Consistent Testing for Structural Changes in Time Series Models via Discrete Fourier Transform*

Zhonghao Fu[a,b], Yongmiao Hong[c], Xia Wang[d]

[a] Fudan University

[b] Shanghai Institute of International Finance and Economics

[c] Chinese Academy of Sciences and University of Chinese Academy of Sciences

[d] Renmin University of China

Abstract

We propose Cramér-von Mises and Kolmogorov-Smirnov type tests for structural changes in linear time series models, possibly with endogenous regressors via a discrete Fourier transform (DFT) approach. If structural changes exist, the conventional OLS and 2SLS estimators are inconsistent for the unknown coefficients. Consequently, the estimated residuals contain the time-varying features of the model parameters. By DFT, we can capture such information in the frequency domain. The proposed tests are powerful against smooth structural changes and abrupt structural breaks and can detect a class of local alternatives at the parametric rate, which is asymptotically more efficient than the existing nonparametric tests. Simulation studies demonstrate the excellent finite sample performance of our tests. In an application to the U.S. Taylor rule, we find significant evidence of structural changes during the post-1979 period, which is treated as a stable period by Clarida et al. (2000).

Keywords: Discrete Fourier transform, Endogeneity, Local power, Structural change

JEL classification: C12; C14

*Fu acknowledges financial support from National Science Foundation of China (NSFC) (No. 71903032), Hong acknowledges financial support from NSFC's Fundamental Scientific Center Project (No. 71988101), and Wang acknowledges financial support from NSFC (No. 71873151). We thank Haiqiang Chen, Liyuan Cui, Jiti Gao, Kristoffer Nimark, and seminar participants in Xiamen University, Cornell University, Chinese University of Hong Kong, and UC Riverside for their helpful comments and suggestions. All remaining errors are solely ours. Correspondance: Zhonghao Fu, *zhfu@fudan.edu.cn*; Yongmiao Hong, *yh20@cornell.edu*; Xia Wang, *wxia@ruc.edu.cn*.

1 Introduction

Time series data usually have structural changes due to varying economic environments, e.g., shocks, policy shifts, technology progress, and preference changes. Many empirical studies have confirmed the prevalence of structural instability in financial and macroeconomic time series. For example, Welch and Goyal (2008) show that the in-sample significant predictability of most financial and macroeconomic variables fails to yield better out-of-sample forecasts of the U.S. equity premium than the simple historical mean of equity returns. One possible reason is the instability of model parameters, as Chen and Hong (2012) confirmed. In labor economics, Hansen (2001) demonstrates evidence of structural breaks in labor productivity between 1992 and 1996, in the 1960s, and the early 1980s. In macroeconomics, Stock and Watson (1996) find substantial instability in 76 representative U.S. monthly post-war macroeconomic time series. Zhang et al. (2008) show that the instability of parameters in the New Keynesian Phillips Curve results in conflicting conclusions about the critical determinant of the short-run inflation dynamics. In the current big data era, the abundance of data is not necessarily equivalent to the sufficiency of relevant information. Specifically, if the underlying DGP changes over time, big data may also result in misleading conclusions. Therefore, detecting structural changes is crucial for econometric modeling and inference.

Testing for structural changes in time series regressions has drawn sustained attention in the literature. Although most existing studies focus on dealing with abrupt structural breaks (e.g., Perron, 2006), estimation and testing for smooth structural changes have drawn increasing attention. Intuitively, it is quite likely that economic agents digest and react to shocks (e.g., policy shifts or unexpected income) gradually. Even when the change is sudden at the individual level, it may appear to be smooth at the aggregate level. As a result, the parameters in a time series model usually change smoothly over time rather than shift abruptly. Smooth structural changes can be modeled parametrically. One example is the Smooth Transition Regression (STR) model developed by Lin and Teräsvirta (1994), where they choose a particular parametric function to model the time-varying parameters. However, economic theories usually do not indicate how parameters evolve. To avoid model misspecification, nonparametric time-varying regressions are introduced by Robinson (1989, 1991) and further studied by Orbe et al. (2000, 2005), Cai (2007), Chen and Hong (2012), Kristensen (2012), Zhang and Wu (2012), Cai et al. (2015), and Xu (2015). Among many others, Chen and Hong (2012) propose a generalized Hausman test for both smooth structural changes and abrupt structural breaks in a linear time series regression model. Kristensen (2012) considers estimation and testing in both mean and variance in a time series dynamics. Cai et al. (2015) test the instability of model parameters when the regressors follow a unit root process. Xu (2015) constructs a CUSUM-type test for smooth structural changes in the regression coefficient when the variance is time-varying.

The aforementioned tests can detect the instability of parameters of unknown forms in a linear time series model with exogenous regressors, but they are not applicable when endogeneity exists, which is quite common in many applications. Therefore, several recent works have considered estimation and testing in a linear time series model with endogenous regressors. For example, Hall et al. (2012) extend Bai and Perron's (1998) approach to estimating and testing for abrupt breaks of linear models with endogeneity using the Two-Stage Least Squares (2SLS). However, their approach requires identifying the structural breaks in the reduced form. If there exist too many breaks or smooth structural changes in the reduced form, their approach is not applicable. Perron and Yamamoto (2014) consider estimation and testing based on 2SLS. Perron and Yamamoto (2015) propose a test for abrupt structural breaks using OLS rather than 2SLS. They show that the OLS-based test is more efficient than those based on the 2SLS in most cases. However, the inference is restricted to the dates and magnitudes of breaks in the reduced form, and their test is not consistent against local alternatives of certain directions. Unlike the studies that focus on abrupt structural breaks, Chen (2015) proposes a Two-Stage Local Linear (2SLL) method for estimation and testing when the unknown parameters change smoothly over time. However, the test statistic is computed via smoothed nonparametric estimation, which involves selecting bandwidths for both the structural function and the first stage reduced form. Moreover, certain smoothness conditions are required to ensure the consistency of smoothed nonparametric estimation, which may be restrictive since we usually have no prior information about the property of unknown parameters over time. Moreover, Chen's (2015) test can detect local alternatives at rate $T^{-1/2}h^{-1/4}$, which depends on the bandwidth h, and is thus slower than $T^{-1/2}$.

In this paper, we propose Cramér-von Mises and Kolmogorov-Smirnov type tests for structural changes in a time series model via discrete Fourier transform (DFT). Unlike the existing tests that focus on time-domain analysis, we investigate structural stability in the frequency domain. The intuition is straightforward: if structural changes exist, then conventional estimation methods such as OLS and 2SLS based on the whole sample will miss them because these estimators cannot capture time-varying features of unknown parameters. Consequently, the estimated residuals will contain such information. By projecting the estimated residuals onto the frequency domain via DFT, we can infer structural changes by examining the corresponding spectrum. Our frequency domain-based tests have the following merits.

To begin with, our tests are consistent against various kinds of smooth structural changes, abrupt structural breaks, or a mixture of both. They can detect a class of local alternatives that converges to the null hypothesis at the parametric rate, which is asymptotically more efficient than the existing tests for smooth structural changes (e.g., Chen and Hong, 2012; Zhang and Wu, 2012; Cai et al., 2015; Chen, 2015). That is an appealing advantage of the DFT, which avoids nonparametric smoothing. Hence, our tests avoid the delicate business of choosing a bandwidth and are more powerful than the existing nonparametric tests.

Besides, our tests work for a linear model with endogenous regressors. Note that our DFT tests detect structural changes using the estimated residuals. When endogeneity exists, one only needs to replace the estimated residuals from OLS with those from 2SLS. Also, our tests allow for possible structural changes in the regressors and instrument variables. That is an improvement over Hall et al. (2012) and Perron and Yamamoto (2014), where the instability in the first stage reduced form has a nontrivial impact on testing and estimation.

Last but not least, our tests do not require trimming of the boundary regions of the sample. In contrast, the existing tests for abrupt structural breaks, such as Andrews' (1993) supremum test and Bai and Perron's (1998) double maximum test, rely on a pre-specified trimming parameter and will miss structural changes in the boundary regions of the sample.

The rest of this paper is organized as follows. We introduce our DFT-based tests in Section 2 and establish the asymptotic theory in Section 3. In Section 4, we extend our tests by considering the case with endogeneity. We then examine their finite sample performance in Section 5 and provide an empirical application to the U.S. Taylor rule in Section 6. We conclude in Section 7. All mathematical proofs are collected in the Mathematical Appendix.

Notation: Throughout this paper, $\mathbf{i}$ denotes the imaginary number such that $\mathbf{i} = \sqrt{-1}$, and "$\equiv$" means "is defined as". For an $m \times n$ complex-valued matrix $A = (a_{ij})$, where a_{ij} is the (i, j)th entry for $i = 1, ..., m; j = 1, ...n$, we denote its complex conjugate as $A^* = (a_{ji}^*)$, its transpose as $A' = (a_{ji})$, its real part as $\mathrm{Re}(A) = (\mathrm{Re}[a_{ij}])$, its Euclidean norm as $\|A\| = (\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2)^{1/2}$, and its Moore-Penrose generalized inverse as A^-. We use $\mathcal{M} \in (0, \infty)$ 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 Hypotheses and Test Statistics

In this section, we introduce the hypotheses of interest and construct the test statistics for structural changes based on the DFT.

2.1 Hypotheses of Interest

We consider testing for structural changes of unknown forms in the following linear time series regression:

$$Y_t = X_t'\theta_t + \varepsilon_t, \tag{2.1}$$

where $\{X_t\}_{t=1}^T$ is an $\mathbb{R}^d$-valued time series with T observations, θ_t is a $d \times 1$ unknown parameter vector that may vary over time, and ε_t is an unobservable error term. We note that X_t can contain the lagged observations of Y_t. Let $\Theta \subset \mathbb{R}^d$ denote a bounded parameter space of time-invariant regression coefficients. We test

$$\mathbb{H}_0 : \theta_t = \theta_0 \text{ for some unknown } \theta_0 \in \Theta \text{ and for all } t = 1, 2, ..., T,$$

against

$$\mathbb{H}_A : \theta_t \neq \theta \text{ for all } \theta \in \Theta \text{ and for some } t \in \mathbb{T} \text{ with } \mathbf{c}(\mathbb{T}) = O(T),$$

where $\mathbb{T}$ denotes the set of time indices that $\theta_t \neq \theta$ and $\mathbf{c}(\cdot)$ denotes a counting measure. Under $\mathbb{H}_0$, (2.1) is the classical linear regression model without structural change. While under $\mathbb{H}_A$, there exist abrupt structural breaks if θ_t is a step function of time, and there exist smooth structural changes if $\theta_t \equiv \theta(t/T)$ is a smooth function of the rescaled time index $t/T \in (0, 1]$.

A straightforward approach to test $\mathbb{H}_0$ is to compare the consistent estimates for θ_t under the null and the alternative hypotheses, e.g., Chen and Hong (2012), Kristensen (2012), Zhang and Wu (2012), and Cai et al. (2015). To avoid model misspecification and capture various kinds of structural changes, the literature mentioned above estimates the model under the alternative hypothesis via non-parametric regression, which involves smoothing over the rescaled time index t/T. As a result, the convergence rates of the estimators and the related tests are slower than the parametric rate; see Section 3.2 for more discussion. Also, the nonparametric regression relies on the smoothing parameters, i.e., a bandwidth. Even though the choice of bandwidth has a trivial impact asymptotically, it is crucial in a finite sample. Two practitioners may obtain conflicting results by choosing different bandwidths.

2.2 Discrete Fourier Transform

To avoid the undesired features of smoothed nonparametric estimation, we construct test statistics based on the DFT. Let $X_t(u) \equiv X_t e^{\mathbf{i}2\pi ut/T}$ be the product of X_t and the Fourier basis function of time, where $X_t(0) = X_t$. We adopt the following notations: $\hat{Q}_{xx}(u) \equiv T^{-1}\sum_{t=1}^{T} X_t(u)X_t'$, $\tilde{Q}_{xx}(u) \equiv T^{-1}\sum_{t=1}^{T} E\left[X_t(u)X_t'\right]$, and $Q_{xx}(u) \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[X_t(u)X_t'\right]$. Obviously, $\tilde{Q}_{xx}(u) = E\hat{Q}_{xx}(u)$, and $Q_{xx}(u)$ is the limit of $\tilde{Q}_{xx}(u)$. For notational convenience, we suppress the dependence of $\hat{Q}_{xx}(u)$ and $\tilde{Q}_{xx}(u)$ on the sample size T. When $u = 0$, we simply use $\hat{Q}_{xx}$, $\tilde{Q}_{xx}$, and Q_{xx} to represent $\hat{Q}_{xx}(0)$, $\tilde{Q}_{xx}(0)$, and $Q_{xx}(0)$, respectively. Analogous rules apply to $\hat{Q}_{x\varepsilon}(u)$, $\tilde{Q}_{x\varepsilon}(u)$, $Q_{x\varepsilon}(u)$, $\hat{Q}_{xy}(u)$, $\tilde{Q}_{xy}(u)$, and $Q_{xy}(u)$.

In this paper, we allow X_t to be nonstationary such that it can change smoothly or shift abruptly. Our definition of nonstationarity is weaker than the asymptotical MSE-stationarity in Hansen (2000), where an array $\{X_t\}_{t=1}^{T}$ is asymptotically MSE-stationary if $T^{-1}\sum_{t=1}^{[\tau T]} X_t X_t' \xrightarrow{p} \tau Q_{xx}$, for all $\tau \in (0, 1]$. Since our analysis does not involve sample splitting, we only require it to hold in the whole sample, i.e., $\tau = 1$, rather than all $\tau \in (0, 1]$. Therefore, we can allow X_t to have structural changes of unknown forms, and that is an improvement over the existing literature.

The main idea of our tests is to capture the time-varying feature of θ_t without estimating it directly. However, θ_t is incorporated in Y_t through X_t. As we do not restrict X_t to be stationary, structural changes in Y_t may arise from the time-varying marginal distribution of X_t. So we need to purge that from Y_t first. Suppose $E(\varepsilon_t|X_t) = 0$ and $\hat{\theta}$ is the OLS estimator. It is straightforward to show that $\hat{\theta} = \tilde{\theta} + O_P(T^{-1/2})$, where $\tilde{\theta} \equiv \tilde{Q}_{xx}^{-1}\left[T^{-1}\sum_{t=1}^{T} E(X_t X_t')\theta_t\right]$ is a weighted average of θ_t. In particular, when X_t is weakly stationary, $\tilde{\theta} = T^{-1}\sum_{t=1}^{T}\theta_t$. Obviously, the OLS estimator cannot capture the time-varying feature of θ_t, which will be contained in the estimated residuals. Consider the estimated residuals

$$\hat{\varepsilon}_t = X_t'(\theta_t - \hat{\theta}) + \varepsilon_t = \varepsilon_t - X_t'\hat{Q}_{xx}^{-1}\hat{Q}_{x\varepsilon} + X_t'\left[\theta_t - \hat{Q}_{xx}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T} X_s X_s' \theta_s\right)\right].$$

It implies that $\hat{\varepsilon}_t$ is decomposed into three parts: the error term ε_t, the estimation uncertainty, and the time-varying behavior of θ_t which is identically zero under

$\mathbb{H}_0$. To investigate θ_t's local feature without estimating it directly, we define the following complex-valued empirical process:

$$\hat{A}(u) = \frac{1}{T}\sum_{t=1}^{T} X_t \hat{\varepsilon}_t e^{\mathbf{i}2\pi ut/T}, \tag{2.2}$$

which is the DFT of $X_t\hat{\varepsilon}_t$. By the definition of $\hat{\varepsilon}_t$, we decompose

$$\hat{A}(u) \;=\; \frac{1}{T}\sum_{t=1}^{T} \hat{M}_t(u) X_t' \theta_t + \frac{1}{T}\sum_{t=1}^{T} \hat{M}_t(u)\varepsilon_t \;\equiv\; \hat{A}_1(u) + \hat{A}_2(u), \tag{2.3}$$

where $\hat{M}_t(u) = X_t(u) - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_t$ is a complex-valued process of u such that $T^{-1}\sum_{t=1}^{T}\hat{M}_t(u)X_t' = 0$ for all $u \in \mathbb{R}$. Intuitively, $\hat{M}_t(u)$ can be viewed as a projection of X_t onto a frequency space which is orthogonal to X_t under $\mathbb{H}_0$.

Under $\mathbb{H}_0$, $\hat{A}_1(u)$ is identically zero for all u. Then the asymptotic behavior of $\hat{A}(u)$ is dominated by $\hat{A}_2(u)$, which is purely a stochastic process that converges to zero in the frequency domain (zero spectrum). While under $\mathbb{H}_A$, $\hat{M}_t(u)$ is not orthogonal to $X_t'\theta_t$. Then $\hat{A}_1(u)$ contains the time-varying property of θ_t. As a result, $\hat{A}(u)$ is dominated by $\hat{A}_1(u)$, and it will converge to a nonzero spectrum. In particular, if X_t is weakly stationary, and $\theta_t = \theta(t/T)$ is a smooth function of the rescaled time index $t/T \in (0, 1]$, then the limit of $\hat{A}_1(u)$ is proportional to the pseudo-covariance of θ_t and $e^{\mathbf{i}2\pi ut/T}$ in the sense that t/T follows the uniform distribution $U[0, 1]$. Intuitively, the DFT is equivalent to running a nonparametric regression of θ_t via the Fourier basis functions at various frequencies. Under $\mathbb{H}_0$, the DFT only captures the noise, and the impact of X_t has been purged. Otherwise, the DFT contains the time-varying feature of θ_t and thus will converge to a nonzero spectrum under $\mathbb{H}_A$. We examine $\hat{A}(u)$ for all $u \in \mathbb{R}$, which is equivalent to checking the DFT at all Fourier frequencies.

