Estimation of a Structural Break Point in Linear Regression Models

Yaein Baek

Korea Institute for International Economic Policy, Sejong, Korea

ABSTRACT

This study proposes a point estimator of the break location for a one-time structural break in linear regression models. If the break magnitude is small, the least-squares estimator of the break date has two modes at the ends of the finite sample period, regardless of the true break location. To solve this problem, I suggest an alternative estimator based on a modification of the least-squares objective function. The modified objective function incorporates estimation uncertainty that varies across potential break dates. The new break point estimator is consistent and has a unimodal finite sample distribution under small break magnitudes. A limit distribution is provided under an in-fill asymptotic framework. Monte Carlo simulation results suggest that the new estimator outperforms the least-squares estimator. I apply the method to estimate the break date in U.S. and U.K. stock return prediction models.

1. Introduction

Researchers in many economic fields extensively address parameter instability in models, which is a common empirical problem in macroeconomics and finance, such as the decrease in output growth volatility in the 1980s, known as “the Great Moderation,” oil-price shocks, labor productivity changes, inflation uncertainty, and stock-return prediction models. It is often reasonable to assume that a change occurs over a long period of time or that some historical event affects the dynamics of a structural model. Hence, the interpretation of structural model dynamics or prediction models relies heavily on the estimation and testing of parameter instability. In econometrics literature, these changes in the underlying data generating process (DGP) of time-series are referenced as a structural break. The timing of the break, as a fraction of the sample size, is called the break point.

Researchers have used estimation methods in the structural break literature to analyze threshold effects and tipping points. Studies on policy change, income inequality dynamics, and social interaction models have used structural break estimation methods. Card et al. (2008) estimate a tipping point of segregation arising in neighborhoods with white preferences. González-Val and Marcén (2012) explore the effect of child custody law reforms and Child Support Enforcement on U.S. divorce rates using the method developed by Bai and Perron (1998, 2003).

Extensive literature describes structural break estimation methods, starting with maximum likelihood estimators (MLE) on break points. Hinkley (1970), Bhattacharya (1987), and Yao (1987) provide an asymptotic theory of the MLE of the break point in a sequence of independent and identically distributed random variables. The asymptotic theory of least-squares (LS) estimation of a one-time break in a linear regression model has been developed by Bai (1994, 1997), with extension to multiple breaks in Bai and Perron (1998) and Bai et al. (1998). The main problem with the LS estimation of the break point is that its finite sample behavior depends on the size of the parameter shift. In many cases, break magnitudes that are empirically relevant are “small” in a statistical sense. For instance, the quarterly U.S. real gross domestic product (GDP) growth rate from 1970Q1 to 2018Q2 has a mean of 0.68 and a standard deviation of 0.8%. A break that decreases the quarterly mean growth rate by 0.25 percentage points is less than half a standard deviation change but is equivalent to a 1 percentage-point decrease in annual growth, which is a significant event for the economy.

In asymptotic analysis, tests have local power against breaks with a magnitude of order $O(T^{-1/2})$. The magnitude represents a small break, shrinking with sample size $T$, so that structural breaks tests have asymptotic power strictly less than one (Elliott and Müller 2007). In the presence of small but detectable breaks, the LS estimator of the break point has a finite sample distribution that exhibits tri-modality with one mode at the true value and two modes at zero and one. Break points at zero or one do not provide any information about a structural break, nor are they likely to be true in practice. Therefore, inference in practical applications based on LS estimation of structural breaks would seem unreliable. Surprisingly, although the methodology is used widely, there are few alternatives for estimating the location of a structural break. Recent literature such as Casini and Perron (2019, 2020) suggest a Laplace-based procedure to provide an estimator of the break point, which is defined by an integration, rather than an optimization-based method.

This study provides an estimator of the structural break point, which is a generalization of LS estimation, and hence, easy to implement in practice. The new estimator resolves the finite...
sample issue of LS estimation; it has a finite sample distribution with a unique mode at the true break and flat tails. This is achieved by imposing weights on the LS objective function. Under small breaks, the LS estimator picks boundaries with high probability due to the functional form of the objective. I construct a weight function of the break point and impose it on the LS objective function to incorporate different estimation uncertainties across potential break points. I provide conditions on the weight function that ensure consistency of the break point estimator and suggest a representative weight function for empirical researchers to use.

The break point estimator is consistent with the same rate of convergence as the LS estimator (Bai 1997) under regularity conditions on the weight functional form in a linear regression model with a structural break on a subset (or all) coefficients. The limit distribution of the break point estimator is derived when the break magnitude is small, under an in-fill asymptotic framework, following the approach of Jiang et al. (2018, 2020). The in-fill asymptotic assumes a fixed time span with shrinking sampling intervals, which leads to a small break magnitude of order $O(T^{-1/2})$. Monte Carlo simulations show that the break point estimator has smaller root mean squared error (RMSE) than the LS estimator in a finite sample under small breaks.

The remainder of this article proceeds as follows: Section 2 constructs the break point estimator for a mean shift in a linear process. Section 3 provides a generalized linear regression framework, following the approach of Jiang et al. (2018, 2020). When the break magnitude is small, under an in-fill asymptotic framework, the limit distribution of the break point estimator is derived following the approach of Elliott and Müller (2007, 2014). I provide theoretical results on the weight function that ensure consistency of the break point estimator. I provide theoretical results in Section 3 under a general linear regression model with multiple regressors, and proves the consistency of the break point estimator. Section 4 presents the in-fill asymptotic theory for stationary and local-to-unit root processes. Monte Carlo simulation results are in Sections 5 and 6 provides an empirical application on the estimation of a structural break on the U.S. and U.K. stock return prediction models. We provide concluding remarks in Section 7. Additional theoretical results and proofs are in the Appendix (supplemental material).

2. Structural Break Point Estimator

In this section, I consider the simplest regression model with a constant term to provide an intuitive explanation of the construction of the break point estimator. I provide theoretical results in Section 3 under a general linear regression model with multiple regressors. Suppose a single break occurs at time $k_0 = [\rho_0 T]$, where $\rho_0 \in (0, 1)$, $[\cdot]$ is the floor function, and $1(t > k_0)$ is an indicator function that equals one if $t > k_0$ and zero otherwise

$$y_t = \mu + \delta 1(t > k_0) + \varepsilon_t, \quad t = 1, \ldots, T. \tag{1}$$

The disturbances $\{\varepsilon_t\}$ are iid with mean zero and $E\varepsilon_t^2 = \sigma^2$. The pre-break mean is $\mu$ and the post-break mean is $\mu + \delta$. Assume we know a one-time break occurs, but the break point $\rho_0$ and parameters $(\mu, \delta, \sigma^2)$ are unknown.

The conventional estimation method of the break location in the literature is least-squares. One obtains the LS estimator by finding a value $k$ that minimizes the objective function $S_T(k)^2$, which is the sum of squared residuals (SSR) under the assumption that $k$ is the break date. $S_T(k)^2 = \sum_{t=1}^{T} (y_t - \bar{y}_k)^2 + \sum_{j=k+1}^{T} (y_t - \bar{y}_k)^2$, where $\bar{y}_k = k^{-1} \sum_{j=1}^{k} y_j$ and $\bar{y}_k = (T-k)^{-1} \sum_{j=k+1}^{T} y_j$ are pre- and post-break LS estimates under break date $k$, respectively. Following Bai’s (1994) expression, I use the identity $\sum_{j=1}^{T} (y_t - \bar{y}_k)^2 = S_T(k)^2 + V_T(k)^2$ (Amemiya 1985, pp. 31–33), where $V_T(k)^2 = (k(T - k)/T) (\bar{y}_k^2 - \bar{y}_k^2)^2$, to substitute for the SSR. Then the LS estimator of the break date is equivalent to

$$\hat{k}_{LS} = \arg\max_{k=1, \ldots, T-1} \{V_T(k)\}, \quad \rho_{LS} = \hat{k}_{LS}/T. \tag{2}$$

An issue with the LS estimator $\hat{\rho}_{LS}$ is that under a small magnitude $|\delta|$, $\hat{\rho}_{LS}$ has a finite distribution that is tri-modal with two modes at the ends of the unit interval and one mode at the true break point $\rho_0$.

