Department of Economics, Hitotsubashi University

Discussion Paper Series #2015-01

Confidence Sets for the Break Date Based on Optimal Tests

Eiji Kurozumi and Yohei Yamamoto

January 2015

Confidence Sets for the Break Date Based on Optimal Tests

Eiji Kurozumi[1] Yohei Yamamoto

Department of Economics, Hitotsubashi University

January 22, 2015

Abstract

This study proposes constructing a confidence set for the date of a one-time structural change using a point optimal test. Following Elliott and Müller (2007), we first construct a test for the break date that maximizes the weighted average of the power function. The confidence set is then obtained by inverting the test statistic. We carefully choose the weights and show by Monte Carlo simulations that the confidence set based on our method has a relatively accurate coverage rate, while the length of our confidence set is significantly shorter than the lengths proposed in the literature.

JEL classification: C12, C22

Keywords: coverage rate, break fraction, hypothesis test, average power

[1]This research was supported by JSPS KAKENHI Grant Numbers 18730142, 23243038 and 25285067. Address correspondence to Eiji Kurozumi, Department of Economics, Hitotsubashi University, 2-1 Naka, Kunitachi, Tokyo 186-8601, Japan; e-mail: kurozumi@stat.hit-u.ac.jp

1. Introduction

This study proposes constructing a confidence set for the break date in linear time series models with a one-time structural change. A frequently used method to construct a confidence set for some parameter is use of the asymptotic distribution of the estimator of that parameter. In the case of structural change, it is known that the distribution of the break point estimator depends on the true process of the error term even asymptotically when the magnitude of the break is fixed, and hence, it is tedious to construct a confidence set in this case. To overcome this problem, Bai (1994) assumed that the magnitude of the break shrinks to zero at a rate slower than $T^{-1/2}$ and derived the limiting distribution of the break point estimator in models with a mean shift. Based on the shrinking shift technique, Bai (1997) further investigated the break point estimator in linear regression models with a one-time break. In addition, multiple breaks were considered by Bai and Perron (1998) for linear regressions, by Bai (2000) for vector autoregressions, by Qu and Perron (2007) for general linear multivariate models, by Boldea Hall and Han (2012) and Perron and Yamamoto (2014) for models with endogenous regressors, and by Boldea and Hall (2013) for nonlinear regressions.

Although the limiting distributions derived in these studies can be used to construct the confidence intervals for the break date(s), Elliott and Müller (2007) showed that the coverage rate based on the break point estimator tends to be too liberal for the small size of the break. That is, the empirical coverage rate is much smaller than the nominal one. To overcome this problem, Elliott and Müller (2007) proposed constructing the confidence set by inverting a test for the break point. For example, we first test the null hypothesis that the break date is given by, say, T_1, and if the null hypothesis is accepted, then T_1 is included in the confidence set; otherwise it is excluded. Elliott and Müller (2007) showed by simulations that the coverage rate of their confidence set is close to the nominal one for various data generating processes.

Although the correct coverage rate is a desirable property of their method, the average length of their confidence set becomes too large in some cases, as pointed out by Chang and Perron (2013), so that no useful information about the break date is included in the confidence set. The main reason for the large confidence set is that long-run variance is estimated with

the break date under the null hypothesis. In this case, the usual kernel-based estimation of the long-run variance works well if the maintained break date is the same as the true one. However, if the maintained break date is different from the true one, then the estimation residuals either before or after the maintained break date include a one-time break. As a result, the long-run variance estimator tends to be larger than the true one and the test statistic cannot reject the null hypothesis, so that such a maintained break date tends to be included in the confidence set. To circumvent this problem, Yamamoto (2014) proposed modifying the estimator of the long-run variance such that the break date is first estimated by minimizing the sum of squared residuals and then the long-run variance is calculated using this estimated break date. Yamamoto (2014) showed that his method leads to an empirical size close to the nominal size, while the confidence set becomes smaller than that based on Elliott and Müller (2007).

In this study, we further extend the methods of Elliott and Müller (2007) and Yamamoto (2014) and construct the confidence set based on an optimal test. Our goal is to construct a confidence set that is as small as possible with a good coverage rate. We first derive a test that maximizes the weighted average power as considered by Elliott and Müller (2007), and then we choose the weighting function carefully. The confidence set can be obtained by inverting the optimal test. We show that our confidence set has a desirable theoretical property while in finite samples, it is smaller than those based on the existing literature, including Bai's (1997) method.

The rest of this paper is organized as follows. Section 2 introduces our model and assumptions and we derive an optimal test under restricted assumptions. We propose several versions of an optimal test and investigate their asymptotic properties. In Section 3, we relax the assumptions and explain how to construct the confidence set in practical analysis. The size of the confidence set is discussed in Section 4 using the limit function of the optimal test under the alternative. The finite sample property is investigated in Section 5 while Section 6 concludes the paper.

2. Optimal Test for the Break Date

2.1. Model and assumptions

We consider the following linear regression model with a one-time structural change:

$$y_t = x_t'\beta + x_t^{\lambda_0\prime}\delta + z_t'\gamma + u_t, \quad (t = 1, 2, \cdots, T) \tag{1}$$

where x_t and z_t are k and p dimensional regressors, respectively, $x_t^{\lambda_0} \equiv 1(t > [\lambda_0 T])x_t$ with $1(\cdot)$ being an indicator function, λ_0 is a true break fraction, $[a]$ denotes the largest integer less than a, u_t is an error term, and β, δ, and γ are k, k, and p dimensional unknown coefficients, respectively. We denote the true break date as $T_0 \equiv [\lambda_0 T]$. In this model, the coefficient of x_t changes from β to $\beta + \delta$ at $t = T_0 + 1$ whereas the coefficient associated with z_t, γ, is stable. In vector form, we denote (1) as

$$y = X\beta + X^{\lambda_0}\delta + Z\gamma + u. \tag{2}$$

We construct a confidence set for the break date T_0 by inverting a test for the break date. In this case, we expect that the confidence set based on an efficient test should be smaller than that based on an inefficient test, and thus, at first, we derive an optimal test following Elliott and Müller (2007). To derive an optimal test, we make the following assumption:

Assumption 1 *(i)* $\{u_t\} \sim i.i.d.N(0, \sigma^2)$.
(ii) The regressors $\{x_t\}$ *and* $\{z_t\}$ *are independent of* $\{u_t\}$.
(iii) The following weak law of large numbers (WLLN) and the functional central limit theorem (FCLT) hold for $q_t = [x_t', z_t']'$*:*

$$\frac{1}{T}\sum_{t=1}^{[rT]} q_t q_t' \xrightarrow{p} r\Sigma_1 \quad \textit{uniformly in } 0 \le r \le \lambda_0, \tag{3}$$

$$\frac{1}{T}\sum_{t=T_0+1}^{[rT]} q_t q_t' \xrightarrow{p} (r - \lambda_0)\Sigma_2 \quad \textit{uniformly in } \lambda_0 < r \le 1, \tag{4}$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[rT]} x_t u_t \Rightarrow \sigma\Sigma_{xx,1}^{1/2} B(r) \quad \textit{for} \quad 0 \le r \le \lambda_0, \tag{5}$$

$$\frac{1}{\sqrt{T}}\sum_{t=T_0+1}^{[rT]} x_t u_t \Rightarrow \sigma\Sigma_{xx,2}^{1/2} \left(B(r) - B(\lambda_0)\right) \quad \textit{for} \quad \lambda_0 < r \le 1, \tag{6}$$

where Σ_1 *and* Σ_2 *are symmetric and positive definite matrices partitioned as* $\Sigma_{ij,1}$ *and* $\Sigma_{ij,2}$ *for* $i, j = x$ *or* z *conformably with* q_t*, respectively,* $B(\cdot)$ *is a* k *dimensional standard Brownian motion, and* $\xrightarrow{p}$ *and* $\Rightarrow$ *denote convergence in probability and weak convergence of the associated probability measures, respectively.*

(iv) $\sup_{0\le r\le 1} \frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} z_t u_t = O_p(1)$.

(v) $0 < \underline{\lambda} \le \lambda_0 \le \overline{\lambda} < 1$.

Assumptions 1(i) and (ii) are supposed to derive an optimal test, but are not required to derive the limiting distribution. Assumptions 1(iii)–(v) are standard in regression models with a one-time break, except that the variance in (5) and (6) is relatively simple. Note that a change in variance of the regressors is allowed by (3) and (4). Assumptions 1(i)–(iii) are restrictive but are made in order to derive an optimal test and its limiting distribution; later, we relax them for practical analysis.

2.2. General form of the optimal test

Suppose that our interest is whether λ_1 is the true break fraction, and consider the following simple testing problem:

$$H_N : \lambda_0 = \lambda_1 \quad \text{vs.} \quad H_A : \lambda_0 = \lambda_2. \tag{7}$$

The corresponding dates are denoted as $T_1 \equiv [\lambda_1 T]$ and $T_2 \equiv [\lambda_2 T]$, respectively.

Following Elliott and Müller (2007), we consider a class of location invariant tests. First, we express model (1) as

$$\begin{aligned} y_t &= x_t'\beta + x_t^{\lambda_0\prime}\delta + z_t'\gamma + u_t \\ &= x_t'\beta + x_t^{\lambda_1\prime}\delta + z_t'\gamma + u_t^{\lambda_{01}} \\ &= w_t^{\lambda_1\prime}\theta + u_t^{\lambda_{01}}, \end{aligned} \tag{8}$$

where $u_t^{\lambda_{01}} = u_t + r_t^{\lambda_{01}\prime}\delta$ with $r_t^{\lambda_{01}} = x_t^{\lambda_0} - x_t^{\lambda_1}$, $w_t^{\lambda_1} = [x_t', x_t^{\lambda_1\prime}, z_t']'$ and $\theta = [\beta', \delta', \gamma']'$. In vector form,

$$y = X\beta + X^{\lambda_1}\delta + Z\gamma + u^{\lambda_{01}} \tag{9}$$

$$= W^{\lambda_1}\theta + u^{\lambda_{01}}, \tag{10}$$

where $W^{\lambda_1} = [X, X^{\lambda_1}, Z]$ and $u^{\lambda_{01}} = u + R^{\lambda_{01}}\delta$ with $R^{\lambda_{01}} = X^{\lambda_0} - X^{\lambda_1}$. As the regression model is expressed as (9), we consider the group of transformations given by

$$y \to y + Xb + X^{\lambda_1}d + Zg, \quad (\beta, \delta, \gamma) \to (\beta + b, \delta + d, \gamma + g), \tag{11}$$

where b, d and g are k, k and p dimensional vectors, respectively. Let $M_{w_1} = I_T - W^{\lambda_1}\left(W^{\lambda_1\prime}W^{\lambda_1}\right)^{-1}W^{\lambda_1\prime}$. Then, there exists a $T \times (T - 2k - p)$ matrix P^{λ_1} such that $P^{\lambda_1\prime}P^{\lambda_1} = I_{T-2k-p}$, $P^{\lambda_1}P^{\lambda_1\prime} = M_{w_1}$ and $P^{\lambda_1\prime}W^{\lambda_1} = 0$. Note that, from (10), $P^{\lambda_1\prime}y$ can be expressed as

$$P^{\lambda_1\prime}y|X, Z = P^{\lambda_1\prime}u + P^{\lambda_1\prime}R^{\lambda_{01}}\delta|X, Z \sim N(P^{\lambda_1\prime}R^{\lambda_{01}}\delta, \sigma^2 I_{T-2k-p}), \tag{12}$$

and we construct an optimal test based on $P^{\lambda_1\prime}y$. The advantage of considering $P^{\lambda_1\prime}y$ is that the test statistic becomes invariant to the magnitude of the break given by δ under the null hypothesis.

