Empirical Likelihood Based Inference for Categorical Varying Coefficient Panel Data Model with Fixed Effects

Luis A. Arteaga-Molina [*]
Departamento de Economía, Universidad de Cantabria

Juan M. Rodriguez-Poo
Departamento de Economía, Universidad de Cantabria

May 22, 2018

Abstract

In this paper local empirical likelihood-based inference for non-parametric categorical varying coefficient panel data models with fixed effects under cross-sectional dependence 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 also a empirical maximum likelihood estimator of the categorical varying coefficient model. We also obtain the asymptotic distribution of this estimator. Furthermore, a non parametric version of the Wilk's theorem is derived. We also illustrated the proposed technique in an application that reports estimates of strike activities from 17 OECD countries for the period 1951 - 1985.

Key Words: Nonparametric regression analysis; Categorical varying coefficient panel data model; Discrete varying coefficient panel data model; fixed effects; empirical likelihood inference.

JEL code: C14, C12, C23

[*]The authors gratefully acknowledge financial support from the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia/Spanish Ministry of Economy and Competitiveness. Ref. ECO2016-76203-C2-1-P. In addition, this work is part of the Research Project APIE 1/2015-17: "New methods for the empirical analysis of financial markets" of the Santander Financial Institute (SANFI) of UCEIF Foundation resolved by the University of Cantabria and funded with sponsorship from Banco Santander.

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, 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 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. 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, for example, Su and Ullah (2011), Rodriguez-Poo and Soberon (2017) and Parmeter and Racine (2018) 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. (2013) 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 in 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. 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. (2013).

As a solution to this problem, in this paper, we propose to use the empirical likelihood techniques to construct confidence intervals/regions. These techniques have acquired importance since they were introduced in Owen (1988) and Owen (1990) because of the advantages over methods such as normal approximation and bootstrap as it combines the reliability of non-parametric methods with the effectiveness of the likelihood approach. Other advantages include, adjustment to the true shape of the underlying distribution, no scale, skewness or limiting variance estimation are necessary, range preserving and transformation respecting among others (see, for example, Hall and La Scala (1990), DiCiccio et al. (1991), Owen (1988) and 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)). For the fixed effect case, in Zhang et al. (2011) and Arteaga-Molina and Rodriguez-Poo (2018) confidence bands based in empirical likelihood techniques are derived for continuous covariates. Unfortunately these type of results are not available for the panel data discrete/categorical varying coefficient setting.

In this paper, 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 non-parametric version of the Wilks' theorem. Besides, we obtain maximum empirical likelihood estimators of the varying parameters. 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 paper is organized as follows. In Section 1 we propose to construct confidence bands for the unknown functions by using what we call a naive empirical likelihood technique. In Section 2, as a by product, we provide an alternative maximum empirical likelihood estimator of the fixed effect categorical varying parameters. In Section 3 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 4 concludes. The proofs of the main results are collected in the Appendix.

1 Naive Empirical Likelihood

We consider the following varying coefficient panel data regression model

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

where Y_{it} is the response, $X_{it} = (X_{it,1}, \dots, X_{it,d})^{\top}$ and $Z_{it} = (Z_{it,1}, \dots, Z_{it,q})^{\top}$ are vectors of dimension d and q respectively, and $\beta(z) = (\beta_1(z), \dots, \beta_d(z))^{\top}$ is a $d \times 1$ vector of unknown functions; here, ω_i stands for the unknown cross-sectional heterogeneity and v_{it} are the random errors. Note that when Z_{it} is a vector of continuous random variables, model (1) stands for the so called varying coefficient panel data model with fixed effects studied by authors such as Rodriguez-Poo and Soberón (2014), Rodriguez-Poo and Soberón (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 (1) is an extension of the cross-sectional varying coefficient model of Li et al. (2013) to the panel data framework as it appears in Feng et al. (2017). Instead of focusing in the consistent estimation of $\beta(z)$, we will obtain confidence bands based in the empirical likelihood principle. As already stated in the introductory section above, this approach presents clear advantages against the standard asymptotically normal approximated confidence bands.

To make the argument for constructing the confidence regions for $\beta(z)$ we can start by noting that, for given z, from (1) we have that

$$E\left[X_{it}\left(Y_{it} - X_{it}^{\top}\beta(Z_{it})\right)\Big| Z_{it} = z\right] \neq 0, \tag{2}$$

because, of the fixed effects. Therefore, the least-squares estimator of $\beta(z)$ would be asymptotically biased. In order to cope with this problem, several transformations have been proposed in the standard literature of panel data models. For example, when Z is continuous, some differencing transformations combined with a Taylor series approximation could be done (see Arteaga-Molina and Rodriguez-Poo (2018)). 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 $1 \leq i,j \leq N$ and $1 \leq t,s \leq T$. Note that $L(Z_{it}, Z_{js}, \gamma)$ represents a kernel function for multivariate discrete spaces

$$L(Z_{it}, z, \gamma) = \prod_{s=1}^{q} l(Z_{it,s}, z_s, \gamma_s) = \prod_{s=1}^{q} \gamma_s^{1(Z_{it,s} \neq z_s)}, \tag{3}$$

