THREE ESSAYS ON TIME SERIES QUANTILE REGRESSION

by

YINI WANG

A thesis submitted to the

Department of Economics

in conformity with the requirements for

the degree of Doctor of Philosophy

Queen's University

Kingston, Ontario, Canada

July, 2012

Abstract

This dissertation considers quantile regression models with nonstationary or nearly nonstationary time series. The first chapter outlines the thesis and discusses its theoretical and empirical contributions. The second chapter studies inference in quantile regressions with cointegrated variables allowing for multiple structural changes. The unknown break dates and regression coefficients are estimated jointly and consistently. The conditional quantile estimator has a nonstandard limit distribution. A fully modified estimator is proposed to remove the second-order bias and nuisance parameters and the resulting limit distribution is mixed normal. A simulation study shows that the fully modified quantile estimator has good finite sample properties. The model is applied to stock index data from the emerging markets of China and several mature markets. Financial market integration is found in some quantiles of the Chinese stock indices.

The third chapter considers predictive quantile regression with a nearly integrated regressor. We derive nonstandard distributions for the quantile regression estimator and t-statistic in terms of functionals of diffusion processes. The critical values are found to depend on both the quantile of interest and the local-to-unity parameter, which is not consistently estimable. Based on these critical values, we propose a valid Bonferroni bounds test for quantile predictability with persistent regressors. We employ this new methodology to test the ability of many commonly employed and highly persistent regressors, such as

the dividend yield, earnings price ratio, and T-bill rate, to predict the median, shoulders, and tails of the stock return distribution.

Chapter Four proposes a cumulated sum (CUSUM) test for the null hypothesis of quantile cointegration. A fully modified quantile estimator is adopted for serial correlation and endogeneity corrections. The CUSUM statistic is composed of the partial sums of the residuals from the fully modified quantile regression. Under the null, the test statistic converges to a functional of Brownian motions. In the application to U.S. interest rates of different maturities, evidence in favor of the expectations hypothesis for the term structure is found in the central part of the distributions of the Treasury bill rate and financial commercial paper rate, but in the tails of the constant maturity rate distribution.

Co-Authorship

Chapter Three of this thesis is a co-authored work with Professor Alex Maynard at University of Guelph, Ontario, Canada and Professor Katsumi Shimotsu at the University of Tokyo, Tokyo, Japan.

Acknowledgments

I am grateful to my supervisors, Morten Nielsen, Katsumi Shimotsu, and Alex Maynard, for their patience, encouragement, and invaluable guidance regarding my research.

I would like to thank Charles Beach, Giuseppe Cavaliere, Ying Chen, Chuan Goh, Christian Gourieroux, Allan Gregory, Bruce Hansen, Michael Jansson, Soren Johansen, Frank Kleibergen, James MacKinnon, Lealand Morin, Benoit Perron, Maria Ponomareva, Zhongjun Qu, Robert Taylor, Tim Vogelsang, Chi Wan, Zhijie Xiao, and conference participants at the Canadian Econometrics Study Group, the Canadian Economic Association, and the Midwest Econometrics Study Group for useful comments and discussion regarding various chapters of this thesis. I am also thankful to Pierre Siklos for providing the stock index data used in Chapter Two. Thanks also to Tolga Cenesizoglu for providing the stock return and predictors data used in Chapter Three.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Least squares methods have been widely used to estimate the conditional mean of the regressand in linear models. However, conventional linear regression models do not provide information on other aspects of the distribution of the response variable. To account for nonlinear and asymmetric relationships between the dependent and independent variables over the entire distribution, this thesis studies the quantile regression estimation method in time series models. The analyses focus on traditional regression, which estimates the mean or quantiles of the response variable conditioning on observed covariates.

Least squares methods are confined to estimating the conditional mean function by minimizing the sum of squared residuals. The median is the 50% quantile, which can be estimated by minimizing the sum of absolute residuals. Koenker and Bassett (1978) extend this idea to estimating a general quantile that divides the population asymmetrically. In particular, the median and other conditional quantiles of the dependent variable y_t are modeled as linear functions of the regressor vector x_t. The quantile regression method is able to estimate any percentage point of the distribution of y_t conditional on the regressors. Generally, asymmetric weights are given to positive and negative residuals, and the minimization problem can be conveniently solved by linear programming methods. Linear conditional quantile models are convenient and valid local approximations to nonlinear quantile regression models which have much more complicated statistical properties[1]. More importantly, quantile regression is able to give a comprehensive picture of the entire conditional distribution of y_t without imposing global distributional assumptions on the errors.

In this thesis, the regressor vector x_t is nonstationary or nearly nonstationary. Both cases are empirically relevant. A lot of economic and financial time series are integrated or nearly integrated. For example, stock prices and interest rates are unit root processes. Typical predictor variables used to forecast stock returns, such as dividend price ratio and earnings price ratio, are persistent. For financial variables, the tails of the distribution are important for portfolio decision. For instance, value at risk (VaR) is a standard measure of market risk, which is defined as the value that a portfolio will lose with a given probability, such as 5% and 1%, over a given time horizon. Engle and Manganelli (2004) interpret VaR as the quantile of future portfolio values conditional on current information. Using quantile regression they propose the conditional autoregressive value at risk (CAViaR) model to estimate VaR. As in the example of estimating VaR, there are several advantages of using quantile regression especially in financial time series applications. First, it avoids strict assumptions such as normality or i.i.d. returns. Second, it models nonlinearity over different quantiles without setting up a complex nonlinear parametric framework. Third, compared to least squares methods, the quantile regression method is more robust to non-Gaussian, leptokurtic, and heavy-tailed distributions, which are common for financial variables such as short term interest rates. Moreover, the conditional quantile model is superior when

[1] Linear quantile regression models may violate the monotonicity requirement in the fitted quantile functions, which is known as the quantile "crossing" problem. However, the linear quantile models can be interpreted as useful local approximations over a region of interest (Koenker and Xiao 2006).

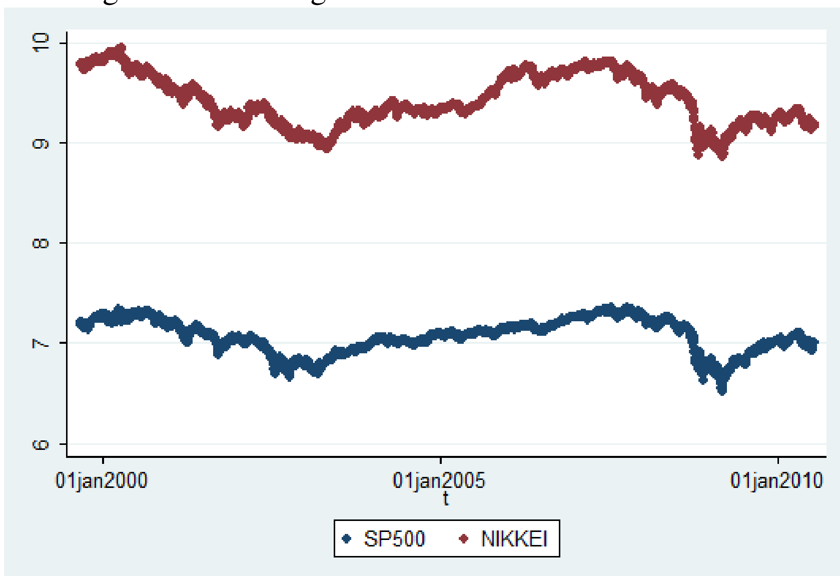
there is strong effect exerted on the least squares fit by the outlying observations (Koenker and Xiao 2004). It is often found that financial data such as stock indices from emerging markets are contaminated by large outliers.

There has been a lot of discussion on international integration among both developed and developing financial markets. In these markets, variables such as stock prices are nonstationary and are often characterized as integrated processes of order one, denoted I(1). If financial integration exists, there should be co-movement among stock indices from different countries. To investigate this world integration, much research has been done on cointegration analysis for stock markets.

Cointegration measures long run equilibrium among nonstationary economic variables. Cointegrated variables should not diverge, at least not in the long run. To put it more precisely, certain linear combination of such variables has a stationary distribution. For example, the expectations hypothesis for the term structure suggests cointegration of long and short term interest rates (Engle and Granger 1987, Stock and Watson 1988, Hansen 1992). A forward-looking solution to the rational expectations model suggests that real stock price and market fundamentals such as dividends should be cointegrated (Campbell and Shiller 1987, 1988a, 1988b). Likewise, there should be a long run equilibrium relationship among exchange rate and its fundamentals (Papell 1997).

In the case of financial market linkage, cointegration among stock indices from different economies can be seen as a sign of equity market integration. For instance, the stock indices from the advanced stock markets of U.S. and Japan tend to move together and are cointegrated. Figure 1.1 shows the cointegration relationship between the daily S&P 500 (SP500) index and NIKKEI. The logarithms of the S&P 500 index and NIKKEI are plotted from August 31, 1999 to July 16, 2010. The upper curve denotes the NIKKEI and the lower curve denotes the S&P 500. The two series are clearly cointegrated.

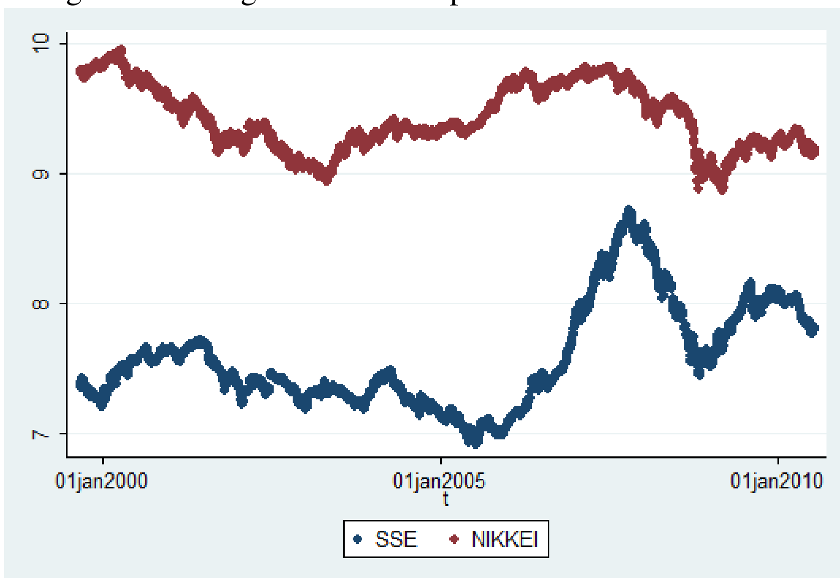
Figure 1.1: Cointegration between SP500 and NIKKEI



Notes: Logarithms of S&P 500 (SP500) and NIKKEI from 08/31/1999 to 07/16/2010.

However, between an emerging market, such as China, and a mature market, the relationship may be quite different. In Figure 1.2, the upper curve is again for the NIKKEI from the mature market of Japan and the lower curve plots the log of the Shanghai composite index (SSE) from the emerging stock market of China. The emerging market may not be well integrated with the rest of the world. Also, the cointegration relationship, if it exists, may have been affected by structural changes. It is possible that the emerging market is only partially cointegrated with the mature market. In the case of the NIKKEI and SSE, although there may not be conventional cointegration in the mean, long run equilibrium relationships may be found in other parts of the distribution of the Shanghai composite index conditional on NIKKEI.

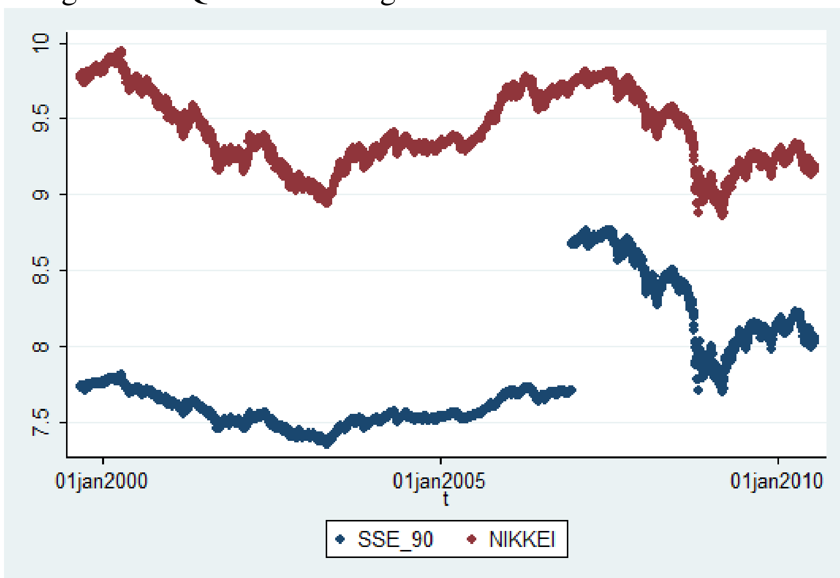
Figure 1.2: Long run relationship between SSE and NIKKEI



Notes: Logarithms of Shanghai composite index (SSE) and NIKKEI from 08/31/1999 to 07/16/2010.

By allowing for a structural change in the cointegrating regression and estimating the conditional quantiles of the distribution of the Shanghai composite index, one can find evidence of long run equilibrium in certain quantiles, which can not be found using least squares methods. In Chapter Two of this thesis, it is found that the high episode of the distribution of the Chinese stock index is cointegrated with some mature stock markets. Hence, locations in the stock index distribution other than the mean actually have an impact. Furthermore, there is a shift in the long run relationship, so the regression coefficients between two stock indices are not stable. In Figure 1.3, the lower curve depicts the 90% quantile of the Shanghai composite index conditional on NIKKEI. The upper tail of the SSE index is cointegrated with the NIKKEI and there is a structural break at the end of 2006.

Figure 1.3: Quantile cointegration between SSE and NIKKEI



Notes: SSE_90 denotes the 90% quantile of the log of SSE conditional on the log of NIKKEI.

This thesis is partially built on the arguments put forth by Xiao (2009), who proposes the quantile cointegrating regression method. The second chapter develops a new asymptotic theory for quantile regression with cointegrated I(1) variables allowing for multiple structural changes of unknown timing. It is important to incorporate structural changes in cointegration analysis. If we ignore structural breaks, tests for cointegration may be biased (Gregory 1994, Gregory and Hansen 1996, Gregory et al. 1996). The framework of the model proposed in Chapter Two is that the regressor x_t has a unit root, and the quantile dependent regression coefficients are subject to structural changes. This model is able to capture the impact of the mature stock market on the dispersion and tail behavior of the emerging market, in cases where the regression coefficients are not time-invariant. The theoretical contributions in this chapter include the following: First, taking the number of breaks as pre-specified, I estimate the unknown break dates and cointegrating vectors jointly. Second, I derive the nonstandard limit distribution of the conditional quantile estimator. Since there exists serial cross correlation and long run endogeneity, the quantile estimator for the cointegrating coefficients is biased and depends on nuisance parameters. To develop standard inference, I propose a Phillips-Hansen type fully modified quantile estimator for serial correlation and endogeneity corrections. The limit distribution of this new estimator is mixed normal. Simulation results show that the fully modified estimator significantly reduces the bias and root mean square error and the confidence intervals have good coverage properties in finite samples. The model is applied to stock index data from the emerging market of China, comparing it to several mature markets to examine financial market integration. When considering one regime shift in the model, I find asymmetric long run equilibria in some quantiles of the conditional distributions of the Chinese stock indices.

The third chapter of this thesis studies predictive quantile regression of a dependent variable y_t on a lagged regressor x_{t-1} when the predictive regressor has a near unit root. This model has important empirical applications, such as testing the predictability of the stock return distribution. There has been an extensive literature on predictive tests for mean returns, which debates whether predictors such as dividend yield can forecast stock returns. Earlier work adopted standard t-tests and found strong evidence of predictability of stock returns by dividend yield, dividend price, and earnings price ratios (Shiller 1984, Campbell and Shiller 1988a, 1988b, Fama and French 1988, Hodrick 1992). However, some regressors in the predictive regressions are stochastic and highly persistent, so that ordinary t-tests of orthogonality are invalid and reject too frequently. Consequently, the predictability of the returns with these nearly nonstationary predictors may be overstated (Shiller 1984, Mankiw and Shapiro 1986, Stambaugh 1986, 1999, Cavanagh et al. 1995).

Conventional predictive tests of stock returns are not valid for two main reasons. Firstly, the predictor variables, such as dividend yield, dividend price, and earnings price ratios, are strongly autocorrelated. Secondly, although pre-determined, typically the predictor is not strictly exogenous, since its innovation is often highly correlated with the error term in the predictive regression. For example, with financial data such as dividend price ratios and returns, it is reasonable to expect strong correlation between the regressor's innovation and the regression disturbance. This tends to inflate the t-statistic, resulting in over-rejection. In linear predictive regressions, it is well-known that when the regressor is not strictly exogenous and has a largest autoregressive root close to one, the limiting distribution of the t-statistic will be nonstandard. In this case, the t-statistic is too large and tests using the standard normal critical values will over-reject the null hypothesis of nonpredictability.

A great deal of attention has been devoted to overcoming such size distortions in linear predictive regressions. For example, some papers provide alternative semi- or nonparametric tests (Campbell and Dufour 1995, 1997, Wright 2000, Lanne 2002, Maynard and Shimotsu 2009). Some authors perform finite sample corrections in tightly parameterized models (Stambaugh 1999, Lewellen 2004, Amihud and Hurvich 2004, Amihud et al. 2004, Nelson and Kim 1993, Goetzmann and Jorion 1993, Wolf 2000). Others consider asymptotic corrections in a local-to-unity framework (Cavanagh et al. 1995, Campbell and Yogo 2006, Jansson and Moreira 2006).

In contrast to this large literature devoted to testing the predictability of the mean of stock returns, there are few works studying the predictability of the entire return distribution using similar lagged predictors. Recent empirical work by Cenesizoglu and Timmermann (2008) employs the quantile regression method and finds a number of pre-determined predictors having asymmetric effects on various quantiles of the return distribution. Quantile regression is useful for analyzing how different positions in the return distribution, such as the median, shoulders, and tails, respond to changes in a predictor. However, there is no theoretical work to date that establishes valid econometric inference methods in quantile predictive regressions with persistent regressors. In Chapter Three, we develop inference in short-horizon predictive quantile regressions with nearly integrated regressors which follow an $AR(p)$ process in general, with $p \geq 1$. We first derive the limit distribution of the quantile regression coefficients by generalizing the results of Xiao (2009) to the local-to-unity setting of Chan and Wei (1987), Chan (1988), Phillips (1987), Phillips (1988a, 1988b), and Nabeya and Sorensen (1994). We then provide a test of quantile predictability based on an asymptotically valid Bonferroni bounds methodology in the spirit of Cavanagh et al. (1995). This Bonferroni technique is applied to testing the predictability of the stock return distribution using various pre-determined predictive regressors. In this empirical application, we consider 16 commonly used predictors, such as dividend yield, dividend price ratio, earnings price ratio, and stock variance. We test whether these variables can predict the median, shoulders, or tails of the stock return distribution. It is found that most of the regressors affect the lower, central, and upper quantiles of the return distribution very differently so that they have asymmetric effects on the return distribution.

The fourth chapter is an extension to Chapter Two and Xiao and Phillips (2002). This chapter also considers the quantile cointegration methodology. Based on the linear cointegrating regression model developed by Engle and Granger (1987), there is a substantial literature on cointegration tests. Among various testing techniques, the unit root type cointegration test is widely used. This residual based procedure performs an (augmented) Dickey-Fuller test on the residuals from the cointegrating regression. Hence, the null hypothesis of interest is no cointegration. If the Dickey-Fuller test rejects that the residual process has a unit root, then we reject the null of no cointegration. However, since cointegration relationship is of particular interest to economists, it is reasonable to develop testing procedures for the null hypothesis of cointegration. Many authors consider such residual based test for the linear case (Park et al. 1988, Park 1990, Shin 1994, Xiao and Phillips 2002). This chapter extends the robust cumulated sum (CUSUM) test proposed by Xiao and Phillips (2002) to quantile regression. The CUSUM tests have been used for testing structural changes. For example, Hao and Inder (1996) test the null hypothesis of parameter constancy in cointegrated models. Under the null, their test statistic has the same asymptotic distribution as that in the model of Xiao and Phillips (2002) under the null of cointegration.

Chapter Four adopts this CUSUM type test for the null hypothesis of cointegration in a quantile regression model. Similar to Chapter Two, the fully modified quantile estimator for the regression coefficients is used for serial correlation and endogeneity corrections. The CUSUM test statistic is composed of the cumulative sums of the residuals from the fully modified quantile regression. Under the null of cointegration, the CUSUM statistic reflects the stationary equilibrium error from the cointegrating regression. The asymptotic representation of the test statistic is a functional of Brownian motions. Under the alternative of no cointegration, the CUSUM test statistic diverges to infinity.

A rich literature discusses the expectations theory of the term structure of interest rates, such as Campbell and Shiller (1987) among others. Many papers examine different yield series on U.S. interest rate, such as the federal funds rate, Treasury bill rate, and commercial paper rate. The empirical results for the expectations hypothesis of term structure is mixed. Chapter Four applies the residual based quantile cointegration test to three sets of daily U.S. data, including Treasury bill rate, constant maturity rate, and financial commercial paper rate. The quantile version of the CUSUM test indicates evidence of cointegration in the central part of the conditional distributions of the Treasury bill rate and financial commercial paper rate, but in the tails of the constant maturity rate distribution.

The remainder of this thesis is organized as follows: Chapter Two studies inference in quantile cointegrating regressions with unknown structural changes. Chapter Three develops inference in predictive quantile regressions with nearly integrated regressors. Chapter Four presents a CUSUM test for the null hypothesis of quantile cointegration. Chapter Five concludes the thesis. All proofs are provided in the Appendix.

Chapter 2

Quantile Cointegration with Structural Changes

2.1 Introduction

Cointegration measures long term equilibrium relationships among nonstationary economic variables. Economic theory suggests that some nonstationary variables should not drift too far apart from each other, at least not in the long run, so that the equilibrium error of certain linear combinations of these variables is stationary. There are a number of possible examples of cointegrated variables, such as stock prices and market fundamentals (Campbell and Shiller 1987, 1988a, 1988b), exchange rate and its fundamentals (Papell 1997), short and long term interest rates (Engle and Granger 1987, Stock and Watson 1988, Hansen 1992), spot and future prices (Leuthold et al. 1989, Chance 1991), wages and prices (Golinelli and Orsi 1994, 2000), aggregate consumption and income (Campbell 1987, Engle and Granger 1987), imports and exports (Arize 2002), money supply and prices, money demand, income, and interest rates (Johansen and Juselius 1990). Furthermore, due to international integration, financial variables such as stock indices or foreign exchange rates (Baillie and Bollerslev 1989, 1994, Aggarwal and Mougoue 1993, Diebold et al. 1994, Lopez 2005) from different countries tend to move together. Since cointegration analysis examines equilibrium in the long run, it normally involves long spans of data which are more likely to be affected by structural changes (Kejriwal and Perron 2008). Without considering the presence of structural breaks, tests for cointegration may be biased (Gregory 1994, Gregory and Hansen 1996, Gregory et al. 1996). Thus, it is important to incorporate structural changes in cointegrated models.

This chapter provides estimation and asymptotic theory in the quantile cointegrating regression context with structural changes. In particular, the possibility of multiple breaks is considered. To make the analysis more general, it is desirable to allow for unknown break dates, since in practice it is unlikely to know the break points a priori. For example, the source of the break may be unknown, and a regime shift may happen after an unknown time lag following some event such as a policy announcement[2].

There is an extensive literature on structural breaks. Most previous work is based on linear models. These models have included stationary regressors, trending variables, variables that exhibit unit roots, cointegrated variables, long memory processes, and so on. Plenty of work has been done to derive estimators, asymptotic distributions, tests for structural changes, and procedures to determine the number of breaks. Perron (2006) provides a comprehensive review of the structural break literature based on linear models. Among recent literature in the context of stationarity, Bai (1996) proposes tests for parameter constancy in weighted empirical distribution functions of estimated residuals which are asymptotically distribution free. Bai and Perron (1998, 2003) estimate and test multiple structural changes

[2] Qu (2008) provides some examples in the stationary context.

in linear models allowing for general forms of serial correlation and heteroskedasticity in the errors, lagged dependent variables, trending regressors, different distributions for the errors and the regressors across segments, as well as partial structural change where not all parameters are subject to regime shifts. There is also much work considering structural change models with cointegrated variables. For example, Hansen (1992), Gregory and Hansen (1996), and Gregory et al. (1996) consider Lagrange multiplier (LM) tests and residual based augmented Dickey-Fuller (ADF) tests for a single structural change in the cointegrating vector. Kejriwal and Perron (2008, 2010) discuss estimation and testing for multiple structural changes that are allowed to occur in the intercept, the cointegrating coefficients, the parameters of the stationary regressors, or any combination of these.

Most existing papers are devoted to estimating and testing structural changes in the conditional mean, while few authors have considered the case of conditional quantiles. Whereas least squares methods are confined to estimating the conditional mean, quantile regression methods, developed by Koenker and Bassett (1978), are able to provide a comprehensive picture of the entire distribution without imposing global assumptions on the distribution function. For example, the conditioning variable may only affect the tails of the distribution of the dependent variable, which leads to changes in the conditional dispersion, leaving the mean unchanged. If so, the conditional mean is no longer informative. Thus, it is useful to examine how structural changes can affect different parts of the conditional distribution as represented by conditional quantiles. The literature on quantile regression has developed rapidly. For example, there is quantile autoregression (Koenker and Xiao 2006), unit root quantile autoregression (Koenker and Xiao 2004), and quantile cointegrating regression (Xiao 2009), among others. In particular, quantile cointegration is a recently developed concept for characterizing quantile varying long run equilibrium relationship between integrated economic variables. Conventional cointegration model focuses on the long run relationship between the means of economic variables, and the cointegrating vector is constant. However, in some cases, it is useful to examine the long run relationship in certain quantiles. For example, the relationship between the Chinese stock market index and the U.S. stock index, such as S&P 500, may be affected by shocks in the Chinese market in each period. In this case, conventional cointegration relationship no longer holds, because the regression coefficients between the two stock indices change over time. However, even in such case, the S&P 500 index may have impacts on the dispersion and tail behavior of the Chinese stock index, and quantile cointegration can properly capture such relationships.

Estimating and testing for structural breaks in the conditional quantiles is a relatively new area. Recent papers have assumed stationarity. Bai (1995) develops the asymptotic theory for least absolute deviation (LAD) estimation of a shift in linear regressions, which is a special case of quantile regression estimation. Also, Bai (1998) extends the analysis to LAD estimation with multiple change points occurring at unknown times. More recently, Su and Xiao (2008) propose a sup-Wald test for parameter stability in a two-parameter quantile regression model with an unknown break point. Qu (2008) proposes two types of tests, namely a subgradient test and a Wald type test, for multiple structural changes at unknown dates in one or multiple conditional quantiles. Oka and Qu (2011) consider estimation in the same stationary context and discuss inference methods in a time series model as well as a model with repeated cross sections.

However, there has been no theoretical work to date that studies structural changes in regression quantiles in cointegrated models. In an attempt to fill this gap, this chapter develops inference in quantile regressions with cointegrated I(1) variables allowing for multiple structural changes occurring at unknown dates. The quantile dependent break dates and cointegrating vectors are estimated jointly by minimizing the absolute deviation loss function over all permissible partitions of the sample when the number of breaks is pre-specified. Consistency of the quantile estimator and the corresponding rates of convergence are derived under the assumption of shrinking shift. In the presence of serial cross correlation between the regression disturbance and the innovation of the I(1) regressors, the conditional quantile estimator for the cointegrating coefficient is second-order biased. Although it does not affect the consistency of the coefficient estimates, the bias affects the centering of the limit distribution and can lead to substantial finite sample bias (Phillips and Hansen 1990). Moreover, due to long run endogeneity, the limit distribution of the cointegrating coefficient estimates is nonstandard and depends on nuisance parameters. Indeed, the simulation study in this chapter shows that the original quantile estimator has large bias and root mean square error in finite samples. Confidence intervals calculated based on this estimator also give misleading results. Thus, a fully modified quantile estimator is required for serial correlation and endogeneity corrections. The resulting limit distribution is mixed normal so that it can provide a useful inference procedure. As shown in the finite sample experiments, the bias and root mean square error are largely reduced after the corrections, and the confidence intervals have adequate coverage rates.

The model has potential for applications to macroeconomic data and financial time series. The empirical example in this chapter involves quantile cointegration between stock indices from different economies, which can be seen as a sign of equity market integration. There is a large literature on international integration among both developed and developing countries. Much research has been done on cointegration analysis for financial markets. In these markets, variables like stock prices are often characterized as integrated processes of order one. If international integration exists, there should be co-movement among stock indices from different countries with financial market linkage. Many recent empirical papers study the equilibrium relationships between emerging stock markets and their mature counterparts. Most papers use linear models. For example, Chen et al. (2002) study the long run relationships between four Latin American stock markets and the U.S. stock market. Voronkova (2004) estimates and tests the cointegration relationships between Central European stock markets and the mature markets of Europe and the United States. Manning (2002) and Azad (2009) have done similar analysis for Asian markets. Some empirical studies adopt the quantile regression method but without formally incorporating structural changes in the model, such as Burdekin and Siklos (2011) who study the quantile cointegration relationships among the stock markets in the Asia Pacific region. In this chapter, the quantile cointegration model is used to study international integration of the emerging stock markets in China with several mature markets. It is important to incorporate structural changes in the cointegration analysis, since assuming time invariant coefficients may conceal the evidence of long run equilibrium relationships. One might mistakenly conclude that there is no cointegration, although it exists, unless the presence of structural changes is taken into account. Allowing for a regime shift in the regression coefficients, the conditional quantile estimation reveals asymmetry in the long run equilibrium relationships between the financial markets, which cannot be found using least squares estimation. Hence, it provides stronger evidence of financial market integration. Furthermore, the estimated break dates accurately capture the boom in the Chinese stock markets during 2006-2007 and closely coincide with policy changes in China.

The remainder of Chapter Two is organized as follows: In Section 2.2, the framework of the problem is established. The asymptotic theory for the conditional quantile estimator is developed and the fully modified estimator is proposed to remove the second-order bias and nuisance parameters. In Section 2.3, the estimation method and asymptotic theory are evaluated by Monte Carlo experiments. In Section 2.4, an empirical application is discussed. Section 2.5 concludes this chapter.

2.2 Theory

2.2.1 The model

Let x_t be a k-dimensional vector of integrated regressors. The elements of x_t are not cointegrated. The DGP of x_t is as follows:

$$x_t = x_{t-1} + v_t. \tag{2.1}$$

In the linear case, the cointegration model with structural changes is the following:

$$y_t = \begin{cases} \alpha_1^0 + x_t'\beta_1^0 + u_t & t = 1, ..., T_1^0 \\ \alpha_2^0 + x_t'\beta_2^0 + u_t & t = T_1^0 + 1, ..., T_2^0 \\ \vdots & \vdots \\ \alpha_{m+1}^0 + x_t'\beta_{m+1}^0 + u_t & t = T_m^0 + 1, ..., T, \end{cases} \tag{2.2}$$

where α_j^0 and β_j^0 $(j = 1, ..., m+1)$ denote the true parameters and T is the sample size. This model allows for m structural changes occurring at the true break dates $(T_1^0, ..., T_m^0)$. When a break occurs, the entire cointegrating vector is subject to change. Define the information set $\mathcal{F}_t = \sigma\{v_t, u_{t-1}, v_{t-1}, u_{t-2}, v_{t-2}, ...\}$, then $(x_t, y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ...) \in \mathcal{F}_t$. Denote $T_0^0 = 0$ and $T_{m+1}^0 = T$. The conditional expectation function of y_t is $E(y_t|\mathcal{F}_t) = \alpha_j^0 + x_t'\beta_j^0$ for $t = T_{j-1}^0 + 1, ..., T_j^0$ and $j = 1, ..., m+1$. Assume $(u_t, v_t')'$ is mean zero stationary with covariance matrix conditional on $\sigma\{u_{t-1}, v_{t-1}, u_{t-2}, v_{t-2}, ...\}$ given by $\begin{bmatrix} \sigma_{uu} & \sigma_{uv} \\ \sigma_{vu} & \sigma_{vv} \end{bmatrix}$.

Least squares methods have been widely used to estimate the conditional mean for linear models. However, the linear model does not provide information on other aspects of the distribution of y_t without further assumptions on the error term. In order to capture the cointegration relationship over the entire distribution, the analysis is generalized to the conditional quantiles of y_t. For example, the median is the 50% quantile that divides the population into two equal portions. In a general case, the random variable at the τth quantile of a distribution is larger than the τ proportion of the population but smaller than the $(1-\tau)$ proportion with $\tau \in (0,1)$.

Let $F(\cdot)$ and $F_t(\cdot) = Pr(u_t < \cdot|\mathcal{F}_t)$ be the unconditional and conditional cumulative distribution functions of u_t, and define the τth unconditional and conditional quantiles of u_t by $Q_{u_t} = F^{-1}(\tau)$ and $Q_{u_t}(\tau|\mathcal{F}_t) = F_t^{-1}(\tau)$, respectively. With the additional assumption that $F_t(\cdot) = F(\cdot)$ for all $1 \leq t \leq T$, the conditional quantile of y_t is $Q_{y_t}(\tau|\mathcal{F}_t) = \alpha_j^0 + F^{-1}(\tau) + x_t'\beta_j^0$ for $t = T_{j-1}^0 + 1, ..., T_j^0$ and $j = 1, ..., m+1$, where $\tau \in (0,1)$ indexes the quantile level. Under this specification, the cointegrating coefficient in each regime β_j^0 is equal across quantiles[3]. This conventional cointegration model is more restrictive than required for the quantile regression estimation method. Within the jth regime, since all quantiles share the same slope coefficient β_j^0, the model only allows for parallel shifts of the conditional quantiles of y_t. It implies that x_t shifts different positions in the conditional distribution of y_t all in the same direction and by exactly the same amount.

However, the cointegrating coefficient can vary across quantiles such that $\beta_j^0(\tau)$ depends on the innovation process u_t (Xiao 2009). The conditional quantile of the error term can be modeled as $Q_{u_t}(\tau|\mathcal{F}_t) = \alpha_j^0(\tau) - \alpha_j^0 + x_t'(\beta_j^0(\tau) - \beta_j^0)$. Then, the quantile function

[3]The cointegration specification is the same as that in section 2 of Xiao (2009).

of y_t conditional on $\mathcal{F}_t$ is $Q_{y_t}(\tau|\mathcal{F}_t) = \alpha_j^0(\tau) + x_t'\beta_j^0(\tau)$. Define $u_t(\tau) = u_t - F_t^{-1}(\tau) = y_t - \alpha_j^0(\tau) - x_t'\beta_j^0(\tau)$ so that $Q_{u_t(\tau)}(\tau|\mathcal{F}_t) = 0$. Consequently, we have a more flexible quantile regression model such that

$$y_t = \alpha_j^0(\tau) + x_t'\beta_j^0(\tau) + u_t(\tau) = z_t'\theta_j^0(\tau) + u_t(\tau), \quad t = T_{j-1}^0(\tau) + 1, ..., T_j^0(\tau), \quad (2.3)$$

where $z_t = (1, x_t')'$ and $\theta_j^0(\tau) = (\alpha_j^0(\tau), \beta_j^{0\prime}(\tau))'$ for $j = 1, ..., m+1$. Also, the break dates, $T_j^0(\tau)$ for $j = 1, ..., m$, are not necessarily the same for all quantiles. For ease of notation, the argument τ for $T_j^0(\tau)$ will be dropped from now on so that the break dates are signified as T_j^0 for $j = 1, ..., m$. This model allows for location shifts as well as changes in the scale and shape of the conditional distribution of y_t. The quantile dependent cointegrating vector $\theta_j^0(\tau)$ can capture the nonlinearity in the long run equilibrium relationship. Nonetheless, for each τ the conditional quantile of y_t is formulated as a linear function of $\theta_j^0(\tau)$ so that the unknown regression coefficients can be estimated using linear programming methods (Koenker and Hallock 2001).

2.2.2 Estimation method

Consider the following objective function with different weights on positive and negative residuals:

$$S_T(\tau, \theta(\tau), T^b) = \sum_{j=0}^{m} \sum_{t=T_j+1}^{T_{j+1}} \rho_\tau(y_t - z_t'\theta_{j+1}(\tau)), \quad (2.4)$$

where $T^b = (T_1, ..., T_m)$ are candidates of the break dates and $\theta(\tau) = (\theta_1(\tau)', ..., \theta_{m+1}(\tau)')'$. Here, $\rho_\tau(\cdot)$ is the asymmetric absolute deviation loss function defined by $\rho_\tau(u) = u(\tau - I(u < 0))$ also known as the check function (Koenker and Bassett 1978). Define $\psi_\tau(u) = \tau - I(u < 0)$, which puts asymmetric weights on the residuals. For the median, $\tau = 0.5$, $\rho_\tau(u) = |u|/2$ is used for the Laplace's median regression function. In this special case, the quantile regression model delivers the least absolute deviation (LAD) estimation. In general, unlike least squares method that minimizes the sum of squared residuals, the conditional quantile estimation minimizes the sum of asymmetrically weighted absolute residuals.

The unknown break dates and quantile regression coefficients are estimated jointly by minimizing the objective function over all permissible break dates. For a single quantile, we have

$$(\hat{\theta}(\tau), \hat{T}^b) = \arg \min_{\theta(\tau), T^b} S_T(\tau, \theta(\tau), T^b), \tag{2.5}$$

where $\hat{T}^b = (\hat{T}_1, ..., \hat{T}_m)$ and $\hat{\theta}(\tau) = (\hat{\theta}_1(\tau)', ..., \hat{\theta}_{m+1}(\tau)')'$ with $\hat{\theta}_j(\tau) = (\hat{\alpha}_j(\tau), \hat{\beta}_j'(\tau))'$ for $j = 1, ..., m+1$. With the number of breaks pre-specified, we take a two-step estimation procedure. First, for a given partition, T^b, we estimate the cointegrating vectors by minimizing $S_T(\tau, \theta(\tau), T^b)$ with respect to $\theta(\tau)$. Second, we search over all possible partitions to find the break dates that give the global minimum.

2.2.3 Assumptions

Let "$\rightarrow^p$" denote convergence in probability, while "$\Rightarrow$" signifies weak convergence of the associated probability measures in the space $D[0,1]$ under the Skorohod metric. To develop the limit theory, the following assumptions are imposed.

Assumption 1 $\xi_t = (\psi_\tau(u_t(\tau)), v_t')'$ *is a zero mean, stationary* $(k+1)$*-dimensional vector satisfying*

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \xi_t \Rightarrow \begin{bmatrix} B_\psi(r) \\ B_v(r) \end{bmatrix} = BM(0, \Omega), \tag{2.6}$$

where $(B_\psi(r), B_v'(r))'$ *is a vector of Brownian motions with covariance matrix*

$$\Omega = \begin{bmatrix} \omega_\psi^2 & \Omega_{\psi v} \\ \Omega_{v\psi} & \Omega_{vv} \end{bmatrix}.$$

Note that the variance of $\psi_\tau(u_t(\tau))$ is $\omega_\psi^2 = \tau(1-\tau)$ and $\Omega_{\psi v} = \Omega'_{v\psi}$. As in Xiao (2009), ξ_t is assumed to satisfy a functional central limit theorem. In general, the long run covariance matrix $\Omega = \lim_{T\to\infty} T^{-1}E(\sum_{t=1}^T \xi_t \sum_{t=1}^T \xi_t') = \Sigma + \Lambda + \Lambda'$, where $\Sigma = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T E(\xi_t\xi_t')$ and $\Lambda = \lim_{T\to\infty} T^{-1}\sum_{i=1}^{T-1}\sum_{t=1}^{T-i} E(\xi_t\xi_{t+i}')$. Also, define $\Delta = \Sigma + \Lambda$.

Assumption 2 *The conditional distribution function,* $F_t(u) = Pr(u_t < u|\mathcal{F}_t)$*, is absolutely continuous and has a continuous density function* $f_t(u)$ *such that* $0 < f_t(F_t^{-1}(\tau)) < \infty$ *for* $t = 1, ..., T$.

Assumption 3 *The density function* $f_t(\upsilon_T)$ *is uniformly integrable for any sequence* $\upsilon_T \to F_t^{-1}(\tau)$ *and* $E[f_t^\gamma(F_t^{-1}(\tau))] < \infty$ *for some* $\gamma > 1$*, for* $t = 1, ..., T$.

Assumptions 2 and 3 are two standard technical assumptions in the quantile regression literature such as Xiao (2009), Qu (2008), and Oka and Qu (2011)[4]. They require that the conditional density function is uniformly continuous, bounded, and integrable in some neighborhood of the τth quantile.

Assumption 4 $T_j^0 = [\lambda_j^0 T]$ *for* $j = 1, ..., m$ *and* $0 < \lambda_1^0 < ... < \lambda_m^0 < 1$.

This is a standard assumption in the structural break literature. The notation is useful when developing the weak convergence conditions of the coefficient and break date estimates.

[4]Assumption 2 corresponds to Assumption B in Xiao (2009) and Assumption 2 in Oka and Qu (2011). Assumption 3 corresponds to Assumption C in Xiao (2009).

Furthermore, the break dates, $(T_1^0, ..., T_m^0)$, are asymptotically distinct. Each regime contains a positive fraction of the whole sample even in the limit so that when $T \to \infty$ each regime will have infinitely many observations. In addition, define $\lambda_0^0 = 0$ and $\lambda_{m+1}^0 = 1$. The break fractions λ_j^0 are bounded away from the end points (Bai 1997). As in equation (2.3), the break fractions are quantile dependent.

Assumption 5 *Define $\tilde{D}_T = diag(1, T^{\frac{1}{2}} I_k)$. There exists an $l_0 > 0$ such that for all $l > l_0$ the minimum eigenvalues of*

$$A_{jl} = \frac{1}{l}\tilde{D}_T^{-1} \sum_{t=T_j^0+1}^{T_j^0+l} z_t z_t' \tilde{D}_T^{-1} \text{ and of } A_{jl}^* = \frac{1}{l}\tilde{D}_T^{-1} \sum_{t=T_j^0-l}^{T_j^0} z_t z_t' \tilde{D}_T^{-1}$$

are bounded away from zero for $j = 1, ..., m+1$.

Assumption 5 is another standard assumption in the structural change literature[5]. The matrices A_{jl} and A_{jl}^* are assumed to be invertible. This condition rules out local collinearity. It requires that there are enough observations around the true break points so that the break dates can be identified.

Assumption 6 *For $j = 1, ..., m$, $\delta_{T,j}(\tau) = \theta_{j+1}^0(\tau) - \theta_j^0(\tau) = \tilde{D}_T^{-1}\delta_j(\tau)\nu_T$ for some $\|\delta_j(\tau)\| > 0$, where $\delta_j(\tau)$ is a vector independent of T. The scalar $\nu_T > 0$ satisfies $\nu_T \to 0$ and $T^{\frac{1}{2}}\nu_T \to \infty$ as $T \to \infty$.*

This assumption imposes that the differences between the parameters in adjacent regimes shrink to zero asymptotically[6]. In particular, the differences associated with the integrated

[5] Assumption 5 is analogous to Assumption A3 from Bai (1997) and Assumption A2 from Bai and Perron (1998).

[6] This assumption is the same as Assumption 7 from Kejriwal and Perron (2008), but weaker than Assumption A6 from Bai and Perron (1998). As in Bai (1997), shrinking shift is assumed instead of fixed shift so that the limiting distribution of the break point is invariant to the exact distribution of the regressors and regression disturbance.

variables shrink faster than that of the intercept term. The positive scalar ν_T goes to zero as $T \to \infty$ so that the magnitudes of the breaks are small. However, the condition that $\sqrt{T}\nu_T$ diverges asymptotically ensures that the break sizes are not too small to be identified. Consistency of the quantile estimator can be derived under this assumption of shrinking shift. It implies that breaks of larger magnitude can also be consistently estimated, because it is easier to identify larger breaks than to identify smaller breaks. Note that ν_T is closely related to the convergence rate of the break point estimates, as shown in the next subsection. Moreover, the rate of convergence derived based on the assumption of shrinking shift holds for more general models with lagged dependent variables and trending regressors (Bai 1997).

