Programa de Doctorado Interuniversitario en Economía:
Instrumentos de Análisis Económico.

Modelos de coeficientes variables: Verosimilitud empírica y tests de estabilidad

Luis Antonio Arteaga Molina

Tesis Doctoral

Directores:
Juan Manuel Rodríguez Poo
Miguel Ángel Delgado González

Universidad de Cantabria, Universidad de Oviedo y
Universidad del País Vasco
Escuela de Doctorado de la Universidad de Cantabria

Santander, 2019

Varying coefficient models: Empirical likelihood and stability tests

Luis Antonio Arteaga Molina

Supervisors:
Juan Manuel Rodríguez Poo
Miguel Ángel Delgado González

Ph.D. Thesis
Department of Economics
University of Cantabria
May, 2019

I would like to dedicate this thesis to my loving family ...

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Luis Antonio Arteaga Molina
May 2019

Acknowledgements

I would like to start by gratefully acknowledge all the people that in one way or another contributed to this Ph.D. thesis. I was not alone in the process, many great friends and colleagues have contributed their fair share to this thesis. You deserve more than a thank you; naming hereafter is the least I can do.

First of all, I would like to express my gratitude to my supervisors, Dr. Juan Manuel Rodriguez Poo and Dr. Miguel Angel Delgado Gonzalez. In fact, this Ph.D. thesis is as much mine as it is theirs. Their guidance, advice and encouragement have been important to pursue this Ph.D. thesis, I can't thank them enough for their support. Juan, thank you for giving me the opportunity to pursue this Ph.D. thesis with you. You have bean a great academic mentor and also extremely supportive, you gave me the freedom and support I needed to accomplish this. Your enthusiasm for good research and the lessons I have learned from you have been a great source of inspiration and gave a me a solid foundation for this thesis. On the personal side, you have always been interested and caring. Thank you for everything. Miguel, although you joined in during the second half of this thesis, your support and rigorous comments of my work have been of great value to this thesis. I admire your honest and straightforward way of communication, which has made working with you a pleasure. Thank you very much for giving me the chance to work with you.

I also want to extend my gratitude to my colleagues of the Department of Economics of the University of Cantabria led by Ruben Sainz Gonzalez, specially to Alexandra Soberón and Particia Moreno. Thank you for your help, support and advice. You have been great colleagues to work with, and I would like to thank you in particular for helping me in gaining relevant teaching experience - I learn some valuable lessons from you in this capacity. I also have to mention my friend and colleague Xosé Luis Fernández, while we work in different areas of study, you always have been supportive and interested in my work.

Further, I would like to thank my office mates Gonzalo, Ana and Marta. Besides being great colleagues they have become friend to me. Our joint times have been great fun and they were always there to discuss important and unimportant matters. We started our Ph.D. almost at the same time, and shared an office for over three years. It has been a pleasure to have you as office mates, and I particularly enjoyed our discussions about our future. Thank you for all the good moments and support.

I also would like to take the opportunity to thank the Department of Economics of the University Carlos III de Madrid, where I spent four months doing a research stay. Here I would like to thank Calos Velasco, the director of the department at the time, and Miguel Angel Delgado Gonzalez, my supervisor during the research stay. I am extremely grateful for giving me the chance to participate of the department's activities.

Next I would like to mention the research project ECO2013-48326-C2-2-P, "Modelos paramétricos y semi-paramétricos de riesgos dependientes: especificación y contraste" funded by Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia / Spanish Ministry of Economy and Competitiveness, thanks to which this PhD thesis has been possible, and thank Professor Jose Maria Sarabia Alegria and all the people involved in the project for their kindness, trust and advice.

My last words of thanks go to those close to me, my friends, family and wife. Many of you have provided me with the needed distraction from work and a shoulder to cry on when thing don't go as planned. In particular, I would like to thank my parents, Sonia and Raul, my brothers, Jose and Jonathan, and my extended family for being there in times, but also for understanding when I needed time for this thesis. Finally, I need to thank the most important person in my life, Andreea. I am truly blessed for having you in my life. The important role that you have played in finishing this thesis can not be expressed in words. Thanks for being you and encourage me to be a better version of myself.

Table of contents

List of figures

List of tables

Abstract

The goal of this doctoral dissertation is to apply and develop inference techniques for varying coefficient models. On the one hand, empirical likelihood based inference for categorical and continuous varying coefficient models, under a panel data with fixed effects framework, is investigated. First, we show that the naive empirical likelihood ratio is asymptotically standard chi-squared. The ratio is self-scale invariant and the plug-in estimate of the limiting variance is not needed. As a by product, we propose empirical maximum likehood estimators for varying coefficients. We also obtain the asymptotic distribution of these estimators and we propose some procedures to calculate the bandwidths empirically. Furthermore, a non parametric version of the Wilk's theorem is derived. To show the feasibility of the technique and to analyse its small sample properties we implement a Monte Carlo simulation exercise and we also illustrated the proposed technique in an empirical analysis.

On the other hand, we propose tests for constancy of coefficients in varying coefficients models under different settings. For exogenous regressors, the testing procedure resembles in spirit the union-intersection parameter stability tests in time series. The test can be applied to model specification checks of interactive effects in linear regression models. Because test statistics are not asymptotically pivotal, critical values and p-values are estimated using a bootstrap technique. For the endogenous case, the testing procedure is defined as a generalized likelihood ratio that focus on the comparison of the restricted and unrestricted sum of squared residuals. As a by product, and resembling the instrumental variable literature, we propose to use a three stages estimation procedure to estimate the varying coefficients; we also establish the asymptotic properties of the estimators. The finite sample properties of the test are investigated by means of Monte Carlo experiments.

Introduction

Varying coefficients models, also called functional coefficient models, have gained importance since they where introduced by Cleveland et al. (1991). These type of models are an important tool to explore dynamic patterns and they appear as a natural extension of the classical parametric models; these models are characterized by allowing the parameters of interest to change smoothly according to the value of some variable proposed by the economic theory. The main advantage of these models is that not prior assumptions of the model's specification is needed. In this sense, these models increase the flexibility of the classical linear regression models as they are able to exploit the information from the data set.

In the past three decades varying coefficient models have experienced a great growth, from both a methodological and a theoretical point of view; the main reason is that they offer a quite general setting to handle many of the specification problems of nonparametric and semiparametric models. To gain a better understanding of the many advantages of varying coefficient models for empirical analysis, we will present some examples of empirical applications.

The first example is provided by the microeconomic and production function literature. Li et al. (2002) conducted a study to analyse the production function of the nonmetal mineral manufacturing industry in China. They conclude that the classical Cobb-Douglas model does not provide an adequate description of the relationship between output and explanatory variables. Also, they show that the marginal returns to labour and capital vary according to the level of the firms' research and development expenses. In this situation a varying coefficient model of the following form may be adequate,

$$y_i = \alpha(z) + \beta_k(z)k_i + \beta_\ell(z)\ell_i + u_i, \text{ for } i = 1, \ldots, N,$$

where $y_i = \ln(Y_i)$ is the added value in thousand renminbi (RMB), $k_i = \ln(K_i)$ is the value of capital assets in thousand RMB, $\ell_i = \ln(L_i)$ is the average number of employees and $z_i = \ln(Z_i)$ is the intermediate production and management expenses in thousand RMB. As it can be seen, this model allows the marginal returns to labour and capital vary according to the level of the firms' research and development expenses.

Under the same cross sectional setting, we find another example in the literature on returns to education. The study of Schultz (2003) shows that the marginal returns to education vary with the level of work experience; therefore, as it is shown in Card (2001), the omission of the nonlinearity of education as well as the interaction impact between education and work experience led us to underestimated outcomes of the education performance. In this situation, a semiparametric varying coefficient model is more convenient as it allows the impact of education to vary with the level of work experience.

Furthermore, under a panel data setting, the international economic literature provides another relevant example. Examining the role of foreign direct investment (FDI) in the economic growth of the countries, Kottaridi and Stengos (2010) show that the positive effect of the FDI on the economic growth only happens in those countries with higher levels of initial income; this translates in a varying coefficient of the FDI that varies with the level initial income of each country. Therefore, the following panel data varying coefficient model is appealing since it allow us to gather this effect,

$$Y_{it} = \alpha_0 + \alpha_1 D_j + \alpha_2 ln\left(\frac{I_{it}^d}{Y}\right) + \alpha_3 ln(n_{it}) + \alpha_4(lnX_{it})\left(\frac{I_{it}^f}{Y}\right) + \alpha_5 h_{it} + \varepsilon_{it},$$

for $i = 1, \ldots, N \quad ; \quad t = 1, \ldots, T$, where Y_{it} is the growth rate of income per capita in country i and period t, I_{it}^d/Y, the domestic investment rate to GDP, n_{it} the population growth rate, h_{it} the human capital, I_{it}^f/Y is the ratio of FDI to GDP and X_{it} is the income per capita at the beginning of each period.

Finally, we find another relevant example in the capital asset pricing literature; here empirical evidence suggest time variation in betas and returns. Authors such as Cho and Engle (1999), Wang (2002), Akdeniz et al. (2003), Wang (2003) and Fraser et al. (2004) among others, find evidence for significance of nonlinearity in the betas and conclude that changes that occur through time in the risk factor are associated with changes in the economic environment. Thus, following Cai et al. (2015), the next semiparametric varying coefficient model is advisable

$$\mathrm{E}\left[\left(1 - m(Z_t) r_{p,t+1}\right) r_{i,t+1} \middle| \Omega_t\right] = 0, \quad t = 1, \ldots, T,$$

where Z_t is a vector of conditioning variables from Ω_t and $r_{p,t+1}$ is the factor.

As it has been shown, varying coefficient models allow the coefficients of the regression model to be unknown functions of some other variables. Therefore, testing on varying coefficients is of great importance as it implies testing on structural information and the underlying economic theory; thus, developing inference devices for varying coefficient models is crucial.

In this context, the goal of this Ph.D. thesis is twofold. On the one hand, to develop confidence bands for varying coefficient models using the empirical likelihood technique. On the other hand, testing that the varying coefficients are constant in the direction of nonparametric alternatives. With these objectives, the Ph.D. thesis is divided into four chapters structured as follows.

In Chapter 1 empirical likelihood based inference for fixed effects varying coefficient panel data models is investigated. Empirical Likelihood is a nonparametric technique of inference based on a data driven likelihood function. The method was introduced in Owen (1990, 1991, 1988, 2001) as a generalization of Thomas and Grunkemeier (1975)'s survival probabilities. In the last three decades, the method have gained importance due to its properties and advantages over other methods such as asymptotic normal based confidence bands, bootstrap or jackknife.

Likelihood methods are often used to find efficient estimators, and to construct tests with good power properties; it is also a flexible method as it offsets or even corrects problems related to incomplete observed data or distorted data. Also, knowledge from outside the data can be incorporated via constrains that restrict the domain of the likelihood function. In parametric likelihood methods, we assume that the data comes from a known joint distribution; however, in practice, we might not know the parametric family and misspecification problems might arise, which could cause inefficient estimates and the test to fail. To avoid this problem, nonparametric methods of inferences appears as a solution; here, besides empirical likelihood, these methods include the jackknife and several versions of the bootstrap. These methods do not use strong distributional assumptions.

Empirical likelihood can be seen as a bootstrap that does not resample and as a likelihood without parametric assumptions. Thus, the main advantage of the empirical likelihood approach is that it combines the reliability of nonparametric methods with the effectiveness of the likelihood approach, (Owen, 2001); other advantages include that there is no need of scale, skewness or limiting variance estimation, (Hall and La Scala, 1990), it is range and transformation respecting, (Hall and La Scala, 1990), and it is Bartlett correctable, (DiCiccio et al., 1991). However, the most appealing property of this method is the asymptotic distribution of the empirical likelihood ratio test statistic follows a chi-squared distribution, which is the same as the one under parametric settings, (Owen, 1990, 1988).

Let $x_1, x_2, \ldots x_N$ be independently and identically distributed observations from an unknown population distribution, F_0, with the mean μ and variance σ^2. Let $p_i = \Pr(X = x_i)$ under some cumulative distribution function $F(x) = \Pr(X \leq x)$. The empirical likelihood function of F is defined as

$$L(F) = \prod_{i=1}^{N} p_i \quad \text{s.t. } p_i \geq 0, \quad \sum_{i=1}^{N} p_i = 1.$$

Note that $L(F)$ is maximized at $p_i = 1/N$. Also note that when a population parameter β defined by $\mathrm{E}[g(\beta, X_i)] = 0$ is of interest, the empirical likelihood, subject to the additional constraint

$$\sum_{i=1}^{N} p_i g(\beta_0, X_i) = 0,$$

reaches its maximum when β is its true value β_0. Therefore the empirical likelihood ratio statistic to test $\beta = \beta_0$ is given by

$$\mathscr{R}(\beta_0) = \left\{ \sum_{i=1}^{N} \log N p_i \middle| p_i > 0, \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i g(\beta_0, X_i) = 0 \right\}.$$

Let, for example, β be the sample mean and β_0 be the population mean, Owen (1988) show that $-2\mathscr{R}(\beta_0) \to \chi^2_{(1)}$ in distribution as $N \to \infty$.

In this context, in chapter 1, we first show that the naive empirical likelihood ratio for the varying coefficient is asymptotically standard chi-squared when undersmoothing is employed. The ratio is self-scale invariant and the plug-in estimate of the limiting variance is not needed. To correct for the bias, mean-corrected and residual-adjusted empirical likelihood ratios are proposed and without undersmoothing, both also have standard chi-squared limit distributions. As a by product, we propose the empirical maximum likehood estimators of the varying coefficient and their derivatives; here, the derivative result can be seen as a way to test constancy of the varying coefficient. We also obtain the asymptotic distribution of these estimators and we propose some procedures to calculate the bandwidths empirically. To show the feasibility of the technique and to analyse its small sample properties we implement a Monte Carlo simulation exercise and we also illustrate the proposed technique in an empirical analysis about the production efficiency of the European Union's companies.

In Chapter 2, and complementing chapter 1, we investigate empirical likelihood based inference for nonparametric categorical varying coefficient panel data models with fixed effects under cross-sectional dependence. The main difference with chapter 1 is that in this case the varying coefficient varies according to a discrete variable and therefore smoothing functions for discrete variables are needed. In this discrete context, we show that the naive empirical likelihood ratio is asymptotically standard chi-squared using a nonparametric version of the Wilks' theorem, (Wilks, 1938). The ratio is still self-scale invariant and, more importantly, no plug-in estimator of the limiting variance is needed; this last feature is important because in this case due to the cross sectional dependence the estimation variance becomes cumbersome. As a by product, we propose also a empirical maximum likelihood estimator of the categorical varying coefficient and we obtain the asymptotic distribution of this estimator. We also illustrated the proposed technique in an application that reports estimates of strike activities from 17 countries of the Organisation for Economic Co-operation and Development (OECD) for the period 1951 – 1985.

In Chapter 3, and developing the idea of testing constancy from chapter 1, we propose tests for constancy of coefficients in semi-varying coefficients models. The testing procedure resembles in spirit the union-intersection (U-I) parameter stability tests in time series, where observations are sorted according to the explanatory variable responsible for the coefficients varying. Parameter stability tests appear in the time series literature to deal with change-point problems (see, e.g., Bhattacharyya and Johnson (1968), Brown et al. (1975), Csörgő and Horváth (1988, 1997), Hawkins (1989), James et al. (1987), Andrews (1993), and Nyblom (1989) among others). In recent years, increasing interest has been shown in problems concerning stability of a regression model because changes in economic factors may cause instability of their initial models over a long period of time. For example, technical progress, changes in policies and regulations, or a different economic environment can induce a change among economic variables, even though no change in the parameters of the structural relationship is present. In this context, our test can be applied to model specification checks of interactive effects in linear regression models. Because test statistics are not asymptotically pivotal, critical values and p-values are estimated using a bootstrap technique. The finite sample properties of the test are investigated by means of Monte Carlo experiments, where the new proposal is compared to existing tests based on smooth estimates of the unrestricted model. We also report an application to returns of education modelling.

In Chapter 4, and complementing chapter 3, we propose a methodology for testing coefficients constancy in varying coefficient with endogenous regressors. The proposed test is applied to conditional capital asset pricing models (CAPM). Conditional asset pricing literature provides a framework in which these models do not provide much information the functional form of the conditional moments. There exist a wide literature that employs parametric techniques to evaluate conditional asset pricing models; for instance, Jagannathan and Wang (1996) find that a conditional CAPM can explain the cross section of stock of returns, while the static CAPM model cannot, also Lettau and Ludvigson (2001) show that the value premium can be explained by a conditional CAPM with time varying price of risk. Nevertheless, other authors such as Lewellen and Nagel (2006) and Nagel and Singleton (2011) suggest that this superior performance of the conditional CAPM is an illusion caused by the low statistical power of standard CAPM. Following these ideas, there exist significant contributions to avoid specifying the conditional distribution of returns and factors by using nonparametric techniques. Here, Nagel and Singleton (2011) estimate nonparametrically first and second conditional moments and then work with parametric CAPM. In contrast, Wang (2003), Orbe et al. (2008), Roussanov (2014) and Peñaranda et al. (2018) consider varying coefficient CAPM.

The testing procedure is defined as a generalized likelihood ratio that focus on the comparison of the restricted and unrestricted sum of squared residuals. As a by product, we have developed a nonparametric method that takes into account the endogenous nature of the regressors to estimate the prices of risk. Resembling the instrumental variable literature, the procedure uses a three stages estimation procedure to estimate the varying coefficient; besides we establish the asymptotic properties of the estimators. Finally, we investigate the finite sample properties of our test by means of Monte Carlo experiments study and using critical values and p-values estimated by the bootstrap technique.

Finally, we conclude by highlighting the main results extracted from these four chapters and possible future research. The proofs of the main results are relegated to the appendix.

Introducción

Modelos de coeficientes variables, también llamados modelos de coeficientes funcionales, han ido ganado importancia desde su introducción en Cleveland et al. (1991). Este tipo de modelos surgen como una extensión de los modelos paramétricos clásicos y una herramienta importante para explorar patrones dinámicos; estos modelos se caracterizan por permitir a los parámetros de interés variar de acuerdo a una variable propuesta por la teoría económica. La principal ventaja de utilizar este tipo de modelos es que no son necesarios supuestos sobre la especificación de la forma funcional de los coeficientes. De esta manera, estos modelos pueden explotar la información del conjunto de datos, con lo que son más flexibles que los modelos de regresión lineales clásicos.

En las tres últimas décadas, la literatura de modelos de coeficientes variables ha experimentado un gran crecimiento, tanto desde un punto de vista metodológico como teórico; la razón principal es que ofrecen un marco bastante general para manejar muchos de los problemas de especificación que aparecen en modelos no paramétricos y semiparamétricos. Para obtener una mejor comprensión de las muchas ventajas que estos modelos ofrecen para el análisis empírico, a continuación, presentaremos algunos ejemplos de aplicaciones empíricas.

El primer ejemplo lo proporciona la literatura microeconómica sobre funciones producción. En el artículo de Li et al. (2002), se realiza un estudio para analizar la función de producción de la industria de fabricación de minerales no metálicos en China. En este estudio se concluye que el modelo Cobb-Douglas claásico no proporciona una descripción adecuada de la relación entre la producción y las variables explicativas. Ellos demuestran que los rendimientos marginales del trabajo y el capital varían según el nivel de los gastos de investigación y desarrollo (I+D) de la empresa. En esta situación, un modelo de coeficiente variables con la siguiente especificación puede ser adecuado,

$$y_i = \alpha(z) + \beta_k(z)k_i + \beta_\ell(z)\ell_i + u_i, \text{ for } i = 1, \ldots, N,$$

donde $y_i = \ln(Y_i)$ es el valor de la produccioón aproximado por el valor añadido en miles de renminbi (RMB), $k_i = \ln(K_i)$ es el valor de los activos de capital en miles de RMB, $\ell_i = \ln(L_i)$ es el número promedio de empleados y $z_i = \ln(Z_i)$ son los gastos intermedios de producción y administración en miles de RMB, que aproximan los gastos de I+D. Como se puede observar, este modelo permite que los rendimientos marginales del trabajo y el capital varíen de acuerdo con el nivel de los gastos de investigación y desarrollo de la empresa.

Bajo el mismo contexto transversal, encontramos otro ejemplo en la literatura sobre rendimientos educativos. El estudio de Schultz (2003) se demuestra que el rendimiento marginal de la educación varía con el nivel de experiencia laboral; por lo tanto, como se demuestra en Card (2001), la omisión de la falta de linealidad en la educación, así como en el impacto de la interacción entre la educación y la experiencia laboral nos lleva a subestimar los resultados del desempeño educativo. En esta situación, un modelo de coeficientes variables semiparamétrico es más conveniente ya que permite que el impacto de la educación varíe con el nivel de experiencia laboral.

Bajo un contexto de datos de panel, la literatura sobre economía internacional proporciona otro ejemplo relevante. Al examinar el papel de la inversión extranjera directa (IED) en el crecimiento económico de los países, Kottaridi and Stengos (2010) demuestran que el efecto positivo de la IED en el crecimiento económico solo ocurre en aquellos países con niveles iniciales de rentas más altos; esto se traduce en un coeficiente asociado a la IED que varía con el nivel de renta inicial de cada país. Por lo tanto, el siguiente modelo de datos de panel con coeficientes variables es aconsejable ya que nos permite recopilar este efecto,

$$Y_{it} = \alpha_0 + \alpha_1 D_j + \alpha_2 ln\left(\frac{I_{it}^d}{Y}\right) + \alpha_3 ln(n_{it}) + \alpha_4 (lnX_{it})\left(\frac{I_{it}^f}{Y}\right) + \alpha_5 h_{it} + \varepsilon_{it},$$

para $i = 1,\ldots,N \quad ; \quad t = 1,\ldots,T$, donde Y_{it} es la tasa de crecimiento de la renta per cápita para el país i en el periodo t , I_{it}^d/Y, es la tasa de inversión interna con respecto al PIB, n_{it} es la tasa de crecimiento de la población, h_{it} representa el capital humano, I_{it}^f/Y es la ratio entre IED y PIB, y X_{it} es la renta per cápita al comienzo de cada periodo.

Para terminar, encontramos otro ejemplo relevante en la literatura de valoración de activos financieros; aquí la evidencia empírica sugiere variación de los betas y los rendimientos en el tiempo. Autores como Cho and Engle (1999), Wang (2002), Akdeniz et al. (2003), Wang (2003) y Fraser et al. (2004) entre otros, encuentran evidencia significativa de no linealidad en los betas y concluyen que los cambios que se producen a través del tiempo están asociados a cambios en el entorno económico. Por lo tanto, y siguiendo a Cai et al. (2015), es recomendable un modelo semiparamétrico de coeficientes variables

$$\mathrm{E}\left[\left(1 - m(Z_t) r_{p,t+1}\right) r_{i,t+1} \middle| \Omega_t\right] = 0, \quad t = 1,\ldots,T,$$

donde Z_t es un vector de variables de condicionamiento de Ω_t y $r_{p,t+1}$ es el factor.

Como se ha demostrado con los ejemplos anteriores, los modelos de coeficientes variables permiten que los coeficientes del modelo de regresión sean funciones desconocidas de alguna otra variable. Por lo tanto, los contrastes sobre coeficientes variables son de gran importancia ya que implican contrastar la información estructural del modelo y la teoría económica subyacente; por lo tanto, el desarrollo de métodos que nos permitan hacer inferencia sobre modelos de coeficientes variables es crucial.

En este contexto, el objetivo de esta tesis docotoral es doble. Por un lado, desarrollar bandas de confianza para modelos de coeficientes variables utilizando la técnica de verosimilitud empírica. Por otro lado, desarrollar tests que nos permitan discernir si los coeficientes variables son constantes en la dirección de alternativas no paramétricas. Con este fin, la tesis se divide en cuatro capítulos estructurados de la siguiente manera.

En el Capítulo 1 se investiga tecnicas de inferencia estadística basadas en la verosimilitud empírica para modelos de datos de panel con efectos fijos y coeficientes variables. La vrosimilitud empírica es una técnica no paramétrica de inferencia que se basa en una función de verosimilitud dada por los datos. Este método fue introducido por Owen (1990, 1991, 1988, 2001) como una generalización de las probabilidades de supervivencia de Thomas and Grunkemeier (1975). En las tres últimas décadas, esta técnica ha ganado importancia debido a sus propiedades y ventajas sobre otros métodos como las bandas de confianza basadas en la distribución asintótica de los estimadores, el bootstrap o el jackknife.

Los métodos de estimación basados en la verosimilitud se utilizan a menudo para encontrar estimadores eficientes y para desarrollar contrastes con buenas propiedades de potencia; también se trata de un método flexible, ya que compensa o incluso corrige problemas relacionados con datos incompletos o distorsionados. Además, conocimiento de información externa a los datos puede incorporarse mediante restricciones que restringen el dominio de la función de verosimilitud. En los métodos de verosimilitud paramétricos, asumimos que los datos provienen de una distribución conjunta conocida; sin embargo, en la práctica, es posible que no sepamos la familia paramétrica de la que provienen los datos y, debido a ello, puedan surgir problemas de especificación, lo que podría provocar estimaciones ineficientes y que el contraste falle. Para evitar este problema, los métodos no paramétricos de inferencia aparecen como una solución factible; en esta categoría, además de la verosimilitud empírica, tenemos el jackknife y varias versiones del bootstrap. Estos métodos no utilizan fuertes supuestos sobre la distribución conjunta de la que provienen los datos.

La verosimilitud empírica se puede entender como un bootstrap que no necesita remuestrear los datos y como una verosimilitud que no hace supuestos sobre la distribución de la que provienen los datos. Por lo tanto, la principal ventaja del enfoque de verosimilitud empírica es que combina la confiabilidad de los métodos no paramétricos con la efectividad de la verosimilitud, (Owen, 2001); otras ventajas incluyen que no necesita estimadores de escala, simetría o varianza, (Hall and La Scala, 1990), respeta el rango y las transformaciones, (Hall and La Scala, 1990), y es corregible por el método Bartlett, (DiCiccio et al., 1991). Pero, la propiedad más atractiva de este método es la distribución asintótica del estadístico de contraste, el ratio de verosimilitud empírica, que sigue a una distribución chi-cuadrada, que es la misma que la que se obtiene para los modelos paramétricos, (Owen, 1990, 1988).

Sea $x_1, x_2, \ldots x_N$ observaciones independiente e identicamente distribuidas de una distribución desconocida, F_0, con media μ y varianza σ^2. Además, sea $p_i = \Pr(X = x_i)$ bajo alguna función de distribución acumulada $F(x) = \Pr(X \leq x)$. La función de verosimilitud empírica de F se define como

$$L(F) = \prod_{i=1}^{N} p_i \quad \text{s.t. } p_i \geq 0, \quad \sum_{i=1}^{N} p_i = 1.$$

Es fácil ver que $L(F)$ se maximiza cuando $p_i = 1/N$. También, cuando un parámetro poblacional β definido por $\mathrm{E}[g(\beta, X_i)] = 0$ es de interés, la verosimilitud empírica, sujeta a la siguiente restricción adicional

$$\sum_{i=1}^{N} p_i g(\beta_0, X_i) = 0,$$

alcanza su máximo cuando β toma su verdadero valor, β_0. Por lo tanto, el estadístico de contraste basado en el ratio de verosimilitud empírica, para contrastar $\beta = \beta_0$, viene dado por

$$\mathscr{R}(\beta_0) = \left\{ \sum_{i=1}^{N} \log N p_i \,\middle|\, p_i > 0, \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i g(\beta_0, X_i) = 0 \right\}.$$

Sea β la media muestral y β_0 la media poblacional, Owen (1988) demuestra que $-2\mathscr{R}(\beta_0) \to \chi^2_{(1)}$ en distribución cuando $N \to \infty$.

En este contexto, en este primer capítulo, primero demostramos que el ratio de verosimilitud empírica para el coeficiente variable es asintóticamente chi-cuadrado cuando se emplea undersmoothing. El ratio es invariable a cambios de escala y no es necesaria la estimación de la varianza. Para corregir por el sesgo se porponen dos modificaciones del ratio de verosimilitud, uno coregido por la media y otro ajustado por los residuos, que sin undersmoothing son también sintóticamente chi-cuadrado. Como subproducto, proponemos los estimadores de máxima verosimilitud empírica de los coeficientes variables y sus derivadas; en este contexto, la derivada puede verse como una forma de contrastar la constancia de los coeficientes variables. También obtenemos la distribución asintótica de estos estimadores y proponemos algunos procedimientos para calcular los bandwidths empíricamente. Para demostrar la viabilidad de la técnica y analizar sus propiedades en muestras finitas, implementamos un ejercicio de simulación de Monte Carlo, y también proponemos un análisis empírico sobre la eficiencia de la producción de las empresas de la Unión Europea.

En el capítulo 2, y complementando lo que se ha desarrollado en el capítulo 1, se investiga tecnicas de inferencia estadística basadas en la verosimilitud empírica para modelos de datos de panel con efectos fijos y coeficientes variables categóricos o discretos. La principal diferencia con el capítulo 1 es que, en este caso, el coeficiente variable varía según una variable de naturaleza categórica o discreta, y por lo tanto se necesitan funciones de suavizado para variables discretas. En este contexto de variables discretas, demostramos que el ratio de verosimilitud empírica es asintóticamente chi-cuadrado, utilizando para ello una versión no paramétrica del teorema de Wilks, (Wilks, 1938). El ratio sigue siedo invariable a cambios en escala y, lo que es más importante, no se necesita un estimador de la varianza; esta última característica es importante porque en este caso, debido a la dependencia de sección cruzad, la estimación de la varianza se complica. Como subproducto, proponemos también un estimador de máxima verosimilitud empírica de los coeficientes variables categóricos y obtenemos su distribución asintótica. También ilustramos la técnica propuesta en una aplicación empírica que reporta estimaciones de las actividades de huelga de 17 países perntenecientes a la Organización para la Cooperación y el Desarrollo Económico (OCDE) para el periodo 1951 – 1985.

En el capítulo 3, y desarrollando la idea que se propuso en el capítulo 1 sobre el contraste de constancia, se propone un test para detectar constancia en los parametros en modelos de coeficientes variables. El procedimiento para relaizar el contraste se asemeja a los contrastes de unión-intersección (U-I) de estabilidad de parámetros en series temporales, donde las observaciones se ordenan de acuerdo con la variable explicativa responsable de la variación de los coeficientes. En el contexto del contraste U-I esta ordenación resulta de manera natural porque los coeficientes varían con el tiempo, por el contrario en nuestro caso usamos concomitantes para ordenar los datos de acuerdo con la variable explicativa responsable de la variación de los coeficientes.

Los contrastes de estabilidad de parámetros aparecen en el nanálisis de series temporales para tratar con problemas de puntos de cambio (chnage-point) (ver, por ejemplo, Bhattacharyya and Johnson (1968), Brown et al. (1975), Csörgő and Horváth (1988, 1997), Hawkins (1989), James et al. (1987), Andrews (1993), y Nyblom (1989) entre otros). En los últimos años, ha crecido el interés en los problemas relacionados con la estabilidad de los parámetros en modelos de regresión debido a que cambios en los factores económicos pueden causar inestabilidad en los modelos inicialmente propuestos durante un largo periodo de tiempo. Por ejemplo, el progreso tecnológico, los cambios en políticas y regulaciones, o un entorno económico diferente pueden inducir un cambio entre las variables económicas, aunque no haya cambios en los parámetros de la relación estructural.

En este contexto, nuestro test puede aplicarse para verificar la especificación de modelización de efectos interactivos en modelos de regresión lineal. Debido a que el estadístico de contraste no es asintóticamente pivotal, los valores críticos y los p-valores se estiman utilizando la técnica del bootstrap. Las propiedades para muestra finitas del test se investigan por medio de un experimento de Monte Carlo, donde nuestra propuesta se compara con contrastes ya existentes basadas en estimaciones con suavizado del modelo no restringido. Además, tmabién ponemos a prueba nuestro test con una aplicación sobre la modelización de los rendimientos educativos.

En el capítulo 4, y complementando lo que se hizo en el capítulo 3, se propone un test para detectar constancia en moelos con coeficientes variables y regresores endógenos. El test, en concreto, se aplica en modelos de valoración de activos financieros (CAPM).

La literatura de valoración de activos financieros condicionales proporciona un marco en el que estos modelos no proporcionan información sobre la forma funcional de los momentos condicionales. Existe una amplia literatura que emplea técnicas paramétricas para evaluar modelos de valoración de activos financieros condicionales; por ejemplo, Jagannathan and Wang (1996) encuentran que un CAPM condicional puede explicar la sección cruzada del stock de rendimientos, mientras que el modelo CAPM estático no puede, también Lettau and Ludvigson (2001) demuestra que el valor premium puede ser explicada por un CAPM condicional con primas de riesgo que varían con el tiempo. Sin embargo, otros autores como Lewellen and Nagel (2006) y Nagel and Singleton (2011) sugieren que este rendimiento superior del CAPM condicional es una ilusión causada por el bajo poder estadístico del CAPM estándar. Siguiendo estas ideas, existen contribuciones significativas para evitar especificar la distribución condicional de los rendimientos y los factores mediante el uso de técnicas no paramétricas. Aquí, Nagel and Singleton (2011) estima el primer y segundo momento condicional no paramétricamente y luego trabaja con un CAPM paramétrico. Por el contrario, Wang (2003), Orbe et al. (2008), Roussanov (2014) y Peñaranda et al. (2018) consideran CAPM con coeficientes variables.

En este contexto, el test se define como un ratio de verosimilitud generalizado que se enfoca en la comparación de la suma de cuadrada de los residuos del modelo restringido y no restringido. Como subproducto, hemos desarrollado un método no paramétrico de estimación que tiene en cuenta la naturaleza endógena de los regresores, con el que hemos podido estimar los betas del modelo CAPM. Mimetizando la literatura de variables instrumentales, proponemos utilizar un procedimiento de estimación en tres etapas para estimar los coeficiente variables; además establecemos las propiedades asintóticas de los estimadores. Pra terminar, investigamos las propiedades en muestras finitas de nuestro test por medio de experimento de Monte Carlo, para lo cual los valores críticos y los p-valores se estiman utilizando la técnica del bootstrap.

Finalmente, concluimos destacando los principales resultados extraídos de estos cuatro capítulos y las posibles investigaciones futuras. Las pruebas de los principales resultados quedan relegadas al apéndice.

Chapter 1

Empirical likelihood based inference for fixed effects varying coefficient panel data models

This chapter also appeared as Arteaga-Molina and Rodriguez-Poo (2018).

1.1 Introduction

Recently nonparametric and semiparametric estimation of panel data models has attracted the attention of many researchers in econometrics. The interest to combine panel data techniques, that somehow alleviate the heterogeneity issue, with nonparametric techniques, that weaken considerably the type of assumptions that are necessary to impose in econometric models, has ended up in a vast literature that is surveyed in Su and Ullah (2011). Although the results are rather promising, it is true that the main drawbacks related to nonparametric techniques also appear when we apply them to panel data econometric models. Among others, the curse of dimensionality (e. g., Härdle (1990)) appears as one of the most important problems. In order to overcome this disadvantage varying coefficient models appear as a reasonable specification that encompasses many alternative models. As for the pure nonparametric case, estimation of varying coefficient models with random effects has been already studied in several papers (e.g., Ruckstuhl et al. (2000); Lin and Carroll (2000); Henderson and Ullah (2005); Su and Ullah (2007)). However, under the setting of fixed effects unfortunately much less results are available. In Henderson et al. (2008) direct estimation of the nonparametric components is undertaken through the use of an iterative version of a profile least squares technique. Already in a varying coefficients context a profile least squares approach is proposed in Sun et al. (2009). For differencing estimators in Rodriguez-Poo and Soberón (2014, 2015) two step backfitting estimators are proposed. Furthermore, a comparison against estimators based in profile least squares techniques is provided. In Cai and Li (2008) a so called nonparametric generalized method of moments is proposed to estimate the varying coefficients. Finally, in Su and Lu (2013) and Li and Liang (2015) profile least squares results are extended towards dynamic models and smooth backfitting methods are applied to estimate the unknown varying coefficients respectively. Eventually, once we have taken care of the estimation process, the next step would be to concentrate in developing inference tools for this type of models. For statistical inference such as confidence region construction or hypothesis testing the most popular techniques are normal approximations and bootstrap methods. In fact, in all above mentioned papers, asymptotic normal approximations are obtained for the different nonparametric estimators. Unfortunately it is well known that, without undersmoothing, the asymptotic distribution will exhibit a bias and a rather cumbersome expression for the variance term. Hence, if the confidence region that is derived from an asymptotic normal distribution is predetermined to be symmetric a bias correction and a plug-in estimate are needed to make the statistic scale invariant. Furthermore, if one wants to use these confidence bands as a testing device it will be necessary to obtain uniform confidence bands such as in Li et al. (2013a).

In this chapter, we propose to use empirical likelihood techniques to construct confidence intervals/regions. These techniques have acquired importance since they were introduced in Owen (1990, 1991, 1988, 2001) because of the advantages of this method over other methods such as normal approximation and bootstrap; for instance, empirical likelihood methods adjust to the true shape of the underlying distribution and do not require the estimation of scale, skewness (Hall and La Scala, 1990) or limiting variance as the studentization is carried out internally via optimization. Therefore, the confidence regions are reliable, range preserving and transformation respecting (Hall and La Scala, 1990). Another advantages is the method's flexibility, as it can be used when the data is incomplete, distorted or tied. Also, DiCiccio et al. (1991) have proved that empirical likelihood regions are Bartlett correctable; thus, it has advantages over the bootstrap and the jackknife methods. Finally, it combines the reliability of non-parametric methods with the effectiveness of the likelihood approach and it has good asymptotic properties and power (Owen, 1990). In fact, empirical likelihood techniques have been already applied to obtain confidence bands for longitudinal data varying coefficient models with random effects (e.g. Xue and Zhu (2007)) but unfortunately these type of results are not available for the fixed effects case. For the fixed effect case, in Zhang et al. (2011) confidence bands based in empirical likelihood techniques are derived under a partially linear model specification. They obtain, under rather restrictive assumptions, maximum empirical likelihood estimators of both parametric and nonparametric components. Furthermore, they obtain an empirical likelihood ratio that is biased if the optimal bandwidth is used.

In this chapter, and starting from a fixed effects varying coefficient model, we obtain maximum empirical likelihood estimators of both the varying parameters and their derivatives. This last result is very interesting for testing constancy of parameter variation. Furthermore, we develop empirical likelihood ratios and we derive a non-parametric version of the Wilks' theorem. In order to obtain an unbiased ratio, we propose two modifications of the empirical likelihood ratio: the mean corrected and the residual adjusted empirical likelihood ratios. Based on these results, we can build up confidence regions for the parameter of interest through a standard chi squared approximation. The rest of this chapter is organized as follows. In Section 1.2 we propose to construct the confidence bands for the unknown functions and their derivatives by using what we call a naive empirical likelihood technique. This technique shows as main drawback sub-optimal rates of convergence. In Section 1.3, as a byproduct, we provide two alternative maximum empirical likelihood estimators of the fixed effect nonparametric varying parameters model and their derivatives. In Section 1.4, and using the estimators that were previously derived, we propose two alternative techniques that enables us to obtain optimal nonparametric rates: Mean corrected and residual-adjusted empirical likelihood ratios. In Section 1.6 we provide a Monte Carlo experiment and in Section 1.7 we undertake an empirical study about the production efficiency of the European Union's companies. Finally Section 1.8 concludes. The proofs of the main results are collected in the Appendix.

1.2 Naive empirical likelihood

Consider the following varying coefficient panel data regression model

$$Y_{it} = X_{it}^{\top} m(Z_{it}) + \mu_i + v_{it}, \quad i = 1,...,N; t = 1,...,T, \tag{1.1}$$

where Y_{it} is the response, Z_{it} and X_{it} are vectors of covariates of dimension q and d respectively, and $m(z) = (m_1(z),...,m_d(z))^{\top}$ is a $d \times 1$ vector of unknown functions; here μ_i stands for heterogeneity of unknown form, that is, individual characteristic that are not observed, and v_{it} are random errors that do variate along time and across individuals. On this econometric model we impose the following standard assumptions,

Assumption 1.2.1. Let $(Y_{it}, X_{it}, Z_{it})_{i=1,...,N;\, t=1,..,T}$ be a set of independent and identically distributed (i.i.d.) $\mathbb{R}^{d+q+1}$ random variables in the subscript i for each fixed t and strictly stationary over t for a fixed i.

Assumption 1.2.2. The random errors v_{it} are independent and identically distributed, with 0 mean and homoscedastic variance $\sigma_v^2 < \infty$. They are also independent of X_{it} and Z_{it} for all i and t. Furthermore, $E|v_{it}|^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption 1.2.3. Let μ_i can be arbitrarily correlated with both X_{it} and Z_{it} with unknown correlation structure.

Assumptions 1.2.1, 1.2.2 and 1.2.3 are rather standard assumptions in the panel data literature. Assumption 1.2.1 is standard in panel data models; we could consider other settings as in Cai and Li (2008), however, since in this chapter we study the asymptotic properties as N tends to infinity and T is fixed, it is enough to assume stationarity. These type of models where T is fixed and N tends to infinity have been proved useful in the analysis of efficiency, where usually there is a large number of individuals during a small period of time. Assumption 1.2.2 is also standard for the conventional within and first difference transformation (Wooldridge (2010) or Hsiao (2014) for the fully parametric case). Independence between the idiosyncratic error and the covariates X_{it} and / or Z_{it} can be assumed without loss of generality, however it can be relaxed assuming some dependence in higher moments. If we allow some dependence, we could transform this estimator to take into consideration more complex structures of the random error contained in the variance-covariance matrix (Martins-Filho and Yao, 2009). Assumptions 1.2.1 and 1.2.2 in some situations, as in Cai and Li (2008), are relaxed by considering that (X_{it}, Z_{it}, v_{it}) are for fixed, i, strictly stationary processes; unfortunately, this set of assumptions is not sufficient to bound the asymptotic variance of the estimator and some further mixing conditions are required to achieve convergence. In this case, T must also tend to infinity. Other cases such as cross sectional dependence also requires both N and T tending to infinity. Finally, assumption 1.2.3 imposes the so called fixed effects; note that we are not willing to assume any constraint in the relationship between the individual heterogeneity μ and the vector of covariates (X, Z).

Rather than focusing in the consistent estimation of $m(z)$ and its vector of derivatives, we will obtain confidence bands for those objects based on the empirical likelihood principle. As already stated in the introductory section above, this approach presents clear advantages against the standard asymptotically aproximated confidence bands. To make the argument for constructing the confidence regions for $m(z)$ and its derivatives we can start by noting that, for a given z, from model (1.1) we have that

$$E\left[X_{it}\left(Y_{it} - X_{it}^{\top} m(Z_{it})\right)\middle| Z_{it} = z\right] \neq 0, \tag{1.2}$$

because of the fixed effects. Therefore, the least-squares estimator of $m(z)$ would be asymptotically biased. In order to cope with this problem, several transformations have been proposed in the standard literature of panel data models. Among them, we can take the so called within transformation. Then we have indeed that,

$$E\left[\ddot{X}_{it}\left(\ddot{Y}_{it} - X_{it}^{\top} m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T} X_{is}^{\top} m(Z_{is})\right)\middle| Z_{i1} = z, \ldots, Z_{iT} = z\right] = 0, \tag{1.3}$$

where $\ddot{X}_{it} = X_{it} - \bar{X}_{i.}$, $\bar{X}_{i.} = T^{-1}\sum_{s=1}^{T} X_{is}$ and $\ddot{Y}_{it} = Y_{it} - \bar{Y}_{i.}$, $\bar{Y}_{i.} = T^{-1}\sum_{s=1}^{T} Y_{is}$. Other transformations are available, for example the so called first differences transformation ends up in the following moment condition,

$$E\left[\Delta X_{it}\left(\Delta Y_{it} - \left(X_{it}^{\top} m(Z_{it}) - X_{i(t-1)}^{\top} m(Z_{i(t-1)})\right)\right)\middle| Z_{it} = z, Z_{i(t-1)} = z\right] = 0. \tag{1.4}$$

In both cases the least squares estimator of $m(z)$ is the solution to either (1.3) or (1.4). If we approximate the unknown function $X_{it}^{\top} m(Z_{it})$ around a value z that is in a close neighborhood of Z_{it} by a linear function $X_{it}^{\top} m(z) + X_{it}^{\top} \otimes (Z_{it} - z)^{\top} vec(D_m(z))$, then the ortogonality conditions (1.3) and (1.4) are approximated respectively by

$$E\left[\tilde{Z}_{it}^{*}\left(\ddot{Y}_{it}-\tilde{Z}_{it}^{*\top}\beta(z)\right)\Big|Z_{i1}=z,\ldots,Z_{iT}=z\right]=0, \tag{1.5}$$

and

$$E\left[\tilde{Z}_{it}\left(\Delta Y_{it}-\tilde{Z}_{it}^{\top}\beta(z)\right)\Big|Z_{it}=z,Z_{i(t-1)}=z\right]=0, \tag{1.6}$$

where $\tilde{Z}_{it}^{\top}=\left(\Delta X_{it}^{\top},X_{it}^{\top}\otimes(Z_{it}-z)^{\top}-X_{i(t-1)}^{\top}\otimes(Z_{i(t-1)}-z)^{\top}\right)$ is $1\times d(q+1)$ vector, $\beta(z)=(m(z),vec(D_m(z)))^{\top}$ is a $d(q+1)\times 1$ vector, and

$$\tilde{Z}_{it}^{*\top}=\left(\ddot{X}_{it}^{\top},X_{it}^{\top}\otimes(Z_{it}-z)^{\top}-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}\otimes(Z_{is}-z)\right)$$

is also a $d(q+1)\times 1$ vector. Also let, $D_m(z)$ be a $d\times q$ matrix of partial derivatives of the $d\times 1$ function $m(z)$ with respect to the elements of the $q\times 1$ vector z, i.e. $D_m(z)=\frac{\partial m(z)}{\partial z}$. Note that equations (1.5) and (1.6) are the first order conditions of the minimization problem $E\left[\left(\ddot{Y}_{it}-\tilde{Z}_{it}^{*\top}\beta(z)\right)^2\Big|z\right]$ and $E\left[\left(\Delta Y_{it}-\tilde{Z}_{it}^{\top}\beta(z)\right)^2\Big|z\right]$ for a given z. Because nonparametric conditional expectations given either $(Z_{i1},\ldots,Z_{iT})$ in (1.5) or $\left(Z_{it},Z_{i(t-1)}\right)$ in (1.6) are involved, a local smoothing method is needed to obtain the sample version of those equations. In order to define the empirical likelihood estimator we employ equation (1.5) or (1.6) as auxiliary random vectors; therefore, the auxiliary random vector for the within transformation is as follows

$$T_{wi}(\beta(z))=\sum_{t=1}^{T}\tilde{Z}_{it}^{*}[\ddot{Y}_{it}-\tilde{Z}_{it}^{*\top}\beta(z)]K_H(Z_{i1}-z)\cdots K_H(Z_{iT}-z), \tag{1.7}$$

and for the first differences transformation

$$T_{fi}(\beta(z))=\sum_{t=2}^{T}\tilde{Z}_{it}[\Delta Y_{it}-\tilde{Z}_{it}^{\top}\beta(z)]K_H(Z_{it}-z)K_H(Z_{i(t-1)}-z). \tag{1.8}$$

In equations (1.7) and (1.8) H is a bandwidth matrix of dimension $q\times q$, $K(\cdot)$ denotes a kernel function in $\mathbb{R}^q$ and

$$K_H(u)=K\left(H^{-1/2}u\right).$$

Note that the $T_{w1}(\beta(z)),\ldots,T_{wN}(\beta(z))$ are independent and, due to assumption 1.2.2, $E(T_{wi})=0$; the same implications remain valid for T_{fi}. Therefore, a naive empirical likelihood ratio function for $m(z)$ and $D_m(z)$ can be defined as the solution to the maximization problem of a multinomial log-likelihood function, i.e.

$$\mathscr{R}_w(\beta(z)) = -2\max\left\{\sum_{i=1}^{N}\log(p_i)\,\middle|\, p_i \geq 0,\ \sum_{i=1}^{N} p_i = 1,\ \sum_{i=1}^{N} p_i T_{wi}(\beta(z)) = 0\right\}, \tag{1.9}$$

where the probabilities $p_i = p_i(z)$, for $i = 1,...,N$. There exists a unique value of $\mathscr{R}_w(\beta(z))$, for a given $\beta(z)$, provided that 0 is inside the convex hull of $(T_{w1}(\beta(z)),...,T_{wN}(\beta(z)))$ (Owen, 1990, 1988). Using the Lagrange multiplier method the probabilities p_i are

$$p_i = \frac{1}{N}\frac{1}{\left(1+\lambda^\top T_{wi}(\beta(z))\right)}.$$

Note that it is necessary that $0 \leq p_i \leq 1$ which implies that λ and $\beta(z)$ must satisfy that $1+\lambda^\top T_{wi}(\beta(z)) \geq N^{-1}$ for each i (see, Owen (2001), Chapter 3). This constraint satisfies the non-negativity condition and it avoids a convex dual problem.

Using p_i's expression and after some calculations equation (1.9) leads to

$$\mathscr{R}_w(\beta(z)) = 2\sum_{i=1}^{N}\log(1+\lambda^\top T_{wi}(\beta(z))), \tag{1.10}$$

where λ is a $d(q+1)\times 1$ vector associated to the constraint $\sum_{i=1}^{N} p_i T_{wi}(\beta(z)) = 0$. It is indeed given as the solution to

$$\sum_{i=1}^{N}\frac{T_{wi}(\beta(z))}{1+\lambda^\top T_{wi}(\beta(z))} = 0. \tag{1.11}$$

Let us now denote $\tilde{D}_w(\beta(z)) = (NT|H|^{T/2})^{-1}\sum_{i=1}^{N} T_{wi}(\beta(z))T_{wi}^\top(\beta(z))$. Using equations (1.10), (1.11) and a Taylor expansion, it can be shown that

$$\mathscr{R}_w(\beta(z)) = \left[\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N} T_{wi}(\beta(z))\right]^\top \left[\tilde{D}_w(\beta(z))\right]^{-1}\left[\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N} T_{wi}(\beta(z))\right] + o_p(1). \tag{1.12}$$

Hence, as expected, $\mathscr{R}_w(\beta(z))$ is asymptotically a standard Chi-squared distribution. To state formally these results, we first introduce some notations and assumptions.

Assumption 1.2.4. The Kernel functions $K(.)$ are compactly supported and bounded kernels such that $\int K(u)du = 1$, $\int uu^\top K(u)du = \mu_2(K_u)I$, and $\int K(u)^2 du = R(K_u)$ where $\mu_2(K_u) \neq 0$, and $R(K_u) \neq 0$ are scalars and I is a $q\times q$ identity matrix. Besides, we will assume that there exist eight-order marginal moment for $K(.)$, i.e.,

$$\int u_1^8 K(u_1,..,u_T)du_1,...,du_T < \infty.$$

Also, the odd-order moments of K, when they exist, are zero, i.e.,

$$\int u_1^{i_1}u_2^{i_2},...,u_T^{i_T}K(u_1,..,u_T)du_1,...,du_T = 0 \quad \text{if} \quad \sum_{j=1}^{T} i_j \quad \text{is odd.}$$

Assumption 1.2.5. Let $f_{Z_{it}}(.)$, $f_{Z_{it},Z_{i(t-1)}}(.,.)$ and $f_{Z_{i1},Z_{i2},Z_{i3}}(.,.,.)$, for $t = 1,..,T$ be respectively the probability density functions of Z_{it}, $(Z_{it},Z_{i(t-1)})$ and (Z_{i1},Z_{i2},Z_{i3}). All density functions are continuously differentiable in all their arguments and they are bounded from above and bellow in any point of their support.

Assumption 1.2.6. Let z be an interior point of $f_{Z_{it}}$. Besides, the third order derivatives of $m_1(.),...,m_d(.)$ are bounded and uniformly continuous.

Assumption 1.2.7. The bandwidth matrix H is symmetric and strictly definite positive. Moreover, each entry of the matrix tends to zero as $N \to \infty$ in such a way that $N|H| \to \infty$.

Assumption 1.2.8. The function $E[\ddot{X}_{it}\ddot{X}_{it}^{\top}|Z_{i1} = z_1,...,Z_{iT} = z_T]$ is positive definite for any interior point of $(z_1,z_2,..z_T)$ in the support of $f_{Z_{i1},...,Z_{iT}}(z_1,z_2,...,z_T)$.

Assumption 1.2.9. Let $||A|| = \sqrt{\text{tr}\left(A^{\top}A\right)}$, then $E[\left|\left|X_{it}X_{it}^{\top}\right|\right|^2|Z_{i1} = z,..,Z_{iT} = z]$ is bounded and uniformly continuous in its support. Furthermore, let the following matrix functions $E[\ddot{X}_{it}X_{it}^{\top}|Z_{i1} = z,..,Z_{iT} = z]$, $E[X_{it}X_{it}^{\top}|Z_{it} = z,\ Z_{i(t-1)} = z]$ and $E[X_{is}X_{is}^{\top}|Z_{i1} = z,...,Z_{iT} = z]$ be bounded and uniformly continuous in their support. Also, $E[\ddot{X}_{it}X_{is}^{\top}|Z_{i1} = z,...,Z_{iT} = z]$ and $E[X_{it}X_{is}^{\top}|Z_{i1} = z,...,Z_{iT} = z]$, for $t \neq s$ and $t = s$, are bounded and uniformly continuous in their support.

Assumption 1.2.10. Let the following functions be bounded and uniformly continuous in any point of its support, $E[|X_{it}v_{it}|^{2+\delta}|Z_{it} = z,\ Z_{i(t-1)=z}]$, $E[|X_{is}v_{it}|^{2+\delta}|Z_{it} = z,\ Z_{i(t-1)=z}]$ and, $E[|\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{it} = z,\ Z_{i(t-1)=z}]$, for some $\delta > 0$.

These assumptions are rather common in the literature of non-parametric regression analysis of panel data models. Similar conditions were used in Xue and Zhu (2007), Su et al. (2013), Rodriguez-Poo and Soberón (2014, 2015). They are basically smoothness and boundedness conditions for the within estimator. There are also assumptions about the kernel functions and about the behavior of the bandwidth matrix.

Under these assumptions, we are able to establish the following results.

Theorem 1.2.1. Assuming that conditions 1.2.1 - 1.2.10 hold and $H \to 0$ in such a way that $NT|H|^{T/2} \to \infty$ and $\sqrt{NT|H|^{T/2}}\mathrm{tr}(H) \to 0$, then $\mathscr{R}_w(\beta(z)) \to_d \chi^2_{d(q+1)}$ as $N \to \infty$ and T is fixed, where $\to_d$ means the convergence in distribution and $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.

Now, following exactly the same steps as for the within transformation and denoting

$$\tilde{D}_f(\beta(z)) = (NT|H|)^{-1} \sum_{i=1}^{N} T_{fi}(\beta(z)) T_{fi}^{\top}(\beta(z)),$$

we obtain

$$\mathscr{R}_f(\beta(z)) = \left[\frac{1}{\sqrt{NT|H|}} \sum_{i=1}^{N} T_{fi}(\beta(z))\right]^{\top} \left[\tilde{D}_f(\beta(z))\right]^{-1} \left[\frac{1}{\sqrt{NT|H|}} \sum_{i=1}^{N} T_{fi}(\beta(z))\right] + o_p(1), \quad (1.13)$$

and, as in the within case, using a non-parametric version of the Wilks' theorem we can provide that $\mathscr{R}_f(\beta(z))$ has, asymptotically, a Chi squared distribution. In fact, in order to show this result we need the following smoothness conditions on moment functional forms,

Assumption 1.2.11. Let $||A|| = \sqrt{\mathrm{tr}\left(A^{\top}A\right)}$, then the function $E[\Delta X_{it} \Delta X_{it}^{\top} | Z_{it} = z,\ Z_{i(t-1)} = z]$ is a positive definite for any interior point of (z,z) in the support of $f_{Z_{it}, Z_{i(t-1)}}(z,z)$.

