Journal of Multivariate Analysis 133 (2015) 95–122

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric estimation of fixed effects panel data varying coefficient models

Juan M. Rodriguez-Poo *, Alexandra Soberón

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

A R T I C L E I N F O

Article history:
Received 7 February 2014
Available online 28 September 2014

JEL classification:
C14
C23

AMS subject classifications:
62G08
62G20
62P20

Keywords:
Varying coefficient models
Fixed effects
Panel data
Local linear regression
Oracle efficient estimator
Within estimator
Profile least squares estimator

A B S T R A C T

In this paper, we consider the nonparametric estimation of a varying coefficient fixed effect panel data model. The estimator is based in a within (un-smoothed) transformation of the regression model and then a local linear regression is applied to estimate the unknown varying coefficient functions. It turns out that the standard use of this technique produces a non-negligible asymptotic bias. In order to avoid it, a high dimensional kernel weight is introduced in the estimation procedure. As a consequence, the asymptotic bias is removed but the variance is enlarged, and therefore the estimator shows a very slow rate of convergence. In order to achieve the optimal rate, we propose a one-step backfitting algorithm. The resulting two-step estimator is shown to be asymptotically normal and its rate of convergence is optimal within its class of smoothness functions. It is also oracle efficient. Further, this estimator is compared both theoretically and by Monte-Carlo simulation against other estimators that are based in a within (smoothed) transformation of the regression model. More precisely the profile least-squares estimator proposed in this context in Sun et al. (2009). It turns out that the smoothness in the transformation enlarges the bias and it makes the estimator more difficult to analyze from the statistical point of view. However, the first step estimator, as expected, shows a bad performance when compared against both the two step backfitting algorithm and the profile least-squares estimator.

1. Introduction

This paper is concerned with the nonparametric estimation and inference of panel data varying coefficient models with fixed effects. In fact, in the random effect setting, direct estimation through the use of standard nonparametric techniques is straightforward and there is only need to care about efficiency issues (see for example [14] or [6]). However, in the fixed effect framework, direct estimation of the functions of interest produces asymptotically biased estimators. This is due to the correlation that exists between the heterogeneity term and the explanatory variables. Traditionally, standard techniques in fixed effect panel data models consist in removing the heterogeneity term by transforming the statistical model of departure. Following Su and Ullah [17] there exist, at least, two different alternative transformations. On one side, the so-called profile least-squares method and, on the other side, the differencing method. Taking first differences, subtracting the equation from time t from that for time 1 or alternatively subtracting the within-group average are all them examples that can be considered differencing techniques. In standard parametric fixed effect panel data models (see [19]) the choice among differencing techniques is related to efficiency issues. For example, if the idiosyncratic errors follow the structure of a random walk, first differences are recommended, however in much general situations such as an i.i.d. or a strictly stationary context the within (fixed effects) estimator is recommended.

* Corresponding author.
E-mail address: rodrigjm@unican.es (J.M. Rodriguez-Poo).

http://dx.doi.org/10.1016/j.jmva.2014.09.008
0047-259X/

In this paper we present an estimation procedure that uses a (un-smoothed) mean deviation transformation of the varying coefficient fixed effect panel data model. Since the transformed model appears as an additive function with the same functional form at different times, the proposals to estimate this type of models are closely related to estimation techniques originally designed for additive models (see [5,9] or [16]). As an alternative, we propose to apply a local approximation on the T additive functions that result from the mean deviation transformation where we denote by T the number of time observations per individual. In this context, the local linear regression estimator exhibits a non-negligible bias in the estimation of the additive components. This is because these techniques approximate the unknown function around a fixed value without considering the sum of the distances between this fixed term and the other values of the sample. This phenomena was already pointed out in [12,8] but unfortunately they did not provide a solution to this problem. In this context, our proposal is to consider a local approximation around the whole vector of time observations for each individual. Unfortunately, although the introduction of the T-variate kernel solves the bias problem, it enlarges the variance. For large T, this can create very slow rates of convergence of our estimator. As a solution, we propose to use a one-step backfitting algorithm. The idea, as already pointed out in [4], is that additional smoothing cannot reduce the bias but it can diminish the variance. Therefore, the additional smoothing that is introduced by the backfitting enables us to achieve optimal nonparametric rates of convergence for the estimators of the unknown functions of interest. The same type of results can be found in [13] for the first differences setting.

The reason to choose the within transformation among others is twofold. First, considering efficiency issues, the resulting estimator will be more efficient than those resulting from other transformations when assuming standard assumptions such as i.i.d. or stationary idiosyncratic errors. Second, note that this transformation consists in removing the fixed effect term by deducting a (un-smoothed) cross-time average from each individual unit. On the contrary, in profile least-squares techniques the heterogeneity term is removed by deducting a smoothed cross-time average. Therefore, since they are rather similar, it can be also of great interest to compare the statistical properties of both estimators, i.e. the one obtained in this paper using the within transformation and the profile least-squares estimator proposed in [18]. Hence, the main interest of the paper is that, to our knowledge, in the framework of fixed effects varying coefficient panel data models this is the first paper where estimators that result from deducting un-smoothed and smoothed cross-time averages from each individual units are compared both from theoretical and simulation results. Furthermore, a nonparametric fixed effect estimator of the varying coefficient model is proposed, its asymptotic properties are obtained and it is also shown that it also exhibits the oracle efficiency property.

The rest of the paper is organized as follows. In Section 2 we set up the model and the estimation procedure. We also provide some comparisons with respect to profile least-squares estimators in very simple situations. In Section 3 we study the main statistical properties of both direct local linear estimator and one-step backfitting estimator for the multivariate case. We also compare both local linear and backfitting estimators against the one proposed in [18]. Finally, in Section 4 we compare empirically the performance in small sample sizes of the same estimators through a Monte Carlo simulation. The proofs of the main results are collected in the Appendix.

2. Statistical model and estimation procedure

We consider the following panel data varying coefficient regression model with fixed effects

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

where X_{it} and Z_{it} are vectors of covariates of dimension $d \times 1$ and $q \times 1$, respectively, $m(Z) = (m_1(Z), \ldots, m_d(Z))$ is a $d \times 1$ vector of unknown functions to estimate, v_{it} is the random error term and μ_i reflects the unknown cross-sectional heterogeneity. Also, we allow for μ_i to be correlated with Z_{it} and/or X_{it} with an unknown correlation structure.

To illustrate the estimation procedure proposed in this paper and to compare it against the profile least-squares estimator proposed in [18] we first focus on the univariate regression model and later we extend the results to the multivariate case.

Consider the linear panel data model, where the dimensions of X and Z are respectively $d = 1$ and $q = 1$,

$$Y_{it} = X_{it} m\left(Z_{it}\right) + \mu_i + v_{it}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T. \tag{2.2}$$

Let $\overline{Y}_{i\cdot} = T^{-1}\sum_{s=1}^{T} Y_{is}$ and $\overline{v}_{i\cdot} = T^{-1}\sum_{s=1}^{T} v_{is}$. The within transformation implies subtracting from time t of (2.2) the within-group mean, i.e.,

$$Y_{it} - \overline{Y}_{i\cdot} = X_{it} m\left(Z_{it}\right) - \frac{1}{T}\sum_{s=1}^{T} X_{is} m\left(Z_{is}\right) + v_{it} - \overline{v}_{i\cdot}, \quad i = 1, \ldots, N;\ t = 1, \ldots, T. \tag{2.3}$$

Instead of taking averages over time for each individual, consider the following corresponding local (smoothed) averages,

$$\widetilde{Y}_{i\cdot}\left(z\right) = \sum_{s=1}^{T} \varpi_{is}\left(z\right) Y_{is}, \qquad \widetilde{X}_{i\cdot}\left(z\right) = \sum_{s=1}^{T} \varpi_{is}\left(z\right) X_{is}$$

$$\widetilde{v}_{i\cdot}\left(z\right) = \sum_{s=1}^{T} \varpi_{is}\left(z\right) v_{is},$$

where

$$\varpi_{is}(z) = \frac{K_g(Z_{is} - z)}{\sum_{r=1}^{T} K_g(Z_{ir} - z)} \quad s = 1, \ldots, T, \tag{2.4}$$

g is a bandwidth and K is a kernel function such as

$$\int K(u)du = 1 \quad \text{and} \quad K_g(u) = \frac{1}{g} K(u/g).$$

Since $\sum_{s=1}^{T} \varpi_{is}(z)\mu_i = \mu_i$, for all i, then, applying the same transformation as for the within estimator we obtain,

$$Y_{it} - \widetilde{Y}_{i\cdot}(Z_{it}) = \left(X_{it} - \widetilde{X}_{i\cdot}(Z_{it})\right) m(Z_{it}) + v_{it} - \widetilde{v}_{i\cdot}(Z_{it}), \quad i = 1, \ldots, N;\ t = 1, \ldots, T. \tag{2.5}$$

Estimation of the quantities of interest can be implemented in (2.5) by considering, for any $z \in A$, where A is a compact subset in a non-empty interior of $\mathbb{R}$, the following Taylor expansion

$$\begin{aligned}\left(X_{it} - \widetilde{X}_{i\cdot}(Z_{it})\right) m(Z_{it}) &\approx \left(X_{it} - \widetilde{X}_{i\cdot}(z)\right) m(z) + \left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)\, m'(z) \\ &\quad + \frac{1}{2}\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)^2 m''(z) + \cdots + \frac{1}{p!}\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)^p m^{(p)}(z) \\ &\equiv \sum_{\lambda=0}^{p} \alpha_\lambda \left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)^\lambda .\end{aligned}$$

This suggests that we estimate $m(z), m'(z), \ldots, m^{(p)}(z)$ by regressing $Y_{it} - \widetilde{Y}_{i\cdot}(z)$ on the terms $\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)^\lambda$, for $\lambda = 1, \ldots, p$, with kernel weights. Then, the quantities of interest can be estimated using a locally weighted linear regression,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - \widetilde{Y}_{i\cdot}(z) - \alpha_0\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right) - \alpha_1\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)\right)^2 K_g(Z_{it} - z); \tag{2.6}$$

see [3,15] or [21].

Let $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ be the minimizers of (2.6). The above exposition suggests as estimators for $m(z)$ and $m'(z)$, $\widehat{m}_h(z) = \widehat{\alpha}_0$ and $\widehat{m}'_h(z) = \widehat{\alpha}_1$, respectively. Furthermore, let us denote by $\alpha = \begin{pmatrix}\alpha_0 & \alpha_1\end{pmatrix}^\top$ and $\dot{Z}_{it}^\top = \left(X_{it} - \widetilde{X}_{i\cdot}(z),\ \left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)(Z_{it} - z)\right)$. Then, the criterion function (2.6) can be rewritten as

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - \widetilde{Y}_{i\cdot}(z) - \dot{Z}_{it}^\top \alpha\right)^2 K_g(Z_{it} - z), \tag{2.7}$$

and $\widehat{\alpha}_0$ and $\widehat{\alpha}_1$ have the following expression

$$\begin{pmatrix}\widehat{\alpha}_0 \\ \widehat{\alpha}_1\end{pmatrix} = \left(\sum_{it} K_g(Z_{it} - z)\dot{Z}_{it}\dot{Z}_{it}^\top\right)^{-1} \sum_{it} K_g(Z_{it} - z)\dot{Z}_{it}\left(Y_{it} - \widetilde{Y}_{i\cdot}(z)\right). \tag{2.8}$$

This estimator is the profile least-squares estimator proposed in [18]. In fact, it turns out that the corresponding local constant regression estimator (consider $\alpha_1 = 0$ in (2.6)) is

$$\widehat{m}_g(z) = \frac{\sum_{it} K_g(Z_{it} - z)\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)\left(Y_{it} - \widetilde{Y}_{i\cdot}(z)\right)}{\sum_{it} K_g(Z_{it} - z)\left(X_{it} - \widetilde{X}_{i\cdot}(z)\right)^2}, \tag{2.9}$$

which corresponds to the estimator proposed in [8].

Following the previous developments, our idea consists in estimating the quantities of interest starting from (2.3) by considering, for any $z \in A$, where A is a compact subset in a non-empty interior of $\mathbb{R}$, the following Taylor expansion

$$\begin{aligned}X_{it} m(Z_{it}) - \frac{1}{T}\sum_{s=1}^{T} X_{is} m(Z_{is}) &\approx \left(X_{it} - \frac{1}{T}\sum_{s=1}^{T} X_{is}\right) m(z) + \left[X_{it}(Z_{it} - z) - \frac{1}{T}\sum_{s=1}^{T} X_{is}(Z_{is} - z)\right] m'(z) \\ &\quad + \frac{1}{2}\left[X_{it}(Z_{it} - z)^2 - \frac{1}{T}\sum_{s=1}^{T} X_{is}(Z_{is} - z)^2\right] m''(z)\end{aligned}$$

$$+\cdots+\frac{1}{p!}\left[X_{it}\left(Z_{it}-z\right)^{p}-\frac{1}{T}\sum_{s=1}^{T}X_{is}\left(Z_{is}-z\right)^{p}\right]m^{(p)}(z)$$
$$\equiv\sum_{\lambda=0}^{p}\beta_{\lambda}\left[X_{it}\left(Z_{it}-z\right)^{\lambda}-\frac{1}{T}\sum_{s=1}^{T}X_{is}\left(Z_{is}-z\right)^{\lambda}\right]. \tag{2.10}$$

This suggests that we estimate $m(z), m'(z), \ldots, m^{(p)}(z)$ by regressing $\ddot{Y}_{it}=Y_{it}-\overline{Y}_{i\cdot}$ on the terms $X_{it}\left(Z_{it}-z\right)^{\lambda}-\frac{1}{T}\sum_{s=1}^{T} X_{is}\left(Z_{is}-z\right)^{\lambda}$, for $\lambda=1,\ldots,p$, with kernel weights. Then, the quantities of interest can be estimated using a locally weighted linear regression,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it}-\beta_0\left(X_{it}-\frac{1}{T}\sum_{s=1}^{T}X_{is}\right)-\beta_1\left[X_{it}\left(Z_{it}-z\right)-\frac{1}{T}\sum_{is}X_{is}\left(Z_{is}-z\right)\right]\right)^{2}K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right); \tag{2.11}$$

where, h is the bandwidth and

$$K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right)=\prod_{\ell=1}^{T}K_h\left(Z_{i\ell}-z\right).$$

Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ be the minimizers of (2.11). The above exposition suggests as estimators for $m(z)$ and $m'(z)$, $\widehat{m}_h(z)=\widehat{\beta}_0$ and $\widehat{m}'_h(z)=\widehat{\beta}_1$, respectively. Furthermore, let us denote by $\ddot{X}_{it}=X_{it}-\overline{X}_{i\cdot}$, $\beta=(\beta_0\;\beta_1)^{\top}$ and $\widetilde{Z}_{it}^{\top}=\left(\ddot{X}_{it},\; X_{it}\left(Z_{it}-z\right)-T^{-1}\sum_{s=1}^{T}X_{is}\left(Z_{is}-z\right)\right)$. Then, the criterion function (2.11) can be rewritten as

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it}-\widetilde{Z}_{it}^{\top}\beta\right)^{2}K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right), \tag{2.12}$$

and $\widehat{\beta}_0$ and $\widehat{\beta}_1$ have the following expression

$$\begin{pmatrix}\widehat{\beta}_0\\ \widehat{\beta}_1\end{pmatrix}=\left(\sum_{it}K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right)\widetilde{Z}_{it}\widetilde{Z}_{it}^{\top}\right)^{-1}\sum_{it}K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right)\widetilde{Z}_{it}\ddot{Y}_{it}. \tag{2.13}$$

For the sake of comparison, it is also easy to show the form of the local constant estimator as

$$\widehat{m}_h(z)=\frac{\sum_{it}K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right)\ddot{X}_{it}\ddot{Y}_{it}}{\sum_{it}K_h\left(Z_{i1}-z,\ldots,Z_{iT}-z\right)\ddot{X}_{it}^{2}}. \tag{2.14}$$

The local constant estimators of $m(z)$ obtained alternatively in (2.9) and (2.14) exhibit two main differences, first, the dimension of the kernel weights. In the profile least-squares case, the dimension of the kernel is univariate whereas in the fixed effects context the dimension is T. This might affect the variance of the fixed effects estimator. Second, the smoothed weights introduced in the profile least-squares estimator do not appear in the fixed effect estimator. This might affect the bias of the former estimator.

Note that in (2.11) or (2.12) it would have been usual to introduce a kernel function around Z_{it}. By doing so, the distance between z and any of the terms $Z_{i1},\ldots,Z_{i(t-1)},Z_{i(t+1)},\ldots,Z_{iT}$ cannot be controlled by a fixed bandwidth and thus the transformed remainder terms cannot be negligible. The consequence of all that is a non-negligible asymptotic bias. Here, we propose to introduce a multivariate kernel function around the vector of values $Z_{i1},\ldots,Z_{iT}$. This modified version of a local linear regression, as it will be shown later, solves the problem of the bias but it considerably enlarges the variance. More precisely, under rather standard conditions in the next section we show that, asymptotically, the bias term is of order $O\left(h^2\right)$ but the variance is of order $O\left(1/Nh^T\right)$. As the reader may notice, this bound for the variance is rather large. In order to reduce the variance term but keeping the bias of the same order we propose to add to both terms in (2.3) the average term $\frac{1}{T}\sum_{s}X_{is}m\left(Z_{is}\right)$ and denote

$$\ddot{Y}_{it}^{*}=\ddot{Y}_{it}+\frac{1}{T}\sum_{s=1}^{T}X_{is}m\left(Z_{is}\right). \tag{2.15}$$

Therefore, combining (2.3) and (2.15) we obtain

$$\ddot{Y}_{it}^{*}=X_{it}m\left(Z_{it}\right)+\ddot{v}_{it},\quad i=1,\ldots,N;\; t=1,\ldots,T, \tag{2.16}$$

where $\ddot{v}_{it}=v_{it}-\frac{1}{T}\sum_{s}v_{is}$. Note that Eq. (2.16) already shows a low dimensional problem where $m\left(\cdot\right)$ could be estimated by a standard nonparametric regression method. Unfortunately, the functions $m\left(Z_{i1}\right),\ldots,m\left(Z_{iT}\right)$ are not observed and the standard locally weighted least-squares procedures would generate unfeasible estimators. To overcome this situation, we propose to replace in (2.15) the $m(Z_{is})$ by their corresponding estimators, $\widehat{m}_h(Z_{is})$, in (2.13). Then, let $\ddot{Y}^b_{it} = \ddot{Y}_{it} + T^{-1}\sum_{s=1}^T X_{is}\widehat{m}_h(Z_{is})$ be the regression problem becomes

$$\ddot{Y}^b_{it} = X_{it} m(Z_{it}) + \ddot{v}^b_{it}, \quad i = 1, \dots, N;\ t = 1, \dots, T, \tag{2.17}$$

where the composed error term is of the form

$$\ddot{v}^b_{it} = \frac{1}{T}\sum_{s=1}^{T} X_{is}\left(\widehat{m}_h(Z_{is}) - m(Z_{is})\right) + \ddot{v}_{it}.$$

The quantities of interest can be obtained by minimizing the following criterion function

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}^b_{it} - \gamma_0 X_{it} - \gamma_1 X_{it}(Z_{it} - z)\right)^2 K_{\widetilde{h}}(Z_{it} - z), \tag{2.18}$$

where $\widetilde{h}$ is the bandwidth of this stage. We denote by $\widetilde{\gamma}_0$ and $\widetilde{\gamma}_1$ the minimizers of (2.18). As previously, we propose as estimators for $m(\cdot)$ and $m'(\cdot)$, $\widetilde{m}_{\widetilde{h}}(z) = \widetilde{\gamma}_0$ and $\widetilde{m}'_{\widetilde{h}}(z) = \widetilde{\gamma}_1$, respectively,

$$\begin{pmatrix}\widetilde{\gamma}_0\\ \widetilde{\gamma}_1\end{pmatrix} = \left(\sum_{it} K_{\widetilde{h}}(Z_{it} - z)\widetilde{Z}^b_{it}\widetilde{Z}^{b\top}_{it}\right)^{-1}\sum_{it} K_{\widetilde{h}}(Z_{it} - z)\widetilde{Z}^b_{it}\ddot{Y}^b_{it}, \tag{2.19}$$

where $\widetilde{Z}^{b\top}_{it} = (X_{it},\ X_{it}(Z_{it} - z))$ is a 2×1-dimensional vector.

Finally, for the sake of comparison the local constant version of the backfitting estimator will be

$$\widetilde{m}_{\widetilde{h}}(z) = \frac{\sum_{it} K_{\widetilde{h}}(Z_{it} - z) X_{it}\ddot{Y}^b_{it}}{\sum_{it} K_{\widetilde{h}}(Z_{i1} - z) X^2_{it}}. \tag{2.20}$$

Taking into account that $\ddot{Y}^b_{it} = \ddot{Y}_{it} + T^{-1}\sum_{s=1}^T X_{is}\widehat{m}_h(Z_{is})$ (2.20) can be written as

$$\widetilde{m}_{\widetilde{h}}(z) = \frac{\sum_{it} K_{\widetilde{h}}(Z_{it} - z) X_{it}\ddot{Y}_{it}}{\sum_{it} K_{\widetilde{h}}(Z_{it} - z) X^2_{it}} + \frac{T^{-1}\sum_{its} K_{\widetilde{h}}(Z_{it} - z) X_{it}X_{is}\widehat{m}_h(Z_{is})}{\sum_{it} K_{\widetilde{h}}(Z_{it} - z) X^2_{it}}. \tag{2.21}$$

3. Asymptotic properties

In this section we extend the above results for the case ($d > 1, q > 1$). Furthermore, we give the asymptotic expressions for the bias and the variance and we calculate the asymptotic distribution of the local linear regression estimator. Finally, we compare theoretically the results obtained in [18] for the profile least-squares estimator against our estimators.

3.1. Local linear estimator

Let us consider (2.12) in its multivariate version,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it} - \widetilde{Z}^\top_{it}\beta\right)^2 \prod_{\ell=1}^{T} K_H(Z_{i\ell} - z), \tag{3.1}$$

where in this case $\beta = \begin{pmatrix}\beta_0^\top & \beta_1^\top\end{pmatrix}^\top$ is a $d(1+q) \times 1$ vector and we denote by $\widetilde{Z}^\top_{it}$ a $1 \times d(1+q)$ dimensional vector of the form

$$\widetilde{Z}^\top_{it} = \left(\ddot{X}^\top_{it},\quad X^\top_{it} \otimes (Z_{it} - z)^\top - T^{-1}\sum_{s=1}^{T} X^\top_{is} \otimes (Z_{is} - z)^\top\right).$$

Let H be a $q \times q$ symmetric positive definite bandwidth matrix, K is the product of q-variate kernels such that for each u it holds

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

Let us denote by $\widehat{\beta}$ the minimizer of (3.1) and assuming $\widetilde{Z}^\top W\widetilde{Z}$ is nonsingular, the solution can be written as

$$\begin{pmatrix}\widehat{\beta}_0\\ \widehat{\beta}_1\end{pmatrix} = \left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\ddot{Y}, \tag{3.2}$$

where $\ddot{Y} = (\ddot{Y}_{11}, \ldots, \ddot{Y}_{NT})$ is a $NT \times 1$ vector while

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

and

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

are $NT \times NT$ and $NT \times d\,(1+q)$ dimensional matrix, respectively.

Then, (3.1) and (3.2) suggest as estimators for $m(z)$ and $D_m(z) = \partial m(z)/\partial z$, $\widehat{m}\,(z; H) = \widehat{\beta}_0$ and $\text{vec}(\widehat{D}_m\,(z; H)) = \widehat{\beta}_1$, respectively. In particular, the local weighted linear least-squares estimator of $m(z)$ is defined as

$$\widehat{m}\,(z; H) = \widehat{\beta}_0 = e_1^T \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W \ddot{Y}, \tag{3.3}$$

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.

Once the estimator in its closed form is defined, let us consider the assumptions required to obtain its asymptotic properties. Consider the data generating process defined in (2.2). Furthermore, we assume the following.

Assumption 3.1. Let $(Y_{it}, X_{it}, Z_{it})_{i=1,\ldots,N;\ t=1,\ldots,T}$ 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 3.2. 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 3.3. The unobserved cross-sectional effect, μ_i, can be arbitrarily correlated with both X_{it} and/or Z_{it} with an unknown correlation structure.

Assumption 3.1 is standard in panel data analysis. We could consider other settings of time-dependence such as strong mixing conditions, as in [1], or nonstationary time series, as in [2]. However, since in this paper we investigate the asymptotic properties of the estimators as N tends to infinity and T is fixed it is enough to assume stationarity. Assumption 3.2 is also standard for the conventional within transformation; see [19] or [7] for the fully parametric case. It also rules out the presence of lagged endogenous variables. Independence between the idiosyncratic error term and the covariates X_{it} and/or Z_{it} is assumed without loss of generality although it can be relaxed assuming some dependence in higher order moments. In particular, if heteroskedasticity of unknown form is allowed in our setting, we could transform this estimator to take into account more complex structures of the random error term contained in the variance–covariance matrix, see [10] or [20] for more details.

Assumptions 3.1 and 3.2 in some situations, as in [1], are relaxed by considering that (X_{it}, Z_{it}, v_{it}) are for fixed, i, strictly stationary processes. Unfortunately, this set of assumptions is not sufficient to bound the asymptotic variance of the estimator and some further mixing conditions are required to achieve convergence. In this case, T must also tend to infinity. Other cases such as cross sectional dependence also requires both N and T tending to infinity. Finally, Assumption 3.3 imposes the so-called fixed effects.

Let $\mathbb{Z} = (Z_{11}, \ldots, Z_{NT})$ and $\mathbb{X} = (X_{11}, \ldots, X_{NT})$ be the observed covariate samples, we also need to impose the following additional assumptions about moments and densities.

Assumption 3.4. Let $f_{Z_{1t}}(\cdot)$ be the probability density function of Z_{1t}, for $t = 1, \ldots, T$. 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 3.5. The function $E\left[\ddot{X}_{it}\ddot{X}_{it}^\top | Z_{i1} = z_1, \ldots, Z_{iT} = z_T\right]$ is positive definite for any interior point of $(z_1, z_2, \ldots, z_T)$ in the support of $f_{Z_{i1},\ldots,Z_{iT}}(z_1, z_2, \ldots, z_T)$.

Assumption 3.6. Let $\|A\| = \sqrt{\text{tr}\left(A^\top A\right)}$, then $E\left[\|X_{it}X_{it}^\top\|^2 | Z_{i1} = z_1, \ldots, Z_{iT} = z_T\right]$ is bounded and uniformly continuous in its support. Furthermore, the matrix functions $E\left[X_{it}X_{is}^\top | Z_{i1} = z_1, \ldots, Z_{iT} = z_T\right]$, for $t = s$ and $t \neq s$, and $E\left[\ddot{X}_{it}X_{is}^\top | Z_{i1} = z_1, \ldots, Z_{iT} = z_T\right]$, for $t = s$ and $t \neq s$, are bounded and uniformly continuous in their support.

Assumption 3.7. Let z be an interior point in the support of $f_{Z_{1t}}$. All second-order derivatives of $m_1(\cdot), m_2(\cdot), \ldots, m_d(\cdot)$ are bounded and uniformly continuous.

Assumption 3.8. 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 $\iota_1, \ldots, \iota_q$ such that their sum is odd.

Assumption 3.9. The bandwidth matrix H is symmetric and strictly positive definite. Furthermore, each entry of the matrix tends to zero as $N \to \infty$ in such a way that $N|H| \to \infty$.

Assumption 3.10. For some $\delta > 0$, the following functions $E\left[|X_{it}v_{it}|^{2+\delta}|Z_{i1} = z_1, \ldots, Z_{iT} = z_T\right]$ and $E\left[|\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1} = z_1, \ldots, Z_{iT} = z_T\right]$ are bounded and uniformly continuous in any point of their support.

