Information combination and forecast (st)ability. Evidence from vintages of time-series data*

Carlo Altavilla†
University of Naples "Parthenope"

Matteo Ciccarelli‡
European Central Bank

Abstract

This paper explores the role of model and vintage combination in forecasting, with a novel approach that exploits the information contained in the revision history of a given variable. We analyse the forecast performance of eleven widely used models to predict inflation and GDP growth, in the three dimensions of accuracy, uncertainty and stability by using the real-time data set for macroeconomists developed at the Federal Reserve Bank of Philadelphia. Instead of following the common practice of investigating only the relationship between first available and fully revised data, we analyse the entire revision history for each variable and extract a signal from the entire distribution of vintages of a given variable to improve forecast accuracy and precision. The novelty of our study relies on the interpretation of the vintages of a real time data base as related realizations or units of a panel data set. The results suggest that imposing appropriate weights on competing models of inflation forecasts and output growth – reflecting the relative ability each model has over different sub-sample periods – substantially increases the forecast performance. More interestingly, our results indicate that augmenting the information set with a signal extracted from all available vintages of time-series consistently leads to a substantial improvement in forecast accuracy, precision and stability.

JEL Classification Codes: C32, C33, C53
Keywords: Real-time data, Forecast Combination, Data and Model Uncertainty

*We are grateful to seminar participants at Columbia University and the 3rd Annual Workshop on "Macroeconomic Forecasting, Analysis and Policy with Data Revision", Cirano (Montreal). Part of the paper was written while the first author was visiting Columbia Business School, whose hospitality is gratefully acknowledged. The views expressed in the paper are those of the authors and do not necessarily reflect the views or the policy of the European Central Bank.

†University of Naples "Parthenope", Via Medina, 40 - 80133 Naples (Italy). E-mail: altavilla@uniparthenope.it; Phone: (+)39 0815474733, fax (+)39 0815474750.

‡Corresponding author. European Central Bank, Kaiserstrasse 29, - 60311 Frankfurt am Main (Germany). E-mail: matteo.ciccarelli@ecb.int; Phone: (+)49 6913448721, fax (+)49 6913446575.

Non-techincal summary

This paper examines the degree of forecast stability, accuracy and precision of a battery of different models which use information contained in vintages of time series data. The main idea is that the combination of vintages and of different forecasts obtained from several models reduce the mismeasurement risks associated with the real-time forecast. Moreover we argue that by imposing appropriate weights on competing forecasting models –reflecting the relative ability each model have over different sub-sample periods– and on different vintages –reflecting the revision process– it is possible to significantly dampen model and data uncertainty. On this basis, we suggest therefore a real-time procedure to combine model and vintages which should help in solving the problems associated with the evaluation of a forecasting model in real-time.

Our contribution relies on the interpretation of the vintages of a real time data base as related realizations of a stochastic process defining a given variable. We exploit two ways (model and vintage combination) in which information can be combined and test whether their joint use improves forecast ability, and focus on the performance of several competing models to forecast GDP growth and GDP deflator inflation.

The combination of forecast is computed using three procedures: a naive scheme with equal weights; an algorithm (AFTER) where the weights are assigned on the base of the past performance history of models; and a new methodology which modifies AFTER, and whose weights are updated after each additional observation to target each time the performance of the best candidate model, defined in terms of an appropriate loss function.

The combination of vintages is obtained in two different ways. First, we exploit information in vintages by using a panel data approach and considering all vintages as the units of the panel. The parameters of the models are estimated by pooling all vintages and exploiting the crucial characteristic of a panel data set that contains repeated observations on the same unit (vintage). Therefore, like panel data the vintage structure of a real-time data set can provide a situation very much close to a controlled experiment. The idea is appealing not only for a forecast evaluation on past realization, where blocks of vintages have the same length, but also in real time, where necessarily different past vintages have different length of time and the panel is unbalanced.

Second, we model the covariability of the vintages in terms of one unobserved common component and an idiosyncratic error term. The common factor is estimated using principal components of several variables over a predefined block of vintages and each model is then augmented with the common factor, and a linear or non linear relationship between the variable to be forecast and the factor is estimated.

Several measures of instability, accuracy and precision are computed for each model in a standard forecasting exercise, where, using every vintage available, a technique similar to the repeated observation forecasting of Stark and Crushore (2002) is employed to exploit the fact that the forecasts for a particular date change as vintage changes.

The results suggest that model and vintage combination might significantly improve the forecast

ability of the models in the three dimensions of accuracy, precision and stability. Some robustness analysis also indicate that, when changing the samples or the actual values used to construct forecast errors, results remain substantially unchanged. Both model and vintage combination should therefore be employed as a standard forecasting procedure. The real-time implications point in the direction of using all information contained in the whole revision history of a variable to forecast it or to measure the associated forecast uncertainty.

Finally, we have compared the forecast performance of the selected models before and after the onset of the Great Moderation. Overall, the results indicate that starting form the mid-80's there has been an increase in predictability, a reduction in both predicted and actual forecast uncertainty, and a rise in the responsiveness of forecast revision to data revision. Understanding how these factors (especially a change in the revision process) contribute to the observed decline in macroeconomic volatility suggests a natural direction for future work.

1 Introduction

Several authors have shown that, depending on the case at hand, data revisions may matter for forecasting. This statement is valid in an ex-ante and an ex-post sense. From an ex-ante point of view, a certain degree of uncertainty is inevitably associated with the forecasts in that (i) forecasts differ, for instance, when using first-available or latest-available data; (ii) these effects vary depending on whether the variables are forecast in level or growth rates; (iii) model selection is sensitive to data revisions; and (iv) the predictive content of variables varies if the variables are subject to revisions (see e.g. Croushore, 2006). From an ex-post point of view, any forecast-model evaluation should in principle be conducted in real-time, precisely because data revisions might have affected the goodness of a given model. Stark and Croushore (2002) show that measures of forecast errors can be lower when using latest-available data rather than real-time data. Therefore "comparisons between the forecasts generated from new model and benchmark forecasts should be based on real-time data". Others (e.g. Koenig et al. 2003) argue, instead, that this approach should generally be avoided and analysis should use data of as many different vintages as there are dates in their sample. Common practice is to gauge forecasting accuracy on the same vintages of data that have been used to build the model.

The ex-ante/ex-post problems run clearly in circle, because the use of real-time data involves a risk of mismeasurement which can then be translated into non optimal choices by policy-makers. Given that data are subject to revision and that revisions affect the forecast and the model evaluation, forecasters should optimally account for the data revision when building and evaluating their models.

In this paper we deal with these issues and examine the degree of forecast stability, accuracy and precision, using a battery of different models which use information contained in vintages of time series data. We check whether both the combination of vintages and the combination of different forecasts obtained from several models reduce the mismeasurement risks associated with the real-time forecast. Our prior is that it should. Moreover, a priori one can also argue that imposing appropriate weights on competing forecasting models –reflecting the relative ability each model have over different sub-sample periods– and on different vintages –reflecting the revision process– should significantly dampen model and data uncertainty. On this basis, one can think of a real-time procedure to combine model and vintages which should help in breaking the circle between the ex-ante and the ex-post problems mentioned above.

The idea of information combination is certainly not new in the literature. The novelty of our paper relies on the interpretation of the vintages of a real time data base as related realizations of a stochastic process defining a given variable. As said, we exploit two ways (model and vintage combination) in which information can be combined and test whether their joint use improves forecast ability.

One of the main justifications for the use of model combination is that forecasts based on a given model may have a high variability if the model has been somehow selected, in the sense that a slight change of the data may result in the choice of a different model. The idea of model combination is that with an appropriate weighting scheme the combined forecast has a smaller variability and the forecast accuracy can be improved relative to the use of a selection criterion. Typically, all exercises on model combination are conducted using only one set of observations for a given variable – either the latest available or the real-time data – with rolling or recursive estimation techniques, and data are perturbed using simulated errors with certain characteristics to simulate forecast and model instability and gauge how model combination performs. The advantage of a real-time data set is that we dispose of such perturbations in a natural way and can solve the problem of irreproducibility of empirical research considering different vintages of data of a given variable as perturbations. The exercise is not a mere speculative one, but has important practical implications because it can show that a revision of data may change the model selection and therefore render the forecast much more unstable. The exercise is also particularly important for those institutions that use always the same model and revise it only with a lower frequency than the data revision.

On the other hand, the existence of multiple vintages of data for a given variables might render incorrect the use of a single vintage when evaluating a model because the stochastic relationship between vintages is not taken into account and therefore the estimated uncertainty is distorted. To avoid a vintage dependence of the forecast accuracy we propose combining the information contained in several (if not all) vintages. Revisions do not seem to be well behaved, for they are large compared to the original variable, they do not have a zero mean, and they are predictable (e.g. Aruoba, 2005). Moreover, the extensive and not necessarily systematic nature of the revision process might make unfeasible the incorporation of the revision's features in a given model. Hence the idea of "extracting a signal" from the entire distribution of vintages of a given variable and using this signal to improve forecast accuracy and precision. This approach shares the same view of previous studies. Guerrero (1993), for instance, proposes to combine historical and preliminary information to obtain timely time series data, using simple regression models that link preliminary and final data. Translated into the language of a real time data-set, the approach relates the final column of data (the latest-available) with the diagonal (real-time data), but disregards all the revision process incorporated in the vintages. Patterson (2003) combines the data generation process and the data measurement process with a nesting model that "comprises the links amongst generic variables and the links within data vintages on the same variable and across variables", and extracts a common trend for each variable and then check whether the common trends cointegrate. The analysis is therefore performed on levels and becomes unfeasible when the number of vintages or variable increases. Moreover, being based on levels, the approach suffer from possible contamination due to benchmark revisions, i.e. those changes that statistical agencies make to their methodologies or statistical changes such as change of base years or seasonal weights.

In this paper we focus on the performance of several competing models to forecast GDP growth and GDP deflator inflation. The combination of forecast is computed using three procedures: a naive scheme with equal weights; the algorithm AFTER (Aggregate Forecast Through Exponential Reweighting) proposed by Yang and Zou (2004), where the weights are assigned on the base of the past performance history of models; and a new methodology which modifies AFTER and whose weights are updated after each additional observation to target each time the performance of the best candidate model, defined in terms of an appropriate loss function.

The combination of vintages is obtained in two different ways. First, we exploit information in vintages by using a panel data approach and considering all vintages as the units of the panel. The parameters of the models are estimated by pooling all vintages and exploiting the crucial characteristic of a panel data set that contains repeated observations on the same unit (vintage). Therefore, like panel data the vintage structure of a real-time data set can provide a situation very much close to a controlled experiment. The idea is appealing not only for a forecast evaluation on past realization, where blocks of vintages have the same length, but also in real time, where necessarily different past vintages have different length of time and the panel is unbalanced.

Second, we model the covariability of the vintages in terms of one unobserved common component and an idiosyncratic error term. The common factor is estimated using principal components of several variables over a predefined block of vintages and each model is then augmented with the common factor, and a linear or non linear relationship between the variable to be forecast and the factor is estimated. The spirit here is the same as in the static factor approach of Stock and Watson (e.g., 2002), but using all available vintages for several variables.

Several measures of instability, accuracy and precision are computed for each model in a standard forecasting exercise, where, using every vintage available, a technique similar to the repeated observation forecasting of Stark and Crushore (2002) is employed to exploit the fact that the forecasts for a particular date change as vintage changes.

The main findings of the paper can be summarised in the following points. First, we show that imposing appropriate weights on competing models of inflation forecasts and output growth – reflecting the relative ability each model has over different sub-sample periods – substantially increases forecast performance. Second, our results indicate that augmenting the information set with a signal extracted from the entire revision history of the selected variables consistently leads to a substantial improvement in forecast accuracy, precision and stability. Both sets of results have important implications for the real-time forecasting of institutions and policymakers, and point in the direction of using all information contained in the whole revision history of a variable to forecast it or to measure the associated forecast uncertainty.

The remainder of the paper is structured as follows. In Section 2 we describe the data set used with a few statistics summarising the properties of the revisions. In Section 3 we illustrate the models and the techniques to combine information from all available vintages, and describe the experimental design. In Section 4 we present the main results and their implications for a real-time forecasting. In Section 5 we conclude and explore possibilities for future works.

2 Data, preliminaries, and notation

The data used come from the real-time data set for macroeconomists, developed at the Federal Reserve Bank of Philadelphia and described in great details for instance in Crushore and Stark (2001).

Our analysis focuses on two variables, the growth rate of real output and the inflation rate based on output deflator. We use quarterly observations of quarterly vintages. Percentages are expressed in annual terms.

Following previous papers on real-time analysis, we will denote with Y_t^v the realization of the v-th vintage of the generic variable Y at time t. Our variable of interest is its growth rate h-period ahead, defined as $y_{t+h}^v = 400/h \left(Y_{t+h}^v / Y_t^v - 1 \right)$. Consequently, we will denote the h-step-ahead forecast of y_t^v, given information up to t, as $\hat{y}_{t+h|t}^v$. The same definition of the dependent variable has been used elsewhere (see e.g. Stock and Watson, 1999).

The table below describes the typical representation of a real-time data set where at the date of a given vintage v_t we observe the realizations $y_{t-1}^{v_t}$ of y. The data on the diagonal of this matrix are usually denoted as *real-time* or *preliminary* data, whereas the last column is referred to as the *latest-available or fully-revised* data.

Scheme of a Real-time data set						
	$\cdots$	v_2	v_3	$\cdots$	v_t	
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\vdots$
1		$y_1^{v_2}$	$y_1^{v_3}$	$\cdots$	$y_1^{v_t}$	$\vdots$
2			$y_2^{v_3}$	$\cdots$	$y_2^{v_t}$	$\vdots$
$\vdots$				$\ddots$	$\vdots$	$\vdots$
$t-1$					$y_{t-1}^{v_t}$	$\vdots$
$\cdots$						$\ddots$

The sample used in the analysis exploits a subset of vintages from 1969q4 to 1997q1. The latest available data are those of the vintage 2007q1. This vintage is used as *actual* data in our benchmark experiment. The historical time series observations for each vintage start in 1959:q3. The choice of the sample depends on several factors. First, we have tried to maximize the availability of historical data for subsequent vintages, whose completeness is not ensured before 1959:q3 for several vintages of both series, particularly in the 1990s. Second, the experiment should also guarantee that a sufficient number of time series observations for each vintage is employed. Our benchmark experiments are based on homogeneous blocks of vintages with ten years of data points (see below Section 3.4). Robustness checks are performed on larger samples and different blocks of vintages. Finally, the blocks of vintages used in the analysis is sufficiently far away from the latest available observations used to measure accuracy, precision and stability. Robustness checks are also performed on different *actual* data, as explained below (see Section 3.4).

One of the main points of our paper is the consideration of **all** vintages of data in the analysis. The existence of multiple vintages of data for a given variables might render incorrect the use of a single vintage when evaluating a model because the stochastic relationship between vintages is not taken into account and therefore the estimated uncertainty is distorted. To avoid a vintage dependence of the forecast accuracy our idea is to combine the information contained in several (if not all) vintages. The main argument is that revisions do not seem to be well behaved, for they are large compared to the original variable, they do not have a zero mean, and they are predictable. These features have been shown, for instance, in Aruoba (2005), who documents the properties of revisions to major macroeconomic variables in the US.

To have a rough and general idea of the revision process for each variable, Figure 1 reports box-plot charts where the (annual averages of) minimum, maximum, interquartile range and the median are computed over all vintage realizations at a given date.

- Figure 1 here -

Leaving aside the last four or five years of observations, for which the numerosity of the vintages is limited, the charts are quite persuasive about the dispersion of the revision process, especially before the 1980's. The distance between minimum and maximum is also eloquent about the approximate size and the distributions are far from being always symmetric.

To put this in perspective, the revision process can be roughly described examining the properties of four simple statistics: the complete revision, the remaining revision, the real-time revision and the standard deviation of the non zero revisions. Revisions are defined on the growth rates of the variables to avoid that results are driven by benchmark revisions, i.e. those changes that statistical agencies make to their methodologies or to the base years or to the seasonal weights.

If we define a revision from a vintage v to a vintage $v+1$ as

$$r_t^v = y_t^{v+1} - y_t^v,$$

then the total (or cumulative) revision is defined as

$$r_t^C = y_t^{final} - y_t^1 = \sum_v r_t^v$$

This simple statistics provides an indication about the total amount of revision between the preliminary data and the fully-revised data.

A "remaining revision" is defined as

$$r_t^R = y_t^{final} - y_t^v$$

and represents the revision remaining after the data release at vintage v.

The standard deviation of the non-zero revisions is easily defined as

$$s_t = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left(r_t^j - \bar{r}_t\right)^2}$$

where j counts the non-zero revisions and $\bar{r}_t$ is their sample mean. This measure gives an indication of the uncertainty surrounding the revision process.

Finally, the real-time revision is the difference between the preliminary data of a given vintage and the corresponding (revised) figure in the subsequent vintage. It is interesting to show this revision because preliminary data are **always** revised in the subsequent vintage and therefore the revision is always different from zero. As we will se, in our experiments of Section 3.4 the forecast revisions are always due at least to real-time revisions.

- Table 1 here -

Table 1 reports the descriptive statistics of these three measures for GDP growth and inflation and Table A1, in the appendix, reports the same statistics for other variables used in the analysis. Specifically (and mimicking Aruoba, 2005) we report the mean, the minimum, the maximum, the standard error also relative to the standard error of the latest available value of the variable (noise/signal), and the first autoregressive coefficient.

Note that revisions – especially those of output and industrial production growth – can be large and have typically a mean positively different from zero. Their standard errors and the minimum-maximum revisions indicate that they are also sizable relatively to the latest available original variable. Finally, the autoregressive structure of most variables seem to suggest predictability or at least some significant, though not very high, persistence. The variability of the revisions as measured by s_t shows a strong positive and significative autocorrelation structure for all series, meaning that the uncertainty surrounding the revision process is not only relatively high – as the mean indicates – but also quite persistent. To put the latter result in a better perspective, Figure 2 shows the time series of s_t from 1965:3 to 1999:4 for output growth and inflation. Clearly the revision process has been far from being constant, though in the last part of the sample, since approximately 1984, the uncertainty shows a negative trend for both variables, more markedly for inflation than for output growth. This reduction might have something to do with the "Great Moderation". We will mention again this later on, but we do not exploit further the idea in the paper and leave it for future research.

- Figure 2 here -

From this preliminary analysis we conclude that the characteristics of the revisions might support the idea that it could be worth exploiting information in all vintages of a certain variable and consider them as realizations of a stochastic process, because the size and the persistence of this process do not seem to be irrelevant as compared to the ones of a White Noise, and might leave some room to improve the forecasting performance of the models used.

3 Model description and experimental design

The information contained in vintages of time-series data can be taken into account in several ways. Our prior idea is that there might be benefits in modeling and forecasting variables by exploiting this information and using also a combination of several models. In this section we describe the models used, the combination approaches and the experimental design.

3.1 Generalities

Suppose that the problem is to forecast a variable (or a vector of variables) y^v_{t+h} at a given vintage v. We start by specifying a model or a data generating process (DGP):

$$y^v_{t+h} = m_j \left[y_t, x_{j,t}; \beta_j; \varepsilon^v_{t+h} \mid \Omega_t \left(v' \right) \right] \tag{1}$$

where a model m_j $(j = 1, 2, ..., J)$ links the variable y^v_{t+h} to own past observations and past observations of other variables, denoted by x. The latter can be model-specific and models can be nested or non-nested. Both sets of variables (y and x) are typically subject to revision. β_j collects all parameters of the model to be estimated, and ε^v_{t+h} is an error term whose properties remain to be specified.

The choice of the information set Ω_t to be used in the construction or the estimation of the model is crucial to the analysis. When $v' = v_t$, the set includes information up to time t relative only to the current vintage v. Alternatively, with $v' = (v_1, v_2, ..., v_t)$, the set might include information up to time t of *current and past* vintages. The idea here is not too much that of modeling the revision process and using its systematic properties to improve the forecast (see e.g. Swanson and van Dijk, 2006), but rather to capture a general stochastic relationship between vintages – without necessarily specifying their DGP – or use all available vintages to estimate the parameters of the model.

3.2 Models

The analysis shows the forecasting properties of eight models over horizons from one to eight quarters, organised so that five competing models for each variable are evaluated and combined. Four models are common to both inflation and output growth. Two additional models are variable-specific.

3.2.1 Common

The first model is a driftless random walk process (RW). This framework is used here as a benchmark to compare all other models. This means that the statistics of interest will be computed as ratios to the same statistics of the RW. The dynamics of the model is as simple as:

$$m_{RW}\left[y_t, x_{j,t}; \beta_j; \varepsilon_{t+h}^v \mid \Omega_t(v')\right] = y_t^v + \sigma\varepsilon_{t+h}^v \tag{2}$$

where the error term is a White Noise with variance σ^2.

The second model is an autoregressive model (AR(p)) where the order p is chosen with standard selection criteria. The dynamic of the model is

$$m_{AR}\left[y_t, x_{j,t}; \beta_j; \varepsilon_{t+h}^v \mid \Omega_t(v')\right] = \mu + \rho(L)\, y_t^v + \sigma\varepsilon_{t+h}^v \tag{3}$$

where $\rho(L)$ is a polynomial in the lag operator L, and parameters are estimated with standard OLS techniques.

The third model is a three-variable vector autoregressive (VAR) model where the lag length has been imposed equal to four:

$$m_{VAR}\left[y_t, x_{j,t}; \beta_j; \varepsilon_{t+h}^v \mid \Omega_t(v')\right] = C + A_1 y_t + A_2 y_{t-1} + A_3 y_{t-2} + A_4 y_{t-3} + \Sigma^{1/2}\varepsilon_{t+h}^v \tag{4}$$

Now y_t^v is a vector containing inflation, output growth, and the Federal Funds Rate (FFR). No assumptions are required on the structure of the variance-covariance matrix and the VAR is estimated in a standard Bayesian fashion *á la* Litterman with a symmetric tightness function. The hyperparameters of the prior are estimated using the latest-available data.

The last common model is an autoregressive Time-Varying Coefficient model (TVC), where the order of the autoregressive structure is imposed equal to two:

$$m_{TVC}\left[y_t, x_{j,t}; \beta_j; \varepsilon_{t+h}^v \mid \Omega_t(v')\right] = \mu_t + \rho_t(L)\, y_t^v + \sigma\varepsilon_{t+h}^v \tag{5}$$

This model, which have been shown to capture both non-linearity and non-normality in the data (e.g. Canova, 1993), is estimated with Kalman filter formulas. We assume that the coefficient vector

$\beta_t = (\mu_t, \rho_{1t}, \rho_{2t})'$ shrinks back towards the mean following an AR process:

$$\beta_t = \gamma\beta_{t-1} + (1-\gamma)\,\bar{\beta} + \eta_t$$

where η_t is assumed independently and identically normally distributed with mean zero and variance-covariance V, estimated via Kalman filter; $\bar{\beta}$ is the OLS estimate of the corresponding non-time varying model computed using the whole series of observations of the latest-available data; and the hyperparameter γ is estimated by ML.