Assumption 1.2.12. Also the following matrix functions $E[\Delta X_{it} X_{it}^{\top} | Z_{it} = z,\ Z_{i(t-1)} = z]$, $E[X_{it} X_{it}^{\top} | Z_{it} = z,\ Z_{i(t-1)} = z]$, $E[X_{i(t-1)} X_{i(t-1)}^{\top} | Z_{it} = z,\ Z_{i(t-1)} = z]$ and $E[\Delta X_{it} X_{i(t-1)}^{\top} | Z_{it} = z,\ Z_{i(t-1)} = z]$ are bounded and uniformly continuous in their support.

Assumption 1.2.13. The functions $E[|\Delta X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z,\ Z_{i(t-1)=z}]$, $E[|X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z,\ Z_{i(t-1)=z}]$ and $E[|X_{i(t-1)} \Delta v_{it}|^{2+\delta} | Z_{it} = z,\ Z_{i(t-1)=z}]$ for some $\delta > 0$, are bounded and uniformly continuous in any point of its support.

These group of conditions substitute assumptions 1.2.8 - 1.2.10 when working with the first differences technique. Then, we are able to show the following result.

Theorem 1.2.2. Assuming that conditions 1.2.1 - 1.2.7 and 1.2.11 - 1.2.13 hold and $H \to 0$ in such a way that $NT|H| \to \infty$ and $\sqrt{NT|H|}\mathrm{tr}(H) \to 0$, then $\mathscr{R}_f(\beta(z)) \to_d \chi^2_{d(q+1)}$ as $N \to \infty$ and T is fixed, where $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.

Using theorems 1.2.1 and 1.2.2 we can approximate α-level confidence regions for $\beta(z)$ as the set of values $\beta(z)$ such that $\mathscr{R}_f(\beta(z)) \leq c_\alpha$ and $\mathscr{R}_w(\beta(z)) \leq c_\alpha$, where c_α is defined such that $\mathbf{Pr}\left(\chi^2_{d(q+1)} \leq c_\alpha\right) = \alpha$.
In the following section we obtain the maximum empirical likelihood estimators using the empirical likelihood ratios defined in this section. Also, as the usual tool to construct confidence bands, we will provide the asymptotic distribution of the estimators.

1.3 Maximum empirical likelihood estimators

We can define the maximum empirical likelihood (MELE) estimator of $\beta(z)$, $\hat{\beta}_w(z)$ as the minimizer of $\mathscr{R}_w(\beta(z))$. From equations (1.10) and (1.12) and following the same lines as Qin and Lawless (1994), $\hat{\beta}_w(z)$ is obtained from the solution of the estimating equation $\left(NT|H|^{T/2}\right)^{-1}\sum_{i=1}^{N} T_{wi}(\beta(z)) = 0$ and, as it will be shown in the proof of Theorem 1.3.1, the remainder term is of smaller order tending to zero as $NT|H|^{T/2}$ tends to infinity. Consequently, the MELE is asymptotically equivalent to the fixed effect estimator using the within transformation. Therefore, if we assume that $\frac{1}{NT|H|^{T/2}}\sum_{it}\prod_{l=1}^{T} K_H(Z_{il}-z)\tilde{Z}_{it}^{*}\tilde{Z}_{it}^{*\top}$ is invertible, then the MELE is as follows

$$\begin{aligned} \hat{\beta}_w(z) &= \left(\frac{1}{NT|H|^{T/2}}\sum_{it}\prod_{l=1}^{T} K_H(Z_{il}-z)\tilde{Z}_{it}^{*}\tilde{Z}_{it}^{*\top}\right)^{-1}\frac{1}{NT|H|^{T/2}}\sum_{it}\prod_{l=1}^{T} K_H(Z_{il}-z)\tilde{Z}_{it}^{*}\ddot{Y}_{it} \\ &+ o_p\left(\frac{1}{\sqrt{NT|H|^{T/2}}}\right). \end{aligned} \tag{1.14}$$

As it has been already pointed out in other works, the leading terms in both bias and variance do not depend on the sample, and therefore we can consider such terms as playing the role of the unconditional bias and variance. For comparison purposes, and in order to build up confidence bands, we state the asymptotic distribution of the estimator in the following theorem.

Theorem 1.3.1. Assuming that conditions 1.2.1 - 1.2.10 hold and $H \to 0$ in such a way that $NT|H|^{T/2} \to \infty$, then

$$\sqrt{NT|H|^{T/2}}\left\{\hat{\beta}_w(z) - \beta(z) - B_w(z)\right\} \to_d \mathscr{N}\left(0, \Sigma_w(z)\right),$$

where

$$\begin{aligned} B_w(z) = \operatorname{diag}&\left\{I_d, \left[\left(1-\frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,\ldots,z)\otimes\mu_2(K_{u_\tau})H\right]^{-1}\right\} \\ &\times\begin{pmatrix} \frac{1}{2}\mu_2(K_{u_\tau})\operatorname{diag}_d\left\{\operatorname{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\} i_d \\ \frac{1}{2}\mu_2(K_{u_\tau})^2 B_{w1}(z) + \frac{1}{3!}B_{w2}(z) \end{pmatrix}, \end{aligned}$$

and

$$\Sigma_w(z) = \sigma_v^2 \text{diag}\left\{I_d, \left[\left(1-\frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,...,z)\otimes\mu_2(K_{u_\tau})H\right]^{-1}\right\}\begin{pmatrix}\Sigma_{w1}(z) & 0 \\ 0 & \Sigma_{w2}(z)\end{pmatrix}$$

$$\times\text{diag}\left\{I_d, \left[\left(1-\frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,...,z)\otimes\mu_2(K_{u_\tau})H\right]^{-1}\right\},$$

where τ is any index between 1 and T. Also, let

$$\begin{aligned}
B_{w1}(z) &= \left(1-\frac{1}{T}\right)\mathscr{D}\mathscr{B}^{\top}_{X_tX_t}(z,...,z)\text{diag}_d\left\{\text{tr}\left\{\mathscr{H}_{m_r}(z)H^2\right\}\right\}i_d \\
&- \left[\mathscr{D}\mathscr{B}_{\ddot{X}\ddot{X}}(z,...,z)\left(I_d\otimes H\right)\right]^{\top}\text{diag}_d\left\{\text{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\}i_d, \\
B_{w2}(z) &= \left(1-\frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,...,z)\otimes\int\left(H^{1/2}u_\tau\right)D_m^3(z,H^{1/2}u_\tau)\prod_{l=1}^{T}K(u_l)du_l, \\
\Sigma_{w1}(z) &= \mathscr{B}^{-1}_{\ddot{X}\ddot{X}}(z,...,z)R(K)^T, \\
\Sigma_{w2}(z) &= \left[\mathscr{D}\mathscr{B}_{\ddot{X}\ddot{X}}(z,...,z)\left(I_d\otimes H\right)\right]^{\top}\mathscr{B}^{-1}_{\ddot{X}\ddot{X}}(z,...,z)R(K)^T\left[\mathscr{D}\mathscr{B}_{\ddot{X}\ddot{X}}(z,...,z)\left(I_d\otimes H\right)\right], \\
\mathscr{B}_{\ddot{X}\ddot{X}}(z,...,z) &= E\left[\ddot{X}_{it}\ddot{X}_{it}^{\top}\middle|Z_{i1}=z,...,Z_{iT}=z\right]f_{Z_{i1},...,Z_{iT}}(z,...,z), \\
\mathscr{B}_{X_tX_t}(z,...,z) &= E\left[X_{it}X_{it}^{\top}\middle|Z_{i1}=z,...,Z_{iT}=z\right]f_{Z_{i1},...,Z_{iT}}(z,...,z).
\end{aligned}$$

Here, $\mathscr{D}\mathscr{B}_{\ddot{X}\ddot{X}}(z,...,z)$ and $\mathscr{D}\mathscr{B}_{X_tX_t}(z,...,z)$ are $d\times dq$ gradient matrix of the form

$$\mathscr{D}\mathscr{B}_{X_tX_t}(z_1,..,z_T) = \begin{pmatrix} \frac{\partial b_{11}^{X_tX_t}(z_1,...,z_T)}{\partial z_1} & \cdots & \frac{\partial b_{1d}^{X_tX_t}(z_1,...,z_T)}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{d1}^{X_tX_t}(z_1,...,z_T)}{\partial z_1} & \cdots & \frac{\partial b_{dd'}^{X_tX_t}(z_1,...,z_T)}{\partial z_1}\end{pmatrix},$$

and

$$b_{dd'}^{X_tX_t}(z_1,...,z_T) = E\left[X_{dit}X_{d'it}|Z_{i1}=z_1,...,Z_{iT}=z_T\right]f_{Z_{i1},...,Z_{iT}}(z_1,...,z_T),$$

$\text{diag}_d\left\{\text{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\}$ stands for a diagonal matrix of elements $\text{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}$, for $r=1,...,d$, where $\mathscr{H}_{m_r}$ is the Hessian matrix of the rth component of $m(.)$ and $D_m^3(z,Z_{it}-z)$ has as general expression, for $k=3$,

$$D_m^k(z,u) = \sum_{i_1,...,i_q} C^k_{i_1,...,i_q}\frac{\partial^k m(z)}{\partial z_1^{i_1},\ldots,\partial z_q^{i_q}}u_1^{i_1},\ldots,u_q^{i_q},$$

where the sums are over all distinct nonnegative integers $i_1,\ldots,i_q$, such that $i_1+\ldots+i_q=k$, and $C^k_{i_1,...,i_q}=k!/(i_1!\ldots i_q!)$. Finally we denote by i_d a $d\times 1$ unit vector

Similarly, if we assume that $\frac{1}{NT|H|}\sum_{it} K_H(Z_{it}-z)K_H(Z_{i(t-1)}-z)\tilde{Z}_{it}\tilde{Z}_{it}^{\top}$ is invertible, we can define the MELE for the first difference approach, $\hat{\beta}_f(d)$, write

$$\begin{aligned}\hat{\beta}_f(z) &= \left(\frac{1}{NT|H|}\sum_{it} K_H(Z_{it}-z)K_H(Z_{i(t-1)}-z)\tilde{Z}_{it}\tilde{Z}_{it}^{\top}\right)^{-1} \\ &\quad\times\frac{1}{NT|H|}\sum_{it} K_H(Z_{it}-z)K_H(Z_{i(t-1)}-z)\tilde{Z}_{it}\Delta Y_{it} + o_p\left(\frac{1}{\sqrt{NT|H|}}\right), \quad (1.15)\end{aligned}$$

where the asymptotic normality of the estimator is as follows

Theorem 1.3.2. Assuming that conditions 1.2.1 - 1.2.7 and 1.2.11 - 1.2.13 hold and $H \to 0$ in such a way that $NT|H| \to \infty$, then

$$\sqrt{NT|H|}\left\{\hat{\beta}_f(z)-\beta(z)-B_f(z)\right\} \to_d \mathscr{N}\left(0,\Sigma_f(z)\right)$$

where,

$$\begin{aligned}B_f(z) &= \mathrm{diag}\left\{I_d, \left[\left(\mathscr{B}_{XX}(z,z)+\mathscr{B}_{X_{-1}X_{-1}}(z,z)\right)\otimes\mu_2(K_u)H\right]^{-1}\right\} \\ &\quad\times\begin{pmatrix}\frac{1}{2}\mu_2(K_u)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\}i_d \\ \frac{1}{2}\mu_2(K_u)^2 B_{f1}(z)+\frac{1}{3!}B_{f2}(z)\end{pmatrix},\end{aligned}$$

and

$$\begin{aligned}\Sigma_f(z) &= 2\sigma_v^2\mathrm{diag}\left\{I_d, \left[\left(\mathscr{B}_{XX}(z,z)+\mathscr{B}_{X_{-1}X_{-1}}(z,z)\right)\otimes\mu_2(K_u)H\right]^{-1}\right\}\begin{pmatrix}\Sigma_{f1}(z) & 0 \\ 0 & \Sigma_{f2}(z)\end{pmatrix} \\ &\quad\times\mathrm{diag}\left\{I_d, \left[\left(\mathscr{B}_{XX}(z,z)+\mathscr{B}_{X_{-1}X_{-1}}(z,z)\right)\otimes\mu_2(K_u)H\right]^{-1}\right\};\end{aligned}$$

also, let

$$
\begin{aligned}
B_{f1}(z) &= \left(\mathscr{D}\mathscr{B}_{XX}(z,z) - \mathscr{D}\mathscr{B}_{X_{-1}X_{-1}}(z,z)\right)^{\top} \operatorname{diag}_d \left\{\operatorname{tr}\left\{\mathscr{H}_{m_r}(z)H^2\right\}\right\} i_d \\
&- \left[\mathscr{D}\mathscr{B}_{\Delta X \Delta X}(z,\ldots,z)\left(I_d \otimes H\right)\right]^{\top} \operatorname{diag}_d \left\{\operatorname{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\} i_d, \\
B_{f2}(z) &= \left(\mathscr{B}_{XX}(z,z) - \mathscr{B}_{X_{-1}X_{-1}}(z,z)\right) \int \left(H^{1/2}u\right) D_m^3(z, H^{1/2}u) K(u) K(v) du dv, \\
\Sigma_{f1}(z) &= \mathscr{B}_{\Delta X \Delta X}(z,z) R(K_u) R(K_v), \\
\Sigma_{f2}(z) &= \left[\mathscr{D}\mathscr{B}_{\Delta X \Delta X}(z,z)\left(I_d \otimes \mu_2(K_u) H\right)\right]^{\top} \mathscr{B}^{-1}_{\Delta X \Delta X}(z,z) R(K_u) R(K_v) \\
&\times \left[\mathscr{D}\mathscr{B}_{\Delta X \Delta X}(z,z)\left(I_d \otimes \mu_2(K_u) H\right)\right], \\
\mathscr{B}_{\Delta X \Delta X}(z,z) &= E\left[\Delta X_{it} \Delta X_{it}^{\top} \middle| Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z), \\
\mathscr{B}_{XX}(z,z) &= E\left[X_{it} X_{it}^{\top} \middle| Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z), \\
\mathscr{B}_{X_{-1}X_{-1}}(z,z) &= E\left[X_{i(t-1)} X_{i(t-1)}^{\top} \middle| Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it}, Z_{i(t-1)}}(z,z).
\end{aligned}
$$

Here, $\mathscr{D}\mathscr{B}_{\Delta X \Delta X}$, $\mathscr{D}\mathscr{B}_{XX}(z,z)$ and $\mathscr{D}\mathscr{B}_{X_{-1}X_{-1}}(z,z)$ are $d \times dq$ gradient matrices defined as in theorem 1.3.1.

The results shown in Theorems 1.3.1 and 1.3.2 somehow correspond, under a different setting, to Theorem 3.1 in Rodriguez-Poo and Soberón (2015) and Theorems 3.1 in Rodriguez-Poo and Soberón (2014) respectively. However, we point out that the results obtained for the vector of derivatives are fully new in this fixed effects panel data setting. An interesting issue that needs to be considered here is the relative asymptotic efficiency of these estimators. Note first, that as the reader surely realizes none of these estimators achieve the optimal rate of convergence in terms of the Mean Integrated Square Error (MISE). Indeed, for this type of problems the optimal rate is $1/NT|H|^{1/2}$ (see, Fan (1993) for details). For the estimator based in the within transformation the rate of convergence in terms of the MISE (see, Theorem 1.3.1) is $1/NT|H|^{T/2}$, whereas for the estimator based in the first differences transformation (see, Theorem 1.3.2) it is $1/NT|H|$. Therefore, the rate of convergence of both Empirical Maximum Likelihood Estimators is suboptimal. However, note that the relative asymptotic efficiency of $\widehat{\beta}_f(z)$ with respect to $\widehat{\beta}_w(z)$ with the same bandwidths is of order $O\left(|H|^{\frac{T}{2}-1}\right)$. If $T > 2$ then $\widehat{\beta}_f(z)$ will exhibit a faster rate of convergence than $\widehat{\beta}_w(z)$. Indeed as far as T gets larger this difference in rates increases. This is due to the so-called curse of dimensionality that is more serious in the case of the estimator based in the within transformation. In fact, in the case of $\widehat{\beta}_f(z)$ we use a kernel function of dimension $2 \times q$ whereas for the other estimator the dimension is $T \times q$. Finally, as an example, consider the estimation of $m(\cdot)$ using both estimators. Using Theorems 1.3.2 and 1.3.1 and some standard calculations note that the bandwidth that minimizes the MISE for the estimator based in the within transformation is of order $(NT)^{-\frac{1}{4+qT}}$ whereas for the estimator based in the first differences transformation converges to zero at the rate $(NT)^{-\frac{1}{4+2q}}$. Substituting these optimal bandwidths in the asymptotic MISE expressions we obtain the following convergence rates: $(NT)^{-\frac{4}{4+qT}}$ for the within estimator, and $(NT)^{-\frac{2}{2+q}}$ for the first differences estimator.

1.4 Bias corrected empirical likelihood

In fact, note that in order to show the convergence of both theorems, theorem 1.2.1 and theorem 1.2.2, we have included one extra condition on the asymptotic behavior of the sequence of bandwidth matrices, i.e. $\sqrt{NT|H|^{T/2}}\text{tr}(H) \to 0$ for the within estimator and $\sqrt{NT|H|}\text{tr}(H) \to 0$ for the first differences transformation. These additional conditions ensure that the smoothness bias becomes negligible as the sample size tends to infinity. Unfortunately, these conditions on H exclude the bandwidth matrix that is optimal, therefore this will end up in suboptimal rates of convergence for both $\mathscr{R}_w(\beta(z))$ and $\mathscr{R}_f(\beta(z))$. In order to avoid this problem we propose two modifications of the Empirical Likelihood ratio that remove the bias term: the Mean-corrected Empirical Likelihood (MCEL) ratio and the Residual-Adjusted Empirical Likelihood (RAEL) ratio. These bias corrections have already been proposed in Xue and Zhu (2007) and what we will do here is to adapt them to our panel data with fixed effect setting.

1.4.1 Mean-corrected empirical likelihood ratio

As we have already pointed out, if H tends to zero at the optimal rate then $\mathscr{R}_w(\beta(z))$ will not converge in distribution to a χ^2 random variable. The main reason is that the smoothness bias will not vanish as $NT|H|^{T/2}$ tends to infinity. However, from the proof of Theorem 1.2.1 we know that, under the assumptions established in theorem 1.2.1, $\sqrt{NT|H|^{T/2}}\left(\frac{1}{NT|H|^{T/2}}\sum_i T_{wi}(\beta(z)) - b_w(z)\right) \to_d \mathscr{N}(0, \upsilon_w(z))$, as $NT|H|^{T/2}$ tends to infinity. Here

$$b_w(z) = \begin{pmatrix} \frac{1}{2}b_{w1}(z) \\ \frac{1}{2}b_{w2}(z) + \frac{1}{3!}b_{w3}(z) \end{pmatrix}, \tag{1.16}$$

and

$$\upsilon_w(z) = \sigma_v^2 d \begin{pmatrix} R(K)^T \mathscr{B}_{\ddot{X}\ddot{X}}(z,...,z) & 0 \\ 0 & \left(1-\frac{1}{T}\right)\mu_2(K_{u_\tau}^2)\prod_{l\neq\tau}^T R(K_{u_l})\mathscr{B}_{X_tX_t}(z,...,z)\otimes H \end{pmatrix}, \tag{1.17}$$

where

$$\begin{aligned}
b_{w1}(z) &= \mu_2(K_{u_\tau})\mathscr{B}_{\ddot{X}\ddot{X}}(z,..,z)\text{diag}_d\left\{\text{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\}i_d,\\
b_{w2}(z) &= \mu_2(K_{u_\tau})^2\left(1-\frac{1}{T}\right)\mathscr{D}\mathscr{B}_{X_tX_t}^{\top}(z,..,z)\text{diag}_d\left\{\text{tr}\left\{\mathscr{H}_{m_r}(z)H^2\right\}\right\}i_d,\\
b_{w3}(z) &= \left(1-\frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,..,z)\otimes\int\left(H^{1/2}u_\tau\right)D_m^3(z,H^{1/2}u_\tau)\prod_{l=1}^{T}K(u_l)du_l.
\end{aligned}$$

Hence, the first proposal is to correct the empirical likelihood ratio, $\mathscr{R}_w(\beta(z))$, by the smoothing bias, given by $\sqrt{NT|H|^{T/2}}\,b_w(z)$. In order to do so we need a consistent estimator of $b_w(z)$. By noting that

$$\frac{1}{|H|^{T/2}}E\left[\tilde{Z}_{it}^*\left(X_{it}^{\top}m(Z_{it})-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}m(Z_{is})-\tilde{Z}_{it}^{*\top}\beta(z)\right)\prod_{l=1}^{T}K_H(Z_{il}-z)\right]=b_w(z)+o_p(1),$$

(see, (A.8) for details) then a consistent estimator of $b_w(z)$ can be naturally defined as

$$\hat{b}_w(z)=\frac{1}{NT|H|^{T/2}}\sum_{it}\tilde{Z}_{it}^*\left(X_{it}^{\top}\hat{m}_w(Z_{it})-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}\hat{m}_w(Z_{is})-\tilde{Z}_{it}^{*\top}\hat{\beta}_w(z)\right)\prod_{l=1}^{T}K_H(Z_{il}-z),\quad(1.18)$$

where $\hat{\beta}^w(z)$ is the MELE defined in (2.18), $\hat{m}_w(z)=e^{\top}\hat{\beta}^w(z)$, and $e=\begin{bmatrix}I_d & \vdots & \mathbf{0}\end{bmatrix}$, I_d is a d-dimensional unit matrix and $\mathbf{0}$ is a $dq\times d$ matrix. Taking into account (1.18), let us denote

$$\begin{aligned}
\tilde{\xi}_w(\beta(z)) &= \sqrt{NT|H|^{T/2}}\,\hat{b}_w(z)^{\top}\left[\tilde{D}_w(\beta(z))\right]^{-1}\\
&\quad\times\left[\frac{2}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N}T_{wi}(\beta(z))-\sqrt{NT|H|^{T/2}}\,\hat{b}_w(z)\right].
\end{aligned}$$

Finally, the mean-corrected empirical likelihood for $\beta(z)$ will be

$$\tilde{\mathscr{R}}_w(\beta(z))=\mathscr{R}_w(\beta(z))-\tilde{\xi}_w(\beta(z)).\qquad(1.19)$$

Similarly, for the first differences transformation, we can define the mean-corrected empirical likelihood as

$$\tilde{\mathscr{R}}_f(\beta(z))=\mathscr{R}_f(\beta(z))-\tilde{\xi}_f(\beta(z)),\qquad(1.20)$$

where

$$\tilde{\xi}_f(\beta(z))=\sqrt{NT|H|}\,\hat{b}_f(z)^{\top}\left[\tilde{D}_f(\beta(z))\right]^{-1}\left[\frac{2}{\sqrt{NT|H|}}\sum_{i=1}^{N}T_{fi}(\beta(z))-\sqrt{NT|H|}\,\hat{b}_f(z)\right].$$

Also, $\hat{b}_f(z)$ is a consistent estimator of $b_f(z)$. In this case, it is easy to show (see, (A.19) for details) that

$$\frac{1}{|H|}E\left[\tilde{Z}_{it}\left(X_{it}^\top m(Z_{it})-X_{i(t-1)}^\top m(Z_{i(t-1)})-\tilde{Z}_{it}^\top\beta(z)\right)K_H(Z_{it}-z,Z_{i(t-1)}-z)\right]=b_f(z)+o_p(1),$$

then the estimator of $b_f(z)$, $\hat{b}_f(z)$ is

$$\frac{1}{NT|H|}\sum_{it}\tilde{Z}_{it}\left(X_{it}^\top\hat{m}_f(Z_{it})-X_{i(t-1)}^\top\hat{m}_f(Z_{i(t-1)})-\tilde{Z}_{it}^\top\hat{\beta}_f(z)\right)K_H\left(Z_{it}-z\right)K_H\left(Z_{i(t-1)}-z\right),$$

where $\hat{\beta}^f(z)$ is the MELE defined in (1.15), and $\hat{m}_f(z)=e^\top\hat{\beta}^w(z)$. Note that from the proof of theorem 1.2.2,

$$b_f(z)=\begin{pmatrix}\frac{1}{2}b_{f1}(z)\\ \frac{1}{2}b_{f2}(z)+\frac{1}{3!}b_{f3}(z)\end{pmatrix}, \tag{1.21}$$

where

$$\begin{aligned}
b_{f1}(z) &= \mu_2(K_u)\mathscr{B}_{\Delta X\Delta X}(z,z)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\}i_d,\\
b_{f2}(z) &= \mu_2(K_u)^2\left(\mathscr{D}\mathscr{B}_{XX}^\top(z,z)-\mathscr{D}\mathscr{B}_{X_{-1}X_{-1}}^\top(z,z)\right)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathscr{H}_{m_r}(z)H^2\right\}\right\}i_d,\\
b_{f3}(z) &= \left(\mathscr{B}_{XX}(z,z)-\mathscr{B}_{X_{-1}X_{-1}}(z,z)\right)\otimes\int\left(H^{1/2}u_\tau\right)D_m^3(z,H^{1/2}u_\tau)K(u)K(v)dudv.
\end{aligned}$$

We state the asymptotic results of these two MCEL ratios in the following theorem.

Theorem 1.4.1. Assuming that conditions 1.2.1 - 1.2.13 hold, and $\beta(z)$ is the true vector of parameters, then $\tilde{\mathscr{R}}_w(\beta(z))\to_d\chi^2_{d(q+1)}$ and $\tilde{\mathscr{R}}_f(\beta(z))\to_d\chi^2_{d(q+1)}$ as $N\to\infty$ and T is fixed, where $\to_d$ means the convergence in distribution and $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.

Note that to state this result we do not impose any extra condition. Here, we need conditions 1.2.1 - 1.2.10 to ensure that $\tilde{\mathscr{R}}_w(\beta(z))\to_d\chi^2_{d(q+1)}$ and conditions 1.2.1 - 1.2.7 and 1.2.11 - 1.2.13 to ensure that $\tilde{\mathscr{R}}_f(\beta(z))\to_d\chi^2_{d(q+1)}$ as $N\to\infty$. Basically these conditions are the same conditions of Theorems 1.2.1 and 1.2.2; however we do not need that $\sqrt{NT|H|^{T/2}}\mathrm{tr}(H)\to 0$ for the within transformation and $\sqrt{NT|H|}\mathrm{tr}(H)\to 0$ for the first differences transformation.

1.4.2 Residual-adjusted empirical likelihood ratio

There exist an alternative method to the MCEL in order to cope with the asymptotic bias. The main idea is to borrow the asymptotic expansion of the empirical likelihood ratio already derived. That is, for the within transformation, let $\hat{T}_{wi}(\beta(z))$ be an adjustment of the weighted residuals, $T_{wi}(\beta(z))$, that is defined as

$$\sum_{t=1}^{T} \tilde{Z}_{it}^{*} \left[\ddot{Y}_{it} - \tilde{Z}_{it}^{*\top} \beta(z) - \left(X_{it}^{\top} \hat{m}_w(Z_{it}) - \frac{1}{T} \sum_{s=1}^{T} X_{is}^{\top} \hat{m}_w(Z_{is}) - \tilde{Z}_{it}^{*\top} \hat{\beta}_w(z) \right) \right] \prod_{l=1}^{T} K_H(Z_{il} - z).$$

Similarly, for the first differences transformation we have that $\hat{T}_{fi}(\beta(z))$ is defined as

$$\sum_{t=2}^{T} \tilde{Z}_{it} \left[\Delta Y_{it} - \tilde{Z}_{it}^{\top} \beta(z) - \left(X_{it}^{\top} \hat{m}_f(Z_{it}) - X_{i(t-1)}^{\top} \hat{m}_f(Z_{i(t-1)}) - \tilde{Z}_{it}^{\top} \hat{\beta}_f(z) \right) \right] K_H(Z_{it} - z) K_H(Z_{i(t-1)} - z).$$

Then, an adjusted empirical log-likelihood ratio function for $\beta(z)$ can be defined, for the within transformation, as

$$\hat{\mathscr{R}}_w(\beta(z)) = -2 \max \left\{ \prod_{i=1}^{N} p_i \,\middle|\, p_i \geq 0,\ \sum_{i=1}^{N} p_i = 1,\ \sum_{i=1}^{N} p_i \hat{T}_{wi}(\beta(z)) = 0 \right\},$$

and, for the first differences transformation, as

$$\hat{\mathscr{R}}_f(\beta(z)) = -2 \max \left\{ \prod_{i=1}^{N} p_i \,\middle|\, p_i \geq 0,\ \sum_{i=1}^{N} p_i = 1,\ \sum_{i=1}^{N} p_i \hat{T}_{fi}(\beta(z)) = 0 \right\}.$$

The asymptotic results for both, $\hat{\mathscr{R}}_w(\beta(z))$ and $\hat{\mathscr{R}}_f(\beta(z))$, are stated in the following theorem

Theorem 1.4.2. Assuming that conditions 1.2.1 - 1.2.13 hold, and $\beta(z)$ is the true parameter value, then $\hat{\mathscr{R}}_w(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ and $\hat{\mathscr{R}}_f(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ as $N \rightarrow \infty$ and T is fixed, where $\rightarrow_d$ means the convergence in distribution and $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.

Note that to state this result we do not impose any extra condition. Here, we need conditions 1.2.1 - 1.2.10 to ensure that $\hat{\mathscr{R}}_w(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ and conditions 1.2.1 - 1.2.7 and 1.2.11 - 1.2.13 to ensure that $\hat{\mathscr{R}}_f(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ as $N \rightarrow \infty$ as $N \rightarrow \infty$; however we do not need that $\sqrt{NT|H|^{T/2}}\mathrm{tr}(H) \rightarrow 0$ for the within transformation and $\sqrt{NT|H|}\mathrm{tr}(H) \rightarrow 0$ for the first differences transformation. Therefore, it is possible now to consider an optimal bandwidth matrix and hence the rate of convergence of the estimators will be also optimal.

As in other nonparametric estimation problems, bandwidth selection is important. Since the previous corrections enable us to use the optimal bandwidth then we can rely on standard data driven bandwidth selection techniques to select a bandwidth matrix. Among them, we propose to use a plug-in rule based on Sheather and Jones (1991). This proposal will be investigated in numerical studies in Section 1.6 and it will be also applied for illustrating the proposed Empirical Likelihood method with an empirical application in Section 1.7. Finally, there exists other data driven bandwidth selection criteria, such as cross-validation or empirical MSE criteria, that can be used alternatively to the plug-in method. They are detailed in the following section. Their main drawback is that their are computationally more demanding.

1.5 Bandwidth Selection

In this section we introduce two alternative procedures to estimate the bandwidths of the varying coefficients and its derivatives. If we are willing to assume that the bandwidth for the function of interest and its derivatives are the same, e.g. $H = hI_q$; we can define a cross validation (CV) criterion function as in Xue and Zhu (2007) and choose the smoothing parameter, h, that minimizes the following CV criterion function

$$CV_w(h) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it} - \tilde{Z}_{it}^{*\top}\hat{\beta}_w^{(-it)}(Z_{it})\right)^2, \tag{1.22}$$

where $\hat{\beta}_w^{(-it)}(Z_{it})$ is the MELE leave-one-out estimator of $\beta(Z_{it})$ as in (1.14). Note that we can also assume that there exist a different bandwidth for the function of interest h_1 and its derivatives h_2; in this case we choose the smoothing parameters h_1 and h_2 that minimizes CV criterion function

$$CV_w(h_1,h_2) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it} - \tilde{Z}_{it}^{*\top}\hat{\beta}_w^{(-it)}(Z_{it})\right)^2, \tag{1.23}$$

where we estimate the function of interest using h_1 and its derivatives using h_2. Note that, we can also replicate this result for the first differences estimator.

A second approach is to consider a modified version of the bandwidth selection criteria proposed in Fan and Gijbels (1995). It exhibits two main advantages against the previous procedure: First, it is originally designed for local polynomial regression and second, it enables us to compute by separate the bandwidths related to the levels and the derivatives of the unknown functions. We propose the following measure of discrepancy,

$$MSE\,(H) = E\left[\tilde{Z}^{*\top}\left(\hat{\beta}_w(Z) - \beta(Z)\right)\right]^2,$$

where $Z = (Z_{11}, \ldots, Z_{NT})^\top$,

$$W = blockdiag\left(K_H(Z_{i1}-z)\prod_{l=2}^{T} K_H(Z_{il}-z), \ldots, K_H(Z_{iT}-z)\prod_{l=1}^{T-1} K_H(Z_{il}-z)\right),$$

and

$$\tilde{Z}^{*\top} = \begin{pmatrix} \ddot{X}_{11}^\top & , X_{11}^\top \otimes (Z_{11}-z)^\top - \frac{1}{T}\sum_{s=1}^{T} X_{1s}^\top \otimes (Z_{1s}-z) \\ \vdots & \vdots \\ \ddot{X}_{NT}^\top & , X_{NT}^\top \otimes (Z_{NT}-z)^\top - \frac{1}{T}\sum_{s=1}^{T} X_{Ns}^\top \otimes (Z_{Ns}-z) \end{pmatrix}.$$

In this Mean Square Error (MSE), the expectation is taken over $Z_1, \cdots, Z_q$; $X_1, \cdots, X_d$. Therefore, for our problem, we can define the optimal bandwidth matrix H_{opt} as the solution to the following minimization problem,

$$H_{opt} = \arg\min_H MSE(H) = \arg\min_H E\left[\tilde{Z}^{*\top}\left(\hat{\beta}_w(Z) - \beta(Z)\right)\right]^2.$$

If $Z_1, \cdots, Z_q; X_1, \cdots, X_d$ are independent of the observed sample $\mathfrak{D} = (X_{11}, Z_{11}, \cdots, X_{NT}, Z_{NT})^T$, but they share the same distribution with (X_{11}, Z_{11}) it is straightforward to show that

$$MSE(H) = E\left[b^T(Z)\,\Omega(Z) b(Z) + \operatorname{tr}\{\Omega(Z) V(Z)\}\right], \tag{1.24}$$

where

$$\begin{aligned} b(Z) &= E\left\{\hat{\beta}_w(Z)|\mathfrak{D}, Z\right\} - \beta(Z), \\ V(Z) &= \operatorname{Var}\left\{\hat{\beta}_w(Z)|\mathfrak{D}, Z\right\}, \quad \text{and} \\ \Omega(Z) &= E\left(\tilde{Z}^*\tilde{Z}^{*\top}|Z\right). \end{aligned}$$

As it can be realized from the expression above, it has been now formalized the idea of choosing a bandwidth matrix H that minimizes the MSE, that is the sum of the squared bias and variance. Note that, the way we have defined the measure of discrepancy determines, in our case, the choice of a global bandwidth. That is, we will choose a bandwidth that remains constant with the location point. Unfortunately, the selection of H_{opt} does not solve all problems in bandwidth selection. In fact, as it can be realized, the MSE depends on some unknown quantities and therefore, our optimal bandwidth matrix can not be estimated from data. There are several alternative solutions to approximate the unknown quantities in the MSE. One alternative is to replace in (1.24) both bias and variance terms by their respective first order asymptotic expressions that were obtained in Theorem 1.3.1. This is the so called 'plug-in' method (see, for details Ruppert et al. (1995)). Another possibility is, as suggested in Fan and Gijbels (1995), to replace directly in (1.24) bias and variance by their exact expressions. That is

$$\begin{aligned} E\left\{\hat{\beta}(Z)|\mathfrak{D},Z\right\}-\beta(Z) &= \left\{E\left\{\hat{\beta}(z)|\mathfrak{D}\right\}-\beta(z)\right|_{z=Z} \\ \operatorname{Var}\left\{\hat{\beta}(z)|\mathfrak{D},Z\right\} &= \operatorname{Var}\left\{\hat{\beta}(z)|\mathfrak{D}\right\}\Big|_{z=Z}, \end{aligned} \tag{1.25}$$

where clearly, according to Theorem 1.3.1

$$\begin{aligned} E\left\{\hat{\beta}(Z)|\mathfrak{D},Z\right\}-\beta(Z) &= \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1}\widetilde{Z}^T W\tau \\ \operatorname{Var}\left\{\hat{\beta}(z)|\mathfrak{D},Z\right\} &= \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1}\widetilde{Z}^T W \mathscr{V} W \widetilde{Z}\left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1}, \end{aligned} \tag{1.26}$$

τ is a NT vector such that, for $i=1,\cdots,N$, $t=1,\cdots,T$,

$$\begin{aligned} \tau_{it} &= X_{it}^{\top} m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T} X_{is}^{\top} m(Z_{is}) \\ &- \left\{X_{it}^{\top} D_m(z)(Z_{it}-z) - \frac{1}{T}\sum_{s=1}^{T} X_{is}^{\top} D_m(z)(Z_{is}-z)\right\} \end{aligned}$$

and $\mathscr{V}$ is a $NT\times NT$ matrix that contains the V_{ij}'s matrices,

$$V_{ij} = E(v_i v_j^T | X_{i1},...,X_{iT},Z_{i1},...,Z_{iT}) = \sigma_v^2 I_T. \tag{1.27}$$

In order to estimate both bias and variance we need to calculate τ and $\mathscr{V}$. Note that for τ, developing a fifth order Taylor expansion of both $m(Z_{it})$ and $m(Z_{is})$ around z a local polynomial regression of order five would guarantee that the proposed bandwidth selection procedure will be $\sqrt{N}$-consistent for the local linear fit (see, Hall et al. (1991) for details). However, for the sake of simplicity a local cubic polynomial regression would be close to a $\sqrt{N}$-consistent selection rule and it will lead to a nice reduction in the computational effort. In this case (for $d=q=1$), the vector $\widehat{\tau}$ will contain the (estimated) expressions for the second and third order derivatives of the local cubic polynomial regression.

On the other side, in order to estimate $\mathscr{V}$, note that, under assumption 1.2.2 we can consistently estimate this quantity by

$$\widehat{\sigma}_{\upsilon}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\ddot{Y}_{it} - X_{it}^{\top}\widehat{m}^{-i}(Z_{it}) + \frac{1}{T}\sum_{s=1}^{T} X_{is}^{\top}\widehat{m}^{-i}(Z_{is})\right\}^2. \tag{1.28}$$

Note that both $\widehat{\tau}$ and $\widehat{\sigma}_{\upsilon}^2$ depend on a bandwidth matrix H that needs to be determined from data. A suitable pilot bandwidth matrix H^* that can be used for these computations can be obtained using the global RSC procedure proposed in Fan and Gijbels (1995). Note that once we have estimated τ and σ_{υ}^2 we can now provide an estimator for $b(H)$, $V(H)$ and $\Omega(H)$. Mainly,

$$\begin{aligned}
\widehat{b}(Z_{it}) &= E\left\{\hat{\beta}_w(Z)|\mathfrak{D},Z\right\} - \beta(Z) = \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W \hat{\tau}, \\
\widehat{V}(Z_{it}) &= \mathrm{Var}\left\{\hat{\beta}_w(Z)|\mathfrak{D},Z\right\} = \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W \widehat{\mathscr{V}} W \widetilde{Z} \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1}, \\
\hat{\Omega}(Z_{it}) &= \frac{\sum_{j\neq i,t} \tilde{Z}_{it}^{*} \tilde{Z}_{it}^{*\top} \prod_{l=1}^{T} K_H\left(Z_{jl} - Z_{it}\right)}{\sum_{j\neq i,t} \prod_{l=1}^{T} K_H\left(Z_{jl} - Z_{it}\right)}.
\end{aligned}$$

The corresponding estimator of the $MSE(H)$, according with (1.24) will be

$$\widehat{MSE}(H) = \frac{1}{NT} \sum_{it} \left[\widehat{b}^T(Z_{it}) \widehat{\Omega}(Z_{it}) \widehat{b}(Z_{it}) + \mathrm{tr}\left\{\widehat{\Omega}(Z_{it}) \widehat{V}(Z_{it})\right\}\right]. \tag{1.29}$$

Then, we define the estimator of H_{opt}, $\widehat{H}_{opt}$ as the solution to the following problem,

$$\widehat{H}_{opt} = \arg\min_{H} \widehat{MSE}(H).$$

Although we do not provide theoretical properties of this bandwidth, in a much simpler context of a varying coefficient model with no heterogenity effect, in Zhang and Lee (2000) it has been studied the theoretical properties of this bandwidth selection criteashe, and we believe it could be extended to our case. The same expressions can be obtained for the Empirical Likelihood Estimator based in first differences. In fact, in a different context, these expressions can be found in Rodriguez-Poo and Soberón (2014).

1.6 Monte Carlo results

In this section we propose a simulation exercise to analyse the small sample behaviour of the empirical likelihood techniques that we have proposed in the previous sections when constructing confidence bands. In order to do so, we consider the following data generating process,

$$Y_{it} = \mu_{qi} + X_{dit}^{\top} m(Z_{qit}) + v_{it},\ i = 1,\ldots,N; t = 1,\ldots,T; d,q = 1,2,$$

where X_{dit} and Z_{qit} are random variables, where $Z_{qit} = w_{qit} + w_{qi(t-1)}$, ($w_{qit}$ are *i.i.d.* $\mathscr{N}(0,\ 1)$) and $X_{dit} = 0.5\zeta_{dit} + 0.5\xi_{dit}$ (ζ_{qit} and ξ_{dit} are *i.i.d.* $\mathscr{N}(0,\ 1)$) and we consider three cases of study

$$\begin{aligned}
a._\ (d=1,q=1) \quad &: \quad Y_{it} = \mu_{1i} + X_{1it}^{\top} m_1(Z_{1it}) + v_{it}, \\
b._\ (d=1,q=2) \quad &: \quad Y_{it} = \mu_{2i} + X_{1it}^{\top} m_1(Z_{1it}, Z_{2it}) + v_{it}, \\
c._\ (d=2,q=1) \quad &: \quad Y_{it} = \mu_{1i} + X_{1it}^{\top} m_1(Z_{1it}) + X_{2it}^{\top} m_2(Z_{1it}) + v_{it}.
\end{aligned}$$

The chosen functional form for $m(.)$ are $m_1(Z_{1it}) = \sin(Z_{1it}\pi)$, $m_1(Z_{1it}, Z_{2it}) = \sin((Z_{1it} + Z_{2it})\pi/2)$, and $m_2(Z_{1it}) = \exp(-Z_{1it}^2)$. We also experiment with tow specifications for the fixed effects

1._ μ_{1i} depends on Z_{1it}, where the dependence is imposed by $\mu_{1i} = c_0\bar{Z}_{1i.} + u_i$ for $i = 1,...,N$ and $\bar{Z}_{1i.} = T^{-1}\sum_t Z_{1it}$,

2._ μ_{2i} depends on Z_{1it} and Z_{2it} by $\mu_{2i} = c_0\bar{Z}_{i.} + u_i$ for $i = 1,...,N$ and $\bar{Z}_{i.} = \frac{1}{2}(\bar{Z}_{1i.} + \bar{Z}_{2i.})$,

where u_i is an i.i.d. $\mathcal{N}(0,\ 1)$ and $c_0 = 0.5$ controls the correlation between the unobservable individual heterogeneity and some of the regressors of the model. Also, let ε_{it} be and $i.i.d.\ \mathcal{N}(0,1)$ and v_{it} a scalar random variable, for each model we work with the following specification of the error term: $v_{it} = \varepsilon_{it}$

In this experiment we use 1000 Monte Carlo replications (M). The number of period (T) is fixed to be 3 and the number of cross-sections (N) take the values 50, 100 and 150. For the calculations we use a Gaussian Kernel and for the bandwidth matrix H we use the standard choice $\hat{H} = \hat{h}I$, where I is the $q \times q$ identity matrix, and $\hat{h} = \hat{\sigma}_z(NT)^{-1/5}$, where $\hat{\sigma}_z$ is the simple standard deviation of $\{Z_{it}\}_{i=1,t=1}^{N,T}$. For any replication we have built up the confidence bands using the empirical likelihood confidence bands and the normal approximation confidence bands introduced before. In table 1.1 we present the point-wise confidence intervals, where NLB = Normal Approximation (NA) Lower Bound, NUB = NA Upper Bound, MELLB = Mean Corrected Empirical Likelihood (MCEL) Lower Bound, MELUB= MCEL Upper Bound, RELLB = Residual Adjusted Empirical Likelihood (RAEL) Lower Bound and RELUB = RAEL Upper Bound.

As the reader may notice, from table 1.1, between the MCEL, the RAEL and the NA, the length of the confidence interval is smaller in the RAEL; also note that, the confidence interval length of the MCEL is smaller than the NA. Also, it is interesting that, as table 1.1 shows, the confidence intervals using NA are wider than ones using empirical likelihood. Therefore we can say that when N goes to infinity the length the confidence bands of the NA are wider that the confidence bands of the MCEL and the RAEL. Thus, we can conclude by saying that the RAEL and MCEL confidence bands behave better than the NA confidence bands. Between RAEL and MCEL confidence bands, simulations results show that the RAEL

Size	Model	NLB	MELLB	RELLB	$\hat{\beta}_1(z)$	RELUB	MELUB	NUB
				Within				
$N = 50$								
	a	-0.99	-0.60	-0.42	0.04	0.48	0.68	1.13
	b	-0.94	-0.62	-0.46	0.00	0.49	0.63	0.97
	c	-0.78	-0.67	-0.51	0.02	0.54	0.70	0.83
$N = 100$								
	a	-0.95	-0.52	-0.28	0.01	0.30	0.54	0.98
	b	-0.98	-0.58	-0.42	-0.03	0.36	0.52	0.90
	c	-0.74	-0.53	-0.28	0.07	0.38	0.65	0.87
$N = 150$								
	a	-0.91	-0.46	-0.21	0.00	0.21	0.47	0.93
	b	-1.02	-0.50	-0.34	0.03	0.39	0.57	1.10
	c	-0.83	-0.52	-0.23	0.00	0.23	0.52	0.78
				First Difference				
$N = 50$								
	a	-0.96	-0.75	-0.41	0.00	0.38	0.74	0.94
	b	-0.85	-0.60	-0.38	0.03	0.43	0.66	0.94
	c	-0.92	-0.78	-0.43	0.00	0.43	0.78	0.92
$N = 100$								
	a	-0.79	-0.64	-0.23	0.01	0.26	0.66	0.81
	b	-0.88	-0.52	-0.31	-0.01	0.29	0.49	0.87
	c	-0.77	-0.70	-0.26	0.00	0.28	0.71	0.80
$N = 150$								
	a	-0.72	-0.58	-0.20	-0.00	0.19	0.56	0.71
	b	-0.93	-0.46	-0.27	0.00	0.26	0.45	0.92
	c	-0.69	-0.61	-0.20	-0.00	0.21	0.62	0.68

Table 1.1 Pointwise Confidence interval for $\beta(z)$ at $z = 0$ based on the MCEL, RAEL and NA, when the nominal level is 95%

confidence bands behave better than the MCEL. Also, by comparing the within method with the first difference method we can conclude that for the NA and the RAEL confidence bands the First Difference method reduces the length of the confidence interval; however the MCEL confidence interval increases it length in comparison to the Within method (table 1.1).

1.7 An empirical aplication

In this section we offer a very simple application where our empirical likelihood based confidence intervals can be of great interest; we consider the estimation of the production efficiency of the EU firms. Conventionally, these type of studies are based on a Cobb-

Douglas stochastic production function. A standard assumption in the literature is that capital and labour elasticities are constant over time; studies conducted under such a restrictive framework present some weaknesses. On the one hand, the estimation procedure can be complicated by the presence of individual heterogeneity together with the inefficiency term; especially, when there exist a correlation between the individual heterogeneity and the covariates of the model. See Greene (2005) or Wang and Ho (2010) among others. On the other hand, there are empirical studies that suggest that capital and labour elasticities vary according to other features of the companies such as the research and development, *R&D*, expenses. See Ahmad et al. (2005) among others, where they prove that varying coefficient models are a natural way to extend these constant elasticities to the functional form. Also, there exist a standard belief that the liquid capital marginal productivity is not affected by the *R&D* expenses. In order to test this fact, we propose the following varying coefficient panel data model

$$y_{it} = w_{it}\beta_1(z_{it}) + l_{it}\beta_2(z_{it}) + k_{it}\beta_3(z_{it}) + \mu_i + v_{it}, \quad i = 1,\ldots,N; \; t = 1,\ldots,T \tag{1.30}$$

where $y_{it} = ln(Y_{it})$, $w_{it} = ln(W_{it})$, $l_{it} = ln(L_{it})$ and $k_{it} = ln(K_{it})$. Also, Y represents the sales of the company, W the liquid capital, L the labour input, K the fixed capital and Z the firms *R&D* expenses. In addition μ_i stands for the individual heterogeneity and $v_{it} = \nu_{it} - u_{it}$ is a composed error term, where ν_{it} is the idiosyncratic error and u_{it} represents the inefficiency that has expected value equal to $E[v_{it}] = -E[u_{it}]$. Note that in the specification (1.30) the *R&D* variable has a neutral effect on the production function by shifting the level of the production frontier but also affects the labour and capital marginal productivities.

Variable	Average	Standard Deviation	Correlations			
Y	6705377.60	25791397.23				
W	1379564.74	5073321.65	0.66			
K	1161082.07	3443125.02	0.79	0.83		
L	17976.52	48686.98	0.59	0.83	0.86	
Z	224303.78	937324.42	0.40	0.60	0.63	0.62

Table 1.2 Statistics of inputs and outputs.

The sample used in this empirical analysis includes 1220 observations divided in 160 companies and 7 time periods, form 2008 to 2014, from the Analyse Major Database form European Sources (AMADEUS). The data contains information about the accounting and financial statements of European firms. Note that we are working with expenses, thus all the variables have been deflected using the implicit index of the GDP. The information related to prices used to deflate the variables was obtained form the Spanish Statistical Office (various years). In Table 1.2 we present summary statistics of the observations, as it can be seen, the standard deviations show that there exist a high degree of heterogeneity.

In figure 1.1 we present our results by plotting the estimated curves against the *R&D* expenses; here the continuous lines denote the non-parametric estimated curve and the doted lines represent the 95% pointwise confidence interval obtained using the MCEL (long-dashed curve) and RAEL (short-dashed curve). The bandwidths, as in Section 5, have been computed by a plug-in technique proposed in Sheather and Jones (1991) and already explained in Section 1.4. Also, note that figure 1.1 shows the results for the marginal productivity of liquid capital (W), fixed capital (K) and labour (L), and the returns to scale defined as $\beta_1(z)+\beta_2(z)+\beta_3(z)$.

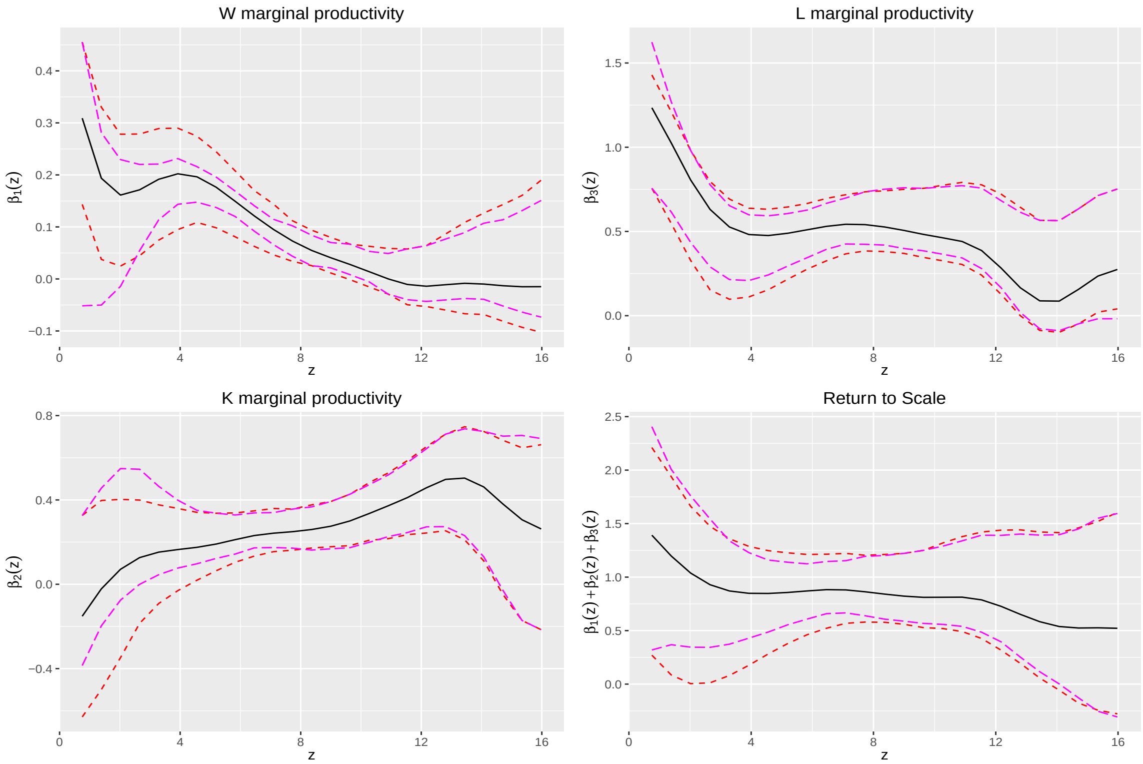

Figure 1.1 Averages of 95 % Confidence Intervals for $\hat{\beta}_j(z)$ for $j=1,\ldots,3$ (Within method), based on MCEL (long-dashed curve) and RAEL (short-dashed curve).

Focusing in the marginal productivity of liquid capital (W marginal productivity) we have realized that it tends to be decreasing; however when it reaches a certain level of *R&D* expenses it tends to be steady and close to zero. Basically, this means that companies with small *R&D* expenses have a decreasing marginal productivity of liquid capital. Analysing the graph, we can see that as the level of *R&D* expenses increases, first the companies see an increase in the marginal productivity of liquid capital; then they experiment a drop of the marginal productivity of liquid capital until it become stable near to zero. On its part, the marginal productivity of fixed capital (*K* marginal productivity) is not a linear function with the level of *R&D* expenses. Clearly, there exist an upward general trend, with a bell shape form for companies with large *R&D* expenses. This bell shape of the marginal productivity of fixed capital curve suggests that, while modest *R&D* expenses can improve the fixed capital productivity, higher *R&D* expenses leads to lower fixed capital productivity.

L marginal productivity shows the results of the labour marginal productivity. Here we observe that the labour marginal productivity is not a linear function of *R&D*; broadly, it decreases with *R&D*, however, with higher levels of *R&D* the marginal productivity of labour becomes to increase. This inverted bell shape suggest that companies with reduced *R&D* tend to have lower labour marginal productivity at the beginning while companies with higher *R&D* are more likely to have an increase in labour marginal productivity. Note that this behaviour is characteristic in companies that use *R&D* to improve the performance of their machines rather than focusing in training their workers. Finally, using these results we can not conclude that the returns to scale are not equal to one because one is within the confidence interval. However, we can conclude that the returns to scale are not linear with *R&D* and they seem to have a negative effect in the behaviour of the returns to scale.

1.8 Conclusions

In this chapter we adapt empirical likelihood techniques to construct confidence bands in a fixed effects varying coefficient panel data model. First we consider a so called naive empirical likelihood technique. As a byproduct we provide two alternative empirical maximum likelihood estimators of the varying coefficients and their derivatives. Since the use of naive empirical likelihood techniques provides sub-optimal rates of convergence we slightly modify the original techniques that enables us to obtain optimal nonparametric rates: Mean corrected and residual adjusted empirical likelihood ratios. Finally we undertake a simulation study and we apply successfully our techniques in a empirical study of of production efficiency of the European Union's companies.

Chapter 2

Empirical likelihood based inference for categorical varying coefficient panel data model with fixed effects

This chapter also appeared as Arteaga-Molina and Rodríguez-Poo (2019).

