Direct Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models

Alexandra Soberón
Universidad de Cantabria,
Avda de los Castros s/n, E-39005 Santander, Spain

Juan M. Rodriguez-Poo
Universidad de Cantabria,
Avda de los Castros s/n, E-39005 Santander, Spain

9th of July, 2012

Abstract

In this paper we present a new technique to estimate varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in a unknown way. The resulting estimator is robust to misspecification in the functional form of the varying parameters and it is shown to be consistent and asymptotically normal. Furthermore, introducing a transformation, it achieves the optimal rate of convergence for this type of problems and it exhibits the so called "oracle" efficiency property. Since the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. Monte Carlo results indicates good performance of the estimator in finite samples. [1]

JEL codes: C14, C23
Key Words: Varying coefficients model, fixed effects, panel data, locally weighted least squares regression, "oracle" estimator.

[1]The authors acknowledge financial support from the Programa Nacional de Formacin de Profesorado Universitario/ Spanish Ministery of Education. Ref. AP-2007-02209.

1 Introduction

In the last five decades, an enormous amount of literature has been devoted to the study of panel data models. Indeed, the availability of this type of data has enriched the complexity of econometric models and therefore, it has enabled us to extract information and inferences that were not possible to obtain with pure cross-section or time series models. Traditionally, (see Baltagi (2005) and Hsiao (2003)) econometric specification of this type of models was concerned with unobserved individuals heterogeneity and its relationship with the explanatory variables. Much less concern was shown about the robustness of parameters estimators to deviations from linearity or other distributional assumptions. However, as it has been already emphasized in Arellano (2003) most part of panel data models rely on rather strong assumptions about functional forms and densities. Although in many cases these assumptions are quite realistic, there exists situations in which the risk of misspecification is high. If this is the case, standard estimators based in moment conditions are biased and their use might lead to wrong inference. Among other possibilities, one approach to overcome these problems is to use fully nonparametric regression techniques. In this context, if the individual effects are assumed to be uncorrelated with the regressors (random effects) Henderson and Ullah (2005), Lin and Carroll (2000), Li and Racine (2007) and Ullah and Roy (1998) propose several alternative nonparametric estimators. If correlation between individual effects and regressors is allowed (fixed effects) then Henderson et al. (2008) and Lee and Mukherjee (2008) propose a nonaprametric estimator that exhibits good asymptotic properties. Although this type of estimators are robust to incorrect specification of the regression function they show a quite unpleasant property that is the so-called "curse of dimensionality". That is, as far as the number of explanatory variables increases the rate of convergence of these estimators becomes dramatically slower. To overcome this problem, semiparametric methods have been proposed in the literature. What is known from previous empirical research or economic theory is modeled parametrically whereas what is unknown for the researcher is specified nonparametrically. Partially index models fall within this class of semiparametric models and there have been widely used in panel data analysis. For the case of random effects see for example Kniesner and Li (2002), Li and Stengos (1996), Li and Hsiao (1998), Su and Ullah (2010) and You et al. (2010) and for fixed effect models see Baltagi and Li (2002) and Su and Ullah (2006).

Recently, some empirical problems such as the estimation of returns to education have motivated the introduction of time varying coefficient models in panel data. As it has been already pointed out in Su and Ullah (2010), varying coefficient models encompass both nonparametric and partially linear models, and therefore they offer a quite general setting to handle a great variety of problems. In the context of varying coefficients, random effect panel data models can be easily estimated through standard nonparametric techniques as for the case of fully nonparametric models. However, under fixed effects the task becomes much cumbersome. In fact, as it has been already remarked in Su and Ullah (2010), traditional techniques for the estimation of this type of models such as standard differencing methods are hard to use in this framework because without further restrictions it is not possible to identify the varying parameters. Motivated by a least squares dummy variable model in parametric panel data analysis, in Sun et al. (2009) is proposed a profile local linear estimator of the varying parameters that is consistent and asymptotically normal. Unfortunately, as in all this type of literature, the estimator relies on a rather strong identification assumption about the heterogeneity component. Trying to overcome this drawback, in this paper we present an estimator that, based in differencing techniques, does not need the type of strong identification assumptions assumed in other contexts and furthermore it does not need of iterative procedures to achieve nice asymptotic properties. The main idea is borrowed from Yang (2002) who proposed it in a completely different context and consists in approximating locally at the same point different values of the same function. The resulting estimator is obtained by a local linear kernel weighted least squares procedure. It turns out to be consistent and asymptotically normal, although it shows a slow rate of convergence. Then, we propose a transformation that allows this estimator to achieve an optimal rate of convergence and a nice "oracle" efficiency property. We also provide a method to compute the bandwidth matrix empirically.

The rest of the paper is organized as follows. In Section 2 we set up the model and the estimation procedure. In Section 3 we study its asymptotic properties and we propose a transformation procedure that provides an estimator that is "oracle" efficient and achieves optimal rates of convergence. Section 4 shows how to estimate the bandwidth matrix empirically and finally in Section 5 we present some simulation results. Finally, Section 6 concludes the paper. The proofs of the main results are collected in the Appendix.

2 Statistical model and Estimation procedure

Let $(X_{it}, Z_{it}, Y_{it})_{i=1,\cdots,N;t=1,\cdots,T}$ be a set of independent and identically $\mathbb{R}^{d+q+1}$-random variables in the subscript i, where the Y_{it} are scalar response variables and $(X_{it}, Z_{it})_{i=1,\cdots,N;t=1,\cdots,T}$ are explanatory $\mathbb{R}^{d+q}$-random variables.

We assume the response variables are generated by the following statistical model,

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

where the function $m(Z) = (m_1(Z), \cdots, m_d(Z))^T$ is unknown to the researcher and needs to be estimated. The random errors v_{it} are independent and identically distributed, with zero mean and homoscedastic variance, $\sigma_v^2 < \infty$. Furthermore, they are independent of μ_i, Z_{is} and X_{is} for all i, j. The unobserved individual effects μ_i are assumed to be independent and identically distributed with zero mean and constant variance $\sigma_\mu^2 < \infty$. Also, μ_i is correlated with the X_{it}'s and/or Z_{it}'s with an unknown correlation structure. Note that, we do not assume further conditions on the μ's as it is the case in Sun et al. (2009) where they assume $\sum_{i=1}^N \mu_i = 0$.

This econometric framework corresponds to the so called fixed effect varying coefficient panel data regression model. Note that correlation of unknown form in the time series is allowed. Furthermore, independence between the v_{it}-errors and the X_{it}-variables and/or the Z_{it}-variables is assumed without loss of generality. We could relax this assumption by assuming some dependence based on second order moments. For example, heteroskedasticity of unknown form can be allowed and in fact, under more complex structures in the variance-covariance matrix a transformation of the estimator proposed in You et al. (2010) can be developed in our setting. This type of assumption also rules out the existence of endogenous explanatory variables and imposes strict exogeneity conditions. In this case, it is also possible to relax these conditions by introducing a instrumental variable method as it is done in Soberon and Rodriguez-Poo (2012). Finally, all our results hold straightforwardly for the random coefficient setting.

Using standard nonparametric estimation methods, consistent estimation of the unknown function, $m(\cdot)$, in (1) is hard to undertake. The main reason is the statistical dependence between μ_i and X_{it} and/or Z_{it}. This dependence produces systematic bias in any standard estimation procedure. As it has been already pointed out by many authors, see in Su and Ullah (2010), there exist two different ways to overcome this problem. One is the so called profile least squares method proposed alternatively in Su and Ullah (2010) and Sun et al. (2009) and the other is based in differencing methods. This last technique consists in removing the individual effects by taking differences in (1). We then obtain the following transformation

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

As it can be clearly remarked, the presence of the function $m(\cdot)$, evaluated at two different values Z_{it} and Z_{it-1}, makes cumbersome the estimation. One possibility that has been already mentioned by some authors is to implement a two step procedure. In the first step, the whole expression $X_{it}^T m(Z_{it}) - X_{i(t-1)}^T m(Z_{i(t-1)})$ is estimated through a multivariate nonparametric smoother. In a second stage, $m(\cdot)$ is estimated using for example marginal integration techniques. Its is clear that this estimation procedure is not very appealing. Among other reasons it is very hard to implement, and to our knowledge no empirical work has been developed based on these estimators. On the other side, profile least squares methods, as we have already pointed out before, need some rather strong assumptions that are not usually considered in panel data models framework. Furthermore, the resulting estimators do not show the so called "oracle" efficiency property and they are computationally intensive. In what follows we present an estimator for $m(\cdot)$ that, based in differencing methods, it is easy to compute and it does not rely on iterative procedures. Furthermore, applying a simple one step backfitting algorithm the estimator achieves the optimal rate of convergence and it exhibits the 'oracle' efficiency property.

The main idea is to approximate linearly around the same element z_0, in the interior of the support of $f(Z)$, both $m(Z_{it})$ and $m\left(Z_{i(t-1)}\right)$. If we do so, then (2) becomes

$$X_{it}^T m(Z_{it}) - X_{i(t-1)}^T m(Z_{i(t-1)}) \simeq \Delta X_{it}^T m(z_0) + \left\{ X_{it}^T \otimes (Z_{it} - z_0)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z_0)^T \right\} D_m(z_0), \tag{3}$$

where $D_m(z) = \text{vec}\left(\frac{\partial m(z)}{\partial z^T}\right)$ is a $dq \times 1$ vector of first derivatives of $m(\cdot)$. The quantity of interest, $m(z)$, can be estimated using a local linear least squares kernel estimator (see Fan and Gijbels (1995), Ruppert and Wand (1994) or Zhan-Qian (1996)),

$$min_{\beta} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it} - \widetilde{Z}_{it}^T \beta \right\}^2 K_H \left(Z_{it} - z_0 \right) K_H \left(Z_{i(t-1)} - z_0 \right) \tag{4}$$

where $\beta = \left[\alpha^T, \delta^T \right]^T$ is a $d(1+q) \times 1$ vector, $\alpha = m(z_0)$ is a $d \times 1$ vector, $\delta = D_m(z_0)$ is a $dq \times 1$ vector and

$$\widetilde{Z}_{it}^T = \left[\Delta X_{it}^T, \quad X_{it}^T \otimes (Z_{it} - z_0)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z_0)^T \right]$$

is a $1 \times d(1+q)$ vector. K is a q-variate kernel such that

$$\int K(u) du = 1 \qquad \text{and} \qquad K_H(u) = \frac{1}{|H|^{1/2}} K \left(H^{-1/2} u \right),$$

where H is a $q \times q$ symmetric positive definite bandwidth matrix.

It is easy to verify that the solution to the minimization problem in (4) can be written in matrix form as

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\delta} \end{bmatrix} = \left(\widetilde{Z}^T W \widetilde{Z} \right)^{-1} \widetilde{Z}^T W \Delta Y, \tag{5}$$

where

$$\begin{aligned} W &= diag \left\{ K_H(Z_{12} - z_0) K_H(Z_{11} - z_0), ..., K_H(Z_{NT} - z_0) K_H(Z_{N(T-1)} - z_0) \right\}, \\ \Delta Y &= [\Delta Y_{12}, ..., \Delta Y_{NT}]^T, \end{aligned}$$

and

$$\widetilde{Z} = \begin{bmatrix} \Delta X_{12}^T & X_{12}^T \otimes (Z_{12} - z_0)^T - X_{11}^T \otimes (Z_{11} - z_0)^T \\ \vdots & \vdots \\ \Delta X_{NT}^T & X_{NT}^T \otimes (Z_{NT} - z_0)^T - X_{N(T-1)}^T \otimes (Z_{N(T-1)} - z_0)^T \end{bmatrix}.$$

The local weighted linear least squares estimator of $m(z_0)$ is then defined as

$$\widehat{m}(z_0; H) = e_1^T \left(\widetilde{Z}^T W \widetilde{Z} \right)^{-1} \widetilde{Z}^T W \Delta Y, \tag{6}$$

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. Note that the dimensions of W and $\widetilde{Z}$ are respectively $N(T-1) \times N(T-1)$ and $N(T-1) \times d(1+q)$.

