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\textsuperscript{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.

\textit{JEL classification:} C12, C22

\textit{Keywords:} coverage rate, break fraction, hypothesis test, average power

\textsuperscript{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 as-
sumptions 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:

\[ y_t = x_t'\beta + x_t^{\lambda_0}\delta + z_t'\gamma + u_t, \quad (t = 1, 2, \ldots, T) \]  

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

\[ y = X\beta + X^{\lambda_0}\delta + Z\gamma + u. \]  

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 \( q_t = [x_t', z_t']' \):

\[ \frac{1}{T} \sum_{t=1}^{[rT]} q_t q_t' \xrightarrow{p} r\Sigma_1 \quad \text{uniformly in} \quad 0 \leq r \leq \lambda_0, \]  

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

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

\[ \frac{1}{\sqrt{T}} \sum_{t=T_0+1}^{[rT]} x_t u_t \Rightarrow \sigma\Sigma_{xx,2}^{1/2}(B(r) - B(\lambda_0)) \quad \text{for} \quad \lambda_0 < r \leq 1, \]
where $\Sigma_1$ and $\Sigma_2$ 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.

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

\[(v) \quad 0 < \underline{\lambda} \leq \lambda_0 \leq \bar{\lambda} < 1.\]

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 $\lambda_1$ is the true break fraction, and consider the following simple testing problem:

\[
H_N : \lambda_0 = \lambda_1 \quad \text{vs.} \quad H_A : \lambda_0 = \lambda_2.
\]

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

\[
y_t = x_t'\beta + x_t'\lambda_1 \delta + z_t'\gamma + u_t
\]

\[
= x_t'\beta + x_t'\lambda_1' \delta + z_t'\gamma + u_t^{\lambda_01}
\]

\[
= w_t^{\lambda_1'} \theta + u_t^{\lambda_01},
\]

where $u_t^{\lambda_01} = u_t + r_t^{\lambda_01} \delta$ with $r_t^{\lambda_01} = x_t^{\lambda_0} - x_t^{\lambda_1}$, $w_t^{\lambda_1} = [x_t', x_t^{\lambda_1'}, z_t']$ and $\theta = [\beta', \delta', \gamma']'$. In vector form,

\[
y = X\beta + X^{\lambda_1} \delta + Z\gamma + u^{\lambda_01}
\]

\[
= W^{\lambda_1} \theta + u^{\lambda_01},
\]

5
where \( W^{\lambda_1} = [X, X^{\lambda_1}, Z] \) and \( u^{\lambda_0} = u + R^{\lambda_0} \delta \) with \( R^{\lambda_0} = X^{\lambda_0} - X^{\lambda_1} \). As the regression model is expressed as (9), we consider the group of transformations given by

\[
y \rightarrow y + Xb + X^{\lambda_1}d + Zg, \quad (\beta, \delta, \gamma) \rightarrow (\beta + b, \delta + d, \gamma + g),
\]

(11)

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

\[
P^{\lambda_1} y | X, Z = P^{\lambda_1} u + P^{\lambda_1} R^{\lambda_0} \delta | X, Z \sim N(P^{\lambda_1} R^{\lambda_0} \delta, \sigma^2 I_{T-2k-p}),
\]

(12)

and we construct an optimal test based on \( P^{\lambda_1} y \). The advantage of considering \( P^{\lambda_1} y \) is that the test statistic becomes invariant to the magnitude of the break given by \( \delta \) 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 \( \varphi \) 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 \( \delta \) and \( \lambda_2 \) given by

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

where \( Q_{\lambda_2}(\delta) \) and \( J(\lambda_2) \) are non-negative measures on \( \mathbb{R}^k \) and \( [\underline{\lambda}, \bar{\lambda}] \), 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 \( \delta \) a normal distribution given by

\[
Q_{\lambda_2}(\delta) \sim N \left( 0, \frac{c \sigma^2}{T \Sigma_{xx,\delta}} \right),
\]

