model, which performs poorly in our comparison except during recessions.

As a final robustness analysis on the relative performance of the more sophisticated time series models for inflation and GDP growth, we use real time datasets rather than the latest available set of values for GDP and CPI. In other words, in each period over the evaluation sample we use the set of data that were available in that period. Croushore and Stark (2001, 2003) and Stark and Croushore (2002) have shown, among others, how using the available information set rather than the final vintage of data can dramatically change the results of a forecasting exercise or of an empirical analysis, see Croushore (2004) for an overview.

The evaluation of the forecasting performance of our large set of non-linear models using real time data is of particular interest since, to the best of our knowledge, this is the first time that such a comparison is conducted using real time data. A priori we might expect that the non-linear models are better suited to handle the revision process and measurement errors in real time data due to their flexibility. On the other hand, if their specification is tailored for a particular vintage of data, the forecasts could be out of track in the presence of major data revisions.

The results reported in Section 6 indicate that the use of real time data induces a deterioration in the forecasting performance of the non-linear models, making the linear specifications systematically the best. This finding is also related to the improved relative forecasting performance of the linear models when the sample starts in the early ‘80s.

Moreover, also in this context and in particular for GDP growth, there are substantial gains from a careful specification of the linear models.

In summary, while real time data are more of interest for practical forecasting purposes than for model validation, their use does not seem to provide any additional advantage for the non-linear models versus simple but carefully specified linear models for growth and inflation.

2 Forecasting methods

A large variety of nonlinear models are now available for modelling and forecasting macroeconomic time series, see, e.g., Terasvirta (2005) and White (2005) for recent overviews. Since we want to propose a benchmark model, it should be relatively easy to estimate and evaluate, and its performance should have already been considered in previous analyses.

We consider several alternative models, along the lines of Marcellino (2004) who further extends the set of specifications analyzed in Stock and Watson (1999b). These two papers evaluate the relative forecasting performance of linear and nonlinear models for a large number of macroeconomic variables for, respectively, the euro area and the US. They find that on average linear models perform best, but for some time series nonlinear specifications can yield substantial gains. We are here interested in evaluating the extent of the gains for US GDP growth and inflation, with a more detailed and comprehensive analysis.

The formulation of a generic forecasting model is

$$y_{t+h}^{h} = f(Z_t; q_{ht}) + e_{t+h}, \qquad (1)$$

where y_t is the log of either real GDP or the CPI index, h indicates the forecast horizon, Z_t is a vector of predictor variables, e_t is an error term, and q_h is a vector of parameters, possibly evolving over time. 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

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

with associated forecast error

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

When y_t is treated as (possibly trend) stationary, it is $y^h_{t+h} = y_{t+h}$, while if y_t is I(1) then $y^h_{t+h} = y_{t+h} - y_t$. We present results for both cases, and for the case where y_t is inflation rather than the (log of the) CPI index. 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 can improve 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).

Notice 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. Also, since $e_{t+h} = (y_{t+h} - y_t) - (\hat{y}_{t+h} - y_t)$ and in our case y_t is expressed in logarithms, the forecast errors can be interpreted as errors in forecasting the h-period growth rate of the variable, e.g., quarter on quarter GDP growth or month on month CPI inflation when h=1.

We focus on 1-step ahead forecasts, but also consider horizons of 2 and 4 quarters for GDP growth and 3, 6 and 12 months for inflation. When the forecast horizon *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), Marcellino, Stock and Watson (2005), but the computational savings in our real time exercise are substantial.[1]

It is also worth noting that a few forecast errors from the more sophisticated non-linear and time-varying methods can be large, due to occasional problems in the estimation of these models. 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.[2]

In the next three subsections we 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 (1999b), Marcellino (2004).

2.1 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. From an economic point of view, the hypothesis is that persistence is the main explanation for the behaviour of the variables.

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.

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 the no change forecast was found to

[1] For example, we have about 300 periods for evaluation of inflation forecasts in the base case and for each period we run 200 bootstrap replications to compute standard errors around the relative loss. Therefore, if were using 1000 Monte Carlo replications to compute a simulation based forecast, overall we should compute about 60 millions forecasts for each of the (about 40) nonlinear models.

[2] The results are fairly robust to modifications in this trimming procedure, such as comparing the absolute value of a forecasted change with a multiple of previously observed changes, or using the average value of the variable rather than a no change forecast.

outperform even forecasts from large-scale structural models, see, e.g., Artis and Marcellino (2001). In our context, it can be better suited for inflation than for GDP growth, due to the higher persistence of the former.

2.2 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)):

$$q_{ht} = q_{ht-1} + u_{ht}, \quad u_{ht} \sim iid(0, l^2 s^2 Q), \tag{4}$$

where s^2 is the variance of the error term e in (1) and $Q = (E(Z_t Z_t^{'}))^{-1}$. From an economic perspective, this model allows for continuous but rather small changes in the dynamics of GDP growth and inflation, and could capture the minor instability common to several US macroeconomic time series, see, e.g., Stock and Watson (1996).

We inspect several values of l: 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 l = 0.005, and then we allow for selection of the number of lags (1,3,6) jointly with the value of l 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

$$y_{t+h}^h = a^{'} z_t + d_t b^{'} z_t + e_{t+h}, \tag{5}$$

where d_t is the logistic function $d_t = 1/(1+\exp(g_0 + g_1 z_t))$. Typically, this specification captures larger changes in the parameters of the model, allowing for a smooth transition from the old to the new parameter values. For example, it could capture the major increase and decrease of inflation in the '70s and '80s better than the AR or ARTV models. The smoothing parameter g_1 regulates the shape of parameter change over time. When $g_1 = 0$ the model becomes linear, while for large values of g_1 the model tends to a self-exciting threshold model with abrupt changes in the parameters (see, e.g., Granger and Terasvirta (1993), Terasvirta (1998) for details), so that it can also approximate the behavior of Markov-switching models à la Hamilton (1989).

For models specified in levels we consider the following choices for the threshold variable in d_t: $z_t = y_t$, $z_t = y_{t-2}$, $z_t = y_{t-5}$, $z_t = y_t - y_{t-6}$, $z_t = y_t - y_{t-12}$. For differenced variables, it can be $z_t = \Delta y_t$, $z_t = \Delta y_{t-2}$, $z_t = \Delta y_{t-5}$, $z_t = y_t - y_{t-6}$, $z_t = y_t - y_{t-12}$. In each case the lag length of the model is either 1 or 3 or 6. We report results for the following models: 3 lags and $z_t = y_t$ (or $z_t = \Delta y_t$ for the I(1) case); 3 lags and $z_t = y_t - y_{t-6}$; AIC or BIC selection of both the number of lags and the specification of z_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 (1999b).

2.3 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), Swanson and White (1997), White (2005) for their use as forecasting devices. They can be even more flexible than the time varying models for capturing the effects of parameter changes, outlying observations, asymmetry in the reaction to shocks or other forms of non-linearity induced by institutional changes or large shocks.

The so called single layer feedforward neural network model with n_1 hidden units (and a linear component) is specified as:

$$y_{t+h}^{h} = b_0' z_t + \sum_{i=1}^{n_1} g_{1i} g(b_{1i}' z_t) + e_{t+h}, \qquad (6)$$

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 $b_{11}' z_t$. A more complex model is the double layer feedforward neural network with n_1 and n_2 hidden units:

$$y_{t+h}^{h} = b_0' z_t + \sum_{j=1}^{n_2} g_{2i} g\left(\sum_{i=1}^{n_1} b_{2ji} g(b_{1i}' z_t) \right) + e_{t+h}. \qquad (7)$$

We report results for the following specifications: n_1=2, n_2=0,p=3 (recall that p is number of lags in z_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 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 (1999b).

Overall, there are 55 models in the forecast comparison exercise, 19 belong to the linear class, 21 are time-varying, and 15 are non-linear. They are summarized in Table 1. We also experimented with different specifications for these models, e.g., different lag lengths or different sets of transition variables in the fixed specifications for the quarterly models, but the following results are quite robust to these modifications. Therefore, for the sake of simplicity, we prefer to stick to the same set of model specifications for both inflation and GDP growth.

3. US GDP growth

The sample for quarterly real GDP covers the period 1959:1-2004:2, and it includes seven business cycles, based on the NBER recession dating. From Figure 1, which graphs the quarter on quarter GDP growth rates, the decline in volatility after the mid '80s is evident, while that in average growth is not so clear cut. However, descriptive statistics indicate that there was a drop of about 20% in the average growth rate, from about 0.0093 over 1959-1979 to 0.0075 over 1980-2004.

The main forecasting period we consider is 1980:1-2004:2, which starts with the twin recessions of 1980 and 1981. To mimic real time situations, in each quarter of the forecasting period we repeat the unit-root test, estimation and specification selection for each forecasting model. Using 1959-1979 as the first estimation sample is sufficient to guarantee a rather accurate estimation of the parameters of the nonlinear models.

The results of the forecast comparison exercise for different forecast horizons, loss functions, sample periods, and estimation methods are summarized in Table 2. We now discuss them in detail, focusing on linear models in the first subsection and on time-varying and non-linear models in the second subsection.

3.1 Linear models

The figures in column 2 of Table 2a show that, when the models are compared on the basis of their 1-quarter ahead mean square forecast error relative to that of an AR model in levels with constant and four lags (ARFC04), our benchmark in this paper, there are some gains by adding a linear trend to the regressors set, while there are only minor differences in the results for different assumptions on the stationarity of GDP growth (I(0), I(1) or pre-test), and number of lags in the AR models (fixed, AIC, BIC, though AIC seems to work better than BIC). The best AR model (i.e., the simplest one with lowest relative mse) is ARFT04, a model with constant, trend and four lags, whose relative mse is 0.91.

Changing the loss function from the mse to the mean absolute error (mae) or mean absolute cubed error (mace) does not alter substantially the ranking of the linear models, compare columns 2, 3, 4 of Table 2a. However, it is interesting to point out that the loss differentials across models increase when a larger weight is assigned to larger errors, i.e., when moving from the mae to the mace loss function. For example, the ARFT04 yields a relative loss of only 0.97 using the mae, while the value becomes 0.89 with the mace. This finding indicates that the ARFT04 performs better than the benchmark in particular when the forecast errors are rather large, which can be important when the forecasts are used in a policy making context.

Increasing the forecast horizon from one to two or four quarters ahead also has limited consequences on the relative ranking of linear models, and the ARFT04 remains the best. However, the gains with respect to the benchmark are larger, the relative mse decreases to 0.82 when h=2 and to 0.73 when h=4, compare columns 2, 5, 6 of Table 2a. The absolute mse increases with the forecast horizon, compare columns 2, 5, 6 of Table 2c, even if, from a theoretical point of view, this is not always necessarily the case when the parameters are estimated.

Repeating the forecast evaluation exercise over subsamples produces some interesting results. We consider two cases: a pure temporal division of the sample, 80-89 and 90-04, and a division based on the NBER recession / expansion classification. During the period 80-89 including a (deterministic or stochastic) trend in the linear models improves the performance in most cases. There is a larger number of models that beat the benchmark, the relative mse decreases to 0.82 for the ARFT04 model, but a slightly lower value (0.80) is obtained for ARFC14, which imposes a unit root rather than a linear trend on the GDP level. The results for the subsample 90-04 are instead similar to the full sample 80-04, ARFT04 is again the best linear model with a relative mse of 0.92, compare columns 2, 7, 8 of Table 2a. From the same columns of Table 2c, in the second evaluation sample there is also an increase of about 30% in the loss of the benchmark model, even larger for the best model. The better performance of trending models in the '80s is attributable to the pattern of declining growth from the peak in '82-'83 to the recession of the early '90s, see Figure 1, while the worse performance in forecasting growth over 1990-2004 is mostly due to the recessionary episodes in this period.

The figures in columns 9 and 10 of Table 2a and 2c indicate that there are even more marked differences in the forecasting performance of linear models during periods of expansion and recession. In expansions, the ranking of models and the values of the relative mse are similar to the full sample, not surprisingly since most of the period 1980-2004 is expansionary, and ARFT04 and ARFC14 are basically equivalent. In recessions, there is a marked deterioration for all models, the loss from the benchmark doubles with respect to the expansionary periods, and the best forecasting model becomes the random walk, with a relative mse of 0.38.

The different behavior of GDP growth over the business cycle relates to the old debate on the characterization of cycles as extrinsic phenomena, i.e. generated by the arrival of external shocks propagated through a linear model, versus intrinsic phenomena, i.e., generated by the nonlinear development of the endogenous variables. If the latter view is correct, as originally supported by Burns and Mitchell (1946) who treated expansions and recessions as two different periods, then the time-varying and nonlinear models should yield forecasting gains with respect to the linear models over the whole forecasting period 1980-04, and present a more stable performance over the two subsamples. We will evaluate whether this is the case in the following subsection.

As an alternative to the use of nonlinear models, rolling estimation is sometimes advocated as a means of robustifying the forecasts in the presence of structural changes in the estimation period (see, e.g., Pesaran and Timmerman (1995)). Therefore, in columns 11 and 12 of Table 2a we report the relative mse for the whole period 80-04 resulting from estimation using a rolling window of either 10 or 15 years (recursively updated). In practice, for a 10-year window, the estimation sample is 1970:1-1979:4 when forecasting 1980:1; it becomes 1970:2-1980:1 when forecasting 1980:2, then 1970:3-1980:2 when forecasting 1980:3, etc.