This estimation procedure, in a especial case where $d = 1$ and $X_{it} = 1$ can be applied to a fully non-parametric fixed effect panel data model as the one considered in Su and Ullah (2006) and Henderson et al. (2008). Furthermore, if we make $X_{it}^T m(Z_{it}) = m_1(Z_{it}) + \tilde{X}_{it}^T \beta_0$, for some real valued function $m_1(\cdot)$ and $(d-1) \times 1$ vector β_0 our estimation technique can be applied to estimate consistently $m_1(\cdot)$ and β_0 in either a fixed effect (see Baltagi and Li (2002)) or a random effect setting (see Li and Stengos (1996)).

3 Asymptotic properties and the "oracle" efficient estimator

In this section we investigate some preliminary asymptotic properties of our estimator. In order to do so, we need the following assumptions

(B.1) Let $f_{Z_{1t}}(z_1)$, $f_{Z_{1t},Z_{1(t-1)}}(z_1,z_2)$, $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(z_1,z_2,z_3)$ be respectively the probability density functions of Z_{1t}, $\left(Z_{1t},Z_{1(t-1)}\right)$ and $\left(Z_{1t},Z_{1(t-1)},Z_{1(t-2)}\right)$. Then for any point z_1, (z_1,z_2), (z_1,z_2,z_3) respectively in the support of $f_{Z_{1t}}(z_1)$, $f_{Z_{1t},Z_{1(t-1)}}(z_1,z_2)$ and $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(z_1,z_2,z_3)$, $f_{Z_{1t}}$, $f_{Z_{1t},Z_{1(t-1)}}$ and $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}$ are continuously differentiable in all their arguments. Furthermore, $f_{Z_{1t}}(z_1)>0$, $f_{Z_{1t},Z_{1(t-1)}}(z_1,z_2)>0$, $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(z_1,z_2,z_3)>0$.

(B.2) Let z be an interior point in the support of $f_{Z_{1t}}$. All third-order derivatives of $m_1(\cdot),m_2(\cdot),\cdots,m_d(\cdot)$ are continuous.

(B.3) Let (z_1,z_2) be an interior point in the support of $f_{Z_{1t},Z_{1(t-1)}}(z_1,z_2)$. All conditional moment functions $E\left[X_{it}X_{it}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$, $E\left[X_{it}X_{i(t-1)}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$, $E\left[X_{i(t-1)}X_{i(t-1)}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$, $E\left[\Delta X_{it}\Delta X_{it}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$, $E\left[\Delta X_{it}X_{it}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$, $E\left[\Delta X_{it}X_{i(t-1)}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$ have compact support and they have continuous first order derivatives in all their arguments. Furthermore, let (z_1,z_2,z_3) be an interior point in the support of $f_{Z_{1t},Z_{1(t-1)},Z_{1(t-2)}}(z_1,z_2,z_3)$. The conditional moment functions $E\left[X_{i(t-1)}X_{i(t-1)}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2,Z_{i(t-2)}=z_3\right]$, $E\left[\Delta X_{it}X_{i(t-1)}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2,Z_{i(t-2)}=z_3\right]$ and $E\left[\Delta X_{it}X_{i(t-2)}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2,Z_{i(t-2)}=z_3\right]$ have also compact support and continuous first order derivatives in all their arguments.

(B.4) The kernel function K is symmetric about zero and compactly supported, bounded kernel such that $\int uu^TK(u)du=\mu_2(K)I$, where $\mu_2(K)\neq 0$ is a scalar and I is a $d\times d$ identity matrix.

(B.5) The moment function

$$\frac{1}{T}\sum_t E\left[\Delta X_{it}\Delta X_{it}^T\middle| Z_{it}=z_1,Z_{i(t-1)}=z_2\right]$$

is definite positive in any interior point (z_1,z_2) in the support of $f_{Z_{1t},Z_{1(t-1)}}(z_1,z_2)$.

Assumptions (B.1) and (B.5) give some smoothness conditions on the functions involved. Assumption (B.4) includes some particularities of the kernel function. Under these assumptions we now establish some results on the conditional mean and the conditional variance of the local linear least squares estimator.

Theorem 3.1 *Under Assumptions (B.1)-(B.5) then, if $H\to 0$ such that $N|H|\to\infty$ as N tends to infinity and T is fixed we get*

$$E\left\{\widehat{m}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\}-m(z_0)=\left(\frac{1}{T}\sum_t\mathcal{B}_t^{\Delta X\Delta X}(z_0,z_0)\right)^{-1}$$

$$\times \frac{1}{T} \sum_t \sum_d \left\{ \mu_2 (K_u) \mathcal{B}_{dt}^{\Delta XX} (z_0, z_0) - \mu_2 (K_v) \mathcal{B}_{dt}^{\Delta XX_{-1}} (z_0, z_0) \right\} \times tr \{ \mathcal{H}_{md} (z_0) H \} + o_p (tr \{H\}),$$

and

$$Var \{ \widehat{m}(z_0; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} = \frac{2\sigma_v^2 R(K_u) R(K_v)}{N |H|} \left(\frac{1}{T} \sum_t \mathcal{B}_t^{\Delta X \Delta X} (z_0, z_0) \right)^{-1} \{1 + o_p(1)\},$$

where

$$\begin{aligned} \mathcal{B}_{dt}^{\Delta XX} (z_0, z_0) &= E \left[\Delta X_{it} X_{dit} \right| Z_{it} = z_0, Z_{i(t-1)} = z_0 \Big] f_{Z_{it}, Z_{i(t-1)}} (z_0, z_0), \\ \mathcal{B}_{dt}^{\Delta XX_{-1}} (z_0, z_0) &= E \left[\Delta X_{it} X_{di(t-1)} \right| Z_{it} = z_0, Z_{i(t-1)} = z_0 \Big] f_{Z_{it}, Z_{i(t-1)}} (z_0, z_0), \\ \mathcal{B}_t^{\Delta X \Delta X} (z_0, z_0) &= E \left[\Delta X_{it} \Delta X_{it}^T \right| Z_{it} = z_0, Z_{i(t-1)} = z_0 \Big] f_{Z_{it}, Z_{i(t-1)}} (z_0, z_0). \end{aligned}$$

The proof of this result is done in the Appendix. As it has already been pointed out in other works the leading terms in both bias and variance do not depend on the sample and therefore, we can consider such terms as playing the role of the unconditional bias and variance. Furthermore, we believe that the conditions established on H are sufficient to show that the other terms are $o_p(1)$ and therefore it is possible to show the following result for the asymptotic distribution of $\widehat{m}(z_0; H)$:

Corollary 3.1 *Assume conditions (B.1)-(B.5) hold. Furthermore, there exists a* $\delta > 0$ *such that* $E|Y|^{2+\delta} < \infty$*. Then, if* $H \to 0$ *in such a way that* $N|H| \to \infty$ *and*

$$\max_d \sum_{r=1}^q \sum_{s=1}^q \frac{\partial^2 m_d (z)}{\partial z_r \partial z_s} h_{sr}^2 \sqrt{N |H|} \longrightarrow 0 \tag{7}$$

then

$$\sqrt{N |H|} \{ \widehat{m} (z_0; H) - m (z_0) \} \to_d \mathcal{N} \left(0, 2\sigma_v^2 R (K_u) R (K_v) \left(\frac{1}{T} \sum_t \mathcal{B}_t^{\Delta X \Delta X} (z_0, z_0) \right)^{-1} \right),$$

as N *tends to infinity.*

The proof of this result is shown in Appendix.

Note that the rate at which our estimator converges is $N |H|$. Under the conditions established in the propositions, our estimator is both consistent and asymptotically normal. However, its rate of convergence is sub-optimal since the lower rate of convergence for this type of estimators is $N |H|^{1/2}$. In order to achieve optimality we propose to reduce the dimensionality of the problem by adding a term that cancels asymptotically with one of the two terms and therefore the resulting estimator shows an asymptotically optimal rate. Let

$$\Delta Y_{it}^{(1)} = \Delta Y_{it} + X_{i(t-1)}^T m(Z_{i(t-1)}) \tag{8}$$

substituting (2) into (8) we get

$$\Delta Y_{it}^{(1)} = X_{it}^T m(Z_{it}) + \Delta v_{it}. \tag{9}$$

As it can be realized in (9), estimation of $m(\cdot)$ is now a q-dimensional problem, and therefore we can use again a local linear least squares estimation procedure with kernel weights. If we do so, the optimization problem to solve is the following:

$$min_{\alpha,\delta} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it}^{(1)} - \left(X_{it}^{T}\alpha + X_{it}^{T} \otimes (Z_{it} - z_0)^T \delta \right) \right\}^2 K_H(Z_{it} - z_0). \tag{10}$$

Let $\widetilde{Z}_{it}^{(1)T} = \left[\Delta X_{it}^{T} \quad X_{it}^{T} \otimes (Z_{it} - z_0)^T \right]$ a $1 \times d(1+q)$ vector and $\beta = [\alpha^T, \delta^T]^T$ a $d(1+q) \times 1$ vector, (10) may be written as

$$min_{\beta} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it}^{(1)} - Z_{it}^{(1)T} \beta \right\}^2 K_H(Z_{it} - z_0), \tag{11}$$

and solving this problem we get

$$\widetilde{m}^{(1)}(z_0) = e_1^T \left[\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)} \right]^{-1} \widetilde{Z}^{(1)T} W^{(1)} \Delta Y^{(1)}, \tag{12}$$

where $\widetilde{Z}^{(1)} = [\widetilde{Z}_{12}^{(1)T}, ..., \widetilde{Z}_{NT}^{(1)T}]^T$ is a $N(T-1) \times d(1+q)$ matrix, $e_1 = (I_d \vdots 0_{dqxd})$ a $d(1+q) \times d$ selection matrix with I_d denoting a $d \times d$ identity matrix and 0_{dqxd} a $dq \times d$ matrix of zeros while $W^{(1)}$ is a $N(T-1) \times N(T-1)$ matrix.

Unfortunately $\widetilde{m}^{(1)}(z_0)$ is an unfeasible estimator for $m(\cdot)$. To avoid this problem we replace $m(\cdot)$ in (8) by a consistent estimator, $\widehat{m}(\cdot; H)$, defined in (6) and then

$$\begin{aligned} \Delta \widetilde{Y}_{it}^{(1)} &= \Delta Y_{it} + X_{i(t-1)}^{T} \widehat{m}(Z_{i(t-1)}; H), && (13) \\ \Delta \widetilde{Y}_{it}^{(1)} &= X_{it}^{T} m(Z_{it}) + X_{i(t-1)}^{T} \left[\widehat{m}(Z_{i(t-1)}; H) - m\left(Z_{i(t-1)}\right) \right] + \Delta v_{it}. && (14) \end{aligned}$$

The local linear least squares estimator of $m(z_0)$ is then

$$\widehat{m}^{(1)}(z_0; H) = e_1^T \left(\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)} \right)^{-1} \widetilde{Z}^{(1)T} W^{(1)} \Delta \widetilde{Y}^{(1)}, \tag{15}$$

where $\Delta \widetilde{Y}^{(1)} = \left[\Delta \widetilde{Y}_{12}^{(1)}, ..., \Delta \widetilde{Y}_{NT}^{(1)} \right]^T$, $W^{(1)} = diag\{K_H(Z_{12} - z_0), ..., K_H(Z_{NT} - z_0)\}$ and

$$\widetilde{Z}^{(1)} = \begin{bmatrix} \Delta X_{12}^{T} & X_{12}^{T} \otimes (Z_{12} - z_0)^T \\ \vdots & \vdots \\ \Delta X_{NT}^{T} & X_{NT}^{T} \otimes (Z_{NT} - z_0)^T \end{bmatrix}.$$

In order to show the asymptotic efficiency of our technique we need the following additional assumptions:

(C.1) The kernel function $K \in L_1$ and $\|u\|^8 K(u) \in L_1$.

