Detecting and Predicting Forecast Breakdowns

Raffaella Giacomini and Barbara Rossi

University of California, Los Angeles and Duke University

April 2005

PRELIMINARY AND INCOMPLETE

Abstract

We propose a theoretical framework for detecting and predicting forecast breakdowns, defined as situations where the out-of-sample performance of a forecast method, judged by some loss function, is systematically worse than its in-sample performance. Our framework can be used to establish the robustness of a forecast method in the past, and to predict whether the forecast method will break down at a future date. We show that main causes of forecast breakdowns are overfitting and structural breaks in the data generating process that have some probability of occurring again in the future ("recurring breaks"). Finally, we find evidence of a forecast breakdown in the Phillips' curve forecasts of U.S. inflation over the past three decades, and link the breakdown to changes in the monetary policy reaction function of the Fed and in inflation variability.

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

Acknowledgements: We would like to thank F. Diebold, G. Elliott, M. McCracken, N. Meddahi, and seminar participants to the 2005 Missouri Economics Conference, the Duke Financial Econometrics Group, the 2005 CIRANO-CIREQ Conference on Forecasting in Macroeconomics and Finance.

1 Introduction

We define a forecast breakdown as a situation where the out-of-sample performance of a forecast method, judged by some loss function, is systematically worse than its in-sample performance and propose a theoretical framework for assessing the robustness of a given forecast method against forecast breakdowns. We show how to: (1) *detect* whether a forecast method broke down in the past, and (2) *predict* whether the forecast method will break down at a future date, the latter prediction possibly conditioned on hypothetical future paths of the economy.

Our notion of forecast breakdown is related to what Clements and Hendry (1998, 1999) called a "forecast failure". These authors characterized a forecast failure as a "deterioration in forecast performance relative to the anticipated outcome based on earlier fit" (Clements and Hendry, 1999, p. 1) and concluded from an extensive number of Monte Carlo studies and empirical analyses that forecast failures are primarily caused by structural breaks in the unconditional mean of the variable of interest. While we build on Clements and Hendry's important work, we depart from their approach in several directions and contribute to the forecast evaluation literature in a number of ways. First, we make the criterion for defining a forecast breakdown more stringent by requiring that the out-of-sample performance of a forecast method be *systematically* worse than its in-sample performance. In practice, this amounts to comparing each out-of-sample loss to a *continuously updated* estimate of the average in-sample loss that also uses recent observations. The definition of Clements and Hendry, instead, would suggest comparing average out-of-sample loss to average in-sample loss estimated only once. Second, we propose a formal testing framework that is generally applicable (e.g., it allows for general estimation procedures and general loss functions). In particular, we derive analytic results for constructing test statistic with the appropriate size in a framework pioneered by West (1996), where we allow both the in-sample and the out-of-sample sizes to be asymptotically a fixed fraction of the total sample size. We also take into account the effects of parameter estimation uncertainty. A third novelty of our approach is the focus on predicting future forecast breakdowns, rather than just detecting past ones. Finally, we formally analyze the causes of forecast breakdowns both analytically in the context of an illustrative example and by Monte Carlo simulations. In particular, we show that our tests can capture forecast breakdowns caused by overfitting (loosely defined as a situation where the model describes the "noise" as well as the "signal" in the sample) as well as *recurring* changes in the data-generating process (defined as changes that have a non-zero probability of occurring again in the future).

The fact that our tests can capture a number of various possible causes of a forecast breakdown at once makes distinguishes our approach from several strands of the literature, that address specific causes individually. Examples are, structural break testing in model parameters (Elliott and Muller (2003)); joint tests for model selection and parameter instability (Rossi, 2005); use of information criteria for model selection to avoid overfitting. These approaches have several drawbacks when used in a forecasting context. Structural break tests, for example, only focus on the past stability of the parameters of the forecast model, and provide no information on the likelihood of future breaks. Furthermore, Elliott (2005) analytically shows, in the context of a single break, that attempts to forecast based on estimates of the break point are unlikely to improve forecasts. With regards to overfitting, information criteria have been found not to be able to systematically improve forecasts. A related class of tests, that, although not specifically designed to detect breaks and overfitting, may nonetheless indirectly capture such effects are forecast optimality tests. We shed some light on the power properties of our forecast breakdown test relative to those of structural break and forecast optimality tests in a Monte Carlo study.

With regards to changes in the data-generating process as a cause of forecast breakdowns, an important feature of our test is that it has power against breaks that have some probability of occurring again in the future. The existing structural break tests, instead, are unable to distinguish between breaks that are an accidental feature of the sample - and thus are irrelevant for forecasting - and recurring breaks, which are of greater interest to forecasters (see Pesaran, Pettenuzzo and Timmermann (2004)). We also discuss the finite-sample implication of this asymptotic result, namely that the choice of the sample split between estimation and evaluation windows in practice affects the test's ability to capture forecast breakdowns due to different causes. We offer some insight into the effects of different choices of sample split on the power of the forecast breakdown test in Monte Carlo simulations.

An innovation of our approach that has useful practical implications is the possibility of making predictions about the likelihood that a forecast method will break down at a future date. In practice, we propose to do so by looking for explanatory variables in the forecaster's information set that can predict the difference between out-of-sample and in-sample performance. If such predictive relationships are found, they can then be used to make probabilistic statements about how much future losses will differ from their current expectations.

To illustrate the methods proposed in this paper, we investigate whether there is evidence of forecast breakdown in the Phillips curve's forecast of inflation in the United States. The Phillips curve relationship has been a useful guide for monetary policy in the U.S. in the past few decades. However, its in-sample stability has recently been challenged by Staiger, Stock and Watson (1997), and its forecasting ability further investigated by Stock and Watson (1999) and Fisher et al. (2002). By using both real-time and revised data, we find that there is striking empirical evidence in favor of a breakdown in the Phillips curve forecasts. We further investigate whether either monetary policy parameters or inflation variability would have been useful to predict the forecast breakdown. We find that changes in the monetary policy behavior of the Fed as well as the degree of inflation variability played a key role in explaining the forecast breakdown. We also offer a truly out-of-sample prediction of the likelihood that the forecasting ability of the Phillips curve will breakdown within a year from the end of our sample.

The paper is organized as follows. Section 1 introduces the main theme of the paper, Sections 2 and 3, respectively, discuss the theoretical foundations of the forecast breakdown test and the test for predicting future forecast breakdown. Section 4 presents a simple motivating example, and Sections 5 provides Monte Carlo experiments to compare our test with various tests available in the literature. Section 6 presents the empirical application to inflation forecasts in the past three decades, and Section 7 concludes.

2 Theory

2.1 Description of the environment

We let $W \equiv \{W_t : \Omega \longrightarrow \mathbb{R}^{s+1}, s \in \mathbb{N}, t = 1, \ldots, T\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and partition the observed vector W_t as $W_t \equiv (Y_t, X_t')'$, where $Y_t : \Omega \to \mathbb{R}$ is the variable of interest and $X_t : \Omega \to \mathbb{R}^s$ is a vector of predictors. We define the information set at time t as $\mathcal{F}_t = \sigma(W_1', ..., W_t')'$. The object of our evaluation is a particular forecast method for forecasting the variable of interest τ steps ahead, $Y_{t+\tau}$. We denote the forecast by $\hat{f}_t(\hat{\beta}_t) \equiv f(W_t, W_{t-1}, ..., W_{t-m+1}; \hat{\beta}_t)$, where f is a measurable function and $\hat{\beta}_t$ is the $k \times 1$ parameter estimate based on the sample $(W_t, W_{t-1}, ..., W_{t-m+1})$.

Let T be the total sample size and m be the size of the first estimation window. We allow for three forecasting schemes, that are the most popular in the literature. A rolling window forecasting scheme computes the first $\tau-$step ahead forecast at time m, using data indexed $1, ..., m$ and compares the forecast to the realization $y_{m+\tau}$. The second forecast is formulated at time $m + 1$, using the previous m observations and it is compared to the realization $y_{m+1+\tau}$. The procedure is iterated and the last forecast is produced at time $T - \tau$, by utilizing the m most recent observations, and compared to y_T. This yields a sequence of $n \equiv T - \tau - m + 1$ out-of-sample forecasts and corresponding forecast errors. To clarify, only a fixed window of the past m observations is used for estimation, so that as t increases from m to T, older observations are discarded. For example, in the linear model $Y_{t+\tau} = X_t'\beta + \varepsilon_{t+\tau}$, we have $\widehat{\beta}_t = \left(\sum_{s=t-m+1}^{t} X_s X_s'\right)^{-1} \sum_{s=t-m+1}^{t} X_s Y_s$. In the recursive scheme, at each time $t = m, m + 1, ...T$ the parameter estimate $\widehat{\beta}_t$ depends on all observations from 1 to t. For example, in the linear model above $\widehat{\beta}_t = \left(\sum_{s=1}^{t} X_s X_s'\right)^{-1} \sum_{s=1}^{t} X_s Y_s$. In the split sample scheme the parameter vector is estimated only once, from observations 1 to m, so that the estimates are never updated even if new observations become available. In the linear model above, $\widehat{\beta}_t = (\sum_{s=1}^{m} X_s X_s')^{-1} \sum_{s=1}^{m} X_s Y_s$.

For a given a forecasting scheme, each forecast corresponds to a sequence of in-sample fitted values $\hat{y}_j(\hat{\beta}_t)$, $j = t - m + 1, ..., t$. For example, for the linear model above and a rolling scheme, the time-t forecast is $\hat{f}_t(\hat{\beta}_t) = X_t'\hat{\beta}_t$ and the corresponding in-sample fitted values are $\hat{y}_j(\hat{\beta}_t) = X_j'\hat{\beta}_t,\ j = t-m+1,...,t-\tau$.

We evaluate the sequence of out-of-sample forecasts by selecting a loss function $L(Y_{t+\tau}, \hat{f}_t)$, that depends on the forecast and on the variable of interest. Notice that the parameter estimate $\hat{\beta}_t$ does not necessarily minimize L, thus allowing for different loss functions for estimation and evaluation (although, for comparability, the same loss function should be used for in-sample and out-of-sample evaluation).

2.2 Definition of Forecast Breakdown

We define a forecast breakdown as a deterioration in the out-of-sample performance of the forecast method relative to its in-sample performance. If the observed out-of-sample loss is much higher than the in-sample average loss, then we suspect that the forecast method is failing to capture important features of the data.

We formalize this idea by defining a "surprise loss" at time $t+\tau$ as the difference between the observed loss at time $t+\tau$ and the in-sample average loss:

$$SL_{t+\tau}(\hat{\beta}_t) = L(Y_{t+\tau}, \hat{f}_t(\hat{\beta}_t)) - \frac{1}{m}\sum_{j=t-m+1}^{t} L(Y_j, \hat{y}_j(\hat{\beta}_t)) \text{ for } t = m, ...T-\tau.$$

Throughout, we require the following set of assumptions which are standard in the literature – see e.g. West (1996).[1] To simplify notation, let $l_t^{oos}\left(\hat{\beta}_t\right)$ denote $L(Y_{t+\tau}, \hat{f}_t(\hat{\beta}_t))$ and $l_j^{is}\left(\hat{\beta}_t\right)$ denote $L(Y_j, \hat{y}_j(\hat{\beta}_t))$, and l_t^{oos}, l_j^{is} denote these quantities evaluated at β^*, where β^* is defined in the following Assumption 1.

Assumptions.

1. *The estimate* $\widehat{\beta}_t$ *satisfies* $\widehat{\beta}_t - \beta^* = B_t H_t$, *where* B_t *is* $(k \times q)$ *and* H_t *is* $(k \times 1)$, *with: (a)* $B_t \underset{as}{\rightarrow} B$, B *a matrix of rank* k; *(b)* H_t *depends on the type of estimator used:* $H_t^{rec} = t^{-1}\sum_{s=1}^{t} h_s(\beta^*)$, $H_t^{roll} = m^{-1}\sum_{s=t-m+1}^{t} h_s(\beta^*)$, $H_t^{split} = m^{-1}\sum_{s=1}^{m} h_s(\beta^*)$, *for a* $(q \times 1)$ *orthogonality condition* $h_s(\beta^*)$; *(c)* $Eh_s(\beta^*) = 0$.

2. *Let* $l_t(\beta) = [l_t^{oos}(\beta), l_t^{is}(\beta)]'$. *In some open neighborhood* N *around* β^*, *and with probability one: (a)* $l_t(\beta)$ *is measurable and twice continuously differentiable with respect to* β. *(b) there exists a constant* $D < \infty$ *such that for all* t, $sup_{\beta \in N}|\partial^2 l_t(\beta)/\partial\beta\partial\beta'| < M_t$ *for a measurable* M_t *for which* $EM_t < D$.

