Does modeling a structural break improve forecast accuracy?

Tom Boot[*] Andreas Pick[†]

August 22, 2018

Abstract

Mean square forecast error loss implies a bias-variance trade-off that suggests that structural breaks of small magnitude should be ignored. In this paper, we provide a test to determine whether modeling a break improves forecast accuracy. The test is near optimal even when the date of a local-to-zero break is not consistently estimable. The results extend to forecast combinations that weight the post-break sample and the full sample forecasts by our test statistic. In a large number of macroeconomic time series, we find that structural breaks that are relevant for forecasting occur much less frequently than existing tests indicate.

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 an anonymous referee, the co-editor, Oliver Linton, Jeremy Chiu, 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 BGSE summer forum, ECB workshop on forecasting techniques, ESEM, IAAE annual conference, NESG meeting, RMSE workshop, and SNDE conference for helpful comments. We thank SURFsara for access to the Lisa Computer Cluster. The opinions expressed are those of the authors and do not necessarily reflect the views of DNB or the European System of Central Banks. The paper previously circulated under the title "A near optimal test for structural breaks when forecasting under square error loss."

1 Introduction

Many macroeconomic and financial time series contain structural breaks as documented by Stock and Watson (1996). Yet, Stock and Watson also find that forecasts are not substantially affected by the presence of structural breaks. Gains from modeling a structural break may be offset by imprecisely estimated break dates and post-break parameters, which Elliott and Müller (2007, 2014) show is a feature of local-to-zero breaks. Additionally, forecasts are typically evaluated using mean square forecast error loss, which implies a bias-variance trade-off. Ignoring rather than modeling small breaks may therefore lead 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 real-time test for equal forecast accuracy that compares the expected mean square forecast error (MSFE) of a forecast from the post-break sample to that from the full sample. The difference in MSFE is a standardized, linear combination of pre- and post-break parameters with weights that depend on the regressors of the forecasting model. As a result, breaks in the parameter vector, which are the focus of the extant literature on structural breaks, do not necessarily imply a break in the forecast.

Full sample and post-break sample forecasts achieve equal forecast accuracy at a critical magnitude of the break. The bias-variance trade-off implies that this critical magnitude is non-zero. The null of our test therefore differs from the null of existing tests where the focus is on testing the absence of any parameter instability, such as Ploberger et al. (1989), Andrews (1993), Andrews and Ploberger (1994), and Dufour et al. (1994). Additionally, the critical magnitude of the break under the null depends on the unknown break date, which is not consistently estimable under local breaks. Using results of Andrews (1993) and Piterbarg (1996), we show that our test is optimal as the size of the test tends to zero. We provide evidence that the power of test remains close to that of 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. This 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. Yet, 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 combination of the forecasts based on the post-break sample and the full sample. The relative weight is a function of the test statistic introduced in this paper. This approach is similar to that proposed by Hansen (2009), who minimizes the in-sample mean square error using weights based on the Mallows criterion.

We find that for small break magnitudes, where the break date is not accurately identified, the combined forecast is less accurate than the full sample forecast. However, compared to the post-break sample forecast, we find that the combined forecast is more accurate for a large area of the parameter space. We therefore propose a second version of our test that compares the forecast accuracy of the combined forecast to 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) assess forecast accuracy *ex post* (see Clark and McCracken (2013) for a review). In contrast, the test we propose in this paper is a real-time 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 forecast accuracy 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. Similiary, Giacomini and Rossi (2010) use a pseudo-out-of-sample period to assess competing models in the presence of instability. Our test, in contrast, does not require a pseudo-out-of-sample period but is a real-time test.

Substantial evidence for structural breaks has been found in macroeconomic and financial time series by, for example, Chib and Kang (2013), Pastor and Stambaugh (2001), Paye and Timmermann (2006), Pesaran and Timmermann (2002), Pettenuzzo and Timmermann (2011), Rapach and Wohar (2006), Rossi (2006), and Stock and Watson (1996, 2007). 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 the breaks suggested by Andrews' sup-Wald test into account. Our paper, therefore, provides theoretical support for the finding of Stock and Watson (1996) that many 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 forecast that combines the full-sample and post-break forecasts based on the derived test statistic. 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

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

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

where

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

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

Denote $\boldsymbol{V}_i = (T_i - T_{i-1})\text{Var}(\hat{\boldsymbol{\beta}}_i)$, for $i = 1,2$, $T_0 = 0$, $T_1 = T_b$, $T_2 = T$ and $\boldsymbol{V}_F = T\text{Var}(\hat{\boldsymbol{\beta}}_F)$ 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.

We would like to test whether the expected mean squared forecast error (MSFE) from the h-step ahead forecast using the full sample, $\hat{y}^F_{T+h} = \mathbf{x}'_{T+h}\hat{\boldsymbol{\beta}}_F$, is smaller or equal to that of the post-break sample, $\hat{y}^P_{T+h} = \mathbf{x}'_{T+h}\hat{\boldsymbol{\beta}}_2$. In this motivating example, we consider $h = 1$, and extend the results to the more general case in Section 4. Note that model (1) is constructed for forecasting purposes. The regressors $\boldsymbol{x}_t$ relate to y_t but are available at $t-1$, which implies that $\boldsymbol{x}_{T+1}$ contains regressors that predict y_{T+1} and are available at time T.

The MSFE using the post-break sample estimate, $\hat{\boldsymbol{\beta}}_2$, conditional on $\mathbf{x}_{T+1}$, is

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

where the first term in the second line represents the estimation uncertainty in the shorter post-break sample and the second term the uncertainty of the disturbance term in the forecast period.

Using the full sample estimate, $\hat{\boldsymbol{\beta}}_F$, we have

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

where, in the last line of the equation, the first term is the square bias that arises from estimating the parameter vector over the two sub-periods, the second term represents the estimation uncertainty in the full sample, and the final term is the uncertainty of the disturbance term in the forecast period.

Define $\tau_b = T_b/T$. Comparing (2) and (3), we see that the full sample forecast is at least as accurate as the post-break sample forecast if

$$\zeta = T\tau_b^2 \frac{\left[\boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\boldsymbol{V}_2}{1-\tau_b} - \boldsymbol{V}_F\right)\boldsymbol{x}_{T+1}} \leq 1 \tag{4}$$

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

$$W(\tau_b) = T\tau_b^2 \frac{\left[\boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\boldsymbol{V}_2}{1-\tau_b} - \boldsymbol{V}_F\right)\boldsymbol{x}_{T+1}} \sim \chi^2(1, \zeta) \tag{5}$$

Details are provided in the online appendix. 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.

Under the assumption that the covariance matrices satisfy $\text{plim}_{T\to\infty}\boldsymbol{V}_i = \boldsymbol{V}$ for $i = 1, 2, F$, a more conventional and asymptotically equivalent form of the test statistic is

$$W(\tau_b) = T \frac{\left[\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\boldsymbol{V}_1}{\tau_b} + \frac{\boldsymbol{V}_2}{1-\tau_b}\right)\boldsymbol{x}_{T+1}} \overset{a}{\sim} \chi^2(1, \zeta) \tag{6}$$

which can be recognized as a Wald test statistic with the regressors for the forecast at $T+1$ 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 $\boldsymbol{x}_{T+1}$ 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 $T \to \infty$, 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. $\boldsymbol{\beta}_{\mathbf{2}} = \boldsymbol{\beta}_1 + \frac{1}{\sqrt{T}}\boldsymbol{\eta}$, 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 $\{\boldsymbol{W}_t = (\boldsymbol{Y}_t, \boldsymbol{X}_t) : 1 \leq t \leq T\}$, $\boldsymbol{Y}_t = (y_1, y_2, \ldots y_t)$, and $\boldsymbol{X}_t = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t)'$. Assumptions can be made with regard to the dependence structure of $\boldsymbol{W}_t$ 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 $\boldsymbol{x}_t$.

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

$$y_t = f_t(\boldsymbol{\beta}_t, \boldsymbol{\delta}; \mathcal{I}_{t-h}) + \varepsilon_t \tag{7}$$

where the function f_t is fixed and differentiable with respect to the parameter vector $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t', \boldsymbol{\delta}')'$, and $\mathcal{I}_t$ is the information set at time t that contains $\boldsymbol{X}_{t'}, \boldsymbol{Y}_{t'}$, $1 \leq t' \leq t$.