Let $f_T(y|X, Z; H_N)$ and $f_T(y|X, Z; H_A)$ be the conditional probability density functions when H_N and H_A, respectively, are true and φ be a test for (7). Following Andrews and Ploberger (1994), Andrews, Lee, and Ploberger (1996), and Elliott and Müller (2006, 2007), we consider maximizing the weighted average of power over δ and λ_2 given by

$$\int\int P\left(\varphi \text{ rejects } H_N|\delta, \lambda_2\right) dQ_{\lambda_2}(\delta)dJ(\lambda_2) = \int \varphi(y)\left[\int\int f_T(y|X, Z; H_A)dQ_{\lambda_2}(\delta)dJ(\lambda_2)\right]dy,$$

where $Q_{\lambda_2}(\delta)$ and $J(\lambda_2)$ are non-negative measures on $\mathbb{R}^k$ and $[\underline{\lambda}, \overline{\lambda}]$, respectively. As the term in the square brackets equals $f_T(y|X, Z; H_N)$ under the null hypothesis, the testing problem can be interpreted such that the density under the null hypothesis is given by $f_T(y|X, Z; H_N)$ while it is $\int\int f_T(y|X, Z; H_A)dQ_{\lambda_2}(\delta)dJ(\lambda_2)$ under the alternative. Then, according to the Neyman-Pearson lemma, the most powerful test rejects the null hypothesis when

$$\widetilde{LR}_T(\lambda_1) \equiv \int\int LR_T(\lambda_1, \lambda_2, \delta)dQ_{\lambda_2}(\delta)dJ(\lambda_2) > a$$

for some value a, where $LR_T(\lambda_1, \lambda_2, \delta) = f_T(y|X, Z; H_A)/f_T(y|X, Z; H_N)$.

Following the literature, we suppose as the weighting function for δ a normal distribution given by

$$Q_{\lambda_2}(\delta) \sim N\left(0, \frac{c\sigma^2}{T}\Sigma_{xx,j}^{-1}\right), \tag{13}$$

where c is a positive constant and $j = 1$ for $\lambda_2 < \lambda_0$ and 2 for $\lambda_0 < \lambda_2$. Note that the local-to-zero variance implies that $\delta = O_p(T^{-1/2})$, which corresponds to the local alternative with non-trivial power.

From (12), noting that $P^{\lambda_1\prime}y|X, Z \sim N(0, \sigma^2 I_{T-2k-p})$ under the null hypothesis ($\lambda_0 = \lambda_1$) and $N(P^{\lambda_1\prime}R^{\lambda_{21}}\delta, \sigma^2 I_{T-2K-p})$ under the alternative ($\lambda_0 = \lambda_2$), we have $LR_T(\lambda_1, \lambda_2, \delta) = \exp\left(\frac{1}{\sigma^2}y'M_{w_1}R^{\lambda_{21}}\delta - \frac{1}{2\sigma^2}\delta'R^{\lambda_{21}\prime}M_{w_1}R^{\lambda_{21}}\delta\right)$. Then, using (13), we can show that

$$
\begin{aligned}
&\int LR_T(\lambda_1, \lambda_2, \delta)dQ_{\lambda_2}(\delta) \\
&= \int (2\pi)^{-k/2}\left|\frac{c\sigma^2}{T}\Sigma_{xx,j}^{-1}\right|^{-1/2}\exp\left(\frac{1}{\sigma^2}y'M_{w_1}R^{\lambda_{21}}\delta - \frac{1}{2\sigma^2}\delta'R^{\lambda_{21}\prime}M_{w_1}R^{\lambda_{21}}\delta - \frac{T}{2c\sigma^2}\delta'\Sigma_{xx,j}\delta\right)d\delta \\
&= \left|I_k + \frac{c}{T}\Sigma_{xx,j}^{-1}R^{\lambda_{21}\prime}M_{w_1}R^{\lambda_{21}}\right|^{-1/2} \\
&\qquad \times \exp\left[\frac{c}{2\sigma^2 T}y'M_{w_1}R^{\lambda_{21}}\left(\Sigma_{xx,j} + \frac{c}{T}R^{\lambda_{21}\prime}M_{w_1}R^{\lambda_{21}}\right)^{-1}R^{\lambda_{21}\prime}M_{w_1}y\right], \qquad (14)
\end{aligned}
$$

where the last equality holds by completing the square in the argument of the exponential and using the fact that the integral of a normal density equals one.

The following theorem gives the limiting distribution of $\widetilde{LR}_T(\lambda_1)$.

Theorem 1 *Suppose Assumption 1 holds. Then, under the null hypothesis H_N ($\lambda_0 = \lambda_1$),*

$$
\begin{aligned}
\widetilde{LR}_T(\lambda_1) &\Rightarrow \widetilde{LR}(\lambda_1) \qquad (15) \\
&\equiv \int (1 + c\omega(\lambda_1, \lambda_2))^{-k/2}\exp\left[\frac{c}{2(1 + c\omega(\lambda_1, \lambda_2))}G(\lambda_1, \lambda_2)'G(\lambda_1, \lambda_2)\right]dJ(\lambda_2),
\end{aligned}
$$

where

$$
\omega(\lambda_1, \lambda_2) = \begin{cases} \frac{\lambda_2(\lambda_1 - \lambda_2)}{\lambda_1} & : \quad \lambda_2 < \lambda_1 \\ \frac{(\lambda_2 - \lambda_1)(1 - \lambda_2)}{1 - \lambda_1} & : \quad \lambda_1 < \lambda_2, \end{cases}
$$

$$
G(\lambda_1, \lambda_2) = \begin{cases} B(\lambda_1) - B(\lambda_2) - \frac{\lambda_1 - \lambda_2}{\lambda_1}B(\lambda_1) & : \quad \lambda_2 < \lambda_1 \\ B(\lambda_1) - B(\lambda_2) + \frac{\lambda_2 - \lambda_1}{1 - \lambda_1}\left(B(1) - B(\lambda_1)\right) & : \quad \lambda_1 < \lambda_2. \end{cases}
$$

In principle, the confidence set for the break date can be constructed by inverting the above-mentioned test; we test for the null hypothesis of $T_0 = T_1$ using the above-mentioned optimal test and include T_1 in the confidence set if the null hypothesis is accepted; otherwise it is not included. We perform this procedure for T_1 from $\underline{T} \equiv [\underline{\lambda}T]$ to $\overline{T} \equiv [\overline{\lambda}T]$. However, we need to deal with the localizing parameter c and the weighting function $J(\lambda_2)$ to obtain the critical values because the limiting distribution $\widetilde{LR}(\lambda_1)$ is determined by both c and $J(\lambda_2)$.

2.3. Specific form of the optimal test

A possible solution for dealing with c is to allow $c \to 0$ or $c \to \infty$, as proposed by Andrews and Ploberger (1994). In terms of the limiting distribution (15), these techniques lead to the well-known average-type and exponential-type forms given by

$$\lim_{c \to 0} \frac{2}{c} \left(\widetilde{LR}(\lambda_1) - 1 \right) = \int G(\lambda_1, \lambda_2)' G(\lambda_1, \lambda_2) dJ(\lambda_2), \tag{16}$$

$$\begin{aligned} &\lim_{c \to \infty} \log(1 + c^2)^{k/2} \widetilde{LR}(\lambda_1) \\ &= \log \int_{\lambda_2 \in \Lambda_\varepsilon} \omega(\lambda_1, \lambda_2)^{-k/2} \exp \left[\frac{1}{2\omega(\lambda_1, \lambda_2)} G(\lambda_1, \lambda_2)' G(\lambda_1, \lambda_2) \right] dJ(\lambda_2), \end{aligned} \tag{17}$$

where in (17), the integral with respect to $J(\lambda_2)$ is taken except for the neighborhood of λ_1, which is given by $\Lambda_\varepsilon \equiv \{\lambda_2 : \ \underline{\lambda} \leq \lambda_2 < \lambda_1 - \varepsilon, \ \lambda_1 + \varepsilon < \lambda_2 \leq \overline{\lambda}_2\}$, to avoid the explosive behavior of $\omega(\lambda_1, \lambda_2)^{-k/2}$ around $\lambda_1 \simeq \lambda_2$. The exclusion of the neighborhood of λ_1 is not expected to have a serious effect on the test as long as ε is small, because λ_1 and λ_2 would be indistinguishable in finite samples if they were too close.

If we choose a uniform measure on $[\underline{\lambda}, \overline{\lambda}]$, (16) and (17) become

$$\text{avg-}LR(\lambda_1) \equiv \int_{\underline{\lambda}}^{\overline{\lambda}} G(\lambda_1, \lambda_2)' G(\lambda_1, \lambda_2) d\lambda_2, \tag{18}$$

$$\exp\text{-}LR(\lambda_1) \equiv \log \int_{\lambda_2 \in \Lambda_\epsilon} \omega(\lambda_1, \lambda_2)^{-k/2} \exp \left[\frac{1}{2\omega(\lambda_1, \lambda_2)} G(\lambda_1, \lambda_2)' G(\lambda_1, \lambda_2) \right] d\lambda_2. \tag{19}$$

The other possibility to make the test independent of c is to focus only on the stochastic part given by $G(\lambda_1, \lambda_2)' G(\lambda_1, \lambda_2)$. In this case, we may consider the sup-type test, as considered in Andrews (1993), given by

$$\text{sup-}LR(\lambda_1) \equiv \sup_{\underline{\lambda} \leq \lambda_2 \leq \overline{\lambda}} G(\lambda_1, \lambda_2)' G(\lambda_1, \lambda_2). \tag{20}$$

For the sup-type test, we can also modify (20) to the weighted version as considered in Bai and Perron (1998). They pointed out that the marginal p values change depending on the given alternative and the test may lose power against some alternative. To avoid this problem, we follow Bai and Perron (1998) and choose the weights for λ_2 such that

the marginal p values become the same for all possible λ_2. In our case, it is easy to see that $G(\lambda_1,\lambda_2)'G(\lambda_1,\lambda_2)/\omega(\lambda_1,\lambda_2) \sim \chi^2_k$ for all λ_2 and thus we also consider the following weighted sup-type test:

$$\text{Wsup-}LR(\lambda_1) \equiv \sup_{\lambda_2 \in \Lambda_\varepsilon} \frac{1}{\omega(\lambda_1,\lambda_2)} G(\lambda_1,\lambda_2)'G(\lambda_1,\lambda_2). \tag{21}$$

However, our preliminary simulation shows that (21) performs similarly to (20) and there is no significance difference. Hence, we do not discuss the weighted version of the test in the rest of the paper.[2]

Before proposing the test statistics corresponding to the limit expressions, we relax Assumption 1 to accommodate more general models.

