An Asymptotically Efficient Test for Functional Coefficient Models with Application to Conditional Asset Pricing Models

(Job Market Paper)

Xingtong Zhang*

November 12, 2019

Abstract

We propose a consistent test for model specification in a functional coefficient model that uses the discrete Fourier transform of a consistent nonparametric estimator of the random coefficient. As a generalization of the conditional moment tests by Bierens (1980, 1982), it is applicable in testing part of the coefficient functions, rather than testing for all of the them jointly. Although a nonparametric estimation step is included, our method is able to detect local alternatives at a rate of $\sqrt{T}$, owing to the U-process structure of the test statistics. Monte Carlo studies demonstrate that our method outperforms current nonparametric tests, such as the generalized likelihood ratio test by Fan et al. (2001) and the Wald-typed tests by Li et al. (2002), especially when the sample size decreases and the dimension of the state variables increases. In application, we employ our test to examine the validity of conditional asset pricing models. We demonstrate that the findings in current literature are misleading due to the small sample problem, which is caused by the coarseness of the state variables. We further check the robustness of the results using various combinations of the state variables.

Keywords: Functional coefficient; Fourier transform; U-process; Conditional CAPM

JEL Classification: C12; C14

*Correspondence: Xingtong Zhang, Department of Economics, Cornell University, Ithaca, NY 14850. Email: xz387@cornell.edu

1 Introduction

Since the work of Hastie and Tibshirani (1993), functional coefficient models have been widely used to capture nonlinear dynamics and changes in economic and financial phenomena. As natural generalizations of classical linear regression models that allow the coefficients to be functions of observable state variables, they are both more flexible and interpretable than other semiparametric models.[1] More importantly, many classical nonlinear models can be regarded as special cases of functional coefficient models. For example, the threshold autoregressive (TAR) model of Tong (1990) assumes that random coefficients are step functions of some state variables. A special case is the self-exciting TAR (SETAR) model of Tong (1983), where the state variables are some lagged order of the series themselves. Another example is the smooth transition autoregressive (STAR) model of Teräsvirta (1994), where the coefficients are typically continuous CDFs of the state variables. Other examples include the FAR of Chen and Tsay (1993), and the exponential autoregressive (EXPAR) model of Haggan and Ozaki (1981). Moreover, many empirical studies have focused on functional coefficient models. For example, researchers in asset pricing use conditional factor models to explain the failures of unconditional asset pricing models when analyzing the cross-sectional returns of different assets. Ideally, the parameters (i.e., the factor loadings) change with the investor's unobservable information set, which can be proxied by state variables.[2] Although Shanken (1990) and Ferson and Harvey (1999) specify parametric forms of the factor loadings, in order to avoid a potential misspecification, nonparametric methods have been proposed for inferences of conditional factor models. These methods include those of French, Schwert, and Stambaugh (1987), Bansal, Hsieh, and Viswanathan (1993), Bansal and Viswanathan (1993), Wang (2003), Lewellen and Nagel (2006), Nagel and Singleton (2011), Ang and Kristensen (2010), Li and Yang (2011), and Roussanov (2014). In the field of labor economics, Card (2001) proposes that the returns on schooling should vary with post-school experience. If wage models assume the additive separability of education and experience, then returns on education could be understated if returns on experience are increasing in education. Therefore, inferences using functional coefficient models are important in economics.

In an important pioneering work, Fan et al. (2001) propose a generalized likelihood ratio (GLR) test in which they compare the likelihood of the parametric specification under the null hypothesis and

[1]Examples include projection pursuit, developed by Huber (1985), the sliced inverse regression of Li (1991), the single index models of Härdle and Stoker (1990), the additive models of Breiman and Friedman (1985) and Hastie and Tibshirani (1987), the low-dimensional interaction models of Friedman (1991), Gu and Wahba (1992), and Stone et al. (1997), the partially linear models of Wahba (1984) and Green and Silverman (1994), and their hybrids, developed by Carroll et al. (1997), Fan et al. (1998), Heckman et al. (1998), Fan et al. (2003), and others.

[2]In another specification, the parameters are deterministic functions of time; see Lewellen and Nagel (2006), Ang and Kristensen (2010), and Li and Yang (2011) for details.

the nonparametric specification (functional coefficient models) under the alternative hypothesis. By assuming that the stochastic errors follow certain parametric distributions, which need not contain the true distribution, they overcome the difficulty of the nonparametric MLE. In this way, they attenuate the difficulty of the nonparametric maximum LR test, and enhance the flexibility of the test by allowing for a range of smoothing parameters. The asymptotic null distribution of the GLR is free of nuisance parameters and nuisance functions. Therefore, it enjoys the appealing Wilks phenomena. Moreover, the rate of alternatives it can detect is $T^{-\frac{4}{9}}$ under suitable choices of smoothing parameters. Ang and Kristensen (2012) use the GLR to test pricing errors in conditional CAPM models. Hong and Lee (2013)[3] improve the efficiency of the GLR using a loss function approach. Here, they reduce the variance of GLR test statistic by eliminating the first-order term by using particular loss functions. Another set of specification tests for functional coefficient models are the Wald-type tests. Li et al. (2002) use the integrated squared difference between the parametric and nonparametric estimators of the random coefficient to test parametric specifications under the null hypothesis. The authors apply the test to the production function of the nonmetal mineral industry in China. Their results show that intermediate production and management expenses play a vital role, and represent an unbalanced determinant of the labor and capital elasticities of output in production. Chen and Hong (2012)[4] consider a generalized Hausman test using the quadratic distance between the parametric and nonparametric estimators. Li and Yang (2011) apply the test to conditional asset pricing models. Li et al. (2015) consider a two-step procedure to estimate conditional factor models. The first step acquires the residuals from the conventional parametric estimation. The second step is based on the nonparametric estimation of the residual from the first step. The test statistic is based on a Wald-type test of the two-step estimator. Fu and Hong (2019) use a discrete Fourier transform method to examine smooth structural changes in time-varying coefficient models (functional coefficient models, where the state variable is time). They argue that the time-varying local feature of the model parameters will remain in the residuals estimated using an ordinary least squares regression.

We propose an alternative approach, using a Fourier transform to construct a convenient and consistent specification test for functional coefficient models. Our test is based on a discrete Fourier transform (DFT) of the consistent nonparametric estimator of the functional coefficient and, thus, has a U-process structure. As pointed out by Powell, Stock, and Stoker (1989), each data point is used to estimate several kernel regressions, and the overlaps render a faster convergence rate. In essence, we transform

[3] Hong and Lee (2013) consider a model in which the conditional expectation is any nonparametrically specified function that can be extended to a functional coefficient model.

[4] Chen and Hong (2012) consider a test for smooth structural changes, where the smoothing is in the time domain. This test can be extended to the state domain as well.

the test problem in the time or state variable domain to one in the frequency domain, achieving a parametric rate at a relatively small cost. Our approach has several advantages over existing parametric and nonparametric tests in literature.

First, compared with nonparametric tests, and despite including a nonparametric estimation step, our approach detects alternatives with parametric rate $T^{-\frac{1}{2}}$ and, thus, does not suffer from the so-called "curse of dimensionality." Owing to the U-process structure of the test statistics, our test significantly improves the power performance in small samples, and when the dimension of the state variable increases. In addition, for particular kinds of null hypotheses (e.g., testing a linear combination of the coefficients), our test is easy to implement by simply imposing a coefficient matrix; other nonparametric tests may incur constrained optimization when estimating the null model. Furthermore, as pointed out by Gao and Gijbels (2008), the performance of nonparametric tests is highly sensitive to the bandwidth selection. Gao and Gijbels (2008) argue that the bandwidth should be chosen based on the trade-off between size and power, rather than on minimizing the mean squared error (MSE) of the estimation. Because of the parametric rate, the power of our test is not affected by the choice of bandwidth; thus, using MSE-based criteria for bandwidth selection is appropriate.

Second, compared with parametric tests, our method can handle a larger class of tests. In particular, it can test one or certain dimensions of the coefficients, as well as any linear combination of the coefficients. For example, the residual-based approach of Fu and Hong (working paper) is computationally convenient, and can achieve a parametric rate. However, because the residuals from the parametric model under the null are derived from a one-step regression of the entire model, they are not able to analyze the behavior of one (or a few) of the coefficients separately. Therefore, these tests cannot be applied to economics and finance problems such as testing the pricing errors of asset pricing models. In contrast, our test can easily handle such problem by computing the DFT of certain coefficients, rather than all of them, and the feature of detecting local alternatives with parametric rates still holds.

To highlight our approach, we apply our tests to examine the validity of various conditional asset pricing models. Using daily, weekly, and monthly U.S. stock market data, we re-examine the validity of conditional versions of well-known asset pricing models, including the CAPM and Fama and French three-factors models. For the case in which the state variable is time, we compare our results with those of Li and Yang (2011) and Ang and Kristensen (2012), who use GLR and Wald-type tests, respectively. When the state variables are random variables, we follow the choices in Nagel and Singleton (2011) and Li et al. (2015). Our method rejects models that are not rejected by the GLR and Wald-type tests, especially when we use all three state variables from the literature.

The paper proceeds as follows. In Section 2, we introduce our framework and propose a new test statistic using DFT. In Section 3, we provide the asymptotic theory, and in Section 4, we discuss the Monte Carlo simulations. In Section 5, we provide an empirical application to test the validity of our approach in various conditional asset pricing models. Conclusions are provided in Section 6. All proofs are relegated to the Appendix.

Notation. Throughout this paper, $\mathbf{i}$ denotes an imaginary number, such that $\mathbf{i} = \sqrt{-1}$. For an $m \times n$ complex matrix A, we denote its complex conjugate as A^*, its transpose as A', its real part as $\text{Re}(A)$, its Euclidean norm as $\|A\| (\equiv [\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2]^{1/2})$, and its Moore–Penrose generalized inverse as A^-. We use $C \in (0, \infty)$ to denote a generic positive constant, which may vary between cases. The operator $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, and $\Rightarrow$ denotes weak convergence.

2 Framework and Approach

In this study, we test special cases of the functional coefficient model in the following linear time series regression form:

$$Y_t = \mathbf{X}_t^\top \beta(\mathbf{Z}_t) + \varepsilon_t, \quad t = 1, 2, ..., T, \tag{2.1}$$

where $\{\mathbf{X}_t \, \mathbf{Z}_t, Y_t\}$ is a $\mathbb{R}^K \times \mathbb{R}^L \times \mathbb{R}$-valued observable random sample, $\beta(\cdot) : \mathbb{R}^L \to \mathbb{R}^K$ is any measurable function of the state variables, and ε_t is an unobservable disturbance, such that $E(\varepsilon_t | \mathbf{X}_t, \mathbf{Z}_t) = 0$, implying that both $\mathbf{X}_t$ and $\mathbf{Z}_t$ are exogenous. However, this might not be true in many applications in economics and finance. When some components of $\mathbf{X}_t$ are endogenous, we use an IV, as in Cai et al. (2006). For identification, we assume $E(\varepsilon_t \mid \mathbf{X}_t, \mathbf{Z}_t) = 0$ and $\Omega(\mathbf{z}) = E[\mathbf{X}_t \mathbf{X}_t^\top \mid \mathbf{Z}_t = \mathbf{z}]$ is nonsingular, for all $\mathbf{z}$. This excludes cases in which $\mathbf{Z}_t = \mathbf{X}_t$, because $\mathbf{X}_t \mathbf{X}_t^\top$ is always singular, unless $\mathbf{X}_t$ is a scalar.

2.1 Testing the Constancy of $\beta(\mathbf{Z}_t)$

In this section, we test the constancy of the parameters as functions of the state variables. The true model is a classical linear regression; the alternative is a general functional coefficient model. We test whether the coefficients change with the state variables. When the state variable is time, we test whether structural changes exist in a time series context, as in Bai and Perron (1998), Hall et al. (2012), Perron and Yamamoto (2014), and Perron and Yamamoto (2015) for abrupt structural changes, and in Chen and Hong (2012), Chen (2015), and Fu and Hong (2019) for smooth structural changes. When the state variables are random variables, the meaning of the hypothesis varies with the economic context. For example, in asset pricing, recent studies have focused on conditional factor models (Shanken 1990;

Ferson and Harvey 1999, Cochrane 1996; Jagannathan and Wang 1996; Lettau and Ludvigson 2001; Lustig and Nieuwerburgh 2005; Santos and Veronesi 2006):

$$r_{i,t} = \alpha_i(\mathbf{Z}_t) + \beta_i(\mathbf{Z}_t)^\top \mathbf{f}_t + \varepsilon_{i,t}, t = 1, 2, ..., T, \tag{2.2}$$

where $r_{i,t}$ is the excess return on asset i at time t, and $\alpha_i(\cdot)$ and $\beta_i(\cdot)$ denote asset i's pricing error and factor loadings, respectively. Then, $\mathbf{f}_t = (f_{1,t}, f_{2,t}, ..., f_{K,t})^\top$ denotes the excess returns for mimicking portfolios.[5] The conditional asset pricing models originate from the failure of unconditional models, and are theoretically more appealing (e.g., Jensen 1968; Dybvig and Ross 1985). Under the conditional factor models, the parameters change with the investor's unobservable information set, which can be proxied by the state variables. The test for the constancy of the parameters can be viewed as a test between unconditional and conditional factor models.

A further example is provided by the wage-education model. Card (2001) proposes a random coefficient model to analyze the return on education. He argues that if we assume additive separability of education and experience, then returns on education could be understated if the returns on experience are increasing in education. As a result, Cai et al. (2006) proposed a functional coefficient wage-education model:

$$Y = \mathbf{Z}_{11}^\top \delta_0 + g_0(Z_{12}) + g_1(Z_{12})X + \varepsilon, \tag{2.3}$$

where Y is the natural logarithm of hourly wage; $\mathbf{Z}_{11}$ includes indicators for marital status, being employed by the government, union status, and other dummy variables; X is a measure of education ("Schooling"); Z_{12} is a measure of work experience; and Z_2 is an instrumental variable, namely, an index of labor market sentiment (Das et al., 2003). Therefore, testing whether the coefficients change with the state variables can be interpreted as testing whether the return on education changes with after-school work experience.

2.1.1 Hypotheses of Interest

The hypothesis of interest is

$$\mathbb{H}_{01} : \beta(\mathbf{Z}_t) = \beta_0, \text{ for unknown } \beta_0 \in \Theta, \tag{2.4}$$

versus

$$\mathbb{H}_{A1} : \beta(\mathbf{Z}_t) \neq \beta, \text{ for any } \beta \in \Theta, \tag{2.5}$$

[5]This includes popular models in the literature, such as the three-factor CAPM of Fama and French (1993), four factors of Hou et al. (2015), five factors of Fama and French (2015), and six factors of Barillas and Shanken (2017).

where $\Theta \subset \mathbb{R}^K$ is a compact parameter space.

A straightforward approach to testing $\mathbb{H}_{01}$ is to first obtain consistent estimates for $\beta(\mathbf{Z}_t)$ under the null and alternative hypothesis, and then to compare the difference between the two. Under the null hypothesis $\mathbb{H}_{01}$, the model is a classical linear regression, and the coefficients can be estimated using the OLS method. Under the alternative hypothesis $\mathbb{H}_{A1}$, the estimation of $\beta(\mathbf{Z}_t)$ can be from any nonparametric local estimation, including the Nadaraya–Watson estimator and local linear estimations. For example, the GLR test by Fan, Zhang, and Zhang (2001) compares the likelihood between parametrically null models and nonparametrically alternative models. The nonparametric MLE is typically infeasible, because the nonparametric likelihood has infinitely many parameters. Thus, they assume that the stochastic errors follow a standard normal distribution,[6] and the GLR test statistic can be based on:

$$\lambda_T = \frac{T}{2}\frac{RSS_0 - RSS_1}{RSS_1} \tag{2.6}$$

where $RSS_0 = \sum_{t=1}^{T}\left(Y_t - X_t'\widehat{\beta}_0\right)^2$ is the residual sum of squares from the OLS, and $RSS_1 = \sum_{t=1}^{T}\left(Y_t - X_t'\widehat{\beta}(Z_t)\right)^2$ is the residual sum of squares from the nonparametric estimation. Under regular conditions, as a sequence of constants $\mu_T \to \infty$ and some constant $r > 0$,

$$\frac{r\lambda_T - \mu_T}{\sqrt{2\mu_T}} \xrightarrow{d} N(0,1). \tag{2.7}$$

The GLR test is a pseudo-likelihood ratio test, because its alternative has infinite dimensions. Therefore, it does not inherit the most powerful property of the Neyman—Pearson lemma (1933). Hong and Lee (2013) argue that the GLR test does not have the optimal power property of the classical LR test. Therefore, they propose a loss function approach that compares the models under the null and alternative hypotheses by specifying a penalty for the discrepancy between the two models, as follows:

$$Q_T = \sum_{t=1}^{T} d[\widehat{m}_h(Z_t)], \tag{2.8}$$

where $\widehat{m}_h(Z_t)$ is a nonparametric estimator of $E(\varepsilon_t \mid X_t, Z_t)$, and d is a class of loss functions. Under regular conditions, let $q_T = Q_T/\widehat{\sigma}_T^2$ and $\widehat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T}[\widehat{\varepsilon}_t - \widehat{m}_h(Z_t)]^2$, where $\widehat{\varepsilon}_t$ is the residual from the OLS. Then,

$$\frac{s(K)q_T - \nu_T}{\sqrt{2\nu_T}} \xrightarrow{d} N(0,1), \tag{2.9}$$

where $s(K)$ and ν_T do not depend on nuisance parameters.

