# Differencing techniques in semi-parametric panel data varying coefficient models with fixed effects: a Monte Carlo study

**Juan M. Rodriguez-Poo** · **Alexandra Soberón**

**Abstract** Recently, some new techniques have been proposed for the estimation of semi-parametric fixed effects varying coefficient panel data models. These new techniques fall within the class of the so-called differencing estimators. In particular, we consider first-differences and within local linear regression estimators. Analyzing their asymptotic properties it turns out that, keeping the same order of magnitude for the bias term, these estimators exhibit different asymptotic bounds for the variance. In both cases, the consequences are suboptimal non-parametric rates of convergence. In order to solve this problem, by exploiting the additive structure of this model, a one-step backfitting algorithm is proposed. Under fairly general conditions, it turns out that the resulting estimators show optimal rates of convergence and exhibit the oracle efficiency property. Since both estimators are asymptotically equivalent, it is of interest to analyze their behavior in small sample sizes. In a fully parametric context, it is well-known that, under strict exogeneity assumptions the performance of both first-differences and within estimators is going to depend on the stochastic structure of the idiosyncratic random errors. However, in the non-parametric setting, apart from the previous issues other factors such as dimensionality or sample size are of great interest. In particular, we would be interested in learning about their relative Average Mean Square Error under different scenarios. The simulation results basically confirm the theoretical findings for both local linear regression and one-step backfitting estimators. However, we have found out that within estimators are rather sensitive to the size of number of time observations.

J.M. Rodriguez-Poo
Universidad de Cantabria, Avda de los Castros s/n, E-39005, Santander, Spain
Tel.:+(34)942201640
E-mail: juan.rodriguez@unican.es

A. Soberón
Universidad de Cantabria
Tel.:+(34)942202068
E-mail: alexandra.soberon@unican.es

## 1 Introduction

Since the last ten years, semi-parametric panel data varying coefficient models with fixed effects have become a very useful tool to handle many statistical problems in empirical studies (see for example Card (2001), Kottaridi and Stengos (2010) and Kuan and Chen (2013)). If individual effects are assumed to be uncorrelated with the explanatory variables (random effects), the smooth functions can be estimated turning to standard non-parametric techniques for panel data models. See Su and Ullah (2011) for an intensive review of this literature. Nevertheless, when the cross-sectional heterogeneity is correlated with some covariates, as it is in our case, the problem is more complex and direct estimation with the aforementioned techniques provides estimators that are asymptotically biased. The reason is that the correlation of the heterogeneity term with some of the explanatory variables causes a non-negligible asymptotic bias.

Recently, in order to cope with this problem, some new procedures have been developed within the framework of differencing techniques. By taking differences one can remove the heterogeneity effect and then estimate the resulting model by standard nonparametric techniques. Unfortunately, this is not as easy as it would look at a first glance since the model in differences appears, for each individual, as an additive function with same functional form at different times. That is the main reason why some proposals to estimate this transformed model are closely related to estimation techniques initially designed for additive models. In Henderson et al. (2008) it is developed an iterative procedure based on a maximum likelihood approach, whereas in Mammen et al. (2009) it is proposed a smooth backfitting algorithm (see Buja et al. (1989) for the original idea of the backfitting). Recently, in Qian and Wang (2012) it is developed a two-step procedure in which in the first-step the whole non-parametric term is estimated through a multivariate non-parametric estimator and later the function, in each point, is obtained via marginal integration techniques. Su and Lu (2013) estimate the unknown function as a solution of a second-order Friedholm integral equation. However, despite the great contributions of these techniques they are not very appealing as they are computationally intensive.

In view of these results, in Rodriguez-Poo and Soberon (2013a,b) it is proposed a direct strategy to estimate the unknown varying coefficients that can be applied either in the context of first differences or the within transformation. The basic idea of both estimators is to approximate the unknown additive functions through a local linear regression technique with higher-dimensional kernel weights. Although the proposed estimation strategies enable us to solve the non-negligible asymptotic bias problem that it is usual in differencing estimators, they reflect the standard dilemma of the non-parametric estimates, i.e. any attempt to hold back the bias is offset by an increase of the variance term and therefore the resulting estimators achieve suboptimal rates of convergence. In order to solve this problem, by exploiting the additive structure of this model, a one-step backfitting algorithm is proposed. Under fairly general conditions, it turns out that the resulting estimators show optimal rates of convergence and exhibit the oracle efficiency property. This is already a well-known result (see Fan and Zhang (1999)): Additional smoothing can reduce the variance without affecting the asymptotic order of the bias.

Since both estimators are asymptotically equivalent, it is of interest to analyze their behavior in small sample sizes under a standard panel data setting, that is, fixed number of time observations and increasing number of individuals. In a fully parametric context, it is well-known that, under strict exogeneity assumptions the performance of both differencing estimators is going to depend on the stochastic structure of the idiosyncratic random errors (see Wooldridge (2002)). Following these ideas, in this paper, we perform a Monte Carlo simulation experiment that is designed to compare the performance of both the first-differences and the within estimator in finite samples under fairly standard conditions. Moreover, in the non-parametric setting, apart from the previous issues other factors such as the number of covariates, the number of time periods, and more importantly the number of individuals, are of great interest. In particular, we would be interested in learning whether, for different scenaria of the idiosyncratic error terms, which estimator is more efficient.

Finally, note that differencing techniques are not always suitable to remove heterogeneity individual effects. For example, in nonseparable panel data models, other estimation strategies need to be undertaken. Altonji and Matzkin (2005), Bester and Hansen (2009), and Hoderlein and White (2012) focus on both the estimation of the structural functional itself or the local average derivatives. They also analyze some identification conditions.

The rest of the paper is organized as follows. In Section 2, we present the local linear estimation procedures for both differencing estimators and we also study their asymptotic properties. As it has been already pointed out before, the estimators at this stage present different and, in both cases, suboptimal rates of convergence. In Section 3, we apply a one-step backfitting algorithm to both estimators allowing them to achieve asymptotically optimal rates. In Section 4, we compare the estimators considered via a Monte Carlo simulation. Finally, we conclude in Section 5. All assumptions and technicalities required to show the asymptotic behavior of the estimators are relegated to the Appendix.

## 2 Local linear estimation procedure

To illustrate the estimation procedures to be compared in this paper we consider a panel data varying coefficient model where some regression coefficients are allowed to be varying depending on some exogenous variables of the form

$$Y_{it} = X_{it}^\top m(Z_{it}) + \mu_i + v_{it}, \qquad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \tag{2.1}$$

where $X$ and $Z$ are $d \times 1$ and $q \times 1$ vector of covariates, respectively, $m(Z)$ is a $d \times 1$ vector of smooth functions to be estimated, $\mu_i$ is the unobserved individual heterogeneity and $v_{it}$ the random disturbance. We assume $\mu_i$ is correlated with $X$ and/or $Z$ with an unknown correlation structure.

As it was already pointed out in the previous section, direct estimation of the unknown function $m(\cdot)$ through nonparametric standard techniques rends asymptotically biased estimators. One solution to this problem are the so called differencing techniques. Among the most popular transformations we consider first differences and the within transformation. Other estimation strategies such as profile likelihood (least squares) techniques are available (see Sun et al. (2009)). However, conditions for consistency of the parameters (functions) of interest are rather strong and this is why we focus on estimators based on differencing transformations.

The first differences transformation implies subtracting from time $t$ of (2.1) that of time $t-1$, i.e.

$$Y_{it} - Y_{i(t-1)} = X_{it}^\top m(Z_{it}) - X_{i(t-1)}^\top m(Z_{i(t-1)}) + v_{it} - v_{i(t-1)}, \qquad i = 1, \cdots, N \quad ; \quad t = 2, \cdots, T, \tag{2.2}$$

whereas the within transformation implies subtracting the within-group mean, i.e.

$$Y_{it} - \frac{1}{T} \sum_{s=1}^{T} Y_{is} = X_{it}^\top m(Z_{it}) - \frac{1}{T} \sum_{s=1}^{T} X_{is}^\top m(Z_{is}) + v_{it} - \frac{1}{T} \sum_{s=1}^{T} v_{is}, \qquad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T. \tag{2.3}$$

However, as we have stated previously, direct non-parametric estimation of $m(\cdot)$ in both (2.2) and (2.3) has been considered as a cumbersome task (see Su and Ullah (2011)). The reason is that, for each individual, the right part of both specifications are linear combinations of $X_{it}^\top m(Z_{it})$ for different time periods $t$. Therefore, to estimate the unknown function it is necessary to consider $m(\cdot)$ as an additive function whose elements share the same functional form.