2.2.4 Theoretical results

If the above assumptions hold, the estimates of the break fractions and regression coefficients will converge to their true values. However, the rates of convergence for the coefficients are very different from those for the break fractions. Denote $D_T = diag(T^{\frac{1}{2}}, TI_k)$. The presence of D_T is due to the different convergence rates for the intercept term and coefficients associated with the integrated variables. For simplicity, in this section the conditional distribution of the regression disturbance u_t is assumed to be time invariant such that $F_t(\cdot) = F(\cdot)$ for all $1 \leq t \leq T$. This additional assumption is not necessary for obtaining consistency with the same rates of convergence or deriving the asymptotic distributions. A less restrictive assumption is discussed in section A.1 of the appendix. The limit distribution of the coefficient estimates is given by the following representation, which is the same as that obtained when the break dates are known.

Theorem 1 *Under Assumptions 1-5, for $j = 1, ..., m+1$, $\hat{\theta}_j(\tau)$ is a consistent estimator* *of* $\theta_j^0(\tau)$ *and*

$$D_T(\hat{\theta}_j(\tau)-\theta_j^0(\tau)) \Rightarrow \left[f(F^{-1}(\tau))\int_{\lambda_{j-1}^0}^{\lambda_j^0}\bar{B}_v\bar{B}_v'\right]^{-1}\left[\int_{\lambda_{j-1}^0}^{\lambda_j^0}\bar{B}_v dB_\psi+(\lambda_j^0-\lambda_{j-1}^0)\bar{\Delta}_{v\psi}\right], \tag{2.7}$$

where $\bar{B}_v=(1,B_v')'$, $\bar{\Delta}_{v\psi}=(0,\Delta_{v\psi}')'$, *and* $\Delta_{v\psi}$ *is the one-sided long run covariance between* v_t *and* $\psi_\tau(u_t(\tau))$.

The asymptotic representation is composed of integrals of Brownian motions and a bias term. In particular, the asymptotic distribution of the cointegrating coefficient estimates, $\hat{\beta}_j(\tau)$ for $j=1,...,m+1$, is as follows:

$$T(\hat{\beta}_j(\tau)-\beta_j^0(\tau)) \Rightarrow \left[f(F^{-1}(\tau))\int_{\lambda_{j-1}^0}^{\lambda_j^0}\underline{B}_v\underline{B}_v'\right]^{-1}\left[\int_{\lambda_{j-1}^0}^{\lambda_j^0}\underline{B}_v dB_\psi+(\lambda_j^0-\lambda_{j-1}^0)\Delta_{v\psi}\right], \tag{2.8}$$

where $\underline{B}_v(r)=B_v(r)-\int_0^1 B_v(r)dr$ is a k-dimensional demeaned Brownian motion.

Due to serial cross correlation, the limit distribution of $\hat{\beta}_j(\tau)$ has a second-order bias term $(\lambda_j^0-\lambda_{j-1}^0)\Delta_{v\psi}$. Moreover, the above limit distribution depends on the nuisance parameter Ω, since the distribution of $\int_{\lambda_{j-1}^0}^{\lambda_j^0}\underline{B}_v dB_\psi$ depends on the correlation between $B_v(r)$ and $B_\psi(r)$, which is unknown in general. In order to develop useful inference procedures, a fully modified estimator is required to correct the bias and endogeneity.

Following Phillips and Hansen (1990), define $\psi_\tau^+(u_t(\tau))=\psi_\tau(u_t(\tau))-\Omega_{\psi v}\Omega_{vv}^{-1}v_t$ so that $\psi_\tau^+(u_t(\tau))$ is uncorrelated with v_t and has variance $\omega_{\psi.v}^2=\omega_\psi^2-\Omega_{\psi v}\Omega_{vv}^{-1}\Omega_{v\psi}$. Then, we have

$$\begin{aligned} T^{-1}\sum_{t=T_{j-1}^0+1}^{T_j^0}x_t\psi_\tau^+(u_t(\tau)) &\Rightarrow \int_{\lambda_{j-1}^0}^{\lambda_j^0}B_v dB_{\psi.v}+(\lambda_j^0-\lambda_{j-1}^0)\Delta_{v\psi}^+ \\ &= \omega_{\psi.v}\int_{\lambda_{j-1}^0}^{\lambda_j^0}B_v dW+(\lambda_j^0-\lambda_{j-1}^0)\Delta_{v\psi}^+, \end{aligned}$$

where $B_{\psi.v}(r) = B_\psi(r) - \Omega_{\psi v}\Omega_{vv}^{-1}B_v(r)$ is a Brownian motion independent of $B_v(r)$, $W(r) = B_{\psi.v}(r)/\omega_{\psi.v}$ is a standard Brownian motion, and $\Delta_{v\psi}^{+} = \Delta_{v\psi} - \Omega_{\psi v}\Omega_{vv}^{-1}\Delta_{vv}$.

Let $\hat{\Delta}_{v\psi}^{+} = \hat{\Delta}_{v\psi} - \hat{\Omega}_{\psi v}\hat{\Omega}_{vv}^{-1}\hat{\Delta}_{vv}$, where $\hat{\Omega}_{\psi v} = \hat{\Omega}'_{v\psi}$, $\hat{\Omega}_{vv}$, $\hat{\Delta}_{v\psi}$, and $\hat{\Delta}_{vv}$ are the kernel estimates,

$$\hat{\Omega}_{v\psi} = \sum_{h=-M}^{M} K(\frac{h}{M})\Gamma_{v\psi}(h), \quad \hat{\Delta}_{v\psi} = \sum_{h=0}^{M} K(\frac{h}{M})\Gamma_{v\psi}(h),$$
$$\hat{\Omega}_{vv} = \sum_{h=-M}^{M} K(\frac{h}{M})\Gamma_{vv}(h), \quad \hat{\Delta}_{vv} = \sum_{h=0}^{M} K(\frac{h}{M})\Gamma_{vv}(h),$$

where $\Gamma_{v\psi}(h) = T^{-1}\sum_{t=1}^{T-h}\Delta x_t\psi_\tau(\hat{u}_{t+h}(\tau))$, $\Gamma_{vv}(h) = T^{-1}\sum_{t=1}^{T-h}\Delta x_t\Delta x'_{t+h}$, $K(\cdot)$ is a kernel function, and M is the bandwidth. As suggested by Andrews (1991) and Hansen (1992), the quadratic spectral kernel is adopted such that

$$K(u) = \frac{25}{12\pi^2u^2}\left[\frac{\sin(6\pi u/5)}{6\pi u/5} - \cos(6\pi u/5)\right]$$

and the plug-in bandwidth estimator is $\hat{M} = 1.3221(aT)^{\frac{1}{5}}$, where

$$a = \sum_{i=1}^{k+1}\frac{4\hat{\rho}_i^2\hat{\sigma}_i^2}{(1-\hat{\rho}_i)^8} \Big/ \sum_{i=1}^{k+1}\frac{\hat{\sigma}_i^2}{(1-\hat{\rho}_i)^4}.$$

To calculate the bandwidth, one should estimate a univariate AR(1) model for each element ξ_{it} $(i = 1, ..., k+1)$ of ξ_t and denote by $(\hat{\rho}_i, \hat{\sigma}_i^2)$ the autoregressive and innovation variance estimates.

The fully modified estimator of the coefficients associated with the I(1) regressors is given by

$$\hat{\beta}_j^{+}(\tau) = \hat{\beta}_j(\tau) - \left[f(\widehat{F^{-1}(\tau)})\sum_{T_{j-1}^0+1}^{T_j^0}\underline{x}_t\underline{x}'_t\right]^{-1}\left[\sum_{T_{j-1}^0+1}^{T_j^0}\underline{x}_t\hat{\Omega}_{\psi v}\hat{\Omega}_{vv}^{-1}\Delta x_t + (\lambda_j^0 - \lambda_{j-1}^0)T\hat{\Delta}_{v\psi}^{+}\right], \tag{2.9}$$

where $f(\widehat{F^{-1}(\tau)})$ is a nonparametric consistent estimator of $f(F^{-1}(\tau))$ and $\underline{x}_t$ denotes the demeaned regressors. In practice, the density function $f(F^{-1}(\tau))$ is estimated using the

Gaussian kernel and Silverman's "rule-of-thumb" bandwidth. Consequently, the estimate $\hat{\beta}_j^+(\tau)$ follows a mixed normal distribution in the limit such that

$$\begin{aligned} T(\hat{\beta}_j^+(\tau) - \beta_j^0(\tau)) &\Rightarrow \left[f(F^{-1}(\tau)) \int_{\lambda_{j-1}^0}^{\lambda_j^0} \underline{B}_v \underline{B}_v' \right]^{-1} \omega_{\psi.v} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \underline{B}_v dW \\ &\sim MN\left(0, \frac{\omega_{\psi.v}^2}{f(F^{-1}(\tau))^2} \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \underline{B}_v \underline{B}_v' \right]^{-1} \right). \end{aligned}$$

Therefore, the asymptotic distribution of the fully modified coefficient estimates $\hat{\theta}_j^+(\tau) = (\hat{\alpha}_j(\tau), \hat{\beta}_j^{+\prime}(\tau))'$ is as follows:

Theorem 2 *Under Assumptions 1-5, for $j = 1, ..., m+1$, $\hat{\theta}_j^+(\tau)$ is a consistent estimator of $\theta_j^0(\tau)$ and*

$$D_T(\hat{\theta}_j^+(\tau) - \theta_j^0(\tau)) \Rightarrow \left[f(F^{-1}(\tau)) \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v' \right]^{-1} \omega_{\psi.v} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dW \quad (2.10)$$

$$\sim MN\left(0, \frac{\omega_{\psi.v}^2}{f(F^{-1}(\tau))^2} \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v' \right]^{-1} \right), \quad (2.11)$$

where $\bar{B}_v = (1, B_v')'$ and $\omega_{\psi.v}^2 = \omega_\psi^2 - \Omega_{\psi v}\Omega_{vv}^{-1}\Omega_{v\psi}$.

The fully modified estimator provides a basis for tests of linear restrictions on the regression coefficients. In particular, to test for q restrictions on the cointegrating coefficients, one may consider the null hypothesis $H_0 : R\beta_j^0(\tau) = r$, where r is a q-dimensional vector. Under the null, we have

$$T(R\hat{\beta}_j^+(\tau) - r) \Rightarrow MN\left(0, \frac{\omega_{\psi.v}^2}{f(F^{-1}(\tau))^2} \left[R \left(\int_{\lambda_{j-1}^0}^{\lambda_j^0} \underline{B}_v \underline{B}_v' \right)^{-1} R' \right] \right).$$

A Wald statistic can be used to test the null hypothesis. Under H_0, the test statistic converges to a chi-squared distribution with q degrees of freedom. We have

$$\mathrm{W}_j(\tau) = \frac{\widehat{f(F^{-1}(\tau))}^2}{\hat{\omega}^2_{\psi.v}}(R\hat{\beta}^+_j(\tau) - r)' \left[R \left(\sum_{t=T^0_{j-1}+1}^{T^0_j} \underline{x}_t \underline{x}'_t \right)^{-1} R' \right]^{-1} (R\hat{\beta}^+_j(\tau) - r) \Rightarrow \chi^2_q. \tag{2.12}$$

In addition, the break fraction estimates, $\hat{\lambda}_j$ for $j = 1, ..., m$, are consistent. More formally, with shrinking shifts, we have the following theorem:

Theorem 3 *Under Assumptions 1-6,* $\hat{\lambda}_j \rightarrow^p \lambda^0_j$ *and* $\nu^2_T T(\hat{\lambda}_j - \lambda^0_j) = O_p(1)$ *for* $j = 1, ..., m$.

Given the rate of convergence $\nu^2_T T$, the asymptotic distribution of the break fraction estimates is given as below:

Theorem 4 *Under Assumptions 1-6, for* $j = 1, ..., m$,

$$\nu^2_T T(\hat{\lambda}_j - \lambda^0_j) \Rightarrow \arg\max_s \begin{cases} \delta'_j(\tau) V_j(s) - \frac{|s|}{2}\pi_j & s \leq 0 \\ \delta'_j(\tau) V_{j+1}(s) - \frac{|s|}{2}\pi_j & s > 0, \end{cases} \tag{2.13}$$

where $V_j(s)$ *is a* $(k+1)$*-dimensional random vector that depends on* j*,* s*, and* $\Omega_{v\psi}$ *and* $\pi_j = f(F^{-1}(\tau))\delta'_j(\tau)\bar{B}_v(\lambda^0_j)\bar{B}'_v(\lambda^0_j)\delta_j(\tau)$ *with* $\bar{B}_v(\lambda^0_j) = (1, B'_v(\lambda^0_j))'$.

This limit distribution is nonstandard and depends on the quantile level τ and the correlation between v_t and $\psi_\tau(u_t(\tau))$. The analytical distribution function cannot be solved. If the regressor vector x_t is strictly exogenous, the asymptotic distribution of $\hat{\lambda}_j$ can be written in the following form:

Corollary 5 *Under Assumptions 1-6 and strict exogeneity of the regressors, for* $j = 1, ..., m$*,*

$$\nu_T^2 T(\hat{\lambda}_j - \lambda_j^0) \Rightarrow \arg\max_s \begin{cases} \eta_j B_j(s) - \frac{|s|}{2}\pi_j & s \leq 0 \\ \eta_j B_{j+1}(s) - \frac{|s|}{2}\pi_j & s > 0, \end{cases} \quad (2.14)$$

where $B_j(s)$ *is a two-sided Brownian motion with variance* ω_ψ^2 *and* $\eta_j = \delta_j'(\tau)\bar{B}_v(\lambda_j^0)$*.*

This convergence condition corresponds to that from Kejriwal and Perron (2008) who consider the least squares regression case. By changing variables, with strictly exogenous regressors, the limit distribution for each $\hat{\lambda}_j$ can be rewritten as

$$\left(\frac{\pi_j}{\omega_\psi \eta_j}\right)^2 \nu_T^2 T(\hat{\lambda}_j - \lambda_j^0) \Rightarrow \arg\max_s \{W(s) - |s|/2\}, \quad (2.15)$$

where $W(s)$ is a standard Brownian motion which is zero when $s = 0$. When $s < 0$, $W(s) = W_j(s)$, and when $s > 0$, $W(s) = W_{j+1}(s)$. Fortunately, the analytical distribution function for the right hand side can be obtained, as in Bai (1997). Assuming that the distribution of the regressors and regression disturbance are the same for all regimes, the distribution of each break fraction is symmetric.

2.3 Simulation Study

2.3.1 Bias and root mean square error

In order to evaluate the accuracy of the estimation method, the bias and root mean square error (RMSE) of the quantile estimators, $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ for $j = 1, 2$, are calculated. The number of Monte Carlo replications is 10000. The sample sizes considered are $T = 120$ and $T = 240$[7]. For simplicity, the simulation study focuses on the single break

[7] For ease of comparison, the sample sizes are the same as those used in the simulation experiments from Kejriwal and Perron (2008).

case with $\lambda_1^0 = 0.50$. In later discussion, another single break case with $\lambda_1^0 = 0.25$ is considered for comparison. The quantile levels of interest are $\tau = 0.1, 0.2, ..., 0.9$. The level of trimming is $\epsilon = 0.15$ so that the unknown break date is searched over the interval of $[\epsilon T, (1-\epsilon)T]$ and the break fraction is in $[\epsilon, (1-\epsilon)]$. The simulated data is generated by the simple DGP, $x_t = x_{t-1} + v_t$, where $v_t \sim$ i.i.d.$N(0,1)$ and

$$y_t = \begin{cases} \alpha_1^0 + x_t'\beta_1^0 + u_t & t = 1, ..., T_1^0 \\ \alpha_2^0 + x_t'\beta_2^0 + u_t & t = T_1^0 + 1, ..., T, \end{cases}$$

where $u_t \sim$ i.i.d.$N(0,1)$, $\sigma_{uv} = corr(u_t, v_t) = 0, 0.5, 0.9$, and $T_1^0 = [\lambda_1^0 T]$. The break size is $\delta_1^0 = 1$[8] and $\alpha_1^0 = 1, \beta_1^0 = 1$ so that after the break $\alpha_2^0 = 2, \beta_2^0 = 2$.

The simulation results for the break fraction and slope coefficient estimates are reported in the following tables. Tables 2.1 and 2.2 report the bias and RMSE of the coefficient estimates, $\hat{\beta}_j(\tau)$ and $\hat{\beta}_j^+(\tau)$ for $j = 1, 2$, assuming the break date is known with $\lambda_1^0 = 0.50$. The sample sizes are $T = 120$ and $T = 240$, respectively. For $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$, the original quantile estimator is unbiased under strict exogeneity. The finite sample bias from the simulations is negligible when $\sigma_{uv} = 0$. When the cross correlation is nonzero, there is a second-order bias of the original quantile estimator and the finite sample bias grows as σ_{uv} increases. When $T = 120$ the largest bias is 0.0771, which occurs for $\hat{\beta}_1(\tau)$ in the 80% and 90% quantiles when $\sigma_{uv} = 0.9$. It is a 7.71% bias, since the true value β_1^0 is one. In general, the bias of $\hat{\beta}_2(\tau)$ is smaller than that of $\hat{\beta}_1(\tau)$ in percentage since β_2^0 is twice as large as β_1^0. The maximum of the RMSE is 0.1278, which occurs for $\hat{\beta}_1(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.9$. In each panel, the RMSE reaches its maximum in either the lower or upper tail. The results are better with larger sample size. When $T = 240$ and $\sigma_{uv} = 0.9$, the largest bias and RMSE are 0.0396 and 0.0649, respectively.

[8] This break size is also studied by Kejriwal and Perron (2008) and Oka and Qu (2011).

Table 2.1: Bias and RMSE of $\hat{\beta}_j(\tau)$ and $\hat{\beta}_j^+(\tau)$ with known T_1^0 ($T = 120$, $\lambda_1^0 = 0.50$)

		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
						$\sigma_{uv} = 0$				
$\hat{\beta}_1(\tau)$	Bias	-0.0007	0.0005	0.0009	0.0009	0.0010	0.0019	0.0013	0.0017	0.0023
	RMSE	0.0963	0.0800	0.0737	0.0711	0.0688	0.0703	0.0728	0.0793	0.0954
$\hat{\beta}_2(\tau)$	Bias	-0.0016	-0.0007	-0.0001	-0.0001	-0.0002	0.0008	-0.0001	0.0008	0.0001
	RMSE	0.0943	0.0774	0.0714	0.0683	0.0683	0.0692	0.0725	0.0788	0.0950
$\hat{\beta}_1^+(\tau)$	Bias	-0.0008	0.0006	0.0008	0.0010	0.0010	0.0019	0.0012	0.0016	0.0023
	RMSE	0.0997	0.0831	0.0763	0.0735	0.0709	0.0727	0.0747	0.0818	0.0986
$\hat{\beta}_2^+(\tau)$	Bias	-0.0018	-0.0007	-0.0002	-0.0003	-0.0004	0.0006	-0.0003	0.0006	-0.0002
	RMSE	0.0980	0.0803	0.0740	0.0707	0.0706	0.0715	0.0753	0.0818	0.0992
						$\sigma_{uv} = 0.5$				
$\hat{\beta}_1(\tau)$	Bias	0.0421	0.0420	0.0424	0.0428	0.0432	0.0437	0.0442	0.0435	0.0442
	RMSE	0.1060	0.0917	0.0866	0.0847	0.0844	0.0850	0.0878	0.0920	0.1064
$\hat{\beta}_2(\tau)$	Bias	0.0422	0.0428	0.0434	0.0431	0.0437	0.0444	0.0442	0.0440	0.0441
	RMSE	0.1062	0.0916	0.0872	0.0848	0.0843	0.0856	0.0876	0.0933	0.1078
$\hat{\beta}_1^+(\tau)$	Bias	0.0093	0.0064	0.0050	0.0046	0.0048	0.0056	0.0072	0.0082	0.0115
	RMSE	0.0955	0.0775	0.0705	0.0669	0.0661	0.0666	0.0702	0.0768	0.0948
$\hat{\beta}_2^+(\tau)$	Bias	0.0101	0.0078	0.0065	0.0052	0.0057	0.0067	0.0076	0.0090	0.0115
	RMSE	0.0956	0.0771	0.0700	0.0671	0.0662	0.0671	0.0696	0.0778	0.0959
						$\sigma_{uv} = 0.9$				
$\hat{\beta}_1(\tau)$	Bias	0.0742	0.0758	0.0760	0.0763	0.0762	0.0767	0.0770	0.0771	0.0771
	RMSE	0.1242	0.1146	0.1110	0.1099	0.1087	0.1090	0.1117	0.1159	0.1278
$\hat{\beta}_2(\tau)$	Bias	0.0762	0.0763	0.0767	0.0773	0.0773	0.0775	0.0778	0.0774	0.0758
	RMSE	0.1265	0.1146	0.1114	0.1106	0.1099	0.1108	0.1129	0.1171	0.1270
$\hat{\beta}_1^+(\tau)$	Bias	0.0150	0.0122	0.0093	0.0080	0.0075	0.0085	0.0103	0.0137	0.0185
	RMSE	0.0876	0.0660	0.0573	0.0528	0.0506	0.0519	0.0571	0.0667	0.0885
$\hat{\beta}_2^+(\tau)$	Bias	0.0176	0.0129	0.0102	0.0090	0.0085	0.0092	0.0110	0.0139	0.0167
	RMSE	0.0869	0.0660	0.0570	0.0534	0.0525	0.0538	0.0577	0.0674	0.0880

Notes: The results are based on 10000 Monte Carlo replications with sample size T and correlation σ_{uv}.

Table 2.2: Bias and RMSE of $\hat{\beta}_j(\tau)$ and $\hat{\beta}_j^+(\tau)$ with known T_1^0 ($T = 240$, $\lambda_1^0 = 0.50$)

		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
						$\sigma_{uv} = 0$				
$\hat{\beta}_1(\tau)$	Bias	-0.0008	-0.0002	0.0002	0.0001	0.0000	0.0003	0.0001	-0.0001	0.0004
	RMSE	0.0475	0.0394	0.0360	0.0348	0.0340	0.0347	0.0362	0.0389	0.0474
$\hat{\beta}_2(\tau)$	Bias	0.0000	-0.0000	0.0001	0.0004	0.0003	0.0002	0.0004	0.0009	0.0008
	RMSE	0.0470	0.0386	0.0356	0.0346	0.0336	0.0343	0.0356	0.0385	0.0467
$\hat{\beta}_1^+(\tau)$	Bias	-0.0008	-0.0001	0.0002	0.0001	-0.0000	0.0002	0.0001	-0.0000	0.0004
	RMSE	0.0486	0.0403	0.0368	0.0355	0.0348	0.0355	0.0369	0.0396	0.0481
$\hat{\beta}_2^+(\tau)$	Bias	0.0001	0.0001	0.0001	0.0005	0.0004	0.0002	0.0004	0.0009	0.0007
	RMSE	0.0477	0.0393	0.0362	0.0352	0.0342	0.0348	0.0362	0.0390	0.0471
						$\sigma_{uv} = 0.5$				
$\hat{\beta}_1(\tau)$	Bias	0.0214	0.0215	0.0218	0.0218	0.0222	0.0222	0.0224	0.0226	0.0224
	RMSE	0.0527	0.0462	0.0435	0.0428	0.0425	0.0430	0.0442	0.0471	0.0539
$\hat{\beta}_2(\tau)$	Bias	0.0219	0.0219	0.0218	0.0221	0.0226	0.0225	0.0224	0.0224	0.0224
	RMSE	0.0536	0.0464	0.0438	0.0431	0.0430	0.0429	0.0441	0.0466	0.0539
$\hat{\beta}_1^+(\tau)$	Bias	0.0025	0.0016	0.0013	0.0009	0.0012	0.0014	0.0018	0.0027	0.0035
	RMSE	0.0457	0.0379	0.0342	0.0331	0.0325	0.0329	0.0342	0.0376	0.0465
$\hat{\beta}_2^+(\tau)$	Bias	0.0029	0.0020	0.0011	0.0011	0.0015	0.0015	0.0017	0.0024	0.0033
	RMSE	0.0464	0.0376	0.0342	0.0328	0.0320	0.0321	0.0340	0.0374	0.0462
						$\sigma_{uv} = 0.9$				
$\hat{\beta}_1(\tau)$	Bias	0.0388	0.0391	0.0394	0.0394	0.0396	0.0393	0.0393	0.0396	0.0393
	RMSE	0.0643	0.0591	0.0573	0.0568	0.0566	0.0566	0.0571	0.0596	0.0648
$\hat{\beta}_2(\tau)$	Bias	0.0391	0.0397	0.0395	0.0399	0.0398	0.0401	0.0398	0.0399	0.0394
	RMSE	0.0645	0.0597	0.0577	0.0572	0.0568	0.0574	0.0576	0.0599	0.0649
$\hat{\beta}_1^+(\tau)$	Bias	0.0046	0.0031	0.0022	0.0016	0.0016	0.0015	0.0022	0.0037	0.0052
	RMSE	0.0414	0.0313	0.0271	0.0253	0.0245	0.0251	0.0268	0.0318	0.0414
$\hat{\beta}_2^+(\tau)$	Bias	0.0046	0.0034	0.0019	0.0017	0.0014	0.0019	0.0023	0.0036	0.0049
	RMSE	0.0422	0.0318	0.0275	0.0254	0.0250	0.0254	0.0271	0.0315	0.0417

Notes: Tables 2.1 and 2.2 report the bias and RMSE of $\hat{\beta}_j(\tau)$ and $\hat{\beta}_j^+(\tau)$ ($j = 1, 2$), with known $T_1^0 = [\lambda_1^0 T]$.

The results for the fully modified quantile estimator, $\hat{\beta}_1^+(\tau)$ and $\hat{\beta}_2^+(\tau)$, are also included in Tables 2.1 and 2.2. When there is no residual cross correlation, it is not necessary to use the fully modified estimator. Since there is no bias under strict exogeneity, the fully modified estimator does not outperform the original quantile estimator in this case. However, the results are not much worse after the modification, which means that the cost of using the fully modified estimator is very small. To investigate the performance of the fully modified quantile estimator under endogeneity, the two cases with nonzero values of σ_{uv} are of particular interest. The density function $f(F^{-1}(\tau))$ is estimated using the Gaussian kernel and Silverman's "rule-of-thumb" bandwidth. With small sample such as $T = 120$, the finite sample bias is significantly reduced for all quantiles considered. The maximum bias is 1.85%, which is for $\hat{\beta}_1(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.9$. The magnitudes of RMSE are reduced to, at least, those in the unbiased case for all quantiles. The largest RMSE is 0.0959, which occurs for $\hat{\beta}_2^+(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.5$. The smallest RMSE is only 0.0506 for $\hat{\beta}_1^+(\tau)$ in the median when $\sigma_{uv} = 0.9$. Similarly, when $T = 240$ both the bias and RMSE are much smaller than those from Table 2.1. After the modification, the largest bias reduces to 0.52%, which is for $\hat{\beta}_1^+(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.9$ and the largest RMSE is 0.0465, which is for $\hat{\beta}_1^+(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.5$. In most cases, the bias and RMSE of the fully modified coefficient estimates are smaller for the inner quantiles than those for the outer quantiles.

Tables 2.3 and 2.4 contain the results for the break fraction and regression coefficient estimates when the true break date is unknown. Overall, the bias and RMSE of the break fraction estimates are small. When $T = 120$ the maximum of the bias of $\hat{\lambda}_1$ is -0.0040, which occurs in the 10% quantile with $\sigma_{uv} = 0.9$. It is less than half observation ($t = 1/2$) biased. The RMSE of $\hat{\lambda}_1$ is small over all quantiles with the maximum equal to 0.0261, which occurs in the 90% quantile when $\sigma_{uv} = 0.9$. The bias and RMSE are smaller when the sample size is larger. When $T = 240$ the largest bias of $\hat{\lambda}_1$ is only -0.0016 with $\sigma_{uv} = 0.9$. The largest RMSE of $\hat{\lambda}_1$ is 0.0110 when $\sigma_{uv} = 0.9$ and $\tau = 0.1$. Moreover, in most cases, the bias and RMSE of the break fraction estimates are smaller for the inner quantiles than for the outer quantiles. Thus, the quantile estimator for the break point is very reliable.

With the estimated break date, $\hat{T}_1 = [\hat{\lambda}_1 T]$, the results for the slope coefficient estimates are similar to those from the known break date cases. For the original quantile estimator, when $\sigma_{uv} = 0$ there is no bias. In this case, the bias and RMSE in finite samples are very small. With the presence of endogeneity such as $\sigma_{uv} = 0.9$, the largest bias is 7.80%, which is for $\hat{\beta}_1(\tau)$ in the 90% quantile and the RMSE can be as large as 0.1319, which occurs for $\hat{\beta}_1(\tau)$ in the 90% quantile. The results improve with larger sample size as expected. However, the finite sample performance of the original quantile estimator $\hat{\beta}_j(\tau)$ is still poor under endogeneity when $T = 240$. The largest bias of $\hat{\beta}_1(\tau)$ is 3.96%, which occurs in the 80% quantile when $\sigma_{uv} = 0.9$. The maximum of RMSE is 0.0654 for $\hat{\beta}_2(\tau)$, which occurs in the 90% quantile when $\sigma_{uv} = 0.9$. After applying the fully modified estimation, the bias and RMSE of $\hat{\beta}_1^+(\tau)$ and $\hat{\beta}_2^+(\tau)$ are largely reduced. With $T = 120$ the maximum bias reduces to 2.10%, which is for $\hat{\beta}_1^+(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.9$ and the maximum RMSE reduces to 0.0994, for $\hat{\beta}_1^+(\tau)$ in the 90% quantile when $\sigma_{uv} = 0.5$. With $T = 240$ the bias is no larger than 0.56% and the RMSE is no larger than 0.0468 for all σ_{uv} and τ considered.

Table 2.3: Bias and RMSE of $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ with unknown T_1^0 ($T = 120$, $\lambda_1^0 = 0.50$)

		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
						$\sigma_{uv} = 0$				
$\hat{\lambda}_1$	Bias	-0.0004	-0.0003	-0.0002	-0.0001	-0.0003	-0.0002	-0.0001	-0.0003	-0.0003
	RMSE	0.0232	0.0192	0.0169	0.0161	0.0167	0.0164	0.0171	0.0192	0.0244
$\hat{\beta}_1(\tau)$	Bias	-0.0021	-0.0004	0.0004	0.0008	0.0010	0.0014	0.0009	0.0013	0.0018
	RMSE	0.0986	0.0819	0.0751	0.0721	0.0700	0.0713	0.0738	0.0808	0.1002
$\hat{\beta}_2(\tau)$	Bias	-0.0018	-0.0007	-0.0002	0.0002	-0.0002	0.0008	0.0000	0.0008	0.0001
	RMSE	0.0955	0.0785	0.0725	0.0694	0.0697	0.0704	0.0739	0.0802	0.0972
$\hat{\beta}_1^+(\tau)$	Bias	-0.0023	-0.0004	0.0004	0.0009	0.0009	0.0014	0.0008	0.0012	0.0017
	RMSE	0.1019	0.0849	0.0778	0.0744	0.0721	0.0734	0.0758	0.0833	0.1025
$\hat{\beta}_2^+(\tau)$	Bias	-0.0022	-0.0008	-0.0003	-0.0000	-0.0004	0.0006	-0.0003	0.0005	-0.0003
	RMSE	0.0991	0.0813	0.0752	0.0716	0.0720	0.0726	0.0766	0.0832	0.1011
						$\sigma_{uv} = 0.5$				
$\hat{\lambda}_1$	Bias	-0.0024	-0.0020	-0.0019	-0.0020	-0.0021	-0.0021	-0.0022	-0.0023	-0.0027
	RMSE	0.0241	0.0197	0.0174	0.0165	0.0167	0.0173	0.0185	0.0210	0.0254
$\hat{\beta}_1(\tau)$	Bias	0.0413	0.0420	0.0426	0.0425	0.0429	0.0438	0.0440	0.0434	0.0447
	RMSE	0.1083	0.0937	0.0885	0.0854	0.0850	0.0859	0.0884	0.0932	0.1105
$\hat{\beta}_2(\tau)$	Bias	0.0429	0.0431	0.0436	0.0433	0.0437	0.0445	0.0443	0.0443	0.0444
	RMSE	0.1078	0.0929	0.0880	0.0856	0.0855	0.0865	0.0886	0.0947	0.1100
$\hat{\beta}_1^+(\tau)$	Bias	0.0095	0.0069	0.0056	0.0047	0.0048	0.0061	0.0074	0.0088	0.0131
	RMSE	0.0985	0.0796	0.0727	0.0681	0.0674	0.0677	0.0715	0.0787	0.0994
$\hat{\beta}_2^+(\tau)$	Bias	0.0112	0.0083	0.0070	0.0059	0.0062	0.0071	0.0080	0.0097	0.0123
	RMSE	0.0968	0.0785	0.0709	0.0680	0.0677	0.0683	0.0707	0.0797	0.0981
						$\sigma_{uv} = 0.9$				
$\hat{\lambda}_1$	Bias	-0.0040	-0.0038	-0.0033	-0.0033	-0.0032	-0.0033	-0.0034	-0.0034	-0.0039
	RMSE	0.0254	0.0219	0.0192	0.0194	0.0193	0.0200	0.0207	0.0212	0.0261
$\hat{\beta}_1(\tau)$	Bias	0.0747	0.0761	0.0764	0.0767	0.0768	0.0771	0.0779	0.0779	0.0780
	RMSE	0.1271	0.1164	0.1120	0.1111	0.1104	0.1106	0.1139	0.1182	0.1319
$\hat{\beta}_2(\tau)$	Bias	0.0768	0.0763	0.0771	0.0776	0.0777	0.0781	0.0782	0.0780	0.0765
	RMSE	0.1287	0.1156	0.1125	0.1117	0.1109	0.1118	0.1140	0.1180	0.1297
$\hat{\beta}_1^+(\tau)$	Bias	0.0173	0.0128	0.0097	0.0084	0.0081	0.0090	0.0114	0.0151	0.0210
	RMSE	0.0912	0.0672	0.0582	0.0537	0.0522	0.0537	0.0589	0.0692	0.0939
$\hat{\beta}_2^+(\tau)$	Bias	0.0192	0.0132	0.0108	0.0097	0.0093	0.0102	0.0117	0.0148	0.0181
	RMSE	0.0891	0.0678	0.0584	0.0552	0.0538	0.0549	0.0592	0.0686	0.0907

Table 2.4: Bias and RMSE of $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ with unknown T_1^0 ($T = 240$, $\lambda_1^0 = 0.50$)

		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$					
$\hat{\lambda}_1$	Bias	-0.0001	-0.0000	-0.0001	-0.0002	-0.0001	-0.0001	-0.0001	-0.0002	-0.0001
	RMSE	0.0095	0.0072	0.0066	0.0063	0.0061	0.0061	0.0065	0.0072	0.0089
$\hat{\beta}_1(\tau)$	Bias	-0.0010	-0.0003	0.0001	0.0001	-0.0000	0.0002	-0.0001	-0.0003	0.0003
	RMSE	0.0479	0.0396	0.0361	0.0350	0.0341	0.0349	0.0365	0.0391	0.0476
$\hat{\beta}_2(\tau)$	Bias	0.0000	0.0001	0.0001	0.0004	0.0004	0.0001	0.0004	0.0009	0.0008
	RMSE	0.0475	0.0388	0.0357	0.0346	0.0338	0.0345	0.0358	0.0387	0.0471
$\hat{\beta}_1^+(\tau)$	Bias	-0.0009	-0.0002	0.0001	0.0001	-0.0001	0.0001	-0.0000	-0.0002	0.0002
	RMSE	0.0490	0.0405	0.0369	0.0357	0.0348	0.0355	0.0371	0.0398	0.0483
$\hat{\beta}_2^+(\tau)$	Bias	0.0001	0.0002	0.0002	0.0005	0.0004	0.0001	0.0004	0.0010	0.0007
	RMSE	0.0483	0.0394	0.0364	0.0352	0.0344	0.0350	0.0364	0.0392	0.0476
					$\sigma_{uv} = 0.5$					
$\hat{\lambda}_1$	Bias	-0.0009	-0.0007	-0.0006	-0.0006	-0.0006	-0.0006	-0.0007	-0.0007	-0.0007
	RMSE	0.0092	0.0072	0.0066	0.0064	0.0064	0.0062	0.0066	0.0071	0.0089
$\hat{\beta}_1(\tau)$	Bias	0.0212	0.0212	0.0217	0.0218	0.0221	0.0222	0.0223	0.0226	0.0223
	RMSE	0.0530	0.0461	0.0435	0.0428	0.0426	0.0431	0.0442	0.0472	0.0540
$\hat{\beta}_2(\tau)$	Bias	0.0221	0.0219	0.0218	0.0222	0.0226	0.0225	0.0224	0.0225	0.0224
	RMSE	0.0539	0.0464	0.0439	0.0432	0.0431	0.0429	0.0442	0.0468	0.0541
$\hat{\beta}_1^+(\tau)$	Bias	0.0024	0.0015	0.0012	0.0009	0.0012	0.0015	0.0019	0.0029	0.0037
	RMSE	0.0462	0.0380	0.0344	0.0333	0.0327	0.0330	0.0344	0.0378	0.0468
$\hat{\beta}_2^+(\tau)$	Bias	0.0031	0.0021	0.0013	0.0014	0.0018	0.0017	0.0019	0.0026	0.0034
	RMSE	0.0467	0.0377	0.0343	0.0329	0.0322	0.0322	0.0341	0.0376	0.0465
					$\sigma_{uv} = 0.9$					
$\hat{\lambda}_1$	Bias	-0.0016	-0.0014	-0.0013	-0.0013	-0.0013	-0.0012	-0.0013	-0.0013	-0.0014
	RMSE	0.0110	0.0094	0.0086	0.0082	0.0081	0.0080	0.0086	0.0088	0.0103
$\hat{\beta}_1(\tau)$	Bias	0.0388	0.0390	0.0393	0.0394	0.0395	0.0393	0.0394	0.0396	0.0393
	RMSE	0.0646	0.0591	0.0572	0.0568	0.0567	0.0566	0.0572	0.0598	0.0648
$\hat{\beta}_2(\tau)$	Bias	0.0393	0.0399	0.0396	0.0400	0.0398	0.0402	0.0399	0.0399	0.0395
	RMSE	0.0649	0.0601	0.0578	0.0574	0.0569	0.0576	0.0578	0.0602	0.0654
$\hat{\beta}_1^+(\tau)$	Bias	0.0050	0.0032	0.0022	0.0016	0.0016	0.0016	0.0023	0.0038	0.0056
	RMSE	0.0420	0.0314	0.0271	0.0254	0.0246	0.0253	0.0269	0.0319	0.0417
$\hat{\beta}_2^+(\tau)$	Bias	0.0050	0.0038	0.0022	0.0021	0.0018	0.0024	0.0026	0.0038	0.0051
	RMSE	0.0425	0.0322	0.0279	0.0258	0.0253	0.0256	0.0274	0.0319	0.0421

Notes: Tables 2.3 and 2.4 report the bias and RMSE of the estimate $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ ($j = 1, 2$).

In general, the break dates can be estimated by minimizing the check function without any modification as in equation (2.5). However, it is worthwhile to apply the fully modified estimator for the coefficients associated with the integrated regressors in order to reduce the bias and RMSE when the residual cross correlation is nonzero.

2.3.2 Finite sample coverage rate

Tables 2.5 and 2.6 report the finite sample coverage properties of the 95% confidence intervals for the slope coefficients, $\beta_1^0(\tau)$ and $\beta_2^0(\tau)$, in the case of known break date. The first two rows in each panel contain the coverage rates of the confidence intervals calculated using the original quantile estimator. When $\sigma_{uv} = 0$ the coverage rates are close to 95%. With nonzero correlation the coverage rates are inadequate especially for large σ_{uv}. As shown in Table 2.5, when $T = 120$ and $\sigma_{uv} = 0.9$ the coverage rate is only 82.54% for $\beta_1^0(\tau)$ in the 90% quantile. It is a 12.46% undercoverage. With $T = 240$ as shown in Table 2.6, the smallest coverage rate is 83.61% for $\beta_1^0(\tau)$ in the median when $\sigma_{uv} = 0.9$, which is a 11.39% undercoverage. This is due to the bias of the original quantile estimator under endogeneity. When the residual cross correlation is large, the centering of the finite sample confidence intervals is severely affected. Consequently, the true value $\beta_j^0(\tau)$ falls outside the confidence intervals too frequently. In this case, the coverage is inadequate for all quantiles. The last two rows in each panel summarize the results when the confidence intervals are calculated using the fully modified estimator. When $\sigma_{uv} = 0$ the coverage rates are slightly lower than those calculated using the original quantile estimator. When $\sigma_{uv} = 0.5$ and 0.9 the coverage rates become adequate after the modification. When $T = 120$ the lowest coverage rate is 90.25%, which is for $\beta_1^0(\tau)$ when $\sigma_{uv} = 0.9$ and $\tau = 0.2$. It is a 4.75% undercoverage. For all quantiles, the coverage rates are above 90%. When $T = 240$ the confidence intervals appear more conservative and closer to 95%. Some entries are slightly less than 95%, while others exceed the level by a small amount. For example, the largest coverage rate is 96.25%, which occurs in the median when $\sigma_{uv} = 0.9$. This is a 1.25% overcoverage. In general, using the fully modified estimator, the confidence intervals for the cointegrating coefficients have adequate coverage rates with slight undercoverage in some cases and they have better coverage properties for the inner quantiles than for the tails.

