Detecting and Predicting Forecast Breakdowns*

Raffaella Giacomini and Barbara Rossi

University of California, Los Angeles and Duke University

December 2005

Abstract

We propose a theoretical framework for assessing whether a forecast model estimated over one period can provide good forecasts over a subsequent period. We formalize this idea by defining a forecast breakdown as a situation in which 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 not only to detect past forecast breakdowns but also to predict future ones. We show that main causes of forecast breakdowns are instabilities in the data generating process and relate the properties of our forecast breakdown test to those of existing 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. As a by-product, we show that our results can be applied to forecast rationality tests and provide the appropriate asymptotic variance estimator that corrects the size distortions of previous forecast rationality tests. The empirical application finds evidence of a forecast breakdown in the Phillips' curve forecasts of U.S. inflation, and links it to inflation volatility and 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, Chris Otrok, Martin Schwerdtfeger, Allan Timmermann as well as participants to the 2005 Missouri Economics Conference, the 2005 CIRANO-CIREQ Conference on Forecasting in Macroeconomics and Finance, the 2005 JAE Annual Conference, the 2005 Conference on Frontiers in Time Series Econometrics, the 2005 NBER-NSF Time Series Conference, the 2005 SEA Meetings, the 2005 Econometric Society World Congress, the 2005 Triangle Econometrics Conference and the 2005 Bank of England forecasting workshop for useful discussions and suggestions. The paper benefited from comments from seminar participants at UCLA, Duke, Berkeley, UCR, UCSD and University of Montreal.

1 Introduction

This paper proposes a new method for evaluating a forecasting model for a macroeconomic or financial variable. There is a large literature claiming that certain models are good at predicting macroeconomic variables such as output growth and inflation (Stock and Watson, 2003b and Clark and McCracken, 2003) and that a range of variables have predictive power for stock market returns (e.g., the references in Goyal and Welch, 2004 and Campbell and Thompson, 2005). These claims are based either on some measure of a model's in-sample fit (most of the literature on stock return predictability), or on the model's out-of-sample performance (e.g., Stock and Watson, 2003b). The robustness of these results has been however recently challenged. On the one hand, Goyal and Welch (2004) showed that for models of stock returns good in-sample fit does not necessarily imply good out-of-sample performance. On the other hand, even models that fare well out-of-sample may not do so when different subsamples of a time series are considered (Stock and Watson, 2003a). Underlying these findings is the possibility that the economy - and the forecasting ability of models - may not be stable over time, as has been forcefully argued by Clements and Hendry (1998, 1999).

From the perspective of the forecaster, it is thus important to know whether a model estimated over one period can provide good forecasts over a subsequent period. The goal of this paper is to develop a formal testing framework for answering this question. Note that our question is different from asking whether the model is a good approximation of the data-generating process. Rather, our concern here is with whether a model's future performance is consistent with what's expected based on its past performance, which fundamentally hinges on the success of the model at adapting to changes in the economy. This in turn reflects a desire to mimic the environment faced by actual forecasters, where models are likely misspecified, variables are inherently difficult to forecast, and data-generating processes may be unstable, so that consistency with expected performance can be viewed as a minimal requirement that a forecasting model should satisfy.

Formally, we define a forecast breakdown as a situation in which the out-of-sample performance of a forecast model, judged by some loss function, is significantly worse than its in-sample performance. We propose a forecast breakdown test for detecting whether a forecast model broke down in the past and further suggest relating the differences between the model's out-of-sample and in-sample performance to economic factors, with the goal of predicting future breakdowns.

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" (Clements and Hendry, 1999, p. 1). We formalize the definition of a forecast breakdown by comparing the model's out-of-sample performance to its 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 that includes all observations from the beginning of the sample ("recursive scheme"). The fixed scheme presumes an interest in comparing performance before and after a specific date, whereas the rolling and recursive schemes mimic adaptive forecasting.

We illustrate how to construct an appropriate estimator for the asymptotic variance for the forecast breakdown test, that depends on the forecasting scheme and that explicitly takes into account the effect of estimation uncertainty in the model's parameters. Our test is valid under general assumptions. In particular, we allow the data to be heterogeneous (e.g., the variables in the model can have time-varying marginal distributions) 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 in which the same loss function is used for estimation and evaluation (e.g., OLS and quadratic loss), estimation uncertainty is asymptotically irrelevant and the asymptotic variance is simpler to compute.

A further contribution aims at understanding the causes of forecast breakdowns. We show that forecast breakdowns are caused by instability in the model's parameters as well as by other instabilities in the data-generating process that result in a non-constant expected loss (e.g., for a quadratic loss, changes in the variance of the disturbances). We also investigate the role of overfitting - which we define as the difference between in-sample and out-of-sample performance present in finite samples when parameter estimates are chosen to minimize the average in-sample loss - and propose a simple correction to the test statistic that eliminates its effects.

The two closest literatures to the present paper are the literature on forecast optimality testing (e.g., Mincer and Zarnowitz, 1969, Patton and Timmermann, 2003, Elliott, Komunjer and Timmermann, 2005) and the literature on structural break testing (e.g., Brown, Durbin and Evans, 1975; Andrews, 1993; Andrews and Ploberger, 1994; Dufour, Ghysels and Hall, 1994; Chu, Hornik and Kuan, 1995a, 1995b; Bai and Perron, 1998; Ghysels and Hall, 1990; Elliott and Muller, 2003; Rossi, 2005). Regarding the former, we point out that the same theory derived here can be applied to forecast optimality testing, after suitably re-definining the loss function and the null hypothesis. For example, a forecast unbiasedness test is related to a forecast breakdown test assessing whether the first moment properties of the forecast errors are consistent in-sample and out-of-sample. Our contribution to this literature is to show that the asymptotic variance estimator to be used in the forecast unbiasedness test (and, more in general, in a forecast rationality test) necessitates a correction in order for the test to have good size properties.

Regarding the structural break testing literature, although our focus 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 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 most structural break tests cannot distinguish between instability in model's parameters and instability in the distribution of the regressors (an exception is the generalized predictive tests proposed by Dufour, Ghysels and Hall, 1994; 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, only parameter instabilities of greater magnitude than those considered by the structural break testing literature cause a forecast breakdown; (3) structural break tests have greater power when instabilities are permanent, whereas the forecast breakdown test can have greater power when there are recurring instabilities that are not captured by the forecast model. A further difference with structural break tests is that they only focus on past breaks and provide no information on the likelihood of future breaks (an exception is Pesaran, Pettenuzzo and Timmermann, 2004). Instead, an innovation of our approach with useful practical implications is the possibility of predicting 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 model of inflation in the United States. Using both real-time and revised data, we find some empirical evidence in favor of a forecast breakdown in the Phillips curve. We further investigate whether monetary policy parameters would have been useful predictors of forecast breakdowns and find that inflation volatility as well as changes in the monetary policy behavior of the Fed played a key role.

2 Detecting forecast breakdowns

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 $Y_t : \Omega \rightarrow \mathbb{R}$ is the variable of interest and $X_t : \Omega \rightarrow \mathbb{R}^s$ is a vector of predictors.

We generate a sequence of $\tau-$step-ahead forecasts of $Y_{t+\tau}$ using an out-of-sample procedure, which involves dividing the sample of size T into an in-sample window of size m and an out-of-sample window of size $n = T - m - \tau + 1$. Which data constitute the in-sample window depends on the forecasting scheme. 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 $t - m + 1, \ldots, t$; and (3) a recursive forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, t$.

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. We assume that multi-step forecasts are produced by the "direct method" (that is, the model specifies the relationship between Y_t and $X_{t-\tau}$). 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 $Y_t = X'_{t-\tau}\beta + \varepsilon_t$ estimated by OLS, the parameter estimate is $\widehat{\beta}_t = \left(\sum_{s=1}^{m-\tau} X_s X_s'\right)^{-1} \sum_{s=1}^{m-\tau} X_s Y_{s+\tau}$ for the fixed scheme; $\widehat{\beta}_t = \left(\sum_{s=t-m+1}^{t-\tau} X_s X_s'\right)^{-1} \sum_{s=t-m+1}^{t-\tau} X_s Y_{s+\tau}$ for the rolling scheme and $\widehat{\beta}_t = \left(\sum_{s=1}^{t-\tau} X_s X_s'\right)^{-1} \sum_{s=1}^{t-\tau} X_s Y_{s+\tau}$ for the recursive scheme. The out-of-sample loss corresponding to the forecast at time t is $L_{t+\tau}(\widehat{\beta}_t) \equiv L(Y_{t+\tau}, X_t'\widehat{\beta}_t)$ and the corresponding in-sample losses are $L_j(\widehat{\beta}_t) \equiv L(Y_{j+\tau}, X_j'\widehat{\beta}_t)$, where $j = 1, \ldots, m-\tau$ for the fixed scheme; $j = t-m+1, \ldots, t-\tau$ for the rolling scheme and $j = 1, \ldots, t-\tau$ for the recursive scheme.

2.2 Assumptions

A1. $\{W_t\}$ is mixing with α of size $-r/(r-2)$, $r > 2$; A2. (a) $L_t(\beta)$ is measurable and twice continuously differentiable with respect to β; (b) Under H_0 in (3) below, in a neighborhood N of β^*, there exists a constant $D < \infty$ such that for all t, $\sup_{\beta \in N} \left|\partial^2 L_t(\beta)/\partial\beta\partial\beta'\right| < m_t$, for a measurable m_t such that $E(m_t) < D$. A3. 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 (nonstochastic) $k \times q$ matrix of rank k, such that $\sup_{t \geq 1} B_t^* < \infty$; $H_t^* = m^{-1} \sum_{s=1}^{m} h_s(\beta^*)$ (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 $E(h_s(\beta^*)) = 0$; A4. $\sup_{t \geq 1} E||[L_t(\beta^*), \partial L_t(\beta^*)/\partial\beta, h_t'(\beta^*)]'||^{2r} < \infty$, where $\partial L_t(\beta^*)/\partial\beta$ is $1 \times k$; A5. $T^{-1} \sum_{t=1}^{T} E(\partial L_t(\beta^*)/\partial\beta) < \infty$ for all T;
A6. $var\left(T^{-1/2} \sum_{t=1}^{T} L_t(\beta^*)\right) > 0$ for all T sufficiently large; A7. $m, n \to \infty$, $\frac{n}{m} \to \pi$, $0 < \pi < \infty$.

Comments: 1. Assumption A1 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. Assumption A2 is the same as Assumption A1 of West (1996), allowing for a number of loss functions typically used in the forecast evaluation literature. The assumption of differentiability is adopted for convenience and can be relaxed along the lines of McCracken (2000).

3. Assumption A3 is related to Assumption A2 of West (1996), permitting a number of estimation 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. The assumption also states that under the null hypothesis of no forecast breakdown the pseudo-true values of the parameters are constant (note that we do not assume correct specification of the

model under the null hypothesis).

4. Assumption A5 is a regularity condition 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 $E\left(\partial L_t(\beta^*)/\partial\beta\right) = 0$ for all t.

5. Assumption A7 shows that our asymptotics assume 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

As motivated in the introduction, 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 in-sample loss computed over the in-sample window implied by the forecasting scheme. We then consider the out-of-sample mean of the surprise losses

$$\overline{SL}_{m,n} \equiv n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t), \tag{2}$$

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

$$H_0 : E\left(n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*)\right) = 0 \text{ for all } m, n. \tag{3}$$

The forecast breakdown test statistic is

$$t_{m,n,\tau} = \sqrt{n}\overline{SL}_{m,n}/\hat{\sigma}_{m,n}, \tag{4}$$

where the expression for the asymptotic variance estimator $\hat{\sigma}^2_{m,n}$ is given in Section 2.5.

