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

Tom Boot^{*} Andreas Pick[†]

February 15, 2016

Abstract

We propose a near optimal test for local 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: forecasting, squared error loss, structural break test

1 Introduction

Structural breaks present a major challenge to forecasters as they require information about the time of the break and accurate parameter estimates for the post-break sample. Mean square forecast error loss implies a biasvariance trade-off, which suggests that ignoring breaks of a smaller magnitude will lead to more accurate forecasts. This is reflected in the literature, where forecasts based on the full sample are frequently found to be more accurate, see for example Pesaran and Timmermann (2005), even if structural instabilities have been documented in many economic series, for example by Stock and Watson (1996). Hence, it is clear that sufficiently small breaks can be ignored, which yields the question: what constitutes sufficiently small?

^{*}Erasmus University Rotterdam, boot@ese.eur.nl

 $^{^{\}dagger}\mathrm{Erasmus}\ \mathrm{University}\ \mathrm{Rotterdam}\ \mathrm{and}\ \mathrm{De}\ \mathrm{Nederlandsche}\ \mathrm{Bank},\ \mathsf{andreas.pick} @\mathsf{cantab.net}$

We develop a test for equal forecast accuracy that compares the expected mean square forecast error (MSFE) of a one-step-ahead forecast that accounts for a break and that of a forecast that uses the full sample. Under a known break date, the largest break size for which a linear model achieves equal predictive accuracy can be calculated to be one standard deviation of the distribution of the parameter estimates. Under local breaks of unknown timing, the uncertainty surrounding the break dates increases the variance of the forecast such that the critical break size is much larger, up to three standard deviations in terms of the distribution of the parameter estimates. Building subsequently on the work by Andrews (1993) and Piterbarg (1996), we derive a weakly optimal test for the critical break size and argue that this test is near optimal for conventional choices of the nominal size due to the size of the breaks that are allowed under the null. This is confirmed in simulations of asymptotic power. In the process, we show that the weak 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), Elliott and Müller (2014), and Elliott et al. (2015). While these 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. Additionally, in line with the forecasting literature, our loss function is the mean squared forecast error, and squared error loss has been shown to yield different conclusions than standard F-tests by Toro-Vizcarrondo and Wallace (1968) and Wallace (1972).

Our test is also 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 evaluates *ex ante* the accuracy of forecasts of models that do or do not account for breaks by comparing the respective expected MSFE.

Closer to our work is the paper by Giacomini and Rossi (2009), which assesses 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 break point test statistic introduced in this paper.

Under a known break, 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, a shrinkage estimator averages between two biased estimators and its forecasting performance compared to the full sample forecast is not immediately clear. Since the estimator does not take break date uncertainty into account, it is expected to put too much weight on the forecast that incorporates the break. 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 derive a second, near-optimal version of our test that compares the forecast.

We apply our test to the macroeconomic time series in the FRED-MD data set by McCracken and Ng (2014), which consists of 135 series divided into eight categories. We find that breaks that are important for forecasting under MSFE loss are between a factor four to five 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 3 using the methodology of Andrews (1993). In Section 4 we derive the test and show its weak optimality in Section 5. Section 6 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. Section 7 considers the optimal weights forecast developed by Pesaran et al. (2013) and studies its asymptotic properties. Finally, an application to the large set of macroeconomic time series in the FRED-MD data set is considered in Section 8.

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 ${\cal T}_b$

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

with

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

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

Denote $V_i = (T_i - T_{i-1}) \operatorname{Var}(\hat{\boldsymbol{\beta}}_i)$ for $i = 1, 2, T_0 = 0, T_1 = T_b, T_2 = T$ and $V_F = T \operatorname{Var}(\hat{\boldsymbol{\beta}}_F)$ as the estimated variances of the vectors of coefficient estimates. The expected mean squared forecast error (MSFE) for the forecast that incorporates the break is

$$R(\boldsymbol{x}_{T+1}'\hat{\boldsymbol{\beta}}_2) = \mathbf{E}\left[\left(\boldsymbol{x}_{T+1}'\hat{\boldsymbol{\beta}}_2 - \boldsymbol{x}_{T+1}'\boldsymbol{\beta}_2 - \varepsilon_{T+1}\right)^2\right]$$

$$= \frac{1}{T - T_b}\boldsymbol{x}_{T+1}\boldsymbol{V}_2\boldsymbol{x}_{T+1} + \sigma^2$$
(3)

The MSFE for forecast using the full sample estimate is

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

Comparing (3) with (4), we see that the full sample forecast is more accurate than the post-break sample forecast if

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

Table 1: Critical values for the $\chi^2(1,1)$ and $\chi^2(1)$ distribution

	0.90	0.95	0.99
$\chi^2(1,1) \ \chi^2(1)$	$5.22 \\ 2.71$	$7.00 \\ 3.84$	$\begin{array}{c} 11.07\\ 6.63\end{array}$

with $\tau_b = T_b/T$ and the third line assumes that the covariance matrices asymptotically satisfy $\operatorname{plim}_{T\to\infty} \mathbf{V}_i \to \mathbf{V}$ for i = 1, 2, F. From (5) it can be observed that, under $H_0: \zeta = 1$, the size of the break $\mathbf{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)$ is symmetric in τ_b . It is also clear from (3) that breaks that occur at the end of the sample will lead tot a larger mean squared forecast error than breaks that occur at the beginning.

A test that takes $H_0: \zeta = 1$ can be carried out by noting that

$$\hat{\zeta}(\tau) = T\tau^{2} \frac{\left[\boldsymbol{x}_{T+1}^{\prime} \hat{\boldsymbol{V}}_{F} \hat{\boldsymbol{V}}_{1}^{-1} (\hat{\boldsymbol{\beta}}_{1} - \hat{\boldsymbol{\beta}}_{2}) \right]^{2}}{\boldsymbol{x}_{T+1}^{\prime} \left(\frac{\hat{\boldsymbol{V}}_{2}}{1 - \tau} - \hat{\boldsymbol{V}}_{F} \right) \boldsymbol{x}_{T+1}}$$

$$= \frac{\left[\boldsymbol{x}_{T+1}^{\prime} (\hat{\boldsymbol{\beta}}_{F} - \hat{\boldsymbol{\beta}}_{2}) \right]^{2}}{\boldsymbol{x}_{T+1}^{\prime} \widehat{\operatorname{Var}} (\hat{\boldsymbol{\beta}}_{F} - \hat{\boldsymbol{\beta}}_{2}) \boldsymbol{x}_{T+1}} \stackrel{a}{\sim} \chi^{2}(1, \zeta)$$
(6)

A more conventional, yet asymptotically equivalent, form of the test statistic is

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

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

It is clear that the results of the test will in general differ for two reasons from the outcomes of the classical Wald test on the difference between the parameter vectors β_1 and β_2 . The first is that the multiplication by \mathbf{x}_{T+1} can cause large breaks to become irrelevant for forecasting or small breaks to become relevant. Especially the first scenario is expected due to the fact that breaks in the coefficients of β can cancel in the inner product $\mathbf{x}'_{T+1}\beta$. The second is that under $H_0: \zeta = 1$, we compare the test statistic against the critical values of the non-central $\chi^2(1,1)$ distribution, instead of the central $\chi^2(1)$ distribution. Table 1 provides a comparison of the critical values of both distributions, which shows that the difference is substantial.

In the next section, we turn to a more general set-up that assumes the break date to be unknown and considers the asymptotic limit where $T \rightarrow \infty$. It is clear from (5) that if the difference in the parameters, $\beta_1 - \beta_2$, converges to zero at a rate $T^{-1/2+\epsilon}$ for some $\epsilon > 0$, the test statistic diverges to infinity. We therefore consider breaks which are local in nature, i.e.

 $\beta_2 = \beta_1 + \frac{1}{\sqrt{T}} \eta$, rendering a finite test statistic in the asymptotic limit. Local breaks have been recently intensively studied, see for example Elliott and Müller (2007), Elliott and Müller (2014) and Elliott et al. (2015). An implication of local breaks is that no consistent estimator for the break date is available. This mimics practical situations, where the break date is rarely known with certainty. As a consequence, the forecast that results from a model that incorporates the break will not be unbiased. One can expect that this favors the full sample forecast, such that larger breaks are allowed under the null of equal forecasting performance between a full sample and post-break sample estimations.