A break magnitude that is small is not necessarily small in an economic sense. For example, quarterly U.S. real GDP growth rate from 1970Q1 to 2018Q2 has a mean of 0.68 and a standard deviation of around 0.8%. A break that decreases the mean quarterly growth rate by 0.3 percentage points (a 1.2 percentage point decrease in yearly growth) is a significant event for the economy. Suppose model (1) has parameter values similar to the U.S. real GDP growth rate; assume $\rho_0 = 0.3$, the pre-break mean is $\mu = 0.88\%$ and the shift in the mean of growth rate is $\delta = -0.29$. The expectation of $y_t$ is $\mu + (1 - \rho_0)\delta = 0.68$, which matches the quarterly U.S. real GDP growth rate. Suppose we have $T = 100$ observations and Gaussian disturbances $\varepsilon_t \sim N(0, 0.8^2)$. The left plot of Figure 1 shows the finite sample distribution of the LS estimator of $\rho$ from a Monte Carlo simulation with 2000 replications. The LS estimator fails to accurately detect the break that occurs in the constant term of a univariate linear regression model. Thus, we expect that in practice, structural breaks that are economically important are not large enough for the LS estimator to detect in many cases.

This study focuses on such empirically relevant breaks that are not "large" enough. I follow the approach of Elliott and Müller (2007) to provide an asymptotic approximation to finite sample properties under this small break magnitude. The break magnitude has the same order as sampling uncertainty, $\delta = T^{-1/2}d$, where $d$ is fixed. These asymptotics reflect an important feature of finite sample properties under moderate breaks, because the $p$ values of tests for breaks are typically significant, but not zero. See Elliott and Müller (2007, 2014) for details on the justification of this break magnitude.

Importantly, in literature, it is standard to trim the boundaries of the optimization space so that $\hat{k}_{LS}$ in (2) is the argmax function across $k = [\alpha T], \ldots, [(1-\alpha) T]$ for some $0 < \alpha < 1/2$. Trimming the optimization space may help reduce the build-up mass at the boundaries of the finite sample distribution; however, this has its own drawbacks. Figure 2 shows the finite sample distribution of the LS estimator of $\rho$ with various fractions of trimming, $\alpha \in \{0, 0.1, 0.15, 0.2\}$, from a Monte Carlo simulation with 2000 replications. Under small break magnitudes $\delta = T^{-1/2}d$, $d \in \{2, 4\}$, the modes at the boundaries remain even after trimming. It is unclear if there is a tradeoff between the break size and how large a trimming is needed. With larger trimming ($\alpha = 0.2$), the mass at the boundaries accumulate even more. In addition, there is no reason to believe a break occurs in a restricted period. Thus, we need an alternative method to resolve this issue.

The finite sample distribution of the LS estimator has a build-up mass at boundaries under small break magnitudes because of the way in which the objective function is constructed. For each
potential break date \( k \), the objective function is constructed by partitioning the sample into two subsamples, before and after \( k \). Each subsample is used to estimate two different means, \( \bar{y}_k \) and \( \bar{y}^*_k \). If \( k \) is near 1, the pre-break subsample size \( k \) is small; similarly, if \( k \) is near \( T - 1 \), the post-break subsample size \( T - k \) is small (estimation theory does not require \( \rho \) to be bounded away from zero and one, provided that a change point is assumed to exist). Hence, when the potential break date \( k \) of \( |V_T(k)| \) is near the boundaries, the estimates of the pre- or post-break mean are imprecise because of the small subsample size. Estimation uncertainty at boundaries distorts picking up the true break location if the break magnitude is small relative to the sampling variability (see Appendix C for the description in plot).

Because the issue arises from the large variance of the objective function at the boundaries, one can think of shrinking the variance accordingly. Suppose there are nonnegative "weights" \( \omega_k \) imposed on the LS objective function such that \( k \) with a large estimation error (of pre or postbreak mean) has smaller weights than \( k \) with a small estimation error. When \( k = 1 \) and \( T - 1 \), weights near zero are imposed, which implies that the variance of the weighted objective function \( \omega_k |V_T(k)| \) would shrink toward zero. If the sample period is normalized to a unit interval, the weights are represented by a continuous function on the unit interval, which is zero at \( \{0, 1\} \), and otherwise has positive values.

A new break point estimator is proposed to maximize the value of the objective function \( |Q_T(k)| \), which is equal to the weights \( \omega_k \) multiplied by the LS objective \( |V_T(k)| \).

\[
\hat{k} = \arg \max_{k=1, \ldots, T-1} |Q_T(k)|, \quad \hat{\rho} = \frac{k}{T} \quad (3)
\]

\[
|Q_T(k)| := \omega_k |V_T(k)| = \omega_k \left( \frac{k(T - k)}{T} \right)^{1/2} |\bar{y}^*_k - \bar{y}_k|.
\]

The weight function reduces the variance of the LS objective function when \( k \) is near the boundaries, and thus, the maximizing value \( \hat{k} \) is less likely to pick either end. The right plot of Figure 1 shows the finite sample distribution of the break point estimator (3) under the DGP calibrated from the U.S. real GDP growth rate. As expected, the break point estimator has flat tails at boundaries with a mode at the true break point \( \rho_0 = 0.3 \), whereas the LS estimator has modes of 0.01 and 0.99.

The break point estimator (3) is easy to implement, as we simply modify the objective function by multiplying the weight function. This is a generalization of the LS estimation because...
the LS estimator is a special case, when \( w_k = 1 \). Moreover, by employing the weight function, we no longer need to trim the search grid, because the boundaries have zero weight. Section 3 provides a set of conditions for the weight function that ensures the consistency of the break point estimator under a general linear regression model.

I suggest a representative weight function \( w_k = (k/T(1 - k/T))^{1/2} \) under model (1) (see Section 3 for its analogue under a model with multiple regressors). This function appears in tests that detect a break, as the scaling that self-normalizes the asymptotic distribution of the statistic (see Bai 1994, p.456). There are several interpretations of employing this function as a weight function. First, the weight function is proportional to the inverse of the standard deviation of the break magnitude estimator. Suppose the errors are iid with \( \sigma^2 \), then the variance of the LS estimator of \( \delta \), assuming \( k/T \) is the break point, is equal to \( \sigma^2(k(T - k))/T \). If the estimation uncertainty of the break size is large at a potential break point, small weights are imposed on the corresponding point. Hence, the weight function reflects the estimation uncertainty of the structural break at potential break points.

Second, it is related to the weighting function of Anderson and Darling (1954), which tests whether the sample is drawn from a particular distribution. A nonnegative weight function is chosen to accentuate the boundaries of the sample space, where the test is desired to have sensitivity. If the cdf under the null hypothesis is \( F(\cdot) \), the weight function is \( |F(x)(1 - F(x))|^{-1} \), which increases as \( x \) approaches the boundaries of the sample space. In contrast, I want to downweight the boundaries of the parameter space \( \rho \in [0,1] \) where \( k = [\rho T] \). If \( \rho \) is a random variable with cdf \( F(\rho) \), the weight is the reciprocal of the weight function in Anderson and Darling (1954), \( F(\rho)/|1 - F(\rho)| \). We further assume that \( x \) is uniformly distributed in the unit interval, \( F(x) = \rho \). We obtain the weight function \( \rho(1 - \rho) \), which reduces the variance of the LS objective function \( V_T(k)^2 \) near the boundaries.

Finally, from a Bayesian perspective, the weight function can be interpreted as a prior on parameters \( \delta \) and \( \rho \). We assume the disturbances in (1) are Gaussian, then \( w_k = (k/T(1 - k/T))^{1/2} \) are equivalent to the square root of the Fisher information up to a constant. The Fisher information is interpreted as a way to measure the amount of information the data gives us about the unknown parameter \( \delta \), given \( \rho \). A prior distribution based on the Fisher information reflects our belief that a structural break is less likely to occur near the boundaries. See Appendix A for details on the Bayesian interpretation and other choices of the weight function.

### 3. Partial Break with Multiple Regressors

This section provides consistency of the break point estimator under a general linear regression model with multiple regressors. The model incorporates a partial break in coefficients and assumes that a one-time break occurs at an unknown date \( k_0 = [\rho_0 T] \) with \( \rho_0 \in (0,1) \). I follow the notations of Bai (1997) by denoting the vector of variables associated with a stable coefficient as \( w \) and the variables associated with coefficients under a break as \( z \). Let \( x_t = (w_t, z_t)' \) be a \( (p \times 1) \) vector and \( z_t \) is a \( (q \times 1) \) vector with \( q \leq p \),

\[
y_t = \begin{cases} x_t' \beta + \epsilon_t & \text{if} \ t = 1, \ldots, k_0 \\ x_t' \beta + z_t' \delta \gamma_t + \epsilon_t & \text{if} \ t = k_0 + 1, \ldots, T, \end{cases}
\]

where \( \epsilon_t \) is a mean zero error term. In general, \( z_t \) can be expressed as a linear function of \( x_t \) so that \( z_t = R x_t \), where \( R \) is a \( (p \times q) \) matrix with full column rank. Let \( Y = (y_1, \ldots, y_T)' \) and define \( X_k := (0, \ldots, 0, x_{k+1}, \ldots, x_T)' \) and \( X_0 := (0, \ldots, 0, x_0, \ldots, x_y)' \), so that \( X_0 \) and \( X_0 \) are \((T \times p)\) matrices. Define \( Z_0 \) and \( Z_0 \) analogously so that \( Z_0 = X_k \) and \( Z_0 = X_k \) are \((T \times q)\) matrices. Let \( M := I - X'(X'X)^{-1}X' \) and use the maximal invariant to eliminate the nuisance parameter \( \beta \).

