

A near optimal test for structural breaks when forecasting under square error loss

Tom Boot* Andreas Pick†

May 12, 2017

Abstract

We propose a near optimal test for structural breaks of unknown timing when the purpose of the analysis is to obtain accurate forecasts under square error loss. A bias-variance trade-off exists under square forecast error loss, which implies that a structural break of small magnitude should be ignored. We quantify this magnitude, assess the relevance of the break location, and provide a test to determine whether modeling a break improves forecast accuracy. Asymptotic critical values and near optimality properties are established allowing for a break under the null, where the critical break magnitude varies with the break location. The results are extended to a class of shrinkage forecasts with our test statistic as shrinkage constant. Empirical results on a large number of macroeconomic time series show that structural breaks that are relevant for forecasting occur much less frequently than indicated by existing tests.

JEL codes: C12, C53

Keywords: structural break test, forecasting, squared error loss

*University of Groningen, t.boot@rug.nl

†Erasmus University Rotterdam, Tinbergen Institute, De Nederlandsche Bank, and CESifo Institute, andreas.pick@cantab.net.

We thank Graham Elliott, Bart Keijsers, Alex Koning, Robin Lumsdaine, Agnieszka Markiewicz, Michael McCracken, Allan Timmermann, participants of seminars at CESifo Institute, Tinbergen Institute, University of Nottingham, and conference participants at ESEM, IAAE annual conference, NESG meeting, RMSE workshop, and SNDE conference for helpful comments.

1 Introduction

Structural breaks present a major challenge to forecasters as they require information about the time of the break and parameter estimates for the post-break sample. Often, both can be estimated only imprecisely (Elliott and Müller, 2007, 2014). Furthermore, forecasts are typically evaluated using mean square error loss, which implies a bias-variance trade-off. This suggests that ignoring rather than modeling small breaks leads to more accurate forecasts (Pesaran and Timmermann, 2005). If sufficiently small breaks can be ignored, the question is: what constitutes sufficiently small?

In this paper, we develop a test for equal forecast accuracy that compares the expected mean square forecast error (MSFE) of the forecast from a post-break sample to that from the full sample. The difference in MSFE depends on a linear combination of the pre- and post-break parameters with weights that are a function of the regressors in the forecast period. As a result, breaks in the parameter vector, which are the focus of the extant literature on structural breaks (such as Ploberger et al., 1989; Andrews, 1993; Andrews and Ploberger, 1994; Elliott and Müller, 2007, 2014; Elliott et al., 2015), do not necessarily imply a break in the forecast.

The full sample and the post-break sample forecasts achieve equal forecast accuracy at a critical magnitude of the break. Below this magnitude, the full sample forecast is preferred; above this magnitude, the forecast based on the post-break sample is more accurate. When the break date is known, the critical magnitude is one standard deviation of the forecast distribution. In contrast, when the date of a local-to-zero break is unknown, the break date cannot be estimated consistently. This increases the variance of the post-break sample forecast, which, in turn, increases the critical magnitude of the break to up to three standard deviations of the forecast distribution.

A complication is that, due to the bias-variance trade-off, equal forecast accuracy will be achieved under a non-zero magnitude of the break. The null of our test is therefore different from that of existing tests that use a null of no instability. Additionally, the magnitude of the break under the null depends on the unknown break date, which under local breaks is not consistently estimable. However, using the work of Andrews (1993) and Piterbarg (1996), we can show that our test is optimal as the size of the test tends to zero. Additionally, we provide evidence that our test remains close to the optimal test for conventional choices of the nominal size. The reason is that critical magnitudes that follow from the MSFE loss function are relatively large, which result in accurate estimates of the break date. The near optimality does not depend on whether our Wald-statistic is used in its homoskedastic form or whether a heteroskedastic version is used, as long as the estimator of the variance is consistent.

The competing forecasts in our test are from the full sample and from the post-break sample. Forecasts based on any test, including ours, will be

based on pre-test estimators, which, in some areas of the parameter space, can have high risk compared to shrinkage estimators. Pesaran et al. (2013) show that forecasts based on post-break samples can be improved by using all observations and weighting them such that the MSFE is minimized. We show that this forecast can be written as a forecast from a shrinkage estimator in the tradition of Thompson (1968), where the shrinkage estimator averages between the full sample estimator and post-break sample estimator with a weight that is equivalent to the test statistic introduced in this paper. This approach is similar to the averaging of estimators proposed by Hansen (2009), which minimized the in-sample mean square error using weights based on the Mallows criterion.

Under a known break date, the performance of shrinkage estimators is well known, see for example Magnus (2002). However, their properties depend critically on the fact that the break date is known, which implies that the estimator from the post-break sample is unbiased. Under a local break, this may not be the case and the forecasting performance of the shrinkage estimator compared to the full sample forecast is not immediately clear. Since the shrinkage estimator does not take break date uncertainty into account, it will likely put too much weight on the post-break sample forecast. We find that for small break magnitudes, where the break date is not accurately identified, the shrinkage forecast is less accurate than the full sample forecast. However, compared to the post-break sample forecast, we find that the shrinkage estimator is more accurate for a large area in the parameter space. We therefore propose a second version of our test that compares the forecast accuracy of the shrinkage estimator and the full sample forecast.

More generally, we propose a testing framework that incorporates the loss function, here the mean square forecast error, into the test. Similar to the work of Trenkler and Toutenburg (1992) and Clark and McCracken (2012), our test is inspired by the in-sample MSE test of Toro-Vizcarrondo and Wallace (1968) and Wallace (1972). However, compared to the tests of Trenkler and Toutenburg (1992) and Clark and McCracken (2012), our testing framework is much simpler in that, under a known break date, our test statistic has a known distribution that is free of nuisance parameters.

Our test shares some similarity with the work of Dette and Wied (2016), who consider CUSUM tests in the spirit of Brown et al. (1975) but allow for a constant parameter differences under the null. They do, however, not consider local-to-zero breaks, which would eliminate break date uncertainty in our asymptotic framework. Also, we show that the critical magnitude of the break depends on the break date and is therefore not identical across samples.

Forecast accuracy tests of the kind suggested by Diebold and Mariano (1995) and Clark and McCracken (2001) (see Clark and McCracken (2013) for a review) are substantially different from our test. These tests assess

forecast accuracy *ex post*. In contrast, the test we propose in this paper is an *ex ante* test of the accuracy of forecasts of models that do or do not account for breaks.

Giacomini and Rossi (2009) assess forecast breakdowns by comparing the in-sample fit and out-of-sample forecasts of a given model. The main focus of their work is on assessing pseudo-out-of-sample forecasts. However, they also consider forecasting the loss differential of in-sample and out-of-sample forecast performance by modeling it with additional regressors. This contrasts with our approach, which targets the out-of-sample period directly in the construction of the test statistic. Of interest for our work is that, while a structural break is only one possible source of forecast breakdowns, Giacomini and Rossi find that it is a major contributor to forecast breakdowns in predicting US inflation using the Phillips curve.

Substantial evidence for structural breaks has also been found in macroeconomic and financial time series by, for example, Stock and Watson (1996), Rapach and Wohar (2006), Rossi (2006), and Paye and Timmermann (2006). We apply our test to macroeconomic and financial time series in the FRED-MD data set of McCracken and Ng (2016). We find that breaks that are important for forecasting under MSFE loss are between a factor two to three less frequent than the sup-Wald test by Andrews (1993) would indicate. Incorporating only the breaks suggested by our test substantially reduces the average MSFE in this data set compared to the forecasts that take all breaks suggested by Andrews' sup-Wald test into account. Our paper, therefore, provides theoretical support for the finding of Stock and Watson (1996) that breaks do not appear to have a substantial effect on forecast accuracy even though they are a prominent feature of macroeconomic data.

The paper is structured as follows. In Section 2, we start with a motivating example using the linear regression model with a break of known timing. The model is generalized in Section 3 using the framework of Andrews (1993). In Section 4, we derive the test, show its near optimality, and extend the test to cover the optimal weights or shrinkage forecast. Simulation results in Section 5 shows that the near optimality of the test is in fact quite strong, with power very close to the optimal, but infeasible, test conditional on the true break date. Finally, the application of our tests to the large set of time series in the FRED-MD data set is presented in Section 6.

2 Motivating example: a structural break of known timing in a linear model

In order to gain intuition, initially consider a linear regression model with a structural break at time T_b

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + \varepsilon_t, \quad \varepsilon_t \sim \text{iid}(0, \sigma^2) \quad (1)$$

where

$$\boldsymbol{\beta}_t = \begin{cases} \boldsymbol{\beta}_1 & \text{if } t \leq T_b \\ \boldsymbol{\beta}_2 & \text{if } t > T_b \end{cases}$$

\mathbf{x}_t is a $k \times 1$ vector of exogenous regressors, $\boldsymbol{\beta}_i$ a $k \times 1$ vector of parameters, and the break date, T_b , is initially assumed to be known. The parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ can be estimated by OLS on the two subsamples. If the break is ignored, a single vector of parameter estimates, $\hat{\boldsymbol{\beta}}_F$, can be obtained using OLS on the full sample.

Denote $\mathbf{V}_i = (T_i - T_{i-1})\text{Var}(\hat{\boldsymbol{\beta}}_i)$, for $i = 1, 2$, $T_0 = 0$, $T_1 = T_b$, $T_2 = T$ and $\mathbf{V}_F = T\text{Var}(\hat{\boldsymbol{\beta}}_F)$ as the covariance matrices of the vectors of coefficient estimates. Initially, assume these matrices to be known; later they will be replaced by their probability limits.

In this paper, we would like to test whether the expected mean squared forecast error (MSFE) from the h -step ahead forecast using the full sample, $\hat{y}_{T+h}^F = \mathbf{x}'_{T+h}\hat{\boldsymbol{\beta}}_F$, is smaller or equal to that of the post-break sample, $\hat{y}_{T+h}^P = \mathbf{x}'_{T+h}\hat{\boldsymbol{\beta}}_2$. In this motivating example, we consider $h = 1$, and extend the results towards the more general case in Section 4.

The MSFE for the forecast from the post-break sample estimate, $\hat{\boldsymbol{\beta}}_2$, is

$$\begin{aligned} \text{MSFE}(\mathbf{x}'_{T+1}\hat{\boldsymbol{\beta}}_2) &= \text{E} \left[\left(\mathbf{x}'_{T+1}\hat{\boldsymbol{\beta}}_2 - \mathbf{x}'_{T+1}\boldsymbol{\beta}_2 - \varepsilon_{T+1} \right)^2 \right] \\ &= \frac{1}{T - T_b} \mathbf{x}'_{T+1} \mathbf{V}_2 \mathbf{x}_{T+1} + \sigma^2 \end{aligned} \quad (2)$$

and that using the full sample estimate, $\hat{\boldsymbol{\beta}}_F$, is

$$\begin{aligned} \text{MSFE}(\mathbf{x}'_{T+1}\hat{\boldsymbol{\beta}}_F) &= \text{E} \left[\left(\mathbf{x}'_{T+1}\hat{\boldsymbol{\beta}}_F - \mathbf{x}'_{T+1}\boldsymbol{\beta}_2 - \varepsilon_{T+1} \right)^2 \right] \\ &= \text{E} \left[\left(\mathbf{x}'_{T+1}\hat{\boldsymbol{\beta}}_F - \mathbf{x}'_{T+1}\boldsymbol{\beta}_2 \right)^2 \right] + \frac{1}{T} \mathbf{x}'_{T+1} \mathbf{V}_F \mathbf{x}_{T+1} + \sigma^2 \\ &= \left[\frac{T_b}{T} \mathbf{x}'_{T+1} \mathbf{V}_F \mathbf{V}_1^{-1} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \right]^2 + \frac{1}{T} \mathbf{x}'_{T+1} \mathbf{V}_F \mathbf{x}_{T+1} + \sigma^2 \end{aligned} \quad (3)$$

Comparing (2) and (3), we see that the full sample forecast is at least as accurate as the post-break sample forecast if

$$\begin{aligned} \zeta &= T\tau_b^2 \frac{[\mathbf{x}'_{T+1} \mathbf{V}_F \mathbf{V}_1^{-1} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)]^2}{\mathbf{x}'_{T+1} \left(\frac{\mathbf{V}_2}{1 - \tau_b} - \mathbf{V}_F \right) \mathbf{x}_{T+1}} \\ &\xrightarrow{p} T\tau_b(1 - \tau_b) \frac{[\mathbf{x}'_{T+1} (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)]^2}{\mathbf{x}'_{T+1} \mathbf{V} \mathbf{x}_{T+1}} \\ &\leq 1 \end{aligned} \quad (4)$$

where $\tau_b = T_b/T$ and the second line assumes that the covariance matrices asymptotically satisfy $\text{plim}_{T \rightarrow \infty} \mathbf{V}_i = \mathbf{V}$ for $i = 1, 2, F$.

To test $H_0 : \zeta = 1$ note that

$$\begin{aligned} \hat{\zeta}(\tau_b) &= T\tau_b^2 \frac{\left[\mathbf{x}'_{T+1} \mathbf{V}_F \mathbf{V}_1^{-1} (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) \right]^2}{\mathbf{x}'_{T+1} \left(\frac{\mathbf{V}_2}{1-\tau} - \mathbf{V}_F \right) \mathbf{x}_{T+1}} \\ &= \frac{\left[\mathbf{x}'_{T+1} (\hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_2) \right]^2}{\mathbf{x}'_{T+1} \text{Var}(\hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_2) \mathbf{x}_{T+1}} \xrightarrow{d} \chi^2(1, \zeta) \end{aligned} \quad (5)$$

Furthermore, given that we are interested in the null of $\zeta = 1$, the test statistic has a $\chi^2(1, 1)$ -distribution under the null, which is free of nuisance parameters.

A more conventional and asymptotically equivalent form of the test statistic is

$$\hat{\zeta}(\tau_b) = T \frac{\left[\mathbf{x}'_{T+1} (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) \right]^2}{\mathbf{x}'_{T+1} \left(\frac{\mathbf{V}_1}{\tau_b} + \frac{\mathbf{V}_2}{1-\tau_b} \right) \mathbf{x}_{T+1}} \xrightarrow{d} \chi^2(1, \zeta) \quad (6)$$

This is a standard Wald test using the regressors at $t = T + 1$ as weights.

The results of the test will, in general, differ from the outcomes of the classical Wald test on the difference between the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ for two reasons. The first is that the multiplication by \mathbf{x}_{T+1} can render large breaks irrelevant. Alternatively, it can increase the importance of small breaks in the coefficient vector for forecasting. The second reason is that under $H_0 : \zeta = 1$, we compare the test statistic against the critical values of the non-central χ^2 -distribution, instead of the central χ^2 -distribution. The critical values of these distributions differ substantially: the $\alpha = 0.05$ critical value of the $\chi^2(1)$ is 3.84 and that of the $\chi^2(1, 1)$ is 7.00.

As is clear from (4), if the difference in the parameters, $\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2$, converges to zero at a rate $T^{-1/2+\epsilon}$ for some $\epsilon > 0$, then the test statistic diverges to infinity as $T \rightarrow \infty$, which is unlikely to reflect the uncertainty surrounding the break date in empirical applications. In the remainder of the paper, we will therefore consider breaks that are local in nature, i.e. $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_1 + \frac{1}{\sqrt{T}} \boldsymbol{\eta}$, rendering a finite test statistic in the asymptotic limit. Local breaks have been intensively studied in the recent literature, see for example Elliott and Müller (2007, 2014) and Elliott et al. (2015). An implication of local breaks is that no consistent estimator for the break date is available. A consequence is that post-break parameters cannot be consistently estimated. This will deteriorate the accuracy of the post-break window forecast compared the full sample forecast, which, in turn, increases the break magnitude that yields equal forecasting performance between full and post-break sample estimation windows.

3 Model and estimation

We consider a possibly non-linear, parametric model, where parameters are estimated using the generalized method of moments. The general estimation framework is that of Andrews (1993). The observed data are given by a triangular array of random variables $\{\mathbf{W}_t = (\mathbf{Y}_t, \mathbf{X}_t) : 1 \leq t \leq T\}$, $\mathbf{Y}_t = (y_1, y_2, \dots, y_t)$, and $\mathbf{X}_t = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t)'$. Assumptions can be made with regard to the dependence structure of \mathbf{W}_t such that the results below apply to a range of time series models. We make the following additional assumption on the noise and the relation between y_t , lagged values of y_t and exogenous regressors \mathbf{x}_t .

Assumption 1 *The model for the dependent variable y_t consists of a signal and additive noise*

$$y_t = f_t(\boldsymbol{\beta}_t, \boldsymbol{\delta}; \mathbf{X}_t, \mathbf{Y}_{t-1}) + \varepsilon_t \quad (7)$$

where the function f_t is fixed and differentiable with respect to the parameter vector $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t', \boldsymbol{\delta}')'$.

In (7), while the parameter vector $\boldsymbol{\delta}$ is constant for all t , the parameter vector $\boldsymbol{\beta}_t$ could be subject to a structural break. When ignoring the break, parameters are estimated by minimizing the sample analogue of the population moment conditions

$$\frac{1}{T} \sum_{t=1}^T \mathbb{E}[m(\mathbf{W}_t, \boldsymbol{\beta}, \boldsymbol{\delta})] = 0$$

which requires solving

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T m(\mathbf{W}_t, \hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}})' \hat{\boldsymbol{\gamma}} \frac{1}{T} \sum_{t=1}^T m(\mathbf{W}_t, \hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) = \\ & \inf_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}} \frac{1}{T} \sum_{t=1}^T m(\mathbf{W}_t, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}})' \hat{\boldsymbol{\gamma}} \frac{1}{T} \sum_{t=1}^T m(\mathbf{W}_t, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}) \end{aligned} \quad (8)$$

where $\hat{\boldsymbol{\beta}}_F$ is estimator based on the full estimation window. Throughout we set the weighting matrix $\boldsymbol{\gamma} = \mathbf{S}^{-1}$ and

$$\mathbf{S} = \lim_{T \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T m(\mathbf{W}_t, \boldsymbol{\beta}, \boldsymbol{\delta}) \right)$$

for which a consistent estimator is assumed to be available.