Another set of nonparametric tests compares the distance between the parametric and nonparametric

[6] This need not contain the true distribution.

estimators of $\beta(\mathbf{Z}_t)$ directly. We refer to these as Wald-type tests. For example, Li et al. (2002) use the integrated squared difference as the basis for the test:

$$I = \int \left[\widehat{\beta}(\mathbf{z}) - \widehat{\beta}\right]^\top \left[\widehat{\beta}(\mathbf{z}) - \widehat{\beta}\right] d\mathbf{z}. \tag{2.10}$$

Similarly, Chen and Hong (2012) and Ang and Kristensen (2012) compare the distance between the parametric and nonparametric estimators of $\beta(\mathbf{Z}_t)$. These are generalized Hausman tests (Hausman, 1978), because the two estimators converge to the same probability limit under $\mathbb{H}_{01}$ only.

The nonparametric tests above follow nonparametric convergence rates, owing to the nonparametric estimation step; thus, they follow a nonparametric convergence rate, which is slower than a parametric rate. This is an undesirable feature, because the low convergence rate may affect the performance of the test, particularly for a finite sample, or when the dimension of the state variables increases (the "curse of dimensionality").

At the same time, it is well known that nonparametric tests are very sensitive to the choice of bandwidth. Typically, the bandwidths are based on cross-validation. This minimizes the MSE from the nonparametric estimation step, which may not be optimal for nonparametric tests. Gao and Gijbels (2008) derive the leading terms of the size and power functions of their test statistic, and choose the bandwidth to maximize the power of a test, given a significance level. Sun, Phillips, and Jin (2008) have a similar idea. Deriving the size and power as a function of the bandwidth is sometimes difficult or infeasible in many nonparametric tests. Therefore, in practice, empirical researchers simply select a bandwidth by minimizing the MSE. However, an inappropriate choice of bandwidth may affect the asymptotic performance of the nonparametric tests, yielding misleading results.

To circumvent these undesirable features, we propose a novel consistent test based on a discrete Fourier transform.

2.1.2 Test Statistic

Under $\mathbb{H}_{01}$, $\beta(\mathbf{Z}_t)$ does not change with $\mathbf{Z}_t$. Therefore, testing $\mathbb{H}_{01}$ is equivalent to testing whether $\beta(\mathbf{Z}_t)$ is independent of $\mathbf{Z}_t$.

Tests of independence have been studied widely in the economics and statistics literature. Because the covariance and correlation capture only a linear relationship between two random variables, researchers generalize these measures to determine the nonlinear dependence structure. For example, Campbell, Lo, and MacKinlay (1997)[7] test the martingale property of a time series using

$$cov[X_t, h(X_{t-j})] = 0, \tag{2.11}$$

for $h(\cdot)$ in a pre-specified family of measurable functions. They conclude that X_t is orthogonal to X_{t-j}, in the sense that $E(X_t \mid X_{t-j}) = \mu$. However, as pointed out by Cucker and Smale (2001), the scope of the family of candidate functions of $h(\cdot)$ induces an approximation error. Therefore, the performance of the test hinges on the pre-specified family of functions.

On the other hand, Hong (1999) provides a generalized covariance approach that uses the characteristic function, or Fourier transform:

$$\sigma_j(u, v) = cov(e^{iuX_t}, e^{ivX_{t-|j|}}). \tag{2.12}$$

A generalized covariance captures the linear relationship between two variables, as well as the nonlinear dependence structure. An advantage of this method is that it avoids the class of potential functions in Campbell, Lo, and Mackinlay (1997) and, thus, circumvents the approximation error.

Therefore, we test $\mathbb{H}_{01}$ based on a generalized covariance between $\beta(\mathbf{Z}_t)$ and $\mathbf{Z}_t$: $cov[\beta(\mathbf{Z}_t), e^{iu^\top \mathbf{Z}_t}]$. To facilitate our method, define

$$A_1(u) \equiv E\left[\beta(\mathbf{Z}_t)\left(e^{iu^T\mathbf{Z}_t} - Ee^{iu^T\mathbf{Z}_t}\right)\right]. \tag{2.13}$$

The following lemma originates from the characteristic function approach of Bierens (1982) for the test of independence. Our test statistic for $\mathbb{H}_{01}$ is based on the following:

Lemma 2.1. *$A_1(u) = 0$, for all $u \in \mathbb{R}^L$, if and only if $\mathbb{H}_{01}$ holds.*

Intuitively, we can take Taylor's expansion of $e^{iu^T\mathbf{Z}_t}$ around 0: $e^{iu^T\mathbf{Z}_t} = 1 + \sum_{k=1}^{\infty} \frac{(iu^T\mathbf{Z}_t)^k}{k!}$. Now, $A_1(u) = 0$ becomes the covariance between $\beta(\mathbf{Z}_t)$ and an infinite series of the moments of $\mathbf{Z}_t$. Thus, linear uncorrelation becomes independence.[8]

In fact, $A_1(u)$ can be viewed as the coefficient of the de-meaned Fourier transform at the frequency u. Under $\mathbb{H}_{01}$, $\beta(\mathbf{Z}_t)$ is a constant function of $\mathbf{Z}_t$ and, thus, the Fourier coefficients should all be equal to zero. The de-meaned $e^{iu'\mathbf{Z}_t}$ is used because we only care whether $\beta(\mathbf{Z}_t)$ is a constant; we do not need to know the value of that constant. Lemma 1 transforms the hypothesis of interest from a time or state variable domain into the frequency domain. The nonparametric estimation of $\beta(\mathbf{Z}_t)$ now becomes

[7]An even older work is that of Granger and Teräsvirta (1993), who determine a maximum correlation coefficient.

[8]A more rigorous proof takes a Fourier transform of $e^{iu^T\mathbf{Z}_t}$; see Bierens (1982).

a parametric estimation of the Fourier coefficient in the frequency domain, which improves the efficiency of our test.

Lemma 1 suggests that we can test $\mathbb{H}_{01}$ based on the sample analog of $A_1(u)$:

$$\widehat{A}_1(u) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\beta}(\mathbf{Z}_t)\left[e^{iu^T\mathbf{Z}_t} - \frac{1}{T}\sum_{s=1}^{T}e^{iu^T\mathbf{Z}_s}\right], \tag{2.14}$$

where $\widehat{\beta}(\mathbf{Z}_t)$ is any consistent nonparametric estimator of $\beta(\mathbf{Z}_t)$.

Various smoothing techniques can be applied here, including local smoothing such as the Nadaraya-Watson estimator, splines, and orthogonal series methods. In this study, we adopt the local linear estimation proposed by Cai, Fan, and Yao (2000). Because we need to demonstrate the performance of our method as the dimension of the state variable increases, we generalize the estimator to the case where $\mathbf{Z}_t$ is multidimensional. The local linear estimator is defined as $\widehat{\beta}(\mathbf{z}_0) = \widehat{\beta}$, where $\{(\widehat{\beta}(\mathbf{z}_0), \widehat{\beta}'(\mathbf{z}_0)\}$ minimizes the sum of the weighted squares:

$$\sum_{t=1}^{T}\left[Y_t - \left(\beta(\mathbf{z}_0)', vec(\beta'(\mathbf{z}_0))'\right)\begin{pmatrix}\mathbf{X}_t \\ (\mathbf{Z}_t - \mathbf{z}_0)\otimes\mathbf{X}_t\end{pmatrix}\right]^2\prod_{i=1}^{d}K_h(Z_{ti} - z_{0i}), \tag{2.15}$$

where $K_h(x) = h^{-1}K(x/h)$, $K(\cdot)$ is a kernel function on $\mathbb{R}$, and $h > 0$ is a bandwidth that controls the degree of smoothing in the estimation. In addition, $\beta'(\mathbf{z}_0)$ is the Jacobian matrix of $\beta(\mathbf{z}_0)$, and $\otimes$ is the Kronecker product. As pointed out by Cai (2009), the estimator $\{(\widehat{\beta}(\mathbf{z}_0), \widehat{\beta}'(\mathbf{z}_0))\}$ can be obtained by performing an OLS on the following model:

$$\left[\prod_{i=1}^{d}K_h(Z_{ti} - z_{0i})\right]^{1/2} Y_t = \left[\prod_{i=1}^{d}K_h(Z_{ti} - z_{0i})\right]^{1/2}\left[\beta(\mathbf{z}_0)'\mathbf{X}_t + vec(\beta'(\mathbf{z}_0))'(\mathbf{Z}_t - \mathbf{z}_0)\otimes\mathbf{X}_t\right] + u_t. \tag{2.16}$$

The sample analog $\widehat{A}_1(u)$ can be viewed as a discrete Fourier transform of $\widehat{\beta}(\mathbf{Z}_t)$ on the random state variable $\mathbf{Z}_t$. At fixed frequency u, it has a V-statistic structure of $\widehat{A}(u)$, which can be approximated by a U-statistic under regularity conditions. As in Powell et al. (1989), although this sample analog invokes the nonparametric estimation, the data points used in the kernel regressions of $\widehat{\beta}(\mathbf{Z}_t)$ will be reused.

Traditionally, the choice of h should satisfy $Th^{d_z} \to \infty$, where the optimal bandwidth $h = c{*}T^{-\frac{1}{d_z+4}}$ minimizes the MSE of the estimation. The constant c is typically set using Silver's rule of thumb, or using a data-driven method, such as cross validation. However, the optimal bandwidth for estimation is not necessarily the best choice for testing, and the bandwidth does impact the performance of the test. In a hypothesis test, the central problem is the trade-off between Type-I and Type-II errors. Gao and Gijbels (2008) derive the leading term of their test statistics, and choose $h = \arg\max_{h\in B_n(\alpha)}\beta_n(h)$, where $B_n(\alpha) = h : \alpha - c_{\min} < \alpha_n(h) < \alpha + c_{\min}$, for some prespecified small constant $c_{\min} \in (0, \alpha)$. Here, $\alpha_n(h)$ and $\beta_n(h)$ represent the size and power, respectively, of the test, given h. The idea is to choose a data-driven bandwidth, with a bootstrap, that maximizes power, while controlling for size within an acceptable range. Sun, Phillips, and Jin (2008) provide a similar approach. In this study, we select the bandwidth by rule of thumb. We leave this issue to further research.

AAA Our approach circumvents this problem due to a As pointed out by Powell, Stock, and Stoker (1989), each data point is used in the estimation of several kernel regressions, and the overlaps render a faster convergence rate. This structure allows us to properly account for the "overlaps" in the local linear estimators by imposing an additional smoothing step. The cost of this approach is that we must check for all $u \in \mathbb{R}^K$, rather than just a subset of $\mathbb{R}^K$.

Under $\mathbb{H}_0$, $A_1(u) = 0$, for any given u. Under $\mathbb{H}_{A1}$, $A_1(u)$ is a nonzero function of u and, thus, determines the power of a test, based on $\widehat{A}_1(u)$. Therefore, we can test $\mathbb{H}_{01}$ by measuring the deviation from zero using the following sample quadratic forms:

$$\widehat{Q}_1 = T \int_{\mathbb{R}^L} \|\widehat{A}_1(u)\|^2 W(u) du, \tag{2.17}$$

where $W : \mathbb{R}^L \to \mathbb{R}^+$ is a nonnegative symmetric weighting function of u. The introduction of $W(u)$ allows us to consider many points for u.

Because Z_t is a $L \times 1$ vector, the test statistic (2.17) involves d_z-dimensional integration, which is usually calculated using numerical integration or approximated using simulation methods. When the dimension Z_t is large, the aforementioned methods would be tremendously computationally costly. We can approximate integration using a finite number of grid points for u. This could reduce the computational cost, but may lead to power loss. Another way to avoid high-dimensional numerical integration is to integrate (2.17) out analytically by choosing some suitable weighting function. Here, we follow Hong, Wang, and Wang (2017), and use the following weighting function, based on the joint independent normal density:

$$W_N(u) = \Pi_{k=1}^{q} \frac{1}{\sqrt{2\pi\xi_k}} exp(-\frac{u_k^2}{2\xi_k^2}), \tag{2.18}$$

where ξ_k can be viewed as the standard deviation of $W_N(u)$ for dimension k. With this weighting function and using the identity:

$$\int_{\mathbb{R}^{d_z}} cos\left(\sum_{i=1}^{d_z}\right) exp\left(-\frac{|u|^2}{2}\right) du = (2\pi)^{\frac{d_z}{2}} exp\left(-\frac{1}{2}|x|^2\right). \tag{2.19}$$

The quadratic form (2.17) can be written as:

$$\widehat{Q}_W = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\widehat{\beta}(Z_s)'V_{st}\widehat{\beta}(Z_t), \tag{2.20}$$

where $V_{st} = e^{-\frac{1}{2}|Z_s-Z_t|^2} - \frac{1}{T}\sum_{l=1}^{T} e^{-\frac{1}{2}|Z_l-Z_t|^2} - \frac{1}{T}\sum_{l=1}^{T} e^{-\frac{1}{2}|Z_s-Z_l|^2} + \frac{1}{T^2}\sum_{m,n=1}^{T} e^{-\frac{1}{2}|Z_m-Z_n|^2}$. The computation of the test statistic $\widehat{Q}$ requires no numerical integration, regardless of the dimension of Z_t. Note that another type of weighting function, based on the Laplace density function, can also avoid numerical integration:

$$W_L(u) = \Pi_{k=1}^{q}\frac{\lambda_k}{2}e^{-\lambda_k|u_k|}, \tag{2.21}$$

where λ_k denotes a scale parameter. Indeed, the choice of weighting function affects the asymptotic power of the test; therefore, those tests that cannot avoid numerical integration may outperform those based on the aforementioned weighting. This comparison needs to be checked using simulations. However, the difference is not significant compared with that between GLR and our test. Note too that the quadratic form (2.17) is not the only way to measure the deviation of $\widehat{A}(u)$ away from zero. For example, $sup_{u\in\mathbb{R}^{d_z}}|\widehat{A}(u)|$ is a classical method in statistics. Chen and Kato (2019) provide a finite-sample approximation to the U-process supremum. We leave these issues to further research.

2.2 Testing a parametric form of $\beta(\mathbf{Z}_t)$

In this section, we are interested in testing a particular parametric form of $\beta(\mathbf{Z}_t)$. A simple case of this type of hypothesis is to test a certain value of the coefficients: $\beta(\mathbf{Z}_t) = \beta_0$ for a pre-specified β_0. This differs from $\mathbb{H}_{01}$ since β_0 is known. In this case, the mean level of the Fourier coefficient is not irrelevant and it leads us to use the regular Fourier transform rather than the demeaned one in $\widehat{A}_1(u)$.

A more complicated case is to test a particular parametric specification of $\beta(\mathbf{Z}_t)$. Depending on the function form in the null hypothesis, the test in this section can be applied to check specification for many classical nonlinear time series models:

1. If $X_t = (Y_{t-1},\cdots,Y_{t-k})'$, $Z_t = (X_{t-d})$ and $\beta(\cdot)$ being the indicator function, then model (2.1) degenerates to the threshold autoregressive model (TAR) proposed by Tong (1990):

   $$Y_t = \phi_1^{(i)}Y_{t-1} + \cdots + \phi_p^{(i)}Y_{t-p} + \varepsilon_t^{(i)}, \quad \text{if } x_{t-d}\in\Omega_i, \quad i = 1,2,\cdots,R$$

   where $\{\Omega_i\}$ form a non-overlapping partition of real line.

2. If $X_t = (Y_{t-1},\cdots,Y_{t-k})'$ and $\beta_j(Z_t) = \alpha_j + (\beta_j+\gamma_j Z_t)\exp(-\theta_j Z_t^2)$ with $Z_t = X_{t-d}$, then model

(2.1) becomes the generalized exponential autoregressive (EXPAR) model proposed by Haggan and Ozaki (1981) and Ozaki (1982):

$$Y_t = \sum_{j=1}^{k} \{\alpha_j + (\beta_j + \gamma_j Y_{t-d}) \exp(-\theta_j X_{t-d}^2)\} X_{t-j} + \varepsilon_t$$

where $\theta_j \geq 0$ for $j = 1, 2, \cdots, K$.

