THREE ESSAYS ON ECONOMETRICS OF NONLINEAR COINTEGRATION AND THRESHOLD EFFECTS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Haiqiang Chen

May 2011

THREE ESSAYS ON ECONOMETRICS OF NONLINEAR COINTEGRATION AND THRESHOLD EFFECTS

Haiqiang Chen, Ph.D.

Cornell University 2011

Traditional linear cointegration models have been widely used to examine long-run relationships between economic variables; however, empirical evidence suggests that the linear structure fails to account for economic changes due to technology improvement, business cycles and policy alterations. Nonlinear cointegration models provide an important means to extend conventional cointegration analysis by incorporating these factors. In the first chapter, I establish a statistical theory for cointegrating regressions with threshold effects. I derive asymptotics of the profiled least square (LS) estimators assuming the size of the threshold effect converges to zero. Depending on how rapidly this sequence converges, the model may be identified or weakly identified. A model-selection procedure is then applied to construct robust confidence intervals, which have approximately correct coverage probability irrespective of the magnitude of the threshold effect.

Using a parametric model, however, one always suffers from the danger of model misspecifications. The standard tests based on parametric models cannot tell us whether the rejection or acceptance of threshold effects is due to real regime shifts or a functional misspecification. In the second chapter, I consider the estimation and testing for threshold effects in regression models with unknown functional forms. I use series expansions to approximate the unknown regression functions and estimate the threshold effect with a profile least square

method. A nice property of the estimator is that it achieves T-convergence rate as in parametric models. I derive the asymptotic distribution of the threshold estimator and design a generalized sup Wald statistic to test the threshold effect.

In the third chapter, I consider an application of threshold cointegration on the price discovery for cross-listed stocks. For cross-listings, the convergence to equilibrium parity between home and guest market prices could be discontinuous, i.e., convergence may be quicker when the price deviation is sufficiently profitable. By considering the concept of threshold cointegration, I modify Harris et al.'s (1995, 2002) common factor approach to estimate the relative extent of market-respective contribution to price discovery. The method is applied to Canadian stocks cross-listed on the New York Stock Exchange (NYSE) and the Toronto Stock Exchange.

BIOGRAPHICAL SKETCH

Haiqiang Chen was born in Hunan, China. In 2003, he received his B.S. degree in Statistics and B.A. degree in Economics from Peking University, China. After that, he entered the department of economics at the Chinese University of Hong Kong, where he received his M.Phil. degree in 2005. He then continued his study in economics at Cornell University and earned his Ph.D. degree in 2011.

This document is dedicated to my parents and my wife.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my advisor, Prof. Yongmiao Hong, for his guidance, support and inspiration. I am very lucky to have learnt so much from him and I deeply appreciate his help. This thesis would not have been completed without him.

I thank Prof. Nicholas M. Kiefer and Prof. George H. Jakubson, for their helpful discussions and encouragement as well as serving on my committee. I thank Prof. Francesca Molinari and Prof. Joerg Stoye for their comments. I also thank Prof. Terence Tai-Leung Chong at the Chinese University of Hong Kong for his advices and help on my research.

Special thanks go to my friends at Cornell. I greatly appreciate Yingxing Li for her help and patience. I would also like to thank my parents for their unconditional love and always being there for me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Cointegrating Regressions with Threshold Effects

Cointegration analysis has been widely applied in economics and finance. Traditional cointegration analysis assumes a linear long-run relationship among integrated processes. In empirical applications, however, little evidence has been found to support this linear cointegration structure, see Park and Hahn (1999) and Xiao (2009). A variety of reasons have been proposed to explain this empirical frustration, leading to many extensions of linear cointegration models. Among these, a major extension is to consider a time-varying cointegrating vector given by

$$y_t = \alpha_t x_t + \varepsilon_t \tag{1.1}$$

where x_t are integrated regressors and ε_t is a stationary process. Such a time-varying cointegration relationship might result from technology improvement, business cycles and policy alterations.

However, in Model (1.1), y_t will not be an $I(1)$ process anymore if α_t is not a constant. Simulations also indicate that y_t could be a time series process with low persistency. Shi and Phillips (2010) view this property as an advantage of nonlinear transformations of integrated processes since they are helpful to model relationships between some weak dependent variables, such as asset returns and highly dependent variables, such as economic fundamentals. Some recent studies call these models as nonlinear or time-varying cointegration. Nevertheless, these definitions lose the economic beauty of the linear cointegration model which implies a common stochastic trend among economic integrated variables.

In the first chapter, following Model (1.1), I modify traditional cointegration models by considering a threshold effect on the cointegrating vector, whose sizes may be small in a statistical sense, but cause the failure of traditional cointegration tests. I assume the size of threshold effect be local to zero so the nonlinear effect only has a negligible impact on the memory property of y_t. This approach not only gives us a flexibility in modeling cointegrating relationships with some deviations from the linear structure, but also avoid the theoretical deficiency of nonlinear cointegration.

The main reason to choose threshold models is they offer a parsimonious approach to capture the time varying properties in economics models. Compared to other nonlinear models, threshold models are more powerful to capture small nonlinear deviations. Threshold effects are also very natural for characterizing some stylized facts of modern economies. For example, threshold models provide a useful framework for modelling the multiple equilibria implied by economic growth models with credit constraints.[1] Modelling asymmetry in economic relationships is another strength, since threshold models avoid the cumbersome cubic and higher order terms necessitated by other parametric nonlinear models.

The basic model I consider has two regimes, which may correspond to expansion and recession stages, normal and crisis periods, aggressive and passive policy regimes in the real world. For asymptotics purposes, the threshold vari-

[1]For example, Azariadis and Smith (1998) show that the economy could switch back and force between two long-run equilibrium regimes according to whether the credit constraint is binding or not. Similar results have been established between inflation rate and long run economy growth rates, inflation rate and financial sector performance, see Boyd et al.(2001).

able is assumed to be stationary ergodic and continuous.[2] I derive asymptotics of the profiled least square (LS) estimators assuming the size of the threshold effect converges to zero. Depending on how rapidly this sequence converges, the model may be identified or weakly identified.[3] In the former case, I show that the profiled least square estimators are consistent and that their confidence intervals (CIs) can be constructed through inversion of certain standard test statistics. For the latter, the estimators are inconsistent and their limiting distributions depend on some inestimable nuisance parameters. The standard method to construct CIs does not control the coverage probability. One way to deal with this problem is to take the supremum of quantiles for all possible values of nuisance parameters and then construct the least favorable CIs. These CIs have the correct asymptotic size under weak identification case, but can be unnecessarily long when the model is identified. Following Cheng (2008) and Shi and Phillips (2010), I apply a model-selection procedure to choose the CIs. It can be shown that the CIs chosen by this method have approximately correct coverage probability irrespective of the magnitude of the threshold effect. This model selection procedure can also be regarded as a pretest to determine the efficacy of the conventional t-test or Chi-square tests.

Endogeneity and serial correlation are common in empirical studies with integrated regressors; an extension of the model allows for these important features. Most previous nonlinear cointegration models assume error terms to be a martingale difference sequence, e.g. Park and Phillips (2001), Cai et al. (2009). This assumption is too restrictive compared to linear cointegrations, where the error terms are assumed to be stationary and the regressors are endogenous. In

[2] The simulations demonstrate that the method works well even the threshold variable is highly persistent.

[3] Threshold effects with fixed sizes can be viewed as special cases with zero convergence rate.

the extended model, I assume the error term to be an AR(1) process, and use leads and lags of innovations as extra regressors to deal with endogeneity. I design a Cochrane-Orcutt-type feasible generalized least square (FGLS) estimator to estimate the model. It is well known that, in linear cointegration models the FGLS estimator cannot improve the estimation, as Phillips and Park (1988) demonstrate by establishing their asymptotic equivalence. However, this equivalence does not hold when there exist regime shifts. I analytically and numerically show that the FGLS estimation improves LS estimation in the presence of serial correlation.

Another attraction of the FGLS estimator is its robustness with respect to different error specifications, including I(1) errors. This robustness allows testing the existence of regime shifts without knowing whether cointegration is present. Compared to Gregory and Hansen (1996), who design a robust cointegration test without knowledge of the existence of a change point, I test the hypotheses of regime shifts and cointegrating relationship in the opposite way. I first design a sup-Wald statistic based on the FGLS estimator to test the existence of regime shifts, and then apply residual-based test statistics to test cointegration given the conclusion from the first step. The model selection procedure is applied to construct robust cointegration tests. Monte Carlo simulations show that these test statistics perform reasonably well.

Finally, I provide an empirical application of my model to the asymmetric effects of monetary policy on real output under different credit conditions. Blinder (1987) develops a model consisting of two equilibria: a Keynesian equilibrium and a credit-rationed equilibrium, showing that the effects of monetary policy could be rather weak in the Keynesian regime and rather strong in the credit-rationed regime. Azariadis and Smith (1998) develop a similar model and claim that the economy could switch back and forth between a Walrasian regime and a credit-rationing regime. Various empirical studies examine this asymmetric relationship; see McCallum (1991), Galbraith (1996), Balke (2000). However, all of these are restricted to the stationary framework due to the lack of theoretical work on the threshold model with integrated regressors. Given the fact that real output and monetary supply variables are very likely to be unit roots, the model presented in this paper can be expected to generate more reliable results. My finding confirms the existence of asymmetric effects between monetary policy and output. Both monetary policy and fiscal policy have larger effects on the real output during the credit-rationed regime than normal regime without credit rationing.

1.2 Threshold Effects in Regression Models with Unknown Functional Forms

Using a parametric model, however, one always suffers from the danger of model misspecifications. The standard tests based on parametric models cannot tell us whether the rejection or acceptance of threshold effects is due to real regime shifts or a functional misspecification. For change-point models, Hidalgo (1995) shows, both theoretically and through Monte Carlo simulations, that when the model is misspecified, the test for structural change will reject the (in fact true) null hypothesis of no structural change with probability tending to one as the sample size increases. Since change-point model is a special case of threshold model, we face a similar risk for threshold models. From these considerations, it is desirable to design a general method by relaxing the parametric assumption on the functional form in the detection of threshold effects.

In the second chapter, I propose a procedure free of any parametric assumption to detect and test threshold effects in regressions. I consider a model as follows

$$y_t = \left\{ \begin{array}{l} g_1(x_t) + u_t, z_t \leq \gamma_0 \\ \\ g_2(x_t) + u_t, z_t > \gamma_0 \end{array} \right\} \tag{1.2}$$

where γ_0 is the threshold and z_t is a random variable. The unknown functions $g_i(x_t)$ for $i = 1, 2$ are assumed to be smooth. Model (1.2) is not uncommon in economics. For example, for a macroeconomic time series, the sample interval could cover both normal and crisis states, with two different relationship between y_t and x_t in these two states. Elliott and Timmermann (2005) point out that, nonlinear models are usually able to adapt rapidly to events with high economic uncertainty, whereas linear models only adjust to these changes sluggishly. Thus, a natural idea in economic modeling is to use more adaptive nonlinear models during crisis periods to capture fast changes, but use stable linear models to get the benefits of more precisely estimated parameters during normal periods. In a nonparametric setting, we can let the data choose a suitable model in each state.

The current study is related to the literature on detecting and testing discontinuities or jumps in a nonparametric model; see Yin (1988), Müller(1992), Delgado and Hidalgo(2000) and Gao et al.(2008). These authors' methods are based on the use of a one-sided kernel smoother, first introduced by Müller(1992). The basic idea is that the left-hand and right-hand side estimates converge to the left and right limit, respectively, at the change points. The difference between these estimates is used to construct the statistic for the detection of a change in conditional mean function $E(y|x)$. In these models, the sample splitting variable z_t is the regression variable x_t itself and the methods focus only on jumps. However, in my approach, I allow z_t to be different from the regressor x_t and the types of change to be more general (see examples in Section 2). This study is also related to the literature on nonparametric regressions with both continuous and categorical regressors. See Li and Racine (2003), Racine et al.(2006) and Li et al.(2009). Briefly, one can define a categorical regressor z_t^* as $z_t^* = 1$ if $z_t \leq \gamma_0$ and 0 otherwise. Testing the threshold effect in the model (1.2) could be viewed as testing whether the discrete variable z_t^* is significant or not. The difference is that, in my case, the categorical regressor z_t^* still depends on another variable z_t and an unknown parameter γ_0; thus, earlier method cannot be applied directly.

I use series expansions to approximate the unknown regression functions and estimate the threshold effect using a profile least square methods.[4] Series estimation is a global smoothing approach which has already been applied to estimate different nonparametric models in economics; see Gallant and Souza (1991) and Newey (1997) for general nonparametric models. For some recent work in specific models, refer to Li (2000) in the case of an additive partially linear model, Baltagi and Li (2002) in the case of a partially linear panel data model and Huang et al (2002) in the case of a varying coefficient model. The related convergence rate and asymptotic normality of the series estimators have been established by Andrews (1991) and Newey (1997). Series estimation has also been used for model specification testing (see Hong and White(1995) and Li et al. (2003) among others). Essentially, series estimation method uses a series expansion to approximate an unknown function as a linear combination of basis functions. The number of basis functions is a smoothing parameter simi-

[4]Chen et al. (2008) and Zhou et al. (2010) apply wavelet analysis to detect jumps and cusps in nonparametric regressions.

lar to the bandwidth for local smoothing methods. Compared to kernel-based smoothing techniques and local polynomial fitting, series estimation has the several advantages (Li , 2000). First, it is very convenient for imposing certain types of restrictions, such as additive separability and shape-preserving. Second, series estimation methods convert nonparametric regression into a many normal means problem, which is simpler, at least for theoretical purposes. A third advantage comes from reduced computational costs, because series estimation only involves least squares and the data are summarized by relatively few estimated coefficients. Finally, the threshold effect can be conveniently tested since the difference of coefficients on basis functions can be used to construct the test statistics directly.

Under some regularity conditions, I show that the profile least square estimator of the threshold value can yield T convergence rate as in parametric models. This super convergence rate enables me to study the asymptotics of the series estimators in each subsample as the true threshold is known. I derive the asymptotic distributions for the threshold estimator and the series estimators in each regime. To test the significance of the threshold effect, I design a generalized super Wald test statistic based on the series estimation in each subsample. This statistic converges to a nonstandard distribution and I generate the critical value table using bootstrap techniques. The results can be extended to allow for multiple threshold effects and I show that all thresholds can be estimated by a sequential method. The Monte Carlo simulations show that the series-expansion based approach has better threshold estimation and test performance than traditional parametric methods.

To illustrate the usefulness of the method, I consider an empirical application examining the convergence hypothesis of economic growth across countries. Many studies have investigated this issue, based on the Mankiw, Romer and Weil (MRW, Mankiw et al. 1992) specification of the Solow (1956) growth model. An important assumption in these studies is that there exists an underlying common linear specification. However, this assumption has been challenged by many recent papers, which find strong evidence of model heterogeneity across countries over time. The sources of heterogeneity can be summarized into three categories: varying parameters, omitted variables and nonlinearity in the production function. Each of these sources corresponds to some modification of the basic MRW model. Durlauf and Johnson (1995) use a tree regression approach and find multiple growth patterns across countries, and later, Hansen (1996, 2000) uses a threshold regression model to test formally for the presence of a regime shift in growth models. These results consider the existence of varying parameters, but ignore the other two types of heterogeneity. By adapting the nonparametric setup to the production function, my approach addresses the third type of model heterogeneity and thus should be more reliable. The empirical results indicate multiple regimes of growth patterns across different countries. Poor countries grow faster than rich countries in general, but they may converge to a different steady state.

1.3 Threshold Cointegration and Price Discovery

In the third chapter, I implement the threshold error correction mechanism in estimating the relative extent of exchange-respective contribution to price discovery of the pairs of cross-listings and their original listings. The existing methods assume linear convergence of relative premiums to parity whereas I hinge on the reality that the premiums disappear quicker when it is profitably arbitrageable than otherwise.

Price discovery is search for an equilibrium price (Schreiber and Schwartz (1986)) and is a key function of a securities exchange. When a security is traded in multiple markets, it is often of interest to determine where and how price discovery occurs. Harris *et al.* (1995) and Hasbrouck (1995) examine the exchange-specific relative contribution to price discovery of fragmented stocks on the NYSE and other U.S. exchanges, and confirm leadership assumed by the NYSE. As for international cross-listing, Bacidore and Sofianos (2002) and Solnik (1996) suggest that price discovery mostly takes place in the home market where substantial information originates. Eun and Sabherwal (2003) report the U.S. host exchanges provide an important feedback effect to affect the prices of Canadian cross-listings, however, to a lesser extent than the Toronto Stock Exchange (TSX) does.

In the literature, there are two broad approaches to estimating the contribution of each market to price discovery of fragmented listings. Hasbrouck's (1995) *innovation variance approach* extracts the information shares by employing variance decomposition based on the vector moving average representation of an error correction model (ECM). Harris *et al.*'s (1995, 2002) *common factor approach* employs Gonzola and Granger's (1995) permanent-transitory decomposition of a cointegrated system to estimate the information share of each market. As Eun and Sabherwal (2003) point out, Hasbrouck's (1995) approach involves Cholesky factorization of the covariance matrix of the innovations to prices on various exchanges and yields multiple information shares. This may cause confounding identification of the venue of price discovery. Hasbrouck's

(2002) modification can be numerically onerous in implementation.[5]

Harris *et al.* (1995) associate error correction dynamics with price discovery of cross-listed pairs which are cointegrated[6] by the law of one price. The cointegrating vectors of the vector ECM (VECM) represent the long-run equilibrium (near-parity condition), while the error correction terms characterize the convergence mechanism, i.e. "the process whereby markets attempt to find equilibrium." Through this representation, one can assess the relative extent of the contribution made by each market to price discovery of fragmented stocks using the estimates of adjustment coefficients. If the price of a Canadian cross-listing on the NYSE responds sensitively to shocks from the TSX whereas the home exchange is largely unaffected by ripples occurring in the host market, price discovery can be deemed as predominantly taking place on the TSX. Harris *al.* (2002) buttress the method earlier formulated in Harris *al.* (1995) by incorporating a microstructure model where the price is assumed to be the sum of an efficient (permanent) price component and a (transitory) error term.[7]

However, an implicit assumption made by Harris *et al.*' (1995, 2002) is that adjustment to parity, the long-run equilibrium, is continuous and linear.[8] Various economic circumstances challenge such a restriction, particularly where transaction costs and policy intervention are present. Given the complexity

[5]See De Jong (2002), Harris *et al.* (2002), and Hasbrouck (2002) for further discussion.

[6]A group of multiple random-walk processes is cointegrated if, by definition, there exists a stationary linear combination of the processes. A time series is (weakly) stationary if the probability laws (of up to the second moments) are time-invariant.

[7]In Harris *et al.* (2002), the efficient price component is unobservable and reflects the underlying fundamentals. Gonzalo and Granger's (1995) permanent-transitory decomposition posits the permanent price as a linear combination of the observable prices where the normalized weights can be market-respective information shares. The higher the normalized weight of an exchange, the bigger the influence of setting the permanent price. It can be shown that the normalized weights are orthogonal to the adjustment coefficient vector, they can be conveniently obtained from an ECM.

[8]This linear convergence is also assumed in Hasbrouck (1995)'s approach.

of trading rules and indirect transaction costs, nonlinear convergence to parity captures the market more accurately. The rationale of nonlinear modeling is straightforward. A relatively small deviation in the price of a cross-listed stock from its parity-implied price can be unarbitrageable if the dollar spread is insufficient to cover the fees, commissions, liquidity shortfalls, and other related costs. In this case, the dollar premium or discount behaves like a near-unit root process and will not converge to parity. Arbitrage forces will activate as the spread widens beyond the "threshold."

To date, there is dearth of articles with a nonlinear framework in the literature. Among the few which have appeared, Rabinovitch et al. (2003) use a nonlinear threshold model to estimate the adjustment dynamics of the return deviations for 20 Chilean and Argentine cross-listings. Koumkwa and Susmel (2008) suggest two nonlinear adjustment models: the exponential smooth transition autoregressive (ESTAR) and the logarithmic smooth transition autoregressive (LSTAR) to delineate the relative premiums of Mexican ADRs. Chung et al. (2005) study the dynamic relationship between the prices of three Taiwanese ADRs and their underlying stocks using a threshold VECM. However, to my best knowledge, there is no paper which considers the nonlinear convergence between two market prices when estimating the information share for each market. Given the existence of nonlinear effects, traditional approaches based on linear ECM may generate biased estimation results and then consequently suggest some misleading conclusions about the importance of each market in the price discovery process.

Motivated by these considerations, I modify Harris et al.'s (1995, 2002) method to estimate exchange-respective information shares of Canadian cross-listed pairs traded on the NYSE and the TSX by considering threshold cointegration as per Balke and Fomby (1997). Departing from linear modeling, the information share is estimated from the outer-regime adjustment coefficients based on a two-regime threshold ECM since the error correction adjustment mechanism only exists in the outer-regime. I also consider a smooth counterpart of the threshold ECM, the smooth transition ECMs, where the transition between two regimes are gradually. The model is estimated nonparametrically and thus avoids the risk of model misspecification.

My method offers in numerous advantages. First, I can theoretically depict and empirically analyze the discrete dynamics of the "bumpy" parity-convergence which is frequently observed in the market due to various risk factors like information asymmetry and market friction. Second, a large deviation (outer regime) is believed to be more susceptible to new information, either public or private. In contrast, a small deviation (inner regime) can be due to noise trading and therefore there is little connection between price discovery and error correction dynamics.[9] The threshold ECM ideally incorporates such a dichotomy while the predecessor linear ECMs may overestimate the information share when there is no cointegration in the unprofitable inner regime.

I develop an equilibrium model for a risky asset cross-listed in two markets. Based on the equilibrium solutions, I show that the short term dynamics could be captured by three different econometric models: standard linear ECM,

[9]A similar idea is illustrated by Gonzalo and Marinz (2006) in a model of price discovery for stocks traded in a single market. In their model, only the new information which implies a profit greater than the transaction cost, measured by bid-ask spread, will be translated into the transaction price. In other words, the shocks that drive the efficient price component must be "big" shocks to the transaction price. The transactions of the uninformed agents cannot generate big inefficient changes in the transaction prices, because the informed traders will arbitrate the situation. Therefore, the shocks driving the pricing error component by uninformed traders must be "small" shocks to the transaction price.

threshold ECM and smooth transition ECM. Each corresponds to an assumption on the demand elasticity of arbitrageurs. I apply these three models to examine the information shares of each market for Canadian stocks cross-listed in TSX and NYSE. All three models generate a conclusion that the home market (TSX) makes a larger contribution than NYSE (guest market) in the price discovery. However, from the estimations of nonlinear error correction models, I get some additional interesting findings. First, there is a larger feedback effect from NYSE on Canadian cross-listed stocks if the price deviations exceed the threshold value. This may be because the arbitrage activities could transfer information from the home market to the guest market. Second, when there exists a negative price premium at NYSE, informed traders tend to trade at NYSE even though the home market usually has better liquidity. As a result, convergence between the two market prices speeds up. Third, my regression analysis shows that information shares are positively affected by the relative degree of private information and market liquidity. The results are consistent with the empirical finding by Eun and Sabherwal (2003) and Chen and Choi (2010).

CHAPTER 2

COINTEGRATING REGRESSION WITH THRESHOLD EFFECTS

2.1 The Basic Model and Assumptions

Consider the following threshold model

$$y_t = \begin{cases} \alpha_1' x_t + e_t, & \text{if} \quad q_t \leq \gamma_0 \\ \alpha_2' x_t + e_t, & \text{otherwise} \end{cases}, \tag{2.1}$$

where y_t and q_t are scalar and x_t is a d_1-dimensional vector of $I(1)$ random variables. In this chapter, I follow Engle and Granger's (1987) single-equation approach with the main aim of investigating the effect of x_t on y_t. One can extend the model to study a cointegrating system where y_t is a vector. The threshold value $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$ is an unknown parameter to be estimated.

Model 2.1 can be regarded as a nonlinear cointegration, which attracts much attention from researchers recently, see Karlsen et al. (2007), Wang and Phillips (2009), Bierens and Martins (2010) and Choi and Saikkonen (2010). Loosely speaking, if the response variable y_t is generated by a nonlinear transformation of integrated regressors x_t and a stationary errors, then there exists a nonlinear cointegrating relationship between y_t and x_t. In such a nonlinear cointegrating relationship, y_t is not necessarily an $I(1)$ process. The stochastic property of y_t depends on the nonlinear transformation. Compared to linear cointegrations, which require y_t to be $I(1)$ process, nonlinear cointegrations have much more flexibility to choose response variables, see Shi and Phillips (2010). In my model, the persistence of the response variable depends on the threshold structure, such as the size of the threshold effect and the frequency of regime switching. [1]

In Model 2.1, the threshold effect leads to a cointegrating vector switching between $(1, -\alpha_1)$ and $(1, -\alpha_2)$. Unlike in Markov-switching models where an unobservable state governs regime switches, I assume regime shifts are induced endogenously by an observable threshold variable q_t, usually chosen according to some economic theories.[2] For example, when modelling the regime-sensitive Taylor rule (Taylor, 1993), q_t can represent lagged inflation rate or GDP growth rate; y_t is short-term interest rate or other measures of the monetary policy; x_t are fundamental economic variables such as inflation rate, growth rate and unemployment rate etc.. The inflation rate targeting theory implies that once inflation exceeds the preset inflation target, monetary policy authorities respond more aggressively to inflation. However, in states when inflation is below its threshold, they turn to output stabilization objects. This relationship can be described by a threshold model.

Model 2.1 can be re-written as

$$y_t = \alpha' x_t + \delta_n' x_t I(q_t \leq \gamma_0) + e_t, \tag{2.2}$$

where $\alpha = \alpha_2$ and $\delta_n = \alpha_1 - \alpha_2$. Here $I(q_t \leq \gamma_0)$ is an indicator function taking the value one if $q_t \leq \gamma_0$ and zero otherwise. The vector $(1, -\alpha)$ can be regarded as a benchmark long-run relationship between y_t and x_t while $\delta_n x_t I(q_t \leq \gamma_0)$ captures

[1]It is easy to see this point by taking difference on Model 2.1:

$$y_t - y_{t-1} = \left\{ \begin{array}{c} \alpha_1' \bigtriangleup x_t + \bigtriangleup e_t, \text{if} q_t \leq \gamma_0, q_{t-1} \leq \gamma_0 \\ \alpha_2' \bigtriangleup x_t + \bigtriangleup e_t, \text{if} q_t > \gamma_0, q_{t-1} > \gamma_0 \\ \alpha_1' \bigtriangleup x_t + \delta x_{t-1} + \bigtriangleup e_t, \text{if} q_t \leq \gamma_0, q_{t-1} > \gamma_0 \\ \alpha_2' \bigtriangleup x_t - \delta' x_{t-1} + \bigtriangleup e_t, \text{if} q_t > \gamma_0, q_{t-1} \leq \gamma_0 \end{array} \right\}$$

where $\delta = \alpha_1 - \alpha_2$.

[2]Both Markov-switching models and threshold models have been used to capture regime shifts in economic time series. However, the statistical inference is hard to implement and the regimes are intractable for the former approach.

deviations from the linear equilibrium relationship under special economic situations where the value of q_t is below γ_0.

The presence of this nonlinear term $\delta_n' x_t I(q_t \leq \gamma_0)$ affords some flexibility in modelling cointegrating relationships. For example, a theoretical and practical issue in finance is about the existence of asset return predictability from such fundamental variables as the dividend-price ratio, earning-price ratio. Linear prediction models have been extensively studied, but have failed to generate any unanimous conclusion (for more detail, refer to Campbell and Yogo, 2006). As is well known, stock returns commonly behave as martingale differences, while fundamental variables are highly persistent (integrated or nearly integrated processes). This discrepancy implies that any predictive relationship should be very weak or short-lived. A small positive threshold effect could be a natural candidate to capture such weak predictability.

Estimation and statistical inference for threshold models with integrated regressors require new asymptotic results, especially when the threshold effect is very small such that the model is only weakly identified. In order to establish these results, various assumptions must be put in place.

2.1.1 Assumptions and Some Preliminary Results

First, I assume the generating mechanism of x_t is integrated process of order one (I(1))

$$x_t = x_{t-1} + v_t, \quad t = 1, 2, ...n,$$

and set $x_0 = 0$ for convenience. Without materially affecting results, the generating mechanism for x_t can be replaced with a nearly integrated process (NI(1)), which has been used in empirical applications to model highly persistent economic and financial variables, see Campbell and Yogo (2006).

A partial sum process of v_t is defined as

$$X_n(s) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{[ns]} v_t$$

where$[ns]$ denotes the integer part of ns. I assume that the $X_n(s)$ satisfies the multivariate invariance principle; more specifically,

$$X_n(s) \Rightarrow X(s) \quad \text{as} \quad n \to \infty \tag{2.3}$$

where $X(\cdot)$ is a d_1-dimensional vector of Brownian motions on $[0,1]$. Furthermore, for any Borel measurable and totally Lebesgue integrable function $F(\cdot)$, I have

$$\frac{1}{n} \sum_{t=1}^{n} F(X_n(s)) \Rightarrow \int_0^1 F(X(s))ds \quad \text{as} \quad n \to \infty.$$

The multivariate invariance principle (or functional central limit theorem) applies for a very wide class of innovation sequences $\{v_t\}_{t=1}^{\infty}$ that are weakly dependent and possibly (conditionally) heterogeneously distributed (see Phillips and Durlauf (1986) and Billingsley (1999) for more discussions about the conditions of $\{v_t\}_{t=1}^{\infty}$). However, the invariance principle is not enough for deriving the convergence rate of the threshold estimator. It is necessary to have a stronger approximation for $X_n(s)$. Park and Hahn (1999) shows that, under some stronger conditions, equation 2.3 can be strengthened to the following approximation result:

$$\sup_{s \in [0,1]} \|X_n(s) - X(s)\| = o_p(1), \quad \text{almost surely.}$$

I follow Park and Hahn's (1999) assumption in the following.

Assumption 2.1.1 *Assume* $v_t = \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i} = \Phi(L)\varepsilon_t$ *where* $\Phi(1)$ *is non-singular and* $\sum i\Phi_i < \infty$. $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ *are i.i.d with* $E(\varepsilon_t \varepsilon_t') > 0$ *and* $E(|\varepsilon_t|^p) < \infty$ *for some* $p > 4$.

Under Assumption 2.1.1, v_t is a general linear process. The conditions on the summability of Φ_i and the moments of $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ are standard and comparable assumptions in the time series literature. Let $a = (p-2)/2p$, where p is the maximal order of the existing moment for ε_t. Note that a approaches 1/2 when $p \to \infty$.

Lemma 2.1.1 *Under Assumption 2.1.1,*

$$\sup_{s \in [0,1]} |X_n(s) - X(s)| = O_p(n^{-a}), \tag{2.4}$$

where $X(s)$ *is a vector of Brownian motions with* Σ_x *as the long run covariance matrix.*

Assumption 2.1.2 Σ_x *is a positive definite matrix.*

Under Assumption 2.1.2, the components of x_t are not cointegrated. This assumption is very typical for cointegration analysis.

Assumption 2.1.3 *Assume the following:*

(i) $E(e_t | F_{t-1}) = 0$ *and* $E(e_t^2 | F_{t-1}) = \sigma^2$, *where* σ^2 *is a positive constant.* F_{t-1} *is the past information set;*

(ii) $E(e_t | q_t, x_t) = 0$.

Under Assumption 2.1.3, e_t is a martingale difference sequence and orthogonal to x_t and q_t. This assumption is commonly used in nonlinear time series models with integrated processes but is too restricted (see Park and Phillips, 2001). Gonzalo and Pitarakis (2006) show that the assumption can be relaxed under certain circumstances for threshold models. For instance, e_t could be generalized to follow a linear moving average process of finite order l. However, a fully generalization of the model to allow for correlated errors would involve a substantial added complexity. For example, it invalids a weak convergence result involving quantities such as $\sum_{t=1}^{[ns]} I_t(q_t \leq \gamma)e_t$, established in Caner and Hansen (2001). Later, I relax this assumption by assuming the error term to be an AR(1) process and designing a Feasible GLS estimator to circumvent the complications.

Under Assumption 2.1.3, heterogeneity is excluded. In the literature, Hansen (1995) considers cointegrating regressions with error variance as a continuous function of a nearly nonstationary AR process. Kim and Park (2010) consider cointegration with time heterogeneity. The extension to allow for heterogeneity in Model (2.1) would be an interesting topic, and is left to the future study.

Assumption 2.1.4 *Assume the followings:*

(i) $\{q_t\}$ *is strictly stationary and* $\rho-$*mixing with* $\rho-$*mixing coefficients* ρ_m *satisfying* $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$. q_t *has a continuous distribution* $F(\cdot)$ *and* $f(\cdot)$ *is the corresponding density function.* $0 < f(\gamma) \leq \overline{f} < \infty$ *for all* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$.

(ii) $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$.

Assumption 2.1.4 is very typical for threshold models. For asymptotic purposes, I assume q_t is stationary. If q_t is nonstationary, such as an $I(1)$, one needs to use another methodology, for example the triangular array asymptotics proposed by Andrews and McDermott (1995).

To obtain some preliminary convergence results, I define the partial sum of the process $I_t(\gamma)e_t$ as

$$W_{[ns]}(\gamma) = \sum_{t=1}^{[ns]} I_t(\gamma)e_t$$

and scale $W_{[ns]}(\gamma)$ as

$$W_n(s,\gamma) = \frac{1}{\sigma\sqrt{n}} W_{[ns]}(\gamma) = \frac{1}{\sigma\sqrt{n}} \sum_{t=1}^{[ns]} I_t(\gamma)e_t$$

Following Caner and Hansen (2001), a two-parameter Brownian motion is defined as below.

Definition 1: *$W(s,\gamma)$ is a two-parameter Brownian motion on $(s,\gamma) \in [0,1]\times(-\infty,\infty)$ if $W(s,\gamma) \sim N(0, sF(\gamma))$ and $E(W(s_1,\gamma_1)W(s_2,\gamma_2)) = (s_1 \wedge s_2)(F(\gamma_1) \wedge F(\gamma_2))$.* The following lemma establishes the convergence results for $W_n(s,\gamma)$.

Lemma 2.1.2 *Under Assumptions 2.1.1-2.1.4, I have $W_n(s,\gamma) \Rightarrow W(s,\gamma)$ on $(s,\gamma) \in [0,1]\times(-\infty,\infty)$ as $n \to \infty$, where $W(s,\gamma)$ is a two-parameter Brownian motion.*

Note that $W_n(s,\infty) = \frac{1}{\sigma\sqrt{n}} \sum_{t=1}^{[ns]} I_t(\infty)e_t = \frac{1}{\sigma\sqrt{n}} \sum_{t=1}^{[ns]} e_t \Rightarrow W(s,\infty)$, which is a one-parameter Brownian motion. For simplicity of notation, I let $W(s) = W(s,\infty)$. The two-parameter Brownian motion is a special tool to derive the limiting distribution in threshold models with integrated processes. This differs from change-point models, where $\sum_{t=1}^{[ns]} I_t(q_t \le \gamma)e_t$ with $q_t = t$ is a martingale process and the limiting results are more easily established.

Using Definition 1, I can define the stochastic integration with respect to $W(s,\gamma)$ on the first argument while holding the second argument as constant as

$$\mathbf{J}_1(\gamma) = \int_0^1 X(s)dW(s,\gamma) = \lim_{n\to\infty} \sum_{t=1}^{n} (X(\frac{t-1}{n})\left(W(\frac{t}{n},\gamma) - W(\frac{t-1}{n},\gamma)\right).$$

Lemma 2.1.3 *Under Assumptions 2.1.1-2.1.4, for any* $\gamma \in [0,1]$,

$$\frac{1}{n}\sum_{t=1}^{n} X_t I_t(\gamma) e_t \Rightarrow \sigma \mathbf{J}_1(\gamma)$$

where $\mathbf{J}_1(\gamma) = \int_0^1 X(s)dW(s,\gamma)$ *is a Gaussian process with almost surely continuous sample path and the covariance kernel*

$$E(\mathbf{J}_1(\gamma_1)\mathbf{J}_1(\gamma_2)) = F(\gamma_1 \wedge \gamma_2)\int_0^1 X(s)X(s)'ds.$$

2.2 Profiled Least Square Estimator

For ease of manipulation, I rewrite Model (2.1) in a more compacted form:

$$y_t = \theta' V_t(\gamma_0) + e_t,$$

where $V_t(\gamma_0) = (x_t', x_t' I(q_t \le \gamma_0))'$ and $\theta = (\alpha', \delta_n')'$. For each $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, the following model is estimated:

$$y_t = \widehat{\theta}(\gamma)' V_t(\gamma) + \widehat{e}_t(\gamma),$$

where $\widehat{\theta}(\gamma)$ is given by

$$\widehat{\theta}(\gamma) = \left[\sum_{t=1}^{n} V_t(\gamma)V_t(\gamma)'\right]^{-1}\left[\sum_{t=1}^{n} V_t(\gamma) y_t\right].$$

The sum of residual square is defined as

$$SSR_n(\gamma) = \sum_{t=1}^{n} \widehat{e}_t(\gamma)^2 = \sum_{t=1}^{n}\left(y_t - \widehat{\theta}(\gamma)' V_t(\gamma)\right)^2$$

and I define the estimator of γ_0 as the value that minimizes $SSR_n(\gamma)$:

$$\widehat{\gamma}_n = \arg\min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR_n(\gamma).$$

Note that $SSR_n(\gamma)$ is not differentiable due to the presence of the indicator functions; thus, I can not write $\widehat{\gamma}_n$ in closed form from first-order conditions. Following Hansen (2000), I adopt a grid-searching method. Particularly, I divide $[\underline{\gamma}, \overline{\gamma}]$ into N quantiles and let $\Gamma_N = \{q_1, q_2, ...q_N\}$. $\widehat{\gamma}_N = \arg\min_{\gamma\in\Gamma_N} SSR_n(\gamma)$ is a good approximation to $\widehat{\gamma}_n$ when N is large enough. The estimations for other parameters are then found by plugging in the point estimate $\widehat{\gamma}_n$ via $\widehat{\theta} = \widehat{\theta}(\widehat{\gamma}_n)$, $\widehat{e}_t = \widehat{e}_t(\widehat{\gamma}_n)$, and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \widehat{e}_t(\widehat{\gamma}_n)^2$ denotes the residual variance from the LS estimation.

2.2.1 Asymptotic Properties

In this subsection, I establish the asymptotic results for the least square estimator $\widehat{\gamma}_n$ and $\widehat{\theta}(\widehat{\gamma}_n)$, under different model identification strengths. Based on asymptotic distributions, I construct confidence intervals. Following the literature of threshold models, I impose the following assumption.

Assumption 2.2.1 $\delta_n = n^{-1/2-\tau}\delta_0$ *where* $-1/2 \leq \tau \leq 1/2$ *and* $\delta_0 \in R$ *is a fixed parameter.*

Under Assumption 2.2.1, the size of the threshold effect converges to zero with rate $n^{-1/2-\tau}$. The value of τ determines the identification strength of γ_0. It can be shown that, if $\tau < 1/2$, γ_0 is identified and can be consistently estimated. However, if $\tau = 1/2$, γ_0 is only weakly identified and the least square estimator converges to a random variable even when the sample size n diverges to infinity. I exclude the case with $\tau > 1/2$ since the nonlinear term is negligible asymptotically. In addition, when $\tau < -1/2$, the nonlinear term is explosive and is also excluded.

Theorem 2.2.1 *Suppose Assumptions 2.1.1-2.2.1 hold and $\delta_0 \neq 0$. Then the following limiting results hold:*

Case 1: if $\tau < 1/2$, then

$$n^{1-2\tau}|\widehat{\gamma}_n - \gamma_0| = O_p(1).$$

Furthermore,

$$n^{1-2\tau}\lambda(\widehat{\gamma}_n - \gamma_0) = r^* \Rightarrow \arg\max_{r\in(-\infty,\infty)}(\Lambda(r) - \frac{1}{2}|r|)$$

where

$$\lambda = \frac{\left(\delta_0' \int_0^1 X(s)X(s)'ds\delta_0\right) f_0}{\sigma^2}, \tag{2.5}$$

and $\Lambda(r)$ is a two-sided Brownian motion on the real line defined as:

$$\Lambda(r) = \begin{cases} \Lambda_1(-r), & \text{if} \quad r < 0 \\ 0, & \text{if} \quad r = 0 \\ \Lambda_2(r), & \text{if} \quad r > 0 \end{cases}. \tag{2.6}$$

$\Lambda_1(r)$ and $\Lambda_2(r)$ are independent standard Brownian motions on $[0, \infty)$.

Case 2: if $\tau = 1/2$, then $\widehat{\gamma}_n \Rightarrow \gamma(\gamma_0, \delta_0)$. $\gamma(\gamma_0, \delta_0)$ is a random variable that maximize $Q(\gamma, \gamma_0, \delta_0)$ where

$$Q(\gamma, \gamma_0, \delta_0) = \frac{1}{F(\gamma)(1 - F(\gamma))}\Gamma_1(\gamma)\left(\int_0^1 X(s)X(s)'ds\right)^{-1}\Gamma_1(\gamma)' \tag{2.7}$$

with

$$\Gamma_1(\gamma) = \Gamma(\gamma) + (F(\gamma \wedge \gamma_0) - F(\gamma)F(\gamma_0))\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 \tag{2.8}$$

and

$$\Gamma(\gamma) = \sigma \int_0^1 X(s)d\left(W(s, \gamma) - F(\gamma)W(s)\right). \tag{2.9}$$

Theorem 2.2.1 shows that the convergence results for $\widehat{\gamma}_n$ depend critically on the value of τ, which characterizes the convergence speed of δ_n. If $\tau < 1/2$, the threshold effect is large enough to be identified and $\widehat{\gamma}_n$ is a consistent estimator. The rate of convergence is $n^{1-2\tau}$, which is decreasing in τ. The reason is that a larger τ decreases the threshold effect δ_n, which decreases the sample information concerning the threshold γ_0 and in turn reduces the precision of the estimator $\widehat{\gamma}$. In the regular case with τ as $-1/2$ such that δ_n is fixed as a constant, the convergence rate of $\widehat{\gamma}_n$ is n^{-2}. This super-consistency rate, resulting from the fast convergence rate of order statistics, makes the model powerful in detecting small regime shifts. In addition, the limiting distribution of $\widehat{\gamma}_n$ has the same form as that found for the stationary threshold model in Hansen (2000), although the scale factor is different. In the present context, the scale factor λ depends on $\int X(s)X(s)'ds$ instead of on a conditional moment matrix. f_0 is the density of q_t at γ_0. Intuitively, larger f_0 implies more data points around γ_0; therefore, $\widehat{\gamma}$ is more accurate.

The confidence intervals are commonly constructed through the inversion of test statistics. Following Hansen (2000), I invert the likelihood ratio statistic $LR_n(\gamma, \widehat{\gamma}_n, \widehat{\theta}_n)$ for the null hypothesis $\gamma = \gamma_0$. Denote $q^I_{\gamma,1-a}$ as the $1-a$ quantile of the limiting distribution of $LR_n(\gamma, \widehat{\gamma}_n, \widehat{\theta}_n)$. Under homoscedasticity assumption, $q^I_{\gamma,1-a}$ is the $1-a$ quantile of the random variable $\max_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$, which is given by the formula $q^I_{\gamma,1-a} = -2\ln(1 - \sqrt{1-a})$. The a–level confidence interval of γ can be expressed as

$$CI^I_{\gamma,n}(\alpha) = \{\gamma : LR_n(\gamma, \widehat{\gamma}_n, \widehat{\theta}_n) \leq q^I_{\gamma,1-a}\}. \tag{2.10}$$

If $\tau = 1/2$, the threshold effect is only weakly identified. The least square estimator $\widehat{\gamma}_n$ converges to a random variable $\gamma(\gamma_0, \delta_0)$, reflecting the lack of information. Since γ_0 and δ_0 are not estimable, any statistical inference based on them is impossible. Following Cheng (2008) and Shi and Phillips (2010), I define the least favorable confidence interval which is large enough for all possible γ_0 and δ_0. Denote $q^W_{\gamma,1-a}(\gamma_0,\delta_0)$ as the $1-a$ quantile of $|\gamma(\gamma_0,\delta_0)-\gamma_0|$ for each $\gamma_0 \in [\underline{\gamma},\overline{\gamma}]$ and $\delta_0 \in R$. The a-level confidence interval given γ_0 and δ_0 is defined as

$$CI^W_{\gamma,n}(1-a,\gamma_0,\delta_0) = \{\gamma : |\widehat{\gamma}_n - \gamma| \leq q^W_{\gamma,1-a}(\gamma_0,\delta_0)\}. \tag{2.11}$$

Since γ_0 and δ_0 are two unknown variables, I define a robust quantile by taking supremum for all possible γ_0 and δ_0. Let

$$q^W_{\gamma,1-a} = \sup_{\gamma_0\in[\underline{\gamma},\overline{\gamma}]} \sup_{\delta_0\in R} q^W_{\gamma,1-a}(\gamma_0,\delta_0). \tag{2.12}$$

The $a-$level least favorable confidence interval is then defined as

$$CI^W_{\gamma,n}(a) = \{\gamma : |\widehat{\gamma}_n - \gamma| \leq q^W_{\gamma,1-a}\}. \tag{2.13}$$

Next, I consider the limiting behavior of $\widehat{\theta}(\widehat{\gamma}_n)$. For any $\gamma \in [\underline{\gamma},\overline{\gamma}]$, define

$$M(\gamma) = \begin{pmatrix} \int_0^1 X(s)X'(s)ds, & F(\gamma)\int_0^1 X(s)X'(s)ds \\ F(\gamma)\int_0^1 X(s)X'(s)ds, & F(\gamma)\int_0^1 X(s)X'(s)ds \end{pmatrix}, \tag{2.14}$$

and

$$\Pi(\gamma,\gamma_0,\delta_0) = -\begin{pmatrix} (F(\gamma)-F(\gamma_0))\int_0^1 X(s)X'(s)ds \\ (F(\gamma)-F(\gamma_0\wedge\gamma))\int_0^1 X(s)X'(s)ds \end{pmatrix}\delta_0. \tag{2.15}$$

Theorem 2.2.2 *Under Assumptions 2.1.1-2.2.1, the following limiting results hold:*

Case 1: if $\tau < 1/2$, *then*

$$n(\widehat{\theta}(\widehat{\gamma}_n) - \theta) \Rightarrow \sigma M(\gamma_0)^{-1}\begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma_0) \end{pmatrix} = N(0,\sigma^2 M(\gamma_0)^{-1}).$$

Case 2: if $\tau = 1/2$, *then*

$$n(\widehat{\theta}(\widehat{\gamma}_n) - \theta) \Rightarrow \sigma M(\widehat{\gamma}_n)^{-1} \begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s, \widehat{\gamma}_n) \end{pmatrix} + M(\widehat{\gamma}_n)^{-1}\Pi(\widehat{\gamma}_n, \gamma_0, \delta_0) \equiv \Psi(\widehat{\gamma}_n, \gamma_0, \delta_0)$$

where

$$\widehat{\gamma}_n \Rightarrow \gamma(\gamma_0, \delta_0) = \arg\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} Q(\gamma, \gamma_0, \delta_0).$$

Theorem 2.2.2 establishes the limiting distribution for the coefficient estimators $\widehat{\theta}(\widehat{\gamma}_n)$. If $\tau < 1/2$, the limiting distribution of the coefficients estimators is mixed normal, which makes conventional t-test and chi-square tests applicable. If $\tau = 1/2$, the limiting result has a bias term $\Pi(\widehat{\gamma}_n, \gamma_0, \delta_0)$ due to the inconsistent estimation of $\widehat{\gamma}_n$. In that case, the standard test statistics are not applicable. Similarly, I discuss the construction of the confidence interval in two cases.

