

Graduate School of Economics, Hitotsubashi University
Discussion Paper Series No. 2014-08

**A Modified Confidence Set for the Structural Break Date
in Linear Regression Models**

Yohei Yamamoto

May 2014

A Modified Confidence Set for the Structural Break Date in Linear Regression Models*

Yohei Yamamoto[†]

Hitotsubashi University

May 7, 2014

Abstract

Elliott and Müller (2007) (EM) provides a method to construct a confidence set for the structural break date by inverting a locally best test statistic. Previous studies show that the EM method produces a set with an accurate coverage ratio even for a small break, however, the set is often overly lengthy. This study proposes a simple modification to rehabilitate their method. Following the literature, we provide an asymptotic justification for the modified method under a nonlocal asymptotic framework. A Monte Carlo simulation shows that like the original method, the modified method exhibits a coverage ratio that is very close to the nominal level. More importantly, it achieves a much shorter confidence set. Hence, when the break is small, the modified method serves as a better alternative to Bai's (1997) confidence set. We apply these methods to a small level shift in post-1980s Japanese inflation data.

JEL Classification Number: C12, C38

Keywords: confidence set, coverage ratio, nonlocal asymptotics, heteroskedasticity and autocorrelation consistent covariance

*The author acknowledges financial support for this work from the MEXT Grants-in-Aid for Scientific Research No. 25870235.

[†]Hitotsubashi University, Department of Economics, 2-1 Naka, Kunitachi, Tokyo, Japan 186-8601 (yohei.yamamoto@econ.hit-u.ac.jp).

1 Introduction

This study provides a simple modification of the confidence set that was proposed by Elliott and Müller (2007) (EM hereafter) for a single structural break date in linear regression models. EM proposed a method of inverting a test statistic that specifies the null hypothesis of the break occurring at a certain date and the alternative hypothesis of it occurring at another date. The idea is to implement such a test for each possible break date and arrive at the collection of dates where the test fails to reject the null hypothesis; this collection of dates is considered a confidence set. Their method has at least a strong theoretical advantage in the sense that it is based on the locally best test and is considered asymptotically efficient under certain conditions, including the break magnitude shrinking to zero at a fast rate. More importantly, their simulation study suggests that this method exhibits very good coverage properties across a wide range of break magnitudes in finite samples. Especially for small breaks where the standard structural break test is marginally significant or insignificant, the existing method of Bai (1997) is known to exhibit undercoverage, whereas the EM method still achieves a coverage ratio that is very close to the nominal level.

However, previous studies show that the EM confidence set is often much longer than the one computed by other methods¹. For example, in their Table 8 (say, the case of $r_0 = 0.5$ and $d = 16$), the coverage ratio of Bai (1997) is a few percentage points lower, whereas that given by the EM method is very close to the nominal level. As a cost of this advantage, the average length of the EM method is about four times that of Bai's method. More recently, Chang and Perron (2013) investigated the same problem by extending the comparison to the set of linear models including lagged dependent variables. They show that EM's confidence set indeed has a better coverage ratio than other methods but often becomes so long that it includes the whole sample period. More interestingly, they discuss a theoretical base for this phenomenon by using a nonlocal asymptotic framework in which the break does not shrink to zero as the sample size increases. Perron and Yamamoto (2013) and Yamamoto and Tanaka (2013) recently showed that the same technique is useful for analyzing the widely known nonmonotonic power problem in structural break tests. (See also Crainiceanu and Vogelsang 2007 and Perron 1991, 2006 for details of the nonmonotonic power problem.)

Following this literature, we provide an asymptotic justification of our confidence set under the nonlocal asymptotic framework with a fixed break magnitude. The results show

¹Eo and Morley (2013) recently proposed a method that inverts the likelihood ratio test statistic. A comparison with this method is beyond the scope of this paper.

that under the null hypothesis, the modified and original tests are asymptotically equivalent. Under the alternative hypothesis, the modified test is an increasing function in the break magnitude, but the original test is not. If serial correlation in the errors is accounted for by the standard heteroskedasticity and autocorrelation consistent (HAC) correction, then the original test may become a decreasing function. This asymptotic property predicts a higher power of the modified test thus it excludes false time points from the confidence set more efficiently in finite samples. Our Monte Carlo simulation shows that the modified method gives a very good coverage ratio, similar to the original method. More importantly, the average length of the modified method is much smaller than that of the original method over a wide range of break magnitudes. This improvement is significant whether or not the HAC correction is conducted. This analysis has a clear practical implication that the modified confidence set serves as a good alternative to Bai’s method when the break is small. If the break is not small, Bai’s method will achieve good coverage with a shorter length than ours and that of the original method on average, so is still recommended.

The remainder of this paper is structured as follows. Section 2 introduces the model and assumptions. Section 3 explains the EM method to construct the confidence set for a single structural break date. Section 4 proposes a modification of the EM method and investigates its asymptotic property using a nonlocal asymptotic framework. Section 5 provides the results of a Monte Carlo simulation to compare finite sample properties of the new and existing methods. Section 6 illustrates the usefulness of our approach by using an empirical example of post-1980s Japanese inflation data and section 7 presents a conclusion. Throughout the paper, the symbols “ \xrightarrow{p} ” and “ \Rightarrow ” denote convergence in probability and convergence in distribution.

2 Model and assumptions

Consider the linear regression model

$$y_t = I(t \leq T_b^0)X_t'\beta_1 + I(t > T_b^0)X_t'\beta_2 + Z_t'\gamma + u_t, \quad t = 1, \dots, T, \quad (1)$$

where y_t is a scalar variable, X_t is a $k \times 1$ vector of regressors with breaking coefficients β_1 and β_2 at a date $t = T_b^0$ and Z_t is a $p \times 1$ vector of regressors with time-invariant coefficients γ . $I(\cdot)$ is the indicator function. The error term u_t has mean zero. The break magnitude is denoted by $\delta = \beta_2 - \beta_1$. Our goal is to compute a confidence set for the unknown break date T_b^0 at a specified confidence level.

To this end, we introduce the following assumptions.

Assumption 1 i) $T_b^0 = [\lambda_0 T]$, where $0 < \lambda_0 < 1$.

ii) $T^{-1/2} \sum_{t=1}^{[sT]} X_t u_t \Rightarrow \Omega_1^{1/2} W(s)$ for $0 \leq s \leq \lambda_0$ and $T^{-1/2} \sum_{t=T_b^0+1}^{[sT]} X_t u_t \Rightarrow \Omega_2^{1/2} (W(s) - W(r_0))$ for $\lambda_0 \leq s \leq 1$ where Ω_1 and Ω_2 are symmetric and positive definite $k \times k$ long-run variance (LRV) matrices and $W(\cdot)$ is a $k \times 1$ standard Wiener process.

iii) Let $W_t = [X_t, Z_t]$. The matrices $j^{-1} \sum_{t=1}^j W_t W_t'$, $j^{-1} \sum_{t=T-j+1}^T W_t W_t'$, $j^{-1} \sum_{t=T_b^0-j+1}^{T_b^0} W_t W_t'$, and $j^{-1} \sum_{t=T_b^0+1}^{T_b^0+j} W_t W_t'$ have minimum eigenvalues bounded away from zero in probability for all $j \geq p + k$.

iv) $\sup_{0 \leq s \leq 1} \left\| T^{-1/2} \sum_{t=1}^{[sT]} Z_t u_t \right\| = O_p(1)$.

v)

$$T^{-1} \sum_{t=1}^{[sT]} W_t W_t' \xrightarrow{p} sQ_1,$$

uniformly in $0 \leq s \leq \lambda_0$ and

$$T^{-1} \sum_{t=T_b^0+1}^{[sT]} W_t W_t' \xrightarrow{p} (s - \lambda_0)Q_2,$$

uniformly in $\lambda_0 \leq s \leq 1$, where Q_1 and Q_2 are full rank.

Parts i), ii), iv), and v) follow Condition 1 of EM. Part iii) follows Bai (1997) A3 and enables to estimate the break date by

$$\hat{T}_b = \arg \min_{T_b \in [\epsilon T, (1-\epsilon)T]} SSR(T_b), \quad (2)$$

where $SSR(T_b)$ is the sum of squared residuals of the regression of y_t on $\{I(t \leq T_b)X_t, I(t > T_b)X_t, Z_t\}$. The quantity ϵ in (2) is the small trimming value ($\epsilon > 0$). We know that such estimator satisfies consistency for the ratio $\hat{T}_b/T \xrightarrow{p} \lambda_0$ as discussed in Bai (1997) under these assumptions. (See also Lemma 1 in the Appendix.) These assumptions also enable us to use the confidence set of Bai (1997) in the form of $[\hat{T}_b - [c_{(1+\alpha)/2}g] - 1, \hat{T}_b - [c_{(1-\alpha)/2}g] + 1]$ where c_α is the $100 \times \alpha$ percentile of the suggested random variable and $g = \hat{\delta}' \hat{\Omega}_1 \hat{\delta} / (\hat{\delta}' \hat{Q}_1 \hat{\delta})^2$.

