THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies Limited (2009) **76**, 669–705 0034-6527/09/00411011$02.00
© 2009 The Review of Economic Studies Limited
doi: 10.1111/j.1467-937X.2009.00545.x

Detecting and Predicting Forecast Breakdowns

RAFFAELLA GIACOMINI
UCL/CEMMAP

BARBARA ROSSI
Duke University

First version received April 2006; *final version accepted August* 2008 *(Eds.)*

We propose a theoretical framework for assessing whether a forecast model estimated over one period can provide good forecasts over a subsequent period. We formalize this idea by defining a forecast breakdown as a situation in which the out-of-sample performance of the model, judged by some loss function, is significantly worse than its in-sample performance. Our framework, which is valid under general conditions, can be used not only to detect past forecast breakdowns but also to predict future ones. We show that main causes of forecast breakdowns are instabilities in the data-generating process and relate the properties of our forecast breakdown test to those of structural break tests. The empirical application finds evidence of a forecast breakdown in the Phillips' curve forecasts of U.S. inflation, and links it to inflation volatility and to changes in the monetary policy reaction function of the Fed.

1. INTRODUCTION

This paper proposes a new method for evaluating a forecasting model for a macroeconomic or financial variable. There is a large literature claiming that certain models are good at predicting macroeconomic variables such as output growth and inflation (Stock and Watson, 2003; Clark and McCracken, 2006) and that a range of variables have predictive power for stock market returns (*e.g.* the references in Goyal and Welch (2004) and Campbell and Thompson (2005)). These claims are based either on some measure of a model's in-sample fit (most of the literature on stock return predictability), or on the model's out-of-sample performance (Stock and Watson, 2003). The robustness of these results has been however recently challenged. On the one hand, Goyal and Welch (2004) showed that, for models of stock returns, good in-sample fit does not necessarily imply good out-of-sample performance. On the other hand, even models that fare well out-of-sample may not do so when different subsamples are considered (Stock and Watson, 2004). Underlying these findings is the possibility that the economy – and the forecasting ability of models – may not be stable over time.

From the perspective of the forecaster, it is thus important to know whether a model estimated over one period can provide good forecasts over a subsequent period. The goal of this paper is to develop a formal testing framework for answering this question. Note that our question is different from asking whether the model is a good approximation of the data-generating process, or whether it produces forecasts that are optimal for a given loss function. Rather, our concern is with whether the model's future performance is consistent with what is expected on the basis of its past performance, which hinges on the success of the model at adapting to changes in the economy. This in turn reflects a desire to mimic the environment

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

faced by actual forecasters, where models are likely misspecified, variables are difficult to forecast, and data-generating processes may be unstable, so that consistency with expected performance can be viewed as a minimal requirement that a forecasting model should satisfy.

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

Our notion of a forecast breakdown is a formalization and generalization of what Clements and Hendry (1998, 1999) called a "forecast failure", described as a "deterioration in forecast performance relative to the anticipated outcome" (Clements and Hendry, 1999, p. 1). We formalize the definition of a forecast breakdown by comparing the model's out-of-sample performance to its in-sample performance computed in one of three ways: (1) over a fixed initial sample ("fixed" scheme); (2) over a rolling window that includes only most recent observations ("rolling scheme"); and (3) over an expanding window that includes all observations from the beginning of the sample ("recursive scheme"). The fixed scheme presumes an interest in comparing performance before and after a specific date, whereas the rolling and recursive schemes mimic forecasting in real time.

We propose a forecast breakdown test based on the intuition that, in the absence of a forecast breakdown, the difference between expected out-of- and in-sample performances should be close to zero. We provide the appropriate estimator for the asymptotic variance needed in the construction of the tests statistic, which depends on the forecasting scheme. Our test is valid under general assumptions. In particular, we allow the data to be heterogeneous (*e.g.* the variables in the model can have time-varying marginal distributions) and impose only weak restrictions on the loss function used for evaluation and on the type of estimators used in constructing the forecasts. In the paper, we focus on the case in which parameter estimation uncertainty is asymptotically irrelevant, which occurs, for example, in the common situation in which the same loss function is used for estimation and evaluation (*e.g.* OLS and quadratic loss). In the Appendix, we present the general result for non-vanishing estimation uncertainty. From a technical point of view, we use a similar asymptotic framework to that developed by West (1996), although we generalize it beyond the covariance stationarity assumption in West (1996). This generalization is of separate theoretical interest in itself, and it is crucial in our framework because our emphasis on structural instability is incompatible with the assumption of covariance stationarity.

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

The two closest literatures to the present paper are the literature on forecast optimality testing (Mincer and Zarnowitz, 1969; Patton and Timmermann, 2006; Elliott, Komunjer and Timmermann, 2005) and the literature on structural break testing (Andrews, 1993; Dufour, Ghysels and Hall, 1990; Elliott and Muller, 2006). Regarding the former, we point out that the same theory derived here can be applied to forecast optimality testing, after suitably redefining the loss function and the null hypothesis. For example, a forecast unbiasedness test is related to a forecast breakdown test assessing whether the first moment properties of the forecast errors are consistent in-sample and out-of-sample. A forecast rationality test can be obtained

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

following our procedure for predicting forecast breakdowns. Our tests for forecast unbiasedness and forecast rationality take estimation uncertainty into account (unlike Elliott, Komunjer and Timmermann (2005), but like West and McCracken (1998)), and extend the validity of West and McCracken (1998) to an environment in which the forecast losses are not necessarily stationary.

Regarding the relationship with the structural break testing literature, we note that the focus of our forecast breakdown test is on stability of forecast performance, which is loss-specific and allows for model misspecification. This makes our test flexible and widely applicable. For a particular loss, for example a quadratic loss, and under correct specification, we show that a forecast breakdown is caused by breaks in the conditional mean parameters and/or in the unconditional variance of the model errors (note that GARCH does not cause a forecast breakdown, as long as the unconditional variance is constant). This means that one could in principle indirectly test for a forecast breakdown by testing jointly for structural breaks in the parameters and in the variance. However, this indirect approach fails to recognize that these two types of breaks could affect the forecast performance in opposite directions and therefore not necessarily cause a deterioration in the forecast performance of the model. For example, a forecast bias induced by a break in parameters could be in part or fully offset by a decrease in the variance of the errors, in a way that leaves the mean squared forecast error unchanged. Further, we show that forecast breakdowns can be caused by larger parameter breaks than those captured by a structural break test, and thus a structural break test may find a break that is too small to affect the forecast performance. A final advantage of the forecast breakdown test is its robustness to the presence of unstable regressors, whereas most structural break tests cannot distinguish between instability in the model's parameters and instability in the distribution of the regressors (Hansen, 2000).

Another difference with structural break tests is that they typically focus on past stability. Instead, an innovation of our approach with useful practical implications is the possibility of predicting future forecast breakdowns. This relates our approach to that in Pesaran, Pettenuzzo and Timmermann (2007) and Koop and Potter (2007), who model breaks in parameters as functions of a latent variable. An advantage of our framework is that it is general and flexible, as it does not require correct specification of the model and it allows the user to directly link the forecast performance for a specific loss function to observable, rather than latent, economic variables.

To illustrate the methods proposed in this paper, we investigate whether there is evidence of a forecast breakdown in the Phillips curve model for predicting inflation in the U.S. Using both real-time and revised data, we find some empirical evidence in favour of a forecast breakdown in the Phillips curve. We further investigate whether monetary policy parameters would have been useful predictors of forecast breakdowns and find that inflation volatility as well as changes in the monetary policy behaviour of the Fed played a key role.

2. DETECTING FORECAST BREAKDOWNS

2.1. *Description of the environment*

Let $\mathrm{W} \equiv \{\mathrm{W}_t : \Omega \longrightarrow \mathbb{R}^{s+1}, s \in \mathbb{N}, t = 1, \ldots, T\}$ be a stochastic process defined on a complete probability space $(\Omega, \mathcal{F}, P)$ and partition the observed vector W_t as $\mathrm{W}_t \equiv (\mathrm{Y}_t, \mathrm{X}_t')'$, where $Y_t : \Omega \to \mathbb{R}$ is the variable of interest and $\mathrm{X}_t : \Omega \to \mathbb{R}^s$ is a vector of predictors.

We generate a sequence of $\tau-$step-ahead forecasts of $\mathrm{Y}_{t+\tau}$ using an out-of-sample procedure, which involves dividing the sample of size T into an in-sample window of size m and an out-of-sample window of size $n = T - m - \tau + 1$. Which data constitute the in-sample

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

window depends on the forecasting scheme. We allow for three forecasting schemes: (1) a fixed forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, m$; (2) a rolling forecasting scheme, where the in-sample window at time t contains observations indexed $t-m+1, \ldots, t$; and (3) a recursive forecasting scheme, where the in-sample window includes observations indexed $1, \ldots, t$.

We let $f_t(\widehat{\beta}_t)$ be the time-t forecast produced by estimating a model over the in-sample window at time t, with $\widehat{\beta}_t$ indicating the $k \times 1$ parameter estimate. We assume that multi-step forecasts are produced by the "direct method" (that is, the model specifies the relationship between Y_t and $\mathrm{X}_{t-\tau}$). Each time t forecast corresponds to a sequence of in-sample fitted values $\hat{y}_j(\widehat{\beta}_t)$, with j varying over the in-sample window.

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

2.2. *Forecast breakdown test*

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

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

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

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

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

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

where β^* (defined formally in Assumption A3) is the pseudo-true value of the parameter estimate, assumed to be constant under the null hypothesis.

The forecast breakdown test statistic is

$$t_{m,n,\tau} = n^{1/2}\, \overline{SL}_{m,n}/\hat{\sigma}_{m,n}, \tag{4}$$

where the expression for the asymptotic variance estimator $\hat{\sigma}^2_{m,n}$ is given in Section 2.6.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

A level α test rejects the null hypothesis whenever $t_{m,n,\tau} > z_\alpha$, where z_α is the $(1-\alpha)$-th quantile of a standard normal distribution. In the remainder of the paper, we focus on a one-sided test to reflect the assumption that a lower-than-expected loss may be desirable and thus does not constitute a forecast breakdown. In certain applications, however, a two-sided test may be desirable. For example, for an investor forming a portfolio based on forecasts of stock returns, the precision of the forecast is a key determinant of how much risk exposure to accept. Hence, if the out-of-sample forecast error variance is smaller than anticipated, this results in an opportunity cost: had the forecaster known about the lower forecast error variance, he could have chosen a different portfolio allocation.[1] The asymptotic justification for the forecast breakdown test is provided by Theorem 1.

2.3. *A step-by-step procedure to implement the forecast breakdown test*

The following step-by-step procedure shows how to implement the forecast breakdown test for a forecast horizon τ, a linear model with k regressors ($\mathrm{Y}_t = \mathrm{X}'_{t-\tau}\beta + \varepsilon_t$), a quadratic loss $L_{t+\tau}(\widehat{\beta}_t) = \left(\mathrm{Y}_{t+\tau} - \mathrm{X}'_t\widehat{\beta}_t\right)^2$, and under the assumption of covariance stationarity of L_t.

- Step 1: Compute the sequence of OLS estimators $\widehat{\beta}_t,\ t = m,\ m+1, \ldots, T-\tau$, by regressing Y_t on $\mathrm{X}_{t-\tau}$, where:
  - Recursive scheme: $\widehat{\beta}_t = \left(\sum_{s=1}^{t-\tau} X_s X'_s\right)^{-1} \sum_{s=1}^{t-\tau} X_s Y_{s+\tau}$.
  - Rolling scheme: $\widehat{\beta}_t = \left(\sum_{s=t-m+1}^{t-\tau} X_s X'_s\right)^{-1} \sum_{s=t-m+1}^{t-\tau} X_s Y_{s+\tau}$.
  - Fixed scheme: $\widehat{\beta}_t = \widehat{\beta}_m = \left(\sum_{s=1}^{m-\tau} X_s X'_s\right)^{-1} \sum_{s=1}^{m-\tau} X_s Y_{s+\tau}$.
- Step 2: Compute the sequence of in-sample average losses, $\overline{L}_t(\widehat{\beta}_t),\ t = m, \ldots, T-\tau$, where:
  - Recursive scheme: $\overline{L}_t(\widehat{\beta}_t) = t^{-1} \sum_{j=1}^{t-\tau} \left(Y_{j+\tau} - X'_j\widehat{\beta}_t\right)^2$.
  - Rolling scheme: $\overline{L}_t(\widehat{\beta}_t) = m^{-1} \sum_{j=t-m+1}^{t-\tau} \left(Y_{j+\tau} - X'_j\widehat{\beta}_t\right)^2$.
  - Fixed scheme: $\overline{L}_t(\widehat{\beta}_t) = m^{-1} \sum_{j=1}^{m-\tau} \left(Y_{j+\tau} - X'_j\widehat{\beta}_t\right)^2$.
- Step 3: Compute the sequence of out-of-sample losses, $L_{t+\tau}(\widehat{\beta}_t)$, for $t = m, \ldots, T-\tau$: $L_{t+\tau}(\widehat{\beta}_t) = \left(Y_{t+\tau} - X'_t\widehat{\beta}_t\right)^2$.
- Step 4: Calculate $\overline{SL}_{m,n} \equiv n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t)$, where $SL_{t+\tau}(\widehat{\beta}_t) = L_{t+\tau}(\widehat{\beta}_t) - \overline{L}_t(\widehat{\beta}_t)$.
- Step 5: Estimate the asymptotic standard deviation $\hat{\sigma}_{m,n}$ as $\sqrt{\lambda S_n^{LL}}$, where: $\lambda = 1 + \frac{n}{m}$ for the fixed scheme; for the rolling scheme, $\lambda = 1 - \frac{1}{3}\left(\frac{n}{m}\right)^2$ if $n < m$ and $\lambda = \frac{2}{3}\frac{m}{n}$ if $n \geq m$; $\lambda = 1$ for the recursive scheme; S_n^{LL} is a heteroskedasticity- and autocorrelation-consistent (HAC) estimator applied to the sequence of demeaned out-of-sample losses, $\widetilde{L}_{t+\tau} \equiv L_{t+\tau}(\widehat{\beta}_t) - n^{-1} \sum_{j=m}^{T-\tau} L_{j+\tau}(\widehat{\beta}_j)$, $t = m, \ldots, T-\tau$, for example: $S_n^{LL} = \sum_{j=-p_n+1}^{p_n-1} \left(1 - |\frac{j}{p_n}|\right) n^{-1} \sum_{t=m+j}^{T-\tau} \widetilde{L}_{t+\tau}\widetilde{L}_{t+\tau-j}$, where p_n is a bandwidth that increases with the sample size (Newey and West, 1987).
- Step 6: Compute the test statistic $t_{m,n,\tau} = n^{1/2}\ \overline{SL}_{m,n}/\hat{\sigma}_{m,n}$.
- Step 7: If a correction for overfitting is desired (Section 4), consider instead the test statistic: $t^c_{m,n,\tau} = \left(n^{1/2}\ \overline{SL}_{m,n} - c\right)/\hat{\sigma}_{m,n}$, where $c = 2\gamma\ \ tr\left((T-\tau)^{-1}\left(\Sigma_{t=1}^{T-\tau} X_t X'_t\right)\ \widehat{V}_T^{\beta}\right)$;

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

$\gamma = n^{1/2}/m$ for the fixed and rolling schemes; $\gamma = n^{-1/2}\ln(1+n/m)$ for the recursive scheme and $\widehat{V}_T^{\beta}$ is a consistent estimator of the asymptotic variance of $\widehat{\beta}_T = \left(\sum_{s=1}^{T-\tau} X_s X_s'\right)^{-1} \sum_{s=1}^{T-\tau} X_s Y_{s+\tau}$. For example, under the additional assumption of conditional homoskedasticity, $c = 2\gamma k\,(T-\tau)^{-1} \sum_{t=1}^{T-\tau} \left(\mathrm{Y}_{t+\tau} - X_t'\widehat{\beta}_T\right)^2$.

2.4. *Relationship with the literature*

Our definition of forecast breakdowns formalizes the notion of reliability of a forecasting model as a systematic difference between the model's in- and out-of-sample performances. Some of the advantages of this definition of reliability are that it is loss-specific and that it allows for model misspecification. This means that our approach is tailored to the forecaster's decision-making problem.

For a specific loss function, and assuming correct specification, one could in principle relate our forecast breakdown test to existing tests for breaks in model parameters. For example, let us consider the simple case of a quadratic loss, a fixed forecasting scheme, a linear model $\mathrm{Y}_t = \mathrm{X}_{t-1}'\beta_t + \varepsilon_t$, with independent and identically distributed regressors and errors, and assume there is a one-time break of size $n^{-1/4}\Delta\beta$ in β_t and a one-time break in the variance of the errors of size $n^{-1/2}\Delta\sigma^2$, occurring at the same time. One can show that in this case the numerator of one of our test statistics (the overfitting-corrected forecast breakdown test) in expectation equals $\Delta\sigma^2 + 0{\cdot}5\Delta\beta' E(\mathrm{X}_t\mathrm{X}_t')\Delta\beta$, which implies that both a "large" break in parameters and a "small" break in error variance can make the test statistic be greater than zero, and thus result in a forecast breakdown. This decomposition suggests that one could in principle test for a forecast breakdown by jointly testing for a break in β and a break in the variance of the errors. This however fails to recognize that these two types of breaks may have opposite effects on the forecast performance, and thus not necessarily result in a forecast breakdown (*e.g.* it could happen that $\Delta\sigma^2 \leq -0{\cdot}5\Delta\beta' E(X_t X_t')\Delta\beta$). One can further see that a forecast breakdown is caused by breaks in parameters of greater magnitude than those considered by previous structural break tests such as Elliott and Muller (2006) (*i.e.* here the breaks are of magnitude $n^{-1/4}\Delta\beta$ rather than $n^{-1/2}\Delta\beta$). As a result, previous tests may detect breaks that do not necessarily cause forecast breakdowns. A final difference is that most existing structural break tests are only valid under the restrictive and unrealistic assumption that the marginal distribution of the regressors is constant over time. Our test, in contrast, is robust to the presence of instability in the marginal distribution of the regressors.