(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 \( P^{\lambda_1}y|X, Z \sim N(0, \sigma^2M - 2K)^{-1} \) under the null hypothesis \( \lambda_0 = \lambda_1 \) and \( N(P^{\lambda_1}R^{\lambda_2} \delta, \sigma^2M - 2K)^{-1} \) under the alternative \( \lambda_0 = \lambda_2 \), we have \( LR_T(\lambda_1, \lambda_2, \delta) = \exp \left( \frac{1}{\sigma^2}y'M_w \lambda_2 \delta - \frac{1}{2\sigma^2}\delta'R^{\lambda_2}M \lambda_2 \delta \right) \). Then, using (13), we can show that

\[
\int LR_T(\lambda_1, \lambda_2, \delta) dQ_\lambda(\delta) = \int (2\pi)^{-k/2} \left| \frac{\sigma^2}{T} \Sigma_{x,x,j} \right|^{-1/2} \exp \left( \frac{1}{\sigma^2}y'M_w \lambda_2 \delta - \frac{1}{2\sigma^2}\delta'R^{\lambda_2}M \lambda_2 \delta - \frac{T}{2\sigma^2}\delta'\Sigma \delta \right) d\delta
\]

\[
= |I_k + \frac{c}{T} \Sigma_{x,x,j} R^{\lambda_2}M \lambda_2 |^{-1/2} \exp \left[ \frac{c}{2\sigma^2}y'M_w \lambda_2 \left( \Sigma_{x,x,j} + \frac{c}{T} R^{\lambda_2}M \lambda_2 \right)^{-1} R^{\lambda_2}M \lambda_2 y \right] ,
\]

(14)

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 \( \tilde{L}R_T(\lambda_1) \).

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

\[
\tilde{L}R_T(\lambda_1) \Rightarrow \tilde{L}R(\lambda_1)
\]

\[
\equiv \int (1 + c \omega(\lambda_1, \lambda_2))^{-k/2} \exp \left[ \frac{c}{2(1 + c \omega(\lambda_1, \lambda_2))} G(\lambda_1, \lambda_2)G(\lambda_1, \lambda_2) \right] dJ(\lambda_2),
\]

(15)

where

\[
\omega(\lambda_1, \lambda_2) = \begin{cases} 
\frac{\lambda_2(1-\lambda_2)}{\lambda_1(1-\lambda_1)} : \lambda_2 < \lambda_1 \\
\frac{(\lambda_2-\lambda_1)(1-\lambda_2)}{1-\lambda_1} : \lambda_1 < \lambda_2,
\end{cases}
\]

\[
G(\lambda_1, \lambda_2) = \begin{cases} 
B(\lambda_1) - B(\lambda_2) - \frac{\lambda_2-\lambda_1}{\lambda_1} B(\lambda_1) : \lambda_2 < \lambda_1 \\
B(\lambda_1) - B(\lambda_2) + \frac{\lambda_2-\lambda_1}{\lambda_1} (B(1) - B(\lambda_1)) : \lambda_1 < \lambda_2.
\end{cases}
\]

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 \( \tilde{T} = [\Delta T] \) to \( \bar{T} = [\bar{\Delta} T] \). 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 \( \tilde{L}R(\lambda_1) \) 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

$$
\lim_{c \to 0} \frac{2}{c} \left( \tilde{L}R(\lambda_1) - 1 \right) = \int G(\lambda_1, \lambda_2)^t G(\lambda_1, \lambda_2) dJ(\lambda_2), \quad (16)
$$

$$
\lim_{c \to \infty} \log(1 + e^{2k/2} \tilde{L}R(\lambda_1)) = \log \int_{\lambda_2 \in \Lambda_{\epsilon}} \omega(\lambda_1, \lambda_2)^{-k/2} \exp \left[ \frac{1}{2\omega(\lambda_1, \lambda_2)} G(\lambda_1, \lambda_2)^t G(\lambda_1, \lambda_2) \right] dJ(\lambda_2), \quad (17)
$$

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

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

$$
\text{avg-LR}(\lambda_1) \equiv \int_{\lambda}^{\lambda} G(\lambda_1, \lambda_2)^t G(\lambda_1, \lambda_2) d\lambda_2, \quad (18)
$$

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

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

$$
\text{sup-LR}(\lambda_1) \equiv \sup_{\lambda_2 \leq \lambda_2 \leq \lambda} G(\lambda_1, \lambda_2)^t G(\lambda_1, \lambda_2). \quad (20)
$$