3 Elliott and Müller's (2007) confidence set

Earlier studies show that Bai's confidence set may exhibit undercoverage when the break is small. To overcome this problem, EM proposes an alternative method that inverts the test statistic $U_T(T_m)$ specifying the null hypothesis of a break occurring at $t = T_m$ and the alternative hypothesis of one occurring at $t \neq T_m$

$$\begin{aligned} U_T(T_m) &= T_m^{-2} \sum_{t=1}^{T_m} \left(\sum_{s=1}^t v_s \right)' \Omega_1^{-1} \left(\sum_{s=1}^t v_s \right) \\ &\quad + (T - T_m)^{-2} \sum_{t=T_m+1}^T \left(\sum_{s=T_m+1}^t v_s \right)' \Omega_2^{-1} \left(\sum_{s=T_m+1}^t v_s \right), \end{aligned}$$

where $v_t = X_t u_t$ and its LRVs Ω_j ($j = 1, 2$) are replaced by their estimates, respectively, as in the following steps.

1. For any $T_m = p + 2k + 1, \dots, T - p - 2k - 1$, compute the least squares regression of $\{y_t\}_{t=1}^T$ on $\{I(t \leq T_m)X_t, I(t > T_m)X_t, Z_t\}_{t=1}^T$. Call the residuals \hat{u}_t .
2. Construct $\{\hat{v}_t\}_{t=1}^T = \{X_t \hat{u}_t\}_{t=1}^T$.
3. Compute the LRV estimators $\hat{\Omega}_1$ and $\hat{\Omega}_2$ of $\{\hat{v}_t\}_{t=1}^{T_m}$ and $\{\hat{v}_t\}_{t=T_m+1}^T$, respectively².
4. Compute $U_T(T_m)$ with v_t replaced by \hat{v}_t and Ω_1 and Ω_2 by $\hat{\Omega}_1$ and $\hat{\Omega}_2$, respectively.
5. Include T_m in the level α confidence set when $U_T(T_m)$ is less than the critical values provided by EM and exclude it otherwise.

In finite samples, this method indeed achieves a coverage ratio that is very close to the nominal level even when the break is small. However, it is also shown that this method often gives an overly lengthy confidence set. For example, in Table 8 of EM (say, the case of $\lambda_0 = 0.5$ and $d = 16$), the coverage ratio of Bai (1997) is a few percentage points lower than, whereas the EM method is very close to, the nominal level. However, as a cost of this advantage, the average length of the EM method is about four times that of Bai's method. More recently, Chang and Perron (2013) investigated the same problem by extending the comparison to linear models including lagged dependent variables. They show that EM's confidence set indeed has a better coverage ratio than other methods but often becomes so long that it includes the whole sample period.

4 A modified method

This section proposes a simple modification to overcome the drawback of the EM confidence set. The basic idea is to estimate the LRV by using the residuals obtained under the alternative hypothesis, that is, the residuals assuming a break at an unknown point $t \neq T_m$. To do so, we use the unknown break date estimation of Bai (1997) that minimizes the sum of squared residuals over possible unknown break dates. If the true break date is consistently estimated, such residuals are asymptotically free from the presence of a break. These residuals are used to estimate the LRV to construct the test. However, the other part of the test

²An attractive choice is to use the automatic bandwidth estimators of Andrews(1991) or Andrews and Monahan (1992). If it is known that $\Omega_1 = \Omega_2$, then it is advisable to rely instead on a single LRV estimator Ω based on $\{v_t\}_{t=1}^T$. This footnote also applies to the step 3 of the modified procedure in the next section.

is constructed as suggested by EM assuming the break at T_m . Hence, under the alternative hypothesis of $T_m \neq T_b^0$, the true break is present in these residuals, which makes the test statistic large. The following steps describe the algorithm.

1. Estimate a break date \hat{T}_b by minimizing the sum of squares in the regression, that is, (2).
2. For any $T_m = p + 2k + 1, \dots, T - p - 2k - 1$, do the following:
 - (a) Compute the least squares regression of $\{y_t\}_{t=1}^T$ on $\{I(t \leq T_m)X_t, I(t > T_m)X_t, I(t > \hat{T}_b)X_t, Z_t\}$. Call the residuals \tilde{u}_t . Note that for T_m such that $|T_m - \hat{T}_b| < k$, omit $\{I(t > \hat{T}_b)X_t\}_{t=1}^T$ from the regressors. Construct $\{\tilde{v}_t\}_{t=1}^T = \{X_t \tilde{u}_t\}_{t=1}^T$.
 - (b) Compute the least squares regression of $\{y_t\}_{t=1}^T$ on $\{I(t \leq T_m)X_t, I(t > T_m)X_t, Z_t\}_{t=1}^T$. Call the residuals \hat{u}_t . Construct $\{\hat{v}_t\}_{t=1}^T = \{X_t \hat{u}_t\}_{t=1}^T$.
3. Compute the LRV estimators $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ of $\{\tilde{v}_t\}_{t=1}^{T_m}$ and $\{\tilde{v}_t\}_{t=T_m+1}^T$, respectively.
4. Compute $U_T(T_m)$ with v_t replaced by \hat{v}_t . Replace Ω_1 and Ω_2 by $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$, respectively. Call it $V_T(T_m)$.
5. Include T_m in the level α confidence set when $V_T(T_m)$ is less than the critical values provided by EM and exclude it otherwise.

Intuitively, this modification only changes the LRV estimate to another consistent estimate, with which the consistency holds under assumption of the break magnitude that shrinks to zero or not. Therefore, EM's Proposition 3 is still valid and the asymptotic critical values are invariant under the null hypothesis. The following theorem is obtained.

Theorem 1 *Suppose Assumption 1 holds. a) Let v_t be independent and identically distributed (i.i.d.) processes and Ω_j ($j = 1, 2$) estimated under the i.i.d. assumption or b) let v_t be strictly stationary processes and Ω_j ($j = 1, 2$) estimated using the HAC estimator. Then, $U_T(T_b^0)$ and $V_T(T_b^0)$ are asymptotically equivalent.*

More importantly, we derive the asymptotic behavior of the tests under the alternative hypothesis. To do so, we follow the literature and use a nonlocal asymptotic framework that assumes the break magnitude δ to be fixed.

Theorem 2 *Suppose Assumption 1 holds. a) Let v_t be i.i.d. processes and Ω_j ($j = 1, 2$) estimated under the i.i.d. assumption. Then, for $T_m \neq T_b^0$,*

$$\begin{aligned} T^{-1}U_T(T_m) &= O_p(1), \\ T^{-1}V_T(T_m) &= O_p(\|\delta\|^2). \end{aligned}$$

b) Let v_t be strictly stationary processes and Ω_j ($j = 1, 2$) estimated using the HAC estimator of Andrews(1991) with the bandwidth chosen by an AR(1) approximation. Then, for $T_m \neq T_b^0$,

$$\begin{aligned} T^{1/\theta-1}U_T(T_m) &= O_p(1), \\ T^{1/\theta-1}V_T(T_m) &= O_p(\|\delta\|^2), \end{aligned}$$

where $\theta = 3$ when the Bartlett kernel is used and $\theta = 5$ when the Quadratic Spectral (or Parzen or Tukey-Hanning) kernel is used.

Theorem 2 gives clear guidance for predicting the behavior of the tests U_T and V_T under the alternative hypothesis. What is important is that the original test U_T is a nonincreasing function in $\|\delta\|$. On the other hand, the modified test V_T is increasing in $\|\delta\|$ whether or not the HAC correction is conducted. This result predicts that in finite samples the modified test has a higher power and excludes false time points $t \neq T_b^0$ from the confidence set more efficiently.

Remark 1 *If the regressor is a constant, then Condition A1 in the Appendix is satisfied. In this case, the selection of bandwidth for constructing the standard nonparametric HAC covariance estimate is affected by the break magnitude, even asymptotically. Suppose that we use the popular Andrews' (1991) AR(1) approximation. Then, the AR(1) coefficient estimate $\hat{\rho}$ approaches one as the break magnitude increases. Given that the rule is $\hat{m} \propto (\alpha T)^{1/\theta}$, where $\alpha = 4\hat{\rho}^2/(1 - \hat{\rho})^4$ and $\theta = 5$ in the case of the Quadratic Spectral kernel, and $\hat{\rho}$ is the OLS estimate for the AR(1) coefficient of \hat{v}_t , the bandwidth is $\hat{m} = O_p(\|\delta\|^{4/\theta} T^{1/\theta})$. This changes the above results for case b) to*

$$T^{1/\theta-1}U_T(T_m) = O_p(\|\delta\|^{-4/\theta}),$$

so that U_T is indeed a decreasing function in $\|\delta\|$. This result closely follows Chang and Perron's (2013) Theorem 1 and applies to the other kernels proposed in Andrews (1991) .

5 Monte Carlo simulation

This section investigates the finite sample property of the proposed confidence set as a comparison of the two existing methods. To this end, we replicate EM's simulation design and consider the following data generating processes:

Model 1:

$$y_t = \beta + dT^{-1/2}I(t > [\lambda_0 T]) + u_t,$$

where $u_t \sim i.i.d.N(0, 1)$.

Model 2:

$$\begin{aligned} y_t &= \beta + dT^{-1/2}I(t > [\lambda_0 T]) + u_t, \\ u_t &= (1 + I(t > [\lambda_0 T])) \varepsilon_t, \end{aligned}$$

where $\varepsilon_t \sim i.i.d.N(0, 1)$.

Model 3:

$$\begin{aligned} y_t &= \beta + dT^{-1/2}I(t > [\lambda_0 T]) + u_t, \\ u_t &= 0.3u_{t-1} + \varepsilon_t, \end{aligned}$$

where $\varepsilon_t \sim i.i.d.N(0, 1)$.

Model 4:

$$\begin{aligned} y_t &= \beta + dT^{-1/2}I(t > [\lambda_0 T]) + u_t, \\ u_t &= \varepsilon_t - 0.3\varepsilon_{t-1}, \end{aligned}$$

where $\varepsilon_t \sim i.i.d.N(0, 2.04)$.

