Multi-step forecasting in the presence of breaks

Hännikäinen, Jari

University of Tampere

7 May 2014

TAMPERE ECONOMIC WORKING PAPERS
NET SERIES

MULTI-STEP FORECASTING IN THE PRESENCE OF BREAKS

Jari Hännikäinen

Working Paper 94
May 2014

SCHOOL OF MANAGEMENT

FI-33014 UNIVERSITY OF TAMPERE, FINLAND

ISSN 1458-1191
ISBN 978-951-44-9497-0

Multi-step forecasting in the presence of breaks*

Jari Hännikäinen†

May 2014

Abstract

This paper analyzes the relative performance of multi-step forecasting methods in the presence of breaks and data revisions. Our Monte Carlo simulations indicate that the type and the timing of the break affect the relative accuracy of the methods. The iterated method typically performs the best in unstable environments, especially if the parameters are subject to small breaks. This result holds regardless of whether data revisions add news or reduce noise. Empirical analysis of real-time U.S. output and inflation series shows that the alternative multi-step methods only episodically improve upon the iterated method.

Keywords: Structural breaks, multi-step forecasting, intercept correction, real-time data

JEL codes: C22, C53, C82

*I would like to thank Henri Nyberg, Jukka Lahtonen, Sami Oinonen, Jari Vainiomäki and participants of the FDPE Econometrics Workshop in Helsinki, Hecer Time Series Econometrics seminar in Helsinki, Allecon seminar in Jyväskylä, and the XXXVI Annual Meeting of the Finnish Economic Association in Kuopio for helpful comments and discussions. Financial support from FDPE is gratefully acknowledged. All mistakes are mine.

†School of Management, University of Tampere, Kanslerinrinne 1, 33014 Tampere, Finland
E-mail: jari.hannikainen@uta.fi
Phone: +358 50 318 5975
Fax: +358 3 3551 7214

1. Introduction

The medium- and long-term prospects of the economy are important for consumers, investors, and policymakers. For example, it is well known that monetary policy affects the economy with a long lag. As a result, central banks conduct forward-looking monetary policy, i.e., central banks' interest rate decisions are based on their forecasts of future output growth, unemployment, and inflation. Given the importance of the medium- and long-term economic outlook, economists provide forecasts of key macroeconomic time series several periods ahead in time. These macroeconomic series are often serially correlated, implying that their own past values are themselves useful predictors. Therefore, autoregressive (AR) models are used extensively in economic forecasting. Despite their parsimonious form, it appears to be difficult to outperform AR models in practice (see, e.g., Elliott and Timmermann, 2008; Rossi, 2013; Stock and Watson, 2003).

When generating a multi-step forecast, a forecaster has to decide whether to use the iterated or direct forecasting strategy. In the iterated approach, forecasts are made using a one-period ahead model, iterated forward for the desired number of periods. A central feature of the iterated approach is that the model specification is the same regardless of the forecast horizon. Direct forecasts, on the other hand, are made using a horizon-specific model. Thus, a forecaster estimates a separate model for each forecast horizon. The theoretical literature analyzing the relative merits of the iterated versus the direct forecast methods includes, e.g., Bao (2007), Brown and Mariano (1989), Chevillon and Hendry (2005), Clements and Hendry (1996b, 1998), Findley (1985), Hoque *et al.* (1988), Ing (2003), Schorfheide (2005), and Weiss (1991). This literature emphasizes that the choice between iterated and direct multi-step forecasts is not clear cut, but rather involves a trade-off between bias and estimation variance. The iterated method uses the largest available data sample in the estimation and thus produces more efficient parameter estimates than the direct method. In contrast, direct forecasts are more robust to model misspecification. Which element, the bias or the estimation variance, dominates in the composition of the mean squared forecast error (MSFE) values in practice depends on the sample size, the forecast horizon, and the (unknown) underlying DGP, and therefore the question of which method to use cannot be decided ex ante on theoretical grounds alone. Hence, the question of which multi-step forecasting method to use is an empirical one. In their empirical analysis of 170 U.S. monthly macroeconomic time series, Marcellino *et al.* (2006) and Pesaran *et al.* (2011) find that the iterated approach typically outperforms the direct approach, especially if the sample size is small, if the forecast horizon is long, and if long lags of the variables are included in the forecasting model.

Although the parameters in many of the macroeconomic time series are unstable over time (Stock and Watson, 1996), work on multi-step forecasting in the presence of breaks has been virtually absent from the literature. However, it is widely accepted that structural breaks play a central role in economic forecasting (see, e.g., Clements and Hendry, 2006; Elliott and Timmermann, 2008; Rossi, 2013). Forecast errors are typically very large after structural breaks. Furthermore, it is possible that a forecasting model that performed well before the break performs poorly after the break. Forecasting models often systematically under- or over-predict in the presence of structural instability. Therefore, one way to improve their forecast accuracy in an unstable environment is to use intercept corrections, advocated by Clements and Hendry (1996a, 1998). Intercept corrections are based on the idea that if the forecasts systematically differ from the true values, i.e., if the forecast errors are systematically either positive or negative, then adjusting the mechanistic, model-based forecast by the previous forecast error (or an average of the most recent errors) should reduce the forecast bias and hence improve forecast performance.

Another issue that has been overlooked in the multi-step forecasting literature is the fact that key macroeconomic data, such as GDP and inflation series, are subject to revisions. The real-time nature of macroeconomic time series is potentially important for the relative performance of multi-step forecasting methods for at least three reasons. First, because data revisions are usually quite large, the parameters estimated on the final revised data may differ considerably from those estimated on the real-time data. Second, data revisions can also affect the dynamic lag structure of the forecasting model. Finally, real-time forecasts are conditioned on the first-release or lightly revised data actually available at each forecast origin, whereas forecasts based on the final revised data are conditioned on the latest available observations of each forecast origin. Practical forecasting is inherently a real-time exercise and thus the relative accuracy of multi-step forecasting methods should be evaluated using real-time data.

The main contributions of this paper are as follows. First, we analyze the relative performance of multi-step forecasting methods in the presence of breaks through Monte Carlo simulations. Our comparison includes the iterated and direct AR models and various forms of intercept correction. We consider several break processes, including changes in the intercept, autoregressive parameter, and error variance. We also examine how the timing of the break affects the accuracy of the methods. Second, we take into account in our simulations that most macroeconomic time series are subject to data revisions. A novelty of our simulation framework is that data revisions can either add news or reduce noise (see, e.g., Mankiw and Shapiro, 1986). The distinction between news and noise revisions allows us to study whether the properties of the revision process matter for the multi-period forecasting problem. Finally, the real-time accuracy of the multi-step forecasting methods for four key U.S. macroeconomic time series, namely, real GDP, industrial production, GDP deflator, and personal consumption expenditures (PCE) inflation, is compared.

The remainder of this paper is organized as follows. Section 2 introduces the notation and the statistical framework. Section 3 provides a brief overview of the multi-step forecasting methods. Section 4 presents the Monte Carlo simulation results and Section 5 presents the empirical results. Section 6 concludes.

2. Statistical framework

Key macroeconomic time series are published with a lag and are subject to revisions. For instance, a forecaster at period $T+1$ has access to the vintage $T+1$ values of GDP up to time period T. In addition, because of data revisions, the first-released value and the final value for a period may differ substantially. These two features of real-time data clearly matter for forecasting. As a result, we incorporate the publication lag and data revisions into our statistical framework. The statistical framework used in this paper follows that adopted in Jacobs and van Norden (2011), Clements and Galvão (2013), and Hännikäinen (2014). It relates a data vintage estimate to the true value plus an error or errors. More specifically, the period $t+s$ vintage estimate of the value of y in period t, denoted by y_t^{t+s} [1], where $s = 1, ..., l$ [2], can be expressed as the sum of the true value $\tilde{y}_t$, a news component v_t^{t+s}, and a noise component ε_t^{t+s}, i.e., $y_t^{t+s} = \tilde{y}_t + v_t^{t+s} + \varepsilon_t^{t+s}$.

In this framework, revisions either add news or reduce noise. Data revisions are news if they are uncorrelated with the previously published vintages, $cov(y_t^{t+k}, v_t^{t+s}) = 0$ $\forall k \leq s$. On the other hand, data revisions reduce noise if each vintage release is equal to the true value plus a noise. Noise revisions are uncorrelated with the true values, $cov(\tilde{y}_t, \varepsilon_t^{t+s}) = 0$. For further discussion of the properties of news and noise revisions, see Croushore (2011) and Jacobs and van Norden (2011).

We stack the l different vintage estimates of y_t, v_t and ε_t into vectors $\boldsymbol{y}_t = (y_t^{t+1}, ..., y_t^{t+l})'$, $\boldsymbol{v}_t = (v_t^{t+1}, ..., v_t^{t+l})'$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_t^{t+1}, ..., \varepsilon_t^{t+l})'$, respectively. Using these

[1] Throughout this section, superscripts refer to vintages and subscripts to time periods.

[2] Following Clements and Galvão (2013), we assume that we observe l different estimates of y_t before the true value, $\tilde{y}_t$, is observed. In practice, however, data may continue to be revised forever, so the true value may never be observed.

vectors we can express each vintage of y_t as follows

$$\boldsymbol{y}_t = \boldsymbol{i}\tilde{y}_t + \boldsymbol{v}_t + \boldsymbol{\varepsilon}_t, \tag{1}$$

where $\boldsymbol{i}$ is an $l \times 1$ vector of ones. For simplicity, we consider an AR(1) process for the true values and assume that a single break has occurred at time T_1[3]

$$\tilde{y}_t = \begin{cases} \rho_1 + \sum_{i=1}^{l} \mu_{v1_i} + \beta_1 \tilde{y}_{t-1} + \sigma_1 \eta_{1t} + \sum_{i=1}^{l} \sigma_{v1_i} \eta_{2t,i}, & \text{for } t \leq T_1, \\ \rho_2 + \sum_{i=1}^{l} \mu_{v2_i} + \beta_2 \tilde{y}_{t-1} + \sigma_2 \eta_{1t} + \sum_{i=1}^{l} \sigma_{v2_i} \eta_{2t,i}, & \text{for } t > T_1, \end{cases} \tag{2}$$

where $v_{j,i,t} = \mu_{vj_i} + \sigma_{vj_i}\eta_{2t,i}$ (for j = 1,2 and i = 1,..,l) denote news and both η_{1t} and $\eta_{2t,i}$ are i.i.d. (0,1) disturbances. This setup allows for changes in the error variance, the intercept, and the slope immediately after the break.

