ESSAYS ON TESTING FUNCTIONAL COEFFICIENT MODELS AND APPLICATIONS IN FINANCE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Xingtong Zhang

August 2020

ESSAYS ON TESTING FUNCTIONAL COEFFICIENT MODELS AND APPLICATIONS IN FINANCE

Xingtong Zhang, Ph.D.

Cornell University 2020

This dissertation consists of three essays on specification tests on functional coefficient models via the Fourier transform and applications on conditional asset pricing models.

The first essay, "An Asymptotically Efficient Test for Functional Coefficient Models", proposes 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, my 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 my 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 the second essay titled "A Consistent Model Specification Test for Functional Coefficient Models", I propose a consistent test for model specifications in functional coefficient models via discrete Fourier transform (DFT). The DFT of the sample score function can extract the local property of unknown parameters over the state variable. Therefore, my test avoids nonparametric estimation

and is asymptotically more efficient than the existing nonparametric tests. It can detect a class of local alternatives at the parametric rate. Furthermore, my test allows the regressors and the state variables to be the same and is also robust to heteroscedasticity and serial correlation. Simulation studies show that the proposed test has reasonable size and excellent power against various misspecifications of coefficient functions. The two essays both aim at improving the efficiency of the tests over existing nonparametric tests in the literature. Rather than the same, they can be viewed as complement to each other. The first involves a step of nonparametric estimation and thus can test part of coefficient functions, rather than the joint test for all of the coefficient functions in the model. The drawback is that it is not tuning parameter free. The second uses a score function approach which is based on the residuals from the parametric estimation and is free of nonparametric estimation. Both have their merits and limitations and together provide a system of more efficient methods that can be applied in various economics circumstances.

The third essay, "How Does Smooth Structural Change Affect Asymmetric Dependence in Foreign Exchange market?", I use a copula approach based on Patton (2006) to model this asymmetric exchange rate dependence and investigate how different but reasonable specification of marginal distribution affect the asymmetric behavior between mark-dollar and yen-dollar exchange rates. Central banks are facing a trade-off between export competitiveness and price stability, which will result in an asymmetric dependence behavior among currencies during joint appreciations versus during joint depreciation. It is plausible that the change pace of the underlying economic mechanism and technological progress can cause the structural change of exchange rate in a country. Furthermore, since the pattern of correlation or dependence structure is determined by second or higher order moment, we would expect the marginal structural change in volatility to change the dependence structure in joint distribution. This chapter can also serve as an empirical evidence of implementing smooth structural change into test for asymmetric correlations, proposed by Hong, Tu and Zhou (2007).

BIOGRAPHICAL SKETCH

Xingtong Zhang was born in Beijing, China in February 1989. He studied Computer Science and Finance at Nankai University of China and earned his B.A. degree in Engineering and Economics in July 2011. He was admitted to Cornell M.S. program in School of Applied Economics and Management in May 2012. He continued his graduate studies in economics in the Department of Economics at Cornell University and will earn his Ph.D. degree in August 2020.

This document is dedicated to my parents Yongshan Zhang and Ying Cui.

ACKNOWLEDGEMENTS

I am deeply indebted to my advisor Prof. Yongmiao Hong for his continued support during my Ph.D. studies. His patience, motivation, and excellent guidance helped me in all the time of research and writing of this thesis.

I would like to express my deepest gratitude to Prof. Nicholas M. Kiefer and Prof. Jörg Stoye, for their helpful discussions and encouragement as well as serving on my committee.

I also thank Xin Wang, Prof. Hong's wife. She treated me like her son and gave me the guidance and encouragement when I need them the most. This thesis could not have been completed without her support.

I also thank Zhonghao Fu and Liyuan Cui, former students of Prof. Hong, for their generous help and valuable suggestions.

Last but not least, I want to thank my parents for their love and support. They are my source of strength during my job market period.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 An Asymptotically Efficient Test for Functional Coefficient Models

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 condi-

[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.

tional 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 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

[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.

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 pa-

[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.

rameters will remain in the residuals estimated using an ordinary least squares regression.

I propose an alternative approach, using a Fourier transform to construct a convenient and consistent specification test for functional coefficient models. My 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, I transform 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. My approach has several advantages over existing parametric and nonparametric tests in literature.

First, compared with nonparametric tests, and despite including a nonparametric estimation step, my 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, my 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), my 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, my 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, my 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 my approach, I apply my test to examine the validity of various conditional asset pricing models. Using daily, weekly, and monthly U.S. stock market data, I 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, I compare my 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, I follow the choices in Nagel and Singleton (2011) and Li et al. (2015). my method rejects models that are not rejected by the GLR and Wald-type tests, especially when I use all three state variables from the literature.

1.2 A Consistent Model Specification Test for Functional Coefficient Models

Functional coefficient models have been extensively applied to capture nonlinear dynamics and evolutions in economic and financial phenomena. As natural generalizations of classical linear regression models, the functional coefficient models allow the coefficients to be functions of observable state variables, which is great useful in capturing such nonlinear features as asymmetric cycles, bi-modality and so on. In addition, they also enjoy great flexibility and interpretability as compared to other semiparametric models including the projection pursuit (Huber 1985), the sliced inverse regression (Li 1991), the single index models(Härdle and Stoker 1990), the additive models (Breiman and Friedman 1985, Hastie and Tibshirani 1987), low-dimensional interaction models (Friedman 1991, Gu and Wahba 1992, Stone et al. 1997) and partially linear models (Wahba 1984, Green and Silverman 1994, Fan et al. 1998, Heckman et al. 1998, Fan et al. 2003).

Many popular nonlinear models can also be regarded as special cases of functions coefficient models. For example, the threshold model considered by Tong (1990) and Hansen (1997, 2000) assumes the coefficient functions to be step functions of some observed state variables. Teräsvirta (1994) models the coefficient functions as smooth functions (e.g. logistic functions) of state variables, and proposes the smooth transition model. If the state variable follows an uniform(0,1) distribution, then the functional coefficient model could be regarded as a linear quantile regression (Koneker, 2005). Due to the meaningful economic interpretations of functional coefficient model, many studies use this model to capture nonlinear relationship among economic and financial variables. For example, literature in asset pricing (e.g.,Shanken 1990, Ferson and Harvey 1999, French et al. 1987, Bansal et al. 1993), Lewellen and Nagel 2006, Roussanov 2014, and so on) models the factor loadings as functions of some state variables, which are proxy variables represent the unobserved information set of investors. By introducing the coefficient function, these studies explain the failures of unconditional asset pricing models in analyzing cross sectional returns of different assets. In labor economics, Card (2001) proposes that the returns to schooling should depend on the after-school work experience in nonlinear form. The linear additive separable setting of education and experience in the model will underestimate returns to education. In international economics, Borensztein et al. (1998) build a threshold regression model to investigate the nonlinear effect of FDI on economic development.

Although a parametric form of coefficient function can greatly facilitate the estimation and testing of functional coefficient model, a parametric setting has the risk of misspecification, which may lead to inconsistent estimation and misleading empirical results. Considering this, nonparametric methods have been proposed for the estimation of functional coefficient model. For example, 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 in time series context. However, by introducing the nonparametric coefficient functions, one avoid the model misspecification problem at the cost of estimate efficiency. That is, the asymptotic convergence rate of the estimated parameters and coefficient functions is much slower than the parametric convergence rate. Moreover, it is usually difficult to provide economic interpretations of the dependence between the unknown regression coefficients β on the state variable Z_t when the dimension of state variable is large. Therefore, testing whether coefficient functions are constant of zero or in a certain parametric form is an important theoretical and empirical issue in fitting functional coefficient models. In an important pioneering work, Fan et al. (2001) propose a generalized likelihood ratio (GLR) test by comparing the likelihood of the parametric specification under the null and the nonparametric specification of coefficient functions under the alternative. Under suitable choice of smoothing parameters, it can detect the local alternative with convergence rate $T^{-\frac{4}{9}}$. Ang and Kristensen (2012) apply the GLR to test pricing errors in conditional CAPM models. Li et al. (2002) measure integrated squared difference between the parametric and nonparametric estimator of the coefficient functions to test parametric specifications under the null hypothesis. Hong and Lee (2013) consider a model where the conditional expectation is nonparametrically specified functions, which can be extended to the functional coefficient models, and improve the efficiency using a loss function approach. The essence is to reduce the variance of GLR test statistic by eliminating the first order term due to the use of particular loss functions.

In this chapter, we propose a unified approach for inference in functional coefficent models via discrete Fourier transform (DFT). Unlike the existing tests that focus on time domain analysis, we investigate the specification of certain functional coefficient models in frequency domain. The intuition is straightforward: if the true model parameter is correctly specified, then the conventional estimation method (depend on the model specification under the null hypothesis) will fail to estimate it consistently. Consequently, such information will be captured by the sample score function. Even though the first order condition guarantees the sample mean of score function is identically 0, by doing a Fourier transform with respect to the state variables, we can extract the local property of unknown parameters over the state variable. Compared with the existing specification tests in functional coefficient models literature, the proposed test has the following appealing features.

First, our test is widely applicable for various specifications of coefficient functions, which is more general than the existing parametric test such as Hansen (1997, 2000) for threshold regression, Teräsvirta (1994, 1998) for smooth transition model. The only input of our test is the sample score function obtained by maximum likelihood estimation. In fact, the sample score function exists as long as the coefficient function is second order differentiable with respect to the unknown parameters that could be consistently estimated with $\sqrt{T}$ convergence rate under the null hypothesis. Even the score function is non-derivable with respect to some sup-convergence parameters, such as the threshold value in threshold regressions, our test is still applicable. Therefore, we could check model specifications for linear models, threshold regressions, smooth transition models and so on.

Second, our test can detect a class of local alternatives that converges to the null hypothesis at a faster rate than the existing nonparametric tests such as Fan et al. (2001), Li et al. (2002), Hong and Lee (2013). The rate of local alternatives that our test can detect is $T^{-1/2}$. In contrast, the upper bound convergence rate of the aforementioned nonparametric test is $T^{-4/9}$, which is lower than our DFT based test statistic. This is a appealing advantage of the DFT. Since we only need to estimate the model under the null hypothesis, we can achieve the parametric convergence rate.

Third, our test does not need the smoothed nonparametric estimation. It avoids the delicate business of choosing a bandwidth and free of the “curse of dimensionality” problem. Therefore, we do not need to impose restrictive assumptions on the dimension of state variables. In comparison, the convergence rate of Fan et al.’s (2001) GLR rely on the dimension of state variables. The power of Fan et al.’s (2001) GLR test also depend on the choice of bandwidth. While they propose a wild bootstrap to relieve this problem the power of their testis still sensitive to the choice of bandwidth in finite sample.

Fourth, we allow the regressors and state variables to be the same. This is of particular importance in testing certain parametric specifications such as SETAR. If the SETAR assumes the coefficients to be indicator functions of the regressors themselves, nonparametric estimations will fail since the functional coefficient model is not even identified in this case.

Finally, our test is robust to conditional heteroskedasticity and serial correlation of unknown form in error terms, and hence broaden the applicability of the proposed test.

1.3 How Does Smooth Structural Change Affect Asymmetric Dependence in Foreign Exchange market?

Asymmetric dependence of stock returns with the market indices have been widely studied in literatures. As reported in Erb et al. (1994), Karolyi and Stulz (1996), Ang and Bekaert (2002), Longin and Solnik (2001), and Ang and Chen (2002), two equity returns exhibit greater correlation during joint appreciations than during joint depreciations. Hong, Tu and Zhou (2007) proposed a model-free test for asymmetric correlations which allows for general distributional assumptions, such as the GARCH process.

Such a dependence structure also exists between foreign exchange rates. For example, Patton (2006) uses a copula approach to investigate the dependence between the Deutsche mark and the yen over the period January 1991 to December 2001. He finds strong evidence that these two exchange rates exhibit stronger dependence during joint depreciation in the pre-euro period, whereas an opposite type of asymmetric dependence pattern in the post-euro period.[5] He also proposes that this asymmetric dependence is possibly caused by asymmetric behavior of central banks in reaction to exchange rate movements: A desire to maintain the competitiveness of their exports to the U.S. with other countries exports to the U.S. would lead the central bank to intervene to ensure a matching depreciation of their currencies against the dollar whenever the other currency depreciated against the dollar.[6]

The study of asymmetric dependence feature is important for many reasons. First, as mentioned in Hong, Tu and Zhou (2007), hedging relies heavily on he dependence structure between the assets hedged and the financial instruments hedged. The presence of asymmetric feature can affect the effectiveness in hedging. Second, in the pricing of financial options or in the calculation of portfolio Value-at Risk (VaR) or in a forecast situation where the loss function is unknown, we require for some feature of conditional joint distribution, say asymmetric dependence. Third, the theory of portfolio diversification might be questionable if assets tend to fall or rise simultaneously.

---

[5]The mark/euro exchange rate was fixed since the inroduction of the euro on January 1, 1999, but the mark was still used for transactions in Germany until the end of 2001.

[6]Such a senario was also considered by Takagi (1999).

Despite suffered from model specification, similar to Hong, Tu and Zhou (2007), Patton (2006) uses generalized GARCH process to model marginal distribution, which can capture the feature of volatility clustering in foreign exchange market. However, GARCH models impose a key assumption of stationarity. Note that the specification of marginal distribution plays an important role in modeling joint distribution (copula approach), especially when modeling dependence structure. In particular, the evolution of marginal structures can be a source of time variation in dependence behavior. Naturally we may ask how does the structural change in marginal distributions affect the asymmetric dependence feature described in Patton (2006).

It is plausible that the change pace of the underlying economic mechanism and technological progress can cause the structural change of exchange rate in a country. Furthermore, since the pattern of correlation or dependence structure is determined by second or higher order moment, we would expect the marginal structural change in volatility to change the dependence structure in joint distribution.

Naturally we are facing the decision between abrupt and smooth changes. In copula literatures, there have been numerous papers discussing abrupt structural change in marginal distributions, e.g. Kramer and Kampen (2011), Na, Lee and Lee (2012). Patton (2006) also considered abrupt structural change at the point of introduction of euro for his marginal distributions. However, Hansen (2001) provide a guidance for this, "...might seem more reasonable to allow a structural change to take a period of time to take effect". Indeed, such changes in foreign exchange market induced by policy switch, preference changes, and technology progress usually exhibit evolutionary changes in a long term. In particular, volatility may take time for the market to achieve some consensus. Therefore, I follow the local QMLE method, proposed by Chen and Hong (2014), to estimate the smooth time-varying GARCH parameters and thus make a modified specification of marginal distribution in Patton (2006).

In this chapter, I follow copula approach in Patton (2006) to model the conditional tail dependence and show how does this smooth structural change affect the dependence structure in foreign exchange market and try to extrapolate the findings. Here, the modification of smooth structural change in GARCH has its own economic intuition. As mentioned above, structural change in exchange rates can be viewed as a evolution of underlying economic conditions. By investigating this effect on dependence structure, we are essentially asking how does the domestic conditions affect central banks' preferences. Furthermore, the nature of copula allows us to separate the variation between local economics and the dependence structure.

CHAPTER 2

AN ASYMPTOTICALLY EFFICIENT TEST FOR FUNCTIONAL COEFFICIENT MODELS

2.1 Framework and Approach

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

$$Y_t = X_t^\top \beta(Z_t) + \varepsilon_t, \quad t = 1, 2, ..., T, \tag{2.1}$$

where $\{X_t\, Z_t, Y_t\}$ 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 $E(\varepsilon_t | X_t, Z_t) = 0$, implying that both X_t and Z_t are exogenous. However, this might not be true in many applications in economics and finance. When some components of X_t are endogenous, we use an IV, as in Cai et al. (2006). For identification, we assume $E(\varepsilon_t \mid X_t, Z_t) = 0$ and $\Omega(Z) = E[X_t X_t^\top \mid Z_t = Z]$ is nonsingular, for all Z. This excludes cases in which $Z_t = X_t$, because $X_t X_t^\top$ is always singular, unless X_t is a scalar. It will be shown in chapter 3 that this constraint can be released by avoiding the nonparametric estimation.

Notation. Throughout this chapter, $\mathbf{i}$ 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 $\mathrm{Re}(A)$, 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.

2.2 Testing the Constancy of $\beta(Z_t)$

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):

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

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.[1] 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

[1] 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).

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:

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

where Y is the natural logarithm of hourly wage; 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.

2.2.1 Hypotheses of Interest

The hypothesis of interest is

$$\mathbb{H}_{01} : \beta(Z_t) = \beta_0, \text{ for unknown } \beta_0 \in \Theta,$$

versus

$$\mathbb{H}_{A1} : \beta(Z_t) \neq \beta, \text{ for any } \beta \in \Theta,$$

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 $\beta(Z_t)$ 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 $\beta(Z_t)$ 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,[2] and the GLR test statistic can be based on:

$$\lambda_T = \frac{T}{2}\frac{RSS_0 - RSS_1}{RSS_1} \tag{2.4}$$

where $RSS_0 = \sum_{t=1}^{T}\left(Y_t - X_t'\hat{\beta}_0\right)^2$ is the residual sum of squares from the OLS, and $RSS_1 = \sum_{t=1}^{T}\left(Y_t - X_t'\hat{\beta}(Z_t)\right)^2$ 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 $r > 0$,

$$\frac{r\lambda_T - \mu_T}{\sqrt{2\mu_T}} \xrightarrow{d} N(0,1). \tag{2.5}$$

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

[2] This need not contain the true distribution.

the two models, as follows:

$$Q_T = \sum_{t=1}^{T} d[\widehat{m}_h(Z_t)], \tag{2.6}$$

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,

$$\frac{s(K)q_T - \nu_T}{\sqrt{2\nu_T}} \xrightarrow{d} N(0,1), \tag{2.7}$$

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 estimators of $\beta(Z_t)$ 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:

$$I = \int \left[\hat{\beta}(Z) - \hat{\beta}\right]^\top \left[\hat{\beta}(Z) - \hat{\beta}\right] dZ. \tag{2.8}$$

Similarly, Chen and Hong (2012) and Ang and Kristensen (2012) compare the distance between the parametric and nonparametric estimators of $\beta(Z_t)$. 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, I propose a novel consistent test based on a discrete Fourier transform.

2.2.2 Test Statistic

Under $\mathbb{H}_{01}$, $\beta(Z_t)$ does not change with Z_t. Therefore, testing $\mathbb{H}_{01}$ is equivalent to testing whether $\beta(Z_t)$ is independent of 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)[3] test the martingale property of a time series using

$$cov[X_t, h(X_{t-j})] = 0, \tag{2.9}$$

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

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:

$$\sigma_j(u, v) = cov(e^{iuX_t}, e^{ivX_{t-|j|}}). \tag{2.10}$$

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 $\beta(Z_t)$ and Z_t: $cov[\beta(Z_t), e^{iu^\top Z_t}]$. To facilitate our method, define

$$A_1(u) \equiv E\left[\beta(Z_t)\left(e^{iu^T Z_t} - Ee^{iu^T Z_t}\right)\right]. \tag{2.11}$$

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

Intuitively, we can take Taylor's expansion of $e^{iu^T Z_t}$ around 0: $e^{iu^T Z_t} = 1 + \sum_{k=1}^{\infty} \frac{(iu^T Z_t)^k}{k!}$. Now, $A_1(u) = 0$ becomes the covariance between $\beta(Z_t)$ and an infinite series of the moments of Z_t. Thus, linear uncorrelation becomes independence.[4]

[4]A more rigorous proof takes a Fourier transform of $e^{iu^T Z_t}$; see Bierens (1982).

In fact, $A_1(u)$ can be viewed as the coefficient of the de-meaned Fourier transform at the frequency u. Under $\mathbb{H}_{01}$, $\beta(Z_t)$ is a constant function of Z_t and, thus, the Fourier coefficients should all be equal to zero. The de-meaned $e^{iu'Z_t}$ is used because we only care whether $\beta(Z_t)$ 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 $\beta(Z_t)$ now becomes a parametric estimation of the Fourier coefficient in the frequency domain, which improves the efficiency of our test.

Lemma 2.2.1 suggests that we can test $\mathbb{H}_{01}$ based on the sample analog of $A_1(u)$:

$$\widehat{A}_1(u) = \frac{1}{T}\sum_{t=1}^{T}\widehat{\beta}(Z_t)\left[e^{iu^T Z_t} - \frac{1}{T}\sum_{s=1}^{T} e^{iu^T Z_s}\right], \tag{2.12}$$

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

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 Z_t is multidimensional. The local linear estimator is defined as $\hat{\beta}(Z_0) = \hat{\beta}$, where $\{(\hat{\beta}(Z_0), \hat{\beta}'(Z_0)\}$ minimizes the sum of the weighted squares:

$$\sum_{t=1}^{T}\left[Y_t - (\beta(Z_0)', vec(\beta'(Z_0))')\begin{pmatrix} X_t \\ (Z_t - Z_0)\otimes X_t \end{pmatrix}\right]^2 \prod_{i=1}^{d} K_h(Z_{ti} - z_{0i}), \tag{2.13}$$

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, $\beta'(Z_0)$ is the Jacobian matrix of $\beta(Z_0)$, and $\otimes$ is the Kronecker product. As pointed out by Cai

(2009), the estimator $\{(\hat{\beta}(Z_0), \hat{\beta}'(Z_0))\}$ can be obtained by performing an OLS on the following model:

$$\left[\prod_{i=1}^{d} K_h(Z_{ti} - z_{0i})\right]^{1/2} Y_t = \left[\prod_{i=1}^{d} K_h(Z_{ti} - z_{0i})\right]^{1/2} [\beta(Z_0)' X_t + vec(\beta'(Z_0))'(Z_t - Z_0) \otimes X_t] + u_t \quad (2.14)$$

The sample analog $\widehat{A}_1(u)$ can be viewed as a discrete Fourier transform of $\widehat{\beta}(Z_t)$ on the random state variable 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 $\widehat{\beta}(Z_t)$ 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.

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:

$$\widehat{Q}_1 = T \int_{\mathbb{R}^L} ||\widehat{A_1}(u)||^2 W(u) du, \tag{2.15}$$

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 (2.15) 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 (2.15) 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:

$$W_N(u) = \Pi_{k=1}^{q} \frac{1}{\sqrt{2\pi\xi_k}} exp(-\frac{u_k^2}{2\xi_k^2}), \tag{2.16}$$

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:

$$\int_{\mathbb{R}^{d_z}} cos\left(\sum_{i=1}^{d_z}\right) exp\left(-\frac{|u|^2}{2}\right) du = (2\pi)^{\frac{d_z}{2}} exp\left(-\frac{1}{2}|x|^2\right). \tag{2.17}$$

The quadratic form (2.15) can be written as:

$$\widehat{Q}_W = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}\hat{\beta}(Z_s)' V_{st}\hat{\beta}(Z_t), \tag{2.18}$$

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:

$$W_L(u) = \Pi_{k=1}^{q}\frac{\lambda_k}{2}e^{-\lambda_k|u_k|}, \tag{2.19}$$

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 (2.15) 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.

