THREE ESSAYS ON ESTIMATION AND TESTING OF NONPARAMETRIC MODELS

A Dissertation

by

GUANGYI MA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Economics

THREE ESSAYS ON ESTIMATION AND TESTING OF NONPARAMETRIC MODELS

A Dissertation

by

GUANGYI MA

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Ke-Li Xu
Committee Members,	Qi Li
	Dennis W. Jansen
	Hwagyun Kim
Head of Department,	Timothy Gronberg

August 2012

Major Subject: Economics

ABSTRACT

Three Essays on Estimation and Testing of Nonparametric Models. (August 2012)

Guangyi Ma, B.A., Harbin Institute of Technology; M.A., Seoul National University
Chair of Advisory Committee: Dr. Ke-Li Xu

In this dissertation, I focus on the development and application of nonparametric methods in econometrics. First, a constrained nonparametric regression method is developed to estimate a function and its derivatives subject to shape restrictions implied by economic theory. The constrained estimators can be viewed as a set of empirical likelihood-based reweighted local polynomial estimators. They are shown to be weakly consistent and have the same first order asymptotic distribution as the unconstrained estimators. When the shape restrictions are correctly specified, the constrained estimators can achieve a large degree of finite sample bias reduction and thus outperform the unconstrained estimators. The constrained nonparametric regression method is applied on the estimation of daily option pricing function and state-price density function.

Second, a modified Cumulative Sum of Squares (CUSQ) test is proposed to test structural changes in the unconditional volatility in a time-varying coefficient model. The proposed test is based on nonparametric residuals from local linear estimation of the time-varying coefficients. Asymptotic theory is provided to show that the new CUSQ test has standard null distribution and diverges at standard rate under the alternatives. Compared with a test based on least squares residuals, the new test enjoys correct size and good power properties. This is because, by estimating the model nonparametrically, one can circumvent the size distortion from potential structural changes in the mean. Empirical results from both simulation experiments and real data applications are presented to demonstrate the test's size and power properties.

Third, an empirical study of testing the Purchasing Power Parity (PPP) hypothesis is conducted in a functional-coefficient cointegration model, which is consistent with equilibrium models of exchange rate determination with the presence of transactions costs in international trade. Supporting evidence of PPP is found in the recent float exchange rate era. The cointegration relation of nominal exchange rate and price levels varies conditioning on the real exchange rate volatility. The cointegration coefficients are more stable and numerically near the value implied by PPP theory when the real exchange rate volatility is relatively lower.

DEDICATION

This dissertation is dedicated to my parents, Yuwen Ma and Min Liu, and my wife, Yunji, for their love, endless support, and encouragement over these many years; and to our wonderful son, Jason, who is the joy of our lives.

ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor Ke-Li Xu. His enthusiasm as a researcher has become a role model for me. His encouragement, guidance, and supervision have enabled me to fulfill my goal. During the past years Dr. Xu has always been willing and available to answer my questions, to share his ideas and experiences. I benefit a lot in the countless discussions with Dr. Xu, and therefore I hope that we can keep co-working in the future. I also wish to express my gratitude to my committee members: Qi Li, Dennis Jansen, and Hwagyun Kim. In particular, Dr. Li has nourished me in a wide range of aspects from econometrics to basketball skills. His diligence, integrity, amiability, and the pursuit of academic excellence have had a huge influence on me. Dr. Jansen has mentored me during my teaching practices, which will definitely be a valuable asset in my future career. I have learned macroeconomics and finance from Dr. Kim, whose classes drove me to the research frontier of those fields I have been interested in.

I would also hope to extend my thanks to other faculty members in the Department of Economics. At this moment I can remember my conversations with many professors during this long journey of academic growth. I am grateful to the secretarial staff for all their assistance. I want to thank my friends and colleagues in the department: Eul Jin, Sung Ick, Graham, Yaojing, Yichen, and Zhongwen, for their useful comments and encouragement from time to time. I was my pleasure to have known all of them.

Last, but obviously not least, a special word of thanks goes out to Dr. Yoon-Jae Whang at Seoul National University, from whom I have learned my first lesson in nonparametric econometrics.

TABLE OF CONTENTS

Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

1. INTRODUCTION AND SUMMARY

In the last two decades, there is a growing literature of nonparametric methods in statistics and econometrics; see the books by, for example, Fan and Gijbels (1996), Pagan and Ullah (1999), and Li and Racine (2007) for general discussion of nonparametric theory as well as applications. Moreover, nonparametric methods have been well recognized as useful tools in economic and financial data analysis because they are less restrictive compared with parametric models, and thus could be attractive when one has little prior information on the underlying data generating process; see the survey articles by Cai and Hong (2003), and Fan (2005). In this dissertation, I contribute to the research literature of nonparametric econometrics in the following aspects.

In the first study (Section 2: Empirical Likelihood-Based Constrained Nonparametric Regression), I develop a constrained nonparametric regression method to estimate a function and its derivatives subject to shape restrictions implied by economic theory. As a data-driven smoothing technique, the standard nonparametric local polynomial regression method produces curve estimates which heavily depend on the observed sample data. In other words, the shape of estimated functions by a standard local polynomial method may not always satisfy certain properties implied by the researcher's economic model. By introducing the empirical likelihood based, reweighted version of the local polynomial estimators, I show that one can accommodate a large range of shape restrictions using the modified estimators and achieve finite sample bias reductions. I also provide asymptotic analysis of the proposed constrained nonparametric estimation method, with a focus on the comparison of the constrained and unconstrained estimators.

In the second study (Section 3: The CUSUM of Squares Test for Volatility Change in a Time-varying Coefficient Model), I consider a testing problem of structural

This dissertation follows the style of Journal of Econometrics.

change in the volatility of time series models. Specifically, I propose a modified Cumulative Sum of Squares (CUSQ) test to test structural changes in the unconditional volatility in a time-varying coefficient model. The reason to consider this model is due to the concern of co-existing structural changes in both the first and second moments. The proposed test is based on nonparametric residuals from local linear estimation of the time-varying coefficients. I provide asymptotic theory to show that the new CUSQ test has standard null distribution and diverges at standard rate under the alternatives. Compared with a test based on least squares residuals, the new test enjoys correct size and good power properties. This is because, by estimating the model nonparametrically, one can circumvent the size distortion from potential structural changes in the mean. I present empirical results from both simulation experiments and real data applications to demonstrate the size and power properties.

In the third study (Section 4: Functional-Coefficient Cointegration Test of Purchasing Power Parity), I test the absolute version of Purchasing Power Parity (PPP) hypothesis in a functional-coefficient cointegration model, which is consistent with equilibrium models of exchange rate determination with the presence of transactions costs in international trade. I find supporting evidence of PPP in the recent float exchange rate era. The cointegration relation of nominal exchange rate and price levels varies conditioning on the real exchange rate volatility. The cointegration coefficients are more stable and numerically near the value implied by PPP theory when the real exchange rate volatility is relatively lower.

2. EMPIRICAL LIKELIHOOD-BASED CONSTRAINED NONPARAMETRIC REGRESSION

2.1 Introduction

Nonparametric regression methods are known to be robust to functional form misspecification, hence they are useful when the researcher does not have a theory specifying the exact relationship between economic variables. However, in many cases, economic theory indicates that the functional relationship between two variables X and Y, say, $Y = m(X)$, should be under certain shape restrictions such as monotonicity, convexity, homogeneity, etc. Because the estimation results from nonparametric regression are not guaranteed to satisfy these ex-ante model restrictions, it is desirable to develop a methodology to accommodate such conventional restrictions in nonparametric estimation.

In previous literature, various approaches to nonparametric regression which satisfy monotonic restriction have been developed. See Matzkin (1994) for a comprehensive survey. A popular approach in the existing research literature is the isotonic regression method. See, e.g., Hansen et al. (1973), Dykstra (1983), Goldman and Rudd (1992), Rudd (1995), etc.[1]. A less desirable feature of the isotonic regression technique is that the estimated function might not be smooth. To produce monotonic yet still smooth estimation results, one can add a kernel-based smoothing step with the isotonic regression. See, e.g., Mukerjee (1988), and Mammen (1991). Recently, Aït-Sahalia and Duarte (2003) proposed a similar two-step procedure to estimate option price function nonparametrically. In the first step, they adopt Dykstra's (1983) constrained least square algorithm to trim the data so that the estimates from the succeeding kernel smoothing step are guaranteed to be monotonic and convex. In this constrained least squares method, the numerical search is performed

[1]Another vast line in the literature is constrained smoothing splines. See, e.g., Yatchew and Bos (1997), etc. Also see a recent survey by Henderson and Parmeter (2009) for this method and other alternatives.

iteratively in a subset of the n-dimensional Euclidean space, where n is the sample size. This algorithm might be computationally intensive when the sample size is large. Consequently, it is practically useful to combine theory-imposed restrictions with a procedure (e.g., kernel smoothing method) which produces smooth functional estimates, meanwhile is able to reduce computational burden.

In this section, I explore the possibility of imposing shape restrictions in the local polynomial regression framework. I construct constrained local quadratic (CLQ) estimators specifically for the functions $m(X)$, $m'(X)$, and $m''(X)$. The proposed estimators can be viewed as a reweighted version of the corresponding standard local quadratic estimators and the weights are determined via empirical likelihood (EL) maximization. Empirical likelihood, proposed by Owen (1988), (1990), and (1991) is a nonparametric likelihood method, in contrast to the widely known parametric likelihood method. See Kitamura (2006) for a comprehensive survey of EL in econometrics. EL can be applied in both parametric and nonparametric models. In parametric estimation, a generalized empirical likelihood estimator is shown to have advantages, in terms of higher order asymptotic properties, against the GMM estimator (see Newey and Smith (2004)). The idea of parametric estimation via EL is to maximize a nonparametric likelihood ratio $\prod_{i=1}^{n}(np_i)$ between a probability measure $\{p_i : i = 1, \cdots, n\}$ given on the sample points and the empirical distribution $\{1/n, \cdots, 1/n\}$, subject to the moment conditions of interest. See Qin and Lawless (1994), and Kitamura et al. (2004).

EL can also be used in combination with nonparametric models. For a given nonparametric estimator, confidence intervals via EL has demonstrated advantages over asymptotic normality-based approaches. See Hall and Owen (1993), Chen (1996) for density function estimation; Chen and Qin (2000), Qin and Tsao (2005) for local linear estimators of conditional mean function; Cai (2002) for conditional distribution and regression quantiles; Xu (2009) for local linear estimators in continuous-time diffusion models. The common approach in this literature is to maximize the nonparametric likelihood ratio $\prod_{i=1}^{n}(np_i)$ subject to the EL-weighted estimating equations which can be viewed as counterparts of the moment conditions in parametric settings.

Following this line, I consider the EL profile $\{p_i : i = 1, \cdots, n\}$ embedded on a set of local quadratic estimators and I maximize $\prod_{i=1}^{n}(np_i)$ under the desired shape restrictions. If the restrictions are true for the underlying data generating process, then the EL profile asymptotically converges to $\{1/n, \cdots, 1/n\}$ as n goes to infinity. Hence the CLQ estimators have the same first order asymptotic distribution as the standard local quadratic estimators. My procedure offers estimation results that are smooth functions, and reduces the dimensions of numerical optimization from sample size n to the number of restrictions. Moreover, the procedure estimates the function $Y = m(X)$ and its first and second derivative simultaneously, so it is particularly useful when one is interested in estimating the derivatives.

When multiple nonparametric functions are jointly estimated, it is common for only some of the restrictions to be violated by the unconstrained estimator. By adjusting those violations to meet the constraints, my EL approach can meanwhile tune other functional estimates at the same location towards the corresponding true values. In addition, the procedure can be applied in more general situations when constraints on the function and its derivatives vary over locations, unlike the constant constraints considered in the existing literature.

Hall and Huang (2001) proposed an EL-based nonparametric regression approach to estimate a function, subject to monotonicity constraints. Under certain assumptions of the weight functions of the original estimator (kernel or local linear weights, etc.), they show the existence of a set of *location independent* EL weights which guarantee the reweighted estimator to be monotonic. Racine et al. (2009) extend Hall and Huang's (2001) approach to multivariate and multi-constraint cases. My study in this section is different from these two papers in several aspects. First, I consider the joint estimation of the regression function and its derivatives and investigate the asymptotic distribution of the EL weighted CLQ estimators, whereas the above mentioned two papers focus on the estimation of the regression function only. Second, my EL weights are *location dependent* so that we can accommodate constraints varying over the domain of X. Third, I allow constraints on the regression function as well as its derivatives, while the theoretical results in Racine et al. (2009) are not directly applicable to such a case where there are multiple constraints on the first and second derivatives with respect to the same explanatory variable X.

As an application, I use the EL-based CLQ estimator to investigate the nonparametric estimation of daily call option prices C as a function of strike prices X. As implied by finance theory, under the assumption of market completeness and no arbitrage opportunities, the price of a call option $C = C(X)$ must be a decreasing and convex function of the option's strike price X. These shape restrictions can be expressed as $C_{t,\tau}(X) \in \left[\max\left(0, S_t e^{-\delta_{t,\tau}\tau} - X e^{-r_{t,\tau}\tau}\right), S_t e^{-\delta_{t,\tau}\tau}\right]$, $C'_{t,\tau}(X) \in [-e^{-r_{t,\tau}\tau}, 0]$, and $C''_{t,\tau}(X) \in [0, \infty)$, where t is the current time, τ is the time-to-expiration, r is the risk free interest rate, and δ is the dividend yield of the underlying asset with price S_t. I estimate $C_{t,\tau}(X)$, $C'_{t,\tau}(X)$, and $C''_{t,\tau}(X)$ under these constraints and compare the results of my CLQ estimation with the results of standard local quadratic estimation. In a simulation study, I adopt the same simulation set-up as in Aït-Sahalia and Duarte (2003), and find that my results are comparable with theirs in this extremely small sample setting, whereas my procedure exhibits potential advantages such as convergent solution and fast computation when the sample size becomes larger. Last, I apply this method to estimate the S&P 500 index options in a typical trading day in May 2009.

The remainder of this section is organized as follows. Section 2.2 introduces the definition of EL for the local quadratic estimators, subject to inequality constraints. Then I show the equivalence of two saddle point problems from the EL formulation to ease the following asymptotic analysis. Next, Section 2.3 studies the asymptotic properties of the EL-based CLQ estimator. Section 2.4 applies the constrained estimation procedure to estimate option price function and the state-price density. The proofs are presented in the Appendix A.

2.2 Empirical Likelihood-Based Constrained Local Quadratic Regression

Suppose one observes a random sample $\{(X_i, Y_i) : i = 1, \cdots, n\}$ generated from a bivariate distribution. I shall denote the conditional mean function of Y given X by $m(x) = E(Y|X = x)$ and the conditional variance function by $\sigma^2(x) = Var(Y|X = x)$, then the nonparametric regression model under consideration is $Y = m(X) + \sigma(X)u$, where $E(u|X) = 0$ and $Var(u|X) = 1$. I shall also denote the marginal density of X by $f(\cdot)$. In this section I develop the empirical likelihood formulation in the context of a local quadratic regression model subject to inequality constraints.

2.2.1 The Local Quadratic Estimator

Because the empirical motivation of this study is to impose theory-motivated constraints on the estimators of the functions $m(\cdot)$, $m'(\cdot)$, and $m''(\cdot)$, I shall focus on the local quadratic regression model, which provides estimators for the three functions simultaneously. The local quadratic estimators can be derived from the following minimization problem

$$\min_{\beta_j(x):\, j=0,1,2} \sum_{i=1}^{n} \left[Y_i - \beta_0(x) - \beta_1(x)(X_i - x) - \beta_2(x)(X_i - x)^2\right]^2 K_i, \tag{2.1}$$

where $K_i = K((X_i - x)/h)$. Hereafter I shall slightly abuse the notation and use $(m_0(\cdot), m_1(\cdot), m_2(\cdot))^\intercal$ to denote $\left(m(\cdot), m^{(1)}(\cdot), m^{(2)}(\cdot)/2\right)^\intercal$, then the local quadratic estimator for $(m_0(x), m_1(x), m_2(x))^\intercal$ can be written as

$$\widehat{\beta}(x) = \left(\widehat{\beta}_0(x), \widehat{\beta}_1(x), \widehat{\beta}_2(x)\right)^\intercal,$$

where for $j = 0, 1, 2,$

$$\widehat{\beta}_j(x) = \frac{1}{h^j}\frac{\frac{1}{n}\sum_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum_{i=1}^{n} W_{0i}(x)},$$

and the corresponding weights are

$$W_{0i}(x) = \left[\left(s_2 s_4 - s_3^2\right) - \left(s_1 s_4 - s_2 s_3\right)\left(\frac{X_i - x}{h}\right) - \left(s_2^2 - s_1 s_3\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i,$$

$$W_{1i}(x) = \left[\left(s_2 s_3 - s_1 s_4\right) - \left(s_2^2 - s_0 s_4\right)\left(\frac{X_i - x}{h}\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i,$$

$$W_{2i}(x) = \left[\left(s_1 s_3 - s_2^2\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right) - \left(s_1^2 - s_0 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i,$$

where

$$s_j = \frac{1}{nh}\sum_{i=1}^{n}\left(\frac{X_i - x}{h}\right)^j K_i \text{ for } j = 0, 1, 2, 3, 4.$$

2.2.2 EL for the Local Quadratic Estimating Equations

In this subsection we construct the empirical likelihood formulation for the local quadratic regression model. Let $\{p_1, \cdots, p_n\}$ be a discrete probability distribution on the sample $\{(X_i, Y_i) : i = 1, \cdots, n\}$. That is, $\{p_1, \cdots, p_n\}$ is a set of nonnegative numbers adding to unity. At a location x in the domain of X, the profile empirical likelihood ratio at a set of candidate values $\beta(x) = (\beta_0(x), \beta_1(x), \beta_2(x))^{\mathsf{T}}$ of

$$E\left[\widehat{\beta}(x)\right] = \left(E\left[\widehat{\beta}_0(x)\right], E\left[\widehat{\beta}_1(x)\right], E\left[\widehat{\beta}_2(x)\right]\right)^{\mathsf{T}}$$

is defined as

$$L(\beta) = \max_{\{p_1, \cdots, p_n\}}\left\{\prod_{i=1}^{n} np_i \mid p_i \geqslant 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i U_i(\beta) = 0\right\}, \qquad (2.2)$$

where $U_i(\beta) = (U_{0i}(\beta), U_{1i}(\beta), U_{2i}(\beta))^\intercal$, $U_{ji}(\beta(x)) = W_{ji}(x)\left[Y_i - (X_i - x)^j \beta_j(x)\right]$. Hereafter, when it is clear we shall omit the explicit dependence of a variable on the location x for brevity of notations. The three equations

$$\sum\nolimits_{i=1}^{n} p_i U_i(\beta) = 0 \tag{2.3}$$

are labeled as *estimating equations* in the empirical likelihood literature.

Heuristically, if we take $\{p_1, \cdots, p_n\} = \{1/n, \cdots, 1/n\}$, then for $j = 0, 1, 2$, the estimating equations become

$$\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x)\left[Y_i - (X_i - x)^j \beta_j(x)\right] = 0,$$

which can be viewed as reformulations of the first order conditions of the weighted least squares problem (2.1). From the above equations, we can solve, for $j = 0, 1, 2$,

$$\beta_j(x) = \frac{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x)(X_i - x)^j} = \frac{1}{h^j}\frac{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x) Y_i}{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{0i}(x)}$$

by recognizing that

$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{0i}(x) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{1i}(x)\left(\frac{X_i - x}{h}\right) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{2i}(x)\left(\frac{X_i - x}{h}\right)^2. \tag{2.4}$$

That is, the candidate values $\beta_j(x)$ coincide with the local quadratic estimators $\widehat{\beta}_j(x)$. In general, the candidate values $\beta_j(x)$ are not fixed at $\widehat{\beta}_j(x)$, and the corresponding $\{p_1, \cdots, p_n\}$ are different from uniform weights $1/n$. As a digression on notation, we will reserve D_n for the common value in (2.4). That is, we denote

$$D_n = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{ji}(x)\left(\frac{X_i - x}{h}\right)^j = s_0\left(s_2 s_4 - s_3^2\right) - s_1\left(s_1 s_4 - s_2 s_3\right) - s_2\left(s_2^2 - s_1 s_3\right).$$

By using the log empirical likelihood ratio, we can modify the EL maximization problem (2.2) to be

$$l(\beta) = \max_{(p_1,\cdots,p_n)} \left\{ \sum\nolimits_{i=1}^{n} \log(np_i) \mid p_i \geqslant 0,\ \sum\nolimits_{i=1}^{n} p_i = 1,\ \sum\nolimits_{i=1}^{n} p_i U_i(\beta) = 0 \right\}. \tag{2.5}$$

Further, by introducing Lagrange multipliers $\lambda(\beta) = (\lambda_0(\beta), \lambda_1(\beta), \lambda_2(\beta))^{\intercal}$ for the estimating equations (2.3) respectively, we can form the Lagrangian as

$$\mathcal{L} = \sum\nolimits_{i=1}^{n} \log(np_i) - \gamma\left(\sum\nolimits_{i=1}^{n} p_i - 1\right) - n\lambda(\beta)^{\intercal} \sum\nolimits_{i=1}^{n} p_i U_i(\beta),$$

and solve for

$$p_i(\beta) = \frac{1}{n\left(1 + \lambda(\beta)^{\intercal} U_i(\beta)\right)}.$$

Now the log empirical likelihood ratio $l(\beta)$ can be expressed as

$$l(\beta) = \min_{\lambda\in\Lambda}\left[-\sum\nolimits_{i=1}^{n} \log\left(1 + \lambda(\beta)^{\intercal} U_i(\beta)\right)\right] = \max_{\lambda\in\Lambda} \sum\nolimits_{i=1}^{n} \log\left(1 + \lambda(\beta)^{\intercal} U_i(\beta)\right), \tag{2.6}$$

and

$$\widetilde{\lambda}(\beta) = \arg\max_{\lambda\in\Lambda} \sum\nolimits_{i=1}^{n} \log\left(1 + \lambda(\beta)^{\intercal} U_i(\beta)\right), \tag{2.7}$$

where

$$\Lambda = \left\{\lambda(\beta) \in \mathbb{R}^3 | 1 + \lambda(\beta)^{\intercal} U_i(\beta) \geqslant 1/n,\ i = 1, \cdots, n\right\}.$$

The domain Λ of $\lambda(\beta)$ is derived from $p_i \in [0,1]$ and it is needed to ensure that the arguments of the logarithm are strictly positive.

2.2.3 EL Formulation under Inequality Constraints

In this section I investigate the log empirical likelihood ratio (2.6) under inequality constraints. In the regression model $Y = m(X) + \sigma(X)u$, one can formulate shape restrictions imposed by economic theory as lower and upper bounds on the function

$m(\cdot)$ and its derivatives. More specifically, let $\underline{b}(x) = (\underline{b}_0(x), \underline{b}_1(x), \underline{b}_2(x))^\intercal$ and $\overline{b}(x) = \left(\overline{b}_0(x), \overline{b}_1(x), \overline{b}_2(x)\right)^\intercal$, then the restrictions can be expressed as

$$\begin{aligned} \underline{b}_0(x) &\leqslant m_0(x) \leqslant \overline{b}_0(x), \\ \underline{b}_1(x) &\leqslant m_1(x) \leqslant \overline{b}_1(x), \\ \underline{b}_2(x) &\leqslant m_2(x) \leqslant \overline{b}_2(x). \end{aligned}$$

For example, in the estimation of option price function and its derivatives, the shape restrictions are given by $\underline{b}(X) = \left(\max\left(0, S_t e^{-\delta\tau} - X e^{-r\tau}\right), -e^{-r\tau}, 0\right)^\intercal$, and $\overline{b}(X) = \left(S_t e^{-\delta\tau}, 0, \infty\right)^\intercal$. The goal is to accommodate these constraints in the nonparametric estimation of $m(\cdot)$ and its derivatives.

Because the log empirical likelihood ratio (2.6) depends on candidate values $\beta(x) = (\beta_0(x), \beta_1(x), \beta_2(x))^\intercal$, one can stack the above inequality constraints and impose them on the candidate values:

$$\underline{b}(x) \leqslant \beta(x) \leqslant \overline{b}(x). \tag{2.8}$$

Then (2.6) is modified as

$$\min_{\underline{b} \leqslant \beta \leqslant \overline{b}} l(\beta) = \min_{\underline{b} \leqslant \beta \leqslant \overline{b}} \max_{\lambda \in \Lambda} G_n(\beta, \lambda), \tag{2.9}$$

where

$$G_n(\beta, \lambda) = \sum_{i=1}^{n} \log\left(1 + \lambda(\beta)^\intercal U_i(\beta)\right).$$

This can be viewed as a saddle point problem, and let its solution be $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$.

Remark 2.2.1. *The objective function in (2.5) corresponds to* -1 *times the Kullback–Leibler distance between the probability distribution* $\{p_1, \cdots, p_n\}$ *and the empirical distribution* $\{1/n, \cdots, 1/n\}$*. Thus maximizing the log empirical likelihood ratio* $l(\beta)$ *can be interpreted as minimizing the Kullback–Leibler distance between* $\{p_1, \cdots, p_n\}$

and $\{1/n, \cdots, 1/n\}$*. On the other hand, it is easy to verify that (2.5) attains its global maximum at* $\{1/n, \cdots, 1/n\}$*, corresponding to the candidate values* $\beta(x)$ *being equal to the standard local quadratic estimators* $\widehat{\beta}(x)$*. Indeed* $\widehat{\beta}(x)$ *are the minimizers of* $l(\beta)$ *without imposing the inequality constraints (2.8). Thus* $\widetilde{\beta}(x)$*, as the minimizers of* $l(\beta)$ *in (2.9), are designed to minimally adjust the standard local quadratic estimators such that the inequality constraints (2.8) are satisfied.*

Remark 2.2.2. *The empirical likelihood formulated so far is for* $E\left[\widehat{\beta}(x)\right] = m(x) +$ *bias, rather than for* $m(x)$*. This point has been observed in previous studies of empirical likelihood-based inference for nonparametric models (Chen and Qin (2000), Qin and Tsao (2005), etc.). To reduce the bias, one can use an under smoothing bandwidth condition* $nh^7 \to 0$*, as recommended in Chen and Qin (2000). I will discuss this point further in the asymptotic analysis.*

To facilitate the asymptotic analysis of the CLQ estimator $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$, I need to introduce another saddle point problem

$$\min_{\beta} \max_{\lambda \in \Lambda, \nu \in \mathbb{R}_+^6} G_n^*(\beta, \lambda, \nu) \tag{2.10}$$

where

$$G_n^*(\beta, \lambda, \nu) = G_n(\beta, \lambda) + n\underline{\nu}^\intercal(\underline{b} - \beta) + n\overline{\nu}^\intercal\left(\beta - \overline{b}\right)$$

and $\nu = (\underline{\nu}^\intercal, \overline{\nu}^\intercal)^\intercal$ is a set of Lagrangian multipliers for the inequalities

$$\underline{b} - \beta \leqslant 0,$$
$$\beta - \overline{b} \leqslant 0.$$

The following lemma states that the two problems (2.9) and (2.10) have the same saddle points.

Lemma 2.2.1. $\left(\widetilde{\beta},\widetilde{\lambda}\right)$ *is a saddle point of* $G_n(\beta,\lambda)$ *and solve (2.9) if and only if* $\left(\widetilde{\beta},\widetilde{\lambda},\widetilde{\nu}\right)$ *is a saddle point of* $G_n^*(\beta,\lambda,\nu)$ *that solves (2.10), where for* $j=0,1,2,$

$$\widetilde{\underline{\nu}}_j\left(x,\widetilde{\beta}\right)=\begin{cases}-\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{\lambda}_j\left(\widetilde{\beta}\right)W_{ji}(x)(X_i-x)^j}{1+\widetilde{\lambda}\left(\widetilde{\beta}\right)^{\intercal}U_i\left(\widetilde{\beta}\right)} & if\ \underline{b}_j-\widetilde{\beta}_j=0,\\ 0 & if\ \underline{b}_j-\widetilde{\beta}_j<0,\end{cases}$$
$$\widetilde{\overline{\nu}}_j\left(x,\widetilde{\beta}\right)=\begin{cases}\frac{1}{n}\sum_{i=1}^n\frac{\widetilde{\lambda}_j\left(\widetilde{\beta}\right)W_{ji}(x)(X_i-x)^j}{1+\widetilde{\lambda}\left(\widetilde{\beta}\right)^{\intercal}U_i\left(\widetilde{\beta}\right)} & if\ \widetilde{\beta}_j-\overline{b}_j=0,\\ 0 & if\ \widetilde{\beta}_j-\overline{b}_j<0.\end{cases}\tag{2.11}$$

Remark 2.2.3. *In practical implementation, one can program according to the saddle point problem (2.9). Essentially, at each evaluating location* x*, searching for* $\widetilde{\beta}(x)$ *is performed in* $\left[\underline{b}(x),\overline{b}(x)\right]$*. This can be viewed as an outer loop. While for each candidate* $\beta(x)$*, searching for* $\lambda(\beta)$ *is done in the inner loop via maximizing* $G_n(\beta,\lambda)$*. Plugging* p_i *in the EL weighted estimating equations* $\sum_{i=1}^n p_iU_i(\beta)=0$*, one can obtain*

$$\frac{1}{n}\sum_{i=1}^n\frac{U_i(\beta)}{1+\lambda(\beta)^{\intercal}U_i(\beta)}=0,$$

which can also be viewed as the first order conditions for $\lambda(\beta)$ *in (2.6) divided by* n*. In the inner loop, given a candidate value of* $\beta(x)$*, one can equivalently solve for* $\lambda(\beta)$ *from these first order conditions.*

Remark 2.2.4. *The saddle point problem (2.10) is useful in the following asymptotic analysis of the CLQ estimators* $\widetilde{\beta}(x)$*. Since (2.9) and (2.10) are equivalent, I shall use the same notation* $l(\beta)$ *in the remaining of this section. That is, I denote*

$$\begin{aligned}l(\beta)&=\max_{\lambda\in\Lambda,\nu\in\mathbb{R}_+^6}G_n^*(\beta,\lambda,\nu)\\&=\max_{\lambda\in\Lambda,\nu\in\mathbb{R}_+^6}\sum_{i=1}^n\log\left(1+\lambda(\beta)^{\intercal}U_i(\beta)\right)+n\underline{\nu}^{\intercal}\left(\underline{b}-\beta\right)+n\overline{\nu}^{\intercal}\left(\beta-\overline{b}\right).\end{aligned}$$

2.3 Asymptotic Analysis of the Constrained Local Quadratic Estimators

In this section, first I show that, under proper regularity conditions, $\lambda(m) = (\lambda_0(m), \lambda_1(m), \lambda_2(m))^\intercal$ converges to zero as $n \to \infty$. This result is presented in Theorem 2.3.1. Then I show in Theorem 2.3.2 that the CLQ estimators $\widetilde{\beta}(x)$ and the standard local quadratic estimators $\widehat{\beta}(x)$ are asymptotically equivalent, that is, $\widetilde{\beta}(x)$ and $\widehat{\beta}(x)$ have the same first-order asymptotic distribution. As a starting point, I shall list the following assumptions:

Assumption 2.3.1. *The kernel function $K(\cdot)$ is a symmetric, bounded density function compactly supported on $[-1, 1]$.*

Assumption 2.3.2. *$f(\cdot)$ and $\sigma(\cdot)$ have continuous derivatives up to the second order in a neighborhood of x, and both $f(x) > 0$ and $\sigma(x) > 0$. Also $m(\cdot)$ has continuous derivatives up to the third order in a neighborhood of x.*

Assumption 2.3.3. *$h \to 0$, $nh \to \infty$, and $nh^7 \to 0$ as $n \to \infty$.*

Lemma 2.3.1. *Under Assumptions 2.3.1, 2.3.2, and 2.3.3, as $n \to \infty$, we have the asymptotic distribution*

$$\sqrt{nh}\left(\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m) - \frac{h^3}{6} m^{(3)}(x) f^3(x) B_U\right) \xrightarrow{d} N\left(0, \sigma^2(x) f^5(x) V_U\right), \quad (2.12)$$

where

$$B_U = \begin{pmatrix} 0 \\ \mu_4^2 - \mu_2^2 \mu_4 \\ 0 \end{pmatrix}, \; V_U = \begin{pmatrix} \omega_0 & 0 & \omega_2 \\ 0 & \omega_3 & 0 \\ \omega_2 & 0 & \omega_5 \end{pmatrix},$$

$$\omega_0 = \mu_2^2\left(\mu_4^2 \nu_0 - 2\mu_2\mu_4\nu_2 + \mu_2^2\nu_4\right),$$

$$\omega_2 = -\mu_2^2\left(\mu_2\mu_4\nu_0 - \left(\mu_4 + \mu_2^2\right)\nu_2 + \mu_2\nu_4\right),$$

$$\omega_3 = \left(\mu_4 - \mu_2^2\right)^2 \nu_2,$$

$$\omega_5 = \mu_2^2\left(\mu_2^2\nu_0 - 2\mu_2\nu_2 + \nu_4\right).$$

Remark 2.3.1. *To investigate the asymptotic behavior of the EL-based CLQ estimators, first I need to find the asymptotic distribution of the estimating equations (2.3). Lemma 2.3.1 presents the asymptotic distribution of (2.3) evaluated at the set of true values* $m(x) = (m_0(x), m_1(x), m_2(x))$. *This result can be derived from the asymptotic distribution of the local quadratic estimators* $\widehat{\beta}(x)$ *because (2.3) can be viewed as a reformulation of* $\widehat{\beta}(x) - m(x)$. *Essentially, from Lemma 2.3.1 we have*

$$\frac{1}{nh}\sum_{i=1}^{n} U_i(m) = O_p\left((nh)^{-1/2} + h^3\right).$$

Remark 2.3.2. *Among the three terms in (2.12), the leading bias of* $\frac{1}{nh}\sum_{i=1}^{n} U_{0i}(m)$ *and* $\frac{1}{nh}\sum_{i=1}^{n} U_{2i}(m)$ *are actually zeros because of the symmetry of the kernel* $K(\cdot)$. *As suggested by Chen and Qin (2000), I use an under smoothing condition* $nh^7 \to 0$ *to reduce the bias of* $\frac{1}{nh}\sum_{i=1}^{n} U_{1i}(m)$. *With this condition, one can still use the optimal bandwidth* $h = O\left(n^{-1/9}\right)$ *for the estimation of the regression function and the second derivative.*

Lemma 2.3.2. *Under Assumptions 2.3.1, 2.3.2, and 2.3.3, we have*

$$\frac{1}{nh}\sum_{i=1}^{n} U_i(m)\, U_i(m)^{\intercal} = \Omega_U + o_p(1),$$

where

$$\Omega_U = f^3(x)\begin{pmatrix} \sigma^2(x)\,\omega_0 & \sigma^2(x)\,\omega_1 & \sigma^2(x)\,\omega_2 \\ \sigma^2(x)\,\omega_1 & \left[\sigma^2(x) + m^2(x)\right]\omega_3 & \left[\sigma^2(x) + m^2(x)\right]\omega_4 \\ \sigma^2(x)\,\omega_2 & \left[\sigma^2(x) + m^2(x)\right]\omega_4 & \left[\sigma^2(x) + m^2(x)\right]\omega_5 \end{pmatrix}.$$

Theorem 2.3.1. *Assume that* $E|Y_i|^s < \infty$ *for some* $s > 2$ *and Assumptions 2.3.1, 2.3.2, and 2.3.3 hold. Then*

$$\lambda(m) = O_p\left((nh)^{-1/2} + h^3\right) = o_p\left(n^{-3/7}\right),$$

also

$$\lambda(m) = \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\, U_i(m)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\right] + o_p\left((nh)^{-1/2} + h^3\right).$$

Remark 2.3.3. *From Theorem 2.3.1, $p_i(m) = n^{-1}\left(1 + \lambda(m)^{\intercal} U_i(m)\right)^{-1}$ converges to $1/n$ with increasing sample size and proper selected bandwidth. Hence the EL-based CLQ estimators*

$$\widetilde{\beta}_j(x) = \frac{1}{h^j}\frac{\sum\nolimits_{i=1}^{n} p_i W_{ji}(x)\, Y_i}{\sum\nolimits_{i=1}^{n} p_i W_{0i}(x)} \quad (j = 0, 1, 2)$$

converge to the unconstrained local quadratic estimators

$$\widehat{\beta}_j(x) = \frac{1}{h^j}\frac{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{ji}(x)\, Y_i}{\frac{1}{n}\sum\nolimits_{i=1}^{n} W_{0i}(x)} \quad (j = 0, 1, 2)$$

Lemma 2.3.3. *Assume that Assumptions 2.3.1, 2.3.2, and 2.3.3 hold, further assume that $nh^5 \to \infty$ as $n \to \infty$. Then $G_n^*(\beta, \lambda, \nu)$ attains its saddle point at $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ where $\widetilde{\beta} = \left(\widetilde{\beta}_0, \widetilde{\beta}_1, \widetilde{\beta}_2\right)$ is such that $\left|\widetilde{\beta}_j(x) - m_j(x)\right| \leqslant h^{2-j}$; $\widetilde{\lambda} = \lambda\left(\widetilde{\beta}\right)$ is given by (2.7); and $\widetilde{\nu}$ is given by (2.11). Further, $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ satisfies*

$$g_{1n}\left(\widetilde{\beta}, \widetilde{\lambda}\right) = 0,\ g_{1n}\left(\widetilde{\beta}, \widetilde{\lambda}\right) = 0,$$

where

$$g_{1n}(\beta, \lambda) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(\beta)}{1 + \lambda(\beta)^{\intercal} U_i(\beta)},$$
$$g_{2n}(\beta, \lambda) = \frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{D_i(x)\,\lambda(\beta)}{1 + \lambda(\beta)^{\intercal} U_i(\beta)},$$

and $D_i(x)$ is a 3×3 matrix such that $D_i(x) = diag\left\{-W_{ji}(x)\left((X_i - x)/h\right)^j\right\}$.

