Detecting and Predicting Forecast Breakdowns*

Raffaella Giacomini and Barbara Rossi

University of California, Los Angeles and Duke University

September 2005

PRELIMINARY AND INCOMPLETE

Abstract

We propose a theoretical framework for assessing the reliability of a forecast model, characterized by: (1) in-sample performance being indicative of out-of-sample performance; and (2) stable performance over time. We accordingly define a forecast breakdown as a situation where the out-of-sample performance of the model, judged by some loss function, is significantly worse than its in-sample performance. Our framework, which is valid under general conditions, can be used to establish the past reliability of the model, and to predict whether the model will break down at a future date. We show that main causes of forecast breakdowns are instabilities in the data generating process and relate the properties of our test to those of previous structural break tests. The main differences are that our test is robust to the presence of unstable regressors and that it has greater power than previous tests to capture systematic forecast errors caused by recurring breaks that are ignored by the forecast model. We find evidence of a forecast breakdown in the Phillips' curve forecasts of U.S. inflation over the past three decades, and link the breakdown to changes in the monetary policy reaction function of the Fed.

J.E.L. Codes: C22, C52, C53

*__Acknowledgements:__ We would like to thank Todd Clark, Frank Diebold, Graham Elliott, Mike McCracken, Ulrich Muller and seminar participants to the 2005 Missouri Economics Conference, the Duke Financial Econometrics Group, the 2005 CIRANO-CIREQ Conference on Forecasting in Macroeconomics and Finance for useful comments and discussions.

1 Introduction

We propose a method for assessing the reliability of a given forecast model, where we characterize "reliability" as a situation where: (1) the model's in-sample performance is representative of its out-of-sample performance; and (2) the forecast performance is stable over time. This definition of reliability is motivated by two empirical findings that have been repeatedly documented by the forecast evaluation literature. The first is that, although forecasting models are typically chosen based on their in-sample fit, a model that fits well in-sample does not necessarily produce good out-of-sample forecasts (Goyal and Welch, 2004; Campbell and Thompson, 2005). The second finding is that the out-of-sample forecast performance of many popular models for macroeconomic and financial forecasting is not stable over different subsamples of a time series (Stock and Watson, 2003a, 2003b). These findings are usually based on informal analyses since none of the currently available methods for forecast evaluation appears adequate for investigating the reliability of a forecast model in a systematic and rigorous manner.

In this paper, we fill this gap by proposing formal methods for assessing the reliability of a forecast model. We do so by defining a forecast breakdown as a situation where the out-of-sample performance of a forecast model, judged by some loss function, is worse than its in-sample performance and propose a testing framework for: (1) *detecting* whether a forecast model broke down in the past, and (2) *predicting* whether the forecast model will break down at a future date.

Our notion of forecast breakdown is a formalization and generalization of what Clements and Hendry (1998, 1999) called a "forecast failure", described as a "deterioration in forecast performance relative to the anticipated outcome based on earlier fit" (Clements and Hendry, 1999, p. 1). We formalize the definition of a forecast breakdown by comparing out-of-sample performance to in-sample performance computed in one of three ways: (1) over a fixed initial sample ("fixed" scheme); (2) over a rolling window that includes only most recent observations ("rolling scheme"); and (3) over an expanding window including all observations from the beginning of the sample ("recursive scheme"). We propose a test based on the intuition that for a reliable model the in-sample and out-of-sample performances should be equal in expectation. We illustrate how to construct appropriate estimators for the asymptotic variance to be used in the test statistic, that explicitly take into account the effect of estimation uncertainty in the model's parameters. Our tests are valid under general assumptions. In particular, we allow the data to be heterogeneous and impose only weak restrictions on the loss function used for evaluation and on the type of estimators used in constructing the forecasts. We show, however, that in the common case where the same loss function is used for estimation and evaluation (e.g., OLS and quadratic loss), estimation uncertainty is asymptotically irrelevant and the test is much simpler to compute. A further contribution is an effort to understand the causes of forecast breakdowns. In particular, we show that our test can capture forecast breakdowns caused by instability in the model's parameters as well as by any other changes in the data-generating process that result in instability of forecast performance (e.g., for a quadratic loss, changes in the variance of the disturbances). We also investigate the role of overfitting - defined as the difference between out-of-sample and in-sample performance induced in finite samples by the fact that parameter estimates minimize the in-sample loss by construction - and propose a simple correction to the test statistic that eliminates its effects.

The two closest literatures to the present papers are the literature on forecast optimality testing (Mincer and Zarnowitz, 1969, Patton and Timmermann, 2003) and the literature on structural break testing (e.g., Andrews, 1993, Andrews and Ploberger, 1994, Dufour, Ghysels and Hall, 1994, Chu, Hornik and Kuan, 1995, Bai and Perron, 1998, Bai, 1999, Hansen, 2000, Elliott and Muller, 2003, Rossi, 2005). Regarding the former, we point out that the same theory developed here applies to forecast optimality testing (for example, a forecast unbiasedness test can be viewed as a a special case of a forecast breakdown test that considers the first moment properties of the forecast errors). Regarding the latter, although the focus of the forecast breakdown test is different from that of structural break tests (stability of forecast performance vs. stability of model's parameters) the two are related since instability in model's parameters is a chief cause of forecast breakdowns. In the paper, we shed some light on the properties of our forecast breakdown test relative to those of structural break tests both analytically and in Monte Carlo simulations. Our main findings can be summarized as follows: (1) the forecast breakdown test is robust to the presence of unstable regressors, whereas structural break tests cannot distinguish between instability in model's parameters and instability in the distribution of the regressors (see also Hansen, 2000); (2) the magnitude of the parameter instabilities that cause forecast breakdowns depend on whether the loss functions used for estimation and evaluation are equal or different; when the losses are equal, parameter instabilities must be of greater magnitude than those captured by previous structural break tests in order to cause a forecast breakdown; (3) structural break tests have greater power when instabilities are permanent, whereas the forecast breakdown test has greater power when there are recurring instabilities (e.g., a Markov-switching data-generating processes) and the forecast model fails to capture them (see also Carrasco, 2002). A further difference with structural break tests is that they only focus on the past stability and provide no information on the likelihood of future breaks. Furthermore, Elliott (2005) analytically shows, in the context of a single break, that attempts to forecast based on estimates of the break point are unlikely to improve forecasts. Instead, an innovation of our approach that has useful practical implications is the possibility of making predictions about the likelihood that a forecast model will break down at a future date.

To illustrate the methods proposed in this paper, we investigate whether there is evidence of a forecast breakdown in the Phillips curve's forecast of inflation in the United States. The Phillips curve relationship has been a useful guide for monetary policy in the U.S. in the past few decades. However, its in-sample stability has been challenged by Staiger, Stock and Watson (1997), and its forecasting ability further investigated by Stock and Watson (1999) and Fisher, Liu and Zhou (2002). By using both real-time and revised data, we find striking empirical evidence in favor of a breakdown in the Phillips curve forecasts. We further investigate whether monetary policy parameters would have been useful to predict the forecast breakdown and find that changes in the monetary policy behavior of the Fed and inflation variability played a key role.

2 Theory

2.1 Description of the environment

Let $W \equiv \{W_t : \Omega \longrightarrow \mathbb{R}^{s+1}, s \in \mathbb{N}, t = 1, \ldots, T\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and partition the observed vector W_t as $W_t \equiv (Y_t, X_t')'$, where $Y_t : \Omega \to \mathbb{R}$ is the variable of interest and $X_t : \Omega \to \mathbb{R}^s$ is a vector of predictors. We are interested in evaluating $\tau-$step-ahead forecasts of the variable $Y_{t+\tau}$.

The forecasts are produced using an out-of-sample procedure, where the total sample of size T is divided into in-sample and out-of-sample windows. Let m be the size of the first in-sample window and $n = T - m - \tau + 1$ be the size of the out-of-sample window. Which data constitute the in-sample window depends on the forecasting scheme used. We allow for three forecasting schemes: (1) a fixed forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, m$; (2) a rolling forecasting scheme, where the in-sample window at time t contains observations indexed $t-m+1, \ldots, t$, so that as t increases older observations are discarded; and (3) a recursive forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, t$, so that all observations from the beginning of the sample are used.

We let $f_t(\widehat{\beta}_t)$ be the time-t forecast produced by estimating a model over the in-sample window at time t, with $\widehat{\beta}_t$ indicating the $k \times 1$ parameter estimate. Each time$-t$ forecast corresponds to a sequence of in-sample fitted values $\hat{y}_j(\widehat{\beta}_t)$, with j varying over the in-sample window.

The forecasts are evaluated by a loss $L(\cdot)$, with each out-of-sample loss $L_{t+\tau}(\widehat{\beta}_t) \equiv L(Y_{t+\tau}, f_t(\widehat{\beta}_t))$ corresponding to in-sample losses $L_j(\widehat{\beta}_t) \equiv L(Y_j, \hat{y}_j(\widehat{\beta}_t))$. For example, for the linear model $Y_t = X_t\beta + \varepsilon_t$ estimated by OLS, the parameter estimate is $\widehat{\beta}_t = (\sum_{s=1}^m X_s X_s')^{-1} \sum_{s=1}^m X_s Y_s$ for the fixed scheme; $\widehat{\beta}_t = \left(\sum_{s=t-m+1}^t X_s X_s'\right)^{-1} \sum_{s=t-m+1}^t X_s Y_s$ for the rolling scheme and $\widehat{\beta}_t = \left(\sum_{s=1}^t X_s X_s'\right)^{-1}$ $\sum_{s=1}^t X_s Y_s$ for the recursive scheme. The out-of-sample loss corresponding to the forecast at time t is $L_{t+\tau}(\widehat{\beta}_t) \equiv L(Y_{t+\tau}, X_{t+\tau}'\widehat{\beta}_t)$ and the corresponding in-sample losses are $L_j(\widehat{\beta}_t) \equiv L(Y_j, X_j'\widehat{\beta}_t)$, where $j = 1, \ldots, m$ for the fixed scheme; $j = t - m + 1, \ldots, t$ for the rolling scheme and $j = 1, \ldots, t$ for the recursive scheme.

2.2 Assumptions

1. $\{W_t\}$ is mixing with α of size $-r/(r-2)$, $r > 2$; 2. $L_t(\beta)$ is measurable and twice continuously differentiable with respect to β; 3. Under H_0, $\widehat{\beta}_t - \beta^* = B_t^* H_t^* + o_p(1)$, where $\widehat{\beta}_t$ is $k \times 1$, B_t^* is a (non-stochastic) $k \times q$ matrix of rank k, such that $\max_t B_t^* < \infty$; $H_t^* = m^{-1} \sum_{s=1}^{m} h_s(\beta^*)$ (fixed scheme), $H_t^* = m^{-1} \sum_{s=t-m+1}^{t} h_s(\beta^*)$ (rolling scheme), $H_t^* = t^{-1} \sum_{s=1}^{t} h_s(\beta^*)$ (recursive scheme) for a $q \times 1$ orthogonality condition $h_s(\beta^*)$ such that $E\left(h_s(\beta^*)\right) = 0$; 4. $\sup_{t \geq 1} E||[L_t(\beta^*), \frac{\partial}{\partial \beta} L_t(\beta^*), h_t'(\beta^*)]'||^{2r} < \infty$, where $\frac{\partial L}{\partial \beta} L_t(\beta^*)$ is $1 \times k$; 5. $T^{-1} \sum_{t=1}^{T} E\left(\frac{\partial}{\partial \beta} L_t(\beta^*)\right) < \infty$ for all T; 6. $var\left(T^{-1/2} \sum_{t=1}^{T} L_t(\beta^*)\right) > 0$ for all T sufficiently large; 7. $m, n \to \infty$, $\frac{n}{m} \to \pi$, $0 < \pi < \infty$.

Comments: 1. Assumption 1 restricts the memory in the data (ruling out, e.g., unit root processes) but allows the data to be heterogeneous, for example permitting the marginal distribution of the regressors to change over time. This is a more general assumption than the assumption of stationarity made in the majority of the structural break testing literature.

2. The assumption of differentiability of the loss function is adopted for convenience and can be relaxed along the lines of McCracken (2000).

3. Assumption 3 states that under the null hypothesis the pseudo-true values of the parameters are constant. The assumption is similar to Assumption 2 of West (1996), and it allows for a number of estimating procedures for the model's parameters, including OLS, (quasi-) maximum likelihood and GMM. For example, for OLS estimation of the parameters in the linear model $Y_s = X_s' \beta^* + \varepsilon_s$, $s = 1, \ldots, t$, we have $B_t^* = \left(E\left(t^{-1} \sum_{s=1}^{t} X_s X_s'\right)\right)^{-1}$ and $h_s(\beta^*) = X_s \varepsilon_s$. For maximum likelihood estimation, B_t^* is the expectation of the inverse of the Hessian evaluated at β^* and H_t^* is the score.

4. Assumption 5 can be viewed as restricting the heterogeneity of the means of the loss derivatives. The condition is trivially satisfied when the loss used for estimation is the same as the loss used for evaluation, in which case $E\left(\frac{\partial}{\partial \beta} L_t(\beta^*)\right) = 0$ for all t.

5. Assumption 6 highlights the fact that the asymptotic distribution is obtained under the assumption that the in-sample and out-of-sample sizes go to infinity at the same rate. This assumption is necessary in order to obtain a non-degenerate asymptotic distribution.

2.3 Forecast breakdown test

We define a forecast breakdown as a deterioration in the out-of-sample performance of the forecast model relative to its in-sample performance. We formalize this idea by defining a "surprise loss" at time $t + \tau$ as the difference between the out-of-sample loss at time $t + \tau$ and the average in-sample loss:

$$SL_{t+\tau}(\widehat{\beta}_t) = L_{t+\tau}(\widehat{\beta}_t) - \bar{L}_t(\widehat{\beta}_t) \text{ for } t = m, \ldots, T - \tau, \tag{1}$$

where $\bar{L}_t(\widehat{\beta}_t)$ is the average loss over the in-sample window implied by the forecasting scheme. We then consider the out-of-sample mean of the surprise losses

$$\overline{SL}_{m,n} \equiv n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t),$$

and propose a test based on the idea that, if a forecast is reliable, this mean should be close to zero. Specifically, we test

$$H_0 : E\left(n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*)\right) = 0 \text{ for all } m, n. \tag{2}$$

One can easily see that H_0 can be interpreted as saying that the forecast performance, measured by the expected loss calculated at the stable pseudo-true parameters, is stable over time.

Our test statistic is

$$t_{m,n,\tau} = \sqrt{n}\overline{SL}_{m,n}/\hat{\sigma}_{m,n}, \tag{3}$$

where the expression for the asymptotic variance estimator $\hat{\sigma}^2_{m,n}$ depends on the forecasting scheme and on the assumptions. Theorem 2 below gives an estimator that is valid under general assumptions. Corollaries 3 and 4 give estimators that are easier to compute under more restrictive conditions.

A level α test rejects the null hypothesis whenever $t_{m,n,\tau} > z_\alpha$, where z_α is the $(1-\alpha)-th$ quantile of a standard normal distribution. The test is one-sided to reflect the fact that decreases in loss may be desirable and thus do not constitute a forecast breakdown.