Just to clarify our proposal to solve the previous problem, we will first focus on the univariate regression model and later we will extend these results to the multivariate case. Consider the first differences transformation in (2.2) with $d = q = 1$. In this case, for any $z \in A$, where $A$ is a compact subsect in a nonempty interior of $\mathbb{R}$, one has the following Taylor expansion

$$X_{it} m(Z_{it}) - X_{i(t-1)} m(Z_{i(t-1)}) \approx m(z) \Delta X_{it} + m'(z) \left[ X_{it}(Z_{it} - z) - X_{i(t-1)} m(Z_{i(t-1)}) \right]$$
$$+ \frac{1}{2} \left[ X_{it}(Z_{it} - z)^2 - X_{i(t-1)}(Z_{i(t-1)} - z)^2 \right] + \cdots + \frac{1}{p!} m^{(p)}(z) \left[ X_{it}(Z_{it} - z)^p - X_{i(t-1)}(Z_{i(t-1)} - z)^p \right]$$
$$\equiv \sum_{\lambda=0}^{p} \beta_{F_\lambda} \left[ X_{it}(Z_{it} - z)^\lambda - X_{i(t-1)}(Z_{i(t-1)} - z)^\lambda \right]. \tag{2.4}$$

Similarly, for the within regression in (2.3), one has the following

$$X_{it} m(Z_{it}) - \frac{1}{T} \sum_{s=1}^{T} X_{is} m(Z_{is}) \approx \left( X_{it} - \frac{1}{T} \sum_{s=1}^{T} X_{is} \right) m(z) + \left[ X_{it}(Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is} m(Z_{is}) \right] m'(z)$$
$$+ \frac{1}{2} \left[ X_{it}(Z_{it} - z)^2 - \frac{1}{T} \sum_{s=1}^{T} X_{is}(Z_{is} - z)^2 \right] m''(z) + \cdots + \frac{1}{p!} \left[ X_{it}(Z_{it} - z)^p - \frac{1}{T} \sum_{s=1}^{T} X_{is}(Z_{is} - z)^p \right] m^{(p)}(z)$$
$$\equiv \sum_{\lambda=0}^{p} \beta_{W_\lambda} \left[ X_{it}(Z_{it} - z)^\lambda - \frac{1}{T} \sum_{s=1}^{T} X_{is}(Z_{is} - z)^\lambda \right]. \tag{2.5}$$

Both expressions (2.4) and (2.5) suggest that we can estimate $m(z), m'(z), \cdots, m^{(p)}(z)$ by regressing, respectively, $\Delta Y_{it}$ on the terms $X_{it}(Z_{it} - z)^\lambda - X_{i(t-1)}(Z_{i(t-1)} - z)^\lambda$ and $\ddot{Y}_i = Y_{it} - \frac{1}{T} \sum_{s=1}^{T} Y_{is}$ on $X_{it}(Z_{it} - z)^\lambda - \frac{1}{T} \sum_{s=1}^{T} X_{is}(Z_{is} - z)^\lambda$, for $\lambda = 1, \cdots, p$, with different kernel weights. The quantities of interest in both cases can be estimated using locally weighted linear regression (see Fan and Gijbels (1995b)). For (2.4) we have

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left( \Delta Y_{it} - \beta_{F_0} \Delta X_{it} - \beta_{F_1} \left[ X_{it}(Z_{it} - z) - X_{i(t-1)}(Z_{i(t-1)} - z) \right] \right)^2 K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z), \tag{2.6}$$

and for (2.5) we have

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it}-\beta_{w_0}\left(X_{it}-\frac{1}{T}\sum_{s=1}^{T}X_{is}\right)-\beta_{w_1}\left[X_{it}\left(Z_{it}-z\right)-\frac{1}{T}\sum_{s=1}^{T}X_{is}\left(Z_{is}-z\right)\right]\right)^2 K_h\left(Z_{i1}-z\right)\cdots K_h\left(Z_{iT}-z\right), \quad (2.7)$$

where $h$ is a bandwidth and $K$ is an univariate kernel such that

$$\int K(u)du=1 \qquad \text{and} \qquad K_h(u)=\frac{1}{h}K\left(u/h\right).$$

Let us denote by $\widehat{\beta}_{F_0}$ and $\widehat{\beta}_{F_1}$ the minimizers of (2.6) and by $\widehat{\beta}_{w_0}$ and $\widehat{\beta}_{w_1}$ the minimizers of (2.7). The above exposition suggests as estimators in the first differences case for $m(\cdot)$ and $m'(\cdot)$, $\widehat{m}_h(z)=\widehat{\beta}_{F_0}$ and $\widehat{m}'_h(z)=\widehat{\beta}_{F_1}$, respectively. For the within regression the estimators will be $\widehat{m}_h(z)=\widehat{\beta}_{w_0}$ and $\widehat{m}'_h(z)=\widehat{\beta}_{w_1}$, respectively.

Note that in (2.6) we propose a bivariate kernel that also contains $Z_{i(t-1)}$ instead of considering only $Z_{it}$. The reason is that, if we consider only a kernel around $Z_{it}$, the transformed regression equation (2.2) would be originally localized around $Z_{it}$ without considering all other values. Consequently, the distance between $Z_{is}$ (for $s\neq t$) and $z$ cannot be controlled by the fixed bandwidth parameter and so that the transformed remainder terms cannot be negligible. The consequence of all that would be a non-degenerated bias in this type of local linear estimator that it is removed by considering a local approximation around the pair $\left(Z_{it},Z_{i(t-1)}\right)$. The same can be said for (2.7). Although there, the non-degenerated bias must be removed by considering a local approximation around the $T\times 1$ vector $\left(Z_{i1},\cdots,Z_{iT}\right)$. The difference in the local approximation makes a substantial difference in terms of the asymptotic variance in both estimators. In fact, in Theorems 1 and 2 it is shown that under similar conditions the order of the bias for the univariate case will be the same, $O(h^2)$, but the variance is for $T>1$ rather different. For the first-differences estimator the variance is of order $O(1/NTh^2)$, whereas for the other estimator is of order $O(1/NTh^T)$.

For $d\neq q\neq 1$, the estimators have the following form. Denote by $\widehat{\beta}_F=\left(\widehat{\beta}_{F_0}^{\top}\ \widehat{\beta}_{F_1}^{\top}\right)$ a $d(1+q)$-vector that minimizes (2.6) in the multivariate case, that is

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Delta Y_{it}-\Delta X_{it}^{\top}\beta_{F_0}-\left[X_{it}\otimes\left(Z_{it}-z\right)-X_{i(t-1)}\otimes\left(Z_{i(t-1)}-z\right)\right]^{\top}\beta_{F_1}\right)^2 K_H\left(Z_{it}-z\right)K_H\left(Z_{i(t-1)}-z\right), \quad (2.8)$$

where $H$ is a $q\times q$ symmetric positive definite bandwidth matrix and $K$ is a $q$-variate kernel.

Let $D(z)=vec(D_m(z))$ be a $dq\times q$ vector and let $D_m(z)=\partial m(z)/\partial z$ be a $d\times q$ matrix of partial derivatives of the $d$th component of $m(z)$ with respect to the elements of the $q\times 1$ vector $z$. Denote $\mathscr{H}_m(z)=\partial m(z)/\partial z\partial z^{\top}$ a $dq\times d$ matrix of the Hessian matrix of the $d$th component of $m(z)$. We suggest as estimators of $m(z)$ and $D_m(z)$, $\widehat{m}_F(z;H)=\widehat{\beta}_{F_0}$ and $vec(\widehat{D}_{F_m}(z)(z;H))=\widehat{\beta}_{F_1}$, respectively. Assuming that $\widetilde{Z}_F^{\top}W_F\widetilde{Z}_F$ is nonsingular, the minimization problem (2.8) has the following solution in matrix form

$$\begin{pmatrix}\widehat{\beta}_{F_0}\\\widehat{\beta}_{F_1}\end{pmatrix}=\left(\widetilde{Z}_F^{\top}W_F\widetilde{Z}_F\right)^{-1}\widetilde{Z}_F^{\top}W_F\Delta Y, \quad (2.9)$$

where $\Delta Y=\left(\Delta Y_{12},\cdots,\Delta Y_{NT}\right)^{\top}$ is a $N(T-1)$ vector while $W_F$ and $\widetilde{Z}_F$ are $N(T-1)\times N(T-1)$ and $N(T-1)\times d(1+q)$ matrix, respectively, of the form