3.2.2 Inflation

Two additional models are used to forecast inflation. The idea here is not so much to be exhaustive in the specification searches, but rather to select a couple of models that have proven to be relatively reliable in other studies.

The first model is a Phillips curve where inflation (π) is a function of its own past and of past dynamics of the unemployment rate (u):

$$m_{PH}\left[\pi_t, x_{j,t}; \beta_j; \varepsilon^v_{t+h} \mid \Omega_t\left(v'\right)\right] = \mu + \rho\left(L\right)\pi^v_t + \gamma\left(L\right)u^v_t + \sigma\varepsilon^v_{t+h} \tag{6}$$

The model has been used, for instance, in Stock and Watson (1999) and Ang et al. (2007), among others, who stress the importance of measures of real activity in a regression forecast for inflation. The characteristics of the unemployment revision process are summarised in the appendix (Table A1).

In the second model, inflation is a function of the term structure as proxied by the spread (SP) between long (10 years) and short run (three months) interest rates:

$$m_{SP}\left[\pi_t, x_{j,t}; \beta_j; \varepsilon^v_{t+h} \mid \Omega_t\left(v'\right)\right] = \mu + \rho\left(L\right)\pi^v_t + \gamma\left(L\right)SP^v_t + \sigma\varepsilon^v_{t+h} \tag{7}$$

Typically such regression models, or a combination of the two where both measures of real activity and the spread are included, provide a good approximation to more sophisticated Phillips curve models of inflation (e.g. Ang et al. 2007).

3.2.3 Output

The two selected models for output have also a long tradition in forecasting exercises and are frequently used, for instance, in central banks. The first model is a simplified version of Okun's law:

$$m_{Okun}\left[g_t, x_{j,t}; \beta_j; \varepsilon^v_{t+h} \mid \Omega_t\left(v'\right)\right] = \mu + \rho\left(L\right)g^v_t + \gamma\left(L\right)u^v_t + \sigma\varepsilon^v_{t+h} \tag{8}$$

where output growth rate (g) is a function of its own past and of past dynamics of unemployment rate (u).

The second model, that we call Leading Indicator (LI), relates output growth to other variables typically used as indicators of output:

$$m_{LI}\left[g_t, x_{j,t}; \beta_j; \varepsilon^v_{t+h} \mid \Omega_t(v')\right] = \mu + \rho(L)\, g^v_t + \gamma_1(L)\, E^v_t + \gamma_2(L)\, IP^v_t + \gamma(L)\, RMB^v_t + \sigma\varepsilon^v_{t+h} \quad (9)$$

Here we use employment (E), Industrial production (IP) and Real Money Balances (RMB). In many institutions this kind of equation is known as belonging to the family of *bridge equations,* because they typically relate quarterly output growth to monthly indicators. We nevertheless use quarterly observations of all variables to estimate and project it. The revision process of the additional variables is described in the appendix (Table A1).

3.3 Information combination

3.3.1 Forecast combination

The idea behind the combination of forecasting techniques is that no forecasting method is fully appropriate for all situations. Typically this means that a single forecasting model might be optimal – given a loss function – only conditional on a given sample realization, information set, model specification or time periods. In our context optimality clearly depends on different vintages. The real-time data set is an ideal framework to check the properties of a forecast combination without having to rely on simulation experiments. To our knowledge, this is the first work that evaluates the properties of a forecast combination using a real-time data set.

One of the main justifications for the use of model combination is that forecasts based on a given model may have a high variability if the model has been somehow selected, in the sense that a slight change of the data may result in the choice of a different model. The change in the data here is not artificial but given directly by the different vintages of the same variable. Therefore the instability of model selection can easily be checked and an appropriate weighting scheme can guarantee that the combined forecast has a smaller variability and that the forecast accuracy can be improved relative to the use of a selection criterion.

To better understand the idea of selection instability we have checked consistency in selection over different vintages. Concretely, if a given model has been selected in a rolling exercise using the fully-revised data, we compute the percentage of times that the same model is selected in rolling blocks of vintages and average it over the number of vintages. In the experiment, which resembles the repeated observation technique of Stark and Crushore (2002), blocks of ten years of vintages – each spanning ten years of historical data – are rolled over time and vintages (see Section 3.4), and for each vintage and each point in time a selection criterion is applied. The rational behind this exercise is simple: changing vintage should not cause much change for a stable procedure in the selection. Analogously, for the same vintage a stable selection criterion should generally select the same model, when slightly changing the sample. The results of this experiment, reported in Table 2, clearly show the difficulty of the model selection criteria in choosing the same model over the years and across vintages. Percentages are too small to claim that a given model is stably chosen according to different criteria if we use fully-revised or single vintages of the same data.

- Table 2 here -

Results also indicate that there would be enough room for model combination to improve upon the forecasting performance of a single model which would have been selected with a given criterion on a given vintage of data.

In this paper we report results relative to three forecast combination methods.

The first is the "naïve" scheme which assigns equal weight (EW henceforth) to all models.

The second is the algorithm called Aggregated Forecast Through Exponential Reweighting (AFTER henceforth) proposed by Yang and Zou (2004), where the weights are assigned on the base of the past performance history of models. In particular, to combine J forecasting models, at each time τ the AFTER algorithm looks at their past performances and assigns weights as follows:

$$
\begin{aligned}
W_{j,1} &= 1/J \\
W_{j,\tau} &= \frac{W_{j,\tau-1}\hat{\sigma}_{j,\tau-1}^{-1/2} \exp\left[-\left(y_{\tau-1} - \hat{y}_{j,\tau-1}\right)^2 / 2\hat{\sigma}_{j,\tau-1}\right]}{\sum_{j'} W_{j',\tau-1}\hat{\sigma}_{j',\tau-1}^{-1/2} \exp\left[-\left(y_{\tau-1} - \hat{y}_{j',\tau-1}\right)^2 / 2\hat{\sigma}_{j',\tau-1}\right]} \qquad \tau \geq 2
\end{aligned}
$$

where $\hat{\sigma}_{j,\tau-1}$ is the standard error of estimate of model j based on information up to $\tau - 1$, and $(y_{\tau-1} - \hat{y}_{j,\tau-1})$ is the forecast error of model j of the previous period.[1]

[1] Yang (2004) examined the theoretical convergence properties of this combination method and found that it has a significant stability advantage in forecasting over some popular model selection criteria. In particular, as already mentioned, the specific relationship imposed insures that a weight attributed to a certain model at time τ is larger the larger its ability to forecast the actual value of the variable of interest in previous periods.

The last combination scheme is our own modification of the previous algorithm (AC henceforth), where the weights are updated after each additional observation to target *each time* the performance of the best candidate model. In other words, the weight attributed to a certain model at time τ is larger the larger its ability to forecast the actual value not in all previous periods, but only at $\tau - 1$. In particular weights are assigned as:

$$
\begin{aligned}
W_{j,1} &= 1/J \\
W_{j,\tau} &= \frac{\exp\left[-\left(y_{\tau-1} - \hat{y}_{j,\tau-1}\right)^2 / 2S_\tau^2\right]}{\sum_{j'} \exp\left[\left(y_{\tau-1} - \hat{y}_{j',\tau-1}\right)^2 / 2S_\tau^2\right]} \qquad \tau \geq 2
\end{aligned}
$$

where S_τ^2 is the sample variance of the dependent variable, $S_\tau^2 = (\tau - 1)^{-1} \sum_{s=1}^{\tau-1} (y_s - \mu_\tau)^2$, and $\mu_\tau = \tau^{-1} \sum_{s=1}^{\tau-1} y_s$.

Our own experience shows that the AFTER weighting scheme after a while might start giving too much weight to one model, thus resembling more a model selection algorithm than a combination method. The reason is that the weights are designed to give importance to the whole history of performances. Therefore if one model has on average a better performance than the others, by the end of the sample it will get a weight equal to one with a high probability. The AC scheme, instead, gives a large weight to model j at period τ only when it forecasted well at period $\tau - 1$. Thus after each additional observation, the weights on the candidate forecasts are updated with more likelihood than in the AFTER scheme.

3.3.2 Vintage combination

One of the innovative ideas of this paper is that the existence of multiple vintages of data for a given variables might render incorrect the use of a single vintage when evaluating a model because the stochastic relationship between vintages is not taken into account. Therefore, to avoid a vintage dependence of the forecast accuracy it seems natural to exploit or combine the information contained in several (if not all) vintages.

The combination of vintages is obtained in two simple ways. In the first one, we exploit information in vintages by using a panel data approach and considering all vintages as the units of the panel. The parameters of the models are estimated by pooling all vintages and exploiting the crucial characteristic of a panel data set that contains repeated observations on the same unit (vintage). Therefore, like panel data, the vintage structure of a real-time data set can provide a situation very much close to a controlled experiment. The idea is appealing not only for a forecast evaluation on past realizations, where blocks of vintages might have the same length, but also in real time, where necessarily different past vintages have different time-length and the panel is unbalanced. In other words, a real-time implication of this approach is that if we want to forecast the variable y^v_{T+h}, instead of using only the historical values $y^v_{t=1,..,T}$ of the same vintage v, we fit a panel data model on $y^{v=v_1,...,v_T}_{t=1,...,T_v}$ and estimate the parameters by pooling all past vintages.

In the case of a general linear model, for instance, it would be:

$$y^v_{t+h} = X^v_t \beta_v + \varepsilon^v_{t+h} \qquad t = 1, 2, ..., T_v \tag{10}$$

where X^v_t combines past values of y and other indicators, and the parameter vector β^v can be assumed equal (pooling regression) or different (fixed or random effect model) across vintages.

In the second approach, we model the covariability of the vintages in terms of one unobserved common component and an idiosyncratic error term. The common factor is estimated using principal components (of possibly several variables) over a predefined block of vintages and each model is then augmented with the common factor, and a linear or non linear relationship between the variable to be forecast and the factor is estimated. The spirit here is the same as in the static or dynamic factor approach of Stock and Watson (e.g., 2002), where the signal is extracted from all available vintages (of several variables). Each model is therefore estimated and projected after being augmented with this signal. The general specification becomes:

$$m_j \left[y_t, x_{j,t}; \beta_j; \varepsilon^v_{t+h} \mid \Omega_t (v') \right] = \mu + \rho (L) y^v_t + \gamma (L) x^v_t + \delta (L) F_t + \varepsilon^v_{t+h} \tag{11}$$

where F_t is a (vector of) factor(s) common to all vintages (and possibly several series) estimated with usual principal component techniques over a number of vintages of a number of indicators. The advantage with respect to usual factor analysis is the consideration of all vintages available to the researcher and not just the last one. The disadvantage with respect to the panel data approach is that the computation burden increases with the number of series involved.

3.4 Experimental design

The experiment we conduct is easily described. Given the sample size (1947q1 to 2006q4) and the available number of vintages (1965q1 to 2007q1), we construct matrices of ten years of vintages each spanning ten years of time series observations. Then we make these matrices rolling "diagonally" over time and vintages so that they have always the same numerosity (ten years of vintages and ten years of data points), and at each rolling step, for every vintage, we estimate a model and use it to forecast the variables of interest at 1, 4, and 8-steps ahead. To better understand the logic, imagine a rolling forecast with one time series where the length of the series is always the same. Here we repeat the exercise not just for one vintage but for several of them.

In our benchmark experiment, the columns (vintages) of the first data-matrix run from 1969q1 to 1978q4, and the rows (historical data) go from 1959q1 to 1968q4. The one-step ahead forecasts, therefore, are those relative to the quarter 1969q1. The columns of the last matrix, instead, go from 1997q2 to 2007q1 and the rows from 1987q2 to 1997q1. Therefore, the one-step ahead forecasts are relative to the period 1997q2. As also remarked above, the choice of this benchmark trades-off between a sufficient number of vintages that contains revisions, and a reasonable number of time series observations for the evaluation measures to be statistically meaningful. In total there are 114 blocks of vintages.

The amount of results that this design generates is huge, but manageable. The advantage of our strategy is that we have two dimensions over which we can compute the relevant statistics: the time series dimension – which, for instance, provides a time series of one-step ahead forecasts spanning from 1969q1 to 1997q2 – and the vintage dimensions –that works as a repetition of a controlled experiment, and gives 40 of such time series.

Moreover, results are organised by model. In total we have eleven models for each variable: the five models described above for inflation and output; the three forecast combination methods, that are additionally considered as alternative models; and three "best" models that are selected by three competitive selection criteria. The latter are the AIC (Akaike, 1974), the SBC, (Schwarz, 1978), and the HQ (Hannan and Quinn, 1979).

In the panel approach, we treat the vintages of the same block as units of the panel and estimate the parameters using a mean group strategy (Pesaran and Smith, 1996). We choose not to pool the vintages because the variability of parameter estimates can be substantial. This is easily appreciated in Figure 3, where we plot the largest and smallest estimates (across ten years of rolling vintages) of the first autoregressive coefficient of the ARMA model for inflation over the period 1969q1-1997q2. The vertical distance between the two lines provides a good measure of this variability. Note that forcing units to be homogeneous with a simple pooling could generate biased and inconsistent estimates given such a variability.

- Figure 3 here -

In the factor approach, we use the following estimation strategy. We take all available variables of the data set and build the same 114 blocks of vintages as described above. For each block, we compute three principal components over the averages across the ten-year vintages of the block for each variable. Then we augment all models with lagged values of these factors as in (11). A list of the variables used to compute the principal components is reported in the appendix (Table A2).

It is worth stressing that ours is a purely evaluation experiment and not a real-time one. In fact both the panel and the factor approach use information contained in the whole block of vintages to estimate and forecast each vintage of the block. However, the same strategies can be employed in real time when only past vintages would be considered in the estimation.

Our benchmark experimental design allows us to compare results across models, for each assumption on the information set – simple, panel, or factor – and across information sets, to check if making use of more information contained in all vintages improves the forecasting performance of each model. Broadly speaking, the strategy is the one of an ensemble forecast methodology commonly used in weather forecast, which consists of designing a number of simulations on a given forecast computed by allowing for small changes to the estimate of the current data used to initialize the simulation. In our context we use the different vintages of the real-time data set as different realizations of a given variable. Then, in order to evaluate the uncertainty in the data measurement process, we use these vintages instead of simulating the current data with an ad-hoc perturbation.

We focus our evaluation exercise predominantly on three dimensions: accuracy, uncertainty and stability. The idea is to check whether information and model combination improve the forecasting performance in these three dimensions. The computation of the statistics in our framework is straightforward. Each rolling block of matrices has always the same number of vintages. Suppose that we have a h-step ahead forecast for the generic vintage v. Because our matrices roll over time, we end up having a time series of h-step ahead forecasts for vintage v. In the case of $h = 1$ the time series spans the period 1969q1-1997q2. For $h = 4$ it is 1969q4-1998q1; and for $h = 8$ it is 1970q4-1999q1. All relevant statistics are computed over these time periods for each h, and replicated for the number of vintages available.

4 Empirical results

In this section we report the empirical results of our experiment along the three dimensions of accuracy (Section 4.1), uncertainty (Section 4.2) and stability (Section 4.3). A summary of the main results will conclude (Section 4.4).

4.1 Accuracy

The most unambiguous measure of accuracy is correct prediction. Accuracy is defined in terms of square root of the mean squared prediction error (RMSPE) for the generic vintage v:

$$RMSPE^{v} = \sqrt{\frac{1}{T}\sum_{t=t_0}^{T+t_0-1}\left(y_{t+h}^{v} - \hat{y}_{t+h|t}^{v}\right)^{2}} \quad (12)$$

where T is the length of the series of forecasts and t_0 is the last observation of the vintage v. As actual value, y_{t+h}^{v}, we take here the latest-available observations.

To compare accuracy across models in a general manner, Table 3 reports the average RMSPE over vintages for all models and information sets. Reported values are relative to the those of a RW model.

- Table 3 here -

Some comments are in order. Note first that, apart from few exceptions, most of the models and methods considered here are on average better than a naïve forecast based on the RW. For inflation, this is almost always true. Second, it seems that combining forecasts using our procedure (AC) always gives a RMSPE which is at least in the first three ranking positions, and in 72% of the cases it is ranked first across all forecast horizons, variables, models and information sets. Moreover, the other combination methods and the other models do not show a comparable performance in such a systematic way, or a stable and systematic pattern in their general forecast ability. This is an important results that shows how combining forecast can lead to a stable improvement when model selection is not necessarily stable – as in our case –, provided that the choice of the combination weights is carefully executed. Finally, augmenting the information set with a signal extracted from all vintages can lead to a substantial improvement with respect to a simple approach. This result is confirmed especially when we use the panel approach, since the comparisons between the RMSPE of each model indicate that for almost all forecast horizons, all models and both variables the panel approach provides lower RMSPE than both the simple and the factor approach. Sometimes the improvement is as considerable as 47% (inflation, comb. AFTER, four-steps ahead) or 37% (output, comb. AC, eight-steps ahead). In fact, overall the panel approach is *consistently* better than the others, across models, variables and forecast horizons. The factor approach does not seem to perform as well as expected. This is possibly due the limited number of variables available in the dataset used to compute the factors.

These results, though valid only on average, are remarkable and confirm our prior intuition that a combination of both information and forecasts from different models might improve the forecast accuracy. Their potential impact should of course be evaluated in real time. The typical pseudo real-time experiment can nevertheless be undertaken by using, for instance, the last vintage of historical data in each block to forecast the variable h-step ahead. In this case, a panel or a factor approach as used in our experiment do not imply using future information as it is when we run the experiment for the first $V-1$ vintages of the block. In Table 4 we report the RMSPE associated with the latest available series. Results are by and large confirmed in both respects that forecast and information combination notably improve forecast accuracy.

- Table 4 here -

As remarked above, these results are informative only on average. Our experiment, however, produces a much greater amount of output which is worth examining in detail. Moreover, we need to check how *statistically* significant the general results are. Therefore, to put them in a better perspective, we test the predictive accuracy across models for each information set – simple, factor and panel – and across information set for each model.

We use an approximately normal test for equal predictive accuracy in nested models as described in a recent paper by Clark and West (2007). The checking procedure is applied vintage by vintage and models are tested bilaterally. For each vintage of forecasts, the null hypothesis of equal accuracy (equal MSPE) is checked against the alternative that the larger model has a smaller MSPE. The null is tested by examining the difference between the MSPE of model 1 (the restricted one) and that of model 2 (the larger one), i.e. $MSPE^1 - \left(MSPE^2 - adj.\right)$, where the "*adj.*ȷ term *"adjust for the upward bias in MSPE produced by estimation of parameters that are zero under the null"*.[2] Repeating

[2] Clark and West (2007) p. 294.

the experiment $\begin{pmatrix} n \\ 2 \end{pmatrix}$ times – where n is either the number of models (11) or the number of ways we combine the information set (3) –, we count the number of times that in total (over bilateral comparisons and vintages) we reject the null in favour of model 2 and report this information in percentage form in Figures 4-6. For all tests significance is set at 0.10 level. For instance, in Figure 4 a bar in correspondence of the model ARMA represents the percentage of times that in bilateral comparisons for all vintages we reject the null of equal accuracy (with a ten percent significance) in favour of the ARMA model. Whereas in Figure 5, fore each model and forecast horizon, a bar in correspondence of PANEL represents the percentage of times that in bilateral comparisons and for all vintages we reject the null of equal accuracy in favour of the PANEL. No bars at all means that all models are equally accurate.

Results statistically confirm and qualify the averages of Table 3. Briefly, in the comparison across models (Fig.4) the forecast combinations computed with the AC method shows always the highest percentage of times that in bilateral comparisons we reject the null in favour of the AC. Only in one case (Output growth, simple eight-step ahead) models are all equally accurate. Analogously, in the comparison across information sets (Fig. 5 and 6) the panel approach is very much competitive especially at one-step ahead for inflation and eight-step ahead for output growth. In several cases, the three different approaches seem all equally accurate from a statistical point of view. The factor approach has some differential forecasting power only at one-step ahead for some models of output growth.

- Figure 4 to 6 here -

4.2 Uncertainty

The concept of forecast uncertainty does not necessarily has a unique empirical counterpart.

From an ex-ante point of view, we know that most uncertainty can arise from extra-model information with respect to the future values of the variables, and is therefore extremely difficult to capture with model-based historical forecast errors. From an ex-post point of view, however, it is inevitable to relate an estimation of the forecast uncertainty to the accuracy of point estimate forecasts, and measure it with the MSPE that combines the bias and the variance in an appealing way. In this respect, the previous subsection would suggest that a forecast combination which also uses the information contained in all vintages provides not only more accurate but also more precise forecasts. This makes sense at least to the extent that every model uses different information variables, and none of them can be considered a priori as a correct description of the DGP. Therefore, a forecast combination, which also exploits efficiently the information contained in all vintages could provide a more accurate and precise forecasts.

We measure uncertainty in two manners, broadly consistently with this ex-ante/ex-post difference and distinguish between a *predicted forecast uncertainty* – the uncertainty anticipated given the model – and an *actual forecast uncertainty* – arising by relating the forecast of the model and the actual data.

The first approach consists of computing the usual standard error of the forecast associated with each model. For instance, in the case of a one-step-ahead forecast for the linear (possibly) dynamic model

$$y^v_{T+1} = x^{v\prime}_T \beta + \varepsilon^v_{T+1},$$

an estimate of the variance of the forecast error is

$$(\hat{\sigma}^v_e)^2 = (\hat{\sigma}^v)^2 \left[1 + x^{v\prime}_T \left(X^{v\prime} X^v\right)^{-1} x^v_T\right] \tag{13}$$

where $(\hat{\sigma}^v)^2$ is the OLS estimate of the model variance relative to vintage v. This measure is available for each vintage of the block and each time period, and is mainly based on the in-sample information needed to estimate σ^2. An average over all vintages

$$\hat{\sigma}^2_e = \frac{1}{V} \sum_v (\hat{\sigma}^v_e)^2$$

can be considered as a measure of *predicted uncertainty* for each model, step-ahead and information set. Given the rolling exercise, this measure varies with t and covers the periods 1969q1–1997q1 for $h = 1$, 1969q4–1998q1 for $h = 4$, 1970q4–1999q1 for $h = 8$.