Besides relating our approach to detecting past forecast breakdowns to previous structural break tests, we can further relate our approach to predicting future forecast breakdowns (Section 5) to the literature that predicts future structural breaks in the model's parameters by modelling the parameter evolution using a meta distribution for the breaks (Pesaran, Pettenuzzo and Timmermann, 2006; Koop and Potter, 2007). A drawback of the latter approach is that it relies on the specification for the meta distribution of the parameters being correct. We instead propose directly relating the difference between in- and out-of-sample performances (for a given loss function) to explanatory variables, and use this relationship to forecast the future behaviour of the forecast losses. This allows us to answer empirically relevant questions such as, for example, whether the reliability of a forecasting model for inflation depends on observable indicators of the monetary policy regime, which is the goal of our empirical application. This question is not readily answered within the structured framework of Pesaran, Pettenuzzo and Timmermann (2006) and Koop and Potter (2007), which assumes that parameter

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

changes are driven by a latent variable that defines different regimes for all the parameters in the model.[2]

Finally, we show that our framework embeds tests for forecast unbiasedness and forecast rationality, as those analysed, among others, by Elliott, Komunjer and Timmermann (2005). Unlike Elliott, Komunjer and Timmermann (2005), however, our tests take into account parameter estimation uncertainty, which, if neglected, can lead to significant size distortions (see also West and McCracken, 1998).

2.5. *Assumptions*

We make the following assumptions:

A1. $\{W_t\}$ is mixing with α of size $-r/(r-2)$, $r > 2$.

A2. (a) $L_t(\beta)$ is measurable and twice continuously differentiable with respect to β; (b) Under H_0 in (3) below, in a neighbourhood N of β^*, there exists a constant $D < \infty$ such that for all t, $\sup_{\beta \in N} \left| \left| \partial^2 L_t(\beta)/\partial\beta\partial\beta' \right| \right| < m_t$, for a measurable m_t such that $E(m_t) < D$.

A3. Under H_0, $\sup_{t \geq m} ||\widehat{\beta}_t - \beta^* - B_t^* H_t^*|| \rightarrow^{a.s.} 0$, where $\widehat{\beta}_t$ is $k \times 1$, B_t^* is a (non-stochastic) $k \times q$ matrix of rank k, such that $\sup_{t \geq 1} |B_t^*| < \infty$; $H_t^* = m^{-1} \sum_{s=1}^{m} h_s(\beta^*)$ (fixed scheme), $H_t^* = m^{-1} \sum_{s=t-m+1}^{t} h_s(\beta^*)$ (rolling scheme), $H_t^* = t^{-1} \sum_{s=1}^{t} h_s(\beta^*)$ (recursive scheme) for a $q \times 1$ orthogonality condition $h_s(\beta^*)$ such that $E(h_s(\beta^*)) = 0$.

A4. $\sup_{t \geq 1} E||[L_t(\beta^*), \partial L_t(\beta^*)/\partial\beta, h_t'(\beta^*)]'||^{2r} < \infty$, where $\partial L_t(\beta^*)/\partial\beta$ is $1 \times k$.

A5. $E(\partial L_t(\beta^*)/\partial\beta)$ is finite and constant for all t.

A6. $var\left(T^{-1/2} \sum_{t=1}^{T} L_t(\beta^*)\right) > 0$ for all T sufficiently large.

A7. $m, n \rightarrow \infty$, $\frac{n}{m} \rightarrow \pi$, $0 \leq \pi < \infty$.

Comments: 1. Assumption A1 restricts the memory in the data (ruling out, *e.g.* unit root processes) but allows the data to be heterogeneous, for example permitting the marginal distribution of the regressors to change over time. This is a more general assumption than the assumption of stationarity made in the majority of the structural break testing literature.

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

3. Assumption A3 is related to Assumption A2 of West (1996), permitting a number of estimation procedures for the model's parameters, including OLS, (quasi-) maximum likelihood and GMM. For example, for OLS estimation of the parameters in the linear model $Y_s = X_s'\beta^* + \varepsilon_s$, $s = 1, \ldots, t$, we have $B_t^* = \left(E\left(t^{-1} \sum_{s=1}^{t} X_s X_s'\right)\right)^{-1}$ and $h_s(\beta^*) = X_s \varepsilon_s$. For maximum likelihood estimation, B_t^* is the expectation of the inverse of the Hessian evaluated at β^* and H_t^* is the score. The assumption also states that under the null hypothesis of no forecast breakdown the pseudo-true values of the parameters are constant (note that we do not assume correct specification of the model under the null hypothesis).

4. Assumption A5 restricts the heterogeneity of the means of the loss derivatives, and is trivially satisfied when the loss used for estimation is the same as the loss used for evaluation,

2. Similar to the forecast breakdown test, for the case of predicting future forecast breakdowns the two approaches capture breaks of different magnitudes, and thus it may happen that the Pesaran, Pettenuzzo and Timmermann (2006) or Koop and Potter (2007) procedure predicts a break in the model's parameters that does not necessarily imply a future forecast breakdown.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

in which case $E\,(\partial L_t(\beta^*)/\partial\beta) = 0$ for all t. The assumption ensures that estimation uncertainty is asymptotically irrelevant, which leads to a simple expression for the asymptotic variance estimator for the forecast breakdown test. Proposition 5 in the Appendix shows how the forecast breakdown test is modified when one relaxes this assumption.

5. Assumption A7 shows that our asymptotic theory assumes that the in- and out-of-sample sizes go to infinity at the same rate, or that the in-sample size grows faster than the out-of-sample size. This assumption ensures that the test statistic has an asymptotically normal distribution for all forecasting schemes. This assumption can in principle be relaxed to let n grow to infinity faster than m, but there are complications that arise in the case of a rolling scheme. We discuss this in greater detail in Section 4.

2.6. *Asymptotic variance estimators*

This section shows how to construct a valid asymptotic variance estimator for the forecast breakdown test statistic (4) and provides the asymptotic justification for the forecast breakdown test. We consider two estimators: a general estimator that allows the losses to be heterogeneous (Theorem 1) and an estimator that is easier to compute, imposing the additional assumption that the losses are covariance stationary (Corollary 1).

The following algorithm shows the steps involved in constructing the general asymptotic variance estimator. The basic intuition is to note that the average surprise loss (2) is a weighted average of in- and out-of-sample losses, with weights depending on m, n and on the forecasting scheme. When estimation uncertainty is asymptotically irrelevant, $\hat{\sigma}^2_{m,n}$ is simply a (rescaled) HAC estimator of the variance of this weighted average. As we show in Proposition 5 in the Appendix, when estimation uncertainty matters, $\hat{\sigma}^2_{m,n}$ contains additional terms that depend on the estimator used.

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

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

and the corresponding vector $\widetilde{L}$ of demeaned losses, where $\widetilde{L}_t \equiv L_t - T^{-1}\sum_{j=1}^{T} L_j$,[3] *(2) a $1 \times T$ vector of weights, depending on the forecasting scheme, with element w_t^L, $t = 1, \ldots, T$:*

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

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

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

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

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

$$a_{m,j} = \frac{1}{m+j}+\frac{1}{m+j+1}+\cdots+\frac{1}{T-\tau};$$

The general HAC asymptotic variance estimator is

$$\hat{\sigma}^2_{m,n} \equiv (T/n)\left(T^{-1}\sum_{t=1}^{T}(w_t^L\widetilde{L}_t)^2 + 2T^{-1}\sum_{j=1}^{p_T}v_{T,j}\sum_{t=j}^{T}w_t^L\widetilde{L}_t w_{t-j}^L\widetilde{L}_{t-j}\right), \tag{5}$$

with weights $v_{T,j}$ and bandwidth p_T appropriately chosen (as in Andrews (1991) or Newey and West (1987)).

Theorem 1 (Asymptotic justification of the forecast breakdown test) *Given Assumptions A1–A7, under H_0 in (3), $t_{m,n,\tau} \xrightarrow{d} N(0,1)$, where $t_{m,n,\tau}$ is defined in (4) and $\hat{\sigma}^2_{m,n}$ in (5).*[4]

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

Corollary 1 (Variance estimator under covariance-stationarity) *Given Assumptions A1–A7, further assume that $\Gamma_j \equiv cov\left(L_t(\beta^*), L_{t-j}(\beta^*)\right)$ depends on j but not on t under H_0.*[5] *Then,*

$$\hat{\sigma}^2_{m,n} = \lambda S_n^{LL}, \tag{6}$$

4. A Matlab code computing $\hat{\sigma}_{m,n}$ can be downloaded from http:\\www.econ.ucla.edu\giacomin or http:\\www.econ.duke.edu\~brossi.

5. In the case of quadratic loss, this rules out time-variation in the unconditional fourth moments of the forecast errors.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

where

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

(7)

and

$$S_n^{LL}=n^{-1}\sum_{t=m}^{T-\tau}\widetilde{L}_{t+\tau}^2+2n^{-1}\sum_{j=1}^{p_n}v_{n,j}\sum_{t=m+j}^{T-\tau}\widetilde{L}_{t+\tau}\widetilde{L}_{t+\tau-j},$$

with $\widetilde{L}_{t+\tau}\equiv L_{t+\tau}(\widehat{\beta}_t)-n^{-1}\sum_{j=m}^{T-\tau}L_{j+\tau}(\widehat{\beta}_j)$ *and* $v_{n,j}$, p_n *appropriately chosen (Andrews (1991) or Newey and West (1987)).*

As we discussed in Section 2.2, if $L(e)=e$, with e the forecast error, the forecast breakdown test becomes a forecast unbiasedness test. In this case, Corollary 1 gives the correct variance estimator for the forecast unbiasedness test and shows that, for a recursive scheme, the estimator does not necessitate an adjustment and is simply an HAC estimator of the variance of the average out-of-sample forecast error. For the fixed and rolling schemes, instead, the estimator must be adjusted.[6]

3. CAUSES OF FORECAST BREAKDOWNS

To gain some insight into the causes of forecast breakdowns, we analyse the expectation of the numerator of the forecast breakdown test statistic (4).[7] For simplicity, in this section we assume that parameters are estimated by maximum likelihood and let $\mathcal{L}(\cdot)$ indicate the loss used for estimation. We further define β_t^* as $E\left(\partial\mathcal{L}_t\left(\beta_t^*\right)/\partial\beta\right)=0$, $t=1,2,\ldots,T$, and let $\overline{\Sigma}_j$ denote the relevant sample average depending on the forecasting scheme: $\overline{\Sigma}_j=t^{-1}\sum_{j=1}^{t}$ for the recursive scheme, $\overline{\Sigma}_j=m^{-1}\sum_{j=t-m+1}^{t}$ for the rolling scheme, and $m^{-1}\sum_{j=1}^{m}$ for the fixed scheme. Also, let $\overline{\widehat{\beta}}_t$, $\overline{\beta}_j^*$ and $\overline{\beta}_{t+\tau}^*$ denote intermediate points between $\left(\widehat{\beta}_t,\beta_t^*\right)$, $\left(\beta_t^*,\beta_j^*\right)$ and $\left(\beta_t^*,\beta_{t+\tau}^*\right)$, respectively. The following proposition decomposes the expectation of the numerator of our test statistic into various components, grouped into the three categories of parameter instabilities, other instabilities and estimation uncertainty.

Proposition 1 (Causes of forecast breakdowns)

$$E\left(n^{-1/2}\sum_{t=m}^{T-\tau}SL_{t+\tau}(\widehat{\beta}_t)\right)$$
$$=\underbrace{E\left(n^{-1/2}\sum_{t=m}^{T-\tau}\left(L_{t+\tau}\left(\beta_{t+\tau}^*\right)-\overline{\sum}_j L_j\left(\beta_j^*\right)\right)\right)}_{\text{“other instabilities”}}$$

6. It is easy to verify that, for the forecast unbiasedness test, our estimator coincides with the estimator proposed by McCracken (2000) for the various forecasting schemes.

7. We implicitly make the assumption that such expectation exists.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

$$
\begin{aligned}
&+\underbrace{n^{-1/2}\sum_{t=m}^{T-\tau} E\left(\frac{\partial L_{t+\tau}\left(\beta_{t+\tau}^{*}\right)}{\partial\beta}\right)\left(\beta_t^{*}-\beta_{t+\tau}^{*}\right)}_{\text{“parameter instabilities } I\text{”}} \\
&\underbrace{-n^{-1/2}\sum_{t=m}^{T-\tau}\overline{\sum}_j E\left(\frac{\partial L_j\left(\beta_j^{*}\right)}{\partial\beta}\right)\left(\beta_t^{*}-\beta_j^{*}\right)}_{\text{“parameter instabilities } I\text{”}} \\
&\underbrace{+0{\cdot}5n^{-1/2}\sum_{t=m}^{T-\tau}\left[\left(\beta_t^{*}-\beta_{t+\tau}^{*}\right)' E\left(\frac{\partial^2 L_{t+\tau}\left(\overline{\beta_{t+\tau}^{*}}\right)}{\partial\beta\partial\beta'}\right)\left(\beta_t^{*}-\beta_{t+\tau}^{*}\right)\right.}_{\text{“parameter instabilities } II\text{”}} \\
&\underbrace{\left.-\overline{\sum}_j\left(\beta_t^{*}-\beta_j^{*}\right)' E\left(\frac{\partial^2 L_j\left(\overline{\beta_j^{*}}\right)}{\partial\beta\partial\beta'}\right)\left(\beta_t^{*}-\beta_j^{*}\right)\right]}_{\text{“parameter instabilities } II\text{”}} \\
&\underbrace{+n^{-1/2}\sum_{t=m}^{T-\tau} E\left[\left(\frac{\partial L_{t+\tau}\left(\beta_t^{*}\right)}{\partial\beta}\right)\left(\widehat{\beta}_t-\beta_t^{*}\right)\right]}_{\text{“estimation uncertainty } I\text{”}} \\
&\underbrace{+n^{-1/2}\sum_{t=m}^{T-\tau} E\left\{\left[\left(\widehat{\beta}_t-\beta_t^{*}\right)'\frac{\partial^2\overline{\mathcal{L}}_t(\overline{\widehat{\beta}}_t)}{\partial\beta\partial\beta'}-\frac{\partial\overline{L}_t\left(\beta_t^{*}\right)}{\partial\beta}+\frac{\partial\overline{\mathcal{L}}_t\left(\beta_t^{*}\right)}{\partial\beta}\right]\left(\widehat{\beta}_t-\beta_t^{*}\right)\right\}}_{\text{“estimation uncertainty } II\text{”}} \\
&\underbrace{+0{\cdot}5n^{-1/2}\sum_{t=m}^{T-\tau} E\left[\left(\widehat{\beta}_t-\beta_t^{*}\right)'\left(\frac{\partial^2 L_{t+\tau}(\overline{\widehat{\beta}}_t)}{\partial\beta\partial\beta'}-\frac{\partial^2\overline{L}_t(\overline{\widehat{\beta}}_t)}{\partial\beta\partial\beta'}\right)\left(\widehat{\beta}_t-\beta_t^{*}\right)\right]}_{\text{“estimation uncertainty } III\text{”}}.
\end{aligned}
\tag{8}
$$

The component “other instabilities” captures any changes in the data-generating process – beyond parameter instabilities – that result in a non-constant expected loss. The “parameter instabilities I” component captures instabilities of the type $\beta_t^{*}-\beta^{*}=O_p\left(n^{1/2}\right)$ (which are the same instabilities considered by the structural break testing literature), whereas the “parameter instabilities II” component captures instabilities of the type $\beta_t^{*}-\beta^{*}=O_p\left(n^{1/4}\right)$. Note that, when the loss functions used for estimation and for evaluation are equal, the component “parameter instabilities I” disappears owing to $E\left(\partial L_{t+\tau}\left(\beta_{t+\tau}^{*}\right)/\partial\beta\right)=0$, implying that forecast breakdowns are in this case caused by parameter instabilities of greater magnitude than those considered by the structural break testing literature. We formally show this result in the next proposition, which compares the forecast breakdown test and Elliott and Muller's (2006) test in the simple situation in which the only source of a forecast breakdown is a break in the model's parameters. Besides showing that Elliott and Muller's (2006) test detects smaller parameter breaks than those causing a forecast breakdown, Proposition 2 illustrates the lack of robustness of Elliott and Muller's (2006) test to breaks in the marginal distribution of the regressors, whereas the forecast breakdown test is proved to be robust.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

Proposition 2 (Comparison with Elliott and Muller's (2006) test) *Suppose* $y = \Xi b + \varepsilon$, *where* $\Xi = diag\left(\left[X_0, \ldots, X_{T-1}\right]\right)$, X_t *scalar and* $\varepsilon_t \sim i.i.d.(0,1)$, *independent of* X_t. *Let* $X = \left[X_0, \ldots, X_{T-1}\right]'$, $y = [y_1, \ldots, y_T]'$, $\varepsilon = [\varepsilon_1, \ldots, \varepsilon_T]'$, $M = I_T - X\left(X'X\right)^{-1}X'$, I_T *the* $(T \times T)$ *identity matrix,* e_T *a* $(T \times 1)$ *vector of ones and* 0_T *a* $(T \times 1)$ *vector of zeros. Consider the scenarios:*

(a) (break in parameters, constant mean of regressors) $\mathrm{X}_t = \widetilde{\mathrm{X}}_t \sim i.i.d.\left(0, \sigma_X^2\right)$ *and the t-th element of b is* $\beta + T^{-\alpha}\Delta\beta \cdot 1\,(t \geq m)$.

(b) (break in mean of regressors, constant parameters) $\mathrm{X}_t = \widetilde{\mathrm{X}}_t + \Delta\mu^X \cdot 1\,(t \geq m)$ *and* $b = \beta \mathrm{e}_T$.

The effect of (a) and (b) on Elliott and Muller's (2006) test and on the forecast breakdown test is:

1. *Elliott and Muller (2006). The test builds on* $\xi_T \equiv T^{-1/2}\left[e'_{[sT]}, 0'_{[(1-s)T]}\right]\left[I_T \otimes V_X^{-1}\right]\Xi' M y$, *where* V_X *is a scaling factor and* $\xi_T \Rightarrow V_X^{-1}\sigma_X\left[B(s) - sB(1)\right] \equiv \xi$ *when* $\Delta\beta = \Delta\mu^X = 0$, *with* $B(.)$ *a standard Brownian Motion. In the scenarios above: (a) If* $\alpha = 1/2$, $\xi_T \Rightarrow \xi + V_X^{-1}\sigma_X^2\left[\int_0^s g_\beta(k)\,dk - s\int_0^1 g_\beta(k)\,dk\right]$ *where* $g_\beta(s) = \Delta\beta \cdot 1\left(s \geq (1+\pi)^{-1}\right)$, $t = [sT]$; *(b)* $\xi_T \Rightarrow V_X^{-1}\int_0^s\left[\sigma_X + g_X(r)\right]dB(r) - V_X^{-1}\left(\int_0^s\left[\sigma_X^2 + g_X^2(r)\right]dr\right)\left(\int_0^1\left[\sigma_X^2 + g_X^2(r)\right]dr\right)^{-1}\left(\int_0^1\left[\sigma_X + g_X(r)\right]dB(r)\right)$, *where* $g_X(s) = \Delta\mu^X \cdot 1\left(s \geq (1+\pi)^{-1}\right)$, $t = [sT]$. *The limiting distribution only equals* ξ *if* $\Delta\mu^X = 0$.
2. *Forecast breakdown test. For a quadratic loss, in Proposition 1 we have:*

   (a) breaks in parameters affect the "parameter instability II" component if $\alpha \leq 1/4$.

