Online appendix to
“Does modeling a structural break improve forecast accuracy?”

Tom Boot[*] Andreas Pick[†]

July 19, 2019

Online appendix

A Distribution of the test statistic $W(\tau_b)$ in Section 2

In order to show that $W(\tau_b) \sim \chi(1,\zeta)$ as stated in (5), it is sufficient to show that

$$\frac{\sqrt{T}\tau_b \boldsymbol{x}'_{T+1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)}{\sqrt{\boldsymbol{x}'_{T+1}\left(\frac{\boldsymbol{V}_2}{1-\tau_b} - \boldsymbol{V}_F\right)\boldsymbol{x}_{T+1}}} \sim N(\sqrt{\zeta}, 1) \tag{A1}$$

Normality follows from the assumptions for model (1), and the mean from the fact that $\mathrm{E}[\hat{\boldsymbol{\beta}}_i] = \boldsymbol{\beta}_i$ for $i = 1, 2$. For the variance, note that $\boldsymbol{V}_F^{-1} = \tau_b\boldsymbol{V}_1^{-1} + (1-\tau_b)\boldsymbol{V}_2^{-1}$. We can rewrite the denominator of the test statistic using the Woodbury matrix identity

$$\begin{aligned}
\frac{1}{1-\tau_b}\boldsymbol{V}_2 - \boldsymbol{V}_F &= \frac{1}{1-\tau_b}\boldsymbol{V}_2^{1/2}\left[\boldsymbol{I} - \left(\frac{\tau_b}{1-\tau_b}\boldsymbol{V}_2^{1/2}\boldsymbol{V}_1^{-1}\boldsymbol{V}_2^{1/2} + \boldsymbol{I}\right)^{-1}\right]\boldsymbol{V}_2^{1/2} \\
&= \frac{\tau_b}{(1-\tau_b)^2}\boldsymbol{V}_2\left(\boldsymbol{V}_1 + \frac{\tau_b}{1-\tau_b}\boldsymbol{V}_2\right)^{-1}\boldsymbol{V}_2 \\
&= \frac{\tau_b}{(1-\tau_b)^2}\boldsymbol{V}_2\boldsymbol{A}^{-1}\boldsymbol{V}_2
\end{aligned} \tag{A2}$$

where $\boldsymbol{A} = \boldsymbol{V}_1 + \frac{\tau_b}{1-\tau_b}\boldsymbol{V}_2$.

The variance of the numerator of (A1) satisfies

$$T\tau_b^2\mathrm{Var}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2) = \tau_b\left(\boldsymbol{V}_1 + \frac{\tau_b}{1-\tau_b}\boldsymbol{V}_2\right) = \tau_b\boldsymbol{A}$$

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

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

Furthermore,

$$\boldsymbol{V}_2^{-1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1} = \left[\boldsymbol{V}_1\boldsymbol{V}_F^{-1}\boldsymbol{V}_2\right]^{-1} = \frac{1}{1-\tau_b}\left[\boldsymbol{V}_1 + \frac{\tau_b}{1-\tau_b}\boldsymbol{V}_2\right]^{-1} = \frac{1}{1-\tau_b}\boldsymbol{A}^{-1}$$

Then,

$$\begin{aligned} T\tau_b^2\boldsymbol{V}_F\boldsymbol{V}_1^{-1}\mathrm{Var}(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2)\boldsymbol{V}_1^{-1}\boldsymbol{V}_F &= \boldsymbol{V}_2\left(\boldsymbol{V}_2^{-1}\boldsymbol{V}_F\boldsymbol{V}_1^{-1}\right)\left(T\tau_b^2\mathrm{Var}(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2)\right)\left(\boldsymbol{V}_1^{-1}\boldsymbol{V}_F\boldsymbol{V}_2^{-1}\right)\boldsymbol{V}_2 \\ &= \boldsymbol{V}_2\frac{1}{1-\tau_b}\boldsymbol{A}^{-1}\tau_b\boldsymbol{A}\frac{1}{1-\tau_b}\boldsymbol{A}^{-1} \\ &= \frac{\tau_b}{(1-\tau_b)^2}\boldsymbol{V}_2\boldsymbol{A}^{-1}\boldsymbol{V}_2 \end{aligned}$$

which together with (A2) shows that variance of (A1) is equal to 1.