In (7), the parameter vector $\boldsymbol{\delta}$ is constant for all t but 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,

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

which requires solving

$$\frac{1}{T}\sum_{t=1}^{T} m(\boldsymbol{W}_t, \hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}})'\hat{\gamma}\frac{1}{T}\sum_{t=1}^{T} m(\boldsymbol{W}_t, \hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) = \inf_{\tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}} \frac{1}{T}\sum_{t=1}^{T} m(\boldsymbol{W}_t, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}})'\hat{\gamma}\frac{1}{T}\sum_{t=1}^{T} m(\boldsymbol{W}_t, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\delta}}) \tag{8}$$

where $\hat{\boldsymbol{\beta}}_F$ is the estimator based on the full estimation window. Throughout, we set the weighting matrix $\boldsymbol{\gamma} = \boldsymbol{S}^{-1}$ and $\boldsymbol{S} = \lim_{T\to\infty} \text{Var}\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(\boldsymbol{W}_t, \boldsymbol{\beta}, \boldsymbol{\delta})\right)$, 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 $\boldsymbol{\eta}(\tau) = \boldsymbol{b}\,\mathrm{I}(\tau < \tau_b)$, $\mathrm{I}(A)$ is the indicator function, which is unity if A is true and zero otherwise, $\boldsymbol{b}$ 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

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

Define

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

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

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

Forecasts are constructed as

$$\hat{y}^F_{T+h} = f_{T+h}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) \tag{9}$$
$$\hat{y}^P_{T+h} = f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) \tag{10}$$

where f_{T+h} is the predictive function for y_{T+h} based on the information set $\mathcal{I}_T$ available at time T, which includes any exogenous and lagged dependent variables that are needed to construct the forecast and are available at time T. 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. As the function f_{T+h} can be non-linear in the parameters, iterated forecasts are covered by our analysis. Direct forecasts, in contrast, leads to residual autocorrelation, which can be addressed using a robust covariance matrix (Pesaran et al., 2011). In order not to complicate the notation further, we do not distinguish between the different forecasts for $h > 1$. The comparison between $\hat{y}^F_{T+h}$ and $\hat{y}^P_{T+h}$ is, however, non-standard as, under a local break, even the parameter estimates of the model that incorporates the break may not be unbiased.

Our aim is to determine whether the full sample forecast (9) is more precise in the MSFE sense than the post-break sample forecast (10). 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,

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

To simplify the notation, define $\bar{\boldsymbol{X}}' = \boldsymbol{M}'\boldsymbol{S}^{-1/2}$ and $\bar{\boldsymbol{Z}}' = \boldsymbol{M}_\delta'\boldsymbol{S}^{-1/2}$.

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

$$\begin{aligned}
\sqrt{T}(\hat{\boldsymbol{\beta}}_1(\tau) - \boldsymbol{\beta}_2) \Rightarrow & \frac{1}{\tau}\left[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{B}(\tau) + \int_0^\tau \boldsymbol{\eta}(s)ds\right] \\
& - (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
\sqrt{T}(\hat{\boldsymbol{\beta}}_2(\tau) - \boldsymbol{\beta}_2) \Rightarrow & \frac{1}{1-\tau}\left[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'(\boldsymbol{B}(1) - \boldsymbol{B}(\tau)) + \int_\tau^1 \boldsymbol{\eta}(s)ds\right] \\
& - (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \Rightarrow & (\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1)
\end{aligned} \tag{11}$$

where $\boldsymbol{B}(\tau)$ is a Brownian motion defined on the interval $[0, 1]$ and $\Rightarrow$ denotes weak convergence. Derivations of (11) and (12) below are in the online appendix. We subtract $\boldsymbol{\beta}_2$ from both estimators $\hat{\boldsymbol{\beta}}_1$ and $\hat{\boldsymbol{\beta}}_2$ as our interest is in forecasting future observations, which are functions of $\boldsymbol{\beta}_2$. The remainder that arises if $\tau \neq \tau_b$ is absorbed in the integral on the right hand side. The convergence in (11) occurs jointly.

For estimators that ignore the break, we have

$$\begin{aligned}
\sqrt{T}\left(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2\right) \Rightarrow & (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{B}(1) + \int_0^1 \boldsymbol{\eta}(s)ds \\
& - (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
\sqrt{T}\left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) \Rightarrow & (\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1)
\end{aligned} \tag{12}$$

Note that for the parameters $\hat{\boldsymbol{\delta}}$, 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

$$\operatorname*{plim}_{T\to\infty} T \ \mathrm{Cov}(\hat{\boldsymbol{\beta}}_2(\tau), \hat{\boldsymbol{\beta}}_F) = \boldsymbol{V} + \tilde{\boldsymbol{H}} = \operatorname*{plim}_{T\to\infty} T \ \mathrm{Var}(\hat{\boldsymbol{\beta}}_F)$$

where $\boldsymbol{P}_{\bar{X}} = \bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'$, $\boldsymbol{M}_{\bar{X}} = \boldsymbol{I} - \boldsymbol{P}_{\bar{X}}$, $\boldsymbol{V} = (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}$, $\boldsymbol{H} = \bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}}$, $\boldsymbol{L} = (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}$, and $\tilde{\boldsymbol{H}} = \boldsymbol{L}\boldsymbol{H}\boldsymbol{L}'$. This corresponds to the result by Hausman (1978) that a consistent and asymptotically efficient estimator should have zero covariance with its difference from an consistent but asymptotically inefficient estimator, i.e. $\operatorname{plim}_{T\to\infty} T\mathrm{Cov}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_2(\tau)) = \boldsymbol{0}$. A difference here is that, under a local structural break, $\hat{\boldsymbol{\beta}}_F$ and $\hat{\boldsymbol{\beta}}_2(\tau)$ are both inconsistent.

4 Testing for a break

Using the estimation framework of the previous section, we will generalize the motivating example of Section 2. We will 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

Based on the information set $\mathcal{I}_T$, which contains the regressor set available at time T that is necessary to construct the forecast, the h-step-ahead forecast is

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

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

$$\zeta = (1-\tau_b)\tau_b \frac{\left(\boldsymbol{f}'_{\beta_2}\boldsymbol{b}\right)^2}{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}} = 1 \tag{13}$$

where we assume that the breaks are local-to-zero, and $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = \boldsymbol{b}/\sqrt{T}$. Details of the derivation can be found in the online appendix. 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, $\zeta^{1/2}$.

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

$$W(\tau_b) = T(1-\tau_b)\tau_b \frac{\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\hat{\omega}} \overset{a}{\sim} \chi^2(1,\zeta) \tag{14}$$

where $\hat{\omega}$ is any consistent estimator of $\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}$. The test statistic, $W(\tau_b)$, 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 (14) to test for equal predictive accuracy between a full-sample and post-break forecast. In this section, we adjust the test statistic to a local-to-zero break of unknown date and provide its asymptotic distribution.

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

$$\sup_{\tau\in\mathcal{T}} W(\tau) = \sup_{\tau\in\mathcal{T}} \left\{ T(1-\tau)\tau \frac{\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1(\tau) - \hat{\boldsymbol{\beta}}_2(\tau))\right]^2}{\hat{\omega}} \right\} \tag{15}$$

with $\mathcal{T} = [\tau_{\min}, \tau_{\max}]$. Since the function $\boldsymbol{f}'_{\beta_2}$ in (15) is fixed, the results in Andrews (1993) and the continuous mapping theorem show that, under local alternatives and as $T \to \infty$, $W(\tau)$ in (15) weakly converges to

$$\begin{aligned} Q^*(\tau) &= \left[\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}} + \sqrt{\frac{1-\tau}{\tau}} \int_0^\tau \eta(s)ds - \sqrt{\frac{\tau}{1-\tau}} \int_\tau^1 \eta(s)ds \right]^2 \\ &= [Z(\tau) + \mu(\tau;\theta_{\tau_b})]^2 \end{aligned} \tag{16}$$

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, $\eta(s) = \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}}$, and

$$\mu(\tau;\theta_{\tau_b}) = \theta_{\tau_b} \left[\sqrt{\frac{1-\tau}{\tau}} \tau_b \mathrm{I}(\tau_b < \tau) + \sqrt{\frac{\tau}{1-\tau}} (1-\tau_b) \mathrm{I}(\tau_b \geq \tau) \right] \tag{17}$$

arises when a structural break is present, where $\theta_{\tau_b} = \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{b}}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} = \sqrt{\zeta/(\tau_b(1-\tau_b))}$. 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

$$\hat{\tau} = \arg\sup_{\tau\in\mathcal{T}} W(\tau) \Rightarrow \arg\sup_{\tau\in\mathcal{T}} Q^*(\tau) \tag{18}$$

4.2.1 MSFE under an unknown break date

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

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

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

Lemma 1 *If the break date is estimated using* (18) *then difference in the standardized mean* *squared forecast error,* $\Delta(\tau_b)$*, is*

$$\begin{aligned}\Delta(\tau_b) &= \lim_{T\to\infty} T\left(\mathrm{E}\left\{\left[f_{T+h}(\hat{\boldsymbol{\beta}}_2(\hat{\tau}),\hat{\boldsymbol{\delta}};\mathcal{I}_T) - f_{T+h}(\boldsymbol{\beta}_2,\boldsymbol{\delta};\mathcal{I}_T)\right]^2\right\}\right.\\ &\qquad \left.-\mathrm{E}\left\{\left[f_{T+h}(\hat{\boldsymbol{\beta}}_F,\hat{\boldsymbol{\delta}};\mathcal{I}_T) - f_{T+h}(\boldsymbol{\beta}_2,\boldsymbol{\delta});\mathcal{I}_T)\right]^2\right\}\right)/\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} \qquad (19)\\ &= \lim_{T\to\infty} T\left(\mathrm{E}\left\{\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_2(\hat{\tau})-\boldsymbol{\beta}_2)\right]^2\right\} - \mathrm{E}\left\{\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_2)\right]^2\right\}\right)/\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}\end{aligned}$$