The news and noise components in (1) before and after the break are specified by

$$\boldsymbol{v}_{1t} = \begin{bmatrix} v_{1t}^{t+1} \\ v_{1t}^{t+2} \\ \vdots \\ v_{1t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^{l} \mu_{v1_i} \\ \sum_{i=2}^{l} \mu_{v1_i} \\ \vdots \\ \mu_{v1_l} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{l} \sigma_{v1_i}\eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v1_i}\eta_{2t,i} \\ \vdots \\ \sigma_{v1_l}\eta_{2t,l} \end{bmatrix}, \boldsymbol{\varepsilon}_{1t} = \begin{bmatrix} \varepsilon_{1t}^{t+1} \\ \varepsilon_{1t}^{t+2} \\ \vdots \\ \varepsilon_{1t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \mu_{\varepsilon 1_1} \\ \mu_{\varepsilon 1_2} \\ \vdots \\ \mu_{\varepsilon 1_l} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon 1_1}\eta_{3t,1} \\ \sigma_{\varepsilon 1_2}\eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon 1_l}\eta_{3t,l} \end{bmatrix} \tag{3}$$

[3]Eklund *et al.* (2013), Hännikäinen (2014), and Pesaran and Timmermann (2005) also focus on an AR(1) model in the presence of breaks.

for $t \leq T_1$ and

$$\boldsymbol{v}_{2t} = \begin{bmatrix} v_{2t}^{t+1} \\ v_{2t}^{t+2} \\ \vdots \\ v_{2t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \sum_{i=1}^{l} \mu_{v2_i} \\ \sum_{i=2}^{l} \mu_{v2_i} \\ \vdots \\ \mu_{v2_l} \end{bmatrix} - \begin{bmatrix} \sum_{i=1}^{l} \sigma_{v2_i} \eta_{2t,i} \\ \sum_{i=2}^{l} \sigma_{v2_i} \eta_{2t,i} \\ \vdots \\ \sigma_{v2_l} \eta_{2t,l} \end{bmatrix}, \boldsymbol{\varepsilon}_{2t} = \begin{bmatrix} \varepsilon_{2t}^{t+1} \\ \varepsilon_{2t}^{t+2} \\ \vdots \\ \varepsilon_{2t}^{t+l} \end{bmatrix} = - \begin{bmatrix} \mu_{\varepsilon 2_1} \\ \mu_{\varepsilon 2_2} \\ \vdots \\ \mu_{\varepsilon 2_l} \end{bmatrix} + \begin{bmatrix} \sigma_{\varepsilon 2_1} \eta_{3t,1} \\ \sigma_{\varepsilon 2_2} \eta_{3t,2} \\ \vdots \\ \sigma_{\varepsilon 2_l} \eta_{3t,l} \end{bmatrix} \tag{4}$$

for $t > T_1$.

The shocks are assumed to be mutually independent. Otherwise stated, if $\boldsymbol{\eta}_t = \left[\eta_{1t}, \boldsymbol{\eta}_{2t}^{'}, \boldsymbol{\eta}_{3t}^{'}\right]^{'}$, then $E(\boldsymbol{\eta}_t) = \mathbf{0}$ and $E(\boldsymbol{\eta}_t \boldsymbol{\eta}_t^{'}) = I$. We assume that $\tilde{y}_t$ is a stationary process, so that $|\beta_j| < 1$ (for j = 1,2). Because $\tilde{y}_t$ is a stationary process and both the news and noise terms are stationary, $\boldsymbol{y}_t$ is also a stationary process. The means of the news and noise terms, denoted by μ_{vj_i} and $\mu_{\varepsilon j_i}$ (for j = 1,2 and i = 1,...,l), are allowed to be non-zero. This is an important feature because the previous literature has found that revisions to macroeconomic data typically have non-zero means (see, e.g., Aruoba, 2008; Croushore, 2011; Clements and Galvão, 2013).

3. Methods for multi-step forecasting

In this section, we explain how the multi-step forecasts are computed in the iterated and direct approaches. We assume that the variable of interest, y_t, is a stationary process. For simplicity, we focus on an AR(1) model. The generalization to AR(p) models is straightforward.

Iterated forecasts are made using a one-period ahead model, iterated forward for the required number of periods. The one-step ahead AR model for y_t, ignoring data revisions, is

$$y_{t+1} = \alpha + \beta y_t + \varepsilon_t. \tag{5}$$

The parameters in (5) are estimated by OLS and the iterated forecast of y_{t+h} is then calculated as follows:

$$\hat{y}^I_{t+h|t} = \hat{\alpha} + \hat{\beta}\hat{y}^I_{t+h-1|t},$$

where $\hat{y}_{j|t} = y_j$ for $j = t$. Note that the same model specification is used for all forecast horizons.

Under the direct approach, the dependent variable in the forecasting model is the multi-step ahead value being forecasted. Thus, a forecaster selects a separate model for each forecast horizon. The direct forecasting model, ignoring data revisions, is

$$y_{t+h} = \phi + \rho y_t + \varepsilon_{t+h}. \tag{6}$$

The parameters in (6) are estimated by OLS using data through period t (i.e., y_t is the last observation on the left-hand side of the multi-step regression). Then, the direct forecast of y_{t+h} is constructed as

$$\hat{y}^D_{t+h|t} = \hat{\phi} + \hat{\rho} y_t.$$

As discussed in the Introduction, intercept corrections offer some protection against structural instability. If the forecasting model systematically either under- or over-predicts after a break, intercept corrections based on the previous forecast errors reduce forecast bias. On the other hand, intercept corrections increase forecast error variance.

Following Clements and Hendry (1996a, 1998), we consider three alternative intercept corrections to the iterated approach. The first strategy is a so-called constant adjustment method, where the adjustment over the forecast period is held constant at the average of the most recent forecast errors, denoted by e_t^*:

$$\tilde{y}^I_{t+h|t} = \hat{\alpha} + \hat{\beta}\tilde{y}^I_{t+h-1|t} + e_t^*,$$

which implies that

$$\tilde{y}^I_{t+h|t} = \hat{y}^I_{t+h|t} + \sum_{i=0}^{h-1} \hat{\beta}^i e_t^*.$$

The second strategy only adjusts the one-step ahead forecast. The iterated forecast generated by this one-off adjustment method is

$$\vec{y}^I_{t+h|t} = \hat{\alpha} + \hat{\beta}\vec{y}^I_{t+h-1|t}, \quad \vec{y}^I_{t+1|t} = \tilde{y}^I_{t+1|t} = \hat{\alpha} + \hat{\beta} y_t + e_t^*,$$

so that

$$\vec{y}^I_{t+h|t} = \hat{y}^I_{t+h|t} + \hat{\beta}^{h-1} e_t^*.$$

The third strategy, called the full-adjustment method, adjusts the model-based forecast by the full amount of the average of the most recent forecast errors:

$$\bar{y}^I_{t+h|t} = \hat{y}^I_{t+h|t} + e_t^*.$$

In addition, we consider a full-adjustment to the direct forecasting method. In this case, the average of the most recent forecast errors from the direct model, denoted by $e_{t,D}^*$, is used to adjust the model-based forecast:

$$\bar{y}^D_{t+h|t} = \hat{y}^D_{t+h|t} + e_{t,D}^*.$$

4. Monte Carlo simulations

In this section, we perform a number of Monte Carlo simulation experiments to evaluate the performance of the multi-step forecasting methods in the presence of breaks. These experiments are based on the statistical framework introduced in Section 2. A sample size of 100 observations, which corresponds to 25 years of quarterly data, is used in the experiments. We assume that a single break has occurred prior to the forecast origin. Because the timing of the break might affect the relative accuracy of the multi-step methods, we consider three different break points: $T_1 = 25$, 50, and 99.

We calibrate the parameter values on actual U.S. data following Hännikäinen (2014). The parameters remain constant over time in experiment 1 (see Table 1). In this case, the selected parameter values imply that the mean of the true process lies between 2.0 and 2.5, which corresponds roughly to the average U.S. annual inflation and real GDP growth over the past 25 years. The parameters in experiment 1 are used as pre-break parameters in the rest of the experiments (with the exceptions of experiments 4–5). We consider several break processes. First, we analyze how moderate (0.25) and large (0.5) changes in the autoregressive parameter in either direction affect the relative performance of the multi-step methods (experiments 2–5). Second, we consider breaks in the error variance. We allow σ to increase from 1.5 to 4.5 (experiment 6) and decrease from 1.5 to 0.5 (experiment 7). Finally, we examine how changes in the constant term affect the accuracy of the methods (experiments 8–9).

We assume that the data revisions are either pure news ($\sigma_{v_i} \neq 0, \sigma_{\varepsilon_i} = 0$ for $i = 1, ..., l$) or pure noise ($\sigma_{v_i} = 0, \sigma_{\varepsilon_i} \neq 0$ for $i = 1, ..., l$). This allows us to analyze whether the properties of the revision process matter for the relative performance of the multi-step forecasting methods. We set $l = 14$, so that we observe 14 different estimates of y_t before the true value, $\tilde{y}_t$, is observed[4]. Consistent with the previous

[4]As discussed in Croushore (2011), GDP and inflation data for period t are subject to annual revisions at the end of July of each of the following three years. Our choice $l = 14$ is motivated by

work in Clements and Galvão (2013) and Hännikäinen (2014), only the first and fifth revisions are assumed to have non-zero means. The means of these revisions are set to four and two percent of the mean of the first-release data, y_t^{t+1}, both before and after the break. Similarly, the standard deviation of the first revision is set to 40 percent of the standard deviation of the first-release data. The standard deviations of revisions 2–13 and 14 are set to 20 and 10 percent of the standard deviation of the first-release data, respectively. For convenience, the parameter values used in the Monte Carlo experiments are shown in Table 1[5].

For simplicity, we focus on forecasting the first-release values and assume that the lag structure of the forecasting model is correctly specified, i.e., the forecasts are generated using an AR(1) model[6]. We estimate the parameters of the forecasting models using the entire data sample from the latest available vintage. Following Clements and Hendry (1996a), the intercept corrections are based on the average of the latest four forecast errors[7]. The iterated multi-step forecasting method is used as a benchmark in our Monte Carlo simulations. For each alternative method we compute MSFE values relative to those produced by the iterated benchmark. Values below (above) unity imply that the candidate method produces more (less) accurate forecasts than the benchmark. Multi-step forecasts are computed for horizons of 2, 4, 8, and 12 periods. The results are based on 10,000 replications and are shown in Tables 2 and 3.

Table 2 shows the relative performance of the multi-step forecasting methods when the data revisions are pure news. The results indicate that the iterated method generates the best forecasts in most of the experiments. In particular, the iterated method

the fact that y_t^{t+15} will have undergone all the regular revisions irrespectively of which quarter of the year t falls in. For a similar approach, see Clements and Galvão (2013).

[5] Appendix A summarizes the means and standard deviations of the first-release and final data for each experiment. The details of the calibration process are presented in Hännikäinen (2014).

[6] The results are qualitatively similar if we use the bias correction method suggested by Clements and Galvão (2013) to forecast the final values or if we consider an AR(2) forecasting model. A full set of results is available upon request.

[7] The general conclusions are the same if the intercept corrections are based on the most recent forecast error or the average of the latest two or three forecast errors.

dominates the other methods when the parameters remain constant over time (experiment 1), or the variance increases (experiment 6), or the intercept increases (experiment 8). The iterated method also performs particularly well when the autoregressive parameter decreases moderately (experiment 3), or when the constant term decreases (experiment 9), although it does not always deliver the most accurate forecasts. In these few cases, however, the best performing alternative makes only a very slight improvement over the iterated approach. By contrast, the iterated method performs poorly when the autoregressive parameter decreases substantially after the break (experiment 5).

The timing of the structural break ($T_1 = 25, 50, 99$) has an impact on the performance of the various approaches. The iterated method appears to be the superior method when the break occurs early ($T_1 = 25$) during the sample, but its performance deteriorates when the break occurs closer to the forecast origin. There is a simple explanation for this finding. Table 4 reports the (squared) forecast bias of each method relative to the MSFE of the benchmark iterated model. As the timing of the break increases, forecasts become more biased, because fewer post-break values are available for estimation. This implies that the importance of the bias component in determining the accuracy of the forecasts increases. The iterated method is more prone to bias than the other methods. Therefore, it is less successful when the break date T_1 gets close to the end of the sample.