The following algorithm shows the steps involved in constructing the general asymptotic variance estimator. The basic intuition is to acknowledge that the average surprise loss is a weighted average of in-sample and out-of-sample losses, with weights depending on m, n and on the forecasting scheme. When estimation uncertainty is asymptotically irrelevant, $\hat{\sigma}^2_{m,n}$ is simply a (rescaled) heteroskedasticity- and autocorrelation-robust (HAC) estimator of the variance of the weighted average of the full-sample losses. When estimation uncertainty matters, $\hat{\sigma}^2_{m,n}$ contains additional terms, which are also given in the algorithm.

Algorithm 1 (General asymptotic variance estimator) *Construct the following: (1) the* $1 \times T$ *vector of in-sample and out-of-sample losses, with element* L_t :

$$L \equiv [\underbrace{L_1(\widehat{\beta}_m), \ldots, L_m(\widehat{\beta}_m)}_{m}, \underbrace{L_{m+1}(\widehat{\beta}_{m+1}), \ldots, L_{m+\tau-1}(\widehat{\beta}_{m+\tau-1})}_{\tau-1}, \underbrace{L_{m+\tau}(\widehat{\beta}_m), \ldots, L_T(\widehat{\beta}_{T-\tau})}_{n}]$$

and the corresponding vector $\widetilde{L}$ *of demeaned losses, where* $\widetilde{L}_t \equiv L_t - T^{-1}\sum_{j=1}^{T} L_j$;[1] *(2) the* $q \times T$

[1] The first m terms of L are in-sample losses from the first estimation window and the last n terms are out-of-sample losses. For the fixed scheme $L \equiv [\underbrace{L_1(\widehat{\beta}_m), ..., L_m(\widehat{\beta}_m)}_{m}, \underbrace{0, ..., 0}_{\tau-1}, \underbrace{L_{m+\tau}(\widehat{\beta}_m), ..., L_T(\widehat{\beta}_m)}_{n}]$;. For the rolling and recursive schemes, each of the middle $\tau - 1$ terms is an in-sample loss from the estimation sample ending at the corresponding date.

matrix of orthogonality conditions, with element h_t :

$$h \equiv [\underbrace{h_1(\widehat{\beta}_m), \ldots, h_m(\widehat{\beta}_m)}_{m}, \underbrace{h_{m+1}(\widehat{\beta}_{m+1}), \ldots, h_{T-\tau}(\widehat{\beta}_{T-\tau})}_{n-1}, \underbrace{0, \ldots, 0}_{\tau}].^2$$

Let $D_{t+\tau} \equiv \frac{\partial L_{t+\tau}}{\partial \beta}(\widehat{\beta}_t) - \frac{\partial \bar{L}_t}{\partial \beta}(\widehat{\beta}_t)$, $t = m, \ldots, T-\tau$ *indicate the sequence of* $1 \times k$ *derivatives of the surprise losses, and let* B_t *be a consistent estimate of* B_t^* *from assumption (iii) that substitutes* $\widehat{\beta}_t$ *for* β^*.[3] *Construct the following weights, depending on the forecasting scheme:*

Fixed: $\underset{1\times T}{w^L} = [\underbrace{-\frac{n}{m}, \ldots, -\frac{n}{m}}_{m}, \underbrace{0, \ldots, 0}_{\tau-1}, \underbrace{1, 1, \ldots, 1}_{n}]$; $\underset{1\times qT}{w^h} = [\underbrace{\frac{B_m \sum_{t=m}^{T-\tau} D_{t+\tau}}{m}, \ldots, \frac{B_m \sum_{t=m}^{T-\tau} D_{t+\tau}}{m}}_{m}, \underbrace{0, \ldots, 0}_{T-m}]$.

Rolling $(n < m)$: $\underset{1\times T}{w^L} = [\underbrace{-\frac{1}{m}, \ldots, -\frac{n}{m}}_{n}, \underbrace{-\frac{n}{m}, .., -\frac{n}{m}}_{m-n}, \underbrace{-\frac{n-1}{m}, \ldots, -\frac{n-\tau+1}{m}}_{\tau-1}, \underbrace{1 - \frac{n-\tau}{m}, \ldots, 1 - \frac{1}{m}}_{n-\tau}, \underbrace{1, \ldots, 1}_{\tau}]$;

$$\begin{aligned}
\underset{1\times qT}{w^h} = & [\underbrace{\frac{D_{m+\tau}B_m}{m}, \frac{D_{m+\tau}B_m + D_{m+\tau+1}B_{m+1}}{m}, \ldots, \frac{\sum_{t=m}^{T-\tau} D_{t+\tau}B_t}{m}}_{n}, \\
& \underbrace{\frac{\sum_{t=m}^{T-\tau} D_{t+\tau}B_t}{m}, \ldots, \frac{\sum_{t=m}^{T-\tau} D_{t+\tau}B_t}{m}}_{m-n}, \underbrace{\frac{\sum_{t=m+1}^{T-\tau} D_{t+\tau}B_t}{m}, \ldots, \frac{D_T B_{T-\tau}}{m}}_{n-1}, \underbrace{0, \ldots, 0}_{\tau}].
\end{aligned}$$

Rolling $(n \geq m)$: $\underset{1\times T}{w^L} = [\underbrace{-\frac{1}{m}, \ldots, -\frac{m}{m}}_{m}, \underbrace{-\frac{m}{m}, \ldots, -\frac{m}{m}}_{\tau-1}, \underbrace{0, \ldots, 0}_{n-m-\tau+1}, \underbrace{1 - \frac{m-1}{m}, \ldots, 1 - \frac{1}{m}}_{m-1}, \underbrace{1, \ldots, 1}_{\tau}]$;

$$\begin{aligned}
\underset{1\times qT}{w^h} = & [\underbrace{\frac{D_{m+\tau}B_m}{m}, \frac{D_{m+\tau}B_m + D_{m+\tau+1}B_{m+1}}{m}, \ldots, \frac{\sum_{t=m}^{2m-1} D_{t+\tau}B_t}{m}}_{m}, \\
& \underbrace{\frac{\sum_{t=m+1}^{2m} D_{t+\tau}B_t}{m}, \frac{\sum_{t=m+2}^{2m+1} D_{t+\tau}B_t}{m}, \ldots, \frac{\sum_{t=n}^{T-\tau} D_{t+\tau}B_t}{m}}_{n-m}, \underbrace{\frac{\sum_{t=n+1}^{T-\tau} D_{t+\tau}B_t}{m}, \ldots, \frac{D_T B_{T-\tau}}{m}}_{m-1}, \\
& \underbrace{0, \ldots, 0}_{\tau}].
\end{aligned}$$

---

[2] The first m terms of h are orthogonality conditions from the first estimation window. For the fixed scheme $h = [\underbrace{h_1(\widehat{\beta}_m), ..., h_m(\widehat{\beta}_m)}_{m}, \underbrace{0, ..., 0}_{T-m}]$. For the rolling and recursive schemes, each of the middle $n-1$ terms is the orthogonality condition from the estimation sample ending at the corresponding date.

[3] For example, for OLS estimation of $Y_s = X_s'\beta^* + \varepsilon_s$, $s = 1, ..., t$, $B_t = (t^{-1}\sum_{s=1}^{t} X_s X_s')^{-1}$. For maximum likelihood estimation, B_t is the inverse of the Hessian evaluated at the parameter estimate.

$$\textit{Recursive:}\qquad \underset{1\times T}{w^L} = [\underbrace{-a_{m,0},\ldots,-a_{m,0}}_{m},\underbrace{-a_{m,1},\ldots,-a_{m,\tau-1}}_{\tau-1},\underbrace{1-a_{m,\tau},1-a_{m,\tau+1},\ldots,1-a_{m,n-1}}_{n-\tau},\underbrace{1,\ldots,1}_{\tau}];$$

$$\underset{1\times qT}{w^h} = [\underbrace{b_{m,0},\ldots,b_{m,0}}_{m},\underbrace{b_{m,1},\ldots,b_{m,n-1}}_{n-1},\underbrace{0,\ldots,0}_{\tau}],\ \text{where}$$

$$\begin{aligned} a_{m,j} &= \left(\frac{1}{m+j}+\frac{1}{m+j+1}+\ldots+\frac{1}{T-\tau}\right); \\ b_{m,j} &= \left(\frac{D_{m+\tau+j}B_{m+j}}{m+j}+\frac{D_{m+\tau+j+1}B_{m+j+1}}{m+j+1}+\ldots+\frac{D_T B_{T-\tau}}{T-\tau}\right). \end{aligned} \tag{4}$$

Let

$$V_T = \begin{pmatrix} V_T^{LL} & V_T^{Lh} \\ V_T^{Lh} & V_T^{hh} \end{pmatrix},\ \text{where} \tag{5}$$

$$V_T^{LL} \equiv T^{-1}\sum_{t=1}^{T}(w_t^L\widetilde{L}_t)^2 + 2T^{-1}\sum_{j=1}^{p_T}\upsilon_{T,,j}\sum_{t=j}^{T}w_t^L\widetilde{L}_t w_{t-j}^L\widetilde{L}_{t-j}; \tag{6}$$

$$V_T^{hh} \equiv T^{-1}\sum_{t=1}^{T}w_t^h h_t h_t' w_t^{h\prime} + T^{-1}\sum_{j=1}^{p_T}\upsilon_{T,,j}\sum_{t=j}^{T}\left(w_t^h h_t h_{t-j}' w_{t-j}^{h\prime} + w_{t-j}^h h_{t-j} h_{t-j}' w_t^{h\prime}\right); \tag{7}$$

$$V_T^{Lh} \equiv T^{-1}\sum_{t=1}^{T}w_t^L\widetilde{L}_t h_t' w_t^{h\prime} + T^{-1}\sum_{j=1}^{p_T}\upsilon_{T,,j}\sum_{t=j}^{T}\left(w_t^L\widetilde{L}_t h_{t-j}' w_{t-j}^{h\prime} + w_{t-j}^L\widetilde{L}_{t-j} h_{t-j}' w_t^{h\prime}\right), \tag{8}$$

with $\{p_T\}$ a sequence of integers such that $p_T \to \infty$ as $T \to \infty$, $p_T = o(T)$ and $\{\upsilon_{T,j} : T = 1,2,\ldots; j = 1,\ldots,p_T\}$ a triangular array such that $|\upsilon_{T,j}| < \infty$, $T = 1,2,\ldots;$ $j = 1,\ldots,p_T$ and $\upsilon_{T,j} \to 1$ as $T \to \infty$ for each $j = 1,\ldots,p_T$ (cf. Andrews, 1991 or Newey and West, 1987).

Theorem 2 (Forecast Breakdown Test) *(a) If $E\left(\frac{\partial L_{t+\tau}}{\partial\beta}(\beta^*)\right)$ is constant for all t, $\hat{\sigma}_{m,n} = \sqrt{(T/n)V_T^{LL}}$, V_T^{LL} given in (6). Then, $t_{m,n,\tau} \xrightarrow{d} N(0,1)$ under H_0 in (2).* [4]

(b) If V_T in (5) is p.d., $\hat{\sigma}_{m,n} = \sqrt{(T/n)\,(V_T^{LL}+V_T^{hh}+2V_T^{Lh})}$, V_T^{LL}, V_T^{hh} and V_T^{Lh} given in (6)-(8). Then, $t_{m,n,\tau} \xrightarrow{d} N(0,1)$ under H_0 in (2).

Comments: Theorem 2-(a) shows that if the loss scores have constant mean under the null hypothesis, then estimation uncertainty is asymptotically irrelevant and the asymptotic variance estimator is easier to compute. Theorem 2-(b) gives the correction to the asymptotic variance estimator needed when estimation uncertainty does not vanish asymptotically. Whether the condition for asymptotic irrelevance in Theorem 2-(a) is satisfied depends in general on the model, the loss function and the estimation procedure, and its plausibility must thus be verified on a case-by-case basis. Corollary 3 below shows that an important case in which this condition is satisfied is when

[4] A Matlab code computing $\hat{\sigma}_{m,n}$ in the case of asymptotically irrelevant estimation uncertainty can be downloaded from http:\\www.econ.ucla.edu\giacomin.

the loss function used for estimation is the same as that used for evaluation. This is a common situation in forecasting applications, since parameters are typically estimated by OLS and forecasts are evaluated using a quadratic loss. Corollary 4 shows that the asymptotic variance estimator can be simplified even further if the losses are covariance-stationary under the null hypothesis.[5]

2. The use of a HAC estimator for the asymptotic variance is motivated by the possible presence of serial correlation in the sequence of forecast losses. This is easy to see for a quadratic loss, in which case serial correlation in the losses is induced by the presence of GARCH in the data.

3. The same theory outlined in Theorem 2 can be applied to forecast unbiasedness tests, by redefining the loss as simply being the forecast error. If the estimation procedure is such that the in-sample errors have zero mean (e.g., when using OLS), the surprise losses (1) coincide with the out-of-sample forecast errors, and the forecast breakdown test then becomes a test of zero mean of the out-of-sample forecast errors, that is, a forecast unbiasedness test.

Corollary 3 (Asymptotic variance estimator under equal loss) *If* $\widehat{\beta}_t = \arg\min_\beta \bar{L}_t(\beta)$, *then* $\hat{\sigma}_{m,n} = \sqrt{(T/n)\, V_T^{LL}}$, V_T^{LL} *given in (6).*

Corollary 4 (Asymptotic variance estimator under equal loss and covariance-stationarity) *Given the assumptions of Theorem 2-(a), further assume that* $\Gamma_j \equiv cov\left(L_t(\beta^*), L_{t-j}(\beta^*)\right)$ *depends on* j *but not on* t *under* H_0. *Then,* $\hat{\sigma}_{m,n} = \sqrt{\lambda S_n^{LL}}$, *where*

Forecasting scheme	λ
Fixed	$1 + n/m$
Rolling, $n < m$	$1 - (1/3)\,(n/m)^2$
Rolling, $n \geq m$	$(2/3)(m/n)$
Recursive	1

(9)

and $S_n^{LL} = n^{-1}\sum_{t=m}^{T-\tau}\widetilde{L}_{t+\tau}^2 + 2n^{-1}\sum_{j=1}^{p_n} v_{n,,j}\sum_{t=m+j}^{T-\tau}\widetilde{L}_{t+\tau}\widetilde{L}_{t+\tau-j}$, *where* $\widetilde{L}_{t+\tau} \equiv L_{t+\tau}(\widehat{\beta}_t) - n^{-1}\sum_{j=m}^{T-\tau} L_{j+\tau}(\widehat{\beta}_j)$ *and with* $\{p_n\}$ *a sequence of integers such that* $p_n \to \infty$ *as* $n \to \infty$, $p_n = o(n)$ *and* $\{v_{n,j} : n = 1, 2, \ldots; j = 1, \ldots, p_n\}$ *a triangular array such that* $|v_{n,j}| < \infty$, $n = 1, 2, \ldots;\ j = 1, \ldots, p_n$ *and* $v_{n,j} \to 1$ *as* $n \to \infty$ *for each* $j = 1, \ldots, p_n$ *(cf. Andrews, 1991 or Newey and West, 1987).*

3 Causes of forecast breakdowns

This section analyzes three possible causes of forecast breakdown: parameter instabilites, other types of instabilities affecting the surprise loss, and overfitting. Define β_t^* as $E\left(\frac{\partial \mathcal{L}_t(\beta_t^*)}{\partial \beta}\right) = 0$, $t = 1, 2, \ldots, T$, where $\mathcal{L}(\beta)$ is the loss used for estimation. Let $\overline{\Sigma}_j$ denote the relevant sample averages