The second approach to measure uncertainty is based on forecast errors. This is a novel approach which exploits the structure of a real-time database. In fact, to the extent that forecast uncertainty reflects the dispersion of possible results relative to a given forecast, the structure of a real-time data set can be of great help in conveying this uncertainty if, as argued in this paper, we consider the different vintages of the data set as repeated outcomes of the same experiment and compare them with given forecasts. A distribution of forecast errors can therefore be obtained by relating the forecast of the variable h-step-ahead for each vintage, to the alternative sequences of "outcomes" that have occurred in the revision process at the corresponding date of the forecast in the subsequent vintages. For instance, suppose that at the vintage 1968:q4, we produced a 1-step-ahead forecast of the variable of interest. Therefore we are forecasting the value of the variable for 1969:q1. We then use the subsequent vintages to forecast the same variable for the same date 1969:1. We store then as many forecasts as available vintages, and compute the sequence of forecast errors using for instance the latest available value of the variables at the date 1969:q1. This gives us a cross section of forecast errors at that date. Using our rolling approach of block of vintages, such cross-sectional distributions can be computed for each h (or date), so that we have a panel of forecast errors, which will form the basis of our final distributions.The variance of such distributions is our measure of *actual forecast uncertainty*.

Figure 7 and 8 report the predicted measure for all models and Table 5 summarises the information by taking an average over the time-span. Results for the combination methods should be taken *cum grano salis* because they assume that forecast errors of different models are independent. The reason is that there is not a trivial way of predicting covariances of forecast errors without actually using the forecast errors. But if we do so, variances and covariances would be compatible only asymptotically and results would be affected by small sample problems. The comparison here, therefore, will be done more over information approaches than over time-series models.

- Figure 7 and 8 and Table 5 here -

Results confirm the idea that forecast uncertainty depends also upon the information available. In this context, the use of additional information plays a crucial role in reducing forecast uncertainty. The models augmented with factors show a systematic lower uncertainty with respect to both “Simple” and “Panel” approaches. The “Panel”, in turn, seems on average slightly better than the “Simple”. These findings are consistent with what generally expected – usually when models are not mis-specified and there are not structural breaks in the data – and are valid for both variables, for all steps, and for all models. In particular the average percentage improvement of the Factor with respect to the Simple approach is around ten percent (with a peak of up to 50% for the BVAR, inflation 8-step ahead), whereas the one of the Panel approach is constantly above three percent. Moreover, improvements due to the use of a factor approach are higher on average in forecasting inflation than output growth, whereas those due to the panel approach are on average higher in forecasting output growth.

Figure 9 and 10 summarise the (panel) distributions of the actual forecast uncertainty for all models. They have been computed pooling all forecast errors of our repetitive block experiment as explained above, from ten years of forecast vintages and corresponding outcomes over the periods 1969q1–1997q1 for $h = 1$, 1969q4–1998q1 for $h = 4$, 1970q4–1999q1 for $h = 8$. The same information is further summarised in the Table 6, where we report the average variance for each model and step, and rank the models according to the variance. The distributions of forecast errors are by and large symmetric around zero except, perhaps, the eight-step-ahead forecast of inflation, which show a consistent downward bias for all models but the TVC. Moreover, the uncertainty surrounding the forecasts of output growth is greater than the one surrounding the inflation forecasts, as it is usually the case. The comparison across models broadly confirm the previous findings. In particular, and consistently with the results described also in the Accuracy section – although using different forecast errors –, our preferred combination method (AC) and the panel approach provide on average the lowest measures of ex-post forecast uncertainty. Interestingly the Factor approach, that would provide a lower anticipated uncertainty than the Panel approach, shows a higher ex-post uncertainty as a consequence of its just-adequate forecasting performance.

- Figure 9 and 10 and Table 6 here -

Overall, our interpretation of the results can be rationalised with a reference to the cathegorization of the sources of model-based forecast errors made for instance by Clements and Hendry (1998, Ch. 7). The sources can be grouped in five categories: (1) Structural changes; (2) Model mis-specification; (3) Data revision and variable mis-measurement; (4) Estimation uncertainty; (5) Accumulation of future shocks to the economy. An approach that combines forecast and information in an efficient way may moderate the effects of at least three of these sources. Specifically it can help (i) reduce the effects of model mis-specification by combining results from several models; (ii) alleviate the inaccuracies in the estimates of model's parameters because they are estimated from a pooling of all vintages; (iii) reduce the effects of mis-measurement of the data by considering the information contained in all vintages.

Note finally that ex-ante measure of uncertainty is much wider than its ex-post counterpart. This aspect has two sides. On the one hand, it says that forecast errors cannot lay too much outside an anticipated range of plausible outcomes. In this sense, the models used in the experiment do not show big specification problems. On the other hand, the ex-post measure would suggest that confidence bands based on an ex-ante measure could be too wide with respect to what in fact happened in the data. These considerations are clearly of great importance, for instance, for policy makers who might want to communicate to the general public not only a point forecast but also their degree of confidence about the point forecast, and have to mediate between ex-ante and ex-post analyses.

4.3 Stability

In Section 3 we have shown that there is a relatively high degree of selection instability and argued that this instability would have also affected accuracy and uncertainty, as confirmed above. As noted elsewhere (e.g. Yang and Zou, 2004) however, stability in selection does not necessarily capture the stability in forecasting because the likelihood of equal accuracy across models can be high, as also seen above. Therefore, in this section we check the degree of forecast stability of each model.

Our definition of stability has to do with the degree of responsiveness of forecast to data revision. The question could be posed in this way: how much does data revision change a model forecast? Or, equivalently: how much responsive a model forecast is to the revision process? The idea of this section, therefore, is to relate the forecast revision obtained with a given model to the data revision and check how much "pass-through" from data revision to forecast revision exists.

To measure the forecast revision we take the absolute deviation of each vintage forecast from the forecast of the fully-revised data, relative to the standard error of estimate of the actual value computed using the same model. That is, we consider

$$\hat{\gamma}^{v}_{t(h)} = \frac{\left|\hat{y}^{final}_{t+h|t} - \hat{y}^{v}_{t+h|t}\right|}{\hat{\sigma}}$$

where $\hat{y}^{final}_{t+h|t}$ and $\hat{\sigma}$ are computed based on the latest-available data, and the time span depends on h, as for the uncertainty measure. To measure the data revision we use the same approach and compute the absolute standardised remaining revision

$$\gamma^{v}_{t(h)} = \frac{\left|y^{final}_{t+h} - y^{v}_{t+h}\right|}{\hat{s}}$$

where $\hat{s}$ is the square root of the sample variance of $y^{final}_{t+h|t}$.[3]

[3] The properties of the "remaining revisions" have been briefly described in Table 1 (see Section 2).

This "pass-through" is checked with a simple OLS regression:

$$\hat{\gamma}^{v}_{t(h)} = \beta^{v}_{0} + \beta^{v}_{1}\gamma^{v}_{t(h)} + u^{v}_{t} \tag{14}$$

and is measured by the estimate of β^{v}_{1}. The rationale behind this measure is that, on average, a good forecasting procedure should pro-actively respond to data revision: the larger the response, the better the model. Table 7 reports the results and ranks the models according to the size of the average β_1, where the average is taken over the ten-year blocks of our experiment. Figure 11-12 report also the evolution of β_1 over the ten years of vintages.

Results show that, apart from rare exceptions, all models have a positive degree of responsiveness. The highest estimates are almost always those corresponding to the forecast combinations, regardless of the information set, though the simple and panel approach seem to have on average higher responsiveness than the factor approach. The evolution of the coefficients in Figure 11 and 12 somehow confirms the expectation that the pass-through might have a tendency to increase as the revision becomes less pronounced, perhaps as a consequence of a learning process, or of the fact that data become more rational – as shown for instance by Swanson and van Dijk (2006) – and therefore the forecast revision adapt more rapidly.

- Figure 11 and 12 and Table 7 here -

The difference between the reaction of any model and those of the forecast combination models are sometimes remarkable, especially for one- and four-step ahead forecasts. This finding, which says that combining forecasts from different sources makes the forecast more capable of adapting to data revision, can be rationalised using the results on accuracy and on selection instability shown before. As remarked above, the idea behind the combination of forecasting techniques is that no forecasting method is fully appropriate for all situations, in the double sense that forecasting methods have different forecast ability, and that for different vintages there is a high degree of selection instability. The combination accounts for the time-varying forecasting ability of alternative models in that a single forecasting model might only be optimal conditional on given realizations, information set, model specification or sample period. By combining methods, instead, we compensate for the weakness of each forecasting model under particular conditions, hence enhancing stability in model selection and improving the accuracy as measured with respect to the fully-revised data.

Consequently, the forecast combinations show a higher degree of responsiveness to data revision. The result is independent on the weights used in the combination method.

4.4 Summary

The results of our experiment with a real time data set are easily summarised. The prior intuition that model and information combination broadly improve the forecasting performance seems to be a valid one in at least three dimensions. First, our preferred forecast combination method provides the best forecast accuracy with average improvements of 27, 25 and 36% respectively at one-, four- and eight-step-ahead horizons for output growth, and of 11, 33 and 32% at one-, four and eigth-steps for inflation. Analogous percentage improvements are shown by the Panel approach in comparison to the simple method. Concretely the improvements are of the order of 11, 16 and 33% for output growth and of 21, 29 and 23% for inflation. This is sufficient evidence of a remarkable forecasting performance, which suggests as a general strategy (i) to use all available vintages to estimate model parameters, and (ii) to exploit several models in a combination where the weights attributed to a certain model at time τ are larger the larger its ability to forecast the actual value in the previous period $\tau - 1$.

Second, the same combination methods (AC and Panel) moderate the effects of at least three sources of actual or ex-post forecast uncertainty, for they reduce the effects of model mis-specification, alleviate the inaccuracies in the estimates of model's parameters, and reduce the effects of data mis-measurement. Moreover, augmenting the models to exploit the covariability of several variables over all vintages reduces dramatically the anticipated or predicted part of forecast uncertainty.

Finally, combining forecasts from different sources makes the forecast more capable of adapting to data revision. This result is fully consistent with the idea that a model combination eases the selection instability due to the revision process. We have shown that a change in the data – as measured by the subsequent vintages – may translate in the choice of different models according to various selection criteria. As a consequence, the forecast based on a certain model might show a high variability (see also Yang and Zou , 2004). The model combination, instead, in reducing the selection instability also reacts more accurately to the data revision.

5 Robustness analysis

In this section we analyze whether the results obtained above depend on the particular assumptions we made concerning the actual value and the sample period considered in the estimation and forecasting exercises.

In the first robustness check, we vary the actual value of both inflation and output growth. This change obviously affects our measures of forecast accuracy. Precisely, when computing forecast errors results crucially depend on what vintage of data is being used to represent the “actual” one. In this section we report results based on two alternative sets of “fully-revised data”, where the latter coincide with the data available 5 and 10 years after each block of vintages.

The results are reported in table 8 and 9. These tables are computed by using the same procedure described when presenting table 3.

The evidence emerging from these tables shows that, in most cases, the three model combination approaches still perform quite well. Moreover, in line with the results reported above, the RMSPE obtained by using the Factor and Panel approaches are consistently lower than the one retrieved with Simple across models, variables and forecasting horizons. Results on accuracy are therefore broadly independent on the choice of the fully revised data used to make the comparisons.

– Table 8 and 9 here –

The second robustness exercise consists of assessing whether the results are influenced by the selected sample. More precisely, in order to evaluate whether the results are sample-dependent, we split the sample into two parts: pre-1983q4 and post-1984q1. As a consequence, the second sub-sample covers the period of the so-called “Great Moderation”.

As documented in Blanchard and Simon (2001), Kim and Nelson (1999), McConnell and Perez-Quiros (2000), and Stock and Watson (2003), since the mid-1980s the variability of quarterly real output growth and inflation rate has declined substantially. This striking features, often termed as the Great Moderation, has a number of possible explanations.

Most studies suggest that a better conduct of monetary policy, good luck, and structural changes might have influenced the decline in macroeconomic volatility. The present study does not provide an alternative explanation for these patterns, which would be out of its scope. However, while examining whether results are robust across different samples, we also check whether there has been a changes in the estimated uncertainty, predictability and stability in quarterly real output growth and inflation since the beginning of the Great Moderation.

First, we explore whether a change in predictability might have taken place after 1984. Table 10 and 11 summarise the results. The relative ranking of the models in terms of RMSPE remains substantially unchanged. Interestingly, before 1984 forecasts constructed from the eleven models for both output growth and inflation were considerably less accurate than those obtained with the same models after 1984. The average improvement in forecast accuracy is about 20% and 30% for output growth and inflation, respectively.

– Table 10 and 11 here –

Second, we quantify the forecast uncertainty associated to each model pre and post 1984. Tables 12 through 15 show the predicted and actual uncertainty, for output growth and inflation, in the two sub-samples at selected forecasting horizons. As accuracy, also uncertainty improved and significantly dropped after the onset of the Great Moderation. This reduction is consistent across all models, variables and forecasting horizons.

Across models, the reductions in the predicted uncertainty are similar to each other, although larger in magnitude for the output growth at 1-step-ahead (0.43 on the average). The average decrease in inflation uncertainty is approximately 0.2. The moderation in forecast uncertainty is less pronounced when measured with our method: On average, actual uncertainty falls by about 0.20 and 0.05 for output growth and inflation, respectively.

The concurrence of the Great Moderation with the decline in the forecast uncertainty suggests that part of the reduction in macroeconomic volatility could also be related to declining uncertainty surrounding the data revision process and the possibility of correctly predicting the time path of output growth and inflation. This is clearly a topic that would deserve further investigation.

– Table 12 to 15 here –

Finally, we examine whether the forecasting models exhibit a different degree of "pass-through" from data revision to forecast revision over the two sub-samples. Table 16 and 17 report the values of the average pass-through coefficients as expressed in equation (14).With the exception of output growth at 1-quarter ahead (where we found no significant difference between the two sub-samples), the results underline that, consistently across models, variables and forecasting horizons, the coefficients are greater in the post- than in the pre- Great Moderation period. On average, the increase in the degree of responsiveness of forecast revision to data revision is about 0.4. This means that the ability of the selected models to adapt their forecast to data revision has increased substantially after 1984.

– Table 16 and 17 here –

In sum, the results are not only robust when changing the “actual value” or the sample, but also suggest that starting form the mid-80’s there has been an increase in predictability, a reduction in both predicted and actual forecast uncertainty, and a rise in the responsiveness of forecast revision to data revision. A thorough discussion of these results as linked to the great moderation goes beyond the scope of the paper, as remarked. However, it is interesting to note here that the literature would not necessarily agree with our finding on the general improvement of forecast accuracy after 1984 (see e.g. Stock and Watson, 2005, and D ’Agostino et al., 2005), a finding that is nonetheless consistent with the reduction in the uncertainty of the revision process documented in Section 2.

6 Conclusion and directions for future work

In this paper we have investigated the forecasting performance of different models designed to capture the time path of inflation and output growth with the help of the real-time data set for macroeconomists, developed at the Federal Reserve Bank of Philadelphia. Precisely, the analysis explored, for each variables, the forecasting properties of eleven time-series models. Moreover, we have analysed two different ways of combining the information coming from the entire revision history of the selected variables, considering the vintages of the real time data set as units of a panel to be used in the estimation of the model. This is a novel approach with important implications for real-time analysis.

The results suggested that model and vintage combination might significantly improve the forecast ability of the models in the three dimensions of accuracy, precision and stability. Robustness analysis also indicate that, when changing the samples or the actual values used to construct forecast errors, results remain substantially unchanged. The real-time implications point in the direction of using all information contained in the whole revision history of a variable to forecast it or to measure the associated forecast uncertainty.

Finally, we have compared the forecast performance of the selected models before and after the onset of the Great Moderation. Overall, the results indicate that starting form the mid-80's there has been an increase in predictability, a reduction in both predicted and actual forecast uncertainty, and a rise in the responsiveness of forecast revision to data revision. Understanding how these factors (especially a change in the revision process) contribute to the observed decline in macroeconomic volatility suggests a natural direction for future work.

Appendix

Summary of revisions

– Table A1 here –

List of variables for the factor analysis

– Table A2 here –

References

[1] Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control* 19 (6), pp. 716–723.

[2] Altavilla, C. and M. Ciccarelli (2007), Inflation Forecasts, Monetary Policy and Unemployment Dynamics: Evidence from the US and the Euro area, *ECB Working Paper*, No.725.

[3] Ang, A., Bekaert, G. and Wei, M. (2007), Do macro variables, asset markets, or surveys forecast inflation better?, *Journal of Monetary Economics*, vol. 54(4), pages 1163-1212, May.

[4] Aruoba, B., (2005), Data Revisions Are Not Well-Behaved, *CEPR Discussion Papers* No.5271

[5] Blanchard, O., and J. Simon (2001). "The Long and Large Decline in U.S. Output Volatility," *Brookings Papers on Economic Activity*, 1, pp. 135-64.

[6] Canova (1993), "Modelling and Forecasting Exchange rates with a Bayesian Time-Varying Coefficient Model", *Journal of Economic Dynamic and Control*, 17, 233-261.

[7] Ciccarelli, M., G. Coenen and J. Henry (2006), A Real-Time DataBase for the euro area, *ECB Research Bulletin* No. 5, ECB.

[8] Clark T. E. and West K. D. (2007), Approximately Normal Tests for Equal Predictive Accuracy in Nested Models, *Journal of Econometrics*, 138(1), pp. 291-311.

[9] Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press

[10] Cole, R. (1969), Data Errors and Forecasting Accuracy, in Jacob Mincer, ed., Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance, New York: National Bureau of Economic Research, pp. 47-82.

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

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

[13] Croushore, D. (2006), Forecasting with real-time macroeconomic data, In: Graham Elliott, Clive W.J. Granger, and Allan Timmermann, eds., Handbook of Economic Forecasting (Amsterdam: North-Holland, 2006), pp. 961–982.

[14] D 'Agostino, A., D. Giannone and P. Surico (2006), (Un)Predictability and Macroeconomic Stability, ECB WP no. 605.

[15] Denton, F. T., and J. Kuiper (1965), The Effect of Measurement Errors on Parameter Estimates and Forecasts: A Case Study Based on the Canadian Preliminary National Accounts, *Review of Economics and Statistics* 47, pp. 198- 206.

[16] Hannan, E. J., and B. G. Quinn (1979), The Determination of the Order of an Autoregression, *Journal of the Royal Statistical Society, B*, 41, 190-195.

[17] Kim, C.-J., and C. Nelson (1999), Has the U.S. Economy become more Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle, *The Review of Economics and Statistics* 81(4): 608- 616.

[18] Koenig E. F., Dolmas S. and Piger J. (2003), The Use and Abuse of Real-Time Data in Economic Forecasting, *The Review of Economics and Statistics*, vol. 85(3), pp. 618-628

[19] McConnell, M., and G. Perez-Quiros (2000) "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?" *American Economic Review* 90(5): 1464-1476.

[20] Guerrero V.M. (1993), Combining historical and preliminary information to obtain timely time series data, *International Journal of Forecasting*, 9, pp.477-485

[21] Orphanides, A. and S. van Norden (2005), The Reliability of Inflation Forecasts Based on Output Gaps in Real Time, *Journal of Money, Credit, and Banking* 37, pp. 583–601.

[22] Patterson K.D. (2003), Exploiting information in vintages of time-series data, *International Journal of Forecasting*, 19, pp. 177–197

[23] Pesaran, M.H. and R.Smith (1996), Estimating long run relationships from dynamic heterogeneous panels, Journal of Econometrics, 68, 79-113.

[24] Schwarz, G. (1978), Estimating the Dimension of a Model, *Annals of Statistics*, 6, pp. 461-464.

[25] Stark, T. and D. Croushore (2002), Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics* 24: 507-531.

[26] Stock, J. and M. Watson, (1999), Forecasting inflation, *Journal of Monetary Economics*, vol. 44, pp. 293-335

[27] Stock, J. and M. Watson (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business & Economic Statistics*, American Statistical Association, vol. 20(2), pages 147-62, April.

[28] Stock, J., and M. Watson (2003), "Has the Business Cycle Changed and Why?" NBER Macroeconomics Annual 2002 17: 159-218.

[29] Stock, J., and M. Watson (2005), Has inflation become harder to forecast?, mimeo.

[30] Swanson, N. (1996), Forecasting Using First-Available Versus Fully Revised Economic Time-Series Data, *Studies in Nonlinear Dynamics and Econometrics* 1, pp. 47-64.

[31] Swanson, N. R. & van Dijk, D.ick (2006), Are Statistical Reporting Agencies Getting It Right? Data Rationality and Business Cycle Asymmetry, *Journal of Business & Economic Statistics*, American Statistical Association, vol. 24, pages 24-42, January

[32] Yang Y. and Zou H. (2004), Combining time series models for forecasting, *International Journal of Forecasting*, vol.20, pp. 69– 84.

[33] Yang Y. (2004), Combining forecasting procedures: some theoretical results, *Econometric Theory*, vol. 20, pp.176-222.