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

$$l(Z_{it,s}, z_s, \gamma_s) = \begin{cases} 1 & \text{if } Z_{it,s} = z_s \\ \gamma_s & \text{if } Z_{it,s} \neq z_s \end{cases}, \tag{4}$$

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 (3) can also be expressed as

$$
\begin{aligned}
L(Z_{it}, z, \gamma) &= \prod_{m=1}^{q} l(Z_{it,m}, z_m, \gamma_m) \\
&= \prod_{m=1}^{q} \{1(Z_{it,m} = z_m) + \gamma_m 1(Z_{it,m} \neq z_m)\} \\
&= \prod_{m=1}^{q} 1(Z_{it,m} = z_m) + \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} + \ldots + \prod_{m=1}^{q} \gamma_m 1(Z_{it,m} \neq z_m) \\
&= 1(Z_{it} = z) + \sum_{m=1}^{q} \gamma_m 1_{m,itz^*} + \ldots + \prod_{m=1}^{q} \gamma_m 1(Z_{it,m} \neq z_m) \qquad (5)
\end{aligned}
$$

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 (3) as follows

$$
L(Z_{js}, Z_{it}, \gamma) = 1_{js,it} + \sum_{m=1}^{q} \gamma_m 1_{m,jsit} + O(||\gamma||^2), \qquad (6)
$$

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 (6) is of great interest because it enables us to apply a modified version of a within transformation in (1) and then remove the fixed effects. Thus, let $T_{it} = \sum_{s=1}^{T} L^p_{it,is,\gamma}$, 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 (1) we obtain

$$
\begin{aligned}
\tilde{Y}_{it} &= X_{it}^{\top} \beta(Z_{it}) + \omega_i + v_{it} - \frac{1}{T_{it}} \sum_{s=1}^{T} \left(X_{is}^{\top} \beta(Z_{is}) + \omega_i + v_{is} \right) L^p_{is,it,\gamma} \\
&= X_{it}^{\top} \beta(Z_{it}) - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^{\top} L^p_{is,it,\gamma} \beta(Z_{it}) + \frac{1}{T_{it}} \sum_{s=1} X_{is}^{\top} L^p_{is,it,\gamma} \beta(Z_{it}) - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^{\top} \beta(Z_{is}) L^p_{is,it,\gamma} + \tilde{v}_{it} \\
&= \tilde{X}_{it}^{\top} \beta(Z_{it}) + \varrho_{it} + \tilde{v}_{it}, \qquad (7)
\end{aligned}
$$

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

$$
\left(\beta(Z_{it}) - \beta(Z_{is})\right) L^p_{is,it,\gamma} = O\left(||\gamma||^p\right) \qquad (8)
$$

uniformly. Therefore, the truncation residual ϱ_{it} is controlled by the bandwidth γ only. Given this result we obtain

$$
E\left[\tilde{X}_{it} \left(\tilde{Y}_{it} - \tilde{X}_{it}^{\top} \beta(Z_{it}) \right) \middle| Z_{it} = z \right] = 0. \qquad (9)
$$

In this case, the least squares estimator of $\beta(z)$ is the solution to (9) when $Z_{it} = z$; therefore, the orthogonality condition (9) for $\beta(z)$ has the following form

$$E\left[\tilde{X}_{it}\left(\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\beta(z)\right)\Big|\, Z_{it} = z\right] = 0. \tag{10}$$

Then, in order to define the empirical likelihood estimator we employ the constraint (10). With this, the auxiliary random vector for the modified within transformation is

$$T_i(\beta(z)) = \sum_{t=1}^{T}\tilde{X}_{it}\left(\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\beta(z)\right)L(Z_{it}, z, \gamma). \tag{11}$$

Note that the $T_1(\beta(z)), \ldots, T_N(\beta(z))$ are independent and that $E[T_i(\beta(z))] = 0$. A naive empirical likelihood ratio function for $\beta(z)$ can be defined as the solution to the maximization problem of a multinomial log-likelihood function, i.e.

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

where $p_i = p_i(z)$, $i = 1, \ldots, N$. There exists a unique value of $\mathcal{R}(\beta(z))$, for a given $\beta(z)$, provided that 0 is inside the convex hull of $(T_1(\beta(z)), \ldots, T_N(\beta(z)))$ (Owen (1988) and Owen (1990)). Using the Lagrange multiplier method the probabilities p_i are

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

which after some calculations leads to

$$\mathcal{R}(\beta(z)) = 2\sum_{i=1}^{N}\log(1 + \lambda^{\top}T_i(\beta(z))); \tag{12}$$

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

$$\sum_{i=1}^{N}\frac{T_i(\beta(z))}{1 + \lambda^{\top}T_i(\beta(z))} = 0, \tag{13}$$

subject to the constraint (see Owen (2001), Chapter 3) $\left(1 + \lambda^{\top}T_i(\beta(z))\right) \geq 1/N$. This constraint satisfies the non-negativity condition and it avoids a convex dual problem.