3 General set-up

The general estimation considered in this paper 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 extra 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

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

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

In the model (8), the parameter vector $\boldsymbol{\delta}$ is known to be constant but the parameter vector $\boldsymbol{\beta}$ could be subject to a subject to a structural break. The parameters can be estimated by minimizing the sample analogue of the population moment conditions

$$\frac{1}{T}\sum_{t=1}^{T} \mathbf{E}[m(\boldsymbol{W}_t, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0)] = 0$$
(9)

which requires solving

$$\frac{1}{T}\sum_{\boldsymbol{\beta},\boldsymbol{\delta}} m(\boldsymbol{W}_{t},\hat{\boldsymbol{\beta}}_{F},\hat{\boldsymbol{\delta}})'\hat{\boldsymbol{\gamma}}\frac{1}{T}\sum_{\boldsymbol{m}} m(\boldsymbol{W}_{t},\hat{\boldsymbol{\beta}}_{F},\hat{\boldsymbol{\delta}}) = \\ \inf_{\boldsymbol{\beta},\boldsymbol{\delta}} \frac{1}{T}\sum_{\boldsymbol{m}} m(\boldsymbol{W}_{t},\boldsymbol{\beta},\boldsymbol{\delta})'\hat{\boldsymbol{\gamma}}\frac{1}{T}\sum_{\boldsymbol{m}} m(\boldsymbol{W}_{t},\boldsymbol{\beta},\boldsymbol{\delta})$$
(10)

where $\hat{\beta}_F$ is estimator that uses the entire sample. We assume throughout the weighting matrix $\gamma = S^{-1}$ and

$$\boldsymbol{S} = \lim_{T \to \infty} \operatorname{Var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} m(\boldsymbol{W}_t, \boldsymbol{\beta}, \boldsymbol{\delta}) \right)$$
(11)

for which a consistent estimator is available.

The *p* parameters given by β might be subject to a structural break, while the *q* parameters in δ are constant over the observed sample. As discussed in the previous paragraph, we consider a null hypothesis that allows local breaks, defined by

$$\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 + \frac{1}{\sqrt{T}} \boldsymbol{\eta}(\tau) \tag{12}$$

where $\eta(\tau) = bI[\tau < \tau_b]$ and **b** is a vector of constants. The pre-break parameters β_1 and the post-break parameters β_2 satisfy partial sample moment conditions given by

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

Define

$$\bar{m}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\delta}, \tau) = \frac{1}{T} \sum_{t=1}^{T\tau} \begin{pmatrix} m(\boldsymbol{W}_t, \boldsymbol{\beta}_1, \boldsymbol{\delta}) \\ \mathbf{0} \end{pmatrix} + \frac{1}{T} \sum_{t=T\tau+1}^{T} \begin{pmatrix} \mathbf{0} \\ m(\boldsymbol{W}_t, \boldsymbol{\beta}_2, \boldsymbol{\delta}) \end{pmatrix}$$
(13)

then, partial sum GMM estimators can be obtained by solving (10) with $m(\cdot)$ replaced by $\bar{m}(\cdot)$ and $\hat{\gamma}$ replaced by

$$\hat{\gamma}(\tau) = \begin{pmatrix} \frac{1}{\tau} \hat{\boldsymbol{S}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{1-\tau} \hat{\boldsymbol{S}}^{-1} \end{pmatrix}$$
(14)

We are interested in comparing the mean squared error of forecasts based on full sample moment conditions for estimation with those using partial sample moment conditions. The forecasts are constructed as

$$\hat{y}_{T+1}^F = f(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \boldsymbol{X}_t, \boldsymbol{Y}_{t-1})$$

$$\hat{y}_{T+1}^P = f(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \boldsymbol{X}_t, \boldsymbol{Y}_{t-1})$$
(15)

Throughout, we condition on the both the exogenous and endogenous variables that are needed to construct the forecast. The comparison between \hat{y}_{T+1}^F and \hat{y}_{T+1}^P 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 (15), 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 as below, are given by Andrews (1993). The asymptotic distributions depend

on the following matrices, for which consistent estimators are assumed to be available,

$$\boldsymbol{M} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \left[\frac{\partial m(\boldsymbol{W}_t, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0)}{\partial \boldsymbol{\beta}} \right]$$
$$\boldsymbol{M}_{\boldsymbol{\delta}} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \left[\frac{\partial m(\boldsymbol{W}_t, \boldsymbol{\beta}_0, \boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} \right]$$
(16)

To simplify the notation, define

$$\bar{\boldsymbol{X}}' = \boldsymbol{M}' \boldsymbol{S}^{-1/2}$$

$$\bar{\boldsymbol{Z}}' = \boldsymbol{M}_{\delta}' \boldsymbol{S}^{-1/2}$$
(17)

Break model The partial-samples estimators converge to the following Gaussian process indexed by τ

$$\begin{aligned}
\sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\beta}}_{1}(\tau) - \boldsymbol{\beta}_{2} \\ \hat{\boldsymbol{\beta}}_{2}(\tau) - \boldsymbol{\beta}_{2} \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} &\to \begin{bmatrix} \tau \bar{\boldsymbol{X}}' \bar{\boldsymbol{X}} & \boldsymbol{0} & \tau \bar{\boldsymbol{X}}' \bar{\boldsymbol{Z}} \\ \boldsymbol{0} & (1 - \tau) \bar{\boldsymbol{X}}' \bar{\boldsymbol{X}} & (1 - \tau) \bar{\boldsymbol{X}}' \bar{\boldsymbol{Z}} \\ \tau \bar{\boldsymbol{Z}}' \bar{\boldsymbol{X}} & (1 - \tau) \bar{\boldsymbol{Z}}' \bar{\boldsymbol{X}} & \bar{\boldsymbol{Z}}' \bar{\boldsymbol{Z}} \end{bmatrix}^{-1} \\
&\times \begin{bmatrix} \bar{\boldsymbol{X}}' \boldsymbol{B}(\tau) + \bar{\boldsymbol{X}}' \bar{\boldsymbol{X}} \int_{0}^{\tau} \boldsymbol{\eta}(s) ds \\ \bar{\boldsymbol{X}}'(\boldsymbol{B}(1) - \boldsymbol{B}(\tau)) + \bar{\boldsymbol{X}}' \bar{\boldsymbol{X}} \int_{0}^{\tau} \boldsymbol{\eta}(s) ds \\ \bar{\boldsymbol{Z}}' \boldsymbol{B}(1) + \bar{\boldsymbol{Z}}' \bar{\boldsymbol{X}} \int_{0}^{1} \boldsymbol{\eta}(s) ds \end{bmatrix}
\end{aligned} \tag{18}$$

where we subtract β_2 from both estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ as our interest lies in forecasting future observations, which are functions of β_2 only, and the remainder that arises if $\tau \neq \tau_b$, is absorbed in the integral on the right hand side.

Define the projection matrix that projects off the columns of \bar{X} as $M_{\bar{X}} = I - \bar{X} (\bar{X}' \bar{X})^{-1} \bar{X}'$ and, additionally,

$$\begin{aligned} \boldsymbol{V} &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \\ \boldsymbol{Q} &= \bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}} \\ \boldsymbol{L} &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1} \\ \tilde{\boldsymbol{Q}} &= (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \end{aligned}$$
(19)

The inverse in (18) can be calculated using blockwise inversion. The result is the asymptotic variance covariance matrix of $(\hat{\boldsymbol{\beta}}_1(\tau)', \hat{\boldsymbol{\beta}}_2(\tau)', \hat{\boldsymbol{\delta}}')'$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{\tau} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ \tilde{\boldsymbol{Q}} & \frac{1}{1-\tau} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & -\boldsymbol{L}' & \boldsymbol{Q}^{-1} \end{pmatrix}$$
(20)

Hence,

$$\begin{aligned} \sqrt{T}(\hat{\boldsymbol{\beta}}_{1}(\tau) - \boldsymbol{\beta}_{2}) &\rightarrow \frac{1}{\tau} \left[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\boldsymbol{B}(\tau) + \int_{0}^{\tau} \boldsymbol{\eta}(s)ds \right] \\ &- (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{X}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{X}\boldsymbol{B}(1) \\ \sqrt{T}(\hat{\boldsymbol{\beta}}_{2}(\tau) - \boldsymbol{\beta}_{2}) &\rightarrow \frac{1}{1-\tau} \left[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'(\boldsymbol{B}(1) - \boldsymbol{B}(\tau)) + \int_{\tau}^{1} \boldsymbol{\eta}(s)ds \right] \quad ^{(21)} \\ &- (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{X}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{X}\boldsymbol{B}(1) \\ \sqrt{T}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) &\rightarrow (\bar{\boldsymbol{Z}}'\boldsymbol{M}_{X}\bar{\boldsymbol{Z}})^{-1}\bar{\boldsymbol{Z}}'\boldsymbol{M}_{X}\boldsymbol{B}(1) \end{aligned}$$