If $\tau < 1/2$, I construct the a-level confidence interval of θ by inverting the t test statistic. Specifically, define

$$CI^I_{\theta,n}(\alpha) = \{\theta : t(\theta, \widehat{\gamma}_n, \widehat{\theta}_n) \leq q^I_{\gamma,1-a}\} \tag{2.16}$$

where $t(\theta_0)$ is t-test statistic for testing $H_0 : \theta = \theta_0$ and $q^I_{\theta,1-a}$ is critical value at $1 - a$ significance level for t-statistic.

If $\tau = 1/2$, I use a similar approach as for $\widehat{\gamma}$ to define the least favorable CI. Denote $q^W_{\theta,1-a}(\gamma_0, \delta_0)$ as the $1 - a$ quantile of $\Psi(\widehat{\gamma}_n, \gamma_0, \delta_0)$ for each $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$ and $\delta_0 \in R$. Let

$$q^W_{\theta,1-a} = \sup_{\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]} \sup_{\delta_0 \in R} q^W_{\theta,1-a}(\gamma_0, \delta_0). \tag{2.17}$$

The least favorable confidence interval for θ is defined as

$$CI^W_{\theta,n}(a) = \{\theta : |n(\widehat{\theta}(\widehat{\gamma}_n) - \theta)| \leq q^W_{\theta,1-a}\}. \tag{2.18}$$

Both $q^W_{\gamma,1-a}(\gamma_0, \delta_0)$ and $q^W_{\theta,1-a}(\gamma_0, \delta_0)$ can be obtained through simulations.

2.2.2 Hypothesis Testing

In empirical studies, one may want to know whether the long-run relationship is linear or not. To answer this question, it is sufficient to test the following null hypothesis

$$H_0 : \delta_n = 0.$$

Under the null, the restricted model is $y_t = \alpha' x_t + e_t$, and under alternative, the unrestricted model is$y_t = \alpha' x_t + \delta_n' x_t I(q_t \leq \gamma_0) + e_t$.

For each $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, let $X(\gamma) = (x_1(\gamma), x_2(\gamma), ..., x_n(\gamma))'$ and $X = (x_1, x_2, ..., x_n)'$. Under homoscedasticity, a Wald-test statistic can be defined as

$$T_n(\gamma) = \widehat{\delta}_n(\gamma)'(X(\gamma)(I - P_n)X(\gamma))\widehat{\delta}_n(\gamma)/\widehat{\sigma}^2 \tag{2.19}$$

where P_n is the projection matrix of X, given by $P_n = X(X'X)^{-1}X'$. Notice that γ is a nuisance parameter which is not identified under the null. Following Hansen(1996), I define a sup-Wald test statistic as

$$T_n = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} T_n(\gamma). \tag{2.20}$$

Theorem 2.2.3 *Under Assumptions 2.1.1-2.1.4 and $\delta_n = 0$, then*

$$T_n \Rightarrow T = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} T(\gamma) = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \frac{1}{\sigma^2 (F(\gamma)(1 - F(\gamma))} \Gamma(\gamma)' \left(\int_0^1 X(s)X(s)'ds \right)^{-1} \Gamma(\gamma) \tag{2.21}$$

where

$$\Gamma(\gamma) = \sigma \int_0^1 X(s)d(W(s, \gamma) - F(\gamma)W(s)). \tag{2.22}$$

Theorem 2.2.3 establishes the limiting distribution of the sup-Wald statistics under the null hypothesis. As shown by Gonzalo and Pitarakis (2006), the limiting distribution T is equivalent to a random variable given by the supremum of a squared normalized Brownian bridge process, whose critical values appear in Andrews (1993). I explore the local power of the sup-Wald statistic in the theorem below.

Theorem 2.2.4 *Under Assumptions 2.1.1-2.2.1 and $\delta_n = n^{-1/2-\tau}\delta_0$, the following limiting results hold:*

i) if $\tau < 1/2$, then $T_n \xrightarrow{p} \infty$ and the power converges to 1.

ii) if $\tau = 1/2$, then

$$T_n \Rightarrow T_1 = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} T_1(\gamma) = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \frac{1}{\sigma^2 (F(\gamma)(1 - F(\gamma))} \Gamma_1(\gamma)' \left(\int_0^1 X(s)X(s)'ds \right)^{-1} \Gamma_1(\gamma) \tag{2.23}$$

where

$$\Gamma_1(\gamma) = \Gamma(\gamma) + (F(\gamma \wedge \gamma_0) - F(\gamma)F(\gamma_0)) \int_0^1 X(s)X(s)'ds\delta_0 \tag{2.24}$$

and the power $\in (0, 1)$.

iii) if $\tau > 1/2$, then $T_n \Rightarrow \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} T(\gamma)$, and the power equals the size.

2.2.3 Robust Confidence Intervals

In empirical studies, τ is unknown, raising the question of which confidence interval should be used. In this subsection, based on a model selection procedure, I construct a robust confidence interval which has approximately correct coverage probability irrespective of the value of τ.

From Theorem 2.2.4, $T_n \xrightarrow{p} \infty$ if $\tau < 1/2$ and $T_n < \infty$ if $\tau = 1/2$. This result enables us to develop the following model selection procedure. I define $\{\kappa_n : n \geq$

1} as a sequence of constants that diverge to infinity as $n \to \infty$. κ_n is referred to as a tuning parameter and I require κ_n satisfy the following assumption

$$\kappa_n^{-1/2} + n^{\nu}\kappa_n^{1/2} \to 0. \tag{2.25}$$

for any $\nu > 0$. Suitable choices of κ_n include $d_1 (\ln(n))^2$, in accordance with *BIC* criterion. The model selection procedure is designed to choose the model with identified threshold effect if $T_n > \kappa_n$ and to choose the model with weakly identified threshold effect otherwise. I use the confidence intervals from the model chosen through this procedure as the final confidence intervals.

For each confidence level a, define

$$CI_{\gamma,n}(a) = \left\{ \begin{array}{l} CI^{I}_{\gamma,n}(a), \text{ if } T_n > \kappa_n \\ CI^{W}_{\gamma,n}(a), \text{ if } T_n \leq \kappa_n \end{array} \right\} \tag{2.26}$$

and

$$CI_{\theta,n}(a) = \left\{ \begin{array}{l} CI^{I}_{\theta,n}(a), \text{ if } T_n > \kappa_n \\ CI^{W}_{\theta,n}(a), \text{ if } T_n \leq \kappa_n \end{array} \right\}. \tag{2.27}$$

I focus on the smallest finite sample coverage probability of $CI_{\gamma,n}(a)$ and $CI_{\theta,n}(a)$ over the whole parameter space, which can be approximated by the following asymptotic size

$$AsySZ_{\theta}(a) = \lim_{n\to\infty} \inf_{\theta\in R} \inf_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \Pr(\theta \in CI_{\theta,n}(a)) \tag{2.28}$$

and

$$AsySZ_{\gamma}(a) = \lim_{n\to\infty} \inf_{\theta\in R} \inf_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \Pr(\gamma \in CI_{\gamma,n}(a)) \tag{2.29}$$

The following theorem shows that the robust confidence intervals have the correct asymptotic size.

Theorem 2.2.5 *Under Assumptions 2.1.1-2.2.1, for any $a \in (0, 1)$, I have $AsySZ_{\theta}(a) = a$ and $AsySZ_{\gamma}(a) = a$.*

2.3 Extension

In many economic applications of cointegration, I may have serially correlated error terms and endogeneity. For linear cointegration models, it is well known that OLS estimator contains a second-order bias. Several efficient estimators have been proposed, such as the fully modified (FM) OLS estimator of Phillips and Hansen (1990), the canonical cointegrating regressions (CCR) estimator of Park (1992) and the dynamic ordinary least square (DOLS) estimator proposed by Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993). In the following, I extend my threshold cointegration model to allow for serial correlation and endogeneity.

Consider the following model

$$y_t = \alpha' x_t + \delta_n' x_t I(q_t \leq \gamma_0) + \xi_t, \tag{2.30}$$

where ξ_t is decomposed into a pure innovation component η_t and a component related to x_t:

$$\begin{aligned} \xi_t &= \sum_{i=1}^{d_1} \sum_{j=-K}^{K} \beta_{ij} v_{i,t-j} + \eta_t = \beta' z_t + \eta_t, \\ \eta_t &= \rho \eta_{t-1} + e_t, \text{ with } \rho \in (-1, 1]. \end{aligned}$$

I assume that the model endogeneity can be fully captured by $\beta' z_t$, where z_t is a $(2K+1)d_1$-dimensional vector of leads and lags of Δx_t. K can diverge to infinity as sample size increases. The idea of using leads and lags to deal with endogeneity in cointegration models was proposed by Saikkonen(1991). I assume β constant to focus on the regime shifts occurring in the cointegrating relationship. The extension allowing β to be regime-sensitive would be interesting and is left to future study. η_t is assumed to be an AR(1) process and ρ

controls the stationarity of η_t. If $\rho = 1$, η_t is a unit root and the model describes a spurious relationship[3], while if $\rho < 1$, η_t is a stationary process and Model (31) is a cointegrating relationship.

To estimate a regression with serial correlation, the Cochrane-Orcutt FGLS procedure is usually adopted. In linear cointegration models, as shown in Phillips and Park (1988), the FGLS estimator and the OLS estimator are equivalent in asympotics. The Cochrane-Orcutt FGLS estimator also works for spurious regressions, as Phillips and Hodgson (1994) demonstrate by proving asymptotic equivalence of the FGLS estimator to the OLS in the differenced regression when the error is an I(1) process. However, in the presence of regime shifts, there is no asymptotic equivalence between FGLS and OLS estimators. The following simple sketch may help to illustrate this difference.

For a linear cointegrating regression, after transformation, I have

$$y_t - \rho y_{t-1} = \alpha'(x_t - \rho x_{t-1}) + (\eta_t - \rho\eta_{t-1}),$$

and it follows that

$$\begin{aligned} n\left(\widehat{\alpha}_{FGLS} - \alpha\right) &= \left(\sum_{t=1}^{n}(x_t - \rho x_{t-1})(x_t - \rho x_{t-1})'\right)^{-1}\left(\sum_{t=1}^{n}(x_t - \rho x_{t-1})(\eta_t - \rho\eta_{t-1})\right) \\ &\Rightarrow \left((1-\rho)^2\int_0^1 X(s)X(s)'ds\right)^{-1}(1-\rho)^2\int_0^1 X(s)dB_\eta(s) \\ &= \left(\int_0^1 X(s)X(s)'ds\right)^{-1}\int_0^1 X(s)dB_\eta(s). \end{aligned}$$

which is the same as the limiting result of OLS estimator. However, for a cointegrating regression with a threshold effect, after transformation, I have,

$$\begin{aligned} y_t - \rho y_{t-1} &= \alpha'(x_t - \rho x_{t-1}) + \delta_n'(x_t(\gamma) - \rho x_{t-1}(\gamma)) + (\eta_t - \rho\eta_{t-1}) \\ &= \alpha'\widetilde{x}_t + \delta_n'\widetilde{x}_t(\gamma) + \widetilde{\eta}_t, \end{aligned}$$

[3]Structural spurious regressions can be due to integrated measurement errors and missing integrated regressors. See Choi et.al (2008).

where ρ can not be canceled in the limiting result because

$$\sum_{t=1}^{n} \widetilde{x}_t(\gamma)\widetilde{x}_t(\gamma)' \xrightarrow{p} \left((1+\rho^2)F(\gamma) - 2\rho F_1(\gamma,\gamma)\right) \int_0^1 X(s)X(s)'ds$$

and

$$\sum_{t=1}^{n} \widetilde{x}_t(\gamma)\widetilde{\eta}_t \Rightarrow \int_0^1 X_1(s)d\left(B_e(s,\gamma) - \rho B_{e,1}(s,\gamma)\right)$$

depend on ρ and the distribution function $F(\cdot)$ in a complex way.

All Assumptions 2.1.1-2.2.1 are applicable for the generalized model. Note that the model can be easily extended to incorporate a linear trend term as a regressor so that $x_{1t} = (1,\ t,\ x_t)$. In that event, a standardized matrix $D_n = diag\{1, n, n^{1/2}I_{d_1}\}$ should be defined to make each regressor converge at the same rate. Meanwhile, I use $X_1(s) = (1,\ s,\ X'(s))'$ to replace $X(s)$ for all assumptions and results. For notational simplicity, I skip the linear trend in the following discussion.

2.3.1 Feasible GLS Estimator

The procedure consists of two steps. In the first step, I estimate the threshold value γ_0 through the profiled least square estimation without considering serial correlation and endogeneity, and then I estimate $\widehat{\rho}$ from the estimated error term. In the second step, I construct the Cochrane-Orcutt type Feasible GLS estimator based on $\widehat{\rho}$. I can estimate γ_0 using the method described in Section 3.1. Specifically,

$$\widehat{\gamma}_n = \arg\min_{\gamma\in[\underline{\gamma},\overline{\gamma}]} SSR_n(\gamma).$$

where $SSR_n(\gamma)$ is the sum of squared residuals for the regression

$$y_t = \widehat{\alpha}_1' x_t + \widehat{\delta}_n' x_t I(q_t \leq \gamma) + \widehat{\xi}_t.$$

The residual term $\widehat{\xi}_t(\widehat{\gamma}_n)$ is defined as

$$\widehat{\xi}_t(\widehat{\gamma}_n) = y_t - \widehat{\alpha}' x_t - \widehat{\delta}_n' x_t I(q_t \leq \widehat{\gamma}_n).$$

By estimating an augmented AR(1) model

$$\widehat{\xi}_t(\widehat{\gamma}_n) = \widehat{\beta}' z_t + \widehat{\rho}\widehat{\xi}_{t-1}(\widehat{\gamma}_n) + \widehat{e}_t,$$

I obtain the OLS estimator $\widehat{\rho}$.

The following proposition establishes the consistency and convergence rate of $\widehat{\rho}$.

Proposition 2.3.1 *Under Assumptions 2.1.1-2.2.1, $\widehat{\rho} \to \rho$ as n increases to infinity. Further, $|\widehat{\rho} - \rho| = O_p(n^{-1/2})$ if $\rho < 1$ and $|\widehat{\rho} - 1| = O_p(n^{-1})$ if $\rho = 1$.*

Proposition 2.3.1 shows that $\widehat{\rho}$ is consistent even when the regression is a spurious relationship. $\widehat{\rho}$ has different convergence rates due to the different convergence speed of integrated and stationary processes. Moreover, the limiting behavior of $\widehat{\rho}$ is not affected by the identification strength of the threshold effect. The intuition is as follow. If $\tau < 1/2$, γ_0 can be consistently estimated. In each regime, if $\rho < 1$, the coefficients can be consistently estimated as well and thus it is obvious that $\widehat{\rho} \xrightarrow{p} \rho$, while if $\rho = 1$, the coefficient estimators are not consistent, however, this inconsistency causes the residual term $\widehat{\xi}_t(\widehat{\gamma}_n)$ to be unit root and I still have $\widehat{\rho} \xrightarrow{p} \rho = 1$. If $\tau = 1/2$, $\widehat{\gamma}_n$ is not consistent as shown in Theorem 4; however, the nonlinear term $\widehat{\delta}_{1n}' X_{1t} I(q_t \leq \gamma)$ decays to zero so fast that it has no impact on the estimation of ρ asymptotically. Following Choi et al. (2008), I can obtain the consistency of $\widehat{\rho}$ as well.

Bases on this consistent estimator $\widehat{\rho}$, I can conduct the Cochrane-Orcutt-type FGLS estimators. Define $\widetilde{y}_t = y_t - \widehat{\rho} y_{t-1}$, and define $\widetilde{z}_t, \widetilde{x}_t, \widetilde{\eta}_t$ in the same way. For each $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, define

$$\widetilde{x}_t(\gamma) = x_t(I(q_t \leq \gamma) - \widehat{\rho} x_{t-1} I(q_{t-1} \leq \gamma).$$

Let $\widetilde{V}_t(\gamma) = (\widetilde{x}_t', \widetilde{x}_t(\gamma)', \widetilde{z}_t')'$; I stack $\widetilde{x}_t$, $\widetilde{y}_t$, $\widetilde{z}_t$, $\widetilde{x}_t(\gamma)$ and $\widetilde{V}_t(\gamma)$ to get $\widetilde{X}$, $\widetilde{Y}$, $\widetilde{Z}$, $\widetilde{X}(\gamma)$, $\widetilde{V}(\gamma)$. After the transformation, I have

$$\widetilde{y}_t = \alpha' \widetilde{x}_t + \delta_n' \widetilde{x}_t(\gamma) + \beta' \widetilde{z}_t + \widetilde{\eta}_t = \widetilde{\theta}' \widetilde{V}_t(\gamma) + \widetilde{\eta}_t.$$

From the transformed regression, for each γ, I can define the OLS estimator

$$\widehat{\widetilde{\theta}}(\gamma) = \left[\sum_{t=2}^{n} \widetilde{V}_t(\gamma) \widetilde{V}_t(\gamma)' \right]^{-1} \left[\sum_{t=2}^{n} \widetilde{V}_t(\gamma) \widetilde{y}_t \right].$$

The FGLS threshold estimator is defined as

$$\widetilde{\gamma}_n = \arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} (S\widetilde{SR}_n(\gamma))$$

where $S\widetilde{SR}n(\gamma)$ is the sum of squared residuals defined as

$$S\widetilde{SR}n(\gamma) = \sum_{t=2}^{n} (\widetilde{y}_t(\gamma) - \widehat{\widetilde{\theta}}(\gamma)' \widetilde{V}_t(\gamma))^2.$$

By plugging in $\widetilde{\gamma}_n$, I obtain FGLS estimator $\widehat{\widetilde{\theta}}(\widetilde{\gamma}_n)$ for the coefficients $\widetilde{\theta}$. In practice, the above procedure can be conducted recursively until $\widetilde{\gamma}_n$ converges.

2.3.2 Asymptotics

Before I continue, I must define some notation. I first define the following joint distribution functions $F_1(\gamma) = \Pr(q_t \leq \gamma, q_{t-1} \leq \gamma)$; $F_2(\gamma) = \Pr(q_t \leq \gamma, q_{t-1} > \gamma)$; $F_3(\gamma) = \Pr(q_t > \gamma, q_{t-1} \leq \gamma)$; $F_4(\gamma) = \Pr(q_t > \gamma, q_{t-1} > \gamma)$; and moment functionals for the stationary regressors z_t

$$h(\gamma) \quad = \quad E(z_t I(q_t \leq \gamma)),$$

$$\begin{aligned}
h_1(\gamma) &= E(z_t I(q_{t-1} \le \gamma)), \\
h_2(\gamma) &= E(z_{t-1} I(q_t \le \gamma)), \\
H &= E(z_t z_t'), \\
H_1 &= E(z_t z_{t-1}'). \qquad (2.31)
\end{aligned}$$

Lemma 2.3.1 *For each $\rho \in (-1, 1]$, there exists a nonrandom weighting matrix $\widetilde{D}_n$ such that*

a) $n^{-1}\widetilde{D}_n^{-1}\widetilde{V}(\gamma)'\widetilde{V}(\gamma)\widetilde{D}_n^{-1} = \widetilde{G}(\gamma) + o_p(1)$;

b) $n^{-3/2}\widetilde{D}_n^{-1}\widetilde{V}(\gamma)'(\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))'\delta_0 = \widetilde{\Pi}(\gamma, \gamma_0, \delta_0) + o_p(1)$;

c) $n^{-1/2}\widetilde{D}_n^{-1}\widetilde{V}(\gamma)'\widetilde{\eta} \Rightarrow \widetilde{\phi}(\gamma)$,

where $\widetilde{G}(\gamma)$, $\widetilde{\Pi}(\gamma, \gamma_0, \delta_0)$ and $\widetilde{\phi}(\gamma)$ are expectation matrices specified in the appendix.

To conform with expression of $\widetilde{V}(\gamma) = (\widetilde{X}, \widetilde{X}(\gamma), \widetilde{Z})$, I express $\widetilde{G}(\gamma)$ and $\widetilde{\phi}(\gamma)$ as

$$\widetilde{G}(\gamma) = \begin{pmatrix} \widetilde{G}_{11}(\gamma), \widetilde{G}_{12}(\gamma), \widetilde{G}_{13}(\gamma) \\ \widetilde{G}_{21}(\gamma), \widetilde{G}_{22}(\gamma), \widetilde{G}_{23}(\gamma) \\ \widetilde{G}_{31}(\gamma), \widetilde{G}_{32}(\gamma), \widetilde{G}_{33}(\gamma) \end{pmatrix} \qquad (2.32)$$

and

$$\widetilde{\phi}(\gamma) = \begin{pmatrix} \widetilde{\phi}_1(\gamma) \\ \widetilde{\phi}_2(\gamma) \\ \widetilde{\phi}_3(\gamma) \end{pmatrix}. \qquad (2.33)$$

Theorem 2.3.1 *Under Assumptions 2.1.1-2.2.1, the following results hold:*

Case 1: if $\tau < 1/2$, then $n^{1-2\tau}|\widetilde{\gamma}_n - \gamma_0| = O_p(1)$. Furthermore,

$$n^{1-2\tau}\widetilde{\lambda}(\widetilde{\gamma}_n - \gamma_0) = r^* \Rightarrow \arg\max_{r\in(-\infty,\infty)}(\Lambda(r) - \frac{1}{2}|r|)$$

where

$$\widetilde{\lambda} = \frac{\left(1+\rho^2\right)\left(\delta_0' \int_0^1 X(s)X'(s)ds\delta_0\right) f_0}{\sigma^2},$$

and $\Lambda(r)$ *is a two-sided Brownian motion on the real line defined as:*

$$\Lambda(r) = \begin{cases} \Lambda_1(-r), & \text{if} \quad r < 0 \\ 0, & \text{if} \quad r = 0 \\ \Lambda_2(r), & \text{if} \quad r > 0 \end{cases}, \tag{2.34}$$

where $\Lambda_1(r)$ *and* $\Lambda_2(r)$ *are independent standard Brownian motions on* $[0, \infty)$.

Case 2: if $\tau = 1/2$, *then* $\widetilde{\gamma}_n \Rightarrow \widetilde{\gamma}(\gamma_0, \delta_0)$. $\widetilde{\gamma}(\gamma_0, \delta_0)$ *is a random variable that maximizes* $\widetilde{Q}(\gamma, \gamma_0, \delta_0)$ *where*

$$\widetilde{Q}(\gamma, \gamma_0, \delta_0) = \widetilde{\Gamma}_1(\gamma)\left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}\right)^{-1} \widetilde{\Gamma}_1(\gamma)'$$

with

$$\widetilde{\Gamma}_1(\gamma) = \widetilde{\Gamma}(\gamma) + \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}\right)\delta_0,$$

and

$$\widetilde{\Gamma}(\gamma) = \widetilde{\phi}_2(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{\phi}_1(\gamma) \\ \widetilde{\phi}_3(\gamma) \end{pmatrix}.$$

Theorem 2.3.1 establishes the convergence results for the FGLS estimator $\widetilde{\gamma}_n$. If $\tau < 1/2$, I can consistently estimate γ_0, and the limiting distribution depends on the persistence parameter ρ. Note that when $\rho = 0$, I get the same limiting distribution defined in Theorem 2.2.2. There is no asymptotic equivalence between the FGLS and LS estimator. $\widetilde{\lambda} = \left(1+\rho^2\right)\lambda$, thus, the FGLS estimator $\widetilde{\gamma}_n$

is more accurate than $\widehat{\gamma}_n$ when $\rho \neq 0$. Simulations also demonstrate that the FGLS estimator performs better than the LS estimator in the presence of serial correlation.

The following theorems establishes the convergence results for $\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n)$.

Theorem 2.3.2 *Under Assumptions 2.1.1-2.2.1 and* $\rho < 1$, *the following limits hold:*

Case 1: if $\tau < 1/2$, *I have*

$$\sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n) - \widetilde{\theta}) \Rightarrow \widetilde{G}(\gamma_0)^{-1}\widetilde{\phi}(\gamma_0) = N(0, \sigma^2\widetilde{G}(\gamma_0)^{-1}).$$

Case 2: if $\tau = 1/2$, *I have*

$$n^{1/2}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n) - \widetilde{\theta}) \Rightarrow \widetilde{G}(\widetilde{\gamma}_n)^{-1}\widetilde{\phi}(\widetilde{\gamma}_n) + \widetilde{G}(\widetilde{\gamma}_n)^{-1}\widetilde{\Pi}(\widetilde{\gamma}_n, \gamma_0, \delta_0).$$

where $\widetilde{\gamma}_n \Rightarrow \widetilde{\gamma}(\gamma_0, \delta_0)$ *and* $\widetilde{\gamma}(\gamma_0, \delta_0)$ *is a random variable that maximizes* $\widetilde{Q}(\gamma, \gamma_0, \delta_0)$.

2.4 Joint Hypothesis Test

Testing the existence of regime shifts in cointegration regression is challenging since it is a joint hypothesis problem (see Balke and Fomby, 1997). Most previous test statistics assume the remaining hypothesis to be true when they test for either regime shifts or cointegration. For example, when testing for the existence of regime shifts, the statistics based on error correction models (ECM) assume the model is a cointegrating regression. Therefore, the rejection of the null hypothesis does not necessarily indicate that there is a regime shift. It may mean the regression is a spurious relationship.

The FGLS estimator is robust under both I(1) and I(0) error terms; thus, I can test the existence of regime shifts using a sup-Wald test statistic based on the FGLS estimators without any knowledge about the presence of a cointegrating relationship. Then I apply residual-based test statistics to test cointegration given the conclusion from the first step.

2.4.1 Testing the Regime Shift

Testing the existence of the regime shift, it is sufficient to test the following null hypothesis

$$H_0 : \delta_n = 0.$$

Under the null, after transformation, the model is

$$\widetilde{y}_t = \alpha' \widetilde{x}_t + \beta' \widetilde{z}_t + \widetilde{\eta}_t.$$

The alternative model can be written as

$$\widetilde{y}_t = \alpha' \widetilde{x}_t + \delta_n' \widetilde{x}_t(\gamma_0) + \beta' \widetilde{z}_t + \widetilde{\eta}_t,$$

or in a compact form

$$\widetilde{y}_t = \widetilde{\theta}' \widetilde{V}_t(\gamma_0) + \widetilde{\eta}_t,$$

For each γ, let $\widetilde{V}_1 = (\widetilde{X}, \widetilde{Z})$. Under Assumption 2.1.3, a standard Wald statistic could be given by

$$\widetilde{T}_n(\gamma) = \widetilde{\delta}_n(\gamma)' (\widetilde{X}(\gamma)(I - \widetilde{P}(\gamma))\widetilde{X}(\gamma))\widetilde{\delta}_n(\gamma)/\widetilde{\sigma}^2$$

where $\widetilde{P}(\gamma)$ is the projection matrix for $\widetilde{V}_1$ and $\widetilde{\sigma}^2 = \frac{S\widetilde{S}Rn(\gamma)}{n}$. Define

$$\widetilde{T}_n = \sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \widetilde{T}_n(\gamma).$$

Theorem 2.4.1 *Under Assumptions 2.1.1-2.2.1 and $H_0 : \delta_n = 0$, the following limiting results hold:*

$$\widetilde{T}_n \Rightarrow \widetilde{T} = \sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \frac{1}{\sigma^2}\widetilde{\Gamma}(\gamma)\left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}\right)^{-1} \widetilde{\Gamma}(\gamma)'$$

where

$$\widetilde{\Gamma}(\gamma) = \widetilde{\phi}_2(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{\phi}_1(\gamma) \\ \widetilde{\phi}_3(\gamma) \end{pmatrix}. \tag{2.35}$$

The limiting distribution of the sup-Wald test statistic is a non-standard distribution and I can generate the critical values using a parametric bootstrap method. I first estimate $\widehat{\widetilde{\theta}}_R$ under the restriction that $\delta_n = 0$. Then, I obtain the residual terms $\{\widehat{\widetilde{\eta}}_t(\widetilde{\gamma}_n)\}_{t=2}^n$ using the unrestricted model. I draw a random variable $\widetilde{\eta}_t^b$ from the sample $\{\widehat{\widetilde{\eta}}_t(\widetilde{\gamma}_n)\}_{t=2}^n$ for all $t = 2, ..., n$, and generate a new sequence $\{\widetilde{y}_t^b\}_{t=1}^n$ by $\widetilde{y}_t^b = \widehat{\alpha}_R'\widetilde{x}_t + \widehat{\beta}_R'\widetilde{z}_t + \widetilde{\eta}_t^b$. Define $y_1^b = y_1$ and $y_t^b = \widetilde{y}_t^b + \widehat{\rho} y_{t-1}^b$ for all $t = 2, ..., n$. Let $\widetilde{T}_n^b$ be the sup-Wald test calculated from the new data set $\{y_t^b, x_t, z_t, q_t\}_{t=2}^n$. Under the null , the distribution of $\widetilde{T}_n^b$ can approximate the distribution of $\widetilde{T}_n$. The bootstrap p-value can be obtained by calculating the frequency of simulated $\widetilde{T}_n^b$ that exceeds $\widetilde{T}_n$ when the number of the simulations is large enough. As shown in Hansen(1996), the generated p-value converges to the true size. A model selection procedure can be constructed based on $\widetilde{T}_n$ and robust confidence intervals can be designed.

2.4.2 Testing against the spurious relationship

In this subsection, I develop test statistics to test a cointegration regression against the alternative of a spurious relationship. In the literature, two approaches have been used to test for cointegration. One takes cointegration as the null hypothesis and noncointegration as the alternative, whereas the other approach reverses the roles of the null and alternative hypotheses. In the nonlinear context, the former approach appears more convenient and thus be adopted here. I develop different cointegration test statistics, based on Kwiatkowski, Phillips, Schmidt, and Shin's (1992; KPSS hereafter) test statistic, for the case with identified or weakly identified threshold nonlinearity respectively. A robust KPSS test statistic is then proposed based on the model selection procedure for practical applications.

The null hypothesis is

$$H_0 : \eta_t \text{ is stationary for some } (\widetilde{\theta}, \gamma_0) \in \Theta;$$

the alternative hypothesis is

$$H_1 : \eta_t \text{ is an I(1) for any } (\widetilde{\theta}, \gamma_0) \in \Theta.$$

For each $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, define the residuals as

$$\widehat{\eta}_t(\gamma) = y_t - \widetilde{\alpha}' x_t - \widetilde{\delta}_n' x_t(\gamma) - \widetilde{\beta}' z_t,$$

where $(\widetilde{\alpha}', \widetilde{\delta}_n', \widetilde{\beta}')$ is the FGLS estimator of the coefficients $\widetilde{\theta} = (\alpha, \delta_n, \beta)$. Define a KPSS-type test statistics as

$$KPSS(\gamma) = n^{-2} \sum_{t=1}^{n} S_t^2 / s^2(L),$$

where $S_t = \sum_{i=1}^{t} \widehat{\eta}_t(\gamma)$ and $s^2(L)$ is a Newey-West estimator of the long-run variance of $\widehat{\eta}_t(\gamma)$.

$$s^2(L) = n^{-1} \sum_{t=1}^{n} \widehat{\eta}_t(\gamma)^2 + 2n^{-1} \sum_{j=1}^{L-1} k(j/L) \sum_{t=j+1}^{n} \widehat{\eta}_t(\gamma) \widehat{\eta}_{t-j}(\gamma),$$

where $k(j/L)$ is a Bartlett kernel function $k(\frac{j}{L}) = 1 - |\frac{j}{L}|$ and L depends on the sample size n. For example, in R software, the truncation lag parameter L is set to be $\frac{3}{13}n^{1/2}$ or $\frac{10}{14}n^{1/2}$.

When the threshold effect is only weakly identified (or completely non-identified), I define a cointegration test statistics as

$$KPSS_1 = \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} KPSS(\gamma). \tag{2.36}$$

Otherwise, I define

$$KPSS_2 = KPSS(\widetilde{\gamma}_n) \tag{2.37}$$

where $\widetilde{\gamma}_n$ is the estimated threshold value. Based on the model selection procedure described in Section 3.3, I can define a robust KPSS test statistic as

$$KPSS^* = \left\{ \begin{array}{c} KPSS_1, \text{if } \widetilde{T}_n \leq d_1(\ln n)^2 \\ KPSS_2, \text{otherwise} \end{array} \right\}. \tag{2.38}$$

The basic idea is that when the threshold effect is large enough to be identified, $\widetilde{\gamma}_n$ is consistent and I just need to check the stationarity of $\widehat{\eta}_t(\widetilde{\gamma}_n)$; otherwise, $KPSS_1$ is used to consider all possible threshold values. It can be shown that, under the null hypothesis, there exist a γ such that $\widehat{\eta}_t(\gamma)$ is stationary and $KPSS_1 \Rightarrow \int_0^1 V^2(s)ds$ where V is a standard Brownian bridge: $V(s) = W(s) - sW(1)$ with $W(s)$ is a standard Brownian motion. The critical value is available from Table 1 in Kwiatkowski et al., (1992), calculated via a direct simulation. Under alternative, $\widehat{\eta}_t(\gamma)$ is an unit root and $KPSS_1$ will diverge at the rate n^2. Therefore, $KPSS_1$ is a consistent test. As for $KPSS_2$, under the null hypothesis and identification conditions, $\widetilde{\gamma}_n$ converges to γ_0 and I have $KPSS_2 \Rightarrow \int_0^1 V^2(s)ds$ as well. Thus, $KPSS_1$ and $KPSS_2$ are identical asymptotically and I can use the same critical value table.

2.5 Simulations

In this section, I will demonstrate the finite sample performance of the estimators and test statistics.

Simulation 1: Through this experiment, I examine the consistency of the profiled least square(LS) estimators and feasible GLS estimators. I also make a comparison between these two estimators. I consider a simple univariate model:

$$y_t = x_t + \delta_n x_t(q_t > 0) + \eta_t,$$

where $x_t = x_{t-1} + e_{1t}$ and the error term $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. The threshold variable q_t is generated by an $AR(1)$ process $q_t = 0.5q_{t-1} + e_{2t}$; e_{1t}, e_{2t} and e_t are $i.i.d.N(0,1)$ and independent of each other. The number of replications is $N = 1000$. I choose $K = 5$ in FGLS estimator to deal with model endogeneity.[4] I consider two choices for δ_n: (i) $\delta_n = 2n^{-0.5}$ and (ii) $\delta_n = 2n^{-1}$, corresponding to the cases with an identified or weakly identified threshold effect respectively.

Table 2.1 reports the mean square error (MSE) of LS and FGLS estimators. From Table 2.1 if the model is identified, I observe that both least square estimators and FGLS estimators are consistent. The endogeneity and serial correlation of error terms do not affect the consistency of the threshold estimators. Furthermore, the FGLS estimator has smaller MSE than the LS estimator, especially when there is a serious serial correlation in error terms. However, when the threshold effect is only weakly identified, both estimators are inconsistent.

Simulation 2: Through this experiment, I show the performance of the sup-Wald test statistics and the model selection procedure. The following data generating process is examined:$y_t = x_t + \delta_n x_t(q_t > 0) + \eta_t$. I use a similar data set up

[4]I try other numbers for K, such as 10 and 15, the results do not change much.

Table 2.1: The Mean Squared Error(MSE) of threshold estimators

			$n = 100$		$n = 200$		$n = 400$	
ρ	τ	ρ_1	MSE($\widehat{\gamma}$)	MSE($\widetilde{\gamma}$)	MSE($\widehat{\gamma}$)	MSE($\widetilde{\gamma}$)	MSE($\widehat{\gamma}$)	MSE($\widetilde{\gamma}$)
0	0	0	.012	.019	.004	.004	.0021	.0022
0	0	.5	.013	.009	.004	.002	.0008	.0003
0	.5	0	.365	.379	.485	.472	.454	.455
0	.5	.5	.519	.509	.518	.518	.51	.454
1	.0	0	.191	.002	.221	.0056	.203	.0002
1	0	.5	.314	.005	.197	.0019	.234	.0002
1	.5	0	1.17	.592	.654	.527	.717	.408
1	.5	.5	.435	.302	.460	.389	.544	.429
.95	0	0	.129	.019	.097	.001	.032	.0001
.95	0	.5	.125	.002	.129	.0031	.031	.0004
.95	.5	0	.586	.410	.572	.38	.522	.404
.95	.5	.5	.375	.359	.557	.44	.494	.367
−.95	0	0	.093	.004	.043	.0002	.027	.0002
−.95	0	.5	.171	.020	.095	.0031	.023	.0003
−.95	.5	0	.800	.608	.422	.364	.606	.412
−.95	.5	.5	.742	.530	.532	.420	.622	.356

Note: The model is $y_t = ax_t + \delta x_t(q_t > 0) + \eta_t$ with $x_t = x_{t-1} + e_{1t}$ and $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. $q_t = 0.5q_{t-1} + e_{2t}$. e_t, e_{1t} and e_{2t} are *i.i.d.* $N(0,1)$. ρ is chosen from $0, 1, 0.95$ and -0.95 to see the impact of serial correlations. $\delta = 2n^{-1/2-\tau}$, where τ is set as 0 or 0.5. ρ_1 is set as 0 or 0.5 to see the impact of the model endogeneity. MSE($\widehat{\gamma}$) is the mean square error(MSE) for the least square(LS) estimator of threshold value γ_0, while MSE($\widetilde{\gamma}$) is for FGLS estimators. n is the sample size. The replication number is 1000.

for x_t, q_t and η_t as Simulation 1. In addition, I consider $\delta_n = 0$ to evaluate the size performance of the test statistics.

Table 2.2 below reports the size performance for the sup-Wald statistics T_n and $\widetilde{T}_n$, which are based on LS estimator and FGLS estimator respectively.

Table 2.2: Size performance of sup-Wald statistics and model selection

			$n = 100$				$n = 200$				$n = 400$			
ρ	δ	ρ_1	10%	5%	1%	$> log^2(n)$	10%	5%	1%	$> log^2(n)$	10%	5%	1%	$> log^2(n)$
$\mathbf{T}_n$														
0	0	0	0.078	0.028	0.005	0	0.114	0.056	0.018	0	0.098	0.046	0.008	0
0	0	.5	0.120	0.092	0.034	0	0.084	0.05	0.008	0	0.076	0.032	0.01	0
1	0	0	0.132	0.082	0.03	0	0.064	0.034	0.014	0.002	0.068	0.044	0.02	0.002
1	0	.5	0.086	0.052	0.022	0	0.134	0.074	0.022	0.01	0.112	0.076	0.018	0
.95	0	0	0.128	0.073	0.042	0	0.178	0.072	0.03	0.04	0.104	0.046	0.008	0
.95	0	.5	0.134	0.075	0.02	0	0.146	0.064	0.004	0.002	0.132	0.068	0.02	0
−.95	0	0	0.12	0.072	0.01	0	0.106	0.06	0.018	0	0.092	0.042	0.016	0
−.95	0	.5	0.07	0.034	0.008	0	0.118	0.064	0.006	0	0.092	0.03	0.006	0
$\widetilde{\mathbf{T}}_n$														
0	0	0	0.11	0.062	0.012	0	0.102	0.072	0.03	0	0.094	0.044	0.01	0
0	0	.5	0.144	0.084	0.028	0	0.076	0.044	0.014	0	0.078	0.036	0.004	0
1	0	0	0.084	0.05	0.02	0	0.112	0.038	0.028	0	0.124	0.074	0.032	0
1	0	.5	0.13	0.062	0.014	0	0.082	0.042	0.034	0	0.116	0.054	0.024	0
.95	0	0	0.114	0.062	0.008	0	0.072	0.036	0.002	0.002	0.072	0.036	0.008	0
.95	0	.5	0.128	0.064	0.032	0	0.132	0.076	0.02	0	0.094	0.038	0.008	0
−.95	0	0	0.086	0.036	0.00	0	0.142	0.078	0.024	0	0.09	0.044	0.01	0
−.95	0	.5	0.116	0.066	0.022	0	0.094	0.046	0.016	0	0.09	0.058	0.018	0

From Table 2.2, I can find that the size performance is reasonably good under various kinds of situations. The model selection procedure also chooses the identification strengths correctly.

Table 2.3-2.5 report the power performance for the sup-Wald statistics T_n and $\widetilde{T}_n$ with different sample size.

Table 2.3: Power performance 1

$n = 100$			$\mathbf{T}_n$				$\widetilde{\mathbf{T}}_n$			
ρ	τ	ρ_1	10%	5%	1%	$> (log(n))^2$	10%	5%	1%	$> (log(n))^2$
0	0	0	1.00	0.994	0.984	.956	1.0	.998	.988	.97
0	0	.5	.992	0.98	0.968	.904	.998	.994	.984	.952
0	0.5	0	0.322	0.224	0.074	.024	.324	.206	.11	.034
0	0.5	.5	0.258	0.15	0.03	.02	.206	.09	.03	.032
1	0	0	0.478	.436	.36	.462	.998	.996	.982	.958
1	0	.5	0.506	.46	.312	.486	1.0	.998	.992	.97
1	0.5	0	.164	.098	.05	0.1	.278	.190	.078	0.044
1	0.5	.5	.09	.043	.01	0.03	.298	.240	.1	0.048
.95	0	0	.796	.72	.564	0.61	.986	.986	.968	0.968
.95	0	.5	.83	.772	.654	.0636	.996	.996	.994	.0978
.95	0.5	0	.07	.034	.01	0.018	.296	.206	.118	0.068
.95	0.5	.5	.322	.234	.072	0.024	.268	.196	.084	0.072
−.95	0	0	.956	.922	.884	0.464	1.0	.998	.994	0.994
−.95	0	.5	.856	.802	.708	0.444	1.0	1.0	.998	0.996
−.95	0.5	0	.296	.236	.146	0.002	.55	.434	.182	0.11
−.95	0.5	.5	.308	.198	.05	0.004	.5	.392	.266	0.088

Note: The model is $y_t = ax_t + \delta x_t(q_t > 0) + \eta_t$ with $x_t = x_{t-1} + e_{1t}$ and $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. $q_t = 0.5q_{t-1} + e_{2t}$. e_t, e_{1t} and e_{2t} are *i.i.d.* $N(0,1)$. ρ is chosen from $0, 1, 0.95$ and -0.95 to see the impact of serial correlations. ρ_1 is set as 0 or 0.5 to see the impact of the model endogeneity. n is the sample size. The replication number is 1000.

From Table 2.3-2.5, if threshold effect is identified, I find that both statistics are

Table 2.4: Power performance 2

$n = 200$			$\mathbf{T}_n$				$\widetilde{\mathbf{T}}_n$			
ρ	τ	ρ_1	10%	5%	1%	$> (log(n))^2$	10%	5%	1%	$> (log(n))^2$
0	0	0	1.0	1.0	.998	.982	1.0	1.0	.998	.984
0	0	.5	1.0	1.0	.996	.974	1.0	1.0	1.0	.99
0	0.5	0	.284	.222	.098	.004	.256	.198	.108	.004
0	0.5	.5	.254	.144	.078	.002	.250	.17	.066	.008
1	0	0	.854	.836	.816	.328	1.0	1.0	1.0	.992
1	0	.5	.668	.614	.54	.346	1.0	1.0	1.0	.988
1	0.5	0	.098	.05	.016	0	.280	.212	.126	0.08
1	0.5	.5	.116	.05	.01	.03	.292	.174	.072	0.014
.95	0	0	.856	.82	.732	.624	1.0	1.0	1.0	0.986
.95	0	.5	.79	.75	.708	.584	1.0	1.0	1.0	.992
.95	0.5	0	.082	.034	.016	0	.428	.336	0.194	0.02
.95	0.5	.5	.078	.044	.002	.034	.308	.228	.068	0.016
−.95	0	0	.962	.95	.912	.506	1.0	1.0	1.0	1.0
−.95	0	.5	.964	.956	.906	.502	1.0	1.0	1.0	1.0
−.95	0.5	0	.318	.22	.124	0	.442	.352	.268	0.034
−.95	0.5	.5	.110	.05	.016	0	.402	.292	.16	0.03

Note: The model is $y_t = ax_t + \delta x_t(q_t > 0) + \eta_t$ with $x_t = x_{t-1} + e_{1t}$ and $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. $q_t = 0.5q_{t-1} + e_{2t}$. e_t, e_{1t} and e_{2t} are *i.i.d.* $N(0,1)$. ρ is chosen from $0, 1, 0.95$ and -0.95 to see the impact of serial correlations. $\delta = 2n^{-1/2-\tau}$, where τ is set as 0 or 0.5. ρ_1 is set as 0 or 0.5 to see the impact of the model endogeneity. n is the sample size. The replication number is 1000.

consistent with power converging to one as sample size increases. In general, $\widetilde{T}_n$ performs better than T_n where there is serial correlation and endogeneity. On the contrary, if the model is only weakly identified, both statistics have low power, even when the sample size is large. This coincides with my theoretical results in Theorem 2.2.4. However, the model selection procedure successfully distinguishes the weak identification cases from the identification cases.

Table 2.5: Power performance 3

n = 400			$\mathbf{T}_n$				$\widetilde{\mathbf{T}}_n$			
ρ	τ	ρ_1	10%	5%	1%	$> (log(n))^2$	10%	5%	1%	$> (log(n))^2$
0	0	0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0	0	.5	1.0	1.0	1.0	.994	1.0	1.0	1.0	.998
0	0.5	0	.294	.196	.042	.002	.264	.166	.052	0
0	0.5	.5	.166	.124	.054	0	.272	.192	.108	.002
1	0	0	.742	.692	.586	.234	1.0	1.0	1.0	1.0
1	0	.5	.546	.488	.41	.266	1.0	1.0	1.0	.998
1	0.5	0	.386	.312	.16	0	.248	.162	.052	0
1	0.5	.5	.02	.012	.002	.012	.228	.146	.05	0.002
.95	0	0	.91	.884	.82	.722	1.0	1.0	1.0	1.0
.95	0	.5	.942	.916	.868	.668	1.0	1.0	1.0	.998
.95	0.5	0	.074	.026	.004	.002	.264	.186	0.068	0.002
.95	0.5	.5	.094	.052	.01	0	.292	.210	.082	0
−.95	0	0	.994	.986	.958	.596	1.0	1.0	1.0	1.0
−.95	0	.5	.996	.984	.958	.648	1.0	1.0	1.0	1.0
−.95	0.5	0	.092	.042	.02	0	.546	.448	.302	0.01
−.95	0.5	.5	.134	.064	.018	0	.320	.204	.134	0.01

Note: The model is $y_t = ax_t + \delta x_t(q_t > 0) + \eta_t$ with $x_t = x_{t-1} + e_{1t}$ and $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. $q_t = 0.5q_{t-1} + e_{2t}$. e_t, e_{1t} and e_{2t} are *i.i.d.* $N(0,1)$. ρ is chosen from $0, 1, 0.95$ and -0.95 to see the impact of serial correlations. $\delta = 2n^{-1/2-\tau}$, where τ is set as 0 or 0.5. ρ_1 is set as 0 or 0.5 to see the impact of the model endogeneity. n is the sample size. The replication number is 1000.

Simulation 3: Through this experiment, I want to show the test performance of $KPSS_1$ and $KPSS_2$. I consider the following simple model

$$y_t = a_1 x_t + \delta x_t(q_t > 0) + \eta_t$$

with $x_t = x_{t-1} + e_{1t}$ and $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. ρ is chosen among 0, 0.5 and 1. When $\rho < 1$ the model is a cointegrating regression, while $\rho = 1$ I get a spurious relationship. The threshold variable $q_t = 0.5q_{t-1} + e_{2t}$. e_t, e_{1t} and e_{2t} are *i.i.d.* $N(0,1)$ and independent of each other. δ is chosen among $4n^{-0.5}$, $4n^{-1}$ and 0 to reflect the impact of the threshold effect with different identification strengths. ρ_1 is chosen from 0 and 0.5. The replication number is 1000. All rejection frequencies are calculated at 5% significance level.

The results are reported in Table 2.6. It is apparent from Table 2.6 that both test statistics have size distortion, which is not surprising given the fact that the conventional KPSS test has size distortion. Generally, $KPSS_2$ has better performance than $KPSS_1$ when the threshold effect is large.[5]

2.6 Empirical Application

In this section, I provide an application to model the asymmetric effects of monetary policy on real output under different credit conditions.