		Mean	Minimum	Maximum	Std. Error	Noise/Sign
Employment	Total revision	0.14	-2.26	2.53	0.95	0.44
	Real-time revision	0.04	-0.92	1.14	0.28	0.13
	Standard deviation	**0.13**[a]	0.07	0.62	0.12	0.06
	Remaining revision					
	v = 2	0.10	-2.23	2.29	0.88	0.41
	v = 6	0.05	-1.90	1.68	0.61	0.28
	v = 10	0.02	-1.12	1.14	0.41	0.19
	v = 20	-0.01	-0.81	0.64	0.26	0.12
	v = 40	0.00	-0.75	0.64	0.24	0.11
Unemployment	Total revision	**0.01**	-0.17	0.23	0.08	0.04
	Real-time revision	0.00	-0.10	0.23	0.03	0.01
	Standard deviation	**0.03**[a]	0.00	0.15	0.03	0.01
	Remaining revision					
	v = 2	**0.01**	-0.17	0.23	0.08	0.03
	v = 6	**0.02**	-0.13	0.17	0.06	0.03
	v = 10	**0.02**	-0.10	0.13	0.05	0.02
	v = 20	**0.01**	-0.03	0.13	0.03	0.01
	v = 40	0.00	-0.03	0.03	0.01	0.00
Industrial Production	Total revision	**0.48**	-7.58	10.11	2.94	1.36
	Real-time revision	**0.16**	-0.96	3.19	0.54	0.25
	Standard deviation	**0.18**[a]	0.46	5.21	0.66	0.30
	Remaining revision					
	v = 2	0.32	-7.62	9.58	2.94	1.36
	v = 6	0.18	-14.08	9.58	3.05	1.41
	v = 10	0.11	-14.08	9.60	3.06	1.41
	v = 20	-0.08	-14.08	9.60	2.74	1.26
	v = 40	-0.27	-14.08	9.60	2.29	1.06
Real Money Balances	Total revision	-0.14	-15.58	15.63	4.61	2.13
	Real-time revision	-0.08	-5.41	3.87	1.09	0.50
	Standard deviation	**0.26**[a]	0.08	2.06	0.45	0.21
	Remaining revision					
	v = 2	-0.23	-15.90	16.28	4.70	2.17
	v = 6	**-0.63**	-13.06	10.41	3.86	1.78
	v = 10	**-0.69**	-11.71	10.08	3.57	1.65
	v = 20	-0.38	-10.20	11.38	2.95	1.36
	v = 40	-0.16	-9.27	5.59	1.87	0.86

(a) Ratio to standard deviation of the latest available.
Sample: 1965:3-1999:4. Values in bold are significantly different fom zero (5%)

Table A1. Descriptive statistics of revisions

	Core Variable	File Name	Column Headers
		Quarterly (NIPA) Variables	
1	Nominal Output	NOUTPUT.xls	NOUTPUTyyQq
2	Real Output	ROUTPUT.xls	ROUTPUTyyQq
3	Output-Price Index	P.xls	PyyQq
4	Real Personal Consumption Expenditures	RCON.xls	RCONyyQq
5	Services	RCONS.xls	RCONSyyQq
6	Nondurables	RCONND.xls	RCONNDyyQq
7	Durables	RCOND.xls	RCONDyyQq
8	Real Investment	N.A.	N.A.
9	Business Fixed	RINVBF.xls	RINVBFyyQq
10	Residential	RINVRESID.xls	RINVRESIDyyQq
11	Change in Inventories	RINVCHI.xls	RINVCHIyyQq
12	Real Government Purchases of G&S	RG.xls	RGyyQq
13	Real Exports of G&S	REX.xls	REXyyQq
14	Real Imports of G&S	RIMP.xls	RIMPyyQq
15	Price Index for Imports	PIMP.xls	PIMPyyQq
16	Nominal Corporate Profits After Tax	NCPROFAT.xls	NCPROFATyyQq
		Quarterly Average Variables	
17	M1	M1.xls	M1yyQq
18	M2	M2.xls	M2yyQq
19	Total Reserves	TRBASA.xls	TRBASAyyQq
20	Nonborrowed Reserves	NBRBASA.xls	NBRBASAyyQq
21	Nonborrowed Reserves + Ext. Credit	NBRECBASA.xls	NBRECBASAyyQq
22	Monetary Base	BASEBASA.xls	BASEyyQq
23	Consumer Price Index	CPI.xls	CPIyyQq
24	Unemployment Rate	RUC.xls	RUCyyQq
25	Employment	Employees on Nonagricultural Payrolls	
26	Industrial Production	Industrial Production Indexes	

Note: All variables have been transformed in annualised quarterly growth rates

Table A2. List of variables used in the factor approach

		Mean	Minimum	Maximum	Std. Error	Noise/Signal	AR(1)
Real Output	Total revision	**0.66**	-5.63	8.56	2.17	0.61	0.03
	Real-time revision	**0.19**	-1.76	2.21	0.79	0.22	-0.06
	Standard deviation	**0.22**[a]	0.32	1.45	0.24	0.07	**0.16**
	Remaining revision						
	v = 2	**0.48**	-5.49	7.71	2.14	0.60	0.00
	v = 6	**0.48**	-5.32	7.75	1.92	0.54	-0.09
	v = 10	**0.34**	-5.34	7.75	1.95	0.55	-0.14
	v = 20	**0.34**	-4.44	5.19	1.64	0.46	**-0.18**
	v = 40	**0.30**	-4.53	4.37	1.38	0.39	**-0.23**
Inflation (output deflator)	Total revision	0.10	-2.29	2.80	0.90	0.36	0.14
	Real-time revision	**0.11**	-1.19	2.02	0.49	0.20	0.16
	Standard deviation	**0.19**[a]	0.09	1.08	0.22	0.09	**0.45**
	Remaining revision						
	v = 2	-0.01	-3.64	2.18	0.96	0.38	0.08
	v = 6	-0.05	-3.64	2.20	0.89	0.35	0.03
	v = 10	-0.07	-2.55	2.14	0.87	0.34	0.05
	v = 20	**-0.15**	-2.67	1.76	0.76	0.30	-0.05
	v = 40	**-0.14**	-1.88	1.97	0.65	0.26	-0.16

(a) Ratio to standard deviation of the latest available.

Sample: 1965:3-1999:4. Values in bold are significantly different fom zero (5%)

Table 1: Descriptive statistic of revision

	Output growth						Inflation					
	ARMA	VAR	TVC	OKUN	LI	RW	ARMA	VAR	TVC	PH	TS	RW
AIC	0.00	0.15	0.00	0.02	0.59	0.00	0.27	0.00	0.01	0.27	0.22	0.01
SBC	0.10	0.00	0.02	0.11	0.50	0.01	0.17	0.00	0.01	0.27	0.18	0.10
HQ	0.01	0.12	0.01	0.04	0.53	0.01	0.25	0.00	0.01	0.27	0.21	0.02

Table 2: Selection instability

Note: The numbers represent the fraction of times that a model selected with a given criterion using the fully revised data has also been selected in ten years of rolling vintages.

Output growth

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
1-step	ARMA	0.86	4	0.78	4	0.69	4
	BVAR	0.88	5	0.77	3	0.75	5
	TVC	0.97	8	0.92	7	0.88	6
	Okun	0.71	3	1.15	10	0.53	1
	Li	0.94	6	0.88	5	0.95	9
	AC	0.67	1	0.62	1	0.56	2
	AFTER	1.07	11	0.99	9	0.94	8
	EW	0.68	2	0.90	6	0.57	3
	AIC	0.97	7	0.75	2	0.97	10
	SBC	1.00	9	1.17	11	0.92	7
	HQ	1.00	10	0.95	8	1.00	11
	mean	0.89		0.90		0.80	
	median	0.94		0.90		0.88	
4-step	ARMA	0.76	5	0.95	7	0.63	4
	BVAR	0.55	1	0.68	2	0.48	1
	TVC	1.00	8	0.94	6	0.91	11
	Okun	0.80	6	0.97	8	0.91	10
	Li	0.93	7	0.98	9	0.76	6
	AC	0.65	2	0.67	1	0.54	2
	AFTER	0.75	4	0.88	5	0.69	5
	EW	0.66	3	0.80	4	0.54	3
	AIC	1.06	10	1.23	10	0.86	9
	SBC	1.09	11	1.27	11	0.84	7
	HQ	1.02	9	0.68	3	0.86	8
	mean	0.84		0.91		0.73	
	median	0.80		0.94		0.76	
8-step	ARMA	1.24	9	0.87	8	0.89	7
	BVAR	0.70	2	0.83	3	0.52	3
	TVC	1.02	6	0.85	6	0.86	5
	Okun	0.90	4	0.84	4	0.63	4
	Li	1.20	8	0.88	10	0.91	8
	AC	0.67	1	0.75	1	0.42	1
	AFTER	0.97	5	0.87	9	0.88	6
	EW	0.70	3	0.84	5	0.48	2
	AIC	1.18	7	0.76	2	0.91	10
	SBC	1.24	10	0.85	7	0.91	9
	HQ	1.24	11	0.91	11	0.92	11
	mean	1.01		0.84		0.76	
	median	1.02		0.85		0.88	

Inflation

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
1-step	ARMA	0.84	2	0.93	2	0.64	1
	BVAR	0.88	7	0.96	8	0.75	9
	TVC	0.89	8	0.94	3	0.75	10
	Ph	0.96	10	0.96	7	0.74	8
	TS	0.92	9	0.96	6	0.72	7
	AC	0.79	1	0.91	1	0.66	3
	AFTER	0.85	5	0.95	4	0.70	4
	EW	0.86	6	0.96	9	0.65	2
	AIC	0.85	4	0.96	5	0.71	6
	SBC	0.85	3	0.98	10	0.71	5
	HQ	1.03	11	1.03	11	1.02	11
	mean	0.88		0.96		0.73	
	median	0.86		0.96		0.71	
4-step	ARMA	0.65	4	0.83	8	0.46	6
	BVAR	0.50	2	0.84	10	0.48	7
	TVC	0.74	7	0.71	2	0.60	8
	Ph	0.65	5	0.83	5	0.42	4
	TS	0.67	6	0.83	6	0.43	5
	AC	0.49	1	0.71	1	0.38	2
	AFTER	0.76	8	0.83	7	0.41	3
	EW	0.55	3	0.84	11	0.35	1
	AIC	0.93	11	0.83	3	0.84	11
	SBC	0.93	10	0.83	4	0.84	10
	HQ	0.93	9	0.84	9	0.83	9
	mean	0.71		0.81		0.55	
	median	0.67		0.83		0.46	
8-step	ARMA	0.68	5	0.90	10	0.42	4
	BVAR	0.68	4	0.91	11	0.66	7
	TVC	0.75	7	0.87	2	0.60	6
	Ph	0.75	8	0.89	9	0.71	8
	TS	0.69	6	0.89	7	0.38	2
	AC	0.50	1	0.67	1	0.35	1
	AFTER	0.56	3	0.89	6	0.56	5
	EW	0.51	2	0.89	5	0.39	3
	AIC	0.93	11	0.89	8	0.81	11
	SBC	0.93	10	0.89	3	0.81	10
	HQ	0.93	9	0.89	4	0.75	9
	mean	0.72		0.87		0.59	
	median	0.69		0.89		0.60	

Table 3: Accuracy. Average RMSPE

Note: The columns labeled RMSPE show for each forecasting horizon and for each model the average root mean square prediction errors calculated as in Eq. (12) in the text as a ratio of the RMSPE of the RW benchmark. The averages are taken over the vintages. The columns labeled Rank simply order the models according to the lowest-to-highest RMSPE.

Output growth

1-step

Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
ARMA	0.70	5	0.64	5	0.56	4
BVAR	0.72	6	0.62	4	0.61	5
TVC	0.79	9	0.75	8	0.71	6
Okun	0.58	3	0.94	10	0.43	1
Li	0.77	7	0.72	6	0.77	9
AC	0.55	2	0.51	1	0.46	2
AFTER	0.59	4	0.58	2	0.77	8
EW	0.55	1	0.73	7	0.46	3
AIC	0.79	8	0.61	3	0.79	10
SBC	0.81	10	0.96	11	0.75	7
HQ	0.82	11	0.77	9	0.81	11
mean	0.70		0.71		0.65	
median	0.72		0.72		0.71	

4-step

Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
ARMA	0.58	5	0.72	7	0.48	4
BVAR	0.42	1	0.51	2	0.36	1
TVC	0.75	8	0.71	6	0.69	11
Okun	0.61	6	0.73	8	0.69	10
Li	0.71	7	0.74	9	0.57	6
AC	0.49	2	0.51	1	0.41	2
AFTER	0.56	4	0.67	5	0.52	5
EW	0.50	3	0.61	4	0.41	3
AIC	0.80	10	0.93	10	0.65	9
SBC	0.82	11	0.96	11	0.63	7
HQ	0.77	9	0.52	3	0.65	8
mean	0.64		0.69		0.55	
median	0.61		0.71		0.57	

8-step

Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
ARMA	0.84	9	0.59	8	0.60	7
BVAR	0.47	2	0.56	3	0.36	3
TVC	0.69	6	0.58	6	0.58	5
Okun	0.61	4	0.57	4	0.43	4
Li	0.81	8	0.60	10	0.61	8
AC	0.45	1	0.51	1	0.28	1
AFTER	0.66	5	0.59	9	0.60	6
EW	0.48	3	0.57	5	0.33	2
AIC	0.80	7	0.51	2	0.62	10
SBC	0.84	10	0.58	7	0.62	9
HQ	0.84	11	0.62	11	0.63	11
mean	0.68		0.57		0.51	
median	0.69		0.58		0.60	

Inflation

1-step

Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
ARMA	0.68	2	0.76	2	0.52	1
BVAR	0.71	7	0.78	8	0.61	9
TVC	0.73	8	0.76	3	0.61	10
Ph	0.78	10	0.78	7	0.61	8
TS	0.75	9	0.78	6	0.58	7
AC	0.64	1	0.74	1	0.54	3
AFTER	0.69	5	0.77	4	0.57	4
EW	0.70	6	0.78	9	0.53	2
AIC	0.69	4	0.78	5	0.58	6
SBC	0.69	3	0.80	10	0.58	5
HQ	0.84	11	0.84	11	0.83	11
mean	0.72		0.78		0.60	
median	0.70		0.78		0.58	

4-step

Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
ARMA	0.62	4	0.80	8	0.44	6
BVAR	0.49	2	0.81	10	0.46	7
TVC	0.72	7	0.68	2	0.58	8
Ph	0.63	5	0.80	5	0.40	3
TS	0.65	6	0.80	6	0.41	5
AC	0.47	1	0.68	1	0.41	4
AFTER	0.73	8	0.80	7	0.39	2
EW	0.53	3	0.81	11	0.34	1
AIC	0.90	11	0.80	3	0.81	11
SBC	0.90	10	0.80	4	0.81	10
HQ	0.90	9	0.81	9	0.80	9
mean	0.68		0.78		0.53	
median	0.65		0.80		0.44	

8-step

Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
ARMA	0.62	5	0.81	10	0.38	4
BVAR	0.62	4	0.82	11	0.60	7
TVC	0.68	7	0.79	2	0.55	6
Ph	0.68	8	0.81	9	0.64	8
TS	0.63	6	0.80	7	0.35	1
AC	0.45	1	0.60	1	0.36	3
AFTER	0.51	3	0.80	6	0.51	5
EW	0.46	2	0.80	5	0.36	2
AIC	0.84	11	0.80	8	0.73	11
SBC	0.84	10	0.80	3	0.73	10
HQ	0.84	9	0.80	4	0.68	9
mean	0.65		0.79		0.53	
median	0.63		0.80		0.55	

Table 4: Accuracy. Last-vintage RMSPE

Note: The columns labeled RMSPE show for each forecasting horizon and for each model the average root mean square prediction errors calculated as in Eq. (12) in the text as a ratio of the RMSPE of the RW benchmark. The averages are taken over the vintages. The columns labeled Rank simply order the models according to the lowest-to-highest RMSPE.

Output growth

1-step

Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
ARMA	3.51	11	3.16	10	3.38	10
BVAR	3.20	8	2.52	3	3.21	8
TVC	3.49	10	3.34	11	3.50	11
Okun	3.10	5	2.92	8	2.93	4
Li	3.39	9	3.03	9	3.27	9
AC	1.53	2	1.40	2	1.25	1
AFTER	2.17	3	2.66	4	2.09	3
EW	1.50	1	1.35	1	1.46	2
AIC	3.10	4	2.74	5	2.97	5
SBC	3.12	7	2.82	7	2.98	7
HQ	3.11	6	2.77	6	2.98	6
mean	2.84		2.61		2.73	
median	3.11		2.77		2.98	

4-step

Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
ARMA	3.77	10	3.49	10	3.64	9
BVAR	3.71	5	2.77	3	3.71	10
TVC	3.86	11	3.67	11	3.86	11
Okun	3.70	4	3.47	9	3.48	4
Li	3.73	6	3.45	8	3.59	8
AC	1.72	2	1.59	2	1.68	2
AFTER	2.98	3	3.02	4	3.36	3
EW	1.68	1	1.51	1	1.64	1
AIC	3.74	7	3.36	5	3.56	5
SBC	3.76	9	3.41	6	3.57	6
HQ	3.75	8	3.41	7	3.57	7
mean	3.31		3.01		3.24	
median	3.73		3.41		3.57	

8-step

Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
ARMA	3.65	9	3.29	6	3.52	9
BVAR	3.81	10	2.41	3	3.81	10
TVC	3.87	11	3.62	11	3.88	11
Okun	3.60	4	3.35	7	3.38	4
Li	3.64	8	3.29	5	3.51	8
AC	1.71	2	1.57	2	1.66	2
AFTER	3.31	3	3.16	4	3.15	3
EW	1.66	1	1.44	1	1.62	1
AIC	3.63	7	3.41	8	3.46	5
SBC	3.63	6	3.45	10	3.46	6
HQ	3.62	5	3.44	9	3.46	7
mean	3.28		2.95		3.17	
median	3.63		3.29		3.46	

Inflation

1-step

Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
ARMA	1.32	10	1.23	7	1.28	9
BVAR	1.25	4	0.88	3	1.21	4
TVC	1.34	11	1.29	11	1.36	11
Ph	1.32	9	1.27	9	1.28	10
TS	1.32	8	1.29	10	1.27	8
AC	0.60	2	0.54	1	0.59	2
AFTER	0.99	3	1.24	8	1.12	3
EW	0.59	1	0.56302	2	0.58	1
AIC	1.27	7	1.17	4	1.23	7
SBC	1.27	6	1.19	6	1.23	6
HQ	1.27	5	1.18	5	1.23	5
mean	1.14		1.08		1.13	
median	1.27		1.19		1.23	

4-step

Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
ARMA	1.65	8	1.33	4	1.60	8
BVAR	1.78	11	0.98	3	1.79	11
TVC	1.73	10	1.56	11	1.76	10
Ph	1.61	7	1.37	6	1.56	7
TS	1.52	3	1.37	5	1.47	3
AC	0.82	2	0.66	2	0.81	2
AFTER	1.70	9	1.46	10	1.63	9
EW	0.74	1	0.60	1	0.73	1
AIC	1.56	6	1.38	9	1.51	5
SBC	1.56	5	1.37	7	1.51	4
HQ	1.56	4	1.38	8	1.52	6
mean	1.48		1.22		1.44	
median	1.56		1.37		1.52	

8-step

Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
ARMA	1.86	9	1.51	4	1.80	8
BVAR	1.94	10	1.00	3	1.95	9
TVC	2.05	11	1.77	11	2.08	11
Ph	1.76	4	1.53	5	1.70	4
TS	1.70	3	1.57	6	1.65	3
AC	0.97	2	0.79	2	0.97	2
AFTER	1.84	8	1.71	10	1.95	10
EW	0.84	1	0.67	1	0.83	1
AIC	1.76	7	1.69	9	1.71	6
SBC	1.76	6	1.67	7	1.71	5
HQ	1.76	5	1.68	8	1.71	7
mean	1.66		1.42		1.64	
median	1.76		1.57		1.71	

Table 5: Predicted forecast Uncertainty

Note: The columns labeled Simple, Factor and Panel show, for each forecasting horizon and for each model, the average predicted forecasted uncertainty calculated as discussed in the text. The averages are taken over the vintages. The columns labeled Rank simply order the models according to the lowest-to-highest uncertainty.

	Output growth							Inflation						
	Model	SIMPLE	Rank	Factor	Rank	Panel	Rank	Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
1-step	ARMA	0.72	3	1.11	8	0.63	3	ARMA	0.35	2	0.42	10	0.29	3
	BVAR	0.78	6	1.39	10	0.71	6	BVAR	0.37	3	0.56	11	0.32	8
	TVC	0.82	7	0.80	4	0.75	7	TVC	0.39	7	0.41	7	0.32	7
	Okun	0.73	5	0.81	5	0.63	4	Ph	0.40	8	0.41	9	0.31	6
	Li	0.95	8	1.01	6	1.02	8	TS	0.39	6	0.40	5	0.30	5
	AC	0.68	1	0.77	2	0.60	1	AC	0.35	1	0.38	2	0.28	2
	AFTER	0.73	4	0.80	3	0.70	5	AFTER	0.38	5	0.41	8	0.30	4
	EW	0.69	2	0.76	1	0.61	2	EW	0.37	4	0.38718	3	0.28	1
	AIC	1.13	11	1.34	9	1.29	11	AIC	0.41	11	0.40	6	0.34	11
	SBC	1.01	9	1.08	7	1.21	9	SBC	0.41	10	0.38	1	0.34	10
	HQ	1.07	10	1.41	11	1.21	10	HQ	0.41	9	0.39	4	0.33	9
	mean	0.85		1.02		0.85		mean	0.38		0.41		0.31	
	median	0.78		1.01		0.71		median	0.39		0.40		0.31	
	Model	SIMPLE	Rank	Factor	Rank	Panel	Rank	Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
4-step	ARMA	0.69	4	1.19	10	0.61	4	ARMA	0.36	6	0.45	6	0.28	6
	BVAR	0.61	1	1.11	7	0.58	1	BVAR	0.25	1	0.52	8	0.23	1
	TVC	0.79	7	0.80	2	0.72	8	TVC	0.40	7	0.41	3	0.32	8
	Okun	0.76	6	0.94	4	0.67	6	Ph	0.35	4	0.42	4	0.26	4
	Li	0.84	8	1.15	9	0.70	7	TS	0.36	5	0.44	5	0.27	5
	AC	0.67	3	0.79	1	0.61	3	AC	0.30	2	0.37	1	0.24	2
	AFTER	0.69	5	1.01	5	0.61	5	AFTER	0.42	8	0.51	7	0.30	7
	EW	0.66	2	0.82	3	0.60	2	EW	0.32	3	0.38	2	0.25	3
	AIC	1.00	11	1.11	6	0.92	11	AIC	0.46	11	0.56	10	0.41	11
	SBC	0.90	10	1.28	11	0.83	10	SBC	0.46	10	0.58	11	0.41	10
	HQ	0.88	9	1.13	8	0.80	9	HQ	0.46	9	0.53	9	0.40	9
	mean	0.77		1.03		0.70		mean	0.38		0.47		0.31	
	median	0.76		1.11		0.67		median	0.36		0.45		0.28	
	Model	SIMPLE	Rank	Factor	Rank	Panel	Rank	Model	SIMPLE	Rank	Factor	Rank	Panel	Rank
8-step	ARMA	0.75	7	1.38	11	0.66	5	ARMA	0.38	6	0.49	7	0.26	6
	BVAR	0.61	1	1.25	9	0.59	2	BVAR	0.25	1	0.54	8	0.23	2
	TVC	0.71	5	0.76	2	0.66	6	TVC	0.40	7	0.40	4	0.31	7
	Okun	0.72	6	0.91	5	0.66	7	Ph	0.35	4	0.45	5	0.25	4
	Li	0.85	8	1.28	10	0.68	8	TS	0.36	5	0.47	6	0.26	5
	AC	0.65	3	0.74	1	0.58	1	AC	0.30	2	0.37	2	0.22	1
	AFTER	0.67	4	0.77	3	0.59	3	AFTER	0.49	8	0.39	3	0.34	8
	EW	0.65	2	0.83	4	0.60	4	EW	0.31	3	0.36	1	0.24	3
	AIC	0.86	9	1.11	8	0.74	10	AIC	0.50	11	0.61	9	0.49	11
	SBC	0.89	11	1.09	7	0.75	11	SBC	0.50	10	0.82	11	0.49	10
	HQ	0.88	10	1.08	6	0.72	9	HQ	0.50	9	0.66	10	0.41	9
	mean	0.75		1.02		0.66		mean	0.39		0.51		0.32	
	median	0.72		1.08		0.66		median	0.38		0.47		0.26	

Table 6: Actual forecast Uncertainty

Note: The columns labeled Simple, Factor and Panel show for each forecasting horizon and for each model the average actual uncertainty computed by pooling all forecast errors of our repetitive block experiment as explained in the text, from ten years of forecast vintages and corresponding outcomes over the periods 1969q1–1997q1 for h=1, 1969q4–1998q1 for h=4, 1970q4–1999q1 for h=8. The columns labeled Rank simply order the models according to the lowest-to-highest uncertainty.