[5] In the case of quadratic loss, this rules out time-variation in the tail fatness of the forecast errors.

depending on the forecasting scheme: $\overline{\Sigma}_j = \frac{1}{t}\sum_{j=1}^{t}$ for the recursive scheme, $\overline{\Sigma}_j = \frac{1}{m}\sum_{j=t-m+1}^{t}$ for the rolling scheme, and $\frac{1}{m}\sum_{j=1}^{m}$ for the fixed scheme. Similar notation holds for the loss used for evaluation, $L_t(\beta)$. A mean value expansion of $\overline{SL}_{m,n}$ around the $\beta_t^{*\prime}s$ gives the following expression:

Proposition 5

$$
\begin{aligned}
& E\left(n^{-1}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t)\right) \\
= \; & \underbrace{E\left(n^{-1}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta_t^*)\right)}_{\text{“other instabilities”}} + \underbrace{\frac{1}{n}\sum_{t=m}^{T-\tau} E\left(\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_{t+\tau}^*\right)\right)\left(\beta_t^*-\beta_{t+\tau}^*\right)}_{\text{“parameter instabilities I”}} \\
& \underbrace{-\frac{1}{n}\sum_{t=m}^{T-\tau}\overline{\sum}_j E\left(\frac{\partial}{\partial\beta}L_j\left(\beta_j^*\right)\right)\left(\beta_t^*-\beta_j^*\right)}_{\text{“parameter instabilities I”}} \qquad (10) \\
& \underbrace{+\frac{1}{2n}\sum_{t=m}^{T-\tau}\left[\left(\beta_t^*-\beta_{t+\tau}^*\right)' E\left(\frac{\partial^2}{\partial\beta^2}L_{t+\tau}\left(\overline{\beta_{t+\tau}^*}\right)\right)\left(\beta_t^*-\beta_{t+\tau}^*\right)\right.}_{\text{“parameter instabilities II”}} \\
& \underbrace{\left.-\overline{\sum}_j\left(\beta_t^*-\beta_j^*\right)' E\left(\frac{\partial^2}{\partial\beta^2}L_j\left(\overline{\beta_j^*}\right)\right)\left(\beta_t^*-\beta_j^*\right)\right]}_{\text{“parameter instabilities II”}} + \underbrace{\frac{1}{n}\sum_{t=m}^{T-\tau} E\left[\left(\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_t^*\right)\right)\left(\widehat{\beta}_t-\beta_t^*\right)\right]}_{\text{“overfit I”}} \\
& \underbrace{+\frac{1}{n}\sum_{t=m}^{T-\tau} E\left\{\left[\left(\widehat{\beta}_t-\beta_t^*\right)'\frac{\partial^2}{\partial\beta^2}\overline{\mathcal{L}}_t(\overline{\widehat{\beta}}_t)-\frac{\partial}{\partial\beta}\overline{L}_t\left(\beta_t^*\right)+\frac{\partial}{\partial\beta}\overline{\mathcal{L}}_t\left(\beta_t^*\right)\right]\left(\widehat{\beta}_t-\beta_t^*\right)\right\}}_{\text{“overfit II”}} \\
& \underbrace{+\frac{1}{2n}\sum_{t=m}^{T-\tau} E\left[\left(\widehat{\beta}_t-\beta_t^*\right)'\left(\frac{\partial^2}{\partial\beta^2}L_{t+\tau}(\overline{\widehat{\beta}}_t)-\frac{\partial^2}{\partial\beta^2}\overline{L}_t(\overline{\widehat{\beta}}_t)\right)\left(\widehat{\beta}_t-\beta_t^*\right)\right]}_{\text{“overfit III”}}
\end{aligned}
$$

The expression (10) reveals that our test's ability to capture parameter instabilities depends on whether the loss functions used for estimation and for evaluation differ or are equal. If the losses are equal, the component "parameter instabilities I" disappears, which implies that forecast breakdowns are not caused by "small breaks" (i.e., those of the type $\beta_t^* - \beta^* = O_p\left(T^{1/2}\right)$), but only by "larger breaks" (i.e., those of the type $\beta_t^* - \beta^* = O_p\left(T^{1/4}\right)$). Note that the former are the breaks considered in the in-sample structural break testing literature. Further, when the estimation and evaluation losses are equal, the "overfit II" component reduces to a quadratic form, and is thus always positive. Intuitively, this is due to the fact that in this case the average in-sample loss computed at the parameter estimates is minimized by construction, which introduces a positive difference between the out-of-sample loss and the average in-sample loss in finite samples.

The following proposition characterizes the causes of forecast breakdowns in the context of a linear regression model and for a quadratic loss.

Proposition 6 (Special case: linear model and quadratic loss) *Consider a quadratic loss : $L(e) = \mathcal{L}(e) = e^2$, a fixed forecasting scheme, and a linear and correctly specified model:*

$$Y_t = X_t' \beta_t + \varepsilon_t, \ \varepsilon_t \sim i.i.d. \left(0, \sigma_t^2\right),$$

where the $k \times 1$ vector X_t is i.i.d. Let $E(X_t X_t') \equiv J$. Suppose there are two breaks: a permanent break in the parameters and a permanent break in the variance of the disturbances, so that $\beta_t = \beta + g_1 \cdot 1(t \geq m) T^{-1/4}$ and $\sigma_t^2 = \sigma^2 + g_2 1(t \geq m) / \sqrt{T}$. We have:

$$E\left(\overline{SL}_{m,n}\right) = \underbrace{g_2/\sqrt{T}}_{\text{"other instabilities"}} + \underbrace{\frac{1}{2} g_1' J g_1 / \sqrt{T}}_{\text{"parameter instabilities II"}} \underbrace{+2\sigma^2 \frac{k}{m}}_{\text{"overfit II"}}. \quad (11)$$

Therefore, a forecast breakdown can be caused by a "small" positive break in the variance of the disturbances and/or a "large" break (positive or negative) in the conditional mean parameters. Overfitting is present only in finite samples and depends on the ratio of the number of parameters to the sample size.

In the following example, we analyze the "parameter instabilities I" component in (10) for the different forecasting schemes, in order to gain some insight into the ability of each forecasting scheme to detect forecast breakdowns due to parameter instability, .

Example 7 (Comparison of different forecasting schemes) *Suppose that $E\left[\frac{\partial}{\partial \beta} L_j\left(\beta_j^*\right)\right] = D$ is constant over time, so that the "parameter instabilities I" component is*

$$D \frac{1}{n} \sum_{t=m}^{T-\tau} \left(\beta_{t+\tau}^* - \overline{\sum}_j \beta_j^*\right).$$

Let $\beta_t^* - \beta^* = g(s)$, where $g(s) = 1(s < \lambda)$, $s = [t/T]$, so that $\overline{\Sigma}_j \left(\beta_j^* - \beta^*\right) \simeq \overline{g}(s)$, where $\overline{g}(s) = \int_0^s g(v)\, dv$ for the recursive scheme, $\overline{g}(s) = \int_{s-(1+\pi)^{-1}}^s g(v)\, dv$ for the rolling scheme, $\overline{g}(s) = \int_0^{(1+\pi)^{-1}} g(v)\, dv$ for the fixed scheme. For each scheme we calculate

$$\frac{1}{n} \sum_{t=m}^{T} \left(\beta_{t+\tau}^* - \overline{\sum}_j \beta_j^*\right) \simeq \sqrt{T} \int_{(1+\pi)^{-1}}^{1} \left[g(s) - \overline{g}(s)\right] ds \quad (12)$$

Let $\widetilde{m} \equiv \left[\frac{m}{T}\right] = (1+\pi)^{-1}$ *and* $\widetilde{n} \equiv \left[\frac{n}{T}\right] = \pi(1+\pi)^{-1}$. *Then (12) becomes:*

$$\begin{aligned}
\textit{Fixed scheme} &: \quad 1(\widetilde{m} \leq \lambda)\frac{(1-\lambda)}{\widetilde{n}} + 1(\widetilde{m} > \lambda)\cdot\frac{\lambda}{\widetilde{m}}; \\
\textit{Rolling scheme} &: \quad 1(\widetilde{m} \leq \lambda)\frac{1}{\widetilde{n}\widetilde{m}}\left[\sum_{s=\lambda}^{\min(\lambda+\widetilde{m},1)}[\lambda-(s-\widetilde{m})]\right] \\
&\quad +1(\widetilde{m} > \lambda)\frac{1}{\widetilde{\widetilde{n}}\widetilde{\widetilde{m}}}\sum_{s=\widetilde{m}}^{\min(\lambda+\widetilde{m},1)}[\lambda-(s-\widetilde{m})] \\
\textit{Recursive scheme} &: \quad 1(\widetilde{m} \leq \lambda)\frac{1}{\widetilde{\widetilde{n}}}[-\lambda\ln\lambda] - 1(\widetilde{m} > \lambda)\frac{1}{\widetilde{\widetilde{n}}}[\lambda\ln\widetilde{m}]
\end{aligned} \tag{13}$$

Figure 1 plots these expressions for the case where $g(s) = 1(s < \lambda)$ *for various values of* λ.[6] *The magnitude of the deviation from the null changes across the different schemes and depends upon* π. *This magnitude is maximum for breaks happening at time* $(1+\pi)^{-1}$. *When* $\pi = 1$, *more weight is given to breaks in the middle of the sample. While a fixed scheme gives symmetric weights before and after the middle of the sample, the recursive and rolling schemes give more weight to breaks occurring after the middle of the sample.*

[FIGURE 1 HERE]

4 An overfitting-corrected forecast breakdown test

We propose a simple correction to the forecast breakdown test (3) that eliminates the overfitting distortion introduced in finite samples by the use of the same loss function for estimation and evaluation. Specifically, we propose subtracting from the numerator of our test statistic an estimate of the "overfitting II" component in (10). Using similar reasonings to those in the proof of Proposition 6, we obtain an estimate of the overfitting correction in the context of a linear model with k regressors - $Y_t = X_t'\beta + \varepsilon_t$ - with quadratic loss used for estimation and evaluation. For computational ease, we also assume here that the $X_t's$ are covariance-stationary. The test statistic is modified as follows:

$$\begin{aligned}
t^c_{m,n,\tau} &= \sqrt{n}(\overline{SL}_{m,n} - c)/\hat{\sigma}_{m,n} \\
c &= 2\cdot\gamma\cdot tr\left(\frac{X'X}{T}\cdot\widehat{V}^{\beta}_T\right),
\end{aligned}$$

where: $\gamma = 1/m$ for the fixed and rolling schemes and $\gamma = \ln(1+n/m)/n$ for the recursive scheme; $X \equiv [X_1', \ldots, X_T']$; $\widehat{V}^{\beta}_T$ is a consistent estimator of the asymptotic variance of the full-sample parameter estimate $\widehat{V}^{\beta}_T = \widehat{asyvar}(\sqrt{T}\widehat{\beta}_T)$; $\hat{\sigma}_{m,n}$ is as in Theorem 2-(b) or Corollary 3.

[6] Note that D only rescales this component, but the shape will remain the same.

It is interesting to note that, for cases where the asymptotic variance of the parameter estimates can be consistently estimated by $\widehat{V}_T^{\beta} = \sigma^2(X'X/T)^{-1}$, the overfitting correction simply becomes

$$c = 2\gamma\sigma^2 k, \tag{14}$$

where $\sigma^2 = var(\varepsilon_t)$. Direct calculations further show that in this case $t^c_{m,n,\tau}$ may be equivalently obtained by re-defining the surprise losses as the difference between the out-of-sample loss and the average in-sample loss penalized using the Akaike's information criterion (AIC).[7]

5 Predicting future forecast breakdowns

The previous section proposed a test for detecting whether a forecast method broke down in the past. A question that may be of further interest to real-time forecasters is whether the forecast method will break down in the future. This is of course related to finding past breakdowns: if the surprise losses had non-zero mean in the past, we expect them to continue to be positive in the future. However, it is possible that one could find additional information available at the time of the prediction - besides a positive mean - that the forecaster could exploit in order to predict whether there will be a forecast breakdown at a specific date in the future. For example, the surprise losses may be persistent (this is easy to see for a quadratic loss function, where the presence of GARCH in the data will induce serial correlation in the surprise losses) or they may be correlated with leading indicators of the state of the economy.

The key idea is to find variables that predict the difference between in-sample and out-of-sample performance by regressing the surprise losses on a number of explanatory variables.

We propose the following method for predicting a forecast breakdown at the target date $T+\tau$.

1. Select a $r \times 1$ information vector Z_t from the information set at time t and estimate the predictive regression
$$SL_{t+\tau} = Z_t^{'}\delta + \varepsilon_{t+\tau}, \tag{15}$$
over the out-of-sample period $t = m, \ldots, T-\tau$. Let $\widehat{\delta}_n$ denote the parameter estimate that minimizes the same loss function used in constructing the surprise losses.[8] Z_t can include a constant; lagged surprise losses; economically meaningful variables such as business cycle leading indicators, measures of stock market volatility, interest rates etc.

2. Test whether δ is significantly different from zero by a Wald test
$$W_{m,n,\tau} = n\widehat{\delta}_n^{'}\hat{\Omega}_{m,n}^{-1}\widehat{\delta}_n \tag{16}$$

[7] To see this, note that (for the fixed scheme) the AIC penalizes the in-sample log-likelihood as $\log \overline{L}_m + 2k/m$, which corresponds to penalizing the in-sample loss as $\overline{L}_m(1+\exp(2k/m)) \simeq \overline{L}_m(1+2k/m)$. The claim then follows from redefining $SL_{t+\tau}$ as $L_{t+\tau} - \overline{L}_m(1+2k/m)$.

[8] We make this assumption for simplicity, since in this case estimation uncertainty is asymptotically irrelevant.

where $\hat{\Omega}_{m,n}^{1/2}$ is an estimator of the asymptotic variance of $\sqrt{n}\widehat{\delta}_n$. Reject the null hypothesis that $\delta = 0$ if $W_{m,n,\tau} > \chi^2_{r,1-\alpha}$, where $\chi^2_{r,1-\alpha}$ is the $(1-\alpha) - th$ quantile of a χ^2_r distribution.

3. In general, one can plot the fitted values from the regression (15) together with a one-sided $(1-\rho)\%$ confidence interval: $\{Z_t'\widehat{\delta}_n - z_\rho Z_t'\hat{\Omega}_{m,n}^{1/2}\}_{t=m}^{T-\tau}$, in order to analyze the behavior of the surprise losses over time. One could further use (15) to construct a $(1-\rho)\%$ prediction interval for $SL_{T+\tau}$,for example as $\left(Z_T'\widehat{\delta}_n - z_\rho\hat{\sigma}_\varepsilon, +\infty\right)$,where $\hat{\sigma}_\varepsilon$ is an estimator of the standard error of $\varepsilon_{t+\tau}$ from (15). If the prediction interval does not contain 0, conclude that the forecast method will break down at time $T+\tau$ with probability $(1-\rho)\%$.

