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

Tom Boot* Andreas Pick†

August 31, 2016

Abstract

We propose a near optimal test for structural breaks of unknown timing when the purpose of the analysis is to obtain accurate forecasts. Under mean squared forecast error loss, a bias-variance trade-off exists where small structural breaks should be ignored. We study critical break sizes, assess the relevance of the break location, and provide a test to determine whether modeling a break will improve forecast accuracy. Asymptotic critical values and weak optimality properties are established allowing for a small break to occur under the null, where the allowed break size varies with the break location. The results are extended to a class of shrinkage forecasts with our test statistics as shrinkage constants. Empirical results on a large number of macroeconomic time series show that structural breaks that are relevant for forecasting occur much less frequently than indicated by existing tests.

JEL codes: C12, C53

Keywords: structural break test, forecasting, squared error loss

*Erasmus University Rotterdam, boot@ese.eur.nl

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

1 Introduction

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

We develop a test for equal forecast accuracy that compares the expected mean square forecast error (MSFE) of a one-step-ahead forecast from the post-break sample to that of a forecast that uses the full sample. Under a known break date, the break size for which post-break sample and full sample forecasts achieve equal predictive accuracy is one standard deviation of the distribution of the parameter estimates. Under a local break of unknown timing, the uncertainty around the break date increases the variance of the post-break sample forecast and the break size of equal forecast accuracy is much larger, up to three standard deviations in terms of the distribution of the parameter estimates.

Building on the work of Andrews (1993) and Piterbarg (1996), we derive a test for the critical break size, which is optimal as the size tends to zero. Simulations of asymptotic power show that our test is near optimal for conventional choices of the nominal size, which is largely due to the size of the breaks that are allowed under the null. In the process, we show that the near optimality of the test follows from an optimality argument of the estimated break date by maximizing a Wald test statistic. This optimality does not depend on whether the 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. We also show that post-test inference following a rejection remains standard if the size of the test is small.

While our test uses much of the asymptotic framework of Andrews (1993), it is substantially different from extant break point tests, such as those of Ploberger et al. (1989), Andrews (1993), Andrews and Ploberger (1994), Elliott and Müller (2007, 2014), and Elliott et al. (2015). While those tests focus on the difference between (sub-)sets of parameters of a model before and after a break date, our measure is the forecast accuracy of the entire model. Our test, therefore, allows for a break in the parameters under the null of equal forecast accuracy.

In line with much of the forecasting literature, our loss function is the mean squared forecast error. Like 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 Trenkler and Toutenburg (1992) and Clark and McCracken (2012), our test is much simpler in that, under a known break date, our test statistic has a known distribution that is free of unknown parameters.

Our test is different from forecast accuracy tests of the kind suggested by Diebold and Mariano (1995) and extended by, among others, Clark and McCracken (2001); a recent review is by Clark and McCracken (2013). These tests assess forecast accuracy *ex post*. In contrast, the test we propose in this paper is an *ex ante* test of the accuracy of forecasts of models that do or do not account for breaks.

Giacomini and Rossi (2009) assess forecast breakdowns in the sense that the forecast performance of a model is not in line with the in-sample fit of the model. They consider forecast breakdowns in historically made forecasts as well as prediction of forecast breakdowns. Our approach is more targeted asking whether a structural beak, which is one of the possible sources of forecast breakdown, needs to be addressed from a forecast perspective.

The competing forecasts in our test are those using the full sample and using the post-break sample. Recently, Pesaran et al. (2013) showed 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 estimator can be written as a shrinkage estimator in the tradition of Thompson (1968), where the shrinkage estimator averages between the full sample estimator and post-break sample estimator with a weight that is equivalent to the test statistic introduced in this paper.

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

Substantial evidence of structural breaks has been found in macroeconomic and financial time series by Stock and Watson (1996), Rapach and Wohar (2006), Rossi (2006), Paye and Timmermann (2006), and others. Hence, we apply our test to the macroeconomic and financial time series and use the FRED-MD data set by McCracken and Ng (2015) We find that breaks that are important for forecasting under MSFE loss are between a factor two to three less frequent than the usual sup-Wald test by Andrews (1993) would indicate. Incorporating only the breaks suggested by our test substantially reduces the average MSFE in this data set compared to the forecasts that take all breaks suggested by Andrew's sup-Wald test into account.

The paper is structured as follows. We start with the motivating example of the linear regression model with one break of known timing in Section 2. The model is generalized in Section **??** using the methodology of Andrews (1993). In Section 3, we derive the test and show its weak optimality. We extend the test to cover the optimal weights or shrinkage forecast in Section 3.4 . Simulation results in Section 4 shows that the weak optimality of the test is in fact quite strong, with power very close to the optimal, but infeasible test that knows 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 5.

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

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

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

where

$$\boldsymbol{\beta}_t = \begin{cases} \boldsymbol{\beta}_1 & \text{if } t \leq T_b \\ \boldsymbol{\beta}_2 & \text{if } t > T_b \end{cases} \tag{2}$$

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

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

In this paper, we would like to test whether the expected mean squared forecast error (MSFE) from the forecast using the full sample, $\hat{y}^F_{T+1} = \mathbf{x}'_{T+1}\boldsymbol{\beta}_F$, is smaller than that of the post-break sample, $\hat{y}^P_{T+1} = \mathbf{x}'_{T+1}\boldsymbol{\beta}_2$.

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

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

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

$$
\begin{aligned}
R(\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F) &= \mathrm{E}\left[\left(\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2 - \varepsilon_{T+1}\right)^2\right] \\
&= \mathrm{E}\left[\left(\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2\right)\right]^2 + \frac{1}{T}\boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{x}_{T+1} + \sigma^2 \\
&= \left[\frac{T_b}{T}\boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2 + \frac{1}{T}\boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{x}_{T+1} + \sigma^2
\end{aligned} \tag{4}
$$

Comparing (3) and (4), we see that the full sample forecast is at least as accurate as the post-break sample forecast if

$$
\begin{aligned}
\zeta &= T\tau_b^2 \frac{\left[\boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\boldsymbol{V}_2}{1-\tau_b} - \boldsymbol{V}_F\right)\boldsymbol{x}_{T+1}} \\
&\overset{a}{\to} T\tau_b(1-\tau_b)\frac{\left[\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1}} \\
&\leq 1
\end{aligned} \tag{5}
$$

where $\tau_b = T_b/T$ and the third line assumes that the covariance matrices asymptotically satisfy $\mathrm{plim}_{T\to\infty}\,\boldsymbol{V}_i \to \boldsymbol{V}$ for $i = 1, 2, F$.

From (5) it can be observed that, under $H_0 : \zeta = 1$, the size of the break $\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$ is symmetric in ζ. Additionally, (3) suggests that breaks that occur at the end of the sample will lead to a larger mean squared forecast error than breaks that occur at the beginning.

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

$$
\begin{aligned}
\hat{\zeta}(\tau) &= T\tau^2 \frac{\left[\boldsymbol{x}'_{T+1}\hat{\boldsymbol{V}}_F\hat{\boldsymbol{V}}_1^{-1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\hat{\boldsymbol{V}}_2}{1-\tau} - \hat{\boldsymbol{V}}_F\right)\boldsymbol{x}_{T+1}} \\
&= \frac{\left[\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_2)\right]^2}{\boldsymbol{x}'_{T+1}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_2)\boldsymbol{x}_{T+1}} \overset{a}{\sim} \chi^2(1,\zeta)
\end{aligned} \tag{6}
$$

using a consistent estimator of the covariance matrix in the denominator.

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

$$
\hat{\zeta}(\tau) = T\frac{\left[\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\boldsymbol{x}'_{T+1}\left(\frac{\hat{\boldsymbol{V}}_1}{\tau} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau}\right)\boldsymbol{x}_{T+1}} \overset{a}{\sim} \chi^2(1,\zeta) \tag{7}
$$

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

The results of the test will, in general, differ from the outcomes of the classical Wald test on the difference between the parameter vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ for two reasons. The first is that the multiplication by $\boldsymbol{x}_{T+1}$ can render large breaks irrelevant for forecasting, or small breaks relevant. The first scenario is more likely due to the fact that breaks in the coefficients of $\boldsymbol{\beta}$ potentially cancel in the inner product $\boldsymbol{x}'_{T+1}\boldsymbol{\beta}$. 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 (5), 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. We will therefore consider breaks that are local in nature, i.e. $\boldsymbol{\beta_2} = \boldsymbol{\beta}_1 + \frac{1}{\sqrt{T}}\boldsymbol{\eta}$, rendering a finite test statistic in the asymptotic limit. Local breaks have been intensively studied in the recent literature, see for example Elliott and Müller (2007, 2014) and Elliott et al. (2015). An implication of local breaks is that no consistent estimator for the break date is available, which mimics practical situations. A consequence is that post-break parameters cannot be consistently estimated. This will deteriorate post-break window forecasts compared the full sample forecast, which, in turn, increases the break size that yields equal forecasting performance between full and post-break sample estimation windows.

We consider a general, possibly non-linear, parametric model, where parameters are estimated using GMM. The general estimation framework is that used by Andrews (1993). The observed data are given by a triangular array of random variables $\{\boldsymbol{W}_t = (\boldsymbol{Y}_t, \boldsymbol{X}_t) : 1 \leq t \leq T\}$, $\boldsymbol{Y}_t = (y_1, \ldots y_t)$ and $\boldsymbol{X}_t = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)'$. Assumptions can be made with regard to the dependency of $\boldsymbol{W}_t$ such that the results below apply to a wide 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*

$$y_t = f(\boldsymbol{\beta}_t, \boldsymbol{\delta}; \boldsymbol{X}_t, \boldsymbol{Y}_{t-1}) + \varepsilon_t \tag{8}$$

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

In the model (8), the parameter vector $\boldsymbol{\delta}$ is known to be constant. 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

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}m(\boldsymbol{W}_t,\hat{\boldsymbol{\beta}}_F,\hat{\boldsymbol{\delta}})'\hat{\gamma}\frac{1}{T}\sum_{t=1}^{T}m(\boldsymbol{W}_t,\hat{\boldsymbol{\beta}}_F,\hat{\boldsymbol{\delta}}) &= \\ \inf_{\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\delta}}}\frac{1}{T}\sum_{t=1}^{T}m(\boldsymbol{W}_t,\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\delta}})'\hat{\gamma}\frac{1}{T}\sum_{t=1}^{T}m(\boldsymbol{W}_t,\tilde{\boldsymbol{\beta}},\tilde{\boldsymbol{\delta}})\end{aligned} \tag{9}$$

where $\hat{\boldsymbol{\beta}}_F$ is estimator based on the full estimation window. We assume throughout the weighting matrix $\boldsymbol{\gamma}=\boldsymbol{S}^{-1}$ and

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

for which a consistent estimator is assumed to be available.

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

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

where $\boldsymbol{\eta}(\tau)=\boldsymbol{b}I[\tau<\tau_b]$, $\boldsymbol{b}$ is a vector of constants, and $\tau=t/T$. The pre-break parameter vector, $\boldsymbol{\beta}_1$, and the post-break parameter vector, $\boldsymbol{\beta}_2$, satisfy the partial sample moment conditions

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

Define

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

then, partial sum GMM estimators can be obtained by solving (9) with $m(\cdot)$ replaced by $\bar{m}(\cdot)$ and $\hat{\boldsymbol{\gamma}}$ replaced by

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

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

$$\begin{aligned}\hat{y}^F_{T+1}&=f(\hat{\boldsymbol{\beta}}_F,\hat{\boldsymbol{\delta}};\boldsymbol{X}_t,\boldsymbol{Y}_{t-1})\\ \hat{y}^P_{T+1}&=f(\hat{\boldsymbol{\beta}}_2,\hat{\boldsymbol{\delta}};\boldsymbol{X}_t,\boldsymbol{Y}_{t-1})\end{aligned} \tag{10}$$

Throughout, we condition on the both the exogenous and lagged dependent variables that are needed to construct the forecast. The comparison between $\hat{y}^F_{T+1}$ and $\hat{y}^P_{T+1}$ is non-standard as, under a local break, even the parameters of the model that incorporates the break are inconsistent.