3. If $X_t = (Y_{t-1}, \cdots, Y_{t-k})'$ and $\beta(Z_t) = \phi_1 G(Z_t; \gamma, c) + \phi_2(1 - G(Z_t; \gamma, c))$, then model (2.1) becomes the smooth transition AR model considered by Teräsvirta (1994):

$$Y_t = \sum_{j=1}^{k} \{\phi_{1j}\} G(Z_t; \gamma, c) Y_{t-j} + \{\phi_{2j}\} \left[1 - G(Z_t; \gamma, c)\right] Y_{t-j} + \varepsilon_t$$

where $G(Z_t; \gamma, c) = \{1 + \exp[-\gamma(Z_t - c)]\}^{-1}$ being the logistic function.

4. If $Z_t \sim U(0, 1)$, then model (2.1) could be regard as a linear quantile regression (Koenker, 2005):

$$Q_Y(\tau | X_t) = X_t' \beta(\tau), \qquad \tau \sim i.i.d. U[0, 1]$$

where $\beta(\tau)$ measures the relationship between X_t and Y_t at quantile τ.

Estimation of $\beta(\cdot)$ has been extensively investigated in the literature. Fan and Zhang (1999) provide an innovative two-step method for independent samples via local polynomial estimation. Cai et al. (2000) adopt local linear regression technique to estimate the coefficient functions $\beta(\cdot)$ in the time series context. However, nonparametric estimation of $\beta(\cdot)$ has some undesired features. For instance, when the dimension of Z_t is large, one will encounter the "curse of dimensionality" problem. As a result, the asymptotic convergence rate of the estimated parameters and coefficient functions is much slower than the parametric convergence rate. Moreover, it is hard to provide economic interpretations of the dependence between the unknown regression coefficients β on the state variable Z_t. Thus, parametric modeling of $\beta(\cdot)$ is quite appealing in practice. For example, Tong (1990) and Hansen (2000) study the theoretical properties of threshold model, while Teräsvirta (1994, 1998) and Granger and Teräsvirta (1993) explore asymptotic results for the smooth transition models. Therefore, it is critical to check whether the parametric form is correctly specified. This leads us to focus on the following hypothesis of interest.

2.2.1 Hypothesis of Interest

The setting of model 2.1 is fairly general. The hypothesis of interest can be written as follows:

$$\mathbb{H}_{02} : \mathbf{R}\beta(\mathbf{Z}_t) = \mathbf{R}\beta(\mathbf{Z}_t, \theta_0), \text{ for some } \theta_0 \in \Theta, \tag{2.22}$$

where $\Theta \in \mathcal{R}'$ is a parameter space, against

$$\mathbb{H}_{A2} : \mathbf{R}\beta(\mathbf{Z}_t) \neq \mathbf{R}\beta(\mathbf{Z}_t, \theta), \quad \forall\theta \in \Theta \tag{2.23}$$

where $\Theta \subset \mathbb{R}^{d_\theta}$ is a compact parameter space. Here, $\mathbf{R}$ is a full rank $m \times K$ matrix, and m represents the number of restrictions. We can form the matrix $\mathbf{R}$ to fit our hypothesis. When $\mathbf{R}$ is the identity matrix, $\mathbb{H}_{01}$ becomes $\beta(\mathbf{Z}_t) = \beta_0$, that is, whether all coefficients are changing with the state variables. This can be regarded as a model specification test for the classical linear regression models against the more general functional coefficient models. More importantly, the choice of $\mathbf{R}$ allows us to test part (or any linear combination) of the random coefficients, while maintaining flexibility in the specifications of the other coefficients. This is important in many circumstances, for example, to test the conditional asset pricing models in 2.2, arbitrage-free pricing theorem (APT) implies the pricing error $\alpha_i(\cdot) = 0$ if there exists risk free asset. Therefore, testing a conditional asset pricing model is equivalent to testing a functional coefficient model by letting $\mathbf{R} = (1, \mathbf{0})$. On the other hand, if the market is without risk free asset, Shanken (1985) and Jagannathan et al. (2009) prove that the condition becomes $\alpha_i(\mathbf{Z}_t) = \gamma(1 - \beta_i(\mathbf{Z}_t))$ where γ is an unknown zero-beta rate. This can be conducted by letting $\mathbf{R} = (1, \gamma)$.

2.2.2 Test Statistic

The test of $\mathbb{H}_{01}$ uses the de-meaned Fourier transform $A_1(u) = 0$. When testing a certain parametric form, we need to check whether $\mathbf{R}\beta(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \theta_0) = 0$. Therefore, we use the Fourier transform rather than the de-meaned transform. Define the following:

$$A_2(u) \equiv E\left[\left(\mathbf{R}\beta(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \theta_0)\right) e^{iu^T \mathbf{Z}_t}\right] \tag{2.24}$$

The test to $\mathbb{H}_{02}$ is based on the following lemma:

Lemma 2.2. *$A_2(u) = 0$ for all $u \in \mathbb{R}^L$ if and only if $\mathbb{H}_{02}$ holds.*

Compared to Lemma 2.1 in section 2.1, besides checking whether the difference $\mathbf{R}\beta(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \theta_0)$ is a constant, we also need to know whether this constant is equal to zero. Therefore, we use the Fourier transform instead of the de-meaned one.$A_2(u)$ can be viewed as the coefficient of the (un-demeaned) Fourier transform of the difference $\mathbf{R}\beta(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \theta_0)$ at the frequency u. Under $\mathbb{H}_{02}$,

$\mathbf{R}\beta(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \theta_0) = 0$ and thus all the Fourier coefficients should be equal to zero. Lemma 2 transforms the hypothesis $\mathbb{H}_{02}$ from the time or state variables' domain into the frequency domain. Again, the estimation of the the Fourier coefficient in $A_2(u)$ follows a parametric convergence rate at each frequency. This gives us the improvement in efficiency of our test.

Lemma 2 suggests that we can test $\mathbb{H}_{02}$ based on the sample analog of $A_2(u)$:

$$\widehat{A}_2(u) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{R}[\widehat{\beta}(\mathbf{Z}_t) - \beta(\mathbf{Z}_t, \widehat{\theta})]e^{iu^T\mathbf{Z}_t}, \tag{2.25}$$

where $\beta(\mathbf{Z}_t, \widehat{\theta})$ is a parametric estimation of the null model in $\mathbb{H}_{02}$.

In the simple case when then hypothesis of interest is $\beta(\mathbf{Z}_t) = \beta_0$ for a pre-specified β_0, there is no need to estimate $\beta(\mathbf{Z}_t, \theta_0)$. We can simply replace $\beta(\mathbf{Z}_t, \widehat{\theta})$ by β_0 in $\widehat{A}_2(u)$. For more general cases, estimation of $\beta(\mathbf{Z}_t, \theta_0)$ depends on the parametric specification in the null hypothesis $\mathbb{H}_{02}$. For example, Tong (1990) and Hansen (2000) use a profile approach to estimate the regression coefficient and the threshold values in the TAR models. The coefficient estimator follows a $\sqrt{T}$ convergence rate while the threshold estimator has a super T rate. It is important to point out that the way of parametric estimation of $\beta(\mathbf{Z}_t, \theta_0)$ does not change the asymptotic results of $\widehat{A}_2(u)$ as long as it is a consistent estimator with a faster convergence rate than the nonparametric estimation.

We can test $\mathbb{H}_{02}$ by measuring the deviation from zero via the following sample quadratic forms:

$$\widehat{Q}_2 = T\int_{\mathbb{R}^L}\|\widehat{A}_2(u)\|^2 W(u)du, \tag{2.26}$$

where $W : \mathbb{R}^L \to \mathbb{R}^+$ is a non-negative symmetric weighting function of u. The introduction of $W(u)$ allows us to consider many points for u.

To avoid numerical integration, we can use the normal weighting function in equation 2.20 and the test statistic in 2.26 can be written as

$$\widehat{Q}_{W2} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}[\mathbf{R}\widehat{\beta}(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \widehat{\theta})]'V_{st}[\mathbf{R}\widehat{\beta}(\mathbf{Z}_t) - \mathbf{R}\beta(\mathbf{Z}_t, \widehat{\theta})] \tag{2.27}$$

where $V_{st} = e^{-\frac{1}{2}||\mathbf{Z}_s - \mathbf{Z}_t||^2}$.

3 Asymptotic Theory

In this section, we will derive the asymptotic null distribution of the test statistic $\widehat{Q}_1$ and $\widehat{Q}_2$ and investigate their asymptotic local power property. We also introduce bootstrap procedures to improve the finite sample performance of the test. Throughout this section, denote "$\xrightarrow{p}$", "$\xrightarrow{d}$" and "$\Rightarrow$" as convergence in probability, convergence in distribution and weak convergence respectively.

3.1 Assumptions

To derive the asymptotic distribution of our test statistics, we first impose some regularity assumptions.

Assumption 1. Let$(\Omega, \mathbb{F}, P)$ be a complete probability space. The stochastic process $\{\mathbf{X}_t, \mathbf{Z}_t\}$, is strictly stationary absolutely regular on $\mathbb{R}^{K+L}$ with β-mixing coefficients satisfying $\sum_{j=1}^{\infty} j^2 \beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < \frac{1}{3}$.

Assumption 2. (a) The joint density $f_{\mathbf{Z}}(\mathbf{z})$ of $\mathbf{Z}_t$ is positive, bounded and continuously differentiable in $z \in \mathbb{G} \subset \mathbb{R}^L$ up to order r, where $\mathbb{G}$ is a compact support of Z_t, (b) The joint density $f_{X,Z}(x, z)$ is positive, bounded and continuously differentiable at fixed $\mathbf{z}$ over the support of $\mathbf{Z}_t$.

Assumption 3. $\{\varepsilon_t\}$ is a martingale difference sequence, that is, $E(\varepsilon_t|I_{t-1}) = 0$ where $I_{t-1} = \{\mathbf{X}_t, \mathbf{Z}_t, \mathbf{X}_{t-1}, \mathbf{Z}_{t-1}, \varepsilon_{t-1}, ...\}$. In addition, $E|\Omega(\mathbf{Z}_t)_{jp}^{-1} X_{tp}\varepsilon_t|^4 < \infty$ for $j, p = 1, ..., K$. Also $v(\mathbf{z}) \equiv E(\varepsilon_t^2 \mid z)$ is continuous in $\mathbf{z}$.

Assumption 4. $\kappa : \mathbb{R}^L \to \mathbb{R}^+$ is a product of some univariate kernel K, i.e., $\kappa(u) = \prod_{i=1}^{L} K(u_i)$, where $K : \mathbb{R} \to \mathbb{R}^+$ satisfies the Lipschitz condition and is symmetric, bounded, and squre-integrable with $\int_{-\infty}^{\infty} u^r K(u)du = C_r < \infty$ for some $r \geq 2$ and $\int_{-\infty}^{\infty} u^l K(u)du = 0$ for $l = 1, ..., r-1$.

Assumption 5. (i) $\beta(\cdot;\theta)$ is measurable function and twice continuously differentiable with respect to $\theta \in \Theta$; $\theta_* \in \mathrm{int}\Theta$ is the probability limit of NLS estimator, such that $\sqrt{T}(\hat{\theta} - \theta_*) = O_P(1)$; (ii) $\beta^{(1)}(Z_t;\theta) \equiv \frac{d\beta(Z_t;\theta)}{d\theta}$ is a $k \times d$ Jacobian matrix such that $\sup_{\theta\in\Theta} E(\|\beta^{(1)}(Z_t;\theta)\|^{4+\delta}) < C$ for some $\delta > 0$; (iii) Let $\beta_j(Z_t;\theta)$ be the jth component in $\beta(Z_t;\theta)$, where $1 \leq j \leq k$, then $\beta_j^{(2)}(Z_t;\theta) \equiv \frac{d^2\beta_j(Z_t;\theta)}{d\theta d\theta'}$ is a $d \times d$ Hessian matrix such that $\sup_{\theta\in\Theta} E(\|\beta_j^{(2)}(Z_t;\theta)\|^{4+\delta}) < C$ for some $\delta > 0$.

Assumption 6. $W : \mathbb{R}^L \to \mathbb{R}^+$ is a nonnegative symmetric integrable function with $\int_{\mathbb{R}^L} \|u\|^4 W(u)du < \infty$.

Assumption 1 imposes regularity conditions on the DGP. Note that, in the literature, there are cases where Z_t and X_t can be nonstationary; see Xiao (2009) for X_t, and Cai, Li, and Park (2009) and Juhl (2005) for Z_t. We leave these issues to further research. The β-mixing condition restricts the degree of temporal dependence in (X_t, Y_t, Z_t), which is generally adopted in the nonparametric time series literature; see, for example, Hjellvik et al. (1998), Su and White (2007,2008), and Chen and Hong (2010).

Assumption 2 imposes regular conditions on the distribution of the marginal density of the state variable and on the joint density of the state variable and the regressors. These are common in the local smoothing literature; see Cai, Fan, and Yao (2000) for local linear estimation. Assumption 2(b) requires only the smoothness condition of $f_{X,Z}(x,z)$ on the state variables, assuming the regressors can be discrete.

Assumption 3 restricts the model disturbance ε_t to be serially uncorrelated. This simplifies our discussion, and can be modified easily to more general cases. In addition, we allow for conditional heteroskedasticity with the smoothness condition on $v(z)$. This is important for the boundness condition in the U-statistic kernel, and is similar to Powell, Stock, and Stoker (1989).

Assumptions 4 and 5 imposes conditions on the kernel and the weighting functions on the frequency. With $r \geq 2$, we allow, but do not require using higher-order kernels. Assumption 4 is a mild condition on $W(u)$, ensuring the existence of the integral in (2.17). Any density function with a finite fourth-order moment satisfies this condition.

3.2 Asymptotic Null Distribution

By construction, $\widehat{A}_1(u)$ is a second order non-degenerate U-process in Serfling (1980), which is based on Hoeffding (1948). Under the regularity conditions, we first state that $\sqrt{T}\widehat{A}_1(u)$ converges to a complex-valued Gaussian process.

Proposition 3.1. *Suppose Assumptions 1 - 4 hold, $Th^L \to \infty$ as $T \to \infty$, then under $\mathbb{H}_{01}$*

$$\sqrt{T}\widehat{A}_1(u) \Rightarrow \mathcal{G}_1(u)$$

where $\mathcal{G}_1(u)$ is a complex-valued Gaussian process with 0 mean and covariance kernel as

$$\begin{aligned} \mathcal{K}_1(u_1,u_2) &\equiv cov(\mathcal{G}_1(u_1), \mathcal{G}_1^*(u_2)) \\ &= 4E\left[r_1(\xi_t,u_1)r_1^*(\xi_t,u_2)'\right] \end{aligned}$$

and

$$r_1(\xi_t,u) \equiv \frac{1}{2}c_0\Omega(\boldsymbol{Z}_t)^{-1}\psi(u,\boldsymbol{Z}_t)\left[\nu_X(\boldsymbol{Z}_t)\nu_Y(\boldsymbol{Z}_t) + \boldsymbol{X}_t Y_t\right]$$

and $\nu_X(\boldsymbol{Z}_t) \equiv E(\boldsymbol{X}_t \mid \boldsymbol{Z}_t)$, $\nu_Y(\boldsymbol{Z}_t) \equiv E(Y_t \mid \boldsymbol{Z}_t)$. And $*$ denotes the complex conjugate.

Proposition 3.1 follows from Powell, Stock and Stoker (1989) to derive an approximation of the asymptotic variance. It is well known that a nonparametric estimator is biased. However, $\widehat{A}_1(u)$ is not affected by bias, because the expectation of the nonparametric estimator has been subtracted. Intuitively, the bias does not matter when testing the constancy of the parameters because we do need to specify these unknown constants.

The statistic $\widehat{A}_2(u)$ involves parametric estimation of $\beta(\mathbf{Z}_t, \theta_0)$ which is of unknown form. Therefore, we impose assumption 5 to restrict the convergence rate of parametric estimator of θ_0. Most parametric estimation satisfies this condition as long as it has a convergence rate faster than $\sqrt{T}$. Thus we state that $\sqrt{T}\widehat{A}_2(u)$ converges to a complex-valued Gaussian process.