This second set of assumptions is more directly related to nonparametric statistics literature. They are basically smoothness and boundedness conditions. Assumption 3.4 imposes smoothness conditions in the probability density function of Z_{1t}, for $t = 1, \ldots, T$. Furthermore, Assumptions 3.5–3.6 are smoothness conditions on moment functionals. Assumptions 3.7–3.9 are standard in the literature of local linear regression where, in particular, Assumption 3.9 contains a standard bandwidth condition for smoothing techniques. Finally, Assumption 3.10 is required to show that the Lyapunov conditions holds for the Central Limit Theorem.

Under these assumptions we obtain the following asymptotic expressions for the conditional bias and conditional variance–covariance matrix of the local weighted linear least-squares estimator.

Theorem 3.1. *Assume conditions 3.1–3.3 and 3.4–3.9 hold, then as $N \to \infty$ and T is fixed we obtain*

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)\left(\mu_2(K_{u_\tau})\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z) - \frac{1}{T}\sum_{s=1}^{T}\mu_2\left(K_{u_s}\right)\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\right) \times \mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}(z)H))\iota_d + o_p\left(\mathrm{tr}(H)\right)$$

and

$$\mathrm{Var}\left(\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right) = \frac{\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)}{N|H|^{T/2}}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)\left(1+o_p(1)\right),$$

where τ is any index between 1 and T,

$$\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z) = E\left[\ddot{X}_{it}X_{it}^\top|Z_{i1} = z,\ldots,Z_{iT} = z\right]f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z),$$
$$\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z) = E\left[\ddot{X}_{it}X_{is}^\top|Z_{i1} = z,\ldots,Z_{iT} = z\right]f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z),$$
$$\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z) = E\left[\ddot{X}_{it}\ddot{X}_{it}^\top|Z_{i1} = z,\ldots,Z_{iT} = z\right]f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z),$$

$\mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}(z)H))$ *stands for a diagonal matrix of elements* $\mathrm{tr}(\mathcal{H}_{m_r}(z)H)$, *for $r = 1, \ldots, d$, where $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the rth component of $m(\cdot)$. Finally, we denote by ι_d is a $d \times 1$ unit vector.*

The proof of this result is done in the Appendix.

This theorem shows that $\widehat{m}(z;H)$ is, conditionally on the sample, a consistent estimator of $m(z)$. Furthermore, as it was already remarked in the previous section, although the bias shows the standard order of magnitude for this type of problems, the variance shows an asymptotic expression that is larger than the expected in this type of problems. In order to achieve an optimal rate of convergence, the variance term must be of order $1/N|H|^{1/2}$ whereas our result shows a bound of order $1/N|H|^{T/2}$. Just to clarify the asymptotic behavior of the estimator we show its properties for the univariate case, $d = q = 1$ and $H = h^2I$.

Corollary 3.1. *Assume conditions 3.1–3.9 hold, then if $h \to 0$ in such a way that $Nh^2 \to \infty$ as N tends to infinity and T is fixed we get*

$$E\left[\widehat{m}(z;H)|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}c(z,z)\,m''(z)h^2 + o_p\left(h^2\right),$$

where

$$c(z,z) = \frac{\mu_2\left(K_{u_\tau}\right)E\left[\ddot{X}_{it}X_{it}|Z_{i1} = z,\ldots,Z_{iT} = z\right] - T^{-1}\sum_{s=1}^{T}\mu_2\left(K_{u_s}\right)E\left[\ddot{X}_{it}X_{is}|Z_{i1} = z,\ldots,Z_{iT} = z\right]}{E\left[\ddot{X}_{it}^2|Z_{i1} = z,\ldots,Z_{iT} = z\right]}.$$

Furthermore, if $\mu_2\left(K_{u_1}\right)=\cdots=\mu_2\left(K_{u_T}\right)=\mu_2\left(K_{u_\tau}\right)=\mu_2(K)$ *then the bias term has the following expression*

$$E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right]-m(z)=\frac{1}{2}\mu_2(K)m''(z)h^2+o_p\left(h^2\right),$$

whereas if $R\left(K_{u_1}\right)=\cdots=R\left(K_{u_T}\right)=R(K)$ *the variance–covariance matrix is*

$$\mathrm{Var}\left(\widehat{m}(z;H)\,|\mathbb{X}\right)=\frac{\sigma_v^2R(K)^T}{Nh^2f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z)\,E\left[\ddot{X}_{it}^2|Z_{i1}=z,\ldots,Z_{iT}=z\right]}\left(1+o_p(1)\right).$$

As a tool to construct asymptotic confidence bands we give also a result that provides the asymptotic distribution of the estimator.

Theorem 3.2. *Assume conditions 3.1–3.3 and 3.4–3.10 hold, then as* $N\to\infty$ *and* T *is fixed we obtain*

$$\sqrt{N|H|^{T/2}}\left(\widehat{m}(z;H)-m(z)\right)\xrightarrow{d}\mathcal{N}\left(b(z),\upsilon(z)\right),$$

where

$$b(z)=\frac{1}{2}\mu_2(K_u)\mathrm{diag}_d\left(\mathrm{tr}\left(\mathcal{H}_{m_r}(z)H\sqrt{N|H|^{T/2}}\right)\right)\iota_d,$$
$$\upsilon(z)=\sigma_v^2R(K)^T\,\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z).$$

The proof of this result is shown in the Appendix.

We can compare the results obtained here with those in [13] for the first differences case. As expected, the bias term presents for both estimators the same linear dependence in the trace of the bandwidth matrix H. However, the variance term differs from one to the other estimator. In the first differences case, see Theorem 3.1 in [13], up to a constant, the variance term exhibits a dependence from the bandwidth matrix H of order $1/N|H|$ whereas in our case it is of order $1/N|H|^{T/2}$. That is, the ratio between the first differences and the deviances from the mean estimators is of order $|H|^{(T-2)/2}$. For $T=2$, the estimators show the same rate of convergence. This is clearly expected. For $T>2$, the first differences estimator under the conditions established above shows a faster rate of convergence for the variance terms as far as the diagonal elements of the bandwidth matrix H tend to zero. This was also expected because the dimensionality of the kernel used in the local linear regression procedure is different in both cases. Of course, efficiency issues are not considered here and they will clearly depend on the stochastic structure of the idiosyncratic errors.

3.2. The backfitting estimator

As we stated previously the function of interest can be consistently estimated by using a local linear regression approach with a high dimensional kernel weight, but at the price of achieving a slow rate of convergence. However, as it is noted in Section 2, we can solve this problem turning to an alternative procedure that enables us to cancel asymptotically all additive terms expected in the model the function of interest.

Let us consider the multivariate version of (2.17) and define

$$\ddot{Y}_{it}^b=X_{it}^\top m\left(Z_{it}\right)+\ddot{v}_{it}^b,\quad i=1,\ldots,N;\ t=1,\ldots,T,\tag{3.4}$$

where

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

The quantities of interest in (3.4) can be estimated by minimizing the following locally weighted linear regression

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\ddot{Y}_{it}-\widetilde{Z}_{it}^{b\top}\gamma\right)^2K_{\widetilde{H}}\left(Z_{it}-z\right),\tag{3.5}$$

where $\widetilde{H}$ is a $q\times q$ symmetric positive definite bandwidth matrix, $\gamma=\left(\gamma_0^\top\quad\gamma_1^\top\right)^\top$ is a $d(1+q)\times1$ vector and $\widetilde{Z}_{it}^{b\top}=\left(X_{it}^\top\ X_{it}^\top\otimes\left(Z_{it}-z\right)^\top\right)$ is a $1\times(1+q)$ vector.

Furthermore, let the vector $\widetilde{\gamma}=\left(\widetilde{\gamma}_0^\top\ \widetilde{\gamma}_1^\top\right)^\top$ be the minimizer of (3.5). As estimators of $m(z)$ and $D_m(z)=\partial m(z)/\partial z$, we suggest $\widetilde{m}(z,\widetilde{H})=\widetilde{\gamma}_0$ and $\mathrm{vec}(\widetilde{D}_m(z;\widetilde{H}))=\widetilde{\gamma}_1$, respectively, i.e.,

$$\widetilde{m}(z;\widetilde{H})=\widetilde{\gamma}_0=e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\ddot{Y}^b,\tag{3.6}$$

where $\ddot{Y}^b=(\ddot{Y}_{11}^b,\ldots,\ddot{Y}_{NT}^b)$ is a NT-vector and W^b and $\widetilde{Z}^b$ are $NT\times NT$ and $NT\times d(1+q)$ dimensional matrix, respectively, of the form

$$W^b=\mathrm{diag}\left(K_{\widetilde{H}}\left(Z_{11}-z\right),\ldots,K_{\widetilde{H}}\left(Z_{NT}-z\right)\right)$$

and

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

We now study the asymptotic behavior of the so called backfitting estimator. At this stage we need the results shown in Theorem 3.1 to hold uniformly in z. In order to do so, we can rely on the well-known results in [11]. In fact, some of the conditions already enounced in Section 3.1 are sufficient to show the uniform rates for $\widehat{m}(z; H)$. However, we need some additional assumptions that relate the bandwidths of both $\widehat{m}(z; H)$ and $\widetilde{m}(z; \widetilde{H})$.

Assumption 3.11. The bandwidth matrix $\widetilde{H}$ is symmetric and strictly positive definite. Furthermore, each entry of the matrix tends to zero as N tends to infinity in such a way that $N|\widetilde{H}| \to \infty$.

Assumption 3.12. The bandwidth matrices H and $\widetilde{H}$ must fulfill that $N\,|H|\,|\widetilde{H}|/\log(N) \to \infty$, and $\mathrm{tr}\,(H)\,/\mathrm{tr}(\widetilde{H}) \to 0$ as N tends to infinity.

These assumptions are needed in order to ensure that both bias and variance terms of the backfitting estimator achieve optimal rates of convergence and they are oracle efficient.

Then, under these assumptions we get the following asymptotic expressions for the conditional bias and conditional variance–covariance matrix of $\widetilde{m}(z; \widetilde{H})$.

Theorem 3.3. *Assume conditions 3.1–3.8 and 3.11–3.12 hold, then as $N \to \infty$ and T is fixed we obtain*

$$E[\widetilde{m}(z; \widetilde{H})|\mathbb{X}, \mathbb{Z}] - m(z) = \frac{1}{2}\mu_2(K_u)\,\mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}(z)\widetilde{H}))\imath_d + o_p(\mathrm{tr}(\widetilde{H}))$$

and

$$\mathrm{Var}(\widetilde{m}(z; \widetilde{H})|\mathbb{X}, \mathbb{Z}) = \frac{\sigma_v^2 R(K)}{NT|\widetilde{H}|^{1/2}} \mathcal{B}_{X_tX_t}^{-1}(z)\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z)\mathcal{B}_{X_tX_t}(z)^{-1}\left(1 + o_p(1)\right),$$

where $\mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{mr}(z)\widetilde{H}))$ *stands for a diagonal matrix of elements* $\mathrm{tr}(\mathcal{H}_{m_r}(z)\widetilde{H})$*, for* $r = 1, \ldots, d$ *and* $\imath_d$ *is a* $d \times 1$ *unit vector.*

The proof of this result is done in the Appendix.

On one hand, we realize that the bias term is influenced by the amount of smoothing, H, as well as the curvature of $m(z)$ at z in a particular direction measured through each entry of $\mathcal{H}_m(z)$. In this way, we can guess that this estimator exhibits a higher conditional bias when there is a higher curvature and more smoothing. On the other hand, from the standpoint of the conditional variance we can see that it is a bit different from the corresponding for the standard case. In particular, it will be increased when the smoothing is lower and sparse data near z but now also depends on the time-demeaned covariates $\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z)$. Regardless, it is proved that the estimation procedure developed in this paper provides a nonparametric estimator in which the variance–covariance matrix of all its components is asymptotically the same as if we would known the rest of components of the mean deviation transformed expression, the so-called oracle efficiency property.

3.3. Comparison of the estimators

As we have already remarked in Section 2, the main difference among the estimators (for their local constant version) consists in the types of averages that are used in order to remove fixed effects. In one case, the one step backfitting algorithm considers un-smoothed averages whereas in the profile least-squares case smoothed weighted averages are preferred. There exists also a difference between the dimension of the kernel weights. All these differences should have an impact in both bias and variances of the estimators and therefore it would be of interest to analyze them, both theoretically and empirically. This subsection will be devoted to analyze the estimators theoretically whereas in Section 4 we will do it empirically through Monte Carlo simulations.

The reader might have noticed that the conditions required to obtain the asymptotic properties of the first step fixed effects estimator and the backfitting estimator (see Theorems 3.1 and 3.2) are rather different from the conditions assumed in [18] to obtain the properties of their estimator. For the sake of comparison, in this section we introduce additional assumptions that will be used to obtain asymptotic terms that can be comparable among the three estimators. In all calculus we will assume that N tends to infinity keeping T fixed. Furthermore, we will remove the strict stationarity assumption established in the previous sections and we will not be willing to impose that $\sum_i \mu_i = 0$. Finally it is important to note that, in the profile least-squares estimator, for fixed T, it is not possible to obtain explicitly the asymptotic bias and variance of the estimator since ϖ_{it} is random.

In order to compare the main statistical properties of these estimators, we extend the above results assuming the following.

Assumption 3.13. Let $(Y_{it}, X_{it}, Z_{it})_{i=1,\ldots,N;t=1,\ldots,T}$ be a set of independent and identically distributed random variables in the subscript i. Furthermore, let $f_t(\cdot)$ the p.d.f. of Z_{it} and $f_t(\cdot,\ldots,\cdot)$ be the p.d.f. of $(Z_{i1},\ldots,Z_{iT})$, for each $z \in \mathbb{R}^q$, $f(z) = \sum_{t=1}^{T} f_t(z) > 0$ and $f(z,\ldots,z) = \sum_{t=1}^{T} f_t(z,\ldots,z) > 0$.

Assumption 3.14. Let $f_t(z)$ be the p.d.f. of Z_{it} and $f_{t,s}(z_1, z_2)$ be the joint p.d.f. of (Z_{it}, Z_{js}) for $t \neq s$ and any i, j. We can assume that $f_t(z)E(X_{it}X_{it}^{\top}|Z_{it} = z)$ and $f_{t,s}(z)E(X_{it}X_{js}^{\top}|Z_{it} = z_1, Z_{is} = z_2)$ are uniformly bounded in the domain of Z and are all twice continuously differentiable at $z \in \mathbb{R}^q$ for all $t \neq s$ and i and j.

