UNIVERSITY OF SOUTHAMPTON

Faculty of Social Science
School of Mathematics Science

Inference of Nonlinear Panel Time Series Modelling with Application to Climate Financial Analysis

by

Lulu Wang

A thesis for the degree of
Doctor of Philosophy

March 2023

University of Southampton

Abstract

Faculty of Social Science
School of Mathematics Science

Doctor of Philosophy

Inference of Nonlinear Panel Time Series Modelling with Application to Climate Financial Analysis

by Lulu Wang

This thesis aims at developing proper threshold panel time series modelling approaches applicable to analysis of climate financial problems. More specifically, we are concerned with the problems of estimation and determining the number of threshold parameters efficiently for nonlinear threshold panel time series models with cross-sectional dependence so that we can study the impact of extreme climate or weather, such as heavy downpours and heatwave, on a stock market.

With the increasing concerns about climate change, there have been a number of studies, in the literature, of the weather effects on the financial markets in the financial journals. However, these studies are based on linear model time series or panel data analysis, and hence the nonlinear effects of a climate or weather variable are unable to be identified and modelled well. With this situation taken into account, we therefore propose to explore the nonlinear relationship between a climate variable such as precipitation and stock returns. By nonparametric analysis of panel stocks in the FTSE100, we first demonstrate why using threshold panel time series model that may characterise the rainfall having a significant impact on the stock market. The analysis results suggest different amounts of rainfall have diverse effects on the stocks in the London Stock Exchange market, which provide a novel insight into the relationship between financial market and weather.

The idea of threshold effect has been popular in nonlinear time series analysis. Although this idea has been extended to panel data analysis, it basically

assumes cross-sectional independence, which cannot facilitate the climate financial analysis of the shocking effect of extreme climate or weather on the stock prices or returns that are actually cross-sectionally dependent in a stock market. We, therefore, propose considering panel threshold time series regression where the processes including both regressors and error terms are allowed to be cross-sectionally dependent. In theory, we have established the asymptotic distribution of the proposed least squares based estimators under the time series length T and cross-section size n tending to infinity. The estimated coefficients are shown to be asymptotically normal with convergence rate of root-nT and asymptotic variance characterised by cross-sectional dependence, which is different from those obtained under cross-sectional independence assumptions in the literature. The asymptotic distribution for the estimators of the threshold parameters is highly non-standard. Moreover, differently from the relevant studies in the literature, we also allow the threshold effects diminishing at different rates of T and n in temporal and cross-sectional directions. Monte Carlo simulations are conducted to demonstrate the finite sample performance of the proposed estimators. Especially, the simulations further suggest that the estimated asymptotic variance ignoring cross-sectional dependence may lead to inaccurate inference, with spurious significance incurred for the estimated parameters. An empirical application to climate financial analysis is also investigated and it concludes that the heavy rainfall has a strongly negative impact on the stock market.

Another important problem arising from panel threshold time series regression analysis is how to determine the number of threshold parameters, especially considering cross-sectional dependence. The existing method suggests using the test statistics by bootstrap but it may only work well under cross-sectional independent conditions owing to potential issues with bootstrap for cross-sectional and time series dependence. In this thesis we consider the estimation and inference of threshold panel time series model via adaptive group fused Lasso letting selection of the thresholds and estimation of the models be done in a simultaneous data-driven manner. Under the assumption of both regressors and error terms allowed to be cross-sectionally dependent, we show that with probability tending to one, the suggested Lasso estimation can correctly determine the number of the thresholds and estimate the panel threshold regression parameters consistently. Monte Carlo simulations demonstrate the Lasso estimation working well in the finite samples. In particular, we further compare it with the bootstrap test statistics and it concludes that the proposed Lasso estimation works better to select threshold parameters under the cross-sectional dependent conditions. We then apply our proposed method to the climate financial problems and find multiple threshold parameters in our model.

This thesis is mainly concerned with the statistical inference for nonlinear panel time series model analysis in theory. On the empirical side, the proposed nonlinear panel time series methods can be applied widely to real data analysis, especially, to deal with the strong dependence in most climate, economic and financial data. In view of the popularity of spatio-temporal panel data analysis, it also has the great potential of extending the overall ideas of this thesis to spatio-temporal panel modelling and empirical applications following the research of this thesis.

Contents

List of Figures

List of Tables

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signed:.. Date:..................

Acknowledgements

Throughout my PhD study, words can hardly express my gratitude to my main supervisor, Professor Zudi Lu. Without the invaluable guidance, encouragement and patience of my supervisor, it would have been impossible to complete my thesis. He generously provides knowledge and expertise in statistics and steers me through the research of nonlinear time series analysis. Many thanks for his support during my research.

I also appreciate the treasured support I received from my second supervisor, Dr Maria Kyriacou, who has now joined in University of Kent. I am grateful to my independent assessors during my annual reviews, Professor Stefanie Biedermann, Professor Wei Liu and Professor Houduo Qi, for their invaluable advice to improve my work. I would also like to express my sincere thanks to my external examiner, Professor Liangjun Su, and internal examiner, Dr Chao Zheng. for their insightful suggestions and comments. In addition, I would like to thank all the staff in School of Mathematics Sciences for their help.

Finally, I would like to express my very profound gratitude to my parents, my little sister and my little brother for their unfailing support and continuous encouragement throughout my years of study. This accomplishment would not have been possible without them.

To my family.

Chapter 1

Introduction

1.1 Research Context and Background

In this project, we are concerned with statistical modelling of nonlinear panel time series data with application to modelling the effect of extreme weather on the stock market in climate financial analysis. Exploring the relationship between climate or climate change and economic activities has become a hot research topic in climate financial analysis recently. Saunders (1993) was probably the first to study the relationship between weather variables and financial market and concluded that the cloudy cover has a negative effect on stock returns. In the last few decades, since panel data analysis can be used to study both common and individual behaviours of groups, it has been developed into one of the most popular econometric tools applied in climate finance, see, e.g.,Hirshleifer and Shumway (2003); Cao and Wei (2005); Chang et al. (2006). Among the wealth of these literatures, the commonly used tool for panel market data analysis is linear panel time series regression model with fixed effect of the form,

$$R_{it} = \alpha_i + \beta X_{it} + \epsilon_{it},\ i = 1, ..., n,\ t = 1, ..., T, \tag{1.1.1}$$

where R_{it} is the return of the i-th individual stock at time t in a financial market, X_{it} is a weather covariate variable corresponding to the i-th individual at time t, which is often taken to be the same weather time series variable at the location of stock exchange, and ϵ_{it} is often called the panel corrected standard errors (PCSE) that are contemporaneously correlated and heteroscedastic across individuals. Here α_i stands for the individual fixed effect and β for the common panel effect of X_{it}.

In reality, the effects of climate or weather variables on a stock market are often rather complex, which may not be simply characterised by a linear model as simple as (1.1.1). In fact, in the literature, it has been documented that they cannot find significant linear relationship between rainfall and stock returns. Especially, Hirshleifer and Shumway (2003) pointed out that the rain did not have any influence on the stock market in their empirical study. In the past few decades, people have often suffered from heavy economic loss that torrential rains bring about. Exploring the significant rainfall effect on stock returns will provide a novel insight for finance policymakers. Therefore such applications call for further research in climate finance analysis. In fact, in Chapter 2 of this thesis, by a nonparametric analysis, we will find that the relationship between rainfall and the stock return is potentially nonlinear. Specifically, some change points are clearly displayed in the nonparametric plots of the relationship between rainfall and stock returns. Therefore this project will propose to use threshold models to study the rainfall impacts on the financial markets.

Although nonparametric regression analysis is more flexible in real data analysis, it requires the sample size sufficiently large. In addition, the nonparametric analysis result is usually not efficient in the sense of statistics and often hard to explain. Therefore nonlinear parametric time series model is often more adorable than nonparametric regression in financial and economic data analysis. Furthermore, the idea of threshold effect as suggested by Tong (1978) has become one of the most popular nonlinear time series methods. In this thesis, the empirical study of the rainfall impact on the stock market in Chapter 2 suggests that extending model (1.1.1) to a threshold panel time series model is necessary, given as follows:

$$R_{it} = \alpha_i + \beta_1 X_{it} I(X_{it} < \gamma) + (\beta_2 + \beta_3 X_{it}) I(X_{it} \geq \gamma) + \epsilon_{it}, \quad (1.1.2)$$

where $I(\cdot)$ is an indicator function, γ is the threshold parameter, X_{it} is the covariate variable, say rainfall, in the following empirical applications. For generality, we will consider the threshold panel time series model, allowing for threshold variable q_{it} that may differ from X_{it}, as follows,

$$R_{it} = \alpha_i + \beta_1 X_{it} I(q_{it} < \gamma) + (\beta_2 + \beta_3 X_{it}) I(q_{it} \geq \gamma) + \epsilon_{it}, \quad (1.1.3)$$

where if $q_{it} = X_{it}$, model (1.1.3) is equal to (1.1.2). This project applies panel threshold model and studies the rainfall effect on the stock market is nonlinear, significantly.

There have been a few research works on threshold panel time series regression analysis in the literature. However, they are also associated with, and suffer from, some shortcomings when applied to climate financial analysis. For example, the panel threshold model proposed by Hansen (1999) was studied under cross-sectional independence. It did not consider cross-sectional dependence, and hence cannot deal with the strong dependence in most economic and finance data. In fact, under a linear model structure, it has been recognised that an approach ignoring cross-sectional dependence leads to serious biased statistical inference results (Bernard, 1987; Hoechle, 2007). On the other hand, Hansen (1999) derived the asymptotic distribution theory with confidence intervals constructed for threshold parameters under a panel framework of cross sectional independence with a short time series length T that is fixed, focusing on cross-sectional size n tending to infinity. In an era of big data, the asymptotic theory under both n and T tending to infinity is desirable for us to study data analysis for climate financial applications, where both the number of individual stocks, n, and time series sample size, T, can be both large for the observations from a financial market. Thus the statistical theory established by Hansen (1999) cannot be applicable to the scenario in this project.

The asymptotic properties of threshold model are considered and established under different conditions in the past few decades. Chan (1993) established asymptotic distribution of threshold autoregressive model for time series. Hansen (2000) developed statistical theory of threshold model allowing for cross-section data or time series data by splitting approach in a non-panel data setting. Hansen (1999) required the exogeneity of both regressors and threshold variables and cross-sectional independence for panel data. Seo and Shin (2016) developed a generalized methods of moments (GMM, hereafter) estimation for dynamic panel threshold model where both regressors and threshold variables are allowed to be endogenous under cross-sectional independence conditions. Miao et al. (2020b) considered the latent group structure under cross-sectional independence and established asymptotic distribution theory. Miao et al. (2020a) moreover studied panel threshold model with interactive fixed effect under conditionally cross-sectional independence, that is, in the setting of model (1.1.3) with $\epsilon_{it} = \lambda_i^0 f_t^0 + e_{it}$ (involving the unobserved common factors $\lambda_i^0 f_t^0$ in their model), it is required in Miao et al. (2020a) that $(X_{it}, q_{it}, e_{it}), i = 1, \cdots, n$, are mutually independent of each other conditional on an unobserved information domain $D = \sigma(F^0, \Lambda^0)$, the minimal sigma-field generated from $F^0 = (f_1^0, \cdots, f_T^0)'$ and $\Lambda^0 = (\lambda_1^0, \cdots, \lambda_n^0)'$. Although this conditionally cross-sectional independence does not rule out the possibility of (unconditional) cross-sectional dependence arising from the unobserved common factors, it appears not to be that realistic to examine or check practically the existence of a common unobserved information domain D that covers all the relevant cross-sectional dependence in $(X_{it}, q_{it}, e_{it}), i = 1, \cdots, n$. Moreover, it appears to exclude the case of cross-sectional dependence like model (1.1.2) in the scenario of climate financial analysis, where X_{it} and ϵ_{it} may not share common factors owing to X_{it} as a weather variable being exogenous to the financial market.

We develop the asymptotic properties of the estimators of the slope and threshold coefficients upon the assumption of cross-sectional dependence (allowing for common factors; c.f., Pesaran (2006); Bai (2009)). For this purpose, we impose certain cross-sectional dependence conditions on the regressors and error terms. As both time series length T and cross-section size n tending to infinity, the estimators of the slope coefficients can achieve an asymptotic normality with a root-nT convergence rate. Its asymptotic variance matrix is derived in a form that is different from that under cross-sectional independence. One difficulty with statistical inference is to study the asymptotic theory of the estimator of threshold parameter since its inference is highly non-standard. We follow the lead of Hansen (1999) together with the assumption of time and cross-sectional dependence and provide the consistent but non-standard asymptotic distribution theory involving two-sided Brownian motions. Differently from those in the literature (c.f., Hansen (1999), Seo and Shin (2016), Miao et al. (2020a)), this non-standard asymptotic distribution is derived allowing the threshold effects diminishing to zero at different rates in T and n. Monte Carol simulation results demonstrate that the efficiency of proposed asymptotic theory under cross-sectional dependence with finite sample size by comparing with asymptotic variance ignoring cross-sectional dependence. We then apply the proposed method to study nonlinear weather effect on stock returns.

One of the most critical steps in multiple threshold model estimation is determining the number of threshold parameters. But the existing methods to estimate the threshold number may not be valid anymore since this project consider dependent variables. Hansen (1999) proposed a test statistics by bootstrap procedure to estimate the threshold number. However, the bootstrap method may not work well for dependent variables. More recently, Chan et al. (2015) developed group Lasso estimation of threshold autoregressive model and consistency results for time series. Li et al. (2016); Qian and Su (2016) studied adaptive group fused Lasso estimation of panel data model with structural breaks but did not consider the extension of threshold model. By far, few considered determining the number of threshold parameter for panel data taking account of cross-sectional dependence, especially, by Lasso estimation. A novel method with better efficiency to determine the number of threshold parameter for dependent variables is in desperate need.

We, therefore, develop Lasso estimation of threshold panel time series under cross-sectional dependence conditions, which eliminate the unnecessary threshold parameters by penalizing the successive difference of coefficient. In this procedure, we also consider different weights of penalty terms and group variables, combining the advantages of adaptive Lasso and group Lasso, which is more efficient than estimation proposed by Chan et al. (2015) only considering group Lasso. We develop the consistency results of Lasso estimators under cross-sectional dependence conditions. With a probability approaching one, the proposed method can correctly determine the number of threshold parameters and estimate coefficients consistently. Monte Carol simulations demonstrate Lasso estimation works well to detect the number of threshold in finite samples. Specifically, compared with the method of determining the number of threshold proposed by Hansen (1999), the developed method is more effective to deal with dependent variables. We apply Lasso to estimate multiple threshold model to study weather effect on stock returns.

1.2 Literature review

1.2.1 Panel threshold regression model

Threshold effect was first introduced in time series model, well known as threshold autoregressive model, by Tong (1978). Several threshold models have been applied in financial analysis (Chen et al., 2005; So et al., 2007; Chen et al., 2008). The panel data models have been applied in various fields such as economics, finance and sociology since they combine the advantages of time series and cross-section data. However, the traditional panel data models assume that there exists a linear relationship among variables. Thus, it is possible to get the wrong model if they are used to analyze the nonlinear relationship among variables. The threshold panel data models which combine the panel data model and nonlinearity were proposed to solve the nonlinear problems. This was first introduced by Hansen, who developed threshold regression methods for non-dynamic panels with individual-specific fixed effects (Hansen, 1999). Compared with static panel data model, dynamic panel data model considers lagged dependent variables. Therefore, the non-dynamic threshold panel data model was built as follows,

$$y_{it} = \alpha_i + \beta_1' x_{it} I(q_{it} \leq c) + \beta_2' x_{it} I(q_{it} > c) + \epsilon_{it}, \tag{1.2.1}$$

where α_i is a fixed effect, the threshold variable q_{it} is scalar, the regressor x_{it} is a k vector, $I(\cdot)$ is the indicator function and c is a threshold parameter. The analysis is asymptotic with fixed T as $n \to \infty$. The threshold variable is exogenous or at least predetermined ($q_{it} = y_{i,t-d}$ with $d \geq 1$). The elements of x_{it} and the threshold variable q_{it} are not time invariant, which means they change with time. The error term ϵ_{it} is assumed to be independent and identically distributed (iid) with $E(\epsilon_{it}) = 0$ and $var(\epsilon_{it}) = \sigma^2$. The iid assumption excludes lagged dependent variables from x_{it}. However, the iid assumption does not work in the following paper. Since the stock data and weather data are dependent over time and cross-sectionally. The model (1.2.1) can be also written as the following form:

$$y_{it} = \begin{cases} \alpha_i + \beta_1' x_{it} + \epsilon_{it}, \ q_{it} \leq c \\ \alpha_i + \beta_2' x_{it} + \epsilon_{it}, \ q_{it} > c. \end{cases} \tag{1.2.2}$$

Then, the panel threshold regression model can be viewed as a heterogeneous and time-varying parameters panel data model, where the marginal effect satisfy

$$\frac{\partial y_{it}}{\partial x_{it}} = \beta_{it} = \begin{cases} \beta_1, \ q_{it} \leq c \\ \beta_2, \ q_{it} > c \end{cases}$$

Hansen estimated the threshold parameter and the slope parameters by using the least square estimation after eliminating the fixed effects. He also developed a non-standard asymptotic theory of inference which allows the construction of confidence intervals and testing of hypotheses. Hansen applied this method to study the influence of financial constraints on investment decisions. Generally, a single threshold parameter cannot meet the needs of practical application. And the panel threshold regression model can be extended with r threshold parameters as follows,

$$y_{it} = \alpha_i + \sum_{j=1}^{r} \beta_j' x_{it} I(c_{j-1} < q_{it} \leq c_j) + \epsilon_{it} \tag{1.2.3}$$

with $c_0 = -\infty$ and $c_{r+1} = +\infty$. Bick, Nautz, et al. (2008) introduced a modified version of panel threshold regression model to study the impact of inflation on relative price variability and found two threshold parameters. Cheikh and Louhichi (2016) researched the influence of exchange rate pass-through on import prices and emphasized inflation regimes. They found two threshold points and splited the data sample into three inflation regimes. Then the panel smooth transition regression model was developed (González et al., 2005). This model can be viewed as a generalization of the panel threshold regression model of Hansen (1999). The panel smooth transition regression model can be defined as the following form:

$$y_{it} = \alpha_i + \beta_0' x_{it} + \sum_{j=1}^{r} \beta_j' x_{it} g_j(q_{it}^{(j)}; \gamma_j, c_j) + \epsilon_{it} \tag{1.2.4}$$

where $g_j(q_{it}^{(j)}; \gamma_j, c_j)$ is a transition function, $q_{it}^{(j)}$ are threshold variables, c_j are location parameters and γ_j are slope parameters. The transition function is a continuous function of the observed variable q_{it} and normalized to be bounded between 0 and 1. In this paper, the scholars considered a logistic transition function:

$$g(q_{it};\gamma,c) = \frac{1}{1+\exp\{-\gamma(q_{it}-c)\}},\ \gamma > 0.$$

There still exists another transition function, exponential transition function:

$$g(q_{it};\gamma,c) = 1-\exp\{-\gamma(q_{it}-c)^2\},\ \gamma > 0.$$

This paper estimated threshold parameters and slope parameters by nonlinear least squares and showed a modelling cycle for the PSTR model, containing tests of homogeneity, parameter constancy and no remaining nonlinearity. Then, they applied the new model to describe firms' investment decisions in the presence of capital market imperfections. Colletaz and Hurlin (2006) studied the threshold effects in the productivity of the public capital stocks by using the panel smooth threshold regression model. And the result showed that regardless of the transition function used, tests strongly rejected the linearity assumption. Moreover, they compared the results of the panel smooth transition regression model with panel threshold regression model and showed that the PSTR model can describe the threshold effect better.

Most of the existing literature proposed to estimate panel threshold model by least squares methods (see, e.g. Hansen (1999); Miao et al. (2020a,b)). Seo and Linton (2007) proposed a smoothed least squares estimation for threshold regression model in the framework of both shrinking and fixed threshold effects. In a recent paper, Seo and Shin (2016) proposed a generalized method of moments (GMM) method to estimate dynamic panel threshold regression with endogenous threshold variables and regressors. The recent works about testing the existence of threshold effect followed the lead of Hansen (1996) by the bootstrap. For example, Hansen (1999) proposed a likelihood ratio statistic for non-dynamic panel threshold model. Another commonly used method to test linearity is to use a supremum Wald statistic under the null hypothesis of no threshold effect (see, e.g. Hansen (1999); Miao et al. (2020a,b)). Yu and Fan (2021) proposed a threshold regression with a threshold boundary and developed an algorithm to ease the computation of the threshold boundary.

1.2.2 Cross-sectional dependence

Cross-sectional dependence problem is an important step in estimating panel data model. Since cross-sectional dependence would influence the estimation of parameters, unit root test and cointegration analysis in the modeling process. Ignoring cross-sectional dependence can lead to serious biased statistical results (Hoechle, 2007). Such models take into account cross-sectional dependence in panel data so that we can deal with the strong cross-sectional dependence in many finance and climate data. For example, carbon dioxide emissions of a city may influence other cities' emission, so we need to consider certain cross-sectional dependence in panel analysis. Similar features defined in empirical research are gross domestic product (GDP), stock prices and so on. Consequently, an econometric theory of cross-sectional dependence is essential to support empirical studies. In the linear modelling setting, see, e.g., Choi (2006); Breitung and Pesaran (2008); Moon and Perron (2007); Fachin (2007) for the evidence of cross- sectional dependence in many real data applications. Breitung and Pesaran (2008) pointed out that the unit root and cointegration tests in panels were developed by assuming that the error term is independent cross-sectionally. And cross-sectional dependence will influence the accuracy of statistics. However, the problem of cross-sectional dependence is difficult to deal with and need further research.

There are different definitions and classifications of cross-sectional dependence. Deistler, Anderson, Filler, Zinner, and Chen (2010) proposed the definitions that the process of variables are strongly and weakly dependent on the cross-sectional dimension. The meanings of weak and strong dependence are under the assumption that the processes are wide sense stationary with absolutely summable covariances. After that, Chudik, Pesaran, and Tosetti (2011) introduced the concepts of weak and strong dependence in large panels under their assumptions.

Assumption 2.1. Let I_t be the information set available at time t. For each $t \in I$, $Z_{Nt} = (z_{1t}, ..., z_{Nt})'$ has the conditional mean, $E(Z_{Nt}|I_{t-1}) = 0$, and the conditional variance, $Var(Z_{Nt}|I_{t-1}) = \Sigma_{Nt}$, where Σ_{Nt} is an $N \times N$ symmetric, non-negative definite matrix. The (i, j)-th element of Σ_{Nt} , denoted by $\sigma_{N,ijt}$ is bounded such that $0 < \sigma_{N,ijt} < K$, for $i = 1, 2, .., N$, where K is a finite constant independent of N.

Assumption 2.2. Let $W_{Nt} = (w_{N,1t}, ..., w_{N,Nt})'$, for $t \in I \subseteq \mathbb{E}$ and $N \in \mathbb{N}$, be a vector of non-stochastic weights. For any $t \in I$, the sequence of weight vectors $\{W_{Nt}\}$ of growing dimension $(N \to \infty)$ satisfies the 'granularity' conditions:

$$\|w_{Nt}\| = O(N^{-\frac{1}{2}}), \frac{w_{N,jt}}{\|W_{Nt}\|} = O(N^{-\frac{1}{2}}) \; for \; any \; j \in \mathbb{N}.$$

The zero mean condition in Assumption 2.1 can be relaxed to $E(Z_{Nt}|I_{t-1}) = \mu_{N,t-1}$, with $\mu_{N,t-1}$ being a pre-determined function of the elements of I_{t-1}. Consider the weighted averages, $\bar{z}_{wt} = \sum_{i=1}^{N} w_{it} z_{it} = W_t' Z_t$, for $t \in I$, where Z_t and W_t satisfy Assumption 2.1 and 2.2.

Definition (Weak and strong cross-section dependence) The process z_{it} is said to be cross-sectionally weakly dependent (CWD) at a given point in time $t \in I$ conditional on the information set I_{t-1}, if for any sequence of weight vectors $\{W_t\}$ satisfying the granularity conditions, we have

$$\lim_{N \to \infty} Var(W_t' Z_t | I_{t-1}) = 0.$$

z_{it} is said to be cross-sectionally strongly dependent (CSD) at a given point in time $t \in I$ conditional on the information set I_{t-1}, if there exists a sequence of weight vectors $\{W_t\}$ satisfying the granularity conditions and a constant K independent of N such that for any N sufficiently large (and as $N \to \infty$)

$$Var(W_t' Z_t | I_{t-1}) \leq K > 0.$$

For the test of cross-sectional dependence, the most popular method is the LM-statistics proposed by Breusch and Pagan (1980), which is based on the correlation coefficients of residuals. But this statistics is only valid for $T \to \infty$ with N fixed. Frees (1995) introduced a statistics based on the sum of squared the rank correlation coefficient. When N is much greater than T, this can be used in testing cross-sectional dependence. Then Pesaran (2021) proposed a CD-statistics based on LM-statistics. Chen, Gao, and Li (2012a) proposed a new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models. In this paper, the assumption of cross-section dependence is considered by imposing some distance function to measure cross-section i and j. Robinson (2012) introduced a nonparametric trending time-varying model for the panel data with cross-sectional dependence. Chen, Gao, and Li (2012b) extended the model by Robinson (2012) to a semiparametric trending panel data model with cross-sectional dependence. In this paper, the explanatory variables and residuals are allowed to be dependent cross-sectionally. Chen (2019) proposed a kernel hierarchical agglomerative clustering method to estimate latent group structures in time-varying coefficient panel data models with heterogeneity as the length of time series tended to be infinite considering cross-sectional dependence. Miao et al. (2020b) considered latent group structures in panel threshold regression models where both slope coefficients and threshold parameters may exhibit latent group structure and studied the asymptotic properties of estimators. However, the regressors, threshold variables and error terms were assumed to be independent of each other across i in this paper. This method only can be applicable to limited empirical studies due to cross-sectional dependence in many real panel data. Gao et al. (2020) considered heterogeneous panel data models with cross-sectional dependence. This paper proposed mean group estimators for the coefficients and showed the asymptotic consistency of proposed estimators.

1.2.3 Variable Selection and Lasso

Variable selection is one of the most important procedures in high-dimensional data analysis. It is common to consider as many explaining variables as possible in case the loss of important information. However, only few parts of these variables can provide efficient information in most cases. Thus how to select the useful variables has become increasingly popular with statisticians and economists. Optimization is also popular in investigating financial solutions (Jagannathan and Ma, 2003; Fan et al., 2012). Penalized regression is the commonly used method of variable selection. For example, consider a sparse coefficient $\beta \in R^p$, Hoerl and Kennard (1970) proposed ridge regression, which is the regression model with a penalty term called $L2$-norm. The expression is as follows,

$$\hat{\beta}_{ridge} = \arg\min \frac{1}{n}\|Y - X\beta\|^2 + \lambda\|\beta\|_2^2. \tag{1.2.5}$$

where λ is a tuning parameter. Although ridge regression shrinks the coefficients close to 0 as λ increases, any of them cannot be 0. Least absolute shrinkage and selection operator (LASSO) overcomes this drawback (Tibshirani, 1996). The Lasso estimation is defined as

$$\hat{\beta}_{lasso} = \arg\min \frac{1}{n}\|Y - X\beta\|^2 + \lambda\|\beta\|_1. \tag{1.2.6}$$

where the penalty term is $L1$-norm, which is the sum of the absolute values of coefficients. Besides, Frank and Friedman (1993) proposed bridge regression, that is defined as

$$\hat{\beta}_{bridge} = \arg\min \frac{1}{n}\|Y - X\beta\|^2 + \lambda \sum_{i=1}^{p} |\beta_i|^{\gamma}. \tag{1.2.7}$$

where $\gamma > 0$. There are two special cases for bridge estimators. If $\gamma = 2$, it is a generalization of ridge regression. The other case when $\gamma = 1$ is considered as Lasso estimation. When $0 < \gamma \leq 1$, Knight and Fu (2000) studied the asymptotic distribution theories of bridge estimators, which can shrink some parameters to exact 0 to achieve the aim of variable selection. Zou (2006) showed that the lasso was inconsistent in some certain cases and proposed adaptive lasso as follows,

$$\hat{\beta}_{adaptive} = \arg\min \frac{1}{n}\|Y - X\beta\|^2 + \lambda \sum_{j=1}^{p} \hat{w}_j |\beta_j|, \tag{1.2.8}$$

where $\hat{w}$ is a known weight vector and defined as $\hat{w} = \frac{1}{|\hat{\beta}|^{\gamma}}$ for $\gamma > 0$. Therefore although the lasso estimation can be inconsistent in some cases, the oracle properties still hold for the adaptive lasso. So far, these penalty methods are designed for the selection of individual variables. Yuan and Lin (2006) extended lasso to group lasso to select groups of variables. It is defined as,

$$\hat{\beta}_{group} = \arg\min \frac{1}{n}\|Y - X\beta\|^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|_{K_j}, \tag{1.2.9}$$

where $\|\eta\|_{K_j} = (\eta' K_j \eta)^{1/2}$. If there is only one group and $K_1 = I$, it is equivalent to ridge regression. If each covariate forms an independent group and $K_j = I$, it reduces to lasso. Tibshirani et al. (2005) proposed fused lasso for the problems with features that can be ordered in some meaningful way, such as ascending order and descending order. The penalty term includes not only the coefficients but also their successive differences. The fused lasso is given by

$$\hat{\beta}_{fused} = \arg\min \frac{1}{n}\|Y - X\beta\|^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=2}^{p} |\beta_j - \beta_{j-1}|, \qquad (1.2.10)$$

where λ_1 and λ_2 are two nonnegative tuning parameters. Friedman, Hastie, Höfling, Tibshirani, et al. (2007) proposed a coordinate-wise descent algorithms for the lasso problems and showed that it is very competitive with LARS procedure. They explored that this algorithm could be applicable to a special case of fused lasso, the fused lasso signal approximator, but did not work for general fused lasso problems. Rinaldo et al. (2009) studied the fused lasso procedure of Friedman, Hastie, Höfling, Tibshirani, et al. (2007) and proposed a different estimator, the fused adaptive lasso. Rinaldo et al. (2009) generalized that the fused adaptive lasso estimator has better asymptotic properties than fused lasso estimator. Hoefling (2010) developed a path algorithm for the fused lasso signal approximator (FLSA). Compared with other available methods, Hoefling (2010) showed that FLSA had significant advantages in speed and accuracy of gathered information. Liu et al. (2010) indicated that due to high computational complexity, the existing algorithms might not be applicable to the large-size fused lasso problem, thereby proposed an efficient fused lasso algorithm based on the fused lasso signal approximator for solving this class of optimization problems.

The idea of Lasso has been a popular method to detect change points in recent years. For example, Lee et al. (2016) proposed to use lasso estimators to select a model between linear and threshold regression models. However, this may only work for single threshold model. In multiple change-point model, Chan et al. (2014) proposed to use group lasso to estimate structural break autoregressive (SBAR) process. Chan et al. (2015) followed the idea of Chan et al. (2014) and applied group lasso to estimate threshold autoregressive model, which may be the fundamental paper of lasso for multiple threshold model. In more recently, adaptive group fused lasso has been very popular to select the number of change point by penalizing the successive differences of coefficients and achieve selection consistency for group variables by combining adaptive lasso and group lasso (see, e.g. Qian and Su (2016); Li et al. (2016)).

1.3 Organisation of the thesis

The overall objective of this project is, therefore, to develop statistical methodology and inference to build effective threshold panel time series models with application to analysing the effect of weather, say precipitation, on the stock market. The structure of the remaining chapters of this report is as follows:

In Chapter 2, we study the nonlinear relationship between precipitation and stock returns. Given the nonparametric plots, we propose to use time series model with threshold effect to study the influence of precipitation on each of the individual stock time series in FTSE 100 first. We interestingly find that there is a significant nonlinear threshold effect of rainfall on the individual stocks in resources and energy industries. According to the siginificant results with resources and energy industries, we model panel threshold time series regression to study the joint rainfall impact on the stocks of resources and energy industries in FTSE 100. Furthermore, we also empirically explore the combined effect of precipitation on all the stocks in FTSE 100 representing the London stock exchange market and conclude that the rainfall have a significant nonlinear impact on the stock market.

The obtained empirical model in Chapter 2 is actually the panel threshold regression model by Hansen (1999) with cross-sectional size n and time series length T separately. The analysis of Hansen (1999) is asymptotic with fixed T as $n \to \infty$. The asymptotic distribution analysis assumes that for each t, the residuals are independent and identically distributed (iid) across i, and for each i, they are iid over t so that it cannot be applied to the empirical study in this project. Motivated by the application, Chapter 3 studies statistical inference of the empirical model assuming residuals are dependent both over time and cross-sectionally. In Chapter 3, we first propose least square estimation method. By imposing some certain dependent conditions, we study the asymptotic distribution of proposed estimators. Some simulated examples are conducted to show the efficiency of the proposed theory in the simulation. It shows that ignoring the cross-sectional dependence would lead to spurious regression outcomes in inference of the significance of the estimated parameters. We also give the empirical application of the rainfall effect on the stock market by panel threshold model with cross-sectional dependence.

Chapter 4 develops the Lasso procedure applicable to the panel threshold regression model with multiple regimes. Hansen (1999) proposed a test statistics to determine the number of thresholds using bootstrap for independent variables. However, this method may work poor for dependent variables. Thus we propose Adaptive Group Fused Lasso estimation of panel threshold model with cross-sectional dependence. Then both consistency and selection consistency theory is established under two cases. One is the number of threshold parameters is fixed as the sample size increases. The other is the number of threshold parameters increases as the sample size increases. Simulated studies including Fused Lasso estimation of multiple threshold models and comparison to test statistics by Hansen (1999) are given. We study the empirical application of determining the number of threshold parameters in climate financial analysis.

Chapter 5 will give the conclusion of my PhD thesis with an outlook of future research.

Chapter 2

Empirical study of the threshold effect of weather on the stock market

2.1 Introduction

In recent years, more and more natural disasters such as floods, storms, droughts and other weather-related events have occurred worldwide because of global warming. As is known to all, global warming has caused a series of environmental issues. Many pieces of evidence show that the global climate is changing at an alarming rate. For example, according to NASA's Goddard Institute for Space Studies report, the average global temperature has increased by more than 1°C since 1880. What calls for special attention is that 2/3 of the rising temperature occurred since 1975, at the rate of $0.15 - 0.2$°C per decade. Over the past 140 years, the global average sea level has risen 8-9 inches. Since 1901, global precipitation has increased at the rate of 0.08 inches per decade. These numbers indicate that the global environment is deteriorating rapidly.

Climate change is also a significant risk to the global economy. Especially in 2019, extreme weather occurred frequently, and it has swept many countries worldwide. Severe weather can have a massive impact on many industries. For example, agriculture and transportation depend much on weather conditions (Adams et al., 1998; Chapman, 2007). Saunders (1993) studied the relationship between the stock price and Wall street weather by using the daily percentage changes in the Dow-Jones Industrial Average and the daily percentage changes in the NYSE index. It found a negative correlation between the stock return and the cloud-cover variable, which can be explained by behaviour finance. Since the weather conditions can affect psychological mood, investor psychology would influence asset prices. The publication of his results has caused intense reactions on Wall Street. After this a lot of literature studied the relationship between the weather variables and stock market by using different stock indexes and methods and obtained the same conclusion (Trombley, 1997; Krämer and Runde, 1997).

Given the previous psychological evidence, Hirshleifer and Shumway (2003) further studied the relationship between sunshine and stock returns by applying linear panel data model. It collected the daily market index and sunshine variables returns across 26 countries from 1982 to 1997. According to the results of the OLS regression and Logit model, the effect of good sunshine on the stock returns generally exists in 26 countries. The conclusion is similar to the research of Saunders (1993), which is the higher the cloud cover is, the lower the stock return is. This paper also added other weather factors in the model, but found that rain and snow were unrelated to the stock return after controlling for the sunshine. Although many literature indicated the negative relationship between the cloud cover and stock return, Keef and Roush (2005) got a different conclusion that the prices of bank bills were influenced by sunshine positively, but the return of the stock index was negatively affected by the wind level in Wellington. Chang et al. (2008) found that cloudy skies were associated with higher volatility in the stock market over the entire trading day and suggested that weather had a significant influence on the intraday trading behaviour of investors.

Cao and Wei (2005) studied the effect of temperature on the stock market mainly by using data from 8 countries and regions. The psychological study showed that temperature could affect mood significantly and mood changes could cause behaviour changes. The temperature change could affect the balance of apathy and aggression. The lower temperature can lead to attack resulting in more risk-taking. Therefore, it examined the statistically significant negative correlation between temperature and stock returns. Chang et al. (2006) showed that the temperature and cloud cover are two important weather factors that significantly influence the price of the Taiwan stock index. Kang et al. (2010) studied the weather effect on the Shanghai stock market by adopting weather condition dummies generated using the 21-day and 31-day moving average and its standard deviation. It revealed that although the strong weather effect on stock returns only existed in the B-share market, the weather conditions significantly influenced the volatility of both the A and B share markets.

Climate change also significantly affects rainfall (Arnbjerg-Nielsen et al., 2013). Some literature studied the influence of rain on economic growth all over the world. However, most papers concluded that little evidence suggested that the rainfall had a statistically significant effect on the increase of GDP. Brown et al. (2013) indicated that the temperature had little significant influence on the economic growth, but the precipitation was the dominant climate influence. Since the brought floods and droughts were associated with an adverse effect on GDP growth. Damania et al. (2020) demonstrated that as the data were aggregated at larger spatial scales, the significance of rainfall effect on the economic growth would decrease or even vanish. Thus, a lot of macro-econometric evidence suggested that rainfall had no significant impact on economic activities. This paper found the influence of rain was substantial in developing countries but insignificant in developed countries due to the proportion of agriculture.

In recent decades, the weather conditions in Britain is getting worse due to the influence of global warming. According to the latest UK state of the climate report, the climate in 2010-2019 has been on average 1% and 5% wetter than 1981-2010 and 1961-1990 separately. More specifically, the rainfall and floods are increased in the UK. For example, the winter of 2013-2014 is the wettest on record in the UK. The highest daily rainfall (13.44 in) and largest river flows (around 1700 cubic meters per sec) recorded in the winter of 2015-2016, according to data released by the center for Ecology and Hydrology. Although climate finance is a popular research topic nowadays, few literature study the rainfall effect on the stock market.

We first study the impact of rainfall on stock returns with our collected data, FTSE100 and daily precipitation in London, by linear time series model. However, few significant results obtained. Then,in this Chapter, with the data by preliminary nonparametric analysis, we interestingly find that some change points appear on the nonparametric plots of the relationship between each stock return in FTSE100 and daily precipitation in London, which show that different regimes of London precipitation may have different significant influence on stock returns. Therefore, we will empirically propose to use a nonlinear threshold model to do the time series analysis of the influence of precipitation on each of the individual stock time series in FTSE100. Also, we will use panel threshold model to study the impact of rainfall on London stock market. The existing panel threshold models still have some unavoidable shortages and defects due to without considering cross-sectional dependence, which leave us a problem to solve in the next chapter.

2.2 Weather and FTSE100 data

Our investigation is based on the influence of weather variables in London on the stock market (London stock exchange). All the weather data are collected from the National Center for Environmental Information (https://www.ncei.noaa.gov/). Here, we use the daily weather data for temperature in Fahrenheit and precipitation including rainfall and melted snow in inch in London from 01/01/2001 to 31/12/2017. All the stock market data are collected from the datastream.

The FTSE 100 constituent companies were chosen as the research subjects. The stocks of 100 representative listed companies were chosen to construct the FTSE 100 index which is an important index to measure the performance of London stock exchange. Considering the collected stock prices are nonstationary and following the conventional approach, daily stock returns are calculated as the logarithmic difference in the daily stock price, $R_t = \ln P_t - \ln P_{t-1}$, where P_t and P_{t-1} are the daily closing price of the stock on the day t and $t-1$, respectively.

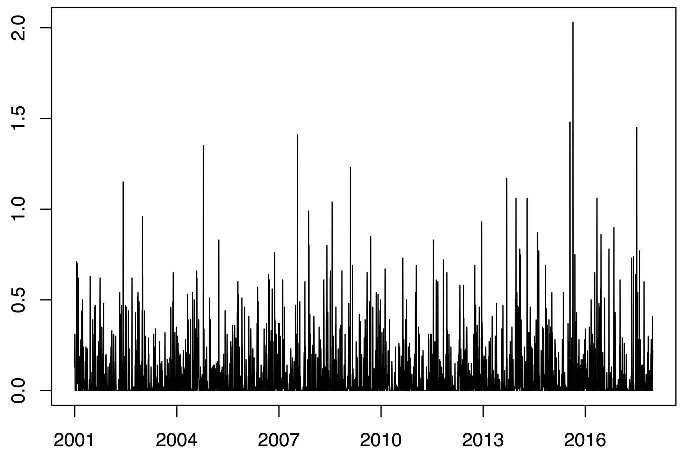
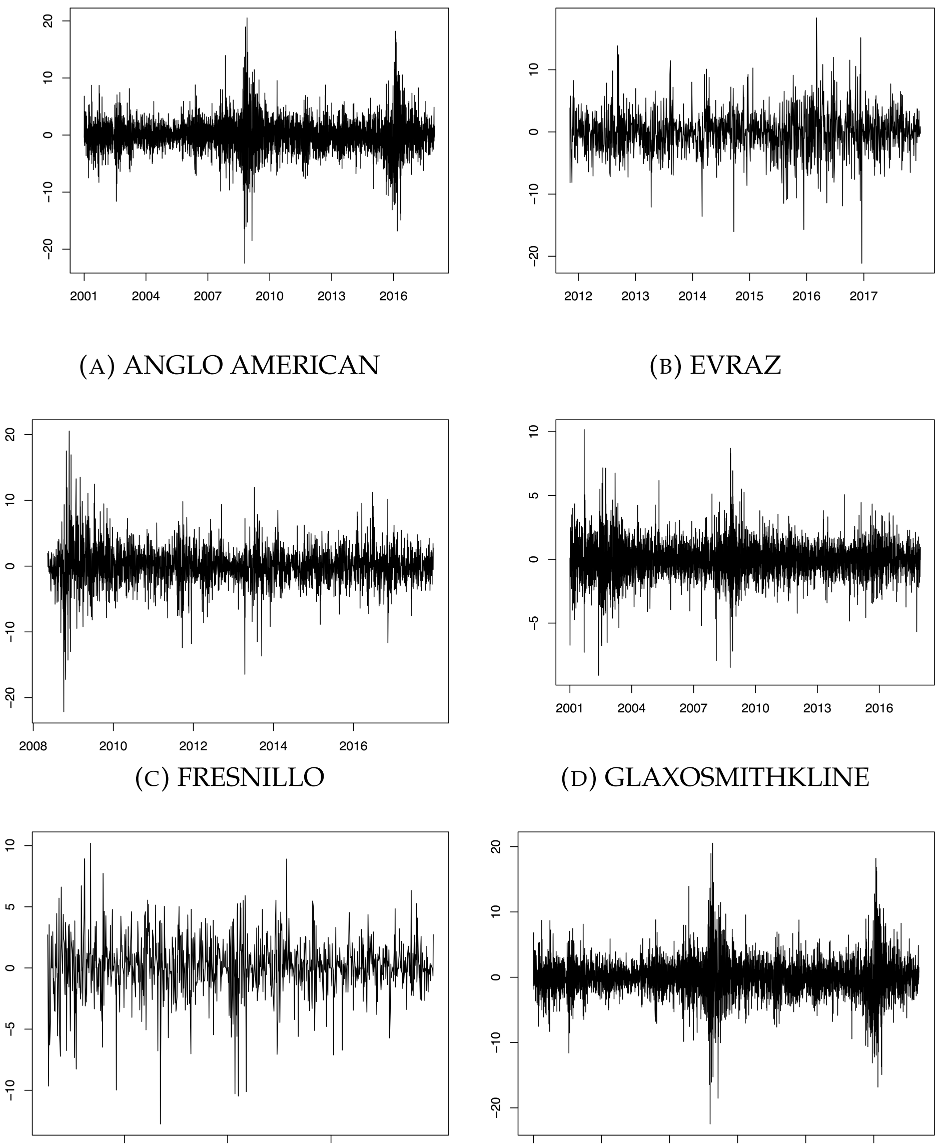


FIGURE 2.1: Plot of daily precipitation in inch in London between 2001 and 2017



(A) ANGLO AMERICAN (B) EVRAZ

(C) FRESNILLO (D) GLAXOSMITHKLINE

(E) JUST.EAT (F) X3I.GROUP

FIGURE 2.2: Plots of daily stock returns

Figure 2.1 shows the precipitation in London from 2001 to 2017. Then, we use descriptive statistics to analyze the sample data. The sample mean and medium of precipitation are 0.0667 inch and 0.0100 inch, respectively. The maximum value of precipitation is 2.03000 and the minimum value is 0. Figure 2.2 gives 6 example of stock returns in FTSE 100 constituent stocks, which is calculated by logarithm difference. Due to different listed date, time series lengths are different.

2.3 Preliminary analysis

If the regression analysis uses the nonstationary data, it may cause spurious regression. Spurious regression is a regression that provides the statistically significant evidence for two unrelated variables. Therefore, we should use the unit root tests to examine the stationarity of the sample data first. Consider the following model:

$$x_t = c_t + \delta x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + \epsilon_t, \tag{2.3.1}$$

where x_t is the time series, and it is the stock return or precipitation in this report. c_t is the certain function of time index t, $\Delta x_j = x_j - x_{j-1}$ is the difference of x_t. Then the unit root hypothesis is $\delta = 1$. The model can be written as the following form:

$$\Delta x_t = c_t + \rho x_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta x_{t-i} + \epsilon_t, \tag{2.3.2}$$

where $\rho = \delta - 1$. Therefore, the hypothesis is $\rho = 0$ equally. Some results of the unit root test can be obtained by running the above Augmented Dickey-Fuller regression. And the ADF value is the t-statistics of ρ. We incorporate the Augmented Dickey-Fuller (ADF) test of precipitation. The Dicker-Fuller value is -14.254, and the p-value is 0.01. The result of Augmented Dicker-Fuller test shows that the precipitation is stationary.

The results of unit root tests show that the stock market returns of every FTSE 100 constituent company and the precipitation in London are all stationary time series. We first use the traditional linear regression model to test the relationship between the stock market returns and precipitation. However, the results of linear regression models are not significant. Therefore, there is no evidence to show the linear relationship between the stock returns and the precipitation. Then, we use the nonparametric method to estimate the nonlinear relationship between the stock returns and the precipitation. Nonparametric regression can solve the unknown relationship problem well and thus widely used in finance area (Aıt-Sahalia and Lo, 2000; Barndorff-Nielsen and Shephard, 2001; Chen and Tang, 2005). Consider the nonparametric regression model as follows,

$$Y_i = g(X_i) + \epsilon_i,\ i = 1, .., n \tag{2.3.3}$$

where the function g is unknown, ϵ_i is the error term with zero mean. Watson (1964) and Nadaraya (1964) proposed the kernel estimate and proved that the regression function is the best estimator based on mean square error. Given a kernel K and a bandwidth h, the Nadaraya-Watson kernel estimator is defined as follows:

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} Y_i K(\frac{X_i - x}{h})}{\sum_{i=1}^{n} K(\frac{X_i - x}{h})}. \tag{2.3.4}$$

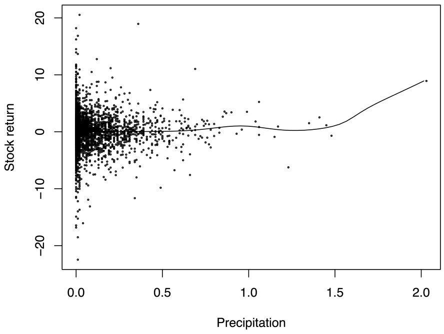
Here, the nonparametric regression method we used in sm package is the local linear estimation. In contrast to Nadaraya-Watson estimator, local linear estimator preserves linear data. More specifically, local linear estimator performs better at boundary than Nadaraya-Watson estimator, since it fits a least square line near the boundary. The local linear estimators are obtained by the following,

$$\min \sum_{i=1}^{n} (Y_i - g(x) - g'(x)(X_i - x))^2 K(\frac{X_i - x}{h}), \tag{2.3.5}$$

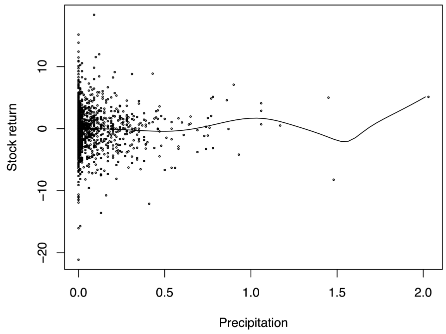
where Y_i is the dependent variables, which is the stock return. X_i is the independent variables, which is the daily precipitation. K is a kernel function. And in this paper, we use the Gaussian kernel function, which is

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}x^2\}. \tag{2.3.6}$$

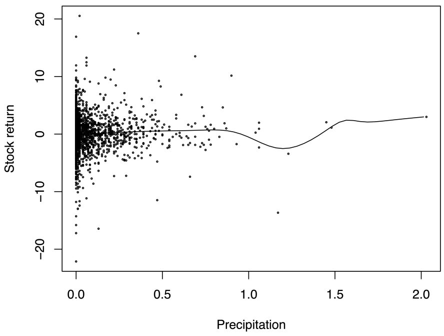
The plots of estimated regression for every FTSE 100 constituent company can be obtained by R (sm package). Figure 2.3 gives several typical nonlinear relationships between the stock returns and precipitation in FTSE 100 companies. From the estimated nonparametric plots, there are some change points obviously, such as 1.0 and 1.5. To some extent, 0.5 is also a change point, although not all plots exhibit this change point. The nonparametric plots indicate that different amounts of rainfall may have different influence on the stock market. Generally, when the daily precipitation is more than 1 inch, the weather can be defined as heavy rain. Therefore, it is evident that the stock market is sensitive to heavy rain. That is to say, the extreme weather has certain effect on stock returns. Due to the evident threshold effect in the plots, we suggest to use the nonlinear threshold model to examine the relationship between the stock returns and the precipitation.