Model 5:

$$\begin{aligned} y_t &= \gamma + x_t\beta + dT^{-1/2}x_t \cdot I(t > [\lambda_0 T]) + u_t, \\ x_t &= 0.5x_{t-1} + \xi_t, \end{aligned}$$

where $\xi_t \sim i.i.d.N(0, 0.75)$.

Model 6:

$$\begin{aligned} y_t &= \gamma + x_t\beta + dT^{-1/2}x_t \cdot I(t > [\lambda_0 T]) + u_t, \\ x_t &= 0.5x_{t-1} + \xi_t, \\ u_t &= \varepsilon_t |x_t|, \end{aligned}$$

where $\xi_t \sim i.i.d.N(0,0.75)$ and $\varepsilon_t \sim i.i.d.N(0,0.333)$.

In every model, we consider $d = [4, 8, 12, 16]$ and $\lambda_0 = [0.5, 0.35, 0.2]$ and set $T = 100$. The following confidence sets are considered i) EM's method assuming equal LRVs in the two regimes (denoted by $U_{.eq}$), ii) EM's method assuming unequal LRVs ($U_{.neq}$), iii) the modified method assuming equal LRVs ($V_{.eq}$), iv) the modified method assuming unequal LRVs ($V_{.neq}$), v) Bai's method assuming equal LRVs ($Bai_{.heq}$), and vi) Bai's method assuming unequal LRVs ($Bai_{.hneq}$).

We report the coverage ratio and average length of the confidence set evaluated by 3,000 replications in Tables 1 through 6 for Models 1 through 6, respectively. These tables exactly correspond to those reported by EM. Since there is no serial correlation in the errors in Models 1, 2, 5, and 6, White's (1980) heteroskedasticity robust covariance is used. For Models 3 and 4, we use the HAC covariance of Andrews(1991) using Quadratic Spectral kernel with the bandwidth selected by an AR(1) approximation of a prewhitened series of Andrews and Monahan (1992) to account for serial correlation in the errors³. We preliminarily point out that the U and Bai confidence sets in Tables 1 through 6 almost completely replicate EM's results. There is a minor difference, because the extent to which Bai undercovers the nominal level in the case of $d = 4$ is somewhat smaller in our simulation than in EM's simulation. However, we can confirm that the EM method gives a very good coverage ratio, whereas Bai shows undercoverage when the break is small ($d = 4$ and 8). In addition, the average length of U is more than Bai in every case as pointed out by earlier studies.

The main findings of this simulation are summarized by the following two points. First, the modified method V gives a very good coverage ratio and is very similar to the original method U . We see a slight undercoverage only in DGP3 with a very small break ($d = 4$), however, the deviation is minor and rare. Furthermore, it quickly achieves the correct nominal level when the break is as large as $d = 8$. Second and more importantly, the average length of V is much smaller than that of U over the range of break magnitudes. However, as the theory suggests, the advantage of V over U is intensified when the break gets larger and/or the HAC correction is conducted. For example, the difference in average lengths is not very large in the case of $d = 4$ in Model 2. However, the average length of V is five times less than that of U in the case of $d = 16$ in Model 3. This confirms Theorem 2.

We also find that in all cases the average length of V falls between U and Bai . This

³In practice, this information is not available and serial correlation in the errors is to be tested prior to form a confidence set. From this point of view, we also implemented the test for which the HAC correction is applied only when the pre-test rejects. However, the results are very close to what are reported here so is not separately reported.

implies that as long as the break is not small, Bai’s method gives good coverage as well as the shortest confidence set, so it is recommended. When the break is small, Bai’s method still gives a short confidence set, but at the risk of undercoverage. Hence, the modified confidence set proposed in this paper must be an attractive alternative method.

6 Empirical illustration : Level shift in the post-1980s Japanese inflation rate

After a long recession, Japanese inflation dynamics have now been attracting much attention. As is common to most developed countries, Japan experienced high inflation period in the 1970s, recording two-digit annual inflation rates in terms of the core consumer price index (CPI). The inflationary pressure came down as late as the 1980s and inflation became fairly stable since the middle of that decade. Of interest is that Japanese economy experienced a strong deflationary pressure in the late 1990s (See De Veirman 2009 and the references therein). The source of this deflation is a controversial issue and detailed investigation is out of the scope of this paper. However, identifying the dates of level shifts should help one understand the major cause.

We use the quarterly annualized Japanese inflation rate $Y_t = 4 \times (\log(P_t) - \log(P_{t-1}))$, where P_t is the core CPI (seasonally adjusted). We take the sample of a relatively stable inflation rate from 1985Q1 through 2013Q4; the series is presented in Figure 1. We are interested in determining whether there is a break in the mean of Y_t , and if so, when it occurred. To see this, we first apply the $\text{Sup}F$ structural change test for the parameter β in the regression model $Y_t = \beta + u_t$. We use the HAC covariance of Andrews(1991) using Quadratic Spectral kernel with the bandwidth chosen by an AR(1) approximation. The $\text{Sup}F$ test gives a value of 18.10, suggesting that there is a break at 1% significance level. We then use the sequential method of Bai and Perron (1998) to see if there is a second significant break. We find that the $\text{Sup}F(2|1)$ test is 3.96 and is insignificant even at the 10% level. We conclude that there is one break in the mean of inflation during this period.

We now compute the confidence intervals by using the EM method (U), the modified method (V), and Bai’s method (Bai). Figure 2 illustrates the 90% confidence set in black and the 95% confidence set in black and gray. The 90% confidence set of the original EM method covers the period from 1993Q2 through 2001Q3, however, the 95% set includes almost the entire sample period and is thus not very informative. However, Bai gives a set centered at 1998Q1 and the 90% and 95% sets span from 1995Q1 through 2001Q2 and 1993Q4 through 2002Q3, respectively. Our method provides a set that covers a time period somewhat earlier than Bai from 1993Q3 through 2000Q2 for 90% and 1993Q1 through

2000Q4 for 95%. The lengths of V and Bai are very similar or V is somewhat shorter in this example. Given that the break is small and Bai's method may be subject to undercoverage, we prioritize the result obtained by our modified method, followed by Bai and U in that order. This example clearly illustrates the usefulness of the proposed method.

7 Conclusion

This paper provides a simple modification of the confidence set of the single break date in linear regression models proposed by Elliott and Müller (2007). The method involves a step that estimates an unknown break point to obtain the residuals under the alternative hypothesis. These are then used to construct the LRV estimate for the test statistic. Following the literature, an asymptotic justification of the proposed method is provided using the nonlocal fixed break asymptotic framework. Our Monte Carlo simulation shows that the modified method gives an equally correct coverage ratio but a significantly shorter confidence set than the original approach. This method is a good alternative to the confidence set proposed by Bai (1997) when the break is small, because it is known that the Bai's method may result in undercoverage in such cases.

Appendix : Proof of Theorems

Consider the LRV estimators

$$\hat{\Omega} = \hat{R}(0) + 2 \sum_{l=1}^{T-1} \kappa(l, \hat{m}) \hat{R}(l) \quad (\text{A.1})$$

and

$$\tilde{\Omega} = \tilde{R}(0) + 2 \sum_{l=1}^{T-1} \kappa(l, \tilde{m}) \tilde{R}(l), \quad (\text{A.2})$$

where $\hat{R}(l)$ and $\tilde{R}(l)$ for $l = 0, 1, \dots$ are the sample autocovariance estimates of $\{\hat{v}_t\}_{t=1}^T$ and $\{\tilde{v}_t\}_{t=1}^T$ of order l , respectively. $\kappa(\cdot, \cdot)$ is a kernel function and \hat{m} and \tilde{m} are the bandwidths. We let \hat{m} and \tilde{m} be selected by the Andrew's (1991) data dependent method with an AR(1) approximation, however, this particular choice does not affect the qualitative results of this paper. For simplicity, we consider the model with $p = 0$ since this part does not affect the final result.

Lemma 1: Suppose δ is a fixed parameter and Assumption 1 holds. Let \hat{T}_b be obtained by (2). Then, $\hat{T}_b - T_b^0 = O_p(\|\delta\|^{-2})$.

Proof of Lemma 1: See Bai (1997), Proposition 1.