3. Construction of the Confidence Set

3.1. Test statistics for the break date

In Section 2, we derived several versions of the optimal test in asymptotic form but Assumption 1 is too restrictive for practical analysis. In this section, we no longer assume normality in $\{u_t\}$, and serial correlation and heteroskedasticity are allowed in the error term. More precisely, we discard Assumptions 1(i) and (ii), and the FCLT in (5) and (6) are replaced by

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{[rT]} x_t u_t \quad \Rightarrow \quad \Omega_1^{1/2} B(r) \quad \text{for} \quad 0 \le r \le \lambda_0, \tag{22}$$

$$\frac{1}{\sqrt{T}} \sum_{t=T_0+1}^{[rT]} x_t u_t \quad \Rightarrow \quad \Omega_2^{1/2} \left(B(r) - B(\lambda_0)\right) \quad \text{for} \quad \lambda_0 < r \le 1, \tag{23}$$

where Ω_1 and Ω_2 are symmetric and positive definite matrices.

In accordance with the generalization given by (22) and (23), we construct the test statistics corresponding to the limiting expressions (18)–(20). Let

$$F_{T_2}(T_1) = \begin{cases} F_{1,T_2}(T_1) \equiv \frac{1}{T} \left(\sum_{t=T_2+1}^{T_1} x_t \hat{u}_t\right)' \hat{\Omega}_1^{-1} \left(\sum_{t=T_2+1}^{T_1} x_t \hat{u}_t\right) & : \quad T_2 < T_1 \quad (\lambda_2 < \lambda_1) \\ F_{2,T_2}(T_1) \equiv \frac{1}{T} \left(\sum_{t=T_1+1}^{T_2} x_t \hat{u}_t\right)' \hat{\Omega}_2^{-1} \left(\sum_{t=T_1+1}^{T_2} x_t \hat{u}_t\right) & : \quad T_1 < T_2 \quad (\lambda_1 < \lambda_2), \end{cases} \tag{24}$$

[2]Similarly, we can consider the weighted versions of the average- and exponential-type tests but again, their performance is very similar to that of the unweighted versions.

where $\hat{u}_t$ are the regression residuals of y_t on x_t, $x_t^{\lambda_1}$, and z_t, $F_{T_2}(T_1) = 0$ for $T_2 = T_1$, and $\hat{\Omega}_1$ and $\hat{\Omega}_2$ are consistent estimators of Ω_1 and Ω_2, respectively. The estimation of long-run variance is explained later. Corresponding to Λ_ε, we define

$$\Gamma_\varepsilon \equiv \left\{T_2 :\ \underline{T} \le T_2 \le T_1 - T_\varepsilon - 1,\ T_1 + T_\varepsilon + 1 \le T_2 \le \overline{T}\right\},$$

where $T_\varepsilon = [\varepsilon T]$. Note that when $\underline{T} \le T_1 \le \underline{T} + T_\varepsilon$, the set $\{T_2 :\ \underline{T} \le T_2 \le T_1 - T_\varepsilon - 1\}$ is empty and thus $\Gamma_\varepsilon \equiv \left\{T_2 :\ T_1 + T_\varepsilon + 1 \le T_2 \le \overline{T}\right\}$ in this case. Similarly, we should carefully treat the case in which $\overline{T} - T_\varepsilon \le T_1 \le \overline{T}$. Then, we propose the following test statistics:

$$\text{avg-}LR_T(T_1) = \frac{1}{T^*} \sum_{T_2=\underline{T}}^{\overline{T}} F_{T_2}(T_1), \tag{25}$$

$$\text{exp-}LR_T(T_1) = \log\left[\frac{1}{T^{**}} \sum_{T_2 \in \Gamma_\varepsilon} \omega(\lambda_1, \lambda_2)^{-k/2} \exp\left(\frac{1}{2\omega(\lambda_1, \lambda_2)} F_{T_2}(T_1)\right)\right], \tag{26}$$

$$\text{sup-}LR_T(T_1) = \max_{\underline{T} \le T_2 \le \overline{T}} F_{T_2}(T_1), \tag{27}$$

where $T^* = \overline{T} - \underline{T} + 1$ and T^{**} is the number of observations included in Γ_ε.

Theorem 2 *Under Assumptions 1 (iii)–(v) with* (5) *and* (6) *replaced by* (22) *and* (23)*, respectively, the avg-$LR_T(T_1)$, exp-$LR_T(T_1)$, and sup-$LR_T(T_1)$ converge to the corresponding distributions given by* (18)–(20)*, respectively, under the null hypothesis.*

Note that the limiting distributions depend not only on k, $\underline{\lambda}$, $\overline{\lambda}$, and the trimming parameter ε but also on the maintained break fraction λ_1 under the null hypothesis. As a result, the table for critical values becomes huge and is inconvenient for practical analysis. To avoid this, we conduct the response surface regressions.

Remark 1 *The test statistic proposed by Elliott and Müller (2007) can be expressed as*

$$U_T(T_1) = \frac{T}{T_1^2} \sum_{T_2=1}^{T_1} F_{1,T_2}(T_1) + \frac{T}{(T-T_1)^2} \sum_{T_2=T_1+1}^{T} F_{2,T_2}(T_1)$$

and then their test can be interpreted as a weighted average-type test with weights given by T/T_1^2 for $t < T_1$ and $T/(T-T_1)^2$ for $t > T_1$. These weights are convenient for practical

analysis because the null limiting distribution of $U_T(T_1)$ does not depend on λ_1. However, as these weights are not chosen by taking the power property into account, the test leads to a large confidence set as seen in the simulation section.

As the limiting distributions are nonstandard, we approximate the critical values via Monte Carlo simulations. We choose the trimming parameter ε as 0.05 and set $\underline{\lambda} = 1 - \overline{\lambda} = \varepsilon$.[3] For a given value of λ_1, the critical values are obtained from 50,000 replications by approximating a k dimensional Brownian motion by the partial sums from 1,000 normal distributions. We obtain these critical values for λ_1 ranging from $\underline{\lambda}$ to $\overline{\lambda}$ step by 0.01. As a result, we obtain $100(1-2\varepsilon)+1$ critical values for one test statistic with a given number of regressors k, ε, and the significance level. Next, we derive the response surfaces of the critical values as a function of λ_1, which, after trying various kinds of functions of λ_1, are given by

$$cval(\lambda_1|k, \varepsilon = 0.05, \text{significance level}) = a_0 + a_{-1} \times \frac{1}{d+1} + a_1 \times d + a_2 \times d^2 + a_3 \times d^3$$

for the sup-LR_T and avg-LR_T, where $d = |\lambda_1 - 0.5|$.

On the other hand, for the exp-LR_T, because the possible break fractions λ_2 within $|\lambda_1 - \lambda_2| \leq \varepsilon$ are trimmed to construct the test statistic, the distribution for a given λ_1 is rather different depending on whether λ_1 is in the middle of possible fractions ($\underline{\lambda} + \varepsilon < \lambda_1 < \overline{\lambda} - \varepsilon$) or close to either end point ($\underline{\lambda} \leq \lambda_1 \leq \underline{\lambda} + \varepsilon$ or $\overline{\lambda} - \varepsilon \leq \lambda_1 \leq \overline{\lambda}$). More precisely, when $\underline{\lambda} \leq \lambda_1 \leq \underline{\lambda} + \varepsilon$, all λ_2 between $\underline{\lambda}$ and λ_1 are trimmed and the test statistic is constructed using $F_{T_2}(T_1)$ only for $T_2 > T_1$ as follows:

$$\text{exp-}LR_T(T_1) = \log\left[\frac{1}{\overline{T} - T_1 - T_\varepsilon} \sum_{T_2 = T_1 + T_\varepsilon + 1}^{\overline{T}} \omega(\lambda_1, \lambda_2)^{-k/2} \exp\left(\frac{1}{2\omega(\lambda_1, \lambda_2)} F_{T_2}(T_1)\right)\right].$$

Similarly, when $\overline{\lambda} - \varepsilon \leq \lambda_1 \leq \overline{\lambda}$,

$$\text{exp-}LR_T(T_1) = \log\left[\frac{1}{T_1 - \underline{T} - T_\varepsilon} \sum_{T_2 = \underline{T}}^{T_1 - T_\varepsilon - 1} \omega(\lambda_1, \lambda_2)^{-k/2} \exp\left(\frac{1}{2\omega(\lambda_1, \lambda_2)} F_{T_2}(T_1)\right)\right].$$

[3] In addition, we obtained the critical values for $\varepsilon = 0.1$ but do not report them to save space.

On the other hand, when $\underline{\lambda}+\varepsilon<\lambda_1<\overline{\lambda}-\varepsilon$, the exponential-type test becomes

$$\begin{aligned}&\text{exp-}LR_T(T_1)\\ &= \log\left[\frac{1}{\overline{T}-\underline{T}-2T_\varepsilon}\left(\sum_{T_2=\underline{T}}^{T_1-T_\varepsilon-1}+\sum_{T_2=T_1+T_\varepsilon+1}^{\overline{T}}\right)\omega(\lambda_1,\lambda_2)^{-k/2}\exp\left(\frac{1}{2\omega(\lambda_1,\lambda_2)}F_{T_2}(T_1)\right)\right].\end{aligned}$$

Therefore, the response surface regression is conducted depending on whether λ_1 is near the boundary, as follows:

$$\begin{aligned}&cval(\lambda_1|k,\varepsilon=0.05,\text{significance level})\\ &= I(d<0.4)\times\left(a_0+a_{-1}\times\frac{1}{d+1}+a_1\times d+a_2\times d^2+a_3\times d^3\right)\\ &\quad +I(d\geq 0.4)\times(b_0+b_1\times d+b_2\times d^2+b_3\times d^3)\end{aligned}$$

for the exp-LR_T.