While the performance of the benchmark is virtually unaffected by the different method of estimation, compare columns 2, 11, 12 of Table 2c, the use of a shorter estimation sample makes the presence of a linear trend in the model redundant. The best model is ARFC14, but its relative mse is close to one, around 0.94, which makes ARFT04 combined with recursive estimation the best linear forecasting tool, when evaluated over the full period 1980-2004.

In summary, there are limited benefits from a more complex specification of a linear model than the benchmark AR(4) in levels with a constant, when the evaluation covers the whole period 1980-2004 and the forecast horizon is one quarter. Recursive estimation of the ARFT04 model yields the lowest mse among the class of linear models, but the gains with respect to the benchmark are lower than 10%. Moreover, the common model where GDP growth is regressed on a constant and four lags is a close second best. Instead, there emerge larger gains for longer forecast horizons, 2 or 4 quarters, and interesting differences over periods of expansion and recession, with the no change forecast performing particularly well in the latter.

3.2 Time varying and non-linear models

A first characteristic that emerges for the time-varying and non-linear models is that in general a fixed specification yields a comparable or lower relative mse than AIC or BIC based specifications, see column 2 of Table 2b. A parsimonious and fixed specification reduces the extent of over-fitting and seems to be preferable also with our rather long estimation sample. There are instead no clear cut results on the usefulness of a careful choice of the stationarity characteristic of GDP.

The best model is LS1103, an LSTAR model for GDP growth, three lags in the autoregressive component, and transition variable $z_t = y_t$, which yields a relative mse of 0.95, larger than the best linear model (ARFT04, with relative mse of 0.91). Actually, the main result emerging from column 2 of Table 2b is that there are no gains from the more sophisticated models.

When a larger (smaller) weight is assigned to large forecast errors, as with the mace (mae) loss function, the relative forecasting performance of the ARTV models is stable, while that of LSTAR and ANN models in general deteriorates (improves), compare columns 2, 3, 4 of Table 2b.. This pattern is due to the fact that more complex specifications can generate sporadic large forecast errors, notwithstanding the automatic trimming, that are amplified (reduced) when using the mace (mae) loss function.

For longer forecast horizons, 2 or 4 quarters, in general the performance of the ANN models worsens while that of the ARTV and LSTAR improves slightly, but not enough to outperform the best linear specifications, compare columns 5, 6 of Tables 2a and 2b.

About the split sample analysis, a few models beat the best linear specifications in the ‘80s, e.g., AN1203 or LSF1a, but the gains are limited and disappear in the ‘90s, compare columns 7 and 8 of Table 2b. From columns 9 and 10 of the same table, the separate evaluation over periods of boom and recession does not indicate any sophisticated model as a serious competitor for the linear specifications. This finding provides support in favour of the extrinsic characterization of business cycles, i.e. they are generated by the arrival of external shocks propagated through a linear model, even though some even more complicated nonlinear model might perform better.

Finally, the use of rolling rather than recursive estimation is basically ineffective for the ARTV and LSTAR models, while the more heavily parameterized ANN models typically suffer from the shorter estimation sample, which increases their relative mse.

Overall, there appear to be virtually no or very limited gains from the use of sophisticated time-varying and/or non-linear models for forecasting GDP growth, while a careful specification of the linear models can yield sizeable gains, in particular for longer forecast horizons.

4. US inflation

The monthly sample for the CPI index covers the period 1959:1-2004:6, and Figure 2 graphs the month on month CPI inflation rate. The graph highlights the irregular behaviour of inflation during the ’60 and ‘70s, which could be better captured by a nonlinear specification; the decline of inflation in the early ‘80s, which is related with the tight monetary policy implemented in this period; and the overall decline in the volatility of the variable, in particular in the ‘90s. The most recent period seems to be characterized by a return of volatility, which is mostly due to energy and food prices.

The main forecasting period for inflation is 1980:1-2004:6, as for GDP growth, and in each month of this period we repeat the unit-root test, estimation and specification selection for each forecasting model.

Table 3 summarizes the results of the forecast comparison exercise for different forecast horizons, loss functions, sample periods, and methods of estimation. We now discuss the figures, focusing first on linear models and then on time-varying and non-linear models. The third subsection conducts additional robustness analyses, evaluating whether the relative ranking of the models changes for a different price index or when the possibility of a unit root in the inflation process is taken into account.

4.1 Linear models

The figures in column 2 of Table 3a show that, when the models are compared on the basis of their 1-month ahead mean square forecast error relative to that of an AR model with constant and four lags, our benchmark also for inflation, there are usually some benefits from excluding a linear trend from the model and imposing the presence of a unit root in CPI (pre-testing yields similar values), while the use of information criteria for lag length selection is basically not relevant. The best AR model (i.e., the simplest model with lowest relative MSE) is ARFC14, which imposes a unit root in a model with constant and 4 lags. However, its relative mse is only 0.91.

Changing the loss function from the mse to the mean absolute error (mae) or mean absolute cubed error (mace) does not alter substantially the ranking of the linear models, compare columns 2, 3, 4 of Table 3a.

Increasing the forecast horizon from one to 12 months improves the performance of specifications based on information criteria, so that the best models become ARFC1a or ARFC1b, with relative mse of 0.78, 0.72 and 0.73, for, respectively, h=3, 6 and 12, compare columns 2, 5, 6, 7 of Table 3a.

From columns 2, 8, 9, 10, 11 of Table 3a, there are also minor changes in the ranking of the linear models when repeating the forecast evaluation exercise over either the two subsamples 80-89 and 90-04, or the two subsamples with, respectively, reserve and inflation targeting. However, the values of the loss function are substantially smaller in the '90s than in the '80s, and during interest rate targeting rather than reserves targeting, mimicking the behavior of the standard deviation of inflation, see Figure 2 and Table 3c.

The use of rolling estimation highlights an interesting feature, i.e., the relevance of a (stochastic or deterministic) trend component in the model increases with the length of the estimation sample. With a 10 year window rather than recursive estimation, no AR models can outperform the benchmark ARFC04. There is also a marked decrease in the loss, about 20%, that makes the ARFC04 (combined with 10 year rolling estimation) the best linear forecasting method, compare columns 2, 12, 13 of Tables 3a and 3c. This finding provides additional evidence on the presence of parameter instability, in particular during the first part of the sample.

4.2 Time varying and non-linear models

In the case of the more sophisticated models, as for GDP growth, the use of information criteria becomes less important with the complexity of the specification. For the ANN models the fixed specifications are systematically better than the AIC or BIC based ones. A similar finding emerges for the ARTV models, while there are some minor gains from the use of AIC or BIC for the LSTAR, compare column 2 of Table 3b.

For all non-linear models there are also little gains from treating the CPI as I(1) rather than I(0). The more flexible specification of these models is sufficient to capture the trending evolution of the CPI.

The best models in this class systematically belong to the ARTV group. This is an interesting finding, since a similar result emerges for the euro area, see Marcellino (2004). The ARTVFC03 model, where the level of CPI is modeled by an AR(3) specification with random walk evolution in the parameters, performs better than any of the linear specifications, with a relative mse of 0.82 (which is though rather close to that of the ARFC14, 0.91). Among the LSTAR models, the LSF1b specification (which models prices as I(1) using BIC model selection) is the best, and the differences with respect to the ARTVFC03 are minor. An interesting feature of these two models is their good performance throughout most of the evaluation exercises we conduct for forecasting inflation.

Changing the loss function by assigning a larger or smaller weight to large forecast errors is not relevant for the relative ranking of the models, see columns 2, 3, 4 of Table 3b.

The ranking is also rather robust to changes in the forecast horizon, but the gains with respect to the benchmark increase with the horizon. In particular, the relative mse for the ARTVFC03 decreases to 0.67 for h=3, 0.60 for h=6, and 0.57 for h=12, compare columns 5, 6, 7 of Table 3b.

About the split sample analysis, some ANN models perform very well in some subsamples, e.g., the AN0213 for the 80s or during reserve targeting, but rather poorly in other subsamples. Instead, the ARTV and LSTAR models have a rather stable behavior, and the ARTVFC03 beats the AR specifications in all cases, see columns 8, 9, 10, 11 of Table 3b.

Finally, contrary to the linear case, the use of rolling rather than recursive estimation has minor effects on forecasts from the ARTV models, suggesting that they are capable of capturing the modifications in the behavior of inflation. The more heavily parameterized LSTAR and ANN models typically suffer from the shorter estimation sample, with an increase in their relative mse compared to recursive estimation, compare columns 2, 12, 13 of Table 3b. The best model remains ARTVFC03, with gains of 10-15% with respect to the best linear specifications.

4.3 Additional robustness analyses

We now evaluate whether there can be additional forecasting gains from treating inflation as I(1) (i.e. the CPI as I(2)), and whether this can alter the ranking of the models in the previous subsection.

The target variable is $y^h_{t+h} = y_{t+h} - y_t$ when the CPI (y_t) is treated as I(1), and $y^h_{t+h} = \sum_{t+1}^{t+h} \Delta y_s - h\Delta^2 y_t = y_{t+h} - y_t - h\Delta^2 y_t$ when the CPI is I(2). This is a convenient formulation because, given that y_t and its lags are known when forecasting, the unknown component of y^h_{t+h} conditional on the available information is equal to y_{t+h} independently of the choice of the order of integration. This makes the mean square forecast error from models for second-differenced variables directly comparable with, for example, that from models for first differences only.

The results of the forecast comparison exercise are reported in columns 14, 15 and 16 of Table 3, for h=1 and different loss functions, using the ARFC04 for I(1) prices as the benchmark, in order to make the values directly comparable with those in columns 2, 3 and 4. Notice that we use the same notation for the different specifications, but it should be remembered that now I(0) and I(1) refer to the inflation rate rather than to the price level.

With this caveat in mind, there are little gains from treating inflation as I(1), and virtually no differences in the ranking of the models. The best linear model from column 14 is ARFC04, where inflation is treated as I(0), with a relative mse of 0.91. This model corresponds to ARFC14 in column 2, which was the best model and treated the price level as I(1) and inflation as I(0), though the number of lags differ due to differencing.

The best overall model for forecasting inflation remains the nonlinear specification ARTVFC03, with a relative mse of 0.83. This specification also yields the lowest mae and mace, compare columns 15 and 16 of Table 3. Such a finding is in line with the good and comparable performance of ARTVFC03 and ARTVFC13 when inflation is treated as stationary (columns 2, 3, 4).

Another possible source of the forecast gains from nonlinearity is the unstable behaviour of inflation in the ‘70s and early ‘80s. In the previous subsection we evaluated the forecasts in the ‘90s only, but the ‘70s and early ‘80s were used for estimation, or using rolling estimation, but the ‘70s and early ‘80s were still used in most of the estimation windows. Therefore, we now consider recursively forecasting inflation over the period 1994-2004, starting the estimation sample in 1984.

The figures in columns 17-20 of Table 3 indicate that the nonlinear models do loose most of their competitive advantage over the linear specifications.

Finally, we noticed that some of the volatility in the inflation rate is due to the food and energy components of the CPI. Hence, the nonlinear models could outperform the linear specification because of a better ability in predicting these erratic components of inflation.

To evaluate whether this is the case, we have repeated the forecasting exercise starting with the series of CPI less food and energy. A first finding that emerges from columns 21 to 26 of Table 3c is that, for virtually all forecast horizons, the losses of the benchmark model are reduced by 50% with respect to the case where food and energy prices are included in the CPI. Moreover, the gains from ARTVFC03 disappear for h=1, 3 and 6, and shrink substantially for h=12. A similar result emerges for the other nonlinear specifications, and overall the figures confirm the importance of the volatile components of the CPI to explain the good performance of nonlinear models.

In summary, some time-varying or nonlinear models can produce better forecasts of inflation than linear specifications, the ARTV models appear to be particularly promising, and the gains can be quantitatively substantial, in particular at longer horizons. The gains do not depend on the assumed degree of persistence in inflation, while they appear to be related to the sample period and to a better performance of the ARTV models in predicting the most erratic component of CPI, i.e., food and energy prices. If the evaluation is conducted on the most recent period or excluding food and energy prices, the linear benchmark ARFC04 can be hardly beaten.

5. Evaluating the relevance of the forecasting gains

In the two previous Sections we have shown that it is possible to reduce the mean square forecast error with respect to the benchmark AR(4) model by a careful specification of the linear models in the case of GDP growth, or of more sophisticated models in the case of inflation over a long evaluation period. Yet, in general the gains are quantitatively limited. In this Section we evaluate first whether the limited gains are nonetheless statistically significant, and then whether they are economically significant, in the sense of changing the conclusions of important studies on the determinants of GDP growth or inflation.

5.1 Statistical evaluation of the forecasting gains

To consider whether the mse reductions relative to the benchmark model are statistically significant, we would like to provide a confidence interval for the relative mse of each model at a certain confidence level, say 90%. The latter can be obtained either analytically or by means of simulation techniques. We prefer the latter method, which provides a more precise answer in finite samples.