Proposition 3.2. *Suppose Assumptions 1 - 5 hold, and $h = cT^{-\lambda}$ for $\frac{1}{2L+2} < \lambda < \frac{1}{L}$, where $0 < c < \infty$, then under $\mathbb{H}_{02}$*

$$\sqrt{T}\widehat{A}_2(u) \Rightarrow \mathcal{G}_2(u)$$

where $\mathcal{G}_1(u)$ is a complex-valued Gaussian process with 0 mean and covariance kernel as

$$\begin{aligned}\mathcal{K}_2(u_1, u_2) &\equiv cov(\mathcal{G}_2(u_1), \mathcal{G}_2^*(u_2)) \\ &= E\left[4r_1(\xi_t, u_1)r_1^*(\xi_t, u_2)' + r_2(\xi_t, u_1)r_2^*(\xi_t, u_2)\right]\end{aligned}$$

where

$$r_2(\xi_t, u) = E\left\{I_l e^{iu'Z_t} - E\left[\beta^{(1)}(Z_t;\theta_*)'X_tX_t'\beta^{(1)}(Z_t;\theta_*)e^{iu'Z_t}\right]E\left[\beta^{(1)}(Z_t;\theta_*)'X_tX_t'\beta^{(1)}(Z_t;\theta_*)\right]^{-1}\right\}\beta^{(1)}(Z_t;\theta_*)'X_t.$$

The bandwidth h is chosen to remove the bias of the nonparametric estimation of $\beta(\mathbf{Z}_t)$. We can also calculate the expression of the bias and subtract it directly.

After providing the test statistics $\widehat{A}_1(u)$ and $\widehat{A}_2(u)$, by continuous mapping theorem, we now show the asymptotic null distributions of $\widehat{Q}_1$ and $\widehat{Q}_2$.

Theorem 3.1. *Suppose Assumptions 1 - 4 and 6 hold, $Th^L \to \infty$ as $T \to \infty$, then under $\mathbb{H}_{01}$*

$$\widehat{Q}_1 \xrightarrow{d} Q_1 \equiv T\int_{R^L} \|\mathcal{G}_1(u)\|^2 W(u)du$$

Theorem 3.2. *Suppose Assumptions 1 - 6 hold, and $h = cT^{-\lambda}$ for $\frac{1}{2L+2} < \lambda < \frac{1}{L}$, where $0 < c < \infty$, then under $\mathbb{H}_{02}$*

$$\widehat{Q}_2 \xrightarrow{d} Q_2 \equiv T\int_{R^L} \|\mathcal{G}_2(u)\|^2 W(u)du$$

Theorem 3.1 and theorem 3.2 provide the asymptotic distribution of $\widehat{Q}_1$ under $\mathbb{H}_{01}$ and $\widehat{Q}_2$ under $\mathbb{H}_{02}$. Because our test statistic does not have a standard asymptotic distribution, we need a resampling method to obtain the critical value of Q_1 and Q_2. We adopt a wild bootstrap procedure to accommodate the heteroskedasticity in the model:

Step 1 Obtain the coefficient estimator $\widehat{\beta}_0$ from an OLS of the null linear regression model, and the nonparametric estimator $\widehat{\beta}(Z_t)$, for all $t = 1, ..., T$.

Step 2 Compute the $\widehat{Q}$ statistic and the residual $\widehat{\varepsilon}_t = Y_t - X_t'\widehat{\beta}(Z_t)$ from the nonparametric model.

Step 3 Draw a bootstrap error $\widehat{\varepsilon}_t^* = \widehat{\varepsilon}_t v_t$, where $\{v_t\}_{t=1}^T$ is an $i.i.d.N(0,1)$ sequence, and compute $Y_t^* = X_t'\widehat{\beta}_0 + \widehat{\varepsilon}_t^*$. This forms a wild bootstrap sample $\{X_t, Z_t, Y_t^*\}_{t=1}^T$.

Step 4 Use the wild bootstrap sample $\{X_t, Z_t, Y_t^*\}_{t=1}^T$ to compute a bootstrap statistic $\widehat{Q}^*$, using the same kernel $K(\cdot)$ and the same bandwidth h as in step 2.

Step 5 Repeat steps 3 and 4 B times, where B is a large number. We then obtain a collection of bootstrap test statistics, $\{\widehat{Q}_l^*\}_{l=1}^B$.

Step 6 Compute the bootstrap P-value $P^* = B^{-1}\sum_{l=1}^B \mathbf{1}(\widehat{Q} < \widehat{Q}_l^*)$. Reject H_0 at a prespecified significance level α if and only if $P^* < \alpha$.

3.3 Asymptotic Power

Next, we show the asymptotic power of our test statistic:

Theorem 3.3. *Suppose Assumptions 1 - 4 and 6 hold, $Th^L \to \infty$ as $T \to \infty$. Then under $\mathbb{H}_{A1}$, for any sequence of nonstochastic constants $\{c_T = o(T)\}$, as $T \to \infty$,*

$$P(\widehat{Q}_1 > c_T) \quad \to \quad 1$$

Theorem 3.4. *Suppose Assumptions 1 - 6 hold, and $h = cT^{-\lambda}$ for $\frac{1}{2L+2} < \lambda < \frac{1}{L}$, where $0 < c < \infty$. Then under $\mathbb{H}_{A2}$, for any sequence of nonstochastic constants $\{c_T = o(T)\}$, as $T \to \infty$,*

$$P(\widehat{Q}_2 > c_T) \quad \to \quad 1$$

These result show the power of $\widehat{Q}_1$ and $\widehat{Q}_2$ against fixed alternatives that approach one as $T \to \infty$. Therefore, $\widehat{Q}_1$ is consistent against both abrupt structural breaks and smooth structural changes, and $\widehat{Q}_2$ is consistent against general functional coefficient models.

We now investigate the local power property of our tests and compare it with existing consistent tests in the literature. Consider the following local alternatives under $\mathbb{H}_{Al}$:

$$\mathbb{H}_{Al} : \mathbf{R}\beta(\mathbf{Z}_t) = \mathbf{R}\beta(\mathbf{Z}, \theta_0) + \alpha(T)\delta(\mathbf{Z}_t), \tag{3.1}$$

where $\alpha(T) \to 0$ as $T \to \infty$. Here, $\mathbf{R}\beta(\mathbf{Z}, \theta_0)$ is equal to an unknown β_0 for $\widehat{Q}_1$.

Theorem 3.5. *Suppose Assumptions 1 - 4 and 6 hold, $Th^L \to \infty$ as $T \to \infty$. Then under $\mathbb{H}_{Al}$ and $\alpha(T) = T^{-\frac{1}{2}}$,*

$$\widehat{Q}_1 \xrightarrow{d} \int |G(u) + \xi(u)|^2 W(u) du. \tag{3.2}$$

where the noncentrality parameter process

$$\xi(u) = cov[\delta(\boldsymbol{Z}_t), e^{iu^{\mathbf{T}}\boldsymbol{Z}}]. \tag{3.3}$$

Theorem 3.6. *Suppose Assumptions 1 - 6 hold, and $h = cT^{-\lambda}$ for $\frac{1}{2L+2} < \lambda < \frac{1}{L}$, where $0 < c < \infty$. Then under $\mathbb{H}_{Al}$ and $\alpha(T) = T^{-\frac{1}{2}}$,*

$$\widehat{Q}_2 \xrightarrow{d} \int |G(u) + \xi(u)|^2 W(u) du, \tag{3.4}$$

where the noncentrality parameter process

$$\xi(u) = cov[\delta(\boldsymbol{Z}_t), e^{iu^{\mathbf{T}}\boldsymbol{Z}}]. \tag{3.5}$$

Theorem 3.5 and 3.6 show that our tests $\widehat{Q}_1$ and $\widehat{Q}_2$ have nontrivial power against a class of local alternatives with the parametric rate $\alpha(T) = T^{-\frac{1}{2}}$. In contrast, the GLR of Fan, Zhang, and Zhang (2001), loss function approach of Hong and Lee (2013), and Wald-type test of Chen and Hong (2012) can detect only a class of local alternatives at a rate of $T^{-\frac{4}{9}}$ under to optimal rate of bandwidth $T^{-\frac{2}{9}}$. This advantage requires an additional smoothing step using the Fourier transform. The cost incurred is that the asymptotic distributions of our tests $\widehat{Q}_1$ and $\widehat{Q}_2$ are not pivotal. At the same time, the parametric test in Fu and Hong (2019) cannot handle $\mathbb{H}_{02}$, owing to the lack of a nonparametric estimation.

4 Monte Carlo Study

We now conduct a Monte Carlo study to assess the finite-sample performance of our test for a specification of functional coefficient models. We compare our model with the GLR test of Fan, Zhang, and Zhang (2001) and the loss function test of Chen and Hong (2012).

For simplicity, we consider the state variables $\mathbf{Z}_t \sim i.i.d.\, U[0,1]$.[9] To demonstrate that our test outperforms the existing tests as the dimension of the state variables increases, for each DGP below, we allow the dimension of $\mathbf{Z}_t$ to vary between one and three.[10]. Furthermore, to examine the robustness of

[9] We also tried other parametric distributions of $\mathbf{Z}_t$, but the performance of the test is not affected significantly. We do not report the results here, for brevity.

[10] The "curse of dimensionality" is known to be serious when the smoothing takes place on more than three variables.

our test, for each DGP below, we consider different specifications for the error terms ε_t, as in Hong and Lee (2013): (i) $i.i.d.\, N(0,1)$;

(ii) ARCH errors:

$$\begin{aligned} \varepsilon_t &= \sqrt{h_t}\nu_t, \\ h_t &= 0.2 + 0.5\varepsilon_{t-1}^2, \\ \nu_t &\sim i.i.d\, N(0,1); \end{aligned}$$

(iii) Conditional heteroskedasticity Errors:

$$\begin{aligned} \varepsilon_t &= \sqrt{h_t}\nu_t, \\ h_t &= 0.2 + 0.5X_t^2, \\ \nu_t &\sim i.i.d\, N(0,1). \end{aligned}$$

To examine the size performance of $\widehat{Q}_1$, we consider the following DGPs, as in Hong and Lee (2013):

DGP S1 (Linear model)

$$\begin{cases} Y_t = 1 + X_t + \varepsilon_t, \\ X_t = 0.5X_{t-1} + v_t, \\ v_t \sim i.i.d.N(0,1). \end{cases}$$

To examine the size performance of $\widehat{Q}_2$, we consider the following DGPs:

DGP S2 (nonlinear model without intercept):

$$\begin{cases} Y_t = (1 + 0.5\mathbf{Z}_t)X_t + \varepsilon_t, \\ X_t = 0.5X_{t-1} + v_t, \\ v_t \sim i.i.d.N(0,1). \end{cases}$$

DGP S3 (STAR(1)):

$$X_t = 0.5X_{t-1} + \Phi\left(\frac{X_{t-1}-1}{2}, ..., \frac{X_{t-L}-1}{2}\right) X_{t-1} + X_{t-2} + u_t,$$

where Φ is the CDF of the multinomial distribution.

DGP S1 satisfies the null hypothesis of constant coefficients, and can be used to check the size of our test statistic $\widehat{Q}_1$. For brevity, we check only the constancy of the entire set of parameters. DGP

Therefore, we consider a maximum dimension of three for the state variables, although the additional smoothing step can ameliorate this problem. Another reason is that most applications consider three state variables

S2 satisfies the null hypothesis of a certain parameter with a particular value. More specifically, the intercept of the model is zero, whereas the other parameters are nonlinear functions of the state variable $\mathbf{Z}_t$. DGP S3 satisfies the null hypothesis for a certain specification of the parameters. Here, we use a STAR(1) model, with a smoothing transition function as the CDF of the multinomial distribution. This can be used to check the performance of our test when the specification in the null is indeed the true model.

To examine the power performance of $\widehat{Q}_1$, we consider the following DGPs:

DGP P1 (nonlinear model)

$$\begin{cases} Y_t = [1 + \theta(\beta(\mathbf{Z}_t) - 1)]\, X_t + \varepsilon_t, \\ \beta(\mathbf{Z}_t) = Z_{1t} + Z_{2t} + ... + Z_{Lt}, \\ X_t = 0.5X_{t-1} + v_t, \\ v_t \sim i.i.d.N(0,1). \end{cases}$$

DGP P2 TAR(1):

$$X_t = \begin{cases} -1.5X_{t-1} - 2X_{t-2} + \varepsilon_{1t} & \text{if } X_{t-1}X_{t-2} < 0, \\ 0.5X_{t-1} + X_{t-2} + \varepsilon_{2t} & \text{if } X_{t-1}, X_{t-2} \geq 0, \\ 10X_{t-1} - 3X_{t-2} + \varepsilon_{3t} & \text{if } X_{t-1}X_{t-2} < 0, \\ -10X_{t-1} - 5X_{t-2} + \varepsilon_{4t} & \text{if } X_{t-1}X_{t-2} < 0, \end{cases}$$

To examine the power performance of $\widehat{Q}_2$, we consider the following DGP:

DGP P3 STAR(1):

$$X_t = 0.5X_{t-1} + \Phi\left(\frac{X_{t-1} - 1}{2}, ..., \frac{X_{t-L} - 1}{2}\right) X_{t-1} + X_{t-2} + u_t,$$

where Φ is CDF of multinomial distribution.

DGP P1 is a nonlinear model where the coefficients of the state variables are nonconstant, and can be used to check the power of our test statistic $\widehat{Q}_1$. Then, θ is a tuning parameter for nonlinearity in the model. To examine how the power of the tests change for different levels of deviation from a linear model, we check the cases where $\theta = 0.2, 0.5$, and 1.0. DGP P2 is a TAR(1) model. This is a functional coefficient model, where the state variables are lagged variables, and the parameters are step functions of the state variables. For the space constraint, we provide an example of two state variables. The case for one and three state variables are similar, except that all thresholds are zero. The null hypothesis for DGP P1 and DGP P2 is still the constancy of all the parameters. DGP P3 is the same as DGP S3, but now the null hypothesis is that the smooth transition function is a CDF of the χ^2_1 distribution.

For each DGP, we simulate 1000 data sets with sample sizes $T = 200, 500, 800$. We use three kernels:

the uniform kernel $K(z) = \frac{1}{2}\mathbf{1}(|z| \leq 1)$, the Epanechnikov kernel $K(z) = \frac{3}{4}(1-u^2)\mathbf{1}(|z| \leq 1)$, and the quartic kernel $K(z) = \frac{15}{16}(1-u^2)^2\mathbf{1}(|z| \leq 1)$. For brevity, we report only those results based on the Epanechnikov kernel. The results for the other two kernels are available from the authors upon request. Our simulation study shows that the choice of kernel function has little effect on the performance of our test. The choice of bandwidth is based on the modified multifold cross-validation criterion in Cai, Fan, and Yao (2000). We choose h to minimize

$$AMS(h) = \sum_{q=1}^{Q} AMS_q(h), \tag{4.1}$$

where, for $q = 1, ..., Q$,

$$AMS_q(h) = \frac{1}{m} \sum_{t=T-qm+1}^{T-qm+m} \left\{ Y_t - \mathbf{X}_t^T \widehat{\beta}_q(\mathbf{Z}_t) \right\}^2, \tag{4.2}$$

and $\widehat{\beta}_q(\cdot)$ are computed from the sample $(Y_t, \mathbf{Z}_t, \mathbf{X}_t), 1 \leq t \leq T - qm$, with bandwidth equal to $h = (\frac{T}{T-qm})^{\frac{1}{5}}$. The idea is to use Q sub-series of lengths $T - qm(q = 1, ..., Q)$ to estimate the unknown coefficient functions, and then to compute the one-step forecasting errors for the next section of the time series of length m, based on the estimated models. In our simulation study, we use $m = [0.1T]$ and $Q = 4$. This cross-validation criterion is used to minimize the MSE in an estimation, which may not be optimal in nonparametric tests. However, because the power of our test is not affected by the selection of the bandwidth, it can be viewed as being optimal in our setup. Other nonparametric tests do not have this feature.

Tables 1–3 show the size performance of the three types of null hypothesis at the 10% and 5% levels. Here, we use only bootstrap critical values for the GLR test and loss function approach, although their test statistics are asymptotically normal. This is because we find the critical values using the bootstrap in our test; thus, we should not take advantage of bootstrap over the asymptotic distribution in small samples. The results show that our test is robust to different specifications of stochastic errors. More importantly, it is well known that nonparametric tests exhibit size distortion, and tend to over-reject, especially for a small sample size. Compared with the GLR test and the Wald-type test, the results show that our test mitigates this problem by means of an additional smoothing step, especially when the dimension of the state variables increases.

Tables 4–6 show the power performance of our test, the GLR test, and the Wald-type test in a finite sample. We can see Q is more powerful than the GLR test and Wald-type test in all DGPs. In particular, Table 4 demonstrates our test has greater power to test constancy when the model parameters are near to constants. The power improvement of our test is more prominent when the sample size is small and the dimension of the state variables is large. Intuitively, this is because of the additional smoothing step and faster convergence rate of our test statistic.

5 Empirical Application

In this section, we provide an empirical application to conditional asset pricing models. In finance, such models have been proposed to explain why unconditional models fail:

$$R_{i,t} = \alpha_{i,t} + \beta_{i,t} f_t + \varepsilon_{i,t}, t = 1, 2, ..., T, \tag{5.1}$$

where $R_{i,t}$ is the excess return for asset i at time t, $f_t = (f_{1,t}, f_{2,t}, ..., f_{K,t})^T$ stands for the excess returns from mimicking portfolios, and $\alpha_{i,t}$ and $\beta_{i,t}$ stand for asset i's pricing error and factor loadings, respectively.