Note that the coefficients are different depending on the test statistics and they are summarized in Tables 1a and 1b for 90 and 95 percentiles, respectively. In all cases, the approximation error relative to the true critical value is less than 0.05.

3.2. Estimation of long-run variance

To construct the test statistics, we need to obtain $\hat{\Omega}_1$ and $\hat{\Omega}_2$, the consistent estimators of long-run variance. Although Elliott and Müller (2007) proposed consistent estimators, the coverage rate of the break date based on their method tends to become very large as pointed out by Chang and Perron (2013) and Yamamoto (2014). To circumvent this problem, we follow the method proposed by Yamamoto (2014), which is given as follows:

1. Estimate the break date by minimizing the sum of squared residuals based on model (8). Denote the break date estimator as $\hat{T}_b$.

$$\hat{T}_b=\arg\min_{T_1}\sum_{t=1}^{T}\left(y_t-x_t'\beta-x_t^{\lambda_1\prime}\delta-z_t'\gamma\right)^2.$$

2. Regress y_t on x_t, $x_t^{\lambda_1}$, $x_t^{\hat{\lambda}_b}$ and z_t where $\hat{\lambda}_b=\hat{T}_b/T$ and denote the regression residuals as $\tilde{u}_t$. Note that $x_t^{\hat{\lambda}_b}$ is omitted from the regression when $|T_1-\hat{T}_b|<k$.

3. Estimate the long-run variance Ω_1 using $X_t\tilde{u}_t$ for $t = 1, \cdots, T_1$. Similarly, estimate Ω_2 using $X_t\tilde{u}_t$ for $t = T_1 + 1, \cdots, T$. Denote the estimators as $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$.

Yamamoto (2014) showed that these long-run variance estimators are consistent under both the null and the fixed alternative.

Proposition 1 *The long-run variance estimators, $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ constructed using the above-mentioned explanation, are consistent estimators of Ω_1 and Ω_2, respectively, under both the null and fixed alternative hypotheses.*

Using $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$, we construct $F_{T_2}(T_1)$ defined in (24). Finally, the test statistics are constructed according to expressions (25)–(27) using $F_{T_2}(T_1)$. Then, the confidence set for the break date is constructed as follows:

1. For a given possible break date T_1, construct the test statistic sup-$LR_T(T_1)$, avg-$LR_T(T_1)$, or exp-$LR_T(T_1)$.

2. If the null hypothesis of $T_0 = T_1$ is accepted at the significance level α, then include T_1 in the confidence set; otherwise exclude it.

3. Continue 1–2 for $\{T_1 : \underline{T} \leq T_1 \leq \overline{T}\}$.

By the above-mentioned procedure, we obtain the confidence set for the break date at the $(1 - \alpha)$ confidence level.

4. Asymptotic Properties under the Alternative

In this section, we investigate the asymptotic property of the confidence set based on the optimal test proposed in Section 3. Because the confidence set is constructed by inverting the test statistic, we see that the more powerful the test statistic is, the smaller the confidence set that can be constructed. That is, the investigation of the power of the test is equivalent to the assessment of the size of the confidence set.

Note that all the optimal tests are functions of $F_{T_2}(T_1)$ defined in (24), and thus, we derive the limiting distribution of $F_{T_2}(T_1)$ under the alternative. We assume that the regressors are

homogeneous through the sample period to simplify the expression. Note that under the alternative, T_0 does not necessarily equal T_2 ($\lambda_0 \neq \lambda_2$ in general).

Theorem 3 *Suppose that* $\Sigma_1 = \Sigma_2 = \Sigma$, $\Omega_1 = \Omega_2 = \Omega$, $\hat{\Omega}_1 \xrightarrow{p} \Omega$ *and* $\hat{\Omega}_2 \xrightarrow{p} \Omega$. *Then, under the alternative hypothesis,* $\lambda_0 \neq \lambda_1$, *we have*

$$\frac{1}{T} F_{T_2}(T_1) \xrightarrow{p} \psi(\lambda_0, \lambda_1, \lambda_2) \times \delta' \Sigma_{xx} \Omega^{-1} \Sigma_{xx} \delta, \tag{28}$$

where for $\lambda_0 < \lambda_1$,

$$\psi(\lambda_0, \lambda_1, \lambda_2) = \begin{cases} \frac{\lambda_2^2(\lambda_1 - \lambda_0)^2}{\lambda_1^2} & : \quad \lambda_2 \leq \lambda_0 < \lambda_1 \\ \frac{\lambda_0^2(\lambda_1 - \lambda_2)^2}{\lambda_1^2} & : \quad \lambda_0 \leq \lambda_2 < \lambda_1 \\ 0 & : \quad \lambda_0 < \lambda_1 < \lambda_2 \end{cases},$$

and for $\lambda_1 < \lambda_0$,

$$\psi(\lambda_0, \lambda_1, \lambda_2) = \begin{cases} 0 & : \quad \lambda_2 < \lambda_1 < \lambda_0 \\ \frac{(\lambda_2 - \lambda_1)^2(1 - \lambda_0)^2}{(1 - \lambda_1)^2} & : \quad \lambda_1 < \lambda_2 \leq \lambda_0 \\ \frac{(\lambda_0 - \lambda_1)^2(1 - \lambda_2)^2}{(1 - \lambda_1)^2} & : \quad \lambda_1 < \lambda_0 < \lambda_2 \end{cases}.$$

Note that, as proved by Yamamoto (2014), $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ are consistent estimators of Ω (see Proposition 1).

From Theorem 3, we find the asymptotic properties of the confidence sets based on the optimal tests. First, all the optimal tests are consistent as $T \to \infty$, and thus, we expect that the confidence sets obtained by inverting the optimal tests become smaller for the fixed magnitude of the break as T goes to infinity. In addition, we see that because the test statistics are asymptotically increasing functions of $\|\delta\|$, the larger the magnitude of the break is, the smaller the confidence sets that are obtained. Regarding the location of the break fraction, if the true break fraction λ_0 is smaller than the fraction under the maintained hypothesis λ_1, then $F_{T_2}(T_1)$ for $T_2 > T_1$ does not contribute to the detection of the true break point. This is a natural result because when the true break point T_0 is located to the left of the maintained hypothetical break point T_1, the optimal tests should not detect the break point after T_1. A similar property is observed for the case in which $\lambda_1 < \lambda_0$.

Furthermore, using expression (28), we obtain the probability limits of the sup-type and the average-type tests by direct calculation given by

$$\begin{aligned}\frac{1}{T}\text{sup-}LR_T(T_1) &\xrightarrow{p} \sup_{\underline{\lambda}\le\lambda_2\le\overline{\lambda}} \psi(\lambda_0,\lambda_1,\lambda_2)\ \delta'\Sigma_{xx}\Omega^{-1}\Sigma_{xx}\delta \\ &= \begin{cases} \frac{\lambda_0^2(\lambda_1-\lambda_0)^2}{\lambda_1^2}\ \delta'\Sigma_{xx}\Omega^{-1}\Sigma_{xx}\delta & : \ \lambda_0<\lambda_1 \\ \frac{(\lambda_0-\lambda_1)^2(1-\lambda_0)^2}{(1-\lambda_1)^2}\ \delta'\Sigma_{xx}\Omega^{-1}\Sigma_{xx}\delta & : \ \lambda_1<\lambda_0 \end{cases}. \end{aligned} \tag{29}$$

and

$$\begin{aligned}\frac{1}{T}\text{avg-}LR_T(T_1) &\xrightarrow{p} \int_{\underline{\lambda}}^{\overline{\lambda}} \psi(\lambda_0,\lambda_1,\lambda_2)d\lambda_2\ \delta'\Sigma_{xx}\Omega^{-1}\Sigma_{xx}\delta \\ &= \begin{cases} \frac{(\lambda_1-\lambda_0)^2(\lambda_0^2\lambda_1-\underline{\lambda}^3)}{3\lambda_1^2}\ \delta'\Sigma_{xx}\Omega^{-1}\Sigma_{xx}\delta & : \ \lambda_0<\lambda_1 \\ \frac{(\lambda_0-\lambda_1)^2[(1-\lambda_0)^2(1-\lambda_1)-(1-\overline{\lambda})^3]}{3(1-\lambda_1)^2}\ \delta'\Sigma_{xx}\Omega^{-1}\Sigma_{xx}\delta & : \ \lambda_1<\lambda_0 \end{cases}. \end{aligned} \tag{30}$$

For example, these functions are shown in Figure 1 as a function of λ_0 when $\lambda_1 = 0.3$, $\underline{\lambda} = 0.05$ and $\overline{\lambda} = 0.95$. Note that, for $\lambda_0 < \lambda_1$, the limiting function (29) is maximized when the true break fraction is located at $\lambda_1/2$, the middle point between 0 and λ_1. Intuitively, this is because the information before and after the true break is balanced at this point (the information after λ_1 is not used in this case). On the other hand, for $\lambda_1 < \lambda_0$, the limit function is maximized at $\lambda_0 = (1+\lambda_1)/2$. A similar result is obtained for the average-type test.

5. Finite Sample Performance

In this section, we investigate the finite sample properties of the confidence sets proposed in the study. We consider six data generating processes (DGP) that are the same as in Elliott and Müller (2007). The first four DGPs are based on

$$y_t = 1 + \frac{d}{\sqrt{T}} I(t \ge T_0) + u_t,$$

where $d = 4$, 8, 12 and 16. The error term $\{u_t\}$ is generated from $i.i.d.N(0,1)$ in the standard case (DGP1); a one-time structural change in variance is allowed in DGP2 such that $u_t \sim i.i.d.N(0,1)$ for $t \le T_0$ and $i.i.d.N(0,4)$ for $t \ge T_0$; DGP3 allows for autoregressive disturbance and $u_t = 0.3u_{t-1} + \epsilon_t$ with $\epsilon_t \sim i.i.d.N(0, 0.49)$, while the moving average structure is supposed in DGP4 as $u_t = \epsilon_t - 0.3\epsilon_{t-1}$ with $\epsilon_t \sim i.i.d.N(0, 1/0.49)$.

In DGPs 5 and 6, a model with a partial structural change is investigated as given by

$$y_t = 1 + x_t + \frac{d}{\sqrt{T}} x_t I(t > T_0) + u_t, \quad \text{where} \quad x_t = 0.5x_{t-1} + \xi_t$$

with $\xi_t \sim i.i.d.N(0, 0.75)$. DGP5 supposes the spherical distribution as $u_t \sim i.i.d.N(0, 1)$ while conditional heteroskedasticity is allowed in DGP6 such that $u_t = |x_t|\epsilon_t$ where $\epsilon_t \sim i.i.d.N(0, 0.333)$. Note that these six DGPs are designed so that the long-run variance Ω_1 becomes equal to one.

