Estimating and Testing Multiple Structural Breaks in Nonparametric Regressions*

Yiqiu Cao[1], Zhonghao Fu[2], Xia Wang[3], Xingtong Zhang[4]

[1] School of Economics and Management, Tsinghua University
[2] School of Economics, Fudan University
[3] School of Economics, Renmin University of China
[4] School of Finance, Renmin University of China

Abstract: We estimate and test for multiple structural breaks in nonparametric regressions. By transforming an unknown regression function into the frequency domain via the Fourier transform, we can investigate structural breaks in a pseudo generalized regression indexed by frequency. As a result, we can avoid smoothed nonparametric estimation for the unknown regression function. A generalized sup-F test statistic is proposed to detect structural breaks. It can detect a class of local alternatives at the parametric rate, which is asymptotically more powerful than the existing smoothed nonparametric tests. In addition, a BIC-type information criterion and a sequential testing procedure are proposed to determine the appropriate number of breaks. Simulation studies demonstrate the excellent finite sample performance of the proposed approach. In an empirical application, we employ our test to examine the stability of the conditional capital asset pricing model and find significant evidence of structural breaks in both factor loadings and pricing errors.

Keywords: Asset pricing, Fourier transform, local power, nonparametric regression, structural break

*Fu acknowledges financial support from the National Science Foundation of China (NSFC) (Nos. 71903032, 72121002). Wang acknowledges financial support from the NSFC (No. 71873151). Please address correspondence to: Xingtong Zhang, School of Finance, Renmin University of China, 59 Zhongguancun St., Beijiing, China, 100872. Email: *xingtongzhang@ruc.edu.cn*.

1 Introduction

Economic and financial time series often exhibit instability; namely, the underlying data generating process (DGP) may have structural breaks over time. Factors such as preference changes, policy switches, and critical social events may cause structural breaks in economic relationships. Existing studies have documented the prevalence of structural breaks in economic and financial relationships and investigated their impact on modeling, inference, and prediction (e.g., Hansen, 2001; Stock and Watson, 2009; Zhang and Wu, 2012). In recent decades, structural breaks in time series have drawn increasing attention. Most existing research investigates structural breaks for a conditional mean dynamic under certain parametric assumptions; see Perron (2006) for a comprehensive review. However, the parametric approach may lead to misleading conclusions due to model misspecification. That is, when testing for structural breaks, the source of rejection may be model misspecification rather than the structural instability of the regression function (e.g., Fu and Hong, 2019). As Hidalgo (1995) remarked, poor specification may lead to a severe over-selection problem. The existing inference methods may falsely regard certain nonlinear features as spurious additional structural breaks. Furthermore, the asymptotic properties of the estimators for break dates rely on the correct specification of a model.

To avoid the misspecification problem, a related strand of the literature employs nonparametric or semiparametric regressions. Hidalgo (1995) develops a nonparametric conditional moment test for structural breaks without specifying the exact form of the conditional mean function. Su and Xiao (2008) propose a CUSUM-type test for structural breaks in a time series regression. Gao et al. (2008) develop a nonparametric testing procedure that can simultaneously test for structural breaks in the conditional mean and the conditional variance. Su and White (2010) propose two CUSUM-type tests for structural breaks in a partially linear model, one of which examines the stability of coefficients in the parametric component, whereas the other focuses on the stability of the entire regression function. Vogt (2015) considers a nonparametric regression with locally stationary regressors and develops a test for smooth structural changes. Fengler et al. (2015) test for structural breaks in a nonparametric additive model via a back-fitting approach. Fu and Hong (2019) propose a model-free approach to detect smooth structural changes in nonparametric regressions with endogeneity. Mohr and Neumeyer (2020) modify the CUSUM-type test of Su and Xiao (2008) with a marked empirical process approach to obtain a simple limiting distribution and attractive consistency properties. Besides the aforementioned tests for structural changes in nonparametric regressions, the literature also considers estimating the locations and magnitude of structural breaks. Among them, a popular approach is to compare the difference between the left-hand and right-hand sides kernel estimates of a regression function at various time points (e.g., Müller, 1992; Delgado and Hidalgo, 2000; Wang, 2007; Yang et al., 2020). However, most of the previous estimators and test statistics rely on the nonparametric estimations of conditional mean functions, which suffer from the "curse of dimensionality" and usually only have a convergence rate slower than the parametric rate.

In this paper, we estimate and test for multiple structural breaks in nonparametric regressions. Using a frequency domain approach, we can convert structural breaks in a nonparametric time series regression into structural breaks in the Fourier transform of the data. Therefore, we can pin down the break dates by estimating structural breaks in a generalized regression indexed by the frequency. We develop a consistent estimator for the break fractions via minimizing the sum of squared generalized residuals (SSGR) and further construct a sup-F test statistic by comparing the SSGRs under the null hypothesis of no structural break and the alternative of a fixed number of breaks. Several essential features distinguish our work from the existing literature.

To begin with, compared with the existing approaches based on certain parametric assumptions such as Bai and Perron (1998), Chen and Hong (2012), Hall et al. (2012), Chen (2015), and Perron and Yamamoto (2015), our approach is free of model misspecification. As shown in the simulation section, when a linear model is correctly specified, approaches designed for a parametric model, such as Bai and Perron (1998), are more efficient than those designed for a nonparametric model. However, when the true model is misspecified, our approach can deliver more reasonable estimation and testing results.

Moreover, albeit designed for a nonparametric regression model, our approach does not require the smoothed nonparametric estimation. As is well known, the smoothed nonparametric estimation not only involves the delicate business of choosing a smoothing parameter but also incurs the "curse of dimensionality" problem. By examining structural breaks in the frequency domain, our approach is fully parametric. As a result, our estimators and test statistics can achieve the parametric convergence rate and are asymptotically more efficient than the existing nonparametric approaches.

Last but not least, our approach is compatible with multiple structural breaks in nonparametric regressions. In contrast, current studies, such as Fengler et al. (2015), mainly focus on a single structural break. Specifically, we propose a BIC-type information criterion and a sequential testing procedure to estimate the number of breaks. Simulation results show that both procedures perform reasonably well in consistently estimating the true number of breaks in nonparametric regressions.

In an empirical application, we employ the proposed approach to examine the stability of the conditional capital asset pricing model for portfolios sorted on book-to-market ratio. We detect structural breaks in both factor loadings and pricing errors for all the portfolios considered. The break dates of the growth portfolio and value portfolio are estimated and reconciled with economic events. Based on the results, we further characterize the nonlinear features of factor loadings and pricing errors in each regime. These new findings suggest that misleading conclusions can be drawn if one fails to account for structural changes in a conditional capital asset pricing model.

The rest of the paper is organized as follows. In Section 2, we introduce the nonparametric regression model and the frequency domain approach. We then provide our estimators for breaks and show their consistency in Section 3. In Section 4, we propose a sup-F test statistic for structural breaks and derive its asymptotic distribution. We develop two procedures to determine the appropriate number of breaks in Section 5 and extend our approach by allowing for structural breaks and endogeneity in regressors in Section 6. We study the finite sample performance of the proposed estimation procedure and tests in Section 7. We provide an empirical application to a conditional capital asset pricing model in Section 8 and conclude in Section 9. All proofs are relegated to the Appendix.

Throughout this paper, we use $\mathbf{i}$ to denote the imaginary number such that $\mathbf{i} = \sqrt{-1}$. For a real-valued scalar a, $\lfloor a \rfloor$ and $\lceil a \rceil$ denote the integer part and nearest larger integer of a, respectively. For an $m \times n$ complex-valued matrix A, we denote its complex conjugate as A^*, its transpose as A', its real part as $\mathrm{Re}(A)$, and its Euclidean norm as $\|A\|(\equiv [\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2]^{1/2})$. We use $C \in (0, \infty)$ to denote a generic positive constant that may vary from case to case. The operators $\xrightarrow{p}$, $\xrightarrow{d}$, and $\Rightarrow$ denote convergence in probability, convergence in distribution, and weak convergence, respectively.

2 Framework and Approach

Consider the following time series regression:

$$Y_t = m_t(X_t) + \varepsilon_t,\ t = 1, ..., T, \tag{2.1}$$

where Y_t is a scalar dependent variable, X_t is a d-dimensional vector of regressors, $m_t(\cdot)$ is an unknown regression function indexed by t which may vary over time, and ε_t is an error term with $E(\varepsilon_t|X_t) = 0$ and $\mathrm{var}(\varepsilon_t|X_t) = \sigma(X_t)$. Note that (2.1) allows for conditional heteroskedasticity. Nevertheless, we require that the conditional variance of the error term is a time-invariant function of X_t.

Now we discuss how to estimate and test for structural breaks in $m_t(\cdot)$ without estimating it directly. Denote the true unknown break dates and number of breaks with a superscript 0.

Suppose there exist M^0 breaks in $m_t(\cdot)$, denoted by $\{T_j^0\}_{j=1}^{M^0}$, where $T_j^0 = \lfloor Tr_j^0 \rfloor$ for $j = 1, ..., M^0$ with $\{r_j^0\}_{j=1}^{M^0}$ being the collection of corresponding break fractions. Following the convention that $T_0^0 = 0$ and $T_{M^0+1}^0 = T$, (2.1) can be rewritten as:

$$Y_t = \begin{cases} h_1(X_t) + \varepsilon_t, & \text{for } t = 1, ..., T_1^0; \\ h_2(X_t) + \varepsilon_t, & \text{for } t = T_1^0 + 1, ..., T_2^0; \\ \vdots & \vdots \\ h_{M^0+1}(X_t) + \varepsilon_t, & \text{for } t = T_{M^0}^0 + 1, ..., T; \end{cases}$$

where $m_t(\cdot) = h_j(\cdot)$ for $t \in [T_{j-1}^0 + 1, T_j^0]$ and $j = 1, ..., M^0 + 1$, and $\{h_j(\cdot)\}_{j=1}^{M^0+1}$ is a collection of time-invariant unknown regression functions. When there is no structural break in $m_t(\cdot)$, $m_t(\cdot) = h(\cdot)$ for all $t = 1, ..., T$ and some time-invariant function $h(\cdot)$.

Unlike the existing approaches that rely on smoothed nonparametric estimation of $m_t(\cdot)$, we propose a novel frequency domain approach to estimate and test structural breaks in (2.1). Multiply $e^{\mathbf{i}u'X_t}$ on both sides of (2.1) and take expectation, then it follows

$$E(Y_t e^{\mathbf{i}u'X_t}) = E[m_t(X_t)e^{\mathbf{i}u'X_t}], \tag{2.2}$$

where $E(\varepsilon_t e^{\mathbf{i}u'X_t}) = 0$ for all u. Here, $u \in \mathbb{R}^d$ is a vector of nuisance parameters. The relationship between X_t and Y_t can be reflected by the patterns of the projection $E(Y_t e^{\mathbf{i}u'X_t})$ at various frequencies u. Via such a transformation, all information in (2.1) is preserved except for ε_t which is orthogonal to X_t. Furthermore, suppose $\{X_t\}$ is strictly stationary. Then, (2.2) implies that the structural breaks in $m_t(\cdot)$ are identical to the change points in the Fourier transform of Y_t. Assuming that the joint distribution of X_t does not change over time is beneficial for illustrating the idea of our approach. We will show that our method is still applicable when X_t has structural breaks. According to (2.2), we can pin down the break dates in $m_t(\cdot)$ by estimating structural breaks in the following generalized regression:

$$Y_t e^{\mathbf{i}u'X_t} = \phi_t(u) + \eta_t(u), \tag{2.3}$$

where $\phi_t(u) \equiv E[m_t(X_t)e^{\mathbf{i}u'X_t}]$, and $\eta_t(u) \equiv Y_t e^{\mathbf{i}u'X_t} - \phi_t(u)$ is a complex-valued generalized error process such that $E[\eta_t(u)] = 0$ for all t and $\text{var}[\eta_t(u)] = E(Y_t^2) - |\phi_t(u)|^2$. We note that (2.3) is an unobservable pseudo regression. It can be interpreted as a functional regression with a time-varying mean process, e.g., Hörmann and Kokoszka (2010).

Given the equivalence between (2.1) and (2.3) in terms of the collection of break dates $\{T_j^0\}_{j=1}^{M^0}$, we have

$$Y_t e^{\mathbf{i}u'X_t} = \begin{cases} \psi_1(u) + \eta_t(u), & \text{for } t = 1, ..., T_1^0; \\ \psi_2(u) + \eta_t(u), & \text{for } t = T_1^0 + 1, ..., T_2^0; \\ \vdots & \vdots \\ \psi_{M^0+1}(u) + \eta_t(u), & \text{for } t = T_{M^0}^0 + 1, ..., T; \end{cases} \tag{2.4}$$

where $\phi_t(u) = \psi_j(u)$ for $t \in [T_{j-1}^0 + 1, T_j^0]$ and $j = 1, ..., M^0 + 1$, and $\{\psi_j(u)\}_{j=1}^{M^0+1}$ is a collection of complex-valued time-invariant functions such that $\psi_j(u) \neq \psi_{j+1}(u)$ for some u in a non-negligible subset of $\mathbb{R}^d$.

By such construction, we convert structural breaks in the unknown regression function $m_t(\cdot)$ to those in a pseudo generalized regression. Notice that when u is fixed, estimating and testing for structural breaks in $\phi_t(u)$ is equivalent to those for a changing mean in a complex-valued univariate time series process. The merit of such a transformation is obvious. Specifically, we avoid the smoothed nonparametric estimation for $m_t(\cdot)$. Given that structural breaks in $m_t(\cdot)$ are equivalent to those in $\phi_t(u)$, we just need to estimate $\{\psi_j(u)\}_{j=1}^{M^0+1}$ consistently which can be achieved using a fully parametric approach. Hence, we avoid the "curse of dimensionality" in smoothed nonparametric estimation and testing.

3 Estimation for Breaks

In this section, we show how to estimate structural breaks in (2.4). We treat the number of breaks M^0 as known in this section and will propose data-driven methods to determine it in Section 5.

Following Bai (1997), we can estimate structural breaks for each fixed u by minimizing the following sum of squared residuals (SSR)

$$\min_{\{r_1,...,r_{M^0}\} \in \Pi_\epsilon} \sum_{j=1}^{M^0+1} \sum_{t=T_{j-1}+1}^{T_j} \left\| Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}_j(u) \right\|^2, \tag{3.1}$$

where $\tilde{\psi}_j(u) \equiv (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} Y_t e^{\mathbf{i}u'X_t}$ is the sample mean of the jth subsample $\{Y_t e^{\mathbf{i}u'X_t}\}_{t=T_{j-1}+1}^{T_j}$ specified by the collection of candidate break dates $\{T_j\}_{j=1}^{M^0}$. Here, $r_j \equiv T_j/T$ denotes the jth break fraction, and $\Pi_\epsilon = \{\{r_1, ..., r_{M^0}\} : r_j - r_{j-1} \geq \epsilon, j = 1, ..., M^0 + 1\}$ for some small positive constant ϵ is a set that contains all candidate break fractions. Intuitively, if two break dates are too close to each other, there is not enough information in the corresponding segment to distinguish them. Therefore, Π_ϵ ensures that the break dates are asymptotically distinct and are distant away from the boundaries of the sample.

Note that the solution to (3.1) depends on u. In order to construct an objection function that can evaluate all $u \in \mathbb{R}^d$, we consider the following minimization problem:

$$\min_{\{r_1,\ldots,r_{M^0}\}\in\Pi_\epsilon} \sum_{j=1}^{M^0+1} \sum_{t=T_{j-1}+1}^{T_j} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}_j(u) \right\|^2 W(u)\mathrm{d}u, \tag{3.2}$$

where $W(\cdot) : \mathbb{R}^d \to \mathbb{R}^+$ is a nonnegative symmetric weighting function such that $\int_{\mathbb{R}^d} W(u)\mathrm{d}u = 1$. We denote the objective function in (3.2) as the sum of squared generalized residuals (SSGR). It can be viewed as a weighted "sum" of the SSRs in (3.1) for all $u \in \mathbb{R}^d$. By solving such an optimization problem, we can consistently pin down the break fractions. An alternative approach is adopting the group-fused Lasso considered by Qian and Su (2016), who estimate the number of breaks, break dates, and regression coefficients simultaneously in a linear regression model. Since at each fixed u, (2.3) is a linear regression with only an intercept, we can adopt the group-fused Lasso to analyse structural breaks for each fixed u and then consider an integration over all $u \in \mathbb{R}^d$.

Suppose the solution to (3.2) is denoted by $\{\hat{r}_j\}_{j=1}^{M^0}$. We now establish its consistency for the true break fractions. Consider the following regularity conditions.

Assumption 1 (i) $\{X_t, \varepsilon_t\}$ is a strong mixing sequence with mixing coefficient $\alpha(s)$ such that $\sum_{s=0}^{\infty}(s+1)^{q/2-1}\alpha(s) < \infty$ and $\sum_{s=1}^{\infty}\alpha(s)^{(q-2)/q} < \infty$ for some $q > 2$; (ii) $\{X_t\}$ is strictly stationary with $E\|X_t\|^{2q} < \infty$; (iii) $\{\varepsilon_t\}$ is weakly stationary with $\max_t E\left(|\varepsilon_t|^{2q}\right) < \infty$; (iv) $E(\varepsilon_t|X_t) = 0$, $\mathrm{var}(\varepsilon_t|X_t) = \sigma(X_t)$, where $\sigma(\cdot)$ is a time-invariant and integrable function with $E\left[\sigma(X_t)\right]^q < \infty$.

Assumption 2 The weighting function $W(\cdot) : \mathbb{R}^d \to \mathbb{R}^+$ is nonnegative, symmetric, and integrable with $\int_{\mathbb{R}^d}\|u\|^4 W(u)\mathrm{d}u < \infty$.

Assumption 3 (i) $\{h_j(\cdot)\}_{j=1}^{M^0+1}$ is a collection of time-invariant regression functions such that $m_t(\cdot) = h_j(\cdot)$ for $t \in [T_{j-1}^0 + 1, T_j^0]$ and $\max_{j\in\{1,\ldots,M^0+1\}} E\left|h_j(X_t)\right|^{2q} < \infty$; (ii) there exists a small positive constant ϵ such that $\min_{j\in\{1,\ldots,M^0+1\}}(r_j^0 - r_{j-1}^0) \geq \epsilon$, where $r_j^0 = T_j^0/T$ and $0 < r_1^0 < r_2^0 \cdots < r_{M^0} < 1$.

Assumption 1(i) restricts the temporal dependence of the data to be weak. The strong mixing condition has been widely adopted in time series analysis. Various time series processes satisfy this condition, such as the autoregressive moving average (ARMA) and autoregressive conditional heteroskedastic (ARCH) processes. It is necessary to establish the limiting results in this paper. Assumption 1(ii) implies that the joint distribution of X_t does not change over time. It rules out the possibility that the instability of the Fourier transform of data is caused by structural breaks in X_t. In fact, we can relax this condition and assume that $\{X_t\}_{t=1}^{T}$ is a piece-wise strictly stationary process such that $\{X_t\}_{t=T_{j-1}^0+1}^{T_j^0}$ is strictly stationary for $j = 1, ..., M^0 + 1$. It implies that we only need X_t to be strictly stationary within each time segment specified by the true break dates. Unlike the existing approaches such as Su and Xiao (2008) and Fengler et al. (2015), we do not require the probability density of X_t to exist. Assumption 1(iii) assumes that the error term ε_t is weakly stationary. Together with Assumption 1(ii), it guarantees the equivalence between structural breaks in $\phi_t(u)$ and those in $m_t(X_t)$. Note that we allow the error term to be serially correlated. Assumption 1(iv) implies that the dependence between X_t and ε_t are time-invariant up to the first two moments of ε_t.

Assumption 2 imposes mild conditions on the weighting function. It ensures that the integral in (3.2) is well-defined. Many functions, such as symmetric joint probability density functions with finite fourth-order moments, satisfy this condition. Intuitively, the weighting function determines the relative impact of the structural breaks at each frequency u. Ideally, the optimal weighting function should assign large weights to the frequencies where the magnitude of structural breaks is big. However, that depends on the very nature of structural breaks, which is unknown *a priori*. In this paper, we suggest using certain weighting functions that can deliver a closed-form expression of the SSGR defined in (3.2). Such weighting functions include the product joint normal probability density function and the joint Laplace probability density function.

Assumption 3(i) imposes regularity conditions on the regression functions $\{h_j(\cdot)\}_{j=1}^{M^0+1}$. Compared with the existing approaches that require smoothed nonparametric regression, we do not impose any smoothness condition on $\{h_j(\cdot)\}_{j=1}^{M^0+1}$. This implies that our approach is applicable to nonparametric regressions with discrete regressors. It also allows for discontinuity in certain regression function $h_j(\cdot)$ or its derivatives. Notice that when there is no structural breaks in $m_t(\cdot)$, we simply have that $h_j(\cdot) = h(\cdot)$ for some time-invariant function $h(\cdot)$ and all j. Then Assumption 3(i) implies that $E|h(X_t)|^{2q} < \infty$. Assumption 3(ii) is a standard condition on the break fractions, see, e.g., Bai and Perron (1998).

Now we show that the estimated break fractions $\{\hat{r}_1, ..., \hat{r}_{M^0}\}$ defined by (3.2) are consistent for the true break fractions.

Theorem 3.1 *Suppose Assumptions 1–3 hold. Then, for* $j = 1, 2, ..., M^0$,
(i) $\hat{r}_j \xrightarrow{p} r_j^0$, *as* $T \to \infty$; *and*
(ii) for every $\gamma > 0$, *there exists a constant* $\delta \in (0, \infty)$, *such that* $P\left[\left|T\left(\hat{r}_j - r_j^0\right)\right| > \delta\right] < \gamma$.

Theorem 3.1(i) shows that $\hat{r}_j$ consistently estimates the corresponding true break fraction for each j. Theorem 3.1(ii) implies that the estimated break fractions converge to the true value at the rate of T. We note that this convergence in probability pertains to the estimated break fraction $\hat{r}_j$ but not the estimated break date $\hat{T}_j$. For the latter, we have $P\left[\left|\hat{T}_j - T_j^0\right| > \delta\right] < \gamma$ for each $j = 1, 2, ..., M^0$, which implies that the deviation of $\hat{T}_j$ from the true break date T_j^0 is bounded by some constant δ with a high probability that is independent of T. The estimators proposed in Delgado and Hidalgo (2000) converge to the true value at the rate of Th^{d-1}, where h and d are the bandwidth and the dimension of regressors, respectively. This rate is slower than that of our estimators.

4 Test for Structural Breaks

This section considers testing the null hypothesis of no breaks in $m_t(\cdot)$ against the alternative of M structural breaks, where M is a prespecified constant. We will introduce the test statistic, derive the asymptotic null distribution, and analyze the power property in this section.

4.1 Test Statistic

Given the equivalence between (2.1) and (2.3) in terms of structural breaks, we can test structural breaks in (2.1) by examining (2.3). With a prespecified number of breaks M, the null hypothesis of no structural breaks can be written as:

$$\mathbb{H}_0 : \phi_t(u) = \psi(u), \text{ for all } t = 1, ..., T, \text{ and all } u \in \mathbb{R}^d,$$

where $\psi(u)$ is an unknown complex-valued time-invariant function. The alternative hypothesis is

$$\mathbb{H}_A : \phi_t(u) = \begin{cases} \psi_1(u), & \text{for } t = 1, ..., T_1; \\ \psi_2(u), & \text{for } t = T_1 + 1, ..., T_2; \\ \vdots & \vdots \\ \psi_{M+1}(u), & \text{for } t = T_M + 1, ..., T; \end{cases}$$

for some $u \in \mathbb{R}^d$, some collection of break dates $\{T_1, ..., T_M\}$, and the corresponding collection of complex-valued functions $\{\psi_j(u)\}_{j=1}^{M+1}$. To test for structural breaks, we construct a generalized sup-F test statistic by comparing the SSGRs under $\mathbb{H}_0$ and $\mathbb{H}_A$.

