Accepted Manuscript

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

Luis A. Arteaga-Molina, Juan M. Rodriguez-Poo

Please cite this article as: Arteaga-Molina L.A., Rodriguez-Poo J.M., Empirical likelihood based inference for fixed effects varying coefficient panel data models. *J. Statist. Plann. Inference* (2017), https://doi.org/10.1016/j.jspi.2017.11.003

Empirical Likelihood Based Inference for Fixed Effects Varying Coefficient Panel Data Models

- A panel data model with varying coefficents and fixed effects.
- Empirical likelihood bands are constructed.
- Simulation results and an application are presented.

Empirical Likelihood Based Inference for Fixed Effects Varying Coefficient Panel Data Models

Luis A. Arteaga-Molina [*]

Departamento de Economía, Universidad de Cantabria

Juan M. Rodriguez-Poo

Departamento de Economía, Universidad de Cantabria and IEE, University of Genève

November 13, 2017

Abstract

In this paper local empirical likelihood-based inference for non-parametric varying coefficient panel data models with fixed effects is investigated. First, we show that the naive empirical likelihood ratio is asymptotically standard chi-squared when undersmoothing is employed. The ratio is self-scale invariant and the plug-in estimate of the limiting variance is not needed. Second, mean-corrected and residual-adjusted empirical likelihood ratios are proposed. The main interest of these techniques is that without undersmoothing, both also have standard chi-squared limit distributions. As a by product, we propose also two empirical maximum likelihood estimators of the varying coefficient models and their derivatives. We also obtain the asymptotic distribution of these estimators. Furthermore, a non parametric version of the Wilk's theorem is derived. To show the feasibility of the technique and to analyse its small sample properties, using empirical likelihood-based inference we implement a Monte Carlo simulation exercise and we also illustrated the proposed technique in an empirical analysis about the production efficiency of the European Union's companies.

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

JEL code: C14, C12, C23

1 Introduction

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

obtain uniform confidence bands such as in Li et al. (2013).

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

In this paper, and starting from a fixed effects varying coefficient model, we obtain maximum empirical likelihood estimators of both the varying parameters and their derivatives. This last result is very interesting for testing constancy of parameter variation. Furthermore, we develop empirical likelihood ratios and we derive a non-parametric version of the Wilks' theorem. In order to obtain an unbiased ratio, we propose two modifications of the empirical likelihood ratio: the mean corrected and the residual adjusted empirical likelihood ratios. Based on these results, we can build up confidence regions for the parameter of interest through a standard chi squared approximation. The rest of this paper is organized as follows. In Section 2 we propose to construct the confidence bands for the unknown functions and their derivatives by using what we call a naive empirical likelihood technique. This technique shows as main drawback sub-optimal rates of convergence. In Section 3, as a byproduct, we provide two alternative maximum empirical likelihood estimators of the fixed effect nonparametric varying parameters model and their derivatives. In Section 4, and using the estimators that were previously derived, we propose two alternative techniques that enables us to obtain optimal

nonparametric rates: Mean corrected and residual-adjusted empirical likelihood ratios. In Section 5 we provide a Monte Carlo experiment and in Section 6 we undertake an empirical study about the production efficiency of the European Union's companies. Finally Section 7 concludes. The proofs of the main results are collected in the Appendix.

2 Naive empirical likelihood

Consider the following varying coefficient panel data regression model

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

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

Assumption 2.1. Let $(Y_{it}, X_{it}, Z_{it})_{i=1,...,N; t=1,...,T}$ be a set of independent and identically distributed (i.i.d.) $\mathbb{R}^{d+q+1}$ random variables in the subscript i for each fixed t and strictly stationary over t for a fixed i.

Assumption 2.2. The random errors v_{it} are independent and identically distributed, with 0 mean and homoscedastic variance $\sigma_v^2 < \infty$. They are also independent of X_{it} and Z_{it} for all i and t. Furthermore, $E|v_{it}|^{2+\delta} < \infty$ for some $\delta > 0$.

Assumption 2.3. Let μ_i can be arbitrarily correlated with both X_{it} and Z_{it} with unknown correlation structure.

Assumptions 2.1, 2.2 and 2.3 are rather standard assumptions in the panel data literature. Assumption 2.1 is standard in panel data models; we could consider other settings as in Cai and Li (2008), however, since in this paper we study the asymptotic properties as N tends to infinity and T is fixed, it is enough to assume stationarity. These type of models where T is fixed and N tends to infinity have been proved useful in the analysis of efficiency, where usually there is a large number of individuals during a small period of time. Assumption 2.2 is also standard for the conventional within and first difference transformation (Wooldridge (2002) or Hsiao (2003) for the fully parametric case). Independence between the idiosyncratic error and the covariates X_{it} and / or Z_{it} can be assumed without loss of generality, however it can be relaxed assuming

some dependence in higher moments. If we allow some dependence, we could transform this estimator to take into consideration more complex structures of the random error contained in the variance-covariance matrix (Martins-Filho and Yao (2009)). Assumptions 2.1 and 2.2 in some situations, as in Cai and Li (2008), are relaxed by considering that (X_{it}, Z_{it}, v_{it}) are for fixed, i, strictly stationary processes; unfortunately, this set of assumptions is not sufficient to bound the asymptotic variance of the estimator and some further mixing conditions are required to achieve convergence. In this case, T must also tend to infinity. Other cases such as cross sectional dependence also requires both N and T tending to infinity. Finally, assumption 2.3 imposes the so called fixed effects; note that we are not willing to assume any constraint in the relationship between the individual heterogeneity μ and the vector of covariates (X, Z).

Rather than focusing in the consistent estimation of $m(z)$ and its vector of derivatives, we will obtain confidence bands for those objects based on the empirical likelihood principle. As already stated in the introductory section above, this approach presents clear advantages against the standard asymptotically aproximated confidence bands. To make the argument for constructing the confidence regions for $m(z)$ and its derivatives we can start by noting that, for a given z, from model (1) we have that

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

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

$$E\left[\ddot{X}_{it}\left(\ddot{Y}_{it} - X_{it}^\top m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T}X_{is}^\top m(Z_{is})\right)\Bigg| Z_{i1} = z, \ldots, Z_{iT} = z\right] = 0, \tag{3}$$

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

$$E\left[\Delta X_{it}\left(\Delta Y_{it} - \left(X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)})\right)\right)\Big| Z_{it} = z, Z_{i(t-1)} = z\right] = 0. \tag{4}$$

In both cases the least squares estimator of $m(z)$ is the solution to either (3) or (4). If we approximate the unknown function $X_{it}^\top m(Z_{it})$ around a value z that is in a close neighborhood of Z_{it} by a linear function $X_{it}^\top m(z) + X_{it}^\top \otimes (Z_{it} - z)^\top vec(D_m(z))$, then the ortogonality conditions (3) and (4) are approximated respectively by

$$E\left[\tilde{Z}_{it}^*\left(\ddot{Y}_{it} - \tilde{Z}_{it}^{*\top}\beta(z)\right)\Big| Z_{i1} = z, \ldots, Z_{iT} = z\right] = 0, \tag{5}$$

5

and

$$E\left[\tilde{Z}_{it}\left(\Delta Y_{it} - \tilde{Z}_{it}^\top \beta(z)\right)\middle| Z_{it} = z, Z_{i(t-1)} = z\right] = 0, \tag{6}$$

where $\tilde{Z}_{it}^\top = \left(\Delta X_{it}^\top, X_{it}^\top \otimes (Z_{it} - z)^\top - X_{i(t-1)}^\top \otimes (Z_{i(t-1)} - z)^\top\right)$ is $1 \times d(q+1)$ vector, $\beta(z) = (m(z), vec(D_m(z)))^\top$ is a $d(q+1) \times 1$ vector, and $\tilde{Z}_{it}^{*\top} = \left(\ddot{X}_{it}^\top, X_{it}^\top \otimes (Z_{it} - z)^\top - \frac{1}{T}\sum_{s=1}^T X_{is}^\top \otimes (Z_{is} - z)\right)$ is also a $d(q+1) \times 1$ vector. Also let, $D_m(z)$ be a $d \times q$ matrix of partial derivatives of the $d \times 1$ function $m(z)$ with respect to the elements of the $q \times 1$ vector z, i.e. $D_m(z) = \frac{\partial m(z)}{\partial z}$. Note that equations (5) and (6) are the first order conditions of the minimization problem $E\left[\left(\ddot{Y}_{it} - \tilde{Z}_{it}^{*\top}\beta(z)\right)^2\middle| z\right]$ and $E\left[\left(\Delta Y_{it} - \tilde{Z}_{it}^\top\beta(z)\right)^2\middle| z\right]$ for a given z. Because nonparametric conditional expectations given either $(Z_{i1}, \ldots, Z_{iT})$ in (5) or $\left(Z_{it}, Z_{i(t-1)}\right)$ in (6) are involved, a local smoothing method is needed to obtain the sample version of those equations. In order to define the empirical likelihood estimator we employ equation (5) or (6) as auxiliary random vectors; therefore, the auxiliary random vector for the within transformation is as follows

$$T_{wi}\left(\beta(z)\right) = \sum_{t=1}^T \tilde{Z}_{it}^*[\ddot{Y}_{it} - \tilde{Z}_{it}^{*\top}\beta(z)]K_H(Z_{i1} - z)\cdots K_H(Z_{iT} - z), \tag{7}$$

and for the first differences transformation

$$T_{fi}\left(\beta(z)\right) = \sum_{t=2}^T \tilde{Z}_{it}[\Delta Y_{it} - \tilde{Z}_{it}^\top\beta(z)]K_H(Z_{it} - z)K_H(Z_{i(t-1)} - z). \tag{8}$$