As we usually do not have prior information about serial correlation and heteroskedasticity in $\{u_t\}$, we estimate the long-run variance using the quadratic spectral kernel with bandwidth selected by the method proposed by Andrews (1991) for all kinds of DGPs. The sample sizes considered are 100 and 300, the break fraction is set to 0.3, 0.5, and 0.7, and the confidence level is 0.95 (the significance level for the tests is 0.05). The number of replications is 5,000 and all computations are conducted using the GAUSS matrix language.

Table 2a reports the coverage rates and the average lengths of the confidence sets when $\lambda_0 = 0.5$ and $T = 100$. The column "EM" corresponds to the result based on the method by Elliott and Müller (2007), while "MEM" implies the modified EM method by Yamamoto (2014) and "Bai" implies the confidence interval based on the asymptotic distribution of $\hat{T}_b$ by Bai (1997).[4] In the standard case (DGP1), the coverage rate of EM is slightly greater than the nominal level for any magnitude of the break, while Bai's method results in an overly liberal coverage rate for small d. On the other hand, the coverage rates of the modified EM and our three methods are close to the nominal confidence level, although they tend to be slightly liberal for small d. Regarding the average lengths of the confidence sets, the MEM method successfully reduces the size of the confidence set compared to the original EM method but the confidence set of the MEM method remains larger than that based on Bai's method. On the other hand, the confidence sets based on our methods are smaller than those based on EM and MEM. Moreover, in some cases, they are almost the same as or smaller than the confidence interval based on Bai's method. For example, when $d = 12$, the length of the confidence set based on Bai's method is 0.168, whereas those based on the sup-type,

[4]For a fair comparison, we exclude the first and last 5% end points from the possible break points and the trimmed version of the EM and MEM tests are considered using the simulated critical values.

the average-type, and the exponential-type tests are 0.169, 0.163, and 0.138, respectively. That is, even though Bai's method leads to a liberal coverage rate, the sizes of the confidence sets based on our methods are almost the same or smaller. This is a great advantage of our methods over the existing ones because the coverage rates of our methods are close to the nominal rate, while our confidence sets are at least as small as those based on the existing methods and, in some cases, are smaller. This implies that the confidence sets based on our methods are constructed more efficiently than those based on the other methods. For DGP2, the coverages rates are similar to DGP1 for all methods but the confidence sets are larger in DGP2 than in DGP1, which is expected because the variance of the error term becomes larger after the break in DGP2.

In the case in which the error term is serially correlated, the positive serial correlation in DGP3 makes the coverage rates smaller compared to the i.i.d. case for all methods except that of Bai (1997), while the negative MA structure in the error term has the opposite effect on the coverage rates (DGP4).

In the case of a partial structural change (DGPs 5 and 6), the EM method leads to a conservative coverage rate while liberal coverage rates are obtained for the other methods.

When $T = 300$, the coverage rates of all the methods move closer to the nominal rate, as is observed in Table 2b, while there is no large difference in the average lengths of the confidence sets compared to the case when $T = 100$.

Tables 3a and 3b report the case when $\lambda_0 = 0.3$.[5] The relative performance is preserved compared to the case when $\lambda_0 = 0.5$.

To summarize, no method dominates another in terms of coverage rates; in some cases our methods perform best but in other cases, the EM and MEM methods work better. However, it seems that Bai's method leads to an overly liberal coverage rate for small d, as pointed out in the literature. On the other hand, our methods can reduce the average size of the confidence sets compared to the other methods.

6. Conclusion

[5]The results when $\lambda_0 = 0.7$ are similar to the case when $\lambda_0 = 0.3$ and we omit the results.

In this study, we proposed constructing the confidence set for the break date by inverting the optimal test. By choosing the weight function carefully, we proposed three optimal tests: the sup-type, the average-type and the exponential-type tests. By Monte Carlo simulations, we found that the coverage rates based on the inversion of our optimal tests are comparable to those based on the methods of Elliott and Müller (2007) and Yamamoto (2014) but are closer to the nominal confidence level than that based on Bai's (1997) method. On the other hand, the confidence sets based on our tests dominate those based on the other methods in terms of the average lengths, even when the coverage rates of those methods are similar. Considering these results, our methods could play an important role in investigating statistical inferences about break dates.

Appendix

Proof of Theorem 1: We first consider the case in which $\lambda_2 < \lambda_1$. Noting that $M_{w_1} = M_{\tilde{w}_1}$ where $M_{\tilde{w}_1} = \tilde{W}^{\lambda_1}(\tilde{W}^{\lambda_1\prime}\tilde{W}^{\lambda_1})^{-1}\tilde{W}^{\lambda_1\prime}$ with $\tilde{W}^{\lambda_1} = [(X - X^{\lambda_1}), X^{\lambda_1}, Z]$, we have, by the WLLN,

$$\frac{1}{T}R^{\lambda_{21}\prime}R^{\lambda_{21}} = \sum_{t=T_2+1}^{T_1} x_t x_t' \xrightarrow{p} (\lambda_1 - \lambda_2)\Sigma_{xx,1}, \tag{31}$$

$$\frac{1}{T}R^{\lambda_{21}\prime}\tilde{W}^{\lambda_1} = \left[\sum_{t=T_2+1}^{T_1} x_t x_t', 0, \sum_{t=T_2+1}^{T_1} x_t z_t'\right] \xrightarrow{p} (\lambda_1 - \lambda_2)\left[\Sigma_{xx,1}, 0, \Sigma_{xz,1}\right], \tag{32}$$

$$\frac{1}{T}\tilde{W}^{\lambda_1\prime}\tilde{W}^{\lambda_1} \xrightarrow{p} \begin{bmatrix} \lambda_1\Sigma_{xx,1} & 0 & \lambda_1\Sigma_{xz,1} \\ 0 & (1-\lambda_1)\Sigma_{xx,2} & (1-\lambda_1)\Sigma_{xz,2} \\ \lambda_1\Sigma_{zx,1} & (1-\lambda_1)\Sigma_{zx,2} & \lambda_1\Sigma_{zz,1} + (1-\lambda_1)\Sigma_{zz,2} \end{bmatrix}. \tag{33}$$

Then, using (31)–(33) and the standard matrix algebra, we have

$$\frac{1}{T}R^{\lambda_{21}\prime}M_{\tilde{w}_1}R^{\lambda_{21}} \xrightarrow{p} \frac{\lambda_2(\lambda_1 - \lambda_2)}{\lambda_1}\Sigma_{xx,1}, \tag{34}$$

so that for the determinant in (14), we obtain

$$\begin{aligned} \left|I_k + \frac{c}{T}\Sigma_{xx,1}^{-1}R^{\lambda_{21}\prime}M_{\tilde{w}_1}R^{\lambda_{21}}\right|^{-1/2} &\xrightarrow{p} \left|\left(1 + c\frac{\lambda_2(\lambda_1 - \lambda_2)}{\lambda_1}\right)I_k\right|^{-1/2} \\ &= (1 + c\omega(\lambda_1, \lambda_2))^{-k/2}, \end{aligned} \tag{35}$$

where $\omega(\lambda_1, \lambda_2)$ is defined in Theorem (1).

To derive the limiting distribution in the argument of the exponential term in (14), we note that $P^{\lambda_1\prime}y = P^{\lambda_1\prime}u$ under the null hypothesis of $\lambda_0 = \lambda_1$ (see also (12)), and thus, $M_{w_1}y = M_{\tilde{w}_1}y = M_{\tilde{w}_1}u$. Then, from (32), (33), and the FCLT, we have

$$\begin{aligned}
\frac{1}{\sqrt{T}}R^{\lambda_{21}\prime}M_{w_1}y &= \frac{1}{\sqrt{T}}R^{\lambda_{21}\prime}u - \left(\frac{1}{T}R^{\lambda_{21}\prime}\tilde{W}^{\lambda_1}\right)\left(\frac{1}{T}\tilde{W}^{\lambda_1\prime}\tilde{W}^{\lambda_1}\right)^{-1}\left(\frac{1}{\sqrt{T}}\tilde{W}^{\lambda_1}u\right) \\
&\Rightarrow \sigma\Sigma_{xx,1}^{1/2}\left(B(\lambda_1) - B(\lambda_2)\right) - \frac{\lambda_1-\lambda_2}{\lambda_1}\sigma\Sigma_{xx,1}^{1/2}B(\lambda_1) \\
&= \sigma\Sigma_{xx,1}^{1/2}G(\lambda_1,\lambda_2), \qquad (36)
\end{aligned}$$

where $G(\lambda_1,\lambda_2)$ is defined in Theorem 1. Then, using (34) and (36), we can see that

$$\begin{aligned}
&\frac{c}{2\sigma^2 T}y'M_{w_1}R^{\lambda_{21}}\left(\Sigma_{xx,1} + \frac{c}{T}R^{\lambda_{21}\prime}M_{w_1}R^{\lambda_{21}}\right)^{-1}R^{\lambda_{21}\prime}M_{w_1}y \\
&\Rightarrow \frac{c}{2(1+c\omega(\lambda_1,\lambda_2))}G(\lambda_1,\lambda_2)'G(\lambda_1,\lambda_2). \qquad (37)
\end{aligned}$$

From (35) and (37), we obtain (15) for $\lambda_2 < \lambda_1$.

In exactly the same manner, we can prove the case in which $\lambda_1 < \lambda_2$.■

Proof of Theorem 2: This can be proved in exactly the same manner as the proof of Theorem 1 using the FCLT supposed in (22) and (23).■

Proof of Theorem 3: Let $\hat{\theta} = [\hat{\beta}', \hat{\delta}', \hat{\gamma}']'$ be the least squares estimator of $\theta = [\beta', \delta', \gamma']'$. We first derive the probability limit of $\hat{\theta}$ under the fixed alternative.