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

and

$$\widetilde{Z}_F=\begin{bmatrix}\Delta X_{12}^{\top} & X_{12}^{\top}\otimes\left(Z_{12}-z\right)^{\top}-X_{11}^{\top}\otimes\left(Z_{11}-z\right)^{\top}\\ \vdots & \vdots\\ \Delta X_{NT}^{\top} & X_{NT}^{\top}\otimes\left(Z_{NT}-z\right)^{\top}-X_{N(T-1)}^{\top}\otimes\left(Z_{N(T-1)}-z\right)^{\top}\end{bmatrix}.$$

Then, the local weighted linear least-squares estimator of $m(z)$ is defined as

$$\widehat{m}_F(z;H)=e_1^{\top}\left(\widetilde{Z}_F^{\top}W_F\widetilde{Z}_F\right)^{-1}\widetilde{Z}_F^{\top}W_F\Delta Y, \quad (2.10)$$

where $e_1 = (I_d \vdots 0_{dq \times d})$ is a $d(1+q) \times d$ selection matrix, $I_d$ is a $d \times d$ identity matrix and $0_{dq \times d}$ a $dq \times d$ matrix of zeros.

We focus now on the within estimator. Let $\widehat{\beta}_w = (\widehat{\beta}_{w_0}^\top \ \widehat{\beta}_{w_1}^\top)^\top$ be a $d(1+q)$- vector that minimizes the expression (2.5) in the multivariate case, i.e.,

$$\prod_{\ell=1}^{T} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( \ddot{Y}_{it} - \left( X_{it} - \frac{1}{T} \sum_{s=1}^{T} X_{is} \right)^\top \beta_{w_0} - \left[ X_{it} \otimes (Z_{it} - z) - \frac{1}{T} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z) \right]^\top \beta_{w_1} \right)^2 K_H(Z_{i\ell} - z), \quad (2.11)$$

where now $K$ is the product of univariate kernels such that $K(u_1, u_2, \cdots, u_T) = \prod_{\ell=1}^{T} K(u_\ell)$ and $u_\ell$ is the $\ell$th component of $u$. We suggest as estimators for $m(z)$ and $D_m(z)$, $\widehat{m}_w(z; H) = \widehat{\beta}_{w_0}$ and $vec(\widehat{D}_{w_m}(z; H)) = \widehat{\beta}_{w_1}$, respectively. Thus, assuming $\widetilde{Z}_w^\top W_w \widetilde{Z}_w$ is nonsingular, the matrix form of the solution of the minimization problem (2.11) can be written as

$$\begin{pmatrix} \widehat{\beta}_{w_0} \\ \widehat{\beta}_{w_1} \end{pmatrix} = \left( \widetilde{Z}_w^\top W_w \widetilde{Z}_w \right)^{-1} \widetilde{Z}_w^\top W_w \ddot{Y}, \quad (2.12)$$

where $\ddot{Y} = (\ddot{Y}_{11}, \cdots, \ddot{Y}_{NT})^\top$ is a $NT$-vector while $W_w$ and $\widetilde{Z}_w$ is a $NT \times NT$ and $NT \times d(1+q)$ matrix, respectively, such that

$$W_w = blockdiag \left( K_H(Z_{i1} - z) \prod_{\ell=2}^{T} K_H(Z_{i\ell} - z), \cdots, K_H(Z_{iT} - z) \prod_{\ell=1}^{T-1} K_H(Z_{i\ell} - z) \right)$$

and

$$\widetilde{Z}_w = \begin{bmatrix} \ddot{X}_{11}^\top & X_{11}^\top \otimes (Z_{11} - z)^\top - T^{-1} \sum_{s=1}^{T} X_{1s}^\top \otimes (Z_{1s} - z)^\top \\ \vdots & \vdots \\ \ddot{X}_{NT}^\top & X_{NT}^\top \otimes (Z_{NT} - z)^\top - T^{-1} \sum_{s=1}^{T} X_{Ns}^\top \otimes (Z_{Ns} - z)^\top \end{bmatrix},$$

where $\ddot{X}_{it} = X_{it} - \frac{1}{T} \sum_{s=1}^{T} X_{is}$.

The local weighted linear least-squares estimator of $m(z)$ for the within regression is then defined as

$$\widehat{m}_w(z; H) = e_1^\top \left( \widetilde{Z}_w^\top W_w \widetilde{Z}_w \right)^{-1} \widetilde{Z}_w^\top W_w \ddot{Y}. \quad (2.13)$$

Note that for the sake of simplicity we use the same bandwidth matrix for these two estimators. As it is well-known in the non-parametric literature, the optimal bandwidth matrix $H$ should be obtained using several standard procedures such as, for example, the residual squares criterion proposed in Fan and Gijbels (1995a). Then, for empirical applications we must not forget that although the resulting bandwidths are very close, they are different.

Once obtained the non-parametric estimators for both the first-differences and the within transformation, the next step is to establish the behavior of the two estimators in large samples. Under some standard assumptions collected in the Appendix, their asymptotic distributions are derived in the next theorems. Conditions for the proof are rather general. Assumption **A.1** characterizes the data-generating process for a panel data model. Assumption **A.2** is a standard strict exogeneity condition and **A.3** imposes the so-called fixed effects. In addition, for conditional moments, densities and kernel functions we need some smoothness and boundedness conditions that are collected in assumptions **A.4**-**A.9** and **A.11**-**A.12**. Finally, assumptions **A.10** and **A.13** are required to show that Lyapunov conditions hold.

Let $\mathbb{X} = (X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT})$ be the observed covariates vector and denote $\mathscr{H}_{m_r}(z)$ the Hessian matrix of the $r$th component of $m(\cdot)$, for $r = 1, \cdots, d$, $diag_d(tr(\mathscr{H}_{m_r}(z) H \sqrt{NT |H|}))$ stands for a diagonal matrix of elements $tr(\mathscr{H}_{m_r}(z) H \sqrt{NT |H|})$ and $\iota_d$ a $d \times 1$ unit vector. Furthermore, $R(K) = \int K^2(u) du$, and $f_{Z_{it}, Z_{i(t-1)}}(z, z)$ is the probability density function of the random variable $(Z_{it}, Z_{i(t-1)})$ and we denote by $f_{Z_{i1}, \cdots, Z_{iT}}(z, \cdots, z)$ the probability density functions of $(Z_{i1}, \cdots, Z_{iT})$ evaluated at point $z$.

In this context, in Rodriguez-Poo and Soberon (2013a) it is shown the following result for the locally weighted least-squares first-differences estimator (2.9):

**Theorem 1** *Assume conditions A.1-A.10 hold. Then,*

$$\sqrt{NT\,|H|}\,(\widehat{m}_F\,(z;H) - m\,(z)) \to_d \mathscr{N}\,(b_F(z), \upsilon_F(z)),$$

*as N tends to infinity and T is fixed, where*

$$b_F(z) = \frac{1}{2}\mu_2\,(K)\,diag_d(tr(\mathscr{H}_{mr}(z)H\sqrt{NT\,|H|}))\iota_d,$$
$$\upsilon_F(z) = 2\sigma_v^2 R^2\,(K)\,\mathscr{B}_{\Delta X \Delta X}^{-1}\,(z,z)$$

*and*

$$\mathscr{B}_{\Delta X \Delta X}\,(z,z) = E\left[\Delta X_{it}\Delta X_{it}^\top \middle| Z_{it} = z, Z_{i(t-1)} = z\right] f_{Z_{it},Z_{i(t-1)}}\,(z,z).$$

On the other hand, for the locally weighted least-squares within estimator (2.12), Rodriguez-Poo and Soberon (2013b) obtain the following asymptotic properties:

**Theorem 2** *Assume conditions A.1-A.7 and A.11-A.13 hold, then as $N \to \infty$ and T remains to be fixed we obtain*

$$\sqrt{NT|H|^{T/2}}\,(\widehat{m}_w\,(z;H) - m\,(z)) \to_d \mathscr{N}\,(b_w(z), \upsilon_w(z)),$$

*where*

$$b_w(z) = \frac{1}{2}\mu_2(K)diag_d(tr(\mathscr{H}_{m_r}(z)H\sqrt{NT|H|^{T/2}}))\iota_d,$$
$$\upsilon_w(z) = \sigma_v^2 R^T\,(K)\,\mathscr{B}_{\ddot{X}\ddot{X}}^{-1}(z,\cdots,z)$$

*and*

$$\mathscr{B}_{\ddot{X}\ddot{X}}\,(z,\cdots,z) = E\left[\ddot{X}_{it}\ddot{X}_{it}^\top \middle| Z_{i1} = z,\cdots,Z_{iT} = z\right] f_{Z_{i1},\cdots,Z_{iT}}\,(z,\cdots,z).$$

As we have already pointed out above in Theorems 1 and 2, the use of a higher dimensional kernel weight enables us to solve the problem of non-negligible asymptotic bias. It provides local linear estimators with a bias term of the same order as the standard results, $O(tr(H))$. However, as it is usual in the non-parametric techniques any attempt to reduce the bias is offset by an enlargement of the variance term. Thus, these two estimators are consistent but exhibit a suboptimal rate of convergence. Note that the standard rate of this type of problems is $NT|H|^{1/2}$. The first-differences estimator exhibits a rate of order $NT|H|$ and the within estimator shows a rate of order $NT|H|^{T/2}$.

