

PHBS WORKING PAPER SERIES

**Estimation of a Structural Break Point
In Linear Regression Models**

Yaein Baek*
Peking University

May 2021

Working Paper 20210510

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. real GDP growth and U.S. and UK stock return prediction models.

Keywords: Change point, parameter instability, structural change

JEL Classification: C13

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China



PHBS 商界学校
北京大学汇丰商学院

Estimation of a Structural Break Point in Linear Regression Models

Yaein Baek*

Peking University HSBC Business School

June 4, 2020

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. real GDP growth and U.S. and UK stock return prediction models.

Keywords: Change point, Parameter instability, Structural change

*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. Website: <http://sites.google.com/view/yaeinbaek>. Email: yibaek@phbs.pku.edu.cn.

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 percent. 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 $O(T^{-1/2})$ ¹. 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. I also 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,

¹Asymptotic analysis under a DGP with a drifting sequence of parameters can be considered as a form of weak identification asymptotics. Literature on estimation and inference with a restricted parameter space under weak identification includes Andrews and Cheng (2012, 2013, 2014), and Han and McCloskey (2019). Their results do not cover an abrupt structural change model, which is our model of interest.

under an in-fill asymptotic framework, following the approach of Jiang et al. (2017, 2018). Monte Carlo simulations show that the break point estimator has smaller root mean squared error (RMSE) than the LS estimator in a finite sample for all break point values considered.

This study provides two empirical applications: estimation of structural breaks on post-war U.S. real GDP growth rate and the U.S. and UK stock return prediction models. For the quarterly U.S. real GDP growth rate under different sample periods, the new method estimates a break in the early 1970s, whereas the LS estimates vary from the 1970s to 1952 or 2000, which are near boundaries of the sample. The break date estimate in the early 1970s is matched with the “productivity growth slowdown” suggested in literature, such as Perron (1989) and Hansen (2001). Thus, my estimation method yields reasonable break point estimates compared to LS estimates, which is sensitive to trimming the sample.

The remainder of this paper 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 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 Section 5 and Section 6 provides three empirical applications of the new structural break estimation method. We provide concluding remarks in Section 7. Additional theoretical results and proofs are in the Appendix.

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 greatest smaller integer function, and $\mathbf{1}\{t > k_0\}$ is an indicator function that equals one if $t > k_0$ and zero otherwise.

$$y_t = \mu + \delta \mathbf{1}\{t > k_0\} + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

The disturbances $\{\varepsilon_t\}$ are independent and identically distributed (i.i.d.) with mean zero and $E\varepsilon_t^2 = \sigma^2$. The pre-break mean y_t is μ and the post-break mean is $\mu + \delta$. Assume we know a one-time break occurs, but the break point ρ_0 and parameters (μ, δ, σ^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, $S_T(k)^2 = \sum_{t=1}^k (y_t - \bar{y}_k)^2 + \sum_{t=k+1}^T (y_t - \bar{y}_k^*)^2$, 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 $\sum_{t=1}^T (y_t - \bar{y})^2 = S_T(k)^2 + TV_T(k)^2$ (Amemiya 1985), where $V_T(k)^2 = k/T(1 - k/T) (\bar{y}_k^* - \bar{y}_k)^2$, to substitute for the SSR. Then the LS estimator of the break date is equivalent to

$$\hat{k}_{LS} = \arg \max_{k=1, \dots, T-1} |V_T(k)|, \quad \hat{\rho}_{LS} = \hat{k}_{LS}/T. \quad (2)$$

Denote $\rho = k/T$ and $\rho_0 = k_0/T$. 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 ρ_0 .

A break magnitude that is statistically 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 percent. 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 $\mu = 0.88$ percent 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 $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 0.8^2)$. The left plot of Figure 1 shows the finite sample distribution of the LS estimator of ρ from a Monte Carlo simulation with 2,000 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.

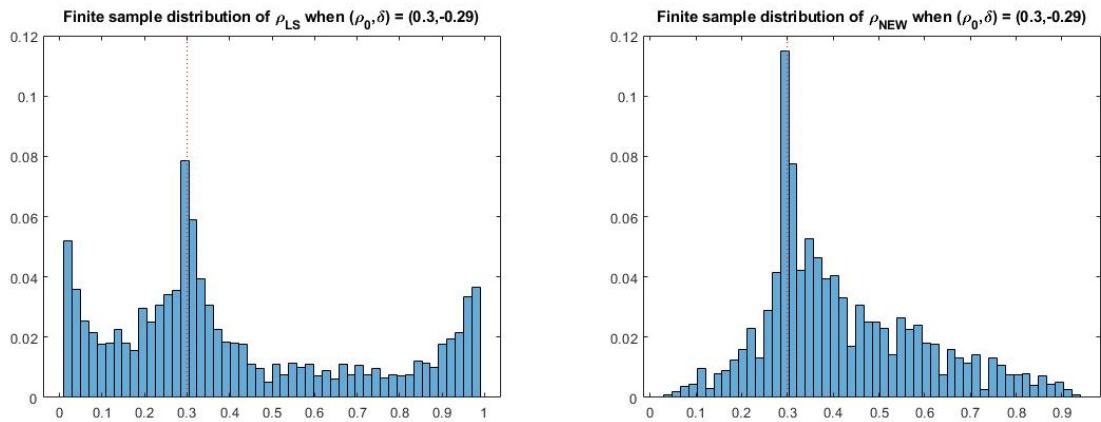


Figure 1: Finite sample distribution of the LS estimator (left) and new estimator (right) of break point ρ with weight function $w(\rho) = (\rho(1 - \rho))^{1/2}$ when $(\rho_0, \delta) = (0.3, -0.29)$, $T = 100$, and $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 0.8^2)$ with 2,000 replications.

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 $k = [\alpha T], \dots, [(1 - \alpha)T]$ 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 ρ with various fractions of trimming, $\alpha \in \{0, 0.1, 0.15, 0.2\}$, from a Monte Carlo simulation with 2,000 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 trade-off 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.

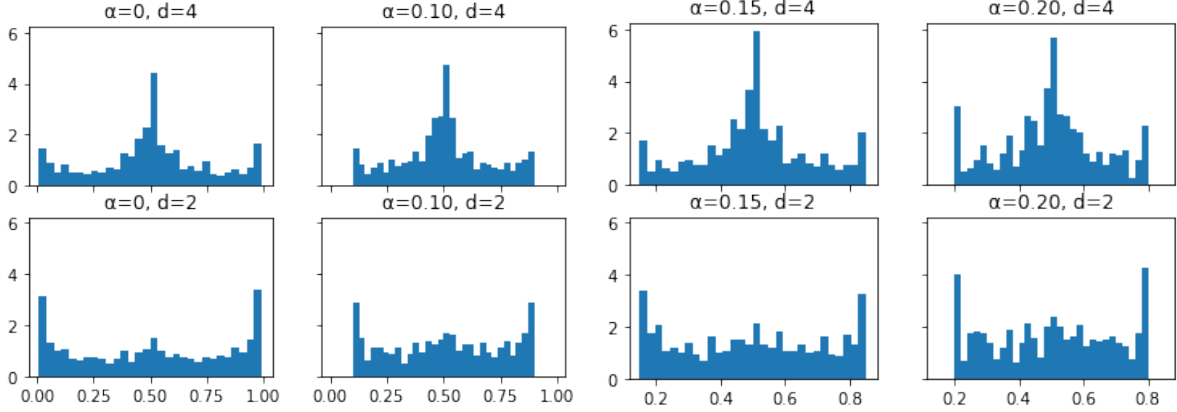


Figure 2: Finite sample distribution of $\hat{\rho}_{LS}$ when $(\rho_0, \delta) = (0.5, 4T^{-1/2})$ (upper plots), $(\rho_0, \delta) = (0.5, 2T^{-1/2})$ (lower plots) $T = 100$, and $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 1)$ with 2,000 replications. Each distribution has different trimmed optimization space $[\alpha, 1 - \alpha]$, with $\alpha \in \{0, 0.1, 0.15, 0.2\}$.

The finite sample distribution of the LS estimator has build-up mass at boundaries under small break magnitudes because of how the objective function is constructed. For each potential break date k , the objective function is constructed by partitioning the sample into two sub-samples, before and after k . Each sub-sample is used to estimate two different means, \bar{y}_k and \bar{y}_k^* . If k is near 1, the pre-break sub-sample size k is small; similarly, if k is near $T - 1$, the post-break sub-sample size $T - k$ is small². Hence, when the potential break date k of $|V_T(k)|$ is near the boundaries, the estimates of pre- or post-break mean are imprecise because of the small sub-sample size. Estimation uncertainty at boundaries distorts picking up the true break location if the break magnitude is small relative to sampling variability.

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 non-negative “weights” ω_k imposed on the LS objective function so that k with a large estimation error has smaller weights than k with a small estimation error. When $k = 1$ and $T - 1$, weights near zero are imposed, which implies the variance of the weighted objective function $\omega_k |V_T(k)|$ would shrink toward zero. If the sample period is normalized into a unit interval, the weights are represented by a continuous function $\omega(\rho)$ on $\rho \in [0, 1]$, which is zero at $\rho \in \{0, 1\}$, and has

²Estimation theory does not require ρ to be bounded away from zero and one, provided that a change point is assumed to exist. However, identification in a finite sample typically needs more than one observation pre- and post-break. In Section 3, I assume the break point exists; see Assumption 1(i).

positive values otherwise. A continuous function with such properties would look like an inverse U-shaped (or concave downward) function on the unit interval.

I propose a new break point estimator to maximize the value of the objective function $|Q_T(k)|$, equal to weights ω_k multiplied by the LS objective $|V_T(k)|$.

$$\hat{k} = \arg \max_{k=1, \dots, T-1} |Q_T(k)|, \quad \hat{\rho} = \hat{k}/T \quad (3)$$

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

The weight function shrinks 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 at 0.01 and 0.99. Section 5 provides additional Monte Carlo simulations comparing the two estimators.

The break point estimator (3) is easy to implement, as we simply modify the objective function by multiplying the weight function. It is a generalization of LS estimation because the LS estimator is a special case, when $\omega(\rho) = 1$. 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 on the weight function that ensures consistency of the break point estimator under a general linear regression model.

I also suggest a representative weight function $\omega(\rho) = (\rho(1-\rho))^{1/2}$ under model (1) (see Section 3 for its analogue under a model with multiple regressors). The weight function has two interpretations. First, it is related to the weighting function from Anderson and Darling (1954), which tests whether the sample is drawn from a particular distribution. A non-negative weight function is chosen to accentuate the boundaries of the sample space, where the test is desired to have sensitivity. If the cumulative distribution function (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 down-weight the boundaries of the parameter space $\rho \in [0, 1]$. If ρ is a random variable with cdf $F(\rho)$,

the weight would be the reciprocal of the weight function in Anderson and Darling (1954), $F(\rho)(1 - F(\rho))$. We further assume that ρ is uniformly distributed on the unit interval, $F(\rho) = \rho$. We obtain the weight function $\rho(1 - \rho)$, which down-weights the variance of the LS objective function $V_T(k)^2$ near the boundaries.

Second, from a Bayesian perspective, the weight function $\omega(\rho)$ can be interpreted as a prior on parameters δ and ρ . We assume the Gaussian disturbances in (1), $\omega(\rho) = (\rho(1 - \rho))^{1/2}$ 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 δ , given ρ . 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.

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 $k_0 = \lceil \rho_0 T \rceil$ 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_t and the variables associated with coefficients under a break as z_t . Let $x_t = (w_t', z_t)'$ be a $(p \times 1)$ vector and z_t is a $(q \times 1)$ vector with $q \leq p$,

$$y_t = \begin{cases} x_t' \beta + \varepsilon_t & \text{if } t = 1, \dots, k_0 \\ x_t' \beta + z_t' \delta_T + \varepsilon_t & \text{if } t = k_0 + 1, \dots, T, \end{cases} \quad (4)$$

where ε_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 $Y = (y_1, \dots, y_T)'$ and define $X_k := (0, \dots, 0, x_{k+1}, \dots, x_T)'$ and $X_0 := (0, \dots, 0, x_{k_0+1}, \dots, x_T)'$. Define Z_k and Z_0 analogously so that $Z_k = X_k R$ and $Z_0 = X_0 R$. Let $M := I - X(X'X)^{-1}X'$ and use the maximal invariant to eliminate the nuisance parameter β .

The subscript on δ_T shows that the break magnitude may depend on the sample size. 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 δ_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 $\bar{S} = Y'MY$, 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 $\bar{S} = S_T(k)^2 + V_T(k)^2$ (Amemiya 1985),

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

$$V_T(k)^2 := \hat{\delta}'_k (Z'_k M Z_k) \hat{\delta}_k,$$

where $\hat{\delta}_k$ is the LS estimate of δ_T by regressing MY on MZ_k . Note that $V_T(k)^2$ is non-negative from the inner product of the vector $(Z'_k M Z_k)^{1/2} \hat{\delta}_k$. The LS objective function is modified by multiplying a $(q \times q)$ positive definite weight matrix Ω_k , which is a generalization of ω_k 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 $(Z'_k M Z_k)^{1/2} \hat{\delta}_k$. Take the inner product and obtain the objective function $Q_T(k)^2 := \hat{\delta}'_k (Z'_k M Z_k)^{1/2} \Omega_k (Z'_k M Z_k)^{1/2} \hat{\delta}_k$. Then the estimator of the break point is

$$\hat{k} = \arg \max_k Q_T(k)^2, \quad \hat{\rho} = \hat{k}/T. \quad (5)$$

An example of the weight matrix is $\Omega_k = T^{-1} Z'_k M Z_k$, 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 $T^{-1} Z'_k M Z_k$ “decreases” as k approaches either end of the sample from the following rearrangement of terms:

$$\begin{aligned} T^{-1} Z'_k M Z_k &= T^{-1} [Z'_k Z_k - Z'_k X (X' X)^{-1} X' Z_k] \\ &= T^{-1} R' (X'_k X_k) (X' X)^{-1} (X' X - X'_k X_k) R. \end{aligned} \quad (6)$$

I prove the consistency of the break point estimator $\hat{\rho}$ in (5) under regularity conditions on model (4) and weight matrix Ω_k . The notation $\|\cdot\|$ denotes the Euclidean norm $\|x\| = (\sum_{i=1}^p x_i^2)^{1/2}$ 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. (i) $k_0 = [\rho_0 T]$, where $\rho_0 \in [\alpha, 1 - \alpha]$, $0 < \alpha < \frac{1}{2}$;

(ii) The data $\{y_{tT}, x_{tT}, z_{tT} : 1 \leq t \leq T, T \geq 1\}$ 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 $q \leq p$;

(iii) The matrices $(j^{-1} \sum_{t=1}^j x_t x_t')$, $(j^{-1} \sum_{t=T-j+1}^T x_t x_t')$, $(j^{-1} \sum_{t=k_0-j+1}^{k_0} x_t x_t')$ and $(j^{-1} \sum_{t=k_0+1}^{k_0+j} x_t x_t')$ 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, $\sup_{j \geq 1} \left\| j^{-1} \sum_{t=1}^j x_t x_t' \right\|$ is stochastically bounded;

(iv) $T^{-1} \sum_{t=1}^{[sT]} x_t x_t' \xrightarrow{P} s \Sigma_x$ uniformly in $s \in [0, 1]$, where Σ_x is a nonrandom positive definite matrix;

(v) For random regressors, $\sup_t E \|x_t\|^{4+\gamma} \leq K$ for some $\gamma > 0$ and $K < \infty$;

(vi) The disturbance ε_t is independent of the regressor x_s for all t and s . For an increasing sequence of σ -fields \mathcal{F}_t , $\{\varepsilon_t, \mathcal{F}_t\}$ form a sequence of L^r -mixingale sequence with $r = 4 + \gamma$ for some $\gamma > 0$ (McLeish (1975) and Andrews (1988)). That is, there exists nonnegative constants $\{c_t : t \geq 1\}$ and $\{\psi_j : j \geq 0\}$, such that $\psi_j \downarrow 0$ as $j \rightarrow \infty$ and for all $t \geq 1$ and $j \geq 0$, we have (a) $E |E(\varepsilon_t | \mathcal{F}_{t-j})|^r \leq c_t^r \psi_j^r$, (b) $E |\varepsilon_t - E(\varepsilon_t | \mathcal{F}_{t+j})|^r \leq c_t^r \psi_{j+1}^r$, (c) $\max_j |c_j| < K < \infty$, (d) $\sum_j j^{1+\kappa} \psi_j < \infty$ for some $\kappa > 0$.

Assumption 2. Ω_k is a positive definite ($q \times q$) matrix ($q = \dim(z_t)$) 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 $k = 1, \dots, T - 1$. In addition, for any nonzero vector $c \in \mathbb{R}^q$,

$$\left\| \Omega_{k_0}^{1/2} (Z_0' M Z_0)^{1/2} c \right\| > \left\| \Omega_k^{1/2} (Z_k' M Z_k)^{-1/2} (Z_k' M Z_0) c \right\|$$

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

The conditions of assumption 1 are similar to assumptions A1 to A6 in Bai (1997), with additional restrictions (iv) and (vi). Assumption 1(vi) allows for general serial correlation in disturbances and requires x_t to be strictly exogeneous. This is because Ω_k depends on the moments of regressors and we want to impose zero weights on the boundaries of the unit interval. For instance, if the second moments of z_t changes at ρ_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 Ω_k converges in probability to a nonrandom matrix that varies across ρ only. Note that if Ω_k is a non-stochastic matrix that satisfies the norm inequality in Assumption 2, consistency holds under weakly exogeneous regressors (see Assumption 4).

Assumption 2 guarantees that the matrix

$$A_T(k) := \frac{1}{|k_0 - k|} \left[(Z_0' M Z_0)^{1/2} \Omega_{k_0} (Z_0' M Z_0)^{1/2} - (Z_0' M Z_k) (Z_k' M Z_k)^{-1/2} \Omega_k (Z_k' M Z_k)^{-1/2} (Z_k' M Z_0) \right] \quad (7)$$

is positive definite, and hence, $\|A_T(k)\| \geq \lambda_{\min}(A_T(k)) > 0$, where λ_{\min} denotes the minimum eigenvalue of $A_T(k)$. The condition can be interpreted as follows: for simplicity, consider the univariate model (1). Assumption 2 is equivalent to $|\omega'(\rho)/\omega(\rho)| < (2\rho(1-\rho))^{-1}$ for all ρ , 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 ρ approaches zero or one. A sufficient condition is the function $\omega(\rho) = (\rho(1-\rho))^\gamma$, with $-1/2 \leq \gamma \leq 1/2$.³ Note that this is sufficient under Assumption 1(i). If α is arbitrarily close to zero, it may restrict the functional of $\omega(\cdot)$. The weight matrix Ω_k may be close to a singular matrix in a finite sample if α is extremely close to zero under model (4). Under Assumption 1(iv), the weight matrix converges in probability to a function of ρ and Σ_x as T increases. Because $\bar{\Omega}(\rho)$ is a differentiable function of ρ element-wise, $\|\Omega_k - \Omega_{k_0}\| \leq b|k - k_0|/T$ for some finite $b > 0$ and all k .