2.3 Testing a Parametric Form of $\beta(Z_t)$

In this section, we are interested in testing a particular parametric form of $\beta(Z_t)$. A simple case of this type of hypothesis is to test a certain value of the coefficients: $\beta(Z_t) = \beta_0$ 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 $\beta(Z_t)$. 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 Ter*ä*svirta (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 Ter*ä*svirta (1994, 1998) and Granger and Ter*ä*svirta (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.

2.3.1 Hypotheses of Interest

The setting of model 2.1 is fairly general. Besides testing the constancy functional coefficients, in some economic circumstances, one may also be interested in checking whether the functional coefficients follow certain parametric forms. In this case, the hypothesis of interest can be written as follows:

$$\mathbb{H}_{02} : \mathbf{R}\beta(Z_t) = \mathbf{R}\beta(Z_t, \theta_0), \text{ for some } \theta_0 \in \Theta, \tag{2.20}$$

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

$$\mathbb{H}_{A2} : \mathbf{R}\beta(Z_t) \neq \mathbf{R}\beta(Z_t, \theta), \quad \forall \theta \in \Theta \tag{2.21}$$

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 $\beta(Z_t) = \beta_0$, 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 $\alpha_i(Z_t) = \gamma(1 - \beta_i(Z_t))$ where γ is an unknown zero-beta rate. This can be conducted by letting $\mathbf{R} = (1, \gamma)$.

2.3.2 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 $\mathbf{R}\beta(Z_t) - \mathbf{R}\beta(Z_t, \theta_0) = 0$. Therefore, we use the Fourier transform rather than the de-meaned transform. Define the following:

$$A_2(u) \equiv E\left[(\mathbf{R}\beta(Z_t) - \mathbf{R}\beta(Z_t, \theta_0))\, e^{iu^T Z_t}\right] \tag{2.22}$$

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

Lemma 2.3.1. $A_2(u) = 0$ *for all* $u \in \mathbb{R}^L$ *if and only if* $\mathbb{H}_{02}$ *holds.*

Compared to Lemma 2.2.1 in section 2.2, besides checking whether the difference $\mathbf{R}\beta(Z_t) - \mathbf{R}\beta(Z_t, \theta_0)$ 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 $\mathbf{R}\beta(Z_t) - \mathbf{R}\beta(Z_t, \theta_0)$ at the frequency u. Under $\mathbb{H}_{02}$, $\mathbf{R}\beta(Z_t) - \mathbf{R}\beta(Z_t, \theta_0) = 0$ and thus all the Fourier coefficients should be equal to zero. Lemma 2.3.1 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 2.3.1 suggests that we can test $\mathbb{H}_{02}$ based on the sample analog of $A_2(u)$:

$$\widehat{A_2}(u) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{R}[\widehat{\beta}(Z_t) - \beta(Z_t, \hat{\theta})]e^{iu^T Z_t}, \tag{2.23}$$

where $\beta(Z_t, \hat{\theta})$ is a parametric estimation of the null model in $\mathbb{H}_{02}$. In this chapter, we adopt the following nonlinear least squares (NLS) estimator

$$\hat{\theta} \;=\; \arg\min_{\theta\in\Theta} \sum_{t=1}^{T} \left[Y_t - X_t'\beta(Z_t;\theta)\right]^2 . \tag{2.24}$$

By FOC, we have

$$\begin{aligned} \frac{d}{d\theta}_{|\theta=\hat{\theta}} \sum_{t=1}^{T} \left[Y_t - X_t'\beta(Z_t;\theta)\right]^2 &= 0 \\ \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t \left[Y_t - X_t'\beta(Z_t;\hat{\theta})\right] &= 0, \end{aligned}$$

where $\beta^{(1)}(Z_t;\hat{\theta}) = \frac{d\beta(Z_t;\theta)}{d\theta}_{|\theta=\hat{\theta}}$ is a $k \times d$ Jacobian matrix.

In the simple case when then hypothesis of interest is $\beta(Z_t) = \beta_0$ for a pre-specified β_0, there is no need to estimate $\beta(Z_t, \theta_0)$. We can simply replace $\beta(Z_t, \hat{\theta})$ by β_0 in $\widehat{A}_2(u)$. For more general cases, estimation of $\beta(Z_t, \theta_0)$ 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 $\beta(Z_t, \theta_0)$ 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:

$$\widehat{Q}_2 = T \int_{\mathbb{R}^L} \|\widehat{A}_2(u)\|^2 W(u) du, \tag{2.25}$$

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. The other weighting functions and measures of deviation introduced in last section are still suitable in this test statistic.

To avoid numerical integration, we can use the normal weighting function in equation 2.16 and the test statistic in 2.25 can be written as

$$\widehat{Q}_{W2} = \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T}[\mathbf{R}\widehat{\beta}(Z_t) - \mathbf{R}\beta(Z_t, \hat{\theta})]' V_{st}[\mathbf{R}\widehat{\beta}(Z_t) - \mathbf{R}\beta(Z_t, \hat{\theta})] \tag{2.26}$$

where $V_{st} = e^{-\frac{1}{2}\|Z_s - Z_t\|^2}$.

2.4 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.

2.4.1 Assumptions

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

Assumption 2.4.1. Let$(\Omega, \mathbb{F}, P)$ be a complete probability space. The stochastic process $\{X_t, Z_t\}$, 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 2.4.2. (a) The joint density $f_Z(Z)$ of 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 Z over the support of Z_t.

Assumption 2.4.3. $\{\varepsilon_t\}$ is a martingale difference sequence, that is, $E(\varepsilon_t \mid I_{t-1}) = 0$ where $I_{t-1} = \{X_t, Z_t, X_{t-1}, Z_{t-1}, \varepsilon_{t-1}, ...\}$ and $v_{xz}(X, Z) = E(\varepsilon_t^2 \mid X_t = X, Z_t = Z)$. In addition, $E|\Omega(Z_t)^{-1}_{jp} X_{tp}\varepsilon_t|^4 < \infty$ for $j, p = 1, ..., K$. Also $v_z(Z) \equiv E(\varepsilon_t^2 \mid z)$ is continuous in Z.

Assumption 2.4.4. $\kappa : \mathbb{R}^L \to \mathbb{R}^+$ 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 2.4.5. (i) $\beta(\cdot;\theta)$ is measurable function and twice continuously differentiable with respect to $\theta \in \Theta$; $\theta_* \in \text{int}\Theta$ 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 2.4.6. $W : \mathbb{R}^L \to \mathbb{R}^+$ is a nonnegative symmetric integrable function with $\int_{\mathbb{R}^L} \|u\|^4 W(u)du < \infty$.

Assumption 2.4.1 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 2.4.2 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 2.4.2(b) requires only the smoothness condition of $f_{X,Z}(x, z)$ on the state variables, assuming the regressors can be discrete.

Assumption 2.4.3 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 2.4.4 and 2.4.5 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 2.4.4 is a mild condition on $W(u)$, ensuring the existence of the integral in (2.15). Any density function with a finite fourth-order moment satisfies this condition.

2.4.2 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 2.4.1. *Suppose Assumptions 2.4.1 - 2.4.4 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*

$$\mathcal{K}_1(u,v) \equiv cov(\mathcal{G}_1(u), \mathcal{G}_1^*(v)) = 4E\left[r_1(\xi_t,u)r_1^*(\xi_t,v)'\right]$$

and

$$r_1(\xi_t,u) \equiv \frac{1}{2}c_0\Omega(Z_t)^{-1}\psi(u,Z_t)\left[\nu_X(Z_t)\nu_Y(Z_t) + X_t\mathbf{Y}_t\right]$$

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 2.4.1 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 $\beta(Z_t,\theta_0)$ 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 2.4.2. *Suppose Assumptions 2.4.1 - 2.4.5 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

$$\mathcal{G}_2(u) = \mathcal{G}_{21}(u) + \mathcal{G}_{22}(u),$$

is a complex-valued Gaussian process with 0 *mean and covariance kernel as*

$$
\begin{aligned}
\mathcal{K}_2(u,v) &\equiv cov(\mathcal{G}_2(u), \mathcal{G}_2^*(v)) \\
&= \mathcal{K}_{21}(u,v) + \mathcal{K}_{22}(u,v) - 2cov\left[G_{21}(u), G_{22}^*(v)\right]
\end{aligned}
$$

where

$$
cov\left[G_{21}(u), G_{22}^*(v)\right] = E\left[c_0\Omega(Z_t)^{-1}\phi(u, Z_t)X_t G_t^*(v)'v(X_t, Z_t)\right]
$$

and $\mathcal{G}_{21}(u)$ *is a complex-valued Gaussian process with* 0 *mean and covariance kernel as*

$$
\begin{aligned}
\mathcal{K}_{21}(u,v) &\equiv cov(\mathcal{G}_{21}(u), \mathcal{G}_{21}^*(v)) \\
&= 4E\left[r_{21}(\xi_t, u) r_{21}^*(\xi_t, v)'\right]
\end{aligned}
$$

and

$$
r_{21}(\xi_t, u) \equiv \frac{1}{2} c_0 \Omega(Z_t)^{-1}\phi(u, Z_t)\left[v_X(Z_t)v_Y(Z_t) + X_t \boldsymbol{Y}_t\right]
$$

and $\mathcal{G}_{22}(u)$ *is a complex-valued Gaussian process with* 0 *mean and covariance kernel as*

$$
\begin{aligned}
\mathcal{K}_{22}(u,v) &\equiv cov(\mathcal{G}_{22}(u), \mathcal{G}_{22}^*(v)) \\
&= E\left[G_t(u) G_s(v)^* \varepsilon_t \varepsilon_s\right],
\end{aligned}
$$

where

$$
G_t(u) = E\left[\beta^{(1)}(Z_t; \theta_0) e^{iu'Z_t}\right] E\left[\beta^{(1)}(Z_t; \hat{\theta})' X_t X_t' \beta^{(1)}(Z_t; \theta_0)\right]^{-1} \beta^{(1)}(Z_t; \hat{\theta})' X_t,
$$

and $*$ *denotes the complex conjugate.*

Proposition 2.4.2 shows that the distribution of $\hat{A}_2(u)$ is determined by two components jointly. The first term $\mathcal{G}_{21}(u)$ is derived from the non-parametric estimation $\hat{\beta}(Z_t)$. It is due to analogous argument of U-process in Proposition 2.4.1. The second part $\mathcal{G}_{22}(u)$ is from the NLS estimation $\beta(Z_t, \hat{\theta})$. Since both follows parametric convergence rate, they jointly determine the distribution of

$\hat{A}_2(u)$. Since both part share a common component $X_t\varepsilon_t$, their interactions, the covariance, also affect the final distribution of $\hat{A}_2(u)$.

The bandwidth h is chosen to remove the bias of the non-parametric estimation of $\beta(Z_t)$. 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 2.4.1. *Suppose Assumptions 2.4.1 - 2.4.4 and 2.4.6 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 2.4.2. *Suppose Assumptions 2.4.1 - 2.4.6 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 2.4.1 and theorem 2.4.2 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 $\hat{\beta}_0$ from an OLS of the null linear regression model, and the nonparametric estimator $\hat{\beta}(Z_t)$, for all $t = 1, ..., T$.

Step 2 Compute the $\widehat{Q}$ statistic and the residual $\widehat{\varepsilon_t} = Y_t - X_t'\hat{\beta}(Z_t)$ 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 $Y_t^* = X_t'\hat{\beta}_0 + \widehat{\varepsilon}_t^*$. 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 $P^* < \alpha$.

Our test is tuning parameter free, and the asymptotic distributions are not pivotal because they depend on the unknown data generating process (DGP). Thus, resampling methods are suitable to obtain critical values.

2.4.3 Asymptotic Power

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

Theorem 2.4.3. *Suppose Assumptions 2.4.1 - 2.4.4 and 2.4.6, $Th^L \to \infty$ 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 2.4.4. *Suppose Assumptions 2.4.1 - 2.4.6 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 $\mathbb{H}_{Al}$:

$$\mathbb{H}_{Al} : \mathbf{R}\beta(Z_t) = \mathbf{R}\beta(Z, \theta_0) + \alpha(T)\delta(Z_t), \tag{2.27}$$

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

Theorem 2.4.5. *Suppose Assumptions 2.4.1 - 2.4.4 and 2.4.6 hold, $Th^L \to \infty$ as $T \to \infty$. Then under $\mathbb{H}_{Al}$ and $\alpha(T) = T^{-\frac{1}{2}}$,*

$$\widehat{Q}_1 \xrightarrow{d} \int ||\mathcal{G}_1(u) + \xi_1(u)||^2 W(u) du. \tag{2.28}$$

where $\xi_1(u)$ is a complex-valued non-centrality parameter process, with covariance kernel

$$\mathcal{K}_{\xi_1}(u, v) = \left[\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\right]^2 \lim_{T\to\infty} \frac{2}{T(T-1)} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} E\left[\delta''(Z_t)'\psi_t(u)\psi_s(v)^*\delta''(Z_s)\right] \tag{2.29}$$

Theorem 2.4.6. *Suppose Assumptions 2.4.1 - 2.4.6 hold, and $h = cT^{-\lambda}$ for $\frac{1}{2L+2} < \lambda < \frac{1}{L}$, where $0 < c < \infty$. Then under $\mathbb{H}_{Al}$ and $\alpha(T) = T^{-\frac{1}{2}}$,*

$$\widehat{Q}_2 \xrightarrow{d} \int ||\mathcal{G}_2(u) + \xi_2(u) + \xi_3(u)||^2 W(u) du, \tag{2.30}$$

where $\xi_2(u)$ is a complex-valued non-centrality parameter process, with covariance kernel

$$\mathcal{K}_{\xi_2}(u, v) = \left[\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\right]^2 \lim_{T\to\infty} \frac{2}{T(T-1)} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} E\left[\delta''(Z_t)'\phi_t(u)\phi_s(v)^*\delta''(Z_s)\right] \tag{2.31}$$

and $\xi_3(u)$ is a complex-valued non-centrality parameter process, with covariance kernel

$$\mathcal{K}_{\xi_3}(u,v) = \lim_{T\to\infty} \frac{2}{T(T-1)} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} E\left[\hat{G}_{3t}(u) X_t' X_s \delta(Z_t)\delta(Z_s)\hat{G}_{3s}(v)^*\right] \tag{2.32}$$

Theorem 2.4.5 and 2.4.6 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 Li et al. (2002) 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 non-parametric estimation. Especially, Theorem 2.4.5 shows the power of $\widehat{Q}_1$ is largely affected by the second derivative of function of deviation $\delta(\cdot)$. It implies the smoothness of alternative model may affect the power of our test, which will be illustrated in the simulation study. On the other hand, Theorem 2.4.6 shows that, in addition to the second derivative part, the deviation also plays a role in the parametric estimation. Unsurprisingly, the power will be less affected by the smoothness of alternative model. The simulation studies also reveal this fact.

2.5 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 $Z_t \sim i.i.d.\ U[0,1]$.[5] 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 Z_t to vary between one and three.[6]. Furthermore, to examine the robustness of 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):

[5] We also tried other parametric distributions of Z_t, but the performance of the test is not affected significantly. We do not report the results here, for brevity.

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

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):

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

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 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 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)

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

DGP P2 TAR(1):

$$X_t = \begin{cases} -0.5X_{t-1} - 0.2X_{t-2} + \varepsilon_{1t} & \text{if } X_{t-2} < 0, X_{t-3} < 0, \\ 0.5X_{t-1} + 0.9X_{t-2} + \varepsilon_{2t} & \text{if } X_{t-2} < 0, X_{t-3} \geq 0, \\ 0.1X_{t-1} - 0.3X_{t-2} + \varepsilon_{3t} & \text{if } X_{t-2} \geq, X_{t-3} < 0, \\ -0.1X_{t-1} - 0.5X_{t-2} + \varepsilon_{4t} & \text{if } X_{t-2} \geq 0X_{t-3} \geq 0, \end{cases}$$

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

$$AMS(h) = \sum_{q=1}^{Q} AMS_q(h), \tag{2.33}$$

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

$$AMS_q(h) = \frac{1}{m} \sum_{t=T-qm+1}^{T-qm+m} \left\{ Y_t - X_t^T \hat{\beta}_q(Z_t) \right\}^2, \tag{2.34}$$

and $\hat{\beta}_q(\cdot)$ are computed from the sample $(Y_t, Z_t, X_t), 1 \leq t \leq T - qm,$ 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 A.1–A.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 non-parametric 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 A.4–A.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 C.3 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.

2.6 Empirical Application

2.6.1 Conditional Asset Pricing Models

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:

$$R_{i,t} = \alpha_{i,t} + \beta_{i,t} f_t + \varepsilon_{i,t}, t = 1, 2, ..., T, \tag{2.35}$$

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: $\alpha_{i,t} = \alpha_i(Z_t)$ and $\beta_{i,t} = \beta_i(Z_t)$. 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.

2.6.2 Data

The state variables examined in our study follow from Nagel and Singleton (2011) and Li, Su, and Xu (2015).[7] 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

[7]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.

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 Z_t summarize the information set at time $t-1$. In our context, we consider seven choices of 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.

2.6.3 Test on Conditional CAPM

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): $\beta_i(Z_t) = \beta_0$, where β_0 is an unknown unconditional beta for portfolio i. Second, we examine whether, when conditioning on Z_t, the conditional CAPM can price a single portfolio and multiple portfolios. This amounts to testing the null hypothesis: $\alpha_i(Z_t = 0$, 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.

Table A.7 provides bootstrap p-values for the tests on conditional alpha. 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.

Table A.8 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.

2.6.4 Test on Conditional FF6 and Q-factor Models

As an extension to our empirical analysis, we further test more frequently adopted conditional asset pricing models. For simplicity, here we check the FF six-factors model (Fama and French, 2018) and Q-factor model (Hou et al., 2015).

For FF6 model (Fama and French, 2018), the factors in model 2.35 are

$$\mathbf{f}_t = (r_m, r_{SMB}, r_{HML}, r_{RMW}, r_{CMA}, r_{UMD}). \quad (2.36)$$

For Q-factor model (Hou et al., 2015), the factors in model 2.35 are

$$\mathbf{f}_t = (r_m, r_{SMB}, r_{I/A}, r_{ROE}). \quad (2.37)$$

Table A.9 provides bootstrap p-values for the tests on conditional alpha in conditional FF-6 model (Fama and French, 2018). The results show that, by all three tests, the conditional FF-6 model is strongly rejected for all portfolios, conditioning on either state variable individually. However, the GLR test and the Wald-type test fail to reject the conditional FF-6 model in all cases. Our results show that the conditional FF-6 model cannot explain pricing anomalies in unconditional FF-6 model.

Table A.10 provides the bootstrap p-values of the tests on the constancy of conditional beta in conditional FF-6 model (Fama and French, 2018). 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 G when conditioning on each state variable individually. When the dimension of the state variable increases, our tests are still able to detect the nonc-onstancy in the conditional beta, whereas the GLR test and the Wald-type test fail to reject the null hypothesis. The results are analogous between the value-weighted and equally weighted portfolios.

Table A.11 provides bootstrap p-values for the tests on conditional alpha in conditional Q-factor model (Hou et al., 2015). The results show that, by all three tests, the conditional Q-factor model (Hou et al., 2015) is strongly rejected for all three B/M portfolios, conditioning on either state variable individually. However, the GLR test and the Wald-type test fail to reject the conditional Q-factor model (Hou et al., 2015) in all cases. Our results show that the conditional Q-factor model (Hou et al., 2015) cannot explain pricing anomalies in unconditional Q-factor model (Hou et al., 2015).

Table A.12 provides the bootstrap p-values of the tests on the constancy of conditional in conditional Q-factor model (Hou et al., 2015). 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 G when conditioning on cay and def. When the dimension of the state variable increases, our tests are still able to detect the non-constancy 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.

2.7 Conclusion

This chapter 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.

CHAPTER 3

A CONSISTENT MODEL SPECIFICATION TEST FOR FUNCTIONAL COEFFICIENT MODELS

3.1 Hypotheses and Test Statistic

In this chapter, we introduce the hypotheses of interest and propose a test statistic to detect model specification in functional coefficient models via a DFT approach. The method in this chapter and last chapter both aim at improving the efficiency of the tests over existing nonparametric tests in the literature. In fact, they both can detect a local alternative at a parametric rate. Rather than the same, they can be viewed as complement to each other. Both have their merits and limitations and together provide a system of more efficient methods that can be applied in various economics circumstances. The method in chapter 2 involve a step of nonparametric estimation and thus can test part of coefficient functions, rather than the joint test for all of the coefficient functions in the model. The drawback is that they depend on the choice tunning parameter, the bandwidth of local linear estimation. The test introduced in this chapter uses a score function approach which is based on the residuals from the parametric estimation and is free of nonparametric estimation. Also, since it only involves parametric estimation, it is computationally easier. However, due to the lack of nonparametric estimation, it can only treat all the parameters in the null model as a whole and therefore cannot test one or part of the parameters. As will be illustrated in chapter 4, it is not suitable in the cases testing the validity of conditonal asset pricing models.

Notation. Throughout this paper, **i** denotes the 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 $\mathrm{Re}(A)$, 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 that may vary from case to case. The operator $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ denotes convergence in distribution, and $\Rightarrow$ denotes weak convergence respectively.

3.1.1 Hypotheses of Interest

Let $\{X_t, Z_t\}_{t=1}^{T}$ be jointly strictly stationary processes. We consider the model specification problem for the following functional coefficient model:

$$Y_t = X_t^\top \beta(Z_t) + \varepsilon_t, \tag{3.1}$$

where X_t is a $k \times 1$ vector of explanatory variables, Z_t is a $l \times 1$ vector of state variable, $\beta(\cdot) : \mathbb{R}^l \to \mathbb{R}^k$ is an unknown function, and the error term satisfies $E(\varepsilon_t | Z_t, X_t) = 0$. It is important to point out that, in this chapter, we allow X_t and Z_t to be overlapping or contain the same variables. The identification condition for functional coefficient models is that $\Omega(Z) = E[X_t X_t^\top \mid Z_t = Z]$ is nonsingular, for all Z. When X_t and Z_t are overlapping or contain the same variables, $\Omega(Z) = E[X_t X_t^\top \mid Z_t = Z]$ is always singular and thus the identification condition fails. In this case, any test involving nonparametric estimation will be invalid since the model is not even nonparametrically identified. This will exclude many classical time series models such as TAR and SETAR with state variables as the regressor themselves. However, as will be illustrated below, since the method in this chapter avoid nonparametric estimation by employing the residuals from parametric estimation under the null model, it can be applied in these special cases.

The setting of model (3.1) is fairly general, and many popular time series models could be viewed as special cases.