Table 2.5: Coverage rates for $\beta_j^0(\tau)$ with known T_1^0 ($T = 120$, $\lambda_1^0 = 0.50$)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$				
$\hat{\beta}_1(\tau)$	0.9030	0.9310	0.9471	0.9480	0.9536	0.9497	0.9450	0.9382	0.9084
$\hat{\beta}_2(\tau)$	0.9123	0.9385	0.9499	0.9545	0.9559	0.9565	0.9464	0.9360	0.9049
$\hat{\beta}_1^+(\tau)$	0.9025	0.9236	0.9368	0.9406	0.9451	0.9409	0.9353	0.9257	0.9070
$\hat{\beta}_2^+(\tau)$	0.9110	0.9309	0.9400	0.9475	0.9493	0.9442	0.9399	0.9274	0.9056
					$\sigma_{uv} = 0.5$				
$\hat{\beta}_1(\tau)$	0.8799	0.9047	0.9140	0.9173	0.9209	0.9156	0.9119	0.9056	0.8788
$\hat{\beta}_2(\tau)$	0.8839	0.9071	0.9124	0.9164	0.9148	0.9192	0.9140	0.9042	0.8768
$\hat{\beta}_1^+(\tau)$	0.9032	0.9266	0.9398	0.9429	0.9452	0.9450	0.9394	0.9318	0.9053
$\hat{\beta}_2^+(\tau)$	0.9108	0.9304	0.9406	0.9450	0.9479	0.9473	0.9419	0.9308	0.9038
					$\sigma_{uv} = 0.9$				
$\hat{\beta}_1(\tau)$	0.8320	0.8402	0.8406	0.8385	0.8402	0.8410	0.8386	0.8366	0.8254
$\hat{\beta}_2(\tau)$	0.8325	0.8369	0.8352	0.8334	0.8373	0.8415	0.8342	0.8323	0.8353
$\hat{\beta}_1^+(\tau)$	0.9049	0.9363	0.9513	0.9583	0.9620	0.9577	0.9532	0.9334	0.9025
$\hat{\beta}_2^+(\tau)$	0.9077	0.9359	0.9489	0.9577	0.9602	0.9530	0.9518	0.9336	0.9083

Table 2.6: Coverage rates for $\beta_j^0(\tau)$ with known T_1^0 ($T = 240$, $\lambda_1^0 = 0.50$)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$				
$\hat{\beta}_1(\tau)$	0.9160	0.9386	0.9498	0.9534	0.9591	0.9549	0.9495	0.9402	0.9215
$\hat{\beta}_2(\tau)$	0.9180	0.9420	0.9498	0.9506	0.9558	0.9550	0.9527	0.9430	0.9180
$\hat{\beta}_1^+(\tau)$	0.9284	0.9353	0.9462	0.9499	0.9542	0.9489	0.9464	0.9408	0.9286
$\hat{\beta}_2^+(\tau)$	0.9241	0.9394	0.9471	0.9482	0.9532	0.9525	0.9480	0.9422	0.9243
					$\sigma_{uv} = 0.5$				
$\hat{\beta}_1(\tau)$	0.9022	0.9124	0.9162	0.9180	0.9204	0.9187	0.9161	0.9077	0.8969
$\hat{\beta}_2(\tau)$	0.8937	0.9146	0.9184	0.9204	0.9204	0.9224	0.9180	0.9118	0.8895
$\hat{\beta}_1^+(\tau)$	0.9314	0.9404	0.9485	0.9510	0.9514	0.9510	0.9485	0.9414	0.9290
$\hat{\beta}_2^+(\tau)$	0.9253	0.9407	0.9471	0.9481	0.9515	0.9525	0.9458	0.9424	0.9283
					$\sigma_{uv} = 0.9$				
$\hat{\beta}_1(\tau)$	0.8490	0.8442	0.8469	0.8433	0.8389	0.8415	0.8427	0.8385	0.8457
$\hat{\beta}_2(\tau)$	0.8442	0.8460	0.8467	0.8422	0.8361	0.8396	0.8418	0.8431	0.8434
$\hat{\beta}_1^+(\tau)$	0.9281	0.9476	0.9552	0.9595	0.9625	0.9602	0.9577	0.9462	0.9300
$\hat{\beta}_2^+(\tau)$	0.9263	0.9441	0.9549	0.9595	0.9617	0.9593	0.9575	0.9499	0.9316

Notes: Tables 2.5 and 2.6 report the finite sample coverage rates of the 95% confidence intervals for the slope coefficients $\beta_j^0(\tau)$ for $j = 1, 2$, when the break date is known. Let $\hat{\beta}_j(\tau)$ denote the case in which the confidence interval is calculated using the original quantile estimator, and $\hat{\beta}_j^+(\tau)$ denotes the case in which the confidence interval is calculated using the fully modified estimator, where $j = 1, 2$.

Tables 2.7 and 2.8 report the coverage rates of the 95% confidence intervals for the break fraction and slope coefficients in the case of unknown break date. The results for the regression coefficients are similar to those from the known break date case. When $\sigma_{uv} = 0$ the confidence intervals calculated using the original quantile estimator have appropriate coverage rates. When σ_{uv} is nonzero the fully modified estimator should be used so that the coverage rates are adequate, otherwise the confidence intervals would have poor coverage properties due to the bias under long run endogeneity.

Table 2.7: Coverage rates for λ_1^0 and $\beta_j^0(\tau)$ with unknown T_1^0 ($T = 120$, $\lambda_1^0 = 0.50$)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$				
$\hat{\lambda}_1$	0.9488	0.9634	0.9686	0.9731	0.9733	0.9730	0.9684	0.9621	0.9483
$\hat{\lambda}_1^+$	0.9447	0.9598	0.9652	0.9695	0.9688	0.9700	0.9647	0.9593	0.9436
$\hat{\beta}_1(\tau)$	0.8978	0.9252	0.9439	0.9463	0.9507	0.9469	0.9420	0.9352	0.9027
$\hat{\beta}_2(\tau)$	0.9020	0.9321	0.9454	0.9510	0.9512	0.9527	0.9414	0.9297	0.8949
$\hat{\beta}_1^+(\tau)$	0.8967	0.9192	0.9341	0.9365	0.9404	0.9372	0.9313	0.9237	0.9010
$\hat{\beta}_2^+(\tau)$	0.9045	0.9265	0.9357	0.9430	0.9430	0.9382	0.9354	0.9222	0.9013
					$\sigma_{uv} = 0.5$				
$\hat{\lambda}_1$	0.9733	0.9830	0.9863	0.9884	0.9889	0.9902	0.9882	0.9829	0.9743
$\hat{\lambda}_1^+$	0.9683	0.9775	0.9827	0.9845	0.9841	0.9867	0.9832	0.9773	0.9690
$\hat{\beta}_1(\tau)$	0.8755	0.9023	0.9116	0.9146	0.9190	0.9142	0.9079	0.9029	0.8747
$\hat{\beta}_2(\tau)$	0.8751	0.8999	0.9077	0.9111	0.9096	0.9134	0.9077	0.8966	0.8656
$\hat{\beta}_1^+(\tau)$	0.8963	0.9248	0.9378	0.9415	0.9411	0.9416	0.9369	0.9283	0.9027
$\hat{\beta}_2^+(\tau)$	0.9021	0.9251	0.9379	0.9377	0.9436	0.9390	0.9358	0.9242	0.8954
					$\sigma_{uv} = 0.9$				
$\hat{\lambda}_1$	0.9814	0.9880	0.9916	0.9933	0.9934	0.9933	0.9936	0.9919	0.9863
$\hat{\lambda}_1^+$	0.9747	0.9808	0.9845	0.9843	0.9852	0.9843	0.9853	0.9833	0.9791
$\hat{\beta}_1(\tau)$	0.8281	0.8391	0.8394	0.8365	0.8364	0.8391	0.8342	0.8339	0.8217
$\hat{\beta}_2(\tau)$	0.8193	0.8302	0.8261	0.8259	0.8280	0.8341	0.8251	0.8216	0.8202
$\hat{\beta}_1^+(\tau)$	0.9022	0.9360	0.9497	0.9567	0.9597	0.9581	0.9493	0.9294	0.8986
$\hat{\beta}_2^+(\tau)$	0.8982	0.9308	0.9441	0.9500	0.9537	0.9462	0.9436	0.9280	0.8989

Table 2.8: Coverage rates for λ_1^0 and $\beta_j^0(\tau)$ with unknown T_1^0 ($T = 240$, $\lambda_1^0 = 0.50$)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$				
$\hat{\lambda}_1$	0.9706	0.9787	0.9833	0.9850	0.9850	0.9850	0.9839	0.9795	0.9687
$\hat{\lambda}_1^+$	0.9694	0.9772	0.9821	0.9837	0.9841	0.9838	0.9816	0.9779	0.9676
$\hat{\beta}_1(\tau)$	0.9126	0.9373	0.9492	0.9514	0.9579	0.9540	0.9477	0.9394	0.9207
$\hat{\beta}_2(\tau)$	0.9134	0.9409	0.9486	0.9497	0.9542	0.9535	0.9504	0.9420	0.9137
$\hat{\beta}_1^+(\tau)$	0.9260	0.9338	0.9432	0.9458	0.9508	0.9461	0.9446	0.9373	0.9275
$\hat{\beta}_2^+(\tau)$	0.9216	0.9364	0.9436	0.9433	0.9478	0.9478	0.9432	0.9399	0.9224
					$\sigma_{uv} = 0.5$				
$\hat{\lambda}_1$	0.9830	0.9875	0.9910	0.9918	0.9942	0.9939	0.9919	0.9887	0.9832
$\hat{\lambda}_1^+$	0.9794	0.9853	0.9888	0.9895	0.9913	0.9916	0.9895	0.9861	0.9812
$\hat{\beta}_1(\tau)$	0.9010	0.9113	0.9158	0.9179	0.9208	0.9173	0.9155	0.9071	0.8962
$\hat{\beta}_2(\tau)$	0.8892	0.9124	0.9158	0.9187	0.9181	0.9205	0.9173	0.9087	0.8851
$\hat{\beta}_1^+(\tau)$	0.9294	0.9387	0.9454	0.9481	0.9484	0.9482	0.9461	0.9399	0.9294
$\hat{\beta}_2^+(\tau)$	0.9225	0.9387	0.9453	0.9452	0.9487	0.9492	0.9440	0.9390	0.9268
					$\sigma_{uv} = 0.9$				
$\hat{\lambda}_1$	0.9865	0.9920	0.9951	0.9960	0.9968	0.9968	0.9966	0.9952	0.9901
$\hat{\lambda}_1^+$	0.9810	0.9849	0.9869	0.9876	0.9895	0.9903	0.9908	0.9898	0.9867
$\hat{\beta}_1(\tau)$	0.8470	0.8442	0.8474	0.8430	0.8389	0.8408	0.8413	0.8380	0.8421
$\hat{\beta}_2(\tau)$	0.8385	0.8402	0.8414	0.8389	0.8332	0.8358	0.8378	0.8398	0.8390
$\hat{\beta}_1^+(\tau)$	0.9263	0.9492	0.9552	0.9583	0.9596	0.9589	0.9574	0.9453	0.9305
$\hat{\beta}_2^+(\tau)$	0.9230	0.9416	0.9502	0.9550	0.9570	0.9543	0.9544	0.9461	0.9264

Notes: Tables 2.7 and 2.8 report the finite sample coverage rates of the 95% confidence intervals for the break fraction λ_1^0 and the slope coefficients $\beta_j^0(\tau)$ for $j = 1, 2$, when the break date is unknown. Let $\hat{\lambda}_1$ and $\hat{\beta}_j(\tau)$ denote the case in which the confidence interval is calculated using the original quantile estimator, and $\hat{\lambda}_1^+$ and $\hat{\beta}_j^+(\tau)$ denote the case in which the confidence interval is calculated using the fully modified estimator, where $j = 1, 2$.

To examine the coverage properties for the break fraction λ_1^0, the confidence intervals are constructed as follows. The cumulative distribution function (CDF) of $s^* = \arg\max_s\{W(s) - |s|/2\}$, where $W(s)$ is a standard two-sided Brownian motion, is

$$G(s^*) = \begin{cases} -(2\pi)^{-\frac{1}{2}}\sqrt{|s^*|}e^{-\frac{|s^*|}{8}} + \frac{1}{2}(|s^*|+5)\Phi(-\frac{\sqrt{|s^*|}}{2}) - \frac{3}{2}e^{|s^*|}\Phi(-\frac{3\sqrt{|s^*|}}{2}) & s^* \leq 0 \\ 1 + (2\pi)^{-\frac{1}{2}}\sqrt{s^*}e^{-\frac{s^*}{8}} - \frac{1}{2}(s^*+5)\Phi(-\frac{\sqrt{s^*}}{2}) + \frac{3}{2}e^{s^*}\Phi(-\frac{3\sqrt{s^*}}{2}) & s^* > 0, \end{cases}$$

where $\Phi(\cdot)$ is the CDF of standard normal distribution and $G(s^*) = 1 - G(-s^*)$, see Bai (1997). The 97.5% quantile of the distribution of s^* is $C_{0.975} = 11$. Due to equation (2.15), when the regressors are strictly exogenous, the confidence interval for a single break point is symmetric[9] such that $[\hat{\lambda}_1 - [C_{0.975}/\hat{L}]/T, \hat{\lambda}_1 + [C_{0.975}/\hat{L}]/T]$, where $\hat{L} = (\frac{\pi_1}{\omega_\psi \eta_1})^2$ and $[C_{0.975}/\hat{L}]$ denotes the integer part of $C_{0.975}/\hat{L}$. When the regressors are not strictly exogenous, endogeneity is corrected by the fully modified estimator and the same calculation applies. In particular, when σ_{uv} is nonzero, instead of $\hat{\theta}_j(\tau)$ and $\hat{\omega}_\psi$, the fully modified estimates $\hat{\theta}_j^+(\tau)$ and $\hat{\omega}_{\psi.v}$ should be used to calculate the confidence interval.

The first row in each panel in Table 2.7 and Table 2.8 contains the coverage rates for the break fraction when the 95% confidence intervals are calculated using the original quantile estimator. In Table 2.7, except for the upper and lower tails in the case of strict exogeneity, all the entries in these rows are larger than 95%. When $\sigma_{uv} = 0$ the coverage rates are not too large. However, when σ_{uv} is nonzero, the confidence intervals are more conservative. For each quantile, the coverage rate increases when the residual cross correlation becomes larger. Moreover, the corresponding entries from Table 2.8 are larger. That is, the coverage rates exceed the 95% level even more with larger sample size. When $T = 240$ there is overcoverage for the break fraction for all σ_{uv} and τ considered. As mentioned before, the confidence interval calculated based on the CDF, $G(\cdot)$, is only valid under strict exogeneity.

[9]For simplicity, the symmetric confidence intervals are adopted, although the limit distribution is in general asymmetric and depends on long run endogeneity.

Thus, when σ_{uv} is nonzero, the fully modified estimator is used to calculate the confidence intervals for λ_1^0. The results are summarized in the second row in each panel in Tables 2.7 and 2.8. Again, except for the tails in the case with zero cross correlation, all the entries in Table 2.7 are larger than 95%. When $T = 240$ the coverage rates are still larger than those from the case with $T = 120$. However, after the modification, the confidence intervals become less conservative in all cases and the improvement is more significant when $\sigma_{uv} = 0.9$. The coverage properties of the confidence intervals for the break fraction are similar to those in the linear case studied by Kejriwal and Perron (2008). In their paper, the finite sample confidence interval for a single break is also found to be conservative, and the overcoverage problem is more severe when the sample size is larger.

In addition, the fully modified quantile estimator has decent finite sample performance in cointegrated models with unknown breaks that occur in various positions in the sample. For example, if the data is generated under $\lambda_1^0 = 0.25$, the simulation results are similar to those in the case of $\lambda_1^0 = 0.50$. Tables 2.9 and 2.10 report the bias and RMSE of $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ for $j = 1, 2$ with the data generated under $\lambda_1^0 = 0.25$. The bias and RMSE of the break fraction are small in all cases, which is similar to the results from Tables 2.3 and 2.4. The bias and RMSE of $\hat{\beta}_1(\tau)$ for the first regime are larger, since in this case the first regime only has one quarter of the observations from the whole sample. On the contrary, the bias and RMSE of $\hat{\beta}_2(\tau)$ are smaller, since the second regime contains more observations. When σ_{uv} is nonzero, the fully modified estimator, $\hat{\beta}_1^+(\tau)$ and $\hat{\beta}_2^+(\tau)$, effectively reduces the bias and RMSE. Tables 2.11 and 2.12 report the coverage rates of the 95% confidence intervals for λ_1^0 and $\beta_j^0(\tau)$ for $j = 1, 2$ using the same simulated data with $\lambda_1^0 = 0.25$. In most cases, the coverage rates for λ_1^0 and $\beta_1^0(\tau)$ are smaller than the corresponding entries in Tables 2.7 and 2.8, but the coverage rates for $\beta_2^0(\tau)$ are slightly larger.

Table 2.9: Bias and RMSE of $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ with unknown T_1^0 ($T = 120$, $\lambda_1^0 = 0.25$)

		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$					
$\hat{\lambda}_1$	Bias	0.0033	0.0018	0.0010	0.0008	0.0009	0.0007	0.0008	0.0016	0.0036
	RMSE	0.0352	0.0267	0.0229	0.0220	0.0222	0.0225	0.0221	0.0278	0.0405
$\hat{\beta}_1(\tau)$	Bias	-0.0028	-0.0012	-0.0013	-0.0012	-0.0000	0.0023	0.0017	0.0010	0.0044
	RMSE	0.2129	0.1735	0.1570	0.1511	0.1492	0.1498	0.1565	0.1707	0.2110
$\hat{\beta}_2(\tau)$	Bias	0.0007	0.0003	0.0010	0.0006	0.0004	0.0007	0.0002	0.0010	0.0015
	RMSE	0.0657	0.0530	0.0481	0.0468	0.0464	0.0471	0.0490	0.0544	0.0702
$\hat{\beta}_1^+(\tau)$	Bias	-0.0032	-0.0016	-0.0014	-0.0015	0.0003	0.0028	0.0020	0.0013	0.0039
	RMSE	0.2202	0.1801	0.1625	0.1574	0.1548	0.1552	0.1617	0.1756	0.2168
$\hat{\beta}_2^+(\tau)$	Bias	0.0005	0.0001	0.0009	0.0005	0.0003	0.0006	0.0002	0.0010	0.0014
	RMSE	0.0680	0.0549	0.0501	0.0484	0.0481	0.0488	0.0508	0.0562	0.0729
					$\sigma_{uv} = 0.5$					
$\hat{\lambda}_1$	Bias	0.0029	0.0001	-0.0013	-0.0013	-0.0010	-0.0010	-0.0007	-0.0005	0.0021
	RMSE	0.0446	0.0328	0.0237	0.0236	0.0259	0.0260	0.0263	0.0296	0.0414
$\hat{\beta}_1(\tau)$	Bias	0.0841	0.0825	0.0814	0.0820	0.0829	0.0835	0.0841	0.0842	0.0870
	RMSE	0.2370	0.1957	0.1780	0.1722	0.1731	0.1767	0.1828	0.1951	0.2360
$\hat{\beta}_2(\tau)$	Bias	0.0307	0.0303	0.0304	0.0303	0.0302	0.0303	0.0304	0.0309	0.0313
	RMSE	0.0772	0.0664	0.0604	0.0591	0.0596	0.0589	0.0604	0.0647	0.0780
$\hat{\beta}_1^+(\tau)$	Bias	0.0247	0.0161	0.0105	0.0092	0.0102	0.0116	0.0146	0.0182	0.0272
	RMSE	0.2219	0.1725	0.1513	0.1433	0.1448	0.1477	0.1556	0.1723	0.2216
$\hat{\beta}_2^+(\tau)$	Bias	0.0088	0.0063	0.0051	0.0045	0.0041	0.0045	0.0054	0.0071	0.0091
	RMSE	0.0685	0.0564	0.0484	0.0463	0.0465	0.0462	0.0479	0.0536	0.0688
					$\sigma_{uv} = 0.9$					
$\hat{\lambda}_1$	Bias	0.0036	-0.0002	-0.0010	-0.0016	-0.0015	-0.0016	-0.0009	0.0001	0.0033
	RMSE	0.0568	0.0404	0.0351	0.0322	0.0319	0.0302	0.0340	0.0389	0.0532
$\hat{\beta}_1(\tau)$	Bias	0.1597	0.1540	0.1527	0.1530	0.1530	0.1520	0.1523	0.1536	0.1601
	RMSE	0.2851	0.2477	0.2337	0.2294	0.2282	0.2276	0.2331	0.2457	0.2867
$\hat{\beta}_2(\tau)$	Bias	0.0553	0.0539	0.0537	0.0539	0.0536	0.0536	0.0539	0.0546	0.0551
	RMSE	0.0981	0.0839	0.0804	0.0784	0.0783	0.0783	0.0800	0.0843	0.0972
$\hat{\beta}_1^+(\tau)$	Bias	0.0525	0.0333	0.0247	0.0213	0.0206	0.0208	0.0255	0.0344	0.0536
	RMSE	0.2232	0.1656	0.1416	0.1305	0.1288	0.1292	0.1392	0.1646	0.2274
$\hat{\beta}_2^+(\tau)$	Bias	0.0156	0.0104	0.0082	0.0072	0.0066	0.0069	0.0082	0.0112	0.0149
	RMSE	0.0705	0.0483	0.0424	0.0383	0.0378	0.0383	0.0412	0.0488	0.0681

Table 2.10: Bias and RMSE of $\hat{\lambda}_1$, $\hat{\beta}_j(\tau)$, and $\hat{\beta}_j^+(\tau)$ with unknown T_1^0 ($T = 240$, $\lambda_1^0 = 0.25$)

		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$					
$\hat{\lambda}_1$	Bias	0.0002	0.0000	0.0000	0.0001	0.0001	0.0001	0.0000	0.0001	0.0000
	RMSE	0.0111	0.0089	0.0082	0.0077	0.0077	0.0076	0.0080	0.0085	0.0105
$\hat{\beta}_1(\tau)$	Bias	-0.0014	-0.0005	-0.0001	0.0000	-0.0000	0.0008	0.0004	0.0009	-0.0004
	RMSE	0.0980	0.0803	0.0740	0.0712	0.0689	0.0705	0.0740	0.0814	0.0977
$\hat{\beta}_2(\tau)$	Bias	-0.0003	-0.0005	-0.0004	-0.0002	-0.0001	0.0000	-0.0001	0.0001	0.0003
	RMSE	0.0316	0.0261	0.0239	0.0233	0.0230	0.0231	0.0241	0.0261	0.0309
$\hat{\beta}_1^+(\tau)$	Bias	-0.0013	-0.0003	-0.0001	-0.0000	-0.0000	0.0007	0.0003	0.0007	-0.0005
	RMSE	0.0998	0.0821	0.0758	0.0726	0.0703	0.0719	0.0754	0.0831	0.0992
$\hat{\beta}_2^+(\tau)$	Bias	-0.0003	-0.0004	-0.0005	-0.0003	-0.0001	-0.0000	-0.0002	0.0001	0.0003
	RMSE	0.0321	0.0264	0.0243	0.0237	0.0235	0.0236	0.0246	0.0265	0.0314
					$\sigma_{uv} = 0.5$					
$\hat{\lambda}_1$	Bias	-0.0010	-0.0009	-0.0009	-0.0009	-0.0009	-0.0008	-0.0010	-0.0010	-0.0011
	RMSE	0.0107	0.0089	0.0080	0.0078	0.0079	0.0078	0.0084	0.0091	0.0114
$\hat{\beta}_1(\tau)$	Bias	0.0418	0.0413	0.0421	0.0422	0.0427	0.0438	0.0434	0.0434	0.0433
	RMSE	0.1076	0.0912	0.0864	0.0846	0.0844	0.0855	0.0882	0.0936	0.1080
$\hat{\beta}_2(\tau)$	Bias	0.0144	0.0147	0.0147	0.0150	0.0150	0.0151	0.0149	0.0149	0.0149
	RMSE	0.0360	0.0315	0.0296	0.0289	0.0286	0.0287	0.0294	0.0310	0.0358
$\hat{\beta}_1^+(\tau)$	Bias	0.0047	0.0020	0.0013	0.0007	0.0010	0.0022	0.0025	0.0041	0.0063
	RMSE	0.0956	0.0761	0.0699	0.0666	0.0658	0.0658	0.0701	0.0772	0.0949
$\hat{\beta}_2^+(\tau)$	Bias	0.0018	0.0014	0.0009	0.0010	0.0009	0.0011	0.0011	0.0015	0.0022
	RMSE	0.0314	0.0257	0.0232	0.0217	0.0214	0.0214	0.0223	0.0247	0.0307
					$\sigma_{uv} = 0.9$					
$\hat{\lambda}_1$	Bias	-0.0018	-0.0016	-0.0015	-0.0015	-0.0015	-0.0015	-0.0014	-0.0016	-0.0016
	RMSE	0.0113	0.0099	0.0088	0.0088	0.0088	0.0088	0.0088	0.0097	0.0120
$\hat{\beta}_1(\tau)$	Bias	0.0753	0.0761	0.0769	0.0773	0.0773	0.0772	0.0778	0.0781	0.0776
	RMSE	0.1268	0.1156	0.1130	0.1118	0.1109	0.1110	0.1133	0.1178	0.1298
$\hat{\beta}_2(\tau)$	Bias	0.0265	0.0269	0.0269	0.0270	0.0269	0.0270	0.0270	0.0270	0.0264
	RMSE	0.0442	0.0407	0.0394	0.0389	0.0387	0.0388	0.0394	0.0408	0.0436
$\hat{\beta}_1^+(\tau)$	Bias	0.0091	0.0053	0.0034	0.0025	0.0022	0.0026	0.0046	0.0076	0.0117
	RMSE	0.0846	0.0629	0.0549	0.0511	0.0490	0.0504	0.0544	0.0642	0.0864
$\hat{\beta}_2^+(\tau)$	Bias	0.0035	0.0024	0.0016	0.0013	0.0012	0.0014	0.0018	0.0026	0.0033
	RMSE	0.0286	0.0215	0.0187	0.0172	0.0172	0.0173	0.0186	0.0217	0.0281

Notes: Same as Tables 2.3 and 2.4. For Tables 2.9 and 2.10, data is generated under $\lambda_1^0 = 0.25$.

Table 2.11: Coverage rates for λ_1^0 and $\beta_j^0(\tau)$ with unknown T_1^0 ($T = 120$, $\lambda_1^0 = 0.25$)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$				
$\hat{\lambda}_1$	0.9355	0.9539	0.9606	0.9636	0.9646	0.9630	0.9613	0.9519	0.9344
$\hat{\lambda}_1^+$	0.9321	0.9490	0.9570	0.9595	0.9608	0.9596	0.9577	0.9476	0.9303
$\hat{\beta}_1(\tau)$	0.8668	0.9065	0.9244	0.9343	0.9382	0.9370	0.9302	0.9192	0.8722
$\hat{\beta}_2(\tau)$	0.9031	0.9334	0.9462	0.9497	0.9524	0.9515	0.9455	0.9304	0.9021
$\hat{\beta}_1^+(\tau)$	0.8790	0.9052	0.9214	0.9289	0.9326	0.9281	0.9246	0.9138	0.8857
$\hat{\beta}_2^+(\tau)$	0.9004	0.9239	0.9367	0.9412	0.9413	0.9404	0.9375	0.9196	0.8992
					$\sigma_{uv} = 0.5$				
$\hat{\lambda}_1$	0.9640	0.9780	0.9839	0.9859	0.9863	0.9845	0.9845	0.9798	0.9699
$\hat{\lambda}_1^+$	0.9545	0.9694	0.9758	0.9791	0.9789	0.9772	0.9769	0.9697	0.9621
$\hat{\beta}_1(\tau)$	0.8440	0.8832	0.8974	0.9040	0.9060	0.9040	0.9011	0.8875	0.8516
$\hat{\beta}_2(\tau)$	0.8752	0.8974	0.9060	0.9091	0.9120	0.9122	0.9097	0.8976	0.8757
$\hat{\beta}_1^+(\tau)$	0.8812	0.9135	0.9348	0.9394	0.9425	0.9380	0.9327	0.9166	0.8833
$\hat{\beta}_2^+(\tau)$	0.8996	0.9247	0.9315	0.9396	0.9429	0.9401	0.9383	0.9253	0.8970
					$\sigma_{uv} = 0.9$				
$\hat{\lambda}_1$	0.9702	0.9831	0.9881	0.9911	0.9907	0.9912	0.9892	0.9858	0.9759
$\hat{\lambda}_1^+$	0.9616	0.9724	0.9760	0.9782	0.9793	0.9815	0.9796	0.9778	0.9671
$\hat{\beta}_1(\tau)$	0.7961	0.8138	0.8167	0.8182	0.8179	0.8163	0.8124	0.8039	0.7933
$\hat{\beta}_2(\tau)$	0.8195	0.8251	0.8241	0.8248	0.8296	0.8306	0.8266	0.8286	0.8224
$\hat{\beta}_1^+(\tau)$	0.8772	0.9273	0.9469	0.9525	0.9533	0.9515	0.9451	0.9219	0.8777
$\hat{\beta}_2^+(\tau)$	0.8926	0.9329	0.9401	0.9529	0.9507	0.9507	0.9441	0.9320	0.9004

Table 2.12: Coverage rates for λ_1^0 and $\beta_j^0(\tau)$ with unknown T_1^0 ($T = 240$, $\lambda_1^0 = 0.25$)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
					$\sigma_{uv} = 0$				
$\hat{\lambda}_1$	0.9566	0.9689	0.9749	0.9793	0.9774	0.9771	0.9737	0.9688	0.9581
$\hat{\lambda}_1^+$	0.9552	0.9670	0.9728	0.9765	0.9762	0.9754	0.9718	0.9678	0.9557
$\hat{\beta}_1(\tau)$	0.9066	0.9298	0.9451	0.9470	0.9523	0.9523	0.9450	0.9320	0.9033
$\hat{\beta}_2(\tau)$	0.9174	0.9404	0.9491	0.9515	0.9539	0.9549	0.9514	0.9434	0.9235
$\hat{\beta}_1^+(\tau)$	0.9190	0.9360	0.9467	0.9434	0.9493	0.9490	0.9445	0.9364	0.9221
$\hat{\beta}_2^+(\tau)$	0.9189	0.9353	0.9456	0.9466	0.9463	0.9457	0.9463	0.9406	0.9281
					$\sigma_{uv} = 0.5$				
$\hat{\lambda}_1$	0.9763	0.9854	0.9878	0.9906	0.9908	0.9925	0.9900	0.9859	0.9785
$\hat{\lambda}_1^+$	0.9710	0.9800	0.9830	0.9859	0.9870	0.9882	0.9851	0.9804	0.9724
$\hat{\beta}_1(\tau)$	0.8807	0.9052	0.9129	0.9179	0.9212	0.9179	0.9088	0.9030	0.8802
$\hat{\beta}_2(\tau)$	0.8880	0.9084	0.9175	0.9194	0.9171	0.9217	0.9150	0.9125	0.8988
$\hat{\beta}_1^+(\tau)$	0.9211	0.9425	0.9489	0.9487	0.9503	0.9524	0.9473	0.9360	0.9201
$\hat{\beta}_2^+(\tau)$	0.9183	0.9355	0.9450	0.9496	0.9542	0.9526	0.9512	0.9412	0.9268
					$\sigma_{uv} = 0.9$				
$\hat{\lambda}_1$	0.9881	0.9917	0.9943	0.9951	0.9947	0.9944	0.9953	0.9938	0.9909
$\hat{\lambda}_1^+$	0.9791	0.9808	0.9837	0.9839	0.9830	0.9835	0.9853	0.9839	0.9827
$\hat{\beta}_1(\tau)$	0.8300	0.8398	0.8376	0.8382	0.8390	0.8375	0.8350	0.8316	0.8230
$\hat{\beta}_2(\tau)$	0.8373	0.8417	0.8423	0.8397	0.8363	0.8367	0.8381	0.8409	0.8476
$\hat{\beta}_1^+(\tau)$	0.9229	0.9449	0.9577	0.9604	0.9621	0.9619	0.9570	0.9460	0.9187
$\hat{\beta}_2^+(\tau)$	0.9225	0.9403	0.9536	0.9588	0.9574	0.9575	0.9527	0.9407	0.9311

Notes: Same as Tables 2.7 and 2.8. For Tables 2.11 and 2.12, data is generated under $\lambda_1^0 = 0.25$.

To summarize, with strictly exogenous regressors, the original quantile estimator is unbiased and the confidence intervals have good coverage properties. When the regressors are not strictly exogenous, it is necessary to employ the fully modified estimator to correct serial correlation and endogeneity so that the bias and RMSE can be reduced and the confidence intervals can have adequate coverage rates in finite samples.

2.4 Empirical Study

The quantile cointegrating regression model with parameter instability is empirically relevant. It is widely accepted that stock prices follow I(1) processes. There may exist long run equilibrium relationships among stock indices from different economies due to market integration. Furthermore, the tails of the distribution of these integrated time series are often of particular interest. Hence, it is useful to estimate the cointegration relationships for various quantiles of the conditional distribution of the stock indices. In particular, this paper estimates the quantile cointegrating vectors between stock prices from different countries allowing for a regime shift, which reflect the degree of international market integration.

In this application, the main focus is the relationship between the stock indices from the emerging markets of China (SSE and SZSE) and those from the advanced markets such as the United States (SP500) and Japan (NIKKEI). This study also considers stock indices from several mature markets on the Asian Pacific rim, including Hong Kong (HANGSENG), Korea (KOSPI), Singapore (STRAITS), Australia (ASX), and New Zealand (NZAO)[10]. The data set comprises daily observations from August 31, 1999 to July 16, 2010, which is the same time span and frequency as in Burdekin and Siklos (2011). The number of observations is 2839. Following the convention in the previous literature, data of the weekends are

[10]SSE denotes the Shanghai Stock Exchange (SSE) composite index and SZSE denotes the Shenzhen Stock Exchange (SZSE) component index in China. SP500 is the S&P 500 index for the U.S. market, NIKKEI is Nikkei 225 for the Japanese market, HANGSENG is the main Hong Kong index, KOSPI is the Korean market index, STRAITS is the Singapore Straits Times index, ASX is the Australian market index, and NZAO is the New Zealand All Ordinaries index.

omitted and missing observations for holidays are filled with the previous day's data.

Note that, as in the quantile cointegrating regression setting from Xiao (2009), the elements of the I(1) vector x_t are assumed not to be cointegrated, otherwise the long run covariance matrix from Assumption 1 will be singular. To avoid this singularity, the scalar x_t case is considered here. The bivariate quantile regression model of interest is $log(Y_t) = \alpha_j(\tau) + \beta_j(\tau)log(X_t) + u_t(\tau)$ for $j = 1, 2$, where Y_t is the stock index from an emerging market and X_t is the stock index from a mature market. The model is estimated under one structural change. The methodology used in this application is similar to that in Voronkova (2004), which adopts a bivariate Phillips-Hansen type fully modified regression with one regime shift in the linear case.

As a pretest to investigate the properties of the stock index series, the augmented Dickey-Fuller (ADF) test does not reject that the logarithm of each stock index has a unit root[11], but rejects a unit root in the first difference of the series at 5% level. Thus, the logarithm of the stock index is I(1). In general, for the linear case, the Johansen cointegration test often finds no cointegration between the emerging stock markets of China and each one of the mature markets if structural changes are not incorporated. Without considering shifts in the parameters, the Engle-Granger Dickey-Fuller type test also finds that the estimated residuals from the least squares estimation are nonstationary. In particular, four lags[12] are included for the Dicky-Fuller test to account for serial correlation in the residuals and the test fails to reject the null of no cointegration at 5% level using the critical values from MacKinnon (2010). Likewise, using quantile regressions with no structural changes, the

[11]The ADF regression corresponds to equation (2.1). For most series considered, the lag length selected based on the Bayes information criterion (BIC) is one, then Δx_{t-1} needs not to be included in equation (2.1). In this case, a Dicky-Fuller test would be sufficient.

[12]The results appear to be insensitive to the number of lags included.

Engle-Granger tests also retain the null of no cointegration for all quantiles. Thus, assuming parameter stability, there is no evidence that the Chinese stock markets are integrated with any of the mature market considered.

Tables 2.13 and 2.14 summarize the results for the Shanghai and Shenzhen markets respectively when the bivariate model is estimated allowing for one regime shift. The estimated break dates $\hat{T}_1$ are reported along with their 95% confidence intervals $[\hat{T}_{1L}, \hat{T}_{1U}]$. The cointegrating coefficients for eleven quantiles, $\tau = 0.1, ..., 0.9$, are estimated using the fully modified estimator $\hat{\beta}_j^+(\tau)$ for $j = 1, 2$. For comparison, the last columns in the tables contain the slope coefficients $\hat{\beta}_j^+$ for $j = 1, 2$, which are from the conditional mean regressions.

In Table 2.13, for the Shanghai market, most estimated break dates are in late 2006 and early 2007. The breaks are likely due to the events that happened in the Chinese stock markets during that period. After the bearish performance in the late 1990s and the first six years of the 21st century, a bullish market emerged in China in 2006[13]. In May 2006 the ban on new IPOs in China, which was put in April 2005, was lifted. In September 2006 China's biggest mainland bank, the Industrial and Commercial Bank of China (ICBC), was launched in the Shanghai and Hong Kong stock markets[14]. It was the world's biggest IPO until 2010 when the Agricultural Bank of China completed the world's largest initial stock listing to date[15]. From the estimation results, the break happened on December 13, 2006 in many cases. In fact, on December 14, 2006 the Shanghai composite index reached its highest point in the history of the Shanghai stock market by then[16].

---

[13]Source: Yao, Shujie and Luo, Dan (2008). Chinese stock market bubble: Inevitable or incidental? Briefing Series 41, China Policy Institute, The University of Nottingham.

[14]Source: "China's ICBC: The world's largest IPO ever". BusinessWeek, 27 September 2006.

[15]Source: "AgBank IPO officially the world's biggest". Financial Times, 13 August 2010.

[16]Source: "Ten biggest news of the Chinese stock market in 2006". http://finance.sina.com.cn/stock/t/20061226/04031121473.shtml, 26 December 2006.

From the least squares estimation results, the Shanghai composite index is not cointegrated with NIKKEI or HANGSENG in either regime, since the cointegrating coefficients $\hat{\beta}_j^+$ for $j = 1, 2$ are insignificant at 5% level. The coefficient $\hat{\beta}_2^+$ is only significant in the second regime with SP500. On the contrary, the conditional mean of the Shanghai stock index is cointegrated with KOSPI, STRAITS, ASX, and NZAO in both regimes. When the analysis is extended to the conditional quantiles, with SP500 and NIKKEI the coefficients $\hat{\beta}_j^+(\tau)$ for $j = 1, 2$ are significant in both regimes only in the 70% and 90% quantiles. The Shanghai stock index is cointegrated with SP500 and HANGSENG in the second regime for all quantiles. Also, with NIKKEI the coefficient $\hat{\beta}_2^+(\tau)$ is significant in the second regime from the 30% to 90% quantiles. However, there is no cointegration relationship between SSE and NIKKEI in the lower tail. The Shanghai market seems well integrated with the Korean and New Zealand markets as indicated by the significant cointegrating coefficients in both regimes not only for the mean but for all quantiles as well. In addition, there are long run equilibrium relationships between SSE and STRAITS in both regimes for the 10% quantile and in the second regime for the entire distribution except for the 80% quantile. There is also co-movement between SSE and ASX in both regimes for several lower quantiles and the upper tail and the coefficient $\hat{\beta}_2^+(\tau)$ in the second regime is also significant from the median up to the 80% quantile. In particular, with KOSPI, ASX, and NZAO the slope coefficients are negative in the first regime and positive in the second regime. In most of the other cases, the coefficients are positive.

In general, without considering structural changes, both linear and quantile regression models give the misleading result that long run equilibria do not exist between the Shanghai stock market and the mature markets. Contrarily, there is evidence of cointegration with a structural change. It is found that the Shanghai market is more integrated with the Korean, Singapore, Australian, and New Zealand markets than with the advanced markets of the U.S. and Japan. The results of cointegration are more robust for certain quantiles than for the mean. Furthermore, co-movement of the stock indices is found more often in the second regime than in the first regime. The long run equilibrium relationships between the stock markets are clearly asymmetric in the distribution and evidence of cointegration is more significant for the upper quantiles than for the lower quantiles.

For the other stock market in China, Shenzhen, as shown in Table 2.14, the estimated break dates also range from late 2006 to early 2007 in most cases, which are similar to those for the Shanghai market. The Shenzhen component index is not cointegrated with SP500, NIKKEI, or HANGSENG in the mean. The conditional mean of SZSE is cointegrated with KOSPI in both regimes, while it is cointegrated with STRAITS, ASX, and NZAO only in the first regime. For the conditional quantiles, the Shenzhen stock index is cointegrated with SP500 and NIKKEI in both regimes in the upper tail. However, with SP500 and NIKKEI in neither regimes is the slope coefficient significant for the lower quantiles. Also, there are cointegration relationships between SZSE and HANGSENG only in the second regime for all quantiles except for the 70% and 80% quantiles where the coefficients are significant in both regimes. Similar to the Shanghai market, the Shenzhen market is well integrated with the Korean market for the whole distribution. In contrast, with STRAITS the coefficients are significant in both regimes only in the lower tail and cointegration relationships exist only in the second regime for most of the other quantiles. More interestingly, the Australian stock index has no impact on the central part of the distribution of SZSE, however it affects the tails of SZSE in both regimes. Also, there are long run equilibrium relationships between SZSE and NZAO in the first regime for all quantiles and the slope coefficients are significant in both regimes for the lower and upper quantiles.