To obtain the empirical distribution of the relative mse, we have generated 200 bootstrap replications of GDP growth and CPI inflation, by using the block bootstrap algorithm with automatic block length selection (Politis and White (2004)).[3] The levels of the series are then recovered by cumulating the growth series, starting with the actual level of each variable.

We prefer this non-parametric approach to the more standard parametric method, where the series are generated using the estimated parameters for the benchmark model. The former permits the derivation of the empirical distribution function of the relative mse without assuming that the benchmark model generates the data. It can therefore provide more reliable results in our context, where there is substantial uncertainty about the optimality characteristics of the benchmark. We would also like to point out that the list of models under comparison in this paper is fixed and taken from previous published work, in order to reduce possible data mining problems.

The 90% confidence intervals for the relative mse are reported in Table 4 for each forecast horizon, in panel A for GDP growth and in B for inflation.

For GDP growth, no models yield a significantly lower mse than the benchmark, except the best model from Section 3, namely ARFT04, when h=1, 2.

For inflation there is a larger set of models whose relative mse falls outside the 90% confidence interval, and it includes the best models identified in Section 4 for each forecast horizon. In particular, both ARTVFC03 and LSF1b are systematically significantly better than the benchmark. However, it is worth noting that here the comparison is conducted with respect to the AR(4) model in levels. If the best linear model is used as a benchmark, usually ARFC1a for inflation, the differences with respect to the nonlinear specifications shrink to about 10% and are only marginally statistically significant.

Two final points that deserve attention are the following. First, if the statistical significance of the forecasting gains were assessed by means of the White's (2000) reality check for data snooping, they would be even less significant. However, it is not clear whether such a method is appropriate in our context. In fact, it might be that some models are better than the simple AR(4) benchmark just by chance but, on the other hand, the models in our comparison are not coming from a specification search but from a list used in previously published papers (Stock and Watson (1999), Marcellino (2004)), where they were not selected on the basis of the forecast accuracy for growth and inflation. Second, the bootstrap algorithm requires stationarity of growth and inflation. The good forecasting performance of stationary AR models for these two variables suggests that this is indeed the case, but the superior performance of some nonlinear models for inflation questions it. However, if the data were resampled parametrically using these nonlinear models, the significance of their forecasting gains would be even larger.

---

[3] Andrew Patton has compiled a Matlab computer code for implementing the above block selection algorithm via flat-top lag-windows; his code is now made publicly available from his website: http://fmg.lse.ac.uk/~ patton/code.html.

5.2 Economic evaluation of the forecasting gains

We now consider whether the results we have obtained can affect the conclusions of economic analyses on the role of macroeconomic and financial variables for understanding the development of GDP growth and inflation.

The article by Stock and Watson (2003) is becoming a reference for forecasting GDP growth due to their careful examination of the role of financial variables and other macroeconomic time series. They find that the results differ during the periods 1971-84 and 1985-99, and across countries.

Focusing on the US over the more recent period, 1985-99, asset prices appear to yield only a limited forecasting gain over the short horizon, h=2, relative to a benchmark AR model with BIC lag length. The relative mse of pooled forecasts using asset prices as regressors is 0.94, compare the final column of Stock and Watson's (2003) table 9C. When the forecast horizon increases to one or two years, h=4 and 8, monetary variables become more important, yielding a relative mse of 0.94 for h=4 and 0.89 for h=8. Other financial series, prices and macroeconomic variables are less useful.

From Section 3, our best forecasting model for the whole period 1980-2004 is ARFT04. Table 5A indicates that overall it remains the best model also over the period 1985-1999. It yields a relative mse with respect to the AR-BIC (ARFC1b in our notation) used by Stock and Watson (2003) of 0.94 for h=2, 0.89 for h=4 and 1.03 for h=8.

Therefore, using our ARFT04 as a benchmark rather than Stock and Watson's AR-BIC makes financial and monetary variables basically irrelevant for up to one year ahead forecasts of GDP growth, while monetary series keep their importance for longer horizons.

There are instead no major changes for the period 1971-84, since in this case it is not possible to beat Stock and Watson's benchmark, compare Table 5A. They found that asset prices were particularly useful for forecasting growth during 1971-84 and this result, combined with their poor performance after 1984, suggests that financial markets could be related with the observed reduction in the volatility of GDP growth, an issue that deserves additional investigation.

The economic importance of the choice of a good benchmark for models of GDP growth is even more evident from the analysis of Ang, Piazzesi and Wei (2004). They focus on the role of the yield curve for forecasting GDP growth and conclude that the former is very important over the period 1990-2001, using an AR(1) as a benchmark in their forecasting exercise, ARFC11 in our notation.

In Table 5B we compare the ARFC11 with other models over the period 1990-2001. It turns out that our favorite model, ARFT04 systematically and substantially outperforms the ARFC11, the relative mses are 0.86 for h=1, 0.80 for h=4, 0.67 for h=8 and 0.54 for h=12.

If Ang's et al. (2004) benchmark is substituted with the ARFT04, all their financial regressors are systematically outperformed, compare the relative mse in their tables 10, 11, and 14.

An interesting by-product of the split sample analysis in Tables 5A and 5B is that it highlights a marked decrease in the mean square error over the more recent subsamples compared with the period 1970-84. For h=1, the mse of the benchmark ARFC04 model is six times smaller in 1985-1999 than in 1970-84. This finding, combined with the good performance of the linear specifications in both subsamples, provides additional evidence in favor of the statement that the size of the shocks matters more than time-variation in the parameters in order to explain the reduction in volatility of GDP growth observed in the more recent period.

Moving now to inflation, the influential and exhaustive analysis of Stock and Watson (1999a), led them to conclude that (from the abstract) "Inflation forecasts produced by the Phillips curve generally have been more accurate than forecasts based on other macroeconomic variables, including interest rates, money and commodity prices. These forecasts can however be improved upon using a generalized Phillips curve based on measures of real aggregate activity other than unemployment, especially a new index of aggregate activity based on 61 real economic indicator".

From table 5C, the best nonlinear model for the subsample 1984-1996 is again ARTVFC03, namely, an AR model specified in levels, with a constant, three lags, and time varying parameters. If we include this model in Stock and Watson's (1999a) comparison, their table 4, it turns out that it has a relative mse of 0.71, which cannot be beaten by *any* of the multivariate specifications, Stock and Watson's preferred model has a relative mse of 0.83.

The results are quite different for the earlier sample, 1970-1983. In this case Stock and Watson's benchmark Phillips curve systematically beats all our time series models.

These two findings, combined with those in Table 3, column 17, indicate that whenever the '70s are included in the estimation period of models for inflation it becomes quite important to allow for parameter time variation (Stock and Watson (1999a) also detected some instability in their estimated Phillips curves). The economic consequences can be substantial. In particular, when the focus is on the more recent period, allowing for some random parameter time variation, as in our model ARTVFC03, matches the forecasting gains obtained from sophisticated versions of the Phillips curve.

Another finding emerging from Table 5C is the different mse in the two subperiods 1970-83 and 1984-1996, the latter is ten times smaller than the former using Stock and Watson's benchmark. This result casts some doubts on forecast evaluation exercises based on a very long forecast period only, such as Stock and Watson (2002). They extract the factors from a very large dataset of macroeconomic variables that includes hundreds of series, use them as regressors in a set of alternative linear specifications for inflation, and compare the resulting forecasts over the period 1970-1998.

Focusing on one year ahead forecasts for inflation, from Table 5D none of our models can beat Stock and Watson's (2002) preferred factor specification. However, this result is just due to the reported different performance of the models over the 70's and 80-90's. The different performance of models for inflation over sub-periods is also emphasized by Atkeson and Ohanian (2001) and Stock and Watson (2005).

A final interesting feature of the ARTV models for inflation is the following. Let us consider the model ARTVFC10, namely,

$$\Delta y_t = q_t + e_t, \qquad q_t = q_{t-1} + u_t, \tag{8}$$

where y_t is the (log) price level, Δy_t is monthly inflation, $e_t \sim iid(0, s_e^2)$, $u_t \sim iid(0, s_u^2)$, and e_t and u_t are uncorrelated at all leads and lags. The model for the change in monthly inflation is

$$\Delta\Delta y_t = u_t + \Delta e_t, \tag{9}$$

which is also equivalent to an IMA(1,1) model such as

$$\Delta\Delta y_t = e_t + a e_{t-1}, \tag{10}$$

where the parameter α is a function of s_e^2 and s_u^2. The model for quarterly inflation, x_t, becomes an IMA(1,3):

$$\Delta x_t = (1 + L + L^2)(e_t + a e_{t-1}), \tag{11}$$

where, denoting as usual the lag operator as L, it is $x_t = (1 + L + L^2)\Delta y_t$. The model for quarterly inflation at the quarterly rather than monthly frequency can be obtained by applying standard techniques for temporal aggregation of ARIMA models (see e.g. Marcellino (1999)). It turns out that it is an IMA(1,1) model:

$$\Delta x_t = (v_t + b v_{t-1}), \tag{12}$$

where τ indexes quarterly time series, and the parameter β is again a function of s_e^2 and s_u^2. This coincides with the reduced form representation of the model suggested by Stock and Watson (2005) for quarterly US inflation. They allow the MA parameter β to be time-varying by making the variance of the permanent and transitory components of inflation also time-varying (s_u^2 and s_e^2, respectively, in our notation). [4]

The ARTVFC10 model in (8) has two main differences with respect to our preferred specification, ARTVFC03. First, it imposes the presence of a unit root. This has minor effects, since it implies only a slight deterioration in the forecasting performance. Second, it has less dynamics. However, the two additional (time-varying) roots in ARTVFC03 are substantially reduced by temporal aggregation (see, e.g. Marcellino (1999)). Therefore, overall, the IMA(1,1) model could represent a good approximation for the model for quarterly inflation implied by the ARTVFC03 specification at the monthly level.

The exact model for quarterly inflation at the quarterly frequency implied by the ARTVFC03 is an ARMA(3,3) model whose parameters have a very complex temporal evolution, even if the variance of the permanent and transitory components of inflation

[4] Stock and Watson (2005) define y_t as the (log of the) average of the price level over three months, but the resulting quarterly inflation series at the quarterly level are very similar, the correlation is higher than 0.95.

(s_u^2 and s_e^2) are constant. It would represent an interesting benchmark, e.g., for Stock and Watson's (2005) and Ang, Bekaert and Wei (2005) analyses of quarterly US inflation.

In summary, the gains from a more careful specification of the benchmark model for GDP growth and inflation appear to be both statistically and economically significant. In the case of GDP growth, focusing on a linear specification suffices, while for inflation evaluating more complex models can yield additional benefits when the estimation sample is long enough to include data before the mid-'80s.

6 A real time analysis

To evaluate the role of real time data, we use quarterly vintages taken from the Philadelphia Federal Reserve website. In order to have a balanced panel without missing values and spanning the same sample period for GDP growth and inflation we use the sample 1960:1-2004:2 for GDP and 1960:1-2004:6 for the CPI. For GDP, in each quarter we use a different data set that coincides with the data on GDP that were available in that quarter. In the case of inflation, since the data are updated only each quarter, we use the same vintage for three consecutive months, until the next vintage becomes available. The latest 42 quarterly vintages are available.

For computational reasons, we adopt rolling estimation with a 15-year window, whose results we saw are very similar to recursive estimation. The forecasts for the last 42 quarters (i.e. 1994:1-2004:2), or corresponding months in the case of inflation, are then compared with the first release of the actual data. We consider the same forecast horizons as is the exercise with the final vintage of data, i.e., h=1, 2, 4 for GDP growth and h=1, 3, 6, 12 for inflation. Notice that an alternative strategy would be to add new data vintages as the sample period is extended, with the data earlier in the sample left untouched. While this approach can yield improved estimators (see e.g. Koening, Dolmas and Piger (2003)), it is rather different from the actual practice of real world forecasting.

The assessment of the forecasting performance with real time data of the large set of non-linear models we consider is of particular interest, since we think that this is the first time that these models are evaluated with real time data. A priori we might expect that the non-linear models are better suited to handle the revision process and measurement errors in real time data. On the other hand, if their specification is tailored for a particular vintage of data, the forecasts could be out of track in the presence of major data revisions.

Starting with the linear models, for GDP growth the best specification remains ARFT04 at all forecast horizons, while for inflation it is virtually impossible to outperform the benchmark ARFC04, compare Table 6a.

From the figures in the columns of Table 6b, we see that for the more complex models there is in general an even larger deterioration in the forecasting performance with respect to the benchmark when using the real time data, the more so the more complicated the specification of the model. For inflation, neither ARTVFCO3 nor LSF1b beat the linear benchmark, and the relative mse can be substantially larger than one at long horizons.

The differences with respect to the first columns of Tables 2 and 3 using the final vintage of data are dramatic. Possible reasons underlying this result were suggested by Elliott (2002) and include differences in the lag structures for real-time and revised data, greater persistence in the latest-available series, and the fact that a wider variety of models are selected using AIC or BIC using real-time data rather than revised data.

