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

Luis A. Arteaga-Molina[a], Juan M. Rodríguez-Poo[b,*]

[a]*Departamento de Economía, Universidad de Cantabria*
[b]*Departamento de Economía, Universidad de Cantabria*

Abstract

In this paper local empirical likelihood-based inference for nonparametric 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 using a nonparametric version of Wilk's theorem. 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 an empirical maximum likelihood estimator of the categorical varying coefficient model 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 OECD countries for the period 1951–85.

Keywords: Categorical varying-coefficient panel data model, Discrete varying-coefficient panel data model, Fixed effects, Empirical likelihood inference, Nonparametric regression analysis.
2010 MSC: Primary 62G10, Secondary 62G05

1. Introduction

In recent years, there has been an increased interest in the study of panel data models combined with nonparametric techniques. The results have been promising, even though the inherent disadvantages of nonparametric techniques such as the curse of dimensionality [16] remain valid in this context. Varying-coefficient models appear as a reasonable avenue to overcome this drawback.

Varying-coefficient models encompass a great variety of simple models applied by econometricians, including partially linear models or fully nonparametric models. In applied microeconomic problems, however, it is often difficult to access all explanatory variables of interest. For this reason, many applied economists have turned their attention to 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 at our disposal 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, e.g.,[28, 33, 34] for comprehensive surveys of the literature. Among others, the problem of considering varying coefficients that depend on discrete data has attracted attention because discrete variables are common in economic analysis. A semiparametric varying coefficient model with purely categorical covariates is proposed in [20] and in [12], this setting is extended to include fixed effects and cross-sectional dependence.

Although extensive results are reported, e.g., in [12, 20] on the asymptotic behavior of estimator, inference is not always an easy task. Typically, asymptotic normal approximations are obtained. In the discrete covariate case, under fairly general conditions, if the bandwidth is selected using a cross-validation criterion, 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 [21]. Unfortunately, the problem becomes much more complex if one also wishes to incorporate cross-sectional dependence. Besides, using confidence bands as a testing device is not straightforward as uniform confidence bands are necessary to do so; see [18].

*Corresponding author. Email address: juan.rodriguez@unican.es

As an alternative, Owen [26] introduced techniques based on the empirical likelihood. This approach, which combines the reliability of nonparametric methods with the effectiveness of the likelihood approach, has several advantages. For instance, no limiting variance estimation is necessary. For further discussion on the advantages of the empirical likelihood technique, see, e.g., [10, 14, 15, 17, 19, 24–27, 30].

Owing to its good properties, the empirical likelihood approach has already been applied to longitudinal data varying-coefficient models with random effects; see, e.g., [37]. As for the fixed-effects case, see [3, 38]. However, we are not aware of any results for the panel data discrete/categorical varying coefficient setting. In [3], empirical likelihood confidence bands are obtained for the varying coefficients, $m(Z)$, under rather strong assumptions such as the continuity of all the vector of covariates, Z, and the assumption of independent and identically distributed idiosyncratic error terms both across units and along time. Although the kernel weights considered in this paper are well suited for continuous data, they are inappropriate for discrete/categorical data. Furthermore, the authors derive the asymptotic theory for T fixed and $N \to \infty$.

In this paper, we develop empirical likelihood ratios and derive a nonparametric version of Wilks' theorem for a fixed-effects varying-coefficient panel data model, where all covariates are assumed to be discrete/categorical. We further derive the maximum empirical likelihood estimator of the varying parameters and its asymptotic theory when cross-sectional dependence in the idiosyncratic error term is allowed. 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 paper is organized as follows. In Section 2 we propose to construct confidence bands for the unknown functions by using a naive empirical likelihood technique. In Section 3, as a by-product, we provide an alternative maximum empirical likelihood estimator of the fixed-effect categorical varying parameters. In Section 4, we illustrate the proposed technique in an application that reports estimates of strike activities from 17 OECD countries for the period 1951–85. Concluding remarks are in Section 5 and the proofs of the main results are in the Appendix.

2. Naive empirical likelihood

We consider the following categorical varying-coefficient panel data regression model

$$Y_{it} = X_{it}^\top \beta(Z_{it}) + \omega_i + v_{it}, \tag{1}$$

where for each $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, Y_{it} is the response, $X_{it} = (X_{it,1}, \ldots, X_{it,d})^\top$ and $Z_{it} = (Z_{it,1}, \ldots, Z_{it,q})^\top$ are vectors of dimension d and q respectively, and $\beta = (\beta_1, \ldots, \beta_d)^\top$ 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 (1) stands for the so-called varying-coefficient panel data model with fixed effects studied e.g., in [6, 9, 31, 32, 34, 35]. In this paper we consider the case where Z is purely categorical and in order to distinguish between X and Z, we will refer to them as the regressor and the covariate, respectively. Note that we are not willing to impose any restriction between ω_i and the pair (X_{it}, Z_{it}).

Model (1) is an extension of the cross-sectional varying-coefficient model of Li et al. [20] to the panel data framework as it appears in [1[illegible]]. First, we will obtain confidence bands for β based on the empirical likelihood approach; to do so, we need the first order condition of the minimization problem for obtaining β. Note that for given z, this condition is, from (1),

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

due to the fixed effects. To deal with this problem, several transformations have been proposed in the literature on panel data models. For example, when Z is continuous, some differencing transformations combined with a Taylor series approximation could be applied; see [3]. Unfortunately, this approach is infeasible if the elements of Z are discrete in nature.

We here 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 [1]. Thus, let $1_{js,it} = \mathbf{1}(Z_{it} = Z_{js})$ and $L_{js,it,\gamma} = L(Z_{it}, Z_{js}, \gamma)$ for all $i, j \in \{1, \ldots, N\}$ and $s, t, \in \{1, \ldots, T\}$. Note that $L(Z_{it}, Z_{js}, \gamma)$ represents a kernel function for multivariate discrete spaces, viz.

$$L(Z_{it}, z, \gamma) = \prod_{s=1}^{q} \ell(Z_{it,s}, z_s, \gamma_s) = \prod_{s=1}^{q} \gamma_s^{\mathbf{1}(Z_{it,s} \neq z_s)}, \tag{2}$$

where $\gamma = (\gamma_1, \ldots, \gamma_q)^\top$, $\mathbf{1}(Z_{it,s} \neq z_s)$ 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 [1] for unordered covariates, where $\gamma_s = 0$ leads to an indicator function and $\gamma_s = 1$ gives a uniform weighted function. We can then conclude that $\gamma_s \in [0, 1]$ for all $s \in \{1, \ldots, q\}$. Also, note that the kernel function (2) can be expressed as

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

where $1_{m,itz^*} = \mathbf{1}(Z_{it,m} \neq z_m) \prod_{n=1, n\neq m}^{q} \mathbf{1}(Z_{it,n} = z_n)$ is an indicator function which takes value 1 if Z_{it} and z differs only in their mth 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) as follows:

$$L(Z_{js}, Z_{it}, \gamma) = 1_{js,it} + \sum_{m=1}^{q} \gamma_m 1_{m,jsit} + O(\|\gamma\|^2). \tag{3}$$