Moreover, the relative performance of the iterated method improves as the forecast horizon increases. This happens for a subtle reason. As the forecast horizon increases, the parameters of the direct model are estimated with fewer observations. The parameters of the iterated model, on the other hand, are estimated with the largest possible sample size regardless of the forecast horizon. Thus, for a fixed sample size, it becomes less desirable to use an inefficient direct method as the forecast horizon lengthens. Intercept corrections reduce the forecast bias at the cost of increased forecast error variance. The additional uncertainty induced by intercept corrections grows with the forecast horizon. Hence, the bias–variance trade-off is less favorable to intercept corrections at long horizons.

The results in Table 2 suggest that various forms of intercept correction yield relatively poor forecasts in the presence of structural instability. The only exception is the case where the slope parameter decreases substantially after the break (experiment 5). In this case, the improvements over the iterated benchmark are very large at longer forecast horizons (i.e., $h = 8$ and 12). Hence it is mainly in situations where a break is believed to decrease substantially the AR parameter (i.e., when both the mean and variance decrease substantially) that intercept corrections can be recommended. In the rest of the experiments, intercept corrections have the most potential when the break has occurred close to the forecast origin (i.e., $T_1 = 99$) and the forecast horizon is short (i.e., $h = 2$ and 4). The one-off adjustment to the iterated method is generally more successful at reducing the MSFE values than the other forms of intercept correction. The constant adjustment to the iterated method and the full adjustment to the direct method perform worst among all the methods. They produce significantly higher MSFE values than the iterated benchmark in most of the experiments.

A comparison of the iterated and direct methods reveals that the iterated method typically delivers more accurate forecasts in the presence of breaks. The direct forecasts only dominate the iterated ones when the autoregressive parameter decreases substantially (experiment 5) and the timing of the break is either $T_1 = 25$ or $T_1 = 50$. Thus, there is only very limited evidence that the direct method helps reduce MSFE values in an unstable environment. The explanation for this finding is again related to the bias–variance trade-off. It appears that in an unstable environment, the reduction in bias obtained from the direct model is less important than the reduction in estimation variance arising from estimating the iterated model.

The results for noise revisions are summarized in Table 3. These results are qualitatively similar to those presented in Table 2. Thus, whether the data revisions add news or reduce noise does not matter much for the relative performance of the multi-period forecasting methods. If anything, the iterated method performs slightly better in relative terms when data revisions reduce noise.

5. Empirical results

Next, we compare the relative performance of the multi-step forecasting methods using actual U.S. real-time data. We consider h-step ahead forecasts of real GDP and industrial production growth, the GDP deflator, and the PCE inflation rate (annualized). All forecasts are out-of-sample. At each forecast origin $t+1$, the $t+1$ vintage estimates of data up to period t are used to estimate the parameters of a forecasting model that is then used to generate a forecast for period $t+h$. Forecasts are generated for horizons of $h = 2$, 4, 8, and 12 quarters. A rolling window of 100 observations is used in the estimation. We consider two fixed lag lengths, namely $p = 1$ and $p = 4$. In addition, we determine the lag length by the Bayes Information Criterion (BIC) and the Akaike Information Criterion (AIC). The possible lag lengths are $p = 1, \ldots, 4$. At each forecast origin the model with the lowest information criteria is chosen. Because the BIC and AIC values are recomputed at each forecast origin, the order of the forecasting model can change from one period to the next[8]. Intercept corrections are based on the average of the four most recent forecast errors[9].For simplicity, we focus on forecasting the first-release values. All real-time data is quarterly and the sample period runs from 1947:Q2 to 2013:Q2. Different vintages are obtained from the Federal Reserve Bank of Philadelphia's real-time database.

[8] Iterated models selected by the AIC on average include two lags for real activity measures and three lags for inflation series. The BIC selects iterated models with only one lag for the real output series and models with two or three lags for the inflation series. For the direct models, the AIC recommends on average one or two lags, whereas the BIC recommends an optimal lag length of one.

[9] The results are qualitatively similar if intercept corrections are based on the most recent forecast error or the average of the latest two or three forecast errors.

We start our analysis by considering the whole out-of-sample period spanning from 1977:Q2 to 2013:Q2. The performance of the various multi-step forecasting methods relative to the iterated benchmark over this period is summarized in Table 5. Panels A and B report the results for the real GDP and industrial production, whereas Panels C and D contain the results for the GDP deflator and PCE inflation. The first row in each Panel provides the root MSFE value of the benchmark iterated estimator. The subsequent rows show the MSFE values of the candidate methods relative to the MSFE value of the benchmark model. The statistical significance is evaluated using the Giacomini and White (2006) test.

The results in Panels A and B indicate that the iterated method typically produces the lowest, or nearly the lowest, MSFE values for both real GDP and industrial production irrespective of which lag method or forecast horizon is employed. Even in the few cases where at least one of the other methods generates more accurate multi-step forecasts, even the best performing alternative provides only modest improvements over the iterated benchmark. For real GDP, the one-off adjustment method systematically dominates the benchmark at $h = 2$. Similarly, when short-lag selection methods ($p = 1$ and BIC) are used, the direct forecast is preferable to the iterated one at the shortest forecast horizon. However, the p-values indicate that these differences in the predictive ability are not statistically significant. When industrial production is forecasted, only the direct estimator outperforms the iterated benchmark in a few cases. Again, the difference in the predictive accuracy in these cases is so small that the null cannot be rejected, suggesting that the improvement from the direct estimator is too small to be of practical forecasting value. For both measures of economic activity, the constant adjustment to the iterated method and the full-adjustment to both the iterated and direct methods perform very poorly and they never improve upon the benchmark. Indeed, the iterated method produces statistically significantly more accurate forecasts than these three forms of intercept correction in the clear majority of cases.

Inspection of Panels C and D reveal that the conclusions are substantially different for the price series. Most importantly, the iterated method performs worse in relative terms when future inflation is forecasted. For the GDP deflator, the one-off and full-adjustment to the iterated model dominate the iterated benchmark, with one exception, regardless of the forecast horizon and lag selection method. These improvements are large and generally statistically significant. In particular, the relative MSFE value at $h = 4$ for the full-adjustment method when an AR(1) specification is used is 0.691, indicating a 30.9% improvement relative to the benchmark. The results also show that the performance of the constant adjustment to the iterated method, the direct method and the full-adjustment to the direct method relative to the iterated benchmark depends on the method of lag selection. The ability of these methods to forecast the future GDP deflator is superior to the iterated benchmark in the majority of cases when the AR(1) model is used. On the other hand, if the results for the AR(1) specification are excluded, the iterated method is almost universally preferred to these three alternative methods. The good performance of these three methods when the AR(1) model is considered is probably due to the fact that low order AR models do not capture the true dynamics of the GDP deflator and are hence misspecified. At least the AR(1) model yields less accurate forecasts than the other lag methods.

The evidence for the one-off and full-adjustment to the iterated method is less convincing when changes in PCE inflation are forecasted. These methods generate smaller forecast errors than the iterated benchmark at $h = 8$ and $h = 12$. Although the improvements are quite large, the null of equal accuracy is rejected at conventional significance levels only for the AR(1) model. In contrast, the one-off and full-adjustment to the iterated method produce higher MSFE values than the benchmark at $h = 2$, sometimes by quite a substantial margin. According to the p-values, the null is rejected in favor of the iterated benchmark at this horizon in six of eight cases. The direct estimator beats the iterated one when the forecasts are computed using an AR(1) model, but using longer lags in the forecasting model eliminates the advantage of the direct estimator, particularly at long horizons ($h = 8$ and $h = 12$). In contrast with the GDP deflator results, the constant-adjustment to the iterated method and the full-adjustment to the direct method never produce better PCE inflation forecasts than the iterated benchmark. Indeed, at the longest horizon $h = 12$, these methods are markedly worse than the benchmark.

All in all, the results in Table 5 indicate that the iterated method provides the most accurate real-time output forecasts, whereas the one-off and full-adjustment to the iterated method help improve the accuracy of the inflation forecast. Thus, there seems to be no single dominant multi-step forecasting method (cf. Marcellino *et al.*, 2006; Pesaran *et al.*, 2011). Figure 1 plots the quarterly growth rates of the four macroeconomic time series (at an annualized rate) over the out-of-sample period. The figure demonstrates that the series have undergone different types of structural breaks. In particular, it is well documented that the volatility of the real GDP and industrial production growth have decreased since the mid-1980s (see, e.g., McConnell and Perez-Quiros, 2000). The simulation results in Section 4 show that when the volatility changes, the iterated method performs well relative to the other multi-step methods. On the other hand, due to changes in monetary policy, both the mean and variance of the two inflation variables have decreased substantially since the early 1980s (Sims and Zha, 2006). The Monte Carlo results show that when both the mean and variance decrease substantially, e.g., when the autoregressive parameter of an AR(1) model decreases substantially (see Appendix A), the iterated method yields rather poor forecasts. Hence, the Monte Carlo results are very helpful in understanding why it is difficult to find a single multi-step method that dominates across all variables.

The results in Section 4 also suggest that the timing of the break affects the accuracy of the multi-step methods, implying that the relative forecasting performance might be time-varying in an unstable environment. To examine this possibility, Figure 2 plots the Giacomini and Rossi (2010) Fluctuation test as well as the two-sided critical values at the 5% significance level (dashed horizontal lines) for an AR(4) model at $h = 4$. The Fluctuation test is implemented by using a centered rolling window of 40 observations. The truncation parameter is set to $P^{1/5} \approx 3$, where P denotes the number of out-of-sample observations. Positive (negative) values of the test indicate that the candidate multi-step forecasting method has produced more (less) accurate forecasts than the iterated benchmark. If the Fluctuation test statistic crosses either the upper or the lower critical value, the null of equal local predictive ability at each point in time is rejected.

Several results stand out. First, despite the large differences in the relative predictive ability reported in Table 5, the Fluctuation test rejects the null of equal accuracy at each point in time only in three cases. Interestingly, the Fluctuation test reveals that the one-off and full-adjustment to the iterated method contain substantial incremental real-time predictive information for the GDP deflator in the early 1980s. However, later in the sample, these two forms of intercept correction give less accurate forecasts than the iterated benchmark. Broadly speaking, these findings are consistent with the aforementioned observation that both the mean and variance of the GDP deflator have decreased substantially in the early 1980s. The simulation results in Tables 2–3 suggest that in the presence of large and recent decrease in both the mean and variance of a series only the one-off and full-adjustment to the iterated method of the five alternatives should dominate the benchmark (see the results for $T_1 = 99$). Furthermore, as time passes after the break, the gains from these two intercept corrections should diminish.

The Fluctuation test for the two output variables show that the track record of the constant adjustment to the iterated method and the full-adjustment to both the iterated and direct method is not good. In fact, the Fluctuation test implies that these methods yield systematically worse forecasts than the iterated benchmark over the whole out-of-sample period (the value of the test statistic is always negative), although the null of equal accuracy at each point in time cannot be rejected. Similarly, the direct estimator almost universally produces larger forecast errors for the price series than the iterated estimator.