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. In the remainder of the paper, we focus on a one-sided test to reflect the assumption that a lower-than-expected loss may be desirable and thus does not constitute a forecast breakdown. In certain applications, however, it might be of interest to consider deviations of the out-of-sample loss from its expected value in either direction, in which case a two-sided test is appropriate. For example, for an investor forming a portfolio based on forecasts of stock returns, the precision of the forecast is a key determinant of how much risk exposure to accept.

Hence, if the out-of-sample forecast error variance is smaller than anticipated, this results in an opportunity cost: had the forecaster known about the lower forecast error variance, he could have chosen a different portfolio allocation.[1] The asymptotic justification for the forecast breakdown test is provided by Theorem 2 in Section 2.5.

2.4 Relationship with the literature

To see how the forecast breakdown test relates to forecast optimality tests, note that letting $L(e) = e$ in (1), where e is the forecast error, yields a test comparing the mean of out-of-sample forecast errors to the mean of in-sample model residuals. When the parameters are estimated by OLS, $\bar{L}_t(\widehat{\beta}_t) = 0$ by construction, so that the numerator of the test statistic (4) coincides with that of a forecast unbiasedness test.[2]

To see how the forecast breakdown test relates to existing tests for structural change, first note that H_0 can be rewritten as $H_0 : E\left[L_t\left(\beta^*\right)\right] =$ constant for all t, and thus one could in principle use structural break tests to test H_0. In particular, for a loss depending on the forecast errors, H_0 postulates stability of some aspect of the distribution of the model's residuals (e.g., their second moment for a quadratic loss), which relates the forecast breakdown test to residual-based tests, such as the CUSUM test (Brown *et al.,* 1975) (related to the forecast breakdown test with a recursive scheme) or the MOSUM test (Chu *et al.,* 1995b) (related to the forecast breakdown test with a rolling scheme). The main differences are that we allow for general transformations of the residuals (through $L_t\left(\cdot\right)$) and compare their in-sample and out-of-sample average properties, rather than comparing the fluctuations of the empirical process based on the cumulative (or moving) sum of residuals to the fluctuations of their limiting process.

Regarding the relationship with structural break tests based on the approach of Chow's (1960), Andrews (1993) and Andrews and Ploberger (1994), note that our fixed test could be related to a Chow's type of test, whereas our recursive test could be related to an Andrews' (1993) type of test. The approaches are similar since both split the sample in two subsamples and compare the properties of regression residuals and/or forecast errors in the two samples. The difference is that the forecast breakdown test compares regression residuals from the first subsample to forecast errors from the second subsample, which are functions of the same parameter estimate based on the first subsample. Chow's (1960) test, instead, compares regression residuals from the first subsample to regression residuals from either the second subsample (Chow's test) or the full sample (Chow's predictive test), obtained by re-estimating the model on the corresponding sample. Since it compares residuals that are functions of different parameter estimates, Chow's test will capture

[1] We thank Allan Timmermann for point out the desirability of two-sided tests in such applications.

[2] Our null hypothesis however slightly differs from that of a forecast unbiasedness test (e.g., West, 1996): we test $E\left(n^{-1}\sum_{t=m}^{T-\tau}\left(L_{t+\tau}(\beta^*) - \bar{L}_t(\beta^*)\right)\right) = 0$ rather than $E\left(n^{-1}\sum_{t=m}^{T-\tau}L_{t+\tau}(\beta^*)\right) = 0$.

not only changes in the model's parameters, but also changes in the marginal distribution of the regressors. This is a drawback of most existing structural break tests, as pointed out by Hansen (2000). The forecast breakdown test, instead, does not suffer from this problem, because it does not involve re-estimating the parameters over different subsamples.

2.5 Asymptotic variance estimators

This section shows how to construct a valid asymptotic variance estimator for the forecast breakdown test statistic (4) and provides the asymptotic justification for the forecast breakdown test.

We provide three estimators: an estimator valid under general assumptions (Theorem 2) and two estimators that are easier to compute under more restrictive conditions (Corollaries 3 and 4).

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 (2) 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 that depend on the estimator used.

Algorithm 1 (General 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* $\widetilde{L}_t \equiv L_t - T^{-1}\sum_{j=1}^{T} L_j$;[3] *(2) the* $q \times T$ *matrix of orthogonality conditions, with element* h_t *:*

$$h \equiv [\underbrace{h_1(\widehat{\beta}_m), \ldots, h_m(\widehat{\beta}_m)}_{m}, \underbrace{h_{m+1}(\widehat{\beta}_{m+1}), \ldots, h_{T-\tau}(\widehat{\beta}_{T-\tau})}_{n-1}, \underbrace{0, \ldots, 0}_{\tau}].^4$$

Let $D_{t+\tau} \equiv \partial L_{t+\tau}(\widehat{\beta}_t)/\partial\beta - \partial\bar{L}_t(\widehat{\beta}_t)/\partial\beta$, $t = m, \ldots, T-\tau$ *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 A3 that substitutes*

[3] 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.

[4] 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.

$\widehat{\beta}_t$ *for* β^*.[5] *Construct the following weights, depending on the forecasting scheme:*

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

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

$$\underbrace{1,\ldots,1}_{\tau}];$$

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

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

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

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

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

$$\begin{aligned} a_{m,j} &= \frac{1}{m+j}+\frac{1}{m+j+1}+\ldots+\frac{1}{T-\tau}; \\ b_{m,j} &= \frac{D_{m+\tau+j}B_{m+j}}{m+j}+\frac{D_{m+\tau+j+1}B_{m+j+1}}{m+j+1}+\ldots+\frac{D_TB_{T-\tau}}{T-\tau}. \end{aligned} \tag{5}$$

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

Let

$$V_T = \begin{pmatrix} V_T^{LL} & V_T^{Lh} \\ V_T^{Lh} & V_T^{hh} \end{pmatrix}, \text{ where} \tag{6}$$

$$V_T^{LL} \equiv T^{-1}\sum_{t=1}^{T}(w_t^L \widetilde{L}_t)^2 + 2T^{-1}\sum_{j=1}^{p_T} \upsilon_{T,j} \sum_{t=j}^{T} w_t^L \widetilde{L}_t w_{t-j}^L \widetilde{L}_{t-j}; \tag{7}$$

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

$$V_T^{Lh} \equiv T^{-1}\sum_{t=1}^{T} w_t^L \widetilde{L}_t h_t' w_t^{h\prime} + T^{-1}\sum_{j=1}^{p_T} \upsilon_{T,j} \sum_{t=j}^{T} \left(w_t^L \widetilde{L}_t h_{t-j}' w_{t-j}^{h\prime} + w_{t-j}^L \widetilde{L}_{t-j} h_{t-j}' w_t^{h\prime} \right), \tag{9}$$

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 (Asymptotic justification of forecast breakdown test) *(a) If* $E\left(\partial L_t(\beta^*)/\partial\beta\right)$ *is constant for all* t, $\hat{\sigma}_{m,n} = \sqrt{(T/n)V_T^{LL}}$, V_T^{LL} *given in (7). Then,* $t_{m,n,\tau} \xrightarrow{d} N(0,1)$ *under* H_0 *in (3).* [6] *(b) If* V_T *in (6) is p.d.,* $\hat{\sigma}_{m,n} = \sqrt{(T/n)\left(V_T^{LL} + V_T^{hh} + 2V_T^{Lh}\right)}$, V_T^{LL}, V_T^{hh} *and* V_T^{Lh} *given in (7)-(9). Then,* $t_{m,n,\tau} \xrightarrow{d} N(0,1)$ *under* H_0 *in (3).*

Comments: 1. Theorem 2-(a) shows that if $\partial L_t(\beta^*)/\partial\beta$ has 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 estimator needed when estimation uncertainty does not vanish asymptotically. Whether the condition for asymptotic irrelevance 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 the loss function used for estimation is the same as that used for evaluation. This is a common situation in forecasting applications, where parameters are typically estimated by OLS and forecasts are evaluated using a quadratic loss.

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.

Corollary 3 (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 (7).*

[6] 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.

Corollary 4 (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.[7] *Then,* $\hat{\sigma}_{m,n} = \sqrt{\lambda S_n^{LL}}$, *where*

Forecasting scheme	λ
Fixed	$1 + \frac{n}{m}$
Rolling, $n < m$	$1 - \frac{1}{3}\left(\frac{n}{m}\right)^2$
Rolling, $n \geq m$	$\frac{2}{3}\frac{m}{n}$
Recursive	1

(10)

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).*

Comment: As we discussed in Section 2.3, when $L(e) = e$, with e the forecast error, the forecast breakdown test is analogous to a forecast unbiasedness test. Corollary 4 gives the correct standard error for the test, provided that the condition for asymptotic irrelevance of estimation uncertainty is satisfied (i.e., if $E\left(\partial L(\beta^*)/\partial\beta\right)$ is constant, which for example is satisfied in a linear regression model with constant-mean regressors). The corollary shows that for a recursive scheme the asymptotic variance estimator does not necessitate a correction and it is simply a HAC estimator of the variance of the average out-of-sample forecast error. For the fixed and rolling schemes, instead, the estimator must be corrected.[8] In Section 5 below, we further provide the correct asymptotic variance estimator for forecast rationality tests, which was not previously available.

3 Causes of forecast breakdowns

To gain some insight into the causes of forecast breakdowns, we analyze the expectation of the numerator of the forecast breakdown test statistic (4)[9]. For simplicity, in this section we assume that parameters are estimated by maximum likelihood and let $L(\cdot)$ indicate the loss used for estimation. We further define β_t^* as $E\left(\partial L_t\left(\beta_t^*\right)/\partial\beta\right) = 0$, $t = 1, 2, \ldots, T$, and let $\overline{\Sigma}_j$ denote the relevant sample average depending on the forecasting scheme: $\overline{\Sigma}_j = t^{-1}\sum_{j=1}^{t}$ for the recursive

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

[8] For cases in which our null hypothesis coincides with the forecast unbiasedness null hypothesis, one can easily see that our estimator for the forecast unbiasedness test coincides with the estimator proposed by McCracken (2000) for the various forecasting schemes.

[9] We implicitly make the assumption that such expectation exists.

scheme, $\overline{\Sigma}_j = m^{-1}\sum_{j=t-m+1}^{t}$ for the rolling scheme, and $m^{-1}\sum_{j=1}^{m}$ for the fixed scheme. Also, let $\overline{\widehat{\beta}}_t$, $\overline{\beta}_j^*$ $\overline{\beta}_{t+\tau}^*$ denote intermediate points between $\left(\widehat{\beta}_t, \beta_t^*\right)$, $\left(\beta_t^*, \beta_j^*\right)$, $\left(\beta_t^*, \beta_{t+\tau}^*\right)$ respectively. We following proposition decomposes the expectation of the numerator of our test statistic into various components, grouped under the three categories of parameter instabilities, other instabilities and estimation uncertainty.