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 $\lambda_2$ such that
the marginal p values become the same for all possible \( \lambda_2 \). 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 \( \lambda_2 \) and thus we also consider the following weighted sup-type test:
\[ W_{\text{sup-LR}}(\lambda_1) \equiv \sup_{\lambda_2 \in \Lambda} \frac{1}{\omega(\lambda_1, \lambda_2)} G(\lambda_1, \lambda_2)'G(\lambda_1, \lambda_2). \] (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

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

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

where \( \Omega_1 \) and \( \Omega_2 \) 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

\[ F_{T_2}(T_1) = \begin{cases} F_{1,T_2}(T_1) \equiv \frac{1}{T} \left( \sum_{t=T_2+1}^{T} x_t \hat{u}_t \right)' \Omega_1^{-1} \left( \sum_{t=T_2+1}^{T_1} x_t \hat{u}_t \right) : T_2 < T_1 \quad (\lambda_2 < \lambda_1) \\ F_{2,T_2}(T_1) \equiv \frac{1}{T} \left( \sum_{t=T_1+1}^{T} x_t \hat{u}_t \right)' \Omega_2^{-1} \left( \sum_{t=T_1+1}^{T_2} x_t \hat{u}_t \right) : T_1 < T_2 \quad (\lambda_1 < \lambda_2), \end{cases} \] (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 $\Omega_1$ and $\Omega_2$, respectively. The estimation of long-run variance is explained later. Corresponding to $\Lambda_\varepsilon$, we define

$$\Gamma_\varepsilon \equiv \{ T_2 : T \leq T_2 \leq T_1 - T_\varepsilon - 1, T_1 + T_\varepsilon + 1 \leq T_2 \leq T \},$$

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

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

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

$$sup-LR_T(T_1) = \max_{T_2 \leq T_2 \leq T} F_{T_2}(T_1),$$

where $T^* = T - T_1 + 1$ and $T^{**}$ is the number of observations included in $\Gamma_\varepsilon$.

**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 $k, \lambda_1, \lambda_2$, and the trimming parameter $\varepsilon$ but also on the maintained break fraction $\lambda_1$ 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

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

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 $\lambda_1$. 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 $\varepsilon$ as 0.05 and set $\Delta = 1 - \overline{\lambda} = \varepsilon$. For a given value of $\lambda_1$, 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 $\lambda_1$ ranging from $\Delta$ 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 $k$, $\varepsilon$, and the significance level. Next, we derive the response surfaces of the critical values as a function of $\lambda_1$, which, after trying various kinds of functions of $\lambda_1$, 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 $\lambda_2$ within $|\lambda_1 - \lambda_2| \leq \varepsilon$ are trimmed to construct the test statistic, the distribution for a given $\lambda_1$ is rather different depending on whether $\lambda_1$ is in the middle of possible fractions ($\Delta + \varepsilon < \lambda_1 < \overline{\lambda} - \varepsilon$) or close to either end point ($\Delta \leq \lambda_1 \leq \Delta + \varepsilon$ or $\overline{\lambda} - \varepsilon \leq \lambda_1 \leq \overline{\lambda}$). More precisely, when $\Delta \leq \lambda_1 \leq \Delta + \varepsilon$, all $\lambda_2$ between $\Delta$ and $\lambda_1$ are trimmed and the test statistic is constructed using $F_{T_2}(T_1)$ only for $T_2 > T_1$ as follows:

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

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

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

\footnote{In addition, we obtained the critical values for $\varepsilon = 0.1$ but do not report them to save space.}
On the other hand, when \( \lambda + \varepsilon < \lambda_1 < \bar{\lambda} - \varepsilon \), the exponential-type test becomes
\[
\exp\text{-}LR_T(T_1) = \log \left[ \frac{1}{T - T - 2T \varepsilon} \left( \sum_{T_2=T}^{T_1-1} + \sum_{T_2=T_1+1}^{T} \right) \omega(\lambda_1, \lambda_2)^{-k/2} \exp \left( \frac{1}{2\omega(\lambda_1, \lambda_2) F_{T_2}(T_1)} \right) \right].
\]
Therefore, the response surface regression is conducted depending on whether \( \lambda_1 \) is near the boundary, as follows:

\[
eval(\lambda_1|k, \varepsilon = 0.05, \text{significance level}) = I(d < 0.4) \times \left( a_0 + a_{-1} \times \frac{1}{d + 1} + a_1 \times d + a_2 \times d^2 + a_3 \times d^3 \right) + I(d \geq 0.4) \times (b_0 + b_1 \times d + b_2 \times d^2 + b_3 \times d^3)
\]
for the exp\text{-}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 \( \hat{T}_b \).
\[
\hat{T}_b = \arg\min_{T_1} \sum_{t=1}^{T} \left( y_t - x_t' \beta - x^\lambda_t \delta - z_t' \gamma \right)^2.
\]