Let us now denote $\tilde{D}(\beta(z)) = (NT)^{-1}\sum_{i=1}^{N}T_i(\beta(z))T_i^{\top}(\beta(z))$. Using equation (12) and a Taylor expansion, it can be shown that

$$\mathcal{R}(\beta(z)) = \left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}T_i(\beta(z))\right]^{\top}\left[\tilde{D}(\beta(z))\right]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}T_i(\beta(z))\right] + o_p(1). \tag{14}$$

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

Assumption 1.1. :

(i) Let $\mathcal{D}$ be the range of values assumed by Z_{it}, then $p(z) = P(Z_{it} = z) > 0 \; \forall z \in \mathcal{D}$. The function $\beta(z)$ is bounded on the support $\mathcal{D}$ of z, i.e. $\max_{z\in\mathcal{D}}||\beta(z)|| < \infty$ and it is not a constant function with respect to z. Denote z_m as the m^{th} component of the q-dimensional vector $z = (z_1, ..., z_q)^\top$, 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 (X_{js}, Z_{js}, v_{js}) is determined by $\alpha_{ij}(|t - s|)$, where

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

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

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

(iv) Denote $\mathcal{X} = \{(X_{js}, Z_{js})\}_{j=1,s=1}^{N,T}$, then $E[v_{it}|\mathcal{X}] = 0$ and $0 < E[v_{it}^2|\mathcal{X}] = \sigma_v^2 < \infty$ almost surely (a.s.) for all $1 \leq i \leq N$ and $1 \leq t \leq T$. For some constants $\delta > 0$ and $0 < a_1 < \infty$, $E[|v_{it}|^{4+\delta} + ||X_{it}||^{4+\delta}] \leq a_1$ uniformly. Also, over the time dimension, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}|E[v_{it}v_{is}|\mathcal{X}]| = O(1)$.

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

Assumption 1.1.(i) is quite standard and similar to Assumption 1.(i) in Li et al. (2013). Note that in order to deal with the case where the cardinality of $\mathcal{D}$ 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 (1) is obtained by normalizing Z_{it} by $u(N, T)$ as follows

$$Y_{it} = X_{it}^\top \beta(Z_{it}/u(N, T)) + \omega_i + v_{it}, \tag{15}$$

were $\beta(.)$ can be treated as a continuous function of covariates; therefore, (15) is just the model proposed by Sun et al. (2009) with $\beta(.)$ 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 1.1.(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 1.1.(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 in this assumption, which is in the same way of Assumption A1 in Chen et al. (2013) and Assumption C in Bai (2009)

Assumption 1.1.(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 1.1.(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 1.1.(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 1.1. *Assuming that conditions 1.1 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* $\mathcal{R}(\beta(z)) \to^d \chi^2_d$. *Here* $\to^d$ *means the convergence in distribution and* χ^2_d *stands for the standard chi-squared distribution with* d *degrees of freedom.*

Note that this result imposes an extra condition on the sequence of bandwidths γ_m, that is $\gamma_m \to 0$ and $\sqrt{NT}\gamma_m \to 0$, which are similar to those 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

$$CV(\gamma) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{Y}_{it} - \tilde{X}_{it}^{\top} \hat{\beta}_{-it}(Z_{it}) \right)^2 \tag{16}$$

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 (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 1.2. :

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

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

where

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

(ii) $\forall z\in\mathcal{D}$, $i=1,\ldots,N$ and $t=1,\ldots,T$, $\Delta_3(z,\gamma)$ and $\Delta_{3\beta}(z,\gamma)$ are uniformly bounded in z. Let us suppose that, together with assumption 1.1(iii)-(iv), the following result holds, $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E||K_{it}||^2=O(1)$ and $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left|\frac{T}{T_{it}}\right|^2=O(1)$ uniformly in $\gamma_m\in[0,1]$ for $m=1,\ldots,q$..

Assumption 1.2.(i) sets restrictions on the unconditional moments as in Assumption 1.1.(iii). Assumption 1.2.(ii) is a panel data version of assumption 2 of Li et al. (2013) 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 (16) converges to zero. UnderAssumptions 1.1 and 1.2 we can state the following results; for further discussion and proofs the reader should refer to Feng et al. (2017).

Lemma 1.1. *Under Assumptions 1.1 and 1.2, 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 (6).

Lemma 1.2. *Assuming that conditions 1.1 and 1.2 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 1.1 will simplify considerably since we will be working with

$$\tilde{T}_i(\beta(z)) = \sum_{t=1}^{T} \tilde{X}_{it}\left(\tilde{Y}_{it} - \tilde{X}_{it}^\top \beta(z)\right) 1(Z_{it} = z) + O_P\left(\frac{1}{NT}\right). \tag{17}$$

Using (17) we can build up an an empirical likelihood ratio function similar to (14), $\tilde{\mathcal{R}}(\beta(z))$ and we can state the following result

Corollary 1.1. *Taking* $\hat{\gamma}$ *to be the minimizer of the cross validation function* (16)*, then assuming that conditions 1.1 and 1.2 hold, and* $(N,T) \to (\infty,\infty)$ *jointly, we get that* $\tilde{\mathcal{R}}(\beta(z)) \to^d \chi^2_d$.