As discussed above, we consider a null hypothesis that allows for local breaks,

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_1 + \frac{1}{\sqrt{T}} \boldsymbol{\eta}(\tau)$$

where $\boldsymbol{\eta}(\tau) = \mathbf{b}\mathbf{I}[\tau < \tau_b]$, $\mathbf{I}(A)$ is the indicator function, which is unity if A is true and zero otherwise, \mathbf{b} is a vector of constants, and $\tau = t/T$. The partial sample parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ satisfy the partial sample moment conditions

$$\frac{1}{\tau T} \sum_{t=1}^{\tau T} m(\mathbf{W}_t, \boldsymbol{\beta}_1, \boldsymbol{\delta}) = \mathbf{0}, \quad \text{and} \quad \frac{1}{(1-\tau)T} \sum_{t=T\tau+1}^T m(\mathbf{W}_t, \boldsymbol{\beta}_2, \boldsymbol{\delta}) = \mathbf{0}$$

Define

$$\bar{m}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\delta}, \tau) = \frac{1}{\tau T} \sum_{t=1}^{\tau T} \begin{pmatrix} m(\mathbf{W}_t, \boldsymbol{\beta}_1, \boldsymbol{\delta}) \\ \mathbf{0} \end{pmatrix} + \frac{1}{(1-\tau)T} \sum_{t=T\tau+1}^T \begin{pmatrix} \mathbf{0} \\ m(\mathbf{W}_t, \boldsymbol{\beta}_2, \boldsymbol{\delta}) \end{pmatrix}$$

Then, the partial sum GMM estimators can be obtained by solving (8) with $m(\cdot)$ replaced by $\bar{m}(\cdot)$ and $\hat{\boldsymbol{\gamma}}$ replaced by

$$\hat{\boldsymbol{\gamma}}(\tau) = \begin{pmatrix} \frac{1}{\tau} \hat{\mathbf{S}}^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{1-\tau} \hat{\mathbf{S}}^{-1} \end{pmatrix}$$

The aim is to determine whether the full sample estimator lead to a more precise h -step-ahead forecast in the mean square forecast error sense than the post-break sample estimator. The forecasts are constructed as

$$\begin{aligned} \hat{y}_{T+h}^F &= f_{T+h}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) \\ \hat{y}_{T+h}^P &= f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) \end{aligned} \tag{9}$$

where \mathcal{I}_T is the information set at time T and it should be noted that it includes the exogenous and lagged dependent variables that are needed to construct the forecast. If $h > 1$, the forecasts can be iterated or direct forecasts and the function f_{T+h} will depend on which type of forecast is chosen. In our analysis, the function f_{T+h} can be a non-linear function of the parameters, which allows for iterated forecasts, and we therefore do not add notation to distinguish the types of forecasts. A direct forecast, in contrast, leads to residual autocorrelation, which is allowed in our analytical framework. The comparison between \hat{y}_{T+h}^F and \hat{y}_{T+h}^P is, however, non-standard as, under a local break, even the parameter estimates of the model that incorporates the break may not be unbiased.

In order to compare the forecasts in (9), we start by providing the asymptotic properties of the estimators in a model that incorporates the break and in a model that ignores the break. The asymptotic distributions derived by Andrews (1993) depend on the following matrices, for which consistent estimators are assumed to be available,

$$\mathbf{M} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E} \left[\frac{\partial m(\mathbf{W}_t, \boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\beta}} \right], \quad \mathbf{M}_\delta = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbf{E} \left[\frac{\partial m(\mathbf{W}_t, \boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\delta}} \right]$$

To simplify the notation, define

$$\begin{aligned}\bar{\mathbf{X}}' &= \mathbf{M}'\mathbf{S}^{-1/2} \\ \bar{\mathbf{Z}}' &= \mathbf{M}'_{\delta}\mathbf{S}^{-1/2}\end{aligned}$$

Partial sample estimator The partial sample estimators converge to the following Gaussian process indexed by τ

$$\begin{aligned}\sqrt{T} \begin{pmatrix} \hat{\beta}_1(\tau) - \beta_2 \\ \hat{\beta}_2(\tau) - \beta_2 \\ \hat{\delta} - \delta \end{pmatrix} &\Rightarrow \begin{bmatrix} \tau\bar{\mathbf{X}}'\bar{\mathbf{X}} & \mathbf{0} & \tau\bar{\mathbf{X}}'\bar{\mathbf{Z}} \\ \mathbf{0} & (1-\tau)\bar{\mathbf{X}}'\bar{\mathbf{X}} & (1-\tau)\bar{\mathbf{X}}'\bar{\mathbf{Z}} \\ \tau\bar{\mathbf{Z}}'\bar{\mathbf{X}} & (1-\tau)\bar{\mathbf{Z}}'\bar{\mathbf{X}} & \bar{\mathbf{Z}}'\bar{\mathbf{Z}} \end{bmatrix}^{-1} \\ &\times \begin{bmatrix} \bar{\mathbf{X}}'\mathbf{B}(\tau) + \bar{\mathbf{X}}'\bar{\mathbf{X}} \int_0^\tau \boldsymbol{\eta}(s)ds \\ \bar{\mathbf{X}}'[\mathbf{B}(1) - \mathbf{B}(\tau)] + \bar{\mathbf{X}}'\bar{\mathbf{X}} \int_\tau^1 \boldsymbol{\eta}(s)ds \\ \bar{\mathbf{Z}}'\mathbf{B}(1) + \bar{\mathbf{Z}}'\bar{\mathbf{X}} \int_0^1 \boldsymbol{\eta}(s)ds \end{bmatrix} \end{aligned} \quad (10)$$

where $\mathbf{B}(\tau)$ is a Brownian motion defined on the interval $[0, 1]$ and \Rightarrow denotes weak convergence. In line with Andrews (1993), we subtract β_2 from both estimators $\hat{\beta}_1$ and $\hat{\beta}_2$. This lines up with our interest in forecasting future observations, which are functions of β_2 only, and the remainder that arises if $\tau \neq \tau_b$, is absorbed in the integral on the right hand side.

Define the projection matrix $\mathbf{P}_{\bar{\mathbf{X}}} = \bar{\mathbf{X}}(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'$, its orthogonal complement as $\mathbf{M}_{\bar{\mathbf{X}}} = \mathbf{I} - \mathbf{P}_{\bar{\mathbf{X}}}$ and

$$\begin{aligned}\mathbf{V} &= (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1} \\ \mathbf{H} &= \bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\bar{\mathbf{Z}} \\ \mathbf{L} &= (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\bar{\mathbf{Z}})^{-1} \\ \tilde{\mathbf{H}} &= \mathbf{L}\mathbf{H}\mathbf{L}'\end{aligned} \quad (11)$$

The inverse in (10) yields the asymptotic variance covariance matrix of $(\hat{\beta}_1(\tau)', \hat{\beta}_2(\tau)', \hat{\delta}')'$

$$\boldsymbol{\Sigma}_P = \begin{pmatrix} \frac{1}{\tau}\mathbf{V} + \tilde{\mathbf{H}} & \tilde{\mathbf{H}} & -\mathbf{L} \\ \tilde{\mathbf{H}} & \frac{1}{1-\tau}\mathbf{V} + \tilde{\mathbf{H}} & -\mathbf{L} \\ -\mathbf{L}' & -\mathbf{L}' & \mathbf{H}^{-1} \end{pmatrix}$$

Hence,

$$\begin{aligned}\sqrt{T}(\hat{\beta}_1(\tau) - \beta_2) &\Rightarrow \frac{1}{\tau} \left[(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\mathbf{B}(\tau) + \int_0^\tau \boldsymbol{\eta}(s)ds \right] \\ &\quad - (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\mathbf{B}(1) \\ \sqrt{T}(\hat{\beta}_2(\tau) - \beta_2) &\Rightarrow \frac{1}{1-\tau} \left[(\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'(\mathbf{B}(1) - \mathbf{B}(\tau)) + \int_\tau^1 \boldsymbol{\eta}(s)ds \right] \\ &\quad - (\bar{\mathbf{X}}'\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\bar{\mathbf{Z}}(\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\mathbf{B}(1) \\ \sqrt{T}(\hat{\delta} - \delta) &\Rightarrow (\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\bar{\mathbf{Z}})^{-1}\bar{\mathbf{Z}}'\mathbf{M}_{\bar{\mathbf{X}}}\mathbf{B}(1)\end{aligned} \quad (12)$$

where the convergence occurs jointly. Several terms can be recognized to be analogous to what would be obtained in a multivariate regression problem using the Frisch-Waugh-Lovell theorem.

Full sample estimator For estimators that ignore the break, we have

$$\sqrt{T} \begin{pmatrix} \hat{\beta}_F - \beta_2 \\ \hat{\delta} - \delta \end{pmatrix} \Rightarrow \begin{bmatrix} \bar{X}'\bar{X} & \bar{X}'\bar{Z} \\ \bar{Z}'\bar{X} & \bar{Z}'\bar{Z} \end{bmatrix}^{-1} \begin{bmatrix} \bar{X}'\mathbf{B}(1) + \bar{X}'\bar{X} \int_0^1 \eta(s)ds \\ \bar{Z}'\mathbf{B}(1) + \bar{Z}'\bar{X} \int_0^1 \eta(s)ds \end{bmatrix} \quad (13)$$

Using the notation defined in (11), the inverse in (13) can be written as

$$\Sigma_F = \begin{pmatrix} \mathbf{V} + \tilde{\mathbf{H}} & -\mathbf{L} \\ -\mathbf{L}' & \mathbf{H}^{-1} \end{pmatrix}$$

and, therefore,

$$\begin{aligned} \sqrt{T} \begin{pmatrix} \hat{\beta}_F - \beta_2 \\ \hat{\delta} - \delta \end{pmatrix} &\Rightarrow (\bar{X}'\bar{X})^{-1} \bar{X}'\mathbf{B}(1) + \int_0^1 \eta(s)ds \\ &\quad - (\bar{X}'\bar{X})^{-1} \bar{X}'\bar{Z}(\bar{Z}'\mathbf{M}_{\bar{X}}\bar{Z})^{-1} \bar{Z}'\mathbf{M}_{\bar{X}}\mathbf{B}(1) \\ \sqrt{T} (\hat{\delta} - \delta) &\Rightarrow (\bar{Z}'\mathbf{M}_{\bar{X}}\bar{Z})^{-1} \bar{Z}'\mathbf{M}_{\bar{X}}\mathbf{B}(1) \end{aligned} \quad (14)$$

Note that for the parameters $\hat{\delta}$, the expression is identical to partial sample estimator.

Later results require the asymptotic covariance between the estimators from the full sample and the break model, which is

$$\text{plim}_{T \rightarrow \infty} T \text{Cov}(\hat{\beta}_2(\tau), \hat{\beta}_F) = \mathbf{V} + \tilde{\mathbf{H}} = \text{plim}_{T \rightarrow \infty} T \text{Var}(\hat{\beta}_F)$$

which corresponds to the results by Hausman (1978) that under the null of no misspecification, a consistent and asymptotically efficient estimator should have zero covariance with its difference from an consistent but asymptotically inefficient estimator, i.e. $\text{plim}_{T \rightarrow \infty} T \text{Cov}(\hat{\beta}_F, \hat{\beta}_F - \hat{\beta}_2(\tau)) = \mathbf{0}$. A difference to the case considered here is that, under a local structural break, $\hat{\beta}_F$ and $\hat{\beta}_2(\tau)$ may both be inconsistent.

4 Testing for a structural break

In this section, we apply the estimation framework in the previous section to generalize the motivating example discussed in Section 2. We briefly consider the case of a known break date and then proceed to the case of an unknown break date. A complication in the testing procedure arises when mapping the null hypothesis of equal predictive accuracy to one based on the break magnitude because the latter varies with the unknown break date. Nevertheless, a test which has correct size and near optimal power can be established.

4.1 A local break of known timing

Conditional on the information set \mathcal{I}_T , which contains the regressor set necessary to construct the forecast, the h -step-ahead forecast is

$$\hat{y}_{T+h} = f_{T+h}(\hat{\beta}_2, \hat{\delta} | \mathcal{I}_T)$$

Denote the derivative of f_{T+h} with respect to a parameter vector θ as \mathbf{f}_θ , where we drop the time subscript of the derivative for notational convenience. Equal predictive accuracy is obtained when the break magnitude satisfies

$$\zeta = T(1 - \tau_b)\tau_b \frac{[\mathbf{f}'_{\beta_2}(\beta_1 - \beta_2)]^2}{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}} = 1 \quad (15)$$

Details of the derivation can be found in Appendix A.1. As in the motivating example of Section 2, the null hypothesis of equal mean squared forecast error maps into a hypothesis on the standardized break magnitude, ζ .

A test for $H_0 : \zeta = 1$ can be derived by noting that, asymptotically, $T\text{Var}(\hat{\beta}_1 - \hat{\beta}_2) \xrightarrow{p} \frac{1}{\tau_b(1-\tau_b)} \mathbf{V}$ and, therefore,

$$\hat{\zeta} = T(1 - \tau_b)\tau_b \frac{[\mathbf{f}'_{\beta_2}(\hat{\beta}_1 - \hat{\beta}_2)]^2}{\hat{\omega}} \xrightarrow{d} \chi^2(1, \zeta) \quad (16)$$

where $\hat{\omega}$ is any consistent estimator of $\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}$. The test statistic, $\hat{\zeta}$, can be compared against the critical values of the $\chi^2(1, 1)$ distribution to test for equal forecast performance.

4.2 A local break of unknown timing

The preceding section motivates the use of the Wald-type test statistic (16) to test for equal predictive accuracy between a full-sample and post-break forecast. In this section, we adjust the test statistic for the fact that the break date is unknown, provide its asymptotic distribution based on the results of Section 3 and, subsequently, show that this can be used to test the null of equal predictive accuracy.

When the break date is unknown, we consider the following test statistic

$$\sup_{\tau \in \mathcal{I}} \hat{\zeta}(\tau) = \sup_{\tau \in \mathcal{I}} \left\{ T(1 - \tau)\tau \frac{[\mathbf{f}'_{\beta_2}(\hat{\beta}_1(\tau) - \hat{\beta}_2(\tau))]^2}{\hat{\omega}} \right\} \quad (17)$$

with $\mathcal{I} = [\tau_{\min}, \tau_{\max}]$. Since the function \mathbf{f}'_{β_2} in (17) is fixed, the results in Andrews (1993) and the continuous mapping theorem show that, under

local alternatives and as $T \rightarrow \infty$, $\hat{\zeta}(\tau)$ in (17) weakly converges to

$$\begin{aligned} Q^*(\tau) &= \left[\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}} + \sqrt{\frac{1-\tau}{\tau}} \int_0^\tau \eta(s) ds - \sqrt{\frac{\tau}{1-\tau}} \int_\tau^1 \eta(s) ds \right]^2 \\ &= [Z(\tau) + \mu(\tau; \theta_{\tau_b})]^2 \end{aligned} \quad (18)$$

where $Z(\tau) = \frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}}$ is a self-normalized Brownian bridge with expectation zero and variance equal to one, and

$$\mu(\tau; \theta_{\tau_b}) = \theta_{\tau_b} \left[\sqrt{\frac{1-\tau}{\tau}} \tau_b \mathbf{I}(\tau_b < \tau) + \sqrt{\frac{\tau}{1-\tau}} (1 - \tau_b) \mathbf{I}(\tau_b \geq \tau) \right] \quad (19)$$

arises when a structural break is present. For a fixed break date, $Q^*(\tau)$ follows a non-central χ^2 -distribution with one degree of freedom and non-centrality parameter $\mu(\tau; \theta_{\tau_b})^2$.

Throughout, we use the following estimate of the break date

$$\hat{\tau} = \arg \sup_{\tau \in \mathcal{I}} \hat{\zeta}(\tau) \xrightarrow{d} \arg \sup_{\tau \in \mathcal{I}} Q^*(\tau) \quad (20)$$

4.2.1 MSFE under an unknown break date

The difference between the expected asymptotic MSFE of the partial sample forecast and that of the full sample forecast, standardized by the variance of the post-break forecast based on the true break date, is denoted by

$$\Delta(\tau_b) = \lim_{T \rightarrow \infty} \left\{ \text{MSFE}(\hat{\beta}_2(\hat{\tau}), \hat{\delta}) - \text{MSFE}(\hat{\beta}_F, \hat{\delta}) \right\} / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}$$

where $\text{MSFE}(\hat{\theta})$ is the asymptotic MSFE under parameter estimates $\hat{\theta}$.

Lemma 1 *If the break date is estimated using (20) then difference in the standardized mean squared forecast error, $\Delta(\tau)$, is*

$$\begin{aligned} \Delta(\tau_b) &= \lim_{T \rightarrow \infty} T \left(\mathbf{E} \left\{ \left[f_{T+h}(\hat{\beta}_2(\hat{\tau}), \hat{\delta} | \mathcal{I}_T) - f_{T+h}(\beta_2, \delta | \mathcal{I}_T) \right]^2 \right\} \right. \\ &\quad \left. - \mathbf{E} \left\{ \left[f_{T+h}(\hat{\beta}_F, \hat{\delta} | \mathcal{I}_T) - f_{T+h}(\beta_2, \delta | \mathcal{I}_T) \right]^2 \right\} \right) / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \quad (21) \\ &= \lim_{T \rightarrow \infty} T \left(\mathbf{E} \left\{ \left[\mathbf{f}'_{\beta_2}(\hat{\beta}_2(\hat{\tau}) - \beta_2) \right]^2 \right\} - \mathbf{E} \left\{ \left[\mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2) \right]^2 \right\} \right) / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \end{aligned}$$

The proof is provided in Appendix A.2. Lemma 1 shows that the difference in the MSFE is not affected by the estimation of the parameter vector δ , which is constant over the sample. Note that if instead of estimating the break date, one considers a fixed value τ , then Lemma 1 holds with $\hat{\tau}$ replaced by the fixed value τ . The difference in the standardized mean squared forecast error is then a function of both τ_b and τ .

Using (12) and (14) we rewrite (21) as

$$\Delta(\tau_b) = \mathbb{E} \left\{ \left[\frac{1}{1-\hat{\tau}} \frac{\mathbf{f}'_{\beta_2} \mathbf{V} \bar{\mathbf{X}}' (\mathbf{B}(1) - \mathbf{B}(\hat{\tau}))}{\sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}}} + \frac{1}{1-\hat{\tau}} \int_{\hat{\tau}}^1 \frac{f'_{\beta_2} \boldsymbol{\eta}(s)}{\sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}}} ds \right]^2 - \left(\int_0^1 \frac{f'_{\beta_2} \boldsymbol{\eta}(s)}{\sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}}} ds \right)^2 - 1 \right\} \quad (22)$$

This does not require assumptions with regard to the form of the instability, which is governed by $\boldsymbol{\eta}(\tau)$. Define $J(\tau) = \int_{\tau}^1 (\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2})^{-1/2} \mathbf{f}'_{\beta_2} \boldsymbol{\eta}(s) ds$ and note that, for fixed \mathbf{f}'_{β_2} , $(\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2})^{-1/2} \mathbf{f}'_{\beta_2} \mathbf{V} \bar{\mathbf{X}}' [\mathbf{B}(1) - \mathbf{B}(\hat{\tau})] = B(1) - B(\hat{\tau})$, where $B(\cdot)$ is a one-dimensional Brownian motion. Then

$$\Delta(\tau_b) = \mathbb{E} \left\{ \left[\frac{1}{1-\hat{\tau}} (B(1) - B(\hat{\tau})) + \frac{1}{1-\hat{\tau}} J(\hat{\tau}) \right]^2 \right\} - J(1)^2 - 1$$

which could be used to test whether the use of a partial sample will improve forecast accuracy compared to the full sample under various forms of parameter instability. The expectation can be evaluated analytically if the size of the partial sample is exogenously set to some fraction of the total number of observations.