Proof of Theorem 1: We consider the case of $T_m = T_b^0$.

$$\hat{v}_t = \begin{cases} X_t u_t - X_t X_t' (\hat{\beta}_1 - \beta_1) & \text{for } t = 1, \dots, T_b^0, \\ X_t u_t - X_t X_t' (\hat{\beta}_2 - \beta_2) & \text{for } t = T_b^0 + 1, \dots, T, \end{cases}$$

where

$$\begin{aligned} \hat{\beta}_1 &= [\sum_1 X_t X_t']^{-1} [\sum_1 X_t Y_t], \\ &= \beta_1 + [\sum_1 X_t X_t']^{-1} [\sum_1 X_t u_t], \\ \hat{\beta}_2 &= [\sum_2 X_t X_t']^{-1} [\sum_2 X_t Y_t], \\ &= \beta_2 + [\sum_2 X_t X_t']^{-1} [\sum_2 X_t u_t], \end{aligned}$$

so that $\hat{\beta}_1 - \beta_1 = o_p(1)$ and $\hat{\beta}_2 - \beta_2 = o_p(1)$ under Assumption 1 ii), respectively. Note that the subscripts 1 and 2 of the summation symbol denote the regimes defined by $t \in [1, T_b^0]$ and $t \in [T_b^0 + 1, T]$ for 1 and 2, respectively. The sample autocovariance of \hat{v}_t in the pre- T_m ($= T_b^0$) period is

$$\begin{aligned} \hat{R}_1(l) &= (T_b^0 - l)^{-1} \sum_1 X_t u_t u_{t-l}' X_{t-l}' \\ &\quad + (\hat{\beta}_1 - \beta_1) [(T_b^0 - l)^{-1} \sum_1 X_t' X_t X_{t-l}' X_{t-l}] (\hat{\beta}_1 - \beta_1)' \\ &\quad + (\hat{\beta}_1 - \beta_1) [(T_b^0 - l)^{-1} \sum_1 X_t' X_t u_{t-l}' X_{t-l}'] \\ &\quad + [(T_b^0 - l)^{-1} \sum_1 X_t u_t X_{t-l}' X_{t-l}] (\hat{\beta}_1 - \beta_1)', \\ &= I + II + IIIa + IIIb. \end{aligned}$$

Term I converges in probability to $R_1(l)$ under the stationary assumption. II , $IIIa$ and $IIIb$ are $o_p(1)$ since $\hat{\beta}_1 - \beta_1 = o_p(1)$. Hence, $\hat{R}_1(l) \xrightarrow{p} R_1(l)$. Similarly,

$$\begin{aligned}\hat{R}_2(l) &= (T - T_b^0 - l)^{-1} \sum_2 X_t u_t u_{t-l} X'_{t-l} \\ &\quad + (\hat{\beta}_2 - \beta_2) [(T - T_b^0 - l)^{-1} \sum_2 X'_t X_t X'_{t-l} X_{t-l}] (\hat{\beta}_2 - \beta_2)' \\ &\quad + (\hat{\beta}_2 - \beta_2) [(T - T_b^0 - l)^{-1} \sum_2 X'_t X_t u_{t-l} X'_{t-l}] \\ &\quad + [(T - T_b^0 - l)^{-1} \sum_2 X_t u_t X'_{t-l} X_{t-l}] (\hat{\beta}_2 - \beta_2)'.\end{aligned}$$

Given that $\hat{\beta}_2 - \beta_2 = o_p(1)$, $\hat{R}_2(l) \xrightarrow{p} R_2(l)$ for $l = 0, 1, 2, \dots$.

Next, suppose we use the modified method. Without loss of generality, let $\hat{T}_b \leq T_b^0$. Then,

$$\tilde{v}_t = \begin{cases} X_t u_t - X_t X'_t (\tilde{\beta}_1 - \beta_1) & \text{for } t = 1, \dots, \hat{T}_b, \\ X_t u_t - X_t X'_t (\tilde{\beta}_2 - \beta_1) & \text{for } t = \hat{T}_b + 1, \dots, T_b^0, \\ X_t u_t - X_t X'_t (\tilde{\beta}_3 - \beta_2) & \text{for } t = T_b^0 + 1, \dots, T, \end{cases}$$

where

$$\begin{aligned}\tilde{\beta}_1 &= [\sum_1 X_t X'_t]^{-1} [\sum_1 X_t Y_t], \\ &= \beta_1 + [\sum_1 X_t X'_t]^{-1} [\sum_1 X_t u_t], \\ \tilde{\beta}_2 &= [\sum_2 X_t X'_t]^{-1} [\sum_2 X_t Y_t], \\ &= \beta_1 + [\sum_2 X_t X'_t]^{-1} [\sum_2 X_t u_t], \\ \tilde{\beta}_3 &= [\sum_3 X_t X'_t]^{-1} [\sum_3 X_t Y_t], \\ &= \beta_2 + [\sum_3 X_t X'_t]^{-1} [\sum_3 X_t u_t].\end{aligned}$$

Note that the subscripts 1, 2, and 3 of the summation symbol denote the regimes $t \in [1, \hat{T}_b]$, $t \in [\hat{T}_b + 1, T_b^0]$, and $t \in [T_b^0 + 1, T]$, respectively. It is straightforward to show that $\tilde{\beta}_1 - \beta_1 = o_p(1)$ and $\tilde{\beta}_3 - \beta_2 = o_p(1)$. In addition, to derive the order of $\tilde{\beta}_2 - \beta_1$, it is reminded that the number of terms in \sum_2 is $O_p(\|\delta\|^{-2})$ by Lemma 1 so that

$$\begin{aligned}\tilde{\beta}_2 - \beta_1 &= [\sum_2 X_t X'_t]^{-1} [\sum_2 X_t u_t], \\ &= [O_p(\|\delta\|^{-2}) \times O_p(1)]^{-1} [O_p(\|\delta\|^{-2}) \times o_p(1)], \\ &= o_p(1).\end{aligned}$$

The sample autocovariances of \tilde{v}_t in the pre- and post- T_m periods are

$$\begin{aligned}
\tilde{R}_1(l) &= (T_b^0 - l)^{-1} \sum_{1,2} X_t u_t u_{t-l} X'_{t-l} \\
&\quad + (\tilde{\beta}_1 - \beta_1) [(T_b^0 - l)^{-1} \sum_1 X'_t X_t X'_{t-l} X_{t-l}] (\tilde{\beta}_1 - \beta_1)' \\
&\quad + (\tilde{\beta}_2 - \beta_1) [(T_b^0 - l)^{-1} \sum_2 X'_t X_t X'_{t-l} X_{t-l}] (\tilde{\beta}_2 - \beta_1)' \\
&\quad + (\tilde{\beta}_1 - \beta_1) [(T_b^0 - l)^{-1} \sum_1 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T_b^0 - l)^{-1} \sum_1 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_1 - \beta_1)' \\
&\quad + (\tilde{\beta}_2 - \beta_1) [(T_b^0 - l)^{-1} \sum_2 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T_b^0 - l)^{-1} \sum_2 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_2 - \beta_1)', \\
&= I + II + III + IVa + IVb + Va + Vb.
\end{aligned}$$

We can show that term $II = o_p(1) \times O_p(1) \times o_p(1) = o_p(1)$, IVa and IVb are $o_p(1) \times O_p(1) = o_p(1)$. Terms III , Va , and Vb are $O_p(T^{-1}) \times o_p(1) = o_p(T^{-1})$. Hence, $\tilde{R}_1(l) \xrightarrow{p} R_1(l)$. Similarly,

$$\begin{aligned}
\tilde{R}_2(l) &= (T - T_b^0 - l)^{-1} \sum_3 X_t u_t u_{t-l} X'_{t-l} \\
&\quad + (\tilde{\beta}_3 - \beta_2) [(T - T_b^0 - l)^{-1} \sum_3 X'_t X_t X'_{t-l} X_{t-l}] (\tilde{\beta}_3 - \beta_2)' \\
&\quad + (\tilde{\beta}_3 - \beta_2) [(T - T_b^0 - l)^{-1} \sum_3 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T - T_b^0 - l)^{-1} \sum_3 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_3 - \beta_2),
\end{aligned}$$

so that given $\tilde{\beta}_3 - \beta_2 = o_p(1)$ the terms except for the first one are $o_p(1)$. Hence, $\tilde{R}_2(l) \xrightarrow{p} R_2(l)$. We now consider the covariance matrix. For part i) (the i.i.d. case), $\Omega_j = R_j(0)$ and the estimators are constructed by $\hat{\Omega}_j = \hat{R}_j(0)$ and $\tilde{\Omega}_j = \tilde{R}_j(0)$. The above results show that $\hat{\Omega}_j \xrightarrow{p} \Omega_j$ and $\tilde{\Omega}_j \xrightarrow{p} \Omega_j$ and Proposition 3 of EM gives the final results. For part ii), the AR(1) coefficient ρ to construct m in $\{\hat{v}_t\}_j$ and $\{\tilde{v}_t\}_j$ is consistently estimated since $R_j(0)$ and $R_j(1)$ are consistently estimated. This means $\hat{\rho} \xrightarrow{p} \rho^*$ and $|\rho^*| < 1$ by stationary assumption. Thus, we obtain $\hat{\Omega}_j \xrightarrow{p} \Omega_j$ and $\tilde{\Omega}_j \xrightarrow{p} \Omega_j$ at the same rate. Finally, Proposition 3 of EM gives the final results.

We now move on to the proof under the alternative hypothesis. Lemma 2 is a preliminary result to derive Theorem 2. Let $\lambda = \lim T_m/T$.