The subscript on \( \delta_T \) shows that the break magnitude may depend on the sample size. Although \( \delta_T \) is referred as the break magnitude, it does not mean \( \delta_T \) has to be positive. We assume the break magnitude is outside the local \( T^{-1/2} \) neighborhood of zero. This is because the break point is not consistently estimable if the break magnitude is in the local \( T^{-1/2} \) neighborhood of zero. This corresponds to the case of small break magnitudes discussed previously, \( \delta_T = O(T^{-1/2}) \), in which structural break tests have asymptotic power strictly less than one. In this section, I proceed by assuming that \( \delta_T \) is fixed or it converges to zero at a rate slower than \( T^{-1/2} \) so that the power of the structural break tests converge to one (Assumption 3).

Let \( S = Y' M Y \), and denote \( S_T(k)^2 \) as the SSR regressing \( MY \) on \( MZ_k \). The LS estimator of break date \( \hat{k}_{LS} \) is the value that minimizes \( S_T(k)^2 \), and thus, maximizes \( V_T(k)^2 \) from the identity \( S = S_T(k)^2 + V_T(k)^2 \) (Amemiya 1985, pp. 31–33),

\[
\hat{k}_{LS} = \arg \max_k V_T(k)^2, \quad \hat{\rho}_{LS} = \hat{k}_{LS}/T
\]

where \( V_T(k)^2 := \hat{\delta}_k'(Z_k' MZ_k) \hat{\delta}_k \) and \( \hat{\delta}_k \) is the LS estimate of \( \delta_T \) assuming that \( k \) is the break date, \( \hat{\delta}_k = (Z_k' MZ_k)^{-1}Z_k' MY \). If \( k = k_0 \), then \( \hat{\delta}_k = \delta_T + (Z_0' MZ_0)^{-1} Z_0' M \eta \). Note that \( V_T(k)^2 \) is nonnegative from the inner product of the vector \((Z_k' MZ_k)^{1/2} \hat{\delta}_k \). The LS objective function is modified by multiplying a \((q \times q)\) positive definite weight matrix \( \Omega_k \), which is a generalization of the square of the weight function, \( \omega_k^2 \) in Section 2 for linear regression models with multiple regressors. Decompose the weight matrix so that \( \Omega_k = \Omega_k^{1/2} \Omega_k^{1/2} \) and multiply \( \Omega_k^{1/2} \) to the vector \((Z_k' MZ_k)^{1/2} \hat{\delta}_k \). Take the inner product and obtain the objective function \( Q_T(k)^2 \), the estimator of the break point is

\[
\hat{k} = \arg \max_k Q_T(k)^2, \quad \hat{\rho} = \hat{k}/T, \quad Q_T(k)^2 := \hat{\delta}_k'(Z_k' MZ_k)^{1/2} \Omega_k (Z_k' MZ_k)^{1/2} \hat{\delta}_k.
\]

An example of the weight matrix is \( \Omega_k = T^{-1} \), which is equal to the square of the representative weight function \( \omega_k^2 = k/T(1 - k/T) \) in model (1) if \( R = I \) and \( X \) is a \((T \times 1)\) vector of ones. Similarly, the matrix \( T^{-1/2} \) decreases as \( k \) approaches either end of the sample from the following rearrangement of terms:

\[
T^{-1} Z_k' MZ_k = T^{-1} R (X_k' X_k) R (X_k' X_k) (X_k' X_k) R.
\]

I prove the consistency of the break point estimator \( \hat{\rho} \) in (5) under regularity conditions on model (4) and weight matrix \( \Omega_k \). The notation \( ||\cdot|| \) denotes the Euclidean norm \( ||x|| = (\sum_{t=1}^p x_t^2)^{1/2} \) for \( x \in \mathbb{R}^p \). For a matrix \( A \), \( ||A|| \) represents the vector induced norm \( ||A|| = \sup_{x} ||Ax||/||x|| \) for \( x \in \mathbb{R}^p \) and \( A \in \mathbb{R}^{p \times p} \).
Assumption 1. 1. $k_0 = [\rho_0 T]$, where $\rho_0 \in [\alpha,1-\alpha]$, $0 < \alpha < \frac{1}{2}$.

2. The data $\{Y_t, X_t, Z_t : 1 \leq t \leq T, T \geq 1\}$ form a triangular array. The subscript $T$ is omitted for simplicity. In addition, $Z_t = R X_t$, where $R$ is $p \times q$, $\text{rank}(R) = q$, $z_t \in \mathbb{R}^q$, $x_t \in \mathbb{R}^p$, and $q \leq p$.

3. The matrices $\left( j^{-1} \sum_{t=1}^{j} x_t x_t^T \right)$, $\left( j^{-1} \sum_{t=-j}^{0} x_t x_t^T \right)$, and $\left( j^{-1} \sum_{t=0}^{1} x_t x_t^T \right)$ have minimum eigenvalues bounded away from zero in probability for all large $j$. For simplicity, assume these matrices are invertible when $j \geq p$. In addition, these four matrices have stochastically bounded norms uniformly in $j$. That is, for example, $\sup_{j \geq 1} \left\| j^{-1} \sum_{t=1}^{j} x_t x_t^T \right\| = O_p(1)$.

4. $T^{-1} \sum_{t=1}^{T} x_t x_t^T \overset{p}{\to} \Sigma_{\infty}$ uniformly in $s \in [0,1]$, where $\Sigma_{\infty}$ is a nonrandom positive definite matrix.

5. For random regressors, $\sup_{t} E \left\| x_t \right\|^{1+\gamma} \leq K$ for some $\gamma > 0$ and $K < \infty$.

6. The disturbance $\epsilon_t$ is independent of the regressor $x_t$ for all $t$ and $s$. For an increasing sequence of $\sigma$-fields $\mathcal{F}_t, \{\epsilon_t, \mathcal{F}_t\}$ form a sequence of $L^r$-mixingale sequence with $r = 4 + \gamma$ for some $\gamma > 0$ (McLeish 1975; Andrews 1988). That is, there exists nonnegative constants $\{c_t : t \geq 1\}$ and $\{\psi_t : t \geq 0\}$, such that $\psi_t \downarrow 0$ as $t \to \infty$ and for all $t \geq 1$ and $j \geq 0$, we have (a) $E \left| E[\epsilon_t|\mathcal{F}_j] \right|^{1+\gamma} \leq c_t^\gamma \psi_t^\gamma$, (b) $E \left| \epsilon_t - E[\epsilon_t|\mathcal{F}_{t+j}] \right|^{1+\gamma} \leq c_t^\gamma \psi_{t+j+1}^\gamma$, (c) $\max_t c_t < K < \infty$, (d) $\sum_t j^{1+\gamma} \psi_t < \infty$ for some $\kappa > 0$.

Assumption 2. $\Omega_k$ is a positive definite $(q \times q)$ matrix ($q = \text{dim}(z_t)$) that is a continuous function of data $\{y_t, x_t, z_t, 1 \leq t \leq T\}$ and have stochastically bounded norms uniformly in $k = 1, \ldots, T - 1$. In addition, for any nonzero vector $c \in \mathbb{R}^q$,

$$\left\| \Omega_k^{1/2} (Z_k^0 M Z_0)^{1/2} c \right\| > \left\| \Omega_k^{1/2} (Z_k^0 M Z_0)^{-1/2} (Z_k^0 M Z_0) c \right\|$$

holds for all $k$ and $k_0$, where $M = I - X(X'X)^{-1}X'$. When $k/T \to \rho$ as $T \to \infty$, then $\Omega_k \overset{p}{\to} \Omega(\rho)$, where $\Omega(\rho)$ is a differentiable function of $\rho$, with $\rho$-wise element.

The conditions of Assumption 1 are similar to those of Assumptions A1 to A6 in Bai (1997), with additional restrictions (iv) and (vi). Assumption 1(vii) allows for a general serial correlation in disturbances and requires $x_t$ to be strictly exogenous. This is because $\Omega_k$ depends on the moments of the regressors, and we want to impose zero weights on $\rho \in [0,1]$. For instance, if the second moments of $z_t$ change at $\rho_0$, the boundaries of the unit interval may have positive weights that depend on the distribution of $z_t$. These cases are avoided under strict exogeneity because $\Omega_k$ converges in probability to a function of $\rho$ and a fixed matrix $\Sigma_x$. Note that if $\Omega_k$ is a nonstochastic matrix that satisfies the norm inequality in Assumption 2, consistency holds under weakly exogenous regressors (see Assumption 4).