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 model Estimators that ignore the break converge to

$$\sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \rightarrow \begin{bmatrix} \bar{\boldsymbol{X}}' \bar{\boldsymbol{X}} & \bar{\boldsymbol{X}}' \bar{\boldsymbol{Z}} \\ \bar{\boldsymbol{Z}}' \bar{\boldsymbol{X}} & \bar{\boldsymbol{Z}}' \bar{\boldsymbol{X}} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\boldsymbol{X}}' \boldsymbol{B}(1) + \bar{\boldsymbol{X}}' \bar{\boldsymbol{X}} \int_0^1 \boldsymbol{\eta}(s) ds \\ \bar{\boldsymbol{Z}}' \boldsymbol{B}(1) + \bar{\boldsymbol{Z}}' \bar{\boldsymbol{Z}} \int_0^1 \boldsymbol{\eta}(s) ds \end{bmatrix}$$
(22)

Using the notation defined in (19), the inverse in (22) can be written as

$$\Sigma_F = \begin{pmatrix} V + \tilde{Q} & -L \\ -L' & Q^{-1} \end{pmatrix}$$
(23)

and, therefore,

$$\begin{aligned} \sqrt{T} \left(\hat{\boldsymbol{\beta}}_{F} - \boldsymbol{\beta}_{2} \right) &\to (\bar{\boldsymbol{X}}' \bar{\boldsymbol{X}})^{-1} \bar{\boldsymbol{X}}' \boldsymbol{B}(1) + \int_{0}^{1} \boldsymbol{\eta}(s) ds \\ &- (\bar{\boldsymbol{X}}' \bar{\boldsymbol{X}})^{-1} \bar{\boldsymbol{X}}' \bar{\boldsymbol{Z}} (\bar{\boldsymbol{Z}}' \boldsymbol{M}_{X} \bar{\boldsymbol{Z}})^{-1} \bar{\boldsymbol{Z}}' \boldsymbol{M}_{X} \boldsymbol{B}(1) \end{aligned} \tag{24} \\ \sqrt{T} \left(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \right) &\to (\bar{\boldsymbol{Z}}' \boldsymbol{M}_{X} \bar{\boldsymbol{Z}})^{-1} \bar{\boldsymbol{Z}}' \boldsymbol{M}_{X} \boldsymbol{B}(1) \end{aligned}$$

Note that for the parameters $\hat{\delta}$, the expression is identical to case where we model the break in β .

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

$$T\text{Cov}(\hat{\boldsymbol{\beta}}_{2}(\tau), \hat{\boldsymbol{\beta}}_{F}) = \boldsymbol{V} + \tilde{\boldsymbol{Q}} = T\text{Var}(\hat{\boldsymbol{\beta}}_{F})$$
(25)

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. $\operatorname{Cov}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\beta}}_F - \hat{\boldsymbol{\beta}}_2(\tau)) = \mathbf{0}$. The difference is that under a local structural break, both $\hat{\boldsymbol{\beta}}_F$ and $\hat{\boldsymbol{\beta}}_2(\tau)$ are not consistent.

4 Testing for a structural break

4.1 A break of known timing

Initially, we will assume that the timing of the break is known in order to illustrate part of our approach, which we will then extend to breaks of unknown timing. 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 $(\boldsymbol{x}_{T+1}, \boldsymbol{z}_{T+1})$

$$\hat{y}_{T+1} = f(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}) \tag{26}$$

where we omit the dependence on the regressors for notational convenience. For a known break date, the results of the previous paragraph imply the following asymptotic distribution of the parameters

$$\sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \to N \begin{bmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{pmatrix} \frac{1}{\tau} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ \tilde{\boldsymbol{Q}} & \frac{1}{1 - \tau} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & -\boldsymbol{L}' & \boldsymbol{Q}^{-1} \end{bmatrix} \end{bmatrix}$$
(27)

The full-sample estimator is given by

$$\hat{\boldsymbol{\beta}}_F = \hat{\boldsymbol{\beta}}_2 + \tau_b (\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) \tag{28}$$

and

$$\sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \to N \begin{bmatrix} \begin{pmatrix} \tau_b (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V} + \tilde{\boldsymbol{Q}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & \boldsymbol{Q}^{-1} \end{bmatrix}$$
(29)

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

$$\sqrt{T} \left(f(\hat{\boldsymbol{\beta}}_{2}, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_{2}, \boldsymbol{\delta}) \right) = \sqrt{T} \left[f_{\beta_{2}}'(\hat{\boldsymbol{\beta}}_{2} - \boldsymbol{\beta}_{2}) + f_{\delta}'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + O(T^{-1}) \right]
\rightarrow N \left(0, f_{\beta_{2}}' \operatorname{Var}(\hat{\boldsymbol{\beta}}_{2}) f_{\beta_{2}} + q \right)
\sqrt{T} \left(f(\hat{\boldsymbol{\beta}}_{F}, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_{2}), \boldsymbol{\delta}) \right) = \sqrt{T} \left[f_{\beta_{2}}'(\hat{\boldsymbol{\beta}}_{F} - \boldsymbol{\beta}_{2}) + f_{\delta}'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + O(T^{-1}) \right]
\rightarrow N \left(\tau_{b}(\boldsymbol{\beta}_{1} - \boldsymbol{\beta}_{2}), f_{\beta_{2}}' \operatorname{Var}(\hat{\boldsymbol{\beta}}_{F}) f_{\beta_{2}} + q \right)$$
(30)

where $q = f'_{\delta} \operatorname{Var}(\hat{\delta}) f_{\delta} + 2f'_{\beta_2} \operatorname{Cov}(\hat{\beta}_F, \hat{\delta}) f_{\delta}$ will drop out later on, and we use that, asymptotically, $\operatorname{Cov}(\hat{\beta}_F, \hat{\delta}) = \operatorname{Cov}(\hat{\beta}_2, \hat{\delta})$. Using previous results on the covariance matrix of the estimators, and the notation in (19), we have

$$f_{\beta_2}^{\prime} \operatorname{Var}(\hat{\boldsymbol{\beta}}_2) f_{\beta_2} = \frac{1}{1 - \tau_b} f_{\beta_2}^{\prime} \boldsymbol{V} f_{\beta_2} + f_{\beta_2}^{\prime} \tilde{\boldsymbol{Q}} f_{\beta_2}$$

$$f_{\beta_2}^{\prime} \operatorname{Var}(\hat{\boldsymbol{\beta}}_F) f_{\beta_2} = f_{\beta_2}^{\prime} \boldsymbol{V} f_{\beta_2} + f_{\beta_2}^{\prime} \tilde{\boldsymbol{Q}} f_{\beta_2}$$
(31)

For the expected MSFEs using β_2 and β_F , we have

$$TE\left[\left(f(\hat{\boldsymbol{\beta}}_{2},\hat{\boldsymbol{\delta}})-f(\boldsymbol{\beta}_{2},\boldsymbol{\delta})\right)^{2}\right] = \frac{1}{1-\tau_{b}}f_{\beta_{2}}'\boldsymbol{V}f_{\beta_{2}} + f_{\beta_{2}}'\tilde{\boldsymbol{Q}}f_{\beta_{2}} + q$$
$$TE\left[\left(f(\hat{\boldsymbol{\beta}}_{F},\hat{\boldsymbol{\delta}})-f(\boldsymbol{\beta}_{2},\boldsymbol{\delta})\right)^{2}\right] = \left[\tau_{b}f_{\beta_{2}}'(\boldsymbol{\beta}_{1}-\boldsymbol{\beta}_{2})\right]^{2} + f_{\beta_{2}}'\boldsymbol{V}f_{\beta_{2}} + f_{\beta_{2}}'\tilde{\boldsymbol{Q}}f_{\beta_{2}} + q$$

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

$$\zeta = T(1-\tau_b)\tau_b \frac{\left[f_{\beta_2}'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2)\right]^2}{f_{\beta_2}' \boldsymbol{V} f_{\beta_2}} \le 1$$
(32)

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

$$\hat{\zeta} = T(1-\tau_b)\tau_b \frac{\left[f'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\hat{\omega}} \sim \chi^2(1,\zeta)$$
(33)

where $\hat{\omega}$ is a consistent estimator of $f'_{\beta_2} V f_{\beta_2}$. 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

$$\zeta = T\tau_b (1 - \tau_b) \frac{\left[\mathbf{x}'_{T+1} (\mathbf{\beta}_1 - \mathbf{\beta}_2) \right]^2}{\mathbf{x}'_{T+1} \mathbf{V} \mathbf{x}_{T+1}} \le 1$$
(34)

identical to the result in (5).