³The function $\omega(\rho) = (\rho(1-\rho))^\gamma$, with $-1/2 \leq \gamma \leq 1/2$, allows a convex function that has a large weight on the boundaries. This is because I assume “large” break magnitudes (Assumption 3) for consistency of the estimator. That is, if the break magnitude is large enough, we no longer have build-up mass at the boundaries and imposing large weights does not matter for consistency.

Assumption 3. $\delta_T \rightarrow 0$ and $T^{1/2-\gamma}\delta_T \rightarrow \infty$ for some $\gamma \in (0, \frac{1}{2})$.

The consistency of the break point estimator is proved by showing that if $\delta_T \neq 0$, then with high probability, $Q_T(k)^2$ can only be maximized near the true break k_0 . The objective function $Q_T(k)^2$ is defined in (5) and $\hat{\delta}_k$ is the LS estimator of the break magnitude, assuming that k is the break date: $\hat{\delta}_k = (Z'_k M Z_k)^{-1}(Z'_k M Z_0)\delta_T + (Z'_k M Z_k)^{-1}Z'_k M \varepsilon$. If $k = k_0$, then $\hat{\delta}_{k_0} = \delta_T + (Z'_0 M Z_0)^{-1}Z'_0 M \varepsilon$.

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

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

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).*

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

$$\left\| \Omega_{k_0}^{1/2} (Z'_0 M Z_0)^{1/2} c \right\| > \left\| \Omega_k^{1/2} (Z'_k M Z_k)^{-1/2} (Z'_k M Z_0) c \right\|$$

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

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

point estimator $\hat{\rho}$ in (5) is consistent.

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

The proof of Theorem 2 is similar to the proof of Theorem 1; hence we have omitted it. Under Assumptions 1(v), 2(i), and 2(ii), the strong law of large numbers holds for $x_t \varepsilon_t$, because the conditions in Hansen (1991) are satisfied. The weight matrix Ω_k in Assumption 4(iii) depends on k/T but not on the data $\{x_t, \varepsilon_t\}$. Thus, by setting $\rho = k/T$, Ω_k is a function of ρ , which is assumed to be differentiable with respect to ρ . 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 follows the same process as in the proof under Assumptions 1 and 2.

Given the consistency of the break point estimator from Theorem 1 or 2, the estimator of the break magnitude corresponding to \hat{k} is consistent and asymptotically normally distributed. Let $\hat{\delta}(\hat{\rho}) = \hat{\delta}_{\hat{k}}$, then the following results hold. The proof is provided in the Appendix.

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

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

where

$$\mathbf{V} := \text{plim}_{T \rightarrow \infty} T^{-1} Z_0' M Z_0, \quad \mathbf{U} := \lim_{T \rightarrow \infty} E \left[(T^{-1/2} Z_0' M \varepsilon)^2 \right].$$

4 In-fill 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 break point, if the second moment of variables associated with coefficients under break (z_t in Section 3) do not change before and after break. I am interested in small breaks in which

the asymptotic distribution depends on the parameters in a complicated manner (Elliott and Müller 2007).

Jiang et al. (2017, 2018) and Casini and Perron (2019) employed a continuous record asymptotic framework to derive the limit distribution of the break point estimator. By assuming that a continuous record is available, a continuous time approximation to the discrete time model is constructed and an in-fill asymptotic distribution is developed. In contrast to the long-span asymptotic, where the time span of the data increases, the in-fill asymptotic assumes a fixed time span with shrinking sampling intervals. The in-fill asymptotic distribution is asymmetric, tri-modal, and dependent on the initial condition. However, the long-span asymptotic distribution of the LS estimator under local-to-unity processes do not depend on the initial condition. 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 before and after the break. I follow the approach of Jiang et al. (2017, 2018) to derive the limit distribution of the break point estimator under a stationary and local-to-unity autoregressive process.

4.1 Partial break in a stationary process

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

$$\Delta_h Y_t = \begin{cases} (\Delta_h X_t)' \beta_h + \Delta_h \mathcal{E}_t & \text{if } t = 1, \dots, [\rho_0 T] \\ (\Delta_h X_t)' \beta_h + (\Delta_h Z_t)' \delta_h + \Delta_h \mathcal{E}_t & \text{if } t = [\rho_0 T] + 1, \dots, T, \end{cases}$$

Divide both sides by \sqrt{h} so that the error term variance is $O(1)$. The parameters β_h and δ_h may depend on the sampling interval, denoted by subscript h . Let $\varepsilon_t := \Delta_h \mathcal{E}_t / \sqrt{h}$,

$y_t := \Delta_h Y_t / \sqrt{h}$, $x_t := \Delta_h X_t / \sqrt{h}$, $z_t := \Delta_h Z_t / \sqrt{h} = R' x_t$,

$$y_t = \begin{cases} x_t' \beta_h + \varepsilon_t & \text{if } t = 1, \dots, [\rho_0 T] \\ x_t' \beta_h + z_t' \delta_h + \varepsilon_t & \text{if } t = [\rho_0 T] + 1, \dots, T. \end{cases} \quad (8)$$

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

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

where $B_1(s)$ is a multivariate Gaussian process on $[0, 1]$ with mean zero and covariance $E[B_1(u)B_1(v)'] = \min\{u, v\}\Xi$, and $\Xi := \lim_{T \rightarrow \infty} E \left[\left(T^{-1/2} \sum_{t=1}^T z_t \varepsilon_t \right)^2 \right]$.

Assumption 6. The break magnitude is $\delta_h = d_0 \lambda_h$, where $d_0 \in \mathbb{R}^q$ is a fixed vector and λ_h is a scalar that depends on the sampling interval h . Assume one of the following cases on λ_h as $h \rightarrow 0$,

(i) $\lambda_h = O(h^{1/2})$ so that $\delta_h = d_0 \sqrt{h}$;

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

Notations from Section 3 are used for model (8): $MY = MZ_0 \delta_h + M\varepsilon$, where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)'$ and $M = I - X(X'X)^{-1}X'$. The objective function of the estimator \hat{k} in (5) is restated as follows:

$$Q_T(k)^2 = \sqrt{T} \hat{\delta}'_k (T^{-1} Z'_k M Z_k)^{1/2} \Omega_k (T^{-1} Z'_k M Z_k)^{1/2} \sqrt{T} \hat{\delta}_k \quad (9)$$

The in-fill asymptotic distribution is derived for the two different magnitudes of δ_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 (β_h, δ_h) . Assumption 1, 2, 5, and 6(i) holds. Then the break point estimator $\hat{\rho} = \hat{k}/T$ defined in (5) has the following

in-fill asymptotic distribution as $h \rightarrow 0$,

$$T \|\delta_h\|^2 \hat{\rho} \xrightarrow{d} \|d_0\|^2 \arg \max_{\rho \in (0,1)} \widetilde{W}(\rho)' \bar{\Omega}(\rho) \widetilde{W}(\rho),$$

with

$$\begin{aligned} \widetilde{W}(\rho) &:= \Sigma_z^{-1/2} \frac{B_1(\rho) - \rho B_1(1)}{\sqrt{\rho(1-\rho)}} - (1-\rho_0) \left(\frac{\rho}{1-\rho} \right)^{1/2} \Sigma_z^{1/2} d_0 && \text{if } \rho \leq \rho_0 \\ &:= \Sigma_z^{-1/2} \frac{B_1(\rho) - \rho B_1(1)}{\sqrt{\rho(1-\rho)}} - \rho_0 \left(\frac{1-\rho}{\rho} \right)^{1/2} \Sigma_z^{1/2} d_0 && \text{if } \rho > \rho_0, \end{aligned}$$

where $B_1(\cdot)$ is a Brownian motion defined in Assumption 5.

An equivalent representation of the in-fill asymptotic distribution is (let $\rho = \rho_0 + u$)

$$T \|\delta_h\|^2 (\hat{\rho} - \rho_0) \xrightarrow{d} \|d_0\|^2 \arg \max_{u \in (\rho_0, 1-\rho_0)} \widetilde{W}(\rho_0 + u)' \bar{\Omega}(\rho_0 + u) \widetilde{W}(\rho_0 + u),$$

where $\widetilde{W}(\cdot)$ is defined in Theorem 3.

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 (β_h, δ_h) . Assumptions 1, 2, 5, and 6(ii) hold. For simplicity, we denote $\bar{\Omega}_0$ for $\bar{\Omega}(\rho_0)$. Then the break point estimator $\hat{\rho} = \hat{k}/T$ defined in (5) has the following in-fill asymptotic distribution as $h \rightarrow 0$,*

$$\lambda_h^2 T (\hat{\rho} - \rho_0) \xrightarrow{d} \frac{(d_0' \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0)}{(d_0' \Sigma_z A_u d_0)^2} \arg \max_{u \in (-\infty, \infty)} (d_0' \Sigma_z A_u d_0)^{-1} \left\{ W(u) - \frac{|u|}{2} \right\},$$

where $A_u = \bar{\Omega}_0 - \text{sgn}(u) \rho_0 (1 - \rho_0) \nabla \bar{\Omega}_0$, $\nabla \bar{\Omega}_0 \equiv \frac{\partial \bar{\Omega}(\rho)}{\partial \rho} \Big|_{\rho=\rho_0}$, $W(u) = W_1(-u)$ 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 $\Omega_k = I_q$, 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 $\Omega_k = T^{-1} Z_k' M Z_k$ and $\rho_0 > 0.5$ so that $\bar{\Omega}(\rho) = \rho(1-\rho) \Sigma_z$ and $\nabla \bar{\Omega}_0 < 0$. 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, $(d'_0 \Sigma_z A_u d_0)^{-1}$ decreases. In contrast, if $u < 0$ decreases such that ρ shifts toward the median, then A_u decreases and $(d'_0 \Sigma_z A_u d_0)^{-1}$ increases. The result is opposite if $\rho_0 < 0.5$. That is, there is larger weight on ρ near the median 0.5 and less weight near the boundaries.

4.2 Break in an autoregressive model

In this section, I derive the in-fill asymptotic distribution of an autoregressive (AR) model with a structural 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 exogeneous regressors by assuming non-stochastic weights. Consider a discrete model closely related to the Ornstein-Uhlenbeck process with a break in the drift function:

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

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

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

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 $t = 1, \dots, T$,

$$y_t = (\beta_1 \mathbf{1}\{t \leq k_0\} + \beta_2 \mathbf{1}\{t > k_0\})y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{i.i.d.}{\sim} (0, \sigma^2), \quad y_0 = x_0/\sqrt{h} = O_p(T^{1/2}). \quad (10)$$

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: $\beta_1 = \exp\{-\mu/T\} \rightarrow 1$ and $\beta_2 = \exp\{-(\mu + \delta)/T\} \rightarrow 1$, as $T \rightarrow \infty$ for any finite (μ, δ) . 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. (2017) provides simulation results that the in-fill asymptotic theory works well even when β_1 and/or β_2 are distant from unity in the finite sample.

The break point estimator and the LS estimator in model (10) takes the form

$$\begin{aligned}
S(k)^2 &= \sum_{t=1}^k \left(y_t - \hat{\beta}_1(k)y_{t-1} \right)^2 + \sum_{t=k+1}^T \left(y_t - \hat{\beta}_2(k)y_{t-1} \right)^2 \\
\hat{k} &= \arg \min_{k=1, \dots, T-1} \omega_k^2 S(k)^2, \quad \hat{\rho} = \hat{k}/T \\
\hat{k}_{LS} &= \arg \min_{k=1, \dots, T-1} S(k)^2, \quad \hat{\rho}_{LS} = \hat{k}_{LS}/T,
\end{aligned} \tag{11}$$

where $\hat{\beta}_1(k) = \sum_{t=1}^k y_t y_{t-1} / \sum_{t=1}^k y_{t-1}^2$ and $\hat{\beta}_2(k) = \sum_{t=k+1}^T y_t y_{t-1} / \sum_{t=k+1}^T y_{t-1}^2$ are LS estimates of β_1 and β_2 under break at k , respectively.

Theorem 5. *Consider the model (10) with fixed parameters (μ, δ) so that $\ln \beta_1 = O(T^{-1})$ and $\ln \beta_2 = O(T^{-1})$. Assume the weight function ω_k is nonrandom and bounded on the unit interval with $\omega_k \rightarrow \omega(\rho)$ as $k/T \rightarrow \rho$. Then, the break point estimator $\hat{\rho} = \hat{k}/T$ in (11) has the in-fill asymptotic distribution as*

$$\hat{\rho} \implies \arg \max_{\rho \in (0,1)} \omega(\rho)^2 \left[\frac{\left(\tilde{J}_0(\rho)^2 - \tilde{J}_0(0)^2 - \rho \right)^2}{\int_0^\rho \tilde{J}_0(r)^2 dr} + \frac{\left(\tilde{J}_0(1)^2 - \tilde{J}_0(\rho)^2 - (1 - \rho) \right)^2}{\int_\rho^1 \tilde{J}_0(r)^2 dr} \right],$$

where $\tilde{J}_0(r)$, for $r \in [0, 1]$ is a Gaussian process defined by

$$d\tilde{J}_0(r) = -(\mu + \delta \mathbf{1}\{r > \rho_0\})\tilde{J}_0(r)dr + dB(r), \tag{12}$$

with the initial condition $\tilde{J}_0(0) = y_0/\sigma = x_0/(\sigma\sqrt{h})$, and $B(\cdot)$ is a standard Brownian motion.

The results of Theorem 5 derive from applying the continuous mapping theorem to the limit distribution $S(k)^2$ in Theorem 4.1 from Jiang et al. (2017). See Appendix B for the proof. The difference between the asymptotic distributions of the two estimators is

the weight function multiplied by the stochastic process in the argmax function. 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; 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 $x_t = z_t = 1$ 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.

$$y_t = \mu + \delta_T \mathbf{1}\{t > [\rho_0 T]\} + \varepsilon_t, \quad (13)$$

where $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$ 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 2⁴. 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 (13), I trim the optimization space by fraction $\alpha = 0.1$ on both ends, following the common practice in the literature.

Table 1 provides the RMSE, the bias, and the standard error for the finite sample distribution. For all ρ_0 and d_0 values considered, the RMSE of the estimator $\hat{\rho}_{NEW}$ is smaller than that of $\hat{\rho}_{LS}$ in the finite sample. A comparison of asymptotic RMSE shows the same results qualitatively (see Appendix C). A trade-off emerges of slightly larger bias but

⁴If the weight function is $\omega_k = (k/T(1 - k/T))^\gamma$, Assumption 2 is satisfied if $-1/2 \leq \gamma \leq 1/2$ for an arbitrary small α in Assumption 1(i). For $\gamma \in \{1/8, 1/4, 3/8\}$, the results (omitted due to space constraints) do not change qualitatively; the probability at the boundaries decrease compared to the finite sample distribution of LS. Because $\omega(\rho) = (\rho(1 - \rho))^\gamma \rightarrow 1$ as $\gamma \rightarrow 0$, the difference between the two estimators finite sample behavior shrinks when γ is close to zero.

a large decrease in standard error for $\hat{\rho}_{NEW}$ compared to $\hat{\rho}_{LS}$, which leads to a decrease in RMSE. When $\rho_0 = 0.5$, both bias and standard error of the new estimator is smaller than the LS estimator.

Figures 3 and 4 show the finite sample distribution of the two estimators under $\rho_0 = 0.30$, and 0.85. Under $\rho_0 = 0.30$, the finite sample distribution of $\hat{\rho}_{LS}$ is tri-modal, whereas the $\hat{\rho}_{NEW}$ has an unique mode at ρ_0 for all d_0 values considered. When the true break point is near the boundaries of the optimization space and the break magnitude is small ($\rho_0 = 0.85$ and $d_0 = 1$), the LS estimator performs particularly worse. The tri-modal LS estimator distribution becomes bi-modal with modes at α and $1 - \alpha$. Under these parameter values, the new estimator has its drawbacks; the finite sample distribution is relatively flat. This is partly due to trimming the optimization space, which is not necessary for our estimation method. Without trimming ($\alpha = 0$), the finite sample distribution of our break point estimator has a unique mode at ρ_0 for all parameter values considered.

In short, the break point estimator is preferable than the LS estimator in terms of RMSE, under small break magnitudes. The weight function biases the estimator toward the median in trade-off to a significant decrease in 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 α and using the new break point estimator.

Table 1: Finite sample RMSE, bias, and the standard error of the new estimator and the LS estimator of the break point under model (13) with parameter values (ρ_0, d_0) and $T = 100$. The number of replications is 5,000.

| ρ_0 | d_0 | RMSE | | Bias | | Standard error | |
|----------|-------|--------|--------|---------|---------|----------------|--------|
| | | NEW | LS | NEW | LS | NEW | LS |
| 0.15 | 1 | 0.4034 | 0.4381 | 0.3428 | 0.3401 | 0.2127 | 0.2762 |
| | 2 | 0.3897 | 0.4211 | 0.3243 | 0.3152 | 0.2161 | 0.2792 |
| | 4 | 0.3455 | 0.3556 | 0.2703 | 0.2323 | 0.2151 | 0.2693 |
| 0.30 | 1 | 0.2853 | 0.3303 | 0.1933 | 0.1898 | 0.2098 | 0.2703 |
| | 2 | 0.2669 | 0.3150 | 0.1741 | 0.1703 | 0.2023 | 0.2649 |
| | 4 | 0.2018 | 0.2435 | 0.1139 | 0.0999 | 0.1666 | 0.2221 |
| 0.50 | 1 | 0.2051 | 0.2681 | -0.0029 | -0.0041 | 0.2051 | 0.2680 |
| | 2 | 0.1876 | 0.2511 | -0.0017 | -0.0029 | 0.1876 | 0.2511 |
| | 4 | 0.1359 | 0.1985 | 0.0002 | -0.0020 | 0.1359 | 0.1985 |
| 0.70 | 1 | 0.2866 | 0.3334 | -0.1940 | -0.1915 | 0.2109 | 0.2729 |
| | 2 | 0.2640 | 0.3104 | -0.1693 | -0.1641 | 0.2025 | 0.2635 |
| | 4 | 0.2043 | 0.2448 | -0.1151 | -0.1029 | 0.1689 | 0.2222 |
| 0.85 | 1 | 0.4018 | 0.4394 | -0.3405 | -0.3386 | 0.2134 | 0.2800 |
| | 2 | 0.3913 | 0.4224 | -0.3265 | -0.3176 | 0.2157 | 0.2785 |
| | 4 | 0.3438 | 0.3524 | -0.2678 | -0.2275 | 0.2155 | 0.2692 |