   (b) breaks in the mean of the regressors only affect the component "parameter estimation uncertainty II", which is asymptotically negligible.

The remaining components in the decomposition in Proposition 1 are owing to estimation uncertainty, and thus do not affect the asymptotic distribution of the forecast breakdown test statistic. It is nonetheless worthwhile to examine their effect in finite samples. First of all, note that, when the estimation and evaluation losses are equal, the "estimation uncertainty II" component is a quadratic form, and is thus always positive. Intuitively, this is because in this case the average in-sample loss computed at the parameter estimates is minimized by construction, and is thus smaller than the expected out-of-sample loss in finite samples. We therefore interpret this component as a measure of "overfitting".

The following proposition illustrates the decomposition in Proposition 1 when there are both breaks in parameters and breaks in the variance of the errors, for the special case of a linear regression model, a fixed forecasting scheme and a quadratic loss. It also shows that the presence of ARCH does not cause a forecast breakdown.

Proposition 3 (Special case: linear model with ARCH and quadratic loss) *Let* $L(e) = \mathcal{L}(e) = e^2$ *and consider a fixed forecasting scheme, and a model* $\mathrm{Y}_t = \mathrm{X}'_{t-1}\beta_t + \varepsilon_t$, *where:* $\varepsilon_t = \sigma_t u_t$; *the* $(k \times 1)$ *vector* X_{t-1} *is i.i.d. with* $E\left(\mathrm{X}_t\mathrm{X}'_t\right) \equiv J$; $\beta_t = \beta + n^{-1/4}\Delta\beta \cdot 1\,(t \geq m)$; $\sigma_t^2 = \sigma^2 + n^{-1/2}\Delta\sigma^2 \cdot 1\,(t \geq m) + \alpha\varepsilon_{t-1}^2$ *($\Delta\sigma^2$ can be negative) and* u_t *is i.i.d.(0,1). This specification allows for ARCH and two types of structural breaks: a break in the conditional mean parameters at time m (from* β *to* $\beta + \Delta\beta$*), and a break in the unconditional variance of*

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

the errors at time m (from $\sigma^2/(1-\alpha)$ *to* $\left(\sigma^2+\Delta\sigma^2\right)/(1-\alpha)$*). We have:*

$$E\left(n^{1/2}\,\overline{SL}_{m,n}\right)=\underbrace{\frac{\Delta\sigma^2}{1-\alpha}}_{\text{“other instabilities”}}+\underbrace{\frac{1}{2}\Delta\beta' J\Delta\beta}_{\text{“parameter instabilities } II\text{”}}+\underbrace{2\frac{n^{1/2}}{m}\frac{\sigma^2}{1-\alpha}k}_{\text{“overfitting”}}. \tag{9}$$

Comments: 1. From (9), we see that a forecast breakdown can be caused by a “small” positive break in the variance of the disturbances and/or a “large” break (positive or negative) in the conditional mean parameters.

2. Expression (9) implies that the breaks in parameters and variance of the errors could have opposite effects on the forecast performance, and thus not necessarily cause a forecast breakdown (*e.g.* if $\Delta\sigma^2 \leq -0{\cdot}5\Delta\beta' J\Delta\beta$). In other words, our test directly captures the bias-variance tradeoff that exists between breaks in the model's parameters (which result in biased forecasts) and breaks in the variance of the errors (which, if negative, lower the forecast error variance). An indirect approach that jointly tests for breaks in parameters and variance may instead detect both breaks and thus incorrectly conclude that the forecast performance of the model necessarily deteriorates.

3. Under Assumption A7, the overfitting component is present only in finite samples and is proportional to the number of parameters, the variance of the disturbances and the factor $n^{1/2}/m$. We discuss the effects of overfitting on the properties of the forecast breakdown test in greater detail in the next section, which proposes an overfitting-corrected version of the test.

4. AN OVERFITTING-CORRECTED FORECAST BREAKDOWN TEST

We propose a simple correction to the forecast breakdown test statistic (4) that eliminates the systematic difference between in- and out-of-sample losses that is present in finite samples when a quadratic loss is used for both estimation and evaluation. As we show in the Monte Carlo simulations in Section 7.1 below, this will substantially improve the finite sample behaviour of the forecast breakdown test.

The overfitting-corrected test consists of subtracting from the numerator of our test statistic an estimate of the “estimation uncertainty II” component in (8), interpreted as a measure of overfitting. Using similar reasonings to those in the proof of Proposition 3, we obtain an estimate of this component in the context of a linear model with covariance-stationary regressors, $\mathrm{Y}_t = \mathrm{X}'_{t-\tau}\beta+\varepsilon_t$. The test statistic is modified as:

$$t^c_{m,n,\tau}=\left(n^{1/2}\,\overline{SL}_{m,n}-c\right)/\hat{\sigma}_{m,n}; \tag{10}$$
$$c=2\gamma\ tr\left(\frac{X'X}{T}\cdot\widehat{V}^{\beta}_T\right),$$

where: $\gamma=n^{1/2}/m$ for the fixed and rolling schemes and $\gamma=n^{-1/2}\ln(1+n/m)$ for the recursive scheme; $X\equiv[X'_1,\ldots,X'_{T-\tau}]$; $\widehat{V}^{\beta}_T$ is a consistent estimator of the asymptotic variance of the full-sample parameter estimate, $\widehat{V}^{\beta}_T=\widehat{var}(T^{1/2}\ \widehat{\beta}_T)$; $\hat{\sigma}_{m,n}$ is as in Theorem 1 or Corollary 1.

The γ component in (10) allows us to discuss the conditions under which overfitting is asymptotically irrelevant in the various schemes. For all schemes, Assumption A7 ensures that $\gamma\to 0$ as n and m grow, but, as noted by a referee, this assumption is in principle stronger than necessary. In an additional appendix, available upon request, we prove that the results of Theorem 1 are still valid for the recursive and fixed schemes when $\pi=\infty$ regardless of the

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

rates at which n and m grow. For the rolling scheme, however, the overfitting component may become asymptotically relevant, depending on the rate at which n goes to infinity relative to m. Intuitively, this is owing to the fact that the test statistic divides the overfitting component by $\sigma_{m,n}$, whose asymptotic behaviour is dictated by λ in (7). One can easily show that for the fixed and recursive schemes the overfitting component always goes to zero when divided by $\sigma_{m,n}$ whereas for the rolling scheme it only goes to zero when n grows slower than $m^{3/2}$. When n grows faster than $m^{3/2}$, the overfitting component becomes asymptotically relevant, and the test statistic diverges. These considerations have implications for the finite sample properties of our test. In particular, we expect the forecast breakdown test to display size distortions for the rolling scheme when n is much larger than m, and the overfitting correction to eliminate such distortions. This will be confirmed by the results in Section 7.1 below.

It is finally worth noting that, under the assumption of conditional homoskedasticity, $\widehat{V}_T^{\beta} = \sigma^2(T^{-1}X'X)^{-1}$ and the overfitting correction simply becomes

$$c = 2\gamma k \ var(\varepsilon_t). \tag{11}$$

Direct calculations show that in this case $t^c_{m,n,\tau}$ may be equivalently obtained by redefining the surprise losses as the difference between the out-of-sample loss and the average in-sample loss penalized using Akaike's information criterion (AIC).[8]

5. PREDICTING FUTURE FORECAST BREAKDOWNS

The forecast breakdown test detects whether a forecast method broke down in the past. A question that may be of further interest to forecasters is whether the forecast method will break down in the future. This is of course related to finding past breakdowns: if the surprise losses had positive mean in the past, one could plausibly expect them to continue being positive in the near future. However, it is possible that one could find additional information that predicts whether there will be a forecast breakdown. For example, the surprise losses may be persistent (in the case of a quadratic loss, for example, the presence of GARCH in the data will induce serial correlation in the surprise losses) or they may be correlated with indicators of the state of the economy.

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

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

$$SL_{t+\tau}(\widehat{\beta}_t) = Z_t'\delta + \varepsilon_{t+\tau} \tag{12}$$

over the out-of-sample period $t = m, \ldots, T - \tau$, where the regression always includes a constant.

In order to verify whether δ is significant in (12), a Wald test can be performed by considering the test statistic $W_{m,n,\tau} = n\widehat{\delta}_n'\hat{\Omega}^{-1}_{m,n}\widehat{\delta}_n$, with $\hat{\Omega}_{m,n}$ given in Proposition 4 below and rejecting H_0 whenever $W_{m,n,\tau} > \chi^2_{r,1-\alpha}$, where $\chi^2_{r,1-\alpha}$ is the $(1-\alpha)$-th quantile of a χ^2_r distribution. Proposition 4 below provides the asymptotic justification for the test.

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

To analyse the behaviour of the surprise losses over time, one may further consider the plot of the fitted values $\{Z_t'\widehat{\delta}_n\}_{t=m}^{T-\tau}$ from the regression (12) together with a one-sided $(1-\alpha)\%$ confidence interval: $\left(Z_t'\widehat{\delta}_n - z_\alpha\left(Z_t'\left(\hat{\Omega}_{m,n}/n\right)Z_t\right)^{1/2}, +\infty\right)$, where z_α is the $(1-\alpha)-th$ quantile of a standard normal distribution.

Proposition 4 (Asymptotic justification of the Wald test) *Let* $Z_t = \left[1, z_t'\right]'$ *and* $\widetilde{z}_t \equiv z_t - \overline{z}$, $\overline{z} \equiv n^{-1}\sum_{t=m}^{T-\tau} z_t$. *Under Assumptions A1–A7, further suppose that, under* H_0*:*

B1. $\{z_t\}_{t=m}^{T-\tau}$ *and* $\{L_t(\beta^*)\}_{t=1}^{T}$ *are fourth order stationary;*
B2. $\overline{z} \underset{p}{\rightarrow} E(z_t)$;
B3. $S_{\widetilde{z}\widetilde{z}} \equiv n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{z}_t' \underset{p}{\rightarrow} \Sigma_{\widetilde{z}\widetilde{z}} \equiv E\left[\widetilde{z}_t\widetilde{z}_t'\right]$ *non-singular;*
B4. For some $d > 1$, $sup_{t\geq 1}E\left|\left|z_t', L_t(\beta^*)\right|\right|^{4d} < \infty$.[9]

Let

$$\widehat{\Omega}_{m,n} = \begin{pmatrix} 1 & -\overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & S_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\sigma}_{m,n}^2 & \Lambda S_{\widetilde{L},\widetilde{z}\widetilde{L}} \\ \Lambda S_{\widetilde{z}\widetilde{L},\widetilde{L}} & S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} \end{pmatrix} \begin{pmatrix} 1 & -\overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1} \\ 0 & S_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix}' \tag{13}$$

where $\widehat{\sigma}_{m,n}^2$ *is defined in Corollary 1,*

$$S_{\widetilde{z}\widetilde{z}} \equiv n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{z}_t' + n^{-1}\sum_{j=1}^{p_n} v_{n,j}\sum_{t=m+j}^{T-\tau}\left(\widetilde{z}_t\widetilde{z}_{t-j}' + \widetilde{z}_{t-j}\widetilde{z}_t'\right);$$

$$S_{\widetilde{z}\widetilde{L},\widetilde{L}}^{-1} \equiv n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_t^2 + n^{-1}\sum_{j=1}^{p_n} v_{n,j}\sum_{t=m+j}^{T-\tau}\left(\widetilde{z}_t\widetilde{L}_t\widetilde{L}_{t-j} + \widetilde{z}_{t-j}\widetilde{L}_{t-j}\widetilde{L}_t\right);$$

$$S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}}^{-1} \equiv n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{z}_t'\widetilde{L}_t^2 + n^{-1}\sum_{j=1}^{p_n} v_{n,j}\sum_{t=m+j}^{T-\tau}\left(\widetilde{z}_t\widetilde{L}_t\widetilde{L}_{t-j}\widetilde{z}_{t-j}' + \widetilde{z}_{t-j}\widetilde{L}_{t-j}\widetilde{L}_t\widetilde{z}_t'\right);$$

$\widetilde{L}_t$, p_n *and* $v_{n,j}$ *as in Algorithm 1;*

$\Lambda \equiv \left[\pi^{-1}\ln(1+\pi)\right]$ *(recursive scheme);*

$\Lambda \equiv 1 - \pi/2$ *(rolling scheme,* $n \leq m$*);*

$\Lambda \equiv (2\pi)^{-1}$ *(rolling scheme,* $n > m$*);*

$\Lambda \equiv 1$ *(fixed scheme).*

Then $W_{m,n,\tau} \xrightarrow{d} \chi_r^2$ *under* $H_0 : E\left(n^{-1}\sum_{t=m}^{T-\tau} Z_t \cdot SL_{t+\tau}(\beta^*)\right) = 0$.[10]

Corollary 2 (Asymptotic justification of the Wald test under conditional homoskedasticity) *Given Assumptions A1–A7, further suppose that, under* H_0,

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

10. The Matlab code to implement the Wald test under the assumptions of Proposition 4 is available at http://www.econ.duke.edu/~brossi/ or http://www.econ.ucla.edu/giacomin/

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

$E\left(\widetilde{L}_t(\beta^*)\,\widetilde{L}_{t-j}(\beta^*) \mid \{z_t\}_{t=m}^{T-\tau}\right) \equiv \gamma_j^{LL}$. *Then:*

$$\widehat{\Omega}_{m,n} = \begin{pmatrix} \sigma_{m,n}^2 + \overline{z}' S_{\widetilde{z}\widetilde{z}}^{-1} S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} S_{\widetilde{z}\widetilde{z}}^{-1} \overline{z} & -\overline{z}' S_{\widetilde{z}\widetilde{z}}^{-1} S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} S_{\widetilde{z}\widetilde{z}}^{-1} \\ -S_{\widetilde{z}\widetilde{z}}^{-1} S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} S_{\widetilde{z}\widetilde{z}}^{-1} \overline{z} & S_{\widetilde{z}\widetilde{z}}^{-1} S_{\widetilde{z}\widetilde{L},\widetilde{z}\widetilde{L}} S_{\widetilde{z}\widetilde{z}}^{-1} \end{pmatrix}. \tag{14}$$

Comments: 1. When a quadratic loss is used for estimation and evaluation, equation (12) can be interpreted as a forecast rationality regression, by letting $L(e) = e$ in (1), where e is the forecast error (as in this case $\overline{L}_t(\widehat{\beta}_t) = 0$). Proposition 4 thus provides the appropriate asymptotic variance estimator for the forecast rationality test, and shows that a correction is only required for the standard error of the intercept (which is the same correction that applies to the forecast unbiasedness test; see the comment after Corollary 1 of West and McCracken, 1998). This result could be easily generalized to the class of loss functions examined in Elliott, Komunjer and Timmermann (2005) by redefining L as $L(e) = [1(e < 0) - \alpha]\,|e|^{p-1}$, where the parameters α and p are defined in their equation (1), and provided the same loss function is used for estimation and evaluation. Our Proposition 4 thus gives an estimator of the asymptotic variance for a forecast rationality test that, unlike that proposed by Elliott, Komunjer and Timmermann (2005), takes into account the effects of parameter estimation uncertainty (this is the same as the estimator suggested by West and McCracken (1998)). Neglecting parameter estimation uncertainty can result in considerable size distortions, as our Monte Carlo simulation in Section 7.2 will show.

2. Note that Z_t having explanatory power in (12) does not necessarily imply a forecast breakdown. This has to do with the fact that equation (12) models the conditional expectation of the surprise losses, whereas a forecast breakdown occurs when the unconditional expectation of the surprise losses is different from zero. The hypothesis of a forecast breakdown can thus still be tested in (12) by a t-test on the intercept. The goal of further modelling the conditional mean of the surprise losses is to be able to forecast how much the future losses differ from their expectation, by relating systematic differences between in- and out-of-sample performances to economic variables. Note that, for a quadratic loss, the losses reflect forecast error variances and thus the idea of expressing surprise losses as functions of explanatory variables is reminiscent of an ARCH-type model where the variance dynamics depend on economic variables, as in Glosten, Jagannathan and Runkle (1993). Our approach, however, not only captures dynamics in scale parameters, but also in location parameters.

3. Our approach to predicting future forecast breakdowns may be further related to Pesaran, Pettenuzzo and Timmermann (2006) and Koop and Potter (2007), who model the breaks in location and scale parameters of a model as functions of a latent variable that defines different regimes. The main difference is that we model directly the difference between out-of- and in-sample performances, for a particular loss function, by relating it to observable explanatory variables and that we do not need to assume that the underlying forecasting model is correctly specified. For a quadratic loss, this may be thought of as trying to contemporaneously characterize breaks in scale and location parameters for the forecasting model, and expressing them as functions of observables, rather than of a latent variable.

6. IMPLICATIONS OF FORECAST BREAKDOWNS

A natural question that arises if a forecast breakdown is detected or predicted is whether the forecast model should be changed or not. In general, the answer to this question depends on the type of forecast (point, interval, density) and on the type of loss function (symmetric or asymmetric). For example, when the forecast is a point forecast and the loss function is

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

symmetric, finding a forecast breakdown does not necessarily imply that the model should be changed. The reason is that the forecast breakdown could be caused by instabilities – such as increases in the variance of the disturbances – that do not affect the optimal forecast (for a symmetric loss, the optimal point forecast does not depend on the variance, unlike for an asymmetric loss, as shown by Christoffersen and Diebold (1997)). As the forecast breakdown test cannot distinguish among the different types of instabilities, the finding of a forecast breakdown does not necessarily suggest changing the model. However, even though a change in the variance may not affect the optimal forecast, it will affect the prediction interval associated with the point forecast, increasing the likelihood of large forecast errors. For a decision maker committed to prevent such large forecast errors, therefore, this would be relevant. In conclusion, we can say that when the loss is asymmetric or when the forecaster is interested in accompanying the point forecast with some measure of its uncertainty, then the finding of a forecast breakdown indicates unreliability of the forecast, regardless of its cause.

7. MONTE CARLO EVIDENCE

We analyse the size and power properties of our forecast breakdown test in finite samples, and compare them to the properties of structural break tests (Elliott and Muller, 2006), henceforth EM. We further compare the size properties of commonly used forecast rationality tests and those of our corrected forecast rationality test (see comment after Corollary 2).