4.2 Forecasting under a local break of unknown timing

If the timing of the break is unknown and $\tau < \tau_b$, then the estimator of β_2 is biased as can be seen from the last term in (21). The difference in expected asymptotic MSFE between the post-break sample based forecast and the full-sample based forecast is

$$\begin{split} \tilde{\Delta} &= R(\hat{\boldsymbol{\beta}}_{2}(\hat{\tau}), \hat{\boldsymbol{\delta}}) - R(\hat{\boldsymbol{\beta}}_{F}, \hat{\boldsymbol{\delta}}) \\ &= T\left\{ \mathbf{E} \left[(f(\hat{\boldsymbol{\beta}}_{2}(\hat{\tau}), \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_{2}, \boldsymbol{\delta}))^{2} \right] - \mathbf{E} \left[(f(\hat{\boldsymbol{\beta}}_{F}, \hat{\boldsymbol{\delta}}) - f(\boldsymbol{\beta}_{2}, \boldsymbol{\delta}))^{2} \right] \right\} \\ &= T\left\{ \mathbf{E} \left[\left(f'_{\beta_{2}}(\hat{\boldsymbol{\beta}}_{2}(\hat{\tau}) - \boldsymbol{\beta}_{2}) \right)^{2} \right] - \mathbf{E} \left[f'_{\beta_{2}}(\hat{\boldsymbol{\beta}}_{F} - \boldsymbol{\beta}_{2}) \right]^{2} - f'_{\beta_{2}} \operatorname{Var}(\hat{\boldsymbol{\beta}}_{F}) f_{\beta_{2}} \right\} \end{split}$$

where $R(\hat{\theta})$ is the asymptotic MSFE under parameter estimates $\hat{\theta}$. The derivations are provided in Appendix A. Standardizing by $f'_{\beta_2} V f_{\beta_2}$ yields

$$\Delta = \frac{R(\hat{\beta}_{2}(\hat{\tau}), \hat{\delta}) - R(\hat{\beta}_{F}, \hat{\delta})}{f_{\beta_{2}}^{\prime} V f_{\beta_{2}}}$$

$$= E\left\{ \left[\frac{1}{1-\hat{\tau}} \frac{f_{\beta_{2}}^{\prime} V \bar{X}^{\prime} (B(1) - B(\hat{\tau}))}{\sqrt{f_{\beta_{2}}^{\prime} V f_{\beta_{2}}}} + \frac{1}{1-\hat{\tau}} \int_{\hat{\tau}}^{1} \frac{f_{\beta_{2}}^{\prime} \eta(s)}{\sqrt{f_{\beta_{2}}^{\prime} V f_{\beta_{2}}}} ds \right]^{2} \right\}$$

$$- \left(\int_{0}^{1} \frac{f_{\beta_{2}}^{\prime} \eta(s)}{\sqrt{f_{\beta_{2}}^{\prime} V f_{\beta_{2}}}} ds \right)^{2} - 1 \qquad (35)$$

$$= E_{f(\hat{\tau})} \left\{ \left[\frac{1}{1-\hat{\tau}} (B(1) - B(\hat{\tau})) + \theta \frac{\tau_{b} - \hat{\tau}}{1-\hat{\tau}} I[\hat{\tau} < \tau_{b}] \right]^{2} \right\} - \theta^{2} \tau_{b}^{2} - 1$$

where $\theta = \frac{f'_{\beta_2} \eta(\tau)}{\sqrt{f'_{\beta_2} V f_{\beta_2}}}$ with the break defined as $\eta(\tau) = bI[\tau < \tau_b]$. By a continuous mapping theorem $\frac{f'_{\beta_2} V \bar{X}'[B(1) - B(\tau)]}{\sqrt{f'_{\beta_2} V f_{\beta_2}}} = B(1) - B(\tau)$ with $B(\cdot)$ a one-dimensional Brownian motion.¹ If $\hat{\tau} = \tau_b$, the critical break size of

a one-dimensional Brownian motion.¹ If $\hat{\tau} = \tau_b$, the critical break size of the previous section is obtained. However, under an unknown break date $\hat{\tau} \neq \tau_b$ in general and (35) cannot be immediately used for testing purposes. However, since Δ is symmetric around $\theta = 0$, and $\Delta > 0$ for $\theta = 0$ and (35) quadratically decreases away from $\theta = 0$, there is a unique break size $|\theta|$ for each τ_b for which $\Delta = 0$. This makes the break size an excellent candidate as a test statistic and consequently we consider as before the Wald test statistic

$$W(\tau) = T \frac{\left[f_{\beta_2}'(\hat{\boldsymbol{\beta}}_2(\tau) - \hat{\boldsymbol{\beta}}_1(\tau)) \right]^2}{f_{\beta_2}'\left(\frac{\hat{\boldsymbol{V}}_1}{\tau} - \frac{\hat{\boldsymbol{V}}_2}{1-\tau} \right) f_{\beta_2}}$$
(37)

analogous to (33) when the break date is known. As we show below, in case of a rejection when testing at small enough size, this test statistic identifies

$$\Delta = \mathcal{E}_{f(\tau)} \left\{ \left[\frac{1}{1 - \tau} (B(1) - B(\tau)) + \frac{1}{1 - \tau} J(\tau) \right]^2 \right\} - J(1)^2 - 1$$
(36)

¹Note that up to the last line of (35), we have not made assumptions regarding the form of the instability, which is governed by $\eta(\tau)$. Defining $J(\tau) = \int_{\tau}^{1} \frac{f'_{\beta_2} \eta(s)}{\sqrt{f'_{\beta_2} V f_{\beta_2}}} ds$, we would have

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 obviously simplifies if the size of the moving window is exogenously set to some fraction of the total number of observations.

the true break date up to a constant that vanishes with decreasing size. This establishes a weak form of optimality of the sup-Wald test. Since the break size θ for which $\Delta = 0$ is different for each τ_b , our null hypothesis will change with the unknown τ_b for which no consistent estimator is available. We will show below how a test can be constructed in this scenario that maintains this weak optimality.

Since the function f is fixed, the results in Andrews (1993) and the continuous mapping theorem show that $W(\tau)$ in (37) converges to

$$Q^{*}(\tau) = \left(\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1 - \tau)}} + \sqrt{\frac{1 - \tau}{\tau}} \int_{0}^{\tau} \eta(s) ds - \sqrt{\frac{\tau}{1 - \tau}} \int_{\tau}^{1} \eta(s) ds\right)^{2} \\ = \left(\frac{B(\tau) - \tau B(1)}{\sqrt{\tau(1 - \tau)}} + \mu(\tau; \theta_{\tau_{b}})\right)^{2}$$
(38)

For the optimality results below require the following assumption.

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

The first term of (38) 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 $\mu(\tau; \theta_{\tau_b})^2$. We will now show that if we test at a small nominal size, rejections are found only for break locations that are close to τ_b . We provide explicit expressions for the region around τ_b in which rejections are found. These results are used to show that we can set up a powerful test even if the null hypothesis depends on the unknown break date.

Specializing to the structural break model, we have

$$\mu(\tau;\theta_{\tau_b}) = \theta_{\tau_b} \left[\sqrt{\frac{1-\tau}{\tau}} \tau_b I[\tau_b < \tau] + \sqrt{\frac{\tau}{1-\tau}} (1-\tau_b) I[\tau_b \ge \tau] \right]$$
(39)

which indeed satisfies Assumption 2. The extremum value is $\mu(\tau_b; \theta_{\tau_b}) = \theta_{\tau_b} \sqrt{\tau_b(1-\tau_b)}$.

5 Weak optimality

The proof of weak optimality will proceed as follows. Using arguments of Piterbarg (1996), we first prove 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 u tends to infinity or, equivalently, 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 . Roughly speaking, if the critical values vary 'slowly' with τ_b , than 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.