Under a structural break, $\boldsymbol{\eta}(\tau) = \mathbf{b} \mathbf{I}[\tau < \tau_b]$, where $\mathbf{b} = \sqrt{T}(\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1)$, and (22) becomes

$$\Delta(\tau_b) = \mathbb{E} \left\{ \left[\frac{1}{1-\hat{\tau}} (B(1) - B(\hat{\tau})) + \theta_{\tau_b} \frac{\tau_b - \hat{\tau}}{1-\hat{\tau}} \mathbf{I}[\hat{\tau} < \tau_b] \right]^2 \right\} - \theta_{\tau_b}^2 \tau_b^2 - 1 \quad (23)$$

where $\theta_{\tau_b} = \frac{\mathbf{f}'_{\beta_2} \mathbf{b}}{\sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}}} = \sqrt{\zeta / (\tau_b(1 - \tau_b))}$.

If the break date is estimated without error, then $\hat{\tau} = \tau_b$ and the critical break magnitude of the previous section is obtained. If τ_b is estimated, then the expectation in (23) has to be taken with respect to both the stochastic process $B(\cdot)$ and the distribution of the estimate $\hat{\tau}$.

The distribution of $\hat{\tau}$ is not analytically tractable and we evaluate (23) for different values of τ_b and θ_{τ_b} via simulation. Since $\Delta(\tau_b) > 0$ for $\theta_{\tau_b} = 0$, and $\Delta(\tau_b) < 0$ when $|\theta_{\tau_b}| \rightarrow \infty$, there is a value of $|\theta_{\tau_b}|$ —and thus for ζ_{τ_b} —for which $\Delta(\tau_b) = 0$ for each τ_b . We find numerically that $\Delta(\tau_b)$ is a monotonically decreasing function of θ_{τ_b} , and that hence, the value of θ_{τ_b} for which $\Delta(\tau_b) = 0$ is unique. We report these results in Appendix A.8. This supports the use of (17) to test $\Delta(\tau_b) = 0$.

The break magnitude θ_{τ_b} that yields $\Delta(\tau_b) = 0$ depends on the unknown break date, τ_b . This implies that critical values $u = u(\tau_b)$ will differ across different values of the unknown break date. However, as we will show, our testing framework remains valid when the critical value $u(\tau_b)$ is replaced with $u(\hat{\tau})$.

4.2.2 Testing under unknown break date

Note that, while the structural break case is our main focus, the results in this section hold for a general form of structural change as long we assume that the change point is identified.

Assumption 2 *The function $\mu(\tau; \theta_{\tau_b})$ has a unique extremum at $\tau = \tau_b$.*

For the structural break model it is easy to verify that Assumption 2 holds. The extremum value of (19) is given by $\mu(\tau_b; \theta_{\tau_b}) = \theta_{\tau_b} \sqrt{\tau_b(1 - \tau_b)} = \zeta_{\tau_b}^{1/2}$.

Under Assumption 2, and for a small nominal size, we show below that rejections are found only for break locations that are close to τ_b . The following theorem shows that the estimated location of the break is close to the true break date.

Theorem 1 (Location concentration) *Suppose $Q^*(\tau) = [Z(\tau) + \mu(\tau; \theta_{\tau_b})]^2$ where $Z(\tau)$ is a zero mean Gaussian process with variance equal to one and $|\mu(\tau; \theta_{\tau_b})|$ satisfies Assumption 2, then as $u \rightarrow \infty$*

$$\mathbb{P} \left(\sup_{\tau \in \mathcal{I}} Q^*(\tau) > u^2 \right) = \mathbb{P} [Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}_1] [1 + o(1)]$$

where $\mathcal{I} = [\tau_{\min}, \tau_{\max}]$, $\mathcal{I}_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$ and $\delta(u) = u^{-1} \log^2 u$.

The proof is presented in Appendix A.3. The location concentration is necessary to show that the proposed test controls size and has near optimal power. Close inspection of the proof of Theorem 1 reveals that for the break magnitudes we find when solving (23) the concentration is expected to hold for conventional choices of the level of the test. This is indeed confirmed by the simulation results in Section 5.

For each break date τ_b and corresponding break magnitude θ_{τ_b} for which (23) equals zero, we can obtain a critical value $u(\tau_b)$ such that $P(\sup_{\tau \in \mathcal{I}} Q^*(\tau) > u(\tau_b)^2) = \alpha$. This yields a sequence of critical values $u(\tau_b)$ that depend on the unknown break date τ_b .

Assumption 3 (Slowly varying critical values) *Suppose that $u(\tau_b)$ is a differentiable function with respect to τ_b , then the critical values are slowly varying with τ_b in comparison to the derivative of the function $\mu(\tau; \theta_{\tau_b})$ with respect to τ on the interval \mathcal{I}_1 , i.e.*

$$\left| \frac{\partial u(\tau_b)}{\partial \tau_b} \right| < \left| \frac{\partial \mu(\tau; \theta_{\tau_b})}{\partial \tau} \right| < \infty$$

In the structural break model, the derivative $\gamma = \frac{\partial \mu(\tau; \theta_{\tau_b})}{\partial \tau} = \theta_{\tau_b} [\tau_b(1 - \tau_b)]^{-1/2}$. The assumptions that critical values vary slowly relates the dependence of the critical values on τ_b to the identification strength of the break

date as the derivative of $\mu(\tau; \theta_{\tau_b})$ with respect to τ scales linearly with the break magnitude. It was shown in Section 2 that $\theta_{\tau_b} \sqrt{\tau_b(1 - \tau_b)} \geq 1$, where the equality holds if the break date is known with certainty. Therefore,

$$\gamma = \frac{\theta_{\tau_b}}{\sqrt{\tau_b(1 - \tau_b)}} \geq \frac{1}{\tau_b(1 - \tau_b)}$$

A sufficient condition for the slowly varying assumption is therefore

$$\left| \frac{\partial u(\tau_b)}{\partial \tau_b} \right| \leq \frac{1}{\tau_b(1 - \tau_b)} \quad (24)$$

This inequality can be verified once critical values are obtained. In Appendix A.7 we show that the inequality holds for the case of the structural break model.

Under the assumption above, the following theorem guarantees that the size of the test is controlled at the desired level once the critical value $u(\tau_b)$ is replaced by the critical value $u(\hat{\tau})$.

Theorem 2 (Size) *Suppose $u(\tau_b)$ is a sequence of critical values such that, for a break of magnitude θ_{τ_b} at time τ_b , we have that*

$$P \left(\sup_{\tau \in \mathcal{I}} Q^*(\tau) > u(\tau_b)^2 \right) = \alpha \quad (25)$$

Then as $u(\tau_b) \rightarrow \infty$

$$P \left(\sup_{\tau \in \mathcal{I}} Q^*(\tau) > u(\hat{\tau})^2 \right) = \alpha \quad (26)$$

where $\hat{\tau}$ is given in (20).

The proof is in Appendix A.4. Using critical values $u(\hat{\tau})$, we can also establish that the test is near optimal in the sense that the power converges to the power of a test conditional on τ_b . Suppose the critical values for the latter test are given by $v(\tau_b)$ such that $P_{H_0} (Q^*(\tau_b) > v(\tau_b)^2) = \alpha$, then we can establish the following theorem.

Theorem 3 (Near optimal power) *Suppose Assumption 3 holds, then*

$$\begin{aligned} & P_{H_a} \left[\sup_{\tau} Q^*(\tau) > u(\hat{\tau})^2 \right] - P_{H_a} [Q^*(\tau_b) > v(\tau_b)^2] \\ & \geq P_{H_a} [Q^*(\tau_b) > u(\tau_b)^2] - P_{H_a} [Q^*(\tau_b) > v(\tau_b)^2] \\ & = 0 \end{aligned} \quad (27)$$

where $\hat{\tau} = \arg \sup_{\tau} Q^*(\tau)$ and P_{H_a} denotes the crossing probability under the alternative.

Appendix A.5 contains the proof.

A test based on the Wald statistic (17) uses critical values that depend on the estimated break date. The following corollary provides a test statistic with critical values that are independent of the break date in the limit where $u \rightarrow \infty$.

Corollary 1 *A test statistic with critical values that are independent of τ_b for $u \rightarrow \infty$ is given by*

$$S(\hat{\tau}) = \sup_{\tau \in \mathcal{I}} \sqrt{T} \frac{\left| \mathbf{f}'_{\beta_2} \left(\hat{\beta}_2(\tau) - \hat{\beta}_1(\tau) \right) \right|}{\sqrt{\mathbf{f}'_{\beta_2} \left(\frac{\hat{V}_1}{\tau} + \frac{\hat{V}_2}{1-\tau} \right) \mathbf{f}_{\beta_2}}} - |\mu(\hat{\tau}; \theta_{\hat{\tau}})| \quad (28)$$

where $\hat{\tau}$ maximizes the first term of S or, equivalently, the Wald statistic (17).

The proof is presented in Appendix A.6.

Finally, following from the location concentration established in Theorem 1, in the limit where $\alpha \rightarrow 0$, inference following a rejection is standard.

Corollary 2 (Corollary 8.1 of Piterbarg (1996)) *As $u \rightarrow \infty$, the distribution of the break location denoted by D converges to a delta function located at $\tau = \tau_b$ for excesses over the boundary u^2 , i.e.*

$$D \left(\hat{\tau} : Q^*(\hat{\tau}) = \sup_{\tau \in \mathcal{I}} Q^*(\tau) \mid \sup_{\tau \in \mathcal{I}} Q^*(\tau) > u^2 \right) \xrightarrow{d} \delta_{\tau_b} \text{ as } u \rightarrow \infty$$

4.2.3 Testing procedure

To summarize, we use the following steps to make the test for $\Delta(\tau_b) = 0$ in (23) operational

1. Using (20), evaluate (23) using simulation to find, for each τ_b , the break magnitude θ_{τ_b} that yields $\Delta(\tau_b) = 0$.
2. For each τ_b and corresponding θ_{τ_b} obtain a critical value $u(\tau_b)$ such that $P(\sup_{\tau \in \mathcal{I}} Q^*(\tau_b) > u(\tau_b)^2) = \alpha$.
3. Now the test statistic $\sup_{\tau \in \mathcal{I}} Q^*(\tau)$ or its finite sample analogue can be compared to the critical value $u(\hat{\tau})^2$ with $\hat{\tau}$ from (20).
 - This test controls size $P(\sup_{\tau \in \mathcal{I}} Q^*(\tau) > u(\hat{\tau})^2) = \alpha$ when α is sufficiently small per Theorem 2.
 - The power of this test approaches that of the infeasible test $P(Q^*(\tau_b) > v(\tau_b)^2)$ per Theorem 3.

The above procedure can also be performed to make the test statistic (28) operational, which leads to critical values that independent of the unknown break date for sufficiently small magnitude. We will present critical values in Section 5.

4.3 Optimal weights or shrinkage forecasts

Pesaran et al. (2013) derive optimal weights for observations in an estimation sample such that, in the presence of a structural break, the MSFE of the one-step-ahead forecast is minimized. Conditional on the break date, the optimal weights take one value for observations in the pre-break regime and one value for observations in the post-break regime. This implies that we can write the optimally weighted forecast as a convex combination of the forecasts from pre-break observations and post-break observations

$$\hat{y}_{T+h}^S(\tau) = \omega f_{T+h}(\hat{\beta}_1) + (1 - \omega) f_{T+h}(\hat{\beta}_2)$$

where the optimal forecast is denoted with subscript S as we will show now that it is equal to a shrinkage forecast that shrinks the post-break sample based forecast in the direction of the full sample based forecast. For ease of exposition, we assume here that all parameters break.

The asymptotic, expected mean square forecast error minus the variance of the forecast period's error is

$$\begin{aligned} & \lim_{T \rightarrow \infty} \text{E} \left[T (\hat{y}_{T+h}^S - f_{T+h}(\beta_2))^2 \right] = \\ & = \lim_{T \rightarrow \infty} \text{E} \left[T \left(\omega \mathbf{f}'_{\beta_2}(\hat{\beta}_1 - \hat{\beta}_2) + \mathbf{f}'_{\beta_2}(\hat{\beta}_2 - \beta_2) \right)^2 \right] + o(1) \\ & = \omega^2 T [\mathbf{f}'_{\beta_2}(\beta_1 - \beta_2)]^2 + \omega^2 \mathbf{f}'_{\beta_2} \left(\frac{1}{\tau_b} + \frac{1}{1 - \tau_b} \right) \mathbf{V} \mathbf{f}_{\beta_2} \\ & \quad - 2\omega \frac{1}{1 - \tau_b} \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} + \frac{1}{\tau_b} \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} + o(1) \end{aligned} \quad (29)$$

where $\mathbf{f}_{\beta_2} = \frac{\partial f_{T+h}(\beta_2)}{\partial \beta_2}$ and the first equality relies on a Taylor expansion and the local-to-zero nature of the breaks. See Appendix A.9 for details.

Maximizing (29) with respect to ω and ignoring the lower order term, yields

$$\omega^* = \tau_b \left[1 + T \frac{[\mathbf{f}'_{\beta_2}(\beta_1 - \beta_2)]^2}{\mathbf{f}'_{\beta_2} \left(\frac{1}{\tau_b} + \frac{1}{1 - \tau_b} \right) \mathbf{V} \mathbf{f}'_{\beta_2}} \right]^{-1} \quad (30)$$

where the denominator contains the Wald statistic, $\zeta(\tau_b)$, derived above.

Alternatively, we can combine the full sample forecast and the post-break sample forecast. Since, $\hat{\beta}_F = \tau_b \hat{\beta}_1 + (1 - \tau_b) \hat{\beta}_2 + o_p(T^{-1/2})$,

$$\begin{aligned} \hat{y}_{T+h}^S & = \omega f_{T+h}(\hat{\beta}_1) + (1 - \omega) f_{T+h}(\hat{\beta}_2) + o_p(T^{-1/2}) \\ & = \frac{\omega}{\tau_b} f_{T+h}(\hat{\beta}_F) + \left(1 - \frac{\omega}{\tau_b} \right) f_{T+h}(\hat{\beta}_2) + o_p(T^{-1/2}) \end{aligned}$$

and after applying a Taylor expansion of the forecast function f_{T+h} , the optimal weight on the full sample forecast is given by

$$\omega_F^* = \frac{\omega^*}{\tau_b} = \frac{1}{1 + \zeta(\tau_b)} \quad (31)$$

The shrinkage estimator is therefore a convex combination of the full sample and post-break sample forecast with weights that are determined by our Wald test statistic.

The empirical results in Pesaran et al. (2013) suggest that uncertainty around the break date substantially deteriorates the accuracy of the optimal weights forecast. As a consequence, Pesaran et al. (2013) derive robust optimal weights by integrating over the break dates, which yield substantially more accurate forecasts in their application. Given the impact that break date uncertainty has on choosing between the post-break and the full sample forecasts, it is not surprising that the same uncertainty should affect the weights. If this uncertainty is not taken into account, the weight on the post-break forecast will be too high. It will therefore be useful to test whether the break date uncertainty is small enough to justify using the shrinkage forecast.

As the Wald statistic in (31) is conditional on the true break date, consider the shrinkage forecast for a general value of τ

$$\begin{aligned}\hat{y}_{T+h}^S(\tau) &= \frac{1}{1+\zeta(\tau)}f_{T+h}(\hat{\beta}_F) + \frac{\zeta(\tau)}{1+\zeta(\tau)}f_{T+h}(\hat{\beta}_2(\tau)) \\ &\Rightarrow \frac{1}{1+Q^*(\tau)}f_{T+h}(\hat{\beta}_F) + \frac{Q^*(\tau)}{1+Q^*(\tau)}f_{T+h}(\hat{\beta}_2(\tau))\end{aligned}\tag{32}$$

where the last line holds by the continuous mapping theorem. The asymptotic expressions for $\hat{\beta}_2$ and $\hat{\beta}_F$ are provided in (12) and (14). The difference in MSFE between the shrinkage forecast and the full sample forecast, after applying again a Taylor expansion on the forecast function f_{T+h} , is given by

$$\begin{aligned}\Delta_s &= TE \left[\left(\frac{1}{1+Q^*(\hat{\tau})} \mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2) + \frac{Q^*(\hat{\tau})}{1+Q^*(\hat{\tau})} \mathbf{f}'_{\beta_2}(\hat{\beta}_2(\hat{\tau}) - \beta_2) \right)^2 \right] \\ &\quad - TE \left[\left(\mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2) \right)^2 \right] + o(1)\end{aligned}\tag{33}$$

where we solve for $\Delta_s = 0$ numerically to obtain the break magnitude that corresponds to equal predictive accuracy. Numerical results in Appendix A.8 show that equal predictive accuracy is associated with a unique break magnitude for each τ_b . The testing procedure outlined in Section 4.2.3 can be directly applied to find the appropriate critical values.

5 Simulations

5.1 Asymptotic analysis for standard size

The theoretical results of the previous section are derived under the assumption that the nominal size tends to zero. In this section, we investigate

the properties of our tests using simulations under conventional choices for nominal size, $\alpha = \{0.10, 0.05, 0.01\}$, while maintaining the assumption that $T \rightarrow \infty$. We will study for which break magnitude the MSFE from the post-break forecast equals that of the full sample forecast. Conditional on this break magnitude, we use simulation to obtain critical values. Finally, we study the size and power properties of the resulting tests.

5.1.1 Implementation

We simulate (18) with (19) for different combinations of the break date and break magnitude $\{\tau_b, \theta_{\tau_b}\}$. Here, we focus on $\tau_b = \{\tau_{\min}, \tau_{\min} + \delta_\tau, \dots, \tau_{\max}\}$ where $\tau_{\min} = 0.15$, $\tau_{\max} = 1 - \tau_{\min}$ and $\delta_\tau = 0.01$. Additionally, we used $\tau_{\min} = 1 - \tau_{\max} = 0.05$ and those results are reported in Appendix B. For the break magnitude, θ_{τ_b} , we consider $\theta_{\tau_b} = \{0, 0.5, \dots, 20\}$. The Brownian motion is approximated by dividing the $[0, 1]$ interval in $n = 1000$ equally spaced parts, generating $\epsilon_i \sim N(0, 1)$ and $B(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n\tau} \epsilon_i$, see, for example, Bai and Perron (1998).

By maximizing (18) we obtain a distribution of the estimated break date $\hat{\tau}$ that can be used to evaluate (22). To approximate the expectation, we use 50,000 repetitions for each break date and break magnitude. For each value of τ_b , a θ_{τ_b} is obtained for which the full sample forecast and the post-break forecast yield equal predictive accuracy using (22). This translates the null hypothesis of equal predictive accuracy into a null hypothesis regarding the break magnitude conditional of the break date τ_b . By simulating under the null hypothesis for each τ_b , we obtain critical values that are conditional on τ_b . Accurate estimation of the break date implies that these critical values can be used for testing without correction. The magnitude of the breaks that we find under the null hypothesis suggest that the estimated break date will, in fact, be quite accurate.