3. *Let* $l_t \equiv l_t(\beta^*)$, $l_{t\beta} \equiv \partial l_t(\beta^*)/\partial\beta$, $F = E\left(l_{t\beta}^{oos} - l_{t\beta}^{is}\right)$. *(a) For some* $d > 1$, $sup_t E||[vec(l_{t\beta})', vec(l_t)', h_t']'||^{4d} < \infty$, *where* $||.||$ *denotes Euclidean norm. (b)* $[vec(l_{t\beta} - F)', vec(l_t - El_t)', h_t']'$ *is strong mixing with mixing coefficients of size* $-3d/(d-1)$. *(c)* $[vec(l_{t\beta})', h_t']'$ *is covariance*

[1] See also McCracken (2000) for weaker assumptions.

stationary, and cov $\left(vec\left(l_t\right)', vec(l_{t-j})'\right)$ *depend on* j *but not on* t*. (d)* l_t^{oos} *and* l_t^{is} *(separately) have a positive long run variance.*[2]

4. $m, n \rightarrow \infty$ *as* $T \rightarrow \infty$*, and* $lim_{T\rightarrow\infty}\left(n/m\right) = \pi$*,* $0 < \pi < \infty$*.*

These assumptions allow a variety estimation of techniques (OLS, GMM, MLE), serial correlation and conditional heteroskedasticity in both $l_t - El_t$ and h_t, and take into account the possibility that parameter estimation uncertainty may affect the forecast errors.

2.3 Forecast Breakdown Test

From the discussion in the previous section, it follows that if a forecast method is reliable, one would expect the out-of-sample surprise losses $\{SL_{t+\tau}\}_{t=m+1}^{T-\tau}$ to have mean zero. From the perspective of a forecaster, decreases in expected out-of-sample losses may also be desirable. Based on this intuition, we construct our forecast breakdown test as a test of the hypothesis that the surprise losses are on average zero, against the alternative that they have a positive mean.

Formally, for a given loss function L, we test the null hypothesis of no forecast breakdown for forecast method f at the target date $t+\tau$:[3]

$$H_0 : E[SL_{t+\tau}\left(\beta^*\right)] = 0 \tag{1}$$

against the one-sided alternative

$$H_A : E[SL_{m,n}\left(\beta^*\right)] > 0 \tag{2}$$

In this section, we will focus on the most common case in which the loss function used for in-sample evaluation is the same as the loss function used for out-of-sample evaluation, $l_t^{is} = l_t^{oos} = l_t$ (it is not necessary, though, that the in-sample loss function used for evaluation be the same as the in sample loss used for estimating the parameters β). In this case, F (defined in Assumption 3) equals zero. Interestingly, this implies that parameter estimation error does not affect the asymptotic distribution of our test statistic, which turns out to be very straightforward to calculate.

Our test statistic, based on the average surprise loss: $\overline{SL}_{m,n} \equiv n^{-1}\sum_{t=m}^{T-\tau} SL_{t+\tau}$, is

$$t_{m,n,\tau} = \frac{\overline{SL}_{m,n}}{\hat{\sigma}_{m,n}^2/\sqrt{n}} \tag{3}$$

where $\hat{\sigma}_{m,n}^2$ is a heteroskedasticity and autocorrelation-consistent (HAC) estimator of the asymptotic variance $\sigma_{m,n}^2 = var[\sqrt{n}\overline{SL}_{m,n}] = \lambda_l S_{ll}$, where $S_{ll} \equiv \sum_{k=-\infty}^{\infty} E\left(l_t - El_t\right)\left(l_{t-k} - El_t\right)$, and

[2] We do not impose covariance stationarity on l_t because otherwise our H_0 would be trivially satisfied.

[3] Note that our null hypothesis could be re-expressed as the hypothesis that the losses are constant over time.

whose precise form depends on λ_l, which differs depending on the type of estimator used:

Scheme	λ_l
Rec	1
Roll, $\pi \leq 1$	$\left[1 - \frac{1}{3}\pi^2\right]$
Roll, $\pi > 1$	$\left[1 - \frac{1}{2}\pi^{-1}\right]$
Split	$(1 + \pi)$

(4)

The long-run variance S_{ll} can be consistently estimated by a HAC estimator along the lines of Newey and West, 1987 and Andrews, 1991. A level α test rejects the null hypothesis of no forecast breakdown whenever $t_{m,n,\tau} > z_\alpha$, where z_α is the $(1-\alpha) - th$ quantile of a standard normal distribution.

The following theorem formalizes the test.[4]

Theorem 1 (Forecast Breakdown Test) *For the given forecast horizon $\tau = 1$ and a first estimation window of size m, under Assumptions 2-4 and when the loss function used for in-sample evaluation is the same as the loss used for out-of-sample evaluation, then under H_0 in (1),*

$$t_{m,n,\tau} \xrightarrow{d} N(0,1) \tag{5}$$

where m, $n \to \infty$ and $\frac{n}{m} \to \pi$ as $T \to \infty$, and where $\hat{\sigma}^2_{m,n}$ is a consistent estimator of $\sigma^2_{m,n} = \lambda_l S_{ll}$ and λ_l is as in (4).

Even if the asymptotic distribution is very simple and independent of parameter estimation error, still note that, unless the forecasting procedure is recursive, the researcher needs to include a correction factor λ_l in the estimation of the asymptotic variance. Failing to do so will result in size distortions, which will be more carefully examined in the Monte Carlo section.

2.4 The Forecast Breakdown test when the losses differ

In the more general case where the loss function used for in-sample evaluation is different from the loss function used for out-of-sample evaluation, $l_t^{is} \neq l_t^{oos}$, parameter estimation error will be important and we have the following result:

Theorem 2 (Forecast Breakdown Test when $l^{is} \neq l^{oos}$) *For the given forecast horizon $\tau = 1$ and a first estimation window of size m, under Assumptions 1-4, then under H_0 in (1),*

$$t_{m,n,\tau} \xrightarrow{d} N(0,1) \tag{6}$$

[4] While we use the theoretical framework of West (1996), however his results cannot be directly applied to our case, as the long-run covariance matrix S (in West's (1996) notation) is singular.

where m, $n \to \infty$ *and* $\frac{n}{m} \to \pi$ *as* $T \to \infty$, *and where* $\hat{\sigma}^2_{m,n}$ *is a consistent estimator of* $\sigma^2_{m,n}$. $\sigma^2_{m,n}$ *is as follows:*

$$\begin{aligned} \sigma^2_{m,n} \quad \equiv \quad & \lambda_{oos} S_{oos} + \lambda_{is} S_{is} + \lambda_{hh} FBS_{hh} B'F' + \lambda_{oos,is} \left(S_{oos,is} + S'_{oos,is} \right) + \\ & \lambda_{oos,h} \left(S_{oos,h} B'F' + FBS'_{oos,h} \right) + \lambda_{is,h} \left(S_{is,h} B'F' + FBS'_{is,h} \right) \end{aligned}$$

where

Scheme	λ_{is}	λ_{oos}	λ_h	$\lambda_{is,oos}$	$\lambda_{is,h}$	$\lambda_{oos,h}$
Recursive	2Π	1	2Π	$-\Pi$	-2Π	Π
Rolling, $\pi \leq 1$	$\pi - \frac{1}{3}\pi^2$	1	$\pi - \frac{1}{3}\pi^2$	$-\frac{1}{2}\pi$	$-\pi + \frac{1}{3}\pi^2$	π
Rolling, $\pi > 1$	$1 - \frac{1}{3}\pi^{-1}$	1	$1 - \frac{1}{3}\pi^{-1}$	$-1 + \frac{1}{2}\pi^{-1}$	$-1 + \frac{1}{3}\pi^{-1}$	$1 - \frac{1}{2}\pi^{-1}$
Split	π	1	π^{-1}	0	-1	0

and $\Pi \equiv \left[1 - \pi^{-1} \ln(1+\pi)\right]$, $S_{oos} \equiv \sum_{k=-\infty}^{\infty} E\left(l_t^{oos} - El_t^{oos}\right)\left(l_{t-k}^{oos} - El_t^{oos}\right)$, $S_{is} \equiv \sum_{k=-\infty}^{\infty} E\left(l_t^{is} - El_t^{is}\right)\left(l_{t-k}^{is} - E\right.$
$S_h \equiv \sum_{k=-\infty}^{\infty} E h_t h'_{t-k}$, $S_{oos,h} \equiv \sum_{k=-\infty}^{\infty} E\left(l_t^{oos} - El_t^{oos}\right) h'_{t-k}$, $S_{is,h} \equiv \sum_{k=-\infty}^{\infty} E\left(l_t^{is} - El_t^{is}\right) h'_{t-k}$,
$S_{oos,is} \equiv \sum_{k=-\infty}^{\infty} E\left(l_t^{oos} - El_t^{oos}\right)\left(l_{t-k}^{is} - El_t^{is}\right)$.

As before, the long-run variances in the theorem above can be consistently estimated by HAC estimators along the lines of Newey and West, 1987 and Andrews, 1991. Again, a level α test rejects the null hypothesis of no forecast breakdown whenever $t_{m,n,\tau} > z_\alpha$, where z_α is the $(1-\alpha) - th$ quantile of a standard normal distribution.

3 Predicting Future Forecast Breakdowns

The previous section proposed a test for detecting whether a forecast method broke down in the past. A question that may be of greater interest to real-time forecasters is whether the forecast method will break down in the future. This is of course related to finding past breakdowns: if the breakdown was caused by overfitting or by structural changes that - as discussed in Remark 1 above - have some probability of occurring again in the future, we may expect the forecast method to break down at some point in the future. In other words, if the surprise losses had non-zero mean in the past, we expect them to continue to be positive in the future. However, it is possible that one could find additional information available at the time of the prediction - besides a positive mean - that the forecaster could exploit in order to predict whether there will be a forecast breakdown at a specific date in the future. For example, the surprise losses may be persistent (this is easy to see for a quadratic loss function, where the presence of GARCH in the data will induce serial correlation in the surprise losses) or they may be correlated with leading indicators of the state of the economy.

We propose the following method to predict whether a forecast method will break down at the target date $T+\tau$.

1. Select a $q \times 1$ information vector h_t from the information set at time t and estimate the predictive regression
$$SL_{t+\tau} = h_t' \delta_\tau + \varepsilon_{t+\tau}, \tag{7}$$
over the out-of-sample period $t = m, ..., T-\tau$. Let $\widehat{\delta}_{n,\tau}$ denote the OLS coefficient. The information vector h_t can include a constant; lagged surprise losses; economically meaningful variables such as business cycle leading indicators, measures of stock market volatility, interest rates etc.

2. Test whether δ_τ is significantly different from zero by a Wald test
$$W_{m,n,\tau} = n\widehat{\delta}_{n,\tau}' \hat{V}_{n,\tau}^{-1} \widehat{\delta}_{n,\tau} \tag{8}$$
where $\hat{V}_{n,\tau}$ is a HAC estimator of the asymptotic variance of $\widehat{\delta}_{n,\tau}$ with truncation lag $\tau - 1$. Reject the null hypothesis that $\delta_\tau = 0$ if $W_{m,n,\tau} > \chi^2_{q,1-\alpha}$, where $\chi^2_{q,1-\alpha}$ is the $(1-\alpha)-th$ quantile of a χ^2_q distribution.

3. In case of rejection, use (7) to construct a $(1-\rho)\%$ prediction interval for $SL_{T+\tau}$. A prediction interval can be computed in a number of ways, using parametric, nonparametric or bootstrap techniques (see e.g. Chatfield, 1993). For example, a simple, approximate prediction interval (assuming normality and ignoring estimation uncertainty in $\widehat{\delta}_{n,\tau}$) can be constructed as
$$h_T' \widehat{\delta}_{n,\tau} \pm z_{\rho/2} \sqrt{\hat{\sigma}^2_{\varepsilon,\tau}},$$
where $\hat{\sigma}^2_{\varepsilon,\tau}$ is a HAC estimator of the variance of the errors $\varepsilon_{t+\tau}$ from regression (7). To obtain a prediction interval that accounts for parameter uncertainty, one could further utilize the correction recently proposed by Hansen (2004). If the prediction interval contains 0, conclude that the forecast method will break down at time $T+\tau$ with probability $(1-\rho)\%$.

4 An illustrative example

The following simple example will allow us to gain insight into the nature of the surprise losses as well as discuss the possible causes of forecast breakdowns and illustrate the motivation for our asymptotic framework.[5]

[5] The asymptotic distribution of our test statistic in Theorem 1 has been obtained under the assumption of covariance stationarity, which holds only under the null hypothesis. We explore in this section its power in the presence of deviations from this hypothesis.

Let $\tau = 1$ and suppose that Y_t is generated by $Y_t = X_t'\mu_t + \varepsilon_t$, ε_t independent, zero-mean process with variance $\sigma^2_{\varepsilon_t}$. Here we abstract for simplicity from conditional mean and variance dynamics, but we allow for heterogeneity in the form of possibly time-varying unconditional mean μ_t and unconditional variance $\sigma^2_{\varepsilon_t}$. Consider the forecasting model $Y_t = X_t'\beta + u_t$, and let $L(e) = e^2$, so that the forecast is the mean calculated over the rolling window: $\hat{\beta}_m = \left(\frac{1}{m} \sum_{j=t-m+1}^{t} X_j X_j' \right) \frac{1}{m} \sum_{j=t-m+1}^{t} X_j Y_j$. In this case we have, for each $t = m, ..., T-1$,

$$E[SL_{t+1}] = \frac{1}{m} \sum_{j=t-m+1}^{t} [E(Y_{t+1} - X_t'\hat{\beta}_m)^2 - E(Y_j - X_t'\hat{\beta}_m)^2],$$

so that the expectation of the surprise loss captures the difference between out-of-sample and in-sample mean squared error. The expected surprise loss can be decomposed as follows.