7.1. *Size properties of forecast breakdown tests*

We investigate the size of the forecast breakdown test, in particular with regard to its robustness to the presence of conditionally heteroskedastic disturbances and to the presence of instability in the marginal distribution of the regressors. We let the data-generating process (DGP) be:

$$\begin{aligned} \mathrm{Y}_t &= 2{\cdot}73 - 0{\cdot}44 X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \sigma_t u_t, \\ \sigma_t^2 &= 1 + \alpha \varepsilon_{t-1}^2, \;\; u_t \sim i.i.d.N(0,1), \end{aligned} \tag{15}$$

and consider two experiment designs. The first (MC1) has $\alpha = 0$ and i.i.d. regressors and errors: X_t, $u_t \sim i.i.d.N(0,1)$, independent of each other. The second (MC2), inspired by our empirical application to the Phillips curve model of U.S. inflation, lets X_t be monthly U.S. unemployment and lets $\alpha = 0{\cdot}5$.[11] The DGP specification and parameters are from Staiger, Stock and Watson (1997). We use an actual time series for unemployment in order to generate data that exhibit realistic and possibly heterogeneous behaviour. Throughout, we restrict attention to the one-step-ahead forecast horizon and use a quadratic loss for both estimation and evaluation.

For each pair of in- and out-of-sample sizes (m, n) and for each of 5000 Monte Carlo replications, we generate $T = m + n$ data as in (15). In MC2, we use the first T data in the unemployment time series, starting from 1948:1. We estimate the model $\mathrm{Y}_t = \beta_1 + \beta_2 \mathrm{X}_{t-1} + e_t$ by OLS using either a fixed, a rolling or a recursive forecasting scheme. We consider the forecast breakdown test for the three forecasting schemes, using either the general asymptotic variance estimator of Theorem 1 ($t_{m,n,\tau}$) or the estimator of Corollary 1 ($t^{\mathrm{stat}}_{m,n,\tau}$) (the truncation lags

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

TABLE 1

Size of FB test. Nominal size 0·05

		MC1					
		$t_{m,n,\tau}$			$t^{\text{stat}}_{m,n,\tau}$		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
50	50	0·113	0·144	0·097	0·064	0·096	0·058
50	100	0·152	0·297	0·121	0·077	0·244	0·071
50	150	0·168	0·492	0·128	0·080	0·440	0·075
100	50	0·072	0·071	0·065	0·049	0·052	0·047
100	100	0·096	0·109	0·081	0·057	0·075	0·055
100	150	0·101	0·143	0·086	0·060	0·117	0·059
150	50	0·044	0·046	0·040	0·036	0·038	0·035
150	100	0·064	0·072	0·058	0·046	0·052	0·043
150	150	0·069	0·087	0·065	0·047	0·066	0·046
		MC2					
		$t_{m,n,\tau}$			$t^{\text{stat}}_{m,n,\tau}$		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
50	50	0·122	0·159	0·111	0·047	0·093	0·051
50	100	0·017	0·276	0·143	0·047	0·191	0·054
50	150	0·197	0·386	0·143	0·040	0·248	0·051
100	50	0·062	0·075	0·062	0·035	0·044	0·036
100	100	0·092	0·100	0·085	0·035	0·051	0·035
100	150	0·116	0·133	0·102	0·036	0·087	0·041
150	50	0·042	0·046	0·041	0·033	0·035	0·032
150	100	0·062	0·071	0·060	0·030	0·035	0·029
150	150	0·076	0·085	0·069	0·029	0·043	0·031

Notes: The table reports rejection frequencies over 5000 Monte Carlo replications of the forecast breakdown test of Section 2.2, using either the asymptotic variance estimator of Algorithm 1 ($t_{m,n,\tau}$) or the estimator of Corollary 1 $\left(t^{\text{stat}}_{m,n,\tau}\right)$, both tests implemented with either a fixed, rolling or recursive scheme. The experiment designs MC1 and MC2 are described in Section 7.1 and m and n denote in-sample and out-of-sample sizes, respectively.

for the HAC estimators are $p_T = p_n = 0$ in MC1 and $p_T = p_n = [n^{1/3}]$ in MC2, where $[\cdot]$ indicates the integer value). Table 1 contains the rejection frequencies of our tests for various (m, n) pairs. Table 2 reports the rejection frequencies for the overfitting-corrected tests.

The forecast breakdown test has good size properties for large in- and out-of-sample sizes ($m, n \geq 100$). The $t^{\text{stat}}_{m,n,\tau}$ test is well sized, if conservative. Both tests (in particular $t_{m,n,\tau}$) tend to over-reject when the in-sample size is small ($m = 50$), especially for the rolling scheme, which may become quite unreliable when m is small. Also note that, for a given m, the size distortions for the rolling scheme become more severe as n grows. This reflects the fact that, when n grows faster than $m^{3/2}$, the test statistic for the rolling scheme diverges owing to the overfitting component becoming asymptotically non-negligible, as we discussed in Section 4. Table 2 also documents that this problem can be overcome by using our overfitting-corrected test, which has good finite sample properties for all forecasting schemes and in- and out-of-sample sizes. Comparing the results from MC1 and MC2, we see that the forecast breakdown test is robust to the presence of possibly heterogeneous regressors and of ARCH errors.

We further perform a small Monte Carlo experiment that illustrates the greater robustness of the forecast breakdown test relative to the EM test in the presence of breaks in the marginal

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

TABLE 2

Size of overfitting-corrected FB test. Nominal size 0·05

		MC1					
		$t^{c}_{m,n,\tau}$			$t^{stat,c}_{m,n,\tau}$		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
50	50	0·064	0·053	0·053	0·031	0·031	0·028
50	100	0·085	0·056	0·066	0·031	0·042	0·032
50	150	0·095	0·068	0·065	0·034	0·053	0·029
100	50	0·043	0·040	0·038	0·029	0·030	0·027
100	100	0·057	0·057	0·052	0·030	0·036	0·031
100	150	0·068	0·055	0·056	0·032	0·041	0·033
150	50	0·031	0·030	0·027	0·024	0·024	0·022
150	100	0·050	0·047	0·046	0·032	0·031	0·030
150	150	0·058	0·053	0·053	0·038	0·035	0·034
		MC2					
		$t^{c}_{m,n,\tau}$			$t^{stat,c}_{m,n,\tau}$		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
50	50	0·097	0·106	0·086	0·032	0·053	0·039
50	100	0·130	0·142	0·107	0·026	0·094	0·035
50	150	0·147	0·184	0·107	0·022	0·119	0·030
100	50	0·053	0·059	0·049	0·033	0·038	0·033
100	100	0·081	0·081	0·074	0·029	0·039	0·028
100	150	0·109	0·092	0·092	0·029	0·062	0·034
150	50	0·038	0·041	0·037	0·032	0·036	0·032
150	100	0·053	0·063	0·051	0·026	0·032	0·027
150	150	0·079	0·079	0·074	0·029	0·038	0·029

Notes: The table reports rejection frequencies over 5000 Monte Carlo replications of the overfitting-corrected forecast breakdown (FB) test of Section 4, using either the asymptotic variance estimator of Algorithm 1 ($t^{c}_{m,n,\tau}$) or the estimator of Corollary 1 $\left(t^{stat,c}_{m,n,\tau}\right)$, both tests implemented with either a fixed, rolling or recursive scheme. The experiment designs MC1 and MC2 are described in Section 7.1 and m and n denote in-sample and out-of-sample sizes, respectively.

distribution of the regressors. We consider the case in which the null hypothesis of no break in the conditional mean parameters is satisfied but there is a break of size $\Delta\beta$ in the mean of the regressor occurring at time τT:

$$\begin{aligned} \mathrm{Y}_t &= \mathrm{X}_{t-1} + \varepsilon_t, \\ \mathrm{X}_t &= \mathrm{Z}_t + \Delta\beta \cdot 1(t \geq \tau T),\ \ \mathrm{Z}_t, \varepsilon_t \sim \text{i.i.d. } N(0,1) \text{ and independent.} \end{aligned}$$

We let $T = 100$, $\Delta\beta$ vary between 1 and 16, and consider $\tau = 0{\cdot}75$ and $\tau = 0{\cdot}95$. We plot the empirical rejection frequencies of both our recursive forecast breakdown test (the overfitting-corrected estimator for $m = 50$ and using the variance estimator from Corollary 1) and the EM test over 5000 Monte Carlo replications in Figure 1, for a 10% nominal size.

The figure clearly shows that the EM test exhibits size distortions which depend not only on the magnitude of the break, but also on its location. The performance of the forecast breakdown test, instead, is affected neither by the presence of a break in the regressor, nor by its location.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

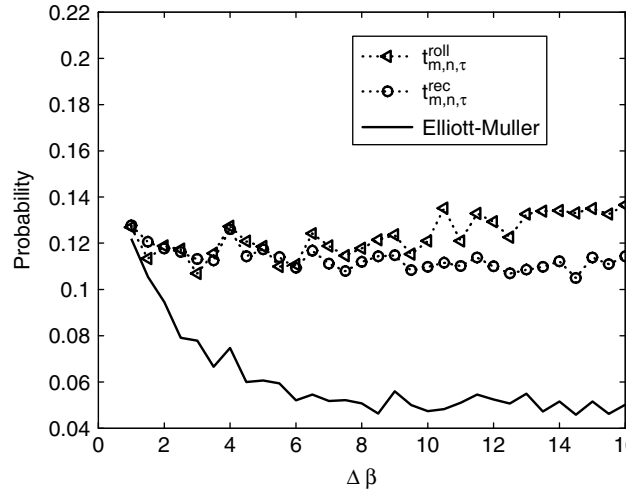

FIGURE 1(a)

Size of FB and EM tests. Break in regressor at time ·75T

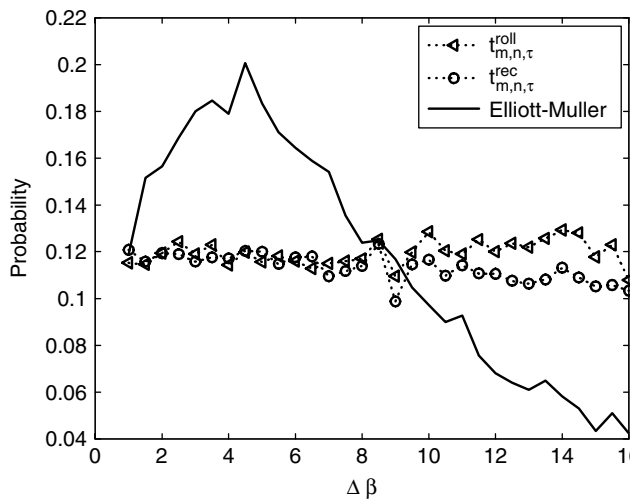

FIGURE 1(b)

Size of FB and EM tests. Break in regressor at time ·95T

7.2. *Size properties of forecast rationality tests*

Finally, we document size distortions of conventional forecast rationality tests and the good size properties of a test based on the variance estimator of Proposition 4 and on the correction for overfitting. The DGP is: $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$, where $\beta_0 = \beta_1 = 0$, $\varepsilon_t \sim i.i.d.N(0, 1)$, $X_t \sim i.i.d.N(0, 1)$ and forecasts are based on a model with a constant and X_t estimated using the various forecasting schemes. The forecast rationality test is performed by estimating the regression: $e_{t+1} = \delta_0 + \delta_1 Z_t + u_t$, where e_{t+1} is the estimated out-of-sample forecast error

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

TABLE 3

Size of forecast rationality tests. Nominal size ·05

		Unadjusted			Adjusted		
m	n	Fixed	Rol.	Rec.	Fixed	Rol.	Rec.
			t_{δ_0}				
50	50	0·172	0·021	0·052	0·054	0·053	0·052
50	100	0·266	0·002	0·050	0·056	0·053	0·050
50	150	0·321	0·000	0·048	0·052	0·058	0·048
100	50	0·111	0·044	0·056	0·056	0·052	0·055
100	100	0·172	0·018	0·053	0·053	0·051	0·053
100	150	0·215	0·004	0·050	0·051	0·048	0·050
150	50	0·101	0·053	0·059	0·061	0·059	0·061
150	100	0·136	0·037	0·054	0·055	0·055	0·054
150	150	0·177	0·016	0·050	0·052	0·047	0·049
			t_{δ_1}				
50	50	0·060	0·062	0·061	0·064	0·062	0·062
50	100	0·055	0·053	0·054	0·051	0·053	0·051
50	150	0·048	0·050	0·049	0·049	0·050	0·049
100	50	0·055	0·054	0·054	0·052	0·052	0·052
100	100	0·056	0·056	0·056	0·056	0·057	0·057
100	150	0·045	0·048	0·045	0·049	0·049	0·049
150	50	0·062	0·062	0·062	0·061	0·060	0·061
150	100	0·057	0·056	0·057	0·057	0·056	0·057
150	150	0·051	0·051	0·050	0·050	0·050	0·051
			Wald				
50	50	0·148	0·040	0·059	0·065	0·071	0·069
50	100	0·220	0·024	0·058	0·055	0·054	0·057
50	150	0·276	0·017	0·051	0·050	0·053	0·048
100	50	0·102	0·050	0·058	0·066	0·061	0·063
100	100	0·146	0·035	0·057	0·057	0·057	0·056
100	150	0·179	0·018	0·048	0·048	0·049	0·048
150	50	0·097	0·062	0·066	0·068	0·072	0·069
150	100	0·115	0·047	0·058	0·060	0·061	0·059
150	150	0·148	0·031	0·053	0·048	0·053	0·052

Notes: The table reports rejection frequencies over 5000 Monte Carlo replications of forecast rationality tests. We consider t-tests of significance of the intercept (t_{δ_0}) and the slope coefficient (t_{δ_1}), as well as a test of joint significance of both coefficients (*Wald*) in the forecast rationality regression (12). Forecast errors are obtained using either a fixed, rolling or recursive scheme and in each case the tests are implemented using either the usual OLS variance estimator ("unadjusted") or the asymptotic variance estimator of Corollary 2 ("adjusted"). The experiment design is described in Section 7.1 and m and n denote in-sample and out-of-sample sizes, respectively.

and $Z_t \sim i.i.d.N(0, 1)$ independent of X_t. Table 3 reports rejection frequencies of a forecast rationality test that uses conventional OLS standard errors (not taking estimation uncertainty into account, as in Elliott, Komunjer and Timmermann (2005)), labelled "unadjusted", and of the corresponding test using our variance estimator (14) with $L(e) = e$, labelled "adjusted". The nominal size is 5%. As the columns labelled "unadjusted" in Table 3 show, both a standard t-test on δ_0 (t_{δ_0}) and a Wald test on both δ_0 and δ_1 (W) have considerable size distortions except for the recursive scheme, whereas a t-test on δ_1 (t_{δ_1}) has no size distortions for any scheme. The columns labelled "adjusted" show instead that our variance estimator yields a test with correct size.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

7.3. *Power properties*

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

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

$$Y_t = \beta_A \cdot 1\,(t > T/2) + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N(0, 1). \tag{16}$$

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

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

$$Y_t = \varepsilon_t, \quad \varepsilon_t \sim i.i.d.N\left(0, \sigma_t^2\right) \tag{17}$$

where $\sigma_t^2 = 1 + \beta_A \cdot 1\,(t > T/2)$. We choose $(T, m) = (300, 150)$. In the recurring changes case, we let σ_t^2 switch between 1 and $(1 + \beta_A)$ every 50 periods, and let $(T, m) = (600, 50)$. In this case, we omit a comparison with Elliott and Muller's (2006) test because their test focuses on breaks in conditional mean parameters rather than variance.

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

$$\begin{aligned} Y_t &= \beta_A \cdot 1\,(t \le T/4) - 3\beta_A \cdot 1\,(T/4 < t \le T/2) + X_t \cdot 1\,(t > T/2) + \varepsilon_t, \\ X_t &= 0{\cdot}6X_{t-1} + \eta_t, \quad \varepsilon_t, \eta_t \sim i.i.d.N\,(0, 1) \text{ independent.} \end{aligned} \tag{18}$$

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

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

8. THE PHILLIPS CURVE AND INFLATION FORECAST BREAKDOWNS

The Phillips curve as a forecasting model of inflation has traditionally been a useful guide for monetary policy in the United States, and its forecasting ability is thus of practical relevance.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

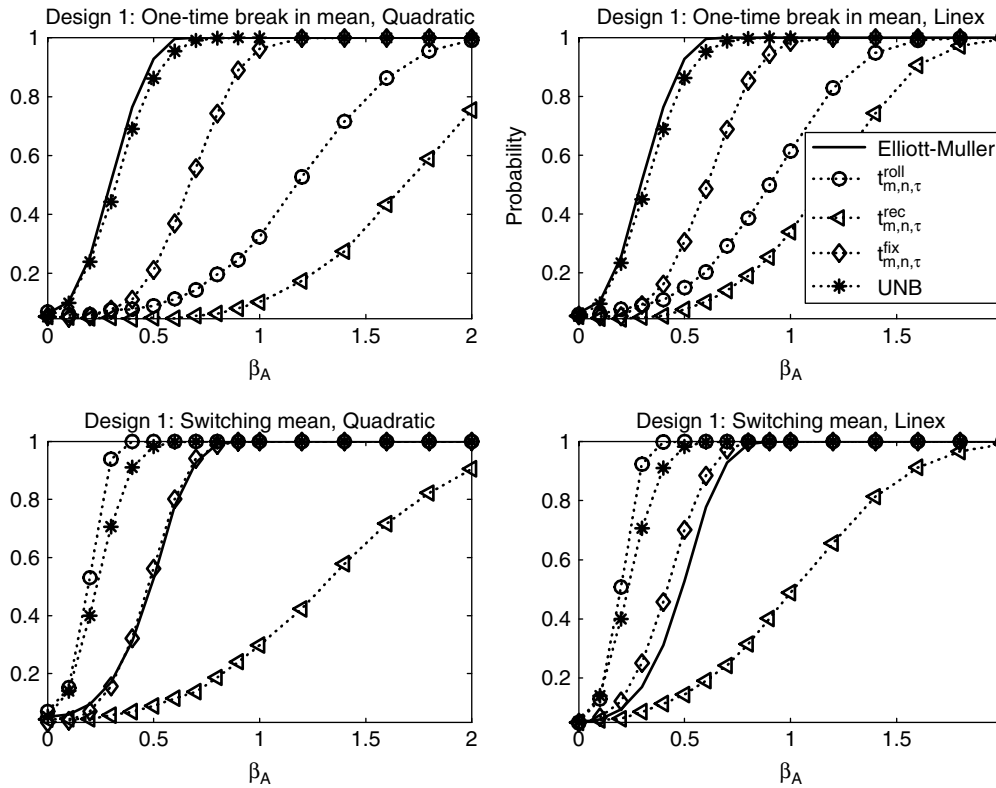

FIGURE 2(a)

Power functions

The model relates changes in inflation to past values of the unemployment gap (the difference between the unemployment rate and the NAIRU) and past values of inflation. The forecasting ability of the Phillips curve as well as its stability have been investigated in a number of works, including Staiger, Stock and Watson (1997), Stock and Watson (1999), and Fisher, Liu and Zhou (2002). The latter, in particular, conclude that the forecasting ability of the Phillips curve depends upon the period; the Phillips curve appears to forecast well 1 year ahead during the 1977–1984 period but not during the 1993–2000 period. Thus, as an empirical application of the methods proposed in this paper, we investigate the robustness of the Phillips curve to forecast breakdowns.