1. If $\beta(Z_t) = \theta$ for some unknown parameter θ, then the model (3.1) degenerate to the linear regression model:

$$Y_t = X_t'\theta + \varepsilon_t.$$

   This is the conventional linear regression model, and the OLS estimators are consistent.

2. If $X_t = (Y_{t-1}, \cdots, Y_{t-k})'$, $Z_t = (X_{t-d})$ and $\beta(\cdot)$ being the indicator function, then model (3.1) 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.

3. 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 (3.1) 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 $j = 1, 2, \cdots, K$.

4. 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 (3.1) becomes the smooth transition AR model considered by Ter*ä*svirta (1994):

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

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

5. If $Z_t \sim U(0,1)$, then model (3.1) 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 Ter*ä*svirta (1994, 1998) and Granger and Ter*ä*svirta (1993) explore asymptotic results for the smooth transition models. The parametric model could be written as

$$Y_t = X_t'\beta(Z_t;\theta) + \varepsilon_t,$$

where $\beta(\cdot;\theta)$ is a parametric function (e.g., threshold model, quantile regression, etc) of an observable state variables Z_t up to an unknown parameter vector $\theta \in \Theta$, where $\Theta \subset \mathbb{R}^d$ is a parameter space. However, a correct specification of Z_t is quite essential for this model. Misleading result will arise when $\beta(Z_t)$ is mistakenly parameterized. Hence, it is desirable to check whether the specification of parametric form is correct or not. Therefore, one should test

$$\mathbb{H}_0 : \beta(Z_t) = \beta(Z_t; \theta_0), \qquad \text{for unknown } \theta_0 \in \Theta$$

against

$$\mathbb{H}_1 : \beta(Z_t) \neq \beta(Z_t; \theta),$$

for all $\theta \in \Theta$.

Notice that here $\mathbb{H}_0$ and $\mathbb{H}_1$ do not involve the restriction matrix $\mathbb{R}$ in chapter 2, which means we cannot test part or linear combinations of the functional coefficients. This is due to the lack of nonparametric estimation. For circumstances that testing part of the coefficients, such as testing the validity of conditional asset pricing models, one can refer to the method introduced in chapter 2.

3.1.2 Test Statistic

Under $\mathbb{H}_0$, we could estimate the unknown parameter θ via maximum likelihood estimation (MLE) or equivalently, the following nonlinear least squares (NLS) estimator

$$\hat{\theta} \;=\; \arg\min_{\theta\in\Theta} \sum_{t=1}^{T} \left[Y_t - X_t'\beta(Z_t;\theta)\right]^2 . \tag{3.2}$$

By taking the first order difference, we have

$$\frac{d}{d\theta}|_{\theta=\hat{\theta}} \sum_{t=1}^{T} \left[Y_t - X_t'\beta(Z_t;\theta)\right]^2 = 0$$
$$\sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t \left[Y_t - X_t'\beta(Z_t;\hat{\theta})\right] = 0,$$

where $\beta^{(1)}(Z_t;\hat{\theta}) = \frac{d\beta(Z_t;\theta)}{d\theta}|_{\theta=\hat{\theta}}$ is a $k \times d$ Jacobian matrix.

Let $S_t(\theta)$ be the $l \times 1$ vector of score function such that

$$S_t(\theta) = \beta^{(1)}(Z_t;\theta)' X_t \left[Y_t - X_t' \beta(Z_t;\theta) \right].$$

Our test statistic is constructed based on the following Discrete Fourier Transform (DFT) of the sample score function

$$\hat{A}(u) = \frac{1}{T} \sum_{t=1}^{T} S_t(\hat{\theta}) e^{iu'Z_t},$$

where $\hat{\theta}$ is the NLS estimator. Under certain regularity conditions, it is straightforward to show that the probability limit of $\hat{\theta}$, denoted as θ_* exists. And θ_* coincides with θ_0 under $\mathbb{H}_0$. (Use White, 1994, Consistency of Extrema Estimators)

To see why $\hat{A}(u)$ is able to test $\mathbb{H}_0$ against $\mathbb{H}_A$, we decompose $\hat{A}(u)$ in the following way:

$$\begin{aligned}
\hat{A}(u) &= \frac{1}{T} \sum_{t=1}^{T} S_t(\hat{\theta}) e^{iu'Z_t} \\
&= \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t \left[Y_t - X_t' \beta(Z_t;\hat{\theta}) \right] e^{iu'Z_t} \\
&= \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \left[\beta(Z_t) - \beta(Z_t;\hat{\theta}) \right] e^{iu'Z_t} + \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t \varepsilon_t e^{iu'Z_t} \\
&= \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \left[\beta(Z_t) - \beta(Z_t;\theta_*) \right] e^{iu'Z_t} \\
&\quad - \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \left[\beta(Z_t;\hat{\theta}) - \beta(Z_t;\theta_*) \right] e^{iu'Z_t} \\
&\quad + \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})' X_t \varepsilon_t e^{iu'Z_t}.
\end{aligned}$$

The above result shows that the DFT $\hat{A}(u)$ contains three parts. The first component captures the deviation of the parametric function $\beta(Z_t;\theta)$ from the true unknown regression coefficient $\beta(Z_t)$ evaluated at $\theta = \theta_*$. Under $\mathbb{H}_0$, $\theta_* = \theta_0$, it equals 0 for all $u \in \mathbb{R}^l$. While under $\mathbb{H}_A$, we can show that it will converge to a non-zero empirical process indexed by u. The second component captures the estimation uncertainty introduced by the NLS estimator. The third component is associated with the regression disturbance. Under certain regularity conditions, we can show that the last two terms will converge to zero for all u at the parametric rate.

To guarantee that our DFT based approach can detect a great range of model mis-specification of unknown form, we shall examine the deviation of $\hat{A}$ from a zero spectrum at each frequency u. Therefore, we propose the following test statistic:

$$\hat{D} = T \int_{\mathbb{R}^l} \|\hat{A}(u)\|^2 W(u) du, \tag{3.3}$$

where $W(\cdot) : \mathbb{R}^l \to \mathbb{R}^+$ is a nonnegative symmetric weighting function of u and satisfies $\int_{\mathbb{R}^d} W(u)du = 1$. The use of $W(u)$ allows us to consider infinitely many points for u. The test statistic 3.3 involves l-dimensional integration, which is usually calculated using numerical integration or approximated by simulation methods. When the dimension d of Y_t is large, the calculation of this integration would be tremendously computational costly. To avoid the high dimensional numerical integration, one can integrate the objective function out analytically by choosing some suitable weighting function. For example, if $W(u)$ is chosen as the joint standard normal density function, the sample analog of the above objective function has a closed-form expression:

$$\hat{D} = \frac{1}{T} \sum_{t=1}^{T} \sum_{r=1}^{T} S_t(\hat{\theta})' S_r(\hat{\theta}) \exp\left[-\frac{(Z_t - Z_r)^2}{2} \right]$$

3.2 Asymptotic Theory

In this section, we derive the asymptotic null distribution of our test statistic and investigate its asymptotic local power property. We also introduce bootstrap procedures to improve the finite sample performance of the test.

3.2.1 Assumptions

To derive the asymptotic distribution of $\hat{D}$, we first impose the following regularity conditions.

Assumption 3.2.1. (i) The stochastic process $\{X_t', Z_t'\}$ is strictly stationary absolutely regular on $\mathbb{R}^{k+l}$ with mixing coefficient such that $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta-1}{\delta}} < C < \infty$ for some $\delta > 1$; (ii) For some $\upsilon > 0$, $E(\|X_t\|^{4+\upsilon}) < C$ and $E(\|Z_t\|^{4+\upsilon}) < C$; (iii) $E(X_t X_t')$ is finite and nonsingular.

Assumption 3.2.2. The error term ε_t is weakly stationary and satisfies: (i) $E(\varepsilon_t | X_t, Z_t) = 0$ almost surely; (ii) The long run variance is $\Omega_\varepsilon = \sum_{j=-\infty}^{\infty} E(\varepsilon_t \varepsilon_{t+j})$; (iii) $E(\varepsilon_t^4) < C$; (iv) $E(X_t X_t' \varepsilon_t)$ is finite and positive definite.

Assumption 3.2.3. The weighting function $W(\cdot) : \mathbb{R}^d \to \mathbb{R}^+$ is nonnegative symmetric integrable function with $\int_{\mathbb{R}^d} W(u)du = 1$ and $\int_{\mathbb{R}^d} \|u\|^4 W(u)du < \infty$.

Assumption 3.2.4. (i) $\beta(\cdot;\theta)$ is measurable function and twice continuously differentiable with respect to $\theta \in \Theta$; $\theta_* \in \text{int}\Theta$ 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 $1 \le j \le k$, 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 3.2.1 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 3.2.2 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 2.4.2(b) requires only the smoothness condition of $f_{X,Z}(x, z)$ on the state variables, assuming the regressors can be discrete.

Assumption 3.2.3 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 3.2.4 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 2.4.4 is a mild condition on $W(u)$, ensuring the existence of the integral in (2.15). Any density function with a finite fourth-order moment satisfies this condition.

3.2.2 Asymptotic Null Distribution

By the property of the score function, we can show that

$$\begin{aligned}\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)X_t'[\beta(Z_t)-\beta(Z_t;\theta_*)] + \frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\varepsilon_t \\ &= \hat{A}_1(u)+\hat{A}_2(u),\end{aligned} \tag{3.4}$$

where $\hat{G}_t(u)$ is a $l \times 1$ complex-valued vector empirical process such that

$$\hat{G}_t(u) = \left\{I_l e^{iu'Z_t} - \left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\right\}\beta^{(1)}(Z_t;\hat{\theta})'X_t.$$

Under $\mathbb{H}_0 : \beta(Z_t) = \beta(Z_t, \theta_0)$, then $\theta_* = \theta_0$, and we can show that $\hat{A}_1(u) = 0$ for all u, and the second term $\hat{A}_2(u)$ converges to a zero spectrum in frequency domain for all u due to the orthogonality condition that $E(\varepsilon_t|X_t, Z_t) = 0$ a.s.. We can show that under $\mathbb{H}_0$, $\sqrt{T}\hat{A}(u)$ converges to a complex-valued Gaussian process.

Proposition 3.2.1. *Suppose Assumptions 3.2.1-3.2.3 hold. Then under $\mathbb{H}_0 : \beta(Z_t) = \beta(Z_t;\theta_0)$, we have*

$$\sqrt{T}\hat{A}(u) \Rightarrow \mathcal{G}(u), \quad \text{as } T \to \infty$$

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

$$\mathcal{K}(u,v) = \lim_{T\to\infty}\frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}E[G_t(u)G_s(v)^*\varepsilon_t\varepsilon_s],$$

and

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

Here we see that the covariance kernel allows for serial correlated errors terms, which makes our approach to be widely applicable.

Now, the asymptotic distribution of our test $\hat{D}$ is stated in the following Theorem.

Theorem 3.2.1. *Suppose Assumptions 3.2.1-3.2.3 hold. Then under* $\mathbb{H}_0 : \beta(Z_t) = \beta(Z_t; \theta_0)$*, we have*

$$\hat{D} \xrightarrow{p} \int_{\mathbb{R}^l} \|\mathcal{G}(u)\|^2 W(u)du.$$

3.2.3 Asymptotic Local Power

Under the $\mathbb{H}_A$, the first component $\hat{A}_1(u)$ is no longer equal to 0 for all $u \in \mathbb{R}^l$ because it contains the deviation of $\beta(Z_t, \theta_*)$ from the true coefficient function $\beta(Z_t)$. The following proposition shows that $\hat{A}(u)$ converges to a non-zero empirical process over $\mathbb{R}^l$. To investigate the asymptotic power property of our tests, we now consider the following local alternatives

$$\mathbb{H}_A(\Delta_T) : \quad \beta(Z_t) = \beta(Z_t, \theta_*) + \Delta_T \phi(Z_t),$$

where $\beta(Z_t, \theta_*)$ is a parametric function evaluated at θ_*, and $\phi(\cdot) : \mathbb{R}^l \to \mathbb{R}^k$ is a nonrandom function of scaled time Z_t satisfying $E\|\phi(Z_t)\| < C$ and $\phi(\cdot) \neq 0$ on an interval with nonzero Borel measure.

Proposition 3.2.2. *Suppose Assumptions 3.2.1-3.2.3 hold. Then under* $\mathbb{H}_A(\Delta_T)$ *with* $\Delta_T = T^{-1/2}$*, as* $T \to \infty$*,*

$$\hat{A}(u) \Rightarrow \tilde{A}(u),$$

where

$$\tilde{A}(u) = E\left[G_t(u) X_t' \{\beta(Z_t) - \beta(Z_t, \theta_*)\}\right].$$

Theorem 3.2.2. *Suppose Assumptions 3.2.1-3.2.3 hold. Then under* $\mathbb{H}_A(\Delta_T)$ *with* $\Delta_T = T^{-1/2}$*, as* $T \to \infty$*,*

$$\hat{D} \xrightarrow{p} \int_{\mathbb{R}^l} \|\xi(u) + \mathcal{G}(u)\|^2 \, W(u) du,$$

where $\mathcal{G}(u)$ *is defined in Proposition 3.1 and* $\xi(u) = E\left[G_t(u) X_t' \phi(Z_t)\right]$.

Theorem 3.2.2 provides the asymptotic distribution of $\hat{D}$ under the local alternative $\mathbb{H}_A(\Delta_T)$. It shows that our test can detect a class of local alternatives with $\xi(u) \neq 0$ at rate $\Delta_T = T^{-1/2}$. In terms of Pitman's criterion, it is asymptotically more efficient than Fan et al.'s (2001) *GLR* test.

3.3 Monte Carlo Simulations

We now study the finite sample performance of the proposed test through Monte Carlo simulations. We also compare our test with Fan et al.'s (2001) *GLR* test.

To examine the size performance, we consider the following DGPs:

DGP.S1 [linear regression]: $Y_t = 1 + X_t + u_t$, $X_t \sim i.i.d.N(0, 1)$

DGP.S2 [AR(1)]: $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$

DGP.S3 [TAR(1)]: $Y_t = -0.5 Y_{t-1} \mathbf{1}(Y_{t-1} > 0) + 0.7 Y_{t-1} \mathbf{1}(Y_{t-1} \leq 0) + u_t$

DGP.S4[Threshold regression]: $Y_t = 1 + X_t \mathbf{1}(X_t > 0) + (1 + \theta) X_t \mathbf{1}(X_t \leq 0) + u_t$, $X_t \sim i.i.d.N(0, 1)$.

Since we want to check the size performance of our test, the model should be correct specified under the null hypothesis. Therefore, we set the parametric form $\beta(Z_t, \theta) = \theta$ for DGP.S1-S2, and set $\beta(Z_t, \theta) = \theta_1 \mathbf{1}(Z_t > c) + \theta_2 \mathbf{1}(Z_t \leq c)$ for $Z_t = Y_{t-1}$ and $Z_t = X_t$ respectively for DGP.S3-S4. That is, we fit a linear regression for DGP.S1-S2 and fit a threshold regression for DGP.S3-S4. In addition, to check the power performance of our test, we set $\beta(Z_t, \theta) = \theta$, i.e., estimate a linear regression of Y_t on $(1, X_t)$ for the following DGP.P1-P4:

DGP.P1 [Quadratic regression]: $Y_t = 1 + X_t + \theta X_t^2 + u_t$

DGP.P2 [Nonlinear regression]: $Y_t = 1 + F(X_t)X_t + 2\left[1 - F(X_t)\right] X_t + u_t$

DGP.P3 [STAR(1)]: $Y_t = -0.5F(Y_{t-1})Y_{t-1} + 0.7\left[1 - F(Y_{t-1})\right] Y_{t-1} + u_t$

DGP.P4 [Threshold regression]: $Y_t = 1 + X_t\mathbf{1}(X_t > 0) + (1 + \theta)X_t\mathbf{1}(X_t \leq 0) + u_t$

where $F(X_t) = \left[1 + \exp(-X_t)\right]^{-1}$.

To assess the power performance of our test under the nonlinear null hypothesis, we fit a threshold regression for DGP.P1-P3. That is, we set $\beta(Z_t, \theta) = \theta_1 \mathbf{1}(Z_t > c) + \theta_2 \mathbf{1}(Z_t \leq c)$ with $Z_t = X_t$ for DGP.P1-P2, and $Z_t = Y_{t-1}$ for DGP.P3, and regard these DGPs as DGP.P5-P7 respectively.

For each DGP, we consider three cases for the error terms u_t: (i) the i.i.d. case, where $u_t \sim i.i.d.N(0, 1)$; (ii) the conditional heteroskedastic case, where $u_t = h_t^{1/2} v_t$, $h_t = 0.2 + 0.3u_{t-1}^2$, $v_t \sim i.i.d.N(0, 1)$; (iii) the serially correlated case, where $u_t = 0.5u_{t-1} + v_t$, $v_t \sim i.i.d.N(0, 1)$. We note that the AR model will suffer from the endogenous regressors problem when the error term exits serial correlation. As a result, the OLS and NLS estimators are inconsistent. Hence, we do not report the results with serial correlation error terms for DGP.S2-S3 and P3.

In addition to our test, we also consider Fan et al.'s (2001) *GLR* test. We use the Epanechnikov kernel and Silverman's rule-of-thumb bandwdith $h = 2.35\sigma_z T^{-1/(4+l)}$ for the *GLR* test, where σ_z is the sample standard deviation of state variable. For each DGP, we simulate 1000 datasets with sample sizes $T = 100, 200, 500$. We apply the wild bootstrap procedure for the cases (i) and (ii), and apply the moving block bootstrap procedure for the case (iii). We set the number of bootstrap iterations $B = 300$.

Table B.1 reports results of DGP.S1-S4. The test statistics of this chapter and Fan et al (2001)'s Generalized Likelihood Ratio (GLR) test are based on the critical value generated by the bootstrap method at the 5% and 10% significance levels. From the results in the table, it can be seen that the test statistics $\hat{D}$ have good performance under different sample sizes and regression disturbance settings, and their experience levels are not much different from the nominal significance levels. The general likelihood ratio (GLR) test statistic has a certain degree of rejection for DGP.S2-S4, and its experience level is significantly lower than the nominal significance level. Table B.2 reportS the empirical rejection probability of the two test statistics for different data generation process models at 5% and 10% significance levels. The first four rows give the test results when the linear regression model is estimated for DGP.P1-P4, last three rows correspond to the test results when the threshold regression model is set for DGP.P1-P4. It can be seen from the results of the test power that, compared with the generalized likelihood ratio (GLR) test statistic based on non-parametric regression, there is a higher probability of $\hat{D}$ rejection under each setting, which also confirms the above. The section discusses the pros and cons of the two types of tests on asymptotic power. In Table B.2, when DGP.P3 is mistakenly set as threshold regression, the test statistics $\hat{D}$ and the Generalized Likelihood Ratio (GLR) test of Fan et al (2001) all have lower rejection rates. This is because the setting form of the function coefficients in the smoothly variable autoregressive model is very similar to the threshold model. Nevertheless, the test statistics in this article show better recognition ability, and the probability of rejecting the null hypothesis is much higher than that of the generalized likelihood ratio test, thus giving researchers more accurate conclusions in empirical research. More importantly, the test statistics proposed in this article do not need to perform non-parametric regression for unknown models, and do not need to select adjustment parameters, which provides researchers with great convenience in practical applications.

3.4 Empirical Application

With the continuous deepening of economic globalization, the world economic situation is changing, and the uncertainties facing the economies of various countries are gradually increasing. Existing literature (e.g. Bloom et al. 2007) shows that the existence of economic uncertainty will affect the effectiveness of monetary policy. Generally speaking, when economic uncertainty increases, decision makers face increased risks, which will cause consumers to postpone consumption behavior, investors delay investment plans to maintain cash liquidity, and decision makers will also be more cautious in making decisions. Aastveit et al. (2017) introduced the interaction term between economic uncertainty and monetary policy in the VAR model, and found that when economic uncertainty is high, the effect of monetary policy is weaker. However, the impact of economic uncertainty on the effect of monetary policy is not necessarily linear. Therefore, this section will apply the functional coefficient model and corresponding test statistics proposed in the previous section to examine the impact of economic uncertainty on the effect of monetary policy.

3.4.1 Data

In order to increase the sample size, this chapter is modeled based on monthly macroeconomic data. The sample interval is from January 1992 to December 2019. The data for some indicators in 1992 are not available, so the calculation starts from the earliest available data.

The first is a measure of the effectiveness of monetary policy regulation. This article mainly examines the impact of monetary policy on economic growth and price levels. Although GDP can fully reflect the comprehensive indicators of economic performance, countries only calculate quarterly GDP and cannot directly obtain the measurement results of monthly GDP. Although some scholars use indicators such as gross industrial production and industrial added value to replace GDP indicators and model based on monthly data, in fact, such alternative indicators cannot comprehensively reflect the overall economic situation, especially as China's industrialization process deepens. As the share of tertiary industry in GDP gradually surpasses industry, the error caused by this alternative indicator will become more serious. Therefore, this chapter refers to Zheng and Wang (2013), Zheng and Wang (2020) to construct a mixing dynamic factor model for GDP, industrial value added, investment, consumption, taxation, import and export, etc., based on which the mixing consistency can be extracted.

Index (MFCI), and get the estimated result of monthly GDP (MGDP). These two estimation series can be regarded as measuring indicators of macroeconomics. In addition, this article also selects the macroeconomic consistency index issued by the Economic Prosperity Monitoring Center of the National Bureau of Statistics as the macroeconomic measurement index. For the price level, this article uses the Consumer Price Index (CPI) minus 100 to measure the price level.

The second is the selection of monetary policy variables. This section will examine two monetary policy tools, price-based and quantity-based, respectively. The selected variables are short-term interest rate (SR) and money supply. For short-term interest rates, referring to the research of most domestic scholars such as Xie et al. (2002), Lu and Zhong (2003), Zhao and Gao Hui (2004), Zheng and Liu(2010), this chapter selects the expiry date The 7-day Interbank Offered Rate is used as a proxy variable for the market interest rate. The 7-day lending rate basically reflects the recent market position fluctuations. For the money supply, this article refers to the research of most domestic scholars and chooses the broad money supply M2 year-on-year growth rate.

The third is the selection of economic uncertainty measurement indicators. With reference to Li and Yang(2015), Li et al. (2016), Gu et al. (2018) and Su et al. (2019), this chapter uses the EPU index constructed by Baker et al. (2016) to measure economic uncertainty. It is true that, as stated in some documents, the EPU index mainly measures the uncertainty of economic policies, and is not equivalent to economic uncertainty. However, the trend of the index is relative to China's main economic timing, and it can measure the uncertainty of China's economic fundamentals to a certain extent.