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 $u \to \infty$

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

where $I = [\tau_{\min}, \tau_{\max}], I_1 = [\tau_b - \delta, \tau_b + \delta]$ and

$$\delta = u^{-1} \log^2 u$$

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

$$\begin{split} P\left(\sup_{\tau\in I}Q^*(\tau)>u^2\right) &= P\left(\sup_{\tau\in I}\sqrt{Q^*(\tau)}>u\right)\\ &= P\left(\sup_{\tau\in I}|Z(\tau)+\mu(\tau;\theta_{\tau_b})|>u\right)\\ &= P\left(\sup_{\tau\times c}[Z(\tau)+\mu(\tau;\theta_{\tau_b})]c>u\right) \qquad \text{with } c=\pm1 \end{split}$$

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

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 $\tau \in I$ and c.

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

$$P(Z(\tau) + \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) = P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in I)$$

$$P(-Z(\tau) - \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) = P(Z(\tau) > u + |\mu(\tau; \theta_{\tau_b})|, \tau \in I)$$
(40)

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

$$\begin{aligned}
P(-Z(\tau) - \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) &= P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|, \tau \in I) \\
P(Z(\tau) + \mu(\tau; \theta_{\tau_b}) > u, \tau \in I) &= P(Z(\tau) > u + |\mu(\tau; \theta_{\tau_b})|, \tau \in I)
\end{aligned} \tag{41}$$

The bounds in the second lines of (40) and (41) 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 $u \to \infty$

$$P\left(\sup_{\tau \times c} [Z(\tau) + \mu(\tau; \theta_{\tau_b})]c > u\right) = P\left(\sup_{\tau \in I} Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})|\right) (1 + o(1))$$
(42)

We now continue to show that in (42), exceedances occur in a small subset of *I*. For the theory to apply to the structural break model, we assume $Z(\tau)$ to be 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 $0 < C(\tau) < \infty$

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

This implies that the correlation function can be approximated as

$$r(\tau, \tau + s) = 1 - C(\tau)|s|^{\alpha} + o(|s|^{\alpha})$$
 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 $u \to \infty$

$$P\left(\sup_{[\tau,\tau+\delta(u)]} Z(t) > u\right) = \frac{1}{\sqrt{2\pi}} \delta(u) u \exp\left(-\frac{1}{2}u^2\right) C(\tau)$$
(43)

Proof: see Hüsler (1990).

Proof of Theorem 1 (continued): The following is based on the approach by Piterbarg (1996). Consider a region close to τ_b defined by $I_1 = [\tau_b - \delta(u), \tau_b + \delta(u)]$. We want to show that, if $\delta(u) \to 0$ as $u \to \infty$ at an appropriate rate, then I_1 contains all relevant information on the probability of crossing the boundary. In I_1 , the minimum value of the boundary is given by

$$c_1 = \inf_{\tau \in I_1} [u - \mu(\tau; \theta_{\tau_b})] = u - |\mu(\tau_b; \theta_{\tau_b})|$$
(44)

so that

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

$$\leq P(Z(\tau) > c_1 \text{ for some } \tau \in I_1)$$

$$= 2\delta(u)c_1 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}c_1^2\right) C(\tau_b - \delta(u))$$

$$= \frac{2\delta(u)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}c_1^2 + \log c_1\right) C(\tau_b - \delta(u))$$

where the third line follows from (43).

Define now also regions that are increasingly further from τ_b as

$$I_{k} = \{ [\tau_{b} - k\delta(u), \tau_{b} - (k-1)\delta(u)], [\tau_{b} + (k-1)\delta(u), \tau_{b} + k\delta(u)] \}$$

In I_k , the minimum value of the boundary satisfies

$$c_k \ge c_{k,min} = u - |\mu(\tau_b + \delta(u); \theta_{\tau_b})| \tag{45}$$

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

$$\mu(\tau_b + \delta(u); \theta_{\tau_b}) = \mu(\tau_b; \theta_{\tau_b}) + \gamma \delta(u) + O(\delta(u)^2)$$
(46)

where $\gamma = \left. \frac{\partial \mu(\tau; \theta_{\tau_b})}{\partial \tau} \right|_{\tau = \tau_b}$. Using (45) we have

$$P_{I_k} = P(Z(\tau) > u - |\mu(\tau; \theta_{\tau_b})| \text{ for some } \tau \in I_k)$$

$$\leq P(Z(\tau) > c_{k,min} \text{ for some } \tau \in I_k)$$

$$= \frac{\delta(u)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}c_1^2 - c_1\gamma\delta(u) - \frac{1}{2}\gamma^2\delta(u)^2 + \log(c_1 + \gamma\delta(u))\right) C(\tau_k)$$
(47)

with $\tau_k \in I_k$, the precise value of τ_k is not essential to the proof. Then

$$\sum_{k \neq 1} P_{I_k} = \frac{\delta(u)}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}c_1^2 - c_1\gamma\delta(u) - \frac{1}{2}\gamma^2\delta(u)^2 + \log(c_1 + \gamma\delta(u))\right) \sum_{k \neq 1} C(\tau_k)$$
(48)

For the sum $\sum_{k\neq 1} C(\tau_k)$ it is sufficient to note that it is independent of c_1 and u. Compare (48) to the probability of a test with a known break date to exceed the critical value

$$P_0 = P(Z(\tau_b) > u - |\mu(\tau_b; \theta_{\tau_b})|) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}c_1^2 - \log(c_1)\right)$$
(49)

where we use that

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

We see that (48) contains an extra term $\exp(-c_1\gamma\delta(u))$ compared to (49). Using (44), this implies that $\sum_{k\neq 1} P_{I_k} = o(P_0)$ if

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

Subsequently, if

$$\delta(u) = u^{-1} \log^2(u) \tag{50}$$

then all intervals outside of I_1 contribute $o(P_0)$ to the probability of crossing the boundary u. Under (50), we have that for P_{I_1} as $u \to \infty$

$$P_{I_1} \le P_I \le P_{I_1} + \sum_{k \ne 1} P_{I_k}$$
$$\le P_{I_1} + o(P_0)$$

We now only need to note that

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

$$\geq P(Z(\tau_b) > u - |\mu(\tau_b; \theta_{\tau_b})|) = P_0$$

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: In (48), the term that ensures that $P_{I_k} = o(P_1)$ is $\exp(c_1\delta(u))^{-\gamma}$. In the structural break model, we see that (46) is given by $\mu(\tau_b + \delta(u); \theta_{\tau_b}) = \theta_{\tau_b}\sqrt{\tau_b(1-\tau_b)} - \frac{1}{2}\theta_{\tau_b}\frac{1}{\sqrt{\tau_b(1-\tau_b)}}\delta(u) + O(\delta(u)^2)$. It is clear that γ scales linearly with the break size. Therefore, if the break size is sufficiently high, we expect the optimality results to extend to the practical case when u is finite. This is confirmed by the simulations of asymptotic power presented in 6.

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.

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

This implies that post-test parameter inference after a rejection of the null is standard. It also provides a weak form of optimality for the break date that results from maximizing finite sample Wald statistics that converge to Q^* given by (38).

The testing problem we consider is complicated by the fact that the null hypothesis depend on the unknown break location. This translates into critical values that will also depend on the unknown break date. The location concentration theorem 1 indicates that plugging in the estimate of the break date that maximizes the Wald statistic could be a viable strategy to obtain a well behaved test. Indeed, 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 critical values that control size for every τ_b , i.e. $u = u(\tau_b)$ and

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

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 $[\tau_b - \delta, \tau_b + \delta]$, 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 γ in (46) is given by $\gamma = \frac{\partial \mu(\tau; \theta_{\tau_b})}{\partial \tau} = \theta_{\tau_b} \frac{1}{\sqrt{\tau_b(1-\tau_b)}}$. 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

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

The slowly varying assumption therefore holds if

$$\frac{\partial u(\tau_b)}{\partial \tau_b} \le \frac{1}{\tau_b(1-\tau_b)} \tag{51}$$

which can be verified once critical values have been obtained.

We now provide a result on the optimality of the presented test under Assumption 3.

Theorem 2 (Weak optimality) Under a slowly varying boundary

$$P_{H_{a}}\left(\sup_{\tau} Q^{*}(\tau) > u(\hat{\tau})^{2}\right) - P_{H_{a}}(Q^{*}(\tau_{b}) > v(\tau_{b}))$$

$$\geq P_{H_{a}}(Q^{*}(\tau_{b}) > u(\tau_{b})) - P_{H_{a}}(Q^{*}(\tau_{b}) > v(\tau_{b}))$$

$$= 0$$
(52)

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

Proof: As before

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

Under the slowly varying assumption, $u(\hat{\tau}) - \mu(\hat{\tau}; \theta_{\tau_b})$ has a unique minimum on I_1 at $\hat{\tau} = \tau_b$. Then taking the supremum necessarily leads to more exceedances than only considering $\tau = \tau_b$, which proves the inequality in (52). The last line of (52) is ensured by the following lemma

Lemma 3 (Convergence of critical values) Let $u(\tau_b)$ be the critical that control size for each τ_b . Furthermore, let $v(\tau_b)$ be the critical values of the test that considers the break date fixed at its true value, then $u(\tau_b) - v(\tau_b) \rightarrow 0$.