Let $\tilde{\psi}(u) = T^{-1} \sum_{t=1}^{T} Y_t e^{\mathbf{i}u'X_t}$ and $\tilde{\psi}_j(u) = (T_j - T_{j-1})^{-1} \sum_{t=T_{j-1}+1}^{T_j} Y_t e^{\mathbf{i}u'X_t}$ be the estimators for $\psi(u)$ and $\psi_j(u)$ under the restricted and unrestricted models, respectively. We denote the

SSGRs under $\mathbb{H}_0$ and $\mathbb{H}_A$ as SSGR_0 and $\text{SSGR}_M(r_1, \cdots, r_M)$, respectively, where

$$\text{SSGR}_0 = \sum_{t=1}^{T} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}(u) \right\|^2 W(u) \mathrm{d}u,$$

and

$$\text{SSGR}_M(r_1, \cdots, r_M) = \sum_{j=1}^{M+1} \sum_{t=T_{j-1}+1}^{T_j} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}_j(u) \right\|^2 W(u) \mathrm{d}u.$$

Then, our test statistic is defined as the following:

$$F_T = \sup_{\{r_1,\dots,r_M\} \in \Pi_\epsilon} \left[\text{SSGR}_0 - \text{SSGR}_M(r_1, \dots, r_M) \right]. \tag{4.1}$$

Intuitively, under $\mathbb{H}_0$, both $\tilde{\psi}(u)$ and $\tilde{\psi}_j(u)$ are unbiased and consistent estimators for $E(Y_t e^{\mathbf{i}u'X_t})$. Hence, the difference between SSGR_0 and $\text{SSGR}_M(r_1, \dots, r_M)$ should be close to 0 under $\mathbb{H}_0$ but will deviate from 0 under $\mathbb{H}_A$. We note that it is possible that M differs from the true number of breaks M^0. However, that does not affect the validity of the proposed test.

4.2 Asymptotic Distribution

Before we provide the limiting distribution of the sup-F test statistic, we establish the following lemma that describes the limiting distribution of the partial sum of generalized errors.

Lemma 4.1 *Suppose Assumptions 1–3 hold. Let $\mathbb{U} = [-c, c]^d$ be a compact subset of $\mathbb{R}^d$ for $c > 0$. Then,*
(i) under $\mathbb{H}_0$,

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor Tr \rfloor} \eta_t(u) \Rightarrow B(u, r),$$

as $T \to \infty$, where $B(u, r)$ is a complex-valued Gaussian process defined on $\mathbb{U} \times [0, 1]$ with mean 0 and covariance kernel $E[B(u, r)B(v, s)^*] = \min\{r, s\}\Omega(u, v)$, with a long-run variance

$$\Omega(u, v) = \sum_{l=-\infty}^{\infty} \text{cov}(Y_t e^{\mathbf{i}u'X_t}, Y_{t-l} e^{-\mathbf{i}v'X_{t-l}});$$

(ii) under $\mathbb{H}_A$ with the collection of break dates being $\{T_j^0\}_{j=1}^{M^0}$, for each $j = 1, ..., M^0 + 1$,

$$\frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+\lfloor (T_j^0 - T_{j-1}^0) r \rfloor} \eta_t(u) \Rightarrow B^{(j)}(u, r),$$

as $T \to \infty$, where $B^{(j)}(u, r)$ is a complex-valued Gaussian process defined on $\mathbb{U} \times [0, 1]$ with mean 0 and covariance kernel $E[B^{(j)}(u, r)B^{(j)}(v, s)^*] = \min\{r, s\}\Omega^{(j)}(u, v)$, with a long-run variance

$$\Omega^{(j)}(u, v) = \sum_{l=-\infty}^{\infty} \text{cov}(Y_t e^{\mathbf{i}u' X_t}, Y_{t-l} e^{-\mathbf{i}v' X_{t-l}})$$

for $T_{j-1}^0 + 1 \leq t \leq T_j^0$.

Lemma 4.1 establishes the limiting distribution related to the generalized error. Under $\mathbb{H}_0$, Y_t and X_t are jointly stationary. Hence, the covariance structure of $\eta_t(u)$ is constant over time, i.e., $E[\eta_t(u)\eta_s(u)^*] = E[Y_t Y_s e^{\mathbf{i}u'(X_t - X_s)}] - |\psi(u)|^2$ only depends on $|t - s|$. Under Assumptions 1 and 3, we can show that the long-run variance $\Omega(u, v)$ exists. However, under $\mathbb{H}_A$, Y_t is only weakly stationary within each time segment $[T_{j-1}^0 + 1, T_j^0]$. Hence, the limiting distribution for the partial sum of the generalized error only holds for $T_{j-1}^0 + 1 \leq t \leq T_j^0$. Note that Lemma 4.1(ii) only holds when the collection of break dates coincides with the true breaks.

Theorem 4.2 *Suppose Assumptions 1–3 hold. Under $\mathbb{H}_0$, as $T \to \infty$,*

$$F_T \xrightarrow{d} \sup_{\{r_1, ..., r_M\} \in \Pi_\epsilon} F(r_1, ..., r_M),$$

where

$$F(r_1, ..., r_M) = \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \|\mathcal{B}(u, r_j) - \mathcal{B}(u, r_{j-1})\|^2 W(u) \mathrm{d}u,$$

with $\mathcal{B}(u, r) \equiv B(u, r) - rB(u, 1)$ being a generalized Brownian bridge and $B(u, r)$ as defined in Lemma 4.1(i).

Theorem 4.2 provides the limiting distribution of our test statistic defined in (4.1). It implies that the asymptotic distribution of the proposed sup-F test statistic depends on the unknown DGP. We will provide a moving block bootstrap procedure to obtain the critical values in Section 4.3.

Next, we conduct a local power analysis to gain additional insights into our test. Under $\mathbb{H}_A$, suppose there are M^0 breaks with break dates $\{T_j^0\}_{j=1}^{M^0}$. Consider the following class of local alternatives:

$$\mathbb{H}_A(a_T) : \psi_j(u) = \psi(u) + a_T \Delta_j(u)$$

where $\Delta_j(u)$ is a complex-valued function of u that characterizes the deviation of the jth regime from a time-invariant complex-valued function $\psi(u)$. We let $\sum_{j=1}^{M^0+1} \Delta_j(u) = 0$ for all u for the identification purpose. Moreover, a_T satisfies that $a_T \to 0$ as $T \to \infty$, which controls the speed at which the class of local alternatives $\mathbb{H}_A(a_T)$ converges to the null hypothesis $\mathbb{H}_0$.

Theorem 4.3 *Suppose Assumptions 1–3 hold. Then, under $\mathbb{H}_A(a_T)$ with $a_T = T^{-1/2}$, as $T \to \infty$,*

$$F_T \xrightarrow{d} \sup_{\{r_1,...,r_M\} \in \Pi_\epsilon} F^A(r_1, ..., r_M),$$

where

$$F^A(r_1, ..., r_M) = \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \| G(u, r_j) - G(u, r_{j-1}) + \Gamma(u, r_j) - \Gamma(u, r_{j-1}) \|^2 W(u) \mathrm{d}u,$$

with

$$\begin{aligned} G(u, r_j) \quad &= \quad \left[\sum_{k=1}^{l} (r_k^0 - r_{k-1}^0)^{1/2} B^{(k)}(u, 1) + (r_{l+1}^0 - r_l^0)^{1/2} B^{(l+1)} \left(u, \frac{r_j - r_l^0}{r_{l+1}^0 - r_l^0} \right) \right] \\ &\quad - r_j \left[\sum_{k=1}^{M^0+1} (r_k^0 - r_{k-1}^0)^{1/2} B^{(k)}(u, 1) \right], \end{aligned}$$

and

$$\Gamma(u, r_j) = \left[\sum_{k=1}^{l} (r_k^0 - r_{k-1}^0) \Delta_k(u) + (r_j - r_l^0) \Delta_{l+1}(u) \right] - r_j \left[\sum_{k=1}^{M^0+1} (r_k^0 - r_{k-1}^0) \Delta_k(u) \right].$$

Here, l denotes that T_j lies in the $(l+1)$th subsample specified by the true break dates, i.e., $r_l^0 < r_j < r_{l+1}^0$, and $B^{(j)}(u, r)$ is defined in Lemma 4.1(ii).

Theorem 4.3 provides the asymptotic distribution of our test under the class of local alternatives $\mathbb{H}_A(a_T)$. It shows that the proposed test has non-trivial power against a class of local alternatives that converges to the null hypothesis at the parametric rate. We note that Fengler et al.'s (2015) test can detect local alternatives at rate $T^{-1/2}h^{-1/4}$, which depends on the bandwidth parameter h and is thus slower than $T^{-1/2}$. Besides, the tests of Hidalgo (1995) and Su and Xiao (2008) only have local power of order $T^{-1/2}$ in certain directions of local alternatives. We note that the validity of our test does not require a correct specification of the number of breaks. Even when a wrong number of breaks is chosen, our test is still powerful.

4.3 Resampling Procedure

Now, we propose the following moving block bootstrap (MBB) procedure to obtain the critical values of our sup-F test. Let $Z_t \equiv (X_t', Y_t)'$.

(i) Given the dataset $\{Z_t\}_{t=1}^T$ and a prespecified number of breaks M, compute SSGR_0 and $\mathrm{SSGR}_M(\hat{r}_1, ..., \hat{r}_M)$, where $\{\hat{r}_j\}_{j=1}^M$ is the collection of estimated break fractions that solves (3.2). Then the test statistic $F_T = \mathrm{SSGR}_0 - \mathrm{SSGR}_M(\hat{r}_1, ..., \hat{r}_M)$.

(ii) Pick a block length $1 < l_T < T$ such that $l_T^{-1} + l_T T^{-1/2} = o(1)$, and construct $N \equiv T - l_T + 1$ sets of block data $\{\mathcal{Z}_n\}_{n=1}^N$, where $\mathcal{Z}_n = \{Z_n, ..., Z_{n+l_T-1}\}$ is a block dataset with length l_T.

(iii) Denoting $K = \lceil T/l_T \rceil$, draw i.i.d. integer-valued random variables $I_1, ..., I_K$ such that I_k, $k = 1, ..., K$, follows a discrete uniform distribution that assigns the probability $1/N$ to each value in the set $\{1, ..., N\}$. Construct a bootstrap dataset $\Phi^\star = \{\mathcal{Z}_1^\star, ..., \mathcal{Z}_K^\star\}$, where $\mathcal{Z}_k^\star = \mathcal{Z}_{I_k}$ for $k = 1, ..., K$.

(iv) Compute the bootstrap test statistic $F_T^\star = \mathrm{SSGR}_0^\star - \mathrm{SSGR}_M^\star(\hat{r}_1^\star, ..., \hat{r}_M^\star)$, where $\mathrm{SSGR}_0^\star$ is the SSGR of no breaks and $\mathrm{SSGR}_M^\star(\hat{r}_1^\star, ..., \hat{r}_M^\star)$ is SSGR under the estimated M breaks $\{\hat{r}_j^*\}_{j=1}^M$ for the bootstrap dataset $\Phi^\star$.

(v) Repeat steps (iii)–(iv) for $\mathbf{B}$ times to obtain $\mathbf{B}$ bootstrap test statistics $\{F_{T,b}^\star\}_{b=1}^{\mathbf{B}}$. Then the bootstrap p-value for the proposed test is given by

$$p_J^{\mathbf{B}} = \frac{1}{\mathbf{B}} \sum_{b=1}^{\mathbf{B}} \mathbf{1}\left(F_T \leq F_{T,b}^\star\right),$$

where $\mathbf{1}(\cdot)$ is an indicator function.

The following theorem establishes the validity of the proposed MBB.

Theorem 4.4 *Suppose Assumptions 1-3 hold. The block length satisfies* $l_T^{-1} + l_T T^{-1/2} = o(1)$. *Then under* $\mathbb{H}_0$, $F_T^\star \Rightarrow^* F_T$ *in probability as* $T \to \infty$. *Under* $\mathbb{H}_A(a_T)$, $P^*(F_T > F_T^\star) \to 1$ *provided* $T^{1/2} a_T l_T^{-1/2} \to \infty$, *where* $\Rightarrow^*$ *and* P^* *denote the weak convergence and probability under the bootstrap probability measure conditional on the observed sample* $\{Z_t\}_{t=1}^T$.

Theorem 4.4 shows that the proposed MBB provides an asymptotic valid approximation to the limiting null distribution of the generalized sup-F test statistic. Under $\mathbb{H}_A(a_T)$, the sup-F test statistic will diverge to infinity at the rate of Ta_T^2. On the contrary, the bootstrap statistic diverges to infinity at the rate of l_T. This guarantees the asymptotic validity of the proposed resampling method. We note that under a non-converging global alternative, the order of magnitude of our sup-F test statistic is $O_P(T)$. That implies that the proposed MBB is always valid given $l_T T^{-1/2} = o(1)$. We will investigate the finite sample performance of our test under the proposed MBB procedure in Section 7.

5 Determining the Number of Breaks

Consistent estimation for break fractions requires that the number of breaks should be correctly specified. This section considers a sequential testing procedure and an information criterion to consistently estimate the true number of breaks.

5.1 Sequential Tests

For the sequential tests, we consider testing the null hypothesis of M breaks against the alternative of $M+1$ breaks. Under the null hypothesis, we estimate the break dates $\{\hat{T}_j\}_{j=1}^M$ by solving (3.2) and denote the corresponding SSGR as $\text{SSGR}_M(\hat{T}_1, \cdots, \hat{T}_M)$. Conditioning on $\{\hat{T}_j\}_{j=1}^M$, we then estimate (3.2) by allowing for an additional break τ, $\hat{T}_{j-1} < \tau < \hat{T}_j$ for some $j = 1, ..., M+1$, and compute the corresponding SSGR denoted by $\text{SSGR}_{M+1}(\hat{T}_1, \ldots, \hat{T}_{j-1}, \tau, \hat{T}_j, \ldots, \hat{T}_M)$. Similar to our sup-$F$ test for structural breaks, we define the following sequential test statistic:

$$F_T(M+1|M) = \text{SSGR}_M\left(\hat{T}_1, \ldots, \hat{T}_M\right) - \min_{1 \le j \le M+1} \inf_{\tau \in \Lambda_{j,\epsilon}} \text{SSGR}_{M+1}\left(\hat{T}_1, \ldots, \hat{T}_{j-1}, \tau, \hat{T}_j, \ldots, \hat{T}_M\right)$$

where $\Lambda_{j,\epsilon} = \{\tau \in \mathbb{N} : \hat{T}_{j-1} + \lfloor(\hat{T}_j - \hat{T}_{j-1})\epsilon\rfloor \le \tau \le \hat{T}_j - \lfloor(\hat{T}_j - \hat{T}_{j-1})\epsilon\rfloor\}$ for some $\epsilon > 0$.

Theorem 5.1 *Suppose Assumptions 1–3 hold. Let* $\mathcal{B}^{(j)}(u,r) \equiv B^{(j)}(u,r) - rB^{(j)}(u,1)$ *be a generalized Brownian bridge, where* $B^{(j)}(u,r)$ *is defined as in Lemma 4.1(ii). Then, under the null* *hypothesis of M structural breaks, as $T \to \infty$,*

$$F_T(M+1|M) \xrightarrow{d} F(M+1|M),$$

where

$$F(M+1|M) = \max_{1 \leq j \leq M+1} \sup_{\epsilon \leq r \leq 1-\epsilon} \int_{\mathbb{R}^d} \frac{\left\| \mathcal{B}^{(j)}(u,r) \right\|^2}{r(1-r)} W(u) \mathrm{d}u.$$

When $M < M^0$, our test should reject the null hypothesis of M breaks and favor the alternative of $M+1$ breaks. The testing procedure stops when we fail to reject the null hypothesis at a certain $\tilde{M}$. We then treat $\tilde{M}$ as the estimate for the true number of breaks M^0. Similar to Theorem 4.2, the asymptotic distribution of the sequential test is data-dependent. Thus, we consider the following MBB to obtain the asymptotic critical values.

(i) Given the dataset $\{Z_t\}_{t=1}^T$ and the hypothesized number of breaks $M = 0, 1, ..., M_{\max}$, where $M_{\max}$ is upper bound for the number of breaks, estimate the break fractions $\{\hat{r}_j\}_{j=1}^M$ and compute the test statistic $F_T(M+1|M)$.

(ii) For each subsample $\{Z_t\}_{t=\hat{T}_{j-1}+1}^{\hat{T}_j}$ specified by the estimated break dates, pick a block length $1 < l_T^j < \hat{T}_j - \hat{T}_{j-1}$ such that $(l_T^j)^{-1} + l_T^j T^{-1/2} = o(1)$, and construct $N_j \equiv \hat{T}_j - \hat{T}_{j-1} - l_T^j + 1$ sets of block data $\{\mathcal{Z}_n^j\}_{n=1}^{N_j}$, where $\mathcal{Z}_n^j = \{Z_{\hat{T}_{j-1}+n}, ..., Z_{\hat{T}_{j-1}+n+l_T^j-1}\}$ is a block dataset with length l_T^j.

(iii) Denoting $K_j = \lceil (\hat{T}_j - \hat{T}_{j-1})/l_T^j \rceil$, for $j = 1, ..., M+1$, draw i.i.d. integer-valued random variables $I_1^j, ..., I_{K_j}^j$ such that I_k^j, $k = 1, ..., K_j$ follows a discrete uniform distribution that assigns the probability $1/N_j$ to each value in the set $\{1, ..., N_j\}$. Construct a bootstrap dataset $\Phi_j^\star = \{\mathcal{Z}_1^{j,b}, ..., \mathcal{Z}_{K_j}^{j,b}\}$, where $\mathcal{Z}_k^{j,b} = \mathcal{Z}_{I_k}^j$ for $k = 1, ..., K_j$, and $j = 1, ..., M+1$. Combine the data contained in each bootstrap subsample j, and then obtain the following bootstrap observations for the whole sample $\{\Phi_1^\star, ..., \Phi_{M+1}^\star\}$;

(iv) Compute the bootstrap test statistic $F_T^\star(M+1|M)$ based on the bootstrap sample $\{\Phi_1^\star, ..., \Phi_{M+1}^\star\}$.

(v) Repeat steps (iii)–(iv) for $\mathbf{B}$ times to obtain $\mathbf{B}$ bootstrap test statistics $\{F_{T,b}^\star(M+1|M)\}_{b=1}^{\mathbf{B}}$. The bootstrap p-value for the sequential test is given by

$$p_S^{\mathbf{B}} = \frac{1}{\mathbf{B}} \sum_{b=1}^{\mathbf{B}} \mathbf{1} \left[F_T(M+1|M) \leq F_{T,b}^\star(M+1|M) \right].$$

Suppose Assumptions 1-3 hold, the validity of this procedure can be established analogously to Theorem 4.4. We note that under the null hypothesis $\mathbb{H}_0 : M^0 = M$, the asymptotic distribution of the sequential test is the same as that of the sup-F test for a single break.

5.2 An Information Criterion

In addition to the sequential test, we propose a BIC-type information criterion (IC) to suggest a data-driven choice of the number of breaks. Define $\hat{\sigma}^2(M) = T^{-1}\,\mathrm{SSGR}_M(\hat{T}_1, \ldots, \hat{T}_M)$. Then, we choose $\hat{M}$ as

$$\hat{M} = \arg\min_{0 \le M \le M_{\max}} \ln\left[\hat{\sigma}^2(M)\right] + \rho_T(M+1) \tag{5.1}$$

where ρ_T is a positive tuning parameter, and $M_{\max}$ is an upper bound of M^0.

To establish the consistency of the model selection approach, we impose the following assumption on the tuning parameter.

Assumption 4 $\rho_T \to 0$ and $T\rho_T \to \infty$ as $T \to \infty$.

Theorem 5.2 *Suppose Assumptions 1-4 hold. Then as* $T \to \infty$*,*

$$P(\hat{M} = M^0) \to 1.$$

Theorem 5.2 shows that the IC procedure can select the true number of breaks consistently. To implement the IC procedure, we let $\rho_T = c_\rho \ln(T)/T$, where c_ρ is a deeper tuning parameter at the practitioner's discretion.

6 Discussions

In the benchmark model, the regressor X_t is assumed to be strictly stationary and exogenous. This section relaxes these assumptions and show that our methods are still applicable in general cases.

6.1 Regressors with structural breaks

Assumption 1(ii) requires that $\{X_t\}_{t=1}^T$ be strictly stationary. That may appear restrictive in practice since the joint distribution of X_t may undergo structural breaks. If that occurs, one can first estimate the break points in the joint distribution of X_t, and then analyze structural breaks in the regression function $m_t(\cdot)$ within each subsample where X_t is strictly stationary. This idea is quite similar to the structural breaks estimation and tests in a linear time series regression model with endogenous covariates proposed by Hall et al. (2012). When the instrumental variables have structural breaks, they need to first estimate structural breaks in the reduced form. Then, they estimate and test for structural breaks in the structural function within each subsample.

Without loss of generality, we assume that $\{X_t\}_{t=1}^T$ has a single structural break $\tau_x \in [T_{j-1}^0, T_j^0]$, i.e.,

$$X_t = \begin{cases} X_{1t}, & \text{for } t = 1, ..., \tau_x; \\ X_{2t}, & \text{for } t = \tau_x + 1, ..., T; \end{cases}$$

where X_{1t} and X_{2t} are two distinct strictly stationary random variables. Given the break dates $\{T_1^0, ..., T_{M^0}^0\}$ of $m_t(\cdot)$ and break point τ_x of $\{X_t\}_{t=1}^T$, model (2.1) can be extended to the following expression:

$$Y_t = \begin{cases} h_1(X_{1t}) + \varepsilon_t, & \text{for } t = 1, ..., T_1^0; \\ h_2(X_{1t}) + \varepsilon_t, & \text{for } t = T_1^0 + 1, ..., T_2^0; \\ \vdots & \vdots \\ h_j(X_{1t}) + \varepsilon_t, & \text{for } t = T_{j-1}^0 + 1, ..., \tau_x; \\ h_j(X_{2t}) + \varepsilon_t, & \text{for } t = \tau_x + 1, ..., T_j^0; \\ \vdots & \vdots \\ h_{M^0+1}(X_{2t}) + \varepsilon_t, & \text{for } t = T_{M^0}^0 + 1, ..., T; \end{cases}$$

where $m_t(\cdot) = h_j(\cdot)$ for $t \in [T_{j-1}^0 + 1, T_j^0]$ and $j = 1, ..., M^0 + 1$, and $\{h_j(X_t)\}_{j=1}^{M^0+1}$ is a collection of time-invariant integrable functions. In addition, the generalized regression in (2.3) can be rewritten as the following

$$Y_t e^{\mathbf{i}u'X_t} = \begin{cases} \psi_1(u) + \eta_t(u), & \text{for } t = 1, ..., T_1^0; \\ \vdots & \vdots \\ \psi_{j-1}(u) + \eta_t(u), & \text{for } t = T_{j-2}^0 + 1, ..., T_{j-1}^0; \\ \psi_{j,1}(u) + \eta_t(u), & \text{for } t = T_{j-1}^0 + 1, ..., \tau_x; \\ \psi_{j,2}(u) + \eta_t(u), & \text{for } t = \tau_x + 1, ..., T_j^0; \\ \psi_{j+1}(u) + \eta_t(u), & \text{for } t = T_j^0 + 1, ..., T_{j+1}^0; \\ \vdots & \vdots \\ \psi_{M^0+1}(u) + \eta_t(u), & \text{for } t = T_{M^0}^0 + 1, ..., T; \end{cases}$$

where $\psi_{j,1}(u) = E\left[h_j^0(X_{1t})e^{\mathbf{i}u'X_{1t}}\right]$, $\psi_{j,2}(u) = E\left[h_j^0(X_{2t})e^{\mathbf{i}u'X_{2t}}\right]$, and

$$\psi_l(u) = \begin{cases} E\left[h_l^0(X_{1t})e^{\mathbf{i}u'X_{1t}}\right], & \text{for } l = 1, ..., j-1; \\ E\left[h_l^0(X_{2t})e^{\mathbf{i}u'X_{2t}}\right], & \text{for } l = j+1, ..., M^0+1. \end{cases}$$

Intuitively, the breaks in X_t will lead to additional breaks in the generalized regression. To purge the impact of the nonstationarity in X_t, we can first estimate structural breaks in the joint distribution of X_t by existing approaches such as Inoue (2001) and obtain the estimated break date $\hat{\tau}_x$ for τ_x. Then we can estimate break fractions $\{r_1^0, ..., r_{M^0}^0\}$ in $\left\{Y_t e^{\mathbf{i}u'X_t}\right\}_{t=1}^T$ by solving the following optimization problem

$$\min_{\{r_1,...,r_{M^0}\}\in\Pi_\epsilon} \left\{\sum_{l=1}^{j-1}\sum_{t=T_{l-1}+1}^{T_l}\int_{\mathbb{R}^d}\|Y_t e^{\mathbf{i}u'X_{1t}} - \tilde{\psi}_l(u)\|^2 W(u)\mathrm{d}u + \sum_{t=T_{j-1}+1}^{\hat{\tau}_x}\int_{\mathbb{R}^d}\|Y_t e^{\mathbf{i}u'X_{1t}} - \tilde{\psi}_{j,1}(u)\|^2 W(u)\mathrm{d}u \right.$$
$$\left. + \sum_{t=\hat{\tau}_x+1}^{T_j}\int_{\mathbb{R}^d}\|Y_t e^{\mathbf{i}u'X_{2t}} - \tilde{\psi}_{j,2}(u)\|^2 W(u)\mathrm{d}u + \sum_{l=j+1}^{M^0+1}\sum_{t=T_{l-1}+1}^{T_l}\int_{\mathbb{R}^d}\|Y_t e^{\mathbf{i}u'X_{2t}} - \tilde{\psi}_l(u)\|^2 W(u)\mathrm{d}u \right\},$$

where $\tilde{\psi}_{j,1}(u) = (\hat{\tau}_x - T_{j-1})^{-1}\sum_{t=T_{j-1}+1}^{\hat{\tau}_x} Y_t e^{\mathbf{i}u'X_{1t}}$, $\tilde{\psi}_{j,2}(u) = (T_j - \hat{\tau}_x)^{-1}\sum_{t=\hat{\tau}_x+1}^{T_j} Y_t e^{\mathbf{i}u'X_{2t}}$, and

$$\tilde{\psi}_l(u) = \begin{cases} (T_l - T_{l-1})^{-1}\sum_{t=T_{l-1}+1}^{T_l} Y_t e^{\mathbf{i}u'X_{1t}}, & \text{for } l = 1, ..., j-1; \\ (T_l - T_{l-1})^{-1}\sum_{t=T_{l-1}+1}^{T_l} Y_t e^{\mathbf{i}u'X_{2t}}, & \text{for } l = j+1, ..., M^0+1. \end{cases}$$

Following analogous arguments established in Sections 3 and 4, we can show the consistency of our estimator for break fractions and derive the limiting distribution of the corresponding sup-F test. Through such a sample splitting approach, we can extend our methodology to the case with nonstationary regressors.