Proposition 5 (Causes of forecast breakdowns)

$$
\begin{aligned}
& E\left(n^{-1/2}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t)\right) \\
= \quad & \underbrace{E\left(n^{-1/2}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta_t^*)\right)}_{\text{“other instabilities”}} + \underbrace{n^{-1/2}\sum_{t=m}^{T-\tau} E\left(\frac{\partial L_{t+\tau}\left(\beta_{t+\tau}^*\right)}{\partial\beta}\right)\left(\beta_t^* - \beta_{t+\tau}^*\right)}_{\text{“parameter instabilities I”}} \\
& \underbrace{-n^{-1/2}\sum_{t=m}^{T-\tau}\overline{\sum}_j E\left(\frac{\partial L_j\left(\beta_j^*\right)}{\partial\beta}\right)\left(\beta_t^* - \beta_j^*\right)}_{\text{“parameter instabilities I”}} \qquad (11) \\
& \underbrace{+\frac{1}{2}n^{-1/2}\sum_{t=m}^{T-\tau}\left[\left(\beta_t^* - \beta_{t+\tau}^*\right)' E\left(\frac{\partial^2 L_{t+\tau}\left(\overline{\beta_{t+\tau}^*}\right)}{\partial\beta\partial\beta'}\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 L_j\left(\overline{\beta_j^*}\right)}{\partial\beta\partial\beta'}\right)\left(\beta_t^* - \beta_j^*\right)\right]}_{\text{“parameter instabilities II”}} \underbrace{+n^{-1/2}\sum_{t=m}^{T-\tau} E\left[\left(\frac{\partial L_{t+\tau}\left(\beta_t^*\right)}{\partial\beta}\right)\left(\widehat{\beta}_t - \beta_t^*\right)\right]}_{\text{“estimation uncertainty I”}} \\
& \underbrace{+n^{-1/2}\sum_{t=m}^{T-\tau} E\left\{\left[\left(\widehat{\beta}_t - \beta_t^*\right)' \frac{\partial^2\overline{L}_t(\overline{\widehat{\beta}}_t)}{\partial\beta\partial\beta'} - \frac{\partial\overline{L}_t\left(\beta_t^*\right)}{\partial\beta} + \frac{\partial\overline{L}_t\left(\beta_t^*\right)}{\partial\beta}\right]\left(\widehat{\beta}_t - \beta_t^*\right)\right\}}_{\text{“estimation uncertainty II”}} \\
& \underbrace{+\frac{1}{2}n^{-1/2}\sum_{t=m}^{T-\tau} E\left[\left(\widehat{\beta}_t - \beta_t^*\right)'\left(\frac{\partial^2 L_{t+\tau}(\overline{\widehat{\beta}}_t)}{\partial\beta\partial\beta'} - \frac{\partial^2\overline{L}_t(\overline{\widehat{\beta}}_t)}{\partial\beta\partial\beta'}\right)\left(\widehat{\beta}_t - \beta_t^*\right)\right]}_{\text{“estimation uncertainty III”}}.
\end{aligned}
$$

The component “other instabilities” captures any changes in the data-generating process - beyond parameter instabilities - that result in a non-constant expected loss. The “parameter instabilities I” component captures instabilities of the type $\beta_t^* - \beta^* = O_p\left(n^{1/2}\right)$ (which are the same instabilities considered by the structural break testing literature), whereas the “parameter instabilities II” component captures instabilities of the type $\beta_t^* - \beta^* = O_p\left(n^{1/4}\right)$. Note that when the loss functions used for estimation and for evaluation are equal the component “parameter instabilities I” disappears due to $E\left(\partial L_{t+\tau}\left(\beta_{t+\tau}^*\right)/\partial\beta\right) = 0$, implying that forecast breakdowns are in this case caused by instabilities of greater magnitude than those considered by the structural break testing literature.

Regarding the remaining components, note that when the estimation and evaluation losses are equal, the "estimation uncertainty II" component is a quadratic form, and is thus always positive. Intuitively, this is because in this case the average in-sample loss computed at the parameter estimates is minimized by construction, and is thus smaller than the expected out-of-sample loss in finite samples. We therefore interpret this component as a measure of "overfitting".

The following proposition characterizes the causes of forecast breakdowns in the special case of a linear regression model, a fixed forecasting scheme and a quadratic loss.

Proposition 6 (Special case: linear model and quadratic loss) *Consider a quadratic loss : $L(e) = L(e) = e^2$, 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 $E\left(X_t X_t'\right) \equiv J$. Suppose there are two breaks: a permanent break in the parameters and a permanent break in the variance of the disturbances, so that $\beta_t = \beta + n^{-1/4} g_1 \cdot 1\left(t \geq m\right)$ and $\sigma_t^2 = \sigma^2 + n^{-1/2} g_2 1\left(t \geq m\right)$. We have:*

$$E\left(\sqrt{n} SL_{m,n}\right) = \underbrace{g_2}_{\text{"other instabilities"}} + \underbrace{\frac{1}{2} g_1' J g_1}_{\text{"parameter instabilities II"}} \underbrace{+2\frac{\sqrt{n}}{m}\sigma^2 k}_{\text{"overfitting"}}. \tag{12}$$

From Proposition 6, we see that a forecast breakdown for a quadratic loss 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 is proportional to the number of parameters, the variance of the disturbances and the relative sizes of in-sample and out-of-sample windows (through the factor $\sqrt{n}/m$).

4 An overfitting-corrected forecast breakdown test

We propose a simple correction to the forecast breakdown test statistic (4) that eliminates the systematic difference between in-sample and out-of-sample loss that is present in finite samples when a quadratic loss is used for both estimation and evaluation. Specifically, we propose subtracting from the numerator of our test statistic an estimate of the "estimation uncertainty II" component in (11), which can be interpreted as a measure of overfitting. Using similar reasonings to those in the proof of Proposition 6, we obtain an estimate of this component in the context of a linear model with k covariance-stationary regressors, $Y_t = X_t'\beta + \varepsilon_t$. The test statistic is modified as:

$$\begin{aligned} t^c_{m,n,\tau} &= \left(\sqrt{n} SL_{m,n} - c\right) / \hat{\sigma}_{m,n}; \\ c &= 2 \cdot \gamma \cdot tr\left(\frac{X'X}{T} \cdot \widehat{V}_T^{\beta}\right), \end{aligned} \tag{13}$$

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

It is interesting to note that, if the asymptotic variance of the parameter estimates can be consistently estimated by $\widehat{V}_T^{\beta} = \sigma^2(T^{-1}X'X)^{-1}$, 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 $t_{m,n,\tau}^c$ 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 Akaike's information criterion (AIC).[10]

5 Predicting future forecast breakdowns

In Section 2.3, we proposed a test for detecting whether a forecast method broke down in the past. A question that may be of further interest to 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 positive mean in the past, we expect them to continue being positive in the future. However, it is possible that one could find additional information that predicts whether there will be a forecast breakdown. For example, the surprise losses may be persistent (in the case of a quadratic loss, for example, the presence of GARCH in the data will induce serial correlation in the surprise losses) or they may be correlated with indicators of the state of the economy.

The idea is to find variables that predict the difference between in-sample and out-of-sample performance by regressing the surprise losses on a set of explanatory variables, including, e.g., a constant, lagged surprise losses, economically meaningful variables such as business cycle leading indicators, measures of stock market volatility, interest rates etc.

Denote by Z_t the $r \times 1$ vector collecting such variables and let $\widehat{\delta}_n$ be the OLS parameter estimate obtained by estimating the predictive regression

$$SL_{t+\tau}(\widehat{\beta}_t) = Z_t'\delta + \varepsilon_{t+\tau} \tag{15}$$

over the out-of-sample period $t = m, \ldots, T-\tau$, where the regression always includes a constant. A Wald test of $H_0 : E\left(n^{-1}\sum_{t=m}^{T-\tau} Z_t SL_{t+\tau}(\beta^*)\right) = 0$ can be performed by considering the test statistic $W_{m,n,\tau} = n\widehat{\delta}_n'\hat{\Omega}_{m,n}^{-1}\widehat{\delta}_n$, with $\hat{\Omega}_{m,n}$ given in Proposition 7 below and rejecting H_0 whenever $W_{m,n,\tau} > \chi_{r,1-\alpha}^2$, where $\chi_{r,1-\alpha}^2$ is the $(1-\alpha)-th$ quantile of a χ_r^2 distribution. Proposition 7 below provides the asymptotic justification for the test.

[10] 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 $L_{t+\tau} - \overline{L}_m(1 + 2k/m)$.

To analyze the behavior of the surprise losses over time, one may further consider the plot of the fitted values $\{Z_t'\widehat{\delta}_n\}_{t=m}^{T-\tau}$ from the regression (15) together with a one-sided $(1-\alpha)\%$ confidence interval: $\left(Z_t'\widehat{\delta}_n - z_\alpha\left(Z_t'\left(\hat{\Omega}_{m,n}/n\right)Z_t\right)^{1/2}, +\infty\right)$, where z_α is the $(1-\alpha)-th$ quantile of a standard normal distribution.

Proposition 7 (Asymptotic justification of the Wald test) *Let* $Z_t = [1, z_t']'$ *and* $\widetilde{z}_t \equiv z_t - \overline{z}$, $\overline{z} \equiv \frac{1}{n}\sum_{t=m}^{T-\tau} z_t$. *Under assumptions A1-A3 and A7, further suppose that the same loss function is used for estimation and evaluation and that, under* H_0*: B1.* $\{z_t\}_{t=m}^{T-\tau}$ *and* $\{L_t(\beta^*)\}_{t=1}^{T}$ *are fourth order stationary; B2.* $\overline{z} \underset{p}{\rightarrow} E(z_t)$*; B3.* $S_{\widetilde{z}\widetilde{z}} \equiv n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{z}_t' \underset{p}{\rightarrow} \Sigma_{\widetilde{z}\widetilde{z}} \equiv E[\widetilde{z}_t\widetilde{z}_t']$ *non-singular; B4. For some* $d > 1$, $sup_{t\geq 1}E\,||z_t', L_t||^{4d} < \infty$.[11] *Let*

$$\widehat{\Omega}_{m,n} = \begin{pmatrix} 1 & -\overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & S_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\sigma}_{m,n}^2 & \Lambda S_{\widetilde{L},\widetilde{z}\widetilde{L}} \\ \Lambda S_{\widetilde{z}\widetilde{L},\widetilde{L}} & S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} \end{pmatrix} \begin{pmatrix} 1 & -\overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & S_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix}' \tag{16}$$

where $\widehat{\sigma}_{m,n}^2$ *is defined in Corollary 4,* $S_{\widetilde{z}\widetilde{z}} \equiv n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{z}_t' + n^{-1}\sum_{j=1}^{p_n}\upsilon_{n,j}\sum_{t=m+j}^{T-\tau}\left(\widetilde{z}_t\widetilde{z}_{t-j}' + \widetilde{z}_{t-j}\widetilde{z}_t'\right)$; $S_{\widetilde{z}\widetilde{L},\widetilde{L}} = n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_t^2 + n^{-1}\sum_{j=1}^{p_n}\upsilon_{n,j}\sum_{t=m+j}^{T-\tau}\left(\widetilde{z}_t\widetilde{L}_t\widetilde{L}_{t-j} + \widetilde{z}_{t-j}\widetilde{L}_{t-j}\widetilde{L}_t\right)$; $S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} = n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{z}_t'\widetilde{L}_t^2 + n^{-1}\sum_{j=1}^{p_n}\upsilon_{n,j}\sum_{t=m+j}^{T-\tau}\left(\widetilde{z}_t\widetilde{L}_t\widetilde{L}_{t-j}\widetilde{z}_{t-j}' + \widetilde{z}_{t-j}\widetilde{L}_{t-j}\widetilde{L}_t\widetilde{z}_t'\right)$, *where* $\widetilde{L}_t$ *is as in Algorithm 1 and* p_n *and* $\upsilon_{n,j}$ *are as in Proposition 4;* $\Lambda = \left[\pi^{-1}\ln(1+\pi)\right]$ *for the recursive scheme,* $\Lambda = 1 - \pi/2$ *for the rolling scheme when* $n \leq m$, $\Lambda = (2\pi)^{-1}$ *for the rolling scheme when* $n > m$, *and* $\Lambda = 1$ *for the fixed scheme. Then* $W_{m,n,\tau} \overset{d}{\rightarrow} \chi_r^2$ *under* $H_0 : E\left(n^{-1}\sum_{t=m}^{T-\tau} Z_t S L_{t+\tau}(\beta^*)\right) = 0$.[12]

Corollary 8 (Asymptotic justification of the Wald test under conditional homoskedasticity) *Under the assumptions of Proposition 7, further suppose that, under* H_0, $E\left(\widetilde{L}_t(\beta^*)\,\widetilde{L}_{t-j}(\beta^*)\,|\,\{z_t\}_{t=m}^{T-\tau}\right) \equiv \gamma_j^{LL}$*. Then:*

$$\widehat{\Omega}_{m,n} = \begin{pmatrix} \sigma_{m,n}^2 + \overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1}S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}}S_{\widetilde{z}\widetilde{z}}^{-1}\overline{z} & -\overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1}S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}}S_{\widetilde{z}\widetilde{z}}^{-1} \\ -S_{\widetilde{z}\widetilde{z}}^{-1}S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}}S_{\widetilde{z}\widetilde{z}}^{-1}\overline{z} & S_{\widetilde{z}\widetilde{z}}^{-1}S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}}S_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix}. \tag{17}$$

Comments: 1. Under the assumption of conditional homoskedasticity, the asymptotic variance of the parameter estimates other than the intercept does not require any correction depending on the forecasting scheme.