(C.2) Then for any point z_1, (z_1, z_2), (z_1, z_2, z_3) respectively in the support of $f_{Z_{1t}}(z_1)$, $f_{Z_{1t}, Z_{1(t-1)}}(z_1, z_2)$ and $f_{Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)}}(z_1, z_2, z_3)$, $f_{Z_{1t}} \le C_1 < \infty$, $f_{Z_{1t}, Z_{1(t-1)}} \le C_2 < \infty$ and $f_{Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)}} \le C_3 < \infty$.

Note that these assumptions are devoted to ensure that both bias and variance rates for $\widehat{m}(z_0)$ are uniform (see Masry (1996)). It is then possible to show the following result

Theorem 3.2 *Under Assumptions (B.1)-(B.5), (C.1)-(C.2) then, if $H \to 0$ such that $N|H| \to \infty$ as N tends to infinity and T is fixed we get*

$$E\left\{\widehat{m}^{(1)}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\} - m(z_0) = \left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\right)^{-1}$$

$$\times\frac{\mu_2(K_u)}{T}\sum_t\sum_d\left\{\mathcal{B}_{dt}^{XX}(z_0)\right\} \times tr\{\mathcal{H}_{md}(z_0)H\} + o_p(tr\{H\}),$$

and

$$Var\left\{\widehat{m}^{(1)}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\} = \frac{2\sigma_v^2 R(K_u)}{N|H|^{1/2}}\left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\right)^{-1}\{1+o_p(1)\},$$

where

$$\begin{aligned}
\mathcal{B}_{dt}^{XX}(z_0) &= E\left[X_{it}X_{dit}|\, Z_{it}=z_0\right] f_{Z_{it}}(z_0),\\
\mathcal{B}_t^{XX}(z_0) &= E\left[\Delta X_{it}\Delta X_{it}^T\middle|\, Z_{it}=z_0\right] f_{Z_{it}}(z_0).
\end{aligned}$$

The proof of the Theorem 3.2 is done in the Appendix.

Then, we see that $\widehat{m}^{(1)}(z_0;H)$ enjoys the same asymptotic distribution as $\widetilde{m}^{(1)}(z_0;H)$. This is the so called "oracle" efficiency property.

Finally, focusing on the relevant terms of bias and variance of Theorems 1 and 2 and following Ruppert and Wand (1994) it can be highlighted that each entry of $\mathcal{H}_m(z_0)$ is a measure of the curvature of $m(\cdot)$ at z_0 in a particular direction. Thus, we can intuitively conclude that the bias is increased when there is a higher curvature and more smoothing is well described by this leading bias term. Meanwhile, in terms of the variance we can conclude that it will be penalized by a higher conditional variance of Y given $Z = z_0$ and sparser data near z_0.

4 Bandwidth selection

In this section we propose a method to estimate the variable bandwidth matrix H for each estimator. Theoretically, one option would be to minimize some measure of discrepancy, as the Mean Square Error (MSE), with respect to H so the optimal bandwidth matrix could be obtained as

$$H_{opt} = \arg\min_H MSE\{\widehat{m}(\cdot;H)\},$$

where

$$MSE\left\{\widehat{m}(\cdot;H)\right\} = E\left\{\sum_{l=1}^{d}\left\{\widehat{m}_l(Z_{it}) - m_l(Z_{it})\right\}X_{lit}\right\}^2.$$

Unfortunately, there are some terms of the MSE unknown so this result is not empirically useful and it is necessary to resort to some alternative approach. In this way and following to Zhang and Lee (2000) we propose to get an estimator of the H optimal that minimize an estimation of the MSE and that we call the estimated variable bandwidth matrix ($\widehat{H}$).

Let $\mathbb{X}, \mathbb{Z}$ be vectors of random variables such that $\mathbb{X} = (X_1, \cdots, X_d)$ and $\mathbb{Z} = (Z_1, \cdots, Z_q)$, and $\mathfrak{D}$ the observed covariates vector, $\mathfrak{D} = (Z_{111}, \cdots, Z_{1NT}, \cdots, Z_{q11}, \cdots, Z_{qNT}, X_{111}, \cdots, X_{1NT}, \cdots, X_{d11}, \cdots, X_{dNT})^T$. Moreover, let $bias\left(\widehat{m}_l(z_0;H)|\mathfrak{D}\right)$ be the conditional bias of $\widehat{m}_l\left(z_0;H\right)$ given $\mathfrak{D}$, $b(z_0;H) = bias\left(\widehat{m}(z_0;H)|\mathfrak{D}\right)$ and the proposed MSE is

$$\begin{aligned} MSE\left\{\widehat{m}(.)\right\} &= E\left(trE\left[\left\{\widehat{m}(Z_{it};H) - m\left(Z_{it}\right)\right\}^T X_{it}X_{it}^T\left\{\widehat{m}(Z_{it};H) - m\left(Z_{it}\right)\right\}\Big|\,\mathbb{Z}\right]\right) \\ &= E\left(b^T(\mathbb{Z};H)\Omega(\mathbb{Z})b(\mathbb{Z};H) + tr\left\{\Omega(\mathbb{Z})Var\left\{\widehat{m}(\mathbb{Z};H)|\mathbb{Z},\mathfrak{D}\right\}\right\}\right), \end{aligned}$$

where $\Omega(\mathbb{Z}) = E\left(X_{it}X_{it}^T|\mathbb{Z}\right)$. Note that

$$Var\left\{\widehat{m}\,\left(\mathbb{Z};H\right)|\,\mathbb{Z},\mathfrak{D}\right\} = Var\left\{\widehat{m}\,\left(z_0;H\right)|\,\mathfrak{D}\right\}|_{\mathbb{Z}=z_0} \quad \text{and} \quad MSE\left\{\widehat{m}(.)\right\} = E\left[E\left\{MSE\left(\widehat{m}(z_0;H)|\mathfrak{D}\right\}|_{\mathbb{Z}=z_0}\,|\mathfrak{D}\right],$$

so we may define the MSE for the local weighted least squares estimator as

$$MSE\left\{\widehat{m}(z_0;H)|\mathfrak{D}\right\} = b^T(z_0;H)\Omega(z_0)b(z_0;H) + tr\left\{\Omega(z_0)Var\left\{\widehat{m}(z_0;H)|\mathfrak{D}\right\}\right\}.$$

However, given that $MSE\left\{\widehat{m}(z_0;H)|\mathfrak{D}\right\}$ also depends on some unknown quantities, to obtain $\widehat{H}$ we need to estimate both the conditional bias and variance previously. Then, taking the idea of Zhang and Lee (2000) as benchmark we need an additional assumption:

(D.1) Let z be an interior point in the support of $f_{Z_{1t}}$. All fifth-order derivatives of $m_1(\cdot), m_2(\cdot), \cdots, m_d(\cdot)$ are continuous.

Thus, in the bandwidth selector procedure that we propose we first calculate the conditional bias based on a Taylor expansion of order $(g+1)$. Combining this approximation with (6)

$$E\left\{\widehat{m}(z_0;H)|\mathfrak{D}\right\} - m(z_0) = e_1^T\left(\widetilde{Z}^TW\widetilde{Z}\right)^{-1}\widetilde{Z}^TW\left(\frac{1}{2}Q_b(z_0)D_{b2}(z_0) + \frac{1}{3!}Q_b^{(g+1)}(z_0)D_{b(g+1)}(z_0) + R_b(z_0)\right),$$

where $D_{b2}(z_0) = vec\left(\mathcal{H}_m^{(2)}(z_0)\right)$ and $D_{b(g+1)}(z_0) = vec\left(\mathcal{H}_m^{(g+1)}(z_0)\right)$ are dq^2 and $dq^{(g+1)}$ vectors that contains the vec operator of the Hessian and the (g+1)th derivative matrix of the d-th component of $m(\cdot)$, respectively, and

$$\begin{aligned} Q_b(z_0) &= S_{b1}(z_0) - S_{b2}(z_0), & Q_b^{(g+1)}(z_0) &= S_{b1}^{(g+1)}(z_0) - S_{b2}^{(g+1)}(z_0) \\ S_{b1}(z_0) &= \left[S_{b1,12}^T(z_0), ..., S_{b1,NT}^T(z_0)\right]^T, & S_{b1}^{(g+1)}(z_0) &= \left[S_{b1,12}^{(g+1)T}(z_0), ..., S_{b1,NT}^{(g+1)T}(z_0)\right]^T, \\ S_{b2}(z_0) &= \left[S_{b2,11}^T(z_0), ..., S_{b2,N(T-1)}^T(z_0)\right]^T, & S_{b2}^{(g+1)}(z_0) &= \left[S_{b2,11}^{(g+1)T}(z_0), ..., S_{b2,N(T-1)}^{(g+1)T}(z_0)\right]^T, \end{aligned}$$

with

$$
\begin{aligned}
S_{b1,it}(z_0) &= \left[\{X_{it} \otimes (Z_{it} - z_0) \otimes (Z_{it} - z_0)\}^T\right], \\
S_{b2,i(t-1)}(z_0) &= \left[\left\{X_{i(t-1)} \otimes \left(Z_{i(t-1)} - z_0\right) \otimes \left(Z_{i(t-1)} - z_0\right)\right\}^T\right], \\
S^{(g+1)}_{b1,it}(z_0) &= \left[\left\{X_{it} \otimes (Z_{it} - z_0)^{(g+1)}\right\}^T\right], \\
S^{(g+1)}_{b2,i(t-1)}(z_0) &= \left[\left\{X_{i(t-1)} \otimes \left(Z_{i(t-1)} - z_0\right)^{(g+1)}\right\}^T\right].
\end{aligned}
$$

Thereby, the conditional bias to estimate is

$$
b(z_0; H) = e_1^T \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W \widetilde{Z}_{(g+1)} D, \tag{16}
$$

where

$$
\widetilde{Z}_{(g+1)} =
$$

$$
\left[\begin{array}{cccc}
\{S_{b1,12}(z_0) - S_{b2,11}(z_0)\} & \{S^3_{b1,12}(z_0) - S^3_{b2,11}(z_0) & \cdots & \{S^{(g+1)}_{b1,12}(z_0) - S^{(g+1)}_{b2,11}(z_0)\} \\
\vdots & \vdots & \ddots & \vdots \\
\{S_{b1,NT}(z_0) - S_{b2,N(T-1)}(z_0)\} & \{S^3_{b1,NT}(z_0) - S^3_{b2,N(T-1)}(z_0)\} & \cdots & \{S^{(g+1)}_{b1,NT}(z_0) - S^{(g+1)}_{b2,N(T-1)}(z_0)\}
\end{array}\right],
$$

and

$$
D = \left[\begin{array}{cccc} D^T_{b2}(z_0), & D^T_{b3}(z_0), & \cdots, & D^T_{b(g+1)}(z_0) \end{array}\right]^T
$$

is a vector of unknown functions that need to be estimated. For convenience we take $g = 2$ so $\widetilde{Z}_3$ is a $N(T-1) \times dq^2(1+q)$ matrix and D_3 a $dq^2(1+q) \times 1$ vector.

