Nonparametric and semi-parametric panel data models: recent developments

Alexandra Soberón* Juan M. Rodriguez-Poo†

July 26, 2016

Abstract

In this paper, we provide an intensive review of the recent developments for semi-parametric and fully nonparametric panel data models that are linearly separable in the innovation and the individual specific term. We analyze these developments under two alternative model specifications: Fixed and random effects panel data models. More precisely, in the random effects setting we focus our attention in the analysis of some efficiency issues that have to do with the so-called *working independence condition*. This assumption is introduced when estimating the asymptotic variance-covariance matrix of nonparametric estimators. In the fixed effects setting, to cope with the so-called *incidental parameters problem*, we consider two different estimation approaches: profiling techniques and differencing methods. Furthermore, we are also interested in the endogeneity problem and how instrumental variables are used in this context. In addition, for practitioners, we also show different ways of avoiding the so-called *curse of dimensionality* problem in pure nonparametric models. In this way, semi-parametric and additive models appear as a solution when the number of explanatory variables is large.[1]

Keywords: nonparametric; semi-parametric; panel data models; random effects; fixed effects

*Departamento de Economía, Universidad de Cantabria, Avda. Los Castros s/n, E-39005 Santander, Spain.

†Departamento de Economía, Universidad de Cantabria, Avda. Los Castros s/n, E-39005 Santander, Spain and Université de Genve, Bd du Pont d'Arve 40, CH - 1211 Genève, Switzerland.

[1]The authors acknowledge financial support from the Programa Nacional de Formación de Profesorado Universitario/ Spanish Ministery of Education. Ref. ECO2013-48326-C2-2-P. In addition, this work is part of the research Project APIE 1/2015-17: "New methods for the empirical analysis of financial markets" of the Santander Financial Institute (SANFI) of UCEIF Foundation resolved by the University of Cantabria and funded with sponsorship from Banco Santander

1 Introduction

In empirical research, the complexity of econometric models has been greatly enriched by the availability of panel data sets. These data are characterized by the observation of a group of individuals (households, consumers, countries, and so on) over time, so they allow us to extract some unknown information about the idiosyncratic characteristics of individuals. From a theoretical point of view, this double index enables us to specify econometric models that account both for the impact of unobserved actions of individuals and observable individual characteristics (explanatory variables). Hence, through the use of panel data econometric models, under some standard assumptions on the data generating process, it is possible to draw inference on the parameters of interest that otherwise would be impossible to obtain. As it is often the case in applied econometrics, we are interested in partial effects of the observable explanatory variables in the population regression (quantile) function but, following the approach in Chamberlain (1984), when there exists time-invariant or/and individual invariant omitted latent variables.

In this context, the statistical properties of the estimators of the unknown parameters are going to depend crucially on the set of assumptions that we are willing to impose on the relationship between the observable explanatory variables and the unobserved effects in the conditioning set. On the one hand, we might consider the unobserved individual heterogeneity as statistically independent from the observed explanatory variables (the so-called random effects case). Then, the individual heterogeneity is just another unobserved factor affecting the explanatory variable that is not systematically related to the observed explanatory variables whose effects are of interest. On the other hand, in empirical applications, many times this assumption is too strong and therefore applied researchers prefer to avoid it by allowing for some type of statistical dependence between individual time-varying heterogeneity and explanatory variables. More precisely, for example, it is commonly assumed that the expected value of the random heterogeneity term, conditionally on the set of values of the explanatory variables, is constant and varies only across individuals. This is the so-called fixed effects model. Under this assumption, if the number of time observations (T) is fixed, it raises the *incidental parameters* issue because when the sample size increases, i.e., when the number of individuals (N) grows, the number of parameters to be estimated also increases (see Lancaster (2000) for a survey). The classical law of large numbers or central limit theorems relies on the assumption that the number of unknown parameters to be estimated remains fixed as the sample size increases. Therefore, in this case they do not apply straightforwardly (see Neyman and Scott (1948)). Under this setting, it is clear that standard estimation techniques for random effects panel data models might result in miss-leading inferential results and hence, more specific estimation techniques are needed.

If panel data models are linearly separable in the innovation and the individual specific term, a simple linear transformation can eliminate the random individual heterogeneity from the transformed model (see Anderson and Hsiao (1981)). However, if this relationship is nonlinear there is no general rule of transformation to eliminate the incidental parameters existence. In this case,

a specific structure for the nonlinear model needs to be specified in order to find an appropriate transformation to eliminate the incidental parameters. We refer to Arellano (2003), Baltagi (2013) or Hsiao (2003) for an intensive review of techniques devoted to estimate panel data models and Maddala (1987) to obtain good arguments about random versus fixed effects. Finally, although from the applied point of view fixed effect models seem to be more popular and useful, random effects models are still important for economics and statistics. For example, in situations where the researcher is interested in estimating time invariant effects (without using instrumental variables) which, in general, do not work with fixed effects models.

Nevertheless, the suitable treatment of these unobserved heterogeneity effects is not enough to guarantee proper statistical properties for the estimators of interest. In most cases, estimation of the parameters of interest also depends on some statistical restrictions imposed on the data generating process. However, sometimes these assumptions are too restrictive with respect to functional forms or densities and the risk of misspecification is high. If this is the case, the resulting estimators can lead us to missleading inference. In this context, nonparametric panel data models are very appealing since they do not make too restrictive assumptions on the specification of the model and they allow data to tailor the shape of the regression function by themselves. However, in some situations this flexibility presents some drawbacks. First, it can be unable to incorporate prior information so the resulting estimator for the unknown function tends to have a higher variance term. Second, it is subject to the so-called curse of dimensionality which practically disables standard nonparametric methods when the number of explanatory variables is high. In order to solve these shortcomings, semi-parametric panel data models appear as a reasonable compromise between fully nonparametric and parametric models. In fact, they enable us to incorporate some prior information coming from economic theory or past experience by keeping at the same time more flexibility in the specification of the model. Furthermore, although there is a nonparametric part that shows a slower rate of convergence, the estimators obtained from the parametric part do exhibit the same statistical properties as if the whole model would be fully parametric. That is the so-called $\sqrt{N}$-consistency property, see among others Robinson (1988) or Speckman (1988). For early discussions on semi-parametric panel data models see Ullah and Roy (1998), while we refer to Ai and Li (2008) for a review about partially linear and limited dependent nonparametric and semi-parametric panel data models.

In this paper, we provide an intensive review of the recent developments for semi-parametric and fully nonparametric panel data models that are linearly separable in the innovation and the individual specific term. Furthermore, we analyze these developments under two alternative settings, the so-called fixed and random effects panel data models. Note that Su and Ullah (2011) focus on similar modelings, although in this case we include the most recent results and pay special attention to the so-called *incidental parameters problem* as well as with endogenous explanatory variables. Meanwhile, in Chen et al. (2013) this type of models are studied when deterministic trends and single-index specifications are present.

The rest of the paper is organized as follows. In Sections 2 and 3, we analyze the literature about

nonparametric panel data models with random and fixed effects, respectively. In Section 4, we focus on semi-parametric models with random effects. In Section 5, we study the corresponding models with fixed effects. Section 6 refers to nonparametric and semi-parametric panel data models when the presence of endogenous explanatory variables is allowed. Finally, Section 7 concludes.

2 Nonparametric panel data models with random effects

The basic nonparametric unobserved effects model can be written, for a randomly drawn cross-section observation i, as

$$Y_{it} = m\left(Z_{it}\right) + \epsilon_{it}, \quad t = 1, \cdots, T, \tag{2.1}$$

where Z_{it} is a $q \times 1$ vector of observable explanatory variables, $m(\cdot)$ is an unknown function that needs to be estimated, and ϵ_{it} is an unobservable error term. Typically, in panel data analysis the error term of the model follows a one-way error component structure of the form

$$\epsilon_{it} = \mu_i + v_{it}, \quad t = 1, \cdots, T, \tag{2.2}$$

where v_{it} is referred as the idiosyncratic error term and μ_i is called the unobserved individual heterogeneity. Through the paper we will assume the following,

$$E\left(Y_{it} \middle| Z_{i1}, \cdots, Z_{iT}, \mu_i\right) = E\left(Y_{it} \middle| Z_{it}, \mu_i\right) = m\left(Z_{it}\right) + \mu_i. \tag{2.3}$$

Note that, the first equality establishes the relationship between the dependent variable Y and the past values of Z. Furthermore, the second equality constrains the regression function to be the sum of a nonparametric function, $m\left(Z_{it}\right)$, plus an unobservable heterogeneity term, μ_i. Using (2.1) and (2.2) note that assumption (2.3) can be stated in terms of the idiosyncratic errors as

$$E\left(v_{it} \middle| Z_{i1}, \cdots, Z_{iT}, \mu_i\right) = 0, \quad t = 1, \cdots, T. \tag{2.4}$$

Let $v_i = \left(v_{i1}, \cdots, v_{iT}\right)^\top$ be a $T \times 1$ vector. The error vector v_i and the heterogeneity term μ_i are such that

$$E\left(v_i v_i^\top \middle| Z_{i1}, \cdots, Z_{iT}, \mu_i\right) = \sigma_v^2 I_T, \quad E\left(\mu_i^2 \middle| Z_{i1}, \cdots, Z_{iT}\right) = \sigma_\mu^2, \tag{2.5}$$

where I_T is a $T \times T$ identity matrix.

Furthermore, let $\epsilon_i = \left(\epsilon_{i1}, \cdots, \epsilon_{iT}\right)^\top$ be a $T \times 1$ vector and let $E(\epsilon_i \epsilon_i^\top)$ be a $T \times T$ matrix. Under the assumptions above note that

$$E(\epsilon_i \epsilon_i^\top) = \sigma_v^2 I_T + \sigma_\mu^2 \imath_T \imath_T^\top, \tag{2.6}$$

and

$$\Omega = E(\epsilon\epsilon^\top) = I_N \otimes \left[\sigma_1^2 \imath_T \imath_T^\top / T + \sigma_v^2 (I_T - \imath_T \imath_T^\top / T)\right] = I_N \otimes V, \tag{2.7}$$

where $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$, $V = \sigma_1^2 \imath_T \imath_T^\top / T + \sigma_v^2 \left(I_T - \imath_T \imath_T^\top / T\right)$ and $\imath_T$ is a T-dimensional vector of ones.

Finally note that for a randomly drawn cross-section observation i, the vector of explanatory variables $Z_{i1}, \cdots, Z_{iT}$ is strictly stationary, whereas for fixed t the vector $Z_{1t}, \cdots, Z_{Nt}$ are independent and identically distributed (*i.i.d.*) random variables. In general, the asymptotic behavior of the estimators that appear in the paper is analyzed in the standard panel data framework where N tends to infinity and T is fixed. In those particular cases where other asymptotic behavior is needed it will be pointed out.

All previous assumptions will be common for both random and fixed effects nonparametric panel data models. Now, to characterize the random effects model, we further assume that

$$E\left(\mu_i | \, Z_{i1}, \cdots, Z_{iT}\right) = E\left(\mu_i\right) = 0. \tag{2.8}$$

Note that using (2.1) and (2.2), applying assumptions (2.3), (2.4), (2.5) and (2.8) and by the law of iterated expectations we obtain that

$$E\left(Y_{it} | \, Z_{i1}, \cdots, Z_{iT}\right) = m\left(Z_{it}\right), \quad t = 1, \cdots, T. \tag{2.9}$$

Hence, the function $m(\cdot)$ and its first order derivatives can be directly estimated through a pooled standard nonparametric technique. However, the resulting estimator is inefficient given that the composed error term is serially correlated by the presence of μ_i in each time period. Hence, it should be possible to improve the efficiency of the estimator by taking into account the information contained in the variance-covariance matrix. Among others, in Ullah and Roy (1998), Lin and Carroll (2000) and Su and Ullah (2007) several nonparametric estimators of $m(\cdot)$ and its derivatives are considered. Furthermore, and with the aim of achieving efficiency, in Ruckstuhl et al. (2000), Wang (2003) and Henderson and Ullah (2005) different strategies are proposed to incorporate the information contained in the disturbances.

2.1 Local linear least-squares versus Nadaraya-Watson estimators

For any $z \in A$, where A is a compact subset in $\mathbb{R}^q$, the basic idea behind the standard nonparametric estimation of $m(z) = E(Y_{it}|Z_{it} = z)$ is to obtain a smoothed average of the Y_{it} values by taking into account the values of Z_{it} contained in a small interval around z. In order to understand further developments, it is useful to start with the analysis of the univariate case, where $q = 1$. Then, taking a Taylor expansion of the unknown smooth function $m(\cdot)$ around z, we obtain,

$$m\left(Z_{it}\right) \approx m(z) + m'(z)\left(Z_{it} - z\right) + \frac{1}{2} m''(z)\left(Z_{it} - z\right)^2 + \cdots + \frac{1}{p!} m^{(p)}(z)\left(Z_{it} - z\right)^p. \tag{2.10}$$

The above exposition suggests that we can estimate $m(z), m'(z), ..., m^{(p)}(z)$ by regressing Y_{it} on the terms $(Z_{it} - z)^\lambda$, for $\lambda = 0, 1, \cdots, p$, with kernel weights. Thus, the quantities of interest can be estimated by minimizing the following criterion function,

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left(Y_{it} - \gamma_0 - \gamma_1 (Z_{it} - z) - \gamma_2 \left(Z_{it} - z\right)^2 - \cdots - \gamma_p \left(Z_{it} - z\right)^p\right)^2 K_h(Z_{it} - z), \tag{2.11}$$

with respect to the values, $\gamma_0, \cdots, \gamma_p$, where $\gamma_0 = m(z)$, $\gamma_1 = m'(z)$ and $\gamma_p = m^{(p)}(z)$. Let us denote by $\widehat{\gamma}_0, \cdots, \widehat{\gamma}_p$ the solution to the minimization problem. Then, the above exposition suggests that $\widehat{m}(z;h) = \widehat{\gamma}_0$, $\widehat{m}'(z;h) = \widehat{\gamma}_1$ and $\widehat{m}^{(p)}(z;h) = \widehat{\gamma}_p$. Note that h is the bandwidth that needs to be selected empirically and $K_h(u) = \frac{1}{h}K(u/h)$ is the so-called kernel function that must fulfill the following conditions,

$$\int K(u)du = 1, \qquad \int K^2(u)du < \infty \qquad \text{and} \qquad \int u^2 K(u)du < \infty. \tag{2.12}$$

The kernel function is a weight function defined in such way that, for fixed h, it takes values close to zero when Z_{it} is far away from z. The solution to the problem above is the so-called local polynomial regression (see Ruppert and Wand (1994), Fan and Gijbels (1995) and Zhan-Qian (1996) for a detailed description of this technique). As it is pointed out in Ullah and Roy (1998), for $p = 0$, (2.11) becomes

$$\sum_{it} \left(Y_{it} - \gamma_0\right)^2 K_h\left(Z_{it} - z\right) \tag{2.13}$$

and the value of γ_0 that minimizes (2.13) is

$$\widehat{\gamma}_0 \doteq \widehat{m}_{NW}(z;h) = \frac{\sum_{i=1}^N \sum_{t=1}^T K_h(Z_{it} - z)Y_{it}}{\sum_{i=1}^N \sum_{t=1}^T K_h(Z_{it} - z)}. \tag{2.14}$$

This is the so-called Naradaya-Watson estimator proposed alternatively in Nadaraya (1964) and Watson (1964). When $p = 1$ and $q > 1$, the previous Taylor expansion can be rewritten as

$$m(Z_{it}) \approx m(z) + (Z_{it} - z)^\top D_m(z),$$

where $D_m(z) = vec(\partial m(z)/\partial z^\top)$ is a $q \times 1$ vector of partial derivatives of the function $m(z)$ with respect to the elements of the q-vector z. Then, as it is suggested in Ullah and Roy (1998), Lin and Carroll (2000), Ruckstuhl et al. (2000), Henderson and Ullah (2005) and Su and Ullah (2007), among others, $m(z)$ and its first-order derivatives are estimated by minimizing the following criterion function

$$\sum_{i=1}^N \sum_{t=1}^T \left(Y_{it} - \gamma_0 - \gamma_1^\top (Z_{it} - z)\right)^2 K_h(Z_{it} - z), \tag{2.15}$$

where we denote by $\widehat{\gamma} = (\widehat{\gamma}_0 \quad \widehat{\gamma}_1)^\top$ a $(1+q)$-vector that minimizes (2.15). Thus, the above exposition suggests $\widehat{m}(z;h) = \widehat{\gamma}_0$ and $\widehat{D}_m(z;h) = \widehat{\gamma}_1$ as estimators for $m(z)$ and $D_m(z)$, respectively.

Assuming $Z_z^\top K_z Z_z$ is nonsingular, the solution (2.15) in matrix form can be written as

$$\begin{pmatrix} \widehat{\gamma}_0 \\ \widehat{\gamma}_1 \end{pmatrix} = (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z Y, \qquad \text{where} \qquad Z_z = \begin{bmatrix} 1 & (Z_{11} - z)^\top \\ \vdots & \vdots \\ 1 & (Z_{NT} - z)^\top \end{bmatrix} \tag{2.16}$$

is a $NT \times (1+q)$ matrix, $K_z = diag\left(K_h(Z_{11} - z), \cdots, K_h(Z_{NT} - z)\right)$ is a $NT \times NT$ diagonal matrix and $Y = (Y_{11}, \cdots, Y_{NT})^\top$ is a NT-dimensional vector. Then, the local linear least-squares (LLLS) estimator for $m(z)$ is

$$\widehat{\gamma}_0 \doteq \widehat{m}_{LLLS}(z;h) = e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z Y, \tag{2.17}$$

where e_1 is a $(1+q)$ selection vector having 1 in the first entry and all other entries 0.

Under the previous assumptions, imposing some smoothness conditions on both $m(\cdot)$ and $f(\cdot)$ and letting $h \to 0$ in such a way that $Nh^q \to \infty$, as $N \to \infty$ and T is fixed, in Kneisner and Li (1996) it is shown that

$$\sqrt{Nh^q}\left(\widehat{m}_{NW}(z;h) - m(z) - b_{NW}(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, v(z)\right) \tag{2.18}$$

and

$$\sqrt{Nh^q}\left(\widehat{m}_{LLLS}(z;h) - m(z) - b_{LLLS}(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, v(z)\right). \tag{2.19}$$

Let $\mathcal{H}_m(z)$ be the Hessian matrix of $m(\cdot)$ and let $D_f(z)$ be the first order derivative vector of the density function of $q \times q$ and $1 \times q$ dimension, respectively, the conditional bias terms are

$$b_{NW}(z) = \frac{h^2}{2T}\mu_2(K)\left[tr\left(\mathcal{H}_m(z)\right) + 2D_m(z)D_f(z)f(z)^{-1}\right], \tag{2.20}$$

$$b_{LLLS}(z) = \frac{h^2}{2T}\mu_2(K)tr\left(\mathcal{H}_m(z)\right). \tag{2.21}$$

Based on these results we can highlight that although the asymptotic variance of these two estimators is the same, the bias is not. More precisely, the bias of the local linear least-squares estimator only depends on the curvature of $m(\cdot)$ at z in a particular direction whereas the bias term of the Nadaraya-Watson estimator emerges mainly from both the curvature of $m(\cdot)$ and the term $D_m(z)D_f(z)f(z)^{-1}$. Furthermore, it is well-known that the local linear estimator usually exhibits a better performance near the boundary of the support of the density function; see Fan (1993) for more details. Note that, under the conditions above, both bias and variance terms converge at the same rate, $\sqrt{Nh^q}$. This makes the comparison of both estimators in terms of relative efficiency rather difficult because we would need to compare asymptotic mean squared errors. Just to avoid it, we choose the sequence of bandwidths, $h \equiv h(N)$, in such a way that $NTh^{4+q} \to 0$, as N tends to infinity. By imposing this rate the variance term dominates the bias asymptotically, and therefore we can compare variance terms only.