In order to compare the forecasts in (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. Proofs for weak convergence of the estimators towards Gaussian processes indexed by the break date τ are given by Andrews (1993). The asymptotic distributions depend on the following matrices, for which consistent estimators are assumed to be available,

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

To simplify the notation, define

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

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

$$\begin{aligned}\sqrt{T}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1(\tau) - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\beta}}_2(\tau) - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\end{pmatrix} \to &\begin{bmatrix} \tau\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} & \boldsymbol{0} & \tau\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}} \\ \boldsymbol{0} & (1-\tau)\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} & (1-\tau)\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}} \\ \tau\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} & (1-\tau)\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} & \bar{\boldsymbol{Z}}'\bar{\boldsymbol{Z}}\end{bmatrix}^{-1} \\ &\times \begin{bmatrix} \bar{\boldsymbol{X}}'\boldsymbol{B}(\tau) + \bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}\int_0^\tau \boldsymbol{\eta}(s)ds \\ \bar{\boldsymbol{X}}'[\boldsymbol{B}(1) - \boldsymbol{B}(\tau)] + \bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}\int_\tau^1 \boldsymbol{\eta}(s)ds \\ \bar{\boldsymbol{Z}}'\boldsymbol{B}(1) + \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}\int_0^1 \boldsymbol{\eta}(s)ds\end{bmatrix}\end{aligned} \tag{11}$$

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

Define the projection matrix that projects onto the columns of $\bar{\boldsymbol{X}}$ as $\boldsymbol{P}_{\bar{X}} = \bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'$, its orthogonal complement as $\boldsymbol{M}_{\bar{X}} = \boldsymbol{I} - \boldsymbol{P}_{\bar{X}}$ and, additionally,

$$\begin{aligned}
\boldsymbol{V} &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \\
\boldsymbol{Q} &= \bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}} \\
\boldsymbol{L} &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1} \\
\tilde{\boldsymbol{Q}} &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}
\end{aligned} \tag{12}$$

The inverse in (11) can be calculated using blockwise inversion. The result is the asymptotic variance covariance matrix of $\left(\hat{\boldsymbol{\beta}}_1(\tau)', \hat{\boldsymbol{\beta}}_2(\tau)', \hat{\boldsymbol{\delta}}'\right)'$

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

Hence,

$$\begin{aligned}
\sqrt{T}(\hat{\boldsymbol{\beta}}_1(\tau) - \boldsymbol{\beta}_2) \to & \frac{1}{\tau}\left[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{B}(\tau) + \int_0^\tau \boldsymbol{\eta}(s)ds\right] \\
& - (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
\sqrt{T}(\hat{\boldsymbol{\beta}}_2(\tau) - \boldsymbol{\beta}_2) \to & \frac{1}{1-\tau}\left[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'(\boldsymbol{B}(1) - \boldsymbol{B}(\tau)) + \int_\tau^1 \boldsymbol{\eta}(s)ds\right] \\
& - (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
\sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \to & (\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1)
\end{aligned} \tag{13}$$

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

Full sample estimator: Estimators that ignore the break converge to

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \to \begin{bmatrix} \bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} & \bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}} \\ \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} & \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\boldsymbol{X}}'\boldsymbol{B}(1) + \bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}\int_0^1 \boldsymbol{\eta}(s)ds \\ \bar{\boldsymbol{Z}}'\boldsymbol{B}(1) + \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}\int_0^1 \boldsymbol{\eta}(s)ds \end{bmatrix} \tag{14}$$

Using the notation defined in (12), the inverse in (14) can be written as

$$\boldsymbol{\Sigma}_F = \begin{pmatrix} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & \boldsymbol{Q}^{-1} \end{pmatrix}$$

and, therefore,

$$\begin{aligned}
\sqrt{T}\left(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2\right) \to & (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{B}(1) + \int_0^1 \boldsymbol{\eta}(s)ds \\
& - (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
\sqrt{T}\left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}\right) \to & (\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1)
\end{aligned} \tag{15}$$

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

$$\begin{aligned}\mathrm{ACov}(\hat{\boldsymbol{\beta}}_2(\tau), \hat{\boldsymbol{\beta}}_F) &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} + (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \\ &= \mathrm{AVar}(\hat{\boldsymbol{\beta}}_F)\end{aligned}$$

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

3 Testing for a structural break

3.1 A break of known timing

Initially, we will assume that the break date is known in order to illustrate our approach. In a second step, we will extend the test to an unknown break date. Following Assumption 1, forecasts are obtained by applying a fixed, differentiable function to the $p+q$ parameters of the model conditional on a set of regressors of dimension $k = p+q$ by $(\boldsymbol{x}_{T+1}, \boldsymbol{z}_{T+1})$

$$\hat{y}_{T+1} = f(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}})$$

where we omit the dependence on the regressors for notational convenience.

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

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \to N\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{1}{\tau}\boldsymbol{V} + \tilde{\boldsymbol{Q}} & \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ \tilde{\boldsymbol{Q}} & \frac{1}{1-\tau}\boldsymbol{V} + \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & -\boldsymbol{L}' & \boldsymbol{Q}^{-1} \end{pmatrix}\right]$$

The full sample estimator is given by

$$\hat{\boldsymbol{\beta}}_F = \hat{\boldsymbol{\beta}}_2 + \tau_b(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)$$

and

$$\sqrt{T}\begin{pmatrix} \hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \to N\left[\begin{pmatrix} \tau_b(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & \boldsymbol{Q}^{-1} \end{pmatrix}\right]$$

Define $f_{\beta_2} = \frac{\partial f(\boldsymbol{\beta}_2,\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_2}$ and $f_\delta = \frac{\partial f(\boldsymbol{\beta}_2,\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}$. Using a Taylor expansion and the fact that the breaks are local in nature, we have that

$$\begin{aligned}
\sqrt{T}\left(f(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_2, \boldsymbol{\delta})\right) &= \sqrt{T}\left[f'_{\beta_2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + f'_\delta(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + O(T^{-1})\right] \\
&\to N\left(0, f'_{\beta_2}\mathrm{Var}(\hat{\boldsymbol{\beta}}_2) f_{\beta_2} + q\right) \\
\sqrt{T}\left(f(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_2), \boldsymbol{\delta})\right) &= \sqrt{T}\left[f'_{\beta_2}(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2) + f'_\delta(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + O(T^{-1})\right] \\
&\to N\left(\tau_b f'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2), f'_{\beta_2}\mathrm{Var}(\hat{\boldsymbol{\beta}}_F) f_{\beta_2} + q\right)
\end{aligned}$$

where $q = f'_\delta \mathrm{Var}(\hat{\boldsymbol{\delta}}) f_\delta + 2f'_{\beta_2}\mathrm{Cov}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) f_\delta$ and we use that, asymptotically, $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) = \mathrm{Cov}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}})$. Using previous results on the covariance matrix of the estimators, and the notation in (12), we have

$$\begin{aligned}
f'_{\beta_2}\mathrm{Var}(\hat{\boldsymbol{\beta}}_2) f_{\beta_2} &= \frac{1}{1-\tau_b} f'_{\beta_2}\boldsymbol{V} f_{\beta_2} + f'_{\beta_2}\tilde{\boldsymbol{Q}} f_{\beta_2} \\
f'_{\beta_2}\mathrm{Var}(\hat{\boldsymbol{\beta}}_F) f_{\beta_2} &= f'_{\beta_2}\boldsymbol{V} f_{\beta_2} + f'_{\beta_2}\tilde{\boldsymbol{Q}} f_{\beta_2}
\end{aligned}$$

For the expected MSFEs using $\boldsymbol{\beta}_2$ and $\boldsymbol{\beta}_F$, we have

$$\begin{aligned}
T\mathrm{E}\left[\left(f(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_2, \boldsymbol{\delta})\right)^2\right] &= \frac{1}{1-\tau_b} f'_{\beta_2}\boldsymbol{V} f_{\beta_2} + f'_{\beta_2}\tilde{\boldsymbol{Q}} f_{\beta_2} + q \\
T\mathrm{E}\left[\left(f(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_2, \boldsymbol{\delta})\right)^2\right] &= \left[\tau_b f'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2 + f'_{\beta_2}\boldsymbol{V} f_{\beta_2} + f'_{\beta_2}\tilde{\boldsymbol{Q}} f_{\beta_2} + q
\end{aligned}$$

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

$$\zeta = T(1-\tau_b)\tau_b \frac{\left[f'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}} \leq 1 \tag{16}$$

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

$$\hat{\zeta} = T(1-\tau_b)\tau_b \frac{\left[f'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\hat{\omega}} \sim \chi^2(1, \zeta) \tag{17}$$

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

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

$$\zeta = T\tau_b(1-\tau_b) \frac{\left[\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1}} \leq 1 \tag{18}$$

identical to the result in (5).

3.2 A local break of unknown timing

If the timing of the break is unknown and $\tau < \tau_b$, then the estimator of $\boldsymbol{\beta}_2$ is biased as can be seen from the last term in (13). The difference between the expected asymptotic MSFE of the partial sample forecast and that of the full sample forecast, standardized by the variance of the partial sample forecast, is

$$\begin{aligned}\Delta &= \left\{R(\hat{\boldsymbol{\beta}}_2(\hat{\tau}), \hat{\boldsymbol{\delta}}) - R(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}})\right\} / f'_{\beta_2}\boldsymbol{V} f_{\beta_2} \\ &= T\left\{\mathrm{E}\left[(f(\hat{\boldsymbol{\beta}}_2(\hat{\tau}), \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_2, \boldsymbol{\delta}))^2\right] - \mathrm{E}\left[(f(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_2, \boldsymbol{\delta}))^2\right]\right\} / f'_{\beta_2}\boldsymbol{V} f_{\beta_2} \\ &= T\left\{\mathrm{E}\left[\left(f'_{\beta_2}(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2)\right)^2\right] - \mathrm{E}\left[f'_{\beta_2}(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2)\right]^2 - f'_{\beta_2}\mathrm{Var}(\hat{\boldsymbol{\beta}}_F) f_{\beta_2}\right\} / f'_{\beta_2}\boldsymbol{V} f_{\beta_2}\end{aligned}$$

where $R(\hat{\boldsymbol{\theta}})$ is the asymptotic MSFE under parameter estimates $\hat{\boldsymbol{\theta}}$. The derivations are provided in Appendix A.1. Using (13) and (15) we obtain

$$\begin{aligned}\Delta &= \frac{R(\hat{\boldsymbol{\beta}}_2(\hat{\tau}), \hat{\boldsymbol{\delta}}) - R(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}})}{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}} \\ &= \mathrm{E}\left\{\left[\frac{1}{1-\hat{\tau}}\frac{f'_{\beta_2}\boldsymbol{V}\bar{\boldsymbol{X}}'\left(\boldsymbol{B}(1) - \boldsymbol{B}(\hat{\tau})\right)}{\sqrt{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}}} + \frac{1}{1-\hat{\tau}}\int_{\hat{\tau}}^{1}\frac{f'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}}}ds\right]^2\right\} \\ &\quad - \left(\int_0^1 \frac{f'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}}}ds\right)^2 - 1 \qquad (19)\end{aligned}$$

Note that (19) makes no assumption about the form of the instability, which is governed by $\boldsymbol{\eta}(\tau)$. Define $J(\tau) = \int_{\tau}^{1} \frac{f'_{\beta_2}\boldsymbol{\eta}(s)}{\sqrt{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}}}ds$ and note that for fixed f'_{β_2} the continuous mapping theorem yields $\frac{f'_{\beta_2}\boldsymbol{V}\bar{\boldsymbol{X}}'[\boldsymbol{B}(1)-\boldsymbol{B}(\tau)]}{\sqrt{f'_{\beta_2}\boldsymbol{V} f_{\beta_2}}} = B(1) - B(\tau)$, where $B(\cdot)$ is a one-dimensional Brownian motion. Then

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

which could be used to test whether the use of a moving window will outperform an expanding window under various forms of parameter instability. The expectation simplifies if the size of the moving window is exogenously set to some fraction of the total number of observations.

Under a structural break, $\boldsymbol{\eta}(\tau) = \boldsymbol{b}I[\tau < \tau_b]$

$$\Delta = \mathrm{E}_{f(\hat{\tau})}\left\{\left[\frac{1}{1-\hat{\tau}}\left(B(1) - B(\hat{\tau})\right) + \theta_{\tau_b}\frac{\tau_b - \hat{\tau}}{1-\hat{\tau}}I[\hat{\tau} < \tau_b]\right]^2\right\} - \theta_{\tau_b}^2\tau_b^2 - 1 \quad (20)$$

where $\theta_{\tau_b} = \frac{f'_{\beta_2}\boldsymbol{b}}{\sqrt{f'_{\beta_2}\boldsymbol{V}f_{\beta_2}}}$. The subscript τ_b is added for notational purposes in the following section. This is related to the standardized break size ζ from the previous section by

$$\theta_{\tau_b} = \sqrt{\frac{\zeta}{\tau_b(1-\tau_b)}} \quad (21)$$

If $\hat{\tau} = \tau_b$, the critical break size of the previous section is obtained. However, under an unknown break date, in general, $\hat{\tau} \neq \tau_b$ and (20) cannot immediately be used for testing purposes.