In equations (7) and (8) H is a bandwidth matrix of dimension $q \times q$, $K(\cdot)$ denotes a kernel function in $\mathbb{R}^q$ and

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

Note that the $T_{w1}\left(\beta(z)\right), \ldots, T_{wN}\left(\beta(z)\right)$ are independent and, due to assumption 2.2, $E\left(T_{wi}\right) = 0$; the same implications remain valid for T_{fi}. Therefore, a naive empirical likelihood ratio function for $m(z)$ and $D_m(z)$ can be defined as the solution to the maximization problem of a multinomial log-likelihood function, i.e.

$$\mathcal{R}_w(\beta(z)) = -2\max\left\{\sum_{i=1}^N \log(p_i)\middle| p_i \geq 0, \ \sum_{i=1}^N p_i = 1, \ \sum_{i=1}^N p_i T_{wi}(\beta(z)) = 0\right\}, \tag{9}$$

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

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

6

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

Using p_i's expression and after some calculations equation (9) leads to

$$\mathcal{R}_w(\beta(z)) = 2 \sum_{i=1}^{N} \log(1 + \lambda^\top T_{wi}(\beta(z))), \tag{10}$$

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

$$\sum_{i=1}^{N} \frac{T_{wi}(\beta(z))}{1 + \lambda^\top T_{wi}(\beta(z))} = 0. \tag{11}$$

Let us now denote $\tilde{D}_w(\beta(z)) = (NT|H|^{T/2})^{-1} \sum_{i=1}^{N} T_{wi}(\beta(z)) T_{wi}^\top(\beta(z))$. Using equations (10), (11) and a Taylor expansion, it can be shown that

$$\mathcal{R}_w(\beta(z)) = \left[\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) \right]^\top \left[\tilde{D}_w(\beta(z)) \right]^{-1} \left[\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) \right] + o_p(1). \tag{12}$$

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

Assumption 2.4. The Kernel functions $K(.)$ are compactly supported and bounded kernels such that $\int K(u)du = 1$, $\int uu^\top K(u)du = \mu_2(K_u)I$, and $\int K(u)^2 du = R(K_u)$ where $\mu_2(K_u) \neq 0$, and $R(K_u) \neq 0$ are scalars and I is a $q \times q$ identity matrix. Besides, we will assume that there exist eight-order marginal moment for $K(.)$, i.e.,

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

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

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

Assumption 2.5. Let $f_{Z_{it}}(.)$, $f_{Z_{it}, Z_{i(t-1)}}(.,.)$ and $f_{Z_{i1}, Z_{i2}, Z_{i3}}(.,.,.)$, for $t = 1, .., T$ be respectively the probability density functions of Z_{it}, $(Z_{it}, Z_{i(t-1)})$ and (Z_{i1}, Z_{i2}, Z_{i3}). All density functions are continuously differentiable in all their arguments and they are bounded from above and bellow in any point of their support.

Assumption 2.6. Let z be an interior point of $f_{Z_{it}}$. The third order derivatives of $m_1(.), ..., m_d(.)$ are bounded and uniformly continuous.

Assumption 2.7. The bandwidth matrix H is symmetric and strictly definite positive. Moreover, each entry of the matrix tends to zero as $N \to \infty$ in such a way that $N|H| \to \infty$.

Assumption 2.8. The function $E[\ddot{X}_{it}\ddot{X}_{it}^\top | Z_{i1} = z_1, ..., Z_{iT} = z_T]$ is positive definite for any interior point of $(z_1, z_2, ..z_T)$ in the support of $f_{Z_{i1}, ..., Z_{iT}}(z_1, z_2, ..., z_T)$.

Assumption 2.9. Let $||A|| = \sqrt{\text{tr}(A^\top A)}$, then $E[\left|\left|X_{it}X_{it}^\top\right|\right|^2 | Z_{i1} = z, .., Z_{iT} = z]$ is bounded and uniformly continuous in its support. Furthermore, let the following matrix functions $E[\ddot{X}_{it}X_{it}^\top | Z_{i1} = z, .., Z_{iT} = z]$, $E[X_{it}X_{it}^\top | Z_{it} = z, \ Z_{i(t-1)} = z]$ and $E[X_{is}X_{is}^\top | Z_{i1} = z, ..., Z_{iT} = z]$ be bounded and uniformly continuous in their support. Also, $E[\ddot{X}_{it}X_{is}^\top | Z_{i1} = z, ..., Z_{iT} = z]$ and $E[X_{it}X_{is}^\top | Z_{i1} = z, ..., Z_{iT} = z]$, for $t \neq s$ and $t = s$, are bounded and uniformly continuous in their support.

Assumption 2.10. The functions $E[|X_{it}v_{it}|^{2+\delta} | Z_{it} = z, \ Z_{i(t-1)=z}]$, $E[|X_{is}v_{it}|^{2+\delta} | Z_{it} = z, \ Z_{i(t-1)=z}]$ and, $E[|\ddot{X}_{it}v_{it}|^{2+\delta} | Z_{it} = z, \ Z_{i(t-1)=z}]$, for some $\delta > 0$, are bounded and uniformly continuous in any point of its support.

These assumptions are rather common in the literature of non-parametric regression analysis of panel data models. Similar conditions were used in Xue and Zhu (2007), Su et al. (2010), Rodriguez-Poo and Soberón (2014) and Rodriguez-Poo and Soberón (2015). They are basically smoothness and boundedness conditions for the within estimator. There are also assumptions about the kernel functions and about the behavior of the bandwidth matrix.

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

Theorem 2.1. *Assuming that conditions 2.1 - 2.10 hold and $H \to 0$ in such a way that $NT|H|^{T/2} \to \infty$ and $\sqrt{NT|H|^{T/2}}tr(H) \to 0$, then $\mathcal{R}_w(\beta(z)) \to_d \chi^2_{d(q+1)}$ as $N \to \infty$ and T is fixed, where $\to_d$ means the convergence in distribution and $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.*

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

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

we obtain

$$\mathcal{R}_f(\beta(z)) = \left[\frac{1}{\sqrt{NT|H|}} \sum_{i=1}^N T_{fi}(\beta(z))\right]^\top \left[\tilde{D}_f(\beta(z))\right]^{-1} \left[\frac{1}{\sqrt{NT|H|}} \sum_{i=1}^N T_{fi}(\beta(z))\right] + o_p(1), \qquad (13)$$

8

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

Assumption 2.11. Let $||A|| = \sqrt{\text{tr}\left(A^\top A\right)}$, then the function $E[\Delta X_{it} \Delta X_{it}^\top | Z_{it} = z, \ Z_{i(t-1)} = z]$ is a positive definite for any interior point of (z, z) in the support of $f_{Z_{it}, Z_{i(t-1)}}(z, z)$.

Assumption 2.12. Also the following matrix functions $E[\Delta X_{it} X_{it}^\top | Z_{it} = z, \ Z_{i(t-1)} = z]$, $E[X_{it} X_{it}^\top | Z_{it} = z, \ Z_{i(t-1)} = z]$, $E[X_{i(t-1)} X_{i(t-1)}^\top | Z_{it} = z, \ Z_{i(t-1)} = z]$ and $E[\Delta X_{it} X_{i(t-1)}^\top | Z_{it} = z, \ Z_{i(t-1)} = z]$ are bounded and uniformly continuous in their support.

Assumption 2.13. The functions $E[|\Delta X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z, \ Z_{i(t-1)=z}]$, $E[|X_{it} \Delta v_{it}|^{2+\delta} | Z_{it} = z, \ Z_{i(t-1)=z}]$ and $E[|X_{i(t-1)} \Delta v_{it}|^{2+\delta} | Z_{it} = z, \ Z_{i(t-1)=z}]$ for some $\delta > 0$, are bounded and uniformly continuous in any point of its support.

These group of conditions substitute assumptions 2.8 - 2.10 when working with the first differences technique. Then, we are able to show the following result.

Theorem 2.2. *Assuming that conditions 2.1 - 2.7 and 2.11 - 2.13 hold and $H \to 0$ in such a way that $NT|H| \to \infty$ and $\sqrt{NT|H|} tr(H) \to 0$, then $\mathcal{R}_f(\beta(z)) \to_d \chi^2_{d(q+1)}$ as $N \to \infty$ and T is fixed, where $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.*

Using theorems 2.1 and 2.2 we can approximate α-level confidence regions for $\beta(z)$ as the set of values $\beta(z)$ such that $\mathcal{R}_f(\beta(z)) \le c_\alpha$ and $\mathcal{R}_w(\beta(z)) \le c_\alpha$, where c_α is defined such that $\mathbf{Pr}\left(\chi^2_{d(q+1)} \le c_\alpha\right) = \alpha$.