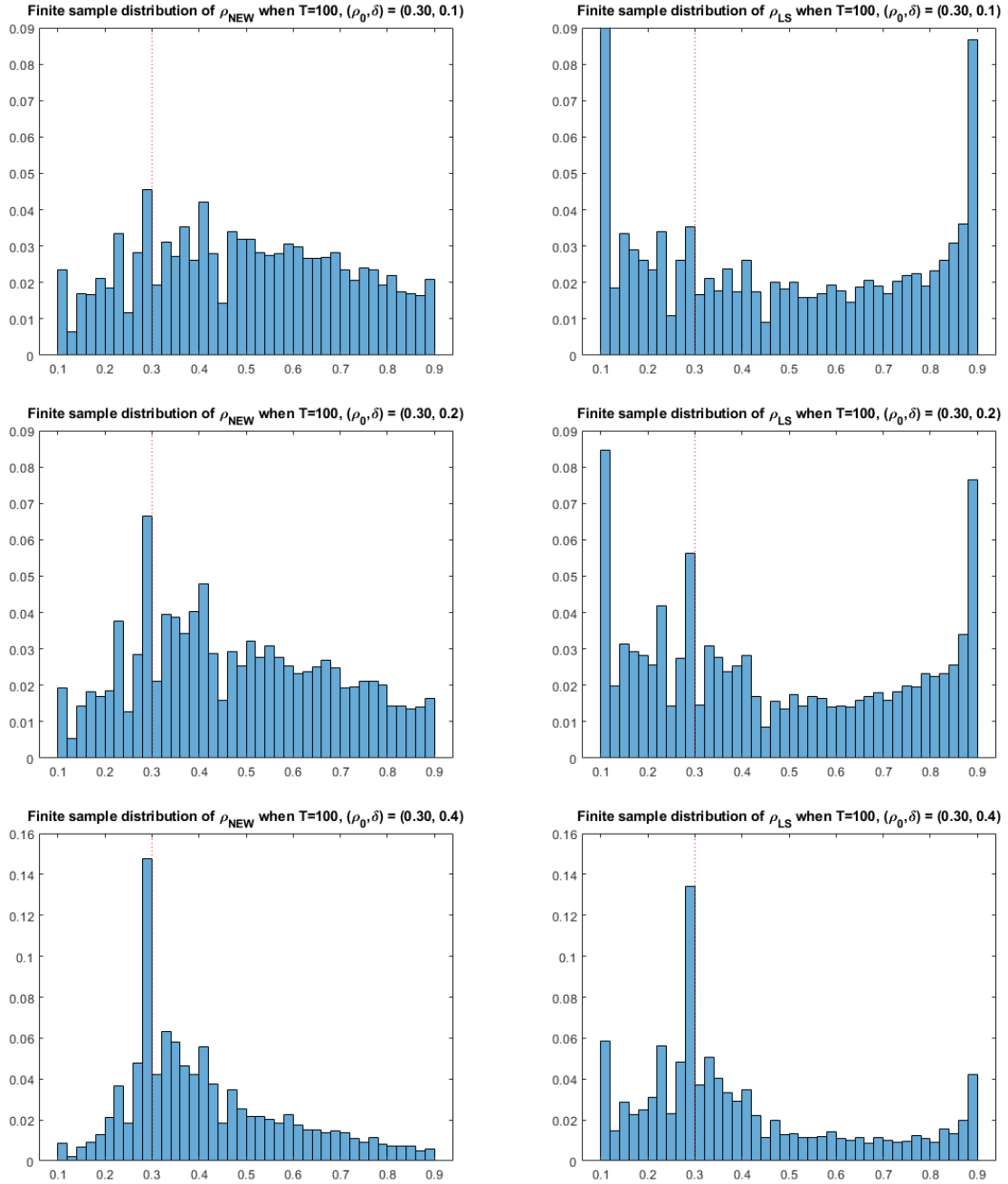


Figure 3: ($\rho_0 = 0.30$) Finite sample distribution of the new estimator $\hat{\rho}_{NEW}$ (left) and the LS estimator $\hat{\rho}_{LS}$ (right) under model (13), with parameter values $(\rho_0, \delta_T) = (0.3, T^{-1/2})$, $(0.3, 2T^{-1/2})$, and $(0.3, 4T^{-1/2})$ and $T = 100$, respectively. The optimization space is trimmed by fraction α on both ends.

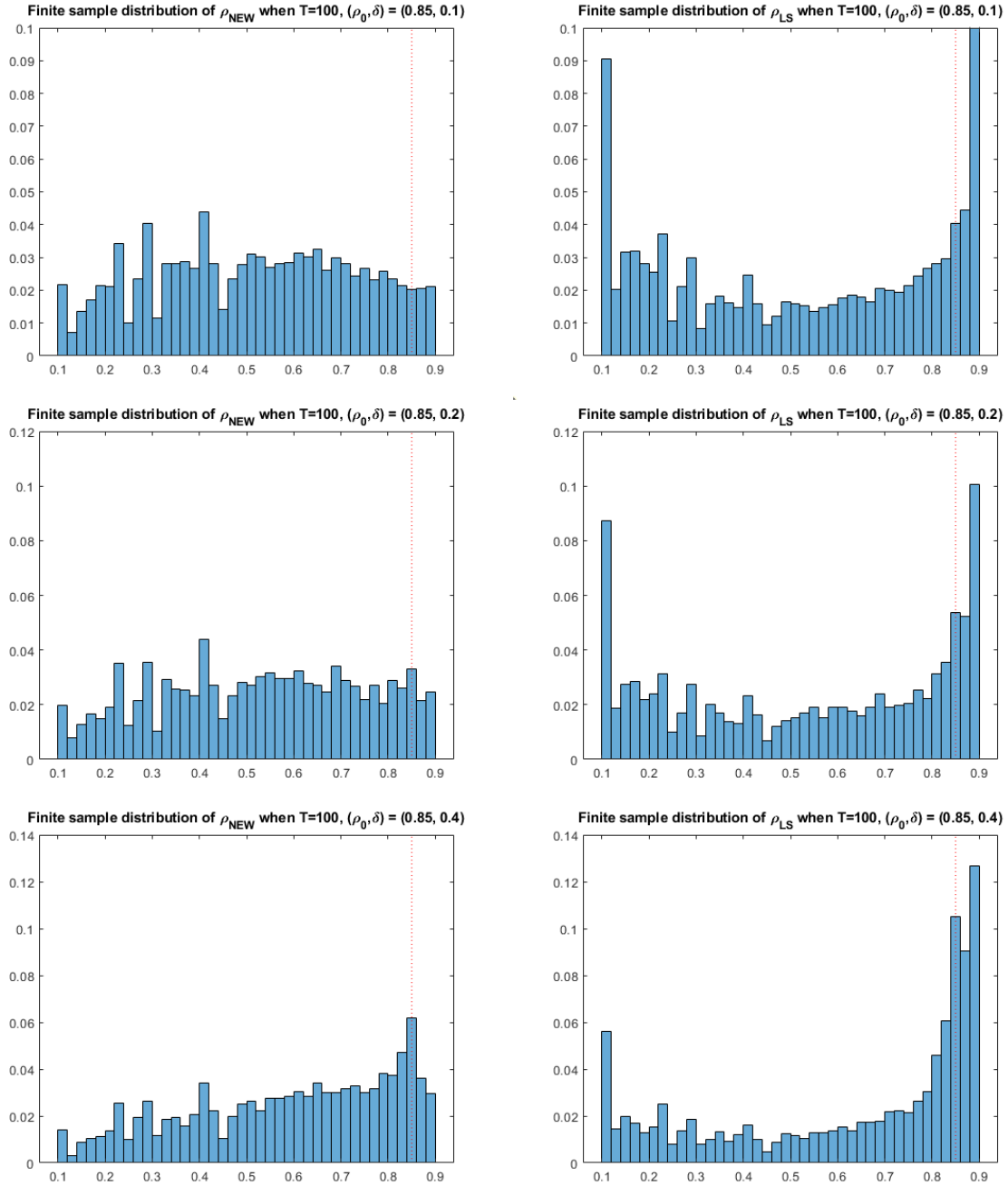


Figure 4: ($\rho_0 = 0.85$) Finite sample distribution of the new estimator $\hat{\rho}_{NEW}$ (left) and the LS estimator $\hat{\rho}_{LS}$ (right) under model (13), with parameter values $(\rho_0, \delta_T) = (0.85, T^{-1/2})$, $(0.85, 2T^{-1/2})$, and $(0.85, 4T^{-1/2})$ and $T = 100$, respectively. The optimization space is trimmed by fraction α on both ends.

5.2 Autoregressive process

For the AR(1) process, I replicate two experiments from Jiang et al. (2017). 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 $h = 1/200$ ($T = 200$), $\sigma = 1$, $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, 1)$, $\rho_0 \in \{0.3, 0.5, 0.7\}$, and different combinations of μ and δ 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-unity 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 $h = 0.005$. The optimization space is trimmed by fraction $\alpha = 0.1$. The break point estimator $\hat{\rho}_{NEW}$ of the AR(1) model is defined in (11) and its asymptotic distribution 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. (2017).

Table 2 provides the RMSE, bias, and the standard error of $\hat{\rho}_{NEW}$ and $\hat{\rho}_{LS}$ for the finite sample, respectively (see Table A.2 in Appendix C for the asymptotic distribution). Similar to results in section 5.1, the RMSE of $\hat{\rho}_{NEW}$ is smaller than that of $\hat{\rho}_{LS}$ for all parameter values $(\beta_1, \beta_2, \rho_0)$ considered. This also holds in the limit. A decrease in RMSE of $\hat{\rho}_{NEW}$ 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 $\hat{\rho}_{LS}$ mode at the true break point is almost negligible, unless it is the median $\rho_0 = 0.5$. In contrast, the estimator $\hat{\rho}_{NEW}$ 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 at the ends, whereas the new estimator has a unique mode at ρ_0 .

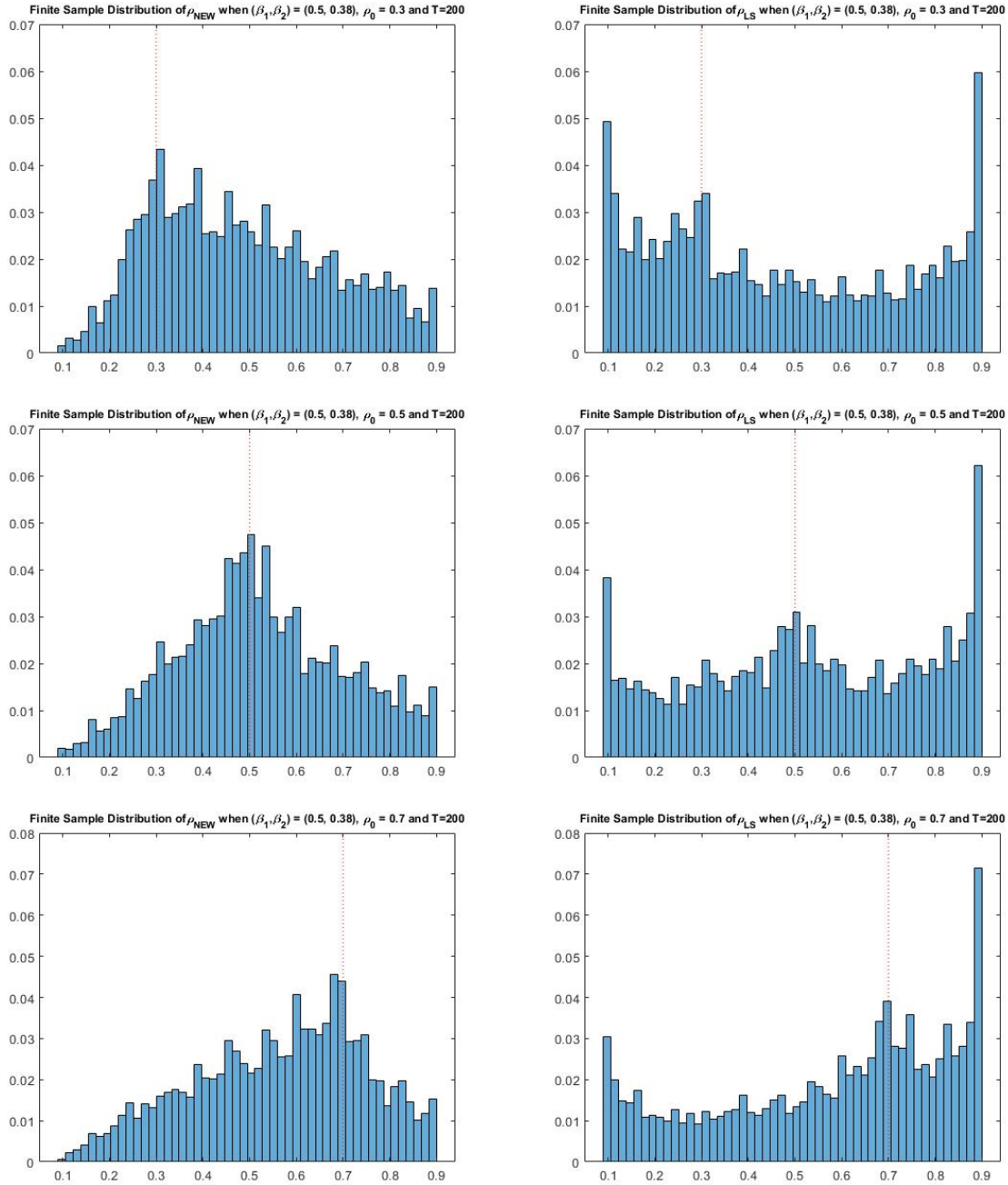


Figure 5: (Stationary to stationary) Finite sample distributions of the new estimator (left) and the LS estimator (right) when the lag coefficient pre- and post-break are $(\beta_1, \beta_2) = (0.5, 0.38)$ at break points $\rho_0 = 0.3, 0.5$, and 0.7 , respectively.

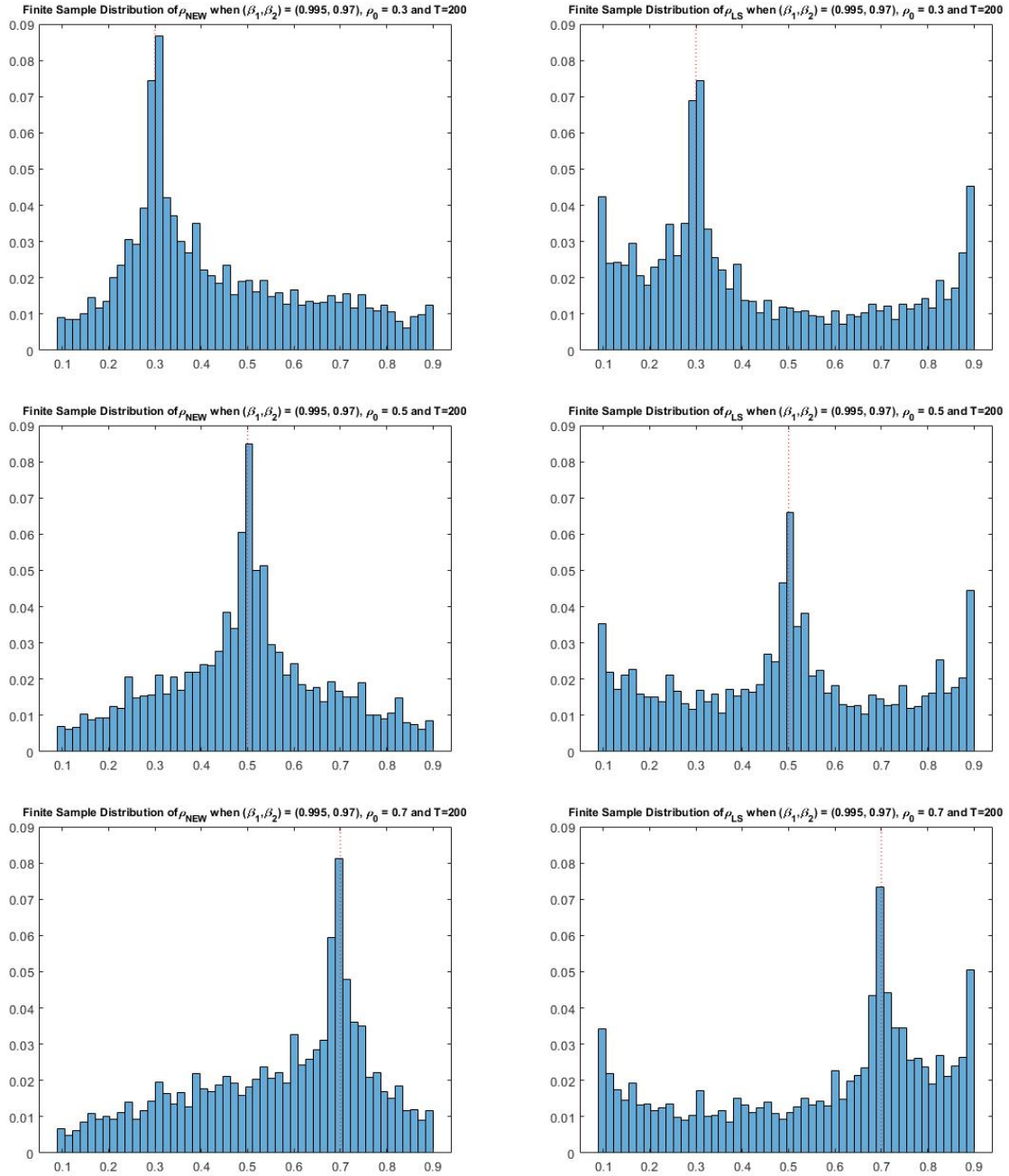


Figure 6: (Local-to-unity to stationary) Finite sample distributions of the new estimator (left) and the LS estimator (right) when the lag coefficient pre- and post-break are $(\beta_1, \beta_2) = (0.995, 0.97)$ at break points $\rho_0 = 0.3, 0.5$, and 0.7 , respectively.

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 $T = 200$. The number of replications is 5,000.

| β_1 | β_2 | ρ_0 | RMSE | | Bias | | Standard error | |
|-----------|-----------|----------|--------|--------|---------|---------|----------------|--------|
| | | | NEW | LS | NEW | LS | NEW | LS |
| 0.5 | 0.38 | 0.3 | 0.2627 | 0.3091 | 0.1821 | 0.1657 | 0.1893 | 0.2610 |
| | | 0.5 | 0.1763 | 0.2452 | 0.0204 | 0.0282 | 0.1751 | 0.2436 |
| | | 0.7 | 0.2285 | 0.2725 | -0.1379 | -0.1223 | 0.1822 | 0.2435 |
| 0.995 | 0.97 | 0.3 | 0.2369 | 0.2780 | 0.1319 | 0.1279 | 0.1967 | 0.2469 |
| | | 0.5 | 0.1754 | 0.2328 | -0.0042 | -0.0047 | 0.1754 | 0.2327 |
| | | 0.7 | 0.2375 | 0.2784 | -0.1358 | -0.1336 | 0.1948 | 0.2442 |

6 Empirical Application

In this section, I estimate the structural breaks in two empirical applications. I analyze the performance of the break point estimator by comparing it with the LS estimator and historical events documented in the literature. Furthermore, I show that the estimator is robust to trimming the sample period, whereas the LS estimator varies significantly depending on the trimmed sample. The first application is about the structural break in postwar U.S. real GDP growth rate. The second application is estimating the break date on the U.S. and UK stock returns using the return prediction model from Paye and Timmermann (2006).