Assumption 3.15. Let g be a bandwidth, the bandwidth matrix G is symmetric and strictly positive definite. Furthermore, each entry of the matrix tends to zero as $N \to \infty$ in such a way that $N|G| \to \infty$.

Note that Assumption 3.13 is a standard data generating condition in this context but stationarity is not allowed. 3.14 is a smoothness assumption and 3.15 is the standard bandwidth condition. For the sake of comparison, we give the results for the univariate case ($d = q = 1$), where now $H = h^2 I$ and $\widetilde{H} = \widetilde{h}^2 I$, and obtain the following results.

Corollary 3.2. *Assume conditions 3.2–3.9 and 3.13 holds, as $N \to \infty$ and T is fixed, then we obtain*

$$E\left[\widehat{m}_h(z)\,|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}\frac{\Lambda_L(z)}{\Psi_L(z)} + o_p(h^2)$$

$$\mathrm{Var}\left(\widehat{m}_h(z)\,|\mathbb{X},\mathbb{Z}\right) = \frac{\sigma_v^2}{Nh^T}\frac{\Gamma_L(z)}{\Psi_L(z)^2} + o_p\left(\frac{1}{Nh^T}\right),$$

where, for any ξ_{it} between Z_{it} and z, $r(\xi_{it}, z) = (Z_{it} - z)^2 \frac{\partial^2 m(\xi_{it})}{\partial z^2}$ and

$$\Psi_L(z) = \frac{1}{Th^T}\sum_{t=1}^{T} E[\ddot{X}_{it}^2 \lambda_i],$$

$$\Gamma_L(z) = \frac{1}{Th^T}\sum_{t=1}^{T} E[\ddot{X}_{it}^2 \lambda_i^2],$$

$$\Lambda_L(z) = \frac{1}{Th^T}\sum_{t=1}^{T} E\left[\ddot{X}_{it}\left(X_{it} r(\xi_{it}, z) - \frac{1}{T}\sum_{s=1}^{T} X_{is} r(\xi_{is}, z)\right)\lambda_i\right] = O_p(h^2),$$

where $\lambda_i = K\left(\frac{Z_{i1}-z}{h}\right) \times \cdots \times K\left(\frac{Z_{iT}-z}{h}\right)$.

In the Appendix, it is shown that under the same conditions established in the corollary, as N tends to infinity we obtain,

$$\Psi_L(z) = T^{-1}\sum_{t=1}^{T} \mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z) + o_p(h^2),$$

$$\Lambda_L(z) = \frac{h^2}{T}\mu_2(K)\sum_{t=1}^{T} \mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)m''(z) + o_p(h^2),$$

$$\Gamma_L(z) = \frac{\sigma_v^2}{T}\sum_{t=1}^{T} \mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z) + o_p(h^2).$$

Corollary 3.3. *Assume conditions 3.2–3.8, 3.11–3.12 and 3.13, holds, as $N \to \infty$ and T is fixed we obtain*

$$E\left[\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X},\mathbb{Z}\right] - m(z) = \frac{1}{2}\frac{\Lambda_b(z)}{\Psi_b(z)} + o_p(\widetilde{h}^2)$$

$$\mathrm{Var}\left(\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X},\mathbb{Z}\right) = \frac{\sigma_v^2}{N\widetilde{h}}\frac{\Gamma_b(z)}{\Psi_b(z)^2} + o_p\left(\frac{1}{N\widetilde{h}}\right),$$

where, let $\widetilde{\lambda}_{it} = K\left(\frac{Z_{it}-z}{\widetilde{h}}\right)$,

$$\Psi_b(z) = \frac{1}{T\widetilde{h}}\sum_{t=1}^{T} E[\ddot{X}_{it}^2 \widetilde{\lambda}_{it}],$$

$$\Gamma_b(z) = \frac{1}{T\widetilde{h}} \sum_{t=1}^{T} E[X_{it}^2 \widetilde{\lambda}_{it}^2],$$

$$\Lambda_b(z) = \frac{1}{T\widetilde{h}} \sum_{t=1}^{T} E[X_{it}^2 r(\xi_{it}, z)\widetilde{\lambda}_{it}] = O_p(\widetilde{h}^2).$$

In the Appendix, it is shown that under the same conditions established in the corollary, as N tends to infinity we obtain,

$$\Psi_b(z) = T^{-1} \sum_{t=1}^{T} \mathcal{B}_{X_t X_t}(z) + o_p(\widetilde{h}^2),$$

$$\Lambda_b(z) = \frac{\widetilde{h}^2}{T} \mu_2(K) \sum_{t=1}^{T} \mathcal{B}_{X_t X_t}(z) m''(z) + o_p(\widetilde{h}^2),$$

$$\Gamma_b(z) = \frac{\sigma_v^2}{T} \sum_{t=1}^{T} \mathcal{B}_{\ddot{X}_t \ddot{X}_t}(z) + o_p(\widetilde{h}^2).$$

The proof of these corollaries is done in the Appendix.

Under this setting, Theorem 3.1 of Sun et al. [18] can be rewritten for a univariate problem as follows.

Corollary 3.4. *Assume conditions 3.2–3.3, 3.7–3.8 and 3.13–3.15 hold, as $N \to \infty$ and T is fixed we obtain*

$$E[\widehat{m}_g(z)|\mathbb{X}, \mathbb{Z}] - m(z) = \frac{1}{2} \frac{\Lambda_p(z)}{\Psi_p(z)} + O_p\left(\frac{g \ln(\ln N)}{\sqrt{N}}\right) + o_p(g^2)$$

$$\mathrm{Var}(\widehat{m}_g(z)|\mathbb{X}, \mathbb{Z}) = \frac{\sigma_v^2}{Ng} \frac{\Gamma_p(z)}{\Psi_p(z)^2} + o_p\left(\frac{1}{Ng}\right),$$

where $\lambda_{it} = K\left(\frac{Z_{it} - z}{h}\right)$ *and*

$$\Psi_p(z) = \frac{1}{Tg} \sum_{t=1}^{T} E[(1 - \varpi_{it}) X_{it}^2 \lambda_{it}],$$

$$\Gamma_p(z) = \frac{1}{Tg} \sum_{t=1}^{T} E[(1 - \varpi_{it})^2 X_{it}^2 \lambda_{it}^2],$$

$$\Lambda_p(z) = \frac{1}{Tg} \sum_{t=1}^{T} E[(1 - \varpi_{it}) X_{it}^2 r(\xi_{it}, z)\lambda_{it}] = O_p(g^2).$$

Note that in [18] it is shown

$$\Psi_p(z) = T^{-1} \sum_{t=1}^{T} \mathcal{B}_{X_t X_t}(z) + o_p(g^2),$$

$$\Lambda_p(z) = \frac{g^2}{T} \mu_2(K) \sum_{t=1}^{T} \mathcal{B}_{X_t X_t}(z) m''(z) + o_p(g^2),$$

$$\Gamma_p(z) = \frac{\sigma_v^2}{T} \sum_{t=1}^{T} \mathcal{B}_{\ddot{X}_t \ddot{X}_t}(z) + o_p(g^2).$$

Under the set of alternative assumptions considered in this section we obtain the results shown in Corollaries 3.2–3.4. Clearly, they coincide with the results shown in Section 3.1. Corollary 3.2 points out the variance is of order $1/Nh^T$ whereas the bias shows a term that is of order $O(h^2)$. Furthermore, the backfitting estimator that is studied in Corollary 3.3 presents the correction in the variance of order $1/N\widetilde{h}$. Furthermore, Assumption 3.12, $h = o(\widetilde{h})$, is crucial to guarantee that the additional bias term vanishes asymptotically. Finally, Corollary 3.4 shows both bias and variance of the profile least-squares estimator in the univariate case. As it can be observed from the expressions the bias shows an additional term of order $O\left(g \ln(\ln(N)) / \sqrt{N}\right)$. This term does not appear in the bias expression of the other estimator. However, the variance shows the standard rate and no further procedure is needed to achieve the optimal rate as it is necessary in our case.

Corollary 3.5. *As $N \to \infty$ and $h \to 0$ and $g \to 0$ we obtain the following bias and variance rates given a finite integer $T > 0$,*

$$\frac{\text{Bias}[\widehat{m}_g(z)|\mathbb{X},\mathbb{Z}]}{\text{Bias}[\widehat{m}_h(z)|\mathbb{X},\mathbb{Z}]} = \frac{\Lambda_p(z)\Psi_L(z)}{\Psi_p(z)\Lambda_L(z)} + O_p\left(\frac{g\ln(\ln N)}{\sqrt{N}}\right) + o_p(g^2),$$

$$\frac{\text{Var}(\widehat{m}_g(z)|\mathbb{X},\mathbb{Z})}{\text{Var}(\widehat{m}_h(z)|\mathbb{X},\mathbb{Z})} = \frac{h^\top}{g}\frac{\Gamma_p(z)\Psi_L(z)^2}{\Psi_p(z)^2\Gamma_L(z)} + o_p(1).$$

Corollary 3.6. *As $N \to \infty$ and $\widetilde{h} \to 0$ and $g \to 0$ we obtain the following bias and variance rate for $\widehat{m}(z;h)$ given a finite integer $T > 0$,*

$$\frac{\text{Bias}[\widehat{m}_g(z)|\mathbb{X},\mathbb{Z}]}{\text{Bias}[\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X},\mathbb{Z}]} = \frac{\Lambda_p(z)\Psi_b(z)}{\Psi_p(z)\Lambda_b(z)} + O_p\left(\frac{g\ln(\ln N)}{\sqrt{N}}\right) + o_p(g^2)$$

$$\frac{\text{Var}(\widetilde{m}_g(z)|\mathbb{X},\mathbb{Z})}{\text{Var}(\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X},\mathbb{Z})} = \frac{\widetilde{h}}{g}\frac{\Psi_b(z)^2\Gamma_p(z)}{\Psi_p(z)^2\Gamma_b(z)} + o_p(1).$$

Corollaries 3.5 and 3.6 show respectively relative bias and variances of the profile least-squares estimator against the local linear fixed effect estimator and the one step backfitting estimator. The ratio $\Lambda_p(z)\Psi_L(z)/\Psi_p(z)\Lambda_L(z)$ in Corollary 3.5 is easily shown to be greater than 1. Therefore, under the conditions established in the corollary, the bias of the profile least-squares estimator is larger than the fixed effect estimator. This difference is increased if we consider the term $O\left(g\ln(\ln(N))/\sqrt{N}\right)$. However, as N tends to infinity the difference between the bias of both estimators diminishes. The relative variance exhibits a term $\Gamma_p(z)\Psi_L(z)^2/\Psi_p(z)^2\Gamma_L(z)$ that is constant but the relative size of the variances of both estimators is determined by the ratio of bandwidths, i.e. h^T/g. For example, if the bandwidths converge to zero at the same rate the variance term of the profile estimator is going to be smaller than the variance term of the local linear regression of the within estimator. Corollary 3.6 shows theoretically the correction introduced by the backfitting algorithm in the variance of the estimators.

4. Monte Carlo simulations

In this section, Monte Carlo simulations are carried out in order to verify the theoretical results of the estimators proposed in this paper under the statistical setting analyzed in the previous sections. Later, we make an empirical comparison about the performance in small samples of the different estimators considered in this paper.

As it is well known, the Mean Squared Error (MSE) is a suitable measure of the estimation accuracy of the proposed estimators. Thus, let us denote φ as the φth replication and Q as the number of replications, for $r = 1, \dots, d$

$$\text{MSE}\left(\widehat{m}_r(z;H)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q} E\left[\left(\left(\widehat{m}_{\varphi r}(z;H) - m_{\varphi r}(z)\right)\right)^2\right]$$

which can be approximated by the Averaged Mean Squared Error (ASME) such as

$$\text{AMSE}\left(\widehat{m}_r(z;H)\right) = \frac{1}{Q}\sum_{\varphi=1}^{Q}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\left(\widehat{m}_{\varphi r}(z;H) - m_{\varphi r}(z)\right)\right)^2.$$

Observations are generated from the following varying coefficient panel data model of unknown form

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

where X_{dit} and Z_{qit} are random variables generated such that $X_{dit} = 0.5X_{di(t-1)} + \xi_{it}$ and $Z_{qit} = w_{qit} + w_{qi(t-1)}$, where w_{qit} is generated as an independent and identically distributed uniform random variable in $[0, \pi/2]$ and ξ_{it} is generated as an independent and identically distributed Gaussian, zero mean, variance one, random variable ($NID(0,1)$). Furthermore, v_{it} is an $NID(0,1)$ random variable and $m(\cdot)$ is a pre-specified function to be estimated.

With the aim of verifying the theoretical results in the Section 3 we consider four different data generating process (DGP)

(1) $Y_{it} = X_{1it}m_1(Z_{1it}) + \mu_{1i} + v_{it}$,
(2) $Y_{it} = X_{1it}m_1(Z_{1it}, Z_{2it}) + \mu_{2i} + v_{it}$,
(3) $Y_{it} = X_{1it}m_1(Z_{1it}) + X_{2it}m_2(Z_{2it}) + \mu_{1i} + v_{it}$,
(4) $Y_{it} = X_{1it}m_1(Z_{1it}, Z_{2it}) + X_{2it}m_2(Z_{1it}, Z_{2it}) + \mu_{2i} + v_{it}$,

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

In addition, we experiment with two specifications for the individual heterogeneity

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

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

		AMSE $m_1(\cdot)$			Relative AMSE $m_1(\cdot)$		
		LLLS	OSB	PLS	OSB/LLLS	PLS/LLLS	PLS/OSB
$T = 10$	$N = 50$	0.277	0.236	0.288	0.852	1.040	1.220
	$N = 100$	0.245	0.183	0.243	0.747	0.992	1.328
	$N = 150$	0.230	0.156	0.225	0.678	0.978	1.442

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

		AMSE $m_1(\cdot)$			Relative AMSE $m_1(\cdot)$		
		LLLS	OSB	PLS	OSB/LLLS	PLS/LLLS	PLS/OSB
$T = 10$	$N = 50$	1.236	0.515	0.533	0.417	0.431	1.035
	$N = 100$	0.793	0.415	0.450	0.523	0.567	1.084
	$N = 150$	0.634	0.382	0.418	0.603	0.659	1.094

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

		LLLS	OSB	PLS	OSB/LLLS	PLS/LLLS	PLS/OSB
		AMSE $m_1(\cdot)$			Relative AMSE $m_1(\cdot)$		
$T = 10$	$N = 50$	0.253	0.284	0.293	1.123	1.158	1.032
	$N = 100$	0.190	0.249	0.249	1.311	1.311	1.000
	$N = 150$	0.162	0.231	0.230	1.426	1.420	0.996
		AMSE $m_2(\cdot)$			Relative AMSE $m_2(\cdot)$		
$T = 10$	$N = 50$	0.098	0.048	0.086	0.490	0.878	1.792
	$N = 100$	0.053	0.025	0.042	0.472	0.792	1.672
	$N = 150$	0.038	0.018	0.029	0.474	0.763	1.611

Table 4
AMSE for $d = 2$ and $q = 2$.