## 3 One-step backfitting procedure

In this section we analyze alternative procedures to obtain non-parametric estimators that exhibit the optimal rate of convergence. Firstly, we focus on the first differences transformation. Later we present the corresponding within estimator. We conclude with a comparison between the asymptotic properties of the resulting estimators.

As it is noted in Fan and Zhang (1999), the variance can be reduced by further smoothing but the bias cannot be reduced by any kind of smoothing. Thus, in order to achieve optimality we propose to combine previous estimators with a one-step backfitting algorithm. Therefore, this estimation strategy allows us to exploit the additive structure of the model in order to cancel asymptotically the additive terms of the model.

Let $\widehat{m}_F\,(z;H)$ be the first-step local weighted linear least-squares first-differences estimator (2.8) and define the variable $\Delta Y_{it}^b$ such that

$$\Delta Y_{it}^b = \Delta Y_{it} + X_{it}^\top \widehat{m}\,(Z_{i(t-1)};H),\tag{3.1}$$

and replace (2.2) in this previous equation obtaining

$$\Delta Y_{it}^b = X_{it}^\top m\,(Z_{it}) + \Delta v_{it}^b, \quad i = 1,\cdots,N \quad ; \quad t = 2,\cdots,T,\tag{3.2}$$

where the composed error term has the form

$$\Delta v_{it}^b = X_{it}^T \left( \widehat{m} \left( Z_{i(t-1)}; H \right) - m \left( Z_{i(t-1)} \right) \right) + \Delta v_{it}.$$

By the same reasoning as before, the quantities of interest of (3.2) can be estimated as a solution for $\gamma_F$ to the following locally weighted linear regression

$$\sum_{i=1}^{N} \sum_{t=2}^{T} \left( \Delta Y_{it}^b - X_{it}^T \gamma_{F_0} - X_{it}^\top \otimes (Z_{it} - z)^\top \gamma_{F_1} \right)^2 K_{\widetilde{H}} (Z_{it} - z), \qquad (3.3)$$

where $\widetilde{H}$ is a $q \times q$ symmetric positive definite bandwidth matrix of this step. Denote by $\widetilde{\gamma}_F = \left( \widetilde{\gamma}_{F_0}^\top \ \widetilde{\gamma}_{F_1}^\top \right)^\top$ a $d(1+q)$-vector that minimizes the expression (3.3). Then, assuming $\widetilde{Z}_F^{b\top} W_F^b \widetilde{Z}_F^b$ is a nonsingular matrix, we suggest as estimators of $m(z)$ and $D_m(z)$, $\widetilde{m}_F(z; \widetilde{H}) = \widetilde{\gamma}_{F_0}$ and $vec(\widetilde{D}_{F_m}(z; \widetilde{H})) = \widetilde{\gamma}_{F_1}$, respectively, i.e.

$$\widetilde{m}_F(z; \widetilde{H}) = \widetilde{\gamma}_{F_0} = e_1^\top \left( \widetilde{Z}_F^{b\top} W_F^b \widetilde{Z}_F^b \right)^{-1} \widetilde{Z}_F^{b\top} W_F^b \Delta Y^b, \qquad (3.4)$$

where $\Delta Y^b = \left( \Delta Y_{12}^b, \cdots, \Delta Y_{NT}^b \right)^\top$, $W_F^b = diag \left( K_{\widetilde{H}} (Z_{it} - z), \cdots, K_{\widetilde{H}} (Z_{NT} - z) \right)$ and

$$\widetilde{Z}_b = \begin{bmatrix} X_{12}^\top & X_{12}^\top \otimes (Z_{12} - z)^\top \\ \vdots & \vdots \\ X_{NT}^\top & X_{NT}^\top \otimes (Z_{NT} - z)^\top \end{bmatrix}.$$

On the other hand, and following a similar procedure as before, we propose a backfitting estimator for the within transformation such as (2.3). Let $\widehat{m}_w(z; H)$ be the first-step within estimator proposed in (2.11), they define $\ddot{Y}_{it}^b = \ddot{Y}_{it} - T^{-1} \sum_{s=1}^{T} X_{is}^\top \widehat{m}_w(Z_{is}; H)$ and replace $\ddot{Y}_{it}$ by (2.3) obtaining

$$\ddot{Y}_{it}^b = X_{it}^\top m(Z_{it}) + \ddot{v}_{it}^b, \qquad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \qquad (3.5)$$

where the error term is

$$\ddot{v}_{it}^b = \frac{1}{T} \sum_{s=1}^{T} X_{is}^\top \left( \widehat{m}_w(Z_{is}; H) - m(Z_{is}) \right) + \ddot{v}_{it}.$$

Denote by $\widetilde{\gamma}_w = \left( \widetilde{\gamma}_{w_0}^\top \ \widetilde{\gamma}_{w_1}^\top \right)^\top$ the $d(1+q)$-vector that minimizes the following problem

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \left( \ddot{Y}_{it}^b - X_{it}^\top \gamma_{w_0} - X_{it}^\top \otimes (Z_{it} - z)^\top \gamma_{w_1} \right)^2 K_{\widetilde{H}} (Z_{it} - z), \qquad (3.6)$$

we propose as estimator of $m(z)$ and $D_m(z)$, $\widetilde{m}_w(z; \widetilde{H}) = \widetilde{\gamma}_{w_0}$ and $vec(\widetilde{D}_{w_m}(z; \widetilde{H})) = \widetilde{\gamma}_{w_1}$, respectively, of the form

$$\widetilde{m}_w \left( z; \widetilde{H} \right) = \widetilde{\gamma}_{w_1} = e_1^\top \left( \widetilde{Z}_w^b W_w^b \widetilde{Z}_w^b \right)^{-1} \widetilde{Z}_w^{b\top} W_w^b \ddot{Y}^b, \qquad (3.7)$$

where $\ddot{Y}^b = \left( \ddot{Y}_{11}^b, \cdots, \ddot{Y}_{NT}^b \right)^\top$, $W_w^b = diag \left( K_{\widetilde{H}} (Z_{11} - z), \cdots, K_{\widetilde{H}} (Z_{NT} - z) \right)$ and

$$\widetilde{Z}_w^b = \begin{bmatrix} X_{11}^\top & X_{11}^\top \otimes (Z_{11} - z)^\top \\ \vdots & \vdots \\ X_{NT}^\top & X_{NT}^\top \otimes (Z_{NT} - z)^\top \end{bmatrix}.$$

In order to show that these two backfitting estimators achieve optimal rates of convergence and furthermore, they are oracle efficient, we need the sampling scheme conditions established in Assumptions **A.1**-**A.3** and the smoothness and boundedness conditions already considered in Assumptions **A.4**-**A.7**, **A.8**-**A.9** and **A.11**-**A.12**. Furthermore, as

they are obtained via a one-step backfitting algorithm we need to ensure that both bias and variance rates of the first-step estimates, $\widehat{m}_F(z;H)$ and $\widehat{m}_w(z;H)$, are uniform. Therefore, following Masry (1996) we impose some assumptions about the bandwidth $\widetilde{H}$ and its relationship with $H$. This is already considered in Assumptions **A.14** and **A.15**.