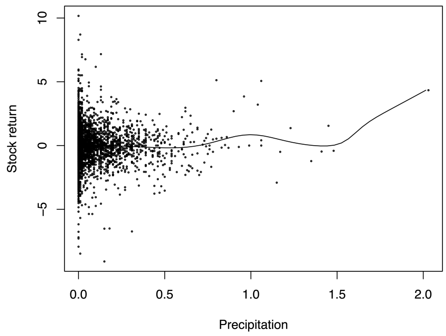
(A) ANGLO AMERICAN



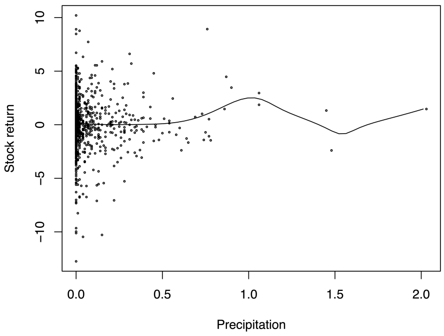
(B) EVRAZ



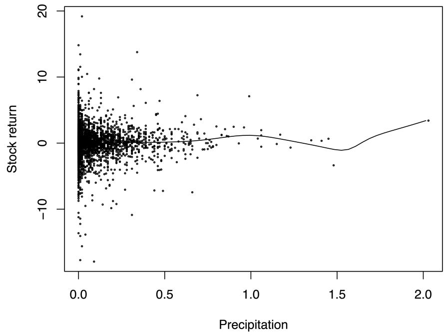
(C) FRESNILLO



(D) GLAXOSMITHKLINE



(E) JUST.EAT



(F) X3I.GROUP

FIGURE 2.3: Preliminary nonparametric analysis between precipitation and stock returns

2.4 Preliminary time series results

In order to study the relationship between the stock returns and the precipitation more precisely, we suggest three threshold parameters about precipitation following Figure 2.3, which are 0.5,1 and 1.5 respectively.These threshold parameters are actually the change points we find at previous analysis. Thus, we propose the following threshold model,

$$\begin{aligned} R_t =& \alpha_1 I(P_t < 0.5) + \beta_1 P_t I(P_t < 0.5) + \alpha_2 I(0.5 \leq P_t < 1) \\ &+ \beta_2 P_t I(0.5 \leq P_t < 1) + \alpha_3 I(1 \leq P_t < 1.5) + \beta_3 P_t I(1 \leq P_t < 1.5) \\ &+ \alpha_4 I(P_t \geq 1.5) + \beta_4 P_t I(P_t \geq 1.5) + \epsilon_t, \end{aligned} \tag{2.4.1}$$

where R_t is the daily stock return of FTSE 100 constituent company, P_t is the daily precipitation in London, and $I(\cdot)$ is the indicator function. Suppose c is a threshold parameter, $I(P_t < c) = 1$ if $P_t < c$, otherwise, $I(P_t < c) = 0$. From the analysis results, it is obvious that the heavy rain has a significant impact on most stocks. And the largest number of significant results is the constant term, α_4.

Table 2.1 shows that 26 stocks have significant results of α_4 in the FTSE 100 constituent companies. Among these 26 companies, most are resource and energy enterprises. The results indicate that if the daily observed rainfall is around 1.5 inch, 25 companies have positive returns on stocks. Only one of the 26 stocks has the negative return. In this project, we try to examine that extreme weather can affect the stock returns. From the analysis results, we also find that if the daily precipitation is between 1.0 inch and 1.5 inch, 2 stocks will be influenced significantly by the heavy rainfall. The following table shows the analysis results of the precipitation between 1.0 inch and 1.5 inch.

Table 2.2 indicates that when the daily rainfall is 1.0 inch, both 2 stocks have positive returns. However, as the precipitation increases until 1.5 inch, both the stock returns will decrease. Pearson plc is a British multinational publishing and education company and its stock return is 11.12156% at the point 1.0 inch. When the rainfall increases by 0.1 inch, this stock return will decrease by 0.983840%. As an insurance company, in the interval [1,1.5), when the precipitation increases by 0.1 inch, the stock return of Phoenix Group will reduce by 0.9209901%. Although we try to examine the relationship between the extreme

TABLE 2.1: Significant Results of α_4 of threshold time series model

Stock	α_4	t-Statistics
Anglo American	8.92^{**}	3.06
Antofagasta	8.53^{**}	3.17
Astrazeneca	4.03^{*}	2.42
BHP Group	8.81^{***}	3.59
BP	5.86^{***}	3.60
Carnival	3.97^{*}	2.00
CRH	5.04^{*}	2.38
Glaxosmithkline	4.34^{**}	3.09
HSBC Holding	4.63^{**}	2.78
ICTL.HTLS.GP.	3.86^{*}	2.13
Johnson Matthey	3.89^{*}	2.02
Marks Spencer Group	4.49^{*}	2.44
National Grid	3.08^{*}	2.35
Next	4.38^{*}	2.35
Phoenix Group HDG.	-3.84^{*}	-2.43
Reckitt Benckiser Group	3.62^{**}	2.66
Rio Tinto	6.55^{*}	2.56
Royal Dutch Shell A	4.86^{**}	3.09
Royal Dutch Shell B	4.56^{**}	2.77
Scottish Mortgage	4.22^{**}	2.84
Smith Nephew	4.14^{*}	2.50
Smiths Group	3.98^{*}	2.44
SSE	2.78^{*}	2.10
Standard Chartered	6.79^{**}	2.77
TUI	3.86^{*}	2.46
Vodafone Group	3.87^{*}	2.11

Note: *p-value < 0.1;**p-value < 0.05;***p-value < 0.01

TABLE 2.2: Significant Results of α_3 and β_3 of threshold time series model

Stock	α_3	β_3
Pearson	11.12	-9.84
	$(2.72)^{**}$	$(-3.00)^{**}$
Phoenix Group HDG.	12.23	-9.21
	$(2.82)^{**}$	$(-2.61)^{**}$

Notes:
1.*p-value < 0.1;**p-value < 0.05;***p-value < 0.01
2.The numbers in parentheses are the t-statistics.

weather and the stock market in this paper, we also find that 7 stocks will be influenced by the moderate rain. Here, we define the weather is moderate rain if the precipitation is between 0.5 and 1.0 inch. Table 2.3 gives the details of the analysis results of moderate rain.

TABLE 2.3: Significant Results of α_2 and β_2 of threshold time series model

Stock	α_2	β_2
British Land	-3.08	4.77
	$(-2.70)^{**}$	$(2.76)^{**}$
Evraz	-7.67	10.80
	$(-2.33)^{*}$	$(2.25)^{*}$
Glaxosmithkline	-2.40	3.59
	$(-2.82)^{**}$	$(2.79)^{**}$
Land Securities Group	-2.28	3.71
	$(-2.25)^{*}$	$(2.42)^{*}$
Prudential	-3.65	5.92
	$(-2.36)^{*}$	$(2.53)^{*}$
Scottish Mortgage	-1.83	2.88
	$(-2.03)^{*}$	$(2.12)^{*}$
Segro	-2.36	3.99
	$(-2.08)^{*}$	$(2.33)^{*}$

Notes:
1.*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01
2.The numbers in parentheses are the t-statistics.

Table 2.3 indicates that all 7 stocks have negative returns at 0.5 inch. However, with the precipitation increasing, all 7 stock returns will increase. There are 3 real estate companies in these 7 stocks, which are British Land, Land Securities Group and Segro, respectively. Therefore with the daily rainfall increasing, the stock returns of real estate companies will increase.

TABLE 2.4: Significant Results of α_1 and β_1 of threshold time series model

Stock	α_1	β_1
Astrazeneca	0.11	-0.64
	$(2.22)^{*}$	$(-1.99)^{*}$
Diageo	0.11	-0.54
	$(2.96)^{**}$	$(-2.15)^{*}$
Severn Trent	0.09	-0.62
	$(2.33)^{*}$	$(-2.45)^{*}$

Notes:
1.*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01
2.The numbers in parentheses are the t-statistics.

Table 2.4 indicates the impact of light rain on the stock market. Here, we define that the weather where the precipitation is between 0 and 0.5 inch is light rain. Table 2.4, we can see only 3 stocks have significant results. Compared with the significant results of other coefficients, the light rain has a relatively little performance impact on these 3 stocks. When the daily precipitation increases by 0.1 inch, the stock returns of 3 companies will reduce by around 0.06%.

2.5 Panel threshold regression analysis

2.5.1 Analysis of resources and energy industry stocks

From the preliminary results, it is obvious that the extreme weather has a significant influence on the stock market. Especially, when the daily precipitation is around 1.5 inch, there are 26 stocks having significant results. And in these 26 stocks, there are 8 stocks which belong to resources and energy industries. They are Anglo American plc, Antofagasta plc, BHP Group plc, BP plc, Rio Tinto, Royal Dutch Shell plc and SSE plc, respectively. Although there are 7 companies, Royal Dutch Shell plc have two stocks.

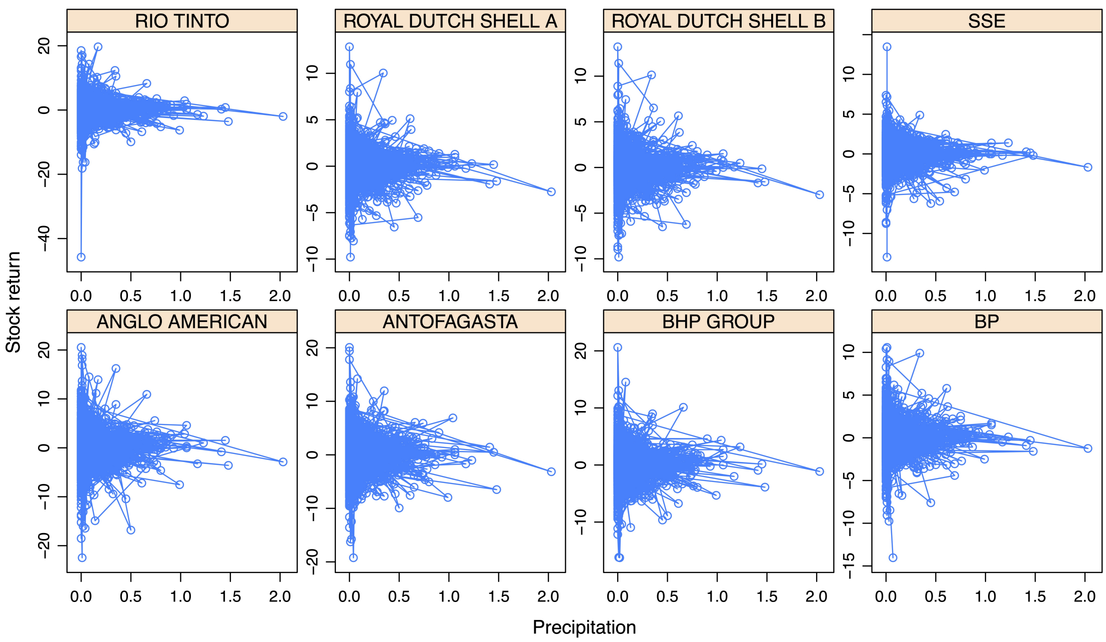


FIGURE 2.4: Scatter plot of resources and energy industry stocks

For analyzing the industrial structure better, we use the panel data model to examine the relationship between the resources and energy industries and the precipitation. In this paper, we use the balanced panel data. Although different stocks hold different time to market, we choose the same time series length for each stock. Here, the price of 8 stocks is from 20/01/2005 to 31/12/2017. As before, the daily return for each stock is calculated as the logarithmic difference in the daily stock price, $r_t = \ln P_t - \ln P_{t-1}$. To avoid spurious regression, the stationarity of the stock return and precipitation in the panel data should be examined by unit root test. Augmented Dickey Fuller test is one of the most common methods of unit root test. The ADF value of stock return in the panel data is -29.609 and the p-value is 0.01. On the other hand, the ADF value of precipitation is -25.339 and the p-value is 0.01, too. It is obvious that both the stock return and precipitation in the panel data are stationary. Figure 2.4 is the scatter plot between the stock return and precipitation of 8 stocks. When the precipitation is smaller than 1.0 inch, all the points are messy. However, when the precipitation is greater than 1.0 inch, there is a certain pattern in the figure. According to the previous section, the same threshold parameters are suggested in panel data model. Therefore, the panel threshold regression model is considered as follows,

$$\begin{aligned} R_{it} =& \alpha_1 I(P_t < 0.5) + \beta_1 P_t I(P_t < 0.5) + \alpha_2 I(0.5 \leq P_t < 1) \\ &+ \beta_2 P_t I(0.5 \leq P_t < 1) + \alpha_3 I(1 \leq P_t < 1.5) + \beta_3 P_t I(1 \leq P_t < 1.5) \\ &+ \alpha_4 I(P_t \geq 1.5) + \beta_4 P_t I(P_t \geq 1.5) + \gamma_i + \epsilon_{it}, \end{aligned} \tag{2.5.1}$$

where R_{it} is the stock return, and P_t is the daily precipitation, $I(\cdot)$ is the indicator function, γ_i is the individual effect. It can be a fixed effect or random effect. If we ignore the influence of the change of individuals or time on error terms, the model is the panel data model with the fixed effect. However, the fixed effect models only consider the certain parameters of individual effect and time effect. In consideration of unobserved uncertain parameters, the panel data model with the random effect can be established. Generally, the regression parameters are fixed in the fixed effect model. It has n intercept terms for n individuals. The random effect model assumes that the individual specific effects are uncorrelated with regressors. The assumption of fixed effect models is that the individual specific effects are correlated with regressors.

Table 2.5 and 2.6 are the analysis results of panel threshold regression model with the fixed effect and the random effect for the resources and energy industries, respectively. The intercept term of the interval greater than 1.5 is deleted

TABLE 2.5: Fixed effect model for the resources and energy industries

Coefficient	estimator	t-statistics	p-value
α_1	2.28045**	2.8080	0.0049916
β_1	-0.29335	-1.6072	0.1080331
α_2	2.51242*	2.5203	0.0117383
β_2	-0.50688	-0.5894	0.5555746
α_3	9.40810***	4.4992	6.878×10^{-6}
β_3	-5.76988***	-3.6826	0.0002318

Notes:*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01

TABLE 2.6: Random effect model for the resources and energy industries

Coefficient	estimator	t-statistics	p-value
α_1	2.28045**	2.8085	0.0049767
β_1	-0.29335	-1.6075	0.1079471
α_2	2.51242*	2.5207	0.0117120
β_2	-0.50688	-0.5895	0.5554947
α_3	9.40810***	4.5000	6.795×10^{-6}
β_3	-5.76988***	-3.6832	0.0002303

Notes:*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01

due to multicollinearity. Since there is only one point where the precipitation is greater than 1.5 inch, the analysis cannot get effective results for the precipitation greater than 1.5 inch. As a result, the tables show no result of α_4 and β_4. From both tables, it is obvious that the value of the coefficients of both models are the same. However, their t-statistics and P-values are a bit different. Therefore, we consider to apply the Hausman test to choose a more effective one.

Generally, Hausman test is applied to choosing a better one from the panel data model with the fixed effect and the random effect (Hausman, 1978). Let the error term of the panel data model $\epsilon_{it} = \lambda_i + u_{it}$, where λ_i is the individual effect in the error term. One of the panel regression assumptions is that the effect is uncorrelated with the explained variables, which is $E(\epsilon_{it}|x_{it}) = 0$. When this assumption cannot be satisfied, the coefficient of the panel data model with the random effect $\hat{\beta}_{RE}$ is biased and inconsistent. But the panel data model with the fixed effect cannot be influenced by it. Therefore, $E(\epsilon_{it}|x_{it}) = 0$ is the null hypothesis of the Hausman test. The alternative hypothesis is $E(\epsilon_{it}|x_{it}) \neq 0$.

Under the null hypothesis of no correlation between the regressors and effects, there should be no difference between the estimators. Given the random effect model estimator $\hat{\beta}_{RE}$ and the fixed effect model estimator $\hat{\beta}_{FE}$, its variance of their difference can be calculated as $\hat{\Sigma} = var(\hat{\beta}_{RE} - \hat{\beta}_{FE})$. The Hausman statistics can be built as follows:

$$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})'\hat{\Sigma}^{-1}(\hat{\beta}_{RE} - \hat{\beta}_{FE}) \sim \chi^2(k). \tag{2.5.2}$$

If H statistics is significant, we reject the null hypothesis and the panel data model with the fixed effect is feasible. Otherwise, we should choose the panel data model with the random effect as the object of study. In this project, based on the Hausman test of these two models, the Hausman statistics is 2.3883×10^{-19} and the p-value is 1. It is obvious that we should not reject the null hypothesis and believe that the panel data model with the random effect is a better model. However, in this empirical study, the estimated coefficient are the same between fixed effect model and random effect model. Hausman test may be not a good choice to a proper model between fixed and random effect models.

From Table 2.6, when the precipitation is around 0 inch and 0.5 inch, the stock returns are all significant. Especially, when the daily rainfall is between 1.0 inch and 1.5 inch, the significant model can be obtained. If the precipitation increases by 0.1 inch, the stock return of resources and energy industries will decrease by 0.576988%.

2.5.2 Analysis of FTSE 100 constituent stocks

We have already examined that heavy rain has a great effect on the resources and energy industries and the analysis result is statistically significant in last section. Due to the above, we use the panel regression model only for the resources and energy industries. A following question is how about the impact of the precipitation on the FTSE 100 constituent stocks representing the London stock exchange market. Similarly, we use the nonparametric estimation method to get the plot of the relationship between the FTSE 100 index return as Figure 2.5.

First, we try to explore the effect of the precipitation on the FTSE 100 index by using data from 01/01/2001 to 31/12/2017. The daily return is calculated as the logarithmic difference in the daily stock index.

According to the unit root test for the FTSE 100 index return, the Dickey-Fuller

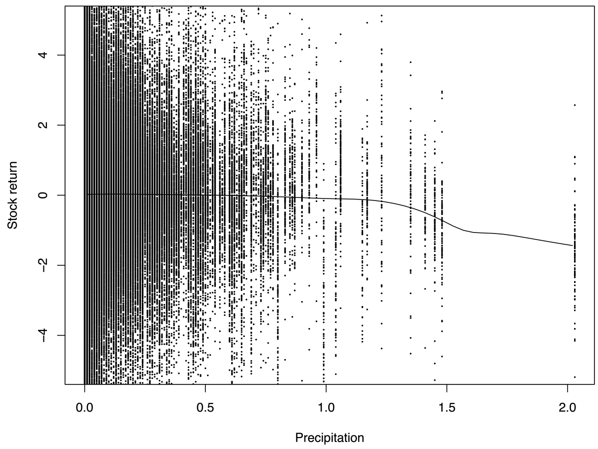


FIGURE 2.5: Plot of nonparametric analysis between precipitation and FTSE 100 index return

statistics is -16.843, and p-value is 0.01. So the FTSE 100 index return is stationary. It is obvious that we can not get new threshold parameters. Therefore, we use the same time series model as Section 2.3 to explore the effect of rainfall on the FTSE 100 index.

TABLE 2.7: Time series results for the FTSE 100 index

Coefficient	estimator	t-statistics	p-value
α_1	0.01517	0.432	0.66572
β_1	-0.07748	-0.337	0.73601
α_2	-1.56103^*	-2.158	0.03104
β_2	2.52744^*	2.312	0.02086
α_3	0.56688	0.207	0.83602
β_3	-0.49615	-0.224	0.82306
α_4	3.49762^{**}	2.927	0.00345

Notes:*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01

From Table 2.7, it is obvious that the moderate rain has a significant impact on the FTSE 100 index. When the precipitation is between 0.5 inch and 1 inch, if the daily rainfall increases by 0.1 inch, the FTSE 100 index return will increase by 0.252744%. However, there is still one significant coefficient. When the daily precipitation is around 1.5 inch, the FTSE 100 index return will be around 3.49762%, which illustrates that the heavy rain also has a certain effect on the FTSE 100 index.

Now we use the same panel model as Section 2.5.1 to explore the threshold effect on the whole FTSE 100 constituent companies. There are 101 stocks in the FTSE 100 index. Because different stocks have different time to market, we choose the latest time to market in 101 stocks to build the balanced panel data. The latest stock in the FTSE 100 constituent companies was on the market on 18/03/2015. Therefore, the chosen panel data is from 18/03/2015 to 31/12/2017. Both the stock return and precipitation are stationary.

TABLE 2.8: Fixed effect model for the FTSE 100 constituent companies

Coefficient	estimator	t-statistics	p-value
α_1	1.50232^{***}	8.4300	2.2×10^{-16}
β_1	-0.35470^{***}	-3.50252	0.0004614
α_2	-0.33093	-1.1008	0.2709764
β_2	2.52703^{***}	7.0145	2.350×10^{-12}
α_3	5.92346^{***}	7.9813	1.494×10^{-15}
β_3	-3.67801^{***}	-6.8585	7.079×10^{-12}

Notes:*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01

TABLE 2.9: Random effect model for the FTSE 100 constituent companies

Coefficient	estimator	t-statistics	p-value
α_1	1.50232^{***}	8.4348	2.2×10^{-16}
β_1	-0.35470^{***}	-3.5045	0.0004574
α_2	-0.33093	-1.1015	0.2706954
β_2	2.52703^{***}	7.0185	2.242×10^{-12}
α_3	5.92346^{***}	7.9858	1.396×10^{-15}
β_3	-3.67801^{***}	-6.8624	6.772×10^{-12}

Notes:*p-value$<$ 0.1;**p-value$<$ 0.05;***p-value$<$ 0.01

Table 2.8 and 2.9 are the analysis results for the FTSE 100 constituent companies. They delete α_4 and β_4 as before. Compared with the time series analysis, it is obvious that the panel threshold analysis can obtain more significant results. For fixed effect models and random effect models, the coefficients are the same. But their standard deviations and t-statistics are different. Only one of the coefficients is not significant in the table. When the precipitation is between 0 and 0.5 inch, if the rainfall increases by 0.1 inch, the overall stock return will decrease by 0.035470%. In the interval between 0.5 and 1.0, when the precipitation increases by 0.1 inch, the stock return will increase by 0.252703%, which is similar to the time series result. When the daily rainfall increases by 0.1 inch in the interval between 1.0 and 1.5, the stock return will reduce by 0.367801%.

2.5.3 Tests for cross-sectional dependence

The previous section mainly used the panel threshold regression model proposed by Hansen (1999), where the error ϵ_{it} is assumed to be identically independent distributed (iid) with mean zero and finite variance σ^2, and the analysis is asymptotic with fixed T as $n \to \infty$. We will consider the existence of cross-sectional dependence and time dependence in panels. Since the stock data is applied in this paper, it is obvious that the data is dependent over time. Then we shall deal with cross-sectional dependence in panels.

Different settings can adopt different test functions, which range from the panels with T fixed and $n \to \infty$ to those with n fixed and $T \to \infty$. All functions are based on the correlation coefficient of the model's residuals, that is defined as

$$\hat{\rho}_{ij} = \frac{\sum_{t=1}^{T} \hat{\epsilon}_{it}\hat{\epsilon}_{jt}}{\sqrt{\sum_{t=1}^{T} \hat{\epsilon}_{it}^2}\sqrt{\sum_{t=1}^{T} \hat{\epsilon}_{jt}^2}}.$$

The Breusch-Pagan LM test is valid for $T \to \infty$ with n fixed, based on the squares of ρ_{ij}^2 (Breusch and Pagan, 1980), that is defined as

$$LM = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} T_{it}\hat{\rho}_{ij}^2,$$

where $T_{ij} = \min(T_i, T_j)$, T_i and T_j are the numbers of observations for individual i and j respectively. This considers the case of unbalanced panel. And if the panel is balanced, $T_{ij} = T$ for each i, j. This test is distributed as $\chi^2_{n(n-1)/2}$. When the n dimension is large, that is the case of $T \to \infty$ and $n \to \infty$, the following test function is considered,

$$SCLM = \sqrt{\frac{1}{n(n-1)}} (\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sqrt{T_{it}\hat{\rho}_{ij}^2}).$$

And this is distributed as a standard Normal.

Pesaran's CD test is based on ρ_{ij} without squaring (Pesaran, 2021), that is defined as

$$CD = \sqrt{\frac{2}{n(n-1)}}(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{T_{it}}\hat{\rho}_{ij}),$$

which is also distributed as a standard Normal. And it is appropriate both in n and T asymptotic settings. It has remarkable properties in samples of any practically relevant size and is robust to a variety of settings.

These tests are originally meant to use the residuals of separate estimation of one time series regression for each cross-sectional unit. First, we apply the cross-sectional dependence test to the panel threshold model for resources and energy industry with fixed effect. The results show that Z-statistics=129.14 and P-value=2.2×10^{-16}, which says that the model exists cross-sectional dependence. Then, we move to the panel threshold model for FTSE 100 with the fixed effect. Since Z-statistics=481.77 and P-value=2.2×10^{-16}, the FTSE 100 panel threshold model also exists cross-sectional dependence. As a result, the residuals in both resources and energy industry model and FTSE 100 model are cross-sectional dependent. It is reasonable to apply the cross-sectional dependence assumption in the residuals.

2.6 Conclusion

This chapter study the rainfall effect on stock returns. Most literature studying the relationship between rainfall and the stock market suggest that the rainfall do not have any influence on the stock returns. Since the existing literature only consider the linear relationship. We first use nonparametric method (sm package) to plot the relationship between rainfall and stock return for every stock in FTSE 100. The plots show nonlinear relationship and obvious change points. Thus we can study the threshold effect of rainfall on stock returns. From the change point in nonparametric plots, we choose 0.5, 1.0 and 1.5 as threshold parameters. Then, we use threshold model for both each time series and panel data.

Panel threshold model is more efficient to study the precipitation effect on the whole stock market (London Stock Exchange). However, due to the dependence of different stocks, we need to consider the cross-sectional dependence in threshold model. Thus the existing panel threshold may be not good enough in this empirical application. A further topic for us is to study threshold panel time series model with cross-sectional dependence.

Chapter 3

Panel threshold regression model with cross-sectional dependence

3.1 Introduction

In recent years, many nonlinear parametric forms of time series model have been explored (Tong, 1990, 1995; Tjøstheim, 1994). Nonlinear time series analysis is widely used in financial data, especially in modelling volatility (Engle, 1982; Bollerslev, 1986). Threshold model proposed by Tong (1978) is one of the most popular nonlinear time series methods. It can explore the influence of different regimes and has good interpretation. Therefore, threshold time series model has been widely applied to the field of economic and biology Tong (1990); Tiao and Tsay (1994). Hansen (1999) considered the threshold effect in panel data model and established asymptotic theory.

The panel threshold regression model proposed by Hansen (1999) is the most common approach to study the threshold effect of panel data in empirical application Floro and Van Roye (2017); Aye and Edoja (2017); Huang et al. (2018). For the case of lagged variables, Seo and Shin (2016) studied the dynamic panel threshold model and proposed two-step least squares estimators when the threshold variable becomes strictly exogenous. It also established the asymptotic theory and developed the testing procedure for threshold effects and the exogeneity of threshold variable . Both Hansen (1999) and Seo and Shin (2016) considered the statistical inference with fixed T as $n \to \infty$, which cannot be applied to the case that sample size is large enough, such as the collected data in Chapter 2.

Therefore, we should consider the case of $n \to \infty$ and $T \to \infty$ in the asymptotic inference. More recently, the threshold effect of panel data model has been further discussed by Miao et al. (2020b,a). Floro and Van Roye (2017) studied the state-dependent response of monetary policy to financial sector-specific stress. They used a factor-augmented dynamic panel threshold regression model with common components to deal with cross-sectional dependence. However, a few literature give any statistical inference of the model with cross-sectional dependence. By far, there may not be proper threshold panel time series model that can be used to study weather effect on the stock market due to lack of statistical theory considering cross-sectional dependence.

In this project, we are therefore investigating cross-sectional dependence conditions for both regressors and errors (allowing for common factors; c.f., Pesaran (2006); Bai (2009)) and then study the asymptotic distribution concerning exogenous (e.g., weather variable) effects in the setting of our threshold models (1.1.3) for panel time series data. The settings of cross-sectional dependence have been discussed in semiparametric time trending models, see, e.g.,Robinson (2012); Chen et al. (2012b); Gao et al. (2020). Compared with semiparametric and nonparametric model, parametric model requires smaller sample size and its result is more explanatory. Different from these references, we consider cross-sectional dependence in parametric panel threshold model.

A growing number of literature would like to exhibit cross-sectional dependent in panel data model Robertson and Symons (2000); Phillips and Sul (2003); Baltagi and Baltagi (2008). Since in the past few decades, economic and financial interactions of countries and financial entities increased rapidly. We need to deal with the strong dependence in most economic and financial data. Hoechle (2007) showed that ignoring cross-sectional dependence could lead to severely biased statistical results. Such models take into account cross-sectional dependence in panel data so that we can deal with the strong cross-sectional dependence in many finance and climate data. For example, carbon dioxide emissions of a city may influence other cities' emission, so we need to consider certain cross-sectional dependence in panel analysis. Similar features defined in empirical research are gross domestic product (GDP), stock prices and so on. Consequently, an econometric theory of cross-sectional dependence is essential to support empirical studies. In the linear modelling setting, see, e.g., Choi (2006); Breitung and Pesaran (2008); Moon and Perron (2007); Fachin (2007) for the evidence of cross- sectional dependence in many real data applications. The asymptotic theory considering cross-sectional dependence was established in semiparametric (Chen et al., 2012b; Dong et al., 2015) and nonparametric models (Cai and Li, 2008; Robinson, 2012). However, few literature studied the statistical theory of parametric model with cross-sectional dependence, especially nonlinear parametric model. In real data analysis, nonlinear parametric is more popular due to the simple computation and good interpretation. Thus developing the statistical theory of nonlinear panel time series can provide one more new option for financial and economic data analysis.

In this project, to efficiently study the rain effect on the stock market, we should consider the existence of both cross-sectional and time dependence in threshold panel time series model with both n and T tending to infinity in the following sections. We establish the asymptotic distribution of estimated slope coefficients and threshold parameters under the assumption of cross-sectional and time dependence in single threshold model. We expand the inference theory to multiple threshold model. The simulated examples are introduced to show the efficiency of proposed statistical theory. We apply the proposed theory to empirical study of climate finance.

3.2 Single threshold model

Consider the following panel threshold regression model:

$$\begin{aligned} y_{it} =& \mu_i + \beta_{11} I(q_{it} \le \gamma) + \beta_{12}' X_{it} I(q_{it} \le \gamma) \\ &+ \beta_{21} I(q_{it} > \gamma) + \beta_{22}' X_{it} I(q_{it} > \gamma) + \epsilon_{it}, \quad i = 1, ..., n;\ t = 1, ..., T \end{aligned} \tag{3.2.1}$$

where the subscript i indexes the individual and the subscript t indexes time. μ_i is a fixed effect, the dependent variable y_{it} is scalar, the threshold variable q_{it} is scalar, the regressor X_{it} is a K-dimensional vector, $I(\cdot)$ is the indicator function and γ is a threshold parameter. The regression coefficient $\beta = (\beta_{11}, \beta_{12}', \beta_{21}, \beta_{22}')'$ is a $(2K+2)$-dimensional vector of unknown parameters. In Hansen's panel threshold regression model, for each t, (q_{it}, X_{it}) are independent and identically distributed (iid) across i. Differently, we assume that there exist cross-sectional dependence in (q_{it}, X_{it}) in this paper. And we also assume that they have time dependence. (q_{it}, X_{it}, y_{it}) should be stationary in further research. An alternative intuitive way of writing (3.2.1) is

$$y_{it} = \begin{cases} \mu_i + \beta_{11} + \beta_{12}' X_{it} + \epsilon_{it}, \ q_{it} \le \gamma, \\ \mu_i + \beta_{21} + \beta_{22}' X_{it} + \epsilon_{it}, \ q_{it} > \gamma. \end{cases}$$

The model (3.2.1) can be written as

$$y_{it} = \mu_i + (1, X_{it}')\beta_1 I(q_{it} \le \gamma) + (1, X_{it}')\beta_2 I(q_{it} < \gamma) + \epsilon_{it}. \tag{3.2.2}$$

where $\beta_1 = (\beta_{11}, \beta_{12}')'$ and $\beta_2 = (\beta_{21}, \beta_{22}')'$. β_1 and β_2 are $k+1$-dimensional vectors.

3.3 Least squares estimation

There are several methods can be adopted to estimate the parameter β and threshold parameter γ. And the ordinary least squares method is commonly used to obtain the estimators. For notational simplicity, we set $\beta_1 = (\beta_{11}, \beta_{12}')'$, $\beta_2 = (\beta_{21}, \beta_{22}')'$, $\beta = (\beta_1', \beta_2')'$, and $Z_{it} = (1, X_{it}')'$, and $Z_{it}(\gamma) = (Z_{it} I(q_{it} \le \gamma)', Z_{it} I(q_{it} > \gamma)')'$. Then model (3.2.1) can be written as follows,

$$Y_{it} = \mu_i + \beta' Z_{it}(\gamma) + \epsilon_{it}, \tag{3.3.1}$$

where β is a $(2K+2)$ dimensional vector of unknown parameters. To simplify the notation in the following sections, we set $Z_{it1}(\gamma) = (1_{it1}(\gamma), X_{it1}(\gamma)')' = Z_{it} \cdot I(q_{it} \le \gamma)$, $Z_{it2}(\gamma) = (1_{it2}(\gamma), X_{it2}(\gamma)')' = Z_{it} \cdot I(q_{it} \le \gamma)$, with $1_{it1}(\gamma) = I(q_{it} \le \gamma)$, $1_{it2}(\gamma) = I(q_{it} > \gamma)$, $X_{it1}(\gamma) = X_{it} \cdot I(q_{it} \le \gamma)$, $X_{it2}(\gamma) = X_{it} \cdot I(q_{it} > \gamma)$. We assume the value of the threshold parameter γ is appropriately taken in a bounded set $\Gamma = [\underline{\gamma}, \overline{\gamma}]$, where $\underline{\gamma}$ and $\overline{\gamma}$ are two appropriately given lower and upper constants. We consider the within-group estimation to eliminate the individual effect μ_i under $\sum_{i=1}^{n} \mu_i = 0$ for model identifiability with the intercepts β_{11} and β_{21} taken account of under different regimes in model (3.2.1). Note that $EY_{it} = \mu_i + \beta' EZ_{it}(\gamma)$, and hence

$$\tilde{Y}_{it} =: Y_{it} - EY_{it} = \beta'[Z_{it}(\gamma) - EZ_{it}(\gamma)] + \epsilon_{it} =: \beta' \tilde{Z}_{it}(\gamma) + \epsilon_{it}. \tag{3.3.2}$$

Then, taking an average over time index t in model (3.3.1), we obtain

$$\bar{Y}_i = \mu_i + \beta'\bar{Z}_i(\gamma) + \bar{\epsilon}_i, \tag{3.3.3}$$

where $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, $\bar{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}$, and

$$\bar{Z}_i(\gamma) = \frac{1}{T}\sum_{t=1}^{T} Z_{it}(\gamma) = \begin{pmatrix} \frac{1}{T}\sum_{i=1}^{T} Z_{it}\cdot I(q_{it}\le\gamma) \\ \frac{1}{T}\sum_{i=1}^{T} Z_{it}\cdot I(q_{it}>\gamma) \end{pmatrix}.$$

The following model can be obtained by taking the difference between (3.3.1) and (3.3.3).

$$Y_{it}^* = \beta' Z_{it}^*(\gamma) + \epsilon_{it}^*, \tag{3.3.4}$$

where $Y_{it}^* = Y_{it} - \bar{Y}_i$, $Z_{it}^*(\gamma) = Z_{it}(\gamma) - \bar{Z}_i(\gamma)$, and $\epsilon_{it}^* = \epsilon_{it} - \bar{\epsilon}_i$. Denote

$$Y_i^* = \begin{pmatrix} Y_{i1}^* \\ \vdots \\ Y_{iT}^* \end{pmatrix},\ Z_i^*(\gamma) = \begin{pmatrix} Z_{i1}^*(\gamma)' \\ \vdots \\ Z_{iT}^*(\gamma)' \end{pmatrix},\ \epsilon_i^* = \begin{pmatrix} \epsilon_{i1}^* \\ \vdots \\ \epsilon_{iT}^* \end{pmatrix}.$$

Let $Y^* = ((Y_1^*)', \cdots, (Y_n^*)')'$, $Z^*(\gamma) = ((Z_1^*(\gamma))', \cdots, (Z_n^*(\gamma))')'$, and $\epsilon^* = ((\epsilon_1^*)', \cdots, (\epsilon_n^*)')'$ denote the data for all individuals. Model (3.3.4) can be written as

$$Y^* = Z^*(\gamma)\beta + \epsilon^*. \tag{3.3.5}$$

If γ is given, an OLS estimator of β is

$$\hat{\beta}(\gamma) = (Z^*(\gamma)'Z^*(\gamma))^{-1}Z^*(\gamma)'Y^*. \tag{3.3.6}$$

which can be seen as the OLS estimator of

$$\beta_0(\gamma) = [E(\tilde{Z}(\gamma)'\tilde{Z}(\gamma))]^{-1}E[\tilde{Z}(\gamma)'\tilde{Y}] \tag{3.3.7}$$

that minimises $L(\beta,\gamma) =: E[(\tilde{Y} - \tilde{Z}(\gamma)\beta)'(\tilde{Y} - \tilde{Z}(\gamma)\beta)]$ with respect to β, following from (3.3.2), where $\tilde{Y}$ and $\tilde{Z}(\gamma)$ are defined from (3.3.2) in a similar manner to Y^* and $Z^*(\gamma)$ defined above. Then we can define the regression residuals to (3.3.5) by

$$\hat{\epsilon}^*(\gamma) = Y^* - Z^*(\gamma)\hat{\beta}(\gamma) = [I - Z^*(\gamma)(Z^*(\gamma)'Z^*(\gamma))^{-1}Z^*(\gamma)']Y^*, \tag{3.3.8}$$

and correspondingly,

$$\tilde{\epsilon}_{it}(\gamma) = \tilde{Y}_{it} - \beta_0(\gamma)'\tilde{Z}_{it}(\gamma). \tag{3.3.9}$$

The sum of squared errors (SSE) or the residual sum of squares (RSS) is

$$S_1(\gamma) = \hat{\epsilon}^*(\gamma)'\hat{\epsilon}^*(\gamma) = Y^{*\prime}(I - Z^*(\gamma)'(Z^*(\gamma)'Z^*(\gamma))^{-1}Z^*(\gamma)')Y^*. \tag{3.3.10}$$

The least square estimator of γ is given by

$$\hat{\gamma} = \arg\min_{\gamma\in\Gamma} S_1(\gamma). \tag{3.3.11}$$

which can be seen as the estimator of $\gamma_0 = \arg\min_{\gamma\in\Gamma} L(\beta_0(\gamma), \gamma)$.

Note that the sum of squared error function is a step function that occurs at less than nT distinct values of the observed threshold variables. Hansen (1999) recommends eliminating the smallest and largest 1% or 5% after sorting the distinct values of the observations. However, this may ignore the influence of extreme values. We keep all distinct values for the estimation of γ in this paper. When nT is large, instead of searching over all values of threshold variables for $\hat{\gamma}$, the search can be implemented by some specific quantiles. Thus the number of regressions performed in the search can be reduced.

After the estimator $\hat{\gamma}$ is obtained, we can define an estimator of β as follows:

$$\hat{\beta} = \hat{\beta}(\hat{\gamma}) = (Z^*(\hat{\gamma})'Z^*(\hat{\gamma}))^{-1}Z^*(\hat{\gamma})'Y^*. \tag{3.3.12}$$

which is the estimator of $\beta_0 = \beta_0(\gamma_0)$. Note that $\tilde{\epsilon}_{it}(\gamma_0) = \epsilon_{it}$.

3.4 Asymptotic properties

In this paper, we consider the cases $n \to \infty$ and $T \to \infty$ with cross-sectional dependence and time dependence. In our discussion, both the residuals and explanatory variables are allowed to be dependent over time and cross-sectionally. Hansen (1999) studied the asymptotic distribution of Panel threshold regression model with $n \to \infty$ and T fixed. He assumed that for each t, (q_{it}, x_{it}, e_{it}) are iid across i. And for each i, e_{it} is iid over t. Hansen (1999) showed that if the threshold estimate $\hat{\gamma}$ were the true value, the statistical inference on β can be proceed. And the estimated slope coefficient is asymptotically normal with a covariance matrix which can be estimated by

$$\hat{V} = (\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}^{*}(\gamma)x_{it}^{*}(\gamma)')^{-1}\hat{\sigma}^2,$$

where $\hat{\sigma}^2$ is the residual variance, and $\hat{\sigma}^2 = \frac{1}{nT}\hat{\epsilon}^{*'}\hat{\epsilon}^{*}$. In the following section, we will study the asymptotic distribution under both time and cross-sectional dependent conditions.

3.4.1 Assumptions

Before studying the asymptotic distribution, we need to introduce a martingale difference, extending identically independent distributed (iid) errors.

Definition (Martingale) Let $(\Omega, \mathcal{F}, P)$ be a probability space: Ω is a set, $\mathcal{F}$ is a σ-field with elements being subsets of Ω, and P a probability measure defined on $\mathcal{F}$. Let T be the set of integers. Let $\{\mathcal{F}_t, t \in T\}$ be an increasing sequence of σ-fields in $\mathcal{F}$. Suppose that $\{S_t, t \in T\}$ is a sequence of random variables on Ω satisfying

(i) S_t is measurable with respect to $\mathcal{F}_t$,

(ii) $E|S_t| < \infty$,

(iii) $E(S_t|\mathcal{F}_k) = S_k$ a.s. for all $k < t$, $t, k \in T$.

Then, the sequence $\{S_t, t \in T\}$ is said to be martingale with respect to $\{\mathcal{F}_t, t \in T\}$. $\epsilon_t = S_t - S_{t-1}$ is a martingale difference sequence and satisfies the following conditions:

(i) $E|\epsilon_t| < \infty$,

(ii) $E[\epsilon_t|\mathcal{F}_{t-1}] = 0$ a.s.

And we also need to introduce α-mixing property.

Definition (α-mixing) For a stochastic process $X_1, X_2, ...,$ define

$$\mathcal{F}_1^n = \sigma(X_1, ..., X_n),$$

the σ-algebra generated by the first n random variables, and

$$\mathcal{F}_n^\infty = \sigma(X_n, X_{n+1}, ...),$$

the σ-algebra generated by the (infinite) family of random variables that is the subsequence starting at n. Let

$$\alpha_n = \sup_{m \in N, A \in \mathcal{F}_1^m, B \in \mathcal{F}_{m+n}^\infty} |P(A \cap B) - P(A)P(B)|.$$

The stochastic process is said to be α-mixing or strongly mixing if $\alpha_n \to 0$.

We first specify the conditions on the implicit cross sectional dependence of the error term $\{\epsilon_{it}\}$ that is a martingale process along time, and more assumptions on the moments of $\{\epsilon_{it}\}$ and $\{X_{it}\}$ in (3.3.4) in Assumption 3.1.

Assumption 3.1.

(i) $\{(\epsilon_{it}, X_{it}, q_{it}) : 1 \le i \le n\}$ as a process in time t is α-mixing stationary, for any n, and $E\|X_{it}\|^{4+\tau} < \infty$, $E\|\epsilon_{it}\|^{4+\tau} < \infty$, for any $\tau > 0$. The α-mixing coefficient satisfies $\sum_{p=1}^{\infty} p^{\phi}(\alpha(p))^{\frac{\tau}{4+\tau}} \le \infty$, for some $\phi > 1$, where $\alpha(p) = \max_i \sup_{A \in \mathcal{F}_{-\infty}^q, B \in \mathcal{F}_{q+p}^\infty} |P(AB) - P(A)P(B)| \to 0$ as $p \to \infty$, where $\mathcal{F}_{-\infty}^q$ and $\mathcal{F}_{q+p}^\infty$ are two σ-fields generated by $\{(\epsilon_{it}, X_{it}, q_{it}), t \le q\}$ and $\{(\epsilon_{it}, X_{it}, q_{it}), t \ge q + p\}$, respectively.

(ii) $\{\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{nt})'\}$ is a sequence of martingale differences in time t. That is, $E[\epsilon_t|\mathcal{F}_{t-1,n}] = 0_n,\ a.s.$, where $\mathcal{F}_{t,n} = \sigma\{(\epsilon_{i\ell}, X_{i,\ell+1}, q_{i,\ell+1}) :\ \ell \leq t, 1 \leq i \leq n\}$ is a σ-field generated by $\{(\epsilon_{i\ell}, X_{i,\ell+1}, q_{i,\ell+1}) :\ \ell \leq t, 1 \leq i \leq n\}$.

(iii) $\{(\epsilon_{it}, X_{it}, q_{it}) : 1 \leq t \leq T\}$ are dependent cross-sectionally and $Var(X_{it}) > 0$. Let C_0 denote any $(K+1) \times 1$ nonrandom vector with $\|C_0\| = 1$, and $X_{it}^C = C_0' X_{it} / \sqrt{Var(C_0' X_{it})}$, as $n \to \infty$,

$$
\begin{aligned}
&E\left|\sum_{i=1}^{n}(X_{it}^C - EX_{it}^C)\right|^{2+\tau} = O(n^{\frac{2+\tau}{2}}), \\
&E\left|\sum_{i=1}^{n}\epsilon_{it}\right|^{2+\tau} = O(n^{\frac{2+\tau}{2}})
\end{aligned}
\tag{3.4.1}
$$

where $\tau > 0$ is as defined in Assumption 3.1(i).

(iv) There exist $(2K+2) \times (2K+2)$ positive definite matrices Σ_γ, Σ_{γ_0}, $\Sigma_{\gamma,\epsilon}$ and $\Sigma_{\gamma_0,\epsilon}$ such that as $n \to \infty$,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} E(\tilde{Z}_{it}(\gamma)\tilde{Z}_{it}(\gamma)') \to \Sigma_\gamma, \\
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E(\tilde{Z}_{it}(\gamma)\tilde{Z}_{jk}(\gamma)'\tilde{\epsilon}_{it}(\gamma)\tilde{\epsilon}_{jk}(\gamma)) \to \Sigma_{\gamma,\epsilon}, \\
&\frac{1}{n}\sum_{i=1}^{n} E(\tilde{Z}_{it}(\gamma_0)\tilde{Z}_{it}(\gamma_0)') \to \Sigma_{\gamma_0}, \\
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E(\tilde{Z}_{it}(\gamma_0)\tilde{Z}_{jt}(\gamma_0)'\sigma_{\epsilon t}(i,j)) \to \Sigma_{\gamma_0,\epsilon},
\end{aligned}
\tag{3.4.2}
$$

where $\sigma_{\epsilon t}(i,j) = E[\epsilon_{it}\epsilon_{jt}|\mathcal{F}_{t-1,n}]\ a.s.$. Additionally, there exists $0 < \sigma_\epsilon^2 < \infty$ such that as $n \to \infty$,

$$
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\sigma_{\epsilon t}(i,j) \to \sigma_\epsilon^2. \tag{3.4.3}
$$

(v) Let C_0 denote any $(K+1) \times 1$ nonrandom vector with $\|C_0\| = 1$, $\zeta_{it} = \frac{C_0' Z_{it}\epsilon_{it}}{\sqrt{E(C_0' Z_{it}\epsilon_{it})^2}}$ and $\tilde{\zeta}_{it} = \frac{(\zeta_{it}^2 - E\zeta_{it}^2}{\sqrt{Var(\zeta_{it}^2)}}$, there exist some constants such that, as both $n \to \infty$, $T \to \infty$,

$$
\frac{1}{n}\sum_{i \neq i_1}^{n} |E\tilde{\zeta}_{it}\tilde{\zeta}_{i_1 t}| \leq C_\rho^1, \quad \frac{1}{nT}\sum_{i \neq i_1}^{n}\sum_{t \neq t_1}^{T} |E\tilde{\zeta}_{it}\tilde{\zeta}_{i_1 t_1}| \leq C_\rho^2,
$$

$$\frac{1}{n^{\frac{3}{2}}}\sum_{i\neq i_1\neq i_2}^{n}\left|E\tilde{\zeta}_{it}\zeta_{i_1t}\zeta_{i_2t}\right|\leq C_\rho^3,\ \frac{1}{n^2}\sum_{i\neq i_1\neq i_2\neq i_3}^{n}\left|E\zeta_{it}\zeta_{i_1t}\zeta_{i_2t}\zeta_{i_3t}\right|\leq C_\rho^4,$$
$$\frac{1}{n^2T}\sum_{i\neq i_1\neq i_2\neq i_3}^{n}\sum_{t\neq t_1}^{T}\left|E\zeta_{it}\zeta_{i_1t}\zeta_{i_2t_1}\zeta_{i_3t_1}\right|\leq C_\rho^5. \tag{3.4.4}$$

Remark 3.1. Note that $\{\epsilon_t, t\geq 1\}$ is assumed to be a stationary martingale difference sequence in Assumption 3.1. The martingale difference condition is useful in establishing asymptotic normality, allowing for conditional heteroscedasticity in financial data, and in general weaker than α-mixing error assumption. We consider $\{\epsilon_{it}\}$ is uncorrelated with $\{X_{it}\}$ and $\{q_{it}\}$, which is implied in Assumption 3.1 (i), and thus the asymptotic results are studied under the exogeneity assumption. We assume cross-sectional dependence in Assumption 3.1 (iii). We impose cross-sectional dependence conditions on $\{\epsilon_{it}\}$ in Assumption 3.1 (v), similar to the condition A2 in Chen et al. (2012b), who proposed certain conditions to measure the distance between cross-sectional units i and j. Unlike the time index that is unilateral for time series, the cross-sectional indices cannot be ordered naturally. To explain the cross-sectional dependence condition in some detail, define a kind of "distance function" among cross-sections of the form

$$\rho(i_1,i_2,...,i_s)=E(\zeta_{i_1,1}\zeta_{i_2,1}...\zeta_{i_s,1}), \tag{3.4.5}$$

and then consider one of the cases where $s=4$. We focus on the case where $1\leq i_1\neq i_2\neq i_3\neq i_4\leq n$. Consider a distance function of the form

$$\rho(i_1,i_2,...,i_4)=\frac{1}{|i_4-i_3|^{\delta_3}|i_3-i_2|^{\delta_2}|i_2-i_1|^{\delta_1}} \tag{3.4.6}$$

for $\delta_i>0$ for all $1\leq i\leq 3$. In this case, the second condition can be verified because

$$\sum_{i_1\neq i_2\neq i_3\neq i_4}^{n}E\zeta_{i_1t}\zeta_{i_2t}\zeta_{i_3t}\zeta_{i_4t}=O(n^{4-\sum_{j=1}^{n}\delta_j})=O(n^2), \tag{3.4.7}$$

when $\sum_{j=1}^{n}\delta_j\geq 2$. Assumption 3.1 (iv) imposes some mild conditions on ζ_{it}. If ζ_{it} is cross-sectionally independent and serially independent or weakly dependent, then (3.4.4) holds.

We also allow the cross sectional dependence of the covariate process $\{X_{it}\}$ that is often defined by adopting a common factor structure,

$$X_{it} = \Lambda_i c_t + v_{it}, i = 1, ..., n, t = 1, ..., T, \tag{3.4.8}$$

where $c_t = (c_{1t}, ..., c_{mt})'$ is a m-dimensional vector of common factors, Λ_i is a $k \times m$ factor loading matrix. As mentioned in Section 1, in a special case, we allow $X_{it} = x_t$ (scalar) across i, where correspondingly $c_t = x_t$, $\Lambda_i = 1$ and $v_{it} = 0$ for all i and t. In general, we impose the following assumptions on X_{it}.

Assumption 3.2.

(i) Let $v_t = (v_{1t}, ..., v_{nt})'$. The sequence $\{v_t, t \geq 1\}$ is a vector of stationary α-mixing time series with zero mean, and $E(\|v_{it}\|^{4+\tau}) < \infty$ uniformly across $1 \leq i \leq n$, for some $\tau > 0$ as given in Assumption 3.1(i). The α-mixing coefficient is the same as that in Assumption 3.1(i).