It is, however, interesting to observe that since Δ is symmetric around $\theta_{\tau_b} = 0$, $\Delta > 0$ for $\theta_{\tau_b} = 0$, and (20) quadratically decreases away from $\theta_{\tau_b} = 0$, there is a break size $|\theta_{\tau_b}|$ for each τ_b for which $\Delta = 0$. Numerical results depicted in Figure 11 clearly show that equal predictive accuracy is attained for a unique break size. This makes it an excellent candidate test statistic. Analogous to the case where the break date is known, we simply use the Wald test statistic

$$W(\tau) = T\frac{\left[f'_{\beta_2}(\hat{\boldsymbol{\beta}}_2(\tau) - \hat{\boldsymbol{\beta}}_1(\tau))\right]^2}{f'_{\beta_2}\left(\frac{\hat{\boldsymbol{V}}_1}{\tau} - \frac{\hat{\boldsymbol{V}}_2}{1-\tau}\right)f_{\beta_2}} \quad (22)$$

We will show below that for sufficiently small size the test statistic in (22) identifies the true break date up to a constant that vanishes with decreasing size, which establishes a weak form of optimality of the sup-Wald test even when the break size for which $\Delta = 0$ is non-constant over τ_b.

Since the function f'_{β_2} is fixed, the results in Andrews (1993) and the continuous mapping theorem show that under local alternatives $W(\tau)$ in (22) converges to

$$\begin{aligned} Q^*(\tau) &= \left(\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}} + \sqrt{\frac{1-\tau}{\tau}}\int_0^\tau \eta(s)ds - \sqrt{\frac{\tau}{1-\tau}}\int_\tau^1 \eta(s)ds\right)^2 \\ &= \left(\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1-\tau)}} + \mu(\tau;\theta_{\tau_b})\right)^2 \end{aligned} \quad (23)$$

The first term of (23) is a self-normalized Brownian bridge with expectation zero and variance equal to one. For a fixed break date, $Q^*(\tau)$ follows a non-central χ^2 distribution with one degree of freedom and non-centrality parameter $\mu(\tau;\theta_{\tau_b})^2$. For the structural break model, we have

$$\mu(\tau;\theta_{\tau_b}) = \theta_{\tau_b}\left[\sqrt{\frac{1-\tau}{\tau}}\tau_b I[\tau_b < \tau] + \sqrt{\frac{\tau}{1-\tau}}(1-\tau_b)I[\tau_b \geq \tau]\right] \tag{24}$$

For the optimality results below the following assumption is made with regard to the non-centrality parameter

Assumption 2 *The function $\mu(\tau;\theta_{\tau_b})$ is maximized (or minimized) if and only if $\tau = \tau_b$*

Under this assumption, for a sufficiently small nominal size, rejections are found only for break locations that are close to τ_b. For the structural break model it is easy to verify that Assumption 2 holds. The extremum value is given by $\mu(\tau_b;\theta_{\tau_b}) = \theta_{\tau_b}\sqrt{\tau_b(1-\tau_b)} = \zeta^{1/2}$.

3.3 Weak optimality

In this section, we will show that the Wald test is weakly optimal when the null hypothesis (and consequently the critical values) depend on the unknown break date. Using arguments of Piterbarg (1996), we show that under a general form of instability, only points in a small neighborhood around the maximum instability point τ_b contribute to the probability of exceeding a constant boundary, u, in the limit where the size of the test tends to zero. In a second step we extend the analysis by considering a null hypothesis that depends on an unknown and weakly identified parameter τ_b. In this case, critical values will also depend τ_b. If the critical values vary sufficiently slow with τ_b, then using an estimate $\hat{\tau}$ leads to a weakly optimal test in the sense that it has larger or equal power compared to the test that knows τ_b in the limit where the size of the test goes to zero. In Section 4 we provide evidence that the power of the test under estimated τ_b is close to that of the infeasible test under known τ_b for values of u corresponding to standard size values.

3.3.1 Location concentration

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

Lemma 1 *Suppose $Z(\tau)$ is a symmetric Gaussian process, i.e. $P(Z(\tau) > u) = P(-Z(\tau) > u)$, then*

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

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

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

$$\begin{aligned} P(Z(\tau) + \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) &= P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in I) \\ P(-Z(\tau) - \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) &= P(Z(\tau) > u + |\mu(\tau; \theta_{\tau_b})|, \tau \in I) \end{aligned} \tag{25}$$

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

$$\begin{aligned} P(-Z(\tau) - \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) &= P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in I) \\ P(Z(\tau) + \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) &= P(Z(\tau) > u + |\mu(\tau; \theta_{\tau_b})|, \tau \in I) \end{aligned} \tag{26}$$

The bounds in the second lines of (25) and (26) 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 $u \to \infty$

$$P\left(\sup_{\tau \times c} [Z(\tau) + \mu(\tau; \theta_{\tau_b})]c > u\right) = P\left(\sup_{\tau \in I} Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|\right)[1 + o(1)] \tag{27}$$

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

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

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

See Hüsler (1990). 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 2 *Suppose $Z(\tau)$ is a locally stationary process with local covariance function $C(\tau)$ then if $\delta(u)u^2 \to \infty$ and $\delta(u) \to 0$ as $u \to \infty$*

$$P\left(\sup_{[\tau, \tau + \delta(u)]} Z(t) > u\right) = \frac{1}{\sqrt{2\pi}} \delta(u) u \exp\left(-\frac{1}{2}u^2\right) C(\tau) \tag{28}$$

Proof: see Hüsler (1990).

Given the above results, we can state the following theorem.

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* $|\mu(\tau;\theta_{\tau_b})|$ *is a function that attains its unique maximum when* $\tau = \tau_b$*, then as* $u \to \infty$

$$P\left(\sup_{\tau\in I} Q^*(\tau) > u^2\right) = P\left(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})| \text{ for some } \tau \in I_1\right)(1 + o(1))$$

where $I = [\tau_{\min}, \tau_{\max}]$*,* $I_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$ *and* $\delta(u) = u^{-1}\log^2 u$ *.*

Proof: We start by noting that for $\tau \in I = [\tau_{\min}, \tau_{\max}]$

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

where the supremum is taken jointly over $\tau \in I$ and c. The last equality follows from Lemma 1. Now we proceed along the lines of Piterbarg (1996).

As in Lemma 2, consider a region close to τ_b defined by $I_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$. In I_1, the minimum value of the boundary is given by

$$\underline{b} = \inf_{\tau\in I_1}[u - \mu(\tau;\theta_{\tau_b})] = u - |\mu(\tau_b;\theta_{\tau_b})| \tag{29}$$

so that

$$\begin{aligned}
P_{I_1} &= P(Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})| \text{ for some } \tau \in I_1) \\
&\leq P(Z(\tau) > \underline{b} \text{ for some } \tau \in I_1) \\
&= 2\delta(u)\underline{b}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\underline{b}^2\right)C(\tau_b) \\
&= \frac{2\delta(u)}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\underline{b}^2 + \log\underline{b}\right)C(\tau_b)
\end{aligned}$$

where the third line follows from (28).

Define the region outside of I_1 as $I_A = I\backslash I_1$. Then in I_A, the minimum value of the boundary is given by

$$\underline{b}_A = u - |\mu(\tau_b + \delta(u);\theta_{\tau_b})| \tag{30}$$

Taking a Taylor expansion of $\mu(\tau_b + \delta(u); \theta_{\tau_b})$ around $\delta(u) = 0$ gives

$$\mu(\tau_b + \delta(u); \theta_{\tau_b}) = \mu(\tau_b; \theta_{\tau_b}) + \gamma\delta(u) + O\left[\delta(u)^2\right] \tag{31}$$

where $\gamma = \frac{\partial\mu(\tau;\theta_{\tau_b})}{\partial\tau}\Big|_{\tau=\tau_b}$. Then $\underline{b}_A = \underline{b} + \gamma\delta(u)$ and

$$\begin{aligned} P_{I_A} &= P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in I_A) \\ &\le P(Z(\tau) > \underline{b}_A \text{ for some } \tau \in I_A) \\ &\le \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\underline{b}^2 - \underline{b}\gamma\delta(u) - \frac{1}{2}\gamma^2\delta(u)^2 + \log(\underline{b} + \gamma\delta(u))\right)\overline{C} \end{aligned} \tag{32}$$

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

$$\sum_{I_k \in I_A} C(k\delta(u))\delta(u) \overset{\delta(u)\to 0}{=} \int_{I_A} C(\tau)d\tau \le \int_I C(\tau)d\tau = \overline{C} < \infty \tag{33}$$

with I_k representing non-overlapping intervals of width $\delta(u)$ such that $\bigcup\limits_{k=2}^{\infty} I_k = I_A$ and $k\delta(u) \in I_k$

Compare (32) to the probability of a test with a known break date to exceed the critical value

$$P_0 = P(Z(\tau_b) > u - |\mu(\tau_b; \theta_{\tau_b})|) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\underline{b}^2 - \log(\underline{b})\right) \tag{34}$$

where we use that

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

Ignoring the lower order terms $-\frac{1}{2}\gamma^2\delta(u)^2 + \log(\underline{b} + \gamma\delta(u))$, equation (32) contains an extra term $\exp(-\underline{b}\gamma\delta(u))$ compared to (34). Using (29), this implies that $P_{I_A} = o(P_0)$ if

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

Subsequently, if

$$\delta(u) = u^{-1}\log^2(u) \tag{35}$$

then all intervals outside of I_1 contribute $o(P_0)$ to the probability of crossing the boundary u. Under (35), we have that for P_{I_1} as $u \to \infty$

$$\begin{aligned} P_{I_1} \le P_I &\le P_{I_1} + P_{I_A} \\ &\le P_{I_1} + o(P_0) \end{aligned}$$

We now only need to note that

$$\begin{aligned} P_{I_1} &= P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in I_1) \\ &\ge P(Z(\tau_b) > u - |\mu(\tau_b; \theta_{\tau_b})|) = P_0 \end{aligned}$$

To conclude that

$$P\left(\sup_{\tau\in I} Z(\tau) > u - |\mu(\tau;\theta_{\tau_b})|, \tau \in I\right) = P_{I_1}(1+o(1)) \text{ as } u \to \infty$$

which completes the proof. ■

Note that, in (32), the term $\exp(\underline{b}\delta(u))^{-\gamma}$ ensures that $P_{I_A} = o(P_1)$. In the structural break model, we see that (31) 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 size. Therefore, for a sufficiently large break, asymptotic optimality results are expected to extend to the practical case when u is finite. The simulations of asymptotic power presented in Section 4 confirm this.

Corollary 1 (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.*

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

This corollary implies that post-test parameter inference after a rejection of the null is in fact standard.

3.3.2 Weak optimality under location dependent boundaries

While the location concentration is essential to our proof, the problem we consider is further complicated by the fact that the size under the null hypothesis depends on the unknown break date. This translates into critical values that will also depend on the unknown break date. Theorem 1 indicates that simply plugging in the estimate of the break date that maximizes the Wald statistic could be a viable strategy to obtain a well behaved test. In fact, this strategy is weakly optimal in the sense of Andrews (1993) if the following assumption is satisfied.

Assumption 3 (Slowly varying critical values) *Suppose we test using a sequence of critical values that control size for every τ_b, i.e. $u = u(\tau_b)$ such that*

$$P\left(\sup_{\tau} Q^*(\tau) > u(\tau^*) \,\middle|\, \tau_b = \tau^*\right) = \alpha$$

then $u(\tau) - \mu(\tau;\theta_{\tau_b})$ should have a unique minimum on $I_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$ at $\tau = \tau_b$.

Suppose that $u(\tau_b)$ a differentiable function with respect to τ_b, then a sufficient condition is that the critical values are slowly varying with τ_b in comparison with the derivative of the function $\mu(\tau;\theta_{\tau_b})$ with respect to τ on the interval I_1, i.e.

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

In the structural break model, the derivative on the right hand side occurs in (31) as $\gamma = \frac{\partial \mu(\tau;\theta_{\tau_b})}{\partial \tau} = \theta_{\tau_b}\frac{1}{\sqrt{\tau_b(1-\tau_b)}}$. The slowly varying condition 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 size. For the break size we know from Section 2 that $\theta_{\tau_b}\sqrt{\tau_b(1-\tau_b)} \geq 1$, where the equality holds if the break date is known with certainty. Then

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

A sufficient condition for the slowly varying assumption is therefore

$$\left|\frac{\partial u(\tau_b)}{\partial \tau_b}\right| \leq \frac{1}{\tau_b(1-\tau_b)} \tag{36}$$

which will show to hold once critical values have been obtained.

We now provide a result on the optimality of the test under Assumption 3. Let $v(\tau_b)$ denote the critical values of the optimal test conditional on the break location, i.e. $P(Q^*(\tau_b) > v(\tau_b)^2) = \alpha$, where $v(\tau_b)$ is the critical value of the test conditional on the true break date, τ_b.