2.1 Introduction

In the last years there has been an increasing interest in the study of panel data models combined with nonparametric techniques. On the one hand, the results are promising, however it is true that the main disadvantages related to nonparametric techniques (e.g., the curse of dimensionality, see Härdle (1990)) continue to appear when we apply them to panel data models. In order to overcome this drawback varying coefficient models appear as a reasonable specification. Varying coefficient models encompass a great variety of other simple models applied by econometricians as partially linear models or the fully nonparametric models. On the other side, in many applied microeconomic problems, the difficulty to have available all explanatory variables of interest has attracted the attention of many applied economists towards panel data models. As it is well known, in a regression model, these techniques enable us to estimate the objects of interest consistently by allowing for individual heterogeneity of unknown form. Nowadays, we have available a pleiad of varying coefficient estimators that exhibit good asymptotic properties under rather different sets of assumptions such as random effects, fixed effects or cross sectional dependence (see, Su and Ullah (2011), Rodriguez-Poo and Soberon (2017) and Parmeter and Racine (2018) among others for comprehensive surveys of the literature). More precisely, the problem of considering varying coefficients that depend on discrete data has attracted some interest because the availability of discrete variables is rather common in economic analysis. In Li et al. (2013b) it is proposed a semiparametric varying-coefficient with purely categorical covariates; furthermore, in Feng et al. (2017) the previous setting is extended to fixed effects and cross-sectional dependence.

Although in the previous papers the authors provide extensive results about the asymptotic behavior of the estimators, inference is not always an easy problem to undertake. In fact, in all above mentioned papers, asymptotic normal approximations are obtained. In the discrete covariates case, under fairly general conditions, if the bandwidth is selected using the cross-validation criteria, the asymptotic bias of the estimator is negligible and therefore inference based on the asymptotic distribution is more feasible than in the continuous covariate case where some undersmoothing is needed (Li and Racine (2007)). Unfortunately, if additionally, we are willing to assume cross-sectional dependence inference becomes much cumbersome. Besides, using confidence bands as a testing device is not straightforward as uniform confidence bands are necessary to do so (see, Li et al. (2013a)).

Aside from the usual tools to make inference (e.g., asymptotic normality), Owen (1988) introduced the empirical likelihood technique; there exist several advantages of this method over the usual ones (e.g., no limiting variance estimation are necessary, it combines the reliability of nonparametric methods with the effectiveness of the likelihood approach, among others). For further discussion on the advantages of the empirical likelihood technique the reader should refer to Owen (1990, 1988), Hall and La Scala (1990), DiCiccio et al. (1991), Owen (1991), Hall and Owen (1993), Kolaczyk (1994), Qin and Lawless (1994), Owen (2001), Li and Van Keilegom (2002), among others. In fact, due to its properties, empirical likelihood have been already applied in longitudinal data varying coefficient models with random effects (e.g., Xue and Zhu (2007)); as for the fixed effects case, see, Zhang et al. (2011) and chapter 1. Unfortunately these type of results are not available for the panel data discrete/categorical varying coefficient setting. In fact, in chapter 1, the varying coefficient, $m(Z)$, varies according to a continuous variable, Z; therefore, the authors use a continuous kernel functions. Also the authors derive the asymptotic theory for T fixed and $N \to \infty$.

In this chapter, and starting from a panel data discrete/categorical varying coefficient model with both fixed effects and cross sectional dependence, we develop empirical likelihood ratios and we derive a nonparametric version of the Wilks' theorem. Besides, we obtain maximum empirical likelihood estimator of the varying parameters and its asymptotic theory. Based on these results, we can build up confidence regions for the parameter of interest through a standard chi square approximation. The rest of this chapter is organized as follows. In Section 2.2 we propose to construct confidence bands for the unknown functions by using what we call a naive empirical likelihood technique. In Section 2.3, as a by product, we provide an alternative maximum empirical likelihood estimator of the fixed effect categorical varying parameters. In Section 2.4 we present the proposed technique in an application that reports estimates of strike activities from 17 OECD countries for the period 1951 – 1985. Finally Section 2.5 concludes. The proofs of the main results are collected in the Appendix.

2.2 Naive Empirical Likelihood

We consider the following categorical varying coefficient panel data regression model

$$Y_{it} = X_{it}^{\top} \beta(Z_{it}) + \omega_i + v_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T, \tag{2.1}$$

where Y_{it} is the response, $X_{it} = \left(X_{it,1}, \ldots, X_{it,d}\right)^{\top}$ and $Z_{it} = \left(Z_{it,1}, \ldots, Z_{it,q}\right)^{\top}$ are vectors of dimension d and q respectively, and $\beta(\cdot) = \{\beta_1(\cdot), \ldots, \beta_d(\cdot)\}^{\top}$ is a $d \times 1$ vector of unknown functions; here, ω_i stands for so called fixed effects and v_{it} are the random errors. Note that when Z_{it} is a vector of continuous random variables, model (2.1) stands for the so called varying coefficient panel data model with fixed effects studied by authors such as Rodriguez-Poo and Soberón (2014, 2015), Cai and Li (2008), Sun et al. (2009), Su and Ullah (2011) and Chen et al. (2013) among others. In this paper we consider the case where Z is purely categorical and in order to distinguish between X and Z we will refer as regressor and covariate respectively. Note that we are not willing to impose any restriction between ω_i an the pair (X_{it}, Z_{it}).

The model (2.1) is an extension of the cross-sectional varying coefficient model of Li et al. (2013b) to the panel data framework as it appears in Feng et al. (2017). First, we will obtain confidence bands for $\beta(\cdot)$ based on the empirical likelihood approach; to do so, we need the first order condition of the minimization problem for obtaining $\beta(\cdot)$. Note that, this condition, for given z, from (2.1) is

$$\mathrm{E}\left[X_{it}\left\{Y_{it} - X_{it}^{\top}\beta(Z_{it})\right\}\middle| Z_{it} = z\right] \neq 0,$$

due to the fixed effects. To deal with this problem, several transformations have been proposed in the standard literature of panel data models. For example, when Z is continuous, some differencing transformations combined with a Taylor series approximation could be done (see, chapter 1). Unfortunately, if the elements of Z are of a discrete nature a Taylor approximation is not feasible.

Here we propose to keep the same idea of using the within transformation but instead of using a continuous kernel we aim to use a kernel function designed for discrete random variables (see, Aitchison and Aitken (1976)). Thus, let $1_{js,it} = 1(Z_{it} = Z_{js})$ and $L_{js,it,\gamma} = L(Z_{it}, Z_{js}, \gamma)$ for $1 \le i, j \le N$ and $1 \le t, s \le T$. Note that $L(Z_{it}, Z_{js}, \gamma)$ represents a kernel function for multivariate discrete spaces

$$L(Z_{it}, z, \gamma) = \prod_{s=1}^{q} \ell(Z_{it,s}, z_s, \gamma_s) = \prod_{s=1}^{q} \gamma_s^{1(Z_{it,s} \neq z_s)}, \tag{2.2}$$

where $\gamma = (\gamma_1, \ldots, \gamma_q)^{\top}$, $1(Z_{it,s} \neq z_s)$ denotes the usual indicator function, which takes the value 1 when $Z_{it,s} \neq z_s$, and 0 otherwise and

$$\ell(Z_{it,s}, z_s, \gamma_s) = \begin{cases} 1 & \text{if } Z_{it,s} = z_s \\ \gamma_s & \text{if } Z_{it,s} \neq z_s \end{cases},$$

is the kernel function of Aitchison and Aitken (1976) for unordered covariates, where $\gamma_s = 0$ leads to an indicator function and $\gamma_s = 1$ gives a uniform weighted function. Thus, we can conclude that $\gamma_s \in [0, 1]$ for $s = 1, \ldots, q$. Also, note that the kernel function (2.2) can also be expressed as

$$\begin{aligned} L(Z_{it}, z, \gamma) &= \prod_{m=1}^{q} \ell(Z_{it,m}, z_m, \gamma_m) \\ &= \prod_{m=1}^{q} \{1(Z_{it,m} = z_m) + \gamma_m 1(Z_{it,m} \neq z_m)\} \\ &= \prod_{m=1}^{q} 1(Z_{it,m} = z_m) + \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} + \ldots + \prod_{m=1}^{q} \gamma_m 1(Z_{it,m} \neq z_m) \\ &= 1(Z_{it} = z) + \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} + \ldots + \prod_{m=1}^{q} \gamma_m 1(Z_{it,m} \neq z_m) \end{aligned}$$

where $1_{m,itz^*} = 1(Z_{it,m} \neq z_m) \prod_{n=1,n\neq m}^{q} 1(Z_{it,n} = z_n)$ is an indicator function which takes value 1 if Z_{it} and z differs only in their m^{th} component and 0 otherwise. Note that if we assume that $\gamma \to 0$ as $(N,T) \to (\infty,\infty)$ it is reasonable to simplify the kernel product function (2.2) as follows

$$L(Z_{js}, Z_{it}, \gamma) = 1_{js,it} + \sum_{m=1}^{q} \gamma_m 1_{m,jsit} + O(||\gamma||^2), \tag{2.3}$$

where $1_{m,jsit} = 1(Z_{js,m} \neq Z_{it,m}) \prod_{n=1,n\neq m}^{q} 1(Z_{js,n} = Z_{it,n})$ and $||.||$ stands for the Frobenius norm.

Expression (2.3) is of great interest because it enables us to apply a modified version of a within transformation in (2.1) and then remove the fixed effects. Thus, let $T_{it} = \sum_{s=1}^{T} L_{it,is,\gamma}^{p}$, where $p \geq 2$ is a finite positive integer and chosen arbitrarily. In practice, the choice of $p = 2$ is enough. Let $\tilde{X}_{it} = X_{it} - T_{it}^{-1} \sum_{s=1}^{T} X_{is} 1_{is,it}$, $\tilde{Y}_{it} = Y_{it} - T_{it}^{-1} \sum_{s=1}^{T} Y_{is} 1_{is,it}$ and $\tilde{v}_{it} = v_{it} - T_{it}^{-1} \sum_{s=1}^{T} v_{is} 1_{is,it}$. Applying this transformation in (2.1) we obtain

$$\begin{aligned} \tilde{Y}_{it} &= X_{it}^{\top} \beta(Z_{it}) + \omega_i + v_{it} - \frac{1}{T_{it}} \sum_{s=1}^{T} \left\{ X_{is}^{\top} \beta(Z_{is}) + \omega_i + v_{is} \right\} L_{is,it,\gamma}^{p} \\ &= X_{it}^{\top} \beta(Z_{it}) - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^{\top} L_{is,it,\gamma}^{p} \beta(Z_{it}) + \frac{1}{T_{it}} \sum_{s=1} X_{is}^{\top} L_{is,it,\gamma}^{p} \beta(Z_{it}) \\ &\quad - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^{\top} \beta(Z_{is}) L_{is,it,\gamma}^{p} + \tilde{v}_{it} \\ &= \tilde{X}_{it}^{\top} \beta(Z_{it}) + \rho_{it} + \tilde{v}_{it}, \end{aligned} \tag{2.4}$$

where $\rho_{it} = T_{it}^{-1} \sum_{s=1}^{T} X_{is}^{\top} \{\beta(Z_{it}) - \beta(Z_{is})\} L_{is,it,\gamma}^{p}$ stands for the truncation residual. Due to the fact that $1^{p}(\cdot) = 1(\cdot)$ and $\{\beta(Z_{it}) - \beta(Z_{is})\} 1(Z_{is} = Z_{it}) = 0$, if $\gamma \to 0$ as $(N,T) \to (\infty,\infty)$

we obtain

$$\{\beta(Z_{it})-\beta(Z_{is})\}\,L^{p}_{is,it,\gamma}=O\left(||\gamma||^{p}\right) \tag{2.5}$$

uniformly. Therefore, due to (2.5), the truncation residual ρ_{it} is controlled by the bandwidth γ only. Given this result we obtain that the first order condition, for given z, from (2.4) is

$$\mathrm{E}\left[\tilde{X}_{it}\left\{\tilde{Y}_{it}-\tilde{X}_{it}^{\top}\beta(Z_{it})\right\}\Big|Z_{it}=z\right]=0. \tag{2.6}$$

In this case, the least squares estimator of $\beta(z)$ is the solution to (2.6) when $Z_{it}=z$; therefore, the orthogonality condition (2.6) for $\beta(z)$ has the following form

$$\mathrm{E}\left[\tilde{X}_{it}\left\{\tilde{Y}_{it}-\tilde{X}_{it}^{\top}\beta(z)\right\}\Big|Z_{it}=z\right]=0. \tag{2.7}$$

Then, employing the constraint (2.7), the auxiliary random vector for the modified within transformation is

$$T_i\{\beta(z)\}=\sum_{t=1}^{T}\tilde{X}_{it}\left\{\tilde{Y}_{it}-\tilde{X}_{it}^{\top}\beta(z)\right\}L(Z_{it},z,\gamma); \tag{2.8}$$

note that, (2.8) is the sample version of (2.7) using a local smoothing method with a discrete kernel function. Also, if $\beta(z)$ is the true parameter, it is easy to show, due to (2.7), that $\mathrm{E}[T_i\{\beta(z)\}]=0$. Therefore, using the information $\mathrm{E}[T_i\{\beta(z)\}]=0$ the naive empirical log-likelihood ratio for $\beta(z)$ is defined as

$$\mathscr{R}\{\beta(z)\}=-2\max\left[\sum_{i=1}^{N}\log(p_i)\middle|p_i\geq 0,\ \sum_{i=1}^{N}p_i=1,\ \sum_{i=1}^{N}p_iT_i\{\beta(z)\}=0\right], \tag{2.9}$$

where $p_i=p_i(z),\ i=1,\ldots,N$. Using the Lagrange multiplier method the probabilities p_i are

$$p_i=\frac{1}{N}\frac{1}{1+\lambda^{\top}T_i\{\beta(z)\}}. \tag{2.10}$$

By (2.9) and (2.10), $\mathscr{R}\{\beta(z)\}$ leads to

$$\mathscr{R}\{\beta(z)\}=2\sum_{i=1}^{N}\log\left[1+\lambda^{\top}T_i\{\beta(z)\}\right]; \tag{2.11}$$

where λ is a $(d\times 1)$ vector of Lagrange multipliers associated to the constraint

$$\sum_{i=1}^{N}p_iT_i\{\beta(z)\}=0$$

and it is given by

$$\sum_{i=1}^{N} \frac{T_i\{\beta(z)\}}{1+\lambda^\top T_i\{\beta(z)\}} = 0, \tag{2.12}$$

subject to the constraint that satisfies the non-negativity condition and avoids a convex dual problem (see, Owen (2001), Chapter 3). Using equation (2.11), (2.12), a Taylor expansion and denoting $\tilde{D}\{\beta(z)\} = (NT)^{-1}\sum_{i=1}^{N} T_i\{\beta(z)\}T_i^\top\{\beta(z)\}$ it can be shown that

$$\mathscr{R}\{\beta(z)\} = \left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} T_i\{\beta(z)\}\right]^\top \left[\tilde{D}\{\beta(z)\}\right]^{-1} \left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} T_i\{\beta(z)\}\right] + o_p(1). \tag{2.13}$$

Hence, it is easy to show using (2.13) that $\mathscr{R}\{\beta(z)\}$ is asymptotically a standard Chi square distribution. In order to formally introduce this result, we need the following assumptions.

Assumption 2.2.1. :

(i) Let $\mathscr{D}$ be the range of values assumed by Z_{it}, then $p(z) = \Pr(Z_{it} = z) > 0\ \forall z \in \mathscr{D}$. The function $\beta(z)$ is bounded on the support $\mathscr{D}$ of z, i.e., $\max_{z\in\mathscr{D}}||\beta(z)|| < \infty$ and it is not a constant function with respect to z. Denote z_m as the m^{th} component of the q-dimensional vector $z = \left(z_1, ..., z_q\right)^\top$, where z_m is assume to take c_m different integer values in $\{0,1,\ldots,c_m-1\}$ for $c_m \geq 2$ and $m = 1,\ldots,q$. Moreover, q is finite and $\max_{1\leq m\leq q} c_m < \infty$.

(ii) Let (X_{it}, Z_{it}, v_{it}) be independent across i for each fixed t. Besides, for each fixed i, the process (X_{it}, Z_{it}, v_{it}) is strictly stationary and α-mixing. The α-mixing coefficient between (X_{it}, Z_{it}, v_{it}) and $\left(X_{js}, Z_{js}, v_{js}\right)$ is determined by $\alpha_{ij}(|t-s|)$, where

$$\alpha(k) = \sup_{\substack{A\in\sigma((X_{is},Z_{is},v_{is}),s\leq t)\\ B\in\sigma((X_{is},Z_{is},v_{is}),s\geq t+k)}} \left|P(A\textstyle\bigcap B) - P(A)P(B)\right|, \quad k \geq 1$$

besides, for a $\delta > 0$, $\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\{\alpha_{ij}(|t-s|)\right\}^{\frac{\delta}{4+\delta}} = O(NT)$

(iii) $\forall z \in \mathscr{D}$, $i = 1,\ldots,N$ and $t = 1,\ldots,T$, let $||\mu_X(z)||$ and $||\Sigma_X(z)||$ be uniformly bounded in z, where $\mu_X(z) = \mathrm{E}\left(X_{it}|Z_{it} = z\right)$ and $\Sigma_X(z) = \mathrm{E}\left(X_{it}X_{it}^\top \middle| Z_{it} = z\right)$.

(iv) Denote $\mathscr{X} = \left\{(X_{js}, Z_{js})\right\}_{j=1,s=1}^{N,T}$, then $\mathrm{E}\left(v_{it}|\mathscr{X}\right) = 0$ and $0 < \mathrm{E}\left(v_{it}^2|\mathscr{X}\right) = \sigma_v^2 < \infty$ almost surely (a.s.) for all $1 \leq i \leq N$ and $1 \leq t \leq T$. For some constants $\delta > 0$ and

$0 < a_1 < \infty$, $E\left(|v_{it}|^{4+\delta} + ||X_{it}||^{4+\delta}\right) \leq a_1$ uniformly. Also, over the time dimension, $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E(v_{it}v_{is}|\mathscr{X})| = O(1)$.

(v) Let ω_i be arbitrarily correlated with both X_{it} and Z_{it} with unknown correlation structure.

Assumption 2.2.1.(i) is quite standard and similar to Assumption 1.(i) in Li et al. (2013b). Note that, in order to deal with the case where the cardinality of $\mathscr{D}$ is infinity, one can work with the normalization used to deal with time varying coefficient model. That is, as in Feng et al. (2017), suppose $q = 1$, $Z_{it} \in \{0,1,2,\ldots,u(N,T)\}$, where $u(N,T) \to \infty$ and $u(N,T)/(NT) \to c$ for $0 \leq c < \infty$ as $(N,T) \to (\infty,\infty)$; then a variant of model (4.1) is obtained by normalizing Z_{it} by $u(N,T)$ as follows

$$Y_{it} = X_{it}^{\top}\beta\left\{\frac{Z_{it}}{u(N,T)}\right\} + \omega_i + v_{it}, \tag{2.14}$$

were $\beta(\cdot)$ can be treated as a continuous function of covariates; therefore, (2.14) is just the model proposed by Sun et al. (2009) with $\beta(\cdot)$ being a continuous function. This normalization is similar to the one employed in Chen et al. (2012b) and Cai (2007) when dealing with time varying coefficients.

Assumptions 2.2.1.(ii) is similar to Assumptions B and C of Bai (2009). The strict stationary assumption goes in the same line as Assumption A4 in Chen et al. (2012a) and Assumption A2 in Chen et al. (2012b). More details and relevant discussion can be found in Feng et al. (2017).

Assumption 2.2.1.(iii) sets restrictions on the unconditional moments as in Assumption $3.3-3.6$ in Rodriguez-Poo and Soberón (2014). Due to the within transformation, we have to assume it holds uniformly across i, which is in the same direction of Assumption A1 in Chen et al. (2013) and Assumption C in Bai (2009)

Assumption 2.2.1.(iv) is the same as that in Arellano (1987) and goes in the same direction as Assumption A2 and A4 Chen et al. (2012b). This assumption sets up the cross-sectional dependence as a weak correlation between individuals by using a spatial error structure, where a general spatial correlation structure has been imposed to link together the cross sectional dependence and the stationary mixing condition. (e.g., Pesaran and Tosetti (2011), Chen et al. (2012a) and Chen et al. (2012b) among others). Here, the last equation in Assumption 2.2.1.(iv) it is a simplified version of the one in Chen et al. (2012a) (A.18); this last equation is needed due to the within transformation.

Finally, Assumption 2.2.1.(v) imposes the so called fixed effects. Note that we are not willing to assume any constraint in the relationship between the random heterogeneity ω and the vector of regressors and covariates, (X,Z).

Having all these assumptions into consideration we can state formally the following theorem.

Theorem 2.2.1. Assuming that condition 2.2.1 hold and if $\gamma_m \to 0$ in such a way that $\sqrt{NT}\gamma_m \to 0$ for $m = 1, \ldots, q$ as $(N,T) \to (\infty, \infty)$ jointly, then $\mathscr{R}\{\beta(z)\} \to^d \chi^2_d$. Here $\to^d$ means the convergence in distribution and χ^2_d stands for the standard chi-square distribution with d degrees of freedom.

Therefore, we can build up the confidence bands using theorem 2.2.1 as follows,

$$R_\alpha = \{\beta(z) : \mathscr{R}\{\beta(z)\} \leq c_\alpha\}, \tag{2.15}$$

where c_α is the $1-\alpha$ quantile of χ^2_d.

Note that this result imposes an extra condition on the sequence of bandwidths γ_m, that is, $\sqrt{NT}\gamma_m \to 0$, which is similar to conditions used in nonparametric regression; as it is well known this last extra condition implies that the rate of convergence is not optimal. As already mentioned in other works (see, Li and Racine (2007)), in the presence of discrete covariates it is possible to improve the rate of convergence by selecting γ_m for $m = 1, \ldots, q$ to be the minimizer of the cross validation (CV) criterion function

$$CV(\gamma) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \tilde{Y}_{it} - \tilde{X}_{it}^\top \hat{\beta}_{-it}(Z_{it}) \right\}^2 \tag{2.16}$$

where $\hat{\beta}_{-it}(Z_{it}) = \left\{ \sum_{js, js \neq it} \tilde{X}_{js} \tilde{X}_{js}^\top L(Z_{js}, Z_{it}, \gamma) \right\}^{-1} \sum_{js, js \neq it} \tilde{X}_{js} \tilde{Y}_{js} L(Z_{js}, Z_{it}, \gamma)$ is the leave-one-out kernel estimator of $\beta(Z_{it})$. We use $\hat{\gamma}_1, \ldots, \hat{\gamma}_q$ to denote the cross-validated choices of $\gamma_1, \ldots, \gamma_q$ that minimize (2.16). In order to state the asymptotic properties of the cross-validated choices $\hat{\gamma}_1, \ldots, \hat{\gamma}_q$ we will need to borrow the following assumption from Feng et al. (2017)

Assumption 2.2.2. :

(i) Define $CV_0(\gamma)$ as

$$\begin{aligned} CV_0(\gamma) &= \sum_{z \in \mathscr{D}} p(z) \{\beta(z) - \eta(z,\gamma)\}^\top \Omega(z,\gamma) \{\beta(z) - \eta(z,\gamma)\} \\ &\quad + \sum_{z \in \mathscr{D}} p(z) \left\{ \Delta_{3\beta}(z,\gamma) - \Delta_3(z,\gamma)^\top \beta(z) \right\}^2 \\ &\quad +2 \sum_{z \in \mathscr{D}} p(z) \{\mu_X(z) - \Delta_3(z,\gamma)\}^\top \{\beta(z) - \eta(z,\gamma)\} \left\{ \Delta_{3\beta}(z,\gamma) - \Delta_3(z,\gamma)^\top \beta(z) \right\} \\ &= CV_{0,1} + CV_{0,2} + CV_{0,3} \end{aligned}$$

where

$$
\begin{aligned}
\Delta_1(z,\gamma) &= \mathrm{E}\{L^p(Z_{is},z,\gamma)|z,\gamma\} \\
\Delta_2(z,\gamma) &= \mathrm{E}\{X_{it}L^p(Z_{is},z,\gamma)|z,\gamma\} \\
\Delta_{2\beta}(z,\gamma) &= \mathrm{E}\{X_{it}\beta(Z_{it})L^p(Z_{is},z,\gamma)|z,\gamma\} \\
\Delta_3(z,\gamma) &= \Delta_2(z,\gamma)/\Delta_1(z,\gamma) \\
\Delta_{3\beta}(z,\gamma) &= \Delta_{2\beta}(z,\gamma)/\Delta_1(z,\gamma) \\
\Omega(z,\gamma) &= \Sigma_X(z)+\Delta_3(z,\gamma)\Delta_3(z,\gamma)^\top-\Delta_3(z,\gamma)\mu_X(z)^\top-\mu_X(z)\Delta_3(z,\gamma) \\
\Sigma_{XX}(z,\gamma) &= \mathrm{E}\{\Omega(z,\gamma)L(Z_{it},z,\gamma)|z,\gamma\} \\
\Sigma_{XX\beta}(z,\gamma) &= \mathrm{E}\{\Omega(z,\gamma)\beta(Z_{it})L(Z_{it},z,\gamma)|z,\gamma\} \\
\eta(z,\gamma) &= \Sigma_{XX}^{-1}(z,\gamma)\Sigma_{XX\beta}(z,\gamma) \\
K_{it} &= \frac{1}{T_{it}}\sum_{s=1}^{T}X_{is}L^p(Z_{is},z,\gamma)-\Delta_3(Z_{it},\gamma).
\end{aligned}
$$

(ii) $\forall z\in\mathscr{D}$, $i=1,\dots,N$ and $t=1,\dots,T$, $\Delta_3(z,\gamma)$ and $\Delta_{3\beta}(z,\gamma)$ are uniformly bounded in z. Let us suppose that, together with assumption 2.2.1(iii)-(iv), the following result holds, $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}||K_{it}||^2=O(1)$ and $(NT)^1\sum_{i=1}^{N}\sum_{t=1}^{T}|T/T_{it}|^2=O(1)$ uniformly in $\gamma_m\in[0,1]$ for $m=1,\dots,q$..

Assumption 2.2.2.(i) sets restrictions on the unconditional moments as in Assumption 2.2.1.(iii). Assumption 2.2.2.(ii) is a panel data version of assumption 2 of Li et al. (2013b) and ensures that $CV_0(\gamma)$ is uniquely optimize at 0. By theorem 2.1 of Newey and McFadden (1994), this assumption implies that $\hat{\gamma}$ obtained by minimizing (2.16) converges to zero. Under Assumptions 2.2.1 and 2.2.2 we can state the following results; for further discussion and proofs the reader should refer to Feng et al. (2017).

Lemma 2.2.1. Under Assumptions 2.2.1 and 2.2.2, as $(N,T)\to(\infty,\infty)$ jointly, $\hat{\gamma}=o_P(1)$

This lemma ensures that γ converges to zero as the sample size increases. Then it is reasonable to assume that γ is sufficiently small and close to zero. Therefore the product kernel function can be simplified as in (2.3).

Lemma 2.2.2. Assuming that conditions 2.2.1 and 2.2.2 hold, as $(N,T)\to(\infty,\infty)$ jointly, $\hat{\gamma}=O_P\left(\frac{1}{NT}\right)$

This lemma gives the rate of convergence for $\hat{\gamma}$; note that this result simplifies considerably the proof of the previous result as we are able to use an indicator function (i.e., $L(Z_{it},z,\gamma)=$

$1(Z_{it} = z)$, letting $\gamma = 0_{q\times 1}$). Note that using these results the proofs of theorem 2.2.1 will simplify considerably since we will be working with

$$\tilde{T}_i\{\beta(z)\} = \sum_{t=1}^{T} \tilde{X}_{it}\left\{\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\beta(z)\right\} 1(Z_{it} = z) + O_P\left(\frac{1}{NT}\right). \tag{2.17}$$

Using (2.17) we can build up an empirical likelihood ratio function similar to (2.13), $\tilde{\mathscr{R}}\{\beta(z)\}$ and we can state the following result.

Corollary 2.2.1. Taking $\hat{\gamma}$ to be the minimizer of the cross validation function (2.16), then assuming that conditions 2.2.1 and 2.2.2 hold, and $(N,T) \to (\infty,\infty)$ jointly, we get that $\tilde{\mathscr{R}}\{\beta(z)\} \to^d \chi_d^2$.

Here we define the confidence bands in the same way as in (2.15), that is, the set of values $\beta(z)$ such that $\tilde{\mathscr{R}}\{\beta(z)\} \le c_\alpha$ where $\Pr\left(\chi_d^2 \le c_\alpha\right) = \alpha$. Note that, using the empirical likelihood technique, it is possible to implement both, theorem 2.2.1 and corollary 2.2.1 without imposing any extra conditions on the random errors.

In the following section we obtain the maximum empirical likelihood estimator (MELE) using the empirical likelihood ratio defined in this section. Also, as the usual tool to construct confidence bands, we will provide the asymptotic distribution of the estimators.

2.3 Maximum empirical likelihood estimator

We define the maximizer of (2.13), $\hat{\beta}(z)$, as the maximum empirical likelihood estimator of $\beta(z)$, that is, $\hat{\beta}(z) = \max_{\beta(z)} \mathscr{R}\{\beta(z)\}$. Using (2.11) and (2.13) and following the same lines as in Qin and Lawless (1994) we can write

$$\hat{\beta}(z) = \left\{\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}^{\top} L(Z_{it}, z, \gamma)\right\}^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{Y}_{it} L(Z_{it}, z, \gamma) + o_P\left(\frac{1}{\sqrt{NT}}\right) \tag{2.18}$$

Consequently, for comparison purposes, we derive the asymptotic distribution of MELE estimator, (2.18), in the following theorem.

Theorem 2.3.1. Assuming that condition 2.2.1 hold, $\gamma \to 0$ and $(N,T) \to (\infty,\infty)$ jointly, then

$$\sqrt{NT}\left\{\hat{\beta}(z) - \beta(z) - \Gamma_1^{-1}(z)b(\gamma)\right\} \to^d \mathscr{N}\left\{0_{d\times 1}, \Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)\right\}$$

where

$$\begin{aligned}
\Gamma_0(z) &= \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{E}\left[v_{it}v_{js}\{X_{it}-\mu_X(z)\}\{X_{js}-\mu_X(z)\}^\top 1(Z_{it}=z)1(Z_{js}=z)\right], \\
\Gamma_1(z) &= p(z)\left\{\Sigma_X(z)-\mu_X(z)\mu_X(z)^\top\right\}+O(||\gamma||), \\
b(\gamma) &= \Gamma_1(z^*)\{\beta(z^*)-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+O\left(||\gamma||^2\right).
\end{aligned}$$

Note that by imposing stronger conditions on the random errors, i.e., v_{it} are i.i.d. over i and t, $\Gamma_0(z)$ is reduced to a simpler expression such as $\Gamma_0(z)=\sigma_v^2 p(z)\left\{\Sigma_X(z)-\mu_X(z)\mu_X(z)^\top\right\}=\sigma_v^2\Gamma_1(z)$, then we can state the following result.

Corollary 2.3.1. Assuming that condition 2.2.1 hold, v_{it} are i.i.d. over i and t, $\gamma\to 0$, and $(N,T)\to(\infty,\infty)$ jointly, then

$$\sqrt{NT}\left\{\hat{\beta}(z)-\beta(z)-\Gamma_1^{-1}(z)b(\gamma)\right\}\to^d \mathcal{N}\left\{0_{d\times 1},\sigma_v^2\Gamma_1^{-1}(z)\right\}.$$

Also note that under unknown sequences of γ and using lemma 2.2.1 and 2.2.2 the proof of theorem 2.3.1 will simplify considerably since we will be working with $\hat{\beta}(z)=\tilde{\beta}(z)+O_P\left(\frac{1}{NT}\right)$, where $\tilde{\beta}(z)$ is a frequency estimator in the same way as in $\hat{\beta}(z)$ when $\gamma_m=0$ $\forall m=1,\dots q$. Therefore, is straightforward to obtain that

$$\sqrt{NT}\left\{\hat{\beta}(z)-\beta(z)\right\}=\sqrt{NT}\left\{\tilde{\beta}(z)-\beta(z)\right\}+O_P\left(\frac{1}{\sqrt{NT}}\right); \tag{2.19}$$

then, we just need to focus on $\sqrt{NT}\left\{\tilde{\beta}(z)-\beta(z)\right\}$.

Theorem 2.3.2. Taking $\hat{\gamma}$ to be the minimizer of the cross validation function (2.16), then assuming that conditions 2.2.1 and 2.2.2 hold, and $(N,T)\to(\infty,\infty)$ jointly, we get that

$$\sqrt{NT}\left\{\tilde{\beta}(z)-\beta(z)\right\}\to^d \mathcal{N}\left\{0_{d\times 1},\Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)\right\}$$

where

$$\begin{aligned}
\Gamma_0(z) &= \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathrm{E}\left[v_{it}v_{js}\{X_{it}-\mu_X(z)\}\{X_{js}-\mu_X(z)\}^\top 1(Z_{it}=z)1(Z_{js}=z)\right], \\
\Gamma_1(z) &= p(z)\left\{\Sigma_X(z)-\mu_X(z)\mu_X(z)^\top\right\}.
\end{aligned}$$

Here, imposing that v_{it} are i.i.d. over i and t, that is, $\Gamma_0(z)=\sigma_v^2\Gamma_1(z)$ will lead us to the following result.

Corollary 2.3.2. Taking $\hat{\gamma}$ to be the minimizer of the cross validation function (2.16), then assuming that conditions 2.2.1 and 2.2.2 hold, v_{it} are i.i.d. over i and t and $(N,T) \to (\infty,\infty)$ jointly, we get that

$$\sqrt{NT}\left\{\tilde{\beta}(z) - \beta(z)\right\} \to^{d} \mathscr{N}\left\{0_{d\times 1}, \sigma_v^2 \Gamma_1^{-1}(z)\right\}.$$

Note that, when using asymptotic normality we need to estimate the variance-covariance matrix and sometimes this estimation is no feasible (see variance expressions in Theorems 2.3.1 and 2.3.2). To cope with this issue, we imposed a stronger condition on the random errors, that is, v_{it} are i.i.d. over i and t; this allowed us to estimate the variance expression using corollary 2.3.1 and 2.3.2. Hence, to construct the confidence bands, by (A.58) it is easy to show that $\hat{\Gamma}_1(z) \to_P \Gamma_1(z)$, where

$$\hat{\Gamma}_1(z) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}^{\top} 1(Z_{it} = z),$$

and if v_{it} are i.i.d. over i and t, $\hat{\sigma}_v^2 \to_P \sigma_v^2$, where

$$\hat{\sigma}_v^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\hat{\beta}(Z_{it})\right\}^2.$$

In the following section we illustrate the proposed technique in an application that reports estimates of strike activities from 17 OECD countries for the period 1951 - 1985

2.4 Empirical Application

The application reports estimates of strike activities from 17 OECD countries for the period 1951 - 1985. Strike activity is defined as the annual number of days lost per 1000 workers though industrial disputes. Strike volume is written as

$$Y_{it} = X_{it}^{\top}\beta(Z_i) + \omega_i + v_{it},$$

where Z_i is a categorical variable containing country codes that do not vary with time; Y_{it} stands for the strike volume of the country i at time t. $X_{it} = (1, U_{it}, I_{it}, P_{it}, UN_{it})^{\top}$ is a 4×1 vector containing U_{it}, unemployment, I_{it}, inflation, P_{it}, left party parliamentary representation, and UN_{it}, a time invariant measure of union centralization. As in Western (1996) we use the log transformation to stabilized the volatility of the strike series.

Empirical likelihood based inference for categorical varying coefficient panel data model with fixed effects

Continuing with our proposed methodology, we first apply the within transformation. Due to the time invariant nature of Z_i and UN_{it} we have

$$\tilde{Y}_{it} = \tilde{X}_{it}^{\top}\beta(Z_i) + \tilde{v}_{it},$$

where $\tilde{X}_{it} = \left(\tilde{U}_{it}, \tilde{I}_{it}, \tilde{P}_{it}\right)^{\top}$ is a 3×1 vector. Now we apply the empirical likelihood approach (corollary 2.2.1) and the asymptotic normality (corollary 2.3.2) to estimate the confidence bands of the parameters of interest. Here, we use corollary 2.2.1 instead of theorem 2.2.1 due to comparison purposes. The results are show in Tables 2.1-2.3; where NUB = Normal Upper Bound, NLB = Normal Lower Bound, LUB = Empirical Likelihood Upper Bound and ELLB = Empirical Likelihood Lower Bound. In Tables 1-3 we can see that the confidence bands using empirical likelihood behave better than the ones estimated using the asymptotic normal distribution.

z	NLB	ELLB	$\hat{\beta}_1(z)$	ELUB	NUB
1	-0.16	-0.02	0.00	0.06	0.16
2	-0.64	-0.49	-0.30	-0.12	0.05
3	-0.22	-0.08	-0.02	0.03	0.17
4	-0.11	-0.15	-0.02	0.10	0.08
5	-0.14	-0.06	0.04	0.15	0.22
6	-0.24	-0.12	-0.08	-0.04	0.08
7	-0.04	-0.05	0.10	0.25	0.25
8	-0.16	-0.07	-0.01	0.05	0.14
9	-0.38	-0.22	-0.19	-0.15	0.01
10	-2.59	-2.12	-1.84	-1.31	-1.09
11	-0.08	-0.14	0.01	0.14	0.10
12	-0.17	0.05	0.09	0.13	0.35
13	-0.40	0.11	0.24	0.47	0.88
14	-0.53	-0.12	0.13	0.40	0.79
15	-0.14	0.74	1.10	1.39	2.34
16	-0.10	0.01	0.05	0.10	0.19
17	-0.47	-0.28	-0.25	-0.21	-0.02

Table 2.1 Confidence bands for $\hat{\beta}_1(z)$

z	NLB	ELLB	$\hat{\beta}_1(z)$	ELUB	NUB
1	-0.00	0.05	0.07	0.13	0.15
2	-0.12	-0.23	-0.04	0.13	0.03
3	-0.03	0.03	0.08	0.14	0.20
4	0.06	0.03	0.16	0.27	0.26
5	0.02	-0.00	0.09	0.21	0.17
6	-0.10	-0.06	-0.02	0.02	0.06
7	-0.14	-0.07	0.08	0.24	0.31
8	-0.00	0.00	0.06	0.12	0.12
9	-0.05	-0.01	0.03	0.07	0.11
10	-0.08	-0.28	0.00	0.54	0.08
11	-0.18	-0.20	-0.05	0.08	0.07
12	0.06	0.09	0.13	0.17	0.21
13	-0.12	-0.15	-0.01	0.21	0.09
14	0.11	-0.04	0.21	0.48	0.31
15	-0.23	-0.38	-0.02	0.26	0.19
16	-0.02	0.02	0.05	0.11	0.12
17	-0.11	-0.03	0.00	0.04	0.11

Table 2.2 Confidence bands for $\hat{\beta}_2(z)$

z	NLB	ELLB	$\hat{\beta}_1(z)$	ELUB	NUB
1	-0.04	-0.02	0.00	0.06	0.05
2	-0.76	-0.77	-0.58	-0.41	-0.40
3	-0.01	-0.04	0.02	0.07	0.04
4	-0.04	-0.06	0.07	0.19	0.19
5	-0.02	0.02	0.11	0.23	0.24
6	-0.03	-0.05	-0.01	0.03	0.02
7	-0.19	-0.25	-0.10	0.06	-0.00
8	-0.11	0.04	0.10	0.16	0.31
9	-0.11	-0.01	0.03	0.07	0.18
10	-0.15	-0.34	-0.06	0.48	0.03
11	-0.16	-0.13	0.02	0.15	0.20
12	-0.05	-0.03	0.00	0.04	0.06
13	0.07	0.07	0.20	0.43	0.33
14	-0.14	-0.17	0.08	0.35	0.30
15	-0.19	-0.23	0.13	0.41	0.45
16	-0.09	-0.06	-0.02	0.04	0.05
17	-0.04	-0.03	0.01	0.05	0.06

Table 2.3 Confidence bands for $\hat{\beta}_3(z)$

2.5 Conclusions

Extending Li et al. (2013b)'s work to the varying coefficient panel data framework with fixed effects, we have shown that the resulting empirical log-likelihood ratio follows a Chi square distribution; therefore we are able to apply empirical likelihood methods to set up confidence bands for the functions of interest. As a by product we provide an alternative empirical maximum likelihood estimator of the categorical varying coefficients and derive its asymptotic theory. Finally we apply successfully our techniques in a empirical study of estimates of strike activities from 17 OECD countries for the period 1951 - 1985.

Chapter 3

Testing constancy in varying coefficient models

This chapter also appeared as Arteaga-Molina and Delgado González (2019).

3.1 Introduction

This chapter proposes a methodology for testing coefficients constancy in semi-varying coefficient models. Let (Y,Z,X_1,X_2) be a $\mathbb{R}^{2+k_1+k_2}-valued$ random vector defined on $(\Omega,\mathscr{F},\mathbb{P})$ such that

$$\mathrm{E}(Y|X,Z) = X_1^{\mathrm{T}}\beta_0(Z) + X_2^{\mathrm{T}}\delta_0 \; a.s., \tag{3.1}$$

where "T" means transpose, $\beta_0 = (\beta_{00},\beta_{01},...,\beta_{0k_1})^{\mathrm{T}}$, $X_1 = (1,X_{11},....,X_{1k_1})^{\mathrm{T}}$ and $\delta_0 = (\delta_{01},...,\delta_{0k_2})^{\mathrm{T}}$ and $X_2 = (X_{21},....,X_{2k_2})^{\mathrm{T}}$, $\beta_0 : \mathbb{R} \to \mathbb{R}^{1+k_1}$ is a vector of unknown functions, and δ_0 is an unknown parameter vector in $\mathbb{R}^{k_2}$. Henceforth, the discussion is centered on the case where the constant term is in X_1, but the procedure also applies to the case where there is a constant intercept, i.e. when $X_1 = (X_{11},....,X_{1k_1})^{\mathrm{T}}$ and $X_2 = (1,X_{21},....,X_{2k_2})^{\mathrm{T}}$. The model with constant slopes, i.e. $Var\left(\beta_{0j}(Z)\right) = 0$ all $j = 1,...,k_1$, is known as partly linear model, and inferences on $\beta_{00}(\cdot)$ and δ_0 have been justified under different regularity conditions by Shiller (1984), Wahba (1985), Engle et al. (1986), Heckman (1986), Schick (1986), Speckman (1988), Chen (1988) and Robinson (1988) among others. This requires estimating the nonparametric regression functions of Y given Z and of each X_2 component given Z. Inferences when all the coefficients are varying, i.e. when $\delta_0 = 0$, have been proposed by Cleveland et al. (1991), Hastie and Tibshirani (1993), Chan and Tsay (1998), McCabe and Tremayne (1995), Wu et al. (1998), Fan and Zhang (1999, 2000), Chiang et al. (2001), Hoover et al. (1998), Cai et al. (2000), Kim (2007), Hoderlein and Sherman (2015) or Feng et al. (2017). The semi-varying coefficient model, with $\delta_0 \neq 0$, has been studied by Zhang et al. (2002), Xia et al. (2004), Ahmad et al. (2005), Fan and Huang (2005), Li et al. (2011b), Li et al. (2011a), Hu and Xia (2012), or Li et al. (2017) among others. All these methods use smooth estimators of the underlying nonparametric functions, generally Nadaraya-Watson kernel regression.

Model (3.1) nests discontinuous regression models where

$$\beta_0(z) = \bar{\beta}_{00} + \bar{\beta}_{01} 1_{\{z \leq z_0\}}, \tag{3.2}$$

for parameter vectors $\bar{\beta}_{00}$ and $\bar{\beta}_{01}$, where the discontinuity is explained by the variable Z, which is the typical alternative to parameter stability hypothesis in time series analysis, with parameters changing at an unknown time point. It is not possible consistently estimating β_0 in model (3.2) using smoothing based methods.

The goal of this article consists of testing that the varying coefficients in model (3.1) are constant in the direction of nonparametric alternatives, i.e., testing

$$H_0 : Var\left(\beta_{0j}\right) = 0 \text{ for all } j = 0,1,...,k_1 \text{vs. } H_1 : Var\left(\beta_{0j}\right) \neq 0 \text{ for some } j = 0,1,...,k_1. \tag{3.3}$$

Model (3.1) also nests a model with

$$X_2 = \left(g_0^{\mathrm{T}}(Z), X_{11}g_1^{\mathrm{T}}(Z), ..., X_{1k_1}g_{k_1}^{\mathrm{T}}(Z)\right)^{\mathrm{T}}, \ \ \delta_0 = \left(\delta_{00}^{\mathrm{T}}, \delta_{01}^{\mathrm{T}}, ..., \delta_{0k_1}^{\mathrm{T}}\right)^{\mathrm{T}} \text{ and } k_2 = \sum_{j=0}^{k_1} m_j, \tag{3.4}$$

where δ_{0j} are unknown $m_j \times 1$ parameter vectors, and $g_j : \mathbb{R} \to \mathbb{R}^{m_j}$ are known functions, $j = 0,...,k_1$. In this case, (3.1) can be expressed as

$$\mathbb{E}\left(Y|X,Z\right) = X_1^{\mathrm{T}}\left[\beta_0(Z) + \mu_0(Z)\right] \ a.s. \tag{3.5}$$

with nonparametric β_0 and parametric

$$\mu_0(\cdot) = \left(g_0^{\mathrm{T}}(\cdot)\delta_{00},\ g_1^{\mathrm{T}}(\cdot)\delta_{01}, ..., g_{k_1}^{\mathrm{T}}(\cdot)\delta_{0k_1}\right)^{\mathrm{T}},$$

for some $\delta_0 = \left(\delta_{00}^{\mathrm{T}}, \delta_{01}^{\mathrm{T}}, ..., \delta_{0k_1}^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^{k_2}$. Therefore, under the maintained hypothesis (3.3) and (3.5) are equivalent to omnibus model checking that the marginal effects of X_1 are $\mu_0(Z)$, which implies a particular parameterization of the interactive effects. Needless to say that model (3.5) is not identifiable in many circumstances, but the test we propose does not need estimating the model under the alternative hypothesis. In particular, our test is in fact a directional specification test for the linear in parameters regression model in the direction of a semi-varying coefficient model. It can also be applied as an omnibus specification test of a simple regression model with explanatory variable Z, i.e. $k_1 = 0$ in (3.5).

Kauermann and Tutz (1999), Cai et al. (2000), Fan and Zhang (2000), Fan et al. (2001), Fan and Huang (2005), Qu and Li (2006) and Cai et al. (2017) have considered testing (3.3) based on the discrepancy between restricted and unrestricted sum of squared residuals using smooth estimates of the varying coefficients. In these proposals, smooth estimates of β_{0j} are needed and, hence, situations like (3.2) are ruled out. Also,these tests are not applicable when the model on the alternative is not identified.

In this chapter we adapt classical parameter stability tests in time series (e.g. Quandt (1958, 1960); Chernoff and Zacks (1964); Bhattacharyya and Johnson (1968); Hinkley (1970); Brown et al. (1975); Sen and Srivastava (1975); Hawkins (1989, 1977); Nyblom

(1989); Andrews (1993); Csörgő and Horváth (1988, 1997); Aue et al. (2008) among many others.)

Given $(Y_i, Z_i, X_{1i}, X_{2i})_{i=1}^n$ *i.i.d.* as (Y, Z, X_1, X_2), we interpret $(Y_i, X_{1i}, X_{2i})_{i=1}^n$ as sequentially observed with respect to the ordered values of $\{Z_i\}_{i=1}^n$. That is, denote by $\left(Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\right)_{i=1}^n$ the $Z-$concomitants, or induced order statistics, of $(Y_i, X_{1i}, X_{2i})_{i=1}^n$, i.e. $\left(Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\right) = \left(Y_j, X_{1j}, X_{2j}\right)$ iff $Z_{(n:i)} = Z_j$, where $Z_{(n:1)} \leq Z_{(n:2)} \leq ... \leq Z_{(n:n)}$ are the ordered statistics of $\{Z_i\}_{i=1}^n$. We propose to adapt union-intersection (U-I) type tests in time series to our context. See Hawkins (1989), Andrews (1993), Horváth and Shao (1995) or Csörgő and Horváth (1997) Section 3.1.5. The test consists of comparing ordinary least squares (OLS) estimators of X_1 coefficients using subsamples $\left(Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\right)_{i=1}^j$ and $\left(Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\right)_{i=j+1}^n$ at each $j-th$ sample $Z-$quantile.

The rest of the chapter is organized as follows. Next section discusses and justifies the testing procedure. Section 3.3 studies the finite sample performance of the test in the context of a Monte Carlo experiment. We report comparisons of existing tests for coefficient constancy based on smooth β_0 estimates, as well as specification CUSUM type tests, as proposed by Stute (1997) and Andrews (1997), which are omnibus, i.e. designed to detect any alternative, much broader than H_1 in (3.1). In Section 3.4 we apply the testing procedure to modeling interactive effects of IQ when studying education returns. Section 3.5 is devoted to conclusions. Mathematical proofs can be found in an appendix at the end of the article.

3.2 Testing Method

Define $M_{\ell j}(u) = \mathbb{E}\left(X_\ell X_j^\intercal 1_{\{F_Z(Z)\leq u\}}\right)$ and $S_j(u) = \mathbb{E}\left(X_j Y 1_{\{F_Z(Z)\leq u\}}\right)$, $j, \ell = 1, 2$, where F_Z is the cdf of Z. Assume that

Assumption 3.2.1. F_Z is continuous.

Assumption 3.2.2.

$$Rank\left\{\begin{bmatrix} M_{11}(u) & M_{12}(u) \\ M_{21}(u) & M_{22}(1)\end{bmatrix}\right\} = k_1 + k_2 + 1 \text{ for all } u \in [0,1].$$

For the sake of exposition assume w.l.o.g. that Z is uniformly distributed on $[0,1]$. An U-I test of H_0 is based on the sample version of $\eta_0(u) = \left(\theta_0^- - \theta_0^+\right)(u)$, where $\theta_0(u) =$

$\left(\theta_0^{-\mathrm{T}}(u),\theta_0^{+\mathrm{T}}(u),\theta_0^{o\mathrm{T}}(u)\right)^{\mathrm{T}}$,

$$\begin{aligned}\theta_0(u) &= \underset{\theta^-,\theta^+,\theta^o}{\arg\min}\Big\{\mathbb{E}\left[\left(Y-X_1^{\mathrm{T}}\theta^- - X_2^{\mathrm{T}}\theta^o\right)1_{\{Z\le u\}}\right]^2 \\ &\qquad\qquad +\mathbb{E}\left[\left(Y-X_1^{\mathrm{T}}\theta^+ - X_2^{\mathrm{T}}\theta^o\right)1_{\{Z> u\}}\right]^2\Big\} \\ &= M^{-1}(u)S(u),\ u\in[0,1],\end{aligned} \tag{3.6}$$

$$M(u)=\begin{bmatrix} M_{11}(u) & 0_{k_1+1} & M_{12}(u) \\ 0_{k_1+1} & M_{11}(1)-M_{11}(u) & M_{12}(1)-M_{12}(u) \\ M_{21}(u) & M_{21}(1)-M_{21}(u) & M_{22}(1)\end{bmatrix},$$

0_m is a $m\times m$ matrix of zeroes, and $S(u)=\left(S_1^{\mathrm{T}}(u),[S_1(1)-S_1(u)]^{\mathrm{T}},S_2^{\mathrm{T}}(1)\right)^{\mathrm{T}}$.

Obviously, $Var\left(\beta_{0j}(Z)\right)=0$ for all $j=0,...,k_1$ implies that $\eta_0(u)=0$ for all $u\in[0,1]$. In relevant circumstances, discussed below, also $Var\left(\beta_{0j}(Z)\right)=0 \ \ iff \ \ \eta_0(u)=0$ for all $u\in[0,1]$.

Remark 3.2.1. Consider $M_{11}(u)=uM_{11}(1)$, which is always satisfied in the partly linear model, and $M_{12}(u)=uM_{12}(1)$ for all $u\in[0,1]$, which is equivalent to $\mathbb{E}\left(X_1X_1^{\mathrm{T}}|Z\right)=M_{11}(1)$ *a.s.* and $\mathbb{E}\left(X_1X_2^{\mathrm{T}}|Z\right)=M_{12}(1)$ *a.s.* Therefore,$S_1(u)=M_{11}(1)\mathbb{E}\left(\beta_0(Z)1_{\{Z\le u\}}\right)$, and applying Lemma A.5 in Andrews (1993),

$$\begin{aligned}\eta_0(u) &= \frac{\mathbb{E}\left(\beta_0(Z)1_{\{Z\le u\}}\right)-u\mathbb{E}\left(\beta_0(Z)\right)}{u(1-u)} \\ &= \frac{1}{u(1-u)}\int_{\{Z\le u\}}\left[\beta_0(Z)-\mathbb{E}\left(\beta_0(Z)\right)\right]d\mathbb{P} \\ &= 0 \text{ for all } u\in[0,1]\Leftrightarrow \beta_0(Z)=\mathbb{E}\left(\beta_0(Z)\right)\ a.s.\end{aligned}$$

Remark 3.2.2. Consider $\delta_0=0$, i.e. a pure varying coefficient model. Then, for all $u\in[0,1]$,

$$\begin{aligned}\eta_0(u) &= M_{11}^{-1}(u)S_1(u)-\left[M_{11}(1)-M_{11}(u)\right]^{-1}\left[S_1(1)-S_1(u)\right] \\ &= \left[M_{11}(1)-M_{11}(u)\right]^{-1}M_{11}(1)\left[M_{11}^{-1}(u)S_1(u)-M_{11}^{-1}(1)S_1(1)\right],\end{aligned}$$

and

$$\begin{aligned}\eta_0(u)=0 \text{ all } u\in[0,1] &\Leftrightarrow S_1(u)-M_{11}(u)M_{11}^{-1}(1)S_1(1)=0 \text{ all } u\in[0,1] \\ &\Leftrightarrow \int_{\{Z\le u\}}J(Z)\left[\beta_0(Z)-\mathbb{E}\left(J(Z)\right)^{-1}\mathbb{E}\left(J(Z)\beta_0(Z)\right)\right]d\mathbb{P}=0.\end{aligned}$$

with $J(Z) = \mathbb{E}\left(X_1 X_1^{\mathrm{T}} | Z\right)$. Hence, if $J(Z)$ is non-singular a.s.,

$$\begin{aligned}\eta_0(u) = 0 \text{ all } u \in [0,1] \quad &\Leftrightarrow \quad \beta_0(Z) = \mathbb{E}\left(J(Z)\right)^{-1} \mathbb{E}\left(J(Z)\beta_0(Z)\right) \; a.s. \\ &\Leftrightarrow \quad Var\left(\beta_{0j}(Z)\right) = 0 \text{ for all } j = 0, ..., k_1.\end{aligned}$$

The above two remarks show that testing

$$\bar{H}_0 : \eta_0(u) = 0 \text{ all } u \in [0,1] \text{ vs. } \bar{H}_1 : \eta_0(u) \neq 0 \text{ some } \; u \in [0,1].$$

is equivalent to (3.3) in pure varying coefficient regression models, as well as in situations where the elements in $X_1 X_1^{\mathrm{T}}$ are mean independent of Z. Also, rejecting $\bar{H}_0$ in the direction of $\bar{H}_1$ implies rejecting H_0 in the direction of H_1 for many other semi-varying coefficient models, as discussed in Section 3.3.