Note that, by assuming just conditional heteroskedastic errors in ϵ, the variance term takes the form

$$v(z) = \frac{\sigma_\epsilon^2(z)R(K)}{Tf(z)}, \tag{2.22}$$

where $\sigma_\epsilon^2(z)$ can be replaced by its consistent estimator, $\widehat{\sigma}_\epsilon^2 = (\imath_{NT}^\top K_z \imath_{NT})^{-1}\imath_{NT}^\top K_z \widehat{\epsilon}^2$. In this case, $\widehat{\epsilon}^2$ is the vector of nonparametric squared residuals. If in the previous expression we impose condition (2.5) then the conditional variance-covariance matrix can be written as

$$e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z \Omega K_z Z_z (Z_z^\top K_z Z_z)^{-1} e_1, \tag{2.23}$$

where Ω is the variance-covariance matrix of the error term defined as in (2.7). Finally, note that under the assumptions above $\widehat{m}_{NW}(z;h)$ and $\widehat{m}_{LLLS}(z;h)$ are equally efficient. As we will show in the following sub-section, a relative efficiency improvement can be made by defining an estimator that accounts for the compounded error term assumed in (2.2).

2.2 Local linear weighted least-squares estimator

With the aim of accounting for the variance-covariance assumed in (2.7), under different specifications of the weighting matrix, in Henderson and Ullah (2005) it is proposed a feasible nonparametric random effects estimators for the two estimators developed in Lin and Carroll (2000) and an alternative version for the estimator introduced in Ullah and Roy (1998).

Following the same lines as in the previous sub-section, it is proposed to obtain estimators for $m(z)$ and $D_m(z)$ by minimizing the following criterion function with respect to γ,

$$(Y - Z_z\gamma)^\top W_z(Y - Z_z\gamma), \tag{2.24}$$

where $\widehat{\gamma}$ and Z_z are defined as in (2.16). Let W_z be a weighting matrix based on the kernel function that contains the information of the error structure, in Henderson and Ullah (2005) it is proposed the following local linear weighted least-squares (LLWLS) estimator for $m(z)$,

$$\widehat{\gamma}_0 \doteq \widehat{m}_{LLWLS}(z;h) = e_1^\top (Z_z^\top W_z Z_z)^{-1} Z_z^\top W_z Y. \tag{2.25}$$

The first step of this procedure is to propose a specific form for W_z. Specifically, Lin and Carroll (2000) use two types of weighting matrices, $W_z = K_z^{1/2}\Omega^{-1}K_z^{1/2}$ and $W_z = \Omega^{-1}K_z$, whereas in Ullah and Roy (1998) it is developed an estimation procedure with $W_z = \Omega^{-1/2}K_z\Omega^{-1/2}$. Note that when Ω is a diagonal matrix, these alternative specifications for W_z are the same.

Furthermore, note that (2.25) is an infeasible estimator for $m(\cdot)$ given that Ω depends on some unknown terms, i.e., σ_v^2 and σ_μ^2. Therefore, in order to get a feasible solution for the minimization problem (2.24), an estimator for this covariance matrix is necessary. Following this idea and based on the spectral decomposition of Ω, in Henderson and Ullah (2005) it is developed a local linear feasible weighted least-squares estimator where the unknown covariance components are replaced by their consistent estimators.

Let $\widehat{\epsilon}_{it} = Y_{it} - \widehat{m}_{LLS}(Z_{it};h)$ be the local linear least-squares residual, in Henderson and Ullah (2005) it is proposed to estimate the unknown terms of the variance-covariance matrix (2.7) as

$$\widehat{\sigma}_1^2 = \frac{T}{N}\sum_{i=1}^{N}\widehat{\bar{\epsilon}}_{i\cdot}^2 \qquad \text{and} \qquad \widehat{\sigma}_v^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{\epsilon}_{it} - \widehat{\bar{\epsilon}}_{i\cdot})^2, \tag{2.26}$$

where $\widehat{\bar{\epsilon}}_{i\cdot} = T^{-1}\sum_{t=1}^{T}\widehat{\epsilon}_{it}$. By plugging these consistent estimators into (2.7) it is obtained

$$\widehat{\Omega} = I_N \otimes \left[\widehat{\sigma}_1^2 \imath_T \imath_T^\top/T + \widehat{\sigma}_v^2(I_T - \imath_T\imath_T^\top/T)\right], \tag{2.27}$$

where $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_v^2$ are defined in (2.26). Then, replacing Ω with $\widehat{\Omega}$ in W_z in (2.25), in Henderson and Ullah (2005) it is proposed the following feasible local linear weighted least-squares (FLLWLS) estimator

$$\widehat{m}_{FLLWLS}(z;h) = e_1^\top (Z_z^\top \widehat{W}_z Z_z)^{-1} Z_z^\top \widehat{W}_z Y, \tag{2.28}$$

where $\widehat{W}_z$ is either $K_z^{1/2}\Omega^{-1}K_z^{1/2}$ or $\Omega^{-1}K_z$ and $\widehat{\Omega}$ is th result of plugging (2.26) into (2.7).

In addition, they show that under some standard regularity conditions, for N large and T fixed, the asymptotic bias and variance of $\widehat{m}_{LLWLS}(z;h)$ are bounded by $O_p(h^2)$ and $O_p\left((Nh^q)^{-1}\right)$, respectively, and they hold for $\widehat{m}_{FLLWLS}(z;h)$. See Lin and Carroll (2000) for a detailed analysis of the proof of these results.

Nevertheless, note that Lin and Carroll (2000) and Henderson et al. (2008) demonstrate that these methods of accounting for the correlation could lead to losses of efficiency in comparison to the working independence method proposed in Lin and Carroll (2000). Specifically, these authors argue that higher efficiency is obtained by assuming independence rather than using the correlation structure. The reason is that since $h \to 0$ asymptotically, the chance of having more than two observations from the same subject in the local estimation procedure is small. Hence, the observations locally will came from different subjects, which are assumed to be independent.

2.3 Local linear two-stage least-squares estimator

In order to develop a procedure that enables us to consider the information of the variance-covariance matrix of the error term for the estimators and, at the same time, to improve the efficiency with respect to the local linear least-squares (LLLS) estimators, in Ruckstuhl et al. (2000) it is proposed a two step nonparametric procedure.

More precisely, these authors argue that the slower rate of convergence of the LLLS estimator is due to the elements of the off-diagonal of Ω^{-1}. To solve it, they propose to estimate a nonparametric model that only depends on both the unknown function $m(\cdot)$ and the error term that is *i.i.d.* Thus, unlike what was proposed in Lin and Carroll (2000) and Ullah and Roy (1998), the intuitive idea of the approach developed in Ruckstuhl et al. (2000) is to multiply both sides of (2.1) by the square-root of Ω^{-1} and add and subtract $m(Z)$ obtaining

$$Y^* = m(Z) + \Omega^{-1/2}\epsilon, \tag{2.29}$$

where $Y^* = \Omega^{-1/2}Y + (I - \Omega^{-1/2})m(Z)$ and $Y^* = (Y_{11}^*, \cdots, Y_{NT}^*)^\top$, $m(Z) = (m(Z_{11}), \cdots, m(Z_{NT}))^\top$ and $\epsilon = (\epsilon_{11}, \cdots, \epsilon_{NT})^\top$ are NT-dimensional vectors. Note that $\Omega^{-1/2}\epsilon$ satisfies the independence condition because it exhibits an identity variance-covariance matrix.

In order to provide feasible estimators of (2.29), in the first step Ruckstuhl et al. (2000) propose to obtain the LLLS estimator for the unknown functions of (2.1) and the corresponding residual term that enables us to compute the matrix $\widehat{\Omega}$ as in the previous section. Later, in the second stage, they use this result to compute $\widehat{Y}^* = \widehat{\Omega}^{-1/2}Y + (I - \widehat{\Omega}^{-1/2})\widehat{m}_{LLLS}(Z;h)$ and regress $\widehat{Y}^*$ on Z through the local polynomial regression method. Thus, these authors provide the following local linear two step least-squares (LL2SLS) estimator

$$\widehat{m}_{LL2SLS}(z;h) = e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z \widehat{Y}^*, \tag{2.30}$$

where Z_z and K_z are defined as in (2.16). Later, in Martins-Filho and Yao (2009) it is used this two step procedure to propose a local linear estimator in a regression model where the error term has a non-spherical covariance structure and the regressors are dependent and heterogeneously distributed.

Alternatively, in Wang (2003) it is shown that the efficiency result obtained in Lin and Carroll (2000) is a natural consequence of how standard kernel methods incorporate the within-subject correlation to control the bias, but at the price of ignoring some input from correlated elements within each individual. In order to consider this information and reduce the variance simultaneously, in Wang (2003) it is proposed a two-step procedure that achieves asymptotic improvements over the working independence technique if the covariance is correctly specified.

In order to efficiently use all the correlated data within a subject, the basic idea proposed in Wang (2003) is as follows: once a data point from one subject is near the estimation point and significantly contributes to the local estimation, all data points from this subject will be used. To avoid the bias, the contributions of all these data points except the data point near the local estimation point are through their residuals. Then, for the nonparametric model (2.1) this two step procedure can be described as follows:

Step 1. Compute an initial nonparametric estimator for $m(z)$, say $\widetilde{m}(z;h)$, using for example the working independence method.

Step 2. Obtain the final estimator for $m(z)$, say $\widehat{m}(z;h)$, by solving the following kernel weighted estimating equation

$$\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it}-z)G_i^{\top}\Omega_i^{-1}[Y_i - m_{it}^*(\gamma)] = 0, \tag{2.31}$$

where Ω_i is the variance-covariance matrix (2.7) for the $i-th$ subject while we define G_i to be a $T\times(q+1)$ matrix with the $t-th$ row to be $Z_{z_{it}}^{\top} = [1,(Z_{it}-z)]$ and 0 otherwise. Thus, the $s-th$ element of $m_{it}^*(\gamma)$ is $Z_{z_{it}}^{\top}\gamma$ when $s=t$ and the $s-th$ element of $m_{it}^*(\gamma)$ is $\widetilde{m}(Z_{is};h)$ when $s\neq t$. Note that $\widehat{\gamma}$ is defined as in (2.16).

3 Nonparametric panel data models with fixed effects

In this section, we maintain assumptions (2.3), (2.4) and (2.5) about the data generating process but we replace (2.8) by

$$E\left(\mu_i|\, Z_{i1},\cdots,Z_{iT}\right) = \mu_i. \tag{3.1}$$

This new assumption introduces a relationship of statistical dependence between the heterogeneity term, μ, and the explanatory variables, $Z_1,\cdots,Z_q$. Using now (2.1) and (2.2), applying assumptions (2.3) to (2.5) and (3.1) and by the law of iterated expectations we obtain that

$$E\left(Y_{it}|\, Z_{i1},\cdots,Z_{iT}\right) = m\left(Z_{it}\right) + \mu_i, \quad t=1,\cdots,T. \tag{3.2}$$

Given the specification in (3.2) it is clear that direct estimation of $m(\cdot)$ through standard nonparametric techniques, as in the previous section, would result in estimators with non-negligible asymptotic bias. As in the fully parametric case, several estimation methods have been developed to estimate nonparametric panel data models with fixed effects of the form (3.2); see Hsiao (2003), Wooldridge (2010) and Baltagi (2013), for example. As we appreciate hereinafter, they can be classified into two broad approaches. On the one hand, there is a first type of nonparametric estimators that use differencing transformations to remove the unobserved individual heterogeneity from the structural model. Thus, the unknown function of the transformed model can be estimated consistently through a direct nonparametric approach. On the other hand, a second type of estimators based on the spirit of the least-squares dummy variable (LSDV) approach are proposed to estimate the function of interest, i.e., $m(\cdot)$. In what follows, we review the latest nonparametric literature based on both approaches. Later, we focus on the resulting estimators for different specifications of these nonparametric models, i.e., allowing for additive structures of the unknown smooth function or the presence of time lagged endogenous explanatory variables.

3.1 Profile least-squares estimators

When we want to estimate directly $m(\cdot)$ in (3.2) we need an estimation procedure that takes into account the information contained in the unobserved individual heterogeneity. Following the idea of the least-squares dummy variable approach, a profile least-squares method can be proposed. In this section, we first analyze this method under the different identification conditions considered in Sun et al. (2009), Su and Ullah (2011), Gao and Li (2013), and Lin et al. (2014) and show why it is so important to impose strong identification conditions in this setting. Later, we focus on alternative feasible forms of the local linear approach according to Li et al. (2013).

Let $Y = (Y_{11}, \cdots, Y_{NT})^\top$, $m(Z) = (m(Z_{11}), \cdots, m(Z_{NT}))^\top$, $v = (v_{11}, \cdots, v_{NT})^\top$ vectors of $NT \times 1$ dimension and denote $\mu = (\mu_1, \cdots, \mu_N)^\top$ to be a N-dimensional vector and $D = (I_N \otimes \imath_T)$ a $NT \times N$ dummy matrix. Proceeding as in Section 2.1, we choose an estimator for $m(z)$ that minimizes

$$(Y - m(Z) - D\mu)^\top K_z (Y - m(Z) - D\mu), \tag{3.3}$$

where K_z is a $NT \times NT$ diagonal kernel weighting matrix.

Let z be an interior point of the neighborhood of Z, replacing $m(Z)$ by $m(z)$, the first-order condition with respect to $m(\cdot)$ yields the following local constant kernel estimator for $m(\cdot)$,

$$\widehat{m}(z; h) = (\imath_{NT}^\top K_z \imath_{NT})^{-1} \imath_{NT}^\top K_z (Y - D\mu). \tag{3.4}$$

However, note that μ is not directly observable so this local constant estimator is infeasible. In order to solve it, we can minimize (3.3) with respect to μ obtaining

$$\widehat{\mu} = (D^\top K_z D)^{-1} D^\top K_z (Y - m(Z)). \tag{3.5}$$

Then, if we substitute (3.5) into (3.3) and rearrange terms we obtain the following (concentrated) weighted least squares criterion function

$$(Y - m(Z))^\top W_z (Y - m(Z)), \tag{3.6}$$

where $W_z = M(z)K_z M(z)$ and $M(z) = I_{NT} - D\left(D^\top K_z D\right)^{-1} D^\top K_z$.

Consider again an interior point, z, of the neighborhood of Z. Replacing $m(Z)$ by $m(z)$, the first-order condition with respect to $m(\cdot)$ in (3.6) yields the following local constant kernel estimator

$$\widehat{m}_{LCLS}(z;h) = (\imath_{NT}^\top W_z \imath_{NT})^{-1} \imath_{NT}^\top W_z Y, \tag{3.7}$$

It is important to highlight that the weighting matrix W_z has been designed to directly remove any time invariant term in the structural model (3.2). To see this, note that $M(z)D = 0$. However, since $\imath_{NT} m(z)$ is time invariant and $W_z \imath_{NT} \equiv 0$, the matrix $\imath_{NT}^\top W_z \imath_{NT}$ is non-invertible so the resulting estimator of this method is infeasible. See Lin et al. (2014) for a detailed description of this problem.

To overcome this situation, it is necessary to use a weighting matrix which removes the unobserved cross-sectional heterogeneity either complete or asymptotically and, at the same time, enables us to select only those values of Z_{it} close to z. In other words, we need a weighting matrix that enables $\imath_{NT}^\top W_z \imath_{NT}$ to be invertible and $(\imath_{NT}^\top W_z \imath_{NT})^{-1} \imath_{NT}^\top W_z D\mu$ asymptotically negligible. If we assume that μ_i is an *i.i.d.* random variable with zero mean and finite variance, in Lin et al. (2014) it is suggested to asymptotically remove the individual effects via the proposal of a new weighting matrix W_{z_0} that satisfies $(\imath_{NT}^\top W_{z_0} \imath_{NT})^{-1} \imath_{NT}^\top W_{z_0} D\mu = N^{-1} \sum_{i=1}^N \mu_i = O_p(N^{-1/2})$, as $N \to \infty$.

Let $\mu_0 = (\mu_2, \cdots, \mu_N)^\top$ be a $(N-1)$-dimensional vector and denote D_0 a $NT \times (N-1)$ matrix of the form $D_0 = (-\imath_{N-1}, I_{N-1})^\top \otimes \imath_T$, Su and Ullah (2006b), Sun et al. (2009) and Lin et al. (2014) propose to replace D with D_0 in (3.3). Then, considering a local linear instead of a local constant approximation, the local linear weighted least squares criterion becomes

$$(Y - Z_z\gamma - D_0\mu_0)^\top K_z (Y - Z_z\gamma - D_0\mu_0), \tag{3.8}$$

where Z_z and γ are defined as in (2.16).

As in (3.3), the quantities of interest can be estimated by minimizing the following criterion function

$$(Y - Z_z\gamma)^\top W_{z_0} (Y - Z_z\gamma), \tag{3.9}$$

where now the weighting matrix is $W_{z_0} = M_0(z)^\top K_z M_0(z)$, with $M_0(z) = I_{NT} - D_0(D_0^\top K_z D_0)^{-1} D_0^\top K_z$, in such a way that $M_0(z)D_0 = 0$. Let $\widehat{\gamma} = (\widehat{\gamma}_0^\top \quad \widehat{\gamma}_1^\top)^\top$ be a $(1+q)$ vector of minimizers of (3.9. Then, the profile local linear weighted least squares estimator is

$$\widehat{\gamma}_{PLLLS} = (Z_z^\top W_{z_0} Z_z)^{-1} Z_z^\top W_{z_0} Y. \tag{3.10}$$

Note that although this new weighting matrix enables us to obtain feasible estimators, the resulting estimator has an extra component in the bias term that comes from the existence of unobserved

cross-sectional heterogeneity. See Theorem 2.1 in Lin et al. (2014) for the profile local constant estimator.