Lastly, the selection of control variables. In order to exclude the influence of fiscal policy and the international environment on the effect of China's monetary policy tools, this paper selects China's government fiscal expenditure (GoV), the U.S. Industrial Production Index (IPUS), the U.S. Federal Benchmark Interest Rate (IRUS), and the U.S. Producer Price Index (PPIUS).) As a control variable, all data are derived from FRED (https://fred.stlouisfed.org/) database. The data frequency of each indicator is monthly. Among them, my country's government fiscal expenditure, the US industrial production index, and the US producer price index are all calculated using logarithmic difference to obtain year-on-year growth rate data.

Table B.3 gives some descriptive statistics of the above indicators. The mean value describes the average value of these indicators, the standard deviation describes the degree of variation of these indicators, and the minimum and maximum values give the fluctuation range of each indicator. Figure B.1 further shows the time series of economic growth indicators, CPI, short-term interest rates, M2 and EPU index.

3.4.2 Model Specification and Results

In order to examine whether the level of economic uncertainty has an impact on the effect of monetary policy control, this chapter considers setting the following functional coefficient model:

$$Y_t = b_0(U_t) + b_1(U_t)X_t + gZ_t + e_t \tag{3.5}$$

where Y_t is the explained variable, which represent output or prices; X_t represents monetary policy variables, U_t represents economic uncertainty, and Z_t represent control variables. In order to avoid the endogeneity of the model, this chapter refers to Aastveit et al. (2017), Su et al. (2019), and selects the lagging one-period variables of each indicator as explanatory variables. Specifically, $Y_t = MGDP_t$, $MFCI_t$, CCI_t or CPI_t, $U_t = EPU_{t-1}$, $X_t = M2_{t-1}$ or SR_{t-1}, $Z_t = (GoV_{t-1}, IPUS_{t-1}, IRUS_{t-1}, PPIUS_{t-1})$. This section tests whether the following three forms of null hypothesis hold:

$$\begin{aligned}
H_0^{(1)} : b_0(U_t) &= b_0, b_1(U_t) = b_1 \\
H_0^{(2)} : b_0(U_t) &= b_{01}1(U_t > c_0) + b_{02}1(U_t > c_0), b_{11}(U_t) = b_{11}1(U_t > c_1) + b_{12}1(U_t > c_1) \\
H_0^{(3)} : b_0(U_t) &= b_{01}G(U_t, g_0, c_0) + b_{02}[1 - G(U_t, g_0, c_0)], \\
b_1(U_t) &= b_{11}G(U_t, g_1, c_1) + b_{12}[1 - G(U_t, g_1, c_1)]
\end{aligned}$$

where $t = 1, 2$, $G(U_t, g, c) = \{1 + exp[-g(U_t - c)]\}^{-1}$ is the logistic function; $1(a)$ is the indicator function. Obviously, the null hypothesis $H_0^{(1)}$ test whether the regression coefficient is a constant, that is, whether the model 3.5 has a linear regression form; the null hypothesis $H_0^{(2)}$ test whether the regression model has a threshold model form; the null hypothesis $H_0^{(3)}$ test whether the regression model has a smooth transition model form.

Similar to the Monte Carlo simulation, for the statistics in this chapter, the weight function $W(u)$ is set as the standard normal distribution density function. For the GLR statistics of Fan et al. (2000) and Fan and Huang (2005), a second-order Gaussian kernel function is used. In order to explore the impact of different bandwidths on the GLR test results, this chapter sets the bandwidth as $h = cT^{-1/5}$, $c = 1/2c_{cv}$ where c_{cv} represents the optimal bandwidth selected by the least squares variational method. Before calculating the statistics, we use the sample mean and sample variance of each data series to standardize it. The critical value of the asymptotic distribution is obtained by the bootstrap method, where the number of bootstraps is $B = 2000$.

Table B.4 reports the test results of whether economic uncertainty affects the effectiveness of monetary policy for three different null hypotheses. As mentioned earlier, this article chooses MGDP, MFCI and CCI to measure output, CPI to measure price, and examines the effectiveness of monetary policy based on the broad money supply M2 and short-term interest rate IR. From the results in the table, the following basic conclusions can be drawn: First, at the 5% significance level, the test statistics proposed in this chapter significantly reject the null hypothesis, indicating that the effect of monetary policy depends on the level of economic uncertainty. For most cases, the GLR test of Fan et al. (2000) can also reject the null hypothesis. However, when short-term interest rates are used to measure monetary policy, it is impossible to reject the assumption that the impact of monetary policy on prices does not depend on uncertainty. According to the Monte Carlo simulation results in this article, this may be caused by the lower power of the GLR test; Then, from the test results, for the three output indicators, price indicators and two monetary policy measurement indicators, the test statistics in this paper are all at the 5%significance level, rejecting the threshold relationship between the effectiveness of monetary policy and policy variables. The null hypothesis. Conversely, GLR statistics can only reject the null hypothesis in a few cases. In particular, in some cases (e.g. $X_t = IR_{t-1}$, $Y_t = MGDP_t$), based on different selection of bandwidth, the GLR test gets different conclusions. The test statistics presented in this chapter do not depend on the selection of bandwidth parameters, so consistent test results are always obtained. Finally, for some cases, our test statistics reject $H_0^{(3)}$ at the 5% significance level, that is, the effectiveness of monetary policy and economic uncertainty have a smooth transition relationship, but the p-value obtained is generally higher than the test result under $H_0^{(1)}$ and $H_0^{(2)}$. This may mean that the smooth transition form described in $H_0^{(3)}$ is closer to the true reaction coefficient than $H_0^{(1)}$ and $H_0^{(2)}$. The GLR test cannot reject $H_0^{(3)}$ in most cases, and the p-value obtained is higher. In addition, we also tries to set the regression coefficient g in formula 3.5 to depend on the form of U_t, that is, $g = g(U_t)$ for all the coefficients, to examine whether the null hypotheses of the form $H_0^{(1)}$, $H_0^{(2)}$ and $H_0^{(3)}$ are established. The test results obtained based on the statistics in this chapter are basically consistent with Table B.4, and will not be repeated here.

3.4.3 Estimation of the Functional Coefficient Model

In order to better characterize the impact of economic uncertainty on the effectiveness of monetary policy, we refer to Fan and Huang (2005), using the Profile Least Square to estimate model 3.5. In this section, we choose $Y_t = MGDP_t$ and CPI_t to examine how the influence of money supply on output and prices changes with economic uncertainty. We standardize economic uncertainty indicators. When estimating the model, a second-order Gaussian kernel function is used and the bandwidth is set as $h = c_{cv}T^{-1/5}$, where c_{cv} is the optimal bandwidth selected by the least squares variational method.

Figure B.2 shows the estimated result of the function coefficient of the influence of broad money supply on output. It can be seen from Figure B.2 that with the increase of economic uncertainty, the effectiveness of monetary policy gradually declines. This result is consistent with the conclusion of Su et al. (2019). In particular, when economic uncertainty is low, the growth rate of broad money supply increases, and output rises accordingly. When the economic uncertainty exceeds 600, the central bank cannot stimulate the economy by increasing the growth rate of the broad money supply, and monetary policy fails. This is also in line with the existing economic theory, that is, when economic uncertainty increases, future risks will intensify, and both investors and consumers will be in a risk-averse mind and make cautious decisions. This greatly reduces the effectiveness of monetary policy. In particular, when economic uncertainty is extremely high, panic and risk-averse psychology dominates, and the central bank's loose monetary policy cannot achieve the effect of stimulating economic recovery.

Figure B.3 depicts the estimated results of the function coefficients of the broad money supply's impact on prices, where the horizontal axis represents the economic uncertainty index. It can be seen from Figure B.3 that the effect of monetary policy on price regulation depends on the economic uncertainty index. When the economic uncertainty is low, the increase in the growth rate of the broad money supply will lead to an increase in prices. However, when the economic uncertainty is high, its effectiveness decreases. Especially when economic uncertainty is very high, the result of $b_1(U - t)$ estimate quickly tends to 0, making monetary policy nearly ineffective. In addition, it should be noted that when the EPU value is greater than 900, the value of $b_1(U-t)$ is less than -0.5, which seems to be different from the actual situation. Note that there are only four samples with an EPU value greater than 900, so the abnormal result may be caused by a large estimation error caused by a small local interval sample size when nonparametric estimation is used.

3.5 Conclusion

Functional coefficient models are useful in capturing nonlinear relationships such as asymmetry among economic and financial variables. Since the general form of functional coefficient model involves nonparametric estimation, which delivers slower convergence rate and lacks definite economic interpretation, many parametric forms of coefficient functions have been developed, such as threshold regression, smooth transition model and so on. However, parametric settings have the risk of misspecification, which will lead to inconsistent estimation and misleading empirical results. Therefore, it is crucial to test whether coefficient functions are in a certain parametric form. In this paper, we propose a unified approach for inference in functional coefficient models via discrete Fourier transform (DFT). Unlike the existing tests that focus on time domain analysis, we investigate the specification of certain functional coefficient models in frequency domain. Our test is widely applicable for various specifications of coefficient functions. More importantly, our test does not involve the smoothed nonparametric estimation. As a result, we can not only avoid the delicate business of choosing a suitable bandwidth, but also detect a class of local alternatives that converges to the null hypothesis at parametric rate, which is faster than the existing nonparametric tests such as Fan et al. (2001). In addition, we allow the regressors and state variables to be the same, which is particular importance in testing certain parametric specifications such as SETAR. Our test is also robust to conditional heteroscedasticity and serial correlation. Monte Carlo simulations show that in comparison with Fan et al.'s (2001) *GLR* test, the proposed test has both reasonable size and excellent power against various misspecification of coefficient functions.

Several interesting extensions are possible. First, 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. Second, it is interesting to consider the model specification problem for coefficient functions with unobserved state variables, such as the Markov regime switching models. Third, one can consider the model specification test for part of coefficient functions, rather than the joint test for all of the coefficient functions in the model. For example, the functional coefficient model $Y_t = \alpha(Z_t) + X_t'\beta(Z_t) + u_t$ under the null hypothesis of $\mathbb{H}_0 : \beta(Z_t) = \beta_0$ degenerates the partially linear model proposed by Robinson (1988) and Stock (1989). All these problems will be pursued in subsequent studies.

CHAPTER 4

HOW DOES SMOOTH STRUCTURAL CHANGE AFFECT ASYMMETRIC DEPENDENCE IN FOREIGN EXCHANGE MARKET?

4.1 The Copula Model

In this section, I introduce the theory of conditional copulas to model the conditional bivariate distribution of the daily Deutsche mark-U.S. dollar and Japanese yen-U.S. dollar exchange rate over the period January 2, 1991, to December 31, 2001. This is the era of post-unification in Germany. The DM-USD and Yen-USD exchange rates have been widely studied for the reason that they are two most heavily traded exchange rates, representing approximately 50 percents of total foreign exchange trading volume. Table C.1 presents some summary statistics of the data.

By Sklar's theorem[1], we are able to decompose a joint distribution into its marginal distributions and its dependence function, or copula:

$$F_{XY}(x, y) = C(F_X(x), F_Y(y)) \tag{4.1}$$

where Equation 4.1 decomposes a bivariate cdf. Under regular conditions, Patton (2006) makes an extension of Sklar's theorem to conditional copulas and shows that for any joint conditional distribution $F_{XY|W}(x, y \mid w)$, there exists a unique conditional copula $C(. \mid w)$ such that:

$$\begin{aligned} &F_{XY|W}(x, y \mid w) = C(F_{X|W}(x \mid w), F_{Y|W}(y \mid w) \mid w), \\ &\forall (x, y) \in \mathbb{R} \times \mathbb{R} \quad \text{and each } w \text{ in its support.} \end{aligned} \tag{4.2}$$

[1]See Sklar (1959). For a thorough review of copulas, see Nelsen (1999) and Joe (1997).

Note that he conditioning variable (or events) must be the same for both marginal distributions and the copula. Specifically, the theorem does not guarantee that $F_{XY|W1,W2}(x, y \mid w1, w2) = C(F_{X|W_1}(x \mid w_1), F_{Y|W_2}(y \mid w_2) \mid w_1, w_2)$. The only exception is when $F_{X|W_1}(x \mid w_1) = F_{X|W_1,W_2}(x \mid w_1, w_2)$ for all $(x, w_1, w_2) \in \mathbb{R}\times\mathbb{W}_1\times\mathbb{W}_2$ and $F_{Y|W_2}(y \mid w_2) = F_{Y|W_1,W_2}(y \mid w_1, w_2)$ for all $(y, w_1, w_2) \in \mathbb{R}\times\mathbb{W}_1\times\mathbb{W}_2$. Fortunately, in this paper, as indicated by Patton (2006), conditional on lags of the DM-USD exchange rate, lags of the Yen-USD exchange rate do not impact the distributional of the DM-USD exchange rate. On the other hand, conditional on lags of the Yen-USD exchange rate, lags of the DM-USD exchange rate do not affect the Yen-USD exchange rate.

A candidate copula should allow for asymmetric dependence structure in either direction and nest symmetric dependence as a special case. The copula used in this paper will be a modification of the "BB7" copula of Joe (1997). [2] Since there is still slight asymmetry in the Joe-Clayton copula even when the two tail dependence measures are equal, we need a symmetritized version of it. In this way, we are able to parameterize the tail dependence by a symmetritized Joe-Clayton copulaconditional univariate extreme events:

$$\begin{aligned} &C_{SJC}(u, v \mid \tau^U, \tau^L) \\ &= 0.5 \cdot (C_{JC}(u, v \mid \tau^U, \tau^L) + C_{JC}(1-u, 1-v \mid \tau^L, \tau^U) + u + v - 1) \end{aligned} \tag{4.3}$$

[2]Patton (2006) also used the normal (or Gaussian) copula. But the essence of this paper is not on the selection of copulas to model tail dependence. Instead, we are arguing the effect of different specification of marginals on joint dependence structure.

where C_{JC} is the Joe-Clayton copula[3]:

$$\begin{aligned} &C_{JC}(u, v \mid \tau^U, \tau^L) = 1 - (1 - \{[1 - (1 - u)^\kappa]^{-\gamma} + [1 - (1 - v)^\kappa]^\gamma - 1\}^{-1/\gamma})^{1/\kappa} \\ &where \quad \kappa = 1/log_2(2 - \tau^U), \quad \gamma = -1/log_2(\tau^L) \\ &and \quad \tau^U \in (0, 1), \quad \tau^L \in (0, 1). \end{aligned} \tag{4.4}$$

By construction, τ^U and τ^L are measure of dependence known as tail dependence under the following definition:

Definition 4.1.1. If the limit

$$\lim_{\varepsilon \to 0} Pr[U \leq \varepsilon \mid V \leq \varepsilon] = \lim_{\varepsilon \to 0} Pr[V \leq \varepsilon \mid U \leq \varepsilon] = \lim_{\varepsilon \to 0} C(\varepsilon, \varepsilon)/\varepsilon = \tau^L$$

exists, then the copula C exhibits lower tail dependence if $\tau^L \in (0, 1]$ and no lower tail dependence if $\tau^L = 0$. Similarly, if the limit

$$\lim_{\delta \to 1} Pr[U > \delta \mid V > \delta] = \lim_{\delta \to 1} Pr[V > \delta \mid U > \delta] = \lim_{\delta \to 1} (1 - 2\delta + C(\delta, \delta))/(1 - \delta) = \tau^U$$

exists, then the copula C exhibits upper tail dependence if $\tau^U \in (0, 1]$ and no upper tail dependence if $\tau^U = 0$.

Note that U and V are derived from the conditional "probability integral transforms" of the original exchange rates conditional on their extreme events respectively. Fisher (1934) and Rosenblatt (1952) showed that U and V have the $Unif(0, 1)$ distribution, regardless of the original distributions. Thus once we specify the marginal distributions of DM-USD and Yen-USD, it is straightforward to transform the original data to the uniformly distributed data, which are easy to implement into the copula model. Patton (2006) models the marginals

[3]A drawback of the Joe-Clayton copula is that even when the two tail dependence parameters are equal, there still exists some slight asymmetry, due to the functional form. This is why we need the symmetrized version to ensure it is symmetric when $\tau^U = \tau^L$.

by a GARCH in Mean process[4][5]:

$$
\begin{aligned}
X_t &= \mu_t + \sqrt{h_t}\varepsilon_t, \\
h_t &= \omega + \beta h_{t-1} + \gamma X_{t-1}^2, \qquad (4.5) \\
\varepsilon_t &\sim i.i.d.N(0,1).
\end{aligned}
$$

The parameter estimates and standard errors for marginal distribution models are presented in Table C.2. In both margin, all parameters changed significantly after the introduction of the euro, implying a structural break in both markets. The test for the structural break at a specified point can was proposed in Patton (2012). It is easy to implement but requires the researcher to have a prior knowledge of when a break in the dependence structure may have occurred. Fortunately, we get a significant structural break on the day of introduction of euro. However, we do not have additional knowledge about other break point in this sample period. Note that smooth structure change cannot be detected from any test for abrupt structural change. On the other hand, we are able to detect abrupt structural change by a smooth change test such as Chen and Hong (2012) and Chen and Hong (2014).

Having the data transformed from the probability integral, we can estimate the tail dependence parameters by maximum likelihood estimation[6]. Indeed, estimating all coefficients simultaneously by a single step QMLE will yield the

[4]Patton (2006) did the model specification test and proposed AR(1)-GARCH(1,1) as marginal distribution for DM-USD exchange rate and AR(1,10)-GARCH(1,1) for Yen-USD. However, in practice, I find that the specification of μ_t has little impact on estimation of GARCH parameter, which is critical for the dependence structure in copula.

[5]QMLE is robust to error term's distribution, but we need it to do "probability integral transforms".Theoretically it has little effect on dependence structure if only the distribution is symmetric. Patton (2006) uses t distribution (the so-called tGarch). Here I use normal distribution for the purpose of staying consistent with estimation in Chen and Hong (2014). Despite we are free from estimating degree of freedom in student-t distribution, the result is very similar to that of tGarch.

[6]Here we do not use the term QMLE because copula model itself describes model's likilihood. There is no need to specify a particular distribution.

most efficient estimates. However, the large number of parameters can result in numerical optimization issue. Under standard conditions, the estimates obtained here are consistent and asymptotically normal. In addition to the constant tail dependence parameter, in order to plot a path to show how the parameters evolve over time, Patton (2006) also proposed a time-varying symmetrized Joe-Clayton copula, which specifies the following evolution equations:

$$\begin{aligned} \tau_t^U &= \Lambda(\omega_U + \beta_U \tau_U^{t-1} + \alpha_U \cdot \frac{1}{10} \sum_{j=1}^{10} \left| u_{t-j} - v_{t-j} \right|) \\ \tau_t^L &= \Lambda(\omega_L + \beta_L \tau_L^{t-1} + \alpha_L \cdot \frac{1}{10} \sum_{j=1}^{10} \left| u_{t-j} - v_{t-j} \right|) \end{aligned} \tag{4.6}$$

where $\Lambda(x) \equiv (1 + e^{-x})^{-1}$ is the logistic transformation, used to keep τ^U and τ^L in $(0, 1)$ at all times.

Figure C.1 shows the degree of asymmetry in the conditional copula by plotting the difference between the upper and lower conditional tail dependence measures implied by both the constant and time-varying SJC copula model. By definition, $\tau^L = \lim_{\epsilon \to 0} C(\epsilon, \epsilon)/\epsilon$ is the limit of quantile dependence when $0 < q \leq 1/2$ and $\tau^U = \lim_{\epsilon \to 1} \frac{1-2q+C(q,q)}{1-q}$ is the limit of quantile dependence when $1/2 < q < 1$[7]. So the tail dependence is actually a measure of the dependence structure between extreme events. In our case, upper (lower) tail dependence measures the dependence between the exchange rates during joint depreciating (appreciating) against the USD. As noted above, it does not matter which of the two currencies one conditions on the dollar having appreciated or depreciated against.

[7]See Patton (2012) for more discussion.

The constant SJC copula results suggest that in the pre-euro period, the limiting probability of the yen appreciating heavily agains the dollar, given that the mark has appreciated heavily against the dollar, is about 86% smaller than the corresponding depreciation probability, meaning that the exchange rates are more dependent during depreciations against the dollar than during appreciations[8]. This implies that export competitiveness preference dominated price stability preference for Bank of Japan and the Bundesbank in the pre-euro period. On the other hand, the time-varying model suggests that the conditional upper tail dependence was greater than conditional lower tail dependence on 98% of days in the pre-euro period[9]. This is also consistent with the finding above. In the post-euro period the asymmetry is reversed. The constant SJC copula suggests that the upper tail dependence is almost zero and the lower tail dependence is 0.012, implying price stability preference dominated export competitiveness preference for the two central banks in the post-euro period. The time-varying model finds similar results: the conditional lower tail dependence was greater than conditional upper tail dependence for the whole period after the introduction of euro.

4.2 The Smooth Time-varying GARCH

As mentioned in the introduction, the specification of marginal distribution plays an important role in modeling joint distribution (copula approach), especially when modeling dependence structure. To illustrate this point, I plot

[8]Despite different from 22% in Patton (2006), it does not change the pattern of asymmetric dependence. When using tGarch as the model of marginal distribution, I found similar results in both constant and time-varing model

[9]Patton (2006) found 92% of days when upper tail dependence is larger than lower tail dependence.

the tail dependence structure using the empirical cumulative distribution function as the marginals. The empirical CDF calculate the cumulative frequencies on each data points and use these frequencies as the transformed data to implement into copula model estimations. However, the empirical CDF approach is not able to capture the feature of volatility clustering in foreign exchange markets. Thus it cannot serve as a correct specification of the marginal distribution. Figure C.1 shows the result.

Despite suffered from model misspecification problem, the empirical CDF indicates that the copula dependence structure is very sensitive to marginal distribution specification: in the pre-euro period, the constant SJC copula shows exactly opposite results as in stationary GARCH marginal specification. The lower tail dependence is about 7% larger than the upper tail dependence, implying that the central banks preferred price stability in this period than export competitiveness. The time-varying model has similar implications: only 31.6% of days in the pre-euro period display larger upper tail dependence, much smaller as compared to 98% in the GARCH specification. If we are using the dependence structure estimated from the SJC copula as an index for hedging, the marginal specification of empirical CDF and stationary GARCH will provide us with totally different guidance in the pre-euro period[10].

Since the copula tail dependence structure relies crucially on the marginal distribution specification, we have to check whether it is appropriate to assume stationarity for the GARCH model and how does different specification affect the dependence pattern.

[10]The dependence structure in the post-euro period is similar to the one using GARCH as marginal distribution: the constant SJC copula indicates the upper tail depence is almost zero and the lower tail dependence is 0.049 while the time-varying SJC copula suggests that the conditional lower tail dependence is greater than conditional upper tail dependence for the whole period after the introduction of euro.