However, the differences with respect to the final vintage of data shrink substantially when the comparison is based on an equal forecast evaluation period. In particular, in the case of inflation, we have seen that the performance of the nonlinear models worsens substantially when evaluated over the subsample 1994-2004 rather than 1980-2004, see columns 17-20 of Table 3. The figures in these columns are much closer to the corresponding ones in Table 6, and a similar finding holds for the values for GDP growth in panel B of Table 5.

Overall, the use of real time data can worsen the forecasting performance of nonlinear models for growth and inflation, but the evaluation period seems to play a larger role.

7. Conclusions

In this paper we have provided an extensive evaluation of the role of sophisticated nonlinear time series models for GDP growth and inflation. Our main conclusion is that in general linear time series models can be hardly beaten if they are carefully specified, and therefore still provide a good benchmark for theoretical models of growth and inflation. This finding is particularly evident when using real time data or considering only the period starting in the mid-‘80s. However, we have also identified some important cases where the adoption of a more complicated benchmark can alter the conclusions of economic analyses about the driving forces of GDP growth and inflation. Therefore, comparing theoretical models also with more sophisticated time series benchmarks can guarantee more robust conclusions.

References

Ang, A., Piazzesi, M. and Wei, M. (2004), "What does the yield curve tell us about GDP growth?", NBER Working Paper 10672.

Ang, A., Bekaert, G. and Wei, M. (2005), "Do macro variables, asset markets or surveys forecast inflation better?", NBER Working Paper 11538.

Artis, M. and Marcellino, M. (2001), "Fiscal forecasting: the track record of IMF, OECD and EC", *Econometrics Journal*, 4, s20-s36.

Atkeson, A. and Ohanian, L.E. (2001), "Are Phillips curves useful for forecasting inflation?", *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2-11.

Banerjee, A. and Marcellino, M. (2005) "Are there any reliable leading indicators for the US inflation and GDP growth?", *International Journal of Forecasting*, forthcoming

Blanchard, O. and Simon, J. (2001), "The Long and Large Decline in U.S. Output Volatility", *Brooking Papers on Economic Activity*, 135-164.

Box, G.E.P. and Jenkins, G.M. (1970), *Time series analysis, forecasting and control*, San Francisco: Holden Day.

Burns, A. F. and W. C. Mitchell (1946), "Measuring business cycles", NBER Studies in Business Cycles no. 2 (New York).

Chong, Y. Y., and Hendry, D. F. (1986), "Econometric evaluation of linear macroeconomic models", *Review of Economic Studies*, 53, 671–690.

Clements, M.P. and Hendry, D.F. (1996), "Multi-step estimation for forecasting", *Oxford Bulletin of Economics and Statistics*, 58, 657-684.

Cogley, T. and Sargent, T.J. (2001), "Evolving Post World War II U.S. Inflation Dynamics", *NBER Macroeconomics Annual*, 16, 331-373.

Croushore, D and Stark, T. (2001) "A Real-Time Data Set for Macroeconomists", *Journal of Econometrics*, 105, 111-130.

Croushore, D. and Stark, T. (2003) "A Real-Time Data Set for Macroeconomists: Does the Data Vintage Matter?", *Review of Economics and Statistics*, 85, 605–617.

Croushore, D. (2004) "Forecasting with real-time macroeconomic data", mimeo.

De Jong, N.D., Liesenfeld, R. and J.-F. Richard (2003a) "A Structural Break in U.S. GDP?", mimeo.

De Jong, N.D., Liesenfeld, R. and J.-F. Richard (2003b) "A Nonlinear Forecasting Model of GDP Growth", *Review of Economics and Statistics*, forthcoming.

Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. (2004), "On the Fit and Forecasting Performance of New Keynesian Models", Working Paper 2004-37, Federal Reserve Bank of Atlanta.

Van Dijk, D.J. and P.H. Franses, (2003), "Selecting a nonlinear time series model using weighted tests of equal forecast accuracy", Econometric Institute Report 315, Erasmus University Rotterdam.

Diebold, F.X. (1989), “Forecast combination and encompassing: Reconciling two divergent literatures”, *International Journal of Forecasting*, 5, 589-592.

Diebold, F.X. and Kilian, L. (2000), “Unit Root Tests are Useful for Selecting Forecasting Models,” *Journal of Business and Economic Statistic*s, 18, 265-273.

Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996), “Efficient tests for an autoregressive unit root”, *Econometrica*, 64, 813-36.

Elliott, G. (2002), “Comments on 'Forecasting with a Real-Time Data Set for Macroeconomists,'” *Journal of Macroeconomics*, 24, 533-539.

Favero, C.A. and Marcellino, M. (2005) “Modelling and Forecasting Fiscal Variables for the euro Area”, *Oxford Bulletin of Economics and Statistics*, 67, 755-783.

Granger, C.W.J. and Terasvirta, T. (1993), *Modelling nonlinear economic relationships*, Oxford: Oxford University Press.

Hamilton, J. D. (1989), “A new approach to the economic analysis of nonstationary time series and the business cycle”, *Econometrica* 57: 357-384.

Hornik, K., Stinchcombe, M. and White, H. (1989), “Multilayer feedforward networks are universal approximators”, *Neural Networks*, 2, 359-66.

Kim, C.J. and Nelson, C.R.(1999), “Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of Business Cycle”, *Review of Economics and Statistics*, 81, 1-10.

Koenig, E.F., Dolmas, S. and Piger, J. (2003), “The Use and Abuse of Real-Time Data in Economic Forecasting**”,** *Review of Economics and Statistics*, 85, 618-628.

Marcellino, M. (1999) “Some consequences of temporal aggregation for empirical analysis”, *Journal of Business and Economic Statistics*, 17, 129-136.

Marcellino, M. (2004) “Forecasting EMU macroeconomic variables”, *International Journal of Forecasting*, 20, 359-72.

Marcellino, M. and Mizon, G.E. (2006), *Progressive Modelling: Encompassing and Hypothesis Testing*, Oxford University Press (*In preparation*)

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. (2005) “A Comparison of Direct and Iterated AR Methods for Forecasting Macroeconomic Series h-Steps Ahead”, *Journal of Econometrics*, forthcoming.

McConnell, M.M. and Perez-Quiros, G. (2000), “Output Fluctuations in the United States: What has Changed Since the Early 1980's?” *American Economic Review*, 90, 1464-1476.

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.

Mizon, G. E., and Richard, J.-F. (1986). "The encompassing principle and its application to non-nested hypothesis tests", *Econometrica*, 54, 657–678.

Nyblom, J. (1989), "Testing for constancy of parameters over time", *Journal of the American Statistical Association*, 84, 223-230.

Pesaran, M. H., and Deaton, A. S. (1978). "Testing non-nested non-linear regression models", *Econometrica*, 46, 677–694.

Pesaran, H. and A. Timmermann (1995), 'Predictability of Stock Returns: Robustness and Economic Significance', *Journal of Finance* 50, 1201-1228.

Pesaran, M.H. and Potter, S.M. (1997), "A Floor and Ceiling Model of US Output", *Journal of Economic Dynamics and Control*, 21, 661-695.

Politis, D. and White, H. (2004) "Automatic Block-Length Selection for the Dependent Bootstrap", *Econometric Reviews*, 23, 53-70.

Sims, C.A. (2001), "Comment on Sargent and Cogley's 'Evolving Post World War U.S. Inflation Dynamics' ", *NBER Macroeconomics Annual*, 16, 373-379.

Sims, C.A. and Zha, T. (2004), "Were there regime switches in U.S. monetary policy?", Working Paper 2004-14, Federal Reserve Bank of Atlanta.

Stark, T. and Croushore, D. (2002) "Forecasting with a Real-Time Data Set for Macroeconomists", *Journal of Macroeconomics*, 24, 507-31.

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. (1999a) "Forecasting Inflation", *Journal of Monetary Economics*, 44, 293-335.

Stock, J.H. and Watson, M.W. (1999b), "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.

Stock, J. H. and M. W. Watson (2002), "Macroeconomic Forecasting Using Diffusion Indexes", *Journal of Business and Economic Statistics*, 20, 147-62.

Stock, J. H. and M. W. Watson (2003), "Forecasting output and inflation: the role of asset prices", *Journal of Economic Literature* 41(3), 788-829.

Stock, J. H. and M. W. Watson (2005), "Has inflation become harder to forecast?", mimeo.

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.

Terasvirta, T. (2005), "Forecasting with Nonlinear Models", in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, forthcoming.

Vuong, Q. H. (1989). "Likelihood ratio tests for model selection and non-nested hypotheses", *Econometrica*, 57, 307–333.

West, K.D. (1996), "Asymptotic inference about predictive ability", *Econometrica*, 64, 1067-1084.

White, H. (2000), "A reality check for data snooping", *Econometrica*, 68, 1097-1126.

White, H. (2005), "Approximate Nonlinear Forecasting Methods" , in G. Elliott, C.W.J. Granger and A. Timmermann (eds.), *Handbook of Economic Forecasting*, forthcoming.

Figure 1: Quarterly GDP growth

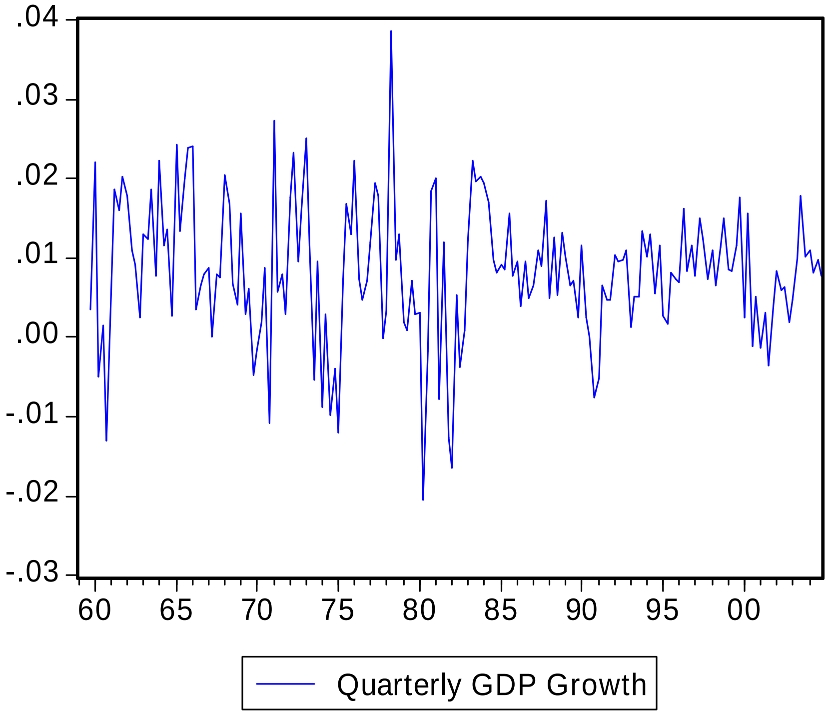

Figure 2: Monthly CPI inflation

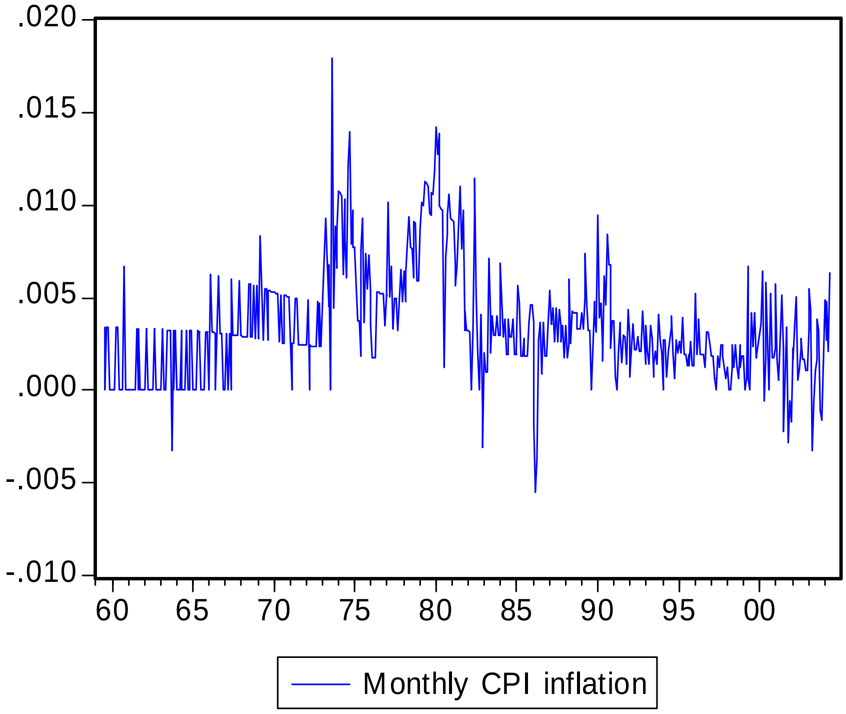

Table 1: Forecasting models

Linear methods

ARF(X,Y,Z)	*Autoregressive models* (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)
NOCHANGE	*No change forecast* (1 model)