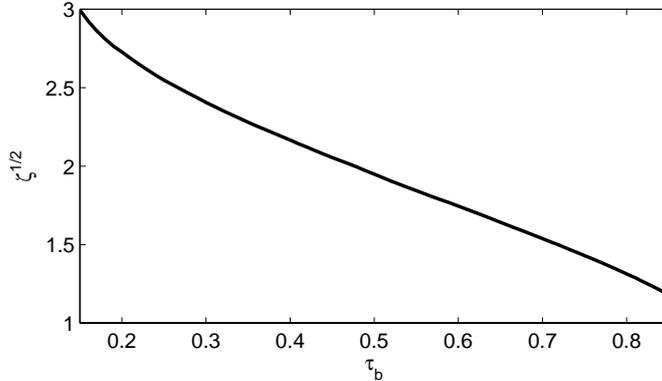
5.1.2 Post-break forecast versus full-sample forecast: break magnitude for equal forecast accuracy

Using (22), we simulate the break magnitude for which the full sample and the post-break sample achieve equal predictive accuracy. Figure 1 shows the combinations of break magnitude and break date for which equal predictive accuracy is obtained. The break magnitude is given in units of the standardized break magnitude,

$$\zeta^{1/2} = \sqrt{T(1 - \tau_b)\tau_b} \frac{\mathbf{f}'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)}{\sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}}} \quad (34)$$

so that it can be interpreted as a standard deviations from a standard normal.

Figure 1: Break magnitude for equal predictive accuracy between post-break and full sample forecasts



Note: The graph shows the standardized break magnitude, $\zeta^{1/2}$, in (34) for which the forecasts based on the post-break sample and the full sample achieve the same MSFE, that is, Δ in (22) equals zero.

The figure shows that for each break date τ_b , the break magnitude of equal forecast accuracy is substantially larger than that under a known break date, which is $\zeta^{1/2} = 1$, as the MSFE of the post-break sample forecast increases due to the uncertainty of the break date estimation. If a break occurs early the sample, $\tau = 0.15$, the post-break forecast is more precise only if the break magnitude is larger than three standard deviations. The break magnitude uniformly decreases as the break date, τ , increases and reaches about 1.2 standard deviations at $\tau = 0.85$.

The intuition for the downward sloping nature of the break magnitude of equal forecast accuracy is as follows. The local-to-zero nature of the break implies that, even asymptotically, the break date is estimated with uncertainty and has a non-degenerate distribution around the true break date. The uncertainty surrounding the break date implies that estimated post-break samples may be too short, increasing the forecast variance, or too long and include a pre-break sample, which leads to a forecast bias. The former leads to an increase in MSFE. The latter can reduce the MSFE as it trades off the increase in the bias for a reduction in variance (Pesaran and Timmermann, 2007). However, this benefit decreases as the post-break sample increases.

Additionally, supremum type test statistics require a trimming of dates over which breaks are allowed. Trimming leads to a truncation of the distribution of break dates at both ends of the sample. From a forecasting perspective, the effect of this truncation is not symmetric over the break dates. If the true break is early in the sample, the distribution is left trun-

Table 1: Critical values and size of the W and S test statistics

Test	α	Critical values					Size				
		0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
W	0.01	30.54	27.29	22.29	18.22	15.82	0.01	0.01	0.01	0.01	0.01
	0.05	23.71	20.99	16.74	13.30	11.37	0.07	0.06	0.06	0.04	0.03
	0.10	20.44	17.99	14.13	11.04	9.36	0.13	0.12	0.11	0.09	0.06
S	0.01	2.76	2.81	2.87	2.80	2.60	0.01	0.01	0.01	0.01	0.01
	0.05	2.12	2.18	2.23	2.14	1.94	0.05	0.05	0.06	0.05	0.04
	0.10	1.78	1.84	1.89	1.80	1.59	0.10	0.10	0.11	0.11	0.08

Note: Reported are critical values and size for, first, W , the Wald test statistic (17) and, second, S , the test statistic (28), which is independent of τ_b when the nominal size tends to zero.

cated and the break date is likely to be, on average, estimated too late. The forecasts are therefore less likely to benefit from the MSFE reduction of a longer sample and more likely to have an estimation sample that is too short, which implies a larger variance without the benefit of a bias reduction. If, in contrast, the true break date is late in the sample, the distribution will be right truncated and therefore lead to an estimated break date that is, on average, too early. The estimation window will likely contain a short pre-break sample that reduces the MSFE and is less likely to be inefficiently short. Therefore, if the break is late, the break magnitude for which the post-break forecast is preferred over the full-sample forecast is smaller compared to the case when the break is early. This is reflected in the downward slope of the critical break magnitude observed in Figure 1.

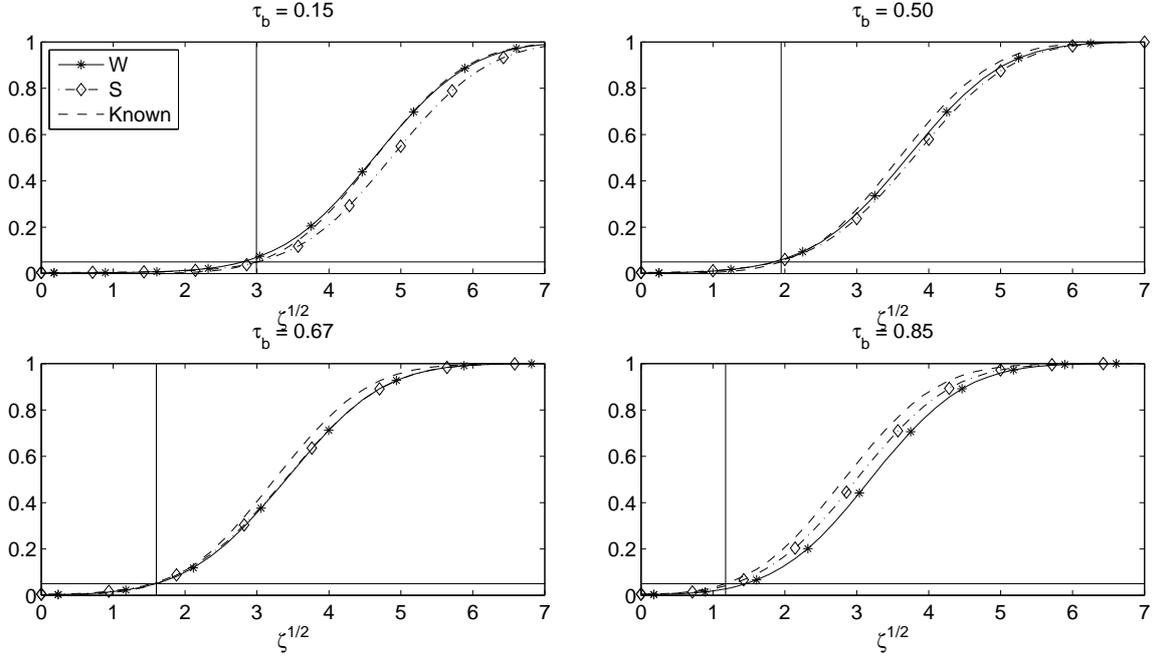
5.1.3 Critical values, size, and power

After finding the break magnitude for which the post-break sample forecast and the full sample forecast yield equal predictive accuracy, we can compute critical values for both the Wald-type test statistic, W , in (17) and the α -asymptotic statistic, S , in (28) for a grid of break dates, τ_b . Condition (24), which is required for the near optimality result does hold for all τ_b —details are available in Appendix A.7.

The first line of the right panel of Table 1 shows that the test has the correct size for $\alpha = 0.01$. For $\alpha = 0.05$ and 0.1 size is still very close to the asymptotic size. At the beginning and the end of the sample, however, some size distortion occurs. Using the corrected test statistic (28) largely remedies these size distortions.

The critical values are given in the left panel of Table 1. Critical values for a finer grid of the true break date can be found in Appendix B. The large

Figure 2: Asymptotic power when testing between a post-break and full-sample forecast at $\alpha = 0.05$



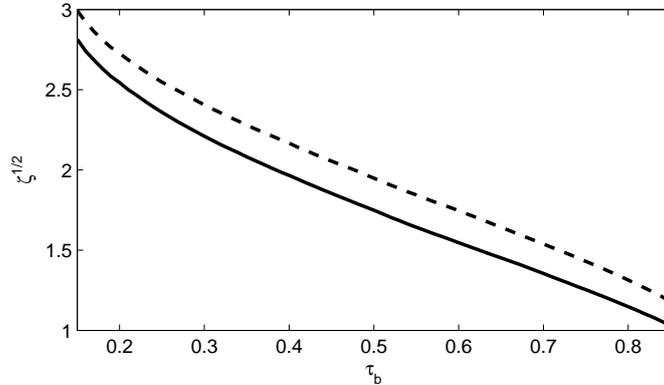
Note: The plots show the power for tests at a nominal size of $\alpha = 0.05$ with the null hypothesis given by the break magnitude depicted in Figure 1. The panels show power for different values of the (unknown) break date. The power of infeasible test conditional on the true break date is given as the dashed line, that of the test statistic W as the solid line with stars, and that of the test statistic S as the dashed line with diamonds. The solid horizontal line indicates the nominal size, and the vertical solid line indicates the break magnitude at which equal predictive accuracy is achieved corresponding to Figure 1.

break magnitude that yields equal forecast accuracy implies a major increase in critical values when using the Wald test statistic (17), compared to the standard values of Andrews (1993). For a nominal size of $[0.10, 0.05, 0.01]$ the critical values in Andrews are equal to $[7.17, 8.85, 12.35]$.

The critical values for the α -asymptotic test statistic, S , in (28) are independent of $\hat{\tau}$ in the limit where $\alpha \rightarrow 0$. Under a known break date, critical values would be from a one-sided normal distribution, that is, they would be $[1.64, 2.33, 2.58]$ for nominal size of $[0.10, 0.05, 0.01]$. The critical values for the corrected test, S , in (28) vary substantially less over $\hat{\tau}$ than those for the Wald statistic, W , in (17). The results in Section 4.2.2 suggest that the differences to the critical values that would be used if the break date is known diminish as $\alpha \rightarrow 0$ and this can be observed in Table 1.

Given that the break magnitudes that lead to equal forecast performance

Figure 3: Break magnitude for equal predictive accuracy of shrinkage and full sample forecasts



Note: The solid line shows the standardized break magnitude for which the shrinkage forecast (32) achieves the same MSFE as the full sample forecast, in which case (33) equals zero. For comparison, the dashed line shows the break magnitude for which the post-break forecast and the full sample forecast achieve equal MSFE.

are reasonably large, we expect the tests to have relatively good power properties. The power curves in Figure 2 show that the power of both tests is close to the power of the optimal test which uses the known break date to test whether the break magnitude exceeds the boundary depicted in Figure 1. The good power properties are true for all break dates. This confirms that the theoretical results for vanishing nominal size extend to conventional choices of the nominal size.

5.1.4 Shrinkage forecast versus full-sample forecast

Figure 3 shows the combination of τ_b and break magnitude for which the shrinkage forecast of Section 4.3 and the full sample forecast that weights observations equally have the same MSFE, which is represented by the solid line in the graph. For comparison, the dashed line gives the combination of post-break forecast and full sample forecast that have the same MSFE, that is, the line from Figure 1. It can be seen that the break magnitude of equal forecast performance for the shrinkage forecast is lower than for the post-break sample forecast. This implies that the shrinkage forecast is more precise than the post-break forecast for smaller break magnitudes for a given break date. However, the difference is relatively small and breaks need to be quite large before the shrinkage estimator is more precise than the full sample estimator.

In order to determine whether to use the shrinkage forecast, critical val-

Table 2: Critical values and size: shrinkage versus full sample forecasts

Test	α	Critical values					Size				
		0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
W	0.01	28.74	25.57	20.74	17.03	15.02	0.02	0.01	0.01	0.01	0.01
	0.05	22.15	19.51	15.43	12.34	10.74	0.07	0.07	0.06	0.04	0.03
	0.10	19.01	16.63	12.95	10.19	8.82	0.14	0.13	0.11	0.08	0.06
S	0.01	2.82	2.87	2.91	2.82	2.63	0.01	0.01	0.01	0.01	0.01
	0.05	2.18	2.24	2.27	2.17	1.98	0.05	0.05	0.06	0.05	0.04
	0.10	1.85	1.90	1.93	1.82	1.63	0.10	0.10	0.11	0.11	0.08

Note: Reported are critical values and size when testing for equal MSFE of the shrinkage forecast (32) and the full sample forecast using, first, W , the Wald test statistic in (17) and, second, S , the test statistic (28) that is independent of τ_b when the nominal size tends to zero.

ues can be obtained in a similar fashion as before and are presented in Table 2. Again, the size is close to the theoretical size with small size disturbances when using W , which are largely remedied when using S . Critical values on a finer grid of the true break date are presented in Appendix B.

Figure 4 displays the power curves of the tests that compare the shrinkage forecast and the full sample, equal weights forecast. Since, the break magnitudes for equal forecast performance are similar to the post-break sample forecast, it is not surprising that the properties in terms of size and power of the tests for the shrinkage forecast are largely the same as those for the post-break forecast.

5.1.5 Shrinkage forecast versus the post-break forecast

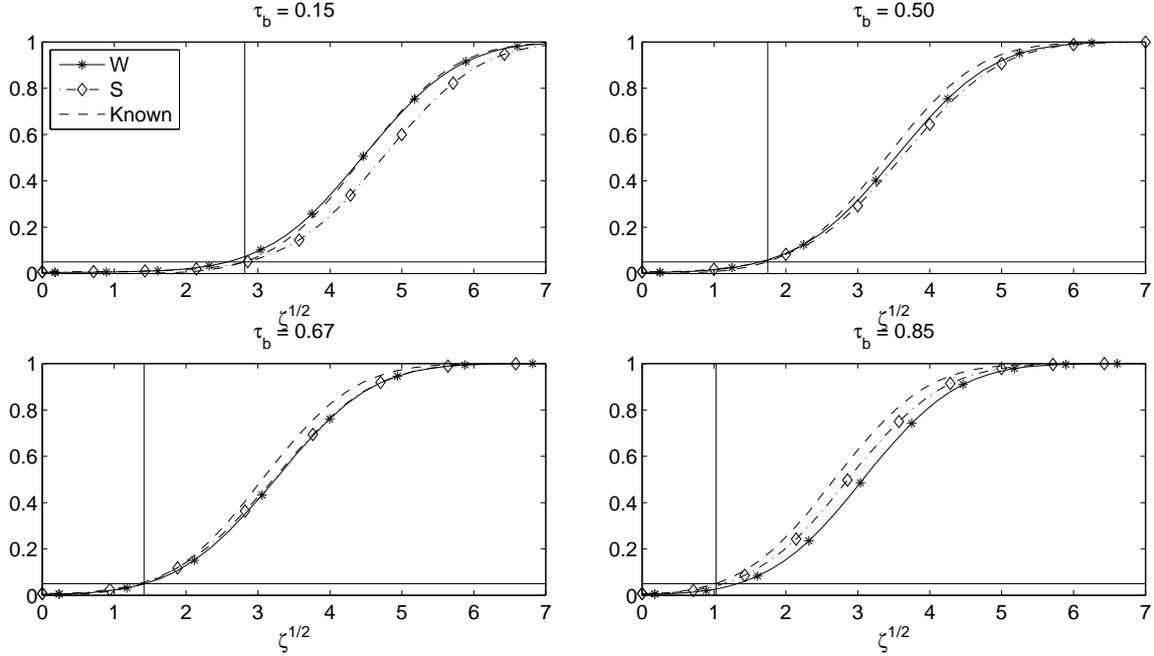
Finally, we investigate the break magnitudes that leads to equal forecast performance of the post-break forecast and the shrinkage forecast. Figure 5 plots the ratio of the MSFE of the shrinkage forecast over that of the post-break forecast. For nearly all break magnitudes and dates, the shrinkage forecast outperforms the post-break forecast. Only when the break occurs at the end of the sample and is relatively large, the post-break forecast is slightly more accurate.

5.2 Finite sample analysis

5.2.1 Set up of the Monte Carlo experiments

We analyze the performance of the tests in finite sample for an AR(1) model with varying degree of persistence. We consider the two tests for equal predictive accuracy between the post-break forecast and the full-sample fore-

Figure 4: Asymptotic power when testing at $\alpha = 0.05$ between the shrinkage and full-sample forecast



Note: The plots show asymptotic power curves when testing for equal predictive accuracy between the shrinkage forecast (32) and the full-sample forecast using the break magnitude depicted in Figure 3 for different values of the break date τ_b . For more information, see the footnote of Figure 2.

cast based on the Wald statistic (17) and on the S -statistic (28). Next, we consider the same test statistics but now test for equal predictive accuracy between the shrinkage forecast (32) and the full-sample, equal weighted forecast. All tests are carried out at a nominal size $\alpha = 0.05$, using sample sizes of $T = \{120, 240, 480\}$ and break dates $\tau_b = [0.15, 0.25, 0.50, 0.75, 0.85]$. Parameter estimates are obtained by least squares, and the results are based on 10,000 repetitions.

The data generating process (DGP) is given by

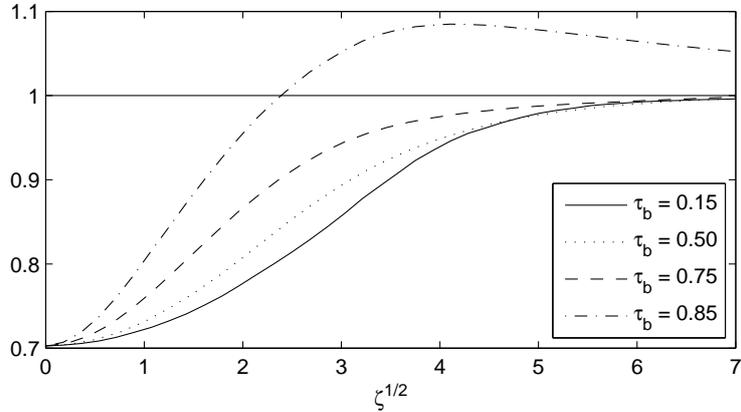
$$y_t = \mu_t + \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \quad (35)$$

where $\sigma^2 = 1$ and

$$\mu_t = \begin{cases} \mu_1 & \text{if } t \leq \tau_b T \\ \mu_2 & \text{if } t > \tau_b T \end{cases}$$

We set $\mu_1 = -\mu_2$ and $\mu_1 = \frac{1}{2\sqrt{T}}\zeta^{1/2}(\tau_b) + \frac{1}{2}\frac{\lambda}{\sqrt{T\tau_b(1-\tau_b)}}$. When $\lambda = 0$ the experiments deliver the finite sample size, whereas $\lambda = \{1, 2\}$ shows the

Figure 5: Relative MSFE of shrinkage and post-break sample forecasts



Note: The graph shows the relative performance of the shrinkage forecast (32) and the post-break sample forecast as a function of the standardized break magnitude $\zeta^{1/2}$ for different values of the break date τ_b . The horizontal solid line corresponds to equal predictive accuracy. Values below 1 indicate that the shrinkage forecast is more precise.

power of the tests. The influence of the degree of persistence on the results is analyzed by varying $\rho = \{0.0, 0.3, 0.6, 0.9\}$.