Remark 2.3.4. *Lemma 2.3.3 shows the existence of a saddle point of* $G_n^*(\beta, \lambda, \nu)$ *in the interior of a (asymptotically shrinking) neighborhood of* $m(x)$,

$$\left\{\beta(x) : |\beta_j(x) - m_j(x)| \leqslant h^{2-j},\ j = 0, 1, 2\right\}. \tag{2.13}$$

This is achieved by establishing a lower bound of $l(\beta)$ *out of (2.13) and then it is shown that this lower bound is of a larger stochastic order than* $l(m)$.

Theorem 2.3.2. *Suppose that the assumptions of Theorem 2.3.1 hold. Also assume that* $nh^5 \to \infty$ *as* $n \to \infty$. *Then for* $j = 0, 1, 2$, *the EL-based CLQ estimator*

$$\widetilde{\beta}_j(x) = \widehat{\beta}_j(x) + o_p\left(\left(nh^{1+2j}\right)^{-1/2} + h^{3-j}\right),$$

where for each j, $\widehat{\beta}_j(x)$ *is the corresponding local quadratic estimator. As* $n \to \infty$, *the asymptotic distribution of* $\widetilde{\beta}(x)$ *is given by*

$$diag\left(\sqrt{nh^{1+2j}}\right)\left(\widetilde{\beta}(x) - m(x) - \frac{h^2}{6} m^{(3)}(x) B_{\widetilde{\beta}}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)}{f(x)} V_{\widetilde{\beta}}\right), \tag{2.14}$$

where

$$B_{\widetilde{\beta}} = \begin{pmatrix} 0 \\ \mu_4/\mu_2 \\ 0 \end{pmatrix},\ V_{\widetilde{\beta}} = \frac{V_U}{\mu_2^2\left(\mu_4 - \mu_2^2\right)^2}.$$

Remark 2.3.5. *Theorem 2.3.2 shows that the EL-based CLQ estimators* $\widetilde{\beta}(x)$ *and the standard local quadratic estimators* $\widehat{\beta}(x)$ *have the same asymptotic distribution up to the first order. This result is naturally expected because, as the sample size increases,* $\widehat{\beta}(x)$ *converges to the true function values which are in the bounded region* $\left[\underline{b}(x), \overline{b}(x)\right]$, *hence the inequality constraints become unbinding* ($\widetilde{\nu} \xrightarrow{p} 0$) *and* $\widetilde{\beta}(x)$ *and* $\widehat{\beta}(x)$ *are asymptotically first-order equivalent.*

2.4 Application: Option Pricing Function Estimation

2.4.1 Restrictions Imposed by Option Pricing Theory

To show the usefulness of the CLQ estimation procedure proposed in this section, I estimate the daily option pricing function and the state-price density function by incorporating various shape restrictions. In summary, given market completeness and no arbitrage assumptions, implication from financial market theory suggests that the price of a call option, as a function of its strike price, must be decreasing and convex.

Let us consider an European call option with price C_t at time t, and expiration time T. Denote by $\tau = T - t$ the maturity, and X the strike price. Also denote by $r_{t,\tau}$ the risk free interest rate and $\delta_{t,\tau}$ the dividend yield of the underlying asset with price S_t. Using these notations, one can write the call option price C_t by

$$C\left(X, S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right) = e^{-r_{t,\tau}\tau} \int_0^{+\infty} \max\left(0, S_T - X\right) f^*\left(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right) dS_T,$$

where $f^*\left(S_T | S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right)$ is the state-price density (SPD), also called the risk-neutral density. The SPD will be denoted as $f^*\left(S_T\right)$ for brevity in what follows. Asset pricing theory imposes no arbitrage bounds for the price function as

$$\max\left(0, S_t e^{-\delta_{t,\tau}\tau} - X e^{-r_{t,\tau}\tau}\right) \leqslant C_{t,\tau}\left(X\right) \leqslant S_t e^{-\delta_{t,\tau}\tau}, \tag{2.15}$$

where $C_{t,\tau}\left(X\right)$ is used to denote $C\left(X, S_t, \tau, r_{t,\tau}, \delta_{t,\tau}\right)$ since the main concern here is the call option price C as a function of X. For the first derivative

$$\frac{\partial C}{\partial X} = -e^{-r_{t,\tau}\tau} \int_X^{+\infty} f^*\left(S_T\right) dS_T,$$

the no-arbitrage assumption requires C to be a decreasing function of X, so the first derivative should be negative; also the first derivative should be larger than $-e^{-r_{t,\tau}\tau}$ because it is the integration of a scaled SPD. Thus we have

$$-e^{-r_{t,\tau}\tau} \leqslant C'_{t,\tau}(X) \leqslant 0 \tag{2.16}$$

from the positivity and integrability to one of the SPD. The second derivative is

$$C''_{t,\tau}(X) = e^{-r_{t,\tau}\tau} f^*(X) \geqslant 0 \tag{2.17}$$

since the SPD must be positive.

Given the data (X_i, C_i) recorded at time t (typically in one trading day) with the same maturity τ, our objective is to estimate $C_{t,\tau}(X)$, $C'_{t,\tau}(X)$, and $C''_{t,\tau}(X)$ under constraints (2.15), (2.16), and (2.17).

2.4.2 Monte-Carlo Simulation

To compare the performance of the EL based CLQ estimation method with other existing approaches, I adopt the simulation setup as that in Aït-Sahalia and Duarte (2003). Specifically, the true call option price function is assumed to be parametric as in the Black-Scholes/Merton model

$$C_{BS}(X, F_{t,\tau}, \tau, r_{t,\tau}, \sigma) = e^{-r_{t,\tau}\tau}\left[F_{t,\tau}\Phi(d_1) - X\Phi(d_2)\right],$$

where $F_{t,\tau} = S_t e^{(r_{t,\tau}-\delta_{t,\tau})\tau}$ is the forward price of the underlying asset at time t and

$$d_1 = \frac{\log(F_{t,\tau}/X)}{\sigma\sqrt{\tau}} + \frac{\sigma\sqrt{\tau}}{2},\ d_2 = \frac{\log(F_{t,\tau}/X)}{\sigma\sqrt{\tau}} - \frac{\sigma\sqrt{\tau}}{2},$$

and $\sigma = \sigma\left(X/F_{t,\tau}, \tau\right)$ is the volatility parameter. To generate data for simulation, I calibrate parameter values from real observations of S&P 500 index options on May 13, 1999. The parameter values and domain of strike prices are set as

$$\begin{aligned} S_t &= 1365, & \tau &= 30/252, \\ r_{t,\tau} &= 4.5\%, & X_i &\in [1000, 1700], \\ \delta_{t,\tau} &= 2.5\%, & \sigma_i &= -X_i/140 + 432/35. \end{aligned}$$

In the first simulation, the strike prices X are equally spaced between 1000 and 1700 with a sample size of 25. That is, in each sample there are 25 distinct strike prices and each of them corresponds to one call option price. In the second simulation, I generate 10 call option prices for each distinct strike price, so the sample size is 250.[2] To generate option prices, in the first simulation ($n = 25$), I add uniform noise to the true option price function, which ranges from 3% of the true price value for deep in the money options ($X = 1000$) to 18% for deep out of the money options ($X = 1700$). I double the noise size in the second simulation ($n = 250$).[3] I use the Epanechnikov kernel in the local quadratic estimation and adopt a rule-of-thumb bandwidth as in Fan and Mancini (2009). In each simulation experiment, I generate and estimate 1000 samples and show the average, 5%, and 95% quantiles as confidence bands in each graph.

For samples with 25 observations in the first simulation, the estimation results are shown in Figure 2.1. The unconstrained estimates are presented in the left column and the constrained estimates in the left column. The sample size in this simulation is tiny, so the unconstrained local quadratic estimators, especially the estimators for first and second derivatives, $\widehat{C'}(X)$ and $\widehat{C''}(X)$, perform poorly and violate the constraints frequently. Although difficult to tell in the graph, the estimator for the

[2] This is similar with the simulation setup in Yatchew and Hardle (2006).

[3] If I use the same noise design in the second simulation, because of the larger sample size ($n = 250$), the unconstrained local quadratic estimates will violate constrains less frequently so the constrained and unconstrained estimation results will be indistinguishable.

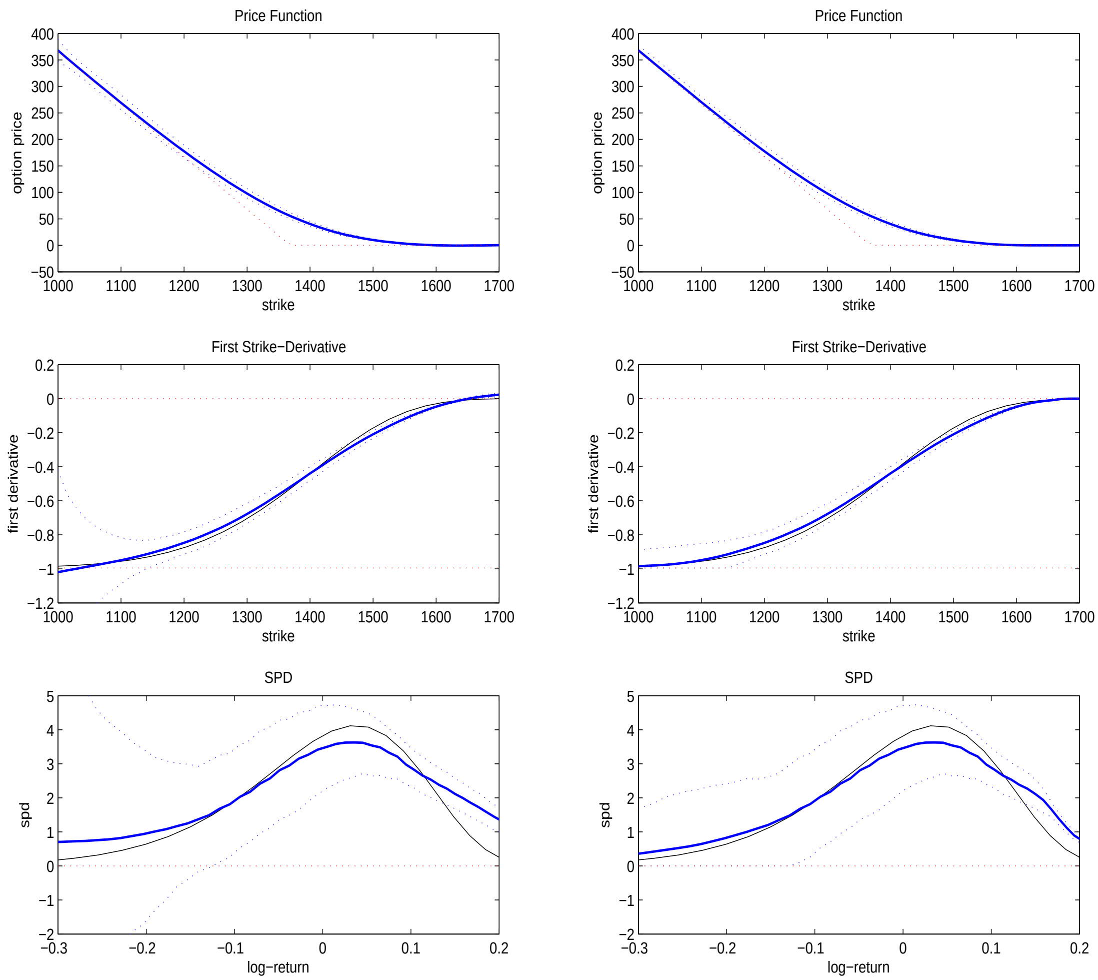

Fig. 2.1. Simulation results for $n = 25$

Left column from top to bottom: *Unconstrained estimates* $\widehat{C'}(X)$, $\widehat{C''}(X)$, and $e^{r_{t,\tau}\tau}\widehat{C''}(X)$. Right column from top to bottom: *Constrained estimates* $\widetilde{C'}(X)$, $\widetilde{C''}(X)$, and $e^{r_{t,\tau}\tau}\widetilde{C''}(X)$. Solid black line: *True function*. Solid blue line: *Average estimate*. Dot blue line: *95% confidence band*. Dot red line: *Constraints*.

option price, $\widehat{C}(X)$, also violates the lower bound when the strike price is low for deep in the money options. This violation of constraint can be adjusted in my

EL-based CLQ estimator, while not in Ait-Sahalia and Duarte (2003). Turning to the constrained estimation by the EL-based procedure, it is obvious that all three estimates, $\widetilde{C}(X)$, $\widetilde{C'}(X)$, and $\widetilde{C''}(X)$, always satisfy the constraints, and the estimates for first and second derivatives have smaller confidence bands in both of the boundary areas of the domain of X. An interesting finding is that, by correcting the violation of constraints in the first derivative estimate, the EL-based procedure also adjusts the second derivative estimate towards to its true function in corresponding boundary areas, although the unconstrained estimate itself, $\widehat{C''}(X)$, may not violate its nonnegative lower bound.[4]

As shown in Figure 2.2, in the second simulation with 250 observations, the unconstrained local quadratic estimators perform better than they did in the previous small sample design, despite of the doubled noise size. With a sample size as large as 250, the unconstrained estimate ($\widehat{C}(X)$) of the option price function and its true value are very close. But for the estimation of derivatives, the unconstrained estimators ($\widehat{C'}(X)$ and $\widehat{C''}(X)$) still violate the constraints when the strike price is very low or very high. In comparison, the EL-based constrained estimators ($\widetilde{C'}(X)$, and $\widetilde{C''}(X)$) are strictly within the constraints and have much narrower confidence bands, specially, in the left boundary area.

Last, I conduct a comparison of the integrated mean squared errors (IMSE) from constrained and unconstrained estimation in Figure 2.3. I focus on the first simulation design with sample size 25. The plots show that the IMSE's are much lower for the constrained estimators in all three functional estimations. Also a U-shaped IMSE curve can be detected in all three cases, showing that there exists an optimal bandwidth minimizing the IMSE.

[4]Note that the 5% quantile of $\widetilde{C'}(X)$ corresponds to the 95% quantile of $\widetilde{C''}(X)$ and vice versa.

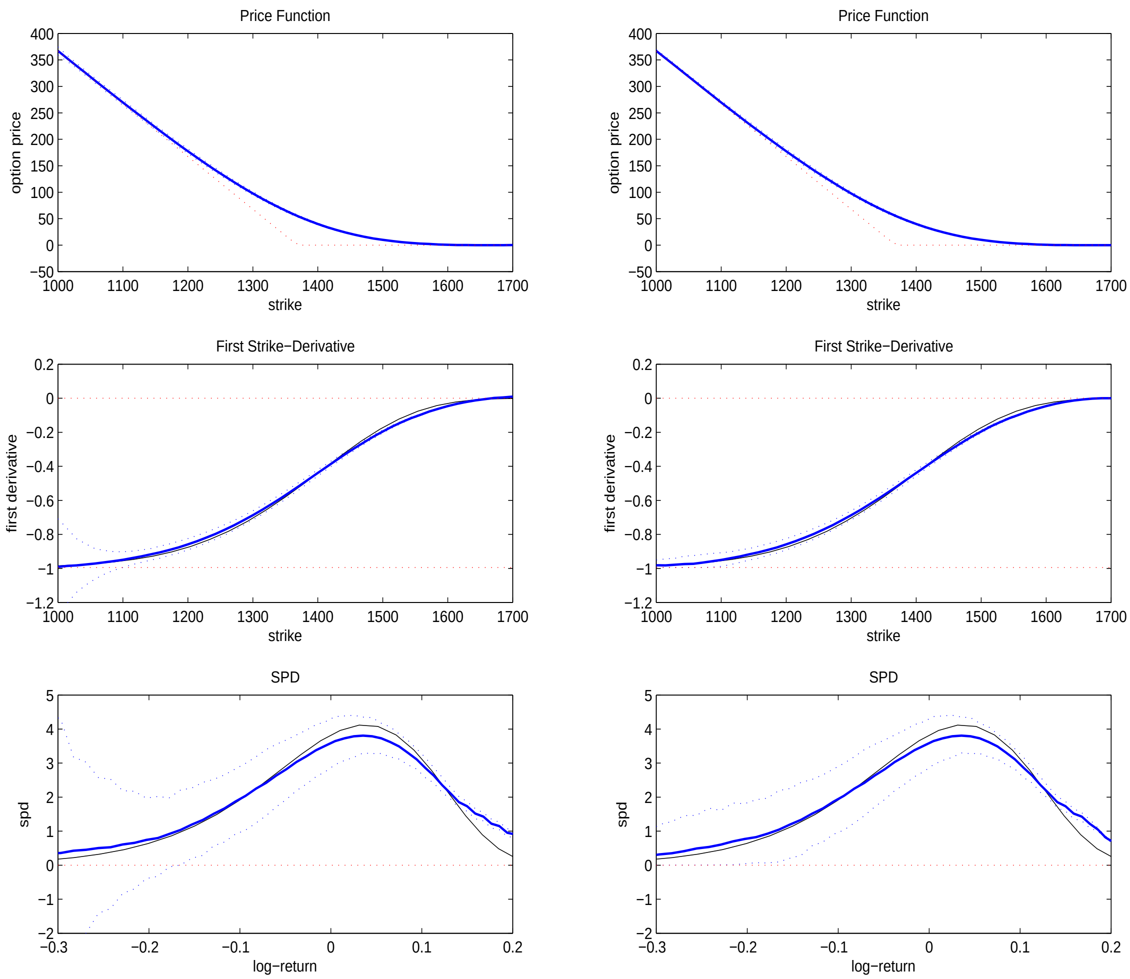

Fig. 2.2. Simulation results for $n = 250$

Left column from top to bottom: *Unconstrained estimates* $\widehat{C'}(X)$, $\widehat{C''}(X)$, and $e^{r_{t,\tau}\tau}\widehat{C''}(X)$. Right column from top to bottom: *Constrained estimates* $\widetilde{C'}(X)$, $\widetilde{C''}(X)$, and $e^{r_{t,\tau}\tau}\widetilde{C''}(X)$. Solid black line: *True function*. Solid blue line: *Average estimate*. Dot blue line: *95% confidence band*. Dot red line: *Constraints*.

2.4.3 Empirical Analysis

To investigate the empirical performance of my EL-based CLQ estimators, I apply this method to S&P 500 index (symbol SPX) based options and estimate the option

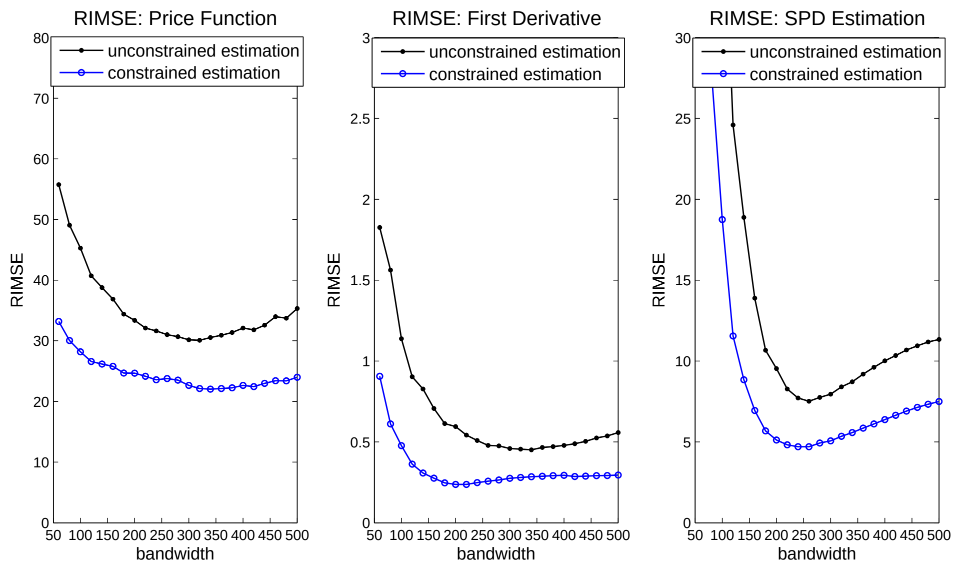

Fig. 2.3. Root of integrated mean squared errors for different bandwidths

price function and the state price density (a scaled second derivative of the option price function) using real data. I focus on closing prices of European call options on the S&P 500 index. The SPX index option is one of the most actively traded options and has been studied extensively in empirical option pricing literature. The data are downloaded from *OptionMetrics*. I collect options on May 18, 2009 for a maturity of 61 days corresponding to the expiration on July 18, 2009. Following Aït-Sahalia and Lo (1998), Fan and Mancini (2009), I use the bid-ask average of closing price as the option price, and delete less liquid options with implied volatility larger than 70%, or price less than or equal to 0.125. Finally I obtain a sample of 81 call option prices with strike prices ranging from 715 to 1150. The closing spot price of the S&P 500 index on that day was 909.71, and the risk free interest rate for the 2-month maturity was 0.51%. The dividend yield is retrieved from the put-call parity. Figure 2.4 presents the estimation results on this data set of daily cross-sectional option prices. From the results we can find that the unconstrained estimate of the

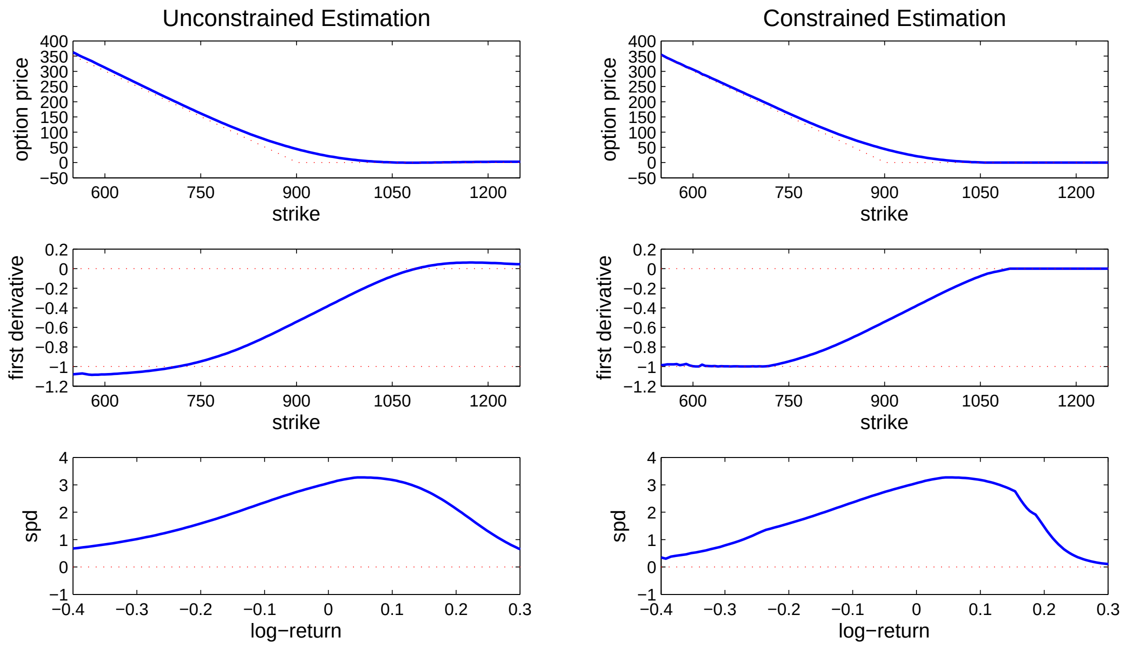

Fig. 2.4. Estimation results of S&P 500 options, July expiration on May 18, 2009

Left column from top to bottom: *Unconstrained estimates* $\widehat{C'}(X)$, $\widehat{C''}(X)$, and $e^{r_{t,\tau}\tau}\widehat{C''}(X)$. Right column from top to bottom: *Constrained estimates* $\widetilde{C'}(X)$, $\widetilde{C''}(X)$, and $e^{r_{t,\tau}\tau}\widetilde{C''}(X)$. Solid blue line: *Average estimate*. Dot red line: *Constraints*.

first derivative significantly violates the constraints at both the in-the-money and the out-of-the-money areas. In contrast, the constrained estimate of this function is bounded in both areas.

3. THE CUSUM OF SQUARES TEST FOR VOLATILITY CHANGE IN A TIME-VARYING COEFFICIENT MODEL

3.1 Introduction

Detecting structural changes in volatility is important in many aspects such as volatility forecast, risk management, etc. Particularly, smooth changes in unconditional volatility has been viewed as a long-run component of the volatility process (Engle and Rangel, 2008, Engle et al., 2009). In the theoretical research on structural change tests in volatility, a large body of the literature has been devoted to the Cumulative Sum (CUSUM) of Squares (hereafter CUSQ) test: Inclán and Tiao (1994), Deng and Perron (2008b), Cavaliere and Taylor (2008), Xu (2008, 2012), etc. Besides, Sensier and van Dijk (2004) applied the CUSQ test on 214 US macroeconomic variables and found 80% of the series have structural breaks. Rapach and Strauss (2008) investigated the empirical consequence of structural breaks for GARCH models of exchange rate volatility. They used the CUSQ test on different bilateral US dollar exchange rate return series, and found seven out of eight containing structural changes.

However, many theoretical and empirical studies of the CUSQ test focus on demeaned data, implicitly viewing the time series has an invariant mean. While our concern is, structural changes in volatility often co-occur with changes in the level. So in this section I reconsider the CUSQ test for structural change in volatility within a nonparametric time-varying coefficient time series model, for example, Cai (2007). The new CUSQ test is based on nonparametric residuals from the local linear estimation of the time-varying coefficient model. By doing this I am able to obtain a test which is robust to potential structural changes in the mean. I provide asymptotic theory to show that the new CUSQ test has standard null distribution and diverges at standard rate under the alternatives. Compared with a test based on least squares residuals, the proposed new CUSQ test enjoys correct size and good power properties. This is because, by estimating the model nonparametrically, we can circumvent the size distortion problem incurred by inconsistent mean estimation. Several simulation experiments and empirical applications are used to investigate the size and power properties.

The remainder of this section is organized as follows. In section 3.2, first, I introduce the time-varying coefficient regression model and briefly review the local linear estimation method; then I define the CUSQ test based on local linear residuals. In section 3.3 I provide limit theory for our test as well as the test based on least squares residuals. Simulation studies are presented in section 3.4. Several empirical examples are given in section 3.5.

3.2 The Model and the CUSQ Test

3.2.1 The Time-varying Coefficient Model

As a general framework for the proposed CUSQ test, first I introduce the time-varying coefficient model for time series data. I consider the model specification as in Cai(2007). For $i = 1, \cdots, n$, suppose one has observed time series data (y_i, X_i) from the regression model:

$$y_i = X_i' \beta(z_i) + u_i, \tag{3.1}$$

where $X_i = (x_{i1}, \cdots, x_{id})'$ with $x_{i1} = 1$, and $\beta(z_i) = (\beta_1(z_i), \cdots, \beta_d(z_i))'$, $z_i = i/n$. The model is featured with time-varying coefficients in the sense that, for each $k = 1, \cdots, d$, $\beta_k(z)$ is a deterministic function on the (rescaled) time domain $(0, 1]$. In the case of $d = 1$, the model (3.1) reduces to a time trend model with the time trend left as an unknown function.

The focus of interest is testing the hypothesis that the unconditional variance of error terms $\{u_i\}$ is a constant over time. To serve this interest, I specify the error term u_i as $u_i = \sigma_i \varepsilon_i$, where $\sigma_i^2 = \sigma^2(z_i)$ is a deterministic function on $(0, 1]$, ε_i is such that $E(\varepsilon_i | X_i) = 0$, $E(\varepsilon_i^2 | X_i) = s^2(X_i)$, $E(\varepsilon_i^2) = 1$. Specifically, here σ_i accounts for the (unconditional) time varying volatility and ε_i accounts for serial correlation and conditional heteroscedasticity. I also assume that $\sigma^2(z)$ is a positive càdlàg function and at least twice differentiable except at a finite number of points of discontinuity, with the second derivative function satisfying a (uniform) first-order Lipschitz condition. Further, I assume that $\{(\varepsilon_i, X_i)\}$ is strictly stationary β-mixing but $\{\varepsilon_i\}$ and $\{X_i\}$ may not be independent. Note that the model specification is largely adopted from Cai (2007), except for the error terms' variance, where the (unconditional) time dependence is explicitly separated from the conditional heteroscedasticity. Then the test hypothesis can be formally stated as

$$\begin{aligned} &H_0 : \sigma_i^2 = \sigma^2, \\ &H_1 : \sigma_i^2 = \sigma^2(z_i), \ z_i = i/n \in (0,1]. \end{aligned}$$

Because $\{u_i\}$ is not observed directly, to test the above hypothesis one must construct a test based on estimated residuals from the model (3.1). Given observed data $\{(y_i, X_i) : i = 1, \cdots, n\}$, one can use the local linear method to estimate $\beta(\cdot)$ nonparametrically. The local linear estimators $\widehat{\beta}(\cdot)$ can be written as

$$\widehat{\beta}(z) = \left(S_0(z) - S_1(t)\, S_2(z)^{-1} S_1(z)\right)^{-1} \left(T_0(z) - S_1(z)\, S_2(t)^{-1} T_1(z)\right),$$

where for $k = 0, 1, 2$,

$$\begin{aligned} S_k(z) &= n^{-1} \sum\nolimits_{i=1}^{n} (z_i - z)^k X_i X_i' K_h(z_i - z), \\ T_k(z) &= n^{-1} \sum\nolimits_{i=1}^{n} (z_i - z)^k X_i y_i K_h(z_i - z), \end{aligned}$$

and $K_h(u) = K(u/h)/h$, where $K(\cdot)$ is a kernel function, $h = h_n > 0$ is a bandwidth satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

3.2.2 The CUSUM of Squares Test Statistic

I use the nonparametric residuals $\widehat{u}_i = y_i - X_i'\widehat{\beta}(z_i)$ from the local linear estimation to construct a Cumulative Sums of Squares (CUSQ) test statistic. Following Xu (2012), I consider a slightly modified CUSQ test statistic given by

$$\widehat{Q} = \max_{1\le m\le n} n^{-1/2} \left| \sum\nolimits_{i=1}^{m} \widehat{u}_i^2 - \frac{m}{n} \sum\nolimits_{i=1}^{n} \widehat{u}_i^2 \right| / \widehat{\omega}, \tag{3.2}$$

where

$$\widehat{\omega}^2 = \sum\nolimits_{l=-n+1}^{n-1} k(l/b)\,\widehat{\gamma}(l). \tag{3.3}$$

If one denotes the long run variance (LRV) of $u_i^2 - \sigma_i^2$ by ω^2, which is given by $\omega^2 = \xi^2 \int_0^1 \sigma^4(t)\,dt$ with $\xi^2 = \sum_{l=-\infty}^{\infty} \psi(l)$ and $\psi(l) = E\left[(\varepsilon_i^2 - 1)\left(\varepsilon_{i-l}^2 - 1\right)\right]$, then $\widehat{\omega}^2$ is an estimator of ω^2. In (3.3), $k(\cdot)$ is another kernel function, $b = b_n > 0$ is another bandwidth (or called a truncation parameter if $k(\cdot)$ has a bounded domain), and $\widehat{\gamma}(l)$ is an estimator of the l-th autocovariance $E\left[(u_i^2 - \sigma_i^2)\left(u_{i-l}^2 - \sigma_{i-l}^2\right)\right]$, specifically, $\widehat{\gamma}(l) = n^{-1}\sum_{i=l+1}^{n}(\widehat{u}_i^2 - \widehat{\sigma}^2)\left(\widehat{u}_{i-l}^2 - \widehat{\sigma}^2\right)$ for $l \geqslant 0$; $\widehat{\gamma}(l) = \widehat{\gamma}(-l)$ for $l < 0$, where $\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\widehat{u}_i^2$.