6.1 U.S. real GDP growth rate

In the macroeconomics literature, shocks that affect the mean growth rate are often modeled as a one-time structural break because of their rare occurrence. However, existing estimation methods fail to capture the graphical evidence of postwar European and U.S. growth, slowing sometime in the 1970s. This is known as the “productivity growth slowdown,” which is widely hypothesized in macroeconomics literature. For instance, Bai, Lumsdaine, and Stock (1998) show that for the U.S., most test statistics reject the no-break hypothesis; however, the estimated confidence interval does not contain the slowdown in the 1970s.

I estimate the structural break of an autoregressive model using the postwar quarterly U.S. real GDP growth rate. I use real GDP in chained dollars (base year 2012) data from the Bureau of Economic Analysis (BEA) website for the sample period 1947Q1-2018Q2, seasonally adjusted at annual rates. Annualized quarterly growth rates are calculated as 400 times the first differences of the natural logarithms of the levels data. I assume that log output has a stochastic trend with a drift and a finite-order representation. Following the Eo and Morley (2015) approach, I use Kurozumi and Tuvaandorj’s (2011) modified Bayesian information criterion (BIC) for lag selection to account for structural breaks. The highest lag order selected is 1 for output growth, given an upper bound of four lags and four breaks. The AR(1) model (14) is estimated under three cases. The first case is a break only in the drift term ($\gamma \neq 0, \delta = 0$), the second case is a break only in the coefficient of lag, the “propagation term” ($\gamma = 0, \delta \neq 0$), and finally, a break in constant and coefficient.

$$\Delta y_t = \beta + \phi \Delta y_{t-1} + \mathbf{1}\{t > k_0\} (\gamma + \delta \Delta y_{t-1}) + \varepsilon_t. \quad (14)$$

I assume the error term $\{\varepsilon_t\}$ is serially uncorrelated with mean zero disturbances. If a structural break occurs in constant and lag coefficients, the long-run growth rate of log output will change from $E[\Delta y_t] = \beta/(1 - \phi)$ to $(\beta + \gamma)/(1 - \phi - \delta)$ and the volatility of the growth rate will change from $Var[\Delta y_t] = \sigma^2/(1 - \phi^2)$ to $\sigma^2/(1 - (\phi + \delta)^2)$ at time k_0 .

Using the notations of model (4), we have $x_t = (1, \Delta y_{t-1})'$, the dependent variable is Δy_t , and for each model $z_t = R'x_t$ is as follows:

- M1: $R = (1, 0)'$, $z_t = 1$
- M2: $R = (0, 1)'$, $z_t = \Delta y_{t-1}$
- M3: $R = (1, 1)'$, $z_t = (1, \Delta y_{t-1})'$

The break point estimator \hat{k}_{NEW} in (5) is obtained using the weight $\Omega_k = \omega_k^2 I_q$, where $\omega_k = (k/T(1 - k/T))^{1/2}$ and $q = \dim(z_t)$ ⁵. I use the full sample 1947Q3-2018Q2 ($T = 284$) to estimate the structural break date, then a shorter sub-sample to see if the break date estimates change. The optimization space of k is the sample trimmed by fraction $\alpha = 0.1$

⁵For consistency of the break point estimator in a AR(1) model, we use $\Omega_k = \omega_k^2 I_q$, where ω_k is a function of k/T only.

at both ends; the grid starts at 1954Q2 and ends at 2011Q1. The second and third columns of Table 3 show the break date estimates for the full sample. For M1 and M3, the two estimates are extremely different from each other, \hat{k}_{NEW} is 1973Q1, whereas \hat{k}_{LS} is 2000Q2. The break point estimates on the unit interval are approximately 0.36 and 0.75, respectively. Without any knowledge of historical events, one might think the finite sample properties of $\hat{\rho}_{LS}$ do not appear here, because it is not close to the boundaries of 0.1 or 0.9.

However, the LS estimate switches to the boundary if we consider a sub-sample that is one decade shorter. Consider a sub-sample that ends at 2007Q1, with starting date 1947Q3, so that the search grid includes \hat{k}_{LS} from all models. The LS estimate of M1 changes drastically to 1953Q1, which is the end of the search grid, $\hat{\rho}_{LS} = 0.1$. In contrast, my estimator under M1 provides the same break date estimate $\hat{k}_{NEW} = 1973Q1$. For M3, both estimates change, so $\hat{k}_{NEW} = 1966Q1$ and $\hat{k}_{LS} = 1958Q1$. Compared to the full sample estimate, the change in the new estimate is 7 years, whereas it is over 40 years for the LS estimate.

The break date estimate $\hat{k}_{NEW} = 1973Q1$ under M1 corresponds to the productivity growth slowdown in the early 1970s. U.S. labor productivity experienced a slowdown in growth after the oil shock in 1973 (see Perron (1989) and Hansen (2001)). None of the models estimate a break date in the 1980s, which is known as “the Great Moderation,” referencing an empirical fact of a large reduction in the volatility of U.S. real GDP growth in 1984Q1, established by Kim and Nelson (1999) and McConnell and Perez-Quiros (2000). I focus on events that affect the mean rather than the volatility of growth rate because the change in volatility is not a linear function of the change in the lag coefficient in model (14).

Table 3: Structural break date estimates of postwar U.S. real GDP growth rate in a AR(1) model. 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).

| Model | 1947Q3-2018Q2 | | 1947Q3-2007Q1 | |
|-------|---------------|--------|---------------|--------|
| | NEW | LS | NEW | LS |
| M1 | 1973Q1 | 2000Q2 | 1973Q1 | 1953Q1 |
| | 0.36 | 0.75 | 0.43 | 0.10 |
| M2 | 1966Q1 | 1966Q1 | 1966Q1 | 1966Q1 |
| | 0.26 | 0.26 | 0.32 | 0.32 |
| M3 | 1973Q1 | 2000Q2 | 1966Q1 | 1958Q1 |
| | 0.36 | 0.75 | 0.32 | 0.18 |

I check the sensitivity of the estimators to trimming the sample, by computing the two estimators for a total of 54 sub-samples, which end at different dates from 2005Q1 to 2018Q2 (the samples used to obtain estimates differ, not the fraction α of the optimization space within the sample). Because the sample is trimmed by one quarter each time, switching to a different estimate, which is farther, implies that the estimator is sensitive to trimming, rather than suggesting multiple breaks⁶. Estimating multiple structural breaks using the weighting scheme is beyond the scope of this study. For LS estimation of multiple breaks see Bai and Perron (1998) and Bai et al. (1998).

Table 4 provides the number of sub-samples that have the same break date estimates; the entries are fractions of the number of sub-samples out of 54 sub-samples. For M1 and M3, there are sub-samples in which the LS estimates are at the boundaries $\hat{\rho}_{LS} = 0.1$ or 0.9 . In contrast our break point estimates are either mid 1960s or early 1970s, which are in the fraction interval $\hat{\rho} \in [0.2, 0.5]$.

In short, estimating a structural break of postwar U.S. real GDP growth rate using our estimation method, provides evidence of a break occurring in 1973Q1, which corresponds to

⁶It is likely that multiple structural breaks exist in the output growth rate because we are considering a sample that is over 70 years. Estimates that vary depending on the sub-sample could be evidence of more than one break in the sample period. The break date estimate $\hat{k}_{LS} = 2000Q2$ can align with the tech bubble, also known as the dot-com crash in 2000. In relation to business cycles, $\hat{k}_{LS} = 1953Q1$ and $1958Q1$ are both recession in 1953, with the end of the Korean war. Under M2, both estimates from the full sample are 1966Q1, and the closest historical event that is likely to affect the output growth rate is the Vietnam war.

the productivity growth slowdown period. However, the LS estimates a break occurs in 2000 or 1953, depending on the time interval. Break date estimates are obtained for sub-samples with end dates 2005Q1 to 2018Q2 for both methods; the LS estimates vary considerably, with $\hat{\rho}_{LS}$ near 0.1 and 0.9 for almost 40% of the sub-samples considered under M1. In contrast, my estimates are 1966Q1 or 1973Q1 for all sub-samples and models. This suggests that the difference in LS estimates, depending on the sample period, are due to their finite sample behavior (tri-modality) rather than evidence of multiple structural breaks.

Table 4: Structural break date estimates of postwar U.S. real GDP growth rate in an AR(1) model using 54 sub-samples. The entries are the fractions of the number of sub-samples that have break date estimates corresponding to the first column. The second column is the interval of break point estimates that depend on sub-sample size. The start date is 1947Q3 and the end dates of sub-samples change across 2005Q1 to 2018Q2.

| Break date | $\hat{\rho}$ interval | M1 | | M2 | | M3 | |
|------------|-----------------------|-----|------|-----|----|------|------|
| | | NEW | LS | NEW | LS | NEW | LS |
| 1953Q1 | 0.10 | | 0.17 | | | | |
| 1958Q1 | [0.17, 0.19] | | | | | | 0.22 |
| 1966Q1 | [0.26, 0.33] | | 0.06 | 1 | 1 | 0.30 | 0.09 |
| 1973Q1 | [0.36, 0.45] | 1 | 0.05 | | | 0.70 | |
| 2000Q2 | [0.74, 0.86] | | 0.52 | | | | 0.48 |
| 2006Q1 | [0.85, 0.90] | | 0.20 | | | | 0.21 |

6.2 Stock return prediction models

Paye and Timmermann (2006) 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 (15) 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 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 $t = 1, \dots, T$ is

$$\begin{aligned}
 Ret_t = & \beta_0 + \beta_1 Div_{t-1} + \beta_2 Tbill_{t-1} + \beta_3 Spread_{t-1} + \beta_4 Def_{t-1} \\
 & + \mathbf{1}\{t > k\} (\delta_0 + \delta_1 Div_{t-1} + \delta_2 Tbill_{t-1} + \delta_3 Spread_{t-1} + \delta_4 Def_{t-1}) + \varepsilon_t,
 \end{aligned} \tag{15}$$

where Ret_t represents the excess return for the international index in question during month t , Div_{t-1} is the lagged dividend yield, $Tbill_{t-1}$ is the lagged local country short interest rate, $Spread_{t-1}$ is the lagged local country spread, and Def_{t-1} is the lagged US default premium. From the notation of model (4), $y_t = Ret_t$ and for the multivariate model, $x_t = z_t = (1, Div_{t-1}, Tbill_{t-1}, Spread_{t-1}, Def_{t-1})$. For the univariate model with dividend yield $x_t = z_t = (1, Div_{t-1})$, which is defined analogously for other univariate models. The weight matrix is $w_k = T^{-1}Z'_k M Z_k$, where $Z_k = (0, \dots, 0, z_{k+1}, \dots, z_T)'$ and $M = I - X(X'X)^{-1}X'$. Following the approach of Paye and Timmermann (2006), I examine univariate models to facilitate the interpretation of coefficients, in addition to the multivariate model (15).

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 UK. 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 the short-term local country 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 sub-sample 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 (15), using data similar to the first dataset from Paye and Timmermann (2006), which is monthly data on the U.S. and the UK 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 5 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 sub-sample. 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 (15), 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 5, which assume a one-time structural break.

Table 5: 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. 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).

| Model | 1952:7-2003:12 | | 1970:1-2003:12 | |
|------------|-----------------|-----------------|----------------|-----------------|
| | NEW | LS | NEW | LS |
| Multi. | 1984:8 0.62 | 1994:12 0.83 | 1982:8 0.37 | 1994:12 0.74 |
| Div. yield | 1982:8 0.59 | 1995:1 0.83 | 1982:8 0.37 | 1996:9 0.79 |
| T-bill | 1974:10 0.43 | 1974:10 0.43 | 1982:8 0.37 | 1975:1 0.15 |
| Spread | 1983:5 0.60 | 1976:2 0.46 | 1987:9 0.52 | 1976:2 0.18 |
| Def.prem. | 1968:12 0.32 | 1965:11 0.26 | 1982:8 0.37 | 1975:7 0.16 |

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 (15), 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 Div_{t-1} (see Figure A.1 in Appendix C).

For UK stock returns, both methods obtain a break date estimate that is (or close to) 1975:1 under all models and sample periods. This is different from the result using the S&P 500 index series, because the excess return for the FTSE All-share index increases nearly 10 standard deviations from 1975:1 to 1975:2 ⁷. Hence, the change in excess returns is large enough for the LS to detect the break point appropriately. Paye and Timmermann (2006) relate the break in the mid-1970s to the large macroeconomic shocks reflecting large oil price increases; breaks in the underlying economic fundamentals process can explain breaks in financial return models. If this is the case, then the break magnitude is large enough for

⁷For the sample period 1952:7-2003:12, the mean excess return of the FTSE All-share index is 0.5949 and the standard deviation is 5.4890. At $t = 1975:1$, the excess return $Ret_t = 0.4556$ and at $t = 1975:2$ we have $Ret_t = 53.2187$; therefore, the change is approximately 9.6 standard deviations.

both methods to accurately estimate the break date 1975:1.

Table 6: Structural break date estimates of the UK (FTSE) stock return prediction model for samples 1952:7-2003:12, and 1970:1-2003:12. 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).

| Model | 1952:7-2003:12 | | 1970:1-2003:12 | |
|------------|----------------|-----------------|----------------|----------------|
| | NEW | LS | NEW | LS |
| Multi. | 1975:1 0.44 | 1975:1 0.44 | 1975:1 0.15 | 1975:1 0.15 |
| Div. yield | 1975:1 0.44 | 1975:1 0.44 | 1975:1 0.15 | 1975:1 0.15 |
| T-bill | 1975:1 0.44 | 1974:12 0.29 | 1975:1 0.15 | 1975:1 0.15 |
| Spread | 1975:1 0.44 | 1975:6 0.45 | 1975:1 0.15 | 1975:3 0.15 |
| Def.prem. | 1979:5 0.52 | 1975:6 0.45 | 1975:1 0.52 | 1975:3 0.15 |

7 Conclusion

This study provides an estimation method of 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 parameters is small in a statistical sense. 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 that 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 derives under an in-fill asymptotic framework, following the approach by Jiang et al. (2017, 2018). For a structural break in a stationary linear process with a small break magnitude (inside the local $T^{-1/2}$ neighborhood of zero), the asymptotic distribution of the new estimator explicitly depends on the weight function. The limit distribution is also derived for a break in a local-to-unit root process, assuming the break magnitude is $O(T^{-1})$. Monte Carlo simulation results show that for a small break, the break point estimator reduces the RMSE compared to the LS estimator for all parameter values considered.

This study provides two empirical applications as follows: structural breaks on the U.S. real GDP growth and the U.S. and UK stock return prediction models. My break point estimator is robust to trimming of the sample, in contrast to the LS. In particular, my method estimates the break date 1973Q1 in U.S. real GDP growth rates, which LS estimation has failed to confirm. In macroeconomics literature, the “productivity growth slowdown” in the early 1970s is a widely known empirical fact.

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 uni-modal 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. The estimation method can be generalized to estimate multiple structural breaks, which is a topic for future research.

Bibliography

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Anderson, T., and Darling, D. (1954). “A test of goodness of fit.” *Journal of the American Statistical Association*, 49(268), 765–769.

- Andrews, D. (1988). “Laws of large numbers for dependent nonidentically distributed random variables.” *Econometric Theory*, 4, 458–466.
- Andrews, D., and Cheng, X. (2012). “Estimation and inference with weak, semi-strong, and strong identification.” *Econometrica*, 80(5), 2153–2211.
- Andrews, D., and Cheng, X. (2013). “Maximum likelihood estimation and uniform inference with sporadic identification failure.” *Journal of Econometrics*, 173(1), 36–56.
- Andrews, D. W. K., and Cheng, X. (2014). “GMM estimation and uniform subvector inference with possible identification failure.” *Econometric Theory*, 30(2), 287–333.
- Bai, J. (1994). “Least squares estimation of a shift in linear processes.” *Journal of Time Series Analysis*, 15(5), 453–472.
- Bai, J. (1997). “Estimation of a change point in multiple regression models.” *The Review of Economics and Statistics*, 79, 551–563.
- 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.
- Bai, J., and Perron, P. (1998). “Estimating and testing linear models with multiple structural changes.” *Econometrica*, 66, 47–78.
- Bai, J., and Perron, P. (2003). “Computation and analysis of multiple structural changes.” *Journal of Applied Econometrics*, 18, 1–22.
- 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.
- Birnbaum, Z. W., and Marshall, A. W. (1961). “Some multivariate chebyshev inequalities with extensions to continuous parameter processes.” *Annals of Mathematical Statistics*, 32, 687–703.
- Card, D., Mas, A., and Rothstein, J. (2008). “Tipping and the dynamics of segregation.” *The Quarterly Journal of Economics*, 123(1), 177–218.

- Casini, A., and Perron, P. (2019). “Continuous record Laplace-based inference about the break Date in structural models.”, arXiv preprint arXiv:1804.00232.
- Casini, A., and Perron, P. (2020). “Generalized Laplace inference in multiple change-point models.”, arXiv preprint arXiv:1803.10871.
- Chong, T. T. L. (2001). “Structural change in AR(1) models.” *Econometric Theory*, 17(1), 87–155.
- 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(2), 1196–1218.
- Elliott, G., and Müller, U. K. (2014). “Pre and post break parameter inference.” *Journal of Econometrics*, 180(2), 141–157.
- Eo, Y., and Morley, J. (2015). “Likelihood-ratio-based confidence sets for the timing of structural breaks.” *Quantitative Economics*, 6, 463–497.
- González-Val, R., and Marcén, M. (2012). “Unilateral divorce versus child custody and child support in the u.s.” *Journal of Economic Behavior & Organization*, 81, 613–643.
- Hájek, R., and Rényi, A. (1955). “Generalization of an inequality of kolmogorov.” *Acta Mathematica Hungarica*, 6, 281–283.
- Han, S., and McCloskey, A. (2019). “Estimation and inference with a (nearly) singular jacobian.” *Quantitative Economics*, 10(3), 1019–01068.
- Hansen, B. E. (1991). “Strong laws for dependent heterogeneous processes.” *Econometric Theory*, 7(2), 213–221.
- Hansen, B. E. (2001). “The new econometrics of structural change: Dating breaks in u.s. labor productivity.” *Journal of Economic Perspective*, 15(4), 117–128.
- Hinkley, D. (1970). “Inference about the change point in a sequence of random variables.” *Biometrika*, 57, 1–17.

- Jiang, L., Wang, X., and Yu, S. (2017). “In-fill asymptotic theory for structural break point in autoregression: A unified theory.”, Singapore Management University, School of Economics.
- Jiang, L., Wang, X., and Yu, S. (2018). “New distribution theory for the estimation of structural break.” *Journal of Econometrics*, 205, 156–176.
- Kim, C. J., and Nelson, C. R. (1999). “Has the u.s. economy become more stable? a bayesian approach based on a markov-switching model of the business cycle.” *Review of Economics and Statistics*, 81(4), 608–616.
- Kurozumi, E., and Tuvaandorj, P. (2011). “Model selection criteria in multivariate models with multiple structural changes.” *Journal of Econometrics*, 164(2), 218–238.
- McConnell, M., and Perez-Quiros, G. (2000). “Output fluctuations in the united states: What has changed since the early 1980’s?” *American Economic Review*, 90(5), 1464–1476.
- McLeish, D. L. (1975). “A maximal inequality and dependent strong laws.” *The Annals of Probability*, 5, 829–839.
- Pang, T., Chong, T., Zhang, D., and Liang, Y. (2018). “Structural change in nonstationary ar (1) models.” *Econometric Theory*, 34(5), 985–1017.
- 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(2), 133–150.
- Paye, B. S., and Timmermann, A. (2006). “Instability of return prediction models.” *Journal of Empirical Finance*, 13, 274–315.
- Perron, P. (1989). “The great crash, the oil-price shock, and the unit-root hypothesis.” *Econometrica*, 57(6), 1361–1401.
- Yao, Y. C. (1987). “Approximating the distribution of the ml estimate of the change-point in a sequence of independent r.v.’s.” *The Annals of Statistics*, 3, 1321–1328.