		LLLS	OSB	PLS	OSB/LLLS	PLS/LLLS	PLS/OSB
		AMSE $m_1(\cdot)$			Relative AMSE $m_1(\cdot)$		
$T = 10$	$N = 50$	1.272	0.548	0.555	0.431	0.436	1.023
	$N = 100$	0.797	0.428	0.461	0.535	0.576	1.077
	$N = 150$	0.637	0.390	0.428	0.612	0.672	1.097
		AMSE $m_2(\cdot)$			Relative AMSE $m_2(\cdot)$		
$T = 10$	$N = 50$	0.823	0.189	0.158	0.230	0.192	0.836
	$N = 100$	0.466	0.088	0.078	0.188	0.167	0.886
	$N = 150$	0.355	0.060	0.053	0.169	0.149	0.883

where in both cases u_i is an *NID* $(0, 1)$ random variable and $c_0 = 0.5$ controls the correlation between the fixed effects and some of the regressors of the model.

In the experiment we use 1000 Monte Carlo replications. The number of time observations T is set up to ten, while the number of cross-sections N is either 50, 100 or 150. The Gaussian kernel has been used and the bandwidth is chosen as $\widehat{H} = \widehat{h}I = \widehat{\sigma}_z(NT)^{-1/3}$, where $\widehat{\sigma}_z$ is the sample standard deviation of $\left\{Z_{qit}\right\}_{i=1,t=1}^{N,T}$, and $\widehat{g} = \widehat{\widetilde{h}} = \widehat{\sigma}_z(NT)^{-1/5}$.

The results from the simulation are presented in Tables 1–4. For the sake of comparison we present the empirical AMSE of the three estimators that we compare in this paper: the local linear least-squares estimator (LLLS), the one-step backfitting estimator (OSB), and the profile least-squares estimator (PLS) proposed in [18].

Table 1 shows the results for DGP(1). This is the simplest case without curse of dimensionality. As expected from our theoretical findings the local linear estimator presents its best result. The profile least-squares estimator, as N grows, seems to perform better than our backfitting estimator. This might be because the second term of the bias, that is related to the fixed effects, diminishes its negative impact on the bias.

Table 2 starts reflecting the problem of the curse of dimensionality. Of course, since the variance of the local linear estimator is of order $1/NTh^{Tq}$, it is expected that the behavior of this estimator with respect to the others, in terms of AMSE, will be worse. This is indeed what we observe in the results. Furthermore, as N grows, the backfitting estimator performs slightly better than the profile least-squares estimator.

Table 3 can be compared against Table 1. In fact, the function $m_1(\cdot)$, which is the same under other DGP's, presents similar results in terms of AMSE. That is, the estimator that presents the better performance is the local linear. On the contrary, the function $m_2(\cdot)$ is better estimated using either the one-step backfitting or the profile least-squares estimators. This can be related with the oracle efficiency property of these estimators.

Table 4 can be compared against Table 2. In fact, we obtain similar conclusions as in the comparison between DGP's 1 and 3. That is, the function $m_1(\cdot)$ is estimated as the same level of accuracy as if $m_2(\cdot)$ were known. Both the profile least-squares and the one-step backfitting estimators perform much better than the local linear estimator. This is the curse of dimensionality. We can say the same for $m_2(\cdot)$ but in this case the profile least-squares estimator performs slightly better then the backfitting estimator.

Acknowledgments

The authors acknowledge financial support from the Programa Estatal de Fomento de la Investigación Cientfica y Técnica de Excelencia/Spanish Ministery of Economy and Competitiveness. Ref. ECO2013-48326-C2-2-P. This paper was presented at the Econometrics Seminar in Universidad Carlos III de Madrid and the Applicable Semiparametrics Conference held at Humboldt Universität zu Berlin in October, 2013. We acknowledge all comments and suggestions made by participants. The authors would also like to thank the Associate Editor and two referees by their very helpful reports. This paper would not have been the same without their help. Of course the errors are ours.

Appendix

Proof of Theorem 3.1. We first focus on the analysis of the conditional bias of the local weighted linear least-squares estimator, $\widehat{m}(z; H)$, and later on the behavior of its conditional variance–covariance matrix. We follow the standard proofs scheme as in [13].

Let $\mathbb{X} = (X_{11}, \dots, X_{NT})$ and $\mathbb{Z} = (Z_{11}, \dots, Z_{NT})$ be the observed covariate vectors. By Assumption 3.2 we know that $E(v_{it}|\mathbb{X}, \mathbb{Z}) = 0$, so the conditional expectation on (3.3) provides

$$E[\widehat{m}(z; H)|\mathbb{X}, \mathbb{Z}] = e_1^\top \left(\widetilde{Z}^\top W \widetilde{Z}\right)^{-1} \widetilde{Z}^\top W M, \tag{A.1}$$

where

$$M = \left[X_{11}^\top m(Z_{11}) - T^{-1}\sum_{s=1}^{T} X_{1s}^\top m(Z_{1s}), \quad \dots, \quad X_{NT}^\top m(Z_{NT}) - T^{-1}\sum_{s=1}^{T} X_{Ns}^\top m(Z_{Ns})\right]^T.$$

Approximating M using the multivariate Taylor theorem we obtain

$$M = \widetilde{Z}\begin{bmatrix} m(z) \\ \mathrm{vec}(D_m(z)) \end{bmatrix} + \frac{1}{2}Q_m(z) + R(z), \tag{A.2}$$

where

$$\begin{aligned} Q_m(z) &= S_m(z) - \overline{S}_m(z), \\ S_m(z) &= \left[S_{m_{11}}^\top(z), \dots, S_{m_{NT}}^\top(z)\right]^\top, \\ \overline{S}_m(z) &= \left[\overline{S}_{m_{11}}^\top(z), \dots, \overline{S}_{m_{NT}}^\top(z)\right]^\top. \end{aligned} \tag{A.3}$$

The corresponding entries of these vectors are

$$\begin{aligned} S_{m_{it}}(z) &= \left[(X_{it} \otimes (Z_{it} - z))^\top \mathcal{H}_m(z)(Z_{it} - z)\right], \\ \overline{S}_{m_{it}}(z) &= \left[\frac{1}{T}\sum_{s=1}^{T}(X_{is} \otimes (Z_{is} - z))^\top \mathcal{H}_m(z)(Z_{is} - z)\right], \end{aligned}$$

where we denote by

$$\mathcal{H}_m(z) = \begin{pmatrix} \mathcal{H}_{m1}(z) \\ \mathcal{H}_{m2}(z) \\ \vdots \\ \mathcal{H}_{md}(z) \end{pmatrix}$$

a $dq \times d$ matrix such that $\mathcal{H}_{md}(z)$ is the Hessian matrix of the dth component of $m(\cdot)$.

On the other hand, the remainder term of the Taylor approximation can be written as

$$\begin{aligned} R(z) &= R_m(z) - \overline{R}_m(z), \\ R_m(z) &= \left[R_{m_{11}}^\top(z), \dots, R_{m_{NT}}^\top(z)\right]^\top, \\ \overline{R}_m(z) &= \left[\overline{R}_{m_{11}}^\top(z), \dots, \overline{R}_{m_{NT}}^\top(z)\right]^\top, \end{aligned} \tag{A.4}$$

where the corresponding entry of each vector are

$$R_{m_{it}}(z) = \left[(X_{it} \otimes (Z_{it} - z))^{\top} \mathcal{R}(Z_{it}; z)(Z_{it} - z)\right],$$
$$\overline{R}_{m_{it}}(z) = \left[\frac{1}{T}\sum_{s=1}^{T}(X_{is} \otimes (Z_{is} - z))^{\top} \mathcal{R}(Z_{is}; z)(Z_{is} - z)\right].$$

We denote by

$$\mathcal{R}(Z_{it}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{it}; z) \\ \mathcal{R}_2(Z_{it}; z) \\ \vdots \\ \mathcal{R}_d(Z_{it}; z) \end{pmatrix}, \qquad \mathcal{R}(Z_{is}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{is}; z) \\ \mathcal{R}_2(Z_{is}; z) \\ \vdots \\ \mathcal{R}_d(Z_{is}; z) \end{pmatrix},$$

and

$$\mathcal{R}_d(Z_{it}; z) = \int_0^1 \left[\frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z + \omega(Z_{it} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z)\right](1 - \omega)\, d\omega, \tag{A.5}$$

$$\mathcal{R}_d(Z_{is}; z) = \int_0^1 \left[\frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z + \omega(Z_{is} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^{\top}}(z)\right](1 - \omega)\, d\omega, \tag{A.6}$$

where ω is a weight function.

If we replace (A.2) in (A.1) we obtain the conditional bias expression consisting in the following two additive terms

$$E\left[\widehat{m}(z; H) | \mathbb{X}, \mathbb{Z}\right] - m(z) = \frac{1}{2} e_1^{\top} \left(\widetilde{Z}^{\top} W \widetilde{Z}\right)^{-1} \widetilde{Z}^{\top} W Q_m(z) + e_1^{\top} \left(\widetilde{Z}^{\top} W \widetilde{Z}\right)^{-1} \widetilde{Z}^{\top} W R(z), \tag{A.7}$$

where we can appreciate that the $\mathrm{vec}(D_m(z))$ term of (A.2) vanishes by the effect of e_1.

As this bias expression has two additive terms, to obtain the conditional bias of this estimator we must analyze both terms of (A.7) separately. Focus first on the analysis of $e_1^{\top} \left(\widetilde{Z}^{\top} W \widetilde{Z}\right)^{-1} \widetilde{Z}^{\top} W Q_m(z)$, we show that is the leading term of the expression of bias and which actually sets the order of this estimator. Later, we study the behavior of $e_1^{\top} \left(\widetilde{Z}^{\top} W \widetilde{Z}\right)^{-1} \widetilde{Z}^{\top} W R(z)$ and explain why this term is asymptotically negligible.

For the sake of simplicity let us denote

$$\lambda_i = K\left(\frac{Z_{i1} - z}{H^{1/2}}\right) \times \cdots \times K\left(\frac{Z_{iT} - z}{H^{1/2}}\right).$$

The inverse term of (A.7) can be rewritten as the following symmetric block matrix

$$(NT)^{-1} \widetilde{Z}^{\top} W \widetilde{Z} = \begin{pmatrix} \mathcal{A}_{NT}^{11} & \mathcal{A}_{NT}^{12} \\ \mathcal{A}_{NT}^{21} & \mathcal{A}_{NT}^{22} \end{pmatrix} \tag{A.8}$$

where,

$$\mathcal{A}_{NT}^{11} = (NT|H|^{T/2})^{-1} \sum_{it} \ddot{X}_{it} \ddot{X}_{it}^{\top} \lambda_i$$

$$\mathcal{A}_{NT}^{12} = (NT|H|^{T/2})^{-1} \sum_{it} \ddot{X}_{it} \left(X_{it} \otimes (Z_{it} - z) - T^{-1} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z)\right)^{\top} \lambda_i,$$

$$\mathcal{A}_{NT}^{21} = (NT|H|^{T/2})^{-1} \sum_{it} \left(X_{it} \otimes (Z_{it} - z) - T^{-1} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z)\right) \ddot{X}_{it}^{\top} \lambda_i,$$

$$\mathcal{A}_{NT}^{22} = (NT|H|^{T/2})^{-1} \sum_{it} \left(X_{it} \otimes (Z_{it} - z) - T^{-1} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z)\right) \left(X_{it} \otimes (Z_{it} - z) - T^{-1} \sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z)\right)^{\top} \lambda_i.$$

Analyzing each of these terms, we first show that as N tends to infinity

$$\mathcal{A}_{NT}^{11} = \mathcal{B}_{\ddot{X}_t \ddot{X}_t}(z, \ldots, z) + o_p(1), \tag{A.9}$$

where

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

With the aim of showing this result, under the stationarity assumption and the law of iterated expectations we get

$$E\left(\mathcal{A}_{NT}^{11}\right) = \int E\left[\ddot{X}_{it}\ddot{X}_{it}^{\top}|Z_{i1} = z + H^{1/2}u_1, \ldots, Z_{iT} = z + H^{1/2}u_T\right] \times f_{Z_{i1},\ldots,Z_{iT}}\left(Z_{i1} = z + H^{1/2}u_1, \ldots, Z_{iT} = z + H^{1/2}u_T\right)\prod_{\ell=1}^{T} K\left(u_\ell\right) du_\ell$$

and by the Taylor expansion of the unknown functions and Assumptions 3.1 and 3.4 the expression (A.9) holds. However, note that to conclude this proof is necessary to turn to a law of large numbers. Therefore, we have to show that $\mathrm{Var}\left(\mathcal{A}_{NT}^{11}\right) \to 0$, as N tends to infinity. Under Assumption 3.1,

$$\mathrm{Var}\left(\mathcal{A}_{NT}^{11}\right) = \frac{1}{NT}\mathrm{Var}\left(\frac{1}{|H|^{T/2}}\ddot{X}_{it}\ddot{X}_{it}^{\top}\lambda_i\right) + \frac{1}{NT^2}\sum_{t=3}^{T}(T-t)\,\mathrm{Cov}\left(\frac{1}{|H|^{T/2}}\ddot{X}_{i2}\ddot{X}_{i2}^{\top}\lambda_i, \frac{1}{|H|^{T/2}}\ddot{X}_{it}\ddot{X}_{it}^{\top}\lambda_i\right).$$

Then, under Assumptions 3.4 and 3.6 we can show that the first element is

$$\mathrm{Var}\left(\frac{1}{|H|^{T/2}}\ddot{X}_{it}\ddot{X}_{it}^{\top}\lambda_i\right) \le \frac{C}{NT|H|^{T/2}}$$

while the second one is

$$\mathrm{Cov}\left(\frac{1}{|H|^{T/2}}\ddot{X}_{i2}\ddot{X}_{i2}^{\top}\lambda_i, \frac{1}{|H|^{T/2}}\ddot{X}_{it}\ddot{X}_{it}^{\top}\lambda_i\right) \le \frac{C'}{N|H|^{T/2}}.$$

Then, if both $NT|H|$ and $N|H|$ tend to infinity the variance term tends to zero and (A.9) follows.