Assumption 2 guarantees that the matrix

$$A_T(k) := \frac{1}{|k_0 - k|} \left[ (Z_0^0 M Z_0)^{1/2} \Omega_{k_0} (Z_0^0 M Z_0)^{1/2} \right] - \left( (Z_0^0 M Z_k)^{1/2} \Omega_k (Z_0^0 M Z_k)^{-1/2} \right)$$

is positive definite; hence, $\left\| A_T(k) \right\| \geq \min_{\kappa} (A_T(k)) > 0$, where $\min_{\kappa}$ denotes the minimum eigenvalue of $A_T(k)$. Assumption 2 provides a sufficient condition for $\Omega_k$ to ensure that $Q_T(k)^2$ can only be maximized near $k_0$ with high probability. In the univariate model (1), the condition is equivalent to $|\omega(\rho)/\omega(\rho)| < (2p (1 - \rho))^{-1}$ for all $\rho$, where $\omega(\rho) = \partial \omega(x) / \partial x |_{x=\rho}$.

The slope magnitude of the logarithm of $\omega(\rho)$ has an upper bound that increases as $\rho$ approaches zero or one. For instance, with the weight function $\omega(\rho) = (\rho (1 - \rho))^2$, Assumption 2 implies $-1/(2p (1 - \rho)) < \gamma (1 - 2p) (\rho(1 - \rho)) < 1/(2p (1 - \rho))$. This condition prevents extreme weights on the boundaries and zero weights in the interior space. Note that we assume the break point is bounded away from endpoints such that $\rho_0 \in [\alpha, 1 - \alpha]$ in Assumption 1.

Theorem 1. Under Assumptions 1 and 2, suppose $\delta_T$ is fixed or shrinking toward zero such that Assumption 3 is satisfied. Then, $\hat{\kappa} = k_0 + O_p(\|\delta_T\|^{-2})$ and the break point estimator $\hat{\rho}$ in (5) is consistent

$$\left| \hat{\rho} - \rho_0 \right| = O_p(T^{-1} \|\delta_T\|^{-2}) = o_p(1).$$

See Appendix B for proof of Theorem 1. For weakly exogenous regressors, the break point estimator is consistent with the same rate of convergence in Theorem 1, under the following conditions that substitute Assumptions 1 and 2.

Assumption 4. Assume the following conditions in model (4) with Assumption 1(i)–(iii) and (v):

1. $(X'X)/T$ converges in probability to a nonrandom positive definite matrix, as $T \to \infty$;
2. $\{\epsilon_t, \mathcal{F}_t\}$ form a sequence of martingale differences for $\mathcal{F}_t = \sigma$-field $\{\epsilon_t, x_{t+1} : s \leq t\}$. Moreover, for all $t$, $E[\epsilon_t^{4+\gamma}] < K$ for some $K < \infty$ and $\gamma > 0$;
3. The weight matrix $\Omega_k$ is a nonrandom $(q \times q)$ positive definite matrix, and for any nonzero vector $c \in \mathbb{R}^q$,

$$\left\| \Omega_k^{1/2} (Z_0^0 M Z_0)^{1/2} c \right\| > \left\| \Omega_k^{1/2} (Z_k^0 M Z_k)^{-1/2} (Z_k^0 M Z_0) c \right\|$$

holds for all $k$ and $k_0$, where $M = I - X(X'X)^{-1}X'$. $\Omega_k$ converges to $\Omega(\rho)$ as $k/T \to \infty$, which is a differentiable function of $\rho$ on the unit interval.

Theorem 2. Under Assumption 4, suppose $\delta_T$ is fixed or shrinking toward zero that satisfies $\delta_T \to 0$ and $T^{1/2-\gamma} \delta_T \to \infty$ for some $\gamma \in (0, \frac{1}{2})$. Then, $\hat{\kappa} = k_0 + O_p(\|\delta_T\|^{-2})$ and the break point estimator $\hat{\rho}$ in (5) is consistent

$$\left| \hat{\rho} - \rho_0 \right| = O_p(T^{-1} \|\delta_T\|^{-2}) = o_p(1).$$
The proof of Theorem 2 is similar to the proof of Theorem 1; hence, we have omitted it. Under Assumptions 1(i), 4(i), and 4(ii), the strong law of large numbers holds for \( x_t \varepsilon_t \) because the conditions in Hansen (1991) are satisfied. The weight matrix \( \Omega_k \) in Assumption 4(iii) depends on \( k/T \) but not on the data \( \{x_t, \varepsilon_t\} \). Thus, by setting \( \rho = k/T \), \( \Omega_k \) is a function of \( \rho \), which is assumed to be differentiable with respect to \( \rho \). Then, for some finite \( c > 0 \), the bound \( \|\Omega_{k_1} - \Omega_{k_2}\| \leq c|k_1 - k_2|/T \) holds for any \( k_1 \) and \( k_2 \). Using these properties, proving the consistency of the estimator under Assumption 4 is the same process as in the proof of Assumptions 1 and 2.

Given the consistency of the breakpoint estimator from Theorems 1 or 2, the estimator of the break magnitude corresponding to \( \hat{k} \) is consistent and asymptotically normally distributed. Let \( \hat{\delta}(\hat{\rho}) = \delta_{\hat{k}} \), then the resulting forms hold. See Appendix B for the proof.

**Corollary 1.** Under Assumptions 1 and 2, suppose \( \delta_T \) is fixed or shrinking toward zero such that Assumption 3 is satisfied. Let \( \hat{\delta}(\hat{\rho}) \) be a consistent estimator of \( \delta_T \) corresponding to \( \hat{k} \), which is defined in (5), and \( Q_0 = (X Z_0) \). Then,

\[
\sqrt{T} \left[ \frac{\hat{\delta} - \delta}{\hat{\delta}(\hat{\rho}) - \delta_T} \right] \xrightarrow{d} N \left( 0, V^{-1} U U^{-1} \right),
\]

where

\[
V := \text{plim}_{T \to \infty} T^{-1} Q_0' Q_0, \quad U := \lim_{T \to \infty} E \left[ \left( T^{-1/2} Q_0' \varepsilon \right)^2 \right].
\]

**4. Asymptotic Distribution**

Bai (1997) provides the limit distribution of the LS estimator assuming large breaks \( \delta = O(T^{-1/2+\epsilon}) \) with \( 0 < \epsilon < 1/2 \). The asymptotic distribution is symmetric at the true breakpoint, if the second moment of variables associated with coefficients under break \( (\varepsilon_t \text{ in Section 3}) \) do not change before and after the break. The limiting distribution has been developed under the long-span asymptotic framework, in which the time span of data is assumed to go to infinity. For the autoregressive model, the statistical inference about the breakpoint is complicated because the long-span asymptotic theory results in a discontinuous distribution that depends on the coefficient before and after the break; see Chong (2001), Pang et al. (2014), and Pang et al. (2018) on the long-span asymptotic distribution of the LS estimator under different settings of the AR root.

In this study we follow the approach of Jiang et al. (2018, 2020) to derive the limit distribution of the breakpoint estimator under a stationary and local-to-uniform autoregressive process. Jiang et al. (2018, 2020) and Casini and Perron (2019) employed a continuous record asymptotic framework to derive the limit distribution of the breakpoint estimator. The in-fill asymptotic assumes a fixed time span with shrinking sampling intervals. This leads to a break magnitude of order \( O(T^{-1/2}) \), which is our primary interest. Furthermore, for the AR model, the in-fill asymptotic distribution is continuous in the underlying parameters and is dependent on the initial condition, in contrast to the discontinuity in the long-span asymptotic distribution. See Jiang et al. (2020) for simulation studies which show that the in-fill asymptotic distribution provide better approximations to the finite sample distribution compared to the long-span asymptotic distribution.