Proposition 3 *(a) If $X_t = 1$ (so that the true DGP includes only a constant) then $E[SL_{t+1}] = A_{t+1} + B_{t+1} + C_{t+1}$, where*

$$\begin{aligned} A_{t+1} &= Var(\varepsilon_{t+1}) - \frac{1}{m} \sum_{j=t-m+1}^{t} Var(\varepsilon_j); \\ B_{t+1} &= (\mu_{t+1} - \bar{\mu}_{m,t})^2 - \frac{1}{m} \sum_{j=t-m+1}^{t} (\mu_j - \bar{\mu}_{m,t})^2, \text{ where } \bar{\mu}_{m,t} = \frac{1}{m} \sum_{i=t-m+1}^{t} \mu_i; \\ C_{t+1} &= \frac{2}{m} \sum_{j=t-m+1}^{t} Cov(\varepsilon_j, \hat{\beta}_m). \end{aligned}$$

(b) If X_t is a $(k \times 1)$ vector but $\mu_t = \mu$ and $\sigma^2_{\varepsilon_t} = \sigma^2_\epsilon$ $\forall t$, and the researcher uses a split sample forecasting scheme then $E\left(\overline{SL}_{m,n}\right) = \overline{A} + \overline{B} + \overline{C}$, where

$$\begin{aligned} \overline{A} &= 0, \\ \overline{B} &= 0 \\ \overline{C} &= \frac{k}{m}\sigma^2_\epsilon \end{aligned}$$

4.1 Causes of forecast breakdowns

The decomposition in Proposition 3 allows us to group the possible causes of a forecast breakdown (i.e., situations where the null hypothesis is not satisfied) in three categories:

1. Changes in variance. Component A captures changes in the variance of the disturbances between in-sample and out-of-sample. An *increase* in the variance of the disturbances at time $t + 1$ relative to the average variances of the in-sample disturbances will result in a positive expected surprise loss. Note that changes in variance of opposite signs will cancel out, and thus need not cause a forecast breakdown.

2. Changes in mean. Component B captures differences between the mean of the variable at time $t+1$ and the means of the in-sample observations. If the mean is constant in-sample, B_{t+1} equals the square of the forecast bias, and thus it is positive if any mean changes occur at time $t+1$. Alternatively, if the mean is not constant in-sample, a positive surprise loss can result if the time $t+1$ mean differs more from the means of the in-sample observations than the in-sample means differ from each other. Note that, unlike the case for the variance, both positive and negative changes in mean will result in a positive expected surprise loss.

3. Overfitting. Component C can be interpreted as a measure of "overfitting", loosely defined as a situation where a model has good in-sample fit because it captures the in-sample "noise" more than it captures the "signal". C measures the covariance between the in-sample disturbances and the fitted values from the forecast model, and it is thus an indication of how much the forecast depends on the particular set of in-sample disturbances. Part (b) shows that if there are no changes in the data-generating process, $\overline{C} = \frac{k}{m}\sigma_\varepsilon^2$, and thus the amount of overfitting increases as k, the number of parameters, increases.

4.2 Motivation of asymptotic framework

The decomposition in Proposition 3 helps us to further discuss the motivation for our test.

4.2.1 Overfitting

Suppose that there are no changes in the data-generating process, so that overfitting is the only possible source of a forecast breakdown. In this case, from (a) we have $E[SL_{t+1}] = \frac{2}{m}\sigma_\varepsilon^2$, which implies that $E[\overline{SL}_{m,n}] = \frac{2}{m}\sigma_\varepsilon^2$. Thus, the null hypothesis is satisfied in (a) when $m \to \infty$. In fact, overfitting is a small sample phenomenon, and when there is only one parameter to estimate with an infinite amount of data ($m \to \infty$) then there is no overfitting. However, part (b) shows that when there is either a large number of parameters to estimate or a small sample, i.e. $\frac{k}{m} \simeq \kappa$, then our test has power against forecast breakdowns due to overfitting.

4.2.2 Structural breaks

Let us separately consider one-time and recurrent breaks in part (a) of Proposition 3.

(1) One-time break. Suppose there is a one-time break in the mean at date $m+\tau$:

$$Y_t = \mu_t + \varepsilon_t,\ \mu_t = \begin{cases} 0 & t = 1, \ldots, m+\tau-1 \\ \mu & t = m+\tau, \ldots \end{cases},\ \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2).$$

Proposition 3 implies that $E[\overline{SL}_{m,n}] = \frac{1}{n}\sum_{t=m}^{T-1}(B_{t+1} + C_{t+1}) \equiv \bar{B} + \bar{C}$, where $\bar{B}$ is the contribution of the break in mean and $\bar{C}$ the contribution of overfitting. We can show the following result.

Proposition 4

$$(a)\ \bar{B} = \frac{1}{n}\left(\frac{m^2+3m+2}{6m}\right)\mu^2$$
$$(b)\ \bar{C} = \frac{2}{m}\sigma_\varepsilon^2.$$

The Proposition shows that, if $\frac{n}{m} \to \pi$, a finite number, $\bar{B} \to \frac{1}{6}\pi^{-1}\mu^2$ and $\bar{C} \to 0$ as $m, n \to \infty$. In other words, the null hypothesis is not satisfied in the case of one time breaks, while the contribution of overfitting dies out as the sample size increases. Our test is thus consistent against alternatives for which there is a one-time break. If instead we let $\pi \to \infty$, we would have that $\bar{B} \to 0$ and $\bar{C} \to 0$ as $n \to \infty$, and thus the test would not capture the one-time break.

(2) Recurring breaks. Consider the case where the mean of the process switches between 0 and μ every τ periods:

$$Y_t = \mu_t + \varepsilon_t,\ \mu_t = \begin{cases} 0 & t = 1,\ldots,\tau; 2\tau+1,\ldots,3\tau;\ldots \\ \mu & t = \tau+1,\ldots,2\tau; 3\tau+1,\ldots,4\tau;\ldots \end{cases},\ \varepsilon_t \sim i.i.d.(0,\sigma_\varepsilon^2).$$

Suppose that there are δ breaks in the sample, and for convenience assume that the total sample size T is divisible by τ, so that $T = (\delta+1)\tau$. From Proposition 3, we have as before that $E[\overline{SL}_{m,n}] = \bar{B} + \bar{C}$, but the expression for $\bar{B}$ is now different, as shown in the following proposition.

Proposition 5

$$(a)\ \bar{B} = \begin{cases} \frac{\delta}{n}\frac{m^2+3m+2}{6m}\mu^2 & \text{if } m \le \tau \\ \frac{\delta}{n}\frac{4\tau-2m+6m\tau-3m^2-m^3-4\tau^3+6m\tau^2}{6m^2}\mu^2 + \frac{1}{n}\frac{2\tau^3-3m\tau^2+m^2\tau}{m^2}\mu^2 & \text{if } \tau < m < 2\tau \end{cases};$$
$$(b)\ \bar{C} = \frac{2}{m}\sigma_\varepsilon^2.$$

If $m,n \to \infty$, $\frac{n}{m} \to \pi$, and δ grows with n (and thus τ is finite), result *(a)* implies that $\bar{B} \simeq \frac{1}{6}\delta\pi^{-1}\mu^2$ is $O(1)$, and our test is consistent against forecast breakdowns due to recurring breaks. Also note that the average surprise loss is in expectation higher the larger the size μ of the break and the higher the number of breaks. From *(b)*, we also see that the break in mean does not alter the contribution of overfitting.

Proposition 5 implies that the severity of the forecast breakdown (i.e., the magnitude of the expected surprise loss) induced by a forecast method that assumes a constant mean whereas the mean is affected by recurring breaks (e.g., the data are generated by a Markov-switching process) depends on the values of m and n. Figure 1 below plots the value of $E[\overline{SL}_{m,n}] = \bar{B} + \bar{C}$ as a function of m when $T = 300$, $\mu = 1$, $\sigma_\varepsilon^2 = 1$ and when there are either 5 or 2 breaks (corresponding to $\tau =$ 50 and $\tau = 100$).

INSERT FIGURE 1

The figure shows that, if the data-generating process is affected by recurring breaks (e.g., the data are generated by a Markov-switching process), the forecast method more likely to induce a breakdown is one that assumes a constant mean and uses an estimation window equal to the length of the stable period or smaller (in which case the effect of overfitting also comes into play). In other words, if the estimation window is large enough to include the first break, the forecast method is less likely to undergo a breakdown.

5 Monte Carlo evidence

In this section we investigate the power of our test against a variety of possible causes of forecast breakdown, as well as the relationship between our test and other tests proposed in the literature. As discussed in the previous sections, our test can capture changes in the data-generating process over time and overfitting. That is, it has power in various dimensions, although, for the same reasons, its power will be diluted relative to the most powerful in-sample test designed to capture a specific alternative. Here we provide a simple Monte Carlo experiment to show that our forecast breakdown test has power in all the aforementioned directions, and compare the power of our test with the optimal tests for each alternative (when available) to evaluate the power loss. Throughout, we consider the following tests: our forecast breakdown test, $t_{m,n,\tau}$; a representative test for a structural breaks, the test proposed by Elliott and Muller (2003); and a forecast unbiasedness regression test. The Elliott and Muller's (2003) test is robust to the presence of multiple structural breaks, and still has remarkably high power even in the case of a single break; it therefore provides a natural benchmark for our analysis, where recurrent breaks are important.[6] Forecast unbiasedness regression tests regress the forecast error onto a constant; a zero constant implies that the forecast is unbiased.[7] In the context of this Monte Carlo with a Mean Squared Error Loss function, unbiasedness is one of the properties of an optimal forecast (cfr. Patton and Timmermann (2003)). Unless otherwise specified, in what follows the number of Monte Carlo simulations is $5,000$, and the total sample size is $T = 300$.

We analyze the following designs.

Design 1: A single structural break in the conditional mean. The DGP is the following:

$$y_t = \beta_A \cdot 1\,(t > T/2) + \epsilon_t \tag{9}$$

[6] 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.

[7] We chose to implement a forecast unbiasedness test rather than a Mincer and Zarnowitz regression because the comparison is between unconditional, rather than conditional, tests. The forecast unbiasedness test is based on the same rolling forecasting scheme as that used for the $t_{m,n,\tau}$ rolling test, and utilizes the same correction proposed in Corollary 2 of this paper to ensure good size properties.

where $\epsilon_t \sim N(0,1)$. Note that in this design the conditional mean of the process changes in the middle of the sample.

Design 2: A single structural break in the variance. The DGP is the following:

$$y_t = \epsilon_t, \quad \epsilon_t \sim N(0, 1 + \beta_A \cdot 1(t > T/2)) \tag{10}$$

In this case, the variance of the process changes over time.

Design 3: Model changing over time (misspecification). The DGP is the following:

$$\begin{aligned} y_t &= \beta_A \cdot 1(t \leq T/2) + x_t \cdot 1(t > T/2) + \epsilon_t, \\ x_t &= \rho x_{t-1} + \eta_t, \end{aligned} \tag{11}$$

where $\epsilon_t, \eta_t \sim N(0,1)$ are independent random variables, and $\rho = 0.6$. In this case, the model changes over time; however, the two models are not nested so that the tests developed for structural breaks are not optimal in this context. Nevertheless, tests for structural breaks might still be powerful.

Design 4: Overfitting. The DGP is the following:

$$y_t = x_t'\beta + \epsilon_t \tag{12}$$

where $\epsilon_t \sim N(0,1)$, $\beta = \mathbf{0}_{(k\times1)}$, x_t is a $(k \times 1)$ vector of candidate explanatory variables, $x_t \sim N_k\left(\mathbf{0}_{(k\times1)}, \mathbf{I}_k\right)$, where $\mathbf{0}_{(k\times1)}$ is a $(k \times 1)$ vector of zeros and $\mathbf{I}_k$ is a k-dimensional identity matrix. Here, as k increases, the model used by the researcher progressively contains an increasing number of insignificant parameters, which may deteriorate its forecasting ability.

In addition, we will consider the following designs for the case of multiple structural breaks:

Design 5: Two offsetting structural breaks. The DGP is the following:

$$y_t = 0 \cdot 1(t \leq T/3) + \beta_A \cdot 1(T/3 < t \leq 2T/3) + 0 \cdot 1(t > 2T/3) + \epsilon_t$$

where $\epsilon_t \sim N(0,1)$. This design has two offsetting breaks that may cause forecast breakdowns but, at the same time, they are difficult to detect by using a simple forecast unbiasedness test as the location of the breaks is such that the average forecast error bias is close to zero.

Design 6: Multiple structural breaks in the variance. The DGP is the following:

$$y_t = \epsilon_t, \quad \epsilon_t \sim N(0, 1 + \beta_A \cdot 1(60 \leq t < 120) + \beta_A \cdot 1(180 \leq t < 240)) \tag{13}$$

In this case, the variance of the process experiences 4 breaks at regular intervals.