Among different specifications, smooth structural GARCH model is of particular interest for two reasons. First, the change pace of the underlying economic mechanism and technological progress can cause the structural change of exchange rate in a country. Since the pattern of correlation or dependence structure is determined by second or higher order moment, we would expect the marginal structural change in volatility to change the dependence structure in joint distribution. Second, structural change in exchange rates can be viewed as a evolution of underlying economic conditions. By investigating this effect on dependence structure, we want to check how does the domestic conditions affect central banks' preferences.

To study the smooth structural GARCH model, I follow the local QMLE method, proposed by Chen and Hong (2014). Note that we admit the abrupt structural change at the date of introduction of euro but only apply smooth change model to the pre-euro and post-euro period separately. The time-varying GARCH model is specified as follows:

$$\begin{aligned} X_t &= \sqrt{h_t}\varepsilon_t, \\ h_t &= \omega_t + \beta_t h_{t-1} + \gamma_t X_{t-1}^2, \\ \varepsilon_t &\sim i.i.d.N(0,1). \end{aligned} \tag{4.7}$$

Let θ_t be the collection of parameters; namely, $\theta_t = (\omega_t, \beta_t, \gamma_t)'$. In the time series literature, it is common to assume θ_t is an unknown smooth function of time in form of:

$$\theta_t = \theta(\frac{t}{T}), \tag{4.8}$$

where $\theta : [0,1] \to \mathbb{R}^3$ is a vector-valued smooth function. The reason for the specification that $\theta(\cdot)$ is a function of ratio $\frac{t}{T}$ rather than time t is that a nonparametric estimator for θ_t will not be consistent unless the amount of data on which it depends increases. The amount of local information must increase as the sample size T increases[11]. To achieve this we can regard θ as the ordinates of the smooth function $\theta(\cdot)$ on an equally spaced grid over $[0, 1]$. which becomes finer and finer as $T \to \infty$. Consider the following locally stationary GARCH process $\{X_t(u)\}$ that is associated with $\{X_t\}$ at the fixed point $u \in [0, 1]$:

$$
\begin{aligned}
&X_t(u) = \sqrt{h_t(u, \theta_u)}\varepsilon_t, \\
&h_t(u, \theta_u) = \omega_t(u) + \beta_t(u)h_{t-1}(u) + \gamma_t(u)X_{t-1}^2(u), \\
&\varepsilon_t \sim i.i.d.N(0, 1), \quad t = 1, ..., T,
\end{aligned} \tag{4.9}
$$

where all coefficients depend on the fixed point u but do not depend on time t.

To transform the original data into the uniformly distributed which are used to estimate the copula, we need the probability distribution at each data point. That is, we need to assume local stationarity at $u = \frac{t}{T}$ for $t = 1, ..., T$. The boundary-corrected[12] local QMLE estimator to θ_t is given by

$$
\hat{\theta}_t^c = \begin{cases} arg \max\limits_{\theta \in \Theta} \dfrac{1}{T} \sum\limits_{s=1}^{T} k_{st} l_s(\theta) & \text{if} \quad t = \lfloor cTb \rfloor, 0 \le c \le 1 \\ \hat{\theta}_t & \text{if} \quad \lfloor Tb \rfloor \le t \le T - \lfloor Tb \rfloor \\ arg \max\limits_{\theta \in \Theta} \dfrac{1}{T} \sum\limits_{s=1}^{T} k_{st} l_s(\theta) & \text{if} \quad t = T - \lfloor cTb \rfloor, 0 \le c \le 1 \end{cases} \tag{4.10}
$$

where $\hat{\theta}_t = arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{s=1}^{T} k_{st} l_s(\theta)$ is the local QMLE at interior points and $l_s(\theta) = -\frac{1}{2}[log h_s(\theta) + \frac{X_s^2}{h_s(\theta)}]$ is the likelihood function. $k_{st} = \frac{1}{b} k(\frac{s-t}{Tb})$, the kernel $k : [-1, 1] \to \mathbb{R}$ is a pre-specified symmetric bounded probability density, and

[11]See Robinson (1989), Phillips and Hansen (1990), Dahlhaus and Subba Rao (2006), Cai (2007), Chen and Hong (2012) for more discussion.

[12]The boundary corrected estimator uses the so called data reflection method, which is very standard to solve the boundary problem in nonparametric estimation. It is necessary for the following test in a sense that it makes the asymptotic biases to be the same order of magnitude at both interior and boundary points and thus improves power of the test.

$b \equiv b(T)$ is a bandwidth such that $b \to 0$ and $Tb \to \infty$ as $T \to \infty$. By applying the extreme estimator lemma (e.g., Amemiya, 1985, Thm. 4.1.1), we can prove the consistency of the local QMLE $\hat{\theta}_t$.

4.3 Empirical Results

Before applying the estimated probability integral transform, we need to check whether there exists smooth structural change in the two series. Chen and Hong (2014) proposed a model free test[13] to dectect the smooth structural change in GARCH parameters. By comparing the difference between the likelihood of constant GARCH and time-varying GARCH, under regular conditions, they are able to prove the test has a null asymptotic $N(0,1)$ distribution:

$$LR = [2T\sqrt{b}(l_U - l_R) - \hat{A}]/\sqrt{\hat{B}} \xrightarrow{d} N(0,1) \tag{4.11}$$

where l_U and l_R are average log likelihood of the (unrestricted) time-varying parameter GARCH model and the (restricted) constant parameter GARCH model respectively. $\hat{A}$ and $\hat{B}$ are centering and scaling factor. Table C.3 shows the test result.

Based on LR test statistics, all series are time-varying GARCH significant at 0.1 level, except for DM-USD in post euro period. The two series in pre euro period are significant at 0.01 level. This might be due to we have fewer data points in post euro period. When using adjusted LR, all the four series are significant at 0.1 level, justifying our use time-varying parameters GARCH model.

As a by-product of the test above, the estimator $\hat{\theta}_t$ for the time-varying

[13] We are still assuming GARCH model here, but allowing the coefficient to be any function of time.

GARCH parameters can be used to derive the marginal distribution at each time $t = 1, ..., T$, which is applicable to transform the original data into the uniformly distributed one. Figure C.2 shows the result.

In the pre-euro period, the constant SJC copula suggests that the limiting probability of the yen appreciating heavily against dollar, given that the mark has appreciated heavily against the dollar, is about 52.7% smaller than the corresponding depreciation probability, as compared to 86% in previous section. This means that the exchange rates are even more dependent during depreciation than during appreciations. The time-varying model verifies the above conclusion: the upper tail dependence was greater than conditional lower tail dependence in the whole pre-euro period. This result implies that we may underestimate the dependence structure if we do not take structural change in marginals into account. This can be possibly explained by endogeneity of monetary policy. When the two central banks intervene to ensure a matching depreciation of the opponent currencies against the dollar, they are implementing expansionary policy and depreciate even more, causing a heavier dependence in joint depreciation. On the other hand, this endogenous effect is smaller in joint appreciation. In this way, ignoring the structural change in marginal distributions may underestimate the asymmetric dependence structure toward upper tail.

However, things are more complicated in the post-euro period. The most striking feature is the inconsistency between constant tail dependence and time-varying tail dependence. The constant SJC copula still suggests that the lower tail dependence was heavier than upper tail dependence, while the time-varying version indicates that the conditional lower tail dependence was greater than conditional upper tail dependence for only 61.3% in the post-euro period, as compared to 100% in previous section. First notice that the constant tail dependence is not the average of the time-varying tail dependence. Instead, it can be viewed as a test of asymmetric dependence similar to Hong, Tu and Zhou (2007). Although they did not allow for nonstationary marginal distribution, the results here implies that the model free test proposed by Hong, Tu and Zhou (2007) should be robust to nonstationary series. On the other hand, the time-varying model indicates that if we are using asymmetric dependence as a factor to form portfolio, simply testing the average level over a certain period may cause troubles, since we are using the opposite strategy on almost 39% days! Also, similar to the pre-euro period, ignoring the smooth structural change in marginals may cause underestimation of the upper tail dependence.

4.4 Conclusion

In this chapter we investigate how does smooth structural change in marginal distribution of mark-dollar and yen-dollar exchange rates affect their joint dependence structure. Abrupt structural change in marginal distribution has been discussed in literature. However, it is more plausible that the structural change takes a period of time to take effect. The specification of time-varying parameters GARCH models is not of no economic implication. By allowing its parameters to change smoothly, we are actually trying to answer how does the change of underlying economic conditions affect their joint asymmetric dependence structure. And the asymmetric dependence behavior can be interpreted as, explained by Patton (2006), the central banks' tradeoff between price stability and export competitiveness.

Based on Patton (2006), we use both the constant and time-varying SJC copula to model the conditional tail dependence of the Deutsche mark- U.S dollar and yen-dollar exchange rates, over the period from January 1991 to December 2001. Justified by a test for smooth structural changes in GARCH models proposed by Chen and Hong (2014), we use the local QMLE method to estimate the marginal distribution for the two series separately over pre-euro period and post-euro period. By doing this, we admit the abrupt structural change on the day of introduction of Euro on January 1, 1999. But different from Patton (2006), we allow for smooth structural changes in the marginal distribution.

As a result, we found ignoring the time-varying structure of marginal distributions may cause under estimation of upper tail dependence in both period. This can be possibly explained by smaller endogenous effect of expansionary monetary policy during joint appreciation than that of contractionary policy during joint depreciation. More importantly, we found that the constant conditional tail dependence did not change much by allowing for smooth structural change in marginal distributions. This indicates the asymmetric test proposed by Hong, Tu and Zhou (2007) is robust to nonstationary series. However, the departure of the time-varying tail dependence from the constant version implies that simply testing the average level of asymmetric dependence over a certain period may lead to misleading result in investment decision.

APPENDIX A

APPENDIX OF CHAPTER 2

A.1 Figures and Tables

Table A.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 Li et al. (2002); (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 A.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 Li et al. (2002); (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 A.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 Li et al. (2002); (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 A.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 Li et al. (2002); (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 A.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 Li et al. (2002); (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 A.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 Li et al. (2002); (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 A.7: Bootstrap p-values for tests on conditional alpha (CAPM)

Zt	V			G			V-G		
	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
panel 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
panel 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 Li et al. (2002); (ii) 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 A.8: Bootstrap p-values for tests on conditional beta (CAPM)

Zt	V			G			V-G		
	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
panel 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
Panel 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 Li et al. (2002); (ii) 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 A.9: Bootstrap p-values for tests on conditional alpha (FF6)

	V			G			V-G		
Zt	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
panel A: value-weighted portfolios									
cay	0.000	0.005	0.005	0.000	0.000	0.005	0.010	0.010	0.005
yc	0.000	0.005	0.005	0.000	0.005	0.005	0.000	0.000	0.005
def	0.000	0.005	0.005	0.005	0.005	0.005	0.010	0.015	0.020
cay, yc	0.020	0.315	0.260	0.045	0.575	0.595	0.040	0.335	0.315
yc, def	0.015	0.520	0.395	0.040	0.460	0.535	0.055	0.330	0.515
def, cay	0.035	0.610	0.485	0.075	0.625	0.730	0.025	0.390	0.365
cay, yc, def	0.040	0.810	0.925	0.025	0.705	0.830	0.065	0.715	0.890
panel 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	0.720	0.037	0.410	0.365
yc, def	0.035	0.515	0.575	0.040	0.420	0.230	0.035	0.570	0.485
def, cay	0.020	0.730	0.455	0.060	0.285	0.730	0.035	0.595	0.590
cay, yc, def	0.055	0.920	0.785	0.035	0.675	0.675	0.045	0.595	0.580

Note: (i) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang(2001); Wald, the Wald-type test in Li et al. (2002); (ii) 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 A.10: Bootstrap p-values for tests on conditional beta (FF6)

	V			G			V-G		
Zt	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
				panel A: value-weighted portfolios					
cay	0.000	0.005	0.005	0.000	0.000	0.005	0.010	0.010	0.005
yc	0.000	0.005	0.005	0.000	0.005	0.005	0.000	0.000	0.005
def	0.000	0.010	0.005	0.005	0.005	0.005	0.020	0.015	0.020
cay, yc	0.025	0.315	0.260	0.040	0.555	0.565	0.040	0.335	0.315
yc, def	0.015	0.520	0.395	0.040	0.460	0.535	0.055	0.330	0.515
def, cay	0.035	0.610	0.485	0.075	0.665	0.720	0.035	0.390	0.365
cay, yc, def	0.040	0.810	0.925	0.025	0.705	0.820	0.065	0.705	0.880
				panel 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	0.720	0.037	0.410	0.365
yc, def	0.035	0.535	0.575	0.040	0.420	0.230	0.035	0.570	0.485
def, cay	0.025	0.720	0.455	0.060	0.285	0.730	0.035	0.595	0.590
cay, yc, def	0.045	0.940	0.765	0.025	0.665	0.665	0.065	0.555	0.590

Note: (i) GLR, the generalized likelihood ratio test in Fan, Zhang, and Zhang (2001); Wald, the Wald-type test in Li et al. (2002); (ii) 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 A.11: Bootstrap p-values for tests on conditional alpha (Q Factor)

	V			G			V-G		
Zt	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
panel A: value-weighted portfolios									
cay	0.000	0.005	0.005	0.000	0.010	0.000	0.005	0.005	0.005
yc	0.005	0.005	0.005	0.000	0.000	0.000	0.000	0.000	0.000
def	0.005	0.010	0.005	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
panel 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 Li et al. (2002) (2012); (ii) 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 A.12: Bootstrap p-values for tests on conditional beta (Q Factor)

Zt	V			G			V-G		
	Q	GLR	Wald	Q	GLR	Wald	Q	GLR	Wald
panel A: value-weighted portfolios									
cay	0.000	0.005	0.005	0.000	0.010	0.000	0.005	0.005	0.000
yc	0.005	0.005	0.005	0.000	0.000	0.005	0.005	0.000	0.000
def	0.005	0.010	0.005	0.005	0.005	0.005	0.010	0.015	0.020
cay, yc	0.020	0.360	0.290	0.045	0.605	0.615	0.035	0.325	0.195
yc, def	0.015	0.520	0.375	0.045	0.460	0.515	0.025	0.310	0.505
def, cay	0.035	0.600	0.495	0.070	0.625	0.730	0.025	0.390	0.455
cay, yc, def	0.040	0.830	0.925	0.025	0.695	0.820	0.060	0.765	0.880
panel 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 Li et al. (2002); (ii) 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.

A.2 Mathematical Proofs

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 2.2.1. 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 2.16. Notice that

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

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 equation (2.2.2). □

Proof of Propostion 2.4.1.

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

Firstly, we show A_1 weakly converges to a gaussian process by the following lemma:

Lemma A.2.1. *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

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

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.

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

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

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 $z \in \mathbb{R}^L$,

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

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

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

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 $u \in \mathbb{R}^L$,

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

where

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

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 A_{11} below. To complete the proof of Proposition 2.4.1, we still need to show

(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 $T \to \infty$.

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

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

then

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

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 5, 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

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

We first show $A_{21} = o_p(1)$.

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

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

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

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

Then we show $A_{22} = \tilde{U} + o_p(1)$.

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

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

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

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

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

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

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

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

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

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.

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

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

$$\begin{aligned}
E\left[\|\Psi(\xi_s, \xi_t)\|^2\right] &= \int \frac{1}{h^{2d}}\left\|\prod_{i=1}^{d} K\left(\frac{Z_{si}-Z_{ti}}{h}\right)\right\|^2 \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}X_s\varepsilon_s\psi_t(u) + \frac{\Omega(Z_s)^{-1}}{f_Z(Z_s)}X_t\varepsilon_t\psi_s(u)\right]^2 \\
&\quad \times f_Z(Z_t)f_Z(Z_s)dZ_t dZ_s \\
&= \int \frac{1}{h^d}\left\|\prod_{i=1}^{d} K(v_i)\right\|^2 \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}X_s\varepsilon_s\psi_t(u) + \frac{\Omega(Z_t+hv)^{-1}}{f_Z(Z_t+hv)}X_t\varepsilon_t\left(e^{iu'(Z_t+hv)}-\phi(u)\right)\right]^2 \\
&\quad \times f_Z(Z_t)f_Z(Z_t+hv)dZ_t dv \\
&= O\left(h^{-d}\right) = O\left(T\frac{1}{Th^d}\right)
\end{aligned}$$

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

$$\begin{aligned}
\Psi(\xi_s) &= \int \frac{1}{h^d}\prod_{i=1}^{d} K\left(\frac{Z_{si}-z_i}{h}\right)\Omega(z)^{-1}X_s\varepsilon_s\left(e^{iu'z}-\phi(u)\right)dz \\
&= \int \prod_{i=1}^{d} K(w_i)\,\Omega(Z_s+hw)^{-1}\psi_{(u)}(Z_s+hw)dwX_s\varepsilon_s \\
&= \Omega(Z_s)^{-1}\psi_s(u)X_s\varepsilon_s + o_p(1)
\end{aligned}$$

Then we have

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

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

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

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

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

and

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

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

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

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

We now turn to verify (A.5), for which it suffices to show that

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

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

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

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,

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

By Cauchy–Schwarz inequality, it is suffices to show

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

where

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

Then

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

by Assumption 2.4.2. The second line follows from Minkowski inequality, and the third line follows triangular inequality.

Let $G_{1t}(u) = c_0\Omega(Z_t)^{-1}\psi_t(u)X_t$, then $A_{22}^{(0)} = \frac{1}{\sqrt{T}}G_{1t}(u)\varepsilon_t + o_p(1)$ uniformly over $u \in \mathbb{R}^L$, we can also have

$$A_1 \Rightarrow \mathcal{G}(u)$$

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

$$\mathcal{K}(u,v) = \lim_{T\to\infty} \frac{2}{T(T-1)} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} E[G_{1t}(u)G_{1s}(v)^* \varepsilon_t \varepsilon_s]$$

□

It remains to show that, under the conditions of Proposition 2.4.1, $A_2 = o_p(1)$, $A_3 = o_p(1)$ and $A_4 = o_p(1)$ uniformly over $u \in \mathbb{R}^L$. Since $\psi_t(u)$ and $\phi(u) - \widehat{\phi}(u)$ converges to 0 pointwisely by law of large numbers, along with the argument establishing stochastic equicontinuous in the proof of Lemma A.2.1, it is straightforward to show these three terms are uniformly negligible. This completes the proof of Proposition 2.4.1.

□

Proof of Theorem 2.4.1. Under $\mathbb{H}_{01}$, Proposition 2.4.1 shows that $\sqrt{T}\hat{A}_1(u)$ weakly converges to a complexed valued Gaussian process $\mathcal{G}_1(u)$. Under Assumption 2.4.4 and continuous mapping theorem, we have

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

as $T \to \infty$. □

Proof of Theorem 2.4.3. Under $\mathbb{H}_{A1}$, it is straightforward to see that uniformly over $u \in \mathbb{R}^L$ and $z \in \mathbb{R}^L$,

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

Then by the proof of Proposition 2.4.1, uniformly over $u \in \mathbb{R}^L$,

$$A_1 = \frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t) \right] \psi_t(u) + \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{h^2}{2} e_p H^{-1} S^{-1}(Z_t) \begin{pmatrix} S_2(Z_t) \\ S_3(Z_t) \end{pmatrix} \beta''(Z_t) + o_p(1)$$

where

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

Notice that the first term $\frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t) \right] \psi_t(u) = O_p(1/\sqrt{T}) = o_p(1)$ by the proof of Proposition 2.4.1. By Cai et. al (2000), the second term $\frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{h^2}{2} e_p H^{-1} S^{-1}(Z_t) \begin{pmatrix} S_2(Z_t) \\ S_3(Z_t) \end{pmatrix} \beta''(Z_t) = O_p(1)$ pointwisely. It is also stochastic equicontinuous using the same argument in the proof of Proposition 2.4.1. The remained terms A_2, A_3 and A_4 are still negligible uniformly over $u \in \mathbb{R}^L$. Therefore,

$$\begin{aligned} \widehat{Q}_1 &= T \int_{\mathbb{R}^L} \|\widehat{A}_1(u)\|^2 W(u) du \\ &= O_p(T). \end{aligned}$$

□

Proof of Theorem 2.4.5. Under $\mathbb{H}_{Al} : \mathbf{R}\beta(Z_t) = \mathbf{R}\beta(Z, \theta_0) + \alpha(T)\delta(Z_t)$, set $\beta(Z, \theta_0) = \beta_0$. We have $\beta(Z_t) = \beta_0 + \alpha(T)\delta(Z_t)$. Notice that uniformly over $u \in \mathbb{R}^L$ and $z \in \mathbb{R}^L$,

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

where

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

Then by the proof of Proposition 2.4.1, uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned} A_1 &= \frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\psi_t(u) + \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{h^2}{2}e_pH^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\beta''(Z_t)\psi_t(u) + o_p(1) \\ &= \frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\psi_t(u) + \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{h^2}{2}e_pH^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\alpha(T)\delta''(Z_t)\psi_t(u) + o_p(1) \\ &= A_{13} + A_{14},\ say. \end{aligned}$$

From the proof of Proposition 2.4.1, the first term $A_{13} \Rightarrow \mathcal{G}_1(u)$. For the second term A_{14}, uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned} \sqrt{T}A_{14} &= \frac{1}{T}\sum_{t=1}^{T}\frac{h^2}{2}e_pH^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\alpha(T)\delta''(Z_t)\psi_t(u) \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{h^2}{2}e_pH^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\alpha(T)\delta''(Z_t)\psi_t(u) \\ &= \frac{1}{T}\sum_{t=1}^{T}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}\delta''(Z_t)\psi_t(u) + o_p(1) \end{aligned}$$

where $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$. Therefore, it is straightforward that

$$\sqrt{T}A_{14} \xrightarrow{p} \frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2}E[\delta''(Z_t)\psi_t(u)] = \xi(u)$$

uniformly over $u \in \mathbb{R}^L$. The remained terms A_2, A_3 and A_4 are still negligible uniformly over $u \in \mathbb{R}^L$. Therefore, under $\mathbb{H}_{Al}$,

$$\sqrt{T}\widehat{A_1}(u) \Rightarrow \xi(u) + \mathcal{G}_1(u),$$

for all $u \neq 0$. By the continuous mapping theorem,

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

□

Proof of Proposition 2.4.2.

$$\begin{aligned}
\sqrt{T}\widehat{A_2}(u) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{\beta}(Z_t)-\beta(Z_t,\hat{\theta})]e^{iu'Z_t} \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\hat{\beta}(Z_t)-\beta(Z_t))-(\beta(Z_t,\hat{\theta})-\beta(Z_t))]e^{iu'Z_t} \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{\beta}(Z_t)-\beta(Z_t)]e^{iu'Z_t}-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\beta(Z_t,\hat{\theta})-\beta(Z_t)]e^{iu'Z_t} \\
&= A_{21}+A_{22}.
\end{aligned}$$

For the first term A_{21}, from the proof of Proposition 2.4.1, it is easy to see that uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned}
A_{21} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}G_{2t}(u)\varepsilon_t+o_p(1) \\
&\Rightarrow \mathcal{G}_2(u)
\end{aligned}$$

where $\mathcal{G}_2(u)$ is a complex-valued Gaussian process with covariance kernel

$$\mathcal{K}(u,v)=\lim_{T\to\infty}\frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}E[G_{2t}(u)G_{2s}(v)^*\varepsilon_t\varepsilon_s],$$

and $G_{2t}(u)=c_0\Omega(Z_t)^{-1}\phi_t(u)X_t$.