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

3 Maximum empirical likelihood estimators

We can define the maximum empirical likelihood (MELE) estimator of $\beta(z)$, $\hat{\beta}_w(z)$ as the minimizer of $\mathcal{R}_w(\beta(z))$. From equations (10) and (12) and following the same lines as Qin and Lawless (1994), $\hat{\beta}_w(z)$ is obtained from the solution of the estimating equation $\left(NT|H|^{T/2}\right)^{-1} \sum_{i=1}^N T_{wi}(\beta(z)) = 0$ and, as it will be shown in the proof of Theorem 3.1, the remainder term is of smaller order tending to zero as $NT|H|^{T/2}$ tends to infinity. Consequently, the MELE is asymptotically equivalent to the fixed effect estimator using the

within transformation. Therefore, if we assume that $\frac{1}{NT|H|^{T/2}} \sum_{it} \prod_{l=1}^{T} K_H(Z_{il} - z)\tilde{Z}_{it}^* \tilde{Z}_{it}^{*\top}$ is invertible, then the MELE is as follows

$$
\begin{aligned}
\hat{\beta}_w(z) &= \left(\frac{1}{NT|H|^{T/2}} \sum_{it}\prod_{l=1}^{T} K_H(Z_{il} - z)\tilde{Z}_{it}^* \tilde{Z}_{it}^{*\top} \right)^{-1} \frac{1}{NT|H|^{T/2}} \sum_{it}\prod_{l=1}^{T} K_H(Z_{il} - z)\tilde{Z}_{it}^* \ddot{Y}_{it} \\
&\quad + o_p \left(\frac{1}{\sqrt{NT|H|^{T/2}}} \right).
\end{aligned} \tag{14}
$$

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

Theorem 3.1. *Assuming that conditions 2.1 - 2.10 hold and $H \to 0$ in such a way that $NT|H|^{T/2} \to \infty$, then*

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

where

$$
\begin{aligned}
B_w(z) = diag &\left\{ I_d, \left[\left(1 - \frac{1}{T} \right) \mathcal{B}_{X_t X_t}(z,...,z) \otimes \mu_2(K_{u_\tau})H \right]^{-1} \right\} \\
&\times \left(\begin{array}{c} \frac{1}{2}\mu_2(K_{u_\tau}) diag_d \left\{ tr\{\mathcal{H}_{m_r}(z)H\} \right\} i_d \\ \frac{1}{2}\mu_2(K_{u_\tau})^2 B_{w1}(z) + \frac{1}{3!}B_{w2}(z) \end{array} \right),
\end{aligned}
$$

and

$$
\begin{aligned}
\Sigma_w(z) = \sigma_v^2 \, diag &\left\{ I_d, \left[\left(1 - \frac{1}{T} \right) \mathcal{B}_{X_t X_t}(z,...,z) \otimes \mu_2(K_{u_\tau})H \right]^{-1} \right\} \left(\begin{array}{cc} \Sigma_{w1}(z) & 0 \\ 0 & \Sigma_{w2}(z) \end{array} \right) \\
&\times diag \left\{ I_d, \left[\left(1 - \frac{1}{T} \right) \mathcal{B}_{X_t X_t}(z,...,z) \otimes \mu_2(K_{u_\tau})H \right]^{-1} \right\},
\end{aligned}
$$

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

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

Here, $\mathcal{DB}_{\ddot{X}\ddot{X}}(z,...,z)$ and $\mathcal{DB}_{X_t X_t}(z,...,z)$ are $d \times dq$ gradient matrix of the form

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

and $b_{dd'}^{X_t X_t}(z_1,...,z_T) = E\left[X_{dit}X_{d'it}|Z_{i1} = z_1,...,Z_{iT} = z_T\right]f_{Z_{i1},...,Z_{iT}}(z_1,...,z_T)$, $diag_d\{tr\{\mathcal{H}_{m_r}(z)H\}\}$ stands for a diagonal matrix of elements $tr\{\mathcal{H}_{m_r}(z)H\}$, for $r = 1,...,d$, where $\mathcal{H}_{m_r}$ is the Hessian matrix of the rth component of $m(.)$ and $D_m^3(z, Z_{it} - z)$ has as general expression, for $k = 3$,

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

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

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

$$
\begin{aligned}
\hat{\beta}_f(z) &= \left(\frac{1}{NT|H|}\sum_{it} K_H(Z_{it} - z)K_H(Z_{i(t-1)} - z)\tilde{Z}_{it}\tilde{Z}_{it}^\top\right)^{-1}\frac{1}{NT|H|}\sum_{it} K_H(Z_{it} - z)K_H(Z_{i(t-1)} - z)\tilde{Z}_{it}\Delta Y_{it} \\
&+ o_p\left(\frac{1}{\sqrt{NT|H|}}\right),
\end{aligned}
\tag{15}
$$

where the asymptotic normality of the estimator is as follows

Theorem 3.2. *Assuming that conditions 2.1 - 2.7 and 2.11 - 2.13 hold and $H \to 0$ in such a way that $NT|H| \to \infty$, then*

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

where,

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

and

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

11

also, let

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

Here, $\mathcal{DB}_{\Delta X \Delta X}$, $\mathcal{DB}_{XX}(z,z)$ and $\mathcal{DB}_{X_{-1}X_{-1}}(z,z)$ are $d \times dq$ gradient matrices defined as in theorem 3.1.

The results shown in Theorems 3.1 and 3.2 somehow correspond, under a different setting, to Theorem 3.1 in Rodriguez-Poo and Soberón (2015) and Theorems 3.1 in Rodriguez-Poo and Soberón (2014) respectively. However, we point out that the results obtained for the vector of derivatives are fully new in this fixed effects panel data setting. An interesting issue that needs to be considered here is the relative asymptotic efficiency of these estimators. Note first, that as the reader surely realizes none of these estimators achieve the optimal rate of convergence in terms of the Mean Integrated Square Error (MISE). Indeed, for this type of problems the optimal rate is $1/NT|H|^{1/2}$ (see Fan (1993) for details). For the estimator based in the within transformation the rate of convergence in terms of the MISE (see Theorem 3.1) is $1/NT|H|^{T/2}$, whereas for the estimator based in the first differences transformation (see Theorem 3.2) it is $1/NT|H|$. Therefore, the rate of convergence of both Empirical Maximum Likelihood Estimators is suboptimal. However, note that the relative asymptotic efficiency of $\widehat{\beta}_f(z)$ with respect to $\widehat{\beta}_w(z)$ with the same bandwidths is of order $O\left(|H|^{\frac{T}{2}-1}\right)$. If $T > 2$ then $\widehat{\beta}_f(z)$ will exhibit a faster rate of convergence than $\widehat{\beta}_w(z)$. Indeed as far as T gets larger this difference in rates increases. This is due to the so-called curse of dimensionality that is more serious in the case of the estimator based in the within transformation. In fact, in the case of $\widehat{\beta}_f(z)$ we use a kernel function of dimension $2 \times q$ whereas for the other estimator the dimension is $T \times q$. Finally, as an example, consider the estimation of $m(\cdot)$ using both estimators. Using Theorems 3.2 and 3.1 and some standard calculations note that the bandwidth that minimizes the MISE for the estimator based in the within transformation is of

order $(NT)^{-\frac{1}{4+qT}}$ whereas for the estimator based in the first differences transformation converges to zero at the rate $(NT)^{-\frac{1}{4+2q}}$. Substituting these optimal bandwidths in the asymptotic MISE expressions we obtain the following convergence rates: $(NT)^{-\frac{4}{4+qT}}$ for the within estimator, and $(NT)^{-\frac{2}{2+q}}$ for the first differences estimator.

4   Bias corrected empirical likelihood

In fact, note that in order to show the convergence of both theorems, theorem 2.1 and theorem 2.2, we have included one extra condition on the asymptotic behavior of the sequence of bandwidth matrices, i.e. $\sqrt{NT|H|^{T/2}}\mathrm{tr}(H) \to 0$ for the within estimator and $\sqrt{NT|H|}\mathrm{tr}(H) \to 0$ for the first differences transformation. These additional conditions ensure that the smoothness bias becomes negligible as the sample size tends to infinity. Unfortunately, these conditions on H exclude the bandwidth matrix that is optimal, therefore this will end up in suboptimal rates of convergence for both $\mathcal{R}_w(\beta(z))$ and $\mathcal{R}_f(\beta(z))$. In order to avoid this problem we propose two modifications of the Empirical Likelihood ratio that remove the bias term: the Mean-corrected Empirical Likelihood (MCEL) ratio and the Residual-Adjusted Empirical Likelihood (RAEL) ratio. These bias corrections have already been proposed in Xue and Zhu (2007) and what we will do here is to adapt them to our panel data with fixed effect setting.