6 Monte Carlo evidence

This section analyzes the size and power properties of our forecast breakdown test in finite samples, relative to the properties of in-sample structural break tests (Elliott and Muller, 2003) and forecast unbiasedness tests.[9]

6.1 Size properties

We investigate the size of our test, in particular with regards to its robustness to the presence of time-variation in the marginal distribution of the regressors and to the presence of conditionally heteroskedastic disturbances. To this end, we let the data-generating process (DGP) be

$$\begin{aligned} Y_t &= 2.73 - 0.44u_{t-1} + \sigma_t\varepsilon_t, \\ \sigma_t^2 &= 1 + \alpha * \varepsilon_{t-1}^2,\ \varepsilon_t \sim i.i.d.N(0,1). \end{aligned} \tag{17}$$

We consider two experiment designs. The first (MC1) has i.i.d. regressors and conditionally homoskedastic errors: $u_t \sim i.i.d.N(0,1)$ and $\alpha = 0$. The second design (MC2), inspired by our empirical application to the Phillips curve model of U.S. inflation, lets u_t be the actual time series of monthly U.S. unemployment and $\alpha = .5$.[10] The DGP parameters and conditional mean specification are from Staiger, Stock and Watson (1997). We use an actual time series in order to generate data that exhibit realistic heterogeneous behavior. Throughout, we restrict attention to the one-step-ahead forecast horizon and use a quadratic loss for both estimation and evaluation.

For each pair of in-sample and out-of-sample sizes (m,n) and for each of 5000 Monte Carlo replications, we generate $T = m+n$ data as in (17). In MC2, we use the last T data in the u_t time series, up to 2005:08. We obtain sequences of out-of-sample forecasts and of relative forecast errors

[9] Andrews' (1991) and Andrews and Ploberger's (1995) test results were qualitatively similar to those obtained by using the Elliott and Muller's (2003) test in the case of a single break, and are therefore not reported.

[10] The unemployment series is the seasonally adjusted civilian unemployment rate from FRED II.

by estimating the model $Y_t = \beta_1 + \beta_2 u_{t-1} + e_t$ by OLS using either a fixed, a rolling or a recursive forecasting scheme.

We consider the following tests: our forecast breakdown test for the three forecasting schemes, where the test statistic (3) is computed using either the general asymptotic variance estimator of Corollary 3 or the estimator of Corollary 4, valid under the assumption of covariance-stationary losses (the truncation lag for both estimators is $p_T = p_n = 0$ in MC1 and $p_T = p_n = n^{1/3}$ in MC2).We denote the two tests by $t_{m,n,\tau}$ and $t^{\text{stat}}_{m,n,\tau}$, respectively. We further consider the test proposed by Elliott and Muller (2003) (denoted EM) and a forecast unbiasedness test (denoted UNB), obtained as a t-test of zero parameter in a regression of the out-of-sample forecast errors (from the recursive scheme) on a constant.

Table 1 contains the rejection frequencies of the above tests for various (m, n) pairs.

[TABLE 1 HERE]

The table reveals that our forecast breakdown test has good size properties for large enough in-sample and out-of-sample sizes ($m, n \geq 100$). This is particularly true for the $t^{\text{stat}}_{m,n,\tau}$ test, which is generally well-sized, if conservative. Both tests (in particular $t_{m,n,\tau}$) tend to over-reject when the in-sample size is small ($m = 50$), partly due to the effects of overfitting. To verify this claim, we report in Table 2 the rejection frequencies of the overfitting-corrected test of Section 4, using the simple correction (14) in both MC1 and MC2.

[TABLE 2 HERE]

As expected, the use of the overfitting correction substantially improves the size properties of the test. The overfitting-corrected test appears well-sized in all cases except for the fixed scheme when the in-sample size is small ($m = 50$). Regarding the different forecasting schemes, the recursive scheme appears to be the most robust whereas the fixed and the rolling schemes can suffer size distortions for small sample sizes (the size distortions for the rolling scheme disappear once using the overfitting correction)..

Comparing the results from MC1 and MC2, we see that the forecast breakdown test is robust to the presence of heterogeneous regressors and of ARCH disturbances. In MC2, our test correctly concludes that the forecasting model is reliable, as there are no changes in the model's parameters or in the distribution of the disturbances. This is in stark contrast with the EM test, which has correct size when the regressor is i.i.d., but (similarly to the majority of existing structural break tests, as documented by Hansen, 2001) erroneously detects instability in model's parameters when the regressor is as heterogeneous as the time series of U.S. unemployment has been over the past 25 years (the EM rejects 100% of the time!). Finally, the forecast unbiasedness test seems to be characterized by good size properties, with a tendency to under-reject in the presence of heterogeneous regressors and ARCH disturbances.

6.2 Power properties

In this section we consider various sources of forecast breakdowns and analyze the power of the tests considered in Section 6.1. Figure 2 shows rejection frequencies over 5000 Monte Carlo replications for the various tests in the following DGPs. Unless otherwise specified, we let $T = 300$, estimate the model $Y_t = \alpha + e_t$ by OLS and use a quadratic loss for evaluation. The in-sample size m used for the forecast breakdown and the unbiasedness tests is specified in each design.

Design 1: Changes in mean. We consider either permanent or transitory changes in mean. The permanent change DGP is

$$Y_t = \beta_A \cdot 1\,(t > T/2) + \varepsilon_t,\ \ \varepsilon_t \sim i.i.d.N(0,1), \tag{18}$$

and we choose an in-sample size $m = 150$. We consider both a quadratic and a linex loss. In the transitory change DGP, we let

$$Y_t = \mu_t + \varepsilon_t,$$

where μ_t switches between 0 and β_A every 50 periods and we choose $m = 50$.

Design 2: Permanent change in variance. The DGP is

$$Y_t = \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N\left(0, 1 + \beta_A \cdot 1\,(t > T/2)\right) \tag{19}$$

We choose $m = 150$.

Design 3: Other DGP changes. Here we consider a case where the true conditional mean undergoes a permanent change but the two specifications are not nested, so that the tests developed for structural breaks are not optimal in this context. We let

$$\begin{aligned} Y_t &= \beta_A \cdot 1\,(t \leq T/2) + X_t \cdot 1\,(t > T/2) + \varepsilon_t, \\ X_t &= .6X_{t-1} + \eta_t,\ \ \varepsilon_t, \eta_t \sim i.i.d.N\,(0,1)\ \text{ independent.} \end{aligned} \tag{20}$$

We let $m = 150$.

Design 4: Overfitting. The DGP is

$$Y_t = \varepsilon_t,\ \ \varepsilon_t \sim i.i.d.N(0,1), \tag{21}$$

and the model is $Y_t = X_t'\alpha + e_t$, where X_t is a $(k \times 1)$ vector of $i.i.d.N(0,1)$ random variables. Here, as k increases, the model used by the researcher progressively contains more insignificant parameters, which may deteriorate its forecasting ability.

[FIGURE 2 HERE]

Figure 2 shows that the forecast breakdown test has power in all four designs. In the case of a permanent break in mean (Design 1, first panel), the forecast breakdown has lower power than both the EM and the UNB tests, but its power grows when the losses used for estimation and evaluation differ (Design 1, second panel). When the permanent change in DGP is as in Design 3, the power loss of the forecast breakdown relative to the EM and UNB test is substantially lower. In the case of transitory breaks in mean (Design 1, third panel), we see that the forecast breakdown test with a rolling scheme has the highest power. The forecast breakdown test for all three forecasting schemes is also the only test to have power against a permanent break in variance (Design 2) and against overfitting (Design 4).

7 Application: the Phillips curve and inflation forecast breakdowns

The Phillips curve inflation forecasting model has traditionally been a useful guide for monetary policy in the United States, and its forecasting ability is thus of practical relevance. The model relates changes in inflation to past values of the unemployment gap (the difference between the unemployment rate and the NAIRU) and past values of inflation. However, the in-sample stability of the Phillips curve relationship has been challenged by Staiger, Stock and Watson (1997), and its forecasting ability further investigated by Stock and Watson (1999) and, more recently, by Fisher et al. (2002). The latter show that the forecasting ability of the Phillips curve depends upon the period. In particular, it appears to forecast well one year ahead during the 1977-1984 period, but its predictive power breaks down in the 1993-2000 period. Thus, as an empirical application of the methods proposed in this paper, we investigate the usefulness of the Phillips curve in forecasting inflation at various horizons. We put a particular emphasis on the 12 months horizon considered by Stock and Watson (1999), and consider whether the Phillips curve experienced forecast breakdowns over that horizon over the past three decades.

Following Stock and Watson (1999), let $\pi_t^h = (1200/h) \ln (P_t/P_{t-h})$ denote the τ-period inflation in the price level P_t reported at an annual rate, π_t denote monthly inflation at an annual rate at time t ($\pi_t \equiv \pi_t^1 = (1200) \ln (P_t/P_{t-1})$), and u_t denote the unemployment rate. Then the Phillips curve can be expressed as:

$$\pi_t^\tau - \pi_t = \theta_0 + \theta_1 (L) u_t + \theta_2 (L) (\pi_t - \pi_{t-1}) + e_{t+\tau} \tag{22}$$

where θ_0 is a constant parameter that implicitly embodies a time-invariant NAIRU, and $\theta_1 (L)$ and $\theta_2 (L)$ are polynomials in the lag operator L with q_u and q_π lags, respectively.

When analyzing in retrospect whether unemployment was a useful predictor for inflation, it is important to assess its actual predictive ability at the historic point in time, that is by using only data that were actually available to the policymakers at that time. For example, Orphanides (2001) and Ghysels et al. (2002) analyze the performance of monetary policy rules in the presence of real-time data, and note their relationship with changes in the Fed Chairmen. For that reason, we use real-time monthly data for the U.S. unemployment rate from the real-time database provided by the Federal Reserve Bank of Philadelphia database. The data are discussed in Croushore and Stark (2001), and are available from January 1947 to April 2004 at quarterly vintages starting from November 1965. While the Federal Reserve Bank of Philadelphia also contains a real-time series of consumer prices, unfortunately that is available only from the 1994 vintage, and is thus not useful for our purposes of investigating the Phillips curve forecasting performance over the past three decades. We use instead the real-time monthly database for consumer prices from the Swanson, van Dijk, and Callan dataset (available at http://econweb.rutgers.edu/nswanson/realtime.htm). We focus on seasonally adjusted inflation, as in Stock and Watson (1999).[11] The data span from January 1961 (with a first vintage in February 1978) until December 2001. Due to the data limitations, we restrict estimation from January 1978 (with a first vintage equal to the first available vintage, February 1978) until December 2001 using quarterly vintages.[12]

The first four columns in Table 2 show the results. We report p-values for both the forecast breakdown test ($t_{m,n,\tau}$) of Section 2.3 as well as the $W_{m,n,\tau}$ test of Section 5, where the latter includes a constant and one lagged value of the surprise loss, SL_{t-1}. $m = 60$, $n = 95$, and the forecast horizons are $\tau = 3$ and 12 months (i.e. 1 and 4 quarters). q_u and q_π are, respectively, the number of lags used for unemployment and inflation; the row labeled "*BIC*" reports instead results for the case in which the lag length is determined at every time t by the Bayesian Information Criterion (since there are multiple regressors, for computational reasons we followed the common practice of requiring that all regressors have the same number of lags). The one-step ahead forecasts begin in 1993:1 (the date has been chosen to correspond to the change in monetary policy identified in Fisher et al., 2002). As the table shows, all the p-values are close to zero at a one year ahead horizon. In some cases, the forecast breakdown test does not reject the null of no forecast breakdown at the three month horizon. Thus, the empirical results reported in Table 3 show striking empirical evidence in favor of a breakdown in the predictive ability of the Phillips curve to predict inflation at the 12 months horizon in the last decade.

[11] Note that Stock and Watson (1999) did not examine real-time data but, on the other hand, investigate many other predictors that could help forecasting inflation beyond unemployment, not only unemployment, as we do.

[12] The sample used in Fisher et al. (2002) begins in January 1977 and that used in Stock and Watson (1999) begins in January 1959. Note that while in the real-time database unemployment is revised at a quarterly frequency, data are still available at a monthly frequency. However, there will be many missing data if one tried to extend the quarterly data to a monthly frequency. For this reason, we calculated the annualized inflation rate at a monthly frequency, then used observations only for February, May, August and November, which correspond to the available vintage quarters.

[TABLE 3 HERE]

Because of small sample concerns associated with the data available for the real-time forecasting exercise, we also examine the forecasting performance of the Phillips curve by using finally revised monthly data. That is, we consider the most recent observations collected by the Philadelphia Fed (August 2004) for both seasonally unadjusted CPI and unemployment. The largest available sample for both variables is from January 1948 until June 2004. The last two columns in Table 2 show the results. The first estimated regression is from 1959:1 to 1985:1 (similar results hold if the first estimated regression is from 1959:1 to 1993:1). The forecast breakdown test finds evidence of a forecast breakdown at the one month horizon.

Given that our test detected a forecast breakdown, we next proceed to empirically investigate its possible economic causes. Fisher et al. (2002) argue that periods of low inflation volatility and periods after regime shifts in monetary policy appear to be associated with the break in the forecasting ability of the standard Phillips curve. Thus, we next analyze whether the forecast breakdown of the Phillips curve could have been detected or predicted by using inflation volatility and a measure of changes in the monetary policy behavior of the Fed. Inflation volatility is simply estimated as the sample variance of inflation (using a rolling window estimator). As a measure of changes in the monetary policy behavior of the Fed, we use rolling estimates of the coefficients of the reaction of the Fed Fund Rate (FFR) on the deviation of inflation from its target level and the unemployment gap in the New-Keynesian forward looking monetary policy reaction function analyzed by Clarida et al. (2000). We recursively estimate by GMM the following moment condition (which describes the Fed's monetary policy reaction function):[13]

$$E\left(r_t-(1-\rho)\left[rr^*-(\beta-1)\,\pi^*+\beta\pi_{t,k}+\gamma x_{t,q}\right]+\rho\left(L\right)r_{t-1}|\Im_t\right)=0 \tag{23}$$

where r_t is the nominal FFR; $\pi_{t,k}$ is the percentage change in the price level between periods t and $t+k$ (expressed in annual rates); $x_{t,q}$ is a measure of the average output gap between period t and $t+q$, with the output gap defined as minus the percentage deviation between actual unemployment rate and the corresponding target (in practice, the target is defined as a fitted quadratic function of time); and $\Im_t$ is the information set at time t. As in Clarida et al. (2000), $\rho\left(L\right)\equiv\rho_1+\rho_2 L$ contains two lags, rr^* is the average FFR over the period, $\rho\equiv\rho\left(1\right)$, and the instrument set comprises a constant and four lags of each of the following variables: inflation, the gap, the FFR, the commodity price inflation, the spread between long-term bond rate and the three-month Treasury Bill rate.[14] The target horizon for both the inflation and the unemployment gaps is 1 quarter.

[13] The first estimation period starts in 1959:1 and ends in 1985:1.