Appendix A

In a Bayesian perspective, the weight function can be interpreted as a prior belief on parameters δ and ρ . Suppose the prior distribution of the break magnitude δ conditional on ρ is normal, with mean zero and variance ν^2 , $\delta|\rho \sim N(0, \nu^2)$. Denote the prior distribution of the break point as $f(\rho)$ and let $\mathbf{y} = (y_1, \dots, y_T)$. We assume Gaussian disturbances, $\varepsilon_t \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. The joint distribution of δ , ρ , and \mathbf{y} is

$$f(\mathbf{y}, \delta, \rho) = (2\pi)^{-1}(\nu^2 I_T^{-1})^{-1/2} \exp \left[-\frac{1}{2} \left\{ \frac{\delta^2}{\nu^2} + \frac{(\bar{y}_k^* - \bar{y}_k - \delta)^2}{I_T^{-1}} \right\} \right] f(\rho),$$

where $I_T = I_T(\delta|\rho)$ is the Fisher information of δ conditional on ρ and $l_T(\delta|\rho)$ is the conditional log-likelihood function.

$$I_T := E \left[-\frac{\partial^2 l_T(\delta|\rho)}{\partial \delta \partial \delta'} \right] = \sigma^{-2} T \rho(1 - \rho).$$

The posterior distribution of ρ is proportional to $f(\mathbf{y}, \delta, \rho)$, integrated with respect to δ .

$$f(\rho|\mathbf{y}) \propto f(\rho)(I_T^{-1} + \nu^2)^{-1/2} \exp \left[-\frac{(\bar{y}_k^* - \bar{y}_k)^2}{2(I_T^{-1} + \nu^2)} \right].$$

We assume ρ is bounded away from $\{0, 1\}$ so that $I_T^{-1} = o(1)$, and for simplicity, assume $\sigma^2 = 1$. Suppose $\nu = T^{-1} I_T = \rho(1 - \rho)$ and $f(\rho) \propto \nu$. In the limit, the posterior distribution is proportional to

$$f(\rho|\mathbf{y}) \propto \exp \left[-\frac{(\bar{y}_k^* - \bar{y}_k)^2}{2\rho^2(1 - \rho)^2} \right]. \quad (\text{A.1})$$

Given the data \mathbf{y} , the argmax function of the monotone transformation of the likelihood (A.1) is asymptotically equivalent to the argmax of $Q_T(k)^2 = \omega_k^2 V_T(k)^2$, with $\omega_k = (k/T(1 - k/T))^{1/2}$. That is, if the weight function is proportional to the square root of the Fisher information of δ conditional on ρ , the mode of the Bayesian posterior distribution (known as the maximum a-posteriori probability (MAP) estimator) is asymptotically equivalent to the break point estimator in (3).

Note that the Fisher information is interpreted as a way to measure the amount of information about the unknown parameter δ , given ρ . Given two different values $\rho_1 \neq \rho_2$,

the inequality $I_T(\delta|\rho_1) > I_T(\delta|\rho_2)$ reflects the fact that observations carry more information on the break magnitude if a break occurs at ρ_1 , compared to ρ_2 . In other words, there is more information of a structural break occurring at ρ_1 than at ρ_2 . If a break occurs with high probability, its magnitude is likely to be far from zero; if it is not likely, then δ is close to zero. If we scale the Fisher information by sample size ($I := T^{-1}I_T$), then the prior distribution $\delta|\rho \sim N(0, I)$ incorporates the amount of information of a break at some fixed ρ . A large I implies the variance of the prior distribution is large, thus δ is more spread out from zero and has large magnitude with high probability. In the opposite case, if there is less information, δ is centered toward mean zero and the break magnitude is likely to be small. Similarly, a prior belief of ρ can be expressed using the Fisher information ($f(\rho) \propto I$ is equivalent to a Beta distribution with shape parameters $(2, 2)$).

Appendix B

Proof of Theorem 1

Proof. By rearranging terms we have

$$Q_T(k)^2 - Q_T(k_0)^2 = -|k_0 - k|G_T(k) + H_T(k), \quad (\text{A.2})$$

where $G_T(k)$ and $H_T(k)$ are defined as follows:

$$\begin{aligned} G_T(k) := & \frac{1}{|k_0 - k|} \delta'_T \left[(Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{1/2} \right. \\ & \left. - (Z'_0 M Z_k)(Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} (Z'_k M Z_0) \right] \delta_T \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} H_T(k) := & \varepsilon' M Z_k (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M \varepsilon \\ & - \varepsilon' M Z_0 (Z'_0 M Z_0)^{-1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_0 M \varepsilon \\ & + 2\delta'_T (Z'_0 M Z_k)(Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M \varepsilon \\ & - 2\delta'_T (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_0 M \varepsilon. \end{aligned} \quad (\text{A.4})$$

Lemma A.2 is equivalent to Bai's (1997) lemma A.3, which is the generalized Hájek-Rényi inequality for martingale differences to mixingales. For the proof see Bai and Perron (1998).

Lemma A.1. *Under Assumptions 1 and 2, for every $\epsilon > 0$, there exists $\lambda > 0$ and $C < \infty$ such that $\inf_{|k-k_0|>C\|\delta_T\|^{-2}} G_T(k) \geq \lambda \|\delta_T\|^2$, with probability at least $1 - \epsilon$.*

Lemma A.2. *Under Assumption 1, there exist a $M < \infty$ such that for every $c > 0$ and $m > 0$,*

$$P\left(\sup_{m \leq k \leq T} \frac{1}{k} \left\| \sum_{t=1}^k z_t \varepsilon_t \right\| > c\right) \leq \frac{M}{c^2 m}.$$

Lemma A.3. *Under Assumptions 1 and 2, suppose δ_T is fixed or shrinking toward zero such that Assumption 3 is satisfied. Then the break point estimator $\hat{\rho}$ in (5) is consistent. That is, for every $\epsilon > 0$ and $\eta > 0$, there exists $T_0 > 0$ such that when $T > T_0$,*

$$P(|\hat{\rho} - \rho_0| > \eta) < \epsilon.$$

Moreover, $|\hat{\rho} - \rho_0| = O_p\left(T^{-1/2} \|\delta_T\| \sqrt{\ln T}\right)$.

The rate of convergence of the break point estimator $\hat{\rho}$ in (5) can be improved from lemma A.3. For a fixed $\epsilon > 0$ and $\eta > 0$, inequality (A.5) holds for any true break point $\rho_0 \in [\alpha, 1 - \alpha]$, when T is large.

$$P\left(\sup_{|k-k_0|>T\eta} Q_T(k)^2 \geq Q_T(k_0)^2\right) < \epsilon. \tag{A.5}$$

This is equivalent to lemma A.3 because given the estimator \hat{k} , $Q_T(\hat{k})^2 - Q_T(k_0)^2 \geq 0$ by definition. This implies that to prove the improved rate of convergence $O_p(T^{-1} \|\delta_T\|^{-2})$, it is sufficient to show that for all $\epsilon > 0$, there exists a finite $C > 0$ so that for all $T > T_\epsilon$,

$$P\left(\sup_{k \in K_T(C)} Q_T(k)^2 \geq Q_T(k_0)^2\right) < \epsilon.$$

Here, $K_T(C) = \{k : |k - k_0| > C \|\delta_T\|^{-2}, |k - k_0| \leq T\eta\}$ for some small fraction η . From identity (A.2), $Q_T(k)^2 \geq Q_T(k_0)^2$ is equivalent to $H_T(k)/|k - k_0| \geq G_T(k)$. From lemma

A.1, it is sufficient to prove that

$$P \left(\sup_{k \in K_T(C)} \left| \frac{H_T(k)}{k_0 - k} \right| > \lambda \|\delta_T\|^2 \right) < \epsilon. \quad (\text{A.6})$$

We use the expression $Z_0 = Z_k - Z_\Delta \text{sgn}(k_0 - k)$ to rewrite the third and fourth terms of $H_T(k)$ given in (A.4) as

$$\begin{aligned} & 2\delta'_T \left[(Z'_0 M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M \varepsilon - (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_0 M \varepsilon \right] \\ &= 2\delta'_T \left[(Z'_k M Z_k)^{1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M \varepsilon - (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_k M \varepsilon \right] \\ &\quad + 2\delta'_T (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} (Z'_\Delta M \varepsilon) \text{sgn}(k_0 - k) \\ &\quad - 2\delta'_T (Z'_\Delta M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} (Z'_k M \varepsilon) \text{sgn}(k_0 - k). \end{aligned} \quad (\text{A.7})$$

Note that for nonsingular matrices S and A with bounded norms, $SAS^{-1} = A + o_p(1)$. Also, we have $(Z'_k M Z_k)^{-1} Z'_k M \varepsilon = O_p(T^{-1/2})$ and $(Z'_0 M Z_0)^{-1} (Z'_k M Z_k) = O_p(1)$ uniformly on $K_T(C)$. We use this to find the order of the first line of the right-side in (A.7).

$$\begin{aligned} & \left\| 2\delta'_T \left\{ (Z'_k M Z_k)^{1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M \varepsilon - (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_k M \varepsilon \right\} \right\| \\ & \leq \left\| 2\delta'_T \left\{ (Z'_k M Z_k)^{1/2} \Omega_k (Z'_k M Z_k)^{1/2} - (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{1/2} O_p(1) \right\} \right\| \\ & \quad \times \left\| (Z'_k M Z_k)^{-1} Z'_k M \varepsilon \right\| + o_p(1) \\ & \leq \|2\delta_T\| \left\| Z'_k M Z_k \Omega_k - Z'_0 M Z_0 \Omega_{k_0} \right\| O_p(T^{-1/2}) + o_p(1) \\ & = \|2\delta_T\| \left\| (Z'_k M Z_k - Z'_0 M Z_0) \Omega_k - Z'_0 M Z_0 (\Omega_{k_0} - \Omega_k) \right\| O_p(T^{-1/2}) + o_p(1). \end{aligned}$$

Then the second norm can be rearranged by subtracting and adding $Z'_0 M Z_k$ to the term

$(Z'_k M Z_k - Z'_0 M Z_0)$ and Assumption 2.

$$\begin{aligned}
& Z'_k M Z_k - Z'_0 M Z_0 \\
&= \begin{cases} R'[X'_\Delta X_\Delta (X'X)^{-1} (X'X - X'_k X_k) - X'_0 X_0 (X'X)^{-1} X'_\Delta X_\Delta] R & \text{if } k \leq k_0 \\ R'[X'_\Delta X_\Delta (X'X)^{-1} X'_k X_k - (X'X - X'_0 X_0) (X'X)^{-1} X'_\Delta X_\Delta] R & \text{if } k > k_0 \end{cases} \\
&= |k_0 - k| O_p(1), \tag{A.8} \\
\Omega_{k_0} - \Omega_k &= |k_0 - k| T^{-1} O_p(1)
\end{aligned}$$

The norm $\|(k_0 - k)^{-1} X'_\Delta X_\Delta\|$ is bounded by assumption, hence the first line of (A.7) has order $|k_0 - k| \|\delta_T\| O_p(T^{-1/2})$. The second and third lines of (A.7) are

$$\begin{aligned}
& 2\delta'_T (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} (Z'_\Delta M \varepsilon) \operatorname{sgn}(k_0 - k) \\
&= 2\delta'_T \Omega_{k_0} (Z'_\Delta M \varepsilon) \operatorname{sgn}(k_0 - k) + o_p(1) \\
&= 2\delta'_T \Omega_{k_0} (Z'_\Delta \varepsilon - Z'_\Delta X (X'X)^{-1} X' \varepsilon) \operatorname{sgn}(k_0 - k) + o_p(1) \\
&= 2\delta'_T \Omega_{k_0} Z'_\Delta \varepsilon \operatorname{sgn}(k_0 - k) + |k_0 - k| T^{-1/2} \|\delta_T\| O_p(1) + o_p(1), \\
& -2\delta'_T (Z'_\Delta M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} (Z'_k M \varepsilon) \operatorname{sgn}(k_0 - k) \\
&= -2\delta'_T (Z'_\Delta M Z_k) \Omega_k (Z'_k M Z_k)^{-1} (Z'_k M \varepsilon) \operatorname{sgn}(k_0 - k) + o_p(1) \\
&= |k_0 - k| T^{-1/2} \|\delta_T\| O_p(1).
\end{aligned}$$

The first and second terms of $H_T(k)$ in (A.4) are $O_p(1)$ uniformly in $K_T(C)$, under Assumptions 1 and 2. Therefore, $H_T(k)$ divided by $|k_0 - k|$ is

$$\frac{H_T(k)}{|k_0 - k|} = 2\delta'_T \Omega_{k_0} \frac{1}{|k_0 - k|} Z'_\Delta \varepsilon \operatorname{sgn}(k_0 - k) + T^{-1/2} \|\delta_T\| O_p(1) + \frac{O_p(1)}{|k_0 - k|}. \tag{A.9}$$

Now we can prove (A.6) using the expression (A.9). Let $1/\|\Omega_{k_0}\| = A$ where $A < \infty$ by Assumption 2. Without loss of generality, consider the case $k < k_0$. The first term of (A.9) is bounded by lemma A.2.

$$\begin{aligned}
& P \left(\sup_{k \in K_T(C)} \left\| 2\delta_T' \Omega_{k_0} \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} z_t \varepsilon_t \right\| > \frac{\lambda \|\delta_T\|^2}{3} \right) \\
& \leq P \left(\sup_{k_0 - k \geq C \|\delta_T\|^{-2}} \left\| \frac{1}{k_0 - k} \sum_{t=k+1}^{k_0} z_t \varepsilon_t \right\| > \frac{\lambda A \|\delta_T\|}{6} \right) \\
& \leq M \left(\frac{\lambda A \|\delta_T\|}{6} \right)^{-2} \frac{1}{C \|\delta_T\|^{-2}} \\
& = \frac{36M}{\lambda^2 A^2 C} < \frac{\epsilon}{3}
\end{aligned}$$

The probability is negligible for large T because we can choose a large C value accordingly. For any $\epsilon > 0$ and $\eta > 0$, we proved that the probability in (A.5) is negligible for large T . Thus we can choose C such that $K_T(C)$ is non-empty and the inequality above is satisfied for all $\epsilon > 0$ and $T > T_\epsilon$. The second term of (A.9) is bounded due to the assumption $(T^{1/2} \|\delta_T\|)^{-1} \rightarrow 0$.

$$P \left(T^{-1/2} \|\delta_T\| O_p(1) > \frac{\lambda \|\delta_T\|^2}{3} \right) = P \left(\frac{O_p(1)}{T^{1/2} \|\delta_T\|} > \frac{\lambda}{3} \right) < \frac{\epsilon}{3}.$$

The third term of (A.9) is bounded for $k_0 - k \geq C \|\delta_T\|^{-2}$, since $O_p(1)/|k_0 - k| \leq O_p(1) \|\delta_T\|^2 / C$,

$$P \left(\sup_{k_0 - k \geq C \|\delta_T\|^{-2}} \frac{O_p(1)}{|k_0 - k|} > \frac{\lambda \|\delta_T\|^2}{3} \right) \leq P \left(\frac{O_p(1)}{C} > \frac{\lambda}{3} \right) < \frac{\epsilon}{3}$$

where $O_p(1)/C$ is small for large T , by choosing a large constant C . Hence the bound (A.6) holds and the rate of convergence of the break point estimator $\hat{\rho} = \hat{k}/T$ in Theorem 1 is proved: $|\hat{\rho} - \rho_0| = O_p(T^{-1} \|\delta_T\|^2)$. \square

Proof of Lemma A.1

Proof. Assumption 2 implies that $A_T(k)$ from (7) is positive definite and thus $G_T(k) = \delta_T' A_T(k) \delta_T \geq \lambda_T(k) \|\delta_T\|^2$, where $\lambda_T(k)$ is the minimum eigenvalue of $A_T(k)$. It is sufficient to argue that $\lambda_T(k)$ is bounded away from zero with probability tending to 1 as $|k_0 - k|$

increases. The matrices Z'_0MZ_0 and Z'_0MZ_k are rearranged as follows, similar to Z'_kMZ_k in (6).

$$\begin{aligned} Z'_0MZ_0 &= R'(X'_0X_0)(X'X)^{-1}(X'X - X'_0X_0)R \\ Z'_0MZ_k &= \begin{cases} R'(X'_0X_0)(X'X)^{-1}(X'X - X'_kX_k)R & \text{if } k \leq k_0 \\ R'(X'X - X'_0X_0)(X'X)^{-1}(X'_kX_k)R & \text{if } k > k_0 \end{cases} \end{aligned} \quad (\text{A.10})$$

Without loss of generality, assume $k \leq k_0$. The second term of $|k_0 - k|A_T(k)$ from (7) is

$$\begin{aligned} &(Z'_0MZ_k)(Z'_kMZ_k)^{-1/2}\Omega_k(Z'_kMZ_k)^{-1/2}(Z'_kMZ_0) \\ &= [R'(X'_0X_0)(X'X)^{-1}(X'X - X'_kX_k)R] [(Z'_kMZ_k)^{1/2}\Omega_k^{-1}(Z'_kMZ_k)^{1/2}]^{-1} \\ &\quad \times [R'(X'X - X'_kX_k)(X'X)^{-1}(X'_0X_0)R]. \end{aligned} \quad (\text{A.11})$$

Define the following matrices.

$$\begin{aligned} F_k &:= (X'_kX_k)^{-1} - (X'X)^{-1} = (X'X)^{-1}(X'X - X'_kX_k)(X'_kX_k)^{-1} \\ F_0 &:= (X'_0X_0)^{-1} - (X'X)^{-1} = (X'X)^{-1}(X'X - X'_0X_0)(X'_0X_0)^{-1} \\ \Omega_{X,k} &:= \begin{bmatrix} I_{(p-q)} & \mathbf{0}_{(p-q) \times q} \\ \mathbf{0}_{q \times (p-q)} & \Omega_k \end{bmatrix}, \quad B := \Omega_{X,k}^{-1/2}F_k^{1/2}X'_kX_k. \end{aligned} \quad (\text{A.12})$$