Lemma 2: Suppose that the LRVs Ω_j ($j = 1, 2$) are known. Then, for $T_m \neq T_b^0$,

$$T^{-1}U_T(T_m) \xrightarrow{p} \Gamma_j(\lambda, \lambda_0)\delta'Q_j\Omega_j^{-1}Q_j\delta,$$

with $j = 1$ if $T_b^0 < T_m$ and $j = 2$ if $T_b^0 > T_m$, where

$$\begin{aligned}\Gamma_1(\lambda, \lambda_0) &= \frac{\lambda_0^2(\lambda_0 - \lambda)^2}{3\lambda^3}, \\ \Gamma_2(\lambda, \lambda_0) &= \frac{(1 - \lambda_0)^2(\lambda_0 - \lambda)^2}{3(1 - \lambda)^3}.\end{aligned}$$

Proof of Lemma 2: The test statistic in this case is

$$\begin{aligned}T^{-1}U_T(T_m) &= \left(\frac{T}{T_m}\right)^2 \underbrace{T^{-1} \sum_{t=1}^{T_m} (T^{-1} \sum_{s=1}^t X_s \hat{u}_s)' \Omega_1^{-1} (T^{-1} \sum_{s=1}^t X_s \hat{u}_s)}_A \\ &\quad + \left(\frac{T}{T - T_m}\right)^2 \underbrace{T^{-1} \sum_{t=T_m+1}^T (T^{-1} \sum_{s=T_m+1}^t X_s \hat{u}_s)' \Omega_2^{-1} (T^{-1} \sum_{s=T_m+1}^t X_s \hat{u}_s)}_B.\end{aligned}$$

We first consider the case of $T_b^0 < T_m$. The residuals are

$$\hat{u}_s = \begin{cases} u_s + X_s \beta_1 - X_s \hat{\beta}_1 & \text{for } s = 1, \dots, T_b^0, \\ u_s + X_s \beta_2 - X_s \hat{\beta}_1 & \text{for } s = T_b^0 + 1, \dots, T_m, \\ u_s + X_s \beta_2 - X_s \hat{\beta}_2 & \text{for } s = T_m + 1, \dots, T, \end{cases}$$

where

$$\begin{aligned}\hat{\beta}_1 &= \left(\sum_{t=1}^{T_m} X_t X_t'\right)^{-1} \left(\sum_{t=1}^{T_m} X_t y_t\right), \\ &= \left(\sum_{t=1}^{T_m} X_t X_t'\right)^{-1} \left(\sum_{t=1}^{T_b^0} X_t X_t' \beta_1 + \sum_{t=T_b^0+1}^{T_m} X_t X_t' \beta_2 + \sum_{t=1}^{T_m} X_t u_t\right), \\ &\rightarrow {}_p \frac{\lambda_0}{\lambda} \beta_1 + \frac{\lambda - \lambda_0}{\lambda} \beta_2,\end{aligned}$$

by Assumption 1 v). It is straightforward to show $\hat{\beta}_2 \xrightarrow{p} \beta_2$ so that $B \xrightarrow{p} 0$. This is because the residuals are consistent estimates of the true errors in the second regime and $T^{-1} \sum_{s=1}^t X_s \hat{u}_s = o_p(1)$. Hence, we consider term A . It can be separated into two terms before and after T_b^0 such that

$$\begin{aligned}A &= T^{-1} \sum_{t=1}^{T_b^0} (T^{-1} \sum_{s=1}^t X_s \hat{u}_s)' \Omega_1^{-1} (T^{-1} \sum_{s=1}^t X_s \hat{u}_s) \\ &\quad + T^{-1} \sum_{t=T_b^0+1}^{T_m} (T^{-1} \sum_{s=1}^t X_s \hat{u}_s)' \Omega_1^{-1} (T^{-1} \sum_{s=1}^t X_s \hat{u}_s).\end{aligned}$$

For $t \leq T_b^0$, the component in A has a limit

$$\begin{aligned}T^{-1} \sum_{s=1}^t X_s \hat{u}_s &= T^{-1} \sum_{s=1}^t X_s u_s + T^{-1} \sum_{s=1}^t X_s X_s' \beta_1 - T^{-1} \sum_{s=1}^t X_s X_s' \hat{\beta}_1, \\ &\rightarrow {}_p 0 + r Q_1 \beta_1 - r Q_1 \left[\frac{\lambda_0}{\lambda} \beta_1 + \frac{\lambda - \lambda_0}{\lambda} \beta_2 \right], \\ &= r \frac{\lambda_0 - \lambda}{\lambda} Q_1 \delta,\end{aligned}$$

and for $t > T_b^0$,

$$\begin{aligned}
T^{-1} \sum_{s=1}^t X_s \hat{u}_s &= T^{-1} \sum_{s=1}^{T_b^0} X_s u_s + T^{-1} \sum_{s=1}^{T_b^0} X_s X'_s \beta_1 - T^{-1} \sum_{s=1}^{T_b^0} X_s X'_s \hat{\beta}_1 \\
&\quad + T^{-1} \sum_{s=T_b^0+1}^t X_s u_s + T^{-1} \sum_{s=T_b^0+1}^t X_s X'_s \beta_2 - T^{-1} \sum_{s=T_b^0+1}^t X_s X'_s \hat{\beta}_1, \\
&\rightarrow p0 + \lambda_0 Q_1 \beta_1 + (r - \lambda_0) Q_1 \beta_2 - r Q_1 \left[\frac{\lambda_0}{\lambda} \beta_1 + \frac{\lambda - \lambda_0}{\lambda} \beta_2 \right], \\
&= \left(r \frac{\lambda_0}{\lambda} - \lambda_0 \right) Q_1 \delta,
\end{aligned}$$

so that

$$\begin{aligned}
A &\rightarrow p \frac{1}{\lambda^2} \int_0^{\lambda_0} r^2 \frac{(\lambda_0 - \lambda)^2}{\lambda^2} \delta' Q_1 \Omega_1^{-1} Q_1 \delta dr + \frac{1}{\lambda^2} \int_{\lambda_0}^{\lambda} \left(r \frac{\lambda_0}{\lambda} - \lambda_0 \right)^2 \delta' Q_1 \Omega_1^{-1} Q_1 \delta dr, \\
&= \frac{(\lambda_0 - \lambda)^2 \lambda_0^3}{3\lambda^4} \delta' Q_1 \Omega_1^{-1} Q_1 \delta - \frac{\lambda_0^2 (\lambda_0 - \lambda)^3}{3\lambda^4} \delta' Q_1 \Omega_1^{-1} Q_1 \delta, \\
&= \frac{\lambda_0^2 (\lambda_0 - \lambda)^2}{3\lambda^3} \delta' Q_1 \Omega_1^{-1} Q_1 \delta.
\end{aligned}$$

For the case of $T_b^0 > T_m$, we follow the same process to reach the stated result (however, term B dominates now, instead of term A). Hence, we have omitted this part of proof to avoid the repetition.

Condition A1: Let L_T be the maximum autoregressive lag order of v_t to be used for bandwidth selection. Then, the following hold:

$$\sup_{l \in [0, \dots, L_T]} \sup_{r \in (0, 1)} \left\| T^{-1} \sum_{t=1}^{\lceil rT \rceil} X_t X'_t X_{t-l} X'_{t-l} - r M_l \right\| = o_p(1),$$

with a positive definite matrix M_l and

$$\frac{\delta' M_1 \delta}{\delta' M_0 \delta} \rightarrow 1 \quad \text{as } \|\delta\| \rightarrow \infty.$$

Proof of Theorem 2: We consider the case of $T_m \neq T_b^0$. Without loss of generality, let $T_b^0 < T_m$. Then,

$$\hat{v}_t = \begin{cases} X_t u_t - X_t X'_t (\hat{\beta}_1 - \beta_1) & \text{for } t = 1, \dots, T_b^0, \\ X_t u_t - X_t X'_t (\hat{\beta}_1 - \beta_2) & \text{for } t = T_b^0 + 1, \dots, T_m, \\ X_t u_t - X_t X'_t (\hat{\beta}_2 - \beta_2) & \text{for } t = T_m + 1, \dots, T, \end{cases}$$

where

$$\begin{aligned}
\hat{\beta}_1 &= [\sum_{12} X_t X_t']^{-1} [\sum_{12} X_t Y_t], \\
&= [\sum_{12} X_t X_t']^{-1} [\sum_1 X_t X_t \beta_1 + \sum_2 X_t X_t \beta_2 + \sum_{12} X_t u_t], \\
&\rightarrow_p \frac{\lambda_0}{\lambda} \beta_1 + \frac{\lambda - \lambda_0}{\lambda} \beta_2, \\
\hat{\beta}_2 &= [\sum_3 X_t X_t']^{-1} [\sum_3 X_t Y_t], \\
&= \beta_2 + [\sum_3 X_t X_t']^{-1} [\sum_3 X_t u_t] \xrightarrow{p} \beta_2,
\end{aligned}$$

by using Assumption 1 v). Note that the subscripts 1, 2, and 3 of the summation symbol denote the regimes $t \in [1, T_b^0]$, $t \in [T_b^0 + 1, T_m]$, and $t \in [T_m + 1, T]$, respectively. The sample autocovariance of \hat{v}_t in the pre- T_m period is

$$\begin{aligned}
\hat{R}_1(l) &= (T_m - l)^{-1} \sum_{12} X_t u_t u_{t-l} X_{t-l}' \\
&\quad + (\hat{\beta}_1 - \beta_1) [(T_m - l)^{-1} \sum_1 X_t' X_t X_{t-l}' X_{t-l}] (\hat{\beta}_1 - \beta_1)' \\
&\quad + (\hat{\beta}_1 - \beta_2) [(T_m - l)^{-1} \sum_2 X_t' X_t X_{t-l}' X_{t-l}] (\hat{\beta}_1 - \beta_2)' \\
&\quad + (\hat{\beta}_1 - \beta_1) [(T_m - l)^{-1} \sum_1 X_t' X_t u_{t-l} X_{t-l}'] \\
&\quad + [(T_m - l)^{-1} \sum_1 X_t u_t X_{t-l}' X_{t-l}] (\hat{\beta}_1 - \beta_1)' \\
&\quad + (\hat{\beta}_1 - \beta_2) [(T_m - l)^{-1} \sum_2 X_t' X_t u_{t-l} X_{t-l}'] \\
&\quad + [(T_m - l)^{-1} \sum_2 X_t u_t X_{t-l}' X_{t-l}] (\hat{\beta}_1 - \beta_2)', \\
&= I + II + III + IVa + IVb + Va + Vb. \tag{A.3}
\end{aligned}$$