5.2.2 Results

The results in Table 3 show that for models with low and moderate persistence, $\rho = 0.0$ and 0.3 , the size of the W and S tests are extremely close to the nominal size irrespective of the sample size and the break date. As persistence increases to $\rho = 0.9$, some size distortions become apparent for $T = 120$. Those do, however, diminish as T increases. These size distortions are similar for W and S and are the result of the small effective sample size in this setting. Power increases with λ . For $T = 120$ it is slightly larger when the break is in the middle of the sample but this effect disappears with increasing T . Overall, differences between W and S are small.

The results for the tests that compare the shrinkage forecast against the full sample, equal weights forecast in Table 4 are very similar to the results for the test with the post-break sample forecast under the alternative. Size is very close to the nominal size for large effective sample sizes and power increases in λ and, mildly, in T .

Overall, the results suggest that the W and S tests have good size and power properties unless the persistence of the time series is very high and this is combined with a small effective T .

Table 3: Finite sample analysis: size and power when testing between post-break and full-sample forecast

ρ	$\lambda \setminus \tau_b$	$T = 120$					$T = 240$					$T = 480$				
		0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
Wald-test (17)																
0.0	0	0.05	0.05	0.06	0.05	0.03	0.06	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.17	0.20	0.22	0.21	0.17	0.21	0.22	0.23	0.21	0.16	0.24	0.24	0.23	0.21	0.16
	2	0.43	0.48	0.52	0.53	0.47	0.52	0.54	0.55	0.53	0.48	0.57	0.56	0.56	0.55	0.49
0.3	0	0.04	0.05	0.06	0.05	0.03	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.13	0.17	0.21	0.21	0.17	0.18	0.20	0.22	0.20	0.16	0.22	0.23	0.22	0.21	0.16
	2	0.33	0.40	0.47	0.50	0.46	0.46	0.50	0.53	0.52	0.47	0.54	0.54	0.55	0.55	0.48
0.6	0	0.03	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.05	0.03	0.05	0.06	0.06	0.05	0.03
	1	0.08	0.12	0.19	0.20	0.16	0.13	0.17	0.20	0.20	0.15	0.18	0.20	0.22	0.21	0.15
	2	0.19	0.26	0.39	0.46	0.43	0.33	0.40	0.47	0.50	0.45	0.47	0.49	0.52	0.53	0.47
0.9	0	0.02	0.05	0.10	0.09	0.06	0.02	0.04	0.08	0.07	0.04	0.03	0.05	0.06	0.06	0.04
	1	0.04	0.07	0.17	0.24	0.20	0.04	0.08	0.16	0.21	0.16	0.07	0.11	0.17	0.20	0.15
	2	0.09	0.12	0.24	0.44	0.44	0.09	0.14	0.28	0.43	0.39	0.16	0.24	0.37	0.46	0.41
S-test (28)																
0.0	0	0.03	0.04	0.06	0.06	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.05	0.06	0.06	0.04
	1	0.13	0.16	0.21	0.23	0.22	0.16	0.18	0.21	0.23	0.21	0.17	0.19	0.21	0.23	0.21
	2	0.34	0.41	0.48	0.56	0.55	0.43	0.48	0.52	0.56	0.56	0.48	0.51	0.53	0.58	0.56
0.3	0	0.03	0.04	0.06	0.06	0.04	0.04	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.06	0.04
	1	0.09	0.14	0.19	0.23	0.22	0.13	0.16	0.20	0.23	0.20	0.16	0.18	0.21	0.23	0.21
	2	0.25	0.34	0.44	0.53	0.54	0.36	0.43	0.50	0.55	0.55	0.44	0.49	0.52	0.58	0.56
0.6	0	0.02	0.04	0.06	0.07	0.05	0.03	0.04	0.05	0.05	0.04	0.04	0.05	0.06	0.06	0.05
	1	0.05	0.09	0.17	0.23	0.21	0.09	0.13	0.19	0.22	0.21	0.13	0.16	0.20	0.23	0.21
	2	0.13	0.21	0.36	0.50	0.52	0.24	0.33	0.44	0.53	0.53	0.37	0.43	0.49	0.56	0.55
0.9	0	0.02	0.04	0.10	0.12	0.08	0.02	0.03	0.07	0.08	0.06	0.02	0.04	0.06	0.07	0.05
	1	0.03	0.05	0.16	0.28	0.26	0.02	0.06	0.14	0.24	0.22	0.04	0.08	0.16	0.23	0.21
	2	0.06	0.08	0.22	0.49	0.54	0.05	0.10	0.25	0.47	0.49	0.10	0.18	0.33	0.50	0.51

Note: The table presents finite sample size and power properties for the test comparing the post-break and full sample based forecasts. The DGP is $y_t = \mu_t + \rho y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim N(0, 1)$, $\mu_1 = -\mu_2$ and $\mu_1 = \frac{1}{2\sqrt{T}}\zeta^{1/2}(\tau_b) + \frac{1}{2} \frac{\lambda}{\sqrt{T\tau_b(1-\tau_b)}}$ where $\zeta^{1/2}(\tau_b)$ corresponds to Figure 1. The empirical size of the tests is obtained when $\lambda = 0$ and power when $\lambda = \{1, 2\}$. Tests are for a nominal size of 0.05.

Table 4: Finite sample analysis: size and power when testing between shrinkage and full-sample forecast

ρ	$\lambda \setminus \tau_b$	$T = 120$					$T = 240$					$T = 480$				
		0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
Wald-test (17)																
0.0	0	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.04	0.03	0.07	0.07	0.06	0.05	0.03
	1	0.18	0.21	0.22	0.21	0.16	0.22	0.23	0.22	0.20	0.15	0.24	0.24	0.23	0.21	0.15
	2	0.45	0.49	0.52	0.52	0.46	0.53	0.55	0.55	0.53	0.47	0.57	0.57	0.56	0.54	0.48
0.3	0	0.05	0.06	0.06	0.05	0.03	0.06	0.06	0.06	0.04	0.03	0.06	0.07	0.06	0.05	0.03
	1	0.15	0.19	0.22	0.20	0.16	0.20	0.21	0.22	0.20	0.15	0.23	0.23	0.22	0.21	0.15
	2	0.36	0.42	0.48	0.51	0.45	0.48	0.51	0.53	0.52	0.46	0.55	0.55	0.55	0.54	0.47
0.6	0	0.04	0.06	0.07	0.05	0.04	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.10	0.14	0.20	0.20	0.16	0.15	0.18	0.21	0.20	0.15	0.20	0.21	0.22	0.20	0.15
	2	0.22	0.30	0.42	0.47	0.43	0.36	0.42	0.49	0.50	0.44	0.48	0.51	0.53	0.52	0.46
0.9	0	0.03	0.07	0.12	0.10	0.07	0.04	0.05	0.09	0.07	0.05	0.04	0.06	0.07	0.06	0.04
	1	0.06	0.09	0.21	0.26	0.21	0.06	0.10	0.19	0.22	0.17	0.09	0.13	0.20	0.21	0.16
	2	0.11	0.15	0.30	0.48	0.45	0.12	0.18	0.34	0.46	0.41	0.20	0.28	0.41	0.47	0.42
S-test (28)																
0.0	0	0.04	0.05	0.06	0.06	0.04	0.04	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.06	0.04
	1	0.13	0.16	0.21	0.24	0.22	0.15	0.18	0.21	0.23	0.20	0.17	0.19	0.21	0.23	0.20
	2	0.34	0.42	0.49	0.56	0.55	0.42	0.47	0.52	0.56	0.55	0.46	0.50	0.53	0.58	0.56
0.3	0	0.03	0.05	0.06	0.06	0.04	0.04	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.06	0.04
	1	0.10	0.14	0.20	0.23	0.22	0.13	0.17	0.20	0.23	0.20	0.16	0.18	0.21	0.23	0.20
	2	0.25	0.35	0.45	0.54	0.54	0.36	0.43	0.50	0.56	0.54	0.43	0.48	0.52	0.58	0.55
0.6	0	0.03	0.05	0.07	0.07	0.05	0.03	0.05	0.06	0.06	0.04	0.04	0.05	0.06	0.06	0.05
	1	0.06	0.10	0.19	0.24	0.22	0.09	0.14	0.19	0.23	0.20	0.13	0.16	0.20	0.23	0.20
	2	0.14	0.23	0.39	0.52	0.52	0.25	0.35	0.45	0.54	0.53	0.37	0.43	0.49	0.56	0.55
0.9	0	0.02	0.05	0.12	0.12	0.09	0.02	0.04	0.09	0.09	0.06	0.03	0.04	0.07	0.08	0.06
	1	0.03	0.06	0.19	0.31	0.28	0.03	0.07	0.17	0.25	0.23	0.05	0.10	0.18	0.25	0.21
	2	0.06	0.11	0.27	0.53	0.56	0.07	0.12	0.30	0.50	0.51	0.12	0.21	0.37	0.52	0.52

Note: The table presents finite sample size and power properties of the tests comparing the shrinkage forecast (32) and the full-sample, equal weights forecast, using a nominal size of 0.05. For further details, see the footnote of Table 3.

6 Application

We investigate the importance of structural breaks for 130 macroeconomic and financial time series from the St. Louis Federal Reserve (FRED-MD) database, which is a monthly updated database. We use the vintage from May 2016. The data are described by McCracken and Ng (2016), who suggest various transformations to render the series stationary and to deal with discontinued series or changes in classification. In the vintage used here, the data start in January 1959 and end in April 2016. After the transformations, all 130 series are available from January 1960 until October 2015. Our first forecast is for July 1970 and we recursively construct one-step ahead forecasts until the end of the sample.

The data are split into 8 groups: *output and income* (OI, 17 series), *labor market* (LM, 32 series), *consumption and orders* (CO, 10 series), *orders and inventories* (OrdInv, 11 series), *money and credit* (MC, 14 series), *interest rates and exchange rates* (IRER, 21 series), *prices* (P, 21 series), and *stock market* (S, 4 series).

Following Stock and Watson (1996), we focus on linear autoregressive models of lag length $p = 1$ and $p = 6$ and test whether the intercept is subject to a break. We estimate parameters on a moving windows of 120 observations to decrease the likelihood of multiple breaks occurring in the estimation sample. Test results are based on heteroskedasticity robust Wald statistics, which use the following estimate of the covariance matrix $\hat{\mathbf{V}}_i = (\mathbf{X}'_i \mathbf{X}_i)^{-1} \mathbf{X}'_i \hat{\mathbf{\Omega}}_i \mathbf{X}_i (\mathbf{X}'_i \mathbf{X}_i)^{-1}$ with $[\hat{\mathbf{\Omega}}_i]_{kl} = \hat{\varepsilon}_k^2 / (1 - h_k)^2$ if $k = l$ and $[\hat{\mathbf{\Omega}}_i]_{kl} = 0$ otherwise, and h_k is the k -th diagonal element of $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. See MacKinnon and White (1985) and Long and Ervin (2000) for discussions of different heteroskedasticity robust covariance matrices. We have also obtained test results and forecasts using a larger window of 240 observations and using the homoskedastic Wald test and, qualitatively, our results do not depend on these choices.

6.1 Structural break test results

In this forecast exercise, we will refer to the test of Andrews (1993) as *supW*, the Wald test statistic (17) as *W*, the test statistic (28) as *S*, and, when the alternative is the shrinkage forecast, as *W^s* and *S^s*. For all tests we use $\alpha = 0.5$ and $\tau_{\min} = 0.15$. In Table 5, we report the fraction of estimation samples where the tests indicate a break. It is clear that a large fraction of the breaks picked up by *supW* are judged as irrelevant for forecasting by *W*, *S*, *W^s*, and *S^s*. The fraction of forecasts for which a break is indicated is lower by a factor of two for the AR(1) and by factor of up to three for the AR(6).

Figure 6 displays the number of estimation samples per series for which the tests were significant when forecasting with the AR(1), where within

Table 5: Fractions of estimation samples with a significant structural break

	supW	W	S	W^s	S^s
AR(1)	0.219	0.102	0.108	0.119	0.126
AR(6)	0.114	0.037	0.042	0.046	0.053

Note: supW refers to the Andrews' (1993) sup-Wald test, W and S refer to the tests developed in this paper that compare post-break and full sample forecasts, and W^s and S^s refer to the tests that compare shrinkage and full sample forecasts. All tests are carried out at $\alpha = 0.05$.

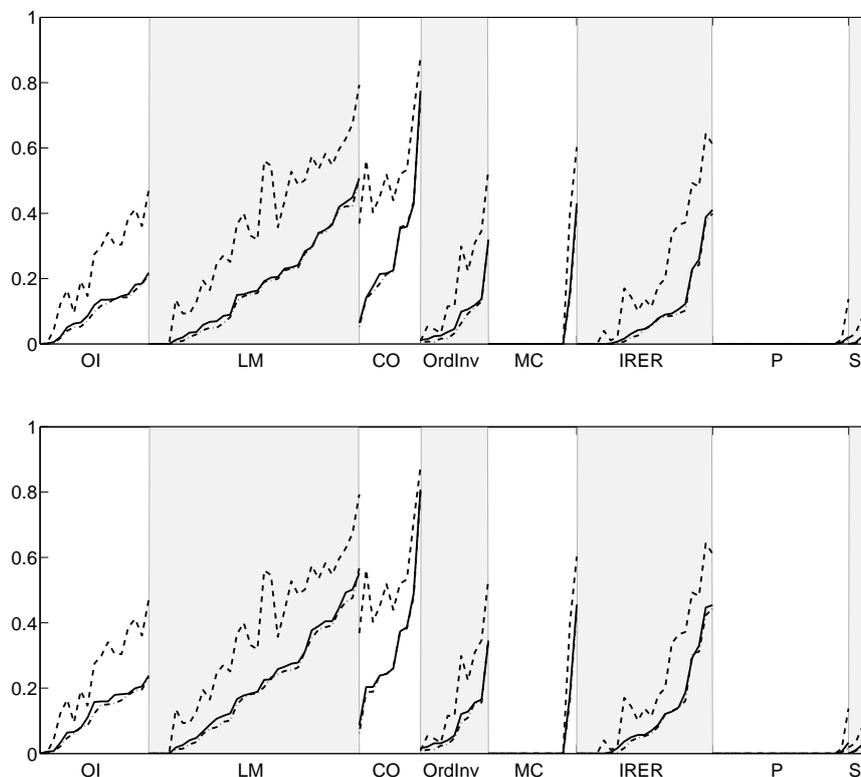
each category we sort the series based on the fraction of breaks found by W . Across all categories the *supW* test is more often significant than the W and S tests for both, post-break and shrinkage forecasts. Yet, we see substantial differences between categories. Whereas in the *labor market* and *consumption and orders* categories some of the series contain a significant break in up to 70% of the estimation samples when the W or S tests are used, the *prices* and *stock market* series hardly show any significant breaks from a forecasting perspective. This finding concurs with the general perception that, for these type of time series, simple linear models are very hard to beat in terms of MSFE.

Figure 7 displays the number of estimation samples with significant breaks for the AR(6) model. Compared to the results for the AR(1) in Figure 6, far fewer estimation samples contain a significant break, and this is true even in the *consumption and orders* category, which contained series with many breaks when using the AR(1). Consistent with the results for the AR(1), however, the W and S tests find fewer estimation samples with breaks than the *supW* test for virtually all series.

Figure 8 shows the occurrence of significant breaks over the different estimation samples when using the AR(1) model, where the end date of the estimation sample is given on the horizontal axis. In the top panel are the results for the test comparing the post-break estimation window with the full estimation window. In the bottom panel are the tests comparing the shrinkage estimator and the full sample, equal weights estimator. It is clear that the *supW* test finds more breaks in for the vast majority of estimation samples, whereas the results from the W and S tests are extremely similar.

A number of interesting episodes can be observed. While in the initial estimation samples the tests find a comparable number of samples with breaks, from 1985 the *supW* test finds many more series that contain breaks that are insignificant for the W and S test. This remains true until 2009 where the W and S tests find the same and, in the case of the shrinkage forecast, even more breaks that are relevant for forecasting than the *supW*

Figure 6: Fraction of significant structural break test statistics per series - AR(1)

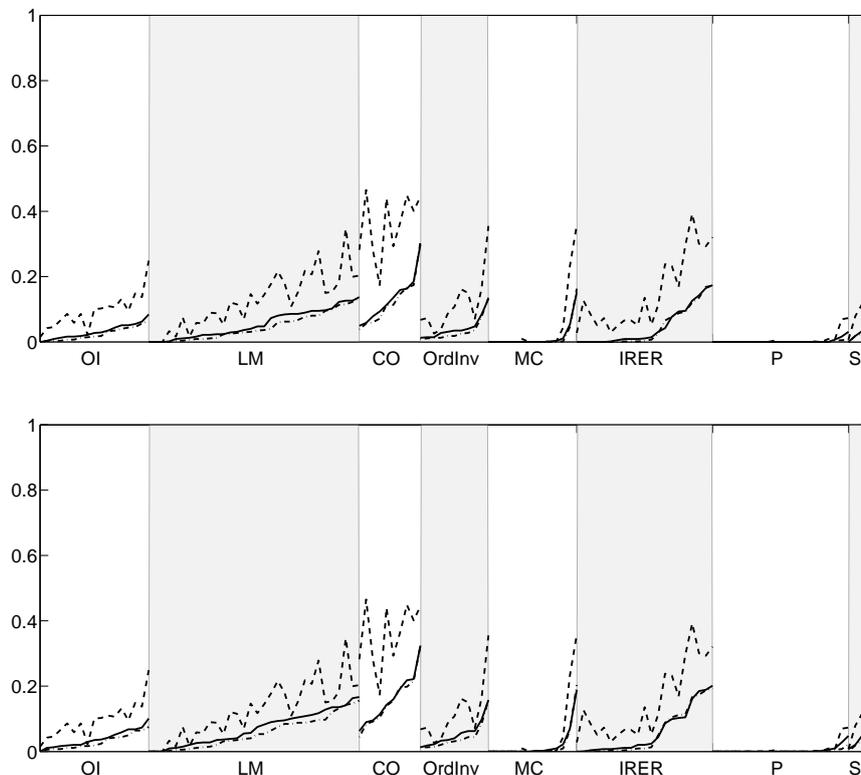


Note: The upper panel depicts the fraction of estimation samples with a significant break when testing under the alternative of the post-break forecast; the lower panel when testing under the alternative of the shrinkage forecast (32). Dashed lines indicate the fraction of estimation samples with significant sup-Wald test, dashed-dotted lines indicate the fraction of estimation samples where the break test W in (17) indicates a break, and solid lines indicate the fraction of estimation samples with significant S test in (28).

test. From 2010 onwards, breaks that are relevant for forecasting decrease sharply, whereas the $supW$ tests continues to find a large number of breaks. The intuition is that, as demonstrated in Figures 1 and 3, breaks early in the sample are less likely to be relevant for forecasting. However, the $SupW$ test does not use this information.