Both F_k and F_0 are positive definite matrices under Assumption 1. Hence, each matrix can be decomposed into $F_k = \left(F_k^{1/2}\right)^2$ and $F_0 = \left(F_0^{1/2}\right)^2$ where $F_k^{1/2}$ and $F_0^{1/2}$ are nonsingular $(p \times p)$ matrices with $p = \dim(x_t)$. $\Omega_{X,k}$ is a $(p \times p)$ matrix where $I_{(p-q)}$ is an identity matrix with rank $(p - q)$ and zeros in non-diagonal block matrices such that $R'\Omega_{X,k}R = \Omega_k$. The projection matrix of BR is $I_p - BR(R'B'BR)^{-1}R'B'$, which is positive semi-definite. If we multiply $R'(X'_0X_0)F_k^{1/2}\Omega_{X,k}^{1/2}$ to the left and its transpose to the right of the projection

matrix, the following inequality is obtained:

$$\begin{aligned}
& R'(X'_0 X_0) F_k^{1/2} \Omega_{X,k} F_k^{1/2} (X'_0 X_0) R \\
& \geq R'(X'_0 X_0) F_k (X'_k X_k) R (R' B' B R)^{-1} R' (X'_k X_k) F_k (X'_0 X_0) R \\
& = R'(X'_0 X_0) (X' X)^{-1} (X' X' - X'_k X_k) R (R' B' B R)^{-1} \\
& \quad \times R' (X' X - X'_k X_k) (X' X)^{-1} (X'_0 X_0) R.
\end{aligned}$$

From (A.12), $R' B' B R = R' (X'_k X_k) F_k^{1/2} \Omega_{X,k}^{-1} F_k^{1/2} (X'_k X_k) R = (Z'_k M Z_k)^{1/2} \Omega_k^{-1} (Z'_k M Z_k)^{1/2}$ and the right side of the inequality is equivalent to (A.11). Therefore, it is sufficient to show that the right-side of inequality (A.13) is bounded away from zero for large $k_0 - k$,

$$A_T(k) \geq \frac{1}{|k_0 - k|} \left[(Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{1/2} - R' (X'_0 X_0) F_k^{1/2} \Omega_{X,k} F_k^{1/2} (X'_0 X_0) R \right]. \quad (\text{A.13})$$

Also from (A.12), $(Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{1/2} = R' (X'_0 X_0) F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} (X'_0 X_0) R$. Thus,

$$\begin{aligned}
|k_0 - k| A_T(k) & \geq R' (X'_0 X_0) \left[F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} - F_k^{1/2} \Omega_{X,k} F_k^{1/2} \right] (X'_0 X_0) R \\
& = |k_0 - k| R' \tilde{A}_T(k) R,
\end{aligned}$$

where $\tilde{A}_T(k) := \frac{1}{|k_0 - k|} (X'_0 X_0) \left[F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} - F_k^{1/2} \Omega_{X,k} F_k^{1/2} \right] (X'_0 X_0)$. Then we have

$$\begin{aligned}
\left\| \tilde{A}_T(k)^{-1} \right\| & = |k_0 - k| \left\| (X'_0 X_0)^{-1} \left[F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} - F_k^{1/2} \Omega_{X,k} F_k^{1/2} \right]^{-1} (X'_0 X_0)^{-1} \right\| \\
& \leq \frac{1}{\|X'_0 X_0\|} \frac{1}{|k_0 - k|^{-1} \left\| F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} - F_k^{1/2} \Omega_{X,k} F_k^{1/2} \right\|} \frac{1}{\|X'_0 X_0\|}.
\end{aligned}$$

Note that for a nonsingular, bounded ($p \times p$) matrix S , the norm does not change by multiplying S on the left and S^{-1} on the right of a matrix: $\|\Omega_{X,k}\| = \|S \Omega_{X,k} S^{-1}\|$. By assumption, $\Omega_{X,k} \geq \lambda_{\min}(\Omega_{X,k}) > 0$ where λ_{\min} denotes the minimum eigenvalue. Therefore, $(F_0^{1/2})^{-1} \Omega_{X,k_0} F_0^{1/2} = \Omega_{X,k_0} + o_p(1) \geq \lambda_{\min}(\Omega_{X,k_0}) + o_p(1)$. By subtracting and adding the matrix $F_k (F_0^{1/2})^{-1} \Omega_{X,k_0} F_0^{1/2}$ to the denominator of the second term, the following inequality

holds:

$$\begin{aligned}
& \left\| F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} - F_k^{1/2} \Omega_{X,k} F_k^{1/2} \right\| \\
&= \left\| (F_0 - F_k) (F_0^{1/2})^{-1} \Omega_{X,k_0} F_0^{1/2} - F_k \left\{ (F_k^{1/2})^{-1} \Omega_{X,k} F_k^{1/2} - (F_0^{1/2})^{-1} \Omega_{X,k_0} F_0^{1/2} \right\} \right\| \\
&\geq \left\| (F_0 - F_k) (F_0^{1/2})^{-1} \Omega_{X,k_0} F_0^{1/2} \right\| - \left\| F_k \left\{ (F_k^{1/2})^{-1} \Omega_{X,k} F_k^{1/2} - (F_0^{1/2})^{-1} \Omega_{X,k_0} F_0^{1/2} \right\} \right\| \\
&= \left\| (F_0 - F_k) \Omega_{X,k_0} \right\| - \left\| F_k (\Omega_{X,k} - \Omega_{X,k_0}) \right\| + o_p(1),
\end{aligned}$$

where the inequality is from the inverse triangular inequality. Let $\tilde{\lambda}$ be the minimum value of $\lambda_{\min}(\Omega_{X,k})$ and $\lambda_{\min}(\Omega_{X,k_0})$. From Assumption 2, we have $|k_0 - k|^{-1} \|\Omega_k - \Omega_{k_0}\| \leq b/T$. Hence,

$$\begin{aligned}
& |k_0 - k|^{-1} \left\| F_0^{1/2} \Omega_{X,k_0} F_0^{1/2} - F_k^{1/2} \Omega_{X,k} F_k^{1/2} \right\| \\
&\geq \left| (k_0 - k)^{-1} \left\| (F_0 - F_k) \tilde{\lambda} \right\| - b/T \|F_k\| \right| + |k_0 - k|^{-1} o_p(1) \\
&\geq \tilde{\lambda} \left| (k_0 - k)^{-1} (F_0 - F_k) \right| + o_p(1).
\end{aligned}$$

Let $X_\Delta := \text{sgn}(k_0 - k)(X_k - X_0)$. By rearranging terms similar to (A.12), $F_0 - F_k = (X'_0 X_0)^{-1} (X'_\Delta X_\Delta) (X'_k X_k)^{-1}$ so that

$$\begin{aligned}
\left\| \tilde{A}_T(k)^{-1} \right\| &\leq \frac{1}{\|X'_0 X_0\|^2} \frac{1}{\tilde{\lambda} |k_0 - k|^{-1} \|F_0 - F_k\|} \\
&\leq \frac{1}{\tilde{\lambda} \|X'_0 X_0\|^2 \|(X'_0 X_0)^{-1} (k_0 - k)^{-1} X'_\Delta X_\Delta (X'_k X_k)^{-1}\|}.
\end{aligned}$$

From Assumptions 1 and 2, the right-side of the inequality is bounded:

$\tilde{\lambda} \|T^{-1} X'_0 X_0\|^2 \|T^2 (X'_0 X_0)^{-1} (X'_k X_k)^{-1}\| < M$, for some $M < \infty$. In addition, the minimum eigenvalue of $(k_0 - k)^{-1} (X'_\Delta X_\Delta)$ is bounded away from zero with large probability so that $1/\|(k_0 - k)^{-1} X'_\Delta X_\Delta\|$ is bounded with large probability for all large $k_0 - k$. Thus $\left\| \tilde{A}_T(k)^{-1} \right\|$ is bounded with large probability for all large $k_0 - k$. This implies that the minimum eigenvalue of $\tilde{A}_T(k)$ is bounded away from zero for all large $k_0 - k$ and this is also true for $A_T(k) = R' \tilde{A}_T(k) R$ because R has full column rank. \square

For the proof of Lemma A.3 we use Proposition 1 and Lemma A.4.

Proposition 1. Let $\varepsilon_1, \varepsilon_2, \dots$, be a sequence of martingale differences with $E[\varepsilon_t^2] = \sigma^2$ and $\{c_k\}$ be a decreasing positive sequence of constants. The Hájek-Rényi inequality takes the following form.

$$P \left(\max_{m \leq k \leq T} c_k \left| \sum_{t=1}^k \varepsilon_t \right| > \alpha \right) \leq \frac{\sigma^2}{\alpha^2} \left(mc_m^2 + \sum_{t=m+1}^T c_t^2 \right).$$

Hájek and Rényi (1955) proved the inequality assuming i.i.d. random variables, and was later generalized to martingales by Birnbaum and Marshall (1961). We use the generalized Hájek-Rényi for martingale difference sequences to prove Lemma A.4, where $\{\varepsilon_t, \mathcal{F}_t\}$ are mixingale sequences under Assumption 1.

Lemma A.4. Under Assumption 1, for every $\alpha > 0$ and $m > 0$ there exists $C < \infty$ such that

$$P \left(\sup_{m \leq k \leq T} \frac{1}{\sqrt{k}} \left| \sum_{t=1}^k z_t \varepsilon_t \right| > \alpha \right) \leq \frac{C \ln T}{\alpha^2},$$

and thus, $\sup_k T^{-1/2} \|Z'_k \varepsilon\| = O_p(\sqrt{\ln T})$.

Proof. Denote $\xi_t = z_t \varepsilon_t$ and proceed. Let $\{\xi_t, \mathcal{F}_t\}$ be $(q \times 1)$ L^r -mixingales, $r = 4 + \gamma$ for some $\gamma > 0$ satisfying Assumption 1(vi). Define $\xi_{jt} := E[\xi_t | \mathcal{F}_{t-j}] - E[\xi_t | \mathcal{F}_{t-j-1}]$. Then $\xi_t = \sum_{j=-\infty}^{\infty} \xi_{jt}$, and hence $\sum_{t=1}^k \xi_t = \sum_{j=-\infty}^{\infty} \sum_{t=1}^k \xi_{jt}$. Denote $\|\cdot\|_s$ for the L^s -norm. For each $T > 0$,

$$P \left(\sup_{m \leq k \leq T} \frac{1}{\sqrt{k}} \left\| \sum_{t=1}^k \xi_t \right\|_2 > \alpha \right) \leq P \left(\sum_{j=-\infty}^{\infty} \sup_{m \leq k \leq T} \frac{1}{\sqrt{k}} \left\| \sum_{t=1}^k \xi_{jt} \right\|_2 > \alpha \right). \quad (\text{A.14})$$

For each j , $\{\xi_{jt}, \mathcal{F}_{t-j}\}$ forms a sequence of martingale difference and the generalized Hájek-Rényi inequality (Proposition 1) holds for this sequence. Let $b_j > 0$ for all j and $\sum_{j=-\infty}^{\infty} b_j = 1$. The right-side of (A.14) is bounded by

$$\begin{aligned} & \sum_{j=-\infty}^{\infty} P \left(\sup_{m \leq k \leq T} \frac{1}{\sqrt{k}} \left\| \sum_{t=1}^k \xi_{jt} \right\|_2 > b_j \alpha \right) \\ & \leq \frac{1}{\alpha^2} \sum_{j=-\infty}^{\infty} \frac{1}{b_j^2} \left(m^{-1} \sum_{i=1}^m E \|\xi_{ji}\|_2^2 + \sum_{i=m+1}^T i^{-1} E \|\xi_{ji}\|_2^2 \right). \end{aligned}$$

Note that $\|\xi_{jt}\|_2 \leq \|\xi_{jt}\|_r$ by Liapounov's inequality. By definition, for $j \geq 0$, we have $\|\xi_{jt}\|_r \leq \|E[\xi_t|\mathcal{F}_{t-j}]\|_r + \|E[\xi_t|\mathcal{F}_{t-j-1}]\|_r$ and for $j < 0$, $\|\xi_{jt}\|_r \leq \|\xi_t - E[\xi_t|\mathcal{F}_{t-j-1}]\|_r + \|\xi_t - E[\xi_t|\mathcal{F}_{t-j}]\|_r$. Hence, from the definition of a mixingale, $E\|\xi_{jt}\|_r^2 \leq 4c_t^2\psi_{|j|}^2$ and with Assumption 1(vi)(c), this implies $E\|\xi_{ji}\|_2^2 \leq 4c_i^2\psi_{|j|}^2 \leq 4K^2\psi_{|j|}^2$. Then the right-side of (A.14) is bounded by

$$\frac{1}{\alpha^2} \sum_{j=-\infty}^{\infty} 4b_j^{-2}K^2\psi_{|j|}^2 \left(1 + \sum_{i=m+1}^T i^{-1}\right) \leq \frac{1}{\alpha^2} \sum_{j=-\infty}^{\infty} 4b_j^{-2}K^2\psi_{|j|}^2 (1 + \ln T). \quad (\text{A.15})$$

We can choose appropriate $\{b_j\}$ so that $\sum_j b_j^{-2}\psi_{|j|}^2$ are bounded. From lemma A.6 of Bai and Perron (1998), let $\nu_0 = 1$ and $\nu_j = j^{-1-\kappa}$ for $j \geq 1$, where $\kappa > 0$ is given in Assumption 1(vi)(d). Let $b_j = \nu_j / (1 + 2\sum_{i=1}^{\infty} \nu_i)$ and $b_{-j} = b_j$ for all $j \geq 0$. Then $\sum_{j=-\infty}^{\infty} b_j = 1$. By Assumption 1(vi)(d), we have

$$\sum_{j=-\infty}^{\infty} b_j^{-2}\psi_{|j|}^2 = \left(\psi_0^2 + 2\sum_{j=1}^{\infty} j^{2+2\kappa}\psi_j^2\right) \left(1 + 2\sum_{j=1}^{\infty} j^{-2-2\kappa}\right) < \infty.$$

Hence, the right-side inequality of (A.15) is bounded by $\frac{C \ln T}{\alpha^2}$ for some $C > 0$ and we obtain the result of Lemma A.4. \square

Proof of Lemma A.3

Proof. We use the expression (A.2); the estimator \hat{k} must satisfy $Q_T(\hat{k})^2 \geq Q_T(k_0)^2$ which is equivalent to $H_T(\hat{k}) \geq |k_0 - \hat{k}|G_T(\hat{k})$. Therefore we have

$$\begin{aligned} P(|\hat{\rho} - \rho_0| > \eta) &= P(|\hat{k} - k_0| > T\eta) \\ &\leq P\left(\sup_{|k-k_0|>T\eta} |H_T(k)| \geq \inf_{|k-k_0|>T\eta} |k_0 - k|G_T(k)\right) \\ &\leq P\left(\sup_{|k-k_0|>T\eta} |H_T(k)| \geq T\eta \inf_{|k-k_0|>T\eta} G_T(k)\right) \\ &\leq P\left(\tilde{G}^{-1} \sup_{p \leq k \leq T-p} T^{-1}|H_T(k)| \geq \eta\right), \end{aligned} \quad (\text{A.16})$$

where $\tilde{G} := \inf_{|k-k_0|>T\eta} G_T(k)$, which is positive and bounded away from zero by Lemma

A.1 and the restriction $p \leq k \leq T - p$ is imposed to guarantee existence of $H_T(k)$. Thus, the consistency follows by showing that $T^{-1} \sup_{p \leq k \leq T-p} |H_T(k)| = o_p(1)$, where

$$\begin{aligned}
T^{-1}|H_T(k)| &\leq |T^{-1}\varepsilon' M Z_k (Z_k' M Z_k)^{-1/2} \Omega_k (Z_k' M Z_k)^{-1/2} Z_k' M \varepsilon| \\
&\quad + |T^{-1}\varepsilon' M Z_0 (Z_0' M Z_0)^{-1/2} \Omega_{k_0} (Z_0' M Z_0)^{-1/2} Z_0' M \varepsilon| \\
&\quad + 2 |T^{-1}\delta_T' (Z_0' M Z_k) (Z_k' M Z_k)^{-1/2} \Omega_k (Z_k' M Z_k)^{-1/2} Z_k' M \varepsilon| \quad (\text{A.17}) \\
&\quad + 2 |T^{-1}\delta_T' (Z_0' M Z_0)^{1/2} \Omega_{k_0} (Z_0' M Z_0)^{-1/2} Z_0' M \varepsilon|.
\end{aligned}$$

Lemma A.4 implies that $\sup_k \|T^{-1/2} Z_k' M \varepsilon\| = O_p(\sqrt{\ln T})$. We use (A.18) to verify uniform convergence for all k .

$$\sup_{p \leq k \leq T-p} \|(Z_k' M Z_k)^{-1/2} Z_k' M \varepsilon\| = O_p(\sqrt{\ln T}). \quad (\text{A.18})$$

We show that the third and fourth terms of (A.17) are $O_p(T^{-1/2} \|\delta_T\| \ln T)$ and $O_p(T^{-1/2} \|\delta_T\|)$, respectively. Denote $D_T := T^{-1/2} (Z_0' M Z_k) (Z_k' M Z_k)^{-1/2}$. From (6) and (A.10), $Z_0' M Z_k \leq Z_k' M Z_k$ for all k , and thus

$$\sup_{p \leq k \leq T-p} D_T' D_T \leq \sup_{p \leq k \leq T-p} T^{-1} Z_k' M Z_k = O_p(1),$$

then the third term of (A.17) is bounded by

$$2 \|T^{-1/2} \delta_T' (T^{-1} Z_k' M Z_k)^{1/2} \Omega_k\| \|(Z_k' M Z_k)^{-1/2} Z_k' M \varepsilon\| = \|\delta_T\| O_p(T^{-1/2} \sqrt{\ln T}).$$

For the true break date k_0 , $|(Z_0' M Z_0)^{-1/2} Z_0' M \varepsilon| = O_p(1)$ under our regularity conditions. Hence, the fourth term of (A.17) has order

$$|T^{-1} \delta_T' (Z_0' M Z_0)^{1/2} \Omega_{k_0} (Z_0' M Z_0)^{-1/2} Z_0' M \varepsilon| = \|\delta_T\| O_p(T^{-1/2}).$$

From (A.18) and boundedness of Ω_k , we have $\sup_k \|\Omega_k^{1/2} (Z_k' M Z_k)^{-1/2} Z_k' M \varepsilon\| = O_p(\sqrt{\ln T})$.