B Convergence of partial-sample and full-sample estimators in (11) and (12) of Section 3

Partial-sample estimators The results in Andrews (1993) show that

$$\sqrt{T}\begin{pmatrix}\hat{\boldsymbol{\beta}}_1(\tau)-\boldsymbol{\beta}_2\\ \hat{\boldsymbol{\beta}}_2(\tau)-\boldsymbol{\beta}_2\\ \hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\end{pmatrix} \Rightarrow \begin{bmatrix}\tau\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} & \boldsymbol{0} & \tau\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}\\ \boldsymbol{0} & (1-\tau)\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} & (1-\tau)\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}\\ \tau\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} & (1-\tau)\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} & \bar{\boldsymbol{Z}}'\bar{\boldsymbol{Z}}\end{bmatrix}^{-1} \times\begin{bmatrix}\bar{\boldsymbol{X}}'\boldsymbol{B}(\tau)+\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}\int_0^\tau\boldsymbol{\eta}(s)ds\\ \bar{\boldsymbol{X}}'[\boldsymbol{B}(1)-\boldsymbol{B}(\tau)]+\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}\int_\tau^1\boldsymbol{\eta}(s)ds\\ \bar{\boldsymbol{Z}}'\boldsymbol{B}(1)+\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}\int_0^1\boldsymbol{\eta}(s)ds\end{bmatrix} \tag{A3}$$

Define the projection matrix $\boldsymbol{P}_{\bar{X}} = \bar{\boldsymbol{X}}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'$, its orthogonal complement as $\boldsymbol{M}_{\bar{X}} = \boldsymbol{I}-\boldsymbol{P}_{\bar{X}}$, $\boldsymbol{V} = (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}$, $\boldsymbol{H} = \bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}}$, $\boldsymbol{L} = (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}(\bar{\boldsymbol{Z}}'\boldsymbol{M}_{\bar{X}}\bar{\boldsymbol{Z}})^{-1}$, and $\tilde{\boldsymbol{H}} = \boldsymbol{L}\boldsymbol{H}\boldsymbol{L}'$. The inverse in (A3) yields the asymptotic variance covariance matrix of $\left(\hat{\boldsymbol{\beta}}_1(\tau)', \hat{\boldsymbol{\beta}}_2(\tau)', \hat{\boldsymbol{\delta}}'\right)'$

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

from which (11) follows.

Full-sample estimator The full-sample estimator weakly converges as

$$\sqrt{T}\begin{pmatrix}\hat{\boldsymbol{\beta}}_F-\boldsymbol{\beta}_2\\ \hat{\boldsymbol{\delta}}-\boldsymbol{\delta}\end{pmatrix} \Rightarrow \begin{bmatrix}\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} & \bar{\boldsymbol{X}}'\bar{\boldsymbol{Z}}\\ \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}} & \bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}\end{bmatrix}^{-1}\begin{bmatrix}\bar{\boldsymbol{X}}'\boldsymbol{B}(1)+\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}\int_0^1\boldsymbol{\eta}(s)ds\\ \bar{\boldsymbol{Z}}'\boldsymbol{B}(1)+\bar{\boldsymbol{Z}}'\bar{\boldsymbol{X}}\int_0^1\boldsymbol{\eta}(s)ds\end{bmatrix} \tag{A4}$$

Using the notation defined above, the inverse in (A4) is

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

which leads to (12).

C A break of known timing

Forecasts are obtained using (9) in the paper

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

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

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

$$\sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \overset{a}{\sim} N \left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \frac{1}{\tau}\boldsymbol{V} + \tilde{\boldsymbol{H}} & \tilde{\boldsymbol{H}} & -\boldsymbol{L} \\ \tilde{\boldsymbol{H}} & \frac{1}{1-\tau}\boldsymbol{V} + \tilde{\boldsymbol{H}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & -\boldsymbol{L}' & \boldsymbol{H}^{-1} \end{pmatrix} \right] \tag{A5}$$