There has been a long debate about how monetary policy affects real economic activity. Policy-makers usually believe that the central bank can manipulate aggregate demand by engineering expansions or contractions of the money supply. In the monetarist view, this story is correct since there is a direct link between money supply and aggregated spending. In the New Keynesian view, the story also holds since adjustments in asset prices brought about by a change in money supply lead to more spending. However, according to the neutral money hypothesis, a core belief of classical economists, a change in the money supply affects only nominal variables in the economy such as prices, wages and exchange rates, but has no effect on real (inflation-adjusted) variables, like employment, real GDP, and real consumption. If the neutral money hypothesis

[5] Choi and Saikkonen (2010) suggests that the subsamples of the regression residual instead of full-sample residual to construct KPSS test statistics may be helpful to enhance the performance of test in nonlinear cointegration models.

Table 2.6: The rejection frequency of KPSS_1 and KPSS_2

			$n = 100$		$n = 200$		$n = 400$	
ρ	δ	ρ_1	$KPSS_1$	$KPSS_2$	$KPSS_1$	$KPSS_2$	$KPSS_1$	$KPSS_2$
0	d_1	0	.01	.02	.01	.015	.01	.03
0	d_1	0.5	.01	.02	.01	.031	.008	.03
0	d_2	0	.02	.026	.02	.035	.036	.04
0	d_2	0.5	.016	.022	.01	.035	.022	.026
0	d_3	0	.012	.026	.03	.026	.028	.036
0	d_3	0.5	.01	.024	.03	.033	.03	.03
0.5	d_1	0	.01	.082	.03	.075	.03	.072
0.5	d_1	0.5	.036	.102	.03	.079	.028	.06
0.5	d_2	0	.048	.06	.07	.084	.06	.066
0.5	d_2	0.5	0.07	.094	.05	.072	.05	.066
0.5	d_3	0	0.078	.96	.078	.065	.052	.056
0.5	d_3	0.5	0.086	.100	.07	.071	.044	.048
1	d_1	0	.56	.700	.764	.79π	.88	.91
1	d_1	0.5	.65	.754	.802	.84	.938	.95
1	d_2	0	.558	.578	.844	.85	.916	.922
1	d_2	0.5	.65	.666	.77	.78	.938	.94
1	d_3	0	.73	.746	.844	.85	.926	.93
1	d_3	0.5	.67	.684	.846	.85	.916	.916

Note: The model is $y_t = ax_t + \delta x_t(q_t > 0) + \eta_t$ with $x_t = x_{t-1} + e_{1t}$ and $\eta_t = \rho\eta_{t-1} + \rho_1 e_{1t} + e_t$. $q_t = 0.5q_{t-1} + e_{2t}$. e_t, e_{1t} and e_{2t} are *i.i.d.* $N(0,1)$. ρ is chosen from $0, 1, 0.95$ and -0.95 to see the impact of serial correlations. ρ_1 is set as 0 or 0.5 to see the impact of the model endogeneity. δ is chosen from $d_1 = 4n^{-0.5}$, $d_2 = 4n^{-1}$ and $d_3 = 0$. The sample size is $n = 100, 200, 400$. In all cases rejection frequencies are at a nominal significance level of five percent and are calculated on the basis of 1000 replications.

is accurate, the central bank cannot affect the real economy by printing money since any increase in the supply of money would be offset by an equal rise in prices and wages.

Even among economists who believe that monetary policy does affect the output, there is still a debate about how monetary fluctuations transmit to real output. In an early paper, Blinder (1987) develops an explanation for how central bank policy affects real economic activity using a credit rationing mechanism. In his model, the economy has two equilibria: a Keynesian equilibrium and a credit-rationed equilibrium. He concludes that the effects of monetary policy, while qualitatively similar in the two regimes, may be rather weak in the Keynesian regime and rather strong in the credit-rationed regime. Translated into real-world terms, a tightening of monetary policy may have strong effects on the real sector when money is already tight, but weak effects when credit is initially plentiful. Azariadis and Smith (1998) develop a similar idea that it is possible for the economy to switch back and forth between a Walrasian regime and a credit-rationing regime. Both papers suggest nonlinear dynamics such as regime switching and asymmetric responses of the monetary shocks to the real economy. Empirical studies have examined this asymmetric relationship (see McCallum (1991), Galbraith (1996), Balke (2000)); however, all of these are restricted to the stationary framework due to the lack of theoretical work on the threshold model with integrated processes. Given that real output and monetary supply variables are unit roots, my model is expected to generate more reliable results.

The following reduced-form output equation is estimated:

$$y_t = \begin{Bmatrix} \mu_1 + b_1 t + \alpha_1 \widetilde{m}_t + \beta_1 g_t + \phi_1 z_t + e_t, \text{if} q_t \leq \gamma_0 \\ \mu_2 + b_2 t + \alpha_2 \widetilde{m}_t + \beta_2 g_t + \phi_2 z_t + e_t, \text{if} q_t > \gamma_0 \end{Bmatrix} \tag{2.39}$$

where y_t is the logarithm of real GDP; $\widetilde{m}_t = \frac{1}{3}\sum_{j=0}^{2} m_{t-j}$ where m_t is the logarithm of detrended M1. I consider a moving average of recent three quarters of m_t due to the lagged effects of monetary policy. g_t is the logarithm of detrended real government expenditure on goods and services; z_t are control variables. In this chapter, z_t are stock market index (sm_t) and unemployment rates (unp_t).

The threshold variable q_t is defined as follows: $q_t = \frac{M8_t - \mu}{\sigma}$ where $M8_t$ is an eight-quarter moving average of the detrended growth rate of the money supply $M1$. μ and σ are the mean and standard deviation of $M8_t$. The definition of q_t is similar to the variable $D1_t$ defined in McCallum (1991), such that $D1_t = 1$ indicates a credit-rationed period if recent monetary policy measured by $M8_t$ has been "one standard deviation tighter than average". I assume the threshold γ_0 to be unknown, allowing data to decide it. When q_t is below the threshold γ_0, I say the economy has a credit rationing.

The data sample period is from 1959-Q1 to 2009-Q2; with the loss of 8 sample points to lags, I are left with 198 observations for estimation. The data for seasonal adjusted real term GDP and government expenditure (2005 dollars) are available from the U.S. National Income and Product Accounts (NIPA). The money supply measure $M1$ is available in OECD data sets.[6] Other control variables such as stock market index and unemployment rate are available from OECD data sets as well.

I first check the persistence of each variable. I use least squares method to estimate the first-order coefficient in an AR(1) model; the estimated results are as follow: the 95% confidence interval of ρ is $[0.983, 0.987]$ for y_t, $[0.987, 0.989]$ for $\widetilde{m}_t$ and $[0.98, 0.983]$ for g_t. Thus it is very likely that all three variables are unit root processes. To confirm this conjecture further, I use the ADF test, PP test, and KPSS test statistics to test whether these variables are unit root processes.

[6]The fullname of OECD dataset is Organization for Economic Co-operation and Development, which is public available from http://www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html.

The results are summarized in the following Table 2.7, which shows that all variables are unit roots (notice that the null hypothesis for KPSS is stationary).

Table 2.7: The results for unit root tests

Test	ADF	Z_α	Z_t	$KPSS$
y_t	$-2.67(0.29)$	$-8.79(0.61)$	$-1.79(0.66)$	$4.98(< 0.01)$
$\widetilde{m}_t$	$-1.90(0.62)$	$-1.22(0.98)$	$-0.57(0.97)$	$5.02(< 0.01)$
g_t	$-3.2(0.076)$	$-9.05(0.59)$	$-2.43(0.39)$	$4.92(< 0.01)$
sm_t	$-2.03(0.56)$	$-7.83(0.66)$	$-1.96(0.58)$	$4.84(< 0.01)$
$unemp_t$	$-2.52(0.36)$	$-12.04(0.43)$	$-2.21(0.48)$	$0.51(0.04)$

Note: y_t is the logarithm of quarterly real GDP; $\widetilde{m}_t = \frac{1}{3}\sum_{j=0}^{2} m_{t-j}$ where m_t is the logarithm of quarterly detrended M1. g_t is the logarithm of quarterly detrended real government expenditure on goods and services; sm_t is quarterly stock market index and $unemp_t$ is quarterly unemployment rate. p-values are reported in brackets.

Next, I estimate a linear cointegration model, obtaining an estimated result

$$\widehat{y_t} = \underset{(0.11)}{8.15} + \underset{(0.0002)}{0.008}\, t - \underset{(0.011)}{0.0074}\, \widetilde{m}_t + \underset{(0.03)}{0.1}\, g_t - \underset{(0.055)}{0.106}\, sm_t - \underset{(0.0002)}{0.023}\, unp_t$$

The numbers below the coefficients are standard deviation calculated using Newey-West estimator. In the linear cointegration model, $\widetilde{m}_t$ is not significant while g_t is significant. The result supports the neutral money hypothesis, suggesting that monetary policy has no significant effect on real output. However, when I check the stationarity of the residual terms, neither ADF nor PP-test statistics can reject the null hypothesis that the residuals are unit roots. Thus, the evidence from the linear cointegration is not reliable.

To check whether there is any threshold effect, I estimate the following threshold model:

$$\widehat{y_t} = \left\{ \begin{array}{l} \underset{(0.33)}{8.08} + \underset{(0.0001)}{0.008}\, t + \underset{(0.012)}{0.124}\, \widetilde{m}_t + \underset{(0.04)}{0.59}\, g_t + \underset{(0.01)}{0.077}\, sm_t - \underset{(0.0047)}{0.025}\, unemp_t, \text{if} q_t \leq -0.34 \\ \underset{(0.045)}{8.15} + \underset{(0.0001)}{0.0075}\, t - \underset{(0.01)}{0.01}\, \widetilde{m}_t - \underset{(0.034)}{0.076}\, g_t - \underset{(0.076)}{0.048}\, sm_t - \underset{(0.0002)}{0.017}\, unemp_t,, \text{if} q_t > -0.34 \end{array} \right\}.$$

The estimated threshold value is −0.34, which is larger than −1 used in McCallum (1991). This result implies that credit rationing may be more likely to happen in the real economy than McCallum(1991)'s model.

In credit rationing periods, I find that the coefficient of $\widetilde{m}_t$ is 0.125 (0.012), larger than −0.01 (0.01) in normal periods. This result confirms the asymmetric effects of monetary policy shocks on real output. The shocks from monetary policy during "tight" credit regime have a larger effect on output than do shocks in the normal or "loose" regime. I also find that government expenditure has a larger effect in the "credit-rationed" regime. As for the stock market, the coefficient is 0.077 (0.01) in credit-rationed periods, larger than −0.048 (0.076) in normal regimes. The significantly positive relationship in tight credit regimes is consistent with the empirical finding that the relationship between stock market returns and fundamental factors is more significant in recessions.

The sequence of Wald statistics values for different thresholds is plotted in Figure 2.1. The sup-Wald statistics value is 426.8. The bootstrapped 95% critical value is 76.94. Thus, the sup-Wald statistics reject the null hypothesis of no threshold effect at 5% significance level. Since $\kappa_n = 167.15$, the model selection procedure chooses the identification case and t-statistics are applicable. To check whether the cointegration relationship is spurious or not, I use KPSS1 and KPSS2 statistics to test the cointegration. The KPSS1 value is 0.138 and the p-value is larger than 0.1. Thus I can not reject the null hypothesis of threshold cointegration regression. The KPSS2 is 0.17 and the p-value is larger than 0.1. Therefore KPSS2 agrees with the test result of KPSS1.

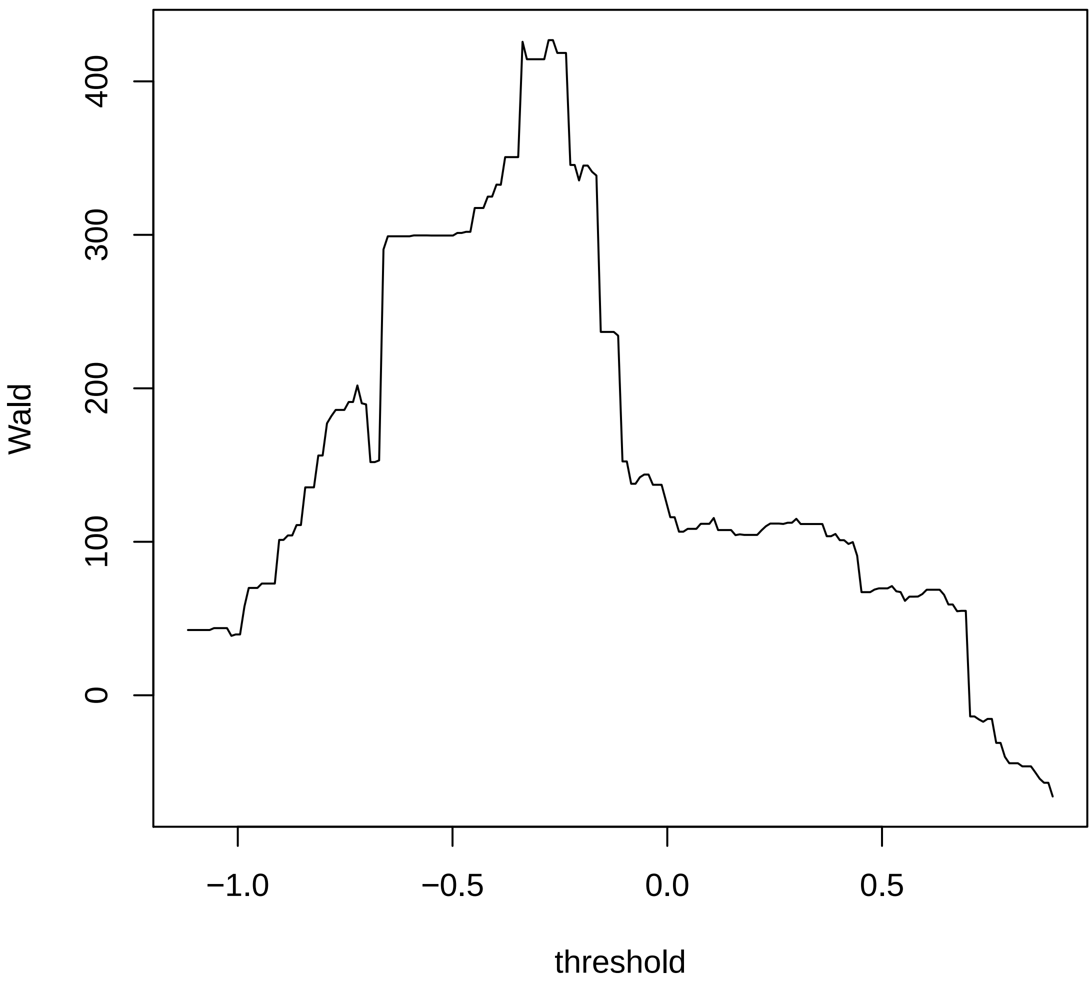

Figure 2.1: Wald statistics for different threshold values

2.7 Conclusions

This chapter can be viewed in two ways: (i) as an attempt to establish a statistical theory for threshold models with nonstationary regressors and (ii) as seeking to extend linear cointegrations by considering regime shifts in cointegrating vectors.

Threshold models have been popularly used to capture nonlinear effects in empirical macroeconomics and finance. In the literature, statistical theory for threshold models with stationary explanatory variables has been well developed. However, in empirical macroeconomics and finance, many explanatory variables are nonstationary, and there have been no econometric theory for nonstationary threshold models in the previous literature. I contribute to the literature by filling this gap. It is shown that the asymptotics depends on the size of the threshold effect. A model selection procedure is then applied to construct robust confidence intervals which have correct coverage no matter what the size of threshold effect is. I allow for model endogeneity and serial correlation, as these are common in regressions with nonstationary variables. A feasible generalized least square (FGLS) estimator is designed and shown to be a robust procedure to different error specifications, including I(1) errors.

This chapter can also be related to the literature of nonlinear cointegration and time-varying cointegration, which provide an important means to extend conventional cointegration analysis. The proposed model offers some flexibility of the cointegrating structure such that I can capture regime shifts in long-run relationships. I develop two test statistics based on KPSS test statistic to test the cointegrating relationship with a threshold effect, under different model identification conditions. The merits of the model and tests have been successfully demonstrated through simulations and an application to the asymmetric effect of monetary policy on real output.

There are several directions open for further work. First, it may be interesting to develop a more general model with multiple regime shifts, each with a different identification strength. A sequential procedure can be applied to determine the number of regimes and their identification strengths. Second, the model can be extended to allow for stationary regressors. For linear cointegrating regressions, Hansen (1995) shows that stationary regressors can improve the performance of cointegration tests. This may be true for nonlinear cointegration models as well. Finally, it would be interesting to deal with nonlinearity in both long-run relationships and short-term dynamics simultaneously under a unified framework.

CHAPTER 3

THRESHOLD EFFECTS IN REGRESSION MODELS WITH UNKNOWN FUNCTIONAL FORMS

3.1 The Model and Assumptions

Consider the following nonparametric model with a threshold effect

$$y_t = g_1(x_t)I(z_t \leq \gamma_0) + g_2(x_t)I(z_t > \gamma_0) + u_t. \tag{3.1}$$

where y_t is the observed dependent variable and $x_t(d \times 1)$ is a vector of explanatory variables, which may contain lagged values of y_t. u_t is the disturbance term. z_t is a random variable. The threshold effect $\gamma_0 \in \Gamma$ where $\Gamma = \left[\underline{\gamma}, \overline{\gamma}\right]$ is a closed set. $I(\cdot)$ is an indicator function[1]. This model can be treated as a semiparametric model since the effect of z_t to y_t is in parametric form, while the effect of x_t is nonparametric. In the current study, my main purpose is to estimate γ_0 and test the threshold effect. This method can be used to capture model heterogeneity across individuals or over time. For example, in my empirical application, I want to test whether the countries in my sample could be grouped according to the initial endowment, measured by per capital GDP. Previous studies assume a Cobb-Douglas production function and that may cause a model misspecification. Through Model (3.1), I can estimate the threshold effect and test it without any parametric specification on the production function.

Very commonly, one may have some information about the model structure based on economic theories for the question at hand. Thus, one may have the

[1]If z_t is t, similar to parametric models, the statistical inference of the structural change point estimator can only be derived for t/T since one needs an infinite amount of information around the change point as sample size increases to infinite.

following specific examples:

Case I: nonparametric regression with mean shift:

$$y_t = \begin{Bmatrix} g(x_t) + u_t, & z_t \leq \gamma_0 \\ g(x_t) + \mu + u_t, & z_t > \gamma_0 \end{Bmatrix} \tag{3.2}$$

Case II: single index models with changes on loadings:

$$y_t = \begin{Bmatrix} g(\beta_1 x_t) + u_t, & z_t \leq \gamma_0 \\ g(\beta_2 x_t) + u_t, & z_t > \gamma_0 \end{Bmatrix} \tag{3.3}$$

Case III: nonparametric models with changes on derivatives (sharp cusp).

$$y_t = \begin{Bmatrix} \int_a^{x_t} g_1(x)dx + u_t, & z_t \leq \gamma_0 \\ \int_a^{x_t} g_2(x)dx + u_t, & z_t > \gamma_0 \end{Bmatrix} \tag{3.4}$$

Case IV: partially linear model with a threshold effect on the linear component:[2]

$$y_t = \begin{Bmatrix} x_{1t}\beta_1 + m(x_{2t}) + u_t, & z_t \leq \gamma_0 \\ x_{1t}\beta_2 + m(x_{2t}) + u_t, & z_t > \gamma_0 \end{Bmatrix}. \tag{3.5}$$

Before going further, I impose some assumptions on the data generating process.

Assumption 3.1.1 *x_t is geometrically ergodic, stationary and $\beta - $mixing with exponential decay; u_t is a martingale difference sequence and $E(u_t|F_{t-1}, x_t, z_t) = 0$; and $\sup_t E|u_t|^{2+\kappa} < \infty$, for some $\kappa > 0$.*

Assumption 3.1.2 *z_t is strictly stationary and has a continuous distribution $F_z(\gamma)$. Let $f(\gamma)$ denote the density function satisfying $f(\gamma) \leq \overline{f} < \infty$ for all $\gamma \in \Gamma = \left[\underline{\gamma}, \overline{\gamma}\right]$ and $f(\gamma_0) > 0$.*

[2]This model is widely used when x_t is of high dimension to circumvent the "curse of dimensionality". See Fan and Li (1996) and Fan et al.(1998).

Assumption 3.1.3 *$Var(u_t|x_t, z_t) = \sigma^2(x_t, z_t)$ is a bounded and smooth function.*

Assumptions 3.1.1 are very standard in the literature of time series models and is trivially satisfied for independent cross-sectional observations. The data are either a random sample or a weakly dependent time series, so that unit roots and stochastic trends are excluded.[3] The martingale difference sequence assumption for the disturbance is necessary for the consistency of the estimator in nonlinear time series models. Assumption 3.1.2 requires z_t to be a continuous random variable and has a bounded density function. The density around γ_0 should be positive so there are observations around the threshold value. Assumption 3.1.3 requires the conditional variance to be bounded, an assumption which is difficult to relax without affecting the convergence rates, but it allows a wide range of conditional heteroscedasticity. In this chapter, I focus on the threshold effect in the conditional mean function. Thus, I assume the conditional function be smooth. The extension to allow for regime shifts in both conditional mean and conditional variance remains for a future study.

3.2 Main Results

In this section, I discuss the estimator and its asymptotic property.

3.2.1 Series Estimation with Known γ_0

As a starting point, I consider the case with known γ_0; thus I only need to estimate the function in each sub-sample split according to γ_0. To estimate $g_1(x)$ and

[3]Different asymptotic results are needed for nonstationary time series.

$g_2(x)$, I use a series estimator which approximates the unknown functions by a series expansion $\sum_{s=1}^{L} p_s^L(x)\beta_s^L$, where $\{p_s^L(x) : s = 1, 2, ...L\}$ is a prespecified family of functions from χ to R. χ is a compact set and $\chi \subset R^d$. Examples of such families include Fourier flexible forms (FFF), polynomials, and regression splines.[4] $\beta^L = (\beta_1^L, ..., \beta_L^L)'$ is an unknown parameter vector and L is the number of basis functions to be used in the approximation, depending on the sample size T.

Define $p^L(x) = (p_1^L(x),\ p_2^L(x), ..., p_L^L(x))'$. For any finite L, one can write the unknown function $g_i(x)$ as follows

$$g_i(x) = p^L(x)'\beta_i + e_t^L(x), \quad \text{for} \quad i = 1, 2.$$

where $e_t^L(x)$ is the remainder residual term. As L grows to infinity, the series expansion becomes a good approximation to $g_i(x)$. Following the literature of series expansion, I assume some regularity conditions on the function $g_i(x)$ and basis functions $\{p_s^L(x) : s = 1, 2, ..., L\}$.

Assumption 3.2.1 *x_t has compact support : $\chi \subset R^d$; the distribution function of x_t : $F_x(x)$ is absolutely continuous with respect to the Lebesgue measure.*

Assumption 3.2.2 *For each L, there is a non-singular matrix B such that for $P^L(x) = Bp^L(x)$: i) $Q_L = E[P^L(x_t)P^L(x_t)']$, which has a smallest eigenvalue bounded away from*

[4] A simple way to construct basis functions for multivariate x_t is to use the tensor product basis. For example, if $p = 2$, we can define $g_{j,k}(\cdot) = \phi_j(x_1)\psi_k(x_2)$ for $j, k \in \{0\} \cup Z^+$. The $g_{j,k}(\cdot)$ forms a basis $g(\cdot)$, called the tensor product basis. Thus, $m_s(\cdot)$ could be expanded in the tensor product basis as

$$\begin{aligned} m_s(x_1, x_2) &= \sum_{j,k=0}^{\infty} \beta_{j,k}\phi_j(x_1)\psi_k(x_2) \\ &= \beta_0 + \sum_{j=1}^{\infty} \beta_{j,0}\phi_j(x_1) + \sum_{k=1}^{\infty} \beta_{0,k}\psi_k(x_2) + \sum_{j,k=1}^{\infty} \beta_{j,k}\phi_j(x_1)\psi_k(x_2). \end{aligned}$$

The basis can be extended to d dimensions in the obvious way.

zero and a bounded largest eigenvalue; ii) there is a sequence of constants $\varsigma_0(L)$ *satisfying* $\sup_{x\in\varkappa} \max_{1\leq s\leq L} \|P_s^L(x)\| \leq \varsigma_0(L)$ *and* $\varsigma_0(L)^4 L^2/T \to 0$ *as* $T \to \infty$.

Assumption 3.2.3 *i) For each function* $g_i(x)$, $i = 1, 2$, *there exist a parameter vector* $\beta_i \in R^L$ *and constants* $\rho_i > 0$ *satisfying* $\sup_{x\in\varkappa} |g_i(x) - p^L(x)\beta_i| = O(L^{-\rho_i})$ *as* $L \to \infty$. *ii)* $\sqrt{T}L^{-\rho_i} \to 0$ *as* $T \to \infty$.

Assumption 3.2.1 is quite standard for series estimation. The support of x_t could be directly restricted to be $[0, 1]^d$ by re-scaling. The absolutely continuous distribution function rules out discrete random variables in x_t.

Assumption 3.2.2 imposes a normalization on the approximation functions, bounding the second moment matrix away from singularity, and restricting the magnitude of the series terms. For regression splines and orthonormal polynomials over a compact support, when the density of x_t is bounded away from zero, Newey(1988) and Andrews (1991) give the primitive conditions $\varsigma(L) = C$ or $C\sqrt{L}$ respectively. Thus, ii) can be expressed as $L^2/T \to 0$ or $L^4/T \to 0$, by substituting $\varsigma(L) = C$ or $C\sqrt{L}$ into $\varsigma_0(L)^4 L^2/T \to 0$.

Assumption 3.2.3 is conventional in the series approximation literature specifying a uniform rate for the approximation of the series approximation to the true function. As pointed out by Newey(1997), ρ_i is related to the smoothness of the true function and the dimensionality of x_t. For regression splines and power series, Newey (1997) shows that the assumption will be satisfied with $\rho = s/d$, where s is the number of continuous derivatives of $g(x)$ and d is the dimension of x_t. In practice, however, we do not know the true function and thus cannot determine ρ_i. ii) requires that L should not increase too slowly so that the approximation error can be ignored when we study the consistency and asymptotics of the estimators.

Remark 1: In Assumption 3.2.2, since the series estimator is invariant under non-singular linear transformations of $p^L(x)$, one can even assume that $B = I$, *i.e.*, $P^L(x) = p^L(x)$. Furthermore, since the smallest eigenvalue of $Q_L = E[P^L(x_t)P^L(x_t)']$ is bounded away from zero, it can be further assumed that $Q_L = I$ if the density of x_t is known. This is because, for any symmetric square root $Q_L^{-1/2}$ of Q_L, $Q_L^{-1/2}p^L(x)$ is a non-singular linear transformation of $p^L(x)$ satisfying all conditions in 3.2.2[5].

Let $Y = (y_1, y_2, ...y_T)', G = (p^L(x_1), p^L(x_2), ...p^L(x_T))'$,and $I_1(\gamma_0)$ be a $T \times T$ diagonal matrix with the $(t,t)^{th}$ element being an indicator function $1(z_t \leq \gamma_0)$. Let $I_2(\gamma_0) = I - I_1(\gamma_0)$. The series estimator for the function $g_i(x)$ can be defined as

$$\widehat{g}_{i,\gamma_0}(x) = p^L(x)'\widehat{\beta}_i(\gamma_0)$$

where $\widehat{\beta}_i(\gamma_0)$ is the ordinary least square estimator given by

$$\widehat{\beta}_i(\gamma_0) = (GI_i(\gamma_0)G)^{-}I_i(\gamma_0)G'Y, \quad \text{for} \quad i = 1, 2.$$

Note that $(\cdot)^-$ denotes the generalized inverse of $(\cdot)$.

Let $f(x|z_t \leq \gamma_0)$ and $f(x|z_t > \gamma_0)$ be the conditional density for x_t in each sub-sample. The following theorem gives a general result on mean-square and uniform convergence rates of the series estimation.

Theorem 3.2.1 *Under Assumptions 3.1.1-3.2.3, the following results hold:*

$$\int_{\mathcal{X}} [g_1(x) - \widehat{g}_{1,\gamma_0}(x)]^2 f_x(x|z_t \leq \gamma_0)dx = O_p(\frac{L_1}{T_{1,\gamma_0}} + L_1^{-2\rho_1}),$$

[5]For example, one can use Gram-Schmidt orthonormalization to get orthogonalized power series. Hermite, Jacobi, Laguerre and Legendre polynomials are examples of orthogonal polynomials. The orthonormal Legendre polynomial basis is defined as $P_n(x) = \sqrt{(2n+1)/2}\frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n$, for $n = 0, 1, ...$

$$\int_{\chi} [g_2(x) - \widehat{g}_{2,\gamma_0}(x)]^2 f_x(x|z_t > \gamma_0)dx = O_p(\frac{L_2}{T_{2,\gamma_0}} + L_2^{-2\rho_2}),$$

and

$$\sup_{x\in\chi} |g_1(x) - \widehat{g}_{1,\gamma_0}(x)| = O_p(\varsigma(L))[(\frac{L_1}{T_{1,\gamma_0}})^{1/2} + L_1^{-\rho_1}],$$

$$\sup_{x\in\chi} |g_2(x) - \widehat{g}_{2,\gamma_0}(x)| = O_p(\varsigma(L))[(\frac{L_2}{T_{2,\gamma_0}})^{1/2} + L_2^{-\rho_2}],$$

where T_{1,γ_0} and T_{2,γ_0} are the sample size in the subsample $\{z \le \gamma_0\}$ and $\{z > \gamma_0\}$ respectively.

It can be shown that the term $L_i/T_{i,\gamma_0}$ corresponds to the variance term and $L_i^{-2\rho_i}$ corresponds to a bias term. One can choose the number of L_i by minimizing the mean square errors, which requires that $L_i/T_{i,\gamma_0}$ and $L_i^{-2\rho_i}$ converge to zero at the same rate and solve for $L_i = O(T^{1/(1+2\rho_i)})$ in each sub-sample. L_i can be different in the two subsamples if the smoothness of the $g_i(x)$ differs. However, in practical applications, the exact value of ρ_i is unknown; thus, it is still impossible to choose L_i through the formula. In the literature, L_i is chosen optimally using data-driven methods, such as the generalized cross-validation method and jump of the estimated residual variance.

Define the following moment functionals

$$\begin{aligned} Q_{1,\gamma} &= E(p^L(x_t)p^L(x_t)'|z_t \le \gamma), \\ Q_{2,\gamma} &= E(p^L(x_t)p^L(x_t)'|z_t > \gamma), \end{aligned}$$

and

$$\begin{aligned} \Sigma_{1,\gamma} &= E(p^L(x_t)p^L(x_t)'u_t^2|z_t \le \gamma), \\ \Sigma_{2,\gamma} &= E(p^L(x_t)p^L(x_t)'u_t^2|z_t > \gamma). \end{aligned}$$

The following theorem establishes the asymptotic normality for the series estimators in each sub-sample.

Theorem 3.2.2 *Under Assumptions 3.1.1-3.2.3, the following results hold: for any fixed $L \times 1$ vector ω satisfying $\|\omega\| = 1$ for every L,*

$$\sqrt{T_{1,\gamma_0}}\omega'\Omega_{1,\gamma_0}^{-1/2}(\widehat{\beta}_1(\gamma_0) - \beta_1) \xrightarrow{d} N(0,1),$$

$$\sqrt{T_{2,\gamma_0}}\omega'\Omega_{2,\gamma_0}^{-1/2}(\widehat{\beta}_2(\gamma_0) - \beta_2) \xrightarrow{d} N(0,1).$$

For each fixed $x \in \chi$,

$$\sqrt{T_{1,\gamma_0}}W_{\gamma_0}^{-1/2}(\widehat{g}_{1,\gamma_0}(x) - g_1(x)) \xrightarrow{d} N(0,1).$$

$$\sqrt{T_{2,\gamma_0}}W_{1-\gamma_0}^{-1/2}(\widehat{g}_{2,\gamma_0}(x) - g_2(x)) \xrightarrow{d} N(0,1).$$

where $\Omega_{i,\gamma_0}^{-1/2} = Q_{i,\gamma_0}^{-1}\Sigma_{i,\gamma_0}Q_{i,\gamma_0}^{-1}$ and $W_{1,\gamma_0} = p^L(x)'Q_{i,\gamma_0}^{-1}\Sigma_{i,\gamma_0}Q_{i,\gamma_0}^{-1}p^L(x)$.

3.2.2 The Estimator of γ_0

In many practical applications, γ_0 is unknown. Thus, I need to estimate the threshold effect before estimating the nonparametric function in each subsample.

By substituting the series expansions of $g_1(x_t)$ and $g_2(x_t)$, one can rewrite Model (3.1) as

$$y_t = p^L(x_t)\beta_1 I(z_t \leq \gamma_0) + p^L(x_t)\beta_2 I(z_t > \gamma_0) + \epsilon_t \tag{3.6}$$

where $\epsilon_t = e_1(x_t)I(z_t \leq \gamma_0) + e_2(x_t)I(z_t > \gamma_0) + u_t$. $e_1(x_t)$ and $e_2(x_t)$ are remainder terms from the series expansions to $g_1(x_t)$ and $g_2(x_t)$ respectively. I can write the model in the following compact form

$$Y = I_1(\gamma_0)G\beta_1 + I_2(\gamma_0)G\beta_2 + \epsilon \tag{3.7}$$

where $Y = (y_1\ y_2\ ...\ y_T)'$ is a $T \times 1$ vector of y_T; $G = (p^L(x_1), ..., p^L(x_T))'$ is a $T \times L$ matrix with $(t, l)^{th}$ element $p_l^L(x_t)$; $I_1(\gamma_0)$ is a $T \times T$ diagonal matrix with the $(t, t)^{th}$ element being an indicator function $1(z_t \leq \gamma_0)$, and $I_2(\gamma_0) = I - I_1(\gamma_0)$.

For each fixed $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$, estimate the following model

$$Y = I_1(\gamma)G\widehat{\beta}_1(\gamma) + I_2(\gamma)G\widehat{\beta}_2(\gamma) + \widehat{\epsilon} \tag{3.8}$$

where $\widehat{\beta}_1(\gamma)$ and $\widehat{\beta}_2(\gamma)$ are $L \times 1$ vectors of OLS estimators defined by:

$$\widehat{\beta}_i(\gamma) = (GI_i(\gamma)G)^{-1}I_i(\gamma)G'Y, \quad \text{for} \quad i = 1, 2. \tag{3.9}$$

The threshold estimator then defined as

$$\widehat{\gamma}_T = \arg\min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR_T(\gamma),$$

where $SSR_T(\gamma)$ denotes the sum of residual squares

$$SSR_T(\gamma) = \left\| Y - I_1(\gamma)G\widehat{\beta}_1(\gamma) - I_2(\gamma)G\widehat{\beta}_2(\gamma) \right\|^2.$$

After obtaining $\widehat{\gamma}_T$, I define the series estimator $\widehat{g}_{1,\widehat{\gamma}_T}(x) = p^L(x_t)\widehat{\beta}_1(\widehat{\gamma}_T)$ and $\widehat{g}_{2,\widehat{\gamma}_T}(x) = p^L(x_t)\widehat{\beta}_2(\widehat{\gamma}_T)$ in each subsample. Define

$$D(\gamma) = E\left(p^L(x_t)p^L(x_t)'|z_t = \gamma\right), \tag{3.10}$$

$$V(\gamma) = E\left(p^L(x_t)p^L(x_t)'u_t^2|z_t = \gamma\right). \tag{3.11}$$

Let $D = D(\gamma_0)$ and $V = V(\gamma_0)$.

Then I use the following assumptions for the consistency of $\widehat{\gamma}_T$.

Assumption 3.2.4 $Q > Q_{1,\gamma} > 0$ *for each* $\gamma \in \Gamma = \left[\underline{\gamma}, \overline{\gamma}\right]$.

Assumption 3.2.5 *$f(\gamma)$, $D(\gamma)$ and $V(\gamma)$ are continuous at $\gamma = \gamma_0$.*

Assumption 3.2.6 *Let $g_1(x) - g_2(x) = \delta(x)$, where $\delta(x)$ can be approximated by the series expansion $p^L(x)\delta$ and δ satisfies Assumption 3.2.3; $\delta' D\delta > 0$ and $\delta' V\delta > 0$.*

Assumption 3.2.4 is the conventional full-rank condition which excludes perfect collinearity. Γ is restricted to be a proper subset of the support of z. Assumption 3.2.5 requires the moment functionals to be continuous so that one can obtain the Taylor expansion around γ_0. This condition excludes regime-dependent heteroscedasticity in Assumption 3.1.3 but it allows a smooth conditional variance function on z_t. Assumption 3.2.6 excludes the continuous threshold model.[6] The following theorem establishes the convergence rate of $\widehat{\gamma}$.

Theorem 3.2.3 *Under Assumptions 3.1.1-3.2.6, $\widehat{\gamma} \xrightarrow{p} \gamma_0$ as T increases to infinity. Furthermore, $T(\widehat{\gamma} - \gamma_0) = O_p(1)$.*

The above theorem shows that $\widehat{\gamma}_T$ converges to the true point γ_0 at rate T, even if the number of regressors grows to infinity. The intuition is that the threshold effect is a parametric part in a semiparametric model where the nonparametric part does not affect its convergence. This super convergence rate ensures that I can derive the asymptotics of $\widehat{\beta}_i(\widehat{\gamma}_T)$ and $\widehat{g}_{i,\widehat{\gamma}_T}(x)$ as the true threshold effect γ_0 is known.

Theorem 3.2.4 *Under Assumptions 3.1.1-3.2.6, $\widehat{\beta}_1(\widehat{\gamma}_T)$, $\widehat{\beta}_2(\widehat{\gamma}_T)$, $\widehat{g}_{1,\widehat{\gamma}_T}(x)$ and $\widehat{g}_{2,\widehat{\gamma}_T}(x)$ have the same asymptotic distribution as $\widehat{\beta}_1(\gamma_0)$, $\widehat{\beta}_2(\gamma_0)$, $\widehat{g}_{1,\gamma_0}(x)$ and $\widehat{g}_{2,\gamma_0}(x)$.*

[6]This paper focuses on the discontinuous threshold effect. For continuous threshold models, one is referred to Chan and Tsay (1998).

3.2.3 Asymptotic Distribution of the Estimator $\widehat{\gamma}_T$

In this section, the asymptotic distribution of the least-squares estimator $\widehat{\gamma}_T$ is derived under the assumption that the magnitude of threshold effect goes to zero at an appropriate rate. As pointed out by Hansen (2000), the assumption of decaying change size is needed in order to obtain an asymptotic distribution of $\widehat{\gamma}_T$ free of nuisance parameters.[7] I replace the Assumption 3.2.6 by the following assumption:

Assumption 3.2.7 *Let $g_1(x) - g_2(x) = T^{-\alpha}\delta(x)$, where $0 < \alpha < \frac{1}{2}$ and $\delta(x)$ can be approximated by the series expansion $p^L(x)\delta$ with δ satisfying 3.2.3; $\delta' D\delta > 0$ and $\delta' V\delta > 0$.*

The following theorem establishes the asymptotic distribution of $\widehat{\gamma}_T$.

Theorem 3.2.5 *Under Assumptions 3.1.1-3.2.5 and 3.2.7, the following result holds:*

$$T^{1-2\alpha}(\widehat{\gamma}_T - \gamma_0) \Rightarrow \omega\Lambda, \tag{3.12}$$

where

$$\omega = \frac{\delta' D\delta}{f_0(\delta' V\delta)^2}$$

and

$$\Lambda = \underset{-\infty<r<\infty}{\arg\max}\left(-\frac{1}{2}|r| + W(r)\right).$$

$W(r)$ is a two-sided Brownian motion on the real line defined as:

$$W(r) = \begin{cases} \Lambda_1(-r), & \text{if} \quad r < 0 \\ 0, & \text{if} \quad r = 0 \\ \Lambda_2(r), & \text{if} \quad r > 0 \end{cases},$$

[7] This approach was first used in the literature of change points (Bai, 1997) and applied to the threshold model by Hansen (2000).

with $\Lambda_i(r)$*,* $i = 1, 2$ *two independent standard Brownian motions on* $[0, \infty)$.

The rate of convergence is $T^{1-2\alpha}$, which is decreasing in α. Intuitively, a large α decreases the change size, in turn reducing the precision of any estimator of γ_0. In the leading case of conditional homoscedasticity, $V = \sigma^2 D$, and I have

$$\omega = \frac{\sigma^2}{f_0 c' D c},$$

where $\sigma^2 = E(u_t^2)$. Hansen (2000) shows that the $1 - c$ quantile of the random variable $\max_{r \in (-\infty, \infty)}(2W(r) - |r|)$ is given by $-2\ln(1 - \sqrt{1-c})$. Therefore, with the estimation of $\widehat{\omega}$, I can calculate the confidence intervals for $\widehat{\gamma}_T$.

3.2.4 Generalized sup-Wald Statistic

After obtaining the estimator $\widehat{\gamma}_T, \widehat{\beta}_1(\widehat{\gamma}_T)$ and $\widehat{\beta}_2(\widehat{\gamma}_T)$, one may wish to test whether or not the economic relationships are really different in each sub-sample. The null hypothesis is:

$$H_0 : g_1(x) = g_2(x) \quad \text{for any} \quad x \in \chi.$$

The alternative is:

$$H_0 : g_1(x) \neq g_2(x) \quad \text{for some} \quad x \in \chi.$$

Following Bai et al. (2008), I define a HAC robust Wald test statistic. Let

$$W_T(\gamma) = \left(\widehat{\beta}_1(\gamma) - \widehat{\beta}_2(\gamma)\right)' \widehat{\Omega}^{-1} \left(\widehat{\beta}_1(\gamma) - \widehat{\beta}_2(\gamma)\right) \tag{3.13}$$

where

$$\widehat{\Omega} = \widehat{\Omega}_{1,\gamma} + \widehat{\Omega}_{2,\gamma},$$

$$\widehat{\Omega}_{1,\gamma} = [GI_1(\gamma)G']^{-1}[(GI_1(\gamma)u)(GI_1(\gamma)u)'][GI_1(\gamma)G']^{-1},$$
$$\widehat{\Omega}_{2,\gamma} = [GI_2(\gamma)G']^{-1}[(GI_2(\gamma)u)(GI_2(\gamma)u)'][GI_2(\gamma)G']^{-1}.$$

The term $\widehat{\Omega}$ is covariance estimator which is robust to heteroscedasticity. One can also employ the Newey-West (1987) estimator to make it robust to serial correlation as well. Furthermore, to increase the power of the test, one can use fixed-b nonparametric covariance matrix estimators as Kiefer and Vogelsange (2005).

However, under the null hypothesis, γ is a nuisance parameter and cannot be identified. Following Hansen (1996, 2000), I take a sup-norm on $W_T(\gamma)$ in a closed set $\Gamma \subset \left[\underline{\gamma}, \overline{\gamma}\right]$. Define

$$W = \sup_{\gamma \in \mathbf{S}} W_T(\gamma)$$

As T increases, L converges to infinity. In that case, I modify the test statistics to a generalized sup-Wald statistic defined as follows

$$M_T(\gamma) = \frac{W_T(\gamma) - L}{\sqrt{2L}}$$

Theorem 3.2.6 *Under Assumptions 3.1.1-3.2.6 and $H_0 : g_1 = g_2$, as $T \to \infty$, for each fixed γ, $M_T(\gamma) \Rightarrow N(0, 1)$.*

Similarly, define the sup-norm of $M_T(\gamma)$ as $M_T = \sup_{\gamma \in \mathbf{S}} M_T(\gamma)$. The limiting distribution of M_T is a non-standard distribution and I generate the critical values using a parametric bootstrapping method. I first estimate $\widehat{g}(x_t)$ using the whole sample under the restriction that $g_1(x) = g_2(x)$. Then, I obtain the residual terms $\{\widehat{\eta}_t(\widehat{\gamma}_T)\}_{t=1}^T$ from the unrestricted model. I draw a random variable $\overline{\eta}_t^b$ from the sample $\{\widehat{\eta}_t(\widehat{\gamma}_t)\}_{t=1}^T$ for all $t = 1, ..., T$, and generate a new sequence $\{y_t^b\}_{t=1}^T$ by

$y_t^b = \widehat{g}(x_t) + \widetilde{\eta}_t^b$. Let M_T^b be the sup-Wald test calculated from the new data set $\{y_t^b$, x_t, $z_t\}_{t=1}^T$. Under the null , the distribution of M_T^b can approximate the distribution of M_T. The bootstrap p-value can be obtained by calculating the frequency of simulated M_T^b that exceeds M_T when the number of the simulations is large enough. As shown in Hansen(1996), the generated p-value converges to the true size.

3.3 Extension to the Model with Multiple Threshold Effects

In empirical studies, it is likely that there exist more than one threshold effects. If the number of threshold effects is known, a global estimation method can be designed to estimate all threshold values simultaneously. However, this number is usually unknown in practical applications. Following Gonzalo and Pitarakis (2002), I estimate the thresholds with a sequential method.

Consider the following nonparametric model with q change points at unknown locations:

$$y_t = \sum_{i=1}^{q+1} I(\gamma_{i-1} \leq z_t < \gamma_i) g_i(x_t) + u_t. \tag{3.14}$$

The model has $q + 1$ regimes defined by q change points: $\gamma_1, ..., \gamma_q$. Throughout, define $\gamma_0 = \underline{\gamma}, \gamma_{q+1} = \overline{\gamma}$. By series expansion, one can re-write the model as

$$y_t = \sum_{t=1}^{q+1} I(\gamma_{t-1} \leq z_t < \gamma_t) p^L(x_t)\beta_t + \varepsilon_t,$$

where $\varepsilon_t = \sum_{i=1}^{q+1} e_i(x_t) I(\gamma_{i-1} \leq z_t < \gamma_i) + u_t$, for $t = 1, 2, ..., T$, and $e_i(x_t)$ is the remainder term from the series expansion of $g_i(x_t)$ using basis functions $p^L(x_t)$. The model can be compactly written as

$$Y = \sum_{i=1}^{q+1} I_i G \beta_t + \epsilon \tag{3.15}$$

where Y and G are defined as before, I_i is a $T \times T$ diagonal matrix with the $(t,t)^{th}$ element being an indicator function $I(\gamma_{i-1} \leq z_t < \gamma_i)$.

The sequential estimation method starts by estimating a model with a threshold effect:

$$Y = I_1(\gamma)G\widehat{\beta}_1(\gamma) + I_2(\gamma)G\widehat{\beta}_2(\gamma) + \widehat{\epsilon} \tag{3.16}$$

where $\widehat{\beta}_1(\gamma)$ and $\widehat{\beta}_2(\gamma)$ are OLS estimators. Define

$$\widehat{\gamma}_T = \arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR_T(\gamma),$$

where $SSR_T(\gamma)$ is the sum of residual squares. Using a similar argument to that of Proposition 2.3 in Gonzalo and Pitarakis (2002), $SSR_T(\gamma)$ converges to a continuous function $R(\gamma)$ uniformly and $R(\gamma)$ takes its minimum value at one of thresholds γ_k. Thus, I can estimate all change points by using the following sequential methods. I first estimate one of the change points $\widehat{\gamma}_T$ from the whole sample, and then split the sample into two sub-samples at the estimated threshold effect $\widehat{\gamma}_T$. Within both sub-samples $[-\infty, \widehat{\gamma}_T]$ and $(\widehat{\gamma}_T, \infty]$, estimate $\widehat{\gamma}'_T = \arg\min SSR_T\left(\gamma, \underline{\gamma}, \widehat{\gamma}_T\right)$ and $\widehat{\gamma}''_T = \arg\min SSR_T(\gamma, \widehat{\gamma}_T, \overline{\gamma})$ to obtain the next two threshold values, assuming each subsample contains a threshold effect. This process continues until the null hypothesis of no threshold effect is accepted in each sub-sample. The generalized sup-Wald test statistic defined in Section 3.4 can be used to test the null hypothesis in each sub-sample.