Proof: By definition of the critical values

$$P(\sup_{\tau} Q^{*}(\tau) > u(\tau_{b})^{2}) = P(Z(\tau) > u(\tau_{b}) - |\mu(\tau; \theta_{\tau_{b}})| \text{ for some } \tau \in I_{1}) = \alpha$$
$$P(Q^{*}(\tau_{b}) > v(\tau_{b})^{2}) = P(Z(\tau_{b}) > v(\tau_{b}) - |\mu(\tau_{b}; \theta_{\tau_{b}})|) = \alpha$$

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, $\max u(\tau_b) - v(\tau_b) = O(\delta(u))$. In fact, by arguments similar to those in the proof of Theorem 1 one could argue that $\max u(\tau_b) - v(\tau_b) = O(\log(u)/u)$ which corresponds to the results in Andrews (1993). In any case, since $\delta(u) = u^{-1} \log^2 u$, we have that $u(\tau_b) - v(\tau_b) \to 0$ as $u \to \infty$.

Corollary 2 (A test statistic with critical values independent of τ_b) A test statistic where the critical values are independent of τ_b for $u \to \infty$ is given by

$$\kappa(\hat{\tau}) = \sup_{\tau \in I} \sqrt{T} \frac{\left| f_{\beta_2}'\left(\hat{\beta}_2(\tau) - \hat{\beta}_1(\tau)\right) \right|}{\sqrt{f_{\beta_2}'\left(\frac{\hat{\mathbf{V}}_1}{\tau} + \frac{\hat{\mathbf{V}}_2}{1-\tau}\right) f_{\beta_2}}} - |\mu(\hat{\tau}; \theta_{\hat{\tau}})| \tag{53}$$

where $\hat{\tau}$ maximizes the first term of κ , or equivalently the Wald statistic (37).

Figure 1: Break size for equal MSFE between the full sample and post-break sample forecast



Proof: The test statistic converges to $\kappa(\hat{\tau}) \to \sup_{\tau} |Z(\tau) + \mu(\tau; \theta_{\tau_b})| - |\mu(\hat{\tau}; \theta_{\hat{\tau}})|$ 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

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

Since the difference $-|\mu(\hat{\tau};\theta_{\tau_b})| + |\mu(\hat{\tau};\theta_{\hat{\tau}})| = O(\delta(u))$, the critical values of $\kappa(\hat{\tau})$ are independent of τ_b in the limit where $u \to \infty$.

6 Simulations

Since our theory considers the case where the nominal size tends to zero, we investigate how the weak optimality translates to conventional choice of the size, $\alpha = \{0.10, 0.05, 0.01, 0.005\}$, using simulations, where we use 200,000 repetitions for each break date and size. The [0, 1] interval is divided in 1,000 equally spaced parts, similar to Bai and Perron (1998). The possible break dates are allowed to be on the interval [τ_{\min}, τ_{\max}] with $\tau_{\min} = 1 - \tau_{\max} = 0.15$.

6.1 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 (35) and Figure 1 shows the combinations of break size and break date for which the same MSFE is obtained. Note that the vertical axis is in units of the standardized break size, so that it can be interpreted as a standard deviations from a standard normal. The plot shows that the break size for $\tau_b = 0.15$ is very large and decreases as τ_b increases. As expected, the break size is uniformly larger than 1, which is the break size under a known break date.

6.2 Critical values, size and power

After finding the line in (τ_b, θ) space for which $\Delta = 0$, we compute critical values by simulating our Wald-type test statistic (37) and the α -asymptotic statistic (53) for pairs of (τ_b, θ) . Condition (51) which should hold for the weak optimality results to apply is verified in Figure 9 in Appendix B. Should either the distribution of $\hat{\tau}$ be very concentrated around τ_b or, alternatively, the dependence of the critical values on τ_b very weak, then, using $\hat{\tau}$, these critical values could be used without further correction. While the approximation using $\hat{\tau}$ is indeed close, applying a small numerical correction leads to size control uniformly over τ_b .

The critical values are displayed in Figure 2a. The large break size that yields equal forecast accuracy implies that a major increase in critical values compared to the standard values of Andrews (1993), which are for a size of [0.10, 0.05, 0.01] equal to [7.17, 8.85, 12.35]. The critical values for the α -asymptotic test statistic (53) are expected to be independent of $\hat{\tau}$ in the limit where $\alpha \to 0$ and can be compared to the usual critical values that would be obtained under a known break date from a one-sided normal distribution, i.e. [2.58, 2.33, 1.64, 1.28]. Figure 2c displays the critical values and it can be seen that for finite α the dependence on $\hat{\tau}$ is considerably smaller than for the Wald statistic. Also, size is controlled better for different values of the true break date, even without numerical corrections. The results in Section 5 suggest that the difference to the critical values diminishes as $\alpha \to 0$, and this is indeed what we can observe. In the plots in Figure 2, the solid lines are obtained by numerically adjusting the critical values to ensure size control over all values of τ_b , which can be seen to be very effective.

Given that the break sizes that lead to equal forecast performance are reasonably large, we expect the test to have relatively good properties in terms of power. This is confirmed in Figure 3, which shows that the power of the test is close to the power of the optimal test which uses a known break date. This is observed for all locations of the break and confirms that the theoretical results for vanishing nominal size extend to conventional choices of nominal size for break sizes considered under the null hypothesis of equal forecast accuracy.

Figure 2: Critical values and size for Wald test statistic (37) and α -asymptotic test statistic (53)



Note: Dashed lines are critical values that control size for each τ_b separately. The continuous lines are obtained when adjusting the critical values to ensure size control uniformly over τ_b . The pairs of lines are for $\alpha = 0.1, 0.05, 0.01, 0.005$.



Figure 3: Power curves for various values of τ_b for $\alpha = 0.05$

Note: Power of a test conditional on the break date is plotted in green. The power of the test based on the Wald statistic (37) and (53) are not distinguishable and plotted in blue.

7 Optimal weights and shrinkage forecasts

Pesaran et al. (2013) derived 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 weights were derived under the assumption of a known break date and size, although they also derived weights that were robust to break date uncertainty by integrating the break date. A feature of the weights is that, conditional on a break date, they take one value for the observations before the break and another value for the observations after the break. It is then easy to show that the forecast that uses the optimal weights can be expressed as a shrinkage forecast of the form

$$egin{aligned} \hat{y}^S_{T+1} &= \omega oldsymbol{x}'_{T+1}oldsymbol{eta}_1 + (1-\omega)oldsymbol{x}'_{T+1}oldsymbol{eta}_2 \ &= \omega oldsymbol{x}'_{T+1}(oldsymbol{eta}_1 - oldsymbol{eta}_2) + oldsymbol{x}'_{T+1}oldsymbol{eta}_2 \end{aligned}$$

Using that $E[\mathbf{x}'_{T+1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)] = \mathbf{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2), T \operatorname{Var} \hat{\boldsymbol{\beta}}_1 = \frac{1}{\tau_b} V, T \operatorname{Var} \hat{\boldsymbol{\beta}}_2 = \frac{1}{1-\tau_b} V$ and $T \operatorname{Cov}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = 0$, we can derive the weights that minimize $E[(\hat{y}_{T+1}(\omega) - y_{T+1})^2]$ as

$$\omega = \frac{\tau_b}{1 + W(\tau)}$$

with $W(\tau) = T\tau_b(1-\tau_b) \frac{\left[\mathbf{z}'_{T+1}(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2) \right]^2}{\mathbf{z}'_{T+1}\mathbf{V}\mathbf{z}_{T+1}}$ is our Wald-type test statistic. Given that $\hat{\boldsymbol{\beta}}_F = \tau_b \hat{\boldsymbol{\beta}}_1 + (1-\tau_b) \hat{\boldsymbol{\beta}}_2$, we can derive the optimal weight that is given to the full sample estimator, which is

$$\omega = \frac{1}{1 + W(\tau_b)} \tag{54}$$

Plugging in estimates of the unknown quantities yields

$$\mathbf{x}_{T+1}'\hat{\boldsymbol{\beta}}_{S}(\hat{\tau}) = \frac{1}{1+W(\hat{\tau})}\mathbf{x}_{T+1}'\hat{\boldsymbol{\beta}}_{F} + \frac{W(\hat{\tau})}{1+W(\hat{\tau})}\mathbf{x}_{T+1}'\hat{\boldsymbol{\beta}}_{2}(\hat{\tau})$$

By a continuous mapping theorem, under local breaks the weights converge as

$$\frac{1}{1+W(\tau)} \stackrel{T \to \infty}{\to} \frac{1}{1+Q^*(\tau)}$$

allowing for asymptotic analysis of these plug-in weights.