Figure 9 shows the results for the AR(6) model. In general, all tests find fewer estimation samples with breaks compared to the AR(1) model. The evolution over the estimation samples is, however, similar to the AR(1) case. In the initial estimation samples up to 1985 all tests agree that a small number of series are subject to a structural break. From 1985 to 1990, however, the $supW$ test finds breaks in up to a third of the estimation samples,

Figure 7: Fraction of significant structural break test statistics per series - AR(6)



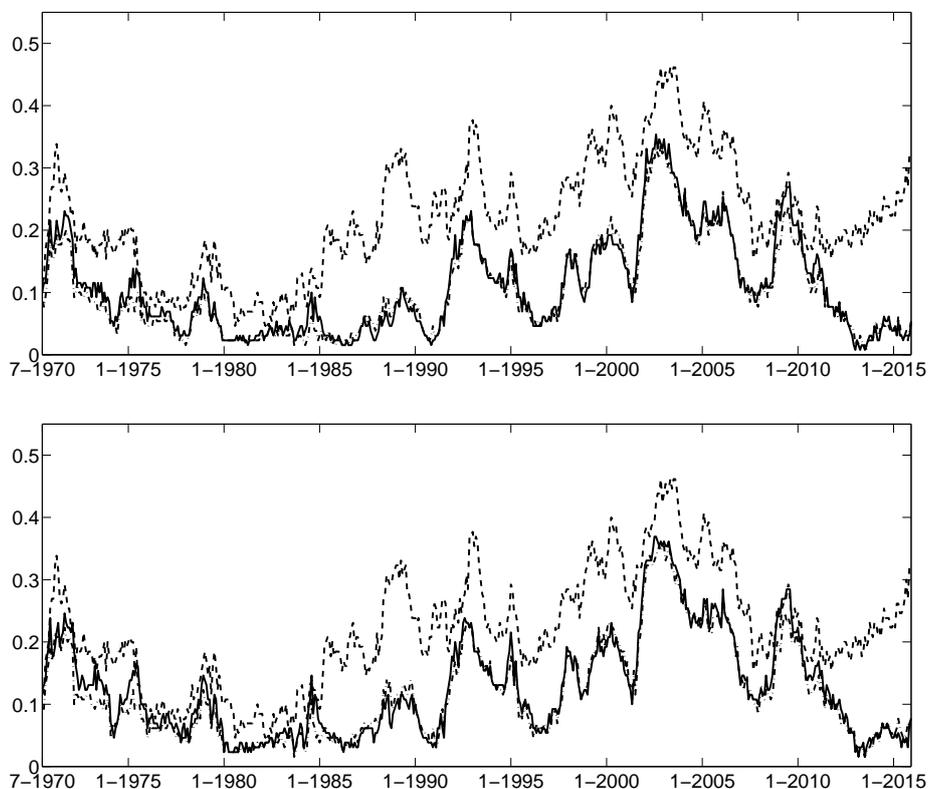
Note: See footnote of Table 6

most of which the W and S tests do not find important for forecasting. The same is true for breaks around 2000. In contrast, in the period following the dot com bubble and following the financial crisis of 2008/9 the W and the S tests find as many and, in the case of the shrinkage forecasts, more series, where taking a break into account will improve forecast accuracy than the $supW$ test. Again, the number of series that should take a break into account declines sharply towards the end of our sample when using the W and S tests but not when using the $supW$ tests.

6.2 Forecast accuracy

Given the different test results, we now investigate whether forecasts conditional on the W and S tests are more accurate than forecasts based on the $supW$ test. We use each test to determine whether to use the post-break or the full sample for forecasting or, alternatively, whether to use the shrinkage or the equal weights forecast and, given these results, we construct the respective forecast.

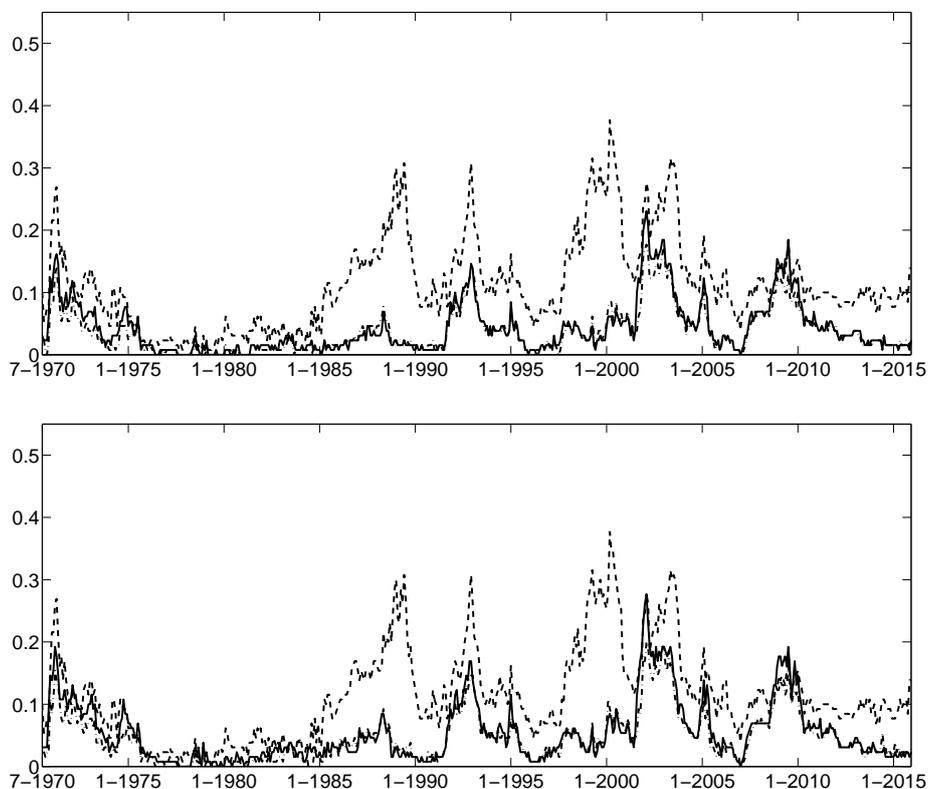
Figure 8: Fraction of significant structural break test statistics over estimation samples – AR(1)



Note: The plots show the fractions of series with a significant break for each estimation sample when using an AR(1) model with a break in intercept. The top panel shows results when testing between the post-break sample based forecast and the full sample based forecast and the lower panel when testing between the shrinkage forecast and the full sample, equal weights forecast. The dashed line indicates the fraction of series when testing using the standard sup-Wald test at $\alpha = 0.05$, the solid line when testing using the S-test in (28), and the dashed-dotted line when testing using the W-test in (17). The dates displayed on the horizontal axis are the end dates of the estimation samples.

Table 6 reports the MSFE of the respective forecasting procedures relative to the MSFE of the forecast based on the supW test of Andrews with the results for the AR(1) in the top panel and those for the AR(6) in the bottom panel. For each model, we report the average relative MSFE over all series in the first line, followed by the average relative MSFE for the series in the different categories. We report only the results for the estimation windows where at least one test finds a break as the estimation samples where no test finds a break will lead to identical full sample forecasts.

Figure 9: Fraction of significant structural break test statistics over estimation samples – AR(6)



Note: The plots show the fractions of series with a significant break for each estimation sample when using an AR(6) model with a break in intercept. For additional details, see the footnote of Figure 8.

The results show that using the W test in place of the $supW$ test leads to a 5.5% improvement in accuracy on average for the AR(1) and a 7.6% improvement in accuracy on average for the AR(6) model. This gain is similar for the S test with improvements of 4.9% and 6.5%. These improvements are found for series in all categories. The only exception is the use of the S test in the AR(1) model on the category ‘prices’. This suggests that the improvements are robust across the different series.

When the shrinkage forecast is used in conjunction with the W^s or S^s test, the accuracy of the forecasts is very similar as those of the post-break forecasts. This can be expected since we reject the test when the Wald statistic, that governs the amount of shrinkage, is relatively large. This implies that upon rejection of the test statistic, a forecast is used that is relatively close to the post-break forecast. The last column shows that using the shrinkage forecast in conjunction with the $supW$ test leads to forecasts

Table 6: Relative MSFE compared to the standard sup-Wald test

		Post-break		Shrinkage		
		W	S	W	S	supW
AR(1)	All series	0.948	0.953	0.948	0.949	0.983
	OI	0.972	0.981	0.970	0.972	0.986
	LM	0.950	0.951	0.948	0.948	0.979
	CO	0.978	0.973	0.975	0.969	0.992
	OrdInv	0.955	0.974	0.955	0.973	0.983
	MC	0.966	0.974	0.971	0.972	0.991
	IRER	0.878	0.891	0.889	0.892	0.974
	P	0.973	1.004	0.969	1.010	0.988
	S	0.924	0.961	0.926	0.928	0.979
AR(6)	All series	0.929	0.938	0.935	0.939	0.982
	OI	0.949	0.978	0.960	0.972	0.983
	LM	0.953	0.961	0.951	0.959	0.978
	CO	0.956	0.954	0.955	0.952	0.989
	OrdInv	0.926	0.953	0.935	0.948	0.983
	MC	0.948	0.957	0.960	0.974	0.990
	IRER	0.851	0.854	0.872	0.870	0.975
	P	0.921	0.940	0.939	0.914	0.985
	S	0.963	0.957	0.961	0.959	0.987

Note: The table reports the average of the ratio of the respective forecasts' MSFE over that of the forecasts resulting from the sup-Wald test of Andrews (1993) at $\alpha = 0.05$. Forecasts for which none of the tests indicate a break are excluded. Results are reported for the test statistic W in (17) and S in (28). 'Post-break' and 'Shrinkage' indicate that under the alternative the post-break forecast, respectively the shrinkage forecast (32), are used. The acronyms in the first column with corresponding series after excluding series without breaks (AR(1)|AR(6)): OI: output and income (16|17 series), LM: labor market (28|29), CO: consumption and orders (10|10), OrdInv: orders and inventories (11|11), MC: money and credit (2|8), IRER: interest rates and exchange rates (17|21), P: prices (2|6), S: stock market (4|4).

that, while more precise than post-break forecasts based on the same test, are clearly dominated by the W^s and S^s tests. In fact, for all categories and both models the W^s test leads to more accurate forecasts, as does the S^s tests with the exception of the AR(1) and prices.

7 Conclusion

In this paper, we formalize the notion that ignoring small breaks may improve the accuracy of forecasts. We quantify the break magnitude that leads to equal forecast accuracy between a forecast based on the full sample and one based on a post-break sample. This break magnitude is substantial, which points to a large penalty that is incurred by the uncertainty around the break date. A second finding is that the break magnitude that leads to equal forecast performance depends on the unknown break date.

We derive a test for equal forecast performance. Under a local break, no consistent estimator is available for the break date. Yet, we are able to prove near optimality of our test in the sense that the power of an infeasible test conditional on the break date is achieved for small enough nominal size. This allows the critical values of the test to depend on the estimated break date. We show that under the break magnitudes we consider under our null hypothesis, this optimality is achieved relatively quickly, that is, for finite nominal size. Simulations confirm this and show only a minor loss of power compared to the test that is conditional on the true break date.

We also consider the optimal weights forecast of Pesaran et al. (2013) and show that it is a shrinkage forecast with our test statistic as the shrinkage coefficient. Our test extends in a straightforward way to test whether the shrinkage forecast will be more accurate than the full sample forecast.

We apply the test to a large set of macroeconomic time series and find that breaks that are relevant for forecasting are rare. Pretesting using the test developed here improves over pretesting using the standard test of Andrews (1993) in terms of MSFE. Similar improvements can be made by considering an optimal weights or shrinkage estimator under the alternative.

References

- Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.
- Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414.
- Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.
- Brown, R. L., Durbin, J., and Evan, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B*, 37(2):142–192.
- Clark, T. E. and McCracken, M. W. (2001). Tests of equal forecast accuracy

- and encompassing for nested models. *Journal of Econometrics*, 105(1):85–110.
- Clark, T. E. and McCracken, M. W. (2012). In-sample tests of predictive ability: A new approach. *Journal of Econometrics*, 170(1):1–14.
- Clark, T. E. and McCracken, M. W. (2013). Advances in forecast evaluation. In Elliott, G. and Timmermann, A., editors, *Handbook of Forecasting*, volume 2, pages 1107–1201. Elsevier.
- Dette, H. and Wied, D. (2016). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society, Series B*, 78(2):371–394.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.
- Elliott, G. and Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics*, 141(2):1196–1218.
- Elliott, G. and Müller, U. K. (2014). Pre and post break parameter inference. *Journal of Econometrics*, 180(2):141–157.
- Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83(2):771–811.
- Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies*, 76(2):669–705.
- Hansen, B. E. (2009). Averaging estimators for regressions with a possible structural break. *Econometric Theory*, 25(6):1489–1514.
- Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.
- Hüsler, J. (1990). Extreme values and high boundary crossings of locally stationary Gaussian processes. *Annals of Probability*, 18(3):1141–1158.
- Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician*, 54(3):217–224.
- MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.
- Magnus, J. R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal*, 5(1):225–236.

- McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.
- Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.
- Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.
- Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129(1):183–217.
- Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.
- Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, volume 148. American Mathematical Soc.
- Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40(2):307–318.
- Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate U.S. stock returns. *Journal of Financial Econometrics*, 4(2):238–274.
- Rossi, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic Dynamics*, 10(1):20–38.
- Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.
- Thompson, J. R. (1968). Some shrinkage techniques for estimating the mean. *Journal of the American Statistical Association*, 63(321):113–122.
- Toro-Vizcarrondo, C. and Wallace, T. D. (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association*, 63(322):558–572.
- Trenkler, G. and Toutenburg, H. (1992). Pre-test procedures and forecasting in the regression model under restrictions. *Journal of Statistical Planning and Inference*, 30(2):249–256.
- Wallace, T. D. (1972). Weaker criteria and tests for linear restrictions in regression. *Econometrica*, 40(4):689–698.

Appendix A Additional mathematical details

A.1 A break of known timing

Forecasts are obtained using (9)

$$\hat{y}_{T+h} = f_{T+h}(\hat{\beta}_2, \hat{\delta} | \mathcal{I}_T)$$

where the information set \mathcal{I}_T contains the regressors required for the forecast.

For a known break date, the results of the previous section imply the following asymptotic distribution of the parameters

$$\sqrt{T} \begin{pmatrix} \hat{\beta}_1 - \beta_1 \\ \hat{\beta}_2 - \beta_2 \\ \hat{\delta} - \delta \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{1}{\tau} \mathbf{V} + \tilde{\mathbf{H}} & \tilde{\mathbf{H}} & -\mathbf{L} \\ \tilde{\mathbf{H}} & \frac{1}{1-\tau} \mathbf{V} + \tilde{\mathbf{H}} & -\mathbf{L} \\ -\mathbf{L}' & -\mathbf{L}' & \mathbf{H}^{-1} \end{pmatrix} \right] \quad (36)$$

For the full sample estimator we have

$$\sqrt{T} \begin{pmatrix} \hat{\beta}_F - \beta_2 \\ \hat{\delta} - \delta \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} \tau_b(\beta_1 - \beta_2) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{V} + \tilde{\mathbf{H}} & -\mathbf{L} \\ -\mathbf{L}' & \mathbf{H}^{-1} \end{pmatrix} \right] \quad (37)$$

and

$$\hat{\beta}_F - [\hat{\beta}_2 + \tau_b(\hat{\beta}_1 - \hat{\beta}_2)] \xrightarrow{p} \mathbf{0}$$

Define $\mathbf{f}_{\beta_2} = \frac{\partial f_{T+h}(\beta_2, \delta | \mathcal{I}_T)}{\partial \beta_2}$ and $\mathbf{f}_{\delta} = \frac{\partial f_{T+h}(\beta_2, \delta | \mathcal{I}_T)}{\partial \delta}$. Using a first order Taylor expansion, (36) and (37), we have that

$$\begin{aligned} \sqrt{T} \left(f_{T+h}(\hat{\beta}_2, \hat{\delta} | \mathcal{I}_T) - f_{T+h}(\beta_2, \delta | \mathcal{I}_T) \right) &= \sqrt{T} \left[\mathbf{f}'_{\beta_2}(\hat{\beta}_2 - \beta_2) + \mathbf{f}'_{\delta}(\hat{\delta} - \delta) + O(T^{-1}) \right] \\ &\xrightarrow{d} N(0, \Sigma_{\beta_2} + \Sigma_r) \\ \sqrt{T} \left(f_{T+h}(\hat{\beta}_F, \hat{\delta} | \mathcal{I}_T) - f_{T+h}(\beta_2, \delta | \mathcal{I}_T) \right) &= \sqrt{T} \left[\mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2) + \mathbf{f}'_{\delta}(\hat{\delta} - \delta) + O(T^{-1}) \right] \\ &\xrightarrow{d} N(\tau_b \mathbf{f}'_{\beta_2}(\beta_1 - \beta_2), \Sigma_{\beta_F} + \Sigma_r) \end{aligned}$$

where

$$\begin{aligned} \Sigma_{\beta_i} &= \text{plim}_{T \rightarrow \infty} T \mathbf{f}'_{\beta_2} \text{Var}(\hat{\beta}_i) \mathbf{f}_{\beta_2}, \quad \text{for } i = 2, F \\ \Sigma_r &= \text{plim}_{T \rightarrow \infty} T \left(\mathbf{f}'_{\delta} \text{Var}(\hat{\delta}) \mathbf{f}_{\delta} + 2 \mathbf{f}'_{\beta_2} \text{Cov}(\hat{\beta}_F, \hat{\delta}) \mathbf{f}_{\delta} \right) \end{aligned} \quad (38)$$

and we use that, asymptotically, $T \left(\text{Cov}(\hat{\beta}_F, \hat{\delta}) - \text{Cov}(\hat{\beta}_2, \hat{\delta}) \right) \xrightarrow{p} \mathbf{0}$. Using previous results on the covariance matrix of the estimators, and the notation in (11), we have

$$\begin{aligned} \Sigma_{\beta_2} &= \frac{1}{1 - \tau_b} \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} + \mathbf{f}'_{\beta_2} \tilde{\mathbf{H}} \mathbf{f}_{\beta_2} \\ \Sigma_{\beta_F} &= \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} + \mathbf{f}'_{\beta_2} \tilde{\mathbf{H}} \mathbf{f}_{\beta_2} \end{aligned}$$

For the expected MSFEs using β_2 and β_F , we have

$$\begin{aligned} \lim_{T \rightarrow \infty} TE \left[\left(f_{T+h}(\hat{\beta}_2, \hat{\delta} | \mathcal{I}_T) - f_{T+h}(\beta_2, \delta | \mathcal{I}_T) \right)^2 \right] &= \frac{1}{1 - \tau_b} \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} + \mathbf{f}'_{\beta_2} \tilde{\mathbf{H}} \mathbf{f}_{\beta_2} + \Sigma_r \\ \lim_{T \rightarrow \infty} TE \left[\left(f_{T+h}(\hat{\beta}_F, \hat{\delta} | \mathcal{I}_T) - f_{T+h}(\beta_2, \delta | \mathcal{I}_T) \right)^2 \right] &= [\tau_b \mathbf{f}'_{\beta_2} (\beta_1 - \beta_2)]^2 \\ &\quad + \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} + \mathbf{f}'_{\beta_2} \tilde{\mathbf{H}} \mathbf{f}_{\beta_2} + \Sigma_r \end{aligned}$$

Hence, the full sample based forecast improves over the post-break sample based forecast if

$$\zeta = T(1 - \tau_b) \tau_b \frac{[\mathbf{f}'_{\beta_2} (\beta_1 - \beta_2)]^2}{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}} \leq 1 \quad (39)$$

This reiterates that the null hypothesis of equal mean squared forecast error translates into a hypothesis on the standardized break magnitude, ζ .