To estimate the vector D_{3b} we use a local polynomial regression of order $g = 3$ ($g > 1$) with a bandwidth matrix H_* so

$$
\widehat{D}_{bk} = e_k^T \left(\widetilde{Z}^T_{(g+1)} W_* \widetilde{Z}_{(g+1)}\right)^{-1} \widetilde{Z}^T_{(g+1)} W_* \Delta Y, \qquad k = 2, \cdots, 4 \tag{17}
$$

where $W_* = diag\left\{K_{H_*}(Z_{12} - z_0) K_{H_*}(Z_{11} - z_0), \cdots, K_{H_*}(Z_{NT} - z_0) K_{H_*}(Z_{N(T-1)} - z_0)\right\}$ is a $N(T-1) \times N(T-1)$ matrix, $e_k^T = \left(0_{dq^k \times d(1+\sum_{g=1}^{k-1} q^g)} \vdots I_{dq^k \times dq^k} \vdots 0_{dq^k \times d(\sum_{g=k+1}^{4} q^g)}\right)$ is a $dq^k \times d(1 + \sum_{g=1}^{4} q^g)$ vector where I is an identity matrix and 0 a matrix of zeros.

Note that the initial bandwidth matrix H_* can be obtained by the minimizer of some residual squares criterion (RSC), as Fan and Gijbels (1995), and then the conditional bias can be already estimated.

Secondly, we calculate the conditional variance that can be written as

$$
Var\left\{\widehat{m}\left(z_0; H\right) | \mathfrak{D}\right\} = e_1^T \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W \mathcal{V} W \widetilde{Z} \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} e_1
$$

where $\mathcal{V} = E\left(\Delta v \Delta v^T | \mathfrak{D}\right)$ is a $N(T-1) \times N(T-1)$ matrix that contains the V_{ij}'s matrices described in (31). Using the information of (31), $\mathcal{V} = \sigma_v^2 A$ where A is a $N(T-1) \times N(T-1)$ matrix of constant, the variance can be rewritten as

$$
Var\left\{\widehat{m}(z_0; H) | \mathfrak{D}\right\} = e_1^T \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W A W \widetilde{Z} \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} e_1 \sigma_v^2. \tag{18}
$$

To calculate (18) we need to estimate the unknown quantity σ_v^2 for which we use the following normalized weighted residual sum of squares from a $(g+1)$th-order polynomial fit

$$\widehat{\sigma}_v^2 = \frac{1}{tr(W_*) - tr\left\{(\widetilde{Z}_{(g+1)}^T W_* \widetilde{Z}_{(g+1)})^{-1} \widetilde{Z}_{(g+1)}^T W_* A W_* \widetilde{Z}_{(g+1)}\right\}} \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta Y_{it} - \Delta \widehat{Y}_{*it}\right)^2 K_{H_*}(Z_{it} - z_0) K_{H_*}(Z_{i(t-1)} - z_0), \tag{19}$$

where $\Delta Y_{*it} = \widetilde{Z}_{(g+1)it}^T D + \Delta v_{*it}$, $g = 3$, Δv_{*it} is the idiosyncratic error term and

$$\Delta \widehat{Y}_* = \left[\Delta \widehat{Y}_{*12}, ..., \Delta \widehat{Y}_{*NT}\right]^T = \widetilde{Z}_{(g-1)} \left[\widetilde{Z}_{(g+1)}^T W_* \widetilde{Z}_{(g+1)}\right]^{-1} \widetilde{Z}_{(g+1)}^T W_* \Delta Y.$$

Once known σ_v^2 the conditional variance can be already calculated.

To get the $\Omega(z_0)$ matrix it is necessary to estimate the $r_{ll'}(z_0)$ element of this matrix, where $l, l' = 1, ..., d$ and $\Omega(z_0) = E(X_{lit} X_{l'it}^T | \mathbb{Z} = z_0)$, through a fully nonparametric model. For that, we can use some standard nonparametric techniques such as the Nadaraya-Watson estimator (see Härdle (1990)).

Finally, using (16)-(19) we obtain the following estimate of $MSE\{\widehat{m}(z_0; H)|\mathfrak{D}\}$,

$$\begin{aligned} \widehat{MSE}\{\widehat{m}(z_0; H)|\mathfrak{D}\} = \widehat{D}^T \widetilde{Z}_{(g+1)}^T W \widetilde{Z} \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \left(\widehat{\Omega} \otimes e_{1+q} e_{1+q}^T\right) \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W \widetilde{Z}_{(g+1)} \widehat{D} + \\ tr\left\{\left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W A W \widetilde{Z} \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \left(\widehat{\Omega} \otimes e_{(1+q)} e_{(1+q)}^T\right) \widehat{\sigma}_v^2\right\} \end{aligned} \tag{20}$$

for $(g+1)$ and the minimizer of this expression is $\widehat{H}$, the estimated variable bandwidth matrix for the local lineal estimator with kernel weights.

Meanwhile, to get the estimated variable bandwidth matrix for the "oracle" estimator we adapt the prior procedure to the particularities of this estimator so now $\widehat{H}$ is selected solving

$$\widehat{H} = \arg\min_H \widehat{MSE}\left\{\widehat{m}^{(1)}(z_0; H)|\mathfrak{D}\right\},$$

where for $(g = 2)$,

$$\begin{aligned} & \widehat{MSE}\left\{\widehat{m}^{(1)}(z_0; H)|\mathfrak{D}\right\} \\ = \;& E\left[b^{(1)T}(z_0; H)\Omega(z_0) b^{(1)}(z_0; H) + tr\left\{\Omega(z_0) Var\{\widehat{m}^{(1)}(z_0; H)|\mathfrak{D}\}\right\}\right] = \\ = \;& \widehat{D}^{(1)T} \widetilde{Z}_{(g+1)}^{(1)T} W^{(1)} \widetilde{Z}^{(1)} \left[\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)}\right]^{-1} \left(\widehat{\Omega} \otimes e_{1+q} e_{1+q}^T\right) \left[\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)}\right]^{-1} \widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}_{(g+1)}^{(1)} \widehat{D}^{(1)} \\ + \;& tr\left\{\left[\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)}\right]^{-1} \widetilde{Z}^{(1)T} W^{(1)} \left(A \widehat{\sigma}_v^2 + \widehat{\mathcal{V}}\right) W^{(1)} \widetilde{Z}^{(1)} \left[\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)}\right]^{-1} \left(\widehat{\Omega} \otimes e_{(1+q)} e_{(1+q)}^T\right)\right\} \end{aligned}$$

being $\widehat{\mathcal{V}} = diag\{E\left(\Delta\widehat{v}_1 \Delta\widehat{v}_1^T|\mathfrak{D}\right), \cdots, E\left(\Delta\widehat{v}_N \Delta\widehat{v}_N^T|\mathfrak{D}\right)\}$, $\Delta\widehat{v}_{it} = X_{i(t-1)}^T\left\{\widehat{m}\left(Z_{i(t-1)}\right) - E\left(\widehat{m}\left(Z_{i(t-1)}\right)|\mathfrak{D}\right)\right\}$ and $b^{(1)}(z_0; H) = bias\left(\widehat{m}^{(1)}(z_0; H)|\mathfrak{D}\right)$.

5 Monte Carlo experiment

In this section we report some Monte Carlos simulation results to examine whether the proposed estimators perform reasonably well in finite samples when μ_i are fixed effects.

We consider the following varying-coefficient nonparametric models,

$$Y_{it} = \mu_i + X_{dit}^T m\left(Z_{qit}\right) + v_{it}, \quad i = 1, \cdots, N; \quad t = 1, \cdots, T; \quad d, q = 1, 2$$

where X_{dit} and Z_{qit} are scalars random variables; v_{it} is an i.i.d.N(0,1) random variable; and m(.) is a pre-specified function to be estimated. The observations follow a data generating process where $Z_{qit} = w_{qit} + w_{qi(t-1)}$, being w_{qit} an i.i.d. uniformly distributed $[0, \Pi/2]$ random variable; and $X_{dit} = 0.5X_{di(t-1)} + \xi_{it}$, with ξ_{it} being an i.i.d.N(0,1).

We consider three different cases of study,

$$\begin{aligned}
(1)\ Y_{it} &= X_{1it} m_1\left(Z_{1it}\right) + \mu_{1i} + v_{it} \\
(2)\ Y_{it} &= X_{1it} m_1\left(Z_{1it}, Z_{2it}\right) + \mu_{2i} + v_{it} \\
(3)\ Y_{it} &= X_{1it} m_1\left(Z_{1it}\right) + X_{2it} m_2\left(Z_{2it}\right) + \mu_{1i} + v_{it},
\end{aligned}$$

where the chosen functionals form are $m_1\left(Z_{1it}\right) = \sin\left(Z_{1it}\Pi\right)$, $m_1\left(Z_{1it}, Z_{2it}\right) = \sin\left(\left(Z_{1it}, Z_{2it}\right)\Pi\right)$ and $m_2\left(Z_{1it}\right) = \exp\left(-Z_{1it}^2\right)$; and we experiment with two specifications for the fixed effects,

(a) μ_{1i} depends on Z_{1it}, where the dependence is imposed by generating $\mu_{1i} = c_0\overline{Z}_{1i.} + u_i$ for $i = 2, \cdots, N$ and $\overline{Z}_{1i.} = T^{-1}\sum_t Z_{1it}$.
(b) μ_{2i} depends on Z_{1it}, Z_{2it} through the generating process $\mu_{2i} = c_0\overline{Z}_{i.} + u_i$ for $i = 2, \cdots, N$ and $\overline{Z}_{i.} = \frac{1}{2}\left(\overline{Z}_{1i.} + \overline{Z}_{2i.}\right)$.

where in both cases u_i is an i.i.d.$N(0,1)$ random variable and $c_0 = 0.5$ controls de correlation between the unobservable individual heterogeneity and some of the regressors of the model.

In the experiment we use 1000 Monte Carlo replications (Q). The number of period (T) is fixed at three, while the number of cross-sections (N) is varied to be 50, 100 and 200. In addition, the Gaussian kernel has been used and, as Henderson et al. (2008), the bandwidth is chosen as $H = \widehat{\sigma}_z(N(T-1))^{-1/5}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\{Z_{qit}\}_{i=1,t=2}^{N,T}$.

We report estimation results for both proposed estimators and as measure of their estimation accuracy we use the Integrated Squared Error (ISE). Thus, denoting the subscript r the rth replication,

$$ISE\left\{\widehat{m}_l\left(z_0; H\right)\right\} = \frac{1}{Q}\sum_{r=1}^{Q}\int\left\{\sum_{l=1}^{d}\left(\widehat{m}_{lr}\left(z_0; H\right) - m_{lr}\left(z_0\right)\right)X_{it,lr}\right\}^2 f(z_0)dz_0$$

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

$$AMSE\left\{\widehat{m}\left(z_0; H\right)\right\} = \frac{1}{Q}\sum_{r=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=2}^{T}\left\{\sum_{l=1}^{d}\left\{\widehat{m}_{lr}(z_0; H) - m_{lr}(z_0)\right\}X_{it,lr}\right\}^2,$$

The simulations results are summarized in Table 1, 2 and 3, respectively.