Then the first and second terms of (A.17) are bounded as below, respectively.

$$\begin{aligned} \sup_k T^{-1} \left\| \Omega_k^{1/2} (Z'_k M Z_k)^{-1/2} Z'_k M \varepsilon \right\|^2 &= O_p(T^{-1} \ln T) \\ T^{-1} \left\| \Omega_{k_0}^{1/2} (Z'_0 M Z_0)^{-1/2} Z'_0 M \varepsilon \right\|^2 &= O_p(T^{-1}). \end{aligned}$$

By combining all four terms, $T^{-1} \sup_{p \leq k \leq T-p} |H_T(k)| = o_p(1)$. Hence, probability (A.16) is negligible for large T . \square

Proof of Corollary 1

Proof. Let \hat{Z}_0 denote Z_k when k is replaced by \hat{k} . Then the LS estimator of $\hat{\delta}(\hat{\rho})$ is obtained by regressing MY on $M\hat{Z}_0$. By multiplying M to model (4) can be rewritten as $MY = M\hat{Z}_0\delta_T + M\varepsilon^*$, where $\varepsilon^* = \varepsilon + (Z_0 - \hat{Z}_0)\delta_T$. Then,

$$\begin{aligned} \sqrt{T} \left(\hat{\delta}(\hat{\rho}) - \delta_T \right) &= (T^{-1} \hat{Z}'_0 M \hat{Z}_0)^{-1} T^{-1/2} \hat{Z}'_0 M \varepsilon^* \\ &= (T^{-1} \hat{Z}'_0 M \hat{Z}_0)^{-1} \left(T^{-1/2} \hat{Z}'_0 M \varepsilon + T^{-1/2} \hat{Z}'_0 M (Z_0 - \hat{Z}_0) \delta_T \right). \end{aligned}$$

We show that the right side converges in probability to the same limit as when \hat{Z}_0 is replaced by Z_0 . First, we show that $\text{plim } T^{-1/2} \hat{Z}'_0 M (Z_0 - \hat{Z}_0) \delta_T = 0$. Without loss of generality, consider $k > k_0$.

$$\begin{aligned} \left\| T^{-1/2} \hat{Z}'_0 M (Z_0 - \hat{Z}_0) \delta_T \right\| &\leq T^{-1/2} \left\| \hat{Z}'_0 (Z_0 - \hat{Z}_0) - \hat{Z}'_0 X (X' X)^{-1} X' (Z_0 - \hat{Z}_0) \right\| \|\delta_T\| \\ &\leq \frac{1}{\sqrt{T} \|\delta_T\|} \left\| \sum_{t=k_0+1}^{\hat{k}} z_t z'_t \right\| \|\delta_T\|^2 \\ &\quad + \left\| \left(\sum_{t=\hat{k}+1}^T z_t x'_t \right) \left(\sum_{t=1}^T x_t x'_t \right)^{-1} \right\| \frac{1}{\sqrt{T} \|\delta_T\|} \left\| \sum_{t=k_0+1}^{\hat{k}} x_t z'_t \right\| \|\delta_T\|^2 \\ &= \frac{1}{\sqrt{T} \|\delta_T\|} O_p(1) = o_p(1). \end{aligned}$$

The last equality is because the sum has $\hat{k} - k_0 = O_p(\|\delta_T\|^{-2})$ terms, hence, $\left\| \sum_{t=\hat{k}+1}^{k_0} z_t z'_t \right\|$

$\times \|\delta_T\|^2 = O_p(1)$. Also,

$$\begin{aligned}
T^{-1} \left\| \hat{Z}'_0 M \hat{Z}_0 - Z'_0 M Z_0 \right\| &\leq T^{-1} \left\| \hat{Z}'_0 \hat{Z}_0 - Z'_0 Z_0 \right\| + T^{-1} \left\| (\hat{Z}_0 - Z_0)' X (X' X)^{-1} X' \hat{Z}_0 \right\| \\
&\quad + T^{-1} \left\| Z'_0 X (X' X)^{-1} X' (\hat{Z}_0 - Z_0) \right\| \\
&\leq \frac{1}{T \|\delta_T\|^2} \left\| \sum_{t=k_0+1}^{\hat{k}} z_t z_t' \right\| \|\delta_T\|^2 + \frac{2}{T \|\delta_T\|^2} \left\| \sum_{t=k_0+1}^{\hat{k}} x_t z_t' \right\| \|\delta_T\|^2 O_p(1) \\
&= \frac{1}{T \|\delta_T\|^2} O_p(1) = o_p(1).
\end{aligned}$$

Thus, $\sqrt{T} \left(\hat{\delta}(\hat{\rho}) - \delta_T \right) = (T^{-1} Z'_0 M Z_0)^{-1} T^{-1/2} Z'_0 M \varepsilon + o_p(1)$, and the normality follows from the central limit theorem. \square

Proof of Theorem 3

Proof. When $\delta_h = d_0 \sqrt{h}$, the break point estimator $\|\delta_h\|^2 (\hat{k} - k_0) = \|d_0\|^2 (\hat{\rho} - \rho_0) = O_p(1)$ has values in the interval $(-\rho_0 \|d_0\|^2, (1 - \rho_0) \|d_0\|^2)$. Therefore, we only need to examine the behavior of the objective function $Q_T(k)^2$ for those k in the neighborhood of k_0 such that $k = \left[k_0 + s \left\| d_0 \sqrt{h} \right\|^{-2} \right]$, with $s \in (-\rho_0 \|d_0\|^2, (1 - \rho_0) \|d_0\|^2)$. Then for any fixed s , if $h \rightarrow 0$ then $T \rightarrow \infty$ with $k/T \rightarrow \rho = \rho_0 + u$, and $T - k \rightarrow \infty$ with $(T - k)/T \rightarrow 1 - \rho = 1 - \rho_0 - u$, where $u = s \|d_0\|^{-2} \in (-\rho_0, 1 - \rho_0)$. From the objective function (9), we have

$$(T^{-1} Z'_k M Z_k)^{1/2} \sqrt{T} \hat{\delta}_k = (T^{-1} Z'_k M Z_k)^{-1/2} (T^{-1} Z'_k M Z_0) d_0 + (T^{-1} Z'_k M Z_k)^{-1/2} T^{-1/2} Z'_k M \varepsilon.$$

Consider each of the terms as $h \rightarrow 0$, which is equivalent to $T \rightarrow \infty$.

$$\begin{aligned}
T^{-1}Z'_kMZ_0 &= T^{-1} \sum_{t=\max\{k,k_0\}+1}^T z_t z'_t - \left(T^{-1} \sum_{t=k+1}^T R'x_t x'_t \right) (T^{-1}X'X)^{-1} \left(T^{-1} \sum_{t=k_0+1}^T x_t x'_t R \right) \\
&\rightarrow (1 - \max\{\rho, \rho_0\}) \Sigma_z - (1 - \rho)(1 - \rho_0) R' \Sigma_x \Sigma_x^{-1} \Sigma_x R \\
&= (\min\{\rho, \rho_0\} - \rho \cdot \rho_0) \Sigma_z \\
T^{-1}Z'_kMZ_k &= T^{-1} \sum_{t=k+1}^T z_t z'_t - \left(T^{-1} \sum_{t=k+1}^T R'x_t x'_t \right) (T^{-1}X'X)^{-1} \left(T^{-1} \sum_{t=k+1}^T x_t x'_t R \right) \\
&\rightarrow \rho(1 - \rho) \Sigma_z \\
T^{-1/2}Z'_kM\varepsilon &= T^{-1/2} \sum_{t=k+1}^T z_t \varepsilon_t - (T^{-1}Z'_kX)(T^{-1}X'X)^{-1} \left(T^{-1/2} \sum_{t=1}^T x_t \varepsilon_t \right) \\
&\Rightarrow B_1(\rho) - \rho B_1(1).
\end{aligned}$$

By assumption, $\Omega_k \xrightarrow{p} \bar{\Omega}(\rho)$ as $k/T \rightarrow \rho$. This implies that for a fixed d_0 , the objective function $Q_T(k)^2$ weakly converges as follows. For $\rho \leq \rho_0$,

$$\begin{aligned}
Q_T(k)^2 &\Rightarrow \frac{1}{\rho(1 - \rho)} [B_1(\rho) - \rho B_1(1) - \rho(1 - \rho_0) \Sigma_z d_0]' \\
&\quad \times \Sigma_z^{-1/2} \bar{\Omega}(\rho) \Sigma_z^{-1/2} [B_1(\rho) - \rho B_1(1) - \rho(1 - \rho_0) \Sigma_z d_0],
\end{aligned}$$

and for $\rho > \rho_0$,

$$\begin{aligned}
Q_T(k)^2 &\Rightarrow \frac{1}{\rho(1 - \rho)} [B_1(\rho) - \rho B_1(1) - \rho_0(1 - \rho) \Sigma_z d_0]' \\
&\quad \times \Sigma_z^{-1/2} \bar{\Omega}(\rho) \Sigma_z^{-1/2} [B_1(\rho) - \rho B_1(1) - \rho_0(1 - \rho) \Sigma_z d_0].
\end{aligned}$$

By continuous mapping theorem, the the in-fill asymptotic distribution of $T \|\delta_h\|^2 \hat{\rho}$ is the argmax functional of the limit of $Q_T(k)^2$, stated in Theorem 3. \square

Proof of Theorem 4

Proof. We omit the proof of consistency of the break point estimator $\hat{\rho} \rightarrow \rho_0$, because it follows the same procedure as the proof of Theorem 1. Given the rate of convergence

$\hat{\rho} - \rho_0 = O_p(T^{-1}\lambda_h^{-2})$, we only need to examine the behavior of $Q_T(k)^2 - Q_T(k_0)^2$ for those k in the neighborhood of k_0 such that $k \in K(C)$, where $K(C) = \{k : |k - k_0| \leq C\lambda_h^{-2}\}$ for some $C > 0$.

Lemma A.5. *Consider the model (8) and the weight matrix Ω_k that satisfies Assumption 2. For the break magnitude $\delta_h = d_0\lambda_h$ that satisfies Assumption 6(ii),*

$$\begin{aligned} Q_T(k)^2 - Q_T(k_0)^2 &= -\lambda_h^2 d_0' Z_{\Delta}' Z_{\Delta} \Omega_{k_0} d_0 - \lambda_h^2 d_0' (Z_0' M Z_0) (\Omega_{k_0} - \Omega_k) d_0 \\ &\quad + 2\lambda_h d_0' \Omega_{k_0} Z_{\Delta}' \varepsilon \operatorname{sgn}(k_0 - k) + o_p(1) \end{aligned}$$

where $Z_{\Delta} := \operatorname{sgn}(k_0 - k)(Z_k - Z_0)$ and $o_p(1)$ is uniform on $K(C)$.

Because $\delta_h = d_0\lambda_h$, for any constant C of $K(C)$, we consider the limiting process of $Q_T(k)^2 - Q_T(k_0)^2$ for $k = [k_0 + \nu\lambda_h^{-2}]$ and $\nu \in [-C, C]$. Consider $\nu \leq 0$ (i.e., $\rho \leq \rho_0$). From lemma A.5,

$$\begin{aligned} Q_T(k)^2 - Q_T(k_0)^2 &= -d_0' \left(\lambda_h^2 \sum_{t=k+1}^{k_0} z_t z_t' \right) \Omega_{k_0} d_0 - d_0' (T^{-1} Z_0' M Z_0) \lambda_h^2 T (\Omega_{k_0} - \Omega_k) d_0 \\ &\quad + 2d_0' \Omega_{k_0} \left(\lambda_h \sum_{t=k+1}^{k_0} z_t \varepsilon_t \right) + o_p(1). \end{aligned}$$

For $k_0 - k = [-\nu\lambda_h^{-2}]$, the partial sum in the first term converges as follows:

$$\lambda_h^2 \sum_{t=k+1}^{k_0} z_t z_t' \xrightarrow{p} |\nu| \Sigma_z.$$

Consider the second term; from the proof of Theorem 3, $T^{-1} Z_0' M Z_0 \xrightarrow{p} \rho_0(1 - \rho_0) \Sigma_z$ and

$$\lambda_h^2 T (\Omega_{k_0} - \Omega_k) = \frac{-\nu(\Omega_{k_0} - \Omega_k)}{-\nu\lambda_h^{-2} T^{-1}} = \frac{-\nu(\Omega_{k_0} - \Omega_k)}{\rho_0 - \rho},$$

where $\Omega_k = \bar{\Omega}(\rho) + o_p(1)$ for fixed ρ . From Assumption 2, $\bar{\Omega}(\rho)$ is a differentiable function of ρ element-wise and by the mean value theorem, $\bar{\Omega}(\rho_0) - \bar{\Omega}(\rho) = \frac{\partial \bar{\Omega}(\rho)}{\partial \rho} \Big|_{\tilde{\rho}} (\rho_0 - \rho)$ for some $\tilde{\rho} \in (\rho, \rho_0)$. Thus, the equation converges to $-\nu \frac{\partial \bar{\Omega}(\rho)}{\partial \rho} \Big|_{\rho_0} \equiv -\nu \nabla \bar{\Omega}_0$, under the break

magnitude 6(ii).

The third term has partial sum of $z_t \varepsilon_t$, which weakly converges to a Brownian motion process $B_1(-\nu)$ on $[0, \infty)$ that has variance $|\nu| \Xi$.

$$\lambda_h \sum_{t=k+1}^{k_0} z_t \varepsilon_t \Rightarrow B_1(-\nu).$$

Therefore,

$$\begin{aligned} Q_T \left([k_0 + \nu \lambda_h^{-2}] \right)^2 - Q_T(k_0)^2 \\ \Rightarrow -|\nu| d'_0 \Sigma_z \bar{\Omega}_0 d_0 + \nu \rho_0 (1 - \rho_0) d'_0 \Sigma_z \nabla \bar{\Omega}_0 d_0 + 2d'_0 \bar{\Omega}_0 B_1(-\nu). \end{aligned}$$

Let $W_1(\cdot)$ and $W_2(\cdot)$ be an Wiener processes that are independent of each other on $[0, \infty)$. Define $A_\nu := \bar{\Omega}_0 - \text{sgn}(\nu) \rho_0 (1 - \rho_0) \nabla \bar{\Omega}_0$.

$$\tilde{G}(\nu) := \begin{cases} -\frac{|\nu|}{2} (d'_0 \Sigma_z A_\nu d_0) + (d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0)^{1/2} W_1(-\nu) & \text{if } \nu \leq 0 \\ -\frac{|\nu|}{2} (d'_0 \Sigma_z A_\nu d_0) + (d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0)^{1/2} W_2(\nu) & \text{if } \nu > 0. \end{cases}$$

From the continuous mapping theorem, the in-fill asymptotic distribution of the break point estimator is $\lambda_h^2 (\hat{k} - k_0) \Rightarrow \arg \max_\nu \tilde{G}(\nu)$. Denote A_ν in terms of u so that $A_u = \bar{\Omega}_0 - \text{sgn}(u) \rho_0 (1 - \rho_0) \nabla \bar{\Omega}_0$. Let $\nu = cu$ where $c = (d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0) / (d'_0 \Sigma_z A_u d_0)^2$ and $u \in (-\infty, \infty)$. For $\nu \leq 0$,

$$\begin{aligned} \arg \max_{\nu \in (-\infty, 0]} \tilde{G}(\nu) &= \arg \max_{cu \in (-\infty, 0]} -\frac{|u|}{2} c (d'_0 \Sigma_z A_u d_0) + c^{1/2} (d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0)^{1/2} W_1(-u) \\ &= \arg \max_{cu \in (-\infty, 0]} (d'_0 \Sigma_z A_u d_0)^{-1} \left\{ W_1(-u) - \frac{|u|}{2} \right\} \\ &= c \arg \max_{u \in (-\infty, 0]} (d'_0 \Sigma_z A_u d_0)^{-1} \left\{ W_1(-u) - \frac{|u|}{2} \right\}, \end{aligned}$$

where the second equality is from

$$c (d'_0 \Sigma_z A_u d_0) = c^{1/2} (d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0)^{1/2} = (d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0) / (d'_0 \Sigma_z A_u d_0),$$

and the numerator does not depend on u . For $\nu > 0$ we have $\arg \max_{\nu \in (0, \infty)} \tilde{G}(\nu) = c \arg \max_{u \in (0, \infty)} (d'_0 \Sigma_z A_u d_0)^{-1} \{W_2(u) - |u|/2\}$. Thus,

$$\lambda_h^2 (\hat{k} - k_0) \Rightarrow \frac{(d'_0 \bar{\Omega}_0 \Xi \bar{\Omega}_0 d_0)}{(d'_0 \Sigma_z A_u d_0)^2} \arg \max_{u \in (-\infty, \infty)} (d'_0 \Sigma_z A_u d_0)^{-1} \left\{ W(u) - \frac{|u|}{2} \right\},$$

for $W(\cdot)$ defined in Theorem 4. □

Proof of Lemma A.5

Proof. Use equation (A.2), $Z_0 = Z_k - Z_\Delta \text{sgn}(k_0 - k)$ and rearrange terms so that

$$\begin{aligned} & |k_0 - k| G_T(k) \\ &= \delta'_h (Z'_0 M Z_0)^{1/2} \Omega_{k_0} (Z'_0 M Z_0)^{1/2} \delta_h \\ &\quad - \delta_h (Z'_0 M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} (Z'_k M Z_0) \delta_h \\ &= \delta'_h [Z'_\Delta M Z_\Delta \Omega_{k_0} + (Z'_k M Z_k) (\Omega_{k_0} - \Omega_k)] \delta_h \\ &\quad + \delta'_h [2 \text{sgn}(k_0 - k) (Z'_\Delta M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{1/2} \\ &\quad\quad - 2 \text{sgn}(k_0 - k) Z'_\Delta M Z_k \Omega_{k_0} - (Z'_\Delta M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M Z_\Delta] \delta_h \\ &= \lambda_h^2 d'_0 [Z'_\Delta M Z_\Delta \Omega_{k_0} + (Z'_k M Z_k) (\Omega_{k_0} - \Omega_k)] d_0 + o_p(1). \end{aligned}$$

The last equality is because $(Z'_\Delta M Z_k) = \|\delta_h\|^{-2} O_p(1)$ uniformly on $K(C)$ and

$$\begin{aligned} & (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{1/2} - \Omega_{k_0} = \Omega_k - \Omega_{k_0} + o_p(1) \leq b |k_0 - k| / T = o_p(1) \\ & (Z'_\Delta M Z_k) (Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} Z'_k M Z_\Delta = \|\delta_h\|^{-4} T^{-1} O_p(1). \end{aligned}$$

Similarly, from $Z_k = Z_0 - Z_\Delta \text{sgn}(k_0 - k)$, we can show that $(Z'_k M Z_k) (\Omega_{k_0} - \Omega_k) = (Z'_0 M Z_0) (\Omega_{k_0} - \Omega_k) + o_p(1)$ and

$$\begin{aligned} \lambda_h^2 d_0 (Z'_\Delta M Z_\Delta) \Omega_{k_0} d_0 &= \lambda_h^2 d_0 (Z'_\Delta Z_\Delta - Z'_\Delta X (X' X)^{-1} X' Z_\Delta) \Omega_{k_0} d_0 \\ &= \lambda_h^2 d_0 Z'_\Delta Z_\Delta \Omega_{k_0} d_0 + o_p(1), \end{aligned}$$

because $Z'_\Delta X = \|\delta_h\|^{-2} O_p(1)$ and $(X'X)^{-1} = O_p(T^{-1})$. Hence,

$$|k_0 - k|G_T(k) = \lambda_h^2 d'_0 Z'_\Delta Z_\Delta \Omega_{k_0} d_0 + \lambda_h^2 d'_0 (Z'_0 M Z_0) (\Omega_{k_0} - \Omega_k) d_0 + o_p(1). \quad (\text{A.19})$$

Next, consider $H_T(k)$ in equation (A.9).

$$H_T(k) = 2\lambda_h d'_0 \Omega_{k_0} Z'_\Delta \varepsilon \operatorname{sgn}(k_0 - k) + T^{-1/2} \|\delta_h\| |k_0 - k| O_p(1) + O_p(1).$$