Let $\text{diag}_d(\text{tr}(\mathscr{H}_{m_r}(z)\widetilde{H}))$ be the diagonal matrix of elements $\text{tr}(\mathscr{H}_{m_r}(z)\widetilde{H})$ and $i_d$ a $d \times 1$ unit vector, Rodriguez-Poo and Soberon (2013a) obtain the following asymptotic expressions for the backfitting first-differences estimator (3.4),

**Theorem 3** *Assume conditions A.1-A.9 and A.14-A.15 holds, then, as N tends to infinity and T is fixed we get*

$$E(\widetilde{m}_F(z;\widetilde{H})|\mathbb{X}) - m(z) = \frac{1}{2}\mu_2(K)\,diag_d(tr(\mathscr{H}_{m_r}(z)\widetilde{H}))\iota_d + o_p(tr(\widetilde{H}))$$

*and*

$$Var(\widetilde{m}_F(z;\widetilde{H})|\mathbb{X}) = \frac{2\sigma_v^2 R(K)}{NT\left|\widetilde{H}\right|^{1/2}}\mathscr{B}_{XX}^{-1}(z)\,(1 + o_p(1)),$$

*where*

$$\mathscr{B}_{XX}(z) = E\left[X_{it}X_{it}^\top \middle| Z_{it} = z\right]f_{Z_{it}}(z).$$

Under similar conditions, in Rodriguez-Poo and Soberon (2013b) are proved the following asymptotic results for the backfitting within estimator (3.4):

**Theorem 4** *Assume conditions A.1-A.7, A.11-A.12 and A.14-A.15 hold, then as $N \to \infty$ and T remains to be fixed we obtain*

$$E(\widetilde{m}_w(z;\widetilde{H})|\mathbb{X}) - m(z) = \frac{1}{2}\mu_2(K)\,diag_d(tr(\mathscr{H}_{m_r}(z)\widetilde{H}))\iota_d + o_p(tr(\widetilde{H}))$$

*and*

$$Var(\widetilde{m}_w(z;\widetilde{H})\middle|\mathbb{X}) = \frac{\sigma_v^2 R(K)}{NT\left|\widetilde{H}\right|^{1/2}}\mathscr{B}_{XX}^{-1}(z)\mathscr{B}_{\ddot{X}\ddot{X}}(z)\mathscr{B}_{XX}(z)^{-1}\,(1 + o_p(1)),$$

*where*

$$\mathscr{B}_{XX}(z) = E\left[X_{it}X_{it}^\top \middle| Z_{it} = z\right]f_{Z_{it}}(z),$$

$$\mathscr{B}_{\ddot{X}\ddot{X}}(z) = E\left[\ddot{X}_{it}\ddot{X}_{it}^\top \middle| Z_{it} = z\right]f_{Z_{it}}(z).$$

With the aim of itemizing the asymptotic behavior of these two backfitting estimators, $\widetilde{m}_F(z;\widetilde{H})$ and $\widetilde{m}_w(z;\widetilde{H})$, we analyze in detail the bias and variance-covariance matrix of Theorems 3 and 4. On one hand, in both cases the conditional bias is very close to the standard one of the local polynomial regression estimators. Thus, as each entry of $\mathscr{H}_{m_r}(z)$ is a measure of the curvature of $m(\cdot)$ at $z$ in a particular direction, we can intuitively conclude that these estimators show a higher conditional bias as far as the unknown function exhibits a higher curvature and more smoothness. On the other hand, regarding to the conditional variance we observe that both estimators achieve the optimal rate of convergence, but they show different constants. Thus, while the first-differences estimator exhibits a variance-covariance matrix which increases when the smoothness becomes lower or the data becomes sparse near $z$, the conditional variance of the within estimator is also influenced by the time-demeaned covariates $\mathscr{B}_{\ddot{X}\ddot{X}}(z)$.

In this way, it is shown that direct estimation techniques allow obtaining estimators with different rates of convergence that depend on the type of differencing transformation. Meanwhile, one-step backfitting procedures provide estimators that achieve the optimal rate of convergence for both transformations. In this situation, the rate of convergence should not be used as an efficiency criterion between both backfitting estimators and, in order to analyze efficiency, it is necessary to study their finite sample behavior.

## 4 Monte Carlo experiment

In this section, we conduct an extensive Monte Carlo simulation with the aim of comparing the small sample behavior of both first-differences and within non-parametric estimators introduced in Sections 2 and 3.

In a fully parametric context, it is well-known that, under strict exogeneity assumptions the performance of both estimators is going to depend on the stochastic structure of the $v_{it}$'s random errors. Furthermore, as it appears in Theorems 1 and 2, the asymptotic bound for the variance term of the first-differences estimator is $O(1/NT|H|)$, whereas the corresponding term for the within estimator is $O(1/NT|H|^{T/2})$. Given that both estimators exhibit the same expression for the bias, for different values of $T$, one might expect a different behavior in their Average Mean Square Error (AMSE). On the opposite, for the one-step backfitting estimators proposed in Section 3 their performance should be affected by $T$ in the same direction, since both bias and variance terms are now of the same order, i.e. $O(\text{tr}(\widetilde{H}))$ and $O(1/NT|\widetilde{H}|^{1/2})$, respectively. Finally, the asymptotic bounds in the non-parametric setting reflect other factors that are also of interest when analyzing the AMSE behavior of both estimators such as the dimension of $q$ or the rate at which the AMSEs tend to zero as $N$ increases.

In this situation, the main goal that we pursue with this simulation experiment is to establish whether the behavior of these estimators depends on the stochastic structure of the error term and to determine if these results hold when facing the curse of dimensionality problem and for different number of time observations. Finally, it is also of interest to check wether the performance of the one-step backfitting estimator is, as expected, better than the corresponding for the local linear regression estimator in small samples sizes.

We propose a Monte Carlo experiment in which observations are generated from the following varying coefficient panel data model:

$$Y_{it} = X_{dit}^{\top} m(Z_{qit}) + \mu_{qi} + v_{it}, \quad i = 1, \cdots, N \quad ; \quad t = 1, \cdots, T, \quad d, q = 1, 2, \tag{4.1}$$

where $X_{dit}$ and $Z_{qit}$ are random variables generated such that $X_{dit} = 0.5\zeta_{dit} + 0.5\xi_{dit}$ ($\zeta_{1it}$ and $\zeta_{2it}$ are $i.i.d.$ $N(0,1)$), $Z_{qit} = \omega_{qit} + \omega_{qi(t-1)}$ ($\omega_{1it}$ and $\omega_{2it}$ are $i.i.d.$ $N(0,1)$) and we consider three different cases of study:

$$(q = 1 \ d = 1): \ Y_{it} = X_{1it} m(Z_{1it}) + \mu_{1i} + v_{it},$$
$$(q = 2 \ d = 1): \ Y_{it} = X_{1it} m(Z_{1it}, Z_{2it}) + \mu_{2i} + v_{it},$$
$$(q = 1 \ d = 2): \ Y_{it} = X_{1it} m(Z_{1it}) + X_{2it} m(Z_{1it}) + \mu_{1i} + v_{1it},$$

where the chosen functional forms are $m_1(Z_{1it}) = sin(Z_{1it}\Pi)$, $m_1(Z_{1it}, Z_{2it}) = \sin\left(\frac{1}{2}(Z_{1it} + Z_{2it})\Pi\right)$ and $m_2(Z_{1it}) = \exp\left(-Z_{1it}^2\right)$.

Treating the cross-sectional heterogeneity as the fixed effect, we allow that the individual effect can be correlated with one or more of the covariates. In particular, the dependence between $\mu_{qi}$ and $Z_{qit}$ is imposed by generating $\mu_{qi} = c_0 \overline{Z}_{i \cdot} + u_i$ and $Z_{i \cdot} = (TQ)^{-1} \sum_{q=1}^{Q} \sum_{t=1}^{T} Z_{qit}$, where $u_i$ is an $i.i.d.$ $N(0,1)$ random variable and $i = 2, \cdots, N$. The correlation between the fixed effects and some of the explanatory variables of the model is controlled by $c_0 = 0.5$. Also, let $\varepsilon_{it}$ be an $i.i.d.$ $N(0,1)$ and $v_{it}$ a scalar random variable, for each model we work with the following three different specifications of the error term:

a) $v_{it} = \varepsilon_{it}$;
b) $v_{it}$ follows a random walk, such us $v_{it} = 1 + v_{i(t-1)} + \varepsilon_{it}$;
c) $v_{it}$ is generated as stationary AR(1) process of the form $v_{it} = \rho v_{i(t-1)} + \varepsilon_{it}$.