2. Regress \( y_t \) on \( x_t, x^\lambda_t, x^\hat{\lambda}_t \) and \( z_t \) where \( \hat{\lambda}_b = \hat{T}_b / T \) and denote the regression residuals as \( \tilde{u}_t \). Note that \( x^\hat{\lambda}_t \) is omitted from the regression when \( |T_1 - \hat{T}_b| < k \).
3. Estimate the long-run variance $\Omega_1$ using $X_t \tilde{u}_t$ for $t = 1, \cdots, T_1$. Similarly, estimate $\Omega_2$ 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 $\Omega_1$ and $\Omega_2$, 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 $\sup-LR_T(T_1)$, $\text{avg}-LR_T(T_1)$, or $\exp-LR_T(T_1)$.

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

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

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

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

where for \( \lambda_0 < \lambda_1 \),

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

and for \( \lambda_1 < \lambda_0 \),

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

Note that, as proved by Yamamoto (2014), \( \hat{\Omega}_1 \) and \( \hat{\Omega}_2 \) are consistent estimators of \( \Omega \) (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 \( \lambda_0 \) is smaller than the fraction under the maintained hypothesis \( \lambda_1 \), 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

\[
\frac{1}{T} \sup-LR_T(T_1) \xrightarrow{p} \sup_{\lambda \leq \lambda_2 \leq \lambda} \psi(\lambda_0, \lambda_1, \lambda_2) \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta
\]

\[
= \begin{cases} 
\frac{(\lambda_1-\lambda_0)^2}{\lambda_1^3} \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta &: \lambda_0 < \lambda_1 \\
(\lambda_0-\lambda_1)^2(1-\lambda_0)^2 \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta &: \lambda_0 < \lambda_1 \\
(\lambda_0-\lambda_1)^2(1-\lambda_0)^2(1-\lambda_1) \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta &: \lambda_1 < \lambda_0 
\end{cases}
\]  

(29)

and

\[
\frac{1}{T} \text{avg-LR}_T(T_1) \xrightarrow{p} \int_{\lambda}^{\Lambda} \psi(\lambda_0, \lambda_1, \lambda_2) d\lambda_2 \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta
\]

\[
= \begin{cases} 
\frac{(\lambda_1-\lambda_0)^2}{\lambda_1^3} \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta &: \lambda_0 < \lambda_1 \\
\frac{(\lambda_0-\lambda_1)^2}{\lambda_1^2(1-\lambda_0)^2(1-\lambda_1)} \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta &: \lambda_0 < \lambda_1 \\
\frac{(\lambda_0-\lambda_1)^2}{\lambda_1^2(1-\lambda_0)^2(1-\lambda_1)} \delta' \Sigma_{xx}^{-1} \Sigma_{xx} \delta &: \lambda_1 < \lambda_0 
\end{cases}
\]  

(30)

For example, these functions are shown in Figure 1 as a function of $\lambda_0$ when $\lambda_1 = 0.3$, $\underline{\lambda} = 0.05$ and $\bar{\lambda} = 0.95$. 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 $\lambda_1$. Intuitively, this is because the information before and after the true break is balanced at this point (the information after $\lambda_1$ 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

\[ y_t = 1 + \frac{d}{\sqrt{T}} I(t \geq T_0) + u_t, \]

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 $t \leq T_0$ and $i.i.d. N(0, 4)$ for $t \geq T_0$; 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 $\epsilon_t \sim i.i.d. N(0, 0.49)$. 