6.2 Endogenous Regressors

Endogeneity is a common phenomenon in time series applications. It may arise due to measurement errors, omitted variables, or simultaneous equation bias. For example, since macroeconomic factors have nonnegligible measurement errors (Connor and Korajczyk, 1986, 1991; Ferson and Harvey, 1999), the asset pricing models incorporating some macroeconomic factors may suffer from the endogeneity problem (Chen et al., 1986).

A few papers discuss the endogeneity issue when analyzing structural breaks. In the parametric framework, Andrews (1993) considers test for structural breaks in a GMM framework which is applicable to cases with endogeneity. Hall et al. (2012) consider a linear time series regression model with endogenous covariates. They extend Bai and Perron's (1998) approach to estimate and test structural breaks via the two-stage least squares (2SLS) estimation. Qian and Su (2014) adopt the group-fused Lasso approach to simultaneously estimate the number and location of breaks in piecewise linear time series models with endogenous regressors. Perron and Yamamoto (2015) advocate an OLS-based approach to analyze structural breaks in a linear time series model with endogeneity. They show that even in the presence of endogenous regressors, the OLS-based approach is still preferable in that it delivers more accurate estimates of the break dates and more powerful tests than those using an instrumental variable (IV) method. Chen (2015) develops a two-stage local linear (2SLL) estimator and a Wald-type test for smooth structural changes, which can be regarded as a generalization of Chen and Hong's (2012) test for smooth structural changes.

The aforementioned tests are proposed under the parametric framework. The results can be misleading when a parametric model is misspecified. Fu and Hong (2019) propose a model-free consistent test for structural changes in nonparametric regressions, which is applicable to endogenous and discrete regressors. Despite being free of misspecification, their approach needs instrumental variables, and they do not consider estimating structural changes.

Besides, when endogeneity exists, i.e., $E(\varepsilon_t|X_t) = \mu(X_t)$, direct nonparametric instrumental estimation for $m_t(\cdot)$ will incur the "ill-posed inverse problem", e.g., Newey and Powell (2003), Darolles et al. (2011), and Horowitz (2011). In contrast, our approach is free of smoothed nonparametric estimation and does not rely on instruments. As long as the endogeneity is time-invariant, it will not affect estimation and testing for structural breaks in the unknown regression function.

In specific, multiply $e^{\mathbf{i}u'X_t}$ on both sides of (2.1) and take expectation, we have

$$E(Y_t e^{\mathbf{i}u'X_t}) = E\{[m_t(X_t) + \mu(X_t)]e^{\mathbf{i}u'X_t}\}, \tag{6.1}$$

where $E\{[\varepsilon_t - \mu(X_t)]e^{\mathbf{i}u'X_t}\} = 0$ for all u given $\mu(X_t) = E(\varepsilon_t|X_t)$. The remaining process is the same as before except the expression of $\phi_t(u)$ in (2.3). Here, $\phi_t(u) \equiv E\{[m_t(X_t) + \mu(X_t)]e^{\mathbf{i}u'X_t}\}$. Compared with the nonparametric test proposed by Fu and Hong (2019), our approach does not need instrumental variables. It is a salient feature in practice since a valid instrumental variable may not be available.

7 Monte Carlo Studies

In this section, we conduct simulation studies to examine the finite sample performance of our proposed estimators and tests. First, we assess the performance of the IC and sequential test in determining the number of breaks and then evaluate the accuracy of our proposed estimators for break dates. At last, the size and power performance of our test are investigated.

Consider the following DGPs:

DGP.S1: $Y_t = 1 + 0.5X_{1t} + 2X_{2t} + \varepsilon_t$;
DGP.S2: $Y_t = 2 + 0.5X_{1t}^2 + X_{2t}^2 + \varepsilon_t$;
DGP.S3: $Y_t = 1.5 + X_{1t} + \varepsilon_t$, $\varepsilon_t = \sqrt{h_t}\nu_t$, $h_t = 0.2 + 0.3\varepsilon_{t-1}^2$;
DGP.S4: $Y_t = (0.5X_{1t})\mathbf{1}(X_{1t} > 0.5) + (0.5 - X_{1t})\mathbf{1}(X_{1t} \leq 0.5) + \varepsilon_t$;

DGP.P1:

$$Y_t = \begin{cases} 2 + X_{1t} + \varepsilon_t, & \text{if } t \leq 0.5T \\ 0.5 - X_{1t} + \varepsilon_t, & \text{otherwise} \end{cases};$$

DGP.P2:

$$Y_t = \begin{cases} X_{1t}X_{2t} + \varepsilon_t, & \text{if } t \leq 0.5T \\ 2 + 0.8X_{1t}X_{2t} + \varepsilon_t, & \text{otherwise} \end{cases};$$

DGP.P3:

$$Y_t = \begin{cases} (1.5 + 2X_{1t})\mathbf{1}(X_{1t} > 0) + (1.5 - 2X_{1t})\mathbf{1}(X_{1t} \leq 0) + \varepsilon_t, & \text{if } t \leq 0.5T \\ (-2X_{1t})\mathbf{1}(X_{1t} > 0.5) + (2X_{1t})\mathbf{1}(X_{1t} \leq 0.5) + \varepsilon_t, & \text{otherwise} \end{cases};$$

DGP.P4:

$$Y_t = \begin{cases} 3 - 0.2X_{1t} + X_{2t} + \varepsilon_t, & \text{if } 1 \leq t \leq 0.3T \\ -0.2X_{1t} + X_{2t} + \varepsilon_t, & \text{if } 0.3T + 1 \leq t \leq 0.6T \\ -3 - 0.2X_{1t} + X_{2t} + \varepsilon_t, & \text{if } 0.6T + 1 \leq t \leq T \end{cases};$$

where $\varepsilon_t \sim i.i.d.N(0,1)$, $\nu_t \sim i.i.d.N(0,1)$, $X_{1t} = 0.5X_{1(t-1)} + \nu_{1t}$, and $X_{2t} = 0.4X_{2(t-1)} + \nu_{2t}$, with $\nu_{1t} \sim i.i.d.N(0,1)$ and $\nu_{2t} \sim i.i.d.N(0,1)$.

DGPs.S1–S4 cover various linear and nonlinear time series models without structural breaks. They allow us to examine the size performance of the proposed test statistics. Specifically, DGPs.S1 and S3 are linear time series models with i.i.d. and ARCH errors, respectively. DGP.S2 is a nonlinear time series model with quadratic polynomials, and DGP.S4 is a threshold regression model. DGPs.P1–P4 depict both linear and nonlinear time series models with structural breaks. Among them, DGPs.P1–P3 admit a single structural break. In particular, DGP.P1 is a linear time series model, while DGPs.P2 and P3 depict two kinds of nonlinear time series models with the regressors in a multiplicative and threshold form, respectively. DGP.P4 is a linear time series model with multiple structural breaks.

To compute the SSGR, we choose the standard normal density function as the weighting function $W(\cdot)$. For each DGP, we generate 1000 datasets with sample size $T = 100, 200$, and 500, respectively. We set the number of bootstrapping $\mathbf{B} = 200$. We consider the most frequently used trimming parameter $\epsilon = 0.1$. For the IC procedure, the tuning parameter is set to be $\rho_T = c_\rho \ln(T)/T$, with $c_\rho = 2$. In addition, we choose the upper bound of the number of breaks as $M_{\max} = 4$.

To evaluate the performance of our IC and sequential tests in determining the number of breaks, we compute the average number of breaks and the percentage of correct selection over 1000 replications. We also compare our results with Bai and Perron's (1998)(BP, hereafter) sequential tests proposed for linear time series regressions.

Table 1 reports the results under various DGPs. For the linear time series regressions specified by DGPs.S1, P1, and P4, our IC and sequential tests do not perform as comparable as the sequential tests of BP when the sample size is small. This is because when a linear model is correctly specified, BP's approach is valid and more efficient than those designed for nonparametric models. However, the performance of our methods improves as the sample size increases. For the nonlinear time series regressions given by DGPs.S2–S4, P2, and P3, our IC and sequential tests work fairly well. The percentage of correct selection for IC grows to 100% and that for sequential tests approaches to the corresponding nominal significance levels. In contrast, BP's sequential tests tend to overestimate the number of breaks, and the percentage of correct selection decreases as the sample size grows. We note that the over-selection problem of BP is caused by the model misspecification. This shows the salient feature of our approach. When a correct specification is not available or the true DGP is nonlinear, our approach can deliver robust estimation for the true number of breaks.

Given the consistent estimate for the number of breaks, we now evaluate the performance of the proposed estimator for break dates under DGPs.P1–P4 with the number of breaks being specified as the true value. We use the bias and root-mean-squared errors (RMSE) to measure the accuracy

Table 1: Performance on determining the number of breaks. $ST5$ and $ST10$ denote BP's sequential test at the 5% and 10% significance levels, respectively; NIC denotes the information criterion proposed in this paper; $NST5$ and $NST10$ denote the sequential tests proposed in this paper at the 5% and 10% significance levels, respectively. Numbers in the main entries are the results based on 1000 replications.

DGP	T	Average number of breaks					Percentage of correct selection				
		$ST5$	$ST10$	NIC	$NST5$	$NST10$	$ST5$	$ST10$	NIC	$NST5$	$NST10$
S1	100	0.051	0.088	0.015	0.054	0.133	95.0	91.4	99.3	95.0	88.8
	200	0.048	0.099	0.006	0.062	0.126	95.2	90.3	99.7	94.2	89.3
	500	0.056	0.119	0.001	0.045	0.133	94.5	88.5	99.9	95.7	88.3
S2	100	0.987	1.238	0.051	0.044	0.121	36.3	25.3	97.9	95.9	89.7
	200	1.193	1.469	0.028	0.052	0.120	27.8	18.2	99.1	95.1	89.4
	500	1.440	1.707	0.004	0.040	0.122	21.2	14.3	99.9	96.1	89.5
S3	100	0.059	0.112	0.030	0.052	0.142	94.2	89.3	98.1	95.1	88.0
	200	0.089	0.145	0.005	0.076	0.171	91.5	86.6	99.8	92.9	85.2
	500	0.124	0.201	0.000	0.055	0.136	88.4	81.9	100	94.6	87.7
S4	100	0.234	0.356	0.093	0.074	0.172	78.5	69.4	94.6	93.1	85.3
	200	0.272	0.413	0.074	0.052	0.125	76.5	65.9	96.1	95.2	89.4
	500	0.354	0.498	0.030	0.075	0.145	69.6	59.9	98.8	93.5	88.5
P1	100	1.040	1.056	1.292	0.926	1.157	96.0	94.4	79.6	80.2	80.7
	200	1.023	1.046	1.261	1.075	1.172	97.9	95.6	84.6	91.7	85.0
	500	1.031	1.063	1.160	1.080	1.135	96.9	93.9	90.3	92.0	87.0
P2	100	1.485	1.615	0.582	0.786	1.081	57.0	47.8	48.8	70.5	82.1
	200	1.620	1.800	1.035	1.036	1.118	47.8	36.1	94.9	95.1	90.0
	500	1.740	1.915	1.036	1.015	1.105	40.4	30.2	98.6	98.5	91.0
P3	100	1.719	1.877	1.415	1.078	1.213	41.7	33.4	72.9	88.2	80.8
	200	1.778	1.924	1.305	1.087	1.177	40.1	32.0	80.6	91.9	84.6
	500	1.871	2.067	1.242	1.055	1.120	34.1	25.0	85.1	94.5	90.0
P4	100	2.047	2.065	1.035	0.875	1.502	95.3	93.5	2.7	12.9	36.5
	200	2.031	2.041	1.639	1.792	2.047	96.9	95.9	63.0	74.8	84.7
	500	2.011	2.025	2.005	2.055	2.125	98.9	97.5	99.5	95.0	88.5

of the estimator, defined as

$$\text{Bias} = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}(\hat{r}_{ji} - r_j^0),$$

$$\text{RMSE} = \sqrt{\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}(\hat{r}_{ji} - r_j^0)^2},$$

where M is the number of breaks, N is the number of replications, r_j^0 is the true value of the jth break fraction, and $\hat{r}_{ji}$ is the estimator for the jth break fraction in the ith replication.

Table 2 reports the performance of our estimator and that of Delgado and Hidalgo (2000) (DH, hereafter). The RMSEs of our estimated break fractions are much lower than those of DH under all DGPs. Besides, our estimator dominates that of DH in terms of Bias under DGPs.P2–P4, especially when the sample size is large. It also shows that our estimators converge faster to the true value than those of DH, in that the RMSEs of our estimators decrease quickly as the sample size increases. Note that the convergence of the DH's estimator is adversely affected by the dimension of regressors. Hence, it delivers an inferior performance than ours in RMSE under all DGPs. Under the univariate time series regression specified by DGP.P1, even though our estimator has a larger Bias (in absolute values), it still outperforms DH in RMSE. This implies the improvement of our approach over the existing ones that rely on smoothed nonparametric regression.

We now examine the size and power performance of our test in finite samples. In addition, we compare ours with the nonparametric tests by Su and Xiao (2008) and Fengler et al. (2015). For the prespecified number of breaks under the alternative, we consider $M = 1, 2$ but only report the results for $M = 1$ to save space. The results under $M = 2$ are quite similar and available from the authors upon request.

Table 3 reports the size of all tests under DGPs.S1–S4 at the 5% and 10% significance levels. Our sup-F test has reasonable size performance since the empirical rejection rates are close to the corresponding nominal levels. Both the Kolmogorov–Smirnov (KS) and Cramér-von Mises (CvM) test statistics of Su and Xiao (2008) have reasonable size under DGPs.S3 and S4 but are oversized under DGPs.S1 and S2. Fengler et al.'s (2015) test tends to over-reject the null hypotheses when the sample size is small, but the over-rejection is mitigated as the sample size increases. Table 4 reports the power of tests under DGPs.P1–P4 at the 5% and 10% significance levels. Our sup-F test is more powerful than Su and Xiao's (2008) tests in detecting structural breaks under DGPs.P2–P4. For the linear time series regression given by DGP.P1, our test performs slightly worse than

Table 2: Performance on estimating break fractions. Numbers in the main entries are Bias and RMSEs $\times 1000$. The bold entries highlight the better performance in each case.

		This Paper		Delgado and Hidalgo (2000)	
		Bias	RMSE	Bias	RMSE
P1	$T = 100$	-34.740	**80.961**	**-9.006**	250.288
	$T = 200$	-16.465	**42.634**	**-1.260**	232.882
	$T = 500$	-6.202	**13.800**	**4.389**	214.875
P2	$T = 100$	13.980	**63.389**	**-9.797**	246.561
	$T = 200$	5.020	**25.274**	**1.579**	239.364
	$T = 500$	**0.560**	**5.171**	-4.762	210.456
P3	$T = 100$	**0.560**	**14.107**	-5.831	241.508
	$T = 200$	**0.285**	**2.797**	3.789	220.410
	$T = 500$	**0.140**	**1.070**	2.913	198.831
P4	$T = 100$	**16.415**	**107.637**	47.401	175.124
	$T = 200$	**-2.110**	**34.483**	31.451	175.095
	$T = 500$	**-1.173**	**9.033**	29.843	167.269

Su and Xiao (2008). Fengler et al.'s (2015) test can detect structural breaks under DGPs.P1 and P3, which belong to the class of additive models. However, it fails to capture structural breaks in DGPs.P2 and P4 because DGP.P2 is a non-additive model, and DGP.P4 only has structural breaks in the intercept.

Table 3: Size of tests under DGPs.S1–S4. F_T denotes our sup-F test for the alternative hypothesis of one break point; $SX08, KS$ and $SX08, CvM$ denote the KS and CvM test statistics of Su and Xiao (2008); $Fe15$ denote Fengler et al.'s (2015) test. The main entries report the percentage of rejection.

		F_T		$SX08, KS$		$SX08, CvM$		$Fe15$	
		5%	10%	5%	10%	5%	10%	5%	10%
S1	$T = 100$	5.5	11.1	12.5	18.8	11.0	16.0	10.4	17.4
	$T = 200$	5.6	11.0	7.8	14.8	8.5	14.7	9.0	14.8
	$T = 500$	6.3	11.7	9.0	14.9	8.9	14.1	4.8	10.1
S2	$T = 100$	6.3	12.6	8.4	15.0	9.0	15.1	11.2	18.9
	$T = 200$	5.5	11.5	8.8	13.5	8.6	12.8	8.3	14.4
	$T = 500$	6.7	12.5	7.9	14.7	8.6	14.6	5.6	11.1
S3	$T = 100$	5.4	13.4	5.3	11.6	6.0	10.2	7.9	14.6
	$T = 200$	5.4	12.3	5.5	12.2	7.0	11.9	6.8	12.1
	$T = 500$	5.2	11.5	5.6	9.9	5.3	10.6	3.8	9.1
S4	$T = 100$	5.6	12.8	6.3	12.9	6.5	11.3	8.4	16.1
	$T = 200$	4.9	11.6	5.4	9.9	5.7	11.1	7.3	12.8
	$T = 500$	5.4	11.5	6.1	10.9	6.2	10.5	4.6	9.1

Table 4: Power of tests under DGPs.P1-P4

		F_T		$SX08, KS$		$SX08, CvM$		$Fe15$	
		5%	10%	5%	10%	5%	10%	5%	10%
P1	$T = 100$	86.2	97.5	99.5	99.6	99.5	99.6	87.7	92.0
	$T = 200$	99.4	99.9	100	100	100	100	99.4	99.6
	$T = 500$	100	100	100	100	100	100	100	100
P2	$T = 100$	76.7	94.7	55.0	68.1	48.7	64.2	11.1	18.4
	$T = 200$	99.4	100	89.2	93.0	86.6	90.9	9.8	16.5
	$T = 500$	100	100	100	100	99.8	100	6.6	12.3
P3	$T = 100$	97.6	99.9	59.9	65.6	57.3	64.7	99.2	99.7
	$T = 200$	100	100	73.8	79.1	71.5	77.3	100	100
	$T = 500$	100	100	87.6	92.5	87.3	90.9	100	100
P4	$T = 100$	73.7	99.2	73.7	80.1	73.2	80.0	11.5	17.0
	$T = 200$	98.4	100	97.5	98.6	97.2	98.0	10.9	17.9
	$T = 500$	100	100	100	100	100	100	7.0	12.7

8 Empirical Application

The capital asset pricing model (CAPM), pioneered by Sharpe (1964) and Lintner (1965), assumes that the expected excess return of an asset co-moves with the expected excess return of the market portfolio. As the cornerstone of theoretical and empirical asset pricing, the CAPM along with its factor loading and intercept term has been thoroughly investigated. The existing literature documents strong evidence that the model coefficients may be affected by certain state variables. For example, Ferson and Harvey (1999) and Petkova and Zhang (2005) model the alpha and beta coefficients as linear functions of certain state variables. To avoid model misspecification, recent studies adopt a nonparametric or semi-parametric approach to capture the nonlinear dependence of the conditional alpha and beta on prespecified state variables. See, for example, Wang (2003), Ferreira et al. (2011), and Cai et al. (2020).

In addition to the fact that the model coefficients depend on certain state variables, they may also suffer from structural breaks. The factors such as fluctuations in investor sentiments, rapid transitions, and frequent reforms in the financial market can lead to structural breaks in the model coefficients. In fact, some papers try to model and test for structural breaks in coefficients under the parametric frameworks. For example, Ghysels (1998) documents significant evidence of structural breaks in beta risk dynamics; Maroney et al. (2004) model the time-varying beta behavior by a logistic function; Aue et al. (2012) assume the beta coefficient to be piecewise constant; Smith and Timmermann (2021) capture the time-varying feature in the beta coefficient by a piecewise linear function.

The literature mentioned above has made fruitful achievements in modeling the instability of the model coefficients in CAPM. However, when a parametric model is misspecified, a rejection of stability can be caused by model misspecification rather than structural breaks (Fu and Hong, 2019). Ghysels (1998) also notes that if the coefficients are misspecified, introducing the time-varying behavior will adversely impact estimation. To avoid model misspecification, we consider the following conditional CAPM that combines both the state dependence and time-varying features:

$$R_{i,t+1} = \alpha_{i,t}(Z_t) + \beta_{i,t}(Z_t)R_{m,t+1} + \epsilon_{t+1}, \quad i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \tag{8.1}$$

where $R_{i,t+1}$ is the excess return of portfolio i, $R_{m,t+1}$ is the market excess return, the intercept term $\alpha_{i,t}(Z_t)$ and the factor loading $\beta_{i,t}(Z_t)$ are unknown and possibly time-varying functions of the state variable Z_t.

Denote $\gamma_{i,t}(Z_t) \equiv (\alpha_{i,t}(Z_t), \beta_{i,t}(Z_t))'$. We now examine the stability of the model coefficients $\gamma_{i,t}(Z_t)$ for ten portfolios constructed by Kenneth French. Specifically, Kenneth French sorts all of the NYSE, AMEX, and NASDAQ stocks by the book-to-market ratio from the lowest to the highest and then split them into ten decile groups. We let the i-th portfolio corresponds to the i-th group. Since the first portfolio has the lowest book-to-market ratio, the stocks that belong to it are called the "growth stocks". The last portfolio has the highest book-to-market ratio, and the stocks that belong to it are called the "value stocks". All portfolios are value-weighted. The excess return of the ith portfolio is calculated as the value-weighted returns of all stocks belonging to portfolio i minus the one-month Treasury bill rate (from Ibbotson Associates). Similarly, the market excess return is constructed by the value-weighted returns of all CRSP firms incorporated in the U.S. and listed on the NYSE, AMEX, or NASDAQ minus the one-month Treasury bill rate. The return data are also available from Kenneth French, which is widely used by many researchers such as Ferreira et al. (2011), Ang and Kristensen (2012), and Guo et al. (2017), among many others.