In this framework, a standard solution is the estimation of the nonparametric regression under further strong identification conditions regarded to the individual effects. Specifically, Mammen et al. (2009) and Su and Ullah (2011) impose $\sum_{i=1}^{N} \mu_i = 0$, while Gao and Li (2013) develop a profile least-squares method under the condition $E(\mu_i) = 0$. As it is proved in Sun et al. (2009) for partially linear models, this stronger identification condition allows us to obtain standard asymptotic properties in the nonparametric framework and, simultaneously, override the individual effects.

Alternatively, in Li et al. (2013) it is proposed a profile method in which it is not necessary to pay special attention to the invertibility problem noted in Lin et al. (2014) and Gao and Li (2013). Again, assuming $\sum_{i=1}^{N} \mu_i = 0$, a profile local linear least-squares method for the nonparametric components of the regression function is proposed. But, unlike the previous methods, in Li et al. (2013) it is assumed that μ_0 is known and a least-squares procedure for the nonparametric components in β can be proposed. In this way, the quantities of interest can be estimated by minimizing the resulting criterion function of a local linear fitting with respect to γ obtaining

$$\widehat{\gamma} = (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z (Y - D_0 \mu_0). \tag{3.11}$$

As previously, this estimator is not feasible but we can multiply the true model by $e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z$ and choose the μ_0 that minimizes the following criterion function

$$\left(Y^* - D_0^* \mu_0\right)^\top \left(Y^* - D_0^* \mu_0\right), \tag{3.12}$$

where $Y^* = \left(I_{NT} - e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z\right) Y$ and $D_0^* = \left(I_{NT} - e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z\right) D_0$. Thus, the minimizer of (3.12) is of the form

$$\widehat{\mu}_{PLLLS} = (D_0^{*\top} D_0^*)^{-1} D_0^{*\top} Y^*, \tag{3.13}$$

and replacing μ_0 with $\widehat{\mu}_{PLLLS}$ in (3.11) the profile local weighted linear least-squares estimator is

$$\widehat{\gamma}_{PLLLS} = (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z (Y - D_0 \widehat{\mu}_{PLLLS}). \tag{3.14}$$

Finally, analyzing the asymptotic normality of (3.14), in Li et al. (2013) it is shown that under standard smoothing conditions, as $N \to \infty$ and $q = 1$,

$$\sqrt{Nh}\left(\widehat{m}_{PLLLS}(z;h) - m(z) - b(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{\overline{\sigma}^2(z) R(K)}{f(z)^2}\right), \tag{3.15}$$

where $m''(z)$ is the second-order derivative of $m(\cdot)$ with respect to z, $f_t(\cdot)$ is the density function of Z_{it}, $f(z) = \sum_{t=1}^{T} f_z(z)$ and

$$\begin{aligned} b(z) &= \frac{h^2}{2} \mu_2(K) m''(z), \\ \overline{\sigma}^2(z) &= \sum_{t=1}^{T} \sigma_t^2(z) f_t(z) = \sum_{t=1}^{T} \mathrm{Var}\left(v_{it} | Z_{it} = z\right) f_t(z). \end{aligned}$$

Note that, as we will detail in later sections, these asymptotic properties are similar to the result obtained in Su and Ullah (2006b) for partially linear panel data models with fixed effects. In fact, although this profiling technique enables to obtain estimators with a negligible asymptotic bias, in most cases additional assumptions such as $\sum_{i=1}^{N} \mu_i = 0$ or $E(\mu_i) = 0$ are needed. Furthermore, in most part of the estimators quoted previously the asymptotic analysis is performed when both N and T tends to infinity. Thus, with the purpose of using standard assumptions as the ones used in the previous sections of this paper and standard asymptotic theory, that is, letting N tends to infinity and fixed T, in the following sub-section differencing transformations are studied.

3.2 Differencing estimators

As it has been pointed out previously, to overcome the main difficulties of the profile techniques and to remove the individual effects from the regression model, differencing transformations are proposed. In this section, we first review the resulting estimators from the first-derivative function proposed in Mundra (2005) and Lee and Mukherjee (2008). Later, the iterative nonparametric kernel estimator based on a profile likelihood approach in Henderson et al. (2008) is analyzed.

As in the fully parametric case, there are several transformations from the model of interest which enable us to remove the heterogeneity of unknown form that does not vary in time. Among the most popular we consider the first differences and the differences from the mean. First differences transformation can be understood as the subtraction from time t of (3.2) that of time $t-1$, i.e.

$$Y_{it} - Y_{i(t-1)} = m(Z_{it}) - m(Z_{i(t-1)}) + v_{it} - v_{i(t-1)}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T \tag{3.16}$$

or that of time 1, i.e.,

$$Y_{it} - Y_{i1} = m(Z_{it}) - m(Z_{i1}) + v_{it} - v_{i1}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \tag{3.17}$$

On the other hand, differences from the mean implies subtracting from time t the within-group mean, i.e.,

$$Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is} = m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T} m(Z_{is}) + v_{it} - \frac{1}{T}\sum_{s=1}^{T} v_{is}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{3.18}$$

As the reader can realize, the right-hand side of either (3.16), (3.17) or (3.18) are linear combinations of $m(Z_{it})$ for different periods t. As noted in Su and Ullah (2011), this makes the estimation of $m(\cdot)$ rather cumbersome because $m(\cdot)$ takes the form of an additive function whose elements share the same functional form.

Assuming $m(\cdot)$ admits some number of derivatives and all assumptions introduced in Section 3 are fulfilled, in Ullah and Roy (1998) it is proposed to use either first differences or mean deviation transformations and then, take a linear approximation of the unknown function $m(\cdot)$ around z. By doing so, they expected that the resulting first-difference and fixed-effects estimators for the

marginal effects of $m(\cdot)$ (i.e., the partial derivatives of $m(z)$ with respect to z) satisfy the standard properties of the local linear regression approach. However, as it is proved in Lee and Mukherjee (2008), this statement is not true because this technique provides estimators with a non-negligible asymptotic bias.

For the sake of simplicity, let us consider the univariate problem ($q = 1$) of the first differences regression model (3.16). By approximating $m(\cdot)$ via a Taylor expansion we obtain

$$\Delta Y_{it} = \Delta Z_{it} m'(z) + \Delta v_{it}(z), \tag{3.19}$$

where $m'(z) = \partial m(z)/\partial z$. Let $m''(z) = \partial^2 m(z)/\partial z^2$, the error of this transformed regression is

$$\Delta v_{it}(z) = \Delta v_{it} + \frac{1}{2} m''(\xi) \left((Z_{it} - z)^2 - (Z_{i(t-1)} - z)^2 \right),$$

for some $\xi \in I\!R$ between Z_{it} and z.

On the contrary, the transformed regression of the mean deviation (i.e., within-group) expression is

$$Y_{it} - \frac{1}{T}\sum_{s=1}^{T} Y_{is} = \left(Z_{it} - \frac{1}{T}\sum_{s=1}^{T} Z_{is} \right) m'(z) + \left(v_{it}(z) - \frac{1}{T}\sum_{s=1}^{T} v_{is}(z) \right), \tag{3.20}$$

where the corresponding error term is

$$v_{it}(z) - \frac{1}{T}\sum_{s=1}^{T} v_{is}(z) = \left(v_{it} - \frac{1}{T}\sum_{s=1}^{T} v_{is} \right) + \frac{1}{2} m''(\xi) \left((Z_{it} - z)^2 - \frac{1}{T}\sum_{s=1}^{T} (Z_{is} - z)^2 \right).$$

For the transformed regression models (3.19) and (3.20), Lee and Mukherjee (2008) propose the following local linear estimators for the first-order derivatives,

$$\widehat{m}'_D(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it} - z)\Delta Z_{it}\Delta Y_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it} - z)\Delta Z_{it}^2} \tag{3.21}$$

and

$$\widehat{m}'_W(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it} - z)\ddot{Z}_{it}\ddot{Y}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it} - z)\ddot{Z}_{it}^2}, \tag{3.22}$$

where we denote $\ddot{Y}_{it} = Y_{it} - (T-1)^{-1}\sum_{s=1, s\neq t}^{T} Y_{is}$. This definition is similar for $\ddot{Z}_{it}$ and $\ddot{v}_{it}$. For the sake of simplicity, when analyzing the asymptotic properties of these estimators Lee and Mukherjee (2008) use the leave-one-out average in (3.20) instead of the within-group mean.

Let $\mathbb{Z} = (Z_{11}, \cdots, Z_{NT})$ be the vector of observed values of the explanatory variable, under the conditions established at the beginning of Section 3, these authors obtain the following conditional bias for these two local linear estimators

$$E[\widehat{m}'_D(z;h) - m(z)|\mathbb{Z}] = \frac{m''(z)\mu_3(z)}{2\mu_2(z)} + O_p(h^2) \tag{3.23}$$

and

$$E[\widehat{m}'_W(z;h) - m(z)|\mathbb{Z}] = \frac{m''(z)[\mu_1(z)\mu_2(z) + \mu_3(z)]}{2[\mu_1^2(z) + \mu_2(z)]} + O_p(h^2), \tag{3.24}$$

where $\mu_j(z) = E(Z_{it} - z)^j < \infty$, for $j = 1, 2, 3$.

Analyzing these results, we can highlight that these two local linear estimators exhibit a non-negligible asymptotic bias. More precisely, as it can be seen in (3.21) and (3.22), the non-degenerated bias is due to the fact that the transformed regression equations are localized around Z_{it}, without taking into account all other values. Consequently, the distance between Z_{is} and z cannot be controlled by a fixed bandwidth parameter h so the residual terms of the Taylor approximation do not vanish. Therefore, it is not possible to assume that $\Delta v_{it}(z)$ and Δv_{it} are close enough so we can conclude that the local linear regression approach provides inconsistent estimators by the correlation between the transformed error terms $\Delta v_{it}(z)$ and the transformed regressors ΔZ_{it}. The same can be said for $\ddot{v}_{it}(z)$ and $\ddot{Z}_{it}$.

To the best of our knowledge, there are two different strategies to overcome this problem. On the one hand, Lee and Mukherjee (2008) propose the estimation of a local within transformation that uses a locally weighted average to remove the fixed effects. On the other hand, Mundra (2005) develops a direct procedure based on the use of higher-dimensional kernel weights. In the following, we detail the main particularities of both techniques.

In order to remove the unobserved individual heterogeneity and, at the same time, propose estimators that take into account all the values of the regressors involved in the estimation, in Lee and Mukherjee (2008) it is proposed to follow a differencing strategy that uses the locally weighted average of Z_{it}, for a given z, to remove the fixed effects. Let

$$\widetilde{Z}_{i\cdot} = \sum_{s=1, s\neq t}^{T} W_{is}(z) Z_{is}, \tag{3.25}$$

where $W_{is}(z)$ is a kernel weight of the form

$$W_{is}(z) = \frac{K_h(Z_{is} - z)}{\sum_{r=1, r\neq t}^{T} K_h(Z_{ir} - z)}. \tag{3.26}$$

We define $\widetilde{Y}_{i\cdot}(z)$ and $\widetilde{v}_{i\cdot}(z)$ in a similar way as the locally weighted averages of Y_{it} and $v_{it}(z)$, respectively.

Note that $\sum_{s=1, s\neq t}^{T} W_{is}(z)\mu_i = \mu_i$ because it holds $W_{is}(z) \geq 0$ and $\sum_{s=1, s\neq t}^{T} W_{is}(z) = 1$ for any z. Then, if we subtract such local averages from (3.2) and denote $Y_{is}^* = Y_{is} - (T-1)^{-1}\sum_{s=1, s\neq t}^{T} Y_{is}$ and $Z_{is}^* = Z_{is} - (T-1)^{-1}\sum_{s=1, s\neq t}^{T} Z_{is}$, the functions of interest can be estimated as the values of β that minimize the following criterion function,

$$\sum_{i=1}^{N} \sum_{s=1, s\neq t}^{T} \left(Y_{is}^* - Z_{is}^*\gamma\right)^2 K_h(Z_{is} - z), \tag{3.27}$$

since it is assumed that $W_{is}(z) \geq 0$ and $\sum_{s=1,s\neq t}^{T} W_{is}(z) = 1$, for any z, so $\sum_{s=1,s\neq t}^{T} W_{is}(z)\mu_i = \mu_i$.

Let us denote by $\widehat{\gamma}$ the value of γ that minimizes (3.27). Proceeding as in other previous local constant regression problems,

$$\widehat{\gamma} \doteq \widehat{m}'_{LWA}(z;h) = \frac{\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T} K_h(Z_{is}-z)Z_{is}^{*}Y_{is}^{*}}{\sum_{i=1}^{N}\sum_{s=1,s\neq t}^{T} K_h(Z_{is}-z)Z_{is}^{*2}}. \tag{3.28}$$

Under the same conditions as in (3.23) and (3.24), in Lee and Mukherjee (2008) it is shown that this local weighted linear estimator $\widehat{m}'_{LWA}(z;h)$ has the following conditional bias and variance

$$E[\widehat{m}'_{LWA}(z;h) - m(z)|\mathbb{Z}] = \frac{h^2}{2}\left(\frac{m'(z)f'(z)}{f(z)}\right)\left(\frac{\kappa_4-\kappa_2^2}{\kappa_2}\right) + o_p(h^2) \tag{3.29}$$

and

$$Var(\widehat{m}'_{LWA}(z;h) - m(z)|\mathbb{Z}) = \frac{1}{NTh^3}\left(\frac{\sigma^2}{f(z)}\right)\left(\frac{\varphi_2}{\kappa_2^2}\right) + o_p\left(\frac{1}{NTh^3}\right), \tag{3.30}$$

where $\kappa_j = \int z^j K(z)dz$, for $j = 2, 4$, and $\varphi_2 = \int z^2 K^2(z)dz$.

Looking at these results, we can point out that both conditional bias and variance terms tend to zero as $h \to 0$ and $NTh^3 \to \infty$, when both N and T tends to infinity. Therefore, $\widehat{m}'_{LWA}(z;h)$ is a consistent estimator for $m'(z)$. However, note that the variance term is of order $O_p\left(1/NTh^3\right)$. This makes the rate of convergence of this estimator rather slow with respect of the standard rate of convergence of these family of estimators that would be of order $O_p\left(1/NTh\right)$.

Another way to overcome this problem is to use a higher-dimensional kernel weight. As it is suggested in Mundra (2005), the bias associated to (3.21) can be removed by considering a local approximation around the pair $(Z_{it}, Z_{i(t-1)})$ obtaining the following first-difference local linear estimator

$$\widehat{m}_{FLL}(z;h) = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it}-z)K_h(Z_{i(t-1)})\Delta Z_{it}\Delta Y_{it}}{\sum_{i=1}^{N}\sum_{t=2}^{T} K_h(Z_{it}-z)K_h(Z_{i(t-1)}-z)\Delta Z_{it}^2}. \tag{3.31}$$

As the reader can appreciate, these procedures are very appealing as they provide estimators for local marginal effects in a framework of differencing models. However, they are unable to identify the function $m(\cdot)$. In this context, in Henderson et al. (2008), to estimate $m(\cdot)$, a nonparametric kernel estimator is proposed based on an iterative profile likelihood approach. More precisely, in their paper they propose the following differencing transformation in (3.2)

$$Y_{it} - Y_{i1} = m(Z_{it}) - m(Z_{i1}) + v_{it} - v_{i1}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \tag{3.32}$$

In order to estimate $m(z)$, following Wang (2003) and Lin and Carroll (2006), in Henderson et al. (2008) it is proposed a profile likelihood approach. In fact, the likelihood function for the i-th individual is defined as

$$\mathcal{L}_i(\cdot) = \mathcal{L}(Y_i, m_i) = -\frac{1}{2}(\ddot{Y}_i - m_i + m_{i1}\imath_{(T-1)})^{\top}\Sigma^{-1}(\ddot{Y}_i - m_i + m_{i1}\imath_{(T-1)}), \tag{3.33}$$

where $\ddot{Y}_i = (\ddot{Y}_{i2}, \cdots, \ddot{Y}_{iT})$, for $\ddot{Y}_{it} = Y_{it} - Y_{i1}$, $m_i = (m(Z_{i2}), \cdots, m(Z_{iT}))^\top$, $m_{i1} = m(Z_{i1})$ and $\imath_{(T-1)}$ is a $(T-1) \times 1$ unit vector. Let $\ddot{v}_i = (\ddot{v}_{i2}, \cdots, \ddot{v}_{iT})^\top$ be a $(T-1)$-dimensional vector. Note that, under (2.5) the variance-covariance matrix of $\ddot{v}_{it} = v_{it} - v_{i1}$, $\Sigma = Cov(\ddot{v}_i | Z_{i1}, \cdots, Z_{iT})$, is equal to $Cov(\ddot{v}_i)$, in such a way that

$$\Sigma = \sigma_v^2 (I_{T-1} + \imath_{T-1}\imath_{T-1}^\top) \qquad \text{and} \qquad \Sigma^{-1} = \sigma_v^{-2}(I_{T-1} - \imath_{T-1}\imath_{T-1}^\top/T). \tag{3.34}$$

Defining $\mathcal{L}_{i,tm} = \partial \mathcal{L}_i(\cdot)/\partial m_{it}$, where $m_{it} = m(Z_{it})$, the unknown function $m(z)$ can be estimated by solving the following first-order condition

$$0 = \sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it} - z) G_{it}(z,h) \mathcal{L}_{i,tm}\left(Y_i, \widehat{m}(Z_{i1}), \cdots, \widehat{m}(z) + \{(Z_{it} - z)/h\}\widehat{m}'(z), \cdots, \widehat{m}(Z_{iT})\right),$$

where $G_{it}(z;h) = \begin{bmatrix} 1 & ((Z_{it} - z)/h)^\top \end{bmatrix}^\top$.