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

Expression (3) 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 an arbitrarily chosen finite positive integer. In practice, the choice of $p = 2$ is enough. Let

$$\tilde{X}_{it} = X_{it} - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is} 1_{is,it}, \quad \tilde{Y}_{it} = Y_{it} - \frac{1}{T_{it}} \sum_{s=1}^{T} Y_{is} 1_{is,it}, \quad \tilde{v}_{it} = v_{it} - \frac{1}{T_{it}} \sum_{s=1}^{T} v_{is} 1_{is,it}.$$

Applying this transformation in (1), we obtain

$$\begin{aligned} \tilde{Y}_{it} &= X_{it}^\top \beta(Z_{it}) + \omega_i + v_{it} - \frac{1}{T_{it}} \sum_{s=1}^{T} \{X_{is}^\top \beta(Z_{is}) + \omega_i + v_{is}\} L^p_{is,it,\gamma} \\ &= X_{it}^\top \beta(Z_{it}) - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^\top L^p_{is,it,\gamma} \beta(Z_{it}) + \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^\top L^p_{is,it,\gamma} \beta(Z_{it}) - \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is}^\top \beta(Z_{is}) L^p_{is,it,\gamma} + \tilde{v}_{it} \\ &= \tilde{X}_{it}^\top \beta(Z_{it}) + \varrho_{it} + \tilde{v}_{it}, \end{aligned} \tag{4}$$

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

$$\{\beta(Z_{it}) - \beta(Z_{is})\} L^p_{is,it,\gamma} = O(\|\gamma\|^p) \tag{5}$$

uniformly. Therefore, due to (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 (4) is

$$\mathrm{E}[\tilde{X}_{it}\{\tilde{Y}_{it} - \tilde{X}_{it}^\top \beta(Z_{it})\} | Z_{it} = z] = 0. \tag{6}$$

In this case, the least squares estimator of $\beta(z)$ is the solution to (6) when $Z_{it} = z$. Therefore, the orthogonality condition (6) for $\beta(z)$ has the following form:

$$\mathrm{E}[\tilde{X}_{it}\{\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\beta(z)\}|Z_{it} = z] = 0. \tag{7}$$

Then, employing the constraint (7), the auxiliary random vector for the modified within-transformation is

$$T_i\{\beta(z)\} = \sum_{t=1}^{T} \tilde{X}_{it}\{\tilde{Y}_{it} - \tilde{X}_{it}^{\top}\beta(z)\}L(Z_{it}, z, \gamma). \tag{8}$$

Eq. (8) is the sample analog of (7) using a local smoothing method with a discrete kernel function. If $\beta(z)$ is the true parameter, it is easy to show, due to (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

$$\mathcal{R}\{\beta(z)\} = -2\max\left[\sum_{i=1}^{N}\ln(p_i) : p_i \geq 0,\ \sum_{i=1}^{N} p_i = 1,\ \sum_{i=1}^{N} p_i T_i\{\beta(z)\} = 0\right], \tag{9}$$

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

$$p_i = \frac{1}{N}\frac{1}{1 + \lambda^{\top}T_i\{\beta(z)\}}. \tag{10}$$

By (9) and (10), $\mathcal{R}\{\beta(z)\}$ leads to

$$\mathcal{R}\{\beta(z)\} = 2\sum_{i=1}^{N}\ln[1 + \lambda^{\top}T_i\{\beta(z)\}], \tag{11}$$

where λ is a $d \times 1$ vector of Lagrange multipliers associated to the constraint $\sum_{i=1}^{N} p_i T_i\{\beta(z)\} = 0$ and it is given by

$$\sum_{i=1}^{N}\frac{T_i\{\beta(z)\}}{1 + \lambda^{\top}T_i\{\beta(z)\}} = 0, \tag{12}$$

subject to the constraint that satisfies the non-negativity condition and avoids a convex dual problem; see Chapter 3 in [27]. Using Eqs. (11)–(12) and a Taylor expansion, and denoting

$$\tilde{D}\{\beta(z)\} = \frac{1}{NT}\sum_{i=1}^{N} T_i\{\beta(z)\}\, T_i^{\top}\{\beta(z)\},$$

it can be shown that

$$\mathcal{R}\{\beta(z)\} = \left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} T_i\{\beta(z)\}\right]^{\top}[\tilde{D}\{\beta(z)\}]^{-1}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N} T_i\{\beta(z)\}\right] + o_p(1). \tag{13}$$

Hence, it is easy to show using (13) that $\mathcal{R}\{\beta(z)\}$ is asymptotically chi-square. In order to formally introduce this result, we need the following assumptions.

Assumption 1. (i) Let $\mathcal{D}$ be the range of values assumed by Z_{it}. Then $p(z) = \Pr(Z_{it} = z) > 0$ for all $z \in \mathcal{D}$. 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 constant with respect to z. Let z_m denote the mth component of the $q \times 1$ vector $z = (z_1, \ldots, z_q)^{\top}$, where z_m is assume to take c_m different integer values in $\{0, \ldots, c_m - 1\}$ for $c_m \geq 2$ and $m \in \{1, \ldots, q\}$. Moreover, q is finite and $\max(c_1, \ldots, c_q) < \infty$.

(ii) Let (X_{it}, Z_{it}, v_{it}) be independent across i for each fixed t. 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 for each integer $k \geq 1$.

$$\alpha(k) = \sup_{\substack{A \in \sigma\{(X_{is}, Z_{is}, v_{is}) \,:\, s \leq t\} \\ B \in \sigma\{(X_{is}, Z_{is}, v_{is}) \,:\, s \geq t+k\}}} |\Pr(A \cap B) - \Pr(A)\Pr(B)|,$$

Furthermore, for some $\delta > 0$,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} \{\alpha_{ij}(|t-s|)\}^{\delta/(4+\delta)} = O(NT).$$

(iii) For all $z \in \mathcal{D}$, $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, $\mu_X(z) = \mathrm{E}(X_{it}|Z_{it} = z)$, and $\Sigma_X(z) = \mathrm{E}(X_{it}X_{it}^{\top}|Z_{it} = z)$, where $\|\mu_X(z)\|$ and $\|\Sigma_X(z)\|$ are uniformly bounded in z.