[14] Unlike in Clarida et al. (2000), the long-term bond rate used here is not FYGL because that series has been discontinued. Our proxy for the long-term bond rate is instead the ten-year monthly rate of interest on government securities provided by the Fed (we checked that in the overlapping portion with FYGL the data look similar). Similar problems lead us to choose the 3-month U.S. Treasury Bills quoted on the secondary market as a proxy for the

Since we recursively estimate the monetary policy reaction function to analyze whether forecast breakdowns can be linked to changes in monetary policy behavior, we need a long sample. In fact, historically, the Chairmen of the Fed were Burns and Miller from 1960:1 to 1979:2, Volker from 1979:3 to 1987:7 and Greenspan from 1987:7 onwards. Also, as argued in Clarida et al. (2000), "...the sample must be sufficiently long in order to identify the slope coefficients in the policy reaction function as well as the target inflation rate. In fact, estimating the rule over a short sample with little variability in inflation can yield highly misleading results" (cfr. p. 154). Thus, we will address this issue by using finally revised data only.[15]

[FIGURE 3 HERE]

Figure 3 shows rolling estimates of the structural parameters in the monetary policy reaction function of the Fed. We use an estimation window $m = 132$ (which is the number of observations for the first regression, from 1959:1 to 1979:1, which corresponds to the Burns-Miller period). Even though our database is different from Clarida et al. (2000), our estimates of the structural parameters are similar to theirs, and are also in line with those in Orphanides (2001). In fact, β changes from being approximately 1 in the 1959:1-1979:1 to being approximately 2 in the second part of the sample, and similar considerations hold for the other parameters except ρ, which we find is slightly higher than in Clarida and Gali (2001) in both sub-samples.

We next proceed to use these estimates as explanatory variables and investigate whether they are useful to predict the inflation forecast breakdown. Table 4 shows estimates of the following equations:

$$SL_{t+\tau} = \delta_{0,\tau} + \delta_{1,\tau} h_t + \epsilon_{t,h} \tag{24}$$

where h_t is either $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$, or $\widehat{\sigma}^2_{\pi,t}$ (the rolling estimates of the parameters in (23) and of the inflation volatility, $\widehat{\sigma}^2_{\pi,t}$), and $\tau = 1$, 3, 12 months. The table reports estimates of the coefficient $\delta_{1,\tau}$ as well as (in parentheses) the p-values associated with testing whether $\delta_{1,\tau}$ equals zero.[16] For

3-month Treasury Bill rate. Finally, for commodity prices we used n.s.a. CPI for all items all urban consumers (U.S. city average) and we collected data for M2 from the Federal Reserve Board database. To make our notation consistent with that in Clarida et al. (2000), we denote the degree of inflation aversion by β. Thus, while in the previous sections β denoted a generic coefficient, in this section β will denote the Fed's inflation aversion.

[15] Note the different treatement of inflation in the forecasting exercise and in the estimation part: in the forecasting exercise, we impose a unit root, whereas in the estimation of the monetary policy reaction function we use inflation in levels. The reason is that imposing a unit root can help in forecasting (see Rossi (2005)), but in the estimation we follow Clarida et al. (2000) and let GMM estimate the degree of serial correlation in inflation. These authors provide a variety of reasons why this is sensible (mainly the fact that stationarity of the inflation rate is a property of many of the theoretical models that rationalize the use of the kind of monetary policy reaction function analyzed by Clarida et al. These authors do not provide robustness checks when the inflation rate is constrained to be I(1)).

[16] The test statistic is implemented with a Newey and West (1987) HAC estimator with a bandwidth equal to $(\tau - 1)$.

comparison purposes, the table also reports the values of the $t_{m,n,\tau}$ test (which correspond to the case in which eq. (24) contains only the constant, $\delta_{0,\tau}$). It is clear that the degree of inflation targeting smoothing operated by the central bank ($\widehat{\rho}_t$) and the degree of inflation volatility ($\widehat{\sigma}^2_{\pi,t}$) have a significant impact in explaining the out of sample forecast breakdown at the 12 month horizon, whereas inflation volatility and the degree of the Fed's risk aversion to the unemployment gap ($\widehat{\gamma}_t$) are significant at the one month horizon. To conclude, we also consider estimating eq. (24) with a constant and all the rolling estimates of the parameters of the monetary policy reaction function of the Fed and jointly test their significance. The column labeled "*Joint*" reports the values of the $W_{m,n,\tau}$ test statistic (p-values in parentheses). It is clear that these variables were jointly significant at conventional significance levels for all horizons.

[TABLE 4 HERE]

8 Conclusions

This paper proposed a new method to both *detect* and *predict* forecast breakdowns. Regarding the detection of forecast breakdowns, we showed that our test can capture various changes in the data generating process and overfitting. Regarding the prediction of future forecast breakdowns, our test has the potential to uncover significant predictors that provide signals of future forecast breakdowns.

Among the advantages of our test, we note that the framework is general and it captures a variety of causes of forecast breakdowns. At the same time, while the test has power against many possible causes of forecast breakdown, it is not necessarily the most powerful test against each of them. Also, in relation to this issue, while our test provides a first step in detecting and predicting forecast breakdowns, an important question that arises when the researcher finds predictive breakdown is: "what to do next?". Providing insights on the causes of the predictive breakdown is an interesting avenue for future research, and is currently under investigation by the authors.

We applied the test to detect and predict forecast breakdowns in the U.S. Phillips curve in the past three decades. Our results substantiate the finding by Fisher et al. (2002) that the monetary policy conduct of the Fed was a significant predictor of the inflation forecast breakdown.

9 References

Andrews, D.W.K. (1991), "Heteroskedasticy and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica* 59, 817-858.

Andrews, D.W.K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica* 61, 821–856.

Andrews, D.W.K. and W. Ploberger (1994), "Optimal Tests when a Nuisance Parameter Is Present Only Under the Alternative", *Econometrica* 62, 1383–1414.

Bai, J. (1999), "Likelihood Ratio Tests for Multiple Structural Changes", *Journal of Econometrics*, 91, 299-323.

Bai, J. and P. Perron (2003), "Estimating and Testing Liner Models with Multiple Structural Changes", *Econometrica*, 66, 47-78.

Campbell, J. H. and S. B. Thompson (2005), "Predicting the Equity Premium Out of Sample: Can Anything Beat the Hystorical Average", *mimeo,* Harvard University

Carrasco, M. (2002), "Misspecified Structural Change, Threshold, and Markov-switching Models", *Journal of Econometrics,* 109, 239-273.

Chao, J., V. Corradi and N. Swanson (2001), "An Out-of-sample Test for Granger Causality", *Macroeconomic Dynamics.*

Chu, C.-S. J., K. Hornik and C.-M. Kuan (1995), "The Moving-Estimates Test for Parameter Stability", *Econometric Theory*, 11, 699-720.

Clark, T. and M. McCracken (2003), "The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence", *mimeo*, Federal Reserve Bank of Kansas City and University of Missouri-Columbia.

Clarida, R., J. Gali and M. Gertler (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory", *Quarterly Journal of Economics*, pp. 147-180.

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

Clements, M. P., and D. F. Hendry (1999), "Some Methodological Implications of Forecast Failure", *mimeo*, Warwick University and Nuffield College

Croushore, D. and T. Stark (2001), "A Real-Time Data Set for Macroeconomists", *Journal of Econometrics* 105(1), pp. 111-130.

Davidson, R., (1994), Stochastic Limit Theory, Oxford University Press, New York.

Diebold, F. X., and R. S. Mariano (1995), "Comparing Predictive Accuracy", *Journal of Business and Economic Statistics*, 13(3), 253-263.

Dufour, J.-M., E. Ghysels and A. Hall (1994), "Generalized Predictive Tests and Structural Change Analysis in Econometrics", *International Economic Review*, 35, 199-229

Elliott, G. (2005), “Forecasting when there is a Single Break”, *mimeo*, UCSD.

Elliott, G. and U. Muller (2003), “Optimally Testing General Breaking Processes in Linear Time Series Models”, *mimeo,* UCSD and Princeton University.

Engle, R. and T. Bollerslev (1986), “Modelling the Persistence of Conditional Variances”, *Econometric Review,* 5, 1–50

Fisher, J. D. M., C. T. Liu, and R. Zhou (2002), “When Can We Forecast Inflation?”, *Economic Perspectives* 10/2002, Federal Reserve Bank of Chicago.

Ghysels, E. and A. Hall (1990), “A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator”, *International Economic Review*, 31, 355-364.

Ghysels, E., N. R. Swanson and M. Callan (2002) “Monetary Policy Rules with Model and Data Uncertainty”, *Southern Economic Journal*, 69, 239-265.

Goyal, A. and I. Welch (2004), “A Comprehensive Look at the Empirical Performance of Equity Premium Prediction”, NBER Working Paper 10483.

Giacomini, R. and H. White (2004), “Tests of Conditional Predictive Ability”, *mimeo*, UCLA and UCSD.

Hansen, B. E. (2000), “Testing for Structural Change in Conditional Models”, *Journal of Econometrics*, 97, 93-115.

Mathai, A. M., and S. B. Provost (1992), *Quadratic Forms in Random Variables, Theory and Applications*, New York: Marcel Dekker.

McCracken, M. (2000), “Robust out-of-sample inference”, *Journal of Econometrics* 99, 195-223.

McLeish, D. L. (1975), “A Maximal Inequality and Dependent Strong Laws”, *Annals of Probability* 3, 826-836.

Mincer J. and V. Zarnowitz (1969), “The Evaluation of Economic Forecasts”, in: J, Mincer (eds.), Economic Forecasts and Expectations, National Bureau of Economic Research Studies in Business Cycles, 19, 3-46, New York, Columbia University Press

Newey, W., and K. West (1987), “A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix”, *Econometrica* 55, 703-708.

Orphanides, A. (2001), “Monetary Policy Rules Based on Real-Time Data”, *American Economic Review* 91(4), 964-985.

Patton, A.J., and A. Timmermann (2003), “Testable Implications of Forecast Optimality”, *mimeo*, LSE and UCSD.

Pesaran, H., D. Pettenuzzo and A. Timmermann (2004), “Forecasting Time Series Subject to Multiple Structural Breaks”, *CESifo Working Paper* No. 1237.

Ploberger, W., W. Kramer and K. Kontrus (1989), “A New Test for Structural Stability in the Linear Regression Model”, *Journal of Econometrics*, 40, 307-318.

Rossi, B. (2005), “Optimal Tests for Nested Model Selection with Underlying Parameter Instability”, *Econometric Theory*.

Swanson, N., K. van Dijk and M. Callan dataset, available at http://econweb. rutgers.edu/ nswanson/ realtime.htm.

Staiger, D., J. H. Stock, and M. W. Watson (1997), “The NAIRU, Unemployment and Monetary Policy”, *Journal of Economic Perspectives* 11, p. 33-51.

Stock, J. H. and M. W. Watson (2003a), “Combination Forecasts of Output Growth in a Seven-Country Data Set,” forthcoming *Journal of Forecasting*

Stock, J. H. and M. W. Watson (2003b), “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, 41, 788-829

Stock, J. H. and M. W. Watson (1999), “Forecasting Inflation”, *Journal of Monetary Economics* 44, p. 293-335.

Stock, J.H. (1994), “Unit Roots, Structural Breaks and Trends”, in Engle, R. and D. McFadden, *Handbook of Econometrics*, Vol. 4.

West, K. (1996), “Asymptotic Inference about Predictive Ability”, *Econometrica* 64(5), 1067-1084.

White, H. (1994), *Estimation, Inference and Specification Analysis*, New York: Cambridge University Press.

White, H. (2001), *Asymptotic Theory for Econometricians*, San Diego: Academic Press.

Wooldridge, J. M., and H. White (1988), “Some Invariance Principles and Central Limit Theorems for Dependent Heterogeneous Processes”, *Econometric Theory* 4, 210-230.

Appendix. Proofs

Notation 8 *Let* $L_t^* \equiv L_t(\beta^*)$, $h_t^* \equiv h_t(\beta^*)$, $\partial L_t^* \equiv \partial L_t(\beta^*)$, $t = 1, \ldots, T$, *with* L_t *and* h_t *as defined in Algorithm 1;* $D_{t+\tau}^* \equiv \partial SL_{t+\tau}(\beta^*)/\partial\beta$, $t = m, \ldots, T-\tau$; $\widetilde{L}_t^* \equiv L_t^* - E(L_t^*)$; $\widetilde{D}_{t+\tau}^* = D_{t+\tau}^* - E(D_{t+\tau}^*)$; $\widetilde{\partial L}_t^* = \partial L_t^* - E(\partial L_t^*)$. *For a matrix* A, $|A| = \max_{i,j}|a_{ij}|$. $sup_t = \sup_{m\le t\le T-\tau}$. *Limits are for* $m, n \to \infty$.

Lemma 9 $n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{D}_{t+\tau}^* B_t^* H_t^* = o_p(1)$.

Proof of Lemma 9. We prove Lemma 9 for the recursive scheme. The proofs for the fixed and rolling schemes are similar and are available upon request. Direct calculations show that $n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{D}_{t+\tau}^* B_t^* H_t^* = n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*$, where

$$\tilde{w}^h = [\underbrace{c_{m,0}, \ldots, c_{m,0}}_{m}, \underbrace{c_{m,1}, \ldots, c_{m,n-1}}_{n-1}, \underbrace{0, \ldots, 0}_{\tau}],\ c_{m,j} = \sum_{i=1}^{n-j}\frac{\widetilde{D}_{m+\tau+j+i-1}^* B_{m+j+i-1}^*}{m+j+i-1}.$$

We will show that (a) $\left|E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)\right| \xrightarrow{p} 0$ and (b) $E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)^2 \xrightarrow{p} 0$ from which the result follows by Chebyshev's inequality.

(a) First note that $\tilde{w}_t^h$ can be written as a weighted average of the scores: $\tilde{w}_t^h = T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_{t,j}$. For example, $\tilde{w}_1^h = c_{m,0} = T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_{1,j}$ with (nonstochastic) weights

$$\begin{aligned} P_1 &= T[\underbrace{d_{m,0}, \ldots, d_{m,0}}_{m}, \underbrace{d_{m,1}, \ldots, d_{m,\tau-1}}_{\tau-1}, \underbrace{\frac{B_m^*}{m} - d_{m,\tau}, \ldots, \frac{B_{m+n-\tau-1}^*}{m+n-\tau-1} - d_{m,n-1}}_{n-\tau}, \\ &\quad \underbrace{\frac{B_{m+n-\tau}^*}{m+n-\tau}, \ldots, \frac{B_{T-\tau}^*}{T-\tau}}_{\tau}],\ \text{where} \\ d_{m,j} &= \sum_{i=1}^{n-j}\frac{B_{m+j+i-1}^*}{(m+j+i-1)^2}. \end{aligned}$$

Similar expressions can be derived for $c_{m,j}$, $j = 1, \ldots, n-1$. Each component of P_1 is bounded since $|Td_{m,0}| \le \sup_t|B_t^*|\sum_{i=m}^{T-\tau}(T/i^2) \le \sup_t|B_t^*|(Tn/m^2) < \infty$ by assumptions 3 and 6. We can similarly show that P_t has bounded components for all t, which allows us to define $P^{\sup} \equiv \sup_t P_t$.