Based on this structure, Henderson et al. (2008) develop an iterative procedure. Specifically, let $\widehat{m}_{[\ell-1]}(z)$ be the current estimator at the $[\ell-1]th$ step, in the next step they propose to estimate $\widehat{m}_{[\ell]}(z) = \widehat{\alpha}_0(z)$, by choosing $(\widehat{\alpha}_0 \quad \widehat{\alpha}_1)$ as the minimizers of

$$\begin{aligned} &\sum_{i=1}^{N}\sum_{t=1}^{T} K_h(Z_{it} - z) G_{it}(z;h) \mathcal{L}_{i,tm}\left(Y_i, \widehat{m}_{[\ell-1]}(Z_{i1}), \cdots, \widehat{\alpha}_0 + ((Z_{it} - z)/h)^\top \widehat{\alpha}_1, \cdots, \widehat{m}_{[\ell-1]}(Z_{iT})\right) \\ &= 0. \end{aligned} \tag{3.35}$$

This procedure will iterate till convergence. Note that as it is pointed out in Henderson and Parmeter (2015) the actual derivative of $m(z)$ for a particular explanatory variable requires you to divide $\widehat{\alpha}_1(z)$ by the bandwidth for that particular explanatory variable. Under the assumption that $h_r \sim N^{-1/(4+q)}$, for $r = 1, \cdots, q$, and by defining $\kappa = \int K^2(u)du$ and $\kappa_2 = \int u^2 K(u)du$, as $N \to \infty$ in such way that $h_r \to 0$ and $Nh_1 \cdots h_q \to \infty$, they obtain

$$\sqrt{Nh_1 \cdots h_q}\left(\widehat{m}(z) - m(z) - \sum_{r=1}^{q} h_r^2 b_r(z)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \kappa^q/\Omega(z)\right), \tag{3.36}$$

where $\Omega(z) = -\sum_{t=1}^{T} f_t(z) E\left(\frac{\partial^2 \mathcal{L}_i(\cdot)}{\partial m_{it}^2}\middle| Z_{it} = z\right)$ and $b_r(z)$ is a bounded and continuous function that is the solution to

$$b_r(z) = \frac{\kappa_2}{2} m_{rr}(z) + \frac{1}{N\sigma_v^2}\sum_{t=1}^{T}\sum_{s\neq t}^{T} d_{ts} f_t(z) E\left(\frac{\partial^2 \mathcal{L}_i(\cdot)}{\partial m_{it}\partial m_{is}} b_r(Z_{is})\middle| Z_{it} = z\right)/\Omega(z)$$

where $d_{ts} = 1$ if $t = 1$ or $s = 1$, and $d_{ts} = -1$ otherwise.

Similarly to other nonparametric estimators developed for differencing models, this iterative estimator has the advantage of completely removing the unobserved individual heterogeneity. However, it is true that the estimator does not achieve the optimal rate of convergence for this type of nonparametric estimators. Alternatively, other authors propose consistent estimators for the $m(\cdot)$ function

in this context of differencing models. On the one hand, Baltagi and Li (2002) use the series approximation to estimate the nonparametric component. On the other hand, Qian and Wang (2012) propose a two step procedure for a partially linear model with fixed effects. As we will detail in Section 5, in the first step the fully nonparametric component, i.e., $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$, is estimated using a multivariate nonparametric estimator. Later, in the second stage they turn to the marginal integration technique proposed originally in Linton and Nielsen (1995). Note that the marginal integration method presents some awkward features such as its high computational cost. Specifically, to obtain an estimator of this type we must compute $O(NT^3h^{1/2})$ operations, i.e., if we have to make NT^2 regressions each one requires $O(NTh^{1/2})$ operations. Therefore, other estimation techniques may be preferred.

3.3 Nonparametric additive panel data models

As it has been already pointed out previously, nonparametric smoothing regression techniques have been intensively used in the last few decades since they enable us to consider some hidden features of the data that cannot be captured by a predetermined parametric model. However, they exhibit an important drawback: *The curse of dimensionality.* That is, the rate of convergence of nonparametric estimators slows down as the number of explanatory variables enlarges. Nevertheless, there are situations where the researcher needs to handle a large number of these variables. In these cases, it is recommended to estimate $m(\cdot)$ in (2.1) by imposing the additional additivity restriction

$$m\left(Z_{it}\right) = \sum_{j=1}^{q} m_j\left(Z_{jit}\right). \tag{3.37}$$

In the random effects setting (see assumptions (2.1), (2.2) and (2.3) to (2.8)), direct application of standard backfitting (see Hastie and Tibshirani (1990)) or marginal integration techniques (see Linton and Nielsen (1995)) provides consistent estimators for $m(\cdot)$ and the additive functions. However, if we introduce the fixed effects setting, (i.e., replacing (2.8) by (3.1)) the estimator that results from applying these techniques does not exhibit the same desirable statistical properties as in the random effects case. In Mammen et al. (2009) it is considered the estimation of a nonparametric additive panel data model under different forms of the unobserved heterogeneity and for two asymptotic frameworks: $N \to \infty$ and T fixed and both $N, T \to \infty$. In order to do so, they rely heavily on the smoothed backfitting approach introduced in Mamen et al. (1999).

The nonparametric model that they propose to estimate presents mainly three differences with respect to the fixed effects panel data model introduced at the beginning of this Section. First, the additivity restriction (3.37) is introduced in the model. Second, temporary effects are also considered. Third, among the explanatory variables there can be included time lagged values of Y_{it}. Then, the model to estimate is

$$Y_{it} = \sum_{j=1}^{q} m_j(Z_{jit}) + \eta_t + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{3.38}$$

Note that we denote by η_t the temporal effects. Just to understand the estimation technique we will introduce some more details in the characterization of the explanatory variables Z_{it}. In Mammen et al. (2009) it is assumed that Z_{it} has a density on $[a,b]=[a_1,b_1]\times\cdots\times[a_q,b_q]$. The conditional density of Z_{it} given that Z_{it} lies in $[a,b]$ is denoted by f_t . For simplicity of notation, $a_1=\cdots=a_q=0$ and $b_1=\cdots=b_q=1$. Thus $[a,b]=[0,1]^q$. Estimation of $m_1(\cdot),\cdots,m_q(\cdot)$ will be considered on $[0,1]$. We denote by n the number of explanatory variables $Z_{it}\in[0,1]^q$ for $i=1,\cdots,N$, $t=1,\cdots,T$. The numbers n_t and n_i are respectively the number of $Z_{it}\in[0,1]^q$ for fixed t and the number $Z_{it}\in[0,1]^q$ for fixed i. The one- and two-dimensional marginals of f^t are denoted by $f_j^t(Z_j)$ or $f_{j,k}^t(Z_j,Z_k)$, respectively. The one- and two-dimensional marginals of f^i are denoted by $f_j^i(Z_j)$ or $f_{j,k}^i(Z_j,Z_k)$, respectively. Also, we put $f(Z)=\sum_{t=1}^T(d_t/d)f^t(Z)$, $f_j(Z_j)=\sum_{t=1}^T(d_t/d)f_j^t(Z_j)$, and $f_{j,k}(Z_j,Z_k)=\sum_{t=1}^T(d_t/d)f_{j,k}^t(Z_j,Z_k)$. Finally, let $d_t=P(Z_{it}\in[0,1]^q)$ and $d_T=\sum_{t=1}^T d_t$.

The local constant smooth backfitting estimator $\widehat{m}_1,\cdots,\widehat{m}_q$ for the functions $m_1,\cdots,m_q$ proposed in Mammen et al. (2009) is based on kernel smoothing. Using the following modification of a convolution kernel,

$$K_h(u,v)=\begin{cases}\frac{K[h^{-1}(u-v)]}{\int_0^1 K[h^{-1}(\omega-v)]d\omega} & \text{if} \quad u,v\in[0,1]\\ 0 & \text{else}\end{cases}$$

the estimators for $m_1,\cdots,m_q$, $\mu_1,\cdots,\mu_N$ and $\eta_1,\cdots,\eta_T$ are defined as minimizers of a smoothed least squares criterion,

$$\sum_{i=1}^N\sum_{t=1}^T\int\left(Y_{it}-\sum_{j=1}^q\widehat{m}_j(u_j)-\widehat{\eta}_t-\widehat{\mu}_i\right)^2K_{h_1}(u_1,Z_{1it})\cdots K_{h_q}(u_q,Z_{qit})du_1\cdots du_q, \tag{3.39}$$

under the following constraints

$$\int\widehat{m}_j(Z_j)\widehat{f}_j(Z_j)dZ_j=0, \tag{3.40}$$

$$\sum_{i=1}^N(n_i/n)\widehat{\mu}_i=0. \tag{3.41}$$

The functions $\widehat{f}_j(\cdot)$ are the following kernel density estimators based on the explanatory variables

$$\widehat{f}_j(Z_j)=n^{-1}\sum_{i=1}^N\sum_{t=1}^T\mathbf{1}(Z_{it}\in[0,1]^q)K_{h_j}(Z_j-Z_{jit}). \tag{3.42}$$

Based on (3.39) note that the estimators only use the values of the Y variable in the smoothing if the corresponding values of the explanatory variables lie in $[0,1]^q$. By using derivatives of the

criterion function (3.39) one gets that the minimizers (estimators) must fulfill

$$\begin{aligned}
\widehat{m}_j(Z_j) &= \widetilde{m}_j(Z_j) - \sum_{t=1}^{T} \frac{n_t}{n} \widehat{\eta}_t \frac{\widehat{f}_j^t(Z_j)}{\widehat{f}_j(Z_j)} - \\
&\quad \sum_{i=1}^{N} \frac{n_i}{n} \widehat{\mu}_i \frac{\widehat{f}_j^i(Z_j)}{\widehat{f}_j(Z_j)} - \sum_{\ell \neq j} \int \widehat{m}_\ell(Z_\ell) \frac{\widehat{f}_{j\ell}(Z_j, Z_\ell)}{\widehat{f}_j(Z_j)} dZ_\ell, \qquad (3.43)\\
\widehat{\eta}_t &= \widetilde{\eta}_t - \sum_{j=1}^{q} \int \widehat{m}_j(Z_j) \widehat{f}_j^t(Z_j) dZ_j, \quad t = 1, \cdots, T, \qquad (3.44)\\
\widehat{\mu}_i &= \widetilde{\mu}_i - \sum_{j=1}^{q} \int \widehat{m}_j(Z_j) \widehat{f}_j^i(Z_j) dZ_j, \quad i = 1, \cdots, N, \qquad (3.45)
\end{aligned}$$

where $\widetilde{m}_j, \widetilde{\eta}_j, \widetilde{\mu}_i$ are the following marginal estimators

$$\begin{aligned}
\widetilde{m}_j(Z_j) &= \frac{1}{n} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h_j}(Z_j, Z_{jit}) Y_{it} / \widehat{f}_j(Z_j), \qquad (3.46)\\
\widetilde{\eta}_t &= \frac{1}{n_t} \sum_{i=1}^{N} \mathbf{1}(Z_{it} \in [0,1]^q) Y_{it}, \qquad (3.47)\\
\widetilde{\mu}_i &= \frac{1}{n_i} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) Y_{it}, \qquad (3.48)
\end{aligned}$$

and the functions $\widehat{f}_{jk}, \widehat{f}_j^t, \widehat{f}_j^i$ are the estimators for the kernel density of the form

$$\begin{aligned}
\widehat{f}_{jk}(Z_j, Z_k) &= \frac{1}{n} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h_j}(Z_{jit} - Z_j) K_{h_k}(Z_{kit} - Z_k), \qquad (3.49)\\
\widehat{f}_j^t(Z_j) &= \frac{1}{n_t} \sum_{i=1}^{N} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h_j}(Z_{jit} - Z_j), \qquad (3.50)\\
\widehat{f}_j^i(Z_j) &= \frac{1}{n_i} \sum_{t=1}^{T} \mathbf{1}(Z_{it} \in [0,1]^q) K_{h_j}(Z_{jit} - Z_j). \qquad (3.51)
\end{aligned}$$

For the sake of clarity, we reproduce here the algorithm for calculating the local constant smoother given in Mammen et al. (2009), p. 448. In fact, our aim here is to estimate the function $m_j(\cdot)$ at some given points Z_j^0. Equations (3.43), (3.44) and (3.45) suggest an iterative calculation of estimators. Application of (3.43) for $j = 1, \cdots, q$ can be used to update $\widehat{m}_j$. In each application one plugs the current values of $\widehat{m}_k$, $j \neq k$, and of both $\widehat{\mu}_i$ and $\widehat{\eta}_t$ into the right-hand side of (3.43). Afterwards, one uses (3.44) and (3.45) for updates of $\widehat{\mu}_i$ and $\widehat{\eta}_t$. Again this is done by using the actual values of $\widehat{m}_j$ on the right-hand side of the equation. Let us call these iterative values $\widehat{m}_j^{[a]}$, $\widehat{\mu}_i^{[a]}$ and $\widehat{\eta}_t^{[a]}$ where a is the number of cycles of the algorithm that has been applied. The starting values are denoted by $\widehat{m}_j^{[0]}$, $\widehat{\mu}_i^{[0]}$ and $\widehat{\eta}_t^{[0]}$. The steps of the algorithm are:

Step 1. Set $a = 0$ and calculate the smoothing weights around Z_j^0.

Step 2. With these weights compute $\widehat{f}_j\left(Z_j^0\right)$, $\widehat{f}_{jk}\left(Z_j^0, Z_k^0\right)$, for $k \neq j$, $\widehat{f}_j^i\left(Z_j^0\right)$, $\widehat{f}_j^t\left(Z_j^0\right)$, $\widetilde{m}_j\left(Z_j^0\right)$, $\widetilde{\mu}_i$, $\widetilde{\eta}_t$. Then set $\widehat{m}_j^{[a]}\left(Z_j^0\right) = \widetilde{m}_j\left(Z_j^0\right)$, $\widehat{\mu}_i^{[a]} = \widetilde{\mu}_i$ and $\widehat{\eta}_t^{[a]} = \widetilde{\eta}_t$.

Step 3. For $j = 1, \cdots, q$

$$\begin{aligned}\widehat{m}_j^{[a+1]}\left(Z_j^0\right) &= \widetilde{m}_j\left(Z_j^0\right) - \sum_{t=1}^{T} \frac{n_t}{n} \widehat{\eta}_t^{[a]} \frac{\widehat{f}_j^t(Z_j^0)}{\widehat{f}_j(Z_j^0)} - \sum_{i=1}^{N} \frac{n_i}{n} \widehat{\mu}_i^{[a]} \frac{\widehat{f}_j^i(Z_j^0)}{\widehat{f}_j(Z_j^0)} - \\ & \sum_{\ell<j} \int \widehat{m}_\ell^{[a+1]}(Z_\ell) \frac{\widehat{f}_{j\ell}(Z_j^0, Z_\ell)}{\widehat{f}_j(Z_j^0)} dZ_\ell - \\ & \sum_{\ell>j} \int \widehat{m}_\ell^{[a]}(Z_\ell) \frac{\widehat{f}_{j\ell}(Z_j^0, Z_\ell)}{\widehat{f}_j(Z_j^0)} dZ_\ell. \end{aligned} \tag{3.52}$$

Step 4. For $t = 1, \cdots, T$,

$$\widehat{\eta}_t^{[a+1]} = \widetilde{\eta}_t - \sum_{j=1}^{q} \int \widehat{m}_j^{[a+1]}(Z_j) \widehat{f}_j^t(Z_j) dZ_j. \tag{3.53}$$

Step 5. For $i = 1, \cdots, N$,

$$\widehat{\mu}_i^{[a+1]} = \widetilde{\mu}_i - \sum_{j=1}^{q} \int \widehat{m}_j^{[a+1]}(Z_j) \widehat{f}_j^i(Z_j) dZ_j. \tag{3.54}$$

Step 5 If a predetermined convergence criterion is fulfilled, stop. Otherwise set a to $a+1$ and go to **Step 3**.

The local constant estimator, $\widehat{m}_j$, exhibits a complicated bias. In order to avoid this problem, in Mammen et al. (2009) it is proposed to use local linear estimators. Then, intercepts $\widehat{m}_1, \cdots, \widehat{m}_q$, slopes $\widehat{m}^1, \cdots, \widehat{m}^p$, and both individual and time effects, $\widehat{\mu}_1, \cdots, \widehat{\mu}_N$ and $\widehat{\eta}_1, \cdots, \widehat{\eta}_T$, are defined as minimizers of

$$\begin{aligned}\sum_{i=1}^{N} \sum_{t=1}^{T} \int \left(Y_{it} - \sum_{j=1}^{q} \widehat{m}_j(u_j) - \frac{Z_{jit} - u_j}{h_j} \widehat{m}_j^j(u_j) - \widehat{\eta}_t - \widehat{\mu}_i \right)^2 \\ \times K_{h_1}(u_1, Z_{1it}) \cdots K_{h_q}(u_q, Z_{qit}) du_1 \cdots du_q, \end{aligned} \tag{3.55}$$

under the following constraints

$$\int \widehat{m}_j\left(Z_j\right) \widehat{f}_j\left(Z_j\right) dZ_j + \int \widehat{m}_j^j\left(Z_j\right) \widehat{f}_j^j\left(Z_j\right) dZ_j = 0, \tag{3.56}$$

$$\sum_{i=1}^{N} \left(n_i/n\right) \hat{\mu}_i = 0. \tag{3.57}$$

By changing accordingly the steps of the algorithm proposed previously for the local constant estimator we obtain estimators for m_j that under the conditions established in Section 3 and some

additional conditions detailed in Mammen et al. (2009) we obtain that, as both T and N tends to infinity in such a way that $T^{3/2}N^{-1+\delta} \to 0$ for some $\delta > 0$,

$$(d_T N)^{2/5} \left(\widehat{m}_j(z_j; h) - m_j(z_j) - b_j(z_j)\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \upsilon_j(z_j)\right) \tag{3.58}$$

where

$$\begin{aligned} b_j(z_j) &= c_j^2 \frac{1}{2} \mu_2(K) m_j''(z_j), \\ \upsilon_j(z_j) &= \frac{\sigma^2}{f_j(z_j) c_j} \int K^2(u) du \end{aligned}$$

and

$$c_j = (d_T N)^{1/5} h_j + o(1).$$

3.4 Nonparametric dynamic panel data models with fixed effects

In the previous section we have analyzed a rather complex situation where lagged endogenous explanatory variables and additive models are allowed for in a fixed effects context. However, the need of considering some dynamics in fixed effects panel data models appears of great interest even if the curse of dimensionality is not present. More precisely, we consider the following fixed effects nonparametric panel data model

$$Y_{it} = m(Y_{i(t-1)}, Z_{it}) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{3.59}$$

where Z_{it} is a $q \times 1$ vector of explanatory variables, $Y_{i(t-1)}$ a scalar lagged dependent variable, μ_i the cross-sectional heterogeneity and v_{it} the error term. All assumptions introduced at the beginning of Section 3 still hold here but with the additional assumption of the presence of lagged endogenous variables as explanatory variables.

Using a first difference transformation to remove the fixed effect we obtain

$$\Delta Y_{it} = m(Y_{i(t-1)}, Z_{it}) - m(Y_{i(t-2)}, Z_{i(t-1)}) + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T. \tag{3.60}$$

Note that after this differencing transformation the error term Δv_{it} has the form of moving average process of order 1, MA(1), that, in general, is correlated with the explanatory variable $Y_{i(t-1)}$. Therefore, in this setting, the conventional kernel estimation based on marginal integration or backfitting procedures does not provide consistent estimators for $m(\cdot)$.