For the second term A_{22}, notice that

$$\begin{aligned}
A_{22} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\beta(Z_t,\hat{\theta})-\beta(Z_t)]e^{iu'Z_t} \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[(\beta(Z_t,\hat{\theta})-\beta(Z_t,\theta_0))-(\beta(Z_t)-\beta(Z_t,\theta_0))]e^{iu'Z_t} \\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\beta(Z_t,\hat{\theta})-\beta(Z_t,\theta_0)]e^{iu'Z_t}-\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\beta(Z_t)-\beta(Z_t,\theta_0)]e^{iu'Z_t} \\
&= A_{22}^{(1)}+A_{22}^{(2)}.
\end{aligned}$$

The second term $A_{22}^{(2)}$ is equal to 0 under $\mathbb{H}_{02}$. For the first term $A_{22}^{(1)}$, by Taylor expansion

$$\beta(Z_t;\hat{\theta}) = \beta(Z_t;\theta_0) + \beta^{(1)}(Z_t;\bar{\theta})(\hat{\theta}-\theta_0),$$

where $\bar{\theta} = \alpha\theta_0 + (1-\alpha)\hat{\theta}$ for some $0 < \alpha < 1$. Furthermore, by the FOC of NLS, we have

$$\begin{aligned} 0 &= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\left[Y_t - X_t'\beta(Z_t;\hat{\theta})\right] \\ &= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\hat{\theta})\right] \\ &= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_0)\right] \\ &\quad - \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t;\hat{\theta}) - \beta(Z_t;\theta_0)\right] \\ &= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_0)\right] \\ &\quad - \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})(\hat{\theta}-\theta_0). \end{aligned}$$

Thus, it follows

$$\begin{aligned} \hat{\theta}-\theta_0 &= \left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1} \\ &\times \left\{\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_0)\right]\right\}. \end{aligned}$$

Therefore,

$$\begin{aligned} A_{22}^{(1)} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\beta(Z_t) - \beta(Z_t,\theta_0)]e^{iu'Z_t} \\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}(\hat{\theta}-\theta_0) \\ &= \frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t \\ &+ \frac{1}{\sqrt{T}}\left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_0)\right] \end{aligned}$$

The second term above is equal to 0 under $\mathbb{H}_{02}$. Let

$$\hat{G}_{3t}(u) = \left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\beta^{(1)}(Z_t;\hat{\theta})'X_t,$$

then under $\mathbb{H}_{02}$,

$$A_{22}^{(1)} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{G}_{3t}(u)\varepsilon_t$$

To complete the proof, we need to show

(i) $\hat{G}_{3t}(u) - G_{3t}(u)$ is stochastically equicontinuous uniformly over $u \in \mathbb{R}^L$;

(ii) $\sup_{u\in\mathbb{R}^L} \|\hat{G}_{3t}(u) - G_{3t}(u)\| \xrightarrow{p} 0$, where

$$G_{3t}(u) = E\left[\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\beta^{(1)}(Z_t;\hat{\theta})'X_t$$

For (i), the property of stochastically equicontinuous is established from the proof of 2.4.1. For (ii), we can use analogous arguments in the proof ??ch3. Therefore, under Assumptions 2.4.1-2.4.4, by FCLT

$$A_{22}^{(1)} \Rightarrow \mathcal{G}_{23}(u),$$

where $\mathcal{G}_{23}(u)$ is a complex-valued Gaussian process, with mean 0 and covariance kernel

$$\mathcal{K}_{23}(u,v) = \lim_{T\to\infty}\frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}E\left[G_{3t}(u)G_{3s}(u)^*\varepsilon_t\varepsilon_s\right]$$

The remained is to compute the covariance term in the variance-covariance kernel in $\mathcal{G}_2(u)$. By the mixing condition in Assumption 2.4.1 and the condition on the error terms in Assumption 2.4.2, it is straightforward to see that

$$cov(\mathcal{G}_2(u),\mathcal{G}_2^*(v)) = E\left[c_0\Omega(Z_t)^{-1}\phi(u,Z_t)X_tG_t^*(v)'v(X_t,Z_t)\right]$$

This completes the proof. □

Proof of Theorem 2.4.2. Under $\mathbb{H}_{02}$, Proposition 2.4.2 shows that $\sqrt{T}\hat{A}_2(u)$ weakly converges to a complexed valued Gaussian process $\mathcal{G}_2(u)$. Under Assumption

2.4.4 and continuous mapping theorem, we have

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

as $T \to \infty$. □

Proof of Theorem 2.4.4. Under $\mathbb{H}_{A2}$, it is straightforward to see that uniformly over $u \in \mathbb{R}^L$ and $z \in \mathbb{R}^L$,

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

Then by the proof of Proposition 2.4.2, uniformly over $u \in \mathbb{R}^L$,

$$A_{21} = \frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t) \right] \phi_t(u) + \frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{h^2}{2} e_p H^{-1} S^{-1}(Z_t) \begin{pmatrix} S_2(Z_t) \\ S_3(Z_t) \end{pmatrix} \beta''(Z_t) \phi_t(u) + o_p(1)$$

where

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

Notice that the first term $\frac{2}{T^{1/2}} \sum_{t=1}^{T} \left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)} Q_T(Z_t) \right] \phi_t(u) = O_p(1/\sqrt{T}) = o_p(1)$ by the proof of Proposition 2.4.1. By Cai et. al (2000), the second term $\frac{1}{T^{1/2}} \sum_{t=1}^{T} \frac{h^2}{2} e_p H^{-1} S^{-1}(Z_t) \binom{S_2(Z_t)}{S_3(Z_t)} \beta''(Z_t) \phi_t(u) = O_p(1)$ pointwisely. It is also stochastic equicontinuous using the same argument in the proof of Proposition 2.4.2. The remained terms are still negligible uniformly over $u \in \mathbb{R}^L$. Therefore,

$$\begin{aligned} \widehat{Q}_2 &= T \int_{\mathbb{R}^L} \|\widehat{A}_2(u)\|^2 W(u) du \\ &= O_p(T). \end{aligned}$$

□

Proof of Theorem 2.4.6. Under $\mathbb{H}_{Al} : \mathbf{R}\beta(Z_t) = \mathbf{R}\beta(Z, \theta_0) + \alpha(T)\delta(Z_t)$, we have $\beta(Z_t) = \beta(Z_t, \theta_0) + \alpha(T)\delta(Z_t)$. Notice that uniformly over $u \in \mathbb{R}^L$ and $z \in \mathbb{R}^L$,

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

where

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

Then by the proof of Proposition 2.4.2, uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned}
A_{21} &= \frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\phi_t(u) + \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{h^2}{2}e_p H^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\beta''(Z_t)\phi_t(u) + o_p(1)\\
&= \frac{2}{T^{1/2}}\sum_{t=1}^{T}\left[\frac{\Omega(Z_t)^{-1}}{f_Z(Z_t)}Q_T(Z_t)\right]\phi_t(u) + \frac{1}{T^{1/2}}\sum_{t=1}^{T}\frac{h^2}{2}e_p H^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\alpha(T)\delta''(Z_t)\phi_t(u) + o_p(1)\\
&= A_{23} + A_{24}, \textit{say}.
\end{aligned}$$

From the proof of Proposition 2.4.2, the first term $A_{23} \Rightarrow \mathcal{G}_2(u)$. For the second term A_{24}, uniformly over $u \in \mathbb{R}^L$,

$$\begin{aligned}
\sqrt{T}A_{24} &= \frac{1}{T}\sum_{t=1}^{T}\frac{h^2}{2}e_p H^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\alpha(T)\delta''(Z_t)\phi_t(u)\\
&= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{h^2}{2}e_p H^{-1}S^{-1}(Z_t)\begin{pmatrix}S_2(Z_t)\\ S_3(Z_t)\end{pmatrix}\alpha(T)\delta''(Z_t)\phi_t(u)\\
&= \frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}\delta''(Z_t)\phi_t(u) + o_p(1)
\end{aligned}$$

where $\mu_j = \int_{-\infty}^{\infty} u^j K(u)du$. Therefore, it is straightforward that

$$\sqrt{T}A_{24} \xrightarrow{p} \frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}E[\delta''(Z_t)\phi_t(u)] = \xi_2(u)$$

uniformly over $u \in \mathbb{R}^L$. The remained terms are still negligible uniformly over $u \in \mathbb{R}^L$.

For A_{22}, notice that

$$\begin{aligned}
A_{22} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{G}_{3t}(u)X_t'\delta(Z_t) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{G}_{3t}(u)\varepsilon_t\\
&\Rightarrow \xi_3(u) + \mathcal{G}_{23}(u),
\end{aligned}$$

where $\xi_3(u)$ is a complex-valued Gaussian process, with non-zero mean and covariance kernel

$$\mathcal{K}(u,v) = \lim_{T\to\infty}\frac{2}{T(T-1)}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}E\left[\hat{G}_{3t}(u)X_t'X_s\delta(Z_t)\delta(Z_s)\hat{G}_{3s}(u)^*\right]$$

Therefore, under $\mathbb{H}_{Al}$,

$$\sqrt{T}\widehat{A}_2(u) \Rightarrow \xi_2(u) + \xi_3(u) + \mathcal{G}_2(u),$$

for all $u \neq 0$. By the continuous mapping theorem,

$$\widehat{Q}_2 \xrightarrow{d} \int \|\xi_2(u) + \xi_3(u) + \mathcal{G}_2(u)\|^2 W(u) du.$$

□

APPENDIX B

APPENDIX OF CHAPTER 3

B.1 Figures and Tables

Figure B.1: Time series of descriptive statistics

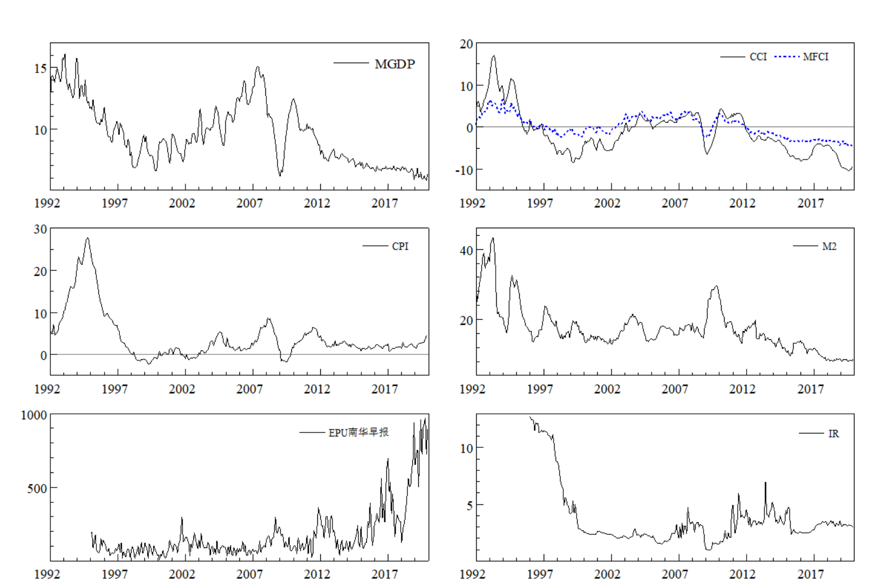


Table B.1: Size of tests under DGP.S1-S4

		i.i.d. error term				ARCH errors				serial correlation errors			
		D		*GLR*		*D*		*GLR*		*D*		*GLR*	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
S1	$T = 100$	5.8	11.0	4.1	8.4	5.2	10.1	3.8	8.0	5.7	10.8	4.6	9.6
	$T = 200$	4.7	9.6	5.0	8.7	5.8	12.2	4.4	9.0	3.8	9.6	4.8	9.4
	$T = 500$	5.2	10.3	4.4	9.6	4.1	8.5	4.5	9.3	5.2	10.5	5.9	10.9
S2	$T = 100$	4.2	9.3	3.6	7.2	5.3	12.0	3.5	7.5	–	–	–	–
	$T = 200$	5.0	11.0	4.0	8.4	6.4	12.5	1.9	6.5	–	–	–	–
	$T = 500$	4.8	10.5	4.7	9.6	3.7	9.7	1.4	4.8	–	–	–	–
S3	$T = 100$	2.7	6.6	0.7	2.2	5.8	11.4	1.0	3.5	–	–	–	–
	$T = 200$	3.7	9.1	1.3	3.3	6.5	11.6	2.1	4.8	–	–	–	–
	$T = 500$	5.1	10.0	1.2	2.9	6.1	11.2	2.1	4.4	–	–	–	–
S4	$T = 100$	2.8	6.2	0.6	1.5	5.1	9.4	1.0	2.7	3.8	8.4	0.9	2.2
	$T = 200$	4.6	9.8	0.4	1.4	5.0	10.2	1.7	3.7	5.2	10.1	1.0	2.9
	$T = 500$	5.4	9.7	0.9	2.8	4.7	9.2	1.8	4.7	5.6	9.8	2.0	3.9

Table B.2: Power of tests under DGP.P1-P4

		i.i.d. error term				ARCH errors				serial correlation errors			
		D		*GLR*		*D*		*GLR*		*D*		*GLR*	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
P1	$T = 100$	98.8	99.5	98.8	99.4	96.9	98.5	72.3	84.6	100	100	100	100
	$T = 200$	100	100	100	100	99.5	99.7	93.4	96.4	100	100	100	100
	$T = 500$	100	100	100	100	100	100	100	100	100	100	100	100
P2	$T = 100$	51.9	64.6	18.0	30.2	41.2	53.4	13.8	22.5	57.5	69.7	32.5	45.1
	$T = 200$	85.8	91.6	37.9	54.4	68.5	80.8	27.5	40.1	87.1	91.7	60.9	71.0
	$T = 500$	100	100	86.5	93.0	99.0	99.0	73.5	83.5	100	100	94.5	97.0
P3	$T = 100$	72.1	81.6	31.8	47.6	30.5	43.8	8.5	17.0	–	–	–	–
	$T = 200$	95.2	97.6	65.3	78.2	50.2	63.3	12.7	22.8	–	–	–	–
	$T = 500$	100	100	99.9	99.9	89.5	92.5	29.5	46.0	–	–	–	–
P4	$T = 100$	71.7	82.6	35.1	47.2	98.7	99.6	86.2	92.3	75.3	83.9	46.3	58.4
	$T = 200$	95.5	97.6	66.1	79.0	100	100	98.7	99.4	96.4	99.0	78.7	86.1
	$T = 500$	99.9	100	99.0	99.5	100	100	100	100	100	100	100	100
P5	$T = 100$	79.4	88.5	48.9	64.6	39.2	55.5	8.2	15.5	90.6	95.2	76.0	83.7
	$T = 200$	96.8	98.2	81.6	89.4	72.1	82.5	23.0	35.2	99.8	99.9	98.1	98.8
	$T = 500$	100	100	100	100	97.5	98.8	70.0	80.5	100	100	100	100
P6	$T = 100$	24.8	40.6	4.7	10.1	32.3	47.7	8.2	15.4	26.7	39.3	7.4	13.4
	$T = 200$	49.4	62.3	14.0	24.7	64.9	78.1	27.1	40.1	52.4	65.1	21.0	32.3
	$T = 500$	88.2	94.2	51.1	65.8	96.4	98.0	72.9	81.9	89.6	93.4	61.6	72.6
P7	$T = 100$	6.2	12.0	0.6	1.5	5.5	12.1	0.9	2.8	–	–	–	–
	$T = 200$	13.4	21.5	1.3	3.6	13.2	22.0	2.4	5.9	–	–	–	–
	$T = 500$	33.1	46.9	8.0	14.3	26.0	36.2	3.6	9.0	–	–	–	–

Table B.3: Descriptive statistics

	Sample interval	Mean	Std	Min	Max
MGDP	1992M1-2019M12	9.540	2.518	5.800	16.111
MFCI	1992M1-2019M12	0.162	2.576	-4.379	6.597
CCI	1992M1-2019M11	-1.088	5.219	-10.420	17.030
CPI	1992M1-2019M12	4.132	5.820	-2.200	27.700
SR	1996M1-2019M12	3.690	2.599	0.990	12.720
M2	1992M1-2019M12	17.379	6.530	7.971	43.382
EPU	1995M1-2019M12	173.771	180.427	9.067	970.830
GoV	1992M1-2019M12	15.198	12.644	-22.141	81.580
IPUS	1992M1-2019M12	1.960	4.041	-16.646	8.198
IRUS	1992M1-2019M12	2.620	2.159	0.070	6.540
PPIUS	1992M1-2019M12	1.926	4.866	-17.505	16.009

Table B.4: Test of the Influence of Economic Uncertainty on the Effectiveness of Monetary Policy

$H_0^{(1)}$: $b_0(U_{t-1}) = b_0$, $b_1(U_{t-1}) = b_1$								
	$X_t = M2_{t-1}$				$X_t = IR_{t-1}$			
	$\hat{D}$	$GLR(h_0)$	$GLR(h_1)$	$GLR(h_{opt})$	$\hat{D}$	$GLR(h_0)$	$GLR(h_1)$	$GLR(h_{opt})$
$Y_t = MGDP_t$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$Y_t = MFCI_t$	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000
$Y_t = CCI_t$	0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.000
$Y_t = CPI_t$	0.010	0.008	0.004	0.017	0.042	0.272	0.097	0.851
$H_0^{(2)}$: $b_0(U_{t-1})$, $b_1(U_{t-1})$ follow threshold model								
$Y_t = MGDP_t$	0.004	0.048	0.055	0.080	0.008	0.640	0.968	0.002
$Y_t = MFCI_t$	0.003	0.023	0.021	0.020	0.000	0.334	0.552	0.472
$Y_t = CCI_t$	0.003	0.697	0.119	0.074	0.001	0.000	0.004	0.000
$Y_t = CPI_t$	0.012	0.132	0.156	0.135	0.008	0.178	0.263	0.191
$H_0^{(3)}$: $b_0(U_{t-1})$, $b_1(U_{t-1})$ follow smooth transition model								
$Y_t = MGDP_t$	0.058	0.974	1.000	0.998	0.050	0.053	0.167	0.025
$Y_t = MFCI_t$	0.019	0.982	1.000	1.000	0.158	0.524	0.564	0.471
$Y_t = CCI_t$	0.006	0.698	0.119	0.075	0.068	0.254	1.000	0.207
$Y_t = CPI_t$	0.045	0.337	0.298	0.347	0.004	0.289	0.417	0.326

Note: The p-value for each test statistic is obtained based on 2000 bootstraps.

Figure B.2: Estimation results of function coefficients of the influence of broad money supply on output

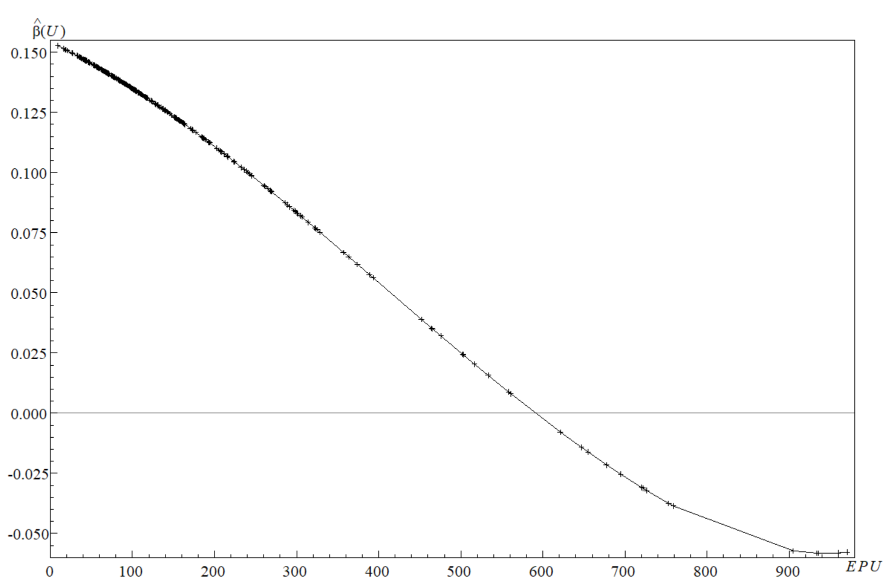
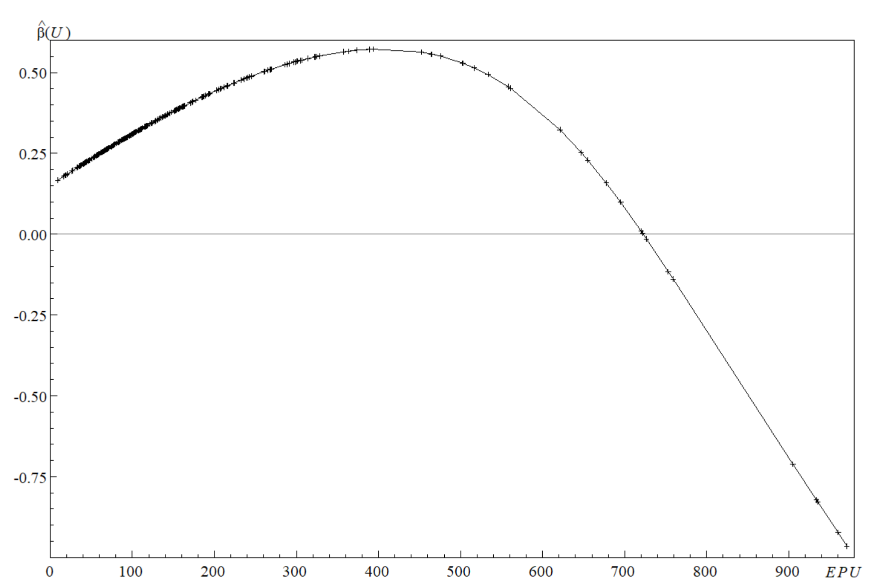


Figure B.3: Estimation results of function coefficients of the influence of broad money supply on prices