We note that the DFT has an alternative theoretical interpretation that it can be viewed as a generalized Hausman test in the frequency domain:

$$\hat{A}(u) \;=\; \hat{Q}_{xx}(u)\left[\hat{Q}_{xx}^{-}(u)\hat{Q}_{xy}(u) - \hat{Q}_{xx}^{-1}\hat{Q}_{xy}\right] \;=\; \hat{Q}_{xx}(u)[\hat{\theta}(u) - \hat{\theta}], \tag{2.4}$$

where $\hat{\theta}(u) \equiv \hat{Q}_{xx}^{-}(u)\hat{Q}_{xy}(u)$ is a weighted least squares (WLS) estimator. Here, we use the Moore-Penrose generalized inverse $\hat{Q}_{xx}(u)^-$ because it is possible that $\hat{Q}_{xx}(u)$ is singular for some u. In fact, the DFT-based test does not require to compute the inverse matrix of $\hat{Q}_{xx}(u)$. Introducing the Moore-Penrose generalized inverse is only for illustration. Under $\mathbb{H}_0 : \theta_t = \theta_0$, it is straightforward to show that $\hat{\theta}(u)$ is consistent for θ_0 at each u: $\hat{\theta}(u) = \theta_0 + \hat{Q}_{xx}^-(u)\hat{Q}_{x\varepsilon}(u) \xrightarrow{p} \theta_0$, given $E(\varepsilon_t|X_t) = 0$. Although $\hat{\theta}(u)$ is a complex-valued function of the nuisance parameter u, the entire complex-valued term vanishes to 0 as $T \to \infty$ under $\mathbb{H}_0$. Therefore the probability limits of the WLS estimator are identical to each other for almost all u. However, if there exist structural changes, the weights $e^{\mathbf{i}u2\pi t/T}$ will capture that information since various u's deliver different weights to each time point t. Thus, under $\mathbb{H}_A : \theta_t \neq \theta_0$, we have $\hat{\theta}(u) = \hat{Q}_{xx}^-(u)\left[\frac{1}{T}\sum_{t=1}^{T} X_t(u)X_t'\theta_t\right] + \hat{Q}_{xx}^-(u)\hat{Q}_{x\varepsilon}(u)$, which will converge to a complex-valued function of u rather than a constant parameter θ_0. It implies that $\hat{A}(u)$ is proportional to the difference between two estimators that converge to the same limit only under $\mathbb{H}_0$. In that sense, $\hat{A}(u)$ is equivalent to a generalized Hausman test in the frequency domain. Unlike the existing generalized Hausman tests that compare the difference between two estimators at each time point t (e.g., Chen and Hong, 2012), our comparison is made at each nuisance parameter u. The DFT projects the time-varying feature of θ_t onto the frequency domain without loss of information. By this device, we can avoid the smoothed nonparametric estimation of θ_t and can test $\mathbb{H}_0$ consistently by checking the behavior of $\hat{A}(u)$.

There is a long history of spectral analysis in time series econometrics, e.g., see Granger and Hatanaka (1964), Hannan (1965, 1967), Engle (1974), Granger and Watson (1984), Choi and Phillips (1993), and Corbae et al. (2002). In recent years, DFT has been used to study nonstationary time series. For example, Yamamoto and Perron (2013) consider estimation and testing for abrupt structural breaks using band spectral regressions. Dwivedi and Subba Rao (2011) and Jentsch and Subba Rao (2015) use the DFT to test the weak stationarity of a time series. They show that the DFT is asymptotically uncorrelated at the canonical frequencies when the time series is weakly stationary. Even though the proposed tests are also based on DFT, the idea behind is quite different. Besides, testing structural change is not necessarily equivalent to testing the weak stationarity. The source of nonstationarity may be the instability of the regressors rather than structural changes in unknown regression coefficients. Therefore, their DFT-based tests cannot be directly applied to the current context.

2.3 Test Statistics

To ensure that the DFT-based tests are consistent against all alternatives, we consider the following Cramér-von Mises (CvM) and Kolmogorov-Smirnov (KS) type test statistics:

$$\hat{C} = T \int_{\mathbb{R}} \|\hat{A}(u)\|^2 W(u) \mathrm{d}u, \tag{2.5}$$

and

$$\hat{K} = T \sup_{u \in \mathbb{U}} \|\hat{A}(u)\|^2, \tag{2.6}$$

where $W(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is a nonnegative symmetric weighting function of u, and $\mathbb{U} = [-c, c]$ with $c > 0$ is a compact subset of $\mathbb{R}$.

The CvM type test involves the weighting function $W(u)$, which allows us to examine the behavior of $\hat{A}(u)$ at all $u \in \mathbb{R}$. Computing $\hat{C}$ generally requires numerical integration over $\mathbb{R}$. However, one can avoid numerical integration by adopting certain weighting functions to obtain a closed-form expression of $\hat{C}$. One choice is the following normal density function $N(0, \xi^2)$ considered by Hong et al. (2017): $W(u) = \frac{1}{\sqrt{2\pi\xi^2}} e^{-\frac{u^2}{2\xi^2}}$. Based on this weighting function, the test statistic (2.5) can be written as: $\hat{C} = T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} X_s' X_t \hat{\varepsilon}_s \hat{\varepsilon}_t e^{-\frac{2\pi^2 (s-t)^2 \xi^2}{T^2}}$, where $\xi > 0$ is the standard deviation that measures the dispersion of weights assigned around 0. In general, the results are insensitive to the choice of ξ. Another candidate is the Laplace density function: $W(u) = \frac{\lambda}{2} e^{-\lambda |u|}$, with which, we have $\hat{C} = T^{-1} \sum_{s=1}^{T} \sum_{t=1}^{T} X_s' X_t \hat{\varepsilon}_s \hat{\varepsilon}_t \left[\frac{\lambda^2 T^2}{\lambda^2 T^2 + 4\pi^2 (s-t)^2} \right]$, where $\lambda > 0$ determines the dispersion of weights assigned around 0. As $1/\lambda$ increases, more weights are assigned to higher frequencies.

For the KS type test, $T\|\hat{A}(u)\|^2$ is the sample periodogram of $X_t\hat{\varepsilon}_t$, and it is proportional to the quadratic difference between the WLS and the OLS estimators at each u. We note that the KS test does not require choosing a weighting function and is constructed for a compact set $\mathbb{U}$ rather than the whole real line $\mathbb{R}$. As a result, it may miss some information of structural instability occurring at a higher frequency. However, as we choosing a sufficiently large set of u, the KS could detect structural changes with probability closing to 1.

3 Asymptotic Properties

In this section, we derive the asymptotic null distribution of our tests and investigate their asymptotic power properties. Consider the following regularity conditions.

Assumption 3.1 (i) $\{X_t, \varepsilon_t\}_{t=1}^T$ is an absolutely regular process on $\mathbb{R}^{d+1}$ with mixing coefficient $\tilde{\beta}(j) = O(j^{-m})$ for some $m > \frac{r}{r-1}$ and $r > 2$, and $\sum_{j=1}^{\infty} j^2\tilde{\beta}(j) \leq \mathcal{M}$; (ii) $\max_t E(\|X_t\|^{2r}) < \mathcal{M}$ and $\max_t E(|\varepsilon_t|^{2r}) < \mathcal{M}$.

Assumption 3.2 $\{\varepsilon_t\}_{t=1}^T$ is a martingale difference sequence (MDS) such that $E(\varepsilon_t|I_{t-1}) = 0$ a.s. for all t, where I_{t-1} denotes a σ-field generated by $\{X_{t-1}, X_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

Assumption 3.3 $E(\varepsilon_t|X_t) = 0$ almost surely for all t.

Assumption 3.4 (i) $V_{xx}(u,v) \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^T E\left[X_t(u)X_t(v)^*\varepsilon_t^2\right]$ is finite and positive semi-definite for almost all $(u,v) \in \mathbb{R}^2$, and $E\left(X_tX_t'\varepsilon_t^2\right)$ is finite and positive definite for all t; (ii) $Q_{xx}(u)$ is finite and nonsingular for almost all $u \in \mathbb{R}$, and $E(X_tX_t')$ is nonsingular and finite for all t.

Assumption 3.5 The weighting function $W(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is nonnegative, symmetric, continuous, and integrable with $\int_{\mathbb{R}} |u|^2 W(u)\mathrm{d}u < \infty$.

Assumption 3.6 (i) Under $\mathbb{H}_A$, the time-varying coefficient θ_t satisfies that $T^{-1}\sum_{t=1}^T \|\theta_t\|^2 < \mathcal{M}$; (ii) $G_{xx}(u) \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^T E\left[X_t(u)X_t'\right]\theta_t$ exists for almost all $u \in \mathbb{R}$.

Assumption 3.1(i) allows the marginal distribution of X_t to change over time but restricts its temporal dependence to be weak uniformly over time. Let $-\infty \le J_1 \le J_2 \le \infty$, and $\mathcal{F}_{J_1}^{J_2}$ be the σ-field of $\{X_t', \varepsilon_t\}$ such that $J_1 \le t \le J_2$. Then for each $m \ge 1$, the uniform absolutely regular mixing coefficient $\tilde{\beta}(m)$ is defined as the following: $\tilde{\beta}(m) = \sup_{t\in\mathbb{Z}} \beta(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+m}^{\infty})$, where $\beta(m)$ is the conventional β-mixing coefficient (e.g., Bradley, 2005). Intuitively, even though the marginal distribution of X_t or ε_t can vary over time, the possible time-varying temporary dependence at each time t is bounded by the uniform mixing rate. It allows us to establish the asymptotic results of our test statistics. The structural changes in X_t can be either smooth or abrupt. It implies that our tests are robust to unknown structural change in the covariates. We note that Assumption 3.1(i) excludes the case that X_t is a unit root or near unit root process. Such cases will be investigated in subsequent studies. Assumption 3.1(ii) imposes mild moment conditions on X_t and ε_t, which allows the higher moments of ε_t to vary over time. Assumption 3.2 restricts the error term ε_t to be serially uncorrelated. This is only for the simplicity of delivering our main result. We can allow ε_t to exhibit serial correlation and adopt a suitable resampling approach that accounts for the serial dependence. Assumption 3.3 excludes endogenous covariates, which will be relaxed in Section 4. It implies that X_t is contemporaneously exogenous, which guarantees the consistency of the OLS under the null hypothesis. Note that when X_t contains the lagged values of Y_t, Assumption 3.3 still holds given that ε_t is an MDS. Assumption 3.4 provides sufficient conditions to ensure that the covariance kernel of the asymptotic distribution exists and is well-defined. It also allows ε_t to be conditionally heteroskedastic, which is common in time series data. Assumption 3.5 imposes some mild conditions on the weighting function, which guarantees that the integral in (2.5) exists. Compared to the existing literature, e.g., Chen and Hong (2012), Kristensen (2012), and Cai et al. (2015), we do not impose any smoothness assumption on the unknown parameter θ_t. We only need it to be square-summable, which is stated in Assumption 3.6(i). Besides, Assumption 3.6(ii) ensures that the limit of the OLS estimator exists under $\mathbb{H}_A$.

3.1 Asymptotic Null Distributions

Before we state the asymptotic null distributions of our test statistics, we derive the limiting results for the DFT $\hat{A}(u)$.

Proposition 1 *Suppose Assumptions 3.1–3.4 and* $\mathbb{H}_0$ *hold. Let* $\mathbb{U} = [-c, c]$ *be a compact subset of* $\mathbb{R}$ *with* $c > 0$*, then as* $T \to \infty$*,* $\sqrt{T}\hat{A}(u) \Rightarrow S(u)$ *on* $\mathbb{U}$*, where* $S(u)$ *is a mean-zero complex-valued Gaussian process with covariance kernel* $\mathcal{K}(u,v) = V_{xx}(u,v) - Q_{xx}(u)Q_{xx}^{-1}V_{xx}(0,v) - V_{xx}(u,0)Q_{xx}^{-1}Q_{xx}(v)^* + Q_{xx}(u)Q_{xx}^{-1}V_{xx}(0,0)Q_{xx}^{-1}Q_{xx}(v)^*$.

Proposition 1 implies that the DFT weakly converges to a complex-valued Gaussian process in the frequency domain under $\mathbb{H}_0$. Intuitively, when there is no structural change, the DFT only contains the estimation uncertainty and the error term ε_t. Due to the orthogonality condition in Assumption 3.3, $\hat{A}(u)$ converges to 0 for each fixed $u \in \mathbb{R}$ at the parametric rate. Furthermore, if X_t and ε_t are jointly weakly stationary, the covariance kernel can be simplified as $\mathcal{K}(u,v) = V_{xx}(0,0)\Gamma(u,v)$, where $\Gamma(u,v) \equiv \int_0^1 e^{\mathbf{i}2\pi(u-v)\tau}\mathrm{d}\tau - \int_0^1 e^{\mathbf{i}2\pi u\tau}\mathrm{d}\tau \int_0^1 e^{-\mathbf{i}2\pi v\tau}\mathrm{d}\tau$ is a pseudo-covariance in the sense that τ follows the $U[0,1]$ distribution. Intuitively, the covariance kernel contains two components. The first component $V_{xx}(0,0) = E(X_tX_t'\varepsilon_t^2)$ is the asymptotic variance of $T^{-1/2}\sum_{t=1}^T X_t\varepsilon_t$, which measures the degree of uncertainty introduced by the OLS estimation. The other component can be viewed as the noise introduced by the discrete Fourier transform. When X_t and ε_t are nonstationary, these two components intertwine with each other.

After stating the asymptotic property of $\hat{A}(u)$, we now derive the asymptotic null distributions of the test statistics $\hat{C}$ and $\hat{K}$.

Theorem 1 *Suppose Assumptions 3.1–3.5 and* $\mathbb{H}_0$ *hold. Let* $\mathbb{U} = [-c, c]$ *be a compact subset of* $\mathbb{R}$ *with* $c > 0$*, then as* $T \to \infty$*,* $\hat{C} \xrightarrow{d} \int_{\mathbb{R}} \|S(u)\|^2 W(u)\mathrm{d}u$*, and* $\hat{K} \xrightarrow{d} \sup_{u\in\mathbb{U}} \|S(u)\|^2$*, where* $S(u)$ *is defined in Proposition 1.*

Given Proposition 1, the asymptotic distribution of $\hat{K}$ follows from the continuous mapping theorem. We note that for $\hat{K}$, we have to restrict u to be within a certain compact subset of $\mathbb{R}$. Because when $|u|$ grows to infinity, the weak convergence of $\sqrt{T}\hat{A}(u)$ does not hold. $\hat{K}$ may be inconsistent against certain alternatives when $\mathbb{U}$ fails to contain the signal of structural change. However, $\hat{K}$ is flexible because it allows the practitioner to examine whether structural changes occur at certain frequency regions. $\hat{C}$ can examine the deviation of $\sqrt{T}\hat{A}(u)$ from a zero spectrum for all $u \in \mathbb{R}$ since the weighting function $W(u)$ assigns small but positive weights to large u's. Proposition 1 establishes the weak convergence of $\sqrt{T}\hat{A}(u)$ on a compact subset of $\mathbb{R}$. For some sufficiently large u, the weak convergence may not hold. However, by introducing the weighting function $W(\cdot)$, the contribution of integrating over the part of $\mathbb{R}$ where the weak convergence does not hold is asymptotically negligible.

Compared to the existing tests for abrupt structural breaks such as Andrews (1993) and Bai and Perron (1998), our test does not need trimming the observations in the boundary regions of the sample. Unlike the tests for smooth structural changes such as Chen and Hong (2012) and Cai et al. (2015), we avoid smoothed nonparametric estimation. Therefore, we do not need to choosing such smoothing parameter as bandwidth or the trimming parameters. That is appealing in practice because there have been no criteria to choose the optimal bandwidth for the nonparametric smoothing tests of Chen and Hong (2012), Zhang and Wu (2012), and Cai et al. (2015) or the trimming parameter for the parametric tests of Andrews (1993) and Bai and Perron (1998).

3.2 Asymptotic Power

In this subsection, we show that the DFT $\hat{A}(u)$ converges to a non-zero spectrum under $\mathbb{H}_A$ in the frequency domain. It is essential for the consistency of our test. Let $M_t(u) \equiv X_t(u) - \tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}X_t$. The following proposition provides the limiting result of the DFT under $\mathbb{H}_A$.