15
In DGPs 5 and 6, a model with a partial structural change is investigated as given by

\[ y_t = 1 + x_t + \frac{d}{\sqrt{T}} x_t I(t > T_0) + u_t, \]

where \( x_t = 0.5 x_{t-1} + \xi_t \) with \( \xi_t \sim \text{i.i.d.} \mathcal{N}(0, 0.75) \). DGP5 supposes the spherical distribution as \( u_t \sim \text{i.i.d.} \mathcal{N}(0, 1) \) while conditional heteroskedasticity is allowed in DGP6 such that \( u_t = |x_t| \epsilon_t \) where \( \epsilon_t \sim \text{i.i.d.} \mathcal{N}(0, 0.333) \). Note that these six DGPs are designed so that the long-run variance \( \Omega_1 \) 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 $\lambda_0 = 0$. 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 \( M_{w_1} = M_{\bar{w}_1} \) where \( M_{\bar{w}_1} = \bar{W}^{\lambda_1} (\bar{W}^{\lambda_1} \bar{W}^{\lambda_1})^{-1} \bar{W}^{\lambda_1}' \) with \( \bar{W}^{\lambda_1} = [(X - X^{\lambda_1}), X^{\lambda_1}, Z] \), we have, by the WLLN,

\[
\frac{1}{T} R_{\lambda_2} R_{\lambda_2} = \sum_{t=T_2+1}^{T_1} x_t x_t' \xrightarrow{p} (\lambda_1 - \lambda_2) \Sigma_{xx,1}, \tag{31}
\]

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

\[
\frac{1}{T} \bar{W}^{\lambda_1} \bar{W}^{\lambda_1} \xrightarrow{p} \left[ \begin{array}{ccc}
\lambda_1 \Sigma_{xx,1} & 0 & \lambda_1 \Sigma_{xx,1} \\
0 & (1 - \lambda_1) \Sigma_{xx,2} & (1 - \lambda_1) \Sigma_{xz,2} \\
\lambda_1 \Sigma_{xz,1} & (1 - \lambda_1) \Sigma_{xz,2} & \lambda_1 \Sigma_{zz,1} + (1 - \lambda_1) \Sigma_{zz,2}
\end{array} \right]. \tag{33}
\]

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

\[
\frac{1}{T} R_{\lambda_2} M_{\bar{w}_1} R_{\lambda_2} \xrightarrow{p} \frac{\lambda_2 (\lambda_1 - \lambda_2)}{\lambda_1} \Sigma_{xx,1}, \tag{34}
\]

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

\[
\left| I_k + \frac{c}{T} \Sigma_{xx,1}^{-1} R_{\lambda_2} M_{\bar{w}_1} R_{\lambda_2} \right|^{-1/2} \xrightarrow{p} \left| \left( 1 + \frac{\lambda_2 (\lambda_1 - \lambda_2)}{\lambda_1} I_k \right)^{-1/2} \right| = (1 + c \omega(\lambda_1, \lambda_2))^{-k/2}, \tag{35}
\]

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 $P^{\lambda_{1}'}y = P^{\lambda_{1}'} u$ under the null hypothesis of $\lambda_0 = \lambda_1$ (see also (12)), and thus, $M_{w_1} y = M_{\tilde{w}_1} u$. Then, from (32), (33), and the FCLT, we have

$$
\frac{1}{\sqrt{T}} R^{\lambda_{21}'} M_{w_1} y = \frac{1}{\sqrt{T}} R^{\lambda_{21}'} u - \left( \frac{1}{T} R^{\lambda_{21}' W^{\lambda_1}} \left( \frac{1}{T} W^{\lambda_{21}' W^{\lambda_1}} \right)^{-1} \left( \frac{1}{\sqrt{T}} W^{\lambda_{1}'} u \right) \right)
\Rightarrow \sigma \Sigma_{xx,1}^{1/2} (B(\lambda_1) - B(\lambda_2)) - \frac{\lambda_1 - \lambda_2}{\lambda_1} \sigma \Sigma_{xx,1}^{1/2} B(\lambda_1)
= \sigma \Sigma_{xx,1}^{1/2} G(\lambda_1, \lambda_2),
$$