The CUSQ test has been considered by Xu (2012), Deng and Perron (2008b), Nielsen and Sohkanen (2011), and many others, but most are based on directly observed data, least squares residuals, or forecast errors from a model. Here the proposed test statistic $\widehat{Q}$ is based on nonparametric residuals.

3.3 Asymptotic Properties of the CUSQ Test

In this section, I analyze the asymptotic properties of the proposed CUSQ test $\widehat{Q}$. Under general conditions for nonparametric estimation, I can show that the test has the same limit distribution (the supremum of a Brownian bridge) with a CUSQ test based on directly observed data. As an illustration, I also clarify the limit behavior of a CUSQ test statistic based on least squares residuals, whereas the data are generated from the time-varying coefficient model (3.1). To begin with, I make the following assumptions.

Assumption 3.3.1. *The regression coefficient functions* $\beta(\cdot) : [0,1] \to \mathbb{R}^d$ *are smooth and bounded, at least twice continuously differentiable, except at a finite number of points of discontinuity on* $[0,1]$, *and* $\sup_{t\in(0,1)} \left\| \lim_{s\to t+} \beta(s) - \lim_{s\to t-} \beta(s) \right\| \leqslant C$ *for some constant* C.

Assumption 3.3.2. *The error term is such that* $u_i = \sigma_i \varepsilon_i$, *where* $\sigma_i^2 = \sigma^2(z_i)$ *and* $E(\varepsilon_i|X_i) = 0$, $E(\varepsilon_i^2|X_i) = s^2(X_i)$, $E(\varepsilon_i^2) = 1$. *The conditional variance* $s^2(\cdot) : [0,1] \to \mathbb{R}_+$ *is continuous and bounded; the unconditional variance* $\sigma^2(\cdot) : [0,1] \to \mathbb{R}_+$ *is a smooth and bounded cadlag function, at least twice continuously differentiable, except at a finite number of points of discontinuity.*

Assumption 3.3.3. *Assume that* $\{(\varepsilon_i, X_i)\}$ *is strictly stationary* β*-mixing. Further, assume that there exists some* $\delta > 0$ *such that* $E\|X_i\|^{2(2+\delta)} < \infty$, $E|\varepsilon_i|^{2(2+\delta)} < \infty$, *and the mixing coefficient* $b(i)$ *is geometrically decreasing, that is,* $b(i) = O(i^{-\tau})$, *and* $\tau > (2-\epsilon)(2+\delta)/\delta$ *for some* δ, $\epsilon > 0$.

Assumption 3.3.4. *Assume that* $n^{-1}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \Omega_0$, *where* $\Omega_0 = E(X_1 X_1')$ *is a positive definite matrix. Furthermore, the LRV* $\xi^2 = \sum_{l=-\infty}^{\infty} E\left[\left(\varepsilon_i^2 - 1\right)\left(\varepsilon_{i-l}^2 - 1\right)\right]$ *is strictly positive and finite.*

Assumption 3.3.5. *For the kernel function and bandwidth used in the local linear estimation of* $\beta(\cdot)$, *assume that,* $K(\cdot)$ *is symmetric and bounded, satisfies the Lipschitz condition, and has a bounded support, say* $[-1,1]$; *the bandwidth satisfies* $h \to 0$, $nh^2 \to \infty$, $nh^8 \to 0$ *as* $n \to \infty$.

Assumption 3.3.6. *For the kernel function and bandwidth used in the LRV estimation, assume that,* $k(\cdot)$ *has a bounded support* $[-1,1]$ *and* $|k(x)| \leqslant 1$ *for all* $x \in \mathbb{R}$, $k(x) = k(-x)$, $k(0) = 1$; $k(x)$ *is continuous at* 0 *and for almost all* $x \in \mathbb{R}$; $|k(x)| \leqslant \left|\overline{k}(x)\right|$ *where* $\overline{k}(\cdot)$ *is a nonincreasing function such that*

$\int_{-\infty}^{\infty} \left| x\overline{k}(x) \right| dx < \infty$; *$b \to \infty$ as $n \to \infty$ and $b = O\left(n^{\vartheta}\right)$, $\vartheta < 1/2 - 1/r$ where r is such that $2 < r < 2 + \delta$.*

Remark 3.3.1. *The assumptions are largely adopted from Cai (2007), with necessary modifications to use results from Hansen (2008), Kristensen (2009), and Cavaliere (2004). Assumption 3.3.1 allows for both smooth and abrupt structural changes in the regression coefficients $\beta(\cdot)$. For possible abrupt structural changes, the location of break points may not be known, but the break size is bounded. The specification of $\sigma^2(\cdot)$ in Assumption 3.3.2 allows for time varying unconditional variance of the error terms.*

Remark 3.3.2. *The β-mixing condition in Assumption 3.3.3 is only necessary for some U-statistic results for dependent sequences, as in Kristensen (2011), which are used to verify a negligible term in the difference between the estimated and true sum of squared residuals (See Lemma B.4.3 in Appendix B). In other parts of the proof, the weaker assumption of α-mixing is sufficient. Specifically, if I put $\epsilon = 1 - \delta$, then Assumption 3.3.3 implies that $\{(\varepsilon_i, X_i)\}$ is α-mixing with the mixing coefficient $\alpha(i)$ satisfies $\alpha(i) = O(i^{-\tau})$, and $\tau > (2+\delta)(1+\delta)/\delta$. This is the α-mixing condition required in Cai (2007).*

Remark 3.3.3. *The above mentioned α-mixing condition further implies Assumption 2 in Hansen (2008), the latter is used to establish the uniform convergence rate of $\widehat{\beta}(z)$ in Lemma B.5.1 in Appendix B. The α-mixing condition also guarantees that $\{\varepsilon_i^2 - 1\}$ satisfies the Assumption E in Cavaliere (2004) if one selects r, p there to be such that $2 < r < 2 + \delta$, $p = 2 + 2\delta$.*

Remark 3.3.4. *The bandwidth conditions in Assumption 3.3.5 are quite general. For small positive δ, ϵ, the optimal bandwidth $h_{opt} = O\left(n^{-1/5}\right)$ minimizing the asymptotic mean square error satisfies Assumption 3.3.5, see Remark 5 in Cai (2007). Assumption 3.3.6 is for the kernel function and bandwidth in the LRV estimator $\widehat{\omega}^2$, as used in Cavaliere (2004) and Xu (2012).*

Before moving towards to the proposed CUSQ test based on nonparametric residuals, I state two theorems in which I characterize the limit behavior of the test statistic based on OLS residuals. If the true regression model is with time-varying coefficients as in (3.1), then in general the least squares estimation will not be consistent. Hence a test based on OLS residuals is expected to have incorrect size. Through the following Theorem 3.3.1 and 3.3.2 one can find what kind of size distortion the test will suffer if the time-varying conditional mean is not consistently estimated. In the first step, I consider a simplified case in which $u_i = \sigma_i \varepsilon_i$, and the terms $\{\varepsilon_i\}$ are independent and identically distributed (i.i.d.). In this special case, one only need to use an estimator of the variance of u_i^2 for standardization. Let $Var\left(u_i^2\right) = E\left[\left(u_i^2 - \sigma_i^2\right)^2\right]$, define the variance estimator as $\overline{\gamma}(0) = n^{-1} \sum_{i=1}^n \left(\overline{u}_i^2 - \overline{\sigma}^2\right)^2$, where $\overline{\sigma}^2 = n^{-1} \sum_{i=1}^n \overline{u}_i^2$ and $\{\overline{u}_i\}$ are the OLS residuals. Then the test statistic is modified as

$$\overline{Q}_1 = \max_{1 \leq m \leq n} n^{-1/2} \left| \sum\nolimits_{i=1}^m \overline{u}_i^2 - \frac{m}{n} \sum\nolimits_{i=1}^n \overline{u}_i^2 \right| / \sqrt{\overline{\gamma}(0)}.$$

Theorem 3.3.1. *Suppose that Assumptions 3.3.1-3.3.4 hold. Let* $m = \lfloor rn \rfloor$ *for* $0 < r \leqslant 1$. *If in the error terms, the part without time-varying variance, that is,* $\{\varepsilon_i\}$ *are i.i.d., then under* $H_1 : \sigma_i^2 = \sigma^2(z_i)$,

$$n^{-1/2} \overline{Q}_1 \xrightarrow{p} \sup_{r \in (0,1]} \left| \int_0^r \sigma^2 - r \int_0^1 \sigma^2 + \Delta(r) \right| / \sqrt{\gamma^*},$$

where

$$\Delta(r) = \int_0^r \beta' \Omega_0 \beta - r^{-1} \left(\int_0^r \beta\right)' \Omega_0 \left(\int_0^r \beta\right) - r \int_0^1 \beta' \Omega_0 \beta + r \left(\int_0^1 \beta\right)' \Omega_0 \left(\int_0^1 \beta\right),$$

and γ^* *is the probability limit of* $\overline{\gamma}(0)$. *Specifically,*

$$\begin{aligned}\gamma^* = {} & [\psi(0) + 1] \int \sigma^4 + 6 \left[\int \sigma^2 \beta' \Omega_\varepsilon \beta - \int \sigma^2 \left(\int \beta\right)' \Omega_\varepsilon \left(\int \beta\right)\right] \\ & - 3\Omega_{40} + 6\Omega_{22} - 4\Omega_{13} + \Omega_{04} - \left(\int \sigma^2 + \int \beta' \Omega_0 \beta - \left(\int \beta\right)' \Omega_0 \left(\int \beta\right)\right)^2,\end{aligned}$$

with

$$\psi(0) = E\left(\varepsilon_1^2 - 1\right)^2,$$
$$\Omega_\varepsilon = E\left(X_1 X_1' \varepsilon_1^2\right),$$
$$\Omega_{40} = E\left[\left(\int\beta\right)' X_1\right]^4,$$
$$\Omega_{22} = E\left[\left(\int\beta\right)' X_1 X_1' \left(\int\beta\right) \int \left(\beta' X_1 X_1' \beta\right)\right],$$
$$\Omega_{13} = E\left[\left(\int\beta\right)' X_1 \int \left(X_1' \beta' X_1 X_1' \beta\right)\right],$$
$$\Omega_{04} = E\left[\int \left(\beta' X_1 X_1' \beta' X_1 X_1' \beta\right)\right].$$

Remark 3.3.5. *From Theorem 3.3.1 one can verify that under both H_0 and H_1, the test statistic based on OLS residuals diverges with the rate $\sqrt{n}$. The test does not have a correct size because of the bias of least squares estimation.*

In the second step, I consider a general case of serially correlated errors, in which the error terms satisfy the mixing condition in Assumption 3.3.3. In this case, I need to use a LRV estimator for standardization. That is, I consider the test statistic given as follows:

$$\overline{Q}_2 = \max_{1 \le m \le n} n^{-1/2} \left| \sum\nolimits_{i=1i}^{m} \overline{u}^2 - \frac{m}{n} \sum\nolimits_{i=1}^{n} \overline{u}_i^2 \right| / \overline{\omega},$$

where $\overline{\omega}^2$ is the LRV estimator based on OLS residuals $\{\overline{u}_i\}$. The construction of $\overline{\omega}^2$ should be straightforward.

Theorem 3.3.2. *Suppose that Assumptions 3.3.1-3.3.4 and 3.3.6 hold. Let $m = \lfloor rn \rfloor$ for $0 < r \leqslant 1$. If the error terms $\{u_i\}$ are serially correlated, then under $H_1 : \sigma_i^2 = \sigma^2(z_i)$,*

$$n^{-1/2} b^{1/2} \overline{Q}_2 \xrightarrow{p} \sup_{r \in (0,1]} \left| \int_0^r \sigma^2 - r \int_0^1 \sigma^2 + \Delta(r) \right| / \omega^*,$$

where

$$\Delta(r) = \int_0^r \beta'\Omega_0\beta - r^{-1}\left(\int_0^r \beta\right)'\Omega_0\left(\int_0^r \beta\right) - r\int_0^1 \beta'\Omega_0\beta + r\left(\int_0^1 \beta\right)'\Omega_0\left(\int_0^1 \beta\right),$$

$$(\omega^*)^2 = \left(\int k(x)\,dx\right)\left[\int \sigma^4 - \left(\int \sigma^2 + \int \beta'\Omega_0\beta - \left(\int \beta\right)'\Omega_0\left(\int \beta\right)\right)^2\right].$$

Remark 3.3.6. *Theorem 3.3.2 shows that again a test based on OLS residuals will diverge so it cannot be a valid test due to the size distortion. This will also be shown in the simulation results in Section 3.4.*

Now in the following Theorem 3.3.3 I present the asymptotic properties of the proposed test based on residuals from nonparametric estimation.

Theorem 3.3.3. *Suppose that Assumptions 3.3.1-3.3.6 hold. Let* $m = \lfloor rn \rfloor$ *for* $0 < r \leqslant 1$. *Under* $H_0 : \sigma_i^2 = \sigma^2$,

$$\widehat{Q} = \max_{1\le m\le n} n^{-1/2}\left|\sum_{i=1}^m \widehat{u}_i^2 - \frac{m}{n}\sum_{i=1}^n \widehat{u}_i^2\right|/\widehat{\omega} \Rightarrow \sup_{r\in(0,1]} |W(r) - rW(1)|.$$

Under $H_1 : \sigma_i^2 = \sigma^2(z_i)$,

$$n^{-1/2}b^{1/2}\widehat{Q} \xrightarrow{p} \sup_{r\in(0,1]} \left|\int_0^r \sigma^2 - r\int \sigma^2\right| \cdot \left[\int \sigma^4 - \left(\int \sigma^2\right)^2\right]^{-1/2}.$$

Remark 3.3.7. *Theorem 3.3.3 shows that, under quite general bandwidth condition in Assumption 3.3.5 ($h \to 0$, $nh^2 \to \infty$, $nh^8 \to 0$ as $n \to \infty$) the test based on local linear residuals will enjoy the same limit distribution as a test based on $\{u_i\}$ does. The limit distribution is the supremum of a standard Brownian bridge, so the critical values can be obtained by simulations. As pointed out by Deng and Perron (2008b), the critical values at the 1%, 5%, and 10% size are 1.63, 1.36, and 1.22, respectively.*

3.4 Simulation Study

3.4.1 Simulation Designs

In this section, I study the finite-sample performance of the proposed test in several simulation experiments. I consider the following simulation design: $y_i = \beta(z_i) + u_i$, where $\beta(z)$ is the time varying mean and the error term is $u_i = \sigma_i \varepsilon_i$. For $z \in (0, 1]$, I use various specification of the mean function considered in the literature, such as

$$\begin{aligned}
M0 &: \beta(z) = 0.2 \exp(\delta_b z - 0.7), \ \delta_b = 0, \\
M1 &: \beta(z) = 0.2 \exp(\delta_b z - 0.7), \ \delta_b = 0.7, \\
M2 &: \beta(z) = 0.2 \exp(\delta_b z - 0.7), \ \delta_b = 1.1, \\
M3 &: \beta(z) = 1.5 + \delta_b \mathbb{I}(z > \tau), \ \delta_b = -0.1, \\
M4 &: \beta(z) = 1.5 + \delta_b \mathbb{I}(z > \tau), \ \delta_b = -0.2, \\
M5 &: \beta(z) = 1.5 + \delta_b \mathbb{I}(z > \tau), \ \delta_b = -0.3,
\end{aligned}$$

where $M0$ is a constant mean: $\beta(z) = 0.2 \exp(-0.7) \approx 0.099$; while $M1$ and $M2$ correspond to smoothly varying mean and $\beta(z) \in (0.099, 0.200)$ in $M1$, $\beta(z) \in (0.099, 0.300)$ in $M2$, respectively; $M3$-$M5$ are cases of constant mean except for a single jump at τ. I use a variety of different jump points to study the effect of mean jump on the tests under investigation. For ε_i, I consider two DGPs as (i) $\varepsilon_i \sim iid\ N(0, 0.01)$ and (ii) GARCH(1,1): $\varepsilon_i = \phi_i \eta_i$, $\phi_i^2 = \mu + \alpha \varepsilon_{i-1}^2 + \beta \phi_{i-1}^2$ with $\mu = 0.005$, $\alpha = 0.1$, $\beta = 0.4$, and $\eta_i \sim iid\ N(0, 1)$. Note that in DGP (ii) I still have

$Var(\varepsilon_i) = 0.01$. For the possible structural change in volatility $\sigma^2(z)$, I consider the following specifications:

$$
\begin{aligned}
&V0 : \sigma^2(z) \equiv 1, \\
&V1 : \sigma^2(z) = 1 + \delta\mathbb{I}(z > 0.5), \ \delta = 0.21, \\
&V2 : \sigma^2(z) = 1 + \delta\mathbb{I}(z > 0.5), \ \delta = 0.44, \\
&V3 : \sigma^2(z) = 1 + \delta\mathbb{I}(z > 0.5), \ \delta = 0.69, \\
&V4 : \sigma^2(z) = 1 + \delta\mathbb{I}(0.4 < z \leqslant 0.6), \ \delta = 0.69, \\
&V5 : \sigma^2(z) = 1.5 - 0.7\exp\left(-3(z-0.5)^2\right),
\end{aligned}
$$

where $V0$ corresponds to no structural change under H_0, $V1$-$V3$ are cases of a single jump in volatility, $V4$ is a non-persistent temporal jump, and $V5$ represents smooth volatility change in a narrow range $(0.80, 1.17)$.

For each of the combinations of mean specification $M0$-$M5$ and volatility specification $V0$-$V5$, I fix two sample sizes $n = 300$ and $n = 600$, and the number of replications is 1000. In each replication, I use the Epanechnikov kernel $K(x) = (3/4)(1 - x^2)\mathbb{I}(|x| \leqslant 1)$ in the local linear estimation of the time varying mean function, and select an optimal bandwidth by minimizing AIC as suggested in Cai (2007). The median of selected bandwidths is reported. The Bartlett kernel $k(x) = (1 - |x|)\mathbb{I}(|x| \leqslant 1)$ is used in the LRV estimation and the truncation parameter b is selected by the AR(1)-based plug-in data dependent bandwidth as suggested by Andrews (1991). Rejection rates at 5% critical value are shown in Table C.1-C.8.

3.4.2 Simulation Results

The simulations results are reported in Table C.1-C.6. In each table, a column labeled as *True* contains rejection rates from the CUSQ test based on simulated data of $\{u_i\}$, as if the error term is observable. The reason of including these results is, this will be a benchmark when I compare the two tests based on different mean estimation. A column labeled as *OLS* stands for the CUSQ test based on OLS residuals; and a column labeled as *LL* stands for the CUSQ test based on local linear residuals. Of course, in the last two cases the test is performed on the $\{y_i\}$ series. In each table, the results from two sample sizes $n = 300$ and $n = 600$ are presented together, so the effects of increasing the sample size are easily seen. From results in the True and LL columns, usually the empirical size under the null is closer to 0.05, and the power under the alternative is larger after the increase of sample size. So I will focus on the larger sample size in the comparison of the three tests. In general, from both the size and power aspects, the test based on local linear residuals behaves very close to the test based on true residuals $\{u_i\}$, no matter whether a smooth mean or a jump mean is added in. But the test based on OLS residuals show severe size distortion when the mean specification is not a constant (in $M1$-$M5$). The size distortion will be larger when the mean change is larger, in both the smooth change ($M0$-$M2$) and jump cases ($M3$-$M5$). Besides the poor size property, the test based on OLS residuals also exhibits unreliable power throughout all the simulation designs. A point needs to be made is, both the size and power distortion can be downward or upward, it seems that the direction of a size or power distortion depends on the numerical specification of mean and variance. For example, despite that one might expect an inflated size in most of the cases, a shrinking size is find in the jump mean cases ($M3$-$M5$) when $\tau = 1/2$ (Table C.3-C.4). Interestingly, the size will be smaller as the mean jump becomes larger. This can be explained as follows.

As in the mean specification $M3$-$M5$, one can write a general form of constant mean with one jump as $\beta(z) = b_0 + \delta_b \mathbb{I}(z > \tau)$, that is, $\beta(z) = b_0$ for $0 < z \leqslant \tau$, and $\beta(z) = b_1$ for $\tau < z \leqslant 1$, with $b_1 = b_0 + \delta_b$. With this specification of mean jump, I can calculate the probability limit of $\overline{Q}$ from Theorem 3.3.1 and 3.3.2. For instance, in the case of i.i.d. error, under H_0 of constant volatility, it is obtained that $n^{-1/2}\overline{Q}_1 \xrightarrow{p} \sup_{r\in(0,1]} |g(r,\tau,\delta_b)| / \sqrt{\gamma^*(\tau)}$, where

$$g(r,\tau,\delta_b) = \int_0^r \beta^2(z)\,dz - r^{-1}\left(\int_0^r \beta(z)\,dz\right)^2 - r\int_0^1 \beta^2(z)\,dz + r\left(\int_0^1 \beta(z)\,dz\right)^2$$

$$= r\tau(1-\tau)\delta_b^2 \mathbb{I}(0 < z \leqslant \tau) + [1-\tau/r - r(1-\tau)]\,\tau\delta_b^2 \mathbb{I}(\tau < z \leqslant 1),$$

$$\gamma^*(\tau) = \psi(0) + 4\sigma^2\tau(1-\tau)\delta_b^2 + \tau(1-\tau)(1-2\tau)^2\delta_b^4,$$

so for any $\tau \in (0,1]$, $\tau \neq 1/2$, $\sup_{r\in(0,1]} |g(r,\tau,\delta_b)| = O(\delta_b^2)$ and $\sqrt{\gamma^*(\tau)} = O(\delta_b^2)$; but for $\tau = 1/2$, $\sup_{r\in(0,1]} |g(r,1/2,\delta_b)| = \delta_b^2/8$, still it is $O(\delta_b^2)$, while $\sqrt{\gamma^*(1/2)} = \sqrt{\psi(0) + \sigma^2\delta_b^2} = O(\delta_b)$. Therefore, for small δ_b less than 1, such as those used in the simulation, $\delta_b = 0.1$, 0.2, 0.3, a mean jump at $\tau = 1/2$ will result in smaller size distortion towards to zero as δ_b increases. But for jump point other than 1/2, the size distortion is shown to be increasing towards to one as δ_b increases. See Table C.7 and C.8 for the details.

3.4.3 Additional Simulations

When applying the test procedure to many different time series variables, I find that smooth mean change and volatility change co-exist in many macroeconomic variables. To provide a more realistic example in simulation, I consider the following design. Again I focus on an univariate variable $y_i = \beta(z_i) + u_i$, this time I fix the time varying mean $\beta(z)$ as

$$M0' : \beta(z) = 0.5z + \exp\left(-4(z-0.5)^2\right),$$

and consider the following volatility changes:

$$V0' : \sigma^2(z) \equiv 1,$$

$$V1' : \sigma(z) = 0.5\beta(z),$$

$$V2' : \sigma(z) = \beta(z),$$

$$V3' : \sigma(z) = 1.5\beta(z).$$

The smoothly varying mean $M0'$ has the same functional form as used in a simulation design in Cai (2007), and $V0'$ is under our H_0, $V1'$-$V3'$ are cases of smooth volatility change proportional to the mean change. The two DGPs (i.i.d. and GARCH(1,1)) for the error terms are the same, except for a rescaled variance, that is, $Var(\varepsilon_i) = 1$ instead for $Var(\varepsilon_i) = 0.01$. In these variance specifications, the magnitude of volatility is manifestly larger than the mean change. For example, a proportional volatility change $\sigma(z) = 0.5\beta(z)$ with GARCH(1,1) error and $n = 300$ is presented in Figure 3.1, in which the smoothly varying mean is near a constant mean due to the large volatility.

The results for this set of simulations are shown in Table C.9. These results are particularly illustrative because now the OLS residuals based test yields higher size and lower power under the same mean specification, compared with the local linear residuals based test. Consequently, the OLS residuals based test cannot be reliable in both the null and the alternative.

3.5 Two Empirical Examples

To show the different implications of a CUSQ test based on different mean estimation method, that is, whether correct for the time-varying mean, I apply the two tests on several empirical time series data sets. The data are obtained from the Federal Reserve Bank of St. Louis web site. As a first example, I consider the CPI

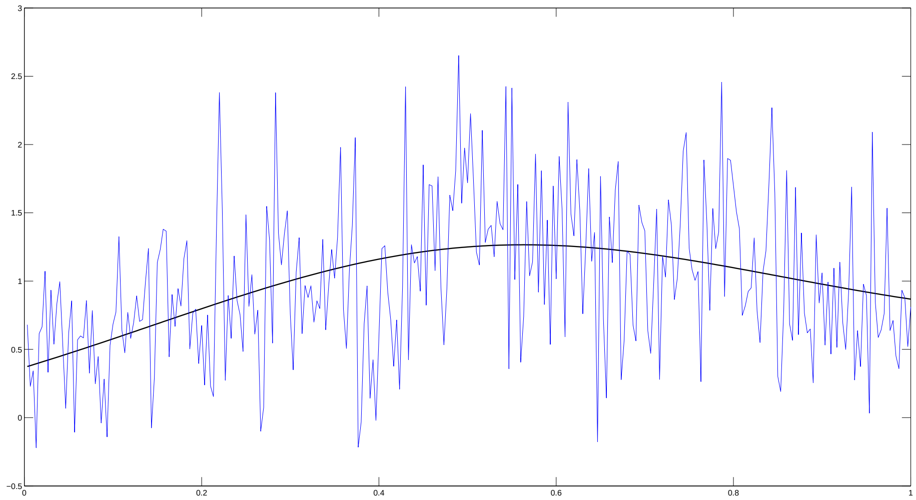

Fig. 3.1. Time-varying mean with proportional variance

based inflation rate. The raw data is a seasonally adjusted series of monthly CPI index (1982-84=100) from 1947/01 to 2012/04 ($n = 783$). As shown in Figure 3.2, the CPI based inflation rate might possess multiple mean changes so that it serves as a good example for our test of volatility change under unstable mean. For this series, the result from CUSQ test based on OLS (demeaned) residuals is 1.278 while the result from test based on local linear residuals is 1.416. So at 5% significant level OLS based test will fail to reject the null of constant volatility while local linear based test will not. The second example I considered is the CBOE S&P 500 3-Month Volatility Index. I use a daily data set which is from 2007/12/04 to 2012/06/13 ($n = 1139$). As seen in Figure 3.3, this series also shows obvious mean changes (well accepted as *jump of volatility*). Then is the volatility of volatility also changing over time? The result from CUSQ test based on OLS (demeaned) residuals is 2.086 while the result from test based on local linear residuals is 1.333. So the test based on OLS residuals

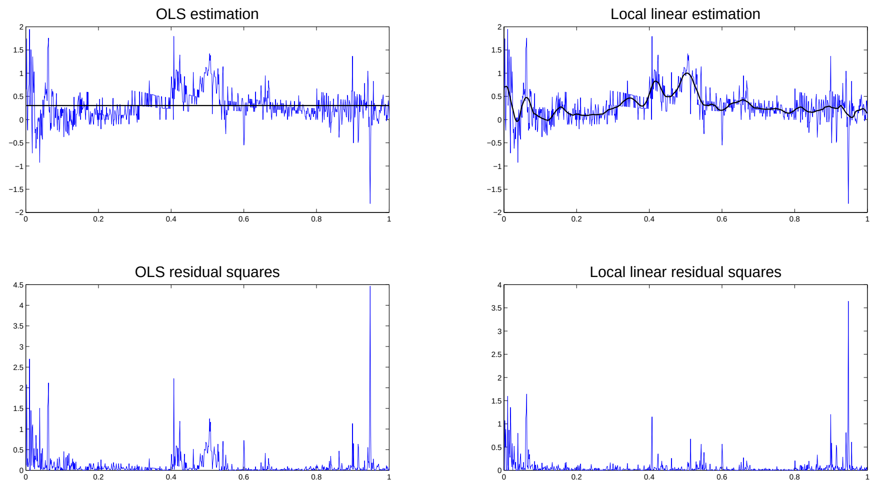

Fig. 3.2. Empirical Example: CPI Based Inflation Rate

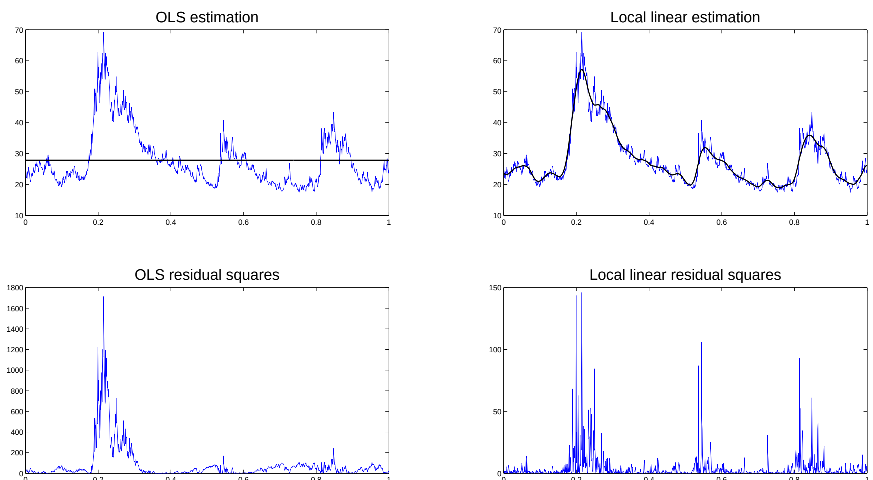

Fig. 3.3. Empirical Example: Option Implied Volatility

will reject the null, but the test based on local linear residuals will not reject, at 5% significant level.

In both the two empirical examples, the two test produce opposite results. As shown in the previous simulation study, it is very possible that the OLS based test leads to a wrong decision because of the co-exist inflated size and low power.

4. FUNCTIONAL-COEFFICIENT COINTEGRATION TEST OF PURCHASING POWER PARITY

4.1 Introduction

Stemming from *the law of one price*, the Purchasing Power Parity (PPP) hypothesis states that nominal exchange rate between two currencies should be determined by the two relevant national price levels so that the price levels, after expressed in a common currency at that rate, equal to each other (Sarno and Taylor, 2002). In other words, the nominal exchange rate should be at the level which causes the purchasing power of a unit of one currency to be the same in both economies.

Let P_t and P_t^* denote the domestic and foreign price levels at time t respectively, also let S_t denote the nominal exchange rate expressed as the foreign price of the domestic currency at time t, then the PPP hypothesis in absolute version can be written as

$$P_t^* = S_t P_t.$$

Given the nonstationarity of the variables involved, empirical tests of the PPP hypothesis are usually performed in the cointegration framework

$$s_t = \beta_0 + \beta_1 p_t^* + \beta_2 p_t + u_t$$

where s_t, p_t, and p_t^* are the natural logarithm of S_t, P_t, and P_t^* respectively, and $(\beta_0, \beta_1, \beta_2)$ is a set of cointegration coefficients expected to be $(0, 1, -1)$. Essentially, a necessary condition for the long-run equilibrium of PPP to hold is that the deviations from PPP, i.e., the residual u_t, is stationary. Hence u_t can be viewed as an *equilibrium error*.

Empirical studies in the literature of cointegration tests of PPP provide mixing results. See surveys such as Taylor and Taylor (2004). Early works in the 1980's either find that the real exchange rate follows a random walk or fail to establish a cointegration relationship (Michael et al., 1997). Later, to overcome the low power of unit root test, long span data and panel data are considered (Sarno and Taylor, 2002). But still the long-established issue is unresolved.

Recently, among other studies, Michael et al. (1997) and Baum et al. (2001) argue that the deviations from PPP should exhibit nonlinear behavior under the presence of transactions costs. They postulate that the deviations may behave similar as a unit root process in a band in which they are small relative to the costs of trading; out of the band, international price arbitrage can cause mean-reverting movement of exchange rate toward the long-run equilibrium of PPP. In practice, they use the exponential smooth transition autoregressive (ESTAR) model on the residual term u_t to capture the above features.