Proposition 2 *Suppose Assumptions 3.1–3.6 and $\mathbb{H}_A$ hold. Let $\mathbb{U} = [-c, c]$ be a compact subset of $\mathbb{R}$ with $c > 0$, then as $T \to \infty$ $\sup_{u\in\mathbb{U}} \|\hat{A}(u) - A(u)\| \xrightarrow{p} 0$, where $A(u) \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E[M_t(u)X_t']\theta_t = G_{xx}(u) - Q_{xx}(u)Q_{xx}^{-1}G_{xx}(0)$.*

Proposition 2 shows that $\hat{A}(u)$ converges to a nonzero spectrum in the frequency domain under $\mathbb{H}_A$. In particular, if X_t is weakly stationary, and $\theta_t = \theta(t/T)$ is a smooth function of the rescaled time index $t/T \in (0,1]$, then $A(u) = Q_{xx}\widetilde{\text{cov}}\left[\theta(\tau), e^{\mathbf{i}2\pi u\tau}\right]$, where $\widetilde{\text{cov}}[\theta(\tau), e^{\mathbf{i}2\pi u\tau}] \equiv \int_0^1 \theta(\tau)e^{\mathbf{i}2\pi u\tau}\text{d}\tau - \int_0^1 \theta(\tau)\text{d}\tau \int_0^1 e^{\mathbf{i}2\pi u\tau}\text{d}\tau$ is a pseudo-covariance in the sense that τ follows the $U[0,1]$ distribution. Proposition 2 shows that $\hat{A}(u)$ is equivalent to the sample analog of a pseudo-covariance between θ_t and $e^{\mathbf{i}2\pi ut/T}$ under $\mathbb{H}_A$. The result also explains why we construct $\hat{A}(u)$ by considering the DFT of $X_t\hat{\varepsilon}_t$ rather than $\hat{\varepsilon}_t$. When X_t is weakly stationary such that $E(X_t) = 0$, then our test will have no power if we only consider the Fourier transform of $\hat{\varepsilon}_t$. For that reason, we construct $\hat{A}(u)$ using the Fourier transform of $X_t\hat{\varepsilon}_t$ to ensure that it converges to a nonzero spectrum under $\mathbb{H}_A$.

Next, we show the asymptotic power of our test statistics.

Theorem 2 *Suppose Assumptions 3.1–3.6 hold. Then for any sequence of nonstochastic constants* $\{c_T = o(T)\}$, *as* $T \to \infty$, $P(\hat{C} > c_T) \to 1$, *and* $P(\hat{K} > c_T) \to 1$, *under* $\mathbb{H}_A$.

Theorem 2 shows our tests $\hat{C}$ and $\hat{K}$ are consistent against any alternatives to $\mathbb{H}_0$ at any significance level. Note that we do not impose any restrictive conditions on the form of $\mathbb{H}_A$. Therefore, $\hat{C}$ and $\hat{K}$ are powerful in capturing various forms of instability in coefficients, including smooth structural changes and abrupt structural breaks with unknown break dates.

To gain more insight into the asymptotic power property of our tests, we now consider the following local alternatives

$$\mathbb{H}_A(\Delta_T): \quad \theta_t = \theta_0 + \Delta_T\phi_t,$$

where θ_0 is a constant coefficient vector, and $\phi_t = \phi(t/T)$ with $\phi(\cdot) : [0,1] \to \mathbb{R}^d$ being a nonrandom function of scaled time t/T, such that $\int_0^1 |\phi(\tau)|^2\text{d}\tau \le \mathcal{M}$, and $\phi(\cdot) \neq 0$ on a set with nonzero Borel measure. $\Delta_T\phi_t$ characterizes the departure of the time-varying coefficient θ_t from θ_0, and the rate Δ_T is the speed at which the deviation vanishes to zero as the sample size $T \to \infty$. In particular, since we do not impose any smooth restriction on $\phi(\cdot)$, such local alternatives allow for smooth structural changes, abrupt structural breaks, or a mixture of both.

Theorem 3 *Suppose Assumptions 3.1–3.6 and $\mathbb{H}_A(\Delta_T)$ hold with $\Delta_T = T^{-1/2}$. Let $\mathbb{U} = [-c, c]$ be a compact subset of $\mathbb{R}$ with $c > 0$. Then as $T \to \infty$, $\hat{C} \xrightarrow{d} \int_{\mathbb{R}} \|\kappa(u) + S(u)\|^2 W(u)\mathrm{d}u$, and $\hat{K} \xrightarrow{d} \sup_{u\in\mathbb{U}} \|\kappa(u) + S(u)\|^2$, where $S(u)$ is defined in Proposition 1 and $\kappa(u) \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E[X_t(u)X_t'](\phi_t - \tilde{\phi})$ with $\tilde{\phi} = \tilde{Q}_{xx}^{-1}[T^{-1}\sum_{t=1}^{T} E(X_tX_t')\phi_t]$.*

Theorem 3 shows that our tests $\hat{C}$ and $\hat{K}$ have nontrivial power against a class of local alternatives with the parametric rate $\Delta_T = T^{-1/2}$. In terms of Pitman's criterion, it is asymptotically more efficient than the smoothed nonparametric tests of Chen and Hong (2012), Zhang and Wu (2012), and Cai et al. (2015), which can only detect the local alternatives with a rate $\Delta_T = T^{-1/2}h^{-1/4}$, where h is a bandwidth. This is an advantage of DFT, which avoids smoothed nonparametric regression. The only price we have to pay is that the asymptotic distributions of our tests $\hat{C}$ and $\hat{K}$ are not pivotal. However, we can use resampling methods to obtain critical values in finite samples. Moreover, although the nonparametric tests are asymptotically pivotal under the null hypothesis, the results based on the asymptotic critical values tend to be sensitive to the choice of bandwidth. As a result, they still rely on suitable bootstrap methods in practice. Moreover, our tests do not require trimming of the data. Thus, we can detect structural changes that occur close to the starting and ending periods of the sample. This is rather appealing because we have no prior information about the possible break dates in practice. In contrast, Andrews' (1993) supremum and Bai and Perron's (1998) double maximum tests do not have uniform power for all $t/T \in (0, 1)$.

3.3 Bootstrap Versions of Our Tests

As shown in Theorem 1, the asymptotical null distributions of $\hat{C}$ and $\hat{K}$ are not pivotal, as they depend on the unknown data generating process. However, we can use resampling methods to obtain critical values in finite sample. Here, we follow Hansen (1996) to propose the following multiplier bootstrap procedure: Step (i).

Use the sample $\{Y_t, X_t'\}_{t=1}^T$ to estimate (2.1) via the OLS, and compute test statistics $\hat{C}$ and $\hat{K}$; Step (ii). Generate $i.i.d.N(0,1)$ random variables $\{v_t^{(b)}\}_{t=1}^T$ and compute $\hat{K}_b$ and $\hat{C}_b$ using $\hat{A}_B^{(b)}(u) = T^{-1}\sum_{t=1}^T \hat{M}_t(u)\hat{\varepsilon}_t v_t^{(b)}$; Step (iii). Repeat Step (ii) for B times to obtain B bootstrap test statistics $\{\hat{C}_b, \hat{K}_b\}_{b=1}^B$; Step (iv). Compute the bootstrap p-values for $\hat{C}$ and $\hat{K}$ respectively, with $\hat{p}_{B,T}^C = B^{-1}\sum_{b=1}^B \mathbb{I}(\hat{C}_b > \hat{C})$ and $\hat{p}_{B,T}^K = B^{-1}\sum_{b=1}^B \mathbb{I}(\hat{K}_b > \hat{K})$, where $\hat{p}_{B,T}^C$ and $\hat{p}_{B,T}^K$ are the bootstrapped p-values for $\hat{C}$ and $\hat{K}$ via resampling for B times, and $\mathbb{I}(\cdot)$ is the indicator function.

Next, we show that the bootstrapped p-values are consistent for the asymptotic p-values. Let $F^C(\cdot)$ and $F^K(\cdot)$ denote the distribution functions of $\int_{\mathbb{R}} \|S(u)\|^2 W(u)\mathrm{d}u$ and $\sup_{u\in\mathbb{U}} \|S(u)\|^2$, respectively. Define $p_T^C \equiv 1 - F^C(\hat{C})$ and $p_T^K \equiv 1 - F^K(\hat{K})$ be the asymptotic p-values, such that $\lim_{T\to\infty} P\{p_T^j \leq \alpha | \mathbb{H}_0\} = \alpha$, for $j = C, K$. Conditional on the sample $\{Y_t, X_t'\}_{t=1}^T$, let $F_T^C(\cdot)$ and $F_T^K(\cdot)$ be the conditional distribution functions of $\hat{C}$ and $\hat{K}$, respectively. Define $\hat{p}_T^C = 1 - F_T^C(\hat{C})$ and $\hat{p}_T^K = 1 - F_T^K(\hat{K})$ to be the bootstrapped p-values, where $\hat{F}_T^C(\cdot)$ and $\hat{F}_T^K(\cdot)$ are generated by replacing $\hat{A}(u)$ with $\hat{A}_B(u) = T^{-1}\sum_{t=1}^T \hat{M}_t(u)\hat{\varepsilon}_t v_t$, where $\{v_t\}_{t=1}^T$ is $i.i.d.N(0,1)$. Then we can show the consistency of the resampling method by the following theorem.

Theorem 4 *Under Assumptions 3.1–3.5 and $\mathbb{H}_0$, $\hat{p}_T^C \Rightarrow p^C$ and $\hat{p}_T^K \Rightarrow p^K$, where p^C and p^K are the true p-values of the test statistics $\hat{C}$ and $\hat{K}$, respectively. Under Assumptions 3.1–3.6 and $\mathbb{H}_A$, $P(\hat{C} > \hat{C}_b) \to 1$ and $P(\hat{K} > \hat{K}_b) \to 1$.*

We note that the proposed resampling approach requires the error term to be serially uncorrelated. However, the DFT test is applicable to serially correlated errors. A different resampling approach that accounts for the serial dependence is needed to obtain the asymptotic critical values.

4 Extensions to Endogenous Regression

In the previous section, we assume that the regressors are exogenous. However, endogeneity is a widespread phenomenon in many time series applications. It may arise due to simultaneous equations, measurement errors, or omitted variables. In this section, we extend our tests to allow for endogenous regressors.

4.1 Test Statistics

Suppose there exists a set of instruments $Z_t \in \mathbb{R}^l$ such that $E(\varepsilon_t|Z_t) = 0$ a.s., and $E(X_t Z_t') \neq 0$, where $l \geq d$. Following analogous notations in Section 2, we define $\hat{Q}_{xz}(u) \equiv T^{-1}\sum_{t=1}^{T} X_t(u)Z_t'$, $\tilde{Q}_{xz}(u) = E\hat{Q}_{xz}(u)$, and $Q_{xz}(u) = \lim_{T\to\infty}\tilde{Q}_{xz}(u)$. Analogously, we let $\hat{Q}_{xz}$ to represent $\hat{Q}_{xz}(0)$. To test $\mathbb{H}_0$, we now consider the following complex-valued empirical process:

$$\hat{A}^{IV}(u) \;=\; \frac{1}{T}\sum_{t=1}^{T}\hat{X}_t\hat{\varepsilon}_t e^{\mathbf{i}2\pi ut/T}.$$

We denote $\hat{X}_t \in \mathbb{R}^d$ as the fitted value from the first stage regression of X_t on the instrumental variables Z_t, such that $\hat{X}_t = \hat{\gamma}' Z_t$ with $\hat{\gamma} = \hat{Q}_{zz}^{-1}\hat{Q}_{xz}$, and $\hat{\varepsilon}_t = Y_t - X_t'\hat{\theta}_{2sls}$ as the estimated residuals with $\hat{\theta}_{2sls} = \hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{Q}_{\hat{x}y}$ being the Two-Stage Least Squares (2SLS) estimator. Furthermore, we let $\tilde{\gamma} = \tilde{Q}_{zz}^{-1}\tilde{Q}_{xz}$ and $\gamma = \lim_{T\to\infty}\tilde{\gamma}$.

Similar to the case with purely exogenous regressors, the estimated residuals from the 2SLS can also be decomposed into three components: the disturbance ε_t, the estimation uncertainty, and the local feature of θ_t over time. Under $\mathbb{H}_0 : \theta_t = \theta_0$ for all t, $\hat{\theta}_{2sls}$ is consistent for θ_0. Then, the third component will be identically zero. While under $\mathbb{H}_A$, $\hat{\theta}_{2sls}$ fails to capture the time-varying feature of θ_t, and such information will be contained in the estimated residuals. Therefore, the DFT will converge to a nonzero spectrum in the frequency domain. In particular, if X_t and Z_t are jointly weakly stationary, and $\theta_t \equiv \theta(t/T)$ is a smooth function of the rescaled time index $t/T \in (0, 1]$, then the probability limit of the DFT under $\mathbb{H}_A$ is proportional to the pseudo-covariance of θ_t and $e^{\mathbf{i}2\pi ut/T}$ in the sense that t/T follows the $U[0, 1]$ distribution.

We note that the true relationship between X_t and Z_t may be unstable such that $X_t = \gamma_t' Z_t + \nu_t$, where γ_t is an unknown coefficient matrix that varies over time. If that is the case, then $\hat{\gamma}$ will be inconsistent for γ_t. However, its impact on our analysis is trivial since our goal is not to estimate γ_t consistently but use $\hat{X}_t$ as a proxy to obtain the 2SLS estimator. The fitted value $\hat{X}_t$ from the first stage regression serves as the regressor in the second stage. It has the same role as X_t in the case where all covariates are exogenous. Thus $\hat{X}_t$ can be nonstationary such that its distribution may change smoothly or break abruptly. As long as X_t and Z_t are not orthogonal to each other, we can achieve the consistency of the 2SLS estimators for θ_0 under $\mathbb{H}_0$. Moreover, we do not restrict either X_t or Z_t to be weakly stationary. The tests by Hall et al. (2012) and Perron and Yamamoto (2014) rely on comparing consistent estimates for θ_t using the 2SLS in subsamples. Therefore, they need to investigate the stability of the reduced form. Our frequency domain-based approach can avoid this issue and is robust to structural changes in covariates and instruments.

Analogously, $\hat{A}^{IV}(u)$ can also be interpreted as a generalized Hausman test in the frequency domain. Letting $\hat{X}_t(u) \equiv \hat{X}_t e^{\mathbf{i}2\pi ut/T} = \hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t e^{\mathbf{i}2\pi ut/T}$, we can rewrite the DFT as $\hat{A}^{IV}(u) = \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left[\hat{\theta}^{IV}(u) - \hat{\theta}_{2sls}\right]$, where we define the following class of linear IV estimators for θ_0:

$$\hat{\theta}^{IV}(u) \equiv \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\right]^{-} \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t Y_t e^{\mathbf{i}2\pi ut/T}\right).$$

Compared to the WLS estimator defined in Section 2, $\hat{\theta}^{IV}$ contains an additional projection to the space spanned by Z_t. In particular, when $u = 0$, $\hat{\theta}^{IV}(0)$ is exactly the 2SLS estimator. Under $\mathbb{H}_0$, we can show $\hat{\theta}^{IV}(u) \xrightarrow{p} \theta_0$ for almost all u. It implies that $\hat{\theta}^{IV}(u)$ is consistent for θ_0 under $\mathbb{H}_0$. Furthermore, it can be viewed as a general class of estimators that contains the 2SLS estimator as a special case. Following analogous arguments, $\hat{\theta}^{IV}(u)$ converges to the same probability limit under $\mathbb{H}_0 : \theta_t = \theta_0$.

To examine the behavior of $\hat{A}^{IV}(u)$ at each u, we consider the following CvM and KS type test statistics: $\hat{C}^{IV} = T\int_{\mathbb{R}} \|\hat{A}^{IV}(u)\|^2 W(u)\mathrm{d}u$, and $\hat{K}^{IV} = T\sup_{u\in\mathbb{U}} \|\hat{A}^{IV}(u)\|^2$, where $\mathbb{U} = [-c, c]$ with $c > 0$ is a compact subset of $\mathbb{R}$, and $W(\cdot)$ is a weighting function that satisfies Assumption 3.5.

4.2 Asymptotic Properties

To provide the asymptotic results for $\hat{C}^{IV}$ and $\hat{K}^{IV}$, we make the following regularity conditions.

Assumption 4.1 (i) $\{X_t', Z_t', \varepsilon_t\}_{t=1}^{'T}$ is an absolutely regular process on $\mathbb{R}^{d+l+1}$ with mixing coefficient $\tilde{\beta}(j) = O(j^{-m})$ for some $m > \frac{r}{r-1}$ and $r > 2$, and $\sum_{j=1}^{\infty} j^2 \tilde{\beta}(j) \leq \mathcal{M}$; (ii) $\max_t E(\|Z_t\|^{2r}) < \mathcal{M}$, $\max_t E(\|X_t\|^{2r}) < \mathcal{M}$, and $\max_t E(|\varepsilon_t|^{2r}) < \mathcal{M}$.

Assumption 4.2 $\{\varepsilon_t\}$ is an MDS such that $E(\varepsilon_t|I_{t-1}) = 0$ a.s., where I_{t-1} is a σ-field generated by $\{Z_{t-1}, Z_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.

Assumption 4.3 $E(\varepsilon_t|X_t) \neq 0$ for some t and $E(Z_t\varepsilon_t) = 0$ a.s. for all t.

Assumption 4.4 (i) $V_{zz}(u, v) \equiv \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E\left[Z_t(u)Z_t(v)^* \varepsilon_t^2\right]$ is finite and positive semi-definite for almost all $(u, v) \in \mathbb{R}^2$, and $E\left(Z_tZ_t'\varepsilon_t^2\right)$ is finite and positive definite for all t; (ii) Q_{zz} is finite and nonsingular, and $E(Z_tZ_t')$ is finite and nonsingular for all t; (iii) Q_{zx} is finite and of full rank for almost all $u \in \mathbb{R}$, and $E(Z_tX_t')$ is finite and of full rank for all t.