While the optimal weights take the estimation uncertainty around the parameters into account, they are conditional on the break date. Ignoring the estimation uncertainty around the break date will imply a too large weight on the post-break forecast, which will suffer from imprecise break date estimation whereas the full sample forecast does not require a break date. The optimal weights forecast will therefore underperform for small, poorly identified breaks. Figure 4: Break size of equal expected MSFE of full sample and optimal weights forecast (green) and full sample and post-break forecast (blue)



Similar to above, we can find the break size for which the optimal weights estimator provides more accurate forecasts than the full sample estimator. Figure 4 plots the combination of τ_b and break size for which the two forecasts have the same MSFE (green line) together with the line where the post-break forecast and the full sample forecast have the same MSFE (blue line). It can be seen that the line for the optimal weights forecast is below the line for the post-break forecast, which implies that the optimal weights 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 optimal weights estimator is more precise than the full sample estimator.

In order to test the applicability of the optimal weights estimator, critical values can be obtained in a similar fashion as before and are presented in Figure 5. Since the break sizes for equal forecast performance in Figure 4 are close, it is not surprising that the properties in terms of size and power of the test are be largely the same as those for the post-break forecast. The critical values given in Figure 5a lead to excellent size control in Figure 5b. Also, the power is near optimal, as shown in Figures 5c and 5d.

Additionally, we can investigate the break sizes that leads to equal forecast performance of the the post-break forecast and the optimal weights forecast. Figure 6 plots the ratio of the MSFE of the optimal weights forecast over that of the post-break forecast. It turns out that for large breaks, the post-break estimator outperforms the optimal weights. However, the improvement in MSFE is relatively small.

Figure 5: Critical values, size and power when testing between optimal weights forecast and full stample forecast



Solid lines in plots (a) and (b) are obtained after a numerical correction to ensure size control uniformly over τ_b .



Figure 6: Relative MSFE of optimal weights forecast to post-break forecast

8 Application

We obtain a set of 135 macroeconomic and financial time series from the St. Louis Federal Reserve database, which is a monthly updated database described in detail by McCracken and Ng (2014).² 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.

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, 14 series), money and credit (MC, 14 series), interest rates and exchange rates (IRER, 22 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 6. We estimate parameters on a moving windows of 120 observations to decrease the likelihood of multiple breaks occurring. Test results are based on heteroskedasticity robust Wald statistics, which use $\hat{V}_i = (X'_i X_i)^{-1} X'_i \hat{\Omega}_i X_i (X'_i X_i)^{-1}$ with $[\hat{\Omega}_i]_{kl} = \hat{\varepsilon}_k^2/(1-h_k)^2$ if k = l and $[\hat{\Omega}_i]_{kl} = 0$ otherwise, and h_k is the k-th diagonal element of $P_X = X(X'X)^{-1}X'$. See MacKinnon and White (1985) and Long and Ervin (2000) for discussions of different heteroskedasticity robust covariance matrices. We have also obtained test results and forecasts using a larger window of 240 observations and using the homoskedastic Wald test and, qualitatively, our results do not depend on these choices.

²https://research.stlouisfed.org/econ/mccracken/fred-databases/

Table 2: Fractions of estimation samples with a significant structural break

AR(p)	supW	f-supW
1	0.306	0.068
6	0.172	0.008

Note: p refers to the number of lags of the estimated autoregressive model. supW refers to the standard sup Wald test, while f-supW refers to the test developed in this paper. All tests are carried out at $\alpha = 0.05$.

Two pre-sample observations are required to transform the data to stationarity and additional p observations are used to initialize the AR(p) model. After accounting for all necessary pre-samples, our first forecast is for September 1979 and we recursively construct one-step ahead forecasts until the end of the sample. If a missing value enters the moving window, no forecast is made.

8.1 Structural break test results

In this forecast exercise, we will refer to the test of Andrews (1993) as supWand to the test developed in this paper as f-sup W. In Table 2, we report the fraction of estimation samples where supW would indicate a break. This is contrasted with the fraction where the two tests indicate a break at a nominal size of $\alpha = 0.05$. It is clear that the majority of the breaks picked up by supW are judged as irrelevant for forecasting by f-sup W. The fraction of forecasts for which a break is indicated is lower by a factor of over four for the AR(1) and by factor of twenty for the AR(6).

Figure 7 displays the number of estimation samples for which the tests were significant, where within each category we sort the series based on the fraction of breaks found by f-sup W. For conciseness, we focus on the AR(1) but the results for the AR(6) are qualitatively similar. Across all categories the sup Wald test is more often significant than the f-sup W 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 40% of the estimation samples when the f-sup W test is 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 8 shows the occurrence of significant breaks over the different estimation samples, with the end date of the estimation sample given on the horizontal axis. It is clear that the sup W test finds more breaks in every single period. The periods of larger number of breaks is more persistent for

Figure 7: Fraction of significant structural break test statistics per series



Note: Dashed line indicates the fraction of estimation samples with significant sup-Wald test, the solid line indicates the fraction of estimation samples with significant f-sup-Wald test.

Figure 8: Fraction of significant structural break test statistics over time



Note: Fraction of forecasts where a breaks per series in the AR(1) model. Dashed line indicates the fraction of estimation samples with significant sup Wald test, the solid line indicates the fraction of estimation samples with significant f-sup Wald test. The dates displayed on the horizontal axis are the end dates of the estimation samples.

the sup W test than for the f-sup W test, which confirms the finding that, for forecasting, breaks are mainly of importance if the are late in the estimation sample, and as time passes the same break will have a smaller influence on the forecast accuracy. An example of this is the recent financial crisis, which leads to an increase in detected breaks around 2009 for both tests. For the f-sup W test, however, after a few months the number of significant breaks decreases again whereas for the sup W test the number of breaks increases further and most significant breaks occur towards the end of our sample.

8.2 Forecast accuracy

We consider a forecasting procedures that use structural break tests to decide whether to model a break. The first uses Andrews' sup W test. If a significant break is found, the forecast will be based on the post-break window, otherwise the full estimation window is used. The second uses our f-sup W test to decide between post-break and full sample windows. Next, we use the optimal weights forecasts of Pesaran et al. (2013). In the first variant we use apply optimal weights irrespective of whether a significant break was found. The second variant applies optimal weights only if the sup W test finds a significant break, and the third variant only if the f-sup Wtest finds a significant break, where all tests are carried out at $\alpha = 0.05$.

Table 3 reports the MSFE of the respective forecasting procedures relative to the MSFE of the forecast based on the supW test of Andrews with the results for the AR(1) in the top panel and those for the AR(6) in the bottom panel. For each model, we report the average relative MSFE over all forecasts in the first line, followed by the average relative MSFE for the series in the different categories. We show the results for all forecast periods on the left side of the table. However, these results contain many forecasts where both tests agree that no break is in the estimation sample and therefore both use the full estimation sample to yield identical results. To bring the difference between the tests out more, we also report the results excluding forecasts when both tests reject the null of a break. These results are on the right side of the table.

The second column of the table shows that using post-break estimation samples using the f-sup W test in place of the sup W test reduces the MSFE for both models on average and for all categories, with the exception of money and credit when the AR(1) is used. Using optimal weight irrespective of the significance of breaks leads to substantially less precise forecasts than using post-break samples suggested by the sup W test, and this is true for both models and nearly all categories. Using optimal weights over post-break window forecasts marginally reduces the MSFE for both models. When using the f-sup W test to decide on optimal weights the gain is much larger.