Similar to Section 2, a test for $H_0 : \zeta = 1$ can be derived by noting that, asymptotically, $T\text{Var}(\hat{\beta}_1 - \hat{\beta}_2) \xrightarrow{p} \frac{1}{\tau_b(1-\tau_b)} \mathbf{V}$ and, therefore,

$$\hat{\zeta} = T(1 - \tau_b) \tau_b \frac{[\mathbf{f}'_{\beta_2} (\hat{\beta}_1 - \hat{\beta}_2)]^2}{\hat{\omega}} \xrightarrow{d} \chi^2(1, \zeta) \quad (40)$$

where $\hat{\omega}$ is a consistent estimator of $\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}$. The test statistic, $\hat{\zeta}$, can be compared against the critical values of the $\chi^2(1, 1)$ distribution to test for equal forecast performance.

The above can be immediately applied to the simple structural break model (1) where $f_{T+1}(\hat{\beta}_2; \mathbf{x}_{T+1}) = \mathbf{x}'_{T+1} \hat{\beta}_2$, and $\mathbf{f}_{\beta_2} = \mathbf{x}_{T+1}$. The full sample forecast is more accurate if

$$\zeta = T \tau_b (1 - \tau_b) \frac{[\mathbf{x}'_{T+1} (\beta_1 - \beta_2)]^2}{\mathbf{x}'_{T+1} \mathbf{V} \mathbf{x}_{T+1}} \leq 1 \quad (41)$$

identical to the result in (4).

A.2 Proof of Lemma 1

Define $\Delta(\hat{\tau}) = \Delta_1 - \Delta_2$ where

$$\begin{aligned} \Delta_1 &= \lim_{T \rightarrow \infty} TE \left[\left(\mathbf{f}'_{\beta_2} (\hat{\beta}_2(\hat{\tau}) - \beta_2) + \mathbf{f}'_{\beta_2} (\hat{\delta} - \delta) \right)^2 \right] / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \\ &= \lim_{T \rightarrow \infty} TE \left[\left(\mathbf{f}'_{\beta_2} (\hat{\beta}_2(\hat{\tau}) - \beta_2) \right)^2 + \left(\mathbf{f}'_{\beta_2} (\hat{\delta} - \delta) \right)^2 + \right. \\ &\quad \left. + 2 \mathbf{f}'_{\beta_2} (\hat{\beta}_2(\hat{\tau}) - \beta_2) \mathbf{f}'_{\beta_2} (\hat{\delta} - \delta) \right] / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \end{aligned} \quad (42)$$

and similarly for Δ_2

$$\begin{aligned}\Delta_2 &= \lim_{T \rightarrow \infty} TE \left[\left(\mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2) + \mathbf{f}'_{\beta_2}(\hat{\delta} - \delta) \right)^2 \right] / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \\ &= \lim_{T \rightarrow \infty} TE \left[\left(\mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2) \right)^2 + \left(\mathbf{f}'_{\beta_2}(\hat{\delta} - \delta) \right)^2 + \right. \\ &\quad \left. + 2\mathbf{f}'_{\beta_2}(\hat{\beta}_F - \beta_2)\mathbf{f}'_{\beta_2}(\hat{\delta} - \delta) \right] / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}\end{aligned}\quad (43)$$

To prove the theorem, we need that

$$\lim_{T \rightarrow \infty} TE \left[\mathbf{f}'_{\beta_2}(\hat{\beta}_2(\hat{\tau}) - \hat{\beta}_F)\mathbf{f}'_{\beta_2}(\hat{\delta} - \delta) \right] / \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} = 0$$

Define

$$\begin{aligned}X(\tau) &= \mathbf{f}'_{\beta_2}(\hat{\beta}_2(\tau) - \hat{\beta}_F) / \sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}} \\ Y &= \mathbf{f}'_{\beta_2}(\hat{\delta} - \delta) / \sqrt{\mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2}}\end{aligned}$$

Note that $X_\tau^2 = \hat{\zeta}(\tau)$, so that $\hat{\tau}$ is found by maximizing X_τ^2 . We know that for given τ , asymptotically these are jointly normally distributed. It is easy to show that

$$E[X(\tau)Y] = 0$$

for any given τ . Together with the joint normality of $X(\tau)$ and Y , this implies independence between $X(\tau)$ and Y for given τ , i.e. $X(\tau) \perp Y$.

However, we need to prove

$$X(\hat{\tau}) \perp Y, \quad \hat{\tau} = \arg \sup_{\tau \in \Pi} X(\tau)^2$$

Denote

$$g(X(\tau)) = \sup_{\tau \in \Pi} X(\tau), \quad h(X(\tau)) = \inf_{\tau \in \Pi} X(\tau)$$

Since $X(\tau)$ is a stochastic process with continuous sample paths, $g(\cdot)$ and $h(\cdot)$ are measurable functions of $X(\tau)$, which implies

$$g(X(\tau)) \perp Y, \quad h(X(\tau)) \perp Y$$

In terms of $g(\cdot)$ and $h(\cdot)$ we can write

$$X(\hat{\tau}) = f(g(\cdot), h(\cdot)) = g(\cdot) + [h(\cdot) - g(\cdot)]I[g(\cdot) + h(\cdot) \leq 0]$$

with $I[\cdot]$ the indicator function. Now $g(X(\tau))$ and $h(X(\tau))$ are measurable functions of $X(\tau)$ and $f(g(\cdot), h(\cdot))$ is a measurable function of $g(\cdot), h(\cdot)$. Since compositions of measurable functions are measurable, $X(\hat{\tau})$ is a measurable function of $X(\tau)$ as well. Since $f(g(X(\tau)), h(X(\tau)))$ and Y are independent if $X(\tau)$ and Y are independent and $f(g(X(\tau)), h(X(\tau)))$ is a measurable function of $X(\tau)$, we have that $X(\hat{\tau})$ is independent of Y . Then $E[X(\hat{\tau})Y] = 0$. \blacksquare

A.3 Proof of Theorem 1

To prove that only points in a small neighborhood of the true break date contribute to the probability of exceeding a distant boundary, we require the following preliminaries.

Lemma 2 *Suppose $Z(\tau)$ is a symmetric Gaussian process, i.e. $P(Z(\tau) > u) = P(-Z(\tau) > u)$, then as $u \rightarrow \infty$*

$$\begin{aligned} & \mathbb{P} \left(\sup_{\tau \times c} [Z(\tau) + \mu(\tau; \theta_{\tau_b})] c > u \right) \\ & = \mathbb{P} (Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}) [1 + o(1)] \end{aligned}$$

where $c = \pm 1$, $\tau \in \mathcal{I} = [\tau_{\min}, \tau_{\max}]$, and the supremum is taken jointly over τ and c .

Proof: Consider first $\mu(\tau; \theta_{\tau_b}) > 0$ then

$$\begin{aligned} \mathbb{P}(Z(\tau) + \mu(\tau; \theta_{\tau_b}) > u, \tau \in \mathcal{I}) &= \mathbb{P}(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in \mathcal{I}) \\ \mathbb{P}(-Z(\tau) - \mu(\tau; \theta_{\tau_b}) > u, \tau \in \mathcal{I}) &= \mathbb{P}(Z(\tau) > u + |\mu(\tau; \theta_{\tau_b})|, \tau \in \mathcal{I}) \end{aligned} \quad (44)$$

where $\tau \in \mathcal{I}$ is shorthand notation for “for some $\tau \in \mathcal{I}$ ”. When $\mu(\tau; \theta_{\tau_b}) < 0$ we have

$$\begin{aligned} \mathbb{P}(-Z(\tau) - \mu(\tau; \theta_{\tau_b}) > u, \tau \in \mathcal{I}) &= \mathbb{P}(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in \mathcal{I}) \\ \mathbb{P}(Z(\tau) + \mu(\tau; \theta_{\tau_b}) > u, \tau \in \mathcal{I}) &= \mathbb{P}(Z(\tau) > u + |\mu(\tau; \theta_{\tau_b})|, \tau \in \mathcal{I}) \end{aligned} \quad (45)$$

The bounds in the second lines of (44) and (45) are equal or larger than the bounds in the first lines. It follows from the results below that the crossing probabilities over the larger bounds are negligible compared to the crossing probabilities over the lower bounds. This implies that for any sign of $\mu(\tau; \theta_{\tau_b})$ as $u \rightarrow \infty$

$$\begin{aligned} & \mathbb{P} \left(\sup_{\tau \times c} [Z(\tau) + \mu(\tau; \theta_{\tau_b})] c > u \right) \\ & = \mathbb{P} (Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}) [1 + o(1)] \end{aligned} \quad (46)$$

as required. ■

In the structural break model, $Z(\tau)$ is a locally stationary Gaussian process with correlation function $r(\tau, \tau + s)$, defined as follows (Hüsler (1990))

Definition 1 (Local stationarity) *A Gaussian process is locally stationary if there exists a continuous function $C(\tau)$ satisfying $0 < C(\tau) < \infty$*

$$\lim_{s \rightarrow 0} \frac{1 - r(\tau, \tau + s)}{|s|^\alpha} = C(\tau) \text{ uniformly in } \tau \geq 0$$

The correlation function can be written as

$$r(\tau, \tau + s) = 1 - C(\tau)|s|^\alpha \text{ as } s \rightarrow 0$$

The standardized Brownian bridge that we encounter in the structural break model is a locally stationary process with $\alpha = 1$ and local covariance function $C(\tau) = \frac{1}{2} \frac{1}{\tau(1-\tau)}$. Since $\tau \in [\tau_{\min}, \tau_{\max}]$ with $0 < \tau_{\min} < \tau_{\max} < 1$, it holds that $0 < C(\tau) < \infty$.

Lemma 3 *Suppose $Z(\tau)$ is a locally stationary process with local covariance function $C(\tau)$ then for $\delta(u) > 0$ if $\delta(u)u^2 \rightarrow \infty$ and $\delta(u) \rightarrow 0$ as $u \rightarrow \infty$*

$$\lim_{u \rightarrow \infty} \mathbb{P} \left(\sup_{[\tau, \tau + \delta(u)]} Z(t) > u \right) = \frac{1}{\sqrt{2\pi}} \delta(u) u \exp \left(-\frac{1}{2} u^2 \right) C(\tau) \quad (47)$$

Proof: see Hüsler (1990).

To prove Theorem 1, we start by noting that for $\tau \in \mathcal{I} = [\tau_{\min}, \tau_{\max}]$

$$\begin{aligned} \mathbb{P} \left(\sup_{\tau \in \mathcal{I}} Q^*(\tau) > u^2 \right) &= \mathbb{P} \left(\sup_{\tau \in \mathcal{I}} \sqrt{Q^*(\tau)} > u \right) \\ &= \mathbb{P} \left(\sup_{\tau \in \mathcal{I}} |Z(\tau) + \mu(\tau; \theta_{\tau_b})| > u \right) \\ &= \mathbb{P} \left(\sup_{\tau \times c} [Z(\tau) + \mu(\tau; \theta_{\tau_b})]c > u \right) \quad \text{with } c = \pm 1 \\ &= \mathbb{P} (Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}) [1 + o(1)] \end{aligned}$$

where the supremum is taken jointly over $\tau \in \mathcal{I}$ and c . The last equality follows from Lemma 2. Now we proceed along the lines of Piterbarg (1996).

Consider a region close to τ_b defined by $\mathcal{I}_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$. In \mathcal{I}_1 , the minimum value of the boundary is given by

$$\underline{b} = \inf_{\tau \in \mathcal{I}_1} [u - |\mu(\tau; \theta_{\tau_b})|] = u - |\mu(\tau_b; \theta_{\tau_b})| \quad (48)$$

and therefore

$$\begin{aligned} \lim_{u \rightarrow \infty} \mathbb{P}_{\mathcal{I}_1} &= \lim_{u \rightarrow \infty} \mathbb{P}(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}_1) \\ &\leq \lim_{u \rightarrow \infty} \mathbb{P}(Z(\tau) > \underline{b} \text{ for some } \tau \in \mathcal{I}_1) \\ &= 2\delta(\underline{b}) \underline{b} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \underline{b}^2 \right) C(\tau_b) \\ &= \frac{2\delta(\underline{b})}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \underline{b}^2 + \log \underline{b} \right) C(\tau_b) \end{aligned}$$

where the third line follows from (47).

Next, define the region outside of \mathcal{I}_1 as $\mathcal{I}_A = \mathcal{I} \setminus \mathcal{I}_1$. Then in \mathcal{I}_A , the minimum value of the boundary is given by

$$\underline{b}_A = u - |\mu(\tau_b + \delta(u); \theta_{\tau_b})| \quad (49)$$

We now expand $-|\mu(\tau_b + \delta(u); \theta_{\tau_b})|$ around $\delta(u) = 0$. Some care must be taken with regard to the difference between approaching τ_b from the left or from the right

$$-|\mu(\tau_b + \delta(u); \theta_{\tau_b})| = -|\mu(\tau_b; \theta_{\tau_b})| + \gamma\delta(u) + O[\delta(u)^2] \quad (50)$$

where $\gamma = \gamma^+ I[\delta(u) > 0] + \gamma^- I[\delta(u) < 0]$, $\gamma^+ = \left. \frac{\partial \mu(\tau; \theta_{\tau_b})}{\partial \tau} \right|_{\tau \downarrow \tau_b}$ and $\gamma^- = \left. \frac{\partial \mu(\tau; \theta_{\tau_b})}{\partial \tau} \right|_{\tau \uparrow \tau_b}$. The important thing to note is that since $\mu(\tau; \theta_{\tau_b})$ achieves a minimum at $\tau = \tau_b$ we have that $\gamma^+ > 0$ and $\gamma^- < 0$, and consequently $\gamma\delta(u) > 0$. Then $\underline{b}_A = \underline{b} + \gamma\delta(u)$ and

$$\begin{aligned} \lim_{u \rightarrow \infty} P_{\mathcal{I}_A} &= \lim_{u \rightarrow \infty} P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}_A) \\ &\leq \lim_{u \rightarrow \infty} P(Z(\tau) > \underline{b}_A \text{ for some } \tau \in \mathcal{I}_A) \\ &\leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\underline{b}^2 - \underline{b}\gamma\delta(u) - \frac{1}{2}\gamma^2\delta(u)^2 + \log(\underline{b} + \gamma\delta(u))\right) \bar{C} \end{aligned} \quad (51)$$

where we define \bar{C} by noting that

$$\sum_{\mathcal{I}_k \in \mathcal{I}_A} C(k\delta(u))\delta(u) \xrightarrow{\delta(u) \rightarrow 0} \int_{\mathcal{I}_A} C(\tau) d\tau \leq \int_{\mathcal{I}} C(\tau) d\tau = \bar{C} < \infty \quad (52)$$

with \mathcal{I}_k representing non-overlapping intervals of width $\delta(u)$ such that $\bigcup_{k=2}^{\infty} \mathcal{I}_k = \mathcal{I}_A$ and $k\delta(u) \in \mathcal{I}_k$

Compare (51) to the probability of a test with a known break date to exceed the critical value

$$P_0 = P(Z(\tau_b) > u - |\mu(\tau_b; \theta_{\tau_b})|) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\underline{b}^2 - \log(\underline{b})\right) \quad (53)$$

where we use that

$$\frac{1}{\sqrt{2\pi}} \int_u^{\infty} \exp\left(-\frac{1}{2}x^2\right) dx \rightarrow \frac{1}{\sqrt{2\pi}u} \exp\left(-\frac{1}{2}u^2\right) \text{ as } u \rightarrow \infty$$

Ignoring the lower order term $-\frac{1}{2}\gamma^2\delta(u)^2 + \log(\underline{b} + \gamma\delta(u))$, equation (51) contains an extra term $\exp(-\underline{b}\gamma\delta(u))$ compared to (53). This term is decreasing as u increases, as we argued above that $\gamma\delta(u) > 0$. Recalling (48), this implies that $P_{\mathcal{I}_A} = o(P_0)$ if

$$\frac{u\delta(u)}{\log u} \rightarrow \infty$$

Then, if

$$\delta(u) = u^{-1} \log^2(u), \quad (54)$$

all intervals outside of \mathcal{I}_1 contribute $o(P_0)$ to the probability of crossing the boundary u . Under (54), we have that for $P_{\mathcal{I}_1}$ as $u \rightarrow \infty$

$$\begin{aligned} P_{\mathcal{I}_1} &\leq P_{\mathcal{I}} \leq P_{\mathcal{I}_1} + P_{\mathcal{I}_A} \\ &\leq P_{\mathcal{I}_1} + o(P_0) \end{aligned}$$

We now only need to note that

$$\begin{aligned} P_{\mathcal{I}_1} &= P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}_1) \\ &\geq P(Z(\tau_b) > u - |\mu(\tau_b; \theta_{\tau_b})|) = P_0 \end{aligned}$$

to conclude that

$$P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in \mathcal{I}) \xrightarrow{u \rightarrow \infty} P_{\mathcal{I}_1}(1 + o(1))$$

which completes the proof. \blacksquare

Note that, in (51), the term $\exp(b\delta(u))^{-\gamma}$ ensures that $P_{\mathcal{I}_A} = o(P_1)$. In the structural break model, we see that (50) is given by $\mu(\tau_b + \delta(u); \theta_{\tau_b}) = \theta_{\tau_b} \sqrt{\tau_b(1-\tau_b)} - \frac{1}{2}\theta_{\tau_b} \frac{1}{\sqrt{\tau_b(1-\tau_b)}}\delta(u) + O[\delta(u)^2]$. It is clear that γ scales linearly with the break magnitude. Therefore, for a sufficiently large break, asymptotic optimality results are expected to extend to the practical case when u is finite. The simulations of asymptotic power presented in Section 5 confirm this.