(ii) There exists a $K \times K$ non-negative definite matrix Ω_v, such that as $n \to \infty$,

$$\frac{1}{n} \sum_{i=1}^{n} E(v_{it} v_{it}') \to \Omega_v, \ \sum_{i=1}^{n} \sigma_{vi} = O(n), \tag{3.4.9}$$

where $\sigma_{vi} = E(v_{i1} v_{i1}') + 2 \sum_{t=2}^{\infty} E(v_{i1} v_{it}')$, and $\| \cdot \|$ is the Euclidean norm.

(iii) The common factor, $\{c_t, t \geq 1\}$, is a stationary α-mixing sequence with $E(\|c_t\|^{4+\tau}) < \infty$ for the same constant $\tau > 0$ in Assumption 3.1(i). The α-mixing coefficient is the same as that in Assumption 3.1(i).

Remark 3.2. Assumption 3.2 imposes some dependent conditions for X_{it}. Assumption 3.2 (i) requires $\{v_t, t \geq 1\}$ to be a stationary α-mixing sequence, where the mixing condition is commonly used in some time series models (see e.g. Fan and Yao, 2003; Gao, 2007; Chen et al., 2012b). Assumption 3.2 (ii) assume $\{v_{it}\}$ to be cross-sectional independence. Therefore $\{v_{it}\}$ is assumed to be cross-sectional independent but serial dependent. For each i, $i \geq 1$, we impose the common factor, $\{c_t\}$, of $\{X_{it}\}$. Thus $\{X_{it}\}$ is assumed to be cross-sectionally dependent by defining $\{c_t\}$.

Further, let $\theta = \beta_2 - \beta_1$ be the threshold (discontinuity) effect of X_{it} at the threshold γ of q_{it}. Differently from the specifications in the literature (c.f., Hansen (1999); Seo and Shin (2016); Miao et al. (2020a,b)), we allow $\theta = O(n^{-\alpha_1} T^{-\alpha_2})$

decaying to zero at different rates of $\alpha_1 \in (0, \frac{1}{2})$ and $\alpha_2 \in (0, \frac{1}{2})$ as the cross-sectional size n and the temporal size T tend to infinity, respectively, in view of the fact that real panel data may increase at different rates in time and cross section. For example, when $\alpha_1 = \alpha_2$, it reduces to the case considered in Miao et al. (2020a,b). Let $f_i(\gamma)$ denote the probability density function of q_{it}, which is independent of t as q_{it} is stationary along time,

$$D(\gamma) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(Z_{it}Z_{it}'|q_{it} = \gamma) f_i(\gamma),$$
$$V(\gamma) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(Z_{it}Z_{it}'\epsilon_{it}^2|q_i = \gamma) f_{it}(\gamma), \tag{3.4.10}$$

and $D = D(\gamma_0)$, $V = V(\gamma_0)$, with γ_0 being the true value of the threshold parameter.

Assumption 3.3.

(i) For some fixed vector of C (of the dimension of θ) satisfying $\|C\| < \infty$ such that $\theta = n^{-\alpha_1}T^{-\alpha_2}C$ for some $0 < \alpha_1, \alpha_2 < \frac{1}{2}$.

(ii) For all $\gamma \in \Gamma$, $\max_i E(\|Z_{it}\epsilon_{it}\|^4|q_{it} = \gamma) \leq A$ and $\max_i E(\|Z_{it}\|^4|q_{it} = \gamma) \leq A$ for some $A < \infty$, and $\max_i f_{it}(\gamma) \leq \bar{f} < \infty$.

(iii) $D(\gamma)$ and $V(\gamma)$ are continuous at $\gamma = \gamma_0$.

(iv) $0 < C'DC, C'VC < \infty$.

Remark 3.3. Assumption 3.3 provides certain conditions on the threshold effect similar to, but extending, the assumptions in Hansen (1999) and Hansen (2000). Assumption 3.3 (i) specifies a diminishing threshold effect that includes the special case $\theta = n^{-\alpha}C \to 0$ as $n \to \infty$, with $\alpha_1 = \alpha$ and $\alpha_2 = 0$ in Assumption 3.3 (i). This special case is commonly used in many threshold panel models, see Hansen (1999), Seo and Shin (2016) and others. Assumption 3.3 (ii) imposes some conditions of the conditional probability density function and moment bounds. Assumption 3.3 (iii) asks as usual for a continuous threshold variable, excluding the threshold effect with a regime variable to have a discrete distribution, as in the literature mentioned. Assumption 3.3 (iv) is a regularity condition on the threshold effect, excluding the case $C = 0$, i.e., without existence of a threshold effect supposed. Under Assumption 3.3 (iii, iv), the threshold regime variable q_{it} is actually assumed to have some kind of continuous distribution with positive densities at γ_0 ensuring positive definiteness for D and V at $\gamma = \gamma_0$.

3.4.2 Asymptotic distribution

The following two theorems give the asymptotic distributions of $\hat{\beta}$ and $\hat{\gamma}$.

Theorem 3.1. *Let Assumption 3.1 and Assumption 3.2 hold. As $n \to \infty$ and $T \to \infty$,*

$$\sqrt{nT}(\hat{\beta}(\gamma) - \beta_0(\gamma)) \xrightarrow{d} N(0_{2K+2}, \Sigma_\gamma^{-1}\Sigma_{\gamma,\epsilon}\Sigma_\gamma^{-1}). \tag{3.4.11}$$

Remark 3.4. The above theorem shows that the proposed estimator, $\hat{\beta}(\gamma)$, follows the normal distribution asymptotically with the $\sqrt{nT}$ convergence rate when γ is given. Note that the estimated asymptotic distribution of $\hat{\beta}(\gamma)$ under cross-sectional independence condition is $N(0_{2K+2}, \Upsilon_\gamma^{-1}\Upsilon_{\gamma,\epsilon}\Upsilon_\gamma^{-1})$, where $\Upsilon_\gamma = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E\tilde{Z}_{it}(\gamma)\tilde{Z}_{it}(\gamma)'$ and $\Upsilon_{\gamma,\epsilon} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n E\tilde{Z}_{it}(\gamma)\tilde{Z}_{it}(\gamma)'\sigma_\epsilon(i,i)$, which is similar to Theorem 3.4 in Miao et al. (2020a). Note the difference of defining $\Upsilon_{\gamma,\epsilon}$, involving $\sigma_\epsilon(i,i)$ only, from our $\Sigma_{\gamma,\epsilon}$ in (3.4.2), involving $\sigma_\epsilon(i,j)$ across i and j, characterising cross-sectional dependence. Compared with the existing literature on panel threshold regression model, the asymptotic variance in Theorem 3.1 captures the cross-sectional dependence.

Theorem 3.2. *Under Assumption 3.1 to Assumption 3.3, as $n \to \infty$ and $T \to \infty$,*

$$n^{1-2\alpha_1}T^{1-2\alpha_2}(\hat{\gamma} - \gamma_0) \xrightarrow{d} w \arg\max_{v\in R}\left[-\frac{|v|}{2} + W(v)\right] \tag{3.4.12}$$

where $w = \frac{C'VC}{(C'DC)^2}$, $W(v)$ is a two-sided Brownian motion.

Remark 3.5. The above theorem shows the asymptotic distribution of threshold parameter is highly non-standard with the convergence rate, $n^{1-2\alpha_1}T^{1-2\alpha_2}$. This result is similar to Theorem 1 of Hansen (2000). In panel data analysis, Hansen (1999) assumed time series length, T, fixed and analyzed the convergence rate,

$n^{1-2\alpha}$, where $\alpha \in (0, \frac{1}{2})$. Our result relaxes the convergence rate to a more general condition so that the theorem can be more applicable.
Finally, by Theorems 3.1 and 3.2, we have the asymptotic property for $\hat{\beta} = \hat{\beta}(\hat{\gamma})$.

Theorem 3.3. *Let Assumption 3.1 to Assumption 3.3 hold. As $n \to \infty$ and $T \to \infty$,*

$$\sqrt{nT}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0_{2K+2}, \Sigma_{\gamma_0}^{-1}\Sigma_{\gamma_0,\epsilon}\Sigma_{\gamma_0}^{-1}). \tag{3.4.13}$$

Remark 3.6. The above theorem establishes the asymptotic distribution of $\hat{\beta}$ by the consistency of $\hat{\gamma}$ in Proposition 3.1 and Theorem 3.1. By this theorem, the asymptotic variance for the estimator $\hat{\beta}$ can hence be estimated through the estimators of

$$\begin{aligned} \hat{\Sigma}_{\hat{\gamma}} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}^{*}(\hat{\gamma})(Z_{it}^{*}(\hat{\gamma}))', \\ \hat{\Sigma}_{\hat{\gamma},\hat{\epsilon}} &= \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} Z_{it}^{*}(\hat{\gamma})(Z_{jt}^{*}(\hat{\gamma}))'\hat{\epsilon}_{it}\hat{\epsilon}_{jt}, \end{aligned} \tag{3.4.14}$$

where $\hat{\epsilon}_{it}$ is the residual of an estimated model of (3.2.1). Also, the estimated variances for the estimators are consistent. Furthermore, the asymptotic variance matrix estimator given by (3.4.14) can be used to calculate confidence intervals for the components of the true parameter vector β_0 of β.

3.5 Testing for the existence of threshold effect

Before making inference for unknown parameters, it is important to verify the existence of threshold effect. This issue has been considered in Hansen (1996). Specifically, recall model (3.2.2) and consider the null hypothesis of no threshold effect,

$$\mathbf{H}_0 : \theta = 0,$$

against the alternative hypothesis,

$$\mathbf{H}_1 : \theta \neq 0.$$

Due to unidentified threshold parameter under the null, testing for the existence of threshold effect is non-standard. To study the local power of our test, we consider the following sequence of Pitman local alternatives,

$$\mathbf{H}_{1nT} : \theta = \frac{c}{\sqrt{nT}}.$$

Obviously, the case $c = 0$ corresponds to the null hypothesis of no threshold effect.

For each $\gamma \in \Gamma = [\underline{\gamma}, \overline{\gamma}]$, we can obtain the estimator $\hat{\theta}(\gamma) = \hat{\beta}_2(\gamma) - \hat{\beta}_1(\gamma)$. The asymptotic variance estimator is defined as

$$\hat{\Sigma}_\gamma = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{it}^*(\gamma) \left(Z_{it}^*(\gamma)\right)',$$

$$\hat{\Sigma}_{\gamma,\epsilon} = \frac{1}{nT} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{t=1}^{T} Z_{it}^*(\gamma) \left(Z_{it}^*(\gamma)\right)' \epsilon_{it}(\gamma)\epsilon_{jt}(\gamma).$$

The sup-Wald statistic is defined as

$$\sup W_{nT} = \sup_{\gamma \in \Gamma} W_{nT}(\gamma),$$

where $W_{nT}(\gamma) = nT\hat{\beta}(\gamma)'(L_2 - L_1)\hat{M}_{nT}^{-1}(\gamma)(L_2 - L_1)\hat{\beta}(\gamma)$ with section matrices $L_1 = [I_{K+1}, 0_{(K+1)\times(K+1)}]'$ and $L_2 = [0_{(K+1)\times(K+1)}, I_{K+1}]'$, and $\hat{M}_{nT}(\gamma) = (L_2 - L_1)'\hat{\Sigma}_\gamma^{-1}\hat{\Sigma}_{\gamma,\epsilon}\hat{\Sigma}_\gamma^{-1}(L_2 - L_1)$. $\hat{M}_{nT}(\gamma)$ is the estimated covariance matrix for $\sqrt{nT}\tilde{\theta}(\gamma)$.

Suppose there exist matrices $F(\gamma)$ and Θ_{γ,γ_0} such that as $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n} E\tilde{Z}_{it}(\gamma)\epsilon_{it} \to F(\gamma)$ and $\frac{1}{n}\sum_{i=1}^{n} E\tilde{Z}_{it}(\gamma)\tilde{Z}_{it}(\gamma_0) \to \Theta_{\gamma,\gamma_0}$. The asymptotic property of the statistic $\sup W_{nT}(\gamma)$ is given in the following theorem.

Theorem 3.4. *Suppose Assumption 3.1 to Assumption 3.3 hold. Then under* $\mathbf{H}_{1nT}$: $\theta = \frac{c}{\sqrt{nT}}$, *we have*

$$\sup W_{nT} \xrightarrow{d} \sup_{\gamma \in \Gamma} W^c(\gamma),$$

where $W^c(\gamma) = [\tilde{F}(\gamma) + \tilde{Q}(\gamma)c]'M^{-1}(\gamma)[\tilde{F}(\gamma) + \tilde{Q}(\gamma)c]$, $\tilde{Q}(\gamma) = \frac{1}{2}(L_2 - L_1)'\Sigma_\gamma^{-1}$ $\Theta_{\gamma,\gamma_0}(L_2 - L_1)$ *and* $\tilde{F}(\gamma) = (L_2 - L_1)'\Sigma_\gamma^{-1}F(\gamma)$ *is a mean zero Gaussian process with covariance kernel* $M(\gamma) = (L_2 - L_1)'\Sigma_\gamma^{-1}\Sigma_{\gamma,\epsilon}\Sigma_\gamma^{-1}(L_2 - L_1)$.

Under $\mathbf{H}_0$, $c = 0$ and $\sup_{\gamma\in\Gamma} W^0(\gamma) = \sup_{\gamma\in\Gamma} \tilde{F}(\gamma)'M^{-1}(\gamma)\tilde{F}(\gamma)$. It is apparently the limiting null distribution of $\sup W_{nT}$ depends on the Gaussian process $\tilde{F}(\gamma)$ and is not pivotal. We cannot tabulate the asymptotic critical values for the above sup-Wald statistic. However, given the simple structure of $\tilde{F}(\gamma)$, we can follow Hansen (1996) and simulate the asymptotic critical values as follows,

1. Generate $\{s_{it}, i = 1, ..., n, t = 1, ..., T\}$ independently from standard normal distribution;
2. Set $\hat{F}_{nT}(\gamma) = \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T Z_{it}^*(\gamma)\hat{e}_{it}^*(\gamma)s_{it}$;
3. Compute $\sup \hat{W}_{nT} = \sup_{\gamma\in\Gamma} \hat{F}_{nT}(\gamma)'\hat{\Sigma}_\gamma^{-1}(L_2 - L_1)\hat{M}_{nT}^{-1}(\gamma)(L_2 - L_1)'\hat{\Sigma}_\gamma^{-1}\hat{F}_{nT}(\gamma)$;
4. Repeat steps 1-3 P times and denote the resulting $\sup \hat{W}_{nT}$ test statistics as $\sup \hat{W}_{nT,j}$ for $j = 1, ..., P$;
5. Calculate the simulate/bootstrap p-value for $\sup W_{nT}$ as $\hat{p}_W = \frac{1}{P} \sum_{j=1}^P I(\sup \hat{W}_{nT,j} > \sup W_{nT})$ and reject the null when $\hat{p}_W$ is smaller than some prescribed nominal level of significance.

3.6 Multiple threshold model

Consider a double threshold model of the form

$$\begin{aligned} Y_{it} =& \mu_i + \beta_{11}I(q_{it} \le \gamma_1) + \beta_{12}'X_{it}I(q_{it} \le \gamma_1) + \beta_{21}I(\gamma_1 < q_{it} \le \gamma_2) \\ &+ \beta_{22}'X_{it}I(\gamma_1 < q_{it} \le \gamma_2) + \beta_{31}I(q_{it} > \gamma_2) + \beta_{32}'X_{it}I(q_{it} > \gamma_2) + \epsilon_{it}, \end{aligned} \tag{3.6.1}$$

where the threshold parameters are ordered as $\gamma_1 < \gamma_2$. The least squares method in Section 3.3 is appropriate for estimating (3.6.1). Let $\beta(\gamma_1, \gamma_2) =$

$(\beta_{11}, \beta'_{12}, \beta_{21}, \beta'_{22}, \beta_{31}, \beta'_{32})'$, and

$$Z_{it}(\gamma_1, \gamma_2) = \begin{pmatrix} Z_{it} \cdot I(q_{it} \le \gamma_1) \\ Z_{it} \cdot I(\gamma_1 < q_{it} \le \gamma_2) \\ Z_{it} \cdot I(q_{it} > \gamma_2) \end{pmatrix}$$

where $Z_{it} = (1, X'_{it})'$.Note that $EY_{it} = \mu_i + \beta' EZ_{it}(\gamma)$, and hence

$$\tilde{Y}_{it} =: Y_{it} - EY_{it} = \beta'[Z_{it}(\gamma_1, \gamma_2) - EZ_{it}(\gamma_1, \gamma_2)] + \epsilon_{it} =: \beta' \tilde{Z}_{it}(\gamma_1, \gamma_2) + \epsilon_{it}. \tag{3.6.2}$$

If γ_1, γ_2 are given, and $\gamma_1 < \gamma_2$, by fixed effect transformation similar to (3.3.5), an OLS estimator of β is obtained

$$\hat{\beta}(\gamma_1, \gamma_2) = (Z^*(\gamma_1, \gamma_2)' Z^*(\gamma_1, \gamma_2))^{-1} Z^*(\gamma_1, \gamma_2)' Y^*, \tag{3.6.3}$$

which can be seen as the OLS estimator of

$$\beta_0(\gamma_1, \gamma_2) = [E(\tilde{Z}(\gamma_1, \gamma_2)' \tilde{Z}(\gamma_1, \gamma_2))]^{-1} E(\tilde{Z}(\gamma_1, \gamma_2)' \tilde{Y}), \tag{3.6.4}$$

which minimize $L(\beta, \gamma_1, \gamma_2) =: E[(\tilde{Y}) - \tilde{Z}(\gamma_1, \gamma_2)\beta)'(\tilde{Y}) - \tilde{Z}(\gamma_1, \gamma_2)\beta)]$. The vector of regression residuals is

$$\hat{\epsilon}^*(\gamma_1, \gamma_2) = Y^* - Z(\gamma_1, \gamma_2) \hat{\beta}(\gamma_1, \gamma_2), \tag{3.6.5}$$

and the sum of squared erros is

$$S(\gamma_1, \gamma_2) = \hat{\epsilon}^*(\gamma_1, \gamma_2)' \hat{\epsilon}^*(\gamma_1, \gamma_2). \tag{3.6.6}$$

The first stage is to estimate the single threshold model given in Section 3.3. Let $S_1(\gamma)$ be the sum of squared errors of single threshold model defined as (3.3.9). The first threshold parameter $\hat{\gamma}_1$ can be estimated by minimizing $S_1(\gamma)$. Suppose the first threshold parameter $\hat{\gamma}_1$ fixed, the sum of squared errors of double threshold model is defined as

$$S_2(\gamma_2) = \begin{cases} S(\hat{\gamma}_1, \gamma_2) & \text{if} \quad \hat{\gamma}_1 < \gamma_2 \\ S(\gamma_2, \hat{\gamma}_1) & \text{if} \quad \gamma_2 < \hat{\gamma}_1 \end{cases} \tag{3.6.7}$$

and the second threshold is estimated by

$$\hat{\gamma}_2 = \arg\min S_2(\gamma_2). \tag{3.6.8}$$

$\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2)'$ is the estimator of $\gamma_0 = (\gamma_{01}, \gamma_{02})' = \arg\min_{\gamma_1 \in \Gamma_1, \gamma_2 \in \Gamma_2} L(\beta_0(\gamma_1, \gamma_2), \gamma_1, \gamma_2)$. Once the threshold parameters are given, the slope coefficients $(\beta_{11}, \beta_{12}', \beta_{21}, \beta_{22}', \beta_{31}, \beta_{32}')$ can be estimated by ordinary least squares. Thus we have the final estimator, $\hat{\beta} = \hat{\beta}(\hat{\gamma}_1, \hat{\gamma}_2)$, of $\beta_0 = \beta_0(\gamma_{01}, \gamma_{02})$. The estimation method can extend to higher order threshold models by fixing $\hat{\gamma}_1$ and $\hat{\gamma}_2$, and estimating more threshold parameters. Both Bai (1997) and Hansen (1999) showed that the second threshold parameter $\hat{\gamma}_2$ is asymptotically efficient, but $\hat{\gamma}_1$ is not. They suggested to fix $\hat{\gamma}_2$ and re-estimate $\hat{\gamma}_1$ by the same method as (3.3.6) to obtain efficient estimator. On the whole, the estimation method of second threshold is similar as the first one since both are estimated by minimizing the squared error. We model the cross-sectional dependence of X_{it} by the common factor function same as (3.4.8) and impose the following assumptions to study asymptotic distributions of multiple threshold model.

To proceed, we need to introduce the necessary assumptions as follows.

Assumption 3.4.

(i) This part is the same as Assumption 3.1 (i).

(ii) This part is the same as Assumption 3.1 (ii).

(iii) There exist $(3K+3) \times (3K+3)$ positive definite matrices $\Sigma_{\gamma_1,\gamma_2}$ and $\Sigma_{\gamma_1,\gamma_2,\epsilon}$ such that as $n \to \infty$,

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} E(\tilde{Z}_{it}(\gamma_1,\gamma_2)\tilde{Z}_{it}(\gamma_1,\gamma_2)') \to \Sigma_{\gamma_1,\gamma_2}, \\ &\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E(\tilde{Z}_{it}(\gamma_1,\gamma_2)\tilde{Z}_{jt}(\gamma_1,\gamma_2)')\sigma_{\epsilon t}(i,j) \to \Sigma_{\gamma_1,\gamma_2,\epsilon}, \end{aligned} \tag{3.6.9}$$

where $\sigma_{\epsilon t}(i,j) = E[\epsilon_{it}\epsilon_{jt}|\mathcal{F}_{t-1,n}]$ *a.s.*. Additionally, there exists $0 < \sigma_\epsilon^2 < \infty$ such that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E\sigma_{\epsilon t}(i,j) \to \sigma_\epsilon^2. \tag{3.6.10}$$

(iv) This part is the same as Assumption 3.1 (iv).

Assumption 3.5. This assumption is the same as Assumption 3.2.

Let $\gamma_0 = (\gamma_{01}, \gamma_{02})'$ be the true vector of the two threshold parameters, and $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2)'$ the estimator of γ_0. Further, denote $\theta_1 = \beta_2 - \beta_1$ and $\theta_2 = \beta_3 - \beta_2$ for the element-wise differences between β_2 and β_1 and between β_3 and β_2, respectively, where $\beta_1 = (\beta_{11}, \beta_{12}')' \in R^{1+K}$, $\beta_2 = (\beta_{21}, \beta_{22}')' \in R^{1+K}$, $\beta_3 = (\beta_{31}, \beta_{32}')' \in R^{1+K}$. Recall that $f_{it}(\gamma)$ denotes the probability density function of q_{it}, which is independent of t as q_{it} is stationary along time. Further, set $D_1 = D(\gamma_{01})$, $D_2 = D(\gamma_{02})$, $V_1 = V(\gamma_{01})$, $V_2 = V(\gamma_{02})$ with γ_{01}, γ_{02} being the two true threshold parameters, where $D(\gamma)$ and $V(\gamma)$, with $\gamma \in R^1$, are defined as.

$$
\begin{aligned}
D(\gamma_1) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(Z_{it}Z_{it}'|q_{it} = \gamma_1) f_i(\gamma_1), \\
D(\gamma_2) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(Z_{it}Z_{it}'|q_{it} = \gamma_2) f_i(\gamma_2), \\
V(\gamma_1) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(Z_{it}Z_{it}'\epsilon_{it}^2|q_{it} = \gamma_1) f_i(\gamma_1), \\
V(\gamma_2) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E(Z_{it}Z_{it}'\epsilon_{it}^2|q_{it} = \gamma_2) f_i(\gamma_2).
\end{aligned}
\tag{3.6.11}
$$

Assumption 3.6.

(i) For some fixed vector of C_1 and C_2 (of the dimension of θ_1 and θ_2) satisfying $\|C_1\| < \infty$ and $\|C_2\| < \infty$ such that $\theta_1 = n^{-\alpha_1}T^{-\alpha_2}C_1$ and $\theta_2 = n^{-\alpha_1}T^{-\alpha_2}C_2$ for some $0 \le \alpha_1, \alpha_2 < \frac{1}{2}$.

(ii) For all $\gamma \in \Gamma$, $\max_i E(\|Z_{it}\epsilon_{it}\|^4|q_{it} = \gamma_1) \le A$, $\max_i E(\|Z_{it}\epsilon_{it}\|^4|q_{it} = \gamma_2) \le A$, $\max_i E(\|Z_{it}\|^4|q_{it} = \gamma_1) \le A$, and $\max_i E(\|Z_{it}\|^4|q_{it} = \gamma_2) \le A$ for some $A < \infty$, and $\max_i f_{it}(\gamma_1) \le \bar{f}_1 < \infty$, $\max_i f_{it}(\gamma_2) \le \bar{f}_2 < \infty$.

(iii) $D_1(\gamma_1) = D(\gamma_1)$ and $V_1(\gamma_1) = V(\gamma_1)$ are continuous at $\gamma_1 = \gamma_{01}$, and $D_2(\gamma_2) = D(\gamma_2)$ and $V_2(\gamma_2) = V(\gamma_2)$ are continuous at $\gamma_2 = \gamma_{02}$.

(iv) $0 < C_1'D_1C_1, C_2'D_2C_2, C_1'V_1C_1, C_2'V_2C_2, < \infty$.

Theorem 3.5. *Let Assumption 3.4 to Assumption 3.6 hold. As $n \to \infty$ and $T \to \infty$,*

$$
\sqrt{nT}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0_{(3K+3)}, \Sigma_{\gamma_{01},\gamma_{02}}^{-1} \Sigma_{\gamma_{02},\gamma_{02},\epsilon} \Sigma_{\gamma_{01},\gamma_{02}}^{-1}). \tag{3.6.12}
$$

Remark 3.7. By this theorem, the asymptotic variance for the estimator $\hat{\beta}$ of multiple threshold model can hence be estimated through the estimators of

$$\hat{\Sigma}_{\hat{\gamma}_1,\hat{\gamma}_2} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} Z^*_{it}(\hat{\gamma}_1,\hat{\gamma}_2)(Z^*_{it}(\hat{\gamma}_1,\hat{\gamma}_2))',$$
$$\hat{\Sigma}_{\hat{\gamma}_1,\hat{\gamma}_2,\hat{\epsilon}} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T} Z^*_{it}(\hat{\gamma}_1,\hat{\gamma}_2)(Z^*_{jt}(\hat{\gamma}_1,\hat{\gamma}_2))'\hat{\epsilon}_{it}\hat{\epsilon}_{jt}, \tag{3.6.13}$$

where $\hat{\epsilon}_{it}$ is the residual of an estimated model of (3.6.1). The estimated variance of $\hat{\beta}$ is consistent. Also, the confidence interval of the components of the true parameter vector β_0 of β can be calculated by using estimated asymptotic variance (3.6.13).

Theorem 3.6. *Under Assumption 3.4 to Assumption 3.6 , as $n \to \infty$ and $T \to \infty$,*

$$n^{1-2\alpha_1}T^{1-2\alpha_2}(\hat{\gamma}-\gamma_0) \xrightarrow{d} \begin{pmatrix} w_1 \arg\max_{v_1\in R}\left[-\frac{|v_1|}{2}+W_1(v_1)\right] \\ w_2 \arg\max_{v_2\in R}\left[-\frac{|v_2|}{2}+W_2(v_2)\right]\end{pmatrix}. \tag{3.6.14}$$

where $w_1 = \frac{C_1'V_1C_1}{(C_1'D_1C_1)^2}$ and $w_2 = \frac{C_2'V_2C_2}{(C_2'D_2C_2)^2}$, $W_1(v_1)$ and $W_2(v_2)$ are two independent two-sided Brownian motions.

3.7 Simulated examples

This section explores the finite sample performance of the OLS estimators with cross-sectional dependence. Here, we consider different cases of the fixed effect and regressors.

3.7.1 Predetermined threshold parameters

Example 3.1. Consider the following model,

$$Y_{it} = \beta_1 X_{it} I(X_{it} \leq \gamma) + \beta_2 X_{it} I(X_{it} > \gamma) + \epsilon_{it}, \tag{3.7.1}$$

where $\beta_1 = 1$, $\beta_2 = 2$ and $\gamma = 0.15$. Let $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{nt})$, which is an n-dimensional vector. Then $\{\epsilon_t, 1 \leq t \leq T\}$ is generated as a sequence of

n-dimensional vector of independent Gaussian variables with zero mean and covariance matrix $(a_{ij})_{n\times n}$, where

$$a_{ij} = 0.8^{|j-i|},\ 1 \le i, j \le n. \tag{3.7.2}$$

Since the cross-sectional dependence is considered, the form of covariance matrix is different from the normal way. And it is obvious that we have

$$\begin{aligned} &E(\epsilon_{it}\epsilon_{js}) = 0 \quad for\ 1 \le i, j \le n,\ t \ne s, \\ &E(\epsilon_{it}\epsilon_{jt}) = 0.8^{|j-i|} \quad for\ 1 \le i, j \le n,\ 1 \le t \le T. \end{aligned} \tag{3.7.3}$$

And the above equations imply that the error term is dependent cross-sectionally. In the previous section, the independent variables consist of the common factor x_t and the error term v_{it}, that is

$$X_{it} = c_t + v_{it}, \quad i = 1, ..., n,\ t = 1, ..., T \tag{3.7.4}$$

where $c_t \sim U(0, 0.2)$ for $1 \le t \le T$. Let $v_t = (v_{1t}, v_{2t}, ..., v_{nt})'$ for $1 \le t \le T$. Consider that $\{v_t, t \ge 1\}$ is a stationary α-mixing sequence, it is generated by the following equation,

$$v_{it} = 0.7v_{i,t-1} + u_{it}, \tag{3.7.5}$$

where $u_{it} \sim N(0, 1)$.

In our experiments, we do the regression 500 times and take the average the parameter estimators. We examine the means, standard deviations and mean squared errors (MSE) of slope coefficients β. The MSE of β is defined as follows,

$$MSE = \frac{1}{500}\sum_{i=1}^{500}(\hat{\beta}^i - \beta)^2. \tag{3.7.6}$$

where $\hat{\beta}^i$ is the estimator of i-th sample. By giving different values of n and T, we compare the influence of sample size on β. For setting $n = 10, 20, 30$ and $T = 10, 20, 30$, the simulation results are reported in Table 3.1 and Table 3.2.

Table 3.1 shows the means and standard deviation of parameter β with 500 iterations. Table 3.2 indicates the mean squared errors of parameter β. We can find that with the sample size, either n or T, increasing, the standard deviation

TABLE 3.1: Means and SDs of slope coefficient estimators of Example 3.1

$n\backslash T$		10	20	30
10	β_1	1.0203 (0.3254)	1.0015 (0.1811)	0.9997 (0.1183)
	β_2	2.0116 (0.2563)	2.0015 (0.1430)	2.0029 (0.1210)
20	β_1	1.0012 (0.2094)	1.0055 (0.1133)	1.0017 (0.0907)
	β_2	2.0045 (0.1734)	1.9897 (0.1213)	2.0033 (0.0757)
30	β_1	1.0019 (0.1507))	1.0003 (0.1117))	0.9983 (0.0719)
	β_2	1.9945 (0.1392)	2.0092 (0.0771)	2.0016 (0.0697)

TABLE 3.2: MSE of slope coefficient estimators of Example 3.1

$n\backslash T$		10	20	30
10	β_1	0.1061	0.0327	0.0140
	β_2	0.0657	0.0204	0.0146
20	β_1	0.0438	0.0128	0.0082
	β_2	0.0300	0.0148	0.0057
30	β_1	0.0227	0.0125	0.0052
	β_3	0.0194	0.0060	0.0048

and mean squared errors of the slope coefficients will decrease.

Example 3.2. The example 3.1 can be seen that the fixed effect is equal to zero. Consider the constant term with the indicator function, we have the following model,

$$Y_{it} = \alpha_i + \beta_0 I(X_{it} \leq \gamma) + \beta_1 X_{it} I(X_{it} \leq \gamma) + \beta_2 X_{it} I(X_{it} > \gamma) + \epsilon_{it}, \quad i = 1, ..., n, t = 1, ...T \tag{3.7.7}$$

where $\beta_1 = 1$, $\beta_2 = 2$ and $\gamma = 0.15$ are the same as Example 3.1. And we set $\beta_0 = 0.5$, $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ for $i = 1, ...n-1$, and $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$. Then, the fixed effects satisfy that $\sum_{i=1}^{n} \alpha_i = 0$ for model identifiability. And the covariate variables X_{it} and the error term ϵ_{it} are generated as Example 3.1 with cross-sectional dependence.

Table 3.3 and Table 3.4 shows similar results to Example 3.1. Although we add the constant term with indicator function and use the different form of fixed

TABLE 3.3: Means and SDs of slope coefficient estimators of Example 3.2

$n\backslash T$		10	20	30
10	β_0	0.4921	0.4857	0.5075
		(0.2885)	(0.2270)	(0.1855)
	β_1	0.9896	0.9833	1.0077
		(0.2789)	(0.1335)	(0.1212)
	β_2	1.9922	2.0029	2.0036
		(0.2120)	(0.1543)	(0.1308)
20	β_0	0.4992	0.4887	0.5006
		(0.2594)	(0.1640)	(0.1466)
	β_1	1.0144	0.9947	1.0037
		(0.1669)	(0.1087)	(0.0916)
	β_2	2.0097	1.9949	2.0026
		(0.1890)	(0.1255)	(0.0922)
30	β_0	0.5023	0.5041	0.4917
		(0.1864)	(0.1507)	(0.1350)
	β_1	1.0127	1.0030	0.9929
		(0.1483)	(0.0953)	(0.0659)
	β_2	1.9951	2.0005	1.9947
		(0.1779)	(0.0976)	(0.0734)

TABLE 3.4: MSE of slope coefficient estimators of Example 3.2

$n\backslash T$		10	20	30
10	β_0	0.0831	0.0516	0.0344
	β_1	0.0777	0.0181	0.0147
	β_2	0.0449	0.0238	0.0171
20	β_0	0.0671	0.0270	0.0215
	β_1	0.0280	0.0118	0.0083
	β_2	0.0357	0.0157	0.0085
30	β_0	0.0347	0.0227	0.0183
	β_1	0.0221	0.0091	0.0044
	β_2	0.0316	0.0095	0.0054

effect, either the cross-section sample size or time series sample size increases, the standard deviations and mean squared errors of parameters will decrease.

Example 3.3. Consider the model in Example 3.2, the fixed effects are generated under the condition that $\alpha_i = \frac{1}{T}\sum_{i=1}^{T} X_{it}$ for $i = 1, ...n-1$, and $\alpha_n = -\sum_{i=1}^{n-1}\alpha_i$. Thus, they satisfy $\sum_{i=1}^{n}\alpha_i = 0$. Another commonly used method to generate the fixed effect is that $\alpha_i \sim N(0,1)$ for $i = 1, ..., n$ is assumed in simulated examples. And the other parameters, independent variables and error terms are generated the same as them in Example 3.2. And we can get the following results in Table

3.5 and Table 3.6.

TABLE 3.5: Means and SDs of slope coefficient estimators of Example 3.3

$n\backslash T$		10	20	30
10	β_0	0.5072 (0.4228)	0.4914 (0.2252)	0.4875 (0.2079)
	β_1	1.0013 (0.2309)	1.0094 (0.1984)	1.0014 (0.1236)
	β_2	1.9987 (0.2553)	1.9843 (0.1845)	1.9919 (0.1472)
20	β_0	0.4978 (0.2270)	0.5116 (0.1795)	0.5021 (0.1592)
	β_1	0.9926 (0.1811)	1.0009 (0.1201)	0.9962 (0.1109)
	β_2	1.9973 (0.1665)	1.9989 (0.1094)	2.0021 (0.0933)
30	β_0	0.4923 (0.2183)	0.5002 (0.1491)	0.5064 (0.1325)
	β_1	0.9982 (0.1183)	1.0013 (0.0889)	1.0016 (0.0774)
	β_2	1.9926 (0.1577)	1.9984 (0.0904)	2.0014 (0.0796)

TABLE 3.6: MSE of slope coefficient estimators of Example 3.3

$n\backslash T$		10	20	30
10	β_0	0.1785	0.0507	0.0433
	β_1	0.0532	0.0394	0.0152
	β_2	0.0651	0.0342	0.0217
20	β_0	0.0514	0.0323	0.0253
	β_1	0.0328	0.0145	0.0123
	β_2	0.0277	0.0119	0.0087
30	β_0	0.0476	0.0222	0.0176
	β_1	0.0140	0.0079	0.0060
	β_2	0.0249	0.0082	0.0063

Despite using different forms of fixed effect from Example 3.2, we can get a similar conclusion. The standard deviations and mean squared errors of parameters will decrease as either the cross-section data sample size or time series sample size increases.

Example 3.4. Consider the model in Example 3.2,

$$y_{it} = \alpha_i + \beta_0 I(X_{it} \leq \gamma) + \beta_1 X_{it} I(X_{it} \leq \gamma) + \beta_2 X_{it} I(X_{it} > \gamma) + \epsilon_{it}, \tag{3.7.8}$$

where $\beta_0 = 0.5$, $\beta_1 = 1$ and $\beta_2 = 2$ set as Example 3.2. ϵ_{it} is generated as Example 3.1 with cross-sectional dependence. The regressor variables are in the form of $X_{it} = c_t + v_{it}$. Differently, c_t is generated by an AR(1) model as follows,

$$c_t = 0.1c_{t-1} + e_t, \quad 1 \leq t \leq T \tag{3.7.9}$$

where the parameter of AR(1) model is 0.1, which means weak dependence. $e_t \sim N(0,1)$, that is generated by the normal distribution. v_{it} is generated as Example 3.1. In this example, we set the threshold parameter is equal to 0.

TABLE 3.7: Means and SDs of slope coefficient estimators of Example 3.4

$n\backslash T$		10	20	30
10	β_0	0.5179 (0.3497)	0.4943 (0.2899)	0.4977 (0.2640)
	β_1	1.0114 (0.1878)	0.994 (0.1554)	0.999 (0.1202)
	β_2	1.9957 (0.2673)	2.0012 (0.2047)	1.9955 (0.1734)
20	β_0	0.5038 (0.2346)	0.4902 (0.1906)	0.4904 (0.1599)
	β_1	1.0033 (0.1299)	0.9989 (0.1121)	0.9946 (0.1086)
	β_2	2.0065 (0.2137)	1.9945 (0.1230)	1.9954 (0.0940)
30	β_0	0.5177 (0.2306)	0.5104 (0.1632)	0.4982 (0.1170)
	β_1	1.0041 (0.1109)	1.0002 (0.0948)	1.0005 (0.08221)
	β_1	2.0079 (0.1930)	1.9994 (0.1136)	1.9977 (0.0788)

Unlike the uniform distribution of x_t in Example 3.1 and Example 3.2, it is generated by AR(1) model and dependent on x_{t-1} weakly. The simulation results in Table 3.7 and Table 3.8 indicate that an increase in either n or T leads to a decrease in SD and MSE.

TABLE 3.8: MSE of slope coefficient estimators of Example 3.4

$n\backslash T$		10	20	30
10	β_0	0.1224	0.0839	0.0695
	β_1	0.0353	0.0241	0.0144
	β_2	0.0713	0.0418	0.0300
20	β_0	0.0550	0.0364	0.0256
	β_1	0.0169	0.0125	0.0118
	β_2	0.0456	0.0151	0.0088
30	β_0	0.0534	0.0267	0.0137
	β_1	0.0123	0.0090	0.0067
	β_2	0.0372	0.0129	0.0062

Example 3.5. Example 3.4 consider the weak dependence of c_t. We consider a larger coefficient in the common factor function, which means the dependence is much stronger than Example 3.4. It is generated as follows,

$$c_t = 0.5c_{t-1} + e_t, \quad 1 \leq t \leq T. \tag{3.7.10}$$

Other parameters and the form of error term is chosen the same as Example 3.4. We can get the following results.

TABLE 3.9: Means and SDs of slope coefficient estimators of Example 3.5

$n\backslash T$		10	20	30
10	β_0	0.5082	0.4926	0.5063
		(0.4205)	(0.3325)	(0.2653)
	β_1	1.0078	0.9970	1.0008
		(0.2596)	(0.1441)	(0.1201)
	β_2	2.0081	2.0018	1.9967
		(0.2905)	(0.1985)	(0.1500)
20	β_0	0.5072	0.4867	0.4960
		(0.2998)	(0.1854)	(0.1571)
	β_1	0.9998	0.9975	0.9971
		(0.1236)	(0.1124)	(0.1039)
	β_2	2.0075	1.9972	1.9978
		(0.2686)	(0.1240)	(0.0895)
30	β_0	0.5083	0.5131	0.4980
		(0.2589)	(0.1639)	(0.1271)
	β_1	0.9993	1.0018	1.0027
		(0.1094)	(0.0864)	(0.0854)
	β_2	2.0049	1.9975	1.9986
		(0.2371)	(0.1002)	(0.0741)

TABLE 3.10: MSE of slope coefficient estimators of Example 3.5

$n\backslash T$		10	20	30
	β_0	0.1765	0.1104	0.0703
10	β_1	0.0673	0.0207	0.0144
	β_2	0.0843	0.0393	0.0225
	β_0	0.0898	0.0345	0.0246
20	β_1	0.0153	0.0126	0.0108
	β_2	0.0721	0.0153	0.0080
	β_0	0.0670	0.0270	0.0161
30	β_1	0.0119	0.0074	0.0073
	β_2	0.0561	0.0100	0.0055

Although a larger parameter of AR(1) model is considered in this example, we can get the same conclusion that an increase in either n or T results in a decrease in standard deviation and mean squared errors.

3.7.2 Estimated threshold parameters

Example 3.6. Consider the following single threshold model:

$$Y_{it} = \alpha_i + \beta_1 X_{it} I(X_{it} \le \gamma) + \beta_2 I(X_{it} > \gamma) + \beta_3 X_{it} I(X_{it} > \gamma) + \epsilon_{it}$$

where $\beta_1 = -0.5$, $\beta_2 = 0.3$, $\beta_3 = 0.7$ and $\gamma = 0$. Considering the cross-sectional dependent and serial independent structure, the error terms ϵ_{it} is generated as follows. For $t \ge 1$, $\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{nT})$ is generated as a sequence of n-dimensional vector of independent Gaussian variables drawn from $N(0, \Sigma_c)$, with the (i, j)-th element of of Σ_c being $c_{ij} = 0.6^{|i-j|}$, $1 \le i, j \le n$. The regressors X_{it} are generated by

$$X_{it} = c_t + v_{it}$$

where v_{it} is independent of ϵ_{it} and follows

$$v_{it} = 0.4 v_{i,t-1} + u_{it}$$

in which u_{it} is serial cross-sectional independent and generated from $N(0, 1)$. $\alpha_i \sim N(0, 1)$; $c_t \sim U(-0.2, 0.2)$ for $1 \le t \le T$.

TABLE 3.11: Means and SDs of estimators of Example 3.6

$n\backslash T$		20	30	40
20	β_1	-0.4671	-0.4767	-0.4956
		(0.1764)	(0.1286)	(0.1072)
	β_2	0.2702	0.2878	0.3001
		(0.3039)	(0.2034)	(0.1745)
	β_3	0.6733	0.6791	0.6878
		(0.1378)	(0.1089)	(0.0823)
	γ	-0.1324	-0.1267	-0.1122
		(0.3609)	(0.2850)	(0.2485)
30	β_1	-0.4823	-0.4778	-0.4904
		(0.1336)	(0.1059)	(0.0908)
	β_2	0.2888	0.2888	0.2972
		(0.2427)	(0.1746)	(0.1412)
	β_3	0.6820	0.6866	0.6902
		(0.0993)	(0.0821)	(0.0649)
	γ	-0.1252	-0.1065	-0.0991
		(0.3044)	(0.2423)	(0.2147)
40	β_1	-0.4868	-0.4897	-0.4970
		(0.1031)	(0.0822)	(0.0716)
	β_2	0.2979	0.2980	0.2999
		(0.1755)	(0.1396)	(0.1137)
	β_3	0.6816	0.6891	0.6962
		(0.0948)	(0.0679)	(0.0586)
	γ	-0.1129	-0.1061	-0.0752
		(0.2550)	(0.2260)	(0.1840)

TABLE 3.12: MSE of estimators of Example 3.6

$n\backslash T$		20	30	40
20	β_1	0.0322	0.0170	0.0115
	β_2	0.0930	0.0414	0.0304
	β_3	0.0197	0.0123	0.0069
	γ	0.1475	0.0971	0.0742
30	β_1	0.0181	0.0117	0.0083
	β_2	0.0589	0.0306	0.0199
	β_3	0.0102	0.0069	0.0043
	γ	0.1082	0.0699	0.0558
40	β_1	0.0108	0.0068	0.0051
	β_2	0.0308	0.0195	0.0129
	β_3	0.0093	0.0047	0.0034
	γ	0.0776	0.0622	0.0394

Example 3.7. Consider the single threshold model in Example 3.6:

$$Y_{it} = \alpha_i + \beta_1 X_{it} I(X_{it} \leq \gamma) + \beta_2 I(X_{it} > \gamma) + \beta_3 X_{it} I(X_{it} > \gamma) + \epsilon_{it},$$

where $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$ for $1 \leq i \leq n-1$ and $\alpha_n = -\sum_{i=1}^{n-1}\alpha_i$; $c_t \sim U(-0.2, 0.2)$ for $1 \leq t \leq n-1$ and $c_T = -\sum_{t=1}^{T-1} c_t$. Other settings are the same as Example 3.6.

TABLE 3.13: Means and SDs of estimators of Example 3.7

$n\backslash T$		20	30	40
20	β_1	-0.4689 (0.1668)	-0.4714 (0.1312)	-0.4831 (0.1079)
	β_2	0.2747 (0.2920)	0.2800 (0.2184)	0.2835 (0.1748)
	β_3	0.6778 (0.1361)	0.6827 (0.0940)	0.6901 (0.0761)
	γ	-0.1752 (0.3466)	-0.1326 (0.3103)	-0.1503 (0.2546)
30	β_1	-0.4808 (0.1206)	-0.4931 (0.1048)	-0.4947 (0.0800)
	β_2	0.2974 (0.2320)	0.2965 (0.1654)	0.3021 (0.1286)
	β_3	0.6793 (0.1111)	0.6864 (0.0677)	0.6875 (0.0673)
	γ	-0.1320 (0.2896)	-0.1156 (0.2582)	-0.0890 (0.2078)
40	β_1	-0.4772 (0.1046)	-0.4814 (0.0825)	-0.5013 (0.0672)
	β_2	0.2749 (0.1782)	0.2882 (0.1428)	0.3033 (0.1150)
	β_3	0.6943 (0.0900)	0.6912 (0.0637)	0.6959 (0.0549)
	γ	-0.1515 (0.2630)	-0.1419 (0.2429)	-0.0820 (0.1846)

TABLE 3.14: MSE of estimators of Example 3.7

$n\backslash T$		20	30	40
20	β_1	0.0287	0.0180	0.0119
	β_2	0.0857	0.0480	0.0308
	β_3	0.0190	0.0091	0.0059
	γ	0.1506	0.1137	0.0873
30	β_1	0.0149	0.0110	0.0064
	β_2	0.0537	0.0273	0.0165
	β_3	0.0127	0.0048	0.0047
	γ	0.1011	0.0799	0.0510
40	β_1	0.0114	0.0071	0.0045
	β_2	0.0323	0.0205	0.0132
	β_3	0.0081	0.0041	0.0030
	γ	0.0920	0.0790	0.0407

Example 3.8. Consider the single threshold model in Example 3.7. We consider AR(1) model of c_t, $c_t = 0.3c_{t-1} + w_t$, in which $w_t \sim N(0, 1)$. Other settings are the same as Example 3.7.

TABLE 3.15: Means and SDs of estimators of Example 3.8

$n\backslash T$		20	30	40
20	β_1	-0.4663	-0.4648	-0.4854
		(0.1668)	(0.1312)	(0.1079)
	β_2	0.2764	0.2685	0.2895
		(0.2695)	(0.2140)	(0.1818)
	β_3	0.6735	0.6883	0.6898
		(0.1361)	(0.0940)	(0.0761)
	γ	-0.1561	-0.1796	-0.1038
		((0.4075)	(0.3173)	(0.2572)
30	β_1	-0.4742	-0.4836	-0.4976
		(0.1262)	(0.1015)	(0.0583)
	β_2	0.2672	0.2820	0.2987
		(0.2186)	(0.1696)	(0.1292)
	β_3	0.6893	0.6928	0.6948
		(0.0793)	(0.0689)	(0.0622)
	γ	-0.1750	-0.1423	-0.0933
		(0.3301)	(0.2763)	(0.2291)
40	β_1	-0.4839	-0.4976	-0.5000
		(0.0816)	(0.0635)	(0.0507)
	β_2	0.3010	0.3015	0.3019
		(0.1681)	(0.1265)	(0.1193)
	β_3	0.6842	0.6983	00.6948
		(0.0790)	(0.0595)	(0.0473)
	γ	-0.1518	-0.0859	-0.0969
		(0.2958)	(0.2202)	(0.2125)

TABLE 3.16: MSE of estimators of Example 3.8

$n\backslash T$		20	30	40
20	β_1	0.0207	0.0139	0.0095
	β_2	0.0731	0.0467	0.0331
	β_3	0.0146	0.0099	0.0043
	γ	0.1901	0.1327	0.0768
30	β_1	0.0166	0.0105	0.0034
	β_2	0.0488	0.0290	0.0167
	β_3	0.0065	0.0048	0.0039
	γ	0.1394	0.0965	0.0611
40	β_1	0.0069	0.0040	0.0026
	β_2	0.0282	0.0160	0.0142
	β_3	0.0064	0.0035	0.0023
	γ	0.1104	0.0558	0.0545

We examine the mean, standard deviation and mean squared error (MSE) of the proposed estimators with 500 iterations. For each example, we consider three cross-sectional sizes, $n = \{20, 30, 40\}$, and three time lengths, $T = \{20, 30, 40\}$, which results in nine sample size combinations. The simulation results are reported in Table 3.11 to Table 3.16. We can find that for each example, an increase in either n or T leads to a decrease in standard deviation. Similar conclusion can be drawn from MSE of estimators, where the MSE of proposed estimators decreases as either n or T increases.

Example 3.9. Consider the double threshold model as follows,

$$\begin{aligned} Y_{it} =& \alpha_i + \beta_1 X_{it} I(X_{it} \leq \gamma_1) + \beta_2 I(\gamma_1 < X_{it} \leq \gamma_2) + \beta_3 X_{it} (\gamma_1 < X_{it} \leq \gamma_2) \\ &+ \beta_4 I(X_{it} > \gamma_2) + \beta_5 X_{it} I(X_{it} > \gamma_2) + \epsilon_{it}, \end{aligned}$$

where $\beta_1 = -0.5$, $\beta_2 = 0.4$, $\beta_3 = 0.7$, $\beta_4 = -0.8$, $\beta_5 = -0.3$ and $\gamma_1 = -1$, $\gamma_2 = 1$. The data generation processes follow that of Example 3.6.