Following a similar procedure we get

$$\mathcal{A}_{NT}^{12} = \mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)\left(I_d \otimes \mu_2(K_{u_\tau})H\right) - \frac{1}{T}\sum_{s=1}^{T}\mathcal{D}\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\left(I_d\otimes\mu_2(K_{u_s})H\right) + o_p(H). \tag{A.10}$$

This is because using the same reasoning,

$$\begin{aligned} E\left(\mathcal{A}_{NT}^{12}\right) = &\int E\left(\ddot{X}_{it}X_{it}^{\top}|Z_{i1} = z + H^{1/2}u_1, \ldots, Z_{iT} = z + H^{1/2}u_T\right) \\ & f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z)\otimes\left(H^{1/2}u_\tau\right)^{\top}\prod_{\ell=1}^{T}K\left(u_\ell\right)du_\ell \\ & -\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\int E\left(\ddot{X}_{it}X_{is}^{\top}\middle| Z_{i1} = z + H^{1/2}u_1, \ldots, Z_{iT} = z + H^{1/2}u_T\right) \\ & \times f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z)\otimes\left(H^{1/2}u_s\right)^{\top}\prod_{\ell=1}^{T}K\left(u_\ell\right)du_\ell \end{aligned}$$

and as N tends to infinity, $\mathrm{Var}\left(\mathcal{A}_{NT}^{12}\right) \to 0$. Then, (A.6) is shown.

Note that $\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(Z_1,\ldots,Z_T)$, for $s = 1,\ldots,T$, is defined in a similar way as in [13]. Thus, $\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(Z_1,\ldots,Z_T)$ is a $d\times dq$ gradient matrix of the form

$$\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(Z_1,\ldots,Z_T) = \begin{pmatrix} \dfrac{\partial b_{11}^{\ddot{X}_tX_t}(Z_1,\ldots,Z_T)}{\partial Z_1^{\top}} & \cdots & \dfrac{\partial b_{1d}^{\ddot{X}_tX_t}(Z_1,\ldots,Z_T)}{\partial Z_1^{\top}} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial b_{d1}^{\ddot{X}_tX_t}(Z_1,\ldots,Z_T)}{\partial Z_1^{\top}} & \cdots & \dfrac{\partial b_{dd'}^{\ddot{X}_tX_t}(Z_1,\ldots,Z_T)}{\partial Z_1^{\top}} \end{pmatrix},$$

and

$$b_{dd'}^{\ddot{X}_tX_t}(Z_1,\ldots,Z_T) = E\left[\ddot{X}_{dit}X_{d'it}\middle| Z_{i1} = Z_1,\ldots,Z_{iT} = Z_T\right]f_{Z_{i1},\ldots,Z_{iT}}(Z_1,\ldots,Z_T).$$

Finally, we obtain that as N tends to infinity

$$\mathcal{A}_{NT}^{22} = \left(1-\frac{1}{T}\right)\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes\mu_u(K_{u_\tau})H + o_p(H), \tag{A.11}$$

where

$$\mathcal{B}_{X_tX_t}(z,\ldots,z) = E\left[X_{it}X_{it}^{\top}\middle| Z_{i1} = z,\ldots,Z_{iT} = z\right]f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z).$$

Then, using the results of (A.9)–(A.11) in (A.8) we obtain

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

where

$$\begin{aligned}
\mathcal{C}_{11} &= \mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)+o_p(1),\\
\mathcal{C}_{12} &= -\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)\left(\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)\left(I_d\otimes\mu_2\left(K_{u_\tau}\right)H\right)-\frac{1}{T}\sum_{s=1}^{T}\mathcal{D}\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\left(I_d\otimes\mu_2\left(K_{u_s}\right)H\right)\right)\\
&\quad\times\left(\left(1-\frac{1}{T}\right)\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes H\mu_2(K_{u_\tau})\right)^{-1}+o_p(1),\\
\mathcal{C}_{21} &= -\left(\left(1-\frac{1}{T}\right)\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes H\mu_2(K_{u_\tau})\right)^{-1}\\
&\quad\times\left(\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)\left(I_d\otimes\mu_2\left(K_{u_\tau}\right)H\right)-\frac{1}{T}\sum_{s=1}^{T}\mathcal{D}\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\left(I_d\otimes\mu_2\left(K_{u_s}\right)H\right)\right)^{\top}\\
&\quad\times\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)+o_p(1),\\
\mathcal{C}_{22} &= \left(\left(1-\frac{1}{T}\right)\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes H\mu_2(K_{u_\tau})\right)^{-1}+o_p\left(H^{-1}\right).
\end{aligned}$$

On the other hand, following the same technique we can show that

$$\begin{aligned}
&(NT)^{-1}\widetilde{Z}^{\top}WS_m(z)\\
&=\begin{pmatrix}\left(NT|H|^{T/2}\right)^{-1}\sum_{it}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{it}-z\right)\lambda_i\\ \left(NT|H|^{T/2}\right)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{it}-z\right)\lambda_i\end{pmatrix}
\end{aligned}$$

are asymptotically equal to

$$\begin{aligned}
&\left(NT|H|^{T/2}\right)^{-1}\sum_{it}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{it}-z\right)\lambda_i\\
&=\mu_2\left(K_{u_\tau}\right)\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)\times\operatorname{diag}_d(\operatorname{tr}(\mathcal{H}_{m_r}(z)H))\iota_d+o_p\left(\operatorname{tr}(H)\right),
\end{aligned}\tag{A.13}$$

where

$$\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)=E\left[\ddot{X}_{it}X_{it}^{\top}|Z_{i1}=z,\ldots,Z_{iT}=z\right]f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z),$$

$\operatorname{diag}_d(\operatorname{tr}(\mathcal{H}_{m_r}(z)H))$ stands for a diagonal matrix of elements $\operatorname{tr}(\mathcal{H}_{m_r}(z)H)$, for $r=1,\ldots,d$, and ι_d is a $d\times 1$ unit vector. In addition,

$$\begin{aligned}
&\left(NT|H|^{T/2}\right)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{it}-z\right)\lambda_i\\
&=\int\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes(H^{1/2}u_\tau)(H^{1/2}u_\tau)^{\top}\mathcal{H}_m(z)(H^{1/2}u_\tau)\prod_{\ell=1}^{T}K(u_\ell)\,du_\ell\\
&\quad-\frac{1}{T}\sum_{s=1}^{T}\int\mathcal{B}_{X_sX_t}(z,\ldots,z)\otimes(H^{1/2}u_s)(H^{1/2}u_\tau)^{\top}\mathcal{H}_m(z)(H^{1/2}u_\tau)\prod_{\ell=1}^{T}K(u_\ell)\,du_\ell\\
&=O_p(H^{3/2}).
\end{aligned}\tag{A.14}$$

Furthermore, the terms of

$$\begin{aligned}
&(NT)^{-1}\widetilde{Z}^{\top}W\overline{S}_m(z)\\
&=\begin{pmatrix}\left(NT^2|H|^{T/2}\right)^{-1}\sum_{its}\ddot{X}_{it}\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{is}-z\right)\lambda_i\\ \left(NT^2|H|^{T/2}\right)^{-1}\sum_{its}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{is}-z\right)\lambda_i\end{pmatrix}
\end{aligned}$$

are of order

$$\left(NT^2|H|^{T/2}\right)^{-1}\sum_{its}\ddot{X}_{it}\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{is}-z\right)\lambda_i$$
$$=\frac{1}{T}\sum_{s=1}^{T}\mu_2\left(K_{u_s}\right)\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\times\mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}(z)H))\imath_d+o_p\left(\mathrm{tr}(H)\right), \tag{A.15}$$

where

$$\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)=E\left[\ddot{X}_{it}X_{is}^{\top}|Z_{i1}=z,\ldots,Z_{iT}=z\right]f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z),$$

and under the stationarity assumption, when $N\to\infty$ and T remains to be fixed we get

$$\left(NT^2|H|^{T/2}\right)^{-1}\sum_{its}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{H}_m(z)\left(Z_{is}-z\right)\lambda_i$$
$$=\int\mathcal{B}_{X_tX_s}(z,\ldots,z)\otimes(H^{1/2}u_\tau)(H^{1/2}u_s)^{\top}\mathcal{H}_m(z)(H^{1/2}u_s)\prod_{\ell=1}^{T}K(u_\ell)du_\ell$$
$$-\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathcal{B}_{X_sX_s}(z,\ldots,z)\otimes(H^{1/2}u_s)(H^{1/2}u_s)^{\top}\mathcal{H}_m(z)(H^{1/2}u_s)\prod_{\ell=1}^{T}K(u_\ell)du_\ell$$
$$=O_p(H^{3/2}). \tag{A.16}$$

Then, replacing (A.13)–(A.16) into (A.3), we can conclude

$$(NT)^{-1}\widetilde{Z}^{\top}WQ_m$$
$$=\begin{pmatrix}\mu_2(K)\left(\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)-\dfrac{1}{T}\displaystyle\sum_{s=1}^{T}\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\right)\times\mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}(z)H))\imath_d+o_p\left(\mathrm{tr}(H)\right)\\ O_p\left(H^{3/2}\right)\end{pmatrix}. \tag{A.17}$$

Focus now on the residual term of (A.7), we use the notation of the beginning of the Appendix in order to write

$$(NT)^{-1}\widetilde{Z}^{\top}WR(z)=\begin{pmatrix}\varepsilon_1(z)\\ \varepsilon_2(z)\end{pmatrix}, \tag{A.18}$$

where

$$\varepsilon_1(z)=\left(NT|H|^{T/2}\right)^{-1}\sum_{it}\ddot{X}_{it}$$
$$\times\left[\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{R}\left(Z_{it};z\right)\left(Z_{it}-z\right)-\frac{1}{T}\sum_{s=1}^{T}\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{R}\left(Z_{is};z\right)\left(Z_{is}-z\right)\right]\lambda_i \tag{A.19}$$

and

$$\varepsilon_2(z)=\left(NT|H|^{T/2}\right)^{-1}\sum_{it}\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)$$
$$\times\left[\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{R}\left(Z_{it};z\right)\left(Z_{it}-z\right)-T^{-1}\sum_{s=1}^{T}\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{R}\left(Z_{is};z\right)\left(Z_{is}-z\right)\right]\lambda_i. \tag{A.20}$$

Note that $\varepsilon_1(z)$ can be decomposed into the following two terms

$$\varepsilon_1(z)=\left(NT|H|^{T/2}\right)^{-1}\sum_{it}\ddot{X}_{it}\Bigg[\left(X_{it}\otimes(Z_{it}-z)\right)^{\top}\mathcal{R}\left(Z_{it};z\right)\left(Z_{it}-z\right)$$
$$-T^{-1}\sum_{s=1}^{T}\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\mathcal{R}\left(Z_{it};z\right)\left(Z_{is}-z\right)\Bigg]\lambda_i$$
$$+\left(NT^2|H|^{T/2}\right)^{-1}\sum_{its}\ddot{X}_{it}\left(X_{is}\otimes(Z_{is}-z)\right)^{\top}\left(\mathcal{R}\left(Z_{it};z\right)-\mathcal{R}\left(Z_{is};z\right)\right)\left(Z_{is}-z\right)\lambda_i$$
$$=\varepsilon_{11}(z)+\varepsilon_{12}(z). \tag{A.21}$$

We want to show that as $N \to \infty$,

$$E(\varepsilon_1(z)) = o_p\,(\mathrm{tr}(H)) \tag{A.22}$$

so, in order to do it, we have to analyze each term of $\varepsilon_1(z)$ separately. Starting from $\varepsilon_{11}(z)$ and by the strict stationarity we have

$$\begin{aligned} E(\varepsilon_{11}(z)) = {} & \mathcal{B}_{\ddot{X}_t X_t}(z + H^{1/2}u_1, \ldots, z + H^{1/2}u_T) \otimes (H^{1/2}u_\tau)^\top \mathcal{R}(z + H^{1/2}u_\tau; z)(H^{1/2}u_\tau) \prod_{\ell=1}^{T} K\,(u_\ell)\, du_\ell \\ & - \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \mathcal{B}_{\ddot{X}_t X_s}(z + H^{1/2}u_1, \ldots, z + H^{1/2}u_T) \otimes (H^{1/2}u_s)^\top \mathcal{R}(z + H^{1/2}u_\tau; z)(H^{1/2}u_s) \prod_{\ell=1}^{T} K\,(u_\ell)\, du_\ell. \end{aligned}$$

By definition (A.5) and Assumption 3.7,

$$\left|\mathcal{R}_d(z + H^{1/2}u_\tau; z)\right| \le \int_0^1 \varsigma(\omega\|H^{1/2}u_\tau\|)\,(1-\omega)\, d\omega, \quad \forall d,$$

where $\varsigma\,(\eta)$ is the modulus of continuity of $\frac{\partial^2 m_r}{\partial z_i \partial z_j}(z)$. Hence, by boundedness of f and $\mathcal{B}_{\ddot{X}_t X_t}$, and Assumption 3.4, for all t we get

$$\begin{aligned} E\,|\varepsilon_{11}(z)| \le {} & C_1 \iint_0^1 |(H^{1/2}u_\tau)^\top|\,|\varsigma(\omega\|H^{1/2}u_\tau\|)|\,|H^{1/2}u_\tau| d\omega \prod_\ell K(u_\ell) du_\ell \\ & + \frac{C_2}{T}\sum_s \iint_0^1 |(H^{1/2}u_s)^\top|\,|\varsigma(\omega\|H^{1/2}u_\tau\|)|\,|H^{1/2}u_s| d\omega \prod_\ell K\,(u_\ell)\, du_\ell \end{aligned}$$

and $E(\varepsilon_{11}(z)) = o_p\,(\mathrm{tr}(H))$ follows by dominated convergence.