The sample analog of (3.6) is $\hat{\theta}_n(u) = \left(\hat{\theta}_n^{-\mathrm{T}}(u), \hat{\theta}_n^{+\mathrm{T}}(u), \hat{\theta}_n^{o\mathrm{T}}(u)\right)^{\mathrm{T}}$, with

$$\begin{aligned}\hat{\theta}_n(u) \quad &= \quad \underset{\theta^-,\theta^+,\theta^o}{\arg\min} \left\{ \sum_{i=1}^{\lfloor nu \rfloor} \left(Y_{[i:n]} - X_{1[i:n]}^{\mathrm{T}} \theta^- - X_{2[i:n]}^{\mathrm{T}} \theta^o \right)^2 \right. \\ &\qquad \left. + \sum_{i=1+\lfloor nu \rfloor}^{n} \left(Y_{[i:n]} - X_{1[i:n]}^{\mathrm{T}} \theta^+ - X_{2[i:n]}^{\mathrm{T}} \theta^o \right)^2 \right\} \\ &= \quad \hat{M}_n^{-1}(u) \hat{S}_n(u), \; u \in [0,1],\end{aligned}$$

where $\lfloor \cdot \rfloor$ means smallest nearest integer, $\hat{S}_n(u) = \left(\hat{S}_{n1}^{\mathrm{T}}(u), \hat{S}_{n1}^{\mathrm{T}}(1) - \hat{S}_{n1}^{\mathrm{T}}(u), \hat{S}_{n2}^{\mathrm{T}}(1)\right)^{\mathrm{T}}$, $\hat{S}_{nj}(u) = \sum_{i=1}^{\lfloor nu \rfloor} X_{j[i:n]} Y_{[i:n]}$, $j = 1, 2$,

$$\hat{M}_n(u) = \begin{bmatrix} \hat{M}_{11n}(u) & 0_{k_1+1} & \hat{M}_{12n}(u) \\ 0_{k_1+1} & \hat{M}_{11n}(1) - \hat{M}_{11n}(u) & \hat{M}_{12n}(1) - \hat{M}_{12n}(u) \\ \hat{M}_{21n}(u) & \hat{M}_{21n}(1) - \hat{M}_{21}(u) & \hat{M}_{n22}(1) \end{bmatrix},$$

and $\hat{M}_{n\ell j}(u) = n^{-1} \sum_{i=1}^{\lfloor nu \rfloor} X_{\ell[i:n]} X_{j[i:n]}^{\mathrm{T}}$, $\ell, j = 1, 2$. Similar expressions can be found in time series parameter stability testing. This suggests test statistics for $\bar{H}_0$ based on suitable functionals of

$$\hat{\eta}_n(u) = \left(\hat{\theta}_n^- - \hat{\theta}_n^+\right)(u) = R \hat{M}_n^{-1}(u) \hat{S}_n(u), \tag{3.7}$$

with $R = \left[I_{k_1+1} \vdots - I_{k_1+1} \vdots \, 0_{k_2} \right]$ and I_m is a $m \times m$ identity matrix, which is the difference between (OLS) estimators of X_1 coefficients under H_0 using subsamples $\left(Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\right)_{i=1}^{j}$ and $\left(Y_{[i:n]}, X_{1[i:n]}, X_{2[i:n]}\right)_{i=j+1}^{n}$.

Notice that

$$\hat{\theta}_n(u) = \theta_0(u) + \hat{M}_n^{-1}(u)\hat{N}_n(u),$$

$\hat{N}_n(u) = \left(\hat{N}_{n1}^{\mathrm{T}}(u), \hat{N}_{n1}^{\mathrm{T}}(1) - \hat{N}_{n1}^{\mathrm{T}}(u), \hat{N}_{n2}^{\mathrm{T}}(u)\right)^{\mathrm{T}}$, and $\hat{N}_{nj}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor} X_{j[i:n]}U_{[i:n]}$, $j = 1,2$ and $U_i = Y_i - X_{1i}^{\mathrm{T}}\beta_0(Z) - X_{2i}^{\mathrm{T}}\delta_0,$. The asymptotic distribution of $\hat{N}_n$ is obtained applying results for partial sums of concomitants in Bhattacharya (1974, 1976), extended by Sen (1976), Stute (1993, 1997) or Davydov and Egorov (2000), among others. Define $N_\infty(u) = (N_{\infty 1}^{\mathrm{T}}(u), N_{\infty 1}^{\mathrm{T}}(1) - N_{\infty 1}^{\mathrm{T}}(u), N_{\infty 2}^{\mathrm{T}}(u))^{\mathrm{T}}$, where $N_{\infty j}$ be $k_j \times 1$, $j = 1,2$, vectors of centered Gaussian processes with $\mathbb{E}\left(N_{\infty \ell}(u)N_{\infty j}^{\mathrm{T}}(v)\right) = \mathbb{E}\left(X_\ell X_j^{\mathrm{T}} U^2 1_{\{Z \le u \wedge v\}}\right)$, $\ell, j = 1,2$. Next assumption suffices to show weak convergence of $\sqrt{n}\hat{N}_n$ and uniform convergence of $\hat{M}_n$.

Assumption 3.2.3. $\mathbb{E}\|XU\|^2 < \infty$.

Henceforth, for any matrix A, $\|A\| = \sqrt{\bar{\lambda}\left(A^{\mathrm{T}}A\right)}$ is the spectral norm, where $\bar{\lambda}(C)$ is the maximum eigenvalue of the matrix C, and "$\to_d$ " means convergence in distribution of random variables, random vectors or random elements in a Skorohov's space $D[a,b]$, $0 \le a < b \le 1$.

Proposition 3.2.1. Assuming that conditions 3.2.1, 3.2.2 and 3.2.3 hold,

$$\sqrt{n}\left(\hat{N}_{n1}^{\mathrm{T}}, \hat{N}_{n2}^{\mathrm{T}}\right)^{\mathrm{T}} \to_d \left(N_{\infty 1}^{\mathrm{T}}, N_{\infty 2}^{\mathrm{T}}\right)^{\mathrm{T}} \text{ in } D[0,1], \tag{3.8}$$

and

$$\lim_{n\to\infty} \sup_{u\in[0,1]} \left\|\left(\hat{M}_{n\ell j} - M_{\ell j}\right)(u)\right\| = 0 \; a.s., \; \ell, j = 1,2. \tag{3.9}$$

Therefore, since

$$\hat{\eta}_n(u) = \left(\theta_0^- - \theta_0^+\right)(u) + R\hat{M}_n^{-1}(u)\hat{N}_n(u),$$

under $\bar{H}_0$ and conditions in Proposition 3.2.1,

$$\sqrt{n}\hat{\eta}_n \to_d \eta_\infty \; in \; D[\varepsilon, 1-\varepsilon], \; \varepsilon \in (0,1),$$

where $\eta_\infty(u) \stackrel{d}{=} R^{\mathrm{T}}M^{-1}(u)N_\infty^0(u)$ with

$$N_\infty^0(u) = \left(N_{\infty 1}^{0\mathrm{T}}(u), N_{\infty 1}^{0\mathrm{T}}(1) - N_{\infty 1}^{0\mathrm{T}}(u), N_{\infty 2}^{0\mathrm{T}}(u)\right)^{\mathrm{T}},$$

and $N_{\infty\ell}^0$, $\ell = 1,2$ is a vector of mean zero Gaussian processes with $\mathbb{E}\left(N_{\infty\ell}^0(u)N_{\infty j}^{0\mathrm{T}}(v)\right) = \mathbb{E}\left(X_\ell X_j^{\mathrm{T}} V^2 1_{\{Z \le u\wedge v\}}\right) =: \Omega_{0\ell j}(u \wedge v)$, with $V = Y - X_1^{\mathrm{T}}\bar{\beta}_0 - X_2^{\mathrm{T}}\bar{\delta}_0$ uncorrelated with the components of $(X_1^{\mathrm{T}}, X_2^{\mathrm{T}})^{\mathrm{T}}$ and $\left(\bar{\beta}_0^{\mathrm{T}}, \bar{\delta}_0^{\mathrm{T}}\right)^{\mathrm{T}} \in \mathbb{R}^{1+k_1+k_2}$. That is $\left(\bar{\beta}_0, \bar{\delta}_0\right)$ are the parameters of the best linear predictor of Y given (X_1, X_2), and $V = U$ a.s. under $\bar{H}_0$. Weak convergence of $\sqrt{n}\hat{\eta}_n$ in $D[0,1]$ is not possible even assuming that Z is independent of X and U, as shown by Chibisov (1964) for the standard empirical process (see subsection 2.5 in Gaenssler and Stute (1979) for discussion).

Therefore, for $\varepsilon \in (0,1)$,

$$\mathbb{E}\left(\eta_\infty(u)\eta_\infty^{\mathrm{T}}(v)\right) = \Sigma_0(u \wedge v) = R^{\mathrm{T}} M^{-1}(u)\Omega_0(u \wedge v)M^{-1}(v)R,\ u, v \in (\varepsilon, 1-\varepsilon),$$

with $\Omega_0(u) = \mathbb{E}\left(N_\infty^0(u) N_\infty^{0\mathrm{T}}(u)\right)$. Applying the U-I testing principle, this suggests tests based on functionals of the empirical process,

$$\hat{\alpha}_n(u) = \hat{\eta}_n^{\mathrm{T}}(u)\hat{\Sigma}_n^{-1}(u)\hat{\eta}_n(u), u \in (\varepsilon, 1-\varepsilon).$$

where

$$\hat{\Sigma}_n(u) = R^{\mathrm{T}}\hat{M}_n^{-1}(u)\hat{\Omega}_n(u)\hat{M}_n^{-1}(u)R,$$

estimates $\Sigma_0(u)$, and

$$\hat{\Omega}_n(u) = \begin{bmatrix} \hat{\Omega}_{n11}(u) & 0_{k_1+1} & \hat{\Omega}_{n12}(u) \\ 0_{k_1+1} & \hat{\Omega}_{n11}(1) - \hat{\Omega}_{n11}(u) & \hat{\Omega}_{n12}(1) - \hat{\Omega}_{n12}(u) \\ \hat{\Omega}_{n21}(u) & \hat{\Omega}_{n21}(1) - \hat{\Omega}_{n21}(u) & \hat{\Omega}_{n22}(1) \end{bmatrix}$$

estimates $\Omega_0(u)$, with $\hat{\Omega}_{n\ell j}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor} X_{\ell[i:n]} X_{j[i:n]}^{\mathrm{T}} \hat{V}_{[i:n]}^2$, and $\ell, j = 1,2$; also let $\hat{V}_i = Y_i - X_{1i}^{\mathrm{T}}\widehat{\theta}_n^{+}(1) - X_{2i}^{\mathrm{T}}\widehat{\theta}_n^{o}(1)$ be the OLS residuals under $\bar{H}_0$. A sufficient condition for consistency of $\hat{\Omega}_n(u)$ is

Assumption 3.2.4. $\mathbb{E}\|X\|^4 < \infty$ and $\mathbb{E}\|V\|^4 < \infty$.

This condition can be relaxed by assuming that $\mathbb{E}\left(V^2 \middle| X, Z\right) = \mathbb{E}\left(V^2\right) = \sigma^2$ *a.s.*, which implies that $\Omega_0(u) = \sigma^2 M_n(u)$ and $\hat{\Sigma}_n(u) = \sigma^2 R^{\mathrm{T}}\hat{M}_n^{-1}(u)R$. Consider the U-I type test,

$$\hat{\varphi}_{n\varepsilon} = \max_{\lfloor n\varepsilon \rfloor + K \le j \le \lfloor n(1-\varepsilon) \rfloor - K} n \cdot \hat{\alpha}_n\left(\frac{j}{n}\right) \text{ for small } \varepsilon \in \left(0, \frac{1}{2} - \frac{K}{n}\right], \text{ with } K < \frac{n}{2},$$

and $K = k_1 + k_2 + 1$. The trimming parameter ε is introduced to avoid boundary points and should be chosen as close to zero as possible in order to detect any possible coefficient variation on all its domain, including those close to the boundary. However, too small ε values can produce serious size distortions (see, section 3.3). The asymptotic distribution of $\hat{\varphi}_{n\varepsilon}$ is derived as an immediate consequence of proposition 3.2.1 after showing uniform consistency of $\hat{\Sigma}_n$. Define the vector of random processes,

$$\{\alpha_\infty(u)\}_{u\in[\varepsilon,1-\varepsilon]} \overset{d}{=} \left\{N_\infty^{0\mathrm{T}}(u)M^{-1}(u)R^{\mathrm{T}}\Sigma_0^{-1}(u)RM^{-1}(u)N_\infty^{0}(u)\right\}_{u\in[\varepsilon,1-\varepsilon]}.$$

Next proposition establishes the asymptotic distribution of $\hat{\varphi}_{n\varepsilon}$ as a consequence of Proposition 3.2.1, after providing consistency of $\hat{\Sigma}_n(u)$ uniformly on $u \in [\varepsilon, 1-\varepsilon]$.

Proposition 3.2.2. Assume that conditions 3.2.1 - 3.2.4 hold. Under $\bar{H}_0$, for any small fixed $\varepsilon \in (0, 1/2 - K/n]$, $K < n/2$

$$\hat{\varphi}_{n\varepsilon} \to_d \varphi_{\infty\varepsilon} \overset{d}{=} \sup_{u\in[\varepsilon,1-\varepsilon]} \alpha_\infty(u).$$

Therefore, a test with α significance level is given by the binary random variable $\hat{\Phi}_{n\varepsilon}(\alpha) = 1_{\{\hat{\varphi}_{n\varepsilon} > c_\varepsilon(\alpha)\}}$, where $c_\varepsilon(\alpha)$ is the $(1-\alpha)-th$ quantile of $\varphi_{\infty\varepsilon}$.

These U-I tests in time series are asymptotically distribution-free under suitable regularity conditions, which has a counterpart in our context assuming that

Assumption 3.2.5. Z is independent of X and U.

Of course, this assumption is not acceptable in practice, but it is worth discussing to illustrate the relation of our proposal with related ones in time series parameter instability testing and the behaviour of our test statistic when ε is too small. Consider the $\delta_0 = 0$ case for simplicity. Under assumption 3.2.5, $M_{1j}(u) = uM_{1j}(1)$, $\Omega_{1j}(u) = \sigma^2 \cdot u \cdot M_{1j}(1)$, $j = 1,2$, $\{N_{\infty 1}(u)\}_{u\in[0,1]} \overset{d}{=} \left\{M_{11}^{1/2}(1)\cdot W_0(u)\right\}_{u\in[0,1]}$, W_0 is a $(1+k_1)\times 1$ vector of independent Wiener's processes and

$$\Sigma_0(u) = \sigma^2 R^{\mathrm{T}} \begin{bmatrix} uM_{11}(1) & 0 \\ 0 & (1-u)M_{11}(1) \end{bmatrix}^{-1} R = \sigma^2 \frac{M_{11}(1)}{u(1-u)}. \tag{3.10}$$

Therefore, under assumption 3.2.5,

$$\varphi_{\infty\varepsilon} \overset{d}{=} \sup_{u\in[\varepsilon,1-\varepsilon]} \frac{B_0(u)}{u(1-u)}, \tag{3.11}$$

where $B_0(u) = [W_0(u) - uW_0(1)]^{\mathrm{T}} [W_0(u) - uW_0(1)]$ is the sum of $1+k_1$ squared independent Brownian bridges. The distribution of $\varphi_{\infty\varepsilon}$ has been tabulated by James et al. (1987) for B_0 scalar and different values of ε, and by Andrews (1993) in the general case.

Under assumption 3.2.5, one can exploit the information in (3.10) and, after estimating σ^2 by $\tilde{\sigma}_n^2 = n^{-1}\sum_{i=1}^n \hat{V}_{ni}^2$, use as test statistic,

$$\tilde{\varphi}_n^{(0)} = n \cdot \max_{K \le j \le n-K} \tilde{\alpha}_n\left(\frac{j}{n}\right) ,$$

with

$$\tilde{\alpha}_n(u) = \hat{\eta}_n^{\mathrm{T}}(u)\,\frac{\hat{M}_{11n}(1)u(1-u)}{\tilde{\sigma}_n^2}\hat{\eta}_n(u),\; u \in [0,1],$$

which resembles the classical U-I tests avoiding any trimming. This statistics, suitably standardized, converges to a extremum distribution, which is proved applying Darling et al. (1956) type results for normalized partial sums. To this end, we need the alternative conditions that replace 3.2.3 and 3.2.4 by,

Assumption 3.2.6. $\mathbb{E}\left|U\right|^{2+\delta} < \infty$ and $\mathbb{E}\left\|X\right\|^{2+\delta} < \infty$ for some $\delta > 0$.

This implies, under condition 3.2.5, that $\mathbb{E}\left\|XU\right\|^{2+\delta} < \infty$, which is stronger than assumption 3.2.6. These type of moment conditions was proposed by Shorack (1979) to extend Darling et al. (1956) result to allow less than three moments. These can be further relaxed using Einmahl (1989) moment condition. Henceforth, $\Gamma(x) = \int_0^\infty y^{x-1}e^{-y}dy$, , and E is a random variable such that $\mathbb{P}(E \le x) = \exp(-2\exp(-x))$, $a(x) = \sqrt{2\log x}$ and $b_m(x) = 2\log x + (m/2)\log\log x - \log\Gamma(m/2)$. The convergence of $\tilde{\varphi}_n^{(0)}$ is slow, which results in a poor size accuracy, and some alternatives may be preferred when condition 3.2.5 is satisfied.

We can also consider the Cramér-von Mises type statistic

$$\tilde{\varphi}_n^{(1)} = \sum_{j=K}^{n-K} \tilde{\alpha}_n\left(\frac{j}{n}\right) ,$$

and the unweighted statistic

$$\tilde{\varphi}_n^{(2)} = \max_{K \le j \le n-K} \frac{j(n-j)}{n}\,\tilde{\alpha}_n\left(\frac{j}{n}\right) ,$$

which both have better size accuracy under condition 3.2.5 than the test based on $\tilde{\varphi}_n^{(0)}$. Next proposition provides the limiting distribution of $\tilde{\varphi}_n^{(j)}$, $j = 0,1,2$ under H_0.

Proposition 3.2.3. Assume $\delta_0 = 0$, and assumptions 3.2.1, 3.2.2, 3.2.6 and 3.2.5 hold, under $\bar{H}_0$,

$$a(\log n)\sqrt{\tilde{\varphi}_n^{(0)}} - b_{1+k_1}(\log n) \xrightarrow{d} E, \tag{3.12}$$

$$\tilde{\varphi}_n^{(1)} \xrightarrow{d} \int_0^1 \frac{B_0(u)}{u(1-u)} du, \tag{3.13}$$

$$\tilde{\varphi}_n^{(2)} \xrightarrow{d} \sup_{u\in[0,1]} B_0(u). \tag{3.14}$$

This suggests that, because the rate of convergence of $\hat{\varphi}_{n\varepsilon}$ changes suddenly at $\varepsilon = 0$, tests based on critical values of the asymptotic approximation (3.12) are expected to exhibitpoor size accuracy. See simulations in section 3.3.

Next, we study the power of the test in the direction of sequences of local alternatives of the form,

$$\bar{H}_{n1} : \beta(Z) = \bar{\beta}_0 + \frac{\tau(Z)}{\sqrt{n}} \; a.s.,$$

for constant $\bar{\beta}_0$ and a function $\tau : \mathbb{R} \to \mathbb{R}^{1+k_1}$ such that $T(u) = \mathbb{E}\left[X_1^{\mathrm{T}}\tau(Z)1_{\{U\le u\}}\right]$ is bounded for all $u \in [0,1]$. Define $T(u) = \left[T^{\mathrm{T}}(u), T^{\mathrm{T}}(1) - T^{\mathrm{T}}(u), 0_{k_2}^{\mathrm{T}}\right]^{\mathrm{T}}$ and the random processes,

$$\left\{\alpha_\infty^1(u)\right\}_{u\in[\varepsilon,1-\varepsilon]} \stackrel{d}{=} \left\{(N_\infty + T)^{\mathrm{T}}(u) M^{-1}(u) R^{\mathrm{T}} \Sigma_0^{-1}(u) R M^{-1}(u) \left(N_\infty + T\right)(u)\right\}_{u\in[\varepsilon,1-\varepsilon]}.$$

In order to study the power of the test under $\bar{H}_{n1}$, we need the following extra assumption.

Assumption 3.2.7. $\mathbb{E}\|X_1\tau(Z)\| < \infty$.

Proposition 3.2.4. Assume that conditions 3.2.1 - 3.2.4 and 3.2.7 hold for $\varepsilon \in \left(0, (n-2K)/2n\right]$,$K < n/2$. Under $\bar{H}_1$,

$$\hat{\varphi}_{n\varepsilon} \to_p \infty, \tag{3.15}$$

and under $\bar{H}_{1n}$,

$$\hat{\varphi}_{n\varepsilon} \to_d \sup_{u\in[\varepsilon,1-\varepsilon]} \alpha_\infty^1(u), \tag{3.16}$$

Therefore, the test does not have trivial power in the direction of $\bar{H}_{n1}$ when $\sup_{u\in[\varepsilon,1-\varepsilon]} \gamma(u) > 0$ with

$$\gamma(u) = T^{\mathrm{T}}(u) M^{-1}(u) R^{\mathrm{T}} \Sigma_0^{-1}(u) R M^{-1}(u) T(u).$$

Under condition 3.2.5,

$$\gamma(u) = \frac{T(u)^{\mathrm{T}} M_{11}^{-1}(1) T(u)}{\sigma^2 \cdot u(1-u)}.$$

This suggests choosing ε as small as possible in order to give more weight to the extreme values of Z.

The bootstrapped test statistic is

$$\hat{\varphi}^*_{n\varepsilon} = n \sup_{K+\lfloor n\varepsilon \rfloor \le j \le n-K-\lfloor n\varepsilon \rfloor} \hat{\alpha}^*_n\left(\frac{j}{n}\right) \text{ for small } \varepsilon \in \left(0, \frac{1}{2} - \frac{K}{n}\right],\ K < \frac{n}{2},$$

with

$$\hat{\alpha}^*_n(u) = \hat{\eta}^{*\mathrm{T}}_n(u)\hat{\Sigma}^{-1}_n(u)\hat{\eta}^*_n(u),$$

and

$$\hat{\eta}^*_n(u) = R\hat{M}^{-1}_n(u)\hat{N}^*_n(u),$$

where $\hat{N}^*_n(u) = \left(\hat{N}^{*\mathrm{T}}_{n1}(u), \hat{N}^{*\mathrm{T}}_{n1}(1) - \hat{N}^{*\mathrm{T}}_{n1}(u), \hat{N}^{*\mathrm{T}}_{n2}(u)\right)^{\mathrm{T}}$, $\hat{N}^*_{nj}(u) = n^{-1}\sum_{i=1}^{\lfloor nu \rfloor} X_{j[i:n]}\hat{V}^*_{[i:n]}$, $j = 1, 2$, $\hat{V}^*_i = \hat{V}_i\xi_i$, $\{\xi_i\}_{i=1}^n$ are *i.i.d.* as ξ, which satisfies that,

Assumption 3.2.8. $\mathbb{E}(\xi) = 0$, $\mathbb{E}\left(\xi^2\right) = 1$ and $\xi \le \kappa < \infty$ a.s

The bootstrap test, justified in next Proposition, is $\hat{\Phi}^*_{n\varepsilon}(\alpha) = 1_{\{\hat{\varphi}_{n\varepsilon} > \hat{c}^*_{\varepsilon n}(\alpha)\}}$, where $\hat{c}^*_{\varepsilon n}(\alpha) = \inf\left\{c : \mathbb{P}_\xi\left(\hat{\varphi}^*_{n\varepsilon} \le c\right) \ge 1 - \alpha\right\}$ and $\mathbb{P}_\xi$ is the induced probability of a random variable ξ.

Proposition 3.2.5. Assume that conditions 3.2.1 - 3.2.4 and 3.2.8 hold for $\varepsilon \in \left(0, (n-2K)/2n\right]$, $K < n/2$. Under $\bar{H}_1$,

$$\lim_{n\to\infty} \mathbb{P}_\xi\left(\hat{\varphi}^*_{n\varepsilon} \le c\right) = \mathbb{P}\left(\varphi_{\infty\varepsilon} \le c\right) \text{ a.s.},$$

and under $\bar{H}_1$ there exists a $C > 0$ such that,

$$\lim_{n\to\infty} \mathbb{P}_\xi\left(\hat{\varphi}^*_{n\varepsilon} > C\right) = 1 \ a.s.$$

This implies that the asymptotic power function takes the value α under $\bar{H}_0$ and 1 under $\bar{H}_1$, i.e. $\lim_{n\to\infty}\mathbb{E}\left[\hat{\Phi}^*_{n\varepsilon}(\alpha)\right] = \alpha$ under $\bar{H}_0$ and $\lim_{n\to\infty}\mathbb{E}\left[\hat{\Phi}^*_{n\varepsilon}(\alpha)\right] = 1$ under $\bar{H}_1$. The test can also be based on the bootstrap $p-values$, $\hat{p}^*_\varepsilon = \mathbb{P}_\xi\left(\hat{\varphi}^*_{n\varepsilon} \ge \hat{\varphi}_{n\varepsilon}\right)$, and we reject $\bar{H}_0$ at $\alpha - level$ of significance when $\hat{p}^*_\varepsilon < \alpha$.

Since $\hat{c}^*_{\varepsilon n}(\alpha)$and $\hat{p}^*_\varepsilon$ are difficult to compute in practice, they can be approximated by Monte Carlo as accurately as desired using the following algorithm.

1. Generate b sets of random numbers $\left\{\xi^{(j)}_i\right\}_{i=1}^n$, $j = 1, \ldots, b$ *i.i.d.* as ξ, with b large.

2. Compute b test statistics $\hat{\varphi}^{(b)*}_{n\varepsilon j}$, $j = 1, \ldots, b$, as $\hat{\varphi}^*_{n\varepsilon}$, using the random numbers in 1.

Approximate the bootstrap critical values $\hat{c}^{*}_{\varepsilon n}(\alpha)$ by

$$\hat{c}^{(b)*}_{\varepsilon n}(\alpha) = \inf\left\{ c : \frac{1}{b}\sum_{j=1}^{b} 1_{\left\{\hat{\varphi}^{(b)*}_{n\varepsilon j} < c\right\}} \geq 1-\alpha \right\},$$

and the corresponding $p-values$, $\hat{p}^{*}_{\varepsilon}$, by

$$\hat{p}^{(b)*}_{\varepsilon} = \frac{1}{b}\sum_{j=1}^{b} 1_{\left\{\hat{\varphi}^{(b)*}_{n\varepsilon j} \geq \hat{\varphi}_{n\varepsilon}\right\}}.$$

The greater b, the better the bootstrap critical values and $p-values$ approximation. The same bootstrap approximations can be performed for tests based on test statistics $\tilde{\varphi}^{(j)}_{n}$, $j=0,1,2$.

3.3 Finite Sample Properties

We generate samples $\left\{Y_i, Z_i, X_{11i}, ..., X_{1k_1i}, X_{21i}, X_{2k_2i}\right\}_{i=1}^{n}$ with

$$Y_i = \beta_{00}(Z_i) + \sum_{j=1}^{k_1}\beta_{0j}(Z_i)X_{1ji} + \sum_{j=1}^{k_2}\delta_{0j}X_{2ji} + U_i,\ i=1,...,n, \tag{3.17}$$

with $\{Z_i\}_{i=1}^{n}$ $i.i.d.$ as uniform in $[0,1]$, $X_{\ell ji} = Z_i + e_{\ell ji}$, $e_{\ell ji}$ iid as uniform in $[0,1]$, $\ell = 1,2$, $j = 1,...,k_\ell$, and

$$U_i = \frac{\varepsilon_i \exp(\tau Z_i/2)}{\sqrt{Var(\varepsilon_i \exp(\tau Z_i/2))}},$$

with ε_i iid $N(0,1)$; that is, $Var(U_i)=1$, and τ governs how severe the heteroskedasticity is. We generate the random coefficients as

$$\beta_{0j}(z) = 1 + \lambda\frac{f(z)}{\sqrt{Var(f(z))}},$$

for all $j = 0,1,...,k_1$, i.e. $Var(\beta_{0j}(Z)) = \lambda^2$, i.e. λ governs how serious is the departure from the null under the following models,

$$\begin{array}{ll} \text{a) } f(z) = z, & \text{b) } f(z) = \left[1+\exp(-\rho z)\right]^{-1}, \\ \text{c) } f(z) = \sin(2\pi z), & \text{d) } f(z) = 1 + 2\cdot 1_{\{z\leq 0.4\}}. \end{array}$$

Model a) is a simple linear model and b) is a nonlinear alternative, almost indistinguishable for $\rho = 1$ when $z \in [0,1]$, the lower ρ, the smaller the departure from linearity. We use model b) to check departures form linearity under different values of ρ. Model c) is harder to fit than a) or b) using smooth methods with moderate sample sizes, and d) is a jump model that cannot be estimated using smoothing methods. We only report results for the 0.4 quantile, but we have also tried other values and the results do not change substantially if the jump is not placed in extreme quantiles. Figure 3.1 represents η_0 for the different models and different λ values.

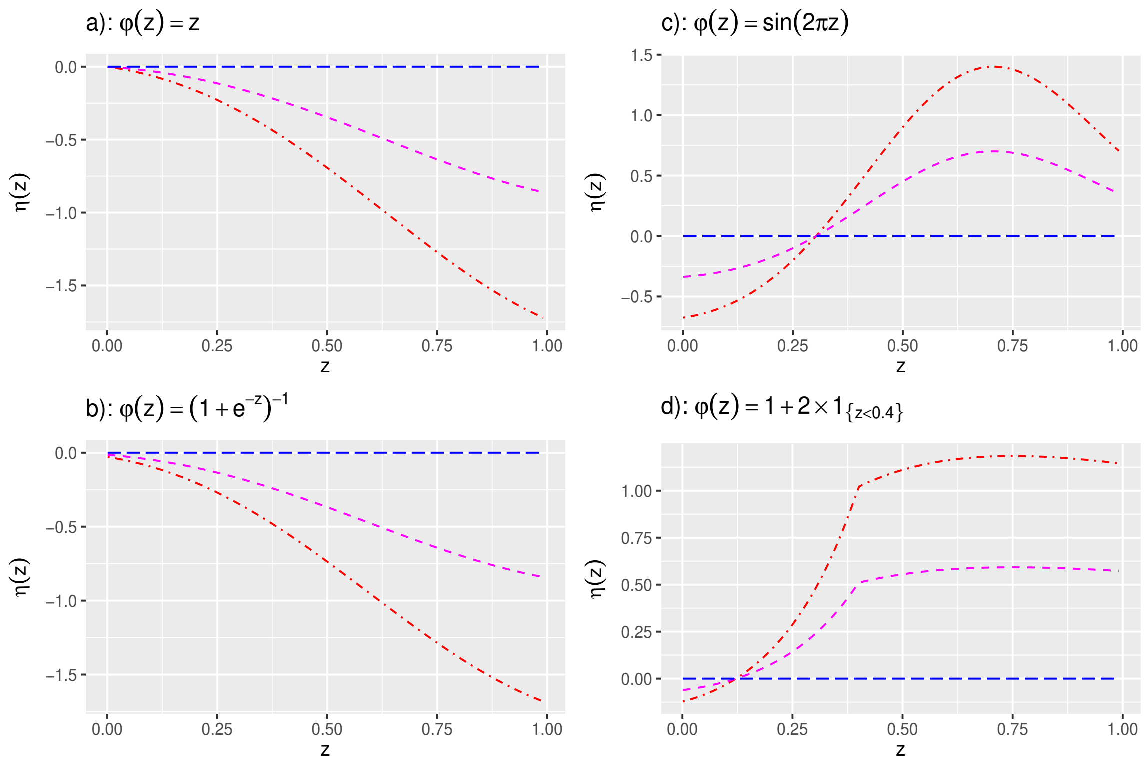

Figure 3.1 Representation of η_0 for different models when $\lambda = 0$ (blue curve), $\lambda = 0.25$ (purple curve), and $\lambda = 0.5$ (red curve).

The simulation study is implemented to provide evidence on the effect of ε's choice on $\hat{\Phi}_{n\varepsilon}(\alpha)$, the accuracy of the bootstrap test, the relative performance of our test with respect to existing alternatives, and the performance of our test for model checking of interactive effects. The Monte Carlo study is based on 1.000 replications and the bootstrap replications are set to 1.000.

Figure 3.2 provides the percentage of rejections for different $\varepsilon's$ for $\alpha = 0.05$. As expected, size accuracy is poor when ε is close to zero. For reasonable ε values, i.e. bigger that 0.1, the level is close to 5%, particularly for the larger sample sizes. On the other hand, under the alternatives, i.e., a), c) and d), the power converges to 1 as n diverges, independently of the value of ε. Of course , the power always increases with λ.

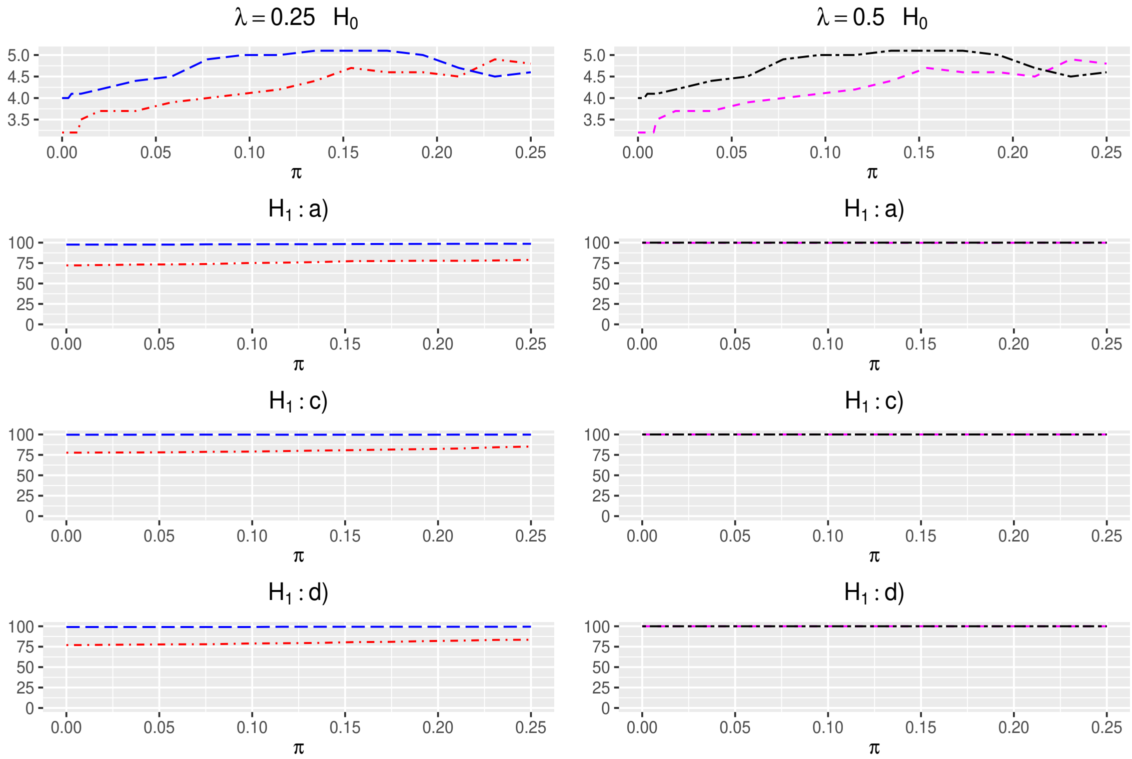

Figure 3.2 Representation of $\hat{\Phi}_{n\varepsilon}(\alpha)$ for different models when $\lambda = 0.25$ (red curve), and $\lambda = 0.5$ (blue curve).

In order to check the level accuracy of the bootstrap test, we compare the percentage of rejections using values of the asymptotic (Proposition 3.2.3) and bootstrap (Proposition 3.2.5) tests when Z is independent of X_1 and U using the test statistics $\tilde{\varphi}_n^{(j)}$, $j = 0, 1, 2$ in a pure varying coefficients model, i.e. with $\delta_0 = 0$. Table 3.1 reports these results. The bootstrap tests exhibit very good size accuracy for the three test statistics. As expected, the asymptotic test based on $\tilde{\varphi}_n^{(0)}$ shows quite poor size properties, particularly for n small. However, the size accuracy of the asymptotic tests based on $\tilde{\varphi}_n^{(1)}$ and $\tilde{\varphi}_n^{(2)}$ is fairly good, but much worse than the corresponding bootstrap tests, as expected.

α	1%				5%				10%			
$n \setminus k_1$	0	1	2	3	0	1	2	3	0	1	2	3
$\tilde{\varphi}_n^{(0)}$ (bootstrap)												
50	0.2	0.3	0.2	0.5	2.5	2.5	2.8	2.4	5.7	6.1	6.4	5.9
100	0.5	0.5	0.7	0.6	3.5	3.2	3.1	2.4	8.5	6.6	6.0	5.8
200	1.2	1.1	0.5	0.4	4.4	4.2	3.6	2.1	8.4	7.9	6.3	4.4
500	0.7	0.7	0.7	1.1	4.1	3.8	3.6	4.5	9.1	7.7	8.4	8.4
$\tilde{\varphi}_n^{(1)}$ (bootstrap)												
50	0.6	0.7	0.1	0.1	4.1	4.1	3.2	3.4	8.7	9.6	8.0	8.3
100	1.2	1.0	0.7	0.5	4.6	4.5	3.9	3.6	9.5	9.4	9.1	8.5
200	1.2	1.2	0.7	0.7	5.3	4.3	5.2	3.9	9.7	10.5	9.2	7.8
500	1.0	0.9	0.6	1.2	4.7	4.1	4.5	5.1	11.1	9.0	8.7	9.5
$\tilde{\varphi}_n^{(2)}$ (bootstrap)												
50	0.7	1.0	0.3	1.3	4.5	5.1	5.2	4.6	9.2	11.6	9.9	11.3
100	1.0	1.0	0.7	0.6	5.1	4.5	4.5	4.4	11.0	9.4	9.2	9.6
200	1.2	1.0	0.7	0.7	5.2	5.1	4.9	2.6	10.3	10.8	9.2	8.4
500	1.2	1.0	1.2	1.5	4.7	5.4	5.7	5.9	9.7	10.0	9.4	10.2
$\tilde{\varphi}_n^{(0)}$ (asymptotic)												
50	0.0	0.1	0.5	3.6	1.7	2.4	6.7	23.3	5.9	8.2	18.1	43.9
100	0.0	0.0	0.1	1.0	1.3	1.1	3.9	9.1	5.3	5.9	10.3	21.8
200	0.0	0.0	0.0	0.4	1.5	1.4	2.5	4.6	4.4	4.6	5.9	13.4
500	0.0	0.0	0.0	0.0	1.5	1.0	2.4	3.3	4.3	3.9	6.2	10.3
$\tilde{\varphi}_n^{(1)}$ (asymptotic)												
50	0.5	0.1	0.0	0.0	2.9	2.7	1.9	1.0	6.8	6.4	5.1	3.3
100	1.0	0.7	0.6	0.1	4.9	3.8	3.7	2.4	10.8	8.5	7.9	6.7
200	1.3	1.2	0.5	0.6	4.6	5.3	4.0	4.1	8.5	10.3	8.8	7.4
500	0.8	1.1	0.8	0.4	4.9	4.5	4.9	4.3	9.5	9.4	9.5	8.7
$\tilde{\varphi}_n^{(2)}$ (asymptotic)												
50	0.2	0.1	0.1	0.0	2.0	1.5	1.2	1.4	4.9	4.0	4.1	3.7
100	0.3	0.2	0.4	0.1	3.3	2.5	2.6	4.6	7.8	5.5	5.0	4.6
200	0.7	0.7	0.4	0.3	4.1	3.5	3.2	1.6	8.2	7.2	6.4	4.1
500	0.7	0.7	0.7	0.8	4.4	3.9	4.0	4.7	8.1	8.3	7.8	8.1

Table 3.1 Percentage of times H_0 was rejected ($k_2 = 0$ and $\tau = 0$)

Now we perform the comparison with existing tests in the context of the partly linear model. We consider the omnibus specification test proposed by Stute (1997) for consistent testing of any nonparametric alternative, which is based on the CUSUM of residuals type process,

$$\hat{\psi}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\hat{U}_i\prod_{j=1}^{k_1}1_{\{X_{1j}\leq x_j\}}\prod_{m=1}^{k_2}1_{\{X_{2m}\leq x_{k_1+m}\}},\ \ x=(x_1,\ldots,x_{k_1+k_2})^{\mathrm{T}}.$$

The CUSUM test is designed for omnibus regression model checking i.e. it detects, in principle, any departure form linearity, including specifications different to the varying coefficient model. We consider the Kolmogorov-Smirnov type statistic,

$$\hat{\phi}_n = \sup_{x \in \mathbb{R}^{k_1+k_2}} \sqrt{n} \left| \hat{\psi}_n(x) \right|.$$

Our test is directional and is expected to be more powerful under H_1. We also consider the LR type bootstrap test of Cai et al. (2000) for testing $\bar{H}_0$ in the direction of $\bar{H}_1$, $\hat{T}_n = (RSS_0/RSS_1) - 1$ that compares restricted and unrestricted sum of squared residuals. LR type tests are asymptotically distribution free by the bandwidth converging to zero at a suitable rate as the sample size diverges (see Fan and Huang (2005) or Cai et al. (2017)). However, tests based on critical values corresponding to the asymptotic distribution exhibit a poor size performance in finite samples. Cai et al. (2017) page 7 lines 15-19 argue that this is because the sensitivity of the test to bandwidth choice and recommend approximating critical values with the assistance of bootstrap. This is why we only report the bootstrap version of Cai et al. (2000)'s test.

Model	H_0			$H_1:a$			$H_1:c$			$H_1:d$		
$n \setminus k_2$	1	2	3	1	2	3	1	2	3	1	2	3
							$\hat{\varphi}_{n0.02}$					
50	3.3	4.4	4.6	11.5	9.5	7.1	13.3	12.5	11.4	15.0	11.2	10.3
100	4.0	5.0	4.6	26.6	15.9	12.4	25.9	23.0	21.2	30.0	20.5	19.8
200	4.5	4.3	3.6	49.1	31.4	22.4	56.0	45.6	40.6	60.9	45.4	38.4
							$\hat{\phi}_n$					
50	4.5	4.4	4.6	12.7	9.4	4.8	14.0	8.1	6.4	14.4	7.9	6.3
100	4.6	5.0	5.4	26.8	10.9	7.8	27.8	16.5	9.9	28.4	14.4	8.5
200	4.4	4.7	4.1	48.1	20.9	11.7	57.0	34.6	18.2	56.9	30.5	15.0
							$\hat{T}_n$					
50	4.7	4.9	6.6	15.8	9.0	7.6	15.0	13.7	7.4	13.2	10.3	9.3
100	3.8	4.0	6.2	32.1	21.6	12.1	31.9	29.0	18.7	29.5	21.4	18.8
200	4.9	5.1	4.2	57.7	40.4	28.3	62.4	55.3	45.5	56.8	41.5	33.6

Table 3.2 Percentage of times H_0 was rejected, 5% of significance ($k_1 = 0$, $\lambda = 0.25$ and $\tau = 1$)

In the following set of simulations we consider different X_2 dimensions, $k_2 = 1,2,3$, $\lambda = 0.25$ and $\tau = 1$. Table 3.2 provides the percentage of rejections in this simulation study. It shows that, under $\bar{H}_1$, our directional test works better than the omnibus CUSUM as k_2 increases because of the curse of dimensionality. For instance, when $k_2 = 3$ and under model d), our test rejects more than twice than the CUSUM test. The smoothing based test has similar power than ours in all models but the jump model d), due to the poor performance of the Nadaraya-Watson estimator for estimating discontinuous regressions.

Table 3.3 reports the percentage of rejections for different X_1 dimensions, $k_1 = 1,2,3$, $\lambda = 0.25$ and $\tau = 1$. Note that, again, our directional test works better than the omnibus CUSUM as k_1 increases. For instance, when $k_1 = 3$ and under model d), the power of our test is almost twice the CUSUM test. The test using $\hat{T}_n$ works similarly to ours in general, but our test performs better when $k_1 = 3$. The smooth test also suffers of the curse of dimensionality; the power decreases as k_1 increases. Also, the LM test detects departures from the null in the direction of jump model d) much less than the other tests, which do not require to estimate the model under the alternative using smoothing.

Under model d) our test also works much better than the LM smoothing based test because of the curse of dimensionality of the Nadaraya-Watson estimator needed to compute $\hat{T}_n$.

Model	H_0			$H_1 : a$			$H_1 : c$			$H_1 : d$		
$n \setminus k_1$	1	2	3	1	2	3	1	2	3	1	2	3
						$\hat{\varphi}_{n0.02}$						
50	2.7	3.4	2.3	18.3	20.5	20.5	26.8	41.6	54.1	25.2	30.6	34.1
100	3.8	4.1	3.1	47.2	59.2	69.7	66.9	92.7	98.5	63.6	85.9	94.6
200	3.9	3.2	4.0	84.1	96.3	98.9	97.2	100	100	97.1	100	100
						$\hat{\phi}_n$						
50	4.4	4.6	5.2	21.3	17.7	16.4	22.9	23.4	22.9	18.8	18.4	16.1
100	5.0	5.4	4.3	41.6	40.5	39.6	55.4	61.6	56.7	45.8	42.3	35.8
200	4.7	4.1	5.9	76.3	83.2	81.4	93.8	96.2	94.7	86.2	84.2	76.7
						$\hat{T}_n$						
50	4.5	4.8	5.7	18.2	20.2	22.7	22.0	48.4	27.2	15.8	42.7	19.8
100	4.2	4.9	4.7	44.8	55.3	36.5	67.0	61.5	42.8	48.8	54.8	39.6
200	4.9	4.8	4.5	71.1	94.0	53.5	97.2	97.7	53.6	89.0	89.8	52.2

Table 3.3 Percentage of times H_0 was rejected, 5% of significance ($k_1 = 1$, $\lambda = 0.25$ and $\tau = 1$)

In the next set of simulations we apply the test as a regression model check of the linearity hypothesis when $k_1 = 0, k_2 = 1$ and $X_2 = Z$. That is, $\bar{H}_0$ is equivalent to omnibus specification testing of the simple regression model $\mathbb{E}(Y|Z) = \bar{\beta}_{00} + Z\delta_{00}$ $a.s.$ The resulting test competes with the CUSUM test based on $\hat{\phi}_n$. Since β_{00} is not identifiable, tests based on comparing fits under the null and the alternative, like the LR test using $\hat{T}_n$ as test statistic, cannot be implemented. We compare our test with the omnibus specification test, designed to detect more general non-linear alternatives. We consider model $b)$ with different ρ values in order to check the performance of the test under small departures from the linearity hypothesis. Table 3.4 shows that our test rejects almost double than the CUSUM test for all ρ values.

λ	0.25						0.5					
$n\setminus\rho$	1	2	3	4	5	15	1	2	3	4	5	15
						$\hat{\varphi}_{n0.02}$						
50	3.8	4.4	5.3	6.6	7.5	11.8	4.1	5.4	8.3	11.6	16.6	35.5
100	4.0	4.7	6.1	7.8	10.6	25.2	3.9	6.2	11.7	21.5	33.6	76.2
200	3.9	4.3	6.4	11.2	18.3	56.3	4.1	6.5	19.0	39.7	61.5	98.7
						$\hat{\phi}_n$						
50	5.6	5.4	5.8	6.0	6.5	7.7	5.6	5.7	6.6	8.6	10.9	19.2
100	4.9	5.6	6.7	7.7	9.0	13.7	5.2	7.2	10.0	14.4	20.8	49.7
200	4.3	4.7	6.7	8.6	12.0	24.8	4.6	6.8	13.9	25.4	40.0	87.1

Table 3.4 Percentage of times H_0 was rejected, 5% of significance ($k_1 = 0$, $k_2 = 1$ and $\tau = 1$)

We also consider the test for model checking of non-linear regression models. We consider testing that $\mathbb{E}(Y|Z) = \bar{\beta}_{00} + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0\ell-1}$ *a.s.* in the direction

$$\mathbb{E}(Y|Z) = \beta_{00}(Z) + \sum_{\ell=1}^{L} Z^{\ell}\delta_{0\ell-1} a.s. \text{ with } Var(\beta_{00}(Z)) \geq 0 \ a.s.$$

and β_{00} unknown. Our test is omnibus for the nonlinear specification hypothesis, since the direction of interest nests any possible departure from the null. This corresponds to applying our test to model (3.5) with $g_j(z) = z^j$, $j = 1, ..., L$. Table 3.5 reports rejections for model b) with $\rho = 15$, which produces a sensitive departure from linearity, for different L values

λ	0.25				0.5			
$n\setminus L$	1	2	3	4	1	2	3	4
				$\hat{\varphi}_{n0.02}$				
50	11.8	7.2	4.4	2.9	35.5	16.4	6.1	2.9
100	25.2	11.8	6.3	4.7	76.2	38.4	11.7	4.9
200	56.3	24.9	7.3	3.9	98.7	77.9	19.6	6.2
				$\hat{\phi}_n$				
50	7.7	5.3	6.2	6.1	19.2	8.2	6.0	5.9
100	13.7	6.0	5.6	6.2	49.7	10.5	5.7	6.1
200	24.8	6.3	4.3	4.2	87.1	18.7	5.4	4.6

Table 3.5 Percentage of times H_0 was rejected, 5% of significance ($k_1 = 0$, $k_2 = 1$, $\rho = 15$ and $\tau = 1$)

Next, we consider the performance of the test as a specification test of interactive effects in the context of model (3.5) with $k_1 > 0$, $L = 1$, $g_0(z) = 1$ and $g_1(z) = z$. That is, our test is implemented for testing the hypothesis

$$\mathbb{E}(Y|X_1,Z) = \beta_{00}(Z) + \delta_{00}Z + \sum_{j=1}^{k_1} \left(\beta_{0j}(Z)X_{1j} + \delta_{0j}X_{1j}Z\right) \ a.s.$$

in the direction

$$\mathbb{E}(Y|X_1,Z) = \bar{\beta}_{00} + \delta_{00}Z + \sum_{j=1}^{k_1} \left(\bar{\beta}_{0j}X_{1j} + \delta_{0j}X_{1j}Z\right) \ a.s..$$

Table 3.6 reports the percentage of rejections for ours and CUSUM test in model b) with $\lambda = 0.5$, different ρ values and $k_1 = 1,2,3$. Our test performs better than CUSUM in most cases.

ρ	1			2			3			15		
$n \setminus k_1$	1	2	3	1	2	3	1	2	3	1	2	3
						$\hat{\varphi}_{n0.02}$						
50	3.6	3.3	2.8	4.2	3.7	3.8	4.8	4.4	5.4	8.5	13.8	16.7
100	3.7	5.4	3.0	4.8	5.8	5.4	6.1	8.8	11.6	19.8	38.8	51.6
200	3.8	3.8	4.9	4.9	6.7	10.4	8.1	16.6	26.7	48.3	84.1	94.0
						$\hat{\phi}_n$						
50	4.5	6.1	7.0	4.6	6.5	6.6	4.9	7.0	7.8	7.8	9.4	10.3
100	5.3	7.1	4.8	6.4	6.9	5.6	7.1	8.7	7.3	12.1	15.5	11.2
200	4.5	5.9	5.1	4.9	7.1	6.8	8.6	9.0	10.6	21.2	34.1	27.5

Table 3.6 Percentage of times H_0 was rejected, 5% of significance ($k_2 = 0$, $\lambda = 0.5$ and $\tau = 1$)

Now, we consider testing non-linear specification of interactive effects in the context of model (3.5) with $k_1 > 0$, $L = 1,2,3,4$, $g_0(z) = 1$ and $g_j(z) = z^j$. Our test is implemented for testing the hypothesis

$$\mathbb{E}(Y|X_1,Z) = \beta_{00}(Z) + \sum_{\ell=1}^{L} Z^\ell \delta_{0\ell-1} + \sum_{j=1}^{k_1} \left(\beta_{0j}(Z)X_{1j} + X_{1j}\sum_{\ell=1}^{L} Z^\ell \delta_{0j+L+\ell-1}\right) a.s$$

in the direction

$$\mathbb{E}(Y|X_1,Z) = \bar{\beta}_{00} + \sum_{\ell=1}^{L} Z^\ell \delta_{0\ell-1} + \sum_{j=1}^{k_1} \left(\bar{\beta}_{0j}X_{1j} + X_{1j}\sum_{\ell=1}^{L} Z^\ell \delta_{0j+L+\ell-1}\right) a.s.$$

Table 3.7 reports the percentage of rejections for both ours and CUSUM tests under model b) with $\lambda = 0.5$, $\rho = 15$, $k_1 = 2$ and different L values. Our test performs better in general.

$n \backslash L$	1	2	3
	$\hat{\varphi}_{n0.02}$		
50	13.8	7.0	4.6
100	38.8	17.3	6.9
200	84.1	47.4	10.4
	$\hat{\phi}_n$		
50	9.4	7.7	8.7
100	15.5	8.5	6.3
200	34.1	14.5	7.5

Table 3.7 Percentage of times H_0 was rejected, 5% of significance ($k_1 = 2$, $k_2 = 0$, $\lambda = 0.5$ $\rho = 15$ and $\tau = 1$)

3.4 An Application to Modeling Education Returns

We complement the previous Monte Carlo study with an application to using IQ as control, or proxy, variable of "ability" in a returns of education model. This is based on Blackburn and Neumark (1995) work, which is used in Wooldridge (2009) textbook (example 9.3). The data consists of 663 observations from the Young Men's Cohort National Longitudinal Survey. The main objective consists of estimating the marginal effect of education on wages, controlling for relevant covariates, which include unobserved "ability". A reasonable model using IQ as proxy variable (Wooldridge (2009), example 9.3) is

$$Log\,(WAGE) = \bar{\beta}_{00} + \bar{\beta}_{01} \cdot EDUC + \bar{\beta}_{02} \cdot IQ + X_2^{\mathrm{T}} \delta_{01} + U, \tag{3.18}$$

where $WAGE$ are USD monthly earnings, $EDUC$ is years of education, IQ is intelligence quotient (proxy of ability), and $X_2^{\mathrm{T}} = (EXPER, TENURE, MARRIED, SOUTH, URBAN, BLACK)^{\mathrm{T}}$, $EXPER$ are years of work experience, $TENURE$ years with current employer, $MARRIED$ a dummy (1 if married), $BLACK$ dummy (1 if black), $SOUTH$ dummy (1 if live in south), $URBAN$ dummy (1 if live in urban area SMSA), and $\delta_{01} = (\delta_{01}, ..., \delta_{06})^{\mathrm{T}}$. The OLS estimators of $\bar{\beta}_{01}$ and $\bar{\beta}_{02}$ in this model (heteroskedasticity robust SE in parenthesis) are 0.054 (0.006) and 0.0036 (0.001), respectively. The OLS estimator of the marginal effect of $EDUC$ ($\bar{\beta}_{01}$) is inconsistent when $\mathbb{E}\,(U|EDUC, IQ, X_2)$ depends on $EDUC$, i.e. IQ is not a good proxy for ability, but also when it only depends on IQ in a nonlinear form. A reasonable alternative to (3.18) is the varying coefficients model

$$Log\,(WAGE) = \beta_{00}(IQ) + \beta_{01}(IQ) \cdot EDUC + X_2^{\mathrm{T}}\delta_0 + U, \tag{3.19}$$

which allows *EDUC* partial effects to be an unknown function of IQ. Figure 3.3 provides estimates of β_{00} and β_{01} varying coefficients using Cai et al. (2000) procedure, which uses a modified manifold cross-validation criterion for choosing the bandwidth. We also provide OLS estimates of the parametric specification $\beta_{0j}(IQ) = \bar{\beta}_{0j}^{(1)} + \bar{\beta}_{0j}^{(2)} IQ + \bar{\beta}_{0j}^{(3)} IQ^2$, $j = 0, 1$.