For the full sample estimator we have

$$\sqrt{T} \begin{pmatrix} \hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2 \\ \hat{\boldsymbol{\delta}} - \boldsymbol{\delta} \end{pmatrix} \overset{a}{\sim} N \left[\begin{pmatrix} \tau_b(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2) \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{V} + \tilde{\boldsymbol{H}} & -\boldsymbol{L} \\ -\boldsymbol{L}' & \boldsymbol{H}^{-1} \end{pmatrix} \right] \tag{A6}$$

and

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

Define $\boldsymbol{f}_{\beta_2} = \frac{\partial f_{T+h}(\boldsymbol{\beta}_2, \boldsymbol{\delta}; \mathcal{I}_T)}{\partial \boldsymbol{\beta}_2}$ and $\boldsymbol{f}_{\delta} = \frac{\partial f_{T+h}(\boldsymbol{\beta}_2, \boldsymbol{\delta}; \mathcal{I}_T)}{\partial \boldsymbol{\delta}}$. Using a first order Taylor expansion, (A5) and (A6), we have that

$$\begin{aligned} \sqrt{T}\left(f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) - f_{T+h}(\boldsymbol{\beta}_2, \boldsymbol{\delta}; \mathcal{I}_T)\right) &= \sqrt{T}\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2) + \boldsymbol{f}'_{\delta}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + O(T^{-1})\right] \\ &\overset{a}{\sim} N\left(0, \Sigma_{\beta_2} + \Sigma_r\right) \\ \sqrt{T}\left(f_{T+h}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) - f_{T+h}(\boldsymbol{\beta}_2), \boldsymbol{\delta}; \mathcal{I}_T)\right) &= \sqrt{T}\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2) + \boldsymbol{f}'_{\delta}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + O(T^{-1})\right] \\ &\overset{a}{\sim} N\left(\tau_b \boldsymbol{f}'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2), \Sigma_{\beta_F} + \Sigma_r\right) \end{aligned}$$

where

$$\begin{aligned} \Sigma_{\beta_i} &= \operatorname*{plim}_{T\to\infty} T \boldsymbol{f}'_{\beta_2} \mathrm{Var}(\hat{\boldsymbol{\beta}}_i) \boldsymbol{f}_{\beta_2}, \quad \text{for } i = 2, F \\ \Sigma_r &= \operatorname*{plim}_{T\to\infty} T \left(\boldsymbol{f}'_{\delta} \mathrm{Var}(\hat{\boldsymbol{\delta}}) \boldsymbol{f}_{\delta} + 2\boldsymbol{f}'_{\beta_2} \mathrm{Cov}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}) \boldsymbol{f}_{\delta} \right) \end{aligned} \tag{A7}$$

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

$$\Sigma_{\beta_2} = \frac{1}{1-\tau_b}\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} + \boldsymbol{f}'_{\beta_2}\tilde{\boldsymbol{H}}\boldsymbol{f}_{\beta_2}, \qquad \Sigma_{\beta_F} = \boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} + \boldsymbol{f}'_{\beta_2}\tilde{\boldsymbol{H}}\boldsymbol{f}_{\beta_2}$$

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

$$\begin{aligned}
\lim_{T\to\infty} T\text{E}\left[\left(f_{T+h}(\hat{\boldsymbol{\beta}}_2, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) - f_{T+h}(\boldsymbol{\beta}_2, \boldsymbol{\delta}; \mathcal{I}_T)\right)^2\right] &= \frac{1}{1-\tau_b}\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} + \boldsymbol{f}'_{\beta_2}\tilde{\boldsymbol{H}}\boldsymbol{f}_{\beta_2} + \Sigma_r \\
\lim_{T\to\infty} T\text{E}\left[\left(f_{T+h}(\hat{\boldsymbol{\beta}}_F, \hat{\boldsymbol{\delta}}; \mathcal{I}_T) - f_{T+h}(\boldsymbol{\beta}_2, \boldsymbol{\delta}; \mathcal{I}_T)\right)^2\right] &= \left[\tau_b \boldsymbol{f}'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2 \\
&\quad + \boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} + \boldsymbol{f}'_{\beta_2}\tilde{\boldsymbol{H}}\boldsymbol{f}_{\beta_2} + \Sigma_r
\end{aligned}$$

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

$$\zeta = T(1-\tau_b)\tau_b \frac{\left[\boldsymbol{f}'_{\beta_2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}} \leq 1 \tag{A8}$$

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

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

$$W(\tau_b) = T(1-\tau_b)\tau_b \frac{\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right]^2}{\hat{\omega}} \overset{a}{\sim} \chi^2(1, \zeta) \tag{A9}$$

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

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

$$\zeta = T\tau_b(1-\tau_b)\frac{\left[\boldsymbol{x}'_{T+1}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2)\right]^2}{\boldsymbol{x}'_{T+1}\boldsymbol{V}\boldsymbol{x}_{T+1}} \leq 1 \tag{A10}$$

which asymptotically coincides with (4) in the paper.