4.1   Mean-corrected empirical likelihood ratio

As we have already pointed out, if H tends to zero at the optimal rate then $\mathcal{R}_w(\beta(z))$ will not converge in distribution to a χ^2 random variable. The main reason is that the smoothness bias will not vanish as $NT|H|^{T/2}$ tends to infinity. However, from the proof of Theorem 2.1 we know that, under the assumptions established in theorem 2.1, $\sqrt{NT|H|^{T/2}}\left(\frac{1}{NT|H|^{T/2}}\sum_i T_{wi}\left(\beta(z)\right) - b_w(z)\right) \to_d \mathcal{N}\left(0, \upsilon_w(z)\right)$, as $NT|H|^{T/2}$ tends to infinity. Here

$$b_w(z) = \begin{pmatrix} \frac{1}{2}b_{w1}(z) \\ \frac{1}{2}b_{w2}(z) + \frac{1}{3!}b_{w3}(z) \end{pmatrix}, \tag{16}$$

and

$$\upsilon_w(z) = \sigma_v^2 d \begin{pmatrix} R(K)^T \mathcal{B}_{\ddot{X}\ddot{X}}(z,...,z) & 0 \\ 0 & \left(1 - \frac{1}{T}\right)\mu_2(K_{u_\tau}^2)\prod_{l \neq \tau}^T R(K_{u_l})\mathcal{B}_{X_t X_t}(z,...,z) \otimes H \end{pmatrix}, \tag{17}$$

13

where

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

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

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

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

$$
\hat{b}_w(z)=\frac{1}{NT|H|^{T/2}}\sum_{it}\tilde{Z}_{it}^*\left(X_{it}^\top\hat{m}_w(Z_{it})-\frac{1}{T}\sum_{s=1}^T X_{is}^\top\hat{m}_w(Z_{is})-\tilde{Z}_{it}^{*\top}\hat{\beta}_w(z)\right)\prod_{l=1}^T K_H(Z_{il}-z), \tag{18}
$$

where $\hat{\beta}^w(z)$ is the MELE defined in (14), $\hat{m}_w(z)=e^\top\hat{\beta}^w(z)$, and $e=\begin{bmatrix}I_d & \vdots & \mathbf{0}\end{bmatrix}$, I_d is a d-dimensional unit matrix and $\mathbf{0}$ is a $dq\times d$ matrix. Taking into account (18), let us denote

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

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

$$
\tilde{\mathcal{R}}_w(\beta(z))=\mathcal{R}_w(\beta(z))-\tilde{\xi}_w(\beta(z)). \tag{19}
$$

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

$$
\tilde{\mathcal{R}}_f(\beta(z))=\mathcal{R}_f(\beta(z))-\tilde{\xi}_f(\beta(z)), \tag{20}
$$

where $\tilde{\xi}_f(\beta(z))=\sqrt{NT|H|}\,\hat{b}_f(z)^\top\left[\tilde{D}_f(\beta(z))\right]^{-1}\left[\frac{2}{\sqrt{NT|H|}}\sum_{i=1}^N T_{fi}(\beta(z))-\sqrt{NT|H|}\,\hat{b}_f(z)\right]$. Also, $\hat{b}_f(z)$ is a consistent estimator of $b_f(z)$. In this case, it is easy to show (see (A.19) for details) that

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

then the estimator of $b_f(z)$ is

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

where $\hat{\beta}^f(z)$ is the MELE defined in (15), and $\hat{m}_f(z) = e^\top \hat{\beta}^w(z)$. Note that from the proof of theorem 2.2,

$$b_f(z) = \begin{pmatrix} \frac{1}{2}b_{f1}(z) \\ \frac{1}{2}b_{f2}(z) + \frac{1}{3!}b_{f3}(z) \end{pmatrix}, \tag{21}$$

where

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

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

Theorem 4.1. *Assuming that conditions 2.1 - 2.13 hold, and $\beta(z)$ is the true vector of parameters, then $\tilde{\mathcal{R}}_w(\beta(z)) \to_d \chi^2_{d(q+1)}$ and $\tilde{\mathcal{R}}_f(\beta(z)) \to_d \chi^2_{d(q+1)}$ as $N \to \infty$ and T is fixed, where $\to_d$ means the convergence in distribution and $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.*

Note that to state this result we do not impose any extra condition. Here, we need conditions 2.1 - 2.10 to ensure that $\tilde{\mathcal{R}}_w(\beta(z)) \to_d \chi^2_{d(q+1)}$ and conditions 2.1 - 2.7 and 2.11 - 2.13 to ensure that $\tilde{\mathcal{R}}_f(\beta(z)) \to_d \chi^2_{d(q+1)}$ as $N \to \infty$. Basically these conditions are the same conditions of Theorems 2.1 and 2.2; however we do not need that $\sqrt{NT|H|^{T/2}}\mathrm{tr}(H) \to 0$ for the within transformation and $\sqrt{NT|H|}\mathrm{tr}(H) \to 0$ for the first differences transformation.

4.2 Residual-adjusted empirical likelihood ratio

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

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

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

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

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

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

15

and, for the first differences transformation, as

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

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

Theorem 4.2. *Assuming that conditions 2.1 - 2.13 hold, and $\beta(z)$ is the true parameter value, then $\hat{\mathcal{R}}_w(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ and $\hat{\mathcal{R}}_f(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ as $N \rightarrow \infty$ and T is fixed, where $\rightarrow_d$ means the convergence in distribution and $\chi^2_{d(q+1)}$ is the standard chi-squared distribution with $d(q+1)$ degrees of freedom.*

Note that to state this result we do not impose any extra condition. Here, we need conditions 2.1 - 2.10 to ensure that $\hat{\mathcal{R}}_w(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ and conditions 2.1 - 2.7 and 2.11 - 2.13 to ensure that $\hat{\mathcal{R}}_f(\beta(z)) \rightarrow_d \chi^2_{d(q+1)}$ as $N \rightarrow \infty$ as $N \rightarrow \infty$; however we do not need that $\sqrt{NT|H|^{T/2}} \text{tr}(H) \rightarrow 0$ for the within transformation and $\sqrt{NT|H|} \text{tr}(H) \rightarrow 0$ for the first differences transformation. Therefore, it is possible now to consider an optimal bandwidth matrix and hence the rate of convergence of the estimators will be also optimal.

As in other nonparametric estimation problems, bandwidth selection is important. Since the previous corrections enable us to use the optimal bandwidth then we can rely on standard data driven bandwidth selection techniques to select a bandwidth matrix. Among them, we propose to use a plug-in rule based on Sheather and Jones (1991). This proposal will be investigated in numerical studies in Section 5 and it will be also applied for illustrating the proposed Empirical Likelihood method with an empirical application in Section 6. Finally, there exists other data driven bandwidth selection criteria, such as cross-validation or empirical MSE criteria, that can be used alternatively to the plug-in method. They are detailed in a supplement. Their main drawback is that their are computationally more demanding.

5 Monte Carlo results

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

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

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

$$a.\_ \quad (d = 1, q = 1) \quad : \quad Y_{it} = \mu_{1i} + X_{1it}^{\top} m_1(Z_{1it}) + v_{it},$$

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

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

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

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

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

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

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

As the reader may notice, from table 1, between the MCEL, the RAEL and the NA, the length of the confidence interval is smaller in the RAEL; also note that, the confidence interval length of the MCEL is smaller than the NA. Also, it is interesting that, as table 1 shows, the confidence intervals using NA are wider than ones using empirical likelihood. Therefore we can say that when N goes to infinity the length the confidence bands of the NA are wider that the confidence bands of the MCEL and the RAEL. Thus, we can conclude by saying that the RAEL and MCEL confidence bands behave better than the NA confidence

Table 1: Pointwise Confidence interval for $\beta(z)$ at $z = 0$ based on the MCEL, RAEL and NA, when the nominal level is 95%

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

bands. Between RAEL and MCEL confidence bands, simulations results show that the RAEL onfidence bands behave better than the MCEL. Also, by comparing the within method with the first difference method we can conclude that for the NA and the RAEL confidence bands the First Difference method reduces the length of the confidence interval; however the MCEL confidence interval increases it length in comparison to the Within method (table 1)

6 An empirical aplication

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

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

$$y_{it} = w_{it}\beta_1(z_{it}) + l_{it}\beta_2(z_{it}) + k_{it}\beta_3(z_{it}) + \mu_i + v_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{22}$$

where $y_{it} = ln(Y_{it})$, $w_{it} = ln(W_{it})$, $l_{it} = ln(L_{it})$ and $k_{it} = ln(K_{it})$. Also, Y represents the sales of the company, W the liquid capital, L the labour input, K the fixed capital and Z the firms $R\&D$ expenses. In addition μ_i stands for the individual heterogeneity and $v_{it} = \nu_{it} - u_{it}$ is a composed error term, where ν_{it} is the idiosyncratic error and u_{it} represents the inefficiency that has expected value equal to $E[v_{it}] = -E[u_{it}]$. Note that in the specification (22) the $R\&D$ variable has a neutral effect on the production function by shifting the level of the production frontier but also affects the labour and capital marginal productivities.