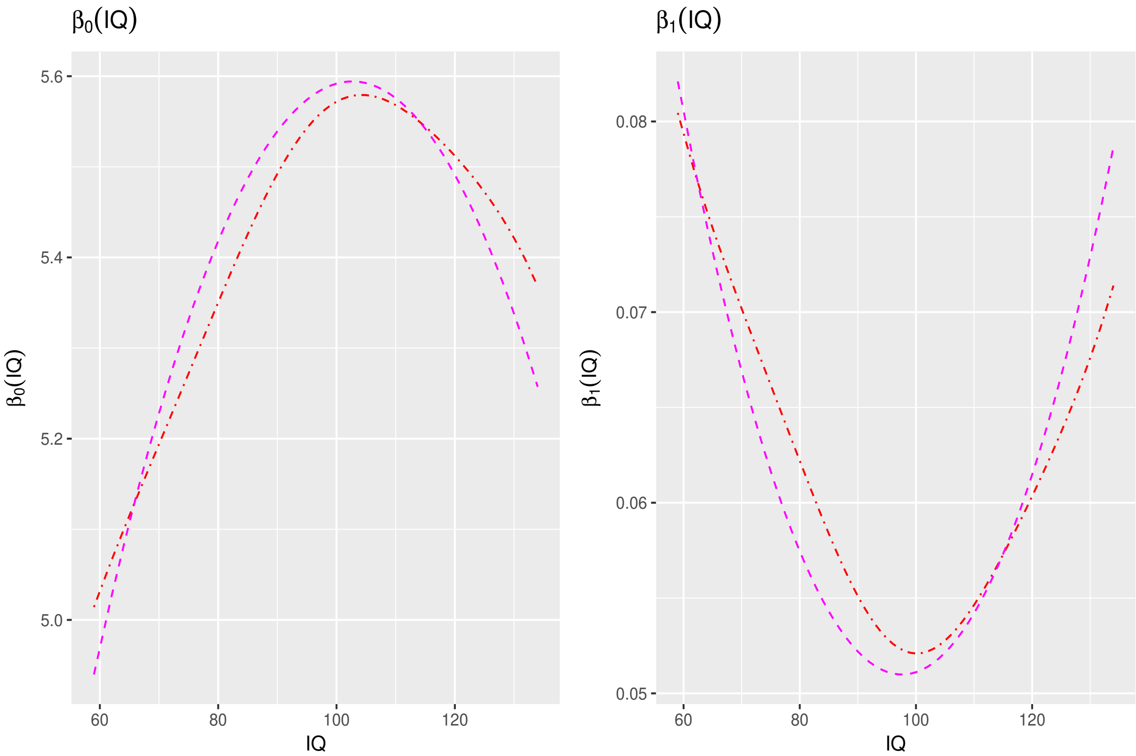

Figure 3.3 Representation of $\beta_{00}(IQ)$ and $\beta_{01}(IQ)$ for the estimates of the varying coefficients using kernels with a plug-in bandwidth (red curve), and OLS estimates of the parametrization (purple curve).

The $p-values$ for testing $H_0 : Var(\beta_{0j}(IQ)) = 0,\ j = 1,2$ versus $H_1 : Var(\beta_{0j}(IQ)) > 0$ some $j = 1,2$, or $H_2 : Var(\beta_{00}(IQ)) = 0$ and $Var(\beta_{01}(IQ)) > 0$ are reported in Table 3.8, where we provide the $p-values$.

We also report the smoothing *LR* test of Cai et al. (2000). Here the CUSUM test is unable to reject the null hypothesis, but the directional tests reject H_0 in the two directions considered. The $p-value$ of our test is the smallest when testing in the direction H_1, but

Test	$H_1 : Var(\beta_{00}(IQ)) > 0$ or $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) = 0$ and $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) > 0$ and $Var(\beta_{01}(IQ)) = 0$
$\hat{\varphi}_{n0.003}$	0.012	0.017	0.08
$\hat{\phi}_n$		0.734	
$\hat{T}_n$	0.041	0.009	0.009

Table 3.8 p-value of testing H_0 versus H_1 and H_2

the corresponding $p-value$ for the smoothing LR test based on $\hat{T}_n$ is the smallest in the direction H_2.

Next, we apply our test as a model check of the interactive effect of $EDUC$. The maintained specification is

$$Log\,(WAGE) = (\beta_{00}(IQ) + \delta_{07}IQ) + (\beta_{01}(IQ) + \delta_{08}) \cdot EDUC + X_2^{\mathrm{T}}\delta_0 + U, \tag{3.20}$$

which is model (3.19) augmented with the explanatory variables $(IQ, EDUC)$ in the constant coefficients terms, *i.e.* X_2^{T} in (3.19) is substituted by $(X_2^{\mathrm{T}}, IQ, EDUC)$ in (3.20). Then H_0 is in fact a specification test of the functional form of the varying coefficients in (3.19).

Test	$H_1 : Var(\beta_{00}(IQ)) > 0$ or $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) = 0$ and $Var(\beta_{01}(IQ)) > 0$	$H_2 : Var(\beta_{00}(IQ)) > 0$ and $Var(\beta_{01}(IQ)) = 0$
$\hat{\varphi}_{n0.003}$	0.6489	0.405	0.484
$\hat{\phi}_n$	0.491	0.653	0.543

Table 3.9 p-value of testing H_0 versus H_1 and H_2

In this case, see table 3.9, we are unable to reject the specification of the interactive effect either with the CUSUM or with our test. We conclude that the specification including IQ and a simple interactive effect $EDUC$ with IQ cannot be rejected.

3.5 Conclusions

We have proposed a test for constancy of coefficients in semi-varying coefficients models, where the variable responsible for the coefficient varying may depend on the rest of explanatory variables in an unknown form. The test, implemented using bootstrap, is based on comparing the OLS coefficients of subsamples of concomitants to the explanatory variable in the varying coefficients. The test is justified under fairly weak regularity conditions, which allow discontinuous random coefficients under the alternative hypothesis. Our test forms a basis for specification testing of parametric varying coefficients and, in particular, for testing the functional form of interactive effects. Simulation results have provided evidence of the good performance of our test in finite samples compared with a CUSUM-type test, designed to omnibus specification testing of linear regression models, and a smooth LR test, designed to test varying coefficients constancy in the direction of smooth alternatives. The CUSUM test, like ours, does not require estimating the model on the alternative, but the LR-type test compares the restricted and unrestricted sum of squared residuals and, hence, requires estimating the nonparametric smooth varying coefficients. Simulations show that, unlike our test, the two competitors suffer of the curse of dimensionality. These also show that the LR smooth test exhibit a lack of power, compared with the two competitors, under alternatives with discontinuous varying coefficients. We have also included a real data application to model interactive effects of IQ in a returns of education model.

The proposed methodology is applicable to testing constancy of a subset of varying coefficients or a linear combination of them. However, since the model under the null must be estimated, smooth estimation of the unrestricted varying coefficients is necessary. A formal justification of the resulting test is technically demanding, but it seems possible to take advantage of existing asymptotic inference results for varying coefficient models.

Chapter 4

Testing beta constancy in capital asset pricing models

4.1 Introduction

Capital asset pricing models (CAPM) are used to reveal how portfolio returns are determined and which factors affect returns. For these type of models, the error-in-variables problem is well known and translates in dependence of the error term with the explanatory variables, which makes the estimator to be inconsistent. In order to overcome this problem, the instrumental variable method is proposed in works such as Amano et al. (2012), Dumas (1994), Jegadeesh et al. (2019), and Roy and Shijin (2018) among others. Instrumental variables (IV) models have attracted the attention of researches in empirical studies due to the possibility to correct potential endogeneity between the regressors and the structural errors (see, e.g., Wooldridge (2010) chapters 5, 8 and 14).

It is common in these type of modes to write the expected returns as a linear function of one or more beta coefficients that measure the asset's systematic risk; however, this linear relationship assumption in asset pricing models have been proven wrong by several studies based on empirical evidence of time variation in betas and expected returns (see, Bansal et al. (1993), Bansal and Viswanathan (1993), Cochrane (1996), Jagannathan and Wang (1996, 2002), Reyes (1999), Ferson and Harvey (1991, 1993, 1997, 1999), Cho and Engle (1999), Wang (2002, 2003), Akdeniz et al. (2003), Ang and Liu (2004), Fraser et al. (2004), Gagliardini et al. (2011), among others). In this context, authors such as Bansal et al. (1993) and Bansal and Viswanathan (1993) assume a nonlinear function and Dittmar (2002), Dumas and Solnik (1995) and Cochrane (1996) assume that the parameters are linear function of some instrumental variables. In these situations, nonparametric and semiparametric models have gained importance as little or no restrictive prior information of the functional form is needed, e.g., Wang (2002, 2003) used a Nadaraya-Watson kernel regression, Gourieroux and Monfort (2007) considered a class of nonlinear parametric and semiparametric models, Cai et al. (2015) proposed an estimation method in the spirit of local generalized estimating equations and Escanciano et al. (2015) used nonparametric estimation in consumption based asset pricing Euler equations.

The combination of both problems, endogeneity and no prior assumptions of the functional form, result in the need of varying coefficient models that take into account the endogeneity of the regressors; here, it is of great relevance the studies of Cai et al. (2006), Cai and Li (2008), Escanciano et al. (2015) and Cai et al. (2017). Their work add to the vast amount of literature on nonparametric estimation of instrumental variable models (see, Blundell and Powell (2003), Newey and Powell (2003), Florens (2005) Hall and Horowitz (2005), Blundell et al. (2007), Horowitz and Lee (2007), Darolles et al. (2011), Florens et al. (2011, 2012), Florens and Simoni (2012), among others). In this context, we propose a nonparametric estimation procedure for varying coefficient models with endogenous regressors.

Conditional asset pricing literature provides a framework in which returns and pricing factors are predictable, in the sense of a significant time variation in their joint conditional distribution; besides, these models do not provide much information the functional form of the conditional moments. There exist a wide literature that employs parametric techniques to evaluate conditional asset pricing models; for instance, Jagannathan and Wang (1996) find that a conditional CAPM can explain the cross section of stock of returns, while the static CAPM model cannot, also Lettau and Ludvigson (2001) show that the value premium can be explained by a conditional CAPM with time varying price of risk. Nevertheless, other authors such as Lewellen and Nagel (2006) and Nagel and Singleton (2011) suggest that this superior performance of the conditional CAPM is an illusion caused by the low statistical power of standard CAPM.

Following these ideas, there exist significant contributions to avoid specifying the conditional distribution of returns and factors by using nonparametric techniques. Here, Nagel and Singleton (2011) estimate nonparametrically first and second conditional moments and then work with parametric CAPM. In contrast, Wang (2003), Orbe et al. (2008), Roussanov (2014) and Peñaranda et al. (2018) consider varying coefficient CAPM; Wang (2003) uses the stochastic discount factor to estimate pricing errors nonparametrically and test if pricing errors are independent of the conditioning variables, unfortunately this test has zero power when there exist a nonlinear dependence between pricing errors and the conditioning variables. Orbe et al. (2008) estimate consistently the time varying parameters of a general conditional beta pricing model using a nonparametric version of the two-pass approach (see, Black et al. (1972), Fama and MacBeth (1973), Shanken (1985, 1992)); they also develop a test to test for invariance of the prices of the risk factors through time or to test whether or not the risk premium can be considered significantly non-zero. Roussanov (2014) estimates nonparametric betas with a focus on consumption based models but he does not develop a formal test; models are evaluated by means of tests of a particular zero average pricing error. Finally, Peñaranda et al. (2018) consider nonparametric estimation and testing of conditional asset pricing models, they present and adaptive omnibus specification test that is robust to functional form misspecification of both conditional moments and prices of risk.

In this context, our proposal is to build up a test to detect constancy of the price of risk or beta constancy, that is, unconditional asset pricing models. The proposed test is based in works such as Kauermann and Tutz (1999), Cai et al. (2000), Fan and Zhang (2000), Fan et al. (2001), Fan and Huang (2005), Qu and Li (2006) and Zhou and Liang (2009) where they proposed a test based on the discrepancy between restricted and unrestricted sum of squared residuals for varying coefficient models without endogenous regressors; as for the case of varying coefficient models with endogenous regressors although under a completely different setting, the reader may refer to Cai et al. (2017). In all the previous works, the test is suppose to detect linearity or significance of the parameters using parametric and nonparametric estimates to build up the test. Although there exists a great literature on detecting constancy in varying coefficient models, little has been applied to CAPM setting; the reader may refer to Wang (2003), Orbe et al. (2008), Roussanov (2014) and Peñaranda et al. (2018) to find inference applied to CAPM models.

The rest of the chapter is organized as follows. Section 4.2 describes our varying coefficient model with endogenous regressors, its estimation procedure and we also present the asymptotic results. Section 4.3 describes the testing procedure and gives the outline of how to obtain the bootstrap p-vales. Section 4.4 presents numerical results based on simulations Finally, section 4.5 concludes the paper. All the proofs are contained in the Appendix.

4.2 Econometric model and estimation procedure

4.2.1 Model

Consider the varying coefficient model with endogenous regressors of the form

$$\begin{aligned} Y_t &= g(X_t, Z_{1t}) + u_t \\ &= X_t^\top \beta(Z_{1t}) + u_t, \quad t = 1, \dots, T, \end{aligned} \tag{4.1}$$

where Y_t is a $m \times 1$ vector of responses, X_t is a $d \times m$ matrix of endogenous explanatory variables, $\beta(\cdot)$ is a $d \times 1$ vector of unknown functions that needs to be estimated, u_t is a $m \times 1$ vector of random errors and $\mathscr{Z}_t$ is a $p+q$ dimensional vector containing a q dimensional vector Z_{1t} of exogenous variables and a p dimensional vector Z_{2t} of instrumental variables. In CAPM models Y_t are the returns of m assets, X_t are the d factors that affect the returns of the m assets and $\beta(Z_{1t})$ are the prices of risk.

Note that this nonparametric model is different from the standard nonparametric model because $\mathrm{E}[u_t | X_t, Z_{1t}] \neq 0$; therefore to estimate $\beta(\cdot)$ we use the assumption

$$\mathrm{E}[u_t | \mathscr{Z}_t] = 0, \tag{4.2}$$

then taking conditional expectations we have

$$\mathrm{E}\left[Y_t|\,\mathscr{Z}_t\right] = \mathrm{E}\left[X_t|\,\mathscr{Z}_t\right]^\top \beta(Z_{1t}) = \pi(\mathscr{Z}_t)^\top \beta(Z_{1t}),$$

where $\pi(\mathscr{Z}_t) = \mathrm{E}\left[X_t|\,\mathscr{Z}_t\right]$ is an unknown $d \times m$ matrix to be estimated. Because $\pi(\mathscr{Z}_t)$ are unknown, we will need a preliminary step to estimate $\beta(Z_{1t})$. Note that, we need to impose $p \geq d$, that is, the number of instruments is larger than the number of endogenous variables; which is a usual identification condition in instrumental variables problems.

4.2.2 Estimation procedure

As $(Y_t, X_t, \mathscr{Z}_t)$ are the only observed data our suggested procedure is a three-stage approach. The first stage is in charge of estimating the conditional expectation $\pi(z) = \mathrm{E}\left[X_t|\,\mathscr{Z}_t = z\right]$ by a regression of X_t on $\mathscr{Z}_t$, the second stage involves estimation of optimal weighting matrix and finally the third stage proceeds by estimating $\beta(\cdot)$. In all three stages, we consider local linear fitting techniques mainly because of its high statical efficiency (see, Fan and Gijbels (1995) and Kniesner and Li (2002) for discussion on the local linear fitting technique).

We begin with the first stage, where we obtain, $\hat{\pi}(\mathscr{Z}_t)$, the fitted value for $\pi(\mathscr{Z}_t)$. Note that we have to estimate $\pi(\mathscr{Z}_t)$ which is $d \times m$ matrix, so in order to simplify the estimation process we will use the *vec* operator; thus, we now need to estimate a $dm \times 1$ vector, $vec\left(\pi(\mathscr{Z}_t)\right) = \left(\pi_1(\mathscr{Z}_t), \ldots, \pi_{dm}(\mathscr{Z}_t)\right)^\top$. Now, assuming that $vec\left(\pi(\mathscr{Z}_t)\right)$ has continuous second derivatives, for z in the neighborhood of $\mathscr{Z}_t$, a Taylor expansion approximates $vec\left(\pi(\mathscr{Z}_t)\right)$ by $\gamma(z) + I_{dm}^\top \otimes (\mathscr{Z}_t - z)^\top vec\left(D_\gamma(z)\right)$, that is,

$$vec\left(\pi(\mathscr{Z}_t)\right) \approx \gamma(z) + I_{dm}^\top \otimes (\mathscr{Z}_t - z)^\top vec\left(D_\gamma(z)\right),$$

where $\otimes$ is the Kronecker product, I_{dm} is $dm \times dm$ identity matrix and $D_\gamma(z)$ is a $dm \times (p+q)$ matrix of partial derivatives of the $dm \times 1$ function $\gamma(z)$ with respect to the elements of the $(p+q) \times 1$ vector z, e.g., $D_\gamma(z) = \partial\gamma(z)/\partial z$. Then, we derive the local linear estimator as the minimizer of

$$\frac{1}{T}\sum_{t=1}^{T}\left[vec\left(X_t\right) - \tilde{Z}_t^{*\top} g(z)\right]^\top \left[vec\left(X_t\right) - \tilde{Z}_t^{*\top} g(z)\right] L_{H_1}(\mathscr{Z}_t - z), \tag{4.3}$$

where $\tilde{Z}_t^* = \left(I_{dm}, I_{dm} \otimes (\mathscr{Z}_t - z)^\top\right)^\top$ is a $dm(p+q+1) \times dm$ matrix of regressors, $g(z) = \left(\gamma(z)^\top, vec\left(D_\gamma(z)\right)^\top\right)^\top$ is a $dm(p+q+1) \times 1$ vector of unknown functions, $L_{H_1}(\cdot) = |H_1|^{-1/2} L(\cdot H_1^{-1/2})$, H_1 is a $(p+q) \times (p+q)$ bandwidth matrix and $L_{H_1}(\cdot)$ is a kernel function;

then, if $\frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_t^*\tilde{Z}_t^{*\top}L_{H_1}(\mathscr{Z}_t - z)$ is invertible, the local linear estimator of $vec(\pi(\mathscr{Z}_t))$, denoted by $vec(\hat{\pi}(z)) = \hat{\gamma}(z) = e_1^\top \hat{g}(z)$ is

$$vec(\hat{\pi}(z)) = e_1^\top \left[\frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_t^*\tilde{Z}_t^{*\top}L_{H_1}(\mathscr{Z}_t - z)\right]^{-1} \frac{1}{T}\sum_{t=1}^{T}\tilde{Z}_t^* vec(X_t)L_{H_1}(\mathscr{Z}_t - z), \tag{4.4}$$

where $e_1 = \left(I_{dm} \vdots 0_{dm(p+q)\times dm}^\top\right)^\top$ is a $dm(p+q+1)\times dm$ selection matrix and $0_{dm(p+q)\times dm}$ is a $dm(p+q)\times dm$ matrix of zeros.

Once we have obtained the estimator of the conditional expectation, we can proceed with the estimation of $\beta(\cdot)$. To do so, as it is usual in IV estimation procedures, we first obtain the estimator of the optimal weighting matrix and then we estimate $\beta(\cdot)$. Therefore, our second stage proceeds by estimating the optimal weighting matrix; to this end, we assume that $\beta(Z_{1t})$ has continuous second order derivatives at any given point z_1, and by a Taylor expansion for Z_{1t} in a neighborhood of z_1, we have

$$\hat{\pi}(\mathscr{Z}_t)^\top\beta(Z_{1t}) \approx \hat{\pi}(\mathscr{Z}_t)^\top\beta(z_1) + \hat{\pi}(\mathscr{Z}_t)^\top \otimes (Z_{1t} - z_1)^\top vec\left(D_\beta(z_1)\right),$$

where $D_\beta(z_1)$ is a $d\times q$ matrix of partial derivatives of the $d\times 1$ function $\beta(z_1)$ with respect to the elements of the $q\times 1$ vector z_1. Then, the local linear estimator comes from the following first order condition

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\left(Y_t - \tilde{\Pi}_t(\mathscr{Z}_t)^{*\top}\eta(z_1)\right)K_{H_2}(Z_{1t} - z_1) = 0, \tag{4.5}$$

where $K_{H_2}(\cdot) = |H_2|^{-1/2}K(\cdot H_2^{-1/2})$, H_2 is a $q\times q$ bandwidth matrix, $K(\cdot)$ is a kernel function, $\tilde{\Pi}_t(\mathscr{Z}_t) = \left(\hat{\pi}(\mathscr{Z}_t)^\top, \hat{\pi}(\mathscr{Z}_t)^\top \otimes (Z_{1t} - z_1)^\top\right)^\top$ is a $d(q+1)\times m$ matrix containing the estimates of the first stage, and $\eta(z_1) = \left(\beta(z_1)^\top, vec\left(D_\beta(z_1)\right)^\top\right)^\top$ is a $d(q+1)\times 1$ vector of unknown functions; thus, if we assume that $\frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\tilde{\Pi}_t(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t} - z_1)$ is nonsingular, the second stage local linear estimator of $\beta(z_1)$, denoted by $\bar{\beta}(z_1) = e_2^\top\bar{\eta}(z_1)$ is

$$\bar{\beta}(z_1) = e_2^\top \left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\tilde{\Pi}_t(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t} - z_1)\right]^{-1} \frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)Y_t K_{H_2}(Z_{1t} - z_1), \tag{4.6}$$

where $e_2 = \left(I_d \vdots 0_{dq\times d}^\top\right)^\top$ is a $d(q+1)\times d$ selection matrix, I_d is a $d\times d$ identity matrix and $0_{dq\times d}$ is a $dq\times d$ matrix of zeros. Now, using the second stage local linear estimator of

$\beta(z_1)$, we define the estimator fir the optimal weighting matrix, $\hat{\Omega}(z)$, as follows

$$\hat{\Omega}(z) = \frac{1}{\hat{f}(z)T}\sum_{t=1}^{T} \bar{u}_t\bar{u}_t^\top L_{H_1}(\mathscr{Z}_t - z), \tag{4.7}$$

where $\bar{u}_t = Y_t - X_t^\top\bar{\beta}(Z_{1t})$ and $\hat{f}(z) = 1/T\sum_{t=1}^{T} L_{H_1}(\mathscr{Z}_t - z)$. Note that for the next stage we need for $\hat{\Omega}(z)$ to be positive definite.

Finally, in the third stage we derive the local linear estimator of $\beta(\cdot)$, using the estimates of $\pi(\mathscr{Z}_t)$ and $\Omega(\mathscr{Z}_t)$, as the minimizer of the sum of weighted least squares

$$\frac{1}{T}\sum_{t=1}^{T}\left[Y_t - \tilde{\Pi}_t(\mathscr{Z}_t)^\top\eta(z_1)\right]^\top \hat{\Omega}(\mathscr{Z}_t)^{-1}\left[Y_t - \tilde{\Pi}_t(\mathscr{Z}_t)^\top\eta(z_1)\right]K_{H_2}(Z_{1t} - z_1); \tag{4.8}$$

then, if we assume that $\frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\hat{\Omega}(\mathscr{Z}_t)^{-1}\tilde{\Pi}_t(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t} - z)$ is invertible, the third stage local linear estimator of $\beta(z_1)$, denoted by $\hat{\beta}(z_1) = e_2^\top\hat{\eta}(z_1)$ is

$$\hat{\beta}(z_1) = e_2^\top\left[\frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\hat{\Omega}(\mathscr{Z}_t)^{-1}\tilde{\Pi}_t(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t} - z)\right]^{-1}\frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\hat{\Omega}(\mathscr{Z}_t)^{-1}Y_tK_{H_2}(Z_{1t} - z). \tag{4.9}$$

Before we continue with the large sample properties of our estimators, we will discuss a particular form of model (4.1). Note that model (4.1) encompasses models such as the partially linear model (see, Robinson (1988)) and the estimators proposed above apply with very little modifications. For instance, if we are willing to assume that $Z_{1t} = \left(Z_{11,t}^\top, Z_{12,t}^\top\right)^\top$, then

$$\begin{aligned} Y_t &= g(X_t, Z_{1t}) + u_t, \\ &= Z_{11,t}^\top\gamma_0 + X_t^\top\beta(Z_{12,t}) + u_t, \quad t = 1, \ldots, T, \end{aligned}$$

and we obtain a semiparametric instrumental variable varying coefficient model. Note that we can impose the same condition on the first stage, e.g., $\pi(\mathscr{Z}_t) = Z_{11,t}^\top\gamma_0 + \pi(Z_{12,t}, Z_{2t})$.

4.2.3 Asymptotic properties

Here we derive the consistency and asymptotic normality of our estimators. First we introduce some notation, $\int k(z)dz = 1$, $\int zk(z)dz = 0$, $\int zz^\top k(z)dz = \mu_2(k)I$, and $\int k(z)^2dz = R(k)$, where $\mu_2(k) \neq 0$ and $R(k) \neq 0$ are scalars and I is the identity matrix. For the first stage estimator, let $v_t = X_t - \pi(\mathscr{Z}_t)$, for the second stage estimator let $\Gamma_*(z_1) = \mathrm{E}\left[\pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top \middle| Z_{1t} = z_1\right]$, for the optimal weighting matrix $\Omega_u(z) = \mathrm{E}\left[u_t ut^\top \middle| \mathscr{Z}_t = z\right]$; for the third stage estimator $\Gamma(z_1) = \mathrm{E}\left[\pi(\mathscr{Z}_t)\Omega(\mathscr{Z}_t)^{-1}\pi(\mathscr{Z}_t)^\top \middle| Z_{1t} = z_1\right]$.

The following conditions underlay the asymptotic theory of our estimators.

Assumption 4.2.1. Let $(Y_t, X_t, \mathscr{Z}_t, u_t)$ be strictly stationary and α-mixing process with $\alpha(t) = O(t^{-\tau})$, where $\tau = (2+\delta)(1+\delta)/\delta$ for some $\delta > 0$.

Assumption 4.2.2. For the random errors, u_t, $\mathrm{E}[u_t|\mathscr{Z}_t] = 0$.

Assumption 4.2.3. Let $f(\cdot)$ and $f_1(\cdot)$ the probability functions of z and z_1 and there exist a compact set D such that $\inf_{z\in D} f(z) > 0$. All density functions are continuously differentiable in all their arguments and they are bounded from above and below in any point of their support.

Assumption 4.2.4. The kernel functions $L(\cdot)$ and $K(\cdot)$ are compactly supported and bounded. Besides, let $|L(z) - L(z')| \leq C|z - z'|$ for all z and z'.

Assumption 4.2.5. The bandwidth matrices H_1 and H_2 are symmetric and strictly definite positive in such a way that $H_1 = o(H_2)$. Moreover, each entry of the matrices tend to zero as $T \to \infty$ in such a way that $T|H_1|^{1/2}/\log T \to \infty$ and $T|H_2|^{1/2} \to \infty$.

Assumption 4.2.6. The second order derivatives of $\gamma_1(\cdot)\ldots\gamma_{dm}(\cdot)$ are bounded and uniformly continuous and satisfy the Lpschitz condition.

Assumption 4.2.7. Let $\mathrm{E}\left[vec(v_t)vec(v_t)^\top \middle| \mathscr{Z}_t = z\right] < \infty$ and $\mathrm{E}\left[|vec(v_t)|^{2+\delta} \middle| \mathscr{Z}_t = z\right]$ be bounded and uniformly continuous in its support for some δ, where δ was defined in assumption 4.2.1.

Assumption 4.2.8. Let the second order derivatives of $\beta_1(\cdot)\ldots\beta_d(\cdot)$ be bounded and uniformly continuous in any point of their support.

Assumption 4.2.9. Let, $\Gamma_*(z_1)$, $\Omega(\mathscr{Z}_t)$ and $\Gamma(z_1)$ be positive definite, continuous and invertible

Assumption 4.2.10. For the same δ defined in assumption 4.2.1, let the matrices $\mathrm{E}|X_t|^{2+\delta} < \infty$, $\mathrm{E}\left[|u_t|^{2+\delta} \middle| Z_{1t} = z_1\right]$ and $\mathrm{E}\left[\left|\pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top\right|^{2+\delta} \middle| Z_{1t} = z_1\right]$ be bounded and uniformly continuous in their support.

Assumption 4.2.11. Let the following matrices $\mathrm{E}\left[\left|\pi(\mathscr{Z}_t)\Omega(\mathscr{Z}_t)^{-1}u_t\right|^{2+\delta} \middle| Z_{1t} = z_1\right]$ and $\mathrm{E}\left[\left|\pi(\mathscr{Z}_t)\Omega(\mathscr{Z}_t)^{-1}\pi(\mathscr{Z}_t)^\top\right|^{2+\delta} \middle| Z_{1t} = z_1\right]$ be bounded and uniformly continuous in their support, for the same δ defined in assumption 4.2.1.

Assumption 4.2.12. $T|H_2|^{\frac{1}{2}\left[1+\frac{2}{1+\delta}\right]} \to \infty$

Assumption 4.2.1 is a standard assumption in the asset pricing literature; the α-mixing condition is one of the weakest mixing conditions for weakly dependent stochastic processes. Many financial time series are α-mixing, see, Cai (2002a), Carrasco and Chen (2002), Chen and Tang (2005) among others for examples. Assumption 4.2.2 imposes the endogeneity of X_t and Z_{1t} and the exogeneity of $\mathscr{Z}_t$. Assumption 4.2.3 requires the densities of Z_{1t} and $\mathscr{Z}_t$ to be bounded and smooth functions , this assumption and assumption 4.2.4 are standard in nonparametric literature. Assumption 4.2.5 goes in the same direction as assumptions 3 and 8 in Cai et al. (2000); these assumptions are common in local fitting and two-stage nonparametric estimation. Note that here we assume that $L_{H_1}(\cdot) = |H_1|^{-1/2}L(\cdot H_1^{-1/2})$ and $K_{H_2}(\cdot) = |H_2|^{-1/2}K(\cdot H_2^{-1/2})$ but we can also assume $L_{h_1}(\cdot) = h_1^{-(p+q)}L(\cdot/h_1)$ and $K_{h_2}(\cdot) = h_2^{-q}K(\cdot/h_2)$; it is easy to show that our results hold with little change if we replace assumptions 4.2.5 and 4.2.12 with 4.2.13 and 4.2.14 respectively.

Assumption 4.2.13. The bandwidths h_1 and h_2 tend to zero as $T \to \infty$ in such a way that $Th_1^{p+q}/\log T \to \infty$ and $Th_2^q \to \infty$. Besides, we also need that $h_1 = o(h_2)$.

Assumption 4.2.14. $Th_2^{q\left[1+\frac{2}{1+\delta}\right]} \to \infty$

Assumptions 4.2.6 and 4.2.8 are common in local linear fitting literature and ensure that the Taylor approximation could carry through. Assumptions 4.2.7, 4.2.9, 4.2.10 and 4.2.11 are moment conditions similar to those in Cai et al. (2015), Cai et al. (2000), Cai et al. (2006), Cai and Li (2008), Fan and Huang (2005) or Cai et al. (2017) among others. Finally, assumption 4.2.12 is necessary for the central limit theorem, note that this assumption is not restrictive (see, Cai et al. (2015) for discussion).

The following theorems establish the main result of our work.

Theorem 4.2.1. Assuming that conditions 4.2.1-4.2.7 hold, then as $T \to \infty$ we obtain

$$vec\left(\hat{\pi}(z)\right) = vec\left(\pi(z)\right) + \frac{1}{2}\mu_2(L)\text{diag}_{dm}\left[\text{tr}\left(\mathscr{H}_{\gamma_r}(z)H_1\right)\right] i_{dm} + o_p\left(\text{tr}(H_1)\right),$$

where $\text{diag}_{dm}\left[\text{tr}\left(\mathscr{H}_{\gamma_r}(z)H_1\right)\right]$ stands for the diagonal matrix of elements $\text{tr}\left(\mathscr{H}_{\gamma_r}(z)H_1\right)$, for $r = 1,\ldots,dm$, and $\mathscr{H}_{\gamma_r}(z)$ is a $(p+q)\times(p+q)$ Hessian matrix of the rth component of $\gamma(z)$.

Theorem 4.2.2. Assuming that conditions 4.2.1-4.2.10 hold, then as $T \to \infty$ we obtain

$$\bar{\beta}(z_1) = \beta(z_1) + \frac{1}{2}\mu_2(K)\text{diag}_d\left[\text{tr}\left(\mathscr{H}_{\beta_r}(z_1)H_2\right)\right] i_d + o_p\left(\text{tr}(H_2)\right),$$

where $\text{tr}\left(\mathscr{H}_{\beta_r}(z_1)H_2\right)$, for $r = 1,\ldots,d$, and $\mathscr{H}_{\beta_r}(z_1)$ is a $q\times q$ Hessian matrix of the rth component of $\beta(z_1)$.

Theorem 4.2.3. Assuming that conditions 4.2.1-4.2.10 hold, then as $T \to \infty$ we obtain

$$\hat{\Omega}(z) \to_P \Omega(z)$$

Theorem 4.2.4. Assuming that conditions 4.2.1-4.2.12 hold, then as $T \to \infty$ we obtain

$$\sqrt{T|H_2|^{1/2}} \left(\hat{\beta}(z_1) - \beta(z_1) - b(z_1) \right) \to \mathscr{N} \left(0_d, R(K) f(z_1)^{-1} \Gamma(z_1)^{-1} \right)$$

where $b(z) = \frac{1}{2}\mu_2(K)\mathrm{diag}_d \left[\mathrm{tr}\left(\mathscr{H}_{\beta_r}(z_1) H_2 \right) \right] i_d$.

As consequence of theorems 4.2.1-4.2.4 it is easy to verify that the estimator is consistent with a coverage rate depending on T and H_2 but not H_1 as long as the condition $H_1 = o(H_2)$ is satisfied. Similar to the standard nonparametric regression (Fan and Gijbels (1996)), the bias appears mainly from the second order derivative of $\pi(\cdot)$ and $\beta(\cdot)$. Indeed, the approximation errors of the functions $\pi(\cdot)$ are transmitted to the bias in estimating $\beta(\cdot)$ but are asymptotically negligible due to our regularity conditions.

4.2.4 Bandwidth selection

Bandwidth selection is a challenging issue in noparametric and because of the nature of the multi-stage estimation. There are two bandwidths involved in the proposed three-stage estimation procedure. As mentioned before H_1 has to fulfill the condition $H_1 = o(H_2)$; that is, H_1 is chosen small enough that the bias term in the first stage is not too large. Then as suggested in Cai (2002b) we use a cross validation function to select the bandwidth $\hat{H}_{01}$; then, use $H_1 = A_0 \hat{H}_{01}$ where $A_0 = 1/2$ or smaller so that we choose a very small bandwidth for the first stage estimator.

For the second and third stage bandwidth, the choice can be done using standard methods of nonparametric regression (see, Li and Racine (2007)); for instance cross validation (Stone (1974)), pre-asymptotic substitution method (Fan and Gijbels (1995)), the plug-in bandwidth selector (Ruppert et al. (1995)), and the empirical bias method (Ruppert (1997)) among others. Unfortunately, there is no existing literature of a data driven bandwidth selection with optimal properties (Newey et al. (1999)).

4.3 Inference

Now, we consider constructing a test statistic on the varying coefficient $\beta(z_1)$. Consider the following general testing problem

$$H_0 : \beta(z_1) = \beta(z_1;\theta) \quad \text{vs} \quad H_0 : \beta(z_1) \neq \beta(z_1;\theta), \tag{4.10}$$

where $\beta(z_1;\theta)$ is a know parametric function of z_1. Note that this problem is different from the omnibus specification test proposed by Peñaranda et al. (2018) in the sense that their model under the null is also nonparametric. However, the above testing problem is a nonparametric test against a parametric form; the test is general enough that we can test, linearity $\beta(z_1;\theta) = \theta$, significance $\beta(z_1;\theta) = 0$, or functional form if $\beta(z_1;\theta)$ is taken to be a given parametric function of z_1. The goal of this article consists of testing that the betas of the CAPM model are constant in the direction of nonparametric alternatives, that is testing

$$\begin{aligned} H_0 \quad &: \quad \mathrm{Var}\left(\beta_j(z_1)\right) = 0 \text{ for all } j = 1,\ldots,d \\ &\text{vs} \\ H_1 \quad &: \quad \mathrm{Var}\left(\beta_j(z_1)\right) \neq 0 \text{ for some } j = 1,\ldots,d \ , \end{aligned} \tag{4.11}$$

As we already stated in the introductory section, testing on varying coefficients is of great interest as it means testing on the structural information and the underlying economic theory. For instance, one can test time variation of betas to support the studies of Bansal et al. (1993), Bansal and Viswanathan (1993), Cochrane (1996), Jagannathan and Wang (1996, 2002), Reyes (1999), Ferson and Harvey (1991, 1993, 1997, 1999), Cho and Engle (1999), Wang (2002, 2003), Akdeniz et al. (2003), Ang and Liu (2004), Fraser et al. (2004), Gagliardini et al. (2011), among others,

Following the idea of Cai and Tiwari (2000), Cai et al. (2000), Fan et al. (2001), Fan and Huang (2005) and Cai et al. (2017) we build up the test statistic based on the ratio of the residual sum of squares (RSS) of the model under the null and under the alternative. Here it is useful to restate the testing problem as

$$H_0 : \mathrm{E}[Y_t | \mathscr{Z}_t] = \pi(\mathscr{Z}_t)\beta \quad \text{vs} \quad H_1 : \mathrm{E}[Y_t | \mathscr{Z}_t] = \pi(\mathscr{Z}_t)\beta(Z_{1t});$$

note that we allow the parameter β to be unknown, but to be $\sqrt{T}$ consistent under the null hypothesis. Then, following Fan et al. (2001), the generalized likelihood ratio statistic is defined as

$$\lambda_T = \frac{T}{2}\log\frac{RSS_0}{RSS_1} \approx \frac{T}{2}\frac{RSS_0 - RSS_1}{RSS_1} \tag{4.12}$$

where

$$
\begin{aligned}
RSS_0 &= \left[Y_t - \hat{\pi}(\mathscr{Z}_t)^\top \hat{\beta}\right]^\top \left[Y_t - \hat{\pi}(\mathscr{Z}_t)^\top \hat{\beta}\right], \\
RSS_1 &= \left[Y_t - \hat{\pi}(\mathscr{Z}_t)^\top \hat{\beta}(z_1)\right]^\top \left[Y_t - \hat{\pi}(\mathscr{Z}_t)^\top \hat{\beta}(z_1)\right],
\end{aligned}
$$

are the residual sum of squares of the model under H_0 and under H_1 respectively. Here $\hat{\beta}$ is the least square estimator under the null hypothesis and $\hat{\beta}(z_1)$ is the nonparametric estimate of the varying coefficient obtained by (4.9). Note that, if we look at the (4.12), and following Fan and Huang (2005), we can use the Wilks phenomenon to derive the asymptotic distribution of the test.

Although we can obtain the asymptotic distribution of the test statistic, λ_T, the test is sensitive to the bandwidths in the finite sample case. It is also true that the studies of Fan et al. (2001), Fan and Huang (2005) and Cai et al. (2017) derive the asymptotic distribution of the test statistic, however in finite samples they prefer to use bootstrap critical values because of the sensitivity to the bandwidths; in the study of Cai et al. (2000) knowing that there exist a problem of bandwidth sensitivity, they prefer not to derive the asymptotic distribution of the test statistic. Therefore, and relaying on the previous studies, to gain better performance we suggest using a bootstrap method to calculate the p-value for the test statistic, λ_T. Here, we adopt the wild bootstrap method that takes account of heteroscedasticity of unknown form proposed by Davidson and MacKinnon (2010) and similar to the one used by Cai et al. (2017); then, the bootstrap approach is as follows:

1. Generate the bootstrap residuals $\{u_t^*, v_t^*\}_{t=1}^T = \{(\hat{u}_t, \hat{v}_t)\, e_t^*\}_{t=1}^T$ with

$$
e_t^* = \begin{cases} -\frac{\sqrt{5}-1}{2}, \text{ with probability } \frac{\sqrt{5}+1}{2\sqrt{5}}, \\ \frac{\sqrt{5}+1}{2}, \text{ with probability } \frac{\sqrt{5}11}{2\sqrt{5}}, \end{cases}
$$

   where

$$
\hat{u}_t = Y_t - X_t^\top \hat{\beta}(Z_{1t}) \quad , \quad \hat{v}_t = X_t - \hat{\pi}(\mathscr{Z}_t),
$$

   and define the bootstrap sample as follows:

$$
\begin{aligned}
Y_t^* &= \pi^*(\mathscr{Z}_t)^\top \hat{\beta} + u_t^*, \\
\pi^*(\mathscr{Z}_t) &= \hat{\pi}(\mathscr{Z}_t) + v_t^*,
\end{aligned}
$$

   where $\hat{\beta}$ is the least square estimator under H_0.

2. Calculate the bootstrap statistic λ_T^* based on the sample $\{Y_t^*, \pi^*(\mathscr{Z}_t), \mathscr{Z}_t\}_{t=1}^T$.

3. Reject the null hypothesis H_0 when λ_T is greater than the upper α point of the conditional distribution of λ_T^* given $\{Y_t, \pi(\mathscr{Z}_t), \mathscr{Z}_t\}_{t=1}^T$.

As for the p-value of the test, this is simply the frequency of the event $\lambda_T^* \geq \lambda_T$ in the replication of the bootstrap sample. Note that for the sake of simplicity we use the same bandwidth in calculating λ_T^* and λ_T. Note that here we bootstrap the residuals from the nonparametric estimation instead of the parametric estimation; that is so because the nonparametric estimates of the residuals are always consistent regardless if we are under the null or the alternative.

4.4 Monte Carlo study

To illustrate the validity of our methodology we conduct a Monte Carlo experiment to examine the finite sample performance of the proposed test. For this purpose, observations are generated according to the following varying coefficient model with endogenous regressors

$$Y_t = X_t\beta(Z_{1t}) + u_t, \quad t = 1, \ldots, T,$$

where $Z_{1t} = 0.05Z_{1(t-1)} + 0.9\varepsilon_{1t}$, $Z_{2t} = 0.05Z_{2(t-1)} + 0.9\varepsilon_{2t}$, ε_{1t} follows a uniform $[2,6]$ and ε_{2t} follows a uniform $[0,4]$. Also let $u_t = 0.05u_{t-1} + 0.9\varepsilon_{1t}$ and the endogenous variable X_t is generated following the next reduced form equation

$$X_t = 2\sin(Z_{1t} + Z_{2t}) + \varepsilon_{2t},$$

and the noises follow

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \to \mathscr{N}\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}\right];$$

here 0.7 controls the correlation between ε_{1t} and ε_{2t}. Finally we generate the random coefficients as

$$\begin{array}{ll} \text{a)}\beta(Z_{1t}) = 1, & \text{b)}\beta(Z_{1t}) = (1 + 0.1Z_{1t})\exp(-(0.5Z_{it} - 1.5)^2), \\ \text{c)}\beta(Z_{1t}) = 1.6 + 4.6I(Z_{1t} \geq 4), & \text{d)}\beta(Z_{1t}) = \cos(Z_{1t}). \end{array}$$

Here model a) represents the null hypothesis and the rest are different alternatives; alternatives b) and d) are smooth alternatives and alternative c) is a discontinuous alternative.

Note that in this study we have set $m = 1$, $d = 1$, $p = 1$, and $q = 1$. The sample size is chosen to be $T = 50$, 100 and 200 and for each sample size, the Monte Carlo study is based on 1000 replications and the bootstrap replications are set to 1000. For the kernel function $K(u)$ we choose the Epanechnikov kernel, $K(u) = 0.75(1-u^2)I(|u| \leq 1)$, and the bandwidth are chosen as it was described in section 4.2.4.

Model	a)	b)	c)	d)
Size		λ_T		
50	6.2	98.2	99.6	99.8
100	4.8	99.8	100	100
200	5.0	100	100	100

Table 4.1 Percentage of times H_0 was rejected, 5% of significance

The results of the simulation study are presented in table 4.1; here, we can see that our test is able to detect parameter constancy.

4.5 Conclusions

In this chapter, we estimate consistently the varying parameters of a conditional asset pricing model. The proposed nonparametric estimation procedure makes possible to estimate prices of risk from observed asset returns without imposing any parametric structure on price of risk function. The proposed test can be seen as a tool to test unconditional beta pricing models so that one can avoid Ghysels' critique (Ghysels (1998)) who states that misspecification of time-varying conditional moments and market prices of risk may induce larger pricing errors than those obtained by unconditional beta pricing models.

Conclusions

Since Cleveland et al. (1991) introduced varying coefficients models, econometric models have been greatly enriched. In the past three decades varying coefficient models have experienced a great growth, from both a methodological and a theoretical point of view, and the main reason is that they offer a quite general setting to handle many of the specification problems of nonparametric and semiparametric models. Varying coefficient models allow the coefficients of the regression model to be unknown functions of some other variables. Therefore, testing on varying coefficients implies testing on structural information and the underlying economic theory; thus, developing inference devices for varying coefficient models is crucial.

In this context, the goal of this Ph.D. thesis is twofold. On the one hand, to develop confidence bands for varying coefficient models using the empirical likelihood technique. On the other hand, testing that the varying coefficients are constant in the direction of nonparametric alternatives. In this way, and as conclusion, we summarize the main objectives pursued in each of the chapter of this dissertation and the main results obtained.

In Chapter 1 we investigate empirical likelihood based inference for fixed effects varying coefficient panel data models to build up confidence bands for the varying coefficients; firstly, and following the idea of removing the fixed effects, (Rodriguez-Poo and Soberón, 2014, 2015), we get rid of the fixed effects using the first different and the within transformation. Then, we show that the naive empirical likelihood ratio is asymptotically chi-squared when undersmoothing is employed. The interesting fact about the empirical likelihood ratio is that it does not need plug-in estimates of the limiting variance as the studentization is carried out internally. As in the work of Xue and Zhu (2007) we correct for the bias, and propose a mean-corrected and residual-adjusted empirical likelihood ratios that without undersmoothing, both have standard chi-squared limit distributions.

As a by product, we give the empirical maximum likelihood estimators of the varying coefficient and their derivatives; the results resembles to those obtained in Rodriguez-Poo and Soberón (2014, 2015), however the result for the derivative of the varying coefficient is brand new. The idea is that this derivative result can be use to test constancy, that is, derivative results that are not statistically significant are equivalent to state that the varying coefficient is constant. We also obtain the asymptotic distribution of these estimators and we propose some procedures to calculate the bandwidths empirically.

To show the feasibility of the technique and to analyse its small sample properties we implement a Monte Carlo simulation exercise; here we conclude the length of the confidence interval is smaller for the residual adjusted empirical likelihood ratio (RAEL), being smaller than the mean corrected empirical likelihood ratio (MCEL) and the asymptotic normal approximation (NA). Thus, we can conclude by saying that the RAEL and MCEL confidence bands behave better than the NA confidence bands. Between RAEL and MCEL confidence bands, simulations results show that the RAEL confidence bands behave better than the MCEL. Also, by comparing the within method with the first difference method we can conclude that for the NA and the RAEL confidence bands the First Difference method reduces the length of the confidence interval; however the MCEL confidence interval increases it length in comparison to the Within method.

In the empirical analysis about the production efficiency of the European Union's companies, we can conclude that the marginal productivity of liquid capital tends to be decreasing; however when it reaches a certain level of *R&D* expenses it tends to be steady and close to zero. Basically, this means that companies with small *R&D* expenses have a decreasing marginal productivity of liquid capital. On its part, the marginal productivity of fixed capital is not a linear function with the level of *R&D* expenses. Clearly, there exist an upward general trend, with a bell shape form for companies with large *R&D* expenses, which means that, while modest *R&D* expenses can improve the fixed capital productivity, higher *R&D* expenses leads to lower fixed capital productivity. For labour marginal productivity we observe that the labour marginal productivity is not a linear function of *R&D* with an inverted bell shape that suggests that companies with reduced *R&D* tend to have lower labour marginal productivity at the beginning while companies with higher *R&D* are more likely to have an increase in labour marginal productivity. Finally, using these results we can not conclude that the returns to scale are not equal to one because one is within the confidence interval. However, we can conclude that the returns to scale are not linear with *R&D* and they seem to have a negative effect in the behaviour of the returns to scale.

In Chapter 2 we investigate empirical likelihood based inference for nonparametric categorical varying coefficient panel data models with fixed effects under cross-sectional dependence. The main difference with chapter 1 is that in this case the varying coefficient varies according to a discrete variable and therefore we need kernel functions for discrete variables. In this context, and following the same outline of chapter 1 we first get rid of the fixed effects using a modified version of the within transformation, (Feng et al., 2017), used in chapter 1. Then we show that the naive empirical likelihood ratio is asymptotically standard chi-squared using a nonparametric version of the Wilks' theorem, (Wilks, 1938). Note that because in this case due to the cross sectional dependence the estimation variance becomes cumbersome, it is a good thing that we do not need plug-in estimates of the limiting variance.

As a by product, we propose also an empirical maximum likelihood estimator of the categorical varying coefficient; the result obtained are similar to the ones obtained in Feng et al. (2017) but we add by giving results when the bandwidth parameter is unknown. We also obtain the asymptotic distribution of this estimator. We also illustrated the proposed technique in an application that reports estimates of strike activities from 17 countries of the OECD for the period 1951 – 1985. Note that because of the cumbersome expression of the variance, the empirical likelihood approach is easier to implement than the asymptotic normality. From the empirical application we can see that the confidence bands using empirical likelihood behave better than the ones estimated using the asymptotic normal distribution.

In Chapter 3, we propose constancy test for coefficients in semi-varying coefficients models. The testing procedure resembles in spirit the union-intersection (U-I) parameter stability tests in time series, where observations are sorted according to the explanatory variable responsible for the coefficients varying. Here, the variable responsible for the coefficient varying may depend on the rest of explanatory variables in an unknown form. In this context, we use induced order statistics or concomitants to sort the data according to the regressor responsible for the varying coefficient.

The test, implemented using bootstrap, is based on comparing the OLS coefficients of subsamples. The test is justified under fairly weak regularity conditions, which allow discontinuous random coefficients under the alternative hypothesis. Our test forms a basis for specification testing of parametric varying coefficients and, in particular, for testing the functional form of interactive effects.

Simulation results have provided evidence of the good performance of our test in finite samples compared with a CUSUM-type test (Stute, 1997), designed to omnibus specification testing of linear regression models, and a smooth LR test, (Cai et al., 2000), designed to test varying coefficients constancy in the direction of smooth alternatives. The CUSUM test, like ours, does not require estimating the model on the alternative, but the LR-type test compares the restricted and unrestricted sum of squared residuals and, hence, requires estimating the nonparametric smooth varying coefficients. Simulations show that, unlike our test, the two competitors suffer of the curse of dimensionality. These also show that the LR smooth test exhibit a lack of power, compared with the two competitors, under alternatives with discontinuous varying coefficients.

We have also included a real data application to model interactive effects of IQ in a returns of education model. We show that the returns to education varies according to IQ, a proxy for ability.

Finally, extending chapter 3 by allowing endogenous explanatory variables, chapter 4 propose a methodology for testing coefficients constancy in varying coefficient capital asset pricing models (CAPM) with endogenous regressors. The testing procedure is defined as a generalized likelihood ratio that focus on the comparison of the restricted and unrestricted sum of squared residuals. The proposed test can be seen as a tool to test unconditional beta pricing models.

As a by product, we have developed a nonparametric method that makes possible to estimate prices of risk from observed asset returns without imposing any parametric structure on price of risk function; besides we establish the asymptotic properties of the estimators. Finally, the Monte Carlo experiments study, using critical values and p-values estimated by the bootstrap technique, provides evidence of the good performance of our test in finite samples.

5.1 Future research

Throughout this thesis and given the advantages of introducing varying coefficient models, some future lines of research have arisen. In this sense, a first line of future research consist in the extension of chapters 1 and 2 by allowing a the varying coefficient to vary according to a mixture of discrete and continuous variables, see, e.g., kernel estimation with mixed data in Li and Racine (2007).

Another line of relevant research is the development of nonparametric tests using empirical likelihood; here it would be interesting to extend the work of chapter 3 and obtain a distribution free test based on the empirical likelihood technique. Some relevant work have been done in Chen et al. (2003), Einmahl and McKeague (2003), Zou et al. (2007) and Liu et al. (2008) under different settings. Also it would be interesting to extend chapter 3 to the panel data framework. We can also extend Chapter 4 by using a test based on empirical likelihood and derive its asymptotic distribution.

Conclusiones

Desde que Cleveland et al. (1991) introdujeron los modelos de coeficientes variables, los modelos econométricos se han enriquecido enormemente. En las tres últimas décadas, estos modelos de coeficientes variables han experimentado un gran crecimiento, tanto desde el punto de vista metodológico como teórico, y la razón principal es que ofrecen un marco bastante general que nos permite lidiar con muchos de los problemas de especificación en modelos no paramétricos y semiparamétricos. Los modelos de coeficientes variables permiten que los coeficientes del modelo de regresión sean funciones desconocidas de otras variables. Por lo que, los contrastes sobre los coeficientes variables implican contrastar la información estructural y la teoría económica subyacente; por lo tanto, el desarrollo de técnicas de inferencia para modelos de coeficientes variables es crucial.

En este contexto, el objetivo de esta tesis docotoral es doble. Por un lado, desarrollar bandas de confianza para modelos de coeficientes variables utilizando la técnica de verosimilitud empírica. Por otro lado, desarrollar tests que nos permitan discernir si los coeficientes variables son constantes en la dirección de alternativas no paramétricas. De esta manera, y como conclusión, resumimos los principales objetivos de cada uno de los capítulos de esta disertación y los principales resultados obtenidos.

En el capítulo 1 se investiga técnicas de inferencia estadística basadas en la verosimilitud empírica para modelos de datos de panel con coeficientes variables y efectos fijos para construir bandas de confianza para los coeficientes; En primer lugar, y siguiendo la idea de eliminar los efectos fijos, (Rodriguez-Poo and Soberón, 2014, 2015), nos deshacemos de los efectos fijos utilizando las ténicas de primeras diferencias y la transformación within. Luego, demostramos que el ratio de verosimilitud empírica es asintóticamente chi-cuadrado cuando se emplea undersmoothing. Un hecho interesante acerca del ratio de verosimilitud empírica es que no necesita estimaciores para la varianza, ya que la estudianteización se lleva a cabo internamente. Al igual que en el trabajo de Xue and Zhu (2007), corregimos por el sesgo y proponemos dos correciones del ratio de verosimilitud empírica, una corregida por la media y otra ajustada por residuos; además demostramos que , sin undersmoothing, ambas tienen distribuciones asintóticas chi cuadrado.

Como subproducto, proporcionamos los estimadores de máxima verosimilitud empírica del coeficiente variable y sus derivadas; los resultados se asemejan a los obtenidos en Rodriguez-Poo and Soberón (2014, 2015), sin embargo, el resultado para las derivadas del coeficiente variable es completamente nuevo. La idea es que este resultado se pueda usar para probar la constancia, es decir, si las derivadas no son estadísticamente significativas, es equivalente a decir que los coeficientes variables son constantes. También obtenemos la distribución asintótica de los estimadores y proponemos algunos procedimientos para calcular los bandwidth empíricamente.

Para mostrar la viabilidad de la técnica y analizar sus propiedades en muestras finitas, implementamos un ejercicio de simulación de Monte Carlo; aquí concluimos que la longitud del intervalo de confianza es menor para el ratio de verosimilitud empírica ajustado por los residuos (RAEL), siendo más pequeña que la del ratio de verosimilitud empírica corregido por la media (MCEL) y la de la aproximación normal asintótica (NA). Por lo tanto, podemos concluir diciendo que las bandas de confianza de RAEL y MCEL se comportan mejor que las bandas de confianza de NA. Entre las bandas de confianza de RAEL y MCEL, los resultados de las simulaciones muestran que las bandas de confianza de RAEL se comportan mejor que el MCEL. Además, al comparar el método within con el método de primeras diferencias, podemos concluir que para las bandas de confianza NA y RAEL, el método de primeras diferencias reduce la longitud del intervalo de confianza; sin embargo, el intervalo de confianza de MCEL aumenta su longitud en comparación con el método within.

En el análisis empírico sobre la eficiencia en la producción de las empresas de la Unión Europea, podemos concluir que la productividad marginal del capital líquido tiende a disminuir; sin embargo, cuando alcanza un cierto nivel de gastos de $I+D$, tiende a ser constante y cercano a cero. Básicamente, esto significa que las empresas con gastos pequeños de $I+D$ tienen una productividad marginal del capital líquido decreciente. Por su parte, la productividad marginal del capital fijo no es una función lineal con el nivel de gastos de $I+D$. Claramente, existe una tendencia ascendente, lo que significa que, mientras que gastos modestos de $I+D$ pueden mejorar la productividad del capital fijo, gastos muy elevados de $I+D$ conducen a una menor productividad del capital fijo. Para la productividad marginal del trabajo, observamos que esta no es una función lineal de $I+D$, con una forma de campana invertida sugiere que las empresas con $I+D$ reducido tienden a tener una productividad marginal del trabajo más baja al principio mientras que las empresas con mayores $I+D$ tienen más probabilidades de tener un aumento en la productividad marginal del trabajo. Finalmente, a la vista de estos resultados, no podemos concluir que los rendimientos a escala no sean iguales a uno porque uno está dentro del intervalo de confianza. Sin embargo, podemos concluir que los rendimientos a escala no son lineales con $I+D$ y parecen tener un efecto negativo en el comportamiento de los rendimientos a escala.