**4.1. Partial Break in a Stationary Process**

Consider the linear regression model (4) with continuous time process \( \{W_s, Z_s, \varepsilon_s\}_{s \geq 0} \) defined on a filtered probability space \( (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P) \), where \( s \) can be interpreted as a continuous time index. Assume we observe at discrete points of time so that \( \{Y_{th}, W_{th}, Z_{th} : t = 0, 1, \ldots, T = N/h\} \), where \( N \) is the time span. We normalize the time span \( N = 1 \) for simplicity. We denote the increment of processes as \( \Delta_{hi} Y_t := Y_{th} - Y_{(t-1)h} \). Let \( X_{hi} = (W_{hi}' , Z_{hi})' \) so that \( Z_{hi} = R' X_{hi} \). The model (4) can be expressed as

\[
\Delta_{hi} Y_t = \left\{ \begin{array}{ll} \Delta_{hi} X_{t}' \beta_h + \Delta_{hi} \varepsilon_t & \text{if } t = 1, \ldots, [\rho_0 T] \\ \Delta_{hi} X_{t}' \beta_h + (\Delta_{hi} Z_{t}) \delta_h + \Delta_{hi} \varepsilon_t & \text{if } t = [\rho_0 T] + 1, \ldots, T. \end{array} \right.
\]

Divide both sides by \( \sqrt{n} \) so that the error term variance is \( O(1) \). The parameters \( \beta_h \) and \( \delta_h \) may depend on the sampling interval, denoted by subscript \( h \). Let \( \varepsilon_t := \Delta_{hi} \varepsilon_t / \sqrt{n} \), \( y_t := \Delta_{hi} Y_t / \sqrt{n} \), \( x_t := \Delta_{hi} X_t / \sqrt{n} \), \( z_t := \Delta_{hi} Z_t / \sqrt{n} = R' x_t \),

\[
y_t = \begin{cases} x_t' \beta_h + \varepsilon_t & \text{if } t = 1, \ldots, [\rho_0 T] \\ x_t' \beta_h + z_t' \delta_h + \varepsilon_t & \text{if } t = [\rho_0 T] + 1, \ldots, T. \end{cases}
\]

**Assumption 5.** \( \{z_t, \varepsilon_t\} \) is a covariance stationary process that satisfies the functional central limit theorem as \( T = 1/h \to \infty \),

\[
T^{-1/2} \sum_{t=1}^{[\rho T]} z_t \varepsilon_t \Rightarrow B_1(s),
\]

where \( B_1(s) \) is a multivariate Gaussian process on \([0, 1]\) with mean zero and covariance

\[
E[B_1(u)B_1(v)'] = \min(u, v) \Xi, \quad \text{and}
\]

\[
\Xi := \lim_{T \to \infty} E \left[ \left( T^{-1/2} \sum_{t=1}^{T} z_t \varepsilon_t \right)^2 \right].
\]

**Assumption 6.** The break magnitude is \( \delta_h = d_0 \lambda_h \), where \( d_0 \in \mathbb{R}^4 \) is a fixed vector and \( \lambda_h \) is a scalar that depends on the sampling interval \( h \). Assume one of the following cases on \( \lambda_h \) as \( h \to 0 \):

1. \( \lambda_h = O(h^{1/2}) \) so that \( \delta_h = d_0 \sqrt{h} \)
2. \( \lambda_h = O(h^{1/2-\gamma}) \), where \( 0 < \gamma < 1/2 \) so that \( \delta_h / \sqrt{n} \to \infty \) simultaneously with \( \delta_h \to 0 \).

Notations from Section 3 are used for model (8): \( MY = MZ_0 \delta_h + M\varepsilon \), where \( \varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)' \). The objective function of the estimator \( \hat{k} \) in (5) is restated as follows:

\[
Q_T(\hat{k})^2 = \sqrt{T} \hat{\delta}_h^' (T^{-1/2} Z_t' M Z_k)^{-1/2} \Omega_k (T^{-1/2} Z_k' M Z_k)^{-1/2} \sqrt{T} \hat{\delta}_h
\]

The in-fill asymptotic distribution is derived for the two different magnitudes of \( \delta_h \) in Assumption 6. Theorem 3 provides the limit distribution under 6(i), which represents small breaks. For proof, see Appendix B.
Theorem 3. Consider the model (8) with unknown parameters $(\beta_h, \delta_t)$. Assumptions 1, 2, 5, and 6(ii) hold. Then the break point estimator $\hat{\rho} = \hat{k}/T$ defined in (5) has the following in-fill asymptotic distribution as $h \to 0$,

$$T\|\delta_h\|^2 \hat{\rho} \xrightarrow{d} \|d_0\|^2 \arg \max_{\rho \in (0,1)} \tilde{W}(\rho) \tilde{\Omega}(\rho) \tilde{W}(\rho),$$

with

$$\tilde{W}(\rho) := \frac{1}{\sqrt{t}} \tilde{B}_t(\rho) - \frac{1}{\sqrt{t}} \tilde{B}_t(1) \frac{1}{\sqrt{1 - \rho}} \frac{1}{\sqrt{t}} \sum_{i=1}^{d_0} d_{0i}^{2} \left\{ W(u) - \frac{|u|}{2} \right\},$$

where $B_t(\cdot)$ is a Brownian motion defined in Assumption 5.

Next, consider the case of Assumption 6(ii). The proof of Theorem 4 is in Appendix B.

Theorem 4. Consider the model (8) with unknown parameters $(\beta_h, \delta_t)$. Assumptions 1, 2, 5, and 6(ii) hold. For simplicity, we denote $\tilde{\Omega}_0$ for $\tilde{\Omega}(\rho_0)$. Then the break point estimator $\hat{\rho} = \hat{k}/T$ defined in (5) has the following in-fill asymptotic distribution as $h \to 0$,

$$\lambda^2 h T(\hat{\rho} - \rho_0) \xrightarrow{d} \frac{d_0^{2} \tilde{\Omega}_0 \Sigma \tilde{\Omega}_0 d_0}{(d_0^{2} \Sigma \tilde{\Omega}_0 d_0)^2} \arg \max_{u \in (-\infty, \infty)} \left\{ W(u) - \frac{|u|}{2} \right\},$$

where $A_u = \tilde{\Omega}_0 - \text{sgn}(u) \rho_0 (1 - \rho_0) \nabla \tilde{\Omega}_0$, $\nabla \tilde{\Omega}_0 \equiv \frac{\partial \tilde{\Omega}(\rho)}{\partial \rho} \bigg|_{\rho = \rho_0}$, $W(u) = W_1(\rho)$ for $u \leq 0$ and $W(u) = W_2(u)$ for $u > 0$. $W_1(\cdot)$ and $W_2(\cdot)$ are two independent Wiener Processes on $[0, \infty)$.

If the weight matrix is $\Omega_k = I_n$, the estimator is equivalent to the LS estimator and the limiting distribution reduces to the distribution in Proposition 3 of Bai (1997). The term $A_u$ shows how the weight matrix down-weights break points near the boundaries. Suppose $\tilde{\Omega}_k = T^{-1/2} Z_{\rho}^T M_{Z_{\rho}} Z_{\rho}$ and $\rho_0 > 0.5$ so that $\tilde{\Omega}(\rho) = \rho (1 - \rho) \tilde{\Omega}_0$ and $\nabla \tilde{\Omega}_0 < 0$. If $u > 0$ increases in a positive direction toward the boundary (i.e., $\rho > \rho_0 > 0.5$), then $A_u$ increases and the term multiplied to the Wiener process with drift, $(d_0^{2} \Sigma \tilde{\Omega}_0 d_0)^{-1}$ decreases. In contrast, if $u < 0$ decreases such that $\rho$ shifts toward the median, then $A_u$ decreases and $(d_0^{2} \Sigma \tilde{\Omega}_0 d_0)^{-1}$ increases. The result is opposite if $\rho_0 < 0.5$; there is larger weight on $\rho$ near the median 0.5 and less weight near the boundaries.

For large break magnitudes (Assumption 6(ii)) we can use bootstrap methods to approximate the distribution of the break point estimator and obtain confidence intervals. The wild bootstrap method of Liu (1988) can be used under model (4) with independent and possibly heteroscedastic errors. To obtain confidence intervals implied by the in-fill asymptotic distribution, Jiang et al. (2020) suggests using the highest density region (Hyndman 1996).