Overall, the Fluctuation test indicates that the alternative multi-step methods only episodically improve upon the iterated benchmark. Therefore, the results over the whole out-of-sample period might give a somewhat misleading picture of their predictive ability. Most notably, the one-off and full-adjustment to the iterated method do not systematically beat the iterated benchmark when GDP deflator is forecasted, but rather they perform particularly well only in the early 1980s. The empirical results, as well as the simulation results, support the view that the iterated method typically produces the most accurate real-time forecasts in unstable environment. However, the results also highlight that if both the mean and variance of the series decrease substantially and the multi-step forecasts are made shortly after the break, the iterated method produces inaccurate forecasts and performs poorly in relative terms. In such a case, an alternative multi-step method, perhaps a one-off adjustment to the iterated method, should be used.

6. Conclusions

This paper analyzes the real-time performance of various multi-step forecasting methods in the presence of structural breaks. Our Monte Carlo and empirical analysis leads us to three main conclusions. First, our results suggest that the iterated method provides the most accurate multi-step forecasts in the presence of structural instability, especially if the parameters are subject to small or medium-size breaks. The good performance of the iterated method suggests that the error component dominates the bias component in the composition of MSFE values in an unstable environment. Second, the alternative multi-step methods, which are less prone to bias, have the most potential when the parameters are subject to large breaks and forecasts are made shortly after the break. Third, in the presence of breaks, the relative performance of the multi-step methods might be time-varying. For instance, it is only in the early 1980s that the one-off and full-adjustment to the iterated method provide more accurate GDP deflator forecasts than the iterated method.

The finding that the type as well as the timing of the break affects the relative merit of the multi-step methods is an intriguing one. The previous literature has found strong evidence for parameter instability in U.S. macroeconomic time series. These series have been subject to different types of breaks at different dates. This observation together with our findings might help explain why it is so difficult to find a single multi-step method that performs well across all variables at all time periods. Clearly, it would be interesting to analyze the time-variations further using the dataset of 170 U.S. monthly macroeconomic time series studied in Marcellino *et al.* (2006) and Pesaran *et al.* (2011).

References

Aruoba SB. 2008. Data revisions are not well-behaved. *Journal of Money, Credit and Banking* **40**: 319–340.

Bao Y. 2007. Finite-sample properties of forecasts from the stationary first-order autoregressive model under a general error distribution. *Econometric Theory* **23**: 767–773.

Brown BW, Mariano RS. 1989. Measures of deterministic prediction bias in nonlinear models. *International Economic Review* **30**: 667–684.

Chevillon G, Hendry DF. 2005. Non-parametric direct multi-step estimation for forecasting economic processes. *International Journal of Forecasting* **21**: 201–218.

Clements MP, Galvão AB. 2013. Real-time forecasting of inflation and output growth with autoregressive models in the presence of data revisions. *Journal of Applied Econometrics* **28**: 458–477.

Clements MP, Hendry DF. 1996a. Intercept corrections and structural change. *Journal of Applied Econometrics* **11**: 475–494.

Clements MP, Hendry DF. 1996b. Multi-step estimation for forecasting. *Oxford Bulletin of Economics and Statistics* **58**: 657–684.

Clements MP, Hendry DF. 1998. *Forecasting Economic Time Series*. Cambridge University Press: Cambridge, UK.

Clements MP, Hendry DF. 2006. Forecasting with breaks. In *Handbook of Economic Forecasting*, Vol. **1**., Elliott G, Granger CWJ, Timmermann A (eds). Elsevier: Amsterdam; 605–657.

Croushore D. 2011. Frontiers of real-time data analysis. *Journal of Economic Literature* **49**: 72–100.

Eklund J, Kapetanios G, Price S. 2013. Robust forecast methods and monitoring during structural change. *Manchester School* **81**: 3–27.

Elliott G, Timmermann A. 2008. Economic forecasting. *Journal of Economic Literature* **46**: 3–56.

Findley DF. 1985. Model selection for multi-step-ahead forecasting. In *Proceedings of the Seventh Symposium on Identification and System Parameter Estimation*, Baker HA, Young PC (eds). Pergamon: Oxford; 1039–1044

Giacomini R, Rossi B. 2010. Forecast comparisons in unstable environments. *Journal of Applied Econometrics* **25**: 595–620.

Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545–1578.

Hoque A, Magnus JR, Pesaran B. 1988. The exact multi-period mean square forecast error for the first-order autoregressive model. *Journal of Econometrics* **39**: 327–346.

Hännikäinen J. 2014. Selection of an estimation window in the presence of data revisions and recent structural breaks. Tampere Economic Working Paper, No. 92, University of Tampere. Downloadable at http://tampub.uta.fi/handle/10024/94782

Ing C-K. 2003. Multistep prediction in autoregressive processes. *Econometric Theory* **19**: 254–279.

Jacobs JPAM, van Norden S. 2011. Modeling data revisions: measurement error and dynamics of "true" values. *Journal of Econometrics* **161**: 101–109.

Mankiw NG, Shapiro MD. 1986. News or noise: an analysis of GNP revisions. *Survey of Current Business* **66**: 20–25.

Marcellino M, Stock JH, Watson MW. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* **135**: 499–526.

McConnell MM, Perez-Quiros G. 2000. Output fluctuations in the United States: what has changed since the early 1980's? *American Economic Review* **90**: 1464–1476.

Pesaran HM, Pick A, Timmermann A. 2011. Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics* **164**: 173–187.

Pesaran MH, Timmermann A. 2005. Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics* **129**: 183–217.

Rossi B. 2013. Advances in forecasting under instability. In *Handbook of Economic Forecasting*, Vol. **2**., Elliott G, Timmermann A (eds). Elsevier: Amsterdam; 1203–1324.

Schorfheide F. 2005. VAR forecasting under misspecification. *Journal of Econometrics* **128**: 99–136.

Sims CA, Zha T. 2006. Were there regime switches in U.S. monetary policy? *American Economic Review* **96**: 54–81.

Stock JH, Watson MW. 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics* **14**: 11–30.

Stock JH, Watson MW. 2003. Forecasting output and inflation: the role of asset prices. *Journal of Economic Literature* **41**: 788–829.

Weiss AA. 1991. Multi-step estimation and forecasting in dynamic models. *Journal of Econometrics* **48**: 135–149.

Table 1: Simulation setup

True process Experiments	ρ_1	ρ_2	β_1	β_2	σ_1	σ_2
1: No break	1	1	0.5	0.5	1.5	1.5
2: Moderate break in β (increase)	1	1	0.5	0.75	1.5	1.5
3: Moderate break in β (decrease)	1	1	0.5	0.25	1.5	1.5
4: Large break in β (increase)	1	1	0.25	0.75	1.5	1.5
5: Large break in β (decrease)	1	1	0.75	0.25	1.5	1.5
6: Increase in post-break variance	1	1	0.5	0.5	1.5	4.5
7: Decrease in post-break variance	1	1	0.5	0.5	1.5	0.5
8: Break in mean (increase)	1	1.5	0.5	0.5	1.5	1.5
9: Break in mean (decrease)	1	0.5	0.5	0.5	1.5	1.5

News Experiments	$\mu_{v}1_{1}$	$\mu_{v}2_{1}$	$\mu_{v}1_{5}$	$\mu_{v}2_{5}$	$\sigma_{v}1_{1}$	$\sigma_{v}2_{1}$	$\sigma_{v}1_{2,..,13}$	$\sigma_{v}2_{2,..,13}$	$\sigma_{v}1_{14}$	$\sigma_{v}2_{14}$
1: No break	0.085	0.085	0.043	0.043	0.783	0.783	0.391	0.391	0.196	0.196
2: Moderate break in β (increase)	0.085	0.195	0.043	0.098	0.783	2.238	0.391	1.119	0.196	0.560
3: Moderate break in β (decrease)	0.085	0.054	0.043	0.027	0.783	0.634	0.391	0.317	0.196	0.158
4: Large break in β (increase)	0.054	0.195	0.027	0.098	0.634	2.238	0.317	1.119	0.158	0.560
5: Large break in β (decrease)	0.195	0.054	0.098	0.027	2.238	0.634	1.119	0.317	0.560	0.158
6: Increase in post-break variance	0.085	0.085	0.043	0.043	0.783	2.348	0.391	1.174	0.196	0.587
7: Decrease in post-break variance	0.085	0.085	0.043	0.043	0.783	0.261	0.391	0.130	0.196	0.065
8: Break in mean (increase)	0.085	0.128	0.043	0.064	0.783	0.783	0.391	0.391	0.196	0.196
9: Break in mean (decrease)	0.085	0.043	0.043	0.021	0.783	0.783	0.391	0.391	0.196	0.196

Noise Experiments	$\mu_{\varepsilon}1_{1}$	$\mu_{\varepsilon}2_{1}$	$\mu_{\varepsilon}1_{2,\ldots,5}$	$\mu_{\varepsilon}2_{2,\ldots,5}$	$\sigma_{\varepsilon}1_{1}$	$\sigma_{\varepsilon}2_{1}$	$\sigma_{\varepsilon}1_{2,4,\ldots,14}$	$\sigma_{\varepsilon}2_{2,4,\ldots,14}$	$\sigma_{\varepsilon}1_{3,5,\ldots,13}$	$\sigma_{\varepsilon}2_{3,5,\ldots,13}$
1: No break	0.113	0.113	0.038	0.038	0.728	0.728	0.188	0.188	0.325	0.325
2: Moderate break in β (increase)	0.113	0.226	0.038	0.075	0.728	0.953	0.188	0.246	0.325	0.426
3: Moderate break in β (decrease)	0.113	0.075	0.038	0.025	0.728	0.651	0.188	0.168	0.325	0.291
4: Large break in β (increase)	0.075	0.226	0.025	0.075	0.651	0.953	0.168	0.246	0.291	0.426
5: Large break in β (decrease)	0.226	0.075	0.075	0.025	0.953	0.651	0.246	0.168	0.426	0.291
6: Increase in post-break variance	0.113	0.113	0.038	0.038	0.728	2.183	0.188	0.564	0.325	0.976
7: Decrease in post-break variance	0.113	0.113	0.038	0.038	0.728	0.243	0.188	0.063	0.325	0.108
8: Break in mean (increase)	0.113	0.170	0.038	0.057	0.728	0.728	0.188	0.188	0.325	0.325
9: Break in mean (decrease)	0.113	0.057	0.038	0.019	0.728	0.728	0.188	0.188	0.325	0.325