Instead of imposing a parametric model on the residuals, I consider a flexible functional-coefficient cointegration (FCC) model in which the cointegration relationship itself may vary according to the degree of adjustment toward PPP. I use the nonparametric estimation and testing method developed in Cai et al. (2009) and Xiao (2009) to test the PPP hypothesis. By constructing a measure of real exchange rate volatility at each time, I empirically test the cointegration suggested by the PPP condition, allowing the cointegration coefficients to vary as a function of real exchange rate volatility. Indeed, the model adopted in this study incorporates the traditional linear cointegration model (with constant coefficients) as a special case.

The rest of this section is organized as follows. Section 4.2 introduces a measure of real exchange rate volatility and establishes the link between this measure and the nonlinear adjustment of real exchange rate toward PPP. Section 4.3 briefly summarizes the functional-coefficient cointegration model used in my empirical study. Section 4.4 reports the empirical results and discusses the implication of the results.

4.2 Nonlinear Adjustment toward PPP and Real Exchange Rate Volatility

Equilibrium models of exchange rate determination with transactions costs, such as Dumas (1992) and Sercu et al. (1995), suggest that persistent deviations of real exchange rate from the PPP level can exist, provided the deviations are relatively smaller than the costs of international trade. Supporting these theoretical arguments, persistence of deviations from PPP has been widely found in empirical studies. This empirical observation was named as the *PPP puzzle* in Rogoff (1996). According to these arguments, certain nonlinear behavior of real exchange rates might be naturally expected: when the deviation from long-run PPP equilibrium level is small, the exchange rate might move more like a random walk in a certain band containing the PPP level; when the deviation is large enough (out of the band) to offset the transactions costs, the international price arbitrage might force the exchange rate to adjust from the present extreme level toward the PPP level.

Consistently, I consider the cointegration relationship suggested by PPP might vary according to the intenseness of the adjustment. To verify my conjecture empirically, I need a measure of the intenseness of the real exchange rate adjustment toward PPP level. Given the behavioral difference of the real exchange rate, I propose to use real exchange rate volatility as that measure. The intuition here is, the average volatility around time t will be larger if the deviation at time t is smaller, say, in the band; on the contrary, if the deviation at time t is very large (out of the band), then the volatility will be relatively small because the exchange rate might move toward the same direction in several adjacent periods with similar magnitude of change.

To illustrate this point more concretely, I denote the real exchange rate as

$$EX_t = \frac{S_t P_t}{P_t^*}$$

then $GR_t = \ln(EX_t/EX_{t-1})$ is essentially the growth rate of real exchange rate. I construct a measure of the volatility of real exchange rate at time t as the standard deviation of GR in a short period before t:

$$z_t = \sqrt{{}^{d}_{i=1} \left(GR_{t-i} - \overline{GR}\right)^2 / (d-1)}$$

where $\overline{GR} ={}^{d}_{i=1} GR_{t-i}/d$. In the following empirical study, I try $d = 3,\ 6,\ 12$ in the computation of z_t as a step of robustness check. Then the functional-coefficient cointegration model used in this empirical study is

$$s_t = \beta_0(z_t) + \beta_1(z_t)\, p_t^* + \beta_2(z_t)\, p_t + u_t.$$

It is expected that the cointegration coefficients will be more stable when the value of z_t is smaller, i.e., when the real exchange rate is experiencing more intensive adjustment.

4.3 The Functional-Coefficient Cointegration Model

I use the functional-coefficient cointegration model proposed by Cai et al. (2009) and Xiao (2009). The model can be written in a general form as

$$Y_t = \beta(Z_t)' X_t + u_t,\ 1 \le t \le T,$$

where Y_t, Z_t and u_t are scalars, $X_t = (X_{t1}, \cdots, X_{td})'$ is a d-dimensional vector of covariates, $\beta(\cdot)$ is a d-dimensional vector of cointegration coefficients varying according to the value of Z. In this model, Y_t and a part of variables in X_t are assumed to be generated from non-stationary $I(1)$ processes, while other variables in X_t, together with Z_t are generated from a stationary process. In my PPP test, $X_t = \left(1, \ln\left(P_t^f\right), \ln\left(P_t^d\right)\right)$, $Y_t = \ln(S_t)$, and Z_t is a measure of real exchange rate volatility.

To estimate the coefficients $\beta(z)$ as a function evaluated at z, I use a nonparametric local linear estimator $\widehat{\beta}(z)$, which is given by

$$\begin{pmatrix} \widehat{\beta}(z) \\ \widehat{\beta}^{(1)}(z) \end{pmatrix} = \left[\sum_{t=1}^{T} \begin{pmatrix} X_t \\ (Z_t - z) X_t \end{pmatrix}^{\otimes 2} K_h (Z_t - z) \right]^{-1} \times \left[\sum_{t=1}^{T} \begin{pmatrix} X_t \\ (Z_t - z) X_t \end{pmatrix} Y_t K_h (Z_t - z) \right],$$

where $K_h(z) = k(z/h)$, h is a bandwidth parameter and $k(\cdot)$ is a kernel function giving more weights on (X_t, Y_t) with Z_t near z. In the empirical study, I use the Gaussian kernel.

To test the existence of cointegration relationship in (X_t, Y_t), i.e. to test the stationarity of u_t, I use the test proposed by Xiao(2009). The test is based on the estimated residual $\widehat{u}_t = Y_t - \widehat{\beta}(Z_t)' X_t$ from above local linear regression. Because a constant is included in the regressors X_t, under the null of stationary u_t, one can expect u_t to have zero mean and constant variance $\sigma_u^2 = E(u_t^2)$. Thus one can run a linear regression

$$\widehat{u}_t^2 = a + bt + e_t$$

and construct a t-ratio statistic given by

$$\tau_T = \widehat{b}/\widehat{s}(b),$$

where $\widehat{b}$ is the OLS estimate of b, and

$$\widehat{s}(b) = \sqrt{\underline{\widehat{\omega}}^2 / {}_{t=1}^{T} \left(t - \overline{t}\right)^2}$$

with $\underline{\widehat{\omega}}^2$ as a consistent nonparametric estimator of the long-run variance of u_t^2. Xiao (2009) shows that under some regularity conditions, this test statistic asymptotically follows the standard normal distribution.

4.4 Empirical Results

4.4.1 The Variables and Data

I investigate bilateral nominal exchange rates (period average) and price indices of US vs Canada, Japan, France, Germany, Italy, UK. For these 6 country-pairs, the consumer price index (CPI) and the product price index (PPI) are used as the measure of national price levels, as usually used in the empirical test of the PPP hypothesis. As advocated in Xu (2003), I also use a traded-goods price index (TPI), which is constructed as the weighted average of export and import price indices. Monthly data from 04/1973 to 12/2008, which corresponds to the post-Bretton Woods era of float exchange rate, of all variables are extracted from IMF International Financial Statistics (*http://www.imfstatistics.org/imf/*). During this period, sample size might be shorter for several pairs in which Europe countries are involved due to data availability.

4.4.2 Preliminary Unit Root Test of Individual Variables

Before forwarding to the cointegration test, I first examine the non-stationarity of each time series. I use the augmented Dickey and Fuller (ADF) test for unit root. Basing on the following regression equation,

$$\Delta y_t = (\rho - 1)\, y_{t-1} + \beta_0 + \beta_1 t + \alpha_1 \Delta y_{t-1} + \cdots + \alpha_p \Delta y_{t-p} + u_t,$$

the ADF test is a t-test of the significance of $\rho - 1$. I also use the usual model selection criterion to include a constant, a time trend, if the coefficients associated with them are significant. Table C.10 in the Appendix C contains results of the ADF test for unit root. In general, the results of unit root test are consistent with those performed in existing researches. The result shows that the null of unit root cannot be rejected at 1% significance level in 22 out of the 26 variables. The *significant* results might be because of the low power of ADF test and the conservative model selection criterion used here. Moreover, even in those significant cases, the estimates of coefficients are close to zero, suggesting a near unit root behavior. Given these results, it can viewed that these series are $I(1)$ processes.

4.4.3 Results of the Linear Cointegration Test

To compare with my functional-coefficient cointegration test, first I conduct the standard linear (constant coefficient) cointegration test pioneered by Enger and Granger. It is essentially an ADF test based on the least squares residuals. The results of the *Enger-Granger cointegration test* are presented in Table C.11 in the Appendix C. In all of the 17 cases, the null of unit root cannot be rejected for the estimated residual series, even at the 10% significance level. Thus based on these test results, each time I find no (linear) cointegration relationship among the three variables under consideration. These results are close to those documented in the literature, that is, focusing on relatively high frequency (monthly) data in the recent float exchange rate, usually one fails to find the existence of (linear) cointegration relationship among the bilateral nominal exchange rate and two price levels.

4.4.4 Results of the Functional-Coefficient Cointegration Test

Table C.12 in the Appendix C contains results of the Xiao's nonparametric test for unit root, based on the estimated local linear residuals from the functional-coefficient cointegration model described in Section 4.3. Being different with the Enger-Granger cointegration test, now the null hypothesis is that the error term is stationary, that is, there is cointegration, possibly with varying coefficients. Under the 5% significance level, the null of cointegration cannot be rejected in 13 ($d = 3$), 14 ($d = 6$), and 14 ($d = 12$) cases out of the total 17 case. These results are strikingly in contrast with those of the linear (constant coefficient) cointegration test. Allowing for varying coefficients, I find strongly supporting evidence of the PPP hypothesis. The cointegration relationship varies according to the degree of real exchange volatility.

To give a more direct comparison of the performance of linear (constant coefficient) cointegration (LC) test and my functional-coefficient cointegration (FCC) test, I further apply the ADF test on the FCC estimated residuals (with z_t computed by using $d = 12$), and compare the test results with those based on the LC estimated residuals. The results are presented in Table C.13 in the Appendix C. The t-values from FCC estimated residuals are larger than those from LC estimated residuals in 14 cases, and the two are very close in the left 3 cases. The ADF test on the FCC estimated residuals yield more significant results.

In addition to the test results, plotting the estimated cointegration coefficients as functions of real exchange rate volatility yields clearer illustration of how the cointegration relationship evolves. Figure 4.1-4.3 present the three estimated coefficients $\widehat{\beta}_0(z)$, $\widehat{\beta}_1(z)$, $\widehat{\beta}_2(z)$ for three PPI-based cases (three countries vs US). The overall impression of these graphs is, when the value of z (a measure of real exchange volatility) is relatively smaller, which might be associated with faster and more adjustment towards to the PPP equilibrium, the cointegration coefficients are more stable and numerically closer to the values suggested by the PPP hypothesis (i.e., $(\beta_0(z), \beta_1(z), \beta_2(z)) = (0, 1, -1)$), especially in the UK vs US case. While when the value of z becomes larger and larger, the cointegration exhibit unstable behavior, which can be understood as weaker or no cointegration when the deviation from PPP is small.

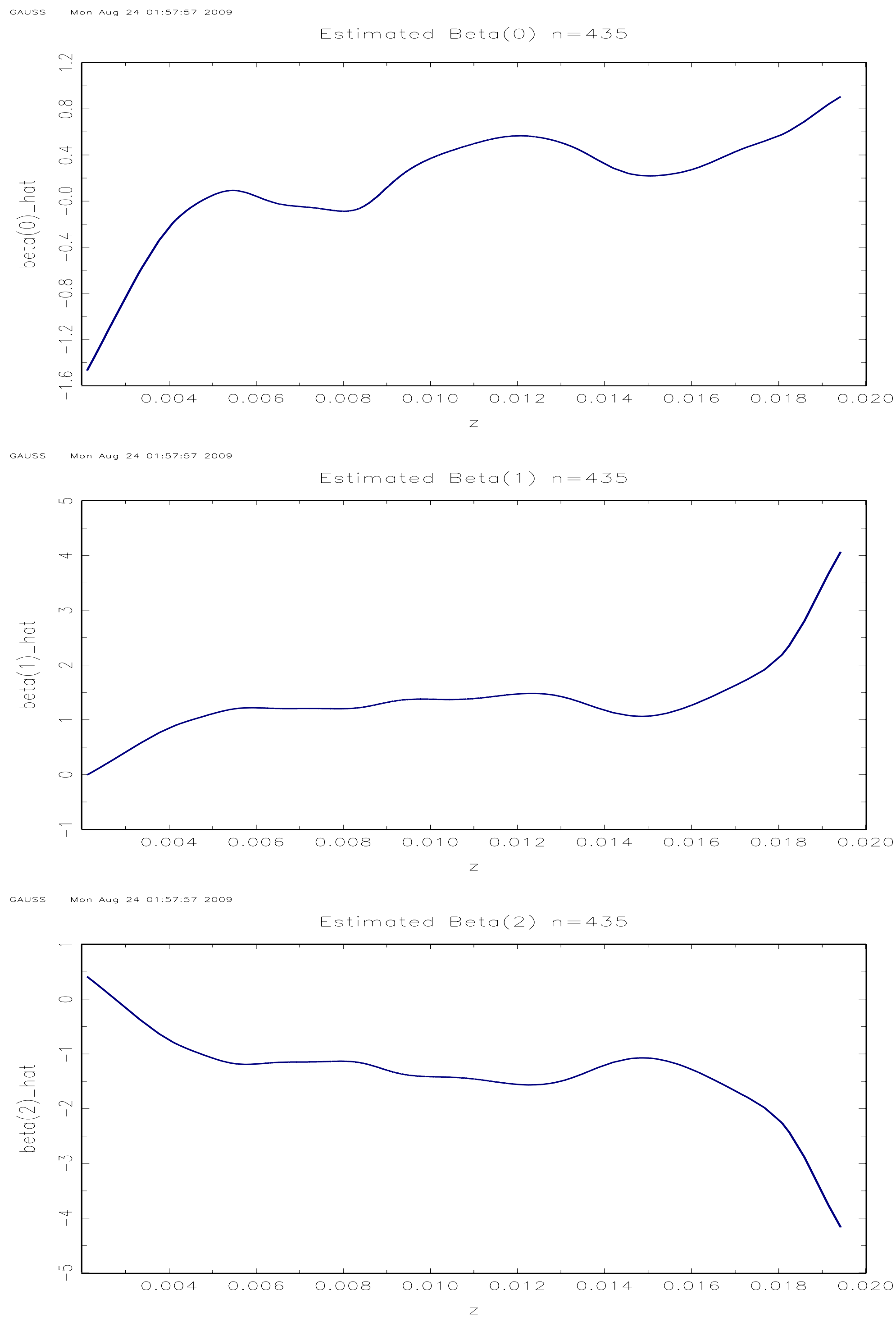

Fig. 4.1. The Estimated Functional Coefficients: Canada vs US, PPI ($d = 6$)

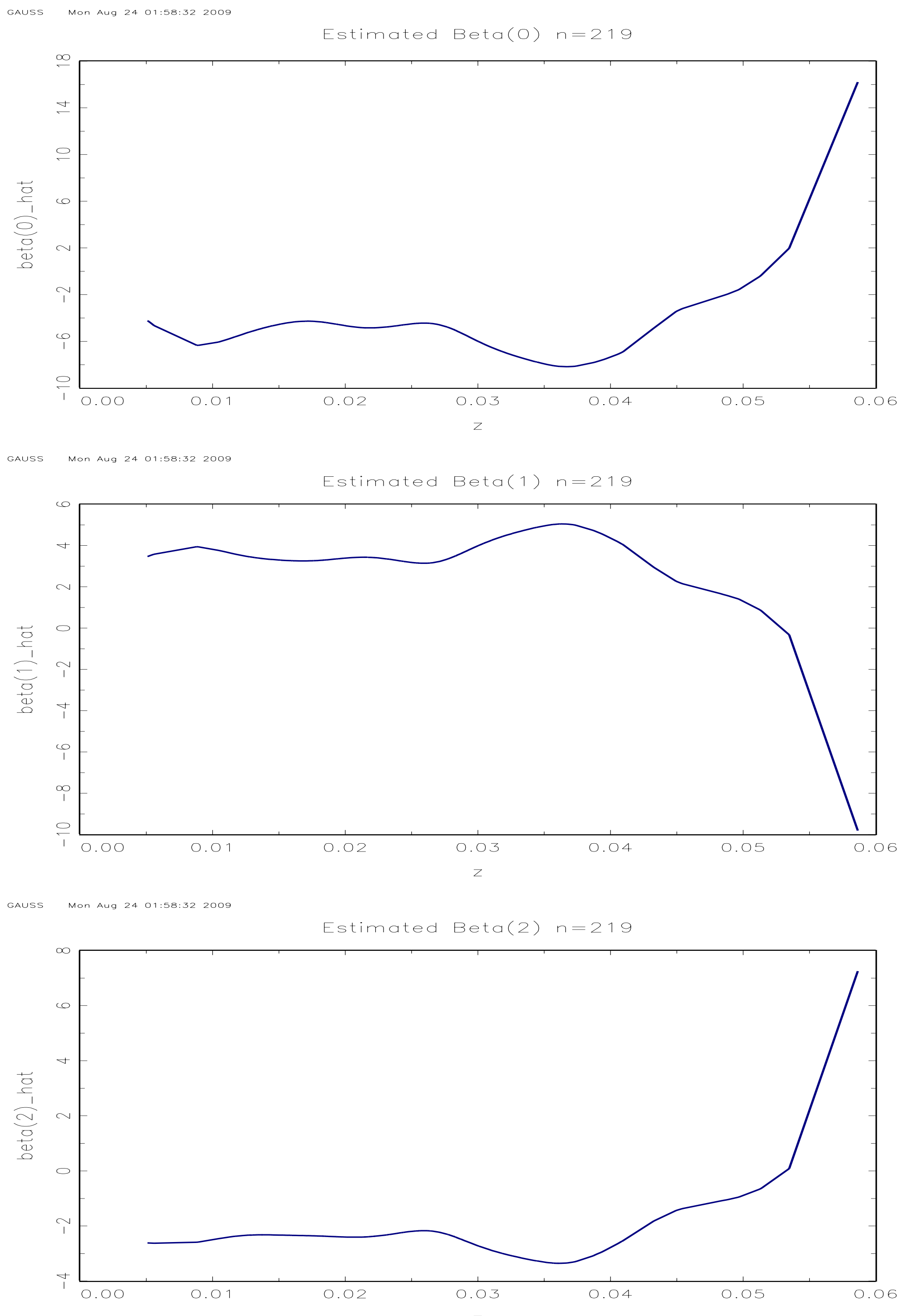

Fig. 4.2. The Estimated Functional Coefficients: Germany vs US, PPI ($d = 6$)

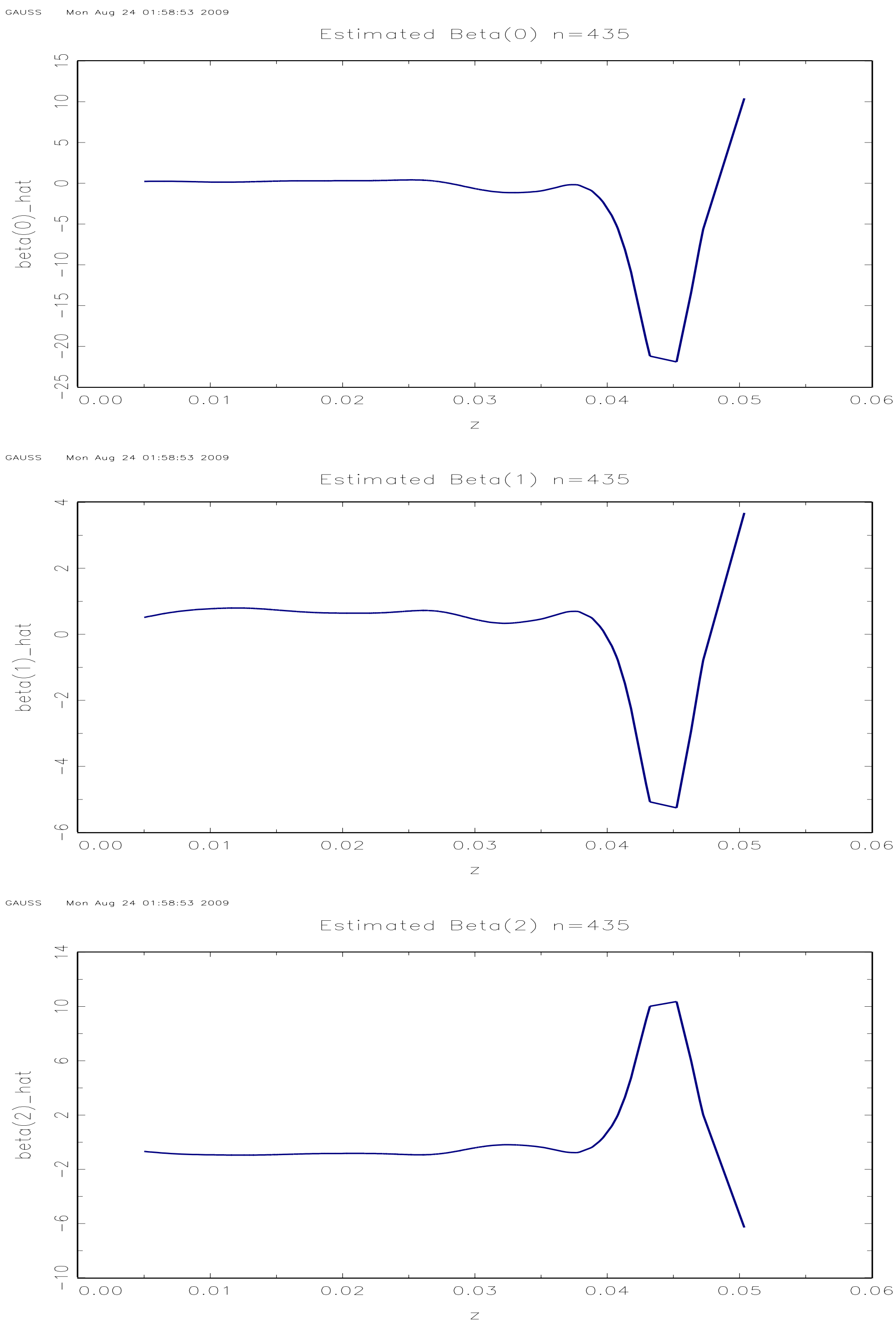

Fig. 4.3. The Estimated Functional Coefficients: UK vs US, PPI ($d = 6$)

5. CONCLUSIONS AND DISCUSSIONS

Economic theory often imply that the functional relationship between economic variables must satisfy certain shape restrictions derived from model assumptions or model properties. In section 2, I develop an empirical likelihood based constrained nonparametric regression method to accommodate such shape restrictions in the estimation of a regression function and its derivatives. Under standard regularity conditions, the proposed constrained local quadratic (CLQ) estimators are shown to be weakly consistent and have the same first order asymptotic distribution as the conventional unconstrained estimators. The CLQ estimators are guaranteed to be within the inequality constraints imposed by economic theory, and display similar smoothness as the unconstrained estimators. At a location where the unconstrained estimator for a curve (e.g., the second derivative) violates a restriction, the corresponding CLQ estimator is adjusted towards to the true function. Interestingly, one can obtain such bias reduction even when the binding effect is from a restriction on another curve (e.g., the first derivative). This finite sample advantage is achieved through the joint estimation of several functions, based on the same empirical likelihood weights. Application on the estimation of daily option pricing function and state-price density function confirms the better performance of the EL-based CLQ estimation method in finite sample.

Several interesting topics for further research are worthy to be mentioned. One direction is to pursue additional analysis on the EL-based CLQ estimators, such as the asymptotic comparison of mean squared errors between the constrained and unconstrained estimators. Another direction which might enrich the scope of this research is to develop tests on shape restrictions such as monotonicity and convexity, based on the asymptotic chi-square distribution of the log EL ratio statistic.

In section 3, I propose a new version of the CUSUM of squares (CUSQ) test. The test is based on nonparametric residuals from a time-varying coefficient series regression model, which is featured with time-varying coefficients to accommodate potential structural changes in the (conditional) mean. I provide asymptotic properties of the proposed test, and compare my test with the CUSQ test based on least squares residuals. This comparison shows the importance of consistent estimation of the possibly mean change in a residual based test such as the CUSQ test under concern. With the presence of a time-varying mean, least squares estimation becomes inconsistent, hence a test based on the least squares residuals suffer from both unreliable size and power. On the other hand, a test based on the nonparametric local linear estimation will be robust to different structural change in the mean. The size and power properties are extensively analyzed in simulation experiments. When applying our CUSQ test to various macro and financial variables, I find important consequence of the changing mean in test of volatility change.

In section 4, I test empirically the Purchasing Power Parity (PPP) hypothesis in the absolute version, focusing on monthly data of seven developed countries in the recent float exchange rate era. Different with existing cointegration tests of PPP, most of which failed to find a cointegration relationship backed by the PPP notion, I use a nonparametric functional-coefficient cointegration model to investigate the possibility of unstable cointegration. The cointegration model adopted in this study can be viewed as a more general framework which contains the conventional linear (constant coefficient) cointegration model as a special case.

As a possible theoretical justification, my proposal of using the functional-coefficient cointegration model can be supported by the argument of nonlinear adjustment toward PPP equilibrium. Consistent with exchange rate determination theory in the presence of transactions costs, the adjustment of real exchange rate toward long-run PPP equilibrium level could be more intensive when the deviation is large enough to activate international price arbitrage. Consequently, one can expect that the cointegration relationship of the nominal exchange rate between two nations and the price levels of the two currencies could as well vary according to the intenseness of the adjustment.

To empirically verify this statement, I construct a measure of the real exchange rate volatility and use this measure to capture the degree of adjustment. By allowing the cointegration coefficients to vary as functions of the real exchange rate volatility, I find supporting evidence of PPP hypothesis in the recent float exchange rate era. Indeed, the cointegration relationship is more stable and numerically near the value implied by PPP theory when the real exchange rate volatility is low. This empirical finding can be understood as, the cointegration relationship is more obvious when the adjustment toward PPP equilibrium is more active, usually characterized by successive similar growth rates, accompanying with a lower real exchange rate volatility.

REFERENCES

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. Journal of Econometrics 116, 9-47.

Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price densities implicit in 2nancial asset prices. Journal of Finance 53, 499-547.

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

Baum, C.F., Barkoulas, J.T., Caglayan, M., 2001. Nonlinear adjustment to purchasing power parity in the post-Bretton Woods era. Journal of International Money and Finance 20, 379-399.

Cai, Z., 2001. Weighted Nadaraya-Watson regression estimation. Statistics and Probability Letters 51, 307-318.

Cai, Z., 2002. Regression quantiles for time series. Econometric Theory 18, 169-192.

Cai, Z., Hong, Y., 2003. Nonparametric methods in continuous-time finance: a selective review. In: Akritas, M.G., Politis, D.M. (Eds.), Recent Advances and Trends in Nonparametric Statistics, Elsevier, North Holland, pp. 283-302.

Cai, Z., 2007. Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163-188.

Cai, Z., Li, Q., Park, J.Y., 2009. Functional-coefficient models for nonstationary time series data. Journal of Econometrics 148, 101-133.

Cavaliere, G., 2004. Unit root tests under time-varying variances. Econometric Review 23, 259-292.

Chen B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. Econometrica (forthcoming).

Chen, S.X., 1996. Empirical likelihood confidence intervals for nonparametric density estimation. Biometrika 83, 329-341.

Chen, S.X., Qin, Y.-S., 2000. Empirical likelihood confidence interval for a local linear smoother. Biometrika 87, 946-953.

Cheng B., Tong, H., 1993. On residual sums of squares in non-parametric autoregression. Stochastic Processes and their Applications 48, 157-174.

Deng, A., Perron, P., 2008. A non-local perspective on the power properties of the CUSUM and CUSUM of squares tests for structural change. Journal of Econometrics 142, 212-240.

Deng, A., Perron, P., 2008. The limit distribution of the CUSUM of squares test under general mixing conditions. Econometric Theory 24, 809-822.

Denker, M., Keller, G., 1983. On U-Statistics and von Mises' statistics for weakly dependent processes. Zeitschrift für Wahrkeinlichkeitstheorie 64, 505-522.

Dumas, B., 1992. Dynamic equilibrium and the real exchange rate in a spatially separated world. Review of Financial Studies 5, 153-180.

Dykstra, R.L., 1983. An algorithm for restricted least squares. Journal of the American Statistical Association 78, 837-842.

Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London.

Fan, J., 2005. A selective overview of nonparametric methods in financial econometrics. Statistical Science 20, 316-354.

Fan, J., Mancini, L., 2009. Option pricing with model-guided nonparametric methods. Journal of the American Statistical Association 104, 1351-1372.

Fan, J., Zhang, W., 2008. Statistical methods with varying coefficient models. Statistics and Its Inference 1, 179-195.

Hall, P., Huang, H., 2001. Nonparametric kernel regression subject to monotonicity constraints. The Annals of Statistics 29, 624-647.

Hall, P., Owen, A.B., 1993. Empirical likelihood confidence bands in density estimation. Journal of Computational and Graphical Statistics 2, 273-289.

Hall, P., Presnell, B. 1999. Intentionally biased bootstrap methods. Journal of the Royal Statistical Society, Series B, 61, 143-158.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

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

Hansen, B., 2008. Uniform convergence rates for kernel estimation with dependent data. Econometric Theory 24, 726-748.

Henderson, D., Parmeter C., 2009. Imposing economic constraints on nonparametric regression: Survey, implementation and extensions. In: Li, Q., Racine, J. S. (Eds.), Advances in Econometrics: Nonparametric Methods, Vol. 25. Emerald Group Publishing Limited, pp. 433-469.

Inclán, C., Tiao, G., 1994. Use of the cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association 89, 913-923.

Juhl, T., Xiao, Z., 2009. Tests for changing mean with monotonic power. Journal of Econometrics 148, 14-24.

Kitamura, Y., 2007. Empirical likelihood methods in econometrics: Theory and practice. In: Blundell, R., Torsten, P., Newey, W.K. (Eds.), Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress, Vol. III. Cambridge University Press, Cambridge, pp. 174-237.

Kitamura, Y., Tripathi G., Ahn H., 2004. Empirical likelihood based inference in conditional moment restriction models, Econometrica 72, 1667-1714.

Kristensen, D., 2009. Uniform convergence rates of kernel estimators with heterogeneous dependent data. Econometric Theory 25, 1433-1445.

Kristensen, D., 2011. Nonparametric detection and estimation of structural change. CREATES research paper.

Li, Q., Racine, J., 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press, Princeton, NJ.

Masry, E., 1996. Multivariate local polynomial regression for time series: uniform strong consistency and rates. Journal of Time Series Analysis 17, 571-599.

Matzkin, R.L., 1994. Restrictions of economic theory in nonparametric methods. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics, Vol. 4. North Holland, Amsterdam, pp. 2523-2558.

Michael, P., Nobay, A.R., Peel, D.A., 1997. Transactions costs and nonlinear adjustment in real exchange rates: an empirical investigation. Journal of Political Economy 105, 862-879.

Moon, H.R., Schorfheide, F., 2009. Estimation with overidentifying inequality moment conditions. Journal of Econometrics 153, 136-154.

Newey, W.K., Smith R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators, Econometrica 72, 219-256.

Nielsen B., Sohkanen, J., 2011. Asymptotic behaviour of the CUSUM of squares test under stochastic and deterministic time trends. Econometric Theory 27, 913-927.

Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237-249.

Owen, A.B., 1990. Empirical likelihood confidence regions. Annals of Statistics 18, 90-120.

Owen, A.B., 1991. Empirical likelihood for linear models. Annals of Statistics 19, 1725-1747.

Owen, A.B., 2001. Empirical Likelihood. Chapman & Hall, New York, NY.

Pagan, A., Ullah, A., 1999. Nonparametric Econometrics. Cambridge University Press, New York, NY.

Perron, P., 2006. Dealing with structural breaks. In Patterson K., Mills T.C., (Eds.), Palgrave Handbook of Econometrics, Vol. 1: Econometric Theory, Palgrave Macmillan, pp. 278-352.

Qin, G., Tsao, M., 2005. Empirical likelihood based inference for the derivative of the nonparametric regression function. Bernoulli 11, 715-735.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300-325.

Qu, Z., Perron, P., 2007. Estimating and testing multiple structural changes in multivariate regressions. Econometrica 75, 459-502.

Racine, J.S., Parmeter, C.F., Du, P., 2009. Constrained nonparametric kernel regression: Estimation and inference. Virginia Tech AAEC Working Paper.

Rogoff, K., 1996. The purchasing power parity puzzle. Journal of Economic Literature 34, 647-668.

Sarno, L., Taylor, M.P., 2002. The Economics of Exchange Rates. Cambridge University Press, Cambridge, UK.

Sercu, P., Uppal, R., van Hulle, C., 1995. The exchange rate in the presence of transaction costs: implications for tests of purchasing power parity. Journal of Finance 50, 1309-1319.

Su, L., Xiao, Z., 2008. Testing structural change in time-series nonparametric regression models. Statistics and Its Inference 1, 347-366.

Taylor, A.M., Taylor, M.P., 2004. The purchasing power parity debate. Journal of Economic Perspectives 18, 135-158.

Xiao, Z., 2001. Testing the null hypothesis of stationarity against an autoregressvie unit root alternative. Journal of Time Series Analysis 22, 87-105.

Xiao, Z., 2009. Functional-coefficient cointegration models. Journal of Econometrics 152, 81-92.

Xu, K.-L., 2009. Empirical likelihood based inference for recurrent nonparametric diffusions. Journal of Econometrics 153, 65-82.

Xu, K.-L., 2009, Re-weighted functional estimation of diffusion models. Econometric Theory 26, 541-563.

Xu, K.-L., 2011. Testing for changing mean, non-monotonic power and testing for volatility breaks. Working paper, Department of Economics, Texas A&M University.

Xu, K.-L., 2012. Powerful tests for structural changes in volatility. Working paper, Department of Economics, Texas A&M University.

Xu, Z., 2003. Purchasing power parity, price indices, and exchange rate forcasts. Journal of International Money and Finance 22, 105-130.

Yatchew, A.J., Bos, L., 1997. Nonparametric regression and testing in economic models. Journal of Quantitative Economics 13, 81-131.

Yatchew, A., Hardle, W., 2006. Nonparametric state price density estimation using constrained least squares and the bootstrap. Journal of Econometrics 133, 579-599.