4.2. Break in an Autoregressive Model

In this section, I derive the in-fill asymptotic distribution of the break point estimator under an autoregressive (AR) model with a break in its lag coefficient, using a deterministic weight function $\omega(\cdot)$. As mentioned in Section 3, Assumption 1 excludes lagged dependent variables, due to the dependence of the weight function on regressors. This condition is relaxed to allow weakly exogenous regressors by assuming non-stochastic weights. Consider a discrete model closely related to the Ornstein-Uhlenbeck process with a break in the drift function:

$$dx(t) = -(\mu + \delta \mathbb{1}[t > \rho_0])x(t)dt + \sigma dB(t),$$

where $t \in [0, 1]$ and $B(\cdot)$ denote a standard Brownian motion. The discrete time model has the form

$$x_t = (\beta_1 \mathbb{1}[t \leq k_0] + \beta_2 \mathbb{1}[t > k_0]) x_{t-1} + \sqrt{h} e_t, \quad e_t \overset{i.i.d.}{\sim} (0, \sigma^2), \quad x_0 = O_p(1),$$

where $\beta_1 = \exp(-\mu/T)$ and $\beta_2 = \exp(-\mu + \delta)/T$ are the AR roots before and after the break. We denote $y_t = x_t/\sqrt{h}$ so that the order of errors is $O_p(1)$ as in model (4). Then, I have for $t = 1, \ldots, T$,

$$y_t = (\beta_1 \mathbb{1}[t \leq k_0] + \beta_2 \mathbb{1}[t > k_0]) y_{t-1} + e_t, \quad e_t \overset{i.i.d.}{\sim} (0, \sigma^2),$$

$$y_0 = x_0/\sqrt{h} = O_p(T^{1/2}).$$

The initial condition of $y_t$ in (10) diverges at rate $T^{1/2}$; thus, the in-fill asymptotic distribution will depend explicitly on the initial value $x_0$. The break size is $\beta_2 - \beta_1 = O(T^{-1})$, whereas the literature on long-span asymptotics assumes $O(T^{-\gamma})$ with $0 < \gamma < 1$. The model (10) is a local-to-unit root process: $\beta_1 = \exp(-\mu/T) \to 1$ and $\beta_2 = \exp(-\mu + \delta)/T \to 1$, as $T \to \infty$ for any finite $(\mu, \delta)$. In contrast, the long-span asymptotic theory incorporates stationary AR(1) processes, where $|\beta_1| < 1$ and $|\beta_2| < 1$. Chong (2001) derives the long-span distribution under $|\beta_2 - \beta_1| = O(T^{-1/2-\gamma})$ with $0 < \gamma < 1/2$. Jiang et al. (2020) provides simulation results that the in-fill asymptotic theory works well even when $\beta_1$ and/or $\beta_2$ are distant from unity in the finite sample. The break point estimator in model (10) is obtained by (5) where $x_t = z_t = y_{t-1}$ in (4).

Theorem 5. Consider the model (10) with fixed parameters $(\mu, \delta)$ so that $\beta_1 = O(T^{-1})$ and $\beta_2 = O(T^{-1})$. Assume the weight function $\omega_k)$ is nonrandom and bounded on the unit interval with $\omega_k \to \omega(\rho)$ as $k/T \to \rho$. Then, the break point estimator $\hat{\rho} = \hat{k}/T$ has the in-fill asymptotic distribution as

$$\hat{\rho} \Rightarrow \arg \max_{\rho \in (0,1)} \omega(\rho)^2 \frac{\left[ (\tilde{\omega}(\rho)^2 - \tilde{\omega}(0)^2 - \rho)^2 + \frac{\tilde{\omega}(1)^2 - \tilde{\omega}(0)^2 - (1 - \rho)^2}{\tilde{\omega}(r)^2 dr} \right]}{\tilde{\omega}(r)^2 dr},$$

where $\tilde{\omega}(r), \rho \in [0, 1]$ is a Gaussian process defined by

$$d\tilde{\omega}(r) = -((\mu + \delta \mathbb{1}[r > \rho_0]) \tilde{\omega}(r) dr + dB(r),$$

with the initial condition $\tilde{\omega}(0) = y_0/\sigma = x_0/(\sigma \sqrt{h})$, and $B(\cdot)$ is a standard Brownian motion.
The results of Theorem 5 derive from applying the continuous mapping theorem to the limit distribution in Theorem 4.1 from Jiang et al. (2020). See Appendix B for the proof. Both estimators are asymmetrically distributed around the true point and biased when \( \rho_0 \neq 1/2 \).

5. Monte Carlo Simulation

This section compares finite sample distributions of the new estimator and the LS estimator using Monte Carlo simulation. It considers two different models with small break magnitudes; a break in the mean of a univariate regression model and a break in the lag coefficient of the AR(1) process. I compare the root mean squared error (RMSE), bias, and standard errors of the two estimators in the finite sample and in-fill asymptotics.

5.1. Univariate Stationary Process

The first model is when a structural break occurs in model (4), where \( x_t = \xi = 1 \) for all \( t \). The break magnitude \( \delta_T = T^{-1/2}d_0 \) is in the local \( T^{-1/2} \) neighborhood of zero to represent small break magnitudes.

\[
y_t = \mu + \delta_T 1\{ t > [\rho_0 T]\} + \epsilon_t, \tag{12}\]

where \( \epsilon_t \overset{iid}{\sim} N(0, \sigma^2) \) and \( \sigma = 1 \). Parameter values are \( \rho_0 \in \{0.15, 0.3, 0.5, 0.7, 0.85\} \), \( \mu = 4 \), \( d_0 \in \{1, 2, 4\} \), and \( T = 100 \) with 5,000 replications. The weight function is \( \omega_k = (k/T(1-k/T))^{1/2} \), which is the representative weight function motivated in Section 2. The break point estimator \( \hat{\rho}_{NEW} \) is defined in (3) and the LS estimator \( \hat{\rho}_{LS} \) in (2). Although it is unnecessary to trim under the simple model (12), I trim the optimization space by fraction \( \alpha = 0.1 \) on both ends, following the common practice in the literature.

Table 1 lists the RMSE, bias, and standard error for the finite sample distribution. For all \( \rho_0 \) and \( d_0 \) values, the RMSE of the estimator \( \hat{\rho}_{NEW} \) is smaller than that of \( \hat{\rho}_{LS} \) in the finite sample. A comparison of the asymptotic RMSE shows the same results qualitatively (see Appendix C). A tradeoff emerges from a slightly larger bias but a large decrease in the standard error for \( \hat{\rho}_{NEW} \) compared to \( \hat{\rho}_{LS} \), which leads to a decrease in the RMSE. When \( \rho_0 = 0.5 \), both the bias and standard error of the new estimator are smaller than those of the LS estimator:

| \( \rho_0 \) | \( d_0 \) | RMSE NEW | Bias NEW | Standard error NEW |
|----------|----------|----------|----------|-------------------|
| 0.15     | 1        | 0.4034   | 0.3481   | 0.3401            |
| 0.30     | 1        | 0.2853   | 0.3303   | 0.1933            |
| 0.50     | 1        | 0.2051   | 0.2681   | -0.0029           |
| 0.70     | 1        | 0.2686   | 0.3334   | -0.1940           |
| 0.85     | 1        | 0.4018   | 0.4394   | -0.3340           |
| 0.15     | 2        | 0.3897   | 0.4211   | 0.3243            |
| 0.30     | 2        | 0.2669   | 0.3150   | 0.1741            |
| 0.50     | 2        | 0.2018   | 0.2435   | 0.1139            |
| 0.70     | 2        | 0.2640   | 0.3104   | -0.1693           |
| 0.85     | 2        | 0.3913   | 0.4224   | -0.3265           |
| 0.15     | 4        | 0.3455   | 0.3556   | 0.2703            |
| 0.30     | 4        | 0.2018   | 0.2435   | 0.1139            |
| 0.50     | 4        | 0.2018   | 0.2435   | 0.1139            |
| 0.70     | 4        | 0.2043   | 0.2448   | -0.1151           |
| 0.85     | 4        | 0.3438   | 0.3524   | -0.2678           |

NOTE: The number of replications is 5000.

The weight function biases the estimator toward the median in a tradeoff to a significant decrease in the standard error. When a break occurs near the boundaries, we need to be careful, because the new estimator can also be problematic. I suggest minimizing the trimming fraction \( \alpha \) and using the new break point estimator.

For comparison under large break magnitudes \( \delta_T = d_0^{-1/2+\gamma} \), \( \gamma = 1/4 \) \( d_0 \in \{1, 2, 4\} \) (see Appendix C). In this case, the results are reversed depending on the true break point. When \( \rho \) is near the boundaries and \( d_0 \) is large, the unnecessary deviation of the new estimator is larger than that of the LS estimator, which leads to a larger RMSE. Without trimming the optimization space, the new estimator has a smaller RMSE than LS for \( d_0 \in \{1, 2\} \) and all \( \rho_0 \).

5.2. Autoregressive Process

For the AR(1) process, I replicate two experiments from Jiang et al. (2020). The first experiment is a break in the lag coefficient so that the stationary process changes to another stationary AR(1) process. The second case is a change from a local-to-unit root to a stationary AR(1) process. Each experiment is generated from model (10) with \( h = 1/200 \) \( (T = 200) \), \( \sigma = 1 \), \( \epsilon_t \overset{iid}{\sim} N(0, 1) \), \( \rho_0 \in \{0.3, 0.5, 0.7\} \), and different combinations of \( \mu \) and \( \delta \) with \( \beta_1 = \exp(-\mu/T) \) and \( \beta_2 = \exp(-\mu/\delta)/T) \).