Example 3.10. Consider the double threshold model in Example 3.10,

$$\begin{aligned} Y_{it} =& \alpha_i + \beta_1 X_{it} I(X_{it} \leq \gamma_1) + \beta_2 I(\gamma_1 < X_{it} \leq \gamma_2) + \beta_3 X_{it} (\gamma_1 < X_{it} \leq \gamma_2) \\ &+ \beta_4 I(X_{it} > \gamma_2) + \beta_5 X_{it} I(X_{it} > \gamma_2) + \epsilon_{it}, \end{aligned}$$

The data generation processes follow that of Example 3.7.

The conclusion of single threshold model can also be applied to double threshold model, where an increase in either n or T results in a decrease in both SD and MSE of proposed estimators.

3.7.3 Comparison

Example 3.11. Consider the following single threshold model:

$$Y_{it} = \alpha_i + \beta_1 X_{it} I(X_{it} \leq \gamma) + \beta_2 I(X_{it} > \gamma) + \beta_3 X_{it} I(X_{it} > \gamma) + \epsilon_{it}$$

where $\beta_1 = -0.5$, $\beta_2 = 0.3$, $\beta_3 = 0.7$ and $\gamma = 0$. The fixed effect α_i is generated from $N(0, 1)$. Considering the cross-sectional dependent and serial independent structure, the error terms ϵ_{it} is generated as follows. For $t \geq 1$,

TABLE 3.17: Means and SDs of estimators of double threshold model

$n\backslash T$		Example 3.9			Example 3.10		
		20	30	40	20	30	40
20	β_1	-0.4881	-0.5085	-0.4956	-0.4865	-0.4690	-0.5018
		(0.2706)	(0.2557)	(0.1826)	(0.2781)	(0.2524)	(0.1958)
	β_2	0.3900	0.4189	0.3922	0.3877	0.3488	0.4028
		(0.4794)	(0.4359)	(0.3290)	(0.5270)	(0.4336)	(0.3325)
	β_3	0.7085	0.6875	0.7044	0.6595	0.6829	0.7034
		(0.1620)	(0.1348)	(0.0974)	(0.2570)	(0.1139)	(0.0956)
	β_4	-0.7633	-0.7040	-0.8022	-0.7198	-0.8461	-0.7273
		(0.6684)	(0.5431)	(0.4119)	(0.6078)	(0.5456)	(0.4298)
	β_5	-0.3272	-0.3494	-0.3040	-0.3515	-0.3048	-0.3357
		(0.2926)	(0.2218)	(0.1555)	(0.2222)	(0.2024)	(0.1728)
	γ_1	-0.7419	-0.9013	-0.9504	-0.7205	-0.9363	-0.9564
		(0.5987)	(0.3097)	(0.2538)	(0.5867)	(0.2816)	(0.2225)
	γ_2	0.8184	0.9701	0.9657	0.8208	0.9769	0.9739
		(0.5543)	(0.2181)	(0.1880)	(0.5308)	(0.1968)	(0.1690)
30	β_1	-0.4934	-0.5175	-0.5017	-0.5029	-0.4910	-0.4945
		(0.2515)	(0.1995)	(0.1426)	(0.2645)	(0.1994)	(0.1606)
	β_2	0.3841	0.4256	0.3996	0.3973	0.3817	0.3928
		(0.4107)	(0.3518)	(0.2518)	(0.4463)	(0.3374)	(0.2710)
	β_3	0.6803	0.6950	0.6886	0.6830	0.7071	0.7003
		(0.1183)	(0.0850)	(0.0791)	(0.1326)	(0.0834)	(0.0739)
	β_4	-0.8062	-0.7193	-0.7556	-0.6985	-0.7830	-0.7418
		(0.4969)	(0.4389)	(0.3375)	(0.5521)	(0.4196)	(0.3707)
	β_5	-0.3055	-0.3200	-0.3240	-0.3533	-0.3163	-0.3339
		(0.1917)	(0.1603)	(0.1455)	(0.2094)	(0.1491)	(0.1457)
	γ_1	-0.9362	-0.9676	-0.9842	-0.8955	-0.9544	-0.9739
		(0.2750)	(0.1854)	(0.1348)	(0.3389)	(0.1610)	(0.1037)
	γ_2	0.9933	0.9907	1.0023	0.9353	0.9888	0.9988
		(0.2190)	(0.0942)	(0.0899)	(0.2612)	(0.0898)	(0.0058)
40	β_1	0.4810	-0.5054	-0.5043	-0.5033	-0.4911	-0.4976
		(0.2066)	(0.1466)	(0.1167)	(0.1592)	(0.1518)	(0.1212)
	β_2	0.3600	0.4090	0.3980	0.4035	0.3831	0.3975
		(0.3526)	(0.2463)	(0.1984)	(0.2945)	(0.2539)	(0.2092)
	β_3	0.6820	0.6943	0.6952	0.7080	0.7045	0.7007
		(0.1100)	(0.0717)	(0.0650)	(0.0939)	(0.0708)	(0.0571)
	β_4	-0.7852	-0.7401	-0.7802	-0.8020	-0.7729	-0.7572
		(0.4489)	(0.3587)	(0.2834)	(0.4124)	(0.3609)	(0.2840)
	β_5	-0.3277	-0.3271	-0.3091	-0.2992	-0.3176	-0.3266
		(0.1589)	(0.1536)	(0.1116)	(0.1606)	(0.1430)	(0.1271)
	γ_1	-0.9507	-1.000	-0.9813	-0.9438	-0.9865	-1.0027
		(0.2433)	(0.0750)	(0.0589)	(0.2717)	(0.1175)	(0.0380)
	γ_2	0.9764	1.003	0.9997	0.9700	0.9911	0.9974
		(0.1899)	(0.0083)	(0.0028)	(0.2356)	(0.0894)	(0.0049)

$\epsilon_t = (\epsilon_{1t}, ..., \epsilon_{nT})$ is generated as a sequence of n-dimensional vector of independent Gaussian variables drawn from $N(0_n, \Sigma_e)$, with the (i, j)-th element of

TABLE 3.18: MSE of estimators of double threshold model

$n\backslash T$		Example 3.9			Example 3.10		
		20	30	40	20	30	40
20	β_1	0.0732	0.0653	.0333	0.0774	0.0645	0.0383
	β_2	0.2294	0.1900	0.1081	0.2773	0.1903	0.1103
	β_3	0.0263	0.0183	0.0095	0.0675	0.0132	0.0092
	β_4	0.4472	0.3036	0.1693	0.3751	0.2992	0.1896
	β_5	0.0862	0.0515	0.0242	0.0519	0.0409	0.0311
	γ_1	0.4244	0.1055	0.0668	0.4216	0.0832	0.0513
	γ_2	0.3396	0.0483	0.0365	0.3133	0.0392	0.0292
30	β_1	0.0632	0.0400	0.0203	0.0698	0.0398	0.0258
	β_2	0.1686	0.1242	0.0633	0.1988	0.1139	0.0734
	β_3	0.0144	0.0072	0.0064	0.0178	0.0070	0.0055
	β_4	0.2465	0.1987	0.1156	0.3145	0.1760	0.1405
	β_5	0.0367	0.0261	0.0217	0.0466	0.0225	0.0223
	γ_1	0.0795	0.0353	0.0184	0.1255	0.0280	0.0114
	γ_2	0.0479	0.0089	0.0081	0.0723	0.0082	0.0000
40	β_1	0.0429	0.0214	0.0136	0.0253	0.0231	0.0147
	β_2	0.1256	0.0606	0.0393	0.0866	0.0647	0.0437
	β_3	0.0124	0.0052	0.0042	0.0089	0.0050	0.0033
	β_4	0.2013	0.1320	0.0806	0.1697	0.1308	0.0823
	β_5	0.0260	0.0243	0.0125	0.0258	0.0207	0.0168
	γ_1	0.0615	0.0056	0.0038	0.0769	0.0140	0.0014
	γ_2	0.0365	0.0000	0.0000	0.0563	0.0081	0.0000

of Σ_e being $e_{ij} = 0.6^{|i-j|}, 1 \le i, j \le n$. The regressors X_{it} are generated by

$$X_{it} = c_t + v_{it}$$

where v_{it} is independent of ϵ_{it} and follows

$$v_{it} = 0.4 v_{i,t-1} + u_{it}$$

in which u_{it} is serial cross-sectional independent and generated from $N(0_n, \Sigma_u)$, where Σ_u is a diagonal matrix and $\Sigma_{u,ii} = \sigma^2_{u,i}$. $c_t = 8\sin(c_{t-1}) + w_t$, in which $w_t \sim N(0,1)$; $\sigma^2_{u,i} = 0$ for all i.

Example 3.12. Consider the single threshold model in Example 3.11:

$$Y_{it} = \alpha_i + \beta_1 X_{it} I(X_{it} \le \gamma) + \beta_2 I(X_{it} > \gamma) + \beta_3 X_{it} I(X_{it} > \gamma) + \epsilon_{it}$$

$c_t = 8\sin(c_{t-1}) + w_t$, in which $w_t \sim N(0,1)$; $\sigma^2_{u,i} = 1$ for all i. Other settings are the same as Example 3.11.

TABLE 3.19: Standard deviations of estimators

$n\backslash T$			Example 3.11			Example 3.12		
			50	100	300	50	100	300
20	β_1	SD_1	0.0359	0.0267	0.0151	0.0310	0.0217	0.0120
		SD_2	0.0323	0.0225	0.0141	0.0283	0.0204	0.0100
		SD_3	0.0169	0.0118	0.0074	0.0175	0.0127	0.0063
	β_2	SD_1	0.2758	0.2322	0.1226	0.2976	0.1989	0.0958
		SD_2	0.2718	0.1888	0.1166	0.2456	0.1627	0.0757
		SD_3	0.1428	0.0992	0.0612	0.1643	0.1069	0.0547
	β_3	SD_1	0.0351	0.0241	0.0148	0.0341	0.0233	0.0115
		SD_2	0.0328	0.0236	0.0141	0.0307	0.0203	0.0105
		SD_3	0.0172	0.0124	0.0074	0.0193	0.0127	0.0067
	γ	SD_1	0.6871	0.5534	0.3548	0.5839	0.3714	0.2266
50	β_1	SD_1	0.0240	0.0175	0.0092	0.0180	0.0138	0.0109
		SD_2	0.0202	0.0161	0.0084	0.0162	0.0122	0.0086
		SD_3	0.0103	0.0082	0.0043	0.0099	0.0074	0.0062
	β_2	SD_1	0.2089	0.1396	0.0800	0.1454	0.1171	0.0563
		SD_2	0.1811	0.1260	0.0713	0.1185	0.0944	0.0395
		SD_3	0.0923	0.0642	0.0363	0.0835	0.0654	0.0339
	β_3	SD_1	0.0237	0.0147	0.0100	0.0194	0.0150	0.0096
		SD_2	0.0233	0.0141	0.0094	0.0167	0.0132	0.0082
		SD_3	0.0119	0.0072	0.0048	0.0104	0.0084	0.0059
	γ	SD_1	0.4913	0.3679	0.2252	0.3343	0.2670	0.1465
100	β_1	SD_1	0.0159	0.0101	0.0081	0.0144	0.0096	0.0073
		SD_2	0.0140	0.0096	0.0073	0.0124	0.0084	0.0059
		SD_3	0.0070	0.0048	0.0037	0.0075	0.0052	0.0043
	β_2	SD_1	0.1439	0.1056	0.0591	0.1248	0.0732	0.0322
		SD_2	0.1220	0.1055	0.0533	0.1007	0.0574	0.0256
		SD_3	0.0616	0.0533	0.0269	0.0692	0.0408	0.0231
	β_3	SD_1	0.0191	0.0151	0.0063	0.0167	0.0092	0.0072
		SD_2	0.0179	0.0143	0.0061	0.0141	0.0082	0.0058
		SD_3	0.0090	0.0072	0.0031	0.0091	0.0051	0.0041
	γ	SD_1	0.4044	0.3017	0.1782	0.2958	0.1859	0.0595

We examine the standard deviation and mean squared error (MSE) of the proposed estimators based 500 repetitions in simulation. For each DGP, we consider three cross-sectional sample sizes, $n = 20, 50, 100$, respectively, and three time lengths, $T = 20, 50, 100$, respectively, which results in nine sample size combinations. The simulation results are reported in Table 3.19 and Table 3.20 for the Monte Carol standard error and the mean squared error (MSE) of the estimated parameters, respectively.

TABLE 3.20: MSE of estimators

$n\backslash T$		Example 3.11			Example 3.12		
		50	100	300	50	100	300
20	β_1	0.0014	0.0007	0.0002	0.0010	0.0005	0.0001
	β_2	0.0765	0.0539	0.0150	0.0884	0.0395	0.0092
	β_3	0.0013	0.0006	0.0002	0.0012	0.0005	0.0001
	γ	0.8857	0.4048	0.1624	0.3959	0.1722	0.0600
50	β_1	0.0006	0.0003	0.0001	0.0003	0.0002	0.0001
	β_2	0.0436	0.0195	0.0064	0.0211	0.0137	0.0032
	β_3	0.0006	0.0002	0.0001	0.0004	0.0002	0.0001
	γ	0.4324	0.2442	0.1057	0.1482	0.0885	0.0245
100	β_1	0.0003	0.0001	0.0001	0.0002	0.0001	0.0001
	β_2	0.0207	0.0111	0.0035	0.0155	0.0054	0.0010
	β_3	0.0004	0.0002	0.0000	0.0003	0.0001	0.0001
	γ	0.2742	0.1410	0.0368	0.1061	0.0404	0.0036

In Table 3.19, we display the different standard errors for the least squares estimators, where SD_1, SD_2 and SD_3 represent the Monte Carlo estimation based standard error, theoretical (i.e., asymptotic distribution based) standard error with cross-sectional dependence and theoretical standard error without consideration (i.e., ignoring) of cross-sectional dependence, respectively. Monte Carlo estimation result is defined as

$$SD_1 = \sqrt{\frac{1}{500}\sum_{i=1}^{500}(\beta_i - \bar{\beta})^2},$$

where β_i is the estimated coefficient in each iteration and $\bar{\beta}$ is the mean value. Theoretical standard error with taking consideration of cross-sectional dependence and without taking consideration of cross-sectional dependence can be calculated by Theorem 3.3 and Remark 3.2. We have the following findings from the simulation result,

(i) Overall, as expected, with increasing of the sample sizes in n and T, the standard deviations (SDs) for the estimators of the coefficients β's and the threshold parameter γ decrease significantly, and hence the estimators become more and more stable, in each case of different Examples.

In particular, it follows from Table 3.19 that the values of the Monte Carol based SD_1 for the estimators of the most important parameters $\beta_2 = -0.5$ and $\beta_3 = 0.7$ in Example 3.11 are approximately close to 0.0359 in the case of the sample size as small as $n = 20$ and $T = 50$, and even turn to be less than 0.0045 in the case of $n = 100$ and $T = 300$, which is close to the sample size of $(n, T) = (101, 339)$ in the real data example below. Relatively, the values of SD_1 for the intercept parameter $\beta_2 = 0.3$ and the threshold parameter $\gamma = 0$ are as large as 0.2758 and 0.6871, respectively, for the Example 3.11 in the case of $n = 20$ and $T = 50$, which decrease to 0.0269 and 0.1782 as the sample size increases to $n = 100$ and $T = 300$. This appears understandable in view of the smaller absolute values of $\beta_2 = 0.3$ and $\gamma = 0$, partially due to potentially weaker signals on them in the data. Similar outcomes can also be observed for another Example in Table 3.19.

(ii) Comparing SD_2 and SD_3 with SD_1 in Table 3.19, we can clearly see that the theoretical SD_2 that takes account of cross-sectional dependence in data is much closer to the SD_1 than the SD_3 that ignores the cross-sectional dependence in data.

This interesting finding warns us that if the cross-sectional dependence is wrongly ignored or specified, the standard deviations of the estimators may be seriously under-estimated by the theoretical SD_3, any inference based on which may hence lead to spurious significance for the estimated parameters. Clearly the SD_2 based on our developed asymptotic theory in Theorem 3.3 provides an accurate estimate of the standard deviations for inference.

(iii) Comparing the SDs under Example 3.11 and 3.12 in Table 3.19, we can appear to see that the SDs become smaller from Example 3.11 to 3.12 in particular when the sample sizes are relatively small. For example, look at the case for estimation of γ under the sample size $(n, T) = (20, 50)$: The SD_1 equals 0.6871 for Example 3.11, decreasing to 0.5839 for Example 3.12. But with increasing of the sample sizes, this phenomenon seems to become less viable.

This phenomenon is basically understandable as different Examples lead to different data of X_{it}'s. Under Example 3.11, the data of X_{it} may be more concentrated, only depending on the variation from the common factor c_t in X_{it}, while the data of X_{it} under Example 3.12 are more varied owing to the variations from both c_t and v_{it} in X_{it}. Somehow, this phenomenon is similar to that for the estimates of a linear regression model with different design matrices.

3.8 Empirical Study

We have studied the threshold effect of climate finance in Chapter 2. However, we also find that due to the cross-sectional dependence, the general panel threshold model may work poor. Thus in this section, we apply our proposed method to study the precipitation effect on stock returns. we first plot the relationship between precipitation and stock returns for each individual stock in FTSE 100 by nonparametric estimation. Figure 3.1 shows the nonparametric results of constitute stocks in FTSE 100. It is obvious that the relationship between precipitation and stock returns is nonlinear and the extreme weather, say heavy rainfall, has a great influence on the stock market. Some change points appear in Figure 3.1 by nonparametric estimation, which indicate that the threshold effects should be considered in estimating the effect of rainfall on the stock market. Therefore, panel threshold model may be an efficient method in this empirical application. Moreover, we only specify the precipitation in London as regressors so that we should consider the cross-sectional dependence in panel data analysis.

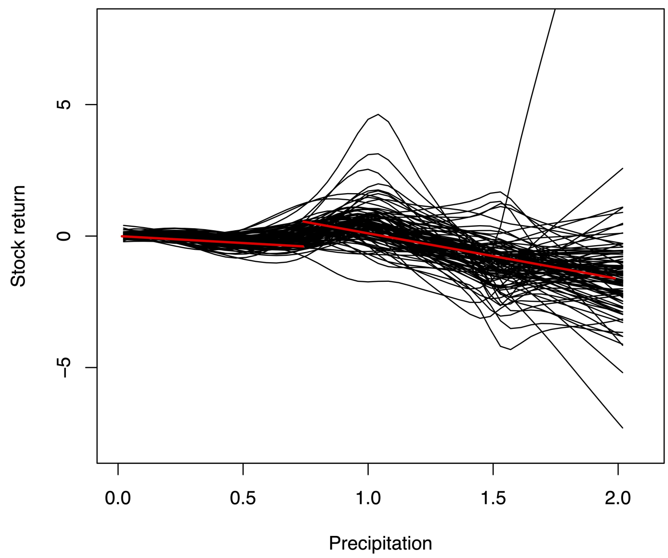


FIGURE 3.1: Plot of relationship between precipitation and stock returns

Motivated by nonparametric analysis, we consider panel threshold model with cross-sectional dependence. We calculate the standard deviation by proposed Econometric theory due to the consideration of cross-sectional dependence rather than that under identically independent distributed conditions. Hansen (1999) suggested to use balanced panels in threshold analysis. Due to different time to market of different stocks, we collect the data from March 19, 2015 to December 29, 2017 to form a balanced panel. Let R_{it} denote stock return and P_t denote precipitation, where $i = 1, 2, ..., 101$ indexes 101 stocks and $t = 1, 2, ..., 339$ indexes 339 days. Table 3.2 and Table 3.3 show some time series plots of panel data.

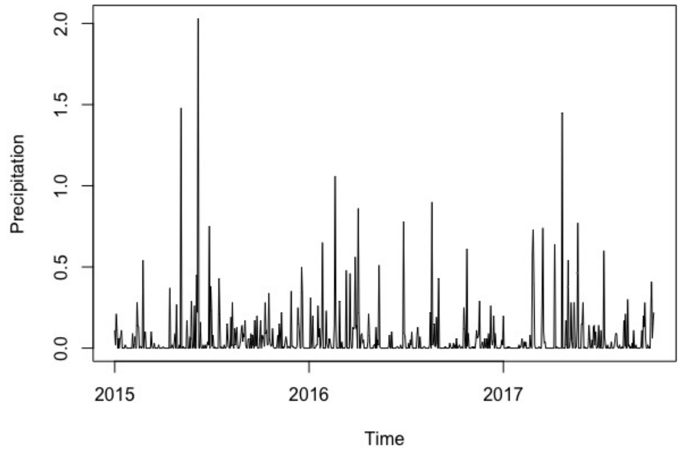
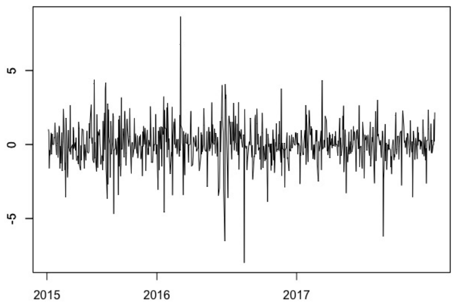
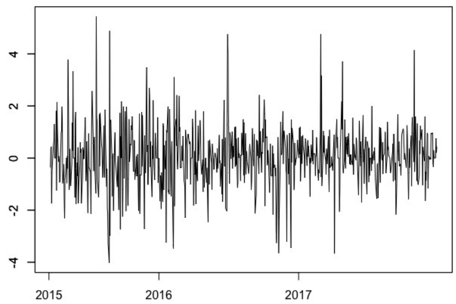


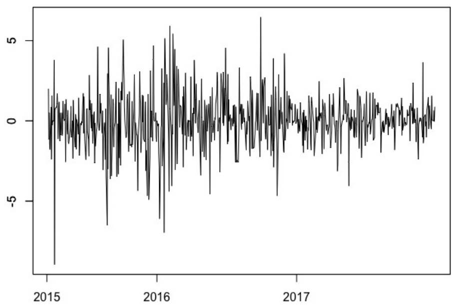
FIGURE 3.2: Plot of precipitation in panel data



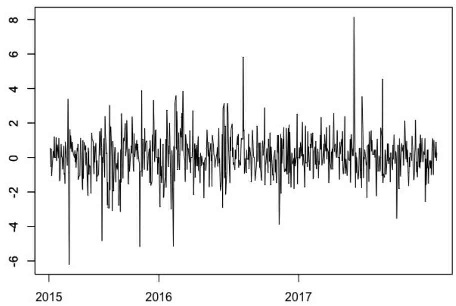
(A) Admiral.Group



(B) Croda.International



(C) Royal.Dutch.Shell.B



(D) Spirax.Sarco.Engr

FIGURE 3.3: Plots of stock returns in panel data

Since we examine the threshold effect of rainfall on stock returns, we only consider the days where precipitation is greater than zero. Summary statistics of the two variables are given in Table 3.21. We consider the following panel data model first.

$$R_{it} = \alpha_i + \beta P_t + \epsilon_{it}. \tag{3.8.1}$$

TABLE 3.21: Summary statistics

	Minimum	25% quantile	Median	Mean	75% quantile	Maximum
R_{it}	-23.3749	-0.7984	0	0.0200	0.8289	19.1055
P_t	0.010	0.020	0.060	0.135	0.150	2.030

TABLE 3.22: Estimates of the slope coefficients for linear panel model

	Coefficient estimate	P-value
Precipitation	−0.382588	0.062278
	(0.20521)[a]	

[a] Standard deviation.

Table 3.22 reports that linear panel model is statistically nonsignificant to study the rainfall effect on London stock market. However, if we apply the theorem ignoring cross-sectional dependence, linear panel model is statistically significant, which is spurious significance we mentioned in Section 3.6. The result of nonparametric analysis clearly shows the existence of threshold effect. We speculate that different precipitation level may have significant influence on the stock market. As a result, the single threshold model we also consider can be written as follows,

$$R_{it} = \alpha_i + \beta_{11} P_t I(P_t \leq \gamma) + \beta_{20} I(P_t > \gamma) + \beta_{21} P_t I(P_t > \gamma) + \epsilon_{it}, \tag{3.8.2}$$

where $I(\cdot)$ is an indicator function and γ the threshold parameter. All the cross-sectional indexes of regressors are the same in (3.8.2). Obviously, there exists cross-sectional dependence among regressors.

Table 3.23 summarizes the estimation results for the panel threshold model with cross-sectional dependence, model (3.8.2). The results for (3.8.2) show that the threshold estimate is 0.74, such that 2.65% of observations fall below the threshold. The coefficient on heavy rainfall is statistically significant, suggesting that the heavy rainfall has a negative effect (-1.731211) on stock returns.

TABLE 3.23: Estimates of the slope and threshold coefficients for single threshold model

	Coefficient estimate	P-value	P-value(ind.)
Threshold (Upper regime)	0.74(2.65%)[a]		
$P_t I(P_t \leq 0.74)$	-0.529989 (0.417136)[b]	0.203900	0.000000
$I(P_t > 0.74)$	1.834443 (1.259033)[b]	0.145119	0.000000
$P_t I(P_t > 0.74)$	-1.731211 (0.849989)[b]	0.041683	0.000000

[a] The percentage of variables falling into the higher regime.
[b] Standard deviation under cross-sectional dependence.

Here the P-values under cross-sectional independence are also reported in Table 3.23, which obviously lead to spurious significances of all coefficients owing to wrongly ignoring cross-sectional dependence.

In this application, it may not be sufficient to only consider single threshold indicated by Figure 3.1. Consequently, we consider the double threshold panel data model as follows,

$$\begin{aligned} R_{it} =& \alpha_i + \beta_{11} P_t I(P_t \leq \gamma_1) + \beta_{20} I(\gamma_1 < P_t \leq \gamma_2) + \beta_{21} P_t I(\gamma_1 < P_t \leq \gamma_2) \\ & + \beta_{30} I(P_t > \gamma_2) + \beta_{31} P_t I(P_t > \gamma_2) + \epsilon_{it}. \end{aligned} \tag{3.8.3}$$

which is a double threshold panel data model with two thresholds $\gamma_1 < \gamma_2$. The estimation results are given in Table 3.24. Furthermore, Figure 3.1 appears to indicate that the change points are located on around 0.5 and 1.0. We hence suppose the threshold parameters are pre-determined, with $\gamma_1 = 0.5$ and $\gamma_2 = 1.0$, to study the double threshold model with predetermined thresholds. The estimate results for this model with predetermined thresholds are reported in Table 3.25.

In order to decide which model is preferred, we have calculated the AIC values together the residual sum of squares (RSS) for all models by the proposed method. Table 3.26 shows the AIC analysis of all models by proposed method.

TABLE 3.24: Estimates of the slope and threshold coefficients for double threshold model

	Coefficient estimate	P-value
Threshold γ_1 (Lower regime)	$0.2246(84.07\%)^a$	
Threshold γ_2 (Upper regime)	$0.7422(2.65\%)^a$	
$P_t I(P_t \leq 0.2246)$	1.2248	0.2830
	(1.1409)	
$I(0.25 < P_t \leq 0.7422)$	-0.0134	0.9653
	(0.3078)	
$P_t I(0.2246 < P_t \leq 0.7422)$	-0.5466	0.3878
	(0.6329)	
$I(P_t > 0.7422)$	1.9271	0.0000
	(0.4353)	
$P_t I(P_t > 0.7422)$	-1.7312	0.0000
	(0.2854)	

[a] The percentage of variables falling into the lower regime.

TABLE 3.25: Estimates of the slope coefficients for predetermined double threshold model

	Coefficient estimate	P-value
$P_t I(P_t \leq 0.5)$	-0.495000	0.517527
	(0.764869)	
$I(0.5 < P_t \leq 1.0)$	-1.408913	0.046905
	(0.708984)	
$P_t I(0.5 < P_t \leq 1.0)$	1.949293	0.050713
	(0.997608)	
$I(P_t > 1.0)$	2.243486	0.002115
	(0.729852)	
$P_t I(P_t > 1.0)$	-1.980918	0.000010
	(0.448164)	

AIC (Akaike, 1998) for the fixed effect estimation is given by

$$AIC = nT \log(\hat{\sigma}_\epsilon^2) + 2K, \tag{3.8.4}$$

where $\hat{\sigma}_\epsilon^2 = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}^2$. However, considering the existence of cross-sectional dependence, we propose the following setting of AIC

$$AIC = nT \log(\hat{\sigma}_\epsilon(i,j)) + 2K, \tag{3.8.5}$$

where $\hat{\sigma}_\epsilon(i,j) = \frac{1}{nT} \sum_{i=1}^n \sum_{j=1}^n \sum_{t=1}^T \hat{\epsilon}_{it}\hat{\epsilon}_{jt}$. The AIC values of estimated single and double threshold model are similar. Both model suggest that only heavy rainfall has significant influence on the stock returns. Thus we conclude that compared with normal weather condition, the extreme weather, saying heavy rainfall, has a negative effect on the stock market.

TABLE 3.26: Results of AIC and RSS analysis

	AIC	RSS
Estimated linear panel model	159296.7	109250
Estimated single threshold model	158908.3	108850
Estimated double threshold model	158550.4	108490
Predetermined double threshold model	159012.7	108950

3.9 Conclusion

We have considered threshold panel time series regression with cross-sectional dependence in both the regressors and the error terms. We proposed a least squares method to estimate both slope coefficients and threshold parameters. As both cross-sectional sample size n and time length T tend to infinity simultaneously, asymptotic theories of estimated slope coefficients and threshold parameter have been derived with convergence rate root-nT and $n^{1-2\alpha_1}T^{1-2\alpha_2}$, respectively. Meanwhile, the estimation method and asymptotic theory have been expanded to multiple threshold model. We have also conducted simulated examples to evaluated the finite sample performance of our estimators by considering four data generating processes regarding to common factor and fixed effect. We have applied our method to study the effect of precipitation on the stock market in London and showed the applicability of our proposed model.

Although we considered multiple threshold model in this paper, we did not study the determination of the number of threshold parameters. The existing literature suggest to determine the threshold number by bootstrap method. However, the bootstrap method has certain limitations on cross-sectional dependence conditions. A future topic is to estimated multiple threshold panel time series via Lasso.

3.10 Appendix

Proof of Theorem 3.1. Suppose γ is given. Note that

$$\begin{aligned}\hat{\beta}(\gamma) - \beta_0(\gamma) =& (Z^*(\gamma)'Z^*(\gamma))^{-1}Z^*(\gamma)'Y^* - \beta_0(\gamma) \\ =& (Z^*(\gamma)'Z^*(\gamma))^{-1}Z^*(\gamma)'\epsilon^*(\gamma). \end{aligned} \tag{3.10.1}$$

To prove Theorem 3.1, it suffices for us to prove

$$\frac{1}{nT}Z^*(\gamma)'Z^*(\gamma) \xrightarrow{p} \Sigma_\gamma, \tag{3.10.2}$$

and

$$\frac{1}{\sqrt{nT}}Z^*(\gamma)'\epsilon^*(\gamma) \xrightarrow{d} N(0_{2K+2}, \Sigma_{\gamma,\epsilon}). \tag{3.10.3}$$

We first consider (3.10.2). By letting $T \to \infty$ (with n fixed first) and then letting $n \to \infty$, we have

$$\begin{aligned}\frac{1}{nT}Z^*(\gamma)'Z^*(\gamma) =& \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Z_{it}(\gamma) - \frac{1}{T}\sum_{t=1}^{T}Z_{it}(\gamma)\right)\left(Z_{it}(\gamma) - \frac{1}{T}\sum_{t=1}^{T}Z_{it}(\gamma)\right)' \\ =& \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Z_{it}(\gamma) - EZ_{it}(\gamma)\right)\left(Z_{it}(\gamma) - EZ_{it}(\gamma)\right)' \\ & - \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{k=1}^{T}(Z_{it}(\gamma) - EZ_{it}(\gamma))(Z_{ik}(\gamma) - EZ_{ik}(\gamma))' \\ =:& \Lambda(1) + \Lambda(2). \end{aligned} \tag{3.10.4}$$

We first consider $\Lambda(2)$,

$$\begin{aligned}\Lambda(2) =& \frac{1}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}(Z_{it}(\gamma) - EZ_{it}(\gamma))(Z_{it}(\gamma) - EZ_{it}(\gamma))' \\ & + \frac{2}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s-t=1}^{T}(Z_{it}(\gamma) - EZ_{it}(\gamma))(Z_{is}(\gamma) - EZ_{is}(\gamma))' \\ =:& \Lambda(2,1) + \Lambda(2,2) = \Lambda(2,2) + o_p(1). \end{aligned}$$

By Lemma 1 in Deo (1973), we have

$$\|E(Z_{it}(\gamma)Z_{is}(\gamma)' - EZ_{it}(\gamma)EZ_{is}(\gamma))'\| \leq C_{a1}[E\|Z_{it}(\gamma)\|^{\frac{4+\tau}{2}}]^{\frac{2}{4+\tau}}[E\|Z_{is}(\gamma)\|^{\frac{4+\tau}{2}}]^{\frac{2}{4+\tau}}\alpha(s-t)^{\frac{\tau}{4+\tau}}.$$

Let $\max_i [E\|Z_{it}(\gamma)\|^{\frac{4+\tau}{2}}]^{\frac{2}{4+\tau}} \leq C_{a2}$,

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{s-t=1}^{T}\|E(Z_{it}(\gamma)-EZ_{it}(\gamma))(Z_{is}(\gamma)-EZ_{is}(\gamma))'\| \leq nTC_{a1}C_{a2}^2\sum_{s=1}^{T}\alpha(s)^{\frac{\tau}{4+\tau}}.$$

Thus we have $\Lambda(2,2) = O_p(\frac{1}{T}) = o_p(1)$. By the law of large numbers for the α-mixing process (Lin and Lu, 1996), and we have

$$\Lambda(1) = \frac{1}{n}\sum_{i=1}^{n} E(\tilde{Z}_{it}(\gamma)\tilde{Z}_{it}(\gamma)') + o_p(1).$$

By Assumption 3.1 (iii), we have

$$\Lambda(1) = \Sigma_\gamma + o_p(1).$$

Thus we can obtain $\frac{1}{nT}Z^*(\gamma)'Z^*(\gamma) \xrightarrow{p} \Sigma_\gamma$.

We next prove (3.10.3). Note that

$$\begin{aligned}\frac{1}{\sqrt{nT}}Z^*(\gamma)'\epsilon^*(\gamma) =& \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Z_{it}(\gamma)-\frac{1}{T}\sum_{t=1}^{T}Z_{it}(\gamma)\right)\left(\epsilon_{it}(\gamma)-\frac{1}{T}\sum_{t=1}^{T}\epsilon_{it}(\gamma)\right)\\ =& \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Z_{it}(\gamma)-EZ_{it}(\gamma)\right)\left(\epsilon_{it}(\gamma)-E\epsilon_{it}(\gamma)\right)\\ &-\frac{1}{T\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{k=1}^{T}[Z_{it}(\gamma)-EZ_{it}(\gamma)][\epsilon_{ik}(\gamma)-E\epsilon_{ik}(\gamma)]\\ =:&\Gamma(1)+\Gamma(2). \end{aligned} \tag{3.10.5}$$

Let $\tilde{\theta}_{it,k}(\gamma) = \tilde{Z}_{it}(\gamma)\tilde{\epsilon}_{ik}(\gamma)$. Denote

$$M_\theta = \max_i (E[\|\tilde{\theta}_{it,k}(\gamma)\|^{\frac{4+\tau}{2}})^{\frac{2}{4+\tau}} \text{ and } \max_i E[\|\tilde{\theta}_{it,k}(\gamma)\|^2] \leq C_\theta.$$

Thus Assumptions 2.1 (i) and 2.2 in Gao et al. (2008) hold. Let C_0 denote a $2K+2$ constant vector with $\|C_0\| = 1$. By Theorem 2.1 in Gao et al. (2008),

$\sum_{i=1}^{n} E\tilde{\theta}_{it,k}(\gamma) = 0$, we have

$$E\left(\sum_{t=1}^{T}\sum_{k=1}^{T}\sum_{i=1}^{n} C_0'\tilde{\theta}_{it,k}(\gamma)\right)^2 \leq nT^2M_\theta^2\sum_{t=1}^{\infty}\alpha(t)^{\frac{\tau}{4+\tau}} + nT^2C_\theta.$$

Thus we have $\Gamma(2) = O_p(\frac{1}{\sqrt{T}}) = o_p(1)$.

$$\begin{aligned} Var(\Gamma_1) = E\Gamma(1)\Gamma(1)' =& E\left(\frac{1}{\sqrt{nT}}\tilde{Z}(\gamma)'\tilde{\epsilon}(\gamma)\right)\left(\frac{1}{\sqrt{nT}}\tilde{Z}(\gamma)'\tilde{\epsilon}(\gamma)\right)' \\ =& \frac{1}{nT}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T}\sum_{k=1}^{T} E\tilde{Z}_{it}(\gamma)\tilde{Z}_{jk}(\gamma)'\tilde{\epsilon}_{it}(\gamma)\tilde{\epsilon}_{jk}(\gamma). \end{aligned}$$

By Assumption 3.1 (iii), it is easy to check

$$Var(\Gamma_1) = \Sigma_{\gamma,\epsilon} + o_p(1).$$

Then $\Gamma(1) \xrightarrow{d} N(0_{2K+2}, \Sigma_{\gamma,\epsilon})$ by the α-mixing central limit theorem (e.g. Fan and Yao (2003)).

Lemma A. 1. *If $\{X_{it}\}$ is strictly stationary and ergodic, $E|\phi(X_{it})| < \infty$, and X_{it} has a continuous distribution, then*

$$\sup_{\gamma\in R}\left|\frac{1}{T}\sum_{t=1}^{T}\phi(X_{it})I(q_{it}\leq\gamma) - E(\phi(X_{it})I(q_{it}\leq\gamma))\right| \to 0, \ a.s.$$

Proof of Lemma A.1. This proof is similar to that of Lemma 1 Hansen (1996). Essentially, this result still holds under cross-sectional dependence conditions. Differently. the proof of Lemma 1 Hansen (1996) uses the law of large numbers for independent variables. Here, we use the law of large numbers for α-mixing processes.

Proposition 3.1. *Suppose that Assumption 3.1 to Assumption 3.3 hold. As $n \to \infty$ and $T \to \infty$, we have $\hat{\gamma} \xrightarrow{p} \gamma_0$.*

Proof of Proposition 3.1 When $\gamma = \gamma_0$, model (3.2.2) is equal to

$$Y_{it} = \mu_i + \beta_1{}'Z_{it}I(q_{it} \leq \gamma_0) + \beta_2{}'Z_{it}I(q_{it} > \gamma_0) + \epsilon_{it}. \tag{3.10.6}$$

When $\gamma \neq \gamma_0$, model (3.2.2) can be rewritten as

$$\begin{aligned} Y_{it} &= \mu_i + \beta_1{}'Z_{it}I(q_{it} \leq \gamma) + \beta_2{}'Z_{it}I(q_{it} > \gamma) \\ &\quad - \beta_1{}'Z_{it}[I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_0)] - \beta_2{}'Z_{it}[I(q_{it} > \gamma) - I(q_{it} > \gamma_0)] + \epsilon_{it} \\ &= \mu_i + \beta_1{}'Z_{it}I(q_{it} \leq \gamma) + \beta_2{}'Z_{it}I(q_{it} > \gamma) \\ &\quad + (\beta_2 - \beta_1)'Z_{it}[I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_0)] + \epsilon_{it} \\ &= \mu_i + \beta_1{}'Z_{it}I(q_{it} \leq \gamma) + \beta_2{}'Z_{it}I(q_{it} > \gamma) \\ &\quad + n^{-\alpha_1}T^{-\alpha_2}C'Z_{it}[I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_0)] + \epsilon_{it} \end{aligned} \tag{3.10.7}$$

The within-group estimation can be applied to (3.10.7) to eliminate the fixed effect to yield

$$Y_{it}^* = \beta_1{}'Z_{it1}^*(\gamma) + \beta_2{}'Z_{it2}^*(\gamma) + n^{-\alpha_1}T^{-\alpha_2}C'Z_{it1}^*(\gamma) - n^{-\alpha_1}T^{-\alpha_2}C'Z_{it1}^*(\gamma_0) + \epsilon_{it}^*. \tag{3.10.8}$$

Thus we have the regression residuals for γ as follows,

$$\hat{\epsilon}_{it}^*(\gamma) = n^{-\alpha_1}T^{-\alpha_2}C'Z_{it1}^*(\gamma) - n^{-\alpha_1}T^{-\alpha_2}C'Z_{it1}^*(\gamma_0) + \epsilon_{it}^*. \tag{3.10.9}$$

Therefore, we have

$$\begin{aligned} S(\gamma) - S(\gamma_0) &= \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}_{it}^*(\gamma)^2 - \sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\epsilon}_{it}^*(\gamma_0)^2 \\ &= \sum_{i=1}^{n}\sum_{t=1}^{T}(n^{-\alpha_1}T^{-\alpha_2}C'Z_{it1}^*(\gamma) - n^{-\alpha_1}T^{-\alpha_2}C'Z_{it1}^*(\gamma_0) + \epsilon_{it}^*)^2 - \sum_{i=1}^{n}\sum_{t=1}^{T}\epsilon_{it}^{*2} \\ &= n^{-2\alpha_1}T^{-2\alpha_2}C'\sum_{i=1}^{n}\sum_{t=1}^{T}(Z_{it1}^*(\gamma) - Z_{it1}^*(\gamma_0))(Z_{it1}^*(\gamma) - Z_{it1}^*(\gamma_0))'C \\ &\quad + 2n^{-\alpha_1}T^{-\alpha_2}C'\sum_{i=1}^{n}\sum_{t=1}^{T}(Z_{it1}^*(\gamma) - Z_{it1}^*(\gamma_0))\epsilon_{it}. \end{aligned} \tag{3.10.10}$$

Note that

$$\frac{n^{2\alpha_1}T^{2\alpha_2}}{nT}(S(\gamma) - S(\gamma_0)) = \frac{1}{nT}C'\sum_{i=1}^{n}\sum_{t=1}^{T}(Z_{it1}^*(\gamma) - Z_{it1}^*(\gamma_0))(Z_{it1}^*(\gamma) - Z_{it1}^*(\gamma_0))'C$$

$$+\frac{2n^{\alpha_1}T^{\alpha_2}}{nT}C'\sum_{i=1}^{n}\sum_{t=1}^{T}(Z^*_{it1}(\gamma)-Z^*_{it1}(\gamma_0))\epsilon_{it}$$
$$=: \Phi(1)+\Phi(2). \tag{3.10.11}$$

According to the central limit theorem for a α-mixing sequence (e.g. Fan and Yao (2003)), it is similar to the proof of Theorem 3.1,

$$\sum_{t=1}^{T}\sum_{i=1}^{n}Z^*_{it}(\gamma)\epsilon_{it}=O_p(\sqrt{nT})\text{ and }\sum_{t=1}^{T}\sum_{i=1}^{n}Z^*_{it}(\gamma_0)\epsilon_{it}=O_p(\sqrt{nT}),$$

For $\alpha_1,\alpha_2\in[0,\frac{1}{2})$, by Lemma 3.4, we have uniformly over $\gamma\in\Gamma$,

$$\Phi(2)=o_p(1).$$

For the part of $\Phi(1)$,

$$\begin{aligned}\Phi(1)=&\frac{1}{nT}C'\sum_{i=1}^{n}\sum_{t=1}^{T}Z^*_{it1}(\gamma)Z^*_{it1}(\gamma)'C-\frac{1}{nT}C'\sum_{i=1}^{n}\sum_{t=1}^{T}Z^*_{it1}(\gamma)Z^*_{it1}(\gamma_0)'C\\&-\frac{1}{nT}C'\sum_{i=1}^{n}\sum_{t=1}^{T}Z^*_{it1}(\gamma_0)Z^*_{it1}(\gamma)'C+\frac{1}{nT}C'\sum_{i=1}^{n}\sum_{t=1}^{T}Z^*_{it1}(\gamma_0)Z^*_{it1}(\gamma_0)'C\\=:&\Phi(1,1)+...+\Phi(1,4).\end{aligned}$$

We first consider $\Phi(1,1)$. Let $\tilde{Z}_{it1}(\gamma)=E(Z_{it1}(\gamma)-EZ_{it1}(\gamma))$,

$$\Phi(1,1)=C'\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{Z}_{it1}(\gamma)\tilde{Z}_{it1}(\gamma)'C+o_p(1).$$

By Lemma A. 1, We have uniformly over $\gamma\in\Gamma$, as $T\to\infty$,

$$\Phi(1,1)\to C'\frac{1}{n}\sum_{i=1}^{n}E\tilde{Z}_{it1}(\gamma)\tilde{Z}_{it1}(\gamma)'C,\ a.s.$$

Similarly, uniformly over $\gamma\in\Gamma$, as $T\to\infty$

$$\begin{aligned}\Phi(1)\to&C'\frac{1}{n}\sum_{i=1}^{n}E(Z_{it1}(\gamma)-EZ_{it1}(\gamma))(Z_{it1}(\gamma)-EZ_{it1}(\gamma))'C\\&-C'\frac{1}{n}\sum_{i=1}^{n}E(Z_{it1}(\gamma)-EZ_{it1}(\gamma))(Z_{it1}(\gamma_0)-EZ_{it1}(\gamma_0))'C\\&-C'\frac{1}{n}\sum_{i=1}^{n}E(Z_{it1}(\gamma_0)-EZ_{it1}(\gamma_0))(Z_{it1}(\gamma)-EZ_{it1}(\gamma))'C\end{aligned}$$

$$
\begin{aligned}
&+C'\frac{1}{n}\sum_{i=1}^{n}E(Z_{it1}(\gamma_0)-EZ_{it1}(\gamma_0)))(Z_{it1}(\gamma_0)-EZ_{it1}(\gamma_0))C\\
=&C'\frac{1}{n}\sum_{i=1}^{n}E[(Z_{it1}(\gamma)-EZ_{it1}(\gamma))-(Z_{it1}(\gamma_0)-EZ_{it1}(\gamma_0))]\\
&E[(Z_{it1}(\gamma)-EZ_{it1}(\gamma))-(Z_{it1}(\gamma_0)-EZ_{it1}(\gamma_0))]C,\ a.s. \qquad (3.10.12)
\end{aligned}
$$

Considering the case $\gamma \in [\gamma_0, \overline{\gamma}]$, by (3.10.12) we have

$$
\begin{aligned}
E\Phi(1) =&C'\frac{1}{n}\sum_{i=1}^{n}E[Z_{it1}I(\gamma_0<q_{it}\leq\gamma)-EZ_{it1}I(\gamma_0<q_{it}\leq\gamma)]\\
&[Z_{it1}I(\gamma_0<q_{it}\leq\gamma)-EZ_{it1}I(\gamma_0<q_{it}\leq\gamma]'C. \qquad (3.10.13)
\end{aligned}
$$

So $E\Phi(1) \geq 0$ and $Var(Z_{it1}) > 0$, when $E\Phi(1) = 0$ if and only if $\gamma = \gamma_0$. Symmetrically, we can show that $E\Phi(1) \geq 0$ over $\gamma \in [\underline{\gamma}, \gamma_0]$. Thus $\frac{n^{2\alpha_1}T^{2\alpha_2}}{nT}E(S(\gamma) - S(\gamma_0))$ has unique minimizer γ_0. Since $\hat{\gamma}$ minimizes $S(\gamma) - S(\gamma_0)$, it follows that $\hat{\gamma} \xrightarrow{p} \gamma_0$.

Proposition 3.2. *Suppose that Assumption 3.1 to Assumption 3.3 hold. As $n \to \infty$ and $T \to \infty$, we have $\hat{\beta} - \beta_0 = o_p(n^{-\alpha_1}T^{-\alpha_2})$.*

Proof of Proposition 3.2 By definition,

$$
\hat{\beta} = (Z^*(\hat{\gamma})'Z^*(\hat{\gamma}))^{-1}Z^*(\hat{\gamma})'Y^* \qquad (3.10.14)
$$

Let the vector form of $Z^*_{it}I(q_{it} \leq \gamma)$ be $Z^{*1}(\gamma)$. Substituting $Y^* = Z^*(\gamma_0)\beta_0 + \epsilon^*$ into the above equation yields,

$$
\begin{aligned}
&n^{\alpha_1}T^{\alpha_2}(\hat{\beta}-\beta_0)\\
=&\left(\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma})\right)^{-1}\left(\frac{1}{nT}Z^*(\hat{\gamma})'(Z^*(\gamma_0)-Z^*(\hat{\gamma}))\beta_0 n^{\alpha_1}T^{\alpha_2}+\frac{n^{\alpha_1}T^{\alpha_2}}{nT}Z^*(\hat{\gamma})'\epsilon^*\right)\\
=&\left(\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma})\right)^{-1}\left(\frac{1}{nT}Z^*(\hat{\gamma})'(Z^*(\gamma_0)-Z^*(\hat{\gamma}))\beta_0 n^{\alpha_1}T^{\alpha_2}\right)\\
&+\left(\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma})\right)^{-1}\left(\frac{n^{\alpha_1}T^{\alpha_2}}{nT}Z^*(\hat{\gamma})'\epsilon\right)\\
=:&\Theta(1)+\Theta(2).
\end{aligned}
$$

We first consider $\Theta(2)$. By α-mixing central limit theorem (e.g.Fan and Yao (2003)), for $\alpha_1, \alpha_2 \in [0, 1/2)$, $\frac{n^{\alpha_1} T^{\alpha_2}}{nT} Z^*(\hat{\gamma})'\epsilon = o_p(1)$. Therefor $\Theta(2) = o_p(1)$.

By $\beta_2 - \beta_1 = n^{-\alpha_1} T^{-\alpha_2} C$,

$$\begin{aligned}(Z^*(\gamma_0) - Z^*(\hat{\gamma}))\beta_0 &= \sum_{i=1}^{n}\sum_{t=1}^{T}[\beta_1'(Z^*_{it1}(\gamma_0) - Z^*_{it1}(\hat{\gamma})) + \beta_2'(Z^*_{it2}(\gamma_0) - Z^*_{it2}(\hat{\gamma}))] \\ &= \sum_{i=1}^{n}\sum_{t=1}^{T}(\beta_1 - \beta_2)'(Z^*_{it1}(\gamma_0) - Z^*_{it1}(\hat{\gamma})) \\ &= n^{-\alpha_1} T^{-\alpha_2} \sum_{i=1}^{n}\sum_{t=1}^{T} C'(Z^*_{it1}(\hat{\gamma}) - Z^*_{it1}(\gamma_0)).\end{aligned}$$

Therefore, $\Theta(2) = \left(\frac{1}{nT} Z^*(\hat{\gamma})' Z^*(\hat{\gamma})\right)^{-1} \left(\frac{1}{nT} Z^*(\hat{\gamma})'(Z^{*1}(\hat{\gamma}) - Z^{*1}(\gamma_0))C\right)$.

By Lemma A.1 and the consistency of $\hat{\gamma} \xrightarrow{p} \gamma_0$, by setting $T \to \infty$ first, and then $n \to \infty$, we have

$$\frac{1}{nT} Z^*(\hat{\gamma})'(Z^{*1}(\gamma_0) - Z^{*1}(\gamma_0))C \to \frac{1}{n}\sum_{i=1}^{n} E(\tilde{Z}_{it}(\gamma_0)(\tilde{Z}_{it1}(\gamma_0) - \tilde{Z}_{it1}(\gamma_0))')C = 0. \tag{3.10.15}$$

Hence we can get $n^{\alpha_1} T^{\alpha_2}(\hat{\beta} - \beta_0) = o_p(1)$.