In this framework, in Su and Lu (2013) it is developed an iterative estimator that is based on a local polynomial regression technique. Let $U_{i(t-2)} = (Y_{i(t-2)} \quad Z_{i(t-1)}^\top)^\top$ and assume $U_{i(t-2)}$ has a positive density on φ, $f_{t-2}(\cdot)$, where φ denotes a compact set on $I\!R^{q+1}$. Then, since Δv_{it} is (conditionally) mean-independent of $U_{i(t-2)}$, by the law of iterated expectations the following conditional moment condition can be obtained

$$E\left[\Delta v_{it} | U_{i(t-2)}\right] = 0, \tag{3.61}$$

and then

$$E\left[\Delta Y_{it} - m(Y_{i(t-1)}, Z_{it}) + m(Y_{i(t-2)}, Z_{i(t-1)})|U_{i(t-2)}\right] = 0. \tag{3.62}$$

Rearranging terms in (3.62) we obtain,

$$\begin{aligned} m(u) &= -E\left[\Delta Y_{it}|U_{i(t-2)} = u\right] + E\left[m(U_{i(t-1)})|U_{i(t-2)} = u\right] \\ &= r_{t|t-2}(u) + \int m(\overline{u}) f_{t-1|t-2}(\overline{u}|u) d\overline{u}, \qquad \text{for} \qquad t = 3, \cdots, T, \end{aligned} \tag{3.63}$$

where $r_{t|t-2}(u) = E\left[\Delta Y_{it}|U_{i(t-2)} = u\right]$, $f_{t-1|t-2}(\cdot|\cdot)$ is the conditional density function of $U_{i(t-1)}$ given $U_{i(t-2)}$ and $\overline{u}$ is the mean value of u.

For the sake of simplicity, let us denote by $\rho_{t-2} = P(U_{i(t-2)} \in \varphi)$ and $\rho = \sum_{t=3}^{T} \rho_{t-2}$, so if we multiply both sides of (3.63) by ρ_{t-2}/ρ and use the fact that $\sum_{t=3}^{T} \rho_{t-2}/\rho = 1$ we get

$$m(u) = r(u) + \int m(\overline{u}) f(\overline{u}|u) d\overline{u}. \tag{3.64}$$

Under certain regularity conditions (3.64) can be rewritten as

$$m = r + \mathbb{A}m, \tag{3.65}$$

where $\mathbb{A}$ is a bounded linear operator defined as $\mathbb{A}m(u) = \int m(\overline{u}) f(\overline{u}|u) d\overline{u}$.

Therefore, from (3.65) we can intuitively conclude that the estimator for the parameter of interest $m(\cdot)$ can be defined as a solution to a second order Fredholm integral equation in an infinite dimensional Hilbert space. However, since both r and $\mathbb{A}m(u)$ are not directly observable, the resulting estimator for (3.65) is infeasible and an iterative procedure is needed. In this situation, in Su and Lu (2013) it is proposed a plug-in estimator for $\widehat{m}(\cdot)$ that is the solution to

$$\widehat{m} = \widehat{r} + \widehat{\mathbb{A}}\widehat{m}, \tag{3.66}$$

where $\widehat{r}$ and $\widehat{\mathbb{A}}$ are nonparametric estimators obtained from a local polynomial regression of pth order. In particular, $r(u)$ can be estimated as the solution to the following criterion function,

$$\sum_{i=1}^{N}\sum_{t=3}^{T}\left(-\Delta Y_{it} - \sum_{0\le|\mathbf{j}|\le p} \gamma_j^{\top}((U_{i(t-2)} - u)/h)^{\mathbf{j}}\right)^2 K_h(U_{i(t-2)} - u)\mathbf{1}(U_{i(t-2)} \in \varphi), \tag{3.67}$$

where $|\mathbf{j}| \equiv \sum_{i=1}^{q} j_i$ and γ stacks the $\gamma_{\mathbf{j}}$'s $(0 \le |\mathbf{j}| \le q)$ that minimizes (3.67) in lexicographic order (with $\gamma_{\mathbf{0}}$ indexed by $\mathbf{0} \equiv (0, \cdots, 0)$ in the first position, the element with index $(0, 0, \cdots, 1)$ next, etc.). Also, note that let $j_0 + j_1 + \cdots + j_q = k$,

$$\sum_{0\le|\mathbf{j}|\le p} \equiv \sum_{k=0}^{p}\sum_{j_0=0}^{k}\cdots\sum_{j_q=0}^{k}.$$

Analogously, $\widehat{\mathbb{A}}m(u)$ is defined as the resulting estimator for $\mathbb{A}m(u)$ when $-\Delta Y_{it}$ is replaced by $m(U_{i(t-1)})$ in the problem to minimize (3.67). However, note that a feasible estimator for $\widehat{\mathbb{A}}$ needs

to observe the $m(\cdot)$ function. In that case, in Su and Lu (2013) it is proposed to resort to the sieve method and, after obtaining $\widehat{m}(u)$, to replace it in the final regression to estimate. See Chen (2007) for an intensive revision of the sieve method.

Let $h! = \prod_{r=0}^{q} h_r^2$ and $||h||^2 = \sum_{r=0}^{q} h_r^2$, under certain standard smoothing conditions, Su and Lu (2013) also establish the uniform consistency and the asymptotic normality of the plug-in estimator when $N \to \infty$, T is fixed, $||h|| \to 0$, $Nh!/logN \to \infty$ and $N||h||^4 h! \to c \in [0, \infty]$. However, before proceeding to the analysis of the asymptotic behavior of this plug-in estimator it is necessary to emphasize that this estimator only uses those observations of $U_{i(t-2)}$ that lie in a compact set φ on $I\!R^{q+1}$. Thus, allowing $U_{i(t-2)}$ to have a non-compact support facilitates the study of the asymptotic peculiarities of these estimators. In this context, they obtain that for the local linear estimator

$$\sqrt{NTh!}(\widehat{m}(u;h) - m(u) - (I - \mathbb{A})^{-1}B(u)) \quad \xrightarrow{d} \quad \mathcal{N}(0, \upsilon(u)), \tag{3.68}$$

where $\mathbb{A}$ is a Hilbert-Schmidt operator and for $\mu_2(K) = \int v^2 K(v)dv$ and $R^{q+1}(K) = \left[\int K(v)^2 dv\right]^{q+1}$,

$$\begin{aligned} B(u) &= \frac{1}{2}\sum_{r=0}^{q} h_r^2 \mu_2(K) \frac{\partial^2 m(u)}{\partial u_r^2}, \\ \upsilon(u) &= \frac{\sigma^2(u)}{f(u)} R^{q+1}(K), \\ \sigma^2(u) &= \sum_{t=3}^{T} \frac{\rho_{t-2}}{\rho} \sigma_{t-2}^2(u) f_{t-2}(u) \end{aligned}$$

and $\sigma_{t-2}^2(u) = E(v_{it}^2 | U_{i(t-2)} = u) + E(v_{i(t-1)}^2 | U_{i(t-2)} = u)$.

Note that the asymptotic variance of this iterative estimator has a similar structure to that presented by a conventional local polynomial estimator for $m(U_{i(t-1)}$ in $\Delta Y_{it} = m(U_{i(t-1)}) - m(U_{i(t-2)}) + \Delta v_{it}$, when $m(U_{i(t-2)})$ is observed. In addition, as regards to the asymptotic bias, we can see that it shows significant variations with reference to the standard results. Specifically, this iterative estimator presents an additional operator, $(I - \mathbb{A})^{-1}$, which reflects the cumulative bias of the iterative process.

Furthermore, in Su and Lu (2013) it is noted that, although consistency and asymptotic normality of the resulting estimator are shown, it is possible to propose an oracle efficient estimator. That is, an iterative estimator that exhibits the same asymptotic properties as if the iterations were evaluated at the true parameter (function) values. Note that this concept of oracle efficiency is different from the standard concept of efficiency. As long as $Y_{i(t-1)}$ and Z_{it} are compactly supported variables and the density function is bounded away from zero to the union of their supports (φ), they propose to use the infra-smoothing kernel estimator obtained previously as an estimator for $m(Y_{i(t-1)}, Z_{it})$. Then, once $\widehat{m}(Y_{i(t-1)}, Z_{it})$ is known, we can perform a local polynomial regression of ΔY_{it}^* on $m(Y_{i(t-2)}, Z_{i(t-1)})$, where $\Delta Y_{it}^* = \Delta Y_{it} - \widehat{m}(Y_{i(t-1)}, Z_{it})$. Finally, note that since the error process is not invertible, it is not possible to apply a similar procedure as in Xiao et al. (2003) and Su and Ullah (2006a) to provide more efficient estimators via the use of the MA(1) structure of this model.

4 Semi-parametric panel data models with random effects

As we have just shown, nonparametric panel data models are very appealing since they do not make too restrictive assumptions on the model specification and they allow data to tailor the shape of the regression function by themselves. However, as it has been pointed out in the introduction, in some situations this flexibility presents some shortcomings. To solve them, semi-parametric panel data models appear as a reasonable compromise between fully nonparametric and parametric models. In fact, they enable us to incorporate some prior information coming from economic theory or past experience by keeping at the same time more flexibility in the specification of the model. Among the most popular semi-parametric panel data model we consider here the so-called partially linear models.

Instead of (2.1), the basic partially linear unobserved effects model can be written, for a randomly drawn cross-section observation i, as

$$Y_{it} = X_{it}^{\top}\beta + m\left(Z_{it}\right) + \epsilon_{it}, \quad t = 1, \cdots, T, \tag{4.1}$$

where X_{it} and Z_{it} are vectors of explanatory variables of $d \times 1$ and $q \times 1$ dimension, respectively, β is a $d \times 1$ vector of unknown parameters and $m(\cdot)$ is an unknown smooth function. Both objects need to be estimated and ϵ_{it} is an unobservable error term. Typically, the error term of the model follows a one-way error component structure as in (2.2). Furthermore, instead of (2.3), now we assume

$$E\left(Y_{it}|\, X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}, \mu_i\right) = E\left(Y_{it}|\, X_{it}, Z_{it}, \mu_i\right) = X_{it}^{\top}\beta + m\left(Z_{it}\right) + \mu_i. \tag{4.2}$$

Note that, this first equality establishes the relationship between Y and the past values of both Z and X, whereas the second one establishes that the regression function is the sum of a parametric term, $X_{it}^{\top}\beta$, a nonparametric function, $m(Z_{it})$, and an unobservable heterogeneity term μ_i. Using (4.1) and (2.2), the assumption in (4.2) can be stated in terms of the idiosyncratic errors as

$$E(v_{it}|X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}, \mu_i) = 0, \quad t = 1, \cdots, T. \tag{4.3}$$

Let $v_i = (v_{i1}, \cdots, v_{iT})^{\top}$ be a $T \times 1$ vector. The error vector v_i and the heterogeneity term μ_i are such that

$$E\left(v_i v_i^{\top}\middle|\, X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}, \mu_i\right) = \sigma_v^2 I_T, \quad E(\mu_i^2|X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}) = \sigma_\mu^2. \tag{4.4}$$

As in the fully nonparametric case, under the assumptions above and since the observations are independent in i and j, the variance-covariance matrix of the composed error term has the standard form as in (2.7). Finally, to characterize the random effects model, instead of (2.8) we assume

$$E(\mu_i|X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}) = E(\mu_i) = 0. \tag{4.5}$$

Note that using (4.1) and (2.2), under assumptions (4.2), (4.3), (4.4) and (4.5) and by the law of iterated expectations,

$$E\left(Y_{it}|\, X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}\right) = X_{it}^{\top}\beta + m\left(Z_{it}\right), \quad t = 1, \cdots, T. \tag{4.6}$$

Given the previous result, root-N consistent estimation of the parameters of interest, β, is possible through the use of standard techniques in partially linear models (see Robinson (1988) and Speckman (1988) among others). Then, following Li and Stengos (1996) and Li and Ullah (1998) the unknown function $m(\cdot)$ is removed from (4.1) by taking the conditional expectation on Z_{it} in (4.1) and assuming that $E\left(\epsilon_{it}\middle| Z_{it}\right) = 0$,

$$E\left(Y_{it}\middle| Z_{it}\right) = \beta^{\top} E\left(X_{it}\middle| Z_{it}\right) + m\left(Z_{it}\right), \quad t = 1, \cdots, T. \tag{4.7}$$

Then, subtracting (4.7) from (4.1) we obtain

$$Y_{it} - E(Y_{it}|Z_{it}) = \beta^{\top}\left(X_{it} - E(X_{it}|Z_{it})\right) + \epsilon_{it}. \tag{4.8}$$

Once the unknown term $m\left(Z_{it}\right)$ has been removed we can estimate the parameters of interest, β, by using a standard OLS techniques and then

$$\widehat{\beta}_{OLS} = (\widetilde{X}^{\top}\widetilde{X})^{-1}\widetilde{X}^{\top}\widetilde{Y}, \tag{4.9}$$

where $\widetilde{X}$ is a $NT \times d$ matrix whose typical row element is $\widetilde{X}_{it} = X_{it} - E(X_{it}|Z_{it})$, and $\widetilde{Y}$ is a $NT \times 1$ vector whose typical row element is of the form $\widetilde{Y}_{it} = Y_{it} - E(Y_{it}|Z_{it})$.

Note that both $E(X_{it}|Z_{it})$ and $E(Y_{it}|Z_{it})$ are unknown terms, and therefore (4.9) is an infeasible estimator. To overcome this problem, we typically substitute the unknown quantities by their estimators, i.e.

$$\widehat{X}_{it} = \widehat{E}(X_{it}|Z_{it}) = \frac{1}{NTh^q}\sum_{j=1}^{N}\sum_{s=1}^{T} X_{js}K_h(Z_{it} - Z_{js})/\widehat{f}_h\left(Z_{it}\right), \tag{4.10}$$

$$\widehat{Y}_{it} = \widehat{E}(Y_{it}|Z_{it}) = \frac{1}{NTh^q}\sum_{j=1}^{N}\sum_{s=1}^{T} Y_{js}K_h(Z_{it} - Z_{js})/\widehat{f}_h\left(Z_{it}\right), \tag{4.11}$$

$$\widehat{f}_h(Z_{it}) = \frac{1}{NTh^q}\sum_{j=1}^{N}\sum_{s=1}^{T} K_h(Z_{it} - Z_{js}). \tag{4.12}$$

For the sake of simplicity, let us denote $\widehat{f}_h\left(Z_{it}\right) = \widehat{f}_{it}$. Replacing the unknown conditional expectations in (4.9) by their respective nonparametric estimators in (4.10)-(4.12) we obtain

$$\widetilde{\beta}_{FOLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \widehat{X}_{it}\right)\left(X_{it} - \widehat{X}_{it}\right)^{\top}\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \widehat{X}_{it}\right)\left(Y_{it} - \widehat{Y}_{it}\right). \tag{4.13}$$

As in other applications of kernel regression, the estimators $\widehat{X}_{it}$ and $\widehat{Y}_{it}$ cause technical difficulties owing to the random denominator, $\widehat{f}_{it}$, that can be small. In order to avoid it, we trim out small values of $\widehat{f}$ as it is done for example in Powell et al. (1989). For a constant $b > 0$ define $\widehat{I}_{it} = \mathbf{1}\left(\left|\widehat{f}_{it}\right| > b\right)$, where $\mathbf{1}(\cdot)$ is the usual indicator function; so the feasible ordinary least-squares (FOLS) estimator for β can be written as

$$\widehat{\beta}_{FOLS} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \widehat{X}_{it}\right)\left(X_{it} - \widehat{X}_{it}\right)^{\top}\widehat{I}_{it}\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \widehat{X}_{it}\right)\left(Y_{it} - \widehat{Y}_{it}\right)\widehat{I}_{it}. \tag{4.14}$$

Under standard regularity conditions that include those assumed at the beginning of Section 4, $h, b \to 0$, as N tends to infinity, in Li and Ullah (1998) it is shown the following result,

$$\sqrt{N}(\widehat{\beta}_{FOLS} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Sigma\Phi^{-1}\right), \tag{4.15}$$

where $\Phi = \frac{1}{T}\sum_{t=1}^{T} E(\widetilde{X}_{it}\widetilde{X}_{it}^{\top} f_{it}^2)$ and $\Sigma = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} E(v_{it}v_{is}\widetilde{X}_{it}\widetilde{X}_{is}^{\top} f_{it}^2 f_{is}^2)$. It can be shown that

$$\widehat{\Phi} = \frac{1}{NT}\sum_{i}\sum_{t}\left(X_{it} - \widehat{X}_{it}\right)\left(X_{it} - \widehat{X}_{it}\right)^{\top}\widehat{I}_{it}$$

and

$$\widehat{\Sigma} = \frac{1}{NT}\sum_{i}\sum_{t}\widehat{I}_{it}\left(X_{it} - \widehat{X}_{it}\right)\widehat{\Omega}_{it}\left(X_{it} - \widehat{X}_{it}\right)^{\top}$$

are consistent estimators of Φ and Σ, respectively. Note that $\widehat{\Omega}_{it}$ is the it-th component of the matrix in (2.27).

Also, this estimation strategy can be extended to other types of situations of interest such as models where endogenous or lagged dependent variables are allowed as explanatory variables. As we will state in Section 6, in Li and Stengos (1996) it is proposed an instrumental variable (IV) method to solve the endogeneity problem. Later, in Kneisner and Li (2002) it is analyzed a dynamic semi-parametric panel data model and, under the assumption that the error term is serially uncorrelated, it is shown that it is possible to obtain a $\sqrt{N}$-consistent estimator for β adapting the previous weighted density problem to this dynamic case.

Unlike other results such as those in Li and Stengos (1996), in Kneisner and Li (2002) it is proposed a two step local linear method to estimate the smooth function $m(\cdot)$. By subtracting the (estimated) fully parametric part in both terms of (4.1) we have that

$$Y_{it} - X_{it}^{\top}\widehat{\beta}_{FOLS} = m(Z_{it}) + X_{it}^{\top}(\beta - \widehat{\beta}_{FOLS}) + \epsilon_{it}. \tag{4.16}$$

Given that $\widehat{\beta}_{FOLS} = \beta + O_p\left(\frac{1}{\sqrt{N}}\right)$, we can write the previous equation such as a standard non-parametric problem, i.e.,

$$Y_{it} - X_{it}^{\top}\widehat{\beta}_{FOLS} = m(Z_{it}) + \epsilon_{it}^{*}, \tag{4.17}$$

where $\epsilon_{it}^{*} = X_{it}^{\top}(\beta - \widehat{\beta}_{FOLS}) + \epsilon_{it}$. Then, $m(Z_{it})$ can be consistently estimated through standard non-parametric regression techniques. We refer to Kneisner and Li (2002) for the asymptotic properties of this type of estimators.