1. Stationary to stationary: \((\mu, \delta) = (138, 55)\), which implies \((\beta_1, \beta_2) = (0.5, 0.38)\);  
2. Local-to-unit to stationary: \((\mu, \delta) = (1, 5)\), which implies \((\beta_1, \beta_2) = (0.995, 0.97)\).

The stochastic integrals of in-fill asymptotic distributions are approximated over a grid size \( h = 0.001 \). The optimization space is trimmed by fraction \( \alpha = 0.1 \). The in-fill asymptotic distribution of the break point estimator \( \hat{\rho}_{NEW} \) is stated in Theorem 3. The in-fill asymptotic distribution of the LS estimator \( \hat{\rho}_{LS} \) is stated in Theorem 4.1 from Jiang et al. (2020).

Table 2 provides the RMSE, bias, and the standard error of \( \hat{\rho}_{NEW} \) and \( \hat{\rho}_{LS} \) for the finite sample, respectively (see Appendix C for the asymptotic distribution). Similar to results in Section 5.1,
the RMSE of \( \hat{\rho}_{\text{NEW}} \) is smaller than that of \( \hat{\rho}_{\text{LS}} \) for all parameter values \((\beta_1, \beta_2, \rho_0)\) considered. This also holds in the limit. A decrease in RMSE of \( \hat{\rho}_{\text{NEW}} \) emerges from the trade-off of a relatively large decrease in variance compared to the increase in the squared bias.

Figures 5 and 6 are finite sample distributions of the break point in the two experiments. For the stationary to another stationary process change, the LS estimator \( \hat{\rho}_{\text{LS}} \) mode at the true break point is almost negligible, unless it is the median \( \rho_0 = 0.5 \). In contrast, the estimator \( \hat{\rho}_{\text{NEW}} \) has a unique mode at the true break point for all \( \rho_0 \in \{0.3, 0.5, 0.7\} \). For the local-to-unit root to a stationary AR(1) change, both estimators have a higher probability at the true break point. However, the LS estimator continues to exhibit tri-modality with modes
Figure 4. \((\rho_0 = 0.85)\) Finite sample distribution of the new estimator \(\hat{\rho}_{\text{NEW}}\) (left) and the LS estimator \(\hat{\rho}_{\text{LS}}\) (right) under model (12), with parameter values \((\rho_0, \delta_T) = (0.85, T^{-1/2}), (0.85, 2T^{-1/2}),\) and \((0.85, 4T^{-1/2})\) and \(T = 100,\) respectively. The optimization space is trimmed by fraction \(\alpha\) on both ends.

at the ends, whereas the new estimator has a unique mode at \(\rho_0.\)

6. Empirical Application

In this section, I estimate the break date on the U.S. and U.K. stock returns using the return prediction model from Paye and Timmermann (2006). The authors studied the instability in models of ex-post predictable components in stock returns by examining structural breaks in the coefficients of state variables. The regression model (13) is specified with four state variables as follows: the lagged dividend yield, short-term interest rate, term spread, and default premium. The model allows for all coefficients to change because no strong reason exists to believe...
Table 2. Finite sample RMSE, bias, and the standard error of the new estimator and the LS estimator of the break point under the AR(1) model (10) with parameter values $(\beta_1, \beta_2, \rho_0)$ and $T = 200$.

| $\beta_1$ | $\beta_2$ | $\rho_0$ | RMSE NEW | Bias NEW | Standard error NEW | RMSE LS | Bias LS | Standard error LS |
|-----------|-----------|---------|----------|-----------|-------------------|--------|--------|-------------------|
| 0.5       | 0.38      | 0.3     | 0.2627   | 0.1821    | 0.1893            | 0.3091 | 0.0204 | 0.1751            |
| 0.5       | 0.38      | 0.5     | 0.2285   | 0.1771    | 0.1822            | 0.2452 | 0.0020 | 0.1751            |
| 0.5       | 0.38      | 0.7     | 0.2369   | 0.1754    | 0.1827            | 0.2780 | 0.0004 | 0.1848            |
| 0.995     | 0.97      | 0.3     | 0.2356   | 0.1867    | 0.1827            | 0.2784 | 0.0004 | 0.1848            |
| 0.995     | 0.97      | 0.5     | 0.2357   | 0.1867    | 0.1827            | 0.2784 | 0.0004 | 0.1848            |
| 0.995     | 0.97      | 0.7     | 0.2357   | 0.1867    | 0.1827            | 0.2784 | 0.0004 | 0.1848            |

NOTE: The number of replications is 5000.

that the coefficient on any of the regressors should be immune from shifts. The multivariate model with a one-time structural break at $k$ with $t = 1, \ldots, T$ is

$$
\text{Ret}_t = \beta_0 + \beta_1 \text{Div}_{t-1} + \beta_2 \text{Tbill}_{t-1} + \beta_3 \text{Spread}_{t-1} + \beta_4 \text{Def}_{t-1} + I(t > k) (\delta_0 + \delta_1 \text{Div}_{t-1} + \delta_2 \text{Tbill}_{t-1} + \delta_3 \text{Spread}_{t-1} + \delta_4 \text{Def}_{t-1}) + \epsilon_t,
$$

where $\text{Ret}_t$ represents the excess return for the international index in question during month $t$, $\text{Div}_{t-1}$ is the lagged dividend yield, $\text{Tbill}_{t-1}$ is the lagged local country short interest rate, $\text{Spread}_{t-1}$ is the lagged local country spread, $\text{Def}_{t-1}$ is the lagged local country spread,

Figure 5. (Stationary to stationary) Finite sample distributions of the new estimator (left) and the LS estimator (right) when the lag coefficient pre- and post-break are $(\beta_1, \beta_2) = (0.5, 0.38)$ at break points $\rho_0 = 0.3, 0.5$, and 0.7, respectively.
and \( \text{Def}_{t-1} \) is the lagged U.S. default premium. From the notation of model (4), \( y_t = \text{Ret}_t \) and for the multivariate model, \( x_t = z_t = (1, \text{Div}_{t-1}, \text{Tbill}_{t-1}, \text{Spread}_{t-1}, \text{Def}_{t-1}) \).

For the univariate model with dividend yield \( x_t = z_t = (1, \text{Div}_{t-1}) \), which is defined analogously for other univariate models. The weight matrix is \( \Omega_k = T^{-1} Z_k' M Z_k \), where \( Z_k = (0, \ldots, 0, z_{k+1}, \ldots, z_T)' \) and \( M = I - XX' \). Following the approach of Paye and Timmermann (2006), I examine univariate models to facilitate the interpretation of coefficients, in addition to the multivariate model (13).

I collected data from Global Financial Data and Federal Reserve Economic Data (FRED). The indices of the total return and dividend yield series are the S&P 500 for the U.S. and the Financial Times Stock Exchange (FTSE) All-share for the United Kingdom (see Appendix C for results on the U.K. stock returns). The dividend yield is expressed as an annual rate and
constructed as the sum of dividends over the preceding 12 months, divided by the current price. For both countries, the three-month Treasury bill (T-bill) rate is used as a measure of the short-term interest rate and the 20-year government bond yield is the measure of the long-term interest rate. Excess returns are the total return stocks in the local currency less the total return on T-bills. The term spread is constructed as the difference between the long- and short-term local interest rate. The U.S. default premium is the differences in yields between Moody’s Baa and Aaa rated bonds. The search grid is obtained by trimming each sample period by fraction $\alpha = 0.15$ (which is equivalent to the trimming window of Paye and Timmermann (2006)). For the full sample, the search grid is 1960:2-1996:3, and for the subsample it is 1975:1-1998:10.

Under the univariate model, with the lagged dividend yield as a single forecasting regressor, the LS estimate of the break point for the S&P 500 is close to the boundary of the search grid. Paye and Timmermann (2006) note that the NYSE or S&P 500 indices have the same estimated break date when the trimming window is shortened, and thus, the discrepancy is not the sole explanation for the timing of the break. However, it is likely that estimates are near the boundaries because of the finite sample behavior of the LS estimator. I check whether the new estimator provides a different break date estimate under model (13), using data similar to the first dataset from Paye and Timmermann (2006), which is monthly data on the U.S. and the U.K. stock returns from 1952:7 to 2003:12. For comparison, I also estimate the break using a shorter period 1970:1-2003:12, which is equivalent to the sample period of their second dataset.