Table 2: Relative MSFE values when revisions add news

Break date		$T_1 = 25$				$T_1 = 50$				$T_1 = 99$			
Forecast horizon		2	4	8	12	2	4	8	12	2	4	8	12
Exp.1	Constant	1.419	1.623	1.713	1.729	–	–	–	–	–	–	–	–
	One-off	1.040	1.003	1.000	1.000	–	–	–	–	–	–	–	–
	Full	1.182	1.175	1.178	1.183	–	–	–	–	–	–	–	–
	Direct	1.009	1.017	1.024	1.027	–	–	–	–	–	–	–	–
	Full direct	1.380	1.481	1.504	1.514	–	–	–	–	–	–	–	–
Exp.2	Constant	1.436	1.821	2.360	2.528	1.427	1.787	2.271	2.483	0.999	1.024	1.058	1.070
	One-off	1.064	1.016	1.004	1.001	1.058	1.015	1.004	1.001	0.981	0.995	1.000	1.000
	Full	1.129	1.097	1.107	1.105	1.121	1.090	1.096	1.104	0.979	0.987	1.004	1.009
	Direct	1.007	1.026	1.047	1.052	1.009	1.024	1.048	1.068	1.006	1.010	1.008	1.008
	Full direct	1.398	1.601	1.723	1.713	1.389	1.592	1.740	1.786	1.007	1.029	1.047	1.060
Exp.3	Constant	1.369	1.406	1.412	1.447	1.344	1.339	1.334	1.341	1.416	1.511	1.464	1.474
	One-off	1.024	0.999	1.000	1.000	1.022	0.993	0.999	1.000	1.044	0.995	0.999	1.000
	Full	1.205	1.167	1.161	1.176	1.166	1.082	1.060	1.070	1.186	1.114	1.059	1.060
	Direct	1.006	1.011	1.011	1.006	1.003	1.004	1.006	0.997	1.009	1.028	1.032	1.038
	Full direct	1.332	1.351	1.346	1.370	1.296	1.261	1.235	1.248	1.406	1.475	1.406	1.413
Exp.4	Constant	1.450	1.838	2.351	2.503	1.396	1.797	2.330	2.544	0.961	0.989	1.004	1.014
	One-off	1.069	1.018	1.006	1.001	1.045	1.015	1.007	1.002	0.984	0.999	1.000	1.000
	Full	1.135	1.101	1.103	1.092	1.102	1.085	1.092	1.094	0.960	0.983	0.995	1.003
	Direct	1.003	1.014	1.032	1.052	1.003	1.031	1.068	1.073	1.006	1.005	1.004	1.003
	Full direct	1.408	1.612	1.722	1.728	1.368	1.634	1.814	1.877	0.973	0.997	1.010	1.018
Exp.5	Constant	1.390	1.471	1.435	1.476	1.235	1.160	1.013	0.997	1.271	1.272	1.267	1.285
	One-off	1.060	0.983	0.984	0.993	0.981	0.919	0.962	0.985	0.949	0.907	0.964	0.987
	Full	1.144	1.023	0.934	0.924	1.019	0.849	0.764	0.752	0.990	0.838	0.770	0.761
	Direct	0.994	0.952	0.875	0.832	0.986	0.944	0.886	0.845	1.028	1.091	1.159	1.186
	Full direct	1.275	1.189	0.981	0.964	1.106	0.861	0.609	0.575	1.617	1.875	1.792	1.750
Exp.6	Constant	1.457	1.657	1.768	1.775	1.450	1.627	1.808	1.795	1.121	1.196	1.240	1.220
	One-off	1.054	1.007	1.000	1.000	1.053	1.004	1.001	1.000	1.013	1.003	1.000	1.000
	Full	1.210	1.194	1.203	1.193	1.200	1.176	1.211	1.200	1.050	1.054	1.062	1.051
	Direct	1.008	1.022	1.024	1.033	1.015	1.026	1.032	1.033	1.003	1.012	1.007	1.014
	Full direct	1.421	1.513	1.552	1.564	1.411	1.496	1.576	1.568	1.068	1.090	1.097	1.097
Exp.7	Constant	1.314	1.477	1.531	1.541	1.264	1.382	1.397	1.403	3.431	4.419	4.663	4.721
	One-off	0.997	0.992	0.998	1.000	0.974	0.984	0.998	1.000	1.212	1.007	0.999	1.000
	Full	1.106	1.096	1.084	1.088	1.063	1.032	1.018	1.014	2.057	1.966	1.905	1.911
	Direct	1.013	1.018	1.021	0.997	1.017	1.025	1.023	1.027	1.032	1.070	1.094	1.109
	Full direct	1.288	1.348	1.355	1.358	1.232	1.263	1.240	1.252	3.494	4.074	4.044	4.091
Exp.8	Costant	1.438	1.663	1.773	1.775	1.412	1.625	1.720	1.704	1.338	1.493	1.535	1.563
	One-off	1.048	1.004	1.000	1.000	1.038	1.001	1.000	1.000	1.026	1.000	1.000	1.000
	Full	1.188	1.174	1.180	1.174	1.168	1.157	1.149	1.140	1.139	1.130	1.116	1.127
	Direct	1.007	1.014	1.021	1.022	1.006	1.013	1.013	1.003	1.010	1.020	1.026	1.031
	Full direct	1.398	1.520	1.526	1.547	1.376	1.485	1.491	1.495	1.327	1.420	1.415	1.440
Exp.9	Costant	1.406	1.652	1.682	1.716	1.361	1.551	1.589	1.601	1.269	1.391	1.400	1.429
	One-off	1.038	1.004	1.000	1.000	1.022	0.994	0.999	1.000	1.004	0.995	0.999	1.000
	Full	1.168	1.174	1.143	1.157	1.134	1.106	1.082	1.076	1.091	1.078	1.057	1.069
	Direct	1.007	1.015	1.014	1.013	1.007	1.017	1.011	1.005	1.011	1.023	1.029	1.031
	Full direct	1.370	1.526	1.477	1.499	1.328	1.423	1.392	1.396	1.260	1.316	1.292	1.326

Notes: The experiments are as defined in Table 1. 'Constant' denotes the method of constant adjustment to the iterated model; 'One-off' denotes the one-off adjustment to the iterated method. 'Full' and 'Full direct' denote full adjustment to the iterated and direct methods, respectively. Intercept corrections are based on the average of the latest 4 forecast errors. The sample size is $T = 100$. The break occurs at $T_1 = 25$, 50, or 99. MSFE values are computed relative to those produced by the iterated forecasting method.

Table 3: Relative MSFE values when revisions reduce noise

Break date		$T_1 = 25$				$T_1 = 50$				$T_1 = 99$			
Forecast horizon		2	4	8	12	2	4	8	12	2	4	8	12
Exp.1	Constant	1.413	1.639	1.680	1.701	–	–	–	–	–	–	–	–
	One-off	1.046	1.006	1.000	1.000	–	–	–	–	–	–	–	–
	Full	1.186	1.193	1.170	1.176	–	–	–	–	–	–	–	–
	Direct	1.008	1.015	1.015	1.022	–	–	–	–	–	–	–	–
	Full direct	1.354	1.470	1.451	1.482	–	–	–	–	–	–	–	–
Exp.2	Constant	1.492	1.877	2.328	2.633	1.431	1.750	2.120	2.234	1.118	1.196	1.296	1.325
	One-off	1.104	1.029	1.005	1.002	1.063	1.006	1.001	1.000	0.970	0.990	1.000	1.000
	Full	1.177	1.124	1.107	1.122	1.132	1.078	1.067	1.065	1.007	1.011	1.055	1.072
	Direct	1.004	1.013	1.028	1.029	1.004	1.019	1.024	1.017	1.018	1.029	1.016	1.012
	Full direct	1.403	1.608	1.718	1.806	1.368	1.583	1.716	1.723	1.112	1.159	1.208	1.227
Exp.3	Constant	1.383	1.424	1.377	1.419	1.350	1.407	1.429	1.423	1.431	1.521	1.498	1.490
	One-off	1.034	1.000	1.000	1.000	1.030	0.999	1.000	1.000	1.060	1.002	0.999	1.000
	Full	1.222	1.189	1.146	1.174	1.177	1.135	1.126	1.123	1.207	1.145	1.099	1.096
	Direct	1.008	1.013	1.008	1.011	1.007	1.012	1.008	1.005	1.006	1.016	1.028	1.031
	Full direct	1.347	1.360	1.299	1.347	1.302	1.314	1.314	1.319	1.393	1.447	1.391	1.402
Exp.4	Constant	1.517	1.902	2.385	2.668	1.345	1.636	1.971	2.095	0.947	1.014	1.056	1.074
	One-off	1.113	1.033	1.006	1.002	1.029	0.995	0.998	0.999	0.958	0.997	1.000	1.000
	Full	1.188	1.132	1.109	1.103	1.080	1.030	1.025	1.016	0.923	0.975	1.011	1.025
	Direct	1.004	1.015	1.028	1.024	0.992	0.995	1.009	1.003	1.026	1.017	1.012	1.009
	Full direct	1.440	1.667	1.778	1.809	1.321	1.582	1.734	1.767	0.975	1.025	1.054	1.066
Exp.5	Constant	1.455	1.569	1.576	1.617	1.430	1.505	1.541	1.489	1.089	1.048	1.005	0.993
	One-off	1.089	1.004	0.996	0.999	1.089	0.983	0.982	0.993	0.956	0.932	0.972	0.990
	Full	1.206	1.105	1.038	1.029	1.161	1.008	0.913	0.883	0.966	0.860	0.806	0.793
	Direct	1.027	1.012	0.952	0.932	1.009	0.962	0.867	0.830	1.012	1.058	1.120	1.137
	Full direct	1.407	1.445	1.327	1.275	1.328	1.263	1.088	0.985	1.133	1.174	1.128	1.109
Exp.6	Constant	1.473	1.644	1.746	1.730	1.484	1.688	1.740	1.811	1.204	1.290	1.295	1.297
	One-off	1.071	1.008	1.000	1.000	1.072	1.010	1.000	1.000	1.036	1.006	1.000	1.000
	Full	1.229	1.196	1.195	1.187	1.232	1.204	1.180	1.205	1.100	1.091	1.073	1.073
	Direct	1.004	1.023	1.027	1.031	1.008	1.023	1.035	1.039	1.006	1.011	1.018	1.016
	Full direct	1.413	1.490	1.503	1.510	1.429	1.510	1.494	1.546	1.118	1.121	1.111	1.108
Exp.7	Constant	1.372	1.532	1.587	1.573	1.329	1.479	1.530	1.543	3.371	4.432	4.723	4.886
	One-off	1.028	0.999	0.999	1.000	1.011	0.993	0.999	1.000	1.232	1.019	0.999	1.000
	Full	1.155	1.135	1.119	1.107	1.122	1.105	1.091	1.086	2.051	2.008	1.948	1.992
	Direct	1.011	1.025	1.021	1.016	1.008	1.025	1.017	1.025	1.020	1.043	1.069	1.087
	Full direct	1.326	1.382	1.373	1.354	1.275	1.348	1.338	1.350	3.510	4.229	4.182	4.288
Exp.8	Constant	1.450	1.691	1.780	1.785	1.426	1.655	1.730	1.757	1.383	1.524	1.602	1.591
	One-off	1.062	1.008	1.000	1.000	1.058	1.008	1.000	1.000	1.045	1.004	1.000	1.000
	Full	1.203	1.192	1.186	1.181	1.185	1.165	1.151	1.149	1.173	1.149	1.153	1.140
	Direct	1.009	1.015	1.018	1.012	1.006	1.010	1.013	1.011	1.005	1.014	1.025	1.028
	Full direct	1.404	1.521	1.529	1.524	1.378	1.492	1.506	1.530	1.351	1.429	1.454	1.440
Exp.9	Constant	1.472	1.686	1.738	1.759	1.417	1.599	1.674	1.655	1.284	1.375	1.416	1.447
	One-off	1.067	1.008	1.000	1.000	1.048	1.000	0.999	1.000	1.016	0.995	0.999	1.000
	Full	1.214	1.187	1.159	1.163	1.171	1.131	1.112	1.101	1.110	1.074	1.070	1.080
	Direct	1.004	1.014	1.013	1.007	1.006	1.005	1.002	1.004	1.010	1.021	1.027	1.026
	Full direct	1.401	1.507	1.478	1.486	1.362	1.429	1.446	1.433	1.264	1.280	1.285	1.309

See the notes to Table 2.