Similarly, analyzing the second term of (A.21) and by strict stationarity we have

$$\begin{aligned} E(\varepsilon_{12}(z)) = {} & \frac{1}{T}\sum_{s=1}^{T} \int \left(\mathcal{B}_{\ddot{X}_t X_s}(z + H^{1/2}u_1, \ldots, z + H^{1/2}u_T) \otimes (H^{1/2}u_s)^\top\right) \\ & \times \left(\mathcal{R}(z + H^{1/2}u_\tau; z) - \mathcal{R}(z + H^{1/2}u_s; z)\right)(H^{1/2}u_s) \prod_{\ell=1}^{T} K\,(u_\ell)\, du_\ell, \end{aligned}$$

where, as previously, we can show

$$|E(\varepsilon_{12}(z))| \le \frac{C_3}{T}\sum_s \iint_0^1 |(H^{1/2}u_s)^\top|\,|\varsigma(\omega\|H^{1/2}u_\tau\| - \omega\|H^{1/2}u_s\|)|\,|H^{1/2}u_s| \prod_\ell K\,(u_\ell)\, du_\ell.$$

Then, proceeding as previously we have that by dominated convergence $E(\varepsilon_{12}(z)) = o_p\,(\mathrm{tr}(H))$.

Once this result (A.22) has been verified, our interest focuses on the second term of (A.21), $\varepsilon_2(z)$, with the aim of showing that as $N \to \infty$,

$$E(\varepsilon_2(z)) = O_p(H^{3/2}). \tag{A.23}$$

In order to prove this result, we follow the same lines as the proof of (A.22) and $\varepsilon_2(z)$ can be decomposed in two terms

$$\varepsilon_2(z) = \varepsilon_{21}(z) + \varepsilon_{22}(z), \tag{A.24}$$

where

$$\begin{aligned} \varepsilon_{21}(z) = {} & \left(NT|H|^{T/2}\right)^{-1}\sum_{it}\left(X_{it} \otimes (Z_{it} - z) - T^{-1}\sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z)\right) \\ & \times \left[(X_{it} \otimes (Z_{it} - z))^\top\, \mathcal{R}\,(Z_{it}; z)\,(Z_{it} - z) - T^{-1}\sum_{s=1}^{T}(X_{is} \otimes (Z_{is} - z))^\top\, \mathcal{R}\,(Z_{is}; z)\,(Z_{is} - z)\right]\lambda_i \end{aligned} \tag{A.25}$$

and

$$\begin{aligned} \varepsilon_{22}(z) = {} & \left(NT^2|H|^{T/2}\right)^{-1}\sum_{its}\left(X_{it} \otimes (Z_{it} - z) - T^{-1}\sum_{s=1}^{T} X_{is} \otimes (Z_{is} - z)\right) \\ & \times (X_{is} \otimes (Z_{is} - z))^\top\,(\mathcal{R}\,(Z_{it}; z) - \mathcal{R}\,(Z_{is}; z))\,(Z_{is} - z)\,\lambda_i. \end{aligned} \tag{A.26}$$

Applying the same arguments as for the proof of (A.22), it is straightforward to show that

$$E(\varepsilon_2(z)) = o_p(H^{3/2}). \tag{A.27}$$

Then, replacing (A.22) and (A.23) in (A.18) we get

$$(NT)^{-1}\widetilde{Z}^\top WR(z) = \begin{pmatrix} o_p\,(\mathrm{tr}(H)) \\ O_p(H^{3/2}) \end{pmatrix} \tag{A.28}$$

and substituting (A.12), (A.17) and (A.28) in (A.7), the asymptotic bias can be written as

$$\begin{aligned} E\left[\widehat{m}\,(z;H)\,|\mathbb{X},\mathbb{Z}\right] - m(z) &= \frac{1}{2}e_1^T\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W(S_m(z) - \overline{S}_m(z)) \\ &= \frac{1}{2}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}\,(z,\ldots,z)\left(\mu_2(K_{u_\tau})\mathcal{B}_{\ddot{X}_tX_t}\,(z,\ldots,z) - \frac{1}{T}\sum_{s=1}^{T}\mu_2\left(K_{u_s}\right)\mathcal{B}_{\ddot{X}_tX_s}\,(z,\ldots,z)\right)\times \mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}\,(z)\,H))\imath_d \\ &\quad + o_p\,(\mathrm{tr}(H))\,. \end{aligned}$$

For the asymptotic expression of the variance term let us define the NT vector $v = (v_1,\ldots,v_N)^\top$, where $v_i = (v_{i1},\ldots,v_{iT})^\top$. Furthermore, let $E\left(vv^\top|\mathbb{X},\mathbb{Z}\right) = \mathcal{V}$ be a $NT\times NT$ matrix that contains the V_{ij}'s matrices. By Assumption 3.2 we obtain

$$V_{ij} = E(v_iv_j^\top|\mathbb{X},\mathbb{Z}) = \sigma_v^2 I_T. \tag{A.29}$$

Denote as $Q_T = I_T - \imath_T\left(\imath_T^\top\imath_T\right)^{-1}\imath_T^\top$ a $T\times T$ symmetric and idempotent matrix with rank $T-1$, where I_T is a $T\times T$ identity matrix and $\imath_T$ a $T\times 1$ unitary vector. Furthermore, let $Q = I_N\otimes Q_T$ an $NT\times NT$ matrix. It is clear that, $\widetilde{Z} = Q\widetilde{Z}^b$ and $\ddot{v} = Qv$.

Then, substituting the previous equalities into

$$\widehat{m}\,(z;H) - E\left[\widehat{m}\,(z;H)\,|\mathbb{X},\mathbb{Z}\right] = e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\ddot{v}, \tag{A.30}$$

we obtain

$$\widehat{m}\,(z;H) - E\left[\widehat{m}\,(z;H)\,|\mathbb{X},\mathbb{Z}\right] = e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}Z^{b\top}Q^\top WQv. \tag{A.31}$$

Since Q is an idempotent matrix, the variance term of $\widehat{m}\,(z;H)$ can be written as

$$\mathrm{Var}\left(\widehat{m}\,(z;H)\,|\mathbb{X},\mathbb{Z}\right) = e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\mathcal{V}W\widetilde{Z}\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}e_1. \tag{A.32}$$

As by Assumption 3.2 the v_{it}'s are i.i.d. in the subscript i, the upper left entry of $(NT)^{-1}\widetilde{Z}^\top W\mathcal{V}W\widetilde{Z}$ is

$$\frac{\sigma_v^2}{NT|H|^T}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}\ddot{X}_{it}^\top\lambda_i^2 = \frac{\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)}{|H|^{T/2}}\mathcal{B}_{\ddot{X}_t\ddot{X}_t}\,(z,\ldots,z)\left(1+o_p(1)\right). \tag{A.33}$$

The upper right block is

$$\begin{aligned} &\frac{\sigma_v^2}{NT|H|^T}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{X}_{it}\left(X_{it}\otimes(Z_{it}-z) - T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^\top\lambda_i^2 \\ &= \frac{\sigma_v^2}{|H|^{T/2}}\int\left(\mathcal{B}_{\ddot{X}_tX_t}(z+H^{1/2}u_1,\ldots,z+H^{1/2}u_T)\otimes(H^{1/2}u_\tau)^\top\right. \\ &\quad\left. -\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}_{\ddot{X}_tX_s}(z+H^{1/2}u_1,\ldots,z+H^{1/2}u_T)\otimes(H^{1/2}u_s)^\top\right)\prod_{\ell=1}^{T}K^2\,(u_\ell)\,du_\ell\left(1+o_p(1)\right) \\ &= O_p(|H|^{-T/2}). \end{aligned} \tag{A.34}$$

Finally, the lower-right block is

$$\begin{aligned} &\frac{\sigma_v^2}{NT|H|^T}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}\otimes(Z_{it}-z) - T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^\top\left(X_{it}\otimes(Z_{it}-z) - T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)^\top\lambda_i^2 \\ &= \left(1-\frac{1}{T}\right)\frac{\sigma_v^2\mu_2(K_{u_\tau}^2)\prod_{\ell\neq\tau}^{T}R(K_{u_\ell})}{|H|^{T/2}}\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes H + O_p(|H|^{-T/2}H). \end{aligned} \tag{A.35}$$

Then, substituting (A.12), (A.33)–(A.35) into (A.32) we get the following conditional covariance matrix result,

$$\operatorname{Var}\left(\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right)=\frac{\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)}{NT|H|^{T/2}}\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}\left(z,\ldots,z\right)\left(1+o_p(1)\right). \quad \blacksquare$$

Proof of Theorem 3.2. With the aim of obtaining the asymptotic distribution of the local weighted linear least-squares estimator $\widehat{m}\left(z;H\right)$ we follow a similar proof scheme as in [13]. For this, let us denote

$$\widehat{m}\left(z;H\right)-m(z)=\left(\widehat{m}\left(z;H\right)-E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right]\right)+\left(E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right]-m(z)\right)\equiv\mathbf{I}_1+\mathbf{I}_2,$$

so in order to obtain the asymptotic distribution of this estimator we must show that as $N\to\infty$ it holds

$$\sqrt{NT|H|^{T/2}}\mathbf{I}_1\to N\left(0,\sigma_v^2\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)\mathcal{B}^{-1}_{\ddot{X}_t\ddot{X}_t}\left(z,\ldots,z\right)\right) \tag{A.36}$$

and

$$E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right]-m(z)=\frac{1}{2}\mu_2\left(K_{u_\tau}\right)\operatorname{diag}_d\left(\operatorname{tr}(\mathcal{H}_{m_r}(z)H)\right)\iota_d+O_p(H^{3/2})+o_p(\operatorname{tr}(H)). \tag{A.37}$$

By Assumption 3.1 we state the variables are i.i.d. in the subscript i but not in T, so the Lindeberg condition cannot be verified directly. Thus, in order to show (A.36) it suffices to check the Lyapunov condition. We have shown that

$$\widehat{m}\left(z;H\right)-E\left[\widehat{m}\left(z;H\right)|\mathbb{X},\mathbb{Z}\right]=e_1^{\top}\left(\widetilde{Z}^{\top}W\widetilde{Z}\right)^{-1}\widetilde{Z}^{\top}Wv. \tag{A.38}$$

The behavior of the inverse term has been analyzed previously, with the aim of proving the result (A.38) we must focus on the asymptotic normality of

$$\frac{1}{\sqrt{NT}}\widetilde{Z}^{\top}Wv. \tag{A.39}$$

As (A.39) is a multivariate vector, with the sake of simplicity we can define a unit vector $d\in\mathbb{R}^{d(1+q)}$ in such a way that

$$\frac{1}{\sqrt{NT}}d^T\widetilde{Z}^{\top}Wv=\frac{1}{\sqrt{NT}}\sum_i\sum_t\phi_{it}, \tag{A.40}$$

where

$$\phi_{it}=|H|^{T/4}d^{\top}\widetilde{Z}_{it}W_{it}v_{it},\quad i=1,\ldots,N;\ t=1,\ldots,T.$$

Following Assumption 3.8, we have that $R\left(K\right)=\int K^2\left(u\right)du=\left(2\pi^{1/2}\right)^{-1}$ and $R\left(K_{u_1}\right)=\cdots=R\left(K_{u_T}\right)$, so $\prod_{\ell=1}^{T}R\left(K_{u_\ell}\right)=R\left(K\right)^T$ holds. Combining these conditions with the results of Theorem 3.1 we can write

$$\operatorname{Var}\left(\phi_{it}\right)=\sigma_v^2d^{\top}\begin{pmatrix}R\left(K\right)^T\mathcal{B}_{\ddot{X}_t\ddot{X}_t}\left(z,\ldots,z\right) & 0\\ 0 & \left(1-\frac{1}{T}\right)\sigma_v^2\mu_2(K_{u_\tau}^2)\prod_{\ell\neq\tau}^{T}R(K_{u_\ell})\mathcal{B}_{X_tX_t}(z,\ldots,z)\otimes H\end{pmatrix}$$
$$\times\, d\left(1+o_p(1)\right), \tag{A.41}$$

whereas

$$\sum_{t=1}^{T}|\operatorname{Cov}\left(\phi_{i1},\phi_{it}\right)|=o_p(1). \tag{A.42}$$

In order to check the Lyapunov condition let us denote $\phi_{n,i}^*=T^{-1/2}\sum_{t=1}^{T}\phi_{it}$ as independent random variables for T fixed and $n=NT$. Then, by the Minkowski inequality and the matrix structure of $\widetilde{Z}_{it}$ we get

$$E\left|\phi_{n,i}^*\right|^{2+\delta}\le CT^{\frac{(2+\delta)}{2}}E\left|\phi_{it}\right|^{2+\delta}=CT^{\frac{(2+\delta)}{2}}E\left|\phi_{1it}+\phi_{2it}\right|^{2+\delta}.$$