Table 2: Statistics of inputs and outputs.

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

The sample used in this empirical analysis includes 1220 observations divided in 160 companies and 7 time periods, form 2008 to 2014, from the Analyse Major Database form European Sources (AMADEUS). The data contains information about the accounting and financial statements of European firms. Note that we are working with expenses, thus all the variables have been deflected using the implicit index of the GDP. The information related to prices used to deflate the variables was obtained form the Spanish Statistical Office (various years). In Table 2 we present summary statistics of the observations, as it can be seen, the standard deviations show that there exist a high degree of heterogeneity.

In figure 1 we present our results by plotting the estimated curves against the $R\&D$ expenses; here the continuous lines denote the non-parametric estimated curve and the doted lines represent the 95% pointwise confidence interval obtained using the MCEL (long-dashed curve) and RAEL (short-dashed curve). The bandwidths, as in Section 5, have been computed by a plug-in technique proposed in Sheather and Jones (1991) and already explained in Section 4. Also, note that figure 1 shows the results for the marginal productivity of liquid capital (W), fixed capital (K) and labour (L), and the returns to scale defined as $\beta_1(z) + \beta_2(z) + \beta_3(z)$.

Figure 1: *Averages of 95 % Confidence Intervals for $\hat{\beta}_j(z)$ for $j = 1, \ldots, 3$ (Within method), based on MCEL (long-dashed curve) and RAEL (short-dashed curve).*
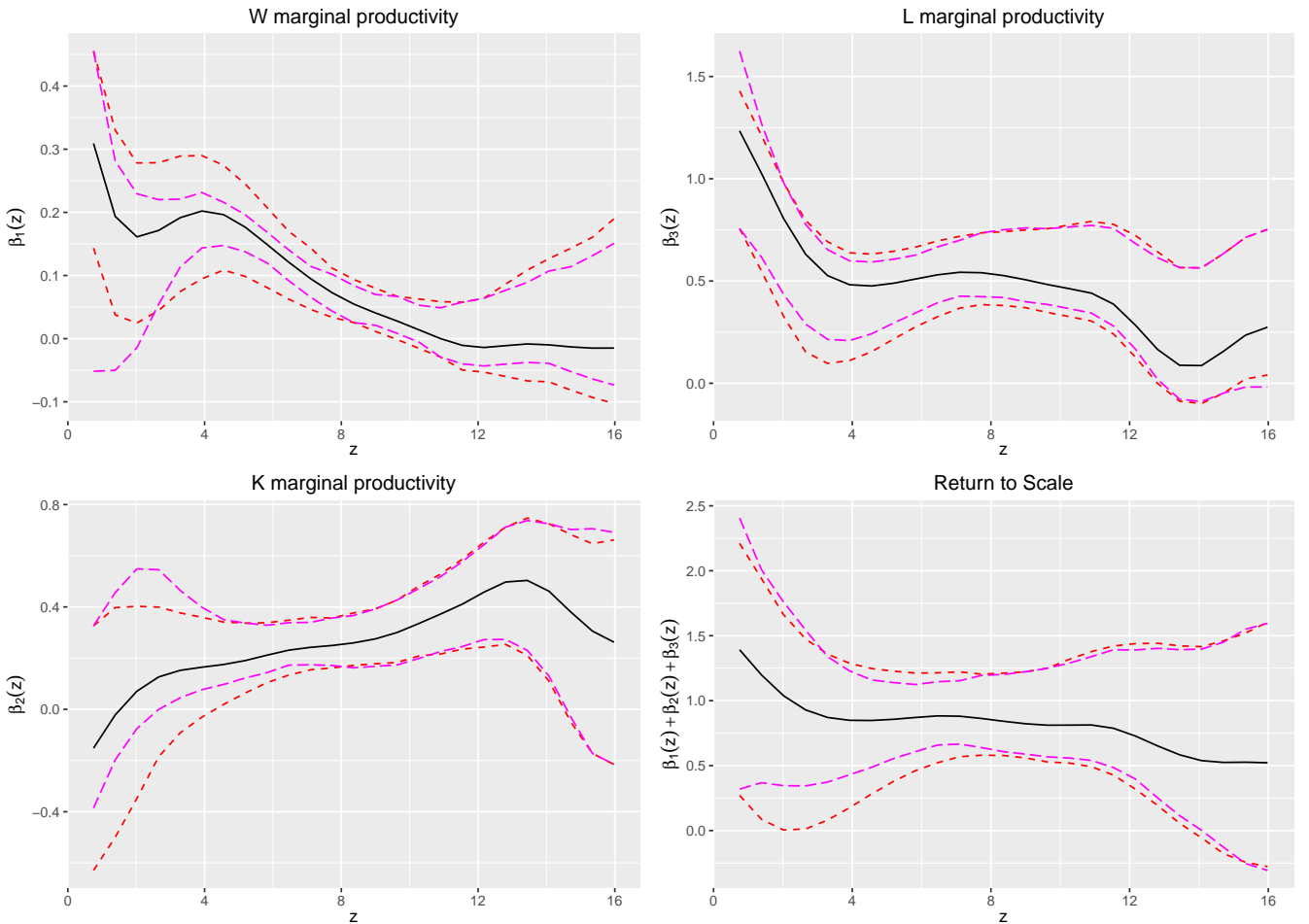
Focusing in the marginal productivity of liquid capital (W marginal productivity) we have realized that it tends to be decreasing; however when it reaches a certain level of $R\&D$ expenses it tends to be steady and close to zero. Basically, this means that companies with small $R\&D$ expenses have a decreasing marginal productivity of liquid capital. Analysing the graph, we can see that as the level of $R\&D$ expenses increases,

first the companies see an increase in the marginal productivity of liquid capital; then they experiment a drop of the marginal productivity of liquid capital until it become stable near to zero. On its part, the marginal productivity of fixed capital (K marginal productivity) is not a linear function with the level of $R\&D$ expenses. Clearly, there exist an upward general trend, with a bell shape form for companies with large $R\&D$ expenses. This bell shape of the marginal productivity of fixed capital curve suggests that, while modest $R\&D$ expenses can improve the fixed capital productivity, higher $R\&D$ expenses leads to lower fixed capital productivity.

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

7 Conclusions

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

8 Appendix: Proofs

8.1 Proof of Theorem 2.1

Note that, $\mathcal{R}_w(\beta(z)) = \left[\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) \right]^{\top} \left[\tilde{D}_w(\beta(z)) \right]^{-1} \left[\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) \right] + o_p(1)$, (see equation (12)) as N tends to infinity. The proof of this result is done in three steps: first, we show the

asymptotic normality of the vector $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z))$, second, we show the consistency of $\tilde{D}_w(\beta(z))$ and finally we use a Cramer-Wold device to close the proof.

In order to obtain the asymptotic distribution of $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z))$ note that

$$
\begin{aligned}
\frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} T_{wi}(\beta(z)) &= \frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} (T_{wi}(\beta(z)) - E[T_{wi}(\beta(z))|\mathbb{X}, \mathbb{Z}]) \\
&+ \frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} E[T_{wi}(\beta(z))|\mathbb{X}, \mathbb{Z}] \equiv U_{1N} + U_{2N},
\end{aligned}
\tag{A.1}
$$

where $\mathbb{X} = (X_{11}, ..., X_{NT})$ and $\mathbb{Z} = (Z_{11}, ..., Z_{NT})$ are the sample covariate values. We first work on the bias term U_{2N}. Then, substituting $T_{wi}(\beta(z))$ by (7) into $E[T_{wi}(\beta(z))|\mathbb{X}, \mathbb{Z}]$, applying Assumption 2.2 and taking Taylor expansion around $X_{it}^{\top} m(Z_{it}) - T^{-1} \sum_s X_{is}^{\top} m(Z_{is})$ we obtain

$$
U_{2N} \equiv \frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} E[T_{wi}(\beta(z))|\mathbb{X}, \mathbb{Z}] = \begin{pmatrix} A_{1.1N} + A_{1.2N} \\ A_{1.3N} + A_{1.4N} + A_{1.5N} \end{pmatrix}.
\tag{A.2}
$$