En el capítulo 2 se investiga tecnicas de inferencia estadística basadas en la verosimilitud empírica para modelos de datos de panel con efectos fijos y coeficientes variables categóricos o discretos bajo dependencia de sección cruzada. La principal diferencia con el capítulo 1 es que, en este caso, el coeficiente variable varía según una variable discreta y, por lo tanto, necesitamos funciones de kernel para variables discretas. En este contexto, y siguiendo el mismo esquema del capítulo 1, primero nos deshacemos de los efectos fijos usando una versión modificada de la transformación within, (Feng et al., 2017), utilizada en el capítulo 1.

Luego demostramos que el ratio de verosimilitud empírica es asintóticamente chi-cuadrado utilizando una versión no paramétrica del teorema de Wilks, (Wilks, 1938).

Como subproducto, proponemos un estimador de máxima verosimilitud empírica del coeficiente variable categórico; los resultados obtenidos son similares a los obtenidos en Feng et al. (2017), pero nuestros resultados se obtienen también cuando el bandwidth es desconocido. Además obtenemos la distribución asintótica de este estimador. También ilustramos la técnica propuesta en una aplicación que informa de las estimaciones de las actividades de huelga de 17 países de la OCDE para el periodo 1951 – 1985. En este caso, debido a la dependencia de sección cruzada, la estimación de la la varianza se vuelve complicada; en este contexto resulta deseable que la técnica de verosimilitu empírica no necesite estimaciones de la varianza. Bajo los supuestos de este capítulo, la verosimilitud empírica resulta ser más sencilla de implementar que la normalidad asintótica. Con lo que se refiere a la aplicación empírica podemos ver que las bandas de confianza que usan la verosimilitud empírica se comportan mejor que las estimadas usando la distribución normal asintótica.

En el capítulo 3, proponemos un test de constancia para coeficientes en modelos de coeficientes semi-variables. El procedimiento de contraste se asemeja en espíritu a las pruebas de estabilidad de parámetros de unión-intersección (U-I) en series temporales, donde las observaciones se clasifican de acuerdo con la variable explicativa responsable de que los coeficientes varíen. En este contexto, la variable responsable de la variación del coeficiente puede depender del resto de variables explicativas en una forma desconocida. Para ordenar los datos de acuerdo con el regresor responsable del coeficiente variable, utilizamos estadísticos de orden inducido o concomitantes.

El test, implementado utilizando bootstrap, se basa en la comparacion de los coeficientes MCO de submuestras. El contraste se justifica bajo condiciones de regularidad bastante débiles, lo que permite coeficientes aleatorios discontinuos bajo la hipótesis alternativa. Nuestro test constituye la base para contrastes de especificación de coeficientes variables paramétricos y, en particular, para contrastar la forma funcional de los efectos interactivos.

Los resultados de las simulaciones han demostrado el buen funcionamiento de nuestro test en muestras finitas en comparación con el test de tipo CUSUM (Stute, 1997), diseñado para contrastes de especificación de tipo omnibus en modelos de regresión lineal, y el test de tipo LR basado en el suavizado, (Cai et al., 2000), diseñado para contrastar la constancia de coeficientes variables en la dirección de alternativas no paramétricas. El test de tipo CUSUM, como el nuestro, no requiere estimar el modelo bajo la alternativa, sin embargo el test de tipo LR compara la suma de cuadrados de los residuso del modelo restringido y no restringido y, por lo tanto, requiere de la estimación de los coeficientes no paramétricos bajo la alternativa. Las simulaciones muestran que, a diferencia de nuestro test, los competidores sufren la maldición de la dimensionalidad. Estos también muestran que el contraste de tipo LR muestra una falta de potencia, en comparación con sus competidores, en alternativas con coeficientes variables discontinuos.

También hemos incluido una aplicación con datos reales para modelizar los efectos interactivos de IQ (una proxy de habilidad) en un modelo de rendimientos educativos. Con nuestro test mostramos que los rendimientos educativos varían de acuerdo con IQ.

Finalmente, y extendiendo el capítulo 3 al permitir variables explicativas endógenas, el capítulo 4 se propone una metodología para contrastar constancia de los coeficientes en modelos de valoración de activos financieros (CAPM) con coeficientes variables y regresores endógenos. El test se define como un ratio de verosimilitud generalizado que se centra en la comparación de la suma de cuadrados de los residuso del modelo restringido y no restringido. El contraste propuesto puede verse como una herramienta para contrastar modelos CAPM incondicionales.

Como producto derivado, hemos desarrollado un método no paramétrico que hace posible estimar los coeficientes a partir de la rentabilidad de los activos observados sin imponer ninguna estructura paramétrica en en ellos; Además establecemos las propiedades asintóticas de los estimadores. Finalmente, el experimento de Monte Carlo, que utiliza valores críticos y p-valores estimados por la técnica bootstrap, proporciona evidencia del buen desempeño de nuestro test en muestras finitas.

6.1 Líneas de investigación futura

A lo largo de esta tesis y dadas las ventajas de los modelos de coeficientes variables, han surgido algunas líneas de investigación futuras. En este sentido, una primera línea de investigación futura consiste en la extensión de los capítulos 1 y 2 al permitir que el coeficiente variable varíe de acuerdo con una mezcla de variables discretas y continuas, ver, por ejemplo, estimación kernel con datos mixtos en Li and Racine (2007).

Otra línea de investigación relevante es el desarrollo de contrastes no paramétricas utilizando la verosimilitud empírica; Aquí sería interesante extender el trabajo del capítulo 3 y obtener una test de distribución libre basado en la técnica de verosimilitud empírica. Algunos trabajos relevantes se han realizado en Chen et al. (2003), Einmahl and McKeague (2003), Zou et al. (2007) y Liu et al. (2008) entre otros, aunque bajo un contexto diferente. También sería interesante extender el capítulo 3 al marco de datos del panel. El capítulo 4 también se puede ampliar usando un contraste basado en la técnica de verosimilitud empírica y derivando su distribución asintótica.

References

Ahmad, I., Leelahanon, S., and Li, Q. (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics*, pages 258–283.

Aitchison, J. and Aitken, C. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, 63(3):413–420.

Akdeniz, L., Altay-Salih, A., and Caner, M. (2003). Time-varying betas help in asset pricing: the threshold capm. *Studies in Nonlinear Dynamics & Econometrics*, 6(4).

Amano, T., Kato, T., and Taniguchi, M. (2012). Statistical estimation for capm with long-memory dependence. *Advances in Decision Sciences*, 2012.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Andrews, D. W. K. (1997). A conditional kolmogorov test. *Econometrica*, 65(5):1097–1128.

Ang, A. and Liu, J. (2004). How to discount cashflows with time-varying expected returns. *The Journal of Finance*, 59(6):2745–2783.

Arellano, M. (1987). Practitioners'corner: Computing robust standard errors for within-groups estimators. *Oxford bulletin of Economics and Statistics*, 49(4):431–434.

Arteaga-Molina, L. A. and Delgado González, M. n. (2019). Testing Constancy in Varying Coefficient Models. UC3M Working papers. Economics 27981, Universidad Carlos III de Madrid. Departamento de Economía.

Arteaga-Molina, L. A. and Rodriguez-Poo, J. M. (2018). Empirical likelihood based inference for fixed effects varying coefficient panel data models. *Journal of Statistical Planning and Inference*, 196:144–162.

Arteaga-Molina, L. A. and Rodríguez-Poo, J. M. (2019). Empirical likelihood based inference for a categorical varying-coefficient panel data model with fixed effects. *Journal of Multivariate Analysis*, 173:110–124.

Aue, A., Horváth, L., Kokoszka, P., and Steinebach, J. (2008). Monitoring shifts in mean: Asymptotic normality of stopping times. *Test*, 17(3):515–530.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

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

Bansal, R. and Viswanathan, S. (1993). No arbitrage and arbitrage pricing: A new approach. *The Journal of Finance*, 48(4):1231–1262.

Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics. *The Annals of Statistics*, 2(5):1034–1039.

Bhattacharya, P. K. (1976). An invariance principle in regression analysis. *The Annals of Statistics*, 4(3):621–624.

Bhattacharyya, G. K. and Johnson, R. A. (1968). Nonparametric tests for shift at an unknown time point. *The Annals of Mathematical Statistics*, 39(5):1731–1743.

Billingsley, P. (1968). *Convergence of probability measures*. John Wiley & Sons.

Black, F., Jensen, M. C., Scholes, M., et al. (1972). The capital asset pricing model: Some empirical tests. *Studies in the theory of capital markets*, 81(3):79–121.

Blackburn, M. L. and Neumark, D. (1995). Are ols estimates of the return to schooling biased downward? another look. *The Review of Economics and Statistics*, 77(2):217–230.

Blundell, R., Chen, X., and Kristensen, D. (2007). Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669.

Blundell, R. and Powell, J. L. (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357.

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192.

Cai, Z. (2002a). Regression quantiles for time series. *Econometric Theory*, 18(1):169–192.

Cai, Z. (2002b). Two-step likelihood estimation procedure for varying-coefficient models. *Journal of Multivariate Analysis*, 82(1):189–209.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Cai, Z., Das, M., Xiong, H., and Wu, X. (2006). Functional coefficient instrumental variables models. *Journal of Econometrics*, 133(1):207–241.

Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956.

Cai, Z., Fang, Y., Lin, M., and Su, J. (2017). Inferences for a partially varying coefficient model with endogenous regressors. *Journal of Business & Economic Statistics*, 0(0):1–13.

Cai, Z. and Li, Q. (2008). Nonparametric estimation of varying coefficient dynamic panel data models. *Econometric Theory*, 24(05):1321–1342.

Cai, Z., Ren, Y., and Sun, L. (2015). Pricing kernel estimation: A local estimating equation approach. *Econometric Theory*, 31(3):560–580.

Cai, Z. and Tiwari, R. C. (2000). Application of a local linear autoregressive model to bod time series. *Environmetrics: The official journal of the International Environmetrics Society*, 11(3):341–350.

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

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18(1):17–39.

Chan, K. S. and Tsay, R. S. (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85(2):413–426.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *The Annals of Statistics*, 16(1):136–146.

Chen, J., Gao, J., and Li, D. (2012a). A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models. *Econometric Theory*, 28(05):1144–1163.

Chen, J., Gao, J., and Li, D. (2012b). Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, 171(1):71–85.

Chen, J., Gao, J., and Li, D. (2013). Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business & Economic Statistics*, 31(3):315–330.

Chen, S. X., Härdle, W., and Li, M. (2003). An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):663–678.

Chen, S. X. and Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of financial econometrics*, 3(2):227–255.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, 35(3):999–1018.

Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96(454):605–619.

Chibisov, D. M. (1964). Some theorems on the limiting behaviour of an empirical distribution function. *Trudy Matematicheskogo Instituta imeni VA Steklova*, 71:104–112.

Cho, Y.-H. and Engle, R. F. (1999). Time-varying betas and asymmetric effect of news: Empirical analysis of blue chip stocks. Technical report, National Bureau of Economic Research.

Chow, Y. S. and Teicher, H. (1988). Probabilty theory.

Cleveland, W., Grosse, E., and Shyu, W. (1991). Local regression models. In Chambers, J. and Hastie, T., editors, *Statistical Models in S*, pages 309–376. Pacific Grove: Wadsworth & Brooks.

Cochrane, J. H. (1996). A cross-sectional test of an investment-based asset pricing model. *Journal of Political Economy*, 104(3):572–621.

Csörgő, M. and Horváth, L. (1988). 20 nonparametric methods for changepoint problems. *Handbook of statistics*, 7:403–425.

Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley series in probability and statistics. Wiley.

Darling, D. A., Erdős, P., et al. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J*, 23(1):143–155.

Darolles, S., Fan, Y., Florens, J.-P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565.

Davidson, R. and MacKinnon, J. G. (2010). Wild bootstrap tests for iv regression. *Journal of Business & Economic Statistics*, 28(1):128–144.

Davydov, Y. and Egorov, V. (2000). Functional limit theorems for induced order statistics. *Mathematical Methods of Statistics*, 9(3):297–313.

DiCiccio, T., Hall, P., and Romano, J. (1991). Empirical likelihood is bartlett-correctable. *The Annals of Statistics*, 19(2):1053–1061.

Dittmar, R. F. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *The Journal of Finance*, 57(1):369–403.

Dumas, B. (1994). A test of the international capm using business cycles indicators as instrumental variables. In *The internationalization of equity markets*, pages 23–58. University of Chicago Press.

Dumas, B. and Solnik, B. (1995). The world price of foreign exchange risk. *The journal of finance*, 50(2):445–479.

Einmahl, J. H. and McKeague, I. W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli*, 9(2):267–290.

Einmahl, U. (1989). The darling-erdős theorem for sums of iid random variables. *Probability Theory and Related Fields*, 82(2):241–257.

Engle, R. F., Granger, C. W. J., Rice, J., and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*, 81(394):310–320.

Escanciano, J. C., Hoderlein, S., Lewbel, A., Linton, O., and Srisuma, S. (2015). Nonparametric euler equation identification and estimation. Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, pages 196–216.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 371–394.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Chapman and Hall.

Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11(6):1031–1057.

Fan, J. and Yao, Q. (2003). *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media.

Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics*, 29(1):153–193.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27(5):1491–1518.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, 27(4):715–731.

Feng, G., Gao, J., Peng, B., and Zhang, X. (2017). A varying-coefficient panel data model with fixed effects: Theory and an application to us commercial banks. *Journal of Econometrics*, 196(1):68–82.

Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of Political Economy*, 99(2):385–415.

Ferson, W. E. and Harvey, C. R. (1993). The risk and predictability of international equity returns. *Review of financial Studies*, 6(3):527–566.

Ferson, W. E. and Harvey, C. R. (1997). Fundamental determinants of national equity market returns: A perspective on conditional asset pricing. *Journal of Banking & Finance*, 21(11-12):1625–1665.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360.

Florens, J. (2005). Engogeneity in nonseparable models. application to treatment models where the outcomes are durations. Technical report, mimeo, University of Toulouse.

Florens, J.-P., Johannes, J., and Van Bellegem, S. (2011). Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory*, 27(3):472–496.

Florens, J.-P., Johannes, J., and Van Bellegem, S. (2012). Instrumental regression in partially linear models. *The Econometrics Journal*, 15(2):304–324.

Florens, J.-P. and Simoni, A. (2012). Nonparametric estimation of an instrumental regression: A quasi-bayesian approach based on regularized posterior. *Journal of Econometrics*, 170(2):458–475.

Fraser, P., Hamelink*, F., Hoesli, M., and Macgregor, B. (2004). Time-varying betas and the cross-sectional return–risk relation: evidence from the uk. *The European Journal of Finance*, 10(4):255–276.

Gaenssler, P. and Stute, W. (1979). Empirical processes: A survey of results for independent and identically distributed random variables. *The Annals of Probability*, 7(2):193–243.

Gagliardini, P., Ossola, E., and Scaillet, O. (2011). Time-varying risk premium in large cross-sectional equidity datasets, swiss finance institute. Technical report, working paper.

Gao, J. (2007). *Nonlinear time series: semiparametric and nonparametric methods*. CRC Press.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573.

Gourieroux, C. and Monfort, A. (2007). Econometric specification of stochastic discount factor models. *Journal of Econometrics*, 136(2):509–530.

Greene, W. (2005). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, 126(2):269–303.

Hall, P. and Heyde, C. C. (2014). *Martingale limit theory and its application*. Academic press.

Hall, P. and Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review/Revue Internationale de Statistique*, 58(2):109–127.

Hall, P. and Owen, A. B. (1993). Empirical likelihood confidence bands in density estimation. *Journal of Computational and Graphical Statistics*, 2(3):273–289.

Hall, P., Sheather, S. J., Jones, M., and Marron, J. S. (1991). On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78(2):263–269.

Härdle, W. (1990). *Applied nonparametric regression*. Number 19. Cambridge university press.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):757–796.

Hawkins, D. L. (1989). A u-i approach to retrospective testing for shifting parameters in a linear model. *Communications in Statistics-Theory and Methods*, 18(8):3117–3134.

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72(357):180–186.

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):244–248.

Henderson, D. J., Carroll, R. J., and Li, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1):257–275.

Henderson, D. J. and Ullah, A. (2005). A nonparametric random effects estimator. *Economics Letters*, 88(3):403–407.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1):1–17.

Hoderlein, S. and Sherman, R. (2015). Identification and estimation in a correlated random coefficients binary response model. *Journal of Econometrics*, 188(1):135–149.

Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.

Horowitz, J. L. and Lee, S. (2007). Nonparametric instrumental variables estimation of a quantile regression model. *Econometrica*, 75(4):1191–1208.

Horváth, L. (1993). Change in autoregressive processes. *Stochastic processes and their Applications*, 44(2):221–242.

Horváth, L. and Shao, Q.-M. (1995). Limit theorems for the union-intersection test. *Journal of statistical planning and inference*, 44(2):133–148.

Hsiao, C. (2014). *Analysis of panel data*, volume 54. Cambridge university press.

Hu, T. and Xia, Y. (2012). Adaptive semi-varying coefficient model selection. *Statistica Sinica*, 22(2):575–599.

Jagannathan, R. and Wang, Z. (1996). The conditional capm and the cross-section of expected returns. *The Journal of finance*, 51(1):3–53.

Jagannathan, R. and Wang, Z. (2002). Empirical evaluation of asset-pricing models: A comparison of the sdf and beta methods. *The Journal of Finance*, 57(5):2337–2367.

James, B., James, K. L., and Siegmund, D. (1987). Tests for a change-point. *Biometrika*, 74(1):71–83.

Jegadeesh, N., Noh, J., Pukthuanthong, K., Roll, R., and Wang, J. (2019). Empirical tests of asset pricing models with individual assets: Resolving the errors-in-variables bias in risk premium estimation. *Journal of Financial Economics*.

Kauermann, G. and Tutz, G. (1999). On model diagnostics using varying coefficient models. *Biometrika*, 86(1):119–128.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *The Annals of Statistics*, 35(1):92–108.

Kniesner, T. J. and Li, Q. (2002). Nonlinearity in dynamic adjustment: Semiparametric estimation of panel labor supply. *Empirical Economics*, 27(1):131–148.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models. *Statistica Sinica*, 4(1):199–218.

Kottaridi, C. and Stengos, T. (2010). Foreign direct investment, human capital and non-linearities in economic growth. *Journal of Macroeconomics*, 32(3):858–871.

Lettau, M. and Ludvigson, S. (2001). Resurrecting the (c) capm: A cross-sectional test when risk premia are time-varying. *Journal of political economy*, 109(6):1238–1287.

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

Li, C. and Liang, Z. (2015). Asymptotics for nonparametric and semiparametric fixed effects panel models. *Journal of Econometrics*, 185(2):420–434.

Li, D., Chen, J., and Lin, Z. (2011a). Statistical inference in partially time-varying coefficient models. *Journal of Statistical Planning and Inference*, 141(2):995 – 1013.

Li, G., Peng, H., and Tong, T. (2013a). Simultaneous confidence band for nonparametric fixed effects panel data models. *Economics Letters*, 119(3):229–232.

Li, G. and Van Keilegom, I. (2002). Likelihood ratio confidence bands in non–parametric regression with censored data. *Scandinavian Journal of Statistics*, 29(3):547–562.

Li, G., Xue, L., and Lian, H. (2011b). Semi-varying coefficient models with a diverging number of components. *Journal of Multivariate Analysis*, 102(7):1166 – 1174.

Li, K., Li, D., Liang, Z., and Hsiao, C. (2017). Estimation of semi-varying coefficient models with nonstationary regressors. *Econometric Reviews*, 36(1–3):354–369.

Li, Q., Huang, C. J., Li, D., and Fu, T. (2002). Semiparametric smooth coefficient models. *Journal of Business and Economic Statistics*, 20(3):412–422.

Li, Q., Ouyang, D., and Racine, J. S. (2013b). Categorical semiparametric varying-coefficient models. *Journal of Applied Econometrics*, 28(4):551–579.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Lin, X. and Carroll, R. J. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American statistical Association*, 95(450):520–534.

Liu, Y., Zou, C., and Zhang, R. (2008). Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics-Theory and Methods*, 37(16):2551–2563.

Martins-Filho, C. and Yao, F. (2009). Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis*, 100(3):309–333.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.

McCabe, B. P. M. and Tremayne, A. R. (1995). Testing a time series for difference stationarity. *The Annals of Statistics*, 23(3):1015–1028.

Nagel, S. and Singleton, K. J. (2011). Estimation and evaluation of conditional asset pricing models. *The Journal of Finance*, 66(3):873–909.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Newey, W. K. and Powell, J. L. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Newey, W. K., Powell, J. L., and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230.

Orbe, S., Gil-Bazo, J., and Ferreira, E. (2008). Nonparametric estimation of conditional beta pricing models. *Working paper*.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1):90–120.

Owen, A. (1991). Empirical likelihood for linear models. *The Annals of Statistics*, pages 1725–1747.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249.

Owen, A. B. (2001). *Empirical likelihood*. CRC press.

Parmeter, C. F. and Racine, J. S. (2018). Nonparametric estimation and inference for panel data models. Department of Economics Working Papers 2018-02, McMaster University.

Peñaranda, F., Rodriguez-Poo, J. M., and Sperlich, S. (2018). Nonparametric specification testing of conditional asset pricing models. *Working paper*.

Pesaran, M. H. and Tosetti, E. (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics*, 161(2):182–202.

Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, 22(1):300–325.

Qu, A. and Li, R. (2006). Quadratic inference functions for varying-coefficient models with longitudinal data. *Biometrics*, 62(2):379–391.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53(284):873–880.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, 55(290):324–330.

Reyes, M. G. (1999). Size, time-varying beta, and conditional heteroscedasticity in uk stock returns. *Review of Financial Economics*, 8(1):1–10.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Rodriguez-Poo, J. M. and Soberón, A. (2014). Direct semi-parametric estimation of fixed effects panel data varying coefficient models. *The Econometrics Journal*, 17(1):107–138.

Rodriguez-Poo, J. M. and Soberón, A. (2015). Nonparametric estimation of fixed effects panel data varying coefficient models. *Journal of Multivariate Analysis*, 133:95–122.

Rodriguez-Poo, J. M. and Soberon, A. (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys*, 31(4):923–960.

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

Roy, R. and Shijin, S. (2018). A six-factor asset pricing model. *Borsa Istanbul Review*, 18(3):205–217.

Ruckstuhl, A., Welsh, A., and Carroll, R. J. (2000). Nonparametric function estimation of the relationship between two repeatedly measured variables. *Statistica Sinica*, 10:51–71.

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *Journal of the American Statistical Association*, 92(439):1049–1062.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

Schick, A. (1986). On asymptotically efficient estimation in semiparametric models. *The Annals of Statistics*, 14(3):1139–1151.

Schultz, T. P. (2003). Human capital, schooling and health. *Economics & Human Biology*, 1(2):207–221.

Sen, A. and Srivastava, M. S. (1975). On tests for detecting change in mean. *The Annals of Statistics*, 3(1):98–108.

Sen, P. K. (1976). A note on invariance principles for induced order statistics. *The Annals of Probability*, 4(3):474–479.

Shanken, J. (1985). Multivariate tests of the zero-beta capm. *Journal of financial economics*, 14(3):327–348.

Shanken, J. (1992). On the estimation of beta-pricing models. *The review of financial studies*, 5(1):1–33.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 683–690.

Shiller, R. J. (1984). Smoothness priors and nonlinear regression. *Journal of the American Statistical Association*, 79(387):609–615.

Shorack, G. R. (1979). The weighted empirical process of row independent random variables with arbitrary distribution functions. *Statistica Neerlandica*, 33(4):169–189.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):413–436.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society. Series B (Methodological)*, pages 111–147.

Stute, W. (1993). U-functions of concomitants of order statistics. *Probab. Math. Statist*, 14(1):143–155.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25(2):613–641.

Stute, W., Manteiga, W. G., and Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, 93(441):141–149.

Su, L. and Lu, X. (2013). Nonparametric dynamic panel data models: kernel estimation and specification testing. *Journal of Econometrics*, 176(2):112–133.

Su, L. and Ullah, A. (2007). More efficient estimation of nonparametric panel data models with random effects. *Economics Letters*, 96(3):375–380.

Su, L. and Ullah, A. (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. *Handbook of Empirical Economics and Finance. Chapman and Hall/CRC*, pages 455–497.

Su, L., Ullah, A., and Wang, Y. (2013). Nonparametric regression estimation with general parametric error covariance: a more efficient two-step estimator. *Empirical economics*, 45(2):1009–1024.

Sun, Y., Carroll, R. J., and Li, D. (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics*, 25:101–129.

Thomas, D. R. and Grunkemeier, G. L. (1975). Confidence interval estimation of survival probabilities for censored data. *Journal of the American Statistical Association*, 70(352):865–871.

Volkonskii, V. and Rozanov, Y. A. (1959). Some limit theorems for random functions. i. *Theory of Probability & Its Applications*, 4(2):178–197.

Wahba, G. (1985). A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, 13(4):1378–1402.

Wang, H.-J. and Ho, C.-W. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics*, 157(2):286–296.

Wang, K. Q. (2002). Nonparametric tests of conditional mean-variance efficiency of a benchmark portfolio. *Journal of Empirical Finance*, 9(2):133–169.

Wang, K. Q. (2003). Asset pricing with conditioning information: A new test. *The journal of finance*, 58(1):161–196.

Western, B. (1996). Vague theory and model uncertainty in macrosociology. *Sociological Methodology*, 26(1):165–192.

Wilks, S. S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Nelson Education.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93(444):1388–1402.

Xia, Y., Zhang, W., and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika*, 91(3):661–681.

Xue, L. and Zhu, L. (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association*, 102(478):642–654.

Zhang, J., Feng, S., Li, G., and Lian, H. (2011). Empirical likelihood inference for partially linear panel data models with fixed effects. *Economics Letters*, 113:165–167.

Zhang, W. and Lee, S.-Y. (2000). Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis*, 74(1):116–134.

Zhang, W., Lee, S.-Y., and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, 82(1):166–188.

Zhou, Y. and Liang, H. (2009). Statistical inference for semiparametric varying-coefficient partially linear models with error-prone linear covariates. *The Annals of Statistics*, 37(1):427–458.

Zou, C., Liu, Y., Qin, P., and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics & probability letters*, 77(4):374–382.

Proofs of Chapter 1

A.1 Proof of Theorem 1.2.1

Note that, $\mathscr{R}_w(\beta(z))$ is given by

$$\left[\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N}T_{wi}(\beta(z))\right]^{\top}\left[\tilde{D}_w(\beta(z))\right]^{-1}\left[\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N}T_{wi}(\beta(z))\right]+o_P(1),$$

as N tends to infinity (see, equation (1.12)). The proof of this result is done in three steps: first, we show the asymptotic normality of the vector $\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N}T_{wi}(\beta(z))$, second, we show the consistency of $\tilde{D}_w(\beta(z))$ and finally we use a Cramer-Wold device to close the proof.

In order to obtain the asymptotic distribution of $\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N}T_{wi}(\beta(z))$ note that

$$\begin{aligned}\frac{1}{NT|H|^{T/2}}\sum_{i=1}^{N}T_{wi}(\beta(z)) &= \frac{1}{NT|H|^{T/2}}\sum_{i=1}^{N}\left(T_{wi}(\beta(z))-E[T_{wi}(\beta(z))|\mathbb{X},\mathbb{Z}]\right)\\ &+ \frac{1}{NT|H|^{T/2}}\sum_{i=1}^{N}E\left[T_{wi}(\beta(z))|\mathbb{X},\mathbb{Z}\right]\equiv U_{1N}+U_{2N},\end{aligned}\tag{A.1}$$

where $\mathbb{X}=(X_{11},...,X_{NT})$ and $\mathbb{Z}=(Z_{11},...,Z_{NT})$ are the sample covariate values. We first work on the bias term U_{2N}. Then, substituting $T_{wi}\left(\beta(z)\right)$ by (2.8) into $E\left[T_{wi}\left(\beta(z)\right)|\mathbb{X},\mathbb{Z}\right]$, applying Assumption 1.2.2 and taking Taylor expansion around $X_{it}^{\top}m(Z_{it})-T^{-1}\sum_s X_{is}^{\top}m(Z_{is})$ we obtain

$$U_{2N}\equiv\frac{1}{NT|H|^{T/2}}\sum_{i=1}^{N}E[T_{wi}(\beta(z))|\mathbb{X},\mathbb{Z}]=\begin{pmatrix}A_{1.1N}+A_{1.2N}\\ A_{1.3N}+A_{1.4N}+A_{1.5N}\end{pmatrix}.\tag{A.2}$$

Here

$$\begin{aligned}
A_{1.1N} &= \frac{1}{2NT|H|^{T/2}}\sum_{it}\ddot{X}_{it}Q_m(z)K_H(Z_i-z),\\
A_{1.2N} &= \frac{1}{NT|H|^{T/2}}\sum_{it}\ddot{X}_{it}R_1(z)K_H(Z_i-z),\\
A_{1.3N} &= \frac{1}{2NT|H|^{T/2}}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)Q_m(z)K_H(Z_i-z),\\
A_{1.4N} &= \frac{1}{3!NT|H|^{T/2}}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)C_m(z)K_H(Z_i-z),\\
A_{1.5N} &= \frac{1}{2NT|H|^{T/2}}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)R_2(z)K_H(Z_i-z),
\end{aligned}$$

where $K_H(Z_i-z)=\prod_{l=1}^{T}K_H\left(Z_{il}-z\right)$ and

$$\begin{aligned}
Q_m(z) &= X_{it}^{\top}\otimes(Z_{it}-z)^{\top}\mathscr{H}_m(z)(Z_{it}-z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}\otimes(Z_{is}-z)^{\top}\mathscr{H}_m(z)(Z_{is}-z),\\
C_m(z) &= X_{it}^{\top}\otimes D_m^3(z,Z_{it}-z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}\otimes D_m^3(z,Z_{is}-z),\\
R_1(z) &= X_{it}^{\top}\otimes(Z_{it}-z)^{\top}\mathfrak{R}(Z_{it};z)(Z_{it}-z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}\otimes(Z_{is}-z)^{\top}\mathfrak{R}(Z_{is};z)(Z_{is}-z),\\
R_2(z) &= X_{it}^{\top}\otimes\mathfrak{R}^*(Z_{it};z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}^{\top}\otimes\mathfrak{R}^*(Z_{is};z).
\end{aligned}$$

The remainder terms in the Taylor expansion are defined as

$$\begin{aligned}
\mathfrak{R}(Z_{it};z) &= \int_0^1\left[\mathscr{H}_m(z+\omega(Z_{it}-z))-\mathscr{H}_m(z)\right](1-\omega)d\omega,\\
\mathfrak{R}^*(Z_{it};z) &= \int_0^1\left[D_m^3(z+\omega(Z_{it}-z),Z_{it}-z)-D_m^3(z,Z_{it}-z)\right](1-\omega)^2d\omega,
\end{aligned}$$

where ω is a weight function. Now we analyze the limit behavior of each of these terms when N tends to infinity and T remains fixed. First we will show that

$$A_{1.1N}=\frac{1}{2}\mu_2(K_{u_\tau})\mathscr{B}_{\ddot{X}\ddot{X}}(z,..,z)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathscr{H}_{m_r}(z)H\right\}\right\}i_d+o_p\left(\mathrm{tr}\left\{H\right\}\right),\qquad\text{(A.5)}$$

and using standard results from nonparametric regression analysis and Assumption 1.2.4 we have that

$$
\begin{aligned}
E\left(A_{1.1N}\right) &= \frac{1}{2}\int E\left[\ddot{X}_{it}X_{it}^{\top}\middle| Z_{i1}=z+H^{1/2}u_1,...,Z_{iT}=z+H^{1/2}u_T\right]\otimes(H^{1/2}u_\tau)^{\top} \\
&\quad\times\mathscr{H}_m(z)(H^{1/2}u_\tau)f(z+H^{1/2}u_1,...,z+H^{1/2}u_T)\prod_{l=1}^{T}K(u_l)du_l \\
&- \frac{1}{2T}\sum_{s=1}^{T}\int E\left[\ddot{X}_{it}X_{is}^{\top}\middle| Z_{i1}=z+H^{1/2}u_1,...,Z_{iT}=z+H^{1/2}u_T\right]\otimes(H^{1/2}u_s)^{\top} \\
&\quad\times\mathscr{H}_m(z)(H^{1/2}u_s)f(z+H^{1/2}u_1,...,z+H^{1/2}u_T)\prod_{l=1}^{T}K(u_l)du_l.
\end{aligned}
$$

Then a straightforward application of a Taylor expansion and assumptions 1.2.1 and 1.2.5 are enough to show that (A.5) holds. Also, note that to show (A.5) we need to prove that $\mathrm{Var}(A_{1.1N})\to 0$, as N tends to infinity and T is fixed. Under Assumption 1.2.1, $\mathrm{Var}(A_{1.1N})=\frac{1}{NT}\mathrm{Var}\left(a_{it}\right)+\frac{1}{NT^2}\sum_{t=3}^{T}(T-t)\mathrm{Cov}\left(a_{i2},a_{it}\right)$, where $a_{it}=\frac{1}{|H|^{T/2}}\ddot{X}_{it}Q_m(z)K_H(Z_i-z)$. Then, under assumptions 1.2.5 and 1.2.9 the first element shows the following bound $\mathrm{Var}\left(a_{it}\right)\leq\frac{C}{NT|H|^{T/2}}$ and $\mathrm{Cov}\left(a_{i2},a_{it}\right)\leq\frac{C'}{N|H|^{T/2}}$. Therefore, if $N|H|^{T/2}$ tends to infinity the variance tends to zero and applying a weak law of large numbers (A.5) follows.

Following a similar procedure, and noting that due to Assumption 1.2.4 the odd order moments of $K(.)$ disappear, it easy to show that

$$
A_{1.3N}=\frac{1}{2}\mu_2(K_{u_\tau})^2\left(1-\frac{1}{T}\right)\mathscr{D}\mathscr{B}_{X_tX_t}^{\top}(z,..,z)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathscr{H}_{m_r}(z)H^2\right\}i_d\right\}+o_p\left(\mathrm{tr}\left\{H^2\right\}\right), \tag{A.6}
$$

and

$$
A_{1.4N}=\frac{1}{3!}\left(1-\frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,..,z)\otimes\int\left(H^{1/2}u_\tau\right)D_m^3(z,H^{1/2}u_\tau)\prod_{l=1}^{T}H(u_l)du_l+o_p\left(H^2\right). \tag{A.7}
$$

Finally focusing on the residual terms $A_{1.2N}$ and $A_{1.5A}$ and using the same procedure as in the proof of (A.18)-(A.28) in Rodriguez-Poo and Soberón (2015) it can be shown that $A_{1.2N}=o_p\left(\mathrm{tr}\left\{H\right\}\right)$ and $A_{1.5N}=o_p\left(\mathrm{tr}\left\{H^2\right\}\right)$. Then by replacing the different asymptotic expressions for the A_N's into U_{2N}, we obtain,

$$
U_{2N}=\frac{1}{NT|H|^{T/2}}\sum_{i=1}^{N}E[T_{wi}(\beta(z))|\mathbb{X},\mathbb{Z}]=\begin{pmatrix}\frac{1}{2}b_{w1}(z)\\ \frac{1}{2}b_{w2}(z)+\frac{1}{3!}b_{w3}(z)\end{pmatrix}+o_p\begin{pmatrix}\mathrm{tr}\left\{H\right\}\\ \mathrm{tr}\left\{H^2\right\}\end{pmatrix}, \tag{A.8}
$$

where $b_{w1}(z)$, $b_{w2}(z)$ and $b_{w3}(z)$ were defined in (1.16).

Now we obtain the limiting distribution of the quantity $\sqrt{NT|H|^{T/2}}U_{1N}$. In order to do so we apply Liapunov's Central Limit Theorem. We do it by obtaining the variance-covariance matrix of the limiting distribution and verifying the so called Liapunov's condition. By substituting (1.7) into U_{1N} we obtain, $U_{1N} \equiv \frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} \left[T_{wi}(\beta(z)) - E\left[T_{wi}(\beta(z))|\mathbb{X},\mathbb{Z}\right]\right] = \frac{1}{NT|H|^{T/2}} \sum_{it} \tilde{Z}_{it}^{*} v_{it} K_H(Z_i - z)$. Now, because of assumptions 1.2.1 and 1.2.2 we have that

$$NT\mathrm{Var}\left(U_{1N}|\mathbb{X},\mathbb{Z}\right) = \frac{\sigma_v^2}{NT|H|^T} \sum_{it} \tilde{Z}_{it}^{*} \tilde{Z}_{it}^{*\top} K_H^2(Z_i - z). \tag{A.9}$$

Applying assumptions 1.2.1 - 1.2.2 and 1.2.4 and mimicking (A.33)-(A.35) in Rodriguez-Poo and Soberón (2015) we obtain the following,

$$NT|H|^{T/2} Var\left(U_{1N}\right) =$$
$$\sigma_v^2 \begin{pmatrix} R(K)^T \mathscr{B}_{\ddot{X}\ddot{X}}(z,\dots,z) & O_p(|H|^T) \\ O_p(|H|^T) & \left(1-\frac{1}{T}\right)\mu_2(K_{u_\tau}^2)\prod_{l\neq\tau}^{T} R(K_{u_l})\mathscr{B}_{X_tX_t}(z,\dots,z)\otimes H \end{pmatrix} (1+o_P(1)). \tag{A.10}$$

Now, we check Liapunov's condition; we must show that for any unit vector $b \in \mathbb{R}^{d(q+1)}$ and some $\delta > 0$, as N tends to infinity, $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{it} E\left[\left|b^\top \tilde{Z}_{it}^{*} v_{it} \prod_{l=1}^{T} K_H(Z_{il} - z)\right|^{2+\delta}\right] \to 0$. To prove this, let us define $\phi_{it} = |H|^{T/4} b^\top \tilde{Z}_{it}^{*} v_{it} \prod_{l=1}^{T} K_H(Z_{il} - z)\ \forall i = 1,\dots,N; t = 1,\dots,T$. Following assumption 1.2.4 we can write

$$\mathrm{Var}(\phi_{it}) =$$
$$\sigma_v^2 b^\top \begin{pmatrix} R(K)^T \mathscr{B}_{\ddot{X}\ddot{X}}(z,\dots,z) & 0 \\ 0 & \left(1-\frac{1}{T}\right)\mu_2(K_{u_\tau}^2)\prod_{l\neq\tau}^{T} R(K_{u_l})\mathscr{B}_{X_tX_t}(z,\dots,z)\otimes H \end{pmatrix} b(1+o_P(1)).$$

and $\sum_{t=1}^{T} |Cov(\phi_{i1}, \phi_{it})| = o_p(1)$. Note also that we can write $\phi_{it} = \phi_{1it} + \phi_{2it}$ where

$$\begin{aligned} \phi_{1it} &= b^\top \ddot{X}_{it} v_{it} \prod_{l=1}^{T} K_H(Z_{il} - z) \\ \phi_{2it} &= b^\top X_{it} \otimes (Z_{it} - z) - b^\top \frac{1}{T} \sum s = 1^T X_{is} \otimes (Z_{is} - z) v_{it} \prod_{l=1}^{T} K_H(Z_{il} - z). \end{aligned}$$

Furthermore, let us define $\phi_{n,i}^{*} = T^{-1/2} \sum_{t=1}^{T} \phi_{it} = T^{-1/2} \sum_{t=1}^{T} (\phi_{1it} + \phi_{2it})$. For fixed T, the $\phi_{n,i}^{*}$ are independent random variables and $n = NT$. Then, using Minkowski inequality and due to the matrix structure of $\tilde{Z}_{it}^{*}$ we get

$$E|\phi_{n,i}^{*}|^{2+\delta} \le CT^{(2+\delta)/2} E|\phi_{it}|^{2+\delta} = CT^{(2+\delta)/2} E|\phi_{1it} + \phi_{2it}|^{2+\delta}.$$

Analysing each term separately, (see, Rodriguez-Poo and Soberón (2015) for details), we obtain

$$(NT)^{-(2+\delta)/2} \sum_{i=1}^{N} E|\phi_{n,i}^{*}|^{2+\delta} \le C(N|H|^{T/2})^{-\delta/2}, \tag{A.11}$$

which tends to zero when $N|H| \to \infty$. Therefore, Liapunov's Central Limit Theorem applies and hence

$$\sqrt{NT|H|^{T/2}}U_{1N} \to_d \mathcal{N}\left(0, \upsilon_w(z)\right). \tag{A.12}$$

Therefore, by substituting (A.8) and (A.12) into (A.1) and imposing the following extra condition $\sqrt{NT|H|^{T/2}}tr(H) \to 0$ we obtain that $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_i T_{wi}\left(\beta(z)\right) \to_d \mathcal{N}\left(0, \upsilon_w(z)\right)$, as N tends to infinity.

Now we prove the consistency of $\tilde{D}_w(\beta(z))$. As N tends to infinity and T is fixed, if conditions 1.2.1 - 1.2.10 hold and, similar to the proof of (A.10), by applying a Law of Large Numbers it is straightforward to show that

$$\tilde{D}_w(\beta(z)) = \frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} T_{wi}(\beta(z))T_{wi}^{\top}(\beta(z)) = \upsilon_w(z) + o_p\left(\mathrm{tr}\{H\}\right), \tag{A.13}$$

where $\upsilon_w(z)$ was defined in (1.17). From (A.8), (A.12) and (A.13), and using the same arguments as in the proof of (2.14) in Owen (1990), we can prove that

$$\lambda = O_p\left((NT|H|^{T/2})^{-1/2}\right), \tag{A.14}$$

where λ was defined in (1.11). Then applying Taylor expansion to (1.10) and invoking (A.8), (A.12) and (A.13), we obtain

$$\mathscr{R}_w(\beta(z)) = 2\sum_{i=1}^{N}\left[T_{wi}^{\top}(\beta(z))\lambda - \left(T_{wi}^{\top}(\beta(z))\lambda\right)^2/2\right] + o_p(1). \tag{A.15}$$

By (1.11) and applying Taylor expansion again it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^{N} \frac{T_{wi}(\beta(z))}{1+\lambda^{\top}T_{wi}(\beta(z))} \\ &= \sum_{i=1}^{N} T_{wi}(\beta(z)) - \sum_{i=1}^{N} T_{wi}(\beta(z))T_{wi}^{\top}(\beta(z))\lambda + \sum_{i=1}^{N} \frac{T_{wi}(\beta(z))(T_{wi}^{\top}(\beta(z))\lambda)^2}{1+\lambda^{\top}T_{wi}(\beta(z))}. \end{aligned}$$

Then, recalling (A.8), (A.12) and (A.13) we can prove that

$$\sum_{i=1}^{N}(T_{wi}^{\top}(\beta(z))\lambda)^2 = \sum_{i=1}^{N} T_{wi}^{\top}(\beta(z))\lambda + o_p(1), \tag{A.16}$$

and

$$\lambda = \left[\sum_{i=1}^{N} T_{wi}(\beta(z)) T_{wi}^{\top}(\beta(z)) \right]^{-1} \sum_{i=1}^{N} T_{wi}(\beta(z)) + o_p\left((NT|H|^{T/2})^{-1/2} \right). \tag{A.17}$$

Now, if we rely on (A.8), (A.12) and (A.13) the proof is concluded by applying the Cramer-Wold device.

A.2 Proof of Theorem 1.2.2

This proof is similar to that of Theorem 1.2.1 and therefore most of the details are omitted. In order to obtain the asymptotic distribution of $\frac{1}{\sqrt{NT|H|}} \sum_{i=1}^{N} T_{fi}(\beta(z))$ we follow similar steps to those in (A.1),

$$\frac{1}{NT|H|} \sum_{i=1}^{N} T_{fi}(\beta(z)) = U_{1N}^* + U_{2N}^*. \tag{A.18}$$

For the bias term U_{2N}^*, defining a similar multivariate Taylor expansion around $X_{it}^{\top} m(Z_{it}) - X_{i(t-1)}^{\top} m(Z_{i(t-1)})$ as the one used in (A.2), and applying Assumption 1.2.2 we obtain

$$U_{2N}^* = \frac{1}{NT|H|} \sum_{i=1}^{N} E[T_{fi}(\beta(z))|\mathbb{X},\mathbb{Z}] = \begin{pmatrix} \frac{1}{2} b_{f1}(z) \\ \frac{1}{2} b_{f2}(z) + \frac{1}{3!} b_{f3}(z) \end{pmatrix} + o_p \begin{pmatrix} \operatorname{tr}\{H\} \\ \operatorname{tr}\{H^2\} \end{pmatrix}, \tag{A.19}$$

where $b_{f1}(z)$, $b_{f2}(z)$ and $b_{f3}(z)$ were defined in (1.21).

Now we obtain the limiting distribution of the quantity $\sqrt{NT|H|}U_{1N}^*$. By substituting (1.8) into U_{1N}^* we obtain that $U_{1N}^* = \frac{1}{NT|H|} \sum_{it} \tilde{Z}_{it} \Delta v_{it} K_H(Z_{it} - z, Z_{i(t-1)} - z)$, and taking into account that because of assumptions 1.2.1 and 1.2.2 we have that

$$E[\Delta v_{it} \Delta v_{i't'}|\mathbb{X},\mathbb{Z}] = \begin{cases} 2\sigma_v^2 & \text{if } i = i', t = t' \\ -\sigma_v^2 & \text{if } i = i', |t - t'| < 2 \\ 0 & \text{otherwise} \end{cases}. \tag{A.20}$$

Then, mimicking (A.30)-(A.36) in Rodriguez-Poo and Soberón (2014) we obtain the following,

$$NT|H|Var(U_{1N}^*) =$$
$$2\sigma_v^2 \begin{pmatrix} R(K_u)R(K_v)\mathscr{B}_{\Delta X,\Delta X}(z,z) & O_p(|H|^2) \\ O_p(|H|^2) & \mu_2(K^2)R(K_u)\left(\mathscr{B}_{XX}(z,z) + \mathscr{B}_{X_{-1}X_{-1}}(z,z)\right) \otimes H \end{pmatrix} (1 + o_P(1)) \tag{A.21}$$

The rest of the proof follows exactly the lines of the proof of Theorem 1.2.1.

A.3 Proof of Theorem 1.3.1

Note that,

$$\hat{\beta}_w(z) - \beta(z) = \left(\hat{\beta}_w(z) - E\left[\hat{\beta}_w(z)\middle|\mathbb{X},\mathbb{Z}\right]\right) + \left(E\left[\hat{\beta}_w(z)\middle|\mathbb{X},\mathbb{Z}\right] - \beta(z)\right) \equiv \mathbf{I}_{1N} + \mathbf{I}_{2N}. \quad \text{(A.22)}$$

To prove the desired result we will show that, under the conditions of this theorem, $\mathbf{I}_{2N} = B_w(z) + o_p(\frac{1}{\sqrt{NT|H|^{T/2}}})$ and $\sqrt{NT|H|^{T/2}}\mathbf{I}_{1N} \to_d \mathcal{N}(0, \Sigma_w(z))$, as N tends to infinity, where $B_w(z)$ and $\Sigma_w(z)$ have been defined in theorem 1.3.1. If we substitute (1.14) into (A.22) and we make a second order Taylor expansion around $X_{it}^\top m(Z_{it}) - \frac{1}{T}\sum_s X_{is}^\top m(Z_{is})$ we obtain that

$$\mathbf{I}_{2N} = \left(\frac{1}{NT|H|^{T/2}}\sum_{it} K_H(Z_i - z)\tilde{Z}_{it}^*\tilde{Z}_{it}^{*\top}\right)^{-1} \begin{pmatrix} A_{1.1N} + A_{1.2N} \\ A_{1.3N} + A_{1.4N} + A_{1.5N} \end{pmatrix}. \quad \text{(A.23)}$$

Note that $A_{1.1N}$, $A_{1.2N}$, $A_{1.3N}$, $A_{1.4N}$ and $A_{1.5N}$ have been already defined in the proof of Theorem 1.2.1. Furthermore, the asymptotic behavior of the second term in (A.23) has been already obtained in (A.8); therefore, all what we need to calculate the asymptotic behavior of $\mathbf{I}_{2N}$ is to study the first term. Proceeding as in Rodriguez-Poo and Soberón (2015) (see, expressions (A.8)-(A.12)), it is straightforward to show that

$$\left(\frac{1}{NT|H|^{T/2}}\sum_{it}\prod_{l=1}^{T} K_H(Z_{il} - z)\tilde{Z}_{it}^*\tilde{Z}_{it}^{*\top}\right)^{-1} = \begin{pmatrix} \mathfrak{C}_{11} & \mathfrak{C}_{12} \\ \mathfrak{C}_{21} & \mathfrak{C}_{22} \end{pmatrix}, \quad \text{(A.24)}$$

where

$$\begin{aligned}
\mathfrak{C}_{11} &= \mathscr{B}_{\ddot{X}\ddot{X}}^{-1}(z,\ldots,z) + o_p(1) \\
\mathfrak{C}_{12} &= -\mathscr{B}_{\ddot{X}\ddot{X}}^{-1}(z,\ldots,z)\left(\mathscr{D}\mathscr{B}_{\ddot{X}\ddot{X}}(z,\ldots,z)\left(I_d \otimes \mu_2(K_{u_\tau})H\right)\right) \\
&\quad \times \left(\left(1 - \frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,..,z) \otimes \mu_2(K_{u_\tau})H\right)^{-1} + o_p(1) \\
\mathfrak{C}_{21} &= -\left(\left(1 - \frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,..,z) \otimes \mu_2(K_{u_\tau})H\right)^{-1} \left(\mathscr{D}\mathscr{B}_{\ddot{X}\ddot{X}}(z,\ldots,z)\left(I_d \otimes \mu_2(K_{u_\tau})H\right)\right)^\top \\
&\quad \times \mathscr{B}_{\ddot{X}\ddot{X}}^{-1}(z,\ldots,z) + o_p(1) \\
\mathfrak{C}_{22} &= \left(\left(1 - \frac{1}{T}\right)\mathscr{B}_{X_tX_t}(z,..,z) \otimes \mu_2(K_{u_\tau})H\right)^{-1} + o_p(H^{-1}),
\end{aligned}$$

Using the terms (A.24) and (A.8) and applying Slutsky's Theorem to (A.23) we finish the proof.

In order to show the asymptotic behavior of $\mathbf{I}_{1N}$ note that by (1.14) we have that

$$\mathbf{I}_{1N} = \hat{\beta}_w(z) - E\left[\hat{\beta}_w(z)\Big|\mathbb{X},\mathbb{Z}\right] = \left(\sum_{it} K_H(Z_i - z)\tilde{Z}^*_{it}\tilde{Z}^{*\top}_{it}\right)^{-1} \sum_{it} K_H(Z_i - z)\tilde{Z}^*_{it} v_{it}, \quad \text{(A.25)}$$

and considering assumptions 1.2.1 and 1.2.2 the variance term of $\hat{\beta}_w(z)$, using the Slutsky's Theorem and previous results can be written as

$$NT|H|^{T/2}\text{Var}\left(\hat{\beta}_w(z)\Big|\mathbb{X},\mathbb{Z}\right) = \Sigma_w(z)(1 + o_p(1)); \quad \text{(A.26)}$$

the Liapunov condition needed to apply a Central Limit Theorem here is the same as the one as in the proof of Theorem 1.2.1 (see A.11) and then a further application of the Cramer-Wold device closes the proof.

A.4 Proof of Theorem 1.3.2

The proof of this theorem is similar to that of Theorem 1.3.1 and therefore the details are omitted.

In order to show Theorems 1.4.1 and 1.4.2 we need the following additional lemma that is proved at the end of this Appendix.

Lemma A.4.1. Assuming that conditions of Theorems 1.4.1 and 1.4.2 hold. Then $\hat{b}_w(z) \to^p b_w(z)$ and $\hat{b}_f(z) \to^p b_f(z)$ respectively, as N tends to infinity.

A.5 Proof of Theorem 1.4.1

If we substitute (1.12) into (1.19) we obtain that $\widetilde{\mathscr{R}}_w(\beta(z))$ is equal to

$$\widetilde{\mathscr{R}}_w(\beta(z)) = L_1(z) + L_2(z) + o_p(1)$$

where

$$\begin{aligned}
L_1(z) &= \left[\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N} T_{wi}(\beta(z)) - \sqrt{NT|H|^{T/2}}\, b_w(z)\right]^\top \tilde{D}_w^{-1}(\beta(z)) \\
&\quad \times \left[\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N} T_{wi}(\beta(z)) - \sqrt{NT|H|^{T/2}}\, b_w(z)\right] \\
L_2(z) &= \left[\sqrt{NT|H|^{T/2}}\left\{\hat{b}_w(z) - b_w(z)\right\}\right]^\top \tilde{D}_w^{-1}(\beta(z)) \\
&\quad \times \left[\sqrt{NT|H|^{T/2}}\ \left\{\hat{b}_w(z) + b_w(z)\right\} - \frac{2}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N} T_{wi}(\beta(z))\right].
\end{aligned}$$

In Theorem 1.2.1, we have already proved that

$$\frac{1}{\sqrt{NT|H|^{T/2}}}\sum_{i=1}^{N} T_{wi}(\beta(z)) \rightarrow^d \mathcal{N}\left(\sqrt{NT|H|^{T/2}}\, b_w(z), \upsilon_w(z)\right)$$

and $\tilde{D}_w(\beta(z)) \rightarrow_p v_w(z)$. Then, together with the proof of Lemma A.4.1 we can conclude that $L_1(z) \rightarrow^d \chi^2_{d(q+1)}$ and $L_2(z) \rightarrow_p 0$. Thus the proof of Theorem 1.4.1 is closed.

A.6 Proof of Theorem 1.4.2

The proof of this theorem is similar to that of Theorem 1.4.1 and therefore the details are omitted.

A.7 Proof of Lemma A.4.1

Let us consider,

$$\begin{aligned}
&\hat{b}_w(z) - b_w(z) \\
= \; & \frac{1}{NT|H|^{T/2}}\sum_{it} Z_{it}^{*}\left(X_{it}^\top\left(\hat{m}_w(Z_{it}) - m(Z_{it})\right) - \frac{1}{T}\sum_{s} X_{is}^\top\left(\hat{m}_w(Z_{is}) - m(Z_{is})\right)\right. \\
& \left. - Z_{it}^{*\top}\left(\hat{\beta}_w(z) - \beta(z)\right)\right) K_H\left(Z_i - z\right) \\
& + \frac{1}{NT|H|^{T/2}}\sum_{it} Z_{it}^{*}\left(X_{it}^\top m\left(Z_{it}\right) - \frac{1}{T}\sum_{s} X_{is}^\top m\left(Z_{is}\right) - Z_{it}^{*\top}\beta(z)\right) K_H\left(Z_i - z\right) \\
& - b_w(z) \\
= \; & L_1^*(z) + L_2^*(z).
\end{aligned}$$

Then, by Theorem 1.2.1, equation (A.8), we have that, as N tends to infinity, $L_2^*(z) \to_p 0$. Furthemore, the conditions of this Theorem guarantee that $\sup_z |\hat{m}_w(z) - m(z)| = o_p(1)$ and $\sup_z \left|\hat{\beta}_w(z) - \beta(z)\right| = o_p(1)$ (see Masry (1996), Theorem C) and jointly with assumption 1.2.9 it is easy to show that $L_1^*(z) \to_p 0$

Proofs of Chapter 2

From here on, we will be using the notation that has been defined in the previous Assumptions 2.2.1 and 2.2.2 and Theorems 2.2.1 and 2.3.1. Also, as in Feng et al. (2017), $O(1)$ denotes some constants which my be different at each appearance.

B.1 Proof of Theorem 2.2.1

Note that , using equation (2.13), the proof of this theorem is completed in three steps: first, we show the asymptotic normality of $(NT)^{-1/2}\sum_{i=1}^{N} T_i\{\beta(z)\}$, second, we show the consistency of $\tilde{D}\{\beta(z)\}$ and finally we use a Cramer-Wold device to close the proof.