1-step

	Output growth					
	SIMPLE		Factor		Panel	
Model	b_1	**Rank**	b_1	**Rank**	b_1	**Rank**
ARMA	0.40	4	0.21	6	0.29	6
BVAR	0.17	10	0.20	7	0.14	11
TVC	0.27	5	0.25	4	0.22	9
Okun	0.18	8	0.23	5	0.15	10
Li	0.21	7	0.18	8	0.33	5
AC	1.02	1	0.57	3	1.01	2
AFTER	1.01	2	0.61	2	1.01	1
EW	0.88	3	0.62	1	0.87	3
AIC	0.14	11	0.08	10	0.24	8
SBC	0.24	6	0.17	9	0.34	4
HQ	0.18	9	0.06	11	0.27	7
mean	0.43		0.29		0.44	
median	0.24		0.21		0.29	

	Inflation					
	SIMPLE		Factor		Panel	
Model	b_1	**Rank**	b_1	**Rank**	b_1	**Rank**
ARMA	0.12	11	-0.01	11	0.13	11
BVAR	0.17	6	0.18	6	0.17	10
TVC	0.19	4	0.17	7	0.17	9
Ph	0.17	7	0.04	10	0.20	4
TS	0.17	5	0.12	8	0.17	8
AC	0.92	3	0.86	2	0.92	3
AFTER	0.97	2	0.83	3	0.94	2
EW	1.07	1	1.12957	1	1.08	1
AIC	0.14	8	0.25	4	0.18	6
SBC	0.14	9	0.11	9	0.18	7
HQ	0.14	10	0.22	5	0.18	5
mean	0.38		0.35		0.39	
median	0.17		0.18		0.18	

4-step

	Output growth					
	SIMPLE		Factor		Panel	
Model	b_1	**Rank**	b_1	**Rank**	b_1	**Rank**
ARMA	0.27	4	0.02	11	0.34	4
BVAR	0.02	11	0.14	9	0.05	11
TVC	0.22	5	0.23	5	0.26	5
Okun	0.11	7	0.16	8	0.17	6
Li	0.09	8	0.14	10	0.11	8
AC	0.36	3	0.58	1	0.41	3
AFTER	0.37	2	0.57	2	0.42	2
EW	0.58	1	0.21	6	0.50	1
AIC	0.05	9	0.29	3	0.05	10
SBC	0.15	6	0.20	7	0.16	7
HQ	0.04	10	0.25	4	0.08	9
mean	0.21		0.25		0.23	
median	0.15		0.21		0.17	

	Inflation					
	SIMPLE		Factor		Panel	
Model	b_1	**Rank**	b_1	**Rank**	b_1	**Rank**
ARMA	0.02	5	0.18	8	0.07	6
BVAR	-0.04	11	0.25	4	0.01	11
TVC	0.05	4	0.10	11	0.08	5
Ph	-0.02	7	0.15	9	0.05	9
TS	-0.02	6	0.14	10	0.02	10
AC	0.66	2	0.32	2	0.68	2
AFTER	0.59	3	0.27	3	0.61	3
EW	1.06	1	0.65	1	0.89	1
AIC	-0.03	8	0.22	5	0.06	7
SBC	-0.03	9	0.19	7	0.06	8
HQ	-0.03	10	0.19	6	0.10	4
mean	0.20		0.24		0.24	
median	-0.02		0.19		0.07	

8-step

	Output growth					
	SIMPLE		Factor		Panel	
Model	b_1	**Rank**	b_1	**Rank**	b_1	**Rank**
ARMA	0.41	2	0.10	11	0.24	5
BVAR	-0.08	11	0.13	9	-0.06	11
TVC	0.40	3	0.32	5	0.35	1
Okun	0.17	10	0.13	10	0.11	10
Li	0.21	9	0.37	3	0.22	7
AC	0.53	1	0.42	2	0.26	2
AFTER	0.29	4	0.32	4	0.25	3
EW	0.29	5	0.43	1	0.25	4
AIC	0.22	8	0.28	6	0.20	9
SBC	0.27	6	0.21	8	0.20	8
HQ	0.24	7	0.23	7	0.23	6
mean	0.27		0.27		0.20	
median	0.27		0.28		0.23	

	Inflation					
	SIMPLE		Factor		Panel	
Model	b_1	**Rank**	b_1	**Rank**	b_1	**Rank**
ARMA	0.07	5	0.23	7	0.04	5
BVAR	-0.05	11	0.34	3	0.00	11
TVC	0.07	4	0.08	10	0.04	4
Ph	-0.01	10	0.22	8	0.01	8
TS	0.02	9	0.01	11	0.03	7
AC	0.13	2	0.41	1	0.16	2
AFTER	0.10	3	0.30	5	0.15	3
EW	0.37	1	0.40	2	0.27	1
AIC	0.03	6	0.32	4	0.01	9
SBC	0.03	7	0.14	9	0.01	10
HQ	0.03	8	0.24	6	0.03	6
mean	0.07		0.24		0.07	
median	0.03		0.24		0.03	

Table 7: Data Revisions pass-through to Forecast Revisions

Note: The columns labeled b_1 illustrate for each forecasting horizon and for each model the OLS estimate of the "pass-through" coefficient in equation (14). The columns labeled Rank simply order the models according to the highest-to-lowest estimate.

	Output growth							Inflation						
		SIMPLE		Factor		Panel			SIMPLE		Factor		Panel	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.76	10	0.69	5	0.75	10	ARMA	0.89	2	0.96	10	0.68	1
	BVAR	0.71	7	0.82	11	0.71	6	BVAR	0.93	7	1.31	11	0.79	10
	TVC	0.76	11	0.74	10	0.76	11	TVC	0.95	8	0.54	2	0.80	11
	Okun	0.74	9	0.65	2	0.74	9	Ph	1.02	10	0.70	8	0.79	9
1-step	Li	0.73	8	0.70	7	0.73	8	TS	0.97	9	0.67	7	0.76	8
	AC	0.69	1	0.65	3	0.66	1	AC	0.84	1	0.51	1	0.70	3
	AFTER	0.70	6	0.66	4	0.72	7	AFTER	0.91	5	0.58	3	0.74	5
	EW	0.70	2	0.64	1	0.70	2	EW	0.91	6	0.62	5	0.70	2
	AIC	0.70	4	0.72	9	0.71	3	AIC	0.91	4	0.61	4	0.76	7
	SBC	0.70	5	0.70	6	0.71	5	SBC	0.91	3	0.81	9	0.76	6
	HQ	0.70	3	0.71	8	0.71	4	HQ	1.10	11	0.64	6	0.74	4
	mean	0.72		0.70		0.72		mean	0.94		0.72		0.75	
	median	0.70		0.70		0.71		median	0.91		0.64		0.76	

		SIMPLE		Factor		Panel			SIMPLE		Factor		Panel	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.92	6	0.98	9	0.92	6	ARMA	0.79	3	1.38	11	0.56	4
	BVAR	0.89	1	1.07	11	0.89	1	BVAR	0.36	1	0.86	4	0.66	5
	TVC	0.96	11	0.95	7	0.95	11	TVC	0.91	7	0.62	1	0.73	6
	Okun	0.90	4	0.90	3	0.90	3	Ph	0.80	4	1.13	8	0.51	2
4-step	Li	0.93	7	0.92	4	0.93	7	TS	0.84	5	1.16	9	0.52	3
	AC	0.89	2	0.88	1	0.90	2	AC	0.89	6	1.08	7	0.77	7
	AFTER	0.92	5	0.94	5	0.91	5	AFTER	0.92	8	1.18	10	0.87	8
	EW	0.90	3	0.89	2	0.90	4	EW	0.68	2	0.95	6	0.44	1
	AIC	0.95	8	1.00	10	0.95	8	AIC	1.14	11	0.78	2	1.03	11
	SBC	0.95	10	0.95	6	0.95	10	SBC	1.14	10	0.81	3	1.03	10
	HQ	0.95	9	0.97	8	0.95	9	HQ	1.14	9	0.88	5	1.02	9
	mean	0.92		0.95		0.92		mean	0.87		0.99		0.74	
	median	0.92		0.95		0.92		median	0.89		0.95		0.73	

		SIMPLE		Factor		Panel			SIMPLE		Factor		Panel	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.89	5	1.03	7	0.89	6	ARMA	0.73	5	1.08	5	0.44	4
	BVAR	0.87	2	1.25	11	0.87	3	BVAR	0.74	7	1.09	6	0.55	6
	TVC	0.96	11	0.98	4	0.96	11	TVC	0.80	8	0.60	1	0.64	7
	Okun	0.91	9	1.00	5	0.90	9	Ph	0.63	2	1.33	11	0.52	5
8-step	Li	0.92	10	1.07	10	0.91	10	TS	0.74	6	1.28	9	0.41	3
	AC	0.85	1	0.96	2	0.84	1	AC	0.65	3	0.89	3	0.37	2
	AFTER	0.88	4	0.98	3	0.87	4	AFTER	0.68	4	0.86	2	0.80	8
	EW	0.87	3	0.95	1	0.86	2	EW	0.56	1	1.20	7	0.31	1
	AIC	0.89	6	1.02	6	0.89	5	AIC	0.99	11	1.29	10	0.86	11
	SBC	0.90	8	1.03	9	0.90	8	SBC	0.99	10	1.05	4	0.86	10
	HQ	0.90	7	1.03	8	0.89	7	HQ	0.99	9	1.22	8	0.80	9
	mean	0.89		1.03		0.89		mean	0.77		1.08		0.60	
	median	0.89		1.02		0.89		median	0.74		1.09		0.55	

Table 8: RMSPE - actual 5 years

Note: The columns labeled RMSPE show for each forecasting horizon and for each model the average root mean square prediction errors calculated as in Eq. (12) in the text as a ratio of the RMSPE of the RW benchmark, and assuming as fully-revised data the values available 5 years after each block of vintages. The averages are taken over the vintages. The columns labeled Rank simply order the models according to the lowest-to-highest RMSPE.

Output growth

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
1-step	ARMA	0.73	10	0.69	5	0.72	10
	BVAR	0.69	7	0.83	11	0.69	7
	TVC	0.74	11	0.72	9	0.74	11
	Okun	0.72	9	0.66	3	0.72	9
	Li	0.71	8	0.71	6	0.72	8
	AC	0.69	5	0.65	2	0.69	2
	AFTER	0.69	6	0.66	4	0.69	6
	EW	0.68	1	0.64	1	0.67	1
	AIC	0.68	4	0.72	10	0.69	5
	SBC	0.68	2	0.71	7	0.69	4
	HQ	0.68	3	0.71	8	0.69	3
	mean	0.70		0.70		0.70	
	median	0.69		0.71		0.69	

Inflation

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
1-step	ARMA	0.89	2	0.96	10	0.68	3
	BVAR	0.93	7	1.31	11	0.79	10
	TVC	0.95	8	0.54	2	0.80	11
	Ph	1.02	10	0.70	8	0.79	9
	TS	0.97	9	0.67	7	0.76	8
	AC	0.84	1	0.51	1	0.70	5
	AFTER	0.91	5	0.58	3	0.63	2
	EW	0.91	6	0.62	5	0.70	4
	AIC	0.91	4	0.61	4	0.76	7
	SBC	0.91	3	0.81	9	0.76	6
	HQ	1.10	11	0.64	6	0.63	1
	mean	0.94		0.72		0.73	
	median	0.91		0.64		0.76	

Output growth

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
4-step	ARMA	0.90	6	0.97	9	0.90	6
	BVAR	0.85	1	1.08	11	0.85	1
	TVC	0.94	11	0.94	6	0.94	11
	Okun	0.87	3	0.89	3	0.87	2
	Li	0.90	7	0.91	4	0.90	7
	AC	0.87	2	0.88	1	0.87	3
	AFTER	0.87	5	0.93	5	0.87	4
	EW	0.87	4	0.88	2	0.88	5
	AIC	0.92	8	0.99	10	0.92	8
	SBC	0.93	10	0.94	7	0.93	10
	HQ	0.93	9	0.96	8	0.93	9
	mean	0.90		0.94		0.90	
	median	0.90		0.94		0.90	

Inflation

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
4-step	ARMA	0.79	3	1.38	10	0.56	5
	BVAR	0.36	1	1.78	11	0.20	1
	TVC	0.91	7	0.62	1	0.73	6
	Ph	0.80	4	1.13	7	0.51	3
	TS	0.84	5	1.16	8	0.52	4
	AC	0.89	6	1.08	6	0.77	7
	AFTER	0.92	8	1.18	9	1.33	11
	EW	0.68	2	0.95	5	0.44	2
	AIC	1.14	11	0.78	2	1.03	10
	SBC	1.14	10	0.81	3	1.03	9
	HQ	1.14	9	0.88	4	1.02	8
	mean	0.87		1.07		0.74	
	median	0.89		1.08		0.73	

Output growth

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
8-step	ARMA	0.88	5	1.01	7	0.87	5
	BVAR	0.83	2	1.25	11	0.83	2
	TVC	0.92	11	0.94	4	0.92	11
	Okun	0.90	8	0.99	5	0.90	9
	Li	0.92	10	1.05	10	0.91	10
	AC	0.82	1	0.94	3	0.82	1
	AFTER	0.85	4	0.94	2	0.84	3
	EW	0.85	3	0.93	1	0.84	4
	AIC	0.89	6	1.01	6	0.88	6
	SBC	0.90	9	1.02	9	0.89	8
	HQ	0.90	7	1.02	8	0.89	7
	mean	0.88		1.01		0.87	
	median	0.89		1.01		0.88	

Inflation

Horizon	Model	SIMPLE RMSPE	SIMPLE Rank	Factor RMSPE	Factor Rank	Panel RMSPE	Panel Rank
8-step	ARMA	0.73	5	0.64	5	0.44	5
	BVAR	0.31	1	0.65	6	0.12	1
	TVC	0.80	7	0.60	2	0.64	8
	Ph	0.66	4	0.53	1	0.56	7
	TS	0.74	6	0.84	10	0.41	4
	AC	0.65	3	0.61	3	0.37	3
	AFTER	1.05	11	0.72	7	0.52	6
	EW	0.56	2	0.77	8	0.31	2
	AIC	0.99	10	0.85	11	0.86	11
	SBC	0.99	9	0.61	4	0.86	10
	HQ	0.99	8	0.80	9	0.80	9
	mean	0.77		0.69		0.54	
	median	0.74		0.65		0.52	

Table 9: RMSPE - actual 10 years

Note: The columns labeled RMSPE show for each forecasting horizon and for each model the average root mean square prediction errors calculated as in Eq. (12) in the text as a ratio of the RMSPE of the RW benchmark, and assuming as fully-revised data the values available 10 years after each block of vintages. The averages are taken over the vintages. The columns labeled Rank simply order the models according to the lowest-to-highest RMSPE.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
		SIMPLE		Factor		Panel		SIMPLE		Factor		Panel	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.94	10	0.84	4	0.93	10	0.68	4	0.95	7	0.68	4
	BVAR	0.89	6	0.96	11	0.89	6	0.69	5	1.22	11	0.69	6
	TVC	0.95	11	0.91	10	0.94	11	0.70	6	0.69	1	0.70	7
	Okun	0.92	9	0.80	2	0.92	8	0.68	3	0.85	5	0.67	3
1-step	Li	0.89	7	0.89	9	0.90	7	0.83	11	0.88	6	0.82	11
	AC	0.86	1	0.80	3	0.86	1	0.67	2	0.78	2	0.66	2
	AFTER	0.92	8	0.84	5	0.93	9	0.71	7	0.85	4	0.68	5
	EW	0.86	2	0.80	1	0.86	2	0.67	1	0.80	3	0.66	1
	AIC	0.87	4	0.86	8	0.87	3	0.77	10	1.02	10	0.76	10
	SBC	0.87	5	0.85	6	0.88	5	0.73	8	0.97	8	0.72	8
	HQ	0.87	3	0.86	7	0.88	4	0.75	9	1.00	9	0.73	9
	mean	0.89		0.86		0.90		0.72		0.91		0.71	
	median	0.89		0.85		0.89		0.70		0.88		0.69	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
		SIMPLE		Factor		Panel		SIMPLE		Factor		Panel	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.89	6	0.97	8	0.88	6	0.67	7	0.70	5	0.67	7
	BVAR	0.84	2	1.05	11	0.84	2	0.63	1	0.88	11	0.63	1
	TVC	0.94	11	0.93	6	0.94	11	0.65	3	0.64	1	0.65	3
	Okun	0.84	1	0.85	1	0.84	1	0.73	11	0.72	6	0.73	11
4-step	Li	0.89	7	0.89	4	0.89	7	0.70	10	0.78	9	0.70	10
	AC	0.85	3	0.87	2	0.85	3	0.65	4	0.66	2	0.65	4
	AFTER	0.86	4	0.93	5	0.86	4	0.65	2	0.67	4	0.63	2
	EW	0.86	5	0.88	3	0.87	5	0.65	5	0.66	3	0.65	5
	AIC	0.92	8	0.98	10	0.93	10	0.66	6	0.87	10	0.66	6
	SBC	0.93	10	0.94	7	0.93	8	0.68	9	0.75	7	0.69	9
	HQ	0.93	9	0.97	9	0.93	9	0.67	8	0.75	8	0.67	8
	mean	0.89		0.93		0.89		0.67		0.73		0.67	
	median	0.89		0.93		0.88		0.66		0.72		0.66	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
		SIMPLE		Factor		Panel		SIMPLE		Factor		Panel	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.87	6	0.96	5	0.86	5	0.57	7	0.81	10	0.57	6
	BVAR	0.81	3	1.23	11	0.81	2	0.57	9	0.91	11	0.57	9
	TVC	0.90	10	0.92	3	0.90	11	0.58	10	0.60	2	0.58	10
	Okun	0.72	1	0.96	6	0.89	6	0.51	1	0.71	5	0.57	7
8-step	Li	0.90	8	1.01	10	0.89	8	0.63	11	0.80	9	0.62	11
	AC	0.80	2	0.91	1	0.79	1	0.54	4	0.63	3	0.54	3
	AFTER	0.83	4	0.92	4	0.83	3	0.57	8	0.60	1	0.57	8
	EW	0.83	5	0.91	2	0.83	4	0.55	5	0.66	4	0.55	5
	AIC	0.89	7	0.97	8	0.89	7	0.53	2	0.73	6	0.52	1
	SBC	0.90	11	0.97	9	0.89	10	0.55	6	0.74	7	0.54	4
	HQ	0.90	9	0.97	7	0.89	9	0.54	3	0.74	8	0.53	2
	mean	0.79		0.98		0.86		0.51		0.72		0.56	
	median	0.87		0.96		0.89		0.55		0.73		0.57	

Table 10: Output Growth, Average RMSPE pre- and post-84

Note: The columns labeled RMSPE show for each forecasting horizon and for each model the average root mean square prediction errors calculated as in Eq. (12) in the text as a ratio of the RMSPE of the RW benchmark. The averages are taken across the vintages, over two different sub-samples. The columns labeled Rank simply order the models according to the lowest-to-highest RMSPE.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
		SIMPLE		**Factor**		**Panel**		**SIMPLE**		**Factor**		**Panel**	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.97	2	1.13	10	0.76	1	0.59	4	0.92	10	0.44	1
	BVAR	1.02	7	1.59	11	0.86	7	0.64	9	1.06	11	0.55	9
	TVC	1.09	10	0.70	4	0.92	9	0.58	3	0.60	2	0.48	4
	Ph	1.09	9	0.80	8	0.87	8	0.72	11	0.80	7	0.54	8
1-step	TS	1.05	8	0.74	6	0.85	4	0.66	10	0.81	8	0.50	7
	AC	0.95	1	0.66	2	0.78	3	0.54	1	0.59	1	0.45	2
	AFTER	1.01	6	0.66	1	1.29	11	0.60	7	0.72	5	0.60	11
	EW	1.01	3	0.73	5	0.77	2	0.61	8	0.71	4	0.46	3
	AIC	1.01	5	0.69	3	0.85	6	0.60	6	0.70	3	0.49	6
	SBC	1.01	4	0.95	9	0.85	5	0.60	5	0.85	9	0.49	5
	HQ	1.34	11	0.74	7	1.29	10	0.55	2	0.75	6	0.60	10
	mean	1.05		0.85		0.92		0.61		0.77		0.51	
	median	1.01		0.74		0.85		0.60		0.75		0.49	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
		SIMPLE		**Factor**		**Panel**		**SIMPLE**		**Factor**		**Panel**	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.81	3	0.99	10	0.55	3	0.58	6	0.89	10	0.44	6
	BVAR	0.28	1	1.54	11	0.19	1	0.33	1	1.04	11	0.16	1
	TVC	0.98	6	0.29	1	0.80	6	0.63	7	0.24	1	0.50	7
	Ph	0.88	4	0.74	6	0.56	4	0.52	5	0.71	9	0.35	4
4-step	TS	0.95	5	0.84	8	0.57	5	0.52	4	0.66	8	0.36	5
	AC	1.04	7	0.75	7	0.95	7	0.48	3	0.49	4	0.35	3
	AFTER	1.82	11	0.92	9	1.71	11	1.37	11	0.59	6	1.39	11
	EW	0.71	2	0.54	5	0.45	2	0.47	2	0.55	5	0.31	2
	AIC	1.26	10	0.40	3	1.21	10	0.75	10	0.40	2	0.60	9
	SBC	1.26	9	0.40	4	1.21	9	0.75	9	0.46	3	0.60	8
	HQ	1.26	8	0.36	2	1.17	8	0.75	8	0.62	7	0.63	10
	mean	1.02		0.71		0.85		0.65		0.60		0.52	
	median	0.98		0.74		0.80		0.58		0.59		0.44	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
		SIMPLE		**Factor**		**Panel**		**SIMPLE**		**Factor**		**Panel**	
	Model	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank	RMSPE	Rank
	ARMA	0.85	6	1.71	10	0.51	5	0.42	5	1.83	10	0.27	5
	BVAR	0.20	1	1.82	11	0.11	1	0.29	1	1.94	11	0.10	1
	TVC	0.87	7	0.10	1	0.72	7	0.54	7	0.84	9	0.42	7
	Ph	0.75	4	0.87	6	0.67	6	0.41	4	0.51	6	0.25	4
8-step	TS	0.84	5	0.70	3	0.41	4	0.46	6	0.55	7	0.31	6
	AC	0.73	3	0.91	7	0.40	3	0.36	3	0.31	2	0.23	3
	AFTER	2.95	11	1.06	8	2.84	11	0.93	11	0.66	8	1.03	11
	EW	0.60	2	0.63	2	0.34	2	0.35	2	0.41	4	0.20	2
	AIC	1.02	10	1.41	9	0.96	10	0.71	10	0.11	1	0.54	9
	SBC	1.02	9	0.77	5	0.96	9	0.71	9	0.42	5	0.54	8
	HQ	1.02	8	0.75	4	0.84	8	0.71	8	0.39	3	0.57	10
	mean	0.99		0.97		0.80		0.54		0.72		0.40	
	median	0.85		0.87		0.67		0.46		0.51		0.31	