(iv) Denote $\mathcal{X} = \{X_{js}, Z_{js}) : j \in \{1, \ldots, N\}, s \in \{1, \ldots, T\}\}$. Then $\mathrm{E}(v_{it}|\mathcal{X}) = 0$ and $0 < \mathrm{E}(v_{it}^2|\mathcal{X}) = \sigma_v^2 < \infty$ almost surely (a.s.) for all $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$. For some constants $\delta > 0$ and $0 < a_1 < \infty$, $\mathrm{E}(|v_{it}|^{4+\delta} + \|X_{it}\|^{4+\delta}) \leq a_1$ uniformly. Also, over the time dimension,

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s=1}^{T} |\mathrm{E}(v_{it} v_{is}|\mathcal{X})| = O(1).$$

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

Assumption 1.(i) is quite standard and similar to Assumption 1.(i) in [20]. Note that to deal with the case where $\mathcal{D}$ is infinite, one can use the same normalization as for the time-varying coefficient model. That is, as in [12], 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 $u(N, T)$, viz.

$$Y_{it} = X_{it}^{\top}\beta\{Z_{it}/u(N, T)\} + \omega_i + v_{it}, \tag{14}$$

where β can be treated as a continuous function of covariates. Therefore, (14) is just the model proposed by Sun et al. [35] with continuous β. This normalization is similar to that of [5, 8] when dealing with time-varying coefficients.

Assumptions 1.(ii) is similar to Assumptions B–C in [4]. The strict stationary assumption is similar to Assumption A4 in Chen et al. [7] and Assumption A2 in Chen et al. [8]. For more details and discussion, see [12].

Assumption 1.(iii) sets restrictions on the unconditional moments as in Assumption 3.3–3.6 of [31]. Due to the within-transformation, we must assume it holds uniformly across i, which is akin to Assumption A1 in [9] and Assumption C in [4]

Assumption 1.(iv) is the same as in [2] and similar to Assumptions A2 and A4 of [8]. 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; see, e.g., [7, 8, 29]. Here, the last equation in Assumption 1.(iv) is a simplified version of (A.18) in [7]; this last equation is needed due to the within-transformation.

Finally, Assumption 1.(v) imposes the so-called fixed effects. Note that we are unwilling 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. *Assume that Condition* 1 *holds and that* $\gamma_m \to 0$ *and* $\sqrt{NT}\gamma_m \to 0$ *for all* $m \in \{1, \ldots, q\}$ *as* $(N, T) \to (\infty, \infty)$. *Then* $\mathcal{R}\{\beta(z)\} \rightsquigarrow \chi_d^2$, *i.e.,* $\mathcal{R}\{\beta(z)\}$ *converges in law to a chi-square random variable with d degrees of freedom.*

Letting c_α stand for the $1 - \alpha$ quantile of χ_d^2, we can then build the confidence bands using Theorem 1 as follows:

$$R_\alpha = \{\beta(z) : \mathcal{R}\{\beta(z)\} \leq c_\alpha\}. \tag{15}$$

Note that this result imposes an extra condition on the sequence of bandwidths γ_m, i.e., $\sqrt{NT}\gamma_m \to 0$, which is similar to conditions used in nonparametric regression. As is well known, the latter condition implies that the rate of convergence is not optimal. As mentioned, e.g., in [21], in the presence of discrete covariates it is possible to improve the rate of convergence by selecting $\gamma_1, \ldots, \gamma_q$ to be the minimizer of the cross-validation (CV) criterion function

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

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 [12].

Assumption 2. (i) Set

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

where

$$\Delta_1(z, \gamma) = \mathrm{E}\{L^p(Z_{is}, z, \gamma) | z, \gamma\}, \quad \Delta_2(z, \gamma) = \mathrm{E}\{X_{it} L^p(Z_{is}, z, \gamma) | z, \gamma\},$$

$$\Delta_{2\beta}(z, \gamma) = \mathrm{E}\{X_{it} \beta(Z_{it}) L^p(Z_{is}, z, \gamma) | z, \gamma\}, \quad \Delta_3(z, \gamma) = \Delta_2(z, \gamma)/\Delta_1(z, \gamma), \quad \Delta_{3\beta}(z, \gamma) = \Delta_{2\beta}(z, \gamma)/\Delta_1(z, \gamma),$$

$$\Omega(z, \gamma) = \Sigma_X(z) + \Delta_3(z, \gamma)\Delta_3(z, \gamma)^\top - \Delta_3(z, \gamma)\mu_X(z)^\top - \mu_X(z)\Delta_3(z, \gamma),$$

$$\Sigma_{XX}(z, \gamma) = \mathrm{E}\{\Omega(z, \gamma) L(Z_{it}, z, \gamma) | z, \gamma\}, \quad \Sigma_{XX\beta}(z, \gamma) = \mathrm{E}\{\Omega(z, \gamma) \beta(Z_{it}) L(Z_{it}, z, \gamma) | z, \gamma\}$$

$$\eta(z, \gamma) = \Sigma_{XX}^{-1}(z, \gamma) \Sigma_{XX\beta}(z, \gamma), \quad K_{it} = \frac{1}{T_{it}} \sum_{s=1}^{T} X_{is} L^p(Z_{is}, z, \gamma) - \Delta_3(Z_{it}, \gamma).$$

(ii) For all $z \in \mathcal{D}$, $i \in \{1, \ldots, N\}$ and $t \in \{1, \ldots, T\}$, $\Delta_3(z, \gamma)$ and $\Delta_{3\beta}(z, \gamma)$ are uniformly bounded in z. Suppose that, together with Assumption 1.(ii)–(iv), one has

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathrm{E}\|K_{it}\|^2 = O(1), \quad \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} |T/T_{it}|^2 = O(1)$$

uniformly in $\gamma_m \in [0, 1]$ for all $m \in \{1, \ldots, q\}$.

Assumption 2.(i) sets restrictions on the unconditional moments as in Assumption 1.(iii). Assumption 2.(ii) is a panel data version of Assumption 2 of [20] which ensures that $CV_0(\gamma)$ is uniquely optimize at 0. By Theorem 2.1 of Newey and McFadden [22], this assumption implies that $\hat{\gamma}$ obtained by minimizing (16) converges to zero. Under Assumptions 1–2 we can state the following results; for further discussion and proofs, refer to [12].

Lemma 1. *Under Assumptions 1–2,* $\hat{\gamma} = o_P(1)$ *as* $(N, T) \to (\infty, \infty)$.

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 (3).

Lemma 2. *If Conditions 1–2 hold,* $\hat{\gamma} = O_P\{1/(NT)\}$ *as* $(N, T) \to (\infty, \infty)$.

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, viz. $L(Z_{it}, z, \gamma) = \mathbf{1}(Z_{it} = z)$, letting $\gamma = 0_{q\times 1}$. Further note that using these results, the proofs of Theorem 1 will simplify considerably since we will be working with

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

Using (17), we can build up an empirical likelihood ratio function similar to (13), $\tilde{\mathcal{R}}\{\beta(z)\}$ and we can state the following result.