2. Note that equation (15) can be interpreted as a forecast rationality regression, by letting $L(e) = e$ in (1), where e is the forecast error, and when parameters are estimated by OLS, so that $\bar{L}_t(\widehat{\beta}_t) = 0$. Proposition 7 thus provides the appropriate asymptotic variance estimator for

[11] The latter assumption ensures that third and fourth order cumulants are finite. This assumption is trivially satisfied if the variables are Normal, and it is a standard assumption – see Brillinger (1981, p. 26, Assumption 2.6.1). Also, note that fourth order stationarity implies covariance stationarity.

[12] Matlab code to implement the Wald test under the assumptions of Proposition 7 is available at http:\\www.econ.duke.edu\~brossi. Matlab code to implement the Wald test under the more general assumption of heterogeneity of the losses is further available at http:\\www.econ.ucla.edu\giacomin.

the forecast rationality test, and shows that a correction is only required for the standard error of the intercept (which is the same correction that applies to the forecast unbiasedness test; see the comment after Corollary 4). The same remarks hold for a Mincer and Zarnowitz (1969) regression, where the regressor is the forecast, provided a two-step estimator is used for the standard errors to account for the generated-regressor problem.

6 Implications of forecast breakdowns

A natural question that arises if a forecast breakdown is detected or predicted is whether the forecast model should be changed or not. In general, the answer to this question depends on the type of forecast (point, interval, density) and on the type of loss function (symmetric or asymmetric). For example, when the forecast is a point forecast and the loss function is symmetric, finding a forecast breakdown does not necessarily imply that the model should be changed. The reason is that the forecast breakdown could be caused by instabilities - such as increases in the variance of the disturbances - that do not affect the optimal forecast (for a symmetric loss, the optimal point forecast does not depend on the variance, unlike for an asymmetric loss, as shown by Christoffersen and Diebold, 1997). Since the forecast breakdown test cannot distinguish among the different types of instabilities, the finding of a forecast breakdown does not in this case suggest a course of action. However, when the loss is asymmetric or when the forecaster is interested in accompanying the point forecast with some measure of its uncertainty (e.g., an interval or a density forecast), then the finding of a forecast breakdown indicates unreliability of the forecast, regardless of its cause.

7 Monte Carlo evidence

We analyze 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).[13] We further compare the size properties of commonly used forecast rationality tests and those of our corrected forecast rationality test (see comment 2 after Corollary 8).

7.1 Size properties of forecast breakdown tests

We investigate the size of the forecast breakdown test, in particular with regards to its robustness to the presence of instability in the marginal distribution of the regressors and to the presence of

[13] 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.

conditionally heteroskedastic disturbances. We let the data-generating process (DGP) be:

$$\begin{aligned} Y_t &= 2.73 - 0.44u_{t-1} + \sigma_t\varepsilon_t, \\ \sigma_t^2 &= 1 + \alpha * \varepsilon_{t-1}^2,\ \varepsilon_t \sim i.i.d.N(0,1), \end{aligned} \tag{18}$$

and 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 (MC2), inspired by our empirical application to the Phillips curve model of U.S. inflation, lets u_t be monthly U.S. unemployment and $\alpha = .5$.[14] The DGP specification and parameters are from Staiger, Stock and Watson (1997). We use an actual time series for unemployment 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 (18). In MC2, we use the last T data in the u_t time series, up to 2005:8. We estimate 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 test proposed by Elliott and Muller (2003) (denoted EM) and our forecast breakdown test for the three forecasting schemes, using either the general asymptotic variance estimator of Corollary 3 ($t_{m,n,\tau}$) or the estimator of Corollary 4 ($t^{\text{stat}}_{m,n,\tau}$) (the truncation lags for the HAC estimators are $p_T = p_n = 0$ in MC1 and $p_T = p_n = n^{1/3}$ in MC2). Table 1(a) contains the rejection frequencies of the tests for various (m, n) pairs.

[TABLES 1(a) AND 1(b) HERE]

The forecast breakdown test has good size properties for large in-sample and out-of-sample sizes $(m, n \geq 100)$. The $t^{\text{stat}}_{m,n,\tau}$ test is 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, Table 1(b) reports the rejection frequencies of the overfitting-corrected test of Section 4, using the simple correction (14) in both MC1 and MC2. As expected, the use of the overfitting correction substantially improves the size properties of the test. The overfitting-corrected test is well-sized in all cases except for the fixed scheme when the in-sample size is small ($m = 50$).

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. The EM test, instead, has correct size when the

[14]The unemployment series is the seasonally adjusted civilian unemployment rate from FRED II. The results are robust to higher values of α, even close to one.

regressor is i.i.d., but erroneously detects instability in model's parameters when the regressor is the actual time series of U.S. unemployment (in this case, the EM rejects 100% of the time).

7.2 Size properties of forecast rationality tests

Finally, we document size distortions of conventional forecast rationality tests and the good size properties of a test based on the "corrected" variance estimator of Proposition 7. The DGP is: $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$, where $\beta_0 = \beta_1 = 0$, $\varepsilon_t \sim i.i.d.N(0,1)$, $X_t \sim i.i.d.N(0,1)$ and forecasts are based on a model with a constant and X_t estimated using the various forecasting schemes. The forecast rationality test is performed by estimating the regression: $e_{t+1} = \delta_0 + \delta_1 Z_t + u_t$, where e_{t+1} is the estimated out-of-sample forecast error and $Z_t \sim i.i.d.N(0,1)$ independent of X_t. Table 2 reports rejection frequencies of forecast rationality tests using conventional OLS standard errors and of the corresponding test using the variance estimator (17) with $L(e) = e$. The nominal size is 5%. As the columns labeled "uncorrected" in Table 2 show, both a standard t-test on δ_0 (t_{δ_0}) and a Wald test on both δ_0 and δ_1 (W) have considerable size distortions except for the recursive scheme, whereas a t-test on δ_1 (t_{δ_1}) has no size distortions for any scheme. The columns labeled "corrected" show instead that our correction performs very well in all cases.

[TABLE 2 HERE]

7.3 Power properties

In this section we consider various sources of forecast breakdowns and analyze the power of the tests considered in Section 7.1 and of a forecast unbiasedness test for the recursive scheme forecasts (UNB). In all designs, we estimate the model $Y_t = \alpha + e_t$ by OLS and consider a quadratic and a linex loss for evaluation. The total sample size T and the in-sample size m for the forecast breakdown and the unbiasedness tests are specified in each design. In all cases, m is set at the time of the first break, which represents the "worst-case scenario" from the perspective of a forecaster.

Design 1: Changes in mean. We consider either one-time or recurring changes in mean. The first corresponds to a single structural break in mean

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

We let $(T, m) = (300, 150)$. In the recurring change DGP, we let $Y_t = \mu_t + \varepsilon_t$, where μ_t switches between $-\beta_A$ and β_A every 50 periods and let $(T, m) = (600, 50)$.

Design 2: Changes in variance. Again, we consider both one-time and recurring changes. The one-time change DGP is

$$Y_t = \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N\left(0, \sigma_t^2\right) \tag{20}$$

where $\sigma_t^2 = 1 + \beta_A \cdot 1(t > T/2)$. and. We choose $(T, m) = (300, 150)$. In the recurring changes case, we let σ_t^2 switch between 1 and $(1 + \beta_A)$ every 50 periods, and let $(T, m) = (600, 50)$.

Design 3: Other DGP changes. Here we assume that the conditional mean undergoes a one-time change but the two specifications are not nested, so that structural break tests are not optimal in this context. We let

$$\begin{aligned} Y_t &= \beta_A \cdot 1(t \leq T/4) - 3\beta_A \cdot 1(T/4 < t \leq T/2) + X_t \cdot 1(t > T/2) + \varepsilon_t, \qquad (21) \\ X_t &= .6X_{t-1} + \eta_t, \ \varepsilon_t, \eta_t \sim i.i.d.N(0,1) \text{ independent.} \end{aligned}$$

We consider $(T, m) = (400, 100)$.

[FIGURE 1 HERE]

For all designs, we obtain power curves by letting β_A vary between 0 and 2 and considering 5000 Monte Carlo replications. Figure 1(a) shows that the forecast breakdown test has power against changes in mean. In the case of a permanent break in mean (upper left panel), the forecast breakdown test has lower power than both the EM and the UNB tests, but its power improves when the losses used for estimation and evaluation differ (upper right panel). In the case of recurring changes in mean (lower panels), the forecast breakdown test with a rolling scheme has the highest power. When the permanent change in DGP is as in Design 3 (Figure 1(c), right panel), the power loss of the forecast breakdown relative to the EM and UNB tests is substantially lower. Figure 1(b) shows that the forecast breakdown test is the only test that has power against changes in variance. The one-sided nature of the test implies that only increases in variance (Figure 1(b), upper panels) or, to a lesser extent, recurring changes in variance (Figure 1(b), lower panels) can cause forecast breakdowns. Decreases in variance, obtained by substituting β_A with $-\beta_A$ in design 2, instead do not cause forecast breakdowns, as can be seen from the left panel of Figure 1(c).

8 The Phillips curve and inflation forecast breakdowns

The Phillips curve as a forecasting model of inflation 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. The forecasting ability of the Phillips curve as well as its stability have been investigated in a number of works, including Staiger, Stock and Watson (1997), Stock and Watson (1999) and Fisher, Liu and Zhou (2002). The latter, in particular, conclude that the forecasting ability of the Phillips curve depends upon the period: the Phillips curve appears to forecast well one year ahead during the 1977-1984 period but not during the 1993-2000 period. Thus, as an empirical application of the methods proposed in this paper, we investigate the robustness of the Phillips curve to forecast breakdowns.

Following Stock and Watson (1999), let $\pi_t^\tau = (1200/\tau)\ln(P_t/P_{t-\tau})$ 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:

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

where θ_0 implicitly embodies a time-invariant NAIRU, and $\theta_1(L)$ and $\theta_2(L)$ are lag polynomials with q_u and q_π lags, respectively.

When analyzing whether unemployment was a useful predictor for inflation, it is important to assess its predictive ability using data that were available to the policymakers at that time. For example, Orphanides (2001) and Ghysels, Swanson and Callan (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 this reason, we use real-time data from the Federal Reserve Bank of Philadelphia database. The data are discussed in Croushore and Stark (2001). Since the real-time series of consumer prices from the same data set is available only from the 1994 vintage, for this series we use 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). The data are from 1961:1 (with a first vintage in 1978:2) until 2001:12. Due to the data limitations, we restrict estimation from 1978:2 until 2001:12, using quarterly vintages.[15]

The first column of Table 3 reports the p-values of the forecast breakdown test of Section 2.3 for a quadratic loss and a rolling scheme with $m = 60$ (so that the one-step ahead forecasts begin in 1993:1, corresponding to the change in monetary policy identified in Fisher et al., 2002). We consider forecast horizons $\tau = 3$ and $\tau = 12$ months and several choices of q_u and q_π. The row labeled "*BIC*" reports results for the case in which the lag length is determined by the Bayesian Information Criterion (BIC) (assuming that all regressors have the same number of lags). The table shows strong evidence of a forecast breakdown at the one year horizon when using real-time data, whereas there is little evidence of forecast breakdowns at shorter horizons. Because of small sample concerns associated with real-time data, we repeat the above exercise using revised monthly

[15] 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 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.

data. We consider the most recent observations collected by the Philadelphia Fed (2004:8) for both seasonally unadjusted CPI and unemployment. The largest available sample for both variables is from 1948:1 until 2004:6. The second column in Table 3 shows that the forecast breakdown test finds some evidence of a forecast breakdown at the one month horizon, but not at longer horizons.