Table 1: AMSE for d=1 and q=1.

	$m_1(z)$	
	Local polynomial	Backfitting
n=150	1.2917536	1.2862442
n=300	0.9158168	0.915888
n=600	0.7278608	0.7136175

Table 2: AMSE for d=1 and q=2.

	$m_1(z)$	
	Local polynomial	Backfitting
n=150	1.7953523	1.5848431
n=300	1.1097905	1.0171259
n=600	0.8181966	0.7656307

Table 3: AMSE for d=2 and q=1.

	$m_1(z)$		$m_2(z)$	
	Local polynomial	Backfitting	Local polynomial	Backfitting
n=150	1.3454139	1.3072503	0.8982404	0.7947426
n=300	0.929184	0.9227399	0.4907434	0.490465
n=600	0.7305442	0.7217713	0.3002903	0.3309637

We further carried out a simulation study to analyze the behavior in finite samples of the multivariate locally estimator with kernels weights, $\widehat{m}\left(z_0; H\right)$, and the "oracle" estimator, $\widehat{m}^{(1)}\left(z_0; H\right)$, proposed in Sections 2 and 3. Looking at Tables 1, 2 and 3 we can highlight the following.

On one hand, as the proposed estimators are based on a first difference transformation, the bias and the variance of both estimators do not depend on the values of the fixed effects so their estimation accuracy are the same for different values of c_0.

On the other hand, from Tables 1, 2 and 3 we can see that both estimators carry out quite well. For all T, as N increases the AMSE of both estimators are lower, as we expected. This is due to the asymptotic properties of the estimators described previously. In addition, these results also allow us to test the hypothesis that the "oracle" estimator generates an improvement in the rate of convergence. Specifically, for the univariate case, Tables 1 and 3, we may appreciate that the achievement of both estimators are quite similar while, on the contrary, in the multivariate case, Table 2, the rate of convergence of the "oracle" estimator is faster that the multivariate locally estimator as we expected. In addition, as we can see in Table 2 results of the local polynomial estimator reflect the "curse of dimensionality" property given that as the dimensionality of Z_{it} increases the AMSE is greater. Thus, the backfitting estimator has an efficiency gain over the local polynomial estimator, as we suspect.

6 Conclusion

This paper introduces a new technique that estimates varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in a unknown way. The resulting estimator is robust to misspecification in the functional form of the varying parameters and we have shown that it is consistent and asymptotically normal. Furthermore we have shown that it achieves the optimal rate of convergence for this type of problems and it exhibits the so called "oracle" efficiency property. Since the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. Monte Carlo results indicates good performance of the estimator in finite samples.

7 Appendix

Proof of Theorem 3.1

Taking conditional expectations in (6) and noting that

$$E(v_{it}|X_{11},...,X_{NT},Z_{11},...,Z_{NT})=0,\quad t=2,\cdots,T,i=1,\cdots,N$$

then

$$E\{\widehat{m}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\}=e_1^T(\widetilde{Z}^TW\widetilde{Z})^{-1}\widetilde{Z}^TWM \tag{21}$$

where $M=\left[\left\{X_{12}^Tm(Z_{12})-X_{11}^Tm(Z_{11})\right\}^T,...,\left\{X_{NT}^Tm(Z_{NT})-X_{N(T-1)}^Tm(Z_{N(T-1)})\right\}^T\right]^T$.

Taylor's Theorem implies that

$$M=\widetilde{Z}\left[\begin{array}{c} m(z_0) \\ D_m(z_0) \end{array}\right]+\frac{1}{2}Q_m(z_0)+R(z_0), \tag{22}$$

where

$$Q_m(z_0)=S_{m1}(z_0)-S_{m2}(z_0), \tag{23}$$

$$\begin{aligned} S_{m1}(z_0) &= [S_{m1,12}^T(z_0),...,S_{m1,NT}^T(z_0)]^T, \\ S_{m2}(z_0) &= [S_{m2,11}^T(z_0),...,S_{m2,N(T-1)}^T(z_0)]^T \end{aligned}$$

and

$$\begin{aligned} S_{m1,it}(z_0) &= \left[\{X_{it}\otimes(Z_{it}-z_0)\}^T\mathcal{H}_m(z_0)(Z_{it}-z_0)\right], \\ S_{m2,i(t-1)}(z_0) &= \left[\{X_{i(t-1)}\otimes(Z_{i(t-1)}-z_0)\}^T\mathcal{H}_m(z_0)(Z_{i(t-1)}-z_0)\right]. \end{aligned}$$

We denote by

$$\mathcal{H}_m(z)=\left(\begin{array}{c} \mathcal{H}_{m1}(z) \\ \mathcal{H}_{m2}(z) \\ \vdots \\ \mathcal{H}_{md}(z) \end{array}\right),$$

a $dq \times q$ matrix such that $\mathcal{H}_{md}(z)$ is the Hessian matrix of the d-th component of $m(\cdot)$ and $R(z)$ is a vector of Taylor series remainder terms. Furthermore, as it is already pointed out in Ruppert and Wand (1994), using (B.1) we get

$$e_1^T(\widetilde{Z}^T W \widetilde{Z})^{-1}\widetilde{Z}^T W R(z_0) = o_p\left(\mathrm{tr}\left\{H\right\}\right),$$

and then,

$$E\left\{\widehat{m}(z_0; H)|X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT}\right\} - m(z_0) = \frac{1}{2}e_1^T(\widetilde{Z}^T W \widetilde{Z})^{-1}\widetilde{Z}^T W Q_m(z_0). \tag{24}$$

Notice that expression $D_m(z_0)$ in (22) vanishes since

$$e_1^T[\widetilde{Z}^T W \widetilde{Z}]^{-1}\widetilde{Z}^T W \begin{bmatrix} m(z_0) \\ D_m(z_0) \end{bmatrix} = e_1^T \begin{bmatrix} m(z_0) \\ D_m(z_0) \end{bmatrix} = m(z_0). \tag{25}$$

For the sake of simplicity let us denote

$$K_{it} = \frac{1}{|H|^{1/2}} K\left(H^{-1/2}\left(Z_{it} - z_0\right)\right),$$

now, define the symmetric block matrix

$$(NT)^{-1}\widetilde{Z}^T W \widetilde{Z} = \begin{pmatrix} \mathcal{A}_{NT}^{11} & \mathcal{A}_{NT}^{12} \\ \mathcal{A}_{NT}^{21} & \mathcal{A}_{NT}^{22} \end{pmatrix} \tag{26}$$

where,

$$\begin{aligned}
\mathcal{A}_{NT}^{11} &= (NT)^{-1}\sum_{it} \Delta X_{it}\Delta X_{it}^T K_{it}K_{i(t-1)}, \\
\mathcal{A}_{NT}^{12} &= (NT)^{-1}\sum_{it} \Delta X_{it}\left\{X_{it}^T \otimes (Z_{it} - z_0)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z_0)^T\right\} K_{it}K_{i(t-1)}, \\
\mathcal{A}_{NT}^{21} &= (NT)^{-1}\sum_{it} \left\{X_{it} \otimes (Z_{it} - z_0) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z_0)\right\} \Delta X_{it}^T K_{it}K_{i(t-1)}, \\
\mathcal{A}_{NT}^{22} &= (NT)^{-1}\sum_{it} \left\{X_{it} \otimes (Z_{it} - z_0) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z_0)\right\} \times \\
&\quad \left\{X_{it}^T \otimes (Z_{it} - z_0)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z_0)^T\right\} K_{it}K_{i(t-1)}.
\end{aligned}$$

Using standard results from density estimation and as the variables are i.i.d. in the subscript i we can show that, as N tends to infinity,

$$\mathcal{A}_{NT}^{11} = \frac{1}{T}\sum_t \mathcal{B}_t^{\Delta X \Delta X}(z_0, z_0) + o_p(1), \tag{27}$$

and

$$\mathcal{B}_t^{\Delta X \Delta X}(z_0, z_0) = E\left[\Delta X_{it}\Delta X_{it}^T | Z_{it} = z_0, Z_{i(t-1)} = z_0\right] f_{Z_{it}, Z_{i(t-1)}}(z_0, z_0),$$

is a $d \times d$ matrix, for $t = 2, \cdots, T$.
Furthermore, using assumptions (B.1) to (B.5), as N tends to infinity

$$\mathcal{A}_{NT}^{12} = \frac{1}{T}\sum_t \left\{\mathcal{D}\mathcal{B}_t^{\Delta X X}(z_0, z_0)\left(I_d \otimes \mu_2\left(K_u\right) H\right) - \mathcal{D}\mathcal{B}_t^{\Delta X X_{-1}}(z_0, z_0)\left(I_d \otimes \mu_2\left(K_v\right) H\right)\right\} + o_p(H). \tag{28}$$

$\mathcal{DB}_t^{\Delta XX}(Z_1, Z_2)$ and $\mathcal{DB}_t^{\Delta XX_{-1}}(Z_1, Z_2)$ are respectively $d \times dq$ gradient matrices defined as

$$\mathcal{DB}_t^{\Delta XX}(Z_1, Z_2) = \begin{pmatrix} \frac{\partial b_{t11}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^T} & \cdots & \frac{\partial b_{t1d}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^T} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{td1}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^T} & \cdots & \frac{\partial b_{tdd'}^{\Delta XX}(Z_1,Z_2)}{\partial Z_1^T} \end{pmatrix},$$

and

$$b_{tdd'}^{\Delta XX}(Z_1, Z_2) = E\left[\Delta X_{dit} X_{d'it} \middle| Z_{it} = Z_1, Z_{i(t-1)} = Z_2\right] f_{Z_{it}, Z_{i(t-1)}}(Z_1, Z_2).$$

The other gradient matrix is

$$\mathcal{DB}_t^{\Delta XX_{-1}}(Z_1, Z_2) = \begin{pmatrix} \frac{\partial b_{t11}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^T} & \cdots & \frac{\partial b_{t1d}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^T} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{td1}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^T} & \cdots & \frac{\partial b_{tdd}^{\Delta XX_{-1}}(Z_1,Z_2)}{\partial Z_1^T} \end{pmatrix},$$

and

$$b_{tdd'}^{\Delta XX_{-1}}(Z_1, Z_2) = E\left[\Delta X_{dit} X_{d'it-1} \middle| Z_{it} = Z_1, Z_{i(t-1)} = Z_2\right] f_{Z_{it}, Z_{i(t-1)}}(Z_1, Z_2).$$

Finally,

$$\mathcal{A}_{NT}^{22} = \frac{1}{T}\sum_t \left[\mathcal{B}_t^{XX}(z_0, z_0) \otimes \mu_2(K_u) H + \mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0) \otimes \mu_2(K_v) H\right] + o_p(H), \tag{29}$$

where

$$\mathcal{B}_t^{XX}(z_0, z_0) = E\left[X_{it} X_{it}^T \middle| Z_{it} = z_0, Z_{i(t-1)} = z_0\right] f_{Z_{it}, Z_{i(t-1)}}(z_0, z_0),$$

and

$$\mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0) = E\left[X_{i(t-1)} X_{i(t-1)}^T \middle| Z_{it} = z_0, Z_{i(t-1)} = z_0\right] f_{Z_{it}, Z_{i(t-1)}}(z_0, z_0).$$