Assumption 4.5 $G_{zx}(u) \equiv \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E\left[Z_t(u)X_t'\right] \theta_t$ exists for almost all $u \in \mathbb{R}$ under $\mathbb{H}_A$.

Similar to Assumption 3.1(i), Assumption 4.1(i) allows both X_t and Z_t to have smooth structural changes or abrupt structural breaks. The temporal dependence has been controlled by the mixing coefficient for possibly nonstationary time series. Assumption 4.1(ii) imposes regular moment conditions on X_t, Z_t, and ε_t in an analogous way as Assumption 3.1(ii). Assumption 4.2 restricts the disturbance to be serially uncorrelated. Assumption 4.3 states the existence of endogeneity and the validity of instrumental variable Z_t. Assumption 4.4 ensures that the covariance kernel exists and is well-defined. Assumption 4.5 guarantees that the limit of the 2SLS estimator exists under $\mathbb{H}_A$.

We now investigate the asymptotic null distributions of $\hat{C}^{IV}$ and $\hat{K}^{IV}$. The following proposition provides the limiting results for the DFT $\hat{A}^{IV}(u)$.

Proposition 3 *Suppose Assumptions 4.1–4.4 and $\mathbb{H}_0$ hold. Let $\mathbb{U} = [-c, c]$ be a compact subset of $\mathbb{R}$ with $c > 0$, then as $T \to \infty$, $\sqrt{T}\hat{A}^{IV}(u) \Rightarrow S^{IV}(u)$*

on $\mathbb{U}$*, where* $S^{IV}(u)$ *is a mean-zero complex-valued Gaussian process with covariance kernel* $\mathcal{K}^{IV}(u_1, u_2) = \gamma' V_{zz}(u_1, u_2)\gamma - \gamma' Q_{zx}(u_1)(\gamma' Q_{zx})^{-1}\gamma' V_{zz}(0, u_2)\gamma - \gamma' V_{zz}(u_1, 0)\gamma(Q_{xz}\gamma)^{-1}[\gamma' Q_{xz}(u_2)]^* + \gamma' Q_{zx}(u_1)(\gamma' Q_{zx})^{-1}\gamma' V_{zz}(0,0)\gamma(Q_{xz}\gamma)^{-1}[\gamma' Q_{xz}(u_2)]^*$.

Similar to Proposition 1, $\sqrt{T}\hat{A}^{IV}(u)$ weakly converges to a complex-valued Gaussian process in the frequency domain under $\mathbb{H}_0$. When endogenous covariates exist, we need to first project X_t onto Z_t, which introduces additional estimation uncertainty to the variance of the DFT. Furthermore, if X_t, Z_t, and ε_t are jointly weakly stationary, the covariance kernel can be simplified as $\mathcal{K}^{IV}(u, v) = [\gamma' V_{zz}(0, 0)\gamma]\,\Gamma(u, v)$, where $\gamma = E(Z_t Z_t')^{-1}E(Z_t X_t')$. Similar to the case with purely exogenous covariates, the covariance kernel is equal to the variance of the score function in the IV regression multiplied by a pseudo-covariance introduced by the Fourier transform.

The following theorem provides the asymptotic distribution of our test statistics in the case of endogenous covariates.

Theorem 5 *Suppose Assumptions 4.1–4.4, 3.6, and* $\mathbb{H}_0$ *hold. Let* $\mathbb{U} = [-c, c]$ *be a compact subset of* $\mathbb{R}$ *with* $c > 0$*, then as* $T \to \infty$*,* $\hat{C}^{IV} \xrightarrow{d} \int_{\mathbb{R}} \|S^{IV}(u)\|^2 W(u)\mathrm{d}u$*, and* $\hat{K}^{IV} \xrightarrow{d} \sup_{u\in\mathbb{U}} \|S^{IV}(u)\|^2$.

Theorem 5 provides the asymptotic distributions of $\hat{C}^{IV}$ and $\hat{K}^{IV}$ under $\mathbb{H}_0$. Similar to Theorem 1, the asymptotic distribution of $\hat{K}^{IV}$ follows directly from the continuous mapping theorem. The choice of $\mathbb{U}$ allows us to detect structural changes in a certain range of frequencies. While $\hat{C}^{IV}$ can examine the behavior of the DFT at all $u \in \mathbb{R}$, and its asymptotic distribution holds due to the conditions on the weighting function $W(u)$.

The following proposition states the asymptotic property of $\hat{A}^{IV}(u)$ under $\mathbb{H}_A$.

Proposition 4 *Suppose Assumptions 4.1–4.5 and 3.6(i) hold. Let* $\mathbb{U} = [-c, c]$ *be a compact subset of* $\mathbb{R}$ *with* $c > 0$*, then under* $\mathbb{H}_A$*, as* $T \to \infty$, $\sup_{u\in\mathbb{U}} \|\hat{A}^{IV}(u) - A^{IV}(u)\| \xrightarrow{p} 0$*, where* $A^{IV}(u) \equiv \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E[M_t^{IV}(u)X_t']\theta_t = \gamma' G_{zx}(u) - \gamma' Q_{zx}(u)(\gamma' Q_{zx})^{-1}\gamma' G_{zx}(0)$ *with* $M_t^{IV}(u) = \tilde{\gamma}' Z_t(u) - \tilde{\gamma}' \tilde{Q}_{zx}(u)(\tilde{\gamma}\tilde{Q}_{zx})^{-1}\tilde{\gamma} Z_t$.

Proposition 4 shows that $\hat{A}^{IV}(u)$ converges to a nonzero spectrum in the frequency domain, which is a weighted average of the unknown parameter θ_t. If we let X_t and Z_t be jointly weakly stationary, and $\theta_t \equiv \theta(\frac{t}{T})$ be a smooth function of the rescaled time index $t/T \in (0,1]$, then the limit of $\hat{A}^{IV}(u)$ is $A^{IV}(u) = \gamma' Q_{zx}\widetilde{\text{cov}}[\theta(\tau), e^{\mathbf{i}2\pi u\tau}]$, where $\widetilde{\text{cov}}[\theta(\tau), e^{\mathbf{i}2\pi u\tau}]$ is a pseudo covariance between θ_t and $e^{\mathbf{i}2\pi ut/T}$ defined in Section 3.

The next theorem provides the asymptotic power property of $\hat{C}^{IV}$ and $\hat{K}^{IV}$.

Theorem 6 *Suppose Assumptions 4.1–4.5, 3.5, and 3.6(i) hold. (i) Under $\mathbb{H}_A$, for any sequence of non-stochastic constants $\{c_T = o(T)\}$, as $T \to \infty$, $P(\hat{C}^{IV} > c_T) \to 1$ and $P(\hat{K}^{IV} > c_T) \to 1$. (ii) Under $\mathbb{H}_A(\Delta_T)$ with $\Delta_T = T^{-1/2}$, as $T \to \infty$, $\hat{K}^{IV} \xrightarrow{d} \sup_{u\in\mathbb{U}} \left\|\kappa^{IV}(u) + S^{IV}(u)\right\|^2$ and $\hat{C}^{IV} \xrightarrow{d} \int_{\mathbb{R}} \left\|\kappa^{IV}(u) + S^{IV}(u)\right\|^2 W(u)\mathrm{d}u$, where $\kappa^{IV}(u) = \lim_{T\to\infty} \tilde{\gamma}'[T^{-1}\sum_{t=1}^T E[Z_t(u)X_t'](\phi_t - \tilde{\phi}^{IV})]$, with $\tilde{\phi}^{IV} = (\tilde{\gamma}'\tilde{Q}_{zx})^{-1}$ $\tilde{\gamma}'[T^{-1}\sum_{t=1}^T E(Z_tX_t'\phi_t)]$.*

Theorem 6(i) and 6(ii) provide the global and local power property of our test statistics respectively. It implies that our tests $\hat{C}^{IV}$ and $\hat{K}^{IV}$ have nontrivial power against a class of local alternatives with a rate $\Delta_T = T^{-1/2}$. This rate is faster than the nonparametric rate $T^{-1/2}h^{-1/4}$ in Chen (2015). Thus, our test is asymptotically more powerful.

Since the asymptotic null distributions of our tests are not pivotal, we need resampling methods to obtain critical values. Like in the OLS case, we follow Hansen (1996) and propose the following resampling procedures. Step (i). Use the sample $\{Y_t, X_t', Z_t'\}_{t=1}^T$ to estimate (2.1) via the 2SLS, and compute $\hat{C}^{IV}$ and $\hat{K}^{IV}$; Step (ii). Generate $i.i.d.N(0,1)$ random variables $\{v_t^{(b)}\}_{t=1}^T$ and compute $\hat{C}_b^{IV}$ and $\hat{K}_b^{IV}$ using $\hat{A}_B^{IV,(b)}(u) = T^{-1}\sum_{t=1}^T \hat{M}_t^{IV}(u)\hat{\varepsilon}_t v_t^{(b)}$, where $\hat{M}_t^{IV}(u) = \hat{X}_t e^{\mathbf{i}2\pi ut/T} - \hat{Q}_{\hat{x}x}(u)\hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{X}_t$ and $\hat{\varepsilon}_t$ is the estimated residuals from the 2SLS; Step (iii). Repeat Step (ii) for B times to obtain B bootstrap test statistics $\{\hat{C}_b^{IV}, \hat{K}_b^{IV}\}_{b=1}^B$; Step (iv). Compute the bootstrap p-values for $\hat{C}^{IV}$ and $\hat{K}^{IV}$ respectively, with $p_{B,T}^{C,IV} = B^{-1}\sum_{b=1}^B \mathbb{I}(\hat{C}_b^{IV} > \hat{C}^{IV})$ and $p_{B,T}^{K,IV} = B^{-1}\sum_{b=1}^B \mathbb{I}(\hat{K}_b^{IV} > \hat{K}^{IV})$.

We note that $p_{B,T}^{C,IV}$ and $p_{B,T}^{K,IV}$ are the bootstrapped p-values for $\hat{C}^{IV}$ and $\hat{K}^{IV}$ via resampling for B times. By analogous proof for Theorem 4, we can show that they are consistent for the true p-values.

5 Monte Carlo Simulations

In this section, we study the finite sample performance of the proposed tests via Monte Carlo simulations. We consider time series regressions with exogenous and endogenous covariates, respectively. For the case of exogenous covariates, we compare our tests with Andrews' (1993) supremum LM test, Bai and Perron's (1998) double maximum test, and Chen and Hong's (2012) generalized Hausman test. For the case of endogenous covariates, we compare our test with Chen's (2015) generalized Hausman test and Hall et al.'s (2012) double maximum F-test.

5.1 Exogenous Covariates

In this subsection, we examine the performance of our tests with purely exogenous covariates. To examine the size and power, we consider the following regressions:
DGP.S1 (No structural change): $Y_t = 1 + 0.5X_t + \varepsilon_t$;
DGP.P1 (A single structural break): $Y_t = \begin{cases} 1 + 0.5X_t + \varepsilon_t, & \text{if } t \leq 0.3T, \\ 1.2 + X_t + \varepsilon_t, & \text{otherwise;} \end{cases}$
DGP.P2 (Multiple structural breaks): $Y_t = \begin{cases} 1 + 0.1X_t + \varepsilon_t, & \text{if } 0.1T \leq t \leq 0.3T, \\ 1 + X_t + \varepsilon_t, & \text{if } 0.7T \leq t \leq 0.9T, \\ 1 + 0.5X_t + \varepsilon_t, & \text{otherwise;} \end{cases}$
DGP.P3 (Monotonic smooth structural change): $Y_t = 1.5 + \theta(t/T)X_t + \varepsilon_t$, where $\theta(\tau) = 0.5 + 0.5\{1 + e^{-20(\tau-0.5)}\}^{-1}$;
DGP.P4 (Non-monotonic smooth structural change): $Y_t = 1+\theta(t/T)X_t+\varepsilon_t$, where $\theta(\tau) = 0.5 + 1.5e^{-3(\tau-0.5)^2}$.

To examine the robustness of tests, we follow Chen and Hong (2012) to consider three cases for the error term $\{\varepsilon_t\}$. (i) i.i.d. case: $\varepsilon_t \sim i.i.d.N(0, 1)$; (ii) ARCH case: $\varepsilon_t = \sqrt{h_t}\nu_t$, $h_t = 0.2 + 0.5\varepsilon_{t-1}^2$, $\nu_t \sim i.i.d.N(0, 1)$; (iii) Heteroskedasticity case: $\varepsilon_t = \sqrt{h_t}\nu_t$, $h_t = 0.2 + 0.5X_t^2$, $\nu_t \sim i.i.d.N(0, 1)$. In addition, we mention that our tests are robust to structural changes in X_t. To show the finite sample performance of our tests, we also consider three cases for the covariate X_t:

Case 1 (No structural change in X_t): $X_t = 0.5X_{t-1} + \eta_t$;

Case 2 (An abrupt structural break in X_t): $X_t = \begin{cases} 1.5 + 0.5X_{t-1} + \eta_t, & \text{if } t \leq T/4, \\ -0.3X_{t-1} + \eta_t, & \text{otherwise;} \end{cases}$

Case 3 (Smooth structural change in X_t): $X_t = 1 + \alpha(t/T)X_{t-1} + \eta_t$, with $\alpha(\tau) = 1.5 - 1.5e^{-3(\tau-0.5)^2}$,

where the innovation $\eta_t \sim i.i.d.N(0,1)$ is independent of the error term ε_t.

DGP.S1 satisfies the null hypothesis and is used to study the size of our tests. Chen and Hong (2012) also use this DGP. Specifically, we examine the performance of our tests under various combinations of error terms and covariates. DGPs.P1–P4 describe various kinds of structural changes, including a single structural break, multiple structural breaks, monotonic and non-monotonic smooth structural changes. These DGPs are the same or quite similar to the DGPs considered by Chen and Hong (2012). We investigate the power property of our tests under DGPs.P1–P4 with various combinations of ε_t and X_t.

For each DGP, we generate 500 data sets of the random sample $\{X_t, Y_t\}_{t=1}^T$ for $T = 100$, 200, and 500. The computation of $\hat{K}$ involves searching over all possible $u \in \mathbb{U}$ which is computationally infeasible. Thanks to the periodicity and symmetry of the trigonometric function, we can compute $\hat{K}$ using a finite interval $\mathbb{U}$ that covers certain full periods of the trigonometric functions. In this simulation study, we choose $\mathbb{U} = [0.01, 0.02, ..., 1]$. For $\hat{C}$, we choose the standard normal density as the weighting function.

The critical values of $\hat{C}$ and $\hat{K}$ are obtained by the resampling method described in Section 3, and we set the number of resampling in each replication to be $B = 200$. For Chen and Hong's (2012) generalized Hausman test, we follow their paper to use the rule-of-thumb bandwidth $h = (1/\sqrt{12})T^{-1/5}$ and the uniform kernel. The critical values of Chen and Hong's (2012) test is computed by the wild bootstrap method provided in their paper. Following Andrews (1993), we choose the trimming parameter $\pi = 0.15$ for the tests of Andrews (1993) and Bai and Perron (1998). For Bai and Perron's (1998) test, we set the upper bound of the number of breaks at 5. We use the asymptotic heteroskedasticity-robust critical values.

Table 1 reports the empirical size of aforementioned tests at the 5% and 10% significant levels. We can see that the proposed tests $\hat{C}$ and $\hat{K}$ are robust to unknown structural change in X_t and unknown conditional volatility dynamics of ε_t. The empirical rejection rates of both $\hat{C}$ and $\hat{K}$ are close to the nominal levels in both small and large samples. For other tests, Chen and Hong's (2012) generalized Hausman test also delivers reasonable size. In contract, Andrews' (1993) sup-LM test tends to under-rejection, while Bai and Perron's (1998) test displays relatively strong over-rejection when sample size is small.