3.4 Simulations

In this section, I carry out Monte Carlo simulation experiments to investigate the finite sample performance of the estimators and tests statistics.[8] For com-

[8]The codes are R language programs and they are available on request.

parison, in each experiment, I consider two models: a linear model and a nonparametric model. I estimate the threshold effects using the profile least square method and construct the sup-Wald statistics for both models. The size and size-corrected power are reported for each experiment to evaluate the test statistics.[9]

Specifically, I estimate the model using the following two model specifications:

Linear Model

$$y_t = \left\{ \begin{array}{ll} a_1 + b_1 x_t + u_t, & \text{if } z_t \leq \gamma \\ a_2 + b_2 x_t + u_t, & \text{otherwise} \end{array} \right\}$$

Nonparametric Model

$$y_t = \left\{ \begin{array}{ll} g_1(x_t) + u_t, & \text{if } z_t \leq \gamma \\ g_2(x_t) + u_t, & \text{otherwise} \end{array} \right\}.$$

I use polynomial power series as the basis functions. The number of basis function L is very important in practical applications. If L is too small, the test will tend to accept the null hypothesis erroneously since the truncated series $\sum_{s=1}^{L} p_s^L(x)\beta_s^L$ is a poor approximation to $g(x)$. If L is too large, $\widehat{g}$ will be very noisy estimator of g and this will tend to cause rejection of a correct H_0. Following the literature of series estimations, I use the generalized cross-validation (GCV) to select L. L is chosen to minimize the value of GCV defined as follows

$$\widehat{L} = \arg\min_{L \in H_t} \frac{T^{-1} SSR_T(\widehat{\gamma})}{(1 - T^{-1} tr(M_T(L))^2},$$

where $SSR_T(\widehat{\gamma})$ is the sum of residual square of the estimated model and $M_T(L) = G(G'G)^{-}G'$.

Experiment 1. This experiment shows the performance of the estimation and tests in a regression model with a threshold effect on the quadratic term.

[9] One can refer to Appendix B for the procedures to generate the size-corrected power.

DGP:

$$y_t = \left\{ \begin{array}{cc} 1 + x_t + 0.2(1 + x_t^2)u_t, & \text{if} \quad z_t \leq 0 \\ 1 + x_t + \delta(x_t - 0.5)^2 + 0.2(1 + x_t^2)u_t, & \text{otherwise} \end{array} \right\}$$

where $u_t \sim i.i.d.N(0,1)$; $x_t \sim i.i.d.U(-1,1)$. $\{x_t\}_{t=1}^T$ and $\{u_t\}_{t=1}^T$ are independent of each other. The threshold variable z_t is *i.i.d.* $N(0,1)$. I allow for conditional heteroscedasticity. The set of sample sizes I consider is $\{T = 100, 200, 400\}$. Replication number of the simulation is $N = 1000$; The Bootstrap repeat number is $BN = 2000$. δ measures the size of the threshold effect.

The results of the estimation and testing are summarized in Table 3.1 and Table 3.2. Table 3.1 reports the size and size-corrected power of the sup-Wald

Table 3.1: Size and size-corrected power for Experiment 1

		Linear Model			**Semiparametric Model**		
Sample Size	δ	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
	0	.147	.126	.040	.12	.071	.006
100	0.5	.584	.480	.249	.979	.951	.894
	1	.997	.957	.872	1	1	.998
	0	.131	.062	.015	.113	.059	.014
200	0.5	.883	.821	.648	.964	.926	.790
	1	1	1	.999	1	1	1
	0	.126	.069	.025	.098	.046	.015
400	0.5	.984	.972	.874	.988	.969	.908
	1	1	1	1	1	1	1

statistics for the linear model and the nonparametric model. One can find that the Wald statistic based on the series estimation methods performs very well, as the sample size increases. For the linear model, the HAC robust Wald-statistic still works in this case, which is consistent with the theoretical prediction from

Table 3.2: MSE of the threshold estimators for Experiment 1

Sample Size	**Linear Model**		**Semiparametric Model**	
	$\delta = 1$	$\delta = 0.5$	$\delta = 1$	$\delta = 0.5$
100	4.44	8.74	.33	2.22
200	.186	1.32	.025	.510
400	.070	.680	.010	.190

Note: All MSE values in this table have been multiplied by 100.

Bai et al (2008). However, the size-corrected power is lower than in the nonparametric approach. Table 3.2 reports the MSE of the estimators. The linear-model approach has far larger MSE.

Experiment 2. This experiment shows the performance of the estimation and tests in a regression model with correlated threshold variable.

DGP:

$$y_t = \begin{Bmatrix} 1 + x_t + 0.2(1 + x_t^2)u_t, & \text{if} \quad z_t \leq 0 \\ 1 + x_t + \delta(x_t - 0.5)^2 + 0.2(1 + x_t^2)u_t, & \text{otherwise} \end{Bmatrix}$$

where $u_t \sim i.i.d.N(0,1)$; $x_t \sim i.i.d.U(-1,1)$. $\{x_t\}_{t=1}^T$ and $\{u_t\}_{t=1}^T$ are independent of each other. The threshold variable $z_t = 0.5x_t + e_t$, where e_t is *i.i.d.* $N(0,1)$. Therefore, z_t and x_t be correlated with each other. I allow for conditional heteroscedasticity. The set of sample sizes I consider is $\{T = 100, 200, 400\}$. Replication number of the simulation is $N = 1000$; The Bootstrap repeat number is $BN = 2000$. δ measures the size of the threshold effect.

The results of the estimation and testing are summarized in Table 3.3 and Table 3.4. Table 3.3 reports the sizes and size-corrected powers. The linear-model approach has a large size distortion even when sample size is large, which im-

Table 3.3: Size and size-corrected power for Experiment 2

		Linear Model			**Semiparametric Model**		
Sample Size	δ	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
	0	.29	.206	.075	.135	.085	.019
100	0.5	.566	.473	.263	.59	.464	.263
	1	.968	.936	.818	.992	.982	.948
	0	.44	.324	.149	.125	.057	.016
200	0.5	.739	.608	.606	.948	.893	.808
	1	1	1	.994	1	1	1
	0	.483	.384	.188	.116	.06	.011
400	0.5	.933	.908	.897	.998	.988	.952
	1	1	1	1	1	1	1

Table 3.4: MSE of the threshold estimators for Experiment 2

Sample Size	**Linear Model**		**Semiparametric Model**	
	$\delta = 1$	$\delta = 0.5$	$\delta = 1$	$\delta = 0.5$
100	9.28	25.4	1.05	14.6
200	3.79	14.7	0.2	4.42
400	1.3	11.2	0.12	2.31

Note: All MSE values in this table have been multiplied by 100.

plies that it may lead to spurious threshold effects. The generalized sup-Wald statistic still performs very well, on both size and power. The results of estimation from Table 3.4 also show that the nonparametric model approach has smaller MSE.

Experiment 3. This experiment shows the performance of the estimation and tests in a regression model with sin and exponential functions.

DGP:

$$y_t = \left\{ \begin{array}{cc} 1 + x + \exp(-x) + 0.1(1 + x_t^2 + 0.5z_t^2)u_t, & \text{if} \quad z_t \leq 0 \\ 1 + x + \delta \sin(x^2)(x - 0.5)^2 + \exp(-x) + 0.1(1 + x_t^2 + 0.5z_t^2)u_t, & \text{otherwise} \end{array} \right\}$$

where $u_t \sim i.i.d.N(0, 1)$; $x_t \sim i.i.d.U(-1, 1)$. $\{x_t\}_{t=1}^T$ and $\{u_t\}_{t=1}^T$ are independent of each other. The threshold variable $z_t = 0.5x_t + e_t$, where e_t is *i.i.d.* $N(0, 1)$. Therefore, z_t and x_t be correlated with each other. I allow for conditional heteroscedasticity. Moreover, the conditional variance depends on both the regressor and threshold variable. Again, the set of sample sizes I consider is $\{T = 100, 200, 400\}$. Replication number of the simulation is $N = 1000$; The Bootstrap repeat number is $BN = 2000$. δ measures the size of the threshold effect.

The results of the estimation and testing are summarized in Table 3.5 and Table 3.6. Table 3.5 reports the sizes and size-corrected powers. The results

Table 3.5: Size and size-corrected power for Experiment 3

		Linear Model			**Semiparametric Model**		
Sample Size	δ	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.99$
	0	.440	.324	.149	.157	.079	.026
100	0.5	.456	.273	.153	.620	.489	.244
	1	.730	.532	.0.36	.81	.691	.495
	0	.506	.396	.202	.119	.074	.018
200	0.5	.554	.419	.199	.711	.605	.523
	1	.940	0.894	.801	1	0.998	.995
	0	.484	.384	.188	.112	.065	.015
400	0.5	.733	.649	.408	.816	.712	.537
	1	.997	.954	.901	1	1	1

are consistent with my expectation that linear-model approach has a large size

Table 3.6: MSE of the threshold estimators for Experiment 3

Sample Size	Linear Model		Semiparametric Model	
	$\delta = 1$	$\delta = 0.5$	$\delta = 1$	$\delta = 0.5$
100	13.8	35.2	2.01	17.2
200	8.71	18.4	0.53	7.22
400	3.81	19.5	0.204	3.67

Note: All MSE values in this table have been multiplied by 100.

distortion, while the semiparametric-model approach has reasonably good performance. Table 3.6 also shows that the semiparametric-model approach has smaller MSE than linear-model approach. Overall, when the functional form of the regression deviates more from the linear model, the semiparametric approach can generate more benefits.

3.5 Empirical Application

Durlauf and Johnson (1995) and Hansen (2000) test the convergence hypothesis by analyzing the relationship between the economic growth rate and the initial endowment of various countries. Their models are linear and parametric. However, a parametric assumption on the model may cause model misspecification, and in turn cause a misleading result. In this section, I apply the nonparametric approach to re-examine the convergence hypothesis using a larger dataset.

The data I use is from Bernanke and Gürkaynank (2001). The data set is drawn from the Summers–Heston Penn World Tables (PWT) version 6.0, which extends the data through 1998 for most of the variables. Following Alfo et al. (2008), I use a 5-year average for each variable during non-overlapping 5-

year periods from 1960 to 1995. The covariates include POP (population growth rate), SEC (human capital measured as the enrollment rate in secondary school), and INV (share of output allocated to investment). The dependent variable is the growth rate of per capita GDP. Following the literature, the choice of 5-year periods is to retain sufficient degrees of freedom while avoiding the negative effects of strong autocorrelation of dependent variables (see Bond et al. 2001). The total number of observations is 784, across 98 countries. After dropping the oil producing countries and the countries with poor quality data and missing data, there are 476 observations left, across 75 countries.

I estimate the following nonparametric model:

$$y_{i,t} = \begin{cases} \alpha_1 + b_1 D_t + \beta_1 \ln(GDP)_{i,1960} + \theta_1 POP_{it} + g_1(INV_{it}, SEC_{it}) + e_{i,t}, & \text{if} \quad z_{i,t} \leq \gamma \\ \alpha_2 + b_2 D_t + \beta_2 \ln(GDP)_{i,1960} + \theta_2 POP_{it} + g_2(INV_{it}, SEC_{it}) + e_{i,t}, & \text{if} \quad z_{i,t} > \gamma \end{cases}$$

In this model, the time dummy variables $D_t's$ and initial per capital GDP at 1960 are used to capture the time effect and the individual effect. If coefficients β_1 and β_2 are significantly negative, we can conclude that the poor countries grow faster than rich countries, and thus the convergence hypothesis holds. In the previous literature, the production function is assumed to be of Cobb-Douglas form, so that $g_j(INV_{it}, SEC_{it})$ is a linear function. However, this assumption may not reflect reality. I assume $g_j(INV_{it}, SEC_{it})$ to be a nonparametric function and use a polynomial series to approximate it. The number of basis functions is chosen by the GCV criterion. The threshold variable $z_{i,t}$ is ln(GDP)$_{i,1960}$. Table 3.5 reports the estimation results for both the linear and nonparametric models.

For the linear model, the threshold estimator is 8.258 ($3854 for the initial per capital GDP). From the second and third columns of Table 3.5, one can find that the coefficients for the initial per capital GDP are not significant. The su-

Table 3.7: Empirical Estimation Results

	Linear				**Semiparametric**			
	ln(GDP)$_{60} \leq \widehat{\gamma}$		**ln(GDP)$_{60} > \widehat{\gamma}$**		**ln(GDP)$_{60} \leq \widehat{\gamma}$**		**ln(GDP)$_{60} > \widehat{\gamma}$**	
Variables	**Coef**	**t-stat**	**Coef**	**t-stat**	**Coef**	**t-stat**	**Coef**	**t-stat**
*Dummy*65	.68	2.89	.38	2.23	.67	1.79	.20	1.01
*Dummy*70	.67	2.85	.37	2.17	.65	1.76	.19	1.05
*Dummy*75	.57	2.42	.23	1.34	.51	1.39	.06	0.36
*Dummy*80	.58	2.51	.30	1.77	.56	1.54	.12	0.67
*Dummy*85	.52	2.26	.33	1.97	.54	1.49	.15	0.83
*Dummy*90	.54	2.33	.29	1.72	.53	1.46	.12	0.65
*Dummy*95	.44	2.13	.23	1.58	.50	1.38	0.07	0.40
ln(GDP)$_{1960}$	−.03	−1.74	−.05	−1.86	−.05	−2.34	−0.04	−2.03

per Wald statistic is 19.67 and the bootstrap 95% critical value is 22.37. Thus, the results do not support the convergence hypothesis. For the nonparametric model, the threshold estimation is 8.23 ($3740.60 for the initial per capital GDP). Notice that the estimation of thresholds in the linear model is very close to that estimated using the nonparametric model. This is consistent with the theoretical prediction of Bai et al.(2008). Moreover, the generalized sup-Wald statistic is 83.26 and the bootstrap critical value is 78.64. The coefficients of ln(GDP)$_{1960}$ are significantly negative. Thus, the results offer some support to the convergence hypothesis and the existence of multiple growth patterns across countries.

3.6 Conclusion

This chapter proposes a method to detect the threshold effect without any parametric assumption on the regression functional forms. The method can avoid the risk of model misspecification which may lead to spurious threshold effect or overlook true threshold effects. The estimation and test statistic are based on series approximation techniques which are very convenient for imposing certain model restrictions, such as additive separability and shape-preserving. I derive the asymptotics for the estimators and develop a generalized sup Wald statistic to test the existence of the threshold effect. A nice property of the estimator is that it achieves the same convergence rate (T-convergence rate) as in parametric models. This super convergence rate enables me to study the asymptotics of the series estimators as the true threshold value is known. The generalized sup Wald statistics is robust to certain conditional heteroscedasticity. I provide an empirical application to test the convergence hypothesis for economic growth across countries over time. The results show that the poor countries grow faster than rich countries in general. However, they may converge to a different steady state.

CHAPTER 4

THRESHOLD COINTEGRATION AND PRICE DISCOVERY

4.1 Price discovery of cross-listings

I first develop an equilibrium model to characterize the interactive dynamics of a cross-listed pair simultaneously traded on two separate exchanges. Arbitrageurs linking the two markets may be subject to market frictions, such as transaction fees, capital constraints etc. Throughout the model, I emphasize the role of arbitrageurs in the process of inter-market price discovery.

I assume that there are two cross-border stock exchanges: the TSX and the NYSE, indexed by $i = 1, 2$. I further assume that there are N_1 participants who trade only in the home market (TSX) and N_2 participants who trade only in the NYSE market, and N_3 arbitrageurs who trade in both markets. The former two groups are one-market traders, and the third group is two-market traders. I focus on the dynamics between two market prices; thus I assume the choice of exchanges for one-market traders is fixed and exogenous to the model. They choose to trade in one specific exchange due to various reasons, such as distance, language, institutional constraints, transaction costs.

The behavior of a one-market trader in market i can be specified in the following manner. At time t, for the trader j, let $E^i_{j,t}$ be her endowment and $\mu^i_{j,t}$ be the reservation price at which she is willing to hold $E^i_{j,t}$ of assets. Given a market price $p_{i,t}$,her demand function can be conjectured as

$$X_{i,j,t} = E^i_{j,t} - \beta(\mu^i_{j,t} - p_{i,t}), \quad \text{for} \quad j = 1, 2, ..., N_i.$$

where $\beta > 0$ is the demand elasticity assumed to be the same for all one-market traders in both markets.

I now consider the demand function of arbitrageurs. Arbitrageurs are initially endowed with no seed money. Arbitrageurs "buy low and sell high" between the two markets; thus their demand function only depends on the cross-border price deviation. Given the market prices $p_{2,t}$ and $p_{1,t}$, arbitrageur j would submit her buy order in market 1 as:

$$X^A_{1,j,t} = \beta^A_{j,t}(p_{2,t} - p_{1,t})$$

$\beta^A_{j,t} > 0$ is the demand elasticity[1]. Since she hedges perfectly,

$$X^A_{2,j,t} = -X^A_{1,j,t} = \beta^A_{j,t}(p_{1,t} - p_{2,t})$$

i.e., her short position in one market always equals her long position in the other market.

In equilibrium, the two exchanges clear as

$$\begin{aligned}
\sum_{j=1}^{N_1} E^1_{j,t} &= \sum_{j=1}^{N_1} X_{1,j,t} + \sum_{j=1}^{N_3} \beta^A_{j,t}(p_{2,t} - p_{1,t}),\\
\sum_{j=1}^{N_2} E^2_{j,t} &= \sum_{j=1}^{N_2} X_{2,j,t} + \sum_{j=1}^{N_3} \beta^A_{j,t}(p_{1,t} - p_{2,t}).
\end{aligned}$$

Solving the market clearing conditions for equilibrium prices of the cross-listed pair yields

$$p_{1,t} = \frac{(\beta N_2 + \sum_{j=1}^{N_3} \beta^A_{j,t})N_1\mu^1_t + \sum_{j=1}^{N_3} \beta^A_{j,t} N_2 \mu^2_t}{(\beta N_2 + \sum_{j=1}^{N_3} \beta^A_{j,t})N_1 + \sum_{j=1}^{N_3} \beta^A_{j,t} N_2} \tag{4.1}$$

$$p_{2,t} = \frac{(\beta N_1 + \sum_{j=1}^{N_3} \beta^A_{j,t})N_2\mu^2_t + \sum_{j=1}^{N_3} \beta^A_{j,t} N_1 \mu^1_t}{(\beta N_1 + \sum_{j=1}^{N_3} \beta^A_{j,t})N_2 + \sum_{j=1}^{N_3} \beta^A_{j,t} N_1} \tag{4.2}$$

where $\mu^1_t = \frac{1}{N_1}\sum_{j=1}^{N_1} \mu^1_{j,t}$ and $\mu^2_t = \frac{1}{N_2}\sum_{j=1}^{N_2} \mu^2_{j,t}$ are market average reservation prices.

[1]Following Garbade and Silber (1983), demand elasticity for arbitrageurs $\beta^A_{j,t}$ is assumed to finite since the market is not frictionless.

In order to derive dynamic price relationships, I further specify a evolution mechanism of the reservation prices $\mu^1_{j,t}$ and $\mu^2_{j,t}$, following Garbade and Silber(1983), as

$$\mu^i_{j,t} = p_{i,t-1} + v_t + \varepsilon^i_{jt}, \quad \text{for} \quad i = 1,2, \quad j = 1,2,...,N_i.$$

As market i clears at the end of the period $t-1$ with a partial equilibrium price $p_{i,t-1}$, each trader decides to hold her share of assets toward her endowment in the subsequent period t, $E^i_{j,t}$. This implies that $p_{i,t-1}$ is the reservation price after the $t-1$ clearing. As new information on the issuer v_t common to all investors in both markets arrives, the trader formulates her new reservation prices $\mu^i_{j,t}$ with an idiosyncratic error ε^i_{jt}. I assume v_t and all ε^i_{jt} are i.i.d normal random variables with mean zero and constant variance.

In aggregate, the market reservation prices μ^1_t and μ^2_t can be expressed as

$$\begin{aligned}
\mu^1_t &= \frac{1}{N_1}\sum_{j=1}^{N_1}\mu^1_{j,t} = p_{1,t-1} + v_t + \frac{1}{N_1}\sum_{j=1}^{N_1}\varepsilon^1_{jt} \\
\mu^2_t &= \frac{1}{N_2}\sum_{j=1}^{N_2}\mu^2_{j,t} = p_{2,t-1} + v_t + \frac{1}{N_2}\sum_{j=1}^{N_2}\varepsilon^2_{jt}.
\end{aligned}$$

Plugging μ^1_t and μ^2_t into the equations (4.1) and (4.2), I have

$$\begin{aligned}
p_{1,t} &= \frac{(\beta N_2 + \sum_{j=1}^{N_3}\beta^A_{j,t})N_1 p_{1,t-1} + \sum_{j=1}^{N_3}\beta^A_{j,t}N_2 p_{2,t-1}}{(\beta N_2 + \sum_{j=1}^{N_3}\beta^A_{j,t})N_1 + \sum_{j=1}^{N_3}\beta^A_{j,t}N_2} + v_t + \widetilde{\varepsilon}^1_t, \\
p_{2,t} &= \frac{(\beta N_1 + \beta^A N_3)N_2 p_{2,t-1} + \sum_{j=1}^{N_3}\beta^A_{j,t}N_1 p_{1,t-1}}{(\beta N_1 + \sum_{j=1}^{N_3}\beta^A_{j,t})N_2 + \sum_{j=1}^{N_3}\beta^A_{j,t}N_1} + v_t + \widetilde{\varepsilon}^2_t,
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{\varepsilon}^1_t &= \frac{(\beta N_2 + \sum_{j=1}^{N_3}\beta^A_{j,t})\sum_{j=1}^{N_1}\varepsilon^1_{jt} + \sum_{j=1}^{N_3}\beta^A_{j,t}\sum_{j=1}^{N_2}\varepsilon^2_{jt}}{(\beta N_2 + \sum_{j=1}^{N_3}\beta^A_{j,t})N_1 + \sum_{j=1}^{N_3}\beta^A_{j,t}N_2}, \\
\widetilde{\varepsilon}^2_t &= \frac{(\beta N_1 + \sum_{j=1}^{N_3}\beta^A_{j,t})\sum_{j=1}^{N_2}\varepsilon^2_{jt} + \sum_{j=1}^{N_3}\beta^A_{j,t}\sum_{j=1}^{N_1}\varepsilon^1_{jt}}{(\beta N_1 + \sum_{j=1}^{N_3}\beta^A_{j,t})N_2 + \sum_{j=1}^{N_3}\beta^A_{j,t}N_1}.
\end{aligned}$$

An equivalent matrix representation prescribes

$$\begin{pmatrix} p_{1,t} \\ p_{2,t} \end{pmatrix} = \begin{pmatrix} 1-a_t, a_t \\ b_t, 1-b_t \end{pmatrix} \begin{pmatrix} p_{1,t-1} \\ p_{2,t-1} \end{pmatrix} - \begin{pmatrix} v_t + \widetilde{\varepsilon}_t^1 \\ v_t + \widetilde{\varepsilon}_t^2 \end{pmatrix} \tag{4.3}$$

where

$$\begin{aligned} a_t &= \frac{\sum_{j=1}^{N_3} \beta_{j,t}^A N_2}{\beta N_2 N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_2}, \\ b_t &= \frac{\sum_{j=1}^{N_3} \beta_{j,t}^A N_1}{\beta N_2 N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_2}. \end{aligned}$$

I can obtain the following bivariate Error Correction Model(VECM) by subtracting $\begin{pmatrix} p_{1,t-1} \\ p_{2,t-1} \end{pmatrix}'$ from both sides:

$$\begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \end{pmatrix} = \begin{pmatrix} -a_t, a_t \\ b_t, -b_t \end{pmatrix} \begin{pmatrix} p_{1,t-1} \\ p_{2,t-1} \end{pmatrix} - \begin{pmatrix} v_t + \widetilde{\varepsilon}_t^1 \\ v_t + \widetilde{\varepsilon}_t^2 \end{pmatrix} \tag{4.4}$$

The above VECM describes the short term dynamics toward the long-run equilibrium given by the cointegrating vector $(1, -1)$. The short term adjustment coefficients a_t and b_t for the prices $p_{1,t}$ and $p_{2,t}$ reflect their responses to deviations from the long-run equilibrium in their respective markets. I can apply the permanent transitory decomposition (Granger and Gonzalo, 1995) to the above VECM: the permanent component is a linear combination of ($p_{1,t}$, $p_{2,t}$), formed by the scaled orthogonal vector of the adjustment coefficient vector (a_t, b_t) Specifically, the permanent component is given by

$$f_t = \frac{b_t}{a_t + b_t} p_{1,t} + \frac{a_t}{a_t + b_t} p_{2,t}.$$

where

$$\frac{b_t}{a_t + b_t} = \frac{N_1}{N_1 + N_2} \tag{4.5}$$

and

$$\frac{a_t}{a_t + b_t} = \frac{N_2}{N_1 + N_2}. \tag{4.6}$$

The quantities $\frac{b_t}{a_t+b_t}$ and $\frac{a_t}{a_t+b_t}$ capture the contribution share of each price to the permanent component: they reflect the respective *information shares* of markets 1 and 2 toward determining the long-run equilibrium price. In other words, they are relative measures of market specific-contribution to price discovery of the cross-listed pair.

Define $\Delta p_t \equiv p_{1t} - p_{2t}$ as the dollar premium on the cross-listing against its original listing. It can be shown that

$$\Delta p_t = \delta_t \Delta p_{t-1} + e_t.$$

where

$$\begin{aligned}
\delta_t &= 1 - a_t - b_t \\
&= 1 - \frac{\sum_{j=1}^{N_3} \beta_{j,t}^A N_2}{\beta N_2 N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_2} - \frac{\sum_{j=1}^{N_3} \beta_{j,t}^A N_1}{\beta N_2 N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_2} \\
&= \frac{\beta N_2 N_1}{\beta N_2 N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_1 + \sum_{j=1}^{N_3} \beta_{j,t}^A N_2}.
\end{aligned}$$

Following Garbade and Silber (1983), δ_t measures the reciprocal convergence speed of the two market prices to their long-run equilibrium. Note that the smaller δ_t is, the faster the convergence occurs between two markets.

4.2 Error correction models

The equilibrium model constructed in Section 4.1 defines the measures of contribution share for each market, which are related to the relative populations of market participants (equations 4.5 and 4.6). This poses an empirical challenge since these numbers are usually unknown. Fortunately, one can estimate the adjustment coefficients a_t and b_t through the error correction model (equation 4.4), which only needs the information of market prices. However, another hurdle is that the adjustment coefficients a_t and b_t are time-varying. In order to estimate the model, some additional restrictions are necessary to characterize time paths of a_t and b_t. In the following three subsections, I discuss three different econometric models: standard linear ECM, threshold ECM, smooth transition ECM, under different assumptions on the demand elasticity of arbitrageurs.

4.2.1 Standard error correction model

I start from a standard error correction model, in which a_t and b_t are constant in equation 4.4. To satisfy this condition, it is sufficient to assume all arbitragers are homogeneous and share a constant demand elasticity, i.e., $\beta^A_{j,t} = \beta^A > 0$ for all j and t. It follows that

$$\begin{aligned} a_t &= \frac{N_3\beta^A N_2}{\beta N_2 N_1 + N_3\beta^A N_1 + N_3\beta^A N_2} \equiv a, \\ b_t &= \frac{N_3\beta^A N_1}{\beta N_2 N_1 + N_3\beta^A N_1 + N_3\beta^A N_2} \equiv b, \end{aligned}$$

and

$$\begin{aligned} \begin{pmatrix} \Delta p_{1,t} \\ \Delta p_{2,t} \end{pmatrix} &= \begin{pmatrix} -a, a \\ b, -b \end{pmatrix} \begin{pmatrix} p_{1,t-1} \\ p_{2,t-1} \end{pmatrix} - \begin{pmatrix} v_t + \widetilde{\varepsilon}_t^1 \\ v_t + \widetilde{\varepsilon}_t^2 \end{pmatrix} \\ &= \begin{pmatrix} -a \\ b \end{pmatrix} (p_{1,t-1} - p_{2,t-1}) - \begin{pmatrix} v_t + \widetilde{\varepsilon}_t^1 \\ v_t + \widetilde{\varepsilon}_t^2 \end{pmatrix}. \end{aligned}$$

Define the dollar premium on the cross-listing against its original listing as

$$\kappa_t \equiv p_{2t} - p_{1t}.$$

For notional convenience, I use α_1 and α_2 to replace $-a$ and b. A standard ECM for the bivariate cointegrated system of the cross-listed pair can be structured as

$$\begin{aligned}
\Delta p_{1t} &= \beta_{10} + \alpha_1 \kappa_{t-1} + \sum_{j=1}^{m_1} \beta_{1j} \Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{1j} \Delta p_{2t-j}, \\
\Delta p_{2t} &= \beta_{20} + \alpha_2 \kappa_{t-1} + \sum_{j=1}^{m_1} \beta_{2j} \Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{2j} \Delta p_{2t-j},
\end{aligned}$$

where κ_{t-1} gives the remaining cross-listing dollar premium or cointegrating residual. α_1 and α_2 are the adjustment coefficients of the TSX and the NYSE, respectively: they describe how much deviation will be subsequently adjusted to restore the long run equilibrium in each series. $\{\{\beta_{1j}, \widetilde{\beta}_{1j}, \beta_{2j}, \widetilde{\beta}_{2j}\}$ are coefficients for short term dynamics. By Granger Representation Theorem (Engle and Granger, 1987), if p_{1t} and p_{2t} are cointegrated, then at least one of α_1 and α_2 must be nonzero. In other words, one or both of p_{1t} and p_{2t}, will adjust fractionally to restore parity in the long run.

Harris et al. (1995, 2000) propose to use this linear ECM's adjustment coefficients to estimate the relative extent of exchange-respective contribution to price discovery (information share) of shares whose order purchases are fragmented across multiple markets. For a Canadian company originally listed on the TSX and cross-listed on the NYSE, the proportion of adjustments that took place on the TSX out of the total adjustments occurring on both exchanges is the share of the home exchange which contributes to setting the long-run equilibrium price as a result of synchronous cross-border stock trading. In an extreme case where there is no feedback from the NYSE so that $\alpha_1 = 0$, then the NYSE has no contribution to price discovery of the cross-listed pair. Eun and Sabherwal (2003) further define the respective information shares of the NYSE and the TSX as

$$\mathrm{IS}^{\mathrm{n}} \equiv \frac{|\alpha_1|}{|\alpha_1| + |\alpha_2|} \quad \text{and} \quad \mathrm{IS}^{\mathrm{t}} \equiv \frac{|\alpha_2|}{|\alpha_1| + |\alpha_2|}.$$

Suppose $p_{1t-1} < p_{2t-1}$ in the previous period $(t-1)$; then a likely scenario to reduce the gap between the two prices is for p_{1t} to increases or p_{2t} to decrease, or both. In this case one can conjecture that α_1 is non-positive and α_2 is non-negative. There two other possibilities: *1.* p_{1t-1} decreases but p_{2t-1} decreases more; or *2.* p_{1t-1} increases but p_{2t-1} increases less.[2] Eun and Sabherwal (2003) assign very low likelihoods to the latter two. One can analogously design a similar adjustment mechanism to show that α_1 is non-positive and α_2 is non-negative for the symmetric situation when $p_{1t-1} > p_{2t-1}$. Based on the above reasons, I can define the exchange-respective information shares of the NYSE and the TSX as

$$\mathrm{IS}^{\mathrm{n}} \equiv \frac{-\alpha_1}{-\alpha_1 + \alpha_2} \text{ and } \mathrm{IS}^{\mathrm{t}} \equiv \frac{\alpha_2}{-\alpha_1 + \alpha_2}.$$

4.2.2 Threshold error correction model

In reality, the market is imperfect due to various sources of market friction such as nonzero transaction costs, direct and indirect trading barriers, etc. Let γ measure the sum of all transaction costs and risk premiums required from arbitrageurs. Arbitrage opportunities exist if and only if

$$\kappa_{t-1} < -\gamma \text{ or } \kappa_{t-1} > \gamma,$$

which becomes $|\kappa_{t-1}| > \gamma$.[3]

[2] These odds may reflect the underreaction to the information share of the market. When information incorporation takes multiple periods, the price adjustment should persist unilaterally during this time.

[3] Transaction costs of cross-border arbitrage are comprised of the bid-ask spreads of the prices on both exchanges and the foreign exchange rate, fixed costs, and liquidity shorfalls. Chen and Choi (2010) find the relative premium of a Canadian cross-listing on the NYSE, on average includes an adverse-selection risk premium due cross-border imbalance in private information on the issuing firm. Along with the asymmetric information component, macroeconomic factors, such as GDP growth rates and interest rates, may also affect determination of the threshold.

I continue to assume all arbitrageurs are homogeneous. The demand elasticity for each arbitrageur can be written as

$$\beta^A_{j,t} = \left\{ \begin{array}{ll} 0, & \text{if} \quad |\kappa_{t-1}| \le \gamma \\ \beta^A > 0, & \text{otherwise} \end{array} \right\}.$$

Now the error correction dynamics become active unless the cross-listing dollar premium sufficiently digresses from parity beyond the threshold. Balke and Fomby (1997) propose this regime-switching mechanism as *threshold cointegration*, and the implied error correction dynamics can be characterized by a threshold ECM, given by

$$\Delta p_{1t} = \left\{ \begin{array}{ll} \beta_{110} + \alpha_{11}\kappa_{t-1} + \sum_{j=1}^{m_1} \beta_{11j}\Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{11j}\Delta p_{2t-j}, & \text{if} \quad |\kappa_{t-1}| \le \gamma \\ \\ \beta_{120} + \alpha_{12}\kappa_{t-1} + \sum_{j=1}^{m_1} \beta_{12j}\Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{12j}\Delta p_{2t-j}, & \text{if} \quad |\kappa_{t-1}| > \gamma \end{array} \right\}$$

and

$$\Delta p_{2t} = \left\{ \begin{array}{ll} \beta_{210} + \alpha_{21}\kappa_{it-1} + \sum_{j=1}^{m_1} \beta_{21j}\Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{21j}\Delta p_{2t-j}, & \text{if} \quad |\kappa_{t-1}| \le \gamma \\ \\ \beta_{220} + \alpha_{22}\kappa_{it-1} + \sum_{j=1}^{m_1} \beta_{22j}\Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{22j}\Delta p_{2t-j}, & \text{if} \quad |\kappa_{t-1}| > \gamma \end{array} \right\}.$$

In the middle regime when $|\kappa_{t-1}| \le \gamma$, there are neither market forces nor arbitrageurs to sustain cointegration of the pair of prices. In other words, unless the pair shows a significant price gap exceeding the threshold minimum profit, the adjustment coefficients are zeroes ($\alpha_{11} = \alpha_{21} = 0$) and, thus, neither price ($p_{1,t}$ nor $p_{2,t}$) appropriately reflects risks. I define the information share, or the relative measure of contribution to price discovery, for respective market using the outer regime coefficient estimates[4] (α_{12} and α_{22}):

$$\text{IS}^{\text{n}}_{out} \equiv \frac{|\alpha_{12}|}{|\alpha_{22}| + |\alpha_{12}|} \quad \text{and} \quad \text{IS}^{\text{t}}_{out} \equiv \frac{|\alpha_{22}|}{|\alpha_{22}| + |\alpha_{12}|}.$$

[4]Eun and Sabherwal (2003) estimate the adjustment coefficients in every period using a linear ECM following Harris *et al.* (1995).

Empirically, one can also define information shares $\mathrm{IS}^{\mathrm{n}}_{in}$ and $\mathrm{IS}^{\mathrm{t}}_{in}$ using the adjustment coefficients α_{11} and α_{21}, from the middle regime.

4.2.3 Smooth Transition Error Correction Model

An important assumption for standard linear ECM and threshold ECM is the homogeneity of arbitrageurs. In reality, arbitrageurs could face different threshold value ($\gamma_j' s$) to establish their positions. For example, fees paid by institutional investors may depend on the arrangement between the investors and the executing brokers. Meanwhile, the opportunity cost faced by capital-constrained arbitrageurs can be another reason for different threshold values: investors with stricter capital constraints will tend to skip small mispricings to wait for larger ones.

More specifically, I assume for arbitrageur j, where $j = 1, 2, ..., N_3$, that the demand elasticity is given by

$$\beta^A_{j,t} = \left\{ \begin{array}{cc} 0, & \text{if } -\gamma_j < \kappa_{t-1} < \gamma_j \\ \beta^A > 0, & \text{otherwise} \end{array} \right\}.$$

The "aggregated" thresholds will be a smooth function of the price deviation such that

$$\sum_{j=1}^{N_3} \beta^A_{j,t} = N_3 E(\beta^A_j) = N_3 \left(\int_{-\infty}^{-|\kappa_{t-1}|} \beta^A dF(\gamma) + \int_{|\kappa_{t-1}|}^{\infty} \beta^A dF(\gamma) \right) \equiv g(\kappa_{t-1}),$$

where $F(\gamma)$ is the probability distribution function of γ_j across all j.

Under this assumption, I have

$$a_t = \frac{g(\kappa_{t-1})N_2}{\beta N_2 N_1 + g(\kappa_{t-1})N_1 + g(\kappa_{t-1})N_2} \equiv \alpha_1(\kappa_{t-1}). \tag{4.7}$$

$$b_t = \frac{g(\kappa_{t-1})N_1}{\beta N_2 N_1 + g(\kappa_{t-1})N_1 + g(\kappa_{t-1})N_2} \equiv \alpha_2(\kappa_{t-1}). \tag{4.8}$$

By plugging (4.7) and (4.8) into equation (4.4), I obtain a smooth transition ECM:

$$\Delta p_{1t} = \beta_{10} + \alpha_1(\kappa_{t-1})\kappa_{t-1} + \sum_{j=1}^{m_1} \beta_{1j}\Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{1j}\Delta p_{2t-j},$$

and

$$\Delta p_{2t} = \beta_{20} + \alpha_2(\kappa_{t-1})\kappa_{t-1} + \sum_{j=1}^{m_1} \beta_{2j}\Delta p_{1t-j} + \sum_{j=1}^{m_2} \widetilde{\beta}_{2j}\Delta p_{2t-j}.$$

An average information share for each market can be defined as

$$\begin{aligned} \mathrm{IS}^{\mathrm{n}} &\equiv \frac{|E(\alpha_1(\kappa_{t-1}))|}{|E(\alpha_2(\kappa_{t-1}))| + |E(\alpha_2(\kappa_{t-1}))|} \\ \mathrm{IS}^{\mathrm{t}} &\equiv \frac{|E(\alpha_2(\kappa_{t-1}))|}{|E(\alpha_2(\kappa_{t-1}))| + |E(\alpha_1(\kappa_{t-1}))|}. \end{aligned}$$

where $E(\alpha_1(\kappa_{t-1}))$ and $E(\alpha_2(\kappa_{t-1}))$ can be estimated by the sample mean $\frac{1}{T}\sum_{t=1}^{T} \alpha_1(\kappa_{t-1})$ and $\frac{1}{T}\sum_{t=1}^{T} \alpha_2(\kappa_{t-1})$ respectively.

In order to see whether informed traders would choose to trade at the market with a lower price, conditional information shares can be defined for cases with a negative or positive price deviation,

$$\begin{aligned} \mathrm{IS}^{\mathrm{n}}_{\kappa>0} &\equiv \frac{|E(\alpha_1(\kappa_{t-1})I(\kappa_{t-1} > 0))|}{|E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} > 0))| + |E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} > 0))|}, \\ \mathrm{IS}^{\mathrm{t}}_{\kappa>0} &\equiv \frac{|E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} > 0))|}{|E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} > 0))| + |E(\alpha_1(\kappa_{t-1})I(\kappa_{t-1} > 0))|}. \end{aligned}$$

$$\begin{aligned} \mathrm{IS}^{\mathrm{n}}_{\kappa<0} &\equiv \frac{|E(\alpha_1(\kappa_{t-1})I(\kappa_{t-1} < 0))|}{|E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} < 0))| + |E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} < 0))|}, \\ \mathrm{IS}^{\mathrm{t}}_{\kappa<0} &\equiv \frac{|E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} < 0))|}{|E(\alpha_2(\kappa_{t-1})I(\kappa_{t-1} < 0))| + |E(\alpha_1(\kappa_{t-1})I(\kappa_{t-1} < 0))|}. \end{aligned}$$

Note that $\kappa_{t-1} > 0$ implies $p_{2t-1} > p_{1t-1}$, or a price premium on the NYSE.

I can also define an estimator of the convergence speed parameter δ_t

$$\delta(\kappa_{t-1}) = 1 + \alpha_1(\kappa_{t-1}) - \alpha_2(\kappa_{t-1}).$$

The average δ_t is defined as $\delta = E(\delta(\kappa_{t-1})) = \frac{1}{T}\sum_{t=1}^{T} \delta(\kappa_{t-1})$ and the conditional δ_t are defined as$\delta_{\kappa>0} = \frac{\sum_{t=1}^{T} \delta(\kappa_{t-1})I(\kappa_{t-1}>0)}{\sum_{t=1}^{T} I(\kappa_{t-1}>0)}$, and $\delta_{\kappa<0} = \frac{\sum_{t=1}^{T} \delta(\kappa_{t-1})I(\kappa_{t-1}<0)}{\sum_{t=1}^{T} I(\kappa_{t-1}<0)}$.

4.3 Data and preliminary results

I estimate the information shares for Canadian stocks cross-listed in the Toronto and New York stock exchanges between January 1, 1998, to December 31, 2000. 56 TSX-NYSE pairs are identified during the sample period.[5] In order to estimate asymmetric-information and market-friction measures, high-frequency data are required for the shares co-listed on the TSX and the NYSE, and for the U.S.-Canada exchange rate. Accordingly, I use the tick-by-tick trade and quote data for the TSX-listed Canadian stocks and the Trade-And-Quote (TAQ) data of their cross-listings on the NYSE through the period. The exchange rate intraday data was purchased from Olson & Associates.

Following Eun and Sabherwal (2003), I use the quoted prices, instead of transaction prices. The mid-points of the U.S.-Canada exchange rate bid and ask quotes are updated every minute, while the bid and ask quotes of the TSX-listed Canadian stocks are matched with their previous minutes' exchange rate quote mid-points and transferred to US$ prices. To reduce the impact of the market microstructure noise, I form the price series by taking the midpoint of the bid and ask quotes at the end of each 10-minute period.

To calculate price deviations between two markets, I require prices observed at the same time in these two markets. The regular trading time of the TSX and NYSE is the same (from 9:30 am to 4:00 pm Eastern time). Thus, for each day, I can observe 40 data points for each stock. Our sample period covers around 772 trading days, but not all stocks have two prices during the whole sample period since some stocks are cross-listed after Jan. of 1998. Our analysis is based on

[5] Eun and Sabherwal (2003)'s sample consists of 62 TSE-listed securities since their sample include those cross-listed in AMEX and Nasdaq.

the stock-year; thus, I require that each year, for each stock, the prices should be observed in two markets at least 6 months continuously. I drop thinly traded stocks in both markets. Our final sample includes just 44 stocks.

4.3.1 Cointegration analysis

I first examine whether pairs of times series on the TSX and NYSE price series are unit roots or not. I use the augmented Dickey and Fuller's (1981) ADF test, which considers lagged first differences of time series in the specification. If the test statistic is too large, then I reject the null hypothesis of unit root and conclude that the time series is stationary. As a result, the null hypothesis is rejected only for four out of 132 firm-years, at a five percent significance level. Thus, I conclude that both price series in my sample are, overall, first-order integrated process I(1).

I subsequently examine whether there exists cointegration between the two price series. As Eun and Sabherwal (2003) find that both S&P TSX Composite and S&P 500 indices (market indices of the TSX and the NYSE, respectively) are not significant in the cointegration system, I consider only the two market price series in each regression equation. Therefore, there is at most one cointegrating vector. I estimate the cointegrating vector for each cross-listed pair in each year. Table 1 reports summary statistics for the normalized estimation of the cointegrating vector[6] for p_{it}^{n} and $\widetilde{p}_{it}^{\mathrm{t}}$. The *t*-statistics for the null hypothesis attest that the cointegrating vector equals $(1, -1)^T$.

From Table 4.1, one can find that the median of the normalized estimates

[6]Normalized such that $b^{\mathrm{n}} = -1$.

Table 4.1: Estimated Cointegrating Vector

Quantiles	b^N	t-statistics
$5\% - ile$	0.9	−5.25
$25\% - ile$	0.995	−1.29
$Median$	0.999	0.25
$75\% - ile$	1.002	0.99
$95\% - ile$	1.011	2.94

throughout the sample is $(1, -1)$ which confirms that the Canadian cross-listed pairs tend to follow the law of one price and are thus cointegrated. Given the estimated cointegrating vector $(1, -b^t)$, the estimated cross-listing dollar premium is $\kappa_{it} \equiv p^{\mathrm{n}}_{it} - b^t \widetilde{p}^{\mathrm{t}}_{it}$. I now test κ_{it} for stationarity using the ADF test and find that only 3 out of 132 samples do not reject the null hypothesis of unit root. In sum, I can conclude that the TSX-NYSE cross-listed pairs are cointegrated with unit cointegrating vector.

4.3.2 Nonlinearity test

The law of one price suggests that two market prices for the same stock should not drift far from each other. This relationship is confirmed by the cointegration analysis in the previous section. However, linear adjustment dynamics is not necessarily prescribed by market efficiency assumptions. Given various market frictions, such as transactions costs and short sale limitations, It is thus more likely that a nonlinear model, such as a threshold cointegration model provides a better description of the convergence procedure between two market prices. In this section, I conduct several nonlinearity test in the course of short-run adjustment dynamics to long-run parity equilibrium.

I estimate a symmetric bivariate threshold ECM model (introduced in Section 4.2.2) and apply the super-Lagrangian multiplier (supLM) test to check the nonlinearity. As Hansen and Seo (2002) suggest, this test also has power to detect smooth transition ECM models. I use Akaike's (1974) and Schwart's (1978) Bayesian information criteria to choose the number of lags, and consistently choose the lag length of 1 ($m_1 = m_2 = 1$). The cointegrating vector is given as $(1, -1)$, following the results of cointegration tests.[7] The model is estimated by the maximum likelihood method described in Appendix A. Estimations are carried out in each year for each pair; results are reported in Table 4.2 below.

Panel A in Table 4.2 displays summary statistics of the threshold estimates and test statistics. The p-values are computed by the parametric bootstrap method suggested by Hansen and Seo (2002). From the table, one can observe that the mean and median of supLM over all samples are equal to 22.32, which exceeds the 95% critical value 22.07. Therefore, on average, I can reject the null hypothesis of no threshold effect. To further confirm the testing results, I apply a combined p-value test on all stock-years. Let p_i be the asymptotic p-value of the sup_LM test for each individual stock-year i, for $i = 1, 2, ..., N$, where N is the total number of stock-years. I combine all p-values using the Z test statistic proposed by Choi (2001)$Z = \frac{1}{2\sqrt{N}} \sum_{i=1}^{N} (-2\ln(p_i) - 2)$. As $N \rightarrow \infty$, under the null hypothesis, one can show that $Z \xrightarrow{d} N(0, 1)$. In the current case, the combined P-value test statistic Z is 10.41, significantly rejecting the null hypothesis (5% critical value is 1.96). Overall, I conclude that there exists nonlinearity in the convergence procedure between two market prices.

[7] I report the estimation and testing results with estimated cointegrating vector in Table2_supp. The estimation results are very similar and I do not find any change on the conclusion of about the test of nonlinearity.