Time-varying methods

ARTVF(X,Y,Z)	*Time-varying AR models* (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)	*Logistic smooth transition* (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) Y = transition variable, 10 ($z_t = y_t$), 06 ($z_t = y_t - y_{t-6}$) Z = 3 (p, lag length)
LSF(X,W)	*Logistic smooth transition* (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)	*Artificial neural network models* (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)	*Artificial neural network models* (6 models) X = 0 (stationary), 1 (I(1)), P (pre-test) S = a (AIC on n_1, n_2, p), b (BIC)

Table 2: Forecasting US GDP growth

		Recursive					Recursive split				Rolling	
		h=1			h=2	h=4	h=1				h=1	
Col.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]
	MODEL	MSE	MAE	MACE	MSE		MSE 80-89	MSE 90-04	MSE Booms	MSE Recessions	window 10 years	window 15 years
A - Linear	**ARFC04**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	ARFT04	0.91	0.97	0.89	0.82	0.73	0.82	0.92	0.87	1.13	1.03	1.02
	ARFC14	0.95	1.00	0.95	0.91	0.91	0.80	0.98	0.87	1.34	0.97	0.94
	ARFT14	0.98	1.00	0.95	1.00	1.02	0.92	0.99	0.97	1.00	1.00	0.97
	ARFCP4	0.95	1.00	0.95	0.91	0.91	0.80	0.98	0.87	1.34	0.97	0.94
	ARFTP4	0.98	1.00	0.95	1.00	1.02	0.92	0.99	0.97	1.00	0.97	0.94
	ARFC0a	1.01	1.00	1.01	1.03	1.01	1.01	1.01	1.00	1.02	1.00	1.00
	ARFT0a	0.91	0.96	0.90	0.82	0.75	0.83	0.93	0.87	1.14	1.07	1.02
	ARFC1a	0.96	0.99	0.97	0.91	0.91	0.81	0.98	0.87	1.39	0.96	0.94
	ARFT1a	0.99	1.00	0.99	1.01	1.03	0.94	1.00	0.98	1.06	1.00	0.97
	ARFCPa	0.96	0.99	0.97	0.91	0.91	0.81	0.98	0.87	1.39	0.96	0.94
	ARFTPa	0.99	1.00	0.99	1.01	1.03	0.94	1.00	0.98	1.06	0.96	0.94
	ARFC0b	1.09	1.03	1.15	1.02	1.01	1.17	1.07	1.07	1.16	1.00	1.00
	ARFT0b	0.98	1.00	0.98	0.84	0.75	1.02	0.97	0.93	1.21	0.97	0.99
	ARFC1b	1.05	1.03	1.13	0.92	0.91	0.93	1.07	0.93	1.62	1.03	1.01
	ARFT1b	1.07	1.02	1.12	1.01	1.00	1.09	1.06	1.04	1.22	1.06	1.04
	ARFCPb	1.05	1.03	1.13	0.92	0.91	0.93	1.07	0.93	1.62	1.03	1.01
	ARFTPb	1.07	1.02	1.12	1.01	1.00	1.09	1.06	1.04	1.22	1.03	1.01
	NOCHANGE	3.33	2.02	5.29	4.30	4.47	4.96	3.08	3.94	0.38	3.33	3.27
B - Time Varying and Nonlinear	**ARTVFC03**	1.03	1.04	1.04	1.01	1.10	0.97	1.04	1.01	1.16	0.99	0.97
	ARTVFC13	1.00	1.03	0.99	0.94	0.97	0.93	1.02	0.95	1.29	0.96	0.94
	ARTVFCP3	1.00	1.03	0.99	0.94	0.97	0.93	1.02	0.95	1.29	0.96	0.94
	ARTVFC0a	1.13	1.08	1.20	1.15	1.29	1.03	1.14	1.14	1.07	0.99	1.12
	ARTVFC1a	1.03	1.03	1.05	1.04	0.90	0.91	1.04	1.00	1.17	0.97	0.96
	ARTVFCPa	1.03	1.03	1.05	1.04	0.90	0.91	1.04	1.00	1.17	0.97	1.60
	ARTVFC0b	1.13	1.08	1.20	1.15	1.23	1.03	1.14	1.14	1.07	0.99	1.39
	ARTVFC1b	1.03	1.02	1.06	0.97	0.86	0.91	1.05	1.00	1.17	0.97	1.92
	ARTVFCPb	1.03	1.02	1.06	0.97	0.86	0.91	1.05	1.00	1.17	0.97	0.96
	LS0103	1.07	1.09	1.11	0.95	0.96	1.16	1.06	1.04	1.21	1.24	1.05
	LS1103	0.95	1.04	0.95	0.89	0.99	0.79	0.98	0.85	1.46	0.95	0.94
	LSP103	0.95	1.00	0.95	0.89	0.99	0.79	0.98	0.85	1.46	0.95	0.94
	LS0063	1.04	1.00	1.01	1.10	1.08	0.84	1.07	1.06	0.91	1.06	1.02
	LS1063	0.99	1.07	0.96	0.92	0.93	0.85	1.01	0.92	1.31	0.97	0.97
	LSP063	0.99	1.03	0.96	0.92	0.93	0.85	1.01	0.92	1.31	0.97	0.97
	LSF0a	1.46	1.03	1.92	1.10	1.53	4.32	1.02	1.66	0.48	1.42	1.22
	LSF1a	1.19	1.24	1.30	0.92	0.85	0.75	1.25	1.13	1.46	1.31	1.29
	LSFPa	1.19	1.13	1.30	0.92	0.85	0.75	1.25	1.13	1.46	1.31	1.29
	LSF0b	1.08	1.13	1.16	1.21	1.41	0.93	1.11	0.95	1.74	1.27	1.07
	LSF1b	1.18	1.07	1.36	1.05	0.85	0.96	1.22	1.07	1.76	1.18	1.07
	LSFPb	1.18	1.08	1.36	1.05	0.85	0.96	1.22	1.07	1.76	1.18	1.07
	AN0203	1.22	1.08	1.29	1.26	1.75	1.28	1.20	1.15	1.51	1.44	3.51
	AN1203	1.30	1.14	1.64	1.01	1.31	0.71	1.39	1.19	1.82	1.91	1.41
	ANP203	1.30	1.14	1.64	1.01	1.31	0.71	1.39	1.19	1.82	1.91	1.41
	AN0213	1.63	1.14	2.21	2.71	5.25	1.88	1.59	1.35	2.96	1.68	1.21
	AN1213	1.55	1.28	1.89	1.28	1.43	1.92	1.49	1.61	1.28	1.78	1.60
	ANP213	1.55	1.28	1.89	1.28	1.43	1.92	1.49	1.61	1.28	1.78	1.60
	AN0223	1.13	1.28	1.29	1.47	1.40	1.00	1.15	0.94	2.07	0.97	1.39
	AN1223	1.81	1.07	2.55	3.15	1.42	1.06	1.92	1.73	2.18	0.99	1.92
	ANP223	1.81	1.36	2.55	3.15	1.42	1.06	1.92	1.73	2.18	1.38	1.92
	ANF0a	1.01	1.36	1.04	1.58	2.07	0.84	1.04	0.92	1.43	1.68	1.67
	ANF1a	1.87	1.02	2.49	2.00	1.30	1.67	1.90	1.86	1.91	2.41	2.48
	ANFPa	1.87	1.44	2.49	2.00	1.30	1.67	1.90	1.86	1.91	2.41	2.48
	ANF0b	1.13	1.44	1.32	1.13	1.08	0.77	1.19	0.94	2.08	1.34	1.40
	ANF1b	1.11	1.05	1.18	1.31	1.13	1.25	1.09	1.05	1.41	1.58	1.71
	ANFPb	1.11	1.09	1.18	1.31	1.13	1.25	1.09	1.05	1.41	1.58	1.71
C - Benchmark Loss (*100.000)		2.529	399.836	0.019	6.971	23.030	1.930	2.655	2.332	4.279	2.531	2.576

Notes:

See Table 1 for the definition of the models. Relative losses wrt ARFC04

Recursive: first estimation sample 1959-1979, forecast sample 1980-2004

Recursive split: first estimation sample 1959-1979, various forecast samples

Rolling: forecast sample 1980-2004, first estimation sample: 1970-1979 or 1965-1979.

Table 3: Forecasting US inflation

Relative to ARFC04 with I(1) CPI (applies to the I(2) columns)