An alternative approach is introduced in Fan and Huang (2005). The main idea in this paper is to transform a semi-parametric problem into a nonparametric one. This is done by subtracting the parametric component to both terms in (4.1). Then,

$$Y_{it} - X_{it}^{\top}\beta = m(Z_{it}) + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T \tag{4.18}$$

If β were known, then $m(\cdot)$ can be estimated by a standard local linear regression problem. Then, let $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ be the minimizers of

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\Big((Y_{it}-X_{it}^{\top}\beta)-\gamma_0-\gamma_1^{\top}(Z_{it}-z)\Big)^2 K_h(Z_{it}-z). \tag{4.19}$$

We suggest as estimators for $m(z)$ and $D_m(z)=vec(\partial m(z)/\partial z^{\top})$, $\widehat{m}(z;h)=\widehat{\gamma}_0$ and $\widehat{D}_m(z;h)=\widehat{\gamma}_1$, respectively,

$$\widehat{\gamma}_0=(1,0_q)(D^{\top}K_zD)^{-1}D^{\top}K_z(Y-X\beta)=S(Y-X\beta) \tag{4.20}$$

and

$$\widehat{\gamma}_1=(0\imath_q)(D^{\top}K_zD)^{-1}D^{\top}K_z(Y-X\beta), \tag{4.21}$$

where 0_q and $\imath_q$ are q-vectors of zeros and ones, respectively, $S=(1,0_q)(D^{\top}K_zD)^{-1}D^{\top}K_z$ is a so-called smoothing matrix, $K_z=diag\,(K_h(Z_{11}-z),\cdots,K_h(Z_{NT}-z))$ is a $NT\times NT$ weighting kernel matrix, $X=(X_{11},\cdots,X_{NT})^{\top}$ is a $NT\times d$ matrix and D is a $NT\times(1+q)$ matrix such as

$$D=\begin{bmatrix}1 & (Z_{11}-z)^{\top}\\ \vdots & \vdots\\ 1 & (Z_{NT}-z)^{\top}\end{bmatrix}.$$

However, because β is a vector of unknown parameters that need to be estimated, we can replace $m(Z_{it})$ with $\widehat{m}(Z_{it};h)=\widehat{\gamma}_0$ in (4.18), so the regression function to estimate now is of the form

$$\widehat{Y}_{it}=\widehat{X}_{it}^{\top}\beta+\epsilon_{it}^{*},\quad i=1,\cdots,N\quad;\quad t=1,\cdots,T, \tag{4.22}$$

where $(\widehat{Y}_{11},\cdots,\widehat{Y}_{NT})^{\top}=(I_{NT}-S)Y$, $(\widehat{X}_{11},\cdots,\widehat{X}_{NT})^{\top}=(I_{NT}-S)X$, and $(\epsilon_{11}^{*},\cdots,\epsilon_{NT}^{*})^{\top}=(I_{NT}-S)\epsilon+(I_{NT}-S)m(Z)$, where $m(Z)=(m(Z_{11}),\cdots,m(Z_{NT}))^{\top}$ and I_{NT} is an identity $NT\times NT$ matrix.

We denote by $\widehat{\beta}_{FLSS}$ as the feasible semi-parametric least-squares estimator for (4.22) of the form

$$\widehat{\beta}_{FLSS}=\left(\widehat{X}^{\top}\widehat{X}\right)^{-1}\widehat{X}^{\top}\widehat{Y}, \tag{4.23}$$

whereas the local linear estimator for $m(\cdot)$ is written as

$$\widehat{m}_{FLSS}(z;h)=(1,0_q)(D^{\top}KD)^{-1}D^{\top}K(Y-X\widehat{\beta}). \tag{4.24}$$

These results are standard in semi-parametric partially linear models (see Robinson (1988) and Speckman (1988)): The presence of nonparametric components, typically estimated at nonparametric rates, does not affect the rate of convergence of $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{FOLS}$ and $\widehat{\beta}_{FLSS}$ that is fully parametric ($\sqrt{N}$-consistency). It is also interesting to note that both OLS and FOLS estimators exhibit the same asymptotic variance that is, in both cases, regardless the fact that nonparametric

components are either estimated or taken at its true values, the asymptotic variance is the same. That is the so-called oracle efficiency property. However, it is true that they do not achieve the semi-parametric efficiency bounds (see Chamberlain (1992)) due to the one-way error component structure that they ignore. In Li and Ullah (1998) it is suggested that the use of the structure of the variance-covariance matrix Ω can be of interest in order to achieve this efficiency bound. As in the parametric case, they propose a feasible generalized least-squares (FGLS) semi-parametric estimator for β.

As an estimator for $\widehat{\Omega}$ they propose

$$\widehat{\sigma}_v^2 = \widehat{\sigma}^2 - \widehat{\sigma}_\mu^2 \qquad \text{and} \qquad \widehat{\sigma}_1^2 = T\widehat{\sigma}_\mu^2 + \widehat{\sigma}_v^2,$$

where $\widehat{\sigma}^2 = \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \widehat{\epsilon}_{it}^2 \widehat{f}_{it}$ and $\widehat{\sigma}_\mu^2 = \frac{1}{NT(T-1)}\sum_{i=1}^N \sum_{t=1}^T \widehat{\epsilon}_{it}\widehat{\epsilon}_{is}\widehat{f}_{it}\widehat{f}_{is}$ and

$$\widehat{\epsilon}_{it} = (Y_{it} - \widehat{Y}_{it}) - (X_{it} - \widehat{X}_{it})^\top \widehat{\beta}_{FOLS}. \tag{4.25}$$

If we replace the unknown terms by these estimators in (2.7), the FGLS estimator for β is

$$\widehat{\beta}_{FGLS} = \left((X - \widehat{X})^\top \widehat{\Omega}^{-1}(X - \widehat{X})\widehat{I}\right)^{-1} \left((X - \widehat{X})\widehat{I}\right)^\top \widehat{\Omega}^{-1}(Y - \widehat{Y}), \tag{4.26}$$

where $(X - \widehat{X})\widehat{I}$ is a matrix of $NT \times d$ dimension with a typical row element $(X_{it} - \widehat{X}_{it})^\top \widehat{I}_{it}$ and $Y - \widehat{Y}$ is a vector of dimension $NT \times 1$. Under some standard regularity conditions that include the assumptions established at the beginning of Section 4 in Li and Ullah (1998) it is shown that

$$\sqrt{N}(\widehat{\beta}_{FGLS} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Sigma\Phi^{-1}/T\right), \tag{4.27}$$

where $\Phi = \frac{1}{T}\sum_{t=1}^T E(\widetilde{X}_{it}\Omega_{it}\widetilde{X}_{it}^\top f_{it}^2)$, $\Sigma = \frac{1}{T^2}\sum_{t=1}^T \sum_{s=1}^T E(\epsilon_{it}\epsilon_{is}\widetilde{X}_{it}\Omega_{it}\widetilde{X}_{it}^\top f_{it}^2 f_{is}^2)$ and Ω_{it} is the $it-th$ element of the Ω matrix defined in (2.7).

As the reader can realize, this estimator achieves the semi-parametric efficiency bound for this type of problem. The estimation strategy developed in Robinson (1988) can be easily extended to other contexts within the framework of partially linear panel data models. However, it is true that the presence of heteroscedastic errors in the model of interest complicates this procedure considerably. In this context, You et al. (2010) propose an alternative method to obtain consistent nonparametric estimators that take into account the one-way error component structure and allow for unequal error variances, i.e., heteroscedastic errors.

More precisely, in You et al. (2010) it is considered a one-way error component structure with heterocedasticity in the idiosyncratic error of the following form

$$\epsilon_{it} = \mu_i + \sigma(Z_{it})v_{it}, \qquad i = 1, \cdots, N \qquad ; \qquad t = 1, \cdots, T, \tag{4.28}$$

where $\sigma_v^2 = 1$ is assumed without loss of generality so the variance-covariance matrix of the error term is written as

$$\Omega = E(\epsilon\epsilon^\top) = \sigma_\mu^2 I_N \otimes (\imath_T \imath_T^\top) + diag\left(\sigma^2(Z_{11}), \cdots, \sigma^2(Z_{NT})\right). \tag{4.29}$$

With this complex variance-covariance structure, we would need an estimator for each error variances and therefore the previous procedures cannot be used directly. In You et al. (2010) it is developed a semi-parametric weighted least-squares estimator for β based on those previous results. Specifically, they propose to estimate both the variance of the error term and the error structure and, later, use this information to obtain an efficient semi-parametric estimator. More precisely, assuming that Z_{it} is a scalar, in You et al. (2010) the following residuals are used

$$\widehat{\epsilon}_{it} = Y_{it} - X_{it}^{\top}\widehat{\beta}_{FLLS} - \widehat{m}_{FLLS}(Z_{it}; h), \tag{4.30}$$

and because $E(\epsilon_{it}\epsilon_{is}) = \sigma_{\mu}^2$, when $t \neq s$, and $E\left(\epsilon_{it}^2 \middle| Z_{it}\right) = \sigma_{\mu}^2 + \sigma^2(Z_{it})$, consistent estimators for σ_{μ}^2 and $\sigma^2(\cdot)$ can be written as

$$\widehat{\sigma}_{\mu}^2 = \frac{1}{NT(T-1)} \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{t\neq s}^{T} \widehat{\epsilon}_{it}\widehat{\epsilon}_{is} \qquad \text{and} \qquad \widehat{\sigma}^2(z) = \sum_{i=1}^{N}\sum_{t=1}^{T} \omega_{it}(z)\widehat{\epsilon}_{it} - \widehat{\sigma}_{\mu}^2, \tag{4.31}$$

where $\omega_{it}(z)$ is some weight function of the local linear estimator such that

$$\omega_{it}(z) = \frac{(Nh)^{-1}K((Z_{it}-z)/h)\left(A_{k2}(z) - (Z_{it}-z)A_{k1}(z)\right)}{A_{k0}(z)A_{k2}(z) - A_{k1}^2(z)}$$

and $A_{ks}(z) = \frac{1}{Nh}\sum_{i=1}^{N}\sum_{t=1}^{T} K\left(\frac{Z_{it}-z}{h}\right)(Z_{it}-z)^{\lambda}$, for $\lambda = 0, 1, 2$.

Consequently, the estimator for Ω^{-1} is given by $\widehat{\Omega}^{-1} = \text{blockdiag}(\widehat{\Sigma}_1^{-1}, \cdots, \widehat{\Sigma}_N^{-1})$, where

$$\begin{aligned}\widehat{\Sigma}_i^{-1} &= diag\left(\widehat{\sigma}^{-2}(Z_{i1}), \cdots, \widehat{\sigma}^{-2}(Z_{iT})\right) - \left(\widehat{\sigma}_{\mu}^{-2} + \sum_{t=1}^{T}\widehat{\sigma}^{-2}(Z_{it})\right)^{-1}\left(\widehat{\sigma}^{-2}(Z_{11}), \cdots, \widehat{\sigma}^{-2}(Z_{iT})\right)^{\top} \\ &\times \left(\widehat{\sigma}^{-2}(Z_{11}), \cdots, \widehat{\sigma}^{-2}(Z_{iT})\right).\end{aligned}$$

Replacing Ω with $\widehat{\Omega}$, the feasible weighted least-squares semi-parametric estimator (WSLSE) is

$$\widehat{\beta}_{WLSS} = \left(X^{\top}(I_{NT}-S)^{\top}\widehat{\Omega}^{-1}(I_{NT}-S)X\right)^{-1}\left(X(I_{NT}-S)\right)^{\top}\widehat{\Omega}^{-1}(I_{NT}-S)Y. \tag{4.32}$$

under the conditions established in Section 2 they show that, as $N \to \infty$,

$$\sqrt{N}(\widehat{\beta}_{WLSS} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Sigma_3^{-1}\right), \tag{4.33}$$

where $\Sigma_3 = \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E(v_i^{\top}\Omega_i^{-1}v_i)$, let Ω_i the $i-th$ element of the Ω matrix defined in (2.7) and $v_i = (v_{i1}, \cdots, v_{iT})^{\top}$ is a $T \times 1$ vector.

5 Semi-parametric panel data models with fixed effects

In this section we are interested in statistical techniques that provide $\sqrt{N}$-consistent estimators of the parameters β in (4.1) when the relationship between the heterogeneity term μ_i and the explanatory variables $X_{i1}, \cdots, X_{iT}$, $Z_{i1}, \cdots, Z_{iT}$ is modeled as

$$E(\mu_i | X_{i1}, \cdots, X_{iT}, Z_{i1}, \cdots, Z_{iT}) = \mu_i. \tag{5.1}$$

Then, using (4.1) and (2.2), under assumptions (4.2), (4.3), (4.4) and (5.1) and by the law of iterated expectations we obtain that

$$E\left(Y_{it}\middle|\, X_{i1},\cdots,X_{iT},Z_{i1},\cdots,Z_{iT}\right)=X_{it}^{\top}\beta+m\left(Z_{it}\right)+\mu_i,\quad t=1,\cdots,T. \tag{5.2}$$

By comparing (4.6) and (5.2) we realize that direct application of the statistical estimation techniques applied for the random effects case provide asymptotically biased estimators of the parameters of interest. As an alternative, we will use similar instruments to those already used in Section 3. More precisely, we will distinguish between the so-called profiling techniques and the differencing methods. Finally, as in the fully fixed effects nonparametric case, we will also consider the problem when the nonparametric object is of high dimension and hence some restriction of additivity is needed to cope with the curse of dimensionality. Profiling techniques in this context have been introduced in Su and Ullah (2006b) and Zhang et al. (2011). On the other side, differencing techniques have been originally proposed in Baltagi and Li (2002) and Qian and Wang (2012). The former paper proposes to estimate $m\left(\cdot\right)$ using series estimators whereas the latter use marginal integration techniques. Finally, we analyze the proposal of Ai et al. (2014) to ameliorate the dimensionality problem related to the explanatory variables through the estimation of an additive version of the semi-parametric regression model as in (3.37).

In Su and Ullah (2006b) it is proposed to profile both the heterogeneity term and the nonparametric part to consistently estimate the parameter vector β. Let $Y=(Y_{11},\cdots,Y_{NT})^{\top}$ be a $NT\times 1$ vector and $X=(X_{11},\cdots,X_{NT})^{\top}$ a matrix of $NT\times d$ dimension. Furthermore, let $\mu_0=(\mu_2,\cdots,\mu_N)^{\top}$ be a $(N-1)$-dimensional vector and $D_d=(I_N\otimes\imath_T)d$ a $NT\times(N-1)$ dimensional matrix, where $d=(-\imath_{N-1}I_{N-1})^{\top}$ is a $N\times(N-1)$ matrix, the standard locally weighted linear least squares regression to estimate the quantities of interest in (4.1) can be written in matrix form as

$$(Y-D_d\mu_0-X\beta-Z_z\gamma)^{\top}K_z(Y-D_d\mu_0-X\beta-Z_z\gamma), \tag{5.3}$$

where $K_z=diag(K_H(Z_{11}-z),\cdots,K_H(Z_{NT}-z))$ is a $NT\times NT$ matrix, $K_H(z)=|H|^{-1}K(H^{-1}z)$, $K\left(\cdot\right)$ is a kernel function, $|H|$ is the determinant of the bandwidth matrix H and Z_z is a $NT\times(1+q)$ matrix of the form,

$$Z_z=\begin{bmatrix}1 & \left(H^{-1}(Z_{11}-z)\right)^{\top}\\ \vdots & \vdots\\ 1 & \left(H^{-1}(Z_{NT}-z)\right)^{\top}\end{bmatrix}.$$

The above exposition suggests as estimators for $m(z)$ and $D_m(z)=vec(\partial m(z)/\partial z^{\top})$, $\widehat{m}(z;H)=\widehat{\gamma}_0$ and $\widehat{D}_m(z;H)=\widehat{\gamma}_1$, respectively,

$$\widehat{m}(z;H)=S(z)(Y-D_d\mu_0-X\beta), \tag{5.4}$$

where $s(z)=e_1^{\top}S(z)$ for $S(z)=(Z_z^{\top}K_zZ_z)^{-1}Z_z^{\top}K_z$, and $e=(1,0_q)^{\top}$ is a $(1+q)\times 1$ selection matrix.

Then, by replacing (5.4) in the following optimization problem

$$(Y - D_d\mu_0 - X\beta - m(Z))^\top (Y - D_d\mu_0 - X\beta - m(Z)), \tag{5.5}$$

the minimizers of (5.5) can be written as

$$\begin{aligned} \widehat{\beta}_{FPL} &= (X^{*\top} M_d^* X^*)^{-1} X^{*\top} M_d^* Y^*, & (5.6)\\ \widehat{\mu}_{FPL} &= (D_d^{*\top} D_d^*)^{-1} D_d^{*\top} (Y^* - X^* \widehat{\beta}_{FPL}), & (5.7) \end{aligned}$$

where $D_d^* = (I_{NT} - S)D_d$, $Y^* = (I_{NT} - S)Y$, $X^* = (I_{NT} - S)X$, $M_d^* = I_{NT} - D_d^*(D_d^{*\top} D_d^*)^{-1} D_d^*$, for $S_i = (s(Z_{i1}), \cdots, s(Z_{iT}))$ being a $T \times T$ smoothing matrix. Note that by identification conditions $\widehat{\mu}_1 = -\sum_{i=2}^N \widehat{\mu}_i$.

Under some standard conditions, in Su and Ullah (2006b) it is obtained the asymptotic distribution of these estimators as $N \to \infty$ and T is fixed,

$$\sqrt{N}(\widehat{\beta}_{FPL} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Omega\Phi^{-1}\right), \tag{5.8}$$

where $\widetilde{X}_{it} = X_{it} - E(X_{it}|Z_{it})$, $\Phi = \sum_t E(\widetilde{X}_{it}(\widetilde{X}_{it} - T^{-1}\sum_{s=1}^T \widetilde{X}_{is})^\top)$ and $\Omega = \sum_{s=1}^T \sum_{t=1}^T E(v_{it}v_{is}\widetilde{X}_{it}(\widetilde{X}_{is} - T^{-1}\sum_{\ell=1}^T \widetilde{X}_{i\ell})^\top)$.

Recently, in Zhang et al. (2011) it is proposed an empirical maximum likelihood estimator for β. This estimator is of the same form as the so-called feasible profile likelihood estimator in Su and Ullah (2006a). In the former paper, and based in the first-order conditions of (5.6), it is proposed the following auxiliary random vector to meet $E(\eta_i(\beta)) = 0$ when β is unknown,

$$\eta_i(\beta) = X_i^{*\top} M_{d_i}^* (Y_i^* - X_i^*\beta), \tag{5.9}$$

where $X_i^* = (X_{i1}^*, \cdots, X_{iT}^*)^\top$, $M_{d_i}^* = (M_{d_{i1}}^*, \cdots, M_{d_{iT}}^*)^\top$ and $Y_i^* = (Y_{i1}^*, \cdots, Y_{iT}^*)^\top$ are $T \times 1$ vectors.