Here

$$
\begin{aligned}
A_{1.1N} &= \frac{1}{2NT|H|^{T/2}} \sum_{it} \ddot{X}_{it} Q_m(z) K_H(Z_i - z), \\
A_{1.2N} &= \frac{1}{NT|H|^{T/2}} \sum_{it} \ddot{X}_{it} R_1(z) K_H(Z_i - z), \\
A_{1.3N} &= \frac{1}{2NT|H|^{T/2}} \sum_{it} \left(X_{it} \otimes (Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z) \right) Q_m(z) K_H(Z_i - z), \\
A_{1.4N} &= \frac{1}{3!NT|H|^{T/2}} \sum_{it} \left(X_{it} \otimes (Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z) \right) C_m(z) K_H(Z_i - z), \\
A_{1.5N} &= \frac{1}{2NT|H|^{T/2}} \sum_{it} \left(X_{it} \otimes (Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z) \right) R_2(z) K_H(Z_i - z),
\end{aligned}
$$

where $K_H(Z_i - z) = \prod_{l=1}^{T} K_H(Z_{il} - z)$ and

$$
\begin{aligned}
Q_m(z) &= X_{it}^{\top} \otimes (Z_{it} - z)^{\top} \mathcal{H}_m(z)(Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is}^{\top} \otimes (Z_{is} - z)^{\top} \mathcal{H}_m(z)(Z_{is} - z), \\
C_m(z) &= X_{it}^{\top} \otimes D_m^3(z, Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is}^{\top} \otimes D_m^3(z, Z_{is} - z), \\
R_1(z) &= X_{it}^{\top} \otimes (Z_{it} - z)^{\top} \mathfrak{R}(Z_{it}; z)(Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is}^{\top} \otimes (Z_{is} - z)^{\top} \mathfrak{R}(Z_{is}; z)(Z_{is} - z), \\
R_2(z) &= X_{it}^{\top} \otimes \mathfrak{R}^*(Z_{it}; z) - \frac{1}{T} \sum_{s=1}^{T} X_{is}^{\top} \otimes \mathfrak{R}^*(Z_{is}; z).
\end{aligned}
$$

The remainder terms in the Taylor expansion are defined as

$$
\begin{aligned}
\mathfrak{R}(Z_{it}; z) &= \int_0^1 \left[\mathcal{H}_m(z + \omega(Z_{it} - z)) - \mathcal{H}_m(z)\right](1 - \omega)d\omega, \\
\mathfrak{R}^*(Z_{it}; z) &= \int_0^1 \left[D_m^3(z + \omega(Z_{it} - z), Z_{it} - z) - D_m^3(z, Z_{it} - z)\right](1 - \omega)^2 d\omega,
\end{aligned}
$$

where ω is a weight function. Now we analyze the limit behavior of each of these terms when N tends to infinity and T remains fixed. First we will show that

$$
A_{1.1N} = \frac{1}{2}\mu_2(K_{u_\tau})\mathcal{B}_{\ddot{X}\ddot{X}}(z, .., z)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathcal{H}_{m_r}(z)H\right\}\right\}i_d + o_p\left(\mathrm{tr}\left\{H\right\}\right), \tag{A.5}
$$

and using standard results from nonparametric regression analysis and Assumption 2.4 we have that

$$
\begin{aligned}
E(A_{1.1N}) &= \frac{1}{2}\int E\left[\ddot{X}_{it}X_{it}^\top \Big| Z_{i1} = z + H^{1/2}u_1, ..., Z_{iT} = z + H^{1/2}u_T\right] \otimes (H^{1/2}u_\tau)^\top \\
&\quad \times \mathcal{H}_m(z)(H^{1/2}u_\tau)f(z + H^{1/2}u_1, ..., z + H^{1/2}u_T)\prod_{l=1}^T K(u_l)du_l \\
&\quad - \frac{1}{2T}\sum_{s=1}^T \int E\left[\ddot{X}_{it}X_{is}^\top \Big| Z_{i1} = z + H^{1/2}u_1, ..., Z_{iT} = z + H^{1/2}u_T\right] \otimes (H^{1/2}u_s)^\top \\
&\quad \times \mathcal{H}_m(z)(H^{1/2}u_s)f(z + H^{1/2}u_1, ..., z + H^{1/2}u_T)\prod_{l=1}^T K(u_l)du_l.
\end{aligned}
$$

Then a straightforward application of a Taylor expansion and assumptions 2.1 and 2.5 are enough to show that (A.5) holds. Also, note that to show (A.5) we need to prove that $\mathrm{Var}(A_{1.1N}) \to 0$, as N tends to infinity and T is fixed. Under Assumption 2.1, $\mathrm{Var}(A_{1.1N}) = \frac{1}{NT}\mathrm{Var}(a_{it}) + \frac{1}{NT^2}\sum_{t=3}^T(T - t)\mathrm{Cov}(a_{i2}, a_{it})$, where $a_{it} = \frac{1}{|H|^{T/2}}\ddot{X}_{it}Q_m(z)K_H(Z_i - z)$. Then, under assumptions 2.5 and 2.9 the first element shows the following bound $\mathrm{Var}(a_{it}) \le \frac{C}{NT|H|^{T/2}}$ and $\mathrm{Cov}(a_{i2}, a_{it}) \le \frac{C'}{N|H|^{T/2}}$. Therefore, if $N|H|^{T/2}$ tends to infinity the variance tends to zero and applying a weak law of large numbers (A.5) follows.

Following a similar procedure, and noting that due to Assumption 2.4 the odd order moments of $K(.)$ disappear, it easy to show that

$$
A_{1.3N} = \frac{1}{2}\mu_2(K_{u_\tau})^2\left(1 - \frac{1}{T}\right)\mathcal{D}\mathcal{B}_{X_tX_t}^\top(z, .., z)\mathrm{diag}_d\left\{\mathrm{tr}\left\{\mathcal{H}_{m_r}(z)H^2\right\}i_d\right\} + o_p\left(\mathrm{tr}\left\{H^2\right\}\right), \tag{A.6}
$$

and

$$
A_{1.4N} = \frac{1}{3!}\left(1 - \frac{1}{T}\right)\mathcal{B}_{X_tX_t}(z, .., z) \otimes \int \left(H^{1/2}u_\tau\right)D_m^3(z, H^{1/2}u_\tau)\prod_{l=1}^T H(u_l)du_l + o_p\left(H^2\right). \tag{A.7}
$$

Finally focusing on the residual terms $A_{1.2N}$ and $A_{1.5A}$ and using the same procedure as in the proof of (A.18)-(A.28) in Rodriguez-Poo and Soberón (2015) it can be shown that $A_{1.2N} = o_p(\text{tr}\{H\})$ and $A_{1.5N} = o_p(\text{tr}\{H^2\})$. Then by replacing the different asymptotic expressions for the A_N's into U_{2N}, we obtain,

$$U_{2N} = \frac{1}{NT|H|^{T/2}} \sum_{i=1}^N E[T_{wi}(\beta(z))|\mathbb{X}, \mathbb{Z}] = \begin{pmatrix} \frac{1}{2}b_{w1}(z) \\ \frac{1}{2}b_{w2}(z) + \frac{1}{3!}b_{w3}(z) \end{pmatrix} + o_p\begin{pmatrix} \text{tr}\{H\} \\ \text{tr}\{H^2\} \end{pmatrix}, \tag{A.8}$$

where $b_{w1}(z)$, $b_{w2}(z)$ and $b_{w3}(z)$ were defined in (16).

Now we obtain the limiting distribution of the quantity $\sqrt{NT|H|^{T/2}}U_{1N}$. In order to do so we apply Liapunov's Central Limit Theorem. We do it by obtaining the variance-covariance matrix of the limiting distribution and verifying the so called Liapunov's condition. By substituting (7) into U_{1N} we obtain, $U_{1N} \equiv \frac{1}{NT|H|^{T/2}} \sum_{i=1}^N [T_{wi}(\beta(z)) - E[T_{wi}(\beta(z))|\mathbb{X}, \mathbb{Z}]] = \frac{1}{NT|H|^{T/2}} \sum_{it} \tilde{Z}_{it}^* v_{it} K_H(Z_i - z)$. Now, because of assumptions 2.1 and 2.2 we have that

$$NT\text{Var}(U_{1N}|\mathbb{X}, \mathbb{Z}) = \frac{\sigma_v^2}{NT|H|^T} \sum_{it} \tilde{Z}_{it}^* \tilde{Z}_{it}^{*\top} K_H^2(Z_i - z). \tag{A.9}$$

Applying assumptions 2.1 - 2.2 and 2.4 and mimicking (A.33)-(A.35) in Rodriguez-Poo and Soberón (2015) we obtain the following,

$$NT|H|^{T/2}Var(U_{1N}) \tag{A.10}$$
$$= \sigma_v^2 \begin{pmatrix} R(K)^T \mathcal{B}_{\ddot{X}\ddot{X}}(z,...,z) & O_p(|H|^T) \\ O_p(|H|^T) & \left(1 - \frac{1}{T}\right)\mu_2(K_{u_\tau}^2)\prod_{l\neq\tau}^T R(K_{u_l})\mathcal{B}_{X_tX_t}(z,...,z) \otimes H \end{pmatrix}(1 + o_p(1)).$$