Table 2.13: Quantile cointegration: Shanghai composite index and mature markets

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
					U.S. (SP500)					
$\hat{T}_1$	12/13/06	12/13/06	1/23/07	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06
$\hat{T}_{1L}$	11/11/05	01/20/06	8/03/06	01/18/06	04/03/06	08/14/06	09/19/06	09/18/06	09/08/2006	11/27/06
$\hat{T}_{1U}$	01/14/08	11/05/07	07/13/07	11/07/07	08/24/07	04/13/07	03/08/07	03/09/07	03/19/07	12/29/06
$\hat{\beta}_1^+(\tau)$	-0.4295	-0.1755	0.0928	0.1907	0.3795	0.4021	0.4803*	0.4814	0.5957*	-0.1704
$\hat{\beta}_2^+(\tau)$	0.8935*	0.9337*	1.1464*	0.9382*	0.8890*	0.9361*	1.1372*	1.2868*	1.2826*	0.8182*
					Japan (NIKKEI)					
$\hat{T}_1$	12/13/06	03/02/07	03/02/07	01/23/07	12/13/06	12/13/06	01/23/07	12/13/06	12/13/06	12/13/06
$\hat{T}_{1L}$	09/08/06	11/14/06	11/20/06	09/01/06	09/28/06	10/02/06	12/06/06	08/31/06	09/29/06	11/06/06
$\hat{T}_{1U}$	03/19/07	06/20/07	06/14/07	06/14/07	02/27/07	02/23/07	03/12/07	03/27/07	02/26/07	01/19/07
$\hat{\beta}_1^+(\tau)$	-0.2143	-0.1163	0.0240	0.0952	0.1436	0.1949	0.3148*	0.3508	0.4315*	-0.1933
$\hat{\beta}_2^+(\tau)$	0.4994	0.7617	0.7186*	0.6936*	0.6951*	0.8092*	1.1436*	1.1237*	1.0977*	0.4777
					Hong Kong (HANGSENG)					
$\hat{T}_1$	11/18/04	11/18/04	10/08/04	10/08/04	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06
$\hat{T}_{1L}$	08/16/04	04/12/04	06/30/04	07/23/04	05/31/06	05/10/06	04/27/06	03/01/06	04/14/2006	07/18/06
$\hat{T}_{1U}$	02/22/05	06/28/05	01/18/05	12/24/04	06/27/07	07/18/07	07/31/07	09/26/07	08/13/2007	05/10/07
$\hat{\beta}_1^+(\tau)$	0.0307	0.0470	0.2378	0.3209	0.1676	0.2224	0.3065	0.3603	0.3541	-0.4266
$\hat{\beta}_2^+(\tau)$	2.2420*	2.2830*	2.2738*	2.1731*	1.0887*	1.2764*	1.2506*	1.2634*	1.2642*	0.7308
					Korea (KOSPI)					
$\hat{T}_1$	05/29/06	11/03/06	11/03/06	11/03/06	11/03/06	09/25/06	05/29/06	05/29/06	04/19/06	09/25/06
$\hat{T}_{1L}$	06/28/05	10/04/06	09/14/06	12/16/05	-	01/13/05	12/19/03	08/10/04	01/11/06	07/05/06
$\hat{T}_{1U}$	04/27/07	12/05/06	12/25/06	09/21/07	-	06/04/08	11/04/08	03/14/08	07/26/06	12/14/06
$\hat{\beta}_1^+(\tau)$	-0.5000*	-0.4318*	-0.4533*	-0.4005*	-0.4302*	-0.5189*	-0.5577*	-0.5242*	-0.5961*	-0.5272*
$\hat{\beta}_2^+(\tau)$	1.9749*	1.2657*	1.2489*	1.1667*	1.1078*	1.2827*	1.6576*	1.6160*	1.5914*	1.2269*

(Table 2.13 cont'd)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
					Singapore (STRAITS)					
$\hat{T}_1$	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06
$\hat{T}_{1L}$	10/17/06	10/23/06	10/16/06	10/19/06	08/14/06	05/19/06	05/11/05	09/02/04	09/20/06	11/17/06
$\hat{T}_{1U}$	02/08/07	02/02/07	02/09/07	02/06/07	04/13/07	07/09/07	07/16/08	03/24/09	03/07/07	01/08/07
$\hat{\beta}_1^+(\tau)$	-0.4130*	-0.3175	-0.2622	-0.1118	0.0246	0.0888	0.1263	0.1611	0.0381	-0.5965*
$\hat{\beta}_2^+(\tau)$	0.8879*	0.8947*	0.9436*	0.9524*	0.9748*	0.9687*	0.9741*	1.0897	1.1532*	0.7229*
					Australia (ASX)					
$\hat{T}_1$	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	11/03/06	11/03/06	12/13/06
$\hat{T}_{1L}$	10/31/06	11/17/06	11/17/06	11/08/06	10/26/06	10/10/06	09/22/06	08/09/06	10/06/06	10/31/06
$\hat{T}_{1U}$	01/25/07	01/08/07	01/08/07	01/17/07	01/30/07	02/15/07	03/05/07	01/30/07	12/01/06	01/25/07
$\hat{\beta}_1^+(\tau)$	-0.5816*	-0.5592*	-0.5220*	-0.4283*	-0.3741	-0.0877	-0.0526	-0.2622	-0.3346*	-0.5126*
$\hat{\beta}_2^+(\tau)$	1.0424*	1.0206*	1.1485*	1.1436*	1.0540*	1.1102*	1.3494*	1.4131*	1.3556*	1.0476*
					New Zealand (NZAO)					
$\hat{T}_1$	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	11/03/06	11/03/06	11/03/06	12/13/06
$\hat{T}_{1L}$	11/06/06	11/07/06	11/22/06	11/24/06	11/22/06	11/09/06	10/06/06	10/10/06	10/10/06	11/30/06
$\hat{T}_{1U}$	01/19/07	01/18/07	01/03/07	01/01/07	01/03/07	01/16/07	12/01/06	11/29/06	11/29/06	12/26/06
$\hat{\beta}_1^+(\tau)$	-0.9125*	-0.7794*	-0.7916*	-0.8104*	-0.8338*	-0.8877*	-1.0119*	-0.7395*	-0.6318*	-0.8674*
$\hat{\beta}_2^+(\tau)$	0.7404*	0.7347*	0.8652*	0.8696*	0.9199*	1.1850*	1.3308*	1.3642*	1.2943*	0.9173*

Table 2.14: Quantile cointegration: Shenzhen component index and mature markets

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
					U.S. (SP500)					
$\hat{T}_1$	01/23/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07
$\hat{T}_{1L}$	01/20/06	04/08/05	05/26/06	10/04/06	12/07/06	11/29/06	11/13/06	12/26/06	12/06/06	02/16/07
$\hat{T}_{1U}$	01/24/08	01/23/09	12/07/07	07/31/07	05/28/07	06/05/07	06/21/07	05/09/07	05/29/07	03/16/07
$\hat{\beta}_1^+(\tau)$	-0.8770	-0.5101	0.0498	0.1766	0.1856	0.4648	0.5912*	0.6279*	0.7073*	-0.4778
$\hat{\beta}_2^+(\tau)$	0.7832	1.1264	0.7850	0.7347	0.8080	0.8745*	0.9381*	0.8653*	0.8102*	0.4561
					Japan (NIKKEI)					
$\hat{T}_1$	04/11/07	04/11/07	04/11/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07	03/02/07
$\hat{T}_{1L}$	11/30/06	10/24/06	12/20/06	11/27/06	11/01/06	09/07/06	08/08/06	11/03/06	11/20/06	01/29/07
$\hat{T}_{1U}$	08/21/07	09/27/07	08/01/07	06/07/07	07/03/07	08/27/07	09/26/07	06/29/07	06/14/07	04/05/07
$\hat{\beta}_1^+(\tau)$	-0.1831	-0.2141	0.0851	0.0348	0.0200	0.1905	0.3107	0.4668*	0.5973*	-0.3182
$\hat{\beta}_2^+(\tau)$	0.8320	0.6852	0.5486	0.4891	0.5309	0.5588	0.5377	0.5234*	0.5136*	0.2267
					Hong Kong (HANGSENG)					
$\hat{T}_1$	07/22/04	07/22/04	01/23/07	01/23/07	01/23/07	01/23/07	03/02/07	03/02/07	03/02/07	01/23/07
$\hat{T}_{1L}$	04/09/04	02/25/04	10/04/06	10/12/06	11/06/06	10/19/06	09/28/06	08/30/06	11/07/05	07/12/06
$\hat{T}_{1U}$	11/03/04	12/17/04	05/14/07	05/04/07	04/11/07	04/27/07	08/06/07	09/04/07	06/26/08	08/06/07
$\hat{\beta}_1^+(\tau)$	0.0854	0.1611	-0.2893	-0.0210	0.0288	0.1041	0.4167*	0.4611*	0.5216	-0.7758
$\hat{\beta}_2^+(\tau)$	2.4635*	2.4827*	1.2390*	1.1125*	1.0263*	0.9847*	1.0283*	1.0218*	1.0693*	0.3208
					Korea (KOSPI)					
$\hat{T}_1$	01/23/07	01/23/07	12/13/06	12/13/06	12/13/06	12/13/06	12/13/06	11/03/06	11/03/06	12/13/06
$\hat{T}_{1L}$	12/26/06	12/14/06	08/22/06	09/20/99	03/17/06	03/23/04	09/04/06	08/07/06	10/19/05	08/18/06
$\hat{T}_{1U}$	02/20/07	03/02/07	04/05/07	-	09/10/07	09/03/09	03/23/07	02/01/07	11/20/07	04/09/07
$\hat{\beta}_1^+(\tau)$	-0.6841*	-0.6460*	-0.6598*	-0.5921*	-0.7269*	-0.5779*	-0.5346*	-0.5246*	-0.5110*	-0.6956*
$\hat{\beta}_2^+(\tau)$	1.6542*	1.4903*	1.4234*	1.3191*	1.1522*	1.3379*	1.3815*	1.3503*	1.3360*	1.1143*

(Table 2.14 cont'd)

	$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	Mean
					Singapore (STRAITS)					
$\hat{T}_1$	03/02/07	03/02/07	01/23/07	03/02/07	01/23/07	01/23/07	01/23/07	03/02/07	01/23/07	01/23/07
$\hat{T}_{1L}$	01/01/07	11/23/06	11/27/06	12/08/06	11/17/06	11/21/06	11/07/06	11/30/05	08/15/06	12/27/06
$\hat{T}_{1U}$	05/03/07	06/11/07	03/21/07	05/25/07	03/29/07	03/27/07	04/10/07	06/03/08	07/03/07	02/19/07
$\hat{\beta}_1^+(\tau)$	-0.6453*	-0.4970	-0.4549	-0.1058	-0.0702	-0.0900	-0.0525	0.3403	-0.0519	-0.8894*
$\hat{\beta}_2^+(\tau)$	1.1591*	1.0927*	0.9907*	1.0200*	0.8699*	0.8371*	0.8483*	0.8475	0.8975*	0.4908
					Australia (ASX)					
$\hat{T}_1$	01/23/07	03/02/07	01/23/07	01/23/07	01/23/07	01/23/07	12/13/06	12/13/06	12/13/06	01/23/07
$\hat{T}_{1L}$	11/27/06	02/06/07	12/21/06	11/10/06	11/02/06	12/01/06	10/12/06	11/03/06	11/21/06	08/21/06
$\hat{T}_{1U}$	03/21/07	03/28/07	02/23/07	04/05/07	04/13/07	03/15/07	02/13/07	01/22/07	01/04/07	06/27/07
$\hat{\beta}_1^+(\tau)$	-0.9545*	-0.8945*	-0.8531*	-0.7108	-0.5667	-0.3926	-0.4332	-0.6417*	-0.6778*	-0.8572*
$\hat{\beta}_2^+(\tau)$	0.7999*	1.1092*	0.9709*	0.8174	0.8264	0.9308	0.9635	0.8785*	0.7929*	0.5904
					New Zealand (NZAO)					
$\hat{T}_1$	01/23/07	03/02/07	01/23/07	01/23/07	12/13/06	12/13/06	12/13/06	12/13/06	11/03/06	12/13/06
$\hat{T}_{1L}$	11/07/06	02/13/07	01/05/07	01/04/07	11/23/06	11/07/06	11/13/06	11/21/06	10/23/06	11/29/06
$\hat{T}_{1U}$	04/10/07	03/21/07	02/08/07	02/09/07	01/02/07	01/18/07	01/12/07	01/04/07	11/16/06	12/27/06
$\hat{\beta}_1^+(\tau)$	-1.3689*	-1.2849*	-1.2669*	-1.3128*	-1.4133*	-1.5347*	-1.5334*	-1.1993*	-1.2128*	-1.5061*
$\hat{\beta}_2^+(\tau)$	0.5877	1.0346*	0.6503*	0.5546	0.4887	0.5216	0.5570	0.6071*	0.5715*	0.3516

Notes: For Tables 2.13 and 2.14, the bivariate quantile regression model is $log(Y_t) = \alpha_j(\tau) + \beta_j(\tau) log(X_t) + u_t(\tau)$ for $j = 1, 2$, where Y_t is the stock index from an emerging market, such as Shanghai composite index (SSE) and Shenzhen component index (SZSE) from China, and X_t is the stock index from a mature market. Let * denote significance at 5% level. The 95% confidence interval of the break date T_1, mm/dd/yy, is $[\hat{T}_{1L}, \hat{T}_{1U}]$. Data is from 08/31/99 to 07/16/10.

To sum up, in empirical cointegration analysis, it is appropriate to relax the assumption of parameter stability, because otherwise the power of the cointegration tests may be deteriorated. In the applications considered in this chapter, without considering structural changes the Engle-Granger tests tend to under-reject the null hypothesis of no cointegration. Assuming that the cointegrating vector is subject to one break, one can find more evidence in support of cointegration. From the results of quantile cointegrating regressions, it is found that there are asymmetric long run equilibrium relationships in the conditional distribution of the dependent variable, which cannot be found in the conditional mean. The quantile cointegration analysis for the stock markets indicates partial integration between China and the mature financial markets, which is only significant in some quantiles. Compared with the linear and quantile regression models without breaks and the linear model with one break, the quantile regression model with one structural change provides the strongest evidence of financial market integration.

2.5 Conclusion

Chapter Two considers estimation and asymptotic theory for quantile cointegrating regressions allowing for multiple structural changes of unknown timing. When the number of structural changes is pre-specified, the unknown break dates and regression coefficients are estimated jointly by minimizing the check function over all possible partitions of the sample. Both the break fraction ($\hat{\lambda}_j$) and parameter ($\hat{\theta}_j(\tau)$) estimates are consistent. The convergence rate of the break fraction estimates is derived under the assumption of shrinking shift. The conditional quantile estimator has a nonstandard distribution in the limit. In particular, the asymptotic distribution of the slope coefficient estimates $\hat{\beta}_j(\tau)$ has a second-order bias and depends on the nuisance parameter Ω. Thus, this chapter provides a fully modified estimator for serial correlation and long run endogeneity corrections. The limit distribution of the fully modified coefficient estimates $\hat{\beta}_j^+(\tau)$ is mixed normal. The simulation results show that both the bias and root mean square error are significantly reduced after applying the fully modification when the residual cross correlation is nonzero. The confidence intervals for the regression coefficients have adequate coverage rates after the modification and the confidence intervals for the break fraction are slightly conservative. Furthermore, this chapter applies the model to investigating the linkage between the financial market in China and several mature markets. In the bivariate quantile regressions of the stock indices, the quantile dependent cointegrating coefficients are subject to one structural change and are found significant only in part of the distribution of the dependent variable, which provides some evidence of partial equity market integration.

As an extension, the model can include trending regressors of the form $(t/T)^l$ for $l > 0$, or more generally, $g(t/T)$, where $g(\cdot)$ is a continuous function (Bai 1997, Bai and Perron 1998). In this case, consistency of the conditional quantile estimator holds with the same rates of convergence derived in this chapter under the assumption of shrinking shift. However, the limit distributions will be different. Also, the regression model can include both stationary and nonstationary regressors. Moreover, this chapter considers the case of pure structural change, in which all regression coefficients are subject to change. A natural extension is to consider a partial break where a subset of the parameters change in the presence of a break.

Also, the estimation can be extended to multiple quantiles. Across multiple quantiles we have

$$(\hat{\theta}(\mathcal{T}_\omega), \hat{T}^b) = \arg \min_{\theta(\mathcal{T}_\omega), T^b} S_T(\mathcal{T}_\omega, \theta(\mathcal{T}_\omega), T^b),$$

where $S_T(\mathcal{T}_\omega, \theta(\mathcal{T}_\omega), T^b) = \sum_{h=1}^{q} \sum_{j=0}^{m} \sum_{t=T_j+1}^{T_{j+1}} \rho_{\tau h}(y_t - z_t' \theta_{j+1}(\tau_h))$ with $\mathcal{T}_\omega = [\omega_1, \omega_2]$ and $\theta(\mathcal{T}_\omega) = (\theta(\tau_1)', ..., \theta(\tau_q)')'$ (Oka and Qu 2011).

Finally, this chapter derives estimates and limit distributions in the quantile cointegration model with structural changes taking the number of breaks as given. When the number of structural changes is unknown, a sequential procedure can be used to determine the number of breaks in the spirit of Bai and Perron (1998). In this case, tests of structural changes or parameter instability for quantile regressions with I(1) variables can be developed.

Chapter 3

Inference in Predictive Quantile Regressions

3.1 Introduction

In this chapter, we develop asymptotic theory in the context of predictive quantile regressions with nearly nonstationary regressors. This has important empirical applications, such as testing the predictability of the stock return distribution. Beginning with influential work by Shiller (1984), Campbell and Shiller (1988a, 1988b), Fama and French (1988), and Hodrick (1992), there has been an extensive literature on predictive tests for mean returns, which debates whether predictors such as dividend yield can forecast stock returns. This has implications, not only for the risk neutral market efficiency hypothesis, but also for portfolio analysis. Indeed, some of the most recent empirical work debates the ability of investors to use predictors, such as dividend yield, to create dynamic asset allocation strategies that outperform the market (Goyal and Welch 2008, Campbell and Thompson 2008). Similarly, asset allocation strategies employing the large and widely successful literature on volatility modeling, may improve the expected welfare of the risk averse investor.

While most of the empirical literature has focused exclusively on predictive means or variances, outside of a very few special cases, such as the mean-variance portfolio models, the portfolio decision depends on the entire return distribution. Likewise, the tails of the distribution are of particular interest to risk managers and are also important to policy makers, who must consider the worst case, as well as base-line, forecast scenarios. Recent empirical work by Cenesizoglu and Timmermann (2008) employs the quantile regression method introduced by Koenker and Bassett (1978), to extract a richer set of return predictions involving not just the center of the predictive density, but also the shoulders and tails. Their empirical results suggest that there is valuable information in the conditional quantiles that cannot be ascertained from the conditional mean and variance alone. For example, they find that a number of pre-determined predictors have an asymmetric effect on various quantiles in the return distribution. Likewise, they find that some predictors, which have little information for the center of the distribution, nonetheless have important implications for the tails.

One reason that the on-going debate over predictive mean regression has lasted so long is that standard inference procedures can be unreliable when predictors are persistent, as is commonly the case in practice. Earlier work adopted standard t-tests and found strong evidence of predictability (Shiller 1984, Campbell and Shiller 1988a, 1988b, Fama and French 1988, Hodrick 1992). However, as Shiller (1984) points out, some regressors in the predictive regressions are stochastic, therefore the conventional t-tests are invalid. In particular, the predictability of the return with these predictors may be overstated. Mankiw and Shapiro (1986) investigate small sample properties of tests for rational expectation models. They find that ordinary tests of orthogonality reject too frequently when the predictive variables are nearly nonstationary. Other papers, such as Stambaugh (1986), Cavanagh et al. (1995), and Stambaugh (1999), also emphasize the poor small sample properties in predictive regressions with highly persistent regressors.

Conventional predictive tests of stock returns are not valid for the following reasons. Firstly, the predictor variables, such as dividend yield, dividend price, and earnings price ratios, are strongly autocorrelated. Secondly, although pre-determined, typically the predictor is not strictly exogenous, since its innovation is often highly correlated with the error term in the predictive regression. For example, with financial data, such as dividend price ratios and returns, it is reasonable to expect strong correlation between the regressor's innovation and the regression disturbance. This tends to inflate the t-statistic, which results in over-rejection. In linear predictive regressions, it is well-known that when the regressor is not strictly exogenous and has a largest autoregressive root close to one, the limiting distribution of the t-statistic will be nonstandard. In this case, the t-statistic is too large and tests using the standard normal critical values will over-reject the null hypothesis of nonpredictability.

Much attention has been devoted to overcoming such size distortions in linear predictive regressions, resulting in a rich literature, with several general approaches having been explored. A number of alternatives to the standard regression based predictive testing approach have been proposed. Campbell and Dufour (1995, 1997) propose the use of tests based on nonparametric sign and sign-rank statistics with exact finite sample size. Maynard and Shimotsu (2009) suggest a semi-parametric covariance based predictive test that is asymptotically normal and allows for an alternative hypothesis with a stationary dependent variable, even when the predictor is near nonstationary. Wright (2000) and Lanne (2002) reinterpret the predictive test as a stationarity test and use this insight to provide conservative testing procedures.

A second strand of the literature applies finite sample corrections to the standard predictive t- or F-test, often motivated under more tightly parameterized models. Stambaugh (1999) derives small sample Bayesian posterior distributions for the regression parameters. Lewellen (2004) performs finite sample correction by calculating a joint significance level from a combination of the conditional and unconditional tests. Amihud and Hurvich (2004) and Amihud et al. (2004) obtain bias reductions using augmented regression methods. Likewise, resampling approaches have also been proposed to improve finite sample inference (Nelson and Kim 1993, Goetzmann and Jorion 1993, Wolf 2000).

A final approach that has proved productive, and which we adapt to the quantile context, is the use of an explicit local-to-unity specification for the predictor, in order to provide for improved estimation and inference in the standard predictive model. Cavanagh et al. (1995) propose corrected critical values based on a local-to-unity model with known values of the local-to-unity parameter c. Since this parameter cannot be consistently estimated, they then propose feasible inference methods using bounds procedures to control size. Campbell and Yogo (2006) instead employ the local-to-unity based setting to correct the asymptotic bias of the predictive regression estimator. The resulting estimator is mixed normal and asymptotically efficient for known c. Since their correction depends on c, a refined bounds procedure is employed for feasible inference. Hjalmarsson (2007) notes that the Campbell and Yogo (2006) procedure can be interpreted as a local-to-unity version of the fully modified estimator of Phillips and Hansen (1990) and proposes a generalization. Jansson and Moreira (2006) derive a test that is both unbiased and conditionally optimal, without knowledge of the localization parameter, using Gaussian asymptotic power envelopes.

In contrast to this large literature devoted to proper inference techniques for predictive mean regression, we are aware of no theoretical work to date, that establishes valid econometric inference methods in quantile predictive regression with persistent regressors. As demonstrated by Cenesizoglu and Timmermann (2008), quantile predictive regressions have considerable empirical interest, yet our preliminary simulations indicate that standard tests based on predictive quantile regressions are subject to similar size distortion as linear predictive tests. In this chapter, we develop proper inference methods for short-horizon predictive quantile regressions with nearly integrated regressors. We first derive the limit distribution of the quantile regression coefficients by generalizing the results of Xiao (2009), who derives inference in a quantile regression with cointegrated time series, to the local-to-unity setting of Chan and Wei (1987), Chan (1988), Phillips (1987), Phillips (1988a, 1988b), and Nabeya and Sorensen (1994)[17]. We then provide a test of quantile predictability based on an asymptotically valid Bonferroni bounds methodology in the spirit of Cavanagh et al. (1995).

Our results contribute to a rapidly developing literature in quantile regression, in which there is substantial recent work in both theoretical and empirical areas. Recent developments of quantile methods for time series data include quantile autoregression (Koenker and Xiao 2006), unit root quantile autoregression (Koenker and Xiao 2004), quantile cointegration (Xiao 2009), conditional quantile estimation for GARCH models (Xiao and Koenker 2009), and Copula-based quantile autoregression (Chen et al. 2009). Chernozhukov and Du (2006) and Chernozhukov (2010) develop the theory of conditional extremal quantiles with applications to value-at-risk and birth weights. Nonparametric quantile methods and their applications have also been developing rapidly (see, for example, Frolich and Melly 2008, Huber 2010, Jun et al. 2009, and Koenker 2010). Other recent contributions include

[17]These papers are all based on nearly nonstationary AR(1) models.

Koenker (2008) who proposes censored quantile regression estimation, Chernozhukov et al. (2009) who derive finite sample inference for quantile regression models under minimal assumptions, and Chernozhukov and Belloni (2010) who study l_1-penalized quantile regression in high-dimensional sparse models.

The remainder of Chapter Three is organized as follows: In Section 3.2, the framework of the problem is established and the asymptotic theory for predictive quantile regression under a local-to-unity specification is developed. In Section 3.3, the Bonferroni method used to correct the size distortion is proposed. In Section 3.4, results from our simulation study are reported. In Section 3.5, the techniques are applied to testing the predictability of the stock return distribution using various pre-determined predictive regressors. Section 3.6 concludes the chapter.

3.2 Theory

3.2.1 Model and assumptions

Denote $\mathcal{F}_t$ as the information set up to time t. The standard mean prediction model is given by

$$y_t = \gamma_0 + \gamma_1 x_{t-1} + \varepsilon_{2t}, \tag{3.1}$$

where $E[\varepsilon_{2t}|\mathcal{F}_{t-1}] = 0$, y_t is typically a financial return, and x_t is a predictor, such as an earnings or dividend price ratio. This yields the mean prediction $E[y_t|\mathcal{F}_{t-1}] = \gamma_0 + \gamma_1 x_{t-1}$, but provides no information on other aspects of the predictive distribution without further assumption on the error term.

Let $F(\cdot)$ and $F_{t-1}(\cdot) = Pr(\varepsilon_{2t} < \cdot|\mathcal{F}_{t-1})$ denote the cumulative and conditional distributions of ε_{2t} and define the τth unconditional and conditional quantiles of ε_{2t} by

$Q_{\varepsilon_{2t}}(\tau) = F^{-1}(\tau)$ and $Q_{\varepsilon_{2t}}(\tau|\mathcal{F}_{t-1}) = F_{t-1}^{-1}(\tau)$, respectively. With the added assumption that $F_{t-1}(\cdot) = F(\cdot)$, it is possible to use (3.1) to forecast quantiles of y_t employing

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \gamma_0 + Q_{\varepsilon_2}(\tau) + \gamma_1 x_{t-1} \tag{3.2}$$

for $\tau \in (0, 1)$, where $Q_{y_t}(\tau|\mathcal{F}_{t-1})$ is defined as the τth quantile of y_t conditional on $\mathcal{F}_{t-1}$. Nonetheless, the quantile forecasts in (3.2) may be viewed as somewhat restrictive since all quantiles share the same slope coefficient γ_1, allowing only for parallel shifts of the conditional quantile. It implies, for example, that the predictor x_{t-1} shifts the center, shoulders, and tails of the distribution all in the same direction and by exactly the same amount.

To obtain a more flexible predictive model, we relax the assumption that the residual distribution in (3.1) is time invariant and allow x_{t-1} to impact not only the mean of y_t, but also the distribution of its error term. In particular, we model the conditional quantile of the error term as

$$Q_{\varepsilon_{2t}}(\tau|\mathcal{F}_{t-1}) = \gamma_0(\tau) - \gamma_0 + (\gamma_1(\tau) - \gamma_1)\, x_{t-1}, \tag{3.3}$$

in which the dependence of $\gamma_1(\tau)$ on τ allows the impact of x_{t-1} to vary across the quantiles of ε_{2t}. Combining (3.2) and (3.3) gives the predictive quantile model for y_t:

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \gamma_0(\tau) + \gamma_1(\tau)x_{t-1}, \tag{3.4}$$

which provides a flexible specification allowing the effect of x_{t-1} to be heterogeneous across the quantiles of y_t.

It is evident from (3.2) that the standard linear mean prediction model (with constant residual distribution) is a special case of the quantile regression model in (3.4). As noted by Cenesizoglu and Timmermann (2008), the quantile predictive model also encompasses a number of other empirical models for financial returns. For example, if x_t is a variable with predictive content for volatility, such as squared returns or realized volatility, we may consider a model of the form

$$y_t = \gamma_0 + \gamma_1 x_{t-1} + (\delta_0 + \delta_1 x_{t-1})\varepsilon_{2t}, \tag{3.5}$$

where $F_{t-1}(\cdot) = F(\cdot)$. The predictive quantile for y_t then takes the form:

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \gamma_0(\tau) + \gamma_1(\tau)x_{t-1}, \tag{3.6}$$

where $\gamma_0(\tau) = \gamma_0 + \delta_0 F^{-1}(\tau)$ and $\gamma_1(\tau) = \gamma_1 + \delta_1 F^{-1}(\tau)$ both depend non-trivially on τ.

We next consider the data generating process for the predictor. Since the majority of predictors employed in practice are highly persistent, we model the regressor x_t as a near unit root process. In particular, following Cavanagh et al. (1995), we assume that the predictor x_t is a finite order autoregressive process:

$$x_t = \alpha_0 + v_t, \qquad (1 - \alpha_1 L)b(L)v_t = \varepsilon_{1t}, \tag{3.7}$$

where $b(L) = \sum_{i=0}^{k} b_i L^i$ with $b_0 = 1$. Assume that the roots of $b(L)$ are fixed and less than one in absolute value. Equation (3.7) can be rewritten as:

$$\Delta x_t = \beta_0 + \beta_1 x_{t-1} + \zeta(L)\Delta x_{t-1} + \varepsilon_{1t}, \tag{3.8}$$

where $\beta_0 = (1 - \alpha_1)b(1)\alpha_0$, $\beta_1 = (\alpha_1 - 1)b(1)$, and $\zeta(L) = -\sum_{j=1}^{k} L^{-1}[b_j - (1 - \alpha_1)\sum_{i=j}^{k} b_i]L^j$. In this augmented Dicky-Fuller representation, it is straightforward that x_t follows an $AR(k+1)$ process with the largest root α_1. Of particular interest in this process is the local-to-unity specification, $\alpha_1 = 1 + \frac{c}{T}$, where $c \leq 0$ and T is the sample size. When $c = 0$ this generalizes to a unit root process. The mean reverting case $c < 0$ provides a useful large sample approximation for the case in which the largest root is very close to, but still less than one. A number of prior studies have used this framework to model predictors such as earnings and dividend price ratios which are highly persistent, but a priori stationary on economic grounds.

Assumption 7 *The error term* $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ *is a martingale difference sequence and* $E(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}) = \Sigma$ *with typical element* σ_{ij}.

Assumption 8 *The cumulative distribution function of* ε_{2t}, $F(\cdot)$, *has a continuous density function* $f(\cdot)$, *which is positive on* $\{\varepsilon_{2t} : 0 < F(\varepsilon_{2t}) < 1\}$.

Assumption 9 *The conditional distribution function* $F_{t-1}(\cdot) = Pr(\varepsilon_{2t} < \cdot | \mathcal{F}_{t-1})$ *has derivative* $f_{t-1}(\cdot)$ *a.s.,* $f_{t-1}(s_n)$ *is uniformly integrable for any sequence* $s_n \to F^{-1}(\tau)$, *and* $E[f_{t-1}^{\xi}(F^{-1}(\tau))] < \infty$ *for some* $\xi > 1$.

Xiao (2009) makes the same assumptions on the distribution functions of the error term.

The standard quantile regression coefficient estimates are given by

$$(\hat{\gamma}_0(\tau), \hat{\gamma}_1(\tau)) = \arg \min_{(\gamma_0, \gamma_1) \in R^2} \sum_{t=1}^{T} \rho_\tau(y_t - \gamma_0 - \gamma_1 x_{t-1}), \tag{3.9}$$

where $\rho_\tau(\cdot)$ is the asymmetric absolute deviation loss function defined by $\rho_\tau(u) = u\psi_\tau(u)$ with $\psi_\tau(u) = \tau - I(u < 0)$ (Koenker and Bassett 1978). For the median $\rho_\tau(u) = 1/2|u|$ is used for Laplaces median regression function. In the next subsection, we derive the asymptotic behavior of these estimates under the local-to-unity specification for x_t.

3.2.2 Theoretical results

Define $\delta = corr(\varepsilon_{1t}, \varepsilon_{2t})$. Let $B = (B_1, B_2)'$ denote a two-dimensional Brownian motion, the covariance matrix of which is $\bar{\Sigma}$ with elements $\bar{\sigma}_{11} = \bar{\sigma}_{22} = 1$ and $\bar{\sigma}_{12} = \bar{\sigma}_{21} = \delta$. We have

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \begin{bmatrix} b^{-1}(L)\varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \Rightarrow \begin{bmatrix} \omega B_1(r) \\ \sigma_{22}^{\frac{1}{2}} B_2(r) \end{bmatrix} = BM(0, \Omega), \tag{3.10}$$

where $\Omega = \begin{bmatrix} \omega^2 & \omega\sigma_{22}^{\frac{1}{2}}\delta \\ \omega\sigma_{22}^{\frac{1}{2}}\delta & \sigma_{22} \end{bmatrix}$ with $\omega = \frac{\sigma_{11}^{\frac{1}{2}}}{b(1)}$.

Let $Z = (Z_1, Z_2)'$ be a two-dimensional Brownian motion such that $Z_1(r) = \omega B_1(r)$ and $Z_2(r) = \sigma_{22}^{\frac{1}{2}} B_2(r)$. Also, denote by J_c the diffusion process defined by $dJ_c(s) = cJ_c(s)ds + dB_1(s)$ with initial condition $J_c(0) = 0$. We distinguish demeaned variables by superscript μ. For example, $x_t^\mu = x_t - (T-1)^{-1}\sum_{t=2}^T x_{t-1}$ and $J_c^\mu(s) = J_c(s) - \int_0^1 J_c(r)dr$. In light of Phillips (1987), under $\alpha_1 = 1 + \frac{c}{T}$, we have

$$\begin{aligned} T^{-\frac{1}{2}} x_{[Tr]}^\mu &\Rightarrow \omega J_c^\mu, \\ T^{-2}\sum (x_{t-1}^\mu)^2 &\Rightarrow \int (\omega J_c^\mu)^2 dr, \\ T^{-1}\sum x_{t-1}^\mu b^{-1}(L)\varepsilon_{1t} &\Rightarrow \int \omega J_c^\mu dZ_1, \end{aligned}$$

where $[\cdot]$ denotes the greatest lesser integer function.

The following standard result (see, for example, Stock (1991) and references within) gives the asymptotic of the t-statistic associated with the largest root of the autoregressive model for x_t in (3.8), when this root is modeled local-to-unity. The result is presented here without proof.

Proposition 1 *The asymptotic representation of the standard t-statistic used to test $H_0 : \beta_1 = 0$ in (3.8) is given by*

$$t_{\beta_1} \Rightarrow \frac{c}{b(1)}\left[\int (J_c^\mu)^2 dr\right]^{\frac{1}{2}} + \frac{\int J_c^\mu dB_1}{[\int (J_c^\mu)^2 dr]^{\frac{1}{2}}}. \tag{3.11}$$

Defining $\varepsilon_{2t\tau} = \varepsilon_{2t} - F_{t-1}^{-1}(\tau) = y_t - \gamma_0(\tau) - \gamma_1(\tau)x_{t-1}$ and $Q_{\varepsilon_{2t\tau}}(\tau|\mathcal{F}_{t-1})$ as the τth quantile of $\varepsilon_{2t\tau}$ conditional on $\mathcal{F}_{t-1}$, we may rewrite (3.4) as

$$y_t = \gamma_0(\tau) + \gamma_1(\tau)x_{t-1} + \varepsilon_{2t\tau} = \gamma(\tau)' z_{t-1} + \varepsilon_{2t\tau}, \tag{3.12}$$

where $Q_{\varepsilon_{2t\tau}}(\tau|\mathcal{F}_{t-1}) = 0$. Since $\psi_\tau(\varepsilon_{2t\tau}) = \tau - I(\varepsilon_{2t} < F^{-1}(\tau))$, we have the conditional expectation $E[\psi_\tau(\varepsilon_{2t\tau})|\mathcal{F}_{t-1}] = 0$ and the variance of the indicator function $I(\cdot)$ is $\tau(1-\tau)$. The following preliminary convergence result, which is comparable to Assumption A of Xiao (2009) and stated without proof, will be needed for the subsequent analysis.

Lemma 1

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \begin{bmatrix} b^{-1}(L)\varepsilon_{1t} \\ \psi_\tau(\varepsilon_{2t\tau}) \end{bmatrix} \Rightarrow \begin{bmatrix} Z_1(r) \\ Z_\psi(r) \end{bmatrix} = BM(0, \Omega_\tau), \tag{3.13}$$

where $\Omega_\tau = \begin{bmatrix} \omega^2 & \omega\sqrt{\tau(1-\tau)}\delta \\ \omega\sqrt{\tau(1-\tau)}\delta & \tau(1-\tau) \end{bmatrix}.$

It follows that $T^{-1}\sum x_{t-1}^\mu \psi_\tau(\varepsilon_{2t\tau}) \to_d \int \omega J_c^\mu dZ_\psi$. The following result provides the limiting distribution of the predictive quantile regression estimator in (3.12).

Proposition 2 *The asymptotic distribution of the regression coefficient estimates is*

$$D_T(\hat{\gamma}(\tau) - \gamma(\tau)) \Rightarrow \frac{1}{f(F^{-1}(\tau))} \left[\int \bar{J}_c \bar{J}_c'\right]^{-1} \left[\int \bar{J}_c dZ_\psi\right]. \tag{3.14}$$

For the slope coefficient, in particular,

$$T(\hat{\gamma}_1(\tau) - \gamma_1(\tau)) \Rightarrow \frac{1}{f(F^{-1}(\tau))} \left[\int (\omega J_c^\mu)^2\right]^{-1} \left[\int \omega J_c^\mu dZ_\psi\right], \tag{3.15}$$

where $D_T = \begin{bmatrix} T^{\frac{1}{2}} & 0 \\ 0 & T \end{bmatrix}$ *and* $\bar{J}_c = (1, \omega J_c)'$.

The distribution is nonstandard. In the case when $c = 0$, it specializes the result of the quantile cointegrating regression (Xiao 2009, Theorem 1) to the case of predictive regression. The extension to $c < 0$ is new to the best of our knowledge. As in the case of cointegrating regression, some further insight into the bias can be gained from the projection of Z_ψ on to B_1, which yields the orthogonal decomposition (see Phillips 1989, pp. 30-31)

$$Z_\psi = \sqrt{\tau(1-\tau)}\left[\omega^{-1}\delta Z_1(r) + \sqrt{1-\delta^2}Z_{\psi\cdot 1}\right] = \sqrt{\tau(1-\tau)}\left[\delta B_1(r) + \sqrt{1-\delta^2}Z_{\psi\cdot 1}\right],$$

where $Z_{\psi\cdot 1} = BM(1)$ and is independent of Z_1. Using this decomposition to substitute for dZ_ψ, we may re-express the numerator of Proposition 2 as

$$\frac{1}{f(F^{-1}(\tau))}\left[\delta\sqrt{\tau(1-\tau)}\left(\int(\omega^2 J_c^\mu)^2\right)^{-1}\int \omega J_c^\mu dB_1 + \omega_{\psi\cdot 1}\left(\int(\omega^2 J_c^\mu)^2\right)^{-1}\int \omega J_c^\mu dZ_{\psi\cdot 1}\right], \tag{3.16}$$

where we define $\omega_{\psi\cdot 1}^2 = (1-\delta^2)\tau(1-\tau)$. The stochastic integral inside the brackets is the local-to-unity generalization of the (de-meaned) Dickey-Fuller distribution and contributes a downward (upward) second-order bias to the estimate of $\hat{\gamma}_1(\tau)$ for $\delta > 0$ ($\delta < 0$). It is analogous to what Phillips and Hansen (1990) refer to as the serial correlation term in the context of cointegration. It is evident that the extent of the bias depends on both δ and on c, as is also true in the linear predictive regression. The second term in brackets is mixed normal, and normal conditional on $\mathcal{F}_1 = \sigma\left(B_1(r), 0 \leq r \leq 1\right)$, due to the independence of J_c and $Z_{\psi\cdot 1}$. As in the case of predictive regression, there is no endogeneity term due to the assumption that ε_{2t} is a martingale difference sequence. The distribution of the estimator depends on τ both directly and through $Z_{\psi\cdot 1}$, which is itself a function of τ.

The standard error of $\hat{\gamma}_1(\tau)$ is given by $\sqrt{\frac{1}{f(\widehat{F^{-1}(\tau)})^2}[\sum(x_{t-1}^\mu)^2]^{-1}}$, where $f(\widehat{F^{-1}(\tau)})$ is a consistent estimator of $f(F^{-1}(\tau))$. The following proposition provides the null limiting distribution of the standard t-statistic in the predictive quantile regression.

Proposition 3 *The asymptotic representation of the t-statistic to test* $H_0 : \gamma_1(\tau) = 0$[18] *for*

[18] *Note that the null hypothesis can be generalized to* $\gamma_1(\tau) = \gamma_{1,0}$.

a given τ is

$$t_{\gamma_1}(\tau) \quad \Rightarrow \quad \frac{\int J_c^\mu dZ_\psi}{[\int (J_c^\mu)^2]^{\frac{1}{2}}} \equiv \sqrt{\tau(1-\tau)} \left[\delta \frac{\int J_c^\mu dB_1}{[\int (J_c^\mu)^2]^{\frac{1}{2}}} + \sqrt{1-\delta^2} z \right], \tag{3.17}$$

where z is a standard normal random variable that is independent of (B_1, J_c).

The asymptotic distribution of $t_{\gamma_1}(\tau)$ depends on (c, δ, τ). The dependence on c and δ is similar to that in the linear regression case, while dependence on τ is new.

3.3 Inference

In equation (3.8), if $\alpha_1 < 1$ and is fixed, the unit root test will reject with probability of one asymptotically. Since x_t is stationary in this case, the quantile regressions of equation (3.1) are well-behaved in large samples. It would be sensible to test $H_0 : \gamma_1(\tau) = 0$ using standard normal critical values. If $\alpha_1 = 1$, which corresponds to $c = 0$, then x_t has a unit root. When $c < 0$ and α_1 is large, the regressor x_t has a near unit root. In the stock return predictability example, many predictors, such as the dividend price ratio, have large autoregressive roots that are very close to one. Hence, it is appropriate to model x_t as a near unit root process. As Proposition 3 shows, in this case the limiting distribution of $t_{\gamma_1}(\tau)$ is nonstandard and dependent on the nuisance parameter c. Moreover, the asymptotic distribution of $t_{\gamma_1}(\tau)$ also depends on δ. If the two error terms in equation (3.8) and (3.1) are uncorrelated, the t-statistic will have a standard normal asymptotic distribution. However, when the two error terms are correlated, the limiting distribution of the quantile regression coefficient is no longer standard normal, which causes size distortion of the predictability test. Thus, standard normal critical values are not reliable for the test of interest, since they tend to over-reject the null. This is a problem in practice, because financial data such as prices and dividends do not satisfy strict exogeneity. The over-rejection is especially severe when the residual cross correlation is large.

Therefore, this chapter proposes a Bonferroni bounds method in the spirit of Cavanagh et al. (1995), which results in asymptotically valid tests. The procedure has two steps. Because the local-to-unity parameter c is a nuisance parameter that cannot be consistently estimated, we first derive a $100(1-\eta_1)\%$ confidence interval for c and denote it by $CI_c(\eta_1)$. This first-stage confidence interval is computed by inverting a unit root test statistic as illustrated in Stock (1991). In the second stage, based on the previously derived asymptotic distribution, we compute the critical value for $t_{\gamma_1}(\tau)$ of size η_2 at each point of c in the first-stage confidence interval. Then, the Bonferroni bounds $[C_l(\eta_1,\eta_2), C_u(\eta_1,\eta_2)]$ give the asymptotically valid critical values for $t_{\gamma_1}(\tau)$, where

$$\begin{aligned} \text{lower bound:} \qquad C_l(\eta_1,\eta_2) &= \min_{c\in CI_c(\eta_1)} C_{t_{\gamma_1},c,\frac{\eta_2}{2}} \\ \text{upper bound:} \qquad C_u(\eta_1,\eta_2) &= \max_{c\in CI_c(\eta_1)} C_{t_{\gamma_1},c,1-\frac{\eta_2}{2}} \end{aligned}$$

and $C_{t_{\gamma_1},c,\frac{\eta_2}{2}}$ is the $100\frac{\eta_2}{2}\%$ quantile of the limiting distribution of $t_{\gamma_1}(\tau)$ for a given δ. Tests based on the Bonferroni bounds are conservative with size less than or equal to η for $\eta = \eta_1 + \eta_2$.

Although the Bonferroni interval is conservative, the size of the test can be adjusted by selecting appropriate (η_1, η_2). More specifically, one can set $\eta_2 = \eta$ and choose η_1 such that the asymptotic size equals to η. In our Monte Carlo experiments, we consider $\eta = 10\%$ as the size of the test and choose η_1 under $T = 1000$ such that the size is close to the benchmark case where c is known.

3.4 Simulation Study

Our simulation study focuses on the demeaned case in which the regression contains a constant term (without a time trend term). The results are based on 2000 Monte Carlo replications with sample size T for several sets of parameters (c, δ, τ), where c ranges from 0 to $-T$, δ ranges from 0 to 0.95, and τ ranges from 0.05 to 0.95. In each replication, the regressor x_t is generated according to equation (3.8). We set $b_j = 0$ for $j > 1$. Under the local-to-unity specification, we have $x_t = (1 + \frac{c}{T})x_{t-1} + \varepsilon_{1t}$. The dependent variable y_t is generated by equation (3.1) under the null $H_0 : \gamma_1 = 0$. We set $\gamma_0 = 0$ so that $y_t = \varepsilon_{2t}$. The errors ε_{1t} and ε_{2t} are both $N(0, 1)$ with correlation coefficient δ. The level of the predictability test is 10%.

With a nearly integrated predictor, conventional t-tests from the predictive quantile regression over-reject the null. As a preliminary experiment, we investigate the size distortion problem associated with the t-test based on standard normal critical values, ± 1.65. As Table 3.1 shows, when $\delta = 0$ the test is conservative for $\tau = 0.2$ to 0.8 with a large sample size such as $T = 1000$. From Table 3.2, even with a small sample size such as $T = 200$, the rejection rates for $\tau = 0.2$ to 0.8 are still close to 10%, which is the level of the test. Hence, when there is no or small correlation between ε_{1t} and ε_{2t}, there is no or very little over-rejection for the inner quantiles. As the correlation coefficient increases, the over-rejection becomes more severe. For $\delta = 0.50$ and 0.95, the null is over-rejected for all quantiles, even when $T = 1000$. Particularly, when the correlation is as large as 0.95 the rejection rate can be over 25%. Similar over-rejection problems are illustrated for linear predictive regressions in Mankiw and Shapiro (1986) and Cavanagh et al. (1995): with $T = 500, 200, 100$, conventional t-tests can lead to substantial size distortion, especially for large correlation coefficients.