B.2 Mathematical Proofs

Proof of (3.4). By Taylor expansion

$$\beta(Z_t;\hat{\theta}) = \beta(Z_t;\theta_*) + \beta^{(1)}(Z_t;\bar{\theta})(\hat{\theta}-\theta_*),$$

where $\bar{\theta} = \alpha\theta_* + (1-\alpha)\hat{\theta}$ for some $0 < \alpha < 1$. Furthermore, by the FOC, we have

$$\begin{aligned}
0 &= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\left[Y_t - X_t'\beta(Z_t;\hat{\theta})\right] \\
&= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\hat{\theta})\right] \\
&= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_*)\right] \\
&\quad - \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t;\hat{\theta}) - \beta(Z_t;\theta_*)\right] \\
&= \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_*)\right] \\
&\quad - \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})(\hat{\theta}-\theta_*).
\end{aligned}$$

Thus, it follows

$$\hat{\theta}-\theta_* = \left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\left\{\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t + \sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t) - \beta(Z_t;\theta_*)\right]\right\}. \tag{B.1}$$

By Taylor expansion, the second term in $\hat{A}(u)$ has

$$\begin{aligned}
&-\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t;\hat{\theta}) - \beta(Z_t;\theta_*)\right]e^{iu'Z_t} \\
&= -\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}(\hat{\theta}-\theta_*).
\end{aligned}$$

Plugging in B.1 to substitute $\hat{\theta} - \theta_*$, we have the following result for the second term of $\hat{A}(u)$:

$$-\left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t$$
$$-\left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'$$
$$\times\left[\beta(Z_t)-\beta(Z_t;\theta_*)\right].$$

Plug it back to $\hat{A}(u)$, then we have

$$\begin{aligned}\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)X_t'[\beta(Z_t)-\beta(Z_t;\theta_*)]\\ &\quad+\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\varepsilon_t\\ &= \hat{A}_1(u)+\hat{A}_2(u),\end{aligned}$$

where $\hat{G}_t(u)$ is a $l\times 1$ complex-valued vector empirical process such that

$$\hat{G}_t(u)=\left\{I_le^{iu'Z_t}-\left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\right\}\beta^{(1)}(Z_t;\hat{\theta})'X_t.$$

□

Lemma B.2.1. *Let Assumptions 3.2.1-3.2.4 hold, then*

(i). $\hat{G}_t(u) - G_t(u)$ *is stochastically equicontinuous over* $u \in \mathbb{R}^l$*;*
(ii). $\sup_{u\in\mathbb{R}^l}\|\hat{G}_t(u)-G_t(u)\| \xrightarrow{p} 0$,
where

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

Proof of Lemma B.2.1. For (i), we need to show that, for any $\epsilon > 0$ and $\kappa > 0$, there exists a $\delta > 0$ such that

$$\lim_{T\to\infty}P\left[\sup_{u_1,u_2\in\mathbb{R}^l:\|u_1-u_2\|<\delta}\left\|\hat{G}_t(u_1)-G_t(u_1)-(\hat{G}_t(u_2)-G_t(u_2))\right\|>\kappa\right]<\epsilon.$$

By Triangle inequality, it suffices to show that $\hat{G}_t(u)$ and $G_t(u)$ are both stochastically equicontinuous. For $\hat{G}_t(u)$, let $\bar{u} = \eta u_1 + (1-\eta)u_2$ for some $\eta \in (0,1)$ such that

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

and $B_t = \left[\frac{1}{T}\sum_{t=1}^T \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta})\right]^{-1} \beta^{(1)}(Z_t;\hat{\theta})' X_t$, then

$$\begin{aligned}
& \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^l:\|u_1-u_2\|<\delta} \left\|\hat{G}_t(u_1) - \hat{G}_t(u_2)\right\| > \kappa\right] \\
\leq\ & \lim_{T\to\infty} P\Bigg[\sup_{u_1,u_2\in\mathbb{R}^l:\|u_1-u_2\|<\delta} \left\|\beta^{(1)}(Z_t;\hat{\theta})' X_t \left(e^{iu_1' Z_t} - e^{iu_2' Z_t}\right)\right\| \\
& + \left\|\left[\frac{1}{T}\sum_{t=1}^T \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) \left(e^{iu_1' Z_t} - e^{iu_2' Z_t}\right)\right] B_t\right\| > \kappa\Bigg] \\
=\ & \lim_{T\to\infty} P\Bigg[\sup_{u_1,u_2\in\mathbb{R}^l:\|u_1-u_2\|<\delta} \left\|\beta^{(1)}(Z_t;\hat{\theta})' X_t i Z_t'(u_1-u_2)e^{i\bar{u}' Z_t}\right\| \\
& + \left\|\left[\frac{1}{T}\sum_{t=1}^T \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) i Z_t'(u_1-u_2)e^{i\bar{u}' Z_t}\right] B_t\right\| > \kappa\Bigg] \\
\leq\ & \lim_{T\to\infty} P\Bigg[\sup_{u_1,u_2\in\mathbb{R}^l:\|u_1-u_2\|<\delta} \left\|\beta^{(1)}(Z_t;\hat{\theta})' X_t Z_t'\right\| \|u_1 - u_2\| \\
& + \left\|\left[\sqrt{\frac{1}{T}\sum_{t=1}^T \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) Z_t'} \sqrt{\frac{1}{T}\sum_{t=1}^T \|(u_1-u_2)e^{i\bar{u}' Z_t}\|^2}\right] B_t\right\| > \kappa\Bigg] \\
\leq\ & \lim_{T\to\infty} P\left[\left\|\beta^{(1)}(Z_t;\hat{\theta})' X_t Z_t'\right\| + \left\|\left[\sqrt{\frac{1}{T}\sum_{t=1}^T \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) Z_t'}\right] B_t\right\| > \kappa/\delta\right],
\end{aligned}$$

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 3.2.3 and 3.2.4(ii), we have $\left\|\beta^{(1)}(Z_t;\hat{\theta})' X_t Z_t'\right\| = O_P(1)$ and $\left\|\left[\sqrt{\frac{1}{T}\sum_{t=1}^T \beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) Z_t'}\right] B_t\right\| = O_P(1)$. Thus, for any $\epsilon > 0$, we can find a $\delta > 0$ small enough such that

$$\lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}^l:\|u_1-u_2\|<\delta} \left\|\hat{G}_t(u_1) - \hat{G}_t(u_2)\right\| > \kappa\right] < \epsilon.$$

The stochastic equicontinuity of $G_t(u)$ can be shown in an analogous way.

Next, we show (ii). By the definition of $\hat{G}_t(u)$ and $G_t(u)$, we decompose

$$
\begin{aligned}
& \hat{G}_t(u) - G_t(u) \\
= \; & \left[\beta^{(1)}(Z_t;\hat{\theta}) - \beta^{(1)}(Z_t;\theta_*)\right]' X_t e^{iu'Z_t} \\
& - \left\{ \left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) e^{iu'Z_t}\right] \left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta})\right]^{-1} \beta^{(1)}(Z_t;\hat{\theta})' X_t \right. \\
& \left. - E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*) e^{iu'Z_t}\right] E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*)\right]^{-1} \beta^{(1)}(Z_t;\theta_*)' X_t \right\} \\
= \; & \hat{a}(u) - a(u) - \left\{ \left[\hat{b}(u) - b(u)\right](\hat{c} - c)\left(\hat{d} - d\right) + \left[\hat{b}(u) - b(u)\right](\hat{c} - c)\, d + \left[\hat{b}(u) - b(u)\right] c \left(\hat{d} - d\right) \right. \\
& \left. + \left[\hat{b}(u) - b(u)\right] cd + b(u)\,(\hat{c} - c)\left(\hat{d} - d\right) + b(u)\,(\hat{c} - c)\, d + b(u) c \left(\hat{d} - d\right) \right\},
\end{aligned}
$$

where we let

$$
\begin{aligned}
\hat{a}(u) &= \beta^{(1)}(Z_t;\hat{\theta})' X_t e^{iu'Z_t} \\
\hat{b} &= \left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta}) e^{iu'Z_t}\right] \\
\hat{c} &= \left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})' X_t X_t' \beta^{(1)}(Z_t;\bar{\theta})\right]^{-1} \\
\hat{d} &= \beta^{(1)}(Z_t;\hat{\theta})' X_t \\
a(u) &= \beta^{(1)}(Z_t;\theta_*)' X_t e^{iu'Z_t} \\
b(u) &= E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*) e^{iu'Z_t}\right] \\
b &= E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*)\right]^{-1} \\
c &= \beta^{(1)}(Z_t;\theta_*)' X_t.
\end{aligned}
$$

By triangle inequality, we have

$$\begin{aligned}
& \sup_{u\in\mathbb{R}^l} \|\hat{G}_t(u) - G_t(u)\| \\
\leq\; & \sup_{u\in\mathbb{R}^l} \|\hat{a}(u) - a(u)\| + \sup_{u\in\mathbb{R}^l} \left\| \left[\hat{b}(u) - b(u)\right](\hat{c} - c)\left(\hat{d} - d\right)\right\| \\
& + \sup_{u\in\mathbb{R}^l} \left\| \left[\hat{b}(u) - b(u)\right](\hat{c} - c)\, d\right\| + \sup_{u\in\mathbb{R}^l} \left\| \left[\hat{b}(u) - b(u)\right] c\left(\hat{d} - d\right)\right\| \\
& + \sup_{u\in\mathbb{R}^l} \left\| \left[\hat{b}(u) - b(u)\right] cd\right\| + \sup_{u\in\mathbb{R}^l} \left\| b(u)(\hat{c} - c)\left(\hat{d} - d\right)\right\| + \sup_{u\in\mathbb{R}^l} \|b(u)(\hat{c} - c)\, d\| + \sup_{u\in\mathbb{R}^l} \left\| b(u) c\left(\hat{d} - d\right)\right\| \\
=\; & \sum_{j=1}^{8} R_j, \text{ say.}
\end{aligned}$$

Then, to show that $\hat{G}_t(u) - G_t(u)$ uniformly converge to 0 over $u \in \mathbb{R}^l$, it is equivalent to show that $R_j \xrightarrow{p} 0$ for $j = 1, .., 8$. For each R_j, we first establish the pointwise convergence for each fixed $u \in \mathbb{R}^l$, and then show that the corresponding empirical process is stochastically equicontinuous. Finally, we show that each R_j converges to 0 in probability.

We start with R_1. For each fixed $u \in \mathbb{R}^l$,

$$\begin{aligned}
\hat{a}(u) - a(u) &= \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*)\right]' X_t e^{iu'Z_t} \\
&\xrightarrow{p} 0,
\end{aligned}$$

by Assumption 3.2.4. For R_2, we show that $\hat{b}(u) - b(u) \xrightarrow{p} 0$, $\hat{c} - c \xrightarrow{p} 0$, and $\hat{d} - d \xrightarrow{p} 0$, respectively.

$$\begin{aligned}
& \hat{b}(u) - b(u) \\
=\; & \frac{1}{T}\sum_{t=1}^{T} \beta^{(1)}(Z_t; \hat{\theta})' X_t X_t' \beta^{(1)}(Z_t; \bar{\theta}) e^{iu'Z_t} - E\left[\beta^{(1)}(Z_t; \theta_*)' X_t X_t' \beta^{(1)}(Z_t; \theta_*) e^{iu'Z_t}\right] \\
=\; & \frac{1}{T}\sum_{t=1}^{T} \beta^{(1)}(Z_t; \theta_*)' X_t X_t' \beta^{(1)}(Z_t; \theta_*) e^{iu'Z_t} - E\left[\beta^{(1)}(Z_t; \theta_*)' X_t X_t' \beta^{(1)}(Z_t; \theta_*) e^{iu'Z_t}\right] + o_P(1) \\
=\; & O_P(T^{-1/2}),
\end{aligned}$$

where the second equality holds given Assumption 3.2.4(i), (iii), and the fact that $\bar{\theta}$ lies between $\hat{\theta}$ and θ_*. The last equality is due to the mixing condition in

Assumption 3.2.1 and the moment conditions in Assumption 3.2.3 and 3.2.4(ii). By analogous arguments, it is straightforward to show $\hat{c} - c \xrightarrow{p} 0$, and $\hat{d} - d \xrightarrow{p} 0$. Therefore, the pointwise convergence for each empirical process in R_j has been established.

Next, to show each empirical process in R_j is stochastically equicontinuous, we can follow similar arguments in Proof of Lemma B.2.1(i).

Lastly, we show the uniform convergence for each R_j. We start with R_1:

$$
\begin{aligned}
R_1 &= \sup_{u \in \mathbb{R}^l} \|\hat{a}(u) - a(u)\| \\
&= \sup_{u \in \mathbb{R}^l} \left\| \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*) \right]' X_t e^{iu'Z_t} \right\| \\
&\leq \sup_{u \in \mathbb{R}^l} \left\| \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*) \right]' X_t \cos(u'Z_t) \right\| + \sup_{u \in \mathbb{R}^l} \left\| \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*) \right]' X_t \sin(u'Z_t) \right\| \\
&= R_{11} + R_{12}, \text{ say.}
\end{aligned}
$$

Next, we show that $R_{11} = o_P(1)$ and $R_{12} = o_P(1)$, respectively.

Let $\mathcal{U}_T = [-s_T, s_T]^l \subset \mathbb{R}^l$, where we let $s_T \to \infty$ as $T \to \infty$, then we want to show for any $\kappa > 0$ and $\epsilon > 0$, there exists a $T(\delta, \kappa)$ such that

$$
P(R_{11} > \kappa) < \epsilon,
$$

for all $T > T(\delta, \kappa)$. Let $\mathcal{U}_T^c = \mathcal{R}^l - \mathcal{U}_T$ be the complement set of $\mathcal{U}_T$, then

$$
\begin{aligned}
P(R_{11} > \kappa) &= P\left(\sup_{u \in \mathbb{R}^l} \left\| \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*) \right]' X_t \cos(u'Z_t) \right\| > \kappa \right) \\
&= \max \left\{ P\left(\sup_{u \in \mathcal{U}_T} \left\| \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*) \right]' X_t \cos(u'Z_t) \right\| > \kappa \right), \right. \\
&\qquad \left. P\left(\sup_{u \in \mathcal{U}_T^c} \left\| \left[\beta^{(1)}(Z_t; \hat{\theta}) - \beta^{(1)}(Z_t; \theta_*) \right]' X_t \cos(u'Z_t) \right\| > \kappa \right) \right\}.
\end{aligned}
$$

Due to the periodicity and boundness of $\cos(\cdot)$, for any given T, we can find a s_T

that is large enough such that

$$
\begin{aligned}
& P\left(\sup_{u\in\mathcal{U}_T}\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\cos(u'Z_t)\right\|<\kappa\right)\\
\geq\; & P\left(\sup_{u\in\mathcal{U}_T^c}\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\cos(u'Z_t)\right\|<\kappa\right),
\end{aligned}
$$

for any $\kappa > 0$. Let $\left\{B_T(u_j,\delta) : j = 1,2,...,J\right\}$ be a finite cover of $\mathcal{U}_T$ such that $u \in B(u_j,\delta)$ if and only if $d(u,u_j) \equiv \|u_j - u\| \leq \delta$. Then, it follows

$$
\begin{aligned}
\lim_{T\to\infty} P(R_{11}>\kappa) \;\leq\; & \lim_{T\to\infty} P\left(\sup_{u\in\mathcal{U}_T}\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\cos(u'Z_t)\right\|>\kappa\right)\\
\leq\; & \lim_{T\to\infty} P\Bigg(\max_{j\leq J}\sup_{\tilde{u}\in B(u_j,\delta)}\Bigg\{\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\left[\cos\left(\tilde{u}'Z_t\right)-\cos\left(u_j'Z_t\right)\right]\right\|\\
& +\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\cos\left(u_j'Z_t\right)\right\|\Bigg\}>\kappa\Bigg)\\
\leq\; & \lim_{T\to\infty} P\left(\sup_{\tilde{u},u_j\in\mathbb{D}:d(\tilde{u},u_j)<\delta}\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\left[\cos\left(\tilde{u}'Z_t\right)-\cos\left(u_j'Z_t\right)\right]\right\|>\frac{\kappa}{2}\right)\\
& +\lim_{T\to\infty} P\left(\max_{j\leq J}\left\|\left[\beta^{(1)}(Z_t;\hat{\theta})-\beta^{(1)}(Z_t;\theta_*)\right]' X_t\cos(u_j'Z_t)\right\|\Bigg\}>\frac{\kappa}{2}\right)\\
=\; & 0,
\end{aligned}
$$

where the last inequality hold since $\hat{a}(u) - a(u)$ is stochastically equicontinuous and $\hat{a}(u) - a(u) \xrightarrow{p} 0$ for any $u \in \mathbb{R}^l$. The derivation for $R_{12} \xrightarrow{p} 0$ is quite similar. The uniform convergence for R_2 to R_8 can be established in an analogous way. We neglect it to save space. □

Proof of Proposition 3.2.1. By definition,

$$
\begin{aligned}
\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T} S_t(\hat{\theta})e^{iu'Z_t} \\
&= \frac{1}{T}\sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t)-\beta(Z_t;\theta_*)\right]e^{iu'Z_t} \\
&\quad -\frac{1}{T}\sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\left[\beta(Z_t;\hat{\theta})-\beta(Z_t;\theta_*)\right]e^{iu'Z_t} \\
&\quad +\frac{1}{T}\sum_{t=1}^{T} \beta^{(1)}(Z_t;\hat{\theta})'X_t\varepsilon_t e^{iu'Z_t} \\
&= \frac{1}{T}\sum_{t=1}^{T} \hat{G}_t(u)X_t'[\beta(Z_t)-\beta(Z_t;\theta_*)] \\
&\quad +\frac{1}{T}\sum_{t=1}^{T} \hat{G}_t(u)\varepsilon_t \\
&= \hat{A}_1(u)+\hat{A}_2(u),
\end{aligned}
$$

where $\hat{G}_t(u)$ is a $l \times 1$ complex-valued vector empirical process such that

$$
\hat{G}_t(u) = \left\{ I_l e^{iu'Z_t} - \left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})e^{iu'Z_t}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\beta^{(1)}(Z_t;\hat{\theta})'X_tX_t'\beta^{(1)}(Z_t;\bar{\theta})\right]^{-1}\right\}\beta^{(1)}(Z_t;\hat{\theta})'X_t.
$$

Under $\mathbb{H}_0 : \beta(Z_t) = \beta(Z_t;\theta_0)$, and $\theta_* = \theta_0$, we have

$$
\hat{A}_1(u) = 0
$$

for all $u \in \mathbb{R}^l$. By Lemma B.2.1(ii), we have

$$
\hat{A}(u) = \frac{1}{T}\sum_{t=1}^{T} G_t(u)\varepsilon_t + o_P(1),
$$

where

$$
G_t(u) = \left\{ I_l e^{iu'Z_t} - E\left[\beta^{(1)}(Z_t;\theta_0)'X_tX_t'\beta^{(1)}(Z_t;\theta_0)e^{iu'Z_t}\right]E\left[\beta^{(1)}(Z_t;\theta_0)'X_tX_t'\beta^{(1)}(Z_t;\theta_0)\right]^{-1}\right\}\beta^{(1)}(Z_t;\theta_0)'X_t.
$$

Lemma B.2.1(i) also establishes that $G_t(u)$ is stochastically equicontinuous, thus, under Assumptions 3.2.1-3.2.4, by FCLT

$$
\sqrt{T}\hat{A}(u) \Rightarrow \mathcal{G}(u),
$$

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

$$\mathcal{K}(u,v) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} E[G_t(u) G_s(v)^* \varepsilon_t \varepsilon_s]$$

□

Proof of Theorem 3.2.1. Under $\mathbb{H}_0$, Proposition 3.2.1 shows that $\sqrt{T}\hat{A}(u)$ weakly converges to a complexed valued Gaussian process $\mathcal{G}(u)$. Under Assumption 3.2.4, by Lemma B.2.1(i) and continuous mapping theorem, we have

$$\hat{D} \xrightarrow{p} \int_{\mathbb{R}^l} \|\mathcal{G}(u)\|^2 W(u) du,$$

as $T \to \infty$. □

Proof of Proposition 3.2.2. By Lemma B.2.1(ii), we have

$$\hat{A}_1(u) \quad = \quad \frac{1}{T} \sum_{t=1}^{T} G_t(u) X_t' \left[\beta(Z_t) - \beta(Z_t, \theta_*)\right] + o_P(1)$$

where

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

Let $\phi(Z_t) = \beta(Z_t) - \beta(Z_t, \theta_*)$, under $\mathbb{H}_A$, we can show that

$$\begin{aligned}
& \frac{1}{T} \sum_{t=1}^{T} G_t(u) X_t' \left[\beta(Z_t) - \beta(Z_t, \theta_*)\right] \\
= \; & \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\theta_*)' X_t X_t' \phi(Z_t) e^{iu' Z_t} \\
& - E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*) e^{iu' Z_t}\right] E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*)\right]^{-1} \frac{1}{T} \sum_{t=1}^{T} \beta^{(1)}(Z_t;\theta_*)' X_t X_t' \phi(Z_t) \\
\Rightarrow \; & E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \phi(Z_t) e^{iu' Z_t}\right] \\
& - E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*) e^{iu' Z_t}\right] E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*)\right]^{-1} E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \phi(Z_t)\right] \\
= \; & E\left[\Xi_t e^{iu' Z_t}\right],
\end{aligned}$$

where we let

$$\Xi_t = \beta^{(1)}(Z_t;\theta_*)' X_t X_t' \phi(Z_t)$$
$$- \beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*) E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \beta^{(1)}(Z_t;\theta_*)\right]^{-1} E\left[\beta^{(1)}(Z_t;\theta_*)' X_t X_t' \phi(Z_t)\right].$$

By Bierens (1982), $E\left[\Xi_t e^{iu'Z_t}\right] = 0$ for all $u \in \mathbb{R}^l$ if and only if $E(\Xi_t|Z_t)$=0 almost surely. □

Proof of Theorem 3.2.2. Under $\mathbb{H}_A(\Delta_T)$, with $\Delta_T = T^{-1/2}$, we have

$$\begin{aligned} \sqrt{T}\hat{A}(u) &= \sqrt{T}\hat{A}_1(u) + \sqrt{T}\hat{A}_2(u) \\ &= \frac{1}{T}\sum_{t=1}^{T} \hat{G}_t(u) X_t' \phi(Z_t) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \hat{G}_t(u)\varepsilon_t \\ &\Rightarrow \xi(u) + \mathcal{G}(u), \end{aligned}$$

where $\xi(u) = E(G_t(u)X_t'\phi(Z_t))$, by Proposition 3.1 and 3.2. By continuous mapping theorem, and Lemma A.1(i), it follows

$$\hat{D} \xrightarrow{p} \int_{\mathbb{R}^l} \|\xi(u) + \mathcal{G}(u)\|^2\, W(u)du.$$