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

$$\pi_{t+\tau}^\tau - \pi_t = \theta_0 + \theta_1(L)\,u_t + \theta_2(L)(\pi_t - \pi_{t-1}) + \varepsilon_{t+\tau} \tag{19}$$

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

When analysing whether unemployment was a useful predictor for inflation, it is important to assess its predictive ability using data that were available to the policy-makers at that time. For example, Ghysels, Swanson and Callan (2002) analyse the performance of monetary

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

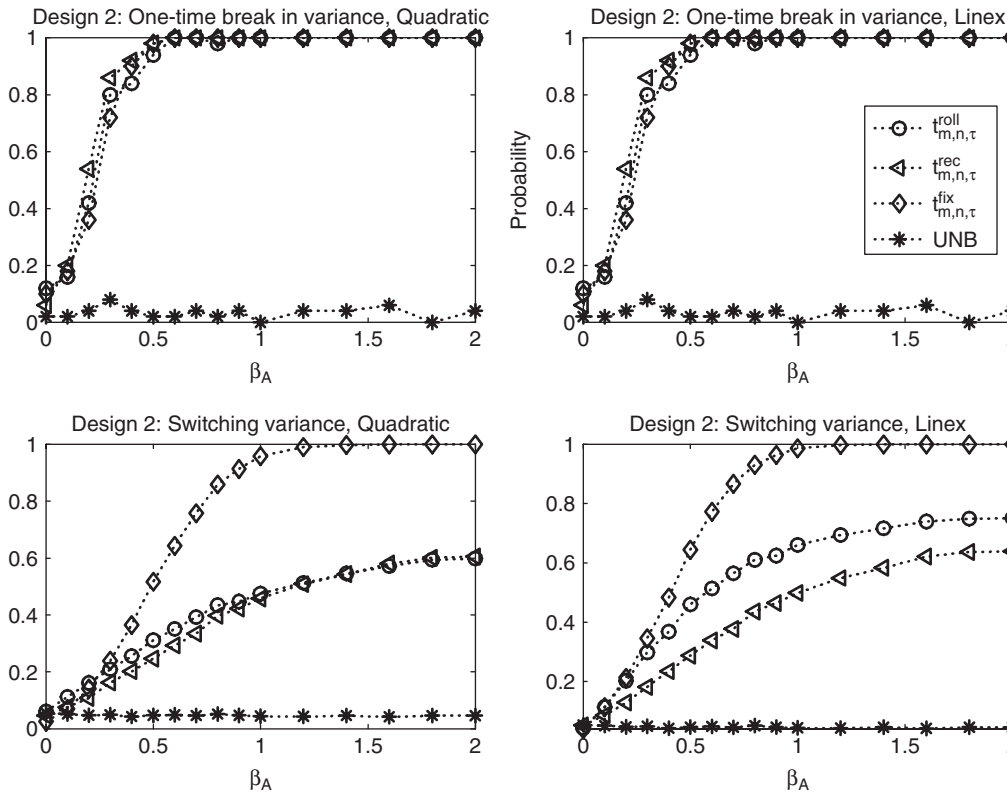

FIGURE 2(b)

Power functions

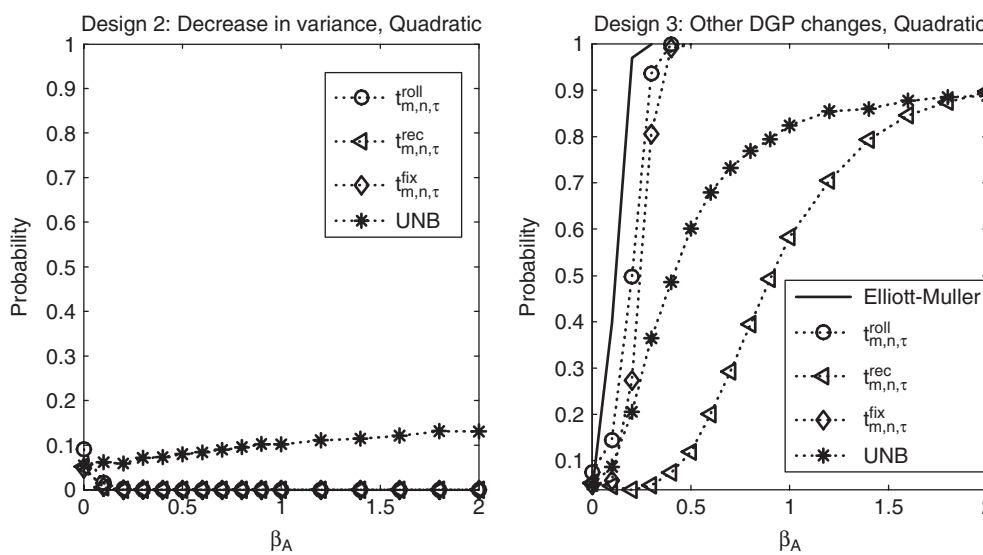

FIGURE 2(c)

Power functions

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

TABLE 4

p-values of forecast breakdown test

		Real-time data	Revised data
q_u	q_π	$t_{m,n,\tau}$	$t_{m,n,\tau}$
		$\tau = 1$	
1	1	–	0·004
1	3	–	0·021
3	1	–	0·009
3	3	–	0·039
	BIC	–	0·021
		$\tau = 3$	
1	1	0·000	0·256
1	3	0·562	0·326
3	1	0·450	0·434
3	3	0·572	0·524
	BIC	0·874	0·475
		$\tau = 12$	
1	1	0·001	0·111
1	3	0·000	0·312
3	1	0·002	0·756
3	3	0·001	0·948
	BIC	0·001	0·591

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

policy rules in the presence of real-time data, and note their relationship with changes in the Fed Chairmen. For this reason, we use real-time data from the Federal Reserve Bank of Philadelphia database. The data are discussed in Croushore and Stark (2001). As the real-time series of consumer prices from the same dataset is available only from the 1994 vintage, for this series we use the Swanson, van Dijk and Callan dataset (available at http://econweb.rutgers.edu/nswanson/realtime.htm). We focus on seasonally adjusted inflation, as in Stock and Watson (1999). The data are from 1961:1 (with a first vintage in 1978:2) until 2001:12. Owing to the data limitations, we restrict estimation from 1978:2 until 2001:12, using quarterly vintages.[12]

The first column of Table 4 reports the p-values of the forecast breakdown test of Section 2.2 for a quadratic loss and a rolling scheme with $m = 60$ (so that the one-step ahead

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

forecasts begin in 1993:1, corresponding to the change in monetary policy identified in Fisher, Liu and Zhou (2002)). We consider forecast horizons $\tau = 3$ and $\tau = 12$ months and several choices of q_u and q_π. The row labelled "BIC" reports results for the case in which the lag length is determined by the Bayesian Information Criterion (BIC) (assuming that all regressors have the same number of lags). The table shows strong evidence of a forecast breakdown at the 1-year horizon when using real-time data, whereas there is little evidence of forecast breakdowns at shorter horizons. Because of small sample concerns associated with real-time data, we repeat the above exercise using revised monthly data. We consider the most recent observations collected by the Philadelphia Fed (2004:8) for both seasonally unadjusted CPI and unemployment. The largest available sample for both variables is from 1948:1 until 2004:6. The second column in Table 4 shows that the forecast breakdown test finds some evidence of a forecast breakdown at the 1-month horizon, but not at longer horizons.

Given the evidence in favour of forecast breakdowns in the Phillips curve, we next investigate its possible economic causes. Fisher, Liu and Zhou (2002) argue that periods of low inflation volatility and periods after regime shifts in monetary policy appear to be associated with changes in the forecasting ability of the Phillips curve. Thus, we construct a forecasting model that relates the surprise losses to inflation volatility and to a measure of changes in the monetary policy behaviour of the Fed. We estimate inflation volatility ($\widehat{\sigma}^2_{\pi,t}$) as the sample variance of the change in the annual inflation over a rolling window of size 241.[13] To measure changes in the monetary policy behaviour of the Fed, we consider rolling two-step efficient GMM estimates (with two-stage least squares in the first step) of the coefficients of the Federal Fund Rate (FFR) reaction function to the output gap and to the deviation of inflation from its target proposed by Clarida, Gali and Gertler (2000), given by

$$E\left(r_t - (1-\rho)\left[rr^* - (\beta - 1)\pi^* + \beta\pi_{t,k} + \gamma x_{t,q}\right] + \rho(L)\, r_{t-1} | \Im_t\right) = 0, \tag{20}$$

with r_t the nominal FFR; $\pi_{t,k}$ the annualized percentage change in the price level between t and $t+k$; $x_{t,q}$ the average output gap between t and $t+q$, defined as minus the percentage deviation of actual unemployment from its target (a fitted quadratic function of time); and $\Im_t$ the information set at time t. As in Clarida, Gali and Gertler (2000), we let $\rho(L) \equiv \rho_1 + \rho_2 L$, rr^* be the average FFR over the estimation window, and we choose as instruments a constant and four lags of the following variables: inflation, output gap, FFR, commodity price inflation, M2 growth rate, spread between the long-term bond rate and the 3-month Treasury Bill rate.[14] k and q are set at 1 quarter. Our measures of changes in monetary policy behaviour are sequences of estimates of β, γ and $\rho \equiv \rho(1)$ in (20) over a rolling window of size 241. Even though our database is different from that of Clarida, Gali and Gertler (2000), our parameter estimates – which we do not report to conserve space – are similar.

We next investigate whether the estimates of the FFR reaction function coefficients and inflation volatility are useful predictors of inflation forecast breakdowns. Table 5 shows

13. That is we use lagged values of the sample variance of $\left(\pi^{\tau}_{t+\tau} - \pi_t\right)$ as a potential predictor.

14. Unlike in Clarida, Gali and Gertler (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 10-year monthly rate of interest on government securities provided by the Fed (we checked that in the overlapping portion with FYGL the data look similar). Similar problems lead us to choose the 3-month U.S. Treasury Bills quoted on the secondary market as a proxy for the 3-month Treasury Bill rate. Finally, for commodity prices we used n.s.a. CPI for all items all urban consumers (U.S. city average) and we collected data for M2 from the Federal Reserve Board database. The abuse of notation in denoting the degree of inflation aversion by β is to make our notation consistent with that of Clarida, Gali and Gertler (2000).

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

TABLE 5

Explaining forecast breakdowns by monetary policy changes and inflation variance

			δ_1				$W_{m,n,\tau}$
τ	q_u	q_π	$z_t = \widehat{\beta}_t$	$z_t = \widehat{\gamma}_t$	$z_t = \widehat{\rho}_t$	$z_t = \widehat{\sigma}^2_{\pi,t}$	$z_t = (\widehat{\beta}_t, \widehat{\gamma}_t, \widehat{\rho}_t)'$
1	1	1	−2·285	1·828	19·770	−1·019	9·533
			(0·156)	(0·018)	(0·795)	(0·024)	(0·023)
	1	3	−2·348	1·612	6·484	−0·892	7·386
			(0·159)	(0·037)	(0·933)	(0·051)	(0·061)
	3	1	−2·306	1·712	13·957	−0·980	8·397
			(0·148)	(0·028)	(0·856)	(0·031)	(0·039)
	3	3	−2·354	1·513	1·977	−0·866	6·623
			(0·153)	(0·050)	(0·980)	(0·059)	(0·085)
		BIC	−2·187	1·654	6·272	−0·855	7·286
			(0·185)	(0·046)	(0·938)	(0·071)	(0·063)
3	1	1	−1·806	−0·404	−114·2	−1·713	1·985
			(0·531)	(0·785)	(0·249)	(0·000)	(0·576)
	1	3	−1·837	−0·267	−122·4	−1·716	2·077
			(0·519)	(0·858)	(0·238)	(0·000)	(0·557)
	3	1	−1·651	−0·568	−128·8	−1·705	2·337
			(0·575)	(0·706)	(0·201)	(0·010)	(0·506)
	3	3	−1·657	−0·415	−136·1	−1·702	2·386
			(0·570)	(0·782)	(0·195)	(0·000)	(0·496)
		BIC	−1·608	−0·642	−141·4	−1·613	2·602
			(0·590)	(0·669)	(0·175)	(0·001)	(0·457)
12	1	1	−1·304	−0·105	−199·5	−1·876	6·268
			(0·578)	(0·942)	(0·040)	(0·000)	(0·099)
	1	3	−1·639	−0·417	−192·0	−1·641	6·778
			(0·480)	(0·776)	(0·032)	(0·000)	(0·079)
	3	1	−0·679	−0·863	−256·5	−1·878	7·162
			(0·797)	(0·592)	(0·026)	(0·000)	(0·067)
	3	3	−0·960	−1·108	−250·9	−1·661	8·445
			(0·708)	(0·488)	(0·017)	(0·000)	(0·038)
		BIC	−0·903	−0·789	−246·5	−1·810	7·308
			(0·729)	(0·620)	(0·024)	(0·000)	(0·063)

Notes: The table reports the coefficient estimates of δ_0 and δ_1 in equation (21), for different choices of z_t. $\widehat{\beta}_t$, $\widehat{\gamma}_t$ and $\widehat{\rho}_t$ are rolling estimates of the structural parameters in the monetary policy reaction function of the Fed described in 20, and $\widehat{\sigma}^2_{\pi,t}$ is a rolling estimate of volatility of inflation changes. The numbers within parentheses are the p-value of the test of significance of the individual coefficient. The last column reports the Wald test statistic $W_{m,n,\tau}$ introduced in Section 5 (with an HAC bandwidth equal to $n^{1/3}$) and its associated p-value (in parentheses). q_u and q_π are, respectively, the number of lags used for unemployment and for inflation. Rows labelled "BIC" report results for the case in which the lag length is determined by the BIC with a maximum of three lags. τ is the forecast horizon.

estimates of the coefficients in the following equation:

$$SL_{t+\tau} = \delta_0 + z_t'\delta_1 + \varepsilon_{t+\tau} \tag{21}$$

where z_t is either $\widehat{\beta}_t$, $\widehat{\gamma}_t$, $\widehat{\rho}_t$ (the rolling estimates of the parameters in (20)), or $\widehat{\sigma}^2_{\pi,t}$, and $\tau = 1$, 3, 12 months. The table reports estimates of δ_1 and (in parentheses) the p-values associated with testing whether δ_1 equals zero.[15] It is clear that the degree of inflation target smoothing operated

15. The test statistic is implemented with Newey and West's (1987) HAC estimator with a bandwidth equal to $n^{1/3}$ and the p-values are calculated from (2).

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

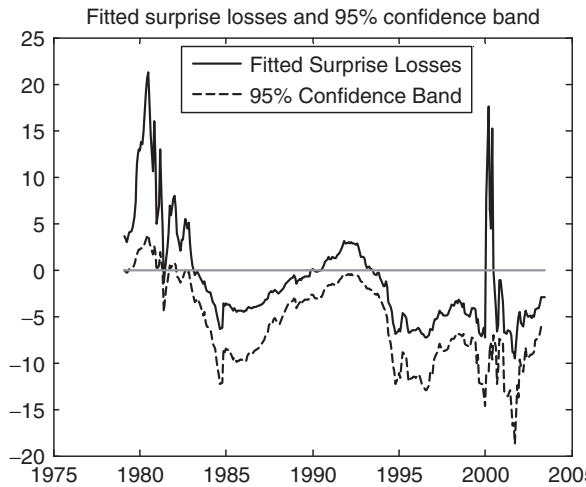

FIGURE 3

Fitted surprise losses

by the central bank ($\widehat{\rho}_t$) and the degree of inflation volatility ($\widehat{\sigma}^2_{\pi,t}$) explain the behaviour of the surprise losses at the 12-month horizon, whereas inflation volatility and the degree of the Fed's risk aversion to the unemployment gap ($\widehat{\gamma}_t$) are significant at the 1-month horizon. We also estimate (21) with $z_t = (\widehat{\beta}_t, \widehat{\gamma}_t, \widehat{\rho}_t)$ and find strong evidence of joint significance at horizons of 1 and 12 months (last column of Table 5). To conclude, Figure 3 plots the sequence of surprise losses $\widehat{SL}_{t+12}$ along with its one-sided 95% confidence band, and shows empirical evidence of forecast breakdowns during the Volker era (1979:3–1987:7) but not during the Greenspan era (1987:7 onwards).

9. CONCLUSION

This paper proposed a method for detecting and predicting forecast breakdowns, defined as a situation in which the out-of-sample performance of a forecast model is significantly worse than its in-sample performance. Unlike the literature evaluating a forecasting model from the perspective of whether it produces optimal forecasts, we focus on whether the model's forecast performance – measured by a general loss function – is consistent with expectations based on the model's earlier fit. The analysis of the possible causes of forecast breakdowns reveals the prime role played by instabilities in the data-generating process in causing forecast breakdowns, thus establishing a link between this paper and the structural break testing literature. Among the differences, we note that our approach is loss-specific and thus directly captures the effect of various types of instabilities on the model's forecast performance, whereas an indirect approach that tests for those instabilities may give misleading conclusions. A further advantage of our approach is that it allows the forecaster to predict future forecast breakdowns, by directly relating the differences between out-of- and in-sample performances to observable economic variables.