Table 4.2: Nonlineaity tests

Panel A:	$Threshold$		$supLM$		$95\% - criticalvalue$		$P - value$	
Mean	0.193		73.516		24.783		0.146	
St.Dev.	0.159		110.81		3.562		0.249	
1%-ile	0.010		10.29		17.861		0	
10%-ile	0.059		14.14		21.113		0	
25%-ile	0.100		19.664		21.920		0	
50%-ile	0.157		31.054		23.101		0.008	
75%-ile	0.242		56.139		28.234		0.142	
90%-ile	0.319		242.699		29.078		0.609	
99%-ile	0.808		509.25		30.407		0.818	
Panel B:	$Wald_{ECM1}$	$P - value$	$Wald_{ECM2}$	$P - value$	$Wald_{DC1}$	$P - value$	$Wald_{DC2}$	$P - value$
Mean	10.118	0.265	32.723	0.260	13.33	0.265	9.349	0.246
St.Dev.	24.471	0.312	87.764	0.324	70.767	0.283	14.630	0.281
1%-ile	0.000	0.000	0.000	0.000	0.158	0.000	0.183	0.000
10%-ile	0.099	0.000	0.054	0.000	0.796	0.002	1.162	0.001
25%-ile	0.44	0.005	0.336	0.000	2.012	0.033	2.222	0.028
50%-ile	2.637	0.104	3.193	0.074	5.676	0.167	5.554	0.108
75%-ile	8.068	0.507	16.106	0.563	9.679	0.428	9.636	0.400
90%-ile	15.839	0.755	75.57	0.818	13.81	0.723	14.601	0.759
99%-ile	132.409	0.984	487.71	0.986	28.32	0.981	62.062	0.920

It may be interesting to examine whether the threshold effect occurs on the coefficients of the error correction term or short dynamic terms. I separately test the threshold effect in these coefficients. Panel B of Table 4.2 reports the test results. The first two columns report the results of Wald statistics for testing the null hypotheses, namely $H_0 : \alpha^{n}_{out} = \alpha^{n}_{in}$ and $\alpha^{t}_{out} = \alpha^{t}_{in}$ i.e., whether the adjustment coefficients are different within and beyond the threshold. The last two columns report the Wald statistics for the null hypothesis: $H_0 : \beta^{n}_{1,out} = \beta^{n}_{2,in}$, $\widetilde{\beta}^{n}_{1,out} = \widetilde{\beta}^{n}_{2,in}$, $\beta^{t}_{1,out} = \beta^{t}_{2,in}$, and $\widetilde{\beta}^{t}_{1,out} = \widetilde{\beta}^{t}_{2,in}$. The combined P-value test statistic Z of the null hypotheses of the error correction terms are: 19.68 and 16.80, while the combined P-value test statistic Z for dynamic coefficients are 14.44 and 13.95. Thus, for both exchanges, I can conclude that there is nonlinearity on both error correction terms and short term dynamics, but it appears that the threshold effect is more likely to take place on the error correction terms.

4.4 Estimation

Before conducting the regression analysis, I offer the following further discussion of the estimation results from the three models of Section 4.3.

4.4.1 Estimation of the threshold γ and convergence speed δ

In threshold ECM models, the threshold γ measures the size of transaction costs and risk premium. The first column of Panel A in Table 4.2 reports the summary statistics of estimated thresholds (γ) ranging from 0.009 to 0.545, with a mean of 0.146. That is to say, on average, when the cross-listing dollar premium/discount records more than 14.6 cents, arbitrageurs begin to take positions on both sides and drive the deviation back into the "no-arbitrage" band.

The convergence speed is measured by δ, defined in Section 4.1. I estimate the smooth transition ECM model using a kernel smoothing estimation method. Panel A of Table 4.3 reports the summary statistics of the estimation of average δ, $\delta_{\kappa<0}$ and $\delta_{\kappa>0}$ over all samples. Panel B of Table 4.3 reports a downward trend of both mean and median of δ, which suggests that NYSE and TSX become more integrated over time. To see how the convergence speed is affected by price deviations, I apply the Wilcoxon signed rank test to test the null hypothesis: $H_0 : \delta_{\kappa<0} \geq \delta_{\kappa>0}$. The p-value is smaller than 0.01; thus I can reject the null hypothesis. In other words, the convergence between two market prices speeds up when there is a negative price premium at NYSE.[8] A possible explanation is that arbitrageurs like to establish short positions in TSX since the stock has better liquidity in the home market.

4.4.2 Estimation of the information share

The information share measures the contribution of each market to the price discovery. I estimate the information share using the three models described in Section 4.3.

The first column of Table 4.4 reports the estimated information share of NYSE from the linear ECM. Eun and Sabherwal (2003) estimate the information share with the same model, but their sample period is shorter (from February to July, 1998). Their estimated information share of the NYSE (IS^n) ranges from

[8] Note that, the smaller δ is, the faster the convergence is.

Table 4.3: Delta estimate of NYSE

Panel A: Statistics Summary			
	Delta	$Delta_{prem}$	$Delta_{disc}$
Mean	0.669	0.688	0.652
St.Dev.	0.105	0.133	0.123
1%-ile	0.495	0.446	0.434
10%-ile	0.55	0.537	0.532
25%-ile	0.585	0.603	0.580
50%-ile	0.654	0.661	0.641
75%-ile	0.734	0.760	0.722
90%-ile	0.827	0.883	0.814
99%-ile	0.897	1.000	0.894

Panel B: Annual estimates						
	Delta		$Delta_{prem}$		$Delta_{disc}$	
	Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
1998	0.709	0.709	0.729	0.724	0.701	0.713
1999	0.653	0.643	0.668	0.644	0.641	0.616
2000	0.650	0.642	0.674	0.616	0.620	0.620

Panel C: Wilcoxon signed rank test		
Hypothesis	Wilcoxon	P-value
$H_0 : Delta_{prem} \geq Delta_{disc}$	1312.0	6.446×10^{-5}
$H_1 : Delta_{prem} < Delta_{disc}$		

0.2% to 98.2%, with an average of 38.1% over their sample. They conclude that price discovery for most cross-listed pairs occurs on the TSX, but there is significant feedback from the NYSE. My results, based on a longer sample period, are consistent with these conclusions: the estimated information share of the NYSE (IS^n) ranges from 1% to 97.5%, with a mean of 40.7%. There is no discernible trend over the sample period as the yearly average estimates of IS^n in

1998, 1999, and 2000 are 39.3%, 48.4%, and 41%, respectively. The linear ECM

Table 4.4: Information Shares of NYSE

Panel A: Summary Statistics						
	Linear ECM	Threshold ECM		Smooth Transition ECM		
	IS	IS_{in}	IS_{out}	IS_{ST}	IS_{prem}	IS_{disc}
Mean	0.430	0.362	0.435	0.374	0.386	0.379
St.Dev.	0.258	0.239	0.259	0.254	0.253	0.264
1%-ile	0.030	0.017	0.02	0.001	0.012	0.003
10%-ile	0.087	0.073	0.106	0.059	0.067	0.068
25%-ile	0.215	0.138	0.215	0.173	0.177	0.161
50%-ile	0.416	0.358	0.418	0.352	0.369	0.360
75%-ile	0.601	0.543	0.626	0.543	0.554	0.536
90%-ile	0.816	0.669	0.804	0.707	0.739	0.797
99%-ile	0.948	0.910	0.980	0.946	0.934	0.981
Panel B: Anuual Estimates						
	Linear ECM	Threshold ECM		Smooth Transition ECM		
	IS	IS_{in}	IS_{out}	IS_{ST}	IS_{prem}	IS_{disc}
1998	0.393	0.367	0.386	0.386	0.413	0.381
1999	0.484	0.368	0.514	0.382	0.378	0.401
2000	0.410	0.352	0.442	0.357	0.350	0.374
Panel C: Wilcoxon signed rank test of smooth transition information share						
Hypothesis		Wilcoxon			P-value	
$H_0 : IS_{prem} \leq IS_{disc}$		2877			0.036	
$H_1 : IS_{prem} > IS_{disc}$						

ignores the nonlinearity of the convergence procedure, as shown in Section 4.2. Thus, the estimation from linear ECM may be biased. Next, I estimate the information share through both threshold ECM (TECM) and smooth transition ECM (STECM).

The second and third columns of Table 4.4 report the results for a bivariate threshold ECM. The estimated information share of the NYSE (IS^n) defined within regimes ranges from 1% to 97.5%, with an average of 38%, while the estimations for the outer regimes range from 2% to 98.5%, with an average of 43.5%. Thus, overall, the NYSE makes a larger contribution to the price discovery in the outer regimes. This may be because arbitrageurs jump into the market when price deviations are very large, and their arbitrage activities can transfer the information from the home market to NYSE (see Fremault 1991).

The last three columns of Table 4.4 report the results from the smooth transition error correction model (STECM). There are three information shares: IS^n, $\text{IS}^n_{\kappa<0}$ and $\text{IS}^n_{\kappa>0}$, which denote the information share defined on whole sample; and on the samples with negative or positive price premium in NYSE. I apply the Wilcoxon signed rank test to examine the null hypothesis $H_0 : Is_{\kappa<0} \leq Is_{\kappa>0}$. The p-value of the test is 0.036, which significantly rejects the null hypothesis. This finding implies that when there is a negative price premium at NYSE, the information share of NYSE is larger, which is evidence that informed traders may choose to trade at the NYSE when it offers a big price discount, even though the home market has better liquidity.

4.5 Regression analysis

This section reports the regression analysis results.

4.5.1 Dataset construction for regression analysis

I construct a panel dataset for regression analyses of the estimates of information shares and thresholds with columns of various indices, dependent variables, explanatory variables, and control variables. *Symbol* is the NYSE ticker of a TSX-NYSE cross-listed pair. *Year* is the year index of an estimated value. *ISLin* is the information share estimate of the NYSE through the linear ECM. *ISIn* and IS^n_{out} are the inner-regime and outer-regime information share estimates of the NYSE from the threshold ECM.

Dependent variables. *ISLin* is the information share estimate of the NYSE through the linear ECM. *ISIn* and *ISOut* are the inner-regime and outer-regime information share estimates of the NYSE from the threshold ECM. *Threshold* is the U.S.$-denominated threshold estimate from threshold ECM. *Delta* is the convergence speed parameter estimated from smooth transition ECM.

Explanatory variables. *PINRat* is the ratio of the PIN of the NYSE over that of the TSX. [9]*PINAvg* is the average PIN of the NYSE and the TSX. *PINDiff* is the PIN of the NYSE minused by that of the TSX. *SpreadRat* is the ratio of the relative quoted bid-ask spread of the NYSE over that of the TSX. *SpreadAvg* is the average relative quoted bid-ask spread of NYSE and the TSX. *SpreadDiff* is the quoted bid-ask spread of the NYSE minused by that of the TSX.[10]

[9]The PINs for TSX- and NYSE-listed Canadian stocks are estimated following Easley, Kiefer, O'Hara, and Paperman (1996) and Easley, Kiefer and O'Hara (1997a,b). Further, I adopt Easley, Engle, O'Hara, and Wu's (2008) log-likelihood function specification for improved numerical stability in computing the PIN.

[10]The bid-ask spreads are adjusted by the mid-quotes and, thus, measure the relative discrepancy between bid and ask quotes free from the exchange rate.

Control variables. USVol is the average daily trading fraction of the NYSE out of both of the NYSE and the TSX following Eun and Sabherwal (2003). *VolAvg* is the average of the log-transformations of average daily trading volume measures of the NYSE and the TSX. *VolDiff* is the difference of the log-transformation of average daily trading volume of the NYSE over that of the TSX. *USDollarVol* is the average daily dollar trading volume of the NYSE out of both of the NYSE and the TSX. *DollarVolAvg* is the sum of log-transformations of average daily dollar trading volume measures of the NYSE and the TSX. *DollarVolDiff* is the difference of the log-transformation of average daily dollar trading volume of the NYSE over that of the TSX. *Governance* is the Report on Business governance index of Canadian firms published by *Globe and Mail* (McFarland 2002). *Indus* equals one if the cross-lister is a manufacturing firm, and zero otherwise. *Size* is the normalized average market capitalization on the TSX and the NYSE.

4.5.2 Regression of the information share

I now conduct a regression analysis on the factors that affect the relative extent of the NYSE's contribution to price discovery. The estimated outer-regime information shares are regressed onto the panel of explanatory and control variables with and without intercept in Panel A and Panel B of Table 4. It turns out that the contribution of the NYSE increases relatively against that of the TSX as the NYSE-based trades become more informative. This is cross-border evidence that informed trades contribute to fostering price discovery, in line with Chen and Choi (2010). Either in quantity or value, the higher the liquidity on the NYSE the more it leads in price discovery. This is consistent with Eun and Sabherwal's (2003) findings: they estimate the information share of the NYSE by using Harris et al.'s (1995, 2002) approach. They find that the information share is directly related to the U.S.'s share of total trading (USVol) as well as to the proportion of informative trades on U.S. exchanges and the TSX, and inversely related to the ratio of bid-ask spreads on U.S. exchanges and the TSX.[11] A Canadian firm which is larger (Size) and offers better investor-protecting (Governance) tends to has more price discovery on the TSX as seen in Panels A and B. The overall explanatory power is significantly higher for models without intercept.

I conduct analogous panel regressions for the inner-regime and linear information shares in Tables 4.5 and 4.6, respectively. Neither alternative measure of exchange-specific contribution to price discovery has a higher explanatory power (adjusted R^2) and statistically significance on regressors. From this perspective, the outer-regime information shares (Table 4.4) have not only proved heuristically appealing but also economically reasonable and statistically robust.

4.5.3 Regression of the estimated threshold

For each cross-listed pair, the threshold includes transactions costs, which consist of bid-ask price spreads on both exchanges and the foreign exchange rate, fixed costs, and liquidity shortfalls. Implicit risk premiums, including those

[11]Hasbrouck (1995) finds a positive and significant correlation between contribution to price discovery made by the NYSE and its market share by trading volume using the U.S. domestic data. Using the same data, Harris et al. (2002) finds evidence that the information share increases when its bid-ask spreads decline relative to the regional exchange.

Table 4.5: Panel regression results of outer-regime information shares

	reg 1	reg 2	reg 3	reg 4	reg 5	reg 6	reg 7	reg 8
Intercept	0.651	0.702	0.632	0.683	0.262	0.307	0.206	0.242
t-stat	5.473	5.961	5.339	5.844	3.531	4.006	2.92	3.282
PinRatio	0.127	0.122	0.133	0.127	0.151	0.136	0.179	0.168
t-stat	2.303	2.156	2.412	2.246	2.661	2.294	3.273	2.938
SpreadRatio	0.001	0.002	0.002	0.002	0.000	−0.001	−0.002	−0.002
t-stat	0.34	0.367	0.572	0.573	−0.093	−0.214	−0.419	−0.555
UsVol	0.386		0.358		0.414		0.454	
t-stat	4.200		3.998		4.600		5.668	
UsDollarVol		0.300		0.277		0.282		0.336
t-stat		3.673		3.486		3.572		4.627
Industry	−0.054	−0.05						
t-stat	−1.282	−1.175						
Governance	−0.005	−0.005	−0.005	−0.005				
t-stat	−3.538	−3.833	−3.717	−3.980				
Size	−0.39	−0.403	−0.353	−0.368	−0.443	−0.473		
t-stat	−2.256	−2.295	−2.063	−2.122	−2.502	−2.585		
Fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Year effect	yes	yes	yes	yes	yes	yes	yes	yes
No.of Obs.	115	115	115	115	115	115	115	115
Adjusted R^2	0.277	0.252	0.273	0.249	0.207	0.154	0.203	0.144

from information asymmetry and macroeconomic uncertainty, can also affect the determination of the threshold. Accordingly, Table 4.8 and 4.9 provide the results of panel regressions of the estimated thresholds onto average (Table 4.8) and difference (Table 4.9) measures of asymmetric information component (PIN) and the inverse of market depth (spread), controlling for liquidity, either in quantity (UsVol) or value (UsDollarVol), firm-level idiosyncratic characteris-

Table 4.6: Panel regression results of inner-regime information shares

	reg 1	reg 2	reg 3	reg 4	reg 5	reg 6	reg 7	reg 8
Intercept	−0.027	0.026	−0.027	−0.026	-0.02	−0.18	−0.022	−0.021
t-stat	−0.649	−0.631	−0.676	−0.661	-1.367	−1.275	−1.672	−1.633
PinRatio	−0.015	−0.016	−0.015	−0.016	−0.007	−0.008	0.008	0.007
t-stat	−0.370	−0.399	−0.372	−0.401	−0.190	−0.221	0.222	0.184
SpreadRatio	0.000	0.000	0.000	0.000	0.000	0.000	−0.001	−0.001
t-stat	−0.129	−0.086	−0.130	−0.087	−0.116	−0.07	−0.213	−0.165
UsVol	0.225		0.225		0.200		0.234	
t-stat	1.516		1.525		1.458		1.886	
UsDollarVol		0.222		0.222		0.213		0.247
t-stat		1.473		1.482		1.467		1.911
Industry	0.000	−0.001						
t-stat	−0.017	−0.034						
Governance	0.000	0.000	0.000	0.000				
t-stat	0.042	0.066	0.042	0.065				
Size	−0.030	−0.036	−0.029	−0.035	−0.033	−0.036		
t-stat	−0.387	−0.471	−0.395	−0.477	−0.464	−0.512		
Fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Year effect	yes	yes	yes	yes	yes	yes	yes	yes
No.of Obs.	115	115	115	115	121	121	131	131
Adjusted R^2	0.015	0.014	0.025	0.023	0.023	0.018	0.044	0.044

tics (Industry, Governance, and Size), and interest rates (yields of 90-day bills and 10-year notes).

As expected, my measure of market friction (relative quoted spread) significantly increases required dollar return of cross-border arbitrage as 8 out of 16 models using average measures (Table 4.8) and all models using difference measures (Table 4.9) agree with it. The better the firm is governed at home, the

Table 4.7: Panel regression results of linear information shares

	reg 1	reg 2	reg 3	reg 4	reg 5	reg 6	reg 7	reg 8
Intercept	0.015	0.014	0.013	0.011	0.014	0.014	0.019	0.019
t-stat	0.257	0.234	0.219	0.190	0.693	0.669	1.006	1.045
PinRatio	0.049	0.055	0.049	0.055	0.052	0.057	0.065	0.071
t-stat	0.868	0.974	0.871	0.978	0.952	1.051	1.258	1.370
SpreadRatio	0.002	0.001	0.002	0.001	0.002	0.001	0.001	0.001
t-stat	0.346	0.248	0.344	0.245	0.353	0.248	0.324	0.248
UsVol	−0.153		−0.151		−0.128		-0.034	
t-stat	0.716		−0.712		−0.658		0.192	
UsDollarVol		-0.35		-0.348		−0.330		-0.189
t-stat		−1.626		−1.672		−1.618		−1.024
Industry	-0.004	−0.005						
t-stat	−0.186	−0.211						
Governance	0.000	0.000	0.000	0.000				
t-stat	0.018	0.046	0.014	0.042				
Size	−0.030	−0.045	−0.025	−0.040	−0.019	−0.034		
t-stat	−0.269	−0.412	−0.236	−0.377	−0.191	−0.34		
Fixed effect	yes	yes	yes	yes	yes	yes	yes	yes
Year effect	yes	yes	yes	yes	yes	yes	yes	yes
No.of Obs.	115	115	115	115	121	121	131	131
Adjusted R^2	0.015	0.006	0.005	0.014	0.010	0.029	0.017	0.025

lower the minimum required profit as all models with the Governance control variable show. Manufacturing firms (when Industry equals 1) tend to require larger relative premiums to be exploited. Overall, difference measures turn out to have a greater influence on the threshold level than the average measures. In summary, the effective break-even point (threshold) of cross-border arbitrage appears to be affected by the relative degree of private information, market fric-

Table 4.8: Panel regression for threshold with average measures

	reg 1	reg 2	reg 3	reg 4	reg 5	reg 6	reg 7	reg 8
Intercept	1.275	2.488	1.085	2.591	0.373	1.852	0.625	1.742
	1.381	1.949	1.134	1.949	0.412	1.531	0.766	1.718
PINAvg	-1.419	-2.152	-1.087	-2.082	-0.053	-0.945	-0.410	-1.131
	-0.917	-1.377	-0.678	-1.279	-0.034	-0.611	-0.280	-0.788
SpreadAvg	15.217	11.387	15.419	11.923	3.959	0.782	2.789	-0.214
	2.735	1.681	2.667	1.690	0.861	0.138	0.657	-0.042
VolAvg	0.003		0.032		0.024		0.008	
	0.049		0.531		0.393		0.157	
DollarVolAvg		-0.066		-0.060		-0.067		-0.056
		-0.981		-0.851		-0.981		-1.020
Industry	0.366	0.370						
	3.090	3.180						
Governance	-0.010	-0.011	-0.010	-0.010				
	-2.627	-2.704	-2.487	-2.533				
Size	0.458	0.789	0.013	0.411	-0.290	0.126		
	0.785	1.257	0.022	0.640	-0.487	0.195		
Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	115	115	115	115	121	121	131	131
Adjusted R2	0.118	0.126	0.048	0.052	-0.034	-0.027	-0.029	-0.021

tion, liquidity measures, and idiosyncratic firm-level characteristics. These economically appealing empirical results lend support to the findings of Gagnon and Karolyi (2010).

Table 4.9: Panel regression for threshold with difference measures

	reg 1	reg 2	reg 3	reg 4	reg 5	reg 6	reg 7	reg 8
Intercept	1.031	1.007	1.278	1.268	0.567	0.574	0.589	0.590
	3.743	3.605	4.424	4.363	4.248	4.592	5.064	5.409
PINDiff	-1.553	-1.427	-1.731	-1.462	-1.206	-1.212	-1.067	-1.048
	-1.875	-1.664	-1.947	-1.594	-1.374	-1.348	-1.389	-1.323
SpreadDiff	10.461	9.386	10.091	10.050	9.299	9.115	7.959	8.064
	3.411	2.840	3.064	2.846	2.822	2.630	2.862	2.664
VolDiff	-0.093		-0.051		-0.013		0.002	
	-2.208		-1.164		-0.315		0.064	
DollarVolDiff		-0.065		-0.019		-0.011		0.004
		-1.666		-0.485		-0.296		0.107
Industry	0.495	0.491						
	4.183	4.053						
Governance	-0.013	-0.011	-0.011	-0.010				
	-3.346	-3.060	-2.707	-2.516				
Size	0.194	0.192	-0.170	-0.132	-0.315	-0.318		
	0.389	0.380	-0.323	-0.247	-0.594	-0.595		
Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of Obs.	115	115	115	115	121	121	131	131
Adjusted R2	0.208	0.193	0.086	0.076	0.031	0.031	0.035	0.036

4.6 Conclusion

In this paper, I contribute to the literature by implementing the threshold error correction mechanism in estimating the relative extent of exchange-respective contribution to price discovery of the pairs of cross-listings and their original listings. The existing methods assume linear convergence of relative premiums to parity whereas I hinge on the reality that the premiums disappear quicker when it is profitably arbitrageable than otherwise. An asset pricing equilibrium model for a stock traded in multiple markets has been developed to illustrate the role of arbitrageurs in the price discovery process. Based on the equilibrium solutions from this equilibrium model, under different assumptions on the demand elasticity of arbitrageurs, I show the short term convergence dynamics could be captured by three econometric models: standard linear ECM, threshold ECM, smooth transition ECM, which may provide different estimation on the contribution share for each market to the price discovery. The latter two are more reliable since their assumptions accommodate the nonlinear convergence in the reality.

I apply these three models to Canadian stocks cross-listed in TSX and NYSE. All three models generate a consistent conclusion that the home market (TSX) makes a larger contribution than NYSE (guest market) in the price discovery. However, from the estimations of nonlinear error correction models, I get some other interesting findings. First, there is a larger feedback effect from NYSE on Canadian cross-listed stocks if the price deviations exceed a threshold value. Second, when there exists a negative price premium at NYSE, informed traders tend to trade at NYSE even though the home market usually has better liquidity. Meanwhile, the convergence between two market prices will speed up. Third, information shares are positively affected by the relative degree of private information and market liquidity. Unlike Grammig et al. (2005), I do not account for exchange-rate market friction in my threshold ECM framework. Additional sources of randomness to the modeling of nonlinear dynamics of cross-listed stocks should be interesting for future studies.

APPENDIX A

APPENDIX OF CHAPTER 2

Throughout the Appendix A, let $|A| = (tr(A'A))^{1/2}$ denote the Euclidean norm of a matrix A. Let "$\Rightarrow$ " denote weak convergence with respect to the uniform metric and "$\xrightarrow{p}$ " denote the convergence in probability. A' is denoted as the transpose of the matrix A. The proofs related to the basic model are conducted in Appendix A.1 and those for the extended model are put in Appendix A.2.

A.1 Mathematical proof for the basic model

Proof of Lemma 2.1.1: The result follows Lemma 1 of Park and Hahn (1999).

Proof of Lemma 2.1.2: The result follows Theorem 1 of Caner and Hansen (2001) by replacing u with $F(\gamma)$.

Proof of Lemma 2.1.3: Following Theorem 2 of Caner and Hansen(2001), under Assumptions 2.1.1-2.1.4, I can easily show that

$$\frac{1}{\sigma n}\sum_{t=1}^{n} x_t I_t(\gamma) e_t \Rightarrow \int_0^1 X(s) dW(s,\gamma).$$

Q.E.D.

Lemma A.1.1 *Under Assumptions 2.1.1-2.1.4, for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, as $n \to \infty$, the following results hold:*

a) $n^{-2}\sum_{t=1}^{n} x_t(\gamma)x_t'(\gamma) = F(\gamma)\int_0^1 X(s)X(s)'ds + o_p(1)$,

b) $n^{-1}\sum_{t=1}^{n} x_t(\gamma)e_t \Rightarrow \int_0^1 X(s)dW(s,\gamma)$,

c) $n^{-2}\sum_{t=1}^{n} V_t(\gamma)V_t'(\gamma) = M(\gamma) + o_p(1)$, *d)* $n^{-1}\sum_{t=1}^{n} V_t(\gamma)e_t \Rightarrow \sigma \begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma) \end{pmatrix}$, *where* $M(\gamma)$ *is defined in (2.14).*

Proof: *a*) can be proved using Lemma 1 and Theorem 3 of Caner and Hansen(2001). *b*) is from Lemma 2.1.3. For *c*) and *d*), using *a*) and *b*), I have

$$
\begin{aligned}
n^{-2}\sum_{t=1}^{n} V_t(\gamma)V_t(\gamma)' &= n^{-2}\begin{pmatrix} \sum_{t=1}^{n} x_t x_t', & \sum_{t=1}^{n} x_t x_t'(\gamma) \\ \sum_{t=1}^{n} x_t(\gamma)x_t', & \sum_{t=1}^{n} x_t(\gamma)x_t'(\gamma) \end{pmatrix} \\
&= M(\gamma) + o_p(1) \qquad \text{(A.1)}
\end{aligned}
$$

and

$$
n^{-1}\sum_{t=1}^{n} V_t(\gamma)e_t = n^{-1}\begin{pmatrix} \sum_{t=1}^{n} x_t e_t \\ \sum_{t=1}^{n} x_t(\gamma)e_t \end{pmatrix} \Rightarrow \sigma \begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma) \end{pmatrix}. \qquad \text{(A.2)}
$$

Q.E.D.

Lemma A.1.2 *If* $\tau < 1/2$, *when* $\gamma = \gamma_0$

$$
n(\widehat{\theta}(\gamma_0) - \theta) \Rightarrow \sigma M(\gamma_0)^{-1}\begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma_0) \end{pmatrix},
$$

while $\gamma \neq \gamma_0$

$$
n^{\tau+1/2}(\widehat{\theta}(\gamma) - \theta) \Rightarrow M(\gamma)^{-1}\Pi(\gamma,\gamma_0,\delta_0).
$$

If $\tau = 1/2$, *when* $\gamma = \gamma_0$,

$$
n(\widehat{\theta}(\gamma_0) - \theta) \Rightarrow \sigma M(\gamma_0)^{-1}\begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma_0) \end{pmatrix},
$$

while $\gamma \neq \gamma_0$,

$$
n(\widehat{\theta}(\gamma) - \theta) \Rightarrow \sigma M(\gamma)^{-1}\begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma) \end{pmatrix} + M(\gamma)^{-1}\Pi(\gamma,\gamma_0,\delta_0)
$$

where $\Pi(\gamma,\gamma_0,\delta_0)$ *is defined in the equation (2.15).*

Proof: I first consider the case with $\tau < 1/2$. Using Lemma A.1.1, when $\gamma = \gamma_0$ I have

$$n(\widehat{\theta}(\gamma_0) - \theta) = (\frac{1}{n^2}\sum_{t=1}^{n} V_t(\gamma_0)V_t(\gamma_0)')^{-1}\frac{1}{n}\sum_{t=1}^{n} V_t(\gamma_0)e_t \Rightarrow M(\gamma_0)^{-1}\begin{pmatrix} \int_0^1 X(s)dW(s) \\ \int_0^1 X(s)dW(s,\gamma_0) \end{pmatrix}. \tag{A.3}$$

When $\gamma \neq \gamma_0$, from the true model, I have

$$y_t = \theta' V_t(\gamma_0) + e_t = \theta' V_t(\gamma) + e_t - \theta'(V_t(\gamma) - V_t(\gamma_0)) = \theta' V_t(\gamma) + e_t - \delta_n'(x_t(\gamma) - x_t(\gamma_0)).$$

If $\tau < 1/2$, then

$$\begin{aligned}
& n^{\tau+1/2}(\widehat{\theta}(\gamma) - \theta) \\
= \ & n^{\tau+1/2}(\sum_{t=1}^{n} V_t(\gamma)V_t(\gamma)')^{-1}\{\sum_{t=1}^{n} V_t(\gamma)e_t - \sum_{t=1}^{n} V_t(\gamma)(x_t(\gamma)' - x_t(\gamma_0)')\delta_n\} \\
= \ & O_p(n^{\tau-1/2}) - M(\gamma)^{-1}\begin{pmatrix} n^{-2}\sum_{t=1}^{n} x_t(x_t(\gamma)' - x_t(\gamma_0)')\delta_0 \\ n^{-2}\sum_{t=1}^{n} x_t(\gamma)(x_t(\gamma)' - x_t(\gamma_0)')\delta_0 \end{pmatrix} + o_p(1) \\
\Rightarrow \ & -M(\gamma)^{-1}\begin{pmatrix} (F(\gamma) - F(\gamma_0))\int_0^1 X(s)X'(s)ds \\ (F(\gamma) - F(\gamma_0 \wedge \gamma))\int_0^1 X(s)X'(s)ds \end{pmatrix}\delta_0 \\
= \ & M(\gamma)^{-1}\Pi(\gamma, \gamma_0, \delta_0).
\end{aligned} \tag{A.4}$$

When $\tau = 1/2$, the proof is very similar and I skip the detail. Q.E.D.

Lemma A.1.3 *If $\tau < 1/2$, I have $\widehat{\gamma}_n \xrightarrow{p} \gamma_0$.*

Proof: To prove the consistency of $\widehat{\gamma}_n$, I need to prove $SSR_n(\gamma)$ uniformly converge to a function which takes global minimum at γ_0. It is equivalent to prove $b_n(\gamma) = n^{2\tau-1}(SSR_n - SSR_n(\gamma))$ uniformly converge to a function which takes global maximum at γ_0. SSR_n is defined as the sum of squared residual by regressing y_t to x_t. After some standard algebra, I have

$$b_n(\gamma) = n^{2\tau-1}(SSR_n - SSR_n(\gamma)) = n^{2\tau-1}\widehat{\delta}_n(\gamma)'(X(\gamma)'(I-P_n)X(\gamma))\widehat{\delta}_n(\gamma)$$

where $X(\gamma) = (x_1(\gamma), x_2(\gamma), ..., x_n(\gamma))'$ and $X = (x_1, x_2, ..., x_n)'$. $P_n = X(X'X)^{-1}X'$, is the projection matrix of X. By plugging in

$$\widehat{\delta}_n(\gamma) = (X'(\gamma)(I-P_n)X(\gamma))^{-1}X'(\gamma)(I-P_n)Y,$$

I have

$$\begin{aligned} n^{2\tau-1}(SSR_n - SSR_n(\gamma)) &= n^{2\tau-1}Y'(I-P_n)X(\gamma)(X'(\gamma)(I-P_n)X(\gamma))^{-1}X'(\gamma)(I-P_n)Y \\ &= \Gamma_n(\gamma)'(n^{-2}X'(\gamma)(I-P_n)X(\gamma))^{-1}\Gamma_n(\gamma) \end{aligned}$$

where

$$\Gamma_n(\gamma) = n^{\tau-3/2}X'(\gamma)(I-P_n)Y.$$

Using Lemma A.1.1, I can show that

$$n^{-2}X'(\gamma)(I-P_n)X(\gamma) \Rightarrow \left(F(\gamma) - F(\gamma)^2\right)\int_0^1 X(s)X(s)'ds. \tag{A.5}$$

Next, I discuss the limiting behavior of $\Gamma_n(\gamma)$. By plugging in the true model $Y = X\alpha + X(\gamma_0)\delta_n + e$, I have

$$\begin{aligned} \Gamma_n(\gamma) &= n^{\tau-3/2}X'(\gamma)(I-P_n)\left(X\alpha + X(\gamma_0)\delta_n + e\right) \\ &= n^{\tau-3/2}X'(\gamma)X(\gamma_0)\delta_n - n^{\tau-3/2}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n + n^{\tau-3/2}X'(\gamma)(I-P_n)e, \end{aligned}$$

where the second equation uses the result that $(I-P_n)X = 0$. Since $\tau < 1/2$, I can show

$$\begin{aligned} n^{\tau-3/2}X'(\gamma)(I-P_n)e &= n^{\tau-1/2}n^{-1}X'(\gamma)e - n^{\tau-1/2}n^{-2}X'(\gamma)X(n^{-2}X'X)^{-1}n^{-1}X'e \\ &= O_p(n^{\tau-1/2}) = o_p(1). \end{aligned}$$

Thus,

$$\Gamma_n(\gamma) \Rightarrow F(\gamma \wedge \gamma_0)\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 - F(\gamma)F(\gamma_0)\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 \equiv \Gamma_2(\gamma). \tag{A.6}$$

Define

$$b(\gamma) = \Gamma_2(\gamma)'\left(\left(F(\gamma) - F(\gamma)^2\right)\int_0^1 X(s)X(s)'ds\right)^{-1}\Gamma_2(\gamma).$$

For any $\gamma \geq \gamma_0$,

$$b(\gamma) = \frac{F(\gamma_0) - F(\gamma)F(\gamma_0)}{(F(\gamma) - F(\gamma)^2)}\delta_0'\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 = \frac{F(\gamma_0)}{F(\gamma)}\delta_0'\left(\int_0^1 X(s)X(s)'ds\right)\delta_0$$

with $\frac{\partial b(\gamma)}{\partial \gamma} < 0$. For any $\gamma \leq \gamma_0$,

$$b(\gamma) = \frac{F(\gamma) - F(\gamma)F(\gamma_0)}{(F(\gamma) - F(\gamma)^2)}\delta_0'\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 = \frac{1 - F(\gamma_0)}{1 - F(\gamma)}\delta_0'\left(\int_0^1 X(s)X(s)'ds\right)\delta_0$$

with $\frac{\partial b(\gamma)}{\partial \gamma} > 0$. Thus, $b(\gamma)$ takes global maximum at $\gamma = \gamma_0$. Using Lemma A.1.1, I can prove that

$$\sup_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} (b_n(\gamma) - b(\gamma)) = o_p(1).$$

In summary, I have

$$\widehat{\gamma}_n = \arg\min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} (SSR_n(\gamma)) = \arg\max_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} (n^{2\tau - 1}(SSR_n - SSR_n(\gamma)) \xrightarrow{p} \gamma_0.$$

Q.E.D.

Lemma A.1.4 *If $\tau < 1/2$, I have $a_n(\widehat{\gamma}_n - \gamma_0) = O_p(1)$, where $a_n = n^{1-2\tau}$.*

Proof: To prove $\widehat{\gamma}_n$ converge to γ_0 with rate a_n, I need to prove that $a_n|\widehat{\gamma}_n - \gamma_0| = O_p(1)$, or for any $\bar{v} > 0$, $\lim_{n\to\infty} \Pr(|\widehat{\gamma}_n - \gamma_0| \leq \bar{v}/a_n) = 1$. For each $B > 0$, define $V_B = \{\gamma : |\gamma - \gamma_0| < B\}$. When n is large enough, I have $\bar{v}/a_n < B$. Since $\widehat{\gamma}_n \xrightarrow{p} \gamma_0$

according to Lemma A.1.3, $\Pr(\{\widehat{\gamma}_n \in V_B\}) \xrightarrow{p} 1$. Therefore, I only need to examine the limiting behavior of γ in V_B.

Define a subset

$$V_B(\overline{v}) = \{\gamma : \overline{v}/a_n < |\gamma - \gamma_0| < B.$$

Thus, $V_B(\overline{v}) \subset V_B$. To prove $\Pr(|\widehat{\gamma}_n - \gamma_0| \leq \overline{v}/a_n) = 1$, I just need to prove $\Pr(\widehat{\gamma}_n \in V_B(\overline{v})) = 0$. Let $\widehat{\theta}$ and $\widehat{\delta}$ as the estimation of $\widehat{\theta}(\widehat{\gamma}_n)$ and $\widehat{\delta}(\widehat{\gamma}_n)$. Define $SSR_n^*(\gamma) = \sum_{t=1}^n (y_t - \widehat{\theta}' V_t(\gamma))^2$ and $SSR_n^*(\gamma_0) = \sum_{t=1}^n (y_t - \widehat{\theta}' V_t(\gamma_0))^2$. From the definition of $\widehat{\gamma}_n$, I have $SSR_n^*(\widehat{\gamma}_n) \leq SSR_n^*(\gamma_0)$. Therefore, it suffices to prove that for any $\gamma \in V_B(\overline{v})$, $SSR_n^*(\gamma) > SSR_n^*(\gamma^0)$ with probability 1.

We consider the case of $\gamma > \gamma^0$ at first. Using an argument of symmetry, I can, without loss of generality, prove the result for the case of $\gamma < \gamma_0$. Given $\gamma > \gamma^0$, it is equivalent to prove

$$\frac{SSR_n^*(\gamma) - SSR_n^*(\gamma_0)}{a_n(\gamma - \gamma_0)} > 0.$$

Note that

$$\begin{aligned}
& SSR_n^*(\gamma) - SSR_n^*(\gamma_0) \\
= & \sum_{t=1}^n (y_t - \widehat{\theta}' V_t(\gamma))^2 - \sum_{t=1}^n (y_t - \widehat{\theta}' V_t(\gamma_0))^2 \\
= & \sum_{t=1}^n \widehat{\delta}'(x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))'\widehat{\delta} - 2\sum_{t=1}^n \widehat{\delta}'(x_t(\gamma) - x_t(\gamma_0))e \\
& + 2\widehat{\delta}'(x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))'(\widehat{\theta} - \theta) \\
= & \sum_{t=1}^n \delta_n'(x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))'\delta_n - 2\widehat{\delta}' \sum_{t=1}^n (x_t(\gamma) - x_t(\gamma_0))e \\
& + 2\widehat{\delta}' \sum_{t=1}^n (x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))'(\widehat{\theta} - \theta) \\
& + 2\sum_{t=1}^n (\widehat{\delta} + \delta_n)'(x_t(\gamma) - x_t(\gamma_0))(x_t(\gamma) - x_t(\gamma_0))'(\widehat{\delta} - \delta_n) \\
\equiv & R_1 - R_2 + R_3 + R_4, \text{say}.
\end{aligned}$$

Next, I will show that $\frac{R_1+R_2+R_3+R_4}{a_n(\gamma-\gamma_0)}$ converge to a positive random variable almost surely. First, I have

$$\begin{aligned}\frac{R_1}{a_n} &= \frac{1}{a_n}\sum_{t=1}^{n}\delta_n'(x_t(\gamma)-x_t(\gamma_0))(x_t(\gamma)-x_t(\gamma_0))'\delta_n \\ &= \delta_0'(F(\gamma)-F(\gamma_0))\int_0^1 X(s)X'(s)ds\delta_0 + o_p(1) \\ &= f(\gamma_0)(\gamma-\gamma_0)\int_0^1 X(s)X'(s)ds\delta_0 + o_p(1),\end{aligned}$$

where the last equation uses the first order Taylor expansion of $F(\gamma)$ around γ_0.

Noting that $\bar{v}/a_n < |\gamma-\gamma_0| < B$ and $a_n = n^{1-2\tau}$ with $\tau < 1/2$, I have $\sqrt{\bar{v}} < \sqrt{a_n}\sqrt{(|\gamma-\gamma_0|)}$. Thus, there exists $k > 0$, such that

$$\frac{R_2}{a_n(\gamma-\gamma_0)} = \frac{2\widehat{\delta}_0'\frac{1}{n}\sum_{t=1}^{n}(x_t(\gamma)-x_t(\gamma_0))e}{\sqrt{a_n}(\gamma-\gamma_0)} = O_p(\frac{1}{\sqrt{a_n}\sqrt{(|\gamma-\gamma_0|)}}) \leq k/\sqrt{\bar{v}}.$$

Furthermore, from Lemma A.1.2, I know $n^{\tau+1/2}((\widehat{\theta}-\theta) = O_p(\widehat{\gamma}_n-\gamma_0)$ and $n^{\tau+1/2}(\widehat{\delta}_n-\delta_n) = O_p(\widehat{\gamma}_n-\gamma_0)$. Hence I can show:

$$\frac{R_3}{a_n(\gamma-\gamma_0)} = \frac{2n^{\tau+1/2}\widehat{\delta}_n'n^{-2}\sum_{t=1}^{n}(x_t(\gamma)-x_t(\gamma_0))(x_t(\gamma)-x_t(\gamma_0))'n^{\tau+1/2}(\widehat{\theta}-\theta)}{(\gamma-\gamma_0)} = O_p(\widehat{\gamma}_n-\gamma_0).$$

$$\begin{aligned}\frac{R_4}{a_n(\gamma-\gamma_0)} &= \frac{2n^{\tau+1/2}(\widehat{\delta}+\delta_n)'n^{-2}\sum_{t=1}^{n}(x_t(\gamma)-x_t(\gamma_0))(x_t(\gamma)-x_t(\gamma_0))'n^{\tau+1/2}(\widehat{\delta}-\delta_n)}{(\gamma-\gamma_0)} \\ &= O_p(n^{\tau+1/2}(\widehat{\delta}-\delta_n) = O_p(\widehat{\gamma}_n-\gamma_0).\end{aligned}$$

For any $B \to 0_+$, there exist $\bar{v} > 0$ and N, such that $k/\sqrt{\bar{v}} < f(\gamma_0)\int_0^1 X(s)X'(s)ds\delta_0$ and $\bar{v}/a_n < B$ when $n > N$. Therefore, for any $\gamma \in V_B(\bar{v})$, I have

$$\frac{R_1}{a_n(\gamma-\gamma_0)} - \frac{R_2}{a_n(\gamma-\gamma_0)} > 0, \tag{A.7}$$

and

$$\frac{R_3}{a_n(\gamma-\gamma_0)} = o_p(1), \tag{A.8}$$

$$\frac{R_4}{a_n(\gamma-\gamma_0)} = o_p(1). \tag{A.9}$$

Combining A.7-A.9, I can show that

$$\frac{SSR_n^*(\gamma) - SSR_n^*(\gamma_0)}{a_n(\gamma - \gamma_0)} > 0$$

with probability 1 for any $\gamma \in V_B(\overline{v})$ and $\gamma > \gamma^0$. Similarly, I can prove $SSR_n^*(\gamma) > SSR_n^*(\gamma^0)$ when $\gamma < \gamma^0$ and $\gamma \in V_B(\overline{v})$ with probability 1. Q.E.D.

Lemma A.1.5 *If $\tau < 1/2$, I have*

$$n^{1-2\tau}\lambda(\widehat{\gamma}_n - \gamma_0) = r^* \Rightarrow \arg\max_{r\in(-\infty,\infty)}(\Lambda(r) - \frac{1}{2}|r|),$$

where λ and $\Lambda(r)$ is defined in the equation (A.6) and (A.7).

Proof: From Lemma A.1.4, I know that $\widehat{\gamma}$ is a consistent estimator with convergence rate $a_n = n^{1-2\tau}$, thus, I can study its asymptotic behavior in the neighborhood of the true thresholds. Let $\gamma = \gamma_0 + \frac{v}{a_n}$. By the definition of $\widehat{\gamma}_n$,

$$a_n(\widehat{\gamma}_n - \gamma_0) = v^* = \arg\min_v \left(SSR_n^*(\gamma_0 + \frac{v}{a_n}) - SSR_n^*(\gamma_0)\right).$$

By the definition of $SSR_n^*(\gamma_0 + \frac{v}{a_n})$ and $SSR_n^*(\gamma_0)$, I have

$$\begin{aligned}
& SSR_n^*(\gamma_0 + \frac{v}{a_n}) - SSR_n^*(\gamma_0) \\
= & \sum_{t=1}^n \left(y_t - \widehat{\theta}' V_t(\gamma_0 + \frac{v}{a_n})\right)^2 - \sum_{t=1}^n \left(y_t - \widehat{\theta}' V_t(\gamma_0)\right)^2 \\
= & \delta_n' \sum_{t=1}^n \left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right)\left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right)' \delta_n - 2\delta_n' \sum_{t=1}^n \left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right) e \\
& + 2\widehat{\delta}_n' \sum_{t=1}^n \left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right)\left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right)' (\widehat{\theta} - \theta) \\
& + 2\sum_{t=1}^n (\widehat{\delta}' + \delta_n') \left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right)\left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right)' (\widehat{\delta} - \delta_n) \\
& + 2(\widehat{\delta}' - \delta_n') \sum_{t=1}^n \left(x_t(\gamma_0 + \frac{v}{a_n}) - x_t(\gamma_0)\right) e \\
\equiv & R_1^* + R_2^* + R_3^* + R_4^* + R_5^*, \text{ say.}
\end{aligned}$$

Next, I turn to consider the limiting behavior of R_i^*, for $i = 1, 2, ..., 5$. We only provide the proof for the case with $v > 0$, and the proof for the other case with $v < 0$ is analogous so I skip the detail.

Given $v > 0$, I have

$$\begin{aligned} R_1^* &= \delta_n' \sum_{t=1}^{n} \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) \right) \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) \right)' \delta_n \\ &= n^{1-2\tau} \delta_0' n^{-2} \sum_{t=1}^{n} \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) \right) \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) \right)' \delta_0 \\ &= n^{1-2\tau} \delta_0' (F(\gamma_0 + \frac{\upsilon}{a_n}) - F(\gamma_0)) \int_0^1 X(s) X'(s) ds \delta_0 + o_p(1) \\ &\xrightarrow{p} f(\gamma_0) v \delta_0' \int_0^1 X(s) X'(s) ds \delta_0 + o(1). \end{aligned} \tag{A.10}$$

The last equation uses the first order Taylor expansion of $F(\gamma_0 + \frac{\upsilon}{a_n})$ around γ_0.

For R_2^*, I have

$$\begin{aligned} R_2^* &= -2 \sum_{t=1}^{n} \widehat{\delta}_n' \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) \right) e \\ &= -2(n^{1/2-\tau}) \delta_0' \frac{1}{n} \sum_{t=1}^{n} \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) \right) e \Rightarrow -2\delta_0' B^*(\upsilon) \end{aligned}$$

where

$$E\left(B^*(1)\, B^*(1)'\right) = f_0 \sigma^2 \int_0^1 X(s) X'(s) ds.$$

From Lemma A.1.2, I know $n^{\tau+1/2}((\widehat{\theta} - \theta) = O_p(\widehat{\gamma} - \gamma_0) = o_p(1)$ and $n^{\tau+1/2}(\widehat{\delta}_n - \delta_n) = O_p(\widehat{\gamma} - \gamma_0) = o_p(1)$; thus, I can show$R_3^* + R_4^* + R_5^* = o_p(1)$. Combining all convergence results, I have

$$SSR_n^*(\gamma) - SSR_n^*(\gamma_0) \Rightarrow f_0 \upsilon \delta_0' \int_0^1 X(s) X'(s) ds \delta_0 - 2\delta_0' B^*(\upsilon).$$

Making the change-of-variables

$$\upsilon = \frac{\sigma^2}{\delta_0' \int_0^1 X(s) X'(s) ds \delta_0 f_0} r,$$

I have

$$SSR_n^*(\gamma) - SSR_n^*(\gamma_0) \Rightarrow 2\sigma^2(\frac{r}{2} - \Lambda_2(r))$$

where $\Lambda_2(r)$ is a standard Brownian motions defined on $[0, \infty)$.