Following the existing literature such as Welch and Goyal (2008), Ferreira et al. (2011), and Lee et al. (2014), the state variable is chosen from the following three variables: (1) long term rate of returns (LTR), i.e., the return on long-term government bonds; (2) default return spread (DFR), i.e., the difference between long-term corporate bond and long-term government bond returns; (3) earnings price ratio (EP), i.e., the difference between the log of earnings and the log of prices. We obtain these variables from Amit Goyal's website (http://www.hec.unil.ch/agoyal). Following Ang and Kristensen (2012) and Guo et al. (2017), we consider the post-1963 sample and use monthly data spanning July 1963 to December 2019, with 678 observations.

Table 5: Bootstrap p-values of structural break test based on $\mathbf{B} = 200$ bootstrap resamples

Portfolio	LTR	DFR	EP	{LTR,DFR}	{LTR,EP}	{DFR,EP}	{LTR,DFR,EP}
1	0.000	0.000	0.040	0.000	0.035	0.035	0.060
2	0.000	0.005	0.015	0.000	0.010	0.015	0.010
3	0.000	0.000	0.025	0.000	0.010	0.025	0.010
4	0.000	0.000	0.030	0.000	0.005	0.025	0.010
5	0.000	0.000	0.025	0.000	0.005	0.025	0.005
6	0.000	0.005	0.020	0.000	0.000	0.010	0.010
7	0.000	0.005	0.055	0.000	0.035	0.070	0.040
8	0.000	0.010	0.050	0.000	0.030	0.050	0.040
9	0.000	0.005	0.080	0.000	0.010	0.020	0.015
10	0.000	0.010	0.095	0.000	0.040	0.085	0.040

Table 5 reports the bootstrap p-values of our joint sup-F test for univariate and multivariate coefficient functions using the above three state variables. For the univariate coefficient functions, we can reject the null hypothesis of time-invariant coefficients for all state variables and all portfolios considered. Specifically for LTR and DFR, the time-invariant coefficient function is rejected at the 1% significance level. Moreover, the strong evidence of structural breaks carries over to the multivariate coefficient functions. Our test can reject the stability of three bivariate coefficient functions and one trivariate coefficient function at the 5% level for all portfolios except for a few cases at the 10% level. According to the results, the alpha and beta coefficient are time-varying, which represent the pricing error and the response of the excess return of book-to-market portfolios to systematic risk, respectively.

Table 6: Estimation of break dates

Panel A: Univariate coefficient function			
Portfolio	LTR	DFR	EP
1 (growth)	{Oct. 2000, Sep. 2002}	Jan. 2003	Nov. 1990
10 (value)	Mar. 2001	{Jun. 2007, Dec. 2008, Apr. 2009}	Dec. 1990

Panel B: Multivariate coefficient function				
Portfolio	{LTR,DFR}	{LTR,EP}	{DFR,EP}	{LTR,DFR,EP}
1 (growth)	Mar. 2001	Nov. 1990	Nov. 1990	Nov. 1990
10 (value)	Mar. 2001	Sep. 1997	Dec. 1990	Sep. 1997

Next, we estimate the number of breaks using the proposed sequential tests. The corresponding locations of breaks are straightforwardly obtained using our method. To save space, we only report the results of the growth portfolio and the value portfolio. As shown in Table 6, all coefficient functions have at least one break point. Among them, the function of LTR for the growth portfolio has two breaks and the function of DFR for the value portfolio has three breaks. Besides, the estimated break dates are reconciled with some important economic and financial events, such as the 1990 economic crisis, 2000 internet bubble, and 2008 global financial crisis, which indicates the rationality of our estimators.

Based on the estimated break dates, we are also interested in estimating the coefficients. Since the model coefficients $\gamma_{i,t}(\cdot)$ are unknown smooth functions of the state variable Z_t in each regime, we use the nonparametric local constant method to estimate them for the subsamples before and after the corresponding break date. Specifically, we choose the Gaussian kernel function $k(w) = \frac{1}{2\pi} e^{-w^2/2}$. To eliminate the bias term in nonparametric regressions, we use the undersmooth bandwidth bandwidth $h_s = \sigma_s T^{-2/9}$, where σ_s is the standard deviation of state variable s. To save space, we only report the results of the univariate coefficient function with DFR for growth portfolio. Given the break date in Jan. 2003, which is now regarded as a crisis throwing the economy back into the 2001 recession, we estimate the conditional alpha and beta before the break (Regime 1) and after the break (Regime 2) separately.

Figure 1 plots the estimated conditional beta and its 95% confidence bands for the growth portfolio before and after the break. The conditional beta is relatively stable after the break, while more fluctuated before the break. This evidence shows that the default spread, as an indicator of the systematic risk, is less volatile to the risk exposure during the financial crisis. Besides, the local smoothing estimates in the two regimes are significantly different at the 5% level, implying that the estimation of the factor loading will be imprecise if ignoring the structural breaks in the conditional beta.

In Figure 2, we plot the estimation of conditional growth alpha and its 95% confidence bands in the two regimes. Again, the estimates, along with the 95% confidence band, show that the pricing error performs differently before and after the break. Especially on the positive DFR horizon, we may conclude there exists a pricing error conditional on the default spread in both regimes with totally different signs. However, when ignoring the structural changes, the whole sample estimation reconciles the difference in these two regimes and possibly leads the researchers to a false conclusion of no pricing error.

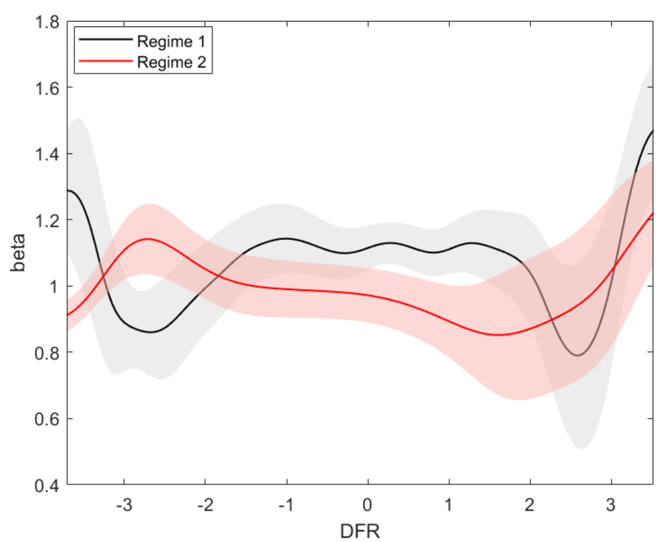

Figure 1: Beta of growth portfolio. The figure shows the estimate of beta function of DFR for growth portfolio. We plot 95% confidence band in shaded area.

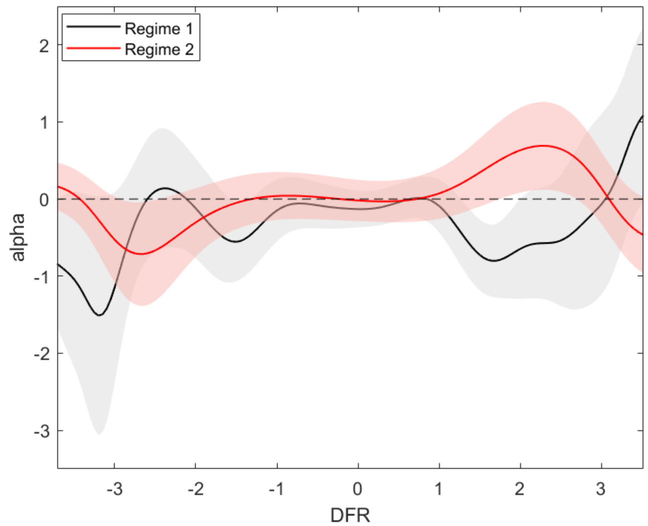

Figure 2: Alpha of growth portfolio. The figure shows the estimate of alpha function of DFR for growth portfolio. We plot 95% confidence band in shaded area.

9 Conclusion

This paper proposes a frequency domain approach to estimate and test multiple structural breaks in nonparametric regressions. Using the equivalence of structural breaks in a nonparametric regression function and those in the Fourier transform of data, we avoid direct smoothed nonparametric estimation of the unknown regression function. Then we can pin down the break dates in nonparametric regression by estimating structural breaks in a generalized regression. By minimizing the SSGR, we can consistently estimate the breaks. A sup-F test statistic is developed for structural breaks by comparing the SSGRs under the null and the alternative. We also provide an information criterion and a sequential testing procedure to determine the number of breaks. Compared to the existing methods, our approach is free of the well-known "curse of dimensionality" problem, and our estimation and inference procedures for structural breaks in nonparametric regressions possess the merits of the parametric procedure. Simulation studies show that our method consistently estimates the break fractions, and the proposed test statistics have reasonable finite sample properties. Our methods are applied to examine the stability of the coefficients in the conditional CAPM. We can reconcile the estimated break dates with economic events and characterize the functional form of model coefficients in each regime.

References

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica 61(4):821-856.

Ang, A., Kristensen, D. (2012). Testing conditional factor models. Journal of Financial Economics 106 (1):132-156.

Aue, A., Hörmann, S., Horváth, L., Huškov á, M., Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. Econometric Theory 28(4):804-837.

Bai, J. (1997). Estimation of a change point in multiple regression models. Review of Economics and Statistics 79(4):551-563.

Bai, J., Perron P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica 66(1):47-78.

Billingsley, P. (1968). Convergence of probability measures. John Wiley & Sons.

Cai, Z., Fang, Y., Xu, Q. (2020). Testing capital asset pricing models using functional-coefficient panel data models with cross-sectional dependence. Journal of Econometrics 227(1):114-133.

Calhoun, G. (2018). Block bootstrap consistency under weak assumptions. Econometric Theory 34(6):1383-1406.

Chen, B. (2015). Modeling and testing smooth structural changes with endogenous regressors. Journal of Econometrics 185(1):196-215.

Chen, B., Hong, Y., (2012). Testing for smooth structural changes in time series models via nonparametric regression. Econometrica 80(3):1157-1183.

Chen, N., Roll, R., Ross, S. (1986). Economic forces and the stock market. Journal of Business 59(3):383-403.

Connor, G., Korajczyk, R. (1986). Performance measurement with the arbitrage pricing theory. Journal of Financial Economics 15(3):373-394.

Connor, G., Korajczyk, R. (1991). The attributes, behavior and performance of US mutual funds. Review of Quantitative Finance & Accounting 1(1):5-26.

Darolles, S., Fan, Y., Florens, J. P., Renault, E. (2011). Nonparametric instrumental regression. Econometrica 79(5):1541-1565.

Delgado, M. A., Hidalgo J. (2000). Nonparametric inference on structural breaks. Journal of Econometrics 96(1):113-144.

Fengler, A., Mammen E., Vogt M. (2015). Specification and structural break tests for additive models with applications to realized variance data. Journal of Econometrics 5(1):88-97.

Ferreira, E., Gil-Bazo, J., Orbe, S. (2011). Conditional beta pricing models: A nonparametric approach. Journal of Banking & Finance 35(12):3362-3382.

Ferson, W. E., Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. Journal of Finance 54(4):1325-1360.

Fu, Z., Hong, Y. (2019). A model-free consistent test for structural change in regression possibly with endogeneity. Journal of Econometrics 211(1):206-242.

Gao, J., Gijbels, I., Van Bellegem, S. (2008). Nonparametric simultaneous testing for structural breaks. Journal of Econometrics 143(1):123-142.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time varying betas help or hurt. Journal of Finance 53(2):549-573.

Guo, H., Wu, C., Yu, Y. (2017). Time-varying beta and the value premium. Journal of Financial and Quantitative Analysis 52(4):1551-1576.

Hall, A. R., Han, S., Boldea, O. (2012). Inference regarding multiple structural changes in linear models with endogenous regressors. Journal of Econometrics 170(2):281-302.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. The Journal of Economic Perspectives 15(4):117-128.

Hidalgo, J. (1995). A nonparametric conditional moment test for structural stability. Econometric Theory 11:671-698.

Hörmann, S., Kokoszka, P. (2010). Weakly dependent functional data. Annals of Statistics 38(3):1845-1884.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. Econometrica 79(2):347-394.

Inoue, A. (2001). Testing for distributional change in time series. Econometric theory 17(1):156-187.

Lee, T. H., Tu, Y., Ullah, A. (2014). Nonparametric and semiparametric regressions subject to monotonicity constraints: estimation and forecasting. Journal of Econometrics 182(1):196-210.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolio and capital budgets. Review of Economics and Statistics 47(1):13-37.

Maroney, N., Naka, A., Wansi, T. (2004). Changing risk, return, and leverage: The 1997 Asian Financial Crisis. Journal of Financial and Quantitative Analysis 39(1):143-166.

Mohr, M., Neumeyer, N. (2020). Consistent nonparametric change point detection combining CUSUM and marked empirical processes. Electronic Journal of Statistics 14(1):2238-2271.

Müller, H. G. (1992). Change points in nonparametric regression analysis. Annals of Statistics 20(2):737-761.

Newey, W. K., Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. Econometrica 71(5):1565-1578.

Perron, P. (2006). Dealing with structural breaks. Palgrave handbook of econometrics 1(2):278-352.

Perron, P., Yamamoto, Y. (2015). Using OLS to estimate and test for structural changes in models with endogenous regressors. Journal of Applied Econometrics 30(1):119-144.

Petkova, R., Zhang, L. (2005). Is value riskier than growth? Journal of Financial Economics 78(1):187-202.

Phillips, P. C. B. (1984). Time series regression with a unit root, Econometrica 55(2):277-301.

Qian, J., Su, L. (2014). Structural change estimation in time series regressions with endogenous variables. Economics Letters 125(3):415-421.

Qian, J., Su, L. (2016). Shrinkage estimation of regression models with multiple structural changes. Econometric Theory 32(6):1376-1433.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance 19:425-442.

Smith, S. C., Timmermann, A. (2021). Break risk. Review of Financial Studies 34(4):2045-2100.

Stock, J., Watson, M. (2009). Forecasting in dynamic factor models subject to structural instability. Oxford University Press, 1-57.

Su, L., Xiao, Z. (2008). Testing structural change in time-series nonparametric regression models. Statistics and Its Interface 1(2):347-366.

Su, L., White, H. (2010). Testing structural change in partially linear models. Econometric Theory 26: 1761-1806.

Vogt, M. (2015). Testing for structural change in time-varying nonparametric regression models. Econometric Theory 31(4):811-859.

Wang, K. Q. (2003). Asset pricing with conditioning information: A new test. Journal of Finance 58(1):161-196.

Wang, L. (2007). Change-point estimation in long memory nonparametric models with applications. Communications in Statistics-Simulation and Computation 37(1):48-61.

Welch, I., Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies 21(4):1455-1508.

Yang, Q., Li, Y. N., Zhang, Y. (2020). Change point detection for nonparametric regression under strongly mixing process. Statistical Papers 61(4):1465-1506.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 52(1):45-57.

Zhang, T., Wu, W. B. (2012). Inference of time-varying regression models. The Annals of Statistics 40(3):1376-1402.

Mathematical Appendix

In this appendix, Sections A and B provide some technical lemmas and the proofs of the main results in the paper, respectively. Throughout the appendix, for $j = 1, ..., M^0 + 1$, we let

$$\eta_t(u) = Y_t e^{\mathbf{i}u'X_t} - \psi_j(u), \text{ for } t \in [T^0_{j-1} + 1, T^0_j],$$
$$\tilde{\eta}_t(u) = Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}^0_j(u), \text{ for } t \in [T^0_{j-1} + 1, T^0_j],$$
$$\hat{\eta}_t(u) = Y_t e^{\mathbf{i}u'X_t} - \hat{\psi}_j(u), \text{ for } t \in [\hat{T}_{j-1} + 1, \hat{T}_j],$$

where $\psi_j(u) = E(Y_t e^{\mathbf{i}u'X_t})$ for $t \in [T^0_{j-1} + 1, T^0_j]$, $\tilde{\psi}^0_j(u) = \left(T^0_j - T^0_{j-1}\right)^{-1} \sum_{t=T^0_{j-1}+1}^{T^0_j} Y_t e^{\mathbf{i}u'X_t}$ for $t \in [T^0_{j-1} + 1, T^0_j]$, and $\hat{\psi}_j(u) = \left(\hat{T}_j - \hat{T}_{j-1}\right)^{-1} \sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j} Y_t e^{\mathbf{i}u'X_t}$ for $t \in [\hat{T}_{j-1} + 1, \hat{T}_j]$.

A Technical Lemmas

Lemma A.1 *Suppose Assumptions 1-3 hold, then for any partition* $\{T_j\}_{j=1}^M$ *and any specified number of breaks* M*,*

$$\frac{1}{T} \int_{\mathbb{R}^d} \left\| \sum_{t=T_{j-1}+1}^{T_j} \eta_t(u) \right\|^2 W(u) du = O_P(1).$$

Proof of Lemma A.1. Given M is a finite number, without loss of generality, we let $j = 1$. By the definition of $\eta_t(u)$, we have

$$\begin{aligned}
E\left[\frac{1}{T}\int_{\mathbb{R}^d}\left\|\sum_{t=1}^{T_1}\eta_t(u)\right\|^2 W(u)\mathrm{d}u\right] &= \int_{\mathbb{R}^d}\frac{1}{T}\sum_{s=1}^{T_1}\sum_{t=1}^{T_1}\mathrm{Re}\left\{E\left[\eta_t(u)\eta_s(u)^*\right]\right\}W(u)\mathrm{d}u \\
&\leq \frac{1}{T}\sum_{s=1}^{T_1}\sum_{t=1}^{T_1}\int_{\mathbb{R}^d}\left\|\mathrm{cov}\left(Y_t e^{\mathbf{i}u'X_t}, Y_s e^{-\mathbf{i}u'X_s}\right)\right\|W(u)\mathrm{d}u \\
&\leq 8r_1\left(\sup_{u\in\mathbb{R}^d}\max_{1\leq t\leq T_1}\|Y_t e^{\mathbf{i}u'X_t}\|_q\right)^2\frac{1}{T_1}\sum_{s=1}^{T_1}\sum_{t=1}^{T_1}\alpha(s-t)^{(q-2)/q}\int_{\mathbb{R}^d}W(u)\mathrm{d}u \\
&\leq 8r_1\sum_{l=-\infty}^{\infty}\alpha(l)^{(q-2)/q}\int_{\mathbb{R}^d}W(u)\mathrm{d}u \\
&= O(1).
\end{aligned}$$

The second to last inequality is due to Davydov's inequality and the fact that $r_1 = T_1/T$. The last inequality holds since

$$\begin{aligned}\sup_{u\in\mathbb{R}^d}\max_{1\le t\le T_1}\|Y_te^{\mathbf{i}u'X_t}\|_q &\le \max_{1\le t\le T_1}\left[\max_{1\le j\le M^0+1}|h_j(X_t)|_q+|\varepsilon_t|_q\right]\sup_{u\in\mathbb{R}^d}\left\|e^{\mathbf{i}u'X_t}\right\|_q\\ &\le \mathcal{C},\end{aligned}$$

for any $q>2$ under Assumptions 1(iii), 1(iv), and 3(i) and the fact that $\sup_{u\in\mathbb{R}^d}\left\|e^{\mathbf{i}u'X_t}\right\|_q\le 1$. And the last equality holds due to Assumptions 1(i) and 2. Hence, by Markov inequality,

$$\frac{1}{T}\int_{\mathbb{R}^d}\left\|\sum_{t=1}^{T_1}\eta_t(u)\right\|^2 W(u)\mathrm{d}u=O_P(1).$$

■

Lemma A.2 *Suppose that there are M^0 structural breaks in the functional form of $m_t(\cdot)$. If the model is under-specified, i.e., the estimated number of breaks $M<M^0$, then we have that the estimated break fractions $\{\hat r_k\}_{k=1}^M$ are consistent for M breaks contained in the collection of true break fractions $\left\{r_j^0\right\}_{j=1}^{M^0}$, such that*

$$\hat r_k\xrightarrow{p} r_j^0$$

for $k=1,\ldots,M$ and some corresponding $j=1,\ldots,M^0$.

Proof of Lemma A.2 Without loss of generality, we assume that the true number of breaks is $M^0=2$ and the corresponding break fractions are r_1^0 and r_2^0, such that $r_1^0<r_2^0$. Suppose (3.2) is solved by setting the number of breaks at $M=1$. Denote the estimated break fraction by $\hat r$, we want to show that $\hat r$ is consistent for r_1^0 or r_2^0. Let $r\in(\epsilon,1-\epsilon)$, consider the following process

$$\begin{aligned}S_T(r) &= \frac{1}{T}\mathrm{SSGR}(rT)\\ &= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\|\tilde\eta_t(u,r)\|^2 W(u)\mathrm{d}u,\end{aligned}$$

where $\tilde\eta_t(u,r)=Y_te^{\mathbf{i}u'X_t}-\tilde\psi_k(u,r)$ with $\tilde\psi_k(u,r)=\frac{1}{T_k-T_{k-1}}\sum_{t=T_{k-1}+1}^{T_k}Y_te^{\mathbf{i}u'X_t}$ for $t\in[T_{k-1},T_k]$, $k=1,2$. Follow the convention that $T_0=0$ and $T_2=T$, we note that $T_1=\lfloor Tr\rfloor$.

Let $\tilde d_t(u,r)=\tilde\psi_k(u,r)-\psi_j(u)$ for $t\in[T_{k-1}+1,T_k]\cap[T_{j-1}^0+1,T_j^0]$ with $k=1,2$ and $j=1,2,3$. Then it follows

$$\begin{aligned}S_T(r) &= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\left\|\eta_t(u)-\tilde d_t(u,r)\right\|^2 W(u)\mathrm{d}u\\ &= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\|\eta_t(u)\|^2 W(u)\mathrm{d}u+\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\left\|\tilde d_t(u,r)\right\|^2 W(u)\mathrm{d}u\end{aligned}$$

$$
\begin{aligned}
& -\frac{2}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\tilde{d}_t(u,r)^*\right]W(u)\mathrm{d}u \\
= & \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\|\eta_t(u)\|^2 W(u)\mathrm{d}u + S_T^{(1)}(r) - 2S_T^{(2)}(r),
\end{aligned}
$$

where

$$
S_T^{(1)}(r) \equiv \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\left\|\tilde{d}_t(u,r)\right\|^2 W(u)\mathrm{d}u,
$$

and

$$
S_T^{(2)}(r) \equiv \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\tilde{d}_t(u,r)^*\right]W(u)\mathrm{d}u.
$$

Apparently, the first term $\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\|\eta_t(u)\|^2 W(u)\mathrm{d}u$ does not dependent on r. Then, it suffices to consider $S_T^{(1)}(r)$ and $S_T^{(2)}(r)$.

We first show that $S_T^{(2)}(r) = o_P(1)$ for all $r \in (\epsilon, 1-\epsilon)$.

$$
\begin{aligned}
S_T^{(2)}(r) &= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\tilde{d}_t(u,r)^*\right]W(u)\mathrm{d}u \\
&= \frac{1}{T}\sum_{k=1}^{2}\sum_{t=T_{k-1}}^{T_k}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\tilde{\psi}_k(u,r)^*\right]W(u)\mathrm{d}u \\
&\quad -\frac{1}{T}\sum_{j=1}^{3}\sum_{t=T_{j-1}^0}^{T_j^0}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\psi_j(u)^*\right]W(u)\mathrm{d}u \\
&= Q_1(r) + Q_2, \text{ say.}
\end{aligned}
$$

Following analogous treatment in the proof of Theorem 3.1, we have $Q_1(r) = O_P(T^{-1/2})$ for all $r \in (\epsilon, 1-\epsilon)$ and $Q_2 = O_P(T^{-1/2})$. Thus, $S_T^{(2)}(r) = o_P(1)$ for all $r \in (\epsilon, 1-\epsilon)$.