The time-variation in the risk and expected returns is expected to explain important asset-pricing anomalies. Two methods are proposed to check the validity of the conditional asset pricing models. First, instead of specifying state variables, we assume the coefficients are functions of time: $\alpha_{i,t} = \alpha_i(t/T)$ and $\beta_{i,t} = \beta_i(t/T)$. Lewellen and Nagel (2006) estimate and test the models using a fixed window size, and Li and Yang (2011) and Ang and Kristensen (2012) argue that inappropriate windows may lead to inconsistent or conflicting inferences. They use the GLR test and the Wald-type test to examine conditional CAPM. The drawback of these methods is that to avoid specifying state variables, we need to make a strong assumption that information is relatively stable within small windows, which is even more difficult to satisfy when the windows are date driven. On the other hand, we can specify certain state variables based on different consumption-based models: $\alpha_{i,t} = \alpha_i(\mathbf{Z}_t)$ and $\beta_{i,t} = \beta_i(\mathbf{Z}_t)$. These include the models of Shanken (1990), Ferson and Harvey (1991), Nagel and Singleton (2011), Li, Su, and Xu (2015), Kelly et al. (2018), Santos and Veronesi (2019), and Nagel et al. (2019). The disadvantage is that, as pointed out by Ghysels (1998) and Harvey (2001), the estimation is highly sensitive to the choice of variables in the information set. Furthermore, many conditioning variables, especially macro and accounting variables, are only available at coarse frequencies, although we can obtain daily stock return data.

In this empirical application, we focus on the state-variable approach and demonstrate that our method can mitigate the aforementioned issues. First, we can solve the problem of sensitivity to the selection of state variables by specifying all of them as state variables. The simulation studies in Section 4 show that our test has good power when the dimension of the state variables increases, compared with that of the GLR test and Wald-type test. Second, because the state variables are typically quarterly data, the sample size is small. Simulations also show that our test has better performance in small samples, owing to a faster convergence rate.

The state variables examined in our study follow from Nagel and Singleton (2011) and Li, Su, and Xu (2015).[11] the consumption—wealth ratio of Lettau and Ludvigson (2001) (cay), labor income-consumption ratio of Santos and Veronesi (2006) (yc), and corporate bond spread, as in Jagannathan and Wang (1996) (def). The cay data are obtained from Martin Lettau's website. Following Santos and Veronesi (2006), we obtain yc as the labor income component of cay. The def series is calculated as the yield difference between Baa- and Aaa-rated bonds, obtained from the Federal Reserve Bank of St. Louis. The data on these state variables run from 1959.Q1 to 2018.Q4. The state variables $\mathbf{Z}_t$ summarize the information set at time $t-1$. In our context, we consider seven choices of $\mathbf{Z}_t$: cay_{t-1}, yc_{t-1}, def_{t-1}, and any combinations thereof.

The returns and other corresponding data are taken from Professor Kenneth French's Website. As in Li and Yang (2011), we divide the stocks into six size and B/M portfolios, and three momentum portfolios: S is the average of the five portfolios in the lowest size quantile; B to be the average of the five portfolios in the highest size quantile; S-B is the difference; G is the average of the five portfolios in the lowest B/M quantile; V is the average of the five portfolios in the highest B/M quantile; G-V is the difference; W is the portfolio with the highest return among the 10 momentum portfolios; L is the portfolio with the lowest return among the 10 momentum portfolios; W-L is the difference. We tried both value weighted and equally weighted portfolios. We compound monthly portfolio returns to obtain quarterly returns that run from 1959.Q1 to 2018.Q4. The sample size is $T = 240$, a relatively small sample.

For simplicity, we evaluate only the performance of the conditional CAPM. We are interested in two types of hypothesis tests. First, are betas time varying (e.g., Bollerslev, Engle, and Wooldridge 1988; Ferson and Harvey 1991; Ferson and Korajczyk 1995): $\beta_i(\mathbf{Z}_t) = \beta_0$, where β_0 is an unknown unconditional beta for portfolio i. Second, we examine whether, when conditioning on $\mathbf{Z}_t$, the conditional CAPM can price a single portfolio and multiple portfolios. This amounts to testing the null hypothesis: $\alpha_i(\mathbf{Z}_t = 0$, for $i = 1, ..., N$. If the conditional CAPM is able to price all N portfolios jointly, the conditional pricing errors associated with any portfolio i should be zero at all time t. The p-value for each test statistic is obtained based on 200 bootstraps using the procedure described in Section 2. The bandwidth selection for all three tests is the same as in the simulations in Section 4.

[11] Santos and Veronesi (2019) provide alternative choices for the state variables: (a) the level of the aggregate premium itself; (b) the level of the firm's expected dividend growth; (c) the firm's fundamental risk, that is, the one pertaining to the covariation of the firm's cash-flow with the aggregate economy. We leave these for further research.

Table 7 provides the bootstrap p-values of the tests on the constancy of conditional beta. Panel A uses returns of value-weighted portfolios, as in Lewellen and Nagel (2006), and Panel B uses the returns of equally weighted portfolios suggested by Boguth et al. (2010). The results show that the constancy of conditional beta is strongly rejected using our method for all V, G, and V-G conditioning on either state variable individually. However, the GLR test and the Wald-type test only reject the null hypothesis for V-G when conditioning on yc and def. When the dimension of the state variable increases, our tests are still able to detect the nonconstancy in the conditional beta, whereas the GLR test and the Wald-type test fail to reject the null hypothesis. The results are similar between the value-weighted and equally weighted portfolios.

Table 8 provides bootstrap p-values for the tests on conditional beta. The results show that, by all three tests, the conditional CAPM is strongly rejected for all three B/M portfolios, conditioning on either state variable individually. This is consistent with the results of Li, Su, and Xu (2015), but are counter to the conclusions of several recent influential studies (e.g., Jagannathan and Wang 1996; Lettau and Ludvigson 2001; Santos and Veronesi 2006), which argue that conditioning dramatically improves the performance of both the simple and the consumption CAPMs. However, the GLR test and the Wald-type test fail to reject the conditional CAPM in all cases. Our results show that the conditional CAPM cannot explain pricing anomalies in unconditional CAPM.

6 Conclusion

This paper proposes a novel DFT-based approach to model specification test in functional coefficient models. We consider two types of hypothesis: constancy and a particular function form of the parameters. Although including a nonparametric estimation, by projecting the nonparametric estimator onto the Fourier basis, our test is able to obtain root-T consistency and, thus, is asymptotically more efficient than existing nonparametric tests. Therefore, our approach can detect a class of local alternatives at the parametric rate. Especially, simulation studies show that our test outperforms other nonparametric tests in small sample and when the dimension of the state variables increases. On the other hand, the implementation of nonparametric estimation generalizes the conditional moment tests in Bierens (1980, 1982) and enables our test to be applicable in broader circumstances. Therefore, our test improves the efficiency of nonparametric tests while maintaining its applicability.

An empirical application to conditional asset pricing models shows us the results by existing nonparametric tests in literature are misleading since they are less powerful in small sample. We also examine the model with more than one state variable and our tests perform good power in these cases. In this way, we avoid the problem of selecting state variables.

Several interesting extensions are possible. First, the construction of our test statistics implies that we can conduct the tests if only we get consistent estimators of the null and alternative models. Therefore, the extension to instrumented functional coefficient models is possible using the consistent estimator in Cai et al. (2006). Second, one can consider model specification test for the endogenous functional coefficient models under the framework of GMM, which is expected to be useful in macroeconomic analysis involving economic expectations. Third, it is interesting to consider the model specification problem for coefficient functions with unobserved state variables, such as the Markov regime switching models. Fourth, our method can be easily extended to the case where the state variable is a deterministic function of time. In this case, we are testing for abrupt or smooth structural changes. All these problems will be pursued in subsequent studies.

Table 1: Empirical size for DGP S.1

	T=200						T=500						T=800					
	Q		GLR		Wald		Q		GLR		Wald		Q		GLR		Wald	
L	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DPG S.1: i.i.d normal errors																		
1	11.3	6.2	10.5	5.9	10.7	5.4	10.0	5.6	10.4	5.3	10.4	5.8	9.7	4.9	8.8	4.9	9.7	5.0
2	11.8	6.2	12.2	6.4	12.5	6.4	8.7	4.0	13.1	6.0	11.6	6.1	10.2	5.0	11.1	4.1	11.6	4.3
3	12.9	6.3	16.8	7.3	15.9	7.8	10.9	5.4	12.9	6.2	13.0	7.0	9.8	4.7	7.9	3.8	10.9	6.9
DPG S.1: ARCH errors																		
1	12.4	6.2	10.5	6.0	11.8	6.6	10.2	5.5	9.8	5.6	9.7	5.6	9.3	4.9	8.5	4.2	8.9	4.3
2	12.5	6.1	12.4	6.2	13.2	6.6	8.9	4.4	13.3	6.0	12.2	5.9	10.8	5.1	11.3	4.0	11.4	5.2
3	12.4	6.5	16.0	7.7	15.9	8.1	10.8	5.9	12.3	6.6	13.0	6.8	9.6	4.8	7.3	3.9	7.7	4.2
DPG S.1: conditional heteroskedasticity errors																		
1	11.7	5.8	10.8	5.9	11.6	5.2	10.4	5.4	10.3	5.2	10.7	5.3	10.0	4.7	14.2	7.9	13.8	7.3
2	12.1	5.9	13.3	6.0	13.5	6.2	11.3	5.7	12.2	6.2	12.6	5.9	9.6	4.4	8.3	3.8	8.6	4.0
3	12.3	6.1	16.9	6.6	16.0	7.0	11.7	5.4	13.4	5.9	14.0	6.1	10.3	5.3	11.8	6.9	12.6	7.1

Note: (i) 1000 iterations; (ii) GLR the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001), Wald the Wald-type test in Chen and Hong (2012); (iii) L is the dimension of state variables; (iv) Critical values for all three tests are computed using the resampling method in Section 3; (v) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type test, and $h = c \cdot T^{-\frac{2}{9}}$ for the GLR, where c is selected based on the CV method in this section.

Table 2: Empirical size for DGP S.2

	T=200						T=500						T=800					
	Q		GLR		Wald		Q		GLR		Wald		Q		GLR		Wald	
L	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DPG S.2: i.i.d normal errors																		
1	11.4	6.1	11.6	6.9	11.7	6.6	10.2	5.6	11.4	5.4	11.4	5.8	10.6	5.1	8.9	4.9	9.2	4.5
2	11.9	6.7	12.9	7.5	12.9	6.9	10.9	6.1	13.3	6.6	12.4	6.1	10.7	5.1	11.7	4.6	11.2	5.2
3	12.8	6.7	16.9	7.6	15.8	7.9	11.9	5.9	15.9	7.2	14.3	7.2	9.9	4.9	8.1	4.2	10.7	6.3
DPG S.2: ARCH errors																		
1	11.6	5.9	12.7	6.2	11.9	6.6	10.6	5.6	11.0	5.7	10.8	5.6	10.2	4.9	10.5	5.1	10.6	5.1
2	12.6	6.4	13.9	6.8	13.7	7.0	11.7	6.2	13.3	6.1	13.2	5.9	10.8	5.2	12.2	5.9	12.4	5.9
3	13.4	6.8	16.2	7.8	16.9	8.2	11.8	5.9	14.3	6.6	13.1	6.6	12.1	5.9	15.1	7.2	15.6	7.1
DPG S.2: conditional heteroskedasticity errors																		
1	10.9	5.8	11.9	5.6	12.3	5.4	9.4	4.9	10.3	5.2	10.7	5.3	9.6	4.8	8.7	3.2	9.0	4.3
2	12.1	6.0	13.5	6.2	14.1	6.5	12.1	5.6	13.1	6.7	12.5	5.8	11.6	4.9	12.3	5.9	13.0	6.2
3	12.7	6.0	15.9	6.4	14.9	6.0	12.8	5.9	14.3	6.9	15.5	6.5	11.9	5.3	14.6	6.8	15.0	6.7

Note: (i) 1000 iterations; (ii) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Chen and Hong (2012); (iii) L is the dimension of state variables; (iv) Critical values for all three tests are computed by the resampling method in Section 3; (v) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and Wald-type, and $h = c \cdot T^{-\frac{2}{9}}$ for the GLR, where c is selected based on the CV method in this section.

Table 3: Empirical size for DGP S.3

	T=200						T=500						T=800					
	Q		GLR		Wald		Q		GLR		Wald		Q		GLR		Wald	
L	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DPG S.3: i.i.d normal errors																		
1	11.6	6.2	11.9	7.0	11.9	6.8	10.4	5.7	11.6	5.7	11.6	5.8	9.6	4.9	8.9	4.4	8.9	3.9
2	12.1	6.6	13.0	7.3	13.3	6.9	11.3	6.5	13.3	6.2	13.4	6.4	11.9	5.6	12.4	5.2	12.2	5.4
3	12.8	6.7	16.9	7.6	15.8	7.9	11.9	5.9	15.9	7.2	14.3	7.2	9.7	4.8	9.1	3.9	10.8	6.2
DPG S.3: ARCH errors																		
1	11.6	6.1	12.9	6.1	12.3	6.6	10.8	5.8	11.1	5.8	11.8	5.9	10.6	5.2	10.5	5.3	10.9	5.2
2	13.6	6.8	13.9	6.9	14.1	7.1	12.2	6.4	13.5	6.3	13.7	6.2	11.8	5.6	12.1	5.8	12.0	5.8
3	13.9	6.9	15.8	7.2	16.7	7.7	13.1	5.8	15.2	7.0	15.5	6.9	12.1	5.4	13.1	6.2	14.9	6.1
DPG S.3: conditional heteroskedasticity errors																		
1	9.4	4.8	9.1	3.3	8.3	4.4	10.4	5.3	11.1	5.6	10.5	5.8	10.4	4.9	10.7	5.5	10.0	5.2
2	13.1	6.2	13.9	5.9	13.8	6.4	12.8	5.6	13.2	6.9	13.1	6.1	11.8	5.2	12.1	5.8	13.2	6.4
3	14.4	6.4	16.1	7.2	15.9	6.6	13.8	5.9	15.2	6.9	15.2	6.7	11.8	5.6	14.9	6.3	15.1	6.4

Note: (i) 1000 iterations; (ii) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Chen and Hong (2012); (iii) L is the dimension of state variables; (iv) Critical values for all three tests are computed by the resampling method in Section 3; (v) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type and $h = c \cdot T^{-\frac{2}{9}}$ for the GLR, where c is selected based on the CV method in this section.