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

APPENDIX: PROOFS

Notation 1. *Let* $L_t^* \equiv L_t(\beta^*)$, $\partial L_t^* \equiv \partial L_t(\beta^*)$, $t = 1, \ldots, T$; $\widetilde{L}_t^* \equiv L_t^* - E\left(L_t^*\right)$; $\partial \widetilde{L}_t^* = \partial L_t^* - E\left(\partial L_t^*\right)$; $D_{t+\tau} \equiv \partial L_{t+\tau}(\widehat{\beta}_t)/\partial\beta - \partial \overline{L}_t(\widehat{\beta}_t)/\partial\beta$, $t = m, \ldots, T-\tau$ *a* $1 \times k$ *vector;* $\mathrm{D}_{t+\tau}^* \equiv \partial SL_{t+\tau}(\beta^*)/\partial\beta$; $\tilde{D}_{t+\tau}^* = D_{t+\tau}^* - E(D_{t+\tau}^*)$. *For a matrix* A, $|\mathrm{A}| = \max_{i,j} |a_{ij}|$. *Limits are for* $m, n \to \infty$. *Let* h *denote the* $q \times T$ *matrix of orthogonality conditions, with element* h_t, $t = 1, \ldots, T$:

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

$h_t^* \equiv h_t(\beta^*)$; B_t *a consistent estimate of* B_t^* *from Assumption A3, substituting* $\widehat{\beta}_t$ *for* β^*;

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

$$b_{m,j} = \frac{D_{m+\tau+j} B_{m+j}}{m+j} + \frac{D_{m+\tau+j+1} B_{m+j+1}}{m+j+1} + \cdots + \frac{D_T B_{T-\tau}}{T-\tau}; \tag{A1}$$

$$\textit{Rolling } (n \geq m): \underset{1\times qT}{w^h} = \Bigg[\underbrace{\frac{D_{m+\tau} B_m}{m}, \ldots, \frac{\sum_{t=m}^{2m-1} D_{t+\tau} B_t}{m}}_{m}, \underbrace{\frac{\sum_{t=m+1}^{2m} D_{t+\tau} B_t}{m}, \ldots, \frac{\sum_{t=n}^{T-\tau} D_{t+\tau} B_t}{m}}_{n-m},$$
$$\underbrace{\frac{\sum_{t=n+1}^{T-\tau} D_{t+\tau} B_t}{m}, \ldots, \frac{D_T B_{T-\tau}}{m}}_{m-1}, \underbrace{0, \ldots, 0}_{\tau}\Bigg].$$

$$\textit{Rolling } (n < m): \underset{1\times qT}{w^h} = \Bigg[\underbrace{\frac{D_{m+\tau} B_m}{m}, \ldots, \frac{\sum_{t=m}^{T-\tau} D_{t+\tau} B_t}{m}}_{n}, \underbrace{\frac{\sum_{t=m}^{T-\tau} D_{t+\tau} B_t}{m}, \ldots, \frac{\sum_{t=m}^{T-\tau} D_{t+\tau} B_t}{m}}_{m-n},$$
$$\underbrace{\frac{\sum_{t=m+1}^{T-\tau} D_{t+\tau} B_t}{m}, \ldots, \frac{D_T B_{T-\tau}}{m}}_{n-1}, \underbrace{0, \ldots, 0}_{\tau}\Bigg].$$

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

$$V_T = \begin{pmatrix} V_T^{LL} & V_T^{Lh} \\ V_T^{Lh} & V_T^{hh} \end{pmatrix}, \textit{ where} \tag{A2}$$

$$V_T^{LL} \equiv T^{-1} \sum_{t=1}^{T} (w_t^L \widetilde{L}_t)^2 + 2T^{-1} \sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} w_t^L \widetilde{L}_t w_{t-j}^L \widetilde{L}_{t-j}; \tag{A3}$$

$$V_T^{hh} \equiv T^{-1} \sum_{t=1}^{T} w_t^h h_t h_t' w_t^{h\prime} + T^{-1} \sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} \left(w_t^h h_t h_{t-j}' w_{t-j}^{h\prime} + w_{t-j}^h h_{t-j} h_{t-j}' w_t^{h\prime}\right); \tag{A4}$$

$$V_T^{Lh} \equiv T^{-1} \sum_{t=1}^{T} w_t^L \widetilde{L}_t h_t' w_t^{h\prime} + T^{-1} \sum_{j=1}^{p_T} v_{T,j} \sum_{t=j}^{T} \left(w_t^L \widetilde{L}_t h_{t-j}' w_{t-j}^{h\prime} + w_{t-j}^L \widetilde{L}_{t-j} h_{t-j}' w_t^{h\prime}\right), \tag{A5}$$

with p_T *and* $v_{T,j}$ *appropriately defined (Andrews, 1991 or Newey and West, 1987).*

Assumption A5′. $T^{-1} \sum_{t=1}^{T} E\left(\partial L_t(\beta^*)/\partial\beta\right) < \infty$ for all T.

Proposition 5 **(Generalization of forecast breakdown test)** . *Given Assumptions A1–A4, A5′, A6, A7, if* V_T *in (A2) is p.d.,* $\hat{\sigma}_{m,n} = \sqrt{(T/n)\,(V_T^{LL} + V_T^{hh} + 2V_T^{Lh})}$, V_T^{LL}, V_T^{hh} *and* V_T^{Lh} *given in (A3)–(A5). Then,* $t_{m,n,\tau} \xrightarrow{d} N(0,1)$ *under* H_0 *in (3).*

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

Lemma 1. *(a)* $R_1 \equiv n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{D}^*_{t+\tau}B^*_t H^*_t = o_p(1)$;
(b) $R_2 \equiv 0{\cdot}5n^{-1/2}\sum_{t=m}^{T-\tau}\left(\widehat{\beta}_t-\beta^*\right)'\left(\partial^2 SL_{t+\tau}(\overline{\beta^*_t})/\partial\beta\partial\beta'\right)\left(\widehat{\beta}_t-\beta^*\right) = o_p(1)$, *where* $\overline{\beta^*_t}$ *is an intermediate point between* $\widehat{\beta}_t$ *and* β^*.

Proof of Lemma 1. *(a)* We focus for simplicity on the recursive scheme. The proofs for the fixed and rolling schemes are similar. Direct calculations show that $R_1^{-1/2}\sum_{t=1}^T \tilde{w}^h_t h^*_t$, where

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

We will show that $E\left(n^{-1/2}\sum_{t=1}^T \tilde{w}^h_t h^*_t\right)^2 \xrightarrow{p} 0$ from which the result follows because convergence in mean square implies convergence in probability.

First note that $\tilde{w}^h_t$ can be written as a weighted average of the scores: $\tilde{w}^h_t = T^{-1}\sum_{j=1}^T \widetilde{\partial L}^*_j P_{t,j}$. For example, $\tilde{w}^h_1 = c_{m,0} = T^{-1}\sum_{j=1}^T \widetilde{\partial L}^*_j P_{1,j}$ with (non-stochastic) weights

$$P_1 = T\Bigg[\underbrace{d_{m,0},\ldots,d_{m,0}}_{m},\underbrace{d_{m,1},\ldots,d_{m,\tau-1}}_{\tau-1},\underbrace{\frac{B^*_m}{m}-d_{m,\tau},\ldots,\frac{B^*_{m+n-\tau-1}}{m+n-\tau-1}-d_{m,n-1},}_{n-\tau}$$

$$\underbrace{\frac{B^*_{m+n-\tau}}{m+n-\tau},\ldots,\frac{B^*_{T-\tau}}{T-\tau}}_{\tau}\Bigg],\ \text{where}$$

$$d_{m,j}=\sum_{i=1}^{n-j}\frac{B^*_{m+j+i-1}}{(m+j+i-1)^2}.$$

Similar expressions can be derived for $c_{m,j}$, $j=1,\ldots,n-1$. Therefore, $E\left(n^{-1/2}\sum_{t=1}^T \tilde{w}^h_t h^*_t\right)^2 = E\left(n^{-1/2}\sum_{t=1}^T\left[T^{-1}\sum_{j=1}^T \widetilde{\partial L}^*_j P_{t,j}\right]h^*_t\right)^2$. We have $E\left(n^{-1/2}\sum_{t=1}^T\left[T^{-1}\sum_{j=1}^T \widetilde{\partial L}^*_j P_{t,j}\right]h^*_t\right)^2 = A_{1T}+A_{2T}+A_{3T}$, where

$$A_{1T}\equiv\left(nT^2\right)^{-1}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^T\sum_{j=1}^T E\left(h^{*\prime}_t h^*_s\right)E\left(\widetilde{\partial L}^*_i P_{t,i}P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right),$$

$$A_{2T}\equiv\left(nT^2\right)^{-1}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^T\sum_{j=1}^T\left[E\left(h^{*\prime}_t P'_{t,i}\widetilde{\partial L}^{*\prime}_i\right)E\left(h^{*\prime}_s P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right)+E\left(h^{*\prime}_t P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right)E\left(h^{*\prime}_s P'_{t,i}\widetilde{\partial L}^{*\prime}_i\right)\right],$$

$$A_{3T}\equiv\left(nT^2\right)^{-1}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^T\sum_{j=1}^T\kappa(t,t-s,t-i,t-j),$$

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

$$\begin{aligned}\kappa(t,t-s,t-i,t-j) &= E\left(h^{*\prime}_t h^*_s\widetilde{\partial L}^*_i P_{t,i}P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right)-E\left(h^{*\prime}_t h^*_s\right)E\left(\widetilde{\partial L}^*_i P_{t,i}P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right)\\ &\quad - E\left(h^{*\prime}_t P'_{t,i}\widetilde{\partial L}^{*\prime}_i\right)E\left(h^{*\prime}_s P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right)-E\left(h^{*\prime}_t P'_{s,j}\widetilde{\partial L}^{*\prime}_j\right)E\left(h^{*\prime}_s P'_{t,i}\widetilde{\partial L}^{*\prime}_i\right).\end{aligned}$$

Note that $|A_{1T}|\le\left(nT^2\right)^{-1}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^T\sum_{j=1}^T\left|E\left(h^{*\prime}_t h^*_s\right)\right|\left|E\left(\widetilde{\partial L}^*_i P^{\sup}_i P^{\sup\prime}_j\widetilde{\partial L}^{*\prime}_j\right)\right|$. Redefining $\widetilde{\partial L}^*_i P^{\sup}_i$ as $\widetilde{\partial L}^*_i$, we thus have $|A_{1T}|\le\left(nT^2\right)^{-1}\sum_{t=1}^T\sum_{s=1}^T\left|E\left(h^{*\prime}_t h^*_s\right)\right|\sum_{i=1}^T\sum_{j=1}^T\left|E\left(\widetilde{\partial L}^*_i\widetilde{\partial L}^{*\prime}_j\right)\right|\le$ $\left(nT^2\right)^{-1}C_2\left(\sum_{j=0}^\infty j\alpha(j)^{1-1/2r}\right)^2$, where C_2 is some positive and finite constant and $\alpha(j)$ are the mixing coefficients. As shown by Davidson (1994, p. 210), $\sum_{j=0}^\infty j\alpha(j)^{1-1/2r}$ is positive and finite, which implies that $A_{1T}\to 0$. A similar argument can be used to show that $A_{2T}\to 0$. For A_{3T}, we have

$$|A_{3T}|\le\left(nT^2\right)^{-1}\sum_{s=1}^\infty\sum_{i=1}^\infty\sum_{j=1}^\infty\sup_{t\ge1}|\kappa(t,t-s,t-i,t-j)|\to 0,$$

as $\sum_{s=1}^\infty\sum_{i=1}^\infty\sum_{j=1}^\infty\sup_{t\ge1}|\kappa(t,t-s,t-i,t-j)|<\infty$, by assumptions A1 and A4, as shown by Andrews (1991).

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

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

$$\begin{aligned}
R_2 &= \left| 0{\cdot}5 n^{-1/2} \sum_{t=m}^{T-\tau} t^{1-a} \left(\widehat{\beta}_t - \beta^*\right)' \left(t^{a-1} \frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'} \right) \left(\widehat{\beta}_t - \beta^*\right) \right| \\
&\leq C \sup_{m \leq t \leq T-\tau} |t^{0{\cdot}5-0{\cdot}5a} \left(\widehat{\beta}_t - \beta^*\right)|^2 n^{-1/2} \sum_{t=m}^{T-\tau} t^{a-1} \left| \frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'} \right| \\
&\leq C \sup_{m \leq t \leq T-\tau} |t^{0{\cdot}5-0{\cdot}5a} \left(\widehat{\beta}_t - \beta^*\right)|^2 n^{-1/2} \sum_{t=m}^{T-\tau} t^{a-1} \left(\left| \frac{\partial^2 L_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'} \right| + \left| \frac{\partial^2 \overline{L}_t(\overline{\beta_t^*})}{\partial\beta\partial\beta'} \right| \right) \\
&\leq C \sup_{m \leq t \leq T-\tau} |t^{0{\cdot}5-0{\cdot}5a} \left(\widehat{\beta}_t - \beta^*\right)|^2 n^{-1/2} \sum_{t=m}^{T-\tau} t^{a-1} \left(m_{t+\tau} + \overline{m}_t\right) = o_p(1)
\end{aligned}$$

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

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

Proof of Lemma 2. We prove Lemma 2 for the recursive scheme. The proofs for the fixed and rolling schemes are similar. First consider $0 < \pi < \infty$. Write $\frac{T}{n} V_T^{LL*} = \text{var}(A_1 + A_2 + A_3 + A_4)$, where $A_1 = -n^{-1/2} a_{m,0} (\widetilde{L}_1^* + \cdots + \widetilde{L}_m^*)$; $A_2 = -n^{-1/2} \left(a_{m,1} \widetilde{L}_{m+1}^* + \cdots + a_{m,\tau-1} \widetilde{L}_{m+\tau-1}^*\right)$; $A_3 = n^{-1/2} \left[\left(1 - a_{m,\tau}\right) \widetilde{L}_{m+\tau}^* + \cdots + \left(1 - a_{m,n-1}\right) \widetilde{L}_{T-\tau}^*\right]$; $A_4 = n^{-1/2} \left(\widetilde{L}_{T-\tau+1} + \cdots + \widetilde{L}_T\right)$. We first show that $|\text{cov}(A_i, A_j)| \to 0$ for $i \neq j$. As $a_{m,j} \leq a_{m,0}$, $|\text{cov}(A_1, A_2)| \leq n^{-1} a_{m,0}^2 |\text{cov}(\sum_{t=1}^{m} \widetilde{L}_t^*, \sum_{t=m+1}^{m+\tau-1} \widetilde{L}_t^*)| | \leq n^{-1} a_{m,0}^2 \sum_{t=1}^{m} \sum_{j=1}^{\tau-1} |E(\widetilde{L}_t^* \widetilde{L}_{t+j}^*)| \leq n^{-1} a_{m,0}^2 C \sum_{j=0}^{\infty} j\alpha(j)^{1-1/2r}$ using Corollary 6.17 of White (2001), where C is some positive and finite constant and $\alpha(j)$ are the mixing coefficients. By Davidson (1994, p. 210), $\sum_{j=0}^{\infty} j\alpha(j)^{1-1/2r}$ is positive and finite. Further, $a_{m,0}^2 \to \ln^2(1+\pi)$, which is finite (West, 1996, p. 1082). As a result, $\text{cov}(A_1, A_2) \to 0$. Using analogous reasonings and the fact that $1 - a_{m,t-m} \leq 1$ for all t, one can show that $|\text{cov}(A_i, A_j)| \to 0$ for the remaining (i, j) pairs. We thus have that $\text{var}\left(n^{-1/2} \sum_{t=1}^{T} w_t^L \widetilde{L}_t^*\right)$ can be approximated by $\sum_{i=1}^{4} \text{var}(A_i)$ and the desired result follows from the fact that, *e.g.* $\text{var}(A_1) = (m/n) a_{m,0}^2 \text{var}(m^{-1/2} \sum_{t=1}^{m} \widetilde{L}_t^*) > 0$ As $m/n \to \pi^{-1} > 0$, $a_{m,0}^2 \to \ln^2(1+\pi) > 0$, and $\text{var}(m^{-1/2} \sum_{t=1}^{m} \widetilde{L}_t^*) > 0$ using Assumption A6. When $\pi = 0$ it is sufficient to show that $\text{var}(A_3) \geq \left(1 - a_{m,0}\right)^2 \text{var}\left(n^{-1/2} \sum_{t=m+\tau}^{T-\tau} \tilde{L}_t^*\right) > 0$ as $\left(1 - a_{m,0}\right) \to 1$ and $\text{var}\left(n^{-1/2} \sum_{t=m+\tau}^{T-\tau} \tilde{L}_t^*\right) > 0$. ‖

Proof of Proposition 1. A second-order mean value expansion of $SL_{t+\tau}(\widehat{\beta}_t) = L_{t+\tau}\left(\widehat{\beta}_t\right) - \overline{L}_t\left(\widehat{\beta}_t\right)$ around β^* gives

$$\begin{aligned}
& n^{1/2} \left[n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t) - E\left(n^{-1} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\beta^*) \right) \right] \qquad \text{(A6)} \\
&= n^{-1/2} \sum_{t=m}^{T-\tau} \left[SL_{t+\tau}\left(\beta^*\right) - E\left(SL_{t+\tau}\left(\beta^*\right)\right) \right] + n^{-1/2} \sum_{t=m}^{T-\tau} \frac{\partial SL_{t+\tau}(\beta^*)}{\partial\beta} \left(\widehat{\beta}_t - \beta^*\right) \\
&\quad + 0{\cdot}5 n^{-1/2} \sum_{t=m}^{T-\tau} \left(\widehat{\beta}_t - \beta^*\right)' \frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'} \left(\widehat{\beta}_t - \beta^*\right) \\
&= n^{-1/2} \sum_{t=m}^{T-\tau} \left[SL_{t+\tau}\left(\beta^*\right) - E\left(SL_{t+\tau}\left(\beta^*\right)\right) \right] + n^{-1/2} \sum_{t=m}^{T-\tau} E\left(D_{t+\tau}^*\right) B_t^* H_t^* \\
&\quad + n^{-1/2} \sum_{t=m}^{T-\tau} \widetilde{D}_{t+\tau}^* B_t^* H_t^* + 0{\cdot}5 n^{-1/2} \sum_{t=m}^{T-\tau} \left(\widehat{\beta}_t - \beta^*\right)' \frac{\partial^2 SL_{t+\tau}(\overline{\beta_t^*})}{\partial\beta\partial\beta'} \left(\widehat{\beta}_t - \beta^*\right) \\
&= n^{-1/2} \sum_{t=m}^{T-\tau} \left[SL_{t+\tau}\left(\beta^*\right) - E\left(SL_{t+\tau}\left(\beta^*\right)\right) \right] + n^{-1/2} \sum_{t=m}^{T-\tau} E\left(D_{t+\tau}^*\right) B_t^* H_t^* + o_p(1)
\end{aligned}$$