INSERT FIGURES 2 AND 3, AND TABLES 1(a) AND 1(b)

For designs 1 to 4, Figure 2 compares the power functions of the following tests: Elliott and Muller's (1991) test, labeled "*Elliott-Muller*";[8] our unconditional test $t_{m,n,\tau}$, labeled "$t_{m,n,\tau}$"; and a forecast unbiasedness regression test, labeled "*UNB*". Our Forecast Breakdown test, $t_{m,n,\tau}$, is implemented with a rolling window with $m = 100$, and the forecast horizon is $\tau = 1$. Ideally, a forecaster with a quadratic loss function would like to detect a deviation from the null hypothesis of no forecast failure in all four cases. As Figure 2 shows, the Elliott and Muller's test is powerful against a single break in mean and model changing over time, but not a single break in the variance, nor overfitting. Similar results are valid for forecast unbiasedness test regressions. Our test is instead powerful in all cases. Table 1(a) shows that our $t_{m,n,\tau}$ test has good size properties for various values of m and n.

Table 1(b) reports instead the empirical sizes of our unconditional test $t_{m,n,\tau}$ implemented with a split sample ("$t_{m,n,\tau}$-*split*") scheme and a forecast unbiasedness regression test ("*UNB*") both when the test statistic has been corrected by the value of λ_l as in Theorem (1), labeled "λ_l", or when there is no correction, that is $\lambda_l = 1$, labeled "not corrected". The "not corrected" FFsplit test is the same as Clements and Hendry's (1998) forecast failure test, whereas the "not corrected" unbiasedness test is the one usually used in the literature. The table shows that, unless π is really small, there are big size distortions in the non-corrected test statistic, which can lead to either under-rejections (for the unbiasedness test) or over-rejections (for the $t_{m,n,\tau}$-*split* test) relative to the nominal level. This conforms from our intuition from Theorem 1, as the correction factor λ_l should be, in this case, $(1 + \pi)$, so that when $\pi \simeq 0$ the corrected and the uncorrected test statistic should be the same.

Figure 3 shows instead the results for designs 5 and 6. Again, a forecaster with a quadratic loss function would like to detect a deviation from the null hypothesis of no forecast failure in both cases. As Figure 3 shows, the Elliott and Muller's test is powerful against multiple breaks in the mean, but not breaks in the variance. Interestingly, unlike in the single break case, the forecast unbiasedness test regressions have now no power against two offsetting breaks in the mean. They also have no power against multiple breaks in the variance. Note that, instead, our $t_{m,n,\tau}$ test is again powerful against all deviations from the null hypothesis.

We therefore conclude that our test is robust to all the causes of forecast breakdown listed above, whereas the other tests are not, in particular they may lack power against overfitting and a change in the variance. Elsewhere, we investigate an optimal test that combines power against these different alternatives in an optimal way.

[8] Note that in this preliminary version of the paper, Elliott and Muller's (2003) test is implemented with only up to 10 regressors in design 4, as their critical values are provided only for those cases.

6 Application: The Phillips curve and inflation forecast breakdowns

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

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

$$\pi_t^h - \pi_t = \theta_0 + \theta_1(L)\,u_t + \theta_2(L)(\pi_t - \pi_{t-1}) + e_{t+h} \tag{14}$$

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

When analyzing in retrospect whether unemployment was a useful predictor for inflation, it is important to assess its actual predictive ability at the historic point in time, that is by using only data that were actually available to the policymakers at that time. For example, Orphanides (2001) and Ghysels et al. (2002) analyze the performance of monetary policy rules in the presence of real-time data, and note their relationship with changes in the Fed Chairmen. For that reason, we use real-time monthly data for the U.S. unemployment rate from the real-time database provided by the Federal Reserve Bank of Philadelphia database. The data are discussed in Croushore and Stark (2001), and are available from January 1947 to April 2004 at quarterly vintages starting from November 1965. While the Federal Reserve Bank of Philadelphia also contains a real-time series of consumer prices, unfortunately that is available only from the 1994 vintage, and is thus not useful for our purposes of investigating the Phillips curve forecasting performance over the past three decades. We use instead the real-time monthly database for consumer prices from the Swanson, van Dijk, and Callan dataset (available at http://econweb.rutgers.edu/nswanson/realtime.htm). We focus on seasonally adjusted inflation, as in Stock and Watson (1999).[9] The data span from January 1961 (with a first vintage in February 1978) until December 2001, and the time series for inflation is depicted in Figure 4. Due to the data limitations, we restrict estimation from January 1978 (with a first vintage equal to the first available vintage, February 1978) until December 2001 using quarterly vintages.[10]

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

INSERT TABLE 2 AND FIGURE 4

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

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

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

first estimated regression is from 1959:1 to 1993:1). Both the unconditional and the conditional tests signal a forecast breakdown at the one month horizon, and, in addition, the conditional test signals a forecast breakdown at both the one quarter and the one year horizons. These results confirm our previous findings with real time data, and point in the direction of the existence of a forecast breakdown in the Phillips curve.

Given that our tests detected a forecast breakdown, we next proceed to empirically investigate its possible economic causes. Fisher et al. (2002) argue that periods of low inflation volatility and periods after regime shifts in monetary policy appear to be associated with the break in the forecasting ability of the standard Phillips curve. Thus, we next analyze whether the forecast breakdown of the Phillips curve could have been detected or predicted by using inflation volatility and a measure of changes in the monetary policy behavior of the Fed. Inflation volatility is simply estimated as the sample variance of inflation (using a rolling window estimator). As a measure of changes in the monetary policy behavior of the Fed, we use rolling estimates of the coefficients of the reaction of the Fed Fund Rate (FFR) on the deviation of inflation from its target level and the unemployment gap in the New-Keynesian forward looking monetary policy reaction function analyzed by Clarida et al. (2000). We recursively estimate by GMM the following moment condition (which describes the Fed's monetary policy reaction function):[11]

$$E\left(r_t - (1-\rho)\left[rr^* - (\beta - 1)\,\pi^* + \beta\pi_{t,k} + \gamma x_{t,q}\right] + \rho\,(L)\, r_{t-1} | \Im_t\right) = 0 \qquad (15)$$

where r_t is the nominal FFR; $\pi_{t,k}$ is the percentage change in the price level between periods t and $t+k$ (expressed in annual rates); $x_{t,q}$ is a measure of the average output gap between period t and $t+q$, with the output gap defined as minus the percentage deviation between actual unemployment rate and the corresponding target (in practice, the target is defined as a fitted quadratic function of time); and $\Im_t$ is the information set at time t. As in Clarida et al. (2000), $\rho\,(L) \equiv \rho_1 + \rho_2 L$ contains two lags, rr^* is the average FFR over the period, $\rho \equiv \rho\,(1)$, and the instrument set comprises a constant and four lags of each of the following variables: inflation, the gap, the FFR, the commodity price inflation, the spread between long-term bond rate and the three-month Treasury Bill rate.[12] The target horizon for both the inflation and the unemployment gaps is 1 quarter.

Since we recursively estimate the monetary policy reaction function to analyze whether forecast

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

[12] 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. To make our notation consistent with that in Clarida et al. (2000), we denote the degree of inflation aversion by β. Thus, while in the previous sections β denoted a generic coefficient, in this section β will denote the Fed's inflation aversion.

breakdowns can be linked to changes in monetary policy behavior, we need a long sample. In fact, historically, the Chairmen of the Fed were Burns and Miller from 1960:1 to 1979:2, Volker from 1979:3 to 1987:7 and Greenspan from 1987:7 onwards. Also, as argued in Clarida et al. (2000), "...the sample must be sufficiently long in order to identify the slope coefficients in the policy reaction function as well as the target inflation rate. In fact, estimating the rule over a short sample with little variability in inflation can yield highly misleading results" (cfr. p. 154). Thus, we will address this issue by using finally revised data only.[13]

INSERT FIGURE 5

Figure 5 shows rolling estimates of the structural parameters in the monetary policy reaction function of the Fed. We use an estimation window $m = 132$ (which is the number of observations for the first regression, from 1959:1 to 1979:1, which corresponds to the Burns-Miller period). Even though our database is different from Clarida et al. (2000), our estimates of the structural parameters are similar to theirs, and are also in line with those in Orphanides (2001). In fact, β changes from being approximately 1 in the 1959:1-1979:1 to being approximately 2 in the second part of the sample, and similar considerations hold for the other parameters except ρ, which we find is slightly higher than in Clarida and Gali (2001) in both sub-samples.

We next proceed to use these estimates as explanatory variables and investigate whether they are useful to predict the inflation forecast breakdown. Table 3 shows estimates of the following equations:

$$SL_{t+\tau} = \delta_{0,\tau} + \delta_{1,\tau} h_t + \epsilon_{t,h} \tag{16}$$

where h_t is either $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$, or $\widehat{\sigma}^2_{\pi,t}$ (the rolling estimates of the parameters in (15) and of the inflation volatility, $\widehat{\sigma}^2_{\pi,t}$), and $\tau = 1$, 3, 12 months. The table reports estimates of the coefficient $\delta_{1,\tau}$ as well as (in parentheses) the p-values associated with testing whether $\delta_{1,\tau}$ equals zero.[14] For comparison purposes, the table also reports the values of the $t_{m,n,\tau}$ test (which correspond to the case in which eq. (16) contains only the constant, $\delta_{0,\tau}$). It is clear that the degree of inflation targeting smoothing operated by the central bank ($\widehat{\rho}_t$) and the degree of inflation volatility ($\widehat{\sigma}^2_{\pi,t}$)

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

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

have a significant impact in explaining the out of sample forecast breakdown at the 12 month horizon, whereas inflation volatility and the degree of the Fed's risk aversion to the unemployment gap ($\widehat{\gamma}_t$) are significant at the one month horizon. To conclude, we also consider estimating eq. (16) with a constant and all the rolling estimates of the parameters of the monetary policy reaction function of the Fed and jointly test their significance. The column labeled "*Joint*" reports the values of the $W_{m,n,\tau}$ test statistic (p-values in parentheses). It is clear that these variables were jointly significant at conventional significance levels for all horizons.

INSERT TABLE 2

Since we found significant economic predictors of future surprise losses, we next proceed to use this information to predict future surprise losses under different scenarios for horizons up to 12 months starting from the last estimation period (July 2004). Results are shown in Figure 6. The figure plots estimates of the forecast losses together with 95% standard error bands under various scenarios: one in which the Fed's aversion to inflation (measured by β) is higher than the one estimated in June 2004, one in which the Fed's reaction to the output gap is lower (measured by γ), and finally one in which the degree of interest rate smoothness decreases (the smoothness is measured by ρ). This exercise is performed as follows. First, we estimate the following regression:

$$SL_{t+\tau} = \delta_{0,\tau} + \{\widehat{\alpha}_t, \widehat{\beta}_t, \widehat{\gamma}_t, \widehat{\rho}_t\}' \boldsymbol{\delta}_{1,\tau} \qquad (17)$$

where $\widehat{\alpha}_t$ is the rolling estimate of the constant term in eq. (15). We consider the estimated values of these parameters in 2004:6, which are, respectively: {0.04, 0.77, 2.86, 0.98}. Then we fix all of them at their estimated values except one, which we modify according to the chosen scenario. We finally forecast the future losses by substituting these values back into eq. (17).[15] The exercise is plausible because the estimated parameters in the Fed's reaction function are structural parameters (cfr. Clarida et al. (2000)).

In the first scenario, we consider a path in which β becomes 0.83 (Figure 6, panel (a)) or 2.15 (Figure 6, panel (b)), which are, respectively, the average values estimated by Clarida et al. (2000) for their Pre-Volker (1960:1-1979:2) and their Volker-Greenspan (1979:3-1996:4) sub-samples (see also Figure 5). This scenario corresponds to an increase in inflation aversion in the Fed's preferences, which could cause a Fed's tighter monetary policy that would not simply be the Fed's reaction to a higher future expected inflation rate. The increase is mild in the first case and large in the second case, as the most recent estimate is 0.77. Figure 6 shows that the forecasted surprise losses are significantly different from zero at horizons of six months and beyond in the first case, and

[15] The 95% standard error bands are calculated by using a HAC (with $(h-1)$ bandwidth) estimate of the variance of the residuals of eq. (17) and then using a normal approximation, as discussed in Section 3.

after one month in the second case. This result confirms our intuition, as we would expect more evidence of forecast breakdown for larger changes in the monetary policy behavior. In a second scenario, depicted in Figure 6, panel (c), we consider a situation in which γ decreases from 2.86 to 0.93, the value estimate by Clarida et al. (2000) for the (1979:3-1996:4) period, and corresponds to lower sensitivity to the output gap. In this case, there is no evidence of forecast breakdown at any horizon. Similar results hold if we let γ be equal to the value estimated by Clarida et al. (2000) during the Pre-Volker period. Finally, in Figure 6, panel (d), we show a scenario corresponding to a lower interest rate smoothing (ρ decreases from 0.98 to 0.79). In this case, there is striking evidence of forecast breakdown at most horizons.