Corollary 1. *Taking $\hat{\gamma}$ to be the minimizer of the cross-validation function (16), then under Conditions 1–2, we have $\tilde{\mathcal{R}}\{\beta(z)\} \rightsquigarrow \chi_d^2$ as $(N, T) \to (\infty, \infty)$.*

Here we define the confidence bands in the same way as in (15), i.e., the set of values $\beta(z)$ such that $\tilde{\mathcal{R}}\{\beta(z)\} \le c_\alpha$ where $\Pr(\chi_d^2 \le c_\alpha) = \alpha$. Note that using the empirical likelihood technique, it is possible to implement both Theorem 1 and Corollary 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.

3. Maximum empirical likelihood estimator

We define the maximizer of (13), $\hat{\beta}(z)$, as the maximum empirical likelihood estimator of $\beta(z)$, i.e., $\hat{\beta}(z) = \max_{\beta(z)} \mathcal{R}\{\beta(z)\}$. Using (11) and (13), and following the same lines as in [30], we can write

$$\hat{\beta}(z) = \left\{\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{X}_{it}^\top L(Z_{it}, z, \gamma)\right\}^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{X}_{it}\tilde{Y}_{it}L(Z_{it}, z, \gamma) + o_P(1/\sqrt{NT}). \tag{18}$$

Consequently, for comparison purposes, we derive the asymptotic distribution of MELE estimator, (18), in the following theorem.

Theorem 2. *Assume that Condition 1 holds, $\gamma \to 0$ and $(N, T) \to (\infty, \infty)$. Then*

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

where

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

$$\Gamma_1(z) = p(z)\{\Sigma_X(z) - \mu_X(z)\mu_X(z)^\top\} + O(\|\gamma\|), \quad b(\gamma) = \Gamma_1(z^*)\{\beta(z^*) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} + O(\|\gamma\|^2).$$

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

Corollary 2. *Assume that Condition 1 holds, v_{it} are iid over i and t, $\gamma \to 0$, and $(N, T) \to (\infty, \infty)$. Then*

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

Note that under unknown sequences of γ and using Lemmas 1–2, the proof of Theorem 2 will simplify considerably since we will be working with $\hat{\beta}(z) = \tilde{\beta}(z) + O_P\{1/(NT)\}$, where $\tilde{\beta}(z)$ is a frequency estimator in the same way as in $\hat{\beta}(z)$ when $\gamma_1 = \cdots = \gamma_q = 0$. Therefore, is straightforward to obtain that

$$\sqrt{NT}\{\hat{\beta}(z) - \beta(z)\} = \sqrt{NT}\{\tilde{\beta}(z) - \beta(z)\} + O_P(1/\sqrt{NT}). \tag{19}$$

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

Theorem 3. *Take $\hat{\gamma}$ to be the minimizer of the cross-validation function* (16)*, assume that Conditions 1–2 hold, and* $(N,T) \to (\infty,\infty)$*. Then*

$$\sqrt{NT}\,\{\tilde{\beta}(z) - \beta(z)\} \rightsquigarrow \mathcal{N}[0_{d\times1}, \Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)],$$

where $\Gamma_1(z) = p(z)\{\Sigma_X(z) - \mu_X(z)\mu_X(z)^\top\}$ *and*

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

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

Corollary 3. *Take $\hat{\gamma}$ to be the minimizer of the cross-validation function* (16)*, assume that Conditions 1 and 2 hold, v_{it} are iid over i and t, and* $(N,T) \to (\infty,\infty)$*. Then*

$$\sqrt{NT}\,t\{\tilde{\beta}(z) - \beta(z)\} \rightsquigarrow \mathcal{N}[0_{d\times1}, \sigma_v^2\Gamma_1^{-1}(z)].$$

Note that to invoke asymptotic normality, we need to estimate the variance-covariance matrix and sometimes this estimation is no feasible; see variance expressions in Theorems 2 and 3. To cope with this issue, we imposed a stronger condition on the random errors, i.e., v_{it} are iid over i and t; this allowed us to estimate the variance expression using Corollaries 2–3. Hence, to construct the confidence bands, by (A.30), it is easy to show that $\hat{\Gamma}_1(z) \to_P \Gamma_1(z)$, where

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

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

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

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 1951–85.

4. Illustration

We report estimates of strike activities from 17 OECD countries for the period 1951–85. 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 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 [36], we use the log transformation to stabilize the volatility of the strike series.

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 $\tilde{X}_{it} = (\tilde{U}_{it}, \tilde{I}_{it}, \tilde{P}_{it})^\top$ is a 3×1 vector. Now we apply the empirical likelihood approach (Corollary 1) and the asymptotic normality (Corollary 3) to estimate the confidence bands of the parameters of interest. Here, we use Corollary 1 instead of Theorem 1 for comparison purposes. The results are shown in Tables 1–3, 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.

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

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	0.2[illegible]
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	[illegible].01
10	−2.59	−2.12	−1.84	−1.3[illegible]	−1.09
11	−0.08	−0.14	0.01	0.1[illegible]	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	1.3[illegible]	2.34
16	−0.10	0.01	0.05	0.10	0.19
17	−0.47	−0.28	−0.2[illegible]	−0.21	−0.02

5. Conclusions

Extending the work of Li et al. [20] 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 were able to apply empirical likelihood methods to set up confidence bands for the functions of interest. As a by-product, we provided an alternative empirical maximum likelihood estimator of the categorical varying coefficients and derive its asymptotic theory. Finally, we applied successfully our techniques to an empirical study of estimates of strike

Table 2: Confidence bands for $\hat{\beta}_2(z)$.

	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 for $\hat{\beta}_3(z)$.

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	0.43	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

activities from 17 OECD countries for the period 1951–85.

Acknowledgments

We thank the Editor-in-Chief, Christian Genest, an Associate Editor and the referees, as well as our financial sponsors. 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. 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. Any errors are ours.

Appendix

From here on, we will be using the notation that has been defined in the previous Assumptions 1 and 2 and Theorems 1 and 2. Also, as in [22], $C(1)$ denotes some constants which may be different at each appearance.

Proof of Theorem 1.

Using Eq. (13), the proof of this theorem is carried out in three steps. First, we show the asymptotic normality of $\sum_{i=1}^{N} T_i\{\beta(z)\}/\sqrt{NT}$; second, we show the consistency of $\tilde{D}\{\beta(z)\}$; and finally, we use a Cramér–Wold device to conclude. In order to obtain the asymptotic distribution of $\sum_{i=1}^{N} T_i\{\beta(z)\}/\sqrt{NT}$ note that

$$\frac{1}{NT}\sum_{i=1}^{N} T_i\{\beta(z)\} = \frac{1}{NT}\sum_{i=1}^{N}\left[T_i\{\beta(z)\} - \mathrm{E}\left[T_i\{\beta(z)\}|\mathcal{X}\right]\right] + \frac{1}{NT}\sum_{i=1}^{N}\mathrm{E}\left[T_i\{\beta(z)\}|\mathcal{X}\right] \equiv U_{1NT} + U_{2NT},$$