In summary, the asymptotic distribution of $\widehat{\gamma}$ can be expressed as

$$n^{1-2\tau}\lambda(\widehat{\gamma} - \gamma_0) = r^* \Rightarrow \arg\max_{r\in(-\infty,\infty)}(\Lambda(r) - \frac{1}{2}|r|)$$

where

$$\lambda = \frac{\left(\delta_0' \int_0^1 X(s)X(s)'ds\delta_0\right) f_0}{\sigma^2},$$

and

$$\Lambda(r) = \begin{cases} \Lambda_1(-r), & \text{if } r < 0 \\ 0, & \text{if } r = 0 \\ \Lambda_2(r), & \text{if } r > 0 \end{cases}.$$

Q.E.D.

Lemma A.1.6 *When $\tau = 1/2, \widehat{\gamma}_n \Rightarrow \gamma(\gamma_0, \delta_0)$ which is a random variable maximizing $Q(\gamma, \gamma_0, \delta_0)$. $Q(\gamma, \gamma_0, \delta_0)$ is defined in the equation (A.8).*

Proof: From the definition of $\widehat{\gamma}_n$, I have

$$\widehat{\gamma}_n = \arg\min_{\gamma\in[\underline{\gamma},\overline{\gamma}]} SSR_n(\gamma) = \arg\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} (SSR_n - SSR_n(\gamma))$$

where $SSR_n - SSR_n(\gamma) = \Gamma_n(\gamma)'(n^{-2}X'(\gamma)(I - P_n)X(\gamma))^{-1}\Gamma_n(\gamma)$ with $\Gamma_n(\gamma) = n^{-1}X'(\gamma)(I - P_n)Y$. It follows that

$$\begin{aligned} \Gamma_n(\gamma) &= n^{-1}X'(\gamma)(I - P_n)(X(\gamma_0)\delta_n + e) \\ &= n^{-1}X'(\gamma)X(\gamma_0)\delta_n - n^{-1}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n + n^{-1}X'(\gamma)(I - P_n)e. \end{aligned}$$

Based on Lemma A.1.2, I have

$$
\begin{aligned}
n^{-1}X'(\gamma)(I-P_n)e &= n^{-1}X'(\gamma)e - n^{-2}X'(\gamma)X(n^{-2}X'X)^{-1}n^{-1}X'e \\
&\Rightarrow \sigma\int_0^1 X(s)d\left(W(s,\gamma)-F(\gamma)W(s)\right) = \Gamma(\gamma).
\end{aligned}
$$

Under Assumption 2.2.1 and $\tau = 1/2$, I have $n\delta_n = \delta_0$ and

$$
\begin{aligned}
& n^{-1}X'(\gamma)X(\gamma_0)\delta_n - n^{-1}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n \\
\Rightarrow\ & (F(\gamma\wedge\gamma_0)-F(\gamma)F(\gamma_0))\left(\int_0^1 X(s)X(s)'ds\right)\delta_0.
\end{aligned}
$$

Thus,

$$
\Gamma_n(\gamma)\Rightarrow\Gamma(\gamma)+(F(\gamma\wedge\gamma_0)-F(\gamma)F(\gamma_0))\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 = \Gamma_1(\gamma).
$$

Moreover, I conclude

$$
n^{2\tau-1}(SSR_n - SSR_n(\gamma))\Rightarrow\Gamma_1(\gamma)\left(\left(F(\gamma)-F(\gamma)^2\right)\int_0^1 X(s)X(s)'ds\right)^{-1}\Gamma_1(\gamma)' = Q(\gamma,\gamma_0,\delta_0).
$$

Q.E.D.

Proof of Theorem 2.2.1: Combining the results from Lemma A.1.3-A.1.6, I complete the proof. Q.E.D.

Proof of Theorem 2.2.2: If $\tau < 1/2$, from Lemma A.1.4, I know $n^{1-2\tau}(\widehat{\gamma}_n-\gamma_0) = O_p(1)$. In the following, I show that the $\widehat{\theta(\widehat{\gamma}_n)}$ and $\widehat{\theta(\gamma_0)}$ are asymptotically equivalent and then I can treat γ_0 as known when I derive the asymptotic distribution for $\widehat{\theta(\widehat{\gamma}_n)}$. Note that

$$
\begin{aligned}
& n(\widehat{\theta(\widehat{\gamma}_n)}-\widehat{\theta(\gamma_0)}) \\
=\ & n(\widehat{\theta(\widehat{\gamma}_n)}-\theta) - n(\widehat{\theta(\gamma_0)}-\theta)) \\
=\ & (n^{-2}\sum_{t=1}^n V_t(\widehat{\gamma}_n)V_t(\widehat{\gamma}_n)')^{-1}n^{-1}\sum_{t=1}^n V_t(\widehat{\gamma}_n)e_t \\
& -(n^{-2}\sum_{t=1}^n V_t(\gamma_0)V_t(\gamma_0)')^{-1}n^{-1}\sum_{t=1}^n V_t(\gamma_0)e_t + O_p(n^{\tau-1/2}).
\end{aligned}
$$

From Lemma A.1.1, I have

$$n^{-2}\sum_{t=1}^{n}V_t(\widehat{\gamma}_n)V_t(\widehat{\gamma}_n)' - n^{-2}\sum_{t=1}^{n}V_t(\gamma_0)V_t(\gamma_0)' \xrightarrow{p} O_p(\widehat{\gamma}_n-\gamma_0) = o_p(1)$$

and

$$n^{-1}\sum_{t=1}^{n}V_t(\widehat{\gamma}_n)e_t - n^{-1}\sum_{t=1}^{n}V_t(\gamma_0)e_t = n^{-1}\sum_{t=1}^{n}(V_t(\widehat{\gamma}_n)-V_t(\gamma_0))e_t = O_p(\sqrt{|\widehat{\gamma}_n-\gamma_0|}) = o_p(1).$$

Thus, I can show that

$$n(\widehat{\theta}(\widehat{\gamma}_n)-\theta) = n(\widehat{\theta}(\gamma_0)-\theta)) + o_p(1) \Rightarrow M(\gamma_0)^{-1}\sigma\begin{pmatrix}\int_0^1 X(s)dW(s)\\ \int_0^1 X(s)dW(s,\gamma_0)\end{pmatrix}.$$

Since

$$Var\left\{\sigma\begin{pmatrix}\int_0^1 X(s)dW(s)\\ \int_0^1 X(s)dW(s,\gamma_0)\end{pmatrix}\right\} = \sigma^2 M(\gamma_0)$$

and $W(s)$ and $W(s,\gamma_0)$ are Brownian motions independent of $X(s)$, I have $n(\widehat{\theta}(\widehat{\gamma}_n)-\theta)$ converges to a mixed normal distribution with variance $\sigma^2 M(\gamma_0)^{-1}$.

If $\tau = 1/2$, from Lemma A.1.2 and Lemma A.1.6, I have

$$n(\widehat{\theta}(\widehat{\gamma}_n)-\theta) \Rightarrow \sigma M(\widehat{\gamma}_n)^{-1}\begin{pmatrix}\int_0^1 X(s)dW(s)\\ \int_0^1 X(s)dW(s,\widehat{\gamma}_n)\end{pmatrix} + M(\widehat{\gamma}_n)^{-1}\Pi(\widehat{\gamma}_n,\gamma_0,\delta_0)$$

where$\widehat{\gamma}_n \Rightarrow \gamma(\gamma_0,\delta_0) = \arg\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]} Q(\gamma,\gamma_0,\delta_0)$. Q.E.D.

Proof of Theorem 2.2.3:

$$\begin{aligned} T_n(\gamma) &= \widehat{\delta}_n(\gamma)'(X(\gamma)(I-P(\gamma))X(\gamma))\widehat{\delta}_n(\gamma)/\widehat{\sigma}^2 \\ &= (I-P_n)Y'X(\gamma)(X'(\gamma)(I-P_n)X(\gamma))^{-1}X'(\gamma)(I-P_n)Y/\widehat{\sigma}^2 \\ &= \Gamma_n(\gamma)'\left(n^{-2}X(\gamma)'X(\gamma) - n^{-2}X(\gamma)'X(X'X)^{-1}XX(\gamma)\right)^{-1}\Gamma_n(\gamma)/\widehat{\sigma}^2. \quad \text{(A.11)}\end{aligned}$$

Using Lemma A.1.1, I have

$$n^{-2}X(\gamma)'X(\gamma) - n^{-2}X(\gamma)'X(X'X)^{-1}XX(\gamma) \Rightarrow \left(F(\gamma)-F(\gamma)^2\right)\int_0^1 X(s)X'(s)ds$$

and

$$\begin{aligned}
\Gamma_n(\gamma) &= \frac{1}{n}X'(\gamma)(I-P_n)Y = \frac{1}{n}X'(\gamma)e - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'e \\
&\Rightarrow \sigma\int_0^1 X(s)dW(s,\gamma) - F(\gamma)\int_0^1 X(s)dW(s) \\
&= \sigma\int_0^1 X(s)d\left(W(s,\gamma)-F(\gamma)W(s)\right) = \Gamma(\gamma).
\end{aligned}$$

Thus,

$$T_n(\gamma) \Rightarrow \frac{1}{\sigma^2}\Gamma(\gamma)'\left(\left(F(\gamma)-F(\gamma)^2\right)\int_0^1 X(s)X'(s)ds\right)^{-1}\Gamma(\gamma).$$

Q.E.D.

Proof of Theorem 2.2.4: Following the equation (A.11), I only need to consider the limiting result for $\Gamma_n(\gamma)$ for different τ. When $\tau < 1/2$, I have

$$\begin{aligned}
\Gamma_n(\gamma) &= \frac{1}{n}X'(\gamma)(I-P_n)Y \\
&= \frac{1}{n}X'(\gamma)e - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'e + \frac{1}{n}X'(\gamma)X(\gamma_0)\delta_n - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n \\
&\Rightarrow \sigma\int_0^1 X(s)dW(s,\gamma) - F(\gamma)\int_0^1 X(s)dW(s) \\
&\quad + n^{1/2-\tau}(F(\gamma\wedge\gamma_0) - F(\gamma)F(\gamma_0))\int_0^1 X(s)X(s)'ds\delta_0 \\
&= O_p(n^{1/2-\tau}) \xrightarrow{p} \infty.
\end{aligned}$$

It follows that $T_n(\gamma) = O_p(n^{1-2\tau}) \xrightarrow{p} \infty$ and power converges to 1.

When $\tau = 1/2$, I have

$$\begin{aligned}
\Gamma_n(\gamma) &= \frac{1}{n}X'(\gamma)e - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'e + \frac{1}{n}X'(\gamma)X(\gamma_0)\delta_n - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n \\
&\Rightarrow \Gamma(\gamma) + (F(\gamma\wedge\gamma_0) - F(\gamma)F(\gamma_0))\left(\int_0^1 X(s)X(s)'ds\right)\delta_0 = \Gamma_1(\gamma).
\end{aligned}$$

It follows that

$$T_n(\gamma) \Rightarrow \frac{1}{\sigma^2}\Gamma_1(\gamma)\left(\left(F(\gamma)-F(\gamma)^2\right)\int_0^1 X(s)X'(s)ds\right)^{-1}\Gamma_1(\gamma)'.$$

When $\tau > 1/2$, I have

$$\begin{aligned}\Gamma_n(\gamma) &= \frac{1}{n}X'(\gamma)e - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'e + \frac{1}{n}X'(\gamma)X(\gamma_0)\delta_n - \frac{1}{n}X'(\gamma)X(X'X)^{-1}X'X(\gamma_0)\delta_n \\ &\Rightarrow \sigma\int_0^1 X(s)d\left(W(s,\gamma) - F(\gamma)W(s)\right) = \Gamma(\gamma).\end{aligned}$$

It follows that $T_n(\gamma) \Rightarrow T$. Q.E.D.

Proof of Theorem 2.2.5: To prove the equality, I only need to prove the following two inequalities: $AsySZ_\gamma(a) \geq 1-a$ and $AsySZ_\gamma(a) \leq 1-a$ hold simultaneously. We first consider the proof of $AsySZ_\gamma(a) \geq 1-a$. By the definition of $AsySZ_\gamma(a)$, I can find a parameter sequence (θ_n, γ_n) such that

$$AsySZ_\gamma(a) = \liminf_{n\to\infty} \Pr_{(\theta_n,\gamma_n)}(\gamma_n \in CI_{\gamma,n}(a)).$$

Let $\{b_n\}$ be a subsequence of $\{n\}$ such that

$$AsySZ_\gamma(a) = \lim_{n\to\infty} \Pr_{(\theta_{b_n},\gamma_{b_n})}(\gamma_{b_n} \in CI_{\gamma,b_n}(a)).$$

Define D_n to be a weight matrix such that $D_n\theta_n = (\alpha, n\delta_n)$. Because the Euclidean space is complete, I can find a subsequence $\{c_n\}$ of $\{b_n\}$ such that $(D_{c_n}\theta_{c_n}, \gamma_{c_n}) \to (\theta_0, \gamma_0)$, where $\theta_0 = (\alpha_0, \delta_0)$ with $\alpha_0 \in R$ and $\delta_0 \in R \cup \{-\infty, \infty\}$; $\gamma_0 \in [\underline{\gamma}, \overline{\gamma}]$.

If $\tau = 1/2$, $\delta_0 \in R$. By Theorem 2.2.4, I have $T_n = O_p(1) < \kappa_n$ with probability one. Thus, $CI_{\gamma,n}(a) = CI^W_{\gamma,n}(a)$, or

$$\begin{aligned}AsySZ_\gamma(a) &= \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(\gamma_{c_n} \in CI_{\gamma,c_n}(a)). = \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(|\widehat{\gamma}_{c_n} - \gamma_{c_n}| \leq \widehat{q}^W_{\gamma,1-a}) \\ &\geq \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(|\widehat{\gamma}_{c_n} - \gamma_{c_n}| \leq \widehat{q}^W_{\gamma,1-a}(\gamma_0,\delta_0)) = 1-a \qquad \text{(A.12)}\end{aligned}$$

The inequality uses the fact that $\widehat{q}^W_{\gamma,1-a} = \sup_{\gamma\in[\underline{\gamma},\overline{\gamma}]} \sup_{\delta\in R} q^W_{\gamma,1-a}(\gamma,\delta)$. The last equation uses the fact that $|\widehat{\gamma}_{c_n} - \gamma_{c_n}|$ will converge to $|\widehat{\gamma}(\gamma_0,\delta_0) - \gamma_0|$ and $\widehat{q}^W_{\gamma,1-a}(\gamma_0,\delta_0)$ is defined as the $(1-a)$ quantile of the limiting distribution of $|\widehat{\gamma}(\gamma_0,\delta_0) - \gamma_0|$.

Notice that $\kappa_n^{-1/2} + n^{\nu}\kappa_n^{1/2} \to 0$, for any $\nu > 0$. If $\tau < 1/2$,I have $T_n = O_p(n^{1-2\tau}) > \kappa_n$ with probability approaching one. Thus,

$$\begin{aligned} AsySZ_{\gamma}(a) &= \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(\gamma_{c_n} \in CI_{\gamma,c_n}(a)) = \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(\gamma_{c_n} \in CI^I_{\gamma,c_n}(a)) \\ &= \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(LR(\gamma_{c_n}, \widehat{\gamma}_{c_n}, \widehat{\theta}_{c_n}) \leq q^I_{1-\alpha}). \end{aligned}$$

By Theorem 2.2.4, I have $n^{1-2\tau}\lambda(\widehat{\gamma}_n - \gamma_n) = r^* \Rightarrow \arg\max_{r\in(-\infty,\infty)}(\Lambda(r) - \frac{1}{2}|r|)$. Then, $LR_n(\gamma_n, \widehat{\gamma}_n, \widehat{\theta}_n) \Rightarrow \max_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$. $LR_{c_n}(\gamma_{c_n}, \widehat{\gamma}_{c_n}, \widehat{\theta}_{c_n})$ converges to $LR_{c_n}(\gamma_0, \widehat{\gamma}_{c_n}, \widehat{\theta}_{c_n})$ and $q^I_{1-\alpha}$ is the $1 - a$ quantile of $LR_{c_n}(\gamma_0, \widehat{\gamma}_{c_n}, \widehat{\theta}_{c_n})$. Thus, I can conclude that

$$AsySZ_{\gamma}(a) = \lim_{n\to\infty} \Pr_{(\theta_{c_n},\gamma_{c_n})}(LR(\gamma_{c_n}, \widehat{\gamma}_{c_n}, \widehat{\theta}_{c_n}) \leq q^I_{1-\alpha}) \geq 1 - a. \tag{A.13}$$

Next, I consider the other side $AsySZ_{\gamma}(a) \leq 1 - a$. Let $\delta_n = \delta_0$ and $\gamma_n = \gamma_0$ with $\delta_0 \in R/\{0\}$. By definition, I have

$$AsySZ_{\gamma}(a) = \liminf_{n\to\infty} \Pr_{(\theta_n,\gamma_n)}(\gamma_n \in CI_{\gamma,n}(a)) \leq \liminf_{n\to\infty} \Pr_{(\theta_0,\gamma_0)}(\gamma_0 \in CI_{\gamma,n}(a)). \tag{A.14}$$

Because δ_n is a fixed constant, I have $T_n = O_p(n^{1-2\tau}) = O_p(n) > \kappa_n$ with probability approaching one. Thus,

$$\begin{aligned} \liminf_{n\to\infty} \Pr_{(\theta_0,\gamma_0)}(\gamma_0 &\in CI_{\gamma,n}(a)) = \liminf_{n\to\infty} \Pr_{(\theta_0,\gamma_0)}(\gamma_0 \in CI^I_{\gamma,n}(a)) \\ &= \liminf_{n\to\infty} (LR_n(\gamma_0, \widehat{\gamma}_n, \widehat{\theta}_n) \leq q^I_{1-\alpha}) \\ &= 1 - a, \end{aligned}$$

where the last equality holds because $LR_n(\gamma_0, \widehat{\gamma}_n, \widehat{\theta}_n) \Rightarrow \max_{r\in(-\infty,\infty)}(2\Lambda(r) - |r|)$ when δ_n is a fixed constant.

The proof of $AsySZ_{\theta}(a) = 1 - a$ can be done using a analogous argument and it is omitted for brevity. Q.E.D.

A.2 Mathematic proof for extended Model

Lemma A.2.1 *Under Assumptions 2.1.1-2.1.4, for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *as* $n \to \infty$,

a) $n^{-3/2} \sum_{t=1}^{n} z_t x_t'(\gamma) = h(\gamma) \int_0^1 X'(s)ds + o_p(1)$;

b) $n^{-3/2} \sum_{t=1}^{n} z_t x_{t-1}'(\gamma) = h_1(\gamma) \int_0^1 X'(s)ds + o_p(1)$;

c) $n^{-3/2} \sum_{t=1}^{n} z_{t-1} x_t'(\gamma) = h_2(\gamma) \int_0^1 X'(s)ds + o_p(1)$;

d) $n^{-1} \sum_{t=1}^{n} z_t z_t' = H + o_p(1)$;

e) $n^{-1} \sum_{t=1}^{n} z_t z_{t-1}' = H_1 + o_p(1)$,

where $h(\gamma)$, $h_1(\gamma)$, $h_2(\gamma)$, H, H_1 *are defined in (2.31).*

Proof: Following Theorem 3 in Caner and Hansen (2001), I can prove $a) - c)$. Proofs of $d)$ and $e)$ can be done by applying the strong law of large numbers for stationary processes under Assumption 2.1.1. Q.E.D.

Lemma A.2.2 *Under Assumptions 2.1.1-2.1.4, for any* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *as* $n \to \infty$, *I have*

a) $n^{-1/2} \sum_{t=1}^{n} I(q_t \leq \gamma) e_t \Rightarrow \sigma W(s, \gamma)$,

b) $n^{-1/2} \sum_{t=1}^{n} I(q_{t-1} \leq \gamma) e_t \Rightarrow \sigma W_1(s, \gamma)$,

c) $n^{-1} \sum_{t=1}^{n} x_t(\gamma)' e_t \Rightarrow \sigma \int_0^1 X'(s) dW(s, \gamma)$

d) $n^{-1} \sum_{t=1}^{n} x_{t-1}(\gamma)' e_t \Rightarrow \sigma \int_0^1 X'(s) dW_1(s, \gamma)$

e) $n^{-1/2} \sum_{t=1}^{n} z_t e_t \Rightarrow \sigma J_2$

f) $n^{-1/2} \sum_{t=1}^{n} z_{t-1} e_t \Rightarrow \sigma J_3$

where J_2 and J_3 are Gaussian random variable with mean zero and variance H. $W(s,\gamma)$ and $W_1(s,\gamma)$ are two-parameter Brownian motions defined in Definition 1.

Proof: Proofs of (a) and (b) follow the results of Lemma 2.1.2. Proofs of c) and d) follow Lemma 2.1.3. Proofs of e) and f) can be done by applying the central limiting theorem for a square integrable stationary martingale difference sequence. Q.E.D.

Lemma A.2.3 *Under Assumptions 2.1.1-2.2.1, if $\tau < 1/2$, I have $\widehat{\gamma}_n \xrightarrow{p} \gamma_0$ for any $\rho \in (-1, 1]$.*

Proof: Rewrite the extended model as a matrix compacted form:

$$Y = X'\alpha + X(\gamma_0)'\delta_n + \xi,$$

where $Y, X, X(\gamma_0)$ and ξ stack $y_t, x_t, x_t(\gamma_0)$ and ξ_t respectively. Denote $X^*(\gamma) = (X(\gamma), X - X(\gamma))$ and define its projection matrix $P^*_\gamma = X^*(\gamma)(X^*(\gamma)'X^*(\gamma))^{-1}X^*(\gamma)'$. After some simple algebra, I have

$$SSR_n(\gamma) = Y'(I - P^*_\gamma)Y = \delta_n'X(\gamma_0)'(I - P^*_\gamma)X(\gamma_0)\delta_n + 2\delta_n'X(\gamma_0)'(I - P^*_\gamma)\xi + \xi'(I - P^*_\gamma)\xi,$$

where the second equation uses the fact that $X'(I - P^*_\gamma) = 0$. It follows that,

$$\begin{aligned}
& n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0)) \\
= \; & n^{-1+2\tau}\delta_n'X(\gamma_0)'(I - P^*_\gamma)X(\gamma_0)\delta_n + n^{-1+2\tau}2\delta_n'X(\gamma_0)'(I - P^*_\gamma)\xi \\
& + n^{-1+2\tau}(\xi'(I - P^*_\gamma)\xi - \xi'(I - P^*_{\gamma_0})\xi) \\
\equiv \; & S_1^* + S_2^* + S_3^*, \text{ say.}
\end{aligned}$$

To prove the consistency of $\widehat{\gamma}_n$, it suffices to show $n^{2\tau-1}(SSR_n(\gamma) - SSR_n(\gamma_0))$ uniformly converge to a function which takes global minimum value at γ_0. In the following, I conduct the proof by considering two cases according to the value of ρ.

Case 1: $\rho < 1$,where $\xi_t = \beta' z_t + \eta_t$ is stationary process. Given $\tau < 1/2$, using Lemma A.2.2, it can be shown that

$$S_2^* = n^{-1/2+\tau} 2 \left(n^{1/2+\tau}\delta_n\right)' \frac{1}{n} X(\gamma_0)'(I - P_\gamma^*)\xi = n^{-1/2+\tau} 2\delta_0' \frac{1}{n} X(\gamma_0)'(I - P_\gamma^*)\xi = O_p(n^{-1/2+\tau}) \tag{A.15}$$

and

$$S_3^* = n^{-1+2\tau}(\xi'(I - P_\gamma^*)\xi - \xi'(I - P_{\gamma_0}^*)\xi) = n^{-1+2\tau}(\xi' P_{\gamma_0}^* \xi - \xi' P_\gamma^* \xi) = O_p(n^{-1+2\tau}). \tag{A.16}$$

Using a similar argument of Lemma A.5 in Hansen(2000), I can show, for any $\gamma \geq \gamma_0$,

$$S_1^* \xrightarrow{p} (F(\gamma_0) - F(\gamma_0)F(\gamma)^{-1}F(\gamma_0))\delta_0' \int_0^1 X(s)X(s)' ds \delta_0 \equiv b_1^*(\gamma) \tag{A.17}$$

uniformly. Since $(F(\gamma_0) - F(\gamma_0)F(\gamma)^{-1}F(\gamma_0)) \geq 0$ and $\int_0^1 X(s)X(s)'ds$ is positive definite matrix, $b_1^*(\gamma) \geq 0$ and the equality holds if and only if $\gamma = \gamma_0$. Combining all convergence results, I have

$$\begin{aligned} & n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0)) \\ \xrightarrow{p} \quad & F(\gamma)^{-1}F(\gamma_0)(F(\gamma) - F(\gamma_0))\delta' \int_0^1 X(s)X(s)' ds \delta_0 \equiv b_1^*(\gamma) \geq 0. \end{aligned}$$

Symmetrically, I can prove for $\gamma \leq \gamma_0$,

$$n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0)) \xrightarrow{p} (F(\gamma_0) - F(\gamma))\delta_0' \int_0^1 X(s)X(s)' ds \delta_0 \equiv b_2^*(\gamma)$$

uniformly, where $b_2^*(\gamma) \geq 0$ and the equality holds if and only if $\gamma = \gamma_0$. Define $b^*(\gamma) = b_1^*(\gamma)I(\gamma \geq \gamma_0) + b_2^*(\gamma)I(\gamma \leq \gamma_0)$. We have

$$n^{-1+2\tau}(SSR_n(\gamma) - SSR_n(\gamma_0)) \xrightarrow{p} b^*(\gamma)$$

uniformly for any $\gamma \in [\underline{\gamma}, \overline{\gamma}]$ and $b^*(\gamma)$ takes global minimum at γ_0. Therefore,

$$\widehat{\gamma}_n = \arg\min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} SSR_n(\gamma) = \arg\min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} \left(n^{-1+2\tau} (SSR_n(\gamma) - SSR_n(\gamma_0)) \right) \xrightarrow{p} \gamma_0.$$

Case 2: $\rho = 1$, where $\xi_t = \beta' z_t + \eta_t$ is nonstationary since η_t is a unit root process. Similarly, I will show that both S_2^* and S_3^* converge to zero and S_1^* uniformly converges to a function which takes global minimum at γ_0.

From Lemma A.2.1, I have

$$n^{-2} X^*(\gamma)' X^*(\gamma) \Rightarrow \begin{pmatrix} F(\gamma) \int_0^1 X(s) X'(s) ds, 0 \\ 0, (1 - F(\gamma)) \int_0^1 X(s) X'(s) ds \end{pmatrix},$$

$$n^{-2} X^*(\gamma)' \eta \Rightarrow \begin{pmatrix} F(\gamma) \int_0^1 X(s) B_1(s) ds \\ (1 - F(\gamma)) \int_0^1 X(s) B_1(s) ds \end{pmatrix},$$

and

$$n^{-2} \eta' \eta \Rightarrow \int_0^1 B_1(s) B_1(s) ds,$$

where $B_1(s)$ is assumed to be a Brownian motion such that

$$\frac{1}{\sqrt{n}} \eta_{[ns]} \Rightarrow B_1(s).$$

It can be shown that

$$S_3^* = n^{-1+2\tau} (\xi'(I - P_\gamma^*)\xi - \xi'(I - P_{\gamma_0}^*)\xi) = n^{-1+2\tau} (\eta'(I - P_\gamma^*)\eta - \eta'(I - P_{\gamma_0}^*)\eta) + o_p(1).$$

Note that

$$\begin{aligned}
& n^{-2} \eta' X^*(\gamma) (X^*(\gamma)' X^*(\gamma))^{-1} X^*(\gamma)' \eta \\
= \; & n^{-2} \eta' X^*(\gamma) (n^{-2} X^*(\gamma)' X^*(\gamma))^{-1} n^{-2} X_1^*(\gamma)' \eta \\
= \; & \begin{pmatrix} F(\gamma) \int_0^1 X(s) B_1(s) ds \\ (1 - F(\gamma)) \int_0^1 X(s) B_1(s) ds \end{pmatrix}' \begin{pmatrix} (\int_0^1 X(s) X'(s) ds)^{-1} \int_0^1 X(s) B_1(s) ds \\ (\int_0^1 X(s) X'(s) ds)^{-1} \int_0^1 X(s) B_1(s) ds \end{pmatrix} + o_p(1) \\
= \; & (\int_0^1 X(s) B_1(s) ds)' (\int_0^1 X(s) X'(s) ds)^{-1} \int_0^1 X(s) B_1(s) ds + o_p(1)
\end{aligned}$$

which is unrelated to γ. Thus,

$$S_3^* = n^{-1+2\tau}(\eta' P_\gamma^* \eta - \eta' P_{\gamma_0}^* \eta) = o_p(1). \tag{A.18}$$

To prove S_2^* converge to zero almost surely, I first consider the case where $\gamma > \gamma_0$. For the case where $\gamma \leq \gamma_0$,the proof is similar and I skip the detail. When $\gamma > \gamma_0$, I have

$$n^{-2}X'(\gamma)X(\gamma_0) = n^{-2}X'(\gamma_0)X(\gamma_0) = F(\gamma_0)\int_0^1 X(s)X'(s)ds + o_p(1);$$

$$n^{-2}X(\gamma_0)'(X - X(\gamma)) = 0;$$

$$n^{-2}(X - X(\gamma_0))'X(\gamma) = (F(\gamma) - F(\gamma_0))\int_0^1 X(s)X'(s)ds + o_p(1).$$

It follows that

$$\begin{aligned}
& n^{-2}X^*(\gamma_0)'X^*(\gamma) \\
= & \begin{pmatrix} n^{-2}\sum_{t=1}^n x_t(\gamma)x_t'(\gamma_0), n^{-2}\sum_{t=1}^n x_t(\gamma_0)(x_t - x_t(\gamma))' \\ n^{-2}\sum_{t=1}^n (x_t - x_t(\gamma))x_t(\gamma_0)', n^{-2}\sum_{t=1}^n (x_t - x_t(\gamma_0))(x_t - x_t(\gamma))' \end{pmatrix} \\
= & \begin{pmatrix} F(\gamma_0)\int_0^1 X(s)X'(s)ds, 0 \\ (F(\gamma) - F(\gamma_0))\int_0^1 X(s)X'(s)ds, (1 - F(\gamma))\int_0^1 X(s)X'(s)ds \end{pmatrix} + o_p(1).
\end{aligned}$$

Furthermore, it can be shown that

$$n^{-1+2\tau}2\delta_n'X(\gamma_0)'\eta = 2n^{1/2+\tau}\delta_0'n^{-2}\sum_{t=1}^n x_t(\gamma_0)\eta \Rightarrow 2n^{1/2+\tau}\delta_0'\begin{pmatrix} F(\gamma_0)\int_0^1 X(s)B_1(s)ds, \\ (1 - F(\gamma_0))\int_0^1 X(s)B_1(s)ds \end{pmatrix};$$

and

$$2n^{-1+2\tau}\delta_n'X(\gamma_0)'P_\gamma^*\eta = 2n^{1/2+\tau}\delta_0'\begin{pmatrix} F(\gamma_0)\int_0^1 X(s)X'(s)ds \\ (1 - F(\gamma_0))\int_0^1 X(s)B_1(s)ds \end{pmatrix} + o_p(1).$$

Thus,

$$S_2^* = n^{-1+2\tau}2\delta_n'X(\gamma_0)'(I - P_\gamma^*)\xi = n^{-1+2\tau}2\delta_n'X(\gamma_0)'\eta - 2n^{-1+2\tau}\delta_n'X(\gamma_0)'P_\gamma^*\eta + o_p(1) = o_p(1).$$

Using a similar argument of Case 1, I can show $S_1^* \xrightarrow{p} b^*(\gamma)$ uniformly and $b^*(\gamma)$ takes global minimum value at γ_0. Combining the convergence results for S_1^*, S_2^* and S_3^*, I complete the proof. Q.E.D.

Proof of Proposition 9 : The model can be written as

$$\begin{aligned} y_t &= \theta_1' x_t^*(\gamma_0) + \xi_t \\ &= \theta_1' x_t^*(\widehat{\gamma}) + \xi_t + \theta_1'(x_t^*(\gamma_0) - x_t^*(\widehat{\gamma})) \\ &= \theta_1' x_t^*(\widehat{\gamma}) + \xi_t^* \end{aligned}$$

where $x_t^*(\widehat{\gamma}) = (x_t(\widehat{\gamma})', x_t' - x_t'(\widehat{\gamma}))'$, $\theta_1 = (\alpha' + \delta_n', \alpha')'$ and $\xi_t^* = \xi_t + \theta_1'(x_t^*(\gamma_0) - x_t^*(\widehat{\gamma}))$.

Noting that $\theta_1'(x_t^*(\gamma_0) - x_t^*(\widehat{\gamma})) = \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma}))$, I have

$$\xi_t^* = \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})).$$

The residual $\widehat{\xi_t}(\widehat{\gamma})$ can be expressed as

$$\widehat{\xi_t}(\widehat{\gamma}) = y_t - \widehat{\theta}_1(\widehat{\gamma})' x_t^*(\widehat{\gamma}) = \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta}_1(\widehat{\gamma}) - \theta_1)' x_t^*(\widehat{\gamma}) \qquad \text{(A.19)}$$

Next, I conduct the proof by considering two cases according to the value of τ.

Case 1: $\tau < 1/2$, from Lemma A.2.3, I have $\widehat{\gamma} \xrightarrow{p} \gamma_0$.

If $\rho = 1$, $\xi_t = \beta' z_t + \eta_t$ is a unit root. Thus,

$$\begin{aligned} \widehat{\theta}_1(\widehat{\gamma}) - \theta_1 &= \left(n^{-2} \sum_{t=1}^{n} x_t^*(\widehat{\gamma}) x_t^*(\widehat{\gamma})' \right)^{-1} n^{-2} \sum_{t=1}^{n} x_t^*(\widehat{\gamma}) \xi_t^* \\ \Rightarrow & \begin{pmatrix} F(\gamma_0) \int X(s)X(s)'ds, 0 \\ 0, (1 - F(\gamma_0)) \int_0^1 X(s)X(s)'ds \end{pmatrix}^{-1} \begin{pmatrix} F(\gamma_0) \int_0^1 X(s)B_1(s)ds \\ (1 - F(\gamma_0)) \int_0^1 X(s)B_1(s)ds \end{pmatrix} \\ = & \begin{pmatrix} (\int_0^1 X(s)X(s)'ds)^{-1} \int_0^1 X(s)B_1(s)ds \\ (\int_0^1 X(s)X(s)'ds)^{-1} \int_0^1 X(s)B_1(s)ds \end{pmatrix} \equiv \begin{pmatrix} \varphi \\ \varphi \end{pmatrix}, \text{say.} \end{aligned}$$

Thus, from equation (A.19), I have

$$\begin{aligned}\widehat{\xi_t}(\widehat{\gamma}) &= \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta_1}(\widehat{\gamma}) - \theta_1)'x_t^*(\widehat{\gamma}) \\ &\xrightarrow{p} \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\varphi, \varphi)x_t^*(\widehat{\gamma}) \\ &= \beta' z_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + \eta_t + \varphi x_t.\end{aligned}$$

It can be shown that[1] $\rho(\widehat{\gamma}) = 1 + O_p(\frac{1}{n})$.

If $\rho < 1$, then $\xi_t = \beta' z_t + \eta_t$ is stationary. We have

$$\delta_n' \left(x_t(\gamma_0) - x_t(\widehat{\gamma})\right) = o_p(1)$$

and

$$\begin{aligned}n(\widehat{\theta_1}(\widehat{\gamma}) - \theta_1) &= (n^{-2}\sum_{t=1}^{n} x_t^*(\widehat{\gamma})x_t^*(\widehat{\gamma})')^{-1} n^{-1}\sum_{t=1}^{n} x_t^*(\widehat{\gamma})\xi_t + o_p(1) \\ \Rightarrow & \begin{pmatrix} F(\gamma_0)\int_0^1 X(s)X(s)'ds, 0 \\ 0, (1-F(\gamma_0))\int_0^1 X(s)X(s)'ds \end{pmatrix}^{-1} \begin{pmatrix} \int_0^1 X(s)dW_\xi(s,\gamma_0) \\ \int_0^1 X(s)d\left(W_\xi(s) - W_\xi(s,\gamma_0)\right) \end{pmatrix} \equiv \phi_\xi(\gamma_0)\end{aligned}$$

say. Thus,

$$\begin{aligned}\widehat{\xi_t}(\widehat{\gamma}) &= \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta_1}(\widehat{\gamma}) - \theta_1)'x_t^*(\widehat{\gamma}) \\ &= \xi_t + o_p(1) + \frac{1}{\sqrt{n}}x_t^*(\widehat{\gamma})'\frac{1}{\sqrt{n}}\phi_\xi(\gamma_0) \xrightarrow{p} \xi_t.\end{aligned}$$

Since ξ_t is a stationary process, it is straightforward to show that

$$\widehat{\rho}(\widehat{\gamma}) = \rho + O_p(\frac{1}{\sqrt{n}}).$$

Case 2: $\tau = 1/2, \widehat{\gamma}$ is not consistent. Noting that $n\delta_n = \delta_0$, I have

$$\delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) = O_p(\frac{1}{\sqrt{n}}) = o_p(1).$$

[1]We skip the detail to save the space, but it is available upon request.

If $\rho = 1$, I have

$$\widehat{\theta_1}(\widehat{\gamma}) - \theta_1 = (n^{-2}\sum_{t=1}^{n} x_t^*(\widehat{\gamma})x_t^*(\widehat{\gamma})')^{-1}n^{-2}\sum_{t=1}^{n} x_t^*(\widehat{\gamma})\xi_t \xrightarrow{p} \begin{pmatrix} \varphi \\ \varphi \end{pmatrix}.$$

Thus,

$$\widehat{\xi_t}(\widehat{\gamma}) = \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta_1}(\widehat{\gamma}) - \theta_1)'x_t^*(\widehat{\gamma}) \xrightarrow{p} \xi_t + (\varphi, \varphi)x_t^*(\widehat{\gamma}) = \xi_t + \varphi x_t$$

which is an I(1) process. It can be shown that

$$\widehat{\rho(\widehat{\gamma})} = 1 + O_p(\frac{1}{n}).$$

If $\rho < 1$, I have

$$\begin{aligned}\widehat{\xi_t}(\widehat{\gamma}) &= \xi_t + \delta_n'(x_t(\gamma_0) - x_t(\widehat{\gamma})) + (\widehat{\theta_1}(\widehat{\gamma}) - \theta_1)'x_t^*(\widehat{\gamma}) \\ &= \xi_t + o_p(1) + \frac{1}{\sqrt{n}}x_t^*(\widehat{\gamma})'\frac{1}{\sqrt{n}}\phi_\xi(\widehat{\gamma}) \xrightarrow{p} \xi_t.\end{aligned}$$

Thus,

$$\rho(\widehat{\gamma}) = \rho + O_p(\frac{1}{\sqrt{n}}).$$

Q.E.D.

Define the following for proving Lemma 2.3.1

$$G(\gamma) = \begin{pmatrix} \int_0^1 X(s)X'(s)ds,\ F(\gamma)\int_0^1 X(s)X'(s)ds,\ 0 \\ F(\gamma)\int_0^1 X(s)X'(s)ds, F(\gamma)\int_0^1 X(s)X'(s)ds,\ \int_0^1 X(s)dsh(\gamma)' \\ 0, \quad h(\gamma)\int_0^1 X'(s)ds, \quad H \end{pmatrix} \tag{A.20}$$

and

$$G_1(\gamma) = \begin{pmatrix} \int_0^1 X(s)X'(s)ds,\ F(\gamma)\int_0^1 X(s)X'(s)ds,\ 0 \\ F(\gamma)\int_0^1 X(s)X'(s)ds, F_1(\gamma)\int_0^1 X(s)X'(s)ds,\ \int_0^1 X(s)dsh_2'(\gamma) \\ 0, \quad h_1(\gamma)\int_0^1 X(s)'ds, \quad H_1 \end{pmatrix} \tag{A.21}$$

Proof of Lemma 2.3.1: Let $V_{1t}(\gamma) = (x_t, x_t(\gamma), z_t)$. Then, $\widetilde{V}_t(\gamma) = V_{1t}(\gamma) - \widehat{\rho} V_{1t-1}(\gamma)$. When $\rho < 1$, let

$$\widetilde{D}_n = diag\{n^{1/2} I_{d_1}, n^{1/2} I_{d_1}, I_{d_2}\},$$

$$\widetilde{G}(\gamma) = G(\gamma) + \rho^2 G(\gamma) - \rho \left(G_1(\gamma) + G_1(\gamma)\right)',$$

and

$$\widetilde{\phi}(\gamma) = \begin{pmatrix} \sigma \int_0^1 X_1(s) d\widetilde{W}(s) \\ \sigma \int_0^1 X_1(s) d\widetilde{W}(s,\gamma) \\ \sigma \widetilde{J}_2 \end{pmatrix} \tag{A.22}$$

where $\widetilde{W}(s) = (1-\rho)W(s)$, $\widetilde{W}(s,\gamma) = W(s,\gamma) - \rho W_1(s,\gamma)$ and $\widetilde{J} = J_2 - \rho J_3$.

Using Lemma A.2.1 and Proposition 9, I have

$$\begin{aligned} & n^{-1} \sum_{t=1}^{n} \widetilde{D}_n^{-1} \widetilde{V}_t(\gamma) \widetilde{V}_t(\gamma)' \widetilde{D}_n^{-1} \\ = \; & n^{-1} \sum_{t=1}^{n} \widetilde{D}_n^{-1} V_{1t}(\gamma) V_{1,t}(\gamma)' \widetilde{D}_n^{-1} + n^{-1} \widehat{\rho}^2 \sum_{t=1}^{n} \widetilde{D}_n^{-1} V_{1t-1}(\gamma) V_{1,t-1}(\gamma)' \widetilde{D}_n^{-1} \\ & - \widehat{\rho} \sum_{t=1}^{n} \widetilde{D}_n^{-1} V_{1t}(\gamma) V_{1,t-1}(\gamma)' \widetilde{D}_n^{-1} - \widehat{\rho} \sum_{t=1}^{n} \widetilde{D}_n^{-1} V_{1t-1}(\gamma) V_{1,t}(\gamma)' \widetilde{D}_n^{-1} \\ & \xrightarrow{p} G(\gamma) + \rho^2 G(\gamma) - \rho \left(G_1(\gamma) + G_1(\gamma)\right)' = \widetilde{G}(\gamma). \end{aligned}$$

Using Lemma A.2.2, I have

$$n^{-1/2} \sum_{t=1}^{n} \widetilde{D}_n^{-1} \widetilde{V}_t(\gamma) e_t = \begin{pmatrix} n^{-1} \sum_{t=1}^{n} \widetilde{x}_t e_t \\ n^{-1} \sum_{t=1}^{n} D_n^{-1} \widetilde{x}_t(\gamma) e_t \\ n^{-1/2} \sum_{t=1}^{n} \widetilde{z}_t e_t \end{pmatrix} \Rightarrow \begin{pmatrix} (1-\rho) \int_0^1 X_1(s) dW_1(s) \\ \sigma \int_0^1 X_1(s) d\widetilde{W}(s,\gamma) \\ \sigma \widetilde{J}_2 \end{pmatrix} = \widetilde{\phi}(\gamma).$$

Define $\widetilde{\Pi}(\gamma, \gamma_0, \delta_0) = p\lim_{n\to\infty} \begin{pmatrix} n^{-2} \sum_{t=1}^{n} \widetilde{x}_t (\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))' \delta_0 \\ n^{-2} \sum_{t=1}^{n} \widetilde{x}_t(\gamma) (\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))' \delta_0 \\ n^{-3/2} \sum_{t=1}^{n} \widetilde{z}_t (\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))' \delta_0 \end{pmatrix}$ which exists based on Lemma A.2.1. Then

$$n^{-3/2}\widetilde{D}_n^{-1}\widetilde{V}(\gamma)'(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))'\delta_0 = \begin{pmatrix} n^{-2}\sum_{t=1}^n \widetilde{x}_t(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))'\delta_0 \\ n^{-2}\sum_{t=1}^n \widetilde{x}_t(\gamma)(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))'\delta_0 \\ n^{-3/2}\sum_{t=1}^n \widetilde{z}_t(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))'\delta_0 \end{pmatrix} \xrightarrow{p} \widetilde{\Pi}(\gamma,\gamma_0,\delta_0).$$

When $\rho = 1$, I can similarly prove the convergence results. Q.E.D.

Lemma A.2.4 *, If $\tau < 1/2$, when $\gamma = \gamma_0$*

$$\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0)-\widetilde{\theta}) \Rightarrow \widetilde{G}(\gamma_0)^{-1}\widetilde{\phi}(\gamma_0)$$

when $\gamma \neq \gamma_0$

$$n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta}) \Rightarrow \widetilde{G}(\gamma)^{-1}\widetilde{\phi}_1(\gamma,\gamma_0).$$

If $\tau = 1/2$, when $\gamma = \gamma_0$, I have

$$\sqrt{n}\widetilde{D}_n(\widehat{\theta}(\gamma_0)-\theta) \Rightarrow \widetilde{G}(\gamma_0)^{-1}\widetilde{\phi}(\gamma_0)$$

when $\gamma \neq \gamma_0$,

$$n^{1/2}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta}) \Rightarrow \widetilde{G}(\gamma)^{-1}\left(\widetilde{\Pi}(\gamma,\gamma_0,\delta_0)+\widetilde{\phi}(\gamma)\right).$$

Proof: Note that

$$\begin{aligned} \widetilde{y}_t &= \widetilde{\theta}'\widetilde{V}_t(\gamma)+\widetilde{\eta}_t+\widetilde{\theta}'(\widetilde{V}_t(\gamma)-\widetilde{V}_t(\gamma_0)) \\ &= \widetilde{\theta}'\widetilde{V}_t(\gamma)+\widetilde{\eta}_t+\delta_n'(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0)) \\ &= \widetilde{\theta}'\widetilde{V}_t(\gamma)+\widetilde{\eta}_t^* \end{aligned}$$

where $\widetilde{\eta}_t^* = \widetilde{\eta}_t + \delta_n'(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))$. It follows that,

$$(\widehat{\widetilde{\theta}}(\gamma)-\theta)$$

$$
\begin{aligned}
&= (\sum_{t=2}^{n}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)')^{-1}\sum_{t=2}^{n}\widetilde{V}_t(\gamma)\widetilde{\eta}_t^{*} \\
&= (\sum_{t=2}^{n}\widetilde{V}_t(\gamma_0)\widetilde{V}_t(\gamma_0)')^{-1}\sum_{t=2}^{n}\widetilde{V}_t(\gamma_0)e_t + (\sum_{t=2}^{n}\widetilde{V}_t(\gamma_0)\widetilde{V}_t(\gamma_0)')^{-1}\sum_{t=2}^{n}\widetilde{V}_t(\gamma_0)(\rho-\widehat{\rho})\eta_{t-1} \\
&\quad +(\sum_{t=2}^{n}\widetilde{V}_t(\gamma_0)\widetilde{V}_t(\gamma_0)')^{-1}\sum_{t=2}^{n}\widetilde{V}_t(\gamma_0)(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))'\delta_n.
\end{aligned}
$$

From Proposition 9, I have $\rho-\widehat{\rho}=o_p(1)$.