[TABLE 3 HERE]

Given the evidence in favor of forecast breakdowns in the Phillips curve, we next 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 changes in the forecasting ability of the Phillips curve. Thus, we construct a forecasting model that relates the surprise losses to inflation volatility and to a measure of changes in the monetary policy behavior of the Fed. We estimate inflation volatility ($\widehat{\sigma}^2_{\pi,t}$) as the sample variance of the change in the annual inflation over a rolling window of size 241.[16] To measure changes in the monetary policy behavior of the Fed, we consider rolling two-step efficient GMM estimates (with 2SLS in the first step) of the coefficients of the Federal Fund Rate (FFR) reaction function to the output gap and to the deviation of inflation from its target proposed by Clarida, Gali and Gertler (2000), given by

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

with r_t the nominal FFR; $\pi_{t,k}$ the annualized percentage change in the price level between t and $t+k$; $x_{t,q}$ the average output gap between t and $t+q$, defined as minus the percentage deviation of actual unemployment from its target (a fitted quadratic function of time); and $\Im_t$ the information set at time t. As in Clarida et al. (2000), we let $\rho\,(L) \equiv \rho_1 + \rho_2 L$, rr^* be the average FFR over the estimation window, and we choose as instruments a constant and four lags of the following variables: inflation, output gap, FFR, commodity price inflation, M2 growth rate, spread between the long-term bond rate and the three-month Treasury Bill rate.[17] k and q are set at 1 quarter. Our measures of changes in monetary policy behavior are sequences of estimates of β, γ and $\rho \equiv \rho(1)$ in (23) over a rolling window of size 241. Even though our database is different from that of Clarida et al. (2000), our parameter estimates - which we do not report to conserve space - are similar.

We next investigate whether the estimates of the FFR reaction function coefficients and inflation volatility are useful predictors of inflation forecast breakdowns. Table 4 shows estimates of the

[16] I.e. we use lagged values of the sample variance of $(\pi^{\tau}_{t+\tau} - \pi_t)$ as a potential predictor.

[17] 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 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. The abuse of notation in denoting the degree of inflation aversion by β is to make our notation consistent with that of Clarida et al. (2000).

coefficients in the following equation:

$$SL_{t+\tau} = \delta_0 + z_t'\delta_1 + \varepsilon_{t+\tau} \tag{24}$$

where z_t is either $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$ (the rolling estimates of the parameters in (23)), or $\widehat{\sigma}^2_{\pi,t}$, and $\tau = 1, 3, 12$ months. The table reports estimates of δ_1 and (in parentheses) the p-values associated with testing whether δ_1 equals zero.[18] It is clear that the degree of inflation targeting smoothing operated by the central bank ($\widehat{\rho}_t$) and the degree of inflation volatility ($\widehat{\sigma}^2_{\pi,t}$) explain the behavior of the surprise losses 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. We also estimate (24) with $z_t = (\widehat{\beta}_t, \widehat{\gamma}_t, \widehat{\rho}_t)$ and find strong evidence of joint significance at horizons of one and twelve months (last column of Table 4). To conclude, Figure 2 plots the sequence of surprise losses $\widehat{SL}_{t+12}$ along with its one-sided 95% confidence band, and shows empirical evidence of forecast breakdowns during the Volker era (1979:3-1987:7) but not during the Greenspan era (1987:7 onwards).

[TABLE 4 AND FIGURE 2 HERE]

9 Conclusion

This paper proposed a method for detecting and predicting forecast breakdowns, defined as a situation in which the out-of-sample performance of a forecast model is significantly worse than its in-sample performance. Unlike the literature evaluating a forecasting model from the perspective of whether it produces optimal forecasts, we focus on whether the model's forecast performance - measured by a general loss function - is consistent with expectations based on the model's earlier fit. The analysis of the possible causes of forecast breakdowns reveals the prime role played by instabilities in the data-generating process in causing forecast breakdowns, thus establishing a link between this paper and the structural break testing literature. Among the differences, we note that our forecast breakdown test is valid under more general assumptions, for example permitting the model to be misspecified and the regressors to be unstable, arguably a closer representation of the environment faced by actual forecasters. Further, our testing framework allows the forecaster to predict the likelihood that the forecast model will break down at a future date, a question that is not typically addressed by the structural break testing literature.

While our method is a first step towards assessing how well a forecasting model adapts to changes in the economy, an important question that we touched upon but that deserves further investigation is what to do in case a forecast breakdown is detected or predicted. We leave this avenue of research for future work.

[18] The test statistic is implemented with a Newey and West (1987) HAC estimator with a bandwidth equal to $n^{1/3}$ and the p-values are calculated from (8).

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

Brillinger (1981), "Time Series: Data Analysis and Theory", San Francisco: Holden Day.

Brown, R.L., J. Durbin and J.M. Evans (1975), "Techniques for Testing Constancy of Regression Relationships Over Time", *Journal of the Royal Statistical Society,* B, 37, 149-163.

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

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

Christoffersen P. F., Diebold, F. X. (1997): "Optimal Prediction under Asymmetric Loss', *Econometric Theory*, 13, 808-817.

Chow, G.C. (1960), "Tests of Equality between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 591-605.

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

Chu, C.-S. J., K. Hornik and C.-M. Kuan (1995b), "MOSUM Tests for Parameter Constancy", *Biometrika,* 82, 603-617.

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.

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., I. Komunjer and A. Timmermann (2005), "Estimation and Testing of Forecast Rationality under Flexible Loss", *Review of Economic Studies*, 72, 1107–1125.

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

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.

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

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

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

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.

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.

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

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.

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

1084.

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

Appendix. Proofs

Notation 9 *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 *defined in Algorithm 1;* $D_{t+\tau}^* \equiv \partial SL_{t+\tau}(\beta^*)/\partial\beta$, $t = m, \ldots, T-\tau$; $\widetilde{L}_t^* \equiv L_t^* - E(L_t^*)$; $\tilde{D}_{t+\tau}^* = D_{t+\tau}^* - E(D_{t+\tau}^*)$; $\widetilde{\partial L}_t^* = \partial L_t^* - E(\partial L_t^*)$. *For a matrix A,* $|A| = \max_{i,j}|a_{ij}|$. *Limits are for* $m, n \to \infty$.

Lemma 10 *(a)* $R_1 \equiv n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{D}_{t+\tau}^* B_t^* H_t^* = o_p(1)$;
(b) $R_2 \equiv .5n^{-1/2}\sum_{t=m}^{T-\tau}\left(\widehat{\beta}_t - \beta^*\right)'\left(\partial^2 SL_{t+\tau}(\overline{\beta_t^*})/\partial\beta\partial\beta'\right)\left(\widehat{\beta}_t - \beta^*\right) = o_p(1)$, *where* $\overline{\beta_t^*}$ *is an intermediate point between* $\widehat{\beta}_t$ *and* β^*.

Proof of Lemma 10. *(a)* We focus for simplicity on the recursive scheme. The proofs for the fixed and rolling schemes are similar and are available upon request. Direct calculations show that $R_1 = n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*$, 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 (i) $\left|E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)\right| \xrightarrow{p} 0$ and (ii) $E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)^2 \xrightarrow{p} 0$ from which the result follows by Chebyshev's inequality.

(i) 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 A3 and A7. 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 $\left|E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)\right| \to 0$.

(ii) From (i), $E\left(n^{-1/2}\sum_{t=1}^{T}\tilde{w}_t^h h_t^*\right)^2 = E\left(n^{-1/2}\sum_{t=1}^{T}\left[T^{-1}\sum_{j=1}^{T}\widetilde{\partial L}_j^* P_{t,j}\right] h_t^*\right)^2$. 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 (i), $\sum_{j=0}^{\infty} j\alpha(j)^{1-1/2r} < \infty$, 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 A1 and A4, as shown by Andrews (1991).

(b) For some a, $0<a<.5$, C a positive constant, m_t defined in assumption A2(b) and denoting by $\overline{m}_t$ the mean of the m_t's over the relevant in-sample window at time t, we have

$$\begin{aligned}
R_2 &= \left|.5n^{-1/2}\sum_{t=m}^{T-\tau} t^{1-a}\left(\widehat{\beta}_t-\beta^*\right)'\left(t^{1-a}\frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'}\right)\left(\widehat{\beta}_t-\beta^*\right)\right| \\
&\leq C\sup_{m\leq t\leq T-\tau}|t^{.5-.5a}\left(\widehat{\beta}_t-\beta^*\right)|^2 n^{-1/2}\sum_{t=m}^{T-\tau} t^{a-1}\left|\frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'}\right| \\
&\leq C\sup_{m\leq t\leq T-\tau}|t^{.5-.5a}\left(\widehat{\beta}_t-\beta^*\right)|^2 n^{-1/2}\sum_{t=m}^{T-\tau} t^{a-1}\left(\left|\frac{\partial^2 L_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'}\right|+\left|\frac{\partial^2 \overline{L}_{t}(\overline{\beta_t^*})}{\partial\beta\partial\beta'}\right|\right) \\
&\leq C\sup_{m\leq t\leq T-\tau}|t^{.5-.5a}\left(\widehat{\beta}_t-\beta^*\right)|^2 n^{-1/2}\sum_{t=m}^{T-\tau} t^{a-1}\left(m_{t+\tau}+\overline{m}_t\right)=o_p(1)
\end{aligned}$$

by Lemmas A1(a) and A3(b) of West (1996), Assumption A2(b) and Markov's inequality. ■

Lemma 11 $\frac{T}{n}V_T^{LL*}=var\left(n^{-1/2}\sum_{t=1}^{T}w_t^L\widetilde{L}_t^*\right)>0$ *for all* T *sufficiently large.*

Proof of Lemma 11. We prove Lemma 11 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 $A_1=-n^{-1/2}a_{m,0}(\widetilde{L}_1^*+\ldots+\widetilde{L}_m^*)$; $A_2=-n^{-1/2}\left(a_{m,1}\widetilde{L}_{m+1}^*+\ldots+a_{m,\tau-1}\widetilde{L}_{m+\tau-1}^*\right)$; $A_3=n^{-1/2}\left[(1-a_{m,\tau})\,\widetilde{L}_{m+\tau}^*+\ldots+(1-a_{m,n-1})\,\widetilde{L}_{T-\tau}^*\right]$; $A_4=n^{-1/2}\left(\widetilde{L}_{T-\tau+1}+\ldots+\widetilde{L}_T\right)$. 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 A6. ■