In order to obtain the asymptotic distribution of $(NT)^{-1/2}\sum_{i=1}^{N} T_i\{\beta(z)\}$ note that

$$\begin{aligned} \frac{1}{NT}\sum_{i=1}^{N} T_i\{\beta(z)\} &= \frac{1}{NT}\sum_{i=1}^{N}[T_i\{\beta(z)\} - \mathrm{E}[T_i\{\beta(z)\}|\mathscr{X}]] + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{E}[T_i\{\beta(z)\}|\mathscr{X}] \\ &\equiv U_{1NT} + U_{2NT}, \end{aligned} \tag{A.27}$$

where $\mathscr{X} = \left\{(X_{js}, Z_{js})\right\}_{j=1,s=1}^{N,T}$. Also, Note that, as we already mentioned, $\gamma \to 0$ as $(N,T) \to (\infty,\infty)$; this allow us, in the same lines as Li and Racine (2007), to simplify the kernel product function as in (2.3) and using the same argument we are able to write

$$\begin{aligned} T_{it}^* = \sum_{s=1}^{T} 1_{isit} + O(||\gamma||^p) \quad &, \quad Y_{it}^* = Y_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} Y_{is}1_{isit}) + o(1) \\ X_{it}^* = X_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is}1_{isit} + o(1) \quad &, \quad v_{it}^* = v_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} v_{is}1_{isit} + o(1) \\ \rho_{it}^* = \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is}^\top\{\beta(Z_{it}) - \beta(Z_{is})\}1_{isit} + o(1) \quad &, \end{aligned} \tag{A.28}$$

where $1_{isit} = 1(Z_{is} = Z_{it})$. We first work on the bias term U_{2NT}; then, substituting $T_i(\beta(z))$ by (2.8) into U_{2NT}, applying Assumption 2.2.1.(iv) and replacing $L(Z_{it}, z, \gamma)$ with (2.3) and using (A.28), we have

$$\begin{aligned} U_{2NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\left[\tilde{X}_{it}^\top\{\beta(Z_{it}) - \beta(z)\} + \rho_{it}\right]L_{it,z,\gamma} \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^*\left[X_{it}^{*\top}\{\beta(Z_{it}) - \beta(z)\} + \rho_{it}^*\right]\left(1_{itz} + \sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right) + O_p\left(||\gamma||^2\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^*X_{it}^{*\top}\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} + O_p\left(||\gamma||^2\right) \end{aligned} \tag{A.29}$$

where $L_{it,z,\gamma} = L(Z_{it}, z, \gamma)$, $1_{itz} = 1(Z_{it} = z)$ and $1_{m,itz^*} = 1(Z_{it,m} \neq z_m) \prod_{n=1,n\neq m}^{q} 1(Z_{it,n} = z_n)$ is an indicator function which takes value 1 if Z_{it} and z differs only in their m^{th} component and 0 otherwise. Note that in the last equality, due to construction, $\{\beta(Z_{it}) - \beta(z)\}\, 1_{itz} = 0_{d\times 1}$ and $\{\beta(Z_{it}) - \beta(Z_{is})\}\, 1\,(Z_{is} = Z_{it}) = 0_{d\times 1}$; therefore, all the terms containing ρ_{it}^* vanish. We continue the analysis of (A.29); to do so, we follow Feng et al. (2017) and use Lemma A2 of Newey and Powell (2003). This lemma is a three steps process given that the cardinality of $\mathscr{D}$ is finite.

Step 1: $[0,1]^q$ is a compact subset of $\mathbb{R}^q$ with Euclidean norm $||.||$

Step 2: Rewrite (A.29) as follows

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*} X_{it}^{*\top} \{\beta(Z_{it}) - \beta(z)\} \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \left(X_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is} 1_{itis} \right) \left(X_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is} 1_{itis} \right)^{\top} \\
& \times \{\beta(Z_{it}) - \beta(z)\} \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} \\
= \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it} X_{it}^{\top} \{\beta(Z_{it}) - \beta(z)\} \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} \\
+ \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{T_{it}^*}\sum_{s_1=1}^{T} X_{is_1} 1_{itis_1} \frac{1}{T_{it}^*}\sum_{s_2=1}^{T} X_{is_2}^{\top} 1_{itis_2} \{\beta(Z_{it}) - \beta(z)\} \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} \\
- \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it} \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is}^{\top} 1_{itis} \{\beta(Z_{it}) - \beta(z)\} \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} \\
- \; & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is} 1_{itis} X_{it}^{\top} \{\beta(Z_{it}) - \beta(z)\} \sum_{m=1}^{q} \gamma_m 1_{m,itz^*},
\end{aligned} \tag{A.30}
$$

For the last two terms of (A.30), note that we can write

$$
\begin{aligned}
& \mathrm{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right. \\
& \left.-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(X_{it})X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\| \\
= & \mathrm{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{it}^{*}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\| \le \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left\|K_{it}^{*}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\| \\
\le & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\mathrm{E}\,||K_{it}^{*}||^2\,\mathrm{E}\left\|X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\|^2\right\}^{1/2} \\
\le & \left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\,||K_{it}^{*}||^2\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left\|X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\|^2\right\}^{1/2} = o_p(||\gamma||)
\end{aligned}
$$

where $K_{it}^{*}=\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}-\mu(Z_{it})$. We now obtain that for any given $z\in\mathscr{D}$ and $\gamma\in[0,1]^q$

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^{\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(Z_{it})\mu(Z_{it})^{\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+o_p(||\gamma||)
\end{aligned}
$$

Similarly, for the second term of (A.30), we have

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s_1=1}^{T}X_{is_1}1_{itis_1}\frac{1}{T_{it}^{*}}\sum_{s_2=1}^{T}X_{is_2}^{\top}1_{itis_2}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(Z_{it})\mu(Z_{it})^{\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+o_p(||\gamma||)
\end{aligned}
$$

According to all the above, we obtain

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_{it}-\mu(Z_{it})\}\{X_{it}-\mu(Z_{it})\}^{\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+o_p(||\gamma||)
\end{aligned}
$$

for any given $z \in \mathscr{D}$ and $\gamma \in [0,1]^q$. We then just need to consider

$$
\begin{aligned}
& \mathrm{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} - p(z^*)\Sigma_X(z^*)\beta(z^*)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\|^2 \\
= \; & \frac{1}{(NT)^2}\sum_{h,\ell=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} \\
& \mathrm{E}\left[\left\{X_{it,h}X_{it,\ell}\beta_h(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} - p(z^*)\Sigma_{X,h\ell}(z^*)\beta_h(z^*)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\}\right. \\
& \left.\times\left\{X_{js,h}X_{js,\ell}\beta_h(Z_{js})\sum_{m=1}^{q}\gamma_m 1_{m,jsz^*} - p(z^*)\Sigma_{X,h\ell}(z^*)\beta_h(z^*)\sum_{m=1}^{q}\gamma_m 1_{m,jsz^*}\right\}\right] \\
\leq \; & O\left(||\gamma||^2\right)\frac{1}{(NT)^2}\sum_{h,\ell=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} c_\delta \left\{\alpha_{ij}\left(|t-s|\right)\right\}^{\frac{\delta}{4+\delta}} \\
\leq \; & O\left(||\gamma||^2\right)\frac{1}{(NT)^2}\sum_{h,\ell=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T} \left\{\alpha_{ij}\left(|t-s|\right)\right\}^{\frac{\delta}{4+\delta}} = O\left(\frac{||\gamma||^2}{NT}\right) \qquad \text{(A.32)}
\end{aligned}
$$

where $c_\delta = 2^{(4+2\delta)/(4+\delta)}(4+\delta)/\delta$; the first inequality comes from using Cauchy-Schwarz inequality, and the second inequality from the fact that $1(Z_{it}=z)$ is uniformly bounded. Also, let $X_{it,h}$ be the h^{th} element of X_{it} and $\Sigma_{X,h\ell}(z^*)$ denotes the $(h,\ell)^{th}$ element of $\Sigma_X(z^*)$ for $h,\ell = 1,\ldots,d$.

Therefore, we have proved that

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
\to_P \; & p(z^*)\left\{\Sigma_X(z^*)-\mu_X(z^*)\mu_X(z^*)^{\top}\right\}\left\{\beta(z^*)-\beta(z)\right\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= \; & \Gamma_1(z^*)\left\{\beta(z^*)-\beta(z)\right\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} = b(\gamma) \qquad \text{(A.33)}
\end{aligned}
$$

for any given $z \in \mathscr{D}$ and $\gamma \in [0,1]^q$. Therefore, (A.29) has the following expression

Step 3: By **Step 2** we can write

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} = b(\gamma)+O_P\left(||\gamma||^2\right),
$$

and for any $\gamma_1, \gamma_2 \in [0,1]^q$, we have

$$||b(\gamma_1) - b(\gamma_2)|| \leq O(1)||\gamma_1 - \gamma_2||,$$

which implies the third condition of Lemma A2 of Newey and Powell (2003) holds. Therefore, we can conclude that

$$U_{2NT} = b(\gamma) + O_p\left(||\gamma||^2\right) \tag{A.34}$$

Now we obtain the limiting distribution of the quantity $\sqrt{NT}U_{1NT}$. By substituting (2.8) into U_{1NT} and replacing $L(Z_{it}, z, \gamma)$ with (2.3) we obtain

$$\begin{aligned} U_{1NT} &= \frac{1}{NT}\sum_{i=1}^{N}\left[T_i\{\beta(z)\} - \mathrm{E}\left[T_i\{\beta(z)\}|\mathscr{X}\right]\right] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}v_{it}^{*}\left(1_{itz} + \sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right) + O_p\left(||\gamma||^2\right), \end{aligned} \tag{A.35}$$

therefore, we first focus on the analysis of $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}v_{it}^{*}1(Z_{it}=z)$, then

$$\begin{aligned} &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}v_{it}^{*}1(Z_{it}=z) \\ =\ &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(v_{it} - \frac{1}{T_{it}^{*}}\sum_{s=1}^{T}v_{is}1_{is,it}\right)1(Z_{it}=z). \end{aligned} \tag{A.36}$$

Applying **Step 2**, we can write the leading term of $\sqrt{NT}U_{1NT}$ as

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}v_{it}^{*}1(Z_{it}=z) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_{it} - \mu_X(z)\}v_{it}1(Z_{it}=z) + o_P\left(1 + ||\gamma||^2\right), \tag{A.37}$$

then we will focus on $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_{it} - \mu_X(z)\}v_{it}1(Z_{it}=z)$. For notational simplicity, denote

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_{it} - \mu_X(z)\}v_{it}1(Z_{it}=z) = \sum_{t=1}^{T}V_{T,N}(t). \tag{A.38}$$

By Assumption 2.2.1.(ii) and construction, $V_{T,N}(t)$ is stationary and α-mixing. Thus, the large-block and small-block technique can be applied in order to prove the normality below (see, Lemma A.1 in Gao (2007), Theorem 2.21 in Fan and Yao (2003) and lemma A.1 in Chen et al. (2012a)).To employ this technique, we partition the set $\{1,\ldots,T\}$ into $2k_T+1$ subsets with large blocks of size ℓ_T, small blocks of size s_T and the remaining set of size

$T - k_T(\ell_t + s_T)$, where ℓ_T and s_T are selected such that

$$s_T \to \infty, \quad s_T/\ell_T \to 0 \quad \ell_T/T \to 0, \quad \text{and} \quad k_T \equiv \{T/(\ell_T + s_T)\} = O(s_T).$$

For instance, for any $\phi > 2$, $\ell_T = T^{\frac{\phi-1}{\phi}}$, $s_T = T^{\frac{1}{\phi}}$; thus $k_T = O(T^{\frac{1}{\phi}}) = O(s_T)$. For $n = 1, \ldots, k_T$ define

$$\tilde{V}_n = \sum_{t=(n-1)(\ell_T+s_T)+1}^{n\ell_T+(n-1)s_T} V_{T,N}(t) \quad , \quad \bar{V}_n = \sum_{t=n\ell_T+(n-1)s_T+1}^{n(\ell_T+s_T)} V_{T,N}(t) \ , \ \hat{V} = \sum_{t=k_T(\ell_T+s_T)+1}^{T} V_{T,N}(t).$$

Besides, note that $\alpha(T) = o(T^{-1})$ and $k_T s_T/T \to 0$; then, by the properties of α-mixing and using similar techniques as the used in the previous results, we obtain that $E\left|\left|\sum_{n=1}^{k_T} \bar{V}_n\right|\right|^2 = O\left(\frac{k_T s_T}{T}\right) = o(1)$, and $E\left|\left|\hat{V}\right|\right|^2 = O\left(\frac{T-k_T\ell_T}{T}\right) = o(1)$. Therefore, we just need to focus the analysis on $\sum_{n=1}^{k_T} \tilde{V}_n$. Using the Feller-Lindeberg central limit theorem, we first need to show that $\{\tilde{V}_n\}_{n=1}^{k_T}$ are independent for each n. By Proposition 2.6 in Fan and Yao (2003) and the condition of α-mixing coefficients, we have

$$\left|\mathrm{E}\left[\exp\left\{\sum_{n=1}^{k_T} ||\tilde{V}_n||\right\}\right] - \prod_{n=1}^{k_T} \mathrm{E}\left[\exp\left\{||\tilde{V}_n||\right\}\right]\right| \leq C(k_T - 1)\alpha(s_T) \to 0, \tag{A.40}$$

where C is a constant and $\alpha(.)$ is the upper bounded of the α-mixing coefficient defined in Assumption 2.2.1.(ii). This upper bound is achievable in the same way as Assumption A.4 of Chen et al. (2012a). Therefore we obtain that $\{\tilde{V}_n\}_{n=1}^{k_T}$ are asymptotically independent. Furthermore, as in the proof of Theorem 2.210(ii) in Fan and Yao (2003), we have to show finite variance (Feller condition)

$$\begin{aligned}
\mathrm{Cov}\left[\tilde{V}_1\right] &= \mathrm{Cov}\left[\sum_{t=1}^{\ell_T} V_{N,T}(t)\right] = \mathrm{Cov}\left[\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{\ell_T} \{X_{it} - \mu_X(z)\} v_{it} 1(Z_{it} = z)\right] \\
&= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{\ell_T} \mathrm{Cov}\left[\{X_{it} - \mu_X(z)\} v_{it} 1(Z_{it} = z)\right] \\
&= \frac{\ell_T}{T} \Gamma_0(z) \{I_d + o(1)\}, \qquad (A.41)
\end{aligned}$$

which implies that

$$\sum_{n=1}^{k_T} \mathrm{Cov}\left(\tilde{V}_n\right) = k_T \mathrm{Cov}\left(\tilde{V}_1\right) = \frac{k_T \ell_T}{T} \Gamma_0(z) \{I_d + o(1)\} \to \Gamma_0(z). \tag{A.42}$$

As a result, the Feller condition is satisfied. Now we just need to check the Lindeberg condition

$$\sum_{n=1}^{k_T} \mathrm{E}\left[||\tilde{V}_n||^2 I\{||\tilde{V}_n|| \geq \varepsilon\}\right] \to^P 0 \tag{A.43}$$

where $\varepsilon > 0$. Using Cauchy-Schwarz inequality, we have

$$\begin{aligned} \mathrm{E}\left[||\tilde{V}_n||^2 I\{||\tilde{V}_n|| \geq \varepsilon\}\right] &\leq \left\{\mathrm{E}||\tilde{V}_n||^3\right\}^{2/3} \left\{\Pr\left(||\tilde{V}_n|| \geq \varepsilon\right)\right\}^{1/3} \\ &\leq C\left\{\mathrm{E}||\tilde{V}_n||^3\right\}^{2/3} \left\{\mathrm{E}||\tilde{V}_n||^2\right\}^{1/3}, \end{aligned} \tag{A.44}$$

and by Lemma B.2 in Chen et al. (2012a)

$$\mathrm{E}||\tilde{V}_n||^3 \leq \left(\frac{\ell_T}{T}\right)^{3/2} \left[\mathrm{E}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{X_{i1} - \mu_X(z)\} v_{i1} 1(Z_{i1} = z)\right|\right|^4\right]^{3/4} < \infty, \tag{A.45}$$

$$\mathrm{E}||\tilde{V}_n||^2 \leq \left(\frac{\ell_T}{T}\right) \left[\mathrm{E}\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{X_{i1} - \mu_X(z)\} v_{i1} 1(Z_{i1} = z)\right|\right|^4\right]^{1/2} < \infty. \tag{A.46}$$

Thus, $\mathrm{E}||\tilde{V}_n||^3 = O\left\{\left(\frac{\ell_T}{T}\right)^{3/2}\right\}$ and , $\mathrm{E}||\tilde{V}_n||^2 = O\left(\frac{\ell_T}{T}\right)$ which, using (A.44), implies

$$\mathrm{E}\left[||\tilde{V}_n||^2 I\{||\tilde{V}_n|| \geq \varepsilon\}\right] \leq O\left\{\left(\frac{\ell_T}{T}\right)^{4/3}\right\} = o\left(\frac{\ell_T}{T}\right), \tag{A.47}$$

therefore,

$$\sum_{n=1}^{k_T} \mathrm{E}\left[||\tilde{V}_n||^2 I\{||\tilde{V}_n|| \geq \varepsilon\}\right] = o\left(\frac{k_T \ell_T}{T}\right) = o(1). \tag{A.48}$$

Consequently, the Lindeberg condition is satisfied; using (A.33), (A.40), (A.42) and (A.48) it is east to see that if $\gamma_m \to 0$ we can conclude that

$$\sqrt{NT} U_{1NT} \to^d \mathscr{N}\left\{0_{d\times 1}, \Gamma_0(z)\right\}. \tag{A.49}$$

as N and T tend to infinity.

Now we prove the consistency of $\tilde{D}\{\beta(z)\}$. Similar to the proof of (A.38) - (A.42), it is straightforward to show that

$$\tilde{D}\{\beta(z)\} = \frac{1}{NT}\sum_{i=1}^{N} T_i\{\beta(z)\} T_i^{\top}\{\beta(z)\} = \Gamma_0(z)\left\{I_d + o_p(1)\right\}. \tag{A.50}$$

From (A.33), (A.49) and (A.50), and using the same arguments as in the proof of (2.14) in Owen (1990), we can prove that

$$\lambda = O_p\left(\frac{1}{\sqrt{NT}}\right), \tag{A.51}$$

where λ was defined in (2.12). Then applying Taylor expansion to (2.11) and invoking (A.33), (A.49) and (A.50), we obtain

$$\mathscr{R}(\beta(z)) = 2\sum_{i=1}^{N}\left[T_i^{\top}\{\beta(z)\}\lambda - \left[T_i^{\top}\{\beta(z)\}\lambda\right]^2/2\right] + o_p(1). \tag{A.52}$$

By (2.12) and applying Taylor expansion again it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^{N}\frac{T_i\{\beta(z)\}}{1+\lambda^{\top}T_i\{\beta(z)\}} \\ &= \sum_{i=1}^{N}T_i\{\beta(z)\} - \sum_{i=1}^{N}T_i\{\beta(z)\}T_i^{\top}\{\beta(z)\}\lambda + \sum_{i=1}^{N}\frac{T_i\{\beta(z)\}\left[T_i^{\top}\{\beta(z)\}\lambda\right]^2}{1+T_i^{\top}\{\beta(z)\}} \end{aligned}$$

Then, recalling (A.33), (A.49) and (A.50) we can prove that

$$\sum_{i=1}^{N}\left[T_i^{\top}\{\beta(z)\}\lambda\right]^2 = \sum_{i=1}^{N}T_i^{\top}\{\beta(z)\}\lambda + o_p(1), \tag{A.53}$$

and

$$\lambda = \left[\sum_{i=1}^{N}T_i\{\beta(z)\}T_i^{\top}\{\beta(z)\}\right]^{-1}\sum_{i=1}^{N}T_i\{\beta(z)\} + o_p\left((NT)^{-1/2}\right). \tag{A.54}$$

Now, if we rely on (A.33), (A.49) and (A.50) the proof is concluded by applying the Cramer-Wold device.

B.2 Proof of Theorem 2.3.1

Note that, without loss of generality, we are able to write

$$\hat{\beta}(z)-\beta(z)=\left[\hat{\beta}(z)-\mathrm{E}\left\{\hat{\beta}(z)\middle|\mathscr{X}\right\}\right]+\left[\mathrm{E}\left\{\hat{\beta}(z)\middle|\mathscr{X}\right\}-\beta(z)\right]\equiv\mathbf{I}_{1NT}+\mathbf{I}_{2NT} \quad (A.55)$$

To prove the desired result, under assumption 2.2.1, we will show first that $\mathbf{I}_{2NT}=\Gamma^{-1}(z)b(\gamma)$ and second that $\sqrt{NT}\mathbf{I}_{1NT}\rightarrow_d \mathscr{N}\left\{0_{d\times 1},\Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)\right\}$, as (N,T) tend to infinity jointly and $\gamma_s\rightarrow 0$. If we substitute (2.18) into (A.55) we obtain

$$\begin{aligned}\mathbf{I}_{2NT} &= \mathrm{E}\left\{\hat{\beta}(z)\middle|\mathscr{X}\right\}-\beta(z) \quad (A.56)\\ &= \left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\right\}^{-1}\left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\left[\tilde{X}_{it}^{\top}\beta(Z_{it})+\rho_{it}-\beta(z)\right]L(Z_{it},z,\gamma)\right\}.\end{aligned}$$

We begin the analysis with the inverse term of (A.56) and by replacing $L(Z_{it},z,\gamma)$ with (2.3) and using (A.32)-(A.33) we obtain

$$\begin{aligned}\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma) &= \frac{1}{NT}\sum_{it}X_{it}^{*}X_{it}^{*\top}1_{itz}+O_p(||\gamma||)\\ &\rightarrow_P p(z)\left\{\Sigma_X(z)-\mu_X(z)\mu_X(z)^{\top}\right\}+O_p(||\gamma||). \quad (A.57)\end{aligned}$$

Then, using (A.57) we have proved that

$$\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\rightarrow_P p(z)\left\{\Sigma_X(z)-\mu_X(z)\mu_X(z)^{\top}\right\}+O_p(||\gamma||)=\Gamma_1(z). \quad (A.58)$$

Now we continue with the second term of (A.56) and using (A.29)-(A.34) we obtain that

$$\begin{aligned}\frac{1}{NT}\sum_{it}\tilde{X}_{it}\left\{\tilde{X}_{it}^{\top}\beta(Z_{it})+\rho_{it}-\beta(z)\right\}L(Z_{it},z,\gamma) &\rightarrow_P \Gamma_1(z^*)\left\{\beta(z^*)-\beta(z)\right\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\\ &\quad +O_p\left(||\gamma||^2\right)\\ &= b(\gamma). \quad (A.59)\end{aligned}$$

In order to show the asymptotic behavior of $\mathbf{I}_{1N}$ note that by (2.18) we have that

$$\mathbf{I}_{1NT}=\hat{\beta}(z)-\mathrm{E}\left\{\hat{\beta}(z)\middle|\mathbb{X},\mathbb{Z}\right\}=\left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\right\}^{-1}\left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{v}_{it}L(Z_{it},z,\gamma)\right\}, \quad (A.60)$$

where the inverse term was already study (see, (A.58)); therefore, we will study the asymptotic behaviour of (A.60) by studying the behaviour of the second term. Based on the results obtained in (A.35) - (A.48) in the proof of theorem 2.2.1 and (A.58) the following result holds

$$\sqrt{NT}\mathbf{I}_{1NT} \to_d \mathscr{N}\left\{0_{d\times 1}, \Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)\right\},$$

and the proof is closed.

B.3 Proof of Corollary 2.2.1

From equation (2.17) we know that the auxiliary random vector $T_i\{\beta(z)\} = \tilde{T}_i\{\beta(z)\}$, where

$$\tilde{T}_i\{\beta(z)\} = \sum_{t=1}^{T}\tilde{X}_{it}\left\{\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\beta(z)\right\}1(Z_{it} = z) + O_P\left(\frac{1}{NT}\right).$$

Then, the proof of Corollary 2.2.1 is similar to the proof of Theorem 2.2.1 but setting $\gamma_m = 0$ for $m = 1, \ldots, q$.

B.4 Proof of Corollary 2.3.2

From equation (2.19) we now that $\sqrt{NT}\left\{\hat{\beta}(z) - \beta(z)\right\}$ can be rewrite as

$$\sqrt{NT}\left\{\hat{\beta}(z) - \beta(z)\right\} = \sqrt{NT}\left\{\tilde{\beta}(z) - \beta(z)\right\} + O_P\left(\frac{1}{\sqrt{NT}}\right);$$

where $\tilde{\beta}(z)$ is a frequency estimator in the same way as in $\hat{\beta}(z)$ when $\gamma_m = 0\ \forall m = 1, \ldots q$. Then, the proof of Corollary 2.3.2 is similar to the proof of Theorem 2.3.1 but setting $\gamma_m = 0$ for $m = 1, \ldots, q$.

Proofs of Chapter 3

C.1 Proof of Proposition 3.2.1

Equation (3.8) follows from Davydov and Egorov (2000) Theorem 1. Recall that Z is distributed as an $U(0,1)$. A typical uniformity argument shows that $\sup_{u\in[0,1]}\left\|\left(\tilde{M}_{jn}-M_j\right)(u)\right\| = o_p(1)$ $a.s.$ with

$$\tilde{M}_{jn}(u) = n^{-1}\sum_{i=1}^{n} X_{ji}X_{ji}^{\mathrm{T}}1_{\{Z_i\leq u\}}, j=1,2.$$

Then (3.9) follows by noticing that $\hat{M}_{jn}(u) = \tilde{M}_{jn}(Z_{n:\lfloor nu\rfloor})$ and that $\sup_{u\in[0,1]}\left|Z_{n:\lfloor nu\rfloor}-u\right| = o(1)$ $a.s.$ by Glivenko-Cantelli theorem.

C.2 Proof of Proposition 3.2.2

Define $X_i(u) = \left[X_{1i}^{\mathrm{T}}(u), X_{1i}^{\mathrm{T}}(1)-X_{1i}^{\mathrm{T}}(u), X_{2i}^{\mathrm{T}}\right]^{\mathrm{T}}$, with $X_{1i}(u) = X_{1i}(1)1_{\{Z_i\leq u\}}$, and $\bar{\theta}_0 = \left(\bar{\beta}_0^{\mathrm{T}}, \bar{\beta}_0^{\mathrm{T}}, \bar{\delta}_0^{\mathrm{T}}\right)^{\mathrm{T}}$. The result is immediate from Proposition 3.2.1 assuming 3.2.3 and 3.2.3, after showing that

$$\sup_{\frac{K}{n}+\varepsilon\leq u\leq\frac{K}{n}-\varepsilon}\left\|\left(\hat{\Omega}_n-\hat{\Omega}_n^0\right)(u)\right\| = o_p(1),$$

with

$$\hat{\Omega}_n^0(u) = \frac{1}{n}\sum_{i=1}^{n} X_i(u)X_i^{\mathrm{T}}(u)V_i^2.$$

Under $\bar{H}_0$, $\hat{\theta}_n(1) = \bar{\theta}_0 + O_p(n^{-1/2})$ and

$$\begin{aligned}\hat{\Omega}_n(u)-\hat{\Omega}_n^0(u) &= \frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor}\left(X_{[i:n]}^{\mathrm{T}}(u)\left(\hat{\theta}_n(1)-\bar{\theta}_0\right)\right)^2 X_{[i:n]}X_{[i:n]}^{\mathrm{T}} \\ &\quad -2\left(\hat{\theta}_n(1)-\bar{\theta}_0\right)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{\lfloor nu\rfloor} X_{[i:n]}X_{[i:n]}X_{[i:n]}^{\mathrm{T}}V_{[i:n]}.\end{aligned}$$

By Proposition 3.2.1, $\left\|\hat{\theta}_n(1)-\bar{\theta}_0\right\| = O_p\left(n^{-1/2}\right)$. Then, by assumption 3.2.3 and 3.2.4, after applying Hölder's inequality,

$$\sup_{\frac{K}{n}+\varepsilon\leq u\leq\frac{K}{n}-\varepsilon}\left\|\left(\hat{\Omega}_n-\hat{\Omega}_n^0\right)(u)\right\| = O_p\left(n^{-1}\right)+O_p\left(n^{-1/2}\right).$$

C.3 Proof of Proposition 3.2.3

First notice that because Z is independent of (U,X), the concomitants $\left\{U_{[i:n]},X_{[i:n]}\right\}_{i=1}^{n}$ are i.i.d. as (U,X). Define

$$\tilde{\varphi}_n^{\dagger}=n\max_{K\le\ell\le n-K}\tilde{\alpha}_n^{\dagger}\left(\frac{\ell}{n}\right)$$

with

$$\begin{aligned}\hat{\eta}_n^{\dagger}(u) &= M_{11}^{-1}(1)\frac{\hat{N}_{1n}(u)-u\hat{N}_{1n}(1)}{u(1-u)},\\ \tilde{\alpha}_n^{\dagger}(u) &= n\cdot\hat{\eta}_n^{\dagger\mathrm{T}}(u)\frac{M_{11}(1)\,u(1-u)}{\sigma^2}\hat{\eta}_n^{\dagger}(u)\end{aligned}$$

Applying the extension of Darling-Erdős' theorem (Darling et al. (1956)) to the vector case, as in Horváth (1993)

$$a(\log n)\sqrt{\tilde{\varphi}_n^{\dagger}}-b_{1+k_1}(\log n)\to_d E.$$

Therefore, it suffices to prove that

$$\sup_{K\le\ell\le n-K}\left|\left(\tilde{\varphi}_n^{(0)}-\tilde{\varphi}_n^{\dagger}\right)\left(\frac{\ell}{n}\right)\right|=o_p(1). \qquad \text{(A.61)}$$

To this end, first, apply the Marcinkiewicz-Zygmund strong law of large numbers (Chow and Teicher (1988), pp 125), to establish that, for the $\delta>0$ in assumption 3.2.6,

$$\left[n\hat{M}_{11}\left(\frac{\ell}{n}\right)-\ell M_{11}(1)\right]=o\left(\ell^{2/(2+\delta)}\right)\text{ a.s. as }\ell\to\infty.$$

$$\begin{aligned}\left[n\left(\hat{M}_{11}(1)-\hat{M}_{11}\left(\frac{\ell}{n}\right)\right)-(n-\ell)M_{11}(1)\right] &= o\left((n-\ell)^{2/(2+\delta)}\right)\text{ a.s. as }\ell\to\infty,\\ \sum_{i=1}^{n}V_i^2-n\sigma^2 &= o\left(n^{2/(2+\delta)}\right)\text{ a.s. as }n\to\infty. \qquad \text{(A.62)}\end{aligned}$$

Hence,

$$\max_{K\le\ell\le n}\left\|\frac{\ell}{n}M_{11}(1)-\hat{M}_{11n}\left(\frac{\ell}{n}\right)\right\| = o_p\left(n^{-\frac{\delta}{(2+\delta)}}\right), \qquad \text{(A.63)}$$

$$\max_{1\le\ell\le n-K}\left\|\frac{n-\ell}{n}M_{11}(1)-\left(\hat{M}_{11n}(1)-\hat{M}_{11n}\left(\frac{\ell}{n}\right)\right)\right\| \;=\; o_p\left(n^{-\frac{\delta}{(2+\delta)}}\right). \quad \text{(A.64)}$$

By assumption 3.2.2,

$$\max_{K\le\ell\le 1}\left\|\frac{\ell}{n}\hat{M}_{11n}^{-1}\left(\frac{\ell}{n}\right)\right\| \;=\; O_p(1), \quad \text{(A.65)}$$

$$\max_{1\le\ell\le n-K}\left\|\frac{n-\ell}{n}\left[\hat{M}_{11n}(1)-\hat{M}_{11n}\left(\frac{\ell}{n}\right)\right]^{-1}\right\| \;=\; O_p(1). \quad \text{(A.66)}$$

Also, by the law of the iterated logarithm for partial sums,

$$\max_{1\le\ell\le n}\left\|\frac{n}{\sqrt{\ell}}\hat{N}_{1n}\left(\frac{\ell}{n}\right)\right\| = O_p\left(\sqrt{\log\log n}\right) \quad \text{(A.67)}$$

and by (A.62), (A.65) and (A.67)

$$\hat{\sigma}_n^2-\sigma^2=\frac{1}{n}\sum_{i=1}^{n}\left(V_i^2-\sigma^2\right)-\hat{S}_n^{\mathrm{T}}(1)\hat{M}_{11n}^{-1}(1)\hat{S}_n(1)=o_p\left(n^{-\delta/(2+\delta)}\right)+O_p\left(\frac{\log\log n}{n}\right).$$

Notice that,

$$\begin{aligned} n\cdot\left(\tilde{\alpha}_n-\tilde{\alpha}_n^{\dagger}\right)\left(\frac{\ell}{n}\right) \;&=\; \left[\sqrt{\frac{\ell(n-\ell)}{n}}\left(\hat{\eta}_n-\hat{\eta}_n^{\dagger}\right)\left(\frac{\ell}{n}\right)\right]^{\mathrm{T}}\left(\frac{M_{11}}{\sigma^2+o_p\left(n^{-\frac{\delta}{2+\delta}}\right)}\right) \\ &\quad\times\left[\sqrt{\frac{\ell(n-\ell)}{n}}\left(\hat{\eta}_n-\hat{\eta}_n^{\dagger}\right)\left(\frac{\ell}{n}\right)\right], \end{aligned}$$

where

$$\begin{aligned} \sqrt{\frac{\ell(n-\ell)}{n}}\left(\hat{\eta}_n-\hat{\eta}_n^{\dagger}\right)\left(\frac{\ell}{n}\right) \;&=\; \sqrt{\frac{\ell(n-\ell)}{n}}\left\{\left[\hat{M}_{11n}^{-1}\left(\frac{\ell}{n}\right)-\frac{n}{\ell}M_{11}^{-1}(1)\right]\hat{N}_{1n}\left(\frac{\ell}{n}\right)\right. \\ &\quad-\left[\left(\hat{M}_{11n}(1)-\hat{M}_{11n}\left(\frac{\ell}{n}\right)\right)^{-1}-\frac{n}{n-\ell}M_{11}^{-1}(1)\right] \\ &\quad\left.\times\left[\hat{N}_{1n}(1)-\hat{N}_{1n}\left(\frac{\ell}{n}\right)\right]\right\} \end{aligned}$$

$$= \sqrt{\frac{(n-\ell)\,\ell^2}{n^3}} \left\{ \frac{n}{\ell} \hat{M}_{11n}^{-1}\left(\frac{\ell}{n}\right) \left[\frac{\ell}{n} M_{11}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right)\right] M_{11}^{-1}(1) \frac{n}{\sqrt{\ell}} \hat{N}_{1n}\left(\frac{\ell}{n}\right) \right\}$$

$$-\sqrt{\frac{\ell\,(n-\ell)^2}{n^3}} \left\{ \frac{n}{n-\ell} \left(\hat{M}_{11n}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right)\right)^{-1} \right.$$
$$\times \left[\frac{n-\ell}{n} M_{11}(1) - \left(\hat{M}_{11n}(1) - \hat{M}_{11n}\left(\frac{\ell}{n}\right)\right)\right]$$
$$\left. \times M_{11}^{-1}(1) \frac{n}{\sqrt{n-\ell}} \left[\hat{N}_{1n}(1) - \hat{N}_{1n}\left(\frac{\ell}{n}\right)\right] \right\}$$

Therefore, by (A.63)-(A.67),

$$\begin{aligned} \sqrt{\frac{\ell\,(n-\ell)}{n}} \max_{K\le \ell \le n-K} \left\| \left(\hat{\eta}_n - \hat{\eta}_n^{\dagger}\right)\left(\frac{\ell}{n}\right) \right\| &\le O_p(1) \times o_p\left(\frac{1}{n^{\frac{\delta}{(2+\delta)}}}\right) \times O_p\left(\sqrt{\log\log n}\right) \\ &= o_p(1), \end{aligned} \tag{A.68}$$

which proves (A.61).

Finally, (3.13) and (3.14) follow by $(A.68)$ and

$$\left\{u(1-u)M^{1/2}(1)\hat{\eta}_n^{\dagger}(u)/\sigma\right\}_{u\in[0,1]} \to_d \{W_0(u) - uW_0(1)\}_{u\in[0,1]} \ \textit{in}\ D[0,1],$$

by Proposition 3.2.1.

C.4 Proof of Proposition 3.2.4

First notice that, by Propositions 3.2.1 and 3.2.2, uniformly in $u \in [\varepsilon, 1-\varepsilon]$,

$$\frac{\tilde{\varphi}_n}{n} \to_p = \sup_{\varepsilon \le u \le 1-\varepsilon} \eta_0^{\mathrm{T}}(u)\Sigma_0^{-1}(u)\eta_0(u),$$

which proves (3.15). In order to prove, (3.16), notice that under $\bar{H}_{1n}$

$$\sqrt{n}\left(\hat{\theta}_n - \bar{\theta}_0\right)(u) = \hat{M}_n^{-1}(u)\left[\frac{1}{n}\sum_{i=1}^{\lfloor nu \rfloor} X_{[i:n]}\tau(Z_{i:n}) + \sqrt{n}\hat{N}_n(u)\right],$$

where, under assumption 3.2.7, $\sup_{\varepsilon\le u\le 1-\varepsilon}\left\|n^{-1}\sum_{i=1}^{\lfloor nu\rfloor} X_{[i:n]}\tau(Z_{i:n}) - T(u)\right\| = o(1)$ *a.s.* using same arguments to prove (3.8) in Proposition 3.2.1. Then, apply (3.8), (3.9) and the continuous mapping Theorem to complete the proof.

C.5 Proof of Proposition 3.2.5

Let $\mathbb{P}_\xi$ be the induced probability distribution of ξ. It suffices to show that for any $c > 0$,

$$\mathbb{P}_\xi\left(\hat{\varphi}^*_{n\varepsilon} \leq c\right) \to \mathbb{P}\left(\varphi_{\infty\varepsilon} \leq c\right) + o(1)\ a.s., \tag{A.69}$$

Notice that uniformly in $u \in [0,1]$,

$$\hat{\eta}^*_n(u) = R\left[M(u) + o(1)\right]^{-1} \hat{N}^*_n(u) \text{ a.s.,}$$

Mimicking the strategy of proof in Stute et al. (1998) (SGQ) for a similar bootstrap statistics, (A.69) follows by showing that conditional to the sample, $\sqrt{n}\hat{N}^*_n$ converges in distribution to N_∞ a.s., i.e. for almost all sample $\{Y_i, X_{1i}, X_{2i}, Z_i\}_{i=1}^n$, by showing the convergence of the finite dimensional distributions (fidis) and tightness. Henceforth, $\mathbb{E}_\xi$ is the expectation operator corresponding to $\mathbb{P}_\xi$. For fidis convergence, first notice that for $u_1, u_2 \in [0,1]$,

$$\begin{aligned} n\mathbb{E}_\xi\left[\hat{N}^*_n(u_1)\hat{N}^{*\mathrm{T}}_n(u_2)\right] &= \frac{1}{n}\sum_{i=1}^{\lfloor n(u_1\wedge u_2)\rfloor} X_{[i:n]}X^{\mathrm{T}}_{[i:n]}\hat{V}^2_{[i:n]} \\ &= \frac{1}{n}\sum_{i=1}^{n} X_iX_i^{\mathrm{T}}V_i^2 1_{\left\{Z_i\leq Z_{\lfloor n(u_1\wedge u_2)\rfloor:n}\right\}} \\ &\quad +\frac{1}{n}\sum_{i=1}^{n}\left[X_i^{\mathrm{T}}\left(\hat{\theta}_n(1)-\bar{\theta}_0\right)\right]^2 X_iX_i^{\mathrm{T}}V_i^2 1_{\left\{Z_i\leq Z_{\lfloor n(u_1\wedge u_2)\rfloor:n}\right\}} \\ &= \Omega_0(u_1\wedge u_2) + o(1)\ a.s. \end{aligned} \tag{A.70}$$

since $\hat{\theta}_n(1) = \bar{\theta}_n + o(1)\ a.s.$, and applying the arguments for proving (3.9) using assumption 3.2.3 and 3.2.4. Then, fixing $u_1, ..., u_q$, by the Cramér-Wold device, it suffices to show that for any $c < \infty$,

$$\mathbb{P}_\xi\left\{\sqrt{n}\sum_{j=1}^{q} a_j b^{\mathrm{T}}\hat{N}^*_n(u_j) \leq c\right\} \to_p \mathbb{P}\left\{\sum_{j=1}^{q} a_j b^{\mathrm{T}} N_\infty(u_j) \leq c\right\}, \tag{A.71}$$

for any $b = (b_1, ..., b_{k+1})^{\mathrm{T}}$ and $a = (a_1, ..., a_p)^{\mathrm{T}}$. Write $W_i = \sum_{j=1}^q a_j b^{\mathrm{T}} X_i 1_{\left\{Z_i\leq Z_{\lfloor nu_j\rfloor:n}\right\}}$

$$\begin{aligned} \sqrt{n}\sum_{j=1}^{q} a_j b^{\mathrm{T}}\hat{N}^*_n(u_j) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i\hat{V}_i\xi_i \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} W_i\hat{V}_i\xi_i. \end{aligned}$$

Then (A.71) follows by checking the Linderberg's condition

$$L_n(\delta) = \frac{1}{n}\sum_{i=1}^{n} W_i^2 \int_{\left\{|W_i\hat{V}_i\xi_i| \geq \delta\sqrt{n}\right\}} W_i^2\hat{V}_i^2\xi_i^2 d\mathbb{P}_\xi \to 0 \; a.s.$$

Define $\bar{W}_i = \sum_{j=1}^{q} a_j b^{\mathrm{T}} X_i$. Since $|\xi| \leq \tau$,

$$\begin{aligned} L_n(\delta) &\leq \frac{\kappa^2}{n}\sum_{i=1}^{n} 1_{\left\{|\bar{W}_i\hat{V}_i| \geq \frac{\delta\sqrt{n}}{\tau}\right\}} \bar{W}_i^2\hat{V}_i^2 \\ &= \frac{\kappa^2}{n}\sum_{i=1}^{n} 1_{\left\{|\bar{W}_i V_i| \geq \frac{\delta\sqrt{n}}{\tau}\right\}} \bar{W}_i^2 V_i^2 + o(1) \; a.s. \\ &= o(1) \; a.s. \end{aligned}$$

using arguments in (A.70). In order to show tightness, it suffices to check Billingsley (1968) Theorem 15.7 as in SGQ Lemma A3. Define

$$\hat{\alpha}_{nb}^*(u) = \sqrt{n} b^{\mathrm{T}} \hat{N}_n^*(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} b^{\mathrm{T}} X_i 1_{\left\{Z_i \leq Z_{\lfloor nu \rfloor :n}\right\}} \hat{V}_i \xi_i.$$

We must show, as in SGQ Lemma 3, that for any $b \in \mathbb{R}^{1+k}$ and $0 \leq u_0 \leq u_1 \leq u_2 \leq 1$,

$$\mathbb{E}_\xi \left[\hat{\alpha}_{nb}^*(u_1) - \hat{\alpha}_{nb}^*(u_0)\right]^2 \left[\hat{\alpha}_{nb}^*(u_2) - \hat{\alpha}_{nb}^*(u_1)\right]^2 \leq C\left[J_n(u_2) - J_n(u_0)\right]^2, \tag{A.72}$$

where $C < \infty$ is a generic constant, J_n monotone $a.s.$, and $J_n \to J$ a.s. Then, applying Lemma 5.1 of Stute (1997),

$$LHS(A.72) \leq \frac{3}{n^2}\sum_{i \neq j}\sum \mathbb{E}_\xi \lambda_i^2 \mathbb{E}_\xi \gamma_i^2,$$

$\lambda_i = b^{\mathrm{T}} X_i \hat{V}_i \xi_i 1_{\left\{Z_{\lfloor nu_0 \rfloor :n} \leq Z_i \leq Z_{\lfloor nu_1 \rfloor :n}\right\}}$ and $\gamma_i = b^{\mathrm{T}} X_i \hat{V}_i \xi_i 1_{\left\{Z_{\lfloor nu_2 \rfloor :n} \leq Z_i \leq Z_{\lfloor nu_1 \rfloor :n}\right\}}$. Then,

$$\begin{aligned} LHS(A.72) &\leq \frac{3}{n^2}\sum_{i \neq j}\sum \left(b^{\mathrm{T}} X_i\right)^2 \left(b^{\mathrm{T}} X_j\right)^2 \hat{V}_i^2 \hat{V}_j^2 1_{\left\{Z_{\lfloor nu_0 \rfloor :n} \leq Z_i \leq Z_{\lfloor nu_1 \rfloor :n}\right\}} 1_{\left\{Z_{\lfloor nu_1 \rfloor :n} \leq Z_i \leq Z_{\lfloor nu_2 \rfloor :n}\right\}} \\ &\leq 3\left[J_n(u_2) - J_n(u_0)\right]^2, \end{aligned}$$

where

$$J_n(u) = \frac{1}{n}\sum_{i=1}^{n} \left(b^{\mathrm{T}} X_i\right)^2 \hat{V}_i^2 1_{\left\{Z_i \leq Z_{\lfloor nu \rfloor :n}\right\}}$$

is monotone and $J_n(u) \to J(u) = \mathbb{E}\left(\left(b^{\mathrm{T}} X\right)^2 V^2 1_{\{Z \leq u\}}\right)$ $a.s.$ uniformly in u.

Proofs of Chapter 4

To prove theorems 4.2.1-4.2.4 we need the following lemmas, which are stated without proof. To find the proof the reader my refer to Hall and Heyde (2014) for lemma D.0.1 and Volkonskii and Rozanov (1959) for lemma D.0.2. Besides, throughout the profs we denote C a generic positive constant, which takes different values at different times. We also use the notation a_t which can also take different values at different times.

Lemma D.0.1 (Davydov's lemma)**.** Suppose that two random variables X and Y are $\mathbb{F}_{-\infty}^{t}$ and $\mathbb{F}_{t+\tau}^{\infty}$ adapted, respectively, and that $||X||_p < \infty$ and $||Y||_q\infty$, where $||X||_p = \{\mathrm{E}|X|^p\}^{1/p}$, $p,q \geq 1$, and $1/p + 1/q < 1$. Then,

$$\sup_t |\mathrm{Cov}(X,Y)| \geq 8\alpha^{1/r}(\tau)\{\mathrm{E}|X|^p\}^{1/p}\{\mathrm{E}|Y|^q\}^{1/q},$$

where $r = (1-1/p-1/q)^{-1}$ and $\alpha(\cdot)$ is the mixing coefficient.

Lemma D.0.2. Let $V_1,\ldots,V_{L_1}$ be α-mixing stationary random variables that are $\mathbb{F}_{i_1}^{j_1},\ldots,\mathbb{F}_{i_{L_1}}^{j_{L_1}}$-measurable, respectively, with $1 \leq i_1 < j_1 < \ldots < j_{L_1}$, $i_{\ell+1} - j_\ell \geq \tau$, and $|V_\ell| \leq 1$ for $\ell = 1,\ldots,L_1$. Then,

$$\left|\mathrm{E}\left(\prod_{\ell=1}^{L_1} V_\ell\right) - \prod_{\ell=1}^{L_1}\mathrm{E}(V_\ell)\right| \geq 16(L_1-1)\alpha(\tau),$$

where $\alpha(\cdot)$ is the mixing coefficient.

D.1 Proof of theorem 4.2.1

Here we focus on the analysis of the conditional bias of the estimator $vec(\hat{\pi}(z))$. Note that from (4.4) we can write

$$vec\left(\hat{\pi}(z)\right) = e_1^\top \mathtt{S}_{T,1}^{-1}(z)\mathtt{T}_{T,1}(z), \tag{A.73}$$

where

$$\mathtt{S}_{T,1}(z) = \begin{pmatrix} \mathtt{S}_{T,1.0} & \mathtt{S}_{T,1.1}^\top \\ \mathtt{S}_{T,1.1} & \mathtt{S}_{T,1.2} \end{pmatrix} \quad , \quad \mathtt{T}_{T,1}(z) = \begin{pmatrix} \mathtt{T}_{T,1.0} \\ \mathtt{T}_{T,1.1} \end{pmatrix},$$

and

$$\begin{aligned}
\mathtt{S}_{T,1.0} &= \frac{1}{T}\sum_{t=1}^{T} I_{dm}L_{H_1}(\mathscr{Z}_t - z),\\
\mathtt{S}_{T,1.1} &= \frac{1}{T}\sum_{t=1}^{T} I_{dm}\otimes(\mathscr{Z}_t - z)L_{H_1}(\mathscr{Z}_t - z),\\
\mathtt{S}_{T,1.2} &= \frac{1}{T}\sum_{t=1}^{T} I_{dm}\otimes(\mathscr{Z}_t - z)(\mathscr{Z}_t - z)^{\top}L_{H_1}(\mathscr{Z}_t - z),
\end{aligned}$$

and

$$\begin{aligned}
\mathtt{T}_{T,1.0} &= \frac{1}{T}\sum_{t=1}^{T} I_{dm}vec(X_t)L_{H_1}(\mathscr{Z}_t - z),\\
\mathtt{T}_{T,1.1} &= \frac{1}{T}\sum_{t=1}^{T} I_{dm}\otimes(\mathscr{Z}_t - z)vec(X_t)L_{H_1}(\mathscr{Z}_t - z).
\end{aligned}$$

We start the analysis studying the inverse term in (A.73), $\mathtt{S}_{T,1}^{-1}(z)$; here we will show that as T to infinity

$$\mathtt{S}_{T,1.0} = I_{dm}f(z) + o_P(1); \tag{A.74}$$

to this aim we will use the following 3 steps procedure.

Step 1: Under the strictly stationarity assumption (see, assumption 4.2.1) and the law of iterated expectations we have

$$\begin{aligned}
\mathrm{E}\left(\mathtt{S}_{T,1.0}\right) &= \frac{1}{|H_1|^{1/2}}\mathrm{E}\left[I_{dm}L\left(\frac{\mathscr{Z}_t - z}{H_1^{1/2}}\right)\right]\\
&= \frac{I_{dm}}{|H_1|^{1/2}}\int f(\mathscr{Z}_t)L\left(\frac{\mathscr{Z}_t - z}{H_1^{1/2}}\right)d\mathscr{Z}_t = \frac{I_{dm}}{|H_1|^{1/2}}\int f(u + H_1^{1/2}z)L(u)\,|H_1|^{1/2}du\\
&= I_{dm}f(z) + o_P(1),
\end{aligned}$$

where $u = \frac{\mathscr{Z}_t - z}{H_1^{1/2}}$ and the last equality comes from assumption 4.2.3 and a Taylor expansion.

Step 2: To conclude we just need to sow that $\mathrm{Var}\left(\mathtt{S}_{T,1.0}\right) \to 0$ as $T \to \infty$. To this end,

$$\begin{aligned}
T|H_1|^{1/2}\mathrm{Var}\left(\mathtt{S}_{T,1.0}\right) &= |H_1|^{1/2}\mathrm{Var}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right]\\
&\quad + \frac{2|H_1|^{1/2}}{T}\sum_{t=1}^{T-1}(T-t)\mathrm{Cov}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right), I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right]\\
&\equiv J_{1,1} + J_{1,2}.
\end{aligned}$$

By assumptions 4.2.1 and 4.2.3,

$$\mathrm{Var}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right] = O_P\left(|H_1|^{1/2}\right),$$

which implies that $J_{1,1} = O_P(1)$. Next we prove that $J_{1,2} \to 0$; to this end, we reformulate $J_{1,2}$ as $J_{1,2} = J_{1,3} + J_{1,4}$, where

$$\begin{aligned} J_{1,3} &= \frac{2|H_1|^{1/2}}{T}\sum_{t=1}^{d_T}(T-t)\mathrm{Cov}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right), I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right], \\ J_{1,4} &= \frac{2|H_1|^{1/2}}{T}\sum_{t=d_T+1}^{T-1}(T-t)\mathrm{Cov}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right), I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right]. \end{aligned}$$

Let $d_T \to \infty$ be a sequence of integers such that $d_T|H_1|^{1/2} \to 0$. First we show that $J_{1,3} \to 0$; conditional on $\mathscr{Z}_1$ and $\mathscr{Z}_t$ and using assumptions 4.2.1, 4.2.3-4.2.4 we obtain that

$$\begin{aligned} \mathrm{Cov}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right), I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right] &\le \mathrm{E}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right)L_{H_1}\left(\mathscr{Z}_t - z\right)\right] \\ &\le \int f(u,v)L(u)L(v)dudv \\ &= O_P(1). \end{aligned}$$

Thus, it follows that $J_{1,3} \le d_T|H_1|^{1/2} \to 0$. We now show the contribution of $J_{1,4}$; then, for an α-mixing process we use Davydov's inequality (see, Lemma D.0.1) to obtain

$$\begin{aligned} \left|\mathrm{Cov}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right), I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right]\right| &\le C\left[\alpha(t)\right]^{\frac{\delta}{2+\delta}}\left\|I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right)\right\|_{2+\delta} \\ &\quad \times\left\|I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right\|_{2+\delta} \end{aligned}$$

By assumptions 4.2.1, 4.2.3-4.2.4

$$\begin{aligned} \mathrm{E}\left|I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right|^{2+\delta} &= \left|\frac{I_{dm}}{|H_1|^{1/2}}\right|^{2+\delta} f(z)|H_1|^{1/2}\int L^{2+\delta}(u)du + o_P\left(|H_1|^{-\frac{1}{2}(1+\delta)}\right) \\ &= O_P\left(|H_1|^{-\frac{1}{2}(1+\delta)}\right). \end{aligned}$$

Thus,

$$\left|\mathrm{Cov}\left[I_{dm}L_{H_1}\left(\mathscr{Z}_1 - z\right), I_{dm}L_{H_1}\left(\mathscr{Z}_t - z\right)\right]\right| = O_P\left(\left[\alpha(t)\right]^{\frac{\delta}{2+\delta}}|H_1|^{-\frac{1+\delta}{2+\delta}}\right),$$

and

$$\left|J_{1,4}\right| = C\frac{2|H_1|^{1/2}}{T}\sum_{t=d_T+1}^{T-1}(T-t)\left[\alpha(t)\right]^{\frac{\delta}{2+\delta}}|H_1|^{-\frac{1+\delta}{2+\delta}} \le C\sum_{t=d_T+1}^{T-1}\left[\alpha(t)\right]^{\frac{\delta}{2+\delta}}|H_1|^{-\frac{1}{2}\frac{\delta}{2+\delta}}.$$

Then by assumption 4.2.1 and choosing $d_T^{2+\delta}|H_1|^{1/2} = O_P(1)$,

$$J_{1,4} = C\sum_{t=d_T+1}^{T-1} [\alpha(t)]^{\frac{\delta}{2+\delta}} |H_1|^{-\frac{1}{2}\frac{\delta}{2+\delta}} = o_P\left(d_T^{-\delta}|H_1|^{-\frac{1}{2}\frac{\delta}{2+\delta}}\right) = o_P(1),$$

where d_T satisfies the requirement that $d_T|H_1|^{1/2} \to 0$. Then by assumption 4.2.5 as $T \to \infty$

$$\mathrm{Var}\left(\mathtt{S}_{T,1.0}\right) = o_P(1).$$

Step 3: Thus (A.74) follows.