Table 4: Empirical power for DGP P.1

		T=200						T=500						T=800					
		Q		GLR		Wald		Q		GLR		Wald		Q		GLR		Wald	
L	θ	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DPG P.1: i.i.d normal errors																			
	0.2	24.4	11.7	17.1	7.3	16.2	7.2	39.4	22.1	26.3	14.2	24.5	13.8	55.5	26.3	49.4	20.1	46.4	19.1
1	0.5	44.7	26.2	33.2	14.4	30.1	12.9	59.3	35.2	45.9	29.6	45.6	27.8	70.8	46.6	60.5	30.1	57.5	29.6
	1.0	80.2	59.7	73.3	52.1	74.1	52.9	98.9	87.2	96.7	80.3	94.5	78.7	100	100	100	100	100	100
	0.2	24.2	11.2	13.3	6.6	12.4	5.6	36.9	19.2	19.3	11.0	18.2	10.4	50.7	23.4	42.4	13.8	40.5	12.2
2	1	42.2	24.5	20.5	9.9	23.6	11.1	56.6	32.9	33.3	20.1	30.8	18.9	66.2	40.1	50.0	19.9	48.7	17.1
	1.0	72.2	53.6	52.4	30.1	50.1	28.5	97.7	84.3	90.0	69.8	88.2	67.3	100	99.8	100	99.2	100	99.3
	0.2	23.2	10.9	11.2	5.4	11.5	5.2	36.0	18.7	18.4	9.0	19.1	9.2	50.7	20.5	39.1	9.8	39.0	9.2
3	1	39.4	22.5	11.9	7.1	12.0	7.6	56.5	28.6	21.4	13.0	23.2	15.1	63.3	38.4	42.0	10.7	38.9	10.0
	1.0	69.9	50.3	37.6	17.1	33.9	16.7	90.8	80.5	70.2	40.9	68.3	36.8	99.9	99.2	87.6	79.9	88.1	81.0
DPG P.1: ARCH errors																			
	0.2	23.5	12.1	16.4	7.9	16.8	8.0	38.8	21.1	25.7	14.1	25.5	13.9	54.2	24.6	48.1	20.0	45.4	18.7
1	0.5	43.6	25.3	31.9	14.0	28.8	12.6	59.2	35.0	44.7	29.0	44.6	27.9	70.2	45.7	60.1	28.9	56.5	29.1
	1.0	79.9	59.3	72.1	51.5	74.0	52.9	98.6	87.0	96.2	80.1	93.8	78.2	100	100	100	100	100	100
	0.2	23.8	11.0	13.1	6.2	12.0	5.5	35.9	19.0	18.7	10.9	18.1	10.2	50.2	23.1	42.8	13.2	40.0	11.5
2	1	42.0	24.6	21.3	9.7	24.0	11.2	56.6	32.8	33.5	19.7	30.5	18.4	66.4	40.1	50.5	19.8	48.9	16.9
	1.0	72.4	53.8	52.3	30.0	51.2	28.9	96.7	83.6	88.9	67.6	86.8	66.8	100	100.0	100	99.7	100	99.6
	0.2	23.0	10.5	11.1	5.5	11.7	5.2	35.5	18.4	18.3	9.0	17.8	8.4	50.5	20.1	39.0	9.6	38.8	9.0
3	1	39.5	22.5	11.7	7.0	12.3	7.7	56.2	28.2	21.1	12.7	23.1	14.6	62.8	37.7	42.0	10.8	38.4	10.2
	1.0	68.9	50.1	37.3	17.0	33.5	16.3	90.3	80.0	68.6	40.1	67.8	36.8	99.9	99.0	87.1	79.9	87.8	80.3
DPG P.1: conditional heteroskedasticity errors																			
	0.2	24.8	11.8	17.6	8.0	16.0	7.1	39.1	21.6	26.0	12.9	24.1	12.6	56.5	26.4	48.4	20.0	46.1	19.0
1	0.5	44.2	26.6	33.1	14.0	29.7	12.5	59.1	35.1	45.1	29.2	45.4	27.8	70.3	45.9	59.7	30.0	56.4	28.6
	1.0	80.6	58.8	72.3	51.3	74.0	52.6	98.2	87.0	96.2	80.0	94.1	78.2	100	100	100	100	100	99.9
	0.2	23.7	10.8	13.1	6.2	12.6	5.6	36.2	19.0	18.9	11.2	18.0	10.3	50.7	24.1	43.2	13.9	40.7	12.8
2	0.5	42.1	24.2	20.1	9.4	23.2	11.0	57.1	33.0	33.1	20.1	30.6	18.6	66.6	40.2	50.4	19.9	48.4	17.0
	1.0	71.9	53.2	52.1	29.5	50.0	28.2	98.1	84.8	87.8	66.9	88.0	67.2	100	99.9	100	99.6	100	99.4
	0.2	23.6	10.9	11.1	5.2	11.8	5.1	36.7	18.9	18.6	9.2	19.0	8.8	50.1	20.1	38.4	9.6	39.1	9.7
3	0.5	39.1	22.4	11.5	7.2	12.6	7.8	56.1	28.2	21.1	13.0	23.1	14.4	62.6	37.4	42.0	10.2	38.3	11.0
	1.0	69.2	50.1	37.2	16.5	34.1	16.0	91.2	80.8	70.0	40.2	68.1	36.2	100	99.1	87.1	79.6	87.3	80.1

Note: (i) 1000 iterations; (ii) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Chen and Hong (2012); (iii) L is the dimension of the state variables; θ is a measure of deviation from constancy; (iv) Critical values for all three tests are computed by the resampling method in Section 3; (v) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type, and $h = c \cdot T^{-\frac{2}{9}}$ for GLR, where c is selected based on the CV method in this section.

Table 5: Empirical power for DGP P.2

	T=200						T=500						T=800					
	Q		GLR		Wald		Q		GLR		Wald		Q		GLR		Wald	
L	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DPG P.2: i.i.d normal errors																		
1	79.3	56.8	63.3	42.5	61.4	41.2	98.9	87.2	96.2	73.2	95.7	72.7	100	100	100	100	100	100
2	72.2	53.6	59.4	30.1	55.8	28.9	97.7	84.3	90.0	69.8	88.4	69.1	100	99.8	100	99.2	100	99.0
3	69.9	50.3	47.6	17.1	47.2	17.0	90.8	80.5	70.2	40.9	71.3	42.5	99.9	99.2	87.6	79.9	87.2	77.8
DPG P.2: ARCH errors																		
1	79.9	57.0	60.1	39.9	59.9	37.8	99.2	89.2	95.4	69.5	92.1	69.4	100	100	100	100	100	100
2	76.6	52.5	55.4	28.2	53.5	27.9	98.7	86.7	90.8	72.1	91.2	72.2	100	100	100	100	100	100
3	70.3	49.2	49.3	17.6	42.6	15.3	96.8	85.2	71.1	44.9	67.2	40.2	100	99.4	95.6	90.8	90.2	85.2
DPG P.2: conditional heteroskedasticity errors																		
1	79.7	57.0	59.2	39.2	59.1	37.3	98.4	87.0	95.5	73.1	95.1	72.4	100	100	100	100	100	99.9
2	76.3	52.1	55.1	28.0	53.2	27.4	97.3	84.1	90.0	69.2	88.1	68.3	100	100	100	100	99.8	99.2
3	70.1	49.2	49.9	17.8	42.2	15.1	90.5	80.1	69.2	40.2	71.1	42.2	100	99.6	96.2	90.9	91.1	85.8

Note: (i) 1000 iterations; (ii) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Chen and Hong (2012); (iii) L is the dimension of the state variables; (iv) Critical values for all three tests are computed by the resampling method in Section 3; (v) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type, and $h = c \cdot T^{-\frac{2}{9}}$ for GLR, where c is selected based on the CV method in this section.

Table 6: Empirical power for DGP P.3

	T=200						T=500						T=800					
	Q		GLR		Wald		Q		GLR		Wald		Q		GLR		Wald	
L	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%
DPG P.3: i.i.d normal errors																		
1	66.8	42.4	43.1	24.6	41.2	20.5	88.7	79.2	78.5	53.4	76.7	52.4	100	99.8	92.5	89.5	91.7	87.3
2	58.3	37.2	28.2	16.3	26.3	15.9	76.9	60.2	50.5	32.3	51.2	33.1	100	99.2	86.5	79.3	83.5	76.8
3	38.9	18.9	12.1	7.2	14.3	9.3	50.5	26.9	28.2	13.3	28.1	12.7	98.9	93.1	77.3	68.5	74.1	65.2
DPG P.3: ARCH errors																		
1	66.3	42.1	42.5	24.2	40.8	20.1	88.2	79.1	77.7	52.9	76.2	52.1	100	99.6	92.1	89.0	91.2	87.1
2	58.1	37.0	27.7	16.1	25.7	15.3	76.1	60.0	50.2	31.6	50.6	32.8	100	99.0	86.2	78.6	83.1	76.2
3	38.4	18.2	11.5	7.1	14.1	8.8	50.2	26.4	28.1	13.3	26.9	12.5	98.2	91.8	78.0	68.9	73.8	64.7
DPG P.3: conditional heteroskedasticity errors																		
1	66.9	43.1	43.0	23.8	41.1	20.4	89.1	79.5	78.2	53.1	76.3	51.8	100	100	92.9	89.9	92.0	88.1
2	58.3	37.0	29.3	15.6	26.1	15.4	76.1	59.4	50.2	32.1	52.2	34.2	100	99.5	87.9	81.2	84.5	76.9
3	38.5	18.5	12.1	7.1	14.2	8.9	50.1	26.2	27.6	13.2	28.1	13.1	97.2	90.3	74.2	66.8	72.9	62.4

Note: (i) 1000 iterations; (ii) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang (2001); Wald, the Wald-type test in Chen and Hong (2012); (iii) L is the dimension of the state variables; (iv) Critical values for all three tests are computed by the resampling method in Section 3; (v) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type, and $h = c \cdot T^{-\frac{2}{9}}$ for GLR, where c is selected based on the CV method in this section.

Table 7: Bootstrap p-values for tests on conditional beta

Zt	V			G			V-G		
	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
Panal A: value-weighted portfolios									
cay	0.035	0.270	0.415	0.135	0.890	0.760	0.030	0.240	0.325
yc	0.005	0.020	0.110	0.000	0.085	0.040	0.000	0.040	0.025
def	0.035	0.125	0.520	0.045	0.060	0.105	0.000	0.000	0.000
cay, yc	0.045	0.530	0.480	0.030	0.215	0.410	0.010	0.315	0.225
yc, def	0.020	0.325	0.265	0.140	0.875	0.980	0.055	0.410	0.610
def, cay	0.055	0.280	0.490	0.070	0.665	0.425	0.045	0.490	0.585
cay, yc, def	0.090	0.885	0.940	0.025	0.260	0.285	0.040	0.525	0.410
Panal B: equally-weighted portfolios									
cay	0.025	0.190	0.095	0.155	0.520	0.450	0.035	0.335	0.365
yc	0.015	0.045	0.110	0.005	0.065	0.060	0.000	0.020	0.020
def	0.060	0.125	0.525	0.045	0.040	0.160	0.000	0.010	0.005
cay, yc	0.040	0.630	0.480	0.045	0.210	0.435	0.040	0.320	0.410
yc, def	0.035	0.315	0.265	0.140	0.820	0.450	0.055	0.470	0.580
def, cay	0.040	0.340	0.510	0.055	0.635	0.425	0.025	0.410	0.560
cay, yc, def	0.085	0.720	0.945	0.040	0.190	0.260	0.120	0.695	0.450

Note: (i) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Chen and Hong (2012); (ii) $\mathbf{Z}_t$ is the set of state variables; (iii) The p-value for each test statistic is obtained based on 200 bootstraps using the procedure described in Section 2; (iv) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type, and $h = c \cdot T^{-\frac{2}{9}}$ for GLR, where c is selected based on the CV method in Section 4.

Table 8: Bootstrap p-values for tests on conditional alpha

	V			G			V-G		
Zt	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
				Panal A: value-weighted portfolios					
cay	0.000	0.005	0.005	0.000	0.000	0.005	0.010	0.010	0.005
yc	0.005	0.010	0.005	0.000	0.000	0.000	0.000	0.000	0.000
def	0.000	0.005	0.000	0.005	0.005	0.005	0.010	0.015	0.020
cay, yc	0.025	0.310	0.290	0.035	0.625	0.615	0.045	0.315	0.225
yc, def	0.015	0.520	0.395	0.040	0.470	0.515	0.025	0.310	0.505
def, cay	0.035	0.610	0.485	0.070	0.625	0.730	0.025	0.390	0.455
cay, yc, def	0.040	0.810	0.925	0.020	0.695	0.810	0.055	0.735	0.890
				Panal B: equally-weighted portfolios					
cay	0.000	0.005	0.005	0.000	0.005	0.005	0.000	0.000	0.005
yc	0.000	0.010	0.005	0.005	0.010	0.005	0.000	0.005	0.000
def	0.000	0.005	0.005	0.000	0.010	0.000	0.005	0.005	0.005
cay, yc	0.015	0.290	0.435	0.045	0.535	0.640	0.045	0.385	0.395
yc, def	0.035	0.515	0.575	0.040	0.310	0.225	0.025	0.610	0.485
def, cay	0.020	0.730	0.455	0.070	0.395	0.730	0.035	0.595	0.570
cay, yc, def	0.055	0.920	0.785	0.025	0.605	0.655	0.035	0.645	0.590

Note: (i) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Chen and Hong (2012); (ii) $\mathbf{Z}_t$ is the set of state variables; (iii) The p-value for each test statistic is obtained based on 200 bootstraps using the procedure described in Section 2; (iv) The Epanechnikov kernel is used for all three tests; the bandwidths $h = c \cdot T^{-\frac{1}{5}}$ for our test Q and the Wald-type, and $h = c \cdot T^{-\frac{2}{9}}$ for GLR, where c is selected based on the CV method in Section 4.

References

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

Bansal, R., Hsieh, D. A., and Viswanathan, S. (1993). A new approach to international arbitrage pricing. *The Journal of Finance*, 48(5):1719–1747.

Borensztein, E., De Gregorio, J., and Lee, J.-W. (1998). How does foreign direct investment affect economic growth? *Journal of international Economics*, 45(1):115–135.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American statistical Association*, 80(391):580–598.

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Fan, J., Härdle, W., Mammen, E., et al. (1998). Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26(3):943–971.

Fan, J., Yao, Q., and Cai, Z. (2003). Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 65(1):57–80.

Fan, J., Zhang, C., Zhang, J., et al. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of statistics*, 29(1):153–193.

Fan, J., Zhang, W., et al. (1999). Statistical estimation in varying coefficient models. *The annals of Statistics*, 27(5):1491–1518.

Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of political economy*, 99(2):385–415.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29.

Friedman, J. H. et al. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 19(1):1–67.

Granger, C. W., Terasvirta, T., et al. (1993). Modelling non-linear economic relationships. *OUP Catalogue*.

Green, P. and Silverman, B. (1994). Nonparametric regression and generalized linear models. number 58 in monographs on statistics and applied probability. *Nonparametric regression and generalized linear models*, (58).

Gu, C., Heckman, N., and Wahba, G. (1992). A note on generalized cross-validation with replicates. *Statistics & probability letters*, 14(4):283–287.

Hansen, B. E. (1997). Inference in tar models. *Studies in Nonlinear Dynamics and Econometrics*, 2:1–14.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603.

Hastie, T. and Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing selection bias using experimental data. Technical report, National bureau of economic research.

Hong, Y., Lee, Y.-J., et al. (2013). A loss function approach to model specification testing and its relative efficiency. *The Annals of Statistics*, 41(3):1166–1203.

Horowitz, J. L. and Härdle, W. (1996). Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association*, 91(436):1632–1640.

Huber, P. J. (1985). Projection pursuit. *The annals of Statistics*, pages 435–475.

Koenker (2005). *Quantile Regression (Econometric Society monographs; no. 38)*. Cambridge university press.

Lewellen, J. and Nagel, S. (2006). The conditional capm does not explain asset-pricing anomalies. *Journal of financial economics*, 82(2):289–314.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327.

Li, Q., Huang, C. J., Li, D., and Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20(3):412–422.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954.

Roussanov, N. (2014). Composition of wealth, conditioning information, and the cross-section of stock returns. *Journal of Financial Economics*, 111(2):352–380.

Shanken, J. (1990). Intertemporal asset pricing: An empirical investigation. *Journal of Econometrics*, 45(1-2):99–120.

Stock, J. H. (1989). Nonparametric policy analysis. *Journal of the American Statistical Association*, 84(406):567–575.

Stone, C. J., Hansen, M. H., Kooperberg, C., Truong, Y. K., et al. (1997). Polynomial splines and their tensor products in extended linear modeling: 1994 wald memorial lecture. *The Annals of Statistics*, 25(4):1371–1470.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, 89(425):208–218.

Teräsvirta, T. et al. (1996). Modelling economic relationships with smooth transition regressions. Technical report, Stockholm School of Economics.

Tong, H. (1990). *Non-linear time series: a dynamical system approach*. Oxford University Press.

Wahba, G. (1984). *Variational methods for multidimensional inverse problems*. University of Wisconsin, Department of Statistics.

White, H. (2014). *Asymptotic theory for econometricians*. Academic press.

Mathematical Appendix

Throughout the appendix, we let "$\xrightarrow{p}$", "$\xrightarrow{d}$" and "$\Rightarrow$" denote convergence in probability, convergence in distribution, and weak convergence respectively. Also define $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$ and $c_0 = \mu_2/(\mu_2 - \mu_1^2)$ and $c_1 = -\mu_1/(\mu_2 - \mu_1^2)$.

Proof of Lemma 1: Notice that

$$\begin{aligned} E[\beta(Z_t)\psi_t(Z_t, u)] &= E[\beta(Z_t)(e^{iu'Z_t} - E(e^{iu'Z_t}))] \\ &= E[[\beta(Z_t) - E(\beta(Z_t))]e^{iu'Z_t}] \end{aligned}$$

By Theorem 1 (I) of Bierens (1982), we have $E[[\beta(Z_t) - E(\beta(Z_t))]e^{iu'Z_t}] = 0$ if and only if $E[\beta(Z_t) - E(\beta(Z_t))|Z_t] = 0$ a.s. The latter is equivalent to $E[\beta(Z_t)|Z_t] = E[\beta(Z_t)]$. That is, $\beta(Z_t)$ is a constant function of Z_t.