		Recursive						Recursive Split				Rolling		I(2)			Estimation 84-93				No food and Energy					
		h=1			h=3	h=6	h=12	MSE, h=1				MSE, h=1		h=1			Forecasts 94-04				h=1			h=3	h=6	h=12
Col.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]	[24]	[25]	[26]
	MODEL	MSE	MAE	MACE	MSE			80-89	90-04	Reserves targeting	interest rate targeting	window 10 years	window 15 years	MSE	MAE	MACE	h=1	h=3	h=6	h=12	MSE	MAE	MACE	MSE		
A - Linear	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	0.91	0.94	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	ARFT04	1.03	1.04	1.01	1.14	1.33	1.71	0.99	1.07	0.94	1.05	1.05	1.04	0.97	1.00	1.06	1.02	1.05	1.12	1.07	1.02	1.05	0.96	1.09	1.32	1.93
	ARFC14	0.91	0.94	0.89	0.87	0.86	0.82	0.86	0.97	0.81	0.93	1.03	0.98	0.96	0.96	1.10	1.04	1.13	1.21	1.27	0.80	0.90	0.67	0.60	0.45	0.43
	ARFT14	0.99	1.00	0.97	0.99	1.02	1.13	1.00	0.98	0.94	1.00	1.00	0.99	0.97	0.96	1.11	1.01	1.01	1.06	1.11	1.00	1.01	0.96	0.97	1.06	1.39
	ARFCP4	0.91	0.94	0.89	0.87	0.86	0.82	0.86	0.97	0.81	0.93	1.03	0.98	0.94	0.95	1.05	1.04	1.13	1.21	1.27	0.80	0.90	0.67	0.60	0.45	0.43
	ARFTP4	0.91	0.94	0.89	0.87	0.86	0.82	0.86	0.97	0.81	0.93	1.03	0.98	0.96	0.96	1.10	1.04	1.13	1.21	1.27	0.80	0.90	0.67	0.60	0.45	0.43
	ARFC0a	0.95	0.96	0.95	0.85	0.80	0.78	0.92	0.98	0.89	0.97	1.00	0.98	0.91	0.93	1.03	1.02	1.01	1.04	1.05	0.93	0.95	0.94	0.79	0.70	0.68
	ARFT0a	0.97	0.99	0.96	0.95	1.03	1.35	0.92	1.03	0.86	1.00	1.06	1.02	0.94	0.96	1.06	1.04	1.05	1.16	1.15	0.95	0.99	0.89	0.87	0.96	1.51
	ARFC1a	0.91	0.93	0.92	0.78	0.72	0.73	0.87	0.94	0.83	0.92	1.01	0.97	0.92	0.93	1.06	1.05	1.11	1.12	1.27	0.79	0.88	0.71	0.56	0.42	0.38
	ARFT1a	0.95	0.96	0.96	0.85	0.79	0.96	0.94	0.95	0.91	0.95	1.00	0.97	0.93	0.92	1.07	1.02	1.03	1.05	1.22	0.92	0.94	0.94	0.80	0.83	1.15
	ARFCPa	0.91	0.93	0.92	0.78	0.72	0.73	0.87	0.94	0.83	0.92	1.01	0.97	0.92	0.93	1.05	1.05	1.11	1.12	1.27	0.79	0.88	0.71	0.56	0.42	0.38
	ARFTPa	0.91	0.93	0.92	0.78	0.72	0.73	0.87	0.94	0.83	0.92	1.01	0.97	0.92	0.93	1.06	1.05	1.11	1.12	1.27	0.79	0.88	0.71	0.56	0.42	0.38
	ARFC0b	0.95	0.96	0.95	0.85	0.80	0.78	0.92	0.98	0.89	0.97	1.04	0.99	0.91	0.93	1.04	1.02	0.99	1.03	0.99	0.98	0.97	1.00	0.82	0.70	0.68
	ARFT0b	1.00	1.00	1.00	0.95	1.03	1.35	0.96	1.03	0.99	1.00	1.08	1.04	0.96	0.97	1.09	1.04	1.05	1.23	1.10	1.00	1.02	0.96	0.89	0.96	1.53
	ARFC1b	0.92	0.93	0.93	0.78	0.72	0.73	0.88	0.94	0.83	0.93	1.05	0.98	0.94	0.94	1.10	1.06	1.16	1.24	1.39	0.81	0.88	0.73	0.56	0.42	0.38
	ARFT1b	0.98	0.98	0.99	0.85	0.79	0.97	1.00	0.95	1.03	0.96	1.04	1.00	0.94	0.94	1.09	1.02	1.00	1.12	1.09	0.97	0.98	0.96	0.86	0.84	1.24
	ARFCPb	0.92	0.93	0.93	0.78	0.72	0.73	0.88	0.94	0.83	0.93	1.05	0.98	0.93	0.94	1.07	1.06	1.16	1.24	1.39	0.81	0.88	0.73	0.56	0.42	0.38
	ARFTPb	0.92	0.93	0.93	0.78	0.72	0.73	0.88	0.94	0.83	0.93	1.05	0.98	0.94	0.94	1.10	1.06	1.16	1.24	1.39	0.81	0.88	0.73	0.56	0.42	0.38
	NOCHANGE	2.28	1.73	2.71	3.29	4.43	4.63	2.24	2.33	1.30	2.49	2.64	2.57	1.13	1.07	1.27	1.91	3.06	5.41	9.40	4.70	2.59	6.60	6.47	8.37	9.08
B - Time Varying and Nonlinear	ARTVFC03	0.82	0.89	0.74	0.67	0.60	0.57	0.78	0.85	0.82	0.81	0.86	0.84	0.83	0.91	0.84	1.00	1.03	1.32	1.92	0.80	0.87	0.75	0.59	0.50	0.27
	ARTVFC13	0.82	0.91	0.74	0.70	0.65	0.59	0.78	0.86	0.81	0.82	0.88	0.85	0.98	0.99	1.08	1.00	1.04	1.06	1.13	0.78	0.87	0.69	0.63	0.53	0.38
	ARTVFCP3	0.82	0.91	0.74	0.70	0.65	0.59	0.78	0.86	0.81	0.82	0.88	0.85	0.89	0.93	0.97	1.00	1.04	1.06	1.13	0.78	0.87	0.69	0.63	0.53	0.38
	ARTVFC0a	0.94	0.95	0.93	0.83	0.77	0.74	0.91	0.97	0.89	0.95	1.88	0.90	0.91	0.93	1.01	0.99	1.07	1.33	2.13	0.90	0.95	0.84	0.75	0.64	0.61
	ARTVFC1a	0.91	0.94	0.90	0.80	0.74	0.79	0.86	0.95	0.82	0.93	0.98	0.96	0.99	0.95	1.18	1.03	1.11	1.12	1.39	0.82	0.91	0.70	0.60	0.48	0.53
	ARTVFCPa	0.91	0.94	0.90	0.80	0.74	0.79	0.86	0.95	0.82	0.93	0.98	0.96	0.91	0.93	1.03	1.03	1.11	1.12	1.39	0.82	0.91	0.70	0.60	0.48	0.53
	ARTVFC0b	0.94	0.95	0.93	0.83	0.77	0.74	0.91	0.97	0.89	0.95	1.83	0.91	0.91	0.93	1.01	0.99	1.08	1.32	2.14	0.90	0.95	0.84	0.75	0.64	0.61
	ARTVFC1b	0.91	0.94	0.90	0.80	0.74	0.79	0.86	0.95	0.82	0.93	0.98	0.96	0.97	0.94	1.13	1.03	1.11	1.12	1.39	0.82	0.91	0.70	0.60	0.48	0.53
	ARTVFCPb	0.91	0.94	0.90	0.80	0.74	0.79	0.86	0.95	0.82	0.93	0.98	0.96	0.91	0.93	1.03	1.03	1.11	1.12	1.39	0.82	0.91	0.70	0.60	0.48	0.53
	LS0103	0.89	0.94	0.83	0.83	1.91	0.62	0.85	0.92	0.90	0.88	0.92	0.90	0.93	0.95	1.03	0.97	0.99	1.15	1.83	0.84	0.92	0.75	0.96	1.85	1.28
	LS1103	0.93	0.95	0.92	0.84	0.91	0.94	0.88	0.99	0.94	0.93	1.27	1.14	0.96	0.96	1.07	1.24	1.47	1.27	1.86	0.77	0.87	0.67	0.62	0.48	0.46
	LSP103	0.93	0.95	0.92	0.84	0.91	0.94	0.88	0.99	0.94	0.93	1.27	1.14	0.94	0.95	1.04	1.24	1.47	1.27	1.86	0.77	0.87	0.67	0.62	0.48	0.46
	LS0063	0.90	0.92	0.92	0.96	0.66	0.68	0.94	0.87	0.87	0.91	0.91	0.89	0.94	0.95	1.07	1.05	1.13	1.64	1.71	0.83	0.88	0.84	0.72	0.62	0.66
	LS1063	0.90	0.91	0.92	0.77	0.67	0.67	0.91	0.89	0.87	0.90	0.98	0.91	0.90	0.93	1.00	1.10	1.07	1.10	1.25	0.78	0.84	0.73	0.58	0.46	0.35
	LSP063	0.90	0.91	0.92	0.77	0.67	0.67	0.91	0.89	0.87	0.90	0.98	0.91	0.92	0.94	1.02	1.10	1.07	1.10	1.25	0.78	0.84	0.73	0.58	0.46	0.35
	LSF0a	0.94	0.97	0.88	1.16	1.60	1.49	0.92	0.96	0.86	0.96	0.98	0.95	0.99	0.95	1.29	1.03	1.05	1.13	2.35	0.87	0.88	1.00	1.20	1.42	1.09
	LSF1a	1.00	0.95	1.15	0.75	0.74	0.75	1.05	0.94	0.88	1.02	1.17	1.21	0.95	0.93	1.12	1.31	1.01	1.11	0.84	0.78	0.84	0.78	0.59	0.50	0.37
	LSFPa	1.00	0.95	1.15	0.75	0.74	0.75	1.05	0.94	0.88	1.02	1.17	1.21	0.91	0.93	1.02	1.31	1.01	1.11	0.84	0.78	0.84	0.78	0.59	0.50	0.37
	LSF0b	0.93	0.95	0.91	1.09	1.62	1.44	0.95	0.92	1.04	0.91	0.95	0.91	0.97	0.94	1.27	1.00	0.94	1.09	1.97	0.90	0.91	0.96	1.22	1.45	1.04
	LSF1b	0.84	0.89	0.81	0.73	0.67	0.67	0.85	0.84	0.82	0.85	1.09	0.97	0.96	0.94	1.14	1.24	1.05	1.01	0.84	0.78	0.84	0.73	0.59	0.49	0.37
	LSFPb	0.84	0.89	0.81	0.73	0.67	0.67	0.85	0.84	0.82	0.85	1.09	0.97	0.89	0.91	1.00	1.24	1.05	1.01	0.84	0.78	0.84	0.73	0.59	0.49	0.37
	AN0203	0.94	0.97	0.89	1.42	1.99	2.14	0.86	1.01	0.89	0.95	1.07	1.03	1.00	0.98	1.19	1.17	1.12	3.67	5.81	0.81	0.87	0.74	1.54	0.77	5.24
	AN1203	1.07	1.02	1.14	0.84	0.96	0.91	1.01	1.13	1.15	1.05	1.23	1.62	1.09	1.01	1.40	1.23	1.32	1.31	1.42	0.90	0.99	0.72	0.71	0.48	0.59
	ANP203	1.07	1.02	1.14	0.84	0.96	0.91	1.01	1.13	1.15	1.05	1.23	1.62	1.05	0.99	1.32	1.23	1.32	1.31	1.42	0.90	0.99	0.72	0.71	0.48	0.59
	AN0213	0.96	1.01	0.83	1.47	3.21	3.89	0.79	1.12	0.55	1.04	1.40	7.75	1.41	1.11	2.24	2.20	1.60	2.49	5.30	1.24	1.09	1.37	9.21	3.83	3.96
	AN1213	1.10	1.01	1.30	1.09	0.98	1.03	1.06	1.13	1.08	1.10	1.36	1.51	1.22	1.07	1.67	1.84	1.45	1.75	1.99	0.96	1.01	0.84	0.70	0.73	0.57
	ANP213	1.10	1.01	1.30	1.09	0.98	1.03	1.06	1.13	1.08	1.10	1.36	1.51	1.38	1.09	2.18	1.84	1.45	1.75	1.99	0.96	1.01	0.84	0.70	0.73	0.57
	AN0223	0.91	0.96	0.84	1.28	2.05	2.63	0.84	0.98	0.80	0.93	1.17	1.16	0.92	0.92	1.07	1.03	1.03	1.77	1.96	1.08	1.05	1.09	1.49	2.21	2.71
	AN1223	0.98	0.97	1.04	0.91	0.95	1.38	0.95	1.01	0.92	1.00	1.87	1.02	0.94	0.96	1.01	1.48	1.64	1.40	1.54	0.99	1.04	0.86	0.77	0.62	0.82
	ANP223	0.98	0.97	1.04	0.91	0.95	1.38	0.95	1.01	0.92	1.00	1.87	1.02	0.88	0.92	0.96	1.48	1.64	1.40	1.54	0.99	1.04	0.86	0.77	0.62	0.82
	ANF0a	1.02	1.04	0.97	1.34	2.00	2.99	0.87	1.17	0.78	1.07	1.24	1.20	1.14	1.04	1.49	1.10	1.51	1.98	3.71	1.02	0.97	1.09	1.26	1.65	2.85
	ANF1a	1.70	1.16	3.25	1.03	0.84	0.96	1.06	2.31	1.18	1.81	1.57	1.34	1.49	1.06	3.18	1.54	1.44	1.65	2.48	0.96	0.99	0.87	0.75	0.66	0.51
	ANFPa	1.70	1.16	3.25	1.03	0.84	0.96	1.06	2.31	1.18	1.81	1.57	1.34	1.36	1.06	2.63	1.54	1.44	1.65	2.48	0.96	0.99	0.87	0.75	0.66	0.51
	ANF0b	1.03	1.02	1.03	1.38	2.06	3.07	0.95	1.11	0.91	1.06	1.13	1.24	0.90	0.93	1.01	1.01	1.27	1.64	3.15	1.19	1.07	1.40	1.35	2.02	3.15
	ANF1b	1.00	0.97	1.12	0.96	0.81	1.05	1.02	0.99	0.67	1.07	1.56	1.78	0.98	0.97	1.11	1.10	1.21	1.44	2.45	0.97	1.02	0.84	0.65	0.48	0.34
	ANFPb	1.00	0.97	1.12	0.96	0.81	1.05	1.02	0.99	0.67	1.07	1.56	1.78	0.92	0.94	1.03	1.10	1.21	1.44	2.45	0.97	1.02	0.84	0.65	0.48	0.34
C - Benchmark Loss (*100.000)		0.473	160.940	0.002	2.264	6.174	21.843	0.657	0.374	2.371	0.407	0.383	0.393	0.473	160.940	0.002	0.393	1.492	3.011	6.927	0.213	109.387	0.001	1.252	3.611	12.905

Notes:
See Table 1 for the definition of the models. Relative losses wrt ARFC04
Recursive: first estimation sample 1959-1979, forecast sample 1980-2004
Recursive split: first estimation sample 1959-1979, various forecast samples
Rolling: forecast sample 1980-2004, first estimation sample: 1970-1979 or 1965-1979.
No food and energy: recursive, first estimation sample 1959-1979, forecast sample 1980-2004