If $\tau<1/2$, I have

$$
\begin{aligned}
\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0)-\theta) &= (\frac{1}{n}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma_0)\widetilde{V}_t(\gamma_0)'\widetilde{D}_n^{-1})^{-1}\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma_0)e_t + o_p(1) \\
&\Rightarrow \widetilde{G}(\gamma_0)^{-1}\widetilde{\phi}(\gamma_0)
\end{aligned}
$$

when $\gamma\neq\gamma_0$,

$$
\begin{aligned}
& n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta}) = n^{\tau-1/2}(\frac{1}{n}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)'\widetilde{D}_n^{-1})^{-1}\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)e_t \\
& -(\frac{1}{n}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)'\widetilde{D}_n^{-1})^{-1}\frac{1}{n}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)(n^{-1/2}\widetilde{x}_t(\gamma)-n^{-1/2}\widetilde{x}_t(\gamma_0)'(n^{\tau+1/2}\delta_{1n})+o_p(1) \\
\Rightarrow\ & \widetilde{G}(\gamma)^{-1}\widetilde{\Pi}(\gamma,\gamma_0,\delta_0).
\end{aligned}
$$

If $\tau=1/2$, using a similar argument, I can prove

$$
\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0)-\widetilde{\theta})\Rightarrow\widetilde{G}(\gamma_0)^{-1}\widetilde{\phi}(\gamma_0)
$$

when $\gamma\neq\gamma_0$, I have

$$
\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta})\Rightarrow\widetilde{G}(\gamma)^{-1}(\widetilde{\Pi}(\gamma,\gamma_0,\delta_0)+\widetilde{\phi}(\gamma)).
$$

Q.E.D.

Lemma A.2.5 *Under Assumptions 2.1.1-2.2.1, if* $\tau<1/2$, *I have* $\widetilde{\gamma}_n\xrightarrow{p}\gamma_0$.

Proof: To prove the consistency of $\widetilde{\gamma}_n$, I need prove to $\Pr(|\widetilde{\gamma}_n - \gamma_0| > \varepsilon) \to 0$ for every $\varepsilon > 0$. Denote $B(\epsilon) = \{\gamma : |\gamma - \gamma_0| > \varepsilon\}$ and $\overline{B}(\epsilon) = [\underline{\gamma}, \overline{\gamma}]\backslash B(\epsilon)$. Noting that

$$\begin{aligned}
\Pr(|\widetilde{\gamma}_n - \gamma_0| > \varepsilon) &= \Pr(\inf_{\gamma \in B(\epsilon)} \widetilde{SSR}_n(\gamma) < \inf_{\gamma \in \overline{B}(\epsilon)} \widetilde{SSR}_n(\gamma)) \\
&\leq \Pr(\inf_{\gamma \in B(\epsilon)} \widetilde{SSR}_n(\gamma) < \widetilde{SSR}_n(\gamma_0)) \\
&= \Pr(\inf_{\gamma \in B(\epsilon)} n^{2\tau-1}(\widetilde{SSR}_n(\gamma) - \widetilde{SSR}_n(\gamma_0)) < 0)
\end{aligned}$$

To prove $\Pr(|\widetilde{\gamma}_n - \gamma_0| > \varepsilon) \to 0$, it suffices to show $\inf_{\gamma \in B(\epsilon)} n^{2\tau-1}(\widetilde{SSR}_n(\gamma) - \widetilde{SSR}_n(\gamma_0)) > 0$ with probability 1. From the definition, I have

$$\begin{aligned}
\widetilde{SSR}_n(\gamma) &= \sum_{t=1}^{n}(\widetilde{y}_t - \widehat{\widetilde{\theta}}(\gamma)'\widetilde{V}_t(\gamma))^2 = \sum_{t=2}^{n}(\widetilde{\theta}'\widetilde{V}_t(\gamma) + \widetilde{\eta}_t^* - \widehat{\widetilde{\theta}}(\gamma)'\widetilde{V}_t(\gamma))^2 \\
&= \sum_{t=2}^{n}(\widetilde{\eta}_t - (\widehat{\widetilde{\theta}}(\gamma) - \widetilde{\theta})'\widetilde{V}_t(\gamma) - \delta_n'(\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0)))^2 \\
&= \sum_{t=2}^{n}\widetilde{\eta}_t^2 + (\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma) - \widetilde{\theta}))'(\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)'\widetilde{D}_n^{-1})\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma) - \theta) \\
&\quad + \delta_n'\sum_{t=2}^{n}(\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))(\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))'\delta_n \\
&\quad - 2(\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma) - \widetilde{\theta}))'(\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)(\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))'\delta_n \\
&\quad - 2(\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma) - \widetilde{\theta}))'\left(\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{\eta}_t\right) - 2\delta_n'\sum_{t=2}^{n}(\widetilde{x}_t(\gamma) - \widetilde{x}_t(\gamma_0))\widetilde{\eta}_t) \\
&\equiv \widetilde{S}_0 + \widetilde{S}_1 + \widetilde{S}_2 - \widetilde{S}_3 - \widetilde{S}_4 - \widetilde{S}_5, \text{say,}
\end{aligned}$$

and

$$\begin{aligned}
\widetilde{SSR}_n(\gamma_0) &= \sum_{t=2}^{n}(\widetilde{y}_t - \widehat{\widetilde{\theta}}(\gamma_0)'\widetilde{V}_t(\gamma_0))^2 = \sum_{t=2}^{n}(\widetilde{\eta}_t - (\widehat{\widetilde{\theta}}(\gamma_0) - \theta)'\widetilde{V}_t(\gamma))^2 \\
&= \sum_{t=1}^{n}\widetilde{\eta}_t^2 + (\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0) - \theta))'(\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)'\widetilde{D}_n^{-1})\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0) - \widetilde{\theta}) \\
&\quad - 2(\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0) - \widetilde{\theta}))'\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{\eta}_t \\
&\equiv \widetilde{S}_0 + \widetilde{S}_6 - \widetilde{S}_7, \text{say.}
\end{aligned}$$

It follows that

$$n^{2\tau-1}(SSR_n(\gamma)-SSR_n(\gamma_0)) = n^{2\tau-1}(\widetilde{S}_1+\widetilde{S}_2-\widetilde{S}_3-\widetilde{S}_4-\widetilde{S}_5-\widetilde{S}_6+\widetilde{S}_7).$$

Next, I show that $n^{2\tau-1}(\widetilde{S}_1+\widetilde{S}_2+\widetilde{S}_3)$ uniformly converges to a function $\widetilde{b}(\gamma)$ which is positive when $\gamma \in B(\epsilon)$, while the left terms converges to zero in probability. By Lemma A.2.5, if $\tau < 1/2$, I have

$$\begin{aligned}
& n^{2\tau-1}(\widetilde{S}_1+\widetilde{S}_2-\widetilde{S}_3) \\
= \;& n^{-1}(n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta}))'(\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)'\widetilde{D}_n^{-1})n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\theta) + \\
& n^{-1}(n^{\tau+1/2}\delta_n)'\sum_{t=2}^{n}\frac{1}{\sqrt{n}}(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))\frac{1}{\sqrt{n}}((\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))'\,n^{\tau+1/2}\delta_n \\
& -2n^{-1}(n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta}))'(\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)(\frac{1}{\sqrt{n}}\widetilde{x}_t(\gamma)-\frac{1}{\sqrt{n}}\widetilde{x}_t(\gamma_0))'n^{\tau+1/2}\delta_n \\
= \;& n^{-1}\sum_{t=2}^{n}\left((n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\widetilde{\theta}))'\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)+\delta_0'\left(\frac{1}{\sqrt{n}}(\widetilde{x}_t(\gamma)-\frac{1}{\sqrt{n}}\widetilde{x}_t(\gamma_0)\right)\right)^2 \\
\Rightarrow \;& \widetilde{b}(\gamma) > 0, \mathit{say}.
\end{aligned}$$

and

$$\begin{aligned}
& n^{2\tau-1}(-\widetilde{S}_6+\widetilde{S}_7) \\
= \;& -n^{2\tau-1}(\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0)-\theta))'(\frac{1}{n}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{V}_t(\gamma)'\widetilde{D}_n^{-1})\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0)-\widetilde{\theta}) \\
& +2n^{2\tau-1}(\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0)-\widetilde{\theta}))'\frac{1}{\sqrt{n}}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{\eta}_t \\
& \xrightarrow{p} -n^{2\tau-1}\widetilde{\lambda}(\gamma_0)'\widetilde{G}(\gamma_0)^{-1}\widetilde{\lambda}(\gamma_0)+n^{2\tau-1}\widetilde{\lambda}(\gamma_0)'\widetilde{G}(\gamma_0)\lambda(\gamma_0) \\
= \;& O_p(n^{2\tau-1}) = o_p(1)
\end{aligned}$$

$$\begin{aligned}
n^{2\tau-1}(-\widetilde{S}_4-\widetilde{S}_5) \;=\; & 2n^{\tau-1/2}(n^{\tau}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma)-\theta))'n^{-1/2}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma)\widetilde{\eta}_t \\
& -2n^{\tau-1/2}\left(n^{\tau+1/2}\delta_n\right)'\frac{1}{\sqrt{n}}\sum_{t=1}^{n}D_n^{-1}(\widetilde{x}_t(\gamma)-\widetilde{x}_t(\gamma_0))\widetilde{\eta}_t) \\
=\; & O_p(n^{\tau-1/2}) = o_p(1)
\end{aligned}$$

Thus, I complete the proof. Q.E.D.

Lemma A.2.6 *Under Assumptions 2.1.1-2.2.1, if* $\tau < 1/2$, *then* $n^{1-2\tau}|\widetilde{\gamma}_n - \gamma_0| = O_p(1)$.

Proof: The proof of the convergence rate for $\widetilde{\gamma}$ to be $a_n = n^{1-2\tau}$ is similar to the proof of Lemma A.1.3. The detail of the proof is available upon request. Q.E.D.

Lemma A.2.7 *Under Assumptions 2.1.1-2.2.1, if* $\tau < 1/2$,*then*

$$n^{1-2\tau}\widetilde{\lambda}(\widetilde{\gamma}_n - \gamma_0) = r^* \Rightarrow \arg \max_{r \in (-\infty,\infty)} (\Lambda(r) - \frac{1}{2}|r|)$$

where

$$\widetilde{\lambda} = \frac{\left(1+\rho^2\right)\left(\delta_0' \int_0^1 X(s)X'(s)ds\delta_0\right)f_0}{\sigma^2}.$$

Proof: The whole proof is similar to that of Lemma A.1.5. We replace all R_i by $\widetilde{R}_i$ for $i = 1, 2, ..., 5$. Note that

$$\begin{aligned}
& n^{-2}\sum_{t=1}^{n}\left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0)\right)\left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0)\right)' \\
= \; & n^{-2}\sum_{t=1}^{n}\left((x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) - \widehat{\rho}(x_{t-1}(\gamma_0 + \frac{\upsilon}{a_n}) - x_{t-1}(\gamma_0))\right) \\
& \left(x_t(\gamma_0 + \frac{\upsilon}{a_n}) - x_t(\gamma_0) - \widehat{\rho}(x_{t-1}(\gamma_0 + \frac{\upsilon}{a_n}) - x_{t-1}(\gamma_0))\right)' \\
\Rightarrow \; & \left(1+\rho^2\right)(F(\gamma_0 + \frac{\upsilon}{a_n}) - F(\gamma_0))\int_0^1 X(s)X'(s)ds - 2\rho(F_1(\gamma_0 + \frac{\upsilon}{a_n}, \gamma_0 + \frac{\upsilon}{a_n}) + \\
& F_1(\gamma_0,\gamma_0) - F_1(\gamma_0 + \frac{\upsilon}{a_n}, \gamma_0) - F_1(\gamma_0, \gamma_0 + \frac{\upsilon}{a_n}))\int_0^1 X(s)X'(s)ds \\
= \; & \left(1+\rho^2\right)f(\gamma_0)\frac{\upsilon}{a_n}\int_0^1 X(s)X'(s)ds + o(1).
\end{aligned}$$

Thus,

$$\widetilde{R}_1 = \delta_n' \sum_{t=2}^{n}\left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0)\right)\left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0)\right)'\delta_n$$

$$
\begin{aligned}
&= n^{1-2\tau} n^{\tau} D_n \delta_{1n}' \frac{1}{n} \sum_{t=1}^{n} D_n^{-1} \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0) \right) D_n^{-1} \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0) \right)' n^{\tau} D_n \delta_n \\
&= a_n \delta_1' \frac{1}{n} \sum_{t=1}^{n} D_n^{-1} \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0) \right) D_n^{-1} \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0) \right)' \delta_1 \\
&\Rightarrow a_n(\left(1+\rho^2\right) f(\gamma_0)(\frac{\upsilon}{a_n}) \delta_1' \int_0^1 X(s)X'(s) ds \delta_1 \\
&= \left(1+\rho^2\right) f_0 |v| \delta_0' \int_0^1 X(s)X'(s) ds \delta_0
\end{aligned}
$$

For $\widetilde{R}_2$, I have

$$
\begin{aligned}
\widetilde{R}_2 &= -2 \sum_{t=1}^{n} \widehat{\delta}_n' \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_{1t}(\gamma_0) \right) \widetilde{\eta} \\
&= -2 n^{1/2-\tau} n^{\tau+1/2} \delta_n' \frac{1}{n} \sum_{t=1}^{n} \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0) \right) e_t + o_p(1) \\
&= -2 \sqrt{a_n} \delta_0' \frac{1}{n} \sum_{t=1}^{n} \left(\widetilde{x}_t(\gamma_0 + \frac{\upsilon}{a_n}) - \widetilde{x}_t(\gamma_0) \right) e_t + o_p(1) \\
&\Rightarrow -2\sigma \sqrt{a_n} \delta_0' \int_0^1 X(s) d \left(W(s, \gamma_0 + \frac{\upsilon}{a_n}) - W(s,\gamma_0) + \rho (W_1(s, \gamma_0 + \frac{\upsilon}{a_n}) - W_1(s,\gamma_0)) \right) \\
&= -2\sigma \delta_0' \widetilde{B}(\upsilon).
\end{aligned}
$$

It can be shown other terms are asymptotically negligible, by which I complete the proof. Q.E.D.

Lemma A.2.8 *Under Assumptions 2.1.1-2.2.1, if $\tau = 1/2$, then $\widetilde{\gamma}_n \Rightarrow \widetilde{\gamma}(\gamma_0, \delta_0)$. $\widetilde{\gamma}(\gamma_0, \delta_0)$ is a random variable that maximizes $\widetilde{Q}(\gamma, \gamma_0, \delta_0)$ where*

$$
\widetilde{Q}(\gamma, \gamma_0, \delta_0) = \widetilde{\Gamma}_1(\gamma) \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix} \right)^{-1} \widetilde{\Gamma}_1(\gamma)'
$$

with

$$
\widetilde{\Gamma}_1(\gamma) = \widetilde{\Gamma}(\gamma) + \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix} \right) \delta_0,
$$

and

$$\widetilde{\Gamma}(\gamma) = \widetilde{\phi}_2(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{\phi}_1(\gamma) \\ \widetilde{\phi}_3(\gamma) \end{pmatrix}.$$

Proof: Let $\widetilde{V}_1 = (\widetilde{X}, \widetilde{Z})$, and $\widetilde{D}_{1n} = diag\{\widetilde{D}_{11}, \widetilde{D}_{33}\}$ where $\widetilde{D}_{11}, \widetilde{D}_{33}$ are components in $\widetilde{D}_n = diag\{\widetilde{D}_{11}, \widetilde{D}_{22}, \widetilde{D}_{33}\}$. After some standard algebra, I have

$$\widetilde{SSR}_n - \widetilde{SSR}_n(\gamma) = \widetilde{\delta}_n(\gamma)'(\widetilde{X}(\gamma)'(I - \widetilde{P}(\gamma))\widetilde{X}(\gamma))\widetilde{\delta}_n(\gamma)$$

where $\widetilde{P}(\gamma)$ is the projection matrix for $\widetilde{V}_1$. By plugging in

$$\widetilde{\delta}_n(\gamma) = (\widetilde{X}(\gamma)'(I - P_n)\widetilde{X}(\gamma))^{-1}\widetilde{X}(\gamma)'(I - \widetilde{P}(\gamma))\widetilde{Y}$$

I have

$$\begin{aligned} \widetilde{SSR}_n - \widetilde{SSR}_n(\gamma) &= (I - \widetilde{P}(\gamma))\widetilde{Y}'\widetilde{X}(\gamma)(\widetilde{X}(\gamma)'(I - \widetilde{P}(\gamma))\widetilde{X}(\gamma))^{-1}\widetilde{X}(\gamma)'(I - \widetilde{P}(\gamma))\widetilde{Y} \\ &= \widetilde{\Gamma}_n(\gamma)'(n^{-2}\widetilde{X}(\gamma)'(I - \widetilde{P}(\gamma))\widetilde{X}(\gamma))^{-1}\widetilde{\Gamma}_n(\gamma) \end{aligned}$$

where

$$\widetilde{\Gamma}_n(\gamma) = \frac{1}{n}\widetilde{X}(\gamma)'(I - \widetilde{P}(\gamma))\widetilde{Y}.$$

From Lemma 2.3.1, I have

$$\frac{1}{n}\widetilde{D}_{1n}^{-1}\widetilde{V}_1'\widetilde{V}_1\widetilde{D}_{1n}^{-1} \xrightarrow{p} \begin{pmatrix} \widetilde{G}_{11}(\gamma), \widetilde{G}_{13}(\gamma) \\ \widetilde{G}_{31}(\gamma), \widetilde{G}_{33}(\gamma) \end{pmatrix},$$

and

$$n^{-2}\widetilde{X}(\gamma)'\widetilde{X}(\gamma) \xrightarrow{p} \widetilde{G}_{22}(\gamma),$$

$$n^{-3/2}\widetilde{X}(\gamma)'\widetilde{V}_1\widetilde{D}_{1n}^{-1} = \begin{pmatrix} n^{-2}\widetilde{X}'(\gamma)\widetilde{X}, \\ n^{-3/2}\widetilde{X}(\gamma)'\widetilde{Z} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix},$$

$$n^{-3/2}\widetilde{D}_{1n}^{-1}\widetilde{V}_1'\widetilde{X}(\gamma) \xrightarrow{p} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}.$$

Thus,

$$n^{-2}\widetilde{X}'(\gamma)(I-\widetilde{P}(\gamma))\widetilde{X}(\gamma) \xrightarrow{p} \widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}.$$

Next, I consider the limiting behavior of $\widetilde{\Gamma}_n(\gamma)$.

$$\begin{aligned}
\widetilde{\Gamma}_n(\gamma) &= \frac{1}{n}\widetilde{X}(\gamma)'(I-\widetilde{P}(\gamma))\widetilde{Y} \\
&= \frac{1}{n}\widetilde{X}(\gamma)'\widetilde{\eta} - \frac{1}{n}\widetilde{X}(\gamma)'\widetilde{V}_1(\widetilde{V}_1'\widetilde{V}_1)^{-1}\widetilde{V}_1'\widetilde{\eta} + \frac{1}{n}\widetilde{X}(\gamma)'(I-\widetilde{P}(\gamma))\widetilde{X}(\gamma_0)\delta_n \\
&= \frac{1}{n}\widetilde{X}(\gamma)'\widetilde{\eta} - \frac{1}{n}\widetilde{X}(\gamma)'\widetilde{V}_1(\widetilde{V}_1'\widetilde{V}_1)^{-1}\widetilde{V}_1'\widetilde{\eta} + \frac{1}{n^2}\widetilde{X}(\gamma)'\widetilde{X}(\gamma_0)\delta_0 \\
&\quad -n^{-3/2}\widetilde{X}(\gamma)'\widetilde{V}_1\widetilde{D}_{1n}^{-1}(n^{-1}\widetilde{D}_{1n}^{-1}\widetilde{V}_1'\widetilde{V}_1\widetilde{D}_{1n}^{-1})^{-1}n^{-3/2}\widetilde{D}_{1n}^{-1}\widetilde{V}_1'\widetilde{X}(\gamma)\delta_0 \\
&\Rightarrow \widetilde{\Gamma}(\gamma) + \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}\right)\delta_0 = \widetilde{\Gamma}_1(\gamma).
\end{aligned}$$

Combining the above convergence results, I have

$$\begin{aligned}
\widetilde{SSR}_n - \widetilde{SSR}_n(\gamma) &\Rightarrow \widetilde{Q}(\gamma,\gamma_0,\delta_0) \\
&= \widetilde{\Gamma}_1(\gamma)'\left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}\right)^{-1}\widetilde{\Gamma}_1(\gamma).
\end{aligned}$$

Thus,

$$\widetilde{\gamma}_n = \arg\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\left(\widetilde{SSR}_n - \widetilde{SSR}_n(\gamma)\right) \Rightarrow \widetilde{\gamma}(\gamma_0,\delta_0) = \arg\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\widetilde{Q}(\gamma,\gamma_0,\delta_0).$$

Q.E.D.

Proof of Theorem 2.3.1: Combining the results from Lemma A.2.5-A.2.8, I complete the proof. Q.E.D.

Proof of Theorem 2.3.2: If $\tau < 1/2$, from Lemma A.2.6, I have $|\widetilde{\gamma}_n - \gamma_0| = o_p(1)$. Next, I will show that

$$\sqrt{n}\widetilde{D}_n(\widehat{\widetilde{\theta}}(\gamma_0) - \widehat{\widetilde{\theta}}(\widetilde{\gamma}_n)) = o_p(1)$$

and then use Lemma A.2.5 to obtain the limiting distribution of $\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n)$.

Note that

$$\begin{aligned}
\sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n) - \widetilde{\widehat{\theta}}(\gamma_0)) &= \sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n) - \widetilde{\theta}) - \sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\gamma_0) - \widetilde{\theta})) \\
&= (n^{-1}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\widetilde{\gamma}_n)\widetilde{D}_n^{-1}\widetilde{V}_t(\widetilde{\gamma}_n)')^{-1}n^{-1/2}\sum_{t=1}^{n}\widetilde{D}_n\widetilde{V}_t(\widetilde{\gamma}_n)\widetilde{\eta}_t \\
&\quad -(n^{-1}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma_0)\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma_0)')^{-1}n^{-1/2}\sum_{t=1}^{n}\widetilde{D}_n\widetilde{V}_t(\gamma_0)\widetilde{\eta}_t + o_p(1).
\end{aligned}$$

From Lemma A.2.1, I have

$$n^{-1}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\widetilde{\gamma}_n)\widetilde{D}_n^{-1}\widetilde{V}_t(\widetilde{\gamma}_n)' - n^{-1}\sum_{t=2}^{n}\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma_0)\widetilde{D}_n^{-1}\widetilde{V}_t(\gamma_0)' = O_p(\widetilde{\gamma}_n - \gamma_0) = o_p(1)$$

and

$$n^{-1/2}\sum_{t=1}^{n}\widetilde{D}_n\widetilde{V}_t(\widetilde{\gamma}_n)\widetilde{\eta}_t - n^{-1/2}\sum_{t=1}^{n}\widetilde{D}_n\widetilde{V}_t(\gamma_0)\widetilde{\eta}_t = O_p(\sqrt{|\widetilde{\gamma}_n - \gamma_0|}) = o_p(1).$$

Thus, I have

$$\sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n) - \widetilde{\widehat{\theta}}(\gamma_0)) = o_p(1).$$

It follows that

$$\sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}_n) - \widetilde{\theta}) = \sqrt{n}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\gamma_0) - \widetilde{\theta}) + o_p(1) \Rightarrow \widetilde{G}(\gamma_0)^{-1}\widetilde{\phi}(\gamma_0).$$

If $\tau = 1/2$, using a similar argument, I have

$$n^{1/2}\widetilde{D}_n(\widetilde{\widehat{\theta}}(\widetilde{\gamma}) - \widetilde{\theta}) \Rightarrow \widetilde{G}(\widetilde{\gamma}_n)^{-1}\left(\widetilde{\Pi}(\widetilde{\gamma}_n, \gamma_0, \delta_0) + \widetilde{\phi}(\widetilde{\gamma}_n)\right).$$

where

$$\widetilde{\gamma}_n \Rightarrow \widetilde{\gamma}(\gamma_0, \delta_1) = \arg\max_{\gamma\in[\underline{\gamma},\overline{\gamma}]}\widetilde{Q}(\gamma, \gamma_0, \delta).$$

Q.E.D.

Proof of Theorem 2.4.1 Let $\widetilde{V}_1 = (\widetilde{X}, \widetilde{Z})$, and $\widetilde{D}_{1n} = diag\{\widetilde{D}_{11}, \widetilde{D}_{33}\}$ where $\widetilde{D}_{11}, \widetilde{D}_{33}$ are components in $\widetilde{D}_n = diag\{\widetilde{D}_{11}, \widetilde{D}_{22}, \widetilde{D}_{33}\}$. By plugging $\widetilde{\delta}_n(\gamma)$, I have

$$\begin{aligned}\widetilde{T}_n(\gamma) &= \widetilde{\delta}_n(\gamma)'(\widetilde{X}(\gamma)'(I - P(\gamma))\widetilde{X}(\gamma))\widetilde{\delta}_n(\gamma)/\widetilde{\sigma}^2 \\ &= (I - P_n)\widetilde{Y}'\widetilde{X}(\gamma)(\widetilde{X}(\gamma)'(I - P_n)\widetilde{X}(\gamma))^{-1}\widetilde{X}(\gamma)'(I - P_n)\widetilde{Y}/\widetilde{\sigma}^2 \\ &= \widetilde{\Gamma}_n(\gamma)'(n^{-2}\widetilde{X}(\gamma)'(I - P_n)\widetilde{X}(\gamma))^{-1}\widetilde{\Gamma}_n(\gamma)/\widetilde{\sigma}^2\end{aligned}$$

where

$$\widetilde{\Gamma}_n(\gamma) = \frac{1}{n}\widetilde{X}(\gamma)'(I - P_n)\widetilde{Y}.$$

From Lemma 2.3.1, I have

$$\frac{1}{n}\widetilde{D}_{1n}^{-1}\widetilde{V}_1'\widetilde{V}_1\widetilde{D}_{1n}^{-1} \xrightarrow{p} \begin{pmatrix} \widetilde{G}_{11}(\gamma), \widetilde{G}_{13}(\gamma) \\ \widetilde{G}_{31}(\gamma), \widetilde{G}_{33}(\gamma) \end{pmatrix}$$

and

$$n^{-2}\widetilde{X}(\gamma)'\widetilde{X}(\gamma) \xrightarrow{p} \widetilde{G}_{22}(\gamma),$$

$$n^{-3/2}\widetilde{X}(\gamma)'\widetilde{V}_1\widetilde{D}_{1n}^{-1} = \begin{pmatrix} n^{-2}\widetilde{X}(\gamma)'\widetilde{X} \\ n^{-3/2}\widetilde{X}(\gamma)'\widetilde{Z} \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix},$$

$$n^{-3/2}\widetilde{D}_{1n}^{-1}\widetilde{V}_1'\widetilde{X}(\gamma) \xrightarrow{p} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}.$$

Thus,

$$n^{-2}\widetilde{X}(\gamma)'(I - P_n)\widetilde{X}(\gamma) \xrightarrow{p} \widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix}.$$

Next, I consider the limiting behavior of $\widetilde{\Gamma}_n(\gamma)$. Under the null hypothesis, $(I - P_n)\widetilde{Y} = (I - P_n)\widetilde{\eta}$, and

$$\widetilde{\Gamma}_n(\gamma) = \frac{1}{n}\widetilde{X}'(\gamma)(I - P_n)\widetilde{Y} = \frac{1}{n}\widetilde{X}(\gamma)'\widetilde{\eta} - \frac{1}{n}\widetilde{X}(\gamma)'\widetilde{V}_1(\widetilde{V}_1'\widetilde{V}_1)^{-1}\widetilde{V}_1'\widetilde{\eta}$$

$$\Rightarrow \quad \widetilde{\phi}_2(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{\phi}_1(\gamma) \\ \widetilde{\phi}_3(\gamma) \end{pmatrix} = \widetilde{\Gamma}(\gamma).$$

Combining the above convergence results, I have

$$\widetilde{T}_n(\gamma) \Rightarrow \frac{1}{\sigma^2}\widetilde{\Gamma}(\gamma)' \left(\widetilde{G}_{22}(\gamma) - \begin{pmatrix} \widetilde{G}_{21}(\gamma) \\ \widetilde{G}_{23}(\gamma) \end{pmatrix}' \begin{pmatrix} \widetilde{G}_{11}, \widetilde{G}_{13} \\ \widetilde{G}_{31}, \widetilde{G}_{33} \end{pmatrix}^{-1} \begin{pmatrix} \widetilde{G}_{12}(\gamma) \\ \widetilde{G}_{32}(\gamma) \end{pmatrix} \right)^{-1} \widetilde{\Gamma}(\gamma).$$

Q.E.D.

APPENDIX B

APPENDIX OF CHAPTER 3

B.1 Procedures to generate Size-corrected power for bootstrap tests:

Suppose W_T is a test statistic and $\{W_T^*\}$ is a set of bootstrapping test statistics $\{W_T^*\}$, which is constructed by applying the same test procedure to artificial samples obtained by drawing observations from the original sample with replacement. An ideal bootstrap test would reject the null if

$$W_T > \widehat{F}^{-1}_{W_T^*}(\alpha).$$

The consistency of the bootstrapping tests is based on the assumption that W_T^* should approximate the distribution of W_T under null very well as the number of bootstrapping grows to infinity. $\widehat{F}^{-1}_{W_T^*}(\alpha)$ is obtained from Monte Carlo simulations. However, the problem is that $\widehat{F}_{W_T^*}$ may be not exactly the same as F_{W_T}, so the rejecting power of the test is not exactly $1-\alpha$. In our case, when the sample size is finite, both linear models and nonparametric models based on finite series expansion are not exactly the true model under null. Thus, there will exist size distortions for both tests. Thus, a size-corrected power should be used. The basic idea is to continue using the $\widehat{F}_{W_T^*}$ distribution, but with the critical value that corresponds to the desired level, in which the size is correct.

Define

$$\alpha^c = \widehat{F}_{W_T^*}(\widehat{F}^{-1}_{W_T}(\alpha)). \tag{B.1}$$

Now the size corrected power bootstrap test would reject the null if

$$W_T > \widehat{F}^{-1}_{W_T^*}(\alpha^c).$$

The practical procedures for performing size-corrected power for a bootstrap test is following

Step1 Estimate $\widehat{F}^{-1}_{W_T}(\alpha)$ by Monte Carlo simulations under the null hypothesis and apply the bootstrapping methods to these simulated data to obtain the estimated distribution: $\widehat{F}_{W_T^*}$.

Step 2 Calculate α^c with the formula in Equation (B.1).

Step 3 generate a different data set under alternative hypothesis and calculate W_T. Then, generate BN bootstrapping estimators:$\{W^*_{T,b}\}_{b=1}^{BN}$. Reject the null if $W_T > \widehat{F}^{-1}_{W^*_{T,b}}(\alpha^c)$.

Step 4 repeat step 3 N times and calculate the size-corrected power using the formula

$$\frac{\#\{W_T > \widehat{F}^{-1}_{W^*_{T,b}}(\alpha^c)\}}{N}.$$

B.2 Mathematical Proof

Through out the appendix, the norm $\|\cdot\|$ for a matrix A is defined by $\|A\| = [tr(A'A)]^{1/2}$, where $tr(\cdot)$ is the trace operator. I also introduce a matrix norm $\|A\|_1 = \sup_{l:\|l\|\leq 1}\|Al\|$. Thus, when A is symmetric and positive definite, $\|A\|_1$ is the largest eigenvalue of A.

Define $\widehat{Q}_L = \frac{1}{T}\sum_{t=1}^{T} p^L(x_t)p^L(x_t)'$ and $\widehat{Q}_L(\gamma) = \sum_{t=1}^{T}\left(p^L(x_t)p^L(x_t)'1(z_t \leq \gamma)\right)/T_{1,\gamma}$ where $T_{1,\gamma} = \sum_{t=1}^{T} 1(z_t \leq \gamma)$ for any $\gamma \in [\gamma, \overline{\gamma}]$. Let Q_L and $Q_L(\gamma)$ be moment functional as follows:

$$Q = E[p^L(x_t)p^L(x_t)'],$$

$$Q_L(\gamma) = E\left[p^L(x_t)p^L(x_t)'1(z_t \le \gamma)\right].$$

Lemma B.2.1 provides the convergence rate of the estimator $\widehat{Q}_L$ and $\widehat{Q}_L(\gamma)$.

Lemma B.2.1 *Under Assumptions 3.1.1-3.2.3, the following results hold:*

$$\begin{aligned}
\|\widehat{Q}_L - Q_L\| &= O_p(\frac{\varsigma^2(L)L}{\sqrt{T}}) = o_p(1), \\
\|\widehat{Q}_L(\gamma) - Q_L(\gamma)\| &= O_p(\frac{\varsigma^2(L)L}{\sqrt{T}}) = o_p(1).
\end{aligned}$$

Proof: By stationarity of x_t, I have

$$\begin{aligned}
& E\|\frac{1}{T}\sum_{t=1}^{T} p^L(x_t)p^L(x_t)' - Q_L\|^2 \\
= & \sum_{i=1}^{L}\sum_{j=1}^{L} E\left(\frac{1}{T}\sum_{t=1}^{T} p_i^L(x_t)p_j^L(x_t) - Q_{ij}\right)^2 \\
= & \sum_{i=1}^{L}\sum_{j=1}^{L} E\left(\frac{1}{T}\sum_{t=1}^{T} (p_i^L(x_t)p_j^L(x_t) - Q_{ij})\right)^2 \\
= & \frac{1}{T}\sum_{i=1}^{L}\sum_{j=1}^{L} E\left(p_i^L(x_t)p_j^L(x_t) - Q_{ij}\right)^2 \\
& + \frac{2}{T}\sum_{i=1}^{L}\sum_{j=1}^{L}\sum_{s=1}^{T-1}(1-\frac{s}{T})cov\left(p_i^L(x_1)p_j^L(x_1), p_i^L(x_{1+s})p_j^L(x_{1+s})\right) \\
= & A_1 + A_2,
\end{aligned}$$

where Q_{ij} is the $(i,j)th$ element of the matrix Q_L. Note that $Q_{ij} = E(p_i^L(x_t)p_j^L(x_t))$, by Assumption 3.2.2, it can be shown that

$$A_1 = \frac{1}{T}\sum_{i=1}^{L}\sum_{j=1}^{L} E\left(p_i^L(x_t)p_j^L(x_t) - Q_{ij}\right)^2 \le \frac{1}{T}\sum_{i=1}^{L}\sum_{j=1}^{L} E[p_i^L(x_t)^2 p_j^L(x_t)^2] = O_p(\frac{\varsigma^4(L)L^2}{T}).$$

As for A_2, since $\beta - mixing$ implies $\alpha - mixing$, Assumption 3.1.1 indicates that $\{x_t\}$ is an $\alpha - mixing$ process with exponential decay. i.e.,

$$\sup_t \sup_{A \in F^t_{-\infty}, B \in F^\infty_{t+s}} |\Pr(A \cap B) - \Pr(A)\Pr(B)| \leq \alpha(s) \text{a.s.}$$

for any $s > 0$, and $\lim_{s \to \infty} E(\alpha(s))) = 0$ at an exponential rate. Note that $F^{t_2}_{t_1}$ is the $\sigma - field$ generated by $\{x_t : t_1 \leq t \leq t_2\}$. Furthermore, $p^L_i(\cdot)$ is Borel measurable for any i, thus, $\{p^L_i(x_t)\}$ is also $\alpha - mixing$ with the same rate (see White and Domowitz, 1984).

Moreover,

$$cov\left(p^L_i(x_1)p^L_j(x_1), p^L_i(x_{1+s})p^L_j(x_{1+s})\right) \leq E\left(p^L_i(x_1)p^L_j(x_1)p^L_i(x_{1+s})p^L_j(x_{1+s})\right) \leq E\left[\alpha(s)\varsigma^4(L)\right]$$

since $|p^L_i(x)p^L_j(x)| \leq \varsigma^2(L)$ for all i and j. It follows that

$$\begin{aligned} A_2 &= \frac{2}{T}\sum_{i=1}^{L}\sum_{j=1}^{L}\sum_{s=1}^{T-1}(1-\frac{s}{T})cov\left(p^L_i(x_1)p^L_j(x_1), p^L_i(x_{1+s})p^L_j(x_{1+s})\right) \\ &\leq \frac{2\varsigma^4(L)L^2}{T}\left(\sum_{s=1}^{T-1}(1-\frac{s}{T})\alpha(s)\right). \end{aligned}$$

Using the Kronecker lemma, $\sum_{s=1}^{T-1}(1-\frac{s}{T})\alpha(s) \to \sum_{s=1}^{\infty}\alpha(s) < \infty$ as $T \to \infty$ since $\alpha(s) \to 0$ at an exponential rate. Therefore,

$$|A_2| = O_p(\frac{\varsigma^4(L)L^2}{T})$$

which complete the proof of first result.

The proof of the second result is very similar and I skip the detail. Q.E.D.

Lemma B.2.2 *Under Assumptions 3.1.1-3.2.3, the following results hold:*

$$\begin{aligned} \|\frac{G'U}{T}\| &= O_p(\frac{\varsigma(L)L^{1/2}}{T^{1/2}}) = o_p(1), \\ \|\frac{G(\gamma)'U}{T}\| &= O_p(\frac{\varsigma(L)L^{1/2}}{T^{1/2}}) = o_p(1). \end{aligned}$$

Proof: Note that $E(p^L(x_t)u_t) = 0$ since $E(u_t|F_t, x_t, z_t) = 0$, thus

$$E(\|\frac{G'U}{T}\|^2) = T^{-2}\sum_{i=1}^{L} E(\sum_{t=1}^{T} p_i^L(x_t)u_t)^2,$$

where $T^{-1}E(\sum_{t=1}^{T} p_i^L(x_t)u_t)^2 \le E(p_i^L(x_t)u_t)^2 + 2\sum_{s=1}^{T-1}(1-\frac{s}{T})|cov(p_i^L(x_1)u_1, p_i^L(x_{1+s})u_{1+s})|$ by stationarity of x_t. Furthermore, the first term is $O_p(\varsigma(L)^2)$ since $E(p_i^L(x_t)u_t)^2 = E(p_i^L(x_t)^2 E(u_t)^2$. For the second term, note that $E(p_i^L(x_s)u_s) = 0$. Using a similar idea as the proof of Lemma B.2.1, I thus have

$$|cov(p_i^L(x_1)u_1, p_i^L(x_{1+s})u_{1+s})| \le E(p_i^L(x_1)u_1 p_i^L(x_{1+s})u_{1+s}) \le E(\alpha(s)^{1-2/r}\varsigma(L)^2(E(|u_{1+s}|^4)^{2/r},$$

where the second inequality uses the Hölder's inequality. Since $\sum_{s=1}^{T-1}(1 - \frac{s}{T})\alpha(s)^{1-2/r} < \infty$ and $(E(|u_s|^r))^2 \le E(|u_s|^{2r} < \infty$ for $r > 2$ by Assumption 3.1.1,

$$T^{-1}E(\sum_{t=1}^{T} p_i^L(x_t)u_t)^2 \le O_p(\varsigma(L)^2) + 2\sum_{s=1}^{T-1}(1-\frac{s}{T})E(\alpha(s)^{1-2/r}\varsigma(L)^2(E(|u_s|^r)^{2/r} = O_p(\varsigma(L)^2).$$

It follows immediately that

$$E(\|\frac{G'U}{T}\|^2) = T^{-2}\sum_{i=1}^{L} E(\sum_{t=1}^{T} p_i^L(x_t)u_t)^2 \le T^{-1}LO_p(\varsigma(L)^2) = O_p(\frac{\varsigma(L)^2 L}{T}).$$

which completes the proofs.

The proof of the second result is very similar and I skip the detail. Q.E.D.

Define $e_1 = (e_{11}, e_{12}, ...e_{1T})'$ and $e_1 = (e_{21}, e_{22}, ...e_{2T})'$, where e_{1t} and e_{2t} are the approximation error for $g_1(x_t)$ and $g_2(x_t)$ using the linear combination $p^L(x_t)\beta_1$. More specifically, $e_{1t} = g_1(x_t) - p^L(x_t)\beta_1$ and $e_{2t} = g_2(x_t) - p^L(x_t)\beta_2$ with β_1 and β_2 are vectors satisfying Assumption 3.2.3.

Lemma B.2.3 *Under Assumption 3.1.1-3.2.3, for $i = 1, 2$, and $\gamma \in [\gamma, \overline{\gamma}]$, the following results hold:*

$$\begin{aligned}
\|\frac{1}{T}G'e_i\| &= \|\frac{1}{T}\sum_{t=1}^{T} p^L(x_t)e_{it}\| = O_p(\frac{\varsigma(L)L^{1/2-\rho_i}}{\sqrt{T}}) = o_p(1), \\
\|\frac{1}{T}G(\gamma)'e_i\| &= \|\frac{1}{T}\sum_{t=1}^{T} p^L(x_t)I(z_t \le \gamma)e_{it}\| = O_p(\frac{\varsigma(L)L^{1/2-\rho_i}}{\sqrt{T}}) = o_p(1).
\end{aligned}$$

Proof: Because $\frac{1}{T}\left(\sum_{t=1}^{T} p_i^L(x_t)\right)^2 = O_p(\varsigma(L)^2)$, it holds that

$$
\begin{aligned}
E(\|\frac{1}{T}\sum_{t=1}^{T} p^L(x_t)e_{it}\|^2) &= \frac{1}{T^2}\sum_{i=1}^{L} E(\sum_{t=1}^{T} p_i^L(x_t)e_{it})^2 \le \frac{1}{T^2}\sum_{i=1}^{L} E\left(\sum_{t=1}^{T} p_i^L(x_t) C L^{-\rho_i}\right)^2 \\
&\le C^2 L^{1-2\rho_i}\varsigma(L)^2/T
\end{aligned}
$$

for some constant $0 < C < \infty$ with a similar argument in Lemma B.2.2. The proof of the second result is very similar and I skip the detail. Q.E.D.

Define $\Sigma_L = E(p^L(x_t)p^L(x_t)'\sigma^2(x_t))$ and $\Sigma_{L,\gamma} = E(p^L(x_t)p^L(x_t)'\sigma^2(x_t)|z_t \le \gamma)$.

Lemma B.2.4 *Under Assumptions 3.1.1-3.2.3, for* $i = 1, 2$, *and* $\gamma \in [\underline{\gamma}, \overline{\gamma}]$

$$
\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \omega'\Sigma_L^{-1/2} p^L(x_t)u_t &\Rightarrow N(0,1), \\
\frac{1}{\sqrt{T_\gamma}}\sum_{t=1}^{T} \omega'\Sigma_{L,\gamma}^{-1} p^L(x_t)I(z_t &\le \gamma)u_t \Rightarrow N(0,1),
\end{aligned}
$$

for some $L \times 1$ *fixed vector* ω *satisfying* $\|\omega\| = 1$.

Proof: Define a random variable $z_t = \omega'(\Sigma_L)^{-1/2}p^L(x_t)u_t$, then $\{z_t\}$ is a martingale difference sequence. Moreover, z_t is at most $\alpha - mixing$ with the same mixing coefficients as $\{x_t\}$. By the martingale central limit theorem, (see White, 1999, Theorem 5.24), I just need to show the following sufficient conditions for the generalized Lindeberg condition hold: for a fixed $\epsilon > 0$, $\frac{1}{T}\sum_{t=1}^{T} E(z_t^2 1(z_t^2 > T\epsilon) \xrightarrow{p} 0$ and $\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1 \xrightarrow{p} 0$. Note that $\|z_t\| \le \|(\Sigma_L)^{-1/2}\| \times \|p^L(x_t)\||u_t/\sigma| \le C_1 L^{1/2}\varsigma(L)|u_t|$ for some finite constant $C_1 > 0$. Thus, $E(z_t^4) \le C_1^4 L^2 \varsigma^4(L) E|u_t|^4 = O_p(L^2\varsigma^4(L))$ since $E|u_t|^4 < \infty$ by Assumption 3.1.2. It follows that

$$
\frac{1}{T}\sum_{t=1}^{T} E(z_t^2 1(z_t^2 > T\epsilon) \le \frac{1}{T}\sum_{t=1}^{T}\frac{E(z_t^4)}{T\epsilon} = \frac{1}{T\epsilon}C_2 L^2\varsigma^4(L)) = O_p(\frac{L^2\varsigma^4(L)}{T}) = o_p(1)
$$

by using the Cauchy-Schwartz and Chebyshev's inequalities.

Note that $E(\frac{1}{T}\sum_{t=1}^{T} z_t^2) = E(z_t^2) = 1$. To prove $\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1 \xrightarrow{p} 0$, I just need to show $E(\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1)^2 \to 0$. Moreover, I have

$$E(\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1)^2 = \frac{1}{T}E\left[(z_t^2 - 1)\right]^2 + 2\frac{1}{T^2}\sum_{s=1}^{T-1}(1 - \frac{s}{T})E\left[(z_1^2 - 1)(z_{1+s}^2 - 1)\right]$$

where

$$\frac{1}{T}E\left[(z_t^2 - 1)\right]^2 = \frac{1}{T}(E(z_t^4) - 2E(z_t^2) + 1) = O_p(\frac{L^2\varsigma^4(L)}{T}) = o_p(1).$$

and

$$E\left[(z_1^2 - 1)(z_{1+s}^2 - 1)\right] = Cov(z_1^2, z_{1+s}^2)$$

since $E(z_1^2) = E(z_{1+s}^2) = 1$. Using an analogy argument in Lemma B.2.1, I have

$$\begin{aligned} cov(z_1^2, z_{1+s}^2) &\leq \alpha(s)^{1-2/r}(E(|z_{1+s}^2|^r)^{2/r} \leq \alpha(s)^{1-2/r}(C_1^{2r}L^r\varsigma^{2r}(L)E|u_t|^{2r})^{2/r} \\ &= \alpha(s)^{1-2/r}C_1^2L^2\varsigma^4(L)(E|u_t|^{2r})^{2/r} \\ &\leq \alpha(s)^{1-2/r}C_2L^2\varsigma^4(L). \end{aligned}$$

Thus, I have

$$\begin{aligned} E(\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1)^2 &= O_p(\frac{L^2\varsigma^4(L)}{T}) + 2\frac{1}{T^2}\sum_{s=1}^{T-1}(1 - \frac{s}{T})E\left[(z_1^2 - 1)(z_{1+s}^2 - 1)\right] \\ &= O_p(\frac{L^2\varsigma^4(L)}{T}) = o_p(1). \end{aligned}$$

Therefore, $\frac{1}{T}\sum_{t=1}^{T} z_t^2 - 1 \xrightarrow{p} 0$. It follows that $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_t \Rightarrow N(0,1)$ by applying the martingale central limit theorem, which completes the proof of the first result. The proof of the second result is very similar and I skip the detail. Q.E.D.

Proof of Theorem 3.2.1: The proof is completed by applying the Theorem 1 of Newey (1997) in each regime. Q.E.D.

Define $G_1(\gamma) = GI_1(\gamma)$ and $G_2(\gamma) = GI_2(\gamma)$. Let $G_1 = GI_1(\gamma_0)$ and $G_2 = GI_2(\gamma_0)$.