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

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

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 $\lambda_0 < \lambda_1$,

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

and for $\lambda_1 < \lambda_0$,

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

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

$$
\hat{\theta} = \theta + \left( W^{\lambda_{1}' W^{\lambda_1}} \right)^{-1} W^{\lambda_{1}' u_{\lambda_01}}
= \theta + \left( \frac{1}{T} W^{\lambda_{1}' W^{\lambda_1}} \right)^{-1} \frac{1}{T} W^{\lambda_{1}' R^{\lambda_{01}}} \delta + O_p \left( \frac{1}{\sqrt{T}} \right).
$$
Since

\[ T^{-1} W^{\lambda_1} W^{\lambda_1} \xrightarrow{P} \begin{bmatrix}
\frac{1}{\lambda_1} \Sigma^{-1}_{xx} + \Sigma^{-1}_{xx} \Sigma^{-1}_{zz} \Sigma^{-1}_{xx} - \frac{1}{\lambda_1} \Sigma^{-1}_{xx} \\
- \frac{1}{\lambda_1} \Sigma^{-1}_{xx} \\
- \Sigma^{-1}_{zz} \Sigma^{-1}_{xx} \\
\lambda_1 (1-\lambda_1) \Sigma^{-1}_{xx} \\
0 \\
\Sigma^{-1}_{zz} x
\end{bmatrix}, \]

where \( \Sigma_{zz} = \Sigma_{zz} - \Sigma_{xx} \Sigma^{-1}_{xx} \Sigma_{zz} \), while for \( \lambda_0 < \lambda_1 \),

\[ \frac{1}{T} W^{\lambda_1} R^{\lambda_1} \xrightarrow{P} [(\lambda_1 - \lambda_0) \Sigma_{xx}, 0, (\lambda_1 - \lambda_0) \Sigma'_{xx}]', \]

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

\[ \frac{1}{T} W^{\lambda_1} R^{\lambda_0} \xrightarrow{P} - [(\lambda_0 - \lambda_1) \Sigma_{xx}, (\lambda_0 - \lambda_1) \Sigma_{xx}, (\lambda_0 - \lambda_1) \Sigma'_{xx}]'. \]

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

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

Since \( \hat{u}_t = u_t^{\lambda_0} - w^{\lambda_1}_t (\hat{\theta} - \theta) = u_t + r_t^{\lambda_0} - w^{\lambda_1}_t (\hat{\theta} - \theta) \), we can using Lemma 1 that

\[ \frac{1}{T} \sum_{t=T_2+1}^{T_1} x_t \hat{u}_t = \frac{1}{T} \sum_{t=T_0+1}^{T_1} x_t x'_t \delta - \frac{1}{T} \sum_{t=T_2+1}^{T_1} x_t w^{\lambda_1}_t (\hat{\theta} - \theta) + O_p \left( \frac{1}{\sqrt{T}} \right) \]

\[ \xrightarrow{P} \left( \lambda_1 - \lambda_0 \right) \Sigma_{xx} \delta - \frac{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_0)}{\lambda_1} \Sigma_{xx} \delta \]

\[ = \frac{\lambda_2(\lambda_1 - \lambda_0)}{\lambda_1} \Sigma_{xx} \delta, \]

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

\[ \frac{1}{T} \sum_{t=T_2+1}^{T_1} x_t \hat{u}_t = \frac{1}{T} \sum_{t=T_2+1}^{T_1} x_t x'_t \delta - \frac{1}{T} \sum_{t=T_2+1}^{T_1} x_t w^{\lambda_1}_t (\hat{\theta} - \theta) + O_p \left( \frac{1}{\sqrt{T}} \right) \]

\[ \xrightarrow{P} \left( \lambda_1 - \lambda_2 \right) \Sigma_{xx} \delta - \frac{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_0)}{\lambda_1} \Sigma_{xx} \delta \]

\[ = \frac{\lambda_0(\lambda_1 - \lambda_0)}{\lambda_1} \Sigma_{xx} \delta. \]

On the other hand, when \( \lambda_1 < \lambda_2 \),

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

20
but because \( r_t^{\lambda_0} = 0 \) for \( t = T_1 + 1, \cdots, T_2 \), we can see that

\[
\frac{1}{T} \sum_{t=T_1+1}^{T_2} x_t \hat{u}_t = 0 - \frac{1}{T} \sum_{t=T_1+1}^{T_2} x_t w_{t}^{\lambda_1}(\hat{\theta} - \theta) + O_p \left( \frac{1}{\sqrt{T}} \right) \xrightarrow{p} 0
\]

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


<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1a: Coefficients of the Response Surface Regressions (90 percentiles)
Table 1b: Coefficients of the Response Surface Regressions (95 percentiles)
Table 2a: Coverage rates and lengths of the confidence sets ($T = 100, \lambda_0 = 0.5$)