Table 3.1: Standard t-tests: Finite-sample size (T = 1000)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0$					
0	0.1240	0.1055	0.0885	0.0900	0.0915	0.0870	0.0915	0.0930	0.1005	0.1055	0.1235
- 1	0.1305	0.1125	0.0855	0.0950	0.0910	0.0915	0.0950	0.0910	0.0940	0.1050	0.1255
- 5	0.1370	0.1085	0.0990	0.0915	0.0855	0.0835	0.0820	0.0910	0.0955	0.1100	0.1225
-10	0.1405	0.1055	0.1015	0.1000	0.0885	0.0880	0.0840	0.0865	0.0925	0.1175	0.1215
-20	0.1340	0.1130	0.0980	0.0905	0.0940	0.0880	0.0835	0.0950	0.0880	0.1140	0.1230
-30	0.1335	0.1150	0.1015	0.0900	0.0915	0.0955	0.0850	0.0900	0.0910	0.1105	0.1245
- T	0.1555	0.1065	0.0965	0.1020	0.0875	0.0935	0.0940	0.0935	0.1100	0.1175	0.1310
						$\delta = 0.50$					
0	0.1440	0.1320	0.1355	0.1365	0.1340	0.1345	0.1320	0.1315	0.1370	0.1405	0.1410
- 1	0.1445	0.1275	0.1215	0.1320	0.1210	0.1185	0.1285	0.1215	0.1205	0.1395	0.1325
- 5	0.1395	0.1140	0.0990	0.1090	0.1095	0.1035	0.1105	0.1045	0.1140	0.1195	0.1190
-10	0.1490	0.1125	0.1010	0.1030	0.0995	0.0985	0.1005	0.0980	0.1080	0.1190	0.1180
-20	0.1525	0.1205	0.1005	0.0975	0.0935	0.0980	0.0960	0.0950	0.1045	0.1215	0.1215
-30	0.1335	0.1160	0.1015	0.0955	0.0940	0.0865	0.0915	0.0965	0.1030	0.1135	0.1220
- T	0.1420	0.1075	0.1020	0.0960	0.0900	0.0945	0.0940	0.1010	0.1110	0.1040	0.1220
						$\delta = 0.95$					
0	0.2030	0.2190	0.2615	0.2765	0.2825	0.2725	0.2610	0.2645	0.2420	0.2075	0.1895
- 1	0.1750	0.1985	0.2180	0.2260	0.2300	0.2245	0.2200	0.2265	0.2175	0.1940	0.1845
- 5	0.1560	0.1465	0.1540	0.1470	0.1380	0.1370	0.1370	0.1480	0.1550	0.1420	0.1545
-10	0.1455	0.1305	0.1280	0.1245	0.1040	0.1100	0.1145	0.1215	0.1310	0.1240	0.1410
-20	0.1370	0.1235	0.1080	0.1035	0.1020	0.0995	0.1005	0.1060	0.1120	0.1155	0.1335
-30	0.1355	0.1225	0.1015	0.1055	0.1025	0.1035	0.1005	0.1030	0.1040	0.1135	0.1330
- T	0.1380	0.1130	0.1025	0.0990	0.0990	0.0915	0.0890	0.1040	0.1115	0.1170	0.1235

Table 3.2: Standard t-tests: Finite-sample size (T = 200)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0$					
0	0.1640	0.1335	0.0995	0.0910	0.0945	0.0820	0.0880	0.1005	0.1040	0.1245	0.1630
- 1	0.1615	0.1305	0.1000	0.0865	0.0870	0.0765	0.0885	0.0955	0.1020	0.1125	0.1600
- 5	0.1625	0.1425	0.1085	0.0915	0.0965	0.0840	0.0835	0.0905	0.1055	0.1230	0.1680
-10	0.1710	0.1325	0.1060	0.0965	0.0940	0.0855	0.0885	0.0980	0.0970	0.1270	0.1680
-20	0.1685	0.1285	0.1025	0.1030	0.0975	0.0790	0.0830	0.0830	0.0955	0.1265	0.1700
-30	0.1665	0.1335	0.1110	0.1000	0.0930	0.0800	0.0875	0.0865	0.0860	0.1245	0.1785
- T	0.1550	0.1325	0.1140	0.0985	0.0880	0.0880	0.0850	0.1005	0.1085	0.1320	0.1875
						$\delta = 0.50$					
0	0.1775	0.1670	0.1500	0.1415	0.1425	0.1315	0.1345	0.1345	0.1400	0.1535	0.1880
- 1	0.1780	0.1565	0.1470	0.1220	0.1295	0.1115	0.1150	0.1190	0.1220	0.1475	0.1805
- 5	0.1615	0.1455	0.1170	0.1075	0.1045	0.1000	0.0970	0.1000	0.1100	0.1325	0.1740
-10	0.1685	0.1385	0.1185	0.1165	0.1035	0.1025	0.1020	0.0950	0.0965	0.1320	0.1595
-20	0.1660	0.1400	0.1130	0.1065	0.0935	0.0955	0.0890	0.0920	0.0975	0.1230	0.1625
-30	0.1640	0.1375	0.1085	0.1030	0.0990	0.0880	0.0910	0.0920	0.0930	0.1275	0.1665
- T	0.1545	0.1305	0.1095	0.0970	0.0880	0.0820	0.0855	0.1050	0.1170	0.1375	0.1810
						$\delta = 0.95$					
0	0.2475	0.2395	0.2505	0.2600	0.2645	0.2610	0.2630	0.2590	0.2470	0.2325	0.2355
- 1	0.2140	0.2140	0.2155	0.2320	0.2310	0.2265	0.2200	0.2150	0.2210	0.1985	0.2190
- 5	0.1810	0.1700	0.1520	0.1585	0.1515	0.1500	0.1480	0.1495	0.1355	0.1515	0.1765
-10	0.1790	0.1545	0.1305	0.1305	0.1240	0.1240	0.1250	0.1175	0.1125	0.1335	0.1635
-20	0.1690	0.1405	0.1200	0.1090	0.1010	0.1050	0.0955	0.1015	0.1030	0.1335	0.1655
-30	0.1670	0.1460	0.1175	0.0990	0.0910	0.0960	0.0950	0.0940	0.1015	0.1295	0.1690
- T	0.1625	0.1305	0.0950	0.0880	0.0900	0.0815	0.0950	0.1075	0.1140	0.1345	0.1885

Notes: The results are based on 2000 Monte Carlo replications with sample size T. Simulated data is generated according to $x_t = (1 + \frac{c}{T})x_{t-1} + \varepsilon_{1t}$ and $y_t = \varepsilon_{2t}$, where ε_{1t} and ε_{2t} are $N(0, 1)$ with correlation coefficient δ. The level of the test is 10%. Tables 3.5 and 3.6 report the rejection frequencies of the standard t-tests.

To correct the size distortion, we employ the asymptotically valid critical values according to the asymptotic representation stated in Section 3.2 for a number of parameter sets. Since the analytic critical values cannot be obtained directly from the distribution, we calculate the simulated critical values under a large sample size and store them in computerized lookup tables. The simulation is based on 49999 Monte Carlo replications with $T = 1000$. Again, the regressor x_t is simulated under the local-to-unity specification $\alpha_1 = 1 + \frac{c}{T}$ and y_t is simulated under the null of no predictability. Since the limiting distribution of $\gamma_1(\tau)$ depends on c, δ, and τ, 42 percentiles of the distribution of $t_{\gamma_1}(\tau)$ are calculated on a grid of 281 values of c, from -100 to 9.5, with the grid most dense on $[-15, 5]$, for all δ and τ of interest.

To investigate the accuracy of the simulated critical values, we calculate the rejection frequencies in cases where c is known using the critical values from the lookup tables. As reported in Table 3.3, when $T = 1000$ the rejection frequencies are close to 10% by construction. We consider this case as the benchmark. When $T = 200$ Table 3.4 shows that the over-rejection problem is lessened. Although for the outer quantiles moderate size distortion still exists when the sample size is small, the rejection rates are significantly reduced, compared to those from the standard t-tests. Thus, for example, if we consider the predictability of the tails of the return distribution, it is useful to perform the size correction using the simulated critical values. For the inner quantiles, when δ is small, the rejection rates are sometimes slightly larger than those from the conventional t-tests; when δ is large, the rejection rates are smaller than those from the conventional t-tests. Hence, the size correction is not beneficial when there is no or small correlation between the error terms, but it is necessary when the correlation is large.

Table 3.3: Known c: Finite-sample size (T = 1000)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0$					
0	0.0950	0.0950	0.0890	0.0955	0.1050	0.0935	0.0955	0.0980	0.1000	0.0925	0.0845
- 1	0.0980	0.1025	0.0895	0.0965	0.1005	0.0995	0.1015	0.0945	0.0940	0.0910	0.0920
- 5	0.1005	0.0980	0.1030	0.0955	0.0995	0.0910	0.0910	0.0945	0.0955	0.0930	0.0895
-10	0.1075	0.0950	0.1025	0.1035	0.0950	0.0975	0.0880	0.0925	0.0925	0.1020	0.0910
-20	0.1010	0.1015	0.1000	0.0955	0.1035	0.0940	0.0885	0.0985	0.0875	0.0980	0.0910
-30	0.0955	0.1045	0.1025	0.0940	0.0985	0.1060	0.0890	0.0930	0.0915	0.0960	0.0945
- T	0.1150	0.0950	0.0995	0.1080	0.1010	0.1040	0.1005	0.0985	0.1095	0.0985	0.0960
						$\delta = 0.50$					
0	0.0995	0.1020	0.0905	0.0950	0.1025	0.1010	0.1030	0.0945	0.0930	0.1020	0.0955
- 1	0.1000	0.0970	0.0900	0.0975	0.1090	0.1005	0.0970	0.0955	0.1010	0.1035	0.0960
- 5	0.1005	0.0950	0.1005	0.1050	0.1085	0.0950	0.1050	0.0975	0.1025	0.1065	0.0880
-10	0.1115	0.0985	0.1020	0.1030	0.1065	0.0975	0.1050	0.1015	0.0990	0.1070	0.0875
-20	0.1100	0.1000	0.0910	0.1015	0.1060	0.1015	0.1000	0.1000	0.0985	0.1045	0.0940
-30	0.1000	0.1025	0.0945	0.0980	0.1020	0.1000	0.0960	0.0985	0.0970	0.0975	0.0965
- T	0.1140	0.0940	0.1015	0.1060	0.1025	0.1015	0.1040	0.1070	0.1105	0.0935	0.0930
						$\delta = 0.95$					
0	0.1040	0.1040	0.0990	0.1060	0.0985	0.1085	0.0965	0.0915	0.0925	0.0935	0.1020
- 1	0.0960	0.1070	0.0915	0.1050	0.0965	0.0900	0.1015	0.0940	0.0980	0.1100	0.0995
- 5	0.1005	0.0940	0.0915	0.1040	0.0955	0.0925	0.0905	0.0925	0.0980	0.1060	0.1120
-10	0.0985	0.1045	0.0900	0.1070	0.0955	0.0875	0.0915	0.0945	0.1030	0.0990	0.1025
-20	0.1015	0.1010	0.0915	0.0950	0.0960	0.0880	0.0935	0.0955	0.0950	0.1025	0.1060
-30	0.0980	0.1085	0.0930	0.0995	0.0990	0.0880	0.0965	0.0985	0.0930	0.0940	0.1040
- T	0.1095	0.1055	0.0985	0.1025	0.1090	0.1030	0.0985	0.1160	0.1130	0.1070	0.1010

Table 3.4: Known c: Finite-sample size (T = 200)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0$					
0	0.1265	0.1185	0.1020	0.0960	0.1035	0.0910	0.0915	0.1055	0.1045	0.1105	0.1280
- 1	0.1275	0.1160	0.1025	0.0900	0.0955	0.0840	0.0925	0.0975	0.1025	0.0995	0.1230
- 5	0.1220	0.1250	0.1120	0.0955	0.1015	0.0885	0.0900	0.0935	0.1045	0.1045	0.1345
-10	0.1375	0.1175	0.1070	0.0995	0.1040	0.0945	0.0935	0.1000	0.0970	0.1075	0.1315
-20	0.1295	0.1190	0.1045	0.1100	0.1055	0.0910	0.0890	0.0870	0.0950	0.1140	0.1305
-30	0.1315	0.1175	0.1140	0.1030	0.1000	0.0895	0.0925	0.0925	0.0865	0.1105	0.1370
- T	0.1185	0.1195	0.1180	0.1050	0.0980	0.0960	0.0895	0.1045	0.1070	0.1185	0.1490
						$\delta = 0.50$					
0	0.1295	0.1180	0.1095	0.0955	0.0990	0.0945	0.0875	0.0970	0.1065	0.1220	0.1295
- 1	0.1255	0.1140	0.1120	0.0945	0.1005	0.0825	0.0850	0.1055	0.1040	0.1195	0.1310
- 5	0.1200	0.1140	0.0985	0.0995	0.0990	0.0950	0.0935	0.1020	0.1090	0.1160	0.1385
-10	0.1290	0.1190	0.1130	0.1090	0.1005	0.1080	0.1000	0.0970	0.1040	0.1165	0.1285
-20	0.1330	0.1170	0.1050	0.1045	0.1030	0.0985	0.0980	0.1010	0.0975	0.1105	0.1330
-30	0.1255	0.1155	0.1075	0.1025	0.1045	0.0940	0.0980	0.1025	0.0895	0.1150	0.1375
- T	0.1285	0.1120	0.1060	0.1005	0.0960	0.0945	0.0985	0.1150	0.1180	0.1210	0.1530
						$\delta = 0.95$					
0	0.1270	0.1200	0.1260	0.1140	0.1025	0.1015	0.1030	0.1045	0.1170	0.1275	0.1380
- 1	0.1300	0.1325	0.1185	0.1130	0.1230	0.1060	0.1030	0.1050	0.1130	0.1215	0.1320
- 5	0.1240	0.1265	0.1070	0.1130	0.1115	0.1145	0.1070	0.1065	0.1070	0.1105	0.1235
-10	0.1170	0.1155	0.1090	0.1080	0.1180	0.1075	0.1010	0.0995	0.0945	0.1145	0.1295
-20	0.1310	0.1110	0.1115	0.1020	0.0970	0.1030	0.0940	0.0980	0.0920	0.1145	0.1335
-30	0.1295	0.1190	0.1030	0.0945	0.0985	0.0995	0.0875	0.0975	0.0955	0.1145	0.1340
- T	0.1325	0.1225	0.0905	0.1025	0.1045	0.0905	0.1095	0.1190	0.1190	0.1235	0.1555

Notes: Tables 3.3 and 3.4 report the rejection frequencies calculated using the critical values from the lookup tables, assuming c is known.

Since c is a nuisance parameter, which is generally unknown in practice, we next follow the two-step procedure described in the previous section. First, we run the OLS regression according to the ADF representation as in equation (3.8) to obtain $\hat{t}_{\beta_1}$. Then, we construct the $100(1-\eta_1)\%$ confidence interval for c by inverting the t-statistic (Stock 1991). Next, we run the quantile regression (3.12) and obtain $\hat{t}_{\gamma_1}(\tau)$ for each τ. Then, we compute the second-stage critical values, which are the $100(1-\eta_2)\%$ Bonferroni bounds for $\hat{t}_{\gamma_1}(\tau)$ over the confidence region of c calculated from the first stage. The rejection frequencies are calculated according to the following rejection rule: if $\hat{t}_{\gamma_1}(\tau)$ lies outside its Bonferroni interval, we reject the null; otherwise, we do not reject. The size of the test is adjusted by choosing η_1 and η_2 according to the previous section so that the test is not too conservative. Given that $\eta_2 = 10\%$ and $T = 1000$, η_1 is selected such that the rejection rates are close to those in the benchmark case. For $\delta = 0.25, 0.50, 0.75, 0.95$, this leads us to select $\eta_1 = 0.7, 0.5, 0.4, 0.3$, respectively. Tables 3.5 and 3.6 summarize the results after performing the size-adjusted Bonferroni technique. For the level of $\eta_2 = 10\%$, we plug in η_1 calculated above for each value of δ. When T=1000 the rejection rates are between 0.0800 and 0.1185, which are close to those from the benchmark case due to the size adjustment. When $T = 200$ there is negligible size distortion for the inner quantiles and moderate over-rejection for the outer quantiles, which is a finite sample problem[19]. Compared to the experiments with known c, after performing the Bonferroni correction, the rejection rates in Table 3.6 are, in general, no larger than those from Table 3.4. Therefore, although using the Bonferroni bounds may not fully solve the over-rejection problem in the tails when the sample size is small, the Bonferroni method is asymptotically valid and can effectively correct the size distortion in finite samples for intermediate quantiles.

[19]When $c = 0$ the normal approximation is poor in extreme quantiles. See, for example, Chernozhukov (2005) for the case in which x_t is $I(0)$.

Table 3.5: Bonferroni correction: Finite-sample size (T = 1000)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
					$\delta = 0$						
0	0.0950	0.0950	0.0890	0.0955	0.1050	0.0935	0.0955	0.0980	0.1000	0.0925	0.0845
- 1	0.0980	0.1025	0.0895	0.0965	0.1005	0.0995	0.1015	0.0945	0.0940	0.0910	0.0920
- 5	0.1005	0.0980	0.1030	0.0955	0.0995	0.0910	0.0910	0.0945	0.0955	0.0930	0.0895
-10	0.1075	0.0950	0.1025	0.1035	0.0950	0.0975	0.0880	0.0925	0.0925	0.1020	0.0910
-20	0.1010	0.1015	0.1000	0.0955	0.1035	0.0940	0.0885	0.0985	0.0875	0.0980	0.0910
-30	0.0955	0.1045	0.1025	0.0940	0.0985	0.1060	0.0890	0.0930	0.0915	0.0960	0.0945
- T	0.1150	0.0950	0.0995	0.1080	0.1010	0.1040	0.1005	0.0985	0.1095	0.0985	0.0960
					$\delta = 0.50$						
0	0.0960	0.0925	0.0860	0.0865	0.0990	0.0920	0.0985	0.0910	0.0875	0.0940	0.0900
- 1	0.0990	0.0930	0.0860	0.0935	0.0980	0.0905	0.0965	0.0915	0.0870	0.0960	0.0845
- 5	0.0980	0.0930	0.0965	0.1020	0.1045	0.0945	0.1070	0.0980	0.0990	0.1015	0.0855
-10	0.1080	0.0965	0.0995	0.1025	0.1085	0.0975	0.1000	0.1030	0.1020	0.1035	0.0870
-20	0.1095	0.0980	0.0920	0.1015	0.1070	0.1025	0.0975	0.0980	0.0975	0.1030	0.0940
-30	0.0985	0.1025	0.0935	0.0975	0.1025	0.1000	0.0955	0.1000	0.0975	0.0980	0.0965
- T	0.1140	0.0940	0.1015	0.1060	0.1025	0.1015	0.1040	0.1070	0.1105	0.0935	0.0930
					$\delta = 0.95$						
0	0.0920	0.0950	0.0930	0.0965	0.0945	0.0955	0.0895	0.0860	0.0870	0.0885	0.0890
- 1	0.0850	0.0875	0.0800	0.0920	0.0850	0.0820	0.0820	0.0855	0.0845	0.0955	0.0860
- 5	0.0880	0.0855	0.0930	0.1025	0.1000	0.1055	0.0985	0.0990	0.1075	0.1020	0.0960
-10	0.0935	0.1000	0.0900	0.1085	0.1005	0.0930	0.0950	0.0955	0.1100	0.0980	0.0960
-20	0.0970	0.0995	0.0915	0.0975	0.1015	0.0915	0.0940	0.0975	0.0970	0.1020	0.0990
-30	0.0980	0.1040	0.0935	0.1015	0.1000	0.0920	0.1000	0.1010	0.0955	0.0950	0.1005
- T	0.1095	0.1055	0.0985	0.1025	0.1090	0.1030	0.0985	0.1160	0.1130	0.1070	0.1010

Table 3.6: Bonferroni correction: Finite-sample size (T = 200)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0$					
0	0.1265	0.1185	0.1020	0.0960	0.1035	0.0910	0.0915	0.1055	0.1045	0.1105	0.1280
- 1	0.1275	0.1160	0.1025	0.0900	0.0955	0.0840	0.0925	0.0975	0.1025	0.0995	0.1230
- 5	0.1220	0.1250	0.1120	0.0955	0.1015	0.0885	0.0900	0.0935	0.1045	0.1045	0.1345
-10	0.1375	0.1175	0.1070	0.0995	0.1040	0.0945	0.0935	0.1000	0.0970	0.1075	0.1315
-20	0.1295	0.1190	0.1045	0.1100	0.1055	0.0910	0.0890	0.0870	0.0950	0.1140	0.1305
-30	0.1315	0.1175	0.1140	0.1030	0.1000	0.0895	0.0925	0.0925	0.0865	0.1105	0.1370
- T	0.1185	0.1195	0.1180	0.1050	0.0980	0.0960	0.0895	0.1045	0.1070	0.1185	0.1490
						$\delta = 0.50$					
0	0.1220	0.1110	0.1090	0.0920	0.0950	0.0920	0.0825	0.0990	0.1010	0.1105	0.1205
- 1	0.1160	0.1025	0.1080	0.0935	0.0965	0.0755	0.0790	0.0930	0.0915	0.1080	0.1215
- 5	0.1160	0.1125	0.0965	0.0970	0.0990	0.0910	0.0915	0.0965	0.1030	0.1090	0.1265
-10	0.1260	0.1155	0.1125	0.1090	0.1050	0.1040	0.0985	0.0940	0.1005	0.1135	0.1210
-20	0.1330	0.1170	0.1065	0.1070	0.1035	0.0995	0.0980	0.1010	0.0960	0.1100	0.1310
-30	0.1245	0.1150	0.1075	0.1020	0.1025	0.0915	0.1005	0.1000	0.0890	0.1120	0.1360
- T	0.1285	0.1120	0.1060	0.1005	0.0960	0.0945	0.0985	0.1150	0.1180	0.1210	0.1530
						$\delta = 0.95$					
0	0.1180	0.1095	0.1030	0.0940	0.1005	0.1000	0.0965	0.0995	0.0990	0.1050	0.1165
- 1	0.1110	0.1100	0.0985	0.1015	0.1035	0.0980	0.0910	0.0945	0.0950	0.0980	0.1135
- 5	0.1080	0.1190	0.1100	0.1150	0.1195	0.1235	0.1130	0.1125	0.1075	0.1035	0.1140
-10	0.1145	0.1130	0.1095	0.1090	0.1205	0.1120	0.1100	0.1115	0.1030	0.1105	0.1190
-20	0.1280	0.1115	0.1110	0.1045	0.1010	0.1065	0.0985	0.1020	0.0930	0.1145	0.1295
-30	0.1250	0.1210	0.1065	0.0965	0.1045	0.1015	0.0885	0.1025	0.0980	0.1185	0.1335
- T	0.1325	0.1225	0.0905	0.1025	0.1045	0.0905	0.1095	0.1190	0.1190	0.1235	0.1555

Notes: Tables 3.5 and 3.6 report the rejection frequencies calculated using the size-adjusted Bonferroni correction. The level of the predictability test, η_2, is 10%. For $\delta = 0.25, 0.50, 0.75, 0.95$, $\eta_1 = 0.7, 0.5, 0.4, 0.3$, respectively. Rejection frequencies for $\delta = 0$ are calculated using simulated i.i.d. critical values.

Also, note that when $\delta = 0$ there is no need for size adjustment. Instead, simulated i.i.d. critical values are used for the t-tests. Similar to producing the lookup tables for nonzero δ, we compute the simulated i.i.d. critical values based on 49999 Monte Carlo replications with $T = 1000$ and $c = -T$ and store them in a lookup table. Consequently, as the first panel in Table 3.5 illustrates, the rejection frequencies range from 0.0845 to 0.1150 when $T = 1000$. When $T = 200$ the rejection frequencies are slightly larger.

The Bonferroni technique is also evaluated by simulations in the cases where the residual cross correlation is unknown. The data is generated under $\delta = 0.5$ and 0.9. The correlation coefficient δ is estimated by $\hat{\delta} = corr(\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t})$, where $\hat{\varepsilon}_{1t}$ is the residual from regressing x_t on a constant and its own lag and $\hat{\varepsilon}_{2t}$ is the residual from regressing y_t on a constant and x_{t-1}. The Bonferroni critical values are calculated by interpolation with respect to $\hat{\delta}$. The results are reported in Tables 3.7 and 3.8. The rejection frequencies are close to those from the cases with known δ. When $T = 1000$ the rejection frequencies are close to 10%. The test is more conservative for c of smaller absolute values. When $T = 200$ the rejection frequencies are close to 10% for the central part of the distribution. There is some moderate size distortion for the outer quantiles. The finite sample over-rejection problem is more significant for the extreme quantiles.

Table 3.7: Bonferroni correction: Finite-sample size with estimated δ (T = 1000)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0.50$					
0	0.0950	0.0915	0.0875	0.0885	0.1005	0.0935	0.0975	0.0920	0.0860	0.0925	0.0895
- 1	0.0980	0.0945	0.0865	0.0940	0.1000	0.0890	0.0975	0.0910	0.0875	0.0970	0.0825
- 5	0.0995	0.0910	0.0975	0.1025	0.1055	0.0940	0.1070	0.0950	0.1010	0.1010	0.0855
-10	0.1070	0.0965	0.0995	0.1025	0.1085	0.0965	0.1015	0.1045	0.1010	0.1025	0.0870
-20	0.1095	0.0980	0.0915	0.1010	0.1055	0.1030	0.0975	0.0975	0.0975	0.1030	0.0940
-30	0.0990	0.1025	0.0935	0.0975	0.1025	0.0990	0.0960	0.1015	0.0975	0.0975	0.0955
- T	0.1140	0.0940	0.1010	0.1055	0.1025	0.1010	0.1045	0.1080	0.1105	0.0935	0.0930
						$\delta = 0.90$					
0	0.0895	0.0965	0.0925	0.0950	0.0970	0.1005	0.0845	0.0865	0.0860	0.0920	0.0900
- 1	0.0920	0.0900	0.0875	0.0905	0.0915	0.0865	0.0870	0.0830	0.0880	0.0935	0.0855
- 5	0.0925	0.0915	0.0880	0.1050	0.0980	0.1040	0.1015	0.0925	0.1085	0.1045	0.0940
-10	0.0935	0.0980	0.0935	0.1075	0.1020	0.0965	0.0980	0.0955	0.1075	0.1020	0.0905
-20	0.0980	0.1020	0.0905	0.1030	0.1030	0.0930	0.0950	0.0925	0.0990	0.1045	0.0960
-30	0.0980	0.1045	0.0940	0.1035	0.1025	0.0910	0.0980	0.0975	0.0950	0.1020	0.0990
- T	0.1085	0.1030	0.1020	0.1060	0.1125	0.1005	0.1050	0.1125	0.1020	0.1075	0.0910

Table 3.8: Bonferroni correction: Finite-sample size with estimated δ (T = 200)

c	$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
						$\delta = 0.50$					
0	0.1245	0.1115	0.1090	0.0920	0.0975	0.0910	0.0865	0.1005	0.1005	0.1145	0.1240
- 1	0.1180	0.1065	0.1090	0.0925	0.0965	0.0750	0.0800	0.0930	0.0935	0.1095	0.1245
- 5	0.1175	0.1135	0.0990	0.0940	0.0970	0.0890	0.0915	0.0985	0.1055	0.1080	0.1270
-10	0.1255	0.1150	0.1120	0.1085	0.1020	0.1030	0.0985	0.0950	0.1020	0.1135	0.1210
-20	0.1320	0.1180	0.1065	0.1070	0.1040	0.1010	0.0945	0.1000	0.0970	0.1105	0.1305
-30	0.1250	0.1155	0.1075	0.1020	0.1040	0.0925	0.1010	0.1005	0.0905	0.1150	0.1350
- T	0.1305	0.1115	0.1065	0.1025	0.0970	0.0945	0.0975	0.1155	0.1185	0.1215	0.1540
						$\delta = 0.90$					
0	0.1120	0.1105	0.1035	0.1035	0.1030	0.0995	0.0920	0.0940	0.1050	0.1060	0.1160
- 1	0.1145	0.1135	0.1000	0.1050	0.1050	0.0940	0.0920	0.0975	0.0995	0.0995	0.1175
- 5	0.1130	0.1150	0.1095	0.1165	0.1240	0.1185	0.1195	0.1150	0.1090	0.1105	0.1155
-10	0.1125	0.1080	0.1115	0.1090	0.1175	0.1145	0.1080	0.1090	0.0970	0.1135	0.1205
-20	0.1195	0.1120	0.1055	0.0995	0.1000	0.1080	0.0980	0.0995	0.0960	0.1110	0.1255
-30	0.1225	0.1195	0.1085	0.1000	0.0945	0.1070	0.0935	0.0975	0.0960	0.1145	0.1370
- T	0.1380	0.1215	0.0960	0.1020	0.1035	0.0915	0.1050	0.1185	0.1195	0.1265	0.1550

Notes: Tables 3.7 and 3.8 report the rejection frequencies calculated using the size-adjusted Bonferroni correction, assuming δ is unknown. The correlation coefficient δ is estimated by $\hat{\delta} = corr(\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t})$, where $\hat{\varepsilon}_{1t}$ and $\hat{\varepsilon}_{2t}$ are the residuals. The relevant Bonferroni critical values are calculated by interpolation with respect to $\hat{\delta}$.

3.5 Empirical Study

Chapter Three applies the Bonferroni technique to testing predictability at different points in the stock return distribution. The univariate quantile regressions of the return on 16 pre-determined predictors are analyzed by Cenesizoglu and Timmermann (2008) using the data from Goyal and Welch (2008). This chapter considers the same data set that comprises monthly observations over the period from February 1871 to December 2005 with the shortest series from May 1937 to December 2002. As in Cenesizoglu and Timmermann (2008), the stock return is measured by the S&P 500 index including dividends. The 16 predictor variables, including dividend price ratio, dividend yield, and earnings price ratio, are listed in Table 3.9.

To examine the properties of the predictor variables, we first run OLS regressions of each predictor on its own lags, as in equation (3.8). As shown in Table 3.9, the lag length is selected by the Bayes information criterion (BIC) with a maximum of twelve lags. In addition, the confidence interval of the local-to-unity parameter c for each predictor is calculated based on Stock (1991). From Table 3.9, the largest autoregressive root α_1 of the predictor is close to but smaller than one in many cases. For some predictors, such as dividend price ratio, dividend yield, and T-bill rate, the upper bounds of the 95% intervals are slightly larger than one. Also, default return spread, long term rate of return, stock variance, and inflation are close to stationary. Besides, the correlation coefficient δ is estimated by $\hat{\delta} = corr(\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t})$, where $\hat{\varepsilon}_{1t}$ is the residual from the ADF regression of equation (3.8) and $\hat{\varepsilon}_{2t}$ is the residual from the predictive regression of equation (3.1). As Table 3.9 shows, the estimate $\hat{\delta}$ is negative for most predictors. Moreover, the residual $\hat{\varepsilon}_{1t}$ is strongly correlated with $\hat{\varepsilon}_{2t}$ for dividend price ratio, earnings price ratio, smoothed earnings price ratio, and book to market ratio. In contrast, for default return spread, long term rate of return, and inflation, the cross correlation coefficients are positive and small.

Table 3.9: Predictor variables

Notation	Variable name	Sample period	Lag length	95% CI on c	$\hat{\delta}$
d/p	Dividend price ratio	02/1871-12/2005	2	[-20.2719, 2.3421]	-0.9645
d/y	Dividend yield	02/1871-12/2005	2	[-20.4209, 2.3135]	-0.1142
e/p	Earnings price ratio	02/1871-12/2005	2	[-, -8.3450]	-0.8987
e10/p	Smoothed earnings price ratio	12/1880-12/2005	7	[-32.9945 -3.9005]	-0.8634
b/m	Book to market ratio	03/1921-12/2005	4	[-19.9380, 2.4088]	-0.8243
tbl	T-bill rate	02/1920-12/2005	3	[-18.9895, 2.6330]	-0.0756
lty	Long term yield	01/1919-12/2005	4	[-7.0319, 4.3407]	-0.1481
tms	Term spread	02/1920-12/2005	2	[-, -29.0576]	-0.0043
dfy	Default yield spread	01/1919-12/2005	4	[-28.7040, 0.5644]	-0.2428
dfr	Default return spread	01/1926-12/2005	1	[-, -]	0.0889
csp	Cross sectional premium	05/1937-12/2002	1	[-25.1078, 1.3438]	-0.0463
ltr	Long term rate of return	01/1926-12/2005	1	[-, -]	0.1329
svar	Stock variance	02/1885-12/2005	8	[-, -]	-0.2966
d/e	Dividend payout ratio	02/1871-12/2005	3	[-, -27.3195]	-0.0238
ntis	Net equity expansion	12/1926-12/2005	1	[-, -13.1848]	-0.0785
infl	Inflation	02/1913-12/2005	4	[-, -]	0.0240

Notes: The data are the same as that considered by Goyal and Welch (2008) and Cenesizoglu and Timmermann (2008). The lag length is selected by the BIC with a maximum of twelve lags. The confidence intervals of c are calculated based on Stock (1991) with simulations for $-38 \leq c \leq 6$. The empty intervals, [-, -], result from the cases where $c < -38$, which implies stationarity. The residual cross correlation estimate is denoted by $\hat{\delta}$.

To test the predictability of different parts in the return distribution, we then run quantile regressions of the return on each predictor for eleven quantiles. The quantile regression coefficients are estimated for all the quantiles. The magnitudes of the coefficient estimates are the same as those from Cenesizoglu and Timmermann (2008), where they have discussed the asymmetric effects of the predictive variables on different positions in the return distribution. For our predictive tests, the relevant Bonferroni critical values are interpolated from the lookup tables with respect to $\hat{\hat{\delta}}$. The test results are reported in Table 3.10. The first row of each panel in Table 3.10 contains the slope coefficient estimates for eleven quantiles of the stock return distribution. The second row reports the corresponding t-statistics. At 10% and 5%[20] significance levels, using the Bonferroni correction we find evidence of predictability at various points in the stock return distribution for most predictor variables.

Some variables have significant effects on the outer quantiles of the return. For example, with default yield spread we find evidence of predictability for the tails and shoulders of the return distribution, but we find no evidence for the inner quantiles, such as $\tau = 0.4, 0.5, 0.6, 0.7$, at 5% level. Interestingly, the quantile regression coefficient estimates are negative and significant for the lower tail, but positive and significant for the upper tail. Similar evidence can be found with stock variance, which also has predictive power for the tails and shoulders, but has no significant effect on the center of the return distribution. The effects of stock variance on the return distribution are significant at least at 10% level. As a volatility measure computed as the sum of squared daily returns, stock variance affects the dispersion of the return distribution and thus captures a predictable component in the riskiness of the stock (Cenesizoglu and Timmermann 2008). When stock variance increases, the return distribution spreads out as expected. Consequently, it is more likely to gain a large positive return or to get a large loss. For smoothed earnings price ratio and book to market ratio, there is only evidence of predictability for the two tails of the return distribution. For default return spread, the null is only rejected at $\tau = 0.05$. For

[20] When the level of the predictability test, η_2, is 5%, for $\delta = 0.25, 0.50, 0.75, 0.95$, we choose $\eta_1 = 0.40, 0.30, 0.25, 0.20$, respectively.

net equity expansion, the null is rejected for the lower tail where $\tau = 0.05, 0.1, 0.2$. A rise in net equity expansion is accompanied by downward shifts of the lower quantiles, which suggests higher probability of large negative return. Thus, net equity expansion is an asymmetric tail predictor. Although many of the pre-determined variables cannot predict the central part of the return distribution which is consistent with the previous literature using linear models, they contain valuable information for other parts of the return distribution especially the tails which is found by using quantile regression. Similar empirical results are also discussed by Cenesizoglu and Timmermann (2008).

Other variables have more effects on the central part of the return distribution. For long term yield we reject the null for the center and right shoulder. For T-bill rate we reject when $\tau \geq 0.4$ and for inflation we reject when $\tau \geq 0.3$. At 5% level, among all these predictors, inflation can predict the most quantiles of the return distribution. Besides, T-bill rate, long term yield, default yield spread, and stock variance also have considerable predictive power for many of the quantiles, compared with the other variables. On the other hand, at 5% level, dividend price ratio, dividend yield, and earnings price ratio fail to predict any quantile of the return distribution. However, at 10% level, the proposed test indicates that dividend price ratio has predictive power only for the median of the stock return.

Table 3.10: Quantile predictive tests of stock return

		$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
d/p	$\hat{\gamma}_1(\tau)$	-0.696	-0.313	-0.245	0.043	-0.234	-0.461†	-0.321	-0.036	-0.038	0.059	0.670
	$t_{\hat{\gamma}_1(\tau)}$	-1.141	-0.614	-0.681	0.127	-0.685	-1.377	-0.919	-0.098	-0.105	0.172	1.748
	p-value	0.254	0.539	0.496	0.899	0.493	0.169	0.358	0.922	0.916	0.863	0.081
d/y	$\hat{\gamma}_1(\tau)$	-0.316	0.155	-0.153	0.124	-0.078	-0.407	-0.254	0.043	-0.037	0.056	0.122
	$t_{\hat{\gamma}_1(\tau)}$	-0.512	0.277	-0.417	0.365	-0.225	-1.195	-0.724	0.118	-0.100	0.164	0.302
	p-value	0.608	0.782	0.677	0.715	0.822	0.232	0.469	0.906	0.920	0.870	0.763
e/p	$\hat{\gamma}_1(\tau)$	1.363	0.962	-0.060	0.280	0.380	-0.023	0.413	0.681	0.222	-0.008	-0.261
	$t_{\hat{\gamma}_1(\tau)}$	1.946	1.736	-0.146	0.714	0.975	-0.058	1.074	1.803	0.572	-0.019	-0.463
	p-value	0.052	0.083	0.884	0.475	0.329	0.954	0.284	0.071	0.568	0.985	0.644
e10/p	$\hat{\gamma}_1(\tau)$	-0.383*	-0.112	-0.059	0.015	0.033	0.005	0.114	0.161*	0.159†	0.281*	0.419*
	$t_{\hat{\gamma}_1(\tau)}$	-2.328	-0.936	-0.666	0.217	0.481	0.073	1.658	2.463	2.235	2.938	3.373
	p-value	0.020	0.349	0.505	0.828	0.631	0.942	0.097	0.014	0.025	0.003	0.001
b/m	$\hat{\gamma}_1(\tau)$	-4.762*	-2.280†	-1.034	-0.898	-0.711	-0.607	0.595	1.097	1.629	2.135	4.929*
	$t_{\hat{\gamma}_1(\tau)}$	-2.236	-1.628	-1.099	-1.021	-0.856	-0.749	0.742	1.424	2.021	2.106	4.260
	p-value	0.025	0.104	0.272	0.308	0.392	0.454	0.458	0.155	0.043	0.035	0.000
tbl	$\hat{\gamma}_1(\tau)$	0.115	0.014	-0.043	-0.104†	-0.146*	-0.139*	-0.135*	-0.158*	-0.127*	-0.190*	-0.225*
	$t_{\hat{\gamma}_1(\tau)}$	1.256	0.173	-0.619	-1.630	-2.428	-2.259	-2.072	-2.377	-2.055	-3.265	-3.004
	p-value	0.209	0.863	0.536	0.103	0.015	0.024	0.038	0.017	0.040	0.001	0.003
lty	$\hat{\gamma}_1(\tau)$	0.192	0.100	0.009	-0.086	-0.119*	-0.114*	-0.139*	-0.123*	-0.120*	-0.180*	-0.099
	$t_{\hat{\gamma}_1(\tau)}$	1.454	1.045	0.127	-1.338	-1.931	-1.832	-2.216	-1.927	-1.903	-2.689	-0.955
	p-value	0.146	0.296	0.899	0.181	0.053	0.067	0.027	0.054	0.057	0.007	0.340
tms	$\hat{\gamma}_1(\tau)$	-0.059	0.150	0.342*	0.126	0.201	0.126	0.091	0.081	0.172	0.511*	0.752*
	$t_{\hat{\gamma}_1(\tau)}$	-0.187	0.648	2.110	0.893	1.587	1.003	0.696	0.572	1.136	2.846	3.533
	p-value	0.852	0.517	0.035	0.372	0.113	0.316	0.486	0.567	0.256	0.004	0.000
dfy	$\hat{\gamma}_1(\tau)$	-4.145*	-2.946*	-1.475*	-0.942*	-0.518	0.039	0.502	0.661†	1.323*	2.108*	3.465*
	$t_{\hat{\gamma}_1(\tau)}$	-7.088	-4.850	-2.812	-2.207	-1.400	0.108	1.410	1.840	2.873	4.483	5.007
	p-value	0.000	0.000	0.005	0.027	0.162	0.914	0.159	0.066	0.004	0.000	0.000

(Table 3.10 cont'd)

		$\tau = 0.05$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
dfy	$\hat{\gamma}_1(\tau)$	-4.145*	-2.946*	-1.475*	-0.942*	-0.518	0.039	0.502	0.661†	1.323*	2.108*	3.465*
	$t_{\hat{\gamma}_1(\tau)}$	-7.088	-4.850	-2.812	-2.207	-1.400	0.108	1.410	1.840	2.873	4.483	5.007
	p-value	0.000	0.000	0.005	0.027	0.162	0.914	0.159	0.066	0.004	0.000	0.000
dfr	$\hat{\gamma}_1(\tau)$	0.713*	0.174	0.270	0.139	0.082	0.093	-0.079	0.072	0.121	0.205	0.178
	$t_{\hat{\gamma}_1(\tau)}$	4.340	0.727	1.227	0.646	0.366	0.407	-0.376	0.363	0.641	0.967	0.711
	p-value	0.000	0.467	0.220	0.518	0.714	0.684	0.707	0.717	0.521	0.334	0.477
csp	$\hat{\gamma}_1(\tau)$	3.807*	2.391†	0.489	2.121*	2.065*	1.196	1.435†	2.298*	1.634†	2.022*	1.402
	$t_{\hat{\gamma}_1(\tau)}$	2.952	1.962	0.451	2.268	2.380	1.407	1.657	2.675	1.962	2.378	1.406
	p-value	0.003	0.050	0.652	0.023	0.017	0.160	0.098	0.008	0.050	0.017	0.160
ltr	$\hat{\gamma}_1(\tau)$	-0.174	0.079	0.034	-0.023	0.006	0.047	0.140*	0.098	0.068	0.180	0.250*
	$t_{\hat{\gamma}_1(\tau)}$	-0.784	0.706	0.341	-0.261	0.065	0.576	1.922	1.352	0.902	1.557	1.819
	p-value	0.433	0.480	0.733	0.794	0.948	0.565	0.055	0.176	0.367	0.120	0.069
svar	$\hat{\gamma}_1(\tau)$	-10.261†	-4.142†	-2.663*	-2.264*	-1.267*	-0.614	0.249	2.180*	2.745*	6.195†	8.517*
	$t_{\hat{\gamma}_1(\tau)}$	-2.035	-2.029	-3.778	-5.498	-3.838	-0.576	0.261	2.172	4.156	1.927	11.924
	p-value	0.042	0.043	0.000	0.000	0.000	0.565	0.794	0.030	0.000	0.054	0.000
d/e	$\hat{\gamma}_1(\tau)$	-0.040*	-0.028*	-0.009	-0.007	-0.009†	-0.014*	-0.014*	-0.014*	-0.004	0.004	0.020†
	$t_{\hat{\gamma}_1(\tau)}$	-3.971	-3.120	-1.540	-1.444	-1.810	-2.916	-2.699	-2.606	-0.649	0.638	1.976
	p-value	0.000	0.002	0.124	0.149	0.070	0.004	0.007	0.009	0.517	0.523	0.048
ntis	$\hat{\gamma}_1(\tau)$	-0.466*	-0.465*	-0.340*	-0.171	-0.153	-0.111	-0.076	-0.068	-0.034	-0.017	0.005
	$t_{\hat{\gamma}_1(\tau)}$	-3.435	-3.087	-2.392	-1.274	-1.285	-0.963	-0.723	-0.732	-0.390	-0.192	0.053
	p-value	0.001	0.002	0.017	0.203	0.199	0.335	0.470	0.464	0.697	0.848	0.958
infl	$\hat{\gamma}_1(\tau)$	0.312	-0.367	-0.493	-1.020*	-1.100*	-0.935*	-0.946*	-0.706*	-0.604*	-0.824*	-1.116*
	$t_{\hat{\gamma}_1(\tau)}$	1.056	-1.297	-1.590	-4.103	-4.699	-3.660	-3.458	-2.413	-2.045	-2.222	-2.467
	p-value	0.291	0.195	0.112	0.000	0.000	0.000	0.001	0.016	0.041	0.026	0.014

Notes: For each quantile τ, $\hat{\gamma}_1(\tau)$ denotes the slope coefficient estimates and $t_{\hat{\gamma}_1(\tau)}$ denotes the corresponding t-statistics. The coefficient estimates for d/p, d/y, e/p, and b/m are multiplied by 100. * denotes significance at 5% level and † denotes significance at 10% level from the tests using Bonferroni correction. The p-values are from the standard t-tests.