The proof is provided in the online appendix. Lemma 1 shows that the difference in the MSFE is not affected by the estimation of the parameter vector $\boldsymbol{\delta}$, 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 (11) and (12) in (19) yields

$$\begin{aligned}\Delta(\tau_b) &= \mathrm{E}\left\{\left[\frac{1}{1-\hat{\tau}}\frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\bar{\boldsymbol{X}}'\left(\boldsymbol{B}(1)-\boldsymbol{B}(\hat{\tau})\right)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} + \frac{1}{1-\hat{\tau}}\int_{\hat{\tau}}^1 \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}}ds\right]^2\right\}\\ &\qquad -\left(\int_0^1 \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}}ds\right)^2 - 1 \qquad (20)\end{aligned}$$

The results in (20) are valid for a general form of instability $\boldsymbol{\eta}(\tau)$. Define $J(\tau) = \int_\tau^1 (\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2})^{-1/2}\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)ds$ and note that, for fixed $\boldsymbol{f}'_{\beta_2}$,

$$(\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2})^{-1/2}\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\bar{\boldsymbol{X}}'\left[\boldsymbol{B}(1)-\boldsymbol{B}(\hat{\tau})\right] = B(1)-B(\hat{\tau})$$

where $B(\cdot)$ is a one-dimensional Brownian motion. Then

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

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 local-to-zero structural break $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = \boldsymbol{b}/\sqrt{T}$ and $\boldsymbol{\eta}(\tau) = \boldsymbol{b}\mathrm{I}(\tau < \tau_b)$,

$$\Delta(\tau_b) = \mathrm{E}\left\{\left[\frac{1}{1-\hat{\tau}}\left(B(1)-B(\hat{\tau})\right) + \theta_{\tau_b}\frac{\tau_b-\hat{\tau}}{1-\hat{\tau}}\mathrm{I}[\hat{\tau}<\tau_b]\right]^2\right\} - \theta_{\tau_b}^2\tau_b^2 - 1 \qquad (21)$$

If the break date is known, 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 (21) 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 (21) 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 $|\theta_{\tau_b}| \to \infty$, there is a value of $|\theta_{\tau_b}|$—and thus for ζ_{τ_b}—for which $\Delta(\tau_b) = 0$ for each τ_b. Numerical results in the online appendix show that $\Delta(\tau_b)$ is a monotonically decreasing function of θ_{τ_b}, which implies that $\Delta(\tau_b) = 0$ for a unique value of θ_{τ_b}. This supports the use of (15) 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

While the structural break case is our main focus, the results in this section hold for a general form of structural change as long as 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 (17) 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 $u \to \infty$*

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

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

The proof is presented in the appendix. 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 (21) 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 (21) equals zero, we can obtain a critical value $u(\tau_b)$ such that $P(\sup_{\tau \in \mathcal{T}} Q^*(\tau) > u(\tau_b)^2) = \alpha$. 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* $\mathcal{T}_1$*, 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 $\frac{\partial\mu(\tau;\theta_{\tau_b})}{\partial\tau} = \theta_{\tau_b}[\tau_b(1-\tau_b)]^{-1/2}$. 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. Assumption 3 can be verified once critical values are obtained. In the online appendix we show that the inequality holds for the case of the structural break model.

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

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*

$$P\left(\sup_{\tau\in\mathcal{T}} Q^*(\tau) > u(\tau_b)^2\right) = \alpha \tag{22}$$

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

$$P\left(\sup_{\tau\in\mathcal{T}} Q^*(\tau) > u(\hat{\tau})^2\right) = \alpha \tag{23}$$

where $\hat{\tau}$ *is given in* (18)*.*

The proof is in the appendix. 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 $P_{H_0}\left(Q^*(\tau_b) > v(\tau_b)^2\right) = \alpha$, then we can establish the following theorem.

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

$$\begin{aligned}
&P_{H_a}\left[\sup_{\tau} Q^*(\tau) > u(\hat{\tau})^2\right] - P_{H_a}\left[Q^*(\tau_b) > v(\tau_b)^2\right] \\
&\geq P_{H_a}\left[Q^*(\tau_b) > u(\tau_b)^2\right] - P_{H_a}\left[Q^*(\tau_b) > v(\tau_b)^2\right] \\
&= 0
\end{aligned} \tag{24}$$

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

The proof is in the appendix. A test based on the Wald statistic (15) 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 $u \to \infty$.

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

$$S(\hat{\tau}) = \sup_{\tau \in \mathcal{T}} \sqrt{T} \frac{\left| \boldsymbol{f}'_{\beta_2} \left(\hat{\boldsymbol{\beta}}_2(\tau) - \hat{\boldsymbol{\beta}}_1(\tau) \right) \right|}{\sqrt{\boldsymbol{f}'_{\beta_2} \left(\frac{\hat{\boldsymbol{V}}_1}{\tau} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau} \right) \boldsymbol{f}_{\beta_2}}} - |\mu(\hat{\tau}; \theta_{\hat{\tau}})| \tag{25}$$

where $\hat{\tau}$ maximizes the first term of S or, equivalently, the Wald statistic (15)*.*

The proof is presented in the appendix. Finally, from the location concentration in Theorem 1 it follows that in the limit where $\alpha \to 0$, inference following a rejection is standard.

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

$$D \left(\hat{\tau} : Q^*(\hat{\tau}) = \sup_{\tau \in \mathcal{T}} Q^*(\tau) \middle| \sup_{\tau \in \mathcal{T}} Q^*(\tau) > u^2 \right) \overset{a}{\sim} \delta_{\tau_b} \text{ as } u \to \infty$$

4.2.3 Testing procedure

To summarize, we use the following steps to make the test for $\Delta(\tau_b) = 0$ in (21) operational

1. Using (18), evaluate (21) by simulation to find the break magnitude θ_{τ_b} that yields $\Delta(\tau_b) = 0$ for each τ_b.

2. For each τ_b and corresponding θ_{τ_b} obtain a critical value $u(\tau_b)$ such that $P(\sup_{\tau \in \mathcal{T}} Q^*(\tau_b) > u(\tau_b)^2) = \alpha$.

3. Now the test statistic $\sup_{\tau \in \mathcal{T}} Q^*(\tau)$ or its finite sample analogue can be compared to the critical value $u(\hat{\tau})^2$ with $\hat{\tau}$ from (18).

   - This test controls size $P(\sup_{\tau \in \mathcal{T}} Q^*(\tau) > u(\hat{\tau})^2) = \alpha$ 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 operationalize test statistic (25), which leads to critical values that are independent of the unknown break date for sufficiently small size. We will present critical values for both test statistics in Section 5.

4.3 Combining post-break and full sample 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

$$\hat{y}^c_{T+h}(\tau) = \omega f_{T+h}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) + (1-\omega) f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T)$$

where the optimal forecast is denoted with superscript c. This forecast can be rewritten as a combination of the post-break sample forecast and the full sample forecast forecast as follows. 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

$$\begin{aligned}
&\lim_{T\to\infty} \mathrm{E}\left[T\left(\hat{y}^c_{T+h} - f_{T+h}(\boldsymbol{\beta}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T)\right)^2\right] = \\
&= \lim_{T\to\infty} \mathrm{E}\left[T\left(\omega \boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) + \boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2)\right)^2\right] \qquad (26)\\
&= \omega^2 \left(\boldsymbol{f}'_{\beta_2}\boldsymbol{b}\right)^2 + \omega^2 \boldsymbol{f}'_{\beta_2}\left(\frac{1}{\tau_b} + \frac{1}{1-\tau_b}\right)\boldsymbol{V}\boldsymbol{f}_{\beta_2} - 2\omega\frac{1}{1-\tau_b}\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} + \frac{1}{\tau_b}\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}
\end{aligned}$$

where $\boldsymbol{f}_{\boldsymbol{\beta}_2} = \frac{\partial f_{T+h}(\boldsymbol{\beta}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T)}{\partial \boldsymbol{\beta}_2}$, and $\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1 = \boldsymbol{b}/\sqrt{T}$, and the first equality relies on a Taylor expansion and the local-to-zero nature of the breaks. Details are in the online appendix.