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 Maximum empirical likelihood estimator

We can define the maximum empirical likelihood estimator of $\beta(z)$, $\hat{\beta}(z)$ as the maximizer of $\mathcal{R}(\beta(z))$

$$\hat{\beta}(z) = \max_{\beta(z)} \mathcal{R}(\beta(z)), \tag{18}$$

using (12) and (14) we can write

$$\hat{\beta}(z) \quad = \quad \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}^\top L(Z_{it}, z, \gamma)\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{Y}_{it} L(Z_{it}, z, \gamma) + o_P\left(\frac{1}{\sqrt{NT}}\right) \tag{19}$$

Here, following the same lines as in Qin and Lawless (1994) for the parametric case, the leading term is obtained from the solution of the estimating equations $(NT)^{-1}\sum_{i=1}^{N} T_i(\beta(z)) = 0$ and, as it will be shown in the proof of Theorem 2.1, the remainder term is of smaller order tending to zero as (N,T) tend to infinity jointly. Consequently, the MELE is asymptotically equivalent to the fixed effect estimator using the within transformation. For comparison purposes, and in order to build up confidence bands, in the following theorem we state the asymptotic distribution of the estimator.

Theorem 2.1. *Assuming that conditions 1.1 hold, $\gamma \to 0$ and $(N,T) \to (\infty,\infty)$ jointly, then*

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

where

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

Note that by imposing stronger conditions, 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

Corollary 2.1. *Assuming that conditions 1.1 hold, $\gamma \to 0$, v_{it} are i.i.d. over i and t 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 1.1 and 1.2 the proof of theorem 2.1 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 $\gamma_m = 0\ \forall m = 1, \ldots q$. Therefore, is straightforward to obtain that

$$\sqrt{NT}\left(\hat{\beta}(z) - \beta(z)\right) = \sqrt{NT}\left(\tilde{\beta}(z) - \beta(z)\right) + O_P\left(\frac{1}{\sqrt{NT}}\right); \tag{20}$$

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

Corollary 2.2. *Taking $\hat{\gamma}$ to be the minimizer of the cross validation function* (16)*, then assuming that conditions 1.1 and 1.2 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

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

$$\Gamma_1(z) = p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)^\top\right).$$

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

3 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.

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 to estimate the confidence bands of the parameters of interest. The results are show in Table 1-3; where NUB = Normal Upper Bound, NLB = Normal Lower Bound, LUB = Empirical Likelihood Upper Bound and ELLB = Empirical Likelihood Lower Bound.

Table 1: Confidence bands $\hat{\beta}_1(z)$

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: Confidence bands $\hat{\beta}_2(z)$

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 3: Confidence bands $\hat{\beta}_3(z)$

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

4 Conclusions

Extending Li et al. (2013)'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-squared distribution; therefore we are able to apply empirical likelihood methods to set up confidence regions for the functions of interest. As a by product we provide an alternative empirical maximum likelihood estimator of the categorical varying coefficients. Finally we apply successfully our techniques in a empirical study of estimates of strike activities from 17 OECD countries for the period 1951 - 1985.

5 Appendix

From here on, we will be using the notation that has been defined in the previous Assumptions 1.1 and 1.2 and Theorems 1.1 and 2.1. Also, as in Feng et al. (2017), $O(1)$ denotes some constants which my be different at each appearance.

Proof of Theorem 1.1

Proof. Note that , $\mathcal{R}(\beta(z)) = \left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} T_i(\beta(z))\right]^{\top}\left[\tilde{D}(\beta(z))\right]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} T_i(\beta(z))\right] + o_p(1)$, (see equation (14)) as $(N,T) \to \infty$. Therefore, the proof of this theorem is completed in three steps: first, we show the asymptotic normality of $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}T_i(\beta(z))$, 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 $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}T_i(\beta(z))$ note that

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}T_i(\beta(z)) &= \frac{1}{NT}\sum_{i=1}^{N}\left(T_i(\beta(z)) - E[T_i(\beta(z))|\mathcal{X}]\right) + \frac{1}{NT}\sum_{i=1}^{N}E\left[T_i(\beta(z))|\mathcal{X}\right] \\ &\equiv U_{1NT} + U_{2NT},\end{aligned} \tag{A.1}$$

where $\mathcal{X} = \{(X_{js}, Z_{js})\}_{j=1,s=1}^{N,T}$. 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 (6) and using the same argument we are able to write

$$\begin{aligned} T_{it}^{*} = \sum_{s=1}^{T}1\left(Z_{is}=Z_{it}\right)+O\left(||\gamma||^{p}\right) \quad &, \quad Y_{it}^{*} = Y_{it} - \frac{1}{T_{it}^{*}}\sum_{s=1}^{T}Y_{is}1\left(Z_{is}=Z_{it}\right)+o(1) \\ X_{it}^{*} = X_{it} - \frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1\left(Z_{is}=Z_{it}\right)+o(1) \quad &, \quad v_{it}^{*} = v_{it} - \frac{1}{T_{it}^{*}}\sum_{s=1}^{T}v_{is}1\left(Z_{is}=Z_{it}\right)+o(1) \\ \varrho_{it}^{*} = \frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}^{\top}\left(\beta(Z_{it})-\beta(Z_{is})\right)1\left(Z_{is}=Z_{it}\right)+o(1) \quad &. \end{aligned} \tag{A.2}$$