Lemma 3.1. *Let* $h_{it}(\gamma_1, \gamma_2) = \|Z_{it}\epsilon_{it}\| |I(q_{it} \le \gamma_2) - I(q_{it} \le \gamma_1)|$ *and* $k_{it}(\gamma_1, \gamma_2) = \|Z_{it}\| |I(q_{it} \le \gamma_2) - I(q_{it} \le \gamma_1)|$*, there is a* $A_1 < \infty$ *such that for* $\underline{\gamma} \le \gamma_1 \le \gamma_2 \le \overline{\gamma}$*, and* $r \le 4$*,*

$$\begin{aligned}\max_i E[h_{it}(\gamma_1, \gamma_2)]^r &\le A_1|\gamma_2 - \gamma_1|, \\ \max_i E[k_{it}(\gamma_1, \gamma_2)]^r &\le A_1|\gamma_2 - \gamma_1|.\end{aligned}$$

Proof of Lemma 3.1 Under Assumption 3.3 (ii),

$$\begin{aligned}\frac{d}{d\gamma} E[\|Z_{it}\epsilon_{it}\|^r I(q_{it} \le \gamma)] &= E[\|Z_{it}\epsilon_{it}\|^r | q_{it} = \gamma)] f_i(\gamma) \\ &\le [E(\|Z_{it}\epsilon_{it}\|^4 | q_{it} = \gamma)]^{\frac{r}{4}} f_i(\gamma)\end{aligned}$$

$$\leq A\overline{f}.$$

Since $I(q_{it} \leq \gamma_2) - I(q_{it} \leq \gamma_1)$ equals either zero or one, this implies with $A_1 = A\overline{f}$,

$$\max_i E[h_{it}(\gamma_1, \gamma_2)]^r \leq A_1|\gamma_2 - \gamma_1|.$$

Analogously, we have $\max_i E[k_{it}(\gamma_1, \gamma_2)]^r \leq A_1|\gamma_2 - \gamma_1|$.

Lemma 3.2. *Let C_0 be a nonrandom $(K+1) \times 1$ vector with $\|C_0\| = 1$, $h^*_{it}(\gamma_1, \gamma_2) = C_0' Z_{it}\epsilon_{it}[I(q_{it} \leq \gamma_2) - I(q_{it} \leq \gamma_1)]$ and $k^*_{it}(\gamma_1, \gamma_2) = C_0' Z_{it}[I(q_{it} \leq \gamma_2) - I(q_{it} \leq \gamma_1)]$, there is a $A_1 < \infty$ such that for There is a $A_2 < \infty$ such that for all $\underline{\gamma} \leq \gamma_1 \leq \gamma_2 \leq \overline{\gamma}$,*

$$E\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}(h^{*2}_{it}(\gamma_1, \gamma_2) - Eh^{*2}_{it}(\gamma_1, \gamma_2))\right|^2 \leq A_2|\gamma_2 - \gamma_1|,$$
$$E\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}(k^{*2}_{it}(\gamma_1, \gamma_2) - Ek^{*2}_{it}(\gamma_1, \gamma_2))\right|^2 \leq A_2|\gamma_2 - \gamma_1|.$$

Proof of Lemma 3.2 Let $\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2) = h^{*2}_{it}(\gamma_1, \gamma_2) - Eh^{*2}_{it}(\gamma_1, \gamma_2))$. Define $\rho_{i,tt_1} = \frac{E\tilde{h}^{*2}_{it}(\gamma_1,\gamma_2)\tilde{h}^{*2}_{it_1}(\gamma_1,\gamma_2)}{\sqrt{E[\tilde{h}^{*2}_{it}(\gamma_1,\gamma_2)]^2}\sqrt{E[\tilde{h}^{*2}_{it_1}(\gamma_1,\gamma_2)]^2}}$, $\rho_{t,ii_1} = \frac{E\tilde{h}^{*2}_{it}(\gamma_1,\gamma_2)\tilde{h}^{*2}_{i_1t}(\gamma_1,\gamma_2)}{\sqrt{E[\tilde{h}^{*2}_{it}(\gamma_1,\gamma_2)]^2}\sqrt{E[\tilde{h}^{*2}_{i_1t}(\gamma_1,\gamma_2)]^2}}$ and $\rho_{it,i_1t_1} = \frac{E\tilde{h}^{*2}_{it}(\gamma_1,\gamma_2)\tilde{h}^{*2}_{i_1t_1}(\gamma_1,\gamma_2)}{\sqrt{E[\tilde{h}^{*2}_{it}(\gamma_1,\gamma_2)]^2}\sqrt{E[\tilde{h}^{*2}_{i_1t_1}(\gamma_1,\gamma_2)]^2}}$,

$$\begin{aligned}
&E\left|\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)\right|^2 \\
=&\frac{1}{n}\sum_{i=1}^{n}E[\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)]^2 + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t\neq t_1}^{T}E\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)\tilde{h}^{*2}_{it_1}(\gamma_1, \gamma_2) \\
&+\frac{1}{n}\sum_{i\neq i_1}^{n}E\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)\tilde{h}^{*2}_{i_1t}(\gamma_1, \gamma_2) + \frac{1}{nT}\sum_{i\neq i_1}^{n}\sum_{t\neq t_1}^{T}E\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)\tilde{h}^{*2}_{i_1t_1}(\gamma_1, \gamma_2) + o(1) \\
=&\frac{1}{n}\sum_{i=1}^{n}E[\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)]^2 + \frac{1}{nT}\sum_{i=1}^{n}\sum_{t\neq t_1}^{T}\rho_{i,tt_1}\sqrt{E[\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)]^2}\sqrt{E[\tilde{h}^{*2}_{it_1}(\gamma_1, \gamma_2)]^2} \\
&+\frac{1}{n}\sum_{i\neq i_1}^{n}\rho_{ii_1,t}\sqrt{E[\tilde{h}^{*2}_{it}(\gamma_1, \gamma_2)]^2}\sqrt{E[\tilde{h}^{*2}_{i_1t}(\gamma_1, \gamma_2)]^2}
\end{aligned}$$

$$+\frac{1}{nT}\sum_{i\neq i_1}^{n}\sum_{t\neq t_1}^{T}\rho_{it,i_1t_1}\sqrt{E[\tilde{h}_{it}^{*2}(\gamma_1,\gamma_2)]^2}\sqrt{E[\tilde{h}_{i_1t_1}^{*2}(\gamma_1,\gamma_2)]^2}$$

By Lemma 1 in Deo (1973), we have

$$\begin{aligned}&|E(h_{it}^{*2}(\gamma_1,\gamma_2)h_{it_1}^{*2}(\gamma_1,\gamma_2))-Eh_{it}^{*2}(\gamma_1,\gamma_2)Eh_{it_1}^{*2}(\gamma_1,\gamma_2)|\\ \leq& C_{b1}[Eh_{it}^{*4+\tau}(\gamma_1,\gamma_2)]^{\frac{2}{4+\tau}}[Eh_{it_1}^{*4+\tau}(\gamma_1,\gamma_2)]^{\frac{2}{4+\tau}}\alpha(t_1-t)^{\frac{2}{4+\tau}}\end{aligned}$$

Let $\max_i[Eh_{it}^{*4+\tau}(\gamma_1,\gamma_2)]^{\frac{2}{4+\tau}}\leq C_{b2}$ and $\min_i[Eh_{it}^{*4}(\gamma_1,\gamma_2)]\geq C_{b3}$,

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{t_1-t=1}^{T}|E(h_{it}^{*2}(\gamma_1,\gamma_2)h_{it_1}^{*2}(\gamma_1,\gamma_2))-Eh_{it}^{*2}(\gamma_1,\gamma_2)Eh_{it_1}^{*2}(\gamma_1,\gamma_2)|\leq nTC_{b1}C_{b2}^2\sum_{t_1=1}^{T}\alpha(t_1)^{\frac{\tau}{4+\tau}}.$$

Thus we have $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t\neq t_1}^{T}|\rho_{i,tt_1}|\leq\frac{C_{b1}C_{b2}^2\sum_{t_1=1}^{T}\alpha(t_1)^{\frac{\tau}{4+\tau}}}{C_{b3}}$. Define $C_\rho^0=\frac{C_{b1}C_{b2}^2\sum_{t_1=1}^{T}\alpha(t_1)^{\frac{\tau}{4+\tau}}}{C_{b3}}$, by Assumption 3.1 (iv) and Lemma 3.1, $\max_i E[\tilde{h}_{it_1}^{*2}(\gamma_1,\gamma_2)]^2\leq\max_i E[h_{it_1}^{*2}(\gamma_1,\gamma_2)]^2\leq A_1|\gamma_2-\gamma_1|$, we have

$$\begin{aligned}E\left|\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{h}_{it}^{2}(\gamma_1,\gamma_2)\right|^2\leq&\frac{1}{n}\sum_{i=1}^{n}A_1|\gamma_2-\gamma_1|+\frac{1}{nT}\sum_{i=1}^{n}\sum_{t\neq t_1}^{T}|\rho_{i,tt_1}|A_1|\gamma_2-\gamma_1|\\&+\frac{1}{n}\sum_{i\neq i_1}^{n}|\rho_{ii_1,t}|A_1|\gamma_2-\gamma_1|+\frac{1}{nT}\sum_{i\neq i_1}^{n}\sum_{t\neq t_1}^{T}|\rho_{it,i_1t_1}|A_1|\gamma_2-\gamma_1|\\ \leq&(1+C_\rho^0+C_\rho^1+C_\rho^2)A_1|\gamma_2-\gamma_1|.\end{aligned}$$

Therefore, there exist a constant A_2 such that

$$E\left|\sum_{t=1}^{T}\sum_{i=1}^{n}\tilde{h}_{it}^{*2}(\gamma_1,\gamma_2)\right|^2\leq A_2|\gamma_2-\gamma_1|.$$

Analogously, we can prove the second result in the lemma.

Lemma 3.3. *Let* $J_{nT}(\gamma)=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}Z_{it}\epsilon_{it}I(q_{it}\leq\gamma)$*, there are constants* K_1 *and* K_2 *such that for all* γ_1, $\varepsilon>0$, $\eta>0$ *and* $\delta\geq\frac{1}{T}$*, if* $\sqrt{T}\geq\frac{K_2}{\eta}$*, then*

$$P\left(\sup_{\gamma_1\leq\gamma\leq\gamma_1+\delta}\|J_{nT}(\gamma)-J_{nT}(\gamma_1)\|>\eta\right)\leq\frac{K_1\delta^2}{\eta^4}.$$

Proof of Lemma 3.3 Let m be an integer satisfying $nT\delta/2 \leq m \leq nT\delta$, which is possible since $nT\delta \geq 1$. Set $\delta_m = \delta/m$. For $k = 1, ..., m+1$, set $\gamma_k = \gamma_1 + \delta_m(k-1)$, $h^*_{it,k} = h^*_{it}(\gamma_k, \gamma_{k+1})$ and $h^*_{it,jk} = h^*_{it}(\gamma_j, \gamma_k)$. Letting $H_{nT,k} = \frac{1}{nT}\sum_{i=1}^n \sum_{t=1}^T h^*_{it,k}$, observe that for $\gamma_k \leq \gamma \leq \gamma_{k+1}$,

$$\begin{aligned}\|J_{nT}(\gamma) - J_{nT}(\gamma_k)\| =& \left\| \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T Z_{it}\epsilon_{it}[I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_k)] \right\| \\ \leq& \frac{1}{\sqrt{nT}} \sum_{i=1}^n \sum_{t=1}^T \|Z_{it}\epsilon_{it}\| |I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_k))| \\ \leq& \sqrt{nT} H_{nT,k} \leq \sqrt{nT}|H_{nT,k} - EH_{nT,k}| + \sqrt{nT}EH_{nT,k}.\end{aligned}$$

Thus

$$\begin{aligned}\sup_{\gamma_1 \leq \gamma \leq \gamma_1+\delta} \|J_{nT}(\gamma) - J_{nT}(\gamma_1)\| \leq& \max_{2\leq k \leq m+1} \|J_{nT}(\gamma_k) - J_{nT}(\gamma_1)\| \\ &+ \max_{1 \leq k \leq m} \sqrt{T}|H_{nT,k} - EH_{nT,k}| + \max_{1\leq k\leq m} \sqrt{T}EH_{nT,k}.\end{aligned} \tag{3.10.16}$$

Let C_0 be a constant vector with $\|C_0\| = 1$, $h^*_{t,jk} = \frac{1}{\sqrt{n}}\sum_{i=1}^n C_0' Z_{it}\epsilon_{it}[I(q_{it} \leq \gamma_k) - I(q_{it} \leq \gamma_j)$. For any $1 \leq j < k \leq m+1$, by Burkholder's inequality (see e.g. Hall and Heyde (2014)),

$$\begin{aligned}E|C_0'(J_{nT}(\gamma_k) - J_{nT}(\gamma_j))|^4 =& E\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^n\sum_{t=1}^T C_0'Z_{it}\epsilon_{it}[I(q_{it}\leq\gamma_k) - I(q_{it}\leq\gamma_j)]\right|^4 \\ \leq& K'E\left|\frac{1}{T}\sum_{t=1}^T\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n C_0'Z_{it}\epsilon_{it}[I(q_{it}\leq\gamma_k) - I(q_{it}\leq\gamma_j)]\right|^2\right|^2 \\ \leq& K'E\left|\frac{1}{T}\sum_{t=1}^T h^{*2}_{t,jk}\right|^2 \\ =& K'E\left|\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T h^{*2}_{it,jk} + \frac{1}{nT}\sum_{i\neq i_1}^n\sum_{t=1}^T h^*_{it,jk}h^*_{i_1t,jk}\right|^2 \\ \leq& 2K'E\left|\frac{1}{nT}\sum_{i=1}^n\sum_{t=1}^T h^{*2}_{it,jk}\right|^2 + 2K'E\left|\frac{1}{nT}\sum_{i\neq i_1}^n\sum_{t=1}^T h^*_{it,jk}h^*_{i_1t,jk}\right|^2 \\ =&: \Pi(1) + \Pi(2).\end{aligned}$$

By Lemma 3.1, $\max_i Eh^r_{it,k} \leq A_1\delta_m$, $\max_i Eh^r_{it,jk} \leq A_1|k-j|\delta_m$ for $r \leq 4$. By

Lemma 3.2, $E\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}(h_{it,jk}^{*2}-Eh_{it,jk}^{*2})\right|^2 \leq A_2(k-j)\delta_m$. For $\Pi(1)$, by $\delta_m \geq \frac{1}{nT}$, we have

$$\begin{aligned}\Pi(1) =&2K'E\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}[(h_{it,jk}^{*2}-Eh_{it,jk}^{*2})+Eh_{it,jk}^{*2}]\right|^2\\ =&2K'E\left[\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(h_{it,jk}^{*2}-Eh_{it,jk}^{*2})\right|^2+\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}Eh_{it,jk}^{*2}\right|^2\right]\\ \leq&2K'\left[\frac{A_2(k-j)\delta_m}{nT}+A_1^2(k-j)^2\delta_m^2\right] \leq 2K'(A_2+A_1^2)(k-j)^2\delta_m^2.\end{aligned} \tag{3.10.17}$$

We next consider $\Pi(2)$,

$$\begin{aligned}\Pi(2) =&\frac{2K'}{n^2T^2}\sum_{t=1}^{T}E\left[\sum_{i\neq i_1}^{n}h_{it,jk}^{*}h_{i_1t,jk}^{*}\right]^2+\frac{2K'}{n^2T^2}\sum_{t\neq t_1}^{T}E\left[\sum_{i\neq i_1}^{n}h_{it,jk}^{*}h_{i_1t,jk}^{*}\right]\left[\sum_{i\neq i_1}^{n}h_{it_1,jk}^{*}h_{i_1t_1,jk}^{*}\right]\\ =&\frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1}^{n}Eh_{it,jk}^{*2}h_{i_1t,jk}^{*2}+\frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1\neq i_2}^{n}Eh_{it,jk}^{*2}h_{i_1t,jk}^{*}h_{i_2t,jk}^{*}\\ &+\frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1\neq i_2\neq i_3}^{n}Eh_{it,jk}^{*}h_{i_1t,jk}^{*}h_{i_2t,jk}^{*}h_{i_3t,jk}^{*}\\ &+\frac{2K'}{n^2T^2}\sum_{t\neq t_1}^{T}\sum_{i\neq i_1\neq i_2\neq i_3}^{n}Eh_{it,jk}^{*}h_{i_1t,jk}^{*}h_{i_2t_1,jk}^{*}h_{i_3t_1,jk}^{*}\\ =&:\Pi(2,1)+\Pi(2,2)+\Pi(2,3)+\Pi(2,4).\end{aligned}$$

For $\Pi(2,1)$, by Assumption 3.1 (iv), Lemma 3.1, and $\delta_m > \frac{1}{nT}$, we have

$$\begin{aligned}\Pi(2,1) =&\frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1}^{n}E(h_{it,jk}^{*2}-Eh_{it,jk}^{*2})(h_{i_1t,jk}^{*2}-Eh_{i_1t,jk}^{*2})+\frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1}^{n}Eh_{it,jk}^{*2}Eh_{i_1t,jk}^{*2}\\ \leq&\frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1}^{n}\rho_{t,ii_1}\sqrt{E(h_{it,jk}^{*2}-Eh_{it,jk}^{*2})^2}\sqrt{E(h_{i_1t,jk}^{*2}-Eh_{i_1t,jk}^{*2})^2}\\ &+\frac{2K'}{n^2T^2}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}Eh_{it,jk}^{*2}\right)\left(\sum_{t=1}^{T}\sum_{i\neq i_1}^{n}Eh_{i_1t,jk}^{*2}\right)\\ \leq&\frac{2K'C_\rho^1A_1(k-j)\delta_m}{nT}+2K'A_1^2(k-j)^2\delta_m^2\\ \leq&(2K'C_\rho^1A_1+2K'A_1^2)(k-j)^2\delta_m^2.\end{aligned} \tag{3.10.18}$$

For $\Pi(2,2)$, by Assumption 3.1 (iv) and Lemma 3.2, we have

$$\begin{aligned}
\Pi(2,2) =& \frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1\neq i_2}^{n} E(h^{*2}_{it,jk}-Eh^{*2}_{it,jk})h^{*}_{i_1t,jk}h^{*}_{i_2t,jk} + \frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1\neq i_2}^{n} Eh^{*2}_{it,jk}Eh^{*}_{i_1t,jk}h^{*}_{i_2t,jk} \\
\leq& \frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1\neq i_2}^{n} \frac{E(h^{*2}_{it,jk}-Eh^{*2}_{it,jk})h^{*}_{i_1t,jk}h^{*}_{i_2t,jk}}{\sqrt{E(h^{2}_{it,jk}-Eh^{*2}_{it,jk})^2}\sqrt{Eh^{*2}_{i_1t,jk}}\sqrt{Eh^{*2}_{i_2t,jk}}} \\
&\cdot\sqrt{E(h^{*2}_{it,jk}-Eh^{*2}_{it,jk})^2}\sqrt{Eh^{*2}_{i_1t,jk}}\sqrt{Eh^{*2}_{i_2t,jk}} \\
&+\frac{2K'}{n^2T^2}\left(\sum_{t=1}^{T}\sum_{i=1}^{n}Eh^{*2}_{it,jk}\right)\left(\sum_{t=1}^{T}\sum_{i_1\neq i_2}^{n}Eh^{*}_{i_1t,jk}h^{*}_{i_2t,jk}\right) \\
\leq& \frac{2K'C_\rho^3(\sqrt{A_1(k-j)\delta_m})^3}{\sqrt{n}T}+2K'C_\rho^2A_1^2(k-j)^2\delta_m^2 \\
\leq& \left(\frac{2K'C_\rho^3(\sqrt{A_1})^3}{\sqrt{T}}+2K'C_\rho^2A_1^2\right)(k-j)^2\delta_m^2. \qquad (3.10.19)
\end{aligned}$$

We next consider $\Pi(2,3)$. Similarly, by Assumption 3.1 (iv),

$$\begin{aligned}
\Pi(2,3) \leq& \frac{2K'}{n^2T^2}\sum_{t=1}^{T}\sum_{i\neq i_1\neq i_2\neq i_3}^{n} \frac{Eh^{*}_{it,jk}h^{*}_{i_1t,jk}h^{*}_{i_2t,jk}h^{*}_{i_3t,jk}}{\sqrt{Eh^{*2}_{it,jk}}\sqrt{Eh^{*2}_{i_1t,jk}}\sqrt{Eh^{*2}_{i_2t,jk}}\sqrt{Eh^{*2}_{i_3t,jk}}} \\
&\cdot\sqrt{Eh^{*2}_{it,jk}}\sqrt{Eh^{*2}_{i_1t,jk}}\sqrt{Eh^{*2}_{i_2t,jk}}\sqrt{Eh^{*2}_{i_3t,jk}} \\
\leq& \frac{2K'C_\rho^4A_1^2(k-j)^2\delta_m^2}{T}. \qquad (3.10.20)
\end{aligned}$$

Finally, we consider $\Pi(2,4)$. Similarly,

$$\Pi(2,4) \leq \frac{2K'C_\rho^5A_1^2(k-j)^2\delta_m^2}{T}. \qquad (3.10.21)$$

Therefore, by (3.10.24), (3.10.19), (3.10.20) and (3.10.21), we have

$$\Pi(2) \leq \left(2K'C_\rho^1A_1+2K'A_1^2+\frac{2K'C_\rho^3(\sqrt{A_1})^3}{\sqrt{T}}+2K'C_\rho^2A_1^2+\frac{2K'C_\rho^4+2K'C_\rho^5}{T}\right)(k-j)^2\delta_m^2.$$

Combined the above bound with $\Pi(1)$, there is a finite K'' such that

$$E\|J_{nT}(\gamma_k)-J_{nT}(\gamma_j)\|^4 \leq K''(k-j)^2\delta_m^2.$$

The above bound and Theorem 10.2 of Billingsley (1999) imply that there is a finite K_3 such that

$$P(\max_{2\leq k\leq m+1} \|J_{nT}(\gamma_k) - J_{nT}(\gamma_1)\| \geq \eta) \leq K_3 \frac{(m\delta_m)^2}{\eta^4} = K_3 \frac{\delta^2}{\eta^4}. \tag{3.10.22}$$

which bounds the first term on the right-hand side of (3.10.16). Similarly, there exist a K_4 such that

$$E|\sqrt{nT}(H_{nT,k} - EH_{nT,k})|^4 \leq E|\sqrt{nT}H_{nT,k}|^4 \leq K_4\delta_m^2.$$

The above bounds and Markov's inequality yield that there is a finite constant K_4 such that

$$P(\max_{1\leq k\leq m} |\sqrt{T}(H_{nT,k} - EH_{nT,k})| \geq \eta) \leq m\frac{K_4\delta_m^2}{\eta^4} \leq \frac{K_4\delta^2}{\eta^4} \tag{3.10.23}$$

Next, we consider the third term, by $\delta_m \leq \frac{2}{nT}$

$$\sqrt{nT}EH_{nT,k} = \sqrt{nT}\frac{1}{nT}\sum_{i=1}^{T}\sum_{t=1}^{T} Eh_{it,k} \leq \sqrt{nT}A_1\delta_m \leq \frac{2A_1}{\sqrt{nT}}. \tag{3.10.24}$$

(3.10.22), (3.10.23) and (3.10.24) imply that when $\frac{2A_1}{\sqrt{nT}} \leq \eta$,

$$P\left(\sup_{\gamma_1\leq\gamma\leq\gamma_1+\delta} \|J_{nT}(\gamma) - J_{nT}(\gamma_1)\| > 3\eta\right) \leq \frac{(K_3+K_4)\delta^2}{\eta^4}.$$

This implies $K_1 = 3^4(K_3 + K_4)$ and $K_2 = 6A_1$.

Lemma 3.4. *Suppose that Assumption 3.1 to Assumption 3.3 hold, we have*

$$J_{nT}(\gamma) \Rightarrow J(\gamma),$$

a mean-zero Gaussian process with almost surely continuous sample paths.

Proof of Lemma 3.4 The proof is an extended version of that of Lemma A.4 in Hansen (2000) with cross-sectional dependence carefully examined. For each γ, $\sum_{i=1}^{n} Z_{it}\epsilon_{it}I(q_{it} \leq \gamma)$ is a square integrable stationary martingale difference, so

$J_{nT}(\gamma)$ converges pointwise to a Gaussian distribution by the central limit theorem. This can be extended to any finite collection of γ to yield the convergence of the finite dimensional distributions.

Fix $\varepsilon > 0$ and $\eta > 0$. Set $\delta = \frac{\varepsilon\eta^4}{K_1}$ and $\overline{nT} = \max[\frac{1}{\delta}, \frac{K_2^2}{\eta^2}]$, where K_1 and K_2 are defined in Lemma 3.3. Then by Lemma 3.3, for any γ_1, if $nT \geq \overline{nT}$,

$$P\left(\sup_{\gamma_1 \leq \gamma \leq \gamma_1+\delta} \|J_{nT}(\gamma) - J_{nT}(\gamma_1)\| > \eta\right) \leq \frac{K_1\delta^2}{\eta^4} = \delta\varepsilon.$$

This implies $J_{nT}(\gamma)$ is tight. Hence we have $J_{nT}(\gamma) \Rightarrow J(\gamma)$.

Lemma 3.5. *Let* $k_{it}(\gamma, \gamma_0) = \|Z_{it}\||I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_0)|$, $K_{nT}(\gamma) = \frac{1}{nT}\sum_{i=1}^{n} \sum_{t=1}^{T} k_{it}(\gamma, \gamma_0)^2$, $G_{nT}(\gamma) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}(C'Z_{it})^2|I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_0)|$. *Suppose that Assumption 3.1 to Assumption 3.3 hold. There exist constants* $B > 0$ *and* $0 < d, k < \infty$*, such that for all* $\eta > 0$ *and* $\delta > 0$*, there exists a* $\bar{v} < \infty$ *such that for large enough* (n, T), $\lambda_{nT} = n^{1-2\alpha_1}T^{1-2\alpha_2}$,

$$P\left(\inf_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{G_{nT}(\gamma)}{|\gamma - \gamma_0|} < (1-\eta)d\right) \leq \varepsilon,$$

$$P\left(\sup_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{K_{nT}(\gamma)}{|\gamma - \gamma_0|} > (1+\eta)k\right) \leq \varepsilon.$$

Proof of Lemma 3.5 The proof is an extended version of that of Lemma A.7 in Hansen (2000) with cross-sectional dependence carefully examined. Note that $EG_{nT}(\gamma) = \frac{1}{n}\sum_{i=1}^{n} E(C'Z_{it})^2|I(q_{it} \leq \gamma) - I(q_{it} \leq \gamma_0)|$. First for $\gamma \geq \gamma_0$,

$$\frac{d}{d\gamma}EG_{nT}(\gamma) = C'D(\gamma)C,$$

where $D(\gamma) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} E(Z_{it}Z_{it}'|\gamma)f_i(\gamma)$. Since $C'D(\gamma_0)C > 0$ and $C'D(\gamma)C$ is continuous at γ_0, there is a $B > 0$ small enough such that

$$d = \min_{|\gamma-\gamma_0| \leq B} C'D(\gamma)C > 0.$$

Since $EG_{nT}(\gamma_0) = 0$, a first-order Taylor series expansion about γ_0 yields

$$\inf_{|\gamma-\gamma_0|\leq B} EG_{nT}(\gamma) \geq d|\gamma - \gamma_0|. \tag{3.10.25}$$

By Lemma 3.2,

$$\begin{aligned} E|G_{nT}(\gamma) - EG_{nT}(\gamma)|^2 \leq& \|C\|^4 E\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(k_{it}^2(\gamma,\gamma_0) - Ek_{it}^2(\gamma,\gamma_0)\right)\right|^2 \\ \leq& \frac{\|C\|^4 A_2|\gamma-\gamma_0|}{nT}. \end{aligned} \tag{3.10.26}$$

For any $\varepsilon, \eta > 0$, set

$$b = \frac{1-\frac{\eta}{2}}{1-\eta} > 1 \text{ and } \bar{v} = \frac{8\|C\|^4 A_2}{\eta^2 d^2(1-1/b)\varepsilon}$$

Since $n, T \to \infty$, $\frac{\bar{v}}{\lambda_{nT}} \leq B$. For $j = 1, 2, ..., m+1$, set $\gamma_j = \gamma_0 + \frac{\bar{v}b^{j-1}}{\lambda_{nT}}$, where m is the integer such that $\gamma_m - \gamma_0 = \frac{\bar{v}b^{m-1}}{\lambda_{nT}} \leq B$ and $\gamma_{m+1} - \gamma_0 > B$. By Markov inequality and (3.10.25) and (3.10.26),

$$\begin{aligned} P\left(\sup_{1\leq j\leq m}\left|\frac{G_{nT}(\gamma_j) - EG_{nT}(\gamma_j)}{EG_{nT}(\gamma_j)}\right| > \frac{\eta}{2}\right) \leq& \left(\frac{2}{\eta}\right)^2\sum_{j=1}^{m}\frac{E|G_{nT}(\gamma_j) - EG_{nT}(\gamma_j)|^2}{|EG_{nT}(\gamma_j)|^2} \\ \leq& \frac{4}{\eta^2}\sum_{j=1}^{m}\frac{\|C\|^4 A_2|\gamma_j-\gamma_0|}{nTd^2|\gamma_j-\gamma_0|^2} \\ =& n^{-2\alpha_1}T^{-2\alpha_2}\frac{4\|C\|^4 A_2}{\eta^2 d^2 \bar{v}}\sum_{j=1}^{m}\frac{1}{b^{j-1}} \\ \leq& \frac{4\|C\|^4 A_2}{\eta^2 d^2\bar{v}}\frac{1}{1-1/b} = \frac{\varepsilon}{2}. \end{aligned}$$

Thus with probability greater than $1 - \frac{\varepsilon}{2}$, for all $1 \leq j \leq m$,

$$\left|\frac{G_{nT}(\gamma_j)}{EG_{nT}(\gamma_j)} - 1\right| \leq \frac{\eta}{2}.$$

$G_{nT}(\gamma)$ as a function of γ is an increasing function for $\gamma > \gamma_0$. For any γ such that $\frac{\bar{v}}{\lambda_{nT}} \leq \gamma - \gamma_0 \leq B$, there is some $j \leq m$ such that $\gamma_j < \gamma < \gamma_{j+1}$,

$$\frac{G_{nT}(\gamma)}{|\gamma-\gamma_0|} \geq \frac{G_{nT}(\gamma_j)}{|\gamma_{j+1}-\gamma_0|} = \frac{G_{nT}(\gamma_j)}{EG_{nT}(\gamma_j)}\frac{EG_{nT}(\gamma_j)}{|\gamma_{j+1}-\gamma_0|} \geq \left(1-\frac{\eta}{2}\right)\frac{d|\gamma_j-\gamma_0|}{|\gamma_{j+1}-\gamma_0|} = (1-\eta)d,$$

since $\frac{|\gamma_j-\gamma_0|}{|\gamma_{j+1}-\gamma_0|} = \frac{1}{b}$ and $b = \frac{1-\frac{\eta}{2}}{1-\eta} > 1$. The probability is greater than $1-\frac{\varepsilon}{2}$. Thus we have

$$P\left(\inf_{\frac{\bar{v}}{\lambda_{nT}}\leq|\gamma-\gamma_0|\leq B}\frac{G_{nT}(\gamma)}{|\gamma-\gamma_0|} < (1-\eta)d\right) \leq \frac{\varepsilon}{2}.$$

A symmetric argument applies to the case $-B \leq \gamma-\gamma_0 \leq -\frac{\bar{v}}{\lambda_{nT}}$.

Lemma 3.6. *Suppose that Assumption 3.1 to Assumption 3.3 hold. There exists some $\bar{v} < \infty$ such that for any $B < \infty$, $\lambda_{nT} = n^{1-2\alpha_1}T^{1-2\alpha_2}$,*

$$P\left(\sup_{\frac{\bar{v}}{\lambda_{nT}}\leq|\gamma-\gamma_0|\leq B}\frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}|\gamma-\gamma_0|} > \eta\right) \leq \varepsilon.$$

Proof of Lemma 3.6 The proof is an extended version of that of Lemma A.8 in Hansen (2000) with cross-sectional dependence carefully examined. Fix $\eta > 0$. For $j = 1,2,...,$ set $\gamma_j-\gamma_0 = \frac{\bar{v}2^{j-1}}{\lambda_{nT}}$, where $\bar{v} < \infty$ will be determined later on. A straight forward calculation shows that Fix $\eta > 0$. For $j = 1,2,...,$ set $\gamma_j-\gamma_0 = \frac{\bar{v}2^{j-1}}{\lambda_{nT}}$, where $\bar{v} < \infty$ will be determined later on. A straight forward calculation shows that

$$\begin{aligned}\sup_{\frac{\bar{v}}{\lambda_{nT}}\leq|\gamma-\gamma_0|\leq B}\frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}|\gamma-\gamma_0|} \leq& 2\sup_{j>0}\frac{\|J_{nT}(\gamma_j)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|} \\ &+2\sup_{j>0}\sup_{\gamma_j\leq\gamma\leq\gamma_{j+1}}\frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_j)\|}{\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|}.\end{aligned} \tag{3.10.27}$$

By Assumption 3.1 (ii), the martingale difference property and Lemma 3.1,

$$\begin{aligned}E\|J_{nT}(\gamma_j)-J_{nT}(\gamma_0)\|^2 \leq& \frac{1}{nT}\sum_{t=1}^{T}E\left\|\sum_{i=1}^{n}Z_{it}\epsilon_{it}(I(q_{it}\leq\gamma_j)-I(q_{it}\leq\gamma_0))\right\|^2 \\ \leq& A_1|\gamma_j-\gamma_0|.\end{aligned}$$

By Markov inequality, we have

$$P\left(2\sup_{j>0}\frac{\|J_{nT}(\gamma_j)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|}>\eta\right)\le\frac{4}{\eta^2}\sum_{j=1}^{\infty}\frac{E\|J_{nT}(\gamma_j)-J_{nT}(\gamma_0)\|^2}{\lambda_{nT}|\gamma_j-\gamma_0|^2}$$
$$\le\frac{4}{\eta^2}\sum_{j=1}^{\infty}\frac{A_1|\gamma_j-\gamma_0|}{\lambda_{nT}|\gamma_j-\gamma_0|^2}$$
$$=\frac{4A_1}{\eta^2}\sum_{j=1}^{\infty}\frac{1}{\bar{v}2^{j-1}}=\frac{8A_1}{\eta^2\bar{v}}. \tag{3.10.28}$$

Set $\delta_j=\gamma_{j+1}-\gamma_j$, and $\eta_j=\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|\eta$. Then

$$P\left(2\sup_{j>0}\sup_{\gamma_j\le\gamma\le\gamma_{j+1}}\frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_j)\|}{\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|}>\eta\right)\le 2\sum_{j=1}^{\infty}P\left(\sup_{\gamma_j\le\gamma\le\gamma_{j+1}}\frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_j)\|}{\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|}>\eta\right)$$
$$\le 2\sum_{j=1}^{\infty}P\left(\sup_{\gamma_j\le\gamma\le\gamma_{j+1}}\|J_{nT}(\gamma)-J_{nT}(\gamma_j)\|>\eta_j\right). \tag{3.10.29}$$

When $\bar{v}\ge 1$, we have $\delta_j\ge\lambda_{nT}^{-1}\ge(nT)^{-1}$. Furthermore, if $\bar{v}\ge\frac{B}{\eta}$, then $\eta_j=\frac{1}{\sqrt{\lambda_{nT}}}2^{j-1}\bar{v}\eta\ge K_2\frac{1}{\sqrt{\lambda_{nT}}}\ge K_2\frac{1}{\sqrt{nT}}$. Thus if $\bar{v}\ge\max[1,K_2/\eta]$ the conditions for Lemma 3.3 hold, (3.10.29) is bounded by

$$2\sum_{j=1}^{\infty}\frac{K_1\delta_j^2}{\eta_j^4}=2\sum_{j=1}^{\infty}\frac{K_1|\gamma_{j+1}-\gamma_j|^2}{\lambda_{nT}^2|\gamma-\gamma_0|^4}=\frac{8K_1}{3\eta^4\bar{v}^2}. \tag{3.10.30}$$

Thus we can prove that if $\bar{v}\ge\max[1,K_2/\eta]$

$$P\left(2\sup_{j>0}\frac{\|J_{nT}(\gamma_j)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}|\gamma_j-\gamma_0|}>2\eta\right)\le\frac{8A_1}{\eta^2\bar{v}}+\frac{8K_1}{3\eta^4\bar{v}^2},$$

by choosing $\bar{v}$ large enough.

Lemma 3.7. *Let* $K_{nT}^*(\gamma)=\frac{1}{nT^2}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\|Z_{it}\||I(q_{it}\le\gamma)-I(q_{it}\le\gamma_0)|\right)^2$, $J_{nT}^*(\gamma)=\frac{1}{n^{\frac{1}{2}}T^{\frac{3}{2}}}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}Z_{it}(I(q_{it}\le\gamma)-I(q_{it}\le\gamma_0))\right)\left(\sum_{t=1}^{T}\epsilon_{it}\right)$. *Suppose that Assumption 3.1 to Assumption 3.3 hold. There exist some* $\bar{v}<\infty$ *and* $B>0$ *such that for any*

$\eta > 0$, $\varepsilon > 0$, $\lambda_{nT} = n^{1-2\alpha_1}T^{1-2\alpha_2}$,

$$P\left(\sup_{\frac{\bar{v}}{\lambda_{nT}}\leq|\gamma-\gamma_0|\leq B}\frac{K^*_{nT}(\gamma)}{|\gamma-\gamma_0|} > \eta\right) \leq \varepsilon,$$

$$P\left(\sup_{\frac{\bar{v}}{\lambda_{nT}}\leq|\gamma-\gamma_0|\leq B}\frac{\|J^*_{nT}(\gamma)\|}{\sqrt{\lambda_{nT}}|\gamma-\gamma_0|} > \eta\right) \leq \varepsilon.$$

Proof of Lemma 3.7 The proof is an extended version of that of Lemma C.9 in Miao et al. (2020b) with cross-sectional dependence carefully examined. We consider the case $\gamma > \gamma_0$. Let $k_{it}(\gamma,\gamma_0) = \|Z_{it}\||I(q_{it}\leq\gamma) - I(q_{it}\leq\gamma_0)|$. By Lemma 3.1 and Lemma 3.2, there is a constant A' such that

$$\begin{aligned}
EK^*_{nT}(\gamma) =& \frac{1}{n}\sum_{i=1}^{n}E\left[\frac{1}{T}\sum_{t=1}^{T}k_{it}(\gamma,\gamma_0)\right]^2 \\
=& \frac{1}{n}\sum_{i=1}^{n}E\left[\frac{1}{T}\sum_{t=1}^{T}k_{it}(\gamma,\gamma_0) - \frac{1}{T}\sum_{t=1}^{T}Ek_{it}(\gamma,\gamma_0)\right]^2 + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}Ek_{it}(\gamma,\gamma_0)\right]^2 \\
\leq& \frac{A'}{nT^2}\sum_{i=1}^{n}\sum_{t=1}^{T}Ek^2_{it}(\gamma,\gamma_0) + \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{T}\sum_{t=1}^{T}Ek_{it}(\gamma,\gamma_0)\right]^2 \\
\leq& \frac{A'A_1}{T}|\gamma-\gamma_0| + A_1^2|\gamma-\gamma_0|^2.
\end{aligned}$$

Choose a $b > 1$, $B < \frac{\varepsilon(b-1)\eta}{4A_1^2b^3}$. For $j = 1,..,m+1$, set $\gamma_j = \gamma_0 + \frac{\bar{v}b^{j-1}}{\lambda_{nT}}$, where m is the integer such that $\gamma_m - \gamma_0 \leq B$ and $\gamma_{m+1} - \gamma_0 > B$. When (n,T) is large enough, we can have $\frac{A'A_1mb}{T\eta} \leq \frac{\varepsilon}{4}$. By Markov's inequality,

$$\begin{aligned}
P\left(\sup_{1\leq j\leq m}\frac{K^*_{nT}(\gamma_{j+1})}{|\gamma_j-\gamma_0|} > \eta\right) \leq& \sum_{j=1}^{m}\frac{EK^*_{nT}(\gamma_{j+1})}{\eta|\gamma_j-\gamma_0|} \\
\leq& \sum_{j=1}^{m}\frac{A'A_1|\gamma_{j+1}-\gamma_0|}{T\eta|\gamma_j-\gamma_0|} + \sum_{j=1}^{m}\frac{A_1^2|\gamma_{j+1}-\gamma_0|^2}{\eta|\gamma_j-\gamma_0|} \\
=& \frac{A'A_1mb}{T\eta} + \frac{A_1^2b^2\bar{v}(b^m-1)}{\eta\lambda_{nT}(b-1)} \\
\leq& \frac{A'A_1mb}{T\eta} + \frac{A_1^2b^3B}{\eta(b-1)} \leq \frac{\varepsilon}{2}.
\end{aligned}$$

For any γ such that $\frac{\bar{v}}{\lambda_{nT}} \leq \gamma - \gamma_0 \leq B$, there is some $j \leq n$ such that $\gamma_j \leq \gamma \leq \gamma_{j+1}$. Since $K^*_{nT}(\gamma)$ is monotonic in γ, we have $\frac{K^*_{nT}(\gamma)}{|\gamma-\gamma_0|} \leq \frac{K^*_{nT}(\gamma_{j+1})}{|\gamma_j-\gamma_0|}$. It follows that

$$P\left(\sup_{\frac{\bar{v}}{\lambda_{nT}} \leq \gamma-\gamma_0 \leq B} \frac{|K^*_{nT}(\gamma)|}{|\gamma-\gamma_0|} > \eta\right) \leq P\left(\sup_{1\leq j\leq m} \frac{K^*_{nT}(\gamma_{j+1})}{|\gamma_j-\gamma_0|} > \eta\right) \leq \frac{\varepsilon}{2}.$$

A symmetric argument gives us the proof for the case $-B \leq \gamma - \gamma_0 \leq -\frac{\bar{v}}{\lambda_{nT}}$. The proof of second result is analogous to that of Lemma 3.6.

Proposition 3.3. *Suppose that Assumption 3.1 to Assumption 3.3 hold. As $n \to \infty$ and $T \to \infty$, we have $\lambda_{nT}(\hat{\gamma} - \gamma_0) = O_p(1)$, where $\lambda_{nT} = n^{1-2\alpha_1}T^{1-2\alpha_2}$.*

Proof of Proposition 3.3. The proof is an extended version of that of Lemma A.9 in Hansen (2000) with cross-sectional dependence carefully examined. Essentially, we use Lemma 3.4 to 3.7 established under cross-sectional dependence conditions rather than Lemma A.5 to A.8 in Hansen (2000). Let B, d and k be defined as in Lemma 3.4 Pick η and κ small enough so that

$$(1-\eta)d - (4\|C\| + 3\kappa)\kappa(1+\eta)k - (\|C\|^2 + 3\kappa^2 + 4\kappa\|C\| + 4\|C\| + 4\kappa)\eta > 0. \tag{3.10.31}$$

Let E_{nT} be the joint event that

(1) $|\hat{\gamma} - \gamma_0| \leq B$,

(2) $(n^{\alpha_1}T^{\alpha_2})\|\hat{\beta} - \beta_0\| \leq \kappa$,

(3) $\inf_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{G_{nT}(\gamma)}{|\gamma-\gamma_0|} > (1-\eta)d$,

(4) $\sup_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{K_{nT}(\gamma)}{|\gamma-\gamma_0|} < (1+\eta)k$,

(5) $\sup_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}|\gamma-\gamma_0|} < \eta$,

(6) $\sup_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{K^*_{nT}(\gamma)}{|\gamma-\gamma_0|} < \eta$,

(7) $\sup_{\frac{\bar{v}}{\lambda_{nT}} \leq |\gamma-\gamma_0| \leq B} \frac{\|J^*_{nT}(\gamma)\|}{\sqrt{\lambda_{nT}}|\gamma-\gamma_0|} < \eta.$

Fix $\delta > 0$, and choose $\bar{v}$ for large enough (n, T) such that $P(E_{nT}) \geq 1-\delta$, by Lemma 3.4 to Lemma 3.7. Since $Y^* = Z^*(\gamma_0)\beta_0 + \epsilon^*$,

$$Y^* - Z^*(\gamma)\hat{\beta} = \epsilon^* - Z^*(\gamma_0)(\hat{\beta}-\beta_0) - (Z^*(\gamma) - Z^*(\gamma_0))\hat{\beta}.$$

Hence

$$\begin{aligned}
&S_{nT}(\hat{\beta},\gamma) - S_{nT}(\hat{\beta},\gamma_0) \\
&= (Y^* - Z^*(\gamma)\hat{\beta})'(Y^* - Z^*(\gamma)\hat{\beta}) - (Y^* - Z^*(\gamma_0)\hat{\beta})'(Y^* - Z^*(\gamma_0)\hat{\beta}) \\
&= \hat{\beta}'(Z^*(\gamma) - Z^*(\gamma_0))'(Z^*(\gamma) - Z^*(\gamma_0))\hat{\beta} - 2\epsilon^{*'}(Z^*(\gamma) - Z^*(\gamma_0))\hat{\beta} \\
&\quad + 2\hat{\beta}'(Z^*(\gamma) - Z^*(\gamma_0))'Z^*(\gamma_0)(\hat{\beta}-\beta_0) \\
&= \beta_0'(Z^*(\gamma) - Z^*(\gamma_0))'(Z^*(\gamma) - Z^*(\gamma_0))\beta_0 - 2\epsilon^{*'}(Z^*(\gamma) - Z^*(\gamma_0))\hat{\beta} \\
&\quad + 2\hat{\beta}'(Z^*(\gamma) - Z^*(\gamma_0))'Z^*(\gamma_0)(\hat{\beta}-\beta_0) \\
&\quad + (\beta_0+\hat{\beta})'(Z^*(\gamma) - Z^*(\gamma_0))'(Z^*(\gamma) - Z^*(\gamma_0))(\hat{\beta}-\beta_0)
\end{aligned}$$

Define $\Delta Z^*(\gamma) = Z^{*1}(\gamma) - Z^{*1}(\gamma_0)$. Suppose $\gamma \in [\gamma_0 + \frac{\bar{v}}{\lambda_{nT}}, \gamma_0 + B]$ and E_{nT} holds. Let $\hat{C} = n^{\alpha_1}T^{\alpha_2}(\hat{\beta}_2 - \hat{\beta}_1)$ so that $\|\hat{C} - C\| \leq \kappa$.

$$\begin{aligned}
&\frac{S_{nT}(\hat{\beta},\gamma) - S_{nT}(\hat{\beta},\gamma_0)}{\lambda_{nT}(\gamma-\gamma_0)} \\
=&\frac{C'\Delta Z^*(\gamma)'\Delta Z^*(\gamma)C}{nT(\gamma-\gamma_0)} + \frac{2\epsilon^{*'}\Delta Z^*(\gamma)\hat{C}}{n^{1-\alpha_1}T^{1-\alpha_2}(\gamma-\gamma_0)} \\
&- \frac{2\hat{C}'\Delta Z^*(\gamma)'Z^*(\gamma_0)(\hat{\beta}-\beta_0)}{n^{1-\alpha_1}T^{1-\alpha_2}(\gamma-\gamma_0)} + \frac{(C+\hat{C})'\Delta Z^*(\gamma)'\Delta Z^*(\gamma)(\hat{C}-C)}{nT(\gamma-\gamma_0)} \\
=&\frac{\sum_{i=1}^n\sum_{t=1}^T C'\Delta Z_{it}(\gamma)\Delta Z_{it}(\gamma)'C}{nT(\gamma-\gamma_0)} + \frac{(C+\hat{C})'\sum_{i=1}^n\sum_{t=1}^T \Delta Z_{it}(\gamma)\Delta Z_{it}(\gamma)'(\hat{C}-C)}{nT(\gamma-\gamma_0)} \\
&- \frac{1}{T}\frac{\hat{C}'\sum_{i=1}^n\left(\sum_{t=1}^T \Delta Z_{it}(\gamma)\right)\left(\sum_{t=1}^T \Delta Z_{it}(\gamma)\right)\hat{C}}{nT(\gamma-\gamma_0)} + \frac{2\sum_{i=1}^n\sum_{t=1}^T \Delta Z_{it}(\gamma)'\epsilon_{it}\hat{C}}{n^{1-\alpha_1}T^{1-\alpha_2}(\gamma-\gamma_0)} \\
&- \frac{2}{T}\frac{\sum_{i=1}^n\left(\sum_{t=1}^T \Delta Z_{it}(\gamma)'\right)\left(\sum_{t=1}^T \epsilon_{it}\right)\hat{C}}{n^{1-\alpha_1}T^{1-\alpha_2}(\gamma-\gamma_0)} - \frac{2\hat{C}'\sum_{i=1}^n\sum_{t=1}^T \Delta Z_{it}(\gamma)Z_{it}'(\gamma_0)(\hat{\beta}-\beta_0)}{n^{1-\alpha_1}T^{1-\alpha_2}(\gamma-\gamma_0)} \\
&+ \frac{2}{T}\frac{\hat{C}'\sum_{i=1}^n\left(\sum_{t=1}^T \Delta Z_{it}(\gamma)\right)\left(\sum_{t=1}^T Z_{it}'(\gamma_0)\right)(\hat{\beta}-\beta_0)}{n^{1-\alpha_1}T^{1-\alpha_2}(\gamma-\gamma_0)}
\end{aligned}$$

$$
\begin{aligned}
\geq & \frac{G_{nT}(\gamma)}{(\gamma-\gamma_0)} - \|C+\hat{C}\|\|\hat{C}-C\|\frac{K_{nT}(\gamma)}{(\gamma-\gamma_0)} - \|\hat{C}\|^2\frac{K^*_{nT}(\gamma)}{(\gamma-\gamma_0)} - 2\|\hat{C}\|\frac{\|J_{nT}(\gamma)-J_{nT}(\gamma_0)\|}{\sqrt{\lambda_{nT}}(\gamma-\gamma_0)} \\
& -2\|\hat{C}\|\frac{\|J^*_{nT}(\gamma)\|}{\sqrt{\lambda_{nT}}(\gamma-\gamma_0)} - 2\|\hat{C}\|(n^{\alpha_1}T^{\alpha_2})\|\hat{\beta}-\beta_0\|\frac{K_{nT}(\gamma)}{(\gamma-\gamma_0)} \\
& -2\|\hat{C}\|(n^{\alpha_1}T^{\alpha_2})\|\hat{\beta}-\beta_0\|\frac{K^*_{nT}(\gamma)}{(\gamma-\gamma_0)} \\
\geq & (1-\eta)d - \|C+\hat{C}\|\|\hat{C}-C\|(1+\eta)k - \|\hat{C}\|^2\eta + 2\|\hat{C}\|\eta - 2\|\hat{C}\|\eta \\
& -2\|\hat{C}\|(n^{\alpha_1}T^{\alpha_2})\|\hat{\beta}-\beta_0\|(1+\eta)k - 2\|\hat{C}\|(n^{\alpha_1}T^{\alpha_2})\|\hat{\beta}-\beta_0\|\eta \\
\geq & (1-\eta)d - (2\|C\|+\kappa)\kappa(1+\eta)k - (\|C\|+\kappa)^2\eta - 4(\|C\|+\kappa)\eta \\
& -2(\|C\|+\kappa)\kappa(1+\eta)k - 2(\|C\|+\kappa)\kappa\eta \\
= & (1-\eta)d - (4\|C\|+3\kappa)\kappa(1+\eta)k - (\|C\|^2+3\kappa^2+4\kappa\|C\|+4\|C\|+4\kappa)\eta > 0.
\end{aligned}
$$

We have shown that on the set E_{nT}, if $\gamma \in [\gamma_0 + \frac{\bar{v}}{\lambda_{nT}}, \gamma_0 + B]$, then $S(\hat{\beta},\gamma) - S(\hat{\beta},\gamma_0) > 0$. We can similarly show that if $\gamma \in [\gamma_0 - B, \gamma_0 - \frac{\bar{v}}{\lambda_{nT}}]$ then $S(\hat{\beta},\gamma) - S(\hat{\beta},\gamma_0) > 0$. Since $S(\hat{\beta},\hat{\gamma}) - S(\hat{\beta},\gamma_0) \leq 0$, we can conclude that when E_{nT} occurs, we have $|\hat{\gamma}-\gamma_0| < \frac{\bar{v}}{\lambda_{nT}}$. As $P(E_{nT}) \geq 1-\delta$, for any $\delta > 0$, there is a constant $\bar{v}$ such that for (n,T) sufficiently large, we have $P(|\hat{\gamma}-\gamma_0| \geq \frac{\bar{v}}{\lambda_{nT}}) < \delta$.