□

APPENDIX C

APPENDIX OF CHAPTER 4

C.1 Figures and Tables

Table C.1: Summary Statistics

	Pre-Euro		Post-Euro	
	DM-USD	Yen-USD	DM-USD	Yen-USD
Mean	0.005	−0.009	0.035	0.017
Std. Dev.	0.673	0.729	0.662	0.693
Skewness	−0.024	−0.789	−0.184	−0.198
Kurtosis	5.065	8.896	4.107	4.903
Jarque-Bera statistic	370.7*	3238*	44.38*	123.2*
ARCH LM statistic	110.5*	205.6*	6.69	20.28*
Linear correlation	0.509		0.116	
Number of obs.	2086		783	

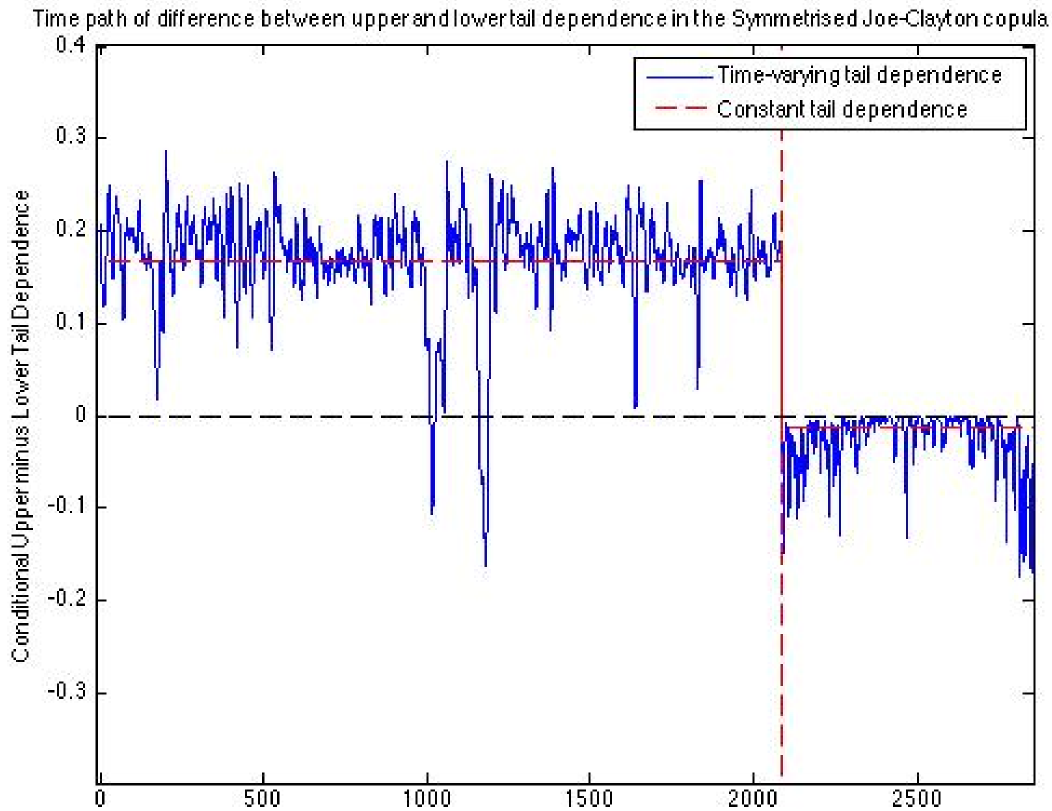
Note: The data are 100 times the log-difference of the daily Deutche mark-U.S. dollar and Japanese yen-U.S. dollar exchange rates. 2,086 in the Pre-Euro sample prio to the introduction of the euro on January 1,1999 and 783 in the Post-Euro sample after the introduction of the euro. The ARCH LM statistic is conducted using 10 lags. An asterisk (*) indicates a rejection of the null at the 0.05 level.

Table C.2: Results For the Marginal Distributions

	DM-USD Margin		Yen-USD Margin	
	Pre-Euro	Post-Euro	Pre-Euro	Post-Euro
Constant	0.008	0.036	−0.006	0.029
	(0.013)	(0.024)	(0.014)	(0.024)
AR(1)	0.022	0.011	0.028	−0.030
	(0.021)	(0.041)	(0.023)	(0.039)
AR(10)	–	–	0.079	0.022
	–	–	(0.024)	(0.040)
GARCH constant	0.005	0.003	0.010	0.006
	(0.001)	(0.003)	(0.001)	(0.004)
Lagged variance	0.950	0.983	0.925	0.956
	(0.007)	(0.012)	(0.006)	(0.015)
Lagged term	0.039	0.009	0.056	0.030
	(0.006)	(0.006)	(0.005)	(0.008)

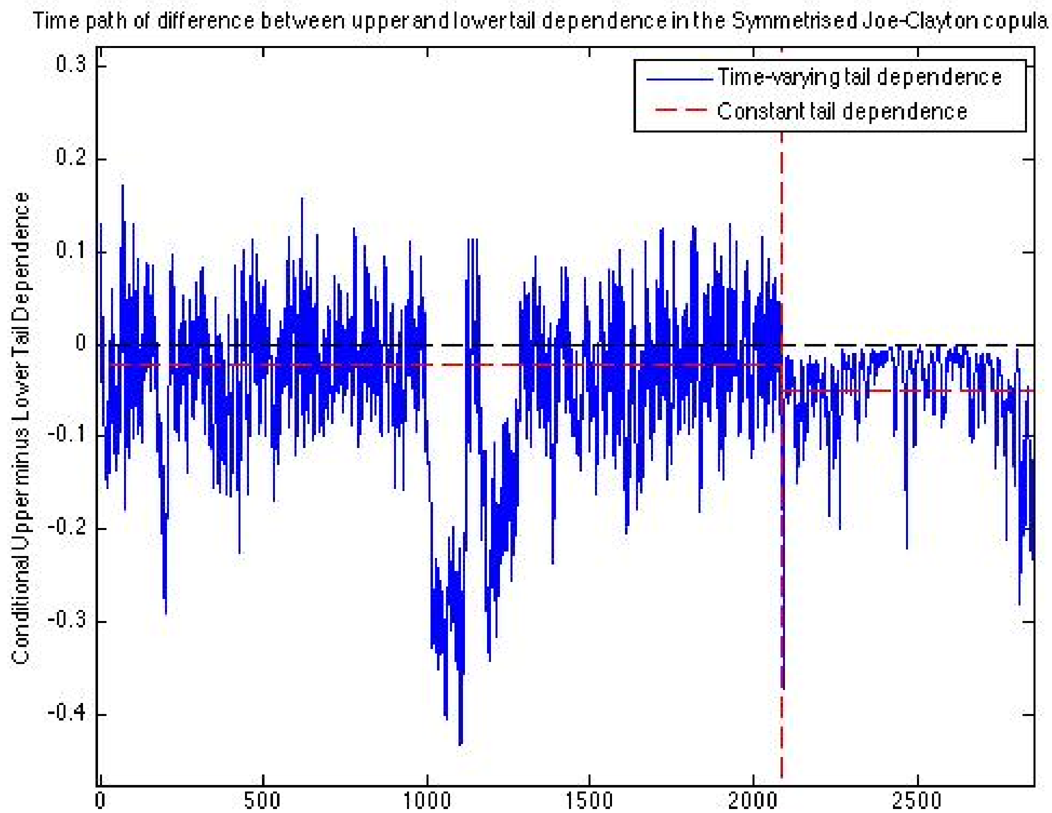
Note: Here we report the QMLE, with asymototic standard errors in parenthese, of the parameters of the marginal distribtution models for the two exchange rates. From the marginal distribution specification tests, the 10th lag was not found to be important for the DM-USD exchange rate. That is why its estimator is not reported here.

Figure C.1: Symmetrised Joe-Clayton copula with stationary marginals



Difference between upper and lower tail dependence from the symmetrized Joe-Clayton copulas allowing for a structural break at the introduction of the euro on January 1, 1999.

Figure C.2: Symmetrised Joe-Clayton copula with non-stationary marginals



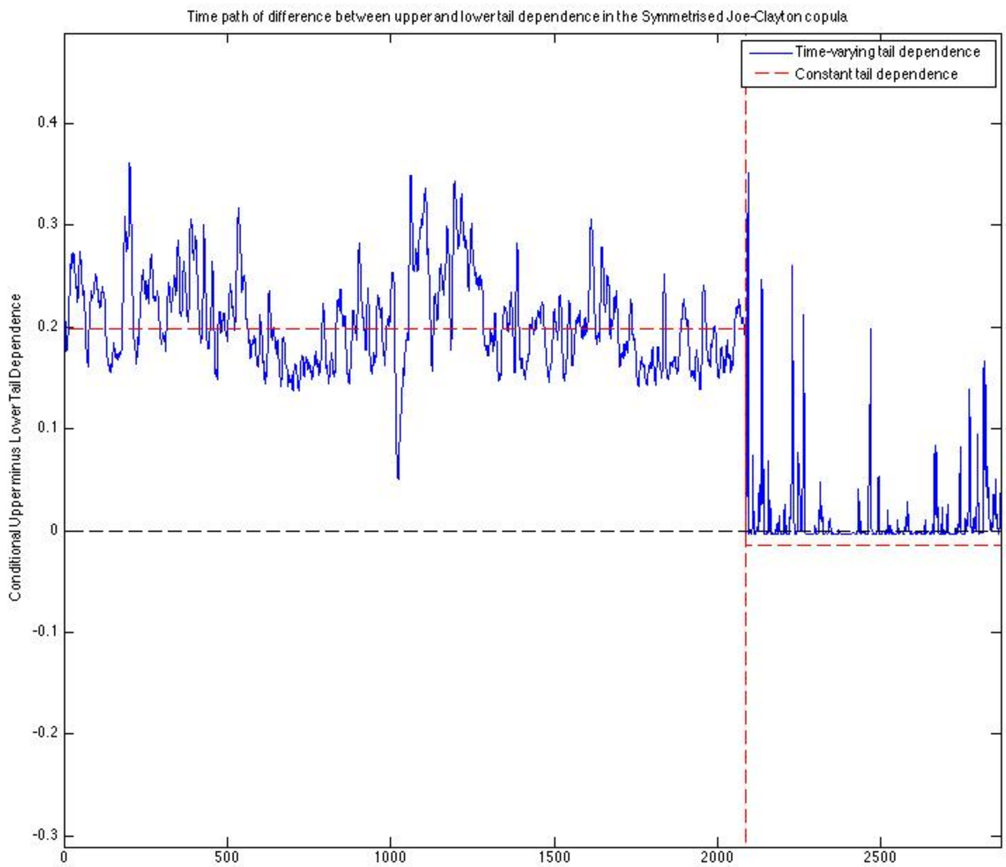
Difference between upper and lower tail dependence from the symmetrized Joe-Clayton copulas using empirical CDF as the marginal distribution.

Table C.3: Test For Smooth Structural Changes in GARCH Models

	LR		Adjusted LR	
	Pre-Euro	Post-Euro	Pre-Euro	Post-Euro
DM-USD	15.430	1.437	12.387	1.732
Yen-USD	4.696	−1.706	5.773	−1.962

Note: Adjusted LR acounts for departure from normal distribution, and thus adjusts LR by esitimated kurtosis, i.e. $\hat{\kappa} = \frac{1}{T}\sum_{t=1}^{T}(\hat{\varepsilon}_t^4 - 3)$.

Figure C.3: Difference between upper and lower tail dependence from the symmetrized Joe-Clayton copulas using time-varying parameter GARCH as the marginal distribution.



Difference between upper and lower tail dependence from the symmetrized Joe-Clayton copulas using time-varying parameter GARCH as the marginal distribution.

BIBLIOGRAPHY

[1] Andrew B. Abel. Asset prices under habit formation and catching up with joneses. *American Economic Review (Papers and Proceedings)*, 80(2):38–42, May 1990.

[2] Donald WK Andrews. Generic uniform convergence. *Econometric theory*, pages 241–257, 1992.

[3] Donald WK Andrews. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 43–72, 1994.

[4] Donald WK Andrews. Empirical process methods in econometrics. *Handbook of econometrics*, 4:2247–2294, 1994.

[5] Donald WK Andrews. Nonparametric kernel estimation for semiparametric models. *Econometric Theory*, pages 560–596, 1995.

[6] Donald WK Andrews 1. An introduction to econometric applications of empirical process theory for dependent random variables. *Econometric Reviews*, 12(2):183–216, 1993.

[7] Andrew Ang and Dennis Kristensen. Testing conditional factor models. *Journal of Financial Economics*, 106(1):132–156, 2012.

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

[9] Nicholas Barberis, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. X-capm: An extrapolative capital asset pricing model. Technical report, National Bureau of Economic Research, 2013.

[10] Nicholas Barberis, Ming Huang, and Tano Santos. Prospect theory and asset prices. *The Quarterly Journal of Economics*, 116(1):1–53, 2001.

[11] Herman J Bierens. Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134, 1982.

[12] Herman J Bierens. A consistent conditional moment test of functional form. *Econometrica: Journal of the Econometric Society*, pages 1443–1458, 1990.

[13] Fischer Black. Noise. *Journal of Finance*, 41(3):529–43, July 1986.

[14] Eduardo Borensztein, Jose De Gregorio, and Jong-Wha Lee. How does foreign direct investment affect economic growth? *Journal of international Economics*, 45(1):115–135, 1998.

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

[16] Zongwu Cai, Mitali Das, Huaiyu Xiong, and Xizhi Wu. Functional coefficient instrumental variables models. *Journal of Econometrics*, 133(1):207–241, 2006.

[17] Zongwu Cai, Jianqing Fan, and Runze Li. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902, 2000.

[18] Zongwu Cai, Jianqing Fan, and Qiwei Yao. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956, 2000.

[19] Zongwu Cai and Qi Li. Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, pages 1321–1342, 2008.

[20] Zongwu Cai, Qi Li, and Joon Y Park. Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113, 2009.

[21] John Y Campbell. Asset prices, consumption, and the business cycle. *Handbook of macroeconomics*, 1:1231–1303, 1999.

[22] John Y Campbell. Asset pricing at the millennium. *The Journal of Finance*, 55(4):1515–1567, 2000.

[23] John Y Campbell and John H Cochrane. Explaining the poor performance of consumption-based asset pricing models. *The Journal of Finance*, 55(6):2863–2878, 2000.

[24] John Y Campbell and Ludger Hentschel. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of financial Economics*, 31(3):281–318, 1992.

[25] John Y Campbell and Robert J Shiller. The dividend-price ratio and expectations of future dividends and discount factors. *Review of financial studies*, 1(3):195–228, 1988.

[26] John Y. Campbell and Robert J. Shiller. Stock prices, earnings, and expected dividend. *Journal of Finance*, 43(3):661–76, July 1988.

[27] Mehmet Caner and Bruce E Hansen. Threshold autoregression with a unit root. *Econometrica*, 69(6):1555–1596, 2001.

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

[29] Stephen G Cecchetti, Pok-Sang Lam, and Nelson Mark. Mean reversion in equilibrium asset prices, 1990.

[30] Stephen G Cecchetti, Pok-sang Lam, and Nelson C Mark. The equity premium and the risk-free rate: Matching the moments. *Journal of Monetary Economics*, 31(1):21–45, 1993.

[31] Stephen G Cecchetti, Pok-sang Lam, and Nelson C Mark. Asset pricing with distorted beliefs: are equity returns too good to be true? Technical report, National Bureau of Economic Research, 1998.

[32] Bin Chen and Yongmiao Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, 2012.

[33] Bin Chen and Yongmiao Hong. Detecting for smooth structural changes in garch models. *Econometric Theory*, 32(3):740, 2016.

[34] Xiaohong Chen and Yanqin Fan. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of econometrics*, 135(1-2):125–154, 2006.

[35] Xiaohong Chen, Yanqin Fan, and Viktor Tsyrennikov. Efficient estimation of semiparametric multivariate copula models. *Journal of the American Statistical Association*, 101(475):1228–1240, 2006.

[36] Xiaohong Chen, Roger Koenker, and Zhijie Xiao. Copula-based nonlinear quantile autoregression. *The Econometrics Journal*, 12:S50–S67, 2009.

[37] John H. Cochrane. How big is the random walk in gnp? *Journal of Political Economy*, 96(5):893–920, October 1998.

[38] John H Cochrane. Asset pricing. 2001.

[39] John H Cochrane. Presidential address: Discount rates. *The Journal of Finance*, 66(4):1047–1108, 2011.

[40] Fabrice Collard and Michel Juillard. Accuracy of stochastic perturbation methods: The case of asset pricing models. *Journal of Economic Dynamics and Control*, 25(6):979–999, 2001.

[41] Andrew J Culham. Computational techniques for solving asset pricing models. 2005.

[42] Martin S Eichenbaum, Lars Peter Hansen, and Kenneth J Singleton. A time series analysis of representative agent models of consumption and leisure choice under uncertainty. *The Quarterly Journal of Economics*, 103(1):51–78, 1988.

[43] Larry G. Epstein and Stanley E. Zin. First order risk aversion and the equity premium puzzle. *Journal of Monetary Economics*, 26(3):387–407, December 1990.

[44] Martin D. D. Evans. Dividend variability and stock market swings. *Review of Economic Studies*, 65(4):711–40, October 1998.

[45] Eugene F Fama and Kenneth R French. Multifactor explanations of asset pricing anomalies. *The journal of finance*, 51(1):55–84, 1996.

[46] Eugene F Fama and Kenneth R French. The capital asset pricing model: Theory and evidence. *Journal of economic perspectives*, 18(3):25–46, 2004.

[47] Eugene F Fama and Kenneth R French. International tests of a five-factor asset pricing model. *Journal of financial Economics*, 123(3):441–463, 2017.

[48] Eugene F. Famma and Kenneth R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–27, October 1988.

[49] Jianqing Fan, Wolfgang Härdle, Enno Mammen, et al. Direct estimation of low-dimensional components in additive models. *The Annals of Statistics*, 26(3):943–971, 1998.

[50] Jianqing Fan, Qiwei Yao, and Zongwu Cai. Adaptive varying-coefficient linear models. *Journal of the Royal Statistical Society: series B (statistical methodology)*, 65(1):57–80, 2003.

[51] Jianqing Fan, Chunming Zhang, Jian Zhang, et al. Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of statistics*, 29(1):153–193, 2001.

[52] Jianqing Fan, Wenyang Zhang, et al. Statistical estimation in varying coefficient models. *The annals of Statistics*, 27(5):1491–1518, 1999.

[53] Wayne E Ferson and Campbell R Harvey. The variation of economic risk premiums. *Journal of political economy*, 99(2):385–415, 1991.

[54] Kenneth R French, G William Schwert, and Robert F Stambaugh. Expected stock returns and volatility. *Journal of financial Economics*, 19(1):3–29, 1987.

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

[56] Clive WJ Granger, Timo Terasvirta, et al. Modelling non-linear economic relationships. *OUP Catalogue*, 1993.

[57] PJ Green and BW Silverman. Nonparametric regression and generalized linear models. number 58 in monographs on statistics and applied probability. *Nonparametric regression and generalized linear models*, (58), 1994.

[58] Robin Greenwood and Andrei Shleifer. Expectations of returns and expected returns. Technical report, National Bureau of Economic Research, 2013.

[59] Chong Gu, Nancy Heckman, and Grace Wahba. A note on generalized cross-validation with replicates. *Statistics & probability letters*, 14(4):283–287, 1992.

[60] James D Hamilton. Regime-switching models. *The new palgrave dictionary of economics*, 2, 2008.

[61] Bruce E Hansen. Lagrange multiplier tests for parameter instability in non-linear models. *University of Rochester*, 1990.

[62] Bruce E Hansen. Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econometric Theory*, 12(2):347–359, 1996.

[63] Bruce E Hansen. Inference in tar models. *Studies in Nonlinear Dynamics and Econometrics*, 2:1–14, 1997.

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

[65] Bruce E Hansen. Regression kink with an unknown threshold. *Journal of Business & Economic Statistics*, 35(2):228–240, 2017.

[66] Lars Peter Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054, 1982.

[67] Lars Peter Hansen, Thomas J Sargent, and Thomas D Tallarini Jr. Robust permanent income and pricing. *Review of Economic Studies*, pages 873–907, 1999.

[68] Lars Peter Hansen and Kenneth J Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica: Journal of the Econometric Society*, pages 1269–1286, 1982.

[69] Trevor Hastie and Robert Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386, 1987.

[70] James Heckman, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. Characterizing selection bias using experimental data. Technical report, National bureau of economic research, 1998.

[71] Yongmiao Hong. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *Journal of the American Statistical Association*, 94(448):1201–1220, 1999.

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

[73] Yongmiao Hong, Xia Wang, and Shouyang Wang. Testing strict stationar-

ity with applications to macroeconomic time series. *International Economic Review*, 58(4):1227–1277, 2017.

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

[75] Kewei Hou, Haitao Mo, Chen Xue, and Lu Zhang. Which factors? *Review of Finance*, 23(1):1–35, 2019.

[76] Kewei Hou, Chen Xue, and Lu Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 2015.

[77] Kewei Hou, Chen Xue, and Lu Zhang. A comparison of new factor models. *Fisher College of Business Working Paper*, (2015-03):05, 2017.

[78] Peter J Huber. Projection pursuit. *The annals of Statistics*, pages 435–475, 1985.

[79] Narayana R. Kocherlakota. The equity premium: It's still a puzzle. *Journal of Economic Literature*, 34(1):42–71, March 1996.

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

[81] Bong-Soo Lee and Beth Fisher Ingram. Simulation estimation of time-series models. *Journal of Econometrics*, 47(2):197–205, 1991.

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

[83] Ker-Chau Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

[84] Qi Li, Cliff J Huang, Dong Li, and Tsu-Tan Fu. Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20(3):412–422, 2002.

[85] Robert E. Lucas. Asset prices in an exchange economy. *Econometrica*, 66(6):1429–45, December 1978.

[86] Daniel McFadden. A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, pages 995–1026, 1989.

[87] Thomas Rietz. The equity premium puzzle: A solution. *Journal of Monetary Economics*, 22(1):117–33, July 1988.

[88] Peter M Robinson. Root-n-consistent semiparametric regression. *Econometrica: Journal of the Econometric Society*, pages 931–954, 1988.

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

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

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

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

[93] Yiguo Sun, Zongwu Cai, and Qi Li. Semiparametric functional coefficient models with integrated covariates. *Econometric Theory*, pages 659–672, 2013.

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

[95] Timo Teräsvirta et al. Modelling economic relationships with smooth transition regressions. Technical report, Stockholm School of Economics, 1996.

[96] Howell Tong. *Non-linear time series: a dynamical system approach*. Oxford University Press, 1990.

[97] Grace Wahba. *Variational methods for multidimensional inverse problems*. University of Wisconsin, Department of Statistics, 1984.

[98] Philippe Weil. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics*, 24(3):401–21, December 1989.

[99] Halbert White. *Asymptotic theory for econometricians*. Academic press, 2014.

[100] Lu Zhang. The investment capm. *European Financial Management*, 23(4):545–603, 2017.