D Proof of Lemma 1 in the paper

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

$$\begin{aligned}\Delta_1 &= \lim_{T\to\infty} T\mathrm{E}\left[\left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2) + \boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2\right] / \boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} \\ &= \lim_{T\to\infty} T\mathrm{E}\left[\left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2)\right)^2 + \left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2 + \right. \\ &\qquad \left. +2\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_2(\hat{\tau}) - \boldsymbol{\beta}_2)\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right] / \boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}\end{aligned} \tag{A11}$$

and similarly for Δ_2

$$\begin{aligned}\Delta_2 &= \lim_{T\to\infty} T\mathrm{E}\left[\left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2) + \boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2\right] / \boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2} \\ &= \lim_{T\to\infty} T\mathrm{E}\left[\left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2)\right)^2 + \left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right)^2 + \right. \\ &\qquad \left. +2\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_F - \boldsymbol{\beta}_2)\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})\right] / \boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}\end{aligned} \tag{A12}$$

To prove the lemma, we need that

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

Define

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

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

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

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

However, we need to prove

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

Denote

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

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

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

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

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

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

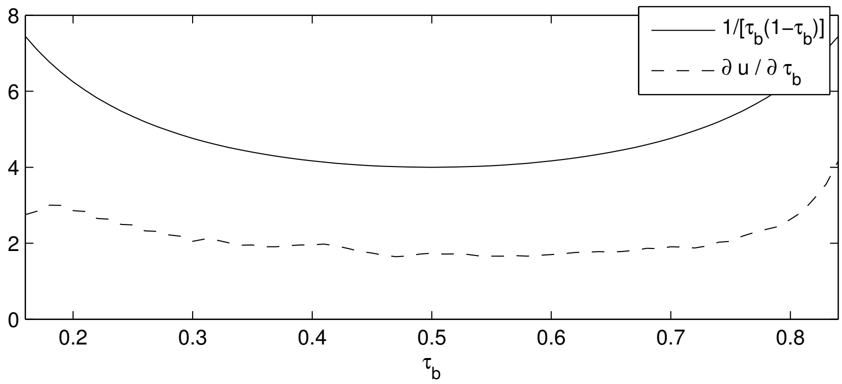
E Verifying Assumption 4 in the paper

In the structural break model, Assumption 4 is satisfied when $|\partial u(\tau_b)/\partial\tau_b| < \theta_{\tau_b}[\tau_b(1-\tau_b)]^{-1/2}$. The break size $\zeta^{1/2}$ depicted in Figure 1 in the paper satisfies $\zeta^{1/2} = \theta_{\tau_b}\sqrt{\tau_b(1-\tau_b)} > 1$. Therefore, a sufficient condition for the slowly varying assumption is

$$\left|\frac{\partial u(\tau_b)}{\partial\tau_b}\right| < \frac{1}{\tau_b(1-\tau_b)} \tag{A13}$$

Observe that, in Figure A, 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 A: Dependence of the critical values on the break date