We first work on the bias term U_{2NT}; then, substituting $T_i\left(\beta(z)\right)$ by (11) into U_{2NT}, applying Assumption 1.1.(iv) and replacing $L(Z_{it}, z, \gamma)$ with (6) and using (A.2), we have

$$\begin{aligned} U_{2NT} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{X}_{it}\left[\tilde{X}_{it}^{\top}\left(\beta(Z_{it})-\beta(z)\right)+\varrho_{it}\right]L_{it,z,\gamma} \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}\left[X_{it}^{*\top}\left(\beta(Z_{it})-\beta(z)\right)+\varrho_{it}^{*}\right]\left(1_{itz}+\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right)+O_p\left(||\gamma||^2\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+O_p\left(||\gamma||^2\right) \end{aligned} \tag{A.3}$$

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, $\left(\beta(Z_{it})-\beta(z)\right)1_{itz} = 0_{d\times 1}$ and $\left(\beta(Z_{it})-\beta(Z_{is})\right)1\left(Z_{is}=Z_{it}\right) = 0_{d\times 1}$; therefore, all the terms containing ϱ_{it}^{*} vanish. We continue the analysis of (A.3); 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 $\mathcal{D}$ is finite.

Step 1: $[0,1]^q$ is a compact subset of $\mathbb{R}^q$ with Euclidean norm $||.||$

Step 2: Rewrite (A.3) as follows

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}X_{it}^{*\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}\right)\left(X_{it}-\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}\right)^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
+ & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s_1=1}^{T}X_{is_1}1_{itis_1}\frac{1}{T_{it}^{*}}\sum_{s_2=1}^{T}X_{is_2}^{\top}1_{itis_2}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
- & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}^{\top}1_{itis}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
- & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*},
\end{aligned}
\tag{A.4}
$$

where $1_{itis}=1(Z_{is}=Z_{it})$. For the last two terms of (A.4), note that we can write

$$
\begin{aligned}
& E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(X_{it})X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\| \\
= & E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_{it}^{*}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\| \le \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left\|K_{it}^{*}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\| \\
\le & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{E\,||K_{it}^{*}||^{2}\,E\left\|X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\|^{2}\right\}^{1/2} \\
\le & \left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\,||K_{it}^{*}||^{2}\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E\left\|X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right\|^{2}\right\}^{1/2}=o_p(||\gamma||)
\end{aligned}
$$

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 $z\in\mathcal{D}$ and $\gamma\in[0,1]^q$

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(Z_{it})\mu(Z_{it})^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+o_p(||\gamma||)
\end{aligned}
$$

Similarly, for the second term of (A.4), we have

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s_1=1}^{T}X_{is_1}1_{itis_1}\frac{1}{T_{it}^{*}}\sum_{s_2=1}^{T}X_{is_2}^{\top}1_{itis_2}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(Z_{it})\mu(Z_{it})^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+o_p(||\gamma||)
\end{aligned}
$$

According to all the above, we obtain

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}X_{it}^{*\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= \quad & \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\mu(Z_{it})\right)\left(X_{it}-\mu(Z_{it})\right)^{\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+o_p(||\gamma||)
\end{aligned}
$$

for any given $z \in \mathcal{D}$ and $\gamma \in [0,1]^q$. We then just need to consider

$$
\begin{aligned}
& E\left|\left|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}-p(z^*)\Sigma_X(z^*)\beta(z^*)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right|\right|^2 \\
= \quad & \frac{1}{(NT)^2}\sum_{h,l=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}E\left[\left(X_{it,h}X_{it,l}\beta_h(Z_{it})\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}-p(z^*)\Sigma_{X,hl}(z^*)\beta_h(z^*)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right)\right. \\
& \left.\times\left(X_{js,h}X_{js,l}\beta_h(Z_{js})\sum_{m=1}^{q}\gamma_m 1_{m,jsz^*}-p(z^*)\Sigma_{X,hl}(z^*)\beta_k(z^*)\sum_{m=1}^{q}\gamma_m 1_{m,jsz^*}\right)\right] \\
\leq \quad & O\left(||\gamma||^2\right)\frac{1}{(NT)^2}\sum_{h,l=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}c_\delta\left(\alpha_{ij}\left(|t-s|\right)\right)^{\frac{\delta}{4+\delta}} \\
\leq \quad & O\left(||\gamma||^2\right)\frac{1}{(NT)^2}\sum_{h,l=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left(\alpha_{ij}\left(|t-s|\right)\right)^{\frac{\delta}{4+\delta}}=O\left(\frac{||\gamma||^2}{NT}\right) \qquad \text{(A.6)}
\end{aligned}
$$

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 $\Sigma_{X,hl}(z^*)$ denotes the $(h,l)^{th}$ element of $\Sigma_X(z^*)$ for $k,l=1,\ldots,d$.