Now, we consider $S_T^{(1)}(r)$. When $r = r_1^0$, we have $\tilde{\psi}_1(u, r_1^0) = \psi_1(u) + \frac{1}{T_1^0}\sum_{t=1}^{T_1^0}\eta_t(u)$, and $\tilde{\psi}_2(u, r_1^0) = \frac{1}{T-T_1^0}\sum_{t=T_1^0+1}^{T} Y_t e^{\mathbf{i}u'X_t} = \frac{T_2^0-T_1^0}{T-T_1^0}\psi_2(u) + \frac{T-T_2^0}{T-T_1^0}\psi_3(u) + \frac{1}{T-T_1^0}\sum_{t=T_1^0+1}^{T}\eta_t(u)$. Then, it follows

$$
\begin{aligned}
S_T^{(1)}(r_1^0) &= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\left\|\tilde{d}_t(u,r_1^0)\right\|^2 W(u)\mathrm{d}u \\
&= \frac{1}{T}\int_{\mathbb{R}^d}\left[T_1^0\left\|\tilde{\psi}_1(u,r_1^0)-\psi_1(u)\right\|^2 + (T_2^0-T_1^0)\left\|\tilde{\psi}_2(u,r_1^0)-\psi_2(u)\right\|^2\right. \\
&\quad \left. +(T-T_2^0)\left\|\tilde{\psi}_2(u,r_1^0)-\psi_3(u)\right\|^2\right]W(u)\mathrm{d}u
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^d} \Bigg\{ r_1^0 \left\| \frac{1}{T_1^0} \sum_{t=1}^{T_1^0} \eta_t(u) \right\|^2 + (r_2^0 - r_1^0) \left\| \frac{T - T_2^0}{T - T_1^0} [\psi_3(u) - \psi_2(u)] + \frac{1}{T - T_1^0} \sum_{t=T_1^0+1}^{T} \eta_t(u) \right\|^2 \\
&\quad + (1 - r_2^0) \left\| \frac{T_2^0 - T_1^0}{T - T_1^0} [\psi_2(u) - \psi_3(u)] + \frac{1}{T - T_1^0} \sum_{t=T_1^0+1}^{T} \eta_t(u) \right\|^2 \Bigg\} W(u) \mathrm{d}u \\
&= \frac{(1 - r_2^0)(r_2^0 - r_1^0)}{1 - r_1^0} \int_{\mathbb{R}^d} \|\psi_2(u) - \psi_3(u)\|^2 W(u) \mathrm{d}u + o_P(1),
\end{aligned}
$$

where the last equality is due to Lemma A.1. Analogously, when $r = r_2^0$, we can show that

$$
S_T^{(1)}(r_2^0) = \frac{r_1^0(r_2^0 - r_1^0)}{r_2^0} \int_{\mathbb{R}^d} \|\psi_1(u) - \psi_2(u)\|^2 W(u) \mathrm{d}u + o_P(1).
$$

Without loss of generality, we assume

$$
\frac{(1 - r_2^0)(r_2^0 - r_1^0)}{1 - r_1^0} \|\psi_2(u) - \psi_3(u)\|^2 < \frac{r_1^0(r_2^0 - r_1^0)}{r_2^0} \|\psi_1(u) - \psi_2(u)\|^2 . \tag{A.1}
$$

That implies r_1^0 is the asymptotic minimizer for $S_T(r)$ relatively to r_2^0. Now, we divide the set $(\epsilon, 1-\epsilon)$ into three subsets $(\epsilon, r_1^0]$, (r_1^0, r_2^0), and $[r_2^0, 1-\epsilon)$.

When $r \in (\epsilon, r_1^0]$, $T_1 < T_1^0$. It follows $\tilde{\psi}_1(u, r) = \psi_1(u) + \frac{1}{T_1} \sum_{t=1}^{T_1} \eta_t(u)$, and $\tilde{\psi}_2(u, r) = \frac{1}{T-T_1} \sum_{t=T_1+1}^{T} Y_t e^{\mathbf{i}u' X_t} = \frac{T_1^0 - T_1}{T-T_1} \psi_1(u) + \frac{T_2^0 - T_1^0}{T-T_1} \psi_2(u) + \frac{T - T_2^0}{T-T_1} \psi_3(u) + \frac{1}{T-T_1} \sum_{t=T_1+1}^{T} \eta_t(u)$. Then, we have

$$
\begin{aligned}
& S_T(r) - S_T(r_1^0) \\
&= S_T^{(1)}(r) - S_T^{(1)}(r_1^0) + o_P(1) \\
&= \frac{1}{T} \int_{\mathbb{R}^d} \Big[T_1 \left\| \tilde{\psi}_1(u, r) - \psi_1(u) \right\|^2 + (T_1^0 - T_1) \left\| \tilde{\psi}_2(u, r) - \psi_1(u) \right\|^2 \\
&\quad + (T_2^0 - T_1^0) \left\| \tilde{\psi}_2(u, r) - \psi_2(u) \right\|^2 + (T - T_2^0) \left\| \tilde{\psi}_2(u, r) - \psi_3(u) \right\|^2 \Big] W(u) \mathrm{d}u \\
&\quad - \frac{(1 - r_2^0)(r_2^0 - r_1^0)}{(1 - r_1^0)} \int_{\mathbb{R}^d} \|\psi_2(u) - \psi_3(u)\|^2 W(u) \mathrm{d}u + o_P(1) \\
&= \int_{\mathbb{R}^d} \Bigg[r \left\| \frac{1}{T_1} \sum_{t=1}^{T_1} \eta_t(u) \right\|^2 + (r_1^0 - r) \left\| \frac{T_1^0 - T}{T - T_1} \psi_1(u) + \frac{T_2^0 - T_1^0}{T - T_1} \psi_2(u) + \frac{T - T_2^0}{T - T_1} \psi_3(u) + \frac{1}{T - T_1} \sum_{t=T_1+1}^{T} \eta_t(u) \right\|^2 \\
&\quad + (r_2^0 - r_1^0) \left\| \frac{T_1^0 - T_1}{T - T_1} \psi_1(u) + \frac{T_2^0 - T_1^0 - T + T_1}{T - T_1} \psi_2(u) + \frac{T - T_2^0}{T - T_1} \psi_3(u) + \frac{1}{T - T_1} \sum_{t=T_1+1}^{T} \eta_t(u) \right\|^2
\end{aligned}
$$

$$
\begin{aligned}
& +(1-r_2^0)\left\|\frac{T_1^0-T_1}{T-T_1}\psi_1(u)+\frac{T_2^0-T_1^0}{T-T_1}\psi_2(u)+\frac{T_1-T_2^0}{T-T_1}\psi_3(u)+\frac{1}{T-T_1}\sum_{t=T_1+1}^{T}\eta_t(u)\right\|^2\Bigg] W(u)\mathrm{d}u \\
& -\frac{(1-r_2^0)(r_2^0-r_1^0)}{(1-r_1^0)}\int_{\mathbb{R}^d}\|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u+o_P(1) \\
= \quad & \int_{\mathbb{R}^d}\left[(r_1^0-r)\left\|\frac{1-r_1^0}{1-r}\left[\psi_2(u)-\psi_1(u)\right]+\frac{1-r_2^0}{1-r}\left[\psi_3(u)-\psi_2(u)\right]\right\|^2\right. \\
& +(r_2^0-r_1^0)\left\|\frac{r_1^0-r}{1-r}\left[\psi_1(u)-\psi_2(u)\right]+\frac{1-r_2^0}{1-r}\left[\psi_3(u)-\psi_2(u)\right]\right\|^2 \\
& \left.+(1-r_2^0)\left\|\frac{r_1^0-r}{1-r}\left[\psi_1(u)-\psi_2(u)\right]+\frac{r_2^0-r}{1-r}\left[\psi_2(u)-\psi_3(u)\right]\right\|^2\right] W(u)\mathrm{d}u \\
& -\frac{(1-r_2^0)(r_2^0-r_1^0)}{(1-r_1^0)}\int_{\mathbb{R}^d}\|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u+o_P(1) \\
= \quad & \int_{\mathbb{R}^d}\left\{\frac{(1-r_1^0)(r_1^0-r)}{1-r}\left\|\psi_1(u)-\psi_2(u)\right\|^2+\frac{(1-r_2^0)(r_2^0-r)}{1-r}\left\|\psi_2(u)-\psi_3(u)\right\|^2\right. \\
& \left.+2\frac{(r_1^0-r)(1-r_2^0)}{1-r}\mathrm{Re}\left(\left[\psi_1(u)-\psi_2(u)\right]\left[\psi_2(u)-\psi_3(u)\right]^*\right)\right\} W(u)\mathrm{d}u \\
& -\frac{(1-r_2^0)(r_2^0-r_1^0)}{(1-r_1^0)}\int_{\mathbb{R}^d}\|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u+o_P(1) \\
= \quad & \int_{\mathbb{R}^d}\left\{\frac{(1-r_1^0)(r_1^0-r)}{1-r}\left\|\psi_1(u)-\psi_2(u)\right\|^2+\frac{(1-r_2^0)^2(r_1^0-r)}{(1-r)(1-r_1^0)}\left\|\psi_2(u)-\psi_3(u)\right\|^2\right. \\
& \left.+2\frac{(r_1^0-r)(1-r_2^0)}{1-r}\mathrm{Re}\left(\left[\psi_1(u)-\psi_2(u)\right]\left[\psi_2(u)-\psi_3(u)\right]^*\right)\right\} W(u)\mathrm{d}u+o_P(1) \\
\xrightarrow{p} \quad & \frac{r_1^0-r}{(1-r)(1-r_1^0)}\int_{\mathbb{R}^d}\left\|(1-r_1^0)[\psi_1^0(u)-\psi_2^0(u)]+(1-r_2^0)[\psi_2(u)-\psi_3(u)]\right\|^2 W(u)\mathrm{d}u\geq 0
\end{aligned}
$$

for distinct $\psi_1(u), \psi_2(u), \psi_3(u)$ and nonnegative $W(u)$.

For $r \in [r_2^0, 1-\epsilon)$, by symmetry, we have $\lim_{T\to\infty}\left[S_T(r)-S_T(r_2^0)\right] \geq 0$ by analogous derviations as above. Then, by (A.1),

$$
\begin{aligned}
\lim_{T\to\infty}\left[S_T(r)-S_T(r_1^0)\right] &= \lim_{T\to\infty}\left[S_T(r)-S_T(r_2^0)\right]+\lim_{T\to\infty}\left[S_T(r_2^0)-S_T(r_1^0)\right] \\
&> 0,
\end{aligned}
$$

for all $r \in [r_2^0, 1-\epsilon)$.

For $r \in (r_1^0, r_2^0)$, by tedious but analogous derivation,

$$
S_T(r)-S_T(r_1^0)=\int_{\mathbb{R}^d}\left[\frac{r_1^0(r-r_1^0)}{r}\left\|\psi_1(u)-\psi_2(u)\right\|^2+\frac{(r_2^0-r)(1-r_2^0)}{1-r}\left\|\psi_2(u)-\psi_3(u)\right\|^2\right] W(u)\mathrm{d}u
$$

$$
\begin{aligned}
&- \frac{(1-r_2^0)(r_2^0-r_1^0)}{(1-r_1^0)} \int_{\mathbb{R}^d} \|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u + o_P(1) \\
&= \int_{\mathbb{R}^d} \left[\frac{r_1^0(r-r_1^0)}{r} \|\psi_1(u)-\psi_2(u)\|^2 - \frac{(r-r_1^0)(1-r_2^0)^2}{(1-r)(1-r_1^0)} \|\psi_2(u)-\psi_3(u)\|^2 \right] W(u)\mathrm{d}u + o_P(1) \\
&= \frac{(r-r_1^0)r_2^0}{r} \int_{\mathbb{R}^d} \left[\frac{r_1^0}{r_2^0} \|\psi_1(u)-\psi_2(u)\|^2 - \frac{r(1-r_2^0)^2}{r_2^0(1-r)(1-r_1^0)} \|\psi_2(u)-\psi_3(u)\|^2 \right] W(u)\mathrm{d}u + o_P(1) \\
&> \frac{(r-r_1^0)r_2^0}{r} \int_{\mathbb{R}^d} \left[\frac{1-r_2^0}{1-r_1^0} \|\psi_2(u)-\psi_3(u)\|^2 - \frac{r(1-r_2^0)^2}{r_2^0(1-r)(1-r_1^0)} \|\psi_2(u)-\psi_3(u)\|^2 \right] W(u)\mathrm{d}u + o_P(1) \\
&= \frac{(r-r_1^0)r_2^0(1-r_2^0)}{r(1-r_1^0)} \left[1 - \frac{r(1-r_2^0)}{r_2^0(1-r)} \right] \int_{\mathbb{R}^d} \|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u + o_P(1) \\
&= \frac{(r-r_1^0)r_2^0(1-r_2^0)}{r(1-r_1^0)} \left[\frac{r_2^0-r}{r_2^0(1-r)} \right] \int_{\mathbb{R}^d} \|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u + o_P(1) \\
&\xrightarrow{p} \frac{(r-r_1^0)(1-r_2^0)(r_2^0-r)}{r(1-r_1^0)(1-r)} \int_{\mathbb{R}^d} \|\psi_2(u)-\psi_3(u)\|^2 W(u)\mathrm{d}u > 0,
\end{aligned}
$$

where the first inequality is by (A.1) and the last one is by $r \in (r_1^0, r_2^0)$. Therefore, we have shown that $S_T(r)$ has a unique asymptotic global minimum at r_1^0 under (A.1). Besides, given $\hat{r}$ is the global minimizer for $S_T(r)$, we have $S_T(\hat{r}) \leq S_T(r_1)$ for all T. Thus, the consistency of $\hat{r} \to r_1^0$ holds. Analogously, we can show that $\hat{r} \xrightarrow{p} r_2^0$ if we assume r_2^0 is an asymptotic minimizer for $S_T(r)$ relatively to r_1^0. We note that the proof for $M^0 > 2$ and $M > 1$ is virtually quite similar, but is much more tedious. For space, we neglect it. ■

Lemma A.3 *Suppose that the number of the breaks M is bounded from above by a finite integer $M_{\max}$, and Assumptions 1-4 hold.Then as $T \to \infty$, we have*

$$
\max_{M^0 \leq M \leq M_{\max}} \left| \hat{\sigma}^2(M) - \hat{\sigma}^2\left(M^0\right) \right| = O_P\left(T^{-1}\right)
$$

where $\hat{\sigma}^2(M) = T^{-1} SSGR_M\left(\hat{r}_1, \ldots, \hat{r}_M\right)$ and $\{\hat{r}_j\}_{j=1}^M$ is the collection of estimated break fractions.

Proof of Lemma A.3 When $M \geq M^0$, following the proof of Theorem 3.1, we can show that the collection $\{\hat{r}_k\}_{k=1}^M$ contains at least M^0 distinct estimated break fractions, say $\hat{r}_{k_1} < \cdots < \hat{r}_{k_{M^0}}$, such that $\hat{r}_{k_j} - r_j^0 = O_P(T^{-1})$ for $j = 1, 2, ..., M^0$, where $\{k_j\}_{j=1}^{M^0}$ is a subset of $\{k\}_{k=1}^M$.

Let $\hat{\sigma}^2(M) = T^{-1}\mathrm{SSGR}_M(\hat{r}_1, ..., \hat{r}_M)$, then

$$
\begin{aligned}
\hat{\sigma}^2(M) &= \frac{1}{T} \sum_{k=1}^{M+1} \sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i} u' X_t} - \hat{\psi}_k(u) \right\|^2 W(u)\mathrm{d}u \\
&= \sum_{k=1}^{M+1} \hat{\sigma}_k^2(M),
\end{aligned}
$$

where $\hat{\psi}_k(u) = \frac{1}{\hat{T}_k - \hat{T}_{k-1}} \sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k} Y_t e^{\mathbf{i}u'X_t}$ for $t \in [\hat{T}_{k-1}+1, \hat{T}_k]$ and

$$\hat{\sigma}_k^2(M) \equiv \frac{1}{T} \sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \hat{\psi}_k(u) \right\|^2 W(u) \mathrm{d}u. \tag{A.2}$$

Consider the break fractions $\hat{r}_{k_1} < \cdots < \hat{r}_{k_{M^0}}$, such that $\hat{r}_{k_j} - r_j^0 = O_P(T^{-1})$ for $j = 1, 2, ..., M^0$. Note that they divide the sample into $M^0 + 1$ segments, the j-th of which is $[\hat{T}_{k_{j-1}} + 1, \hat{T}_{k_j}]$, for $j = 1, ..., M^0 + 1$. Here, we let $k_0 = 0$, and $k_{M^0+1} = M + 1$. Following the convention, it follows $\hat{T}_{k_0} = \hat{T}_0 = 0$, and $\hat{T}_{k_{M^0+1}} = \hat{T}_{M+1} = T$.

It is obvious that each interval $[\hat{T}_{k_{j-1}} + 1, \hat{T}_{k_j}]$ can contain L_j sub-intervals specified by $\hat{r}_{k_{j-1}} < \hat{r}_{k_{j-1}+1} \cdots < \hat{r}_{k_j - 1} < \hat{r}_{k_j}$, where $L_j \geq 1$. Then, we have

$$\hat{\sigma}^2(M) = \sum_{j=1}^{M^0+1} \sum_{l=1}^{L_j} \hat{\sigma}_{k_{j-1}+l}^2(M),$$

such that $k_{j-1} + L_j = k_j$ for $L_j \geq 1$ and $j = 1, .., M^0 + 1$.

Now, consider the SSGR under M^0 breaks.

$$\begin{aligned} \hat{\sigma}^2(M^0) &= \frac{1}{T}\mathrm{SSGR}(M^0) \\ &= \sum_{l=1}^{M^0+1} \frac{1}{T} \sum_{t=\hat{T}_{l-1}+1}^{\hat{T}_l} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \hat{\psi}_l(u) \right\|^2 W(u) \mathrm{d}u, \end{aligned}$$

where $\{\hat{T}_l\}_{l=1}^{M^0}$ is the collection of estimated breaks and $\hat{\psi}_l(u)$ is the feasible empirical CF given $\{\hat{T}_l\}_{l=1}^{M^0}$. Denote

$$\hat{\varsigma}_l^2(M_0) = \frac{1}{T} \sum_{t=\hat{T}_{l-1}+1}^{\hat{T}_l} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \hat{\psi}_l(u) \right\|^2 W(u) \mathrm{d}u. \tag{A.3}$$

Then

$$\begin{aligned} \hat{\sigma}^2(M_0) - \hat{\sigma}^2(M) &= \sum_{j=1}^{M^0+1} \left[\hat{\varsigma}_j^2(M_0) - \sum_{l=1}^{L_j} \hat{\sigma}_{k_{j-1}+l}^2(M) \right] \\ &= \sum_{j=1}^{M^0+1} \left[\tilde{\varsigma}_j^2(M_0) - \sum_{l=1}^{L_j} \tilde{\sigma}_{k_{j-1}+l}^2(M) \right] + O_P(T^{-1}), \end{aligned}$$

where $\tilde{\sigma}_{k_{j-1}+l}^2(M)$ and $\tilde{\varsigma}_j^2(M_0)$ are defined as (A.2) and (A.3) with each $\hat{T}_{k_j}$ and $\hat{T}_j$ replaced by the corresponding true break date T_j^0 for $j = 1, ..., M^0$. The last equality holds since $\hat{r}_{k_j} - r_j^0 = O_P(T^{-1})$

for the estimated break fractions $\{\hat{r}_{k_j}\}_{j=1}^{M_0}$ under M breaks and $\hat{r}_j - r_j^0 = O_P(T^{-1})$ for the estimated break fractions $\{\hat{r}_j\}_{j=1}^{M^0}$ under M^0 breaks.

Now, it remains to show

$$\tilde{\varsigma}_j^2(M_0) - \sum_{l=1}^{L_j} \tilde{\sigma}_{k_{j-1}+l}^2(M) = O_P(T^{-1}).$$

Consider the j-th segment specified by the true break dates $[T_{j-1}^0 + 1, T_j^0]$, then

$$\tilde{\varsigma}_j^2(M_0) = \frac{1}{T} \sum_{t=T_{j-1}^0+1}^{T_j^0} \int_{\mathbb{R}^d} \left\| Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}_j^0(u) \right\|^2 W(u)\mathrm{d}u,$$

Given

$$Y_t e^{\mathbf{i}u'X_t} = \psi_j(u) + \eta_t(u),$$

for $t \in [T_{j-1}^0 + 1, T_j^0]$. Obviously, $T\tilde{\varsigma}_j^2(M_0)$ is equivalent to the SSGR for a sample with no breaks, and $T\sum_{l=1}^{L_j} \tilde{\sigma}_{k_{j-1}+l}^2$ is equivalent to the SSGR by setting the number of breaks at L^j for that sample. Then, it implies that $T\tilde{\varsigma}_j^2(M_0) - T\sum_{l=1}^{L_j} \tilde{\sigma}_{k_{j-1}+l}^2$ is equivalent to the sup-F test in Theorem 4.2. Thus,

$$\tilde{\varsigma}_j^2(M_0) - \sum_{l=1}^{L_j} \tilde{\sigma}_{k_{j-1}+l}^2 = O_P(T^{-1}),$$

for all $j = 1, ..., M^0$. Therefore,

$$\hat{\sigma}^2(M_0) - \hat{\sigma}^2(M) = O_P(T^{-1}),$$

for all $M^0 \leq M \leq M_{\max}$, where $M_{\max}$ is a finite integer. ■

B Proof of Main Results

Proof of Theorem 3.1. We first show (i). Since $\{\hat{r}_1, \cdots, \hat{r}_{M^0}\}$ minimize the objective function in (3.2) for any partition $\{r_j\}_{j=1}^{M^0}$, we have

$$\frac{1}{T}\sum_{t=1}^{T} \int_{\mathbb{R}^d} \|\hat{\eta}_t(u)\|^2 W(u)\mathrm{d}u \leq \frac{1}{T}\sum_{t=1}^{T} \int_{\mathbb{R}^d} \|\tilde{\eta}_t(u)\|^2 W(u)\mathrm{d}u, \tag{B.1}$$

By the definition of $\eta_t(u)$ and $\tilde{\eta}_t(u)$, it follows

$$\begin{aligned}\tilde{\eta}_t(u) &= Y_t e^{\mathrm{i}u'X_t} - \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} Y_t e^{\mathrm{i}u'X_t} \\ &= \eta_t(u) - \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u)\end{aligned}$$

for $t \in [T_{j-1}^0 + 1, T_j^0]$. Hence,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d} \|\tilde{\eta}_t(u)\|^2 W(u)du &= \frac{1}{T}\sum_{j=1}^{M^0+1}\sum_{t=T_{j-1}^0+1}^{T_j^0}\int_{\mathbb{R}^d} \|\tilde{\eta}_t(u)\|^2 W(u)du \\ &= \frac{1}{T}\sum_{j=1}^{M^0+1}\sum_{t=T_{j-1}^0+1}^{T_j^0}\int_{\mathbb{R}^d} \left\{ \|\eta_t(u)\|^2 + \left\| \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) \right\|^2 \right. \\ &\quad \left. - \frac{2}{T_j^0 - T_{j-1}^0} \sum_{s=T_{j-1}^0+1}^{T_j^0} \mathrm{Re}[\eta_t(u)\eta_s(u)^*] \right\} W(u)du \\ &= \frac{1}{T}\sum_{j=1}^{M^0+1}\sum_{t=T_{j-1}^0+1}^{T_j^0}\int_{\mathbb{R}^d} \|\eta_t(u)\|^2 W(u)du \\ &\quad - \sum_{j=1}^{M^0+1} (r_j^0 - r_{j-1}^0) \int_{\mathbb{R}^d} \left\| \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) \right\|^2 W(u)\mathrm{d}u.\end{aligned}$$

Given Lemma A.1,

$$\int_{\mathbb{R}^d} \left\| \frac{1}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) \right\|^2 W(u)du = O_P(T^{-1}), \tag{B.2}$$

for each $j = 1, ..., M^0 + 1$. This in conjunction with (B.1) implies

$$\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T} \|\hat{\eta}_t(u)\|^2 W(u)\mathrm{d}u \leq \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T} \|\eta_t(u)\|^2 W(u)\mathrm{d}u + o_P(1). \tag{B.3}$$

Let

$$d_t(u) = \hat{\psi}_k(u) - \psi_j(u)$$

for $t \in [\hat{T}_{k-1} + 1, \hat{T}_k] \cap [T_{j-1}^0 + 1, T_j^0]$ with $k, j = 1, 2, ..., M^0 + 1$. Then $\hat{\eta}_t(u) = \eta_t(u) - d_t(u)$, we have

$$\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\hat{\eta}_t(u)\|^2 W(u)\mathrm{d}u = \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\eta_t(u)\|^2 W(u)\mathrm{d}u + \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|d_t(u)\|^2 W(u)\mathrm{d}u - \frac{2}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\mathrm{Re}\left[\eta_t(u)d_t(u)^*\right] W(u)\mathrm{d}u.$$

To proceed, we make the following claims:

$$\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\mathrm{Re}\left[\eta_t(u)d_t(u)^*\right] W(u)\mathrm{d}u = O_P(T^{-1/2}), \tag{B.4}$$

and if $\hat{r}_j \stackrel{p}{\nrightarrow} r_j^0$ for some $j = 1, 2, ..., M^0$, then for any $0 < c_0 < 1$, there exists a $\delta > 0$, such that

$$P\left(\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|d_t(u)\|^2 W(u)\mathrm{d}u > \delta\int_{\mathbb{R}^d}\|\psi_{j+1}(u)-\psi_j(u)\|^2 W(u)\mathrm{d}u\right) > c_0. \tag{B.5}$$

We first show (B.4).