Table 1: Empirical size for DGP.S1 with exogenous regressors

		No structural change in X_t						Abrupt structural break in X_t						Smooth structural change in X_t					
		$T = 100$		$T = 200$		$T = 500$		$T = 100$		$T = 200$		$T = 500$		$T = 100$		$T = 200$		$T = 500$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d.	$\hat{K}$	4.2	10.2	4.8	11.0	5.0	11.8	5.4	11.6	6.2	11.8	4.4	9.0	6.4	12.6	4.4	7.8	6.2	11.8
	$\hat{C}$	4.4	10.2	5.6	11.0	5.0	11.2	5.4	10.4	6.0	10.0	4.8	8.8	6.6	12.8	4.0	8.0	7.0	11.4
	$\hat{H}$	4.4	9.2	4.4	9.6	5.6	11.8	6.4	13.8	7.2	14.2	4.0	9.2	6.4	11.8	4.0	8.8	4.2	8.4
	sup-LM	2.0	5.0	2.2	5.0	4.2	10.2	2.2	5.0	3.2	7.0	4.0	8.6	1.8	5.8	3.0	6.8	4.0	9.0
	UDMax	14.6	23.2	8.0	16.4	7.0	14.2	9.4	15.2	7.8	11.4	2.2	4.2	11.6	17.4	6.2	10.0	5.6	9.2
ARCH	$\hat{K}$	4.6	9.6	3.8	9.4	7.2	13.2	7.0	12.8	6.6	11.4	4.6	8.6	5.4	13.0	6.0	11.2	5.0	10.0
	$\hat{C}$	5.6	9.8	5.4	11.0	6.8	11.8	6.4	11.6	6.4	10.4	4.4	8.2	5.4	11.2	6.0	10.6	5.0	10.0
	$\hat{H}$	8.0	13.2	5.8	9.4	7.6	13.2	8.0	13.2	9.0	14.4	5.0	10.4	6.8	13.8	6.0	11.0	6.4	10.8
	sup-LM	2.0	6.2	2.2	5.4	4.2	9.6	1.2	4.8	3.0	6.4	4.6	8.2	1.8	4.4	2.0	5.8	3.4	7.2
	UDMax	14.2	22.0	8.2	13.6	7.6	13.2	10.6	15.2	6.6	10.4	3.0	4.8	11.8	18.8	6.8	11.2	4.4	8.4
Heter.	$\hat{K}$	5.2	13.0	4.8	12.2	3.8	8.6	6.6	13.0	5.8	11.8	4.0	9.6	7.2	13.6	8.0	12.2	5.8	11.4
	$\hat{C}$	5.2	10.8	4.8	10.8	3.2	9.2	5.2	9.4	5.2	11.0	4.4	9.0	8.2	12.4	7.6	12.2	6.0	10.6
	$\hat{H}$	5.2	9.8	5.2	9.4	2.6	8.0	4.8	9.2	7.6	13.2	4.4	9.8	3.8	9.2	5.4	10.2	6.8	10.8
	sup-LM	1.4	2.4	1.6	5.0	3.6	7.2	1.0	2.6	2.4	6.2	3.4	7.8	1.2	4.2	1.2	3.6	2.6	6.8
	UDMax	20.4	32.4	15.8	23.6	7.8	14.6	14.6	21.0	8.2	12.8	2.0	4.6	12.6	18.2	9.4	13.0	6.8	10.2

Notes: (i) $\hat{K}$ denotes the proposed KS type test computed using a grid of $\mathbb{U} = [0.01, 0.02, ..., 1]$; (ii) $\hat{C}$ is the proposed CvM type test computed using the standard normal weighting function; (iii) $\hat{H}$ denotes Chen and Hong's (2012) generalized Hausman test with bandwidth $h = (1/\sqrt{12})T^{-1/5}$; (iv) sup-LM denotes Andrews' (1993) supremum LM test; (v) UDMax denotes Bai and Perron's (1998) double maximum test; (vi) The first row lists three cases of covariates X_t, and the first column lists three cases of error term ε_t; (vii) The results are based on 500 replications for each cases and 200 resampling for tests involving bootstrap; (viii) The main entries report the percentage of rejection.

Table 2 shows the empirical power of tests under DGPs.P1–P4 at the 5% and 10% significant levels. To save space, we only report the results under the case with i.i.d. error terms. The results under the cases with ARCH and conditional heteroskedastic error terms are quite similar and are available upon request. Under a single structural break specified by DGP.P1, Bai and Perron's (1998) test is most powerful for $T = 100$ and 200 when the covariate X_t is stationary. When there exists structural breaks in X_t, our tests are more powerful than the others. Intuitively, structural change in X_t may contribute to the deviation of $\hat{A}(u)$ from a zero spectrum, making the signal of structural change stronger at certain range of u. Under the multiple structural breaks given by DGP.P2, the proposed tests $\hat{C}$ and $\hat{K}$ dominate all the other tests. Chen and Hong's (2012) test performs better than Andrews' (1993) test and Bai and Perron's (1998) test. Next, we consider the monotonic smooth structural changes given by DGP.P3. The $\hat{C}$ and $\hat{K}$ tests are most powerful except for $T = 100$ when X_t is stationary. When the sample size increases, our tests outperform other tests. At last, under the non-monotonic smooth structural changes given by DGP.P4, our tests outperform other tests and the improvement is significant. Andrews' (1993) test almost has no power when the sample size is small and X_t is stationary. To sum up, our test is powerful to detect both smooth structural changes and abrupt structural breaks. They outperform Chen and Hong's (2012) generalized Hausman test in all cases, which is consistent with our analysis on the relative efficiency between our tests and Chen and Hong's (2012) test.

5.2 Endogenous Regressors

In this subsection, we examine the finite sample performance of our tests $\hat{C}^{IV}$ and $\hat{K}^{IV}$ with endogenous covariates under DGP.S1 and DGPs.P1–P4 given in the last subsection. To allow for structural change in the relationship between the endogenous covariate X_t and the instrumental variable Z_t, we consider the following three cases:

Case 1 (No structural change): $X_t = 0.5Z_t + \eta_t$;

Table 2: Empirical power for DGPs.P1–P4 with exogenous regressors (i.i.d. errors)

		No structural change in X_t						Abrupt structural break in X_t						Smooth structural change in X_t					
		$T=100$		$T=200$		$T=500$		$T=100$		$T=200$		$T=500$		$T=100$		$T=200$		$T=500$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$\hat{K}$	44.6	57.0	74.8	84.0	99.2	99.6	92.2	96.0	99.8	99.8	100	100	98.6	100	100	100	100	100
	$\hat{C}$	47.8	59.6	79.2	86.4	99.6	99.6	92.0	95.4	99.8	100	100	100	99.4	100	100	100	100	100
	$\hat{H}$	38.4	51.8	71.2	80.6	99.0	99.4	76.0	83.8	98.6	99.6	100	100	95.2	97.0	100	100	100	100
	sup-LM	23.2	39.4	69.8	79.8	99.2	99.6	72.0	84.0	99.0	99.8	100	100	93.2	96.8	100	100	100	100
	UDMax	57.8	69.0	81.4	86.6	99.4	99.8	48.6	59.4	74.2	84.4	99.8	100	97.4	98.4	100	100	100	100
DGP.P2	$\hat{K}$	53.4	65.8	82.0	87.0	99.6	99.8	84.0	90.2	98.2	99.2	100	100	91.4	96.2	99.8	100	100	100
	$\hat{C}$	53.8	65.6	81.8	87.2	99.4	99.8	84.6	91.4	98.8	99.4	100	100	92.2	96.2	99.6	100	100	100
	$\hat{H}$	37.6	52.4	73.6	84.6	99.4	99.6	63.0	75.4	96.6	98.8	100	100	86.8	91.8	99.6	100	100	100
	sup-LM	22.0	36.0	60.2	73.6	98.4	99.0	37.8	54.2	83.0	92.0	100	100	65.0	75.0	96.0	98.2	100	100
	UDMax	49.6	61.6	72.6	81.2	97.8	98.8	32.2	40.6	64.6	73.8	88.4	93.4	62.4	73.4	89.4	94.0	100	100
DGP.P3	$\hat{K}$	51.4	61.8	83.8	90.4	99.8	99.8	74.8	82.8	97.2	99.0	100	100	99.2	99.8	100	100	100	100
	$\hat{C}$	52.4	61.0	84.0	90.6	99.8	99.8	73.0	79.0	96.6	99.0	100	100	99.2	100	100	100	100	100
	$\hat{H}$	35.4	48.0	72.0	81.6	98.6	99.4	54.4	66.8	88.4	93.2	99.8	100	89.4	96.0	100	100	100	100
	sup-LM	27.4	43.8	71.2	82.4	99.6	99.8	40.0	57.0	88.4	94.8	100	100	89.2	95.2	100	100	100	100
	UDMax	53.2	64.2	83.8	88.0	99.6	99.8	32.8	41.2	40.8	55.2	93.4	97.4	96.0	97.2	99.8	100	100	100
DGP.P4	$\hat{K}$	52.8	64.2	86.6	91.6	100	100	72.4	83.8	98.8	99.8	100	100	91.4	96.0	100	100	100	100
	$\hat{C}$	48.0	58.4	83.4	88.6	100	100	65.4	77.4	98.0	99.2	100	100	89.2	93.6	100	100	100	100
	$\hat{H}$	36.0	51.0	76.4	84.2	99.6	99.8	49.4	63.4	95.0	97.6	100	100	82.8	89.6	99.8	99.8	100	100
	sup-LM	6.6	16.0	26.4	44.6	91.2	97.6	22.2	35.6	75.2	87.4	100	100	24.6	41.0	76.2	87.8	99.8	100
	UDMax	49.2	59.8	74.6	84.6	99.4	99.8	42.4	53.6	71.6	79.6	99.4	99.6	63.2	74.6	87.0	92.2	99.6	99.6

Notes: see the notes in Table 1.

Case 2 (An abrupt structural break): $X_t = \begin{cases} 0.8Z_t + \eta_t, & \text{if } t \leq T/4, \\ 0.4Z_t + \eta_t, & \text{otherwise;} \end{cases}$

Case 3 (Smooth structural change): $X_t = \gamma(t/T)Z_t + \eta_t$, where $\gamma(\tau) = 3.5e^{-(4u-1)^2} + 3.5e^{-(4u-3)^2} - 1.5$. We generate the instrument variable Z_t by $Z_t = 1 + 0.5Z_{t-1} + u_t$, where $u_t \sim i.i.d.N(0,1)$ is independent of ε_t and η_t. To check the robustness of our test, we also consider three cases for the error terms ε_t and η_t. (i) i.i.d. case: $\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$; (ii) ARCH case: $\varepsilon_t = \sqrt{h_t}\nu_t$,

$h_t = 0.2+0.5\varepsilon_{t-1}^2$, $\begin{pmatrix} \nu_t \\ \eta_t \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$; (iii) Heteroskedasticity case: $\varepsilon_t = \sqrt{h_t}\nu_t$, $h_t = 0.2 + 0.5X_t^2$, $\begin{pmatrix} \nu_t \\ \eta_t \end{pmatrix} \sim i.i.d.N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$. Here, the parameter ρ measures the degree of endogeneity. We set $\rho = 0.6$ in this subsection. We have also examined results when $\rho = 0.2$ and $\rho = 0.8$ respectively. Although the rejection rates of all tests are affected by this parameter, the conclusions are quite similar.

We compare the finite sample performance of $\hat{C}^{IV}$ and $\hat{K}^{IV}$ with the smoothed nonparametric test $\hat{H}^{IV}$ proposed by Chen (2015) and the double maximum test of Hall et al. (2012). The critical values of $\hat{C}^{IV}$ and $\hat{K}^{IV}$ are obtained by the resampling method described in Section 4.2. And we set the number of resampling in each replication at $B = 200$. Chen's (2015) test involves two different bandwidths, which are determined by the cross-validation method in Chen (2015). For Hall et al.'s (2012) test, we set the upper bound of breaks to be 5, the trimming parameter $\epsilon = 0.15$, and use the asymptotic heteroskedasticity-robust critical values. For each DGP, we generate 500 data sets of the random sample $\{X_t, Z_t, Y_t\}_{t=1}^T$ for $T = 100$, 200, and 500 respectively.

Table 3 reports the empirical size of these tests at both the 5% and 10% significance levels. As shown in the table, the proposed tests $\hat{C}^{IV}$ and $\hat{K}^{IV}$ are robust to unknown structural changes in the covariates. Our tests deliver reasonable sizes for all cases. Chen's (2015) test also has reasonable size for the cases when X_t has no structural change, but it shows slight under-rejection when X_t suffers from structural changes. Specifically, the under-rejection is alleviated when the error term is conditional heteroskedastic. Hall et al.'s (2012) double maximum test delivers reasonable empirical rejection rates when X_t does not have structural changes. It tends to over-reject the null when X_t has either an abrupt structural break or smooth structural changes. It implies that their test is not robust to structural changes in the marginal distribution of the regressors when using the asymptotic critical values. We note that the fixed regressor bootstrap of Hansen (2000) could solve this issue.

Table 3: Empirical size for DGP.S1 with endogenous regressors

		No structural change in X_t						Abrupt structural break in X_t						Smooth structural change in X_t					
		$T=100$		$T=200$		$T=500$		$T=100$		$T=200$		$T=500$		$T=100$		$T=200$		$T=500$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d.	$\hat{K}^{IV}$	5.6	10.4	4.8	9.0	5.8	10.8	5.0	11.0	5.4	10.4	5.0	9.4	6.0	9.8	5.2	9.0	5.4	10.0
	$\hat{C}^{IV}$	5.4	9.0	5.0	8.4	5.4	11.4	4.8	9.2	6.0	10.2	5.0	9.0	4.6	10.4	5.6	8.6	4.8	9.2
	$\hat{H}^{IV}$	3.4	8.8	2.8	6.8	3.8	6.8	4.4	8.6	3.0	6.8	4.2	11.0	3.0	6.2	1.6	4.8	2.6	3.8
	UDMax^{IV}	6.2	11.0	4.2	9.2	5.8	9.4	24.6	33.2	49.8	61.4	64.2	78.2	68.0	77.2	90.6	97.2	96.2	99.8
ARCH	$\hat{K}^{IV}$	5.8	9.4	4.6	9.0	5.4	9.8	5.4	10.4	6.0	11.0	5.6	9.8	6.0	11.0	4.6	10.2	5.2	9.6
	$\hat{C}^{IV}$	5.0	9.6	4.8	9.0	5.2	10.2	5.4	9.2	6.0	10.0	4.0	8.2	5.6	11.8	5.0	9.2	4.8	10.2
	$\hat{H}^{IV}$	3.0	7.8	3.6	8.4	4.8	9.8	4.0	6.0	3.6	6.4	4.0	8.4	2.6	6.4	1.4	5.8	3.0	6.8
	UDMax^{IV}	6.4	11.0	7.8	14.6	5.8	12.4	43.0	52.8	80.0	88.0	92.2	97.0	91.6	96.4	92.8	97.6	91.6	98.8
Heter.	$\hat{K}^{IV}$	6.2	11.4	2.8	7.2	4.6	8.8	6.6	12.4	4.8	11.0	5.4	10.0	6.0	12.4	5.8	10.8	5.2	10.6
	$\hat{C}^{IV}$	5.4	11.2	3.6	6.8	4.6	9.2	5.2	11.0	5.0	10.2	4.6	9.8	6.0	12.2	5.4	11.4	5.0	10.2
	$\hat{H}^{IV}$	3.2	8.8	3.4	7.8	4.0	7.6	5.0	10.0	4.2	9.0	4.8	9.8	7.8	10.2	5.6	9.4	5.8	10.2
	UDMax^{IV}	8.8	15.8	7.0	13.0	6.8	12.8	27.2	34.2	43.4	54.8	54.2	61.0	34.4	45.6	47.6	53.8	60.6	71.8

Notes: (i) $\hat{K}^{IV}$ denotes the proposed KS type test for endogenous regressors computed using a grid of $\mathbb{U} = [0.01, 0.02, ..., 1]$; (ii) $\hat{C}^{IV}$ is the proposed CvM type test for endogenous regressors computed using the standard normal weighting function; (iii) UDMax^{IV} denotes Hall et al.'s (2012) double maximum F-test; (iv) $\hat{H}^{IV}$ denotes Chen's (2015) heteroskedasticity-robust generalized Hausman test; (v) The first row lists three cases of covariates X_t, and the first column lists three cases of error term ε_t; (vi) The results are based on 500 replications for each cases and 200 resampling for tests involving bootstrap; (vii) The main entries report the percentage of rejection.