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

$$
\begin{aligned}
& n^{1/2}\left[n^{-1}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t) - E\left(n^{-1}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*)\right)\right] \qquad (26)\\
= \; & n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}(\beta^*) - E\left(SL_{t+\tau}(\beta^*)\right)\right] + n^{-1/2}\sum_{t=m}^{T-\tau}\frac{\partial SL_{t+\tau}(\beta^*)}{\partial\beta}\left(\widehat{\beta}_t-\beta^*\right)\\
& +\frac{1}{2}n^{-1/2}\sum_{t=m}^{T-\tau}\left(\widehat{\beta}_t-\beta^*\right)'\frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'}\left(\widehat{\beta}_t-\beta^*\right)\\
. \;= \; & n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}(\beta^*) - E\left(SL_{t+\tau}(\beta^*)\right)\right] + n^{-1/2}\sum_{t=m}^{T-\tau}E\left(D_{t+\tau}^*\right)B_t^*H_t^* +\\
& n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{D}_{t+\tau}^*B_t^*H_t^* + \frac{1}{2}n^{-1/2}\sum_{t=m}^{T-\tau}\left(\widehat{\beta}_t-\beta^*\right)'\frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'}\left(\widehat{\beta}_t-\beta^*\right)\\
= \; & n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}(\beta^*) - E\left(SL_{t+\tau}(\beta^*)\right)\right] + n^{-1/2}\sum_{t=m}^{T-\tau}E\left(D_{t+\tau}^*\right)B_t^*H_t^* + o_p(1)
\end{aligned}
$$

where $\overline{\beta_t^*}$ is some intermediate point between $\widehat{\beta}_t$ and β^* and where we have used assumption A3 and Lemma 10. We show that, under H_0,

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

with V_T defined in (6), from which the theorem follows. Direct calculations show that $\left(\frac{T}{n}V_T\right)^{-1/2} n^{-1/2}\left[\sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*), \sum_{t=m}^{T-\tau}E\left(D_{t+\tau}^*\right)B_t^*H_t^*\right]' = V_T^{-1/2}T^{-1/2}\left[\sum_{t=1}^{T} w_t^L L_t^*, \sum_{t=1}^{T} w_t^{h*}h_t^*\right]'$, where w_t^{h*} equals w^h defined in 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 A4 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 A7; 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 11. 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 10-(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 A5, by $P_{t,j}$ having bounded components (as shown in the proof of Lemma 10-(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 A4. 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) $E\left(D_{t+\tau}^*\right) = E\left(\partial SL_{t+\tau}(\beta^*)/\partial\beta\right) = E\left(\partial L_{t+\tau}(\beta^*)/\partial\beta\right) - E\left(\partial \overline{L}_t(\beta^*)/\partial\beta\right) = 0$, and thus expression (26) reduces to $n^{-1/2} \sum_{t=m}^{T-\tau} \left[SL_{t+\tau}(\beta^*) - E\left(SL_{t+\tau}(\beta^*)\right)\right] + o_p(1)$. The result then follows from reasonings analogous to those in part (b) above and from Lemma 11. ■

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

Lemma 12 *For $a_{m,j}$ as defined in (5), we have: (i)* $a_{m,j} \simeq \ln(m+n-1/(m+j))$; *(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 12. *(i)* $a_{m,j} = \sum_{i=j}^{n-1} (m+i)^{-1} \simeq \int_j^{n-1} (m+x)^{-1} dx = \ln(m+n-1/(m+j))$; *(ii)* $n^{-1} \sum_{j=\tau}^{n-1} a_{m,j} \simeq n^{-1} \int_\tau^{n-1} \ln(m+n-1/(m+x))\, dx =$
$n^{-1}\left[n - 1 - \tau - (m-\tau)\ln(m+n-1/(m+\tau))\right] \to 1 - \pi^{-1}\ln(1+\pi)$;
(iii) $n^{-1} \sum_{j=\tau}^{n-1} a_{m,j}^2 \simeq n^{-1} \int_\tau^{n-1} \ln^2(m+n-1/(m+x))\, dx =$
$n^{-1}\left[2(n-\tau) - 2(m+\tau)\ln(m+n-1/(m+\tau)) - (m+\tau)\ln^2(m+n-1/(m+\tau))\right]$
$\to 2\left[1 - \pi^{-1}\ln(1+\pi)\right] - \pi^{-1}\ln(1+\pi)$. ■

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 11, $var(n^{-1/2} \sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \sum_{i=1}^{4} var(A_i)$. 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 12-*(i)*. Further, $var(A_2) = n^{-1} var\left(a_{m,1}\widetilde{L}_{m+1}^* + \ldots + a_{m,\tau-1}\widetilde{L}_{m+\tau-1}^*\right) \to 0$ 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 12, $n^{-1} d_0 = (n-\tau)/n - 2n^{-1} \sum_{j=\tau}^{n-1} a_{m,j} + n^{-1} \sum_{j=\tau}^{n-1} a_{m,j}^2 \to 1 - \pi^{-1}\ln(1+\pi)$, and thus lim $var(A_3) = \left[1 - \pi^{-1}\ln(1+\pi)\right] \sum_{j=-\infty}^{\infty} \Gamma_j$. Finally, $var(A_4) = n^{-1} var(\widetilde{L}_{T-\tau+1} + \ldots + \widetilde{L}_T) \to 0$

since τ is fixed. In sum, we have $var(n^{-1/2}\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \sum_{j=-\infty}^{\infty}\Gamma_j$ and thus $\lambda^* = 1$. The proofs for the fixed and rolling schemes follow from similar reasonings. ■

Proof of Proposition 5. A mean value expansion of $n^{-1/2}\sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t) \equiv n^{-1/2}\sum_{t=m}^{T-\tau}\left[L_{t+\tau}\left(\widehat{\beta}_t\right) - \overline{L}_t\left(\widehat{\beta}_t\right)\right]$ around β_t^* gives:

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

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

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

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

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

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

Substituting (30) into (27) gives:

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

Note that, since $0 = \partial\overline{L}_t\left(\widehat{\beta}_t\right)/\partial\beta = \partial\overline{L}_t\left(\beta_t^*\right)/\partial\beta + \left(\partial^2\overline{L}_t(\overline{\widehat{\beta}}_t)/\partial\beta\partial\beta'\right)\left(\widehat{\beta}_t - \beta_t^*\right)$, then $\partial L_{t+\tau}\left(\beta_t^*\right)/\partial\beta - \partial\overline{L}_t\left(\beta_t^*\right)/\partial\beta = \partial L_{t+\tau}\left(\beta_t^*\right)/\partial\beta\ -\partial\left(\overline{L}_t\left(\beta_t^*\right) - \overline{L}_t\left(\beta_t^*\right)\right)/\partial\beta +$ $\left(\widehat{\beta}_t - \beta_t^*\right)'\left(\partial^2\overline{L}_t(\overline{\widehat{\beta}}_t)/\partial\beta\partial\beta'\right)$. Therefore, by taking expectations of (31), we have (11). ■

Proof of Proposition 6. Since $E\left(\partial L_t\left(\beta_t\right)/\partial\beta - \partial L_t\left(\beta_t\right)/\partial\beta\right) = 0\ \forall t$, the "parameter instabilities I" component is zero. The "parameter instabilities II" component is $(1/2)n^{-1/2}\sum_{t=m}^{T-\tau} E\left[\left(\beta - \left(\beta + n^{-1/4}g_1\right)\right)' J\left(\beta - \left(\beta + n^{-1/4}g_1\right)\right)\right] = (1/2)g_1' J g_1$ and the "other instabilities" component is g_2. Since $\partial L_{t+\tau}\left(\beta_t\right)/\partial\beta = -2X_{t+\tau}\left(Y_{t+\tau} - X_{t+\tau}'\beta_t\right)$ is uncorrelated with $\left(\widehat{\beta}_t - \beta_t\right)$, the "estimation uncertainty I" component is zero. Since $E\left(\partial^2 L_j\left(\beta\right)/\partial\beta\partial\beta'\right) =$ $E\left(\partial^2 L_j\left(\beta\right)/\partial\beta\partial\beta'\right) = 2J\ \forall j$, the "estimation uncertainty III" component in (11) is also zero. Finally, the "estimation uncertainty II" component equals $\sqrt{n}E\left(\widehat{\beta}_m - \beta\right)'\left(2m^{-1}\sum_{s=1}^m X_s X_s'\right)\left(\widehat{\beta}_m - \beta\right) =$ $2(\sqrt{n}/m)E\left(m^{-1/2}\sum_{s=1}^m X_s\varepsilon_s\right)'\left(m^{-1}\sum_{s=1}^m X_s X_s'\right)^{-1}\left(m^{-1/2}\sum_{s=1}^m X_s\varepsilon_s\right) \underset{p}{\to} 2\left(\sqrt{n}/m\right)\sigma^2 E\left(\chi_k^2\right) =$ $2\left(\sqrt{n}/m\right)\sigma^2 k$. ■

Proof of Proposition 7. We focus on the recursive scheme and, for simplicity, assume that z_t is scalar. Let $\widehat{\delta}_n^* \equiv \begin{pmatrix} 1 & -\overline{z}' S_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & S_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}\begin{pmatrix} \overline{SL}_{m,n}^* \\ \frac{1}{n}\sum_{t=m}^{T-\tau}\widetilde{z}_t(SL_{t+\tau}^* - \overline{SL}_{m,n}^*)\end{pmatrix}$. Given assumptions B2 and B3, $var\left(\sqrt{n}\widehat{\delta}_n^*\right) = \begin{pmatrix} 1 & -E\left(z_t\right)'\Sigma_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix} var\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\left(SL_{t+\tau}^* - E\left(SL_{t+\tau}^*\right)\right) \\ \frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^* - E\left(SL_{t+\tau}^*\right)\right)\end{pmatrix}\begin{pmatrix} 1 & -E\left(z_t\right)'\Sigma_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}'$ $+o_p(1)$. As shown in Corollary 4, the upper diagonal element $\sigma_{m,n}^2$ of $var\left(\sqrt{n}\widehat{\delta}_n^*\right)$ can be consistently estimated under H_0 by $\hat{\sigma}_{m,n}^2$, given in the same corollary. Letting $\widetilde{L}_t \equiv L_t - E\left(L_t\right)$, the remaining elements are as follows: (I) $var\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^* - E\left(SL_{t+\tau}^*\right)\right)\right) = var\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*\right)$

$+var\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(\frac{1}{t}\sum_{j=1}^{t}\widetilde{L}_j^*\right)\right)-$
$2cov\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*,\ \frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(\frac{1}{t}\sum_{j=1}^{t}\widetilde{L}_j^*\right)\right)$. Each element of the second term goes to zero by arguments similar to those in Lemma A4(a) of West (1994) under Assumption B1. The typical element of the third term is $\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\sum_{j=1}^{t}\frac{1}{s}\widetilde{\kappa}_4(j,s,\tau)$, where $\widetilde{\kappa}_4(j,s,\tau)\equiv E\left[\widetilde{z}_{it}\widetilde{L}_{t+\tau}^*\widetilde{L}_j^*\widetilde{z}_{is}\right]$ is the fourth order cumulant, and τ is fixed. Therefore, $\left|\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\sum_{j=1}^{t}\frac{1}{s}\widetilde{\kappa}_4(j,s,\tau)\right|\leq$ $\frac{1}{n}\sum_{t=m}^{T-\tau}\frac{1}{m}\sum_{s=m}^{T-\tau}\sum_{j=1}^{t}|\widetilde{\kappa}_4(j,s,\tau)|\leq\frac{1}{m}\sum_{s=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}|\widetilde{\kappa}_4(j,s,\tau)|\underset{p}{\rightarrow}0$ by Assumptions A7 and B4. Hence, $var\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)\underset{p}{\rightarrow}\sum_{j=-\infty}^{\infty}E\left(\widetilde{z}_t\widetilde{L}_{t+\tau}^*\widetilde{L}_{t+\tau-j}^*\widetilde{z}_{t-j}\right)$.