APPENDIX A

PROOFS IN SECTION 2

A.1 Proof of Lemma 2.2.1

Proof. (i) Let $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ be a saddlepoint of $G_n^*\left(\beta, \lambda, \nu\right)$ solving (2.10). First we look at the upper bounds $\bar{b}$. Suppose there is $j \in \{0, 1, 2\}$ such that $\widetilde{\beta}_j > \bar{b}_j$, then there must exist $\overline{\nu}'_j > \widetilde{\overline{\nu}}_j \geqslant 0$ such that $\overline{\nu}'_j\left(\widetilde{\beta}_j - \bar{b}_j\right) > \widetilde{\overline{\nu}}_j\left(\widetilde{\beta}_j - \bar{b}_j\right)$, so $G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\underline{\nu}}, \widetilde{\overline{\nu}}_{-j}, \overline{\nu}'_j\right) > G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$, which contradicts with the definition of $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$. Therefore $\widetilde{\beta}_j \leqslant \bar{b}_j$ for all $j = 0, 1, 2$. This implies that $\widetilde{\overline{\nu}}^{\mathsf{T}}\left(\widetilde{\beta} - \bar{b}\right) \leqslant 0$ since $\widetilde{\overline{\nu}} \geqslant 0$. Further, if $\widetilde{\beta}_j < \bar{b}_j$, then $\widetilde{\overline{\nu}}_j = 0$. Together we have $\widetilde{\overline{\nu}}^{\mathsf{T}}\left(\widetilde{\beta} - \bar{b}\right) = 0$. Similarly we can show that $\widetilde{\underline{\nu}}^{\mathsf{T}}\left(\underline{b} - \widetilde{\beta}\right) = 0$. So

$$G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right) = G_n\left(\widetilde{\beta}, \widetilde{\lambda}\right) \geqslant G_n\left(\beta, \lambda\right)$$

for any $\beta \in \left(\underline{b}, \bar{b}\right)$ and $\lambda \in \Lambda$. That is, $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ is a saddlepoint of $G_n\left(\beta, \lambda\right)$ that solves (2.9).

(ii) Let $\left(\widetilde{\beta}, \widetilde{\lambda}\right)$ be a saddlepoint of $G_n\left(\beta, \lambda\right)$ solving (2.9) and $\widetilde{\nu}$ as defined in the lemma. We want to show that $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ is a saddlepoint of (2.10). First, since $\widetilde{\overline{\nu}}^{\mathsf{T}}\left(\widetilde{\beta} - \bar{b}\right) = 0$ and $\widetilde{\underline{\nu}}^{\mathsf{T}}\left(\underline{b} - \widetilde{\beta}\right) = 0$ by the definition of $\widetilde{\nu}$, also since $\widetilde{\beta} \in \left[\underline{b}, \bar{b}\right]$, we have $\widetilde{\overline{\nu}}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right) \geqslant \overline{\nu}^{\intercal}\left(\widetilde{\beta}-\overline{b}\right)$ and $\widetilde{\underline{\nu}}^{\intercal}\left(\underline{b}-\widetilde{\beta}\right) \geqslant \underline{\nu}^{\intercal}\left(\underline{b}-\widetilde{\beta}\right)$ for any $\nu = (\underline{\nu}^{\intercal}, \overline{\nu}^{\intercal})^{\intercal} \in \mathbb{R}_{+}^{6}$, hence

$$G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right) \geqslant G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \nu\right). \tag{A.1}$$

Second, let $\widetilde{\beta} = \left(\widetilde{\beta}_j, \widetilde{\beta}_{-j}\right)$ such that $\widetilde{\beta}_j = \overline{b}_j$ (or $\widetilde{\beta}_j = \underline{b}_j$) and $\widetilde{\beta}_{-j} \in \left(\underline{b}_{-j}, \overline{b}_{-j}\right)$. Then for any $\beta \in \left[\underline{b}, \overline{b}\right]$, we make the same partition $\beta = (\beta_j, \beta_{-j})$ and have

$$G_n^*\left(\beta, \widetilde{\lambda}, \widetilde{\nu}\right) = G_n^*\left(\beta_j, \beta_{-j}, \widetilde{\lambda}, \widetilde{\nu}\right) = G_n^*\left(\beta_j, \widetilde{\beta}_{-j}, \widetilde{\lambda}, \widetilde{\nu}\right),$$

where the second equality holds because by definition the part in $\widetilde{\nu}$ corresponding to β_{-j} are zeros. Further, we have

$$G_n^*\left(\beta_j, \widetilde{\beta}_{-j}, \widetilde{\lambda}, \widetilde{\nu}\right) \geqslant G_n^*\left(\widetilde{\beta}_j, \widetilde{\beta}_{-j}, \widetilde{\lambda}, \widetilde{\nu}\right)$$

since $G_n^*(\beta, \lambda, \nu)$ is globally convex in β_j, and by definition of $\widetilde{\nu}$,

$$\frac{\partial G_n^*(\beta, \lambda, \nu)}{\partial \beta_j}\Big|_{\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}} = 0.$$

Together we have

$$G_n^*\left(\beta, \widetilde{\lambda}, \widetilde{\nu}\right) \geqslant G_n^*\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right). \tag{A.2}$$

Finally, (A.1) and (A.2) imply that $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ is a saddlepoint of (2.10). ∎

A.2 Proof of Lemma 2.3.1

For the general local polynomial estimators, the asymptotic conditional bias and variance terms are discussed in Fan and Gijbels (1996), Theorem 3.1. Following their notations, we denote, in the case of local quadratic estimator,

$$S = \begin{pmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{pmatrix}, \ S^* = \begin{pmatrix} \nu_0 & \nu_1 & \nu_2 \\ \nu_1 & \nu_2 & \nu_3 \\ \nu_2 & \nu_3 & \nu_4 \end{pmatrix}, \ c_2 = \begin{pmatrix} \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}, \ \widetilde{c}_2 = \begin{pmatrix} \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix},$$

where $\mu_j = \int u^j K(u)\, du$, $\nu_j = \int u^j K^2(u)\, du$. Note that $\mu_0 = 1$, and for a symmetric kernel, $\mu_1 = \mu_3 = \mu_5 = \nu_1 = \nu_3 = 0$. Then the asymptotic bias is given by

$$Bias\left(\widehat{\beta}_j(x)\,|\mathbf{X}\right) = e_{j+1}^{\intercal} S^{-1} c_2 \frac{m^{(3)}(x)}{6} h^{3-j} + o_p\left(h^{3-j}\right)$$

for $j = 1$, and

$$Bias\left(\widehat{\beta}_j(x)\,|\mathbf{X}\right) = e_{j+1}^{\intercal} S^{-1} \widetilde{c}_2 \frac{1}{24}\left(m^{(4)}(x) + 4m^{(3)}(x)\frac{f^{(1)}(x)}{f(x)}\right) h^{4-j} + o_p\left(h^{4-j}\right)$$

for $j = 0, 2$. The asymptotic variances are given by

$$Var\left(\widehat{\beta}_j(x)\,|\mathbf{X}\right) = e_{j+1}^{\intercal} S^{-1} S^* S^{-1} e_{j+1} \frac{\sigma^2(x)}{f(x)\, nh^{1+2j}} + o_p\left(\frac{1}{nh^{1+2j}}\right)$$

for $j = 0, 1, 2$. It is known that the leading term in the asymptotic bias is of a smaller order for j being even than in the case for j being odd. Explicitly, we have

$$
\begin{aligned}
Bias\left(\widehat{\beta}_0(x)\,|\mathbf{X}\right) &= \frac{h^4}{24}\frac{\mu_4^2-\mu_2\mu_6}{\mu_4-\mu_2^2}\left(m^{(4)}(x)+4m^{(3)}(x)\frac{f^{(1)}(x)}{f(x)}\right)+o_p\left(h^4\right),\\
Bias\left(\widehat{\beta}_1(x)\,|\mathbf{X}\right) &= \frac{h^2}{6}\frac{\mu_4}{\mu_2}m^{(3)}(x)+o_p\left(h^2\right),\\
Bias\left(\widehat{\beta}_2(x)\,|\mathbf{X}\right) &= \frac{h^2}{24}\frac{\mu_6-\mu_2\mu_4}{\mu_4-\mu_2^2}\left(m^{(4)}(x)+4m^{(3)}(x)\frac{f^{(1)}(x)}{f(x)}\right)+o_p\left(h^2\right),\\
Var\left(\widehat{\beta}_0(x)\,|\mathbf{X}\right) &= \frac{1}{nh}\frac{\mu_4^2\nu_0-2\mu_2\mu_4\nu_2+\mu_2^2\nu_4}{\left(\mu_4-\mu_2^2\right)^2}\frac{\sigma^2(x)}{f(x)}+o_p\left(\frac{1}{nh}\right),\\
Var\left(\widehat{\beta}_1(x)\,|\mathbf{X}\right) &= \frac{1}{nh^3}\frac{\nu_2}{\mu_2^2}\frac{\sigma^2(x)}{f(x)}+o_p\left(\frac{1}{nh^3}\right),\\
Var\left(\widehat{\beta}_2(x)\,|\mathbf{X}\right) &= \frac{1}{nh^5}\frac{\mu_2^2\nu_0-2\mu_2\nu_2+\nu_4}{\left(\mu_4-\mu_2^2\right)^2}\frac{\sigma^2(x)}{f(x)}+o_p\left(\frac{1}{nh^5}\right).
\end{aligned}
$$

To derive the asymptotic distribution for the estimating equations, we need to introduce more notations. Let $S^{-1} = T/D$, where

$$
T = \begin{pmatrix} t_0 & t_1 & t_2 \\ t_1 & t_3 & t_4 \\ t_2 & t_4 & t_5 \end{pmatrix} = \begin{pmatrix} \mu_2\mu_4-\mu_3^2 & \mu_2\mu_3-\mu_1\mu_4 & \mu_1\mu_3-\mu_2^2 \\ \mu_2\mu_3-\mu_1\mu_4 & \mu_0\mu_4-\mu_2^2 & \mu_1\mu_2-\mu_0\mu_3 \\ \mu_1\mu_3-\mu_2^2 & \mu_1\mu_2-\mu_0\mu_3 & \mu_0\mu_2-\mu_1^2 \end{pmatrix},
$$

$$
D = \det(S) = \mu_0\left(\mu_2\mu_4-\mu_3^2\right)-\mu_1\left(\mu_1\mu_4-\mu_2\mu_3\right)-\mu_2\left(\mu_2^2-\mu_1\mu_3\right),
$$

then

$$
S^{-1}S^*S^{-1} = \frac{1}{D^2}TS^*T.
$$

Note that we have already denoted

$$D_n = \frac{1}{nh}\sum\nolimits_{i=1}^{n} W_{0i}\left(x\right)$$
$$= s_0\left(s_2 s_4 - s_3^2\right) - s_1\left(s_1 s_4 - s_2 s_3\right) - s_2\left(s_2^2 - s_1 s_3\right),$$

thus we have $D_n \xrightarrow{p} f^3\left(x\right) D$ because $s_j \xrightarrow{p} f\left(x\right)\mu_j$ for $j = 0, 1, 2, 3, 4$.

Evaluated at the true values $\left(m_0\left(x\right), m_1\left(x\right), m_2\left(x\right)\right)^{\intercal}$, the three estimating equations are

$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}\left(m\right) = \left(\widehat{\beta}_0\left(x\right) - m_0\left(x\right)\right) D_n,$$
$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{1i}\left(m\right) = \left(\widehat{\beta}_1\left(x\right) - m_1\left(x\right)\right) h D_n,$$
$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{2i}\left(m\right) = \left(\widehat{\beta}_2\left(x\right) - m_2\left(x\right)\right) h^2 D_n,$$

so we can derive that, by assuming $h \to 0$, $nh \to \infty$, $nh^7 \to 0$ as $n \to \infty$,

$$\sqrt{nh}\left(\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i\left(m\right) - \frac{h^3}{6} m^{(3)}\left(x\right) f^3\left(x\right) T c_2\right) \xrightarrow{d} N\left(0, \sigma^2\left(x\right) f^5\left(x\right) T S^* T\right),$$

where

$$Tc_2 = \begin{pmatrix} \mu_3\left(\mu_2\mu_4 - \mu_3^2\right) - \mu_4\left(\mu_1\mu_4 - \mu_2\mu_3\right) - \mu_5\left(\mu_2^2 - \mu_1\mu_3\right) \\ \mu_3\left(\mu_2\mu_3 - \mu_1\mu_4\right) - \mu_4\left(\mu_2^2 - \mu_0\mu_4\right) - \mu_5\left(\mu_0\mu_3 - \mu_1\mu_2\right) \\ \mu_3\left(\mu_1\mu_3 - \mu_2^2\right) - \mu_4\left(\mu_0\mu_3 - \mu_1\mu_2\right) - \mu_5\left(\mu_1^2 - \mu_0\mu_2\right) \end{pmatrix},$$

$$TS^*T = \begin{pmatrix} \omega_0 & \omega_1 & \omega_2 \\ \omega_1 & \omega_3 & \omega_4 \\ \omega_2 & \omega_4 & \omega_5 \end{pmatrix},$$

and

$$\omega_0 = t_0^2\nu_0 + 2t_0t_1\nu_1 + \left(2t_0t_2 + t_1^2\right)\nu_2 + 2t_1t_2\nu_3 + t_2^2\nu_4,$$

$$\omega_1 = t_0t_1\nu_0 + \left(t_0t_3 + t_1^2\right)\nu_1 + \left(t_0t_4 + t_1t_2 + t_1t_3\right)\nu_2 + \left(t_1t_4 + t_2t_3\right)\nu_3 + t_2t_4\nu_4,$$

$$\omega_2 = t_0t_2\nu_0 + \left(t_0t_4 + t_1t_2\right)\nu_1 + \left(t_0t_5 + t_1t_4 + t_2^2\right)\nu_2 + \left(t_2t_4 + t_1t_5\right)\nu_3 + t_2t_5\nu_4,$$

$$\omega_3 = t_1^2\nu_0 + 2t_1t_3\nu_1 + \left(2t_1t_4 + t_3^2\right)\nu_2 + 2t_3t_4\nu_3 + t_4^2\nu_4,$$

$$\omega_4 = t_1t_2\nu_0 + \left(t_1t_4 + t_2t_3\right)\nu_1 + \left(t_1t_5 + t_2t_4 + t_3t_4\right)\nu_2 + \left(t_3t_5 + t_4^2\right)\nu_3 + t_4t_5\nu_4,$$

$$\omega_5 = t_2^2\nu_0 + 2t_2t_4\nu_1 + \left(2t_2t_5 + t_4^2\right)\nu_2 + 2t_4t_5\nu_3 + t_5^2\nu_4.$$

For a symmetric kernel $K(\cdot)$, remind that $\mu_1 = \mu_3 = \mu_5 = \nu_1 = \nu_3 = 0$, so $t_0 = \mu_2\mu_4$, $t_2 = -\mu_2^2$, $t_3 = \mu_4 - \mu_2^2$, $t_5 = \mu_2$, $t_1 = t_4 = 0$, so

$$Tc_2 = \begin{pmatrix} 0 \\ \mu_4^2 - \mu_2^2\mu_4 \\ 0 \end{pmatrix}, \; TS^*T = \begin{pmatrix} \omega_0 & 0 & \omega_2 \\ 0 & \omega_3 & 0 \\ \omega_2 & 0 & \omega_5 \end{pmatrix},$$

where

$$\begin{aligned}
\omega_0 &= \mu_2^2\left(\mu_4^2\nu_0 - 2\mu_2\mu_4\nu_2 + \mu_2^2\nu_4\right), \\
\omega_2 &= \mu_2^2\left(-\mu_2\mu_4\nu_0 + \left(\mu_4 + \mu_2^2\right)\nu_2 - \mu_2\nu_4\right), \\
\omega_3 &= \mu_2^2\nu_2\left(\mu_4/\mu_2 - \mu_2\right)^2, \\
\omega_5 &= \mu_2^2\left(\mu_2^2\nu_0 - 2\mu_2\nu_2 + \nu_4\right).
\end{aligned}$$

A.3 Proof of Lemma 2.3.2

Lemma 2.3.2 states the stochastic order for the squared sums of $U_i(m)$. The proof is similar as that of Lemma 2 in Qin and Tsao (2005). Using the same notation as in Section 2.2, we let $K_i = K\left(\left(X_i - x\right)/h\right)$, and

$$\begin{aligned}
W_{0i}(x) &= \left[\left(s_2 s_4 - s_3^2\right) - \left(s_1 s_4 - s_2 s_3\right)\left(\frac{X_i - x}{h}\right) - \left(s_2^2 - s_1 s_3\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i \\
&= \left[T_0 + T_1\left(\frac{X_i - x}{h}\right) + T_2\left(\frac{X_i - x}{h}\right)^2\right] K_i, \\
W_{1i}(x) &= \left[\left(s_2 s_3 - s_1 s_4\right) - \left(s_2^2 - s_0 s_4\right)\left(\frac{X_i - x}{h}\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i \\
&= \left[T_1 + T_3\left(\frac{X_i - x}{h}\right) + T_4\left(\frac{X_i - x}{h}\right)^2\right] K_i, \\
W_{2i}(x) &= \left[\left(s_1 s_3 - s_2^2\right) - \left(s_0 s_3 - s_1 s_2\right)\left(\frac{X_i - x}{h}\right) - \left(s_1^2 - s_0 s_2\right)\left(\frac{X_i - x}{h}\right)^2\right] K_i \\
&= \left[T_2 + T_4\left(\frac{X_i - x}{h}\right) + T_5\left(\frac{X_i - x}{h}\right)^2\right] K_i,
\end{aligned}$$

then for $j = 0, 1, 2$,

$$\sum\nolimits_{i=1}^{n} U_{ji}(m) = \sum\nolimits_{i=1}^{n} W_{ji}(x)\left[Y_i - m_j(x)\left(X_i - x\right)^j\right].$$

The conclusion in Lemma 2.3.2 can be verified as follows. Under Assumption 2.3.1, 2.3.2, and 2.3.3, we state the follow Lemma A.3.1-Lemma A.3.6.

Lemma A.3.1. $\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}^2(m_0) = \sigma^2(x) f^3(x)\omega_0 + o_p(1)$.

Proof. Write

$$\begin{aligned}
\frac{1}{nh}\sum U_{0i}^{2}\left(m_{0}\right) &= \frac{1}{nh}\sum W_{0i}^{2}\left(x\right)\left[Y_{i}-m\left(x\right)\right]^{2} \\
&= \frac{1}{nh}\sum W_{0i}^{2}\left(x\right)\left[Y_{i}-m\left(X_{i}\right)\right]^{2} \\
&+ \frac{2}{nh}\sum W_{0i}^{2}\left(x\right)\left[Y_{i}-m\left(X_{i}\right)\right]\left[m\left(X_{i}\right)-m\left(x\right)\right] \\
&+ \frac{1}{nh}\sum W_{0i}^{2}\left(x\right)\left[m\left(X_{i}\right)-m\left(x\right)\right]^{2} \\
&= J_{1}+2J_{2}+J_{3}.
\end{aligned}$$

First,

$$\begin{aligned}
J_{1} &= T_{0}^{2}\frac{1}{nh}\sum K_{i}^{2}\sigma^{2}\left(X_{i}\right)u_{i}^{2}+2T_{0}T_{1}\frac{1}{nh}\sum\left(\frac{X_{i}-x}{h}\right)K_{i}^{2}\sigma^{2}\left(X_{i}\right)u_{i}^{2} \\
&+ \left(2T_{0}T_{2}+T_{1}^{2}\right)\frac{1}{nh}\sum\left(\frac{X_{i}-x}{h}\right)^{2}K_{i}^{2}\sigma^{2}\left(X_{i}\right)u_{i}^{2} \\
&+ 2T_{1}T_{2}\frac{1}{nh}\sum\left(\frac{X_{i}-x}{h}\right)^{3}K_{i}^{2}\sigma^{2}\left(X_{i}\right)u_{i}^{2}+T_{2}^{2}\frac{1}{nh}\sum\left(\frac{X_{i}-x}{h}\right)^{4}K_{i}^{2}\sigma^{2}\left(X_{i}\right)u_{i}^{2},
\end{aligned}$$

since for $j=0,1,2,3,4,$

$$\begin{aligned}
\frac{1}{nh}\sum\left(\frac{X_{i}-x}{h}\right)^{j}K_{i}^{2}\sigma^{2}\left(X_{i}\right)u_{i}^{2} &= E\left[\left(\frac{X_{1}-x}{h}\right)^{j}\frac{1}{h}K_{1}^{2}\sigma^{2}\left(X_{1}\right)u_{1}^{2}\right]+o_{p}\left(1\right) \\
&= \sigma^{2}\left(x\right)f\left(x\right)\int u^{j}K^{2}\left(u\right)du+o_{p}\left(1\right) \\
&= \sigma^{2}\left(x\right)f\left(x\right)\nu_{j}+o_{p}\left(1\right),
\end{aligned}$$

and for $j = 0, 1, 2, 3, 4,$

$$T_j = f^2(x)\, t_j + o_p(1),$$

so

$$\begin{aligned} J_1 &= \sigma^2(x)\, f^3(x)\left[t_0^2\nu_0 + 2t_0t_1\nu_1 + \left(2t_0t_2 + t_1^2\right)\nu_2 + 2t_1t_2\nu_3 + t_2^2\nu_4\right] + o_p(1) \\ &= \sigma^2(x)\, f^3(x)\, \omega_0 + o_p(1). \end{aligned}$$

Second, $J_2 = o_p(1)$ since for $j = 0, 1, 2, 3, 4,$

$$\begin{aligned} &\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 \sigma(X_i)\, u_i\left(m(X_i) - m(x)\right) \\ &= E\left[\left(\frac{X_1 - x}{h}\right)^j \frac{1}{h} K_1^2 \sigma(X_1)\, u_1\left(m(X_1) - m(x)\right)\right] + o_p(1) \\ &= E\left[\left(\frac{X_1 - x}{h}\right)^j \frac{1}{h} K_1^2 \sigma(X_1)\, E(u_1|X_1)\left(m(X_1) - m(x)\right)\right] + o_p(1) \\ &= o_p(1). \end{aligned}$$

Third, $J_3 = o_p(1)$ since for $j = 0, 1, 2, 3, 4,$

$$\begin{aligned} &\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2\left[m(X_i) - m(x)\right]^2 \\ &= \frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2\left[m^{(1)}(x)\,(X_i - x) + o_p(h)\right]^2 \end{aligned}$$

$$= \frac{\left(m^{(1)}(x)\right)^2 h}{n} \sum \left(\frac{X_i - x}{h}\right)^{j+2} K_i^2 + o_p(h) \frac{2m^{(1)}(x)}{n} \sum \left(\frac{X_i - x}{h}\right)^{j+1} K_i^2$$

$$+ o_p\left(h^2\right) \frac{1}{nh} \sum \left(\frac{X_i - x}{h}\right)^{j} K_i^2$$

$$= \left(m^{(1)}(x)\right)^2 h^2 \left[\nu_{j+2} f(x) + o_p(1)\right] + o_p(h)\, 2m^{(1)}(x)\, h \left[\nu_{j+1} f(x) + o_p(1)\right]$$

$$+ o_p\left(h^2\right) \left[\nu_j f(x) + o_p(1)\right]$$

$$= O_p\left(h^2\right) + o_p\left(h^2\right) = o_p(1).$$

where the first equality is because the kernel function is bounded in $[-1, 1]$. ■

Lemma A.3.2. $\frac{1}{nh}\sum_{i=1}^{n} U_{1i}^2(m_1) = \left[\sigma^2(x) + m^2(x)\right] f^3(x)\, \omega_3 + o_p(1)$.

Proof. Write

$$\begin{aligned}
\frac{1}{nh}\sum U_{1i}^2(m_1) &= \frac{1}{nh}\sum W_{1i}^2(x)\left[Y_i - m^{(1)}(x)(X_i - x)\right]^2 \\
&= \frac{1}{nh}\sum W_{1i}^2(x)\left[Y_i - m(X_i)\right]^2 \\
&\quad + \frac{2}{nh}\sum W_{1i}^2(x)\left[Y_i - m(X_i)\right]\left[m(X_i) - m^{(1)}(x)(X_i - x)\right] \\
&\quad + \frac{1}{nh}\sum W_{1i}^2(x)\left[m(X_i) - m^{(1)}(x)(X_i - x)\right]^2 \\
&= J_1 + 2J_2 + J_3,
\end{aligned}$$

where $J_1 = \sigma^2(x) f^3(x)\omega_3 + o_p(1)$, $J_2 = o_p(1)$ because of similar proof for corresponding parts in Lemma A.3.1. Next,

$$\begin{aligned}
J_3 &= \frac{m^2(x)}{nh}\sum W_{1i}^2(x) + \frac{2m(x)}{nh}\sum W_{1i}^2(x)\left[m(X_i) - m(x) - m^{(1)}(x)(X_i - x)\right] \\
&\quad + \frac{1}{nh}\sum W_{1i}^2(x)\left[m(X_i) - m(x) - m^{(1)}(x)(X_i - x)\right]^2 \\
&= m^2(x) J_4 + 2m(x) J_5 + J_6,
\end{aligned}$$

where $m^2(x) J_4 = m^2(x) f^3(x)\omega_3 + o_p(1)$, since for $j = 0, 1, 2, 3, 4,$

$$\frac{1}{nh}\sum\left(\frac{X_i - x}{h}\right)^j K_i^2 = E\left[\left(\frac{X_1 - x}{h}\right)^j \frac{1}{h} K_1^2\right] + o_p(1) = f(x)\nu_j + o_p(1),$$

and for $j = 0, 1, 2, 3, 4,$

$$T_j = f^2(x) t_j + o_p(1).$$

Also,

$$\begin{aligned}
J_5 &= \frac{1}{nh}\sum W_{1i}^2(x)\left[\frac{1}{2}m^{(2)}(x)(X_i - x)^2 + o_p\left(h^2\right)\right] \\
&= \frac{m^{(2)}(x)}{2nh}\sum W_{1i}^2(x)(X_i - x)^2 + o_p\left(h^2\right) J_4 \\
&= O_p\left(h^2\right) + o_p\left(h^2\right) = o_p(1), \\
J_6 &= \frac{1}{nh}\sum W_{1i}^2(x)\left[\frac{1}{2}m^{(2)}(x)(X_i - x)^2 + o_p\left(h^2\right)\right]^2 \\
&= O_p\left(h^4\right) + o_p\left(h^4\right) = o_p(1).
\end{aligned}$$

■

Lemma A.3.3. $\frac{1}{nh}\sum_{i=1}^{n} U_{2i}^2(m_2) = [\sigma^2(x) + m^2(x)] f^3(x)\omega_5 + o_p(1)$.

Proof. Write

$$\begin{aligned}
\frac{1}{nh}\sum U_{2i}^2(m_2) &= \frac{1}{nh}\sum W_{2i}^2(x)\left[Y_i - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right]^2 \\
&= \frac{1}{nh}\sum W_{2i}^2(x)[Y_i - m(X_i)]^2 \\
&+ \frac{2}{nh}\sum W_{2i}^2(x)[Y_i - m(X_i)]\left[m(X_i) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right] \\
&+ \frac{1}{nh}\sum W_{2i}^2(x)\left[m(X_i) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right]^2 \\
&= J_1 + 2J_2 + J_3,
\end{aligned}$$

where $J_1 = \sigma^2(x) f^3(x)\omega_5 + o_p(1)$, $J_2 = o_p(1)$ because of similar proof for corresponding parts in Lemma A.3.1. Next, let

$$\begin{aligned}
A_1 &= m(X_i) - m(x) - m^{(1)}(x)(X_i - x) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2 = o_p\left(h^2\right), \\
A_2 &= m(x) + m^{(1)}(x)(X_i - x),
\end{aligned}$$

then

$$\begin{aligned}
J_3 &= \frac{1}{nh}\sum W_{2i}^2(x) A_1^2 + \frac{2}{nh}\sum W_{2i}^2(x) A_1 A_2 + \frac{1}{nh}\sum W_{2i}^2(x) A_2^2 \\
&= J_4 + 2J_5 + J_6,
\end{aligned}$$

where

$$
\begin{aligned}
J_4 &= \frac{1}{nh}\sum W_{2i}^2(x)\left[o_p\left(h^2\right)\right]^2 = o_p\left(h^4\right),\\
J_5 &= \frac{m(x)}{nh}\sum W_{2i}^2(x)\left[o_p\left(h^2\right)\right] + \frac{m^{(1)}(x)}{nh}\sum W_{2i}^2(x)\left[o_p\left(h^2\right)(X_i - x)\right]\\
&= o_p\left(h^2\right) + o_p\left(h^3\right),\\
J_6 &= \frac{m^2(x)}{nh}\sum W_{2i}^2(x) + \frac{2m(x)\,m^{(1)}(x)}{nh}\sum W_{2i}^2(x)(X_i - x)\\
&+ \frac{\left[m^{(1)}(x)\right]^2}{nh}\sum W_{2i}^2(x)(X_i - x)^2\\
&= J_7 + O_p(h) + O_p\left(h^2\right),
\end{aligned}
$$

and $J_7 = m^2(x)\, f^3(x)\,\omega_5 + o_p(1)$ as J_4 in Lemma A.3.2. So $J_3 = m^2(x)\, f^3(x)\,\omega_5 + o_p(1)$. ■

Lemma A.3.4. $\frac{1}{nh}\sum_{i=1}^{n} U_{0i}(m_0)\, U_{1i}(m_1) = \sigma^2(x)\, f^3(x)\,\omega_1 + o_p(1)$.

Proof. Write

$$\begin{aligned}
&\frac{1}{nh}\sum U_{0i}\left(m_0\right)U_{1i}\left(m_1\right)\\
&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[Y_i-m\left(x\right)\right]\left[Y_i-m^{(1)}\left(x\right)\left(X_i-x\right)\right]\\
&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[Y_i-m\left(X_i\right)\right]^2\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[Y_i-m\left(X_i\right)\right]\left[m\left(X_i\right)-m^{(1)}\left(x\right)\left(X_i-x\right)\right]\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[Y_i-m\left(X_i\right)\right]\left[m\left(X_i\right)-m\left(x\right)\right]\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[m\left(X_i\right)-m\left(x\right)\right]\left[m\left(X_i\right)-m^{(1)}\left(x\right)\left(X_i-x\right)\right]\\
&=J_1+J_2+J_3+J_4,
\end{aligned}$$

where $J_1=\sigma^2\left(x\right)f^3\left(x\right)\omega_1+o_p\left(1\right)$, $J_2=J_3=o_p\left(1\right)$ because of similar proof for corresponding parts in Lemma A.3.1, and

$$\begin{aligned}
J_4&=\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[m\left(X_i\right)-m\left(x\right)\right]^2\\
&+\frac{1}{nh}\sum W_{0i}\left(x\right)W_{1i}\left(x\right)\left[m\left(X_i\right)-m\left(x\right)\right]\left[m\left(x\right)-m^{(1)}\left(x\right)\left(X_i-x\right)\right]\\
&=J_{41}+J_{42},
\end{aligned}$$

where $J_{41} = o_p(1)$ as J_3 in Lemma A.3.1, and

$$\begin{aligned}
J_{42} &= \frac{1}{nh}\sum W_{0i}(x)\,W_{1i}(x)\left[m^{(1)}(x)(X_i - x) + o_p(h)\right]\left[m(x) - m^{(1)}(x)(X_i - x)\right] \\
&= \frac{m(x)}{nh}\sum W_{0i}(x)\,W_{1i}(x)\left[m^{(1)}(x)(X_i - x) + o_p(h)\right] \\
&- \frac{m^{(1)}(x)}{nh}\sum W_{0i}(x)\,W_{1i}(x)\left[m^{(1)}(x)(X_i - x) + o_p(h)\right](X_i - x) \\
&= O_p(h) + o_p(h) + O_p\left(h^2\right) + o_p\left(h^2\right) = o_p(1).
\end{aligned}$$

■

Lemma A.3.5. $\frac{1}{nh}\sum_{i=1}^{n} U_{0i}(m_0)\,U_{2i}(m_2) = \sigma^2(x)\,f^3(x)\,\omega_2 + o_p(1)$.

Proof. Write

$$\begin{aligned}
&\frac{1}{nh}\sum U_{0i}(m_0)\,U_{2i}(m_2) \\
&= \frac{1}{nh}\sum W_{0i}(x)\,W_{2i}(x)\left[Y_i - m(x)\right]\left[Y_i - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right] \\
&= \frac{1}{nh}\sum W_{0i}(x)\,W_{2i}(x)\left[Y_i - m(X_i)\right]^2 \\
&+ \frac{1}{nh}\sum W_{0i}(x)\,W_{2i}(x)\left[Y_i - m(X_i)\right]\left[m(X_i) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right] \\
&+ \frac{1}{nh}\sum W_{0i}(x)\,W_{2i}(x)\left[Y_i - m(X_i)\right]\left[m(X_i) - m(x)\right] \\
&+ \frac{1}{nh}\sum W_{0i}(x)\,W_{2i}(x)\left[m(X_i) - m(x)\right]\left[m(X_i) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right] \\
&= J_1 + J_2 + J_3 + J_4,
\end{aligned}$$

where $J_1 = \sigma^2(x) f^3(x)\omega_2 + o_p(1)$, $J_2 = J_3 = o_p(1)$ because of similar proof for corresponding parts in Lemma A.3.1, and

$$\begin{aligned} J_4 &= \frac{1}{nh}\sum W_{0i}(x) W_{2i}(x)\left[m(X_i) - m(x)\right]^2 \\ &+ \frac{1}{nh}\sum W_{0i}(x) W_{2i}(x)\left[m(X_i) - m(x)\right]\left[m(x) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right] \\ &= J_{41} + J_{42}, \end{aligned}$$

where $J_{41} = o_p(1)$ as J_3 in Lemma A.3.1, and

$$\begin{aligned} J_{42} &= \frac{1}{nh}\sum W_{0i}(x) W_{2i}(x)\left[m^{(1)}(x)(X_i - x) + o_p(h)\right]\left[m(x) - \frac{1}{2}m^{(2)}(x)(X_i - x)^2\right] \\ &= \frac{m(x)}{nh}\sum W_{0i}(x) W_{2i}(x)\left[m^{(1)}(x)(X_i - x) + o_p(h)\right] \\ &- \frac{m^{(2)}(x)}{2nh}\sum W_{0i}(x) W_{2i}(x)\left[m^{(1)}(x)(X_i - x) + o_p(h)\right](X_i - x)^2 \\ &= O_p(h) + o_p(h) + O_p\left(h^3\right) + o_p\left(h^3\right) = o_p(1). \end{aligned}$$

■

Lemma A.3.6. $\frac{1}{nh}\sum_{i=1}^{n} U_{1i}(m_1) U_{2i}(m_2) = \left[\sigma^2(x) + m^2(x)\right] f^3(x)\omega_4 + o_p(1)$.