Table 4 reports the empirical power of our tests, Hall et al.'s (2012) and Chen's (2015) tests at the 5% and 10% significance levels when the sample size $T = 100$, 200, and 500. From the table, we have the following findings. First, our $\hat{C}^{IV}$ and $\hat{K}^{IV}$ tests are powerful in detecting all forms of structural changes given by DGPs.P1–P4 and the simulation results are consistent with our theoretical conclusion that our tests are able to detect both abrupt structural breaks and smooth structural changes. Second, our tests are more powerful than Hall et al.'s (2012) and Chen's (2015) tests under all DGPs. In fact, the nonparametric test proposed by Chen (2015) relies on nonparametric estimations for both the structural equation and the first stage reduced form. While our tests are based on

Table 4: Empirical power for DGPs.P1–P4 with endogenous regressors (i.i.d. errors)

		No structural change in X_t						Abrupt structural break in X_t						Smooth structural change in X_t					
		$T=100$		$T=200$		$T=500$		$T=100$		$T=200$		$T=500$		$T=100$		$T=200$		$T=500$	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
DGP.P1	$\hat{K}^{IV}$	82.6	88.6	99.0	99.4	100	100	77.4	86.0	95.4	97.6	100	100	99.6	99.8	100	100	100	100
	$\hat{C}^{IV}$	85.2	91.8	99.4	99.6	100	100	81.6	87.4	97.0	98.2	100	100	99.8	99.8	100	100	100	100
	$\hat{H}^{IV}$	29.8	42.8	66.0	74.2	80.2	94.0	56.8	68.0	87.2	92.6	100	100	99.4	99.6	100	100	100	100
	UDMax^{IV}	44.6	54.6	59.2	70.4	80.6	96.0	21.4	28.8	21.0	33.4	40.2	56.8	95.0	97.2	100	100	100	100
DGP.P2	$\hat{K}^{IV}$	64.6	73.2	90.4	94.0	100	100	56.0	69.0	80.2	87.0	99.0	99.6	99.8	99.8	100	100	100	100
	$\hat{C}^{IV}$	64.8	73.8	90.2	93.8	100	100	57.2	69.0	80.6	87.6	99.6	99.6	99.6	98.8	100	100	100	100
	$\hat{H}^{IV}$	30.8	42.2	60.0	71.0	91.8	97.8	48.4	61.6	72.4	80.2	97.8	98.8	99.8	99.8	100	100	100	100
	UDMax^{IV}	40.6	50.4	55.2	66.0	66.8	80.2	18.6	28.6	17.2	27.2	20.2	30.8	94.6	97.0	100	100	100	100
DGP.P3	$\hat{K}^{IV}$	70.4	79.6	93.8	96.0	100	100	60.6	69.6	83.2	87.8	99.4	99.6	100	100	100	100	100	100
	$\hat{C}^{IV}$	71.4	80.0	93.8	95.6	100	100	61.8	69.4	82.8	88.2	99.4	99.6	100	100	100	100	100	100
	$\hat{H}^{IV}$	28.4	37.2	63.2	72.0	98.4	98.6	47.0	59.6	69.4	78.6	98.2	99.0	94.0	96.0	100	100	100	100
	UDMax^{IV}	35.4	47.6	50.2	61.6	60.2	82.0	16.6	24.8	11.2	18.8	15.8	20.2	94.4	96.8	99.6	99.6	100	100
DGP.P4	$\hat{K}^{IV}$	93.2	96.0	100	100	100	100	92.6	97.0	100	100	100	100	96.8	97.6	100	100	100	100
	$\hat{C}^{IV}$	91.4	95.6	100	100	100	100	87.6	94.4	99.8	100	100	100	96.8	98.4	100	100	100	100
	$\hat{H}^{IV}$	26.8	43.4	49.6	61.6	88.2	94.6	46.2	58.2	84.2	88.8	98.6	99.8	95.4	97.6	100	100	100	100
	UDMax^{IV}	77.8	84.6	97.0	98.8	100	100	81.8	89.4	98.4	99.6	100	100	82.2	88.8	99.0	100	100	100

Notes: See notes in Table 3.

the 2SLS using the whole sample. Third, the unknown structural change in the covariates and instruments may help to increase the power of our tests. However, we find that Hall et al.'s (2012) test has lower power when X_t has an abrupt structural break under DGPs.P1–P3.

6 Empirical Application to the Taylor Rule

In this section, we apply our tests to examine the stability of the U.S. Taylor rule. Proposed by Taylor (1993), the Taylor rule describes how a central bank adjusts the interest rate to account for inflation and unemployment. It has been a benchmark for central banks when conducting monetary policy to stabilize price levels and smooth output fluctuations. Clarida et al. (1998, 2000) extend the Taylor rule by incorporating a central bank's forward-looking behavior and its tendency to smooth the fluctuations in interest rates. Recent studies show that the reaction function of the interest rate to inflation and output gap can be time-varying, which can depend on certain state variables. So they use nonlinear models to estimate the Taylor rule. For example, Kim and Nelson (2006) and Boivin (2006) adopt time-varying parameter models to fit the U.S. data during the past 50 years. Brüggemann and Riedel (2011) use logistic smooth transition regression (LSTR) models with time-varying parameters to estimate the U.K. monetary policy reaction function, and Zheng et al. (2012) introduce a regime-switching model to examine the reaction of China's monetary policy. In addition to the nonlinearity, the monetary policy reaction may suffer from structural changes due to the changes of the Chairman of the Federal Reserve System and some other factors, see Clarida et al. (2000). Even though the aforementioned studies allow central banks' reaction to inflation and output gap to be time-varying, their models cannot capture structural changes as described by (2.1), in which the time-varying parameter depends on the deterministic time indices rather than certain underlying state variables.

Following Clarida et al. (1998, 2000) and Kim and Nelson (2006), the forward-looking Taylor rule with smooth interest rate adjustment can be formulated as the following:

$$r_t = (1-\rho)(c + \theta\pi_{t+p} + \gamma y_{t+q}) + \rho r_{t-1} + \varepsilon_t, \tag{6.1}$$

where r_t is the short-term nominal interest rate at time t, π_{t+p} is the inflation rate at time $t+p$, y_{t+q} is the output gap at time $t+q$, and ρ is a smoothing parameter that measures the partial adjustment of interest rate. The error $\varepsilon_t = -(1-\rho)\{\theta[\pi_{t+p} - E_t(\pi_{t+p})] + \gamma[y_{t+q} - E_t(y_{t+q})]\} + m_t$, where E_t is the conditional expectation operator and m_t is a random error term caused by a central bank's control of interest rate. Conditional on information available to the monetary authority at time t, $E_t(\pi_{t+p})$ is the p-th periods ahead forecast for inflation rate at time t. $c = \bar{r} - (\theta - 1)\pi^*$, where $\bar{r}$ is the equilibrium real interest rate, and π^*

is the target level of inflation. Suppose we define $\theta_0 = (1-\rho)c$, $\theta_1 = (1-\rho)\theta$, $\theta_2 = (1-\rho)\gamma$ and $\theta_3 = \rho$, then (6.1) becomes:

$$r_t = \theta_0 + \theta_1\pi_{t+p} + \theta_2 y_{t+q} + \theta_3 r_{t-1} + \varepsilon_t. \tag{6.2}$$

If $p < 0$ and $q < 0$, then (6.2) is the backward-looking Taylor rule with $\varepsilon_t = m_t$. In this case, the parameters in (6.2) can be consistently estimated via OLS. If $p = 0$ and $q = 0$, then (6.2) is the contemporary Taylor rule. If $p > 0$ and $q > 0$, then (6.2) is the forward-looking Taylor rule. As shown by Kim and Nelson (2006), the contemporary and forward-looking Taylor rules suffer from endogeneity problem since the error term ε_t contains forecasting errors, and should be estimated by IV estimation. As in Clarida et al. (1998, 2000) and Kim and Nelson (2006), we consider $p = q = 1$ for the forward-looking Taylor rule.

Using the U.S. quarterly data from 1960Q1 to 2018Q1, we investigate the stability of the backward-looking, contemporary, and forward-looking monetary reaction functions. We apply the tests proposed in Section 3, Andrews' (1993) supremum LM test, and Chen and Hong's (2012) generalized Hausman test to examine the stability of backward-looking Taylor rule under $p = q = -1$ and apply the tests proposed in Section 4 and Chen's (2015) generalized Hausman test to check the stability of contemporary and forward-looking Taylor rules under $p = q = 0$ and $p = q = 1$, respectively. As in Clarida et al. (1998, 2000) and Kim and Nelson (2006), the interest rate is measured by the annualized average Federal Funds rate in the first month of each quarter, the inflation rate is measured by the annualized change rate of the GDP deflators between two subsequent quarters, and the output gap is constructed by the Congressional Budget Office. The set of instruments includes four lags of the following variables: the Federal Funds rate, output gap, inflation rate, commodity price inflation, M2 growth, and the spread between the long-term bond rate and the three-month Treasury Bill rate.

Table 5 reports the results of the proposed tests $\hat{C}$ and $\hat{K}$, Andrews' (1993) supremum LM test as well as Chen and Hong's (2012) or Chen's (2015) generalized Hausman test. We first examine the stability of the Taylor rule for the whole

Table 5: Testing for structural changes in the Taylor rule

	$\hat{C}$	$\hat{K}$	sup-LM	$\hat{H}$
	1960Q1–2018Q1: the whole sample			
$p = -1, q = -1$	0.003	0.019	0.041	0.004
$p = 0, q = 0$	0.000	0.002	–	0.872
$p = 1, q = 1$	0.000	0.000	–	0.908
	1960Q1–1996Q4: Clarida et al. (2000)			
$p = -1, q = -1$	0.002	0.014	0.002	0.012
$p = 0, q = 0$	0.000	0.001	–	0.824
$p = 1, q = 1$	0.000	0.004	–	0.918
	1960Q1–2001Q2: Kim and Nelson (2006)			
$p = -1, q = -1$	0.000	0.006	0.001	0.008
$p = 0, q = 0$	0.000	0.000	–	0.876
$p = 1, q = 1$	0.000	0.002	–	0.884
	1960Q1–1979Q2: Clarida et al.'s (2000) pre-1979			
$p = -1, q = -1$	0.028	0.039	0.000	0.002
$p = 0, q = 0$	0.091	0.127	–	0.872
$p = 1, q = 1$	0.534	0.393	–	0.964
	1979Q3–1996Q4: Clarida et al.'s (2000) post-1979			
$p = -1, q = -1$	0.043	0.043	0.001	0.006
$p = 0, q = 0$	0.008	0.006	–	0.364
$p = 1, q = 1$	0.046	0.039	–	0.482

Notes: (i) $\hat{C}$ and $\hat{K}$ denote the proposed DFT-based tests, which are calculated as in Section 3 when $p = q = -1$, and are calculated as in Section 4 when $p = q = 0$ and $p = q = 1$; (ii) sup-LM is Andrews' (1993) supremum LM test; (iii) $\hat{H}$ is the generalized Hausman test, which is calculated according to Chen and Hong (2012) when $p = q = -1$, and is calculated according to Chen (2015) when $p = q = 0$ and $p = q = 1$; (iv) the p-values of $\hat{C}$, $\hat{K}$ and Chen and Hong's (2012) test are based on 2000 bootstrap iterations, while the p-values of Chen's (2015) test are based on 500 bootstrap iterations; (v) $\hat{K}$ is calculated using a grid of $\mathbb{U} = [0.01, 0.02, \cdots, 1]$ and $\hat{C}$ is computed using the standard normal weighting function; (vi) The main entries report p-values.

sample from 1960Q1 to 2018Q1. The p-values of the proposed DFT tests are below 5% under all specifications. These results document the existence of structural changes in various specifications of the Taylor rule. We also check the stability of the Taylor rule for the sub-samples 1960Q1–1996Q4 and 1960Q1–2001Q2, which are the samples considered by Clarida et al. (2000) and Kim and Nelson (2006), respectively. The p-values of the proposed tests reported in the second and third panels of Table 5 are also less than 5% for all cases, indicating that there exists substantial evidence of structural instabilities of the Taylor rule during these periods. Clarida et al. (2000) point out that there is a significant difference in the way that monetary policy was conducted pre- and post-1979, the year that Paul Volcker was appointed the Chairman of the Board of Governors of the Federal Reserve System. Therefore, they divide the sample into two sub-periods: the pre-1979 period from 1960Q1 to 1979Q2 and the post-1979 period from 1979Q3 to 1996Q4, and then estimate the stable forward-looking Taylor rule within each subsample. The fourth and fifth panels of Table 5 report the test results for these two periods. According to the results of the proposed tests, we find no significant evidence against the stability of the contemporary and forward-looking Taylor rules for the first period, which encompasses the tenures of William M. Martin, Arthur Burns and G. William Miller as Federal Reserve chairmen. It seems that the structural change that we find in the backward-looking Taylor rule may be due to the model misspecification, i.e., the backward-looking Taylor rule may not be suitable to characterize the U.S. monetary reaction function. However, the results reveal the existence of structural changes for the Taylor rule during the second period, which is treated as stable by Clarida et al. (2000).

Table 5 also reports the results of Chen and Hong's (2012) generalized Hausman test and Andrews' (1993) supremum LM test for the backward-looking Taylor rule and the results of Chen's (2015) generalized Hausman test for the contemporary and forward-looking Taylor rules. We note that the results of Andrews' (1993) test are affected by the choice of trimming parameter. For example, if we set the trimming parameter $\pi_0 = 0.15$, the p-value for the whole sample is 0.041, while it changes to 0.063 if we set $\pi_0 = 0.05$, which fails to reject the null hypothesis of no structural change at the 5% level. Table 5 reports the results by setting $\pi_0 = 0.15$. For the backward-looking Taylor rule, the results of Chen and Hong's (2012) generalized Hausman test and Andrews' (1993) supremum LM test are consistent with our frequency domain-based tests for the exogenous variable. However, the results of Chen's (2015) test can not detect the possible structural changes in contemporary and forward-looking Taylor rules for any samples. This is consistent with our theoretical results and simulation studies that our tests are asymptotically more powerful than Chen's (2015) test.

To sum up, our results reveal the existence of structural changes in the Taylor rule, which cannot be accommodated by nonlinear models. In particular, we find that the Taylor rule in the post-1979 period from 1979Q3 to 1996Q4 suffer from structural changes, which are treated as stable by Clarida et al. (2000).

7 Conclusion

This paper proposes DFT-based tests to detect structural changes in a linear time series regression model, which is consistent against both abrupt structural breaks and smooth structural changes. Our tests avoid smoothed nonparametric estimation of the unknown model parameters. Therefore, they do not rely on a prespecified bandwidth and can detect a class of local alternatives at the parametric rate, which is asymptotically more efficient than the existing smoothed nonparametric tests. Our approach is applicable for linear time series regression models with both exogenous and endogenous covariates. In particular, it is robust to structural changes of unknown type in both regressors and instruments. Simulation studies show that in comparison with the tests of Andrews (1993), Bai and Perron (1998), Chen and Hong (2012), Hall et al. (2012), and Chen (2015), our DFT-based tests have both reasonable size and excellent power against various alternatives of abrupt structural breaks and smooth structural changes. In an application to the U.S. Taylor rule, we document significant evidence of structural changes during the post-1979 period, which the existing tests may ignore.

Mathematical Appendix

Due to space constraints, this appendix provides the proofs for the main results of the paper. When proving these results, we only present the statements of lemmas to save space. The proofs for the remaining results are contained in the online supplement.

Proof of Proposition 1. We first state two lemmas.

Lemma 1 *Suppose Assumption 3.1 holds,* $\sup_{u\in\mathbb{U}}\|\hat{Q}_{xx}(u)-\tilde{Q}_{xx}(u)\| = o_P(1)$ *where* $\mathbb{U}=[-c,c]$ *with* $c>0$ *is a compact subset of* $\mathbb{R}$.

Lemma 2 *Suppose Assumptions 3.1–3.4 hold. Let* $\mathbb{U}=[-c,c]$ *be a compact subset of* $\mathbb{R}$ *with* $c>0$. *Then as* $T\to 0$, $\sup_{u\in\mathbb{U}}\left\|\sqrt{T}\hat{A}_2(u)-\hat{S}(u)\right\| \xrightarrow{p} 0$, *where* $\hat{S}(u)\equiv T^{-1/2}\sum_{t=1}^{T}M_t(u)\varepsilon_t$ *with* $M_t(u)=X_t(u)-\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}X_t$.

Under $\mathbb{H}_0:\theta_t=\theta_0$, it is straightforward to show that

$$\hat{A}_1(u) \;=\; \frac{1}{T}\sum_{t=1}^{T}\hat{M}_t(u)X_t'\theta_0 \;=\; \left[\frac{1}{T}\sum_{t=1}^{T}X_t(u)X_t'-\hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T}X_tX_t'\right]\theta_0 \;=\; 0.$$

Then, by Lemma 2, $\sup_{u\in\mathbb{U}}\left\|\sqrt{T}\hat{A}(u)-\hat{S}(u)\right\| = \sup_{u\in\mathbb{U}}\left\|\sqrt{T}\hat{A}_2(u)-\hat{S}(u)\right\| = o_P(1)$, where $\hat{S}(u)=T^{-1/2}\sum_{t=1}^{T}M_t(u)\varepsilon_t$. To show the desired weak convergence, we show that (i) $\hat{S}(u)$ converges in distribution to a normal distribution for almost all fixed $u\in\mathbb{U}$ and (ii) $\hat{S}(u)$ is asymptotically tight.