Maximizing (26) with respect to ω yields

$$\omega^* = \tau_b \left[1 + \frac{\left[\boldsymbol{f}'_{\beta_2}\boldsymbol{b}\right]^2}{\boldsymbol{f}'_{\beta_2}\left(\frac{1}{\tau_b} + \frac{1}{1-\tau_b}\right)\boldsymbol{V}\boldsymbol{f}'_{\beta_2}}\right]^{-1} \qquad (27)$$

where the denominator contains the standardized break magnitude ζ defined in (13).

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

$$\begin{aligned}
\hat{y}^c_{T+h} &= \omega f_{T+h}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) + (1-\omega) f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) + o_p(T^{-1/2}) \\
&= \frac{\omega}{\tau_b} f_{T+h}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) + \left(1 - \frac{\omega}{\tau_b}\right) f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) + o_p(T^{-1/2})
\end{aligned}$$

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

$$\omega^*_F = \frac{\omega^*}{\tau_b} = \frac{1}{1+\zeta} \qquad (28)$$

The forecast $\hat{y}^c_{T+h}$ is therefore a convex combination of the full sample and post-break sample forecast with weights that are determined by the standardized break magnitude. In practice, this is replaced by the Wald test statistic (14).

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 combined forecast.

As the Wald statistic (14) is conditional on the true break date, consider the combined forecast for a general value of τ

$$\begin{aligned}\hat{y}^c_{T+h}(\tau) &= \frac{1}{1+W(\tau)} f_{T+h}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) + \frac{W(\tau)}{1+W(\tau)} f_{T+h}(\hat{\boldsymbol{\beta}}_2(\tau), \hat{\boldsymbol{\delta}}; \mathcal{I}_T) \\ &\Rightarrow \frac{1}{1+Q^*(\tau)} f_{T+h}(\hat{\boldsymbol{\beta}}^a_F, \hat{\boldsymbol{\delta}}^a) + \frac{Q^*(\tau)}{1+Q^*(\tau)} f_{T+h}(\hat{\boldsymbol{\beta}}^a_2(\tau), \hat{\boldsymbol{\delta}}^a)\end{aligned} \tag{29}$$

where $\hat{\boldsymbol{\beta}}^a_F$, $\hat{\boldsymbol{\beta}}^a_2(\tau)$ and $\hat{\boldsymbol{\delta}}^a$ denote the asymptotic distributions provided in (11) and (12). The last line holds by the continuous mapping theorem. The difference in MSFE between the combined forecast and the full sample forecast, after applying a Taylor expansion on the forecast function f_{T+h}, is given by

$$\begin{aligned}\Delta_c =& \mathrm{E}\Bigg\{\Bigg[\frac{1}{1+Q^*(\tau)}\left(\frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\bar{\boldsymbol{X}}'\boldsymbol{B}(1)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} + \int_0^1 \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} ds\right) \\ &+ \frac{Q^*(\hat{\tau})}{1+Q^*(\tau)}\left(\frac{1}{1-\hat{\tau}}\frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\bar{\boldsymbol{X}}'\left(\boldsymbol{B}(1)-\boldsymbol{B}(\hat{\tau})\right)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} + \frac{1}{1-\hat{\tau}}\int_{\hat{\tau}}^1 \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} ds\right)\Bigg]^2\Bigg\} \\ &- \left(\int_0^1 \frac{\boldsymbol{f}'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}}} ds\right)^2 - 1\end{aligned} \tag{30}$$

where we solve for $\Delta_c = 0$ numerically to obtain the break magnitude that corresponds to equal predictive accuracy. Numerical results in the online appendix 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 applied to find the appropriate critical values, which are tabulated in the next section.

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 $T \to \infty$. 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 (16) with (17) for different combinations of break date and magnitude $\{\tau_b, \theta_{\tau_b}\}$. Here, we focus on $\tau_b = \{\tau_{\min}, \tau_{\min} + \delta_\tau, \ldots, \tau_{\max}\}$ where $\tau_{\min} = 0.15$, $\tau_{\max} = 1 - \tau_{\min}$ and $\delta_\tau = 0.01$. Additionally, we ran simulations for $\tau_{\min} = 1 - \tau_{\max} = 0.05$ and those results are reported in the online appendix. For the break magnitude, θ_{τ_b}, we consider $\theta_{\tau_b} = \{0, 0.5, \ldots, 20\}$. 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 Bai and Perron (1998).

By maximizing (16) we obtain a distribution of the estimated break date $\hat{\tau}$ that can be used to evaluate (20). To approximate the expectation, we use 50,000 repetitions for each break date and break magnitude. For each value of τ_b, we obtain the θ_{τ_b} that yields equal predictive accuracy in (20) for full sample and post-break forecasts. 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 versus full-sample forecast: break magnitude for equal forecast accuracy

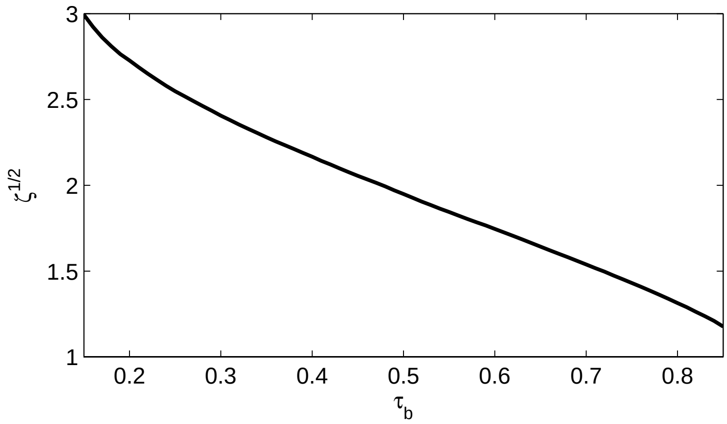
Using (20), we simulate the break magnitude for which the full sample and the post-break sample achieve equal predictive accuracy. Figure 1 shows the combinations of break magnitude and break date for which equal predictive accuracy is obtained. The break magnitude is given in units of the standardized break magnitude,

$$\zeta^{1/2} = \sqrt{T(1-\tau_b)\tau_b} \frac{\boldsymbol{f}'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)}{\sqrt{\boldsymbol{f}'_{\beta_2} \boldsymbol{V} \boldsymbol{f}_{\beta_2}}} \tag{31}$$

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

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 $\zeta^{1/2} = 1$, because the uncertainty of the break date estimation increases the MSFE of the post-break sample forecast. If a break occurs early the sample, $\tau = 0.15$, then 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

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 (31) for which forecasts from post-break and full sample achieve the same MSFE, that is, Δ in (20) equals zero.

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, introducing 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 truncated 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.

Table 1: Critical values and size of the W and S test statistics: Post-break versus full sample

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
α	*W test statistic* (15)														
	Critical values														
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
0.05	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
0.10	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
	Size														
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
0.05	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03
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
α	*S test statistic* (25)														
	Critical values														
0.01	2.76	2.79	2.81	2.83	2.85	2.86	2.86	2.87	2.86	2.86	2.85	2.83	2.80	2.74	2.60
0.05	2.12	2.16	2.18	2.20	2.21	2.22	2.23	2.23	2.22	2.22	2.20	2.18	2.14	2.08	1.94
0.10	1.78	1.82	1.84	1.86	1.87	1.88	1.89	1.89	1.88	1.87	1.86	1.83	1.80	1.73	1.59
	Size														
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
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04
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	0.10	0.08

Note: Note: Reported are critical values and size for, first, W, the Wald test statistic (15) and, second, S, the test statistic (25), which is independent of τ_b when the nominal size tends to zero. The null hypothesis is equal MSFE of the post-break and full sample forecasts.