INSERT FIGURE 6

7 Conclusions

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

Among the advantages of our test, we note that the framework is very general and allows for very general estimation and forecasting procedures, as well as forecasters' loss functions; it is very easy to implement; and it captures a variety of causes of forecast breakdowns. At the same time, while the test has power against many possible causes of forecast breakdown, it is not necessarily the most powerful test against each of them. Also, in relation to this issue, while our test provides a first step in detecting and predicting forecast breakdowns, an important question that arises when the researcher finds predictive breakdown is: "what are the causes of predictive breakdown, and what to do next?". Deriving such optimal tests and providing insights on the causes of the predictive breakdown is an interesting avenue for future research, and is currently under investigation by the authors.

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

8 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.

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

Chatfield, C. (1993), "Calculating Interval Forecasts", *Journal of Business and Economic Statistics* 11, 121-144.

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.

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

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

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

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

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

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

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

Hansen, B. (2004), "Interval Forecasts and Parameter Uncertainty", *mimeo*, University of Wisconsin-Madison.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix. Proofs

Notation and definitions. Let

$$A_l \equiv n^{-1/2} \sum_{j=m+1}^{T} SL_t = \overline{SL}_{m,n}$$

$$\text{and } A_h \equiv n^{-1/2} \sum_{j=m+1}^{T} H_t,$$

and let superscripts *rec*, *roll* or *split* denote the method used, so that:

$$\begin{aligned}
SL_t^{rec}\left(\widehat{\beta}_t\right) &= L_{t+1}^{oos}\left(\widehat{\beta}_t\right) - \frac{1}{m}\sum_{j=1}^{t} L_t^{is}\left(\widehat{\beta}_t\right), \\
SL_t^{roll}\left(\widehat{\beta}_t\right) &= L_{t+1}^{oos}\left(\widehat{\beta}_t\right) - \frac{1}{m}\sum_{j=t-m+1}^{t} L_t^{is}\left(\widehat{\beta}_t\right), \text{ and} \\
SL_t^{split}\left(\widehat{\beta}_t\right) &= L_{t+1}^{oos}\left(\widehat{\beta}_t\right) - \frac{1}{m}\sum_{j=1}^{m} L_t^{is}\left(\widehat{\beta}_t\right),
\end{aligned}$$

where we allow the possibility that the in-sample loss may be different from the out-of-sample loss, and that the in-sample loss used for evaluation may be different from the loss function used in the estimation of the parameters β.

Direct calculations show that

$$\begin{aligned}
A_l &= A_{is} + A_{oos}, \\
A_{is} &= n^{-1/2}\mathbf{a}_{is}'\mathbf{l}_{is}, \\
A_{oos} &= n^{-1/2}\mathbf{a}_{oos}'\mathbf{l}_{oos}, \\
A_h &= n^{-1/2}\mathbf{a}_h'\mathbf{h},
\end{aligned}$$

where $\mathbf{l}^{is}$, $\mathbf{l}^{oos}$ and $\mathbf{h}$ are vectors that contain the time series observations of the in-sample loss, the out-of-sample loss, and the $h_t's$:

$$\begin{aligned}
\mathbf{l}_{is} &= [l_1^{is}, l_2^{is}, ..., l_m^{is}, l_{m+1}^{is}, l_{m+2,}^{is} ... l_T^{is}]'; \\
\mathbf{l}_{oos} &= [l_1^{oos}, l_2^{oos}, ..., l_m^{oos}, l_{m+1}^{oos}, l_{m+2}^{oos}, ..., l_T^{oos}]'; \\
\mathbf{h} &= [h_1, h_2, ..., h_m, h_{m+1}, h_{m+2}, ..., h_T]';
\end{aligned}$$

Instead, the exact expressions for $\mathbf{a}_{is}', \mathbf{a}_{oos}', \mathbf{a}_h'$ depend on the type of estimator used and are given below:

Recursive case:

$\mathbf{a}_{is}^{rec} = [-a_{m,0}, -a_{m,0}, ... - a_{m,0}, -a_{m,1}, -a_{m,2}, ... - a_{m,n}]$,

$\mathbf{a}_{oos}^{rec} = [0, 0, ..., 0, 1, 1, ...1]'$;
$\mathbf{a}_{h}^{rec} = [a_{m,0}, a_{m,0}, ...a_{m,0}, a_{m,1}, a_{m,2}, ...a_{m,n}]$;
where $a_{m,j} = \left(\frac{1}{m+j} + \frac{1}{m+j+1} + ... + \frac{1}{T}\right)$
Rolling case, $\pi < 1$:

$$\mathbf{a}_{is}^{roll} = [\underbrace{-\frac{1}{m}, -\frac{2}{m}, ..., -\frac{n}{m}}_{n \text{ terms}}, \underbrace{-\frac{n}{m}, .., -\frac{n}{m}}_{m-n \text{ terms}}, \underbrace{-\frac{m+n-(m+1)}{m} ... -\frac{m+n-(m+n-1)}{m}}_{n-1 \text{ terms}}];$$

$$\mathbf{a}_{oos}^{roll} = [\underbrace{0, 0, ..., 0, 0, ..., 0}_{m \text{ terms}}, \underbrace{1, ..., 1}_{n-1 \text{ terms}}];$$

$$\mathbf{a}_{h}^{roll} = [\underbrace{\frac{1}{m}, \frac{2}{m}, ..., \frac{n}{m}}_{n \text{ terms}}, \underbrace{\frac{n}{m}, ..., \frac{n}{m}}_{m-n \text{ terms}}, \underbrace{\frac{(n-1)}{m}, ... \frac{1}{m}}_{n-1 \text{ terms}}];$$

Rolling case, $\pi > 1$:

$$\mathbf{a}_{is}^{roll} = [\underbrace{-\frac{1}{m}, -\frac{2}{m}, ..., -\frac{(m-1)}{m}, -\frac{m}{m}}_{m \text{ terms}}, \underbrace{-1, ..., -1}_{n-m-1 \text{ terms}}, \underbrace{-\frac{m}{m}, -\frac{m-1}{m}, ..., -\frac{1}{m}}_{m \text{ terms}}];$$

$$\mathbf{a}_{oos}^{roll} = [\underbrace{0, 0, ..., 0}_{m \text{ terms}}, \underbrace{1, ..., 1, 1, ..., 1}_{n-1 \text{ terms}}];$$

$$\mathbf{a}_{h}^{roll} = [\underbrace{\frac{1}{m}, \frac{2}{m}, ... \frac{m}{m}}_{m \text{ terms}}, \underbrace{1, ..., 1}_{n-m-1 \text{ terms}}, \underbrace{\frac{m}{m}, \frac{(m-1)}{m}, ... \frac{1}{m}}_{m \text{ terms}}];$$

Split sample case:

$$\mathbf{a}_{is}^{split} = [\underbrace{-\frac{n}{m}, -\frac{n}{m}, ..., -\frac{n}{m}}_{m \text{ terms}}, \underbrace{0, 0...0}_{(n-1) \text{ terms}}];$$

$\mathbf{a}_{oos}^{split} = [0, 0, ...0, 1, 1, ...1]$;
$\mathbf{a}_{h}^{split} = [1, 1, ...1, 0, 0, ...0]$.

Note that under H_0 we are in a covariance stationary framework. As in West (1996), we take a mean value expansion of $SL_t(\widehat{\beta}_t)$ around β^* (the proof is similar to that provided in West (1996), eq. (4.1), and is therefore omitted):

$$\begin{aligned} & n^{1/2}[SL_t(\widehat{\beta}_t) - E(SL_t)] \\ = \; & n^{-1/2} \sum_{t=m+1}^{T} [SL_t(\beta_t^*) - E(SL_t)] + FBn^{-1/2} \sum_{t=m+1}^{T} H_t + o_p(1) \\ = \; & A_{oos} + A_{is} + FBA_h \end{aligned}$$

which shows that $n^{1/2}[SL_t(\widehat{\beta}_t) - E(SL_t)]$ is ultimately a function of $l_t \equiv l_t(\beta^*)$ and $h_t = h_t(\beta^*)$.

Let $\Gamma_{wz}(k) \equiv E(w_t - E(w_t))(z_{t-k} - E(z_{t-k}))$ where w_t and z_t can be either l_t^{in}, l_t^{oos} or h_t, and let $S_{wz} \equiv \sum_{k=-\infty}^{\infty} \Gamma_{wz}(k)$.

In what follows, we will use the following Lemma 1, whose proof follows from the fact that the covariances are absolutely summable, and the particular structure of the problem:

Lemma 1. $\frac{1}{m^2 n} \sum_{k=-\infty}^{\infty} (d_k^{wz} - \lambda_{wz}) \Gamma_{wz}(k) \to 0$ *as* $m, n \to \infty$*, where* $\lambda_{wz} \equiv \sum_{j=k+1}^{T} A_w(j) A_z(j)$.

Lemma 2.

$$[n^{-1/2}\sum_{t=m+1}^{T}\left[l_t^{oos}-E\left(l_t^{oos}\right)\right], n^{-1/2}\sum_{t=m+1}^{T}\left[L_t^{is}-E\left(L_t^{is}\right)\right], n^{-1/2}\sum_{t=m+1}^{T}H_t]' \overset{A}{\sim} N\left(0,V\right),$$

where L_t^{is} is the average in-sample loss at time t evaluated at β^*, the $\lambda's$ have been defined in Theorem 1, and

$$V\equiv\begin{pmatrix}\lambda_{oos}S_{oos} & \lambda_{oos,is}S_{oos,is} & \lambda_{oos,h}S_{oos,h}\\ \lambda_{oos,is}\left(S_{oos,is}\right)' & \lambda_{is}S_{is} & \lambda_{is,h}S_{is,h}\\ \lambda_{oos,h}\left(S_{oos,h}\right)' & \lambda_{is,h}\left(S_{is,h}\right)' & \lambda_{hh}S_{hh}\end{pmatrix}$$

Proof. By definition, $Cov\left(A_w,A_z\right)=n^{-1}\sum_{k=-(T-1)}^{T-1}\Gamma_{wz}\left(k\right)d_k$, where $d_k^{wz}=\sum_{j=k+1}^{T}A_w\left(j\right)A_z\left(j-k\right)$. It follows from Lemma 1 that $Cov\left(A_w,A_z\right)=\lambda_{wz}\frac{1}{n}\sum_{k=-(T-1)}^{T-1}\Gamma_{wz}\left(k\right)+o_p\left(1\right)=\lambda_{wz}S_{wz}$. We proceed by analytically calculating λ_{wz} for each type of estimator and for $w,z=l^{is},l^{oos},h$. Note that in the special case where the loss function used for in-sample and out-of-sample evaluation is the same, we need only $Cov\left(A_{is}+A_{oos}\right)=Var\left(A_{is}\right)+Var\left(A_{oos}\right)+2Cov\left(A_{is},A_{oos}\right)=\lambda_lS_l$, where S_l is the long run variance of the losses, $l=l^{is}=l^{oos}$, and this case will be separately calculated as well.

(i) Recursive case:

$\mathbf{a}_{is}^{rec}=[-a_{m,0},-a_{m,0},...-a_{m,0},-a_{m,1},-a_{m,2},...-a_{m,n}]$,

$\mathbf{a}_{oos}^{rec}=[0,0,...,0,1,1,...1]'$;

$\mathbf{a}_{h}^{rec}=[a_{m,0},a_{m,0},...a_{m,0},a_{m,1},a_{m,2},...a_{m,n}]$

$\lambda_{is}=n^{-1}\mathbf{a}_{is}^{rec\prime}\mathbf{a}_{is}^{rec}=n^{-1}\left(ma_{m,0}^2+\sum_{j=1}^{n}a_{m,j}^2\right)$

$\lambda_{oos}=n^{-1}\mathbf{a}_{oos}^{rec\prime}\mathbf{a}_{oos}^{rec}=n^{-1}\left(n\right)$

$\lambda_{h}=n^{-1}\mathbf{a}_{h}^{rec\prime}\mathbf{a}_{h}^{rec}=n^{-1}\left(ma_{m,0}^2+\sum_{j=1}^{n}a_{m,j}^2\right)$

$\lambda_{is,oos}=n^{-1}\mathbf{a}_{is}^{rec\prime}\mathbf{a}_{oos}^{rec}=n^{-1}\left(-\sum_{j=1}^{n}a_{m,j}\right)$

$\lambda_{is,h}=n^{-1}\mathbf{a}_{is}^{rec\prime}\mathbf{a}_{h}^{rec}=n^{-1}\left(-ma_{m,0}-\sum_{j=1}^{n}a_{m,j}^2\right)=-Cov\left(A_{is},A_{is}\right)$

$\lambda_{oos,h}=n^{-1}\mathbf{a}_{h}^{rec\prime}\mathbf{a}_{oos}^{rec}=n^{-1}\left(\sum_{j=1}^{n}a_{m,j}\right)=-Cov\left(A_{is},A_{oos}\right)$

Direct calculations show that:[16]

$n^{-1}\sum_{j=1}^{n}a_{m,j}\simeq 1-\pi^{-1}\ln\left(1+\pi\right)$,

$n^{-1}\sum_{j=1}^{n}a_{m,j}^2\simeq 2\left[1-\pi^{-1}\ln\left(1+\pi\right)\right]-\pi^{-1}\ln^2\left(1+\pi\right)$,

$a_{m,0}\simeq\ln\left(1+\pi\right)$

Thus:

[16] These approximations follow from the same calculations as in West (1996):
$n^{-1}\sum_{j=1}^{n}a_{m,j}=n^{-1}\sum_{j=1}^{n-1}\frac{j}{m+j}\simeq n^{-1}[(n-1)-m\ln\left(m+n-1\right)+m\ln m]=n^{-1}[(n-1)-m\ln\left(\frac{m+n-1}{m}\right)]\simeq 1-\pi^{-1}\ln\left(1+\pi\right)$

$n^{-1}\sum_{j=1}^{n}a_{m,j}^2=n^{-1}\sum_{j=1}^{n}\left(\sum_{k=j}^{n-1}\frac{1}{m+k}\right)^2\simeq n^{-1}\int_0^n\ln^2\left(\frac{m+n}{m+j}\right)\simeq 2\left[1-\pi^{-1}\ln\left(1+\pi\right)\right]-\pi^{-1}\ln^2\left(1+\pi\right)$

$\lambda_{is} = n^{-1}\left(ma_{m,0}^2 + \sum_{j=1}^{n} a_{m,j}^2\right) = 2\left[1 - \pi^{-1}\ln(1+\pi)\right]$

$\lambda_{oos} = 1$

$\lambda_h = 2\left[1 - \pi^{-1}\ln(1+\pi)\right]$

$\lambda_{is,oos} = -\left[1 - \pi^{-1}\ln(1+\pi)\right]$

$\lambda_{is,h} = -2\left[1 - \pi^{-1}\ln(1+\pi)\right]$

$\lambda_{oos,h} = \left[1 - \pi^{-1}\ln(1+\pi)\right]$

In the special case where the loss function used for in-sample and out-of-sample evaluation is the same: $\lambda_l = 2\left[1 - \pi^{-1}\ln(1+\pi)\right] + 1 - 2\left[1 - \pi^{-1}\ln(1+\pi)\right] = 1$

(ii) Rolling case, $\pi \leq 1$:

Note that: $\sum_{j=W}^{Z} j^2 \simeq \frac{1}{3}\left(Z^3 - W^3\right)$ and $\sum_{j=W}^{Z} j \simeq \frac{1}{2}\left(Z^2 - W^2\right)$. Therefore:

$\lambda_{is} = n^{-1}\left[2\sum_{j=1}^{n-1}\left(\frac{j}{m}\right)^2 + 1 + (m-n)\left(\frac{n}{m}\right)^2\right] \simeq \pi - \frac{1}{3}\pi^2$

$\lambda_{oos} = n^{-1}(n-1) \simeq 1$

$\lambda_h = Var(A_{is}) \simeq \pi - \frac{1}{3}\pi^2$

$\lambda_{is,oos} = -n^{-1}\sum_{j=1}^{n-1}\left(\frac{j}{m}\right) \simeq -\frac{1}{2}\pi$

$\lambda_{is,h} = -Var(A_{is}) \simeq -\pi + \frac{1}{3}\pi^2$

$\lambda_{oos,h} = -Cov(A_{is}, A_{oos}) \simeq \pi$

In the special case where the loss function used for in-sample and out-of-sample evaluation is the same: $\lambda_l \simeq 1 - \frac{1}{3}\pi^2$

(iii) Rolling case, $\pi > 1$:

$\lambda_{is} = n^{-1}\left[2\sum_{j=1}^{m}\left(\frac{j}{m}\right)^2 + (n-m-1)\right] \simeq 1 - \frac{1}{3}\pi^{-1}$

$\lambda_{oos} = n^{-1}(n-1) \simeq 1$

$\lambda_h = Var(A_{is}) = 1 - \frac{1}{3}\pi^{-1}$

$\lambda_{is,oos} = n^{-1}\left[-(n-m-1) - \sum_{j=1}^{m}\left(\frac{j}{m}\right)\right] \simeq -1 + \frac{1}{2}\pi^{-1}$

$\lambda_{is,h} = -Var(A_{is}) = -\left(1 - \frac{1}{3}\pi^{-1}\right)$

$\lambda_{oos,h} = -Cov(A_{is}, A_{oos}) = 1 - \frac{1}{2}\pi^{-1}$

In the special case where the loss function used for in-sample and out-of-sample evaluation is the same: $\lambda_l \simeq \frac{2}{3}\pi^{-1}$

(iv) Split sample case:

$\lambda_{is} = n^{-1}m\left(\frac{n}{m}\right)^2 \simeq \pi$

$\lambda_{oos} = 1$

$\lambda_h = \pi^{-1}$

$\lambda_{is,oos} = 0$

$\lambda_{is,h} = -1$

$\lambda_{oos,h} = 0$

In the special case where the loss function used for in-sample and out-of-sample evaluation is the same: $\lambda_l \simeq 1+\pi$ ■

Proof of Theorem 2. From Lemma 2, we have

$$\begin{aligned} & n^{1/2}\left[\overline{SL}\left(\widehat{\beta}\right) - E\left(SL\left(\beta^*\right)\right)\right] \\ = \; & [1,-1,FB][n^{-1/2}\sum_{t=m+1}^{T}\left[l_t^{oos} - E\left(l_t^{oos}\right)\right], n^{-1/2}\sum_{t=m+1}^{T}\left[L_t^{is} - E\left(L_t^{is}\right)\right], n^{-1/2}\sum_{t=m+1}^{T}H_t]' \\ & \overset{A}{\sim} N(0,\sigma^2_{m,n}) \end{aligned}$$

where

$$\begin{aligned} \sigma^2_{m,n} \; \equiv \; & \lambda_{oos}S_{oos} + \lambda_{is}S_{is} + \lambda_{hh}FBS_{hh}B'F' + \lambda_{oos,is}\left(S_{oos,is} + S'_{oos,is}\right) + \\ & \lambda_{oos,h}\left(S_{oos,h}B'F' + FBS'_{oos,h}\right) + \lambda_{is,h}\left(S_{is,h}B'F' + FBS'_{is,h}\right), \end{aligned}$$

and the asymptotic normality follows from McLeish (1975) and Wooldridge and White (1988), as in West (1996). ■

Proof of Theorem 1. It follows directly as a special case of Theorem 2. ■

Proof of Proposition 3. (a) $E[SL_{t+1}] = \frac{1}{m}\sum_{j=t-m+1}^{t}[E(Y_{t+1}-\hat{\beta}_m)^2 - E(Y_j - \hat{\beta}_m)^2] \equiv \frac{1}{m}\sum_{j=t-m+1}^{t}[MSE^{oos} - MSE_j^{is}]$. We have

$$\begin{aligned} MSE^{oos} \; = \; & Var\left(Y_{t+1} - \frac{1}{m}\sum_{j=t-m+1}^{t}Y_j\right) + \left(E\left(Y_{t+1} - \frac{1}{m}\sum_{j=t-m+1}^{t}Y_j\right)\right)^2 = \\ & Var(\varepsilon_{t+1}) + Var\left(\frac{1}{m}\sum_{j=t-m+1}^{t}\varepsilon_j\right) + \left(\mu_{t+1} - \frac{1}{m}\sum_{j=t-m+1}^{t}\mu_j\right)^2; \end{aligned}$$

$$\begin{aligned} MSE_j^{is} \; = \; & Var\left(Y_j - \frac{1}{m}\sum_{i=t-m+1}^{t}Y_i\right) + \left(E\left(Y_j - \frac{1}{m}\sum_{i=t-m+1}^{t}Y_i\right)\right)^2 = \\ & Var(\varepsilon_j) + Var\left(\frac{1}{m}\sum_{i=t-m+1}^{t}\varepsilon_i\right) - 2Cov(\varepsilon_j,\hat{\beta}_m) + \left(\mu_j - \frac{1}{m}\sum_{i=t-m+1}^{t}\mu_i\right)^2, \end{aligned}$$

from which the proposition is easily verified.

(b) That $A_{t+1} = B_{t+1} = 0$ immediately follows from above. Collect ε_t and X_t into the vector and matrices ε and X. Note that:

$MSE^{is} = \frac{1}{m}\widehat{\varepsilon}'\widehat{\varepsilon} = \frac{1}{m}\varepsilon'\varepsilon - 2\frac{1}{m}\varepsilon'X\left(\frac{1}{m}X'X\right)^{-1}\frac{1}{m}X'\varepsilon + \left(\hat{\beta}_m - \mu\right)'\left(\frac{1}{m}X'X\right)\left(\hat{\beta}_m - \mu\right) = \frac{1}{m}\varepsilon'\varepsilon - 2\frac{1}{m}\varepsilon'X\left(\frac{1}{m}X'X\right)^{-1}\frac{1}{m}X'\varepsilon + \frac{1}{m}\varepsilon'X\left(\frac{1}{m}X'X\right)^{-1}\frac{1}{m}X'\varepsilon = \frac{1}{T}\varepsilon'\varepsilon - \frac{1}{T}\left(\frac{1}{\sqrt{T}}X'\varepsilon\right)'\left(\frac{1}{T}X'X\right)^{-1}\left(\frac{1}{\sqrt{T}}X'\varepsilon\right)$

$E\left(MSE^{is}\right) = \sigma_\epsilon^2 - \sigma_\epsilon^2 E\left(\chi_k^2\right) = \left(1 - \frac{k}{T}\right)\sigma_\epsilon^2.$
$E\left(MSE^{oos}\right) = \sigma_\epsilon^2$
$E\left(MSE^{oos} - MSE^{is}\right) = \frac{k}{m}\sigma_\epsilon^2$ ■

Proof of Proposition 4. *(a)* From Proposition 3,

$$\frac{1}{n}\sum_{t=m}^{T-1} B_{t+1} = \frac{1}{n}\sum_{t=m}^{T-1}\left[(\mu_{t+1} - \bar{\mu}_{m,t})^2 - \frac{1}{m}\sum_{j=t-m+1}^{t}(\mu_j - \bar{\mu}_{m,t})^2\right], \tag{18}$$

where $\bar{\mu}_{m,t} = \frac{1}{m}\sum_{i=t-m+1}^{t}\mu_i$. We have

$$\bar{\mu}_{m,t} = \begin{cases} 0 & t \le m+\tau-1 \\ \frac{1}{m}\sum_{j=m+\tau}^{t}\mu = \frac{t-m-\tau+1}{m}\mu & m+\tau \le t \le 2m+\tau-2 \\ \mu & t \ge 2m+\tau-1 \end{cases};$$

$$(\mu_j - \bar{\mu}_{m,t})^2 = \begin{cases} 0 & t \le m+\tau-1 \\ (\mu - \frac{t-m-\tau+1}{m}\mu)^2 & m+\tau \le t \le 2m+\tau-2 \\ 0 & t \ge 2m+\tau-1 \end{cases};$$

$$\frac{1}{m}\sum_{j=t-m+1}^{t}(\mu_j - \bar{\mu}_{m,t})^2 = \begin{cases} 0 & t \le m+\tau-1 \\ \frac{2m+\tau-t-1}{m}(\frac{t-m-\tau+1}{m}\mu)^2 + & \\ \frac{t-m-\tau+1}{m}(\mu - \frac{t-m-\tau+1}{m}\mu)^2 & m+\tau \le t \le 2m+\tau-2 \\ 0 & t \ge 2m+\tau-1 \end{cases};$$

$$(\mu_{t+1} - \bar{\mu}_{m,t})^2 = \begin{cases} 0 & t \le m+\tau-2 \\ \mu^2 & t = m+\tau-1 \\ (\mu - \frac{t-m-\tau+1}{m}\mu)^2 & m+\tau \le t \le 2m+\tau-2 \\ 0 & t \ge 2m+\tau-1 \end{cases}.$$

We therefore have

$$B_{t+1} = \begin{cases} 0 & t \le m+\tau-2 \\ \mu^2 & t = m+\tau-1 \\ (\mu - \frac{t-m-\tau+1}{m}\mu)^2(1 - \frac{t-m-\tau+1}{m}) & \\ \quad -\frac{2m+\tau-t-1}{m}(\frac{t-m-\tau+1}{m}\mu)^2 & m+\tau \le t \le 2m+\tau-2 \\ 0 & t \ge 2m+\tau-1 \end{cases},$$

so that

$$\bar{B} = \frac{1}{n}\left\{\mu^2 + \sum_{t=m+\tau}^{2m+\tau-2}\left[(\mu - \frac{t-m-\tau+1}{m}\mu)^2(1 - \frac{t-m-\tau+1}{m}) - \frac{2m+\tau-t-1}{m}(\frac{t-m-\tau+1}{m}\mu)^2\right]\right\},$$

which simplifies to *(a)*.