Lemma 3 (Convergence of critical values) *Let $u(\tau)$ be the critical value that controls size for a given τ and let $v(\tau_b)$ be the critical value of the test on the true break date, then $u(\tau_b) - v(\tau_b) \to 0$.*

Proof: By definition of the critical values

$$P\left[\sup_\tau Q^*(\tau) > u(\tau_b)^2\right] = P\left[Z(\tau) > u(\tau_b) - |\mu(\tau;\theta_{\tau_b})| \text{ for some } \tau \in I_1\right] = \alpha$$
$$P\left[Q^*(\tau_b) > v(\tau_b)^2\right] = P\left[Z(\tau_b) > v(\tau_b) - |\mu(\tau_b;\theta_{\tau_b})|\right] = \alpha$$

Since τ in the first line is contained in I_1, we have by a Taylor series expansion of $\mu(\tau;\theta_{\tau_b})$ around τ_b that $\max|\mu(\tau;\theta_{\tau_b})| - |\mu(\tau_b;\theta_{\tau_b})| = O[\delta(u)]$ and consequently, $\max u(\tau_b) - v(\tau_b) = O(\delta(u))$. Since $\delta(u) \to 0$ as $u \to \infty$, the difference in the critical values $u(\tau_b) - v(\tau_b) \to 0$ as $u \to \infty$. ■

Theorem 2 (Weak optimality) *Under a slowly varying boundary*

$$\begin{aligned}
&P_{H_a}\left[\sup_\tau Q^*(\tau) > u(\hat{\tau})^2\right] - P_{H_a}(Q^*\left[\tau_b) > v(\tau_b)^2\right] \\
&\geq P_{H_a}\left[Q^*(\tau_b) > u(\tau_b)\right] - P_{H_a}(Q^*\left[\tau_b) > v(\tau_b)^2\right] \\
&= 0
\end{aligned} \tag{37}$$

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

Proof: As before

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

Under the slowly varying assumption, $u(\hat{\tau})-\mu(\hat{\tau};\theta_{\tau_b})$ has a unique minimum on I_1 at $\hat{\tau} = \tau_b$. Taking the supremum therefore necessarily leads to at least as many exceedances as considering $\tau = \tau_b$ alone, which proves the inequality in (37). The last line of (37) is follows from Lemma 3. ■

3.3.3 Testing with critical values independent of the break date

A test based on the Wald statistic (22) requires critical values that dependent on the estimated break date. It is, however, straightforward to derive a test statistic with critical values that are independent of the break date in the limit where $u \to \infty$.

Corollary 2 *A test statistic with critical values that are independent of τ_b for $u \to \infty$ is given by*

$$S(\hat{\tau}) = \sup_{\tau\in I} \sqrt{T}\frac{\left|f'_{\beta_2}\left(\hat{\boldsymbol{\beta}}_2(\tau) - \hat{\boldsymbol{\beta}}_1(\tau)\right)\right|}{\sqrt{f'_{\beta_2}\left(\frac{\hat{\boldsymbol{V}}_1}{\tau} + \frac{\hat{\boldsymbol{V}}_2}{1-\tau}\right) f_{\beta_2}}} - |\mu(\hat{\tau};\theta_{\hat{\tau}})| \tag{38}$$

where $\hat{\tau}$ maximizes the first term of S or, equivalently, the Wald statistic (22).

Proof: 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

$$\begin{aligned} P(S(\hat{\tau}) > u) &= P\left(\sup_{I_1} |Z(\tau) + \mu(\tau;\theta_{\tau_b})| - |\mu(\hat{\tau};\theta_{\hat{\tau}})| > u\right) \\ &= P(Z(\hat{\tau}) > u - |\mu(\hat{\tau};\theta_{\tau_b})| + |\mu(\hat{\tau};\theta_{\hat{\tau}})|) \end{aligned}$$

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 $u \to \infty$. ■

3.4 Optimal weights or shrinkage forecasts

Pesaran et al. (2013) derive optimal weights for observations in the estimation sample such that, in the presence of a structural break, the MSFE of the one step ahead forecast is minimized. These optimal weights are derived under the assumption of a known break date and break size. 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 the forecast can be written as a weighted average of pre-break and post-break parameter estimates.

We can therefore write the optimally weighted forecast as a convex combination of the forecasts from pre-break observations and post-break observations

$$\hat{y}^S_{T+1}(\tau) = \omega \boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_1 + (1-\omega)\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2$$

where the optimal forecast is denoted with subscript S as we will see below that it is equal to a shrinkage forecast that shrinks the post-break sample based forecast in the direction of the full sample based forecast.

The expected mean squared error is

$$\begin{aligned}
\mathrm{E}\left[T\left(\hat{y}^S_{T+1} - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2\right)^2\right] &= \mathrm{E}\left[T\left(\omega \boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_1 + (1-\omega)\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2 - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2\right)^2\right] \\
&= \omega^2 T\left[\boldsymbol{x}'_{T+1}\left(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\right)\right]^2 + \\
&\quad + \omega^2 \boldsymbol{x}'_{T+1}\left(\frac{1}{\tau_b} + \frac{1}{1-\tau_b}\right)\boldsymbol{V}\boldsymbol{x}_{T+1} \\
&\quad - 2\omega\frac{1}{1-\tau_b}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} + \frac{1}{\tau_b}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1}
\end{aligned} \tag{39}$$

see Appendix A.4 for details.

Maximizing (39) with respect to ω yields

$$\omega^* = \tau_b\left[1 + T\frac{\left(\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right)^2}{\boldsymbol{x}'_{T+1}\left(\frac{1}{\tau_b} + \frac{1}{1-\tau_b}\right)\boldsymbol{V}\boldsymbol{x}_{T+1}}\right]^{-1} \tag{40}$$

where the denominator contains the Wald statistic derived, W, in (5) and (18).

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

$$\begin{aligned}
\hat{y}^S_{T+1} &= \omega \boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_1 + (1-\omega)\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2 \\
&= \frac{\omega}{\tau_b}\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F + \left(1 - \frac{\omega}{\tau_b}\right)\hat{\boldsymbol{\beta}}_2
\end{aligned}$$

and the optimal weight on the full sample forecast is

$$\omega^*_F = \frac{\omega^*}{\tau_b} = \frac{1}{1 + W(\tau_b)} \tag{41}$$

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

The empirical results in Pesaran et al. (2013) suggest that uncertainty around the break date substantially deteriorates the accuracy of the optimal weights forecast in applications. As a consequence, Pesaran et al. (2013) derive robust optimal weight by integrating over the break dates, which yield substantially more accurate forecasts. 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 impact the weights. If this uncertainty is not taken into account, the weight on the post-break forecast will be too high. However, the lack of analytic expressions for the break date uncertainty complicates an analytic weighting scheme. Alternatively, this uncertainty can be taken into account by testing whether the break date uncertainty is small enough to justify using the shrinkage forecast.

As the Wald statistic in (41) requires the true break date, consider the shrinkage forecast for a general value of τ

$$\begin{aligned}\hat{y}^S_{T+1}(\tau) &= \frac{1}{1+W(\tau)}\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F + \frac{W(\tau)}{1+W(\tau)}\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2(\tau) \\ &\to \frac{1}{1+Q^*(\tau)}\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_F + \frac{Q^*(\tau)}{1+Q^*(\tau)}\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2(\tau)\end{aligned} \tag{42}$$

where the last line holds by the continuous mapping theorem. The asymptotic expressions for $\hat{\boldsymbol{\beta}}_2$ and $\hat{\boldsymbol{\beta}}_F$ are provided in (13) and (15). The difference in MSFE between the shrinkage forecast and the full sample forecast depends on the distribution of the break date

$$\begin{aligned}\Delta_s \;=\; & \mathrm{E}_{\hat{\tau}}\left[\left(\frac{1}{1+Q^*(\hat{\tau})}\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_2) + \frac{Q^*(\hat{\tau})}{1+Q^*(\hat{\tau})}\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_2(\hat{\tau})-\boldsymbol{\beta}_2)\right)^2\right] \\ & -\mathrm{E}\left[\left(\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_2)\right)^2\right]\end{aligned} \tag{43}$$

where we solve for $\Delta_s = 0$ numerically to obtain the break size that corresponds to equal predictive accuracy. Numerical results in Appendix A.3 show that equal predictive accuracy is associated with a unique break size.

4 Simulations

4.1 Asymptotic analysis

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

4.1.1 Implementation

We simulate (23) with (24) for different combinations of the break date and break size $\{\tau_b, \theta_{\tau_b}\}$. Here, we focus on $\tau_b = \{\tau_{\min}, \tau_{\min} + \delta, \ldots, \tau_{\max}\}$ where $\tau_{\min} = 0.15$, $\tau_{\max} = 1 - \tau_{\min}$ and $\delta = 0.01$. Additional results for a wider grid with $\tau_{\min} = 0.05$ are presented in Appendix B. For the break size parameter 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 $n = 1,000$ equally spaced parts, generating $\epsilon_i \sim N(0, 1)$ and $B(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n\tau} \epsilon_i$, see, for example, Bai and Perron (1998).

By maximizing (23) we obtain a distribution of the estimated break date $\hat{\tau}$ that can be used to evaluate (19). To approximate the expectation, we use 50,000 repetitions for each break date and break size. For each value of τ_b, a break size θ_{τ_b} is obtained for which the full sample forecast and the post-break forecast yield equal predictive accuracy using (19). This translates the null hypothesis of equal predictive accuracy into a null hypothesis regarding the break size conditional of the break date τ_b. By simulating under the null hypothesis for each τ_b, we obtain critical values that are dependent on τ_b. If the break date is estimated with sufficient accuracy, these critical values can be used for testing without correction. The size of the breaks that we find under the null hypothesis suggest that the estimated break date will, indeed, be quite accurate.

4.1.2 Post-break forecast versus full-sample forecast: break size for equal forecast accuracy

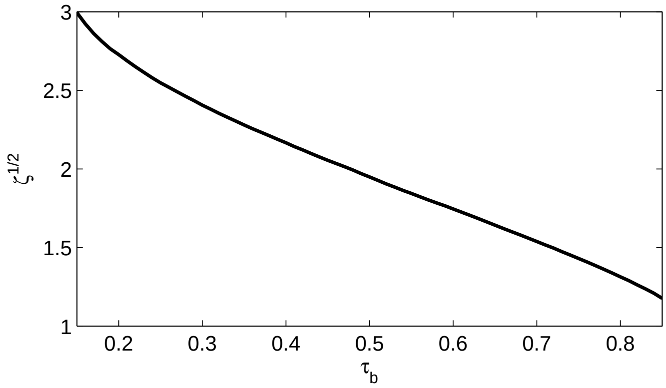
The break size for which the full sample and the post-break sample achieve equal predictive accuracy can be simulated using (19) as outlined above. Figure 1 shows the combinations of break size and break date for which equal predictive accuracy is obtained. The break size is given in units of the standardized break size,

$$\zeta^{1/2} = \sqrt{T(1-\tau_b)\tau_b} \frac{\boldsymbol{f}'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)}{\sqrt{\boldsymbol{f}'_{\beta_2} \boldsymbol{V} \boldsymbol{f}_{\beta_2}}} \tag{44}$$

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 size is substantially larger than the break size under a known break date, which yielded

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



Note: The graph shows the standardized break size, $\zeta^{1/2}$, in (44) for which the forecasts based on the post-break sample and the full sample achieve the same MSFE, that is, Δ in (19) equals zero.

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

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

Note: Reported are critical values and size for, first, W, the Wald test statistic (17) and, second, S, the test statistic (38), which is independent of τ_b when the nominal size tends to zero.

$\zeta^{1/2} = 1$. This illustrates the increase in the MSFE of the post-break sample forecast due to the fact that the break date needs to be estimated. The importance of this effect is clear. If a break occurs in the beginning of the sample, then we choose for the post-break forecast if the break size larger than three standard deviations. For breaks that occur closer to the end of the sample, this break size uniformly decreases. This provides further evidence for the intuition that breaks that occur at the end of the sample are the main reason for forecast failure.

4.1.3 Critical values, size and power

After finding the break size for which the post sample forecast and the full sample forecast yield equal predictive accuracy, we can compute critical values for both the Wald-type test statistic, W, in (22) and the α-asymptotic statistic, S, in (38) for a grid of break dates τ_b. Condition (36), which is required for the weak optimality result does hold for all τ_b—details are available in Appendix A.2.

The third line of the right panel of Table 1 shows that the test has the correct size for $\alpha = 0.01$. For $\alpha = 0.05$ and 0.1 size is still very close to the asymptotic size, only at the beginning and the end of the sample some size distortion occurs. However, using the corrected test statistic (38) largely remedies these size distortions.