Table 11: Inflation, Average RMSPE pre- and post-84

Note: The columns labeled RMSPE show for each forecasting horizon and for each model the average root mean square prediction errors calculated as in Eq. (12) in the text as a ratio of the RMSPE of the RW benchmark. The averages are taken across the vintages, over two different sub-samples. The columns labeled Rank simply order the models according to the lowest-to-highest RMSPE.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
1-step	ARMA	3.79	10	3.36	10	3.65	9	3.19	11	2.93	10	3.08	10
	BVAR	3.40	8	2.68	4	3.41	8	2.98	9	2.35	3	2.99	9
	TVC	3.78	9	3.62	11	3.79	11	3.17	10	3.02	11	3.18	11
	Okun	3.33	5	3.08	8	3.15	4	2.85	6	2.73	9	2.69	4
	Li	3.80	11	3.33	9	3.66	10	2.94	8	2.70	7	2.84	8
	AC	1.67	2	1.51	2	1.63	2	1.36	2	1.26	2	1.33	2
	AFTER	1.95	3	2.61	3	2.10	3	1.58	3	2.71	8	1.90	3
	EW	1.62	1	1.45	1	1.58	1	1.36	1	1.23	1	1.32	1
	AIC	3.33	4	2.89	5	3.17	5	2.84	4	2.57	4	2.73	5
	SBC	3.35	7	2.98	7	3.19	7	2.86	7	2.63	6	2.75	7
	HQ	3.34	6	2.91	6	3.18	6	2.84	5	2.62	5	2.74	6
	mean	3.03		2.77		2.96		2.54		2.43		2.51	
	median	3.34		2.91		3.18		2.85		2.63		2.74	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
4-step	ARMA	3.92	10	3.65	10	3.78	9	3.54	10	3.27	7	3.41	9
	BVAR	3.89	6	2.94	3	3.89	10	3.44	4	2.55	3	3.45	10
	TVC	4.00	11	3.89	11	4.00	11	3.63	11	3.37	11	3.63	11
	Okun	3.87	5	3.59	9	3.64	4	3.45	5	3.29	8	3.24	4
	Li	3.85	4	3.58	8	3.71	8	3.52	8	3.26	5	3.39	8
	AC	1.70	2	1.57	2	1.66	2	1.36	2	1.25	2	1.33	2
	AFTER	2.71	3	3.43	7	3.00	3	2.70	3	2.67	4	2.95	3
	EW	1.62	1	1.45	1	1.58	1	1.36	1	1.23	1	1.32	1
	AIC	3.90	8	3.40	4	3.69	6	3.50	6	3.26	6	3.36	5
	SBC	3.90	7	3.42	6	3.68	5	3.53	9	3.33	9	3.39	7
	HQ	3.91	9	3.42	5	3.70	7	3.52	7	3.33	10	3.38	6
	mean	3.39		3.12		3.30		3.05		2.80		2.99	
	median	3.89		3.42		3.69		3.50		3.26		3.38	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
8-step	ARMA	3.62	8	3.37	4	3.50	8	3.49	6	3.10	7	3.37	6
	BVAR	3.85	11	2.43	3	3.85	11	3.57	10	2.31	3	3.58	10
	TVC	3.83	10	3.69	10	3.84	10	3.72	11	3.41	11	3.73	11
	Okun	3.53	4	3.46	9	3.33	4	3.46	4	3.09	6	3.26	4
	Li	3.63	9	3.41	6	3.50	9	3.48	5	3.05	4	3.36	5
	AC	1.74	2	1.69	2	1.70	2	1.36	2	1.28	2	1.33	2
	AFTER	2.91	3	3.71	11	3.02	3	2.94	3	3.07	5	3.01	3
	EW	1.62	1	1.45	1	1.58	1	1.36	1	1.23	1	1.32	1
	AIC	3.57	5	3.41	5	3.37	5	3.52	9	3.26	8	3.39	9
	SBC	3.59	7	3.44	8	3.40	7	3.50	7	3.30	9	3.37	7
	HQ	3.58	6	3.42	7	3.39	6	3.50	8	3.30	10	3.38	8
	mean	3.23		3.04		3.13		3.08		2.76		3.01	
	median	3.58		3.41		3.39		3.49		3.09		3.37	

Table 12: Output Growth, Predicted Uncertainty, pre- and post-84

Note: The columns labeled Simple, Factor and Panel show, for each forecasting horizon and for each model, the average predicted forecasted uncertainty calculated as discussed in the text. The averages are taken across the vintages, over two different sub-samples. The columns labeled Rank simply order the models according to the lowest-to-highest uncertainty.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	1.49	10	1.42	7	1.44	9	1.14	7	1.01	4	1.10	6
	BVAR	1.37	4	1.04	3	1.38	3	1.12	6	0.69	3	1.13	9
	TVC	1.52	11	1.48	10	1.54	10	1.15	9	1.09	8	1.16	10
	Ph	1.47	9	1.44	8	1.42	8	1.16	10	1.09	9	1.12	8
1-step	TS	1.47	8	1.47	9	1.42	7	1.15	8	1.10	10	1.11	7
	AC	0.68	2	0.64	2	0.66	2	0.46	1	0.39	1	0.45	1
	AFTER	1.24	3	1.49	11	1.56	11	1.17	11	1.11	11	1.18	11
	EW	0.65	1	0.62	1	0.64	1	0.51	2	0.45	2	0.50	2
	AIC	1.43	7	1.31	4	1.38	6	1.10	5	1.02	5	1.07	5
	SBC	1.43	6	1.34	6	1.38	5	1.10	4	1.03	7	1.07	4
	HQ	1.43	5	1.32	5	1.38	4	1.10	3	1.03	6	1.07	3
	mean	1.29		1.23		1.29		1.02		0.91		1.00	
	median	1.43		1.34		1.38		1.12		1.03		1.10	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	1.77	9	1.58	8	1.71	9	1.48	9	1.05	4	1.44	9
	BVAR	1.87	11	1.12	3	1.87	10	1.64	11	0.82	3	1.65	11
	TVC	1.86	10	1.73	11	1.89	11	1.56	10	1.35	11	1.58	10
	Ph	1.74	8	1.60	10	1.68	8	1.44	8	1.11	5	1.39	8
4-step	TS	1.71	7	1.59	9	1.65	7	1.29	4	1.11	6	1.25	4
	AC	0.69	2	0.67	2	0.68	2	0.51	2	0.45	2	0.51	2
	AFTER	1.44	3	1.43	4	1.45	3	1.14	3	1.11	7	1.15	3
	EW	0.65	1	0.62	1	0.64	1	0.51	1	0.45	1	0.50	1
	AIC	1.69	6	1.55	6	1.63	5	1.40	7	1.19	10	1.36	6
	SBC	1.69	5	1.55	5	1.63	4	1.40	6	1.17	8	1.36	5
	HQ	1.69	4	1.55	7	1.64	6	1.40	5	1.19	9	1.36	7
	mean	1.53		1.36		1.50		1.25		1.00		1.23	
	median	1.69		1.55		1.64		1.40		1.11		1.36	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	1.88	9	1.58	5	1.82	9	1.73	5	1.38	5	1.68	5
	BVAR	1.96	10	1.08	3	1.97	10	1.81	10	0.89	3	1.81	10
	TVC	2.07	11	1.78	11	2.11	11	1.91	11	1.67	9	1.94	11
	Ph	1.67	4	1.60	8	1.62	4	1.74	9	1.40	6	1.68	6
8-step	TS	1.73	8	1.63	10	1.68	8	1.59	4	1.42	7	1.54	4
	AC	0.72	2	0.68	2	0.71	2	0.52	2	0.46	2	0.51	2
	AFTER	1.51	3	1.57	4	1.51	3	1.13	3	1.11	4	1.10	3
	EW	0.65	1	0.62	1	0.64	1	0.51	1	0.45	1	0.50	1
	AIC	1.69	7	1.61	9	1.64	6	1.74	8	1.68	11	1.69	9
	SBC	1.69	6	1.60	7	1.64	5	1.74	7	1.66	8	1.69	8
	HQ	1.69	5	1.60	6	1.65	7	1.74	6	1.68	10	1.69	7
	mean	1.57		1.40		1.54		1.47		1.25		1.44	
	median	1.69		1.60		1.64		1.74		1.40		1.68	

Table 13: Inflation, Predicted Uncertainty, pre- and post-84

Note: The columns labeled Simple, Factor and Panel show, for each forecasting horizon and for each model, the average predicted forecasted uncertainty calculated as discussed in the text. The averages are taken across the vintages, over two different sub-samples. The columns labeled Rank simply order the models according to the lowest-to-highest uncertainty.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	0.38	2	0.42	6	0.38	1	0.21	3	0.25	10	0.21	5
	BVAR	0.40	5	0.43	9	0.39	4	0.22	9	0.27	11	0.22	7
	TVC	0.40	6	0.40	3	0.40	6	0.23	11	0.23	6	0.23	8
	Okun	0.41	9	0.40	5	0.39	3	0.21	4	0.22	3	0.21	3
1-step	Li	0.41	8	0.45	11	0.41	8	0.21	2	0.22	5	0.22	6
	AC	0.38	1	0.39	2	0.40	7	0.21	6	0.21	1	0.19	1
	AFTER	0.40	4	0.40	4	0.39	5	0.21	5	0.21	2	0.21	4
	EW	0.39	3	0.38	1	0.38	2	0.21	1	0.22	4	0.21	2
	AIC	0.41	10	0.43	10	0.43	9	0.22	7	0.24	8	0.23	10
	SBC	0.41	7	0.43	7	0.48	10	0.22	10	0.25	9	0.24	11
	HQ	0.42	11	0.43	8	0.67	11	0.22	8	0.23	7	0.23	9
	mean	0.40		0.42		0.43		0.22		0.23		0.22	
	median	0.40		0.42		0.40		0.21		0.23		0.22	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	0.39	2	0.44	11	0.39	2	0.21	1	0.26	11	0.21	1
	BVAR	0.38	1	0.40	4	0.38	1	0.21	2	0.23	5	0.21	5
	TVC	0.42	8	0.43	8	0.41	6	0.23	10	0.23	8	0.22	9
	Okun	0.47	10	0.40	3	0.42	9	0.21	5	0.24	9	0.21	3
4-step	Li	0.42	7	0.43	9	0.41	7	0.22	8	0.25	10	0.22	7
	AC	0.40	3	0.39	2	0.40	4	0.22	6	0.21	1	0.21	6
	AFTER	0.41	5	0.42	7	0.40	5	0.21	3	0.23	7	0.21	2
	EW	0.40	4	0.39	1	0.40	3	0.21	4	0.23	6	0.21	4
	AIC	0.41	6	0.42	6	0.43	10	0.23	9	0.22	4	0.24	10
	SBC	0.45	9	0.43	10	0.51	11	0.22	7	0.22	2	0.22	8
	HQ	0.48	11	0.42	5	0.41	8	0.23	11	0.22	3	0.25	11
	mean	0.42		0.42		0.41		0.22		0.23		0.22	
	median	0.41		0.42		0.41		0.22		0.23		0.21	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	0.39	5	0.47	10	0.39	5	0.23	7	0.26	10	0.23	7
	BVAR	0.38	2	0.40	1	0.38	2	0.22	3	0.30	11	0.22	3
	TVC	0.40	8	0.41	3	0.40	9	0.22	4	0.22	2	0.22	5
	Okun	0.42	11	0.52	11	0.42	11	0.23	9	0.24	8	0.23	8
8-step	Li	0.40	7	0.44	8	0.39	8	0.23	6	0.24	9	0.23	6
	AC	0.38	1	0.42	5	0.39	3	0.21	1	0.23	6	0.21	1
	AFTER	0.38	3	0.41	2	0.39	4	0.22	2	0.22	1	0.22	4
	EW	0.38	4	0.42	4	0.38	1	0.22	5	0.24	7	0.22	2
	AIC	0.40	6	0.43	7	0.39	7	0.24	10	0.23	3	0.24	9
	SBC	0.40	9	0.44	9	0.40	10	0.23	8	0.23	4	0.25	11
	HQ	0.40	10	0.43	6	0.39	6	0.24	11	0.23	5	0.24	10
	mean	0.39		0.44		0.39		0.23		0.24		0.23	
	median	0.40		0.43		0.39		0.23		0.23		0.23	

Table 14: Output Growth, Actual Uncertainty, pre- and post-84

Note: The columns labeled Simple, Factor and Panel show for each forecasting horizon and for each model the average actual uncertainty computed by pooling all forecast errors of our repetitive block experiment as explained in the text, from ten years of forecast vintages and corresponding outcomes over two sub-samples. The columns labeled Rank simply order the models according to the lowest-to-highest uncertainty.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	0.21	3	0.24	7	0.23	9	0.18	2	0.19	8	0.17	1
	BVAR	0.22	9	0.25	10	0.21	5	0.19	6	0.20	9	0.18	7
	TVC	0.23	11	0.27	11	0.22	7	0.20	11	0.20	11	0.18	11
	Ph	0.21	4	0.23	6	0.23	8	0.20	10	0.18	5	0.17	6
1-step	TS	0.21	2	0.22	3	0.21	3	0.19	7	0.18	3	0.17	4
	AC	0.21	6	0.22	5	0.22	6	0.18	1	0.18	6	0.17	2
	AFTER	0.21	5	0.21	1	0.20	1	0.19	9	0.20	10	0.18	8
	EW	0.21	1	0.21	2	0.21	4	0.19	8	0.18	7	0.17	5
	AIC	0.22	7	0.22	4	0.21	2	0.19	5	0.18	1	0.18	10
	SBC	0.22	10	0.24	8	0.23	10	0.19	4	0.18	2	0.18	9
	HQ	0.22	8	0.25	9	0.24	11	0.19	3	0.18	4	0.17	3
	mean	0.22		0.23		0.22		0.19		0.19		0.17	
	median	0.21		0.23		0.22		0.19		0.18		0.17	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	0.22	9	0.23	7	0.23	10	0.18	3	0.17	5	0.17	4
	BVAR	0.21	3	0.26	11	0.21	1	0.16	1	0.19	8	0.16	1
	TVC	0.21	4	0.23	5	0.21	6	0.20	11	0.20	10	0.18	11
	Ph	0.23	11	0.23	9	0.22	9	0.18	6	0.19	9	0.17	6
4-step	TS	0.21	7	0.24	10	0.21	4	0.18	5	0.17	6	0.17	5
	AC	0.21	1	0.23	4	0.21	3	0.18	4	0.17	3	0.16	3
	AFTER	0.21	2	0.21	1	0.21	7	0.20	10	0.20	11	0.17	7
	EW	0.21	5	0.23	8	0.21	2	0.18	2	0.18	7	0.16	2
	AIC	0.21	6	0.23	6	0.21	5	0.18	9	0.17	4	0.18	10
	SBC	0.23	10	0.22	3	0.24	11	0.18	8	0.16	1	0.18	9
	HQ	0.22	8	0.22	2	0.22	8	0.18	7	0.16	2	0.17	8
	mean	0.22		0.23		0.22		0.18		0.18		0.17	
	median	0.21		0.23		0.21		0.18		0.17		0.17	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
	Model	Simple	Rank	Factor	Rank	Panel	Rank	Simple	Rank	Factor	Rank	Panel	Rank
	ARMA	0.23	9	0.22	1	0.25	10	0.18	6	0.18	6	0.17	6
	BVAR	0.23	7	0.26	10	0.23	7	0.16	1	0.18	4	0.16	1
	TVC	0.22	3	0.30	11	0.22	3	0.19	8	0.19	7	0.18	7
	Ph	0.22	4	0.22	3	0.22	5	0.17	5	0.18	5	0.17	4
8-step	TS	0.23	10	0.24	9	0.23	8	0.17	3	0.17	3	0.17	5
	AC	0.23	6	0.23	6	0.23	6	0.17	2	0.17	2	0.16	3
	AFTER	0.21	1	0.23	7	0.21	1	0.19	7	0.19	8	0.19	8
	EW	0.22	2	0.22	2	0.22	4	0.17	4	0.17	1	0.16	2
	AIC	0.22	5	0.24	8	0.22	2	0.20	11	0.20	9	0.22	11
	SBC	0.24	11	0.23	4	0.24	9	0.20	10	0.22	10	0.22	10
	HQ	0.23	8	0.23	5	0.25	11	0.20	9	0.25	11	0.19	9
	mean	0.22		0.24		0.23		0.18		0.19		0.18	
	median	0.23		0.23		0.23		0.18		0.18		0.17	

Table 15: Inflation, Actual Uncertainty, pre- and post-84

Note: The columns labeled Simple, Factor and Panel show for each forecasting horizon and for each model the average actual uncertainty computed by pooling all forecast errors of our repetitive block experiment as explained in the text, from ten years of forecast vintages and corresponding outcomes over two sub-samples. The columns labeled Rank simply order the models according to the lowest-to-highest uncertainty.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
		SIMPLE		**Factor**		**Panel**		**SIMPLE**		**Factor**		**Panel**	
	Model	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank
	ARMA	0.27	4	0.10	6	0.19	4	0.17	7	0.11	11	0.14	10
	BVAR	0.11	6	-0.02	9	0.08	9	0.14	9	0.29	4	0.16	8
	TVC	0.19	5	0.19	4	0.14	6	0.17	8	0.19	10	0.16	9
	Okun	0.06	8	0.13	5	0.05	11	0.23	6	0.40	3	0.11	11
1-step	Li	0.05	9	0.08	7	0.12	7	0.37	3	0.58	1	0.28	2
	AC	0.85	2	0.37	3	0.82	2	0.07	10	0.24	5	0.22	6
	AFTER	0.88	1	0.48	2	0.86	1	0.05	11	0.24	6	0.20	7
	EW	0.76	3	0.55	1	0.64	3	0.27	5	0.42	2	0.28	1
	AIC	-0.01	11	-0.08	10	0.08	10	0.40	2	0.21	9	0.25	4
	SBC	0.11	7	0.01	8	0.18	5	0.36	4	0.24	7	0.23	5
	HQ	0.03	10	-0.13	11	0.11	8	0.43	1	0.23	8	0.25	3
	mean	0.30		0.15		0.30		0.24		0.29		0.21	
	median	0.11		0.10		0.14		0.23		0.24		0.22	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
		SIMPLE		**Factor**		**Panel**		**SIMPLE**		**Factor**		**Panel**	
	Model	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank
	ARMA	0.11	1	-0.02	9	0.17	1	0.25	9	-0.54	11	0.12	10
	BVAR	0.04	3	0.05	7	-0.01	4	0.38	8	0.50	1	0.26	7
	TVC	0.07	2	0.11	4	0.11	2	0.18	10	0.11	9	0.15	8
	Okun	-0.06	4	0.05	6	0.01	3	0.12	11	0.07	10	0.10	11
4-step	Li	-0.13	6	0.04	8	-0.08	5	0.47	7	0.27	4	0.14	9
	AC	-0.21	10	0.32	1	-0.30	11	0.78	4	0.26	7	1.10	2
	AFTER	-0.21	9	0.32	2	-0.29	10	0.80	3	0.33	2	1.09	3
	EW	-0.12	5	-0.03	10	-0.12	7	1.53	1	0.26	5	1.25	1
	AIC	-0.17	8	0.12	3	-0.10	6	0.73	5	0.31	3	0.27	6
	SBC	-0.25	11	-0.04	11	-0.15	9	0.83	2	0.26	6	0.49	4
	HQ	-0.17	7	0.06	5	-0.12	8	0.66	6	0.24	8	0.31	5
	mean	-0.10		0.09		-0.08		0.61		0.19		0.48	
	median	-0.13		0.05		-0.10		0.66		0.26		0.27	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
		SIMPLE		**Factor**		**Panel**		**SIMPLE**		**Factor**		**Panel**	
	Model	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank
	ARMA	0.12	3	0.00	9	0.02	9	0.54	2	0.30	6	0.37	2
	BVAR	0.05	9	-0.22	11	0.00	10	0.36	5	0.47	5	0.15	11
	TVC	0.21	1	0.28	2	0.12	1	0.34	7	-0.11	11	0.17	9
	Okun	0.08	6	0.07	7	0.04	8	0.11	11	0.24	9	0.19	8
8-step	Li	0.05	10	0.31	1	0.08	5	0.32	8	0.78	1	0.21	7
	AC	0.11	4	0.13	5	0.11	2	0.19	10	0.72	2	0.24	4
	AFTER	0.17	2	0.22	4	0.08	4	0.21	9	0.54	4	0.16	10
	EW	-0.04	11	0.23	3	-0.02	11	0.78	1	0.69	3	0.75	1
	AIC	0.06	8	0.11	6	0.07	6	0.38	4	0.27	7	0.21	6
	SBC	0.09	5	-0.02	10	0.07	7	0.44	3	0.25	8	0.27	3
	HQ	0.07	7	0.03	8	0.10	3	0.35	6	0.23	10	0.22	5
	mean	0.09		0.10		0.06		0.37		0.40		0.27	
	median	0.08		0.11		0.07		0.35		0.30		0.21	

Table 16: Output Growth, Stability, pre- and post-84

Note: The columns labeled b_1 illustrate for each forecasting horizon and for each model the OLS estimate of the "pass-through" coefficient in equation (14). The columns labeled Rank simply order the models according to the highest-to-lowest estimate. he same information set.