(II) $cov\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right),\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)=A_{1n}+A_{2n}-A_{3n}-A_{4n}$, where
$A_{1n}\equiv\frac{1}{n}cov\left(\sum_{t=m}^{T-\tau}\widetilde{L}_{t+\tau}^*,\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*\right)$, $A_{2n}\equiv\frac{1}{n}cov\left(\sum_{t=m}^{T-\tau}\frac{1}{t}\sum_{j=1}^{t}\widetilde{L}_j^*,\sum_{t=m}^{T-\tau}\frac{1}{t}\sum_{j=1}^{t}\widetilde{z}_t\widetilde{L}_j^*\right)$, $A_{3n}\equiv$ $\frac{1}{n}cov\left(\sum_{t=m}^{T-\tau}\frac{1}{t}\sum_{j=1}^{t}\widetilde{L}_j^*,\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*\right)$, $A_{4n}\equiv\frac{1}{n}cov\left(\sum_{t=m}^{T-\tau}\widetilde{L}_{t+\tau}^*,\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\frac{1}{t}\sum_{j=1}^{t}\widetilde{z}_t\widetilde{L}_j^*\right)$. Consider each term separately: (i) $A_{1n}=\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_{s+\tau}^*\right)\underset{p}{\rightarrow}\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)$;
(ii) $|A_{2n}|=\left|\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\sum_{s=m}^{T-\tau}\sum_{k=1}^{s}\frac{1}{s}\frac{1}{t}E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\leq\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\sum_{s=m}^{T-\tau}\sum_{k=1}^{s}\frac{1}{s}\frac{1}{t}\left|E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|$ $\leq\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\frac{1}{m^2}\sum_{j=1}^{t}\sum_{k=1}^{s}\left|E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\leq\frac{1}{n}\left(\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\frac{1}{m^2}\right)\left(\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\left|E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\right)$ $\underset{p}{\rightarrow}0$ by Assumptions A7 and B4; (iii) $A_{3n}=\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\sum_{s=m}^{T-\tau}\frac{1}{t}E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_{s+\tau}^*\right)\underset{p}{\rightarrow}\left[1-\pi^{-1}\ln(1+\pi)\right]$ $\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)$ from similar reasonings to those in Lemma A6 in West (1996); (iv) Letting $\widetilde{\kappa}_3(j,s)\equiv E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_j^*\right)$, we have $|A_{4n}|=\left|\frac{1}{n}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\sum_{j=1}^{s}\frac{1}{s}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_j^*\right)\right|\leq$
$\frac{1}{n}\sum_{s=m}^{T-\tau}\frac{1}{s}\sum_{t=m}^{T-\tau}\sum_{j=1}^{s}\left|E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_j^*\right)\right|\leq\frac{1}{m}\sum_{s=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}|\widetilde{\kappa}_3(j,s)|\underset{p}{\rightarrow}0$ by Assumption B4 and Lemma A1(a) of West (1996).
Therefore, $cov\left(\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right),\frac{1}{\sqrt{n}}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)$
$\underset{p}{\rightarrow}\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)-\left[1-\pi^{-1}\ln(1+\pi)\right]\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)=$
$\left[\pi^{-1}\ln(1+\pi)\right]\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)$. We have therefore shown that

$$\Omega_{m,n}\equiv var\left(\sqrt{n}\widehat{\delta}_n^*\right)=\begin{pmatrix}1 & -E(z_t)'\Sigma_{\widetilde{z}\widetilde{z}}^{-1}\\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}\begin{pmatrix}\sigma_{m,n}^2 & \Lambda\Sigma_{\widetilde{L}^*,\widetilde{z}\widetilde{L}^*}\\ \Lambda\Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{L}^*} & \Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*}\end{pmatrix}\begin{pmatrix}1 & -E(z_t)'\Sigma_{\widetilde{z}\widetilde{z}}^{-1}\\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}'$$

where $\Lambda=\left[\pi^{-1}\ln(1+\pi)\right]$, $\Sigma_{\widetilde{L}^*,\widetilde{z}\widetilde{L}^*}\equiv\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)$, $\Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*}\equiv$ $\sum_{j=-\infty}^{\infty}E\left(\widetilde{z}_t\widetilde{L}_{t+\tau}^*\widetilde{L}_{t+\tau-j}^*\widetilde{z}_{t-j}\right)$. Consistency of $\widehat{\Omega}_{m,n}$ for $\Omega_{m,n}$ and the asymptotic distribution under H_0 then follow from reasonings analogous to those in the proof of Corollary 4. ■

Proof of Corollary 8. When the losses are conditionally homoskedastic, then $A_{1n}\underset{p}{\rightarrow}0$, and

$A_{3n} \underset{p}{\rightarrow} 0$ in the proof of Proposition 7, which implies $\Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{L}^*} = 0$. Thus,

$$\begin{aligned}\Omega_{m,n} &= \begin{pmatrix} 1 & -E(z_t)' \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix} \begin{pmatrix} \sigma_{m,n}^2 & 0 \\ 0 & \Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*} \end{pmatrix} \begin{pmatrix} 1 & -E(z_t)' \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix}' \\ &= \begin{pmatrix} \sigma_{m,n}^2 + E(z_t)' \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*} \Sigma_{\widetilde{z}\widetilde{z}}^{-1} E(z_t) & -E(z_t)' \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*} \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \\ -\Sigma_{\widetilde{z}\widetilde{z}}^{-1} \Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*} \Sigma_{\widetilde{z}\widetilde{z}}^{-1} E(z_t) & \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{z}\widetilde{L}^*} \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix}.\end{aligned}$$

Confidence bands for $SL_{t+\tau}$ can be easily obtained from $[1, z_t']' \widehat{\delta} | z_t \sim N([1, z_t']' \delta, \sigma_{m,n}^2 + [z_t - E(z_t)]' \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \Sigma_{\widetilde{z}L} \Sigma_{\widetilde{z}\widetilde{z}}^{-1} [z_t - E(z_t)])$. If furthermore data are iid, $\Sigma_{\widetilde{z}\widetilde{z}}^{-1} \Sigma_{\widetilde{z}L} \Sigma_{\widetilde{z}\widetilde{z}}^{-1} = \Sigma_{\widetilde{z}\widetilde{z}}^{-1} \gamma_0^{LL}$. ∎