Proof of Theorem 3.2.2 First note that

$$\widehat{\beta_i} = (G_i'G_i)^{-1}G_i'Y = \beta_i + (G_i'G_i)^{-1}G_i'\varepsilon$$

where $\epsilon = I_1(\gamma_0)e_1 + I_2(\gamma_0)e_2 + U$. Thus,

$$\begin{aligned}
\sqrt{T_{1,\gamma_0}}\Omega_{1,\gamma_0}^{-1/2}(\widehat{\beta_1} - \beta_1) &= \sqrt{T_{1,\gamma_0}}\Omega_{1,\gamma_0}^{-1/2}(G_1'G_1)^{-1}G_1'\varepsilon \\
&= \sqrt{T_{1,\gamma_0}}\Omega_{1,\gamma_0}^{-1/2}(G_1'G_1)^{-1}G_1'e_1 + \sqrt{T_{1,\gamma_0}}\Omega_{1,\gamma_0}^{-1/2}(G_1'G_1)^{-1}G_1'u \\
&= H_1 + H_2. \\
\sqrt{T_{2,\gamma_0}}\Omega_{2,\gamma_0}^{-1/2}(\widehat{\beta_1} - \beta_1) &= \sqrt{T_{2,\gamma_0}}\Omega_{2,\gamma_0}^{-1/2}(G_2'G_2)^{-1}G_2'\varepsilon \\
&= \sqrt{T_{2,\gamma_0}}\Omega_{2,\gamma_0}^{-1/2}(G_2'G_2)^{-1}G_2'e_2 + \sqrt{T_{2,\gamma_0}}\Omega_{2,\gamma_0}^{-1/2}(G_2'G_2)^{-1}G_2'u \\
&= H_3 + H_4.
\end{aligned}$$

It can be shown that H_1 and H_3 is asymptotically negligible since

$$\begin{aligned}
H_1 = \| \sqrt{T_{1,\gamma_0}}\Omega_{1,\gamma_0}^{-1/2}(G_1'G_1)^{-1}G_1'e_1\| &\leq \|\Omega_{1,\gamma_0}^{-1/2}(G_1'G_1)^{-1}G_1'\| * \| \sqrt{T_{1,\gamma_0}}e_1\| \\
&= O_p(L^{-\rho_1}\sqrt{T_{1,\gamma_0}}) \to 0
\end{aligned}$$

and

$$\begin{aligned}
H_3 = \| \sqrt{T_{2,\gamma_0}}\Omega_{2,\gamma_0}^{-1/2}(G_2'G_2)^{-1}G_2'e_2\| &\leq \|\Omega_{2,\gamma_0}^{-1/2}(G_2'G_2)^{-1}G_2'\| * \| \sqrt{T_{2,\gamma_0}}e_2\| \\
&= O_p(L^{-\rho_2}\sqrt{T_{2,\gamma_0}}) \to 0.
\end{aligned}$$

by Cauchy-Schwartz inequality.

By Lemma B.2.1 and Lemma B.2.4,

$$\begin{aligned}
H_2 &= \sqrt{T_{1,\gamma_0}}\omega'\Omega_{1,\gamma_0}^{-1/2}(G_1'G_1)^{-1}G_1'u \xrightarrow{d} N(0,1), \\
H_4 &= \sqrt{T_{2,\gamma_0}}\omega'\Omega_{2,\gamma_0}^{-1/2}(G_2'G_2)^{-1}G_2'u \xrightarrow{d} N(0,1).
\end{aligned}$$

It follows immediately that

$$\sqrt{T_{1,\gamma_0}}\omega'\Omega_{1,\gamma_0}^{-1/2}(\widehat{\beta_1}(\gamma_0) - \beta_1) \xrightarrow{d} N(0,1),$$

$$\sqrt{T_{2,\gamma_0}}\omega'\Omega_{2,\gamma_0}^{-1/2}(\widehat{\beta}_2(\gamma_0)-\beta_2)\xrightarrow{d}N(0,1).$$

Using a similar argument, I can show

$$\sqrt{T_{1,\gamma_0}}W_{1,\gamma_0}^{-1/2}(\widehat{g}_{1,\gamma_0}(x)-g_1(x))\xrightarrow{d}N(0,1),$$

$$\sqrt{T_{2,\gamma_0}}W_{2,\gamma_0}^{-1/2}(\widehat{g}_{2,\gamma_0}(x)-g_2(x))\xrightarrow{d}N(0,1).$$

Q.E.D.

Lemma B.2.5 *Under Assumptions 3.1.1-3.2.4,* $\widehat{\gamma}_T\xrightarrow{p}\gamma_0$.

Proof: The true model can be rewritten as

$$\begin{aligned}
Y &= I_1(\gamma_0)(G\beta_1+e_1)+I_2(\gamma_0)(G\beta_2+e_2)+U,\\
&= I_1(\gamma_0)G\beta_1+I_2(\gamma_0)G\beta_2+\epsilon\\
&= G\beta_2+I_1(\gamma_0)G(\beta_1-\beta_2)+\epsilon\\
&= G\beta_2+G_1\delta+\epsilon
\end{aligned}$$

where $\epsilon=I_1(\gamma_0)e_1+I_2(\gamma_0)e_2+U$. For a given $\gamma\in\Gamma=[\underline{\gamma},\overline{\gamma}]$, I estimate the following model

$$Y=G\widehat{\beta}_2(\gamma)+G_1(\gamma)\widehat{\delta}(\gamma)+\widehat{\epsilon}$$

where $\widehat{\beta}_2(\gamma)$ and $\widehat{\delta}(\gamma)$ are $L\times 1$ vectors of OLS coefficient estimators.

The estimator $\widehat{\gamma}_T=\arg\min_{\gamma\in[\underline{\gamma},\overline{\gamma}]}SSR_T(\gamma)$,where $S_T(\gamma)$ denotes the residual sum of squares $SSR_T(\gamma)=\left\|Y-G\widehat{\beta}_2(\gamma)-G_1(\gamma)\widehat{\delta}(\gamma)\right\|^2$. To prove the consistency of $\widehat{\gamma}$, I just need to prove that $SSR_T(\gamma)$ will uniformly converge to a function $R(\gamma)$ which takes minimum value at the true break point γ_0. It is equivalent to prove

$R_T(\gamma) = T^{-1}(SSR_T(\gamma) - u'u)$ uniformly converge to a function which takes global minimum at γ_0.

Let $G^*(\gamma) = [G, G_1(\gamma)]$ and $P^*_\gamma = G^*(\gamma)(G^*(\gamma)'G^*(\gamma))^{-1}G^*(\gamma)'$. After some standard algebra, I have

$$
\begin{aligned}
R_T(\gamma) &= T^{-1}(SSR_T(\gamma) - \epsilon'\epsilon) = T^{-1}(Y'(I - P^*_\gamma)Y' - \epsilon'\epsilon) \\
&= T^{-1}(-\epsilon' P^*_\gamma \epsilon' + 2\delta' G_1(\gamma_0)'(I - P^*_\gamma)\epsilon + \delta'(G_1(\gamma_0)'(I - P^*_\gamma)G_1(\gamma_0))\delta).
\end{aligned}
$$

By Lemma B.2.2 and Lemma B.2.3, it can be shown that, for any $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$,

$$
\begin{aligned}
T^{-1}(SSR_T(\gamma) - \epsilon'\epsilon) &= T^{-1}(Y'(I - P^*_\gamma)Y' - \epsilon'\epsilon) \\
&= T^{-1}\delta'(G_1(\gamma_0)'(I - P^*_\gamma)G_1(\gamma_0))\delta + o_p(1).
\end{aligned}
$$

Note that the projection matrix P^*_γ can be written as the projection matrix onto $[G_1(\gamma), G_2(\gamma)]$ where $G_2(\gamma) = G - G_1(\gamma)$. Given $\gamma > \gamma_0$, $G_1(\gamma_0)'G_2(\gamma) = 0$ and $G_1(\gamma_0)'G_1(\gamma) = G_1(\gamma_0)'G_2(\gamma_0)$, by Lemma B.2.1 , it can be further shown that

$$
T^{-1}\delta'(G_1(\gamma_0)'(I - P^*_\gamma)G_1(\gamma_0))\delta = \delta'(Q_1(\gamma_0) - Q_1(\gamma_0)Q_1^{-1}(\gamma)Q_1(\gamma_0))\delta \equiv R_1(\gamma).
$$

For any $\gamma > \gamma_0$,

$$
Q_1(\gamma_0) - Q_1(\gamma_0)Q_1^{-1}(\gamma)Q_1(\gamma_0) = Q_1(\gamma_0)Q_1^{-1}(\gamma)(Q_1(\gamma) - Q_1(\gamma_0))
$$

is positive definite matrix since all three matrices: $Q_1(\gamma_0), Q_1^{-1}(\gamma), (Q_1(\gamma) - Q_1(\gamma_0)$ are positive definite based on Assumption 3.2.4. Thus, $R_1(\gamma) > 0$ for any $\gamma > \gamma_0$.

Symmetrically, when $\gamma < \gamma_0$, I can show that $T^{-1}(SSR_T(\gamma) - \epsilon'\epsilon) \to R_2(\gamma)$, and $R_2(\gamma) > 0$ for any $\gamma < \gamma_0$. Moreover, when $\gamma = \gamma_0$, $T^{-1}(SSR_T(\gamma) - \epsilon'\epsilon) \to 0$. Define a function

$$
R(\gamma) = \left\{ \begin{array}{c} R_2(\gamma), \text{if} \gamma < \gamma_0 \\ 0, \text{if} \gamma = \gamma_0 \\ R_1(\gamma), \text{otherwise} \end{array} \right\}.
$$

Combining the above results, I have

$$T^{-1}(SSR_T(\gamma) - \epsilon'\epsilon) \to R(\gamma)$$

uniformly for any $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$ and $R(\gamma)$ takes minimum value at $\gamma = \gamma_0$ uniquely. In summary,

$$\widehat{\gamma} = \arg \min_{\gamma \in [\underline{\gamma}, \overline{\gamma}]} (SSR_T(\gamma)) \xrightarrow{p} \gamma_0.$$

Q.E.D.

Lemma B.2.6 *Under Assumptions 3.1.1-3.2.4,* $T(\widehat{\gamma}_T - \gamma_0) = O_p(1)$.

Proof: To prove $\widehat{\gamma}_T$ converge to γ_0 with rate T, I only need to prove, for any $\overline{v} > 0$,

$$\lim_{T \to \infty} \Pr(|\widehat{\gamma}_T - \gamma_0| \leq \overline{v}/T) = 1.$$

For each $B > 0$, define $V_B = \{\gamma : |\gamma - \gamma_0| < B\}$. When T is large enough, I have $\overline{v}/T < B$. Since $\widehat{\gamma}_T \xrightarrow{p} \gamma_0$ according to Lemma A.4, $\Pr(\{\widehat{\gamma}_T \in V_B\}) \xrightarrow{p} 1$. Therefore, I only need to examine the limiting behavior of γ in V_B. Define a subset

$$V_B(\overline{v}) = \{\gamma : \overline{v}/T < |\gamma - \gamma_0| < B\}.$$

and $V_B(\overline{v}) \subset V_B$. To prove $\Pr(|\widehat{\gamma}_T - \gamma_0| \leq \overline{v}/T) = 1$, I just need to prove $\Pr(\widehat{\gamma}_T \in V_B(\overline{v})) = 0$. Let $\widehat{\beta}_2$ and $\widehat{\delta}$ as the estimation of $\widehat{\beta}_2(\widehat{\gamma}_T)$ and $\widehat{\delta}(\widehat{\gamma}_T)$. Define $SSR_T^*(\gamma_0) = \left\| Y - G\widehat{\beta}_2 - G_1(\gamma_0)\widehat{\delta} \right\|^2$ and $SSR_T^*(\gamma) = \left\| Y - G\widehat{\beta}_2 - G_1(\gamma)\widehat{\delta} \right\|^2$. From the definition of $\widehat{\gamma}_T$, I have $SSR_T^*(\widehat{\gamma}_T) \leq SSR_T^*(\gamma_0)$. Therefore, it suffices to prove that for any $\gamma \in V_B(\overline{v})$, $SSR_T^*(\gamma) > SSR_T^*(\gamma_0)$ with probability 1.

Now, I consider the case with $\gamma > \gamma_0$. Using an argument of symmetry, I can, without loss of generality, prove the result for the case of $\gamma < \gamma_0$. Given $\gamma > \gamma_0$, it is equivalent to prove

$$\frac{SSR_T^*(\gamma) - SSR_T^*(\gamma_0)}{T(\gamma-\gamma_0)} > 0.$$

Let $p^L(x_t,\gamma) = p^L(x_t)I(z_t \le \gamma)$. Then

$$\begin{aligned}
& SSR_T^*(\gamma) - SSR_T^*(\gamma_0) \\
= & \sum_{t=1}^T (y_t - \widehat{\beta}_2' p^L(x_t) - \widehat{\delta}' p^L(x_t,\gamma))^2 - \sum_{t=1}^T (y_t - \widehat{\beta}_2' p^L(x_t) - \widehat{\delta}' p^L(x_t,\gamma_0))^2 \\
= & \sum_{t=1}^T \widehat{\delta}'(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))'\widehat{\delta} - 2\sum_{t=1}^T \widehat{\delta}'(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))\varepsilon \\
& +2\widehat{\delta}'(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))'(\widehat{\theta}-\theta) \\
= & \sum_{t=1}^T \delta'(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))'\delta - 2\widehat{\delta}'\sum_{t=1}^T (p^L(x_t,\gamma) - p^L(x_t,\gamma_0))\varepsilon \\
& +2\widehat{\delta}'\sum_{t=1}^T (p^L(x_t,\gamma) - p^L(x_t,\gamma_0))(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))'(\widehat{\theta}-\theta) + \\
& 2\sum_{t=1}^T (\widehat{\delta}+\delta)'(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))'(\widehat{\delta}-\delta) \\
\equiv & R_1 - R_2 + R_3 + R_4, \text{say.}
\end{aligned}$$

Next, I will show that

$$\frac{R_1 + R_2 + R_3 + R_4}{T(\gamma-\gamma_0)} > 0$$

almost surely. First, I have

$$\begin{aligned}
\frac{R_1}{T} &= \frac{1}{T}\sum_{t=1}^T \delta'(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))(p^L(x_t,\gamma) - p^L(x_t,\gamma_0))'\delta \\
&= \delta'(Q_L(\gamma) - Q_L(\gamma_0))\delta + o_p(1) = \delta' D\delta(\gamma-\gamma_0),
\end{aligned}$$

where the last equation uses the first order Taylor approximation of $Q_L(\gamma)$ around γ_0. Noting that $\bar{v}/T < |\gamma-\gamma_0| < B$, I have $\sqrt{\bar{v}} < \sqrt{T}\sqrt{(|\gamma-\gamma_0|)}$. Thus, there exists $k > 0$, such that

$$\frac{R_2}{T(\gamma-\gamma_0)} = \frac{2\widehat{\delta}'\frac{1}{\sqrt{T}}\sum_{t=1}^T (p^L(x_t,\gamma) - p^L(x_t,\gamma_0))\varepsilon}{\sqrt{T}(\gamma-\gamma_0)} = O_p(\frac{L}{\sqrt{T}\sqrt{(|\gamma-\gamma_0|)}}) \le kL/\sqrt{\bar{v}}.$$

Furthermore, from Lemma B.2.2, $T^{1/2}((\widehat{\beta}_2 - \beta_2) = O_p(\gamma - \gamma_0)$ and $T^{1/2}(\widehat{\delta} - \delta) = O_p(\gamma - \gamma_0)$. Thus, I can show:

$$\begin{aligned}
\frac{R_3}{T(\gamma - \gamma_0)} &= \frac{2\widehat{\delta}' T^{-1} \sum_{t=1}^{T} (p^L(x_t, \gamma) - p^L(x_t, \gamma_0))(p^L(x_t, \gamma) - p^L(x_t, \gamma_0))'(\widehat{\theta} - \theta)}{(\gamma - \gamma_0)} \\
&= O_p(\frac{L(\gamma - \gamma_0)}{\sqrt{T}}).
\end{aligned}$$

$$\begin{aligned}
\frac{R_4}{a_T(\gamma - \gamma_0)} &= \frac{2(\widehat{\delta} + \delta)' T^{-1} \sum_{t=1}^{T} (p^L(x_t, \gamma) - p^L(x_t, \gamma_0))(p^L(x_t, \gamma) - p^L(x_t, \gamma_0))'(\widehat{\delta} - \delta)}{(\gamma - \gamma_0)} \\
&= O_p(\frac{L(\gamma - \gamma_0)}{\sqrt{T}}).
\end{aligned}$$

For any $B \to 0_+$, there exist $\overline{v} > 0$ and N, such that $k/\sqrt{\overline{v}} < \frac{\delta' D_L(\gamma_0)(\gamma-\gamma_0)\delta}{L}$ and $\overline{v}/T < B$ when $T > N$. Therefore, for any $\gamma \in V_B(\overline{v})$, I have

$$\frac{R_1}{T(\gamma - \gamma_0)} - \frac{R_2}{T(\gamma - \gamma_0)} > 0,$$

and

$$\begin{aligned}
\frac{R_3}{T(\gamma - \gamma_0)} &= o_p(1), \\
\frac{R_4}{T(\gamma - \gamma_0)} &= o_p(1).
\end{aligned}$$

Combining the above results, I can show that

$$\frac{SSR_T^*(\gamma) - SSR_T^*(\gamma_0)}{T(\gamma - \gamma_0)} > 0$$

with probability 1 for any $\gamma \in V_B(\overline{v})$ and $\gamma > \gamma_0$. Similarly, I can prove $SSR_T^*(\gamma) > SSR_T^*(\gamma_0)$ when $\gamma < \gamma_0$ and $\gamma \in V_B(\overline{v})$ with probability 1. Q.E.D.

Proof of Theorem 3.2.3: Lemma B.2.5 shows the convergence of the estimator and Lemma B.2.6 establishes the result about the convergence rate. I complete the proof by combining Lemma B.2.5 and B.2.6. Q.E.D.

Proof of Theorem 3.2.4: From Lemma B.2.5, $\widehat{\gamma}_T - \gamma_0 = O_p(\frac{1}{T})$. In the following, I will show that the $\widehat{\beta}_1(\widehat{\gamma}_T)$ and $\widehat{\beta}_1(\gamma_0)$ are asymptotically equivalent and use a analogy argument to prove the equivalence of other estimators. Note that

$$\begin{aligned}
& \widehat{\beta}_1(\widehat{\gamma}_T) - \widehat{\beta}_1(\gamma_0) \\
= \; & (\widehat{\beta}_1(\widehat{\gamma}_T) - \beta_1) - (\widehat{\beta}_1(\gamma_0) - \beta_1) \\
= \; & (G_1'(\widehat{\gamma})G_1(\widehat{\gamma}))^{-1}G_1'(\widehat{\gamma})\varepsilon - (G_1'(\gamma_0)G_1(\gamma_0))^{-1}G_1'(\gamma_0)\varepsilon
\end{aligned}$$

From Lemma B.2.1, I have

$$T^{-1}G_1'(\widehat{\gamma})G_1(\widehat{\gamma}) - T^{-1}G_1'(\gamma_0)G_1(\gamma_0) \xrightarrow{p} Q_L(\widehat{\gamma}) - Q_L(\gamma_0).$$

Thus, I can show that

$$\begin{aligned}
\|\widehat{\beta}_1(\widehat{\gamma}_T) - \widehat{\beta}_1(\gamma_0)\| \; \leq \; & \|(T^{-1}G_1'(\gamma_0)G_1(\gamma_0))^{-1}\| * \|T^{-1}\left(G_1'(\widehat{\gamma}) - G_1'(\gamma_0)\right)\varepsilon\| \\
& + \|(T^{-1}G_1'(\widehat{\gamma})G_1(\widehat{\gamma}))^{-1} - \left(T^{-1}G_1'(\gamma_0)G_1(\gamma_0)\right)^{-1}\| * \|T^{-1}G_1'(\widehat{\gamma})\varepsilon\| \\
= \; & O_p(\varsigma(L)L\sqrt{|\widehat{\gamma}_T - \gamma_0|}) + O_p(\varsigma(L)|\widehat{\gamma}_T - \gamma_0|L^2) \\
= \; & O_p(\frac{\varsigma(L)L}{\sqrt{T}}) + O_p(\frac{\varsigma(L)L^2}{T}) = o_p(1).
\end{aligned}$$

Q.E.D.

Proof of Theorem 3.2.5: From Lemma B.2.5, $\widehat{\gamma}_T$ is a consistent estimator, thus, I can study its asymptotic behavior in the neighborhood of the true thresholds. Let $\gamma = \gamma_0 + \frac{\upsilon}{a_T}$, where $a_T = T^{1-2\alpha}$. From the definition of $\widehat{\gamma}_T$, I have

$$a_T(\widehat{\gamma}_T - \gamma_0) = v^* = \arg\min_v \left(SSR_T^*(\gamma_0 + \frac{\upsilon}{a_T}) - SSR_T^*(\gamma_0) \right).$$

From the definition of $SSR_T^*(\gamma_0 + \frac{\upsilon}{a_T})$ and $SSR_T^*(\gamma_0)$, I have

$$\begin{aligned}
& SSR_T^*(\gamma_0 + \frac{\upsilon}{a_T}) - SSR_T^*(\gamma_0) \\
= \; & \sum_{t=1}^{T}\left(y_t - \widehat{\beta}_2' p^L(x_t) - \widehat{\delta}' p^L(x_t, \gamma_0 + \frac{\upsilon}{a_T})\right)^2 - \sum_{t=1}^{T}\left(y_t - \widehat{\beta}_2' p^L(x_t) - \widehat{\delta}' p^L(x_t, \gamma_0)\right)^2
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=1}^{T}\widehat{\delta}'_T\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'\widehat{\delta}_T \\
&\quad -2\sum_{t=1}^{T}\widehat{\delta}'_T\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\varepsilon \\
&\quad +2\widehat{\delta}'_T\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'(\widehat{\theta}-\theta) \\
&= \delta'_T\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'\delta_T \\
&\quad -2\delta'_T\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\varepsilon \\
&\quad +2\widehat{\delta}'_T\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'(\widehat{\theta}-\theta) \\
&\quad +2\sum_{t=1}^{T}(\widehat{\delta}'+\delta'_T)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'(\widehat{\delta}-\delta_T) \\
&\quad +2(\widehat{\delta}'-\delta'_T)\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\varepsilon \\
&\equiv R_1^*+R_2^*+R_3^*+R_4^*+R_5^*, \text{ say.}
\end{aligned}
$$

Next, I turn to consider the limiting behavior of R_i^*, for $i = 1, 2, ..., 5$. I only provide the proof for the case with $v > 0$, and the proof for the other case with $v < 0$ is analogous so I skip the detail. Given $v > 0$, I have

$$
\begin{aligned}
R_1^* &= \delta'_T\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'\delta_T \\
&= T^{1-2\alpha}\delta' T^{-1}\sum_{t=1}^{T}\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)'\delta \\
&= T^{1-2\alpha}\delta'(Q_L(\gamma_0+\frac{\upsilon}{a_T})-Q_L(\gamma_0))+o_p(1) \\
&\xrightarrow{p} v\delta' D\delta,
\end{aligned}
$$

where the last equation uses the first order Taylor expansion of $Q_L(\gamma)$ around γ_0 and $a_T = T^{1-2\alpha}$. For R_2^*, I have

$$
R_2^* = -2\sum_{t=1}^{T}\widehat{\delta}'_T\left(p^L(x_t,\gamma_0+\frac{\upsilon}{a_T})-p^L(x_t,\gamma_0)\right)\varepsilon
$$

$$= \quad -2(T^{1/2-a})\delta_0' \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left(p^L(x_t, \gamma_0 + \frac{\upsilon}{a_T}) - p^L(x_t, \gamma_0) \right) \varepsilon \Rightarrow -2\delta_0' B^*(\upsilon)$$

where

$$E\left(B^*(1)\, B^*(1)'\right) = f_0 V.$$

Moreover, using a similar argument in Lemma B.2.6, it can be shown

$$R_3^* + R_4^* + R_5^* = o_p(1).$$

Combining all convergence results, I have

$$SSR_T^*\left(\gamma_0 + \frac{\upsilon}{a_T}\right) - SSR_T^*(\gamma_0) \Rightarrow v\delta' D\delta - 2\delta' B^*(\upsilon).$$

Making the change-of-variables

$$\upsilon = \frac{\delta' V \delta}{(\delta' D \delta)^2} r,$$

I have

$$SSR_T^*\left(\gamma_0 + \frac{\upsilon}{a_T}\right) - SSR_T^*(\gamma_0) \Rightarrow 2\sigma^2 (\frac{r}{2} - \Lambda_2(r))$$

where $\Lambda_2(r)$ is a standard Brownian motions defined on $[0, \infty)$.

In summary, the asymptotic distribution of $\widehat{\gamma}$ can be expressed as

$$T^{1-2\alpha}\lambda(\widehat{\gamma} - \gamma_0) = r^* \Rightarrow \arg \max_{r \in (-\infty,\infty)} (\Lambda(r) - \frac{1}{2}|r|)$$

where

$$\lambda = \frac{(\delta' D \delta)^2 f_0}{\delta' V \delta},$$

and

$$\Lambda(r) = \begin{cases} \Lambda_1(-r), \text{ if } & r < 0 \\ 0, \quad \text{if } & r = 0 \\ \Lambda_2(r), \text{ if } & r > 0 \end{cases}.$$

Q.E.D.

Proof of Theorem 3.2.6 The proof can be completed by applying the results of Theorem 3.2.4. Q.E.D.

APPENDIX C

APPENDIX OF CHAPTER 4

C.1 MLE estimation for threshold ECM

For convenience, the firm indicator i is selectively omitted in the following discussion if no misunderstanding will be caused. The threshold ECM aforementioned in Section 4.2 can be represented as follows:

$$\Delta x_t = A_1' X_{t-1} d_{1t}(\gamma) + A_2' X_{t-1} d_{2t}(\gamma) + u_t,$$

where $\Delta x_t = (p_{it}^{\mathrm{n}}, \widetilde{p}_{it}^{\mathrm{t}})$, $X_{t-1} = [1, \kappa_{t-1}, \Delta x_{t-1}, \Delta x_{t-2}, ..\Delta x_{t-m}]'$, $d_{1t}(\gamma) = 1(|\kappa_{it-1}| \leq \gamma_i)$ and $d_{1t}(\gamma) = 1(|\kappa_{it-1}| > \gamma_i)$. $1(\cdot)$ denotes the indicator function. A_1' and A_2' contains the parameters to be estimated; and γ is the threshold parameter to be estimated.

The threshold VECM model can be estimated using the MLE method proposed by Hansen and Seo (2002). Assuming that the error term u_t are i.i.d. Gaussian, the likelihood function is

$$\mathcal{L}_n(A_1, A_2, \Sigma, \gamma) = -\frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{n} u_t\,(A_1, A_2, \gamma)'\,\Sigma^{-1} u_t\,(A_1, A_2, \gamma)\,,$$

where $u_t(A_1, A_2, \gamma) = \Delta x_t - A_1' X_{t-1} d_{1t}(\gamma) - A_2' X_{t-1} d_{2t}(\gamma)$. The covariance matrix Σ is identity matrix due to the i.i.d. Gaussian assumption of the error term. For a fixed γ, A_1 and A_2 could be estimated by an OLS regression, thus

$$\widehat{A}_1(\gamma) = \left(\sum_{t=1}^{n} X_{t-1} X_{t-1}' d_{1t}(\gamma)\right)^{-1} \sum_{t=1}^{n} X_{t-1} \Delta x_t' d_{1t}(\gamma),$$

$$\widehat{A}_2(\gamma) = \left(\sum_{t=1}^{n} X_{t-1} X_{t-1}' d_{2t}(\gamma)\right)^{-1} \sum_{t=1}^{n} X_{t-1} \Delta x_t' d_{2t}(\gamma),$$

and then $\widehat{u}_t(\gamma) = \Delta x_t - \widehat{A}_1' X_{t-1} d_{1t}(\gamma) - \widehat{A}_2' X_{t-1} d_{2t}(\gamma)$. By plugging $\widehat{u}_t(\gamma)$, the likelihood function $L_n(A_1, A_2, \Sigma, \gamma)$ is simplified to be a function of γ:

$$\mathcal{L}_n(\gamma) = \frac{-n}{2}\ln\left(\frac{1}{n}\sum_{t=1}^{n}\widehat{u}_t(\gamma)\widehat{u}_t(\gamma)'\right) - \frac{n(m+2)}{2}.$$

Following Hansen (2000), the grid search method could be used to estimate the γ in an preset interval $[\underline{\gamma}, \overline{\gamma}]$. The MLE estimator for A_1 and A_2 could be obtained by inserting $\widehat{\gamma}$. To further confirm the threshold effect, I need to test the following hypothesis:

$$H_0 : A_1 = A_2 \quad \text{for any } \gamma \in [\underline{\gamma}, \overline{\gamma}].$$

The alternative is

$$H_1 : A_1 \neq A_2 \quad \text{for some } \gamma \in [\underline{\gamma}, \overline{\gamma}].$$

I use the super-Lagrange Multiplier (supLM) test (Hansen and Seo, 2002) to test above hypothesis. The LM statistic is

$$\mathcal{LM}(\gamma) = \left(\widehat{A}_1(\gamma) - \widehat{A}_2(\gamma)\right)' \left(\widehat{V}_1(\gamma) + \widehat{V}_2(\gamma)\right)^{-1} \left(\widehat{A}_1(\gamma) - \widehat{A}_2(\gamma)\right),$$

where $\widehat{V}_1(\gamma) = M_j(\gamma)^{-1}\Omega_j(\gamma)M_j(\gamma)^{-1}$, $M_j(\gamma) = I_{m+2} \otimes \Pi_j(\gamma)'\Pi_j(\gamma)$, $\Omega_j(\gamma) = \Gamma_j(\gamma)'\Gamma_j(\gamma)$, and $\Pi_j(\gamma), \Gamma_j(\gamma)$ are matrices of the stacked rows $X_{t-1}d_{jt}(\gamma)$ and $\widehat{u}_t(\gamma) \otimes X_{t-1}d_{jt}(\gamma)$ respectively. Define

$$\text{sup}\mathcal{LM} = \text{sup}_{\gamma \in [\underline{\gamma}, \overline{\gamma}]}\mathcal{LM}(\gamma).$$

A bootstrap method is used to generate the critical value since the asymptotic distribution is not standard.

BIBLIOGRAPHY

Akdeniz L., Altay-Salih L. and M.Caner (2003). Time-varying betas help in asset pricing: the threshold CAPM. *Studies in Nonlinear Dynamics & Econometrics* 6(4), 1-16.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.

Alfo, M., G. Trovato and R. J. Waldmann (2008). Testing for country heterogeneity in growth models using a finite mixture approach. *Journal of Applied Econometrics*, 23(4), 487-514.

Altissimo, F. and V. Corradi (2003). Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics* 117, 207-244.

Andrews, D.W.K.(1991). Asymptotic Normality of series estimators for nonparametric and semiparametric models. *Econometrica* 59,307-345.

Andrews, D.W.K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821-856.

Andrews, D.W.K. and W. Ploberger(1994). Inference in nonlinear econometric models with structural change. *Review of Economics Studies* 55 615-640.

Andrews, D.W.K. and C. J. McDermott (1995). Nonlinear Econometric Models with Deterministically Trending Variables. *Review of Economic Studies*, 62(3), 343-60.

Azariadis, C. and B. Smith (1998). Financial intermediation and regime switching in business cycles. *American Economic Review*, 88, 516-536.

Bai .J, H.Chen, T.T.L. Chong and S. Wang (2008). Generic Consistency of the Break-Point Estimators under Specification Errors in a Multiple-Break Model. *Econometrics Journal*, 11, 2, 287-307.

Balke, N. and T. Fomby (1997). Threshold Cointegration. *International Economics Review*, 8 627-645.

Balke, N.(2000) Credit and economic activity: credit regimes and nonlinear propagation of shocks. *Review of Economics and Statistics*, 82 344-349.

Baltagi B.H. and D. Li (2002). Series Estimation of Partially Linear Panel Data Models with Fixed Effects. *Annals of Economics and Finance* 3 103-116.

Bacidore J., Sofianos G., 2002. Liquidity provision and specialist trading in NYSE-listed non-U.S. stocks. *Journal of Financial Economics* 63, 133-158.

Bernanke B.S. and R.S. Gürkaynak 2001. Is growth exogenous? Taking Mankiw, Romer and Weil seriously. NBER Working paper 8365, Cambridge, MA.

Bierens H. and L. Martins (2010). Time-varying cointegration. *Econometric Theory*, 26, 1453-1490.

Billingsley, P. (1999). Convergence of Probability Measures, 2nd ed, Wiley, New York.

Blinder, A.S. (1987). Credit rationing and effective supply failures. *Economic Journal* 97, 327-352.

Bond S., A. Hoeffler and J. Temple (2001). GMM estimation of empirical growth models. University of Bristol discussion paper no. 01/525.

Boyd, J.H., R. Levine and B.D. Smith (2001). The impact of inflation on financial sector performance. *Journal of Monetary Economics* 47 (2), 221-248

Busetti, F. and A. Harvey (1998). Testing for the presence of a random walk in series with structural breaks. *Discussion paper No. Em/98/365* , London School of Economics and Political Science.

Cai, Z., Q. Li and J.Y. Park (2009) Functional-coefficient models for nonstationary time series data. *Journal of Econometrics,* 148 101-113.

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

Caner, M. and B.E. Hansen (2001). Threshold Autoregression with a unit root. *Econometrica* 69,1555-1596.

Caner, M. and B.E. Hansen (2004). Instrumental Variable estimation of a threshold model, *Econometric Theory*, 20, 813–843.

Caner, M. (2007). Boundedly pivotal structural change tests in continuous updating GMM with strong, weak identification and completely unidentified cases. *Journal of Econometrics*, 137, 28-67.

Caner, M. (2011). Pivotal structural change tests in linear simultaneous equations with weak identification. *Econometric Theory*, 1-14.

Chan, K., Menkveld, A.J., Yang, Z., 2008. Information asymmetry and asset prices: Evidence from the China foreign share discount. *Journal of Finance* 63, 159-196.

Chen, H., Choi, P.M.S., 2010. A trans-Niagara tale of informed traders. Working paper, Cornell University.

Chen, G., Choi, Y.K., Zhou, Y., 2008. Detections of changes in return by a wavelet smoother with conditional heteroscedastic volatility. *Journal of Econometrics* 143, 227 262

Cheng, X. (2008). Robust Confidence Intervals in Nonlinear Regression under Weak Identification. Manuscript, Department of Economics, Yale University.

Choi, C.Y., L. Hu and M. Ogaki (2008). Robust estimation for structural spurious regressions and a Hausman-type cointegration test. *Journal of Econometrics* 142, 327-351.

Choi, In and P. Saikkonen (2010). Test of nonlinear cointegration. *Econometric Theory*, forthcoming.

Choi, I. 2001. Unit Root Tests for Panel Data. *Journal of International Money and Finance* 20, 249-272.

Chung, H., Ho, T.-W., Wei, L.-J., 2005. The dynamic relationship between the prices of ADRs and their underlying stocks: Evidence from the threshold vector error correction model. *Applied Economics* 37, 2387-2394.

Delgado, M.A. and Hidalgo,J. (2000). Nonparametric inference on structural breaks, *Journal of Econometrics* 96 113-144.

De Jong, F., 2002. Measures of contributions to price discovery: A comparison. *Journal of Financial Markets* 5, 323-327.

Dickey, D.A., Fuller, W.A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Doidge, C., Karolyi, G.A., Stulz, R.M., 2009. Has New York become less competitive in global markets? Evaluating foreign listing choices over time. *Journal of Financial Economics* 91, 253-287.

Durlauf, S. N. and P. A. Johnson (1995). Multiple Regimes and Cross-country Growth Behavior. *Journal of Applied Econometrics* 10, 365-384.

Duarte, J., Young, L., 2009. Why is PIN priced? *Journal of Financial Economics* 91, 119-138.

Easley, D., Engle, R.F., O' Hara, M., Wu, L., 2008. Time-varying arrival rates of informed and uninformed trades. *Journal of Financial Ecnometrics* 6, 171-207.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? *Journal of Finance* 57, 2185-2221.

Easley, D., Kiefer, N.M., O'Hara, M., 1997. One day in the life of a very common stock. *Review of Financial Studies* 10, 805-835.

Easley, D., Kiefer, N.M., O'Hara, M., 1997. The information content of the trading process. *Journal of Empirical Finance* 4, 159-186.

Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51, 1405-1436.

Easley, D., O'Hara, M., Saar, G., 2001. How stock splits affect trading: A microstructure approach. *Journal of Financial and Quantitative Analysis* 36, 25-51.

Elliott, G, and A. Timmermann (2005). Optimal Forecast Combination Under Regime Switching , *International Economic Review* 46(4), 1081-1102.

Elliott, G. and U.K. Muller (2007). Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics* 141(2), 1196-1218.

Engle, R.F. and C.W.J. Granger (1987). Cointegration and error correction: Representation, Estimation and Testing. *Econometrica* 55, 251-276.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: Representation, estimation and testing. Econometrica 35, 251-276.

Engle, R.F., Yoo, B.S., 1987. Forecasting and testing in co-integrated systems. *Journal of Econometrics* 35, 143-159.

Eun, C., Sabherwal, S., 2003. Cross-border listings and price discovery: Evidence from U.S.-listed Canadian stocks. *Journal of Finance* 58, 549-574.

Figuerola-Ferretti, I., Gonzalo, J., 2009. Modelling and measuring price discovery in commodity markets. *Journal of Econometrics*, forthcoming.

Fuller, K., Van Ness, B., Van Ness, R. Is information risk priced for NASDAQ-listed securities? Working paper, University of Mississippi.

Gagnon, L., Karolyi, G. A., 2010. Multi-market trading and arbitrage. *Journal of Financial Economics*, forthcoming.

Garbade, K.D., Silber, W.L., 1983. Price movements and price discovery in futures and cash markets. *Review of Economics and Statistics* 65, 289-297.

Galbraith, J.W. (1996). Credit Rationing and Threshold effects in the relation between money and output. *Journal of Applied Econometrics*, 11, 419-429.

Gao, J, I. Gijbels and S. Van Bellegem (2008). Nonparametric simultaneous testing for structural breaks, *Journal of Econometrics*, 143(1), 123-142.

Garcia, R. and P. Perron (1996), An analysis of the real interest rate under regime shifts. *Review of Economics and Statistics* 78, 111-125.

Gonzalo, J., Granger, C.W.J., 1995. Estimation of common long-memory components in cointegrated systems. *Journal of Business and Economic Statistics* 13, 1-9.

Gonzalo, J. and O. Martínez, 2006. Large shocks vs. small shocks (Or does size matter? May be so). *Journal of Econometrics* 135, 311–347

Gonzalo, J. and J.Y. Pitarakis (2006). Threshold Effects in Cointegrating Regressions. *Oxford Bulletin of Economics and Statistics*, 813-833.

Gonzaloa, J. and J. Pitarakis (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics,* 110, 2, 319-352.

Grammig, J., Melvin, M., Schlag, C., 2005. Internationally cross-listed stock prices during overlapping trading hours: Price discovery and exchange rate effects. *Journal of Empirical Finance* 12, 139-164.

Grossman, S., Stiglitz, J, 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70, 393-408.

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

Fremault A., (1991). Stock index futures and index arbitrage in a rational expectation model. *Journal of Business* 64 523-547.

Hansen, B.E. (1992). Testing for parameter instability in regressions with I(1) processes. *Journal of Business and Economics Statistics* 10, 321-335.

Hansen, B.E. (1995). Rethinking the Univariate Approach to Unit Root Testing: Using Covariates to Increase Power. *Econometric Theory* 11(05), 1148-1171.

Hansen, B.E. (1996). Inference when a Nuisance Parameter is not Identified under the Null Hypothesis. *Econometrica* 64, 413-430.

Hansen, B.E. (2000). Sample Splitting and Threshold Estimation. *Econometrica* 68, 575-603.

Hansen, B.E., Seo, B., 2002. Testing for two-regime threshold cointegration in vector error correction models. *Journal of Econometrics* 110, 293-318.

Harris, F.H.deB., McInish, T.H., Shoesmith, G., Wood, R.A., 1995. Cointegration, error correction, and price discovery on informationally linked security markets. *Journal of Financial and Quantitative Analysis* 30, 563-579.

Harris, F.H.deB., McInish, T.H., Wood, R.A., 1997. Common long-memory components of intraday stock prices: A measure of price discovery. Working paper, Wake Forest University.

Harris, F.H.deB., McInish, T.H., Wood, R.A., 2002. Security price adjustment across exchanges: An investigation of common factor components for Dow stocks. *Journal of Financial Markets* 5, 277-308.

Hasbrouck, J., 1995. One security, many markets: determining the contributions to price discovery. *Journal of Finance* 50, 1175-1201.

Hasbrouck, J., 2002. Stalking the efficient price in market microstructure specifications: An overview. *Journal of Financial Markets* 5, 329-339.

Hidalgo,J. (1995). A nonparametric conditional moment test for structural stability, *Econometric Theory*, 11, 671-698.

Hong, Y. and H. White (1995). Consistent speci cation testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

Hong, S.H. and P.C.B. Phillips. Testing Linearity in Cointegrating Relations with an Application to Purchasing Power Parity. *Journal of Business and Economic Statistics*, forthcoming.

Huang, J.Z.,C.O. Wu and L. Zhou (2002). Varying-Coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements. *Biometrika*, 89,1,111-128.

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

Juhl, T. and Z. Xiao (2005). Testing for cointegration using partially linear models. *Journal of Econometrics* 124(2), 363-394.

Karlsen H.A., T. Myklebust and D.TjØstheim (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35 (1), 252-299.

Khan, M. S. and A. S. Senhadji (2001). Threshold effects in the relationship between inflation and growth. IMF Staff Papers 48(1), 1–21.

Kiefer, N.M. and T. Vogelsang (2005). A New Asymptotic Theory For Heteroskedasticity-Autocorrelation Robust Tests, *Econometric Theory*, 21(06),1130-1164.

Kim C.S. and J.Y. Park (2010). Cointegrating regressions with time heterogeneity. *Econometric Reviews* 29(4), 397-438.

Koumkwa, S., Susmel, R., 2005. Arbitrage and convergence: evidence from Mexican ADRs. Working paper, University of Houston.

Krugman, P.R., 1991. Target zones and exchange rate dynamics. *Quarterly Journal of Economics* 106, 669-682.

Kyle, A., 1985. Continuous auctions and insider trading. *Econometrica* 53, 1315-1335.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are I that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), 159-178.

Lee, C.M.C., Ready, M.J., 1991. Inferring trade direction from intraday data. *Journal of Finance* 46, 733-746.

Lee, S., M.H. Seo and Y. Shin (2010). Testing for threshold effects in regression models. CeMMAP working papers, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Li, Q., E. Maasoumi and J.S. Racine (2009). A Nonparametric Test for Equality of Distributions with Mixed Categorical and Continuous Data, *Journal of Econometrics*, 148, 2, 186-200.

Li, Q., C. Hsiao and J. Zinn (2003). Consistent specification tests for semiparametric/nonparametric models based on series estimation methods. *Journal of Econometrics*, 112, 295-325

Li, Q. (2000). Efficient estimation of additive partially linear models. *International Economic Reviews* 41, 1073–1092..

Lin, Z., D.Li and J. Chen (2008), threshold effect estimators by local polynomial fits under a dependence assumption, *Journal of Multivariate Analysis*, 99, 10, 2339-2355.

Mankiw N.G., D. Romer and D.N. Weil (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics* 107: 407–437.

McCallum, J.(1991). Credit rationing and the monetary transmission mechanism. *American Economic Review*, 81, 946-51.

McFarland, J., 2002. Report of Business. Globe and Mail, 7 October, B6.

Mũller, H.G. (1992). Change-points in nonparametric regression analysis. *Annal of statistics* 20 737-761.

Newey, W. K. and K.D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Newey, W. K.(1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* 79 147-168.

Park, J.Y. and P.C.B. Phillips (1988). Statistical inference in regressions with integrated processes, Part 1. *Econometric Theory*, 4, 468-97.

Park, J.Y. and P.C.B. Phillips (1989). Statistical inference in regressions with integrated processes, Part 2. *Econometric Theory*, 4, 468-97.

Park, J.Y. (1992). Canonical Cointegrating Regressions. *Econometrica,* 60,119-143.

Park, J.Y. and S. Hahn (1999). Cointegration regression with time varying coefficients. *Econometric Theory* 15, 664-703.

Park, J.Y. and P.C.B. Phillips (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* 5, 95-131.

Park, J.Y. and P.C.B. Phillips (2001). Nonlinear Regressions with Integrated Time Series. *Econometrica* 69(1), 117-161.

Pesaran, M. H. and A. Pick (2007). Econometric issues in the analysis of contagion. *Journal of Economic Dynamics and Control* 31(4), 1245–1277.

Phillips, P.C.B. and S.N. Durlauf (1986). Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 474-95.

Phillips, P.C.B. and J.Y. Park (1988). Asymptotic Equivalence of Ordinary Least Squares and Generalized Least Squares in Regressions With Integrated Regressors. *Journal of the American Statistical Association* 83, 111-115.

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 Loretan, M. (1991). Estimating long-run economic equilibria. *Review of Economic Studies* 58, 407–436.

Phillips, P.C.B. and D.J. Hodgson (1994). Spurious regression and generalized least squares. *Econometric Theory* 10, 957–958.

Quintos, C. and P.C.B. Phillips (1993). Parameter constancy in cointegrating regressions. *Empirical Economics* 18, 675-706.

Rabemananjara, R., and Zakoian, J. M. (1993). Threshold ARCH Models and Asymmetries in Volatility, *Journal of Applied Econometrics*, 8, 31-49.

Racine, J.S., J. Hart and Q. Li (2006), Testing the Significance of Categorical Predictor Variables in Nonparametric Regression Models. *Econometric Reviews*, 25, 4, 523-544.

Rabinovitch, R., Silva, A.C., Susmel, R., 2003. Impact of capital controls and transaction costs on the return distribution of dually traded securities: Evidence from Chile and Argentina. Working paper, University of Houston.

Robinson,P.M.(1989). Nonparametric estimation of time varying parameters. *Statistical analysis and forecasting of economic structural change*, Springer, Berlin.

Saikkonen, P. (1991). Asymptotically efficient estimation of cointegration regressions. *Econometric Theory*, 7 1-20.

Saikkonen, P. (2005). Stability results for nonlinear error correction models. *Journal of Econometrics* 127, 69–81.

Saikkonen, P. (2008). Stability of regime switching error correction models under linear cointegration. *Econometric Theory*, 24, 294-318.

Schreiber, P.S., Schwartz, R.A., 1986. Price discovery in securities markets. *Journal of Portfolio Management* 12, 43-48.

Schultz, P., Shive, S., 2008. Mispricing of dual class shares: Profit opportunities, arbitrage, and trading. Working paper, University of Notre Dame.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461-464.

Seo, B. (1998). Tests for structural change in cointegrated systems. *Econometric Theory*, 14, 222-259.

Shi, X. and P.C.B. Phillips (2010). Nonlinear Cointegration Regression under Weak Identification. Manuscript, Department of Economics, Yale University.

Solnik, B., Boucrelle, C., Le Fur, Y., 1996. International market correlation and volatility. *Financial Analysts Journal* 52, 17-34.

Solow, Robert M. (1956). A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics*, 70 (1): 65–94

Stock, J.H. and M.W. Watson (1993). A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61,783-820.

Stock and Watson (1996), Evidence on Structural Instability in Macroeconomic Time Series Relations, *Journal of Business and Economic Statistics*. January 1996; 14(1): 11-30

Stone, C.J.(1982). Optimal global rates of convergence for nonparametric regression. *Annal of statistics* 10, 1040-1053.

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

Wang, Q. and P.C.B. Phillips (2009). Structural nonparametric cointegrating regression. *Econometrica* 77 (6), 1901-1948.

White, H. (1980). A heteroskedasticity-consistent covariance matrix and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

White, H. (1999). Asymptotic Theory for Econometricians, Rev. ed., San Diego: Academic Press.

White, H., and I. Domowitz (1984). Nonlinear regression with dependent observations, *Econometrica*, 52, 143-161.

Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*152 (2), 81-92.

Yin, Y.Q. (1988). Detection of the number,locations and magnitudes of jumps. *Communications in statistics: stochastic models* 4, 45-455.

Zhou, Y., A. Wan, S. Xie and X.Wang (2010). Wavelet analysis of change-points in a non-parametric regression with heteroscedastic variance. *Journal of Econometrics*, 159 183-201.