Using the the results shown in (27), (28) and (29) we obtain

$$NT\left(\tilde{Z}^T W \tilde{Z}\right)^{-1} = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} \\ \mathcal{C}_{21} & \mathcal{C}_{22} \end{pmatrix}, \tag{30}$$

where

$$\begin{aligned}
\mathcal{C}_{11} &= \left(\frac{1}{T}\sum_t \mathcal{B}_t^{\Delta X \Delta X}(z_0, z_0)\right)^{-1} + o_p(1), \\
\mathcal{C}_{12} &= -\left(\frac{1}{T}\sum_t \mathcal{B}_t^{\Delta X \Delta X}(z_0, z_0)\right)^{-1} \left[\frac{1}{T}\sum_t \left\{\mathcal{DB}_t^{\Delta XX}(z_0, z_0)\left(I_d \otimes \mu_2\left(K_u\right) H\right) - \mathcal{DB}_t^{\Delta XX_{-1}}(z_0, z_0)\left(I_d \otimes \mu_2\left(K_v\right) H\right)\right\}\right] \\
&\quad \times \left(\frac{1}{T}\sum_t \left[\mathcal{B}_t^{XX}(z_0, z_0) \otimes \mu_2\left(K_u\right) H + \mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0) \otimes \mu_2\left(K_v\right) H\right]\right)^{-1} + o_p(1), \\
\mathcal{C}_{21} &= \left(\frac{1}{T}\sum_t \left[\mathcal{B}_t^{XX}(z_0, z_0) \otimes \mu_2\left(K_u\right) H + \mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0) \otimes \mu_2\left(K_v\right) H\right]\right)^{-1} \\
&\quad \times \left[\frac{1}{T}\sum_t \left\{\mathcal{DB}_t^{\Delta XX}(z_0, z_0)\left(I_d \otimes \mu_2\left(K_u\right) H\right) - \mathcal{DB}_t^{\Delta XX_{-1}}(z_0, z_0)\left(I_d \otimes \mu_2\left(K_v\right) H\right)\right\}\right]^T \left(\frac{1}{T}\sum_t \mathcal{B}_t^{\Delta X \Delta X}(z_0, z_0)\right)^{-1} \\
&\quad + o_p(1), \\
\mathcal{C}_{22} &= \left(\frac{1}{T}\sum_t \left[\mathcal{B}_t^{XX}(z_0, z_0) \otimes \mu_2(K_u) H + \mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0) \otimes \mu_2(K_v) H\right]\right)^{-1} + o_p\left(H^{-1}\right).
\end{aligned}$$

Also it is straightforward to show that the terms in

$$(NT)^{-1}\widetilde{Z}^T W S_{m1}(z_0)$$

$$= \begin{pmatrix} (NT)^{-1}\sum_{it}\Delta X_{it}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0)K_{it}K_{i(t-1)} \\ \\ (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z_0)\right\}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0)K_{it}K_{i(t-1)} \end{pmatrix},$$

are asymptotically equal to

$$\begin{aligned} & (NT)^{-1}\sum_{it}\Delta X_{it}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0)K_{it}K_{i(t-1)} \\ = \; & \frac{\mu_2\left(K_u\right)}{T}\sum_t\sum_d E\left[\Delta X_{it}X_{dit}\middle|\, Z_{it}=z_0, Z_{i(t-1)}=z_0\right] f_{Z_{it},Z_{i(t-1)}}\left(z_0,z_0\right)\times\operatorname{tr}\left\{\mathcal{H}_{md}\left(z_0\right)H\right\}+o_p\left(\operatorname{tr}\left\{H\right\}\right), \end{aligned}$$

and

$$\begin{aligned} & (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z_0)\right\}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0)K_{it}K_{i(t-1)} \\ = \; & \frac{1}{T}\sum_t\int\mathcal{B}_t^{XX}\left(z_0,z_0\right)\otimes\left(H^{1/2}u\right)\left(H^{1/2}u\right)^T\mathcal{H}_m\left(z_0\right)\left(H^{1/2}u\right)K\left(u\right)K\left(v\right)dudv \\ - \; & \frac{1}{T}\sum_t\int\mathcal{B}_t^{X_{-1}X}\left(z_0,z_0\right)\otimes\left(H^{1/2}v\right)\left(H^{1/2}u\right)^T\mathcal{H}_m\left(z_0\right)\left(H^{1/2}u\right)K\left(u\right)K\left(v\right)dudv+o_p\left(H^{3/2}\right) \\ = \; & O_p\left(H^{3/2}\right). \end{aligned}$$

Finally, the terms in

$$(NT)^{-1}\widetilde{Z}^T W S_{m2}(z_0) =$$

$$\begin{pmatrix} (NT)^{-1}\sum_{it}\Delta X_{it}\left\{X_{i(t-1)}^T\otimes(Z_{i(t-1)}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{i(t-1)}-z_0)K_{it}K_{i(t-1)} \\ (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z_0)\right\}\left\{X_{i(t-1)}^T\otimes(Z_{i(t-1)}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{i(t-1)}-z_0)K_{it}K_{i(t-1)} \end{pmatrix}$$

are of order

$$\begin{aligned} & (NT)^{-1}\sum_{it}\Delta X_{it}\left\{X_{i(t-1)}^T\otimes(Z_{i(t-1)}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{i(t-1)}-z_0)K_{it}K_{i(t-1)} \\ = \; & \frac{\mu_2\left(K_v\right)}{T}\sum_t\sum_d E\left[\Delta X_{it}X_{di(t-1)}\middle|\, Z_{it}=z_0, Z_{i(t-1)}=z_0\right] f_{Z_{it},Z_{i(t-1)}}\left(z_0,z_0\right)\times\operatorname{tr}\left\{\mathcal{H}_{md}\left(z_0\right)H\right\}+o_p\left(\operatorname{tr}\left\{H\right\}\right), \end{aligned}$$

and

$$\begin{aligned} & (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)-X_{i(t-1)}\otimes(Z_{i(t-1)}-z_0)\right\}\left\{X_{i(t-1)}^T\otimes(Z_{i(t-1)}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{i(t-1)}-z_0)K_{it}K_{i(t-1)} \\ = \; & \frac{1}{T}\sum_t\int\mathcal{B}_t^{XX_{-1}}\left(z_0,z_0\right)\otimes\left(H^{1/2}u\right)\left(H^{1/2}v\right)^T\mathcal{H}_m\left(z_0\right)\left(H^{1/2}v\right)K\left(u\right)K\left(v\right)dudv \\ - \; & \frac{1}{T}\sum_t\int\mathcal{B}_t^{X_{-1}X_{-1}}\left(z_0,z_0\right)\otimes\left(H^{1/2}v\right)\left(H^{1/2}v\right)^T\mathcal{H}_m\left(z_0\right)\left(H^{1/2}v\right)K\left(u\right)K\left(v\right)dudv+o_p\left(H^{3/2}\right) \\ = \; & O_p\left(H^{3/2}\right). \end{aligned}$$

Taking into account (23), (24) and the previous results the asymptotic bias can be written as

$$\begin{aligned}
&E\left\{\widehat{m}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\}-m(z_0)\\
=\;&\frac{1}{2}e_1^T\left(\widetilde{Z}^TW\widetilde{Z}\right)^{-1}\widetilde{Z}^TW\left\{S_{m_1}(z_0)-S_{m_2}(z_0)\right\}+o_p\left(\mathrm{tr}\left\{H\right\}\right)=\left(\frac{1}{T}\sum_t\mathcal{B}_t^{\Delta X\Delta X}(z_0,z_0)\right)^{-1}\\
&\times\frac{1}{T}\sum_t\sum_d\left\{\mu_2(K_u)E\left[\Delta X_{it}X_{dit}\right|Z_{it}=z_0,Z_{i(t-1)}=z_0\right]-\mu_2(K_v)E\left[\Delta X_{it}X_{di(t-1)}\right|Z_{it}=z_0,Z_{i(t-1)}=z_0\left.\right]\right\}\\
&\times f_{Z_{it},Z_{i(t-1)}}(z_0,z_0)\,\mathrm{tr}\left\{\mathcal{H}_{md}(z_0)H\right\}+o_p\left(\mathrm{tr}\left\{H\right\}\right).
\end{aligned}$$

To obtain an asymptotic expression for the variance let us first define the $(N(T-1)\times 1)$-vector $\Delta v=(\Delta v_1,\cdots,\Delta v_N)^T$ where $\Delta v_i=(\Delta v_{i2},...,\Delta v_{iT})^T$ and let $E\left(\Delta v\Delta v^T\right)=\mathcal{V}$ be a $N(T-1)\times N(T-1)$ matrix that contains the V_{ij}'s matrices

$$V_{ij}=E(\Delta v_i\Delta v_j^T|X_{i1},...,X_{iT},Z_{i1},...,Z_{iT})=\begin{cases}2\sigma_v^2, & \text{for}\quad i=j,\quad t=s\\ -\sigma_v^2, & \text{for}\quad i=j,\quad |t=s|<2.\\ 0, & \text{for}\quad i=j,\quad |t=s|\geq 2.\end{cases}\tag{31}$$

Then, taking into account that

$$\widehat{m}(z_0;H)-E\left\{\widehat{m}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\}=e_1^T\left(\widetilde{Z}^TW\widetilde{Z}\right)^{-1}\widetilde{Z}^TW\Delta v,\tag{32}$$

the variance of $\hat{m}(z_0;H)$ can be written as

$$\begin{aligned}
&Var\left\{\widehat{m}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\}\\
&=e_1^T(\widetilde{Z}^TW\widetilde{Z})^{-1}\widetilde{Z}^TW\mathcal{V}W^T\widetilde{Z}(\widetilde{Z}^TW\widetilde{Z})^{-1}e_1.
\end{aligned}\tag{33}$$

The upper left entry of $\frac{1}{NT}\widetilde{Z}^TW\mathcal{V}W^T\widetilde{Z}$ is

$$\begin{aligned}
&\frac{2\sigma_v^2}{NT}\sum_{it}\Delta X_{it}\Delta X_{it}^TK_{it}^2K_{i(t-1)}^2-\frac{\sigma_v^2}{NT}\sum_{it}\Delta X_{it}\Delta X_{i(t-1)}^TK_{it}K_{i(t-1)}^2K_{i(t-2)}\\
&=\frac{2\sigma_v^2R(K_u)R(K_v)}{T|H|}\sum_t\mathcal{B}_t^{\Delta X\Delta X}(z_0,z_0)\left\{1+o_p(1)\right\}+O_p\left(|H|^{1/2}\right).
\end{aligned}\tag{34}$$

The upper right block is

$$\begin{aligned}
&\frac{2\sigma_v^2}{NT}\sum_{it}\Delta X_{it}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T-X_{i(t-1)}^T\otimes(Z_{i(t-1)}-z_0)^T\right\}K_{it}^2K_{i(t-1)}^2\\
&-\frac{\sigma_v^2}{NT}\sum_{it}\Delta X_{it}\left\{X_{i(t-1)}^T\otimes(Z_{i(t-1)}-z_0)^T-X_{i(t-2)}^T\otimes(Z_{i(t-2)}-z_0)^T\right\}K_{it}K_{i(t-1)}^2K_{i(t-2)}\\
=\;&\mathbf{I}_1-\mathbf{I}_2.
\end{aligned}\tag{35}$$