where $\mathcal{X} = \{(X_{js}, Z_{js}) : j \in \{1,\ldots,N\}, s \in \{1,\ldots,T\}\}$. Also note that, as we already mentioned, $\gamma \to 0$ as $(N,T) \to (\infty,\infty)$; this allows us, along the same lines as [21], to simplify the kernel product function as in (3) and using the same argument we are able to write

$$T_{it}^{*} = \sum_{s=1}^{T}\mathbf{1}(Z_{is} = Z_{it}) + O\left(\|\gamma\|^{p}\right), \quad Y_{it}^{*} = Y_{it} - \sum_{s=1}^{T} Y_{is}\mathbf{1}(Z_{is} = Z_{it})/T_{it}^{*} + o(1),$$

$$X_{it}^* = X_{it} - \sum_{s=1}^{T} X_{is}\mathbf{1}\left(Z_{is} = Z_{it}\right)/T_{it}^* + o(1), \quad v_{it}^* = v_{it} - \sum_{s=1}^{T} v_{is}\mathbf{1}(Z_{is} = Z_{it})/T_{it}^* + o(1),$$

$$\varrho_{it}^* = \sum_{s=1}^{T} X_{is}^\top\{\beta(Z_{it}) - \beta(Z_{is})\}\mathbf{1}(Z_{is} = Z_{it})/T_{it}^* + o(1). \tag{A.1}$$

We first work on the bias term U_{2NT}; then, substituting $T_i\{\beta(z)\}$ by (8) into U_{2NT}, applying Assumption 1.(iv) and replacing $L(Z_{it}, z, \gamma)$ with (3) and using (A.1), we have

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

where $L_{it,z,\gamma} = L(Z_{it}, z, \gamma)$, $1_{itz} = \mathbf{1}(Z_{it} = z)$ and $1_{m,itz^*} = \mathbf{1}(Z_{it,m} \neq z_m)\prod_{n=1,n\neq m}^{q}\mathbf{1}(Z_{it,n} = z_n)$ is an indicator function which takes value 1 if Z_{it} and z differs only in their mth component and 0 otherwise. Note that in the last equality, by construction, $\{\beta(Z_{it}) - \beta(z)\}\,1_{itz} = 0_{d\times 1}$ and $\{\beta(Z_{it}) - \beta(Z_{is})\}\,\mathbf{1}(Z_{is} = Z_{it}) = 0_{d\times 1}$; therefore, all the terms containing ϱ_{it}^* vanish. We continue the analysis of (A.2); to do so, we follow [12] and use Lemma A2 of [23]. This lemma is a three-step 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 $\|\cdot\|$.

Step 2: Rewrite (A.2) as

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^*X_{it}^{*\top}\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T}X_{is}1_{itis}\right)\left(X_{it} - \frac{1}{T_{it}^*}\sum_{s=1}^{T}X_{is}1_{itis}\right)^\top\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}^\top\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
&\quad + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^*}\sum_{s_1=1}^{T}X_{is_1}1_{itis_1}\frac{1}{T_{it}^*}\sum_{s_2=1}^{T}X_{is_2}^\top 1_{itis_2}\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}\frac{1}{T_{it}^*}\sum_{s=1}^{T}X_{is}^\top 1_{itis}\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} \\
&\quad - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^*}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^\top\{\beta(Z_{it}) - \beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*},
\end{aligned} \tag{A.3}$$

where $1_{itis} = \mathbf{1}(Z_{is} = Z_{it})$. For the last two terms of (A.3), note that we can write

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

which can then be bounded above by

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\mathrm{E}\left\|K_{it}^{*}\right\|^{2}\mathrm{E}\left\|X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}\right\|^{2}\right\}^{1/2}$$

$$\leq\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left\|K_{it}^{*}\right\|^{2}\right\}^{1/2}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}\left\|X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}\right\|^{2}\right\}^{1/2}=o_{p}(\|\gamma\|),\quad\text{(A.4)}$$

where $K_{it}^{*}=\sum_{s=1}^{T}X_{is}1_{itis}/T_{it}^{*}-\mu(Z_{it})$. We now obtain that, for any given $z\in\mathcal{D}$ and $\gamma\in[0,1]^{q}$,

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{itis}X_{it}^{\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(Z_{it})\mu(Z_{it})^{\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}+o_{p}(\|\gamma\|).$$

Similarly, for the second term of (A.3), we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{T_{it}^{*}}\sum_{s_{1}=1}^{T}X_{is_{1}}1_{itis_{1}}\frac{1}{T_{it}^{*}}\sum_{s_{2}=1}^{T}X_{is_{2}}^{\top}1_{itis_{2}}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mu(Z_{it})\mu(Z_{it})^{\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}+o_{p}(\|\gamma\|).$$

In view of all the above, we obtain

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}^{*}X_{it}^{*\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}$$

$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{X_{it}-\mu(Z_{it})\right\}\left\{X_{it}-\mu(Z_{it})\right\}^{\top}\left\{\beta(Z_{it})-\beta(z)\right\}\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}+o_{p}(\|\gamma\|)$$

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

$$\mathrm{E}\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}^{\top}\beta(Z_{it})\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}-p(z^{*})\Sigma_{X}(z^{*})\beta(z^{*})\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}\right\|^{2}$$

$$=\frac{1}{(NT)^{2}}\sum_{h,\ell=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\mathrm{E}\left[\left\{X_{it,h}X_{it,\ell}\beta_{h}(Z_{it})\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}-p(z^{*})\Sigma_{X,h\ell}(z^{*})\beta_{h}(z^{*})\sum_{m=1}^{q}\gamma_{m}1_{m,itz^{*}}\right\}\right.$$

$$\left.\times\left\{X_{js,h}X_{js,\ell}\beta_{h}(Z_{js})\sum_{m=1}^{q}\gamma_{m}1_{m,jsz^{*}}-p(z^{*})\Sigma_{X,h\ell}(z^{*})\beta_{h}(z^{*})\sum_{m=1}^{q}\gamma_{m}1_{m,jsz^{*}}\right\}\right],$$

which can be bounded above by

$$O(\|\gamma\|^{2})\frac{1}{(NT)^{2}}\sum_{h,\ell=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}c_{\delta}\{\alpha_{ij}(|t-s|)\}^{\delta/(4+\delta)}$$

$$\leq O(\|\gamma\|^{2})\frac{1}{(NT)^{2}}\sum_{h,\ell=1}^{d}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\{\alpha_{ij}(|t-s|)\}^{\delta/(4+\delta)}=O\{\|\gamma\|^{2}/(NT)\},\quad\text{(A.5)}$$

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 $\mathbf{1}(Z_{it}=z)$ is uniformly bounded. Also, let $X_{it,h}$ be the hth element of X_{it} and $\Sigma_{X,h\ell}(z^*)$ denotes the (h,ℓ)th element of $\Sigma_X(z^*)$ for $h,\ell \in \{1,\ldots,d\}$. Therefore, we have proved that

$$\begin{aligned}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}X_{it}^{*\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} & \\ \to_P p(z^*)\{\Sigma_X(z^*)-\mu_X(z^*)\mu_X(z^*)^\top\}\{\beta(z^*)-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} & \\ = \Gamma_1(z^*)\{\beta(z^*)-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} = b(\gamma) & \quad \text{(A.6)}\end{aligned}$$

for any given $z \in \mathcal{D}$ and $\gamma \in [0,1]^q$. Therefore, (A.2) has the required expression.