Considering the right hand side of the table, it is clear that the differences are much larger when forecasts from estimation sample that contain no

	All forecasts				Forecasts with sign. break		
	post-break	optimal weights		post-break	optimal weights		
	f-supW	all	supW	f-supW	f-supW	supW	f-supW
AR(1)							
All series	0.981	1.016	0.995	0.981	0.942	0.987	0.938
OI	0.977	1.059	0.996	0.978	0.950	0.992	0.952
LM	0.971	1.042	0.998	0.967	0.937	0.994	0.929
CO	0.961	0.988	0.968	0.955	0.945	0.957	0.936
OrdInv	0.982	1.008	0.992	0.984	0.955	0.980	0.959
MC	1.001	1.080	1.001	1.001	1.016	1.005	1.011
IRER	0.988	0.997	1.000	0.987	0.921	1.007	0.917
Р	0.998	1.049	0.999	0.998	0.932	0.987	0.941
\mathbf{S}	0.983	1.023	0.998	0.984	0.904	0.986	0.913
AR(6)							
All series	0.961	1.205	0.994	0.963	0.819	0.979	0.831
OI	0.941	1.216	0.985	0.941	0.762	0.952	0.761
LM	0.971	1.244	0.996	0.972	0.846	0.978	0.853
CO	0.902	1.108	0.972	0.907	0.779	0.939	0.789
OrdInv	0.945	1.137	0.996	0.945	0.827	0.998	0.828
MC	0.979	1.366	0.989	0.979	0.842	0.986	0.834
IRER	0.969	1.207	1.002	0.968	0.792	1.008	0.798
Р	0.975	1.256	0.996	0.979	0.887	0.994	0.916
\mathbf{S}	0.976	1.179	1.003	0.997	0.836	1.020	1.004

Table 3: Relative MSFE of different forecast methods

Note: Reported are the MSFEs of the respective forecast relative to that based on the postbreak forecast using the Andrews test at $\alpha = 0.05$. The second columns, "f-supW", gives the relative MSFE of post-break sample forecasts based on our break-point test. "Optimal weights" denotes forecasts from applying optimal weights to observations, where "all" indicates that the optimal weights were used for all forecasts irrespective of significance of the break point test, "supW" indicates that optimal weights were used when the Andrews test was significant, and "f-supW" when our test was significant. The right side of the table excludes observations where both tests were insignificant, in which case the forecasts are identically based on the full sample. The acronyms in the first column: OI: output and income (17 series), LM: labor market (32), CO: consumption and orders (10), OrdInv: orders and inventories (14), MC: money and credit (14), IRER: interest rates and exchange rates (22), P: prices (21), S: stock market (4). break according to both tests. Using the f-sup W test in place of the sup W test leads to a 6% improvement in a accuracy on average for the AR(1) and a 18% improvement in accuracy on average for the AR(6) model. These large improvements are found for series in all categories with the exception of the money and credit with the AR(1). Even in this category, the improvement from using the f-sup W for the AR(6) is 16%. The results for the optimal weights forecasts are similar in that the forecasts using the f-sup W are substantially more precise than those using the sup W.

9 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 a 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 argue that under the break sizes we consider under our null hypothesis, this optimality might be 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 also improves over pretesting using the standard test of Andrews (1993) in terms of MSFE. Further improvements can be made by considering an optimal weights or shrinkage estimator under the alternative. Applying the optimal weights/shrinkage estimator without pretesting is argued not to be a fruitful strategy as it suffers too much from imprecise break date estimation in the case of small breaks.

A Derivation of (35)

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

$$\Delta_{1} = T \mathbf{E} \left[\left(\partial_{\beta_{2}} f'(\hat{\boldsymbol{\beta}}_{2}(\hat{\tau}) - \boldsymbol{\beta}_{2}) + \partial_{\delta} f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right)^{2} \right]$$

$$= T \mathbf{E} \left[\left(\partial_{\beta_{2}} f'(\hat{\boldsymbol{\beta}}_{2}(\hat{\tau}) - \boldsymbol{\beta}_{2}) \right)^{2} + \left(\partial_{\delta} f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right)^{2} + 2 \partial_{\beta_{2}} f'(\hat{\boldsymbol{\beta}}_{2}(\hat{\tau}) - \boldsymbol{\beta}_{2}) \partial_{\delta} f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right]$$

(55)

and similarly for Δ_2

$$\Delta_{2} = T \mathbf{E} \left[\left(\partial_{\beta_{2}} f'(\hat{\boldsymbol{\beta}}_{F}(\hat{\tau}) - \boldsymbol{\beta}_{2}) + \partial_{\delta} f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right)^{2} \right]$$

$$= T \mathbf{E} \left[\left(\partial_{\beta_{2}} f'(\hat{\boldsymbol{\beta}}_{F} - \boldsymbol{\beta}_{2}) \right)^{2} + \left(\partial_{\delta} f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right)^{2} + 2 \partial_{\beta_{2}} f'(\hat{\boldsymbol{\beta}}_{F} - \boldsymbol{\beta}_{2}) \partial_{\delta} f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \right]$$

(56)

In addition, we define

$$a = \frac{1}{1 - \hat{\tau}} \left[\partial_{\beta_2} f'(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \bar{\boldsymbol{X}}'(\boldsymbol{B}(1) - \boldsymbol{B}(\hat{\tau})) + \int_{\hat{\tau}}^{1} \partial_{\beta_2} f'\boldsymbol{\eta}(s) ds \right]$$

$$b = \partial_{\beta_2} f'(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \bar{\boldsymbol{X}}' \bar{\boldsymbol{Z}} (\bar{\boldsymbol{Z}}' \boldsymbol{M}_{\bar{\boldsymbol{X}}} \bar{\boldsymbol{Z}})^{-1} \bar{\boldsymbol{Z}}' \boldsymbol{M}_{\boldsymbol{X}} \boldsymbol{B}(1)$$

$$c = \partial_{\delta} f'(\bar{\boldsymbol{Z}}' \boldsymbol{M}_{\boldsymbol{X}} \bar{\boldsymbol{Z}})^{-1} \bar{\boldsymbol{Z}}' \boldsymbol{M}_{\bar{\boldsymbol{X}}} \boldsymbol{B}(1)$$

$$d = \partial_{\beta_2} f'(\bar{\boldsymbol{X}}' \bar{\boldsymbol{X}})^{-1} \bar{\boldsymbol{X}}' \boldsymbol{B}(1) + \int_{\beta_2}^{1} \partial_{\beta_2} f' \boldsymbol{\eta}(s) ds$$
(57)

Schematically, we have from (21) and (24) that $\sqrt{T}\partial_{\beta_2}f'(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}) = a - b$, $\sqrt{T}\partial_{\beta_2}f'(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2) = d - b$ and $\sqrt{T}\partial_{\delta}f'(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) = c$ Then

$$\Delta_1 = \mathbf{E}[a^2 + b^2 - 2ab + c^2 + 2ca - 2cb]$$

$$\Delta_2 = \mathbf{E}[d^2 + b^2 - 2db + c^2 + 2cd - 2cb]$$
(58)

Now E[db] = E[cd] = 0 by the fact that $E[\boldsymbol{B}(1)] = 0$, $E[\boldsymbol{B}(1)\boldsymbol{B}(1)'] = \boldsymbol{I}$ and $\boldsymbol{M}_{\bar{\boldsymbol{X}}}\bar{\boldsymbol{X}} = \boldsymbol{O}$. Furthermore E[ab] = E[ac] = 0 since $E\left[\frac{1}{1-\hat{\tau}}(\boldsymbol{B}(1) - \boldsymbol{B}(\hat{\tau}))\boldsymbol{B}(1)'\right] = \text{cst} \cdot \boldsymbol{I}$ by the fact that the elements of $\boldsymbol{B}(\cdot)$ are identically and independently distributed. Then

$$\Delta_1 - \Delta_2 = \mathbf{E}[a^2 - d^2] \tag{59}$$

B Verifying condition (51)

Figure 9: Dependence of the critical values on the break date



Note: equation (51) states that for weak optimality $\partial u(\tau_b)/\partial \tau_b < 1/\tau_b(1-\tau_b)$. The green line depicts the derivative of the critical values depicted in Figure 2 for $\alpha = 0.05$ as a function of the break date τ_b . The blue line depicting the upper bound is clearly sufficiently high such that (51) indeed is satisfied.

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. (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. and Ng, S. (2014). FRED-MD: A monthly database for macroeconomic research. *December, available at http://www.columbia.edu/sn2294/papers/freddata.pdf.*
- 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.
- 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.
- Wallace, T. D. (1972). Weaker criteria and tests for linear restrictions in regression. *Econometrica*, 40(4):689–698.