Table 4: Squared bias relative to the MSFE of the iterated benchmark when revisions add news

Break date		$T_1 = 25$				$T_1 = 50$				$T_1 = 99$			
Forecast horizon		2	4	8	12	2	4	8	12	2	4	8	12
Exp.1	Iterated	0.003	0.003	0.004	0.004	–	–	–	–	–	–	–	–
	Constant	0.000	0.000	0.000	0.000	–	–	–	–	–	–	–	–
	One-off	0.002	0.003	0.004	0.004	–	–	–	–	–	–	–	–
	Full	0.001	0.001	0.001	0.001	–	–	–	–	–	–	–	–
	Direct	0.004	0.003	0.004	0.004	–	–	–	–	–	–	–	–
	Full direct	0.000	0.000	0.000	0.000	–	–	–	–	–	–	–	–
Exp.2	Iterated	0.002	0.004	0.004	0.006	0.014	0.025	0.033	0.036	0.072	0.118	0.158	0.174
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.057	0.098	0.135	0.150
	One-off	0.000	0.003	0.004	0.005	0.005	0.019	0.031	0.035	0.067	0.116	0.157	0.174
	Full	0.000	0.001	0.002	0.003	0.002	0.009	0.015	0.017	0.062	0.107	0.146	0.162
	Direct	0.002	0.004	0.004	0.003	0.014	0.026	0.033	0.030	0.073	0.121	0.160	0.177
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.067	0.117	0.156	0.173
Exp.3	Iterated	0.020	0.027	0.022	0.027	0.058	0.071	0.087	0.078	0.155	0.210	0.236	0.222
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.052	0.063	0.074	0.066
	One-off	0.011	0.026	0.022	0.027	0.030	0.064	0.087	0.078	0.116	0.198	0.235	0.222
	Full	0.001	0.003	0.002	0.003	0.005	0.009	0.016	0.013	0.079	0.119	0.142	0.131
	Direct	0.019	0.023	0.014	0.014	0.056	0.064	0.073	0.061	0.161	0.225	0.251	0.235
	Full direct	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.001	0.074	0.111	0.129	0.118
Exp.4	Iterated	0.005	0.010	0.017	0.016	0.024	0.037	0.051	0.053	0.127	0.191	0.240	0.253
	Constant	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.104	0.166	0.213	0.226
	One-off	0.002	0.008	0.017	0.016	0.009	0.028	0.048	0.052	0.122	0.191	0.240	0.253
	Full	0.001	0.004	0.010	0.009	0.004	0.012	0.022	0.024	0.109	0.172	0.220	0.232
	Direct	0.005	0.010	0.013	0.008	0.023	0.035	0.045	0.039	0.129	0.194	0.242	0.254
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.111	0.176	0.224	0.235
Exp.5	Iterated	0.066	0.136	0.185	0.190	0.163	0.324	0.421	0.439	0.494	0.623	0.687	0.693
	Constant	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.091	0.076	0.059	0.051
	One-off	0.024	0.105	0.175	0.186	0.050	0.236	0.392	0.429	0.287	0.512	0.654	0.682
	Full	0.010	0.047	0.082	0.086	0.021	0.113	0.196	0.216	0.220	0.351	0.428	0.441
	Direct	0.055	0.091	0.089	0.061	0.150	0.275	0.325	0.299	0.515	0.686	0.790	0.804
	Full direct	0.000	0.000	0.000	0.002	0.001	0.001	0.006	0.016	0.215	0.355	0.448	0.455
Exp.6	Iterated	0.001	0.000	0.000	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.001	0.000
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	One-off	0.000	0.000	0.000	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.001	0.000
	Full	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Direct	0.001	0.000	0.000	0.001	0.000	0.002	0.001	0.000	0.000	0.000	0.001	0.000
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Exp.7	Iterated	0.023	0.031	0.033	0.027	0.023	0.025	0.024	0.030	0.015	0.018	0.020	0.025
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003
	One-off	0.012	0.028	0.033	0.027	0.011	0.022	0.023	0.030	0.009	0.016	0.020	0.025
	Full	0.003	0.008	0.009	0.006	0.002	0.004	0.004	0.007	0.004	0.006	0.008	0.011
	Direct	0.024	0.034	0.036	0.028	0.024	0.028	0.025	0.032	0.015	0.019	0.023	0.027
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003
Exp.8	Iterated	0.001	0.003	0.002	0.002	0.020	0.027	0.033	0.040	0.119	0.162	0.177	0.185
	Constant	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.091	0.123	0.133	0.140
	One-off	0.000	0.003	0.002	0.002	0.008	0.023	0.033	0.040	0.109	0.159	0.177	0.185
	Full	0.000	0.001	0.000	0.001	0.002	0.005	0.008	0.011	0.100	0.141	0.155	0.163
	Direct	0.001	0.002	0.000	0.000	0.019	0.025	0.026	0.026	0.121	0.169	0.183	0.190
	Full direct	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.112	0.168	0.189	0.194
Exp.9	Iterated	0.018	0.022	0.021	0.020	0.058	0.075	0.083	0.089	0.173	0.231	0.260	0.257
	Constant	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.081	0.109	0.123	0.121
	One-off	0.008	0.019	0.021	0.020	0.027	0.064	0.082	0.089	0.140	0.222	0.259	0.257
	Full	0.002	0.004	0.004	0.004	0.008	0.015	0.020	0.023	0.107	0.159	0.183	0.181
	Direct	0.017	0.019	0.014	0.009	0.058	0.072	0.071	0.069	0.179	0.243	0.273	0.268
	Full direct	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001	0.101	0.150	0.172	0.168

Notes: The table reports the squared bias of the different methods as a ratio of the MSFE of the iterated benchmark model.

Table 5: MSFE values relative to the iterated benchmark based on the same lag selection method

Forecast horizon		AR(1)				AR(4)				BIC				AIC			
		2	4	8	12	2	4	8	12	2	4	8	12	2	4	8	12
(A) *GDP*																	
	Iterated	*2.664*	*2.719*	*2.723*	*2.688*	*2.716*	*2.786*	*2.720*	*2.682*	*2.673*	*2.724*	*2.720*	*2.685*	*2.663*	*2.724*	*2.722*	*2.686*
	Constant	1.227	1.343	1.518**	1.742***	1.257	1.366	1.660**	1.908***	1.279	1.461*	1.654**	1.923***	1.275	1.426	1.603**	1.873***
	One-off	0.978	0.995	1.000	1.000	0.995	0.993	1.003	1.001	0.993	1.004	0.999	1.000	0.995	1.002	0.999	1.000
	Full	1.075	1.098	1.205	1.338***	1.114	1.082	1.239	1.385***	1.127	1.142	1.219	1.355***	1.127	1.126	1.199	1.337***
	Direct	0.978	1.019	1.015	0.999	1.022	1.010	1.040	1.015	0.971	1.023	1.017	1.001	1.006	1.038	1.021	1.005
	Full direct	1.217	1.387	1.375*	1.483***	1.236	1.347	1.491**	1.466***	1.208	1.384*	1.379*	1.486***	1.270*	1.405*	1.424**	1.481***
(B) *Industrial production*																	
	Iterated	*6.639*	*6.744*	*6.748*	*6.859*	*6.705*	*6.798*	*6.767*	*6.832*	*6.618*	*6.744*	*6.749*	*6.860*	*6.781*	*6.716*	*6.752*	*6.831*
	Constant	1.504**	1.787***	1.575**	1.872***	1.424*	1.714**	1.653**	1.938***	1.505**	1.786***	1.571**	1.865***	1.395*	1.693**	1.619**	1.880***
	One-off	1.057	1.009	1.000	1.000	1.033	1.028	1.002	1.001	1.057	1.009	1.000	1.000	1.025	1.015	1.001	1.001*
	Full	1.272*	1.359**	1.195	1.388***	1.208	1.297*	1.253*	1.429***	1.271*	1.359**	1.195	1.386***	1.191	1.301*	1.247*	1.436***
	Direct	1.004	1.004	1.020	0.978	0.997	0.982	1.065	1.004	1.012	1.000	1.019	0.981	0.976	1.019	1.015	0.997
	Full direct	1.472**	1.596**	1.408**	1.657***	1.380*	1.497**	1.569***	1.593***	1.486**	1.584**	1.400**	1.655***	1.438*	1.594**	1.405**	1.684***
(C) *GDP deflator*																	
	Iterated	*1.479*	*1.779*	*2.279*	*2.444*	*1.231*	*1.396*	*1.851*	*2.122*	*1.289*	*1.436*	*1.910*	*2.186*	*1.246*	*1.412*	*1.861*	*2.124*
	Constant	0.971	0.725	0.811	1.114	1.151	1.158	1.471*	2.040***	1.123	1.122	1.384	1.907**	1.123	1.142	1.474*	2.063***
	One-off	0.872**	0.838***	0.948***	0.983***	0.967	0.891***	0.937**	0.968*	0.955	0.878***	0.926***	0.964**	0.957	0.889***	0.935**	0.967*
	Full	0.857*	0.691***	0.718***	0.764***	1.015	0.866*	0.853**	0.900*	0.992	0.831**	0.813***	0.862**	0.995	0.856**	0.847**	0.896*
	Direct	0.886***	0.675***	0.756***	0.895	1.019	1.076***	1.196***	1.292**	0.965	1.037	1.080**	1.231*	1.005	1.059**	1.160***	1.318**
	Full direct	0.869	0.665*	0.857	1.127	1.096	1.155	1.487	1.757*	0.998	1.023	1.225	1.504	1.062	1.094	1.381	1.720*
(D) *PCE inflation*																	
	Iterated	*1.898*	*2.051*	*2.482*	*2.678*	*1.830*	*1.955*	*2.323*	*2.555*	*1.847*	*2.016*	*2.389*	*2.609*	*1.823*	*1.960*	*2.337*	*2.568*
	Constant	1.364***	1.345*	1.397	1.629*	1.485***	1.658***	1.985***	2.404***	1.385***	1.473**	1.669**	2.078**	1.446***	1.604***	1.892***	2.334***
	One-off	1.061	0.940*	0.954***	0.980**	1.080**	1.012	0.969	0.982	1.061	0.990	0.962*	0.983	1.067*	1.005	0.969	0.984
	Full	1.129**	0.949	0.879**	0.867***	1.247***	1.085	0.986	0.958	1.168**	1.025	0.936	0.925	1.221**	1.078	0.976	0.953
	Direct	1.009	0.947	0.959	0.983	0.994	1.018	1.162**	1.146	1.003	0.967	1.036	1.070	0.997	1.016	1.121*	1.125
	Full direct	1.241**	1.079	1.127	1.446	1.333***	1.244	1.553*	1.785*	1.305**	1.166	1.220	1.552	1.330**	1.230	1.419	1.630

Notes: Forecast period spans from 1977:Q2 to 2013:Q2. The first row in each panel shows the root mean squared forecast error for the iterated benchmark. Subsequent rows report the ratio of the MSFE of each candidate multi-step method relative to the MSFE of the iterated benchmark. Intercept corrections are based on the average of the latest 4 forecast errors. Asterisks mark rejection of the two-sided Giacomini and White (2006) test at the 1%(***), 5%(**), and 10%(*) significance levels, respectively.