Now, we check Liapunov's condition; we must show that for any unit vector $b \in \mathbb{R}^{d(q+1)}$ and some $\delta > 0$, as N tends to infinity, $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{it} E\left[\left|b^\top \tilde{Z}_{it}^* v_{it} \prod_{l=1}^T K_H(Z_{il} - z)\right|^{2+\delta}\right] \to 0$. To prove this, let us define $\phi_{it} = |H|^{T/4}b^\top \tilde{Z}_{it}^* v_{it} \prod_{l=1}^T K_H(Z_{il} - z) \, \forall i = 1,...,N; t = 1,...,T$. Following assumption 2.4 we can write

$$\text{Var}(\phi_{it}) = \sigma_v^2 b^\top \begin{pmatrix} R(K)^T \mathcal{B}_{\ddot{X}\ddot{X}}(z,...,z) & 0 \\ 0 & \left(1 - \frac{1}{T}\right)\mu_2(K_{u_\tau}^2)\prod_{l\neq\tau}^T R(K_{u_l})\mathcal{B}_{X_tX_t}(z,...,z) \otimes H \end{pmatrix} b(1 + o_p(1)),$$

and $\sum_{t=1}^T |Cov(\phi_{i1}, \phi_{it})| = o_p(1)$. Note also that we can write $\phi_{it} = \phi_{1it} + \phi_{2it}$ where $\phi_{1it} = b^\top \ddot{X}_{it} v_{it} \prod_{l=1}^T K_H(Z_{il} - z)$ and $\phi_{2it} = b^\top X_{it} \otimes (Z_{it} - z) - b^\top \frac{1}{T}\sum s = 1^T X_{is} \otimes (Z_{is} - z)v_{it}\prod_{l=1}^T K_H(Z_{il} - z)$. Furthermore, let us define $\phi_{n,i}^* = T^{-1/2}\sum_{t=1}^T \phi_{it} = T^{-1/2}\sum_{t=1}^T(\phi_{1it} + \phi_{2it})$. For fixed T, the $\phi_{n,i}^*$ are independent random variables and $n = NT$. Then, using Minkowski inequality and due to the matrix structure of $\tilde{Z}_{it}^*$ we get $E|\phi_{n,i}^*|^{2+\delta} \leq CT^{(2+\delta)/2}E|\phi_{it}|^{2+\delta} = CT^{(2+\delta)/2}E|\phi_{1it} + \phi_{2it}|^{2+\delta}$. Analysing each term by separate, (see Rodriguez-Poo and Soberón (2015) for details), we obtain

$$(NT)^{-(2+\delta)/2} \sum_{i=1}^N E|\phi_{n,i}^*|^{2+\delta} \leq C(N|H|^{T/2})^{-\delta/2}, \tag{A.11}$$

which tends to zero when $N|H| \to \infty$. Therefore, Liapunov's Central Limit Theorem applies and hence

$$\sqrt{NT|H|^{T/2}}U_{1N} \to_d \mathcal{N}(0, \upsilon_w(z)). \tag{A.12}$$

Therefore, by substituting (A.8) and (A.12) into (A.1) and imposing the condition $\sqrt{NT|H|^{T/2}}tr(H) \to 0$ we obtain that $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_i T_{wi}(\beta(z)) \to_d \mathcal{N}(0, \upsilon_w(z))$, as N tends to infinity.

Now we prove the consistency of $\tilde{D}_w(\beta(z))$. As N tends to infinity and T is fixed, if conditions 2.1 - 2.10 hold and, similar to the proof of (A.10), by applying a Law of Large Numbers it is straightforward to show that

$$\tilde{D}_w(\beta(z)) = \frac{1}{NT|H|^{T/2}} \sum_{i=1}^{N} T_{wi}(\beta(z))T_{wi}^\top(\beta(z)) = \upsilon_w(z) + o_p(\text{tr}\{H\}), \tag{A.13}$$

where $\upsilon_w(z)$ was defined in (17). From (A.8), (A.12) and (A.13), and using the same arguments as in the proof of (2.14) in Owen (1990), we can prove that

$$\lambda = O_p\left((NT|H|^{T/2})^{-1/2}\right), \tag{A.14}$$

where λ was defined in (11). Then applying Taylor expansion to (10) and invoking (A.8), (A.12) and (A.13), we obtain

$$\mathcal{R}_w(\beta(z)) = 2\sum_{i=1}^{N}\left[T_{wi}^\top(\beta(z))\lambda - \left(T_{wi}^\top(\beta(z))\lambda\right)^2/2\right] + o_p(1). \tag{A.15}$$

By (11) and applying Taylor expansion again it follows that

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

Then, recalling (A.8), (A.12) and (A.13) we can prove that

$$\sum_{i=1}^{N}(T_{wi}^\top(\beta(z))\lambda)^2 = \sum_{i=1}^{N} T_{wi}^\top(\beta(z))\lambda + o_p(1), \tag{A.16}$$

and

$$\lambda = \left[\sum_{i=1}^{N} T_{wi}(\beta(z))T_{wi}^\top(\beta(z))\right]^{-1} \sum_{i=1}^{N} T_{wi}(\beta(z)) + o_p\left((NT|H|^{T/2})^{-1/2}\right). \tag{A.17}$$

Now, if we rely on (A.8), (A.12) and (A.13) the proof is concluded by applying the Cramer-Wold device.

∎

8.2 Proof of Theorem 2.2

This proof is similar to that of Theorem 2.1 and therefore most of the details are omitted. In order to obtain the asymptotic distribution of $\frac{1}{\sqrt{NT|H|}} \sum_{i=1}^{N} T_{fi}(\beta(z))$ we follow similar steps to those in (A.1),

$$\frac{1}{NT|H|} \sum_{i=1}^{N} T_{fi}(\beta(z)) = U_{1N}^* + U_{2N}^*. \tag{A.18}$$

For the bias term U_{2N}^*, defining a similar multivariate Taylor expansion around $X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)})$ as the one used in (A.2), and applying Assumption 2.2 we obtain

$$U_{2N}^* = \frac{1}{NT|H|} \sum_{i=1}^{N} E[T_{fi}(\beta(z))|\mathbb{X}, \mathbb{Z}] = \begin{pmatrix} \frac{1}{2} b_{f1}(z) \\ \frac{1}{2} b_{f2}(z) + \frac{1}{3!} b_{f3}(z) \end{pmatrix} + o_p \begin{pmatrix} \operatorname{tr}\{H\} \\ \operatorname{tr}\{H^2\} \end{pmatrix}, \tag{A.19}$$

where $b_{f1}(z)$, $b_{f2}(z)$ and $b_{f3}(z)$ were defined in (21).

Now we obtain the limiting distribution of the quantity $\sqrt{NT|H|} U_{1N}^*$. By substituting (8) into U_{1N}^* we obtain that $U_{1N}^* = \frac{1}{NT|H|} \sum_{it} \tilde{Z}_{it} \Delta v_{it} K_H(Z_{it} - z, Z_{i(t-1)} - z)$, and taking into account that because of assumptions 2.1 and 2.2 we have that

$$E[\Delta v_{it} \Delta v_{i't'}|\mathbb{X}, \mathbb{Z}] = \begin{cases} 2\sigma_v^2 & \text{if } i = i', t = t' \\ -\sigma_v^2 & \text{if } i = i', |t - t'| < 2 \\ 0 & \text{otherwise} \end{cases}. \tag{A.20}$$

Then, mimicking (A.30)-(A.36) in Rodriguez-Poo and Soberón (2014) we obtain the following,

$$NT|H| Var\left(U_{1N}^*\right) \tag{A.21}$$

$$= 2\sigma_v^2 \begin{pmatrix} R(K_u)R(K_v)\mathcal{B}_{\Delta X, \Delta X}(z,z) & O_p(|H|^2) \\ O_p(|H|^2) & \mu_2(K^2)R(K_u)\left(\mathcal{B}_{XX}(z,z) + \mathcal{B}_{X_{-1}X_{-1}}(z,z)\right) \otimes H \end{pmatrix} (1 + o_p(1)).$$

The rest of the proof follows exactly the lines of the proof of Theorem 2.1.

∎

8.3 Proof of Theorem 3.1

Note that,

$$\hat{\beta}_w(z) - \beta(z) = \left(\hat{\beta}_w(z) - E\left[\hat{\beta}_w(z)\big| \mathbb{X}, \mathbb{Z}\right]\right) + \left(E\left[\hat{\beta}_w(z)\big| \mathbb{X}, \mathbb{Z}\right] - \beta(z)\right) \equiv \mathbf{I}_{1N} + \mathbf{I}_{2N}. \tag{A.22}$$

To prove the desired result we will show that, under the conditions of this theorem, $\mathbf{I}_{2N} = B_w(z) + o_p(\frac{1}{\sqrt{NT|H|^{T/2}}})$ and $\sqrt{NT|H|^{T/2}} \mathbf{I}_{1N} \to_d \mathcal{N}(0, \Sigma_w(z))$, as N tends to infinity, where $B_w(z)$ and $\Sigma_w(z)$ have

been defined in theorem 3.1. If we substitute (14) into (A.22) and we make a second order Taylor expansion around $X_{it}^\top m(Z_{it}) - \frac{1}{T}\sum_s X_{is}^\top m(Z_{is})$ we obtain that

$$\mathbf{I}_{2N} = \left(\frac{1}{NT|H|^{T/2}}\sum_{it} K_H(Z_i - z)\tilde{Z}_{it}^* \tilde{Z}_{it}^{*\top}\right)^{-1} \left(\begin{array}{c} A_{1.1N} + A_{1.2N} \\ A_{1.3N} + A_{1.4N} + A_{1.5N} \end{array}\right). \tag{A.23}$$