<table>
<thead>
<tr>
<th>DGP</th>
<th>Coverage rates</th>
<th>Lengths of the confidence sets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>DGP</td>
</tr>
<tr>
<td></td>
<td>d</td>
<td></td>
</tr>
<tr>
<td>DGP1</td>
<td>4</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.961</td>
</tr>
<tr>
<td>DGP2</td>
<td>4</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.961</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.961</td>
</tr>
<tr>
<td>DGP3</td>
<td>4</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.940</td>
</tr>
<tr>
<td>DGP4</td>
<td>4</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.991</td>
</tr>
<tr>
<td>DGP5</td>
<td>4</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.970</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.970</td>
</tr>
<tr>
<td>DGP6</td>
<td>4</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.981</td>
</tr>
<tr>
<td>DGP</td>
<td>Coverage rates</td>
<td>Lengths of the confidence sets</td>
</tr>
<tr>
<td>------</td>
<td>----------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td></td>
<td>d EM MEM Bai sup exp EM MEM Bai sup exp</td>
<td>exp</td>
</tr>
<tr>
<td>DGP1</td>
<td>4</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.955</td>
</tr>
<tr>
<td>DGP2</td>
<td>4</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.955</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.955</td>
</tr>
<tr>
<td>DGP3</td>
<td>4</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.940</td>
</tr>
<tr>
<td>DGP4</td>
<td>4</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.984</td>
</tr>
<tr>
<td>DGP5</td>
<td>4</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.959</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.959</td>
</tr>
<tr>
<td>DGP6</td>
<td>4</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.966</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.966</td>
</tr>
</tbody>
</table>

Table 2b: Coverage rates and lengths of the confidence sets ($T = 300, \Delta = 0.5$)
<table>
<thead>
<tr>
<th>DGP</th>
<th>d</th>
<th>EM</th>
<th>MEM</th>
<th>Bai</th>
<th>sup</th>
<th>avg</th>
<th>exp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0.966</td>
<td>0.919</td>
<td>0.708</td>
<td>0.937</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0</td>
<td>0.966</td>
<td>0.929</td>
<td>0.859</td>
<td>0.950</td>
<td>0.939</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0</td>
<td>0.966</td>
<td>0.939</td>
<td>0.921</td>
<td>0.961</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0</td>
<td>0.966</td>
<td>0.952</td>
<td>0.948</td>
<td>0.966</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0.966</td>
<td>0.931</td>
<td>0.613</td>
<td>0.930</td>
<td>0.920</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0</td>
<td>0.966</td>
<td>0.934</td>
<td>0.742</td>
<td>0.939</td>
<td>0.927</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0</td>
<td>0.966</td>
<td>0.939</td>
<td>0.855</td>
<td>0.945</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0</td>
<td>0.966</td>
<td>0.946</td>
<td>0.906</td>
<td>0.951</td>
<td>0.940</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0.951</td>
<td>0.855</td>
<td>0.669</td>
<td>0.897</td>
<td>0.879</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0</td>
<td>0.951</td>
<td>0.887</td>
<td>0.828</td>
<td>0.927</td>
<td>0.904</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0</td>
<td>0.951</td>
<td>0.911</td>
<td>0.901</td>
<td>0.949</td>
<td>0.922</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0</td>
<td>0.951</td>
<td>0.928</td>
<td>0.940</td>
<td>0.962</td>
<td>0.930</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>0.973</td>
<td>0.868</td>
<td>0.649</td>
<td>0.886</td>
<td>0.891</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0</td>
<td>0.973</td>
<td>0.914</td>
<td>0.797</td>
<td>0.945</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0</td>
<td>0.973</td>
<td>0.934</td>
<td>0.908</td>
<td>0.959</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0</td>
<td>0.973</td>
<td>0.952</td>
<td>0.947</td>
<td>0.973</td>
<td>0.953</td>
</tr>
</tbody>
</table>