Table 4: Evaluating the statistical significance of the forecast gains

		A - GDP Growth									B - Inflation											
			h=1			h=2			h=4			h=1			h=3			h=6			h=12	
	Col. [1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]
	MODEL	no bootstrap	5%	95%	no bootstrap	5%	95%	no bootstrap	5%	95%	No bootstrap	5%	95%	No bootstrap	5%	95%	No bootstrap	5%	95%	No bootstrap	5%	95%
A - Linear	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
	ARFT04	0.91	0.92	1.11	0.82	0.86	1.25	0.73	0.72	1.54	1.03	0.99	1.03	1.14	0.95	1.10	1.33	0.89	1.21	1.71	0.78	1.73
	ARFC14	0.95	0.92	1.05	0.91	0.85	1.08	0.91	0.69	1.15	0.91	0.96	1.01	0.87	0.90	1.02	0.86	0.85	1.03	0.82	0.78	1.07
	ARFT14	0.98	0.94	1.04	1.00	0.92	1.06	1.02	0.88	1.09	0.99	0.96	1.01	0.99	0.94	1.01	1.02	0.93	1.02	1.13	0.91	1.06
	ARFCP4	0.95	0.92	1.05	0.91	0.85	1.08	0.91	0.69	1.15	0.91	0.96	1.01	0.87	0.90	1.02	0.86	0.85	1.03	0.82	0.78	1.07
	ARFTP4	0.98	0.94	1.04	1.00	0.92	1.06	1.02	0.88	1.09	0.91	0.96	1.01	0.87	0.90	1.02	0.86	0.85	1.03	0.82	0.78	1.07
	ARFC0a	1.01	0.96	1.05	1.03	0.97	1.06	1.01	0.95	1.07	0.95	0.95	1.01	0.85	0.92	1.02	0.80	0.92	1.02	0.78	0.94	1.02
	ARFT0a	0.91	0.92	1.13	0.82	0.85	1.25	0.75	0.72	1.64	0.97	0.95	1.03	0.95	0.90	1.10	1.03	0.86	1.23	1.35	0.77	1.65
	ARFC1a	0.96	0.91	1.07	0.91	0.84	1.07	0.91	0.70	1.16	0.91	0.94	1.01	0.78	0.88	1.01	0.72	0.83	1.01	0.73	0.76	1.07
	ARFT1a	0.99	0.94	1.07	1.01	0.93	1.07	1.03	0.88	1.11	0.95	0.95	1.02	0.85	0.91	1.02	0.79	0.90	1.02	0.96	0.90	1.08
	ARFCPa	0.96	0.91	1.07	0.91	0.84	1.07	0.91	0.70	1.16	0.91	0.94	1.01	0.78	0.88	1.01	0.72	0.83	1.01	0.73	0.76	1.07
	ARFTPa	0.99	0.94	1.07	1.01	0.93	1.07	1.03	0.88	1.11	0.91	0.94	1.01	0.78	0.88	1.01	0.72	0.83	1.01	0.73	0.76	1.07
	ARFC0b	1.09	0.95	1.08	1.02	0.94	1.07	1.01	0.93	1.07	0.95	0.97	1.03	0.85	0.93	1.03	0.80	0.93	1.04	0.78	0.96	1.04
	ARFT0b	0.98	0.92	1.14	0.84	0.86	1.25	0.75	0.73	1.57	1.00	0.96	1.06	0.95	0.90	1.10	1.03	0.87	1.23	1.35	0.79	1.65
	ARFC1b	1.05	0.90	1.06	0.92	0.83	1.10	0.91	0.70	1.14	0.92	0.95	1.03	0.78	0.89	1.03	0.72	0.84	1.04	0.73	0.78	1.08
	ARFT1b	1.07	0.94	1.08	1.01	0.92	1.10	1.00	0.88	1.11	0.98	0.96	1.03	0.85	0.92	1.04	0.79	0.91	1.05	0.97	0.91	1.09
	ARFCPb	1.05	0.90	1.06	0.92	0.83	1.10	0.91	0.70	1.14	0.92	0.95	1.03	0.78	0.89	1.03	0.72	0.84	1.04	0.73	0.78	1.08
	ARFTPb	1.07	0.94	1.08	1.01	0.92	1.10	1.00	0.88	1.11	0.92	0.95	1.03	0.78	0.89	1.03	0.72	0.84	1.04	0.73	0.78	1.08
	NOCHANGE	3.33	1.35	2.90	4.30	1.54	4.01	4.47	1.70	6.07	2.28	2.33	5.96	3.29	3.13	7.02	4.43	3.63	7.19	4.63	3.27	7.47
B- Time Varying and Nonlinear	ARTVFC03	1.03	0.95	1.13	1.01	0.93	1.18	1.10	0.87	1.28	0.82	0.94	1.08	0.67	0.89	1.38	0.60	0.94	1.58	0.57	0.91	1.90
	ARTVFC13	1.00	0.92	1.12	0.94	0.87	1.11	0.97	0.77	1.14	0.82	0.95	1.11	0.70	0.91	1.41	0.65	0.87	1.54	0.59	0.84	1.63
	ARTVFCP3	1.00	0.92	1.12	0.94	0.87	1.11	0.97	0.77	1.14	0.82	0.95	1.11	0.70	0.91	1.41	0.65	0.87	1.54	0.59	0.84	1.63
	ARTVFC0a	1.13	0.94	1.26	1.15	0.92	1.35	1.29	0.84	1.47	0.94	0.99	1.42	0.83	0.94	1.74	0.77	0.95	1.98	0.74	0.93	1.90
	ARTVFC1a	1.03	0.95	1.25	1.04	0.91	1.34	0.90	0.77	1.45	0.91	0.95	1.16	0.80	0.90	1.45	0.74	0.85	1.65	0.79	0.79	1.42
	ARTVFCPa	1.03	0.95	1.25	1.04	0.91	1.34	0.90	0.77	1.45	0.91	0.95	1.16	0.80	0.90	1.45	0.74	0.85	1.65	0.79	0.79	1.42
	ARTVFC0b	1.13	0.94	1.26	1.15	0.92	1.35	1.23	0.84	1.43	0.94	0.99	1.41	0.83	0.94	1.75	0.77	0.94	2.00	0.74	0.93	1.93
	ARTVFC1b	1.03	0.94	1.15	0.97	0.89	1.22	0.86	0.76	1.29	0.91	0.95	1.15	0.80	0.90	1.42	0.74	0.85	1.66	0.79	0.80	1.40
	ARTVFCPb	1.03	0.94	1.15	0.97	0.89	1.22	0.86	0.76	1.29	0.91	0.95	1.15	0.80	0.90	1.42	0.74	0.85	1.66	0.79	0.80	1.40
	LS0103	1.07	0.97	1.36	0.95	0.94	1.45	0.96	0.89	1.96	0.89	0.99	1.26	0.83	0.99	1.88	1.91	0.96	2.07	0.62	0.99	2.31
	LS1103	0.95	0.87	1.19	0.89	0.81	1.21	0.99	0.71	1.24	0.93	0.90	1.11	0.84	0.86	1.12	0.91	0.85	1.29	0.94	0.79	1.20
	LSP103	0.95	0.87	1.19	0.89	0.81	1.21	0.99	0.71	1.24	0.93	0.90	1.11	0.84	0.86	1.12	0.91	0.85	1.29	0.94	0.79	1.20
	LS0063	1.04	0.96	1.24	1.10	0.95	1.41	1.08	0.93	1.62	0.90	0.92	1.07	0.96	0.88	1.39	0.66	0.95	1.92	0.68	0.98	2.13
	LS1063	0.99	0.91	1.18	0.92	0.85	1.23	0.93	0.74	1.24	0.90	0.91	1.03	0.77	0.84	1.11	0.67	0.81	1.18	0.67	0.75	1.19
	LSP063	0.99	0.91	1.18	0.92	0.85	1.23	0.93	0.74	1.24	0.90	0.91	1.03	0.77	0.84	1.11	0.67	0.81	1.18	0.67	0.75	1.19
	LSF0a	1.46	0.98	1.52	1.10	0.98	1.85	1.53	0.94	2.79	0.94	0.95	1.29	1.16	0.94	2.12	1.60	0.97	2.52	1.49	0.99	3.14
	LSF1a	1.19	0.91	1.38	0.92	0.87	1.44	0.85	0.75	1.46	1.00	0.92	1.22	0.75	0.85	1.30	0.74	0.82	1.73	0.75	0.79	1.47
	LSFPa	1.19	0.91	1.38	0.92	0.87	1.44	0.85	0.75	1.46	1.00	0.92	1.22	0.75	0.85	1.30	0.74	0.82	1.73	0.75	0.79	1.47
	LSF0b	1.08	0.96	1.41	1.21	0.97	1.72	1.41	0.89	2.59	0.93	0.94	1.21	1.09	0.98	1.87	1.62	1.01	2.76	1.44	0.99	3.24
	LSF1b	1.18	0.89	1.25	1.05	0.84	1.30	0.85	0.72	1.42	0.84	0.92	1.12	0.73	0.86	1.24	0.67	0.81	1.35	0.67	0.77	1.41
	LSFPb	1.18	0.89	1.25	1.05	0.84	1.30	0.85	0.72	1.42	0.84	0.92	1.12	0.73	0.86	1.24	0.67	0.81	1.35	0.67	0.77	1.41
	AN0203	1.22	1.01	1.63	1.26	1.02	2.26	1.75	1.00	3.46	0.94	1.01	3.20	1.42	1.06	3.94	1.99	1.17	4.16	2.14	1.15	5.24
	AN1203	1.30	0.89	1.37	1.01	0.80	1.29	1.31	0.71	1.29	1.07	0.92	1.14	0.84	0.88	1.21	0.96	0.85	1.26	0.91	0.78	1.32
	ANP203	1.30	0.89	1.37	1.01	0.80	1.29	1.31	0.71	1.29	1.07	0.92	1.14	0.84	0.88	1.21	0.96	0.85	1.26	0.91	0.78	1.32
	AN0213	1.63	1.06	1.90	2.71	1.07	2.80	5.25	1.30	5.46	0.96	1.09	3.73	1.47	1.29	4.44	3.21	1.53	5.85	3.89	1.71	9.26
	AN1213	1.55	0.87	1.57	1.28	0.87	1.47	1.43	0.77	1.57	1.10	0.93	1.32	1.09	0.88	1.30	0.98	0.87	1.40	1.03	0.80	1.42
	ANP213	1.55	0.87	1.57	1.28	0.87	1.47	1.43	0.77	1.57	1.10	0.93	1.32	1.09	0.88	1.30	0.98	0.87	1.40	1.03	0.80	1.42
	AN0223	1.13	1.02	1.46	1.47	1.03	1.97	1.40	1.05	3.18	0.91	1.02	1.58	1.28	1.06	2.86	2.05	1.18	4.31	2.63	1.11	5.13
	AN1223	1.81	0.86	1.47	3.15	0.86	1.47	1.42	0.73	1.51	0.98	0.89	1.23	0.91	0.87	1.24	0.95	0.85	1.28	1.38	0.82	1.40
	ANP223	1.81	0.86	1.47	3.15	0.86	1.47	1.42	0.73	1.51	0.98	0.89	1.23	0.91	0.87	1.24	0.95	0.85	1.28	1.38	0.82	1.40
	ANF0a	1.01	1.01	1.59	1.58	1.03	2.41	2.07	1.14	4.33	1.02	1.02	1.52	1.34	1.10	3.32	2.00	1.14	4.40	2.99	1.38	6.02
	ANF1a	1.87	0.97	1.70	2.00	0.93	1.84	1.30	0.80	1.81	1.70	0.90	1.47	1.03	0.91	1.49	0.84	0.90	1.54	0.96	0.82	1.63
	ANFPa	1.87	0.97	1.70	2.00	0.93	1.84	1.30	0.80	1.81	1.70	0.90	1.47	1.03	0.91	1.49	0.84	0.90	1.54	0.96	0.82	1.63
	ANF0b	1.13	0.96	1.36	1.13	0.97	1.78	1.08	0.97	3.21	1.03	0.99	1.28	1.38	1.08	3.07	2.06	1.16	4.26	3.07	1.35	6.27
	ANF1b	1.11	0.87	1.36	1.31	0.85	1.27	1.13	0.72	1.26	1.00	0.92	1.25	0.96	0.87	1.28	0.81	0.84	1.39	1.05	0.82	1.50
	ANFPb	1.11	0.87	1.36	1.31	0.85	1.27	1.13	0.72	1.26	1.00	0.92	1.25	0.96	0.87	1.28	0.81	0.84	1.39	1.05	0.82	1.50

Notes:

See Table 1 for the definition of the models. Relative losses wrt ARFC04

The intervals are based on the empirical distribution of the relative losses over 200 (non parametric) bootstrap replications using recursive estimation with first estimation sample 1959-1979, forecast sample 1980-2004