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)d_t(u)^*\right] W(u)\mathrm{d}u &= \frac{1}{T}\sum_{k=1}^{M^0+1}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\hat{\psi}_k(u)^*\right] W(u)\mathrm{d}u \\ &\quad -\frac{1}{T}\sum_{j=1}^{M^0+1}\sum_{t=T_{j-1}^0+1}^{T_j^0}\int_{\mathbb{R}^d}\mathrm{Re}\left[\eta_t(u)\psi_j(u)^*\right] W(u)\mathrm{d}u \\ &\equiv Q_1 - Q_2.\end{aligned}$$

Consider Q_1. By the definition of $\hat{\psi}_k(u)$,

$$\begin{aligned} Q_1 &= \frac{1}{T}\sum_{k=1}^{M^0+1}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\int_{\mathbb{R}^d}\mathrm{Re}\left\{\eta_t(u)\left[\frac{1}{\hat{T}_k-\hat{T}_{k-1}}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}(\phi_t(u)+\eta_t(u))\right]^*\right\} W(u)\mathrm{d}u \\ &= \frac{1}{T}\sum_{k=1}^{M^0+1}\int_{\mathbb{R}^d}\mathrm{Re}\left\{\left[\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\eta_t(u)\right]\left[\frac{1}{\hat{T}_k-\hat{T}_{k-1}}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\phi_t(u)\right]^*\right\} W(u)\mathrm{d}u \\ &\quad +\frac{1}{T}\sum_{k=1}^{M^0+1}\frac{1}{\hat{T}_k-\hat{T}_{k-1}}\int_{\mathbb{R}^d}\left\|\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\eta_t(u)\right\|^2 W(u)\mathrm{d}u \\ &\equiv Q_{11} + Q_{12},\end{aligned}$$

where $\phi_t(u) \equiv E(Y_t e^{\mathbf{i}u'X_t})$. By Cauchy-Schwarz inequality,

$$\begin{aligned}
|Q_{11}| &\leq \frac{1}{T}\sum_{k=1}^{M^0+1}\left[\int_{\mathbb{R}^d}\left\|\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\eta_t(u)\right\|^2 W(u)\mathrm{d}u\right]^{1/2}\left[\int_{\mathbb{R}^d}\left\|\frac{1}{\hat{T}_k-\hat{T}_{k-1}}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\phi_t(u)\right\|^2 W(u)\mathrm{d}u\right]^{1/2} \\
&\leq \frac{1}{\sqrt{T}}\sum_{k=1}^{M^0+1}\left[\frac{1}{T}\int_{\mathbb{R}^d}\left\|\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\eta_t(u)\right\|^2 W(u)\mathrm{d}u\right]^{1/2}\left[\int_{\mathbb{R}^d}\left[\frac{1}{\hat{T}_k-\hat{T}_{k-1}}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\|\phi_t(u)\|\right]^2 W(u)\mathrm{d}u\right]^{1/2} \\
&= O_P(T^{-1/2}),
\end{aligned}$$

where the last equality holds by Lemma A.1 and the fact that

$$\frac{1}{\hat{T}_k-\hat{T}_{k-1}}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}\|\phi_t(u)\| \leq \frac{1}{\hat{T}_k-\hat{T}_{k-1}}\sum_{t=\hat{T}_{k-1}+1}^{\hat{T}_k}E\left\|Y_t e^{\mathbf{i}u'X_t}\right\| \leq \mathcal{C}.$$

Analogously, $Q_{12} = O_P(T^{-1})$ by Lemma A.1. Hence, $Q_1 = O_P(T^{-1/2})$. Consider Q_2. By Cauchy-Schwarz inequality,

$$\begin{aligned}
Q_2 &= \frac{1}{T}\sum_{j=1}^{M^0+1}\int_{\mathbb{R}^d}\mathrm{Re}\left[\sum_{t=T_{j-1}^0+1}^{T_j^0}\eta_t(u)\psi_j(u)^*\right]W(u)\mathrm{d}u \\
&\leq \frac{1}{\sqrt{T}}\sum_{j=1}^{M^0+1}\left(\frac{1}{T}\int_{\mathbb{R}^d}\left\|\sum_{t=T_{j-1}^0+1}^{T_j^0}\eta_t(u)\right\|^2 W(u)\mathrm{d}u\right)^{1/2}\left(\int_{\mathbb{R}^d}\|\psi_j(u)\|^2 W(u)du\right)^{1/2} \\
&= O_P(T^{-1/2}),
\end{aligned}$$

by Lemma A.1 and Assumptions 1(iii), 1(iv), and 3(i). Hence, we have established (B.4).

Now, we show (B.5). If there exists a break, say r_j^0, which is not consistently estimated, then for some positive probability $0 < c_0 < 1$, there exists a $\kappa > 0$ such that no estimated break dates fall into the interval $\left[T_j^0 - \lfloor\kappa T\rfloor + 1, T_j^0 + \lfloor\kappa T\rfloor\right]$ for a subsequence of T. Suppose this interval is classified into the k-th regime, i.e., $\hat{T}_{k-1} \leq T_j^0 - \lfloor\kappa T\rfloor + 1 < T_j^0 + \lfloor\kappa T\rfloor \leq \hat{T}_k$, then

$$d_t(u) = \begin{cases} \hat{\psi}_k(u) - \psi_j(u), & \text{for } t \in \left[T_j^0 - \lfloor\kappa T\rfloor + 1, T_j^0\right]; \\ \hat{\psi}_k(u) - \psi_{j+1}(u), & \text{for } t \in \left[T_j^0 + 1, T_j^0 + \lfloor\kappa T\rfloor\right]. \end{cases}$$

By the definition of $d_t(u)$,

$$\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|d_t(u)\|^2 W(u)\mathrm{d}u$$

$$
\begin{aligned}
&\geq \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=T_j^0-\lfloor\kappa T\rfloor+1}^{T_j^0+\lfloor\kappa T\rfloor}\|d_t(u)\|^2 W(u)\mathrm{d}u \\
&= \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=T_j^0-\lfloor\kappa T\rfloor+1}^{T_j^0}\|d_t(u)\|^2 W(u)\mathrm{d}u + \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=T_j^0+1}^{T_j^0+\lfloor\kappa T\rfloor}\|d_t(u)\|^2 W(u)\mathrm{d}u \\
&= \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=T_j^0-\lfloor\kappa T\rfloor+1}^{T_j^0}\left\|\hat{\psi}_k(u)-\psi_j(u)\right\|^2 W(u)\mathrm{d}u + \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=T_j^0+1}^{T_j^0+\lfloor\kappa T\rfloor}\left\|\hat{\psi}_k(u)-\psi_{j+1}(u)\right\|^2 W(u)\mathrm{d}u \\
&= \kappa\int_{\mathbb{R}^d}\left[\left\|\hat{\psi}_k(u)-\psi_j(u)\right\|^2+\left\|\hat{\psi}_k(u)-\psi_{j+1}(u)\right\|^2\right]W(u)\mathrm{d}u \\
&\geq \frac{1}{2}\kappa\int_{\mathbb{R}^d}\|\psi_{j+1}(u)-\psi_j(u)\|^2 W(u)\mathrm{d}u,
\end{aligned}
$$

by triangle inequality. Define $\delta = \frac{1}{2}\kappa$, we then get the desired result.

Based on (B.4) and (B.5), we know that if some break fraction r_j^0 is not consistently estimated, then the following inequality

$$
\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\hat{\eta}_t(u)\|^2 W(u)\mathrm{d}u \geq \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\eta_t(u)\|^2 W(u)\mathrm{d}u + \delta\int_{\mathbb{R}^d}\|\psi_{j+1}(u)-\psi_j(u)\|^2 W(u)\mathrm{d}u + o_P(1)
$$

holds with probability no less that some $0 < c_0 < 1$. This is contradictory to the inequality in (B.3), which holds with probability 1 for all T. Hence, all break fractions are consistently estimated.

For (ii), the proof is quite similar to the proof of Proposition 2 in Bai and Perron (1998). Hence we omit it. ■

Proof of Lemma 4.1. To establish the desired results in (i), we first show that for each fixed $u \in \mathbb{U}$,

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\eta_t(u) \Rightarrow B(u,r),
$$

and then show that $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\eta_t(u)$ is asymptotically tight in $\mathbb{U}$.

Notice that for $u \in \mathbb{R}^d$, $E[\eta_t(u)] = 0$, and

$$
E\|\eta_t(u)\|^q \;=\; E\left\|Y_t e^{\mathbf{i}u'X_t} - E(Y_t e^{\mathbf{i}u'X_t})\right\|^q \leq \mathcal{C},
$$

under Assumptions 1(iii), 1(iv), and 3(i) and the fact $\|e^{\mathbf{i}u'X_t}\|^q$ is bounded. Consider

$$
\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[\eta_t(u)\eta_s(u)^*] \;=\; \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{cov}(Y_t e^{\mathbf{i}u'X_t}, Y_s e^{-\mathbf{i}u'X_s})
$$

$$\to \quad \Omega(u,u),$$

where $\Omega(u,u)=\sum_{j=-\infty}^{\infty}\text{cov}(Y_t e^{\mathbf{i}u'X_t}, Y_{t-j}e^{-\mathbf{i}u'X_{t-j}})$ that exists under Assumptions 1 and 3(i). By Lemma 2.2 of Phillips (1984),

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\eta_t(u)\Rightarrow B(u,r),$$

where $B(u,r)$ is a Brownian motion with covariance kernel $E[B(u,r)B(u,s)^*]=\min\{r,s\}\Omega(u,u)$ for each fixed u. Next, we show asymptotic tightness. For $u_1,u_2\in\mathbb{U}$, we apply mean value theorem to obtain

$$\sum_{t=1}^{\lfloor Tr\rfloor}[\eta_t(u_1)-\eta_t(u_2)] \quad = \quad \sum_{t=1}^{\lfloor Tr\rfloor}\Upsilon_t(\bar{u})'(u_1-u_2),$$

where $\bar{u}$ lies between u_1 and u_2, and

$$\Upsilon_t(u)=\frac{\mathrm{d}\eta_t(u)}{\mathrm{d}u} \quad = \quad \mathbf{i}\left[X_tY_te^{\mathbf{i}u'X_t}-E(X_tY_te^{\mathbf{i}u'X_t})\right].$$

Then,

$$\begin{aligned} E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}[\eta_t(u_1)-\eta_t(u_2)]\right\|^2 &= E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\Upsilon_t(\bar{u})'(u_1-u_2)\right\|^2 \\ &\leq \frac{1}{T}\sum_{t=1}^{\lfloor Tr\rfloor}\sum_{s=1}^{\lfloor Tr\rfloor}\left\|\text{cov}\left(X_tY_te^{\mathbf{i}\bar{u}'X_t},X_sY_se^{\mathbf{i}\bar{u}'X_s}\right)\right\|\|u_1-u_2\|^2 \\ &\leq \mathcal{C}\|u_1-u_2\|^2, \end{aligned}$$

by the mixing condition in Assumption 1(i) and the moment conditions in Assumptions 1(ii), 1(iii), and 3(i). Hence, $\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\eta_t(u)$ is asymptotically tight on $\mathbb{U}$. By Theorem 15.6 of Billingsley (1968),

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor Tr\rfloor}\eta_t(u)\Rightarrow B(u,r),$$

as $T\to\infty$, where $B(u,r)$ is a complex-valued Gaussian defined on $\mathbb{U}\times[0,1]$ with mean 0 and covariance kernel $E[B(u,r)B(v,s)^*]=\min\{r,s\}\Omega(u,v)$, with a long-run variance

$$\Omega(u,v)=\sum_{l=-\infty}^{\infty}\text{cov}(Y_te^{\mathbf{i}u'X_t},Y_{t-l}e^{-\mathbf{i}v'X_{t-l}}).$$

For (ii), notice that under $\mathbb{H}_A$, the observations within each time segment $[T_{j-1}^0, T_j^0]$ are weakly stationary. Following analogous arguments for showing (i), (ii) can be established for each $j = 1, ..., M^0 + 1$. ■

Proof of Theorem 4.2. Under $\mathbb{H}_0 : \phi_t(u) = \psi(u)$, it follows

$$Y_t e^{\mathbf{i}u'X_t} = \psi(u) + \eta_t(u),$$

for all $t = 1, ..., T$. We denote

$$\tilde{\psi}^{(R)}(u) = \frac{1}{T}\sum_{t=1}^{T} Y_t e^{\mathbf{i}u'X_t} = \psi(u) + \frac{1}{T}\sum_{t=1}^{T}\eta_t(u),$$

and

$$\tilde{\psi}_j^{(U)}(u) = \frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j} Y_t e^{\mathbf{i}u'X_t} = \psi(u) + \frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\eta_t(u)$$

for $j = 1, ..., M + 1$, where $\{T_j\}_{j=1}^M$ is the collection of break dates under $\mathbb{H}_A$. Let

$$F_T(r_1, ..., r_M) = \text{SSGR}_0 - \text{SSGR}_M(r_1, ..., r_M),$$

then

$$\begin{aligned}
& F_T(r_1, ..., r_M) \\
= \; & \sum_{j=1}^{M+1}\sum_{t=T_{j-1}+1}^{T_j}\int_{\mathbb{R}^d}\left[\left\|Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}^{(R)}(u)\right\|^2 - \left\|Y_t e^{\mathbf{i}u'X_t} - \tilde{\psi}_j^{(U)}(u)\right\|^2\right]W(u)\mathrm{d}u \\
= \; & \sum_{j=1}^{M+1}\sum_{t=T_{j-1}+1}^{T_j}\int_{\mathbb{R}^d}\left[\left\|\tilde{\psi}^{(R)}(u)\right\|^2 - \left\|\tilde{\psi}_j^{(U)}(u)\right\|^2 - 2\mathrm{Re}\left\{Y_t e^{\mathbf{i}u'X_t}\tilde{\psi}^{(R)}(u)^*\right\} + 2\mathrm{Re}\left\{Y_t e^{\mathbf{i}u'X_t}\tilde{\psi}_j^{(U)}(u)^*\right\}\right]W(u)\mathrm{d}u \\
= \; & \sum_{j=1}^{M+1}(T_j - T_{j-1})\int_{\mathbb{R}^d}\left\{\left\|\tilde{\psi}^{(R)}(u)\right\|^2 + \left\|\tilde{\psi}_j^{(U)}(u)\right\|^2 - 2\mathrm{Re}\left[\tilde{\psi}_j^{(U)}(u)\tilde{\psi}^{(R)}(u)^*\right]\right\}W(u)\mathrm{d}u \\
= \; & \sum_{j=1}^{M+1}(T_j - T_{j-1})\int_{\mathbb{R}^d}\left\|\tilde{\psi}_j^{(U)}(u) - \tilde{\psi}^{(R)}(u)\right\|^2 W(u)\mathrm{d}u.
\end{aligned}$$

Under $\mathbb{H}_0$, we have

$$F_T(r_1, ..., r_M) = \sum_{j=1}^{M+1}(T_j - T_{j-1})\int_{\mathbb{R}^d}\left\|\frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\eta_t(u) - \frac{1}{T}\sum_{t=1}^{T}\eta_t(u)\right\|^2 W(u)\mathrm{d}u$$

$$= \sum_{j=1}^{M+1}(r_j - r_{j-1})\int_{\mathbb{R}^d}\left\|\frac{1}{r_j - r_{j-1}}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_j}\eta_t(u) - \frac{1}{\sqrt{T}}\sum_{t=1}^{T_{j-1}}\eta_t(u)\right] - \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t(u)\right\|^2 W(u)\mathrm{d}u$$

$$= \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{R}^d}\|[\mathcal{L}_T(u,r_j) - \mathcal{L}_T(u,r_{j-1})] - (r_j - r_{j-1})\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u,$$

where we let $\mathcal{L}_T(u,r_j) \equiv T^{-1/2}\sum_{t=1}^{T_j}\eta_t(u)$. Define the following

$$\mathcal{A}_{1,T}(r_1,...,r_M) = \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{U}}\|[\mathcal{L}_T(u,r_j) - \mathcal{L}_T(u,r_{j-1})] - (r_j - r_{j-1})\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u,$$

$$\mathcal{A}_{2,T}(r_1,...,r_M) = \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{U}^c}\|[\mathcal{L}_T(u,r_j) - \mathcal{L}_T(u,r_{j-1})] - (r_j - r_{j-1})\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u,$$

$$\mathcal{A}_1(r_1,...,r_M) = \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{U}}\|[B(u,r_j) - B(u,r_{j-1})] - (r_j - r_{j-1})B(u,1)\|^2 W(u)\mathrm{d}u,$$

$$\mathcal{A}_2(r_1,...,r_M) = \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{U}^c}\|[B(u,r_j) - B(u,r_{j-1})] - (r_j - r_{j-1})B(u,1)\|^2 W(u)\mathrm{d}u,$$

where $\mathbb{U}$ is any compact subset of $\mathbb{R}^d$ and $\mathbb{U}^c$ is its complement set. Obviously, $F_T(r_1,...,r_M) = \mathcal{A}_{1,T}(r_1,...,r_M)+\mathcal{A}_{2,T}(r_1,...,r_M)$. Furthermore, we denote $F(r_1,...,r_M) \equiv \mathcal{A}_1(r_1,...,r_M)+\mathcal{A}_2(r_1,...,r_M)$. Now, we show that $F_T \xrightarrow{d} \sup_{\{r_1,...,r_M\}} F(r_1,...,r_M)$.

Note that for any fixed constant $\iota > 0$, and fixed collection of breaks $\{r_1,...,r_M\} \in \Pi_\epsilon$, there exists a compact subset $\mathbb{U}$ that depends on ι, such that

$$\begin{aligned}
& E\left[\mathcal{A}_{2,T}(r_1,...,r_M)\right] \\
= & \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{U}^c}E\|[\mathcal{L}_T(u,r_j) - \mathcal{L}_T(u,r_{j-1})] - (r_j - r_{j-1})\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u \\
\leq & \sum_{j=1}^{M+1}\frac{1}{(r_j - r_{j-1})}\int_{\mathbb{U}^c}2E\|\mathcal{L}_T(u,r_j) - \mathcal{L}_T(u,r_{j-1})\|^2 + 2E\|(r_j - r_{j-1})\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u \\
= & 2\sum_{j=1}^{M+1}\left[\int_{\mathbb{U}^c}E\left\|\frac{1}{\sqrt{T(r_j - r_{j-1})}}\sum_{t=1}^{\lfloor T(r_j - r_{j-1})\rfloor}\eta_t(u)\right\|^2 + (r_j - r_{j-1})E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t(u)\right\|^2\right]W(u)\mathrm{d}u \\
\leq & 2(M+1)\mathcal{C}\int_{\mathbb{U}^c}W(u)\mathrm{d}u < \iota,
\end{aligned}$$

where the last equality is by stationarity under $\mathbb{H}_0$, and the second to last inequality is by Lemma A.1, the fact that $0 < r_j - r_{j-1} < 1$ for all j, and the boundedness of $\eta_t(u)$. Analogous result holds for $E[\mathcal{A}_2(r_1,...,r_M)]$. Under Lemma 4.1, $\mathcal{L}(u,r_j) \Rightarrow B(u,r_j)$ on $\mathbb{U}\times[0,1]$ for any compact subset $\mathbb{U}$ of $\mathbb{R}^d$. Thus, by continuous mapping theorem, $\mathcal{A}_{1,T}(r_1,...,r_M) \Rightarrow \mathcal{A}_1(r_1,...,r_M)$.