Lemma 1 *Under the assumption of Theorem 3, we have, for* $\lambda_0 < \lambda_1$,

$$\hat{\theta} - \theta \xrightarrow{p} \begin{bmatrix} \frac{\lambda_1-\lambda_0}{\lambda_1}\delta \\ -\frac{\lambda_1-\lambda_0}{\lambda_1}\delta \\ 0 \end{bmatrix},$$

and for $\lambda_1 < \lambda_0$,

$$\hat{\theta} - \theta \xrightarrow{p} \begin{bmatrix} 0 \\ -\frac{\lambda_1-\lambda_0}{1-\lambda_1}\delta \\ 0 \end{bmatrix}.$$

Proof of Lemma 1: From (10), the least squares estimator can be expressed as

$$\begin{aligned}
\hat{\theta} &= \theta + \left(W^{\lambda_1\prime}W^{\lambda_1}\right)^{-1}W^{\lambda_1\prime}u^{\lambda_{01}} \\
&= \theta + \left(\frac{1}{T}W^{\lambda_1\prime}W^{\lambda_1}\right)^{-1}\frac{1}{T}W^{\lambda_1\prime}R^{\lambda_{01}}\delta + O_p\left(\frac{1}{\sqrt{T}}\right).
\end{aligned}$$

Since

$$T^{-1}W^{\lambda_1\prime}W^{\lambda_1} \xrightarrow{p} \begin{bmatrix} \frac{1}{\lambda_1}\Sigma_{xx}^{-1} + \Sigma_{xx}^{-1}\Sigma_{xz}\Sigma_{zz\cdot x}^{-1}\Sigma_{zx}\Sigma_{xx}^{-1} & -\frac{1}{\lambda_1}\Sigma_{xx}^{-1} & -\Sigma_{xx}^{-1}\Sigma_{xz}\Sigma_{zz\cdot x}^{-1} \\ -\frac{1}{\lambda_1}\Sigma_{xx}^{-1} & \frac{1}{\lambda_1(1-\lambda_1)}\Sigma_{xx}^{-1} & 0 \\ -\Sigma_{zz\cdot x}^{-1}\Sigma_{zx}\Sigma_{xx}^{-1} & 0 & \Sigma_{zz\cdot x}^{-1} \end{bmatrix},$$

where $\Sigma_{zz\cdot x} = \Sigma_{zz} - \Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xz}$, while for $\lambda_0 < \lambda_1$,

$$\frac{1}{T}W^{\lambda_1\prime}R^{\lambda_{01}} \xrightarrow{p} \left[(\lambda_1 - \lambda_0)\Sigma_{xx}, 0, (\lambda_1 - \lambda_0)\Sigma_{zx}'\right]',$$

we obtain the result. Similarly, the lemma is proved for $\lambda_1 < \lambda_0$ because

$$\frac{1}{T}W^{\lambda_1\prime}R^{\lambda_{01}} \xrightarrow{p} -\left[(\lambda_0 - \lambda_1)\Sigma_{xx}, (\lambda_0 - \lambda_1)\Sigma_{xx}, (\lambda_0 - \lambda_1)\Sigma_{zx}'\right]'$$

in this case.■

Consider the case in which $\lambda_0 < \lambda_1$. When $\lambda_2 \leq \lambda_0 < \lambda_1$, we have

$$\frac{1}{T}F_{T_2}(T_1) = \frac{1}{T}F_{1,T_2}(T_1) = \left(\frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t\hat{u}_t\right)' \hat{\Omega}_1^{-1} \left(\frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t\hat{u}_t\right).$$

Since $\hat{u}_t = u_t^{\lambda_{01}} - w_t^{\lambda_1\prime}(\hat{\theta} - \theta) = u_t + r_t^{\lambda_{01}\prime}\delta - w_t^{\lambda_1\prime}(\hat{\theta} - \theta)$, we can see using Lemma 1 that

$$\begin{aligned} \frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t\hat{u}_t &= = \frac{1}{T}\sum_{t=T_0+1}^{T_1} x_t x_t'\delta - \frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t w_t^{\lambda_1\prime}(\hat{\theta} - \theta) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\xrightarrow{p} (\lambda_1 - \lambda_0)\Sigma_{xx}\delta - \frac{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_0)}{\lambda_1}\Sigma_{xx}\delta \\ &= \frac{\lambda_2(\lambda_1 - \lambda_0)}{\lambda_1}\Sigma_{xx}\delta, \end{aligned}$$

because $r_t^{\lambda_{01}} = x_t$ for $t = T_0 + 1, \cdots, T_1$ and $r_t^{\lambda_{01}} = 0$ for other t; in particular, it is zero for $t = T_2 + 1, \cdots, T_0$. Similarly, when $\lambda_0 < \lambda_2 < \lambda_1$,

$$\begin{aligned} \frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t\hat{u}_t &= = \frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t r_t^{\lambda_{01}\prime}\delta - \frac{1}{T}\sum_{t=T_2+1}^{T_1} x_t w_t^{\lambda_1\prime}(\hat{\theta} - \theta) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\xrightarrow{p} (\lambda_1 - \lambda_2)\Sigma_{xx}\delta - \frac{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_0)}{\lambda_1}\Sigma_{xx}\delta \\ &= \frac{\lambda_0(\lambda_1 - \lambda_0)}{\lambda_1}\Sigma_{xx}\delta. \end{aligned}$$

On the other hand, when $\lambda_1 < \lambda_2$,

$$\frac{1}{T}F_{T_2}(T_1) = \frac{1}{T}F_{2,T_2}(T_1) = \left(\frac{1}{T}\sum_{t=T_1+1}^{T_2} x_t\hat{u}_t\right)' \hat{\Omega}_2^{-1} \left(\frac{1}{T}\sum_{t=T_1+1}^{T_2} x_t\hat{u}_t\right),$$

but because $r_t^{\lambda_{01}} = 0$ for $t = T_1 + 1, \cdots, T_2$, we can see that

$$\begin{aligned} \frac{1}{T}\sum_{t=T_1+1}^{T_2} x_t \hat{u}_t &= 0 - \frac{1}{T}\sum_{t=T_1+1}^{T_2} x_t w_t^{\lambda_1 \prime}(\hat{\theta} - \theta) + O_p\left(\frac{1}{\sqrt{T}}\right) \\ &\xrightarrow{p} 0 \end{aligned}$$

by using Lemma 1 again. Then, we obtain the result for $\lambda_0 < \lambda_1$.

The result for the case in which $\lambda_1 < \lambda_0$ can be obtained in exactly the same manner and we omit the proof.■

References

[1] Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817-858.

[2] Andrews, D. W. K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61, 821-856 (Erratum, 71, 395-397).

[3] Andrews, D. W. K., I. Lee and W. Ploberger (1996). Optimal Changepoint Tests for Normal Linear Regression. *Journal of Econometrics* 70, 9-38.

[4] Andrews, D. W. K. and W. Ploberger (1994). Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative, *Econometrica* 62, 1383-1414.

[5] Bai, J. (1994). Least Squares Estimation of a Shift in Linear Processes. *Journal of Time Series Analysis* 15, 453-472.

[6] Bai, J. (1997). Estimation of a Change Point in Multiple Regression Models. *Review of Economics and Statistics* 79, 551-563.

[7] Bai, J. (2000). Vector Autoregressive Models with Structural Changes in Regression Coefficients and in Variance-Covariance Matrices. *Annals of Economics and Finance* 1, 303-339.

[8] Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66, 47-78.

[9] Boldea, O. and A. R. Hall (2013). Estimation and Inference in Unstable Nonlinear Least Squares Models. *Journal of Econometrics* 172, 158-167.

[10] Boldea, O., A. R. Hall and S. Han (2012). Asymptotic Distribution Theory for Break Point Estimators in Models Estimated via 2SLS. *Econometric Reviews* 31, 1-33.

[11] Chang, S. and P. Perron (2013). A Comparison of Alternative Methods to Construct Confidence Intervals for the Estimate of a Break Date in Linear Regression Models. Manuscript, Boston University.

[12] Elliott, G. and U. K. Müller (2006). Efficient Tests for General Persistent Time Variation in Regression Coefficients. *Review of Economic Studies* 73, 907-940.

[13] Elliott, G. and U. K. Müller (2007). Confidence Sets for the Date of a Single Break in Linear Time Series Regressions. *Journal of Econometrics* 141, 1196-1218.

[14] Perron, P. and Y. Yamamoto (2014). A Note on Estimating and Testing for Multiple Structural Changes in Models with Endogenous Regressors via 2SLS. *Econometric Theory* 30, 491-507.

[15] Qu, Z. and P. Perron (2007). Estimating and Testing Structural Changes in Multivariate Regressions. *Econometrica* 75, 459-502.

[16] Yamamoto, Y. (2014). A Modified Confidence Set for the Structural Break Date in Linear Regression Models. Discussion Paper No.2014-08, Hitotsubashi University.

Table 1a: Coefficients of the Response Surface Regressions (90 percentiles)

	k								
	1	2	3	4	5	6	7	8	9
sup-LR_T									
a_0	0.674	−30.640	−54.062	−85.748	−117.243	−124.059	−132.381	−149.405	−154.744
a_{-1}	0.209	31.850	55.540	87.467	119.183	126.214	134.739	151.961	157.491
a_1	0.061	31.527	55.135	87.069	118.676	125.835	134.462	151.888	157.638
a_2	4.920	−20.676	−39.811	−67.321	−93.940	−99.061	−105.622	−120.808	−125.261
a_3	−5.025	7.028	15.879	30.168	43.305	45.159	48.031	56.150	58.330
avg-LR_T									
a_0	1.543	−0.955	0.694	0.888	−3.583	−8.504	−2.659	−6.329	−11.900
a_{-1}	−1.376	1.245	−0.290	−0.373	4.206	9.234	3.492	7.265	12.941
a_1	−1.405	1.202	−0.300	−0.364	4.193	9.172	3.497	7.254	12.881
a_2	2.524	1.042	2.969	3.612	0.136	−3.697	1.850	−1.026	−5.545
a_3	−1.354	−0.519	−1.333	−1.517	0.481	2.647	−0.101	1.645	4.307
exp-LR_T									
a_0	49.562	112.723	211.010	361.636	522.817	668.605	764.361	824.621	961.868
a_{-1}	−46.806	−107.625	−203.673	−352.114	−511.146	−654.784	−748.385	−806.502	−941.597
a_1	−46.656	−107.328	−203.159	−350.727	−509.000	−651.751	−744.948	−802.756	−936.899
a_2	42.886	100.339	191.442	328.004	475.894	607.590	694.296	748.323	870.439
a_3	−28.593	−67.563	−127.173	−209.419	−299.126	−375.810	−425.702	−456.143	−523.122
b_0	10.670	44.425	29.559	23.697	−24.928	16.270	51.952	−49.026	−64.895
b_1	−57.856	−285.167	−168.789	−113.392	243.104	−33.761	−260.345	456.068	588.899
b_2	137.294	681.472	415.857	287.101	−548.063	104.479	611.721	−1046.975	−1373.591
b_3	−109.402	−543.617	−342.838	−243.880	407.329	−105.401	−482.692	796.157	1062.856