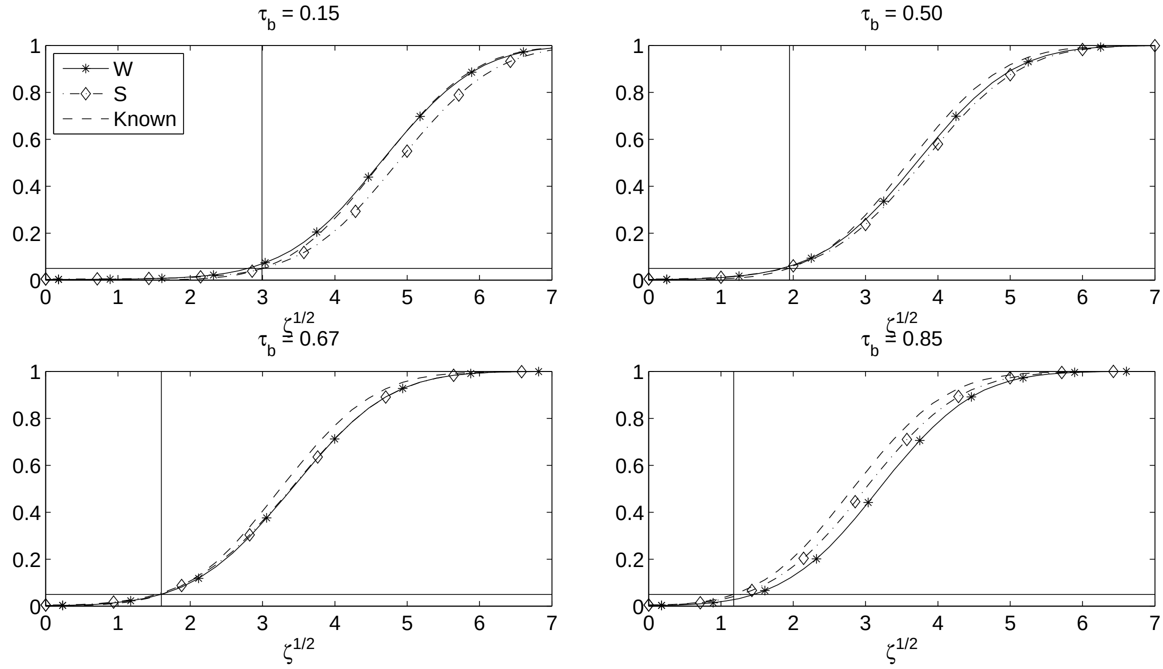
5.1.3 Critical values, size, and power

After finding the break magnitude for which post-break sample and full sample forecasts yield equal predictive accuracy, we can compute critical values for both the Wald-type test statistic in (15), which we will denote as W for simplicity from here, and the α-asymptotic statistic in (25), denoted as S, for a grid of break dates, τ_b. Assumption 3, which is required for the near optimality result does hold for all τ_b—details are available in the online appendix.

The first line of the second 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 (25) largely remedies these size distortions as the results in the bottom panel show.

The critical values are given in the first and third panel of Table 1. Critical values for a wider grid of the true break date can be found in the online appendix. The large break magnitude that yields equal forecast accuracy implies a major increase in critical values when using the Wald test statistic (15), compared to the standard values of Andrews (1993). For comparison, for a nominal size of $[0.10, 0.05, 0.01]$ the critical values in Andrews are equal to $[7.17, 8.85, 12.35]$.

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

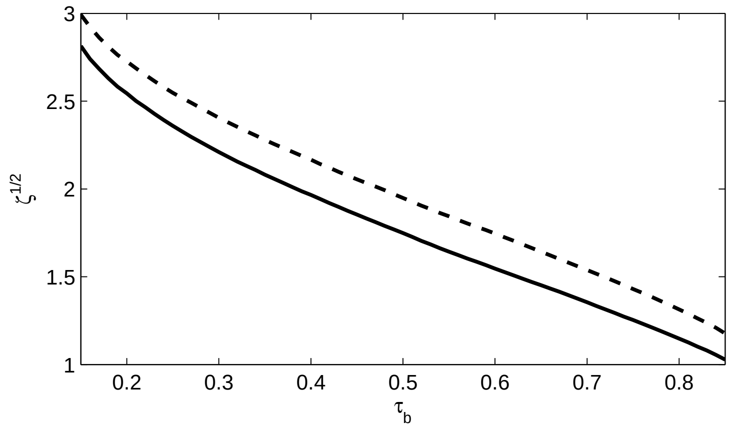


Note: The plots show the power for tests at a nominal size of $\alpha = 0.05$ with the null hypothesis given by the break magnitude depicted in Figure 1. The panels show power for different values of the (unknown) break date. The power of infeasible test conditional on the true break date is given as 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.

The critical values for the α-asymptotic test statistic, S, in (25) are independent of $\hat{\tau}$ in the limit where $\alpha \to 0$. 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, vary substantially less over $\hat{\tau}$ than those for the Wald statistic, W. 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 $\alpha \to 0$ and this can be observed in Table 1.

Given that the break magnitudes that lead to equal forecast performance 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.

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



Note: The solid line shows the standardized break magnitude for which the forecast combination (29) achieves the same MSFE as the full sample forecast, in which case (30) 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.

5.1.4 Forecast combination versus full-sample forecast

Figure 3 shows the combination of τ_b and break magnitude for which the forecast combination 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 and full sample forecasts 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 combined forecast is lower than for the post-break sample forecast. This implies that combining the post-break and full sample forecasts offers improvements over 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 combined forecast is more precise than the full sample forecast.

In order to determine whether to use the combined forecast, critical values can be obtained as before and are presented in Table 2. Again, the size is close to the theoretical size with small size distortions when using W, which are largely remedied when using S. Critical values on a larger grid of the true break date are presented in the online appendix.

Figure 4 displays the power curves of the tests that compare the combined 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 combined forecast are largely the same as those for the post-break forecast.

Table 2: Critical values and size of the W and S test statistics: Forecast combination versus full sample

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
α	*W test statistic* (15)														
	Critical values														
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
0.05	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
0.10	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
	Size														
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
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
0.10	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
α	*S test statistic* (25)														
	Critical values														
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
0.05	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
0.10	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
	Size														
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
0.05	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
0.10	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

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 forecast combination (29) and the full sample forecast.

5.1.5 Forecast combination versus the post-break forecast

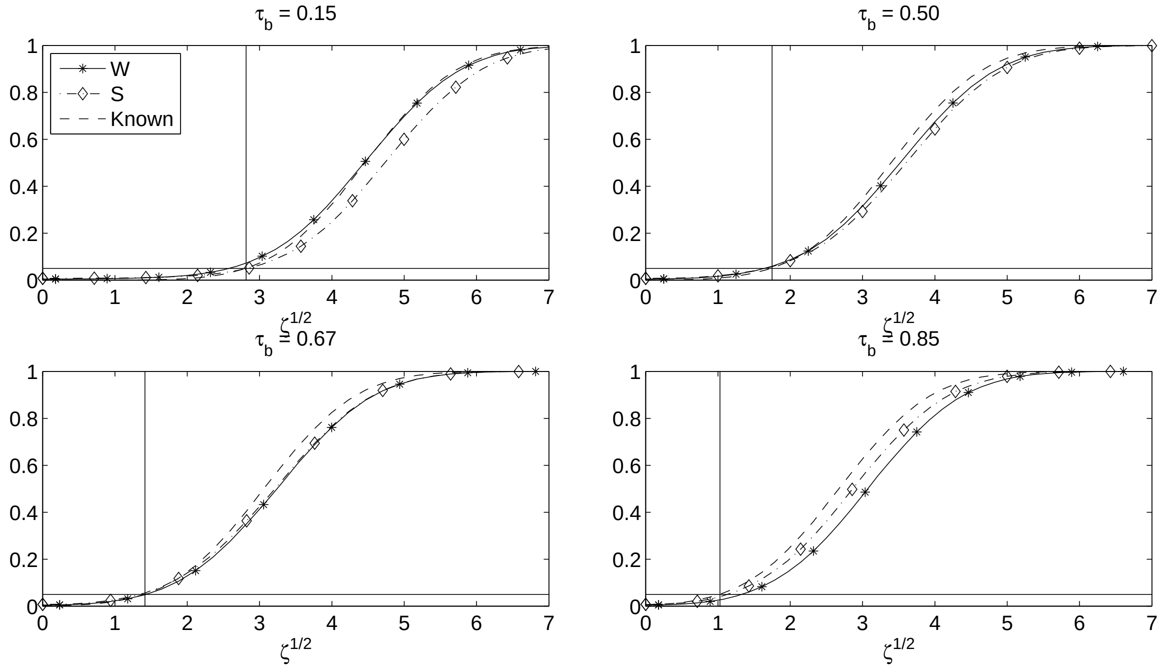
Finally, we investigate the break magnitudes that leads to equal forecast performance of the post-break forecast and the forecast that combines the post-break with the full sample forecast. Figure 5 plots the ratio of the MSFE of the combined forecast over that of the post-break forecast. For nearly all break magnitudes and dates, the combined 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 forecast based on the Wald statistic (15) and on the S-statistic (25). Next, we consider the same test statistics but now test for equal predictive accuracy between the forecast combination (29) and the full-sample, equal weighted forecast. All tests are carried out at a nominal size $\alpha = 0.05$, using sample sizes of $T = \{120, 240, 480\}$

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



Note: The plots show asymptotic power curves when testing for equal predictive accuracy between the forecast combination (29) 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.

and break dates $\tau_b = [0.15, 0.25, 0.50, 0.75, 0.85]$. While the sample sizes may appear large, note that $\tau_b = 0.15$ and $T = 120$ yield only 18 post-break observations. Parameter estimates are obtained by least squares, and the results are based on 10,000 repetitions.