The critical values are given in the left panel of Table 1. Critical values for a finer grid of the true break date can be found in Appendix B. The large break size that yields equal forecast accuracy implies a major increase in critical values when using the Wald test statistic (22), compared to the standard values of Andrews (1993). For a nominal size of $[0.10, 0.05, 0.01]$ the critical values in Andrews are equal to $[7.17, 8.85, 12.35]$.

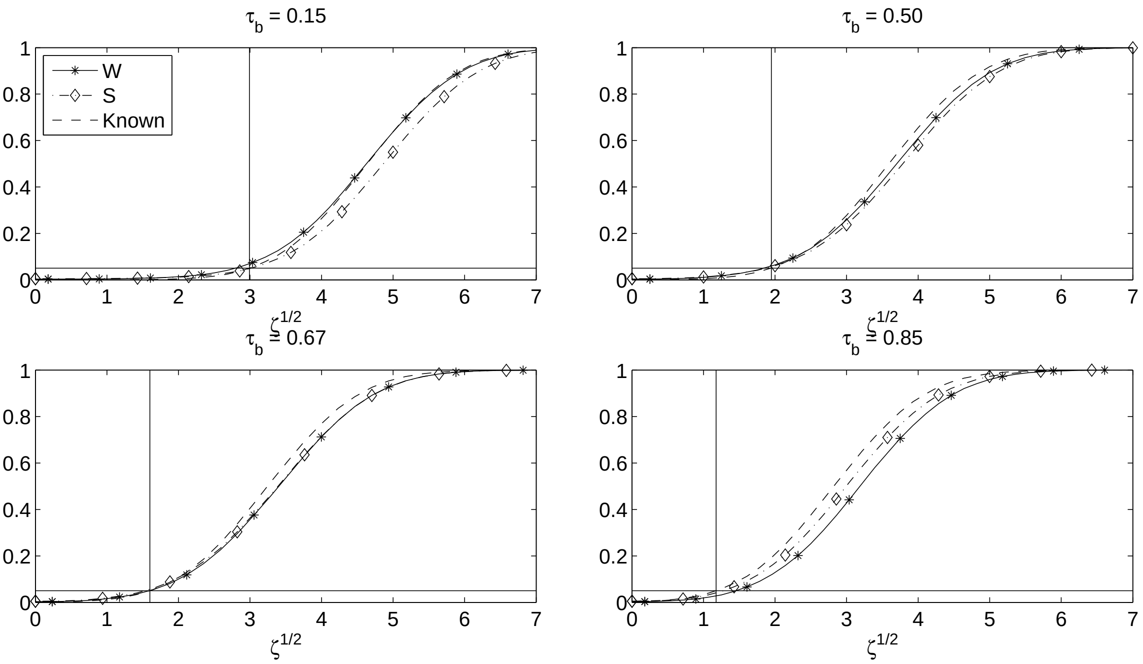
The critical values for the α-asymptotic test statistic, S, in (38) are independent of $\hat{\tau}$ in the limit where $\alpha \to 0$. Under a known break date, critical values would be from a one-sided normal distribution, that is, they would be $[1.64, 2.33, 2.58]$ for nominal size of $[0.10, 0.05, 0.01]$. The critical values for the corrected test, S, in (38) vary substantially less over $\hat{\tau}$ than those for the Wald statistic, W, in (22). The results in Section 3.3 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 observe in Table 1.

Given that the break sizes 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 size 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.

4.1.4 Shrinkage forecast versus full-sample forecast

We now turn to the shrinkage forecast of Section 3.4. Figure 3 shows the combination of τ_b and break size for which the shrinkage forecast and full sample forecast that weights observations equally have the same MSFE, which is represented by the solid line in the graph. For comparison, the dashed line gives the combination of post-break forecast and full sample forecast that have the same MSFE. It can be seen that the break size for

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

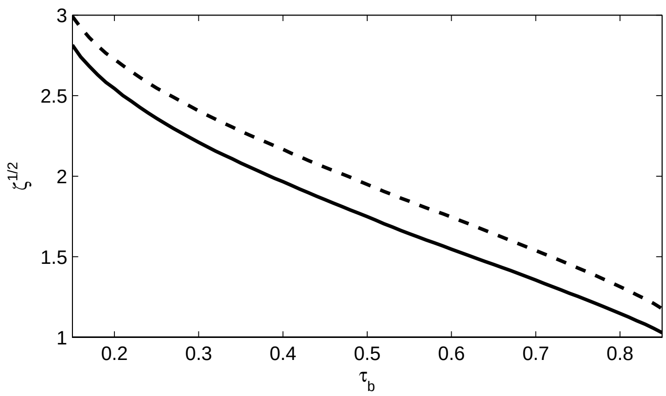


Note: The plots show the power for tests at a nominal size of $\alpha = 0.05$ with the null hypothesis given by the break size depicted in Figure 1. The panels show power for different values of the (unknown) break date. The power of infeasible test conditional on the true break date is given as the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

equal forecast performance for the shrinkage forecast is lower than for the post-break sample forecast. This implies that the shrinkage forecast is more precise than the post-break forecast for smaller break sizes for a given break date. However, the difference is relatively small and breaks need to be quite large before the shrinkage estimator is more precise than the full sample estimator.

In order to determine whether to use the shrinkage forecast, critical values can be obtained in a similar fashion as before and are presented in Table 2. Again, the size is close to the theoretical size with small size disturbances when using W, which are largely remedied when using S. Critical values on a finer grid of the true break date are presented in Appendix B. Figure 4 displays the power curves of the tests that compare the shrinkage forecast and the full sample, equal weights forecast. Since, the break sizes for equal forecast performance are similar to the post-break sample forecast,

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



Note: The solid line shows the standardized break size for which the shrinkage forecast (42) achieves the same MSFE as the full sample forecast, in which case (43) equals zero. For comparison, the dashed line shows the break size for which the post-break forecast and the full sample forecast achieve equal MSFE.

it is not surprising that the properties in terms of size and power of the tests for the shrinkage forecast are largely the same as those for the post-break forecast.

4.1.5 Shrinkage forecast versus the post-break forecast

In addition to comparing post-break sample and shrinkage forecasts to the full sample forecast, we can investigate the break sizes that leads to equal forecast performance of the post-break forecast and the shrinkage forecast. Figure 5 plots the ratio of the MSFE of the shrinkage forecast over that of the post-break forecast. Nearly for all break sizes and dates, the shrinkage forecast outperforms the post-break forecast. Only when the break occurs at the end of the sample and is relatively large, the post-break forecast is slightly more accurate. This suggests that one can improve over the post-break estimator in a wide range of settings.

4.2 Finite sample analysis

4.2.1 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 (17) and on the S-statistic (38). Next,

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

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

Note: Reported are critical values and size when testing for equal MSFE of the shrinkage forecast (42) and the full sample forecast using, first, W, the Wald test statistic in (17) and, second, S, the test statistic (38) that is independent of τ_b when the nominal size tends to zero.

we consider the same test statistics but now test for equal predictive accuracy between the shrinkage forecast (42) and the full-sample, equal weighted forecast. All tests are carried out at a nominal size $\alpha = 0.05$, using sample sizes of $T = \{120, 240, 480\}$ and break dates $\tau_b = [0.15, 0.25, 0.50, 0.75, 0.85]$. Parameter estimates are obtained by least squares, and the results are based on 10,000 repetitions.

The DGP is given by

$$y_t = \mu_t + \rho y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2) \tag{45}$$

where $\sigma^2 = 1$ and

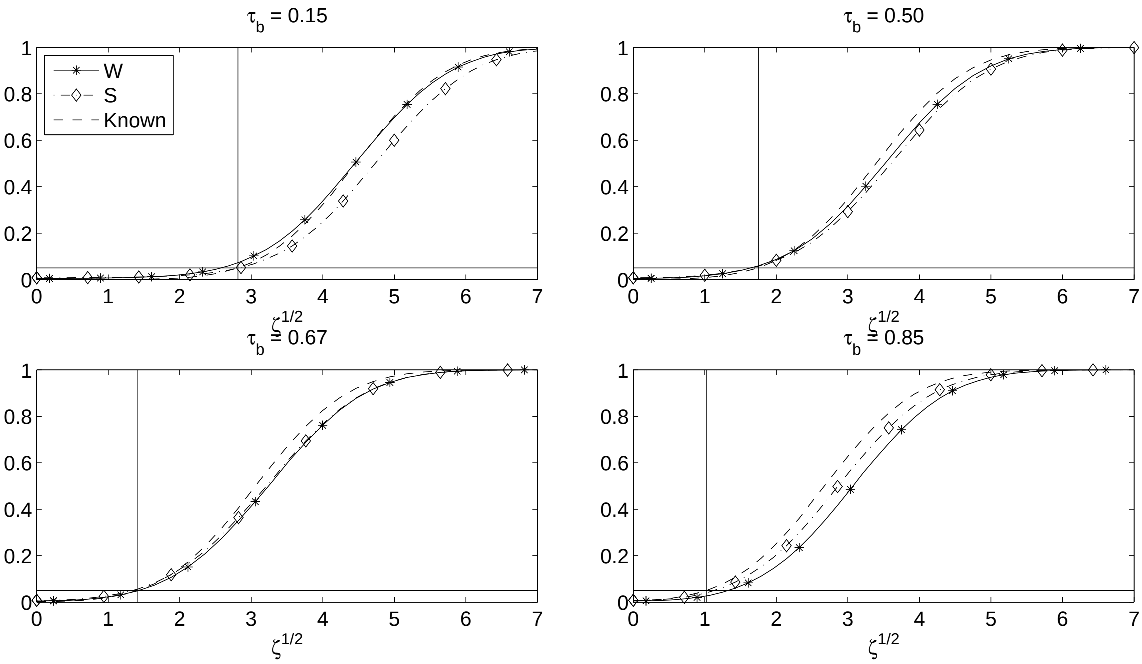
$$\mu_t = \begin{cases} \mu_1 & \text{if } t \leq \tau_b T \\ \mu_2 & \text{if } t > \tau_b T \end{cases}$$

We set $\mu_1 = -\mu_2$ and $\mu_1 = \frac{1}{2\sqrt{T}}\zeta^{1/2}(\tau_b) + \frac{1}{2}\frac{\lambda}{\sqrt{T\tau_b(1-\tau_b)}}$. To investigate the finite sample size of the tests, we choose $\lambda = 0$ which yields the asymptotic the break size from Figure 1. To investigate power, we choose $\lambda = \{1, 2\}$. The influence of the degree of persistence on the results is analyzed by varying $\rho = \{0.0, 0.3, 0.6, 0.9\}$.

4.2.2 Results

The results in Table 3 show that for models with low and moderate persistence, that is, $\rho = 0.0$ or 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. When persistent increases to $\rho = 0.9$, some size distortions become apparent for $T = 120$. This does, however, diminish as T increases. These size distortions are similar for W and S and are the result of the small effective sample

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



Note: The plots show asymptotic power curves when testing for equal predictive accuracy between the shrinkage forecast (42) and the full-sample forecast using the break size depicted in Figure 3 for different values of the break date τ_b. For more information, see the footnote of Figure 2.

size in this setting. Power increases with λ. For $T = 120$ it is slightly larger when the break is in the middle of the sample but this effect disappears with increasing T. Overall, differences between W and S are small.

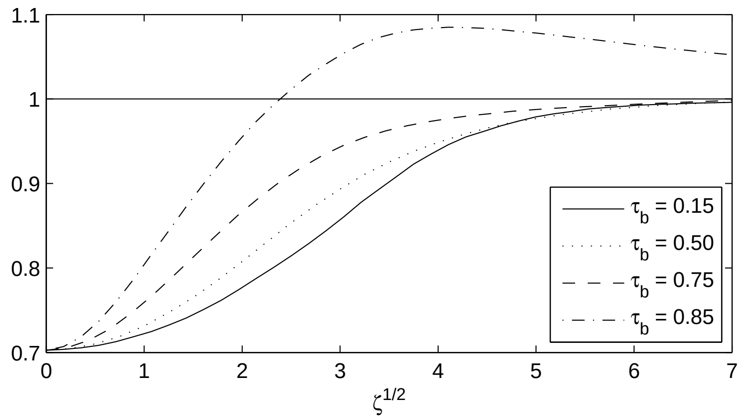
The results for the tests that compare the shrinkage forecast against the full sample, equal weights forecast in Table 4 are very similar to the results for the test with the post-break sample forecast under the alternative. Size is very close to the nominal size for large effective sample sizes and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

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

5 Application

We investigate the importance of structural breaks for 130 macroeconomic and financial time series from the St. Louis Federal Reserve database, which

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



Note: The graph shows the relative performance of the shrinkage forecast (42) and the post-break sample forecast as a function of the standardized break size $\zeta^{1/2}$ for different values of the break date τ_b. The horizontal solid line corresponds to equal predictive accuracy. Values below 1 indicate that the shrinkage forecast is more precise.

is a monthly updated database. The data are described by McCracken and Ng (2015), who suggest various transformations are applied to render the series stationary and to deal with discontinued series or changes in classification. In the vintage used here, the data start in 1959M01 and end in 2015M10. After the transformations, all 130 series are available from 1960M01 onwards.