We thus have

$$
\begin{aligned}
\left|E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)\right| &= \left|E\left(n^{-1/2}\sum_{t=1}^{T}\left[T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_{t,j}\right]h_t^*\right)\right| \\
&\leq \left|E\left(n^{-1/2}\sum_{t=1}^{T}\left[T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_j^{\sup}\right]h_t^*\right)\right| \\
&= \left|E\left(\left[T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^*\right]n^{-1/2}\sum_{t=1}^{T}h_t^*\right)\right| \qquad (25)\\
&\leq T^{-1}n^{-1/2}\sum_{j=1}^{T}\sum_{t=1}^{T}|E(\widetilde{\partial L}_j^* h_{t+j}^*)|
\end{aligned}
$$

where we redefined $\widetilde{\partial L}_j^* P_j^{\sup}$ as $\widetilde{\partial L}_j^*$ in (25) without loss of generality. By Corollary 6.17 of White (2001), $T^{-1}n^{-1/2}\sum_{j=1}^{T}\sum_{t=1}^{T}|E(\widetilde{\partial L}_j^* h_{t+j}^*)| \leq T^{-1}n^{-1/2}C_1\sum_{j=0}^{\infty}j\alpha(j)^{1-1/2r}$, where C_1 is some positive and finite constant and $\alpha(j)$ are the mixing coefficients. By Davidson (1994), p. 210, $\sum_{j=0}^{\infty}j\alpha(j)^{1-1/2r}$ is positive and finite, which implies that $\left|E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)\right| \to 0$.

(b) From (a), $E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)^2 = E\left(n^{-1/2}\sum_{t=1}^{T}\left[T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_{t,j}\right]h_t^*\right)^2$. We have $E\left(n^{-1/2}\sum_{t=1}^{T}\left[T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_{t,j}\right]h_t^*\right)^2 = A_{1T}+A_{2T}+A_{3T}$, where

$$
\begin{aligned}
A_{1T} &\equiv \left(nT^2\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{T}\sum_{j=1}^{T}E\left(h_t^{*\prime}h_s^*\right)E\left(\widetilde{\partial L}_i^* P_{t,i}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right), \\
A_{2T} &\equiv \left(nT^2\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\left[E\left(h_t^{*\prime}P_{t,i}^{\prime}\widetilde{\partial L}_i^{*\prime}\right)E\left(h_s^{*\prime}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right)+E\left(h_t^{*\prime}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right)E\left(h_s^{*\prime}P_{t,i}^{\prime}\widetilde{\partial L}_i^{*\prime}\right)\right], \\
A_{3T} &\equiv \left(nT^2\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{T}\sum_{j=1}^{T}\kappa(t,t-s,t-i,t-j),
\end{aligned}
$$

where $\kappa(t,t-s,t-i,t-j)$ is the fourth cumulant

$$
\begin{aligned}
\kappa(t,t-s,t-i,t-j) &= E\left(h_t^{*\prime}h_s^*\widetilde{\partial L}_i^* P_{t,i}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right)-E\left(h_t^{*\prime}h_s^*\right)E\left(\widetilde{\partial L}_i^* P_{t,i}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right) \\
&\quad -E\left(h_t^{*\prime}P_{t,i}^{\prime}\widetilde{\partial L}_i^{*\prime}\right)E\left(h_s^{*\prime}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right)-E\left(h_t^{*\prime}P_{s,j}^{\prime}\widetilde{\partial L}_j^{*\prime}\right)E\left(h_s^{*\prime}P_{t,i}^{\prime}\widetilde{\partial L}_i^{*\prime}\right).
\end{aligned}
$$

Note that $|A_{1T}| \leq \left(nT^2\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{T}\sum_{j=1}^{T}|E\left(h_t^{*\prime}h_s^*\right)|\left|E\left(\widetilde{\partial L}_i^* P_i^{\sup}P_j^{\sup\prime}\widetilde{\partial L}_j^{*\prime}\right)\right|$. Redefining $\widetilde{\partial L}_i^* P_i^{\sup}$ as $\widetilde{\partial L}_i^*$, we thus have $|A_{1T}| \leq \left(nT^2\right)^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}|E\left(h_t^{*\prime}h_s^*\right)|\sum_{i=1}^{T}\sum_{j=1}^{T}\left|E\left(\widetilde{\partial L}_i^*\widetilde{\partial L}_j^{*\prime}\right)\right| \leq \left(nT^2\right)^{-1}C_2\left(\sum_{j=0}^{\infty}j\alpha(j)^{1-1/2r}\right)^2$, where C_2 is some positive and finite constant and $\alpha(j)$ are the mixing coefficients. As shown in point (a), $\sum_{j=0}^{\infty}j\alpha(j)^{1-1/2r}<\infty$, which implies that $A_{1T}\to 0$. A similar argument can be used to show that $A_{2T}\to 0$. For A_{3T}, we have

$$
|A_{3T}| \leq \left(nT^2\right)^{-1}\sum_{s=1}^{\infty}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sup_{t\geq 1}|\kappa(t,t-s,t-i,t-j)| \to 0,
$$

since $\sum_{s=1}^{\infty}\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\sup_{t\geq 1}|\kappa(t,t-s,t-i,t-j)|<\infty$, by assumptions 1 and 4, as shown by Andrews (1991). ■

Lemma 10 $\frac{T}{n}V_T^{LL*}=var\left(n^{-1/2}\sum_{t=1}^{T}w_t^L\widetilde{L}_t^*\right)>0$ *for all T sufficiently large.*

Proof of Lemma 10. We prove Lemma 10 for the recursive scheme. The proofs for the fixed and rolling schemes are similar and are available upon request. Write $\frac{T}{n}V_T^{LL*}=var(A_1+A_2+A_3+A_4)$, where $A_1=-n^{-1/2}a_{m,0}(\widetilde{L}_1^*+\ldots+\widetilde{L}_m^*)$; $A_2=-n^{-1/2}\left(a_{m,1}\widetilde{L}_{m+1}^*+\ldots+a_{m,\tau-1}\widetilde{L}_{m+\tau-1}^*\right)$; $A_3=n^{-1/2}\left[\left(1-a_{m,\tau}\right)\widetilde{L}_{m+\tau}^*+\ldots+\left(1-a_{m,n-1}\right)\widetilde{L}_{T-\tau}^*\right]$; $A_4=n^{-1/2}\left(\widetilde{L}_{T-\tau+1}+\ldots+\widetilde{L}_T\right)$. We first show that $|cov(A_i,A_j)|\to 0$ for $i\neq j$. Since $a_{m,j}\leq a_{m,0}$, $|cov(A_1,A_2)|\leq n^{-1}a_{m,0}^2$ $|cov(\sum_{t=1}^{m}\widetilde{L}_t^*,\sum_{t=m+1}^{m+\tau-1}\widetilde{L}_t^*)|\,|\leq n^{-1}a_{m,0}^2\sum_{t=1}^{m}\sum_{j=1}^{\tau-1}|E(\widetilde{L}_t^*\widetilde{L}_{t+j}^*)|\leq n^{-1}a_{m,0}^2C\sum_{j=0}^{\infty}j\alpha(j)^{1-1/2r}$ by Corollary 6.17 of White (2001), where C is some positive and finite constant and $\alpha(j)$ are the mixing coefficients. By Davidson (1994), p. 210, $\sum_{j=0}^{\infty}j\alpha(j)^{1-1/2r}$ is positive and finite. Further, $a_{m,0}^2\to\ln^2(1+\pi)$, which is finite (cf. West, 1996, pg. 1082). As a result, $cov(A_1,A_2)\to 0$. Using analogous reasonings and the fact that $1-a_{m,t-m}\leq 1$ for all t, one can show that $|cov(A_i,A_j)|\to 0$ for the remaining (i,j) pairs. We thus have that $var\left(n^{-1/2}\sum_{t=1}^{T}w_t^L\widetilde{L}_t^*\right)$ can be approximated by $\sum_{i=1}^{4}var(A_i)$ and the desired result follows from the fact that, e.g., $var(A_1)=(m/n)a_{m,0}^2var(m^{-1/2}\sum_{t=1}^{m}\widetilde{L}_t^*)>0$ since $m/n\to\pi^{-1}>0$, $a_{m,0}^2\to\ln^2(1+\pi)>0$, and $var(m^{-1/2}\sum_{t=1}^{m}\widetilde{L}_t^*)>0$ by assumption 6. ■

Proof of Theorem 2. (b) A second order mean value expansion of $SL_{t+\tau}(\widehat{\beta}_t)=L_{t+\tau}\left(\widehat{\beta}_t\right)-\bar{L}_t\left(\widehat{\beta}_t\right)$ around β^* gives

$$\begin{aligned}
& n^{1/2}\left[n^{-1}\sum_{t=m}^{T-\tau}SL_{t+\tau}(\widehat{\beta}_t)-E\left(n^{-1}\sum_{t=m}^{T-\tau}SL_{t+\tau}(\beta^*)\right)\right] \qquad (26)\\
= \; & n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}\left(\beta^*\right)-E\left(SL_{t+\tau}\left(\beta^*\right)\right)\right]+n^{-1/2}\sum_{t=m}^{T-\tau}\left(\partial SL_{t+\tau}(\beta^*)/\partial\beta\right)\left(\widehat{\beta}_t-\beta^*\right)+o_p\left(1\right)\\
= \; & n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}\left(\beta^*\right)-E\left(SL_{t+\tau}\left(\beta^*\right)\right)\right]+n^{-1/2}\sum_{t=m}^{T-\tau}E\left(D_{t+\tau}^*\right)B_t^*H_t^*+\\
& n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{D}_{t+\tau}^*B_t^*H_t^*+o_p\left(1\right),
\end{aligned}$$

where we used assumption 3. We show that, under H_0,

$$\left(\frac{T}{n}V_T\right)^{-1/2}n^{-1/2}\left[\sum_{t=m}^{T-\tau}SL_{t+\tau}\left(\beta^*\right),\sum_{t=m}^{T-\tau}D_{t+\tau}^*B_t^*H_t^*\right]'\xrightarrow{d}N(0,I_2),$$

with V_T defined in (5), from which the theorem follows. Direct calculations show that $\left(\frac{T}{n}V_T\right)^{-1/2}n^{-1/2}\left[\sum_{t=m}^{T-\tau}SL_{t+\tau}\left(\beta^*\right),\sum_{t=m}^{T-\tau}D_{t+\tau}^*B_t^*H_t^*\right]'=V_T^{-1/2}T^{-1/2}\left[\sum_{t=1}^{T}w_t^LL_t^*,\sum_{t=1}^{T}w_t^{h*}h_t^*\right]'$, where w_t^{h*} equals w^h defined Algorithm 1 with $\widehat{\beta}_t$, B_t, $D_{t+\tau}$ replaced respectively by β^*, B_t^* and $E(D_{t+\tau}^*)$. Under H_0, we have $T^{-1/2}\sum_{t=1}^T w_t^L L_t^* = T^{-1/2}\sum_{t=1}^T w_t^L \widetilde{L}_t^*$, since $T^{-1/2}\sum_{t=1}^T w_t^L E\left(L_t^*\right) = nT^{-1/2}E\left(n^{-1}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*)\right) = 0$. We show that

$$V_T^{*-1/2}T^{-1/2}\left[\sum_{t=1}^T w_t^L \widetilde{L}_t^*, \sum_{t=1}^T w_t^{h*}h_t^*\right]' \xrightarrow{d} N(0, I_2),$$

where $V_T^* = var\left(T^{-1/2}\left[\sum_{t=1}^T w_t^L \widetilde{L}_t^*, \sum_{t=1}^T w_t^{h*}h_t^*\right]'\right)$. The result follows from the fact that $V_T - V_T^* \xrightarrow{p} 0$, due to consistency of $\widehat{\beta}_t$ for β^* under H_0. We verify that the zero-mean vector sequence $\{\left[V_T^{*-1/2}w_t^L\widetilde{L}_t^*, V_T^{*-1/2}w_t^{h*}h_t^*\right]'\}$ satisfies the conditions of Wooldridge and White's (1988) CLT for mixing processes. Since $Z_t \equiv \left[V_T^{*-1/2}w_t^L\widetilde{L}_t^*, V_T^{*-1/2}w_t^{h*}h_t^*\right]$ is a function of only a finite number of leads and lags of W_t, it follows from Lemma 2.1 of White and Domowitz (1984) that it is mixing of the same size as W_t. For the first component of Z_t, we have $E|V_T^{*-1/2}w_t^L\widetilde{L}_t^*|^{2r} < \infty$ by assumption 4 and by the fact that V_T is p.d. and $|w_t^L| < \infty$ for all t (for the fixed and rolling schemes, this follows from assumption 7; for the recursive scheme, it follows from the fact that $a_{m,j} \leq a_{m,0} \to \ln(1+\pi) < \infty$, as shown in the proof of Lemma 10. For the second component of Z_t, writing $w_t^{h*} = T^{-1}\sum_{j=1}^T E\left(\partial L_j^*\right)P_{t,j}$ - using similar reasonings as those in the proof of Lemma 9-(a) - we have $E|V_T^{*-1/2}w_t^{h*}h_t^*|^{2r} = E|V_T^{*-1/2}T^{-1}\sum_{j=1}^T E\left(\partial L_j^*\right)P_{t,j}h_t^*|^{2r} \equiv E|\lambda_t h_t^*|^{2r}$. Note that $|\lambda_{t,i}| < \infty$ for all t, i, by assumption 5, by $P_{t,j}$ having bounded components (as shown in the proof of Lemma 9-(a)) and by V_T^* p.d. Further, by Minkowski's inequality,

$$E|V_T^{*-1/2}w_t^{h*}h_t^*|^{2r} = E|\lambda_t' h_t^*|^{2r} = E|\sum_{i=1}^q \lambda_{t,i}h_{t,i}^*|^{2r} \leq [\sum_{i=1}^q |\lambda_{t,i}|(E|h_{t,i}^*|^{2r})^{1/2r}]^{2r} < \infty$$

by assumption 4. This implies that $V_T^{*-1/2}T^{-1/2}\left[\sum_{t=1}^T w_t^L\widetilde{L}_t^*, \sum_{t=1}^T w_t^{h*}h_t^*\right]' \xrightarrow{d} N(0, I_2)$. The desired result then follows from consistency of V_T for V_T^* due to $\widehat{\beta}_t - \beta^* \xrightarrow{p} 0$ under H_0.

(a) $E\left(\partial SL_{t+\tau}(\beta^*)/\partial\beta\right) = E\left(\partial L_{t+\tau}(\beta^*)/\partial\beta\right) - E\left(\partial\overline{L}_t(\beta^*)/\partial\beta\right) = 0$, and thus expression (26) reduces to $n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}\left(\beta^*\right) - E\left(SL_{t+\tau}\left(\beta^*\right)\right)\right] + o_p\left(1\right)$. The result then follows from reasonings analogous to those in part (b) above and from Lemma 10. ■

Proof of Corollary 3. Follows from the fact that, under H_0, $E\left(\frac{\partial\overline{L}_t}{\partial\beta}(\beta^*)\right) = 0$ for all t, which implies that the condition of Theorem 2-*(a)* is satisfied. ■

Lemma 11 *For $a_{m,j}$ as defined in (4), we have: (i)* $a_{m,j} \simeq \ln(\frac{m+n-1}{m+j})$; *(ii)* $n^{-1}\sum_{j=\tau}^{n-1}a_{m,j} \simeq 1 - \pi^{-1}\ln(1+\pi)$; *(iii)* $n^{-1}\sum_{j=\tau}^{n-1}a_{m,j}^2 \simeq 2\left[1 - \pi^{-1}\ln(1+\pi)\right] - \pi^{-1}\ln(1+\pi)$.