The log-likelihood function is

$$\ell(\beta) = \sum_{i=1}^N log(N\rho_i). \tag{5.10}$$

with the constraints

$$\begin{aligned} \rho_i &\geq 0, \quad i = 1, \cdots, N, & (5.11)\\ \sum_{i=1}^N \rho_i &= 1, & (5.12)\\ \sum_{i=1}^N \rho_i \eta_i(\beta) &= 0. & (5.13) \end{aligned}$$

To find the values $\rho_1, \cdots, \rho_N$, we maximize the log-likelihood function (5.10) subject to the constraints (5.11) to (5.13). Using the Lagrange multiplier method, we obtain

$$\rho_i = \frac{1}{N}\frac{1}{1 + \lambda^\top(\beta)\eta_i(\beta)}, \quad i = 1, \cdots, N, \tag{5.14}$$

and substituting (5.14) into (5.10) we obtain

$$\ell(\beta) = -\sum_{i=1}^{N} log\left(1 + \lambda^{\top}(\beta)\eta_i(\beta)\right), \tag{5.15}$$

where $\lambda(\beta)$, the Lagrange multiplier, is determined by

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\eta_i(\beta)}{1+\lambda^{\top}(\beta)\eta_i(\beta)} = 0.$$

Now, we define the value of β that maximizes (5.15) as the maximum empirical likelihood estimator (MELE) of β, i.e.

$$\widehat{\beta}_{MELE} = \left(\sum_{i=1}^{N} X_i^{*\top} M_{d_i}^{*} X_i^{*}\right)^{-1} \sum_{i=1}^{N} X_i^{*\top} M_{d_i}^{*} Y_i^{*}, \tag{5.16}$$

so from this expression they obtain

$$\begin{aligned} \widehat{\mu}_{MELE} &= (D_d^{*\top} D_d^{*})^{-1} D_d^{*\top}(Y^{*} - X^{*}\widehat{\beta}_{MELE}), & (5.17)\\ \widehat{m}_{MELE}(z;H) &= s(z)(Y - D_d\widehat{\mu}_{MELE} - X\widehat{\beta}_{MELE}). & (5.18) \end{aligned}$$

Note that as it is pointed out by these authors, the maximum empirical likelihood estimator $\widehat{\beta}_{MELE}$ is identical to the profile likelihood estimator $\widehat{\beta}_{FPL}$ in Su and Ullah (2006b). In addition, following standard conditions and similar definitions as in Su and Ullah (2006b), in Zhang et al. (2011) it is obtained

$$\sqrt{N}(\widehat{\beta}_{MELE} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Lambda\Phi^{-1}\right), \tag{5.19}$$

as N tends to infinity, where $\Phi = \sum_{i=1}^{N} E\left(\widetilde{X}_{it}(\widetilde{X}_{it} - T^{-1}\sum_{s=1}^{T}\widetilde{X}_{is})^{\top}\right)$ and $\Lambda = E(\widetilde{X}_i^{\top}\Omega_i\widetilde{X}_i)^{\top}$, for $\widetilde{X}_i = (\widetilde{X}_{i1}, \cdots, \widetilde{X}_{iT})^{\top}$ and $\widetilde{X}_{it} = X_{it} - E(X_{it}|Z_{it})$.

An alternative approach to the so-called profiling methods are the differencing techniques. Using the first differences transformation in (4.1) we obtain

$$\Delta Y_{it} = \Delta X_{it}^{\top}\beta + m(Z_{it}, Z_{i(t-1)}) + \Delta v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{5.20}$$

where $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$. Estimation of β, as it is suggested in Li and Stengos (1996), can be implemented by conditioning (5.20) in $\left(Z_{it}, Z_{i(t-1)}\right)$. Assuming $E\left(\Delta v_{it}|\, Z_{it}, Z_{i(t-1)}\right) = 0$,

$$E\left(\Delta Y_{it}|\, Z_{it}, Z_{i(t-1)}\right) = \beta^{\top} E\left(\Delta X_{it}|\, Z_{it}, Z_{i(t-1)}\right) + m(Z_{it}, Z_{i(t-1)}). \tag{5.21}$$

Then subtracting (5.21) to (5.20) we obtain,

$$\Delta Y_{it} - E\left(\Delta Y_{it}|\, Z_{it}, Z_{i(t-1)}\right) = \beta^{\top}\left(\Delta X_{it} - E\left(\Delta X_{it}|\, Z_{it}, Z_{i(t-1)}\right)\right) + \Delta v_{it}, \tag{5.22}$$

where β can be estimated using the standard FOLS technique, where conditional expectations are replaced by conventional nonparametric estimators (see (4.9)-(4.12) for details).

However, as it is pointed out in Baltagi and Li (2002) this technique presents some weaknesses. On the one hand, taking conditional expectations on $(Z_{it}, Z_{i(t-1)})$ implies having to deal with the curse of dimensionality problem. In that case, it is necessary to estimate the nonparametric regression of $\Delta Y_{it} - E(\Delta Y_{it}|Z_{it}, Z_{i(t-1)})$ on $\Delta X_{it} - E(\Delta X_{it}|Z_{it}, Z_{i(t-1)})$ by the kernel method. This estimator has to be defined on $I\!R^{2q}$ rather than $I\!R^{q}$. On the other hand, and although these authors suggest how to estimate $m(Z_{it}, Z_{i(t-1)})$, they ignore the additive structure of (5.20) and do not provide a nonparametric estimator for $m(Z_{it})$. In this framework, Baltagi and Li (2002) develop an estimation method based on the series approach which enables us to impose the additive structure characteristic of first differences regression models and propose a nonparametric estimator for $m(Z_{it})$. Alternatively, in Qian and Wang (2012) it is presented a method based on marginal integration techniques to provide an estimator for this smooth function allowing for the presence of some endogenous explanatory variables.

If $m(\cdot)$ is a two times differentiable function twice differentiable in the interior of its support $\mathcal{A}$, being $\mathcal{A}$ a compact subset in $I\!R^{q}$, and $E[m''(z)] = E[\partial^2 m(z)/\partial z^2] < \infty$, the unknown function $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$ belongs to the class of additive functions $\mathcal{M}$ ($m \in \mathcal{M}$). Then, with the aim of taking into account the restriction that both additive functions share the same functional form, Baltagi and Li (2002) propose to approximate $m(z)$ through the series $\rho^L(z)$ of $L \times 1$ dimension, where $L = L(N)$.

Note that, as Baltagi and Li (2002) emphasize, the approximation function $\rho^L(z)$ has to meet a number of special features for the series method that can be summarized in the following

i) $\rho^L(z) \in \mathcal{M}$,

ii) as far as L increases, there is a linear combination of $\rho^L(z)$ that may approximate any $m \in \mathcal{M}$ arbitrarily well in mean square error.

In this way, $\rho^L(z)$ approximates $m(z)$ and $\rho^L(Z_{it}, Z_{i(t-1)}) = \rho^L(Z_{it}) - \rho^L(Z_{i(t-1)})$ approximates $m(Z_{it}, Z_{i(t-1)}) = m(Z_{it}) - m(Z_{i(t-1)})$, where

$$\rho^L(Z_{it}, Z_{i(t-1)}) = \begin{bmatrix} \rho_1(Z_{it}) - \rho_1(Z_{i(t-1)}) \\ \rho_2(Z_{it}) - \rho_2(Z_{i(t-1)}) \\ \vdots \\ \rho_L(Z_{it}) - \rho_L(Z_{i(t-1)}) \end{bmatrix}. \tag{5.23}$$

For any scalar or vector function $W(z)$, $E_{\mathcal{M}}(W(z))$ denotes an element which belongs to $\mathcal{M}$ and that is the closest function to $W(z)$ among all the functions in $\mathcal{M}$. Denote $P = (\rho^L_{11}, \cdots, \rho^L_{NT})$ a $NT \times L$ matrix, where $\rho^L_{it} = \rho^L(Z_{it}, Z_{i(t-1)})$. For the sake of simplicity, let us define $\theta(z) = E(X|Z = z)$ and $m(z) = E_{\mathcal{M}}(\theta(z))$ so the expression (5.20) can be written in matrix form as

$$\Delta Y = \Delta X\beta + M + \Delta v, \tag{5.24}$$

where ΔY and M are NT-dimensional vectors with a typical element ΔY_{it} and $m(Z_{it}, Z_{i(t-1)})$, respectively. This is similar to ΔX and Δv.

Multiplying both sides by $\overline{P} = P(P^\top P)^{-1}P^\top$ and subtracting the resulting expression from (5.24),

$$\Delta Y - \Delta\widetilde{Y} = (\Delta X - \Delta\widetilde{X})\beta + (M - \widetilde{M}) + (\Delta v - \Delta\widetilde{v}), \tag{5.25}$$

where $\Delta\widetilde{Y} = \overline{P}\Delta Y = P\gamma_{\Delta Y}$, and $\gamma_{\Delta Y} = (P^\top P)^{-1}P^\top\Delta Y$. This definition is similar for $\widetilde{M}$, $\Delta\widetilde{X}$ and $\Delta\widetilde{v}$.

Thus, the least-squares estimator for β is defined as

$$\widehat{\beta} = \left((X - \widetilde{X})^\top(X - \widetilde{X})\right)^{-1}(X - \widetilde{X})^\top(Y - \widetilde{Y}), \tag{5.26}$$

whereas for the smooth function $m(z)$ they propose $\widehat{m}(z) = \rho^L(z)^\top\widehat{\gamma}$, where

$$\widehat{\gamma} = (P^\top P)^{-1}P^\top(Y - X\widehat{\beta}). \tag{5.27}$$

Under standard conditions of the series approach, in Baltagi and Li (2002) it is established that defining $\xi_{it} = X_{it} - W(Z_{it})$, where $W(Z_{it}) = E_{\mathcal{M}}(\theta(Z_{it}))$,

$$\sqrt{N}(\widehat{\beta} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Omega\Phi^{-1}\right), \tag{5.28}$$

where $\Phi = T^{-1}\sum_{t=1}^{T} E(\xi_{it}\xi_{it}^\top)$ and $\Omega = T^{-1}\sum_{t=1}^{T} E\left(\sigma^2_{\Delta v}(X_{it}, Z_{it})\xi_{it}\xi_{it}^\top\right)$, being $\sigma^2_{\Delta v}(X_{it}, Z_{it}) = E\left(\Delta v_{it}^2 | X_{it} = x, Z_{it} = z\right)$.

We refer to the appendix in Baltagi and Li (2002) for the proofs of these results. Finally, note that they extend these results to the situation in which endogenous explanatory variables are allowed. As we will see in Section 6, they use a semi-parametric regression model with instrumental variables to avoid the endogeneity problem and to provide consistent estimators.

An alternative approach to Baltagi and Li (2002) can be found in Qian and Wang (2012). In this paper, they propose an estimator for the nonparametric component, $m(\cdot)$, that does not suffer from the curse of dimensionality. Let $\Delta Y_{it}^* = \Delta Y_{it} - \Delta X_{it}^\top\widehat{\beta}$, where

$$\widehat{\beta} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta\widehat{X}_{it}\Delta\widehat{X}_{it}^\top\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta\widehat{X}_{it}\Delta\widehat{Y}_{it}, \tag{5.29}$$

where $\widehat{X}_{it} = X_{it} - \widehat{E}\left(X_{it}|\, Z_{it}, Z_{i(t-1)}\right)$ and $\widehat{Y}_{it} = Y_{it} - \widehat{E}\left(Y_{it}|\, Z_{it}, Z_{i(t-1)}\right)$ and

$$\widehat{E}\left(X_{it}|Z_{it}, Z_{i(t-1)}\right) = \frac{1}{NTh^q}\sum_{j=1}^{N}\sum_{s=1}^{T} X_{js}K_h(Z_{it} - Z_{js})K_h(Z_{i(t-1)} - Z_{js})/\widehat{f}_h\left(Z_{it}, Z_{i(t-1)}\right), \tag{5.30}$$

$$\widehat{E}\left(Y_{it}|Z_{it}, Z_{i(t-1)}\right) = \frac{1}{NTh^q}\sum_{j=1}^{N}\sum_{s=1}^{T} Y_{js}K_h(Z_{it} - Z_{js})K_h(Z_{i(t-1)} - Z_{js})/\widehat{f}_h\left(Z_{it}, Z_{i(t-1)}\right), \tag{5.31}$$

$$\widehat{f}_h(Z_{it}, Z_{i(t-1)}) = \frac{1}{NTh^q}\sum_{j=1}^{N}\sum_{s=1}^{T} K_h(Z_{it} - Z_{js})K_h(Z_{i(t-1)} - Z_{js}). \tag{5.32}$$

Then, (4.1) can be written as

$$\Delta Y_{it}^* = m(Z_{it}, Z_{i(t-1)}) + \Delta v_{it}^*, \qquad i = 1, \cdots, N \qquad ; \qquad t = 2, \cdots, T. \tag{5.33}$$

where $\Delta v_{it}^* = \Delta v_{it} - (\widehat{\beta} - \beta)^\top \Delta X_{it}$. However, as it has been pointed out before, the estimation of $m(\cdot)$ is cumbersome due to the fact that $m(Z_{it}, Z_{i(t-1)})$ is an additive function.

In Qian and Wang (2012) it is proposed a non-iterative method based on the marginal integration technique. More precisely, they develop a two step procedure in which they first use conventional multivariate nonparametric techniques such as the Nadaraya-Watson or the local linear regression, and later the function $m(\cdot)$ is obtained through the marginal integration of the previous estimator. Thus, using the local linear regression procedure to estimate $m(Z_{it}, Z_{i(t-1)})$, Qian and Wang (2012) propose to solve the following locally weighted linear least-squares problem for α,

$$\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\Delta Y_{it}^* - \alpha - (Z_{it} - z_1)^\top \gamma_0 - (Z_{i(t-1)} - z_2)^\top \gamma_1\right)^2 K_H(Z_{it} - z_1) K_H(Z_{i(t-1)} - z_2), \tag{5.34}$$

where z_1 and z_2 are points in the interior of the support of $f(\cdot)$.

Let $\widehat{\alpha}$ be a minimizer of (5.34), the estimator for $m(Z_{it}, Z_{i(t-1)})$ is of the form

$$\widehat{m}(z_1, z_2; H) = \widehat{\alpha} = e_1^\top (Z_z^\top K_z Z_z)^{-1} Z_z^\top K_z \Delta Y^*, \tag{5.35}$$

where now K_z and Z_z are $N(T-1) \times N(T-1)$ and $N(T-1) \times (1+2q)$ matrices, respectively, of the following form

$$K_z = diag\left(K_H(Z_{12} - z_1) K_H(Z_{11} - z_2), \cdots, K_H(Z_{NT} - z_1) K_H(Z_{N(T-1)} - z_2)\right)$$

and

$$Z_z = \begin{bmatrix} 1 & (Z_{12} - z_1)^\top & (Z_{11} - z_2)^\top \\ \vdots & \vdots & \vdots \\ 1 & (Z_{NT} - z_1)^\top & (Z_{N(T-1)} - z_2)^\top \end{bmatrix}.$$

Note that if our interest is the estimation of the partial derivatives of $m(\cdot)$, i.e. $D_{m_1}(z) = vec(\partial m(z_1, z_2)/\partial z_1^\top)$ and $D_{m_2}(z) = vec(\partial m(z_1, z_2)/\partial z_2^\top)$, it would be enough to minimize (5.34) for γ_0 and γ_1. Thus, we could propose as estimators for γ_0 and γ_1, $vec(\widehat{D}_{m_1}(z_1; H)) = \widehat{\gamma}_0$ and $vec(\widehat{D}_{m_2}(z_2; H)) = \widehat{\gamma}_1$, respectively. However, since the objective of these authors is to provide an estimator for the unknown function, $m(Z_{it})$, they propose to integrate marginally the estimated function $\widehat{m}(z_1, z_2)$, i.e.,

$$\widehat{m}(z_1; H) = \int \widehat{m}(z_1, z_2) q(z_2) dz_2, \tag{5.36}$$

where $q(\cdot)$ is a predetermined density function.

With the aim of avoiding strict usual identification restrictions of the marginal integration technique, such as the assumption $\int m(z_1) q(z_1) dz_1 = 0$ proposed in Hengartner and Sperlich (2005),

or numerical integration methods such as Simpson's or Trapezoidal rules, Qian and Wang (2012) develop an alternative strategy. In particular, they propose to generate *i.i.d.* samples of the $q(\cdot)$ distribution as Z_k^*, for $k = 1, \cdots, NT$, and compute

$$\widehat{m}_{MC}(z_1; H) = \frac{1}{NT} \sum_{k=1}^{N} \widehat{m}(z_1, z_k^*). \tag{5.37}$$

As it is emphasized in Qian and Wang (2012), if NT is large enough $\widehat{m}_{MC}(\cdot)$ approximates considerably well to $\widehat{m}(\cdot)$ and we choose $q(\cdot)$ to be the density function of Z_{it}, the sample version of (5.35) can be used rather than (5.36), i.e.,

$$\widehat{m}_S(z_1; H) = \frac{1}{N(T-1)} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{m}(z_1, Z_{it}). \tag{5.38}$$

Under standard conditions of the marginal integration technique, these authors show that the nonparametric estimator (5.38) behaves asymptotically equal to (5.36) when $q(\cdot)$ is the density function of Z_{it}, bounded and twice differentiable and when it satisfies $\int m(z_1)q(z_1) = 0$. Thus, they obtain

$$N^{2/(4+q)}(\widehat{m}_S(z_1; H) - m(z_1) - B(z_1)) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, V(z_1)\right), \tag{5.39}$$

where

$$\begin{aligned} B(z_1) &= \frac{1}{2}\mu_2(K)\left(tr(H\mathcal{H}_m(z_1)) - \int tr(H\mathcal{H}_f(z_2))q(z_2)dz_2\right), \\ V(z_1) &= \frac{\overline{\sigma}^2 R^q(K)}{T|H|^{1/2}}\left(\int \frac{q^2(z_2)}{f_2(z_1, z_2)} q(z_2)dz_2\right), \end{aligned}$$

$\overline{\sigma}^2 = T^{-1}\sum_{t=2}^{T}\sigma_t^2$ and $\mathcal{H}_m(z_1)$ is the Hessian matrix of $m(\cdot)$ evaluated at z_1.

Analyzing in detail these asymptotic results, in Qian and Wang (2012) it is pointed out that if Z_{it} is *i.i.d.* across t, as well as for i, and $q(\cdot) = f(\cdot)$, the asymptotic variance takes the conventional form $\frac{\overline{\sigma}^2 R^q(K)}{T|H|^{1/2}} f(z_1)^{-1}$. In addition, when Z_{it} is accurately predictable by $Z_{i(t-1)}$ the conditional density function $f(z_1|z_2)$ is close to zero, except in a small neighborhood of z_2, and this method can fail. Finally, note that if $\widehat{m}(z_1, z_2)$ is estimated using the Nadaraya-Watson kernel smoothing the asymptotic variance remains without change but the asymptotic bias is different. See Qian and Wang (2012) for further details.