Figure 1(a). Power functions

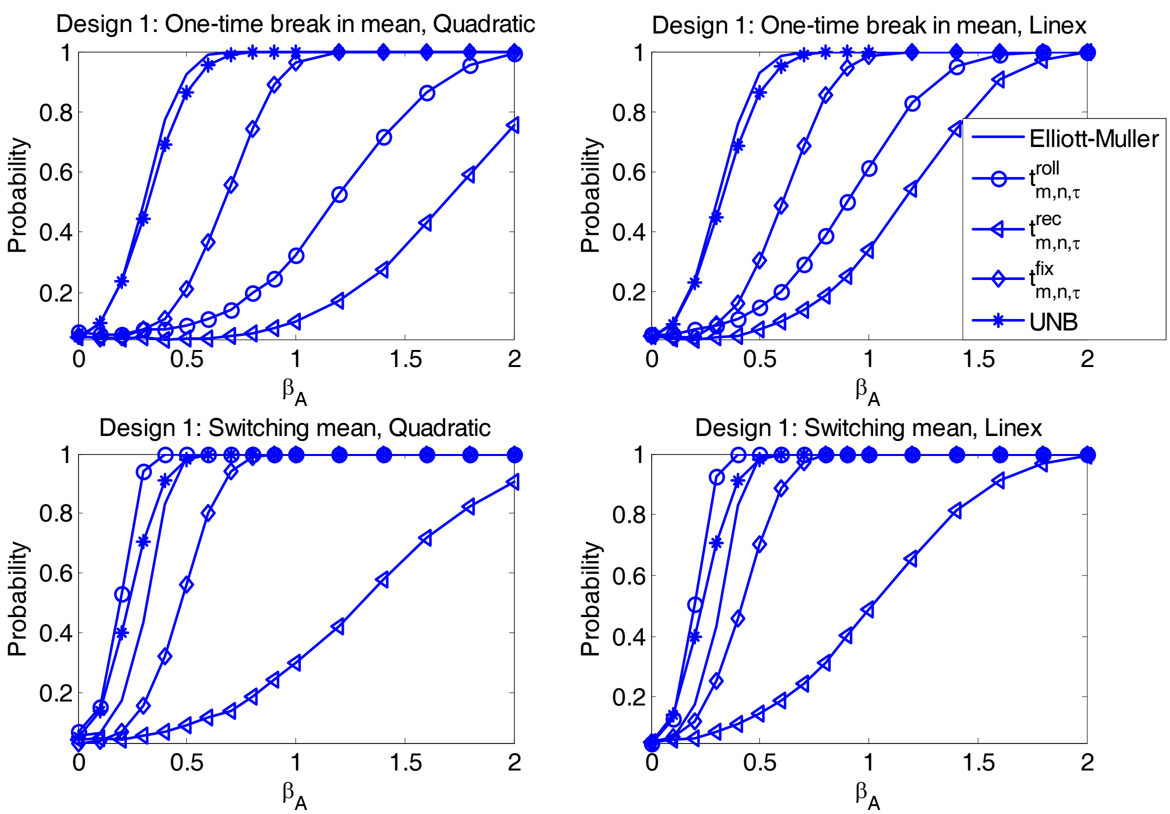

Figure 1(b). Power functions

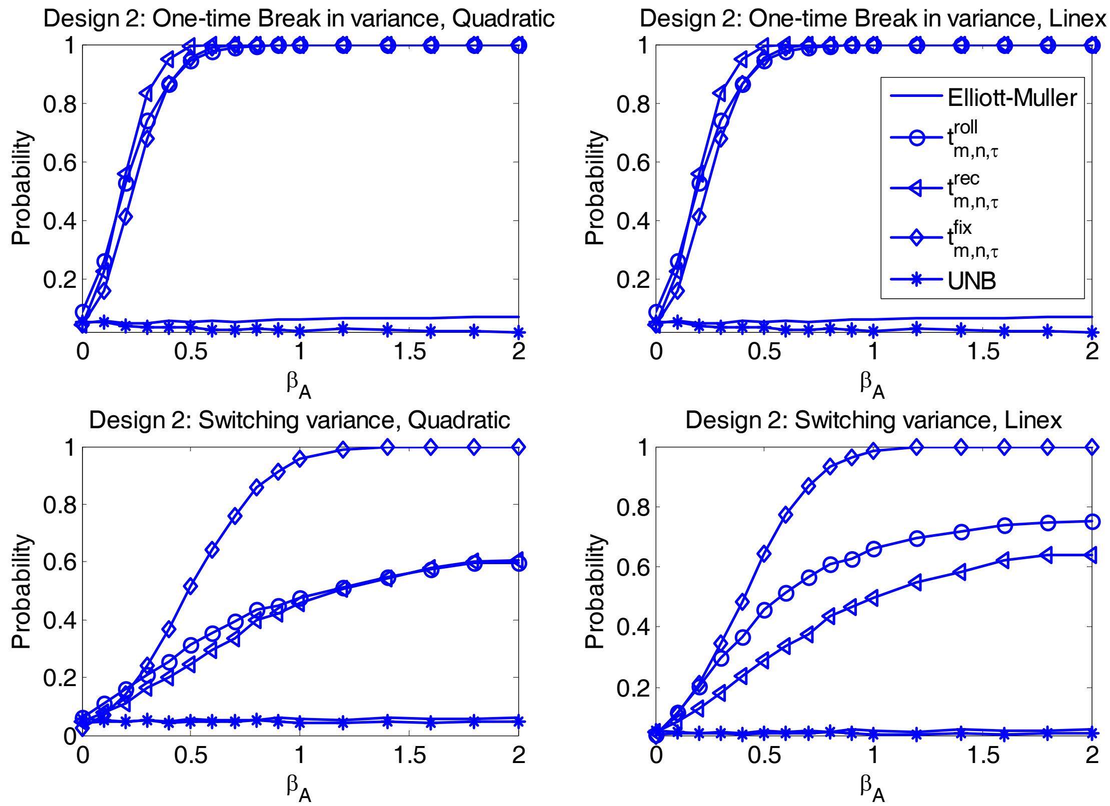

Figure 1c. Power functions

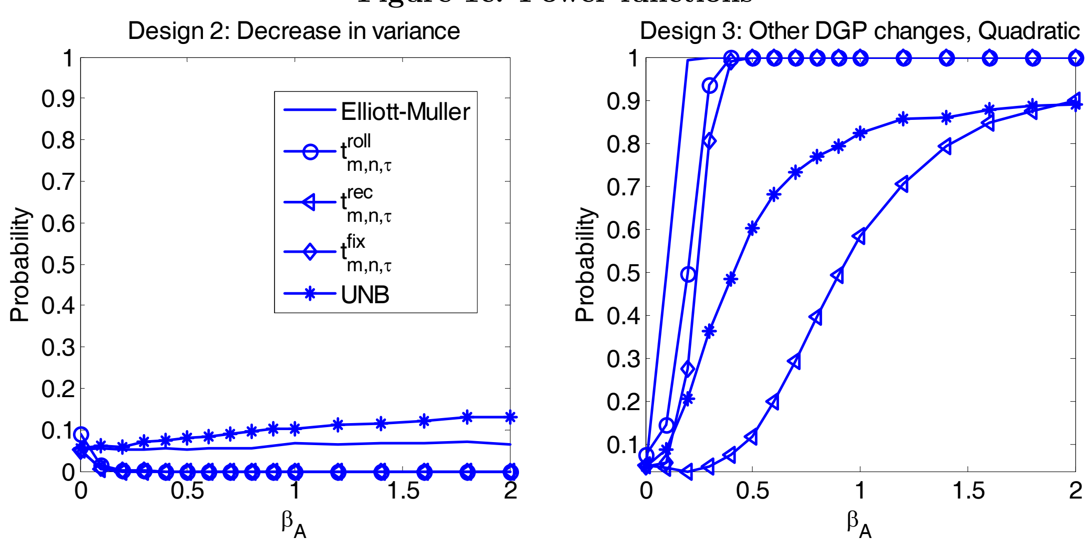

Figure 2. Fitted surprise losses

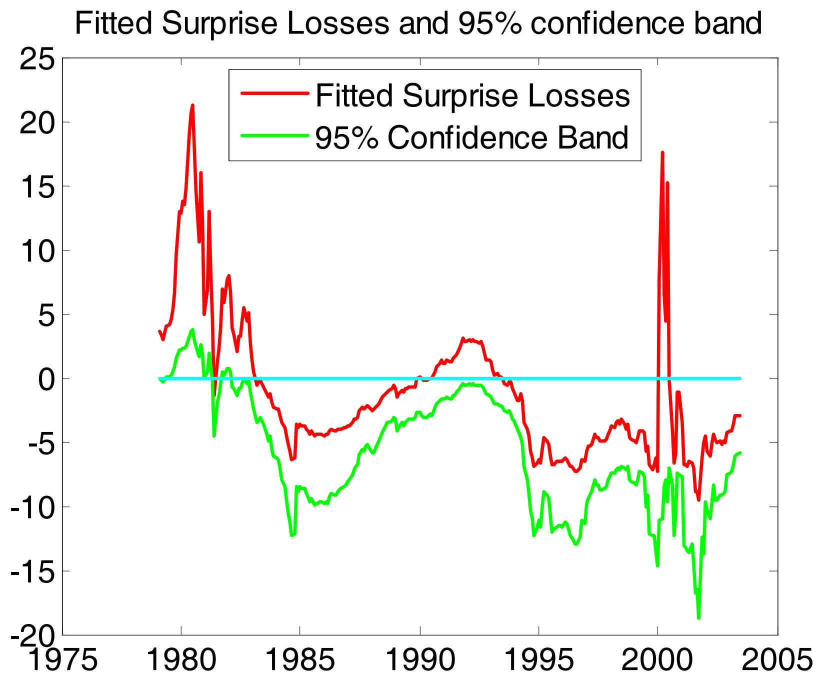

Table 1(a). Size of FB test and structural break tests. Nominal size .05

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

Notes to Table 1(a). The table reports rejection frequencies over 5000 Monte Carlo replications of the forecast breakdown (FB) 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 and of Elliott and Muller's (2003) test (EM). The experiment designs MC1 and MC2 are described in Section 7.1 and m and n denote in-sample and out-of-sample sizes, respectively.

Table 1(b). Size of overfitting-corrected FB test. 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 to Table 1(b). The table reports rejection frequencies over 5000 Monte Carlo replications of the overfitting-corrected forecast breakdown (FB) test of Section 4, using either the asymptotic variance estimator of Corollary 3 ($t^{\mathrm{c}}_{m,n,\tau}$) 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 7.1 and m and n denote in-sample and out-of-sample sizes, respectively.

Table 2. Size of forecast rationality tests. Nominal size .05

		Uncorrected			Corrected		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
				t_{δ_0}			
50	50	0.172	0.021	0.052	0.054	0.053	0.052
50	100	0.266	0.002	0.050	0.056	0.053	0.050
50	150	0.321	0.000	0.048	0.052	0.058	0.048
100	50	0.111	0.044	0.056	0.056	0.052	0.055
100	100	0.172	0.018	0.053	0.053	0.051	0.053
100	150	0.215	0.004	0.050	0.051	0.048	0.050
150	50	0.101	0.053	0.059	0.061	0.059	0.061
150	100	0.136	0.037	0.054	0.055	0.055	0.054
150	150	0.177	0.016	0.050	0.052	0.047	0.049
				t_{δ_1}			
50	50	0.060	0.062	0.061	0.064	0.062	0.062
50	100	0.055	0.053	0.054	0.051	0.053	0.051
50	150	0.048	0.050	0.049	0.049	0.050	0.049
100	50	0.055	0.054	0.054	0.052	0.052	0.052
100	100	0.056	0.056	0.056	0.056	0.057	0.057
100	150	0.045	0.048	0.045	0.049	0.049	0.049
150	50	0.062	0.062	0.062	0.061	0.060	0.061
150	100	0.057	0.056	0.057	0.057	0.056	0.057
150	150	0.051	0.051	0.050	0.050	0.050	0.051
				$Wald$			
50	50	0.148	0.040	0.059	0.065	0.071	0.069
50	100	0.220	0.024	0.058	0.055	0.054	0.057
50	150	0.276	0.017	0.051	0.050	0.053	0.048
100	50	0.102	0.050	0.058	0.066	0.061	0.063
100	100	0.146	0.035	0.057	0.057	0.057	0.056
100	150	0.179	0.018	0.048	0.048	0.049	0.048
150	50	0.097	0.062	0.066	0.068	0.072	0.069
150	100	0.115	0.047	0.058	0.060	0.061	0.059
150	150	0.148	0.031	0.053	0.048	0.053	0.052

Notes to Table 2. The table reports rejection frequencies over 5000 Monte Carlo replications of forecast rationality tests. We consider t-tests of significance of the intercept (t_{δ_0}) and the slope coefficient (t_{δ_1}), as well as a test of joint significance of both coefficients ($Wald$) in the forecast rationality regression (15).

Forecast errors are obtained using either a fixed, rolling or recursive scheme and in each case the tests are implemented using either the usual OLS variance estimator ("uncorrected") or the asymptotic variance estimator of Corollary 8 ("corrected"). The experiment design is described in Section 7.1 and m and n denote in-sample and out-of-sample sizes, respectively.

Table 3. P-values of forecast breakdown test

		Real-time data	Revised data
q_u	q_π	$t_{m,n,\tau}$	$t_{m,n,\tau}$
		$\tau = 1$	
1	1	- -	0.004
1	3	- -	0.021
3	1	- -	0.009
3	3	- -	0.039
BIC		- -	0.021
		$\tau = 3$	
1	1	0.000	0.256
1	3	0.562	0.326
3	1	0.450	0.434
3	3	0.572	0.524
BIC		0.874	0.475
		$\tau = 12$	
1	1	0.001	0.111
1	3	0.000	0.312
3	1	0.002	0.756
3	3	0.001	0.948
BIC		0.001	0.591

Notes to Table 3. The table reports p-values for the forecast breakdown test ($t_{m,n,\tau}$) of Theorem 2(a). We used a rolling scheme with $m = 60$, $n = 95$ for real-time data, and $m = 241$ and $T = 546$ for revised data. The forecast horizons are $\tau = 1$, 3 and 12 months (since real-time data are only available at a quarterly frequency, in this case we only report results for $\tau = 3$ months and $\tau = 12$ months). q_u and q_π are the number of lags used for unemployment and for inflation, respectively; the row labeled "BIC" reports results for the case in which the lag length is determined by the BIC with a maximum of three lags.

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

			δ_1				$W_{m,n,\tau}$
τ	q_u	q_π	$z_t = \widehat{\beta}_t$	$z_t = \widehat{\gamma}_t$	$z_t = \widehat{\rho}_t$	$z_t = \widehat{\sigma}^2_{\pi,t}$	$z_t = (\widehat{\beta}_t, \widehat{\gamma}_t, \widehat{\rho}_t)'$
1	1	1	-2.285	1.828	19.770	-1.019	9.533
			(0.156)	(0.018)	(0.795)	(0.024)	(0.023)
	1	3	-2.348	1.612	6.484	-0.892	7.386
			(0.159)	(0.037)	(0.933)	(0.051)	(0.061)
	3	1	-2.306	1.712	13.957	-0.980	8.397
			(0.148)	(0.028)	(0.856)	(0.031)	(0.039)
	3	3	-2.354	1.513	1.977	-0.866	6.623
			(0.153)	(0.050)	(0.980)	(0.059)	(0.085)
	BIC		-2.187	1.654	6.272	-0.855	7.286
			(0.185)	(0.046)	(0.938)	(0.071)	(0.063)
3	1	1	-1.806	-0.404	-114.2	-1.713	1.985
			(0.531)	(0.785)	(0.249)	(0.000)	(0.576)
	1	3	-1.837	-0.267	-122.4	-1.716	2.077
			(0.519)	(0.858)	(0.238)	(0.000)	(0.557)
	3	1	-1.651	-0.568	-128.8	-1.705	2.337
			(0.575)	(0.706)	(0.201)	(0.010)	(0.506)
	3	3	-1.657	-0.415	-136.1	-1.702	2.386
			(0.570)	(0.782)	(0.195)	(0.000)	(0.496)
	BIC		-1.608	-0.642	-141.4	-1.613	2.602
			(0.590)	(0.669)	(0.175)	(0.001)	(0.457)
12	1	1	-1.304	-0.105	-199.5	-1.876	6.268
			(0.578)	(0.942)	(0.040)	(0.000)	(0.099)
	1	3	-1.639	-0.417	-192.0	-1.641	6.778
			(0.480)	(0.776)	(0.032)	(0.000)	(0.079)
	3	1	-0.679	-0.863	-256.5	-1.878	7.162
			(0.797)	(0.592)	(0.026)	(0.000)	(0.067)
	3	3	-0.960	-1.108	-250.9	-1.661	8.445
			(0.708)	(0.488)	(0.017)	(0.000)	(0.038)
	BIC		-0.903	-0.789	-246.5	-1.810	7.308
			(0.729)	(0.620)	(0.024)	(0.000)	(0.063)

Notes to Table 4. The table reports the coefficient estimates of δ_0 and δ_1 in equation (24), for different choices of z_t. $\widehat{\beta}_t$, $\widehat{\gamma}_t$ and $\widehat{\rho}_t$ are rolling estimates of the structural parameters in the monetary policy reaction function of the Fed described in 23, and $\widehat{\sigma}^2_{\pi,t}$ is a rolling estimate of volatility of inflation changes. The numbers within parentheses are the p-value of the test of significance of the individual coefficient. The last column reports the Wald test statistic $W_{m,n,\tau}$ introduced in Section 5 (with a HAC bandwidth equal to $n^{1/3}$) and its associated p-value (in parentheses). 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 by the BIC with a maximum of three lags. τ is the forecast horizon.