Therefore, we have proved that

$$
\begin{aligned}
& \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}X_{it}^{*\top}\left(\beta(Z_{it})-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
\to_P \quad & p(z^*)\left(\Sigma_X(z^*)-\mu_X(z^*)\mu_X(z^*)^{\top}\right)\left(\beta(z^*)-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
= \quad & \Gamma_1(z^*)\left(\beta(z^*)-\beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}=b(\gamma) \qquad \text{(A.7)}
\end{aligned}
$$

for any given $z \in \mathcal{D}$ and $\gamma \in [0,1]^q$. Therefore, (A.3) 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

$$U_{2NT} = b(\gamma) + O_p\left(||\gamma||^2\right) \tag{A.8}$$

Now we obtain the limiting distribution of the quantity $\sqrt{NT}U_{1NT}$. By substituting (11) into U_{1NT} and replacing $L(Z_{it}, z, \gamma)$ with (6) we obtain

$$\begin{aligned} U_{1NT} &= \frac{1}{NT}\sum_{i=1}^{N}\left(T_i(\beta(z)) - E[T_i(\beta(z))|\mathcal{X}]\right) \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^* v_{it}^* \left(1_{itz} + \sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right) + O_p\left(||\gamma||^2\right), \end{aligned} \tag{A.9}$$

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

$$\begin{aligned} &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^* v_{it}^* 1(Z_{it}=z) \\ = &\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} X_{is}1_{is,it}\right)\left(v_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T} v_{is}1_{is,it}\right)1(Z_{it}=z). \end{aligned} \tag{A.10}$$

Applying **Step 2**, we can write the leading term of $\sqrt{NT}U_{1NT}$ as

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^* v_{it}^* 1(Z_{it}=z) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \mu_X(z)\right)v_{it}1(Z_{it}=z) + o_P\left(1 + ||\gamma||^2\right), \tag{A.11}$$

then we will focus on $\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \mu_X(z)\right)v_{it}1(Z_{it}=z)$. For notational simplicity, denote

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \mu_X(z)\right)v_{it}1(Z_{it}=z) = \sum_{t=1}^{T} V_{T,N}(t). \tag{A.12}$$

By Assumption 1.1.(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 $T - k_T(l_t + s_T)$, where l_T and s_T are selected such that

$$s_T \to \infty, \quad s_T/l_T \to 0 \quad l_T/T \to 0, \quad \text{ and } \quad k_T \equiv [T/(l_T + s_T)] = O(s_T).$$

For instance, for any $\phi > 2$, $l_T = T^{\frac{\phi-1}{\phi}}$, $s_T = T^{\frac{1}{\phi}}$; thus $k_T = O(T^{\frac{1}{\phi}}) = O(s_T)$. For $n = 1, \ldots, k_T$ define

$$\tilde{V}_n = \sum_{t=(n-1)(l_T+s_T)+1}^{nl_T+(n-1)s_T} V_{T,N}(t) \quad , \quad \bar{V}_n = \sum_{t=nl_T+(n-1)s_T+1}^{n(l_T+s_T)} V_{T,N}(t) \quad , \quad \hat{V} = \sum_{t=k_T(l_T+s_T)+1}^{T} V_{T,N}(t).$$

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 $E\left|\left|\hat{V}\right|\right|^2 = O\left(\frac{T-k_T l_T}{T}\right) = o(1)$. 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

$$\left| E\left[\exp\left\{\sum_{n=1}^{k_T} ||\tilde{V}_n||\right\}\right] - \prod_{n=1}^{k_T} E\left[\exp\left\{||\tilde{V}_n||\right\}\right]\right| \leq C(k_T - 1)\alpha(s_T) \to 0, \tag{A.14}$$

where C is a constant and $\alpha(.)$ is the upper bounded of the α-mixing coefficient defined in Assumption 1.1.(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)

$$\begin{aligned}
\mathrm{Cov}\left[\tilde{V}_1\right] &= \mathrm{Cov}\left[\sum_{t=1}^{l_T} V_{N,T}(t)\right] = \mathrm{Cov}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{l_T}\left(X_{it} - \mu_X(z)\right) v_{it} 1(Z_{it} = z)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{l_T} \mathrm{Cov}\left[\left(X_{it} - \mu_X(z)\right) v_{it} 1(Z_{it} = z)\right] = \frac{l_T}{T}\Gamma_0(z)\left[I_d + o(1)\right],
\end{aligned} \tag{A.15}$$

which implies that

$$\sum_{n=1}^{k_T} \mathrm{Cov}\left[\tilde{V}_n\right] = k_T \mathrm{Cov}\left[\tilde{V}_1\right] = \frac{k_T l_T}{T}\Gamma_0(z)\left[I_d + o(1)\right] \to \Gamma_0(z). \tag{A.16}$$

As a result, the Feller condition is satisfied. Now we just need to check the Lindeberg condition

$$\sum_{n=1}^{k_T} E\left[\left|\left|\tilde{V}_n\right|\right|^2 I\left\{\left|\left|\tilde{V}_n\right|\right| \geq \varepsilon\right\}\right] \to^P 0 \tag{A.17}$$

where $\varepsilon > 0$. Using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
E\left[\left|\left|\tilde{V}_n\right|\right|^2 I\left\{\left|\left|\tilde{V}_n\right|\right| \geq \varepsilon\right\}\right] &\leq \left\{E\left|\left|\tilde{V}_n\right|\right|^3\right\}^{2/3}\left\{P\left(\left|\left|\tilde{V}_n\right|\right| \geq \varepsilon\right)\right\}^{1/3} \\
&\leq C\left\{E\left|\left|\tilde{V}_n\right|\right|^3\right\}^{2/3}\left\{E\left|\left|\tilde{V}_n\right|\right|^2\right\}^{1/3},
\end{aligned} \tag{A.18}$$

and by Lemma B.2 in Chen et al. (2012a)

$$E\left|\left|\tilde{V}_n\right|\right|^3 \leq \left(\frac{l_T}{T}\right)^{3/2}\left\{E\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(X_{i1} - \mu_X(z)\right) v_{i1} 1(Z_{i1} = z)\right|\right|^4\right\}^{3/4} < \infty, \tag{A.19}$$

$$E\left|\left|\tilde{V}_n\right|\right|^2 \leq \left(\frac{l_T}{T}\right)\left\{E\left|\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(X_{i1} - \mu_X(z)\right) v_{i1} 1(Z_{i1} = z)\right|\right|^4\right\}^{1/2} < \infty. \tag{A.20}$$