Table 3a: Coverage rates and lengths of the confidence sets ($T = 100, \lambda = 0.3$).
| Table 3b: Coverage rates and lengths of the confidence sets (T = 300, λ0 = 0.3) |
|----------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                                | Coverage rates | Lengths of the confidence sets |
|                                | d EM MEM Bai sup avg exp | d EM MEM Bai sup avg exp |
| **DGP1**                        |                 |                               |                 |                 |                 |                 |
| 4                              | 0.953 0.941 0.835 | 0.953 0.835 0.835          | 0.743 0.649 0.649 | 0.520 0.465 0.465 |
| 8                              | 0.956 0.942 0.915 | 0.956 0.915 0.915          | 0.473 0.395 0.395 | 0.376 0.311 0.311 |
| 12                             | 0.958 0.944 0.935 | 0.958 0.935 0.935          | 0.241 0.188 0.188 | 0.199 0.157 0.157 |
| 16                             | 0.960 0.947 0.935 | 0.960 0.935 0.935          | 0.160 0.131 0.131 | 0.136 0.101 0.101 |
| **DGP2**                        |                 |                               |                 |                 |                 |                 |
| 4                              | 0.953 0.942 0.625 | 0.953 0.625 0.625          | 0.839 0.726 0.726 | 0.587 0.463 0.463 |
| 8                              | 0.956 0.944 0.744 | 0.956 0.744 0.744          | 0.726 0.642 0.642 | 0.463 0.383 0.383 |
| 12                             | 0.958 0.946 0.901 | 0.958 0.901 0.901          | 0.529 0.419 0.419 | 0.335 0.251 0.251 |
| 16                             | 0.960 0.949 0.921 | 0.960 0.921 0.921          | 0.340 0.261 0.261 | 0.224 0.158 0.158 |
| **DGP3**                        |                 |                               |                 |                 |                 |                 |
| 4                              | 0.939 0.910 0.683 | 0.939 0.683 0.683          | 0.757 0.684 0.684 | 0.476 0.327 0.327 |
| 8                              | 0.939 0.918 0.832 | 0.939 0.832 0.832          | 0.474 0.364 0.364 | 0.277 0.183 0.183 |
| 12                             | 0.939 0.924 0.894 | 0.939 0.894 0.894          | 0.262 0.200 0.200 | 0.140 0.100 0.100 |
| 16                             | 0.939 0.932 0.921 | 0.939 0.921 0.921          | 0.186 0.131 0.131 | 0.123 0.082 0.082 |
| **DGP4**                        |                 |                               |                 |                 |                 |                 |
| 4                              | 0.983 0.972 0.762 | 0.983 0.762 0.762          | 0.837 0.598 0.598 | 0.598 0.383 0.383 |
| 8                              | 0.983 0.977 0.939 | 0.983 0.939 0.939          | 0.590 0.383 0.383 | 0.383 0.241 0.241 |
| 12                             | 0.983 0.979 0.955 | 0.983 0.955 0.955          | 0.201 0.116 0.116 | 0.116 0.066 0.066 |
| 16                             | 0.983 0.979 0.966 | 0.983 0.966 0.966          | 0.186 0.116 0.116 | 0.116 0.066 0.066 |
| **DGP5**                        |                 |                               |                 |                 |                 |                 |
| 4                              | 0.957 0.937 0.760 | 0.957 0.760 0.760          | 0.839 0.501 0.501 | 0.501 0.305 0.305 |
| 8                              | 0.957 0.936 0.835 | 0.957 0.835 0.835          | 0.426 0.277 0.277 | 0.277 0.183 0.183 |
| 12                             | 0.957 0.944 0.895 | 0.957 0.895 0.895          | 0.218 0.158 0.158 | 0.158 0.100 0.100 |
| 16                             | 0.957 0.949 0.925 | 0.957 0.925 0.925          | 0.143 0.093 0.093 | 0.093 0.056 0.056 |
| **DGP6**                        |                 |                               |                 |                 |                 |                 |
| 4                              | 0.965 0.913 0.770 | 0.965 0.770 0.770          | 0.885 0.592 0.592 | 0.592 0.357 0.357 |
| 8                              | 0.965 0.939 0.909 | 0.965 0.909 0.909          | 0.485 0.292 0.292 | 0.292 0.183 0.183 |
| 12                             | 0.965 0.948 0.929 | 0.965 0.929 0.929          | 0.279 0.191 0.191 | 0.191 0.125 0.125 |
| 16                             | 0.965 0.948 0.939 | 0.965 0.939 0.939          | 0.194 0.134 0.134 | 0.134 0.085 0.085 |
Figure 1: Limit Functions of the Test Statistics under the Fixed Alternative ($\lambda_1 = 0.3$, $\lambda_2 = 0.05$, $\lambda_3 = 0.95$)