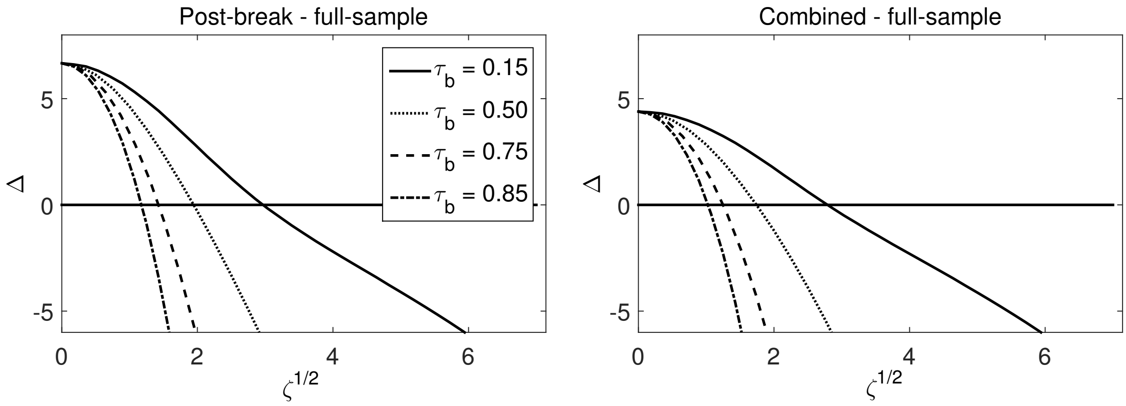


Note: The dashed line depicts the derivative of the critical values for $\alpha = 0.05$ as a function of the break date τ_b. The solid line depicting the upper bound $[\tau_b(1-\tau_b)]^{-1}$.

F Uniqueness of the break magnitude that yields equal forecast accuracy

In order to ensure the uniqueness of the break magnitude that leads to equal forecast accuracy, we evaluate Δ in (20) in the paper and Δ_c in (31) in the paper numerically using the simulation set-up described in Section 5. The results in Figure B show that the value of $|\theta_{\tau_b}|$ that leads to equal forecast accuracy is unique.

Figure B: Difference in asymptotic MSFEs, Δ and Δ_c



Note: The left panel shows the difference in the asymptotic MSFE between the post-break forecast and the full-sample forecast as a function of the standardized break magnitude $\zeta^{1/2}$ in (20) for $\tau_b = \{0.15, 0.50, 0.75, 0.85\}$. The right panel shows the difference in MSFE between the combined forecast and the full-sample forecast in (31) in the paper.

G Derivation of equation (27) in the paper

From a Taylor series expansion it follows that

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

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

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

$$\begin{aligned}\lim_{T\to\infty} \mathrm{E}\left[T\left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right)^2\right] &= \lim_{T\to\infty} \mathrm{E}\left[T\left(\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right)\right]^2 + \lim_{T\to\infty} T\mathrm{Var}\left[\boldsymbol{f}'_{\beta_2}(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_2)\right] \\ &= \left(\boldsymbol{f}'_{\beta_2}\boldsymbol{b}\right)^2 + \boldsymbol{f}'_{\beta_2}\left(\frac{1}{\tau_b} + \frac{1}{1-\tau_b}\right)\boldsymbol{V}\boldsymbol{f}_{\beta_2}\end{aligned}$$

since $\lim_{T\to\infty} T\mathrm{Cov}(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = 0$.

The term linear in ω is given by

$$\lim_{T\to\infty} \boldsymbol{f}'_{\beta_2}\mathrm{E}\left[T\left(\hat{\boldsymbol{\beta}}_1-\hat{\boldsymbol{\beta}}_2\right)\left(\hat{\boldsymbol{\beta}}_2-\boldsymbol{\beta}_2\right)\right]\boldsymbol{f}_{\beta_2} = -\lim_{T\to\infty} \boldsymbol{f}'_{\beta_2}\mathrm{E}\left[T\left(\boldsymbol{\beta}_1-\boldsymbol{\beta}_2\right)\boldsymbol{\beta}'_2\right]\boldsymbol{f}_{\beta_2} + \lim_{T\to\infty} \boldsymbol{f}'_{\beta_2}\mathrm{E}\left[T\hat{\boldsymbol{\beta}}_1\hat{\boldsymbol{\beta}}'_2-\hat{\boldsymbol{\beta}}_2\hat{\boldsymbol{\beta}}'_2\right]\boldsymbol{f}_{\beta_2} = -\frac{1}{1-\tau_b}\boldsymbol{f}'_{\beta_2}\boldsymbol{V}\boldsymbol{f}_{\beta_2}$$

Using these two expressions yields (27) in the paper.

H Application results with rolling windows of 240 observations

This section reports the result of our application but when the estimation window contains 240 instead of 120 observations. Table A gives the fractions of the time series and estimation windows where a significant structural break is found. Compared to Table 5 marginally fewer breaks are found but the pattern across the different tests is the same.

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

	supW	W	S	W^c	S^c
AR(1)	0.183	0.087	0.096	0.103	0.112
AR(6)	0.086	0.028	0.036	0.035	0.045

Note: This table is the analogue of Table 5 with estimates based on a rolling window size of $T = 240$.