Then using standard kernel density results, under (A.1) to (A.5) we get

$$\begin{aligned}
\mathbf{I}_1\;=\;&\frac{2\sigma_v^2}{T|H|}\sum_t\int\left\{\mathcal{B}_t^{\Delta XX}\left(z_0+H^{1/2}u,z_0+H^{1/2}v\right)\otimes\left(H^{1/2}u\right)^T\right.\\
&\left.-\mathcal{B}_t^{\Delta XX_{-1}}\left(z_0+H^{1/2}u,z_0+H^{1/2}v\right)\otimes\left(H^{1/2}v\right)^T\right\}K^2(u)K^2(v)dudv\left\{1+o_p(1)\right\}
\end{aligned}$$

$$
\begin{aligned}
\mathbf{I}_2 &= \frac{2\sigma_v^2}{T\,|H|^{1/2}} \sum_t \int \left\{ \mathcal{B}_t^{\Delta XX_{-1}}\left(z_0 + H^{1/2}u, z_0 + H^{1/2}v, z_0 + H^{1/2}w\right) \otimes \left(H^{1/2}v\right)^T \right. \\
&\quad \left. -\mathcal{B}_t^{\Delta XX_{-2}}\left(z_0 + H^{1/2}u, z_0 + H^{1/2}v, z_0 + H^{1/2}w\right) \otimes \left(H^{1/2}w\right)^T \right\} K(u)K^2(v)K(w)dudvdw\left\{1 + o_p(1)\right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\mathcal{B}_t^{\Delta XX}(z_0, z_0) &= E\left[\Delta X_{it}X_{it}^T \middle| Z_{it} = z_0, Z_{i(t-1)} = z_0\right] f_{Z_{it},Z_{i(t-1)}}(z_0, z_0), \\
\mathcal{B}_t^{\Delta XX_{-1}}(z_0, z_0) &= E\left[\Delta X_{it}X_{i(t-1)}^T \middle| Z_{it} = z_0, Z_{i(t-1)} = z_0\right] f_{Z_{it},Z_{i(t-1)}}(z_0, z_0), \\
\mathcal{B}_t^{\Delta XX_{-1}}(z_0, z_0, z_0) &= E\left[\Delta X_{it}X_{i(t-1)}^T \middle| Z_{it} = z_0, Z_{i(t-1)} = z_0, Z_{i(t-2)} = z_0\right] f_{Z_{it},Z_{i(t-1)}}(z_0, z_0, z_0), \\
\mathcal{B}_t^{\Delta XX_{-2}}(z_0, z_0, z_0) &= E\left[\Delta X_{it}X_{i(t-2)}^T \middle| Z_{it} = z_0, Z_{i(t-1)} = z_0, Z_{i(t-2)} = z_0\right] f_{Z_{it},Z_{i(t-1)}}(z_0, z_0, z_0).
\end{aligned}
$$

It is straightforward to show that $\mathbf{I}_1 = O_p\left(\frac{1}{|H|}\|H\|\right) + o_p\left(\frac{1}{|H|}\|H\|\right)$ and $\mathbf{I}_2 = O_p\left(\frac{1}{|H|^{1/2}}\|H\|\right) + o_p\left(\frac{1}{|H|^{1/2}}\|H\|\right)$, as N tends to infinity. $\|\cdot\|$ is a certain norm, for example $\|H\| = \left(\sum_{i,j} h_{ij}^2\right)^{1/2}$. Finally, the lower-right block is

$$
\begin{aligned}
&\frac{2\sigma_v^2}{NT} \sum_{it} \left\{X_{it} \otimes (Z_{it} - z_0) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z_0)\right\} \left\{X_{it}^T \otimes (Z_{it} - z_0)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z_0)^T\right\} K_{it}^2 K_{i(t-1)}^2 \\
&-\frac{\sigma_v^2}{NT} \sum_{it} \left\{X_{it} \otimes (Z_{it} - z_0) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z_0)\right\} \\
&\times \left\{X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z_0)^T - X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z_0)^T\right\} K_{it} K_{i(t-1)}^2 K_{i(t-2)} = \mathbf{I}_1 - \mathbf{I}_2, \qquad (36)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{I}_1 &= \frac{2\sigma_v^2 \mu_2\left(K^2\right) R\left(K_v\right)}{T\,|H|} \sum_t \mathcal{B}_t^{XX}(z_0, z_0) \otimes H + \frac{2\sigma_v^2 \mu_2\left(K^2\right) R\left(K_u\right)}{T\,|H|} \sum_t \mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0) \otimes H + O_p\left(|H|\right), \\
\mathbf{I}_2 &= \frac{\sigma_v^2 \mu_2\left(K^2\right)}{T\,|H|^{1/2}} \sum_t \mathcal{B}_t^{X_{-1}X_{-1}}(z_0, z_0, z_0) \otimes H + O_p\left(|H|^{1/2}\right).
\end{aligned}
$$

So now, substituting (30), (34), (35) and (36) into (33) we obtain

$$
Var\left\{\widehat{m}(z_0; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT}\right\} = \frac{2\sigma_v^2 R\left(K_u\right) R\left(K_v\right)}{N\,|H|} \left(\frac{1}{T} \sum_t \mathcal{B}_t^{\Delta X \Delta X}(z_0, z_0)\right)^{-1} \left\{1 + o_p(1)\right\}.
$$

∎

Proof of Corollary 3.1

Let

$$
\begin{aligned}
\widehat{m}(z_0; H) - m(z_0) &= \left\{\widehat{m}(z_0; H) - E\left[\widehat{m}(z_0; H) \middle| X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT}\right]\right\} \\
&\quad + \left\{E\left[\widehat{m}(z_0; H) \middle| X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT}\right] - m(z_0)\right\} \equiv \mathbf{I}_1 + \mathbf{I}_2.
\end{aligned}
$$

We will first show that under conditions of the corollary $\mathbf{I}_2 = o_p\left(\frac{1}{\sqrt{N|H|}}\right)$, as N tends to infinity. In order to do so, recall that under conditions in Theorem 3.1

$$E\left\{\left.\widehat{m}\left(z_0;H\right)\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\} - m(z_0)$$

$$= \left(\frac{1}{T}\sum_t \mathcal{B}_t^{\Delta X \Delta X}(z_0,z_0)\right)^{-1}\frac{1}{T}\sum_t\sum_d\left\{\mu_2\left(K_u\right)\mathcal{B}_{dt}^{\Delta XX}\left(z_0,z_0\right) - \mu_2\left(K_v\right)\mathcal{B}_{dt}^{\Delta XX_{-1}}\left(z_0,z_0\right)\right\} \quad (37)$$
$$\times \mathrm{tr}\left\{\mathcal{H}_{md}\left(z_0\right)H\right\} + O_p\left(H^{3/2}\right) + o_p\left(\mathrm{tr}\left\{H\right\}\right).$$

By the law of iterated expectations,

$$E\left\{\widehat{m}\left(z_0;H\right)\right\} = \int E\left\{\left.\widehat{m}\left(z_0;H\right)\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\} dF\left(X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right).$$

The leading term in (37) does not depend on the sample and furthermore, under the conditions established in the corollary the remainder terms are $o_p(1)$. Hence,

$$\mathbf{I}_2 = E\left\{\left.\widehat{m}\left(z_0;H\right)\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\} - m\left(z_0;H\right) + o_p(1).$$

Condition (7) applies and the proof is done.
Now we show that

$$\sqrt{N\left|H\right|}\mathbf{I}_1 \to \mathcal{N}\left(0, 2\sigma_v^2 R\left(K_u\right)R\left(K_v\right)\left(\frac{1}{T}\sum_t \mathcal{B}_t^{\Delta X \Delta X}\left(z_0,z_0\right)\right)^{-1}\right),$$

as N tends to infinity.
In order to show this let

$$\widehat{m}\left(z_0;H\right) - E\left\{\left.\widehat{m}\left(z_0;H\right)\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\} = e_1^T\left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1}\widetilde{Z}^T W \Delta v, \quad (38)$$

where $\Delta v = \left[\Delta v_{11},\cdots,\Delta v_{NT}\right]^T$. Under conditions (B.1) to (B.5) and recalling that (X_{it}, Z_{it}) are i.i.d random variables in the sub-index i, we directly apply the Lindeberg-Levy Central Limit Theorem and we obtain

$$\frac{1}{\sqrt{N\left|H\right|}}\widetilde{Z}^T W \Delta v \to_d \mathcal{N}\left(0,\mathcal{D}\right), \quad N \to \infty,$$

where $\mathcal{D}$ has been already defined in (34), (35) and (36). Finally, using (30) and applying the Cramer-Wold device the proof is done.

∎

Proof of Theorem 3.2

The proof of this result follows the same lines as in the proof of Theorem 3.1. Let

$$\widehat{m}^{(1)}(z_0;H) = e_1^T(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)})^{-1}\widetilde{Z}^{(1)T}W^{(1)}\Delta\widetilde{Y}^{(1)}. \quad (39)$$

Then proceeding as before in the proof of Theorem 3.1 we get

$$E\left\{\widehat{m}^{(1)}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\} = e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}\left[M^{(1)}+M^{(2)}\right], \quad (40)$$

where

$$\begin{aligned} M^{(1)} &= \left[\left\{X_{12}^T m(Z_{12})\right\}^T, \cdots, \left\{X_{NT}^T m(Z_{NT})\right\}^T\right]^T, \\ M^{(2)} &= \left[\left\{X_{11}^T\left\{E\left\{\widehat{m}(Z_{11};H)|X_{11},...,Z_{NT}\right\} - m(Z_{11})\right\}\right\}^T, \cdots, \right. \\ & \left.\left\{X_{N(T-1)}^T\left\{E\left\{\widehat{m}(Z_{N(T-1)};H)|X_{11},...,Z_{NT}\right\} - m(Z_{N(T-1)})\right\}\right\}^T\right]^T \end{aligned}$$

are $N(T-1)\times 1$ vectors. We can approximate $M^{(1)}$ through a Taylor's expansion, i.e.

$$M^{(1)} = \widetilde{Z}^{(1)}\begin{bmatrix} m(z_0) \\ D_m(z_0)\end{bmatrix} + \frac{1}{2}Q_m^{(1)}(z_0) + R(z_0),$$

where,

$$Q_m^{(1)}(z_0) = \left[S_{m,12}^{(1)T}(z_0), \cdots, S_{m,NT}^{(1)T}(z_0)\right]^T,$$

and

$$S_{m,it}^{(1)}(z_0) = \left\{X_{it}^T \otimes (Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0).$$

Using (B.1) we get

$$e_1^T(\widetilde{Z}^{(1)T}W\widetilde{Z})^{-1}\widetilde{Z}^{(1)T}W^{(1)}R(z_0) = o_p\left(\text{tr}\left\{H\right\}\right),$$

and therefore,

$$\begin{aligned} & E\left\{\widehat{m}^{(1)}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\} - m(z_0) \\ &= e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}\left\{\frac{1}{2}Q_m^{(1)}(z_0) + M^{(2)}\right\} + o_p\left(\text{tr}\left\{H\right\}\right). \end{aligned} \quad (41)$$

To obtain an asymptotic expression for the bias we first calculate

$$\frac{1}{NT}\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)} =$$

$$\begin{pmatrix} (NT)^{-1}\sum_{it}X_{it}X_{it}^TK_{it} & (NT)^{-1}\sum_{it}X_{it}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}K_{it} \\ (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)\right\}X_{it}^TK_{it} & (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)\right\}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}K_{it}\end{pmatrix}.$$

Using standard properties of kernel density estimators, under conditions (A.1) to (A.5) and as N tends to infinity,

$$\begin{aligned} (NT)^{-1}\sum_{it}X_{it}X_{it}^TK_{it} &= \frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0) + o_p(1), \\ (NT)^{-1}\sum_{it}X_{it}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}K_{it} &= \frac{1}{T}\sum_t \mathcal{D}\mathcal{B}_t^{XX}(z_0)\left(I_d\otimes\mu_2(K_u)H\right) + o_p(H), \\ (NT)^{-1}\sum_{it}\left\{X_{it}\otimes(Z_{it}-z_0)\right\}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}K_{it} &= \frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\otimes\mu_2(K_u)H + o_p(H). \end{aligned}$$