Hence, for each fixed $\{r_1,...,r_M\}$,

$$F_T(r_1,...,r_M) \xrightarrow{d} F(r_1,...,r_M).$$

Now, we show tightness. Without loss of generality, we consider $M=1$. Let $0<r_1<s_1<1$, and $q>2$, it follows

$$\begin{aligned}
& E\left|F_T(s_1)-F_T(r_1)\right|^q \\
= \; & E\left|\frac{1}{s_1(1-s_1)}\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)-s_1\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u - \frac{1}{r_1(1-r_1)}\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,r_1)-r_1\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right|^q \\
\leq \; & \frac{2^{q-1}}{s_1^q(1-s_1)^q}E\left|\int_{\mathbb{R}^d}\left\{\|\mathcal{L}_T(u,s_1)-s_1\mathcal{L}_T(u,1)\|^2-\|\mathcal{L}_T(u,r_1)-r_1\mathcal{L}_T(u,1)\|^2\right\}W(u)\mathrm{d}u\right|^q \\
& +2^{q-1}\left[\frac{(s_1-r_1)(s_1+r_1-1)}{s_1r_1(1-s_1)(1-r_1)}\right]^q E\left|\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,r_1)-r_1\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right|^q \\
= \; & \mathcal{I}_1(s_1,r_1)+\mathcal{I}_2(s_1,r_1), \text{ say.}
\end{aligned}$$

Consider $\mathcal{I}_1(s_1,r_1)$. By Cauchy-Schwarz inequality,

$$\begin{aligned}
& \left|\int_{\mathbb{R}^d}\left\{\|\mathcal{L}_T(u,s_1)-s_1\mathcal{L}_T(u,1)\|^2-\|\mathcal{L}_T(u,r_1)-r_1\mathcal{L}_T(u,1)\|^2\right\}W(u)\mathrm{d}u\right| \\
= \; & \left|\int_{\mathbb{R}^d}\mathrm{Re}\left\{\left[\mathcal{L}_T(u,s_1)-\mathcal{L}_T(u,r_1)-(s_1-r_1)\mathcal{L}_T(u,1)\right]\left[\mathcal{L}_T(u,s_1)+\mathcal{L}_T(u,r_1)-(s_1+r_1)\mathcal{L}_T(u,1)\right]^*\right\}W(u)\mathrm{d}u\right| \\
\leq \; & \left[\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)-\mathcal{L}_T(u,r_1)-(s_1-r_1)\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right]^{1/2} \\
& \times\left[\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)+\mathcal{L}_T(u,r_1)-(s_1+r_1)\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right]^{1/2}.
\end{aligned}$$

Then, it follows

$$\begin{aligned}
\mathcal{I}_1(s_1,r_1) \leq \; & \frac{2^{q-1}}{s_1^q(1-s_1)^q}E\left\{\left[\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)-\mathcal{L}_T(u,r_1)-(s_1-r_1)\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right]^{q/2}\right. \\
& \left.\times\left[\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)+\mathcal{L}_T(u,r_1)-(s_1+r_1)\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right]^{q/2}\right\} \\
\leq \; & \frac{2^{q-1}}{s_1^q(1-s_1)^q}\left(E\left[\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)-\mathcal{L}_T(u,r_1)-(s_1-r_1)\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right]^q\right)^{1/2} \\
& \times\left(E\left[\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)+\mathcal{L}_T(u,r_1)-(s_1+r_1)\mathcal{L}_T(u,1)\|^2 W(u)\mathrm{d}u\right]^q\right)^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{2^{q-1}}{s_1^q(1-s_1)^q}\left(E\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)-\mathcal{L}_T(u,r_1)-(s_1-r_1)\mathcal{L}_T(u,1)\|^{2q}W(u)\mathrm{d}u\right)^{1/2}\\
&\quad\times\left(E\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)+\mathcal{L}_T(u,r_1)-(s_1+r_1)\mathcal{L}_T(u,1)\|^{2q}W(u)\mathrm{d}u\right)^{1/2}\\
&= \frac{2^{q-1}}{s_1^q(1-s_1)^q}\mathcal{I}_{11}(s_1,r_1)\mathcal{I}_{12}(s_1,r_1),
\end{aligned}
$$

where the second inequality is by Cauchy-Schwarz inequality, and the last inequality is by Jensen's inequality. By stationarity,

$$
\begin{aligned}
& E\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,s_1)-\mathcal{L}_T(u,r_1)-(s_1-r_1)\mathcal{L}_T(u,1)\|^{2q}W(u)\mathrm{d}u\\
= {} & T^{-q}\int_{\mathbb{R}^d}E\left\|\sum_{t=1}^{\lfloor T(s_1-r_1)\rfloor}\eta_t(u)-(s_1-r_1)\sum_{t=1}^{T}\eta_t(u)\right\|^{2q}W(u)\mathrm{d}u\\
\leq {} & 2^{2q-1}T^{-q}\left[\int_{\mathbb{R}^d}E\left\|\sum_{t=1}^{\lfloor T(s_1-r_1)\rfloor}\eta_t(u)\right\|^{2q}W(u)\mathrm{d}u+\int_{\mathbb{R}^d}E\left\|(s_1-r_1)\sum_{t=1}^{T}\eta_t(u)\right\|^{2q}W(u)\mathrm{d}u\right]\\
\leq {} & 2^{2q-1}T^{-q}\mathcal{C}\left[T(s_1-r_1)\right]^q+2^{2q-1}T^{-q}\mathcal{C}(s_1-r_1)^{2q}T^q\\
\leq {} & \mathcal{C}(s_1-r_1)^q,
\end{aligned}
$$

where the second to last inequality is by Theorem 2 of Yokoyama (1980) and Assumption 2. Hence, $\mathcal{I}_{11}(s_1,r_1)\leq\mathcal{C}(s_1-r_1)^{q/2}$. By analogous arguments, given $0<r_1<s_1<1$, we have $\sup_{s_1,r_1}|\mathcal{I}_{12}(s_1,r_1)|\leq\mathcal{C}$. Finally, for $\mathcal{I}_2(s_1,r_1)$, we have

$$
\begin{aligned}
& E\left|\int_{\mathbb{R}^d}\|\mathcal{L}_T(u,r_1)-r_1\mathcal{L}_T(u,1)\|^2W(u)\mathrm{d}u\right|^q\\
\leq {} & \int_{\mathbb{R}^d}E\|\mathcal{L}_T(u,r_1)-r_1\mathcal{L}_T(u,1)\|^{2q}W(u)\mathrm{d}u\\
\leq {} & 2^{2q-1}\int_{\mathbb{R}^d}E\|\mathcal{L}_T(u,r_1)\|^{2q}W(u)\mathrm{d}u+2^{2q-1}r_1^{2q}\int_{\mathbb{R}^d}E\|\mathcal{L}_T(u,1)\|^{2q}W(u)\mathrm{d}u\\
\leq {} & 2^{2q-1}T^{-q}\mathcal{C}(Tr_1)^q+2^{2q-1}r_1^{2q}T^{-q}\mathcal{C}T^q\\
\leq {} & \mathcal{C},
\end{aligned}
$$

given $0<r_1<1$. Besides, given $0<r_1<s_1<1$,

$$
\left[\frac{(s_1-r_1)(s_1+r_1-1)}{s_1r_1(1-s_1)(1-r_1)}\right]^q\leq\mathcal{C}(s_1-r_1)^q.
$$

Combining the results, we have

$$
E\left|F_T(s_1)-F_T(r_1)\right|^q\leq\mathcal{C}(s_1-r_1)^{q/2}.
$$

By Theorem 15.6 of Billingsley (1968),

$$\sup_{r_1} F_T(r_1) \xrightarrow{d} \sup_{r_1} F(r_1),$$

when $M = 1$. When $M > 1$,

$$\sup_{\{r_1,...,r_M\}} F_T(r_1,...,r_M) \xrightarrow{d} \sup_{\{r_1,...,r_M\}} F(r_1,...,r_M),$$

where

$$\begin{aligned} F(r_1,...,r_M) &= \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \|[B(u,r_j) - B(u,r_{j-1})] - (r_j - r_{j-1})B(u,1)\|^2 W(u)\mathrm{d}u \\ &= \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \|\mathcal{B}(u,r_j) - \mathcal{B}(u,r_{j-1})\|^2 W(u)\mathrm{d}u, \end{aligned}$$

where $\mathcal{B}(u,r) \equiv B(u,r) - rB(u,1)$ is a generalized Brownian bridge. ■

Proof of Theorem 4.3. Let $\{T_j\}_{j=1}^M$ be the specified collection of break dates under $\mathbb{H}_A$. By proof of Theorem 4.2,

$$\begin{aligned} F_T(r_1,...,r_M) &= \sum_{j=1}^{M+1} (T_j - T_{j-1}) \int_{\mathbb{R}^d} \left\| \tilde{\psi}_j^{(U)}(u) - \tilde{\psi}^{(R)}(u) \right\|^2 W(u)\mathrm{d}u \\ &= \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \Bigg\| \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} Y_t e^{\mathbf{i}u'X_t} - \frac{r_j}{\sqrt{T}} \sum_{t=1}^{T} Y_t e^{\mathbf{i}u'X_t} \right) \\ &\quad - \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T_{j-1}} Y_t e^{\mathbf{i}u'X_t} - \frac{r_{j-1}}{\sqrt{T}} \sum_{t=1}^{T} Y_t e^{\mathbf{i}u'X_t} \right) \Bigg\|^2 W(u)\mathrm{d}u. \end{aligned}$$

Let $\{T_k^0\}_{k=1}^{M^0}$ denote the the collection of true breaks. Under $\mathbb{H}_A(a_T) : \psi_k(u) = \psi(u) + a_T\Delta_k(u)$ for the k-th regime, $k = 1,..., M^0 + 1$, then

$$Y_t e^{\mathbf{i}u'X_t} = \psi(u) + a_T\Delta_k(u) + \eta_t(u),$$

for $t = T_{k-1}^0 + 1, ..., T_k^0$, and $k = 1, ..., M^0 + 1$. Given $a_T = T^{-1/2}$, it follows

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} Y_t e^{\mathbf{i}u'X_t} &= \frac{T_j}{\sqrt{T}}\psi(u) + \frac{1}{\sqrt{T}} \sum_{k=1}^{l} \sum_{t=T_{k-1}^0+1}^{T_k^0} [a_T\Delta_k(u) + \eta_t(u)] + \frac{1}{\sqrt{T}} \sum_{t=T_l^0+1}^{T_j} [a_T\Delta_{l+1}(u) + \eta_t(u)] \\ &= r_j\sqrt{T}\psi(u) + \left[\sum_{k=1}^{l} (r_k^0 - r_{k-1}^0)\Delta_k(u) + (r_j - r_l^0)\Delta_{l+1}(u) \right] \end{aligned}$$

$$+\left[\sum_{k=1}^{l}\frac{\sqrt{r_k^0-r_{k-1}^0}}{\sqrt{T_k^0-T_{k-1}^0}}\sum_{t=T_{k-1}^0+1}^{T_k^0}\eta_t(u)+\frac{\sqrt{r_{l+1}^0-r_l^0}}{\sqrt{T_{l+1}^0-T_l^0}}\sum_{t=T_l^0+1}^{T_j}\eta_t(u)\right],$$

where $r_l^0 < r_j < r_{l+1}^0$, and

$$\frac{r_j}{\sqrt{T}}\sum_{t=1}^{T}Y_t e^{\mathbf{i}u'X_t} = r_j\sqrt{T}\psi(u)+r_j\sum_{k=1}^{M^0+1}(r_k^0-r_{k-1}^0)\Delta_k(u)+r_j\sum_{k=1}^{M^0+1}\frac{\sqrt{r_k^0-r_{k-1}^0}}{\sqrt{T_k^0-T_{k-1}^0}}\sum_{t=T_{k-1}^0+1}^{T_k^0}\eta_t(u).$$

By Lemma 4.1(ii),

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T_j}Y_t e^{\mathbf{i}u'X_t}-\frac{r_j}{\sqrt{T}}\sum_{t=1}^{T}Y_t e^{\mathbf{i}u'X_t}\Rightarrow G(u,r_j)+\Gamma(u,r_j),$$

where

$$\begin{aligned}G(u,r_j) &= \left[\sum_{k=1}^{l}(r_k^0-r_{k-1}^0)^{1/2}B^{(k)}(u,1)+(r_{l+1}^0-r_l^0)^{1/2}B^{(l+1)}\left(u,\frac{r_j-r_l^0}{r_{l+1}^0-r_l^0}\right)\right]\\ &\quad -r_j\left[\sum_{k=1}^{M^0+1}(r_k^0-r_{k-1}^0)^{1/2}B^{(k)}(u,1)\right],\end{aligned}$$

and

$$\Gamma(u,r_j)=\left[\sum_{k=1}^{l}(r_k^0-r_{k-1}^0)\Delta_k(u)+(r_j-r_l^0)\Delta_{l+1}(u)\right]-r_j\left[\sum_{k=1}^{M^0+1}(r_k^0-r_{k-1}^0)\Delta_k(u)\right].$$

Note that $\Gamma(u,r_j)$ is a continuous function of r_j, and $\int_{\mathbb{R}^d}\|\Gamma(u,r_j)\|^2 W(u)\mathrm{d}u<\infty$ for each j. Given Assumptions 1-3, by analogous arguments to proof of Theorem 4.2,

$$\sup_{\{r_1,...,r_M\}\in\Pi_\epsilon}F_T(r_1,...,r_M)\xrightarrow{d}\sup_{\{r_1,...,r_M\}\in\Pi_\epsilon}F^A(r_1,...,r_M),$$

where

$$F^A(r_1,...,r_M) = \sum_{j=1}^{M+1}\frac{1}{r_j-r_{j-1}}\int_{\mathbb{R}^d}\|G(u,r_j)-G(u,r_{j-1})+\Gamma(u,r_j)-\Gamma(u,r_{j-1})\|^2 W(u)\mathrm{d}u.$$

■

Proof of Theorem 4.4 Without loss of generality, we assume $Kl_T = T$. Let the reconstructed sample obtained via the moving block bootstrap be $\{Z_1^\star, Z_2^\star, ..., Z_T^\star\}$, where $Z_t^\star = Z_{t_i}$ for some

$1 \leq t_i \leq T$. Under $\mathbb{H}_0$ of no structural breaks,

$$Y_t^\star e^{\mathbf{i}u'X_t^\star} = \psi(u) + \eta_t^\star(u),$$

where $\psi(u) = E(Y_t e^{\mathbf{i}u'X_t})$ for all t. It implies that resampling Z_t is equivalent to resampling the generalized error $\eta_t(u)$ under $\mathbb{H}_0$. Consider the following sup-F test statistic based on the bootstrap sample

$$\begin{aligned} F_T^\star(r_1, ..., r_M) &= \text{SSGR}_0^\star - \text{SSGR}_M^\star \\ &= \sum_{j=1}^{M+1} (T_j - T_{j-1}) \int_{\mathbb{R}^d} \left\| \frac{1}{T_j - T_{j-1}} \sum_{s=T_{j-1}+1}^{T_j} Y_s^\star e^{\mathbf{i}u'X_s^\star} - \frac{1}{T} \sum_{s=1}^{T} Y_s^\star e^{\mathbf{i}u'X_s^\star} \right\|^2 W(u)\mathrm{d}u \\ &= \sum_{j=1}^{M+1} (T_j - T_{j-1}) \int_{\mathbb{R}^d} \left\| \frac{1}{T_j - T_{j-1}} \sum_{s=T_{j-1}+1}^{T_j} \eta_s^\star(u) - \frac{1}{T} \sum_{s=1}^{T} \eta_s^\star(u) \right\|^2 W(u)\mathrm{d}u, \end{aligned}$$

where

$$\text{SSGR}_0^\star = \sum_{t=1}^{T} \int_{\mathbb{R}^d} \left\| Y_t^\star e^{\mathbf{i}u'X_t^\star} - \frac{1}{T} \sum_{s=1}^{T} Y_s^\star e^{\mathbf{i}u'X_s^\star} \right\|^2 W(u)\mathrm{d}u,$$

and

$$\text{SSGR}_M^\star = \sum_{j=1}^{M+1} \sum_{t=T_{j-1}+1}^{T_j} \int_{\mathbb{R}^d} \left\| Y_t^\star e^{\mathbf{i}u'X_t^\star} - \frac{1}{T_j - T_{j-1}} \sum_{s=T_{j-1}+1}^{T_j} Y_s^\star e^{\mathbf{i}u'X_s^\star} \right\|^2 W(u)\mathrm{d}u.$$

By analogous steps as proof of Theorem 4.2, we have

$$\begin{aligned} F_T^\star(r_1, ..., r_M) = \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \Bigg\| & \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \eta_t^\star(u) - \frac{r_j}{\sqrt{T}} \sum_{t=1}^{T} \eta_t^\star(u) \right] \\ & - \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T_{j-1}} \eta_t^\star(u) - \frac{r_{j-1}}{\sqrt{T}} \sum_{t=1}^{T} \eta_t^\star(u) \right] \Bigg\|^2 W(u)\mathrm{d}u. \end{aligned}$$

Let $E^\star[\bar{\eta}^\star(u)] \equiv E^\star[T^{-1} \sum_{t=1}^{T} \eta_t^\star(u)]$ be the expectation of the sample average of $\eta_t^\star(u)$ conditioning on the observable sample, we have

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \eta_t^\star(u) - \frac{r_j}{\sqrt{T}} \sum_{t=1}^{T} \eta_t^\star(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} [\eta_t^\star(u) - E^\star[\bar{\eta}^\star(u)]] - \frac{r_j}{\sqrt{T}} \sum_{t=1}^{T} [\eta_t^\star(u) - E^\star[\bar{\eta}^\star(u)]] \\ &= \mathcal{L}_T^\star(u, r_j) - r_j \mathcal{L}_T^\star(u, 1), \end{aligned}$$

where

$$\mathcal{L}_T^\star(u, r_j) \equiv \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \left[\eta_t^\star(u) - E^\star[\bar{\eta}^\star(u)]\right].$$

Then, it follows

$$F_T^\star(r_1, ..., r_M) \quad = \quad \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \left\| [\mathcal{L}_T^\star(u, r_j) - r_j \mathcal{L}_T^\star(u, 1)] - [\mathcal{L}_T^\star(u, r_{j-1}) - r_{j-1} \mathcal{L}_T^\star(u, 1)] \right\|^2 W(u) \mathrm{d}u.$$

Note that proof of Theorem 4.2 implies

$$F_T(r_1, ..., r_M) \quad = \quad \sum_{j=1}^{M+1} \frac{1}{r_j - r_{j-1}} \int_{\mathbb{R}^d} \left\| [\mathcal{L}_T(u, r_j) - r_j \mathcal{L}_T(u, 1)] - [\mathcal{L}_T(u, r_{j-1}) - r_{j-1} \mathcal{L}_T(u, 1)] \right\|^2 W(u) \mathrm{d}u,$$

where

$$\mathcal{L}_T(u, r_j) = T^{-1/2} \sum_{t=1}^{T_j} \eta_t(u) \Rightarrow B(u, r_j)$$

under Lemma 4.1(i). To show the validity of the bootstrap test statistic, it suffices to show that

$$\mathcal{L}_T^\star(u, r_j) \Rightarrow B(u, r_j),$$

in probability over $\mathbb{U} \times [0, 1]$. We first establish (a) the convergence of $\mathcal{L}_T^\star(u, r_j)$ for each fixed $u \in \mathbb{U}$ and then show that (b) $\mathcal{L}_T^\star(u, r_j)$ is stochastically equicontinuous in $\mathbb{U}$.

We use Theorem 2 of Calhoun (2018) to establish (a). By Assumption 1(i) and $\sup_u \|Y_t e^{\mathbf{i}u'X_t}\|_2 \leq \mathcal{C}$, for each fixed $u \in \mathbb{U}$, $\eta_t(u) = Y_t e^{\mathbf{i}u'X_t} - \psi(u)$ is L_2 Near-Epoch-Dependent process of size $-1/2$ since the process $\{Y_t\}$ is strong mixing with mixing coefficient $\alpha(s) = O(s^{-\frac{q}{q-2}})$ for some $q > 2$. Hence, condition 1 in Theorem 2 of Calhoun (2018) is satisfied. Note that $E[\eta_t(u)] = 0$, and $\eta_t(u)$ is uniformly L_q bounded under Assumptions 1 and 3. Besides,

$$\sqrt{T} \left\| \frac{1}{T} \sum_{t=1}^{T} \eta_t(u) \right\| \xrightarrow{p} \Omega(u, u)^{1/2}$$

for each fixed u, where $\Omega(u, u)$ is defined in Lemma 4.1(i). Then, conditions 2 and 3 in Theorem 2 of Calhoun (2018) are satisfied. Finally, condition 4 in Theorem 2 of Calhoun (2018) holds under the condition on the block length l_T. Furthermore,

$$\frac{1}{T} \sum_{s=1}^{T_j} \sum_{t=1}^{T_j} \operatorname{cov}\left[\eta_t(u), \eta_s(u)^*\right] \to r_j \Omega(u, u)$$

under Lemma 4.1(i). Then, by Theorem 2 of Calhoun (2018),

$$\mathcal{L}_T^\star(u, r_j) \Rightarrow B(u, r_j),$$

for each fixed $u \in \mathbb{U}$.

Now, we show (b). For any $u_1, u_2 \in \mathbb{U}$,

$$\begin{aligned} E^\star \left\| \mathcal{L}_T^\star(u_1, r_j) - \mathcal{L}_T^\star(u_2, r_j) \right\|^2 &\leq 2E^\star \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \left[\eta_t^\star(u_1) - \eta_t^\star(u_2) \right] \right\|^2 + 2Tr_j^2 \left\| E^\star \left[\bar{\eta}^\star(u_1) - \bar{\eta}^\star(u_2) \right] \right\|^2 \\ &\leq 2I_1 + 2I_2, \text{ say.} \end{aligned}$$

By mean value theorem,

$$\sum_{t=1}^{T_j} [\eta_t^\star(u_1) - \eta_t^\star(u_2)] = \sum_{t=1}^{T_j} \Upsilon_t^\star(\bar{u})'(u_1 - u_2),$$

where $\bar{u}$ lies between u_1 and u_2, and

$$\Upsilon_t^\star(u) = \frac{\mathrm{d}\eta_t^\star(u)}{\mathrm{d}u} = \mathbf{i}\left[X_t^\star Y_t^\star e^{\mathbf{i}u'X_t^\star} - E(X_t Y_t e^{\mathbf{i}u'X_t}) \right].$$

Let $\mathcal{Q}_i(u) \equiv \Upsilon_i(u) + \cdots \Upsilon_{i+l_T-1}(u)$ for the ith block, $i = 1, ..., N$, and $\mathcal{Q}_k^\star(u) \equiv \Upsilon_{(k-1)l_T+1}^\star(u) + \cdots \Upsilon_{kl_T}^\star(u)$ for the kth resampled block, $k = 1, ..., K$. Then,

$$\begin{aligned} EI_1 &= E\left[E^\star \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T_j} \Upsilon_t^\star(\bar{u})'(u_1 - u_2) \right\|^2 \right] \\ &\leq E\left[E^\star \left\| \frac{1}{\sqrt{T}} \sum_{k=1}^{\lfloor Kr_j \rfloor} \mathcal{Q}_k^\star(\bar{u}) \right\|^2 \right] \|u_1 - u_2\|^2 \\ &= E\left[\frac{1}{T} \sum_{k=1}^{\lfloor Kr_j \rfloor} E^\star \left\| \mathcal{Q}_k^\star(\bar{u}) \right\|^2 \right] \|u_1 - u_2\|^2 \\ &= E\left[\frac{\lfloor Kr_j \rfloor}{T} \frac{1}{N} \sum_{i=1}^{N} \left\| \mathcal{Q}_i(\bar{u}) \right\|^2 \right] \|u_1 - u_2\|^2 \\ &= \frac{r_j}{N} \sum_{i=1}^{N} E \left\| l_T^{-1/2} \mathcal{Q}_i(\bar{u}) \right\|^2 \|u_1 - u_2\|^2 \\ &\leq \mathcal{C} \|u_1 - u_2\|^2. \end{aligned}$$

The second equality holds given that the blocks are i.i.d. conditioning on the observed data. The third equality holds since the blocks are drawn following a discrete uniform distribution. The last inequality holds given that $\sup_{u \in \mathbb{U}} \left\| l_T^{-1/2} \mathcal{Q}_i(u) \right\| = O_P(1)$ for each i, which can be established analogously to Lemma A.1.

For I_2, by mean value theorem, there exists $\tilde{u}$ that lies between u_1 and u_2, such that

$$\begin{aligned}
EI_2 &= Tr_j^2 E\left\|E^{\star}\left[\frac{1}{T}\sum_{t=1}^{T}\Upsilon_t^{\star}(\tilde{u})'(u_1-u_2)\right]\right\|^2 \\
&\leq r_j^2 E\left\|\frac{1}{\sqrt{T}}\sum_{k=1}^{K}E^{\star}\left[\mathcal{Q}_k^{\star}(\tilde{u})\right]\right\|^2 \|u_1-u_2\|^2 \\
&= r_j^2 E\left\|\frac{K}{\sqrt{T}}\frac{1}{N}\sum_{i=1}^{N}\mathcal{Q}_i(\tilde{u})\right\|^2 \|u_1-u_2\|^2 \\
&= r_j^2 E\left\|\frac{\sqrt{T}}{N}\sum_{t=1}^{T}\Upsilon_t(\tilde{u})-\frac{\sqrt{T}}{N}\sum_{k=1}^{l_T-1}\left(1-\frac{k}{l_T}\right)\left[\Upsilon_k(\tilde{u})+\Upsilon_{T-k+1}(\tilde{u})\right]\right\|^2 \|u_1-u_2\|^2 \\
&\leq 2r_j^2(I_{21}+I_{22})\|u_1-u_2\|^2,
\end{aligned}$$

where

$$I_{21} \equiv E\left\|\frac{\sqrt{T}}{N}\sum_{t=1}^{T}\Upsilon_t(\tilde{u})\right\|^2 = O(1),$$

uniformly for all $\tilde{u}\in\mathbb{U}$ by that $N=T-l_T+1$ and the mixing condition in Assumption 1, and

$$I_{22} \equiv E\left\|\frac{\sqrt{T}}{N}\sum_{k=1}^{l_T-1}\left(1-\frac{k}{l_T}\right)\left[\Upsilon_k(\tilde{u})+\Upsilon_{T-k+1}(\tilde{u})\right]\right\|^2 = O(T^{-1}l_T^2)=o(1),$$

under the moment restrictions in Assumption 1 and fact that $T^{-1/2}l_T = o(1)$. Thus, $EI_2 \leq \mathcal{C}\|u_1-u_2\|^2$. Now, (b) is established. Given $\mathbb{U}\times[0,1]$ is a compact set, by continuous mapping theorem, and analogous arguments in proof of Theorem 4.2,

$$F_T^{\star}(r_1,...,r_M) \Rightarrow F(r_1,...,r_M),$$

in probability.