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

Following Stock and Watson (1996), we focus on linear autoregressive models of lag length $p = 1$ and $p = 6$ and test whether the intercept is subject to a break. We estimate parameters on a moving windows of 120 observations to decrease the likelihood of multiple breaks occurring in the estimation sample. Test results are based on heteroskedasticity robust Wald statistics, which use the following estimate of the covariance matrix $\hat{\boldsymbol{V}}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\hat{\boldsymbol{\Omega}}_i\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$ with $[\hat{\boldsymbol{\Omega}}_i]_{kl} = \hat{\varepsilon}_k^2/(1-h_k)^2$ if $k = l$ and $[\hat{\boldsymbol{\Omega}}_i]_{kl} = 0$ otherwise, and h_k is the k-th diagonal element of $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. 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

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

		$T=120$					$T=240$					$T=480$				
ρ	λ	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
		Wald-test (17)														
0.0	0	0.05	0.05	0.06	0.05	0.03	0.06	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.17	0.20	0.22	0.21	0.17	0.21	0.22	0.23	0.21	0.16	0.24	0.24	0.23	0.21	0.16
	2	0.43	0.48	0.52	0.53	0.47	0.52	0.54	0.55	0.53	0.48	0.57	0.56	0.56	0.55	0.49
0.3	0	0.04	0.05	0.06	0.05	0.03	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.13	0.17	0.21	0.21	0.17	0.18	0.20	0.22	0.20	0.16	0.22	0.23	0.22	0.21	0.16
	2	0.33	0.40	0.47	0.50	0.46	0.46	0.50	0.53	0.52	0.47	0.54	0.54	0.55	0.55	0.48
0.6	0	0.03	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.05	0.03	0.05	0.06	0.06	0.05	0.03
	1	0.08	0.12	0.19	0.20	0.16	0.13	0.17	0.20	0.20	0.15	0.18	0.20	0.22	0.21	0.15
	2	0.19	0.26	0.39	0.46	0.43	0.33	0.40	0.47	0.50	0.45	0.47	0.49	0.52	0.53	0.47
0.9	0	0.02	0.05	0.10	0.09	0.06	0.02	0.04	0.08	0.07	0.04	0.03	0.05	0.06	0.06	0.04
	1	0.04	0.07	0.17	0.24	0.20	0.04	0.08	0.16	0.21	0.16	0.07	0.11	0.17	0.20	0.15
	2	0.09	0.12	0.24	0.44	0.44	0.09	0.14	0.28	0.43	0.39	0.16	0.24	0.37	0.46	0.41
		S-test (38)														
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.

depend on these choices.

To initialize the AR(p) model, we require p observations. Our first fore-

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

		$T = 120$					$T = 240$					$T = 480$				
ρ	h	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85	0.15	0.25	0.50	0.75	0.85
								Wald-test (17)								
0.0	0	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.04	0.03	0.07	0.07	0.06	0.05	0.03
	1	0.18	0.21	0.22	0.21	0.16	0.22	0.23	0.22	0.20	0.15	0.24	0.24	0.23	0.21	0.15
	2	0.45	0.49	0.52	0.52	0.46	0.53	0.55	0.55	0.53	0.47	0.57	0.57	0.56	0.54	0.48
0.3	0	0.05	0.06	0.06	0.05	0.03	0.06	0.06	0.06	0.04	0.03	0.06	0.07	0.06	0.05	0.03
	1	0.15	0.19	0.22	0.20	0.16	0.20	0.21	0.22	0.20	0.15	0.23	0.23	0.22	0.21	0.15
	2	0.36	0.42	0.48	0.51	0.45	0.48	0.51	0.53	0.52	0.46	0.55	0.55	0.55	0.54	0.47
0.6	0	0.04	0.06	0.07	0.05	0.04	0.05	0.06	0.06	0.04	0.03	0.06	0.06	0.06	0.05	0.03
	1	0.10	0.14	0.20	0.20	0.16	0.15	0.18	0.21	0.20	0.15	0.20	0.21	0.22	0.20	0.15
	2	0.22	0.30	0.42	0.47	0.43	0.36	0.42	0.49	0.50	0.44	0.48	0.51	0.53	0.52	0.46
0.9	0	0.03	0.07	0.12	0.10	0.07	0.04	0.05	0.09	0.07	0.05	0.04	0.06	0.07	0.06	0.04
	1	0.06	0.09	0.21	0.26	0.21	0.06	0.10	0.19	0.22	0.17	0.09	0.13	0.20	0.21	0.16
	2	0.11	0.15	0.30	0.48	0.45	0.12	0.18	0.34	0.46	0.41	0.20	0.28	0.41	0.47	0.42
								S-test (38)								
0.0	0	0.04	0.05	0.06	0.06	0.04	0.04	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.06	0.04
	1	0.13	0.16	0.21	0.24	0.22	0.15	0.18	0.21	0.23	0.20	0.17	0.19	0.21	0.23	0.20
	2	0.34	0.42	0.49	0.56	0.55	0.42	0.47	0.52	0.56	0.55	0.46	0.50	0.53	0.58	0.56
0.3	0	0.03	0.05	0.06	0.06	0.04	0.04	0.05	0.06	0.05	0.04	0.04	0.05	0.06	0.06	0.04
	1	0.10	0.14	0.20	0.23	0.22	0.13	0.17	0.20	0.23	0.20	0.16	0.18	0.21	0.23	0.20
	2	0.25	0.35	0.45	0.54	0.54	0.36	0.43	0.50	0.56	0.54	0.43	0.48	0.52	0.58	0.55
0.6	0	0.03	0.05	0.07	0.07	0.05	0.03	0.05	0.06	0.06	0.04	0.04	0.05	0.06	0.06	0.05
	1	0.06	0.10	0.19	0.24	0.22	0.09	0.14	0.19	0.23	0.20	0.13	0.16	0.20	0.23	0.20
	2	0.14	0.23	0.39	0.52	0.52	0.25	0.35	0.45	0.54	0.53	0.37	0.43	0.49	0.56	0.55
0.9	0	0.02	0.05	0.12	0.12	0.09	0.02	0.04	0.09	0.09	0.06	0.03	0.04	0.07	0.08	0.06
	1	0.03	0.06	0.19	0.31	0.28	0.03	0.07	0.17	0.25	0.23	0.05	0.10	0.18	0.25	0.21
	2	0.06	0.11	0.27	0.53	0.56	0.07	0.12	0.30	0.50	0.51	0.12	0.21	0.37	0.52	0.52

Note: The table presents finite sample size and power properties of the tests comparing the shrinkage forecast (42) and the full-sample, equal weights forecast, using a nominal size of 0.05. For further details, see the footnote of Table 3.

cast is therefore for July 1970 and we recursively construct one-step ahead forecasts until the end of the sample.

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

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

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

5.1 Structural break test results

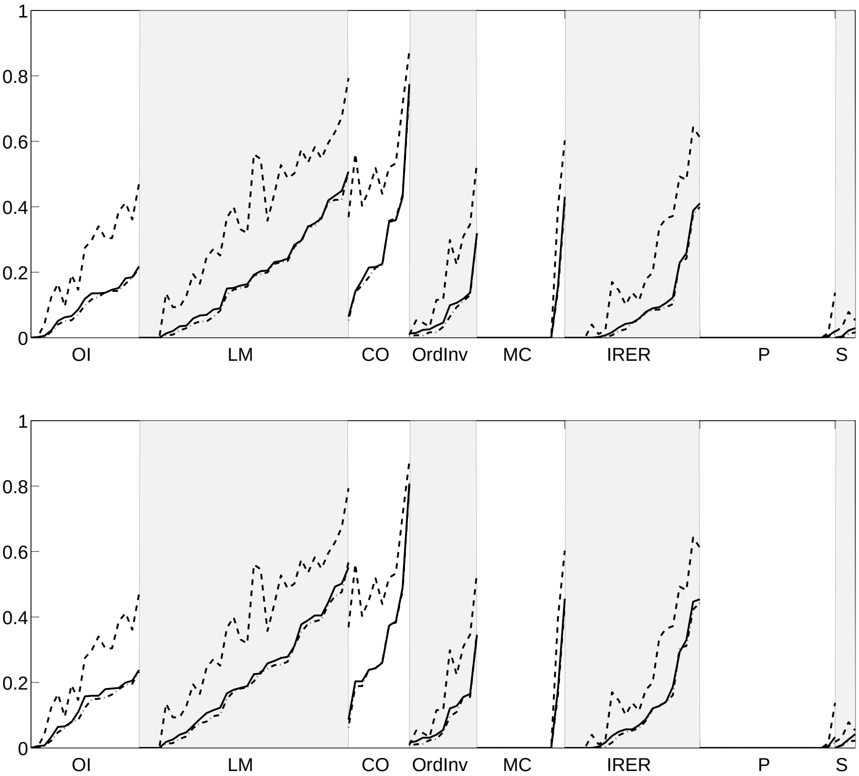
In this forecast exercise, we will refer to the test of Andrews (1993) as *sup W*, the Wald test statistic (17) as W, the test statistic (38) as S, and, when the alternative is the shrinkage forecast, these tests as W^s and S^s. In Table 5, we report the fraction of estimation samples where *sup W* would indicate a break at a nominal size of $\alpha = 0.05$. This is contrasted with the fraction where our tests indicate a break also at a nominal size of $\alpha = 0.05$. It is clear that a large fraction of the breaks picked up by *sup W* are judged as irrelevant for forecasting by W and S. The fraction of forecasts for which a break is indicated is lower by a factor of over 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 the *sup W* test is more often significant than the W and S test. Yet, we see substantial differences between categories. Whereas in the *labor market* and *consumption and orders* categories some of the series contain a significant breaks in up to 70% of the estimation samples when the W or S tests are used, the *prices* and *stock market* series hardly show any significant breaks from a forecasting perspective. This finding concurs with the general perception that, for these type of time series, simple linear models are very hard to beat in terms of MSFE.

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

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

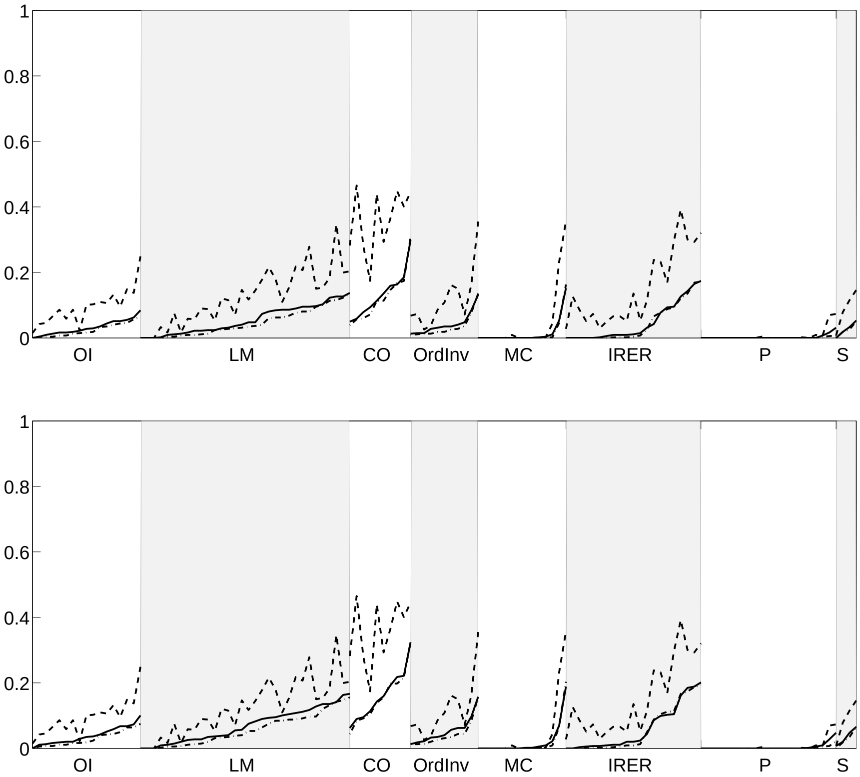


Note: The upper panel depicts the fraction of estimation samples with a significant break when testing under the alternative of the post-break forecast; the lower panel when testing under the alternative of the shrinkage forecast (42). Dashed lines indicate the fraction of estimation samples with significant sup-Wald test, dashed-dotted lines indicate the fraction of estimation samples where the break test W in (17) indicates a break, and solid lines indicate the fraction of estimation samples with significant S test in (38).

estimation sample is given on the horizontal axis. In the top panel are the results for the test comparing the post-break estimation window with the full estimation window. In the bottom panel are the tests comparing the shrinkage estimator and the full sample, equal weights estimator. It is clear that the $sup W$ test finds more breaks in for the vast majority of estimation samples, whereas the results from the W and S tests are extremely similar.