Lemma 3.8. *Let* $\tilde{G}_{nT}(v) = \lambda_{nT}G_{nT}(\gamma_0 + \frac{v}{\lambda_{nT}})$ *and* $\tilde{K}_{nT}(v) = \lambda_{nT}K_{nT}(\gamma_0 + \frac{v}{\lambda_{nT}})$. *Suppose that Assumption 3.1 to Assumption 3.3 hold. Then we have uniformly* $v \in \Psi$,

$$\tilde{G}_{nT}(v) \xrightarrow{p} C'DC|v|,\ \tilde{K}_{nT}(v) \xrightarrow{p} D|v|$$

where Ψ *is a compact set.*

Proof of Lemma 3.8 The proof is an extended version of that of Lemma A.10 in Hansen (2000) with cross-sectional dependence carefully examined. Fix $v \in \Psi$, by (3.4.10)

$$E\tilde{G}_{nT}(v) = \lambda_{nT}C'E\left[Z_{it}Z'_{it}\left|I\left(q_{it} \leq \gamma_0 + \frac{v}{\lambda_{nT}}\right) - I(q_{it} \leq \gamma_0)\right|\right]C \to C'DC|v|,$$

By (3.10.26)

$$E|\tilde{G}_{nT}(v) - E\tilde{G}_{nT}(v)|^2 \leq' \frac{\lambda^2_{nT}}{nT}\|C\|^4 A_2\left|\frac{v}{\lambda_{nT}}\right| = o(1).$$

Thus we have $\tilde{G}_{nT}(v) \xrightarrow{p} C'DC|v|$.

Suppose $\Psi = [0, \bar{v}]$. Since $\tilde{G}_{nT}(v)$ is monotonically increasing on Ψ and the limit function is continuous, the convergence is uniform over Ψ. To see this, set $\tilde{G}(v) = C'DCv$. Pice any $\varepsilon > 0$, then set $J = \frac{\bar{v}C'DC}{\varepsilon}$ and for $j = 0, 1, ..., J$, set $v_j = \frac{\bar{v}C'DCj}{J}$. Then pick n and T large enough so that $\max_{j \le J} |\tilde{G}_{nT}(v_j) - \tilde{G}(v_j)| \le \varepsilon$ with probability greater than $1 - \varepsilon$, which is possible by pointwise consistency. For any $j \ge 1$, take any $v \in (v_{j-1}, v_j)$. Both $\tilde{G}_{nT}(v)$ and $\tilde{G}(v)$ lie in the interval $[\tilde{G}(v_{j-1}) - \varepsilon, \tilde{G}(v_j) + \varepsilon]$ with probability greater than $1 - \varepsilon$, which has length bounded by 3ε. Since v is arbitrary, $|\tilde{G}_{nT}(v) - \tilde{G}(v)| \le 3\varepsilon$ uniformly over Ψ.

An identical argument yields uniformity over sets of the form $[-\bar{v}, 0]$, and thus for arbitrary compact sets Ψ.

Lemma 3.9. *Let $R_{nT}(v) = \sqrt{\lambda_{nT}}(J_{nT}(\gamma_0 + \frac{v}{\lambda_{nT}}) - J_{nT}(\gamma_0))$. Suppose that Assumption 3.1 to Assumption 3.3 hold. Then on any compact set Ψ,*

$$R_{nT}(v) \Rightarrow B(v).$$

where $B(v)$ is a vector Brownian motion with covariance matrix $E[B(1)B(1)'] = V$.

Proof of Lemma 3.9 First, we establish the convergence of the finite dimensional distributions of $R_{nT}(v)$, $R_{nT}(v) \xrightarrow{d} N(0, |v|V)$, and then show $R_{nT}(v)$ is tight.

Let $\xi_{nT}(v) = \frac{\sqrt{\lambda_{nT}}}{\sqrt{nT}} \sum_{i=1}^{n} \sum_{t=1}^{T} Z_{it}\epsilon_{it}[I(q_{it} \le \gamma_0 + \frac{v}{\lambda_{nT}}) - I(q_{it} \le \gamma_0)]$ and $u_{it}(v) = \sqrt{\lambda_{nT}} Z_{it}\epsilon_{it}[I(q_{it} \le \gamma_0 + \frac{v}{\lambda_{nT}}) - I(q_{it} \le \gamma_0)]$. By martingale difference array central limit theorem (for example, Theorem 24.3 of Davidson (1994)), it suffices to verify that

$$\xi_{nT}(v)\xi_{nT}(v)' \xrightarrow{p} |v|V \text{ and } \frac{1}{\sqrt{nT}} \sum_{i=1}^{n} \|u_{it}(v)\| = o_p(1).$$

Note that

$$E[\xi_{nT}(v)\xi_{nT}(v)']$$

$$
\begin{aligned}
=&\frac{\lambda_{nT}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}E\left(Z_{it}Z_{it}'\epsilon_{it}\epsilon_{it}\left[I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right]\right)\\
&+\frac{\lambda_{nT}}{nT}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{t=1}^{T}E\left(Z_{it}Z_{jt}'\epsilon_{it}\epsilon_{jt}\left[I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right]\right.\\
&\left.\cdot\left[I\left(q_{jt}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{jt}\leq\gamma_0)\right]\right)\\
=:&\Xi(1)+\Xi(2).
\end{aligned}
$$

Similar to the proof of Theorem 3.1, we can easily show that $E\|\sum_{i=1}^{n}\sum_{t=1}^{T}Z_{it}\epsilon_{it}\|^2=O(nT)$. By Lemma 3.1, $\max_i E\|Z_{it}\epsilon_{it}\|\left|I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right|=O(\frac{1}{\lambda_{nT}})$, we have $E\|\Xi(2)\|=O(\frac{1}{\lambda_{nT}})=o(1)$. In addition, it is easy to verify that $Var[\Xi(2)]=o_p(1)$. For $\Xi(1)$, we can use similar calculation as that in the proof of Lemma 3.1,

$$
\frac{E\left[Z_{it}Z_{it}'\epsilon_{it}\epsilon_{it}\left[I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right]\right]}{v/\lambda_{nT}}\to E[Z_{it}Z_{it}'\epsilon_{it}\epsilon_{it}|q_{it}=\gamma_0]f_i(\gamma_0).
$$

Therefore, as $T\to\infty$ by (3.4.10), we can prove that $\xi_{nT}(v)\xi_{nT}(v)'\xrightarrow{p}|v|V$, where $V=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}E[Z_{it}Z_{it}'\epsilon_{it}\epsilon_{it}|q_{it}=\gamma_0]f_i(\gamma_0)$

Then we verify that $\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\|u_{it}(v)\|=o_p(1)$.

$$
\begin{aligned}
&E\left|\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\|u_{it}(v)\|\right|^4\\
\leq&\frac{1}{T}E\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\|u_{it}(v)\|\right)^4\\
=&\frac{1}{T}E\left(\frac{\lambda_{nT}}{n}\sum_{i=1}^{n}\|Z_{it}\epsilon_{it}\|^2\left[I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right]+o_p(1)\right)^2\\
=&\frac{\lambda_{nT}^2}{n^2T}E\left(\sum_{i=1}^{n}\|Z_{it}\epsilon_{it}\|^4\left[I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right]\right.\\
&+\sum_{i=1}^{n}\sum_{i\neq j}\|Z_{it}\epsilon_{it}\|^2\|Z_{jt}\epsilon_{jt}\|^2\left[I\left(q_{it}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{it}\leq\gamma_0)\right.\\
&\left.\cdot\left[I\left(q_{jt}\leq\gamma_0+\frac{v}{\lambda_{nT}}\right)-I(q_{jt}\leq\gamma_0)\right]\right)\\
=&O(\lambda_{nT}(nT)^{-1})+O((nT)^{-1}).
\end{aligned}
$$

Thus we conclude that $R_{nT}(v) \xrightarrow{d} N(0, |v|V)$. This argument can be extended to include any finite collection $[v_1, ..., v_k]$ to yield the convergence of the finite dimensional distributions of $R_{nT}(v)$ to those of $B(v)$.

We now show tightness. Fix $\varepsilon > 0$ and $\eta > 0$. Set $\delta = \frac{\varepsilon\eta^4}{K_1}$ and $\bar{n} = (\max[\frac{1}{\sqrt{\delta}}, \frac{K_2}{\eta}])^{\frac{1}{\alpha_1}}$, $\bar{T} = (\max[\frac{1}{\sqrt{\delta}}, \frac{K_2}{\eta}])^{\frac{1}{\alpha_2}}$, where K_1 and K_2 are defined in Lemma 3.3. Set $\gamma_1 = \gamma_0 + \frac{v_1}{\lambda_{nT}}$. By Lemma 3.3, for $n \geq \bar{n}$ and $T \geq \bar{T}$,

$$\begin{aligned} &P\left(\sup_{v_1 \leq v \leq v_1+\delta} |R_{nT}(v) - R_{nT}(v_1)| > \eta\right) \\ =&P\left(\sup_{\gamma_1 \leq \gamma \leq \gamma_1 + \frac{\delta}{\lambda_{nT}}} \|J_{nT}(\gamma) - J_{nT}(\gamma_1)\| > \frac{\eta}{\sqrt{\lambda_{nT}}}\right) \leq \frac{K_1(\delta/\lambda_{nT})^2}{\lambda_{nT}^{-2}\eta^2} \leq \delta\varepsilon. \end{aligned}$$

The conditions for Lemma 3.3 are met since $\frac{\delta}{\lambda_{nT}} \geq \frac{1}{nT}$ when $n^{\alpha_1}T^{\alpha_2} \geq \frac{1}{\sqrt{\lambda_{nT}}}$, and $\frac{\eta}{\sqrt{\lambda_{nT}}} \geq \frac{K_2}{\sqrt{nT}}$ when $n^{\alpha_1}T^{\alpha_2} \geq \frac{K_2}{\eta}$, and these hold for $n \geq \bar{n}$ and $T \geq \bar{T}$. This imply that $R_{nT}(\gamma)$ is tight.

Lemma 3.10. *Let* $\tilde{K}^*_{nT}(v) = \lambda_{nT}K^*_{nT}(\gamma_0 + \frac{v}{\lambda_{nT}})$ *and* $\tilde{J}^*_{nT}(v) = \lambda_{nT}J^*_{nT}(\gamma_0 + \frac{v}{\lambda_{nT}})$. *Suppose that Assumption 3.1 to Assumption 3.3 hold. Then* $\tilde{K}^*_{nT}(v) \xrightarrow{p} 0$ *and* $\tilde{J}^*_{nT}(v) \xrightarrow{p} 0$ *uniformly in* $v \in \Psi$, *where* Ψ *is a compact set.*

Proof of Lemma 3.10 By the proof of Lemma 3.7, we have

$$E[\tilde{K}^*_{nT}(v)] = O\left(\frac{1}{T}\right) + O\left(\frac{1}{\lambda_{nT}}\right) = o(1).$$

Thus we have $\tilde{K}^*_{nT}(v) = o_p(1)$ for each $v \in \Psi$. This result, in conjunction with the monotonicity of $\tilde{K}^*_{nT}(v)$ in either the half line $[0, \infty)$ or half line $(-\infty, 0]$, implies that $\tilde{K}^*_{nT}(v) \xrightarrow{p} 0$ uniformly in $v \in \Psi$. See Lemma 3.8.

For $\tilde{J}^*_{nT}(v)$, we can follow the above arguments and show that $\tilde{J}^*_{nT}(v) = o_p(1)$ for each $v \in \Psi$. Following Lemma 3.9, we can readily show the tightness of the process $\{\tilde{J}^*_{nT}(v)\}$. Thus we have $\tilde{J}^*_{nT}(v) \xrightarrow{p} 0$ uniformly in $v \in \Psi$.

Lemma 3.11. *Let* $Q_{nT}(v) = S_{nT}(\hat{\beta}, \gamma_0) - S_{nT}(\hat{\beta}, \gamma_0 + \frac{v}{\lambda_{nT}})$, *where* $\lambda_{nT} = n^{1-2\alpha_1}T^{1-2\alpha_2}$. *Suppose that Assumption 3.1 to Assumption 3.3 hold. Then on any compact set* Ψ,

$$Q_{nT}(v) \Rightarrow Q(v) = -C'DC|v| + 2\sqrt{C'VC}W(v)$$

Proof of Lemma 3.11 Let $\Delta Z_{it}(v) = Z_{it1}\left(\gamma_0 + \frac{v}{\lambda_{nT}}\right) - Z_{it1}(\gamma_0)$, We have

$$\begin{aligned}
Q_{nT}(v) =& S_{nT}(\hat{\beta}, \gamma_0) - S_{nT}\left(\hat{\beta}, \gamma_0 + \frac{v}{\lambda_{nT}}\right) \\
=& -n^{-2\alpha_1}T^{-2\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T} C'\Delta Z_{it}(v)\Delta Z_{it}(v)'C - 2n^{-\alpha_1}T^{-\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T} C'\Delta Z_{it}(v)\epsilon_{it} \\
&+ L_{nT}(v),
\end{aligned}$$

where

$$\begin{aligned}
L_{nT}(v) =& 2(C-\hat{C})'R_{nT}(v) + 2\hat{C}\tilde{K}_{nT}(v)n^{\alpha_1}T^{\alpha_2}(\hat{\beta}-\beta_0) \\
&- (\hat{C}-C)'\tilde{K}_{nT}(v)(\hat{C}+C) + \hat{C}'\tilde{K}^{*}_{nT}(v)\hat{C} \\
&+ 2\hat{C}'\tilde{J}^{*}_{nT}(v) - 2n^{-\alpha_1}T^{-\alpha_2-1}\hat{C}'\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\Delta Z_{it}(v)\right)\left(\sum_{t=1}^{T} Z_{it}(\gamma_0)'\right)(\hat{\beta}-\beta_0) \\
=& L_{1,nT}(v) + \ldots + L_{6,nT}(v).
\end{aligned}$$

By Lemma 3.8, we have

$$\begin{aligned}
& n^{-2\alpha_1}T^{-2\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T} C'\Delta Z_{it}(v)\Delta Z_{it}(v)'C \\
=& \frac{\lambda_{nT}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} C'\Delta Z_{it}(v)\Delta Z_{it}(v)'C \\
=& \tilde{G}_{nT}(v) \Rightarrow C'DC|v|.
\end{aligned}$$

By Lemma 3.9, we have

$$\begin{aligned}
& n^{-\alpha_1}T^{-\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T} C'\Delta Z_{it}(v)\epsilon_{it} \\
=& \sqrt{\lambda_{nT}}C'\left(J_{nT}\left(\gamma_0 + \frac{v}{\lambda_{nT}}\right) - J_{nT}(\gamma_0)\right) \\
=& C'R_{nT}(v) \Rightarrow C'B(v) = \sqrt{C'VC}W(v).
\end{aligned}$$

By the fact that $n^{\alpha_1}T^{\alpha_2}(\hat{\beta}-\beta_0)=o_p(1)$, Assumption 3.3 (i), and Lemma 3.8, we have $L_{g,nT}(v)=o_p(1)$ uniformly in v for $g=1,2,3$. By Lemma 3.10, we have that $L_{4,nT}(v)=o_p(1)$ and $L_{5,nT}(v)=o_p(1)$. For $L_{6,nT}(v)$, we have uniformly in $v\in\Psi$,

$$\begin{aligned}|L_{6,nT}(v)|\leq&2\|\hat{C}\|\left[n^{\alpha_1}T^{\alpha_2}\|\hat{\beta}-\beta_0\|\right]\frac{\lambda_{nT}}{nT^2}\left\|\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\Delta Z_{it}(v)\right)\left(\sum_{t=1}^{T}Z_{it}(\gamma_0)'\right)\right\|\\=&O_p(1)o_p(1)o_p(1),\end{aligned}$$

as we can follow the proofs of Lemmas 3.8 and 3.10 and show that $\frac{\lambda_{nT}}{nT^2}\left\|\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\Delta Z_{it}(v)\right)\left(\sum_{t=1}^{T}Z_{it}(\gamma_0)'\right)\right\|=o_p(1)$ uniformly in $v\in\Psi$. Thus we have $Q_{nT}(v)\Rightarrow -C'DC|v|+2\sqrt{C'VC}W(v)$ on any compact set Ψ.

Proof of Theorem 3.2. By Proposition 3.3, $\lambda_{nT}(\hat{\gamma}-\gamma_0)=\arg\max_v Q_{nT}(v)=O_p(1)$, and by Lemma A.10, $Q_{nT}(v)\Rightarrow Q(v)$. The limit functional $Q(v)$ is continuous, has a unique maximum, and $\lim_{|v|\to\infty}Q(v)=-\infty$ almost surely. It therefore satisfies the conditions of Theorem 2.7 of Kim and Pollard (1990),

(i) $Q_{nT}(v)\Rightarrow Q(v)$;

(ii) $\lambda_{nT}(\hat{\gamma}-\gamma_0)=O_p(1)$;

(iii) $Q_{nT}(\lambda_{nT}(\hat{\gamma}-\gamma_0))\geq\sup_v Q_{nT}(v)-a_{nT}$ for random variables $\{a_{nT}\}$ of order $o_p(1)$.

Then $\lambda_{nT}(\hat{\gamma}-\gamma_0)\xrightarrow{d}\arg\max Q(v)$.

This implies

$$\lambda_{nT}(\hat{\gamma}-\gamma_0)\xrightarrow{d}\arg\max_{v\in R}Q(v)$$

Setting $w=\frac{C'VC}{(C'DC)^2}$, we can rewrite the asymptotic distribution as

$$\arg\max_{v\in R}\left[-C'DC|v|+2\sqrt{C'VC}W(v)\right]=w\arg\max_{v\in R}\left[-\frac{|v|}{2}+W(v)\right].\quad(3.10.32)$$

This completes the proof of Theorem 3.2.

Proof of Theorem 3.3 By definition,

$$\hat{\beta}(\hat{\gamma}) = (Z^*(\hat{\gamma})'(Z^*(\hat{\gamma}))^{-1}(Z^*(\hat{\gamma})'Y^*.$$

Substituting $Y^* = Z^*(\gamma_0)\beta_0 + \epsilon^*$,

$$\begin{aligned}\hat{\beta} - \beta_0 =&\hat{\beta}(\hat{\gamma}) - \beta_0(\gamma_0)\\ =&\left(\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma})\right)^{-1}\left(\frac{1}{nT}Z^*(\hat{\gamma})'(Z^*(\gamma_0) - Z^*(\hat{\gamma}))\beta_0\right)\\ &+\left(\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma})\right)^{-1}\left(\frac{1}{nT}Z^*(\hat{\gamma})'\epsilon\right)\\ =:&\Pi(1) + \Pi(2).\end{aligned}$$

By Lemma A.1 and the consistency of $\hat{\gamma}$, $\hat{\gamma}\longrightarrow\gamma_0$, by setting $T\to\infty$ first and then setting $n\to\infty$, we have

$$\frac{1}{nT}Z^*(\hat{\gamma})'(Z^*(\gamma_0) - Z^*(\hat{\gamma}))\beta_0 \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}E[\tilde{Z}_{it}(\gamma_0)(\tilde{Z}_{it}(\gamma_0) - \tilde{Z}_{it}(\gamma_0))'\beta_0] = 0.$$

Therefore, $\Pi(1) = o_p(1)$. In the proof of Theorem 3.1, we showed that

$$\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma}) \to \frac{1}{n}\sum_{i=1}^{n}E[\tilde{Z}_{it}(\hat{\gamma})\tilde{Z}_{it}(\hat{\gamma})'].$$

By Lemma A.1 and the consistency of $\hat{\gamma}$, $\hat{\gamma}\longrightarrow\gamma_0$, by setting $T\to\infty$ first and then setting $n\to\infty$, and by Assumption 3.1 (iii), we have

$$\frac{1}{nT}Z^*(\hat{\gamma})'Z^*(\hat{\gamma}) \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}E[\tilde{Z}_{it}(\gamma_0)\tilde{Z}_{it}(\gamma_0)'] \to \Sigma_{\gamma_0}. \quad (3.10.33)$$

Similarly, by the proof of Theorem 3.1 and the properties of martingale difference, we have

$$Var\left(\frac{1}{\sqrt{nT}}Z^*(\hat{\gamma})'\epsilon\right) \to \frac{1}{n}\sum_{i=1}^{n}E[\tilde{Z}_{it}(\hat{\gamma})\tilde{Z}_{jt}(\hat{\gamma})'\epsilon_{it}\epsilon_{jt}].$$

By Lemma A.1 and the consistency of $\hat{\gamma}$, $\hat{\gamma}\longrightarrow\gamma_0$, by setting $T\to\infty$ first and then setting $n\to\infty$, and by Assumption 3.1 (iii), we have

$$Var\left(\frac{1}{\sqrt{nT}}Z^*(\hat{\gamma})'\epsilon\right) \xrightarrow{p} \frac{1}{n}\sum_{i=1}^{n}E[\tilde{Z}_{it}(\gamma_0)\tilde{Z}_{jt}(\gamma_0)'\epsilon_{it}\epsilon_{jt}] \to \Sigma_{\gamma_0,\epsilon}. \quad (3.10.34)$$

By (3.10.33) and (3.10.34), we can prove that $\sqrt{nT}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0_{2K+2}, \Sigma_{\gamma_0}^{-1}\Sigma_{\gamma_0,\epsilon}\Sigma_{\gamma_0}^{-1})$.

Proof of Theorem 3.4 Note that

$$\begin{aligned}\sqrt{nT}(\hat{\beta}(\gamma)-\beta_0) =&\hat{\Sigma}_{\gamma}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}^*(\gamma)(Z_{it}^*(\gamma_0)-Z_{it}^*(\gamma))'\beta_0\\ &+\hat{\Sigma}_{\gamma}^{-1}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} Z_{it}^*(\gamma)\epsilon_{it}.\end{aligned}$$

Recall that $F_{nT}(\gamma)=\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}Z_{it}^*(\gamma)\epsilon_{it}$ and let $\hat{\Theta}_{\gamma_1,\gamma_2}=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}Z_{it}^*(\gamma_1)(Z_{it}^*(\gamma_2))'$. By the proof of Theorem 3.1 and Lemma A.1, we have $\hat{\Sigma}_{\gamma}\xrightarrow{p}\Sigma_{\gamma}$, $F_{nT}(\gamma)\xrightarrow{p} F(\gamma)$ and $\hat{\Theta}_{\gamma,\gamma_0}\xrightarrow{p}\Theta_{\gamma,\gamma_0}$. By $(L_2-L_1)'\beta_0=\theta$, $\sqrt{nT}\theta=c$ and $c=\frac{1}{2}(L_2-L_1)'(L_2-L_1)c$, we have

$$\begin{aligned}\sqrt{nT}(L_2-L_1)'\hat{\beta}(\gamma) =&\sqrt{nT}(L_2-L_1)'\hat{\Sigma}_{\gamma}^{-1}\hat{\Theta}_{\gamma,\gamma_0}\beta_0+(L_2-L_1)'\hat{\Sigma}_{\gamma}^{-1}F_{nT}(\gamma)\\ =&\frac{1}{2}(L_2-L_1)'\hat{\Sigma}_{\gamma}^{-1}\hat{\Theta}_{\gamma,\gamma_0}(L_2-L_1)c+(L_2-L_1)'\hat{\Sigma}_{\gamma}^{-1}F_{nT}(\gamma)\\ =&\frac{1}{2}(L_2-L_1)'\Sigma_{\gamma}^{-1}\Theta_{\gamma,\gamma_0}(L_2-L_1)c+(L_2-L_1)'\Sigma_{\gamma}^{-1}F(\gamma)+o_p(1)\\ =&\tilde{Q}(\gamma)c+\tilde{F}(\gamma),\end{aligned}$$

where $\tilde{Q}(\gamma)=\frac{1}{2}(L_2-L_1)'\Sigma_{\gamma}^{-1}\Theta_{\gamma,\gamma_0}(L_2-L_1)$ and $\tilde{F}(\gamma)=(L_2-L_1)'\Sigma_{\gamma}^{-1}F(\gamma)$.

Next, it is standard to show that $\hat{M}_{nT}(\gamma)\xrightarrow{p}(L_2-L_1)'\Sigma_{\gamma}^{-1}\Sigma_{\gamma,\epsilon}\Sigma_{\gamma}^{-1}(L_2-L_1)=M(\gamma)$ uniformly in γ.

Then by the continuous mapping theorem, we have $W_{nT}(\gamma)\Rightarrow W^c(\gamma)$.
Proof of Theorem 3.5. The proof is similar to that of Theorem 3.3.

Proof of Theorem 3.6. Let $Q_{nT}(v_1,v_2)=S_{nT}(\hat{\beta},\gamma_{01},\gamma_{02})-S_{nT}(\hat{\beta},\gamma_{01}+\frac{v_1}{\lambda_{nT}},\gamma_{02}+\frac{v_2}{\lambda_{nT}})$. Let $\Delta Z_{it}(v_1)=Z_{it1}\left(\gamma_{01}+\frac{v_1}{\lambda_{nT}}\right)-Z_{it1}(\gamma_{01})$ and $\Delta Z_{it}(v_2)=Z_{it1}\left(\gamma_{02}+\frac{v_2}{\lambda_{nT}}\right)-Z_{it1}(\gamma_{02})$ We have

$$\begin{aligned}Q_{nT}(v_1,v_2)=&S_{nT}(\hat{\beta},\gamma_{01},\gamma_{02})-S_{nT}\left(\hat{\beta},\gamma_{01}+\frac{v_1}{\lambda_{nT}},\gamma_{02}+\frac{v_2}{\lambda_{nT}}\right)\\ =&-n^{-2\alpha_1}T^{-2\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T}C_1'\Delta Z_{it}(v_1)\Delta Z_{it}(v_1)'C_1-2n^{-\alpha_1}T^{-\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T}C'\Delta Z_{it}(v_1)\epsilon_{it}\\ &-n^{-2\alpha_1}T^{-2\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T}C_2'\Delta Z_{it}(v_2)\Delta Z_{it}(v_2)'C_2-2n^{-\alpha_1}T^{-\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T}C'\Delta Z_{it}(v_2)\epsilon_{it}\end{aligned}$$

$$+ L_{nT}(v_1, v_2),$$

where

$$\begin{aligned}
L_{nT}(v) =& - n^{-2\alpha_1}T^{-2\alpha_2}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{C}_1'\Delta Z_{it}(v_1)\Delta Z_{it}(v_2)'\hat{C}_2 \\
&+ n^{-2\alpha_1}T^{-2\alpha_2-1}\sum_{i=1}^{n}\hat{C}_1'\left(\sum_{t=1}^{T}\Delta Z_{it}(v_1)\right)\left(\sum_{t=1}^{T}\Delta Z_{it}(v_2)'\right)\hat{C}_2 \\
&+ 2(C_1-\hat{C}_1)'R_{nT}(v_1) + 2(C_1-\hat{C}_1)'R_{nT}(v_1) + 2\hat{C}_1\tilde{K}_{nT}(v_1)n^{\alpha_1}T^{\alpha_2}(\hat{\beta}-\beta_0) \\
&+ 2\hat{C}_2\tilde{K}_{nT}(v_2)n^{\alpha_1}T^{\alpha_2}(\hat{\beta}-\beta_0) - (\hat{C}_1-C_1)'\tilde{K}_{nT}(v_1)(\hat{C}_1+C_1) \\
&- (\hat{C}_2-C_2)'\tilde{K}_{nT}(v_2)(\hat{C}_2+C_2) + \hat{C}_1'\tilde{K}^*_{nT}(v_1)\hat{C}_1 + \hat{C}_2'\tilde{K}^*_{nT}(v_2)\hat{C}_2 \\
&+ 2\hat{C}_1'\tilde{J}^*_{nT}(v_1) + 2\hat{C}_2'\tilde{J}^*_{nT}(v_2) \\
&- 2n^{-\alpha_1}T^{-\alpha_2-1}\hat{C}_1'\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\Delta Z_{it}(v_1)\right)\left(\sum_{t=1}^{T}Z_{it}(\gamma_0)'\right)(\hat{\beta}-\beta_0) \\
&- 2n^{-\alpha_1}T^{-\alpha_2-1}\hat{C}_2'\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\Delta Z_{it}(v_2)\right)\left(\sum_{t=1}^{T}Z_{it}(\gamma_0)'\right)(\hat{\beta}-\beta_0) \\
=& L_{1,nT}(v_1,v_2) + \ldots + L_{14,nT}(v_1,v_2).
\end{aligned}$$

We can follow the proof of Lemma 3.11 and show that $L_{g,nT}(v_1, v_2) = o_p(1)$ uniformly in v for $g = 3, 4, ..., 14$.

For $L_{1,nT}(v_1, v_2)$, we have

$$\begin{aligned}
L_{1,nT}(v_1,v_2) =& \frac{\lambda_{nT}}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{C}_1'Z_{it}Z_{it}'\hat{C}_2\left[I\left(q_{it}\le\gamma_{01}+\frac{v_1}{\lambda_{nT}}\right)-I(q_{it}\le\gamma_{01})\right] \\
&\cdot\left[I\left(q_{it}\le\gamma_{02}+\frac{v_2}{\lambda_{nT}}\right)-I(q_{it}\le\gamma_{02})\right] \\
=& O_p\left(\frac{1}{\lambda_{nT}}\right) = o_p(1)
\end{aligned}$$

Following the proof of Lemma 3.10, we have

$$L_{2,nT}(v_1,v_2) = \frac{\lambda_{nT}}{nT^2}\sum_{i=1}^{n}\hat{C}_1'\left(\sum_{t=1}^{T}\Delta Z_{it}(v_1)\right)\left(\sum_{t=1}^{T}\Delta Z_{it}(v_2)'\right)\hat{C}_2 = o_p(1).$$

Thus we can prove that $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are asymptotically independent. The proof of asymptotic distribution is similar to that of Theorem 3.2, we omit it here.

Chapter 4

Adaptive Group Fused LASSO for panel threshold regression model

4.1 Introduction

In the past few decades, the idea of nonlinear time series analysis has been very popular (see Lu (1998); Lu and Jiang (2001)). We considered the nonlinear structure of panel time series model and studied the asymptotic distribution of panel threshold regression model with cross-sectional dependence in Chapter 3. Determining the number of thresholds is essential in real panel data analysis. There are only few literature studying how to determine the number of thresholds.

Lee, Seo, and Shin (2016) proposed a threshold estimation method for a high dimensional regression model with homoscedastic normal errors and with deterministic covariates via LASSO. Lee, Liao, Seo, and Shin (2018) then studied a high dimensional quantile regression model with a change-point which allowed heteroscedastic nonnormal errors and stochastic covariates. Both LASSO method for high dimensional models select only one change-point. However, it is more applicable to study multiple thresholds for many datasets nowadays, where the number of change point is indeterminable in most cases. Although multiple thresholds estimation is important for empirical studies, only limited literature studied how to estimate multiple-regime models. Tsay (1989) proposed to use scatter plots of various statistics to locate the thresholds and introduced that graphics did provide useful information in determining the number of thresholds. Gonzalo and Pitarakis (2002) introduced a joint and sequential model selection procedure in estimating single and multiple threshold models and studied its asymptotic and finite sample properties. Lasso has also been extended to nonlinear time series analysis. For example, Al-Sulami et al. (2019) considered selecting lagging variables in time series and space-time model via adaptive Lasso to study the nonlinear relationships between covariate variables and a response variable.

As for common breaks analysis, both Qian and Su (2016) and Li, Qian, and Su (2016) considered estimation of multiple structural breaks in panel data models via adaptive group fused lasso. Qian and Su (2016) proposed two estimation methods of common breaks. They are penalized least squares estimation and penalized GMM estimation for first-differenced models, where both approaches can estimate common break dates consistently. However, since this model dose not allow the existence of cross-sectional dependence, its applicability in empirical studies is restricted. Li et al. (2016) proposed a penalized principal component estimation to determine the number of breaks in panel data models with cross-sectional dependence and established the asymptotic distribution theory, which can also be applied to dynamic panel data models. Although there exist multiple regimes in both threshold models and structural break models, the changes take place in different variables, which occur in threshold variables and common dates in threshold models and structural break models respectively. These literature only considered change-point model via Lasso estimation.

In change-point analysis, Chan, Yau, and Zhang (2014) applied Group LASSO to structural break time series. They proposed a two-step LASSO procedure for multiple change-point estimation in time series. The number of change points can be estimated by Group LASSO in the first step. However, this is usually larger than the true number of change points in time series. Then, some prescribed information criterion can be adopted to select a subset of this over-estimated break set, which can estimate the number and locations of change-points appropriately. Compared with the change-point model, the change in the threshold autoregressive model happens to be the lagged observation rather than the time index. Therefore, the two-step LASSO estimation method cannot be applied to the threshold autoregressive model directly. Chan, Yau, and

Zhang (2015) have hence developed two-step procedure by sorting the threshold parameter so that it can be applicable for the threshold model.

As for the estimation of the number of threshold parameters in panel threshold model, Hansen (1999) proposed a test statistics to determine the number of thresholds. However, both regressors and error terms are assumed to be cross-sectionally independent so that the bootstrap can be applicable to approximate p-value. In this chapter, we assume both regressors and error terms to be dependent over time and cross-sectionally. The existing bootstrap methods for dependent data may not be valid for approximating p-value anymore. Thus there may be no efficient method to determine the number of thresholds for panel threshold model with cross-sectional dependence so far.

This chapter develops adaptive group fused lasso method for estimating panel threshold model with cross-sectional dependence since fused lasso can penalize the successive difference of slope coefficients and eliminate the extra threshold parameters. We consider consistency and selection consistency results of Lasso estimators. The simulation result show that the proposed Lasso estimation can correctly determine the number of threshold parameter, especially compared with test statistics proposed by Hansen (1999). We apply it to study the weather effect on stock returns.

4.2 Multiple threshold model

Consider the following panel threshold regression model,

$$Y_{it} = \mu_i + \sum_{j=1}^{m+1} \beta_j' X_{it} I(\gamma_{j-1} < q_{it} \leq \gamma_j) + \epsilon_{it},\ i = 1, ..., n, t = 1, ...T \tag{4.2.1}$$

where $\beta_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{jK})' \in \mathbf{R}^K$, $j = 1, ..., m+1$, is a K-dimensional vector, μ_i is the individual effect, $I(\cdot)$ is the indicator function, q_{it} is the threshold variable, the vector $\gamma = (\gamma_1, ..., \gamma_m)$, $-\infty < \gamma_1 \leq \gamma_2 \leq ... \leq \gamma_m < \infty$, is the threshold parameter, which divides the threshold variable process. The error term is assumed to be time dependent and cross-sectionally dependent. Hansen (1999) proposed a test statistic to determine the threshold number by bootstrap method. However, this method may only work for the assumption of identically independent distribution. Thus it may work poor in empirical studies due to the time dependence and cross-sectional dependence in most real panel data. We introduce Fused Lasso method to estimate panel threshold model with cross-sectional dependence in this chapter, where Fused Lasso can shrink the successive difference of coefficients to zero as the threshold parameter does not exist actually.

4.3 Lasso estimation

Differently from Chan et al. (2015) for time series data, in this section, we develop a Group Fused LASSO procedure to be applicable for panel threshold model.

Since the threshold parameter in threshold autoregressive model is a lagged observation, Chan et al. (2015) sorted the thresholds as the the order statistics. By reformulating the threshold autoregressive model as a high dimensional model, group Lasso can be applied to estimate the potential thresholds. However, according to the properties of order statistics, we cannot arrange all panel data sample from the smallest to the largest simply in cross-sectional direction. We assume that threshold variables, $(q_{11}, q_{12}, ..., q_{1T}, ..., q_{n1}, q_{n2}, ..., q_{nT})$, are observed. We sort the value of threshold variables according to ascending order. Let $\boldsymbol{\rho} = (\rho_{(1)}, \rho_{(2)}, ..., \rho_{(nT)})$ be the ordered threshold parameters from smallest to largest. Consider the following panel threshold model

$$Y_{it} = \mu_i + \sum_{l=1}^{nT} \theta_l' X_{it} I(\rho_{(l-1)} < q_{it} \leq \rho_{(l)}) + \epsilon_{it},\ i = 1, ..., n,\ t = 1, ..., T \tag{4.3.1}$$

where X_{it} is a K-dimensional vector of regressors. θ_l is a K-dimensional vector of unknown coefficients. μ_i is the individual effect. ϵ_{it} is the time and cross-sectional dependent error terms. $I(\cdot)$ is the indicator function. q_{it} is the threshold variable. The vector $\boldsymbol{\rho} = (\rho_{(1)}, ..., \rho_{(nT)})$, $-\infty < \rho_{(1)} < ... < \rho_{(nT)} < \infty$, is the threshold parameter. By convention, $\rho_{(0)} = -\infty$ and $\rho_{(nT+1)} = \infty$. We assume $\{\theta_1, ..., \theta_{nT}\}$ exhibit certain sparsity such that the total number of distinct vectors in the set is given by $m + 1$, which is unknown but typically much smaller than nT. Thus the actual threshold parameter is given by $(\gamma_1, ..., \gamma_m)$.

More specifically, we assume that

$$\theta_l = \beta_j \text{ for } l = L_{j-1}, ..., L_j - 1, \text{ and } j = 1, ..., m+1$$

As for the estimation of threshold parameters, the model (4.2.1) can be written as,

$$Y_{it} = \mu_i + \sum_{j=1}^{m+1} \beta_j' X_{it}(\gamma_j) + \epsilon_{it} \tag{4.3.2}$$

To eliminate the individual effect, we consider the average of model (4.3.2) over time index t,

$$\bar{Y}_i = \mu_i + \sum_{j=1}^{m+1} \beta_j' \bar{X}_i(\gamma_j) + \bar{\epsilon}_i \tag{4.3.3}$$

where $\bar{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, $\bar{X}_i(\gamma_j) = \frac{1}{T}\sum_{t=1}^{T} X_{it}(\gamma_j) = \frac{1}{T}\sum_{t=1}^{T} X_{it} I(\gamma_{j-1} < q_{it} \leq \gamma_j)$ and $\bar{\epsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \epsilon_{it}$. By taking the difference between model (4.3.2) and (4.3.3), the following model can be obtained,

$$Y_{it}^* = \sum_{j=1}^{m+1} \beta_j' X_{it}^*(\gamma_j) + \epsilon_{it}^* \tag{4.3.4}$$

where $Y_{it}^* = Y_{it} - \bar{Y}_i$, $X_{it}^*(\gamma_j) = X_{it}(\gamma_j) - \bar{X}_i(\gamma_j)$ and $\epsilon_{it}^* = \epsilon_{it} - \bar{\epsilon}_i$. Let $X_{it}^*(\gamma) = (X_{it}^*(\gamma_1)', ..., X_{it}^*(\gamma_{m+1})')'$ be the $(m+1)k$-dimensional vector. Let $Y^* = (Y_{11}^*, ..., Y_{1T}^*, ..., Y_{n1}^*, ..., Y_{nT}^*)'$, $X^*(\gamma) = (X_{11}^*(\gamma), ..., X_{1T}^*(\gamma), ..., X_{n1}^*(\gamma), ..., X_{nT}^*(\gamma))'$, and $\epsilon^* = (\epsilon_{11}^*, ..., \epsilon_{1T}^*, ..., \epsilon_{n1}^*, ..., \epsilon_{nT}^*)'$. Using this notation, (4.3.4) is equivalent to

$$Y^* = X^*(\gamma)\beta + \epsilon^*. \tag{4.3.5}$$

For any given threshold parameters, the estimated slope coefficients, $\tilde{\beta}(\gamma)$, are obtained by ordinary least squares. The vector of residuals is

$$\hat{\epsilon}^*(\gamma) = Y^* - X^*(\gamma)\tilde{\beta}(\gamma)$$

Consider the panel data model with single threshold parameter, where $m = 1$,

$$Y_{it} = \mu_i + \beta_1' X_{it} I(q_{it} \leq \gamma_1) + \beta_2' X_{it} I(q_{it} > \gamma_1) + \epsilon_{it}(\gamma_1)$$

The above model can be rewritten as

$$Y_{it} = \mu_i + (\beta_1', \beta_2') Z_{it}(\gamma_1) + \epsilon_{it}(\gamma_1)$$

where $Z_{it}(\gamma_1) = (X_{it}I(q_{it} \leq \gamma_1)', X_{it}I(q_{it} > \gamma_1)')'$ is of $2K$ dimension. After within transformation, the sum of squared errors of single threshold model is

$$S_1(\gamma) = \hat{\epsilon}^*(\gamma_1)'\hat{\epsilon}^*(\gamma_1) = Y^{*'}(I - Z^*(\gamma_1)'(Z^*(\gamma_1)'X^*(\gamma_1))^{-1}Z^*(\gamma_1)')Y^*. \tag{4.3.6}$$

The first threshold parameters can be estimated by minimizing the sum of squared errors as Hansen's estimation. Thus the first estimated threshold parameter, $\hat{\gamma}_1$, is

$$\tilde{\gamma}_1 = \arg\min_{\gamma} S_1(\gamma). \tag{4.3.7}$$

Once $\tilde{\gamma}_1$ is obtained, we have the residual variance defined as

$$\hat{\sigma}^2 = \frac{1}{nT}\hat{\epsilon}^*(\tilde{\gamma}_1)'\hat{\epsilon}^*(\tilde{\gamma}_1) = \frac{1}{nT}S_1(\hat{\gamma}_1). \tag{4.3.8}$$

$S_1(\gamma)$ is the sum of squared errors of single threshold model. In this section, we consider multiple thresholds model. Consider the panel data model with double threshold parameter,

$$Y_{it} = \mu_i + (\beta_1', \beta_2', \beta_3')Z_{it}(\gamma_1, \gamma_2) + \epsilon_{it}(\gamma_1, \gamma_2)$$

where $Z_{it}(\gamma_1, \gamma_2) = (X_{it}I(q_{it} \leq \gamma_1)', X_{it}I(\gamma_1 < q_{it} \leq \gamma_2)', X_{it}I(q_{it} > \gamma_2)')'$ is of $3K$ dimension. Similarly, the sum of squared errors of double thresholds model is defined as

$$S_2(\gamma_2) = \begin{cases} S(\tilde{\gamma}_1, \gamma_2) & \text{if} \quad \hat{\gamma}_1 < \gamma_2 \\ S(\gamma_2, \hat{\gamma}_1) & \text{if} \quad \gamma_2 < \tilde{\gamma}_1 \end{cases} \tag{4.3.9}$$

Therefore, the second threshold is estimated by

$$\tilde{\gamma}_2 = \arg\min_{\gamma_2} S_2(\gamma_2) \tag{4.3.10}$$

This is similar to estimation of the first threshold. First, we estimate the threshold in any given regimes. Then, the estimated threshold is chosen by minimizing the sum of squared errors. The third threshold, the fourth threshold and so on can be estimated by the same method. However, both Bai (1997) and Hansen (1999) showed that the second threshold estimator $\tilde{\gamma}_2$ is asymptotically efficient, but $\tilde{\gamma}_1$ is not. They suggested to fix $\tilde{\gamma}_2$ and estimate γ_1 again to solve the problem.

Thus we propose to estimate $\boldsymbol{\beta} = (\beta_1', ..., \beta_{m+1}')'$ by the following adaptive group fused lasso equation

$$\hat{\boldsymbol{\beta}} = \arg\min \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}^* - \sum_{j=1}^{m+1}\beta_j' X_{it}^*(\gamma_j)\right)^2 + \lambda \sum_{j=1}^{m} w_j \|\beta_{j+1} - \beta_j\| \quad (4.3.11)$$

where λ is a tuning parameter, and w_j is a data-driven weight defined by

$$w_j = \|\tilde{\beta}_{j+1} - \tilde{\beta}_j\|^{-\kappa},\ j = 1, ..., m \quad (4.3.12)$$

$\tilde{\beta}_j$ are preliminary estimates of β_j, and κ is an user-specified positive constant that usually takes value 2 in the literature. We can suppress the dependence of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, ..., \hat{\beta}_m) = (\hat{\beta}_1(\lambda), ..., \hat{\beta}_m(\lambda))$ on λ as long as no confusion arises. Note that $\|\tilde{\beta}_{j+1} - \tilde{\beta}_j\| = 0$, $j \geq 1$, there is no change in the coefficients, $\{\beta_j, j = 1, ..., m+1\}$. The threshold parameters γ_j, $j = 1, ..., m$, can be estimated when $\hat{\beta}_{j+1} - \hat{\beta}_j$ is non-zero. Unlike Chan et al. (2015) assigning the same weight to penalize the successive difference of coefficients, our panel regression allows us to apply adaptive weights $\{w_j\}$ in case that the Lasso is inconsistent for model selection.

For a given solution $\hat{\beta}_j$, the estimated threshold parameters are given by $\hat{\Gamma}_{\hat{m}} = (\hat{\gamma}_1, ..., \hat{\gamma}_{\hat{m}})$, where $\tilde{\gamma}_1 \leq \hat{\gamma}_1 < ... < \hat{\gamma}_{\hat{m}} \leq \tilde{\gamma}_m$ such that $\|\hat{\beta}_{j+1} - \hat{\beta}_j\| \neq 0$ at $\tilde{\gamma}_j = \hat{\gamma}_s$ for some $s \in \{1, ..., \hat{m}\}$, where $\tilde{\gamma}_j, j = 1, ..., m$ are the preliminary estimates of γ. $\hat{\Gamma}_{\hat{m}}$ divide the estimated threshold parameters into $\hat{m}+1$ regimes such that the slope estimators remain constant within each regime. Let $\hat{\gamma}_0 = -\infty$, $\hat{\gamma}_{\hat{m}+1} = \infty$ and $\hat{\beta}(\hat{\gamma}_j) = \hat{\beta}_j$ be the estimated slope coefficient of $X^*(\hat{\gamma}_j)$. Define $\hat{\alpha}_s = \hat{\alpha}(\hat{\gamma}_s) = \hat{\beta}(\hat{\gamma}_s)$ as the estimates of α_s for $s = 1, ..., \hat{m}+1$.

Before determining the number of threshold parameters by Lasso estimation, it is really important to test the existence of threshold effect. Then, we can study Lasso estimation of panel threshold regression. The details of testing the existence of threshold effect has been discussed in Section 3.5. We omit it here.

4.4 Asymptotic properties

4.4.1 Assumptions

Let $\{\gamma_i^0, i = 1, ..., m_0\}$ be the true threshold parameters and β_j^0 be the true parameter vector in the jth regime, $j = 1, ..., m_0 + 1$. Define the following common factor function of X_{it},

$$X_{it} = \Lambda_i c_t + v_{it}, i = 1, ..., n, t = 1, ..., T. \tag{4.4.1}$$

where $c_t = (c_{1t}, ..., c_{pt})'$ is a p-dimensional common factor, Λ_i is a $k \times m$ factor loading matrix. Let $\Delta_{\min} = \min_{1\leq j\leq m_0} \|\beta_{j+1}^0 - \beta_j^0\|$ and $\Delta_{\max} = \max_{1\leq j\leq m_0} \|\beta_{j+1}^0 - \beta_j^0\|$ denote the minimum and maximum jump size, respectively. To discuss the asymptotic properties for the Lasso estimation, we first impose the following assumptions,

Assumption 4.1.

1. This part is the same as Assumption 3.1 (i).
2. This part is the same as Assumption 3.1 (ii).
3. This part is the same as Assumption 3.1 (iii).
4. This part is the same as Assumption 3.2 (i).
5. This part is the same as Assumption 3.2 (ii).
6. This part is the same as Assumption 3.2 (iii).

Assumption 4.2

1. $\Delta_{\max} = O(1)$ and $\sqrt{nT}\Delta_{\min} \to C_\Delta \in (0, \infty]$ as $(n, T) \to \infty$.
2. $\frac{m_0\lambda}{\sqrt{nT}}\Delta_{\min}^{-\kappa} = O(1)$ as $(n, T) \to \infty$.
3. $(nT)^{\frac{\kappa+1}{2}}\lambda \xrightarrow{p} \infty$ as $(n, T) \to \infty$.

Remark 4.1 Assumption 4.1 requires time dependence and cross-sectional dependence of X_{it} and ϵ_{it} similar to Assumption 3.1 and Assumption 3.2. We assume both regressors and error terms to be time dependent by introducing α-mixing sequence and martingale difference sequence similar to Assumption 3.1 (i) and (ii). In Assumption 4.1, we impose cross-sectional dependence by common factor function the same as Assumption 3.2. Some certain conditions of cross-sectional dependence required in the following proof is also similar to Assumption 3.2 (iii). Assumption 4.2 mainly impose conditions on m_0, λ and $\Delta_{\min}$ that mainly help us to study the consistent results of lasso estimators.

4.4.2 Consistency

In this section, we consider the consistency of estimators in two case. One is m fixed, the other is $m < nT$. Since lasso estimator can select the number of threshold parameter by penalizing the successive difference of estimated slope coefficients. We mainly study the consistency of slope coefficient and selection consistency. As for the consistency of threshold parameters, we studied it in Chapter 3 and the result still holds since lasso can not penalize them.

Theorem 4.1 Under Assumptions 4.1 to 4.2 (1) and (2), for fixed m_0, as $n, T \to \infty$, $\sqrt{nT}\|\hat{\beta} - \beta^0\| = O_p(1)$.

Remark 4.2 The above theorem establish the the convergence rate of $\hat{\beta}$ as both n and T tending to infinity but m fixed.