Step 3: By Step 2, we can write

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}X_{it}^{*\top}\{\beta(Z_{it})-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*} = b(\gamma) + O_P(\|\gamma\|^2),$$

and for any $\gamma_1,\gamma_2 \in [0,1]^q$, we have $\|b(\gamma_1)-b(\gamma_2)\| \le O(1)\,\|\gamma_1-\gamma_2\|$, which implies the third condition of Lemma A2 of [23] holds. Therefore, we can conclude that

$$U_{2NT} = b(\gamma) + O_p(\|\gamma\|^2). \quad \text{(A.7)}$$

Now we obtain the limiting distribution of the quantity $\sqrt{NT}\,U_{1NT}$. By substituting (8) into U_{1NT} and replacing $L(Z_{it},z,\gamma)$ with (3), we obtain

$$U_{1NT} = \frac{1}{NT}\sum_{i=1}^{N}\left[T_i\{\beta(z)\} - \mathrm{E}\left[T_i\{\beta(z)\}|\mathcal{X}_i\right]\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}v_{it}^{*}\left(1_{itz}+\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}\right) + O_p(\|\gamma\|^2). \quad \text{(A.8)}$$

Therefore, we first focus on the analysis of $\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}v_{it}^{*}\mathbf{1}(Z_{it}=z)/(NT)$. We have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}v_{it}^{*}\mathbf{1}(Z_{it}=z) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}X_{is}1_{is,it}\right)\left(v_{it}-\frac{1}{T_{it}^{*}}\sum_{s=1}^{T}v_{is}1_{is,it}\right)\mathbf{1}(Z_{it}=z). \quad \text{(A.9)}$$

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

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}^{*}v_{it}^{*}\mathbf{1}(Z_{it}=z) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_{it}-\mu_X(z)\}\,v_{it}\mathbf{1}(Z_{it}=z) + o_P(1+\|\gamma\|^2). \quad \text{(A.10)}$$

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

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\{X_{it}-\mu_X(z)\}\,v_{it}\mathbf{1}(Z_{it}=z) = \sum_{t=1}^{T} V_{T,N}(t). \quad \text{(A.11)}$$

By Assumption 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 [13], Theorem 2.21 in [11] and Lemma A.1 in [7]. 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

$$s_T \to \infty, \quad s_T/\ell_T \to 0, \quad \ell_T/T \to 0, \quad k_T \equiv \{T/(\ell_T+s_T)\} = O(s_T).$$

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

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

Note that $\alpha(T) = o(1/T)$ 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 find

$$\mathrm{E}\|\bar{V}_1 + \cdots + \bar{V}_{k_T}\|^2 = O\{(k_T s_T)/T\} = o(1), \quad \mathrm{E}\|\hat{V}\|^2 = O\{(T - k_T\ell_T)/T\} = o(1).$$

Therefore, we just need to focus the analysis on $\tilde{V}_1 + \cdots + \tilde{V}_{k_T}$. Using the Feller–Lindeberg Central Limit Theorem, we first need to show that $\tilde{V}_1 + \cdots + \tilde{V}_{k_T}$ are asymptotically mutuallly independent. By Proposition 2.6 in [11] and the condition of α-mixing coefficients, we have

$$\left|\mathrm{E}(\exp\|\tilde{V}_1 + \cdots + \tilde{V}_{k_T}\|) - \prod_{n=1}^{k_T} \mathrm{E}(\exp\|\tilde{V}_n\|)\right| \le C(k_T - 1)\alpha(s_T) \to 0, \tag{A.12}$$

where C is a constant and α is the upper bounded of the α-mixing coefficient defined in Assumption 1.(ii). This upper bound is achievable in the same way as Assumption A.4 of [7]. Therefore we obtain that $\tilde{V}_1, \ldots, \tilde{V}_{k_T}$ are asymptotically independent. Furthermore, as in the proof of Theorem 2.10(ii) in [11], we have to show Feller's finite variance condition. We have

$$\begin{aligned}\mathrm{cov}(\tilde{V}_1) &= \mathrm{cov}\left\{\sum_{t=1}^{\ell_T} V_{N,T}(t)\right\} = \mathrm{cov}\left[\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{\ell_T}\{X_{it} - \mu_X(z)\}\, v_{it}\mathbf{1}(Z_{it} = z)\right] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{\ell_T}\mathrm{cov}\left[\{X_{it} - \mu_X(z)\}\, v_{it}\mathbf{1}(Z_{it} = z)\right] = \ell_T\Gamma_0(z)\left\{I_d + o(1)\right\}/T,\end{aligned} \tag{A.13}$$

which implies that

$$\sum_{n=1}^{k_T}\mathrm{cov}(\tilde{V}_n) = k_T\,\mathrm{cov}(\tilde{V}_1) = k_T\ell_T\Gamma_0(z)\left\{I_d + o(1)\right\}/T \to \Gamma_0(z). \tag{A.14}$$

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

$$\sum_{n=1}^{k_T}\mathrm{E}\{\|\tilde{V}_n\|^2\mathbf{1}(\|\tilde{V}_n\| \ge \varepsilon)\} \to^p 0, \tag{A.15}$$

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

$$\mathrm{E}\{\|\tilde{V}_n\|^2\mathbf{1}(\|\tilde{V}_n\| \ge \varepsilon)\} \le (\mathrm{E}\|\tilde{V}_n\|^3)^{2/3}\{\Pr(\|\tilde{V}_n\| \ge \varepsilon)\}^{1/3} \le C(\mathrm{E}\|\tilde{V}_n\|^3)^{2/3}(\mathrm{E}\|\tilde{V}_n\|^2)^{1/3}, \tag{A.16}$$

and by Lemma B.2 in [7],

$$\mathrm{E}\|\tilde{V}_n\|^3 \le (\ell_T/T)^{3/2}\left[\mathrm{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{X_{i1} - \mu_X(z)\}\, v_{i1}\mathbf{1}(Z_{i1} = z)\right\|^4\right]^{3/4} < \infty, \tag{A.17}$$

$$\mathrm{E}\|\tilde{V}_n\|^2 \le (\ell_T/T)\left[\mathrm{E}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{X_{i1} - \mu_X(z)\}\, v_{i1}\mathbf{1}(Z_{i1} = z)\right\|^4\right]^{1/2} < \infty. \tag{A.18}$$