Because $|k_0 - k| \leq C \|\delta_h\|^{-2}$ on $K(C)$, the second term in the above equation is bounded by $CT^{-1/2} \|\delta_h\|^{-1} O_p(1) = o_p(1)$. The last term $O_p(1)$ is $o_p(1)$ uniformly on $K(C)$, which can be verified by rearranging terms using $Z_0 = Z_k - Z_\Delta \operatorname{sgn}(k_0 - k)$.

$$\begin{aligned} & \varepsilon' M Z_k (Z_k M Z_k)^{-1/2} \Omega_k (Z_k M Z_k)^{-1/2} Z'_k M \varepsilon - \varepsilon' M Z_0 (Z_0 M Z_0)^{-1/2} \Omega_{k_0} (Z_0 M Z_0)^{-1/2} Z'_0 M \varepsilon \\ &= \varepsilon' M Z_k [(Z'_k M Z_k)^{-1/2} \Omega_k (Z'_k M Z_k)^{-1/2} - (Z'_0 M Z_0)^{-1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2}] Z'_k M \varepsilon \\ & \quad + \varepsilon' M Z_k (Z'_0 M Z_0)^{-1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_\Delta M \varepsilon \operatorname{sgn}(k_0 - k) \\ & \quad + \varepsilon' M Z_0 (Z'_0 M Z_0)^{-1/2} \Omega_{k_0} (Z'_0 M Z_0)^{-1/2} Z'_\Delta M \varepsilon \operatorname{sgn}(k_0 - k) \\ &= O_p(T^{-1} \|\delta_h\|^{-2}) + O_p(T^{-1/2} \|\delta_h\|^{-1}) + o_p(1). \end{aligned}$$

The first line is $O_p(T^{-1} \|\delta_h\|^{-2})$ is uniformly on $K(C)$,

$$\begin{aligned} & \varepsilon' M Z_k [(Z'_k M Z_k)^{-1} \Omega_k - (Z'_0 M Z_0)^{-1} \Omega_{k_0}] Z'_k M \varepsilon + o_p(1) \\ &= \varepsilon' M Z_k (Z'_k M Z_k)^{-1} (Z'_0 M Z_0 - Z'_k M Z_k) (Z'_0 M Z_0)^{-1} \Omega_k Z'_k M \varepsilon \\ & \quad + \varepsilon' M Z_k (Z'_0 M Z_0)^{-1} (\Omega_k - \Omega_{k_0}) Z'_k M \varepsilon + o_p(1) \\ &= T^{-1/2} \|\delta_h\|^{-2} O_p(T^{-1/2}) + o_p(1), \end{aligned}$$

from (A.8) and $(Z'_k M Z_k)^{-1} Z'_k M \varepsilon = O_p(T^{-1/2})$ uniformly on $K(C)$. The second and third lines are $O_p(T^{-1/2} \|\delta_h\|^{-1})$ from $Z'_\Delta M \varepsilon \operatorname{sgn}(k_0 - k) = |k_0 - k|^{1/2} O_p(1) = O_p(\|\delta_h\|^{-1})$. Hence,

$$H_T(k) = 2\lambda_h d'_0 \Omega_{k_0} Z'_\Delta \varepsilon \operatorname{sgn}(k_0 - k) + o_p(1).$$

Combine this with (A.19), we obtain the expression in Lemma A.5. \square

Proof of Theorem 5

Proof. Lemma B.1 from Jiang et al. (2017) is restated below, and used without proof.

Lemma A.6. *For the process y_t defined in (10), the following equations hold when $T = 1/h \rightarrow \infty$ with a fixed $\rho_0 = k_0/T$, for any $\rho \in [0, 1]$,*

$$(a) \quad T^{-1} \sum_{t=1}^{\lfloor \rho T \rfloor} y_{t-1} \varepsilon_t \Rightarrow \sigma^2 \int_0^\rho \tilde{J}_0(r) dB(r);$$

$$(b) \quad T^{-2} \sum_{t=1}^{\lfloor \rho T \rfloor} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^\rho [\tilde{J}_0(r)]^2 dr;$$

$$(c) \quad [\tilde{J}_0(\rho)]^2 - [\tilde{J}_0(0)]^2 = 2 \int_0^\rho \tilde{J}_0(r) dB(r) - 2 \int_0^\rho (\mu + \delta \mathbf{1}\{r > \rho_0\}) [\tilde{J}_0(r)]^2 dr + \rho;$$

$$(d) \quad [\tilde{J}_0(1)]^2 - [\tilde{J}_0(\rho)]^2 = 2 \int_\rho^1 \tilde{J}_0(r) dB(r) - 2 \int_\rho^1 (\mu + \delta \mathbf{1}\{r > \rho_0\}) [\tilde{J}_0(r)]^2 dr + (1 - \rho);$$

where $\tilde{J}_0(r)$ for $r \in [0, 1]$ is a Gaussian process defined in (12) and $B(\cdot)$ is a standard Brownian motion.

Define the $(T \times 2)$ matrix $Y(k) = [Y_1(k) : Y_2(k)]$ with $Y_1(k) = (y_0, \dots, y_{k-1}, 0, \dots, 0)'$, $Y_2(k) = (0, \dots, 0, y_k, \dots, y_{T-1})'$ and $Y = (y_1, \dots, y_T)'$. Then the LS objective function can be expressed as $S(k)^2 = Y' M Y$ where

$$M = I_T - Y_1(k)[Y_1(k)'Y_1(k)]^{-1}Y_1(k)' - Y_2(k)[Y_2(k)'Y_2(k)]^{-1}Y_2(k)',$$

where I_T is a $(T \times T)$ identity matrix. The model (10) can be written as

$$y_t = \beta_1 y_{t-1} + (\beta_2 - \beta_1) \mathbf{1}\{t > k_0\} y_{t-1} + \varepsilon_t = \beta_1 y_{t-1} + \eta_t,$$

where $\eta_t := (\beta_2 - \beta_1) \mathbf{1}\{t > k_0\} y_{t-1} + \varepsilon_t$. Let $Y_- = (y_0, \dots, y_{T-1})'$ and $\eta = (\eta_1, \dots, \eta_T)'$. Then we have $Y = Y_- \beta_1 + \eta$, and the LS objective function is

$$\begin{aligned} S(k)^2 &= (Y_- \beta_1 + \eta)' M' M (Y_- \beta_1 + \eta) \\ &= \eta' \eta - \eta' Y_1(k) [Y_1(k)' Y_1(k)]^{-1} Y_1(k)' \eta - \eta' Y_2(k) [Y_2(k)' Y_2(k)]^{-1} Y_2(k)' \eta \end{aligned}$$

because M is a idempotent matrix and $MY_- = \mathbf{0}$. Note that

$$\eta' \eta = \sum_{t=1}^{k_0} \eta_t^2 + \sum_{t=k_0+1}^T \eta_t^2 = \sum_{t=1}^{k_0} \varepsilon_t^2 + \sum_{t=k_0+1}^T ((\beta_2 - \beta_1)y_{t-1} + \varepsilon_t)^2$$

which holds regardless of the choice of k , and

$$\begin{aligned} \eta' Y_1(k) [Y_1(k)' Y_1(k)]^{-1} Y_1(k)' \eta &= \frac{\left(\sum_{t=1}^k y_{t-1} \eta_t \right)^2}{\sum_{t=1}^k y_{t-1}^2}, \\ \eta' Y_2(k) [Y_2(k)' Y_2(k)]^{-1} Y_2(k)' \eta &= \frac{\left(\sum_{t=k+1}^T y_{t-1} \eta_t \right)^2}{\sum_{t=k+1}^T y_{t-1}^2}. \end{aligned}$$

Therefore, the break point estimator is

$$\begin{aligned} \hat{\rho} &= \arg \max_{\rho \in (0,1)} \omega(\rho)^2 \mathcal{V}(\rho), \tag{A.20} \\ \mathcal{V}(\rho) &:= \left[\frac{\left(\sum_{t=1}^{[\rho T]} y_{t-1} \eta_t \right)^2}{\sum_{t=1}^{[\rho T]} y_{t-1}^2} + \frac{\left(\sum_{t=[\rho T]+1}^T y_{t-1} \eta_t \right)^2}{\sum_{t=[\rho T]+1}^T y_{t-1}^2} \right]. \end{aligned}$$

When $\rho \leq \rho_0$, the terms in the numerator and denominator of $\mathcal{V}(\rho)$ weakly converges as follows.

$$T^{-1} \sum_{t=1}^{[\rho T]} y_{t-1} \eta_t = T^{-1} \sum_{t=1}^{[\rho T]} y_{t-1} \varepsilon_t \Rightarrow \sigma^2 \int_0^\rho \tilde{J}_0(r) dB(r).$$

From Lemma A.6,

$$\begin{aligned} T^{-1} \sum_{t=[\rho T]+1}^T y_{t-1} \eta_t &= T^{-1} \left[\sum_{t=[\rho T]+1}^{[\rho_0 T]} y_{t-1} \eta_t + \sum_{t=[\rho_0 T]+1}^T y_{t-1} \eta_t \right] \\ &= T^{-1} \sum_{t=[\rho T]+1}^T y_{t-1} \varepsilon_t + T(\beta_2 - \beta_1) T^{-2} \sum_{t=[\rho_0 T]+1}^T y_{t-1}^2 \\ &\Rightarrow \sigma^2 \int_\rho^1 \tilde{J}_0(r) dB(r) - \delta \sigma^2 \int_{\rho_0}^1 [\tilde{J}_0(r)]^2 dr, \end{aligned}$$

$$T^{-2} \sum_{t=1}^{[\rho T]} y_{t-1}^2 \Rightarrow \sigma^2 \int_0^\rho [\tilde{J}_0(r)]^2 dr, \quad \text{and} \quad T^{-2} \sum_{t=[\rho T]+1}^T y_{t-1}^2 \Rightarrow \sigma^2 \int_\rho^1 [\tilde{J}_0(r)]^2 dr.$$

Then the LS objective function $\mathcal{V}(\rho)$ in (A.20) weakly converges to

$$\mathcal{V}(\rho) \Rightarrow \sigma^2 \left[\frac{\left(\int_0^\rho \tilde{J}_0(r) dB(r) \right)^2}{\int_0^\rho [\tilde{J}_0(r)]^2 dr} + \frac{\left(\int_\rho^1 \tilde{J}_0(r) dB(r) - \delta \int_{\rho_0}^1 [\tilde{J}_0(r)]^2 dr \right)^2}{\int_\rho^1 [\tilde{J}_0(r)]^2 dr} \right].$$

Lemma A.6 (c) and (d) implies that each term is rearranged as follows.

$$\begin{aligned} \frac{\left(\int_0^\rho \tilde{J}_0(r) dB(r) \right)^2}{\int_0^\rho [\tilde{J}_0(r)]^2 dr} &= \frac{\left([\tilde{J}_0(\rho)]^2 - [\tilde{J}_0(0)]^2 - \rho + 2\mu \int_0^\rho [\tilde{J}_0(r)]^2 dr \right)^2}{4 \int_0^\rho [\tilde{J}_0(r)]^2 dr} \\ &= \frac{\left([\tilde{J}_0(\rho)]^2 - [\tilde{J}_0(0)]^2 - \rho \right)^2}{4 \int_0^\rho [\tilde{J}_0(r)]^2 dr} + \mu^2 \int_0^\rho [\tilde{J}_0(r)]^2 dr \\ &\quad + \mu \left([\tilde{J}_0(\rho)]^2 - [\tilde{J}_0(0)]^2 - \rho \right) \end{aligned}$$

$$\begin{aligned} \frac{\left(\int_\rho^1 \tilde{J}_0(r) dB(r) - \delta \int_{\rho_0}^1 [\tilde{J}_0(r)]^2 dr \right)^2}{\int_\rho^1 [\tilde{J}_0(r)]^2 dr} &= \frac{\left([\tilde{J}_0(1)]^2 - [\tilde{J}_0(\rho)]^2 - (1 - \rho) \right)^2}{4 \int_\rho^1 [\tilde{J}_0(r)]^2 dr} \\ &\quad + \mu^2 \int_\rho^1 [\tilde{J}_0(r)]^2 dr + \mu \left([\tilde{J}_0(1)]^2 - [\tilde{J}_0(\rho)]^2 - (1 - \rho) \right). \end{aligned}$$

As a result, the objective function of the break point estimator in (A.20) weakly converges to

$$\begin{aligned} \frac{\omega(\rho)^2 \mathcal{V}(\rho)}{\sigma^2} &\Rightarrow \frac{\omega(\rho)^2 \left([\tilde{J}_0(\rho)]^2 - [\tilde{J}_0(0)]^2 - \rho \right)^2}{4 \int_0^\rho [\tilde{J}_0(r)]^2 dr} + \frac{\omega(\rho)^2 \left([\tilde{J}_0(1)]^2 - [\tilde{J}_0(\rho)]^2 - (1 - \rho) \right)^2}{4 \int_\rho^1 [\tilde{J}_0(r)]^2 dr} \\ &\quad + \mu^2 \omega(\rho)^2 \int_0^1 [\tilde{J}_0(r)]^2 dr + \mu \omega(\rho)^2 \left([\tilde{J}_0(1)]^2 - [\tilde{J}_0(0)]^2 - 1 \right). \end{aligned}$$

Following the same procedure above, $\omega(\rho)^2 \mathcal{V}(\rho) / \sigma^2$ has the same limit when $\rho > \rho_0$. Therefore, by deleting the terms which are independent of the choice of ρ , the in-fill

asymptotic distribution of $\hat{\rho}$ in (A.20) is

$$\hat{\rho} = \arg \max_{\rho \in (0,1)} \omega(\rho)^2 \mathcal{V}(\rho)$$

$$\Rightarrow \arg \max_{\rho \in (0,1)} \frac{\omega(\rho)^2 \left([\tilde{J}_0(\rho)]^2 - [\tilde{J}_0(0)]^2 - \rho \right)^2}{\int_0^\rho [\tilde{J}_0(r)]^2 dr} + \frac{\omega(\rho)^2 \left([\tilde{J}_0(1)]^2 - [\tilde{J}_0(\rho)]^2 - (1 - \rho) \right)^2}{\int_\rho^1 [\tilde{J}_0(r)]^2 dr},$$

which is identical to the distribution in Theorem 5. \square

Appendix C

Table A.1: In-fill asymptotic RMSE, bias, and the standard error of the new estimator and the LS estimator of the break point under model (13) with parameter values (ρ_0, d_0) and $T = 100$. The number of replications is 5,000.

| ρ_0 | d_0 | RMSE | | Bias | | Standard error | |
|----------|-------|--------|--------|---------|---------|----------------|--------|
| | | NEW | LS | NEW | LS | NEW | LS |
| 0.15 | 1 | 0.4112 | 0.4284 | 0.3270 | 0.3174 | 0.2493 | 0.2876 |
| | 2 | 0.3763 | 0.3898 | 0.2895 | 0.2729 | 0.2403 | 0.2784 |
| | 4 | 0.3159 | 0.3071 | 0.2285 | 0.1807 | 0.2181 | 0.2483 |
| 0.30 | 1 | 0.3004 | 0.3225 | 0.1865 | 0.1741 | 0.2355 | 0.2715 |
| | 2 | 0.2595 | 0.2959 | 0.1527 | 0.1471 | 0.2097 | 0.2567 |
| | 4 | 0.1930 | 0.2277 | 0.1059 | 0.0890 | 0.1614 | 0.2096 |
| 0.50 | 1 | 0.2226 | 0.2635 | 0.0086 | 0.0081 | 0.2224 | 0.2634 |
| | 2 | 0.1894 | 0.2425 | 0.0081 | 0.0140 | 0.1893 | 0.2421 |
| | 4 | 0.1322 | 0.1887 | 0.0094 | 0.0122 | 0.1319 | 0.1883 |
| 0.70 | 1 | 0.2796 | 0.3116 | -0.1560 | -0.1533 | 0.2320 | 0.2713 |
| | 2 | 0.2437 | 0.2837 | -0.1282 | -0.1240 | 0.2072 | 0.2551 |
| | 4 | 0.1840 | 0.2188 | -0.0890 | -0.0676 | 0.1611 | 0.2081 |
| 0.85 | 1 | 0.3986 | 0.4116 | -0.3114 | -0.2981 | 0.2489 | 0.2838 |
| | 2 | 0.3616 | 0.3767 | -0.2717 | -0.2518 | 0.2386 | 0.2802 |
| | 4 | 0.2982 | 0.3003 | -0.2083 | -0.1651 | 0.2134 | 0.2508 |

Table A.2: In-fill asymptotic 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 $T = 200$. The number of replications is 5,000.

| β_1 | β_2 | ρ_0 | RMSE | | Bias | | Standard error | |
|-----------|-----------|----------|--------|--------|---------|---------|----------------|--------|
| | | | NEW | LS | NEW | LS | NEW | LS |
| | | 0.3 | 0.1909 | 0.2046 | 0.1888 | 0.0145 | 0.0290 | 0.2041 |
| 0.5 | 0.38 | 0.5 | 0.0270 | 0.2368 | -0.0053 | -0.0751 | 0.0264 | 0.2246 |
| | | 0.7 | 0.1995 | 0.3164 | -0.1970 | -0.1878 | 0.0316 | 0.2546 |
| | | 0.3 | 0.1284 | 0.1737 | 0.0529 | 0.0392 | 0.1170 | 0.1692 |
| 0.995 | 0.97 | 0.5 | 0.0934 | 0.1749 | -0.0145 | -0.0260 | 0.0923 | 0.1730 |
| | | 0.7 | 0.1759 | 0.2484 | -0.0934 | -0.1128 | 0.1490 | 0.2213 |

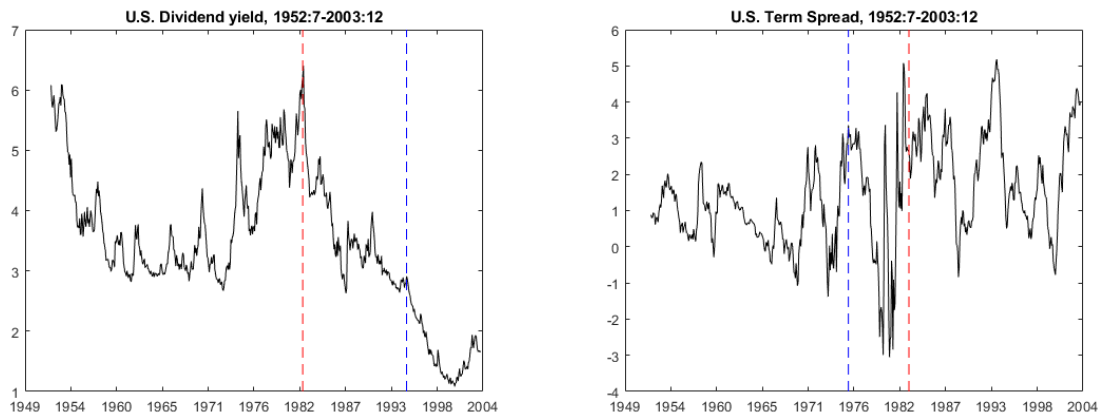


Figure A.1: U.S. dividend yield (left) and term spread (right), 1952:7-2003:12. Red and blue dotted lines are the new and LS break date estimates from the univariate model, respectively.