Table B shows the MSFE of the post-break forecasts of our tests relative to the test of Andrews. It can be seen that the forecasts from our test are generally more precise even if compared to the results in Table 5 the improvements are slightly smaller.

I Combination of tests with optimal window selection

Table C reports the results of forecasting with the optimal window of Inoue et al. (2017) either based on the test of Andrews (which in the case of one break is equivalent to the test of Bai and Perron (1998)) and on our test. The approach of Inoue et al. (2017) has two steps: first, using one of the two tests it is established whether an estimation sample has a structural break. Second, once a structural break has been found, forecasts are made using the optimal estimation window.

The results in Table C are the MSFE of each forecast method relative that of the post break sample based forecast using Andrews' test. It can be seen that uniformly using our Wald-type test leads to further improvements of the optimal windows suggested by Inoue et al. (2017).

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

		Post-break		Combination		
		W	S	W	S	$supW$
AR(1)	All series	0.986	0.987	0.981	0.980	0.990
	OI	0.997	0.994	0.990	0.987	0.991
	LM	0.989	0.994	0.984	0.982	0.988
	CO	0.974	0.971	0.968	0.968	0.991
	OrdInv	1.003	0.999	0.995	0.997	0.992
	MC	1.033	1.035	1.019	1.037	0.997
	IRER	0.954	0.954	0.954	0.951	0.988
	P	0.879	0.976	0.897	1.059	0.986
	S	0.960	0.989	0.972	0.992	0.995
AR(6)	All series	0.972	0.981	0.969	0.987	0.991
	OI	0.992	1.013	0.982	1.000	0.993
	LM	0.980	0.995	0.976	1.006	0.993
	CO	0.991	0.986	0.986	0.985	0.994
	OrdInv	0.982	1.014	0.988	1.019	0.993
	MC	0.981	1.052	1.005	1.084	0.993
	IRER	0.930	0.922	0.924	0.938	0.985
	P	0.942	0.983	0.959	0.972	0.989
	S	1.008	1.019	1.005	1.008	0.993

Note: This table is the analogue of Table 6 with a rolling window size of $T = 240$.

References

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

Inoue, A., Jin, L., and Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, 196(1):55–67.

Table C: MSFE of optimal estimation window of Inoue et al. (2017)

		OptR1-A	OptR2-A	OptR3-A	OptR1-W	OptR1-W	OptR1-W
AR(1)	All series	0.916	0.916	0.916	0.898	0.898	0.898
	OI	0.929	0.929	0.929	0.924	0.924	0.924
	LM	0.896	0.895	0.895	0.883	0.882	0.882
	CO	0.956	0.956	0.956	0.951	0.951	0.951
	OrdInv	0.918	0.918	0.918	0.901	0.901	0.901
	MC	0.998	0.997	0.997	0.972	0.971	0.971
	IRER	0.887	0.887	0.887	0.833	0.833	0.833
	P	0.939	0.939	0.939	0.939	0.939	0.939
	S	0.944	0.946	0.946	0.902	0.902	0.902
AR(6)	All series	0.918	0.918	0.918	0.890	0.890	0.890
	OI	0.927	0.927	0.927	0.899	0.899	0.899
	LM	0.894	0.893	0.893	0.879	0.879	0.879
	CO	0.955	0.955	0.955	0.937	0.937	0.937
	OrdInv	0.909	0.909	0.909	0.883	0.883	0.883
	MC	0.977	0.977	0.977	0.946	0.946	0.946
	IRER	0.890	0.891	0.891	0.832	0.832	0.832
	P	0.941	0.941	0.941	0.899	0.899	0.899
	S	0.963	0.963	0.963	0.955	0.955	0.955

Note: OptR1-A, OptR2-A, and OptR3-A denote the forecasts from optimal windows of Inoue et al. (2017) based on Andrews' test and OptR1-W, OptR2-W, and OptR3-W the equivalent forecasts based on our Wald-type test. All are MSFEs as a ratio of the post-break sample based forecasts using the break date from Andrews' test.

Table D: 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 (15) in the paper																			
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 (26) in the paper																			
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 ζ 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 in the paper.

Table E: Forecast combination 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 (15) in the paper																			
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 (26) in the paper																			
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 ζ and S test statistics that test the null hypothesis of equal MSFE of the forecast combination (30) in the paper and the full sample forecast. For additional information, see the footnote of Table 2 in the paper.