The data generating process (DGP) is given by

$$y_t = \mu_t + \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \tag{32}$$

where $\sigma^2 = 1$ and

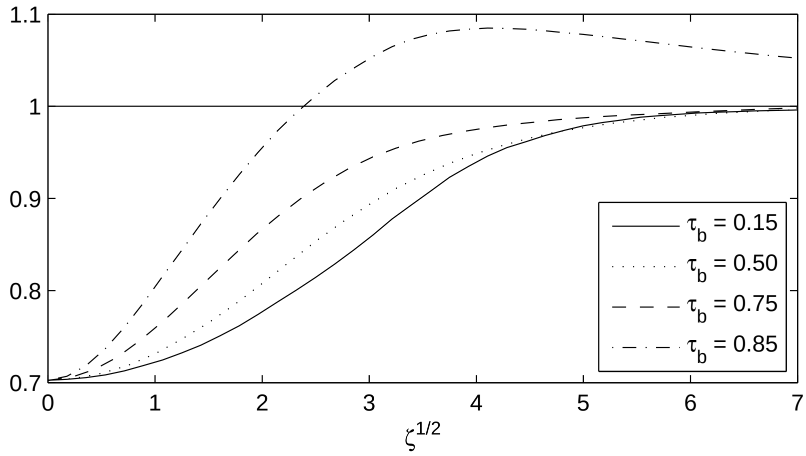
$$\mu_t = \begin{cases} \mu_1 & \text{if } t \leq \tau_b T \\ \mu_2 & \text{if } t > \tau_b T \end{cases}$$

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 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 persistent increases to $\rho = 0.9$, some size distortions become apparent for $T = 120$. Those do, however, diminish as T increases. These size

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



Note: The graph shows the relative performance of the forecast combination (29) 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 forecast combination is more precise.

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 forecast combination 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 the effective sample size is small.

6 Application

We investigate the importance of structural breaks for 130 macroeconomic and financial time series from the FRED-MD database (McCracken and Ng, 2016). We use the vintage from May 2016 and transform the data as suggested by McCracken and Ng (2016). 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 & income* (OI, 17 series), *labor market* (LM, 32 series), *consumption & orders* (CO, 10 series), *orders & inventories* (OrdInv, 11 series), *money & credit* (MC, 14 series), *interest rates & exchange rates* (IRER, 21 series), *prices* (P, 21 series), and *stock market* (S, 4 series).

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

		$T = 120$					$T = 240$					$T = 480$				
ρ	$\lambda \backslash \tau_b$	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 (15)									
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 (25)									
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.

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 $\hat{\boldsymbol{V}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\hat{\boldsymbol{\Omega}}_i\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$ with $[\hat{\boldsymbol{\Omega}}_i]_{kl} = \hat{\varepsilon}_k^2/(1-h_k)^2$ if $k = l$ and $[\hat{\boldsymbol{\Omega}}_i]_{kl} = 0$

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

		$T = 120$					$T = 240$					$T = 480$				
ρ	$\lambda\backslash\tau_b$	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 (15)									
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 (25)									
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 forecast combination (29) and the full-sample, equal weights forecast, using a nominal size of 0.05. For further details, see the footnote of Table 3.

otherwise, and h_k is the k-th diagonal element of $\boldsymbol{P_X} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. 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.

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

	sup W	W	S	W^c	S^c
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: *sup W* 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^c and S^c refer to the tests that compare combined and full sample forecasts. All tests are carried out at $\alpha = 0.05$.

6.1 Structural break test results

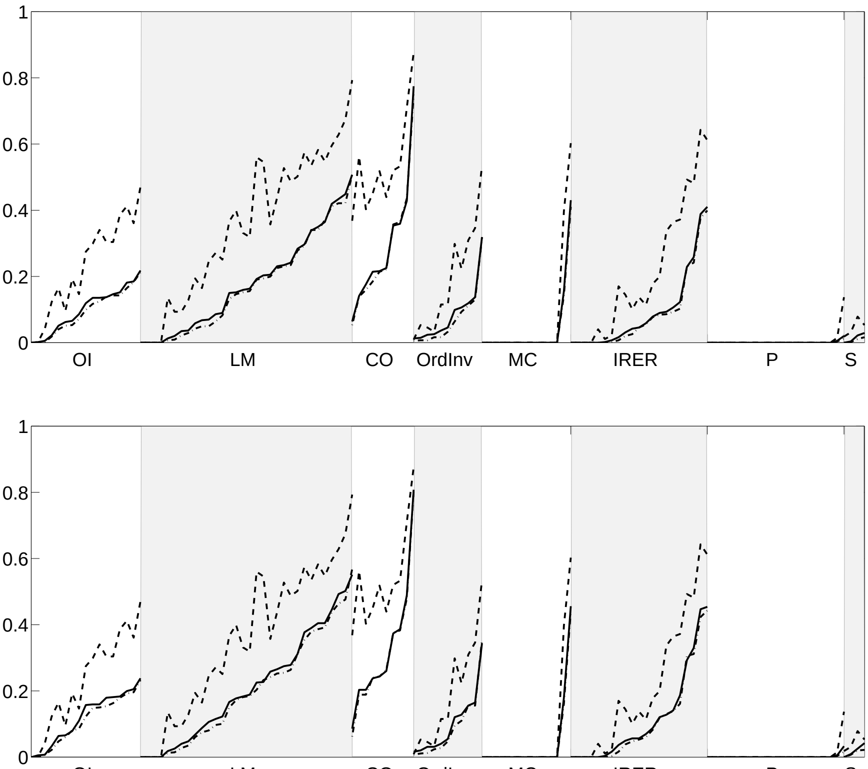
In this forecast exercise, we compare our Wald test statistic, W, in (15), the S-test in (25), the tests based on the combined forecast, which we denote as W^c and S^c, and the *sup W* of Andrews (1993). For all tests we use $\alpha = 0.5$ and $\tau_{\min} = 0.15$. In Table 5, we report the fraction of estimation samples where the tests indicate a break. It is clear that a large fraction of the breaks picked up by Andrews' *sup W* are judged as irrelevant for forecasting by W, S, W^c, and S^c. 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 each category we sort the series based on the fraction of breaks found by W. Across all categories Andrews' *sup W* test is more often significant than the W and S tests for both, post-break and combined forecast. 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 Andrews' *sup W* 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 combined forecast and the full sample, equal weights forecast. It is clear that Andrews' *sup W* test finds more breaks in for the vast majority of estimation samples,

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



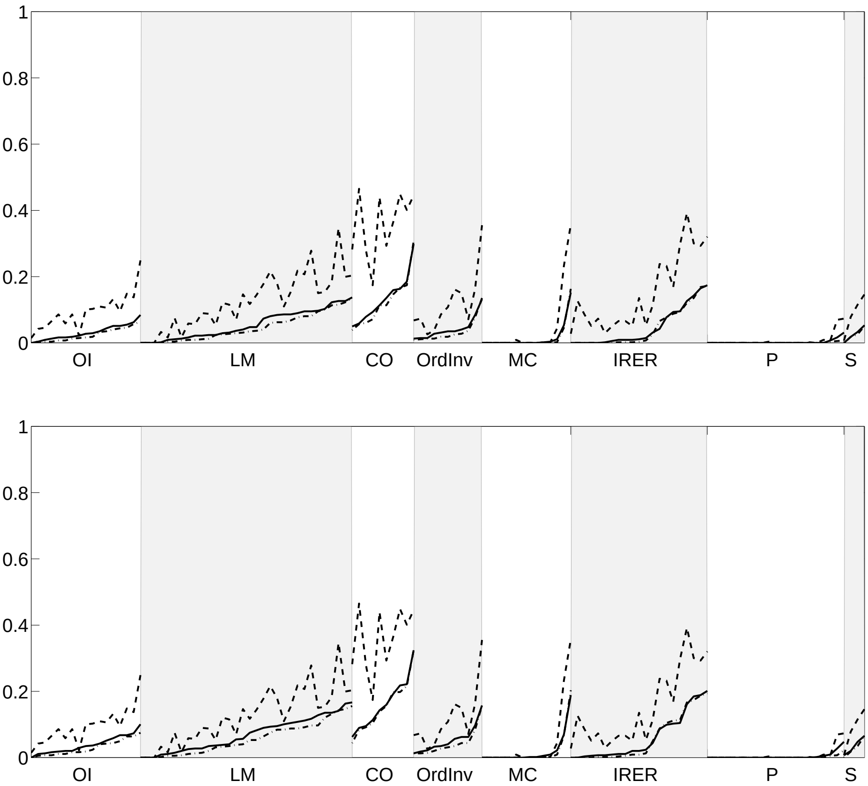
Note: The upper panel depicts the fraction of estimation samples with a significant break when testing under the alternative of the post-break forecast; the lower panel when testing under the alternative of the forecast combination (29). Dashed lines indicate the fraction of estimation samples with significant Andrews' $supW$ test, dashed-dotted lines indicate the fraction of estimation samples where the break test W in (15) indicates a break, and solid lines indicate the fraction of estimation samples with significant S test in (25).