Theorem 4.2 Under Assumptions 4.1 to 4.2 (1) and (3), for $m_0 = m_{0,nT}$, as $n, T \to \infty$, if $\lambda = 2kc_0\sqrt{\frac{logm}{nT}}$ for some $c_0 > 0$, then with some $C > 0$ and probability greater than $1 - C(c_0^2\Delta_{\min}^{-2\kappa}\log m)^{-1-\frac{\delta}{2}}$,

$$\frac{1}{nT}\|X^*(\gamma)(\hat{\beta} - \beta^0)\|^2 \leq 4kc_0\sqrt{\frac{m_0 \log m}{nT}}\Delta_{\min}^{-\kappa}\|\hat{\beta} - \beta^0\|. \tag{4.4.2}$$

The above theorem can be also written as $\|\hat{\beta} - \beta^0\| = O_p\left(\sqrt{\frac{m_0 \log m}{nT}}\right)$.

Remark 4.3 Theorem 4.2 is the consistent results in terms of the prediction error, where m may increase as sample size increase.

Theorem 4.3 Under Assumptions 4.1 to 4.2 (1) and (2), for fixed m_0, as $n, T \to \infty$, $P(\hat{m} = m_0) = 1$.

Remark 4.4 The above theorem show the selection consistency under m fixed as sample size increase.

Theorem 4.4 Under Assumptions 4.1 to 4.2 (1) and (3), for $m_0 = m_{0,nT}$, as $n, T \to \infty$, $P(\hat{m} = m_0) \to 1$.

Remark 4.5 The above theorem gives the selection consistency result where m can increase as sample size increase.

4.5 Simulated examples

4.5.1 Computational algorithm

The block coordinate descent algorithm (BCD) and the least angle regression (LAR) algorithm are two commonly used computation algorithms for the implementation of Lasso. However, the existing literature show that coordinate descent based algorithms have better performance under high dimensional settings (see, e.g. Friedman et al. (2007); Wu and Lange (2008)). BCD algorithm optimize the objective function with respect to each block-coordinate of the variables, while keeping all the other coordinates fixed. In other word, set $f(\beta) = f(\beta_1, ..., \beta_m)$. At the k-th iteration, choose an appropriate $j \in \{1, ..., m\}$ and set

$$\beta_j^{k+1} = \arg\min_{\beta_j \in R} f(\beta_1^{k+1}, ..., \beta_{j-1}^{k+1}, \beta_j, \beta_{j+1}^k, ..., \beta_m^k).$$

By setting $\theta_j = \beta_{j+1} - \beta_j$, $j = 1, ..., m$ and $\theta_0 = \beta_1$, model (4.3.11) can be rewritten as

$$\hat{\theta} = \arg\min \frac{1}{nT} \left\| Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1}) \sum_{l=0}^{j} \theta_l \right\|^2 + \lambda \sum_{j=1}^{m} w_j \|\theta_j\|, \tag{4.5.1}$$

We define the squared error loss as follows,

$$\rho(\theta) = \frac{1}{nT} \left\| Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1}) \sum_{l=0}^{j} \theta_l \right\|^2 .$$

The objective function can be written as

$$Q(\theta) = \rho(\theta) + \lambda \sum_{j=1}^{m} w_j \|\theta_j\|.$$

Recall the Karush-Kuhn-Tucker (KKT) conditions in Lemma 4.3 in Appendix, we have

$$-\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \left(\sum_{j=0}^{m} X^*(\gamma_{j+1})\right)\left(\sum_{l=0}^{j} \hat{\theta}_l\right)\right) + \frac{1}{2} nT\lambda w_p \frac{\hat{\theta}_p}{\|\hat{\theta}_p\|} = 0,$$
$$j = 1, ..., m, \text{ where } \hat{\theta}_p \neq 0, \tag{4.5.2}$$

and

$$\left\| \left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \left(\sum_{j=0}^{m} X^*(\gamma_{j+1})\right)\left(\sum_{l=0}^{j} \hat{\theta}_l\right)\right) \right\| \leq \frac{1}{2} nT\lambda w_p,$$
$$j = 1, ..., m, \text{ where } \hat{\theta}_p = 0, \tag{4.5.3}$$

As we mentioned before, we optimize the objective function with respect to the corresponding block while keeping all but the current parameters corresponding to a group fixed. At the k-th iteration, for the parameter θ_s, $s = 1, ..., m$, the objective function is defined as

$$Q(\theta_1^{k+1}, ..., \theta_{s-1}^{k+1}, \theta_s^k, ...\theta_m^k) = \rho(\theta_1^{k+1}, ..., \theta_{s-1}^{k+1}, \theta_s^k, ...\theta_m^k) + +\lambda \sum_{j=1}^{m} w_j \|\theta_j\|.$$

The block coordinate descent algorithm for adaptive group fused Lasso is as follows,

Algorithm 1 Block Coordinate Descent Algorithm

1: Let $\theta_0 \in R^K$ be an initial parameter vector. Set $s = 0$.
2: **repeat**
3: Increase s by one: $s \leftarrow s + 1$.
Denote by θ_j^s the parameter of block j at the s-th iteration through the block coordinates $j \in \{1, ..., m\}$.
4: If $\left\| \left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right) \left(Y^* - \left(\sum_{j=0}^{p-1} X^*(\gamma_{j+1})\right) \left(\sum_{l=0}^{j} \theta_l^s\right) - \left(\sum_{j=p+1}^{m} X^*(\gamma_{j+1})\right) \left(\sum_{l=0}^{j} \theta_l^{s-1}\right)\right)\right\| \leq \frac{1}{2} nT\lambda w_p$: set $\theta_p^s = 0$,
otherwise $\theta_p^s = \arg\min_{\theta_p^s} Q(\theta_1^{k+1}, ..., \theta_{s-1}^{k+1}, \theta_s^k, ...\theta_m^k)$.
5: **until** numerical convergence

By the block coordinate descent algorithm, we can calculate the estimated parameters $\hat{\theta} = (\hat{\theta}_0, ..., \hat{\theta}_m)$. Thus $\hat{\beta}$ can be calculated by $\hat{\beta}_j = \sum_{l=0}^{j-1} \hat{\theta}_l, j \in \{1, ..., m\}$. If $\|\hat{\beta}_{j+1} - \hat{\beta}_j\| = 0$, which is $\hat{\theta}_{j+1} \neq 0, j \in \{1, ..., m\}$, the threshold parameter γ_j does not exist. The total number of threshold parameter is actually the number of $\hat{\theta}_j \neq 0, j \in \{1, ..., m\}$.

4.5.2 Fused LASSO simulated examples

This section explores the finite sample performance of Fused Lasso estimators with cross-sectional dependence. We compare the threshold effect of Fused Lasso estimators with OLS estimators, too. Here, we consider different cases of the fixed effect and regressors.

Example 4.1. Consider the following model,

$$Y_{it} = \beta_1 X_{it} I(X_{it} \leq \gamma) + \beta_2 X_{it} I(X_{it} > \gamma) + \epsilon_{it},\ i = 1, ..., n, t = 1, ..., T \quad (4.5.4)$$

Let $\epsilon_t = (\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{nt})$, which is an n-dimensional vector. Then $\{\epsilon_t, 1 \leq t \leq T\}$ is generated as a sequence of n-dimensional vector of independent Gaussian variables with zero mean and covariance matrix $(a_{ij})_{n\times n}$, where

$$a_{ij} = 0.8^{|j-i|},\ 1 \leq i, j \leq n. \quad (4.5.5)$$

Since the cross-sectional dependence is considered, the form of covariance matrix is different from the normal way. And it is obvious that we have

$$\begin{aligned} &E(\epsilon_{it}\epsilon_{js}) = 0 \quad for\ 1 \le i,j \le n,\ t \ne s, \\ &E(\epsilon_{it}\epsilon_{jt}) = 0.8^{|j-i|} \quad for\ 1 \le i,j \le n,\ 1 \le t \le T. \end{aligned} \tag{4.5.6}$$

And the above equations imply that the error term is dependent cross-sectionally. In the previous section, the independent variables consist of the common factor c_t and the error term v_{it}, that is

$$X_{it} = c_t + v_{it}, \quad i = 1, ..., n,\ t = 1, ..., T. \tag{4.5.7}$$

where $c_t \sim U(0, 0.2)$ for $1 \le t \le T$. Let $v_t = (v_{1t}, v_{2t}, ..., v_{nt})'$ for $1 \le t \le T$. Consider that $\{v_t, t \ge 1\}$ is a stationary α-mixing sequence, it is generated by the following equation,

$$v_{it} = 0.7v_{i,t-1} + u_{it}, \tag{4.5.8}$$

where u_{it} is generated from N(0,1). It is obvious that v_{it} is dependent over time. By setting $\gamma = 0.15$, we have the following model,

$$y_{it} = \beta_1 x_{it} I(x_{it} \le 0.15) + \beta_2 x_{it} I(x_{it} > 0.15) + \epsilon_{it},\ i = 1, ..., n, t = 1, ..., T \tag{4.5.9}$$

The model (4.5.6) is equivalent to

$$\begin{aligned} y_{it} = \theta_1 x_{it} I(x_{it} \le 0.1) + \theta_2 x_{it} I(0.1 < x_{it} \le 0.15) + \theta_3 x_{it} I(x_{it} > 0.15) + \epsilon_{it}, \\ i = 1, ..., n, t = 1, ..., T \end{aligned} \tag{4.5.10}$$

where $\theta_1 = \theta_2 = \beta_1$ and $\theta_3 = \beta_2$. We set $\beta_1 = 1$ and $\beta_2 = 2$. Thus $\theta_1 = \theta_2 = 1$ and $\theta_3 = 3$. In our experiments, we do the both Fused Lasso and panel regression 500 times. We examine the probability of only one threshold in Fused Lasso regression, which means $\hat{\theta}_1 = \hat{\theta}_2$. Furthermore, we compare means, standard deviation, and mean squared error (MSE) of difference of one regime coefficient, $\hat{\theta}_1 - \hat{\theta}_2$ of fused lasso regression with panel data regression. The MSE of

β is defined as follows,

$$MSE = \frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_1^i - \hat{\theta}_2^i)^2. \tag{4.5.11}$$

where $\hat{\theta}_1^i$ and $\hat{\theta}_2^i$ are the estimators of i-th sample. By giving different values of n and T, we compare the influence of sample size on θ. For setting $n = 20, 40, 60$ and $T = 20, 40, 60$, the simulation results are reported in Table 4.1, Table 4.2, and Table 4.3.

TABLE 4.1: The probability of correct estimation for the number of regimes of Example 4.1

n/T	20	40	60
20	92.2	98.4	99.8
40	96.4	99.8	99.8
60	98.6	99.8	100

TABLE 4.2: Means and SDs of $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.1

n/T		20	40	60
20	LASSO	-0.0793	-0.0159	-0.0021
		(0.2754)	(0.1248)	(0.0479)
	OLS	0.2147	0.0337	0.1051
		(3.4639)	(2.1585)	(1.7801)
40	LASSO	-0.0362	-0.0021	-0.0019
		(0.1886)	(0.0466)	(0.0428)
	OLS	-0.0102	0.0642	0.1148
		(2.4727)	(1.5646)	(1.6066)
60	LASSO	-0.0148	-0.0019	0
		(0.1242)	(0.0424)	(0)
	OLS	0.0157	0.0014	-0.0111
		(1.9420)	(1.4392)	(1.1638)

TABLE 4.3: MSE of the $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.1

n/T		20	40	60
20	LASSO	0.0820	0.0158	0.0023
	OLS	12.0209	4.6510	3.1733
40	LASSO	0.0368	0.0022	0.0018
	OLS	6.1021	2.4473	2.5893
60	LASSO	0.0156	0.0018	0
	OLS	3.7641	2.0671	1.3518

Table 4.1 indicates the probability of correct estimation for only one threshold parameter, which is the percentage of $\hat{\theta}_1 = \hat{\theta}_2$. The results show that the probability of correct estimation grows as the sample size increases generally. However, there is one exception, when $T = 40$, the probability decreases as n increases to 40 from 20. Table 4.2 and Table 4.3 are the results of means, standard deviation and mean squared errors of $\hat{\theta}_1 - \hat{\theta}_2$ separately. We can find that with sample size, either n or T increasing, both standard deviation and mean squared errors of LASSO and OLS estimators will decrease.

Example 4.2. Example 4.1 is the panel threshold regression model without individual effect. By setting the same value of the threshold parameter as Example 4.1, consider the following model with fixed effect,

$$\begin{aligned} Y_{it} = \alpha_i + \beta_1 X_{it} I(X_{it} \leq 0.15) + \beta_2 X_{it} I(X_{it} > 0.15) + \epsilon_{it}, \\ i = 1, ..., n, t = 1, ..., T \end{aligned} \tag{4.5.12}$$

where the setting of covariate variables x_{it} and error terms ϵ_{it} are the same as Example 4.1. They both are generated with cross-sectional dependence. We set the fixed effects, $\alpha_i = \frac{1}{T} \sum_{t=1}^{T} x_{it}$ for $i = 1, ..., n-1$, and $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$. The fixed effects satisfy that $\sum_{i=1}^{n} \alpha_i = 0$ for model identifiability. Thus we have the following panel threshold model to estimate,

$$\begin{aligned} y_{it} =& \alpha_i + \theta_1 x_{it} I(x_{it} \leq 0.1) + \theta_2 x_{it} I(1 < x_{it} \leq 0.15) \\ &+ \theta_3 x_{it} I(x_{it} > 0.15) + \epsilon_{it}, \ i = 1, ..., n, t = 1, ..., T \end{aligned} \tag{4.5.13}$$

TABLE 4.4: The probability of correct estimation for the number of regimes of Example 4.2

n/T	20	40	60
20	89.8	96.6	99
40	90.2	99.8	100
60	100	100	100

Table 4.4 show that as the sample size increases, fused lasso can estimate the number of regimes 100 percent correctly. Thus there is no bias between one regime parameters, that is $\hat{\theta}_1 = \hat{\theta}_2$. Other conclusions are similar to Example 4.1, either cross-section sample size or time series sample size increases, both LASSO and OLS estimated SDs and MSE of difference of one regime coefficients

TABLE 4.5: Means and SDs of $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.2

n/T		20	40	60
20	LASSO	-0.0984	-0.0325	-0.0089
		(0.2969)	(0.1739)	(0.0890)
	OLS	-0.1841	-0.1142	-0.1667
		(2.8967)	(2.3839)	(2.0764)
40	LASSO	-0.0944	-0.0023	0
		(0.2885)	(0.0521)	(0)
	OLS	-0.1314	0.0129	-0.1221
		(2.6471)	(1.8640)	(1.4512)
60	LASSO	0	0	0
		(0)	(0)	(0)
	OLS	-0.0382	0.0685	-0.0030
		(1.8715)	(1.3490)	(1.0762)

TABLE 4.6: MSE of $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.2

n/T		20	40	60
20	LASSO	0.0976	0.0312	0.0080
	OLS	8.4079	5.6848	4.3307
40	LASSO	0.0920	0.0027	0
	OLS	7.0106	3.4676	2.1167
60	LASSO	0	0	0
	OLS	3.4970	1.8207	1.1559

will decrease.

Example 4.3. Consider the model in Example 4.2,

$$y_{it} = \alpha_i + \beta_1 x_{it} I(x_{it} \leq 0.15) + \beta_2 x_{it} I(x_{it} > 0.15) + \epsilon_{it}, \quad i = 1, ..., n, t = 1, ..., T \tag{4.5.14}$$

The error term ϵ_{it} is generated as Example 4.1 with time dependence and cross-sectional dependence. The regressor is in form of $x_{it} = c_t + v_{it}$. Differently, c_t is generated as the following AR(1) model,

$$c_t = 0.1c_{t-1} + e_t, \ 1 \leq t \leq T \tag{4.5.15}$$

where AR(1) parameter is 0.1. There exists weak time dependence in c_t. $e_t \sim N(0, 1)$ is generated by normal distribution. The setting of v_{it} is the same as Example 4.1. Thus the model need to be estimated is (4.5.10).

TABLE 4.7: The probability of correct estimation for the number of regimes of Example 4.3

n/T	20	40	60
20	86.4	98	99.6
40	97.4	100	100
60	97.4	100	100

TABLE 4.8: Means and SDs of $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.3

n/T		20	40	60
20	LASSO	-0.1338	-0.0200	-0.0036
		(0.3425)	(0.1406)	(0.0572)
	OLS	-0.2059	-0.0055	0.01590
		(4.7789)	(2.6922)	(2.2886)
40	LASSO	-0.0250	0	0
		(0.1551)	(0)	(0)
	OLS	0.0039	0.0064	-0.0299
		(2.8844)	(1.5208)	(1.7025)
60	LASSO	-0.0253	0	0
		(0.1558)	(0)	(0)
	OLS	-0.1203	0.0484	0.0415
		(2.6201)	(1.2886)	(1.0986)

TABLE 4.9: MSE of $\hat{\theta}_1 - \hat{\theta}_2$ of Examole 4.3

n/T		20	40	60
20	LASSO	0.1350	0.0201	0.0033
	OLS	22.8349	7.2334	5.2274
40	LASSO	0.0246	0	0
	OLS	8.3030	2.3082	2.8936
60	LASSO	0.0249	0	0
	OLS	6.8656	1.6594	1.2061

The simulation results are similar to Example 4.2. Table 4.7 indicates that the probability of correct estimation can be 100 percent as sample size increases. Table 4.8 and Table 4.9 show that an increase in either n or T leads to a decrease in standard deviations and mean squared errors of LASSO estimated difference of one regime coefficients. However, this conclusion is inapplicable to $\hat{\theta}_1 - \hat{\theta}_2$ by OLS estimators.

Example 4.4. Example 4.3 consider the weak time dependence of c_t. In this example , we consider a stronger dependence, that is a larger AR(1) parameter to be assumed as follows,

$$c_t = 0.5c_{t-1} + e_t, \ 1 \leq t \leq T \tag{4.5.16}$$

Other settings such as regressors and error terms are the same as Example 4.3. Table 4.10, Table 4.11, and Table 4.12 show the simulation results.

TABLE 4.10: The probability of correct estimation for the number of regimes of Example 4.4

n/T	20	40	60
20	86.8	96.8	100
40	94.6	99.6	100
60	99.4	100	100

TABLE 4.11: Means and SDs of $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.4

n/T		20	40	60
20	LASSO	-0.1221	-0.0287	0
		(0.3197)	(0.1603)	(0)
	OLS	-0.1506	-0.1065	-0.1224
		(4.7072)	(2.9485)	(2.1459)
40	LASSO	-0.0561	-0.0035	0
		(0.2358)	(0.0548)	(0)
	OLS	-0.1456	0.0589	-0.0122
		(3.3866)	(2.1461)	(1.7590)
60	LASSO	-0.0057	0	0
		(0.0732)	(0)	(0)
	OLS	0.0866	-0.1115	-0.0177
		(2.3710)	(1.4002)	(1.5421)

TABLE 4.12: MSE of $\hat{\theta}_1 - \hat{\theta}_2$ of Example 4.4

n/T		20	40	60
20	LASSO	0.1169	0.0265	0
	OLS	22.1361	8.6875	4.6107
40	LASSO	0.0587	0.0030	0
	OLS	11.4672	4.6001	3.0880
60	LASSO	0.0054	0	0
	OLS	5.6179	1.9690	2.3735

Although we assume a stronger dependence in this example, we can get a similar conclusion as Example 4.3. With sample size increasing, the probability of correct estimation of fused lasso can be 100 percent. The SDs and MSE of OLS estimated difference may not reduce as the sample size increases. But an increase in either n or T results in a decrease in those of LASSO estimated difference.

4.5.3 Comparison to Hansen's test statistics

In real data examples, the actual number of threshold parameters is unknown. Therefore, it is important to determine the number of threshold in empirical studies. Hansen (1999) proposed a test statistic to determine the number of thresholds. First, he introduced F_1 as a test of no thresholds against one threshold. It is defined as

$$F_1 = \frac{S_0 - S_1(\hat{\gamma}_1)}{\hat{\sigma}^2}. \tag{4.5.17}$$

where S_0 is the sum of squared errors of panel data model without thresholds, $\hat{\sigma}^2$ is defined as (4.3.8). F_1 dominates the χ^2_k distribution. Hansen suggested using bootstrap method to approximate the asymptotic p-value. If F_1 reject the null hypothesis of no threshold, he proposed to take further test to determine the number of threshold between one and two. The statistic, F_2, for the test of one threshold versus two thresholds can be based on

$$F_2 = \frac{S_1(\hat{\gamma}_1) - S_2(\hat{\gamma}_2)}{\hat{\sigma}^2} \tag{4.5.18}$$

where $\hat{\sigma}^2 = S_2(\hat{\gamma}_2)/nT$. If the null hypothesis is rejected, the panel threshold model is in favour of two thresholds. Thus we need further test to determine the threshold number similar to the test statistics, F_1 and F_2.

Hansen generated the bootstrap sample by holding the regressors X_{it} and threshold variables q_{it} fixed. The bootstrap error terms would be drawn from residuals and generated new dependent variables. The error term is assumed to be time and cross-sectional independent in Hansen's model. Thus the traditional bootstrap method can be applicable to the residuals. However, we assume both regressors and error terms to be time and cross-sectional dependent. General bootstrap methods cannot be applied to generate samples. Therefore, we consider block bootstrap method for dependent variables.

Kunsch (1989) and Liu, Singh, et al. (1992) introduced moving block bootstrap that can be applicable to weakly dependent stationary observations. Politis and Romano (1994) proposed a new resampling procedure appropriate for stationary weakly dependent time series, the stationary bootstrap, by constructing resampling blocks of random length. Bühlmann and Künsch (1999) proposed a fully data driven method to select block length for block bootstrap and showed that the estimated block length is asymptotically close to the optimum. The following simulated examples show the behaviour of block bootstrap for panel threshold models, especially for determining the number of threshold parameters.

Example 4.5. We consider single threshold model as follows first.

$$y_{it} = \alpha_i + \beta_0 x_{it}^0 + \beta_1 x_{it}^1 I(x_{it}^1 \leq 0) + \beta_2 x_{it}^1 I(x_{it}^1 > 0) + \epsilon_{it}, \quad i = 1, ..., n, t = 1, ..., T \tag{4.5.19}$$

This is similar to the panel threshold model we used in Example 4.4. But we add a regime independent variables, x_{it}^0, which distribute a standard normal $N(0,1)$. Both regime dependent variables and error terms are assumed to be time dependent and cross-sectionally dependent. The fixed effect is generated by $\alpha_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$ for $i = 1, ..., n-1$, and $\alpha_n = -\sum_{i=1}^{n-1} \alpha_i$. By setting $\beta_0 = 1$, $\beta_1 = 1$, and $\beta_2 = 2$, we study the efficiency of test statistics with different sample size. 100 bootstrap replications are used for each test. The test statistics F_1, F_2, F_3 and their P-values are shown in the following table.

Since we consider a single threshold model, if the test can determine the number of threshold efficiently, the test statistics, F_2, must reject the null hypothesis. However, Table 4.13 shows that even if the sample size increase, it is possible that the test cannot determine the true number of threshold. There exist not only one threshold in most empirical studies. Thus we need to study the efficiency of the test for double and triple threshold models.

Example 4.6. Consider the double threshold model as follows,

$$\begin{aligned} y_{it} =& \alpha_i + \beta_0 x_{it}^0 + \beta_1 x_{it}^1 I(x_{it}^1 \leq 0) + \beta_2 x_{it}^1 I(0 < x_{it}^1 \leq 1) \\ &+ \beta_3 x_{it}^1 I(x_{it}^1 > 1) + \epsilon_{it}, \ i = 1, ..., n, t = 1, ..., T \end{aligned} \tag{4.5.20}$$

where $\beta_3 = -1$, other settings keep same as Example 4.5.

Example 4.7. Consider the triple threshold model as follows,

$$\begin{aligned} y_{it} =& \alpha_i + \beta_0 x_{it}^0 + \beta_1 x_{it}^1 I(x_{it}^1 \leq -1) + \beta_2 x_{it}^1 I(-1 < x_{it}^1 \leq 0) \\ &+ \beta_3 x_{it}^1 I(0 < x_{it}^1 \leq 1) + \beta_4 x_{it}^1 I(x_{it}^1 > 1) + \epsilon_{it},\ i = 1, ..., n, t = 1, ..., T \end{aligned} \tag{4.5.21}$$

where $\beta_0 = 1$, $\beta_1 = -2$, $\beta_2 = 1$, $\beta_3 = 2$, and $\beta_4 = -1$. The regressors and error terms are generated same as Example 4.5. The simulated results of tests for double and triple threshold models are shown in Table 4.14 and Table 4.15 separately. The simulated result in Table 4.13 indicated that although the accuracy of determine the number of threshold is not 100%, it can determine one threshold for single threshold model. However, as the number of threshold increases, the test fails to determine the threshold number. It may be due to the bootstrap method. Although we consider the dependence among residuals and use the block bootstrap method, the resampling residuals still cannot show the real correlated relationship. Thus we cannot use the bootstrap to build the test statistics, thereby determining the number of threshold for panel data model with time and cross-sectional dependence.

The simulated results of test statistics of Hansen (1999) with block bootstrap showed that it can determine the number of threshold in single threshold model. But the test is no longer valid for multiple thresholds model. Actually, there often exist more than one thresholds in empirical studies. Thus it is necessary to find an efficient method to determine the number of thresholds for multiple thresholds model. This paper developed fused lasso for estimating the unknown number of thresholds for panel threshold model with multiple regimes. The efficiency of this method for single threshold model has been presented in section 3.1. The following examples shows fused lasso can detect the correct number of threshold for multiple threshold model.

Example 4.8. Consider the following double thresholds model,

$$\begin{aligned} y_{it} = \alpha_i + \beta_1 x_{it} I(x_{it} \leq -1) + \beta_2 x_{it} I(-1 < x_{it} \leq 1) + \beta_3 x_{it} I(x_{it} > 1) + \epsilon_{it}, \\ i = 1, ..., n, t = 1, ..., T \end{aligned} \tag{4.5.22}$$

where the regressors, error terms, and fixed effects are generated as Example 4.4. By setting $\beta_1 = 1$, $\beta_2 = 2$, and $\beta_3 = -1$, we compare the above model with the following model by adding an additional threshold,

$$
\begin{aligned}
y_{it} =& \alpha_i + \theta_1 x_{it} I(x_{it} \leq -1) + \theta_2 x_{it} I(-1 < x_{it} \leq 0) + \theta_3 x_{it} I(0 < x_{it} \leq 1) \\
& + \theta_4 x_{it} I(x_{it} > 1) + \epsilon_{it},\ i = 1, ..., n, t = 1, ..., T.
\end{aligned}
\tag{4.5.23}
$$

If the fused lasso can determine the correct number of threshold, θ_2 should be equal to θ_3. We do the regression 500 times. The results of simulated example 4.8 are shown in Table 4.16 and Table 4.17. In both tables, the results of estimated Lasso estimator of θ_2 and θ_3 are the same, which means the threshold parameter, 0, is eliminated by Lasso estimation. Although there only is small difference between the OLS estimators of θ_2 and θ_3, it still cannot estimate the true model since the extra threshold parameters is estimated by OLS. Thus the simulation result show that Lasso estimation can estimate the true number of thresholds parameters under the dependent conditions. It also shows that as either n or T increases, the standard deviation and mean square error will decrease.

Example 4.9. Consider the following triple thresholds model,

$$
\begin{aligned}
y_{it} =& \alpha_i + \beta_1 x_{it} I(x_{it} \leq -2) + \beta_2 x_{it} I(-2 < x_{it} \leq -1) + \beta_3 x_{it} I(-1 < x_{it} \leq 1) \\
& + \beta_4 x_{it} I(x_{it} > 1) + \epsilon_{it},\ i = 1, ..., n, t = 1, ..., T
\end{aligned}
\tag{4.5.24}
$$

Set $\beta_1 = -2$, $\beta_2 = 1$, $\beta_3 = 2$, and $\beta_4 = -1$. Other settings are same as Example 4.8. Thus we compare (4.5.21) with the following model,

$$
\begin{aligned}
y_{it} =& \alpha_i + \theta_1 x_{it} I(x_{it} \leq -2) + \theta_2 x_{it} I(-2 < x_{it} \leq -1) + \theta_3 x_{it} I(-1 < x_{it} \leq 0) \\
& + \theta_4 x_{it} I(0 < x_{it} \leq 1) + \theta_5 x_{it} I(x_{it} > 1) + \epsilon_{it},\ i = 1, ..., n, t = 1, ..., T.
\end{aligned}
\tag{4.5.25}
$$

The analysis is similar to Example 4.8. If the fused lasso is valid, the value of estimated θ_4 will be equal to θ_5. Table 4.18 and Table 4.19 are the simulated results. Similar to the conclusion of Example 4.8, the simulation result show that compared with OLS, Lasso can determine the true number of threshold parameters.

Generally, we need to consider more than one threshold parameters in real data analysis. In this section, by assuming that both regressors and error terms are dependent over time and cross-sectionally, we compare the behaviour of Hansen's test statistics and Lasso estimation of multiple threshold model. Hansen's test statistics works well in single threshold model. However, it cannot determine the exact number of threshold parameters of multiple threshold model, which may not be applied widely in real data analysis. By studying the simulated examples of multiple threshold model, it is obvious that Fused Lasso can eliminate extra threshold parameter efficiently for both single and multiple threshold models and help to deal with the strong dependence in most climate and finance data.

4.6 Empirical study

We studied the rainfall effect on the stock returns in Chapter 3 by panel threshold regression with cross-sectional dependence. We concluded that single threshold model is the most efficient one by AIC analysis. However, we propose a more advanced method in this chapter, Group Fused Adaptive Lasso, to determine the number of threshold parameters. Thus, in this section, we apply Lasso method to study the precipitation effect on stock market. Given the collected data in Chapter 3 we first consider the follow panel threshold model with 5 threshold parameters,

$$\begin{aligned} R_{it} =& \alpha_i + \beta_1 P_t I(P_t \le \gamma_1) + \beta_2 P_t I(\gamma_1 < P_t \le \gamma_2) + \beta_3 P_t I(\gamma_2 < P_t \le \gamma_3) \\ &+ \beta_4 P_t I(\gamma_3 < P_t \le \gamma_4) + \beta_5 P_t I(\gamma_4 < P_t \le \gamma_5) + \beta_6 P_t I(P_t > \gamma_5) + \epsilon_{it}. \end{aligned} \tag{4.6.1}$$

Table 4.20 gives the estimation results of the above model (4.6.1). The estimation results show that the first regime is not significant statistically. But the coefficients of the remaining regimes are significant. The second regime show that the precipitation between 0.21 inch and 0.25 inch has a positive effect on stock returns. The precipitation between 0.25 inch an 0.74 inch, the third and fourth regime, has a negative effect on stock returns. The precipitation between 0.74 inch and 1.01 inch has a positive effect on stock returns. The last regime, precipitation greater than 1.01 inch, has a negative effect on stock returns. Generally, information criterion analysis and cross-validation are the most widely

TABLE 4.13: Test for threshold effects of single threshold model

	n=20, T=20			n=30, T=30			n=40, T=40		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
statistics	197.9960	4.7849	73.4298	278.0257	6.8212	33.2757	553.2182	5.0519	7.8933
p-value	(0.00)	(0.58)	(0.00)	(0.00)	(0.35)	(0.00)	(0.00)	(0.54)	(0.06)
statistics	155.6048	20.2366	66.7947	246.0012	5.3056	10.4539	592.7134	5.8963	294.2330
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.45)	(0.02)	(0.00)	(0.31)	(0.00)
statistics	103.2943	9.0731	18.1472	183.5780	3.5065	7.7571	732.3135	12.8162	17.0332
p-value	(0.00)	(0.11)	(0.00)	(0.00)	(0.87)	(0.08)	(0.00)	(0.00)	(0.00)
statistics	170.9439	4.8687	149.6452	273.4963	3.8918	14.3474	468.0108	11.0043	18.9517
p-value	(0.00)	(0.53)	(0.00)	(0.00)	(0.78)	(0.00)	(0.00)	(0.07)	(0.00)
statistics	78.4446	4.1842	37.3908	488.1686	5.9901	8.8621	576.9287	5.7346	11.1936
p-value	(0.00)	(0.58)	(0.00)	(0.00)	(0.50)	(0.03)	(0.00)	(0.53)	(0.00)
statistics	100.2746	10.7235	13.4597	391.9444	7.7663	15.9286	628.9040	4.5075	360.4989
p-value	(0.00)	(0.05)	(0.00)	(0.00)	(0.22)	(0.00)	(0.00)	(0.74)	(0.00)
statistics	93.8218	6.2072	60.5965	508.7347	6.7595	152.4009	455.1636	7.8346	260.4559
p-value	(0.00)	(0.43)	(0.00)	(0.00)	(0.38)	(0.00)	(0.00)	(0.24)	(0.00)
statistics	117.5125	12.6293	20.2971	392.3973	4.4789	7.0209	626.0193	5.2244	516.0557
p-value	(0.00)	(0.05)	(0.00)	(0.00)	(0.72)	(0.15)	(0.00)	(0.54)	(0.00)
statistics	81.0393	7.3301	30.2410	224.5868	6.4298	10.1372	530.0942	8.1004	12.8980
p-value	(0.00)	(0.33)	(0.00)	(0.00)	(0.31)	(0.01)	(0.00)	(0.16)	(0.01)
statistics	121.2494	8.5759	20.8064	336.3827	10.4472	17.3184	481.0263	4.8584	86.3880
p-value	(0.00)	(0.19)	(0.00)	(0.00)	(0.06)	(0.00)	(0.00)	(0.66)	(0.00)

TABLE 4.14: Test for threshold effects of double threshold model

	n=20, T=20				n=30, T=30				n=40, T=40			
	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4	F_1	F_2	F_3	F_4
statistics	655.8599	30.7979	36.6336	477.7828	1672.3802	27.9443	31.2669	36.1338	3869.4105	79.2567	83.4617	87.8542
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	915.5168	16.9715	26.6338	79.9662	1885.1604	27.8883	33.7176	39.6029	4022.2166	48.3285	52.4870	818.3584
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	594.5003	22.8015	29.2477	151.2386	1692.6500	41.9139	54.9432	57.3459	3578.2716	53.4376	58.9055	65.3595
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	812.9696	14.9534	17.8650	19.1814	1932.4338	44.4582	51.0245	1519.9622	4138.5488	53.6464	66.6236	2709.7812
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	545.8291	26.2365	31.5928	36.4452	1519.9106	41.7070	59.2048	72.4361	4434.3102	43.72900	48.0194	50.5245
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	685.8144	18.7506	25.2742	31.7638	2033.0457	15.1944	17.9092	1423.5505	4185.5044	88.3202	93.5478	98.1278
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	888.7379	8.8169	13.7131	539.5233	1779.1624	14.2859	17.7634	1252.1275	4567.7659	71.4711	77.1347	2399.7268
p-value	(0.00)	(0.15)	(0.01)	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	726.1951	14.8378	18.7730	21.3619	1579.0954	28.7214	32.6092	36.0902	3572.5467	46.2658	65.2261	69.0565
p-value	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	727.9611	23.6915	37.2317	639.3814	1552.8694	47.4114	51.1834	55.8926	3476.3382	84.4096	95.6926	101.8168
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	787.0069	17.4945	26.8421	36.1253	1695.8477	42.1799	50.8326	1459.4025	3929.0232	25.2946	31.8998	2238.0639
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

TABLE 4.15: Test for threshold effects of triple threshold model

	n=20, T=20					n=30, T=30					n=40, T=40				
	F_1	F_2	F_3	F_4	F_5	F_1	F_2	F_3	F_4	F_5	F_1	F_2	F_3	F_4	F_5
statistics	220.4745	297.2959	688.4299	701.8433	698.1071	237.1343	621.0820	1065.9609	1072.3129	1079.5682	512.0453	1140.0025	2002.4969	2011.5562	2019.1916
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	142.6580	330.0808	615.0826	654.6282	686.3672	221.2613	675.3850	1094.7664	1125.0202	1120.8466	604.6167	1144.9612	2092.1174	2105.8848	2122.6268
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	207.5364	229.9941	548.0721	571.7887	584.0700	225.1180	524.6083	890.0206	900.9535	905.9683	558.2695	1158.3622	2072.4681	2092.2363	2104.5002
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	131.0053	217.5679	465.1120	475.7119	479.2420	205.6846	860.0917	1312.5259	1327.8268	1337.0866	780.1314	889.5715	1933.7885	1943.1266	2009.5017
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	149.3046	296.1959	551.4875	583.3553	589.8112	162.1122	549.1850	851.3916	861.0965	865.6439	550.0501	1144.6913	2002.3406	2024.2119	2039.6499
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	142.9777	359.7328	638.0548	647.7928	654.7026	223.7660	728.3839	1167.9914	1175.4254	1169.9159	647.880	1285.516	2358.494	2370.892	2387.912
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	127.8499	298.4656	525.5880	521.9925	529.4169	240.2430	655.0768	1133.0574	1159.4859	1139.3671	478.5152	1051.9459	1772.3571	1764.4861	1772.1135
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	106.8088	225.0192	407.4198	415.9253	428.2121	329.8482	717.7683	1359.4463	1371.1129	1384.0696	450.2932	1111.3177	1822.2244	1829.7299	1883.8796
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	161.4966	374.0280	705.9598	721.3819	722.7292	196.2325	878.7935	1279.5685	1294.1426	1304.5137	564.8077	1232.7945	2167.3493	2182.3582	2254.9353
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
statistics	154.4894	341.0032	639.1005	648.6007	648.1264	231.9771	953.6621	1468.4888	1493.6921	1504.8105	469.5109	1135.1535	1878.2560	1899.5580	1911.7541
p-value	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

TABLE 4.16: Means and SDs of slope coefficients of Example 4.8

n/T		20		30		40	
		LASSO	OLS	LASSO	OLS	LASSO	OLS
20	$\hat{\theta}_1$	1.0007	0.9864	1.0061	0.9992	1.0069	1.0059
		(0.1107)	(0.1186)	(0.0786)	(0.0891)	(0.0577)	(0.0669)
	$\hat{\theta}_2$	1.6677	1.9868	1.7101	1.9959	1.7112	2.0220
		(0.1581)	(0.2443)	(0.1361)	(0.2297)	(0.1312)	(0.2172)
	$\hat{\theta}_3$	1.6677	1.9961	1.7101	1.9950	1.7112	1.9770
		(0.1581)	(0.2883)	(0.1361)	(0.2406)	(0.1312)	(0.2270)
	$\hat{\theta}_4$	-0.9817	-0.9989	-0.9936	-1.0029	-0.9860	-1.0056
		(0.0808)	(0.0943)	(0.0701)	(0.0786)	(0.0562)	(0.0641)
30	$\hat{\theta}_1$	0.9993	0.9966	1.0196	1.0000	1.0046	0.9992
		(0.0733)	(0.0846)	(0.0648)	(0.0768)	(0.0531)	(0.0589)
	$\hat{\theta}_2$	1.7320	1.9882	1.7304	1.9913	1.7533	1.9889
		(0.1676)	(0.2490)	(0.1181)	(0.2203)	(0.1064)	(0.1788)
	$\hat{\theta}_3$	1.7320	1.9946	1.7304	1.9899	1.7533	1.9993
		(0.1676)	(0.2707)	(0.1181)	(0.1879)	(0.1064)	(0.1775)
	$\hat{\theta}_4$	-0.9888	-1.0048	-0.9942	-1.0021	-0.9939	-0.9994
		(0.0810)	(0.0950)	(0.0589)	(0.0695)	(0.0422)	(0.0481)
40	$\hat{\theta}_1$	0.9996	0.9967	1.0106	1.0013	1.0103	1.0047
		(0.0692)	(0.0774)	(0.0529)	(0.0589)	(0.0469)	(0.0522)
	$\hat{\theta}_2$	1.7722	1.9988	1.7708	1.9972	1.8201	2.0078
		(0.1176)	(0.1961)	(0.1047)	(0.1844)	(0.0798)	(0.1522)
	$\hat{\theta}_3$	1.7722	1.9895	1.7708	1.9978	1.8201	1.9967
		(0.1176)	(0.1928)	(0.1047)	(0.1747)	(0.0798)	(0.1395)
	$\hat{\theta}_4$	-0.9760	-1.0009	-1.0017	-1.0025	-0.9989	-1.0031
		(0.0663)	(0.0753)	(0.0538)	(0.0596)	(0.0443)	(0.0496)

TABLE 4.17: MSE of slope coefficients of Example 4.8

n/T		20		30		40	
		LASSO	OLS	LASSO	OLS	LASSO	OLS
20	$\hat{\theta}_1$	0.0122	0.0142	0.0062	0.0079	0.0034	0.0045
	$\hat{\theta}_2$	0.1354	0.0597	0.1025	0.0527	0.1006	0.0476
	$\hat{\theta}_3$	0.1354	0.0829	0.1025	0.0578	0.1006	0.0520
	$\hat{\theta}_4$	0.0068	0.0089	0.0049	0.0062	0.0033	0.0041
20	$\hat{\theta}_1$	0.0054	0.0071	0.0046	0.0059	0.0028	0.0035
	$\hat{\theta}_2$	0.0999	0.0620	0.0866	0.0485	0.0722	0.0320
	$\hat{\theta}_3$	0.0999	0.0732	0.0866	0.0353	0.0722	0.0314
	$\hat{\theta}_4$	0.0067	0.0090	0.0035	0.0048	0.0018	0.0023
20	$\hat{\theta}_1$	0.0048	0.0060	0.0029	0.0035	0.0023	0.0027
	$\hat{\theta}_2$	0.0657	0.0384	0.0635	0.0339	0.0387	0.0232
	$\hat{\theta}_3$	0.0657	0.0372	0.0635	0.0306	0.0387	0.0194
	$\hat{\theta}_4$	0.0050	0.0057	0.0029	0.0036	0.0020	0.0025

TABLE 4.18: Means and SDs of slope coefficients of Example 4.9

n/T		20		30		40	
		LASSO	OLS	LASSO	OLS	LASSO	OLS
20	$\hat{\theta}_1$	-1.9856	-2.0132	-1.9874	-2.0008	-1.9798	-1.9965
		(0.1121)	(0.1184)	(0.0803)	(0.0888)	(0.0743)	(0.0808)
	$\hat{\theta}_2$	0.9612	0.9831	0.9929	0.9985	0.9972	1.0033
		(0.1533)	(0.1644)	(0.1038)	(0.1219)	(0.0865)	(0.1045)
	$\hat{\theta}_3$	1.6323	1.9845	1.6228	1.9954	1.7154	2.0157
		(0.2306)	(0.2595)	(0.1406)	(0.2395)	(0.1331)	(0.2098)
	$\hat{\theta}_4$	1.6323	1.9983	1.6228	1.9956	1.7154	1.9998
		(0.2306)	(0.2835)	(0.1406)	(0.2391)	(0.1331)	(0.1892)
	$\hat{\theta}_5$	-0.9794	-0.9983	-0.9899	-1.0028	-0.9864	-1.0008
		(0.0825)	(0.0933)	(0.0691)	(0.0767)	(0.0569)	(0.0668)
30	$\hat{\theta}_1$	-1.9918	-2.0035	-1.9939	-2.0045	-1.9859	-1.9966
		(0.0737)	(0.0842)	(0.0601)	(0.0667)	(0.0538)	(0.0602)
	$\hat{\theta}_2$	0.9761	0.9991	0.9921	0.9954	0.9983	1.0016
		(0.1054)	(0.1279)	(0.0839)	(0.1010)	(0.0719)	(0.0842)
	$\hat{\theta}_3$	1.6500	1.9902	1.7250	1.9907	1.7804	2.0039
		(0.1777)	(0.2692)	(0.1204)	(0.1882)	(0.0909)	(0.1643)
	$\hat{\theta}_4$	1.6500	1.9925	1.7250	1.9930	1.7804	2.0149
		(0.1777)	(0.2734)	(0.1204)	(0.2078)	(0.0909)	(0.1647)
	$\hat{\theta}_5$	-0.9822	-1.0054	-0.9879	-1.0001	-0.9909	-1.0009
		(0.0802)	(0.0960)	(0.0607)	(0.0693)	(0.0478)	(0.0545)
40	$\hat{\theta}_1$	-1.9843	-1.9975	-1.9856	-1.9987	-1.9948	-2.0020
		(0.0721)	(0.0804)	(0.0536)	(0.0589)	(0.0445)	(0.0494)
	$\hat{\theta}_2$	0.9903	1.0023	1.0047	0.9983	0.9980	0.9962
		(0.1097)	(0.1288)	(0.0812)	(0.0976)	(0.0657)	(0.0748)
	$\hat{\theta}_3$	1.6869	2.0049	1.7224	1.9950	1.8337	1.9882
		(0.1522)	(0.2351)	(0.1039)	(0.2049)	(0.0856)	(0.1568)
	$\hat{\theta}_4$	1.6869	1.9929	1.7224	2.0000	1.8337	2.0142
		(0.1522)	(0.2259)	(0.1039)	(0.1759)	(0.0856)	(0.1428)
	$\hat{\theta}_5$	-0.9943	-1.0063	-1.0012	-1.0020	-0.9894	-0.9948
		(0.0664)	(0.0777)	(0.0538)	(0.0610)	(0.0464)	(0.0542)

used methods to choose tuning parameter. In this section, we propose to use Bayesian information criterion (BIC) method to choose λ. Schwarz (1978) proposed BIC as follows,

$$BIC = nT\ln(\frac{RSS}{nT}) + k\ln(nT). \tag{4.6.2}$$

By BIC analysis, we choose $\lambda = 48.03045$. Table 4.21 show the Lasso estimation results.

The above table indicate that the first threshold parameter is not essential. Lasso

TABLE 4.19: MSE of slope coefficients of Example 4.9

n/T		20		30		40	
		LASSO	OLS	LASSO	OLS	LASSO	OLS
20	$\hat{\theta}_1$	0.0127	0.0142	0.0066	0.0079	0.0059	0.0065
	$\hat{\theta}_2$	0.0250	0.0272	0.0108	0.0148	0.0075	0.0109
	$\hat{\theta}_3$	0.1883	0.0674	0.1620	0.0573	0.0987	0.0442
	$\hat{\theta}_4$	0.1883	0.0802	0.1620	0.0570	0.0987	0.0357
	$\hat{\theta}_5$	0.0072	0.0087	0.0049	0.0059	0.0034	0.0045
30	$\hat{\theta}_1$	0.0055	0.0071	0.0036	0.0045	0.0031	0.0036
	$\hat{\theta}_2$	0.0117	0.0163	0.0071	0.0102	0.0052	0.0071
	$\hat{\theta}_3$	0.1540	0.0724	0.0901	0.0355	0.0565	0.0270
	$\hat{\theta}_4$	0.1540	0.0746	0.0901	0.0431	0.0565	0.0273
	$\hat{\theta}_5$	0.0067	0.0092	0.0038	0.0048	0.0024	0.0030
40	$\hat{\theta}_1$	0.0054	0.0065	0.0031	0.0035	0.0020	0.0024
	$\hat{\theta}_2$	0.0121	0.0166	0.0066	0.0095	0.0043	0.0056
	$\hat{\theta}_3$	0.1212	0.0552	0.0878	0.0419	0.0350	0.0247
	$\hat{\theta}_4$	0.1212	0.0510	0.0878	0.0309	0.0350	0.0205
	$\hat{\theta}_5$	0.0044	0.0061	0.0029	0.0037	0.0023	0.0030

TABLE 4.20: OLS estimates of the slope and threshold coefficients

	Coefficient estimate	P-value
Threshold γ_1	0.21	
Threshold γ_2	0.25	
Threshold γ_3	0.51	
Threshold γ_4	0.74	
Threshold γ_5	1.01	
$P_t I(P_t \leq 0.21)$	0.307337	0.1103
	(0.192472)	
$P_t I(0.21 < P_t \leq 0.25)$	2.649024	0.0000
	(0.266481)	
$P_t I(0.25 < P_t \leq 0.51)$	-0.945552	0.0000
	(0.100533)	
$P_t I(0.51 < P_t \leq 0.74)$	-0.442112	0.0000
	(0.093353)	
$P_t I(0.74 < P_t \leq 1.01)$	0.550204	0.0000
	(0.086015)	
$P_t I(P_t > 1.01)$	-0.675827	0.0000
	(0.061767)	

estimation penalize the coefficients of the first and second regimes to the same level. That means the precipitation smaller than 0.25 inch has a positive effect

TABLE 4.21: Lasso estimates of the slope and threshold coefficients

	Lasso estimate	OLS estimate
$P_t I(P_t \leq 0.21)$	0.7050924	0.307337
$P_t I(0.21 < P_t \leq 0.25)$	0.7050924	2.649024
$P_t I(0.25 < P_t \leq 0.51)$	-0.6531067	-0.945552
$P_t I(0.51 < P_t \leq 0.74)$	-0.4173103	-0.442112
$P_t I(0.74 < P_t \leq 1.01)$	0.3537766	0.550204
$P_t I(P_t > 1.01)$	-0.6101472	-0.675827

on stock returns. The effect of precipitation of other levels are similar to the conclusion we obtain from OLS estimates.

4.7 Conclusion

In this chapter, we consider panel threshold model via adaptive group fused Lasso. In the model, we assume both regressors and error terms are dependent over time and cross-sectionally. We study consistency and selection consistency in two cases: one is the number of threshold parameters is fixed as sample size increases, the other is the number of threshold parameters increases as sample size increases. We show that Lasso estimators are consistent in both cases. In the meanwhile, the probability of the estimated number of threshold parameters equal to the true number of threshold parameter is tending to 1 as both n and T tending to infinity.

We also study the simulation of Lasso estimation of panel threshold regression model where both regressors and error terms are assumed to be time dependent and cross-sectional dependent. We show that Lasso estimation can determine the number of threshold parameter efficiently. Hansen (1999) proposed to use a test statistic to determine the number of threshold parameters. In simulation, we study the test statistic under the condition that both regressors and error terms are dependent over time and cross-sectionally. The results show that the test statistic may be not efficient under dependent conditions. We apply the Lasso estimation to study the rainfall effect on the stock market in empirical study. In this application, we propose to use BIC choosing the tuning parameter and determine the number of threshold parameters in the multiple threshold model for studying the threshold effect of precipitation on stock market.

4.8 Appendix

Lemma 4.1 Suppose Assumptions 4.1 hold. Then for fixed m_0, as $n, T \to \infty$, $\tilde{\beta}_j - \beta_j^0 = O_p((nT)^{-\frac{1}{2}})$for each $j = 1, ..., m$, where $\tilde{\beta}$ is the ordinary least square estimators.