Next, we show that the proposed MBB is asymptotic valid under $\mathbb{H}_A$. Let $\phi_t(u)\equiv E(Y_t e^{\mathbf{i}u'X_t})$ and $\phi_t^{\star}(u)\equiv E(Y_t^{\star}e^{\mathbf{i}u'X_t^{\star}})$. Given $Y_t^{\star}e^{\mathbf{i}u'X_t^{\star}}=\phi_t^{\star}(u)+\eta_t^{\star}(u)$ under $\mathbb{H}_A$, we have

$$\begin{aligned}
F_T^{\star}(r_1,...,r_M) &= \mathrm{SSGR}_0^{\star}-\mathrm{SSGR}_M^{\star} \\
&= \sum_{j=1}^{M+1}(T_j-T_{j-1})\int_{\mathbb{R}^d}\left\|\frac{1}{T_j-T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}Y_t^{\star}e^{\mathbf{i}u'X_t^{\star}}-\frac{1}{T}\sum_{t=1}^{T}Y_t^{\star}e^{\mathbf{i}u'X_t^{\star}}\right\|^2 W(u)\mathrm{d}u \\
&\leq 2\sum_{j=1}^{M+1}(T_j-T_{j-1})\int_{\mathbb{R}^d}\left\|\frac{1}{T_j-T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\eta_t^{\star}(u)-\frac{1}{T}\sum_{t=1}^{T}\eta_t^{\star}(u)\right\|^2 W(u)\mathrm{d}u
\end{aligned}$$

$$
\begin{aligned}
& +2\sum_{j=1}^{M+1}(T_j - T_{j-1})\int_{\mathbb{R}^d}\left\|\frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\phi_t^\star(u) - \frac{1}{T}\sum_{t=1}^{T}\phi_t^\star(u)\right\|^2 W(u)\mathrm{d}u \\
= \; & 2\mathcal{R}_1(r_1,...,r_M) + 2\mathcal{R}_2(r_1,...,r_M), \text{ say.}
\end{aligned}
$$

By the mixing condition in Assumption 1(i) and analogous proof in Lemma A.1, we can show $\sup_{r_1,...,r_M}|\mathcal{R}_1(r_1,...,r_M)| = O_{P^\star}(1)$. We consider $\mathcal{R}_2(r_1,...,r_M)$. By analogous steps, we have

$$
\begin{aligned}
& \mathcal{R}_2(r_1,...,r_M) \\
= \; & \sum_{j=1}^{M+1}(T_j - T_{j-1})\int_{\mathbb{R}^d}\left\|\frac{1}{T_j - T_{j-1}}\sum_{t=T_{j-1}+1}^{T_j}\phi_t^\star(u) - \frac{1}{T}\sum_{t=1}^{T}\phi_t^\star(u)\right\|^2 W(u)\mathrm{d}u \\
= \; & \sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{R}^d}\left\|\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_j}\phi_t^\star(u) - \frac{r_j}{\sqrt{T}}\sum_{t=1}^{T}\phi_t^\star(u)\right] - \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_{j-1}}\phi_t^\star(u) - \frac{r_{j-1}}{\sqrt{T}}\sum_{t=1}^{T}\phi_t^\star(u)\right]\right\|^2 W(u)\mathrm{d}u \\
\leq \; & 2\sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{R}^d}\left\|\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_j}\left\{\phi_t^\star(u) - E^\star[\bar{\phi}^\star(u)]\right\} - \frac{r_j}{\sqrt{T}}\sum_{t=1}^{T}\left\{\phi_t^\star(u) - E^\star[\bar{\phi}^\star(u)]\right\}\right]\right\|^2 W(u)\mathrm{d}u \\
& +2\sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{R}^d}\left\|\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_{j-1}}\left\{\phi_t^\star(u) - E^\star[\bar{\phi}^\star(u)]\right\} - \frac{r_{j-1}}{\sqrt{T}}\sum_{t=1}^{T}\left\{\phi_t^\star(u) - E^\star[\bar{\phi}^\star(u)]\right\}\right]\right\|^2 W(u)\mathrm{d}u \\
= \; & 2\sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{R}^d}\left\|\mathcal{N}_T^\star(u,r_j) - r_j\mathcal{N}_T^\star(u,1)\right\|^2 W(u)\mathrm{d}u \\
& +2\sum_{j=1}^{M+1}\frac{1}{r_j - r_{j-1}}\int_{\mathbb{R}^d}\left\|\mathcal{N}_T^\star(u,r_{j-1}) - r_{j-1}\mathcal{N}_T^\star(u,1)\right\|^2 W(u)\mathrm{d}u,
\end{aligned}
$$

where $E^\star[\bar{\phi}^\star(u)] \equiv E^\star[T^{-1}\sum_{t=1}^{T}\phi_t^\star(u)]$ is the expectation of the sample average of $\phi_t^\star(u)$ conditioning on the observable sample, and $\mathcal{N}_T^\star(u,r_j) \equiv T^{-1/2}\sum_{t=1}^{T_j}\left\{\phi_t^\star(u) - E^\star[\bar{\phi}^\star(u)]\right\}$. Let $\mathcal{X}_i(u) = \phi_i(u) + \cdots + \phi_{i+l_T-1}(u)$ for the i-th block, $i = 1,...,N$, and $\mathcal{X}_k^\star(u) = \phi_{(k-1)l_T+1}^\star(u) + \cdots + \phi_{kl_T}^\star(u)$ for the k-th resampled block, $k = 1,...,K$. Conditioning on the data, $\mathcal{X}_1^\star(u), \mathcal{X}_2^\star(u), ..., \mathcal{X}_K^\star(u)$ are independent and identically distributed. Given $T = l_T K$, and $N = T - l_T + 1$, then it follows

$$
\begin{aligned}
E^\star\left[\mathcal{N}_T^\star(u,r_j)\right] \; &= \; \sqrt{T}E^\star\left[\frac{1}{l_T K}\sum_{t=1}^{\lfloor l_T K r_j\rfloor}\phi_t^\star(u)\right] - \sqrt{T}r_j E^\star\left[\frac{1}{l_T K}\sum_{t=1}^{l_T K}\phi_t^\star(u)\right] \\
&= \; \frac{\sqrt{T}}{l_T K}\sum_{k=1}^{\lfloor K r_j\rfloor}E^\star\left[\mathcal{X}_k^\star(u)\right] - \frac{\sqrt{T}r_j}{l_T K}\sum_{k=1}^{K}E^\star\left[\mathcal{X}_k^\star(u)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\sqrt{T}}{l_T K}\sum_{k=1}^{\lfloor Kr_j \rfloor}\frac{1}{N}\sum_{i=1}^{N}\mathcal{X}_i(u) - \frac{\sqrt{T}r_j}{l_T K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\mathcal{X}_i(u) \\
&= \frac{\sqrt{T}r_j}{Nl_T}\sum_{i=1}^{N}\mathcal{X}_i(u) - \frac{\sqrt{T}r_j}{Nl_T}\sum_{i=1}^{N}\mathcal{X}_i(u) \\
&= 0,
\end{aligned}
$$

for all $\{r_j\}_{j=1}^{M+1}$. And

$$
\begin{aligned}
E^\star\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_j}\left\{\phi_t^\star(u) - E^\star[\bar{\phi}^\star(u)]\right\}\right]^2 &= \text{var}^\star\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T_j}\phi_t^\star(u)\right] \\
&= \frac{1}{T}\text{var}^\star\left[\sum_{k=1}^{\lfloor Kr_j \rfloor}\mathcal{X}_k^\star(u)\right] \\
&= \frac{Kr_j}{T}\text{var}^\star\left[\mathcal{X}_k^\star(u)\right] \\
&= \frac{Kr_j}{T}\frac{1}{N}\sum_{i=1}^{N}\|\mathcal{X}_i(u)\|^2 - \frac{Kr_j}{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\mathcal{X}_i(u)\right\|^2 \\
&= O_P(l_T).
\end{aligned}
$$

It implies that $\mathcal{N}_T^\star(u, r_j) = O_{P^\star}(l_T^{1/2})$ for all u and r_j. The uniform results can be established in a similar way as in Proof of Theorem 4.2. Then, we have that under $\mathbb{H}_A$, the bootstrap joint test statistic $\sup F_T^\star(r_1, ..., r_M) = O_P(l_T)$.

Note that Theorem 4.3 implies our test statistic

$$
\begin{aligned}
\sup F_T(r_1, ..., r_M) &= \text{SSGR}_0 - \text{SSGR}_M(\hat{r}_1, ..., \hat{r}_M) \\
&= O_P(Ta_T^2),
\end{aligned}
$$

under $\mathbb{H}_A(a_T)$. Hence, if $Ta_T^2 l_T^{-1} \to \infty$, we have that $P(\sup F > \sup F^\star) \to 1$ as $T \to \infty$. ■

Proof of Theorem 5.1 Under the null of M breaks, we let $D(T_j, T_k)$ to be the SSGR using the data within the segment specified by $[T_j + 1, T_k]$ for $0 \le j < k \le M + 1$, then it is obvious that

$$
\text{SSGR}_M(T_1^0, ..., T_M^0) = \sum_{k=1}^{M+1} D(T_{k-1}^0, T_k^0),
$$

and

$$
\text{SSGR}_{M+1}\left(T_1^0, ..., T_{j-1}^0, \tau, T_j^0, ..., T_M^0\right) = \sum_{k=1}^{j-1} D(T_{k-1}^0, T_k^0) + D(T_{j-1}^0, \tau) + D(\tau, T_j^0) + \sum_{k=j+1}^{M+1} D(T_{k-1}^0, T_k^0),
$$

for some $1 < j < M$. We note that

$$\begin{aligned}\mathrm{SSGR}_{M+1}\left(T_1^0,...,T_{j-1}^0,\tau,T_j^0,...,T_M^0\right) &\equiv \mathrm{SSGR}_{M+1}\left(\tau,T_1^0,...,T_M^0\right)\\ &= D(1,\tau)+D(\tau,T_1^0)+\sum_{k=2}^{M+1}D(T_{k-1}^0,T_k^0),\end{aligned}$$

when $j = 1$, and

$$\begin{aligned}\mathrm{SSGR}_{M+1}\left(T_1^0,...,T_{j-1}^0,\tau,T_j^0,...,T_M^0\right) &\equiv \mathrm{SSGR}_{M+1}\left(T_1^0,...,T_M^0,\tau\right)\\ &= \sum_{k=1}^{M}D(T_{k-1}^0,T_k^0)+D(T_M^0,\tau)+D(\tau,T),\end{aligned}$$

when $j = M$. Let

$$\begin{aligned}F_T^0(M+1|M) &= \mathrm{SSGR}_M\left(T_1^0,...,T_M^0\right)-\min_{1\le j\le M+1}\inf_{\tau\in\Lambda_{j,\epsilon}^0}\mathrm{SSGR}_{M+1}\left(T_1^0,...,T_{j-1}^0,\tau,T_j^0,...,T_M^0\right)\\ &= \max_{1\le j\le M+1}\sup_{\tau\in\Lambda_{j,\epsilon}^0}\left[\mathrm{SSGR}_M\left(T_1^0,...,T_M^0\right)-\mathrm{SSGR}_{M+1}\left(T_1^0,...,T_{j-1}^0,\tau,T_j^0,...,T_M^0\right)\right]\\ &= \max_{1\le j\le M+1}\sup_{\tau\in\Lambda_{j,\epsilon}^0}\left[D(T_{j-1}^0,T_j^0)-D(T_{j-1}^0,\tau)-D(\tau,T_j^0)\right],\end{aligned}$$

where

$$\Lambda_{j,\epsilon}^0=\left\{\tau:T_{j-1}^0+\lfloor(T_j^0-T_{j-1}^0)\epsilon\rfloor\le\tau\le T_j^0-\lfloor(T_j^0-T_{j-1}^0)\epsilon\rfloor\right\},$$

for some arbitrary small $\epsilon > 0$. It is obvious to see that

$$\sup_{\tau\in\Lambda_{j,\epsilon}^0}\left[D(T_{j-1}^0,T_j^0)-D(T_{j-1}^0,\tau)-D(\tau,T_j^0)\right]$$

is equivalent to the sup-F test statistic in Theorem 4.2 in a subsample specified by $[T_{j-1}^0, T_j^0]$ with $M = 1$ in the alternative hypothesis. It follows

$$\begin{aligned}&D(T_{j-1}^0,T_j^0)-D(T_{j-1}^0,\tau)-D(\tau,T_j^0)\\ =\;&(\tau-T_{j-1}^0)\int_{\mathbb{R}^d}\left\|\frac{1}{T_j^0-T_{j-1}^0}\sum_{t=T_{j-1}^0+1}^{T_j^0}\eta_t(u)-\frac{1}{\tau-T_{j-1}^0}\sum_{t=T_{j-1}^0+1}^{\tau}\eta_t(u)\right\|^2W(u)\mathrm{d}u\\ &+(T_j^0-\tau)\int_{\mathbb{R}^d}\left\|\frac{1}{T_j^0-T_{j-1}^0}\sum_{t=T_{j-1}^0+1}^{T_j^0}\eta_t(u)-\frac{1}{T_j^0-\tau}\sum_{t=\tau+1}^{T_j^0}\eta_t(u)\right\|^2W(u)\mathrm{d}u.\end{aligned}$$

Let $r = \frac{\tau - T_{j-1}^0}{T_j^0 - T_{j-1}^0}$. Given Lemma 4.1(ii), we have for $t \in [T_{j-1}^0, T_j^0]$,

$$\frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{\tau} \eta_t(u) = \frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0 + \lfloor (T_j^0 - T_{j-1}^0) r \rfloor} \eta_t(u) \Rightarrow B^{(j)}(u, r).$$

Hence,

$$\begin{aligned}
& D(T_{j-1}^0, T_j^0) - D(T_{j-1}^0, \tau) - D(\tau, T_j^0) \\
= \; & \frac{1}{\tau - T_{j-1}^0} \int_{\mathbb{R}^d} \left\| \frac{\tau - T_{j-1}^0}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \sum_{t=T_{j-1}^0+1}^{\tau} \eta_t(u) \right\|^2 W(u) \mathrm{d}u \\
& + \frac{1}{T_j^0 - \tau} \int_{\mathbb{R}^d} \left\| \frac{T_j^0 - \tau}{T_j^0 - T_{j-1}^0} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \sum_{t=\tau+1}^{T_j^0} \eta_t(u) \right\|^2 W(u) \mathrm{d}u \\
= \; & \frac{1}{\lfloor r(T_j^0 - T_{j-1}^0) \rfloor} \int_{\mathbb{R}^d} \left\| r \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \sum_{t=T_{j-1}^0+1}^{\tau} \eta_t(u) \right\|^2 W(u) \mathrm{d}u \\
& + \frac{1}{\lfloor (1-r)(T_j^0 - T_{j-1}^0) \rfloor} \int_{\mathbb{R}^d} \left\| (1-r) \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \left[\sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \sum_{t=T_{j-1}^0+1}^{\tau} \eta_t(u) \right] \right\|^2 W(u) \mathrm{d}u \\
= \; & \frac{1}{r} \int_{\mathbb{R}^d} \left\| r \frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{\tau} \eta_t(u) \right\|^2 W(u) \mathrm{d}u \\
& + \frac{1}{1-r} \int_{\mathbb{R}^d} \left\| r \frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{T_j^0} \eta_t(u) - \frac{1}{\sqrt{T_j^0 - T_{j-1}^0}} \sum_{t=T_{j-1}^0+1}^{\tau} \eta_t(u) \right\|^2 W(u) \mathrm{d}u \\
\Rightarrow \; & \frac{1}{r(1-r)} \int_{\mathbb{R}^d} \left\| B^{(j)}(u, r) - r B^{(j)}(u, 1) \right\|^2 W(u) \mathrm{d}u.
\end{aligned}$$

It follows

$$\sup_{\tau \in \Lambda_{j,\epsilon}^0} \left[D(T_{j-1}^0, T_j^0) - D((T_{j-1}^0, \tau) - D(\tau, T_j^0) \right] \xrightarrow{d} \sup_{\epsilon \le r \le 1-\epsilon} \int_{\mathbb{R}^d} \frac{\left\| B^{(j)}(u, r) - r B^{(j)}(u, 1) \right\|^2}{r(1-r)} W(u) \mathrm{d}u.$$

Hence, we have

$$F_T^0(M+1|M) = \max_{1 \le j \le M+1} \sup_{\tau \in \Lambda_{j,\epsilon}^0} \left[D(T_{j-1}^0, T_j^0) - D(T_{j-1}^0, \tau) - D(\tau, T_j^0) \right]$$

$$\xrightarrow{d} \max_{1\leq j\leq M+1} \sup_{\epsilon\leq r\leq 1-\epsilon} \int_{\mathbb{R}^d} \frac{\left\|\mathcal{B}^{(j)}(u,r)\right\|^2}{r(1-r)} W(u)\mathrm{d}u,$$

where $\mathcal{B}^{(j)}(u,r) \equiv B^{(j)}(u,r) - rB^{(j)}(u,1)$ is a generalized Brownian bridge under the j-th subsample, $j = 1, ..., M+1$. Under the null hypothesis, Theorem 3.1 indicates that $\hat{r}_j = r_j^0 + O_P(T^{-1})$. Based on this result, we know the above equation also holds with T_{j-1}^0 and T_j^0 replaced by $\hat{T}_{j-1}$ and $\hat{T}_j$, respectively. Hence, the limiting distribution of $F_T(M+1|M)$ is the same as $F_T^0(M+1|M)$. ■

Proof of Theorem 5.2. Let $\mathcal{M}_0 = \left\{M \in \mathcal{M} : M = M^0\right\}, \mathcal{M}_- = \left\{M \in \mathcal{M} : M < M^0\right\}$, and $\mathcal{M}_+ = \left\{M \in \mathcal{M} : M > M^0\right\}$ denote the subset of $\mathcal{M} = \{1, 2, \cdots, M_{\max}\}$ which correctly estimate, under-estimate, and over-estimate the true number of breaks. We show that neither the under-fitted model nor the over-fitted model can minimize the information criterion function, i.e., $P\left(\min_{M\in(\mathcal{M}_-\cup\mathcal{M}_+)} IC(M) > IC\left(M^0\right)\right) \to 1$ as $T \to \infty$, where

$$IC(M) \equiv \ln\left[\hat{\sigma}^2(M)\right] + \rho_T(M+1)$$

with $\hat{\sigma}^2(M) = T^{-1}\,\mathrm{SSGR}_M\left(\hat{r}_1, \ldots, \hat{r}_M\right)$.

Under Assumption 4, it follows that

$$\begin{aligned}
IC\left(M^0\right) &= \ln\left[\hat{\sigma}^2\left(M^0\right)\right] + \rho_T\left(M^0+1\right) \\
&= \ln\left[\frac{1}{T}\sum_{j=1}^{M^0+1}\sum_{t=\hat{T}_{j-1}+1}^{\hat{T}_j}\int_{\mathbb{R}^d}\left\|Y_t e^{\mathbf{i}uX_t} - \hat{\psi}_j(u)\right\|^2 W(u)\mathrm{d}u\right] + o(1) \\
&= \ln\left[\frac{1}{T}\sum_{j=1}^{M^0+1}\sum_{t=T_{j-1}^0+1}^{T_j^0}\int_{\mathbb{R}^d}\left\|Y_t e^{\mathbf{i}uX_t} - \psi_j^0(u)\right\|^2 W(u)\mathrm{d}u\right] + o_P(1) \\
&= \ln\left[\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^d}\left\|\eta_t(u)\right\|^2 W(u)\mathrm{d}u\right] + o_P(1)
\end{aligned}$$

When the model is under-fitted, i.e., $M < M^0$, Lemma A.2 implies that the estimated break fractions $\{\hat{r}_1, \cdots, \hat{r}_M\}$ are consistent for M breaks in the collection of true break fractions $\left\{r_j^0\right\}_{j=1}^{M^0}$. Then there must exist $M^0 - M$ break fractions in $\left\{r_j^0\right\}_{j=1}^{M^0}$ that can not be identified.

Without loss of generality, we assume the j-th break r_j^0 can not be consistently estimated. Then with some positive probability $0 < c_0 < 1$, there exists an $\eta > 0$ such that no estimated break fraction falls in the interval $\left[T_j^0 - \lfloor\eta T\rfloor + 1, T_j^0 + \lfloor\eta T\rfloor\right]$. Suppose this interval is classified into the k-th regime, i.e., $\hat{T}_{k-1} \leq T_j^0 - \lfloor\eta T\rfloor + 1 < T_j^0 + \lfloor\eta T\rfloor \leq \hat{T}_k$. Then based on the proof of

Theorem 3.1, we have

$$\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\hat{\eta}_t(u)\|^2 W(u)\mathrm{d}u \geq \frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\eta_t(u)\|^2 W(u)\mathrm{d}u + \delta\int_{\mathbb{R}^d}\left\|\psi_{j+1}^0(u)-\psi_j^0(u)\right\|^2 W(u)\mathrm{d}u + o_P(1)$$

with probability no less than c_0 for any $M < M^0$, where $\delta = \frac{1}{2}\eta$. It follows

$$\begin{aligned}\min_{M\in\mathcal{M}-} IC(M) - IC\left(M^0\right) &= \min_{M\in\mathcal{M}-}\left[\ln\left(\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\hat{\eta}_t(u)\|^2 W(u)\mathrm{d}u\right) + \rho_T(M+1)\right] \\ &\quad - \left[\ln\left(\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\eta_t(u)\|^2 W(u)\mathrm{d}u\right) + \rho_T\left(M^0+1\right) + o_P(1)\right] \\ &\geq \ln\left[1 + \frac{\delta\int_{\mathbb{R}^d}\left\|\psi_{j+1}^0(u)-\psi_j^0(u)\right\|^2 W(u)\mathrm{d}u}{\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\eta_t(u)\|^2 W(u)\mathrm{d}u}\right] + o_P(1)\end{aligned}$$

Therefore, for any positive probability $0 < c_0 < 1$, there exists an $\eta = 2\delta$ such that

$$P\left(\min_{M\in\mathcal{M}_-} IC(M) - IC\left(M_0\right) > \Delta\right) = c_0$$

for T sufficiently large, where $0 < \Delta < \ln\left[1 + \frac{\delta\int_{\mathbb{R}^d}\left\|\psi_{j+1}^0(u)-\psi_j^0(u)\right\|^2 W(u)\mathrm{d}u}{\frac{1}{T}\int_{\mathbb{R}^d}\sum_{t=1}^{T}\|\eta_t(u)\|^2 W(u)\mathrm{d}u}\right]$. Then

$$P\left(\min_{M\in\mathcal{M}_-} IC(M) > IC\left(M^0\right)\right) \to 1$$

as $T \to \infty$.

When the model is over-fitted, by Lemma A.3, we have

$$\begin{aligned}&P\left(\min_{M\in\mathcal{M}_+} IC(M) > IC\left(M^0\right)\right) \\ =&P\left(\min_{M\in\mathcal{M}_+}\left[T\ln\left(\hat{\sigma}^2(M)/\hat{\sigma}^2\left(M^0\right)\right) + T\rho_T\left(M-M^0\right)\right] > 0\right) \\ =&P\left(\min_{M\in\mathcal{M}_+}\left[T\left(\hat{\sigma}^2(M)-\hat{\sigma}^2\left(M^0\right)\right)/\hat{\sigma}^2\left(M^0\right) + T\rho_T\left(M-M^0\right)\right] + o_P(1) > 0\right) \\ \to&1 \text{ as } T\to\infty\end{aligned}$$

Therefore, we have proved that

$$P\left(\min_{M\in\mathcal{M}-\cup\mathcal{M}_+} IC(M) > IC\left(M^0\right)\right) \to 1 \text{ as } T\to\infty$$