Table 1b: Coefficients of the Response Surface Regressions (95 percentiles)

	k								
	1	2	3	4	5	6	7	8	9
sup-LR_T									
a_0	−20.905	−63.920	−83.800	−137.578	−129.430	−145.183	−144.461	−150.787	−160.708
a_{-1}	21.967	65.326	85.485	139.524	131.600	147.581	147.063	153.594	163.723
a_1	21.664	64.896	85.187	138.957	131.409	147.509	147.335	154.230	164.406
a_2	−12.647	−48.889	−65.838	−112.373	−104.312	−117.634	−116.824	−122.841	−130.591
a_3	3.152	20.890	29.513	53.739	48.405	54.928	54.306	57.755	60.991
avg-LR_T									
a_0	1.110	−8.903	−10.709	−9.273	−16.557	−10.028	−6.826	−9.501	−10.939
a_{-1}	−0.904	9.242	11.168	9.848	17.245	10.825	7.731	10.514	12.058
a_1	−0.945	9.162	11.105	9.825	17.162	10.816	7.741	10.534	12.078
a_2	2.871	−5.393	−6.502	−4.700	−10.524	−4.275	−0.873	−2.988	−3.845
a_3	−1.948	2.737	3.596	2.760	5.929	2.722	0.933	2.448	3.130
exp-LR_T									
a_0	61.014	86.903	166.740	387.611	571.898	730.617	859.267	1000.949	1060.976
a_{-1}	−57.764	−81.201	−158.722	−377.371	−559.456	−715.967	−842.436	−981.932	−1039.758
a_1	−57.690	−80.910	−158.522	−375.523	−556.871	−712.410	−838.187	−976.609	−1033.971
a_2	53.541	75.498	150.786	349.621	519.651	662.240	778.851	904.607	956.344
a_3	−34.743	−52.308	−103.460	−220.661	−324.528	−406.022	−473.472	−542.723	−568.376
b_0	52.623	3.953	76.318	−7.857	31.925	51.379	68.956	6.763	−48.162
b_1	−351.866	6.559	−486.740	121.603	−153.426	−264.089	−374.432	81.533	467.209
b_2	831.743	−8.168	1146.652	−281.839	387.094	619.160	881.009	−195.727	−1061.980
b_3	−655.202	−0.630	−901.924	214.845	−327.476	−487.175	−694.128	152.407	798.697

Table 2a: Coverage rates and lengths of the confidence sets ($T = 100$, $\lambda_0 = 0.5$)

	Coverage rates						Lengths of the confidence sets					
d	EM	MEM	Bai	sup	avg	exp	EM	MEM	Bai	sup	avg	exp
DGP1												
4	0.961	0.918	0.714	0.933	0.917	0.910	0.784	0.695	0.552	0.661	0.642	0.651
8	0.961	0.934	0.866	0.953	0.932	0.935	0.492	0.362	0.323	0.316	0.304	0.302
12	0.961	0.946	0.936	0.969	0.944	0.953	0.305	0.187	0.168	0.169	0.163	0.138
16	0.961	0.953	0.951	0.976	0.952	0.962	0.252	0.126	0.104	0.117	0.114	0.088
DGP2												
4	0.961	0.921	0.658	0.936	0.919	0.914	0.842	0.778	0.570	0.757	0.751	0.731
8	0.961	0.928	0.769	0.944	0.926	0.926	0.715	0.602	0.476	0.551	0.534	0.544
12	0.961	0.936	0.863	0.956	0.933	0.938	0.520	0.383	0.341	0.335	0.322	0.327
16	0.961	0.942	0.908	0.962	0.939	0.948	0.371	0.239	0.225	0.212	0.203	0.191
DGP3												
4	0.940	0.852	0.677	0.882	0.851	0.840	0.776	0.598	0.486	0.578	0.557	0.548
8	0.940	0.884	0.839	0.921	0.882	0.889	0.562	0.287	0.272	0.262	0.248	0.236
12	0.940	0.907	0.914	0.952	0.905	0.925	0.475	0.152	0.143	0.145	0.136	0.114
16	0.940	0.919	0.946	0.970	0.917	0.946	0.546	0.105	0.090	0.102	0.096	0.076
DGP4												
4	0.991	0.971	0.755	0.970	0.970	0.962	0.857	0.795	0.646	0.749	0.735	0.757
8	0.991	0.977	0.902	0.980	0.976	0.974	0.623	0.488	0.428	0.408	0.400	0.418
12	0.991	0.982	0.954	0.986	0.982	0.981	0.369	0.258	0.228	0.217	0.215	0.194
16	0.991	0.986	0.966	0.990	0.985	0.986	0.253	0.168	0.139	0.147	0.147	0.116
DGP5												
4	0.970	0.901	0.650	0.912	0.900	0.869	0.817	0.682	0.498	0.657	0.637	0.622
8	0.970	0.925	0.815	0.941	0.924	0.908	0.610	0.386	0.317	0.346	0.329	0.322
12	0.970	0.941	0.858	0.961	0.941	0.931	0.449	0.219	0.170	0.199	0.191	0.173
16	0.970	0.956	0.877	0.978	0.955	0.954	0.378	0.153	0.106	0.144	0.138	0.121
DGP6												
4	0.981	0.876	0.663	0.877	0.874	0.822	0.807	0.614	0.468	0.589	0.568	0.539
8	0.981	0.916	0.834	0.929	0.915	0.883	0.564	0.325	0.265	0.295	0.279	0.268
12	0.981	0.946	0.912	0.960	0.945	0.924	0.409	0.189	0.141	0.178	0.167	0.152
16	0.981	0.963	0.955	0.979	0.962	0.952	0.354	0.137	0.089	0.132	0.124	0.109

Table 2b: Coverage rates and lengths of the confidence sets ($T = 300$, $\lambda_0 = 0.5$)

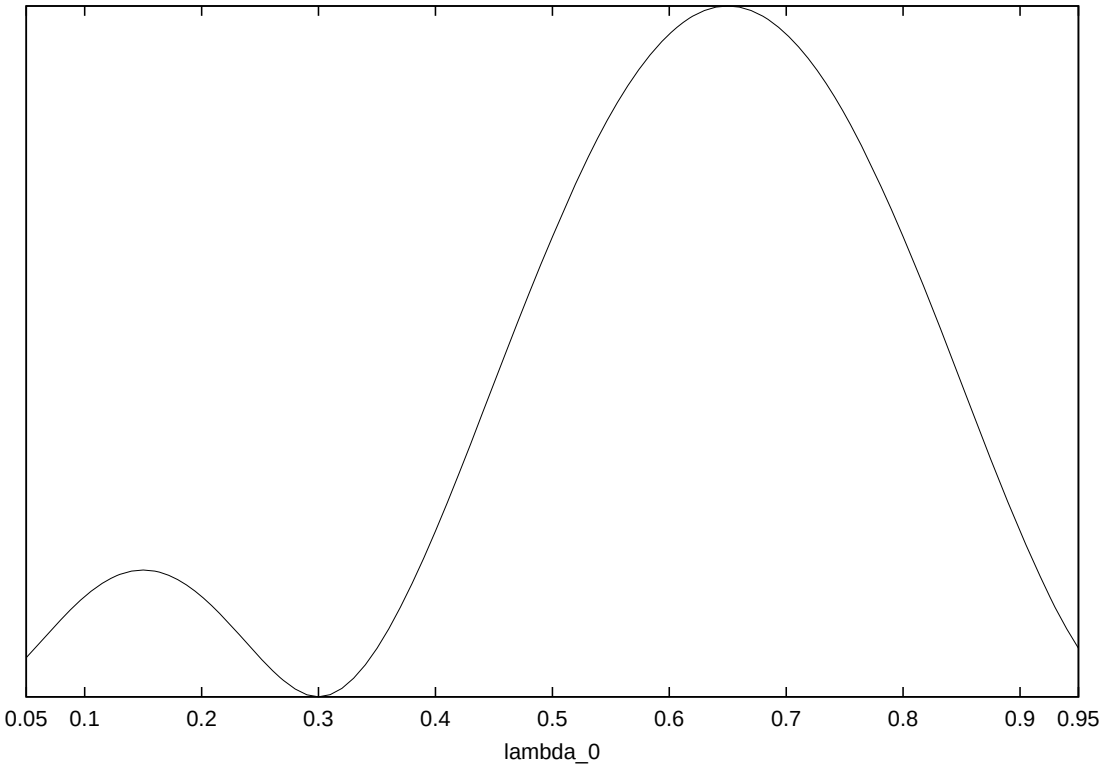
	Coverage rates						Lengths of the confidence sets					
d	EM	MEM	Bai	sup	avg	exp	EM	MEM	Bai	sup	avg	exp
DGP1												
4	0.955	0.943	0.713	0.952	0.942	0.936	0.763	0.734	0.558	0.684	0.673	0.696
8	0.955	0.947	0.872	0.959	0.946	0.946	0.435	0.394	0.320	0.326	0.323	0.329
12	0.955	0.950	0.926	0.961	0.949	0.949	0.229	0.199	0.158	0.172	0.172	0.146
16	0.955	0.951	0.940	0.964	0.950	0.952	0.158	0.132	0.092	0.118	0.118	0.091
DGP2												
4	0.955	0.945	0.674	0.953	0.944	0.941	0.827	0.807	0.585	0.773	0.775	0.765
8	0.955	0.948	0.780	0.957	0.947	0.943	0.683	0.646	0.486	0.574	0.569	0.587
12	0.955	0.949	0.859	0.960	0.948	0.948	0.462	0.418	0.341	0.347	0.345	0.357
16	0.955	0.951	0.907	0.962	0.949	0.950	0.293	0.256	0.218	0.216	0.215	0.203
DGP3												
4	0.940	0.903	0.679	0.919	0.902	0.897	0.746	0.674	0.511	0.633	0.621	0.629
8	0.940	0.916	0.851	0.934	0.914	0.915	0.435	0.338	0.286	0.289	0.284	0.277
12	0.940	0.923	0.903	0.945	0.922	0.927	0.244	0.173	0.141	0.155	0.152	0.126
16	0.940	0.928	0.926	0.952	0.928	0.934	0.180	0.116	0.083	0.107	0.105	0.082
DGP4												
4	0.984	0.976	0.762	0.978	0.976	0.974	0.829	0.801	0.636	0.746	0.738	0.770
8	0.984	0.978	0.906	0.982	0.978	0.978	0.546	0.488	0.398	0.391	0.393	0.418
12	0.984	0.980	0.947	0.984	0.979	0.981	0.294	0.250	0.200	0.205	0.209	0.187
16	0.984	0.981	0.957	0.985	0.980	0.982	0.197	0.163	0.117	0.140	0.143	0.112
DGP5												
4	0.959	0.934	0.683	0.940	0.933	0.921	0.779	0.724	0.532	0.681	0.666	0.676
8	0.959	0.942	0.848	0.948	0.941	0.935	0.481	0.397	0.317	0.337	0.331	0.332
12	0.959	0.948	0.898	0.961	0.947	0.946	0.273	0.210	0.160	0.183	0.182	0.161
16	0.959	0.951	0.899	0.964	0.951	0.952	0.194	0.142	0.093	0.129	0.127	0.105
DGP6												
4	0.966	0.914	0.693	0.920	0.913	0.885	0.774	0.688	0.523	0.650	0.629	0.629
8	0.966	0.930	0.848	0.945	0.929	0.912	0.457	0.363	0.294	0.316	0.304	0.300
12	0.966	0.943	0.910	0.963	0.942	0.933	0.258	0.195	0.148	0.176	0.170	0.149
16	0.966	0.952	0.935	0.971	0.951	0.945	0.185	0.133	0.087	0.124	0.120	0.099