Again, I converges in probability to $R_1(l)$. It is shown that II and III are $O_p(\|\delta\|^2)$ since $p \lim(\hat{\beta}_1 - \beta_1) = (1 - \frac{\lambda_0}{\lambda})\delta$ and $p \lim(\hat{\beta}_1 - \beta_2) = -\frac{\lambda_0}{\lambda}\delta$. Similarly, IV and V are $O_p(\|\delta\|)$. Hence,

$$\hat{R}_1(l) = R_1(l) + O_p(\|\delta\|^2).$$

The sample autocovariance in the post- T_m period is,

$$\begin{aligned}
\hat{R}_2(l) &= (T - T_m - l)^{-1} \sum_3 X_t u_t u_{t-l} X_{t-l}' \\
&\quad + (\hat{\beta}_2 - \beta_2) [(T - T_m - l)^{-1} \sum_3 X_t' X_t X_{t-l}' X_{t-l}] (\hat{\beta}_2 - \beta_2)' \\
&\quad + (\hat{\beta}_2 - \beta_2) [(T - T_m - l)^{-1} \sum_3 X_t' X_t u_{t-l} X_{t-l}'] \\
&\quad + [(T - T_m - l)^{-1} \sum_3 X_t u_t X_{t-l}' X_{t-l}] (\hat{\beta}_2 - \beta_2)' \xrightarrow{p} R_2(l).
\end{aligned}$$

given $\hat{\beta}_2 \xrightarrow{p} \beta_2$.

We now consider the modified method. We also assume $\hat{T}_b \leq T_b^0$ without loss of generality. Then,

$$\tilde{v}_t = \begin{cases} X_t u_t - X_t X_t' (\tilde{\beta}_1 - \beta_1) & \text{for } t = 1, \dots, \hat{T}_b, \\ X_t u_t - X_t X_t' (\tilde{\beta}_2 - \beta_1) & \text{for } t = \hat{T}_b + 1, \dots, T_b^0, \\ X_t u_t - X_t X_t' (\tilde{\beta}_2 - \beta_2) & \text{for } t = T_b^0 + 1, \dots, T_m, \\ X_t u_t - X_t X_t' (\tilde{\beta}_3 - \beta_2) & \text{for } t = T_m + 1, \dots, T, \end{cases}$$

where

$$\begin{aligned} \tilde{\beta}_1 &= [\sum_1 X_t X_t']^{-1} [\sum_1 X_t Y_t], \\ &= \beta_1 + [\sum_1 X_t X_t']^{-1} [\sum_1 X_t u_t] \xrightarrow{p} \beta_1, \\ \tilde{\beta}_2 &= [\sum_{23} X_t X_t']^{-1} [\sum_{23} X_t Y_t], \\ &= \left[\underbrace{\sum_2 X_t X_t' + \sum_3 X_t X_t'}_{=O_p(\|\delta\|^{-2})} \right]^{-1} \left[\underbrace{\sum_2 X_t X_t' \beta_1 + \sum_3 X_t X_t' \beta_2}_{=O_p(\|\delta\|^{-2})} + \sum_{23} X_t u_t \right] \xrightarrow{p} \beta_2, \\ \tilde{\beta}_3 &= [\sum_3 X_t X_t']^{-1} [\sum_3 X_t Y_t], \\ &= \beta_2 + [\sum_3 X_t X_t']^{-1} [\sum_3 X_t u_t] \xrightarrow{p} \beta_2. \end{aligned}$$

Note that the subscripts 1, 2, 3, and 4 of the summation symbol denote the regimes $t \in [1, \hat{T}_b]$, $t \in [\hat{T}_b + 1, T_b^0]$, $t \in [T_b^0 + 1, T_m]$, and $t \in [T_m + 1, T]$, respectively (Σ_4 appears later). The sample autocovariance of \tilde{v}_t in the pre- T_m period is

$$\begin{aligned}
\tilde{R}_1(l) &= (T_m - l)^{-1} \sum_{123} X_t u_t u_{t-l} X'_{t-l} \\
&\quad + \underbrace{(\tilde{\beta}_1 - \beta_1)}_{=o_p(1)} \underbrace{[(T_m - l)^{-1} \sum_1 X'_t X_t X'_{t-l} X_{t-l}]}_{=O_p(1)} (\tilde{\beta}_1 - \beta_1)' \\
&\quad + \underbrace{(\tilde{\beta}_2 - \beta_1)}_{=O_p(\|\delta\|)} \underbrace{[(T_m - l)^{-1} \sum_2 X'_t X_t X'_{t-l} X_{t-l}]}_{=O_p(T^{-1}\|\delta\|^{-2})} (\tilde{\beta}_2 - \beta_1)' \\
&\quad + \underbrace{(\tilde{\beta}_2 - \beta_2)}_{=o_p(1)} \underbrace{[(T_m - l)^{-1} \sum_3 X'_t X_t X'_{t-l} X_{t-l}]}_{=O_p(1)} (\tilde{\beta}_2 - \beta_2)' \\
&\quad + (\tilde{\beta}_1 - \beta_1) [(T_m - l)^{-1} \sum_1 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T_m - l)^{-1} \sum_1 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_1 - \beta_1)' \\
&\quad + (\tilde{\beta}_2 - \beta_1) [(T_m - l)^{-1} \sum_2 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T_m - l)^{-1} \sum_2 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_2 - \beta_1)' \\
&\quad + (\tilde{\beta}_2 - \beta_2) [(T_m - l)^{-1} \sum_3 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T_m - l)^{-1} \sum_3 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_2 - \beta_2)', \\
&= I + II + III + IV + Va + Vb + VIa + VIb + VIIa + VIIb.
\end{aligned}$$

Term I converges in probability to $R_1(l)$ and II , IV , Va , Vb , $VIIa$, and $VIIb$ are $o_p(1)$. Term III is $O_p(T^{-1})$. Terms VIa and VIb are $O_p(T^{-1} \|\delta\|^{-1})$. Therefore, $\tilde{R}_1(l) \xrightarrow{p} R_1(l)$. Further,

$$\begin{aligned}
\tilde{R}_2(l) &= (T - T_m - l)^{-1} \sum_4 X_t u_t u_{t-l} X'_{t-l} \\
&\quad + (\tilde{\beta}_3 - \beta_2) [(T - T_m - l)^{-1} \sum_4 X'_t X_t X'_{t-l} X_{t-l}] (\tilde{\beta}_3 - \beta_2)' \\
&\quad + (\tilde{\beta}_3 - \beta_2) [(T - T_m - l)^{-1} \sum_4 X'_t X_t u_{t-l} X'_{t-l}] \\
&\quad + [(T - T_m - l)^{-1} \sum_4 X_t u_t X'_{t-l} X_{t-l}] (\tilde{\beta}_3 - \beta_2)' \xrightarrow{p} R_2(l),
\end{aligned}$$

since $\tilde{\beta}_3 \xrightarrow{p} \beta_2$.

We now consider the LRV. For both methods, the LRV estimate in the first regime dominates that in the second, with respect to the break magnitude. (This is without loss of generality, because the LRV estimate in the second regime dominates in turn if we assume $T_m < T_b^0$.) For part a), the estimators in the first regime are $\hat{\Omega}_1 = \hat{R}_1(0) = R_1(0) + O_p(\|\delta\|^2)$ and $\tilde{\Omega}_1 = \tilde{R}_1(0) = R_1(0) + o_p(1)$. Hence, we can use Lemma 2 by replacing the denominator to obtain the results of U_T and V_T , respectively.

We move on to part b). For V_T , the AR(1) coefficients in $\{\tilde{v}_i\}_j$ are consistently estimated, because $R_j(0)$ and $R_j(1)$ are consistently estimated as shown above. Hence, Andrews' (1991)

method is appropriately applied to obtain $\tilde{\Omega}_j \xrightarrow{p} \Omega_j$. We use Lemma 2 by replacing the denominator and obtain the result. For U_T , we invoke the HAC covariance estimator (A.1), where \hat{m} is chosen by Andrews' (1991) AR(1) approximation so that $\hat{m} \propto (\alpha T)^{1/\theta}$ where $\alpha = 4\hat{\rho}^2/(1 - \hat{\rho})^4$ or $\alpha = 4\hat{\rho}^2/(1 - \hat{\rho}^2)^2$ with $\hat{\rho} = \hat{R}_1(1)/\hat{R}_1(0)$. Because the terms that dominate in $\hat{R}_1(0)$ and $\hat{R}_1(1)$ are *II* and *III*, respectively, in (A.3), $\hat{\rho} \xrightarrow{p} \rho^*$ where $|\rho^*| < 1$ and $\hat{m} = O_p(T^{1/\theta})$ if Condition A1 is not satisfied. We combine the fact $\sum_{l=1}^{T-1} \kappa(l, m) = O(m)$ and obtain the following:

$$\begin{aligned}\hat{\Omega} &= \Omega + O_p(\|\delta\|^2) \times O(\hat{m}), \\ &= \Omega + O_p(\|\delta\|^2 T^{1/\theta}).\end{aligned}$$

Finally, replacing Ω_2 in Lemma 2 by this quantity gives the result for U_T .