Thus, $E\left|\left|\tilde{V}_n\right|\right|^3 = O\left(\left(\frac{l_T}{T}\right)^{3/2}\right)$ and , $E\left|\left|\tilde{V}_n\right|\right|^2 = O\left(\frac{l_T}{T}\right)$ which, using (A.18), implies

$$E\left[\left|\left|\tilde{V}_n\right|\right|^2 I\left\{\left|\left|\tilde{V}_n\right|\right| \geq \varepsilon\right\}\right] \leq O\left(\left(\frac{l_T}{T}\right)^{4/3}\right) = o\left(\frac{l_T}{T}\right), \quad \text{(A.21)}$$

therefore,

$$\sum_{n=1}^{k_T} E\left[\left|\left|\tilde{V}_n\right|\right|^2 I\left\{\left|\left|\tilde{V}_n\right|\right| \geq \varepsilon\right\}\right] = o\left(\frac{k_T l_T}{T}\right) = o(1). \quad \text{(A.22)}$$

Consequently, the Lindeberg condition is satisfied; using (A.7), (A.14), (A.16) and (A.22) it is east to see that if $\gamma_m \to 0$ we can conclude that

$$\sqrt{NT}U_{1NT} \to^d \mathcal{N}(0, \Gamma_0(z)). \quad \text{(A.23)}$$

as N and T tend to infinity.

Now we prove the consistency of $\tilde{D}(\beta(z))$. Similar to the proof of (A.12) - (A.16), it is straightforward to show that

$$\tilde{D}_w(\beta(z)) = \frac{1}{NT}\sum_{i=1}^{N} T_i(\beta(z))T_i^\top(\beta(z)) = \Gamma_0(z)\left(I_d + o_p(1)\right). \quad \text{(A.24)}$$

From (A.7), (A.23) and (A.24), and using the same arguments as in the proof of (2.14) in Owen (1990), we can prove that

$$\lambda = O_p\left((NT)^{-1/2}\right), \quad \text{(A.25)}$$

where λ was defined in (13). Then applying Taylor expansion to (12) and invoking (A.7), (A.23) and (A.24), we obtain

$$\mathcal{R}(\beta(z)) = 2\sum_{i=1}^{N}\left[T_i^\top(\beta(z))\lambda - \left(T_i^\top(\beta(z))\lambda\right)^2/2\right] + o_p(1). \quad \text{(A.26)}$$

By (13) and applying Taylor expansion again it follows that

$$\begin{aligned} 0 &= \sum_{i=1}^{N}\frac{T_i(\beta(z))}{1+\lambda^\top T_i(\beta(z))} \\ &= \sum_{i=1}^{N} T_i(\beta(z)) - \sum_{i=1}^{N} T_i(\beta(z))T_i^\top(\beta(z))\lambda + \sum_{i=1}^{N}\frac{T_i(\beta(z))(T_i^\top(\beta(z))\lambda)^2}{1+T_i^\top(\beta(z))} \end{aligned}$$

Then, recalling (A.7), (A.23) and (A.24) we can prove that

$$\sum_{i=1}^{N}(T_i^\top(\beta(z))\lambda)^2 = \sum_{i=1}^{N} T_i^\top(\beta(z))\lambda + o_p(1), \quad \text{(A.27)}$$

and

$$\lambda = \left[\sum_{i=1}^{N} T_i(\beta(z))T_i^{\top}(\beta(z))\right]^{-1} \sum_{i=1}^{N} T_i(\beta(z)) + o_p\left((NT)^{-1/2}\right). \tag{A.28}$$

Now, if we rely on (A.7), (A.23) and (A.24) the proof is concluded by applying the Cramer-Wold device.

□

Proof of Theorem 2.1

Proof. Note that, without loss of generality, we are able to write

$$\hat{\beta}(z) - \beta(z) = \left(\hat{\beta}(z) - E\left[\hat{\beta}(z)\middle|\mathcal{X}\right]\right) + \left(E\left[\hat{\beta}(z)\middle|\mathcal{X}\right] - \beta(z)\right) \equiv \mathbf{I}_{1NT} + \mathbf{I}_{2NT} \tag{A.29}$$