Note that $\mathcal{B}_t^{XX}(z_0)$ and $\mathcal{D}\mathcal{B}_t^{XX}(z_0)$ are defined as in the proof of Theorem 3.1 but the moment functions now are taken conditionally only to $Z_{it}=z_0$.
Using the previous results,

$$NT\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}=\begin{pmatrix}\mathcal{C}_{11}^{(1)} & \mathcal{C}_{12}^{(1)}\\ \mathcal{C}_{21}^{(1)} & \mathcal{C}_{22}^{(1)}\end{pmatrix}, \tag{42}$$

where

$$\begin{aligned}
\mathcal{C}_{11}^{(1)} &= \left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\right)^{-1}+o_p(1),\\
\mathcal{C}_{12}^{(1)} &= -\left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\right)^{-1}\left[\frac{1}{T}\sum_t \mathcal{D}\mathcal{B}_t^{XX}(z_0)\right]\left(\left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\right)^{-1}\otimes I_q\right)+o_p(1),\\
\mathcal{C}_{22}^{(1)} &= \left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}(z_0)\otimes\mu_2(K_u)H\right)^{-1}+o_p\left(H^{-1}\right).
\end{aligned}$$

Furthermore the terms in

$$(NT)^{-1}\frac{1}{2}\widetilde{Z}^{(1)T}W^{(1)}Q_m^{(1)}(z_0)= \tag{43}$$

$$\begin{pmatrix}(NT)^{-1}\sum_{it}X_{it}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0)K_{it}\\ (NT)^{-1}\sum_{it}\{X_{it}\otimes(Z_{it}-z_0)\}\left\{X_{it}^T\otimes(Z_{it}-z_0)^T\right\}\mathcal{H}_m(z_0)(Z_{it}-z_0)K_{it}\end{pmatrix}$$

are of order

$$\frac{\mu_2(K_u)}{T}\sum_t\sum_d E\left[X_{it}X_{dit}\right|Z_t=z_0]\,f_{Z_{it}}(z_0)\times\mathrm{tr}\left\{\mathcal{H}_{md}(z_0)\right\}+o_p\left(\mathrm{tr}\{H\}\right)$$

and $O_p\left(H^{3/2}\right)$, respectively. In order to evaluate the asymptotic bias of the last term we have to calculate

$$(NT)^{-1}\widetilde{Z}^{(1)T}W^{(1)}M^{(2)}= \tag{44}$$

$$\begin{pmatrix}(NT)^{-1}\sum_{it}X_{it}X_{i(t-1)}^T\left(E\left\{\widehat{m}(Z_{i(t-1)})\right|X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}-m(Z_{i(t-1)})\Big)K_{it}\\ (NT)^{-1}\sum_{it}\{X_{it}\otimes(Z_{it}-z_0)\}X_{i(t-1)}^T\left(E\left\{\widehat{m}(Z_{i(t-1)})\right|X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}-m(Z_{i(t-1)})\Big)K_{it}\end{pmatrix}.$$

It is straightforward to show that

$$(NT)^{-1}\sum_{it}X_{it}X_{i(t-1)}^T\left(E\left\{\widehat{m}(Z_{i(t-1)})\right|X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}-m(Z_{i(t-1)})\Big)K_{it}=O_p\left(\mathrm{tr}\{H\}\right),$$

as N tends to infinity, and

$$(NT)^{-1}\sum_{it}\{X_{it}\otimes(Z_{it}-z_0)\}X_{i(t-1)}^T\left(E\left\{\widehat{m}(Z_{i(t-1)})\right|X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}-m(Z_{i(t-1)})\Big)K_{it}$$
$$=o_p\left(\mathrm{tr}\{H\}\right),$$

as N tends to infinity.
Now substitute the asymptotic expressions for (42), (43) and (44) into (41) apply that $|H|\to 0$ in such a way that $N|H|\to\infty$ and we have shown that the asymptotic bias in $\widehat{m}^{(1)}(z_0;H)$ is of the same

order as it is in $\widetilde{m}^{(1)}(z_0; H)$.

For the variance term, recall that substituting (13) into (15) and taking conditional expectations on the sample values

$$\widehat{m}^{(1)}(z_0;H) - E\left\{\widehat{m}^{(1)}(z_0;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\} = e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}\Delta v$$
$$+e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}\widehat{v},$$

where $\widehat{v} = (\widehat{v}_1, \cdots, \widehat{v}_N)^T$ is a $(N(T-1)\times 1)$-vector, such that

$$\widehat{v}_i = \left(\left\{X_{i0}^T r\left(Z_{i0}; H\right)\right\}^T, \cdots, \left\{X_{i(T-1)}^T r\left(Z_{i(T-1)}; H\right)\right\}^T\right)^T,$$

$i = 1, \cdots, N$, and

$$r\left(Z_{i(t-1)}; H\right) = \widehat{m}\left(Z_{i(t-1)}; H\right) - E\left\{\widehat{m}\left(Z_{i(t-1)}; H\right)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\right\},$$

$i = 1, \cdots, N; t = 2, \cdots, T$.

Then, the variance of $\widehat{m}^{(1)}(z_0; H)$ takes the form

$$\mathrm{Var}\left\{\widehat{m}^{(1)}(z_0;H)\Big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\} \tag{45}$$

$$\begin{aligned}
&= e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}\mathcal{V}W^{(1)T}\widetilde{Z}^{(1)}\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}e_1\\
&+ e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}E\left\{\widehat{v}\widehat{v}^T\big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}W^{(1)T}\widetilde{Z}^{(1)}\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}e_1\\
&+ 2e_1^T\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}\widetilde{Z}^{(1)T}W^{(1)}E\left\{\widehat{v}\Delta v^T\big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}W^{(1)T}\widetilde{Z}^{(1)}\left(\widetilde{Z}^{(1)T}W^{(1)}\widetilde{Z}^{(1)}\right)^{-1}e_1.\\
&\equiv \mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3.
\end{aligned}$$

Following exactly the same lines as in the proof of the variance term in Theorem 3.1 we get, as N tends to infinity,

$$\mathbf{I}_1 = \frac{2\sigma_v^2 R\left(K_u\right)}{N\left|H\right|^{1/2}}\left(\frac{1}{T}\sum_t \mathcal{B}_t^{XX}\left(z_0\right)\right)^{-1}\left\{1+o_p(1)\right\}. \tag{46}$$

In order to calculate the asymptotic order of $\mathbf{I}_2$, we just need to calculate

$$\frac{1}{NT}\widetilde{Z}^{(1)T}W^{(1)}E\left\{\widehat{v}\widehat{v}^T\big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}W^{(1)T}\widetilde{Z}^{(1)}. \tag{47}$$

The upper left entry is

$$(NT)^{-1}\sum_i\sum_{ts}X_{it}X_{i(t-1)}^T E\left\{r\left(Z_{i(t-1)};H\right)r\left(Z_{i(s-1)};H\right)^T\Big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}X_{i(s-1)}X_{is}^T K_{it}K_{is}. \tag{48}$$

Applying the Cauchy-Schwarz inequality for covariance matrices then (48) is bounded by

$$(NT)^{-1}\sum_i\sum_{ts}X_{it}X_{i(t-1)}^T\mathrm{vec}^{1/2}\left\{\mathrm{diag}\left(E\left\{r\left(Z_{i(t-1)};H\right)r\left(Z_{i(t-1)};H\right)^T\Big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}\right)\right\}$$
$$\times\mathrm{vec}^{1/2}\left\{\mathrm{diag}\left(E\left\{r\left(Z_{i(s-1)};H\right)r\left(Z_{i(s-1)};H\right)^T\Big| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}\right)\right\}^T X_{i(s-1)}X_{is}^T K_{it}K_{is}.$$

Now, note that under the conditions of the Theorem 3.1

$$\text{vec}\left\{\text{diag}\left(E\left\{r\left(z;H\right)r\left(z;H\right)^T\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\right\}\right)\right\} = O_p\left(\frac{1}{N\left|H\right|}\right),$$

uniformly in z, and therefore (48) is of order $O_p\left(\frac{1}{N|H|}\right)$.
Following the same lines, it is easy to show that the upper right entry of (47) is

$$(NT)^{-1}\sum_i\sum_{ts} X_{it}X_{i(t-1)}^T E\left\{r\left(Z_{i(t-1)};H\right)r\left(Z_{i(s-1)};H\right)^T\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\Big\}$$
$$\times X_{i(s-1)}\left\{X_{is}\otimes\left(Z_{is}-z_0\right)\right\}^T K_{it}K_{is} = o_p\left(\frac{1}{N\left|H\right|}\right),$$

and finally the lower right entry of (47) is

$$(NT)^{-1}\sum_i\sum_{ts}\left\{X_{it}\otimes\left(Z_{it}-z_0\right)\right\}X_{i(t-1)}^T E\left\{r\left(Z_{i(t-1)};H\right)r\left(Z_{i(s-1)};H\right)^T\right| X_{11},\cdots,X_{NT},Z_{11},\cdots,Z_{NT}\Big\}$$
$$\times X_{i(s-1)}\left\{X_{is}\otimes\left(Z_{is}-z_0\right)\right\}^T K_{it}K_{is} = O_p\left(\frac{1}{N\left|H\right|}\right).$$

Now, combining results in (42) and (47) we show that $\mathbf{I}_2 = o_p\left(\frac{1}{N|H|}\right)$. Finally a standard Cauchy-Schwarz inequality is enough to show that $\mathbf{I}_3 = o_p\left(\frac{1}{N|H|}\right)$ and then the proof of the result is closed.

∎

References

Arellano, M. (2003). *"Panel Data Econometrics"*. Oxford University Press.

Baltagi, B. (2005). *"Econometrics Analysis of Panel Data (2nd edition)"*. Wiley, New York.

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

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

Härdle, W. (1990). *"Applied Nonparametric Regression"*. Cambridge University Press.

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

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

Hsiao, C. (2003). *"Analysis of Panel Data, 2nd Edition"*. Econometric Society Monographs.

Kniesner, T. and Q. Li (2002). "Semiparametric panel data models with heterogeneous dynamic adjustment: theoretical consideration and an application to labor supply". *Empirical Economics 27*, 131–148.

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". Working Paper.

Li, Q. and C. Hsiao (1998). "Testing serial correlation in semiparametric panel data models". *Journal of Econometrics 87*, 207–237.

Li, Q. and J. Racine (2007). "Nonparametric Econometrics: Theory and Practice". Princeton: Princeton University Press.

Li, Q. and T. Stengos (1996). "Semiparametric estimation of partially linear panel data models". *Journal of Econometrics 71*, 389–397.

Lin, X. and 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.

Masry, E. (1996). "Multivariate regression estimation: local polynomial fitting for time series". *Stochastic Processes and their Applications 65*, 81–101.

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

Soberon, A. and J. Rodriguez-Poo (2012). "Semiparametric estimation of panel data varying coefficient instrumental variables models". Preprint.

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

Su, L. and A. Ullah (2010). "Nonparametric and semiparametric panel econometric models: estimation and testing". Working Paper.

Sun, Y., R. 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 Handbook of Applied Economic Statistics, ed. A. Ullah and D.E.A. Giles, New York: Marcel Dekker.

Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statistica Sinica 12*, 801–821.

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, W. and S. Lee (2000). "Variable bandwdith selection in varying-coefficient models". *Journal of Multivariate Analysis 74*, 116–134.