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 beaks, from 1985 Andrews' $supW$ test finds many more series that contain breaks than the W and S tests. This remains true until 2009 where the W and S tests find the same and, in the case of the combined forecast, even more breaks that are relevant for forecasting than Andrews' $supW$ test. From 2010 onwards, breaks that are relevant for forecasting decrease sharply, whereas Andrews' $supW$ 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, Andrews' $SupW$ 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, Andrews' $supW$ test finds breaks in up to a third of the estimation samples, most

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



Note: See footnote of Table 6

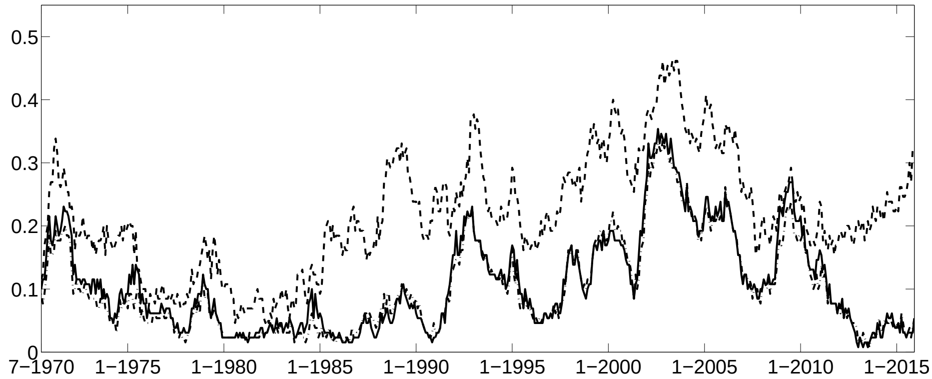
of which the W and S tests do not find important for forecasting. The same is true for breaks around 2000. In contrast, in the period following the dot com bubble and following the financial crisis of 2008/9 the W and the S tests find as many and, in the case of the combined forecasts, more series where taking a break into account will improve forecast accuracy than Andrews' $supW$ 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 Andrews' $supW$ tests.

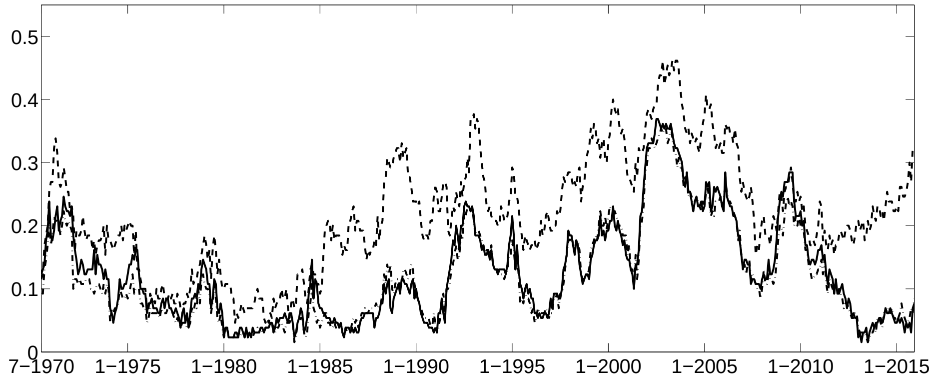
6.2 Forecast accuracy

Given the different test results, we now investigate whether forecasts conditional on the W and S tests are more accurate than forecasts based on Andrews' $supW$ 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 combined or the full sample forecast and, given these results, we construct the respective forecast.

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

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





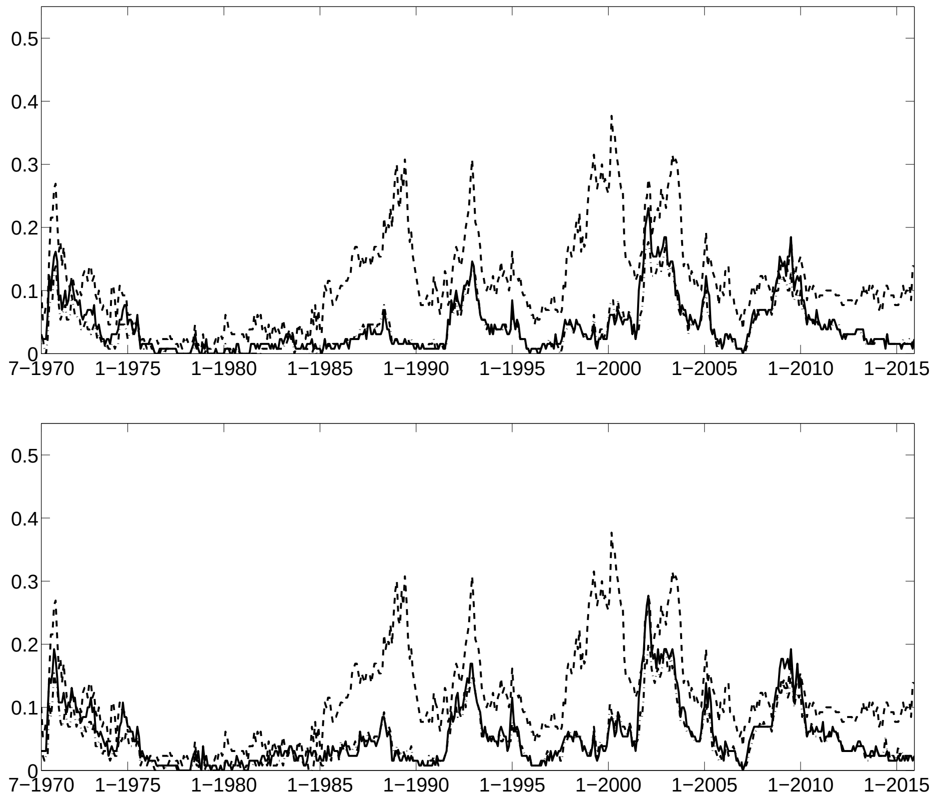
Note: The plots show the fractions of series with a significant break for each estimation sample when using an AR(1) model with a break in intercept. The top panel shows results when testing between the post-break sample based forecast and the full sample based forecast and the lower panel when testing between the combined forecast and the full sample, equal weights forecast. The dashed line indicates the fraction of series when testing using Andrews' $supW$ test at $\alpha = 0.05$, the solid line when testing using the S-test in (25), and the dashed-dotted line when testing using the W-test in (15). The dates displayed on the horizontal axis are the end dates of the estimation samples.

where at least one test finds a break as the estimation samples where no test finds a break will to lead to identical full sample forecasts.

The results show that using the W test in place of Andrews' $supW$ 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 combined forecast is used in conjunction with the W^c or S^c 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 combination weights, is relatively large. This implies that upon rejection of the test statistic, a forecast is used that

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.

is relatively close to the post-break forecast. The last column shows that using the combined forecast in conjunction with Andrews' $supW$ test leads to forecasts that, while more precise than post-break forecasts based on the same test, are clearly dominated by the W^c and S^c tests. In fact, for all categories and both models the W^c test leads to more accurate forecasts, as does the S^c 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 forecasts based on the full sample and based on a post-break sample with an estimated break date. This break magnitude is substantial, which points to a large penalty that is incurred by the uncertainty around the estimated break date. Additionally, 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

Table 6: Relative MSFE compared to Andrews' $supW$ test

		Post-break		Combination		
		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 Andrews' $supW$ test 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 (15) and S in (25). 'Post-break' and 'Combination' indicate that under the alternative the post-break forecast, respectively the forecast combination (29), 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).

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 combination of the post-break and full sample forecasts, with our test statistic governing the combination weights. Our test extends in a straightforward way to test whether the combined forecast will be more accurate than the full sample forecast.

We apply the test to a large set of macroeconomic and financial time series and find that breaks that are relevant for forecasting are rarer than the test of Andrews (1993). 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 with optimal weights or forecast combination 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.

Chib, S. and Kang, K. H. (2013). Change-points in affine arbitrage-free term structure models. *Journal of Financial Econometrics*, 11(2):302–334.

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.

Dufour, J.-M., Ghysels, E., and Hall, A. (1994). Generalized predictive tests and structural change analysis in Econometrics. *International Economic Review*, 35(1):199–229.

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.

Giacomini, R. and Rossi, B. (2010). Forecast comparison in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620.

Hansen, B. E. (2009). Averaging estimators for regressions with a possible stuctural 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.

McCracken, M. W. and Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Pastor, L. and Stambaugh, R. F. (2001). The equity premium and structural breaks. *Journal of Finance*, 56(4):1207–1231.

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., Pick, A., and Timmermann, A. (2011). Variable selection, estimation and inference for multi-period forecasting problems. *Journal of Econometrics*, 164(1):173–187.

Pesaran, M. H. and Timmermann, A. (2002). Market timing and return prediction under model instability. *Journal of Empirical Finance*, 9(5):495–510.

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.

Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60–78.

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.

Stock, J. H. and Watson, M. W. (2007). Why has U.S. inaflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(1):3–33.