Table 3 provides estimates of the two samples using the S&P 500 index. One notable feature is that the LS estimates that a break occurred in December 1994, with break point $\hat{\rho}_{LS} = 0.83$, whereas my method estimates a break in the mid-1980s and $\hat{\rho}_{NEW} = 0.62$. Although the LS estimate is close to the boundaries of the grid, it gives the same break estimate in the subsample. This suggests that a break may have occurred multiple times. Paye and Timmermann (2006) use the Bai and Perron (1998) method and find that two structural breaks occur in the return model (13), using the S&P 500, where each break occurs at 1987:7 and 1995:3. They note that the break in 1987 appears to be an isolated break, not appearing in other international markets. These two break date estimates are similar to estimates in Table 3, which assume a one-time structural break.

### Table 3. Structural break date estimates of the U.S. stock return (S&P 500) prediction model for samples 1952:7-2003:12 and 1970:1-2003:12.

| Model | 1952:7-2003:12 | 1970:1-2003:12 |
|-------|----------------|----------------|
|       | NEW | LS | NEW | LS |
| Multi. | 1984:8 | 1994:12 | 1982:8 | 1994:12 |
|        | 0.62 | 0.83 | 0.37 | 0.79 |
| Div. yield | 1982:8 | 1995:1 | 1982:8 | 1996:9 |
|         | 0.59 | 0.83 | 0.37 | 0.79 |
| T-bill | 1974:10 | 1974:10 | 1982:8 | 1975:1 |
|         | 0.43 | 0.43 | 0.37 | 0.15 |
| Spread | 1983:5 | 1976:2 | 1987:9 | 1976:2 |
|         | 0.60 | 0.46 | 0.32 | 0.18 |
| Def. prem. | 1968:12 | 1965:11 | 1982:8 | 1975:7 |
|          | 0.32 | 0.26 | 0.37 | 0.16 |

NOTE: For each model, the first row is the break date estimate and the second row is the break point estimate (fraction in the corresponding sample).

An alternative explanation of the break in the early 1980s is that the estimation method captures a change in the individual state variable itself rather than the coefficient of the prediction model (13), because the noisy nature of stock market returns makes it extremely difficult to detect a break. For instance, the estimate $\hat{\beta}_1$ could be capturing noise caused by the movement in $D_{t-1}$ (see Appendix C).

### 7. Conclusion

This study provides an estimation method for the structural break point in multivariate linear regression models when a one-time break occurs in a subset of (or all) coefficients. In particular, this study focuses on break magnitudes that are empirically relevant. In practice, it is likely that the shift in the parameters is statistically small. The LS estimation widely used in the literature fails to accurately estimate the break point under small break magnitudes, which motivates us to construct the estimation method in this study.

I construct a weight function on the sample period normalized to a unit interval, which imposes small weights on the LS objective for potential break points with large estimation uncertainty. The break point estimator is the argmax of the objective function, which is equal to the LS objective multiplied by a weight function. The break point estimator is consistent under regularity conditions on a general weight function, with the same rate of convergence as the LS estimator from Bai (1997). The limit distribution under a small break magnitude is derived using the in-fill asymptotic theory, following the approach of Jiang et al. (2018, 2020).

Unlike the existing LS estimator, the break point estimator does not randomly select boundary values. However, it has drawbacks in certain cases. Monte Carlo simulation results show that if a structural break occurs near the boundaries of the optimization space, both the LS and the new estimation method have difficulty in accurately estimating the break point (small breaks) or the RMSE of the new estimator is larger than that of the LS (large breaks).

In short, this study provides an alternative estimation method that estimates the timing of a structural break in linear regression models under empirically relevant break magnitudes. My estimator shows a unimodal finite sample distribution under statistically small break magnitudes. To my knowledge, this is the first study to widen the class of break point estimators by generalizing least-squares. I provide theoretical results of the consistency of the estimator and an asymptotic distribution that represents finite sample behavior. If the break magnitude is small, my estimator outperforms the LS estimation in terms of RMSE. Thus, under statistically small but empirically relevant breaks, the estimator described in this study provides reliable inferences of the change point in models.

### Supplementary Materials

Appendices: Appendix A provides an explanation of the Bayesian framework in Section 2, suggestions on other weight functions and the derivation that the estimator in (3) is a special case of (5). Appendix B includes all proofs of Sections 3 and 4. Additional tables and figures are in Appendix C. (pdf file)
Acknowledgments

I am very grateful for helpful comments and suggestions from Graham Elliott, James Hamilton, Brendan Beare, Kaspar Wuthrich, Yixiao Sun, Juwon Seo, Jungmo Yoon, and all seminar participants in UC San Diego Department of Economics and Peking University HSBC Business School. All remaining errors are my own.

Disclosure Statement

The author report there are no competing interests to declare.

ORCID

Yaein Baek  http://orcid.org/0000-0001-5862-448X

References

Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press. [96,98]

Anderson, T., and Darling, D. (1954), "A Test of Goodness of Fit," *Journal of the American Statistical Association*, 49, 765–769. [98]

Andrews, D. (1988), "Laws of Large Numbers for Dependent Nonidentically Distributed Random Variables," *Econometric Theory*, 4, 458–466. [99]

Bai, J. (1994), "Least Squares Estimation of a Shift in Linear Processes," *Journal of Time Series Analysis*, 15, 453–472. [95,96,98]

——— (1997), "Estimation of a Change Point in Multiple Regression Models," *The Review of Economics and Statistics*, 79, 551–563. [95,96,98,99,100,101,107]

Bai, J., Lumsdaine, R. L., and Stock, J. (1998), "Testing for and Dating Common Breaks in Multivariate Time Series," *Review of Economic Studies*, 65, 395–432. [95]

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78. [95,107]

——— (2003), "Computation and Analysis of Multiple Structural Changes," *Journal of Applied Econometrics*, 18, 1–22. [95]

Bhattacharya, P. K. (1987), "Maximum Likelihood Estimation of a Change Point in the Distribution of Independent Random Variables," *Journal of Multivariate Analysis*, 23, 183–208. [95]

Card, D., Mas, A., and Rothstein, J. (2008), "Tipping and the Dynamics of Segregation," *The Quarterly Journal of Economics*, 123, 177–218. [95]

Casini, A., and Perron, P. (2019), "Continuous Record Laplace-based Inference about the Break Date in Structural Models," arXiv preprint arXiv:1804.00232. [95,100]

——— (2020), "Generalized Laplace Inference in Multiple Change-Point Models," arXiv preprint arXiv:1803.10871. [95]

Chong, T. T. L. (2001), "Structural Change in AR(1) Models," *Econometric Theory*, 17, 87–155. [100,101]

Elliott, G., and Müller, U. K. (2007), "Confidence Sets for the Date of a Single Break in Linear Time Series Regression," *Journal of Econometrics*, 141, 1196–1218. [95,96]

——— (2014), "Pre and Post Break Parameter Inference," *Journal of Econometrics*, 180, 141–157. [96]

González-Val, R., and M. Marcén (2012), "Unilateral Divorce versus Child Custody and Child Support in the U.S.," *Journal of Economic Behavior & Organization*, 81, 613–643. [95]

Hansen, B. E. (1991), "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221. [100]

Hinkley, D. (1970), "Inference about the Change Point in a Sequence of Random Variables," *Biometrika*, 57, 1–17. [95]

Hyndman, R. (1996), "Computing and Graphing Highest Density Regions," *The American Statistician*, 50, 120–126. [101]

Jiang, L., Wang, X., and Yu, S. (2018), "New Distribution Theory for the Estimation of Structural Break," *Journal of Econometrics*, 205, 156–176. [96,100,107]

——— (2020), "In-fill Asymptotic Theory for Structural Break Point in Autoregressions," *Econometric Reviews*, 40, 1–28. [96,100,101,102,107]

Liu, R. Y. (1988), "Bootstrap Procedures under Some non-i.i.d. Models," *Annals of Statistics*, 16, 1696–1708. [101]

McLeish, D. L. (1975), "A Maximal Inequality and Dependent Strong Laws," *The Annals of Probability*, 5, 829–839. [99]

Pang, T., Chong, T., Zhang, D., and Liang, Y. (2018), "Structural Change in Nonstationary AR (1) Models," *Econometric Theory*, 34, 985–1017. [100]

Pang, T., Zhang, D., and Chong, T. (2014), "Asymptotic Inferences for an AR (1) Model with a Change Point: Stationary and Nearly Non-stationary Cases," *Journal of Time Series Analysis*, 35, 133–150. [100]

Paye, B. S., and A. Timmermann (2006), "Instability of Return Prediction Models," *Journal of Empirical Finance*, 13, 274–315. [104,106,107]

Yao, Y. C. (1987), "Approximating the Distribution of the ML Estimate of the Change-Point in a Sequence of Independent rxs," *The Annals of Statistics*, 3, 1321–1328. [95]