(b) $C_{t+1} = \frac{2}{m} \sum_{j=t-m+1}^{t} Cov(\varepsilon_j, \hat{\beta}_m) = \frac{2}{m^2} \sum_{j=t-m+1}^{t} Var(\varepsilon_j) = \frac{2}{m}\sigma_\varepsilon^2$ and thus $\bar{C} = \frac{2}{m}\sigma_\varepsilon^2$. ■

Proof of Proposition 5. *(a)* From Proposition 3,

$$\bar{B} = \frac{1}{n}\sum_{t=m}^{T-1}\left[(\mu_{t+1} - \bar{\mu}_{m,t})^2 - \frac{1}{m}\sum_{j=t-m+1}^{t}(\mu_j - \bar{\mu}_{m,t})^2\right],$$

where $\bar{\mu}_{m,t} = \frac{1}{m}\sum_{i=t-m+1}^{t}\mu_i$. We have that

$$\bar{\mu}_{m,t} = \begin{cases} 0 & t = m \\ \frac{1}{m}\sum_{j=m+1}^{t}\mu = \frac{t-m}{m}\mu & m+1 \le t \le 2m-1 \\ \mu & t = 2m \\ \frac{1}{m}\sum_{j=t-m+1}^{2m}\mu = \frac{3m-t}{m}\mu & 2m+1 \le t \le 3m-1 \\ 0 & t = 3m \\ \frac{1}{m}\sum_{j=3m+1}^{t}\mu = \frac{t-3m}{m}\mu & 3m+1 \le t \le 4m-1 \\ \dots & \end{cases}$$

and thus

$$\frac{1}{m}\sum_{j=t-m+1}^{t}(\mu_j - \bar{\mu}_{m,t})^2 = \begin{cases} 0 & t = m \\ \frac{1}{m}\left[\sum_{j=t-m+1}^{m}\frac{(t-m)^2}{m^2}\mu^2 + \sum_{j=m+1}^{t}(\mu - \frac{t-m}{m}\mu)^2\right] & m+1 \le t \le 2m-1 \\ 0 & t = 2m \\ \frac{1}{m}\left[\sum_{j=t-m+1}^{2m}(\mu - \frac{3m-t}{m}\mu)^2 + \sum_{j=2m+1}^{t}\frac{(3m-t)^2}{m^2}\mu^2\right] & 2m+1 \le t \le 3m-1 \\ 0 & t = 3m \\ \frac{1}{m}\left[\sum_{j=t-m+1}^{3m}\frac{(t-3m)^2}{m^2}\mu^2 + \sum_{j=3m+1}^{t}(\mu - \frac{t-3m}{m}\mu)^2\right] & 3m+1 \le t \le 4m-1 \\ \dots & \end{cases}.$$

Further,

$$(\mu_{t+1} - \bar{\mu}_{m,t})^2 = \begin{cases} \mu^2 & t = m \\ (\mu - \frac{t-m}{m}\mu)^2 & m+1 \leq t \leq 2m-1 \\ \mu^2 & t = 2m \\ \frac{(3m-t)^2}{m^2}\mu^2 & 2m+1 \leq t \leq 3m-1 \\ \mu^2 & t = 3m \\ (\mu - \frac{t-3m}{m}\mu)^2 & 3m+1 \leq t \leq 4m-1 \\ \ldots & \end{cases}$$

which implies

$$B_{t+1} = \begin{cases} \mu^2 & t = m \\ \frac{(2m-t)(3m-2t)}{m^2}\mu^2 & m+1 \leq t \leq 2m-1 \\ \mu^2 & t = 2m \\ \frac{(3m-t)(5m-2t)}{m^2}\mu^2 & 2m+1 \leq t \leq 3m-1 \\ \mu^2 & t = 3m \\ \frac{(4m-t)(7m-2t)}{m^2}\mu^2 & 3m+1 \leq t \leq 4m-1 \\ \ldots & \end{cases}$$

We thus have

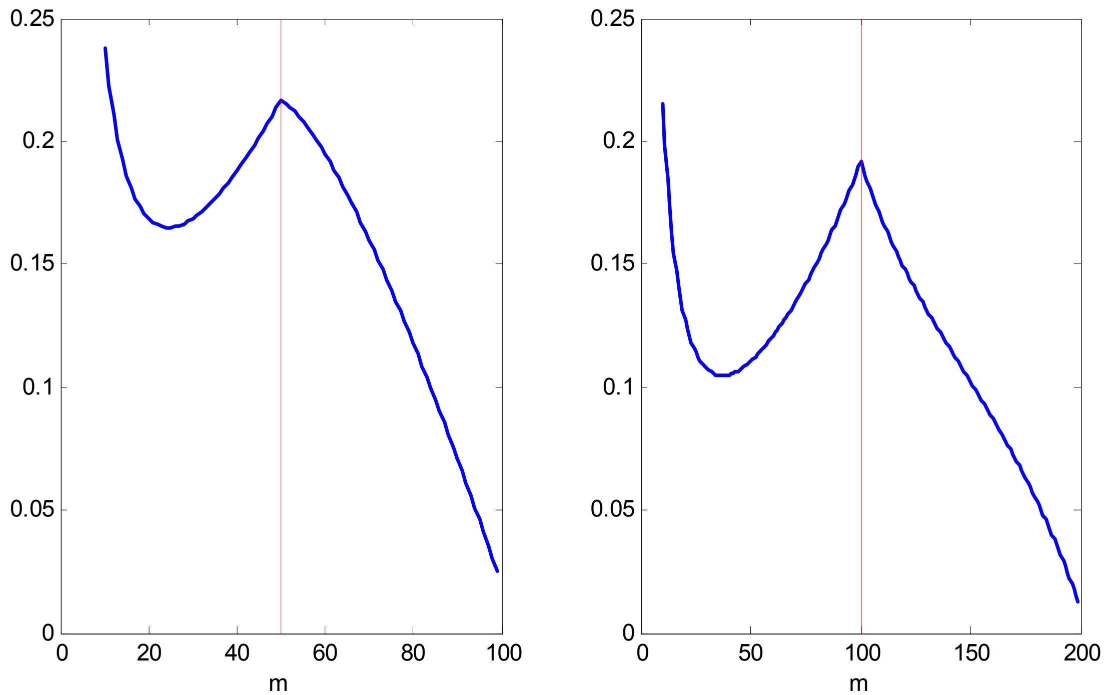
$$\begin{aligned} \bar{B} &= \frac{1}{n}\left\{\delta\mu^2 + \sum_{k=1}^{\delta}\sum_{t=km+1}^{(k+1)m-1}\frac{[(k+1)m-t][(2k+1)m-2t]}{m^2}\mu^2\right\} \\ &= \frac{1}{n}\left\{\delta\mu^2 + \sum_{k=1}^{\delta}\left[(m-1)(k+1)(2k+1)\mu^2 - \frac{4k+3}{m}\mu^2\sum_{t=km+1}^{(k+1)m-1}t + \frac{2}{m^2}\mu^2\sum_{t=km+1}^{(k+1)m-1}t^2\right]\right\} \quad (19) \end{aligned}$$

Using the fact that $\sum_{t=a}^{b} t = \frac{b(b+1)}{2} - \frac{(a-1)a}{2}$ and that $\sum_{t=a}^{b} t^2 = \frac{b(b+1)(2b+1)}{6} - \frac{(a-1)a(2a-1)}{6}$ and simplifying terms gives the results in *(a)*.

(b) Same as Proposition 4 *(b)*.

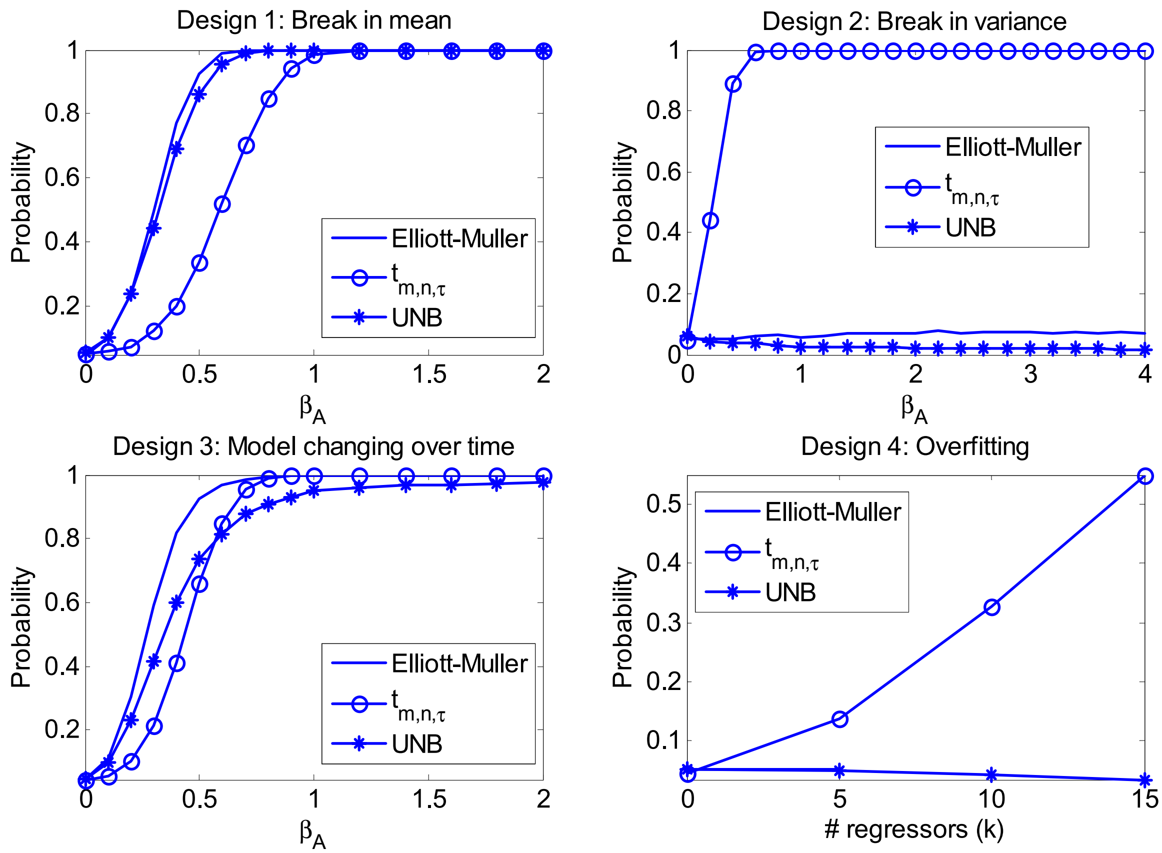
■

Figure 1. Expected surprise loss in the presence of recurring breaks



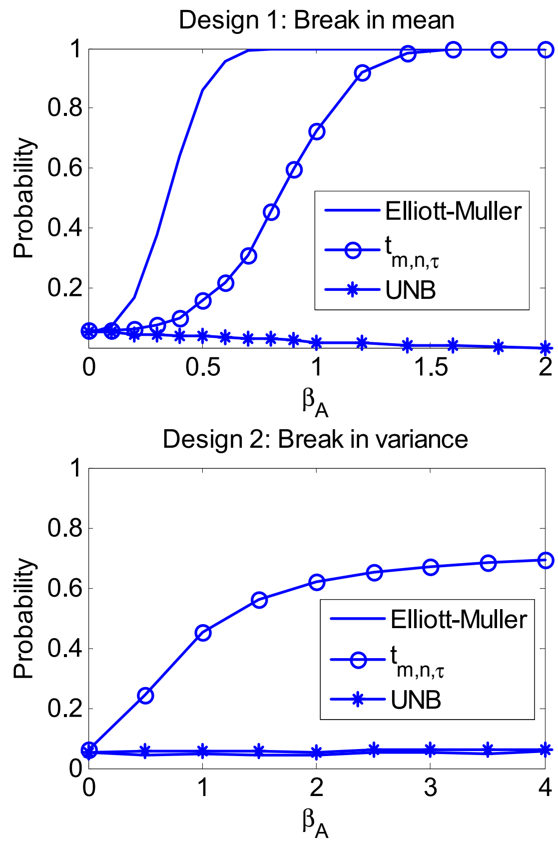
Notes to Figure 1. The figure plots the expression for $E[\overline{SL}_{m,n}]$ derived in Proposition 5 as a function of m when $T = 300$, $\mu = 1$, σ_{ε}^{2} and $\tau = 50$, 100 (corresponding respectively to 5 and 2 breaks). The vertical line in each panel is the first break date.

Figure 2. Monte Carlo comparison of power functions (single break)



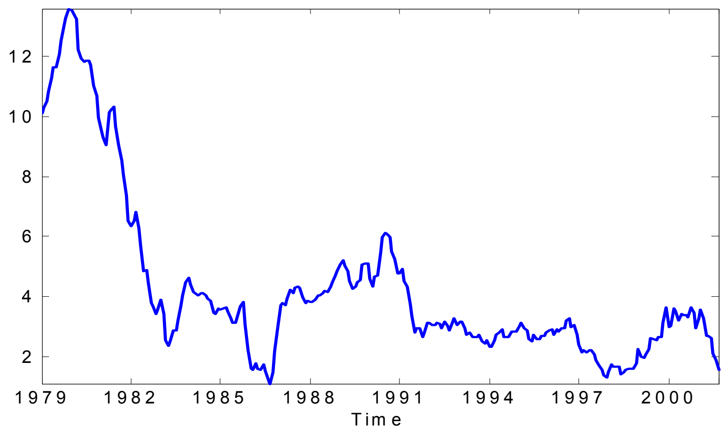
Notes to Figure 2. The figure shows power functions of the following tests: Elliott and Muller's (2003) test, labeled "$Elliott - Muller$"; our unconditional test $t_{m,n,\tau}$ (implemented with a rolling window with $m = 100$, $n = 300$, $\tau = 1$), labeled "$t_{m,n,\tau}$"; and a forecast unbiasedness regression test, labeled "UNB".