Proof of Lemma 4.1. The proof is an extended version of that of Lemma B.1 in Qian and Su (2016) with cross-sectional dependence carefully considered. Let $Lambda_{nT}(\beta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}^{*} - \sum_{j=1}^{m+1}\beta_j' X_{it}^{*}(\gamma_j)\right)^2$. Let $\beta_j = \beta_j^0 + \frac{1}{\sqrt{nT}} b_j$ for $j = 1, ..., m+1$ and $\mathbf{b} = (b_1', ..., b_{m+1}')'$. Note that $\boldsymbol{\beta} = \boldsymbol{\beta}^0 + \frac{1}{\sqrt{nT}}\mathbf{b}$. Let $\tilde{b}_j = \sqrt{nT}(\tilde{\beta}_j - \beta_j^0)$ and $\tilde{\mathbf{b}} = (\tilde{b}_1', ..., \tilde{b}_T')$, we have

$$
\begin{aligned}
&\Lambda_{nT}(\beta) - \Lambda_{nT}(\beta^0) \\
&= \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\left(\epsilon_{it}^{*} - \frac{1}{\sqrt{nT}}\sum_{j=1}^{m+1} b_j' X_{it}^{*}(\gamma_j)\right)^2 - \epsilon_{it}^{*2}\right] \\
&= \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{j=1}^{m+1} b_j' X_{it}^{*}(\gamma_j)\right)^2 - \frac{2}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{m+1} b_j' X_{it}^{*}(\gamma_j)\epsilon_{it}^{*} \\
&= \mathbf{b}'\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^{*}(\gamma) X_{it}^{*}(\gamma)'\right)\mathbf{b} - \mathbf{b}'\left(\frac{2}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^{*}(\gamma)\epsilon_{it}^{*}\right). \qquad (4.8.1)
\end{aligned}
$$

By the central limit theorem of α-mixing sequence (e.g. Fan and Yao (2003)), we can prove that

$$\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^{*}(\gamma)\epsilon_{it}^{*} = o_p(\sqrt{nT}). \qquad (4.8.2)$$

This has been proved in Theorem 3.1 in Chapter 3.

Thus we have, as both $n \to \infty$ and $T \to \infty$,

$$\Lambda_{nT}(\beta) - \Lambda_{nT}(\beta^0) = \mathbf{b}'\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^{*}(\gamma) X_{it}^{*}(\gamma)'\right)\mathbf{b} > 0 \qquad (4.8.3)$$

Let $\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^{*}(\gamma)' X_{it}^{*}(\gamma) = \Sigma_\gamma$. We use $\mu_{\max}(\Sigma_\gamma)$ and $\mu_{\min}(\Sigma_\gamma)$ to denote its largest and smallest eigenvalues. Therefore,

$$\Lambda_{nT}(\beta) - \Lambda_{nT}(\beta^0) = \mathbf{b}'\Sigma_\gamma \mathbf{b} \geq \lambda_{\min}(\tilde{\Sigma}_\gamma)\|\mathbf{b}\|^2 > 0 \qquad (4.8.4)$$

If $\|\mathbf{b}\|$ is sufficiently large, $\Lambda_{nT}(\beta)$ cannot be minimized in this case. This further implies that $\|\tilde{\mathbf{b}}\|$ must be stochastically bounded.

Theorem 4.1 establish the consistent result of $\hat{\beta}$ as $(n, T) \to \infty$, but m fixed. The convergence rate is root-nT.

Proof of Theorem 4.1. The proof is an extended version of that of Theorem 3.2 (i) in Qian and Su (2016) with cross-sectional dependence carefully considered. Define

$$\Lambda_{nT,\lambda}(\beta) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}^* - \sum_{j=1}^{m+1}\beta_j' X_{it}^*(\gamma_j)\right)^2 + \lambda\sum_{j=1}^{m} w_j\|\beta_{j+1} - \beta_j\|. \tag{4.8.5}$$

Let $\beta_j = \beta_j^0 + \frac{1}{\sqrt{nT}}b_j$, for $j = 1, ..., m$ and $\mathbf{b} = (b_1', ..., b_{m+1}')'$. Note that $\beta = \beta_0 + \frac{1}{\sqrt{nT}}\mathbf{b}$. Let $\hat{b}_j = \sqrt{nT}(\hat{\beta}_j - \beta_j^0)$ and $\hat{\mathbf{b}} = \sqrt{nT}(\hat{\beta} - \beta^0)$. Note that $Y_{it}^* = \sum_{j=1}^{m+1}\beta_j^{0'} X_{it}(\gamma_j)^* + \epsilon_{it}^*$, we have

$$\begin{aligned}
&\Lambda_{nT,\lambda}(\hat{\beta}) - \Lambda_{nT,\lambda}(\beta^0) \\
&-\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}^* - \sum_{j=1}^{m+1}\beta_j^{0'} X_{it}^*(\gamma_j)\right)^2 - \lambda\sum_{j=1}^{m} w_j\|\beta_{j+1}^0 - \beta_j^0\| \\
=&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{j=1}^{m+1}(\beta_j^0 - \hat{\beta}_j)' X_{it}^*(\gamma_j)\right)^2 + \frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{m+1}(\beta_j^0 - \hat{\beta}_j)' X_{it}^*(\gamma_j)\epsilon_{it}^* \\
&+\lambda\sum_{j=1}^{m} w_j\|\hat{\beta}_{j+1} - \hat{\beta}_j\| - \lambda\sum_{j=1}^{m} w_j\|\beta_{j+1}^0 - \beta_j^0\| \\
=&\mathbf{b}'\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^*(\gamma)X_{it}^*(\gamma)'\right)\mathbf{b} - \mathbf{b}'\left(\frac{2}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^*(\gamma)\epsilon_{it}^*\right) \\
&+\lambda\sum_{j=1}^{m} w_j\|\hat{\beta}_{j+1} - \hat{\beta}_j\| - \lambda\sum_{j=1}^{m} w_j\|\beta_{j+1}^0 - \beta_j^0\|.
\end{aligned} \tag{4.8.6}$$

By the central limit theorem of α-mixing sequence, we can prove that

$$\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^*(\gamma_j)\epsilon_{it}^* = o_p(\sqrt{nT}). \tag{4.8.7}$$

Let $\frac{1}{nT}X^*(\gamma)'X^*(\gamma) \rightarrow \Sigma_\gamma$. Let $\hat{\beta}_j = \beta_j^0 + \frac{1}{\sqrt{nT}}b_j$ for $j = 1, ..., m$ and $\mathbf{b} = (b_1', ..., b_m')'$. Let $\mathcal{A} = \{j : \beta_{j+1}^0 - \beta_j^0 \neq 0\}$, we have

$$\begin{aligned}
&\Lambda_{nT,\lambda}(\hat{\beta}) - \Lambda_{nT,\lambda}(\beta^0) \\
&= \mathbf{b}'\Sigma_\gamma\mathbf{b} + \lambda\sum_{j=1}^{m} w_j[\|\beta_{j+1}^0 - \beta_j^0 + \frac{1}{\sqrt{nT}}(b_{j+1} - b_j)\| - \|\beta_{j+1}^0 - \beta_j^0\|] \\
&= \mathbf{b}'\Sigma_\gamma\mathbf{b} + \lambda\sum_{j\in\mathcal{A}} w_j[\|\beta_{j+1}^0 - \beta_j^0 + \frac{1}{\sqrt{nT}}(b_{j+1} - b_j)\| - \|\beta_{j+1}^0 - \beta_j^0\|] \\
&+ \lambda\sum_{j\in\mathcal{A}^c} w_j\|\frac{1}{\sqrt{nT}}(b_{j+1} - b_j)\| \\
&=: A(1) + A(2) + A(3). \qquad (4.8.8)
\end{aligned}$$

By Lemma 4.1 and Assumption 4.2 (1), $\max_{j\in\mathcal{A}} w_j = \max_{j\in\mathcal{A}} \|\tilde{\beta}_{j+1} - \tilde{\beta}_j\|^{-\kappa} = \max_{j\in\mathcal{A}} \|\beta_{j+1}^0 - \beta_j^0 + O_p(\frac{1}{\sqrt{nT}})\|^{-\kappa} = O_p(\Delta_{\min}^{-\kappa})$. By Jensen, triangle and Cauchy-Schwarz inequalities, and Assumption 4.2 (2),

$$\begin{aligned}
A(2) &\leq \frac{m_0\lambda}{\sqrt{nT}}\max_{s\in\mathcal{A}} w_s\left\{\frac{1}{m_0}\sum_{j\in\mathcal{A}}\|b_{j+1} - b_j\|\right\} \\
&\leq \frac{m_0\lambda}{\sqrt{nT}}\max_{s\in\mathcal{A}} w_s\left\{\frac{1}{m_0}\sum_{j\in\mathcal{A}}\|b_{j+1} - b_j\|^2\right\}^{\frac{1}{2}} \\
&\leq 2\frac{\sqrt{m_0}\lambda}{\sqrt{nT}}\max_{s\in\mathcal{A}} w_s\|\mathbf{b}\| \\
&= O_p(\frac{\sqrt{m_0}\lambda}{\sqrt{nT}}\Delta_{\min}^{-\kappa})\|\mathbf{b}\| \\
&= O_p(1)\|\mathbf{b}\| \qquad (4.8.9)
\end{aligned}$$

In addition, $A(3) > 0$. Thus we have

$$\begin{aligned}
\Lambda_{nT,\lambda}(\hat{\beta}) - \Lambda_{nT,\lambda}(\beta^0) &\geq \mathbf{b}'\Sigma_\gamma\mathbf{b} - O_p(1)\|\mathbf{b}\| \\
&\geq \mu_{\min}\Sigma_\gamma\|\mathbf{b}\|^2 - O_p(1)\|\mathbf{b}\| > 0. \qquad (4.8.10)
\end{aligned}$$

It follows that $\mathbf{b}$ has to be stochastically bounded. Otherwise, $\mathbf{b}$ cannot minimize $\Lambda_{nT,\lambda}(\hat{\beta})$. This implies that $\|\mathbf{b}\| = O_p(1)$ and Theorem 4.1 holds.

Lemma 4.2 Under Assumptions 4.1, for any $c_0 > 0$, there exists some constant

$C > 0$ such that

$$P\left(\max_{1\le j\le m+1}\max_{1\le l\le k}\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^{*l}(\gamma_j)\epsilon_{it}^{*}\right|\ge c_0\sqrt{\frac{\log m}{nT}}\Delta_{\min}^{-\kappa}\right)$$
$$\le C(c_0^2\Delta_{\min}^{-2\kappa}\log m)^{-1-\frac{\delta}{2}}, \tag{4.8.11}$$

where $X_{it}^{*l}(\gamma_j) = X_{it}^{l}I(\gamma_{j-1} < q_{it} \le \gamma_j) - \frac{1}{T}\sum_{t=1}^{T}X_{it}^{l}I(\gamma_{j-1} < q_{it} \le \gamma_j)$ for $j = 1,...,m+1$ and $l = 1,...,k$.

Proof of Lemma 4.2 Following Theorem 4 of Kim (1993) and Assumption 4.1 (2), by Markov inequality, we have

$$\begin{aligned}
&P\left(\max_{1\le j\le m+1}\left|\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^{*l}(\gamma_j)\epsilon_{it}^{*}\right|\ge c_0\sqrt{nT\log m}\Delta_{\min}^{-\kappa}\right)\\
\le\ &\frac{E\left(\max_{1\le j\le m+1}\left|\sum_{i=1}^{n}\sum_{t=1}^{T}X_{it}^{*l}(\gamma_j)\epsilon_{it}^{*}\right|^{2+\delta}\right)}{(c_0\sqrt{nT\log m}\Delta_{\min}^{-\kappa})^{2+\delta}}\\
\le\ &\frac{C(nT)^{1+\frac{\delta}{2}}}{(c_0\sqrt{nT\log m}\Delta_{\min}^{-\kappa})^{2+\delta}}\\
=\ &\frac{C}{(c_0^2\Delta_{\min}^{-2\kappa}\log m)^{1+\frac{\delta}{2}}}.
\end{aligned}\tag{4.8.12}$$

Since k is finite, Lemma 4.2 holds.

Proof of Theorem 4.2 By the definition of $\hat{\beta}$, we have

$$\begin{aligned}
&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}^{*}-\sum_{j=1}^{m+1}\hat{\beta}_j'X_{it}^{*}(\gamma_j)\right)^2+\lambda\sum_{j=1}^{m}w_j\|\hat{\beta}_{j+1}-\hat{\beta}_j\|\\
\le\ &\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(Y_{it}^{*}-\sum_{j=1}^{m+1}\beta_j^{0\prime}X_{it}^{*}(\gamma_j)\right)^2+\lambda\sum_{j=1}^{m}w_j\|\beta_{j+1}^{0}-\beta_j^{0}\|.
\end{aligned}\tag{4.8.13}$$

Let $\mathcal{A}_0 = \{j : \beta_{j+1}^{0}-\beta_j^{0}\neq 0\}$. Then, applying $Y_{it}^{*} = \sum_{j=1}^{m+1}\beta_j^{0\prime}X_{it}(\gamma_j)^{*}+\epsilon_{it}^{*}$ to (4.2.1) gives

$$\begin{aligned}
&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{j=1}^{m+1}(\beta_j^{0}-\hat{\beta}_j)'X_{it}^{*}(\gamma_j)\right)^2\\
\le\ &\frac{2}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j=1}^{m+1}(\hat{\beta}_j-\beta_j^{0})'X_{it}^{*}(\gamma_j)\epsilon_{it}^{*}+\lambda\sum_{j=1}^{m}w_j\|\beta_{j+1}^{0}-\beta_j^{0}\|-\lambda\sum_{j=1}^{m}w_j\|\hat{\beta}_{j+1}-\hat{\beta}_j\|
\end{aligned}$$

$$
\begin{aligned}
&= 2\sum_{j=1}^{m+1}(\hat{\beta}_j - \beta_j^0)'\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^*(\gamma_j)\epsilon_{it}^*\right) + \lambda\sum_{j\in\mathcal{A}_0} w_j(\|\beta_{j+1}^0 - \beta_j^0\| - \|\hat{\beta}_{j+1} - \hat{\beta}_j\|) \\
&\quad - \lambda\sum_{j\in\mathcal{A}_0^c} w_j\|\hat{\beta}_{j+1} - \hat{\beta}_j\| \\
&\le 2k\left(\sum_{j=1}^{m+1}\|\hat{\beta}_j - \beta_j^0\|\right)\left(\max_{1\le j\le m+1}\max_{1\le l\le k}\left|\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} X_{it}^{*l}(\gamma_j)\epsilon_{it}^*\right|\right) \\
&\quad + \lambda\sum_{j\in\mathcal{A}_0} w_j(\|\beta_{j+1}^0 - \beta_j^0\| - \|\hat{\beta}_{j+1} - \hat{\beta}_j\|) - \lambda\sum_{j\in\mathcal{A}_0^c} w_j\|\hat{\beta}_{j+1} - \hat{\beta}_j\|. \qquad (4.8.14)
\end{aligned}
$$

Thus, by Lemma 4.2, we have with probability greater than $1 - C(c_0^2\Delta_{\min}^{-2\kappa}\log m)^{-1-\frac{\delta}{2}}$ that

$$
\begin{aligned}
&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{j=1}^{m+1}(\beta_j^0 - \hat{\beta}_j)'X_{it}^*(\gamma_j)\right)^2 \\
&\le 2kc_0\sqrt{\frac{\log m}{nT}}\Delta_{\min}^{-\kappa}\sum_{j=1}^{m+1}\|\hat{\beta}_j - \beta_j^0\| + \lambda\sum_{j\in\mathcal{A}_0} w_j(\|\beta_{j+1}^0 - \beta_j^0\| - \|\hat{\beta}_{j+1} - \hat{\beta}_j\|) \\
&\quad - \lambda\sum_{j\in\mathcal{A}_0^c} w_j\|\hat{\beta}_{j+1} - \hat{\beta}_j\| \\
&\le \lambda\Delta_{\min}^{-\kappa}\sum_{j=1}^{m+1}\|\hat{\beta}_j - \beta_j^0\| + \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0}(\|\beta_{j+1}^0 - \beta_j^0\| - \|\hat{\beta}_{j+1} - \hat{\beta}_j\|) \\
&\quad - \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0^c}\|\hat{\beta}_{j+1} - \hat{\beta}_j\| \qquad (4.8.15)
\end{aligned}
$$

By the triangle inequality and the Cauchy-Schwarz inequality, we can prove that

$$
\begin{aligned}
&\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\sum_{j=1}^{m+1}(\beta_j^0 - \hat{\beta}_j)'X_{it}^*(\gamma_j)\right)^2 \\
&\le \lambda\Delta_{\min}^{-\kappa}\sum_{j=1}^{m+1}\|\hat{\beta}_j - \beta_j^0\| + \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0}\|\hat{\beta}_j - \beta_j^0\| \\
&\quad - \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0^c}\|(\hat{\beta}_{j+1} - \beta_{j+1}^0) - (\hat{\beta}_j - \beta_j^0)\| \\
&\le \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0}\|\hat{\beta}_j - \beta_j^0\| + \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0^c}\|\hat{\beta}_j - \beta_j^0\| + \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0}\|\hat{\beta}_j - \beta_j^0\| \\
&\quad - \lambda\Delta_{\min}^{-\kappa}\sum_{j\in\mathcal{A}_0^c}\|(\hat{\beta}_{j+1} - \beta_{j+1}^0) - (\hat{\beta}_j - \beta_j^0)\|
\end{aligned}
$$

$$\leq 2\lambda\Delta_{\min}^{-\kappa}\sqrt{m_0}\left(\sum_{j\in\mathcal{A}_0}\|\hat{\beta}_j-\beta_j^0\|^2\right)^{\frac{1}{2}}$$
$$=4kc_0\sqrt{\frac{m_0\log m}{nT}}\Delta_{\min}^{-\kappa}\|\hat{\beta}-\beta^0\|. \tag{4.8.16}$$

Next we study the selection consistency of adaptive group fused Lasso estimators. By setting $\theta_j=\beta_{j+1}-\beta_j$, $j=1,...,m$ and $\theta_0=\beta_1$, model (4.3.11) can be rewritten as

$$\hat{\theta}=\arg\min\frac{1}{nT}\left\|Y^*-\sum_{j=0}^{m}X^*(\gamma_{j+1})\sum_{l=0}^{j}\theta_l\right\|^2+\lambda\sum_{j=1}^{m}w_j\|\theta_j\|, \tag{4.8.17}$$

The following lemma provides the Karush-Kuhn-Tucker (KKT) conditions of the group fused Lasso estimator (e.g. Bühlmann and Van De Geer (2011)).

Lemma 4.3 Let $\hat{\theta}$ and θ be defined as in (4.8.17), we have

$$-\left(\sum_{l=p}^{m}X^*(\gamma_{l+1})'\right)\left(Y^*-\sum_{j=0}^{m}X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right)+\frac{1}{2}nT\lambda w_p\frac{\hat{\theta}_p}{\|\hat{\theta}_p\|}=0,$$
$$j=1,...,m,\ \text{where}\ \hat{\theta}_p\neq 0, \tag{4.8.18}$$

and

$$\left\|\left(\sum_{l=p}^{m}X^*(\gamma_{l+1})'\right)\left(Y^*-\sum_{j=0}^{m}X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right)\right\|\leq\frac{1}{2}nT\lambda w_p,$$
$$j=1,...,m,\ \text{where}\ \hat{\theta}_p=0, \tag{4.8.19}$$

Proof of Lemma 4.3 For the first statement, we invoke subdifferential calculus. Define

$$\Lambda_{nT,\lambda}(\theta)=\arg\min\frac{1}{nT}\left\|Y^*-\sum_{j=0}^{m}X^*(\gamma_{j+1})\sum_{l=0}^{j}\theta_l\right\|^2+\lambda\sum_{j=1}^{m}w_j\|\theta_j\|$$

Thus we have, for $\hat{\theta}_p \neq 0$,

$$\frac{\partial \Lambda_{nT,\lambda}(\hat{\theta})}{\partial \theta_p} = -\frac{2}{nT}\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right) + \lambda w_p \frac{\hat{\theta}_p}{\|\hat{\theta}_p\|} = 0. \tag{4.8.20}$$

On the other hand, if $\hat{\theta}_p = 0$, the subdifferential at $\hat{\theta}$ has to include the zero elements, that is

$$-\frac{2}{nT}\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right) + \lambda w_p e = 0, \text{ for some } e \in \{-1,1\}. \tag{4.8.21}$$

Thus we have

$$\left\|\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right)\right\| \leq \frac{1}{2}nT\lambda w_p. \tag{4.8.22}$$

Proof of Theorem 4.3 We consider two cases: (i) $p \in \mathcal{A}_0$ and (ii) $p \in \mathcal{A}_0^c$. In case (i), for $p \in \hat{\mathcal{A}}$, Theorem 4.1 indicate that $\hat{\theta}_p = \hat{\beta}_{p+1} - \hat{\beta}_p = \beta^0_{p+1} - \beta^0_p + O_p(\frac{1}{\sqrt{nT}}) = \theta^0 + O_p(\frac{1}{\sqrt{nT}})$. Thus by the consistency of $\hat{\theta}_p$, $\sqrt{nT}(\hat{\theta}_p - \theta^0) = O_p(1)$, $P(p \in \mathcal{A}_0) \to 1$. In case (ii), it suffices to show that for $p' \notin \hat{\mathcal{A}}$, $P(p' \in \mathcal{A}_0) \to 0$. Consider the event $p' \in \mathcal{A}_0$. By Lemma 4.3, we have

$$-\left(\sum_{l=p'}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right) + \frac{1}{2}nT\lambda w_{p'}\frac{\hat{\theta}_{p'}}{\|\hat{\theta}_{p'}\|} = 0.$$

There exists $r \in \{1,..,k\}$ such that $\|\hat{\theta}_{p',r}\| = \max\{|\hat{\theta}_{p',q}|, q = 1,...,k\}$, where for any $k \times 1$ vector $a_{p'}$, $a_{p',q}$ denotes its q-th element. Without loss of generality assume that $r = k$, implying that $\frac{|\hat{\theta}_{p',k}|}{\|\hat{\theta}_{p'}\|} \geq \frac{1}{\sqrt{k}}$. In view the fact that $w_{p'} = O_p((nT)^{\frac{\kappa}{2}})$ and Assumption 4.2 (3), we have

$$\frac{1}{2}\sqrt{nT}\lambda w_{p'}\frac{|\hat{\theta}_{p',k}|}{\|\hat{\theta}_{p'}\|} \geq \frac{1}{2}\sqrt{nT}\lambda w_{p'}/\sqrt{k} \to \infty.$$

On the other hand,

$$\frac{\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1})\sum_{l=0}^{j}\hat{\theta}_l\right)}{\sqrt{nT}}$$

$$
\begin{aligned}
&= \frac{\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right) X^*(\gamma)(\beta^0 - \hat{\beta})}{\sqrt{nT}} + \frac{\sum_{l=p'}^{m} X^*(\gamma_{p'+1})'\epsilon^*}{\sqrt{nT}} \\
&= \frac{\left(\sum_{l=p}^{m} X^*(\gamma_{l+1})'\right) X^*(\gamma)\sqrt{nT}(\beta^0 - \hat{\beta})}{nT} + \frac{\sum_{l=p'}^{m} X^*(\gamma_{p'+1})'\epsilon^*}{\sqrt{nT}} \\
&=: B(1) + B(2).
\end{aligned}
$$

By the proof of Theorem 4.1, we have $B(1) = O_p(1)$ and $B(2) = O_p(1)$. Therefore,

$$
\begin{aligned}
&P(p' \in \mathcal{A}_0) \\
\leq\, & P\left(-\left(\sum_{l=p'}^{m} X^*(\gamma_{l+1})'\right)\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_{j+1}) \sum_{l=0}^{j} \hat{\theta}_l\right) + \frac{1}{2} nT\lambda w_{p'} \frac{\hat{\theta}_{p'}}{\|\hat{\theta}_{p'}\|} = 0\right) \to 0.
\end{aligned}
$$

Consequently, we can prove Theorem 4.3.

Proof of Theorem 4.4 In Lemma 4.3, $\sum_{l=p}^{m} X_{it}^*(\gamma_{l+1}) = X_{it}^* I(q_{it} > \gamma_p)$. Setting $X_{it}^*(\gamma_j') = X_{it}^* I(q_{it} > \gamma_j)$ for $j = 1, ..., m$ and $X_{it}^*(\gamma_j') = X_{it}^*$ for $j = 0$, model (4.8.17) can be written as

$$
\hat{\theta} = \arg\min \frac{1}{nT} \left\| Y^* - \sum_{j=0}^{m} X^*(\gamma_j')\theta_j \right\| + \lambda \sum_{j=1}^{m} w_j \|\theta_j\|. \tag{4.8.23}
$$

The Karush-Kuhn-Tucker (KKT) conditions of above model can be written as

$$
-X^*(\gamma_p')'\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_j')\hat{\theta}_j\right) + \frac{1}{2} nT\lambda w_p \frac{\hat{\theta}_p}{\|\hat{\theta}_p\|} = 0, \text{ for } \hat{\theta}_p \neq 0,
$$

and

$$
\left\| X^*(\gamma_p')'\left(Y^* - \sum_{j=0}^{m} X^*(\gamma_j')\hat{\theta}_j\right) \right\| \leq \frac{1}{2} nT\lambda w_p, \text{ for } \hat{\theta}_p = 0.
$$

Let $\zeta = \left(\frac{1}{2} nTw_p \frac{\hat{\theta}_p}{\|\hat{\theta}_p\|}, p \in \mathcal{A}_0\right)$ and $X_1^*(\gamma') = \left(X^*(\gamma_p'), p \in \mathcal{A}_0\right)$, we have

$$
\begin{aligned}
\hat{\theta}_{nT1} =& (X_1^*(\gamma')'X_1^*(\gamma'))^{-1}(X_1^*(\gamma')'Y^* - \lambda\zeta) \\
=& \theta_{01} + (X_1^*(\gamma')'X_1^*(\gamma'))^{-1}(X_1^*(\gamma')'\epsilon^* - \lambda\zeta).
\end{aligned}
$$

Therefore, we need to prove that

$$\hat{\theta}_{nT1} =_{\mathcal{A}} \theta_{01},$$
$$\left\| X^*(\gamma'_p)' \left(Y^* - X_1^*(\gamma')\hat{\theta}_{nT1} \right) \right\| \le \frac{1}{2} nT\lambda w_p, \ \forall p \notin \mathcal{A}_0. \tag{4.8.24}$$

Following $Y^* - X_1^*(\gamma')\hat{\theta}_{nT1} = \epsilon^* - X_1^*(\gamma')(\hat{\theta}_{nT1} - \theta_{01})$, (4.8.24) is equal to

$$\|\theta_p^0 - \hat{\theta}_{nTp}\| \le \|\theta_p^0\|, \forall p \in \mathcal{A}_0$$
$$\left\| X^*(\gamma'_p)' \left(\epsilon^* - X_1^*(\gamma')(\hat{\theta}_{nT1} - \theta_{01}) \right) \right\| \le \frac{1}{2} nT\lambda w_p, \ \forall p \notin \mathcal{A}_0. \tag{4.8.25}$$

Let $H_{nT} = I - X_1^*(\gamma')(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}X_1^*(\gamma')'$, we have

$$\left\| e'_p(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}(X_1^*(\gamma')'\epsilon^* - \lambda\zeta) \right\| \le \|\theta_p^0\|, \forall p \in \mathcal{A}_0$$
$$\left\| X^*(\gamma'_p)'(H_{nT}\epsilon^* + X_1^*(\gamma')(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\lambda\zeta) \right\| \le \frac{1}{2} nT\lambda w_p, \ \forall p \notin \mathcal{A}_0. \tag{4.8.26}$$

where e_p is the unit vector in the direction of the p-th coordinate. When

$$1 - \|X^*(\gamma'_p)'X_1^*(\gamma')(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\zeta\|$$
$$\ge 1 - \|X^*(\gamma'_p)'X_1^*(\gamma')(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\frac{1}{2}nT\lambda\Delta_{\min}^{-\kappa}\zeta_1\| \ge \eta,$$

where $\zeta_1 = (\frac{\hat{\theta}_p}{\|\hat{\theta}_p\|}, p \in \mathcal{A}_0)$ for some $\eta > 0$. Hence

$$\|e'_p(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}X_1^*(\gamma')'\epsilon^*\| + \lambda\|e'_p(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\zeta\| \le \|\theta_p^0\|, p \in \mathcal{A}_0, \forall p \in \mathcal{A}_0$$
$$\|X^*(\gamma'_p)'H_{nT}\epsilon^*\| \le \frac{1}{2}nT\lambda\Delta_{\min}^{-\kappa}\eta, \ \forall p \notin \mathcal{A}_0. \tag{4.8.27}$$

Define

$$\Lambda_1 = \left\{ \max_{p \in \mathcal{A}_0} \left(\|e'_p(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}X_1^*(\gamma')'\epsilon^*\| + \lambda\|e'_p(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\zeta\| \right) \ge \Delta_{\min} \right\}$$
$$\Lambda_2 = \left\{ \max_{p \notin \mathcal{A}_0} \|X^*(\gamma'_p)'H_{nT}\epsilon^*\| \ge \frac{1}{2}nT\lambda\Delta_{\min}^{-\kappa}\eta \right\}$$

To prove $P(\hat{m} = m_0) \to 1$, we need to prove that $P(\Lambda_1) \to 0$ and $P(\Lambda_2) \to 0$. According to the Assumption 4.2 (3) and the proof of Lemma 4.2, we have

$$P(\Lambda_2) = P\left(\max_{p \notin \mathcal{A}_0} \|X^*(\gamma'_p)'H_{nT}\epsilon^*\| \ge \frac{1}{2}nT\lambda\Delta_{\min}^{-\kappa}\eta \right)$$

$$\leq \frac{\max_{p\notin\mathcal{A}_0}\|X^*(\gamma_p')'H_{nT}\epsilon^*\|}{\frac{1}{2}nT\lambda\Delta_{\min}^{-\kappa}\eta}$$
$$\leq \frac{C_1\sqrt{nT}}{\frac{1}{2}nT\lambda\Delta_{\min}^{-\kappa}\eta}\to 0. \tag{4.8.28}$$

Now consider $P(\Lambda_1)$, we have

$$P(\Lambda_1)\leq P\left(\max_{p\in\mathcal{A}_0}\|e_p'(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}X_1^*(\gamma')'\epsilon^*\|\geq\frac{\Delta_{\min}}{2}\right)$$
$$+P\left(\lambda\max_{p\in\mathcal{A}_0}\|e_p'(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\zeta\|\geq\frac{\Delta_{\min}}{2}\right)$$

Denote the smallest eigenvalue of $\frac{X_1^*(\gamma')'X_1^*(\gamma')}{nT}$ by c_1, for the second term of the above function, we have

$$\|e_p'(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}\zeta\|\leq\|e_p\|\cdot\left\|\left(\frac{X_1^*(\gamma')'X_1^*(\gamma')}{nT}\right)^{-1}\right\|\cdot\left\|\frac{1}{2}\Delta_{\min}^{-\kappa}\zeta_1\right\|$$
$$\leq\frac{1}{2}\Delta_{\min}^{-\kappa}c_1^{-1}\sqrt{m_0}. \tag{4.8.29}$$

Thus we want

$$\Delta_{\min}^{\kappa+1}>\lambda c_1^{-1}\sqrt{m_0}. \tag{4.8.30}$$

For the first term, we have,

$$P\left(\max_{p\in\mathcal{A}_0}\|e_p'(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}X_1^*(\gamma')'\epsilon^*\|\geq\frac{\Delta_{\min}}{2}\right)$$
$$\leq P\left(\max_{p\in\mathcal{A}_0}c_1^{-1}\|e_p'X_1^*(\gamma')'\epsilon^*\|\geq\frac{nT\Delta_{\min}}{2}\right)$$
$$\leq\frac{\max_{p\in\mathcal{A}_0}c_1^{-1}\|e_p'X_1^*(\gamma')'\epsilon^*\|}{\frac{1}{2}nT\Delta_{\min}}$$
$$\leq\frac{c_1^{-1}C_2\sqrt{nT}}{\frac{1}{2}nT\Delta_{\min}} \tag{4.8.31}$$

Combining (4.8.30) and Assumption 4.2 (3),

$$P\left(\max_{p\in\mathcal{A}_0}\|e_p'(X_1^*(\gamma')'X_1^*(\gamma'))^{-1}X_1^*(\gamma')'\epsilon^*\|\geq\frac{\Delta_{\min}}{2}\right)\leq\frac{C_2}{\frac{1}{2}\sqrt{nT}\lambda\Delta_{\min}^{-\kappa}\sqrt{m_0}}\to 0. \tag{4.8.32}$$

Therefore, the proof of Theorem 4.4 is completed.

Chapter 5

Conclusion and Outlook

Climate financial analysis has been a hot topic in the past few years. However, only few literature study the weather effect on stock market. Saunders (1993) may be the first to study the relationship between weather and stock returns and conclude that the cloud cover has a negative effect on stock returns. Hirshleifer and Shumway (2003) apply panel time series to study the weather effect on stock returns and obtain a similar conclusion, which sunshine has a positive effect on stock returns. It also point that the rainfall does not have any influence on the stock market. But in this project, we propose to study the effect of weather, saying precipitation, on the stock market. The major contributions of this research are as follows. First, we empirically study that the rainfall has a nonlinear impact on the stock market, significantly. Then, we develop an asymptotic theory of panel threshold model under the assumption of both error terms and regressors dependent cross-sectionally. Finally, we develop adaptive group fused Lasso estimation of panel threshold regression to determine the number of threshold parameters for dependent variables.

In Chapter 2, we have studied that the rainfall has a significant impact on the stock market. More specifically, by using nonparametric method to plot the relationship between precipitation and stock return for each individual stock in FTSE 100, we interestingly find that the relationship is nonlinear. Some change points can be obviously found in the plots. Thus we propose to use threshold time series model with threshold parameter, 0.5, 1.0 and 1.5 in this application. Analysis results of individual stocks show the threshold effects of rainfall on some individual stocks are significant. We then suggest to apply panel threshold regression model to study the rainfall on the whole stock market and conclude that different amounts of rainfall have different significant influence on stock returns. However, due to the correlation of different stocks, panel threshold model proposed by Hansen (1999) may be not good enough in empirical study due to ignoring cross-sectional dependence, which is the research topic in the following theoretical statistical analysis.

Motivated by empirical application in Chapter 2, we develop a statistical theory for threshold panel time series model with cross-sectional dependence in Chapter 3. We develop a lease square estimation to estimate regression coefficients and threshold parameters. Under the dependent conditions of regressors and error terms, we establish the asymptotic properties of regression coefficient and threshold parameters. The asymptotic distribution of regression coefficient is a normal distribution with convergence rate, root-nT. The distribution of threshold parameters is non-standard with convergence rate, $n^{1-\alpha_1}T^{1-\alpha_2}$, where $\alpha_1, \alpha_2 \in [0, \frac{1}{2})$. We expand the asymptotic properties to multiple threshold model. The simulation studies show the the proposed asymptotic theory considering cross-sectional dependence is better than the asymptotic theory ignoring cross-sectional dependence. We apply our method to study the rainfall effect on stock returns and find that the heavy rainfall has a negative effect on stock market. Both simulation and empirical study indicate that ignoring cross0-sectional dependence lead to spurious significance.

A further topic left over by Chapter 3 is how to determine the number of threshold parameters in multiple threshold model. We develop adaptive group fused Lasso estimation for multiple threshold panel time series model with cross-sectional dependence in Chapter 4. We consider two cases in consistency study: one is the number of threshold parameter is fixed as sample size increases, the other is the number of threshold parameter increases as sample size increases. In both cases, the proposed method can correctly determine the number of threshold parameters with probability approaching 1. The simulations are conducted to evaluate the finite sample performance of Lasso estimators. Compared with the test statistics proposed by Hansen (1999), the proposed Lasso estimator can correctly detect the number of threshold parameters for dependent variables. We finally apply our method to study the rainfall effect on stock returns and detect four threshold parameters.

However, there are still quite some questions to be worth investigation following this thesis.

In this project, we have focused on the threshold effects of the exogenous covariates such as climate or weather variable on the response of individual stock returns, leaving other factors as nuisances in the error terms in Chapters 2-4 under cross-sectional dependence. It would be interesting to include, for example, like Miao et al. (2020a), the effects of other factors on the market returns, such as the market factor in the capital asset pricing model or others in multi-factor models, or even to study panel threshold model with interactive fixed effect of latent factors under cross-sectional dependence, by setting the model (1.1.3) with $\epsilon_{it} = \lambda_i^0 f_t^0 + e_{it}$, with estimation of $F^0 = (f_1^0, \cdots, f_T^0)'$ and $\Lambda^0 = (\lambda_1^0, \cdots, \lambda_n^0)'$ further considered; c.f., Miao et al. (2020a).

In addition, spatio-temporal model is one of the most popular models in climate and financial data analysis. In the last decade, the idea of nonlinear spatio-temporal model has been commonly used in real data analysis. A further topic for us is expanding threshold effect to spatio-temporal model considering spatial dependence. In fact, in many empirical examples in the literature (c.f., Lu et al. (2009),Al-Sulami et al. (2017)), the threshold model structure identified by nonparametric estimation can significantly improve the prediction accuracy. Different with the cross-sectional dependence, we may consider spatial dependence by measuring the distance between different spatial locations so that we can deal with strong dependence raised in spatial data.

References

Richard M Adams, Brian H Hurd, Stephanie Lenhart, and Neil Leary. Effects of global climate change on agriculture: an interpretative review. *Climate research*, 11(1):19–30, 1998.

Yacine Aıt-Sahalia and Andrew W Lo. Nonparametric risk management and implied risk aversion. *Journal of econometrics*, 94(1-2):9–51, 2000.

Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*, pages 199–213. Springer, 1998.

D. Al-Sulami, Z. Jiang, Z. Lu, and J. Zhu. Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econometrics and Statistics*, 2:22–35, 2017.

Dawlah Al-Sulami, Zhenyu Jiang, Zudi Lu, and Jun Zhu. On a semiparametric data-driven nonlinear model with penalized spatio-temporal lag interactions. *Journal of Time Series Analysis*, 40(3):327–342, 2019.

Karsten Arnbjerg-Nielsen, P Willems, Jonas Olsson, S Beecham, A Pathirana, I Bülow Gregersen, H Madsen, and V-T-V Nguyen. Impacts of climate change on rainfall extremes and urban drainage systems: a review. *Water science and technology*, 68(1):16–28, 2013.

Goodness C Aye and Prosper Ebruvwiyo Edoja. Effect of economic growth on co2 emission in developing countries: Evidence from a dynamic panel threshold model. *Cogent Economics & Finance*, 5(1):1379239, 2017.

Jushan Bai. Estimating multiple breaks one at a time. *Econometric theory*, 13(3): 315–352, 1997.

Jushan Bai. Panel data models with interactive fixed effects. *Econometrica*, 77(4): 1229–1279, 2009.

Badi Hani Baltagi and Badi H Baltagi. *Econometric analysis of panel data*, volume 4. Springer, 2008.

Ole E Barndorff-Nielsen and Neil Shephard. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001.

Victor L Bernard. Cross-sectional dependence and problems in inference in market-based accounting research. *Journal of Accounting Research*, 25(1):1–48, 1987.

Alexander Bick, Dieter Nautz, et al. Inflation thresholds and relative price variability: evidence from us cities. *International Journal of Central Banking*, 4(3): 61–76, 2008.

Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 1999.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

Jörg Breitung and M Hashem Pesaran. Unit roots and cointegration in panels. In *The econometrics of panel data*, pages 279–322. Springer, 2008.

Trevor S Breusch and Adrian R Pagan. The lagrange multiplier test and its applications to model specification in econometrics. *The review of economic studies*, 47(1):239–253, 1980.

Casey Brown, Robyn Meeks, Yonas Ghile, and Kenneth Hunu. Is water security necessary? an empirical analysis of the effects of climate hazards on national-level economic growth. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(2002):20120416, 2013.

Peter Bühlmann and Hans R Künsch. Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310, 1999.

Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Zongwu Cai and Qi Li. Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, 24(5):1321–1342, 2008.

Melanie Cao and Jason Wei. Stock market returns: A note on temperature anomaly. *Journal of Banking & Finance*, 29(6):1559–1573, 2005.

Kung-Sik Chan. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *The annals of statistics*, 21(1): 520–533, 1993.

Ngai Hang Chan, Chun Yip Yau, and Rong-Mao Zhang. Group lasso for structural break time series. *Journal of the American Statistical Association*, 109(506): 590–599, 2014.

Ngai Hang Chan, Chun Yip Yau, and Rong-Mao Zhang. Lasso estimation of threshold autoregressive models. *Journal of Econometrics*, 189(2):285–296, 2015.

Shao-Chi Chang, Sheng-Syan Chen, Robin K Chou, and Yueh-Hsiang Lin. Weather and intraday patterns in stock returns and trading activity. *Journal of Banking & Finance*, 32(9):1754–1766, 2008.

Tsangyao Chang, Chien-Chung Nieh, Ming Jing Yang, and Tse-Yu Yang. Are stock market returns related to the weather effects? empirical evidence from taiwan. *Physica A: Statistical Mechanics and its Applications*, 364:343–354, 2006.

Lee Chapman. Transport and climate change: a review. *Journal of transport geography*, 15(5):354–367, 2007.

Nidhaleddine Ben Cheikh and Waël Louhichi. Revisiting the role of inflation environment in exchange rate pass-through: A panel threshold approach. *Economic Modelling*, 52:233–238, 2016.

Cathy WS Chen, Mike KP So, and Richard H Gerlach. Assessing and testing for threshold nonlinearity in stock returns. *Australian & New Zealand Journal of Statistics*, 47(4):473–488, 2005.

Cathy WS Chen, Richard Gerlach, and Edward MH Lin. Volatility forecasting using threshold heteroskedastic models of the intra-day range. *Computational Statistics & Data Analysis*, 52(6):2990–3010, 2008.

Jia Chen. Estimating latent group structure in time-varying coefficient panel data models. *The Econometrics Journal*, 22(3):223–240, 2019.

Jia Chen, Jiti Gao, and Degui Li. A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models. *Econometric Theory*, 28(5): 1144–1163, 2012a.

Jia Chen, Jiti Gao, and Degui Li. Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85, 2012b.

Song Xi Chen and Cheng Yong Tang. Nonparametric inference of value-at-risk for dependent financial returns. *Journal of financial econometrics*, 3(2):227–255, 2005.

In Choi. Nonstationary panels. *Econometric theory*, 1:511–539, 2006.

Alexander Chudik, M Hashem Pesaran, and Elisa Tosetti. Weak and strong cross-section dependence and estimation of large panels, 2011.

Gilbert Colletaz and Christophe Hurlin. Threshold effects of the public capital productivity: an international panel smooth transition approach. *Econometric Society European Meeting*, 2006.

Richard Damania, Sebastien Desbureaux, and Esha Zaveri. Does rainfall matter for economic growth? evidence from global sub-national data (1990–2014). *Journal of Environmental Economics and Management*, 102:102335, 2020.

James Davidson. *Stochastic limit theory: An introduction for econometricians*. OUP Oxford, 1994.

Manfred Deistler, Brian DO Anderson, Alexander Filler, Ch Zinner, and Weitan Chen. Generalized linear dynamic factor models: An approach via singular autoregressions. *European Journal of Control*, 16(3):211–224, 2010.

Chandrakant M Deo. A note on empirical processes of strong-mixing sequences. *The Annals of Probability*, 1(5):870–875, 1973.

Chaohua Dong, Jiti Gao, and Bin Peng. Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics*, 188(1):301–312, 2015.

Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 50(4):987–1007, 1982.

Stefano Fachin. Long-run trends in internal migrations in italy: A study in panel cointegration with dependent units. *Journal of Applied Econometrics*, 22 (2):401–428, 2007.

Jianqing Fan and Qiwei Yao. *Nonlinear time series: nonparametric and parametric methods*. Springer, 2003.

Jianqing Fan, Jingjin Zhang, and Ke Yu. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498): 592–606, 2012.

Danvee Floro and Björn Van Roye. Threshold effects of financial stress on monetary policy rules: A panel data analysis. *International Review of Economics & Finance*, 51:599–620, 2017.

LLdiko E Frank and Jerome H Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, 35(2):109–135, 1993.

Edward W Frees. Assessing cross-sectional correlation in panel data. *Journal of econometrics*, 69(2):393–414, 1995.

Jerome Friedman, Trevor Hastie, Holger Höfling, Robert Tibshirani, et al. Pathwise coordinate optimization. *Annals of applied statistics*, 1(2):302–332, 2007.

Jiti Gao. *Nonlinear time series: semiparametric and nonparametric methods*. CRC Press, 2007.

Jiti Gao, Zudi Lu, and Dag Tjøstheim. Moment inequalities for spatial processes. *Statistics & probability letters*, 78(6):687–697, 2008.

Jiti Gao, Kai Xia, and Huanjun Zhu. Heterogeneous panel data models with cross-sectional dependence. *Journal of Econometrics*, 219(2):329–353, 2020.

A González, T Teräsvirta, and D Van Dijk. Panel smooth transition regression model and an application to investment under credit constraints. Technical report, Working Paper, Stockholm School of Economics, 2005.

Jesús Gonzalo and Jean-Yves Pitarakis. Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, 110 (2):319–352, 2002.

Peter Hall and Christopher C Heyde. *Martingale limit theory and its application*. Academic press, 2014.

Bruce E Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, 64(2):413–430, 1996.

Bruce E Hansen. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of econometrics*, 93(2):345–368, 1999.

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

Jerry A Hausman. Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 46(6):1251–1271, 1978.

David Hirshleifer and Tyler Shumway. Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032, 2003.

Daniel Hoechle. Robust standard errors for panel regressions with cross-sectional dependence. *The stata journal*, 7(3):281–312, 2007.

Holger Hoefling. A path algorithm for the fused lasso signal approximator. *Journal of Computational and Graphical Statistics*, 19(4):984–1006, 2010.

Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

Junbing Huang, Qiang Liu, Xiaochen Cai, Yu Hao, and Hongyan Lei. The effect of technological factors on china's carbon intensity: New evidence from a panel threshold model. *Energy Policy*, 115:32–42, 2018.

Ravi Jagannathan and Tongshu Ma. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance*, 58(4):1651–1683, 2003.

Sang Hoon Kang, Zhuhua Jiang, Yeonjeong Lee, and Seong-Min Yoon. Weather effects on the returns and volatility of the shanghai stock market. *Physica A: Statistical Mechanics and its Applications*, 389(1):91–99, 2010.

Stephen P Keef and Melvin L Roush. Influence of weather on new zealand financial securities. *Accounting & Finance*, 45(3):415–437, 2005.

Jeankyung Kim and David Pollard. Cube root asymptotics. *The Annals of Statistics*, 18(1):191–219, 1990.

Tae Yoon Kim. A note on moment bounds for strong mixing sequences. *Statistics & probability letters*, 16(2):163–168, 1993.

Keith Knight and Wenjiang Fu. Asymptotics for lasso-type estimators. *Annals of statistics*, 28(5):1356–1378, 2000.

Walter Krämer and Ralf Runde. Stocks and the weather: An exercise in data mining or yet another capital market anomaly? *Empirical Economics*, 22(4): 637–641, 1997.

Hans R Kunsch. The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, 17(3):1217–1241, 1989.

Sokbae Lee, Myung Hwan Seo, and Youngki Shin. The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 78(1):193, 2016.

Sokbae Lee, Yuan Liao, Myung Hwan Seo, and Youngki Shin. Oracle estimation of a change point in high-dimensional quantile regression. *Journal of the American Statistical Association*, 113(523):1184–1194, 2018.

Degui Li, Junhui Qian, and Liangjun Su. Panel data models with interactive fixed effects and multiple structural breaks. *Journal of the American Statistical Association*, 111(516):1804–1819, 2016.

Z Lin and C Lu. *Limit theorems of mixing dependent random variables*. Science Press, Kluwer, Academic Pub: New York, Dordrecht, 1996.

Jun Liu, Lei Yuan, and Jieping Ye. An efficient algorithm for a class of fused lasso problems. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 323–332, 2010.

Regina Y Liu, Kesar Singh, et al. Moving blocks jackknife and bootstrap capture weak dependence. *Exploring the limits of bootstrap*, 225:248, 1992.

Z. Lu, D.J. Steinskog, D. Tjøstheim, and Q. Yao. Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(4):859–880, 2009.

Zudi Lu. On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica*, pages 1205–1217, 1998.

Zudi Lu and Zhenyu Jiang. L1 geometric ergodicity of a multivariate nonlinear ar model with an arch term. *Statistics & Probability Letters*, 51(2):121–130, 2001.

Ke Miao, Kunpeng Li, and Liangjun Su. Panel threshold models with interactive fixed effects. *Journal of Econometrics*, 219(1):137–170, 2020a.

Ke Miao, Liangjun Su, and Wendun Wang. Panel threshold regressions with latent group structures. *Journal of Econometrics*, 214(2):451–481, 2020b.

Hyungsik Roger Moon and Benoit Perron. An empirical analysis of nonstationarity in a panel of interest rates with factors. *Journal of Applied Econometrics*, 22(2):383–400, 2007.

Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.

M Hashem Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012, 2006.

M Hashem Pesaran. General diagnostic tests for cross-sectional dependence in panels. *Empirical Economics*, 60(1):13–50, 2021.

Peter CB Phillips and Donggyu Sul. Dynamic panel estimation and homogeneity testing under cross section dependence. *The econometrics journal*, 6(1):217–259, 2003.

Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.

Junhui Qian and Liangjun Su. Shrinkage estimation of common breaks in panel data models via adaptive group fused lasso. *Journal of Econometrics*, 191(1):86–109, 2016.

Alessandro Rinaldo et al. Properties and refinements of the fused lasso. *Annals of Statistics*, 37(5B):2922–2952, 2009.

Donald Robertson and James Symons. *Factor residuals in SUR regressions: estimating panels allowing for cross sectional correlation*. Number 473. Centre for Economic Performance, London School of Economics and Political ..., 2000.

Peter M Robinson. Nonparametric trending regression with cross-sectional dependence. *Journal of Econometrics*, 169(1):4–14, 2012.

Edward M Saunders. Stock prices and wall street weather. *The American Economic Review*, 83(5):1337–1345, 1993.

Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

Myung Hwan Seo and Oliver Linton. A smoothed least squares estimator for threshold regression models. *Journal of Econometrics*, 141(2):704–735, 2007.

Myung Hwan Seo and Yongcheol Shin. Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2):169–186, 2016.

Mike KP So, Cathy WS Chen, Thomas C Chiang, and Doris SY Lin. Modelling financial time series with threshold nonlinearity in returns and trading volume. *Applied Stochastic Models in Business and Industry*, 23(4):319–338, 2007.

George C Tiao and Ruey S Tsay. Some advances in non-linear and adaptive modelling in time-series. *Journal of forecasting*, 13(2):109–131, 1994.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.

Dag Tjøstheim. Non-linear time series: a selective review. *Scandinavian Journal of Statistics*, 21(2):97–130, 1994.

Howell Tong. On a threshold model. *Pattern Recognition and Signal Processing*, 1978.

Howell Tong. *Non-linear time series: a dynamical system approach*. Oxford university press, 1990.

Howell AM Tong. *Chaos and Forecasting-Proceedings of the Royal Society Discussion Meeting*, volume 2. World Scientific, 1995.

Mark A Trombley. Stock prices and wall street weather: Additional evidence. *Quarterly Journal of Business and Economics*, 36(3):11–21, 1997.

Ruey S Tsay. Testing and modeling threshold autoregressive processes. *Journal of the American statistical association*, 84(405):231–240, 1989.

Geoffrey S Watson. Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964.

Tong Tong Wu and Kenneth Lange. Coordinate descent algorithms for lasso penalized regression. 2008.

Ping Yu and Xiaodong Fan. Threshold regression with a threshold boundary. *Journal of Business & Economic Statistics*, 39(4):953–971, 2021.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429, 2006.