Proof. Write

$$
\begin{aligned}
&\frac{1}{nh}\sum U_{1i}(m_1)\,U_{2i}(m_2)\\
&=\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[Y_i-m^{(1)}(x)(X_i-x)\right]\left[Y_i-\frac{1}{2}m^{(2)}(x)(X_i-x)^2\right]\\
&=\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[Y_i-m(X_i)\right]^2\\
&+\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[Y_i-m(X_i)\right]\left[m(X_i)-\frac{1}{2}m^{(2)}(x)(X_i-x)^2\right]\\
&+\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[Y_i-m(X_i)\right]\left[m(X_i)-m^{(1)}(x)(X_i-x)\right]\\
&+\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[m(X_i)-m^{(1)}(x)(X_i-x)\right]\left[m(X_i)-\frac{1}{2}m^{(2)}(x)(X_i-x)^2\right]\\
&=J_1+J_2+J_3+J_4,
\end{aligned}
$$

where $J_1=\sigma^2(x)\,f^3(x)\,\omega_4+o_p(1)$, $J_2=J_3=o_p(1)$ because of similar proof for corresponding parts in Lemma A.3.1, and

$$
\begin{aligned}
J_4&=\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[m(X_i)-m(x)-m^{(1)}(x)(X_i-x)+m(x)\right]\\
&\quad\left[m(X_i)-m(x)-m^{(1)}(x)(X_i-x)-\frac{1}{2}m^{(2)}(x)(X_i-x)^2+m(x)+m^{(1)}(x)(X_i-x)\right]\\
&=\frac{1}{nh}\sum W_{1i}(x)\,W_{2i}(x)\left[o_p(h)+m(x)\right]\left[o_p\left(h^2\right)+m(x)+m^{(1)}(x)(X_i-x)\right]\\
&=\frac{m^2(x)}{nh}\sum W_{1i}(x)\,W_{2i}(x)+\frac{m(x)\,m^{(1)}(x)}{nh}\sum W_{1i}(x)\,W_{2i}(x)(X_i-x)+o_p(h)\\
&=J_5+O_p(h)+o_p(h),
\end{aligned}
$$

where $J_5=m^2(x)\,f^3(x)\,\omega_4+o_p(1)$ as J_4 in Lemma A.3.2. ■

A.4 Proof of Theorem 2.3.1

Proof. Write $\lambda(m) = \rho\theta$ where $\rho \geqslant 0$ and $\|\theta\| = 1$. Also we need to denote $\overline{U} = \sum_{j=0}^{2}\left[\frac{1}{nh}\sum_{i=1}^{n} U_{ji}(m_j)\right]$ for $j = 0, 1, 2$, and $\overline{\overline{U}} = \frac{1}{nh}\sum_{i=1}^{n} U_i(m)\, U_i(m)^{\intercal}$. Note that

$$p_i = \frac{1}{n\left(1+\rho\theta^{\intercal}U_i(m)\right)} \in [0,1],$$

from which we have $1+\rho\theta^{\intercal}U_i(m) > 0$. From the three EL weighted estimating equations, $\sum_{i=1}^{n} p_i U_i(m) = 0$, we have

$$\begin{aligned}
0 &= \left\|\frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(m)}{1+\rho\theta^{\intercal}U_i(m)}\right\| \\
&\geqslant \left|\frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{\theta^{\intercal}U_i(m)}{1+\rho\theta^{\intercal}U_i(m)}\right| \\
&= \left|\theta^{\intercal}\left(\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m) - \rho\frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(m)\left[\theta^{\intercal}U_i(m)\right]}{1+\rho\theta^{\intercal}U_i(m)}\right)\right| \\
&\geqslant \rho\theta^{\intercal}\left(\frac{1}{nh}\sum\nolimits_{i=1}^{n} \frac{U_i(m)\,U_i(m)^{\intercal}}{1+\rho\theta^{\intercal}U_i(m)}\right)\theta - \left|\overline{U}\right| \\
&\geqslant \frac{\rho}{1+\rho Z_n}\theta^{\intercal}\overline{\overline{U}}\theta - \left|\overline{U}\right|,
\end{aligned}$$

where in the right hand side of the last inequality, $Z_n = \max\limits_{1\leqslant i\leqslant n}\|U_i(m)\|$ so $Z_n \geqslant \theta^{\intercal}U_i(m)$ for each i. Therefore

$$\frac{\rho}{1+\rho Z_n}\theta^{\intercal}\overline{\overline{U}}\theta \leqslant \left|\overline{U}\right|$$

implies

$$\rho\left(\theta^{\intercal}\overline{\overline{U}}\theta - Z_n\left|\overline{U}\right|\right) \leqslant \left|\overline{U}\right|.$$

Since (i) by Lemma 2.3.1, $\left|\overline{U}\right| = O_p\left((nh)^{-1/2} + h^3\right)$, (ii) by Lemma 2.3.2, $\overline{\overline{U}} = \Omega_U + o_p(1)$, (iii) $Z_n = o_p\left(n^{1/s}\right)$ from the assumption of $E\left|Y_i\right|^s < \infty$ for $s > 2$, we have

$$\left\|\lambda(m)\right\| = \rho = O_p\left((nh)^{-1/2} + h^3\right).$$

Moreover, by a Taylor expansion of the EL weighted estimating equations at $\lambda = 0$, we have

$$0 = \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m) - \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\,U_i(m)^{\intercal}\right]\lambda(m) + o\left(\left\|\lambda(m)\right\|\right),$$

hence

$$\lambda(m) = \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\,U_i(m)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\right] + o_p\left((nh)^{-1/2} + h^3\right).$$

■

A.5 Proof of Lemma 2.3.3

Proof. Without losing generality, for the saddlepoint $\left(\widetilde{\beta}, \widetilde{\lambda}, \widetilde{\nu}\right)$ of $G_n^*(\beta, \lambda, \nu)$, we only consider the case $\widetilde{\nu} = 0$. That is, the inequality constraints $\underline{b} \leqslant \beta \leqslant \overline{b}$ are not binding in the large sample context. Therefore the "inner" optimization problem

$$\max_{\lambda \in \Lambda, \nu \in \mathbb{R}_+^6} \sum_{i=1}^{n} \log\left(1 + \lambda(\beta)^\intercal U_i(\beta)\right) + n\underline{\nu}^\intercal\left(\underline{b} - \beta\right) + n\overline{\nu}^\intercal\left(\beta - \overline{b}\right)$$

is simplified as

$$l(\beta) = \max_{\lambda \in \Lambda} \sum_{i=1}^{n} \log\left(1 + \lambda(\beta)^\intercal U_i(\beta)\right).$$

We point out that the following proof also holds without this simplification.

Denote $\overline{\beta} = \left(\overline{\beta}_0, \overline{\beta}_1, \overline{\beta}_2,\right)^\intercal$, and for $j = 0, 1, 2$, $\overline{\beta}_j = m_j - h^{2-j}u_j$, where $u_j \in \mathbb{R}$ is such that $u = (u_0, u_1, u_2)^\intercal$, $\|u\| = 1$. First, following the argument in the proof of Lemma 1 in Qin and Lawless(1994), we establish a lower bound for $l(\beta)$ at $\overline{\beta}$. To do this, notice that:

(i) by Lemma 2.3.1,

$$\begin{aligned}
\frac{1}{nh}\sum_{i=1}^{n} U_{ji}\left(\overline{\beta}_j\right) &= h^2 u_j\left[\frac{1}{nh}\sum_{i=1}^{n} W_{ji}(x)\left(\frac{X_i - x}{h}\right)^j\right] + \frac{1}{nh}\sum_{i=1}^{n} U_{ji}(m_j) \\
&= h^2 u_j f^3(x) D + o_p\left(h^2\right) + O_p\left((nh)^{-1/2} + h^3\right) \\
&= h^2 u_j f^3(x) D + o_p\left(h^2\right),
\end{aligned}$$

since $D_n = \frac{1}{nh}\sum_{i=1}^{n} W_{ji}(x)\left((X_i - x)/h\right)^j = f^3(x) D + o_p(1)$;

(ii) by Lemma 2.3.2,

$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i\left(\overline{\beta}\right) U_i\left(\overline{\beta}\right)^{\intercal} = \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i\left(m\right) U_i\left(m\right)' + o_p\left(1\right) = \Omega_U + o_p\left(1\right),$$

where

$$\begin{aligned}
\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}^2\left(\overline{\beta}_0\right) &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}^2\left(m_0\right) + O_p\left(h^3\right),\\
\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{1i}^2\left(\overline{\beta}_1\right) &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{1i}^2\left(m_1\right) + O_p\left(h^2\right),\\
\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{2i}^2\left(\overline{\beta}_2\right) &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{2i}^2\left(m_2\right) + O_p\left(h^2\right),\\
\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}\left(\overline{\beta}_0\right) U_{1i}\left(\overline{\beta}_1\right) &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}\left(m_0\right) U_{1i}\left(m_1\right) + O_p\left(h^2\right),\\
\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}\left(\overline{\beta}_0\right) U_{2i}\left(\overline{\beta}_2\right) &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{0i}\left(m_0\right) U_{2i}\left(m_2\right) + O_p\left(h^2\right),\\
\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{1i}\left(\overline{\beta}_1\right) U_{2i}\left(\overline{\beta}_2\right) &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_{1i}\left(m_1\right) U_{2i}\left(m_2\right) + O_p\left(h^2\right).
\end{aligned}$$

As in the proof of Theorem 2.3.1, from (i) and (ii), we have

$$\begin{aligned}
\lambda\left(\overline{\beta}\right) &= \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i\left(\overline{\beta}\right) U_i\left(\overline{\beta}\right)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i\left(\overline{\beta}\right)\right] + o_p\left(h^2\right) \qquad \text{(A.3)}\\
&= O_p\left(h^2\right).
\end{aligned}$$

Therefore by a Taylor expansion at $\lambda = 0$ and by (A.3),

$$
\begin{aligned}
l\left(\overline{\beta}\right) &= nh\{\lambda\left(\overline{\beta}\right)^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)\right] - \frac{1}{2}\lambda\left(\overline{\beta}\right)^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)U_i\left(\overline{\beta}\right)^{\intercal}\right]\lambda\left(\overline{\beta}\right) \\
&+ o_p\left(\left\|\lambda\left(\overline{\beta}\right)\right\|^2\right)\} \\
&= \frac{nh}{2}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)\right]^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)U_i\left(\overline{\beta}\right)^{\intercal}\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(\overline{\beta}\right)\right] \\
&+ o_p\left(nh^5\right) \\
&= \frac{nh}{2}\left[h^2uf^3\left(x\right)D + o_p\left(h^2\right)\right]^{\intercal}\Omega_U^{-1}\left[h^2uf^3\left(x\right)D + o_p\left(h^2\right)\right] + o_p\left(nh^5\right) \\
&\geqslant nh^5\left(c - \epsilon\right),
\end{aligned}
$$

where $c - \epsilon > 0$ and c is the smallest eigenvalue of $f^6\left(x\right)D^2\left(u^{\intercal}\Omega_U^{-1}u\right)$.

Similarly,

$$
\begin{aligned}
l\left(m\right) &= \frac{nh}{2}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)\right]^{\intercal}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)U_i\left(m\right)'\right]^{-1}\left[\frac{1}{nh}\sum\nolimits_{i=1}^{n}U_i\left(m\right)\right] \\
&+ o_p\left(nh\left(\left(nh\right)^{-1/2} + h^3\right)^2\right) \\
&= \frac{nh}{2}O_p\left(\left(nh\right)^{-1/2} + h^3\right)^{\intercal}\Omega_U^{-1}O_p\left(\left(nh\right)^{-1/2} + h^3\right) + o_p\left(nh^7\right) \\
&= O_p\left(nh^7\right).
\end{aligned}
$$

Since $l\left(\beta\right)$ is continuous in the interior of

$$
\left\{\beta\left(x\right) : \left|\beta_j\left(x\right) - m_j\left(x\right)\right| \leqslant h^{2-j},\ j = 0, 1, 2\right\}, \tag{A.4}
$$

$l(\beta)$ attains minimum value $\widetilde{\beta}$ in (A.4). Moreover, we have

$$\begin{aligned}\left.\frac{\partial l(\beta)}{\partial \beta}\right|_{\beta=\widetilde{\beta}} &= \left(\partial \lambda(\beta)^{\intercal}/\partial \beta\right) \sum_{i=1}^{n} \left.\frac{U_i(\beta)}{1+\lambda(\beta)^{\intercal} U_i(\beta)}\right|_{\beta=\widetilde{\beta}} \\ &\quad + \sum_{i=1}^{n} \left.\frac{\left(\partial U_i(\beta)^{\intercal}/\partial \beta\right)\lambda(\beta)}{1+\lambda(\beta)^{\intercal} U_i(\beta)}\right|_{\beta=\widetilde{\beta}} \\ &= 0\end{aligned} \tag{A.5}$$

Note that we already have

$$g_{1n}\left(\widetilde{\beta},\widetilde{\lambda}\right) = \frac{1}{nh}\sum_{i=1}^{n} \left.\frac{U_i(\beta)}{1+\lambda(\beta)^{\intercal} U_i(\beta)}\right|_{\beta=\widetilde{\beta}} = 0$$

as discussed in Remark 2.2.1. Therefore by (A.5),

$$\sum_{i=1}^{n} \left.\frac{\left(\partial U_i(\beta)^{\intercal}/\partial \beta\right)\lambda(\beta)}{1+\lambda(\beta)^{\intercal} U_i(\beta)}\right|_{\beta=\widetilde{\beta}} = 0,$$

where $\partial U_i(\beta)^{\intercal}/\partial \beta = diag\left\{-W_{ji}(x)(X_i - x)^j\right\}$. Denote $H_3 = diag\{h^j\}$, then $D_i(x) = \left(\partial U_i(\beta)^{\intercal}/\partial \beta\right) H_3^{-1}$ and

$$g_{2n}\left(\widetilde{\beta},\widetilde{\lambda}\right) = \frac{1}{nh}\sum_{i=1}^{n} \left.\frac{D_i(x)\lambda(\beta)}{1+\lambda(\beta)^{\intercal} U_i(\beta)}\right|_{\beta=\widetilde{\beta}} = 0.$$

■

A.6 Proof of Theorem 2.3.2

Proof. Taking derivatives of $g_{1n}(\beta,\lambda)$ and $g_{2n}(\beta,\lambda)$ and evaluating at $(m,0)$, we have

$$\begin{aligned}
\frac{\partial g_{1n}(m,0)}{\partial\beta^{\intercal}} &= \frac{1}{nh}\sum\nolimits_{i=1}^{n}\left(\partial U_i(\beta)^{\intercal}/\partial\beta\right) = \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} D_i(x)\right] H_3,\\
\frac{\partial g_{1n}(m,0)}{\partial\lambda^{\intercal}} &= -\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\,U_i(m)^{\intercal},\\
\frac{\partial g_{2n}(m,0)}{\partial\beta^{\intercal}} &= 0,\\
\frac{\partial g_{2n}(m,0)}{\partial\lambda^{\intercal}} &= \frac{1}{nh}\sum\nolimits_{i=1}^{n} D_i(x).
\end{aligned}$$

Note that $\frac{1}{nh}\sum_{i=1}^{n} D_i(x) = -D_n I_3$ since $D_n = \frac{1}{nh}\sum_{i=1}^{n} W_{ji}(x)\left((X_i - x)/h\right)^j$ for $j = 0,1,2$. By Taylor expanding $g_{1n}\left(\widetilde{\beta},\widetilde{\lambda}\right)$ and $g_{2n}\left(\widetilde{\beta},\widetilde{\lambda}\right)$ at $(m,0)$, we have

$$\begin{aligned}
0 &= g_{1n}\left(\widetilde{\beta},\widetilde{\lambda}\right)\\
&= g_{1n}(m,0) + \frac{\partial g_{1n}(m,0)}{\partial\beta^{\intercal}}\left(\widetilde{\beta}-m\right) + \frac{\partial g_{1n}(m,0)}{\partial\lambda^{\intercal}}\left(\widetilde{\lambda}-0\right) + o_p(\delta)\\
&= \frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m) - D_n H_3\left(\widetilde{\beta}-m\right) - \left[\frac{1}{nh}\sum\nolimits_{i=1}^{n} U_i(m)\,U_i(m)^{\intercal}\right]\left(\widetilde{\lambda}-0\right) + o_p(\delta),\\
0 &= g_{2n}\left(\widetilde{\beta},\widetilde{\lambda}\right)\\
&= g_{2n}(m,0) + \frac{\partial g_{2n}(m,0)}{\partial\beta^{\intercal}}\left(\widetilde{\beta}-m\right) + \frac{\partial g_{2n}(m,0)}{\partial\lambda^{\intercal}}\left(\widetilde{\lambda}-0\right) + o_p(\delta)\\
&= 0 + 0\left(\widetilde{\beta}-m\right) - D_n I_3\left(\widetilde{\lambda}-0\right) + o_p(\delta),
\end{aligned}$$

where $\delta = \left\| H_3 \left(\widetilde{\beta} - m \right) \right\| + \left\| \widetilde{\lambda} \right\|$. Hence we have

$$\begin{pmatrix} H_3 \left(\widetilde{\beta} - m \right) \\ \widetilde{\lambda} \end{pmatrix} = \Omega_g^{-1} \begin{pmatrix} \frac{1}{nh} \sum_{i=1}^n U_i (m) + o_p (\delta) \\ o_p (\delta) \end{pmatrix},$$

where

$$\Omega_g = \begin{pmatrix} D_n I_3 & \frac{1}{nh} \sum_{i=1}^n U_i (m) U_i (m)^{\intercal} \\ 0 & D_n I_3 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} f^3 (x) D I_3 & -\Omega_U \\ 0 & f^3 (x) D I_3 \end{pmatrix}.$$

By this result, and $\frac{1}{nh} \sum_{i=1}^n U_i (m) = O_p \left((nh)^{-1/2} + h^3 \right)$, it can be shown that $\delta = O_p \left((nh)^{-1/2} + h^3 \right)$. For the limit distribution of $\widetilde{\beta}$, we have

$$H_3 \left(\widetilde{\beta} - m \right) = D_n^{-1} \left[\frac{1}{nh} \sum\nolimits_{i=1}^n U_i (m) \right] + o_p \left((nh)^{-1/2} + h^3 \right),$$

that is, for $j = 0, 1, 2,$

$$\begin{aligned} \widetilde{\beta}_j (x) - m_j (x) &= h^{-j} \frac{\frac{1}{nh} \sum_{i=1}^n W_{ji} (x) \left[Y_i - m_j (x) (X_i - x)^j \right]}{\frac{1}{nh} \sum_{i=1}^n W_{ji} (x) \left(\frac{X_i - x}{h} \right)^j} + h^{-j} o_p \left((nh)^{-1/2} + h^3 \right) \\ &= \widehat{\beta}_j (x) - m_j (x) + o_p \left(\left(nh^{1+2j} \right)^{-1/2} + h^{3-j} \right). \end{aligned}$$

Thus

$$
\begin{aligned}
&\sqrt{nh^{1+2j}}\left(\widetilde{\beta}_j\left(x\right)-m_j\left(x\right)\right)\\
&=\sqrt{nh^{1+2j}}\left(\widehat{\beta}_j\left(x\right)-m_j\left(x\right)\right)+\sqrt{nh^{1+2j}}o_p\left(\left(nh^{1+2j}\right)^{-1/2}+h^{3-j}\right)\\
&=\sqrt{nh^{1+2j}}\left(\widehat{\beta}_j\left(x\right)-m_j\left(x\right)\right)+o_p\left(1+\sqrt{nh^7}\right).
\end{aligned}
$$

■

APPENDIX B

PROOFS IN SECTION 3

Throughout the following proofs, I use $\|\cdot\|$ to denote the Euclidean norm for a vector or a matrix, that is, $\|X\| = \sqrt{\sum_{k=1}^{d} x_k^2}$, $\|M\| = \sqrt{\sum_{k=1}^{d}\sum_{l=1}^{d} m_{kl}^2}$. I also use C as a generic constant. For notational simplicity, I will use β_i instead of $\beta\left(z_i\right)$, and $\int \beta$ instead of $\int \beta\left(z\right) dz$ when it does not cause confusion.

B.1 Proof of Theorem 3.3.1

Proof. The conclusion of Theorem 3.3.1 follows from two results. First,

$$n^{-1}\left|\sum_{i=1}^{m}\overline{u}_i^2 - \frac{m}{n}\sum_{i=1}^{n}\overline{u}_i^2\right| \xrightarrow{p} \sup_{r\in(0,1]}\left|\int_0^r \sigma^2 - r\int_0^1 \sigma^2 + \Delta\left(r\right)\right| \tag{B.1}$$

with

$$\Delta\left(r\right) = \int_0^r \beta'\Omega_0\beta - r^{-1}\left(\int_0^r \beta\right)'\Omega_0\left(\int_0^r \beta\right) - r\int_0^1 \beta'\Omega_0\beta + r\left(\int_0^1 \beta\right)'\Omega_0\left(\int_0^1 \beta\right).$$

Second,

$$\begin{aligned}\overline{\gamma}\left(0\right) \xrightarrow{p} &\left[\psi\left(0\right)+1\right]\int \sigma^4 + 6\left[\int \sigma^2\beta'\Omega_\varepsilon\beta - \int \sigma^2\left(\int \beta\right)'\Omega_\varepsilon\left(\int \beta\right)\right] \\ &- 3\Omega_{40} + 6\Omega_{22} - 4\Omega_{13} + \Omega_{04} - \left(\int \sigma^2 + \int \beta'\Omega_0\beta - \left(\int \beta\right)'\Omega_0\left(\int \beta\right)\right)^2.\end{aligned} \tag{B.2}$$

We will show them in the following.

Proof of (B.1). The OLS estimator can be written as

$$\begin{aligned}\overline{\beta} &= \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i y_i \\ &= \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i X_i' \beta_i + \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i u_i,\end{aligned}$$

and the OLS residuals are $\overline{u}_i = y_i - X_i'\overline{\beta}$. By standard calculation, we have

$$\begin{aligned}\sum_{i=1}^n \overline{u}_i^2 &= \sum_{i=1}^n u_i^2 - 2\sum_{i=1}^n \left(\overline{\beta} - \beta_i\right)' X_i u_i + \sum_{i=1}^n \left(\overline{\beta} - \beta_i\right)' X_i X_i' \left(\overline{\beta} - \beta_i\right) \\ &= \sum_{i=1}^n u_i^2 - 2\sum_{i=1}^n \beta_i' X_i X_i' \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i u_i + 2\sum_{i=1}^n \beta_i' X_i u_i \\ &+ \sum_{i=1}^n \beta_i' X_i X_i' \beta_i - \sum_{i=1}^n \beta_i' X_i X_i' \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i X_i' \beta_i \\ &- \sum_{i=1}^n X_i' u_i \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i u_i,\end{aligned}$$

so $n^{-1/2}\left(\sum_{i=1}^m \overline{u}_i^2 - \frac{m}{n}\sum_{i=1}^n \overline{u}_i^2\right) = T_0 - 2T_1 + 2T_2 + T_3 - T_4 - T_5$, where

$$\begin{aligned}T_0 &= n^{-1/2}\left[\sum_{i=1}^m u_i^2 - \frac{m}{n}\sum_{i=1}^n u_i^2\right], \\ T_1 &= n^{-1/2}\sum_{i=1}^m \beta_i' X_i X_i' \left(\sum_{i=1}^m X_i X_i'\right)^{-1} \sum_{i=1}^m X_i u_i \\ &\quad - n^{-1/2}\frac{m}{n}\sum_{i=1}^n \beta_i' X_i X_i' \left(\sum_{i=1}^n X_i X_i'\right)^{-1} \sum_{i=1}^n X_i u_i, \\ T_2 &= n^{-1/2}\left[\sum_{i=1}^m \beta_i' X_i u_i - \frac{m}{n}\sum_{i=1}^n \beta_i' X_i u_i\right], \\ T_3 &= n^{-1/2}\left[\sum_{i=1}^m \beta_i' X_i X_i' \beta_i - \frac{m}{n}\sum_{i=1}^n \beta_i' X_i X_i' \beta_i\right],\end{aligned}$$

$$T_4 = n^{-1/2}\textstyle\sum_{i=1}^{m}\beta_i' X_i X_i' \left(\sum_{i=1}^{m} X_i X_i'\right)^{-1} \sum_{i=1}^{m} X_i X_i' \beta_i$$

$$- n^{-1/2}\frac{m}{n}\textstyle\sum_{i=1}^{n}\beta_i' X_i X_i' \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i X_i' \beta_i,$$

$$T_5 = n^{-1/2}\textstyle\sum_{i=1}^{m} X_i' u_i \left(\sum_{i=1}^{m} X_i X_i'\right)^{-1} \sum_{i=1}^{m} X_i u_i$$

$$- n^{-1/2}\frac{m}{n}\textstyle\sum_{i=1}^{n} X_i' u_i \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} X_i u_i.$$

To verify the limit results of these terms, we need: (i) $n^{-1}\sum_{i=1}^{n} X_i X_i' \xrightarrow{p} \Omega_0$ due to the assumption that $\{X_i\}$ are ergodic for second moments; (ii) $n^{-1}\sum_{i=1}^{n} X_i X_i' \beta_i \xrightarrow{p} \Omega_0 \int \beta$ since $E\left(n^{-1}\sum_{i=1}^{n} X_i X_i' \beta_i\right) = \Omega_0 n^{-1}\sum_{i=1}^{n}\beta_i = \Omega_0 \left(\int \beta + O\left(n^{-1}\right)\right)$ by the Riemann sum approximation of an integral, and for $k, l = 1, \cdots, d$, $Var\left(n^{-1}\sum_{i=1}^{n} x_{ik} x_{il} \beta_l\left(z_i\right)\right) \to 0$ by a similar argument as in Cai (2007), Lemma 2; (iii) $n^{-1}\sum_{i=1}^{n}\beta_i' X_i X_i' \beta_i \xrightarrow{p} \int \beta' \Omega_0 \beta$; (iv) $n^{-1}\sum_{i=1}^{n} X_i u_i \xrightarrow{p} 0$; (v) $n^{-1}\sum_{i=1}^{n}\beta_i' X_i u_i \xrightarrow{p} 0$, by the LLN for martingale difference sequence. Then by (i)-(v), we have,

$$n^{-1/2} T_0 \xrightarrow{p} \textstyle\int_0^r \sigma^2 - r\int_0^1 \sigma^2,$$

$$n^{-1/2} T_1 \xrightarrow{p} 0,$$

$$n^{-1/2} T_2 \xrightarrow{p} 0,$$

$$n^{-1/2} T_3 \xrightarrow{p} \textstyle\int_0^r \beta' \Omega_0 \beta - r\int_0^1 \beta' \Omega_0 \beta,$$

$$n^{-1/2} T_4 \xrightarrow{p} r^{-1}\left(\textstyle\int_0^r \beta\right)' \Omega_0 \left(\int_0^r \beta\right) - r\left(\int_0^1 \beta\right)' \Omega_0 \left(\int_0^1 \beta\right),$$

$$T_5 = n^{-1/2} O_p(1) \xrightarrow{p} 0.$$

Now it can be seen that T_0, T_3, and T_4 dominate the other terms, so (B.1) holds.

Proof of (B.2). In the case of i.i.d. errors, instead of using the LRV estimator, we only need to use an estimator of the variance of $\{u_i^2\}$ to standardize the CUSQ statistic. Define the variance estimator based on OLS residuals as

$$\overline{\gamma}(0) = n^{-1} \sum\nolimits_{i=1}^{n} \left(\overline{u}_i^2 - \overline{\sigma}^2\right)^2 = n^{-1} \sum\nolimits_{i=1}^{n} \overline{u}_i^4 - \left(\overline{\sigma}^2\right)^2, \tag{B.3}$$

where $\overline{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \overline{u}_i^2$. For $\overline{\sigma}^2$, again by (i)-(v), we have

$$\overline{\sigma}^2 \xrightarrow{p} \textstyle\int \sigma^2 + \int \beta' \Omega_0 \beta - \left(\int \beta\right)' \Omega_0 \left(\int \beta\right). \tag{B.4}$$

For the first term in (B.3), write

$$\begin{aligned} \frac{1}{n} \sum\nolimits_{i=1}^{n} \overline{u}_i^4 = {} & \frac{1}{n} \sum\nolimits_{i=1}^{n} u_i^4 - \frac{4}{n} \sum\nolimits_{i=1}^{n} \left(\overline{\beta} - \beta_i\right)' X_i u_i^3 + \frac{6}{n} \sum\nolimits_{i=1}^{n} \left[\left(\overline{\beta} - \beta_i\right)' X_i\right]^2 u_i^2 \\ & - \frac{4}{n} \sum\nolimits_{i=1}^{n} \left[\left(\overline{\beta} - \beta_i\right)' X_i\right]^3 u_i + \frac{1}{n} \sum\nolimits_{i=1}^{n} \left[\left(\overline{\beta} - \beta_i\right)' X_i\right]^4, \end{aligned}$$

where

$$\begin{aligned} & \frac{1}{n} \sum\nolimits_{i=1}^{n} \left[\left(\overline{\beta} - \beta_i\right)' X_i\right]^4 \\ & = \frac{1}{n} \sum\nolimits_{i=1}^{n} \left(\overline{\beta}' X_i\right)^4 - \frac{4}{n} \sum\nolimits_{i=1}^{n} \left(\overline{\beta}' X_i\right)^3 \left(\beta_i' X_i\right) + \frac{6}{n} \sum\nolimits_{i=1}^{n} \left(\overline{\beta}' X_i\right)^2 \left(\beta_i' X_i\right)^2 \\ & - \frac{4}{n} \sum\nolimits_{i=1}^{n} \left(\overline{\beta}' X_i\right) \left(\beta_i' X_i\right)^3 + \frac{1}{n} \sum\nolimits_{i=1}^{n} \left(\beta_i' X_i\right)^4. \end{aligned}$$

Let $\Omega_\varepsilon = E\left(X_1 X_1' \varepsilon_1^2\right)$, then under H_1 of time varying unconditional volatility, $E\left(X_i X_i' u_i^2\right) = \sigma_i^2 \Omega_\varepsilon$ for each $i = 1, \cdots, n$. Also, let

$$\Omega_{40} = E\left[\left(\int \beta\right)' X_1\right]^4,$$

$$\Omega_{22} = E\left[\left(\int \beta\right)' X_1 X_1' \left(\int \beta\right) \int \left(\beta' X_1 X_1' \beta\right)\right],$$

$$\Omega_{13} = E\left[\left(\int \beta\right)' X_1 \int \left(X_1' \beta' X_1 X_1' \beta\right)\right],$$

$$\Omega_{04} = E\left[\int \left(\beta' X_1 X_1' \beta' X_1 X_1' \beta\right)\right],$$

then because

$$\frac{1}{n}\sum_{i=1}^n u_i^4 \xrightarrow{p} \left[E\left(\varepsilon_1^2 - 1\right)^2 + 1\right] \int \sigma^4,$$

$$\frac{1}{n}\sum_{i=1}^n \left(\overline{\beta} - \beta_i\right)' X_i u_i^3 \xrightarrow{p} 0,$$

$$\frac{1}{n}\sum_{i=1}^n \left[\left(\overline{\beta} - \beta_i\right)' X_i\right]^2 u_i^2 \xrightarrow{p} \int \sigma^2 \beta' \Omega_\varepsilon \beta - \int \sigma^2 \left(\int \beta\right)' \Omega_\varepsilon \left(\int \beta\right),$$

$$\frac{1}{n}\sum_{i=1}^n \left[\left(\overline{\beta} - \beta_i\right)' X_i\right]^3 u_i \xrightarrow{p} 0,$$

and $\frac{1}{n}\sum_{i=1}^n \left(\overline{\beta}' X_i\right)^4 \xrightarrow{p} \Omega_{40}$, $\frac{1}{n}\sum_{i=1}^n \left(\overline{\beta}' X_i\right)^3 \left(\beta_i' X_i\right) \xrightarrow{p} \Omega_{40}$, $\frac{1}{n}\sum_{i=1}^n \left(\overline{\beta}' X_i\right)^2 \left(\beta_i' X_i\right)^2 \xrightarrow{p} \Omega_{22}$, $\frac{1}{n}\sum_{i=1}^n \left(\overline{\beta}' X_i\right) \left(\beta_i' X_i\right)^3 \xrightarrow{p} \Omega_{13}$, $\frac{1}{n}\sum_{i=1}^n \left(\beta_i' X_i\right)^4 \xrightarrow{p} \Omega_{04}$, we have (B.2) holds. ■

B.2 Proof of Theorem 3.3.2

Proof. In the case of serially correlated errors, (B.1) still holds[1]. For the denominator of the CUSQ statistic, we want to show that under Assumption 3.3.6,

$$b^{-1}\overline{\omega}^2 \xrightarrow{p} \left(\int k(x)\,dx\right)\left[\int\sigma^4 - \left(\int\sigma^2 + \int\beta'\Omega_0\beta - \left(\int\beta\right)'\Omega_0\left(\int\beta\right)\right)^2\right]. \qquad \text{(B.5)}$$

Consider the LRV estimator $\overline{\omega}^2 = \sum_{l=-n+1}^{n-1} k(l/b)\,\overline{\gamma}(l)$ based on OLS residuals, where $\overline{\gamma}(l) = n^{-1}\sum_{i=l+1}^{n}\left(\overline{u}_i^2 - \overline{\sigma}^2\right)\left(\overline{u}_{i-l}^2 - \overline{\sigma}^2\right)$, $\overline{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\overline{u}_i^2$. We decompose $\overline{\gamma}(l)$ as

$$\overline{\gamma}(l) = n^{-1}\sum_{i=l+1}^{n}\left(\overline{u}_i^2 - \overline{\sigma}^2\right)\left(\overline{u}_{i-l}^2 - \overline{\sigma}^2\right) = T_{1l} + T_{2l} + T_{3l} + T_{4l} + T_{5l},$$

where

$$\begin{aligned}
T_{1l} &= n^{-1}\sum_{i=l+1}^{n}\left(u_i^2 - \sigma_i^2\right)\left(u_{i-l}^2 - \sigma_{i-l}^2\right),\\
T_{2l} &= n^{-1}\sum_{i=l+1}^{n}\left(u_i^2 - \sigma_i^2\right)\left(\sigma_{i-l}^2 - \overline{\sigma}^2\right),\\
T_{3l} &= n^{-1}\sum_{i=l+1}^{n}\left(\sigma_i^2 - \overline{\sigma}^2\right)\left(u_{i-l}^2 - \sigma_{i-l}^2\right),\\
T_{4l} &= n^{-1}\sum_{i=l+1}^{n}\left(\sigma_i^2 - \overline{\sigma}^2\right)\left(\sigma_{i-l}^2 - \overline{\sigma}^2\right),\\
T_{5l} &= n^{-1}\sum_{i=l+1}^{n}\left(\overline{u}_i^2 - u_i^2\right)\left(\overline{u}_{i-l}^2 - u_{i-l}^2\right) + n^{-1}\sum_{i=l+1}^{n}\left(\overline{u}_i^2 - u_i^2\right)\left(u_{i-l}^2 - \overline{\sigma}^2\right)\\
&\quad + n^{-1}\sum_{i=l+1}^{n}\left(u_i^2 - \overline{\sigma}^2\right)\left(\overline{u}_{i-l}^2 - u_{i-l}^2\right).
\end{aligned}$$

[1]The only change in the proof is that we need to use a LLN for mixingales to establish (iv) and (v).