We also prove the statement in Remark 1. If Condition A1 is satisfied, then $\hat{\rho} = \frac{\hat{R}_1(1)}{\hat{R}_1(0)} \xrightarrow{p} 1$ as $\|\delta\| \rightarrow \infty$. In this case, the bandwidth inflates as the break magnitude increases. If we use the Quadratic Spectral kernel, then $\hat{m} = O_p(\|\delta\|^{4/\theta} T^{1/\theta})$. Hence,

$$\begin{aligned}\hat{\Omega} &= \Omega + O_p(\|\delta\|^2) \times O(\hat{m}), \\ &= \Omega + O_p(\|\delta\|^{2+4/\theta} T^{1/\theta}).\end{aligned}$$

Replacing Ω_2 in Lemma 2 by this quantity confirms the statement in Remark 1.

References

- Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817-858.
- Andrews, D.W.K., and Monahan, J. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953-966.
- Bai, J. (1997). Estimation of a change point in multiple regressions. *Review of Economics and Statistics* 79, 551-563.
- Bai, J., and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47-78.
- Chang, S., and Perron, P. (2013). A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. Working Paper Boston University.
- Crainiceanu, C., and Vogelsang, T. (2007). Nonmonotonic power for tests of mean shift in a time series. *Journal of Statistical Computation and Simulation* 77, 457-476.
- De Veirman, E. (2009). What makes the output-inflation trade-off change? The absence of accelerating deflation in Japan. *Journal of Money, Credit and Banking* 41, 1117-1140.
- Elliott, G., and Müller, U.K. (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141, 1196-1218.
- Eo, Y., and Morley, J. (2013). Likelihood-based confidence sets for the timing of structural breaks. Unpublished Working Paper, University of Sydney.
- Perron, P. (1991). A test for changes in a polynomial trend function for a dynamic time series. Econometric Research Program Research Memorandum No. 363, Princeton University.
- Perron, P. (2006). Dealing with structural breaks, Patterson, K. Ed. Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory. Palgrave Macmillan pp. 278-352.
- Perron, P., and Yamamoto, Y. (2013). On the usefulness or lack thereof of optimality criteria for structural change tests. forthcoming in *Econometric Reviews*.
- Yamamoto, Y., and Tanaka, S. (2013). Testing for factor loading structural change under common breaks. Discussion Paper, Graduate School of Economics, Hitotsubashi University, 2013-17.

Table 1: Empirical coverage ratio and average length of confidence set :

		Model 1							
		coverage ratio				average length			
		4	8	12	16	4	8	12	16
$\lambda_0=0.5$	U.eq	0.95	0.95	0.95	0.96	76.45	42.10	22.41	15.39
	U.neq	0.96	0.96	0.95	0.96	76.92	42.67	23.23	16.20
	V.eq	0.94	0.95	0.94	0.95	73.06	37.69	19.68	13.27
	V.neq	0.94	0.95	0.94	0.95	73.17	37.64	19.66	13.25
	Bai.het	0.82	0.90	0.94	0.96	67.53	33.00	16.31	9.71
	Bai.hneq	0.82	0.89	0.94	0.96	67.86	33.29	16.44	9.80
$\lambda_0=0.35$	U.eq	0.96	0.95	0.95	0.96	78.19	44.46	22.47	15.19
	U.neq	0.96	0.96	0.95	0.96	78.82	45.28	23.33	16.13
	V.eq	0.95	0.94	0.94	0.96	75.02	39.91	19.71	13.09
	V.neq	0.95	0.94	0.94	0.96	75.10	39.77	19.61	13.07
	Bai.het	0.82	0.89	0.94	0.96	67.71	34.05	16.19	9.63
	Bai.hneq	0.81	0.89	0.93	0.96	68.03	34.26	16.32	9.72
$\lambda_0=0.2$	U.eq	0.94	0.95	0.95	0.95	80.38	54.89	27.24	15.60
	U.neq	0.95	0.96	0.96	0.96	81.17	56.75	29.16	16.96
	V.eq	0.93	0.95	0.94	0.95	77.41	49.56	23.22	13.28
	V.neq	0.92	0.94	0.94	0.95	77.18	49.10	22.93	13.15
	Bai.het	0.82	0.91	0.94	0.96	66.07	35.22	16.85	9.76
	Bai.hneq	0.81	0.90	0.93	0.95	66.20	35.28	16.86	9.81

Table 2: Empirical coverage ratio and average length of confidence set :

		Model 2							
		coverage ratio				average length			
		4	8	12	16	4	8	12	16
$\lambda_0=0.5$	U.eq	0.94	0.94	0.93	0.95	83.09	67.49	46.06	29.92
	U.neq	0.96	0.96	0.95	0.96	84.02	67.45	45.47	29.81
	V.eq	0.92	0.93	0.92	0.93	80.70	63.29	41.35	26.33
	V.neq	0.94	0.95	0.94	0.95	81.26	62.88	40.26	25.46
	Bai.het	0.77	0.80	0.85	0.89	75.25	55.93	35.37	22.29
	Bai.hneq	0.82	0.85	0.89	0.92	78.56	58.90	36.98	23.04
$\lambda_0=0.35$	U.eq	0.97	0.96	0.96	0.97	85.71	73.09	52.79	35.04
	U.neq	0.96	0.96	0.95	0.96	85.79	71.25	49.14	31.65
	V.eq	0.96	0.95	0.95	0.96	83.84	69.45	48.02	31.19
	V.neq	0.95	0.94	0.94	0.96	83.44	66.57	43.35	26.83
	Bai.het	0.72	0.79	0.86	0.90	80.32	64.64	41.36	26.04
	Bai.hneq	0.77	0.81	0.87	0.92	80.99	62.21	38.07	23.35
$\lambda_0=0.2$	U.eq	0.97	0.98	0.98	0.98	87.28	80.93	67.06	48.93
	U.neq	0.95	0.96	0.96	0.96	86.88	79.32	63.99	45.32
	V.eq	0.96	0.97	0.97	0.98	85.55	78.16	62.42	43.41
	V.neq	0.93	0.95	0.94	0.95	84.42	74.94	56.82	37.13
	Bai.het	0.68	0.78	0.85	0.91	82.42	71.16	49.09	31.56
	Bai.hneq	0.72	0.79	0.85	0.91	81.05	65.38	41.76	25.59

Table 3: Empirical coverage ratio and average length of confidence set :

		Model 3							
		coverage ratio				average length			
		4	8	12	16	4	8	12	16
$\lambda_0=0.5$	U.eq	0.96	0.97	0.96	0.96	81.67	57.45	37.46	29.30
	U.neq	0.97	0.97	0.97	0.97	81.38	58.51	42.43	37.86
	V.eq	0.91	0.93	0.94	0.95	68.10	34.98	18.43	12.42
	V.neq	0.90	0.93	0.94	0.96	66.74	34.43	18.35	12.47
	Bai.het	0.81	0.89	0.93	0.96	64.47	31.46	15.61	9.30
	Bai.hneq	0.77	0.87	0.92	0.95	65.80	32.99	16.39	9.74
$\lambda_0=0.35$	U.eq	0.96	0.96	0.96	0.96	81.62	58.99	39.03	29.53
	U.neq	0.97	0.97	0.97	0.97	81.77	61.56	46.15	40.50
	V.eq	0.92	0.92	0.94	0.95	68.85	36.64	18.92	12.34
	V.neq	0.90	0.91	0.94	0.95	66.86	35.70	18.64	12.26
	Bai.het	0.80	0.87	0.94	0.96	63.42	32.25	15.77	9.32
	Bai.hneq	0.75	0.85	0.92	0.95	64.20	33.41	16.47	9.70
$\lambda_0=0.2$	U.eq	0.96	0.97	0.97	0.97	84.58	67.57	46.63	34.09
	U.neq	0.97	0.98	0.97	0.97	85.36	72.84	59.92	54.62
	V.eq	0.91	0.93	0.94	0.95	73.52	45.43	21.47	12.90
	V.neq	0.89	0.91	0.92	0.94	70.70	42.84	20.17	12.18
	Bai.het	0.82	0.88	0.93	0.96	64.66	34.05	16.30	9.53
	Bai.hneq	0.77	0.85	0.90	0.95	64.51	34.32	16.37	9.65

Table 4: Empirical coverage ratio and average length of confidence set :

		Model 4							
		coverage ratio				average length			
		4	8	12	16	4	8	12	16
$\lambda_0=0.5$	U.eq	0.97	0.98	0.98	0.97	81.95	54.92	30.70	20.72
	U.neq	0.97	0.98	0.98	0.98	81.70	55.30	32.60	23.16
	V.eq	0.96	0.96	0.97	0.97	76.14	44.60	23.32	15.29
	V.neq	0.95	0.96	0.97	0.97	75.75	44.41	23.37	15.36
	Bai.het	0.82	0.91	0.95	0.96	70.68	38.90	19.45	11.61
	Bai.hneq	0.81	0.90	0.94	0.96	71.85	40.12	19.99	11.92
$\lambda_0=0.35$	U.eq	0.97	0.97	0.97	0.97	82.73	55.97	31.69	20.83
	U.neq	0.98	0.97	0.98	0.98	82.65	56.92	34.09	23.72
	V.eq	0.95	0.96	0.96	0.97	77.06	45.56	23.76	15.27
	V.neq	0.95	0.95	0.96	0.97	76.38	45.19	23.59	15.27
	Bai.het	0.82	0.90	0.95	0.96	69.89	38.44	19.68	11.57
	Bai.hneq	0.81	0.90	0.94	0.96	71.20	39.56	20.22	11.84
$\lambda_0=0.2$	U.eq	0.97	0.97	0.97	0.97	85.71	66.08	38.69	22.90
	U.neq	0.97	0.97	0.97	0.97	86.05	68.57	44.49	29.34
	V.eq	0.95	0.96	0.96	0.96	81.33	55.08	27.28	15.76
	V.neq	0.94	0.95	0.95	0.95	80.18	53.76	26.66	15.45
	Bai.het	0.83	0.92	0.95	0.96	70.56	38.53	19.40	11.83
	Bai.hneq	0.81	0.89	0.94	0.95	71.30	39.49	19.94	12.11