Proof of Lemma 11. *(i)* $a_{m,j} = \sum_{i=j}^{n-1}(m+i)^{-1} \simeq \int_j^{n-1}(m+x)^{-1}dx = \ln\left(\frac{m+n-1}{m+j}\right)$; *(ii)* $n^{-1}\sum_{j=\tau}^{n-1}a_{m,j} \simeq n^{-1}\int_\tau^{n-1}\ln\left(\frac{m+n-1}{m+x}\right)dx = n^{-1}\left[n-1-\tau-(m-\tau)\ln(\frac{m+n-1}{m+\tau})\right] \to 1 -$

$\pi^{-1}\ln(1+\pi)$; *(iii)* $n^{-1}\sum_{j=\tau}^{n-1} a_{m,j}^2 \simeq n^{-1}\int_{\tau}^{n-1}\ln^2\left(\frac{m+n-1}{m+x}\right)dx =$
$n^{-1}\left[2(n-\tau)-2(m+\tau)\ln(\frac{m+n-1}{m+\tau})-(m+\tau)\ln^2(\frac{m+n-1}{m+\tau})\right] \rightarrow 2\left[1-\pi^{-1}\ln(1+\pi)\right]-\pi^{-1}\ln(1+\pi)$. ■

Proof of Proposition 4. We show that $\lim var(n^{-1/2}\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \lambda^* \sum_{j=-\infty}^{\infty}\Gamma_j$, where $\lambda^* = 1+\pi$ for the fixed scheme; $\lambda^* = 1-(1/3)\,\pi^2$ for the rolling $(n<m)$ scheme; $\lambda^* = (2/3)\,\pi^{-1}$ for the rolling $(n \geq m)$ scheme; $\lambda^* = 1$ for the recursive scheme. The desired result then follows from λS_n^{LL} being a consistent estimator of $\lambda^* \sum_{j=-\infty}^{\infty}\Gamma_j$ under H_0. For conciseness, we focus on the recursive scheme. As shown in the proof of Lemma 10, $var(n^{-1/2}\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \sum_{i=1}^{4} var(A_i)$. We have $var(A_1) = (m/n)\,a_{m,0}^2 var(m^{-1/2}\sum_{t=1}^{m}\widetilde{L}_t^*)$ and thus $\lim var(A_1) = \pi^{-1}\ln(1+\pi)\sum_{j=-\infty}^{\infty}\Gamma_j$ by Lemma 11-*(i)*. Further, $var\,(A_2) = n^{-1}var\left(a_{m,1}\widetilde{L}_{m+1}^* + \ldots + a_{m,\tau-1}\widetilde{L}_{m+\tau-1}^*\right) \rightarrow 0$ since τ is fixed. For A_3, it follows from West (1996), pg. 1082-1083, (with $(1-a_{m,j})$ substituting $a_{m,j}$) that $var(A_3) = n^{-1}d_0\sum_{j=-n+2}^{n-2}\Gamma_j + o(1)$, where $d_0 = \sum_{j=\tau}^{n-1}(1-a_{m,j})^2$. By Lemma 11, $n^{-1}d_0 = \frac{n-\tau}{n} - 2n^{-1}\sum_{j=\tau}^{n-1}a_{m,j} + n^{-1}\sum_{j=\tau}^{n-1}a_{m,j}^2 \rightarrow 1-\pi^{-1}\ln(1+\pi)$, and thus $\lim var(A_2) = \left[1-\pi^{-1}\ln(1+\pi)\right]\sum_{j=-\infty}^{\infty}\Gamma_j$. Finally, $var(A_4) = n^{-1}var(\widetilde{L}_{T-\tau+1} + \ldots + \widetilde{L}_T) \rightarrow 0$ since τ is fixed. In sum, $var(n^{-1/2}\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \sum_{j=-\infty}^{\infty}\Gamma_j$ and thus $\lambda^* = 1$. The proofs for the fixed and rolling schemes follow from similar reasonings (for each scheme, λ^* can be computed as the limit of $n^{-1}w^{L\prime}w^L$). ■

Proof of Proposition 5. A mean value expansion of $\overline{SL}\left(\widehat{\beta}_t\right) \equiv \frac{1}{n}\sum_{t=m+1}^{T}\left[L_{t+\tau}\left(\widehat{\beta}_t\right) - \overline{L}_t\left(\widehat{\beta}_t\right)\right]$ around β_t^* gives:

$$\begin{aligned}\overline{SL}_{m,n}\left(\widehat{\beta}_t\right) &= \overline{SL}_{m,n}\left(\beta_t^*\right) + \frac{1}{n}\sum_{t=m}^{T-\tau}\left\{\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_t^*\right) - \frac{\partial}{\partial\beta}\overline{L}_t\left(\beta_t^*\right)\right\}'\left(\widehat{\beta}_t - \beta_t^*\right) + \\ &+ \frac{1}{2}\frac{1}{n}\sum_{t=m}^{T-\tau}\left(\widehat{\beta}_t - \beta_t^*\right)'\left\{\frac{\partial^2}{\partial\beta^2}L_{t+\tau}(\overline{\widehat{\beta}}_t) - \frac{\partial^2}{\partial\beta^2}\overline{L}_t(\overline{\widehat{\beta}}_t)\right\}\left(\widehat{\beta}_t - \beta_t^*\right) \qquad (27)\end{aligned}$$

where $\overline{\widehat{\beta}}_t$ is an intermediate point between β_t^* and $\widehat{\beta}_t$. Note also that:

$$\begin{aligned}L_{t+\tau}\left(\beta_t^*\right) &= L_{t+\tau}\left(\beta_{t+\tau}^*\right) + \frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_{t+\tau}^*\right)'\left(\beta_t^* - \beta_{t+\tau}^*\right) + \qquad (28) \\ &+ \frac{1}{2}\left(\beta_t^* - \beta_{t+\tau}^*\right)'\frac{\partial^2}{\partial\beta^2}L_{t+\tau}\left(\overline{\beta^*}_{t+\tau}\right)\left(\beta_t^* - \beta_{t+\tau}^*\right)\end{aligned}$$

$$\begin{aligned}L_j\left(\beta_t^*\right) &= L_j\left(\beta_j^*\right) + \frac{\partial}{\partial\beta}L_j\left(\beta_j^*\right)'\left(\beta_t^* - \beta_j^*\right) + \qquad (29) \\ &+ \frac{1}{2}\left(\beta_t^* - \beta_j^*\right)'\frac{\partial^2}{\partial\beta^2}L_j\left(\overline{\beta^*}_j\right)\left(\beta_t^* - \beta_j^*\right)\end{aligned}$$

where $\overline{\beta^*}_{t+\tau}$ is an intermediate point between β_t^* and $\beta_{t+\tau}^*$, and $\overline{\beta^*}_j$ is an intermediate point between

β_t^* and β_j^*. From (28) and (29) above, it follows that

$$\begin{aligned} SL_{t+\tau}\left(\beta_t^*\right) &= L_{t+\tau}\left(\beta_{t+\tau}^*\right) - \overline{\sum}_j L_j\left(\beta_j^*\right) + \\ &+\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_{t+\tau}^*\right)'\left(\beta_t^* - \beta_{t+\tau}^*\right) - \overline{\sum}_j\frac{\partial}{\partial\beta}L_j\left(\beta_j^*\right)'\left(\beta_t^* - \beta_j^*\right) \\ &+\frac{1}{2}\left[\left(\beta_t^* - \beta_{t+\tau}^*\right)'\frac{\partial^2}{\partial\beta^2}L_{t+\tau}\left(\overline{\beta^*}_{t+\tau}\right)\left(\beta_t^* - \beta_{t+\tau}^*\right)\right] \\ &-\frac{1}{2}\overline{\sum}\left(\beta_t^* - \beta_j^*\right)'\frac{\partial^2}{\partial\beta^2}L_j\left(\overline{\beta^*}_j\right)\left(\beta_t^* - \beta_j^*\right) \end{aligned} \tag{30}$$

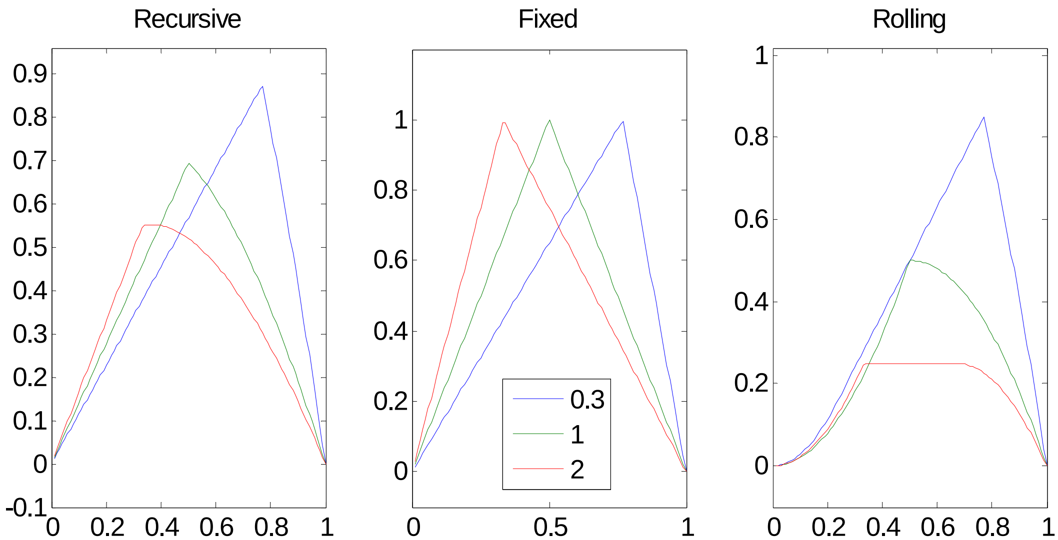
Let $\overline{SL}_{m,n}\left(\beta_{1:T}^*\right) \equiv \frac{1}{n}\sum_{t=m}^{T-\tau}\left[L_{t+\tau}\left(\beta_{t+\tau}^*\right) - \overline{\sum}_j L_j\left(\beta_j^*\right)\right]$, sum (30) from m to $T-\tau$, divide it by n, and substitute the result into (27) to obtain:

$$\begin{aligned} \overline{SL}_{m,n}\left(\widehat{\beta}_t\right) &= \overline{SL}_{m,n}\left(\beta_{1:T}^*\right) + \frac{1}{n}\sum_{t=m}^{T-\tau}\left[\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_{t+\tau}^*\right)'\left(\beta_t^* - \beta_{t+\tau}^*\right) - \overline{\sum}_j\frac{\partial}{\partial\beta}L_j\left(\beta_j^*\right)'\left(\beta_t^* - \beta_j^*\right)\right] \\ &+\frac{1}{2}\frac{1}{n}\sum_{t=m}^{T-\tau}\left[\left(\beta_t^* - \beta_{t+\tau}^*\right)'\frac{\partial^2 L_{t+\tau}\left(\overline{\beta^*}_{t+\tau}\right)}{\partial\beta^2}\left(\beta_t^* - \beta_{t+\tau}^*\right)\right. \\ &\left.-\overline{\sum}_j\left(\beta_t^* - \beta_j^*\right)'\frac{\partial^2}{\partial\beta^2}L_j\left(\overline{\beta^*}_j\right)\left(\beta_t^* - \beta_j^*\right)\right] \qquad (31) \\ &+\frac{1}{n}\sum_{t=m}^{T-\tau}\left\{\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_t^*\right) - \frac{\partial}{\partial\beta}\overline{L}_t\left(\beta_t^*\right)\right\}'\left(\widehat{\beta}_t - \beta_t^*\right) \qquad (32) \\ &+\frac{1}{2}\frac{1}{n}\sum_{t=m}^{T-\tau}\left(\widehat{\beta}_t - \beta_t^*\right)'\left[\frac{\partial^2}{\partial\beta^2}L_{t+\tau}(\overline{\widehat{\beta}}_t) - \frac{\partial^2}{\partial\beta^2}\overline{L}_t(\overline{\widehat{\beta}}_t)\right]\left(\widehat{\beta}_t - \beta_t^*\right) \end{aligned}$$

Note that, since $0 = \frac{\partial}{\partial\beta}\overline{\mathcal{L}}_t\left(\widehat{\beta}_t\right) = \frac{\partial}{\partial\beta}\overline{\mathcal{L}}_t\left(\beta_t^*\right) + \left[\frac{\partial^2}{\partial\beta^2}\overline{\mathcal{L}}_t(\overline{\widehat{\beta}}_t)\right]'\left(\widehat{\beta}_t - \beta_t^*\right)$, then $\left\{\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_t^*\right) - \frac{\partial}{\partial\beta}\overline{L}_t\left(\beta_t^*\right)\right\}' = \left(\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_t^*\right)\right)' - \left(\frac{\partial}{\partial\beta}\left[\overline{L}_t\left(\beta_t^*\right) - \overline{\mathcal{L}}_t\left(\beta_t^*\right)\right]\right)' + \left(\widehat{\beta}_t - \beta_t^*\right)'\frac{\partial^2}{\partial\beta^2}\overline{\mathcal{L}}_t(\overline{\widehat{\beta}}_t)$. Therefore, by taking expectations of (31), we have (10). ■

Proof of Proposition 6. Since $E\left(\frac{\partial}{\partial\beta}L_t\left(\beta_t\right) - \frac{\partial}{\partial\beta}\mathcal{L}_t\left(\beta_t\right)\right) = 0\ \forall t$, the "parameter instabilities I" component is zero. The "parameter instabilities II" component is $\frac{1}{2}\frac{1}{n}\sum_{t=m}^{T-\tau}E\left[\left(\beta - \left(\beta + g_1 T^{-1/4}\right)\right)'J\left(\beta - \left(\beta + g_1 T^{-1/4}\right)\right)\right] = \frac{1}{2}g_1'Jg_1/\sqrt{T} + o_p(1)$ and the "other instabilities" component is $g_2/\sqrt{T}$. Since $\frac{\partial}{\partial\beta}L_{t+\tau}\left(\beta_t\right) = -2X_{t+\tau}\left(Y_{t+\tau} - X_{t+\tau}'\beta_t\right)$ is uncorrelated with $\left(\widehat{\beta}_t - \beta_t\right)$, the "overfit I" component is zero. Since $E\left(\frac{\partial^2}{\partial\beta^2}L_j\left(\beta\right)\right) = E\left(\frac{\partial^2}{\partial\beta^2}\mathcal{L}_j\left(\beta\right)\right) = 2J$ $\forall j$, the "overfitting III" component in (10) is also zero. Finally, the "overfit II" component equals $E\left(\widehat{\beta}_m - \beta\right)'\left(\frac{2}{m}\sum_{s=1}^m X_s X_s'\right)\left(\widehat{\beta}_m - \beta\right) = \frac{2}{m}E\left(\frac{1}{\sqrt{m}}\sum_{s=1}^m X_s\varepsilon_s\right)'\left(\frac{1}{m}\sum_{s=1}^m X_s X_s'\right)^{-1}\left(\frac{1}{\sqrt{m}}\sum_{s=1}^m X_s\varepsilon_s\right)$ $\underset{p}{\to} \frac{2}{m}\sigma^2 E\left(\chi_k^2\right) = 2\sigma^2\frac{k}{m}$.
■

Figure 1. Weighting schemes



Notes. The formulas for the various forecasting schemes are provided in (13).