The p-values in the third row of each panel in Table 3.10 are from the standard t-tests. In many cases, our test results are similar to the results from the standard t-tests. However, there are some discrepancies between the proposed test and conventional t-test. For example, at 10% level, with dividend price ratio, earnings price ratio, smoothed earnings price ratio, and book to market ratio, conventional t-tests over-reject for some quantiles due to the large residual cross correlations. This corresponds to our simulation results: using the Bonferroni procedure with the simulated critical values is especially beneficial when the residual cross correlation is large. At 5% level, with smoothed earnings price ratio, book to market ratio, cross sectional premium, stock variance, and dividend earnings ratio, standard t-tests also over-reject for some parts of the return distribution. On the contrary, with long term yield and long term rate of return, our tests over-reject, which corresponds to the simulation results in the cases where $\hat{\delta}$ is small.

Furthermore, our empirical study has some similarities to that conducted by Cenesizoglu and Timmermann (2008). In their paper, the Bonferroni p-values are used as a "summary measure" in a joint test across all the quantiles of interest. They consider the null hypothesis that a given predictor does not predict any of the quantiles of the return distribution. According to their findings, only eight of the predictors are significant at 5% level. In particular, the p-values are one for dividend price ratio and dividend yield. It means that neither of the two variables has effect on any points in the return distribution. Unlike their approach, the Bonferroni methodology we proposed is used for a different purpose. Instead of testing the significance of each predictor across all quantiles jointly, we adopt a Bonferroni bounds method in the local-to-unity context and test whether the quantile regression coefficient is zero for each individual τ. At 10% level, we find evidence of predictability for the median of the return distribution with dividend price ratio. However, at 5% level, our test indicates that dividend price ratio and dividend yield are insignificant, which is consistent with the results from Cenesizoglu and Timmermann (2008). In addition, our results also show that earnings price ratio has no predictive power for the return distribution.

3.6 Conclusion

Chapter Three develops inference in the predictive quantile regression with a nearly nonstationary regressor. We consider the quantile regression of a dependent variable y_t on a lagged regressor x_{t-1}, which is possibly strongly autocorrelated and not strictly exogenous. We derive the limiting distributions of the quantile regression coefficient and its corresponding t-statistic under the local-to-unity specification such that $\alpha_1 = 1 + \frac{c}{T}$. The asymptotic distributions depend on the nuisance parameter c, the residual cross correlation coefficient δ, and the quantile level τ. According to the asymptotic representation of the test statistic, we create computerized lookup tables consisting of simulated critical values for a number of (c, δ, τ) sets. We also present an asymptotically valid Bonferroni bounds methodology to test whether x_{t-1} can predict the τth quantile of the conditional distribution of y_t using the critical values from our lookup tables. The simulation results based on 2000 Monte Carlo replications show that it is worthwhile to adopt the Bonferroni procedure to correct the size distortion especially when there is considerable residual cross correlation. Furthermore, despite the substantial literature on the predictability of the mean of stock return, few research has studied the predictability of the entire return distribution. Thus, we apply the asymptotic theory and Bonferroni technique to testing the predictability of different points in the return distribution using 16 pre-determined predictor variables. At 5% significance level, except for dividend price ratio, dividend yield, and earnings price ratio, all of the predictors have asymmetric effects on the return distribution.

Chapter 4

CUSUM Test for Quantile Cointegration

4.1 Introduction

The cointegration methodology developed by Engle and Granger (1987) has given rise to numerous studies of long run co-movements among nonstationary economic variables. There is a substantial literature on cointegration tests for time series models. Traditional cointegration testing procedures introduced by Engle and Granger (1987) and Johansen (1988) are commonly used in bivariate and multivariate settings. Engle and Granger (1987) propose a unit root type cointegration test on the residuals from the cointegrating regression, which allows for heteroskedasticity. If the Dickey-Fuller test rejects that the residuals contain a unit root, then we reject the null hypothesis of no cointegration. In the case of serial correlation, the augmented Dickey-Fuller test can be employed. Such residual based test is simple to implement, but the asymptotic distribution of the test statistic is nonstandard. Phillips and Ouliaris (1990) analyze the asymptotic properties of the Engle-Granger type tests. Both Engle and Granger (1987) and Phillips and Ouliaris (1990) report critical values for the tests. More accurate simulated critical values of the unit root and cointegration tests are provided by MacKinnon (1991, 2010). In contrast, Johansen (1988, 1991, 1995) proposes maximum likelihood test statistics that follow multivariate unit root distributions.

Among various approaches to investigating cointegration relationships among time series, the residual based Engle-Granger type tests have been popular due to their computational convenience. For example, in a bivariate setting, the residuals $\hat{u}_t$ are calculated from the OLS regression of y_t on x_t. If the errors (residuals $\hat{u}_t$, in practice) follow an I(0) process, then y_t and x_t are cointegrated. If the error process has a unit root, then we have a "spurious" regression. In this case, the residual is nonstationary, thus there is no long run equilibrium relationship between y_t and x_t.

In the Engle-Granger type tests, the null hypothesis is no cointegration and the alternative is cointegration. However, since long run equilibrium relationship is of particular interest to economists, some authors, such as Park et al. (1988), Park (1990), and Shin (1994), focus on the null hypothesis of cointegration by using residual based procedures. More recently, Xiao and Phillips (2002) apply the conventional cumulated sum (CUSUM) test for structural change to cointegrating regression residuals and develop a consistent residual based test for the null of cointegration. Particularly, the fully modified approach proposed by Phillips and Hansen (1990) is used for serial correlation and endogeneity corrections. To examine the fluctuation in the equilibrium error from the cointegrating regression, Xiao and Phillips (2002) calculate the cumulative sum of the residuals from the fully modified OLS regression. Under the null of cointegration, the residuals are expected to follow a stable process and the cumulative sum test statistic converges to a functional of Brownian motions that is free of nuisance parameters. Under the alternative when there is no cointegration relationship, the fluctuations in the residuals should have larger order of magnitude and the test statistic diverges to infinity asymptotically. This robust test is easy to implement and has good finite sample performance. Furthermore, Xiao (2009) extends the cointegration methodology to quantile regression. In the quantile cointegrating regression, leads and lags of the integrated regressors are included in the model to account for endogeneity. The cumulative sum of the resulting residuals has the same asymptotic behavior as that from the fully modified regression considered by Xiao and Phillips (2002). However, in the dynamic model, selecting the lengths of leads and lags can be an issue.

Tests based on the CUSUM statistics were originally introduced to investigate whether a regression relationship is stable over time (Brown and Durbin 1968). Brown et al. (1975) test constancy of the regression coefficients based on the cumulated sum of squared recursive residuals. Ploberger and Kramer (1992) apply the CUSUM test of parameter stability to ordinary least squares residuals from a stationary model. They study the limiting null distribution and local power of the resulting test statistic. Hao and Inder (1996) propose a diagnostic OLS based CUSUM test for structural change in the context of cointegrated regression models. In particular, they derive the asymptotic distribution of the CUSUM statistic and tabulate its critical values. Although used for different purposes, it should be emphasized that the cointegration models studied by Hao and Inder (1996) and Xiao and Phillips (2002) have the same behavior under the null hypotheses, but have different behavior under the alternatives.

This chapter extends the analysis of Xiao and Phillips (2002) to the case of conditional quantiles, since the long run relationship among nonstationary time series may not be uniform. In practice, locations in the distribution other than the mean may matter for cointegration analysis. To examine the equilibrium relationships across different quantiles of the distribution of the response variable, the CUSUM test is employed to test the null hypothesis of quantile cointegration. To make the model parsimonious, instead of including leads and lags of the nonstationary regressors as in the quantile cointegrating regression considered by Xiao (2009), this chapter exploits the Phillips-Hansen type fully modified estimator. Similar to the OLS regression, in quantile regression with I(1) regressors, it is reasonable to expect serial cross correlation and endogeneity. In the presence of serial correlation between the regression disturbance and the innovation of the integrated regressors, the quantile estimator for the cointegrating coefficients is second-order biased. Also, due to long run endogeneity in the data, the limit distribution of the cointegrating coefficient estimates depends on nuisance parameters. In this case, it is difficult to make inference. As suggested by Phillips and Hansen (1990) and Phillips and Loretan (1991), the fully modified estimator displays superior properties than those of the usual estimator[21]. The fully modified quantile estimator can effectively correct serial correlation and endogeneity. The resulting limit distribution of the regression coefficients is mixed normal so that it can provide a useful standard inference procedure, as in Xiao and Phillips (2002). The CUSUM test statistic is composed of partial sums of the residuals from the fully modified quantile regression. Under the null of quantile cointegration, the test statistic converges to a functional of Brownian motions in the limit, which is the same as in the fully modified OLS regression case.

The seminal paper of Campbell and Shiller (1987) demonstrates that present value models of the term structure imply cointegration of short and long term interest rates. A great number of previous studies, such as Engle and Granger (1987), Stock and Watson (1988),

[21] Phillips and Loretan (1991) show the advantage of the fully modified OLS estimator over the usual OLS estimator by Monte Carlo simulations. In the quantile regression case, as shown in Chapter Two, the finite sample bias and root mean square error of the fully modified quantile estimator are significantly smaller than those of the usual conditional quantile estimator.

Boothe (1991), Hansen (1992), Hall et al. (1992), Mandeno and Giles (1995), and Downing and Oliner (2007) among others, also discuss the expectations theory of the term structure of interest rates. A more complete review of work on the expectations hypothesis of the term structure is provided in Iacone (2009). These papers consider various U.S. interest rate series, including different yield series on the federal funds rate, Treasury bill rate, and commercial paper rate. Many empirical papers find that the interest rate process is I(1) and test the hypothesis of cointegration of interest rates of different terms to maturity. The results are mixed. This chapter applies the residual based quantile cointegration test to several sets of U.S. interest rate data. The quantile version of the CUSUM test rejects the expectations hypothesis of the term structure in certain quantiles of the conditional distributions of the interest rates.

The remainder of Chapter Four is organized as follows: In Section 4.2, the model is set up. The asymptotic theory for the fully modified conditional quantile estimator and test statistic is developed. In Section 4.3, the empirical application is discussed. Section 4.4 concludes the chapter.

4.2 Theory

4.2.1 The model

Let $\{w_t\}$ be an m-vector time series generated by $w_t = \Phi d_t + w_t^s$ for $t = 1, ..., T$, where Φ is a coefficient, d_t is the deterministic trend such that

$$d_t = \begin{cases} 0 & \text{no trend} \\ 1 & \text{constant term} \\ (1,t)' & \text{linear trend,} \end{cases} \tag{4.1}$$

and w_t^s is the stochastic component such that $w_t^s = w_{t-1}^s + \xi_t$ with $w_0^s = 0$, where ξ_t satisfies Assumption L, as in Xiao and Phillips (2002)[22]. We have

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \xi_t \Rightarrow B_w(r) = BM(0, \Omega), \tag{4.2}$$

where $B_w(r)$ is a vector of Brownian motions with covariance matrix

$$\Omega = C(1)\Sigma_\varepsilon C(1)' = \begin{bmatrix} \omega_{yy}^2 & \Omega_{yx} \\ \Omega_{xy} & \Omega_{xx} \end{bmatrix}.$$

Consider y_t as a scalar, x_t as a k-dimensional vector, and $m = k + 1$. We have

$$w_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix}, \qquad \xi_t = \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix}, \qquad B_w(r) = \begin{bmatrix} B_y(r) \\ B_x(r) \end{bmatrix}. \tag{4.3}$$

If $u_t = y_t^s - \beta' x_t^s$ is stationary with continuous spectral density $f_{uu}(\lambda)$, then y_t^s and x_t^s are cointegrated. In linear regression $y_t = \alpha' d_t + \beta' x_t + u_t = \theta' z_t + u$, where $\theta = (\alpha', \beta')'$ and $z_t = (d_t', x_t')'$, the residuals are $\hat{u}_t = y_t - \hat{\theta}' z_t$.

[22]Assumption L: $\xi_t = C(L)\varepsilon_t$, where ε_t is a white noise process with zero mean and variance matrix $\Sigma_\varepsilon > 0$, and $C(L) = \sum_{j=0}^{\infty} C_j L^j$, $C(1) \neq 0$, and $\sum_{j=1}^{\infty} j^2 |C_j| < \infty$.

Unit root type cointegration tests involve the null hypothesis of no cointegration. If the Dickey-Fuller test rejects that the cointegrating regression residual process has a unit root, then we reject the null of no cointegration. Alternatively, one can employ a CUSUM cointegration test, as in Xiao and Phillips (2002). By calculating the partial sum of the residuals from the cointegrating regression, the residual-based procedure tests the null hypothesis of cointegration. Under the null, the test statistic converges to a functional of Brownian motions, while under the alternative of no cointegration, the statistic diverges. Since there is serial correlation and endogeneity in the regression with integrated regressors, the fully modified estimator should be used to remove bias and nuisance parameters. In particular, for the linear model with an intercept and no time trend, the fully modified residuals are computed as $\hat{u}_t^+ = y_t^+ - \hat{\alpha} - x_t'\hat{\beta}^+$, where $y_t^+ = y_t - \hat{\Omega}_{ux}\hat{\Omega}_{xx}^{-1}\Delta x_t$ and $\hat{\beta}^+$ is the fully modified OLS estimator of β. In this case, the test statistic of interest is $CS_T = \max_{n=1,\ldots,T} \frac{1}{\hat{\omega}_{u.x}\sqrt{T}} |\sum_{t=1}^{n} \hat{u}_t^+|$, where $\hat{\omega}_{u.x}^2 = \hat{\omega}_u^2 - \hat{\Omega}_{ux}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{xu}$ with consistent kernel estimates $\hat{\omega}_u^2$, $\hat{\Omega}_{ux}$, $\hat{\Omega}_{xx}$, and $\hat{\Omega}_{xu}$ for the elements from the long run covariance matrix of $(u_t, \xi_{2t}')'$. The critical values of the CUSUM statistic are provided in Hao and Inder (1996) and Xiao and Phillips (2002).

Least squares methods are confined to estimating conditional mean functions for linear models. In order to account for the nonlinearity in the long run relationship, the analysis is generalized to the entire conditional distribution of y_t, which is represented by quantile τ with $\tau \in (0, 1)$. Denote $\mathcal{F}_t$ as the information set up to time t, so $x_t \in \mathcal{F}_t$. Let $F(\cdot)$ and $F_t(\cdot) = Pr(u_t < \cdot|\mathcal{F}_t)$ be the unconditional and conditional cumulative distribution functions of u_t and define the τth unconditional and conditional quantiles of u_t by $Q_{u_t} = F^{-1}(\tau)$ and $Q_{u_t}(\tau|\mathcal{F}_t) = F_t^{-1}(\tau)$, respectively. With the additional assumption that $F_t(\cdot) = F(\cdot)$ for all $1 \leq t \leq T$, the conditional quantile of y_t is $Q_{y_t}(\tau|\mathcal{F}_t) =$

$\alpha' d_t + F^{-1}(\tau) + \beta' x_t$ for $t = 1, ..., T$. Under this restrictive specification, all quantiles share the same slope coefficient β. However, the slope coefficient may depend on the innovation process u_t. In this case, the conditional quantile of the error term can be modeled as $Q_{u_t}(\tau|\mathcal{F}_t) = d_t'(\alpha(\tau) - \alpha) + x_t'(\beta(\tau) - \beta)$. Then, the conditional quantile function of y_t is $Q_{y_t}(\tau|\mathcal{F}_t) = \alpha'(\tau) d_t + \beta'(\tau) x_t$. Define $u_t(\tau) = u_t - F_t^{-1}(\tau) = y_t - \alpha'(\tau) d_t - \beta'(\tau) x_t$ so that $Q_{u_t(\tau)}(\tau|\mathcal{F}_t) = 0$. Consequently, we have the following quantile regression model:

$$y_t = \alpha'(\tau) d_t + \beta'(\tau) x_t + u_t(\tau) = \theta'(\tau) z_t + u_t(\tau), \tag{4.4}$$

where $\theta(\tau) = (\alpha'(\tau), \beta'(\tau))'$. The regression coefficient vector $\theta(\tau)$ is quantile dependent so that it can capture the possibly nonlinear long run relationships between y_t and the regressors.

4.2.2 Estimation method

Consider the following objective function with asymmetric weights on positive and negative residuals:

$$S_T(\tau, \theta(\tau)) = \sum_{t=1}^{T} \rho_\tau(y_t - z_t'\theta(\tau)). \tag{4.5}$$

As in the previous chapters, we define $\rho_\tau(u) = u(\tau - I(u < 0))$ and $\psi_\tau(u) = \tau - I(u < 0)$. For the τth quantile, the unknown regression coefficients are estimated by minimizing the sum of asymmetrically weighted absolute residuals such that

$$\hat{\theta}(\tau) = \arg\min_{\theta(\tau)} S_T(\tau, \theta(\tau)), \tag{4.6}$$

where $\hat{\theta}(\tau) = (\hat{\alpha}'(\tau), \hat{\beta}'(\tau))'$.

4.2.3 Assumptions

The following assumptions are imposed in order to derive the asymptotic distribution of the test statistic.

Assumption 10 $(\psi_\tau(u_t(\tau)), \xi'_{2t})' = C_\tau(L)\varepsilon_{\tau t}$, *where* $\varepsilon_{\tau t}$ *is a white noise process with zero mean and variance matrix* $\Sigma_{\tau\varepsilon} > 0$ *and* $C_\tau(L) = \sum_{j=0}^{\infty} C_{\tau j} L^j$ *with* $C_\tau(1) \neq 0$ *and* $\sum_{j=1}^{\infty} j^2 |C_{\tau j}| < \infty$.

This assumption corresponds to Assumption L in Xiao and Phillips (2002). Note that $\psi_\tau(u_t(\tau))$ can be regarded as a variant of the error term $u_t(\tau)$ from the quantile regression.

Assumption 11 *The conditional distribution function,* $F_t(u) = Pr(u_t < u|\mathcal{F}_t)$, *is absolutely continuous and has a continuous density function* $f_t(u)$ *such that* $0 < f_t(F_t^{-1}(\tau)) < \infty$ *for* $t = 1, ..., T$.

Assumption 12 *The density function* $f_t(\upsilon_T)$ *is uniformly integrable for any sequence* $\upsilon_T \rightarrow F_t^{-1}(\tau)$ *and* $E[f_t^{\gamma}(F_t^{-1}(\tau))] < \infty$ *for some* $\gamma > 1$, *for* $t = 1, ..., T$.

As in Chapter Two, Assumptions 11 and 12 ensure that the conditional density function is uniformly continuous, bounded, and integrable in some neighborhood of the τth quantile.

4.2.4 Theoretical results

If the above assumptions hold, the estimates of the regression coefficients will converge to their true values. Let D_T be a diagonal matrix. The presence of D_T is due to the different convergence rates of the coefficients associated with the deterministic trend and integrated regressors. When the regression model contains an intercept such that $d_t = 1$, we have $D_T = diag(T^{\frac{1}{2}}, TI_k)$. In the case of a linear trend such that $d_t = (1, t)'$, we have

$diag(1, T^{-1})d_{[Tr]} \Rightarrow B_d(r) = (1, r)'$. Since the convergence rate of the coefficient on the trending variable is T, the diagonal matrix becomes $D_T = diag(T^{\frac{1}{2}}, TI_{k+1})$. Similar to Theorem 1 in Xiao (2009), under the null hypothesis of cointegration denoted H_0, the limit distribution of the coefficient estimates is given by the following representation:

Theorem 6 *Under H_0 and Assumptions 10-12, $\hat{\theta}(\tau)$ is a consistent estimator of $\theta(\tau)$ and*

$$D_T(\hat{\theta}(\tau) - \theta(\tau)) \Rightarrow \left[f(F^{-1}(\tau)) \int_0^1 B_z B_z'\right]^{-1} \left[\int_0^1 B_z dB_\psi + \bar{\Delta}_{x\psi}\right], \qquad (4.7)$$

where $B_z = (B_d', B_x')'$, $\bar{\Delta}_{v\psi} = (0, \Delta_{x\psi}')'$, and $\Delta_{x\psi} = \sum_{t=0}^{\infty} E(\xi_{2t}\psi_\tau(u_0(\tau)))$ is the one-sided long run covariance between ξ_{2t} and $\psi_\tau(u_t(\tau))$.

The asymptotic representation is composed of integrals of Brownian motions and a bias term. In particular, the asymptotic distribution of the cointegrating coefficient estimates $\hat{\beta}(\tau)$ is as follows:

$$T(\hat{\beta}(\tau) - \beta(\tau)) \Rightarrow \left[f(F^{-1}(\tau)) \int_0^1 \underline{B}_{xd}\underline{B}_{xd}'\right]^{-1} \left[\int_0^1 \underline{B}_{xd} dB_\psi + \Delta_{x\psi}\right], \qquad (4.8)$$

where $\underline{B}_{xd} = B_x - (\int_0^1 B_x d')(\int_0^1 dB_d')^{-1} B_d$ is a k-dimensional demeaned or detrended Brownian motion.

Due to serial cross correlation, the limit distribution of $\hat{\beta}(\tau)$ has a second-order bias term $\Delta_{x\psi}$. Moreover, the above limit distribution depends on the nuisance parameter Ω, since the distribution of $\int_0^1 \underline{B}_{xd} dB_\psi$ depends on the correlation between $B_x(r)$ and $B_\psi(r)$, which is unknown in general. In order to develop useful inference procedures, a fully modified estimator is required to correct the bias and endogeneity.

Following Phillips and Hansen (1990), define $\psi_\tau^+(u_t(\tau)) = \psi_\tau(u_t(\tau)) - \Omega_{\psi x}\Omega_{xx}^{-1}\xi_{2t}$ so that $\psi_\tau^+(u_t(\tau))$ is uncorrelated with ξ_{2t} and has variance $\omega_{\psi.x}^2 = \omega_\psi^2 - \Omega_{\psi x}\Omega_{xx}^{-1}\Omega_{x\psi}$. Then we have

$$T^{-1}\sum_{t=1}^{T} x_t\psi_\tau^+(u_t(\tau)) \Rightarrow \int_0^1 B_x dB_{\psi.x} + \Delta_{x\psi}^+$$
$$= \omega_{\psi.x}\int_0^1 B_x dW + \Delta_{x\psi}^+,$$

where $B_{\psi.x}(r) = B_\psi(r) - \Omega_{\psi x}\Omega_{xx}^{-1}B_x(r)$ is a Brownian motion independent of $B_x(r)$, $W(r) = B_{\psi.x}(r)/\omega_{\psi.x}$ is a standard Brownian motion, and $\Delta_{x\psi}^+ = \Delta_{x\psi} - \Omega_{\psi x}\Omega_{xx}^{-1}\Delta_{xx}$ is the one-sided long run covariance between ξ_{2t} and $\psi_\tau^+(u_t(\tau))$.

Let $\hat{\Delta}_{x\psi}^+ = \hat{\Delta}_{x\psi} - \hat{\Omega}_{\psi x}\hat{\Omega}_{xx}^{-1}\hat{\Delta}_{xx}$, where $\hat{\Omega}_{\psi x} = \hat{\Omega}'_{x\psi}$, $\hat{\Omega}_{xx}$, $\hat{\Delta}_{x\psi}$, and $\hat{\Delta}_{xx}$ are the kernel estimates,

$$\hat{\Omega}_{x\psi} = \sum_{h=-M}^{M} K(\frac{h}{M})\hat{\Gamma}_{x\psi}(h), \quad \hat{\Delta}_{x\psi} = \sum_{h=0}^{M} K(\frac{h}{M})\hat{\Gamma}_{x\psi}(h),$$
$$\hat{\Omega}_{xx} = \sum_{h=-M}^{M} K(\frac{h}{M})\hat{\Gamma}_{xx}(h), \quad \hat{\Delta}_{xx} = \sum_{h=0}^{M} K(\frac{h}{M})\hat{\Gamma}_{xx}(h),$$

where $\hat{\Gamma}_{x\psi}(h) = T^{-1}\sum_{t=1}^{T-h}\Delta x_t\psi_\tau(\hat{u}_{t+h}(\tau))$, $\hat{\Gamma}_{xx}(h) = T^{-1}\sum_{t=1}^{T-h}\Delta x_t\Delta x'_{t+h}$, $K(\cdot)$ is a kernel function, and M is the bandwidth. As suggested by Andrews (1991) and Xiao and Phillips (2002), the Bartlett kernel and a plug-in bandwidth are adopted such that

$$K(u) = \begin{cases} 1-|u| & \text{for } |u| \le 1 \\ 0 & \text{otherwise} \end{cases}$$

and $M = O(T^{\frac{1}{3}})$.

The regression coefficient estimates after the modification, $\hat{\theta}^+(\tau) = (\hat{\alpha}'(\tau), \hat{\beta}'^+(\tau))'$, are

$$\hat{\theta}^+(\tau) = \hat{\theta}(\tau) - \left[f(\widehat{F^{-1}(\tau)})\sum z_t z_t'\right]^{-1}\left[\sum z_t\hat{\Omega}_{\psi x}\hat{\Omega}_{xx}^{-1}\Delta x_t + T\bar{\hat{\Delta}}_{x\psi}^+\right], \tag{4.9}$$

where $\hat{\bar{\Delta}}^{+}_{x\psi} = (0, \Delta'^{+}_{x\psi})'$. In particular, the fully modified estimator of the coefficients associated with the I(1) regressors is given by

$$\hat{\beta}^{+}(\tau) = \hat{\beta}(\tau) - \left[f(\widehat{F^{-1}(\tau)}) \sum \underline{x}^{d}_{t} \underline{x}'^{d}_{t}\right]^{-1} \left[\sum \underline{x}^{d}_{t} \hat{\Omega}_{\psi x} \hat{\Omega}^{-1}_{xx} \Delta x_{t} + T\hat{\Delta}^{+}_{x\psi}\right], \quad (4.10)$$

where $\underline{x}^{d}_{t}$ denotes the demeaned or detrended regressors and $f(\widehat{F^{-1}(\tau)})$ is a nonparametric consistent estimator of the density function $f(F^{-1}(\tau))$. In practice, this density function can be estimated using the Gaussian kernel and Silverman's "rule-of-thumb" bandwidth.

Consequently, the fully modified estimate $\hat{\beta}^{+}(\tau)$ follows a mixed normal distribution in the limit such that

$$\begin{aligned} T(\hat{\beta}^{+}(\tau) - \beta(\tau)) &\Rightarrow \left[f(F^{-1}(\tau)) \int_0^1 \underline{B}_{xd} \underline{B}'_{xd}\right]^{-1} \omega_{\psi.x} \int_0^1 \underline{B}_{xd} dW \\ &\sim MN\left(0, \frac{\omega^2_{\psi.x}}{f(F^{-1}(\tau))^2} \left[\int_0^1 \underline{B}_{xd} \underline{B}'_{xd}\right]^{-1}\right), \end{aligned}$$

where $\omega^2_{\psi.x}$ is estimated by $\hat{\omega}^2_{\psi.x} = \hat{\omega}^2_{\psi} - \hat{\Omega}_{\psi x} \hat{\Omega}^{-1}_{xx} \hat{\Omega}_{x\psi}$, which is calculated by nonparametric kernel method. Therefore, the asymptotic distribution of $\hat{\theta}^{+}(\tau)$ is as follows:

Theorem 7 *Under H_0 and Assumptions 10-12, $\hat{\theta}^{+}(\tau)$ is a consistent estimator of $\theta(\tau)$ and*

$$D_T(\hat{\theta}^{+}(\tau) - \theta(\tau)) \Rightarrow \left[f(F^{-1}(\tau)) \int_0^1 B_z B'_z\right]^{-1} \omega_{\psi.x} \int_0^1 B_z dW \quad (4.11)$$

$$\sim MN\left(0, \frac{\omega^2_{\psi.x}}{f(F^{-1}(\tau))^2} \left[\int_0^1 B_z B'_z\right]^{-1}\right), \quad (4.12)$$

where $B_z = (B'_d, B'_x)'$ and $\omega^2_{\psi.x} = \omega^2_{\psi} - \Omega_{\psi x} \Omega^{-1}_{xx} \Omega_{x\psi}$.

In the linear regression case, the cumulated sum of the residuals, which is $T^{-\frac{1}{2}} \sum_{t=1}^{n} \hat{u}_t$, converges under the null of cointegration and diverges to infinity under the alternative. Analogously, in the quantile regression model, the cumulated sum of $\hat{\psi}_{\tau}(u_t(\tau))$ satisfies the following conditions:

$$\max_{n=1,\ldots,T} T^{-\frac{1}{2}} \left| \sum_{t=1}^{n} \hat{\psi}_\tau(u_t(\tau)) \right| = \begin{cases} O_p(1) & \text{under } H_0 \\ O_p(T^{\frac{3}{2}}) & \text{under } H_1. \end{cases} \tag{4.13}$$

Under H_0, the cumulated sum of the fully modified residuals converges to a functional of Brownian motions. For the quantile regression model, the residuals from the fully modified quantile regression is calculated as $\hat{u}_t^+(\tau) = y_t^+ - z_t' \hat{\theta}^+(\tau)$, where $y_t^+ = y_t - \hat{\Omega}_{\psi x} \hat{\Omega}_{xx}^{-1} \Delta x_t$. The CUSUM test statistic is given by

$$CS_T(\tau) = \max_{n=1,\ldots,T} \frac{1}{\hat{\omega}_{\psi.x}\sqrt{T}} \left| \sum_{t=1}^{n} \psi_\tau(\hat{u}_t^+(\tau)) \right|, \tag{4.14}$$

where $\psi_\tau(u_t^+(\tau)) = \tau - I(\hat{u}_t^+(\tau) < 0)$.

In particular,

$$T^{-\frac{1}{2}} \sum_{t=1}^{[Tr]} \psi_\tau(\hat{u}_t^+(\tau)) \Rightarrow B_{\psi.x} - \left[\int_0^1 dB_{\psi.x} B_z' \right] \left[\int_0^1 B_z B_z' \right]^{-1} \int_0^r B_z(s) \tag{4.15}$$

$$= \omega_{\psi.x} \left\{ W_1 - \left[\int_0^1 dW_1 S' \right] \left[\int_0^1 SS' \right]^{-1} \int_0^r S \right\}, \tag{4.16}$$

where $S(r) = (B_d'(r), W_2'(r))'$ and $W_1(r)$ and $W_2(r)$ are one and k-dimensional standard Brownian motions that are independent of each other.

Define

$$\underline{W}(r) = W_1 - \left[\int_0^1 dW_1 S' \right] \left[\int_0^1 SS' \right]^{-1} \int_0^r S.$$

Then, for a certain quantile τ, the asymptotic representation of the CUSUM test statistic is as follows:

Theorem 8 *Under H_0 and Assumptions 10-12,*

$$CS_T(\tau) \Rightarrow \sup_{0 \le r \le 1} |\underline{W}(r)|. \tag{4.17}$$

For each quantile level τ, the asymptotic distribution is the same as that from the linear case. Simulated critical values are tabulated in Tables 1 and 2 from Hao and Inder (1996) and Table 1 from Xiao and Phillips (2002). Hao and Inder (1996) also consider the CUSUM statistic, however, for testing the null hypothesis of parameter constancy. Under the null of no structural change, the test statistic from their model has the same asymptotic distribution as the cumulative sum statistic here. The limit distributions are different under the alternatives. The empirical size and power properties of the CUSUM test for cointegration are evaluated via simulation by Xiao and Phillips (2002). In the case of quantile regression, the finite sample properties of the test are similar to those from the fully modified OLS regression.

4.3 Empirical Study

The theoretical model can be applied to examining the term structure of interest rates, and, more specifically, to testing the hypothesis of cointegration of short and long term interest rates. The expectations theory of the term structure of interest rates suggests that interest rates should be cointegrated if can be characterized as I(1) processes.

According to the generalized-present-value (GPV) model, any long term yield can be expressed as a function of current and expected short term yields (Boothe 1991). Engle and Granger (1987) and Campbell and Shiller (1987) test the model of the term structure and find that the yields of domestic bonds that differ only by term to maturity are intimately linked. Since interest rates in the United States are usually characterized as nonstationary processes, the spread between these interest rates of different terms to maturity should be stationary if the expectations theory of the term structure holds. This means that short and long term rates are cointegrated.

Many papers have tested the expectations hypothesis for the term structure using U.S. yield series on the federal funds rate (Hansen 1992), Treasury bill rate (Stock and Watson 1988, Boothe 1991, Hansen 1992, Hall et al. 1992, Mandeno and Giles 1995), and commercial paper rate (Downing and Oliner 2007). In this chapter, three sets of data are considered. The data comprises daily observations of interest rates of different yields from the U.S. Department of the Treasury and the Board of Governors of the Federal Reserve System. The interest rate statistics include daily Treasury bill rates[23], U.S. government securities/Treasury constant maturities (nominal) interest rates[24], and financial commercial paper rates[25]. The frequency is one business day. Unavailable or missing observations of holidays are omitted. The longest series, which is the Treasury bill constant maturity rate, is from January 4, 1982 to April 30, 2012 and contains 7584 observations. The shortest series, which is the Treasury bill rate, is from January 2, 2002 to April 30, 2012 and contains 2585 observations.

[23]Data source is the U.S. Department of the Treasury. U.S. daily 4-, 13-, 26-week Treasury bill rates are from 01/02/2002 to 04/30/2012. The number of observation is 2585. Both bank discount and coupon equivalent rates are considered. Secondary market quotations on the most recently auctioned Treasury Bills are obtained at approximately 3:30 PM each business day by the Federal Reserve Bank of New York. "The Bank Discount rate is the rate at which a Bill is quoted in the secondary market and is based on the par value, amount of the discount and a 360-day year. The Coupon Equivalent, also called the Bond Equivalent, or the Investment Yield, is the bill's yield based on the purchase price, discount, and a 365- or 366-day year. The Coupon Equivalent can be used to compare the yield on a discount bill to the yield on a nominal coupon bond that pays semiannual interest." (U.S. Department of the Treasury, http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=billrates, 07 May 2012.)

[24]Data source is the Board of Governors of the Federal Reserve System. U.S. daily 3-, 6-, 12-month Treasury bill constant maturity rates are from 01/04/1982 to 04/30/2012. The number of observation is 7584. The market yields are quoted on investment basis and are interpolated by the Treasury from the daily Treasury yield curve. The yield curve relates the yield on a security to its time to maturity. It is based on the closing market bid yields on actively traded Treasury securities in the over-the-counter market. The market yields are calculated from composites of quotations obtained by the Federal Reserve Bank of New York. The constant maturity Treasury rates, or CMTs, are also referred to as the Treasury yield curve rates. The Treasury will restrict the use of negative input yields for securities used in deriving interest rates for the nominal constant maturity Treasury series. "Any CMT input points with negative yields will be reset to zero percent prior to use as inputs in the CMT derivation." (U.S. Department of the Treasury, http://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield, 07 May 2012.)

[25]Data source is the Board of Governors of the Federal Reserve System. U.S. daily 30-, 60-, and 90-day AA financial commercial paper rates are from 01/02/1997 to 04/30/2012. The number of observation is 3748.

The regression model contains an intercept term but no time trend ($d_t = 1$), as a result of which $y_t = \alpha(\tau) + \beta(\tau)x_t + u_t(\tau)$. Nine representative quantile levels, $\tau = 0.1, 0.2, ..., 0.9$, are considered. The bandwidth for the fully modified kernel estimator is $M = 2T^{\frac{1}{3}}$.

As a pretest to investigate the properties of the interest rate series, the augmented Dickey-Fuller (ADF) test does not reject that the logarithm of each rate has a unit root, but strongly rejects a unit root in the first difference of the series. Thus, the interest rates are I(1).

The estimation and test results are summarized in Tables 4.1 to 4.3. For each quantile τ, the fully modified coefficient estimate $\hat{\beta}^{+}(\tau)$ and the CUSUM test statistic $CS_T(\tau)$ are reported. Let * denote the case where the CUSUM test rejects the null of cointegration at 5% level and let ** denote rejection at 1% level. For each instrument, the evidence in support of the expectations hypothesis for the term structure is mixed over different quantiles, since the CUSUM test fails to reject cointegration in various positions over the interest rate distributions.

As Tables 4.1 and 4.2 show, short and long term bank discount and coupon equivalent Treasury bill rates and financial commercial rates are generally cointegrated in the central part of the distribution, but not in the tails. In the first panel of Table 4.1, 4-, 13-, and 26-week Treasury bill bank discount rates are cointegrated with each other for the inner quanitles. Also, the 90% quantile of the 13-week bank discount rate is cointegrated with the 4-week rate. Similarly, the 90% quantile of the 26-week rate and the 4-week rate are cointegrated. From the second panel of Table 4.1, the corresponding coupon equivalent Treasury bill rates follow similar pattern. There is evidence of cointegration when the quantile level is between 0.4 and 0.6. The upper quantiles of the 13- and 26-week coupon equivalent rates are cointegrated with the 4-week rate.

Table 4.1: Cointegration among Treasury bill rates (yields: 4, 13, 26 weeks)

y-x		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
						Bank Discount				
4-13	$\hat{\beta}^{+}(\tau)$	0.9277	0.9519	0.9657	0.9743	0.9823	0.9917	1.0041	1.0185	1.0306
	$CS_T(\tau)$	1.3405*	1.4009*	1.4586**	1.1296	0.5869	1.1693*	2.5175**	4.8285**	6.8210**
4-26	$\hat{\beta}^{+}(\tau)$	0.8977	0.9112	0.9283	0.9578	0.9699	0.9783	0.9972	1.0198	1.0446
	$CS_T(\tau)$	1.4817**	1.4740**	1.4375**	1.5938**	1.2855*	0.9978	1.1742*	1.1906*	1.6425**
13-4	$\hat{\beta}^{+}(\tau)$	0.9702	0.9798	0.9949	1.0044	1.0161	1.0258	1.0314	1.0433	1.0702
	$CS_T(\tau)$	8.5412**	4.9614**	3.3197**	1.3636*	0.6999	1.0054	1.3845*	1.2737*	1.1013
13-26	$\hat{\beta}^{+}(\tau)$	0.9725	0.9673	0.9662	0.9780	0.9891	0.9928	1.0019	1.0081	1.0124
	$CS_T(\tau)$	2.0737**	1.5570**	1.3513*	1.5347**	1.0337	0.9155	1.9900**	3.0591**	4.8101**
26-4	$\hat{\beta}^{+}(\tau)$	0.9564	0.9726	0.9982	1.0113	1.0242	1.0339	1.0543	1.0618	1.0496
	$CS_T(\tau)$	3.7237**	1.8984**	1.2797*	0.9947	1.1424	1.4907**	1.3302*	1.3294*	1.0870
26-13	$\hat{\beta}^{+}(\tau)$	0.9878	0.9921	0.9963	1.0029	1.0088	1.0145	1.0288	1.0234	1.0129
	$CS_T(\tau)$	4.8220**	2.9392**	1.8578**	0.8833	0.9543	1.4324**	1.1671	1.3898*	1.8279**
						Coupon Equivalent				
4-13	$\hat{\beta}^{+}(\tau)$	0.9210	0.9445	0.9587	0.9665	0.9748	0.9851	0.9998	1.0101	1.0238
	$CS_T(\tau)$	1.3901*	1.4046*	1.2724*	0.9406	0.7558	1.0460	3.4397**	4.3309**	6.8739**
4-26	$\hat{\beta}^{+}(\tau)$	0.8845	0.8960	0.9117	0.9390	0.9503	0.9604	0.9752	1.0020	1.0222
	$CS_T(\tau)$	1.4312**	1.5441**	1.4277**	1.3336*	1.1337	1.2139*	1.2262*	1.2755*	2.1178**
13-4	$\hat{\beta}^{+}(\tau)$	0.9777	0.9861	1.0015	1.0133	1.0231	1.0343	1.0406	1.0512	1.0768
	$CS_T(\tau)$	8.6429**	5.1868**	3.4836**	1.3518*	0.8191	1.0880	1.2547*	1.3962*	1.1125
13-26	$\hat{\beta}^{+}(\tau)$	0.9609	0.9555	0.9553	0.9673	0.9755	0.9798	0.9880	0.9927	0.9979
	$CS_T(\tau)$	2.3203**	1.5408**	1.2494*	1.4563**	1.0802	0.8867	1.7646**	3.1131**	4.6295**
26-4	$\hat{\beta}^{+}(\tau)$	0.9758	0.9926	1.0156	1.0320	1.0473	1.0552	1.0755	1.0807	1.0732
	$CS_T(\tau)$	3.8391**	2.2514**	1.5080**	1.3153*	1.3036*	1.3447*	1.3707*	1.3316*	0.9762
26-13	$\hat{\beta}^{+}(\tau)$	1.0004	1.0045	1.0087	1.0165	1.0218	1.0286	1.0412	1.0360	1.0269
	$CS_T(\tau)$	4.1117**	2.9041**	1.7448**	0.8042	1.0676	1.3622*	1.2202*	1.3708*	1.8768**

Notes: U.S. daily Treasury bill data is from 01/02/2002 to 04/30/2012. The number of observation is 2585. * denotes rejection of the null of cointegration at 5% level and ** denotes rejection at 1% level.

In Table 4.2, for the 30-, 60-, and 90-day financial commercial rates, the null of cointegration is retained around the median in most cases. However, the distributions of the 30- and 90-day rates are not cointegrated in either direction. In addition, the fully modified coefficient estimates $\hat{\beta}^+(\tau)$ are highly significant in the cases when cointegration is not rejected.

Table 4.2: Cointegration among financial commercial paper rates (yields: 30, 60, 90 days)

y-x		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
30-60	$\hat{\beta}^+(\tau)$	1.0056	1.0036	1.0026	1.0039	1.0041	1.0041	1.0041	1.0052	1.0156
	$CS_T(\tau)$	1.5898**	2.4501**	2.0793**	1.1205	1.1426	1.7477**	3.7407**	6.6798**	4.6846**
30-90	$\hat{\beta}^+(\tau)$	1.0069	1.0062	1.0047	1.0061	1.0082	1.0096	1.0090	1.0101	1.0326
	$CS_T(\tau)$	1.5954**	2.3190**	2.5919**	2.3279**	1.7541**	1.8461**	2.5133**	4.4216**	2.9653**
60-30	$\hat{\beta}^+(\tau)$	0.9828	0.9947	0.9961	0.9962	0.9962	0.9963	0.9964	0.9950	0.9926
	$CS_T(\tau)$	6.0914**	6.9691**	4.4223**	2.2500**	1.2470*	1.1233	1.6134**	2.0012**	1.7048**
60-90	$\hat{\beta}^+(\tau)$	1.0095	1.0063	1.0043	1.0041	1.0043	1.0042	1.0047	1.0049	1.0102
	$CS_T(\tau)$	1.4585**	2.3730**	1.7650**	1.0464	0.9105	1.5000**	3.8005**	6.4077**	6.6321**
90-30	$\hat{\beta}^+(\tau)$	0.9656	0.9885	0.9893	0.9903	0.9922	0.9936	0.9928	0.9906	0.9898
	$CS_T(\tau)$	3.3449**	4.2920**	2.7547**	1.9796**	1.4799**	2.1144**	2.3797**	2.1914**	1.6938**
90-60	$\hat{\beta}^+(\tau)$	0.9877	0.9950	0.9954	0.9960	0.9960	0.9943	0.9940	0.9938	0.9896
	$CS_T(\tau)$	7.7150**	7.0606**	4.1914**	1.8809**	0.8963	1.5375**	1.4186*	2.0230**	1.9401**

Notes: U.S. daily financial commercial paper rate data is from 01/02/1997 to 04/30/2012. The number of observation is 3748.