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

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

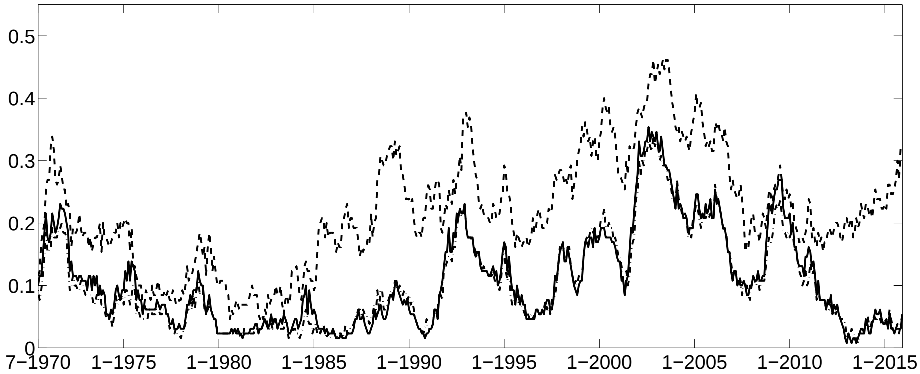


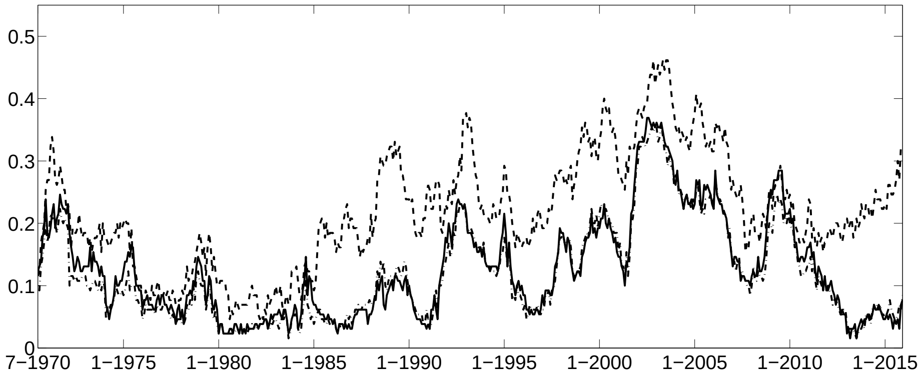
Note: See footnote of Table 6

forecast, even more breaks that are relevant for forecasting than the *supW* test. From 2010 onwards, breaks that are relevant for forecasting decrease sharply, whereas the *supW* tests continues to find a large number of breaks.

Figure 9 shows the results but for the AR(6) model. All tests find fewer estimation samples with breaks compared to the AR(1) model. The evolution over the estimation samples is, however, similar to the AR(1) case. In the initial estimation samples up to 1985 all tests agree that a small number of series are subject to a structural break. From 1985 to 1990, however, the *supW* test finds breaks in up to a third of the estimation samples, which the W and S tests do not find important for forecasting. The same is true for breaks around 2000. In contrast, in the period following the dot com bubble and following the financial crisis of 2008/9 the W and the S tests find as many and, in the case of the shrinkage forecasts, more series, where taking a break into account will improve forecast accuracy than the *supW* test. Again, the number of series that should take breaks into account declines sharply towards the end of our sample when using the W and S tests but not when using the *supW* tests.

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





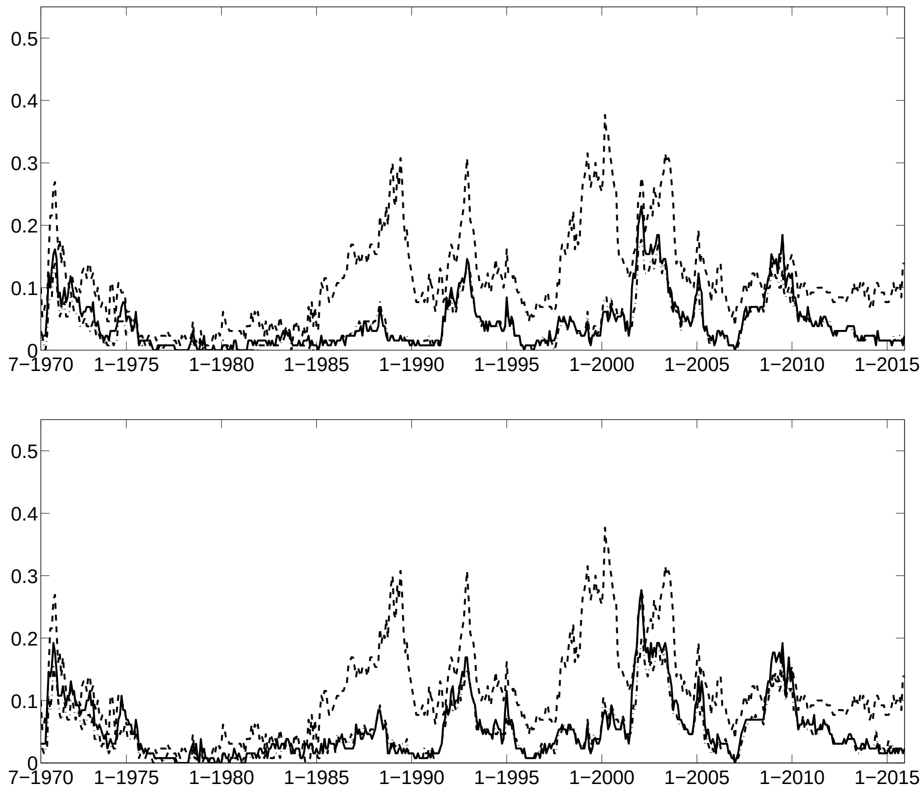
Note: The plots show the fractions of series with a significant break for each estimation sample when using an AR(1) model with a break in intercept. The top panel shows results when testing between the post-break sample based forecast and the full sample based forecast and the lower panel when testing between the shrinkage forecast and the full sample, equal weights forecast. The dashed line indicates the fraction of series when testing using the standard sup-Wald test at $\alpha = 0.05$, the solid line when testing using the S-test in (38), and the dashed-dotted line when testing using the W-test in (17). The dates displayed on the horizontal axis are the end dates of the estimation samples.

5.2 Forecast accuracy

In the next step, we are going to investigate whether forecasts conditional on the W and S tests are more accurate than forecasts based on the $supW$ test. We use each test to determine whether to use the post-break or the full sample for forecasting or, alternatively, whether to use the shrinkage or the equal weights forecast. All tests are carried out at $\alpha = 0.05$.

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

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



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

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

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

When the shrinkage forecast is used in conjunction with the W^s or S^s

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

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

Note: The table reports the average of the ratio of the respective forecasts' MSFE over that of the forecasts resulting from the sup-Wald test of Andrews (1993) at $\alpha = 0.05$. Forecasts for which none of the tests indicate a break are excluded. Results are reported for the test statistic W in (17) and S in (38). 'Post-break' and 'Shrinkage' indicate that under the alternative the post-break forecast, respectively the shrinkage forecast (42), 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, the accuracy of the forecasts is very similar as those of the post-break forecasts. This can be expected since we reject the test when the Wald statistic, that governs the amount of shrinkage, is relatively large. This implies that upon rejection of the test statistic, a forecast is used that is relatively close to the post-break forecast. The last column shows that using the shrinkage forecast in conjunction with the $supW$ test leads to forecasts that, while more precise than post-break forecasts based on the same test, are clearly dominated by the W^s and S^s tests. In fact, for all categories and both models the W^s test leads to more accurate forecasts and the S^s tests for all categories and both models, with the exception of the AR(1) and prices.

6 Conclusion

In this paper, we formalize the notion that small breaks might be better left ignored when forecasting. We quantify the break size that leads to equal forecast performance between a model based on the full sample and one based on a post-break sample. This break size is substantial, which points to a large penalty that is incurred by the uncertainty around the break date. A second finding is that the break size 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 weak optimality, in the sense that the power of an infeasible test conditional on the break date is achieved when we consider a 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 sizes we consider under our null hypothesis, this optimality is achieved relatively quickly, i.e. for finite nominal size. Simulations confirm this argument and show only a minor loss of power compared to the test is conditional on the true break date.

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

References

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6):1383–1414.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

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.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):134–144.

Elliott, G. and Müller, U. K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics*, 141(2):1196–1218.

Elliott, G. and Müller, U. K. (2014). Pre and post break parameter inference. *Journal of Econometrics*, 180(2):141–157.

Elliott, G., Müller, U. K., and Watson, M. W. (2015). Nearly optimal tests when a nuisance parameter is present under the null hypothesis. *Econometrica*, 83(2):771–811.

Giacomini, R. and Rossi, B. (2009). Detecting and predicting forecast breakdowns. *Review of Economic Studies*, 76(2):669–705.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Hüsler, J. (1990). Extreme values and high boundary crossings of locally stationary Gaussian processes. *Annals of Probability*, 18(3):1141–1158.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *American Statistician*, 54(3):217–224.

MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.

Magnus, J. R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *Econometrics Journal*, 5(1):225–236.

McCracken, M. W. and Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, (forthcoming).

Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.

Pesaran, M. H., Pick, A., and Pranovich, M. (2013). Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 177(2):134–152.

Pesaran, M. H. and Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129(1):183–217.

Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, volume 148. American Mathematical Soc.

Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40(2):307–318.

Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate U.S. stock returns. *Journal of Financial Econometrics*, 4(2):238–274.

Rossi, B. (2006). Are exchange rates really random walks? Some evidence robust to parameter instability. *Macroeconomic Dynamics*, 10(1):20–38.

Stock, J. H. and Watson, M. W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30.

Thompson, J. R. (1968). Some shrinkage techniques for estimating the mean. *Journal of the American Statistical Association*, 63(321):113–122.

Toro-Vizcarrondo, C. and Wallace, T. D. (1968). A test of the mean square error criterion for restrictions in linear regression. *Journal of the American Statistical Association*, 63(322):558–572.

Trenkler, G. and Toutenburg, H. (1992). Pre-test procedures and forecasting in the regression model under restrictions. *Journal of Statistical Planning and Inference*, 30(2):249–256.

Wallace, T. D. (1972). Weaker criteria and tests for linear restrictions in regression. *Econometrica*, 40(4):689–698.

Appendix A Additional mathematical details

A.1 Derivation of (19)

Define $\Delta = \Delta_1 - \Delta_2$ where

$$\begin{aligned}
\Delta_1 &= T\mathrm{E}\left[\left(\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2) + \partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2\right] \\
&= T\mathrm{E}\left[\left(\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2)\right)^2 + \left(\partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2 + \right. \\
&\qquad \left. + 2\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2)\partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right]
\end{aligned} \tag{46}$$

and similarly for Δ_2

$$\begin{aligned}
\Delta_2 &= T\mathrm{E}\left[\left(\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_F(\hat{\tau}) - \boldsymbol{\beta}_2) + \partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2\right] \\
&= T\mathrm{E}\left[\left(\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2)\right)^2 + \left(\partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2 + \right. \\
&\qquad \left. + 2\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2)\partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right]
\end{aligned} \tag{47}$$

In addition, we define

$$\begin{aligned}
a &= \frac{1}{1-\hat{\tau}}\left[\partial_{\beta_2} f'(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'(\boldsymbol{B}(1) - \boldsymbol{B}(\hat{\tau})) + \int_{\hat{\tau}}^1 \partial_{\beta_2} f'\boldsymbol{\eta}(s)ds\right] \\
b &= \partial_{\beta_2} f'(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
c &= \partial_\delta f'(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\boldsymbol{B}(1) \\
d &= \partial_{\beta_2} f'(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{B}(1) + \int_0^1 \partial_{\beta_2} f'\boldsymbol{\eta}(s)ds
\end{aligned} \tag{48}$$

We have from (13) and (15) that $\sqrt{T}\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) \to a - b$, $\sqrt{T}\partial_{\beta_2} f'(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2) \to d - b$ and $\sqrt{T}\partial_\delta f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \to c$ Then

$$\begin{aligned}
\Delta_1 &\to \mathrm{E}[a^2 + b^2 - 2ab + c^2 + 2ca - 2cb] \\
\Delta_2 &\to \mathrm{E}[d^2 + b^2 - 2db + c^2 + 2cd - 2cb]
\end{aligned} \tag{49}$$