Note that the four terms T_{1l}, T_{2l}, T_{3l}, and T_{4l} are the same as defined in Xu (2012, proof of Theorem 1) except for that our $\overline{\sigma}^2$ is based on $\overline{u}_i$ instead of u_i. So we follow the notation there and let

$$\overline{\omega}^2 = \overline{\gamma}(0) + 2\sum_{l=1}^{n-1} k(l/b)\,\overline{\gamma}(l) = T_1 + T_2 + T_3 + T_4 + T_5,$$

where $T_i = T_{i0} + 2\sum_{l=1}^{n-1} k(l/b)\,T_{il}$, $i = 1,2,3,4,5$. As shown in Xu (2012, proof of Theorem 1), we have $T_1 \xrightarrow{p} \omega^2$, $b^{-1}T_2 = O_p\left(n^{-1/2}\right)$, $b^{-1}T_3 = O_p\left(n^{-1/2}\right)$. For T_4, since for $0 \leqslant l \leqslant b$, $l \to \infty$, $l/n \to 0$ as $n \to \infty$, also by (B.4),

$$\begin{aligned} T_{4l} &= n^{-1}\sum_{i=l+1}^{n} \sigma_i^2\sigma_{i-l}^2 - \overline{\sigma}^2 n^{-1}\sum_{i=l+1}^{n} \sigma_i^2 - \overline{\sigma}^2 n^{-1}\sum_{i=l+1}^{n} \sigma_{i-l}^2 + \left(\overline{\sigma}^2\right)^2 (n-l)/n \\ &\xrightarrow{p} \int\sigma^4 - \left(\int\sigma^2\right)^2 + \left(\int\beta'\Omega_0\beta - \left(\int\beta\right)'\Omega_0\left(\int\beta\right)\right)^2, \end{aligned}$$

so

$$b^{-1}T_4 \xrightarrow{p} \left(\int k(x)\,dx\right)\left[\int\sigma^4 - \left(\int\sigma^2\right)^2 + \left(\int\beta'\Omega_0\beta - \left(\int\beta\right)'\Omega_0\left(\int\beta\right)\right)^2\right], \tag{B.6}$$

where we have used

$$\begin{aligned} b^{-1}\sum_{l=-n+1}^{n-1} k(l/b) &\overset{l=\lfloor ns\rfloor}{=} nb^{-1}\sum_{l=-n+1}^{n-1}\int_{l/n}^{(l+1)/n} k\left(\lfloor ns\rfloor/b\right) ds \\ &= nb^{-1}\int_{-1+1/n}^{1} k\left(\lfloor ns\rfloor/b\right) ds \overset{x=ns/b}{=} \int_{(-n+1)/b}^{n/b} k\left(\lfloor xb\rfloor/b\right) dx \to \int k(x)\,dx. \end{aligned}$$

Then it remains to find the limit of T_5. Denote

$$e_{2i}=\left(\overline{\beta}-\beta_i\right)' X_i u_i,\ e_{3i}=\left(\overline{\beta}-\beta_i\right)' X_i X_i' \left(\overline{\beta}-\beta_i\right),$$

$$A_{n2,l}=n^{-1}\sum_{i=l+1}^{n}\left(e_{2i}+e_{2i-l}\right),\ A_{n3,l}=n^{-1}\sum_{i=l+1}^{n}\left(e_{3i}+e_{3i-l}\right),$$

$$A_{n12}=n^{-1}\textstyle\sum_{i=l+1}^{n}\left(e_{2i}u_{i-l}^2+e_{2i-l}u_i^2\right),\ A_{n13}=n^{-1}\textstyle\sum_{i=l+1}^{n}\left(e_{3i}u_{i-l}^2+e_{3i-l}u_i^2\right),$$

$$A_{n22}=n^{-1}\textstyle\sum_{i=l+1}^{n}e_{2i}e_{2i-l},\ A_{n23}=n^{-1}\textstyle\sum_{i=l+1}^{n}\left(e_{2i}e_{3i-l}+e_{2i-l}e_{3i}\right),$$

$$A_{n33}=n^{-1}\textstyle\sum_{i=l+1}^{n}e_{3i}e_{3i-l},$$

then

$$T_{5l}=\overline{\sigma}^2\left(2A_{n2,l}-A_{n3,l}\right)-2A_{n12}+A_{n13}+4A_{n22}-2A_{n23}+A_{n33},$$

and we can show that

$$A_{n2,l}\xrightarrow{p}0,$$

$$A_{n3,l}\xrightarrow{p}2\left[\textstyle\int\beta'\Omega_0\beta-\left(\int\beta\right)'\Omega_0\left(\int\beta\right)\right],$$

$$A_{n12}\xrightarrow{p}0,$$

$$A_{n13}\xrightarrow{p}\textstyle\int\beta' E\left(X_1X_1'u_{1-l}^2\right)\beta-\left(\int\beta\right)'\int E\left(X_1X_1'u_{1-l}^2\right)\left(\int\beta\right)$$

$$+\textstyle\int\beta' E\left(X_1X_1'u_{1+l}^2\right)\beta-\left(\int\beta\right)'\int E\left(X_1X_1'u_{1+l}^2\right)\left(\int\beta\right),$$

$$A_{n22}\xrightarrow{p}\textstyle\int\beta' E\left(X_1X_{1-l}'u_1u_{1-l}\right)\beta-\left(\int\beta\right)'\int E\left(X_1X_{1-l}'u_1u_{1-l}\right)\left(\int\beta\right),$$

$$A_{n23}\xrightarrow{p}E\left[\left(\textstyle\int\beta\right)'\int X_1u_1\int\left(\beta'X_{1-l}X_{1-l}'\beta\right)\right]-E\left[\textstyle\int\beta'X_1u_1\beta'X_{1-l}X_{1-l}'\beta\right]$$

$$+E\left[\left(\textstyle\int\beta\right)'\int X_{1-l}u_{1-l}\int\left(\beta'X_1X_1'\beta\right)\right]-E\left[\textstyle\int\beta'X_{1-l}u_{1-l}\beta'X_1X_1'\beta\right],$$

and

$$
\begin{aligned}
A_{n33} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\overline{\beta}-\beta_i\right)' X_i X_i' \left(\overline{\beta}-\beta_i\right) \left(\overline{\beta}-\beta_{i-l}\right)' X_{i-l} X_{i-l}' \left(\overline{\beta}-\beta_{i-l}\right) \\
&= a_{11} - 2a_{12} + a_{13} - 2a_{21} + 4a_{22} - 2a_{23} + a_{31} - 2a_{32} + a_{33}, \\
a_{11} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\overline{\beta}' X_i X_i' \overline{\beta}\right) \left(\overline{\beta}' X_{i-l} X_{i-l}' \overline{\beta}\right) \\
&\xrightarrow{p} E\left[\left(\int\beta\right)' X_1 X_1' \left(\int\beta\right) \left(\int\beta\right)' X_{1-l} X_{1-l}' \left(\int\beta\right)\right], \\
a_{12} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\overline{\beta}' X_i X_i' \overline{\beta}\right) \left(\beta_{i-l}' X_{i-l} X_{i-l}' \overline{\beta}\right) \\
&\xrightarrow{p} E\left[\left(\int\beta\right)' X_1 X_1' \left(\int\beta\right) \left(\int\beta\right)' X_{1-l} X_{1-l}' \left(\int\beta\right)\right], \\
a_{13} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\overline{\beta}' X_i X_i' \overline{\beta}\right) \left(\beta_{i-l}' X_{i-l} X_{i-l}' \beta_{i-l}\right) \\
&\xrightarrow{p} E\left[\left(\int\beta\right)' X_1 X_1' \left(\int\beta\right) \int \left(\beta' X_{1-l} X_{1-l}' \beta\right)\right], \\
a_{21} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\beta_i' X_i X_i' \overline{\beta}\right) \left(\overline{\beta}' X_{i-l} X_{i-l}' \overline{\beta}\right) \\
&\xrightarrow{p} E\left[\left(\int\beta\right)' X_1 X_1' \left(\int\beta\right) \left(\int\beta\right)' X_{1-l} X_{1-l}' \left(\int\beta\right)\right], \\
a_{22} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\beta_i' X_i X_i' \overline{\beta}\right) \left(\beta_{i-l}' X_{i-l} X_{i-l}' \overline{\beta}\right) \\
&\xrightarrow{p} E\left[\left(\int\beta\right)' X_1 \left(\int X_1' \beta' X_{1-l}\right) X_{1-l}' \left(\int\beta\right)\right], \\
a_{23} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\beta_i' X_i X_i' \overline{\beta}\right) \left(\beta_{i-l}' X_{i-l} X_{i-l}' \beta_{i-l}\right) \\
&\xrightarrow{p} E\left[\left(\int\beta\right)' X_1 \int \left(X_1' \beta' X_{1-l} X_{1-l}' \beta\right)\right], \\
a_{31} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\beta_i' X_i X_i' \beta_i\right) \left(\overline{\beta}' X_{i-l} X_{i-l}' \overline{\beta}\right) \\
&\xrightarrow{p} E\left[\int \left(\beta' X_1 X_1' \beta\right) \left(\int\beta\right)' X_{1-l} X_{1-l}' \left(\int\beta\right)\right], \\
a_{32} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\beta_i' X_i X_i' \beta_i\right) \left(\beta_{i-l}' X_{i-l} X_{i-l}' \overline{\beta}\right) \\
&\xrightarrow{p} E\left[\left(\int \beta' X_1 X_1' \beta' X_{1-l}\right) X_{1-l}' \left(\int\beta\right)\right], \\
a_{33} &= n^{-1}\textstyle\sum_{i=l+1}^{n} \left(\beta_i' X_i X_i' \beta_i\right) \left(\beta_{i-l}' X_{i-l} X_{i-l}' \beta_{i-l}\right) \xrightarrow{p} E\left[\int \beta' X_1 X_1' \beta' X_{1-l} X_{1-l}' \beta\right].
\end{aligned}
$$

Together, for $T_5 = T_{50} + 2\sum_{l=1}^{n-1} k\left(l/b\right)T_{5l}$, we have

$$b^{-1}T_5 \xrightarrow{p} \left(\int k\left(x\right)dx\right)\left\{-2\left[\int\sigma^2 + \int\beta'\Omega_0\beta - \left(\int\beta\right)'\Omega_0\left(\int\beta\right)\right]\right.$$
$$\left.\times\left[\int\beta'\Omega_0\beta - \left(\int\beta\right)'\Omega_0\left(\int\beta\right)\right]\right\}. \tag{B.7}$$

Finally, since $b^{-1}T_i \xrightarrow{p} 0$, $i = 1, 2, 3$, and by (B.6) and (B.7), we can get (B.5). ■

B.3 Proof of Theorem 3.3.3

Proof. The CUSQ test statistic based on local linear residuals $\{\widehat{u}_i\}$ is

$$\widehat{Q} = \max_{1\leq m\leq n} n^{-1/2}\left|\sum_{i=1}^{m}\widehat{u}_i^2 - \frac{m}{n}\sum_{i=1}^{n}\widehat{u}_i^2\right|/\widehat{\omega}.$$

Consider the numerator first. Note that $\widehat{u}_i = y_i - X_i'\widehat{\beta}_i$, where $\widehat{\beta}_i = \widehat{\beta}\left(z_i\right)$ is the local linear estimator evaluated at z_i, we have

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\widehat{u}_i^2 \\
&= n^{-1}\sum_{i=1}^{n}u_i^2 - 2n^{-1}\sum_{i=1}^{n}\left(\widehat{\beta}_i - \beta_i\right)'X_iu_i + n^{-1}\sum_{i=1}^{n}\left(\widehat{\beta}_i - \beta_i\right)'X_iX_i'\left(\widehat{\beta}_i - \beta_i\right) \\
&= A_{n1} - 2A_{n2} + A_{n3}.
\end{aligned}$$

By Lemma B.4.2 and Lemma B.4.3 below, $\sqrt{n}A_{n2} = o_p\left(1\right)$, $\sqrt{n}A_{n3} = o_p\left(1\right)$, so

$$n^{-1/2}\sum_{i=1}^{n}\widehat{u}_i^2 = n^{-1/2}\sum_{i=1}^{n}u_i^2 + o_p\left(1\right).$$

Then by Lemma 1 in Cavaliere (2004), we can show that, for $m = 1, \cdots, n$,

$$n^{-1/2}\sum_{i=1}^{m} u_i^2 - \frac{m}{n}\left(n^{-1/2}\sum_{i=1}^{n} u_i^2\right) \Rightarrow \omega\left[W_\sigma(r) - rW_\sigma(1)\right],$$

where $\omega^2 = \xi^2 \int_0^1 \sigma^4$ and $\xi^2 = \sum_{l=-\infty}^{\infty} E\left[\left(\varepsilon_i^2 - 1\right)\left(\varepsilon_{i-l}^2 - 1\right)\right]$ is the LRV of $\varepsilon_i^2 - 1$, $W_\sigma(r) = \left(\int_0^1 \sigma^4\right)^{-1/2} \int_0^r \sigma^2 dW$ is a Brownian functional, also called a time-deformed Brownian motion in the sense that $W_\sigma(r) = W(\eta(r))$, with $\eta(r) = \int_0^r \sigma^4 / \int_0^1 \sigma^4$. Note that under H_0, $W_\sigma(r)$ reduces to a standard Brownian motion $W(r)$ and $\omega^2 = \sigma^4 \xi^2$. So under H_0,

$$\begin{aligned} n^{-1/2}\left(\sum_{i=1}^{m} \widehat{u}_i^2 - \frac{m}{n}\sum_{i=1}^{n} \widehat{u}_i^2\right) &= n^{-1/2}\left(\sum_{i=1}^{m} u_i^2 - \frac{m}{n}\sum_{i=1}^{n} u_i^2\right) + o_p(1) \\ &\Rightarrow \omega\left[W(r) - rW(1)\right], \end{aligned}$$

and

$$\max_{1 \le m \le n} n^{-1/2}\left|\sum_{i=1}^{m} \widehat{u}_i^2 - \frac{m}{n}\sum_{i=1}^{n} \widehat{u}_i^2\right| \Rightarrow \omega \sup_{r \in (0,1]} \left|W(r) - rW(1)\right|. \tag{B.8}$$

Under H_1, since $A_{n2} = o_p(1)$, $A_{n3} = o_p(1)$ by Lemma B.4.2 and Lemma B.4.3, we have

$$n^{-1}\sum_{i=1}^{n} \widehat{u}_i^2 = n^{-1}\sum_{i=1}^{n} u_i^2 + o_p(1) = \int_0^1 \sigma^2 + o_p(1).$$

Then as shown in the proof of Theorem 1 in Xu (2012), under H_1,

$$n^{-1}\left(\sum_{i=1}^{m}\widehat{u}_i^2-\frac{m}{n}\sum_{i=1}^{n}\widehat{u}_i^2\right)=n^{-1}\left(\sum_{i=1}^{m}u_i^2-\frac{m}{n}\sum_{i=1}^{n}u_i^2\right)+o_p(1) \tag{B.9}$$

$$\xrightarrow{p}\int_0^r\sigma^2-r\int_0^1\sigma^2. \tag{B.10}$$

For the denominator, as in the proof of Theorem 3.3.2, we decompose $\widehat{\gamma}(l)$ as

$$\widehat{\gamma}(l)=n^{-1}\sum_{i=l+1}^{n}\left(\widehat{u}_i^2-\widehat{\sigma}^2\right)\left(\widehat{u}_{i-l}^2-\widehat{\sigma}^2\right)=T_{1l}+T_{2l}+T_{3l}+T_{4l}+T_{5l},$$

where

$$\begin{aligned}
T_{1l}&=n^{-1}\sum_{i=l+1}^{n}\left(u_i^2-\sigma_i^2\right)\left(u_{i-l}^2-\sigma_{i-l}^2\right),\\
T_{2l}&=n^{-1}\sum_{i=l+1}^{n}\left(u_i^2-\sigma_i^2\right)\left(\sigma_{i-l}^2-\widehat{\sigma}^2\right),\\
T_{3l}&=n^{-1}\sum_{i=l+1}^{n}\left(\sigma_i^2-\widehat{\sigma}^2\right)\left(u_{i-l}^2-\sigma_{i-l}^2\right),\\
T_{4l}&=n^{-1}\sum_{i=l+1}^{n}\left(\sigma_i^2-\widehat{\sigma}^2\right)\left(\sigma_{i-l}^2-\widehat{\sigma}^2\right),\\
T_{5l}&=n^{-1}\sum_{i=l+1}^{n}\left(\widehat{u}_i^2-u_i^2\right)\left(\widehat{u}_{i-l}^2-u_{i-l}^2\right)+n^{-1}\sum_{i=l+1}^{n}\left(\widehat{u}_i^2-u_i^2\right)\left(u_{i-l}^2-\widehat{\sigma}^2\right)\\
&\quad+n^{-1}\sum_{i=l+1}^{n}\left(u_i^2-\widehat{\sigma}^2\right)\left(\widehat{u}_{i-l}^2-u_{i-l}^2\right),
\end{aligned}$$

and let

$$\widehat{\omega}^2=\widehat{\gamma}(l)+2\sum_{l=1}^{n-1}k(l/b)\,\widehat{\gamma}(l)=T_1+T_2+T_3+T_4+T_5,$$

where $T_i = T_{i0} + 2\sum_{l=1}^{n-1} k\left(l/b\right) T_{il},\ i = 1, 2, 3, 4, 5$. Again we have $T_1 \xrightarrow{p} \omega^2$. For the other terms, first, note that

$$\widehat{\sigma}^2 = A_{n1} - 2A_{n2} + A_{n3} = \int \sigma^2 + o_p\left(1\right)$$

since $A_{n1} = n^{-1}\sum_{i=1}^{n} u_i^2 = \int \sigma^2 + o_p\left(1\right)$ by a LLN for mixing sequences, and $A_{n2} = o_p\left(1\right)$, $A_{n3} = o_p\left(1\right)$ by Lemma B.4.2 and Lemma B.4.3. So we have $b^{-1}T_2 = O_p\left(n^{-1/2}\right)$, $b^{-1}T_3 = O_p\left(n^{-1/2}\right)$. For T_4, since

$$\begin{aligned} T_{4l} &= n^{-1}\sum_{i=l+1}^{n} \sigma_i^2\sigma_{i-l}^2 - \widehat{\sigma}^2 n^{-1}\sum_{i=l+1}^{n} \sigma_i^2 - \widehat{\sigma}^2 n^{-1}\sum_{i=l+1}^{n} \sigma_{i-l}^2 + \left(\widehat{\sigma}^2\right)^2\left(n-l\right)/n \\ &\xrightarrow{p} \int \sigma^4 - \left(\int \sigma^2\right)^2, \end{aligned}$$

we have $b^{-1}T_4 \xrightarrow{p} \left(\int k\left(x\right)dx\right)\left[\int \sigma^4 - \left(\int \sigma^2\right)^2\right]$, with the limit being zero under H_0. For T_5, again we write

$$T_{5l} = \widehat{\sigma}^2\left(2A_{n2,l} - A_{n3,l}\right) - 2A_{n12} + A_{n13} + 4A_{n22} - 2A_{n23} + A_{n33},$$

where we have used the same notations A_{n12}, A_{n13}, A_{n22}, A_{n23}, A_{n33} as in the proof of Theorem 3.3.2, with $\widehat{\beta}_i$ instead of $\overline{\beta}$ in these terms. $\widehat{\sigma}^2\left(2A_{n2,l} - A_{n3,l}\right) = o_p\left(1\right)$ by Lemma B.4.2 and Lemma B.4.3. Let $E = \cup_{j=0}^{J}\left(\tau_j + h, \tau_{j+1} - h\right)$ with $\tau_0 = 0$, $\tau_{J+1} = 1$, and $\left\{\tau_j \in (0,1) : j = 1, \cdots, J\right\}$ being points of discontinuity for $\beta\left(\cdot\right)$.

Then by Lemma B.4.1, B.4.2, and B.4.3 and the stationarity of $\{(\varepsilon_i, X_i)\}$, under both H_0 and H_1 we have

$$|A_{n12}| \leqslant \sup\left\|\widehat{\beta}_i - \beta_i\right\| n^{-1}\textstyle\sum_{i=l+1,t_i\in E}^{n}\left(\|X_i\|\,|u_i|\,u_{i-l}^2 + \|X_{i-l}\|\,|u_{i-l}|\,u_i^2\right) + n^{-1}O_p(h),$$

$$|A_{n13}| \leqslant \sup\left\|\widehat{\beta}_i - \beta_i\right\|^2 n^{-1}\textstyle\sum_{i=l+1,t_i\in E}^{n}\left(\|X_iX_i'\|\,u_{i-l}^2 + \left\|X_{i-l}X_{i-l}'\right\|u_i^2\right) + n^{-1}O_p(h),$$

$$|A_{n22}| \leqslant \sup\left\|\widehat{\beta}_i - \beta_i\right\|^2 n^{-1}\textstyle\sum_{i=l+1,t_i\in E}^{n}\|X_i\|\,|u_i|\,\|X_{i-l}\|\,|u_{i-l}| + n^{-1}O_p(h),$$

$$|A_{n23}| \leqslant \sup\left\|\widehat{\beta}_i - \beta_i\right\|^3 n^{-1}\textstyle\sum_{i=l+1,t_i\in E}^{n}\left(\|X_i\|\,|u_i|\left\|X_{i-l}X_{i-l}'\right\| + \|X_{i-l}\|\,|u_{i-l}|\,\|X_iX_i'\|\right)$$
$$+ n^{-1}O_p(h),$$

$$|A_{n33}| \leqslant \sup\left\|\widehat{\beta}_i - \beta_i\right\|^4 n^{-1}\textstyle\sum_{i=l+1,t_i\in E}^{n}\left(\|X_iX_i'\|\left\|X_{i-l}X_{i-l}'\right\|\right) + n^{-1}O_p(h).$$

So $T_{5l} = o_p(1)$, and therefore $T_5 = o_p(1)$. Combining the results of T_1 to T_5, we have

$$\widehat{\omega}^2 \xrightarrow{p} \omega^2 \tag{B.11}$$

under H_0, where $\omega^2 = \sigma^4\xi^2$, $\xi^2 = \sum_{l=-\infty}^{\infty}\psi(l)$, $\psi(l) = E\left[\left(\varepsilon_i^2 - 1\right)\left(\varepsilon_{i-l}^2 - 1\right)\right]$; and

$$\widehat{\omega}^2 = \omega^2 + b\left(\int k(x)\,dx\right)\left[\textstyle\int\sigma^4 - \left(\int\sigma^2\right)^2\right] + O_p\left(bn^{-1/2}\right) \tag{B.12}$$

under H_1.

The conclusion of Theorem 3.3.3 under H_0 follows from (B.8) and (B.11), and the conclusion under H_1 follows from (B.9) and (B.12). ■

B.4 Proof of Lemmas

Lemma B.4.1. *Suppose that Assumptions 3.3.1-3.3.6 hold, also assume* $\lambda_{\max}(\Omega_0) \leqslant C$ *for some* $0 < C < \infty$*, where* $\lambda_{\max}(\Omega_0)$ *is the largest eigenvalue of* Ω_0*. Let* $E = \cup_{j=0}^{J}(\tau_j + h, \tau_{j+1} - h)$ *with* $\tau_0 = 0$*,* $\tau_{J+1} = 1$*,* $\{\tau_j \in (0,1) : j = 1, \cdots, J\}$ *are points of discontinuity for* $\beta(\cdot)$*. Then*

$$\sup_{t \in E} \left\| \widehat{\beta}(z) - \beta(z) \right\| = O_p\left(\sqrt{\frac{\log n}{nh}}\right) + O_p\left(h^2\right).$$

Proof. The proof is similar to the proof of Theorem 10 in Hansen (2008). Let $D(z) = S_0(z) - S_1(z) S_2(z)^{-1} S_1(z)$, and let $\mu_k = \int u^k K(u)\, du$ for $k = 0, 1, 2$. By Lemma B.5.1 and Lemma B.5.2, we have

$$\sup_{z \in E} \|T_0(z) - S_0(z)\beta(z)\| = O_p\left(\sqrt{\log n/(nh)}\right) + O_p\left(h^2\right),$$

$$\sup_{z \in E} \left\|h^{-1}(T_1(z) - S_1(t)\beta(z))\right\| = O_p\left(\sqrt{\log n/(nh)}\right) + O_p(h),$$

$$\sup_{z \in (h, 1-h)} \left\|h^{-k} S_k(z) - \mu_k \Omega_0\right\| = O_p\left(\sqrt{\log n/(nh)}\right),\ k = 0, 1, 2,$$

$$\sup_{z \in (h, 1-h)} \|D(z) - \Omega_0\| = O_p\left(\sqrt{\log n/(nh)}\right).$$

Since

$$\widehat{\beta}(z) - \beta(z) = D(t)^{-1}\left[(T_0(z) - S_0(z)\beta(z)) - S_1(z) S_2(z)^{-1}(T_1(z) - S_1(z)\beta(z))\right],$$

we have

$$\sup_{z\in E}\left\|\widehat{\beta}(z)-\beta(z)\right\| \leqslant \Omega_0^{-1}\sup_{z\in E}\left\|T_0(z)-S_0(z)\beta(z)\right\|$$

$$+\,\Omega_0^{-1}\sup_{z\in E}\left\|h^{-1}S_1(z)\right\|\mu_2^{-1}\Omega_0^{-1}\sup_{z\in E}\left\|h^{-1}\left(T_1(z)-S_1(z)\beta(z)\right)\right\|$$

$$=O_p\left(\sqrt{\log n/(nh)}\right)+O_p\left(h^2\right).$$

■

Lemma B.4.2. *Suppose that Assumptions 3.3.1-3.3.6 hold. Then*

$$\sqrt{n}\left|A_{n3}\right|=O_p\left(\log n/\left(\sqrt{n}h\right)\right)+O_p\left(\sqrt{n}h^4\right)=o_p(1).$$

Proof. Since $A_{n3}=n^{-1}\sum_{i=1}^{n}\left(\widehat{\beta}_i-\beta_i\right)'X_iX_i'\left(\widehat{\beta}_i-\beta_i\right)$, by Lemma B.4.1,

$$\sqrt{n}\left|A_{n3}\right|\leqslant\sqrt{n}\sup_{z\in E}\left\|\widehat{\beta}(z)-\beta(z)\right\|^2 n^{-1}\sum_{i=1,z_i\in E}^{n}\left\|X_iX_i'\right\|+n^{-1/2}O_p(h)$$

$$=\sqrt{n}\left(O_p\left(\frac{\log n}{nh}\right)+O_p\left(h^4\right)\right)O_p(1)+o_p(1)=o_p(1),$$

given that $h\to 0$, $nh^2\to\infty$, $nh^8\to 0$ as $n\to\infty$. ■

Lemma B.4.3. *Suppose that Assumptions 3.3.1-3.3.6 hold. Then*

$$\sqrt{n}\left|A_{n2}\right|=o_p(1).$$

Proof. We follow the same steps as in Kristensen (2011, proof of Lemma B.2). Here A_{n2} is analogous to the term $S^{w}_{g-\widehat{g},e}$ in that lemma except for that we use the local linear estimator of $\beta(z)$ while Kristensen (2011) uses the local constant estimator. To start with, we introduce the following notations. For $k = 0, 1, 2$, define

$$S_k^*(z) = n^{-1}\sum\nolimits_{i=1}^{n} (z_i - z)^k X_i X_i' \beta_i K_h(z_i - z),$$
$$T_k^*(z) = n^{-1}\sum\nolimits_{i=1}^{n} (z_i - z)^k X_i u_i K_h(z_i - z).$$

Note that $A_{n2} = n^{-1}\sum_{i=1}^{n}\left(\widehat{\beta}_i - \beta_i\right)' X_i u_i$, where because

$$\begin{aligned}
&\widehat{\beta}(z) - \beta(z)\\
&= D(t)^{-1}\left[T_0(z) - S_1(z) S_2(z)^{-1} T_1(z) - D(z)\beta(z)\right]\\
&= D(z)^{-1}\left[T_0(z) - S_0(z)\beta(z)\right] - D(z)^{-1} S_1(z) S_2(z)^{-1}\left[T_1(z) - S_1(z)\beta(z)\right]\\
&= \Omega_0^{-1}\left[S_0^*(z) - \Omega_0\beta(z)\right] + \Omega_0^{-1} T_0^*(z) - \Omega_0^{-1}\left(S_0(z)\beta(z) - \Omega_0\beta(z)\right)\\
&- D(z)^{-1}\left(D(z) - \Omega_0\right)\Omega_0^{-1}\left[T_0(z) - S_0(z)\beta(z)\right]\\
&- D(z)^{-1} S_1(z) S_2(z)^{-1}\left[T_1(z) - S_1(z)\beta(z)\right],
\end{aligned}$$

we can write $A_{n2} = B_{n1} + B_{n2} - B_{n3} - B_{n4} - B_{n5}$, where

$$B_{n1} = n^{-1}\textstyle\sum_{i=1}^{n} \left[S_0^*(z_i) - \Omega_0\beta(z_i)\right]' \Omega_0^{-1} X_i u_i,$$

$$B_{n2} = n^{-1}\textstyle\sum_{i=1}^{n} T_0^*(z_i)' \Omega_0^{-1} X_i u_i,$$

$$B_{n3} = n^{-1}\textstyle\sum_{i=1}^{n} \left[S_0(z_i)\beta(z_i) - \Omega_0\beta(z_i)\right]' \Omega_0^{-1} X_i u_i,$$

$$B_{n4} = n^{-1}\textstyle\sum_{i=1}^{n} \left[T_0(z_i) - S_0(z_i)\beta(z_i)\right]' \Omega_0^{-1} \left(D(z_i) - \Omega_0\right) D(z_i)^{-1} X_i u_i,$$

$$B_{n5} = n^{-1}\textstyle\sum_{i=1}^{n} \left[T_1(z_i) - S_1(z_i)\beta(z_i)\right]' S_2(z)^{-1} S_1(z) D(z_i)^{-1} X_i u_i.$$

We now prove that for some $\epsilon, \delta > 0$:

$$B_{n1} = O_p\left(n^{-1+\epsilon/2} h^{-(1+\delta)/(2+\delta)}\right) + O_p\left(n^{-1/2} h^2\right), \tag{B.13}$$

$$B_{n2} = O_p\left(n^{-1+\epsilon/2} h^{-(1+\delta)/(2+\delta)}\right), \tag{B.14}$$

$$B_{n3} = O_p\left(n^{-1+\epsilon/2} h^{-(1+\delta)/(2+\delta)}\right) + O_p\left(n^{-1/2} h^2\right), \tag{B.15}$$

$$B_{n4} = O_p\left(\log n/(nh)\right) + O_p\left(h^4\right), \tag{B.16}$$

$$B_{n5} = O_p\left(\log n/(nh)\right) + O_p\left(\sqrt{h \log n/n}\right). \tag{B.17}$$

Then under the conditions $n \to \infty$, $h \to 0$, $n^{1-\epsilon/2} h^{(1+\delta)/(2+\delta)} \to \infty$ (which is trivially implied by $nh \to \infty$ for small ϵ and δ), we have $\sqrt{n}B_{n1} = \sqrt{n}B_{n2} = \sqrt{n}B_{n3} = o_p(1)$; further, under $nh^8 \to 0$, $\sqrt{n}B_{n4} = \sqrt{n}B_{n5} = o_p(1)$. This will complete the proof of Lemma B.4.3.