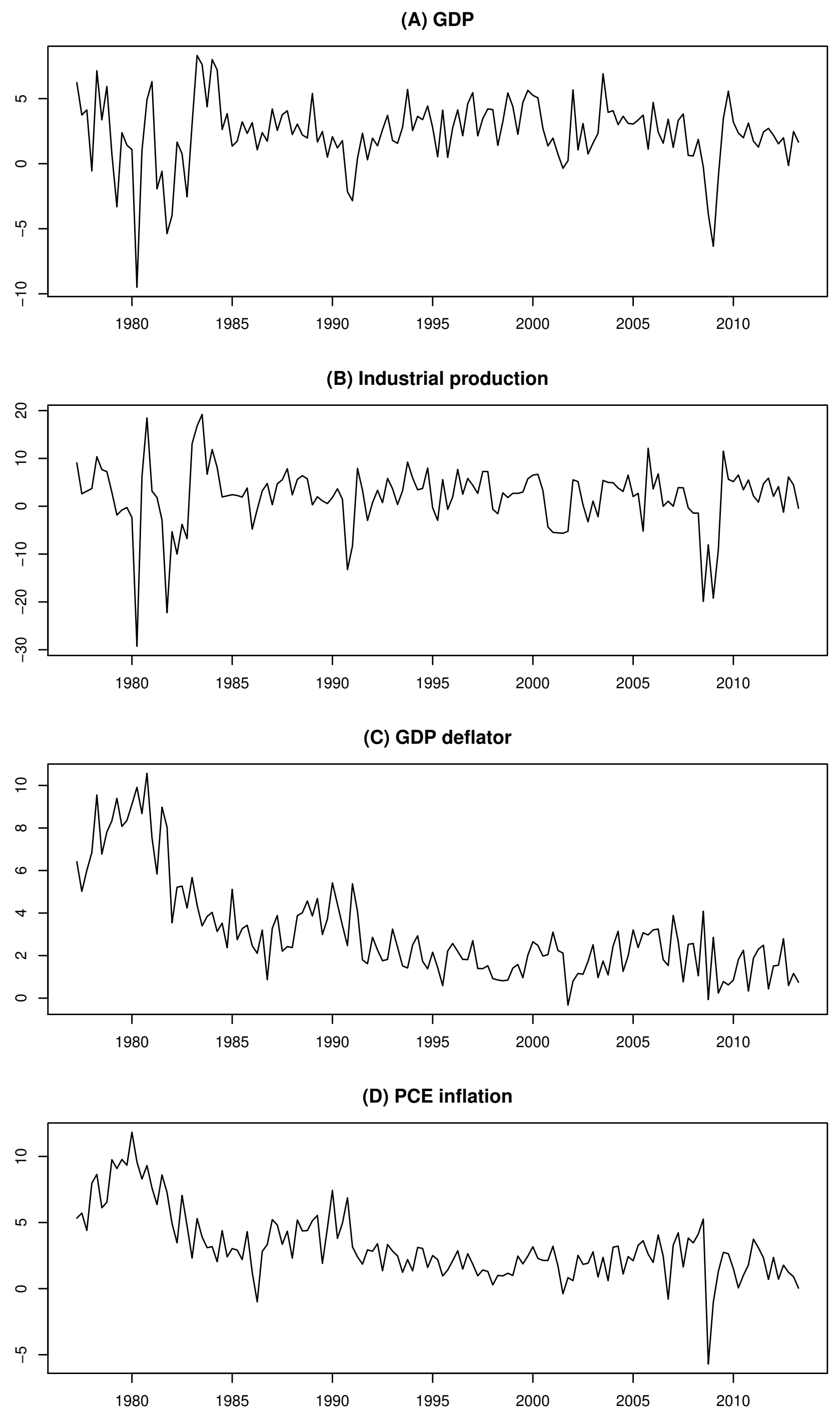
Appendix A

Table 6: Means and standard deviations

News Experiment	$E(\tilde{y}_{1t})$	$E(\tilde{y}_{2t})$	$E(y_{1t}^{t+1})$	$E(y_{2t}^{t+1})$	$\sigma_{\tilde{y}_{1t}}$	$\sigma_{\tilde{y}_{2t}}$	$\sigma_{y_{1t}^{t+1}}$	$\sigma_{y_{2t}^{t+1}}$
1	2.255	2.255	2.128	2.128	2.514	2.514	1.957	1.957
2	2.255	5.171	2.128	4.878	2.514	7.187	1.957	5.595
3	2.255	1.442	2.128	1.361	2.514	2.035	1.957	1.584
4	1.442	5.171	1.361	4.878	2.035	7.187	1.584	5.595
5	5.171	1.442	4.878	1.361	7.187	2.035	5.595	1.584
6	2.255	2.255	2.128	2.128	2.514	7.541	1.957	5.871
7	2.255	2.255	2.128	2.128	2.514	0.838	1.957	0.652
8	2.255	3.383	2.128	3.191	2.514	2.514	1.957	1.957
9	2.255	1.128	2.128	1.064	2.514	2.514	1.957	1.957

Noise Experiment	$E(\tilde{y}_{1t})$	$E(\tilde{y}_{2t})$	$E(y_{1t}^{t+1})$	$E(y_{2t}^{t+1})$	$\sigma_{\tilde{y}_{1t}}$	$\sigma_{\tilde{y}_{2t}}$	$\sigma_{y_{1t}^{t+1}}$	$\sigma_{y_{2t}^{t+1}}$
1	2.000	2.000	1.887	1.887	1.732	1.732	1.879	1.879
2	2.000	4.000	1.887	3.774	1.732	2.268	1.879	2.460
3	2.000	1.333	1.887	1.258	1.732	1.549	1.879	1.680
4	1.333	4.000	1.258	3.774	1.549	2.268	1.680	2.460
5	4.000	1.333	3.774	1.258	2.268	1.549	2.460	1.680
6	2.000	2.000	1.887	1.887	1.732	5.196	1.879	5.636
7	2.000	2.000	1.887	1.887	1.732	0.577	1.879	0.626
8	2.000	3.000	1.887	2.830	1.732	1.732	1.879	1.879
9	2.000	1.000	1.887	0.943	1.732	1.732	1.879	1.879