Note that $A_{1.1N}$, $A_{1.2N}$, $A_{1.3N}$, $A_{1.4N}$ and $A_{1.5N}$ have been already defined in the proof of Theorem 2.1. Furthermore, the asymptotic behavior of the second term in (A.23) has been already obtained in (A.8); therefore, all what we need to calculate the asymptotic behavior of $\mathbf{I}_{2N}$ is to study the first term. Proceeding as in Rodriguez-Poo and Soberón (2015) (see expressions (A.8)-(A.12)), it is straightforward to show that

$$\left(\frac{1}{NT|H|^{T/2}}\sum_{it}\prod_{l=1}^{T} K_H(Z_{il} - z)\tilde{Z}_{it}^* \tilde{Z}_{it}^{*\top}\right)^{-1} = \left(\begin{array}{cc} \mathfrak{C}_{11} & \mathfrak{C}_{12} \\ \mathfrak{C}_{21} & \mathfrak{C}_{22} \end{array}\right), \tag{A.24}$$

where

$$
\begin{aligned}
\mathfrak{C}_{11} &= \mathcal{B}_{\ddot{X}\ddot{X}}^{-1}(z,...,z) + o_p(1) \\
\mathfrak{C}_{12} &= -\mathcal{B}_{\ddot{X}\ddot{X}}^{-1}(z,...,z)\left(\mathcal{DB}_{\ddot{X}\ddot{X}}(z,...,z)\left(I_d \otimes \mu_2(K_{u_\tau})H\right)\right) \\
&\quad \times \left(\left(1 - \frac{1}{T}\right)\mathcal{B}_{X_t X_t}(z,..,z) \otimes \mu_2(K_{u_\tau})H\right)^{-1} + o_p(1) \\
\mathfrak{C}_{21} &= -\left(\left(1 - \frac{1}{T}\right)\mathcal{B}_{X_t X_t}(z,..,z) \otimes \mu_2(K_{u_\tau})H\right)^{-1}\left(\mathcal{DB}_{\ddot{X}\ddot{X}}(z,...,z)\left(I_d \otimes \mu_2(K_{u_\tau})H\right)\right)^\top \\
&\quad \times \mathcal{B}_{\ddot{X}\ddot{X}}^{-1}(z,...,z) + o_p(1) \\
\mathfrak{C}_{22} &= \left(\left(1 - \frac{1}{T}\right)\mathcal{B}_{X_t X_t}(z,..,z) \otimes \mu_2(K_{u_\tau})H\right)^{-1} + o_p(H^{-1}),
\end{aligned}
$$

Using the terms (A.24) and (A.8) and applying Slutsky's Theorem to (A.23) we finish the proof.

In order to show the asymptotic behavior of $\mathbf{I}_{1N}$ note that by (14) we have that

$$\mathbf{I}_{1N} = \hat{\beta}_w(z) - E\left[\hat{\beta}_w(z)\Big|\mathbb{X},\mathbb{Z}\right] = \left(\sum_{it} K_H(Z_i - z)\tilde{Z}_{it}^* \tilde{Z}_{it}^{*\top}\right)^{-1}\sum_{it} K_H(Z_i - z)\tilde{Z}_{it}^* v_{it}, \tag{A.25}$$

and considering assumptions 2.1 and 2.2 the variance term of $\hat{\beta}_w(z)$, using the Slutsky's Theorem and previous results can be written as

$$NT|H|^{T/2}\mathrm{Var}\left(\hat{\beta}_w(z)\Big|\mathbb{X},\mathbb{Z}\right) = \Sigma_w(z)(1 + o_p(1)); \tag{A.26}$$

the Liapunov condition needed to apply a Central Limit Theorem here is the same as the one as in the proof of Theorem 2.1 (see A.11) and then a further application of the Cramer-Wold device closes the proof.

■

8.4 Proof of Theorem 3.2

The proof of this theorem is similar to that of Theorem 3.1 and therefore the details are omitted.

∎

In order to show Theorems 4.1 and 4.2 we need the following additional lemma that is proved at the end of the Appendix.

Lemma 8.1. Assuming that conditions of Theorems 4.1 and 4.2 hold. Then $\hat{b}_w(z) \to^p b_w(z)$ and $\hat{b}_f(z) \to^p b_f(z)$ respectively, as N tends to infinity.

8.5 Proof of Theorem 4.1

If we substitute (12) into (19) we obtain that $\widetilde{\mathcal{R}}_w(\beta(z))$ is equal to

$$
= \left[\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) - \sqrt{NT|H|^{T/2}}\, b_w(z) \right]^{\top} \tilde{D}_w^{-1}(\beta(z)) \left[\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) - \sqrt{NT|H|^{T/2}}\, b_w(z) \right]
$$

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

$$
\equiv L_1(z) + L_2(z) + o_p(1).
$$

In Theorem 2.1, we have already proved that $\frac{1}{\sqrt{NT|H|^{T/2}}} \sum_{i=1}^{N} T_{wi}(\beta(z)) \to^d \mathcal{N}\left(\sqrt{NT|H|^{T/2}}\, b_w(z), \upsilon_w(z) \right)$ and $\tilde{D}_w(\beta(z)) \to_p \upsilon_w(z)$. Then, together with the proof of Lemma 8.1 we can conclude that $L_1(z) \to^d \chi^2_{d(q+1)}$ and $L_2(z) \to_p 0$. Thus the proof of Theorem 4.1 is closed.

∎

8.6 Proof of Theorem 4.2

The proof of this theorem is similar to that of Theorem 4.1 and therefore the details are omitted. ∎

28

8.7   Proof of Lemma 8.1

Let us consider,

$$
\begin{aligned}
&\hat{b}_w(z) - b_w(z)\\
&= \frac{1}{NT|H|^{T/2}} \sum_{it} Z_{it}^* \left(X_{it}^\top (\hat{m}_w(Z_{it}) - m(Z_{it})) - \frac{1}{T} \sum_s X_{is}^\top (\hat{m}_w(Z_{is}) - m(Z_{is})) \right.\\
&\quad \left. - Z_{it}^{*\top} \left(\hat{\beta}_w(z) - \beta(z) \right) \right) K_H (Z_i - z)\\
&\quad + \frac{1}{NT|H|^{T/2}} \sum_{it} Z_{it}^* \left(X_{it}^\top m(Z_{it}) - \frac{1}{T} \sum_s X_{is}^\top m(Z_{is}) - Z_{it}^{*\top} \beta(z) \right) K_H (Z_i - z)\\
&\quad - b_w(z)\\
&= L_1^*(z) + L_2^*(z).
\end{aligned}
$$

Then, by Theorem 2.1, equation (A.8), we have that, as N tends to infinity, $L_2^*(z) \to_p 0$. Furthemore, the conditions of this Theorem guarantee that $\sup_z |\hat{m}_w(z) - m(z)| = o_p(1)$ and $\sup_z \left| \hat{\beta}_w(z) - \beta(z) \right| = o_p(1)$ (see Masry (1996), Theorem C) and jointly with assumption 2.9 it is easy to show that $L_1^*(z) \to_p 0$

∎

References

Ahmad, I., S. Leelahanon, and Q. Li (2005). Efficient estimation of a semiparametric partially linear varying coefficient model. *Annals of Statistics 33*, 258–283.

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

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

Fan, J. (1993). Local linear regression smoothers and their minimax efficiency. *The Annals of Statistics 21*, 196–216.

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

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

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

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

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

Hsiao, C. (2003). *Analysis of panel data*, Volume 34. Cambridge university press.

Li, C. and Z. Liang (2015). "Asymptotics for nonparametric and semiparametric fixed effects panel models". *Journal of Econometrics 185*, 420–434.

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

Lin, X. and R. J. Carroll (2000). "Nonparametric function estimation for clustered data when the predictor is measured without/with error". *Journal of the American Statistical Association 95*, 520–534.

Martins-Filho, C. and F. Yao (2009). Nonparametric regression estimation with general parametric error covariance. *Journal of Multivariate Analysis 100*(3), 309–333.

Masry, E. (1996). "Multivariate local polynomial regression for time series: uniform strong consistency rate". *Journal of Time Series Analysis 17*, 571–599.

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.

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

Ruckstuhl, A. F., A. H. Welsh, and R. J. Carroll (2000). "Nonparametric function estimation of the relationship between two repeatedly measured variables". *Statistica Sinica 10*, 51–71.

Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 683–690.

Su, L. and X. Lu (2013). "Nonparametric dynamic panel data models: Kernel estimation and specification testing". *Journal of Econometrics 176*, 112–133.

Su, L. and A. Ullah (2007). "More efficient estimation of nonparametric panel data models with random effects". *Economics Letters 96*, 375–380.

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.

Su, L., A. Ullah, and Y. Wang (2010). A note on nonparametric regression with general parametric error covariance. *Empirical Economics 45*, 1009–1024.

Sun, Y., R. J. Carroll, and D. Li (2009). "Semiparametric estimation of fixed effects panel data varying coefficient models". *Advances in Econometrics 25*, 101–130.

Wang, H.-J. and C.-W. Ho (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics 157*(2), 286–296.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data.* The MIT press.

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