To prove the desired result, under assumption 1.1, we will show first that $\mathbf{I}_{2NT} = \Gamma^{-1}(z)b(\gamma)$ and second that $\sqrt{NT}\mathbf{I}_{1NT} \to_d \mathcal{N}\left(0_{d\times 1}, \Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)\right)$, as (N,T) tend to infinity jointly and $\gamma_s \to 0$. If we substitute (19) into (A.29) we obtain

$$\begin{aligned} \mathbf{I}_{2NT} &= E\left[\hat{\beta}(z)\middle|\mathcal{X}\right] - \beta(z) \\ &= \left(\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\right)^{-1}\left(\frac{1}{NT}\sum_{it}\tilde{X}_{it}\left[\tilde{X}_{it}^{\top}\beta(Z_{it}) + \varrho_{it} - \beta(z)\right]L(Z_{it},z,\gamma)\right). \end{aligned} \tag{A.30}$$

We begin the analysis with the inverse term of (A.30) and by replacing $L(Z_{it},z,\gamma)$ with (6) and using (A.6)-(A.7) we obtain

$$\begin{aligned} \frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma) &= \frac{1}{NT}\sum_{it}X_{it}^{*}X_{it}^{*\top}1_{itz} + O_p\left(||\gamma||\right) \\ &\to_P p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)^{\top}\right) + O_p\left(||\gamma||\right). \end{aligned} \tag{A.31}$$

Then, using (A.31) we have proved that

$$\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma) \to_P p(z)\left(\Sigma_X(z) - \mu_X(z)\mu_X(z)^{\top}\right) + O_p\left(||\gamma||\right) = \Gamma_1(z). \tag{A.32}$$

Now we continue with the second term of (A.30) and using (A.3)-(A.8) we obtain that

$$\begin{aligned} \frac{1}{NT}\sum_{it}\tilde{X}_{it}\left[\tilde{X}_{it}^{\top}\beta(Z_{it}) + \varrho_{it} - \beta(z)\right]L(Z_{it},z,\gamma) &\to_P \Gamma_1(z^*)\left(\beta(z^*) - \beta(z)\right)\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} + O_p\left(||\gamma||^2\right) \\ &= b(\gamma). \end{aligned} \tag{A.33}$$

In order to show the asymptotic behavior of $\mathbf{I}_{1N}$ note that by (19) we have that

$$\mathbf{I}_{1NT} = \hat{\beta}(z) - E\left[\hat{\beta}(z)\middle|\mathbb{X},\mathbb{Z}\right] = \left(\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\right)^{-1}\left(\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{v}_{it}L(Z_{it},z,\gamma)\right), \tag{A.34}$$

where the inverse term was already study (see (A.32)); therefore, we will study the asymptotic behaviour of (A.34) by studying the behaviour of the second term. Based on the results obtained in (A.9) - (A.22) in the proof of theorem 1.1 and (A.32) the following result holds

$$\sqrt{NT}\mathbf{I}_{1NT} \to_d \mathcal{N}\left(0_{d\times 1}, \Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)\right),$$

and the proof is closed. □

Proof of Corollary 1.1

Proof. From equation (17) we know that the auxiliary random vector $T_i\left(\beta(z)\right) = \tilde{T}_i(\beta(z))$, 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 1.1 is similar to the proof of Theorem 1.1 but setting $\gamma_m = 0$ for $m = 1, \ldots, q$. □

Proof of Corollary 2.2

Proof. From equation (20) 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 2.2 is similar to the proof of Theorem 2.1 but setting $\gamma_m = 0$ for $m = 1, \ldots, q$. □

References

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

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

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

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

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

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

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

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

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

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

Feng, G., J. Gao, B. Peng, and X. Zhang (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.

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

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

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge Univ Press.

Li, G., H. Peng, and T. Tong (2013). Simultaneous confidence band for nonparametric fixed effects panel data models. *Economics Letters 119*(3), 229–232.

Li, Q., D. Ouyang, and J. S. Racine (2013). Categorical semiparametric varying-coefficient models. *Journal of Applied Econometrics 28*(4), 551–579.

Li, Q. and J. S. Racine (2007). *Nonparametric econometrics: theory and practice.* Princeton University Press.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics 18*(1), 90–120.

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 J. S. Racine (2018, January). Nonparametric estimation and inference for panel data models. Department of Economics Working Papers 2018-02, McMaster University.

Pesaran, M. H. and E. Tosetti (2011). Large panels with common factors and spatial correlation. *Journal of Econometrics 161*(2), 182–202.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Rodriguez-Poo, J. M. and A. Soberón (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 A. Soberón (2015). Nonparametric estimation of fixed effects panel data varying coefficient models. *Journal of Multivariate Analysis 133*, 95–122.

Rodriguez-Poo, J. M. and A. Soberon (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys 31*(4), 923–960.

Su, L. and A. Ullah (2011). Nonparametric and semiparametric panel econometric models: estimation and testing. *Handbook of Empirical Economics and Finance. Chapman and Hall/CRC*, 455–497.

Sun, Y., R. J. Carroll, and D. Li (2009). Semiparametric estimation of fixed effects panel data varying coefficient models. *Advances in Econometrics 25*, 101–129.

Western, B. (1996). Vague theory and model uncertainty in macrosociology. *Sociological Methodology 26*(1), 165–192.

Xue, L. and L. Zhu (2007). Empirical likelihood for a varying coefficient model with longitudinal data. *Journal of the American Statistical Association 102*(478), 642–654.

Zhang, J., S. Feng, G. Li, and H. Lian (2011). Empirical likelihood inference for partially linear panel data models with fixed effects. *Economics Letters 113*, 165–167.