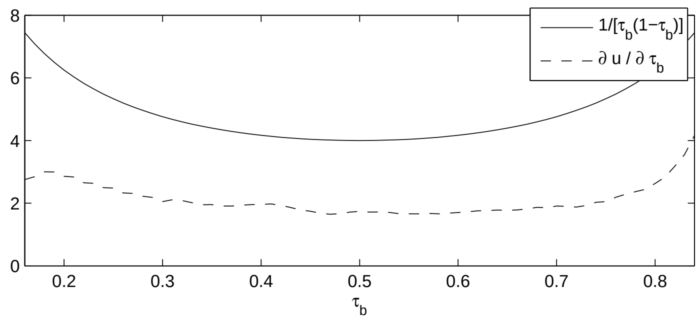
Now $\mathrm{E}[db] = \mathrm{E}[cd] = 0$ by the fact that $\mathrm{E}[\boldsymbol{B}(1)] = 0$, $\mathrm{E}[\boldsymbol{B}(1)\boldsymbol{B}(1)'] = \boldsymbol{I}$ and $\boldsymbol{M}_{\bar{\boldsymbol{X}}}\bar{\boldsymbol{X}} = \boldsymbol{O}$. Furthermore, as the distribution of the test statistic is independent of the estimation of $\boldsymbol{\delta}$ (Andrews, 1993), we can regard $\hat{\tau}$ as independent of $\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}$. This yields $\mathrm{E}[(a-d)c] = 0$. Concluding, we have

$$\Delta_1 - \Delta_2 = \mathrm{E}[a^2 - d^2] \tag{50}$$

A.2 Verifying condition (36)

In order to very that (36) holds, that is, that the condition for weak optimality, $\partial u(\tau_b)/\partial \tau_b < 1/[\tau_b(1-\tau_b)]$, holds. Observe that, in Figure 10, the dashed line, which depicts the derivative of the critical values for $\alpha = 0.05$ as a function of the break date τ_b and is obtained via simulation, is clearly below the solid line, which depicts the upper bound $[\tau_b(1-\tau_b)]^{-1}$.

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



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

A.3 Uniqueness of the break size that yields equal forecast accuracy

In order to ensure the uniqueness of the break size that leads to equal forecast accuracy, we evaluate Δ in (19) and Δ_s in (43) numerically using the simulation set-up described in Section 4. The results in Figure 11 show that the break size that leads to equal forecast accuracy is, in fact, unique.

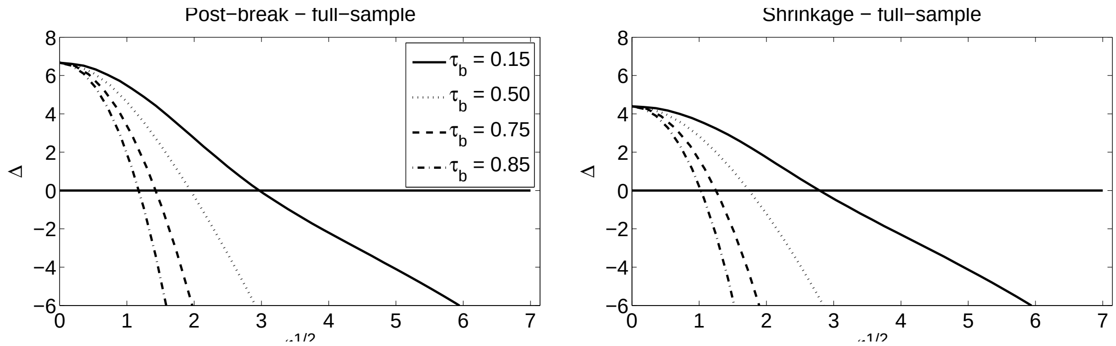
A.4 Derivation of equation (39)

We start by noting that (39)

$$\begin{aligned} \mathrm{E}\left[T\left(\hat{y}^S_{T+1} - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2\right)^2\right] &= \mathrm{E}\left[T\left(\omega\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_1 + (1-\omega)\boldsymbol{x}'_{T+1}\hat{\boldsymbol{\beta}}_2 - \boldsymbol{x}'_{T+1}\boldsymbol{\beta}_2\right)^2\right] \\ &= \omega^2\mathrm{E}\left[T\left(\boldsymbol{x}'_{T+1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right)^2\right] + \frac{1}{\tau_b}\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1} \\ &\quad + 2\omega\boldsymbol{x}'_{T+1}\mathrm{E}\left[T\left(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2\right)\left(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2\right)\right]\boldsymbol{x}_{T+1} \end{aligned} \tag{51}$$

We analyze the first and third term of the second equality separately. Using a bias-variance decomposition, the expectation in the first term can be

Figure 11: Difference in MSFE between the post-break forecast and full-sample forecast



Note: The left panel shows the difference in the asymptotic MSFE between the post-break forecast and the full-sample forecast as a function of the standardized break size $\zeta^{1/2}$ in (19) for $\tau_b = \{0.15, 0.50, 0.75, 0.85\}$. The right panel shows the difference in MSFE between the shrinkage forecast and the full-sample forecast in (43).

calculated as

$$\begin{aligned}\mathrm{E}\left[T\left(\boldsymbol{x}_{T+1}'(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2)\right)^2\right] &= \mathrm{E}\left[T\left(\boldsymbol{x}_{T+1}'(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2)\right)\right]^2 + T\mathrm{Var}\left[\boldsymbol{x}_{T+1}'(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2)\right] \\ &= T\left(\boldsymbol{x}_{T+1}'\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\right)\right)^2 + \boldsymbol{x}_{T+1}'\left(\frac{1}{\tau_b}+\frac{1}{1-\tau_b}\right)\boldsymbol{V}\boldsymbol{x}_{T+1}\end{aligned} \tag{52}$$

using that $\mathrm{Cov}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = 0$. The term linear in ω is given by

$$\begin{aligned}\boldsymbol{x}_{T+1}'\mathrm{E}\left[T\left(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2\right)\left(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2\right)\right]\boldsymbol{x}_{T+1} &= -\boldsymbol{x}_{T+1}'\mathrm{E}\left[T\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\right)\boldsymbol{\beta}_2'\right]\boldsymbol{x}_{T+1} \\ &\quad + \boldsymbol{x}_{T+1}'\mathrm{E}\left[T\hat{\boldsymbol{\beta}}_1\hat{\boldsymbol{\beta}}_2' - \hat{\boldsymbol{\beta}}_2\hat{\boldsymbol{\beta}}_2'\right]\boldsymbol{x}_{T+1} \\ &= -\frac{1}{1-\tau_b}\boldsymbol{x}_{T+1}'\boldsymbol{V}\boldsymbol{x}_{T+1}\end{aligned} \tag{53}$$

Appendix B   Tables with critical values

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

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

τ_b	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
$\zeta^{1/2}$	2.99	2.73	2.55	2.41	2.28	2.17	2.06	1.95	1.84	1.75	1.64	1.54	1.43	1.31	1.18
Wald test statistic (22)															
0.10	20.44	19.16	17.99	17.05	16.22	15.49	14.79	14.13	13.49	12.91	12.30	11.68	11.04	10.32	9.36
0.05	23.71	22.29	20.99	19.95	19.04	18.24	17.46	16.74	16.03	15.38	14.71	14.02	13.30	12.48	11.37
0.01	30.54	28.84	27.29	26.07	25.00	24.06	23.15	22.29	21.46	20.70	19.89	19.08	18.22	17.23	15.82
0.10	0.13	0.13	0.12	0.12	0.12	0.12	0.11	0.11	0.11	0.10	0.10	0.09	0.09	0.08	0.06
0.05	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S test statistic (38)															
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
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
0.01	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
0.10	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
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
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

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

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

τ_b	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
$\zeta^{1/2}$	2.81	2.54	2.36	2.21	2.08	1.97	1.85	1.75	1.64	1.55	1.45	1.35	1.25	1.15	1.03
Wald test statistic (22)															
0.10	19.01	17.78	16.63	15.71	14.92	14.22	13.56	12.95	12.34	11.81	11.28	10.74	10.19	9.58	8.82
0.05	22.15	20.78	19.51	18.48	17.60	16.84	16.10	15.43	14.76	14.16	13.57	12.97	12.34	11.64	10.74
0.01	28.74	27.08	25.57	24.35	23.30	22.40	21.53	20.74	19.95	19.23	18.53	17.81	17.03	16.18	15.02
0.10	0.14	0.13	0.13	0.12	0.12	0.12	0.11	0.11	0.11	0.10	0.10	0.09	0.08	0.07	0.06
0.05	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.03
0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
S test statistic (38)															
0.10	1.85	1.88	1.90	1.92	1.93	1.93	1.93	1.93	1.91	1.90	1.88	1.86	1.82	1.76	1.63
0.05	2.18	2.22	2.24	2.25	2.26	2.27	2.27	2.27	2.26	2.25	2.23	2.20	2.17	2.11	1.98
0.01	2.82	2.85	2.87	2.89	2.90	2.90	2.91	2.91	2.90	2.89	2.87	2.85	2.82	2.76	2.63
0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.10	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Note: The table reports critical values and size for the W and S test statistics that test the null hypothesis of equal MSFE of the shrinkage forecast (42) and the full sample forecast. For additional information, see the footnote of Table 2.

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

τ_b	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
$\zeta^{1/2}$	4.17	3.69	3.44	3.25	3.10	2.97	2.85	2.73	2.63	2.53	2.43	2.32	2.21	2.10	1.97	1.84	1.68	1.48	1.19
Wald test statistic (22)																			
0.10	31.31	27.59	25.36	23.74	22.49	21.43	20.49	19.59	18.85	18.11	17.38	16.65	15.90	15.16	14.39	13.56	12.62	11.52	9.82
0.05	35.41	31.37	28.96	27.22	25.85	24.69	23.69	22.70	21.88	21.07	20.27	19.47	18.64	17.83	16.97	16.04	15.00	13.74	11.75
0.01	43.80	39.15	36.41	34.41	32.87	31.54	30.37	29.23	28.29	27.33	26.38	25.44	24.47	23.52	22.51	21.40	20.15	18.60	16.04
0.10	0.11	0.11	0.11	0.11	0.12	0.11	0.12	0.11	0.12	0.12	0.12	0.11	0.11	0.11	0.11	0.10	0.10	0.08	0.05
0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.04	0.03
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00
S test statistic (38)																			
0.10	1.55	1.64	1.68	1.72	1.74	1.77	1.79	1.80	1.82	1.83	1.85	1.86	1.86	1.87	1.87	1.86	1.84	1.79	1.56
0.05	1.90	1.98	2.03	2.06	2.08	2.10	2.13	2.14	2.16	2.17	2.18	2.19	2.20	2.21	2.21	2.20	2.19	2.14	1.91
0.01	2.55	2.63	2.66	2.70	2.72	2.74	2.76	2.77	2.79	2.80	2.82	2.83	2.84	2.84	2.85	2.85	2.83	2.79	2.57
0.10	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

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

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

τ_b	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
$\zeta^{1/2}$	4.04	3.56	3.30	3.10	2.95	2.81	2.68	2.56	2.46	2.35	2.24	2.13	2.02	1.90	1.78	1.65	1.49	1.31	1.05
									Wald test statistic (22)										
0.10	30.00	26.40	24.15	22.47	21.24	20.17	19.17	18.33	17.58	16.84	16.14	15.43	14.70	14.01	13.29	12.54	11.70	10.76	9.40
0.05	34.01	30.09	27.65	25.85	24.49	23.32	22.25	21.31	20.49	19.67	18.90	18.11	17.32	16.55	15.74	14.90	13.95	12.86	11.24
0.01	42.23	37.71	34.93	32.84	31.31	29.96	28.71	27.61	26.66	25.70	24.77	23.84	22.89	21.99	21.04	20.01	18.85	17.48	15.35
0.10	0.11	0.11	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.11	0.11	0.10	0.10	0.09	0.08	0.05
0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.04	0.02
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00
									S test statistic (38)										
0.10	1.57	1.67	1.73	1.76	1.79	1.82	1.84	1.85	1.87	1.88	1.90	1.90	1.91	1.91	1.91	1.89	1.87	1.81	1.59
0.05	1.92	2.02	2.07	2.10	2.13	2.16	2.17	2.19	2.21	2.22	2.23	2.24	2.24	2.25	2.24	2.24	2.21	2.16	1.95
0.01	2.57	2.66	2.70	2.74	2.77	2.79	2.81	2.82	2.84	2.85	2.86	2.87	2.88	2.88	2.88	2.88	2.85	2.81	2.60
0.10	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.08
0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.04
0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

Note: The table reports critical values and size for the W and S test statistics that test the null hypothesis of equal MSFE of the shrinkage forecast (42) and the full sample forecast. For additional information, see the footnote of Table 2.