Table 5: Evaluating the economic significance of the forecast gains

		A - Growth								B - Growth				C - Inflation		D - Inflation	
		85-99				71-84				90-01				70-83	84-96	70-98	
Col.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[16]	[17]	[14]	[15]
	MODEL	h=1	h=2	h=4	h=8	h=1	h=2	h=4	h=8	h=1	h=4	h=8	h=12	h=12	h=12	h=6	h=12
A - Linear	**ARFC11**	-	-	-	-	-	-	-	-	1.00	1.00	1.00	1.00	-	-	-	-
	ARFC04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	-	-	-	-	3.92	1.30	1.74	2.36
	ARFT04	0.75	0.78	0.58	0.38	1.00	0.95	0.93	0.36	0.86	0.80	0.67	0.54	1.63	2.04	1.48	1.38
	ARFC14	0.78	0.83	0.66	0.38	1.02	0.97	0.94	0.45	0.91	0.99	1.01	1.00	2.54	0.97	1.48	1.56
	ARFT14	0.80	1.01	1.02	0.95	1.06	1.04	1.08	1.08	0.90	1.10	1.36	1.91	1.56	1.54	1.26	1.19
	ARFCP4	0.78	0.83	0.66	0.38	1.02	0.97	0.94	0.45	0.91	0.99	1.01	1.00	2.54	0.97	1.48	1.56
	ARFTP4	0.80	1.01	1.02	0.95	1.06	1.04	1.08	1.08	0.90	1.10	1.36	1.91	2.54	0.97	1.48	1.56
	ARFC0a	0.83	1.03	1.23	1.00	1.08	1.02	1.07	0.96	0.91	1.07	1.39	2.09	4.70	0.97	1.52	2.68
	ARFT0a	0.75	0.78	0.60	0.40	1.03	0.92	0.92	0.40	0.86	0.81	0.68	0.53	1.58	1.54	1.25	1.23
	ARFC1a	0.77	0.82	0.65	0.37	1.08	1.00	1.03	0.46	0.91	1.00	1.03	1.00	3.45	0.84	1.68	2.00
	ARFT1a	0.81	0.99	1.21	0.90	1.11	1.11	1.20	1.00	0.91	1.12	1.33	1.97	1.55	1.27	1.17	1.12
	ARFCPa	0.77	0.82	0.65	0.37	1.08	1.00	1.03	0.46	0.91	1.00	1.03	1.00	3.45	0.84	1.68	2.00
	ARFTPa	0.81	0.99	1.21	0.90	1.11	1.11	1.20	1.00	0.91	1.12	1.33	1.97	3.45	0.84	1.68	2.00
	ARFC0b	0.89	1.03	0.97	1.00	0.98	0.96	1.05	1.00	1.00	1.08	1.39	1.88	4.83	0.97	1.52	2.75
	ARFT0b	0.83	0.79	0.60	0.43	0.97	0.92	0.92	0.41	0.91	0.81	0.73	0.62	1.62	1.54	1.25	1.25
	ARFC1b	0.83	0.83	0.65	0.37	0.99	0.92	0.90	0.45	1.00	1.00	1.03	1.00	3.04	0.84	1.68	1.79
	ARFT1b	0.87	0.99	0.91	0.90	1.04	0.99	1.13	1.09	1.00	1.09	1.33	1.78	1.61	1.30	1.19	1.15
	ARFCPb	0.83	0.83	0.65	0.37	0.99	0.92	0.90	0.45	1.00	1.00	1.03	1.00	3.04	0.84	1.68	1.79
	ARFTPb	0.87	0.99	0.91	0.90	1.04	0.99	1.13	1.09	1.00	1.09	1.33	1.78	3.04	0.84	1.68	1.79
	NOCHANGE	3.07	5.00	4.76	5.08	1.23	1.34	1.30	1.80	2.71	4.31	6.54	10.95	11.42	6.35	9.32	7.47
B - Time Varying and Nonlinear	**ARTVFC03**	0.83	0.90	0.81	0.74	1.06	1.14	1.29	1.14	0.93	1.21	1.60	2.06	4.71	0.71	1.98	2.62
	ARTVFC13	0.81	0.84	0.71	0.52	1.05	1.05	1.04	0.64	0.92	1.07	1.19	1.28	3.24	0.75	1.87	1.88
	ARTVFCP3	0.81	0.84	0.71	0.52	1.05	1.05	1.04	0.64	0.92	1.07	1.19	1.28	3.24	0.75	1.87	1.88
	ARTVFC0a	0.91	1.03	0.93	0.89	1.22	1.39	1.57	1.59	0.98	1.40	1.75	2.54	4.47	0.92	2.05	2.55
	ARTVFC1a	0.83	1.11	0.99	0.53	1.09	1.08	0.98	0.55	0.93	0.97	0.99	0.96	3.34	0.92	2.90	1.97
	ARTVFCPa	0.83	1.11	0.99	0.53	1.09	1.08	0.98	0.55	0.93	0.97	0.99	0.96	3.34	0.92	2.90	1.97
	ARTVFC0b	0.91	1.03	0.88	0.85	1.22	1.39	1.56	1.59	0.98	1.32	1.72	2.62	4.57	0.92	2.12	2.60
	ARTVFC1b	0.86	1.00	0.89	0.53	1.09	1.08	0.98	0.55	0.96	0.93	0.99	0.96	3.13	0.92	2.90	1.86
	ARTVFCPb	0.86	1.00	0.89	0.53	1.09	1.08	0.98	0.55	0.96	0.93	0.99	0.96	3.13	0.92	2.90	1.86
	LS0103	0.88	1.19	0.76	0.84	1.90	1.86	1.42	2.72	1.13	1.07	1.48	1.63	5.76	1.61	3.44	3.37
	LS1103	0.78	0.84	0.69	0.43	1.08	1.08	0.93	0.39	0.97	1.09	1.23	1.01	2.94	1.12	2.01	1.81
	LSP103	0.78	0.84	0.69	0.43	1.08	1.08	0.93	0.39	0.97	1.09	1.23	1.01	2.94	1.12	2.01	1.81
	LS0063	0.83	1.26	1.28	1.14	0.97	0.97	1.18	1.19	0.90	1.15	1.15	2.43	2.68	0.88	1.71	1.63
	LS1063	0.80	0.90	0.69	0.37	1.02	1.14	0.95	0.48	0.86	1.01	1.05	1.04	1.32	0.85	1.93	0.90
	LSP063	0.80	0.90	0.69	0.37	1.02	1.14	0.95	0.48	0.86	1.01	1.05	1.04	1.32	0.85	1.93	0.90
	LSF0a	1.47	1.26	2.03	2.16	1.42	2.41	3.03	1.13	0.95	1.70	2.71	3.81	7.37	2.62	6.87	4.46
	LSF1a	0.98	1.01	0.76	0.75	1.40	1.45	1.38	0.75	1.08	0.92	1.25	1.00	2.57	0.97	3.51	1.57
	LSFPa	0.98	1.01	0.76	0.75	1.40	1.45	1.38	0.75	1.08	0.92	1.25	1.00	2.57	0.97	3.51	1.57
	LSF0b	0.87	1.19	1.73	2.28	1.39	2.00	3.15	1.13	1.01	1.61	2.71	3.37	5.86	2.18	6.53	3.57
	LSF1b	0.98	0.99	0.67	0.75	1.28	1.11	1.57	0.35	1.08	0.92	1.17	0.98	2.34	0.84	3.58	1.43
	LSFPb	0.98	0.99	0.67	0.75	1.28	1.11	1.57	0.35	1.08	0.92	1.17	0.98	2.34	0.84	3.58	1.43
	AN0203	0.93	1.48	2.39	1.85	1.15	1.68	1.83	2.75	1.15	1.98	4.53	1.47	7.91	5.22	6.54	5.41
	AN1203	1.07	1.48	1.04	0.49	1.49	1.36	1.36	0.44	1.58	1.39	1.21	1.07	2.75	1.14	2.65	1.72
	ANP203	1.07	1.48	1.04	0.49	1.49	1.36	1.36	0.44	1.58	1.39	1.21	1.07	2.75	1.14	2.65	1.72
	AN0213	1.28	2.73	4.18	8.52	1.61	2.58	1.90	2.53	1.43	5.28	3.26	9.71	13.12	3.96	6.89	7.76
	AN1213	1.43	1.90	0.78	0.52	1.83	1.81	1.23	1.04	3.03	1.44	1.26	1.22	3.04	1.49	2.57	1.95
	ANP213	1.43	1.90	0.78	0.52	1.83	1.81	1.23	1.04	3.03	1.44	1.26	1.22	3.04	1.49	2.57	1.95
	AN0223	0.93	1.24	2.24	0.65	1.77	1.52	1.56	1.09	0.95	1.47	1.58	5.60	11.42	3.56	5.54	6.77
	AN1223	1.75	1.61	1.15	0.43	1.16	0.96	1.12	0.54	1.38	1.38	1.27	1.35	2.41	1.18	1.98	1.56
	ANP223	1.75	1.61	1.15	0.43	1.16	0.96	1.12	0.54	1.38	1.38	1.27	1.35	2.41	1.18	1.98	1.56
	ANF0a	0.80	1.30	3.28	1.01	1.95	2.09	1.83	2.43	0.95	2.04	6.71	3.48	12.92	3.65	7.40	7.59
	ANF1a	1.76	2.42	2.23	0.69	1.70	1.32	1.02	0.73	1.73	1.38	1.06	1.99	3.25	1.19	2.98	1.99
	ANFPa	1.76	2.42	2.23	0.69	1.70	1.32	1.02	0.73	1.73	1.38	1.06	1.99	3.25	1.19	2.98	1.99
	ANF0b	0.92	1.11	1.31	0.99	1.26	2.18	1.73	2.17	1.14	1.21	3.54	3.90	12.92	3.08	7.73	7.46
	ANF1b	0.92	1.59	1.71	0.43	1.26	1.08	1.39	0.83	1.00	1.19	1.09	1.02	2.80	1.38	3.10	1.79
	ANFPb	0.92	1.59	1.71	0.43	1.26	1.08	1.39	0.83	1.00	1.19	1.09	1.02	2.80	1.38	3.10	1.79
Benchmark Loss		2.71	44.83	133.55	95.84	17.03	6.19	24.67	470.10	3.04	24.38	60.56	83.10	22.57	2.55	1.74	10.40
	*** 100.000**																

Notes:
See Table 1 for the definition of the models. Relative losses wrt ARFC04 for A, to or ARFC11 for B, to Stock and Watson's benchmark for C and D.
Different estimation and forecast samples for comparison with:
A: Stock and Watson (2003) B: Ang, Piazzesi, Wei (2004)
C: Stock and Watson (1999a) D: Stock and Watson (2002)

Table 6: Real Time Analysis

		GDP			CPI			
		h=1	h=2	h=4	h=1	h=3	h=6	h=12
Col.	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
A-Linear	**ARFC04**	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	ARFT04	0.77	0.64	0.45	1.00	1.23	1.72	2.95
	ARFC14	0.87	0.74	0.54	1.03	1.18	1.41	1.89
	ARFT14	0.89	0.81	0.72	1.00	0.99	1.01	1.10
	ARFCP4	0.87	0.74	0.54	1.03	1.18	1.41	1.89
	ARFTP4	0.87	0.74	0.54	1.03	1.18	1.41	1.89
	ARFC0a	0.99	0.96	1.00	1.03	1.07	1.01	1.00
	ARFT0a	0.82	0.65	0.47	1.03	1.32	1.73	2.95
	ARFC1a	0.86	0.73	0.53	1.06	1.30	1.52	1.99
	ARFT1a	0.92	0.86	0.74	1.04	1.16	1.10	1.10
	ARFCPa	0.86	0.73	0.53	1.06	1.30	1.52	1.99
	ARFTPa	0.86	0.73	0.53	1.06	1.30	1.52	1.99
	ARFC0b	1.09	0.96	1.02	1.08	1.07	0.97	1.01
	ARFT0b	0.85	0.68	0.47	1.09	1.34	1.81	2.92
	ARFC1b	0.90	0.72	0.51	1.05	1.24	1.68	2.17
	ARFT1b	0.94	0.82	0.71	1.10	1.09	1.11	1.14
	ARFCPb	0.90	0.72	0.51	1.05	1.24	1.68	2.17
	ARFTPb	0.90	0.72	0.51	1.05	1.24	1.68	2.17
	NOCHANGE	3.20	4.06	4.08	1.92	4.08	6.45	9.89
B-Time Varying and Nonlinear	**ARTVFC03**	1.14	1.30	1.45	0.99	1.06	1.28	1.34
	ARTVFC13	1.04	0.96	0.77	0.99	1.16	1.42	1.84
	ARTVFCP3	1.04	0.96	0.77	0.99	1.16	1.42	1.84
	ARTVFC0a	1.31	1.50	1.57	0.93	1.08	1.33	1.44
	ARTVFC1a	1.09	1.14	0.84	1.04	1.33	1.49	1.82
	ARTVFCPa	1.09	1.14	0.84	1.04	1.33	1.49	1.82
	ARTVFC0b	1.31	1.50	1.57	0.93	1.07	1.32	1.43
	ARTVFC1b	1.09	1.14	0.84	1.04	1.33	1.48	1.74
	ARTVFCPb	1.09	1.14	0.84	1.04	1.33	1.48	1.74
	LS0103	1.16	1.84	3.91	1.15	1.20	1.54	1.87
	LS1103	0.90	0.94	0.63	1.01	1.38	1.79	1.96
	LSP103	0.90	0.94	0.63	1.01	1.38	1.79	1.96
	LS0063	1.99	1.28	1.54	1.20	1.72	1.87	2.00
	LS1063	1.15	1.01	0.80	1.05	1.30	1.65	2.43
	LSP063	1.15	1.01	0.80	1.05	1.30	1.65	2.43
	LSF0a	2.14	2.96	2.92	1.04	1.17	1.52	2.02
	LSF1a	1.15	1.18	0.65	0.98	1.07	1.63	1.48
	LSFPa	1.15	1.18	0.65	0.98	1.07	1.63	1.48
	LSF0b	1.30	2.35	2.34	1.08	1.22	1.50	1.90
	LSF1b	0.95	0.82	0.61	1.05	1.17	1.53	1.41
	LSFPb	0.95	0.82	0.61	1.05	1.17	1.53	1.41
	AN0203	1.59	2.36	3.21	1.17	1.82	2.96	6.15
	AN1203	1.35	1.23	0.66	1.47	1.86	2.08	2.31
	ANP203	1.35	1.23	0.66	1.47	1.86	2.08	2.31
	AN0213	1.99	3.26	4.49	1.91	3.05	3.08	7.03
	AN1213	1.53	1.77	1.44	1.82	1.74	2.43	2.93
	ANP213	1.53	1.77	1.44	1.82	1.74	2.43	2.93
	AN0223	1.34	1.98	3.32	1.53	1.47	3.66	6.55
	AN1223	1.57	1.40	1.23	1.08	1.60	1.81	2.76
	ANP223	1.57	1.40	1.23	1.08	1.60	1.81	2.76
	ANF0a	1.97	2.50	3.77	1.29	1.65	3.22	6.15
	ANF1a	2.47	2.27	1.28	1.54	1.24	2.02	2.57
	ANFPa	2.47	2.27	1.28	1.54	1.24	2.02	2.57
	ANF0b	1.38	1.78	3.40	1.28	1.60	3.07	5.93
	ANF1b	1.75	1.63	0.66	1.10	1.46	2.29	2.79
	ANFPb	1.75	1.63	0.66	1.10	1.46	2.29	2.79
Benchmark Loss		2.49	7.11	26.33	0.47	1.18	2.66	6.39
	*100000							

Notes:

See Table 1 for the definition of the models. Relative losses wrt ARFC04

Rolling estimation with 15 year window. Forecast sample is 1994-2004.