In each experiment we use 1000 Monte Carlo replications ($M$). The number of period ($T$) is varied to be 3 and 5, whereas the number of cross-sections ($N$) takes the values 50, 100 and 150. For the calculations we use a Gaussian kernel and the bandwidth is chosen as $\widehat{H} = \widehat{h}I$, and $\widehat{h} = \widehat{\sigma}_z (NT)^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{qit}\}_{i=1, t=1}^{N,T}$.

In order to state the performance of the first-differences and within estimator, we use the Mean Square Error (MSE) as a measure of their estimation accuracy. Thus, denoting the $\varphi$th replication by the subscript $\varphi$,

$$MSE(\widehat{m}(z; \widehat{H})) = \frac{1}{M} \sum_{\varphi=1}^{M} E\left[\left(\sum_{r=1}^{d} (\widehat{m}_{\varphi r}(z; \widehat{H}) - m_{\ell r}(z)) X_{it, \varphi r}\right)^2\right],$$

which can be approximated by the Averaged Mean Squared Error (AMSE),

$$AMSE(\widehat{m}(z; \widehat{H})) = \frac{1}{M} \sum_{\varphi=1}^{M} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\sum_{r=1}^{d} (\widehat{m}_{\varphi r}(z; \widehat{H}) - m_{\varphi r}(z)) X_{it, \varphi r}\right)^2.$$

Simulation results are summarized in Tables 1-9 that are relegated to the Appendix. Specifically, Tables 1-3 contain the AMSE obtained for each of the three varying coefficient specifications proposed in the simulation when the error term is *i.i.d.*, Tables 4-6 focus on the results for the random walk case, whereas Tables 7-9 contain the simulation results when the idiosyncratic errors are generated according to an $AR(1)$ stationary structure. In every table, we present results for the two differencing estimators, $\widehat{m}_F(z;H)$ and $\widehat{m}_w(z;H)$, and for both one-step backfitting estimators. We consider the cases when $T = 3, 5$ and $N = 50, 100, 150$. Note that our asymptotic results hold as $N$ becomes larger whereas $T$ is kept fixed.

Based on these simulation results, we first try to determine how the structure of the stochastic error term affects the AMSE behavior of the proposed estimators. Later, we focus on the impact that both curse of dimensionality and number of time observations have on the behavior of the AMSE. We conclude the analysis by checking whether, regardless of the structure of the error term, the one-step backfitting estimator exhibits a better performance than the local linear regression estimator, as we expected from their statistical properties.

According to our theoretical findings and taking into account some standard results in fully parametric settings, in the *i.i.d.* case, the within estimator should perform better, in terms of the AMSE, than the first-differences one, whereas the later should be preferred when the disturbance follows a random walk. A quick look to Table 1 confirms in general our theoretical findings. As expected, as $N$ increases all AMSEs tend to zero but the rates of convergence are not similar. In fact, if we look at the relative AMSE, defined as $AMSE(\widehat{m}_F(z;\widehat{H}))/AMSE(\widehat{m}_w(z;\widehat{H}))$, we find out that the rate at which the AMSE of the within estimator converges to zero is faster than the convergence of the first-differences estimator. Therefore, in this stochastic setting, the within estimator is preferred.

On the other hand, for the random walk case, some standard results for the fully parametric case point out that first-differences estimator should exhibit a better performance. This is true if we compare the AMSE of the first-differences estimator in Tables $1 - 3$ against their counterparts in Tables $4 - 6$. In all cases, the AMSE is smaller when the idiosyncratic errors are generated as a random walk.

Finally, in Tables $7 - 9$ we find out that when the random errors follow an autoregressive process $AR(1)$, the within estimator performs better than the first-differences estimator in terms of their AMSE, although its performance becomes worse if we compare it against other error settings. Moreover, the results of the first-differences estimator are better than the other ones obtained in the random walk setting. This is somehow unexpected.

Now, to establish the impact of the curse of dimensionality on the performance of the proposed estimators we compare the AMSE of $\widehat{m}_F(z;H)$ and $\widehat{m}_w(z;H)$ when $q = 1$ (Tables 1, 4 and 7), against their counterparts when $q = 2$ (Tables 2, 5 and 8). As expected, we point out that when $q = 2$ the relative performance in terms of the AMSE of the proposed estimators is worse in all error settings, although the within estimator is the most affected by the curse of dimensionality. When the idiosyncratic error term is *i.i.d.* the relative AMSE for $N = 150$ and $T = 3$ goes from 4.463 ($q = 1$) to 1.688 ($q = 2$), whereas for the random walk case the relative AMSE for $N = 150$ and $T = 3$ goes from 3.228 ($q = 1$) to 1.190 ($q = 2$). The same can be said when the error term is generated according to an $AR(1)$ stationary structure. Finally, the results shown in Tables 3, 6 and 9 indicate that the $d$-dimension of the vector of covariates $X$ does not affect the asymptotic behavior of the estimators.

On the contrary, if we focus now on the impact of the number of time observations we find out that the within estimator $\widehat{m}_w(z;H)$ is much more sensitive to $T$ than the other. As we can realize in all tables, as the number of time observations ($T$) increases the relative performance of $\widehat{m}_w(z;H)$ becomes worse. In Table 1, for example, if we set $N = 150$ its relative AMSE goes from 4.46 ($T = 3$) to 3.33 ($T = 5$). This effect can be explained in terms of the asymptotic bounds of both estimators. In Tables $7 - 9$ we can see that the within estimator is much more sensitive to the size of $T$ than the first-differences estimator when the random error term follows an autoregressive structure.

As a summary, the results of $\widehat{m}_F(z;H)$ are much more stable across different specifications of the error term. The same cannot be said about the $\widehat{m}_w(z;H)$. In fact, it performs quite well with $q = 1$ and under *i.i.d.* or an AR(1) stationary process, but it shows a much worse performance when the errors are generated following a random walk process. Therefore, we can conclude that the within estimator is preferred when the idiosyncratic error term is *i.i.d.* or exhibits an autoregressive structure, whereas when $v_{it}$ follows a random walk the first-differences estimator has a better performance.

Finally, in all error settings, the performance of the one-step backfitting estimator is better than its corresponding local linear regression estimator. However, the improvement is higher for the within estimator. Also, as we expected, the curse of dimensionality is overridden. This is of course the main reason why we have applied this second stage estimation procedure. Furthermore, the rate at which the AMSE tends to zero for both estimators seems to be faster, according to the predictions of the asymptotic results. However, for small sample sizes, although both estimators do have the same rates of convergence the constants are different. This is considered in simulations under different scenarios of the error terms. Hence, under the *i.i.d.* and the $AR(1)$ setting the first-step backfitting algorithm of the within estimator performs better than the first-differences one. This can be realized by analyzing the relative AMSE. On the contrary, under the random walk specification, the performance is better in the opposite sense.

We finish this section by highlighting that as $N$ increases, that is, asymptotically, the AMSE tends to converge for each estimator under different specifications of the error term. For different values of $T$, the AMSE of the first-

differences estimator tends to dominate in terms of the AMSE of the within estimator. This can be also observed by looking at the relative AMSE values.

## 5 Conclusions

Recently, some new techniques have been proposed for the estimation of semi-parametric within varying coefficient panel data models. These new techniques fall within the class of the so-called differencing estimators. In particular, we consider first-differences and within local linear regression estimators. Analyzing their asymptotic properties it turns out that, keeping the same order of magnitude for the bias term, these estimators exhibit different asymptotic bounds for the variance. In both cases, the consequences are suboptimal non-parametric rates of convergence. In order to solve this problem, by exploiting the additive structure of this model, a one-step backfitting algorithm is proposed. Under fairly general conditions, it turns out that the resulting estimators show optimal rates of convergence and exhibit the oracle efficiency property. Since both estimators are asymptotically equivalent, it is of interest to analyze their behavior in small sample sizes. In a fully parametric context, it is well-known that, under strict exogeneity assumptions the performance of both first-differences and within estimators is going to depend on the stochastic structure of the idiosyncratic random errors. However, in the non-parametric setting, apart from the previous issues other factors such as dimensionality or sample size are of great interest. In particular, we would be interested in learning about their relative Average Mean Square Error under different scenarios. The simulation results basically confirm the theoretical findings for both local linear regression and one-step backfitting estimators. However, we have found out that within estimators are rather sensitive to the size of number of time observations.