Figure 3. Monte Carlo comparison of power functions (multiple breaks)



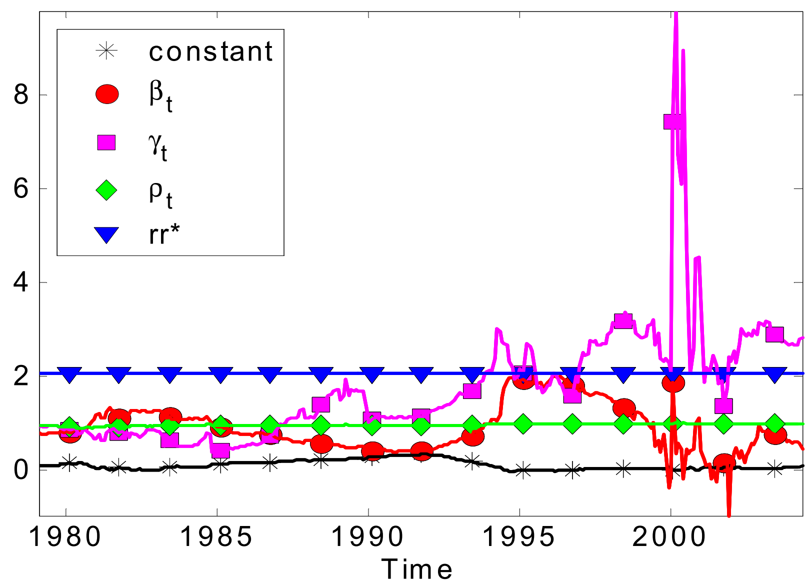
Notes to Figure 3. The figure shows power functions of the following tests: Elliott and Muller's (2003) test, labeled "$Elliott - Muller$"; our unconditional test $t_{m,n,\tau}$ (implemented with a rolling window with $m = 100$, $n = 300$, $\tau = 1$), labeled "$t_{m,n,\tau}$"; and a forecast unbiasedness regression test, labeled "UNB".

Figure 4. Inflation with real-time data



Notes to Figure 4. The figure shows seasonally adjusted inflation at the monthly frequency.

Figure 5. Rolling estimates of the structural parameters of the Fed's reaction function



Notes to Figure 5. The figure shows rolling estimates of the structural parameters in the monetary policy reaction function of the Fed, eq. (15).

Figure 6. Forecasted surprise losses under different scenarios.

Figure 6(a). The Fed's higher inflation aversion scenario ($\beta = 0.83$)

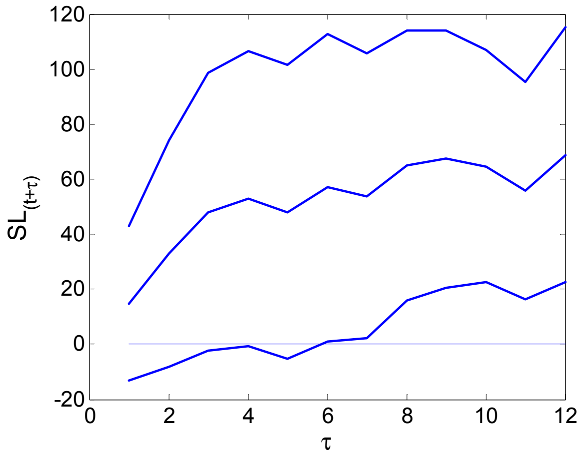


Figure 6(b). The Fed's higher inflation aversion scenario ($\beta = 2.15$)

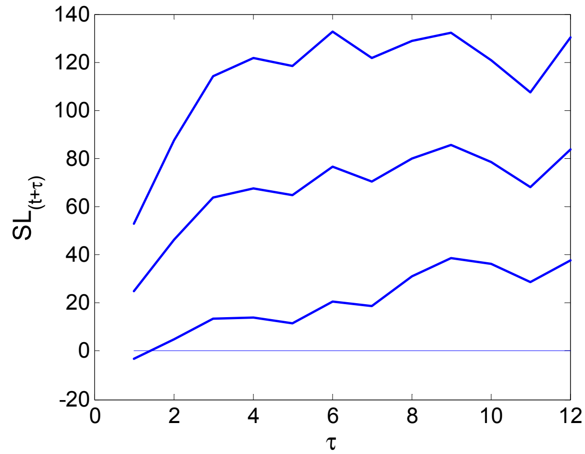


Figure 6(c). The Fed's lower sensitivity to the output gap scenario ($\gamma = 0.93$)

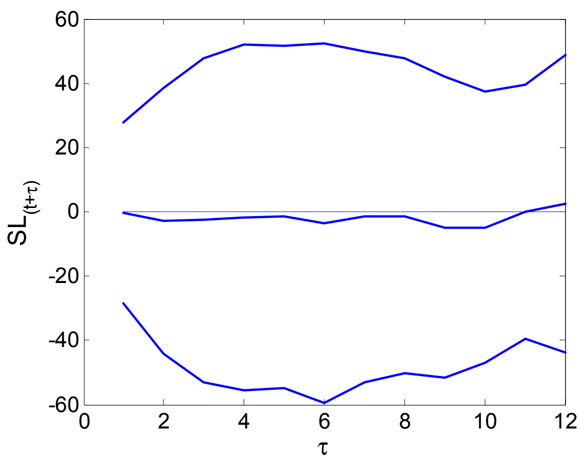
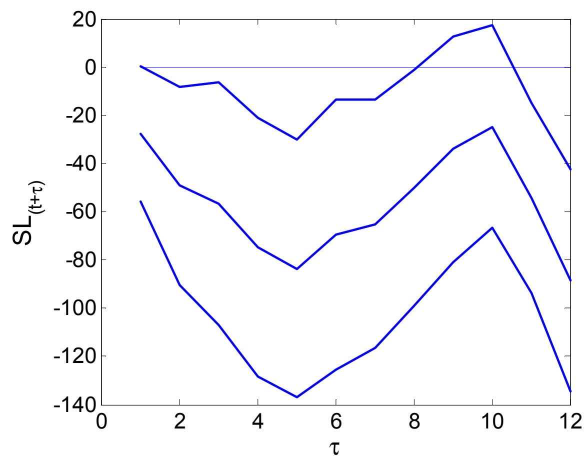


Figure 6(d). The Fed's lower interest rate smoothing scenario ($\rho = 0.79$)



Notes to Figure 6. The figure shows forecasted surprise losses for the out-of-sample period at various forecast horizons, from one to twelve months (corresponding to the period 2004:7-2005:6) under: the Fed's higher inflation aversion scenario ($\beta = 0.83$ – i.e. Pre-Volker – panel (a), and $\beta = 2.15$ – i.e. (1979:3-1996:4) – panel (b); our most recent rolling estimate is $\beta = 0.77$), the lower sensitivity to the output gap scenario ($\gamma = 0.93$, panel (c) – the current estimated value is 2.86), and the lower interest rate smoothing ($\rho = 0.79$, panel (d) – the current estimated value is 0.98) together with 95% standard error bands. For visual comparison purposes, we also plot a horizontal line at zero. Here, we used $q_u = q_\pi = 3$.

Table 1(a). Size of Forecast Breakdown test

m	n	π	*Elliott-Muller*	$t_{m,n,\tau}$ *seq.*	*split*	*roll*	*UNB* *roll*
100	33	0.330	0.055	0.095	0.094	0.093	0.058
100	50	0.500	0.052	0.073	0.073	0.070	0.059
100	100	1.000	0.047	0.060	0.056	0.054	0.051
100	150	1.500	0.054	0.056	0.056	0.059	0.051
100	200	2.000	0.050	0.048	0.047	0.062	0.051
100	300	3.000	0.049	0.049	0.053	0.072	0.050
150	50	0.33	0.048	0.081	0.081	0.077	0.059
150	75	0.5	0.049	0.065	0.065	0.063	0.055
150	150	1	0.055	0.055	0.055	0.054	0.049
150	225	1.5	0.048	0.053	0.052	0.059	0.051
150	300	2	0.050	0.053	0.051	0.063	0.054
150	450	3	0.054	0.057	0.050	0.076	0.057
200	67	0.34	0.046	0.066	0.064	0.063	0.051
200	100	0.5	0.046	0.061	0.059	0.056	0.051
200	200	1	0.041	0.055	0.057	0.053	0.054
200	300	1.5	0.053	0.055	0.056	0.055	0.046
200	400	2	0.050	0.051	0.052	0.052	0.043
200	600	3	0.053	0.052	0.046	0.063	0.053
300	100	0.33	0.054	0.068	0.067	0.068	0.054
300	150	0.5	0.050	0.054	0.054	0.055	0.050
300	300	1	0.049	0.054	0.053	0.058	0.051
300	450	1.5	0.049	0.055	0.054	0.056	0.047
300	600	2	0.048	0.047	0.048	0.057	0.048
300	900	3	0.049	0.057	0.056	0.063	0.053

Note to Table 1(a). The table reports empirical size of the following tests: Elliott and Muller's (2003) test, labeled "$Elliott - Muller$"; our unconditional test $t_{m,n,\tau}$ implemented with either a sequential ("seq"), a split sample ("split") or a rolling ("roll") scheme; and a forecast unbiasedness regression test, labeled "UNB".

Table 1(b). Size of Forecast Breakdown test

			$t_{m,n,\tau}$-*split*		*UNB*	
m	n	π	λ_{ll}	*not corrected*	λ_{ll}	*not corrected*
100	100	1	0.058	0.173	0.054	0.019
100	150	1.5	0.052	0.210	0.055	0.004
100	200	2	0.054	0.268	0.05	0.001
100	300	3	0.052	0.325	0.053	0
150	150	1	0.056	0.168	0.051	0.016
150	225	1.5	0.049	0.217	0.053	0.002
150	300	2	0.052	0.261	0.051	0.001
150	450	3	0.050	0.328	0.049	0
200	67	0.335	0.076	0.114	0.056	0.054
200	100	0.5	0.060	0.119	0.058	0.047
200	200	1	0.055	0.176	0.056	0.018
200	300	1.5	0.053	0.217	0.049	0.004
200	400	2	0.051	0.269	0.046	0.001
200	600	3	0.050	0.329	0.054	0
300	100	0.333	0.063	0.100	0.056	0.050
300	150	0.5	0.055	0.117	0.052	0.040
300	300	1	0.053	0.172	0.050	0.016
300	450	1.5	0.056	0.220	0.050	0.005
300	600	2	0.056	0.260	0.045	0.001
300	900	3	0.050	0.330	0.049	0
500	167	0.334	0.055	0.090	0.052	0.048
500	250	0.5	0.058	0.113	0.052	0.038

Note to Table 1(b). The table reports empirical size of our unconditional test $t_{m,n,\tau}$ implemented with a split sample ("$t_{m,n,\tau}$-split") scheme and a forecast unbiasedness regression test ("*UNB*") both when the test statistic has been corrected by the value of λ_l as in Theorem (1), labeled "λ_l", or when there is no correction, that is $\lambda_l = 1$, labeled "not corrected".

Table 2. P-values of Forecast Breakdown tests

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

Notes to Table 2. The table reports p-values for both the unconditional ($t_{m,n,\tau}$) and the conditional ($W_{m,n,\tau}$) tests. The $W_{m,n,\tau}$ test is implemented with a constant and one lagged value of the surprise loss, SL_{t-1}. We used $m = 60$, $n = 95$ in the real-time data case, and $m = 241$ and $n = 546$ in the revised data case. The forecast horizons are $\tau = 1$, 3 and 12 months (note that real-time data are available only at a quarterly frequency, so in that case only results for $\tau = 3$ months (1 quarter) and $\tau = 12$ months (4 quarters) are reported). q_u and q_π are, respectively, the number of lags used for unemployment and for inflation; the row labeled "BIC" reports results for the case in which the lag length is determined at every t by the BIC with a maximum of three lags. The one-step ahead forecasts begin in 1993:1 (the date has been chosen to correspond to a change in monetary policy according to Fisher et al. (2002)).

Table 3. The conditional Forecast Breakdown test with monetary policy and inflation variance explanatory variables, eq. (16).

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

Notes to Table 3. The table reports the empirical results for the estimation of equation (16). The first regression dates have been selected according to the findings in Fisher et al. (2002). The regressions for the Forecast Breakdown test $W_{m,n,\tau}$ contain a constant and each of the following regressors: SL_t (the lagged value of the surprise loss), $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$ (the rollingly estimated structural parameters in the monetary policy reaction function of the Fed), and $\widehat{\sigma}^2_{\pi,t-1}$ (the inflation volatility). The column labeled "Joint" reports instead the joint test on a constant and all the parameters $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$. P-values of the $W_{m,n,\tau}$ test statistic (with a HAC bandwidth equal to $(\tau - 1)$) for testing whether the explanatory variable is insignificant are reported in parentheses. For comparison purposes, we also report results for the unconditional $t_{m,n,\tau}$ test. q_u and q_π are, respectively, the number of lags used for unemployment and for inflation; rows labeled "BIC" report results for the case in which the lag length is determined at every t by the BIC with a maximum of three lags. The horizons are one month ($\tau = 1$), one quarter ($\tau = 3$), and one year ($\tau = 12$).