		Sample 1969:1 - 1983:4						Sample 1984:1 - 1997:1					
		SIMPLE		Factor		Panel		SIMPLE		Factor		Panel	
	Model	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank
1-step	ARMA	0.05	6	-0.05	9	0.08	10	0.59	4	-0.35	11	0.52	4
	BVAR	0.08	5	0.02	8	0.09	8	0.32	7	0.20	8	0.28	7
	TVC	0.12	4	0.11	5	0.11	7	0.21	11	0.17	9	0.18	11
	Ph	-0.01	11	-0.06	10	0.09	9	0.49	5	0.12	10	0.37	5
	TS	0.03	10	0.03	7	0.06	11	0.40	6	0.26	6	0.32	6
	AC	0.31	3	0.26	2	0.37	3	1.56	3	1.33	3	1.52	3
	AFTER	0.38	2	0.21	3	0.44	2	1.57	2	1.40	2	1.56	2
	EW	0.49	1	0.45	1	0.56	1	1.99	1	1.94	1	2.08	1
	AIC	0.04	7	0.17	4	0.11	5	0.25	8	0.25	7	0.24	8
	SBC	0.04	8	-0.13	11	0.11	6	0.25	9	0.49	4	0.24	9
	HQ	0.04	9	0.08	6	0.12	4	0.25	10	0.35	5	0.23	10
	mean	0.14		0.10		0.19		0.72		0.56		0.69	
	median	0.05		0.08		0.11		0.40		0.26		0.32	

		Sample 1969:4 - 1983:4						Sample 1984:1 - 1998:1					
		SIMPLE		Factor		Panel		SIMPLE		Factor		Panel	
	Model	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank
4-step	ARMA	-0.11	6	0.01	5	-0.01	6	0.48	4	1.14	3	0.42	5
	BVAR	-0.04	3	-0.11	8	-0.01	5	0.13	9	0.85	5	0.09	11
	TVC	-0.07	5	-0.02	6	-0.01	4	0.25	8	0.28	11	0.20	8
	Ph	-0.12	8	0.05	3	-0.02	7	0.10	10	0.43	7	0.12	10
	TS	-0.11	7	0.01	4	-0.05	8	0.08	11	0.40	9	0.14	9
	AC	0.03	2	-0.42	10	0.09	2	1.48	2	1.15	2	1.52	2
	AFTER	-0.05	4	-0.47	11	0.04	3	1.23	3	1.13	4	1.30	3
	EW	0.34	1	-0.26	9	0.35	1	2.45	1	1.84	1	2.18	1
	AIC	-0.20	9	0.08	1	-0.14	10	0.28	5	0.33	10	0.36	6
	SBC	-0.20	10	-0.03	7	-0.14	11	0.28	6	0.53	6	0.36	7
	HQ	-0.20	11	0.05	2	-0.09	9	0.28	7	0.42	8	0.43	4
	mean	-0.06		-0.10		0.00		0.64		0.77		0.65	
	median	-0.11		-0.02		-0.01		0.28		0.53		0.36	

		Sample 1970:4 - 1983:4						Sample 1984:1 - 1999:1					
		SIMPLE		Factor		Panel		SIMPLE		Factor		Panel	
	Model	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank	β_1	Rank
8-step	ARMA	-0.07	7	0.02	3	-0.05	7	0.14	8	0.73	2	0.09	9
	BVAR	-0.01	3	0.30	1	0.01	2	0.09	9	0.45	6	0.04	11
	TVC	0.00	2	-0.04	6	-0.04	4	0.17	6	0.42	8	0.15	6
	Ph	-0.08	8	0.12	2	-0.05	6	0.06	10	0.61	4	0.14	8
	TS	-0.07	6	-0.12	8	-0.01	3	0.15	7	0.63	3	0.15	7
	AC	-0.06	5	-0.14	11	-0.13	9	0.20	5	0.49	5	0.35	1
	AFTER	-0.04	4	-0.12	9	-0.04	5	-0.01	11	0.29	10	0.05	10
	EW	0.07	1	-0.07	7	0.04	1	0.28	4	1.32	1	0.32	2
	AIC	-0.11	9	-0.02	5	-0.16	10	0.33	1	0.28	11	0.25	4
	SBC	-0.11	10	-0.13	10	-0.16	11	0.33	2	0.43	7	0.25	5
	HQ	-0.11	11	0.01	4	-0.10	8	0.33	3	0.39	9	0.28	3
	mean	-0.06		-0.02		-0.06		0.19		0.55		0.19	
	median	-0.07		-0.04		-0.05		0.17		0.45		0.15	

Table 17: Inflation, Stability, pre- and post-84

Note: The columns labeled b_1 illustrate for each forecasting horizon and for each model the OLS estimate of the "pass-through" coefficient in equation (14). The columns labeled Rank simply order the models according to the highest-to-lowest estimate. he same information set.

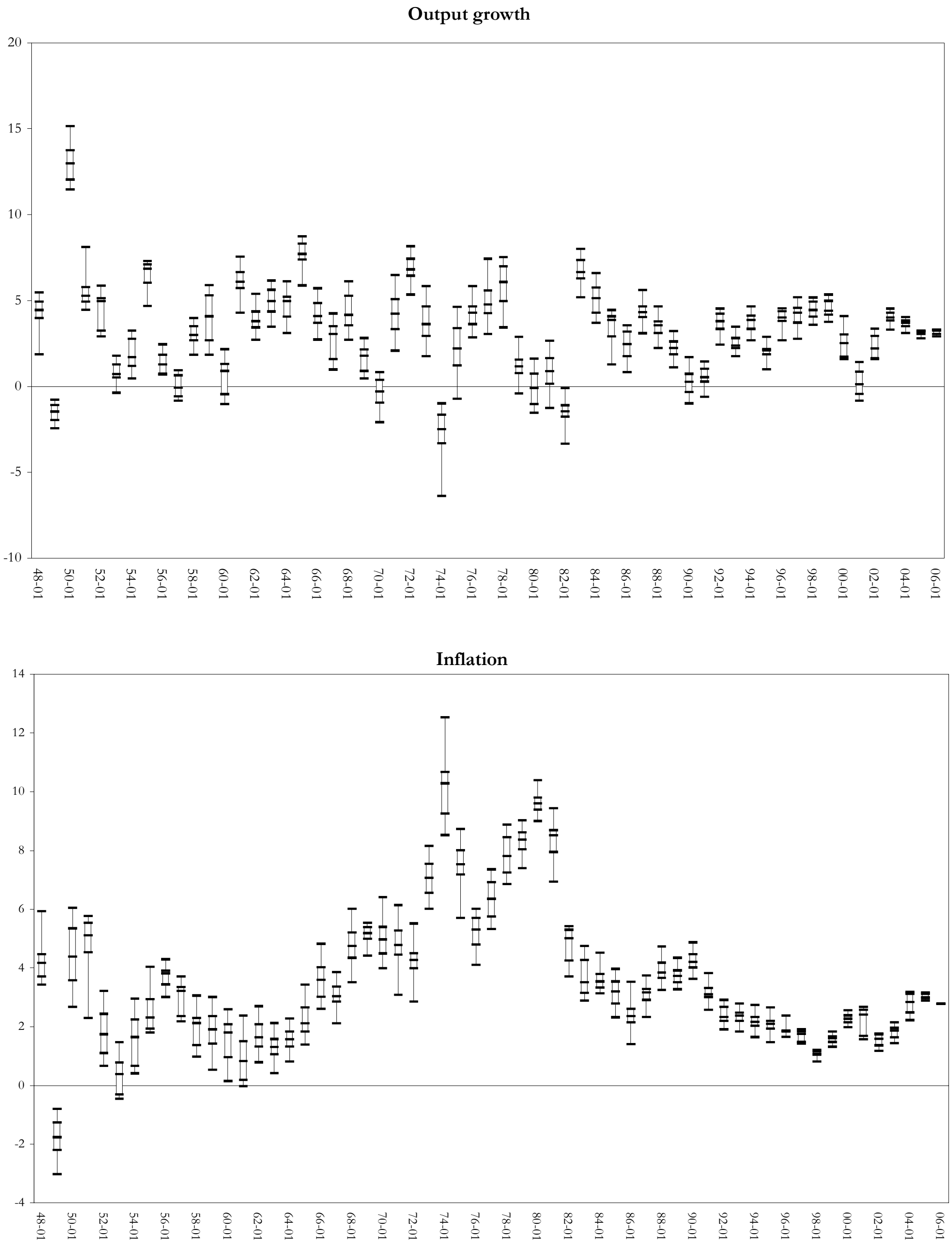

Figure 1: The revision process

Note: Each box plot depicts the (annual averages of) minimum, maximum, interquartile range and the median over all vintage realizations at each date.

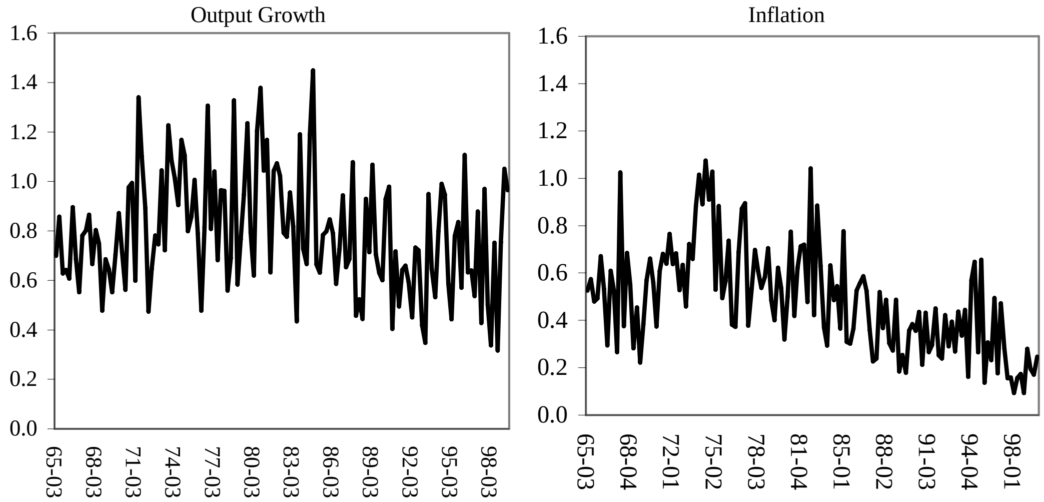

Figure 2: Standard deviation of the revisions

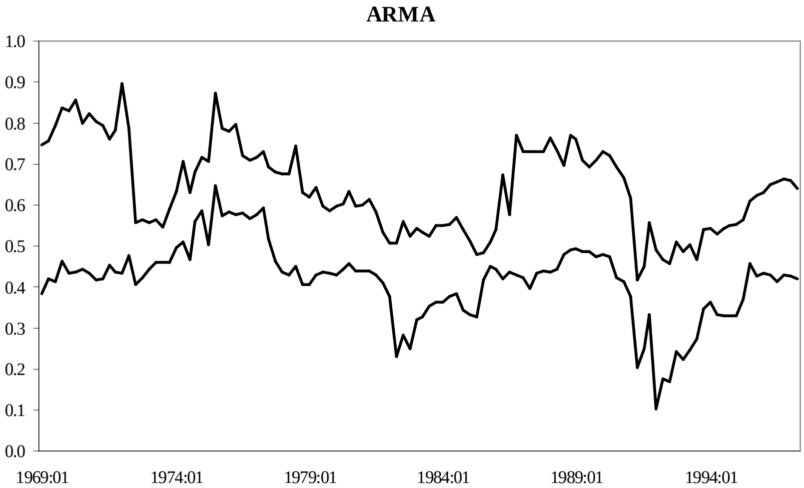

Figure 3: First AR coefficient for inflation. Min and Max.

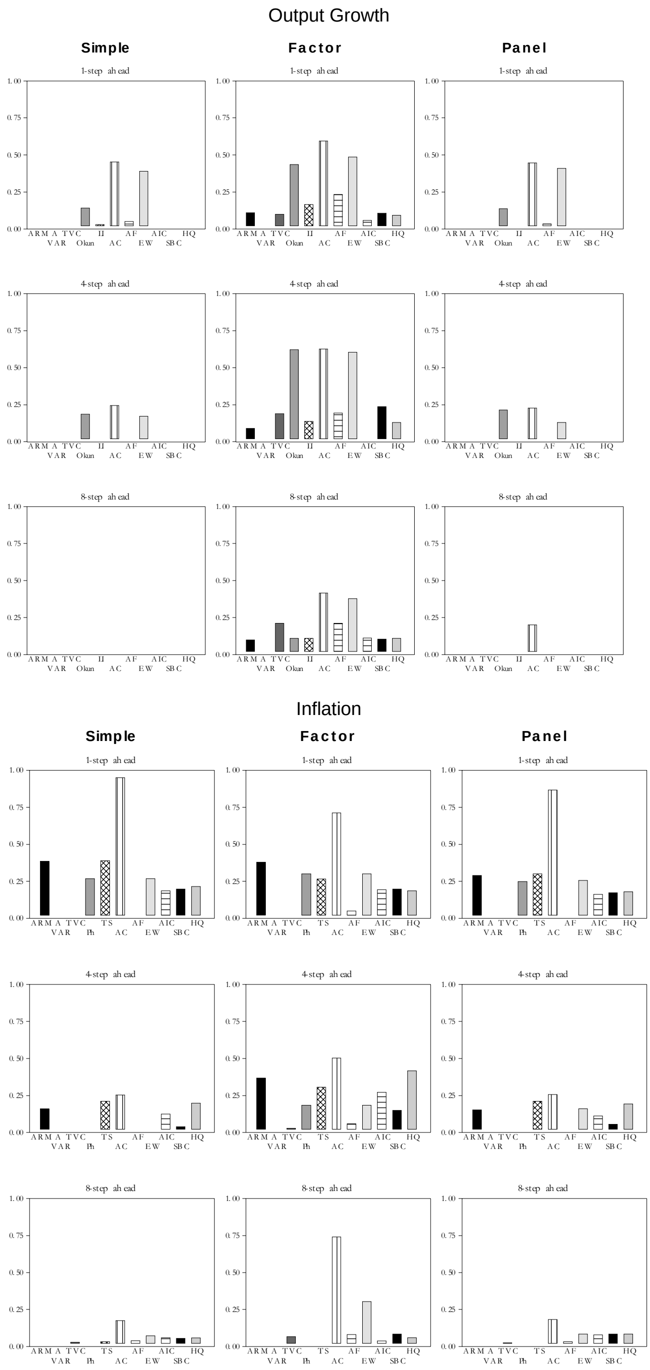

Figure 4: CW test of equal accuracy

Note: For each model the bar represents the percentage of times that in bilateral comparisons (across models) for all vintages we reject the null of equal predictive accuracy in favor of the model labeled on the horizontal axis according to the Clark and West (2007) test. For all tests significance is set at 0.10 confidence level.

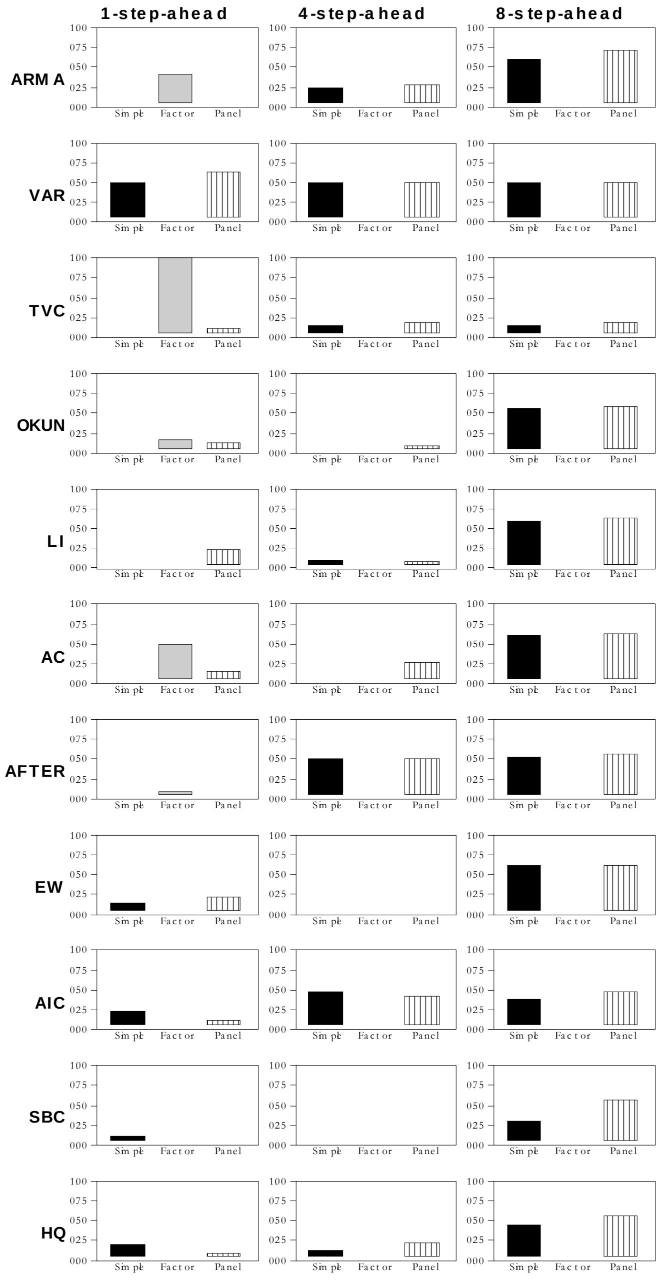

Figure 5: Output growth - CW test of equal accuracy

Note: For each information set (Simple, Factor, Panel), the bar represents the percentage of times that in bilateral comparisons (across information sets) for all vintages we reject the null of equal predictive accuracy in favor of the information sets labeled on the horizontal axis according to the Clark and West (2007) test. For all tests significance is set at 0.10 confidence level.

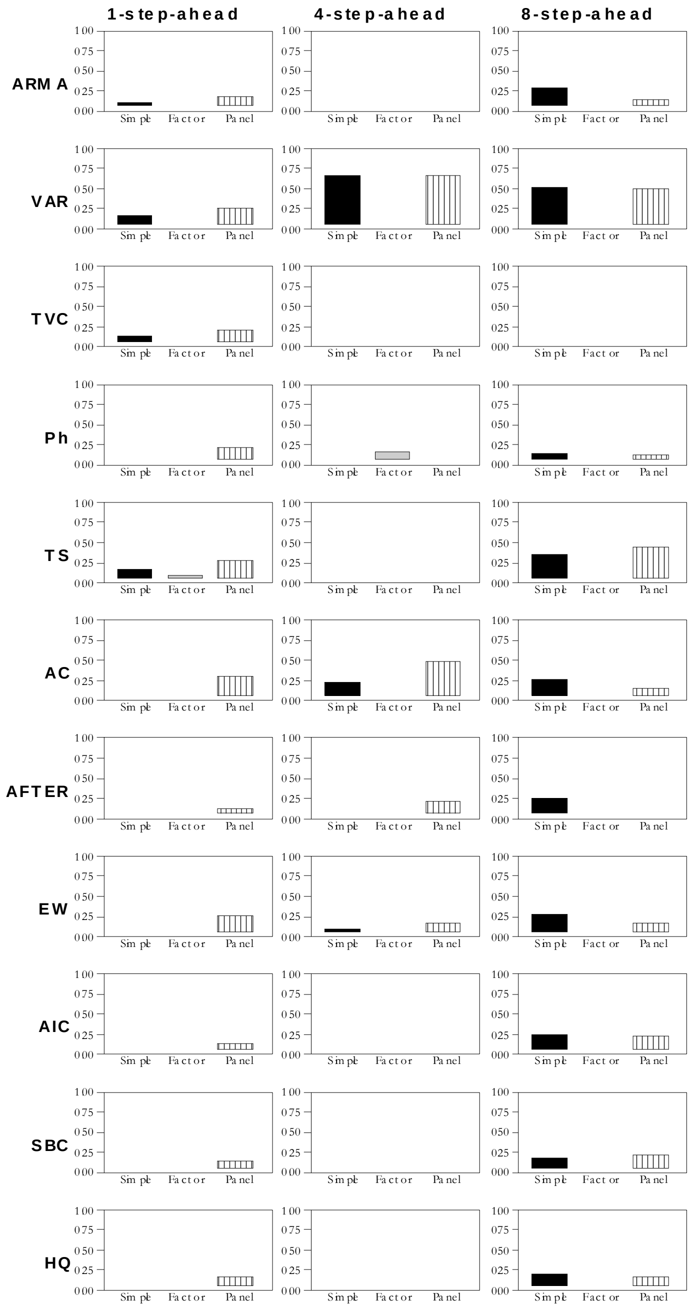

Figure 6: Inflation - CW test of equal accuracy

Note: For each information set (Simple, Factor, Panel), the bar represents the percentage of times that in bilateral comparisons (across information sets) for all vintages we reject the null of equal predictive accuracy in favor of the information sets labeled on the horizontal axis according to the Clark and West (2007) test. For all tests significance is set at 0.10 confidence level.