Proof of equation (2.20): Notice that

$$\begin{aligned} \widehat{Q} &= \frac{1}{T}\int_{\mathbb{R}^{d_z}} \left| \sum_{t=1}^{T} \widehat{\beta}(Z_t)\widehat{\psi}_t(u) \right|^2 W(u)du \\ &= \frac{1}{T}\int_{\mathbb{R}^{d_z}} \sum_{s,t=1}^{T} \widehat{\beta}(Z_s)'\widehat{\beta}(Z_t)\widehat{\psi}_s(u)\widehat{\psi}_t(u)^* W(u)du \\ &= \frac{1}{T}\sum_{s,t=1}^{T} \widehat{\beta}(Z_s)'\widehat{\beta}(Z_t)\int_{\mathbb{R}^{d_z}} \left(e^{iu'Z_s} - \frac{1}{T}\sum_{l=1}^{T} e^{iu'Z_l} \right)\left(e^{-iu'Z_t} - \frac{1}{T}\sum_{l=1}^{T} e^{-iu'Z_l} \right) W(u)du \end{aligned}$$

Denote

$$\begin{aligned} V_{st} &\equiv \int_{\mathbb{R}^{d_z}} \left(e^{iu'Z_s} - \frac{1}{T}\sum_{l=1}^{T} e^{iu'Z_l} \right)\left(e^{-iu'Z_t} - \frac{1}{T}\sum_{l=1}^{T} e^{-iu'Z_l} \right) W(u)du \\ &= \int_{\mathbb{R}^{d_z}} \left[e^{iu'(Z_s - Z_t)} - \frac{1}{T}\sum_{l=1}^{T} e^{iu'(Z_l - Z_t)} - \frac{1}{T}\sum_{l=1}^{T} e^{iu'(Z_s - Z_l)} + \frac{1}{T^2}\sum_{m,n=1}^{T} e^{iu'(Z_m - Z_n)} \right] W(u)du \\ &= \int_{\mathbb{R}^{d_z}} \left[\cos u'(Z_s - Z_t) - \frac{1}{T}\sum_{l=1}^{T} \cos u'(Z_l - Z_t) - \frac{1}{T}\sum_{l=1}^{T} \cos u'(Z_s - Z_l) + \frac{1}{T^2}\sum_{m,n=1}^{T} \cos u'(Z_m - Z_n) \right] W(u)du \\ &= e^{-\frac{1}{2}|Z_s - Z_t|^2} - \frac{1}{T}\sum_{l=1}^{T} e^{-\frac{1}{2}|Z_l - Z_t|^2} - \frac{1}{T}\sum_{l=1}^{T} e^{-\frac{1}{2}|Z_s - Z_l|^2} + \frac{1}{T^2}\sum_{m,n=1}^{T} e^{-\frac{1}{2}|Z_m - Z_n|^2} \end{aligned}$$

by (2.1.2).

Proof of Proposition 3.1:

$$
\begin{aligned}
\sqrt{T}\widehat{A}_1(u) &= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left[\widehat{\beta}(Z_t)-\beta(Z_t)+\beta(Z_t)\right]\left[\widehat{\psi_t}(u)-\psi_t(u)+\psi_t(u)\right] \\
&= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left[\widehat{\beta}(Z_t)-\beta(Z_t)+\beta(Z_t)\right]\left[\phi(u)-\widehat{\phi}(u)+\psi_t(u)\right] \\
&= \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left[\widehat{\beta}(Z_t)-\beta(Z_t)\right]\psi_t(u)+\frac{1}{T^{1/2}}\sum_{t=1}^{T}\beta(Z_t)\psi_t(u) \\
&+ \frac{1}{T^{1/2}}\sum_{t=1}^{T}\left[\widehat{\beta}(Z_t)-\beta(Z_t)\right]\left[\phi(u)-\widehat{\phi}(u)\right]+\frac{1}{T^{1/2}}\sum_{t=1}^{T}\beta(Z_t)\left[\phi(u)-\widehat{\phi}(u)\right] \\
&= A_1+A_2+A_3+A_4,\ \text{say.}
\end{aligned}
$$

To show Theorem **??**, it is sufficient to show Propositions 1 to 4 as follows.

Proposition 1. Under the conditions of Theorem 1, $A_1 \Rightarrow \mathcal{G}(u)$ where $\mathcal{G}(u)$ is a complex-valued Gaussian process with 0 mean and covariance kernel as

$$
\begin{aligned}
\mathcal{K}(u_1,u_2) &\equiv cov(\mathcal{G}(u_1),\mathcal{G}^*(u_2)) \\
&= 4E\left[r(\xi_t,u_1)r^*(\xi_t,u_2)'\right])
\end{aligned}
$$

where

$$
r(\xi_t,u) \equiv \frac{1}{2}c_0\Omega(Z_t)^{-1}\psi(u,Z_t)\left[\nu_X(Z_t)\nu_Y(Z_t)+X_tY_t\right] \tag{.1}
$$

and $\nu_X(Z_t)\equiv E(X_t \mid Z_t)$, $\nu_Y(Z_t)\equiv E(Y_t \mid Z_t)$. And $*$ denotes the complex conjugate.

Proposition 2. Under the conditions of Theorem 1, $A_2 =?$.

Proposition 3. Under the conditions of Theorem 1, $A_3 = o_p(1)$ uniformly over $u \in \mathbb{R}^L$.

Proposition 4. Under the conditions of Theorem 1, $A_4 = o_p(1)$ uniformly over $u \in \mathbb{R}^L$.

Proof of Proposition 2: By Cai, Fan and Yao (2001), the local linear estimator for functional coefficient models can be expressed as:

$$
\widehat{\beta}(z) = e_pH^{-1}S_T^{-1}(z)\Gamma_T(z)
$$

where

$$
S_T(z) = \begin{bmatrix} S_{T,0}(z) & S_{T,1}(z) \\ S_{T,1}(z) & S_{T,2}(z) \end{bmatrix}
$$

$$= \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t-z)X_tX_t' & \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t-z)X_tX_t'\left(\frac{Z_t-z}{h}\right) \\ \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t-z)X_tX_t'\left(\frac{Z_t-z}{h}\right) & \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t-z)X_tX_t'\left(\frac{Z_t-z}{h}\right)^2 \end{bmatrix}$$

and

$$\Gamma_T(z) = \begin{bmatrix} S_{T,0}(z) \\ S_{T,2}(z) \end{bmatrix} = \begin{bmatrix} \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t-z)X_tY_t \\ \frac{1}{T}\sum_{t=1}^{T} K_h(Z_t-z)X_tY_t'\left(\frac{Z_t-z}{h}\right) \end{bmatrix}$$

Since the coefficient $\beta(z)$ is are conducted in the neighborhood of $|Z_t - z| < h$, by Taylor's expansions, it is straightforward to see that uniformly over $u \in \mathbb{R}^L$ and $z \in \mathbb{R}^L$,

$$\widehat{\beta}(z) - \beta(z) = e_p H^{-1} S_T^{-1}(z)\Gamma_T^*(z) + \frac{h^2}{2} e_p H^{-1} S_T^{-1}(z) \begin{pmatrix} S_{T,2}(z) \\ S_{T,3}(z) \end{pmatrix} \beta^{''}(z) + o_p(1)$$

Under $\mathbb{H}_{01}$, $\beta^{''}(z) = 0$. Then uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned} A_1 &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[e_p H^{-1} S_T^{-1}(Z_t)\Gamma_T^*(Z_t)\right] \psi_t(u) + o_p(1) \\ &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[e_p H^{-1} S^{-1}(Z_t) S(Z_t) S_T^{-1}(Z_t)\Gamma_T^*(Z_t)\right] \psi_t(u) + o_p(1) \\ &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[e_p H^{-1} S^{-1}(Z_t) \left[I - (S(Z_t) - S_T(Z_t))S^{-1}(Z_t)\right]^{-1} \Gamma_T^*(Z_t)\right] \psi_t(u) + o_p(1) \end{aligned}$$

It is easy to see that $|\lambda_i| < 1$ for each eigenvalue λ_i of $(S(Z_t) - S_T(Z_t))S^{-1}(Z_t)$, then by geometric expansion,

$$\begin{aligned} \left[I - (S(Z_t) - S_T(Z_t))S^{-1}(Z_t)\right]^{-1} &= I + (S(Z_t) - S_T(Z_t))S^{-1}(Z_t) + O(|(S(Z_t) - S_T(Z_t))S^{-1}(Z_t)|) \\ &= I + (S(Z_t) - S_T(Z_t))S^{-1}(Z_t) + o_p(1) \end{aligned}$$

for all t, where we use the uniform convergence result of $\sup_{z\in\mathbb{R}^L} |S(z) - S_T(z)| = o_p(1)$ by Theorem 5, Hansen (2008). Then uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned} A_1 &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[e_p H^{-1} S^{-1}(Z_t)\Gamma_T^*(Z_t)\right] \psi_t(u) + \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[e_p H^{-1} S^{-1}(Z_t)(S(Z_t) - S_T(Z_t))S^{-1}(Z_t)\Gamma_T^*(Z_t)\right] \psi_t(u) + o_p(1) \\ &= \frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t)\right] \psi_t(u) + \frac{1}{T^{1/2}} \sum_{t=1}^{T} \left[e_p H^{-1} S_T(Z_t) S^{-1}(Z_t)\Gamma_T^*(Z_t)\right] \psi_t(u) + o_p(1) \\ &= A_{11} + A_{12} + o_p(1), \text{ say.} \end{aligned}$$

where

$$Q_T(Z_t) = \frac{1}{T}\sum_{s=1}^{T} X_s \left[c_0 + c_{11}\left(\frac{Z_{s1}-Z_{t1}}{h}\right) + ... + c_{1d}\left(\frac{Z_{sd}-Z_{td}}{h}\right)\right]\kappa_h(Z_s - Z_t)\varepsilon_s$$

Notice that A_{12} is a third-order non-degenerate U-process and $A_{12} = o_p(1)$ uniformly over $u \in \mathbb{R}^L$, which is involved in the proof of A_{11} term, we refer the reader to the proof of ?. To prove Proposition 1, it suffices to show the following lemma.

Lemma 1. Under the conditions of Theorem 1,

(i) A_{11} is stochastically equicontinuous over $u \in \mathbb{R}^L$;

(ii) For each $u \in \mathbb{R}^L$, if $Th^{d_z} \to \infty$ as $T \to \infty$, then $A_{11} \xrightarrow{d} MN(0, 4E\left[r(\xi_t, u)r^*(\xi_t, u)'\right] - 4A(u)A^*(u)')$ as $T \to \infty$.

Proof of Lemma 1: For (i), we need to show that, for any $\epsilon > 0$ and $\kappa > 0$, there exists a $\delta > 0$ such that

$$\lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta}\left\|\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\psi_t(u_1) - \frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\psi_t(u_2)\right\| > \kappa\right] < \epsilon.$$

Let $\bar{u} = \eta u_1 + (1-\eta)u_2$ for some $\eta \in (0,1)$ such that

$$e^{iu_1'Z_t} = e^{iu_2'Z_t} + iZ_t'(u_1-u_2)e^{i\bar{u}'Z_t},$$

then

$$\begin{aligned}
& \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta}\left\|\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\psi_t(u_1) - \frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\psi_t(u_2)\right\| > \kappa\right] \\
= & \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta}\left\|\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right](e^{iu_1'Z_t} - e^{iu_2'Z_t})\right\| > \kappa\right] \\
= & \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta}\left\|\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]iZ_t'e^{i\bar{u}'Z_t}(u_1-u_2)\right\| > \kappa\right] \\
\leq & \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta}\sqrt{\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left\|\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right\|^2}\sqrt{\frac{2}{T^{1/2}}\sum_{t=1}^{T}\|iZ_t'e^{i\bar{u}'Z_t}(u_1-u_2)\|^2} > \kappa\right] \\
= & \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta}\sqrt{\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left\|\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right\|^2}\sqrt{\frac{2}{T^{1/2}}\sum_{t=1}^{T}\|iZ_t'e^{i\bar{u}'Z_t}(u_1-u_2)\|^2} > \kappa\right] \\
\leq & \lim_{T\to\infty} P\left[\sqrt{\frac{2}{T^{1/2}}\sum_{t=1}^{T}\left\|\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right\|^2}\sqrt{\frac{2}{T^{1/2}}\sum_{t=1}^{T}\|Z_t\|^2} > \kappa/\delta\right]
\end{aligned}$$

where the second to last inequality is by Cauchy-Schwarz inequality and the last equality is due to the fact that $\|e^{i\bar{u}'Z_t}\|^2 = 1$ and $\|u_1 - u_2\| \leq \delta$. Given the moment conditions in Assumptions ?, we have $\frac{2}{T^{1/2}} \sum_{t=1}^{T} \|Z_t\|^2 = O_p(1)$ and $\frac{2}{T^{1/2}} \sum_{t=1}^{T} \left\| \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t) \right\|^2 = O_p(1)$. Thus, for any $\epsilon > 0$, we can find a $\delta > 0$ small enough such that

$$\lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^L:\|u_1-u_2\|<\delta} \left\| \frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t)\right] \psi_t(u_1) - \frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t)\right] \psi_t(u_2) \right\| > \kappa \right] < \epsilon.$$

Next we show (ii). Notice that

$$\begin{aligned}
A_2 &= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t)\psi_t(u) \\
&= \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} \left[\frac{1}{T}\sum_{s=1}^{T} X_s \left[c_0 + c_{11}\left(\frac{Z_{s1}-Z_{t1}}{h}\right) + ... + c_{1d}\left(\frac{Z_{sd}-Z_{td}}{h}\right)\right] \kappa_h(Z_s - Z_t)\varepsilon_s\right] \psi_t(u) \\
&= \frac{1}{T^{3/2}} \sum_{s,t=1}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \left[c_0 + c_{11}\left(\frac{Z_{s1}-Z_{t1}}{h}\right) + ... + c_{1d}\left(\frac{Z_{sd}-Z_{td}}{h}\right)\right] \kappa_h(Z_s - Z_t)\varepsilon_s\psi_t(u) \\
&= \frac{1}{T^{3/2}} \sum_{t=1}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_t c_0 \frac{1}{h^d} K(0)^d \varepsilon_t\psi_t(u) \\
&+ \frac{1}{T^{3/2}} \sum_{s\neq t}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \left[c_0 + c_{11}\left(\frac{Z_{s1}-Z_{t1}}{h}\right) + ... + c_{1d}\left(\frac{Z_{sd}-Z_{td}}{h}\right)\right] \kappa_h(Z_s - Z_t)\varepsilon_s\psi_t(u) \\
&= A_{21} + A_{22}
\end{aligned} \tag{.2}$$

The proof of Proposition 2 consists of the proofs of Lemma 2-3 below.

Lemma 2. Let A_{21} be defined as in (6). Then $A_{21} = o_p(1)$.

Lemma 3. Let A_{22} be defined as in (6). Then $A_{22} = \tilde{U} + o_p(1)$.

Proof of Lemma 2

$$A_{21} = c_0 K(0)^d \frac{1}{T^{3/2}h^d} \sum_{t=1}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_t\varepsilon_t\psi_t(u) \equiv c_0 K(0)^d \frac{1}{T^{3/2}h^d} \sum_{t=1}^{T} \eta_t$$

By orthogonality condition $E(\varepsilon_t \mid X_t, Z_t) = 0$, we have $E(A_{21}) = 0$, and

$$\begin{aligned}
var(A_{21}) &= c_0^2 K(0)^{2d} T^{-3} h^{-2d} \sum_{t=1}^{T} var(\eta_t) + c_0^2 K(0)^{2d} T^{-2} h^{-2d} \sum_{i=1}^{T-1} |cov\left[\eta_1, \eta_{1+i}\right]| \\
&= O(T^{-2}h^{-2d}) + O(T^{-2}h^{-2d}) \sum_{j=1}^{\infty} j^2 \beta(j)^{\delta/(1+\delta)} = O(T^{-2}h^{-2d})
\end{aligned}$$

We have $A_{21} = o_p(1)$ by Chebyshev's inequality.