However, note that despite the great advantages offered for the empirical analysis by the procedures above, the dimensionality problem characteristic of the nonparametric models is unsolved. As we have stated previously, when the dimension of the nonparametric component is large we have to deal with the curse of dimensionality. In order to avoid the slower rates of convergence of these nonparametric estimators, a possible solution is to analyze an additive alternative expression for $m(\cdot)$. More precisely, substituting (3.37) into (4.1) we obtain,

$$Y_{it} = X_{it}^\top \beta + m_1(Z_{1it}) + \cdots + m_q(Z_{qit}) + \mu_i + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{5.40}$$

where now $m(\cdot) = (m_1(\cdot), \cdots, m_q(\cdot))$ is a vector of unknown functions to estimate and the remaining components are defined as in (4.1).

In this context, in Ai et al. (2014) it is proposed to combine the polynomial spline series approximation with the profile least-squares procedure to obtain a semi-parametric least-squares dummy variables (SLSDV) estimator for the parametric component, and a series estimator for the nonparametric term. Under very weak conditions, these authors show that the semi-parametric least-squares dummy variables estimator is asymptotically normal and the series estimator achieves the optimal rate of convergence of the nonparametric regression. Later, with the aim of obtaining estimators that exhibit the oracle efficiency property, a two step local polynomial procedure is developed based on a series method that makes it possible to impose the additive structure of the $m(\cdot)$ function. Since the nonparametric smoothing spline technique is beyond the scope of this study, we refer to Ai et al. (2014) for a detailed analysis of the proposed procedure and the study of the main asymptotic properties of the resulting estimators.

6 Semi-parametric panel data models with endogeneity

As we have already remarked in the fully nonparametric case, there exists many applied problems where it is necessary to include lagged dependent variables as explanatory variables. Furthermore, the presence of endogeneity is also frequent in applied econometrics. In order to solve these problems in most part of cases it is common to use instrumental variables techniques. For example, in Li and Stengos (1996) and Baltagi and Li (2002) it is considered the estimation of partially linear dynamic panel data models using instrumental variables (IV) methods.

6.1 Endogenous partially linear panel data models with random effects

Consider the partially linear panel data model introduced in Section 4 with the random effects specification. Instead of assuming $E\left(\upsilon_{it}|\, X_{it}, Z_{it}, \mu_i\right) = 0$, we are willing to assume only $E\left(\upsilon_{it}|\, Z_{it}, \mu_i\right) = 0$. In this context, in Li and Stengos (1996) it is developed an IV technique that follows the proposal in Robinson (1988). Thus, these authors use a kernel estimation method with the aim of removing $m(\cdot)$ before proposing an estimator for β. Taking conditional expectations given Z_{it} in both sides of (4.1) and subtracting the resulting expression from (4.1), the regression model to estimate is

$$Y_{it} - E(Y_{it}|Z_{it}) = \left(X_{it} - E(X_{it}|Z_{it})\right)^\top \beta + \epsilon_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T \tag{6.1}$$

and assuming there is a vector of instruments, $W_{it} \in I\!R^d$, which holds $E(\epsilon_{it}|Z_{it}, W_{it}) = 0$ and $E(X_{it}W_{it}^\top) \neq 0$, the endogeneity problem can be avoided using the IV approach, i.e.,

$$\widehat{\beta}_{IV} = (\widetilde{W}^\top \widetilde{X})^{-1} \widetilde{W}^\top \widetilde{Y}, \tag{6.2}$$

where $\widetilde{W}$ and $\widetilde{X}$ are $NT \times d$ matrices whose typical row element is $\widetilde{W}_{it} = W_{it} - E(W_{it}|Z_{it})$ and $\widetilde{X}_{it} = X_{it} - E(X_{it}|Z_{it})$, respectively, whereas $\widetilde{Y}$ is a NT-dimensional vector whose typical row element is $\widetilde{Y}_{it} = Y_{it} - E(Y_{it}|Z_{it})$.

However, as in Li and Ullah (1998), the conditional expectations $E(W_{it}|Z_{it})$, $E(X_{it}|Z_{it})$, and $E(Y_{it}|Z_{it})$ are some unknown terms that can be replaced with their consistent estimators, i.e., $\widehat{W}_{it}$, $\widehat{X}_{it}$, and $\widehat{Y}_{it}$, respectively, to obtain feasible IV estimators for the parametric component of (4.1). Thus, in Li and Stengos (1996) it is proposed the following feasible IV estimator

$$\widetilde{\beta}_{FIV} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(W_{it} - \widehat{W}_{it} \right) \left(X_{it} - \widehat{X}_{it} \right)^{\top} \right) \sum_{i=1}^{N} \sum_{t=1}^{T} \left(W_{it} - \widehat{W}_{it} \right) (Y_{it} - \widehat{Y}_{it}), \tag{6.3}$$

where

$$\widehat{W}_{it} = \widehat{E}(W_{it}|Z_{it}) = \frac{1}{NTh^q} \sum_{j=1}^{N} \sum_{s=1}^{T} W_{js} K_h(Z_{it} - Z_{js}) / \widehat{f}_h(Z_{it})$$

and $\widehat{X}_{it}$, $\widehat{Y}_{it}$ and $\widehat{f}_h(Z_{it})$ are defined as in (4.10), (4.11) and (4.12), respectively.

As it has been pointed out in Section 4, in order to avoid the technical difficulties owning to the random denominator, $\widehat{f}_{it}$, we trim out again small values of $\widehat{f}_{it}$. Then, for a constant $b > 0$, we define $\widehat{I}_{it} = \mathbf{1}(\left|\widehat{f}_i\right| > b)$, where $\mathbf{1}(\cdot)$ is the usual indicator function. Therefore, the feasible IV estimator for β can be written as

$$\widehat{\beta}_{FIV} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \left(W_{it} - \widehat{W}_{it} \right) \left(X_{it} - \widehat{X}_{it} \right)^{\top} \widehat{I}_{it} \right) \sum_{i=1}^{N} \sum_{t=1}^{T} \left(W_{it} - \widehat{W}_{it} \right) (Y_{it} - \widehat{Y}_{it}) \widehat{I}_{it}. \tag{6.4}$$

Under some standard regularity conditions, these authors provide the following asymptotic distribution for this IV estimator,

$$\sqrt{N}(\widehat{\beta}_{FIV} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Phi^{-1}\Gamma\Phi^{-1}\right), \tag{6.5}$$

where $\Phi = T^{-1} \sum_{t=1}^{T} E\left(\widetilde{W}_{it}\widetilde{X}_{it}^{\top} f_{it}^2\right)$ and $\Gamma = T^{-2} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left(\epsilon_{it}\epsilon_{is}\widetilde{W}_{it}\widetilde{W}_{is}^{\top} f_{it}^2 f_{is}^2\right)$.

Similarly, following this procedure and allowing for the presence of lagged dependent variables in the vector X_{it}, in Baltagi and Li (2002) it is proposed an alternative estimator that pays special attention to the choice of the instruments in order to avoid the existence of weak instrumental variables. Finally, note that in both studies the authors leave the estimation of the nonparametric component for future research.

6.2 Endogenous partially linear panel data models with fixed effects

Considering a partially linear model as in (4.1) that fulfills all conditions established in Section 5. Furthermore, as in Section 6.1, instead of assuming $E(v_{it}|X_{it}, Z_{it}, \mu_i) = 0$, we are willing to assume only $E(v_{it}|Z_{it}, \mu_i) = 0$. Then, in order to avoid the incidental parameters problem, first differences

are used in (4.1). Taking as a benchmark the technique developed in Li and Stengos (1996), in Qian and Wang (2012) it is proposed to estimate the linear component β in a regression model such as

$$\Delta Y_{it} - E(\Delta Y_{it}|Z_{it}, Z_{i(t-1)}) = \left(\Delta X_{it} - E(\Delta X_{it}|Z_{it}, Z_{i(t-1)})\right)^\top \beta + \Delta v_{it}. \tag{6.6}$$

Assuming there is a vector of instruments, $W_{it} \in I\!R^d$, and replacing the unknown parameters $E(\Delta Y_{it}|Z_{it}, Z_{i(t-1)})$, $E(\Delta X_{it}|Z_{it}, Z_{i(t-1)})$ and $E(\Delta W_{it}|Z_{it}, Z_{i(t-1)})$ by their consistent estimators, the feasible IV estimator is of the form

$$\widehat{\beta}_{FDIV} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta W_{it} - \Delta\widehat{W}_{it})(\Delta X_{it} - \Delta\widehat{X}_{it})^\top \widehat{I}_{it}\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T}(\Delta W_{it} - \Delta\widehat{W}_{it})(\Delta Y_{it} - \Delta\widehat{Y}_{it})\widehat{I}_{it}, \tag{6.7}$$

where $\widehat{I}_{it}$ is defined as in (4.14),

$$\begin{aligned} \Delta\widehat{Y}_{it} &= \widehat{E}(\Delta Y_{it}|Z_{it}, Z_{i(t-1)}) \\ &= \frac{1}{NTh^{2q}}\sum_{j=1}^{N}\sum_{s=1}^{T}\Delta Y_{js}K_h(Z_{it} - Z_{js})K_h(Z_{i(t-1)} - Z_{j(s-1)})/\widehat{f}_h(Z_{it}, Z_{i(t-1)}), \quad (6.8) \\ \widehat{f}_h(Z_{it}, Z_{i(t-1)}) &= \frac{1}{NTh^{2q}}\sum_{j=1}^{N}\sum_{s=1}^{T}K_h(Z_{it} - Z_{js})K_h(Z_{i(t-1)} - Z_{j(s-1)}), \quad (6.9) \end{aligned}$$

and $\Delta\widehat{X}_{it}$ and $\Delta\widehat{W}_{it}$ are defined as in (6.8).

Note that this technique makes it possible to avoid the random denominator problem usual in the nonparametric estimation of the regression model, but at the cost of having to define the resulting estimator on $I\!R^{2q}$ rather than $I\!R^q$. Adapting the assumptions in Li and Stengos (1996) and imposing that $f_h(Z_{it}, Z_{i(t-1)})$ is a bounded density function and at least first-order partially differentiable with a remainder term that is Lipschitz-continuous, in Qian and Wang (2012) it is obtained that as N tends to infinity,

$$\sqrt{N}(\widehat{\beta}_{FDIV} - \beta) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \Psi^{-1}\Gamma\Psi^{-1}\right), \tag{6.10}$$

where $\Psi = T^{-1}\sum_{t=2}^{T} E\left((\Delta W_{it} - E(\Delta W_{it}|Z_{it}, Z_{i(t-1)}))(\Delta X_{it} - E(\Delta X_{it}|Z_{it}, Z_{i(t-1)}))^\top f_{it}^2\right)$ and $\Gamma = T^{-2}\sum_{t=2}^{T}\sum_{s=2}^{T} E\left(\Delta v_{1t}\Delta v_{1s}(\Delta W_{1t} - (\Delta W_{1t}|Z_{it}, Z_{i(t-1)}))(\Delta W_{1t} - (\Delta W_{1s}|Z_{is}, Z_{i(s-1)}))^\top f_{it}^2 f_{is}^2\right)$, let $f_{it} = f_h(Z_{it}, Z_{i(t-1)})$ and $f_{is} = f_h(Z_{is}, Z_{i(s-1)})$.

7 Conclusion

In this paper, we have made an intensive review of the recent developments for semi-parametric and fully nonparametric panel data models that are linearly separable in the innovation and the individual specific term. We have analyzed these developments under two alternative model specifications: Fixed and random effects panel data models. More precisely, in the random effects setting we have focused our attention in the analysis of some efficiency issues that have to do with the so-called

working independence condition. In the fixed effects setting, to cope with the so-called *incidental parameters problem*, we have consider two different estimation approaches: profiling techniques and differencing methods. We have been also interested in the endogeneity problem and in the use of instrumental variables in this setting. In addition, for practitioners, we have also shown different ways of avoiding the so-called *curse of dimensionality* problem in pure nonparametric models. In this way, semi-parametric and additive models appear as a solution when the number of explanatory variables becomes large. Note that Su and Ullah (2011) and Chen et al. (2013) focus on similar models, although in this case we include the most recent results and pay special attention to the so-called *incidental parameters problem* as well as with endogenous explanatory variables.

References

Ai, C. and Q. Li (2008). "Semiparametric and nonparametric methods in panel data models". In L. Mátyás and P. Sevestre (Eds.), *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (3rd ed.)., pp. 451–478. Springer.

Ai, C., J. You, and Y. Zhou (2014). "Estimation of fixed effects panel data partially linear additive regression models". *The Econometrics Journal 17*, 83–106.

Anderson, T. W. and C. Hsiao (1981). "Estimation of dynamic models with error components". *Journal of the American Statistical Association 76*, 598–606.

Arellano, M. (2003). *"Panel data econometrics"*. Oxford University Press.

Baltagi, B. (2013). *"Econometric analysis of panel data"* (2nd ed.). Wiley, New York.

Baltagi, B. H. and Q. Li (2002). "Series estimation of partially linear panel data models with fixed effects". *Annals of Economic and Finance 3*, 103–116.

Chamberlain, G. (1984). "Panel Data". In Z. Griliches and M. D. Intrilligator (Eds.), *Handbook of Econometrics*, Volume 2, pp. 1247–1318. North Holland, Amsterdam.

Chamberlain, G. (1992). "Efficiency Bounds for Semiparametric Regression". *Econometrica 60*, 567–596.

Chen, J., D. Li, and J. Gao (2013). "Non- and semi-parametric panel data models: a selective review". Working Paper 18/13 Monash University.

Chen, X. (2007). "Large sample sieve estimation fo semi-nonparametric models". In J. J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6, pp. 5549–5632. North Holland, Amsterdam.

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

Fan, J. and I. Gijbels (1995). *"Local polynomial modelling and its applications"*. Chapman & Hall.

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

Gao, Y. and K. Li (2013). "Nonparametric estimation of fixed effects panel data models". *Journal of Nonparametric Statistics 25*(3), 679–693.

Hastie, T. and R. Tibshirani (1990). *"Generalized additive models"*. Chapman and Hall, New York.

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

Henderson, D. J. and C. F. Parmeter (2015). *"Applied nonparametric econometrics"*. Cambridge University Press.

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

Hengartner, N. W. and S. Sperlich (2005). "Rate optimal estimation with the integration method in the presence of many covariates". *Journal of Multivariate Analysis 95*, 246–272.

Hsiao, C. (2003). *"Analysis of panel data"* (2nd ed.). Econometric Society Monographs.

Kneisner, T. and Q. Li (1996). "Semiparametric panel data model with dynamic adjustment: theoretical considerations and an application to labor supply". Manuscript, University of Guelph.

Kneisner, T. and Q. Li (2002). "Nonlinearity in dynamic adjustment: semiparametric estimation of panel labor supply". *Empirical Economics 27*, 131–148.

Lancaster, T. (2000). "The incidental parameter problem since 1948". *Journal of Econometrics 95*, 391–413.

Lee, Y. and D. Mukherjee (2008). "New nonparametric estimation of the marginal effects in fixed-effects panel models: an application on the environmental Kuznets curve". Revise and resubmit invited from *Journal of Econometrics*.

Li, G., H. Peng, and T. Tong (2013). "Simultaneous confidence band for nonparametric fixed effects panel data models". *Economics Letters 19*, 229–232.

Li, Q. and T. Stengos (1996). "On the root-n-consistent semiparametric esitmation of partially linear models". *Journal of Econometrics 71*, 389–397.

Li, Q. and A. Ullah (1998). "Estimating partially linear panel data models with one-way error components". *Econometric Reviews 17*(2), 145–166.

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.

Lin, X. and R. J. Carroll (2006). "Semiparametric estimation in general repeated measures problems". *Journal of the Royal Statistical Society 68*(Series B), 68–88.

Lin, Z., Q. Li, and Y. Sun (2014). "A consistent nonparametric test of parametric regression functional form in fixed effects panel data models". *Journal of Econometrics 178*(1), 167–179.

Linton, O. and J. P. Nielsen (1995). "A kernel method of estimating structured nonparametric regression based on marginal integration". *Biometrika 82*, 93–101.

Maddala, G. S. (1987). "Recent developments in the econometrics of panel data analysis". *Transportation Research 21*, 303–326.

Mamen, E., O. Linton, and J. P. Nielsen (1999). "The existence and asymptotic properties of a backfitting projection algorithm under weak conditions". *Annals of Statistics 27*, 1443–1490.

Mammen, E., B. Støve, and D. Tjøstheim (2009). "Nonparametric additive models for panels of time series". *Econometric Theory 25*, 442–481.

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

Mundra, K. (2005). "Nonparametric slope estimation for fixed-effect panel data". Working paper, Department of Economics, San Diego State University.

Nadaraya, E. A. (1964). "On estimating regression". *Theory of Probability and its Applications 9*(1), 141–142.

Neyman, J. and E. Scott (1948). "Consistent estimates based on partially consistent observations". *Econometrica 16*, 1–32.

Powell, J. L., J. H. Stock, and T. M. Stocker (1989). "Semiparametric estimation of the index coefficients". *Econometrica 57*, 1043–1430.

Qian, J. and L. Wang (2012). "Estimating semiparametric panel data models by marginal integration". *Journal of Econometrics 167*, 483–493.

Robinson, P. M. (1988). "Root-n-consistent semiparametric regression". *Econometrica 56*(4), 931–954.

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.

Ruppert, D. and M. P. Wand (1994). "Multivariate locally weighted least squares regression". *The Annals of Statistics 22*(3), 1346–1370.

Speckman, P. (1988). "Kernel smoothing in partial linear models". *Journal of the Royal Statistical Society: Series B 50*, 413–436.

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 (2006a). "More efficient estimation in nonparametric regression with nonparametric autocorrelated errors". *Econometric Theory 22*, 98–126.

Su, L. and A. Ullah (2006b). "Profile likelihood estimation of partially linear panel data models with fixed effects". *Economics Letters 92*, 75–81.

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". In A. Ullah and D. E. Giles (Eds.), *Handbook of Empirical Economics and Finance*, pp. 455–497. New York: Taylor and Francis Group.

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.

Ullah, A. and N. Roy (1998). "Nonparametric and semiparametric econometrics of panel data". In A. Ullah and D. E. Giles (Eds.), *Handbook of Applied Economic Statistics*, pp. 579–604. New York: Marcel Dekker.

Wang, N. (2003). "Marginal nonparametric kernel regression accounting for within-subject correlation". *Biometrika 90*, 43–52.

Watson, G. S. (1964). "Smooth regression analysis". *Sankhya: The Indian Journal of Statistics 26*(4), 359–372.

Wooldridge, J. (2010). *"Econometric analysis of cross section and panel data models"*. The MIT Press, Cambridge, Massachusets.

Xiao, Z., O. B. Linton, and R. J. Mammen (2003). "More efficient local polynomial estimation in nonparametric regression with autocorrelated errors". *Journal of American Statistical Association 98*, 980–992.

You, J., X. Zhou, and Y. Zhou (2010). "Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors". *Journal of Multivariate Analysis 101*, 1079–1101.

Zhan-Qian, L. (1996). "Multivariate locally weighted polynomial fitting and partial derivative estimation". *Journal of Multivariate Analysis 59*, 187–205.

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.