Table 5: Empirical coverage ratio and average length of confidence set :

		Model 5							
		coverage ratio				average length			
		4	8	12	16	4	8	12	16
$\lambda_0=0.5$	U.eq	0.96	0.96	0.96	0.96	80.20	52.78	31.64	22.34
	U.neq	0.96	0.96	0.97	0.97	81.07	53.91	33.51	24.53
	V.eq	0.93	0.94	0.95	0.96	72.80	41.67	22.92	16.34
	V.neq	0.92	0.93	0.94	0.95	71.65	40.53	22.33	15.89
	Bai.het	0.77	0.87	0.89	0.90	62.99	33.21	16.17	9.51
	Bai.hneq	0.74	0.86	0.88	0.89	62.84	33.65	16.32	9.62
$\lambda_0=0.35$	U.eq	0.96	0.95	0.95	0.96	80.75	55.12	32.56	22.37
	U.neq	0.97	0.97	0.96	0.96	81.78	56.90	35.00	24.96
	V.eq	0.93	0.94	0.94	0.96	73.72	43.11	23.50	16.22
	V.neq	0.91	0.93	0.94	0.95	72.57	41.70	22.56	15.65
	Bai.het	0.79	0.86	0.89	0.90	63.28	32.65	16.23	9.48
	Bai.hneq	0.75	0.84	0.88	0.89	62.95	32.96	16.30	9.52
$\lambda_0=0.2$	U.eq	0.95	0.95	0.96	0.96	83.10	63.96	41.20	27.74
	U.neq	0.96	0.97	0.97	0.97	84.80	67.54	46.20	32.93
	V.eq	0.93	0.93	0.95	0.95	77.28	52.52	29.47	18.72
	V.neq	0.91	0.92	0.94	0.95	75.94	50.25	27.62	17.27
	Bai.het	0.79	0.87	0.90	0.90	64.06	35.13	16.79	9.83
	Bai.hneq	0.74	0.84	0.88	0.88	63.00	34.81	16.60	9.69

Table 6: Empirical coverage ratio and average length of confidence set :

		Model 6							
		coverage ratio				average length			
		4	8	12	16	4	8	12	16
$\lambda_0=0.5$	U.eq	0.96	0.96	0.96	0.96	79.19	48.53	28.57	20.51
	U.neq	0.97	0.97	0.97	0.97	79.98	49.07	29.91	22.51
	V.eq	0.91	0.93	0.94	0.96	67.53	36.77	20.65	15.08
	V.neq	0.90	0.92	0.94	0.96	65.68	34.67	19.47	14.40
	Bai.het	0.81	0.89	0.94	0.97	61.83	29.63	14.41	8.50
	Bai.hneq	0.75	0.85	0.92	0.96	62.20	30.04	14.59	8.61
$\lambda_0=0.35$	U.eq	0.96	0.96	0.96	0.96	79.54	50.55	29.25	20.54
	U.neq	0.97	0.97	0.97	0.97	80.62	51.60	30.88	22.42
	V.eq	0.91	0.93	0.94	0.95	68.58	37.67	21.16	15.10
	V.neq	0.90	0.92	0.93	0.95	66.35	34.98	19.63	14.09
	Bai.het	0.82	0.89	0.94	0.97	62.36	29.41	14.33	8.51
	Bai.hneq	0.76	0.86	0.92	0.96	62.44	29.65	14.33	8.50
$\lambda_0=0.2$	U.eq	0.96	0.96	0.96	0.97	82.36	60.61	36.86	24.91
	U.neq	0.98	0.98	0.98	0.98	84.01	63.03	40.46	28.34
	V.eq	0.91	0.93	0.95	0.96	72.81	47.16	26.19	17.26
	V.neq	0.89	0.91	0.93	0.95	70.24	42.51	22.62	14.97
	Bai.het	0.81	0.88	0.94	0.97	63.99	31.88	15.05	8.85
	Bai.hneq	0.75	0.84	0.91	0.96	63.04	31.07	14.48	8.53

Figure 1 Japanese inflation rate (core CPI)

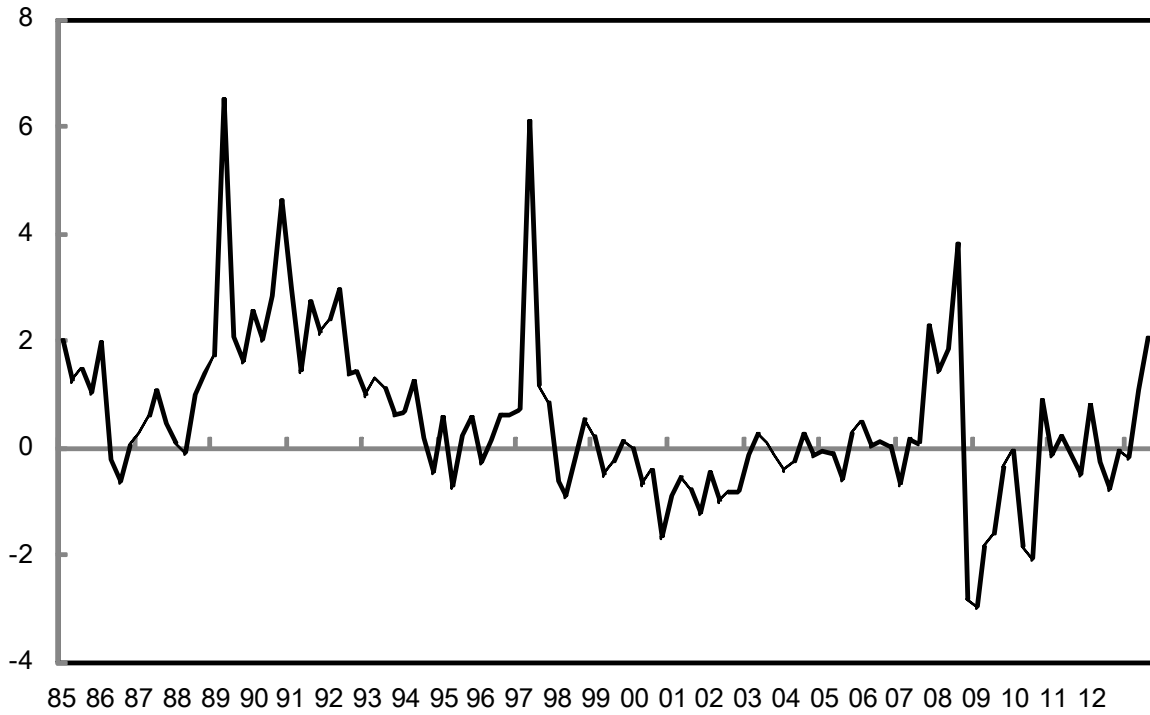
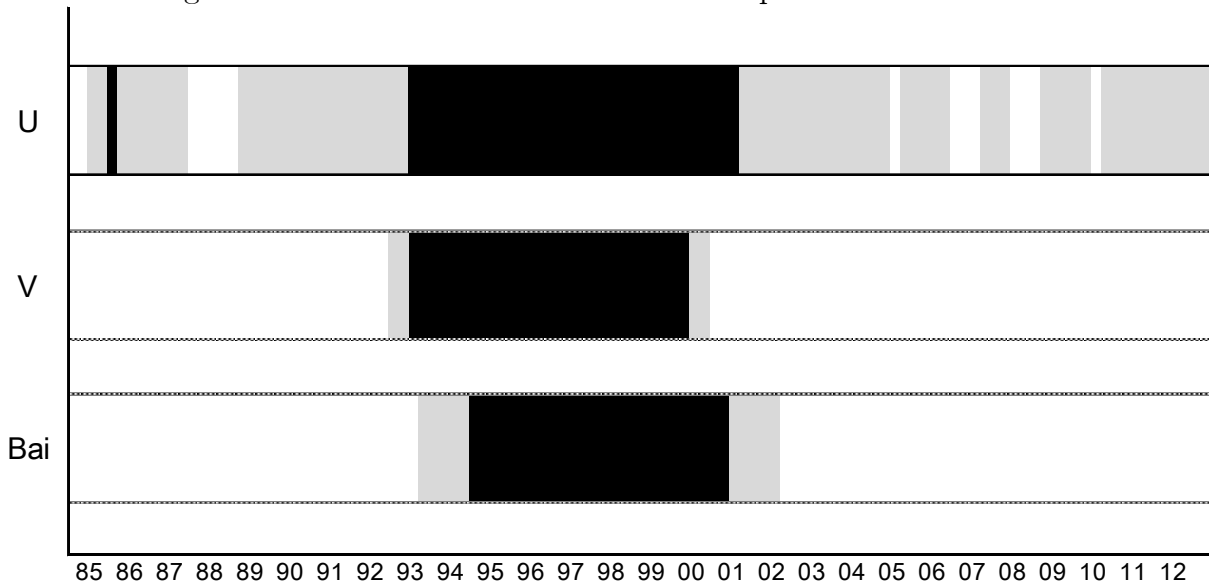


Figure 2 Confidence set of a level shift in Japanese inflation rate



Note: The 90% confidence sets are in black. The 95% confidence sets are in black and gray.