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

where $\overline{\beta_t^*}$ is some intermediate point between $\widehat{\beta}_t$ and β^* and where we have used Assumption A3 and Lemma 1. We show that, under H_0,

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

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

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

where $V_T^* = \text{var}\left(T^{-1/2}\left[\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*, \sum_{t=1}^{T} w_t^{h*} h_t^*\right]'\right)$. The result follows from the fact that $V_T - V_T^* \xrightarrow{p} 0$, owing to consistency of $\widehat{\beta}_t$ for β^* under H_0. We verify that the zero-mean vector sequence $\left\{\left[V_T^{*-1/2} w_t^L \widetilde{L}_t^*, V_T^{*-1/2} w_t^{h*} h_t^*\right]'\right\}_{t=1}^{T}$ satisfies the conditions of Wooldridge and White's (1988) Central Limit Theorem for mixing processes. As $Z_t \equiv \left[V_T^{*-1/2} w_t^L \widetilde{L}_t^*, V_T^{*-1/2} w_t^{h*} h_t^*\right]$ is a function of only a finite number of leads and lags of W_t, it follows from Lemma 2.1 of White and Domowitz (1984) that it is mixing of the same size as W_t. For the first component of Z_t, we have $E|V_T^{*-1/2} w_t^L \widetilde{L}_t^*|^{2r} < \infty$ using Assumption A4 and by the fact that V_T^* is p.d. and $|w_t^L| < \infty$ for all t (for the fixed and rolling schemes, this follows from Assumption A7; for the recursive scheme, it follows from the fact that $a_{m,j} \leq a_{m,0} \to \ln(1+\pi) < \infty$, as shown in the proof of Lemma 2). For the second component of Z_t, writing $w_t^{h*} = T^{-1}\sum_{j=1}^{T} E\left(\partial L_j^*\right) P_{t,j}$ – using similar reasonings as those in the proof of Lemma 1(a) – we have $E|V_T^{*-1/2} w_t^{h*} h_t^*|^{2r} = E|V_T^{*-1/2} T^{-1}\sum_{j=1}^{T} E\left(\partial L_j^*\right) P_{t,j} h_t^*|^{2r} \equiv E|\lambda_t h_t^*|^{2r}$. Note that $|\lambda_{t,i}| < \infty$ for all t, i, using Assumption A5′, by $P_{t,j}$ having bounded components (as shown in the proof of Lemma 1(a)) and by V_T^* p.d. Further, using Minkowski's inequality,

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

using Assumption A4. This implies that $V_T^{*-1/2} T^{-1/2}\left[\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*, \sum_{t=1}^{T} w_t^{h*} h_t^*\right]' \xrightarrow{d} N(0, I_2)$. The desired result then follows from consistency of V_T for V_T^* owing to $\widehat{\beta}_t - \beta^* \xrightarrow{p} 0$ under H_0. ‖

Proof of Theorem 1. Given A5, $E\left(D_{t+\tau}^*\right) = E\left(\partial SL_{t+\tau}(\beta^*)/\partial\beta\right) = E\left(\partial L_{t+\tau}(\beta^*)/\partial\beta\right) - E\left(\partial \overline{L}_t(\beta^*)/\partial\beta\right) = 0$, expression (A8) reduces to $n^{-1/2}\sum_{t=m}^{T-\tau}\left[SL_{t+\tau}\left(\beta^*\right) - E\left(SL_{t+\tau}\left(\beta^*\right)\right)\right] + o_p(1)$. The result then follows from reasonings analogous to those in the proof of Proposition 5 and from Lemma 2. ‖

Lemma 3. *For $a_{m,j}$ as defined in Algorithm 1, we have: (i) $a_{m,j} \simeq \ln(m+n-1/(m+j))$; (ii) $n^{-1}\sum_{j=\tau}^{n-1} a_{m,j} \simeq 1 - \pi^{-1}\ln(1+\pi)$; (iii) $n^{-1}\sum_{j=\tau}^{n-1} a_{m,j}^2 \simeq 2\left[1 - \pi^{-1}\ln(1+\pi)\right] - \pi^{-1}\ln(1+\pi)$.*

Proof of Lemma 3. *(i)* $a_{m,j} = \sum_{i=j}^{n-1}(m+i)^{-1} \simeq \int_j^{n-1}(m+x)^{-1}dx = \ln(m+n-1/(m+j))$; *(ii)* $n^{-1}\sum_{j=\tau}^{n-1} a_{m,j} \simeq n^{-1}\int_\tau^{n-1}\ln(m+n-1/(m+x))\,dx = n^{-1}\left[n-1-\tau-(m-\tau)\ln(m+n-1/(m+\tau))\right] \to 1-\pi^{-1}\ln(1+\pi)$; *(iii)* $n^{-1}\sum_{j=\tau}^{n-1} a_{m,j}^2 \simeq n^{-1}\int_\tau^{n-1}\ln^2(m+n-1/(m+x))\,dx = n^{-1}\left[2(n-\tau) - 2(m+\tau)\ln(m+n-1/(m+\tau)) - (m+\tau)\ln^2(m+n-1/(m+\tau))\right] \to 2\left[1-\pi^{-1}\ln(1+\pi)\right] - \pi^{-1}\ln(1+\pi)$. ‖

Proof of Corollary 1. We show that $\lim \text{var}(n^{-1/2}\sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \lambda^* \sum_{j=-\infty}^{\infty}\Gamma_j$, where $\lambda^* = 1+\pi$ for the fixed scheme; $\lambda^* = 1-(1/3)\pi^2$ for the rolling $(n < m)$ scheme; $\lambda^* = (2/3)\pi^{-1}$ for the rolling $(n \geq m)$ scheme;

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

$\lambda^* = 1$ for the recursive scheme. The desired result then follows from λS_n^{LL} being a consistent estimator of $\lambda^* \sum_{j=-\infty}^{\infty} \Gamma_j$ under H_0. For conciseness, we focus on the recursive scheme. As shown in the proof of Lemma 2, $var(n^{-1/2} \sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \sum_{i=1}^{4} \text{var}(A_i)$. We have $\text{var}(A_1) = (m/n)\, a_{m,0}^2 \text{var}(m^{-1/2} \sum_{t=1}^{m} \widetilde{L}_t^*)$ and thus $\lim \text{var}(A_1) = \pi^{-1} \ln(1+\pi) \sum_{j=-\infty}^{\infty} \Gamma_j$ using Lemma 3*(i)*. Further, $\text{var}\,(A_2) = n^{-1} \text{var}\left(a_{m,1} \widetilde{L}_{m+1}^* + \cdots + a_{m,\tau-1} \widetilde{L}_{m+\tau-1}^*\right) \to 0$ as τ is fixed. For A_3, it follows from West (1996, pp. 1082–3), (with $(1 - a_{m,j})$ substituting $a_{m,j}$) that $\text{var}(A_3) = n^{-1} d_0 \sum_{j=-n+2}^{n-2} \Gamma_j + o(1)$, where $d_0 = \sum_{j=\tau}^{n-1} (1 - a_{m,j})^2$. Using Lemma 3, $n^{-1} d_0 = (n-\tau)/n - 2n^{-1} \sum_{j=\tau}^{n-1} a_{m,j} + n^{-1} \sum_{j=\tau}^{n-1} a_{m,j}^2 \to 1 - \pi^{-1} \ln(1+\pi)$, and thus $\lim \text{var}(A_3) = \left[1 - \pi^{-1} \ln(1+\pi)\right] \sum_{j=-\infty}^{\infty} \Gamma_j$. Finally, $\text{var}(A_4) = n^{-1} var(\widetilde{L}_{T-\tau+1} + \cdots + \widetilde{L}_T) \to 0$ as τ is fixed. In sum, we have $\text{var}(n^{-1/2} \sum_{t=1}^{T} w_t^L \widetilde{L}_t^*) = \sum_{j=-\infty}^{\infty} \Gamma_j$ and thus $\lambda^* = 1$. The proofs for the fixed and rolling schemes follow from similar reasonings. ‖

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

$$n^{-1/2} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t) = n^{-1/2} \sum_{t=m}^{T-\tau} SL_{t+\tau}\left(\beta_t^*\right) + n^{-1/2} \sum_{t=m}^{T-\tau} \left(\frac{\partial L_{t+\tau}\left(\beta_t^*\right)}{\partial \beta} - \frac{\partial \overline{L}_t\left(\beta_t^*\right)}{\partial \beta}\right)\left(\widehat{\beta}_t - \beta_t^*\right)$$
$$+ \cdot 5 n^{-1/2} \sum_{t=m}^{T-\tau} \left(\widehat{\beta}_t - \beta_t^*\right)' \left(\frac{\partial^2 L_{t+\tau}(\overline{\widehat{\beta}}_t)}{\partial \beta \partial \beta'} - \frac{\partial^2 \overline{L}_t(\overline{\widehat{\beta}}_t)}{\partial \beta \partial \beta'}\right)\left(\widehat{\beta}_t - \beta_t^*\right) \tag{A7}$$

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

$$L_{t+\tau}\left(\beta_t^*\right) = L_{t+\tau}\left(\beta_{t+\tau}^*\right) + \frac{\partial L_{t+\tau}\left(\beta_{t+\tau}^*\right)}{\partial \beta}\left(\beta_t^* - \beta_{t+\tau}^*\right) + 0{\cdot}5\left(\beta_t^* - \beta_{t+\tau}^*\right)' \frac{\partial^2 L_{t+\tau}\left(\overline{\beta_{t+\tau}^*}\right)}{\partial \beta \partial \beta'}\left(\beta_t^* - \beta_{t+\tau}^*\right) \tag{A8}$$

$$L_j\left(\beta_t^*\right) = L_j\left(\beta_j^*\right) + \frac{\partial L_j\left(\beta_j^*\right)}{\partial \beta}\left(\beta_t^* - \beta_j^*\right) + 0{\cdot}5\left(\beta_t^* - \beta_j^*\right)' \frac{\partial^2 L_j\left(\overline{\beta_j^*}\right)}{\partial \beta \partial \beta'}\left(\beta_t^* - \beta_j^*\right) \tag{A9}$$

where $\overline{\beta_{t+\tau}^*}$ is an intermediate point between β_t^* and $\beta_{t+\tau}^*$, and $\overline{\beta_j^*}$ is an intermediate point between β_t^* and β_j^*. From (A8) and (A9) above, it follows that

$$SL_{t+\tau}\left(\beta_t^*\right) = L_{t+\tau}\left(\beta_{t+\tau}^*\right) - \overline{\sum}_j L_j\left(\beta_j^*\right) + \frac{\partial L_{t+\tau}\left(\beta_{t+\tau}^*\right)}{\partial \beta}\left(\beta_t^* - \beta_{t+\tau}^*\right) - \overline{\sum}_j \frac{\partial L_j\left(\beta_j^*\right)}{\partial \beta}\left(\beta_t^* - \beta_j^*\right)$$
$$+ 0{\cdot}5\left[\left(\beta_t^* - \beta_{t+\tau}^*\right)' \frac{\partial^2 L_{t+\tau}\left(\overline{\beta_{t+\tau}^*}\right)}{\partial \beta \partial \beta'}\left(\beta_t^* - \beta_{t+\tau}^*\right)\right] \tag{A10}$$
$$- 0{\cdot}5 \overline{\sum}_j \left(\beta_t^* - \beta_j^*\right)' \frac{\partial^2 L_j\left(\overline{\beta_j^*}\right)}{\partial \beta \partial \beta'}\left(\beta_t^* - \beta_j^*\right).$$

Substituting (A10) into (A7) gives:

$$n^{-1/2} \sum_{t=m}^{T-\tau} SL_{t+\tau}(\widehat{\beta}_t) = n^{-1/2} \sum_{t=m}^{T-\tau} \left(L_{t+\tau}\left(\beta_{t+\tau}^*\right) - \overline{\sum}_j L_j\left(\beta_j^*\right)\right) \tag{A11}$$
$$+ n^{-1/2} \sum_{t=m}^{T-\tau} \left[\frac{\partial L_{t+\tau}\left(\beta_{t+\tau}^*\right)}{\partial \beta}\left(\beta_t^* - \beta_{t+\tau}^*\right) - \overline{\sum}_j \frac{\partial L_j\left(\beta_j^*\right)}{\partial \beta}\left(\beta_t^* - \beta_j^*\right)\right]$$
$$+ 0{\cdot}5 n^{-1/2} \sum_{t=m}^{T-\tau} \left[\left(\beta_t^* - \beta_{t+\tau}^*\right)' \frac{\partial^2 L_{t+\tau}\left(\overline{\beta_{t+\tau}^*}\right)}{\partial \beta \partial \beta'}\left(\beta_t^* - \beta_{t+\tau}^*\right)\right.$$
$$\left. - \overline{\sum}_j \left(\beta_t^* - \beta_j^*\right)' \frac{\partial^2 L_j\left(\overline{\beta_j^*}\right)}{\partial \beta \partial \beta'}\left(\beta_t^* - \beta_j^*\right)\right]$$

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

$$+ n^{-1/2} \sum_{t=m}^{T-\tau} \left(\frac{\partial L_{t+\tau}\left(\beta_t^*\right)}{\partial \beta} - \frac{\partial \overline{L}_t\left(\beta_t^*\right)}{\partial \beta} \right) \left(\widehat{\beta}_t - \beta_t^*\right)$$
$$+ 0{\cdot}5 n^{-1/2} \sum_{t=m}^{T-\tau} \left(\widehat{\beta}_t - \beta_t^*\right)' \left(\frac{\partial^2 L_{t+\tau}(\overline{\widehat{\beta}}_t)}{\partial \beta \partial \beta'} - \frac{\partial^2 \overline{L}_t(\overline{\widehat{\beta}}_t)}{\partial \beta \partial \beta'} \right) \left(\widehat{\beta}_t - \beta_t^*\right).$$

Note that, as $0 = \partial \overline{\mathcal{L}}_t\left(\widehat{\beta}_t\right)/\partial\beta = \partial \overline{\mathcal{L}}_t\left(\beta_t^*\right)/\partial\beta + \left(\partial^2 \overline{\mathcal{L}}_t(\overline{\widehat{\beta}}_t)/\partial\beta\partial\beta'\right)\left(\widehat{\beta}_t - \beta_t^*\right)$, then $\partial L_{t+\tau}\left(\beta_t^*\right)/\partial\beta - \partial \overline{L}_t\left(\beta_t^*\right)/\partial\beta = \partial L_{t+\tau}\left(\beta_t^*\right)/\partial\beta - \partial\left(\overline{L}_t\left(\beta_t^*\right) - \overline{\mathcal{L}}_t\left(\beta_t^*\right)\right)/\partial\beta + \left(\widehat{\beta}_t - \beta_t^*\right)'\left(\partial^2 \overline{\mathcal{L}}_t(\overline{\widehat{\beta}}_t)/\partial\beta\partial\beta'\right)$. Therefore, by taking expectations of (A11), we have (8). ‖

Proof of Proposition 3. 1. Consider first Elliott and Muller's (2006) qLL test, where $qLL = \widehat{v}'\left[G_a - M_e\right]\widehat{v}$, $\widehat{v} = V_X^{-1/2}\Xi M y$ and G_a and M_e are deterministic matrices.[16] Note that, when $\Delta\beta = \Delta\mu^X = 0$, the DGP satisfies Elliott and Muller's (2006) Conditions 2 and 3. Their Theorem 4 shows that qLL involves functionals of $T^{-1/2}\left[e'_{[sT]}, 0'_{[(1-s)T]}\right]\left[I_T \otimes V_X^{-1}\right]\Xi' M y$, where $V_X = E\left(T^{-1}\sum_{j=1}^T \mathrm{X}_t^2 \varepsilon_t^2\right)$, which in the case $\Delta\beta = \Delta\mu^X = 0$ becomes $T^{-1/2}\left[e'_{[sT]}, 0'_{[(1-s)T]}\right]\left[I_T \otimes V_X^{-1}\right]\Xi' M\varepsilon \Rightarrow \xi$, so that the asymptotic distribution of qLL under the null hypothesis follows directly by applying their Lemma 2. Under our two scenarios we have: (a) $My = M\Xi\left(\Delta\beta\iota_T + \beta e_T\right) + M\varepsilon = M\Xi\Delta\beta\iota_T + M\varepsilon$, where ι_T is a $(T \times 1)$ vector s.t. its tth component is one if $t \geq m$ and zero otherwise. The result follows by using results similar to those in Rossi (2005, Section 4). (b) $My = M\Xi\beta e_T + M\varepsilon = M\varepsilon$. The result follows from

$$\xi_T = T^{-1/2} V_X^{-1} \sum_{j=1}^{[sT]} X_j \varepsilon_j - T^{-1} V_X^{-1} \sum_{j=1}^{[sT]} X_j^2 \left(T^{-1} \sum_{j=1}^{T} X_j^2 \right)^{-1} T^{-1/2} \sum_{j=1}^{T} X_j \varepsilon_j$$
$$\Rightarrow V_X^{-1} \int_0^s \left[\sigma_X + g_X(r)\right] dB(r)$$
$$- V_X^{-1} \left(\int_0^s \left[\sigma_X^2 + g_X^2(r)\right] dr \right) \left(\int_0^1 \left[\sigma_X^2 + g_X^2(r)\right] dr \right)^{-1} \left(\int_0^1 \left[\sigma_X + g_X(r)\right] dB(r) \right),$$

using Lemma 1 in Cavaliere (2004) and results from the proof of Theorem 4 in Elliott and Muller (2006). As in this case $V_X = \sigma_X^2 + \int_0^1 g_X^2(s)\,ds$, the limiting behaviour of ξ_T becomes the same as Elliott and Muller (2006) if $\Delta\mu^X = 0$ as:

$$V_X^{-1} \int_0^s \left[\sigma_X + g_X(r)\right] dB(r) - V_X^{-1} \left(\int_0^s \left[\sigma_X^2 + g_X^2(r)\right] dr \right) \left(\int_0^1 \left[\sigma_X^2 + g_X^2(r)\right] dr \right)^{-1} \left(\int_0^1 \left[\sigma_X + g_X(r)\right] dB(r) \right)$$
$$= V_X^{-1}\sigma_X \left[\int_0^s dB(r) - s \int_0^1 dB(r) \right] = V_X^{-1}\sigma_X \left[B(s) - sB(1)\right].$$

2. Using the decomposition of the forecast breakdown test in Proposition 2 we have: (a) as the variance of ε_t is constant, the "other instabilities" component is zero; as $E\left(\partial L_t\left(\beta_t^*\right)/\partial\beta\right) = 0\ \forall t$, the "parameter instabilities I" component is zero; all "estimation uncertainty" components vanish asymptotically, which means that only the "parameter instabilities II" component remains, which is: $(1/2)n^{-1/2}\sum_{t=m}^{T-\tau} E\left[\left(\beta - \left(\beta + n^{-\alpha}\Delta\beta\right)\right)'\sigma_X^2\left(\beta - \left(\beta + n^{-\alpha}\Delta\beta\right)\right)\right] = (1/2)n^{1/2-2\alpha}\sigma_X^2\Delta\beta^2$. This component does not vanish asymptotically provided $\alpha \leq 1/4$. (b) The "other instabilities" term does not depend on X_t; the "parameter instabilities" components are unaffected because $\beta_t^* = \beta\ \forall t$; as $\partial L_{t+\tau}(\beta)/\partial\beta = -2\mathrm{X}_{t+\tau}\left(\mathrm{Y}_{t+\tau} - \mathrm{X}'_{t+\tau}\beta\right)$ is uncorrelated with $\left(\widehat{\beta}_t - \beta\right)$, the "estimation uncertainty I" component is zero. Given that the loss function for estimation and evaluation is the same, the only component that (b) will affect is "estimation uncertainty II", which becomes:

$$n^{1/2} E\left(\widehat{\beta}_m - \beta\right)' \left(2m^{-1}\sum_{s=1}^{m} X_s^2\right)\left(\widehat{\beta}_m - \beta\right) = 2(n^{1/2}/m) E\left(m^{-1}\sum_{s=1}^{m} X_s^2\right)^{-1}\left(m^{-1/2}\sum_{s=1}^{m} X_s\varepsilon_s\right)^2.$$

16. In order to develop intuition, we abstract from the fact that V_X should be estimated, and we standardize all variances to equal one.