The results from the model with the Treasury constant maturities interest rates or the Treasury yield curve rates are different. In many cases, the null of cointegration is retained in either the lower or upper tail. In Table 4.3, the Treasury constant maturities interest rates of shorter terms, such as the 3- and 6-month rates, are cointegrated with rates of longer terms in the lower tail of the distribution. The 6-month constant maturities rate is cointegrated with the 3-month rate in the upper tail. Similarly, there is cointegration relationship between the 90% quantile of the 12-month rate and the 6-month rate. However, the distribution of the 12-month rate is not cointegrated with the 3-month rate. Also, when cointegration is not rejected, the fully modified coefficient estimates $\hat{\beta}^{+}(\tau)$ are highly significant.

Table 4.3: Cointegration among Treasury constant maturity rates (yields: 3, 6, 12 months)

y-x		$\tau = 0.1$	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3-6	$\hat{\beta}^{+}(\tau)$	0.9465	0.9605	0.9660	0.9687	0.9730	0.9780	0.9828	0.9880	0.9924
	$CS_T(\tau)$	0.8859	1.1161	1.3232*	1.2241*	1.2130*	1.2300*	1.1889*	1.3899*	1.4494**
3-12	$\hat{\beta}^{+}(\tau)$	0.9071	0.9338	0.9462	0.9499	0.9560	0.9642	0.9742	0.9879	1.0071
	$CS_T(\tau)$	0.8836	1.2439*	1.4801**	1.4236*	1.3526*	1.4633**	1.6802**	1.7879**	1.7508**
6-3	$\hat{\beta}^{+}(\tau)$	1.0056	1.0104	1.0161	1.0210	1.0265	1.0305	1.0334	1.0373	1.0517
	$CS_T(\tau)$	1.9344**	1.6523**	1.3564*	1.2062*	1.2235*	1.2491*	1.2896*	1.1003	1.0039
6-12	$\hat{\beta}^{+}(\tau)$	0.9594	0.9730	0.9772	0.9807	0.9845	0.9899	0.9956	1.0028	1.0133
	$CS_T(\tau)$	1.0650	1.0816	1.3913*	1.6695**	1.6807**	1.9304**	2.0034**	2.0739**	2.0344**
12-3	$\hat{\beta}^{+}(\tau)$	0.9846	1.0064	1.0211	1.0293	1.0363	1.0432	1.0507	1.0605	1.0839
	$CS_T(\tau)$	1.9336**	1.8389**	1.6954**	1.7005**	1.5964**	1.6644**	1.7136**	1.6424**	1.2180*
12-6	$\hat{\beta}^{+}(\tau)$	0.9808	0.9952	1.0037	1.0079	1.0111	1.0173	1.0209	1.0253	1.0377
	$CS_T(\tau)$	2.2030**	2.1713**	2.0231**	2.0425**	2.0922**	1.7259**	1.5582**	1.2095*	0.9516

Notes: U.S. daily Treasury bill constant maturity data is from 01/04/1982 to 04/30/2012. The number of observation is 7584.

In general, the expectations hypothesis for the term structure is supported in the inner quantiles for Treasury bill rates and financial commercial rates. For the constant maturity Treasury rates, cointegration relationships are found in the tails of the interest rate distributions. One explanation may be that each interest rate itself exhibits asymmetric adjustment dynamics over the business cycle[26] (Koenker and Xiao 2004). Hence, results from the multifactor (term structure) model also display some asymmetry in the median, lower, or upper quartiles of the interest rate distributions.

4.4 Conclusion

This chapter provides a cumulated sum test for the null hypothesis of quantile cointegration. In order to correct serial correlation and long run endogeneity, a Phillips-Hansen type fully modified quantile estimator for the cointegrating coefficients is employed to remove the second-order bias and nuisance parameters. For this semiparametric correction, the long run covariance between the regression disturbance and the innovation of the I(1) regressors is estimated using the Bartlett kernel with the plug-in bandwidth, as suggested by Andrews (1991) and Xiao and Phillips (2002). The CUSUM test statistic is composed of the partial sums of the residuals from the fully modified quantile regression. For each quantile, under the null of cointegration, the test statistic converges to a functional of Brownian motions.

The model is applied to several sets of daily U.S. interest rate data. The U.S. interest rates are found to be nonstationary and integrated of order one. Over various quantiles, evidence of cointegration is mixed. In the conditional quantile context, the expectations hypothesis for the term structure is retained only in part of the interest rate distribution for certain data sets.

In addition, several extensions can be made to the model. For example, it is relevant to incorporate structural changes in the quantile regression, since the long run relationship among the nonstationary variables may not be time-invariant. Also, ignoring the possibility

[26] Koenker and Xiao (2004) apply quantile autoregression based unit root tests to several interest rate series in the U.S.. They find that short term interest rate series are not constant unit root processes over different quantiles. There is significant support for asymmetry in the business cycle dynamics of the interest rates.

of structural break can affect the power of cointegration tests (Gregory 1994, Gregory and Hansen 1996). In empirical studies, for example, Hansen (1992), Hall et al. (1992), and Mandeno and Giles (1995) all consider a possible regime shift in the term structure data due to the Federal Reserve's policy change. By allowing for structural changes, one can estimate and test the quantile cointegration relationship, which is possibly unstable over time.

Chapter 5

Conclusion

This thesis studies quantile regression with cointegrated or nearly integrated variables. In Chapter Two, I analyze quantile cointegrating regressions allowing for multiple structural changes of unknown timing. When the number of breaks is pre-specified, the break dates and cointegrating vectors are estimated jointly by minimizing the check function over all permissible partitions of the sample. The rate of convergence of the break fraction estimates is obtained under the assumption of shrinking shift. The quantile estimator has a nonstandard limit distribution. In particular, to develop standard inference procedure for the cointegrating coefficients, a fully modified estimator is proposed to remove the second-order bias and nuisance parameters. The resulting limit distribution is mixed normal. The simulation study shows that the fully modified estimator has good finite sample bias and coverage properties, compared with the original quantile estimator. In the empirical application, I examine the linkage between the emerging stock markets of China and several mature markets, allowing for a regime shift in the quantile cointegrating vectors. The estimated break dates accurately capture the boom in the stock markets in China during

2006-2007 and closely coincide with the policy changes in that period. I also find evidence of financial integration in some conditional quantiles of the Chinese stock indices. The quantile cointegration relationships are asymmetric in the distribution, which cannot be found using least squares estimation.

In Chapter Three, we develop inference in the predictive quantile regression with a persistent regressor. Under the local-to-unity specification, the limit distributions of the predictive quantile regression coefficient estimate and the corresponding t-statistic are nonstandard and depend on the local-to-unity parameter, residual cross correlation, and quantile level. In general, the local-to-unity parameter is unknown and cannot be consistently estimated. Therefore, we propose a Bonferroni bounds method based on Cavanagh et al. (1995). We create computerized lookup tables consisting of simulated critical values for a number of parameter sets. The Monte Carlo experiments suggest that it is worthwhile to adopt the Bonferroni procedure to reduce the size distortion especially when there is considerable residual cross correlation. We test the predictability at different quantiles of the stock return distribution using 16 pre-determined predictor variables. At 5% significance level, except for dividend price ratio, dividend yield, and earnings price ratio, all of the variables have some predictive power for a component of the return distribution.

In Chapter Four, I propose a CUSUM test for the null hypothesis of quantile cointegration. Similar to Chapter Two, the fully modified quantile estimator for the cointegrating coefficients is employed to correct serial correlation and long run endogeneity. The CUSUM test statistic is composed of the cumulative sums of the residuals from the fully modified quantile regression. For each quantile, under the null of cointegration, the test statistic converges to a functional of Brownian motions, as in Xiao and Phillips (2002). When applied to three U.S. interest rate data sets, the robust residual based test rejects the null hypothesis of cointegration in various quantiles of the interest rate distributions.

Several extensions can be made to the models proposed in this thesis. First, the quantile cointegrating regression model with breaks can include trending regressors. In this case, consistency and the rates of convergence will remain the same, but the limit distributions will be different. Also, the regression model can include both stationary and nonstationary regressors. Second, Chapter Two considers the case of pure structural change. It is also useful to consider a partial break where only part of the coefficients change when a break occurs. Third, the second chapter focuses on estimation and asymptotic theory in the quantile cointegration model with a given number of structural breaks. When the number of breaks is unknown, tests of structural changes for quantile regressions with I(1) variables can be developed. Fourth, the estimation can be extended to multiple quantiles. In this case, the unknown parameters can be obtained jointly by minimizing the sum of the check functions across the quantiles of interest and all regimes.

Moreover, instead of using the fully modified estimator, leads and lags of the integrated regressors can be included in the model to account for endogeneity. When a break occurs, a subset of the parameters, possibly including the intercept and coefficients associated with the stationary and integrated regressors, is subject to change. Two cases of partial structural changes are interesting. The first case is that the cointegrating coefficients are allowed to change. The second is that only the intercept and coefficients of the stationary regressors vary across regimes. The break fraction estimates are consistent with the same rate of convergence in both cases. However, the limit distributions are different. When the number of breaks is unknown, a sequential procedure can be used to detect structural changes. In this case, a sup-Wald statistic can be adopted to test for structural breaks in the model.

Furthermore, it is relevant to combine Chapters Two and Four in order to develop tests for cointegration in quantile regression model with structural breaks. Particularly, the CUSUM test for the null hypothesis of quantile cointegration can be extended to the fully modified quantile regression with shifting regimes. In the application to the term structure data, regime shifts due to Federal Reserve's policy change can be incorporated. Also, the quantile cointegration model has potential for applications to other financial and macroeconomic data, such as foreign exchange rates and spot and futures prices.

For the theoretical model from Chapter Three, the asymptotic bias in the estimated coefficient associated with the persistent regressor can be removed using a specialization of the fully modified approach to the predictive quantile regression framework. In particular, a local-to-unity version of the Phillips-Hansen fully modified estimator and a refined bounds procedure (Campbell and Yogo 2006, Hjalmarsson 2007) can be generalized to conditional quantiles to develop feasible inference.

In addition, it is also interesting to analyze the long-horizon version of the predictive quantile regression. It can be assumed that the return horizon grows as a fixed fraction of the sample size.

Finally, the predictive model can be applied to testing the predictability of excess exchange rate returns using forward exchange rate premium (or, equivalently, the home foreign interest rate differential). The null hypothesis is typically referred to as forward rate unbiasedness, since it implies that the forward rate is an unbiased predictor of the future spot rate. The strength of the rejections typical in these regressions suggests that the forward rate is instead a perverse predictor, predicting changes in the wrong direction. This gives rise to the forward premium anomaly, a well known puzzle in international finance. The quantile regression framework can be adopted to account for the asymmetric and nonlinear relationship between exchange rate return and its predictors, which are often highly persistent. Interesting empirical findings in the context of conditional quantiles may provide new explanations to the puzzle.

Bibliography

Aggarwal, R. and M. Mougoue (1993): "Cointegration among Southeast Asian and Japanese currencies. Preliminary evidence of a Yen bloc?", *Economics Letters* 41, 161-166.

Amihud, Y. and C. M. Hurvich (2004): "Predictive regressions: A reduced-bias estimation method", *Journal of Financial and Quantitative Analysis* 39, 813-841.

Amihud, Y., C. M. Hurvich, and Y. Wang (2004): "Hypothesis testing in predictive regressions", Mimeographed, New York University.

Andrews, D. W. K. (1991): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica* 59(3), 817-858.

Arize, A. C. (2002): "Imports and exports in 50 countries tests of cointegration and structural breaks", *International Review of Economics and Finance* 11, 101-115.

Azad, A. S. M. S. (2009): "Efficiency, cointegration and contagion in equity markets: Evidence from China, Japan and South Korea", *Asian Economic Journal* 23(1), 93-118.

Bai, J. (1995): "Least absolute deviation estimation of a shift", *Econometric Theory* 11(3), 403-436.

Bai, J. (1996): "Testing for parameter constancy in linear regressions: An empirical distribution function approach", *Econometrica* 64(3), 597-622.

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

Bai, J. (1998): "Estimation of multiple-regime regressions with least absolutes deviation", *Journal of Statistical Planning and Inference* 74, 103-134.

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

Bai, J. and P. Perron (2003): "Computation and analysis of multiple structural change models", *Journal of Applied Econometrics* 18, 1-22.

Baillie, R. T. and T. Bollerslev (1989): "Common stochastic trends in a system of exchange rates", *Journal of Finance* 44, 167-181.

Baillie, R. T. and T. Bollerslev (1994): "Cointegration, fractional cointegration, and exchange rate dynamics", *The Journal of Finance* 49(2), 737-745.

Brown, R. L. and J. Durbin (1968): "Methods of investigating whether a regression relationship is constant over time", Selected Statistical Papers, European Meeting, Mathematical Centre Tracts No. 26, Amsterdam.

Brown, R. L., J. Durbin, and J. Evans (1975): "Techniques for testing the constancy of regression relationship over time", *Journal of Royal Statistical Society*, Series (B) 37, 149-163.

Burdekin, R. C. K. and P. L. Siklos (2011): "Enter the dragon: Interactions between Chinese, US and Asia-Pacific equity markets, 1995-2010", Working paper.

Campbell, B. and J.-M. Dufour (1995): "Exact nonparametric orthogonality and random walk tests", *The Review of Economics and Statistics* 77, 1-16.

Campbell, B. and J.-M. Dufour (1997): "Exact nonparametric tests of orthogonality and random walk in the presence of a drift parameter", *International Economic Review* 38, 151-173.

Campbell, J. Y. (1987): "Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis", *Econometrica* 55, 1249-1273.

Campbell, J. Y. and R. J. Shiller (1987): "Cointegration and tests of present value models", *Journal of Political Economy* 95, 1062-1088.

Campbell, J. Y. and R. J. Shiller (1988a): "The dividend price ratio and expectations of future dividends and discount factors", *Review of Financial Studies* 1, 195-227.

Campbell, J. Y. and R. J. Shiller (1988b): "Stock prices, earnings, and expected dividends", *Journal of Finance* 43, 661-676.

Campbell, J. Y. and S. B. Thompson (2008): "Predicting the equity premium out of sample: Can anything beat the historical average?", *Review of Financial Studies* 21, 1509-1531.

Campbell, J. Y. and M. Yogo (2006): "Efficient tests of stock return predictability", *Journal of Financial Economics* 81(1), 27-60.

Cavanagh, C., G. Elliott, and J. H. Stock (1995): "Inference in models with nearly integrated regressors", *Econometric Theory* 11, 1131-1147.

Cenesizoglu, T. and A. Timmermann (2008): "Is the distribution of stock returns predictable?", Working paper, HEC Montreal and University of California at San Diego.

Chan, N. H. and C. Z. Wei (1987): "Asymptotic inference for nearly nonstationary AR(1) processes", *Annals of Statistics* 15(3), 1050-1063.

Chan, N. H. (1988): "The parameter inference for nearly nonstationary time series", *Journal of the American Statistical Association* 83(403), 857-862.

Chance, D. M. (1991): *An Introduction to Options and Futures*. Dryden Press, Chicago.

Chen, G., M. Firth, and O. M. Rui (2002): "Stock market linkages: Evidence from Latin America", *Journal of Banking and Finance* 26(6), 1113-1141.

Chen, X., R. Koenker, and Z. Xiao (2009): "Copula-based quantile autoregression", *Econometrics Journal* 12(1), 50-67.

Chernozhukov, V. (2005): "Extremal quantile regression", *The Annals of Statistics* 33, 806-839.

Chernozhukov, V. (2010): "Inference for extremal conditional quantile models", *Review of Economic Studies*, forthcoming.

Chernozhukov, V. and A. Belloni (2010): "l_1-Penalized quantile regression in high dimensional sparse models", *Annals of Statistics*, forthcoming.

Chernozhukov, V. and S. Du (2006): "Extremal quantiles and value-at-risk", MIT Department of Economics Working Paper No. 07-01.

Chernozhukov, V., C. Hansen, and M. Jansson (2009): "Finite sample inference for quantile regression models", *Journal of Econometrics* 152(2), 93-103.

Chuang, C.-C., C.-M. Kuan, and H. Lin (2009): "Causality in quantiles and dynamic stock return-volume relations", *Journal of Banking and Finance* 33, 1351-1360.

Diebold, F. X., J. Gardeazabal, and K. Yilmaz (1994): "On cointegration and exchange rate dynamics", *The Journal of Finance* 49(2), 727-735.

Downing, C. and S. Oliner (2007): "The term structure of commercial paper rate", *Journal of Financial Economics* 83, 59-86.

Engle, R. F. and C. W. J. Granger (1987): "Co-integration and error correction: Representation, estimation, and testing", *Econometrica* 55(2), 251-276.

Engle, R. F. and S. Manganelli (2004): "CAViaR: Conditional autoregressive value at risk by regression quantiles", *Journal of Business and Economic Statistics* 22, 367-381.

Fama, E. and K. French (1988): "Permanent and temporary components of stock prices", *Journal of Political Economy* 96, 246-273.

Frolich, M. and B. Melly (2008): "Quantile treatment effects in the regression discontinuity design", IZA Discussion Papers No. 3638.

Goetzmann, W. and P. Jorion (1993): "Testing the predictive power of dividend yields", *Journal of Finance* 48(2), 663-670.

Golinelli, R. and R. Orsi (1994): "Price-wage dynamics in a transition economy: The case of Poland", *Economics of Planning* 27, 293-313.

Golinelli, R. and R. Orsi (2000): "Testing for structural change in cointegrated relationships: Analysis of pricewages models for Poland and Hungary", *Economics of Planning* 33, 19-51.

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

Gregory, A. W. (1994): "Testing for cointegration in linear quadratic models", *Journal of Business and Economic Statistics* 12(3), 347-360.

Gregory, A. W. and B. E. Hansen (1996): "Residual-based tests for cointegration in models with regime shifts", *Journal of Econometrics* 70, 99-126.

Gregory, A. W., J. M. Nason, and D. G. Watt (1996): "Testing for structural breaks in cointegrated relationships", *Journal of Econometrics* 71, 321-341.

Hall, A. D., H. M. Anderson, and C. W. J. Granger (1992): "A cointegration analysis of treasury bill yields", *The Review of Economics and Statistics* 74(1), 116-126.

Hansen, B. E. (1992): “Tests for parameter instability in regressions with I(1) processes”, *Journal of Business and Economic Statistics* 10(3), 321-335.

Hansen, B. E. and P. C. B. Phillips (1990): “Estimation and inference in models of cointegration: A simulation study”, *Advances in Econometrics* 8, 225-248.

Hao, K. and B. Inder (1996): “Diagnostic test for structural change in cointegrated regression models”, *Economics Letters* 50, 179-187.

Hjalmarsson, E. (2007): “Fully modified estimation with nearly integrated regressors”, *Finance Research Letters* 4, 92-94.

Hodrick, R. (1992): “Dividend yields and expected stock returns: Alternative procedures for inference and measurement”, *Review of Financial Studies* 5, 357-368.

Huber, M. (2010): “Testing for covariate balance using nonparametric quantile regression and resampling methods”, University of St. Gallen Department of Economics working paper series 2010 No. 2010-18.

Iacone, F. (2009): “A semiparametric analysis of the term structure of the US interest rates”, *Oxford Bulletin of Economics and Statistics* 71(4), 475-490.

Jansson, M. and M. J. Moreira (2006): “Optimal inference in regression models with nearly integrated regressors”, *Econometrica* 74, 681-714.

Johansen, S. (1988): “Statistical analysis of cointegrating vectors”, *Journal of Economic Dynamics and Control* 12, 231-254.

Johansen, S. (1991): “Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models”, *Econometrica* 59, 1551-1580.

Johansen, S. (1995): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.

Johansen, S. and K. Juselius (1990): "Maximum likelihood estimation and inference on cointegration—with applications to the demand for money", *Oxford Bulletin of Economics and Statistics* 52(2), 169-210.

Jun, S. J., J. Pinkse, and H. Xu (2009): "Tighter bounds in triangular systems", Preprint.

Kejriwal, M. and P. Perron (2008): "The limit distribution of the estimates in cointegrated regression models with multiple structural changes", *Journal of Econometrics* 146, 59-73.

Kejriwal, M. and P. Perron (2010): "Testing for multiple structural changes in cointegrated regression models", *Journal of Business and Economic Statistics* 28(4), 503-522.

Knight, K. (1989): "Limit theory for autoregressive-parameter estimates in an infinite-variance random walk", *The Canadian Journal of Statistics* 17(3), 261-278.

Koenker, R. (2008): "Censored quantile regression redux", *Journal of Statistical Software* 27(6), 1-25.

Koenker, R. (2010): "Additive models for quantile regression: Model selection and confidence bandaids", Preprint.

Koenker, R and G. Bassett (1978): "Regression quantiles", *Econometrica* 46, 33-49.

Koenker, R. and F. K. Hallock (2001): "Quantile regression", *Journal of Economic Perspectives* 15(4), 143-156.

Koenker, R. and Z. Xiao (2004): "Unit root quantile autoregression inference", *Journal of the American Statistical Association* 99, 775-787.

Koenker, R. and Z. Xiao (2006): "Quantile autoregression", *Journal of the American Statistical Association* 101, 980-990.

Lanne, M. (2002): "Testing the predictability of stock returns", *The Review of Economics and Statistics* 84(3), 407-415.

Leuthold, R. M., J. C. Junkus, and J. E. Cordier (1989): *The Theory and Practice of Futures Markets*. Lexington Books, D. C. Health and Co., Lexington, MA.

Lewellen, J. (2004): "Predicting returns with financial ratios", *Journal of Financial Economics* 74(2), 209-235.

Lopez, J. A. (2005): "Exchange rate cointegration across central bank regime shifts", *Research in Finance* 22, 327-356.

MacKinnon, J. G. (1991): "Critical values for cointegration tests". In: R. F. Engle and C. W. J. Granger (Eds.), Long run economic relations: Readings in cointegration. Berlin: Oxford University Press.

MacKinnon, J. G. (2010): "Critical values for cointegration tests", Queen's Economics Department Working Paper No. 1227.

Mandeno, R. J. and D. E. A. Giles (1995): "The expectations theory of the term structure: A cointegration/causality analysis of US interest rates", *Applied Financial Economics* 5, 273-283.

Mankiw, N. G. and M. Shapiro (1986): "Do we reject too often? Small sample properties of tests of rational expectations models", *Economics Letters* 20, 139-145.

Manning, N. (2002): "Common trends and convergence? South East Asian equity markets, 1998-1999", *Journal of International Money and Finance* 21, 183-202.

Maynard, A. and K. Shimotsu (2009): "Covariance-based orthogonality tests for regressors with unknown persistence", *Econometric Theory* 25(1), 63-116.

Nabeya, S. and B. E. Sorensen (1994): "Asymptotic distributions of the least-squares estimators and test statistics in the near unit root model with non-zero initial value and local drift and trend", *Econometric Theory* 10, 937-966.

Nelson, C. and M. Kim (1993): "Predictable stock returns: The role of small sample bias", *Journal of Finance* 48, 641-661.

Oka, T. and Z. Qu (2011): "Estimating structural changes in regression quantiles", *Journal of Econometrics* 162(2), 248-267.

Papell, D. H. (1997): "Cointegration and exchange rate dynamics", *Journal of International Money and Finance* 16(3), 445-460.

Park, J. (1990): "Testing for unit roots and cointegration by variable addition". In: Fomby and Rhodes (Eds.), Advances in Econometrics, 107-133. JAI press.

Park, J., S. Ouliaris, and B. Choi (1988): "Spurious regressions and tests for cointegration", Mimeo, Cornell University.

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

Phillips, P. C. B. (1987): "Toward a unified asymptotic theory for autoregression", *Biometrika* 74, 535-547.

Phillips, P. C. B. (1989): "Partially identified econometric models", *Econometric Theory* 5, 181-240.

Phillips, P. C. B. and B. E. Hansen (1990): "Statistical inference in instrumental variables regression with I(1) processes", *Review of Economic Studies* 57, 99-125.

Phillips, P. C. B. and M. Loretan (1991): "Estimating long run economic equilibria", *Review of Economic Studies* 58(3), 407-436.

Phillips, P. C. B. and S. Ouliaris (1990): "Asymptotic properties of residual-based tests for cointegration", *Econometrica* 58, 165-193.

Ploberger, W. and W. Kramer (1992): "The CUSUM test with OLS residuals", *Econometrica* 60, 271-285.

Pollard, D. (1991): "Asymptotics for least absolute deviation regression estimators", *Econometric Theory* 7(2), 186-199.

Qu, Z. (2008): "Testing for structural change in regression quantiles", *Journal of Econometrics* 146, 170-184.

Shiller, R. (1984): "Stock prices and social dynamics", *Brookings papers on Economic Activity* 2, 457-498.

Shin, Y. (1994): "A residual based test of the null of cointegration against the alternative of no cointegration", *Econometric Theory* 10, 91-115.

Stambaugh, R. F. (1986): "Bias in regressions with lagged stochastic regressors", Center for Research in Security Prices Working Paper 156, University of Chicago.

Stambaugh, R. F. (1999): "Predictive regressions", *Journal of Financial Economics* 54, 375-421.

Stock, J. H. (1991): "Confidence intervals for the largest autoregressive root in U.S. economic time series", *Journal of Monetary Economics* 28(3), 435-460.

Stock, J. H. and M. W. Watson (1988): "Testing for common trends", *Journal of the American Statistical Association* 83, 1097-1107.

Su, L. and Z. Xiao (2008): "Testing for parameter stability in quantile regression models", *Statistics and Probability Letters* 78, 2768-2775.

Voronkova, S. (2004): "Equity market integration in Central European emerging markets: A cointegration analysis with shifting regimes", *International Review of Financial Analysis* 13, 633-647.

Wolf, M. (2000): "Stock returns and dividend yields revisited: A new way to look at an old problem", *Journal of Business and Economic Statistics* 18(1), 18-30.

Wright, J. (2000): "Confidence sets for cointegrating coefficients based on stationarity tests", *Journal of Business and Economic Statistics* 18(2), 211-222.

Xiao, Z. (2009): "Quantile cointegrating regression", *Journal of Econometrics* 150, 248-260.

Xiao, Z. and P. C. B. Phillips (2002): "A CUSUM test for cointegration using regression residuals", *Journal of Econometrics* 108, 43-61.

Xiao, Z. and R. Koenker (2009): "Conditional quantile estimation for GARCH models", Boston College Working Papers in Economics No. 725.

Appendix A

Proofs for Chapter 2

A.1 Proof of Theorem 1

The regression coefficient and break fraction estimates are consistent such that $D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau)) = O_p(1)$ with $j = 1, ..., m+1$ and $\nu_T^2(\hat{T}_j - T_j^0) = O_p(1)$ with $j = 1, ..., m$, see Kejriwal and Perron (2008). Moreover, consistency of the regression coefficient estimates will also follow from the present proof and consistency of the break fraction estimates will be shown in section A.3.

Rewrite the minimization problem of equation (2.5) as

$$\begin{aligned} & \inf_{T^b} \inf_{\theta_j(\tau)} S_T(\tau, \theta(\tau), T^b) \\ = \; & \inf_{T^b} \inf_{\theta_j(\tau)} \{S_T(\tau, \theta(\tau), T^0) + [S_T(\tau, \theta(\tau), T^b) - S_T(\tau, \theta(\tau), T^0)]\}. \end{aligned}$$

The limit distribution of the coefficient estimates is derived from the following expression:

$$\inf_{\theta_j(\tau)} S_T(\tau, \theta(\tau), T^0) = \inf_{\theta_j(\tau)} \sum_{j=1}^{m+1} \sum_{t=T_{j-1}^0+1}^{T_j^0} \rho_\tau(y_t - z_t'\theta_j(\tau)).$$

Note that the asymptotic distribution of the coefficient estimates is the same as that obtained when the break dates are known. The rest of the proof follows the procedure in section A.1 from Xiao (2009), since in each regime the model in Chapter Two has the same asymptotic behavior as the quantile cointegrating regression considered by Xiao (2009). For each j the minimization problem is $\min_{\theta_j(\tau)} \sum_{t=T_{j-1}^0+1}^{T_j^0} \rho_\tau(y_t - z_t'\theta_j(\tau))$, which is equivalent to

$$\min_{\phi_j} \sum_{t=T_{j-1}^0+1}^{T_j^0} \{\rho_\tau(u_t(\tau) - (D_T^{-1}\phi_j)'z_t) - \rho_\tau(u_t(\tau))\},$$

where $\hat{\phi}_j = D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau))$. Consider the objective function

$$G_T(\phi_j) = \sum_{t=T_{j-1}^0+1}^{T_j^0} \{\rho_\tau(u_t(\tau) - (D_T^{-1}\phi_j)'z_t) - \rho_\tau(u_t(\tau))\}$$

and $\hat{\phi}_j$ is a minimizer of $G_T(\phi_j)$. Similar to the convex random function discussed in Knight (1989), the objective function $G_T(\phi_j)$ is also convex. According to Knight (1989) and Pollard (1991), if the finite-dimensional distributions of $G_T(\phi_j)$ converge weakly to those of $G(\phi_j)$ and $G(\phi_j)$ has a unique minimum, the convexity of $G_T(\phi_j)$ implies that $\hat{\phi}_j$ converges in distribution to the minimizer of $G(\phi_j)$ (Xiao 2009). Furthermore, if $\theta_j^0(\tau)$ is the unique minimizer of $G(\phi_j)$, then $\hat{\theta}_j(\tau)$ is a consistent estimator of $\theta_j^0(\tau)$.

Thus, to derive the asymptotic distribution of $\hat{\theta}_j(\tau)$, according to equation (24) from Xiao (2009), for $u \neq 0$ we have $\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + (u - v)[I(0 > u > v) - I(0 < u < v)]$. Then, with $T_{j-1}^0 + 1 \leq t \leq T_j^0$ we have

$$\begin{aligned} G_T(\phi_j) &= -\sum_t (D_T^{-1}\phi_j)'z_t\psi_\tau(u_t(\tau)) + \sum_t (u_t(\tau) - (D_T^{-1}\phi_j)'z_t) \\ &\quad \times[I(0 > u_t(\tau) > (D_T^{-1}\phi_j)'z_t) - I(0 < u_t(\tau) < (D_T^{-1}\phi_j)'z_t)] \\ &= -(a) + (b), \end{aligned}$$

where

$$
\begin{aligned}
(a) &= \phi_j' D_T^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} z_t \psi_\tau(u_t(\tau)) \\
&= \phi_j' \begin{bmatrix} T^{-\frac{1}{2}} \sum_{t=T_{j-1}^0+1}^{T_j^0} \psi_\tau(u_t(\tau)) \\ T^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} x_t \psi_\tau(u_t(\tau)) \end{bmatrix} \\
&= \phi_j' \begin{bmatrix} T^{-\frac{1}{2}} \sum_{t=1}^{[\lambda_j^0 T]} \psi_\tau(u_t(\tau)) - T^{-\frac{1}{2}} \sum_{t=1}^{[\lambda_{j-1}^0 T]} \psi_\tau(u_t(\tau)) \\ T^{-1} \sum_{t=1}^{[\lambda_j^0 T]} x_t \psi_\tau(u_t(\tau)) - T^{-1} \sum_{t=1}^{[\lambda_{j-1}^0 T]} x_t \psi_\tau(u_t(\tau)) \end{bmatrix} \\
&\Rightarrow \phi_j' \begin{bmatrix} B_\psi(\lambda_j^0) - B_\psi(\lambda_{j-1}^0) \\ \int_{\lambda_{j-1}^0}^{\lambda_j^0} B_v dB_\psi + (\lambda_j^0 - \lambda_{j-1}^0)\Delta_{v\psi} \end{bmatrix} \\
&\Rightarrow \phi_j' \begin{bmatrix} \int_{\lambda_{j-1}^0}^{\lambda_j^0} dB_\psi \\ \int_{\lambda_{j-1}^0}^{\lambda_j^0} B_v dB_\psi + (\lambda_j^0 - \lambda_{j-1}^0)\Delta_{v\psi} \end{bmatrix} \\
&= \phi_j' \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dB_\psi + (\lambda_j^0 - \lambda_{j-1}^0)\bar{\Delta}_{v\psi} \right].
\end{aligned}
$$

For (b) denote $(b1) = \sum_t (\phi_j' D_T^{-1} z_t - u_t(\tau)) I(0 < u_t(\tau) < \phi_j' D_T^{-1} z_t) = W_T(\phi_j) = \sum_t w_t(\phi_j)$, where $w_t(\phi_j) = (\phi_j' D_T^{-1} z_t - u_t(\tau)) I(0 < u_t(\tau) < \phi_j' D_T^{-1} z_t)$. Also, denote $W_{Tc}(\phi_j) = \sum_t w_{tc}(\phi_j)$, where $w_{tc}(\phi_j) = (\phi_j' D_T^{-1} z_t - u_t(\tau)) I(0 < u_t(\tau) < \phi_j' D_T^{-1} z_t) \times I(\phi_j' D_T^{-1} z_t \leq c)$ is the truncation of $\phi_j' D_T^{-1} z_t$ at some finite number $c > 0$. Let $\bar{w}_{tc}(\phi_j) = E[w_{tc}(\phi_j)|\mathcal{F}_t]$ and $\bar{W}_{Tc}(\phi_j) = \sum_t \bar{w}_{tc}(\phi_j)$, where $\mathcal{F}_t$ is information set of the σ-algebra generated by $(v_t, u_{t-1}, v_{t-1}, u_{t-2}, v_{t-2}, ...)$ and $z_t \in \mathcal{F}_t$. From Assumption 3, we have $F_t(u) = Pr(u_t < u|\mathcal{F}_t)$ and its derivative $f_t(u)$ a.s., and $f_t(u)$ is uniformly integrable in some neighborhood of $F_t^{-1}(\tau)$. For simplicity, the conditional distribution of u_t is assumed to be time invariant, such that $F_t(\cdot) = F(\cdot)$ for $t = 1, ..., T$. Then $u_t(\tau) = u_t - F^{-1}(\tau)$.

Thus, we have

$$
\begin{aligned}
\bar{W}_{Tc}(\phi_j) &= \sum_t E[(\phi_j' D_T^{-1} z_t - u_t(\tau)) I(0 < u_t(\tau) < \phi_j' D_T^{-1} z_t) I(\phi_j' D_T^{-1} z_t \leq c) | \mathcal{F}_t] \\
&= \sum_t E[(\phi_j' D_T^{-1} z_t + F^{-1}(\tau) - u_t) I(F^{-1}(\tau) < u_t < \phi_j' D_T^{-1} z_t + F^{-1}(\tau)) \\
&\quad \times I(\phi_j' D_T^{-1} z_t \leq c) | \mathcal{F}_t] \\
&= \sum_t \int_{F^{-1}(\tau)}^{[\phi_j' D_T^{-1} z_t + F^{-1}(\tau)] I(\phi_j' D_T^{-1} z_t \leq c)} [\int_r^{[\phi_j' D_T^{-1} z_t + F^{-1}(\tau)] I(\phi_j' D_T^{-1} z_t \leq c)} ds] \\
&\quad \times f_t(r) dr \\
&= \sum_t \int_{F^{-1}(\tau)}^{[\phi_j' D_T^{-1} z_t + F^{-1}(\tau)] I(\phi_j' D_T^{-1} z_t \leq c)} [\int_{F^{-1}(\tau)}^{s} f_t(r) dr] ds \\
&= \sum_t \int_{F^{-1}(\tau)}^{[\phi_j' D_T^{-1} z_t + F^{-1}(\tau)] I(\phi_j' D_T^{-1} z_t \leq c)} [s - F^{-1}(\tau)] \left[\frac{F_t(s) - F_t(F^{-1}(\tau))}{s - F^{-1}(\tau)} \right] ds \\
&= \sum_t \int_{F^{-1}(\tau)}^{[\phi_j' D_T^{-1} z_t + F^{-1}(\tau)] I(\phi_j' D_T^{-1} z_t \leq c)} [s - F^{-1}(\tau)] f_t(F^{-1}(\tau)) ds + o_p(1) \\
&= \sum_t f_t(F^{-1}(\tau)) \left[\frac{[s - F^{-1}(\tau)]^2}{2} \Big|_{F^{-1}(\tau)}^{[\phi_j' D_T^{-1} z_t + F^{-1}(\tau)] I(\phi_j' D_T^{-1} z_t \leq c)} \right] + o_p(1) \\
&= \frac{1}{2} \sum_t f_t(F^{-1}(\tau)) \phi_j' D_T^{-1} z_t z_t' D_T^{-1} \phi_j I(\phi_j' D_T^{-1} z_t \leq c) + o_p(1).
\end{aligned}
$$

Since $\{f_t(F^{-1}(\tau))\}$ is stationarity, for some $\epsilon > 0$ it holds that

$$
\sup_{0 \leq r \leq 1} \left| \frac{1}{T^{1-\epsilon}} \sum_{t=1}^{[Tr]} [f_t(F^{-1}(\tau)) - f(F^{-1}(\tau))] \right| \to_d 0.
$$

Therefore, we have

$$
\begin{aligned}
\bar{W}_{Tc}(\phi_j) &= \frac{1}{2T}\sum_{t=1}^{[\lambda_j^0 T]} f_t(F^{-1}(\tau))\phi_j'[\sqrt{T}D_T^{-1}z_t z_t' D_T^{-1}\sqrt{T}]\phi_j I(\phi_j' D_T^{-1}z_t \le c) \\
&\quad -\frac{1}{2T}\sum_{t=1}^{[\lambda_{j-1}^0 T]} f_t(F^{-1}(\tau))\phi_j'[\sqrt{T}D_T^{-1}z_t z_t' D_T^{-1}\sqrt{T}]\phi_j I(\phi_j' D_T^{-1}z_t \le c) \\
&\quad +o_p(1) \\
&\to_d \frac{1}{2}f(F^{-1}(\tau))\phi_j'\left[\int_0^{\lambda_j^0}\bar{B}_v\bar{B}_v'\right]\phi_j I(0<\phi_j'\bar{B}_v \le c) \\
&\quad -\frac{1}{2}f(F^{-1}(\tau))\phi_j'\left[\int_0^{\lambda_{j-1}^0}\bar{B}_v\bar{B}_v'\right]\phi_j I(0<\phi_j'\bar{B}_v \le c) \\
&= \frac{1}{2}f(F^{-1}(\tau))\phi_j'\left[\int_{\lambda_{j-1}^0}^{\lambda_j^0}\bar{B}_v\bar{B}_v'\right]\phi_j I(0<\phi_j'\bar{B}_v \le c).
\end{aligned}
$$

Since $(\phi_j' D_T^{-1}z_t)I(0 \le \phi_j' D_T^{-1}z_t \le c) \to_p 0$ uniformly in t, we have $\sum_t E[w_{tc}(\phi_j)^2|\mathcal{F}_t] \le \max\{(\phi_j' D_T^{-1}z_t)I(0 \le \phi_j' D_T^{-1}z_t \le c)\}\sum \bar{w}_{tc}(\phi_j) \to_p 0$. We thus obtain $\sum_t[w_{tc}(\phi_j) - \bar{w}_{tc}(\phi_j)] \to_p 0$, where $w_{tc}(\phi_j) - \bar{w}_{tc}(\phi_j)$ is a martingale difference sequence (MDS). By the asymptotic equivalence lemma, the limit distribution of $\sum_t w_{tc}(\phi_j)$ is the same as that of $\sum_t \bar{w}_{tc}(\phi_j)$ i.e.

$$
\begin{aligned}
W_{Tc}(\phi_j) &\Rightarrow \frac{1}{2}f(F^{-1}(\tau))\phi_j'\left[\int_{\lambda_{j-1}^0}^{\lambda_j^0}\bar{B}_v\bar{B}_v'\right]\phi_j I(0<\phi_j'\bar{B}_v \le c) \\
&\Rightarrow \frac{1}{2}f(F^{-1}(\tau))\phi_j'\left[\int_{\lambda_{j-1}^0}^{\lambda_j^0}\bar{B}_v\bar{B}_v'\right]\phi_j I(\phi_j'\bar{B}_v > 0) \quad \text{as } c \to \infty.
\end{aligned}
$$

Due to the following

$$
\begin{aligned}
& Pr[|W_T(\phi_j) - W_{Tc}(\phi_j)| \ge 0] \\
= \ & Pr[\sum_t (\phi_j' D_T^{-1}z_t - u_t(\tau))I(0 < u_t(\tau) < \phi_j' D_T^{-1}z_t)I(\phi_j' D_T^{-1}z_t > c)] \\
\le \ & Pr[\bigcup_t \{\phi_j' D_T^{-1}z_t > c\}] \\
= \ & Pr[\max_t\{\phi_j' D_T^{-1}z_t\} > c]
\end{aligned}
$$

and $\lim_{c\to\infty} Pr[\sup_{0\leq r\leq 1} \phi_j' \bar{B}_v(r) > c] = 0$, we have

$$\lim_{c\to\infty} \limsup_{T\to\infty} Pr[|W_T(\phi_j) - W_{Tc}(\phi_j)| \geq \epsilon] = 0.$$

Thus, $W_T(\phi_j)$ and $W_{Tc}(\phi_j)$ have the same limit distribution i.e.

$$(b1) = W_T(\phi_j) \Rightarrow \frac{1}{2} f(F^{-1}(\tau))\phi_j' \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v'\right] \phi_j I(\phi_j' \bar{B}_v > 0).$$

Similarly, we can show that

$$\begin{aligned} (b2) &= \sum_t (u_t(\tau) - \phi_j' D_T^{-1} z_t) I(0 > u_t(\tau) > \phi_j' D_T^{-1} z_t) \\ &\Rightarrow \frac{1}{2} f(F^{-1}(\tau))\phi_j' \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v'\right] \phi_j I(\phi_j' \bar{B}_v \leq 0). \end{aligned}$$

Consequently,

$$(b) = (b1) + (b2) \Rightarrow \frac{1}{2} f(F^{-1}(\tau))\phi_j' \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v'\right] \phi_j$$

and

$$\begin{aligned} G_T(\phi_j) &= -(a) + (b) \\ &\Rightarrow -\phi_j' \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dB_\psi + (\lambda_j^0 - \lambda_{j-1}^0)\bar{\Delta}_{v\psi}\right] + \frac{1}{2} f(F^{-1}(\tau))\phi_j' \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v'\right] \phi_j \\ &:= G(\phi_j), \end{aligned}$$

where ":=" signifies definitional equality. By the convexity lemma of Pollard (1991) and Knight (1989), both $G_T(\phi_j)$ and $G(\phi_j)$ are minimized at $\hat{\phi}_j = D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau))$ (Xiao 2009). By Lemma A of Knight (1989), from the first order condition,

$$\frac{\partial G(\phi_j)}{\partial \phi_j'} = -\left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dB_\psi + (\lambda_j^0 - \lambda_{j-1}^0)\bar{\Delta}_{v\psi}\right] + \frac{1}{2} f(F^{-1}(\tau)) \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v'\right] \phi_j = 0,$$

we obtain the limit distribution of the regression coefficient estimates such that

$$D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau)) \Rightarrow \left[f(F^{-1}(\tau)) \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v'\right]^{-1} \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dB_\psi + (\lambda_j^0 - \lambda_{j-1}^0)\bar{\Delta}_{v\psi}\right].$$

In addition, the weak convergence conditions in section 2.2 are derived under the added assumption that $F_t(\cdot) = F(\cdot)$ for all t. Nonetheless, this is a sufficient condition for consistency and asymptotic distributions of the regression coefficient and break fraction estimates. The conditional distribution of the regression disturbance does not need to be time invariant to obtain the same rates of convergence. Analogous to assumption A5(b) from Oka and Qu (2011), one can assume that $D_T^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[sT]} f_t(F_t^{-1}(\tau)) z_t z_t' D_T^{-1} \to^p H_j^0(\tau, s)$ and $D_T^{-1} \sum_{t=T_{j-1}^0+1}^{T_{j-1}^0+[sT]} z_t z_t' D_T^{-1} \to^p J_j^0(s)$ hold uniformly in $0 \leq s \leq \lambda_j^0 - \lambda_{j-1}^0$ as $T \to \infty$, where $H_j^0(\tau, s)$ and $J_j^0(s)$ are random almost surely positive definite matrices. This assumption imposes some restriction on possible heteroskedasticity. Note that the matrices $H_j^0(\tau, s)$ and $J_j^0(s)$ are not necessarily the same for all j and they are nonlinear in s. Let "$\to^d$" signify convergence in distribution. In this case, we have $D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau)) \to^d (H_j^0(\tau, s))^{-1} \kappa_j(\tau)$, where $H_j^0(\tau, s) = plim_{t\to\infty} D_T^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} f_t(F_t^{-1}(\tau)) z_t z_t' D_T^{-1}$ with $s = \lambda_j^0 - \lambda_{j-1}^0$ and $\kappa_j(\tau)$ is the limit distribution of $D_T^{-1} \sum_{t=T_{j-1}^0+1}^{T_j^0} z_t \psi_\tau(u_t(\tau))$.