We first show (i). Let Λ be a $d\times 1$ nonstochastic vector such that $\Lambda'\Lambda=1$. Define $U_T(u)=\Lambda'\hat{S}(u)$, and $\Psi_t(u)=\Lambda'M_t(u)\varepsilon_t$, then $U_T(u)=\Lambda'\hat{S}(u)=T^{-1/2}\sum_{t=1}^{T}\Psi_t(u)$. For each fixed u, $E[\Psi_t(u)]=\Lambda'E[M_t(u)\varepsilon_t]=0$, and

$$\begin{aligned}
\max_t E\left|\Psi_t(u)\right|^r &= \max_t E\left|\Lambda'X_t(u)\varepsilon_t-\Lambda'\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}X_t\varepsilon_t\right|^r\\
&\le 2^{r-1}\left[\max_t E\left|\Lambda'X_t\varepsilon_t e^{\mathbf{i}2\pi ut/T}\right|^r+\left\|\Lambda'\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}\right\|^r\max_t E\left\|X_t\varepsilon_t\right\|^r\right]\\
&\le 2^{r-1}\left[\|\Lambda\|^r\max_t E\left\|X_t\varepsilon_t\right\|^r\left|e^{\mathbf{i}2\pi ut/T}\right|^r+\|\Lambda\|^r\left\|\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}\right\|^r\max_t E\left\|X_t\varepsilon_t\right\|^r\right]\\
&\le \mathcal{M},
\end{aligned}$$

given $\max_t E\left\|X_t\right\|^{2r} < \infty$, $\max_t E\left|\varepsilon_t\right|^{2r} < \infty$ under Assumption 3.1(ii), $\sup_{u\in\mathbb{R}}\left|e^{\mathbf{i}2\pi ut/T}\right| = 1$, and $\tilde{Q}_{xx}(u) = O(1)$ for all u under Assumption 3.4(ii). Denote $\zeta_T^2(u) = \mathrm{var}[U_T(u)]$. Then

$$\zeta_T^2(u) = \Lambda'\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[M_t(u)M_s(u)^*\varepsilon_t\varepsilon_s\right]\right)\Lambda = \frac{1}{T}\sum_{t=1}^{T}E\left|\Lambda' M_t(u)\varepsilon_t\right|^2 > 0$$

for all most all $u \in \mathbb{U}$, where the last equality holds under Assumption 3.3. Then, given the mixing condition in Assumption 3.1, by Theorem 5.20 of White (2001), $\zeta_T(u)^{-1}U_T(u) \xrightarrow{d} N(0,1)$, almost all $u \in \mathbb{U}$ and all Λ. Let $\tilde{\mathcal{K}}(u,v) \equiv T^{-1}\sum_{t=1}^{T}E\left[M_t(u)M_t(v)^*\varepsilon_t^2\right]$. By the Cramer-Wold device, we have for each fixed u, $\tilde{\mathcal{K}}(u,u)^{-1/2}\hat{S}(u) \xrightarrow{d} N(0,\mathbf{I}_d)$, where $\mathbf{I}_d$ denotes a $d\times d$ identity matrix. Furthermore, we can show that the pointwise limit of $\tilde{\mathcal{K}}(u,u)$: $\mathcal{K}(u,u) = \lim_{T\to\infty}\tilde{\mathcal{K}}(u,u) = V_{xx}(u,u) - Q_{xx}(u)Q_{xx}^{-1}V_{xx}(0,u) - V_{xx}(u,0)Q_{xx}^{-1}Q_{xx}(u)^* + Q_{xx}(u)Q_{xx}^{-1}V_{xx}(0,0)Q_{xx}^{-1}Q_{xx}(u)^*$, is well-defined under Assumption 3.4. In particular, when $\{X_t,\varepsilon_t\}$ are weakly stationary, we have $\mathcal{K}(u,u) = E(X_tX_t'\varepsilon_t^2)\left(1-\int_0^1 e^{\mathbf{i}2\pi u\tau}\mathrm{d}\tau\right)^2$. We see that $\mathcal{K}(u,u)$ exists and is positive definite for almost all $u \in \mathbb{R}$.

Next, we show (ii). By the mean value theorem, for any $u_1, u_2 \in \mathbb{U}$,

$$\sum_{t=1}^{T}X_t\varepsilon_t\left(e^{\mathbf{i}2\pi u_1 t/T} - e^{\mathbf{i}2\pi u_2 t/T}\right) = \mathbf{i}2\pi\sum_{t=1}^{T}X_t\varepsilon_t\left(\frac{t}{T}e^{\mathbf{i}2\pi\bar{u}t/T}\right)(u_1-u_2), \text{ and}$$

$$\tilde{Q}_{xx}(u_1) - \tilde{Q}_{xx}(u_2) = \frac{\mathbf{i}2\pi}{T}\sum_{t=1}^{T}E(X_tX_t')\left(\frac{t}{T}e^{\mathbf{i}2\pi\tilde{u}t/T}\right)(u_1-u_2),$$

for some $\bar{u}$ and $\tilde{u}$ that lie between u_1 and u_2. Then

$$\begin{aligned}
& E\left\|\hat{S}(u_1)-\hat{S}(u_2)\right\|^2 \\
\leq\; & 2E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\left(e^{\mathbf{i}2\pi u_1 t/T} - e^{\mathbf{i}2\pi u_2 t/T}\right)\right\|^2 + 2E\left\|\left[\tilde{Q}_{xx}(u_1)-\tilde{Q}_{xx}(u_2)\right]\tilde{Q}_{xx}^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\right\|^2 \\
\leq\; & 8\pi^2 E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\left(\frac{t}{T}e^{\mathbf{i}2\pi\bar{u}t/T}\right)(u_1-u_2)\right\|^2 \\
& +8\pi^2\left\|\frac{1}{T}\sum_{t=1}^{T}E(X_tX_t')\left(\frac{t}{T}e^{\mathbf{i}2\pi\tilde{u}t/T}\right)(u_1-u_2)\right\|^2\left\|\tilde{Q}_{xx}^{-1}\right\|^2 E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\right\|^2 \\
\leq\; & \mathcal{M}(u_1-u_2)^2,
\end{aligned}$$

given $E\left\|T^{-1/2}\sum_{t=1}^{T}X_t\varepsilon_t\left(\frac{t}{T}e^{\mathbf{i}2\pi\bar{u}t/T}\right)\right\|^2 = T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}E(X_t'X_s\varepsilon_t\varepsilon_s)\frac{st}{T^2}e^{\mathbf{i}2\pi\bar{u}(t-s)/T} = T^{-1}\sum_{t=1}^{T}E(X_t'X_t\varepsilon_t^2)\frac{t^2}{T^2} = O(1)$, $\left\|T^{-1}\sum_{t=1}^{T}E(X_tX_t')\left(\frac{t}{T}e^{\mathbf{i}2\pi\tilde{u}t/T}\right)\right\| \leq T^{-1}\sum_{t=1}^{T}\|E(X_tX_t')\| = O(1)$, $Q_{xx}^{-1} = O(1)$, and $E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\varepsilon_t\right\|^2 = O(1)$. Then, the asymptotic tightness is established. Furthermore, given that the Fourier series $e^{\mathbf{i}2\pi ut/T}$ is a continuous function of u, then it is straightforward to show that $\sup_{u,v\in\mathbb{U}}\left\|\tilde{\mathcal{K}}(u,v)-\mathcal{K}(u,v)\right\| = o(1)$, under Assumption 3.4. Given $\mathbb{U}$ is a compact set, we have $\sqrt{T}\hat{A}(u) \Rightarrow S(u)$, where $S(u)$ is a mean-zero complex-valued Gaussian process and covariance kernel $\mathcal{K}(u,v)$ such that $\mathcal{K}(u,v) = E[S(u)S(v)^*] = V_{xx}(u,v) - Q_{xx}(u)Q_{xx}^{-1}V_{xx}(0,v) - V_{xx}(u,0)Q_{xx}^{-1}Q_{xx}(v)^* + Q_{xx}(u)Q_{xx}^{-1}V_{xx}(0,0)Q_{xx}^{-1}Q_{xx}(v)^*$.

Furthermore, if both X_t and ε_t are weakly stationary, the covariance kernel can be simplified as $\mathcal{K}(u,v) = E(X_tX_t'\varepsilon_t^2)\Gamma(u,v)$, where $\Gamma(u,v) = \int_0^1 e^{\mathbf{i}2\pi(u-v)\tau}\mathrm{d}\tau - \int_0^1 e^{\mathbf{i}2\pi u\tau}\mathrm{d}\tau\int_0^1 e^{-\mathbf{i}2\pi v\tau}\mathrm{d}\tau$ is a pseudo-covariance in the sense that τ follows the $U[0,1]$ distribution. ■

Proof of Theorem 1. Given Proposition 1, for any fixed $\mathbb{U}$, $\hat{K} \xrightarrow{d} \sup_{u\in\mathbb{U}}\|S(u)\|^2$, and $\hat{C}_{\mathbb{U}} \equiv \int_{\mathbb{U}}\left\|\hat{S}(u)\right\|^2 W(u)\mathrm{d}u \xrightarrow{d} \int_{\mathbb{U}}\|S(u)\|^2W(u)\mathrm{d}u$, by the continuous mapping theorem. It remains to show that $\hat{C} = \int_{\mathbb{R}}\left\|\hat{S}(u)\right\|^2 W(u)\mathrm{d}u \xrightarrow{d} \int_{\mathbb{R}}\|S(u)\|^2W(u)\mathrm{d}u$. To proceed, we need to check uniform integrability (UI) of $\{\hat{C}_{\mathbb{U}}\}$ by verifying the sufficient condition that $E|\hat{C}_{\mathbb{U}}|^{r/2} \leq \mathcal{M}$ for some $r > 2$. By the definition of $M_t(u)\varepsilon_t$ and under Assumption 3.1,

$$\begin{aligned}\max_t\sup_{u\in\mathbb{R}}E\|M_t(u)\varepsilon_t\|^r &\leq \max_t\sup_{u\in\mathbb{R}}E\left\|X_t\varepsilon_te^{\mathbf{i}2\pi ut/T}\right\| + \max_t\sup_{u\in\mathbb{R}}\left\|\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}\right\|^r E\|X_t\varepsilon_t\|^r \\ &= \max_t E\|X_t\varepsilon_t\|^r + \left\|\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}\right\|^r\max_t E\|X_t\varepsilon_t\|^r \;\leq\; \mathcal{M}.\end{aligned}$$

By the moment bounds for non-stationary mixing sequences (e.g., Kim, 1994), $\sup_{u\in\mathbb{R}}E\left\|\hat{S}(u)\right\|^r \leq \mathcal{M}\max_t\sup_{u\in\mathbb{R}}E\|M_t(u)\varepsilon_t\|_r^r < \infty$. By Jensen's inequality and Fubini theorem,

$$E\left|\hat{C}_{\mathbb{U}}\right|^{r/2} = E\left[\int_{\mathbb{U}}\left\|\hat{S}(u)\right\|^2W(u)\mathrm{d}u\right]^{r/2} \leq \int_{\mathbb{U}}E\left\|\hat{S}(u)\right\|^rW(u)\mathrm{d}u \leq \mathcal{M},$$

under Assumption 3.5. Thus, $\{\hat{C}_{\mathbb{U}}\}$ is uniformly integrable. It implies that $E\left[\hat{C}_{\mathbb{U}}\right] \to \int_{\mathbb{U}}E\|S(u)\|^2W(u)\mathrm{d}u$. Similarly, $\{\hat{S}(u)\}$ also satisfies UI for each fixed

u. This along with the results in Proposition 1 implies $E\|\hat{S}(u)\|^2 \to E\|S(u)\|^2$ for each fixed u. Under Assumption 3.5, for any positive value ϵ, we can find a subset $\mathbb{U}$ of $\mathbb{R}$ large enough such that $\int_{\mathbb{U}^c} E\|S(u)\|^2 W(u)\mathrm{d}u < \epsilon^2/2$, where $\mathbb{U}^c$ is a complementary set of $\mathbb{U}$ in $\mathbb{R}$. Then it follows $E\left(\int_{\mathbb{U}^c}\left\|\hat{S}(u)\right\|^2 W(u)\mathrm{d}u\right) = \int_{\mathbb{U}^c} E\left\|\hat{S}(u)\right\|^2 W(u)\mathrm{d}u \to \int_{\mathbb{U}^c} E\|S(u)\|^2 W(u)\mathrm{d}u < \epsilon^2/2$. For this $\mathbb{U}$, define $\hat{C}_1 = \int_{\mathbb{U}^c}\left\|\hat{S}(u)\right\|^2 W(u)\mathrm{d}u$, $\hat{C}_2 = \int_{\mathbb{U}}\left\|\hat{S}(u)\right\|^2 W(u)\mathrm{d}u$, $C_1 = \int_{\mathbb{U}^c}\|S(u)\|^2 W(u)\mathrm{d}u$, and $C_2 = \int_{\mathbb{U}}\|S(u)\|^2 W(u)\mathrm{d}u$. Then $\hat{C}_2 \xrightarrow{d} C_2$, $E(C_1) < \epsilon^2/2$, and there exists a T_0 large enough such that $E(\hat{C}_1) < \epsilon^2$ for $T \geq T_0$. For any $x \in \mathbb{R}$ and arbitrary small $\epsilon > 0$, we have

$$
\begin{aligned}
P(C_1 + C_2 \leq x - \epsilon) - \epsilon \;&\leq\; P(C_2 \leq x - \epsilon) - \epsilon = \liminf_{T\to\infty} P(\hat{C}_2 \leq x - \epsilon) - \epsilon \\
&= \liminf_{T\to\infty}\left[P(\hat{C}_1 + \hat{C}_2 \leq x) + P(\hat{C}_1 \geq \epsilon)\right] - \epsilon \\
&\leq \liminf_{T\to\infty} P(\hat{C}_1 + \hat{C}_2 \leq x) + \epsilon - \epsilon \;\leq\; \limsup_{T\to\infty} P(\hat{C}_1 + \hat{C}_2 \leq x) \\
&\leq \limsup_{T\to\infty} P(\hat{C}_2 \leq x) \;\leq\; P(C_2 \leq x) \\
&\leq P(C_1 + C_2 \leq x + \epsilon) + P(C_1 \geq \epsilon) \;\leq\; P(C_1 + C_2 \leq x + \epsilon) + \epsilon/2.
\end{aligned}
$$

Letting $\epsilon \to 0$, we have $\lim_{T\to\infty} P(\hat{C}_1 + \hat{C}_2 \leq b) = P(C_1 + C_2 \leq b)$. As a result, $\hat{C} \xrightarrow{d} \int_{\mathbb{R}}\|S(u)\|^2 W(u)\mathrm{d}u$. ■

Proof of Proposition 2. Under $\mathbb{H}_A$, $\hat{A}_1(u) = T^{-1}\sum_{t=1}^{T}\hat{M}_t(u)X_t'\theta_t$. Let $\tilde{A}(u) \equiv \frac{1}{T}\sum_{t=1}^{T} E\left[M_t(u)X_t'\right]\theta_t$, We want to show that $\sup_{u\in\mathbb{U}}\left\|\hat{A}(u) - \tilde{A}(u)\right\| \leq \sup_{u\in\mathbb{U}}\left\|\hat{A}_1(u) - \tilde{A}(u)\right\| + \sup_{u\in\mathbb{U}}\left\|\hat{A}_2(u)\right\| = o_P(1)$. By Lemma 2 and Proposition 1, it is straightforward to show $\sup_{u\in\mathbb{U}}\left\|\hat{A}_2(u)\right\| = o_P(1)$. We consider

$$
\begin{aligned}
&\sup_{u\in\mathbb{U}}\left\|\hat{A}_1(u) - \tilde{A}(u)\right\| \\
\leq\;& \sup_{u\in\mathbb{U}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\hat{M}_t(u)X_t' - M_t(u)X_t'\right]\theta_t\right\| + \sup_{u\in\mathbb{U}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{M_t(u)X_t' - E\left[M_t(u)X_t'\right]\right\}\theta_t\right\| \\
=\;& \sup_{u\in\mathbb{U}}\left\|\left[\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\right]\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\theta_t\right\| \\
&+\sup_{u\in\mathbb{U}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{X_t(u)X_t' - E[X_t(u)X_t']\right\}\theta_t - \tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T}[X_tX_t' - E(X_tX_t')]\theta_t\right\|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{u\in\mathbb{U}}\left\|\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}-\hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}X_tX_t'\theta_t\right\|+\sup_{u\in\mathbb{U}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left\{X_t(u)X_t'-E[X_t(u)X_t']\right\}\theta_t\right\| \\
&\quad+\sup_{u\in\mathbb{U}}\left\|\tilde{Q}_{xx}(u)\right\|\left\|\tilde{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T}[X_tX_t'-E(X_tX_t')]\theta_t\right\| \\
&= R_1+R_2+R_3,\ \text{say.}
\end{aligned}
$$

For R_1, we have shown that $\sup_{u\in\mathbb{U}}\left\|\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}-\hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\right\| = o_P(1)$ in the proof of Lemma 2. And $\left\|\frac{1}{T}\sum_{t=1}^{T}X_tX_t'\theta_t\right\| = O_P(1)$ follows from Assumptions 3.1(ii) an 3.6(i). Then we have $R_1 = o_P(1)$. To show $R_2 = o_P(1)$, we need to show that: (i) $T^{-1}\sum_{t=1}^{T}\{X_tX_t'-E[X_tX_t']\}\theta_te^{\mathbf{i}2\pi ut/T}$ converges to 0 pointwisely, and (ii) $T^{-1}\sum_{t=1}^{T}\{X_tX_t'-E[X_tX_t']\}\theta_te^{\mathbf{i}2\pi ut/T}$ is stochastically equicontinuous over $\mathbb{U}$. Under Assumption 3.6(i), the desired results can be established analogously to the proof of Lemma 1. Finally, $R_3 = o_P(1)$ is due to $\sup_{u\in\mathbb{U}}\left\|\tilde{Q}_{xx}(u)\right\| = O_P(1)$ and $\tilde{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T}[X_tX_t'-E(X_tX_t')]\theta_t = O_P(T^{-1/2})$ by Chebyshev's inequality under Assumptions 3.1 and 3.6.