It remains to show (B.13)-(B.17). By following the same arguments as in Kristensen (2011, proof of Claim B.2.1), we will rely on some results of U-statistics for weakly dependent processes to prove (B.13)-(B.15). We briefly describe the proof for (B.13), then (B.14) and (B.15) will follow trivially. Define $w_i = (z_i, X_i, u_i)$ and

$$\begin{aligned} a(w_i, w_j) &= \left[K_h(z_j - z_i) X_j X_j' \beta_j - \Omega_0 \beta_i\right]' \Omega_0^{-1} X_i u_i, \\ \phi(w_i, w_j) &= a(w_i, w_j) + a(w_j, w_i), \\ U_n &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i} \phi(w_i, w_j). \end{aligned}$$

By viewing t_i as i.i.d. uniform random variables independent of (X_i, u_i), we can write

$$\begin{aligned} B_{n1} &= n^{-1} \sum_{i=1}^{n} \left[S_0^*(z_i) - \Omega_0 \beta(z_i)\right]' \Omega_0^{-1} X_i u_i = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} a(w_i, w_j) \\ &= n^{-1}(n-1) U_n + n^{-2} \sum_{i=1}^{n} a(w_i, w_i), \end{aligned}$$

where U_n can be viewed as an U-statistic, and $n^{-2} \sum_{i=1}^{n} a(w_i, w_i) = O_p(n^{-1})$. For the U-statistic U_n, we can further use the Hoeffding decomposition to write

$$U_n = 2n^{-1} \sum_{i=1}^{n} \overline{\phi}(w_i) + R_n,$$

where

$$\overline{\phi}(w) = E[\phi(w, w_j)] = E[a(w, w_j)] + E[a(w_j, w)]$$
$$= \left[\int_0^1 K_h(s - z)\,\beta(s)\,ds - \beta(z)\right]' Xe + 0,$$

with $w = (z, X, e)$, and R_n is a remainder. It can be shown that $R_n = O_p\left(n^{-1+\epsilon/2} s_{n,\delta}\right)$, where $s_{n,\delta} = \sup_{i,j} E\left[|\phi(w_i, w_j)|^{2+\delta}\right]^{1/(2+\delta)} = O\left(h^{-(1+\delta)/(2+\delta)}\right)$.[2] For the first term in the Hoeffding decomposition, we can show that

$$n^{-1}\sum_{i=1}^{n}\overline{\phi}(w_i) = O_p\left(n^{-1/2}h^2\right).$$

This is verified by noticing $E\left[n^{-1}\sum_{i=1}^{n}\overline{\phi}(w_i)\right] = 0$, and

$$E\left[\left(n^{-1}\sum_{i=1}^{n}\overline{\phi}(w_i)\right)^2\right] = n^{-2}\sum_{i=1}^{n}E\left[\overline{\phi}(w_i)^2\right] + n^{-2}\sum_{i=1}^{n}\sum_{j\neq i}E\left[\overline{\phi}(w_i)\,\overline{\phi}(w_j)\right]$$
$$= B_{n6} + B_{n7},$$

where we can show that $B_{n6} = O\left(n^{-1}h^4\right)$ and $B_{n7} = O\left(n^{-1}h^{4/(1+\delta)}\right)$.

For (B.16), we have

$$|B_{n4}| \leqslant \sup_{z\in E}\|T_0(z) - S_0(z)\,\beta(z)\| \sup_{z\in E}\left\|\Omega_0^{-1}(D(z) - \Omega_0)\right\| *$$
$$n^{-1}\sum_{i=1,z_i\in E}^{n}\left\|D(z_i)^{-1}\right\|\|X_i\|\,|u_i| + n^{-1}O_p(h),$$

[2]In Kristensen's proof, $s_{n,\delta} = O\left(h^{-(1+\delta)/(2+\delta)}\right)$ is shown but then viewed as $O(1)$ by mistake.

where by Lemma B.5.1,

$$\sup_{z\in E}\|T_0(z)-S_0(z)\beta(z)\| = O_p\left(\sqrt{\log n/(nh)}\right)+O_p\left(h^2\right),$$

and by Lemma B.5.2,

$$\sup_{z\in E}\left\|\Omega_0^{-1}(D(z)-\Omega_0)\right\| = O_p\left(\sqrt{\log n/(nh)}\right),$$

also, $n^{-1}\sum_{i=1}^{n}\left\|D(z_i)^{-1}\right\|\|X_i\||u_i| = O_p(1)$. So $|B_{n4}| = O_p(\log n/(nh))+O_p(h^4)$. For (B.17), we have

$$|B_{n5}| \leqslant \sup_{z\in E}\left\|h^{-1}[T_1(z)-S_1(z)\beta(z)]\right\|\sup_{z\in E}\left\|\left(h^{-2}S_2(z)\right)^{-1}h^{-1}S_1(z)\right\| * \\ n^{-1}\sum_{i=1,z_i\in E}^{n}\left\|D(z_i)^{-1}\right\|\|X_i\||u_i| + n^{-1}O_p(h),$$

where by Lemma B.5.1, $\sup_{z\in E}\|h^{-1}[T_1(z)-S_1(z)\beta(z)]\| = O_p\left(\sqrt{\log n/(nh)}\right)+O_p(h)$, and by Lemma B.5.2, $\sup_{z\in E}\left\|(h^{-2}S_2(z))^{-1}h^{-1}S_1(z)\right\| = O_p\left(\sqrt{\log n/(nh)}\right)$. So $|B_{n5}| = O_p(\log n/(nh))+O_p\left(h\sqrt{\log n/(nh)}\right)$. ■

B.5 Proof of Auxiliary Lemmas

We provide the following results of uniform convergence rate.

Lemma B.5.1. *Under Assumptions 3.3.1-3.3.6, for* $k = 0, 1$,

$$\sup_{z \in E} \left\| T_k(z) - S_k(z)\beta(z) \right\| = O_p\left(h^k \sqrt{\frac{\log n}{nh}} \right) + O_p\left(h^2\right),$$

where $E = \cup_{j=0}^{J} (\tau_j + h, \tau_{j+1} - h)$ *with* $\tau_0 = 0$, $\tau_{J+1} = 1$, $\{\tau_j \in (0,1) : j = 1, \cdots, J\}$ *are points of discontinuity for* $\beta(\cdot)$.

Proof. Let $T_k(z) - S_k(z)\beta(z) = M_k(z) + T_k^*(z)$, where

$$M_k(z) = n^{-1} \sum\nolimits_{i=1}^{n} (z_i - z)^k X_i X_i' (\beta(z_i) - \beta(z)) K_h(z_i - z),$$
$$T_k^*(z) = n^{-1} \sum\nolimits_{i=1}^{n} (z_i - z)^k X_i u_i K_h(z_i - z),$$

then under H_0, by Theorem 2 in Hansen (2008), or under H_1, by Theorem 1 in Kristensen (2009),

$$\sup_{z \in E} \left\| h^{-k} M_k(z) - E\left[h^{-k} M_k(z)\right] \right\| = O_p\left(\sqrt{\log n / (nh)}\right),$$
$$\sup_{z \in E} \left\| h^{-k} T_k^*(z) - E\left[h^{-k} T_k^*(z)\right] \right\| = O_p\left(\sqrt{\log n / (nh)}\right).$$

Now it is sufficient to show that $\sup_{z \in E} \|E[M_k(z)]\| = O(h^2)$ since $E[T_k^*(z)] = 0$. Uniformly over $z \in E$, by change of variable and Taylor expansion arguments,

$$\begin{aligned} E\left[h^{-k} M_k(z)\right] &= \int_0^1 \Omega_0 (\beta(s) - \beta(z)) \left(\frac{s-z}{h}\right)^k K_h(s-z)\, ds + O\left(n^{-1}\right) \\ &= h\beta^{(1)}(z)\,\Omega_0 \int_{-z/h}^{(1-z)/h} u^{k+1} K(u)\, du + O\left(h^2\right) \\ &= h\beta^{(1)}(z)\,\Omega_0 \mu_{k+1} + O\left(h^2\right), \end{aligned}$$

so for $k = 0, 1$, $\sup_{z \in E} \|E[M_k(z)]\| = O(h^2)$ since $\mu_0 = 1$, $\mu_1 = 0$. ■

Lemma B.5.2. *Under Assumptions 3.3.1-3.3.6, we have*

$$\sup_{z \in (h,1-h)} \|D(z) - \Omega_0\| = O_p\left(\sqrt{\frac{\log n}{nh}}\right).$$

Proof. Note that for $k = 0, 1, 2$,

$$h^{-k} S_k(z) = n^{-1} \sum\nolimits_{i=1}^n X_i X_i' \left(\frac{z_i - z}{h}\right)^k K_h(z_i - z),$$

where as in the proof of Lemma B.5.1, $\sup_{z \in (h,1-h)} \left\|h^{-k} S_k(z) - E\left[h^{-k} S_k(z)\right]\right\| = O_p\left(\sqrt{\log n/(nh)}\right)$. Also, by the Riemann sum approximation of an integral, uniformly over $z \in (h, 1-h)$, $E\left[h^{-k} S_k(z)\right] = \mu_k \Omega_0 + O(n^{-1})$. Again by $\mu_0 = 1$,

$\mu_1 = 0$, we have $E\left[S_0(z)\right] = \Omega_0 + O(n^{-1})$, $E\left[h^{-1}S_1(z)\right] = O(n^{-1})$, $E\left[h^{-2}S_2(z)\right] = \mu_2\Omega_0 + O(n^{-1})$. Thus

$$\sup_{z\in(h,1-h)} \|S_0(z) - \Omega_0\| = O_p\left(\sqrt{\log n/(nh)}\right),$$

$$\sup_{z\in(h,1-h)} \left\|h^{-1}S_1(z)\right\| = O_p\left(\sqrt{\log n/(nh)}\right),$$

$$\sup_{z\in(h,1-h)} \left\|h^{-2}S_2(z) - \mu_2\Omega_0\right\| = O_p\left(\sqrt{\log n/(nh)}\right).$$

So

$$\begin{aligned}
&\sup_{z\in(h,1-h)} \|D(z) - \Omega_0\| \\
&\leqslant \sup_{z\in(h,1-h)} \|S_0(z) - \Omega_0\| + \sup_{z\in(h,1-h)} \left\|h^{-1}S_1(z)\left[h^{-2}S_2(z)\right]^{-1} h^{-1}S_1(z)\right\| \\
&= O_p\left(\sqrt{\log n/(nh)}\right).
\end{aligned}$$

■

APPENDIX C

TABLES

Table C.1

Size and power under mean specification M0-M2: IID error

DGP		n = 300				n = 600			
		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M0	V0	0.034	0.032	0.034	1.00	0.040	0.039	0.040	1.00
	V1	0.142	0.137	0.141	1.00	0.282	0.282	0.279	1.00
	V2	0.443	0.444	0.434	1.00	0.792	0.780	0.781	1.00
	V3	0.782	0.777	0.760	1.00	0.981	0.980	0.983	1.00
	V4	0.119	0.116	0.120	1.00	0.297	0.302	0.302	1.00
	V5	0.090	0.091	0.092	1.00	0.159	0.158	0.157	1.00
M1	V0	0.026	0.043	0.025	0.98	0.031	0.065	0.035	0.90
	V1	0.135	0.155	0.125	1.00	0.252	0.316	0.248	0.92
	V2	0.421	0.413	0.425	1.00	0.786	0.791	0.788	0.94
	V3	0.785	0.764	0.777	1.00	0.980	0.982	0.979	1.00
	V4	0.108	0.061	0.106	1.00	0.289	0.164	0.291	1.00
	V5	0.094	0.145	0.095	0.96	0.147	0.307	0.134	0.86
M2	V0	0.034	0.232	0.033	0.76	0.042	0.609	0.035	0.44
	V1	0.110	0.387	0.113	0.76	0.275	0.802	0.266	0.48
	V2	0.413	0.584	0.409	0.80	0.786	0.935	0.784	0.48
	V3	0.776	0.823	0.768	0.80	0.990	0.994	0.986	0.50
	V4	0.104	0.088	0.106	0.76	0.274	0.220	0.281	0.46
	V5	0.091	0.440	0.087	0.70	0.139	0.902	0.138	0.42

Note: "True" stands for the CUSQ test based on the error term series (without mean); "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. "*h_med*" stands for the median of selected bandwidths by AIC. The mean specifications are M0: $\beta(z) = 0.2\exp(-0.7) \approx 0.099$; M1: $\beta(z) = 0.2\exp(0.7z - 0.7) \in (0.099, 0.200)$; M2: $\beta(z) = 0.2\exp(1.1z - 0.7) \in (0.099, 0.300)$.

Table C.2

Size and power under mean specification M0-M2: GARCH(1,1) error

		$n = 300$				$n = 600$			
DGP		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M0	V0	0.076	0.079	0.076	1.00	0.069	0.067	0.069	1.00
	V1	0.140	0.140	0.141	1.00	0.271	0.271	0.271	1.00
	V2	0.394	0.381	0.389	1.00	0.682	0.682	0.685	1.00
	V3	0.676	0.667	0.664	1.00	0.934	0.931	0.932	1.00
	V4	0.115	0.117	0.123	1.00	0.259	0.252	0.252	1.00
	V5	0.114	0.108	0.108	1.00	0.195	0.201	0.198	1.00
M1	V0	0.062	0.069	0.057	0.96	0.071	0.082	0.073	0.94
	V1	0.140	0.157	0.139	1.00	0.239	0.301	0.236	0.94
	V2	0.398	0.403	0.385	1.00	0.666	0.688	0.666	0.96
	V3	0.676	0.684	0.668	1.00	0.952	0.951	0.948	0.94
	V4	0.093	0.062	0.095	1.00	0.261	0.162	0.266	0.94
	V5	0.098	0.137	0.098	0.94	0.158	0.265	0.149	0.90
M2	V0	0.051	0.251	0.056	0.66	0.067	0.559	0.065	0.46
	V1	0.140	0.394	0.142	0.74	0.267	0.710	0.263	0.46
	V2	0.398	0.575	0.396	0.76	0.695	0.898	0.686	0.48
	V3	0.672	0.756	0.657	0.76	0.944	0.976	0.944	0.56
	V4	0.104	0.102	0.115	0.72	0.230	0.196	0.229	0.48
	V5	0.108	0.419	0.107	0.66	0.184	0.806	0.181	0.42

Note: "True" stands for the CUSQ test based on the error term series (without mean); "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. "*h_med*" stands for the median of selected bandwidths by AIC. The mean specifications are M0: $\beta(z) = 0.2\exp(-0.7) \approx 0.099$; M1: $\beta(z) = 0.2\exp(0.7z - 0.7) \in (0.099, 0.200)$; M2: $\beta(z) = 0.2\exp(1.1z - 0.7) \in (0.099, 0.300)$.

Table C.3

Size and power under mean specification M3-M5: $\tau = 0.5$, IID error

		$n = 300$				$n = 600$			
DGP		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M3	V0	0.042	0.023	0.041	0.14	0.049	0.027	0.049	0.10
	V1	0.133	0.076	0.141	0.16	0.269	0.154	0.268	0.10
	V2	0.443	0.305	0.436	0.16	0.794	0.645	0.801	0.10
	V3	0.781	0.655	0.767	0.16	0.982	0.955	0.983	0.12
	V4	0.092	0.049	0.119	0.16	0.290	0.166	0.317	0.10
	V5	0.097	0.044	0.077	0.14	0.179	0.091	0.142	0.10
M4	V0	0.040	0.005	0.041	0.08	0.041	0.006	0.051	0.06
	V1	0.136	0.021	0.150	0.08	0.275	0.043	0.309	0.06
	V2	0.456	0.104	0.472	0.08	0.811	0.274	0.816	0.06
	V3	0.791	0.318	0.780	0.08	0.982	0.776	0.986	0.06
	V4	0.097	0.019	0.137	0.08	0.267	0.047	0.380	0.06
	V5	0.085	0.011	0.057	0.08	0.156	0.024	0.094	0.04
M5	V0	0.030	0.003	0.054	0.06	0.038	0.010	0.069	0.04
	V1	0.144	0.010	0.161	0.06	0.288	0.012	0.347	0.04
	V2	0.444	0.032	0.472	0.06	0.799	0.107	0.833	0.04
	V3	0.780	0.100	0.789	0.06	0.970	0.377	0.984	0.04
	V4	0.102	0.001	0.182	0.06	0.326	0.021	0.440	0.04
	V5	0.086	0.009	0.053	0.04	0.154	0.013	0.076	0.04

Note: "True" stands for the CUSQ test based on the error term series (without mean); "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. "*h_med*" stands for the median of selected bandwidths by AIC. The mean specifications are M3: $\beta(z) = 1.5 - 0.1\mathbb{I}(z > 0.5)$; M4: $\beta(z) = 1.5 - 0.2\mathbb{I}(z > 0.5)$; M5: $\beta(z) = 1.5 - 0.3\mathbb{I}(z > 0.5)$.

Table C.4

Size and power under mean specification M3-M5: $\tau = 0.5$, GARCH(1,1) error

		$n = 300$				$n = 600$			
DGP		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M3	V0	0.053	0.031	0.047	0.14	0.061	0.036	0.068	0.10
	V1	0.179	0.113	0.168	0.16	0.263	0.172	0.254	0.10
	V2	0.415	0.321	0.424	0.16	0.682	0.577	0.671	0.12
	V3	0.723	0.596	0.697	0.16	0.924	0.873	0.926	0.12
	V4	0.124	0.065	0.124	0.14	0.277	0.192	0.319	0.10
	V5	0.118	0.078	0.110	0.14	0.194	0.123	0.157	0.10
M4	V0	0.059	0.009	0.057	0.08	0.049	0.014	0.064	0.06
	V1	0.143	0.031	0.146	0.08	0.271	0.057	0.278	0.06
	V2	0.378	0.094	0.371	0.08	0.684	0.310	0.702	0.06
	V3	0.692	0.306	0.688	0.08	0.928	0.703	0.934	0.06
	V4	0.101	0.023	0.131	0.08	0.239	0.058	0.327	0.06
	V5	0.136	0.038	0.098	0.08	0.182	0.046	0.134	0.04
M5	V0	0.072	0.006	0.059	0.06	0.074	0.004	0.094	0.04
	V1	0.160	0.009	0.162	0.06	0.296	0.026	0.328	0.04
	V2	0.415	0.045	0.421	0.06	0.689	0.115	0.717	0.04
	V3	0.665	0.135	0.654	0.06	0.941	0.384	0.946	0.04
	V4	0.104	0.007	0.160	0.06	0.220	0.022	0.334	0.04
	V5	0.121	0.009	0.081	0.06	0.186	0.022	0.117	0.04

Note: "True" stands for the CUSQ test based on the error term series (without mean); "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. "*h_med*" stands for the median of selected bandwidths by AIC. The mean specifications are M3: $\beta(z) = 1.5 - 0.1\mathbb{I}(z > 0.5)$; M4: $\beta(z) = 1.5 - 0.2\mathbb{I}(z > 0.5)$; M5: $\beta(z) = 1.5 - 0.3\mathbb{I}(z > 0.5)$.

Table C.5

Size and power under mean specification M3-M5: $\tau = 0.45$, IID error

		$n = 300$				$n = 600$			
DGP		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M3	V0	0.039	0.031	0.037	0.14	0.038	0.048	0.046	0.10
	V1	0.143	0.025	0.132	0.14	0.296	0.047	0.261	0.10
	V2	0.456	0.169	0.423	0.16	0.798	0.421	0.765	0.10
	V3	0.775	0.482	0.741	0.16	0.984	0.857	0.980	0.12
	V4	0.113	0.077	0.139	0.14	0.306	0.226	0.336	0.10
	V5	0.078	0.066	0.065	0.14	0.159	0.142	0.119	0.10
M4	V0	0.034	0.118	0.041	0.08	0.038	0.351	0.042	0.06
	V1	0.163	0.019	0.127	0.08	0.271	0.035	0.262	0.06
	V2	0.448	0.009	0.399	0.08	0.801	0.017	0.745	0.06
	V3	0.759	0.051	0.684	0.08	0.983	0.111	0.978	0.06
	V4	0.097	0.174	0.151	0.08	0.284	0.458	0.391	0.06
	V5	0.094	0.159	0.058	0.08	0.175	0.458	0.112	0.06
M5	V0	0.034	0.526	0.058	0.06	0.040	0.957	0.073	0.04
	V1	0.145	0.180	0.145	0.06	0.297	0.561	0.290	0.04
	V2	0.430	0.042	0.383	0.06	0.799	0.128	0.780	0.04
	V3	0.749	0.007	0.686	0.06	0.989	0.016	0.977	0.04
	V4	0.107	0.465	0.186	0.06	0.300	0.904	0.439	0.04
	V5	0.092	0.592	0.058	0.06	0.146	0.982	0.089	0.04

Note: "True" stands for the CUSQ test based on the error term series (without mean); "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. "*h_med*" stands for the median of selected bandwidths by AIC. The mean specifications are M3: $\beta(z) = 1.5 - 0.1\mathbb{I}(z > 0.45)$; M4: $\beta(z) = 1.5 - 0.2\mathbb{I}(z > 0.45)$; M5: $\beta(z) = 1.5 - 0.3\mathbb{I}(z > 0.45)$.

Table C.6

Size and power under mean specification M3-M5: $\tau = 0.45$, GARCH(1,1) error

		$n = 300$				$n = 600$			
DGP		True	OLS	LL	h_med	True	OLS	LL	h_med
M3	V0	0.063	0.056	0.053	0.14	0.059	0.084	0.069	0.10
	V1	0.148	0.054	0.140	0.14	0.261	0.077	0.234	0.10
	V2	0.390	0.176	0.366	0.16	0.668	0.358	0.638	0.10
	V3	0.670	0.422	0.623	0.18	0.923	0.745	0.915	0.12
	V4	0.093	0.095	0.115	0.16	0.252	0.219	0.301	0.10
	V5	0.129	0.096	0.112	0.14	0.187	0.166	0.151	0.10
M4	V0	0.069	0.142	0.060	0.08	0.067	0.358	0.081	0.06
	V1	0.152	0.032	0.136	0.08	0.244	0.060	0.215	0.06
	V2	0.385	0.023	0.347	0.08	0.669	0.021	0.615	0.06
	V3	0.659	0.070	0.608	0.08	0.928	0.112	0.910	0.06
	V4	0.113	0.166	0.151	0.08	0.243	0.443	0.299	0.06
	V5	0.126	0.197	0.089	0.08	0.172	0.475	0.124	0.04
M5	V0	0.070	0.491	0.077	0.06	0.061	0.939	0.079	0.04
	V1	0.177	0.182	0.140	0.06	0.271	0.576	0.253	0.04
	V2	0.426	0.057	0.370	0.06	0.692	0.175	0.659	0.04
	V3	0.677	0.016	0.626	0.06	0.932	0.033	0.917	0.04
	V4	0.109	0.450	0.169	0.06	0.223	0.871	0.346	0.04
	V5	0.113	0.576	0.096	0.04	0.176	0.974	0.127	0.04

Note: "True" stands for the CUSQ test based on the error term series (without mean); "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. "h_med" stands for the median of selected bandwidths by AIC. The mean specifications are M3: $\beta(z) = 1.5 - 0.1\mathbb{I}(z > 0.45)$; M4: $\beta(z) = 1.5 - 0.2\mathbb{I}(z > 0.45)$; M5: $\beta(z) = 1.5 - 0.3\mathbb{I}(z > 0.45)$.

Table C.7

Size distortion under different mean jump: IID error

		$n = 300$							
		$\tau = 0.3$		$\tau = 0.4$		$\tau = 0.45$		$\tau = 0.5$	
DGP		OLS	LL	OLS	LL	OLS	LL	OLS	LL
M3	V0	0.258	0.034	0.078	0.036	0.035	0.040	0.023	0.041
M4	V0	1.000	0.056	0.744	0.042	0.111	0.049	0.005	0.041
M5	V0	1.000	0.066	1.000	0.057	0.509	0.054	0.003	0.054
		$n = 600$							
		$\tau = 0.3$		$\tau = 0.4$		$\tau = 0.45$		$\tau = 0.5$	
DGP		OLS	LL	OLS	LL	OLS	LL	OLS	LL
M3	V0	0.580	0.038	0.166	0.041	0.055	0.051	0.027	0.049
M4	V0	1.000	0.068	0.991	0.053	0.325	0.065	0.006	0.051
M5	V0	1.000	0.075	1.000	0.082	0.980	0.085	0.010	0.069

Note: "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. The mean specifications are M3: $\beta(z) = 1.5 - 0.1\mathbb{I}(z > \tau)$; M4: $\beta(z) = 1.5 - 0.2\mathbb{I}(z > \tau)$; M5: $\beta(z) = 1.5 - 0.3\mathbb{I}(z > \tau)$.

Table C.8

Size distortion under different mean jump: GARCH(1,1) error

		$n = 300$							
		$\tau = 0.3$		$\tau = 0.4$		$\tau = 0.45$		$\tau = 0.5$	
DGP		OLS	LL	OLS	LL	OLS	LL	OLS	LL
M3	V0	0.281	0.059	0.105	0.062	0.056	0.054	0.031	0.047
M4	V0	0.999	0.079	0.670	0.080	0.149	0.063	0.009	0.057
M5	V0	1.000	0.088	1.000	0.087	0.535	0.071	0.006	0.059
		$n = 600$							
		$\tau = 0.3$		$\tau = 0.4$		$\tau = 0.45$		$\tau = 0.5$	
DGP		OLS	LL	OLS	LL	OLS	LL	OLS	LL
M3	V0	0.523	0.068	0.193	0.065	0.076	0.062	0.036	0.068
M4	V0	1.000	0.071	0.974	0.070	0.343	0.065	0.014	0.064
M5	V0	1.000	0.100	1.000	0.093	0.942	0.088	0.004	0.094

Note: "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. The mean specifications are M3: $\beta(z) = 1.5 - 0.1\mathbb{I}(z > \tau)$; M4: $\beta(z) = 1.5 - 0.2\mathbb{I}(z > \tau)$; M5: $\beta(z) = 1.5 - 0.3\mathbb{I}(z > \tau)$.

Table C.9

Smooth mean change and proportional volatility change in H_1

IID error									
		$n = 300$				$n = 600$			
DGP		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M0'	V0'	0.034	0.055	0.033	0.32	0.029	0.074	0.029	0.26
	V1'	0.504	0.058	0.500	0.22	0.923	0.167	0.924	0.20
	V2'	0.498	0.208	0.484	0.32	0.925	0.582	0.922	0.28
	V3'	0.502	0.350	0.497	0.40	0.927	0.793	0.930	0.32
GARCH(1,1) error									
		$n = 300$				$n = 600$			
DGP		True	OLS	LL	*h_med*	True	OLS	LL	*h_med*
M0'	V0'	0.056	0.061	0.053	0.32	0.058	0.094	0.059	0.26
	V1'	0.416	0.072	0.414	0.23	0.817	0.174	0.813	0.20
	V2'	0.402	0.206	0.402	0.32	0.812	0.468	0.811	0.28
	V3'	0.399	0.296	0.404	0.40	0.825	0.661	0.824	0.32

Note: "OLS" stands for the test based on OLS residuals; "LL" stands for the test based on local linear residuals. The mean specification is M0': $\beta(z) = 0.5z + \exp\left(-4(z-0.5)^2\right)$. The variance specifications are V0': $\sigma^2(z) \equiv 1$; V1': $\sigma(z) = 0.5\beta(z)$; V2': $\sigma(z) = \beta(z)$; V3': $\sigma(z) = 1.5\beta(z)$.

Table C.10
Results of the Unit Root Test on Each Variable

		Test Statistics				Test Statistics	
Variable	Model	coef.	t-value	Variable	Model	coef.	t-value
	CANADA				JAPAN		
ER	C(12)	-0.013	-2.523	ER	CT(11)	-0.022	-2.834*
CPI	C(12)	-0.002	-3.374**	CPI	CT(7)	-0.022	-6.228***
PPI	C(13)	-0.004	-3.612***	PPI	C(13)	-0.006	-2.752*
TPI	C(13)	-0.008	-3.139**	TPI	C(13)	-0.013	-2.180
	FRANCE				GERMANY		
ER	C(3)	-0.015	-2.134	ER	C(1)	-0.003	-1.267
CPI	CT(7)	-0.001	-1.139	CPI	C(9)	-0.002	-1.905
PPI	No Data Available			PPI	C(6)	-0.004	-2.529
TPI	N(12)	0.000	-1.762	TPI	C(14)	-0.009	-1.933
	ITALY				UK		
ER	C(1)	-0.008	-1.963	ER	C(8)	-0.022	-2.814*
CPI	CT(13)	-0.001	-1.000	CPI	CT(13)	-0.006	-3.569***
PPI	CT(2)	-0.022	-3.266**	PPI	C(14)	-0.005	-6.420***
TPI	C(14)	-0.008	-3.215**	TPI	C(9)	-0.007	-2.204
	US						
CPI	C(13)	-0.001	-2.687*				
PPI	CT(14)	-0.011	-2.848*				
TPI	C(13)	-0.005	-2.904**				

Note:

1. ER represents nominal exchange rates; CPI, PPI, TPI represent consumer price index, producer price index, trade price index, respectively. All variables are in the log form.

2. Critical values are -2.57 for 10%(*), -2.87 for 5%(**), -3.44 for 1%(***) Hamilton (1994) p763 Table B.6.

3. "CT", "C", "N" represent "*constant and time trend*", "*constant only*", "*none*", respectively.

4. Numbers in the parentheses are orders of lagged dependent variables, selected by a data-driven lag selection procedure in Ng and Perron (1995).

Table C.11
Results of the Linear Cointegration Test

		OLS Est. of Coef.s			Residual Based Test	
	# of obs.	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	coef.	t-value
			CANADA			
CPI-based	429	-0.489	0.733	-0.567	-0.020	-2.826
PPI-based	429	0.290	1.347	-1.368	-0.026	-2.369
TPI-based	429	0.761	0.999	-1.107	-0.032	-2.119
			JAPAN			
CPI-based	429	7.900	0.136	-0.840	-0.022	-2.675
PPI-based	429	4.683	1.124	-1.150	-0.020	-2.620
TPI-based	429	4.412	1.032	-0.964	-1.789	-0.035
			FRANCE			
CPI-based	309	2.407	2.359	-2.574	-0.029	-2.422
PPI-based			No Data Available			
TPI-based	108	11.556	-0.984	-1.150	-0.135	-2.963
			GERMANY			
CPI-based	225	-15.644	7.278	-4.114	-0.045	-2.377
PPI-based	213	-5.347	3.592	-2.445	-0.033	-1.834
TPI-based	309	-3.148	2.305	-1.570	-0.078	-3.031
			ITALY			
CPI-based	309	9.377	1.437	-1.902	-0.022	-2.345
PPI-based	216	9.190	0.638	-1.069	-0.025	-2.119
TPI-based	309	9.047	1.266	-1.570	-0.062	-2.986
			UK			
CPI-based	429	0.425	1.118	-1.326	-0.029	-3.016
PPI-based	429	0.197	0.697	-0.863	-0.037	-3.429
TPI-based	348	1.368	1.254	-1.682	-0.050	-2.914

Note:

1. Critical values are -3.45 for 10%, -3.77 for 5%, and -4.31 for 1% if all the three variables (nominal exchange rate, domestic price index, foreign price index) do not have a time trend component. (Refer to Hamilton (1994) p766 Table B.9 Case II)

2. Critical values are -3.52 for 10%, -3.80 for 5%, and -4.36 for 1% if at least one of the three variables has a time trend component. (Refer to Hamilton (1994) p766 Table B.9 Case III)

Table C.12
Results of the Functional-Coefficient Cointegration Test

	$d=3$	$d=6$	$d=12$	$d=3$	$d=6$	$d=12$
	CANADA			GERMANY		
CPI-based	4.507	4.124	2.634	0.384	0.891	-1.313
PPI-based	-0.982	-0.724	-0.947	3.147	3.208	2.858
TPI-based	0.403	0.015	-0.082	3.298	2.338	2.369
	JAPAN			ITALY		
CPI-based	-1.193	-1.219	-1.555	0.810	0.793	0.677
PPI-based	1.325	1.044	0.821	-1.334	-1.044	-1.352
TPI-based	0.640	0.819	0.346	0.609	0.169	-0.661
	FRANCE			UK		
CPI-based	-0.026	0.387	0.321	-1.030	-1.142	-1.370
PPI-based	No Data Available			-1.840	-1.927	-1.911
TPI-based	0.484	1.340	1.295	-2.088	-1.825	-0.877

Note:

1. The values of t-statistic of the test suggested in Xiao (2009) are reported in the above table for 3 differently computed variables z_t: $d = 3$, $d = 6$, and $d = 12$.

2. Under the null of stationary residual (i.e. there is cointegration, possibly with varying coefficients), the test statistic asymptotically follows the standard normal distribution. Critical values are 1.645 for 10%, 1.960 for 5%, and 2.576 for 1% significance level.

Table C.13

Linear vs Functional-Coefficient Cointegration Test: A Comparison

	CC Residual		FC Residual		CC Residual		FC Residual	
	coef.	t-value	coef.	t-value	coef.	t-value	coef.	t-value
	CANADA				GERMANY			
CPI-based	-0.020	-2.826	-0.047	-3.187	-0.045	-2.377	-0.079	-2.156
PPI-based	-0.026	-2.369	-0.041	-2.395	-0.033	-1.834	-0.039	-1.931
TPI-based	-0.032	-2.119	-0.140	-5.618	-0.078	-3.031	-0.112	-4.238
	JAPAN				ITALY			
CPI-based	-0.022	-2.675	-0.033	-2.636	-0.022	-2.345	-0.050	-2.308
PPI-based	-0.020	-2.620	-0.068	-3.886	-0.025	-2.119	-0.090	-3.083
TPI-based	-1.789	-0.035	-0.064	-2.832	-0.062	-2.986	-0.168	-5.368
	FRANCE				UK			
CPI-based	-0.029	-2.422	-0.103	-3.584	-0.029	-3.016	-0.064	-3.930
PPI-based	No Data Available				-0.037	-3.429	-0.056	-3.526
TPI-based	-0.135	-2.963	-0.597	-4.827	-0.050	-2.914	-0.101	-3.690

Note: Refer to the note after Table C.11.