Thus, $\mathrm{E}\|\tilde{V}_n\|^3 = O\{(\ell_T/T)^{3/2}\}$ and , $\mathrm{E}\|\tilde{V}_n\|^2 = O(\ell_T/T)$ which, using (A.16), implies

$$\mathrm{E}\{\|\tilde{V}_n\|^2\mathbf{1}(\|\tilde{V}_n\| \ge \varepsilon)\} \le O\{(\ell_T/T)^{4/3}\} = o(\ell_T/T). \tag{A.19}$$

Therefore,

$$\sum_{n=1}^{k_T} \mathrm{E}\{\|\tilde{V}_n\|^2 \mathbf{1}(\|\tilde{V}_n\| \geq \varepsilon)\} = o(k_T \ell_T / T) = o(1). \tag{A.20}$$

Consequently, the Lindeberg condition is satisfied; using (A.6), (A.12), (A.14) and (A.20), it is easy to see that if $\gamma_m \to 0$ we can conclude that, as $(N, T) \to (\infty, \infty)$,

$$\sqrt{NT}\, U_{1NT} \rightsquigarrow \mathcal{N}[0_{d\times 1}, \Gamma_0(z)]. \tag{A.21}$$

Now we prove the consistency of $\tilde{D}\{\beta(z)\}$. Similar to the proof of (A.11)–(A.14), it is straightforward to show that

$$\tilde{D}\{\beta(z)\} = \frac{1}{NT}\sum_{i=1}^{N} T_i\{\beta(z)\}\, T_i^\top\{\beta(z)\} = \Gamma_0(z)\{I_d + o_p(1)\}. \tag{A.22}$$

From (A.6), (A.21) and (A.22), and using the same arguments as in the proof of (2.14) in [24], we can prove that

$$\lambda = O_p(1/\sqrt{NT}), \tag{A.23}$$

where λ was defined in (12). Then applying Taylor expansion to (11) and invoking (A.6), (A.21) and (A.22), we get

$$\mathcal{R}\{\beta(z)\} = 2\sum_{i=1}^{N}[T_i^\top\{\beta(z)\}\,\lambda - [T_i^\top\{\beta(z)\}\,\lambda]^2/2] + o_p(1). \tag{A.24}$$

By (12), and applying Taylor expansion again, it follows that

$$0 = \sum_{i=1}^{N}\frac{T_i\{\beta(z)\}}{1+\lambda^\top T_i\{\beta(z)\}} = \sum_{i=1}^{N} T_i\{\beta(z)\} - \sum_{i=1}^{N} T_i\{\beta(z)\}\, T_i^\top\{\beta(z)\}\,\lambda + \sum_{i=1}^{N}\frac{T_i\{\beta(z)\}\,[T_i^\top\{\beta(z)\}\,\lambda]^2}{1+T_i^\top\{\beta(z)\}}.$$

Then, recalling (A.6), (A.21) and (A.22), we can prove that

$$\sum_{i=1}^{N}[T_i^\top\{\beta(z)\}\,\lambda]^2 = \sum_{i=1}^{N} T_i^\top\{\beta(z)\}\,\lambda + o_p(1), \tag{A.25}$$

and

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

Relying on (A.6), (A.21)–(A.26), we can conlude the proof of Theorem 1 by applying the Cramér–Wold device. □

Proof of Theorem 2. Note that, without loss of generality, we can write

$$\hat{\beta}(z) - \beta(z) = [\hat{\beta}(z) - \mathrm{E}\{\hat{\beta}(z)|\mathcal{X}\}] + [\mathrm{E}\{\hat{\beta}(z)|\mathcal{X}\} - \beta(z)] \equiv \mathbf{I}_{1NT} + \mathbf{I}_{2NT}. \tag{A.27}$$

To prove the desired result, under Assumption 1, we will show first that $\mathbf{I}_{2NT} = \Gamma^{-1}(z)b(\gamma)$ and second that $\sqrt{NT}\,\mathbf{I}_{1NT} \rightsquigarrow \mathcal{N}[0_{d\times 1}, \Gamma_1^{-1}(z)\Gamma_0(z)\Gamma_1^{-1}(z)]$, as $(N, T) \to (\infty, \infty)$ and $\gamma_s \to 0$. If we substitute (18) into (A.27), we obtain

$$\mathbf{I}_{2NT} = \mathrm{E}\{\hat{\beta}(z)|\mathcal{X}\} - \beta(z) = \left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^\top L(Z_{it}, z, \gamma)\right\}^{-1}\left[\frac{1}{NT}\sum_{it}\tilde{X}_{it}\{\tilde{X}_{it}^\top\beta(Z_{it}) + \varrho_{it} - \beta(z)\}L(Z_{it}, z, \gamma)\right]. \tag{A.28}$$

We begin with the inverse term in (A.28). Replacing $L(Z_{it}, z, \gamma)$ with (3), and using (A.5)–(A.6), we get

$$\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^\top L(Z_{it}, z, \gamma) = \frac{1}{NT}\sum_{it} X_{it}^* X_{it}^{*\top} 1_{itz} + O_p(\|\gamma\|) \to_P p(z)\{\Sigma_X(z) - \mu_X(z)\mu_X(z)^\top\} + O_p(\|\gamma\|). \tag{A.29}$$

Then, using (A.29) we find that

$$\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\rightarrow_{P} p(z)\{\Sigma_X(z)-\mu_X(z)\mu_X(z)^{\top}\}+O_p(\|\gamma\|)=\Gamma_1(z). \tag{A.30}$$

Continuing with the second term of (A.28), and using (A.2)–(A.7), we obtain

$$\frac{1}{NT}\sum_{it}\tilde{X}_{it}\{\tilde{X}_{it}^{\top}\beta(Z_{it})+\varrho_{it}-\beta(z)\}L(Z_{it},z,\gamma)\rightarrow_{P}\Gamma_1(z^*)\{\beta(z^*)-\beta(z)\}\sum_{m=1}^{q}\gamma_m 1_{m,itz^*}+O_P(\|\gamma\|^2)=b(\gamma). \tag{A.31}$$

In order to show the asymptotic behavior of $\mathbf{I}_{1N}$ note that by (18) we have

$$\mathbf{I}_{1NT}=\hat{\beta}(z)-\mathrm{E}\{\hat{\beta}(z)|\mathbb{X},\mathbb{Z}\}=\left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{X}_{it}^{\top}L(Z_{it},z,\gamma)\right\}^{-1}\left\{\frac{1}{NT}\sum_{it}\tilde{X}_{it}\tilde{v}_{it}L(Z_{it},z,\gamma)\right\}, \tag{A.32}$$

where the inverse term was already study; see (A.30). Therefore, we will study the asymptotic behavior of (A.32) by studying the behavior of the second term. Based on the results obtained in (A.8)–(A.20) in the proof of Theorem 1 and (A.30) the following result holds

$$\sqrt{NT}\,\mathbf{I}_{1NT}\rightsquigarrow\mathcal{N}[0_{d\times 1},\Gamma_1^{-1}(z)\Gamma_2(z)\Gamma_1^{-1}(z)].$$