Thus, we have $\sup_{u\in\mathbb{U}}\left\|\hat{A}_1(u)-\tilde{A}(u)\right\| = o_P(1)$. Next, we need to show that the limit of $\tilde{A}(u)$ exists. Note that under $\mathbb{H}_A$, the OLS estimator $\hat{\theta} = \tilde{\theta}+O_P(T^{-1/2})$, where the limit of $\tilde{\theta}$ exists under Assumption 3.4(ii) and 3.6(ii) by letting $u=0$.

$$
\begin{aligned}
\lim_{T\to\infty}\tilde{A}(u) &= \lim_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}E[X_t(u)X_t']\theta_t-\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T}E(X_tX_t')\theta_t\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E[X_t(u)X_t']\left(\theta_t-\tilde{\theta}\right)
\end{aligned}
$$

exists since Assumption 3.4(ii) ensures that $\lim_{T\to\infty}\tilde{Q}_{xx}(u)$ and $\lim_{T\to\infty}\tilde{Q}_{xx}^{-1}$ exist. And Assumption 3.6(ii) ensures that $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E[X_t(u)X_t']\theta_t$ exists. Furthermore, when X_t is weakly stationary, we have $\tilde{A}(u) = \frac{1}{T}\sum_{t=1}^{T}E\left[M_t(u)X_t'\right]\theta_t = E(X_tX_t')\frac{1}{T}\sum_{t=1}^{T}\left[e^{\mathbf{i}2\pi ut/T}\theta_t-\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}2\pi ut/T}\theta_t\right] \to E(X_tX_t')\widetilde{\text{cov}}[\theta(\tau),e^{\mathbf{i}2\pi u\tau}]$, where $\widetilde{\text{cov}}[\theta(\tau),e^{\mathbf{i}2\pi u\tau}] = \int_0^1\theta(\tau)e^{\mathbf{i}2\pi u\tau}\mathrm{d}\tau-\int_0^1\theta(\tau)\mathrm{d}\tau\int_0^1 e^{\mathbf{i}2\pi u\tau}\mathrm{d}\tau$ is a pseudo-covariance in the sense that τ follows the $U[0,1]$ distribution. ■

Proof of Theorem 2. Under $\mathbb{H}_A$, by Proposition 2, $\hat{A}(u) = \hat{A}_1(u)+\hat{A}_2(u) = \tilde{A}(u)+o_P(1)$. Given Lemma 1, it is easy to show that $\tilde{A}(u)$ is stochastically equicontinuous. Therefore, $\hat{K} = T\sup_{u\in\mathbb{U}}\|\hat{A}(u)\|^2 = O_P(T)$ and $\hat{C} = T\int_{\mathbb{R}}\|\hat{A}(u)\|^2W(u)\mathrm{d}u =$

$O_P(T)$. ■

Proof of Theorem 3. Under $\mathbb{H}_A(\Delta_T): \theta_t = \theta_0 + \Delta_T \phi_t$,

$$\hat{A}(u) \;=\; \frac{1}{T}\sum_{t=1}^{T} \hat{M}_t(u) X_t'(\theta_0 + \Delta_T \phi_t) + \hat{A}_2(u) \;=\; \frac{\Delta_T}{T}\sum_{t=1}^{T} \hat{M}_t(u) X_t' \phi_t + \hat{A}_2(u),$$

where the last equality holds by $\sum_{t=1}^{T} \hat{M}_t(u) X_t' = 0$. Given $\Delta_T = T^{-1/2}$, $\sqrt{T}\hat{A}(u) = T^{-1}\sum_{t=1}^{T} \hat{M}_t(u) X_t' \phi_t + \sqrt{T}\hat{A}_2(u)$. By analogous derivations in Proof of Proposition 2, we can show $\sup_{u\in\mathbb{U}} \left\| \frac{1}{T}\sum_{t=1}^{T} \hat{M}_t(u) X_t' \phi_t - \kappa(u) \right\| \xrightarrow{p} 0$ where

$$\kappa(u) \;=\; \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[M_t(u) X_t'] \phi_t \;=\; \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left[X_t(u) X_t'\right] \left(\phi_t - \tilde{\phi}\right),$$

with $\tilde{\phi} = \tilde{Q}_{xx}^{-1}\left[T^{-1}\sum_{t=1}^{T} E(X_t X_t' \phi_t)\right]$. By Proposition 1, we have $\sqrt{T}\hat{A}_2(u) \Rightarrow S(u)$, over $u \in \mathbb{U}$ where $S(u)$ is a complex-valued Gaussian process defined in Proposition 1. Therefore, under $\mathbb{H}_A(\Delta_T)$, $\sqrt{T}\hat{A}(u) \Rightarrow \kappa(u) + S(u)$, over $u \in \mathbb{U}$. By the continuous mapping theorem and analogous arguments in the proof of Theorem 1, we have $\hat{C} \xrightarrow{d} \int_{\mathbb{R}} \|\kappa(u) + S(u)\|^2 W(u)\mathrm{d}u$ and $\hat{K} \xrightarrow{d} \sup_{u\in\mathbb{U}} \|\kappa(u) + S(u)\|^2$. ■

Proof of Theorem 4. By the definition of $\hat{A}_B(u)$, we have

$$\begin{aligned}
\sqrt{T}\hat{A}_B(u) \;&=\; \frac{1}{\sqrt{T}}\sum_{t=1}^{T} M_t(u)\varepsilon_t v_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{M}_t(u) - M_t(u)\right]\varepsilon_t v_t \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} M_t(u)\left(\hat{\varepsilon}_t - \varepsilon_t\right) v_t + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{M}_t(u) - M_t(u)\right]\left(\hat{\varepsilon}_t - \varepsilon_t\right) v_t \\
&=\; R_1(u) + R_2(u) + R_3(u) + R_4(u), \text{ say.}
\end{aligned}$$

Let $E_\omega(\cdot) \equiv E(\cdot|\omega)$ denote the conditional expectation on the sample realization $\omega \in \mathcal{W}$. And $\mathcal{W}$ denotes the set of samples ω for which $\hat{Q}_{xx}(u) - \tilde{Q}_{xx}(u) \to$ and $\limsup_{T\to\infty} T^{-1}\sum_{t=1}^{T} \sup_{u\in\mathbb{U}} \|M_t(u)\|^2 \varepsilon_t^2 < \infty$ uniformly over $u \in \mathbb{U}$, and $\hat{\mathcal{K}}(u_1, u_2) \to \mathcal{K}(u_1, u_2)$, uniformly over $(u_1, u_2) \in \mathbb{U}^2$, where we let $\hat{\mathcal{K}}(u_1, u_2) = T^{-1}\sum_{t=1}^{T} M_t(u_1) M_t(u_2)^* \varepsilon_t^2$, and $\mathcal{K}(u_1, u_2) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left[M_t(u_1) M_t(u_2)^* \varepsilon_t^2\right]$, as is defined in the proof of Proposition 1. Under Assumptions 3.1–3.5, we have $P(\mathcal{W}) = 1$. We first show that $R_1(u) \Rightarrow S(u)$ over $\mathbb{U} = [-c, c]$. Note that $\{\nu_t\}$ is i.i.d. N(0,1), then conditional on ω, $R_1(u)$ is a mean-zero Gaussian process with the following covariance function $E_\omega\left[R_1(u_1)R_1(u_2)^*\right] = T^{-1}\sum_{t=1}^T E\left[M_t(u_1)M_t(u_2)^*\varepsilon_t^2 v_t^2\right] = T^{-1}\sum_{t=1}^T M_t(u_1)M_t(u_2)^*\varepsilon_t^2 = \hat{\mathcal{K}}(u_1,u_2)$. Since for $\omega \in \mathcal{W}$, $\hat{\mathcal{K}}(u_1,u_2) \to \mathcal{K}(u_1,u_2)$, we have $R_1(u)$ converges to $S(u)$ in distribution.

Next, by the definition of $\mathcal{W}$, we have that the envelop of $M_t(u)\varepsilon_t$ is L^2-integrable. We have

$$\limsup_{T\to\infty} T^{-1}\sum_{t=1}^T E_\omega\left[\sup_{u\in\mathbb{U}}\|M_t(u)\|^2\varepsilon_t^2\right] = \limsup_{T\to\infty} T^{-1}\sum_{t=1}^T \sup_{u\in\mathbb{U}}\|M_t(u)\|^2\varepsilon_t^2 < \infty.$$

Furthermore, $\left(E_\omega\|R_1(u_1)-R_1(u_2)\|^2\right)^{1/2} = \left(T^{-1}\sum_{t=1}^T\|M_t(u_1)\varepsilon_t - M_t(u_2)\varepsilon_t\|^2\right)^{1/2} \to \left(\mathrm{tr}\left\{\mathcal{K}(u_1,u_1)+\mathcal{K}(u_2,u_2)-\mathcal{K}(u_1,u_2)-\mathcal{K}(u_2,u_1)\right\}\right)^{1/2} = \|R_1(u_1)-R_1(u_2)\|$, uniformly over $(u_1,u_2)\in\mathbb{U}^2$ under $\mathcal{W}$. Then, by analogous arguments on L^2 bracketing numbers in Proof of Theorem 2, Hansen (1996), we have $R_1(u)$ is stochastically equicontinuous over $\mathbb{U}$. Given $P(\mathcal{W})=1$, we have $R_1(u) \Rightarrow S(u)$.

It remains to show that $R_2(u)$, $R_3(u)$, and $R_4(u)$ are asymptotically negligible under $\mathbb{H}_0$. For $R_2(u)$, we have $\sup_{u\in\mathbb{U}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T\left[\hat{M}_t(u)-M_t(u)\right]\varepsilon_t v_t\right\| \leq \sup_{u\in\mathbb{U}}\left\|\tilde{Q}_{xx}(u)\tilde{Q}_{xx}^{-1}-\hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\right\|\frac{1}{\sqrt{T}}\sum_{t=1}^T X_t\varepsilon_t v_t \Rightarrow_p 0$, conditional on ω since we pick $\mathcal{W}$ such that $\hat{Q}_{xx}(u)-\tilde{Q}_{xx}(u)\to 0$, uniformly over $u\in\mathbb{U}$. For $R_3(u)$, we have $\sup_{u\in\mathbb{U}}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^T M_t(u)\left(\hat{\varepsilon}_t-\varepsilon_t\right)v_t\right\| \leq \sup_{u\in\mathbb{U}}\left\|\frac{1}{T}\sum_{t=1}^T M_t(u)X_t'v_t\right\|\left\|\sqrt{T}(\theta_0-\hat{\theta})\right\|$ $\Rightarrow_p 0$, under $H_0:\theta_0=\theta_t$, since conditional on ω, $T^{-1}\sum_{t=1}^T M_t(u)X_t'v_t \Rightarrow_p 0$. Combining the results of $R_2(u)$ and $R_3(u)$, it is quite straightforward to show that $R_4(u)\Rightarrow_p 0$. Therefore, we have $\sqrt{T}\hat{A}_B(u)\Rightarrow_p S(u)$. By the continuous mapping theorem, and similar arguments in the Proof of Theorem 1, we can establish the consistency of the bootstrapped p-value under $\mathbb{H}_0$.

Under $\mathbb{H}_A$, it is straightforward to show that $E_\omega(\hat{A}_B(u))=0$ and $E_\omega\|\hat{A}_B(u)\|^2 = O_p(1)$ given that v_t is i.i.d.$N(0,1)$. By Chebyshev's inequality, we have that the bootstrapped test statistics are bounded in probability conditional on the sample. Given that $\sup_u\|\sqrt{T}\hat{A}(u)\| = O_p(\sqrt{T})$ by Proposition 2, it follows that $P(\hat{C}>\hat{C}_b)\to 1$ and $P(\hat{K}>\hat{K}_b)\to 1$. ∎

References

Andrews, D. W. K. (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2247–2294. North Holland.

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

Boivin, J. (2006): "Has U.S. Monetary Policy Changed? Evidence from Drifting Coefficients and Real-Time Data," *Journal of Money, Credit and Banking*, 38, 1149–1173.

Bradley, R. C. (2005): "Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions," *Probability Surveys*, 2, 107–144.

Brüggemann, R., and J. Riedel (2011): "Nonlinear Interest Rate Reaction Functions for the UK," *Economic Modelling*, 28, 1174–1185.

Cai, Z. (2007): "Trending Time-Varying Coefficient Time Series Models with Serially Correlated Errors," *Journal of Econometrics*, 136, 163–188.

Cai, Z., Y. Wang, and Y. Wang (2015): "Testing Instability in a Predictive Regression Model with Nonstationary Regressors," *Econometric Theory*, 31, 953–980.

Chen, B. (2015): "Modeling and Testing Smooth Structural Changes with Endogenous Regressors," *Journal of Econometrics*, 185, 196–215.

Chen, B., and Y. Hong (2012): "Testing for Smooth Structural Changes in Time Series Models via Nonparametric Regression," *Econometrica*, 80, 1157–1183.

Choi, I., and P. C. B. Phillips (1993): "Testing for a Unit Root by Frequency Domain Regression," *Journal of Econometrics*, 59, 263–286.

Clarida, R., J. Galí, and M. Gertler (1998): "Monetary Policy Rules in Practice: Some International Evidence," *European Economic Review*, 42, 1033–1067.

——— (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *The Quarterly Journal of Economics*, 115, 147–180.

Corbae, D., S. Ouliaris, and P. C. B. Phillips (2002): "Band Spectral Regression with Trending Data," *Econometrica*, 70, 1067–1109.

Dwivedi, Y., and S. Subba Rao (2011): "A Test for Second-Order Stationarity of a Time Series Based on the Discrete Fourier Transform," *Journal of Time Series Analysis*, 32, 68–91.

Engle, R. F. (1974): "Band Spectrum Regression," *International Economic Review*, 15, 1–11.

Granger, C. W. J., and M. Hatanaka (1964): *Spectral Analysis of Economic Time Series*. Princeton University Press.

Granger, C. W. J., and M. W. Watson (1984): "Time Series and Spectral Methods in Econometrics," in *Handbook of Econometrics*, vol. 2, pp. 979–1022. Elsevier.

Hall, A. R., S. Han, and O. Boldea (2012): "Inference Regarding Multiple Structural Changes in Linear Models with Endogenous Regressors," *Journal of Econometrics*, 170, 281–302.

Hannan, E. J. (1965): "The Estimation of Relationships Involving Distributed Lags," *Econometrica*, 33, 206–224.

——— (1967): "The Estimation of a Lagged Regression Relation," *Biometrika*, 54, 409–418.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

Hansen, B. E. (2000): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115.

——— (2001): "The New Econometrics of Structural Change: Dating Breaks in US Labor Productivity," *The Journal of Economic Perspectives*, 15, 117–128.

Hong, Y., X. Wang, and S. Wang (2017): "Testing Strict Stationarity with Applications to Macroeconomic Time Series," *International Economic Review*, 58, 1227–1277.

Jentsch, C., and S. Subba Rao (2015): "A Test for Second Order Stationarity of a Multivariate Time Series," *Journal of Econometrics*, 185, 124–161.

Kim, C.-J., and C. R. Nelson (2006): "Estimation of a Forward-Looking Monetary Policy Rule: A Time-Varying Parameter Model Using Ex Post Data," *Journal of Monetary Economics*, 53, 1949–1966.

Kim, T. Y. (1994): "Moment Bounds for Non-Stationary Dependent Sequences," *Journal of Applied Probability*, 31, 731–742.

Kristensen, D. (2012): "Non-Parametric Detection and Estimation of Structural Change," *The Econometrics Journal*, 15, 420–461.

Lin, C.-F. J., and T. Teräsvirta (1994): "Testing the Constancy of Regression Parameters against Continuous Structural Change," *Journal of Econometrics*, 62, 211–228.

Orbe, S., E. Ferreira, and J. Rodríguez-póo (2000): "A Nonparametric Method to Estimate Time Varying Coefficients under Seasonal Constraints," *Journal of Nonparametric Statistics*, 12, 779–806.

——— (2005): "Nonparametric Estimation of Time Varying Parameters under Shape Restrictions," *Journal of Econometrics*, 126, 53–77.

Perron, P. (2006): "Dealing with Structural Breaks," in *Palgrave Handbook of Econometrics*, vol. 1, pp. 278–352.

Perron, P., and Y. Yamamoto (2014): "A Note on Estimating and Testing for Multiple Structural Changes in Models with Endogenous Regressors via 2SLS," *Econometric Theory*, 30, 491–507.

——— (2015): "Using OLS to Estimate and Test for Structural Changes in Models with Endogenous Regressors," *Journal of Applied Econometrics*, 30, 119–144.

Robinson, P. M. (1989): "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. by P. Hackl, pp. 253–264. Springer Berlin Heidelberg.

——— (1991): "Time-Varying Nonlinear Regression," in *Economic Structural Change*, pp. 179–190. Springer Berlin Heidelberg.

Stock, J. H., and M. W. Watson (1996): "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

Taylor, J. B. (1993): "Discretion versus Policy Rules in Practice," in *Carnegie-Rochester Conference Series on Public Policy*, vol. 39, pp. 195–214. Elsevier.

Welch, I., and A. Goyal (2008): "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, 21, 1455–1508.

White, H. (2001): *Asymptotic Theory for Econometricians.* Academic Press, rev. edn.

Xu, K.-L. (2015): "Testing for Structural Change under Non-Stationary Variances," *The Econometrics Journal*, 18, 274–305.

Yamamoto, Y., and P. Perron (2013): "Estimating and Testing Multiple Structural Changes in Linear Models Using Band Spectral Regressions," *The Econometrics Journal*, 16, 400–429.

Zhang, C., D. R. Osborn, and D. H. Kim (2008): "The New Keynesian Phillips Curve: From Sticky Inflation to Sticky Prices," *Journal of Money, Credit and Banking*, 40, 667–699.

Zhang, T., and W. B. Wu (2012): "Inference of Time-Varying Regression Models," *The Annals of Statistics*, 40, 1376–1402.

Zheng, T., X. Wang, and H. Guo (2012): "Estimating Forward-Looking Rules for China's Monetary Policy: A Regime-Switching Perspective," *China Economic Review*, 23, 47–59.