## Appendix I: assumptions

To analyze the asymptotic distributions of the above estimators we need to establish some assumptions

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

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

**Assumption A.3** The unobserved cross-sectional effect, $\mu_i$, can be arbitrarily correlated with both $X_{it}$ and/or $Z_{it}$ with an unknown correlation structure.

**Assumption A.4** Let $f_{Z_{1t}}(\cdot)$, $f_{Z_{1t},Z_{1(t-1)}}(\cdot,\cdot)$, $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(\cdot,\cdot,\cdot)$ be respectively the probability density function of $Z_{1t}$, $(Z_{1t}, Z_{1(t-1)})$ and $(Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)})$. All density functions are continuously differentiable in all their arguments and they are bounded from above and below in any point of their support.

**Assumption A.5** Let $z$ an interior point in the support of $f_{Z_{1t}}$. All second-order derivatives of $m_1(\cdot), m_2(\cdot), \cdots, m_d(\cdot)$ are bounded and uniformly continuous.

**Assumption A.6** The $q$-variate Kernel functions $K$ are compactly supported, bounded kernel such that $\int uu^\top K(u)du = \mu_2(K)I$ and $\int K^2(u)du = R(K)$, where $\mu_2(K) \neq 0$ and $R(K) \neq 0$ are scalars and $I$ is the $q \times q$ identity matrix. In addition, all odd-order moments of $K$ vanish, that is $\int u_1^{\iota_1} \cdots u_q^{\iota_q} K(u)du = 0$, for all nonnegative integers $\iota_1, \cdots, \iota_q$ such that their sum is odd.

**Assumption A.7** The bandwidth matrix $H$ is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as $N \to \infty$ in such a way that $N|H| \to \infty$.

**Assumption A.8** Let $\|A\| = \sqrt{tr(A^\top A)}$, then $E\left\{\|X_{it}X_{it}^\top\|^2 \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2\right\}$ is bounded and uniformly continuous in its support. Furthermore, let

$$\mathscr{X}_{it} = \left(X_{it}^\top \; X_{i(t-1)}^\top\right)^\top \quad \text{and} \quad \Delta\mathscr{X}_{it} = \left(\Delta X_{it}^\top \; \Delta X_{i(t-1)}^\top\right)^\top$$

.

Also, the following matrix functions $E\left(\mathscr{X}_{it}\mathscr{X}_{it}^\top \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2\right)$, $E\left(\mathscr{X}_{it}\mathscr{X}_{it}^\top \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2, Z_{i(t-2)} = z_3\right)$, $E\left(\Delta\mathscr{X}_{it}\Delta\mathscr{X}_{it}^\top \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2\right)$ and $E\left(\mathscr{X}_{it}\Delta\mathscr{X}_{it}^\top \middle| Z_{it} = z_1, Z_{i(t-1)} = z_2, Z_{i(t-2)} = z_3\right)$ are bounded and uniformly continuous in their support.

**Assumption A.9** The function $E\left[\Delta X_{it}\Delta X_{it}^{\top}\middle|Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$ is positive definite for any interior point of $(z_1,z_2)$ in the support of $f_{Z_{it},Z_{i(t-1)}}(z_1,z_2)$.

**Assumption A.10** For some $\delta > 0$, the functions $E\left\{|X_{it}\Delta v_{it}|^{2+\delta}\middle|Z_{it}=z_1,Z_{i(t-1)}=z_2\right\}$, $E\left\{|X_{i(t-1)}\Delta v_{it}|^{2+\delta}\middle|Z_{it}=z_1,Z_{i(t-1)}=z_2\right\}$ and $E\left\{|\Delta X_{it}\Delta v_{it}|^{2+\delta}\middle|Z_{it}=z_1,Z_{i(t-1)}=z_2\right\}$ are bounded and uniformly continuous in any point of their support.

**Assumption A.11** The function $E\left(\ddot{X}_{it}\ddot{X}_{it}^{\top}\middle|Z_{i1}=z,\cdots,Z_{iT}=z\right)$ is definite positive for any interior point of $(z_1,z_2,\cdots,z_T)$ in the support of $f_{Z_{i1},\cdots,Z_{iT}}(z_1,z_2,\cdots,z_T)$.

**Assumption A.12** Let $\|A\|=\sqrt{tr(A^{\top}A)}$, then $E\left\{\|X_{it}X_{it}^{\top}\|^2\middle|Z_{i1}=z,\cdots,Z_{iT}=z\right\}$ is bounded and uniformly continuous in its support. Furthermore, let $\mathscr{X}_{it}=\left(X_{it}^{\top}\;X_{is}^{\top}\right)^{\top}$ the following matrix functions $E\left(\mathscr{X}_{it}\mathscr{X}_{it}^{\top}\middle|Z_{i1}=z_1,\cdots,Z_{iT}=z_T\right)$, $E\left(\ddot{X}_{it}\ddot{X}_{it}^{\top}\middle|Z_{i1}=z_1,\cdots,Z_{iT}=z_T\right)$ and $E\left(\ddot{X}_{it}\mathscr{X}_{it}^{\top}\middle|Z_{i1}=z_1,\cdots,Z_{iT}=z_T\right)$ are bounded and uniformly continuous in their support.

**Assumption A.13** The functions $E\left\{|X_{it}v_{it}|^{2+\delta}|Z_{i1}=z,\cdots,Z_{iT}=z\right\}$, $E\left\{|X_{is}v_{it}|^{2+\delta}|Z_{i1}=z,\cdots,Z_{iT}=z\right\}$ and $E\left\{|\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1}=z,\cdots,Z_{iT}=z\right\}$ are bounded and uniformly continuous in any point of their support, for some $\delta > 0$.

**Assumption A.14** The bandwidth matrix $\widetilde{H}$ is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as $N$ tends to infinity in such a way that $N\left|\widetilde{H}\right|\to\infty$.

**Assumption A.15** The bandwidth matrices $H$ and $\widetilde{H}$ must fulfill that $N|H|\left|\widetilde{H}\right|/log(N)\to\infty$, and $tr(H)/tr\left(\widetilde{H}\right)\to 0$ as $N$ tends to infinity.

## Appendix II: Monte Carlo results

The results of the Monte Carlo experiments are summarized in the following figures.

**Table 1** AMSE for $d = 1$ and $q = 1$ when $v_{it}$ is *i.i.d.* $N(0,1)$.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.501 | 0.267 | 1.879 | 0.516 | 0.089 | 5.749 |
| | N = 100 | 0.453 | 0.167 | 2.707 | 0.463 | 0.054 | 8.502 |
| | N = 150 | 0.431 | 0.096 | 4.463 | 0.444 | 0.034 | 12.88 |
| T=5 | N = 50 | 0.431 | 0.232 | 1.854 | 0.441 | 0.053 | 8.366 |
| | N = 100 | 0.415 | 0.166 | 2.507 | 0.428 | 0.034 | 12.28 |
| | N = 150 | 0.411 | 0.123 | 3.333 | 0.428 | 0.023 | 18.49 |

**Table 2** AMSE for $d = 1$ and $q = 2$ when $v_{it}$ is *i.i.d.* $N(0,1)$.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.876 | 0.713 | 1.228 | 0.526 | 0.258 | 2.044 |
| | N = 100 | 0.741 | 0.507 | 1.460 | 0.403 | 0.171 | 2.361 |
| | N = 150 | 0.676 | 0.400 | 1.688 | 0.348 | 0.123 | 2.794 |
| T=5 | N = 50 | 0.583 | 0.859 | 0.679 | 0.677 | 0.160 | 4.221 |
| | N = 100 | 0.532 | 0.648 | 0.821 | 0.533 | 0.111 | 4.782 |
| | N = 150 | 0.496 | 0.497 | 0.998 | 0.469 | 0.079 | 5.890 |

**Table 3** AMSE for $d = 2$ and $q = 1$ when $v_{it}$ is *i.i.d.* $N(0,1)$.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\hat{H}))}{AMSE(\widehat{m}_W(z;\hat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_W(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.673 | 0.405 | 1.662 | 0.687 | 0.146 | 4.849 |
| | N = 100 | 0.556 | 0.268 | 2.075 | 0.565 | 0.089 | 6.353 |
| | N = 150 | 0.509 | 0.178 | 2.867 | 0.515 | 0.053 | 6.686 |
| T=5 | N = 50 | 0.529 | 0.539 | 0.980 | 0.529 | 0.087 | 6.083 |
| | N = 100 | 0.489 | 0.415 | 1.181 | 0.497 | 0.056 | 8.945 |
| | N = 150 | 0.474 | 0.270 | 1.752 | 0.480 | 0.035 | 13.84 |