Thus the proof of Theorem 2 is complete. □

Proof of Corollary 1. From Eq. (17) we know that $T_i\{\beta(z)\}=\tilde{T}_i\{\beta(z)\}$, where

$$\tilde{T}_i\{\beta(z)\}=\sum_{t=1}^{T}\tilde{X}_{it}\{\tilde{Y}_{it}-\tilde{X}_{it}^{\top}\beta(z)\}\mathbf{1}(Z_{it}=z)+O_P\{1/(NT)\}.$$

Then, the proof of Corollary 1 is similar to the proof of Theorem 1 by setting $\gamma_1=\cdots=\gamma_q=0$. □

Proof of Corollary 3. From Eq. (19) we now that $\sqrt{NT}\{\hat{\beta}(z)-\beta(z)\}$ can be rewrite as

$$\sqrt{NT}\{\hat{\beta}(z)-\beta(z)\}=\sqrt{NT}\{\tilde{\beta}(z)-\beta(z)\}+O_P(1/\sqrt{NT}),$$

where $\tilde{\beta}(z)$ is a frequency estimator in the same way as in $\hat{\beta}(z)$ when $\gamma_1=\cdots=\gamma_q=0$. Then, the proof of Corollary 3 is similar to the proof of Theorem 2 by setting $\gamma_1=\cdots=\gamma_q=0$. □

References

[1] J. Aitchison, C.G. Aitken, Multivariate binary discrimination by the kernel method, Biometrika 63 (1976) 413–420.
[2] M. Arellano, Practitioners' corner: Computing robust standard errors for within-groups estimators, Oxf. Bull. Econ. Stat. 49 (1987) 431–434.
[3] L.A. Arteaga-Molina, J.M. Rodríguez-Poo, Empirical likelihood based inference for fixed effects varying coefficient panel data models, J. Statist. Plann. Inference 196 (2018) 144–162.
[4] J. Bai, Panel data models with interactive fixed effects, Econometrica 77 (2009) 1229–1279.
[5] Z. Cai, Trending time-varying coefficient time series models with serially correlated errors, J. Econometrics 136 (2007) 163–188.
[6] Z. Cai, Q. Li, Nonparametric estimation of varying coefficient dynamic panel data models, Econometric Theory 24 (2008) 1321–1342.
[7] J. Chen, J. Gao, D. Li, A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models, Econometric Theory 28 (2012) 1144–1163.
[8] J. Chen, J. Gao, D. Li, Semiparametric trending panel data models with cross-sectional dependence, J. Econometrics 171 (2012) 71–85.
[9] J. Chen, J. Gao, D. Li, Estimation in partially linear single-index panel data models with fixed effects, J. Bus. Econom. Statist. 31 (2013) 315–330.
[10] T. DiCiccio, P. Hall, J. Romano, Empirical likelihood is bartlett-correctable, Ann. Statist. 19 (1991) 1053–1061.
[11] J. Fan, Q. Yao, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer Science & Business Media, 2003.
[12] G. Feng, J. Gao, L. Peng, X. Zhang, A varying-coefficient panel data model with fixed effects: Theory and an application to us commercial banks, J. Econometrics 196 (2017) 68–82.
[13] J. Gao, Nonlinear Time Series: Semiparametric and Nonparametric Methods, CRC Press, Boca Raton, FL, 2007.

[14] P.J. Hall, B. La Scala, Methodology and algorithms of empirical likelihood, Internat. Statist. Rev. 58 (1990) 109–127.

[15] P.J. Hall, A.B. Owen, Empirical likelihood confidence bands in density estimation, Journal of Computational and Graphical Statistics 2 (1993) 273–289.

[16] W.K. Härdle, Applied Nonparametric Regression, 19, Cambridge University Press, 1990.

[17] E.D. Kolaczyk, Empirical likelihood for generalized linear models, Statist. Sinica 4 (1994) 199–218.

[18] G. Li, H. Peng, T. Tong, Simultaneous confidence band for nonparametric fixed effects panel data models, Economics Lett. 119 (2013) 229–232.

[19] G. Li, I. Van Keilegom, Likelihood ratio confidence bands in non–parametric regression with censored data, Scand. J. Statist. 29 (2002) 547–562.

[20] Q. Li, D. Ouyang, J.S. Racine, Categorical semiparametric varying-coefficient models, J. Appl. Econometrics 28 (2013) 551–579.

[21] Q. Li, J.S. Racine, Nonparametric Econometrics: Theory and Practice, Princeton University Press, 2007.

[22] W.K. Newey, D. McFadden, Large sample estimation and hypothesis testing, Handbook of Econometrics 4 (1994) 2111–2245.

[23] W.K. Newey, J.L. Powell, Instrumental variable estimation of nonparametric models, Econometrica 71 (2003) 1565–1578.

[24] A.B. Owen, Empirical likelihood ratio confidence regions, Ann. Statist. 18 (1990) 90–120.

[25] A.B. Owen, Empirical likelihood for linear models, Ann. Statist. (1991) 1725–1747.

[26] A.B. Owen, Empirical likelihood ratio confidence intervals for a single functional, Biometrika 75 (1988) 237–249.

[27] A.B. Owen, Empirical Likelihood, CRC Press, Boca Raton, FL, 2001.

[28] C.F. Parmeter, J.S. Racine, Nonparametric Estimation and Inference for Panel Data Models, Department of Economics Working Papers 2018-02, McMaster University, Hamilton, ON, Canada, 2018.

[29] M.H. Pesaran, E. Tosetti, Large panels with common factors and spatial correlation, J. Econometrics 161 (2011) 182–202.

[30] J. Qin, J.F. Lawless, Empirical likelihood and general estimating equations, Ann. Statist. 22 (1994) 300–325.

[31] J.M. Rodríguez-Poo, A. Soberón, Direct semi-parametric estimation of fixed effects panel data varying coefficient models, Econometrics J. 17 (2014) 107–138.

[32] J.M. Rodríguez-Poo, A. Soberón, Nonparametric estimation of fixed effects panel data varying coefficient models, J. Multivariate Anal. 133 (2015) 95–122.

[33] J.M. Rodríguez-Poo, A. Soberón, Nonparametric and semiparametric panel data models: Recent developments, J. Economic Surveys 31 (2017) 923–960.

[34] L. Su, A. Ullah, Nonparametric and emiparametric panel econometric models: Estimation and testing, Handbook of Empirical Economics and Finance, Chapman and Hall/CRC, Boca Faton, FL, 2011, pp.455–497.

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

[36] B. Western, Vague theory and model uncertainty in macrosociology, Sociological Methodol. 26 (1996) 165–192.

[37] L. Xue, L. Zhu, Empirical likelihood for a varying coefficient model with longitudinal data, J. Amer. Statist. Assoc. 102 (2007) 642–654.

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