Figure 2. Power functions

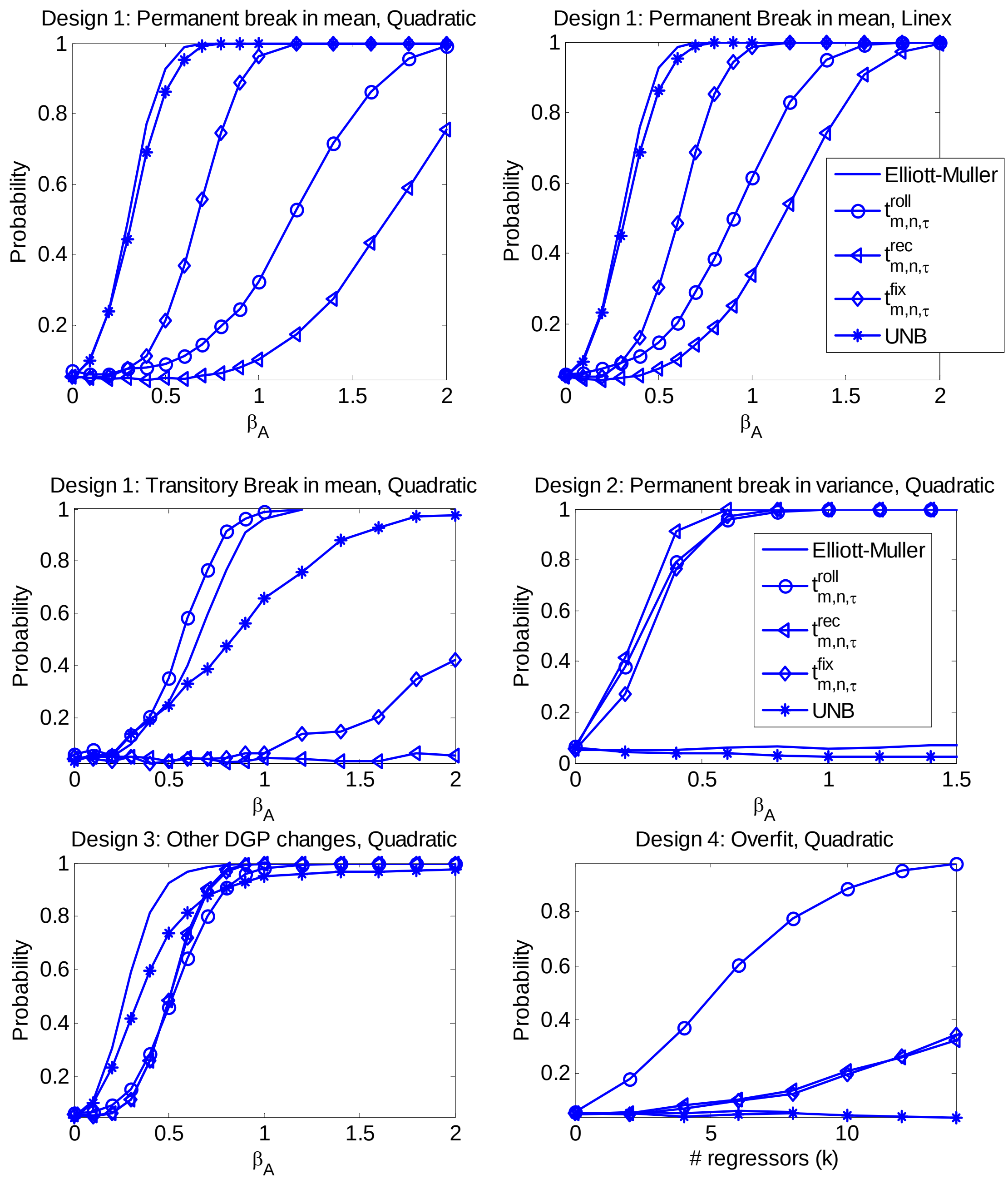
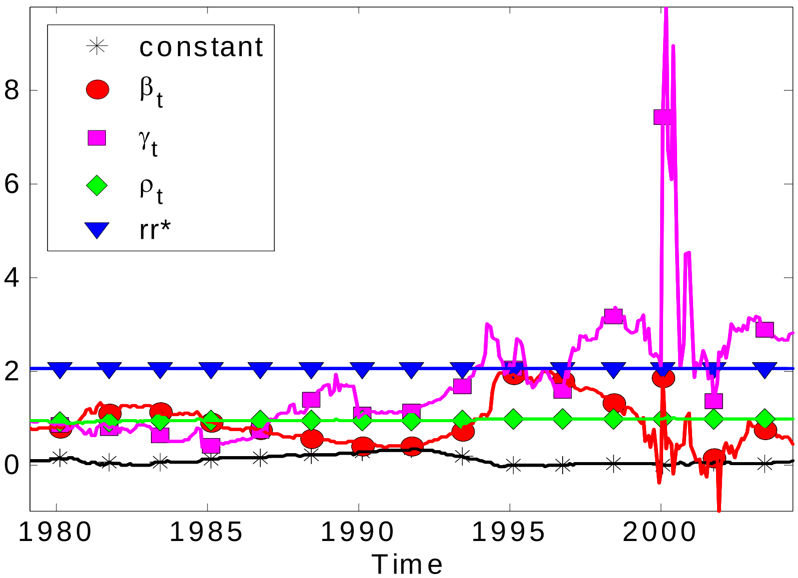


Figure 3. Rolling estimates of the structural parameters of the Fed's reaction function



Notes to Figure 3. The figure shows rolling estimates of the structural parameters in the monetary policy reaction function of the Fed, eq. (23).

Table 1. Size. Nominal size .05

				MC1					
		$t_{m,n,\tau}$			$t^{\text{stat}}_{m,n,\tau}$				
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.	EM	UNB
50	50	.113	.144	.097	.064	.096	.058	.057	.055
50	100	.152	.297	.121	.077	.244	.071	.057	.051
50	150	.168	.492	.128	.080	.440	.075	.055	.049
100	50	.072	.071	.065	.049	.052	.047	.053	.053
100	100	.096	.109	.081	.057	.075	.055	.055	.053
100	150	.101	.143	.086	.060	.117	.059	.059	.053
150	50	.044	.046	.040	.036	.038	.035	.054	.052
150	100	.064	.072	.058	.046	.052	.043	.052	.053
150	150	.069	.087	.065	.047	.066	.046	.049	.050
				MC2					
		$t_{m,n,\tau}$			$t^{\text{stat}}_{m,n,\tau}$				
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.	EM	UNB
50	50	.272	.165	.120	.187	.090	.054	1	.040
50	100	.178	.293	.130	.050	.179	.042	1	.024
50	150	.183	.415	.122	.036	.268	.042	1	.014
100	50	.047	.056	.046	.031	.036	.030	1	.028
100	100	.087	.098	.079	.036	.054	.034	1	.062
100	150	.115	.105	.092	.040	.066	.034	1	.016
150	50	.030	.032	.028	.024	.024	.022	1	.059
150	100	.062	.069	.058	.033	.036	.031	1	.042
150	150	.077	.079	.069	.033	.041	.032	1	.029

Notes. The table reports rejection frequencies over 5000 Monte Carlo replications of the following tests: the forecast breakdown test of Section 2.3, using either the asymptotic variance estimator of Corollary 3 $(t_{m,n,\tau})$ or the estimator of Corollary 4 $\left(t^{\text{stat}}_{m,n,\tau}\right)$, both tests implemented with either a fixed, rolling or recursive scheme; Elliott and Muller's (2003) test (EM) and a forecast unbiasedness test (UNB). The experiment designs MC1 and MC2 are described in Section 6.1 and m and n denote in-sample and out-of-sample sizes, respectively.

Table 2. Size of overfitting-corrected tests. Nominal size .05

		MC1					
		$t^{\mathrm{c}}_{m,n,\tau}$			$t^{\mathrm{stat,c}}_{m,n,\tau}$		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
50	50	.064	.053	.053	.031	.031	.028
50	100	.085	.056	.066	.031	.042	.032
50	150	.095	.068	.065	.034	.053	.029
100	50	.043	.040	.038	.029	.030	.027
100	100	.057	.057	.052	.030	.036	.031
100	150	.068	.055	.056	.032	.041	.033
150	50	.031	.030	.027	.024	.024	.022
150	100	.050	.047	.046	.032	.031	.030
150	150	.058	.053	.053	.038	.035	.034
		MC2					
		$t^{\mathrm{c}}_{m,n,\tau}$			$t^{\mathrm{stat,c}}_{m,n,\tau}$		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
50	50	.256	.080	.079	.189	.039	.037
50	100	.122	.083	.069	.042	.050	.027
50	150	.096	.073	.067	.023	.053	.023
100	50	.044	.045	.043	.031	.031	.030
100	100	.071	.059	.057	.035	.033	.029
100	150	.088	.045	.066	.033	.030	.028
150	50	.031	.029	.028	.028	.029	.028
150	100	.057	.049	.047	.035	.027	.028
150	150	.062	.043	.050	.029	.026	.026

Notes. The table reports rejection frequencies over 5000 Monte Carlo replications of the overfitting-corrected forecast breakdown test of Section 4, using either the asymptotic variance estimator of Corollary 3 $\left(t^{\mathrm{c}}_{m,n,\tau}\right)$ or the estimator of Corollary 4 $\left(t^{\mathrm{stat,c}}_{m,n,\tau}\right)$, both tests implemented with either a fixed, rolling or recursive scheme. The experiment designs MC1 and MC2 are described in Section 6.1 and m and n denote in-sample and out-of-sample sizes, respectively.

Table 3. P-values of Forecast Breakdown test

q_u	q_π	Real-time data $t_{m,n,\tau}$	Real-time data $W_{m,n,\tau}$	Revised data $t_{m,n,\tau}$	Revised data $W_{m,n,\tau}$
$\tau = 1$		*1 month horizon*			
1	1	- -	- -	0.037	0.004
1	3	- -	- -	0.093	0.016
3	1	- -	- -	0.061	0.007
3	3	- -	- -	0.134	0.026
BIC		- -	- -	0.102	0.0121
$\tau = 3$		*1 quarter horizon*			
1	1	0	0	0.408	0.000
1	3	0.585	0.000	0.474	0.000
3	1	0.477	0.440	0.568	0.000
3	3	0.595	0.011	0.643	0.001
BIC		0.882	0.088	0.621	0.000
$\tau = 12$		*1 year horizon*			
1	1	0.001	0.004	0.238	0.001
1	3	0	0	0.454	0.006
3	1	0.002	0.008	0.818	0.001
3	3	0.001	0	0.962	0.009
BIC		0.001	0	0.644	0.002

Notes to Table 3. The table reports p-values for both the forecast breakdown test ($t_{m,n,\tau}$) and the $W_{m,n,\tau}$ test of Section 5. The $W_{m,n,\tau}$ test is implemented with a constant and one lagged value of the surprise loss, SL_{t-1}. We used $m = 60$, $n = 95$ in the real-time data case, and $m = 241$ and $n = 546$ in the revised data case. The forecast horizons are $\tau = 1$, 3 and 12 months (note that real-time data are available only at a quarterly frequency, so in that case only results for $\tau = 3$ months (1 quarter) and $\tau = 12$ months (4 quarters) are reported). q_u and q_π are, respectively, the number of lags used for unemployment and for inflation; the row labeled "BIC" reports results for the case in which the lag length is determined at every t by the BIC with a maximum of three lags. The one-step ahead forecasts begin in 1993:1 (the date has been chosen to correspond to a change in monetary policy according to Fisher et al. (2002)).

Table 4. Explaining forecast breakdowns by monetary policy changes and inflation variance

			$t_{m,n,\tau}$	$W_{m,n,\tau}$				
τ	q_u	q_π		$\widehat{\beta}_t$	$\widehat{\gamma}_t$	$\widehat{\rho}_t$	*Joint*	$\widehat{\sigma}^2_{\pi,t}$
1	1	1	2.090	2.285	-1.828	-19.770	16.88	0.991
			(0.04)	(0.17)	(0.01)	(0.60)	(0.00)	(0.00)
	1	3	1.678	2.348	-1.612	-6.484	14.09	0.860
			(0.09)	(0.16)	(0.02)	(0.86)	(0.00)	(0.02)
	3	1	1.875	2.306	-1.712	-13.957	14.97	0.947
			(0.06)	(0.17)	(0.01)	(0.71)	(0.00)	(0.01)
	3	3	1.500	2.354	-1.513	-1.977	12.68	0.830
			(0.13)	(0.16)	(0.03)	(0.96)	(0.00)	(0.02)
	BIC		1.633	2.186	-1.654	-6.272	14.15	0.830
			(0.10)	(0.19)	(0.02)	(0.87)	(0.00)	(0.02)
3	1	1	0.828	1.806	0.404	114.281	4.64	1.478
			(0.41)	(0.57)	(0.76)	(0.11)	(0.03)	(0.01)
	1	3	0.716	1.837	0.267	122.4	5.68	1.482
			(0.47)	(0.55)	(0.84)	(0.08)	(0.02)	(0.01)
	3	1	0.570	1.651	0.568	128.8	4.81	1.464
			(0.57)	(0.61)	(0.67)	(0.08)	(0.03)	(0.02)
	3	3	0.464	1.657	0.415	136.1	5.93	1.467
			(0.64)	(0.60)	(0.75)	(0.06)	(0.01)	(0.01)
	BIC		0.494	1.608	0.642	141.4	5.72	1.363
			(0.62)	(0.62)	(0.63)	(0.05)	(0.02)	(0.02)
12	1	1	1.180	1.304	0.105	199.5	10.84	1.389
			(0.24)	(0.76)	(0.95)	(0.03)	(0.00)	(0.01)
	1	3	0.749	1.639	0.417	192.0	9.03	1.143
			(0.45)	(0.69)	(0.81)	(0.03)	(0.00)	(0.04)
	3	1	0.230	0.679	0.863	256.5	12.03	1.328
			(0.82)	(0.88)	(0.66)	(0.01)	(0.00)	(0.04)
	3	3	-0.048	0.960	1.108	250.9	11.78	1.117
			(0.96)	(0.83)	(0.55)	(0.01)	(0.00)	(0.07)
	BIC		0.462	0.903	0.789	246.5	11.61	1.261
			(0.64)	(0.84)	(0.68)	(0.01)	(0.00)	(0.04)

Notes to Table 4. The table reports the empirical results for the estimation of equation (24). The first regression dates have been selected according to the findings in Fisher et al. (2002). The regressions for the Forecast Breakdown test $W_{m,n,\tau}$ contain a constant and each of the following regressors: SL_t (the lagged value of the surprise loss), $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$ (the rollingly estimated structural parameters in the monetary policy reaction function of the Fed), and $\widehat{\sigma}^2_{\pi,t-1}$ (the inflation volatility). The column labeled "Joint" reports instead the joint test on a constant and all the parameters $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$. P-values of the $W_{m,n,\tau}$ test statistic (with a HAC bandwidth equal to $(\tau - 1)$) for testing whether the explanatory variable is insignificant are reported in parentheses. For comparison purposes, we also report results for the unconditional $t_{m,n,\tau}$ test. q_u and q_π are, respectively, the number of lags used for unemployment and for inflation; rows labeled "BIC" report results for the case in which the lag length is determined at every t by the BIC with a maximum of three lags. The horizons are one month ($\tau = 1$), one quarter ($\tau = 3$), and one year ($\tau = 12$).