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: Proofs

Proof of Theorem 1 The proof that only points in a small neighborhood of the true break date contribute to the probability of exceeding a distant boundary, requires 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 $u \to \infty$*

$$P\left(\sup_{\tau \times c}[Z(\tau) + \mu(\tau;\theta_{\tau_b})]c > u\right) = P\left(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})| \quad \textit{for some } \tau \in \mathcal{T}\right)[1 + o(1)]$$

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

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

$$\begin{aligned} P(Z(\tau) + \mu(\tau;\theta_{\tau_b}) > u, \tau \in \mathcal{T}) &= P(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})|, \tau \in \mathcal{T}) \\ P(-Z(\tau) - \mu(\tau;\theta_{\tau_b}) > u, \tau \in \mathcal{T}) &= P(Z(\tau) > u + |\mu(\tau;\theta_{\tau_b})|, \tau \in \mathcal{T}) \end{aligned} \tag{33}$$

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

$$\begin{aligned} P(-Z(\tau) - \mu(\tau;\theta_{\tau_b}) > u, \tau \in \mathcal{T}) &= P(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})|, \tau \in \mathcal{T}) \\ P(Z(\tau) + \mu(\tau;\theta_{\tau_b}) > u, \tau \in \mathcal{T}) &= P(Z(\tau) > u + |\mu(\tau;\theta_{\tau_b})|, \tau \in \mathcal{T}) \end{aligned} \tag{34}$$

The bounds in the second lines of (33) and (34) are equal or larger then 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 $u \to \infty$

$$P\left(\sup_{\tau \times c}[Z(\tau) + \mu(\tau;\theta_{\tau_b})]c > u\right) = P\left(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})| \text{ for some } \tau \in \mathcal{T}\right)[1 + o(1)] \tag{35}$$

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* $0 < C(\tau) < \infty$

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

The correlation function can be written as

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

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* $\delta(u)u^2 \to \infty$ *and* $\delta(u) \to 0$ *as* $u \to \infty$

$$\lim_{u \to \infty} P\left(\sup_{[\tau, \tau+\delta(u)]} Z(t) > u\right) = \frac{1}{\sqrt{2\pi}} \delta(u) u \exp\left(-\frac{1}{2}u^2\right) C(\tau) \tag{36}$$

Proof: see Hüsler (1990).

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

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

where the supremum is taken jointly over $\tau \in \mathcal{T}$ 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 $\mathcal{T}_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$. In $\mathcal{T}_1$, the minimum value of the boundary is given by

$$\underline{b} = \inf_{\tau \in \mathcal{T}_1} [u - |\mu(\tau; \theta_{\tau_b})|] = u - |\mu(\tau_b; \theta_{\tau_b})| \tag{37}$$

and therefore

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

where the third line follows from (36).

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

$$\underline{b}_A = u - |\mu(\tau_b + \delta(u);\theta_{\tau_b})| \tag{38}$$

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

$$-|\mu(\tau_b + \delta(u);\theta_{\tau_b})| = -|\mu(\tau_b;\theta_{\tau_b})| + \gamma\delta(u) + O\left[\delta(u)^2\right] \tag{39}$$

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

$$
\begin{aligned}
\lim_{u\to\infty} P_{\mathcal{T}_A} &= \lim_{u\to\infty} P(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})| \text{ for some } \tau \in \mathcal{T}_A) \\
&\leq \lim_{u\to\infty} P(Z(\tau) > \underline{b}_A \text{ for some } \tau \in \mathcal{T}_A) \\
&\leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\underline{b}^2 - \underline{b}\gamma\delta(u) - \frac{1}{2}\gamma^2\delta(u)^2 + \log(\underline{b} + \gamma\delta(u))\right) \overline{C}
\end{aligned}
\tag{40}
$$

where we define $\overline{C}$ by noting that

$$\sum_{\mathcal{T}_k \in \mathcal{T}_A} C(k\delta(u))\delta(u) \overset{\delta(u)\to 0}{\longrightarrow} \int_{\mathcal{T}_A} C(\tau)d\tau \leq \int_{\mathcal{T}} C(\tau)d\tau = \overline{C} < \infty \tag{41}$$

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

Compare (40) to the probability of a test with a known break date to exceed the critical value

$$P_0 = P(Z(\tau_b) > u - |\mu(\tau_b;\theta_{\tau_b})|) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\underline{b}^2 - \log(\underline{b})\right) \tag{42}$$

where we use that

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

Ignoring the lower order term $-\frac{1}{2}\gamma^2\delta(u)^2 + \log(\underline{b} + \gamma\delta(u))$, equation (40) contains an extra term $\exp(-\underline{b}\gamma\delta(u))$ compared to (42). This term is decreasing as u increases, as we argued above that $\gamma\delta(u) > 0$. Recalling (37), this implies that $P_{\mathcal{T}_A} = o(P_0)$ if $\frac{u\delta(u)}{\log u} \to \infty$. Then, if

$$\delta(u) = u^{-1}\log^2(u), \tag{43}$$

all intervals outside of $\mathcal{T}_1$ contribute $o(P_0)$ to the probability of crossing the boundary u. Under (43), we have that for $P_{\mathcal{T}_1}$ as $u \to \infty$, $P_{\mathcal{T}_1} \leq P_{\mathcal{T}} \leq P_{\mathcal{T}_1} + P_{\mathcal{T}_A} \leq P_{\mathcal{T}_1} + o(P_0)$. We now only need to note that

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

to conclude that

$$P\left(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})|, \tau \in \mathcal{T}\right) \overset{u\to\infty}{\longrightarrow} P_{\mathcal{T}_1}(1 + o(1))$$

which completes the proof. ■

Note that, in (40), the term $\exp(\underline{b}\delta(u))^{-\gamma}$ ensures that $P_{\mathcal{T}_A} = o(P_1)$. In the structural break model, we see that (39) is given by $\mu(\tau_b+\delta(u);\theta_{\tau_b}) = \theta_{\tau_b}\sqrt{\tau_b(1-\tau_b)} - \frac{1}{2}\theta_{\tau_b}\frac{1}{\sqrt{\tau_b(1-\tau_b)}}\delta(u) + O\left[\delta(u)^2\right]$. 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 in Section 5 confirm this.

Proof of Theorem 2 Within the interval $\mathcal{T}_1$, we have $u_- \leq u(\tau_b) \leq u_+$ and $u_- \leq u(\hat{\tau}) \leq u_+$. The lower and upper bounds satisfy

$$\begin{aligned} u_- &= u(\tau_b) - \left|\frac{\partial u(\tau_b)}{\partial \tau_b}\right| \delta(u) + O(\delta(u)^2) \geq u(\tau_b) - C\delta(u) + O(\delta(u)^2) \\ u_+ &= u(\tau_b) + \left|\frac{\partial u(\tau_b)}{\partial \tau_b}\right| \delta(u) + O(\delta(u)^2) \leq u(\tau_b) + C\delta(u) + O(\delta(u)^2) \end{aligned} \tag{44}$$

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

$$\begin{aligned} \epsilon &= P(\sup_\tau Q^*(\tau) > u_-^2) - P(\sup_\tau Q^*(\tau) > u_+^2) \\ &= \frac{1}{\sqrt{2\pi}}\delta(u)u(\tau_b)\exp\left(-\frac{1}{2}u(\tau_b)^2\right)\left[\exp(-C\delta(u)) - \exp(+C\delta(u))\right]C(\tau_b) + o(\cdot) \\ &\to 0 \end{aligned} \tag{45}$$

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

$$P\left(\sup_\tau Q^*(\tau) > u_+^2\right) \leq P\left(\sup_\tau Q^*(\tau) > u(\tau_b)^2\right) \leq P\left(\sup_\tau Q^*(\tau) > u_-^2\right)$$

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

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

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* (15) *is used as a test statistic. Let $v(\tau_b)$ be the critical value when using the test statistic with $\tau = \tau_b$, then $u(\tau_b) - v(\tau_b) \to 0$.*

Proof: By definition of the critical values

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

Since τ in the first line is contained in $\mathcal{T}_1$, 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 $\delta(u) \to 0$ as $u \to \infty$, the difference in the critical values $u(\tau_b) - v(\tau_b) \to 0$ as $u \to \infty$. ■

A proof of Theorem 3 readily follows. With $\hat{\tau}$ from (18) we have

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

Under the slowly varying assumption, $u(\hat{\tau}) - \mu(\hat{\tau};\theta_{\tau_b})$ has a unique minimum on $\mathcal{T}_1$ 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 (24). The last line of (24) follows from Lemma 4. ■

Proof of Corollary 1 The test statistic converges to $S(\hat{\tau}) \to \sup_\tau |Z(\tau) + \mu(\tau;\theta_{\tau_b})| - |\mu(\hat{\tau};\theta_{\hat{\tau}})|$ 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 $\delta(u) \to 0$ as $u \to \infty$. Then

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

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 $u \to \infty$. ■