A.4 Proof of Theorem 2

Within the interval \mathcal{I}_1 , we have $u_- \leq u(\tau_b) \leq u_+$ and $u_- \leq u(\hat{\tau}) \leq u_+$. The lower and upper bounds satisfy

$$\begin{aligned} u_- &= u(\tau_b) - \left| \frac{\partial u(\tau_b)}{\partial \tau_b} \right| \delta(u) + O(\delta(u)^2) \\ &\geq u(\tau_b) - C\delta(u) + O(\delta(u)^2) \\ u_+ &= u(\tau_b) + \left| \frac{\partial u(\tau_b)}{\partial \tau_b} \right| \delta(u) + O(\delta(u)^2) \\ &\leq u(\tau_b) + C\delta(u) + O(\delta(u)^2) \end{aligned} \quad (55)$$

where $C < \infty$ and we used Assumption 3. Then

$$\begin{aligned} \epsilon &= P(\sup_{\tau} Q^*(\tau) > u_-^2) - P(\sup_{\tau} Q^*(\tau) > u_+^2) \\ &= \frac{1}{\sqrt{2\pi}} \delta(u) u(\tau_b) \exp\left(-\frac{1}{2}u(\tau_b)^2\right) [\exp(-C\delta(u)) - \exp(+C\delta(u))] C(\tau_b) + o(\cdot) \\ &\rightarrow 0 \end{aligned} \quad (56)$$

where $o(\cdot)$ contains lower order terms and the last line uses $\delta(u) = u^{-1} \log^2(u)$, which was shown in Theorem 1. Since

$$\begin{aligned} P\left(\sup_{\tau} Q^*(\tau) > u_+^2\right) &\leq P\left(\sup_{\tau} Q^*(\tau) > u(\tau_b)^2\right) \leq P\left(\sup_{\tau} Q^*(\tau) > u_-^2\right) \\ P\left(\sup_{\tau} Q^*(\tau) > u_+^2\right) &\leq P\left(\sup_{\tau} Q^*(\tau) > u(\hat{\tau})^2\right) \leq P\left(\sup_{\tau} Q^*(\tau) > u_-^2\right) \end{aligned}$$

(56) implies that $P(\sup_{\tau} Q^*(\tau) > u(\tau_b)^2) = P(\sup_{\tau} Q^*(\tau) > u(\hat{\tau})^2)$. ■

A.5 Proof of Theorem 3

To prove Theorem 3, we require the following lemma

Lemma 4 (Convergence of critical values) *Let $u(\tau_b)$ be the critical value that controls size when a break occurs at τ_b and (17) is used as a test statistic. Let $v(\tau_b)$ be the critical value when using the test statistic with $\tau = \tau_b$, then $u(\tau_b) - v(\tau_b) \rightarrow 0$.*

Proof: By definition of the critical values

$$\begin{aligned} P\left[\sup_{\tau} Q^*(\tau) > u(\tau_b)^2\right] &= P[Z(\tau) > u(\tau_b) - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in \mathcal{I}_1] = \alpha \\ P\left[Q^*(\tau_b) > v(\tau_b)^2\right] &= P[Z(\tau_b) > v(\tau_b) - |\mu(\tau_b; \theta_{\tau_b})|] = \alpha \end{aligned}$$

Since τ in the first line is contained in \mathcal{I}_1 , we have by a Taylor series expansion of $\mu(\tau; \theta_{\tau_b})$ around τ_b that $\max |\mu(\tau; \theta_{\tau_b})| - |\mu(\tau_b; \theta_{\tau_b})| = O[\delta(u)]$ and consequently, $\max u(\tau_b) - v(\tau_b) = O(\delta(u))$. Since $\delta(u) \rightarrow 0$ as $u \rightarrow \infty$, the difference in the critical values $u(\tau_b) - v(\tau_b) \rightarrow 0$ as $u \rightarrow \infty$. ■

A proof of Theorem 3 readily follows. With $\hat{\tau}$ from (20) we have

$$P_{H_a} \left[\sup_{\tau} Q^*(\tau) > u(\hat{\tau})^2 \right] = P_{H_a} [Z(\hat{\tau}) > u(\hat{\tau}) - \mu(\hat{\tau}; \theta_{\tau_b})]$$

Under the slowly varying assumption, $u(\hat{\tau}) - \mu(\hat{\tau}; \theta_{\tau_b})$ has a unique minimum on \mathcal{I}_1 at $\hat{\tau} = \tau_b$. Taking the supremum therefore necessarily leads to at least as many exceedances as considering $\tau = \tau_b$ alone, which proves the inequality in (27). The last line of (27) follows from Lemma 4. ■

A.6 Proof of Corollary 1

The test statistic converges to $S(\hat{\tau}) \rightarrow \sup_{\tau} |Z(\tau) + \mu(\tau; \theta_{\tau_b})| - |\mu(\hat{\tau}; \theta_{\hat{\tau}})|$ where $\hat{\tau}$ maximizes the first term. As shown before, exceedances of a high

boundary are concentrated in the region $[\tau_b - \delta(u), \tau_b + \delta(u)]$ where $\delta(u) \rightarrow 0$ as $u \rightarrow \infty$. Then

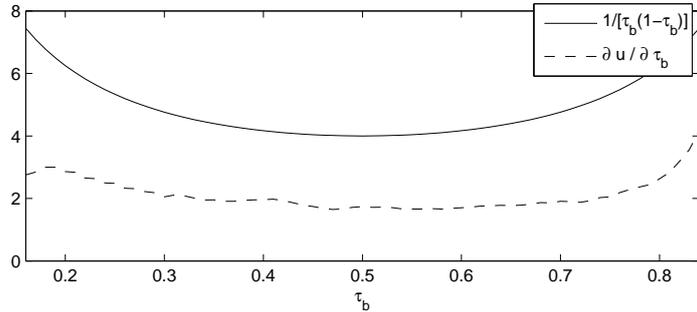
$$\begin{aligned} \lim_{u \rightarrow \infty} \mathbb{P}(S(\hat{\tau}) > u) &= \lim_{u \rightarrow \infty} \mathbb{P} \left(\sup_{\mathcal{I}_1} |Z(\tau) + \mu(\tau; \theta_{\tau_b})| - |\mu(\hat{\tau}; \theta_{\hat{\tau}})| > u \right) \\ &= \lim_{u \rightarrow \infty} \mathbb{P}(Z(\hat{\tau}) > u - |\mu(\hat{\tau}; \theta_{\tau_b})| + |\mu(\hat{\tau}; \theta_{\hat{\tau}})|) \end{aligned}$$

Under the slowly varying assumption, the difference $-|\mu(\hat{\tau}; \theta_{\tau_b})| + |\mu(\hat{\tau}; \theta_{\hat{\tau}})| = O[\delta(u)]$. This implies that the critical values of $S(\hat{\tau})$ are independent of τ_b in the limit where $u \rightarrow \infty$. ■

A.7 Verifying condition (24)

In order to verify that (24) holds, that is, that the condition for near optimality, $\partial u(\tau_b)/\partial \tau_b < 1/[\tau_b(1 - \tau_b)]$, holds. Observe that, in Figure 10, the dashed line, which depicts the derivative of the critical values for $\alpha = 0.05$ as a function of the break date τ_b and is obtained via simulation, is clearly below the solid line, which depicts the upper bound $[\tau_b(1 - \tau_b)]^{-1}$.

Figure 10: Dependence of the critical values on the break date

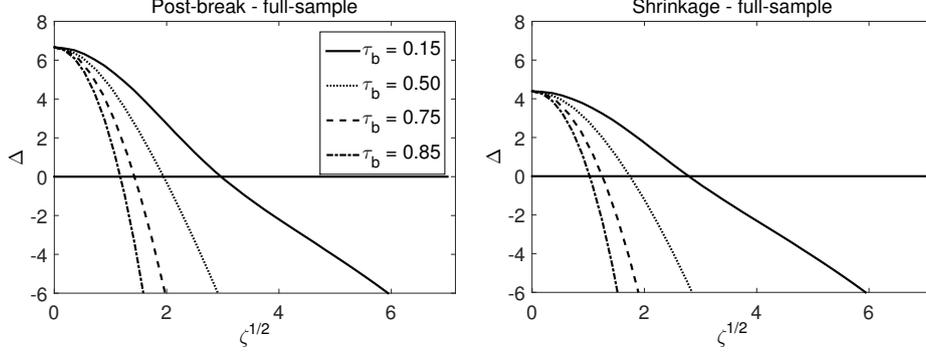


Note: The dashed line depicts the derivative of the critical values for $\alpha = 0.05$ as a function of the break date τ_b . The solid line depicting the upper bound $[\tau_b(1 - \tau_b)]^{-1}$.

A.8 Uniqueness of the break magnitude that yields equal forecast accuracy

In order to ensure the uniqueness of the break magnitude that leads to equal forecast accuracy, we evaluate Δ in (22) and Δ_s in (33) numerically using the simulation set-up described in Section 5. The results in Figure 11 show that the value of $|\theta_{\tau_b}|$ that leads to equal forecast accuracy is unique.

Figure 11: Difference in asymptotic MSFEs, Δ and Δ_s



Note: The left panel shows the difference in the asymptotic MSFE between the post-break forecast and the full-sample forecast as a function of the standardized break magnitude $\zeta^{1/2}$ in (22) for $\tau_b = \{0.15, 0.50, 0.75, 0.85\}$. The right panel shows the difference in MSFE between the shrinkage forecast and the full-sample forecast in (33).

A.9 Derivation of equation (29)

From a Taylor series expansion it follows that

$$\begin{aligned} \mathbb{E} \left[T \left(\hat{y}_{T+h}^S - f_{T+h}(\beta_2) \right)^2 \right] &= \mathbb{E} \left[T \left(\omega \mathbf{f}'_{\beta_2} \hat{\beta}_1 + (1 - \omega) \mathbf{f}'_{\beta_2} \hat{\beta}_2 - \mathbf{f}'_{\beta_2} \beta_2 \right)^2 \right] + o(1) \\ &= \omega^2 \mathbb{E} \left[T \left(\mathbf{f}'_{\beta_2} (\hat{\beta}_1 - \hat{\beta}_2) \right)^2 \right] + \frac{1}{\tau_b} \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \\ &\quad + 2\omega \mathbf{f}'_{\beta_2} \mathbb{E} \left[T \left(\hat{\beta}_1 - \hat{\beta}_2 \right) \left(\hat{\beta}_2 - \beta_2 \right) \right] \mathbf{f}_{\beta_2} + o(1) \end{aligned}$$

We analyze the first and third term of the second equality separately.

Using a bias-variance decomposition, the expectation in the first term can be calculated as

$$\begin{aligned} \mathbb{E} \left[T \left(\mathbf{f}'_{\beta_2} (\hat{\beta}_1 - \hat{\beta}_2) \right)^2 \right] &= \mathbb{E} \left[T \left(\mathbf{f}'_{\beta_2} (\hat{\beta}_1 - \hat{\beta}_2) \right) \right]^2 + T \text{Var} \left[\mathbf{f}'_{\beta_2} (\hat{\beta}_1 - \hat{\beta}_2) \right] \\ &= T \left(\mathbf{f}'_{\beta_2} (\beta_1 - \beta_2) \right)^2 + \mathbf{f}'_{\beta_2} \left(\frac{1}{\tau_b} + \frac{1}{1 - \tau_b} \right) \mathbf{V} \mathbf{f}_{\beta_2} \end{aligned}$$

since $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = 0$.

The term linear in ω is given by

$$\begin{aligned} \mathbf{f}'_{\beta_2} \mathbb{E} \left[T \left(\hat{\beta}_1 - \hat{\beta}_2 \right) \left(\hat{\beta}_2 - \beta_2 \right) \right] \mathbf{f}_{\beta_2} &= -\mathbf{f}'_{\beta_2} \mathbb{E} \left[T \left(\beta_1 - \beta_2 \right) \beta_2' \right] \mathbf{f}_{\beta_2} \\ &\quad + \mathbf{f}'_{\beta_2} \mathbb{E} \left[T \hat{\beta}_1 \hat{\beta}_2' - \hat{\beta}_2 \hat{\beta}_2' \right] \mathbf{f}_{\beta_2} \\ &= -\frac{1}{1 - \tau_b} \mathbf{f}'_{\beta_2} \mathbf{V} \mathbf{f}_{\beta_2} \end{aligned}$$

Using these two expressions yields (29).

Appendix B Tables with critical values

Tables 7–8 contain critical values when the break is in the range $\tau_b = 0.15$ to 0.85, where Table 7 considers post-break sample and full sample based forecasts and Table 8 considers shrinkage forecast and full sample based forecasts. Tables 9–10 contain the critical values when the break can be in the range $\tau_b = 0.05$ to 0.95 for the same comparisons.

Table 7: Post-break versus full sample: critical values and size

τ_b	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
$\zeta^{1/2}$	2.99	2.73	2.55	2.41	2.28	2.17	2.06	1.95	1.84	1.75	1.64	1.54	1.43	1.31	1.18
Wald test statistic (17)															
0.10	20.44	19.16	17.99	17.05	16.22	15.49	14.79	14.13	13.49	12.91	12.30	11.68	11.04	10.32	9.36
0.05	23.71	22.29	20.99	19.95	19.04	18.24	17.46	16.74	16.03	15.38	14.71	14.02	13.30	12.48	11.37
0.01	30.54	28.84	27.29	26.07	25.00	24.06	23.15	22.29	21.46	20.70	19.89	19.08	18.22	17.23	15.82
0.10	0.13	0.13	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.10	0.10	0.09	0.09	0.08	0.06
0.05	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.03
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S test statistic (28)															
0.10	1.78	1.82	1.84	1.86	1.87	1.88	1.88	1.89	1.89	1.88	1.87	1.86	1.83	1.80	1.59
0.05	2.12	2.16	2.18	2.20	2.21	2.22	2.22	2.23	2.23	2.22	2.22	2.20	2.18	2.14	1.94
0.01	2.76	2.79	2.81	2.83	2.85	2.86	2.86	2.87	2.87	2.86	2.86	2.85	2.83	2.80	2.60
0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.10	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Note: The table reports critical values and size for the W and S test statistics that test the null hypothesis of equal MSFE of the post-break and full sample forecasts. For additional information, see the footnote of Table 1.

Table 8: Shrinkage versus full sample: critical values and size

τ_b	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
$\zeta^{1/2}$	2.81	2.54	2.36	2.21	2.08	1.97	1.85	1.75	1.64	1.55	1.45	1.35	1.25	1.15	1.03
Wald test statistic (17)															
0.10	19.01	17.78	16.63	15.71	14.92	14.22	13.56	12.95	12.34	11.81	11.28	10.74	10.19	9.58	8.82
0.05	22.15	20.78	19.51	18.48	17.60	16.84	16.10	15.43	14.76	14.16	13.57	12.97	12.34	11.64	10.74
0.01	28.74	27.08	25.57	24.35	23.30	22.40	21.53	20.74	19.95	19.23	18.53	17.81	17.03	16.18	15.02
0.10	0.14	0.13	0.13	0.12	0.12	0.12	0.11	0.11	0.11	0.10	0.10	0.09	0.08	0.07	0.06
0.05	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03
0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S test statistic (28)															
0.10	1.85	1.88	1.90	1.92	1.93	1.93	1.93	1.93	1.91	1.90	1.88	1.86	1.82	1.76	1.63
0.05	2.18	2.22	2.24	2.25	2.26	2.27	2.27	2.27	2.26	2.25	2.23	2.20	2.17	2.11	1.98
0.01	2.82	2.85	2.87	2.89	2.90	2.90	2.91	2.91	2.90	2.89	2.87	2.85	2.82	2.76	2.63
0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.10	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Note: The table reports critical values and size for the W and S test statistics that test the null hypothesis of equal MSFE of the shrinkage forecast (32) and the full sample forecast. For additional information, see the footnote of Table 2.

Table 9: Post-break versus full sample: critical values and size when searching [0.05, 0.95]

	τ_b	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
$\zeta^{1/2}$	4.17	3.69	3.44	3.25	3.10	2.97	2.85	2.73	2.63	2.53	2.43	2.32	2.21	2.10	1.97	1.84	1.68	1.48	1.19		
Wald test statistic (17)																					
0.10	31.31	27.59	25.36	23.74	22.49	21.43	20.49	19.59	18.85	18.11	17.38	16.65	15.90	15.16	14.39	13.56	12.62	11.52	9.82		
0.05	35.41	31.37	28.96	27.22	25.85	24.69	23.69	22.70	21.88	21.07	20.27	19.47	18.64	17.83	16.97	16.04	15.00	13.74	11.75		
0.01	43.80	39.15	36.41	34.41	32.87	31.54	30.37	29.23	28.29	27.33	26.38	25.44	24.47	23.52	22.51	21.40	20.15	18.60	16.04		
0.10	0.11	0.11	0.11	0.11	0.12	0.11	0.12	0.11	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.10	0.10	0.08	0.05		
0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.03		
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	

S test statistic (28)

0.10	1.55	1.64	1.68	1.72	1.74	1.74	1.77	1.79	1.80	1.82	1.83	1.85	1.86	1.86	1.87	1.87	1.86	1.84	1.79	1.56
0.05	1.90	1.98	2.03	2.06	2.08	2.10	2.13	2.14	2.14	2.16	2.17	2.18	2.19	2.20	2.21	2.21	2.20	2.19	2.14	1.91
0.01	2.55	2.63	2.66	2.70	2.72	2.74	2.76	2.77	2.77	2.79	2.80	2.82	2.83	2.84	2.84	2.85	2.85	2.83	2.79	2.57
0.10	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Note: The table reports critical values and size for the W and S test statistics that test the null hypothesis of equal MSFE of the post-break and full sample forecasts. For additional information, see the footnote Table 1.

Table 10: Shrinkage versus full sample: critical values and size when searching [0.05, 0.95]

	τ_b	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	
$\zeta^{1/2}$	4.04	3.56	3.30	3.10	2.95	2.81	2.68	2.56	2.46	2.35	2.24	2.13	2.02	1.90	1.78	1.65	1.49	1.31	1.05	0.95	
Wald test statistic (17)																					
0.10	30.00	26.40	24.15	22.47	21.24	20.17	19.17	18.33	17.58	16.84	16.14	15.43	14.70	14.01	13.29	12.54	11.70	10.76	9.40	9.40	
0.05	34.01	30.09	27.65	25.85	24.49	23.32	22.25	21.31	20.49	19.67	18.90	18.11	17.32	16.55	15.74	14.90	13.95	12.86	11.24	11.24	
0.01	42.23	37.71	34.93	32.84	31.31	29.96	28.71	27.61	26.66	25.70	24.77	23.84	22.89	21.99	21.04	20.01	18.85	17.48	15.35	15.35	
0.10	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11	0.10	0.10	0.09	0.08	0.05	
0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.02	
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	

S test statistic (28)

0.10	1.57	1.67	1.73	1.76	1.79	1.82	1.84	1.85	1.87	1.88	1.90	1.90	1.91	1.91	1.91	1.91	1.89	1.87	1.81	1.59
0.05	1.92	2.02	2.07	2.10	2.13	2.16	2.17	2.19	2.21	2.22	2.23	2.24	2.24	2.24	2.25	2.24	2.24	2.21	2.16	1.95
0.01	2.57	2.66	2.70	2.74	2.77	2.79	2.81	2.82	2.84	2.85	2.86	2.87	2.88	2.88	2.88	2.88	2.88	2.85	2.81	2.60
0.10	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Note: The table reports critical values and size for the W and S test statistics that test the null hypothesis of equal MSFE of the shrinkage forecast (32) and the full sample forecast. For additional information, see the footnote of Table 2.