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

As $\left(m^{-1}\sum_{s=1}^{m}X_sX_s'\right)^{-1}\left(m^{-1/2}\sum_{s=1}^{m}X_s\varepsilon_s\right)^2 \Rightarrow \left(\int_0^{1/(1+\pi)}\left[\sigma_X^2+g_X^2(r)\right]dr\right)^{-1}\times\left(\int_0^{1/(1+\pi)}\left[\sigma_X+g_X(r)\right]dB(r)\right)^2$, the component vanishes asymptotically provided $(n^{1/2}/m)\to 0$. ‖

Proof of Proposition 4. As $E\left(\partial L_t(\beta_t)/\partial\beta-\partial\mathcal{L}_t(\beta_t)/\partial\beta\right)=0\ \forall t$, the "parameter instabilities I" component is zero. The "parameter instabilities II" component is $(1/2)n^{-1/2}\sum_{t=m}^{T-\tau}E\left[\left(\beta-\left(\beta+n^{-1/4}\Delta\beta\right)\right)'J\left(\beta-\left(\beta+n^{-1/4}\Delta\beta\right)\right)\right]=(1/2)\Delta\beta'J\Delta\beta$ and the "other instabilities" component is $n^{-1/2}\sum_{t=m}^{T-\tau}E\left(\varepsilon_{t+\tau}^2-\overline{\Sigma}_j\varepsilon_j^2\right)^{-1/2}=n^{-1/2}\sum_{t=m}^{T-\tau}\left(\left(\sigma^2+n^{-1/2}\Delta\sigma^2\right)/(1-\alpha)-\overline{\Sigma}_j\sigma^2/(1-\alpha)\right)=n^{-1/2}\sum_{t=m}^{T-\tau}n^{-1/2}\Delta\sigma^2/(1-\alpha)=\Delta\sigma^2/(1-\alpha)$. As $\partial L_{t+\tau}(\beta_t)/\partial\beta=-2X_{t+\tau}\left(Y_{t+\tau}-X_{t+\tau}'\beta_t\right)$ is uncorrelated with $\left(\widehat{\beta}_t-\beta_t\right)$, the "estimation uncertainty I" component is zero. As $E\left(\partial^2L_j(\beta)/\partial\beta\partial\beta'\right)=E\left(\partial^2\mathcal{L}_j(\beta)/\partial\beta\partial\beta'\right)=2J\ \forall j$, the "estimation uncertainty III" component in (8) is also zero. Finally, the "estimation uncertainty II" component equals

$$n^{1/2}E\left(\widehat{\beta}_m-\beta\right)'\left(2m^{-1}\sum_{s=1}^{m}X_sX_s'\right)\left(\widehat{\beta}_m-\beta\right)=2(n^{1/2}/m)E\left(m^{-1/2}\sum_{s=1}^{m}X_s\varepsilon_s\right)'\left(m^{-1}\sum_{s=1}^{m}X_sX_s'\right)^{-1}\left(m^{-1/2}\sum_{s=1}^{m}X_s\varepsilon_s\right)$$
$$\simeq 2\left(n^{1/2}/m\right)\sigma^2k/(1-\alpha)$$

as $(1-\alpha)\sigma^{-2}\left(m^{-1/2}\sum_{s=1}^{m}X_s\varepsilon_s\right)'\left(m^{-1}\sum_{s=1}^{m}X_sX_s'\right)^{-1}\left(m^{-1/2}\sum_{s=1}^{m}X_s\varepsilon_s\right)\simeq\chi_k^2$ (strictly speaking, the equality holds under a normality assumption on the $\varepsilon's$). ‖

Proof of Proposition 5. We focus on the recursive scheme and, for simplicity, assume that z_t is scalar. Let $\widehat{\delta}_n^*\equiv\begin{pmatrix}1 & -\overline{z}'S_{\widetilde{z}\widetilde{z}}^{-1}\\ 0 & S_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}\begin{pmatrix}\overline{SL}_{m,n}^*\\ n^{-1}\sum_{t=m}^{T-\tau}\widetilde{z}_t(SL_{t+\tau}^*-\overline{SL}_{m,n}^*)\end{pmatrix}$. Given B2 and B3, $\operatorname{var}\left(n^{1/2}\widehat{\delta}_n^*\right)=\begin{pmatrix}1 & -E(z_t)'\Sigma_{\widetilde{z}\widetilde{z}}^{-1}\\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}$ $\operatorname{var}\begin{pmatrix}n^{-1/2}\sum_{t=m}^{T-\tau}\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\\ n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\end{pmatrix}\begin{pmatrix}1 & -E(z_t)'\Sigma_{\widetilde{z}\widetilde{z}}^{-1}\\ 0 & \Sigma_{\widetilde{z}\widetilde{z}}^{-1}\end{pmatrix}'+o_p(1)$. As shown in Corollary 1, the upper diagonal element $\sigma_{m,n}^2$ of $\operatorname{var}\left(n^{1/2}\delta_n^*\right)$ can be consistently estimated under H_0 by $\hat{\sigma}_{m,n}^2$, given in the same corollary. Letting $\widetilde{L}_t\equiv L_t-E(L_t)$, the remaining elements are as follows: (I) $\operatorname{var}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)=\operatorname{var}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*\right)+\operatorname{var}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(t^{-1}\sum_{j=1}^{t}\widetilde{L}_j^*\right)\right)-2\operatorname{cov}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*,\ n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(t^{-1}\sum_{j=1}^{t}\widetilde{L}_j^*\right)\right)$. Each element of the second term goes to zero by arguments similar to those in Lemma A4(a) of West (1994) under Assumption B1. The typical element of the third term is $n^{-1}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\sum_{j=1}^{t}s^{-1}\widetilde{\kappa}_4(j,s,\tau)$, where $\widetilde{\kappa}_4(j,s,\tau)\equiv E\left[\widetilde{z}_{it}\widetilde{L}_{t+\tau}^*\widetilde{L}_j^*\widetilde{z}_{is}\right]$ is the fourth-order cumulant, and τ is fixed. Therefore, $\left|n^{-1}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\sum_{j=1}^{t}s^{-1}\widetilde{\kappa}_4(j,s,\tau)\right|\le n^{-1}\sum_{t=m}^{T-\tau}m^{-1}\sum_{s=m}^{T-\tau}\sum_{j=1}^{t}\left|\widetilde{\kappa}_4(j,s,\tau)\right|\le m^{-1}\sum_{s=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\left|\widetilde{\kappa}_4(j,s,\tau)\right|\underset{p}{\to}0$ using Assumptions A7 and B4. Hence, $\operatorname{var}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)\underset{p}{\to}\sum_{j=-\infty}^{\infty}E\left(\widetilde{z}_t\widetilde{L}_{t+\tau}^*\widetilde{L}_{t+\tau-j}^*\widetilde{z}_{t-j}\right)$. (II) $\operatorname{cov}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right),\ n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)=A_{1n}+A_{2n}-A_{3n}-A_{4n}$, where $A_{1n}\equiv n^{-1}\operatorname{cov}\left(\sum_{t=m}^{T-\tau}\widetilde{L}_{t+\tau}^*,\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*\right)$, $A_{2n}\equiv n^{-1}\operatorname{cov}\left(\sum_{t=m}^{T-\tau}t^{-1}\sum_{j=1}^{t}\widetilde{L}_j^*,\ \sum_{t=m}^{T-\tau}t^{-1}\sum_{j=1}^{t}\widetilde{z}_t\widetilde{L}_j^*\right)$, $A_{3n}\equiv n^{-1}\operatorname{cov}\left(\sum_{t=m}^{T-\tau}t^{-1}\sum_{j=1}^{t}\widetilde{L}_j^*,\sum_{t=m}^{T-\tau}\widetilde{z}_t\widetilde{L}_{t+\tau}^*\right)$, $A_{4n}\equiv n^{-1}\operatorname{cov}\left(\sum_{t=m}^{T-\tau}\widetilde{L}_{t+\tau}^*,\ \sum_{t=m}^{T-\tau}\sum_{j=1}^{t}t^{-1}\sum_{j=1}^{t}\widetilde{z}_t\widetilde{L}_j^*\right)$. Consider each term separately: (i) $A_{1n}=n^{-1}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_{s+\tau}^*\right)\underset{p}{\to}\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)$; (ii) $|A_{2n}|=\left|n^{-1}\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\sum_{s=m}^{T-\tau}\sum_{k=1}^{s}s^{-1}t^{-1}E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\le n^{-1}\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\sum_{s=m}^{T-\tau}\sum_{k=1}^{s}s^{-1}t^{-1}\left|E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\le n^{-1}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}m^{-2}\sum_{j=1}^{t}\sum_{k=1}^{s}\left|E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\le n^{-1}\left(\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}m^{-2}\right)\left(\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\left|E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_k^*\right)\right|\right)\underset{p}{\to}0$ using Assumptions A7 and B4; (iii) $A_{3n}=n^{-1}\sum_{t=m}^{T-\tau}\sum_{j=1}^{t}\sum_{s=m}^{T-\tau}t^{-1}E\left(\widetilde{L}_j^*\widetilde{z}_s'\widetilde{L}_{s+\tau}^*\right)\underset{p}{\to}\left[1-\pi^{-1}\ln(1+\pi)\right]\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)$ from similar reasonings to those in Lemma A6 in West (1996); (iv) Letting $\widetilde{\kappa}_3(j,s)\equiv E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_j^*\right)$, we have $|A_{4n}|=\left|n^{-1}\sum_{t=m}^{T-\tau}\sum_{s=m}^{T-\tau}\sum_{j=1}^{s}s^{-1}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_j^*\right)\right|\le n^{-1}\sum_{s=m}^{T-\tau}s^{-1}\sum_{t=m}^{T-\tau}\sum_{j=1}^{s}\left|E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_s'\widetilde{L}_j^*\right)\right|\le m^{-1}\sum_{s=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\left|\widetilde{\kappa}_3(j,s)\right|\underset{p}{\to}0$ using Assumption B4 and Lemma A1(a) of West (1996). Therefore, $\operatorname{cov}\left(n^{-1/2}\sum_{t=m}^{T-\tau}\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right),n^{-1/2}\sum_{t=m}^{T-\tau}\widetilde{z}_t\left(SL_{t+\tau}^*-E\left(SL_{t+\tau}^*\right)\right)\right)\underset{p}{\to}\sum_{j=-\infty}^{\infty}E\left(\widetilde{L}_{t+\tau}^*\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^*\right)-\left[1-\pi^{-1}\ln(1+\pi)\right]$

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

$\sum_{j=-\infty}^{\infty} E\left(\widetilde{L}_{t+\tau}^{*}\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^{*}\right) = \left[\pi^{-1}\ln(1+\pi)\right]\sum_{j=-\infty}^{\infty} E\left(\widetilde{L}_{t+\tau}^{*}\widetilde{z}_{t-j}'\widetilde{L}_{t+\tau-j}^{*}\right)$. We have therefore shown that

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

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

Proof of Corollary 2. When the losses are conditionally homoskedastic, then $A_{1n} \underset{p}{\to} 0$, and $A_{3n} \underset{p}{\to} 0$ in the proof of Proposition 4, which implies $\Sigma_{\widetilde{z}\widetilde{L}^*,\widetilde{L}^*} = 0$. Thus,

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

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

Acknowledgements. We would like to thank the editor, Bernard Salanié, two anonymous referees, Todd Clark, Frank Diebold, Graham Elliott, Mike McCracken, Ulrich Muller, Chris Otrok, Martin Schwerdtfeger, Allan Timmermann as well as participants to the 2005 Missouri Economics Conference, the 2005 CIRANO-CIREQ Conference on Forecasting in Macroeconomics and Finance, the 2005 JAE Annual Conference, the 2005 Conference on Frontiers in Time Series Econometrics, the 2005 NBER-NSF Time Series Conference, the 2005 SEA Meetings, the 2005 Econometric Society World Congress, the 2005 Triangle Econometrics Conference, the 2005 Bank of England forecasting workshop, and the 2006 SAMSI workshop for useful discussions and suggestions. The paper benefited from comments from seminar participants at UCLA, Duke, Berkeley, UCR, UCSD, Texas-Austin, Pittsburgh and University of Montreal. Support from the NSF is gratefully acknowledged.

REFERENCES

ANDREWS, D. W. K. (1991), "Heteroskedasticity 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.

BRILLINGER, D. R. (1981), *Time Series: Data Analysis and Theory* (San Francisco: Holden Day).

CAMPBELL, J. H. and THOMPSON, S. B. (2005), "Predicting the Equity Premium Out of Sample: Can Anything Beat the Historical Average?" (Mimeo, Harvard University).

CAVALIERE, G. (2004), "Unit Root Tests Under Time-Varying Variances", *Econometric Reviews*, **23** (3), 259–292.

CHRISTOFFERSEN P. F. and DIEBOLD, F. X. (1997), "Optimal Prediction under Asymmetric Loss", *Econometric Theory*, **13**, 808–817.

CLARK, T. and MCCRACKEN, M. (2006), "The Predictive Content of the Output Gap for Inflation: Resolving In-Sample and Out-of-Sample Evidence", *Journal of Money, Credit and Banking*, **38** (5), 1127–1148.

CLARIDA, R., GALI, J. and GERTLER, M. (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory", *Quarterly Journal of Economics*, **115** (1), 147–180.

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

CLEMENTS, M. P. and HENDRY, D. F. (1999), "Some Methodological Implications of Forecast Failure" (Mimeo, Warwick University and Nuffield College).

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

DAVIDSON, H. E. H. (1994), *Stochastic Limit Theory* (Oxford: Oxford University Press).

DUFOUR, J.-M., GHYSELS, E. and HALL, A. (1994), "Generalized Predictive Tests and Structural Change Analysis in Econometrics", *International Economic Review*, **35**, 199–229.

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

© 2009 The Review of Economic Studies Limited

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

ELLIOTT, G. and MULLER U. (2006), "Efficient Tests for General Persistent Time Variation in Regression Coefficients", *Review of Economic Studies*, **73**, 907–940.

FISHER, J. D. M., LIU, C. T. and ZHOU, R. (2002), "When Can We Forecast Inflation?", *Economic Perspectives*, **26** (1), 30–42.

GHYSELS, E. and HALL, A. (1990), "A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator", *International Economic Review*, **31**, 355–364.

GHYSELS, E., SWANSON, N. R. and CALLAN, M. (2002), "Monetary Policy Rules with Model and Data Uncertainty", *Southern Economic Journal*, **69**, 239–265.

GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. E. (1993), "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks", *The Journal of Finance*, **48** (5), 1779–1801.

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

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

KOOP, G. M. and POTTER, S. (2007), "Forecasting and Estimating Multiple Change-point Models with an Unknown Number of Change-points", *Review of Economic Studies*, **74** (3), 763–789.

McCRACKEN, M. (2000), "Robust Out-of-sample Inference", *Journal of Econometrics*, **99**, 195–223.

MINCER J. and ZARNOWITZ, V. (1969), "The Evaluation of Economic Forecasts", in J. Mincer (eds.) *Economic Forecasts and Expectations*, *National Bureau of Economic Research Studies in Business Cycles 19* (New York: Columbia University Press) 3–46.

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

PATTON, A. J. and TIMMERMANN, A. (2007), "Testing Forecast Optimality Under Unknown Loss", *Journal of the American Statistical Association*, **102** (480), 1172–1184 .

PESARAN, H., PETTENUZZO, D. and TIMMERMANN, A. (2006), "Forecasting Time Series Subject to Multiple Structural Breaks", *Review of Economic Studies*, **73** (4), 1057–1084.

ROSSI, B. (2005), "Optimal Tests for Nested Model Selection with Underlying Parameter Instability", *Econometric Theory*, **21** (5), 962–990.

STAIGER, D., STOCK, J. H. and WATSON, M. W. (1997), "The NAIRU, Unemployment and Monetary Policy", *Journal of Economic Perspectives*, **11**, 33–51.

STOCK, J. H. and WATSON, M. W. (1999), "Forecasting Inflation", *Journal of Monetary Economics*, **44**, 293–335.

STOCK, J. H. and WATSON, M. W. (2003), "Forecasting Output and Inflation: The Role of Asset Prices", *Journal of Economic Literature*, **41**, 788–829.

STOCK, J. H. and WATSON, M. W. (2004), "Combination Forecasts of Output Growth in a Seven-Country Data Set", *Journal of Forecasting*, **23** (6), 405–430.

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

WEST, K. and McCRACKEN, M. (1998), "Regression-Based Tests of Predictive Ability", *International Economic Review*, **39** (4), 817–840.

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

WHITE, H. and DOMOWITZ, I. (1984), "Nonlinear Regression with Dependent Observations", *Econometrica*, **52**, 143–162.

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

Downloaded from http://restud.oxfordjournals.org/ at Deutsches Institut fuer Wirtschaftsforschung / Bibliothek on May 4, 2015

© 2009 The Review of Economic Studies Limited