Using the same 3 steps technique we get that

$$\mathtt{S}_{T,1.1} = 0_{dm(p+q)\times dm}, \tag{A.75}$$

where $0_{dm(p+q)\times dm}$ is a $dm(p+q)\times dm$ matrix of zeros, and

$$\mathtt{S}_{T,1.2} = I_{dm}f(z)\otimes H_1\mu_2(L) + o_P(H_1). \tag{A.76}$$

Therefore by (A.74), (A.75) and (A.76),

$$\mathtt{S}_{T,1}(z) = I_{dm}f(z)\otimes \mathrm{diag}\left\{1+o_P(1), H_1\mu_2(L)+o_P(H_1)\right\},$$

and

$$\mathtt{S}_{T,1}^{-1}(z) = I_{dm}f^{-1}(z)\otimes \mathrm{diag}\left\{1+o_P(1), H_1^{-1}\mu_2^{-1}(L)+o_P(H_1^{-1})\right\}. \tag{A.77}$$

Now, we just need to study $\mathtt{T}_{T,1}(z)$; to this end, we need to center the vector $\mathtt{T}_{T,1}(z)$ by replacing $vec(X_t)$ with $vec(X_t) - \tilde{Z}_t^{*\top}\mathscr{A}(z)$; note that as we did to estimate (4.4) we need to approximate $vec(\pi(\mathscr{Z}_t))$ using the multivariate Taylor expansion

$$vec(\pi(\mathscr{Z}_t)) = \tilde{Z}_t^{*\top}g(z) + Q_T + R_T$$

where

$$\begin{aligned} Q_T &= \frac{1}{2}I_{dm}\otimes(\mathscr{Z}_t - z)^\top\mathscr{H}_\gamma(z)(\mathscr{Z}_t - z),\\ R_T &= I_{dm}\otimes(\mathscr{Z}_t - z)^\top\mathscr{R}(\mathscr{Z}_t; z)(\mathscr{Z}_t - z), \end{aligned}$$

and $\mathscr{H}_\gamma(z)$ is a $dm(p+q)\times(p+q)$ matrix such that $\mathscr{H}_{\gamma_r}(z)$ is the Hessian matrix of the rth component of $\gamma(z)$; also, let rth component of $\mathscr{R}(\mathscr{Z}_t;z)$ be

$$\mathscr{R}_r(\mathscr{Z}_t;z)=\int_0^1\left[\frac{\partial^2\alpha_r}{\partial z\partial z^\top}(z+\omega(\mathscr{Z}_t-z))-\frac{\partial^2\alpha_r}{\partial z\partial z^\top}(z)\right](1-\omega)d\omega,$$

where ω is a weight function. Then let

$$\mathtt{T}^*_{T,1}(z)\quad=\quad\begin{pmatrix}\mathtt{T}^*_{T,1.01}+\mathtt{T}^*_{T,1.02}+\mathtt{T}^*_{T,1.03}\\ \mathtt{T}^*_{T,1.11}+\mathtt{T}^*_{T,1.12}+\mathtt{T}^*_{T,1.13}\end{pmatrix},$$

where

$$\begin{aligned}
\mathtt{T}^*_{T,1.01} &= \frac{1}{2T}\sum_{t=1}^T I_{dm}\otimes(\mathscr{Z}_t-z)^\top\mathscr{H}_\gamma(z)(\mathscr{Z}_t-z)L_{H_1}(\mathscr{Z}_t-z),\\
\mathtt{T}^*_{T,1.02} &= \frac{1}{T}\sum_{t=1}^T I_{dm}\otimes(\mathscr{Z}_t-z)^\top\mathscr{R}(\mathscr{Z}_t;z)(\mathscr{Z}_t-z)L_{H_1}(\mathscr{Z}_t-z),\\
\mathtt{T}^*_{T,1.03} &= \frac{1}{T}\sum_{t=1}^T I_{dm}vec(v_t)L_{H_1}(\mathscr{Z}_t-z),\\
\mathtt{T}^*_{T,1.11} &= \frac{1}{2T}\sum_{t=1}^T I_{dm}\otimes(\mathscr{Z}_t-z)(\mathscr{Z}_t-z)^\top\mathscr{H}_\gamma(z)(\mathscr{Z}_t-z)L_{H_1}(\mathscr{Z}_t-z),\\
\mathtt{T}^*_{T,1.12} &= \frac{1}{T}\sum_{t=1}^T I_{dm}\otimes(\mathscr{Z}_t-z)(\mathscr{Z}_t-z)^\top\mathscr{R}(\mathscr{Z}_t;z)(\mathscr{Z}_t-z)L_{H_1}(\mathscr{Z}_t-z),\\
\mathtt{T}^*_{T,1.13} &= \frac{1}{T}\sum_{t=1}^T I_{dm}\otimes(\mathscr{Z}_t-z)vec(v_t)L_{H_1}(\mathscr{Z}_t-z).
\end{aligned}$$

Note that by construction and the condition (4.3), $\mathrm{E}\left[\left(\mathtt{T}^{*\top}_{T,1.03},\mathtt{T}^{*\top}_{T,1.13}\right)^\top\right]=0_{dm(p+q+1)}$; using similar arguments to those used to prove (A.74) and using assumptions 4.2.1-4.2.7 we can prove that,

$$T|H_1|^{1/2}\mathrm{Var}\left[\begin{pmatrix}\mathtt{T}^*_{T,1.03}\\ \mathtt{T}^*_{T,1.13}\end{pmatrix}\right]=\Omega_v(z)f(z)\otimes\mathrm{diag}\left\{R(L)+o_P(1),H_1\mu_2^2(L)+o_P(H_1)\right\}+o_P(1)\quad\text{(A.78)}$$

We continue the analysis studying the rest of the terms in $\mathtt{T}^*_{T,1}(z)$, so using similar techniques to those used to prove (A.74) we get

$$\mathtt{T}^*_{T,1.01}=\frac{1}{2}I_{dm}f(z)\mu_2(L)\mathrm{diag}_{dm}\left[\mathrm{tr}\left(\mathscr{H}_{\gamma_r}(z)H_1\right)\right]i_{dm}+o_P\left(\mathrm{tr}\left(H_1\right)\right),\quad\text{(A.79)}$$

and by the same token, we can show that the variance of $H_1^{-1}\mathtt{T}^*_{T,1.01}$ converges to zero. Similarly, we can show that

$$\mathtt{T}^*_{T,1.02} = o_P\left(\mathrm{tr}\left(H_1\right)\right), \tag{A.80}$$

$$\mathtt{T}^*_{T,1.11} = O_P\left(H_1^{3/2}\right), \tag{A.81}$$

$$\mathtt{T}^*_{T,1.12} = O_P\left(H_1^{3/2}\right). \tag{A.82}$$

Therefore by (A.79), (A.80), (A.81) and (A.82)

$$\mathtt{T}^*_{T,1}(z) = \begin{pmatrix} \frac{1}{2}I_{dm}f(z)\mu_2(L)\mathrm{diag}_{dm}\left[\mathrm{tr}\left(\mathscr{H}_{\gamma_r}(z)H_1\right)\right]i_{dm} + o_P\left(\mathrm{tr}\left(H_1\right)\right) \\ O_P\left(H_1^{3/2}\right) \end{pmatrix}. \tag{A.83}$$

Thus, using (A.77) and (A.83) into (A.73) we close the proof.

Before proving the next theorem, we first consider the term $\hat{\pi}(z) - \pi(z)$. It follows from the proof of Theorem 4.2.1 that we can write, for $1 \leq j \leq d$ and $1 \leq k \leq m$, the $(j,k)-th$ component of (4.4) as $\hat{\pi}_{(j,k)}(z) = \frac{1}{T}\sum_{t=1}^{T} W_{T,1}^{H_1}(\mathscr{Z}_t - z)X_{(j,k),t}$, where

$$W_{T,1}^{H_1}(\mathscr{Z}_t - z) = e_3^{\top}\{S_{T,1}^{*}(z)\}^{-1}\begin{pmatrix} 1 \\ \mathscr{Z}_t - z \end{pmatrix} L_{H_1}(\mathscr{Z}_t - z),$$

$e_3 = \left(1, 0_{p+q}\right)^{\top}$, 0_{p+q} is a $(p+q) \times 1$ vector of zeros, and

$$S_{T,1}^{*}(z) = \begin{pmatrix} 1 & (\mathscr{Z}_t - z)^{\top} \\ (\mathscr{Z}_t - z) & (\mathscr{Z}_t - z)(\mathscr{Z}_t - z)^{\top} \end{pmatrix} L_{H_1}(\mathscr{Z}_t - z).$$

By (A.77) we can infer that $S_{T,1}^{*}(z) = f(z)\mathrm{diag}\{1 + o_P(1), H_1\mu_2(L) + o_P(H_1)\}$, then

$$W_{T,1}^{H_1}(\mathscr{Z}_t - z) = \frac{1}{f(z)} L_{H_1}(\mathscr{Z}_t - z)\{1 + o_P(1)\} \tag{A.84}$$

holds and

$$\hat{\pi}_{(j,k)}(z) = \frac{1}{Tf(z)}\sum_{t=1}^{T} L_{H_1}(\mathscr{Z}_t - z)X_{(j,k),t}\{1 + o_P(1)\};$$

see Fan and Gijbels (1996) for more details. Thus,

$$\hat{\pi}_{(j,k)}(z) - \pi_{(j,k)}(z) = \frac{1}{T}\sum_{t=1}^{T} W_{T,1}^{H_1}(\mathscr{Z}_t - z)\left(\pi_{(j,k)}(\mathscr{Z}_t) - \pi_{(j,k)}(z)\right) + \frac{1}{T}\sum_{t=1}^{T} W_{T,1}^{H_1}(\mathscr{Z}_t - z)v_{(j,k),t},$$

where taking Taylor expansion, for z in the neighborhood of $\mathscr{Z}_t$ and taking into account the proof of Theorem 4.2.1

$$\hat{\pi}_{(j,k)}(z) - \pi_{(j,k)}(z) = \frac{1}{T}\sum_{t=1}^{T} W_{T,1}^{H_1}(\mathscr{Z}_t - z)\left\{\frac{1}{2}(\mathscr{Z}_t - z)^{\top}\mathscr{H}_{\pi_{(j,k)}}(z)(\mathscr{Z}_t - z) + v_{(j,k),t}\right\} + o_P\left(\mathrm{tr}\{H_1\}\right). \tag{A.85}$$

D.2 Proof of theorem 4.2.2

Here, as in the proof of theorem 4.2.1, we focus on the analysis of the conditional bias of the estimator $\bar{\beta}(z_1))$. Note that from (4.6) we can write

$$\bar{\beta}(z_1)) = e_2^{\top}\mathtt{S}_{T,2}^{-1}(z_1)\mathtt{T}_{T,2}(z_1), \tag{A.86}$$

where

$$\mathtt{S}_{T,2}(z) = \begin{pmatrix} \mathtt{S}_{T,2.0} & \mathtt{S}_{T,2.1}^{\top} \\ \mathtt{S}_{T,2.1} & \mathtt{S}_{T,2.2} \end{pmatrix} \quad , \quad \mathtt{T}_{T,2}(z) = \begin{pmatrix} \mathtt{T}_{T,2.0} \\ \mathtt{T}_{T,2.1} \end{pmatrix},$$

and

$$\begin{aligned} \mathtt{S}_{T,2.0} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^{\top} K_{H_2}(Z_{1t}-z_1), \\ \mathtt{S}_{T,2.1} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^{\top} \otimes (Z_{1t}-z) K_{H_2}(Z_{1t}-z_1), \\ \mathtt{S}_{T,2.2} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^{\top} \otimes (Z_{1t}-z)(Z_{1t}-z)^{\top} K_{H_2}(Z_{1t}-z_1), \end{aligned}$$

and

$$\begin{aligned} \mathtt{T}_{T,2.0} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t) Y_t K_{H_2}(Z_{1t}-z_1), \\ \mathtt{T}_{T,2.1} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t) \otimes (Z_{1t}-z) Y_t K_{H_2}(Z_{1t}-z_1). \end{aligned}$$

We start the analysis studying the inverse term in (A.86), $\mathtt{S}_{T,2}^{-1}(z)$; here we will show that as $T \to \infty$

$$\mathtt{S}_{T,2.0} = \Gamma_*(z_1) f(z) + o_P(1), \tag{A.87}$$

where $\Gamma_*(z_1) = \mathrm{E}\left[\pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^{\top} \middle| Z_{1t} = z_1\right]$; to this end let

$$\mathtt{S}_{T,2.0} = \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^{\top} K_{H_2}(Z_{1t}-z_1) = J_{2,1} + J_{2,2} + J_{2,2}^{\top} + J_{2,4},$$

where

$$\begin{aligned} J_{2,1} &= \frac{1}{T}\sum_{t=1}^{T} \pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^{\top} K_{H_2}(Z_{1t}-z_1), \\ J_{2,2} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\left[\hat{\pi}(\mathscr{Z}_t) - \pi(\mathscr{Z}_t)\right]^{\top} K_{H_2}(Z_{1t}-z_1), \\ J_{2,3} &= \frac{1}{T}\sum_{t=1}^{T} \left[\hat{\pi}(\mathscr{Z}_t) - \pi(\mathscr{Z}_t)\right]\left[\hat{\pi}(\mathscr{Z}_t) - \pi(\mathscr{Z}_t)\right]^{\top} K_{H_2}(Z_{1t}-z_1). \end{aligned}$$

We first consider $J_{2,1}$, and as in **Step 1** in the proof of theorem 4.2.1, under the strictly stationarity assumption and the law of iterated expectations we have

$$\begin{aligned}
\mathrm{E}\left(J_{2,1}\right) &= \frac{1}{|H_2|^{1/2}}\mathrm{E}\left[\pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top K\left(\frac{Z_{1t}-z}{H_2^{1/2}}\right)\right] \\
&= \frac{\Gamma_*(z_1)}{|H_2|^{1/2}}\int f(Z_{1t})K\left(\frac{Z_{1t}-z}{H_2^{1/2}}\right)dZ_{1t} = \frac{\Gamma_*(z_1)}{|H_1|^{1/2}}\int f(u+H_2^{1/2}z_1)K(u)\,|H_2|^{1/2}du \\
&= \Gamma_*(z_1)f(z_1)+o_P(1), \qquad \text{(A.88)}
\end{aligned}$$

where $u=\frac{Z_{1t}-z_1}{H_2^{1/2}}$ and the last equality comes from assumption 4.2.3 and a Taylor expansion. To conclude we just need to sow that $\mathrm{Var}\left(J_{2,1}\right)\to 0$ as $T\to\infty$, as in **Step 2**. To this end and for notational simplicity let $a_t=\pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top$, then

$$\begin{aligned}
T|H_1|^{1/2}\mathrm{Var}\left(J_{2,1}\right) &= |H_2|^{1/2}\mathrm{Var}\left[a_tK_{H_2}(Z_{1t}-z_1)\right] \\
&\quad +\frac{2|H_2|^{1/2}}{T}\sum_{t=1}^{T-1}(T-t)\mathrm{Cov}\left[a_1K_{H_2}(Z_{11}-z_1),a_tK_{H_2}(Z_{1t}-z_1)\right] \\
&\equiv J_{2,4}+J_{2,5}.
\end{aligned}$$

By assumptions 4.2.1 and 4.2.3,$\mathrm{Var}\left[\pi(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t}-z_1)\right]=O_P\left(|H_2|^{1/2}\right)$, which implies that $J_{2,4}=O_P(1)$. Next we prove that $J_{2,5}\to 0$; to this end, we reformulate $J_{2,5}$ as $J_{2,5}=J_{2,6}+J_{2,7}$, where $J_{2,6}=\frac{2|H_2|^{1/2}}{T}\sum_{t=1}^{d_T}(\ldots)$ and $J_{2,7}=\frac{2|H_2|^{1/2}}{T}\sum_{t=d_T+1}^{T-1}(\ldots)$.

Let $d_T\to\infty$ be a sequence of integers such that $d_T|H_2|^{1/2}\to 0$. First we show that $J_{2,6}\to 0$; conditional on Z_1 and Z_t and using assumptions 4.2.1, 4.2.3-4.2.4 and 4.2.9 we obtain that

$$\begin{aligned}
\mathrm{Cov}\left[a_1K_{H_2}(Z_{11}-z_1),a_tK_{H_2}(Z_{1t}-z_1)\right] &\leq \mathrm{E}\left[a_1a_t^\top K_{H_2}(Z_{11}-z_1)K_{H_2}(Z_{1t}-z_1)\right] \\
&\leq \int f(u,v)K(u)K(v)dudv \\
&= O_P(1).
\end{aligned}$$

Thus, it follows that $J_{2,6}\leq d_T|H_2|^{1/2}\to 0$. We now show the contribution of $J_{2,7}$; then, for an α-mixing process we use Davydov's inequality (see, Lemma D.0.1) to obtain

$$\begin{aligned}
\left|\mathrm{Cov}\left[a_1K_{H_2}(Z_{11}-z_1),a_tK_{H_2}(Z_{1t}-z_1)\right]\right| &\leq C\left[\alpha(t)\right]^{\frac{\delta}{2+\delta}}\left\|a_1K_{H_2}(Z_{11}-z_1)\right\|_{2+\delta} \\
&\quad \times\left\|a_tK_{H_2}(Z_{1t}-z_1)\right\|_{2+\delta}
\end{aligned}$$

By assumptions 4.2.1, 4.2.3-4.2.4 and 4.2.9-4.2.10

$$\begin{aligned}\mathrm{E}\left|a_t K_{H_2}(Z_{1t}-z_1)\right|^{2+\delta} &= \left|\frac{\Gamma_*(z_1)}{|H_2|^{1/2}}\right|^{2+\delta} f(z_1)|H_2|^{1/2}\int K^{2+\delta}(u)du + o_P\left(|H_2|^{-\frac{1}{2}(1+\delta)}\right)\\ &= O_P\left(|H_2|^{-\frac{1}{2}(1+\delta)}\right).\end{aligned}$$

Thus,

$$\left|\mathrm{Cov}\left[a_1 K_{H_2}(Z_{11}-z_1), a_t K_{H_2}(Z_{1t}-z_1)\right]\right| = O_P\left([\alpha(t)]^{\frac{\delta}{2+\delta}}\,|H_2|^{-\frac{1+\delta}{2+\delta}}\right),$$

and

$$\left|J_{2,7}\right| = C\frac{2|H_2|^{1/2}}{T}\sum_{t=d_T+1}^{T-1}(T-t)\,[\alpha(t)]^{\frac{\delta}{2+\delta}}\,|H_2|^{-\frac{1+\delta}{2+\delta}} \le C\sum_{t=d_T+1}^{T-1}[\alpha(t)]^{\frac{\delta}{2+\delta}}\,|H_2|^{-\frac{1}{2}\frac{\delta}{2+\delta}}.$$

Then by assumption 4.2.1 and choosing $d_T^{2+\delta}|H_2|^{1/2} = O_P(1)$,

$$J_{2,7} = C\sum_{t=d_T+1}^{T-1}[\alpha(t)]^{\frac{\delta}{2+\delta}}\,|H_2|^{-\frac{1}{2}\frac{\delta}{2+\delta}} = o_P\left(d_T^{-\delta}|H_2|^{-\frac{1}{2}\frac{\delta}{2+\delta}}\right) = o_P(1),$$

where d_T satisfies the requirement that $d_T|H_2|^{1/2}\to 0$. Note that in assumption 4.2.5 we assume that $H_2\to 0$ and $T|H_2|^{1/2}\to\infty$ as $T\to\infty$; then

$$\mathrm{Var}\left(J_{2,1}\right) = o_P(1),$$

and (A.88) follows, as in **Step 3** .

Next, we study $J_{2,2}$, by assumption 4.2.1, 4.2.3,4.2.5, 4.2.6 and 4.2.10 and applying Theorem 6 of Masry (1996), we have

$$\max_t|\hat{\pi}(\mathscr{Z}_t)-\pi(\mathscr{Z}_t)| = O_P\left\{\left(\frac{T|H_1|^{1/2}}{\log T}\right)^{-1/2}\right\} + O_P(|H_1|) = o_P(1).$$

Then, since $\frac{1}{T}\sum_{t=1}^{T}K_{H_1}(Z_{1t}-z_1)\,|\pi(\mathscr{Z}_t)| = O_P(1)$,

$$\left|J_{2,2}\right| \le \frac{1}{T}\sum_{t=1}^{T}K_{H_1}(Z_{1t}-z_1)\,|\pi(\mathscr{Z}_t)|\max_t|\hat{\pi}(\mathscr{Z}_t)-\pi(\mathscr{Z}_t)| = o_P(1).$$

Similarly, we can show that $J_{2,3} = o_P(1)$. Therefore (A.87) follows.

Using similar arguments we can sow that

$$\mathtt{S}_{T,2.1} = 0_{dq\times d}, \tag{A.89}$$

where $0_{dq\times d}$ is a $dq\times d$,and finally we obtain

$$\mathtt{S}_{T,2.2} = \Gamma_*(z_1)f(z_1)\otimes H_2\mu_2(K) + o_P(H_2). \tag{A.90}$$

Therefore by (A.87), (A.89) and (A.90),

$$\mathtt{S}_{T,2}^{-1}(z) = \Gamma_*^{-1}(z_1)f^{-1}(z_1)\otimes \operatorname{diag}\left\{1+o_P(1), H_2^{-1}\mu_2^{-1}(K)+o_P(H_2^{-1})\right\}. \tag{A.91}$$

Now, we just need to study $\mathtt{T}_{T,2}(z)$; to this end, as in the proof of theorem 4.2.1 we need to center the vector $\mathtt{T}_{T,2}(z)$ by replacing Y_t with $Y_t - \Pi_t(\mathscr{Z}_t)^\top \eta(z_1)$; then, observe that

$$\begin{aligned}
\mathtt{T}^*_{T,2}(z) &= \frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\left[Y_t - \Pi_t(\mathscr{Z}_t)^\top\eta(z_1)\right]K_{H_2}(Z_{1t}-z_1)\\
&= \frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\left[X_t^\top\beta(Z_{1t}) - \Pi_t(\mathscr{Z}_t)^\top\eta(z_1)+u_t\right]K_{H_2}(Z_{1t}-z_1)\\
&= \mathtt{T}_{T,2.1}+\mathtt{T}_{T,2.2}.
\end{aligned} \tag{A.92}$$

where

$$\begin{aligned}
\mathtt{T}_{T,2.1} &= \frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\left[X_t^\top\beta(Z_{1t}) - \Pi_t(\mathscr{Z}_t)^\top\eta(z_1)\right]K_{H_2}(Z_{1t}-z_1),\\
\mathtt{T}_{T,2.2} &= \frac{1}{T}\sum_{t=1}^{T}\tilde{\Pi}_t(\mathscr{Z}_t)\left[u_t\right]K_{H_2}(Z_{1t}-z_1).
\end{aligned}$$

Note that we can express $\mathtt{T}_{T,2.1}$ as

$$\mathtt{T}_{T,2.1}(z_1) = \begin{pmatrix}\mathtt{T}_{T,2.11}+\mathtt{T}_{T,2.12}\\ \mathtt{T}_{T,2.13}+\mathtt{T}_{T,2.14}\end{pmatrix},$$

where, taking onto account that $K_{H_2,t} = K_{H_2}(Z_{1t} - z_1)$ and $\pi_{z_t} = \pi(\mathscr{Z}_t)$,

$$\begin{aligned}
\mathrm{T}_{T,2.11} &= \frac{1}{T}\sum_{t=1}^{T} \pi_{z_t}\left[X_t^\top \beta(Z_{1t}) - \Pi_{t,z_t}^\top \eta(z_1)\right] K_{H_2,t}, \\
\mathrm{T}_{T,2.12} &= \frac{1}{T}\sum_{t=1}^{T} [\hat{\pi}_{z_t} - \pi_{z_t}]\left[X_t^\top \beta(Z_{1t}) - \Pi_{t,z_t}^\top \eta(z_1)\right] K_{H_2,t}, \\
\mathrm{T}_{T,2.13} &= \frac{1}{T}\sum_{t=1}^{T} \pi_{z_t} \otimes (Z_{1t} - z_1)\left[X_t^\top \beta(Z_{1t}) - \Pi_{t,z_t}^\top \eta(z_1)\right] K_{H_2,t}, \\
\mathrm{T}_{T,2.14} &= \frac{1}{T}\sum_{t=1}^{T} [\hat{\pi}_{z_t} - \pi_{z_t}] \otimes (Z_{1t} - z_1)\left[X_t^\top \beta(Z_{1t}) - \Pi_{t,z_t}^\top \eta(z_1)\right] K_{H_2,t}.
\end{aligned}$$

Then, under the strictly stationarity assumption and the law of iterated expectations we have that

$$\begin{aligned}
\mathrm{E}(\mathrm{T}_{T,2.11}) &= \mathrm{E}\left(\mathrm{E}\left[\pi_{z_t}\left\{\mathrm{E}(X_t|\mathscr{Z}_t)^\top \beta(Z_{1t}) - \Pi_{t,z_t}^\top \eta(z_1)\right\} K_{H_2,t}\middle| Z_{1t=z_1}\right]\right) \\
&= \mathrm{E}(\mathrm{E}[\pi_{z_t}\{Q_T + R_T\} K_{H_2,t}|Z_{1t=z_1}]) \\
&= \mathrm{E}\left(\mathrm{E}\left[\mathrm{T}_{T,2.111}\middle|Z_{1t=z_1}\right]\right) + \mathrm{E}\left(\mathrm{E}\left[\mathrm{T}_{T,2.112}\middle|Z_{1t=z_1}\right]\right),
\end{aligned}$$

where the second equality comes from the fact that $\mathrm{E}(X_t|\mathscr{Z}_t) = \pi(\mathscr{Z}_t)$ and approximating $\pi(\mathscr{Z}_t)^\top \beta(Z_{1t})$ using the multivariate Taylor expansion

$$\begin{aligned}
\pi(\mathscr{Z}_t)^\top \beta(Z_{1t}) &= \Pi_t(\mathscr{Z}_t)^\top \eta(z_1) + \frac{1}{2}\hat{\pi}(\mathscr{Z}_t)^\top \otimes (Z_{1t} - z_1)^\top \mathscr{H}_\beta(z_1)(Z_{1t} - z_1) \\
&\quad + \hat{\pi}(\mathscr{Z}_t)^\top \otimes (Z_{1t} - z_1)^\top \mathscr{R}(Z_{1t}; z_1)(Z_{1t} - z_1) \\
&= \Pi_t(\mathscr{Z}_t)^\top \eta(z_1) + Q_T + R_T.
\end{aligned}$$

Now using the 3 steps technique we can show that

$$\mathrm{E}\left(\mathrm{E}\left[\mathrm{T}_{T,2.111}\middle|Z_{1t=z_1}\right]\right) = \frac{1}{2}\Gamma_*(z_1) f(z_1)\mu_2(K)\mathrm{diag}_d\left[\mathrm{tr}\left(\mathscr{H}_{\beta_r}(z_1)H_2\right)\right] i_d + o_P(\mathrm{tr}(H_2)). \tag{A.93}$$

and

$$\mathrm{E}\left(\mathrm{E}\left[\mathrm{T}_{T,2.112}\middle|Z_{1t=z_1}\right]\right) = o_P(\mathrm{tr}(H_2)); \tag{A.94}$$

then using (A.93) and (A.94) we get that

$$\mathrm{T}_{T,2.11} = \frac{1}{2}\Gamma_*(z_1) f(z_1)\mu_2(K)\mathrm{diag}_d\left[\mathrm{tr}\left(\mathscr{H}_{\beta_r}(z_1)H_2\right)\right] i_d + o_P(\mathrm{tr}(H_2)). \tag{A.95}$$

Similarly, we can show that

$$\mathtt{T}_{T,2.13} = O_P\left(H_2^{3/2}\right), \tag{A.96}$$

We now analyse $\mathtt{T}_{T,2.12}$; to this end, let $\pi_{(j,k),z_t}$ be the $(j,k)-th$ element of $\pi(\mathscr{Z}_t)$; then

$$\mathtt{T}_{T,2.12}(j,k) = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{\pi}_{(j,k),z_t} - \pi_{(j,k),z_t}\right]\left[X_{(j,k)t}\beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top}\eta_j(z_1)\right]K_{H_2,t},$$

where $\Pi_{(j,k),t,z_t} = \left(\pi_{(j,k),z_t}, \pi_{(j,k),z_t}\otimes(Z_{1t}-z_1)^{\top}\right)^{\top}$ and $\eta_j(z_1) = \left(\beta_1(z_1), D_{\beta_j}(z_1)^{\top}\right)^{\top}$. Note that using (A.85) and (A.84) we get that

$$\mathtt{T}_{T,2.12}(j,k) \quad = \quad J_{2,2.1}(j,k) + J_{2,2.2}(j,k) + J_{2,2.3}(j,k),$$

where, naming $a_{st} = (\mathscr{Z}_s - \mathscr{Z}_t)^{\top}\mathscr{H}_{\pi_{(j,k)}}(\mathscr{Z}_t)(\mathscr{Z}_s - \mathscr{Z}_t)$, and $L_{H_1,st} = L_{H_1}(\mathscr{Z}_s - \mathscr{Z}_t)$

$$\begin{aligned}
J_{2,2.1}(j,k) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2Tf(\mathscr{Z}_t)}\sum_{s=1}^{T}a_{st}L_{H_1,st}\left[X_{(j,k)t}\beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top}\eta_j(z_1)\right]K_{H_2,t}\\
J_{2,2.2}(j,k) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{Tf(\mathscr{Z}_t)}\sum_{s=1}^{T}v_{(j,k),s}L_{H_1,st}\left[X_{(j,k)t}\beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top}\eta_j(z_1)\right]K_{H_2,t}\\
J_{2,2.3}(j,k) &= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2Tf(\mathscr{Z}_t)}\sum_{s=1}^{T}O_P\left(\mathrm{tr}\{H_1\}\right)L_{H_1,st}\left[X_{(j,k)t}\beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top}\eta_j(z_1)\right]K_{H_2,t}.
\end{aligned}$$

We look at each term separately, note that $\mathrm{E}\left(J_{2,2.2}(j,k)\right) = 0$ and following Martins-Filho and Yao (2009),

$$\begin{aligned}
T|H_2|^{1/2}\mathrm{Var}\left(H_2^{-1}J_{2,2.1}(j,k)\right) &= \frac{|H_2|^{1/2}}{T}\sum_{s=1}^{T}\sum_{s'=1}^{T}\mathrm{E}\left[v_{(j,k),s}v_{(j,k),s'}a_s a_{s'}\right]\\
&= s_1 + s_2 + s_3 + s_4,
\end{aligned}$$

where $a_s = \sum_{t=1}^{T}\frac{1}{Tf(\mathscr{Z}_t)}\left[X_{(j,k)t}\beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top}\eta_j(z_1)\right]L_{H_1,st}K_{H_2,t}$. Then we need to study several cases, first $s = s'$ and $t = t'$

$$\begin{aligned}
s_1 &= |H_2|^{1/2}H_2^{-2}\mathrm{E}\left[v_{(j,k),s}^2\sum_{t=1}^{T}\frac{1}{T^2f^2(\mathscr{Z}_t)}\left[X_{(j,k)t}\beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top}\eta_j(z_1)\right]^2 L_{H_1,st}^2K_{H_2,t}^2\right]\\
&= O_P\left(\frac{1}{T|H_1|^{1/2}}\right),
\end{aligned}$$

$s = s'$ and $t \neq t'$

$$\begin{aligned} s_2 &= \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=1}^{T} (T-t) \mathrm{Cov}\left[v_{(j,k),s} \sum_{t=1}^{T} \frac{1}{T f(\mathscr{Z}_t)} \left[X_{(j,k)t} \beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top} \eta_j(z_1) \right] L_{H_1,st} K_{H_2,t}, \right. \\ & \left. v_{(j,k),s} \sum_{t=1}^{T} \frac{1}{T f(\mathscr{Z}_1)} \left[X_{(j,k)1} \beta_j(Z_{11}) - \Pi_{(j,k),1,z_1}^{\top} \eta(z_1) \right] L_{H_1,s1} K_{H_2,1} \right] \\ &= s_{2.1} + s_{2.2} = o_P(1), \end{aligned}$$

where $s_{2.1} = \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=1}^{d_T} (\ldots)$ and $s_{2.2} = \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=d_T+1}^{T-1} (\ldots)$. Next $s \neq s'$ and $t = t'$

$$\begin{aligned} s_3 &= \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=1}^{T} (T-t) \mathrm{E}\left[v_{(j,k),s} v_{(j,k),1} \sum_{t=1}^{T} \frac{1}{T^2 f^2(\mathscr{Z}_t)} \left[X_{(j,k)t} \beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top} \eta_j(z_1) \right]^2 \right. \\ & \left. \times L_{H_1,st} L_{H_1,1t} K_{H_2,t}^2 \right] \\ &= O_P(1), \end{aligned}$$

$s \neq s'$ and $t \neq t'$

$$\begin{aligned} s_4 &= \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=1}^{T} (T-t) \mathrm{Cov}\left[v_{(j,k),s} \sum_{t=1}^{T} \frac{1}{T f(\mathscr{Z}_t)} \left[X_{(j,k)t} \beta_j(Z_{1t}) - \Pi_{(j,k),t,z_t}^{\top} \eta_j(z_1) \right] L_{H_1,st} K_{H_2,t}, \right. \\ & \left. v_{(j,k),1} \sum_{t=1}^{T} \frac{1}{T f(\mathscr{Z}_1)} \left[X_{(j,k)1} \beta_j(Z_{11}) - \Pi_{(j,k),1,z_1}^{\top} \eta_j(z_1) \right] L_{H_1,11} K_{H_2,1} \right] \\ &= s_{4.1} + s_{4.2} = o_P(1), \end{aligned}$$

where $s_{4.1} = \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=1}^{d_T} (\ldots)$ and $s_{4.2} = \frac{|H_2|^{1/2} H_2^{-2}}{T} \sum_{s=d_T+1}^{T-1} (\ldots)$. Then it is easy to show that $\mathrm{Var}\left(H_2^{-1} J_{2,2.1}(j,k) \right) = o_P(1)$; therefore, $J_{2,2.2}(j,k) = o_P\left((T|H_2|^{1/2})^{-1/2} \right)$.

We continue our analysis by showing that

$$\begin{aligned} J_{2,2.1}(j,k) &= \mathrm{E}\left[\mathrm{tr}\left\{ \mu_2(L) \mathscr{H}_{\pi_{(j,k)}}(\mathscr{Z}_t) H_1 \right\} \pi_{(j,k),z_t} \middle| Z_{1t} = z_1 \right] \mathrm{tr}\left\{ \mu_2(K) \mathscr{H}_{\beta_j}(z_1) H_2 \right\} \\ & \quad + o_P\left(\mathrm{tr}\{H_1\} \mathrm{tr}\{H_2\} \right) \\ &= O_P\left(\mathrm{tr}\{H_1\} \mathrm{tr}\{H_2\} \right); \end{aligned}$$

and that $\mathrm{Var}\left(H_1^{-1} H_2^{-1} J_{2,2.1}(j,k) \right) = o_P(1)$. Similarly we can show that $J_{2,2.3}(j,k) = O_P\left(\mathrm{tr}\{H_1\} \mathrm{tr}\{H_2\} \right)$.

Therefore we can conclude that

$$\mathtt{T}_{T,2.12}(j,k) = O_P\left(\mathrm{tr}\{H_1\} \mathrm{tr}\{H_2\} \right) + o_P\left((T|H_2|^{1/2})^{-1/2} \right) + O_P\left(\mathrm{tr}\{H_1\} \mathrm{tr}\{H_2\} \right), \quad \text{(A.104)}$$

and similarly we can show that

$$\mathtt{T}_{T,2.14}(j,k) \quad = \quad O_P\left(\operatorname{tr}\{H_1\}H_2^{3/2}\right) + o_P\left((T|H_2|^{1/2})^{-1/2}\right) + O_P\left(\operatorname{tr}\{H_1\}H_2^{3/2}\right), \quad \text{(A.105)}$$

Next we study

$$\begin{aligned}
\mathtt{T}_{T,2.2} &= \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\pi(\mathscr{Z}_t)\\ \pi(\mathscr{Z}_t)\otimes(Z_{1t}-z_1)\end{pmatrix}u_tK_{H_2}(Z_{1t}-z_1)\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix}\hat{\pi}(\mathscr{Z}_t)-\pi(\mathscr{Z}_t)\\ (\hat{\pi}(\mathscr{Z}_t)-\pi(\mathscr{Z}_t))\otimes(Z_{1t}-z_1)\end{pmatrix}u_tK_{H_2}(Z_{1t}-z_1)\\
&= \mathtt{T}_{T,2.21}+\mathtt{T}_{T,2.22};
\end{aligned}$$

it is clear that $\mathrm{E}\left(\mathtt{T}_{T,2.21}\right) = 0_{d(q+1)}$ and applying **Step 2**

$$T|H_2|^{1/2}\mathrm{Var}\left(\mathtt{T}_{T,2.21}\right) = \tilde{\Gamma}(z_1)f(z_1)\otimes\operatorname{diag}\left\{R(K)+o_P(1), H_2\mu_2^2(K)+o_P(H_2)\right\}, \quad \text{(A.107)}$$

where $\tilde{\Gamma}(z_1) = \mathrm{E}[\pi(\mathscr{Z}_t)\Omega(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top|Z_{1t}=z_1]$, also applying the same techniques as in (A.104)

$$\mathtt{T}_{T,2.22} = \begin{pmatrix} o_P\left((T|H_2|^{1/2})^{-1/2}\right)\\ o_P\left((T|H_2|^{1/2})^{-1/2}\right)\end{pmatrix}. \quad \text{(A.108)}$$

Therefore by (A.91), (A.95), (A.96), (A.104), (A.105), and (A.108) we close the proof.

D.3 Proof of theorem 4.2.3

Note that we can write $\bar{u}_t = u_t - X_t^\top\left(\bar{\beta}(Z_{1t})-\beta(Z_{1t})\right)$; then we can rewrite (4.7) as

$$\begin{aligned}
\hat{\Omega}(z) &= \frac{1}{\hat{f}(z)T}\sum_{t=1}^{T}u_tu_t^\top L_{H_1}(\mathscr{Z}_t-z)\\
&\quad+\frac{1}{\hat{f}(z)T}\sum_{t=1}^{T}X_t^\top\left(\bar{\beta}(Z_{1t})-\beta(Z_{1t})\right)\left(\bar{\beta}(Z_{1t})-\beta(Z_{1t})\right)^\top X_tL_{H_1}(\mathscr{Z}_t-z)\\
&= U_{T1}+U_{T2}. \qquad \text{(A.109)}
\end{aligned}$$

It is clear that $U_{T1} \to_P \Omega(z)$, therefore we just need to study the asymptotic behaviour of U_{T2}; note that we can write

$$\begin{aligned}\left(\bar{\beta}(Z_{1t}) - \beta(Z_{1t})\right) &= e_2^\top \mathtt{S}_{T,2}^{-1}(Z_{1t})\frac{1}{T}\sum_{s=1}^{T}\tilde{\Pi}_s(\mathscr{Z}_s)\,(Y_s - \Pi_s(\mathscr{Z}_s)\eta(Z_{1t}))\,K_{H2} \\ &= e_2^\top \mathtt{S}_{T,2}^{-1}(Z_{1t})\frac{1}{T}\sum_{s=1}^{T}\tilde{\Pi}_s(\mathscr{Z}_s)\,(Q_s + R_s + u_s)\,K_{H2},\end{aligned}$$

then using similar techniques to those used in the proof of theorem 4.2.2 we can say that $U_{T2} \to 0_{m\times m}$ as $T|H_2|^{1/2} \to \infty$

D.4 Proof of theorem 4.2.4

Here, as in the proof of theorem 4.2.1, we can write (4.6) as

$$\hat{\beta}(z_1)) = e_2^\top \mathtt{S}_{T,3}^{-1}(z_1)\mathtt{T}_{T,3}(z_1), \tag{A.110}$$

where

$$\mathtt{S}_{T,3}(z) = \begin{pmatrix} \mathtt{S}_{T,3.0} & \mathtt{S}_{T,3.1}^\top \\ \mathtt{S}_{T,3.1} & \mathtt{S}_{T,3.2}\end{pmatrix} \quad , \quad \mathtt{T}_{T,3}(z) = \begin{pmatrix}\mathtt{T}_{T,3.0} \\ \mathtt{T}_{T,3.1}\end{pmatrix},$$

and

$$\begin{aligned}\mathtt{S}_{T,3.0} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}(\mathscr{Z}_t)\hat{\Omega}^{-1}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t} - z_1), \\ \mathtt{S}_{T,3.1} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}(\mathscr{Z}_t)\hat{\Omega}^{-1}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^\top \otimes (Z_{1t} - z)K_{H_2}(Z_{1t} - z_1), \\ \mathtt{S}_{T,3.2} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}(\mathscr{Z}_t)\hat{\Omega}^{-1}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^\top \otimes (Z_{1t} - z)(Z_{1t} - z)^\top K_{H_2}(Z_{1t} - z_1),\end{aligned}$$

and

$$\begin{aligned}\mathtt{T}_{T,3.0} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}(\mathscr{Z}_t)\hat{\Omega}^{-1}(\mathscr{Z}_t)Y_t K_{H_2}(Z_{1t} - z_1), \\ \mathtt{T}_{T,3.1} &= \frac{1}{T}\sum_{t=1}^{T}\hat{\pi}(\mathscr{Z}_t)\otimes(Z_{1t} - z)\hat{\Omega}^{-1}(\mathscr{Z}_t)Y_t K_{H_2}(Z_{1t} - z_1).\end{aligned}$$

We start the analysis studying the inverse term in (A.110), $\mathtt{S}_{T,3}^{-1}(z_1)$; here we will show that as $T \to \infty$

$$\mathtt{S}_{T,3.0} = \Gamma(z_1)f(z) + o_P(1), \tag{A.111}$$

where $\Gamma(z_1) = \mathrm{E}\left[\pi(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)\pi(\mathscr{Z}_t)^\top \middle| Z_{1t} = z_1\right]$; to this end, let

$$\begin{aligned}
\mathtt{S}_{T,3.0} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\hat{\Omega}^{-1}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t}-z_1) \\
&= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\left[\Omega(\mathscr{Z}_t)+\hat{\Omega}(\mathscr{Z}_t)-\Omega(\mathscr{Z}_t)\right]^{-1}\hat{\pi}(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t}-z_1) \\
&= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\left[\Omega(\mathscr{Z}_t)+o_P(1)\right]^{-1}\hat{\pi}(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t}-z_1) \\
&= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)\hat{\pi}(\mathscr{Z}_t)^\top K_{H_2}(Z_{1t}-z_1) + o_P(1) \qquad \text{(A.112)}
\end{aligned}$$

where the third equality comes from the fact that $\hat{\Omega}(\mathscr{Z}_t) - \Omega(\mathscr{Z}_t) \to 0_{m\times m}$ and $T|H_2|^{1/2} \to \infty$, (see, proof of theorem 4.2.3). The rest of the proof is similar to the proof of theorem 4.2.2; therefore we can conclude that

$$\mathtt{S}_{T,3.1} = 0_{dq\times d}, \tag{A.113}$$

and

$$\mathtt{S}_{T,2.2} = \Gamma(z_1)f(z_1) \otimes H_2\mu_2(K) + o_P(H_2). \tag{A.114}$$

Then, by (A.111), (A.113) and (A.114),

$$\mathtt{S}_{T,3}^{-1}(z) = \Gamma^{-1}(z_1)f^{-1}(z_1) \otimes \operatorname{diag}\left\{1+o_P(1), H_2^{-1}\mu_2^{-1}(K) + o_P(H_2^{-1})\right\}. \tag{A.115}$$

Now, we just need to study $\mathtt{T}_{T,2}(z)$; note that by the same reasoning as (A.112) we can write

$$\begin{aligned}
\mathtt{T}_{T,3.0} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)Y_t K_{H_2}(Z_{1t}-z_1) + o_P(1), \\
\mathtt{T}_{T,3.1} &= \frac{1}{T}\sum_{t=1}^{T} \hat{\pi}(\mathscr{Z}_t)\otimes(Z_{1t}-z)\Omega^{-1}(\mathscr{Z}_t)Y_t K_{H_2}(Z_{1t}-z_1) + o_P(1).
\end{aligned}$$

Then, as in the proof of theorem 4.2.2 we center the vector $\mathtt{T}_{T,2}(z)$ by replacing Y_t with $Y_t - \Pi_t(\mathscr{Z}_t)^\top \eta(z_1)$ and observe that

$$\begin{aligned}
\mathtt{T}^*_{T,3}(z_1) &= \frac{1}{T}\sum_{t=1}^{T} \tilde{\Pi}_t(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)\left[Y_t - \Pi_t(\mathscr{Z}_t)^\top \eta(z_1)\right] K_{H_2}(Z_{1t}-z_1) \\
&= \frac{1}{T}\sum_{t=1}^{T} \tilde{\Pi}_t(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)\left[X_t^\top \beta(Z_{1t}) - \Pi_t(\mathscr{Z}_t)^\top \eta(z_1) + u_t\right] K_{H_2}(Z_{1t}-z_1) \\
&= \frac{1}{T}\sum_{t=1}^{T} \tilde{\Pi}_t(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)\left[X_t^\top \beta(Z_{1t}) - \Pi_t(\mathscr{Z}_t)^\top \eta(z_1)\right] K_{H_2}(Z_{1t}-z_1) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T} \tilde{\Pi}_t(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t) u_t K_{H_2}(Z_{1t}-z_1) \\
&= \mathtt{T}_{T,3.1} + \mathtt{T}_{T,3.2}. \qquad \text{(A.116)}
\end{aligned}$$

Following the same lines as in the proof of theorem 4.2.2

$$\begin{aligned}
\mathtt{T}_{T,3.1} &= \begin{pmatrix} \frac{1}{2}\Gamma(z_1)f(z_1)\mu_2(K)\mathrm{diag}_d\left[\mathrm{tr}\left(\mathscr{H}_{\beta_r}(z_1)H_2\right)\right] i_d + o_P(\mathrm{tr}(H_2)) \\ O_P\left(H_2^{3/2}\right) \end{pmatrix} \\
&+ \begin{pmatrix} O_P(\mathrm{tr}\{H_1\}\,\mathrm{tr}\{H_2\}) + o_P\left((T|H_2|^{1/2})^{-1/2}\right) + O_P(\mathrm{tr}\{H_1\}\,\mathrm{tr}\{H_2\}) \\ O_P\left(\mathrm{tr}\{H_1\} H_2^{3/2}\right) + o_P\left((T|H_2|^{1/2})^{-1/2}\right) + O_P\left(\mathrm{tr}\{H_1\} H_2^{3/2}\right) \end{pmatrix} \qquad \text{(A.117)}
\end{aligned}$$

and

$$\begin{aligned}
\mathtt{T}_{T,3.2} &= \frac{1}{T}\sum_{t=1}^{T} \Pi_t(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t) u_t K_{H_2}(Z_{1t}-z_1) \\
&+ \begin{pmatrix} o_P\left((T|H_2|^{1/2})^{-1/2}\right) \\ o_P\left((T|H_2|^{1/2})^{-1/2}\right) \end{pmatrix}. \qquad \text{(A.118)}
\end{aligned}$$

It is clear that $\mathrm{E}\left(\mathtt{T}_{T,3.2}\right) = 0_{d(q+1)}$ and applying **Step 2**

$$T|H_2|^{1/2}\mathrm{Var}\left(\mathtt{T}_{T,2.21}\right) = \Gamma(z_1)f(z_1) \otimes \mathrm{diag}\left\{R(K) + o_P(1), H_2\mu_2^2(K) + o_P(H_2)\right\}. \qquad \text{(A.119)}$$

Now we need to establish the asymptotic normality of

$$\begin{aligned}\sqrt{T|H_2|^{1/2}}i_{d(q+1)}^{\top}\mathrm{T}_{T,3.2} &= \sqrt{T|H_2|^{1/2}}\frac{1}{T}\sum_{t=1}^{T}i_{d(q+1)}^{\top}\Pi_t(\mathscr{Z}_t)\Omega^{-1}(\mathscr{Z}_t)u_tK_{H_2}(Z_{1t}-z_1)+o_P(1)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}V_t(t)+o_P(1),\end{aligned} \tag{A.120}$$

where

$$\mathrm{Var}\left(V_t(t)\right)=i_{d(q+1)}^{\top}\Gamma(z_1)f(z_1)\otimes\mathrm{diag}\left\{R(K)+o_P(1),H_2\mu_2^2(K)+o_P(H_2)\right\}i_{d(q+1)}=\Sigma(z_1)$$

and $\sum_{t=2}^{T}\left|\mathrm{Cov}\left(V_t(1),V_t(t)\right)\right|=o_P(1)$. Note that by construction and assumption 4.2.1 $V_t(t)$ is strictly stationary and α-mixing, thus the large-block and small-block technique can be applied in order to prove the asymptotic normality. To do so, we partition the set $\{1,\ldots,T\}$ into $2k_T+1$ subsets with large blocks of size ℓ_T, small blocks of size s_T and the remaining set of size $T-k_T(\ell_T+s_T)$, where

$$s_T\to\infty,\qquad \frac{s_T}{\ell_T}\to 0,\qquad \frac{\ell_T}{T}\to 0,\qquad \frac{k_Ts_T}{T}\to 0. \tag{A.121}$$

For instance, for any $\phi>2$, $\ell_T=\lfloor(T|H_2|^{1/2})^{\frac{\phi-1}{\phi}}\rfloor$, $s_T=\lfloor(T|H_2|^{1/2})^{\frac{1}{\phi}}/\log T\rfloor$; thus $k_T=o(s_T)$. Then, for $s=1,\ldots,k_T$ define

$$\tilde{V}_s=\sum_{t=(s-1)(\ell_T+s_T)+1}^{s\ell_T+(s-1)s_T}V_T(t);\qquad \bar{V}_s=\sum_{t=s\ell_T+(s-1)s_T+1}^{s(\ell_T+s_T)}V_T(t);\qquad \hat{V}=\sum_{t=k_T(\ell_T+s_T)+1}^{T}V_T(t).$$

Then

$$\begin{aligned}\sqrt{T|H_2|^{1/2}}i_{d(q+1)}^{\top}\mathrm{T}_{T,3.2} &= \frac{1}{\sqrt{T}}\left\{\sum_{s=1}^{k_T}\tilde{V}_s+\sum_{s=1}^{k_T}\bar{V}_s+\hat{V}\right\}\\ &= \frac{1}{\sqrt{T}}\left\{J_{T,1}+J_{T,2}+J_{T,3}\right\},\end{aligned} \tag{A.122}$$

and we will show that as $T\to\infty$

$$\frac{1}{T}\mathrm{E}\left(J_{T,2}\right)^2\to 0,\qquad \frac{1}{T}\mathrm{E}\left(J_{T,3}\right)^2\to 0, \tag{A.123}$$

$$\left|\mathrm{E}\left[\exp\left(J_{T,1}\right)\right]-\prod_{s=1}^{k_T}\mathrm{E}\left[\exp\left(\tilde{V}_s\right)\right]\right|\to 0, \tag{A.124}$$

$$\frac{1}{T}\sum_{s=1}^{k_T}\mathrm{E}\left(\tilde{V}_s^2\right) \to \Sigma(z_1), \tag{A.125}$$

and that for every $\varepsilon > 0$

$$\frac{1}{T}\sum_{s=1}^{k_T}\mathrm{E}\left[\tilde{V}_s^2 I\left\{|\tilde{V}_s| \geq \varepsilon\Sigma^{1/2}(z_1)\sqrt{T}\right\}\right] \to 0. \tag{A.126}$$

Clearly, these four conditions imply that the sum over the small and residual blocks are asymptotically negligible in probability and that $\tilde{V}_s$ are asymptotically independent. Besides, we have the standard Lindeberg-Feller conditions for normality of $T^{-1/2}J_{T,1}$.

We start the analysis by showing (A.123) and (A.125). It is easy to show that

$$\mathrm{E}\left(J_{T,2}^2\right) = \sum_{s=1}^{k_T}\mathrm{Var}\left(\bar{V}_s\right) + 2\sum_{0\leq j\leq s\leq k_T}\mathrm{Cov}\left(\bar{V}_j,\bar{V}_s\right) = J_{3.1} + J_{3.2},$$

and by stationary and assumption 4.2.5,

$$J_{3.1} = k_T\mathrm{Var}\left(\bar{V}_1\right) = k_T\mathrm{Var}\left(\sum_{t=1}^{s_T}V_T(t)\right) = k_T s_T\left[\Sigma(z_1) + o(1)\right],$$

and

$$|J_{3.2}| \leq 2\sum_{s_1=1}^{T-\ell_T}\sum_{s_2=s_1+\ell_T}^{T}|\mathrm{Cov}\left(V_T(s_1),V_t(s_1)\right)| \leq 2T\sum_{s=\ell_T+1}^{T}|\mathrm{Cov}\left(V_T(1),V_t(s)\right)| = o(T),$$

Hence, by (A.121), $k_T s_T = o(T)$, so that $\mathrm{E}\left(J_{T,2}\right)^2 = k_T s_T\Sigma(z_1) + o(T) = o(T)$. Besides, by the stationary condition, (A.121) and (A.119) we get

$$\mathrm{Var}\left(J_{T,3}\right) = \mathrm{Var}\left(\sum_{t=1}^{T-k_T(\ell_T+s_T)}V_T(t)\right) = O\left(T - k_T(\ell_T+s_T)\right) = o(T).$$

From Lemma D.0.2,

$$\left|\mathrm{E}\left[\exp\left(J_{T,1}\right)\right] - \prod_{s=1}^{k_T}\mathrm{E}\left[\exp\left(\tilde{V}_s\right)\right]\right| \leq C(k_T-1)\alpha(s_T) \to 0,$$

which proofs (A.124). Now it remains to prove(A.126)

$$\begin{aligned} \mathrm{E}\left[\tilde{V}_s^2 I\left\{|\tilde{V}_s| \geq \varepsilon\Sigma^{1/2}(z_1)\sqrt{T}\right\}\right] &\leq CT^{-\delta/2}\mathrm{E}\left[|\tilde{V}_s|^{2+\delta}\right] \\ &\leq CT^{-\delta/2}\ell_T^{1+\delta/2}\left[\mathrm{E}|\tilde{V}_s|^{2(1+\delta)}\right]^{(2+\delta)/(2(1+\delta))}, \end{aligned}$$

and it is easy to show that

$$\mathrm{E}|\tilde{V}_s|^{2(1+\delta)} \leq C|H_2|^{-\delta/2},$$

then by the definition of ℓ_T

$$\begin{aligned} \frac{1}{T}\sum_{s=1}^{k_T}\mathrm{E}\left[\tilde{V}_s^2 I\left\{|\tilde{V}_s| \geq \varepsilon\Sigma^{1/2}(z_1)\sqrt{T}\right\}\right] &= O\left(T^{-\delta/2}\ell_T^{\delta/2}|H_2|^{-1/2(2+\delta)\delta/(2(1+\delta))}\right) \\ &= O\left(T^{-\delta/2}(T|H_2|^{1/2})^{\delta/4}|H_2|^{-1/2(2+\delta)\delta/(2(1+\delta))}\right) \\ &= O\left(T^{-\delta/4}|H_2|^{-\frac{1}{2}\left(1+\frac{2}{1+\delta}\right)\frac{\delta}{4}}\right) \to 0, \end{aligned}$$

by assumption 4.2.12. This proves Theorem 4.2.4

Resumen

El objetivo de esta tesis doctoral es aplicar y desarrollar técnicas de inferencia estadística para modelos de coeficientes variables. Por un lado, se investigan técnicas de inferencia estadística basadas en la verosimilitud empírica para modelos de coeficientes variables continuos y discretos, en un contexto de datos de panel con efectos fijos. Primero, demostramos que el ratio de verosimilitud empírica para el coeficiente variable es asintóticamente chi-cuadrado. El ratio es invariable a cambios de escala y no es necesaria la estimación de la varianza. Como subproductos, proponemos los estimadores de máxima verosimilitud empírica de los coeficientes variables. También obtenemos la distribución asintótica de estos estimadores y proponemos algunos procedimientos para calcular los bandwidths empíricamente. Para demostrar la viabilidad de la técnica y analizar sus propiedades en muestras finitas, implementamos un ejercicio de simulación de Monte Carlo, y también proponemos un análisis empírico. Sería interesante ampliar los resultados obtenidos a modelos de coeficientes variables con datos mixtos.

Por otro lado, se propone un test para detectar constancia de parametros en modelos de coeficientes variables. Para regresores exógenos, el procedimiento para relaizar el contraste se asemeja a los contrastes de unión-intersección (U-I) de estabilidad de parámetros en series temporales. El test puede aplicarse para verificar la especificación de modelización de efectos interactivos en modelos de regresión lineales. Debido a que el estadístico de contraste no es asintóticamente pivotal, los valores críticos y los p-valores se estiman utilizando la técnica del bootstrap. Para regresores endógenos, el test se define como un ratio de verosimilitud generalizado que se enfoca en la comparación de la suma de cuadrada de los residuos del modelo restringido y no restringido. Como subproducto, y mimetizando la literatura de variables instrumentales, proponemos utilizar un procedimiento de estimación en tres etapas para estimar los coeficiente variables; también establecemos las propiedades asintóticas de los estimadores. Pra terminar, investigamos las propiedades en muestras finitas de nuestro test por medio de experimentos de Monte Carlo. Aunque los resultados son prometedores, serí interesante ampliar los resultados al marco de datos de panel.