Proof of Lemma 3

$$\begin{aligned}
A_{22} &= \frac{1}{T^{3/2}} c_0 \sum_{s\neq t}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \kappa_h(Z_s - Z_t)\varepsilon_s \psi_t(u) \\
&+ \frac{1}{T^{3/2}} c_{11} \sum_{s\neq t}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \left(\frac{Z_{s1} - Z_{t1}}{h}\right) \kappa_h(Z_s - Z_t)\varepsilon_s \psi_t(u) \\
&+ \frac{1}{T^{3/2}} c_{12} \sum_{s\neq t}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \left(\frac{Z_{s2} - Z_{t2}}{h}\right) \kappa_h(Z_s - Z_t)\varepsilon_s \psi_t(u) \\
&... \\
&+ \frac{1}{T^{3/2}} c_{1d} \sum_{s\neq t}^{T} \frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \left(\frac{Z_{sd} - Z_{td}}{h}\right) \kappa_h(Z_s - Z_t)\varepsilon_s \psi_t(u) \\
&= A_{22}^{(0)} + A_{22}^{(1)} + A_{22}^{(2)} + ... + A_{22}^{(d)}
\end{aligned}$$

We shall first prove $A_{22}^{(m)} = o_p(1)$ for $m = 1, 2, ..., d$. Define

$$\begin{aligned}
p_T(\xi_s, \xi_t) = &\prod_{i=1}^{d} K\left(\frac{Z_{si} - Z_{ti}}{h}\right) \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \left(\frac{Z_{sm} - Z_{tm}}{h}\right) \varepsilon_s \psi_t(u)\right] \\
&+ \prod_{i=1}^{d} K\left(\frac{Z_{ti} - Z_{si}}{h}\right) \left[\frac{\Omega(Z_s)^{-1}}{f_Z(Z_s)} X_t \left(\frac{Z_{tm} - Z_{sm}}{h}\right) \varepsilon_t \psi_s(u)\right]
\end{aligned}$$

Then we have

$$A_{22}^{(m)} = c_{1m} \frac{1}{T^{3/2} h^d} \sum_{1\leq s<t\leq T} p_T(\xi_s, \xi_t)$$

Define

$$p_T(\xi_s) = E(p_T(\xi_s, \xi_t)|\xi_s)$$

Then we have

$$\begin{aligned}
p_T(\xi_s) &= \int p_T(\xi_s, \xi) dP(\xi) \\
&= \int \prod_{i=1}^{d} K\left(\frac{Z_{si} - z_i}{h}\right) \left[\frac{\Omega(z)^{-1}}{f_Z(z)} X_s \left(\frac{Z_{sm} - z_m}{h}\right) \varepsilon_s (e^{iu'z} - \phi(u))\right] f_Z(z) dz \\
&= \int \prod_{i=1}^{d} K\left(\frac{Z_{si} - z_i}{h}\right) \left[\Omega(z)^{-1} X_s \left(\frac{Z_{sm} - z_m}{h}\right) \varepsilon_s (e^{iu'z} - \phi(u))\right] dz
\end{aligned}$$

where we have used the fact that $E(\varepsilon_t \mid X_t, Z_t) = 0$. Notice that $E(p_T(\xi_s)) = 0$, then

$$\begin{aligned}
A_{22}^{(m)} &= \frac{c_{1m}}{T^{3/2} h^d} \sum_{1\leq s<t\leq T} [p_T(\xi_s, \xi_t) - p_T(\xi_s) - p_T(\xi_t)] + \frac{c_{1m}(T-1)}{T^{3/2} h^d} \sum_{t=1}^{T} p_T(\xi_t) \\
&= R_{22}^{(1)} + R_{22}^{(2)}
\end{aligned}$$

Obviously, $E\left[p_T(\xi_s,\xi_t)-p_T(\xi_s)-p_T(\xi_t)\right]=0$, which implies $E(R_{22}^{(1)})=0$. By Lemma A(ii) of Hjellvik et al. (1998), we have

$$
\begin{aligned}
var(R_{22}^{(1)}) &\le \frac{C}{T^3h^{2d}}T^2E\left[|p_T(\xi_s,\xi_t)-p_T(\xi_s)-p_T(\xi_t)|^{2(1+\delta)}\right]^{\frac{1}{1+\delta}}\sum_{j=1}^{\infty}j^2\beta(j)^{\delta/(1+\delta)}\\
&= O\left(\frac{1}{Th^{2d}}h^{\frac{d}{1+\delta}}\right)=O\left(\frac{1}{Th^{\frac{1+2\delta}{1+\delta}d}}\right)=o(1)
\end{aligned}
$$

Then $R_{22}^{(1)}=o_p(1)$ by Chebyshev's inequality. In addition,

$$
\begin{aligned}
var(R_{22}^{(2)}) &\le \frac{C(T-1)^2}{T^3h^{2d}}\sum_{t=1}^{T}var\left[p_T(\xi_t)\right]+\frac{C(T-1)^2}{T^3h^{2d}}T\sum_{j=1}^{T-1}\left|cov\left[p_T(\xi_1),p_T(\xi_{1+j})\right]\right|\\
&\le Ch^{-2d}O\left(h^{2(d+1)}\right)+Ch^{-2d}O\left(h^{2(d+1)}\right)\\
&= O\left(h^2\right)=o(1)
\end{aligned}
$$

Then $R_{22}^{(2)}=o_p(1)$ follows from Chebyshev's inequality. Hence we have proved $A_{22}^{(m)}=o_p(1)$ for $m=1,2,...,d$, if only $Th^{\frac{1+2\delta}{1+\delta}d}\to\infty$ as $T\to\infty$.

Here we have used the fact that $E\left[(p_T(\xi_t))^2\right]=O(h^{2(d+1)})$. To prove this, we define $\frac{Z_{si}-z_i}{h}=v_i$ for $i=1,2,...,d$. And we focus on the first integration inside.

$$
\begin{aligned}
p_T(\xi_s) &= \int\prod_{i=1}^{d}K\left(\frac{Z_{si}-z_i}{h}\right)\left[\Omega(z)^{-1}X_s\left(\frac{Z_{sm}-z_m}{h}\right)\varepsilon_s(e^{iu'z-\phi(u)})\right]dz\\
&= h^d\int\prod_{i=1}^{d}K(v_i)\left[\Omega(Z_s+hv)^{-1}X_sv_m\varepsilon_s\left(e^{iu'(Z_s+hv)-\phi(u)}\right)\right]dv\\
&= h^d\int\prod_{i=1}^{d}K(v_i)\left[\Omega(Z_s)^{-1}X_sv_m\varepsilon_s\left(e^{iu'Z_s-\phi(u)}\right)\right]dv\\
&+ h^dh\int\prod_{i=1}^{d}K(v_i)\left[\sum_{j=1}^{d}\Omega_i(Z_s)^{-1}X_sv_i\right]v_m\varepsilon_s\left(e^{iu'Z_s-\phi(u)}\right)'dv(1+o(1))\\
&= h^d\Omega(Z_s)^{-1}X_s\varepsilon_s\left(e^{iu'Z_s-\phi(u)}\right)\prod_{i\neq m}^{d}\int K(v_i)dv_i\int K(v_m)v_mdv_m\\
&+ h^{d+1}\left(e^{iu'Z_s-\phi(u)}\right)'\varepsilon_s\prod_{i=1}^{d}\Omega_i(Z_s)^{-1}X_s\int v_iK(v_i)v_mdv(1+o(1))\\
&= h^{d+1}\left(e^{iu'Z_s-\phi(u)}\right)'\varepsilon_s\Omega_m(Z_s)^{-1}X_s\int v_m^2K(v_m)dv_m(1+o(1))=O(h^{d+1})
\end{aligned}
$$

where we have used the fact that $\int vK(v)dv = 0$ and $\Omega_i(Z_s)^{-1} = \frac{\partial}{\partial x_i}\Omega(Z_s)^{-1}$. Then

$$\begin{aligned}E\left[(p_T(\xi_s))^2\right] &= \int \left[h^{d+1}\left(e^{iu'Z_s-\phi(u)}\right)' \varepsilon_s \Omega_m(Z_s)^{-1} X_s \int v_m^2 K(v_m)dv_m(1+o(1))\right]^2 f_Z(Z_s)dZ_s \\ &= O(h^{2(d+1)})\end{aligned}$$

Next we prove $A_{22}^{(0)} = \tilde{U} + o_p(1)$. Let

$$\xi_t = (X_t^{'}, Z_t^{'}, Y_t)$$

and define

$$\begin{aligned}\Psi(\xi_s, \xi_t) &= \frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{Z_{si}-Z_{ti}}{h}\right)\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \varepsilon_s \psi_t(u) \\ &+ \frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{Z_{ti}-Z_{si}}{h}\right)\frac{\Omega(Z_s)^{-1}}{f_Z(Z_s)} X_t \varepsilon_t \psi_s(u)\end{aligned}$$

Then we have

$$A_{22}^{(0)} = \frac{1}{T^{3/2}} c_0 \sum_{1\le s<t\le T} \Psi(\xi_s, \xi_t)$$

It is obvious that

$$A_{22}^{(0)} = \frac{c_0}{2}\sqrt{T}\frac{2}{T(T-1)} \sum_{1\le s<t\le T} \Psi(\xi_s, \xi_t) + o_p(1)$$

So we now concentrate on $U_T \equiv \frac{2}{T(T-1)}\sum_{1\le s<t\le T}\Psi(\xi_s, \xi_t)$. Define

$$\begin{aligned}\Psi(\xi_s) &= E\left[\Psi(\xi_s, \xi_t)|\xi_s\right] \\ &= E\left[\frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{Z_{si}-Z_{ti}}{h}\right)\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} X_s \varepsilon_s \psi_t(u)|\xi_s\right] \\ &= \int \frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{Z_{si}-z_i}{h}\right)\Omega(z)^{-1} X_s \varepsilon_s \left(e^{iu'z} - \phi(u)\right) dz\end{aligned}$$

and define

$$\widehat{U}_T = \frac{2}{T}\sum_{s=1}^{T}\Psi(\xi_s)$$

Next we will reconcile the relatively slow convergence properties of nonparametric kernel estimator with the classical properties of sample average. Note that U_T is a second-order U-statistic: this structure permits proper accounting of the "overlaps" in the kernel estimators that comprise U_T. To establish $\sqrt{T}$-consistency and asymptotic normality of U_T, we follow Lemma 3.1 of Powell, Stack and Stoker (1998) in a time series context to show the asymptotic equivalence of U_T and $\widehat{U}_T$:

$$\sqrt{T}(U_T - \widehat{U}_T) = o_p(1)$$

if only $E\left[\|\Psi(\xi_s,\xi_t)\|^2\right]=o(T)$.

To see this, notice that

$$E\left[\|\Psi(\xi_s,\xi_t)\|^2\right]=\int \frac{1}{h^{2d}}\left\|\prod_{i=1}^{d}K\left(\frac{Z_{si}-Z_{ti}}{h}\right)\right\|^2\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}X_s\varepsilon_s\psi_t(u)+\frac{\Omega(Z_s)^{-1}}{f_Z(Z_s)}X_t\varepsilon_t\psi_s(u)\right]^2$$

$$\times f_Z(Z_t)f_Z(Z_s)dZ_tdZ_s$$

$$=\int \frac{1}{h^{d}}\left\|\prod_{i=1}^{d}K\left(v_i\right)\right\|^2\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}X_s\varepsilon_s\psi_t(u)+\frac{\Omega(Z_t+hv)^{-1}}{f_Z(Z_t+hv)}X_t\varepsilon_t\left(e^{iu'(Z_t+hv)}-\phi(u)\right)\right]^2$$

$$\times f_Z(Z_t)f_Z(Z_t+hv)dZ_tdv$$

$$=O\left(h^{-d}\right)=O\left(T\frac{1}{Th^d}\right)$$

where the second equality uses the change-of-variables from (Z_t,Z_s) to $(Z_t,v=(Z_s-Z_t)/h)$, with Jacobian h^{-d} and the third equality uses the continuity of f. Consequently, we have $E\left[\|\Psi(\xi_s,\xi_t)\|^2\right]=o(T)$ if and only if $Th^d\to\infty$ as $h\to 0$.

Thus we have

$$\sqrt{T}U_T\sim\sqrt{T}\widehat{U}_T=\frac{1}{\sqrt{T}}\sum_{s=1}^{T}2\Psi(\xi_s)$$

and

$$\Psi(\xi_s)=\int\frac{1}{h^d}\prod_{i=1}^{d}K\left(\frac{Z_{si}-z_i}{h}\right)\Omega(z)^{-1}X_s\varepsilon_s\left(e^{iu'z}-\phi(u)\right)dz$$

$$=\int\prod_{i=1}^{d}K\left(w_i\right)\Omega(Z_s+hw)^{-1}\psi_{(u)}(Z_s+hw)dwX_s\varepsilon_s$$

$$=\Omega(Z_s)^{-1}\psi_s(u)X_s\varepsilon_s+o_p(1)$$

Then we have

$$A_{22}^{(0)}=\frac{1}{\sqrt{T}}\sum_{s=1}^{T}c_0\Omega(Z_s)^{-1}\psi_s(u)X_s\varepsilon_s+o_p(1)$$

and complete the proof of Lemma 3.

Proof of Propostion 8.

Let $S_t=c_0\Omega(Z_t)^{-1}\psi_t(u)X_t\varepsilon_t$. By Assumption 2, it is easy to show that $\{S_t,\mathcal{F}_t\}$ is an adapted stationary martingale difference sequence. Define

$$V_n\equiv var\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}S_t\right)=var(S_t)\equiv V$$

The equality above comes from stationarity of S_t. In order to make use of Cramer-Wold device, we define any real $k\times 1$ vector λ such that $\lambda'\lambda=1$. From now on we concentrate on $\lambda'V^{-1/2}S_t$. It is obvious that $\lambda'V^{-1/2}S_t$ is also a stationary martingale difference sequence. Therefore one can apply

Brown's (1971) theorem. Notice that

$$var(\lambda' V^{-1/2} S_t) = \lambda' V^{-1/2} V V^{-1/2} \lambda = 1 \tag{.3}$$

By Brown's theorem, $\frac{\sum_{t=1}^{T} \lambda' V^{-1/2} S_t}{var\left(\sum_{t=1}^{T} \lambda' V^{-1/2} S_t\right)} \xrightarrow{d} N(0,1)$ if

$$\frac{1}{var\left(\sum_{t=1}^{T} \lambda' V^{-1/2} S_t\right)} \sum_{t=1}^{T} E\{(\lambda' V^{-1/2} S_t)^2 \mathbf{1}\left[|\lambda' V^{-1/2} S_t| > \epsilon \cdot std(\sum_{t=1}^{T} \lambda' V^{-1/2} S_t)\right]\} \to 0, \forall \epsilon > 0 \tag{.4}$$

and

$$\frac{1}{T} \sum_{t=1}^{T} \lambda' V^{-1/2} S_t S_t' V^{-1/2} \lambda - 1 \xrightarrow{p} 0 \tag{.5}$$

Given .3 and stationariy assumption, it suffices for .4

$$E\{(\lambda' V^{-1/2} S_t)^2 \mathbf{1}\left[|\lambda' V^{-1/2} S_t| > \epsilon\sqrt{T}\right]\} \to 0 \tag{.6}$$

Since $E\left[\left(\lambda' V^{-1/2} S_t\right)^2\right] = 1 < \infty$, and $\mathbf{1}\left[|\lambda' V^{-1/2} S_t| > \epsilon\sqrt{T}\right] \xrightarrow{p} 0$, (.6) is satisfied by dominated convergence theorem. This proves the lindeberg condition in Brown's martingale central limit theorem. We now turn to verify (.5), for which it suffices to show that

$$\frac{1}{T} \sum_{t=1}^{T} \lambda' V^{-1/2} S_t S_t' V^{-1/2} \lambda \xrightarrow{p} E\left(\lambda' V^{-1/2} S_t S_t' V^{-1/2} \lambda\right) = 1 \tag{.7}$$

It is straightfoward to show (.7) by stationary and ergodic theorem (e.g. White) if the following condition is satisfied

$$E|\lambda' V^{-1/2} S_t S_t' V^{-1/2} \lambda| < \infty \tag{.8}$$

Since λ and V are finite, it is equivalent to show the expected absolute value of each element of $E|S_t S_t'|$ is finite, that is,

$$E|S_t S_t'|_{i,j} < \infty \quad \text{for} \quad i, j = 1, ..., d_x \tag{.9}$$

By Cauchy–Schwarz inequality, it is suffices to show

$$E(|S_{tj}|^2) < \infty \quad \text{for} \quad j = 1, ..., d_x \tag{.10}$$

where

$$S_{tj} = \left[\sum_{p=1}^{d_x} \Omega(Z_t)^{-1}_{jp} X_{tp} \varepsilon_t\right] \phi_t(u)$$

Then

$$
\begin{aligned}
E(|S_{tj}|^2) &= E\left(\left|\sum_{p=1}^{d_x} \Omega(Z_t)^{-1}_{jp} X_{tp}\varepsilon_t\right|^2 |\phi_t(u)|^2\right) \\
&\le \left[\left(E\left|\sum_{p=1}^{d_x} \Omega(Z_t)^{-1}_{jp} X_{tp}\varepsilon_t\right|^4\right)^{\frac{1}{4}}\right]^2 \left(E|\phi_t(u)|^4\right)^{\frac{1}{2}} \\
&\le \left(\sum_{p=1}^{d_x} E\left|\sum_{p=1}^{d_x} \Omega(Z_t)^{-1}_{jp} X_{tp}\varepsilon_t\right|^4\right)^{\frac{1}{2}} \left(E|\phi_t(u)|^4\right)^{\frac{1}{2}} < \infty
\end{aligned}
$$

by Assumption 2. The second line follows from Minkowski inequality, and the third line follows triangular inequality. This complete the proof of Theorem 1.