**Table 4** AMSE for $d = 1$ and $q = 1$ when $v_{it}$ follows a random walk.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\hat{H}))}{AMSE(\widehat{m}_W(z;\hat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_W(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.483 | 0.347 | 1.393 | 0.526 | 0.118 | 4.446 |
| | N = 100 | 0.449 | 0.218 | 2.054 | 0.466 | 0.074 | 6.298 |
| | N = 150 | 0.427 | 0.132 | 3.228 | 0.444 | 0.045 | 9.863 |
| T=5 | N = 50 | 0.416 | 0.217 | 1.917 | 0.431 | 0.050 | 8.553 |
| | N = 100 | 0.404 | 0.162 | 2.492 | 0.421 | 0.033 | 12.66 |
| | N = 150 | 0.404 | 0.119 | 3.385 | 0.423 | 0.023 | 18.75 |

**Table 5** AMSE for $d = 1$ and $q = 2$ when $v_{it}$ follows a random walk.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_W(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_W(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.808 | 0.948 | 0.852 | 0.518 | 0.345 | 1.505 |
| | N = 100 | 0.722 | 0.706 | 1.023 | 0.399 | 0.235 | 1.684 |
| | N = 150 | 0.652 | 0.548 | 1.190 | 0.342 | 0.162 | 2.116 |
| T=5 | N = 50 | 0.502 | 0.911 | 0.551 | 0.571 | 0.150 | 3.799 |
| | N = 100 | 0.474 | 0.613 | 0.773 | 0.476 | 0.107 | 4.468 |
| | N = 150 | 0.445 | 0.466 | 0.956 | 0.439 | 0.077 | 5.732 |

**Table 6** AMSE for $d = 2$ and $q = 1$ when $v_{it}$ follows a random walk.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_W(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_W(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.638 | 0.592 | 1.078 | 0.690 | 0.210 | 3.291 |
| | N = 100 | 0.544 | 0.435 | 1.250 | 0.564 | 0.129 | 4.357 |
| | N = 150 | 0.502 | 0.272 | 1.849 | 0.516 | 0.077 | 6.708 |
| T=5 | N = 50 | 0.492 | 0.522 | 0.943 | 0.504 | 0.081 | 6.236 |
| | N = 100 | 0.457 | 0.384 | 1.189 | 0.469 | 0.034 | 13.89 |
| | N = 150 | 0.453 | 0.026 | 17.33 | 0.468 | 0.034 | 13.84 |

**Table 7** AMSE for $d = 1$ and $q = 1$ when $v_{it}$ is $AR(1)$ with $\rho = 0.5$.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.454 | 0.174 | 2.604 | 0.475 | 0.071 | 6.669 |
| | N = 100 | 0.426 | 0.119 | 3.584 | 0.442 | 0.047 | 9.479 |
| | N = 150 | 0.418 | 0.076 | 5.501 | 0.435 | 0.031 | 14.23 |
| T=5 | N = 50 | 0.420 | 0.550 | 764 | 0.444 | 0.113 | 3.986 |
| | N = 100 | 0.410 | 0.449 | 0.913 | 0.432 | 0.074 | 5.867 |
| | N = 150 | 0.406 | 0.299 | 1.354 | 0.429 | 0.043 | 9.909 |

**Table 8** AMSE for $d = 1$ and $q = 2$ when $v_{it}$ is $AR(1)$ with $\rho = 0.5$.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.670 | 0.605 | 1.008 | 0.443 | 0.215 | 2.056 |
| | N = 100 | 0.600 | 0.429 | 1.397 | 0.361 | 0.148 | 2.441 |
| | N = 150 | 0.556 | 0.343 | 1.623 | 0.324 | 0.107 | 3.033 |
| T=5 | N = 50 | 0.563 | 2.153 | 0.262 | 0.678 | 0.364 | 1.864 |
| | N = 100 | 0.505 | 1.631 | 0.309 | 0.526 | 0.249 | 2.109 |
| | N = 150 | 0.475 | 1.295 | 0.367 | 0.468 | 0.171 | 2.739 |

**Table 9** AMSE for $d = 2$ and $q = 1$ when $v_{it}$ is $AR(1)$ with $\rho = 0.5$.

| | | Local Linear Estimator | | | One-Step Backfitting Estimator | | |
|---|---|---|---|---|---|---|---|
| | | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ | First-Differences | Within | $\frac{AMSE(\widehat{m}_F(z;\widehat{H}))}{AMSE(\widehat{m}_w(z;\widehat{H}))}$ |
| T=3 | N = 50 | 0.581 | 0.393 | 1.479 | 0.607 | 0.138 | 4.409 |
| | N = 100 | 0.512 | 0.220 | 2.328 | 0.524 | 0.045 | 11.67 |
| | N = 150 | 0.481 | 0.137 | 3.503 | 0.495 | 0.045 | 11.02 |
| T=5 | N = 50 | 0.510 | 1.839 | 0.277 | 0.534 | 0.238 | 2.245 |
| | N = 100 | 0.475 | 1.230 | 0.386 | 0.493 | 0.141 | 3.508 |
| | N = 150 | 0.458 | 0.884 | 0.518 | 0.481 | 0.087 | 5.501 |

## References

Altonji, J. and R. Matzkin (2005). "Cross section and panel data estimators for nonseparable models with endogenous regressors". *Econometrica 73*, 1053–1102.

Bester, A. and C. Hansen (2009). "Identification of marginal effects in a nonparametric correlated random effects model". *Journal of Business and Economic Statistics 27*, 235–250.

Buja, A., T. Hastie, and R. Tibshirani (1989). "Linear Smoothers and Additive Models". *The Annals of Statistics 17*, 453–510.

Card, D. (2001). "Estimating the Return to Schooling: Progress on some Persistent Econometric Problems". *Econometrica 69*, 1127–1160.

Fan, J. and I. Gijbels (1995a). "Data-driven Bandwidth Selection in Local Polynomial Fitting: Variable Bandwidth and Spatial Adaption". *Journal of the Royal Statistical Society Series B*(57), 371–394.

Fan, J. and I. Gijbels (1995b). *"Local Polynomial Modelling and its Applications"*. Chapman and Hall.

Fan, J. and W. Zhang (1999). "Statistical Estimation in Varying Coefficient Models". *Annals of Statistics 27*, 1491–1518.

Henderson, D., R. Caroll, and Q. Li (2008). "Nonparametric Estimation and Testing of Fixed Effects Panel Data Model". *Journal of Econometrics 144*, 257–275.

Hoderlein, S. and H. White (2012). "Nonparametric identification in nonseparable panel data models with generalized fixed effects". *Journal of Econometrics 168*, 300–314.

Kottaridi, C. and T. Stengos (2010). "Foreign Direct Investment, Human Capital and Nonlinearities in Economics Growth". *Journal of Macroeconomics 32*, 858–871.

Kuan, C. and C. Chen (2013). "Effects of national health insurance on precautionary saving: new evidence from Taiwan". *Empirical Economics 44*, 942–943.

Mammen, E., B. Støve, and D. Tjøstheim (2009). "Nonparametric Additive Models for Panels of Time Series". *Econometric Theory 25*, 442–481.

Masry, E. (1996). "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency Rate". *Journal of Time Series Analysis 17*, 571–599.

Qian, J. and L. Wang (2012). "Estimating Semiparametric Panel Data Models by Marginal Integration". *Journal of Econometrics 167*, 483–493.

Rodriguez-Poo, J. and A. Soberon (2013a). Direct Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models. *The Econometrics Journal 17*(1), 107–138.

Rodriguez-Poo, J. and A. Soberon (2013b). Efficient Nonparamteric Estimation of Fixed Effects Panel Data Varying Coefficient Models. Preprint.

Su, L. and X. Lu (2013). "Nonparametric Dynamic Panel Data Models: Kernel Estimation and Specification Testing". *Journal of Econometrics 176*(2), 112–133.

Su, L. and A. Ullah (2011). "Nonparametric and Semiparametric Panel Econometric Models: Estimation and Testing". in Handbook of Empirical Economics and Finance. A. Ullah and D.E.A. Giles (eds), New York: Taylor and Francis Group.

Sun, Y., R. Carroll, and D. Li (2009). "Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models". *Advances in Econometrics 25*, 101–130.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Camdridge, Massachusets.