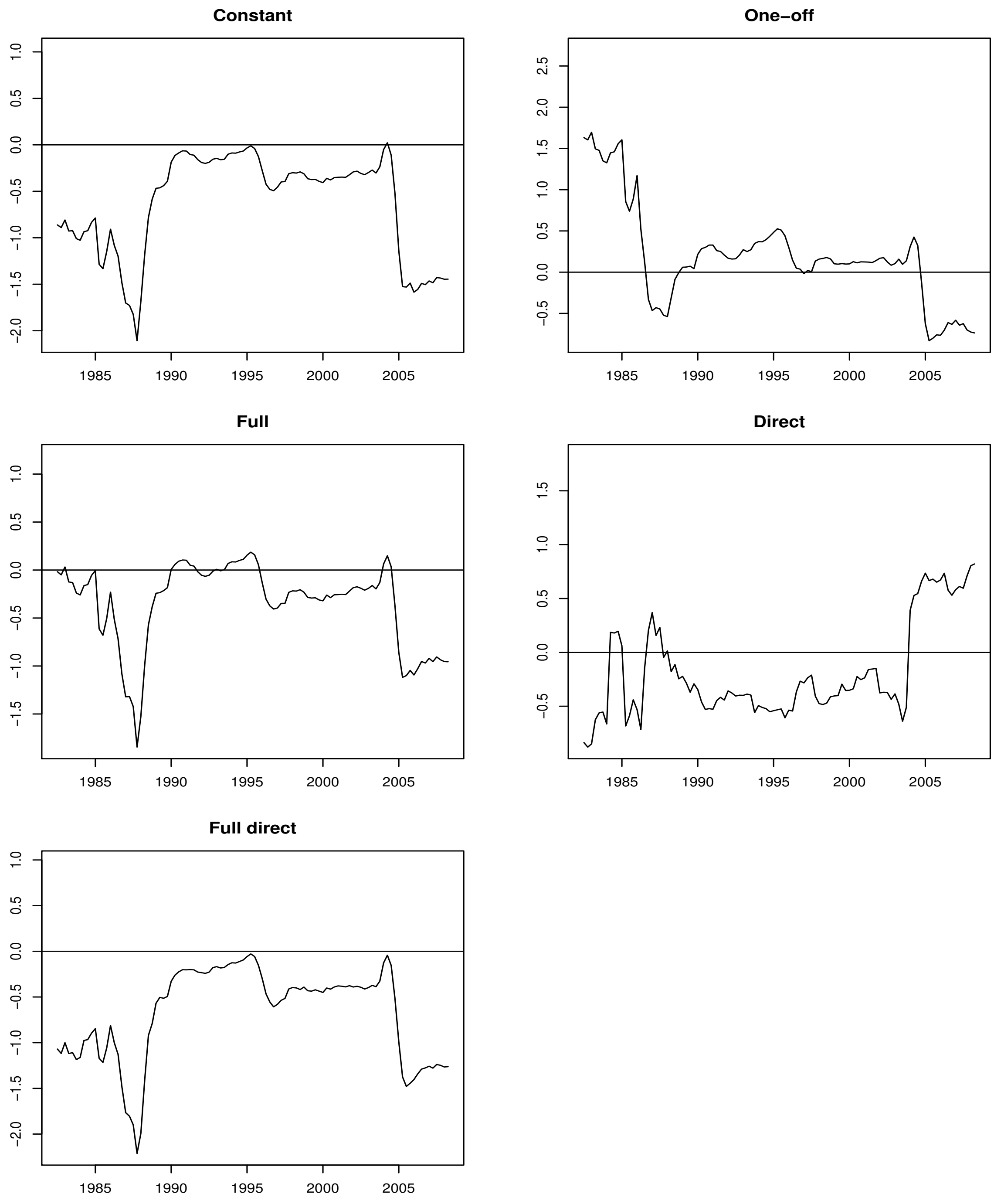
Figure 1: Quarterly growth rates



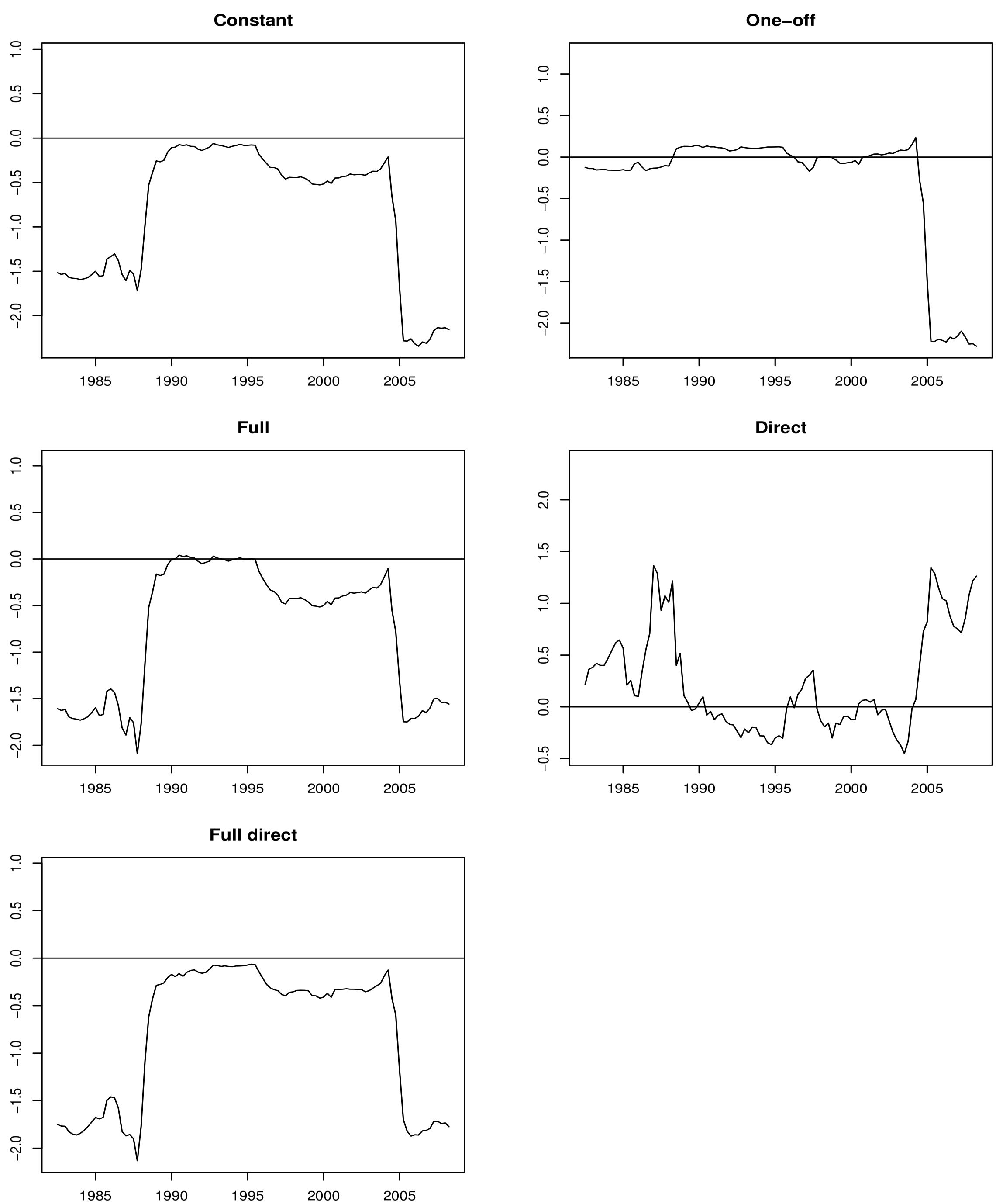
Notes: Sample period 1977:Q2—2013:Q2. The Figure plots the first-release growth rates, annualized.

Figure 2: Fluctuation test for equal out-of-sample predictability at $h = 4$

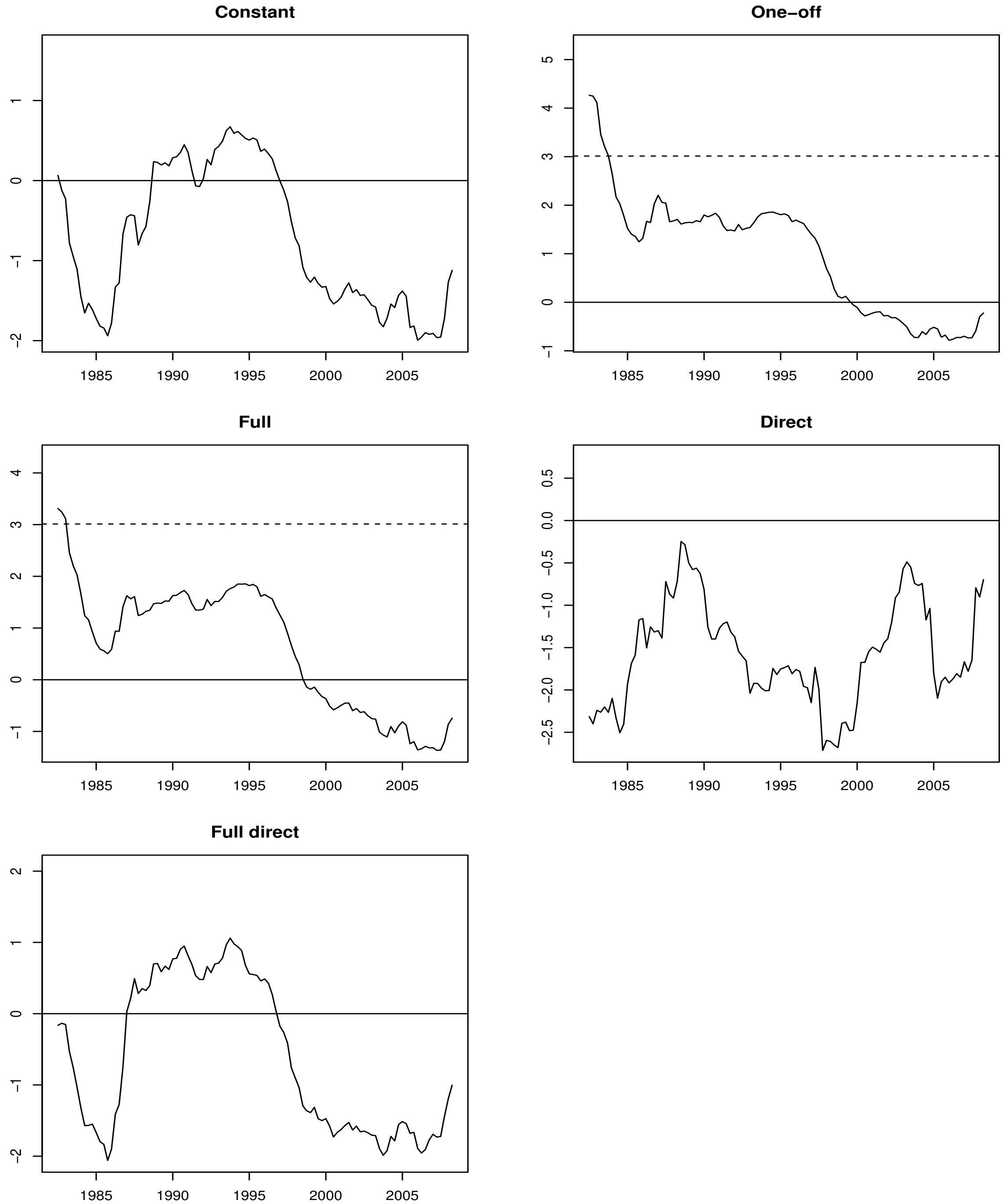
(A) GDP



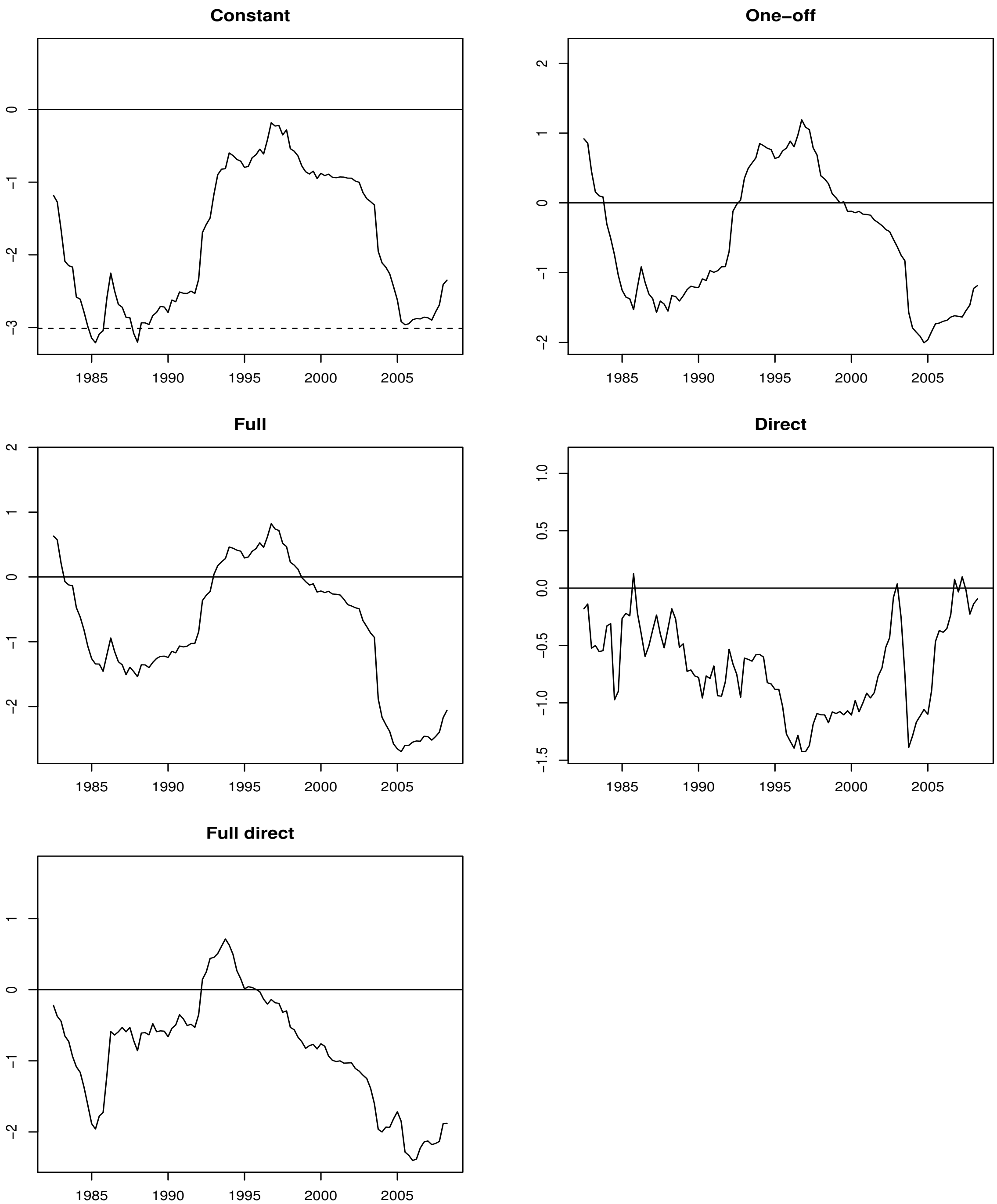
(B) Industrial production



(C) GDP deflator



(D) PCE inflation



Notes: The Figure plots the two-sided Giacomini and Rossi (2010) Fluctuation test based on sequences of the Giacomini and White (2006) unconditional test statistic for AR(4) specification. The test is implemented by using a centered rolling window of 40 observations. The sample period spans from 1977:Q4 to 2013:Q2. Positive (negative) values indicate that the candidate method has produced more (less) accurate forecasts than the benchmark. The dashed lines represent critical values at the 5% level. If the absolute value of the Fluctuation test exceeds the critical value, the null that the two multi-step methods have equal predictive ability at each point in time is rejected.