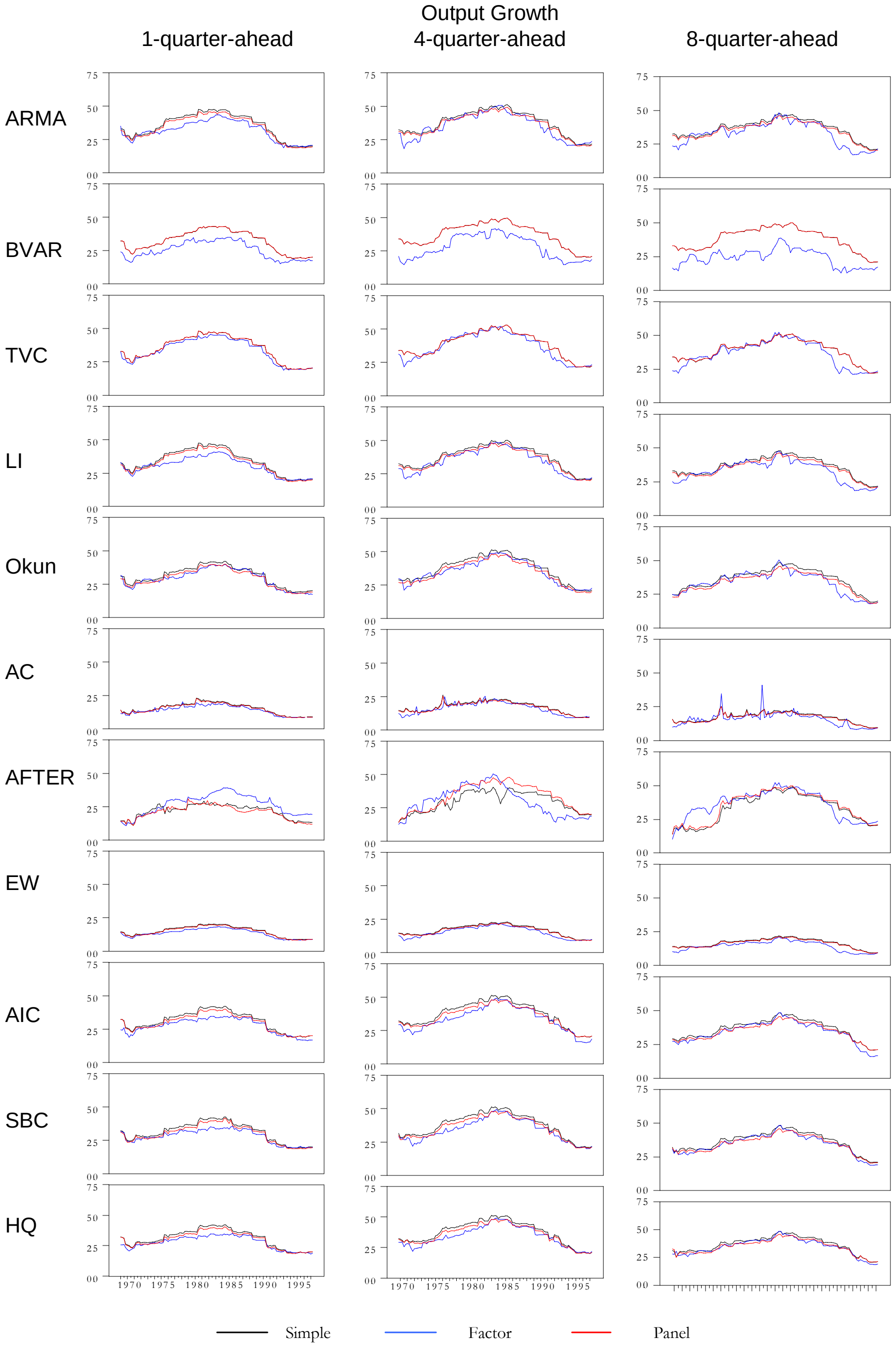

Figure 7: Predicted Uncertainty Output Growth

Note: Each graph reports (for all models and forecasting horizons) the evolution of the predicted uncertainty for the three different information sets over the sample 1969q1-1997q1 for h=1, 1969q4-1998q1 for h=4, 1970q4-1999q1 for h=8 calculated as discussed in Section 4.2.

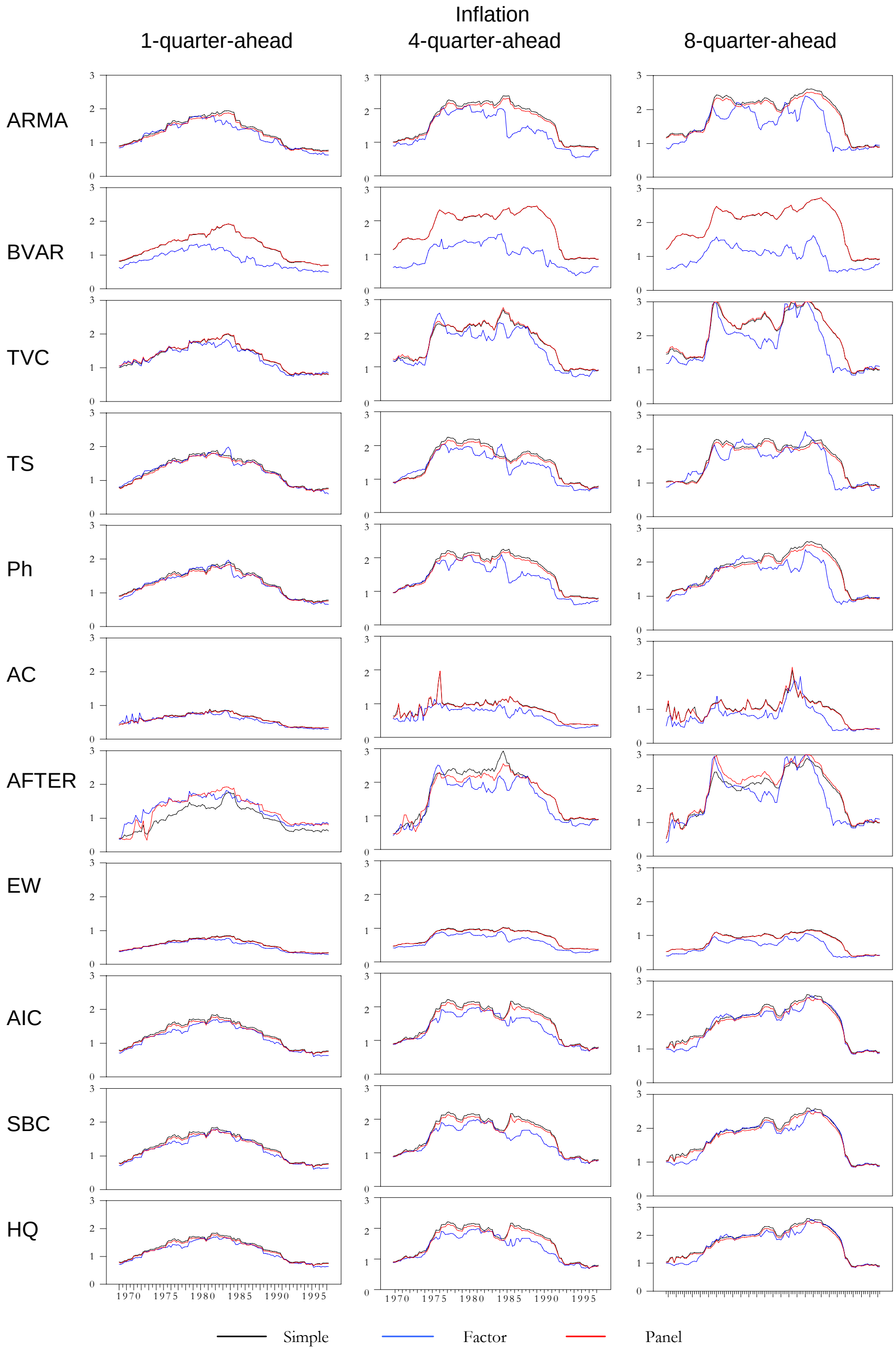

Figure 8: Predicted Uncertainty Inflation

Note: Each graph reports (for all models and forecasting horizons) the evolution of the predicted uncertainty for the three different information sets over the sample 1969q1-1997q1 for h=1, 1969q4-1998q1 for h=4, 1970q4-1999q1 for h=8 calculated as discussed in Section 4.2.

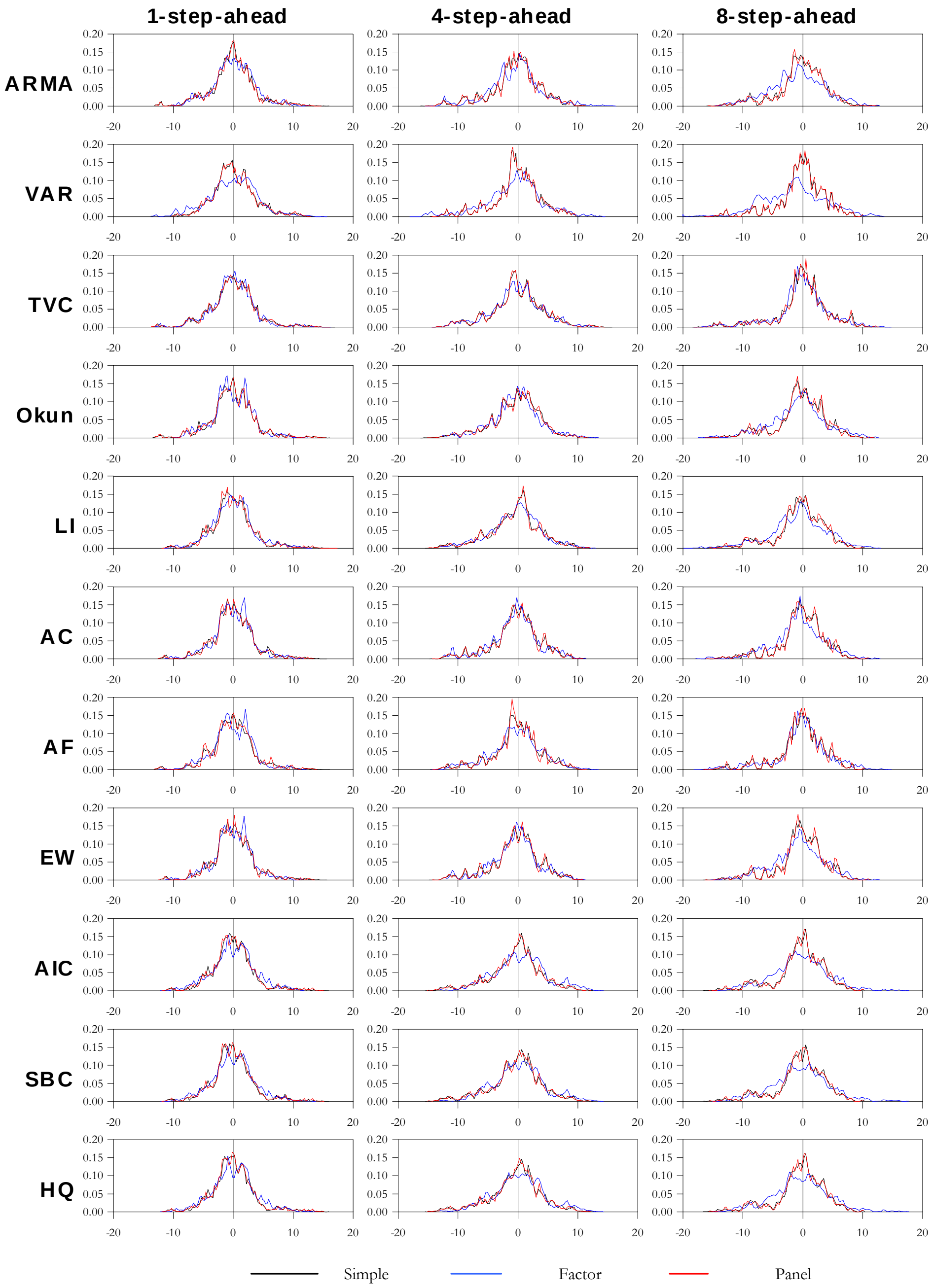

Figure 9: Actual Uncertainty Output Growth

Note: Each graph reports (for all models and forecasting horizons) the distribution of the actual uncertainty for the three different information sets over the sample 1969q1-1997q1 for h=1, 1969q4-1998q1 for h=4, 1970q4-1999q1 for h=8 calculated as discussed in section 4.2.

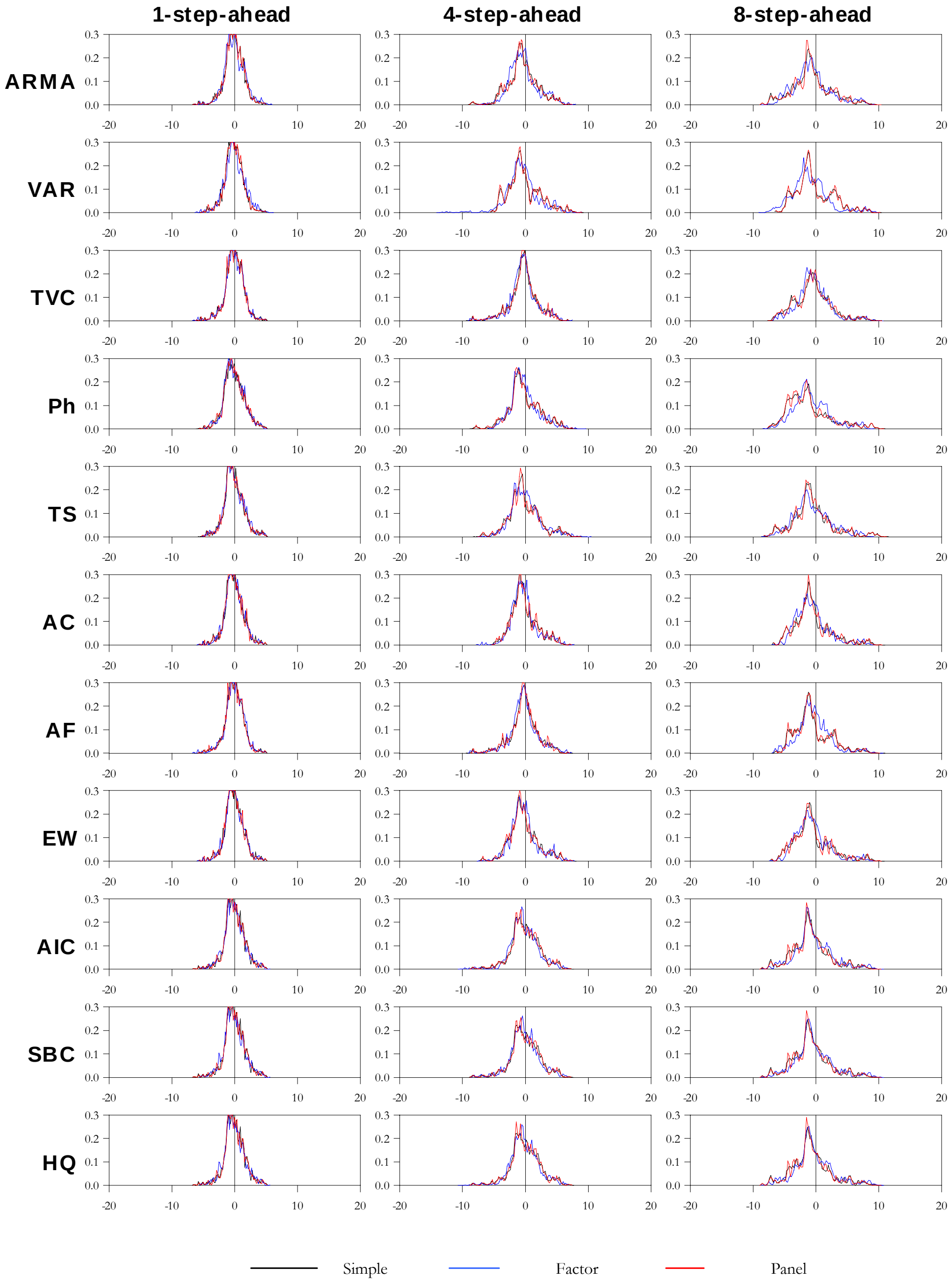

Figure 10: Actual Uncertainty Inflation

Note: Each graph reports (for all models and forecasting horizons) the distribution of the actual uncertainty for the three different information sets over the sample 1969q1-1997q1 for h=1, 1969q4-1998q1 for h=4, 1970q4-1999q1 for h=8 calculated as discussed in section 4.2.

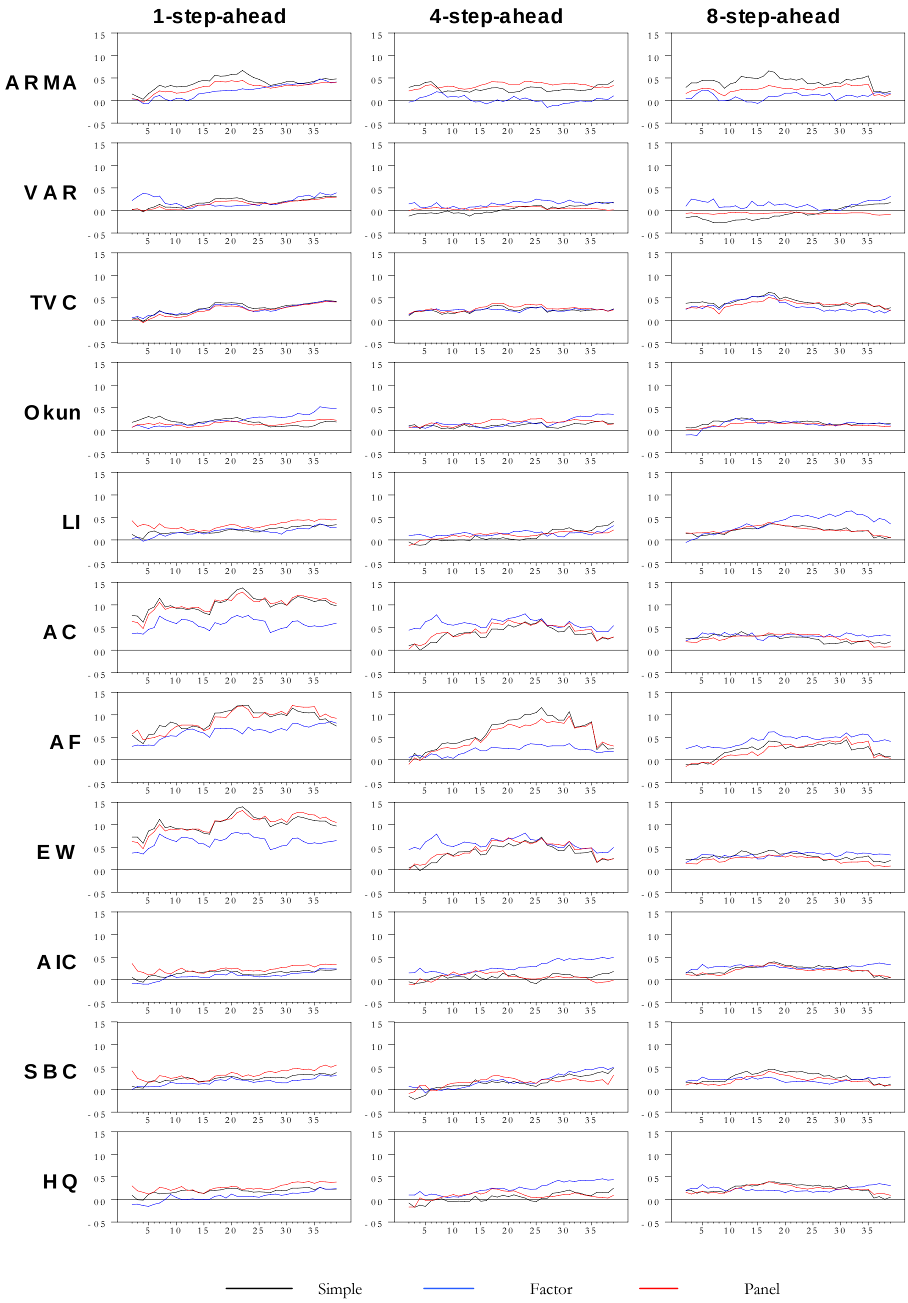

Figure 11: Stability - Output Growth

Note: Each graph reports (for all models and forecasting horizons) the evolution of the OLS estimate of the "pass-through" coefficient in equation (14) for the three different information sets over the selected sample.

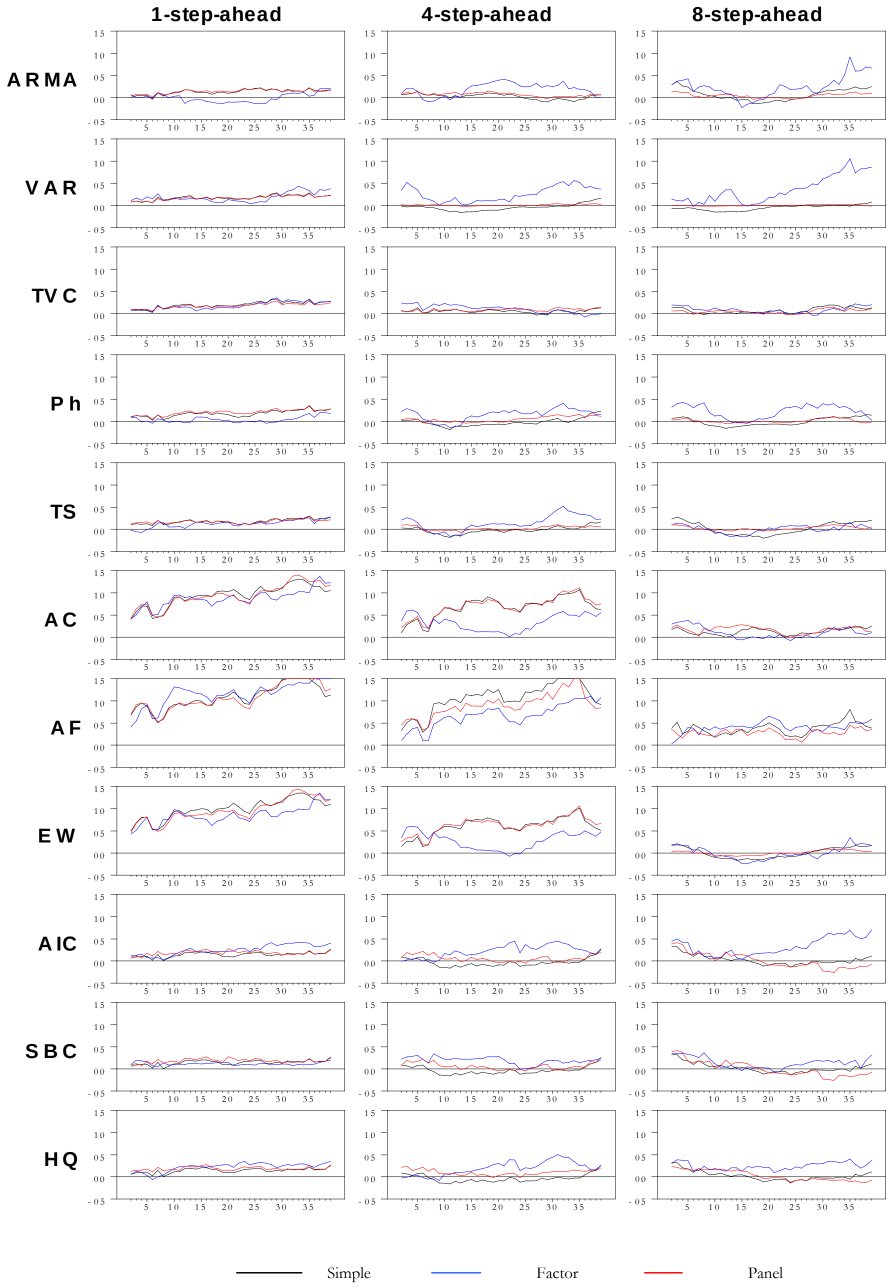

Figure 12: Stability - Inflation

Note: Each graph reports (for all models and forecasting horizons) the evolution of the OLS estimate of the "pass-through" coefficient in equation (14) for the three different information sets over the selected sample.