Table 3a: Coverage rates and lengths of the confidence sets ($T = 100$, $\lambda_0 = 0.3$)

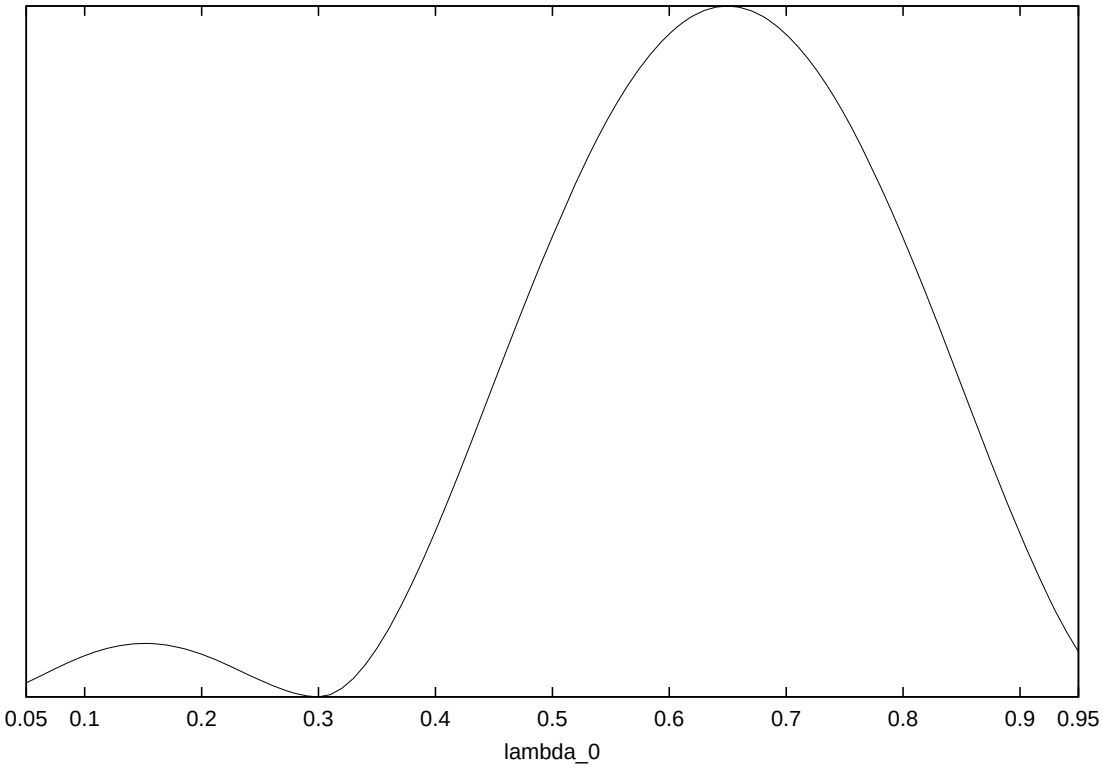
	Coverage rates						Lengths of the confidence sets					
d	EM	MEM	Bai	sup	avg	exp	EM	MEM	Bai	sup	avg	exp
DGP1												
4	0.966	0.919	0.708	0.937	0.927	0.909	0.801	0.711	0.520	0.694	0.674	0.670
8	0.966	0.929	0.859	0.950	0.939	0.928	0.544	0.390	0.313	0.362	0.352	0.326
12	0.966	0.939	0.921	0.961	0.947	0.944	0.342	0.193	0.166	0.194	0.194	0.143
16	0.966	0.952	0.948	0.966	0.951	0.957	0.284	0.125	0.103	0.133	0.136	0.091
DGP2												
4	0.966	0.931	0.613	0.930	0.920	0.920	0.856	0.798	0.570	0.791	0.786	0.772
8	0.966	0.934	0.742	0.939	0.927	0.929	0.760	0.654	0.456	0.626	0.610	0.617
12	0.966	0.939	0.855	0.945	0.934	0.940	0.599	0.446	0.335	0.408	0.397	0.389
16	0.966	0.946	0.906	0.951	0.940	0.948	0.445	0.274	0.232	0.255	0.253	0.216
DGP3												
4	0.951	0.855	0.669	0.897	0.879	0.841	0.793	0.613	0.458	0.610	0.588	0.566
8	0.951	0.887	0.828	0.927	0.904	0.889	0.616	0.306	0.263	0.299	0.287	0.250
12	0.951	0.911	0.901	0.949	0.922	0.922	0.545	0.155	0.140	0.164	0.161	0.116
16	0.951	0.928	0.940	0.962	0.930	0.944	0.608	0.104	0.089	0.115	0.115	0.077
DGP4												
4	0.990	0.970	0.752	0.971	0.968	0.954	0.866	0.806	0.609	0.777	0.766	0.774
8	0.990	0.974	0.893	0.979	0.975	0.965	0.676	0.527	0.413	0.466	0.460	0.459
12	0.990	0.979	0.946	0.984	0.980	0.974	0.410	0.271	0.228	0.251	0.256	0.208
16	0.990	0.982	0.966	0.987	0.982	0.980	0.266	0.169	0.139	0.170	0.177	0.121
DGP5												
4	0.967	0.890	0.635	0.920	0.914	0.867	0.828	0.694	0.465	0.686	0.667	0.637
8	0.967	0.914	0.797	0.945	0.933	0.901	0.653	0.411	0.301	0.390	0.375	0.346
12	0.967	0.932	0.848	0.961	0.943	0.929	0.499	0.226	0.170	0.224	0.221	0.181
16	0.967	0.946	0.870	0.968	0.950	0.947	0.421	0.154	0.106	0.160	0.159	0.122
DGP6												
4	0.973	0.868	0.649	0.886	0.891	0.819	0.819	0.629	0.437	0.618	0.602	0.555
8	0.973	0.903	0.818	0.930	0.921	0.880	0.607	0.340	0.252	0.330	0.313	0.280
12	0.973	0.934	0.908	0.959	0.943	0.922	0.451	0.191	0.138	0.195	0.188	0.152
16	0.973	0.952	0.947	0.973	0.953	0.945	0.392	0.135	0.087	0.143	0.140	0.108

Table 3b: Coverage rates and lengths of the confidence sets ($T = 300$, $\lambda_0 = 0.3$)

	Coverage rates						Lengths of the confidence sets					
d	EM	MEM	Bai	sup	avg	exp	EM	MEM	Bai	sup	avg	exp
DGP1												
4	0.953	0.937	0.712	0.950	0.943	0.931	0.774	0.743	0.520	0.710	0.700	0.710
8	0.953	0.941	0.859	0.953	0.946	0.937	0.473	0.425	0.310	0.376	0.377	0.359
12	0.953	0.944	0.915	0.956	0.947	0.942	0.241	0.206	0.157	0.199	0.206	0.153
16	0.953	0.947	0.935	0.957	0.949	0.946	0.160	0.132	0.092	0.136	0.144	0.093
DGP2												
4	0.953	0.941	0.625	0.947	0.941	0.935	0.839	0.820	0.587	0.800	0.804	0.798
8	0.953	0.942	0.744	0.950	0.944	0.937	0.726	0.689	0.463	0.642	0.641	0.655
12	0.953	0.944	0.846	0.952	0.945	0.941	0.529	0.478	0.335	0.419	0.423	0.418
16	0.953	0.946	0.901	0.953	0.946	0.944	0.340	0.293	0.224	0.261	0.269	0.229
DGP3												
4	0.939	0.896	0.683	0.924	0.918	0.887	0.757	0.684	0.476	0.661	0.649	0.641
8	0.939	0.910	0.832	0.937	0.928	0.907	0.474	0.364	0.277	0.333	0.332	0.301
12	0.939	0.918	0.894	0.945	0.933	0.920	0.262	0.178	0.140	0.178	0.182	0.131
16	0.939	0.926	0.921	0.947	0.936	0.930	0.186	0.116	0.082	0.123	0.128	0.083
DGP4												
4	0.983	0.972	0.762	0.972	0.968	0.971	0.837	0.810	0.598	0.768	0.763	0.784
8	0.983	0.975	0.895	0.976	0.969	0.975	0.590	0.526	0.383	0.448	0.453	0.458
12	0.983	0.977	0.939	0.978	0.970	0.978	0.317	0.265	0.200	0.239	0.250	0.201
16	0.983	0.979	0.955	0.979	0.971	0.982	0.201	0.165	0.117	0.162	0.173	0.116
DGP5												
4	0.957	0.930	0.680	0.946	0.937	0.916	0.792	0.736	0.501	0.713	0.697	0.694
8	0.957	0.936	0.835	0.951	0.942	0.925	0.523	0.426	0.305	0.385	0.380	0.357
12	0.957	0.944	0.885	0.958	0.947	0.937	0.297	0.218	0.158	0.210	0.214	0.167
16	0.957	0.949	0.896	0.962	0.951	0.944	0.204	0.143	0.093	0.146	0.152	0.106
DGP6												
4	0.965	0.913	0.690	0.936	0.922	0.885	0.788	0.699	0.492	0.680	0.659	0.643
8	0.965	0.928	0.837	0.951	0.935	0.911	0.496	0.385	0.282	0.357	0.347	0.318
12	0.965	0.939	0.899	0.962	0.943	0.927	0.279	0.200	0.145	0.198	0.198	0.152
16	0.965	0.948	0.929	0.966	0.948	0.939	0.194	0.134	0.085	0.140	0.142	0.099



(i) sup-type test



(ii) average-type test

Figure 1: Limit Functions of the Test Statistics under the Fixed Alternative ($\lambda_1 = 0.3$, $\underline{\lambda} = 0.05$, $\overline{\lambda} = 0.95$)