A.2 Proof of Theorem 2

Because the regression coefficient estimator is consistent and the modification is asymptotically negligible, the fully modified quantile estimator is consistent, which is also shown in the following derivation. The fully modified slope coefficient estimate $\hat{\beta}_j^+(\tau)$ is given by

$$\begin{aligned} \hat{\beta}_j^+(\tau) &= \hat{\beta}_j(\tau) - \left[f(\widehat{F^{-1}(\tau)}) \sum_{T_{j-1}^0+1}^{T_j^0} \underline{x}_t \underline{x}_t' \right]^{-1} \\ &\quad \times \left[\sum_{T_{j-1}^0+1}^{T_j^0} \underline{x}_t \hat{\Omega}_{\psi v} \hat{\Omega}_{vv}^{-1} \Delta x_t + (\lambda_j^0 - \lambda_{j-1}^0) T \hat{\Delta}_{v\psi}^+ \right]. \end{aligned}$$

The regression coefficient estimates after the modification $\hat{\theta}_j^{+}(\tau) = (\hat{\alpha}_j(\tau), \hat{\beta}_j'^{+}(\tau))'$ are

$$\begin{aligned}\hat{\theta}_j^{+}(\tau) &= \hat{\theta}_j(\tau) - \left[f(\widehat{F^{-1}(\tau)}) \sum_{T_{j-1}^0+1}^{T_j^0} z_t z_t' \right]^{-1} \\ &\quad \times \left[\sum_{T_{j-1}^0+1}^{T_j^0} z_t \hat{\Omega}_{\psi v} \hat{\Omega}_{vv}^{-1} \Delta x_t + (\lambda_j^0 - \lambda_{j-1}^0) T \bar{\hat{\Delta}}_{v\psi}^{+} \right],\end{aligned}$$

where $\bar{\hat{\Delta}}_{v\psi}^{+} = (0, \Delta_{v\psi}'^{+})'$. Under Assumptions 1-5, Theorem 1 holds so that

$$\begin{aligned}D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau)) &\Rightarrow \left[f(F^{-1}(\tau)) \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v' \right]^{-1} \\ &\quad \times \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dB_{\psi.v} + \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v d\Omega_{\psi v} \Omega_{vv}^{-1} B_v + \bar{\Delta}_{v\psi} \right].\end{aligned}$$

Then, we have

$$\begin{aligned}D_T(\hat{\theta}_j^{+}(\tau) - \theta_j^0(\tau)) &= D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau)) - D_T \left[f(\widehat{F^{-1}(\tau)}) \sum_{T_{j-1}^0+1}^{T_j^0} z_t z_t' \right]^{-1} \\ &\quad \times \left[\sum_{T_{j-1}^0+1}^{T_j^0} z_t \hat{\Omega}_{\psi v} \hat{\Omega}_{vv}^{-1} \Delta x_t + T \bar{\hat{\Delta}}_{v\psi}^{+} \right] \\ &\Rightarrow \left[f(F^{-1}(\tau)) \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v' \right]^{-1} \int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v dB_{\psi.v} \\ &\sim MN\left(0, \frac{\omega_{\psi.v}^2}{f(F^{-1}(\tau))^2} \left[\int_{\lambda_{j-1}^0}^{\lambda_j^0} \bar{B}_v \bar{B}_v' \right]^{-1} \right),\end{aligned}$$

where $\bar{B}_v = (1, B_v')'$ and $\omega_{\psi.v}^2 = \omega_{\psi}^2 - \Omega_{\psi v} \Omega_{vv}^{-1} \Omega_{v\psi}$.

A.3 Proof of Theorem 3

The sketch of the proof is based on the appendix from Oka and Qu (2011). Details are omitted.

In the context of quantile cointegration, the regression coefficient estimates in each regime satisfies that $D_T(\hat{\theta}_j(\tau) - \theta_j^0(\tau)) = O_p(1)$ for $j = 1, ..., m+1$, see Theorem 1. According to Lemma 6 in Bai (2000) and Lemma A.5 in Oka and Qu (2011), the regression coefficient estimates using the pooled data from two regimes are close to the coefficients of "the dominating regime" (Oka and Qu 2011). Consider the case with sample size T and one structural change occurring at time T_1^0, we have

$$y_t = \begin{cases} z_t'\theta_1^0(\tau) + u_t(\tau) & t = 1, ..., T_1^0 \\ z_t'\theta_2^0(\tau) + u_t(\tau) & t = T_1^0 + 1, ..., T. \end{cases}$$

Assume $T - T_1^0 \leq [T^{\frac{1}{2}}\nu_T^{-1}]$ and denote $\hat{\theta}(\tau)$ as the coefficient estimate using the whole sample, we have $D_T(\hat{\theta}(\tau) - \theta_1^0(\tau)) = O_p(1)$.

Suppose there is no structural break and let subscript κ denote estimates or values using sample $t = 1, ..., \kappa$. Define $q_{\tau,t}(\phi) = \rho_\tau(u_t^0(\tau) - z_t'\phi) - \rho_\tau(u_t^0(\tau))$ and $Q_\kappa(\tau, \phi) = \sum_{t=1}^{\kappa} q_{\tau,t}(\phi)$, where $\phi \in R^{k+1}$. With cointegrated I(1) regressors x_t, we have $D_T(\hat{\theta}_\kappa(\tau) - \theta^0(\tau)) = O_p(1)$. Similar to the arguments from Lemma A.4 in Oka and Qu (2011), the following results hold:

1. $\sup_{1\leq\kappa\leq T} \sup_{\|D_T\phi\|\leq A} |Q_\kappa(\tau, \phi)| = O_p(1)$ for any $A > 0$.

2. Let h_T and d_T be positive sequences such that h_T is nondecreasing, $h_T \to \infty$, $d_T \to \infty$ and $(h_T d_T^2)/T \to h$ with $0 < h < \infty$. Then for each $\epsilon > 0$ and $B > 0$ there exists an $A > 0$ such that when T is large enough, $P(\inf_{Ah_T\leq\kappa\leq T} \inf_{\|D_T\phi\|\geq d_T} Q_\kappa(\tau, \phi) < B) < \epsilon$.

Under Assumption 6, we have $\theta_{j+1}^0(\tau) - \theta_j^0(\tau) = \tilde{D}_T^{-1} O_p(\nu_T)$ for $j = 1, ..., m$, and moreover, $\theta_i^0(\tau) - \theta_j^0(\tau) = \tilde{D}_T^{-1} O_p(\nu_T)$ for all i and j. We can also prove that $P(|\hat{T}_j - T_j^0| \leq T^{\frac{1}{2}}\nu_T^{-1}) \to 1$ as $T \to \infty$. For simplicity, consider the case of two structural changes with break dates T_j^0 for $j = 1, 2$, which are estimated by $T^b = (\hat{T}_1, \hat{T}_2)$. We prove $\nu_T^2(\hat{T}_j - T_j^0) = O_p(1)$ by contradiction. Suppose for any $s > 0$, $|\hat{T}_2 - T_2^0| \geq s\nu_T^{-2}$ holds with positive probability. Without loss of generality, assume $\hat{T}_2 > T_2^0$. Consider the partition using $T^{b*} = (\hat{T}_1, T_2^0)$. We have

$$\begin{aligned} & S_T(\tau, \hat{\theta}(\tau), T^b) - S_T(\tau, \hat{\theta}^*(\tau), T^{b*}) \\ \geq\ & S_T(\tau, \hat{\theta}(\tau), T^b) - S_T(\tau, \hat{\theta}(\tau), T^{b*}) \\ =\ & \sum_{t=T_2^0+1}^{\hat{T}_2} q_{\tau,t}(\hat{\theta}_2(\tau) - \theta_3^0(\tau)) - \sum_{t=T_2^0+1}^{\hat{T}_2} q_{\tau,t}(\hat{\theta}_3(\tau) - \theta_3^0(\tau)). \end{aligned}$$

Since $\|D_T(\hat{\theta}_2(\tau) - \theta_3^0(\tau))\| = O_p(T^{\frac{1}{2}}\nu_T)$ and $\|D_T(\hat{\theta}_3(\tau) - \theta_3^0(\tau))\| = O_p(1)$, we can obtain that the first summation on the right hand side of the above equality is strictly positive and dominates the second term. Thus, the partition T^{b*} yields a smaller objective function value, which contradicts the fact that T^b minimizes the objective function. Consequently, we have $\nu_T^2(\hat{T}_j - T_j^0) = O_p(1)$ for each j. Hence, the break fraction estimate $\hat{\lambda}_j$ is a consistent estimator of λ_j^0 and the rate of convergence is $\nu_T^2 T$.

A.4 Proof of Theorem 4

As Theorem 3 states, $\hat{\lambda}_j$ is a consistent estimator of λ_j^0 and $\nu_T^2 T(\hat{\lambda}_j - \lambda_j^0) = O_p(1)$ for $j = 1, ..., m$. The limit distribution of the break date estimates are given by the following minimization problem:

$$\inf_{T^b \in \Upsilon} [S_T(\tau, \theta(\tau), T^b) - S_T(\tau, \theta(\tau), T^0)],$$

where $\Upsilon = \{T_j : T_j = T_j^0 + [s\nu_T^{-2}] \text{ and } |s| \leq A < \infty, j = 1, ..., m\}$.

Consider the case where $\hat{T}_j \leq T_j^0$ for all $j = 1, ..., m$. The objective function is

$$\begin{aligned} & [S_T(\tau, \theta(\tau), T^b) - S_T(\tau, \theta(\tau), T^0)] \\ = & \sum_{j=1}^{m} \sum_{t=\hat{T}_j+1}^{T_j^0} \{\rho_\tau(y_t - z_t'\theta_{j+1}^0(\tau)) - \rho_\tau(y_t - z_t'\theta_j^0(\tau))\} + o_p(1). \end{aligned}$$

Then, $\sum_{t=\hat{T}_j+1}^{T_j^0} \{\rho_\tau(y_t - z_t'\theta_{j+1}^0(\tau)) - \rho_\tau(y_t - z_t'\theta_j^0(\tau))\}$ is the objective function for each j that delivers the limit distribution of $\hat{T}_j$. Moreover, we have

$$\begin{aligned} & \rho_\tau(y_t - z_t'\theta_{j+1}^0(\tau)) - \rho_\tau(y_t - z_t'\theta_j^0(\tau)) \\ = & \rho_\tau(z_t'\theta_j^0(\tau) + u_t^0(\tau) - z_t'\theta_{j+1}^0(\tau)) - \rho_\tau(u_t^0(\tau)) \\ = & -z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))\psi_\tau(u_t^0(\tau)) \\ & +(u_t^0(\tau) - z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))) \\ & \times\{I[0 > u_t^0(\tau) > z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))] - I[0 < u_t^0(\tau) < z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))]\}. \end{aligned}$$

Thus, the objective function for each j is

$$\begin{aligned} - & (\theta_{j+1}^0(\tau) - \theta_j^0(\tau))' \sum_{t=T_j^0+[s\nu_T^{-2}]+1}^{T_j^0} z_t\psi_\tau(u_t^0(\tau)) \\ + & \sum_{t=T_j^0+[s\nu_T^{-2}]+1}^{T_j^0} (u_t^0(\tau) - z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))) \\ & \times\{I[0 > u_t^0(\tau) > z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))] - I[0 < u_t^0(\tau) < z_t'(\theta_{j+1}^0(\tau) - \theta_j^0(\tau))]\}. \end{aligned}$$

For the first term, we have

$$
\begin{aligned}
& (\theta^0_{j+1}(\tau) - \theta^0_j(\tau))' \sum_{t=T^0_j+[s\nu_T^{-2}]+1}^{T^0_j} z_t \psi_\tau(u^0_t(\tau)) \\
= \quad & \delta'_j(\tau) \left[\begin{array}{c} \nu_T \sum_{t=T^0_j+[s\nu_T^{-2}]+1}^{T^0_j} \psi_\tau(u^0_t(\tau)) \\ T^{-\frac{1}{2}} \nu_T \sum_{t=T^0_j+[s\nu_T^{-2}]+1}^{T^0_j} x_t \psi_\tau(u^0_t(\tau)) \end{array} \right] \\
\Rightarrow \quad & \delta'_j(\tau) \left[\begin{array}{c} B_j(s) \\ Q_j(s) \end{array} \right] \\
= \quad & \delta'_j(\tau) V_j(s),
\end{aligned}
$$

where $B_j(s)$ is a two-sided Brownian motion with variance ω_ψ, $Q_j(s)$ is a k-dimensional random vector that depends on j, s, and $\Omega_{v\psi}$, and $V_j(s) = (B_j(s), Q_j(s)')'$.

If the regressors are strictly exogenous, we have the following simplified convergence condition[27]:

$$
\begin{aligned}
& (\theta^0_{j+1}(\tau) - \theta^0_j(\tau))' \sum_{t=T^0_j+[s\nu_T^{-2}]+1}^{T^0_j} z_t \psi_\tau(u^0_t(\tau)) \\
= \quad & \delta'_j(\tau) \left[\begin{array}{c} \nu_T \sum_{t=T^0_j+[s\nu_T^{-2}]+1}^{T^0_j} \psi_\tau(u^0_t(\tau)) \\ T^{-\frac{1}{2}} \nu_T \sum_{t=T^0_j+[s\nu_T^{-2}]+1}^{T^0_j} x_t \psi_\tau(u^0_t(\tau)) \end{array} \right] \\
\Rightarrow \quad & \delta'_j(\tau) \left[\begin{array}{c} B_j(s) \\ B_v(\lambda^0_j) B_j(s) \end{array} \right] \\
= \quad & \delta'_j(\tau) \bar{B}_v(\lambda^0_j) B_j(s),
\end{aligned}
$$

where $\bar{B}_v(\lambda^0_j) = (1, B'_v(\lambda^0_j))'$.

For the second term, using similar arguments as those from the proof of Theorem 1, we

[27] The convergence conditions are similar to the arguments on page 63 from Kejriwal and Perron (2008).

have

$$\sum_{t=T_j^0+[s\nu_T^{-2}]+1}^{T_j^0} (z_t'(\theta_{j+1}^0(\tau)-\theta_j^0(\tau))-u_t^0(\tau))I[0<u_t^0(\tau)<z_t'(\theta_{j+1}^0(\tau)-\theta_j^0(\tau))]$$
$$\Rightarrow \frac{|s|}{2}\delta_j'(\tau)f(F^{-1}(\tau))\bar{B}_v(\lambda_j^0)\bar{B}_v'(\lambda_j^0)\delta_j(\tau)I[\delta_j'(\tau)\bar{B}_v'>0]$$

and

$$\sum_{t=T_j^0+[s\nu_T^{-2}]+1}^{T_j^0} (u_t^0(\tau)-z_t'(\theta_{j+1}^0(\tau)-\theta_j^0(\tau)))I[0>u_t^0(\tau)>z_t'(\theta_{j+1}^0(\tau)-\theta_j^0(\tau))]$$
$$\Rightarrow \frac{|s|}{2}\delta_j'(\tau)f(F^{-1}(\tau))\bar{B}_v(\lambda_j^0)\bar{B}_v'(\lambda_j^0)\delta_j(\tau)I[\delta_j'(\tau)\bar{B}_v'\leq 0].$$

Thus, the limit distribution of the second term is

$$\frac{|s|}{2}\delta_j'(\tau)f(F^{-1}(\tau))\bar{B}_v(\lambda_j^0)\bar{B}_v'(\lambda_j^0)\delta_j(\tau).$$

Therefore, when $s \leq 0$ we have the limit distribution of the break point estimates as follows:

$$\begin{aligned}\hat{s} &= \nu_T^2(\hat{T}_j-T_j^0)=\nu_T^2T(\hat{\lambda}_j-\lambda_j^0)\\ &\Rightarrow \arg\max_s\left\{\delta_j'(\tau)V_j(s)-\frac{|s|}{2}\delta_j'(\tau)f(F^{-1}(\tau))\bar{B}_v(\lambda_j^0)\bar{B}_v'(\lambda_j^0)\delta_j(\tau)\right\}.\end{aligned}$$

If under strict exogeneity, we have

$$\begin{aligned}\hat{s} &= \nu_T^2(\hat{T}_j-T_j^0)=\nu_T^2T(\hat{\lambda}_j-\lambda_j^0)\\ &\Rightarrow \arg\max_s\left\{\delta_j'(\tau)\bar{B}_v(\lambda_j^0)B_j(s)-\frac{|s|}{2}\delta_j'(\tau)f(F^{-1}(\tau))\bar{B}_v(\lambda_j^0)\bar{B}_v'(\lambda_j^0)\delta_j(\tau)\right\}.\end{aligned}$$

Similarly, for the case where $\hat{T}_j > T_j^0$, the first term becomes

$$
\begin{aligned}
& (\theta_{j+1}^0(\tau) - \theta_j^0(\tau))' \sum_{t=T_j^0+1}^{T_j^0+[s\nu_T^{-2}]} z_t \psi_\tau(u_t^0(\tau)) \\
= \quad & \delta_j'(\tau) \begin{bmatrix} \nu_T \sum_{t=T_j^0+1}^{T_j^0+[s\nu_T^{-2}]} \psi_\tau(u_t^0(\tau)) \\ T^{-\frac{1}{2}} \nu_T \sum_{t=T_j^0+1}^{T_j^0+[s\nu_T^{-2}]} x_t \psi_\tau(u_t^0(\tau)) \end{bmatrix} \\
\Rightarrow \quad & \delta_j'(\tau) \begin{bmatrix} B_{j+1}(s) \\ Q_{j+1}(s) \end{bmatrix} \\
= \quad & \delta_j'(\tau) V_{j+1}(s).
\end{aligned}
$$

Thus, when $s > 0$ the limit distribution of the break point estimates is given by

$$
\begin{aligned}
\hat{s} \quad = \quad & \nu_T^2(\hat{T}_j - T_j^0) = \nu_T^2 T(\hat{\lambda}_j - \lambda_j^0) \\
\Rightarrow \quad & \arg\max_s \left\{ \delta_j'(\tau) V_{j+1}(s) - \frac{|s|}{2} \delta_j'(\tau) f(F^{-1}(\tau)) \bar{B}_v(\lambda_j^0) \bar{B}_v'(\lambda_j^0) \delta_j(\tau) \right\}.
\end{aligned}
$$

If under strict exogeneity, we have

$$
\begin{aligned}
\hat{s} \quad = \quad & \nu_T^2(\hat{T}_j - T_j^0) = \nu_T^2 T(\hat{\lambda}_j - \lambda_j^0) \\
\Rightarrow \quad & \arg\max_s \left\{ \delta_j'(\tau) \bar{B}_v(\lambda_j^0) B_{j+1}(s) - \frac{|s|}{2} \delta_j'(\tau) f(F^{-1}(\tau)) \bar{B}_v(\lambda_j^0) \bar{B}_v'(\lambda_j^0) \delta_j(\tau) \right\}.
\end{aligned}
$$

Appendix B

Proofs for Chapter 3

B.1 Proof of Proposition 2

This proof follows the procedure in Xiao (2009). Denote $\hat{u} = D_T(\hat{\gamma}(\tau) - \gamma(\tau))$, we have

$$\begin{aligned}\rho_\tau(y_t - \hat{\gamma}(\tau)' z_{t-1}) &= \rho_\tau(\gamma(\tau)' z_{t-1} + \varepsilon_{2t\tau} - \hat{\gamma}(\tau)' z_{t-1}) \\ &= \rho_\tau(\varepsilon_{2t\tau} - (\hat{\gamma}(\tau)' - \gamma(\tau)') z_{t-1}) \\ &= \rho_\tau(\varepsilon_{2t\tau} - (D_T^{-1}\hat{u})' z_{t-1}).\end{aligned}$$

Thus, the minimization problem can be rewritten as $\min_u G_T(u)$, where the objective function $G_T(u) = \sum_{t=1}^{T} [\rho_\tau(\varepsilon_{2t\tau} - (D_T^{-1}u)' z_{t-1}) - \rho_\tau(\varepsilon_{2t\tau})]$. From Xiao (2009), for any $u \neq 0$,

$$\rho_\tau(u - v) - \rho_\tau(u) = -v\psi_\tau(u) + (u - v)I(0 > u > v) - I(0 < u < v).$$

Then, we have

$$\begin{aligned} G_T(u) &= -\sum_{t=1}^{T}(D_T^{-1}u)'z_{t-1}\psi_\tau(\varepsilon_{2t\tau}) + \sum_{t=1}^{T}(\varepsilon_{2t\tau} - (D_T^{-1}u)'z_{t-1}) \\ &\quad \times[I(0 > \varepsilon_{2t\tau} > (D_T^{-1}u)'z_{t-1}) - I(0 < \varepsilon_{2t\tau} < (D_T^{-1}u)'z_{t-1})] \\ &= -\sum_{t=1}^{T}(D_T^{-1}u)'z_{t-1}\psi_\tau(\varepsilon_{2t\tau}) + \sum_{t=1}^{T}((D_T^{-1}u)'z_{t-1} - \varepsilon_{2t\tau}) \\ &\quad \times[I(0 < \varepsilon_{2t\tau} < (D_T^{-1}u)'z_{t-1}) - I(0 > \varepsilon_{2t\tau} > (D_T^{-1}u)'z_{t-1})]. \end{aligned}$$

For the first term, since $T^{-1}\sum x_{t-1}\psi_\tau(\varepsilon_{2t\tau}) \to_d \int \omega J_c dZ_\psi$, the following condition holds

$$D_T^{-1}\sum_{t=1}^{T} z_{t-1}\psi_\tau(\varepsilon_{2t\tau}) = \begin{bmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T}\psi_\tau(\varepsilon_{2t\tau}) \\ T^{-1}\sum_{t=1}^{T} x_{t-1}\psi_\tau(\varepsilon_{2t\tau}) \end{bmatrix} \to_d \begin{bmatrix} \int dZ_\psi \\ \int \omega J_c dZ_\psi \end{bmatrix}.$$

For the second term, we start with $W_T(u) = \sum_{t=1}^{T}(u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 < \varepsilon_{2t\tau} < u'D_T^{-1}z_{t-1})$. Denote $\xi_t(u) = (u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 < \varepsilon_{2t\tau} < u'D_T^{-1}z_{t-1})$. Define $W_{Tm}(u) = \sum_{t=1}^{T}\xi_{tm}(u)$, where

$$\xi_{tm}(u) = (u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 < \varepsilon_{2t\tau} < u'D_T^{-1}z_{t-1})I(u'D_T^{-1}z_{t-1} \leq m)$$

is the truncation of $u'D_T^{-1}z_{t-1}$ at some finite number $m > 0$. Denote the conditional expectation $\bar{\xi}_{tm}(u) = E[\xi_{tm}(u)|\mathcal{F}_{t-2}]$ and $\bar{W}_{Tm}(u) = \sum_{t=1}^{T}\bar{\xi}_{tm}(u)$, where $\mathcal{F}_{t-2}$ is information set up to time $t-1$, and $z_{t-1} \in \mathcal{F}_{t-2}$. Thus, we have

$$
\begin{aligned}
&\bar{W}_{Tm}(u) \\
= &\sum_{t=1}^{T} E[(u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 < \varepsilon_{2t\tau} < u'D_T^{-1}z_{t-1})I(u'D_T^{-1}z_{t-1} \leq m)|\mathcal{F}_{t-2}] \\
= &\sum_{t=1}^{T} E[(u'D_T^{-1}z_{t-1} + F^{-1}(\tau) - \varepsilon_{2t})I(F^{-1}(\tau) < \varepsilon_{2t} < u'D_T^{-1}z_{t-1} + F^{-1}(\tau)) \\
&\times I(u'D_T^{-1}z_{t-1} \leq m)|\mathcal{F}_{t-2}] \\
= &\sum_{t=1}^{T} \int_{F^{-1}(\tau)}^{[u'D_T^{-1}z_{t-1}+F^{-1}(\tau)]I(u'D_T^{-1}z_{t-1}\leq m)} [\int_{r}^{[u'D_T^{-1}z_{t-1}+F^{-1}(\tau)]I(u'D_T^{-1}z_{t-1}\leq m)} ds] \\
&\times f_{t-1}(r)dr \\
= &\sum_{t=1}^{T} \int_{F^{-1}(\tau)}^{[u'D_T^{-1}z_{t-1}+F^{-1}(\tau)]I(u'D_T^{-1}z_{t-1}\leq m)} [\int_{F^{-1}}^{s} f_{t-1}(r)dr]ds \\
= &\sum_{t=1}^{T} \int_{F^{-1}(\tau)}^{[u'D_T^{-1}z_{t-1}+F^{-1}(\tau)]I(u'D_T^{-1}z_{t-1}\leq m)} [s - F^{-1}(\tau)][\frac{F_{t-1}(s) - F_{t-1}(F^{-1}(\tau))}{s - F^{-1}(\tau)}]ds \\
= &\sum_{t=1}^{T} \int_{F^{-1}(\tau)}^{[u'D_T^{-1}z_{t-1}+F^{-1}(\tau)]I(u'D_T^{-1}z_{t-1}\leq m)} [s - F^{-1}(\tau)]f_{t-1}(F^{-1}(\tau))ds + op(1) \\
= &\sum_{t=1}^{T} f_{t-1}(F^{-1}(\tau)) \left[\frac{[s - F^{-1}(\tau)]^2}{2} |_{F^{-1}(\tau)}^{[u'D_T^{-1}z_{t-1}+F^{-1}(\tau)]I(u'D_T^{-1}z_{t-1}\leq m)} \right] + op(1) \\
= &\frac{1}{2}\sum_{t=1}^{T} f_{t-1}(F^{-1}(\tau))[u'D_T^{-1}z_{t-1}]^2 I(u'D_T^{-1}z_{t-1} \leq m) + op(1) \\
= &\frac{1}{2T}\sum_{t=1}^{T} f_{t-1}(F^{-1}(\tau))u'[\sqrt{T}D_T^{-1}z_{t-1}z_{t-1}'D_T^{-1}\sqrt{T}]uI(u'D_T^{-1}z_{t-1} \leq m) + op(1).
\end{aligned}
$$

By Assumption 9 and the stationarity of $\{f_{t-1}(F^{-1}(\tau))\}$, for some $\epsilon > 0$,

$$\sup_{0\leq r\leq 1} |\frac{1}{T^{1-\epsilon}} \sum_{t=1}^{[Tr]} [f_{t-1}(F^{-1}(\tau)) - f(F^{-1}(\tau))]| \to_d 0.$$

Therefore, we have

$$
\begin{aligned}
\bar{W}_{Tm}(u) \quad &\to_d \quad \frac{1}{2}f(F^{-1}(\tau))u'[\int B_z B_z']I(0 < u'B_z \leq m)u \\
&:= \quad \eta_m,
\end{aligned}
$$

where $B_z = (1, \omega J_c)'$. Since $(u'D_T^{-1}z_{t-1})I(0 \leq u'D_T^{-1}z_{t-1} \leq m) \rightarrow_p 0$ uniformly in t,

$$\begin{aligned}\sum_{t=1}^{T} E[\xi_{tm}(u)^2|\mathcal{F}_{t-2}] &\leq \max\{(u'D_T^{-1}z_{t-1})I(0 \leq u'D_T^{-1}z_{t-1} \leq m)\} \times \sum \bar{\xi}_{tm}(u) \\ &\rightarrow_p 0.\end{aligned}$$

Thus, $\sum_{t=1}^{T}[\xi_{tm}(u) - \bar{\xi}_{tm}(u)] \rightarrow_p 0$, where $\xi_{tm}(u) - \bar{\xi}_{tm}(u)$ is a MDS.
By the Asymptotic Equivalence lemma, the limiting distribution of $\sum_{t=1}^{T} \xi_{tm}(u)$ is the same as that of $\sum_{t=1}^{T} \bar{\xi}_{tm}(u)$ i.e. $W_{Tm}(u) \rightarrow_d \eta_m$. As $m \rightarrow \infty$, we have

$$\eta_m \rightarrow_d \frac{1}{2} f(F^{-1}(\tau))u'[\int B_z B_z']uI(u'B_z > 0) = \eta.$$

We have $\lim_{m\rightarrow\infty} \limsup_{n\rightarrow\infty} Pr[|W_T(u) - W_{Tm}(u)| \geq \epsilon] = 0$, since

$$\begin{aligned}&Pr[|W_T(u) - W_{Tm}(u)| \geq 0] \\ =\ & Pr[\sum_{t=1}^{T}(u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 < \varepsilon_{2t\tau} < u'D_T^{-1}z_{t-1})I(u'D_T^{-1}z_{t-1} > m)] \\ \leq\ & Pr[\bigcup_t \{u'D_T^{-1}z_{t-1} > m\}] \\ =\ & Pr[\max_t \{u'D_T^{-1}z_{t-1}\} > m]\end{aligned}$$

and $\lim_{m\rightarrow\infty} Pr[\sup_{0\leq r\leq 1} u'B_z(r) > m] = 0$. Also, we have $W_T \rightarrow_d \eta$ i.e.

$$\begin{aligned}&\sum_{t=1}^{T}(u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 < \varepsilon_{2t\tau} < u'D_T^{-1}z_{t-1}) \\ \rightarrow_d\ & \frac{1}{2} f(F^{-1}(\tau))u'[\int B_z B_z']uI(u'B_z > 0).\end{aligned}$$

Similarly,

$$\begin{aligned}&\sum_{t=1}^{T}(u'D_T^{-1}z_{t-1} - \varepsilon_{2t\tau})I(0 > \varepsilon_{2t\tau} > u'D_T^{-1}z_{t-1}) \\ \rightarrow_d\ & \frac{1}{2} f(F^{-1}(\tau))u'[\int B_z B_z']uI(u'B_z < 0).\end{aligned}$$

Thus, the second term of the objective function converges to $\frac{1}{2}f(F^{-1}(\tau))u'[\int B_z B_z']u$ in distribution. The limiting distribution of the objective function is as follows:

$$G_T(u) \to_d -u' \begin{bmatrix} \int dZ_\psi \\ \int \omega J_c dZ_\psi \end{bmatrix} + \frac{1}{2}f(F^{-1}(\tau))u' \begin{bmatrix} 1 & \int \omega J_c \\ \int \omega J_c & \int (\omega J_c)^2 \end{bmatrix} u := G(u).$$

Consider the first order condition of the minimization problem of $G(u)$:

$$\frac{\partial G(u)}{\partial u'} = - \begin{bmatrix} \int dZ_\psi \\ \int \omega J_c dZ_\psi \end{bmatrix} + f(F^{-1}(\tau)) \begin{bmatrix} 1 & \int \omega J_c \\ \int \omega J_c & \int (\omega J_c)^2 \end{bmatrix} u = 0,$$

the solution of which is the following:

$$u = \frac{1}{f(F^{-1}(\tau))} \begin{bmatrix} 1 & \int \omega J_c \\ \int \omega J_c & \int (\omega J_c)^2 \end{bmatrix}^{-1} \begin{bmatrix} \int dZ_\psi \\ \int \omega J_c dZ_\psi \end{bmatrix}.$$

Therefore, $G_T(u)$ is minimized at $\hat{u} = D_T(\hat{\gamma}(\tau) - \gamma(\tau))$ and $\hat{u} \to_d u$ (Knight 1989, Xiao 2009). Thus,

$$D_T(\hat{\gamma}(\tau) - \gamma(\tau)) \to_d \frac{1}{f(F^{-1}(\tau))} \begin{bmatrix} 1 & \int \omega J_c \\ \int \omega J_c & \int (\omega J_c)^2 \end{bmatrix}^{-1} \begin{bmatrix} \int dZ_\psi \\ \int \omega J_c dZ_\psi \end{bmatrix}.$$

B.2 Proof of Proposition 3

We have $\psi_\tau(\varepsilon_{2t\tau}) = \tau - I(\varepsilon_{2t\tau} < 0) = \tau - I(\varepsilon_{2t} < F^{-1}(\tau))$. The mean and variance of $\psi_\tau(\varepsilon_{2t\tau})$ are $E[\psi_\tau(\varepsilon_{2t\tau})] = 0$ and $var[\psi_\tau(\varepsilon_{2t\tau})] = \tau(1-\tau)$. Then, we have

$$\Omega_\tau = \begin{bmatrix} \omega^2 & \omega\sqrt{\tau(1-\tau)}\delta \\ \omega\sqrt{\tau(1-\tau)}\delta & \tau(1-\tau) \end{bmatrix}.$$

Thus,

$$\begin{aligned} Z_\psi &= \sqrt{\tau(1-\tau)}B_\psi, \quad \text{where } B_\psi \text{ is a standard BM} \\ &= \sqrt{\tau(1-\tau)}[\delta B_1 + \sqrt{1-\delta^2}\tilde{B}_\psi], \quad \text{where } \tilde{B}_\psi \text{ is independent of } B_1. \end{aligned}$$

Therefore,

$$
\begin{aligned}
t_{\gamma_1}(\tau) &\to_d \frac{\int J_c^\mu dZ_{2\tau}}{[\int (J_c^\mu)^2]^{\frac{1}{2}}} \\
&= \frac{\sqrt{\tau(1-\tau)}\int J_c^\mu d(\delta B_1 + \sqrt{1-\delta^2}\tilde{B}_\psi)}{[\int (J_c^\mu)^2]^{\frac{1}{2}}} \\
&= \sqrt{\tau(1-\tau)}\left[\delta\frac{\int J_c^\mu dB_1}{[\int (J_c^\mu)^2]^{\frac{1}{2}}} + \sqrt{1-\delta^2}\frac{\int J_c^\mu d\tilde{B}_\psi}{[\int (J_c^\mu)^2]^{\frac{1}{2}}}\right] \\
&= \sqrt{\tau(1-\tau)}\left[\delta\frac{\int J_c^\mu dB_1}{[\int (J_c^\mu)^2]^{\frac{1}{2}}} + \sqrt{1-\delta^2}z\right],
\end{aligned}
$$

where z is a standard normal random variable that is independent of (B_1, J_c).

Appendix C

Proofs for Chapter 4

C.1 Proof of Theorem 6

Theorem 6 can be proved following the procedure in section A.1 from Xiao (2009), since under the null hypothesis of cointegration the model in Chapter Four has the same asymptotic behavior as the quantile cointegrating regression considered by Xiao (2009). For a single quantile τ, the unknown quantile regression coefficients $\hat{\theta}(\tau)$ are estimated by the minimization problem:

$$\hat{\theta}(\tau) = (\hat{\alpha}'(\tau), \hat{\beta}'(\tau))' = \arg\min_{\theta(\tau)} \sum_{t=1}^{T} \rho_\tau(y_t - z_t'\theta(\tau)),$$

which is equivalent to

$$\min_{\phi} \sum \{\rho_\tau(u_t(\tau) - (D_T^{-1}\phi)'z_t) - \rho_\tau(u_t(\tau))\},$$

where $\hat{\phi} = D_T(\hat{\theta}(\tau) - \theta(\tau))$. Consider the objective function

$$G_T(\phi) = \sum \{\rho_\tau(u_t(\tau) - (D_T^{-1}\phi)'z_t) - \rho_\tau(u_t(\tau))\}$$

and $\hat{\phi}$ is a minimizer of $G_T(\phi)$. Similar to the convex random function discussed in Knight (1989), the objective function $G_T(\phi)$ is also convex. According to Knight (1989) and Pollard (1991), if the finite-dimensional distributions of $G_T(\phi)$ converge weakly to those of $G(\phi)$ and $G(\phi)$ has a unique minimum, the convexity of $G_T(\phi)$ implies that $\hat{\phi}$ converges in distribution to the minimizer of $G(\phi)$. This implies the consistency of $\hat{\theta}(\tau)$.

To derive the asymptotic distribution of the consistent estimator $\hat{\theta}(\tau)$, we have

$$\begin{aligned} G_T(\phi) &= -\sum_t (D_T^{-1}\phi)' z_t \psi_\tau(u_t(\tau)) + \sum_t (u_t(\tau) - (D_T^{-1}\phi)' z_t) \\ &\quad \times [I(0 > u_t(\tau) > (D_T^{-1}\phi)' z_t) - I(0 < u_t(\tau) < (D_T^{-1}\phi)' z_t)]. \end{aligned}$$

For the first term,

$$\begin{aligned} \phi' D_T^{-1} \sum z_t \psi_\tau(u_t(\tau)) &\Rightarrow \phi' \begin{bmatrix} \int B_d dB_\psi \\ \int B_x dB_\psi + \Delta_{x\psi} \end{bmatrix} \\ &= \phi' \left[B_z dB_\psi + \bar{\Delta}_{x\psi} \right], \end{aligned}$$

where $B_z = (B_d', B_x')'$, $\bar{\Delta}_{v\psi} = (0, \Delta_{x\psi}')'$, and $\Delta_{x\psi} = \sum_{t=0}^{\infty} E(\xi_{2t}\psi_\tau(u_0(\tau)))$ is the one-sided long run covariance between ξ_{2t} and $\psi_\tau(u_t(\tau))$.

Similar to Xiao (2009), the second term converges to $\frac{1}{2} f(F^{-1}(\tau))\phi' \left[\int B_z B_z'\right] \phi$. Consequently,

$$\begin{aligned} G_T(\phi) &\Rightarrow -\phi' \left[\int B_z dB_\psi + \bar{\Delta}_{x\psi} \right] + \frac{1}{2} f(F^{-1}(\tau))\phi' \left[\int B_z B_z' \right] \phi \\ &:= G(\phi), \end{aligned}$$

where ":=" signifies definitional equality. By the convexity lemma of Pollard (1991) and Knight (1989), both $G_T(\phi)$ and $G(\phi)$ are minimized at $\hat{\phi} = D_T(\hat{\theta}(\tau) - \theta(\tau))$ (Xiao 2009). By Lemma A of Knight (1989), from the first order condition

$$\frac{\partial G(\phi)}{\partial \phi'} = -\left[\int B_z dB_\psi + \bar{\Delta}_{x\psi} \right] + \frac{1}{2} f(F^{-1}(\tau)) \left[\int B_z B_z' \right] \phi = 0,$$

we obtain the limit distribution of the regression coefficient estimates such that

$$D_T(\hat{\theta}(\tau) - \theta(\tau)) \Rightarrow \left[f(F^{-1}(\tau)) \int_0^1 B_z B_z' \right]^{-1} \left[\int_0^1 B_z dB_\psi + \bar{\Delta}_{x\psi} \right].$$

C.2 Proof of Theorem 7

Similar to the case of fully modified OLS regression, the fully modified quantile estimator is consistent, which is also shown in the following derivation. The regression coefficient estimates after the modification $\hat{\theta}^+(\tau) = (\hat{\alpha}'(\tau), \hat{\beta}'^+(\tau))'$ are

$$\hat{\theta}^+(\tau) = \hat{\theta}(\tau) - \left[f(\widehat{F^{-1}(\tau)}) \sum z_t z_t' \right]^{-1} \left[\sum z_t \hat{\Omega}_{\psi x} \hat{\Omega}_{xx}^{-1} \Delta x_t + T \hat{\bar{\Delta}}_{x\psi}^+ \right],$$

where $\hat{\bar{\Delta}}_{x\psi}^+ = (0, \Delta_{x\psi}'^+)'$. Under cointegration and Assumptions 10-12, Theorem 6 holds so that

$$\begin{aligned} & D_T(\hat{\theta}(\tau) - \theta(\tau)) \\ \Rightarrow \; & \left[f(F^{-1}(\tau)) \int_0^1 B_z B_z' \right]^{-1} \left[\int_0^1 B_z dB_{\psi.x} + \int_0^1 B_z d\Omega_{\psi x} \Omega_{xx}^{-1} B_x + \bar{\Delta}_{x\psi} \right]. \end{aligned}$$

Then, we have

$$\begin{aligned} D_T(\hat{\theta}^+(\tau) - \theta(\tau)) &= D_T(\hat{\theta}(\tau) - \theta(\tau)) \\ & \quad - D_T \left[f(\widehat{F^{-1}(\tau)}) \sum z_t z_t' \right]^{-1} \left[\sum z_t \hat{\Omega}_{\psi x} \hat{\Omega}_{xx}^{-1} \Delta x_t + T \hat{\bar{\Delta}}_{x\psi}^+ \right] \\ & \Rightarrow \left[f(F^{-1}(\tau)) \int_0^1 B_z B_z' \right]^{-1} \int_0^1 B_z dB_{\psi.x} \\ & \sim MN \left(0, \frac{\omega_{\psi.x}^2}{f(F^{-1}(\tau))^2} \left[\int_0^1 B_z B_z' \right]^{-1} \right), \end{aligned}$$

where $\omega_{\psi.x}^2 = \omega_\psi^2 - \Omega_{\psi x} \Omega_{xx}^{-1} \Omega_{x\psi}$. Similar to $\hat{\theta}(\tau)$, the fully modified estimator $\hat{\theta}^+(\tau)$ is also consistent, with $D_T(\hat{\theta}^+(\tau) - \theta(\tau)) = O_p(1)$.

C.3 Proof of Theorem 8

For the fully modified quantile regression, we have $y_t^+ = y_t - \hat{\Omega}_{\psi x}\hat{\Omega}_{xx}^{-1}\Delta x_t$ and $\hat{u}_t^+(\tau) = y_t^+ - z_t'\hat{\theta}^+(\tau)$. The residuals of the quantile regression are transformed to $\psi_\tau(\hat{u}_t^+(\tau)) = \tau - I(\hat{u}_t^+(\tau) < 0)$. Under cointegration and Assumptions 10-12, the partial sum of $\psi_\tau(\hat{u}_t^+(\tau))$ is

$$\begin{aligned} T^{-\frac{1}{2}}\sum_{t=1}^{[Tr]}\psi_\tau(\hat{u}_t^+(\tau)) &\Rightarrow B_{\psi.x} - \left[\int_0^1 dB_{\psi.x}B_z'\right]\left[\int_0^1 B_zB_z'\right]^{-1}\int_0^r B_z(s) := \underline{B}_{\psi.x}(r) \\ &= \omega_{\psi.x}\left\{W_1 - \left[\int_0^1 dW_1S'\right]\left[\int_0^1 SS'\right]^{-1}\int_0^r S\right\} := \omega_{\psi.x}\underline{W}(r), \end{aligned}$$

where in general $S(r) = (B_d'(r), W_2'(r))'$ and $W_1(r)$ and $W_2(r)$ are one and k-dimensional standard Brownian motions that are independent of each other. If the regression contains an intercept but no time trend, then $S(r) = (1, W_2'(r))'$. Also, by definition $\underline{W}(r) = W_1 - \left[\int_0^1 dW_1S'\right]\left[\int_0^1 SS'\right]^{-1}\int_0^r S$.

The CUSUM test statistic is given by

$$\begin{aligned} CS_T(\tau) &= \max_{n=1,\ldots,T}\frac{1}{\hat{\omega}_{\psi.x}\sqrt{T}}\left|\sum_{t=1}^{n}\psi_\tau(\hat{u}_t^+(\tau))\right| \\ &\Rightarrow \sup_{0\le r\le 1}\left|\left\{W_1 - \left[\int_0^1 dW_1S'\right]\left[\int_0^1 SS'\right]^{-1}\int_0^r S\right\}\right|, \end{aligned}$$

where $\hat{\omega}_{\psi.x}^2 = \hat{\omega}_\psi^2 - \hat{\Omega}_{\psi x}\hat{\Omega}_{xx}^{-1}\hat{\Omega}_{x\psi}$.