Analyzing each term separately we obtain

$$\begin{aligned}E\left|\phi_{1it}\right|^{2+\delta}&\le E\left|\,|H|^{-T/4}d^{\top}\ddot{X}_{it}v_{it}\lambda_i\right|^{2+\delta}=|H|^{-T(2+\delta)/4}E\left[E\left(|d^{\top}\ddot{X}_{it}v_{it}|^{2+\delta}|\mathbb{X},\mathbb{Z}\right)\lambda_i^{2+\delta}\right]\\&=\frac{1}{|H|^{T\delta/4}}\int E\left(|d^{\top}\ddot{X}_{it}v_{it}\,|^{2+\delta}\,|Z_{i1}=z+H^{1/2}u_1,\ldots,Z_{iT}=z+H^{1/2}u_T\right)\end{aligned}$$

$$\times f_{Z_{i1},\ldots,Z_{iT}}(z+H^{1/2}u_1,\ldots,z+H^{1/2}u_T)\prod_{\ell=1}^{T}K^{2+\delta}(u_\ell)\,du_\ell$$

$$=|H|^{-T\delta/4}E\left(|d^\top\ddot{X}v_{it}|^{2+\delta}|Z_{i1}=z,\ldots,Z_{iT}=z\right)f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z)\prod_{\ell=1}^{T}\int K^{2+\delta}(u_\ell)\,du_\ell+o_p(|H|^{-T\delta/4}).$$

$$\begin{aligned}E|\lambda_{2it}|^{2+\delta}&\le E\left||H|^{-T/4}d^\top\left(X_{it}\otimes(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}\otimes(Z_{is}-z)\right)v_{it}\lambda_i\right|^{2+\delta}\\
&\le|H|^{-T(2+\delta)/4}E\left[E\left(|d^\top\ddot{X}_{it}v_{it}|Z_{i1},\ldots,Z_{iT}\right)\otimes|Z_{it}-z|^{2+\delta}\lambda_i^{2+\delta}\right]\\
&\quad+|H|^{-T(2+\delta)/4}\frac{1}{T}\sum_{s=1}^{T}E\left[E\left(|d^\top X_{is}v_{it}|Z_{i1},\ldots,Z_{iT}\right)\otimes|Z_{is}-z|^{2+\delta}\lambda_i^{2+\delta}\right]\\
&=|H|^{-T\delta/4}E\left(|d^\top\ddot{X}_{it}v_{it}|^{2+\delta}|Z_{i1}=z,\ldots,Z_{iT}=z\right)f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z)\otimes\int|H^{1/2}u|^{2+\delta}\prod_{\ell=1}^{T}K^{2+\delta}(u_\ell)\,du_\ell\\
&\quad+|H|^{-T\delta/4}\frac{1}{T}\sum_{s=1}^{T}E\left(|d^\top\ddot{X}_{is}v_{it}|^{2+\delta}|Z_{i1}=z,\ldots,Z_{iT}=z\right)f_{Z_{i1},\ldots,Z_{iT}}(z,\ldots,z)\\
&\quad\otimes\int|H^{1/2}u_s|^{2+\delta}\prod_{\ell=1}^{T}K^{2+\delta}(u_\ell)\,du_\ell+o_p(|H|^{1-(T-2)\delta/4}).\end{aligned}$$

In this way, we can write

$$(NT)^{-\frac{2+\delta}{2}}\sum_{i=1}^{N}E\left|\phi_{n,i}^*\right|^{2+\delta}\le C(N|H|^{T/2})^{-\delta/2},\tag{A.43}$$

and given that when $N|H|\to\infty$ this term tends to zero it is proved that the Lyapunov condition holds. Then, using (A.12), (A.33)–(A.35) and the Cramer–Wold device, the proof of the result (A.36) is done.

On the other hand, focus on the proof of (A.37) we know that by the law of iterated expectations

$$E\left[\widehat{m}(z;H)\right]=\int E\left[\widehat{m}(z;H)\,|\mathbb{X},\mathbb{Z}\right]dF(\mathbb{X},\mathbb{Z}).$$

Then, we can turn to the bias expression of the estimator collected in Theorem 3.1 and the proof is closed. ■

Proof of Theorem 3.3. The proof of this theorem follows the pattern set by the Theorem 3.1. The estimator to analyze is

$$\widetilde{m}(z;\widetilde{H})=e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\ddot{Y}^b,\tag{A.44}$$

we can write

$$E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]=e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\left[M^{(1)}+E\left(M^{(2)}|\mathbb{X},\mathbb{Z}\right)\right],\tag{A.45}$$

where

$$M^{(1)}=\left[\left(X_{11}^\top m(Z_{11})\right)^\top,\ldots,\left(X_{NT}^\top m(Z_{NT})\right)^\top\right]^\top,$$

$$M^{(2)}=\left[\left(T^{-1}\sum_{s=1}^{T}X_{1s}^\top(\widehat{m}(Z_{1s};H)-m(Z_{1s}))\right)^\top,\ldots,\left(T^{-1}\sum_{s=1}^{T}X_{Ns}^\top(\widehat{m}(Z_{Ns};H)-m(Z_{Ns}))\right)^\top\right]^\top\otimes\imath_T^\top.$$

The Taylor theorem implies that we can approximate $M^{(1)}$ as

$$M^{(1)}=\widetilde{Z}^b\begin{bmatrix}m(z)\\\mathrm{vec}(D_m(z))\end{bmatrix}+\frac{1}{2}Q_m^b(z)+R^b(z).\tag{A.46}$$

Following a similar nomenclature as in Theorem 3.1,

$$Q_m^b(z)=\left[S_{m_{11}}^{b\top},\ldots,S_{m_{NT}}^{b\top}\right]^\top,$$
$$R^b(z)=\left[R_{m_{11}}^b(z),\ldots,R_{m_{NT}}^b(z)\right]^\top,$$

where $R^b(z)$ is the remainder term of this approximation. Then, the corresponding entries of these vectors are

$$S^b_{m_{it}} = \left[(X_{it} \otimes (Z_{it} - z))^\top \mathcal{H}_m(z)\,(Z_{it} - z)\right]$$
$$R^b_{it}(z) = \left[(X_{it} \otimes (Z_{it} - z))^\top \mathcal{R}\,(Z_{it}; z)\,(Z_{it} - z)\right],$$

where $\mathcal{R}\,(Z_{it}; z)$ has already been defined in (A.5).

If we replace (A.46) in (A.45) the bias expression is then

$$E[\widetilde{m}(z; \widetilde{H})|\mathbb{X}, \mathbb{Z}] - m(z) = \frac{1}{2} e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b Q^b_m(z) + e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b E\left(M^{(2)}|\mathbb{X}, \mathbb{Z}\right) + o_p(\mathrm{tr}(\widetilde{H})), \tag{A.47}$$

given that following [15] and Assumption 3.1,

$$e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}\right)^{-1} \widetilde{Z}^{b\top} W^b R^b(z) = O_p(\mathrm{tr}(\widetilde{H})).$$

As you can see in (A.47), this bias expression is formed by two additive terms. The first one refers to the approximation error of the estimates, whereas the second one reflects the potential estimation error dragged from the first stage. Within this context, our aim is to show that this second term converges in probability to zero, so it is the first element which provides the asymptotic distribution of the backfitting estimator. For the sake of simplicity let us denote

$$\widetilde{\lambda}_{it} = K\left(\widetilde{H}^{-1/2}\,(Z_{it} - z)\right).$$

Focus first on the behavior of the inverse term of (A.47) we analyze

$$(NT)^{-1} \widetilde{Z}^{b\top} W^b \widetilde{Z}^b = \begin{pmatrix} \left(NT|\widetilde{H}|^{1/2}\right)^{-1} \sum_{it} X_{it} X_{it}^T \widetilde{\lambda}_{it} & \left(NT|\widetilde{H}|^{1/2}\right)^{-1} \sum_{it} X_{it}\,(X_{it} \otimes (Z_{it} - z))^\top \widetilde{\lambda}_{it} \\ \left(NT|\widetilde{H}|^{1/2}\right)^{-1} \sum_{it} (X_{it} \otimes (Z_{it} - z))\, X_{it}^\top \widetilde{\lambda}_{it} & \left(NT|\widetilde{H}|^{1/2}\right)^{-1} \sum_{it} (X_{it} \otimes (Z_{it} - z))\,(X_{it} \otimes (Z_{it} - z))^\top \widetilde{\lambda}_{it} \end{pmatrix}$$

and as it is proved in [13], using standard properties of kernel density estimators, conditions 3.1–3.3 and 3.4–3.10, as $N \to \infty$ we get

$$NT \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} = \begin{pmatrix} \mathcal{B}^{-1}_{X_tX_t}(z) + o_p(1) & -\mathcal{B}^{-1}_{X_tX_t}(z)\left[\mathcal{D}\mathcal{B}_{X_tX_t}(z)\right]\left(\mathcal{B}^{-1}_{X_tX_t}(z) \otimes I_q\right) + o_p(1) \\ -\left(\mathcal{B}^{-1}_{X_tX_t}(z) \otimes I_q\right)^\top \left[\mathcal{D}\mathcal{B}_{X_tX_t}(z)\right]^\top \mathcal{B}^{-1}_{X_tX_t}(z) + o_p(1) & \left(\mathcal{B}_{X_tX_t}(z) \otimes \mu_2(K)\widetilde{H}\right)^{-1} + o_p(\widetilde{H}^{-1}) \end{pmatrix}, \tag{A.48}$$

where $\mathcal{B}_{X_tX_t}(z)$ and $\mathcal{D}\mathcal{B}_{X_tX_t}(z)$ has been already defined in the proof of Theorem 3.1 conditioning only to $Z_{it} = z$.

Furthermore,

$$(NT)^{-1} \widetilde{Z}^{b\top} W^b Q^b_m(z) = \begin{pmatrix} \left(NT|\widetilde{H}|^{1/2}\right)^{-1} \sum_{it} X_{it}\,(X_{it} \otimes (Z_{it} - z))^\top \mathcal{H}_m(z)\,(Z_{it} - z)\,\widetilde{\lambda}_{it} \\ \left(NT|\widetilde{H}|^{1/2}\right)^{-1} \sum_{it} (X_{it} \otimes (Z_{it} - z))\,(X_{it} \otimes (Z_{it} - z))^\top \mathcal{H}_m(z)\,(Z_{it} - z)\,\widetilde{\lambda}_{it} \end{pmatrix} \tag{A.49}$$

are of order

$$\mu_2\,(K_u)\,\mathcal{B}_{X_tX_t}(z) \times \mathrm{diag}_d(\mathrm{tr}(\mathcal{H}_{m_r}(z)\widetilde{H}))\iota_d + o_p(\mathrm{tr}(\widetilde{H}))$$

and $O_p(\widetilde{H}^{3/2})$, respectively. Substituting these latter results and (A.48) in the first term of (A.47) we obtain

$$\frac{1}{2} e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b Q^b_m(z) = \frac{1}{2}\mu_2\,(K)\,\mathcal{B}^{-1}_{X_tX_t}(z)\,\mathcal{B}_{X_tX_t}(z) \times \mathrm{diag}_d\left(\mathrm{tr}(\mathcal{H}_{m_r}(z)\widetilde{H})\right)\iota_d + o_p(\mathrm{tr}(\widetilde{H})). \tag{A.50}$$

Focus now on the behavior of the second term of (A.47),

$$\left(NT|\widetilde{H}|^{1/2}\right)^{-1} \widetilde{Z}^{b\top} W^b E\left(M^{(2)}|\mathbb{X}, \mathbb{Z}\right) = \begin{pmatrix} \left(NT^2|\widetilde{H}|^{1/2}\right)^{-1} \sum_{its} X_{it} X_{is}^\top \left(E\left[\widehat{m}\,(Z_{is}; H)|\,\mathbb{X}, \mathbb{Z}\right] - m\,(Z_{is})\right)\widetilde{\lambda}_{it} \\ \left(NT^2|\widetilde{H}|^{1/2}\right)^{-1} \sum_{its} (X_{it} \otimes (Z_{it} - z))\, X_{is}^\top \left(E\left[\widehat{m}\,(Z_{is}; H)|\,\mathbb{X}, \mathbb{Z}\right] - m\,(Z_{is})\right)\widetilde{\lambda}_{it} \end{pmatrix} \tag{A.51}$$

and analyzing both terms separately we can show that as N tends to infinity

$$\left(NT^2|\widetilde{H}|^{1/2}\right)^{-1}\sum_{its}X_{it}X_{is}^\top\left(E\left[\widehat{m}\left(Z_{is};H\right)|\,\mathbb{X},\mathbb{Z}\right]-m\left(Z_{is}\right)\right)\widetilde{\lambda}_{it}=o_p(\mathrm{tr}(\widetilde{H}))$$

and

$$\left(NT^2|\widetilde{H}|^{1/2}\right)^{-1}\sum_{its}\left(X_{it}\otimes\left(Z_{it}-z\right)\right)X_{is}^\top\left(E\left(\widehat{m}\left(Z_{is};H\right)|\,\mathbb{X},\mathbb{Z}\right)-m\left(Z_{is}\right)\right)\widetilde{\lambda}_{it}=o_p(\mathrm{tr}(H)\mathrm{tr}(\widetilde{H})).$$

Under Assumptions 3.1–3.3, 3.10 and 3.12, this latter expression is $o_p\left(\mathrm{tr}(H)\right)$ and the rate is uniform in z; see [11] for more details.

Replacing these results in the second term of (A.47),

$$e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bM^{(2)}=o_p(\mathrm{tr}(\widetilde{H})).\tag{A.52}$$

Finally, substituting (A.50) and (A.52) in (A.47) the proof of the conditional bias is done. Also, it is proved that the asymptotic bias of $\widetilde{m}(z;\widetilde{H})$ is the same order as $\widehat{m}(z;H)$, given that $\mathrm{tr}(H)\to 0$, $\mathrm{tr}(\widetilde{H})\to 0$ in such a way that $N|H|\to\infty$ and $N|\widetilde{H}|\to\infty$.

From the standpoint of the variance, let us denote $\widehat{\upsilon}=\left(\widehat{\upsilon}_1,\ldots,\widehat{\upsilon}_N\right)^\top$ as a NT-dimensional vector such that

$$\widehat{\upsilon}_i=\left(T^{-1}\sum_{s=1}^T\left(X_{is}^\top\left(\widehat{m}\left(Z_{is};H\right)-E\left[\widehat{m}\left(Z_{is};H\right)|\,\mathbb{X},\mathbb{Z}\right]\right)\right)^\top,\ldots,T^{-1}\sum_{s=1}^T\left(X_{is}^\top\left(\widehat{m}\left(Z_{is};H\right)-E\left[\widehat{m}\left(Z_{is};H\right)|\,\mathbb{X},\mathbb{Z}\right]\right)\right)^\top\right)^\top.$$

As we know, the conditional variance–covariance matrix of the estimator has the following form

$$\mathrm{Var}(\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z})=E\left[\left(\widetilde{m}(z;\widetilde{H}|\mathbb{X},\mathbb{Z})-E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]\right)\left(\widetilde{m}(z;\widetilde{H})-E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]\right)^\top\middle|\,\mathbb{X},\mathbb{Z}\right]$$

where

$$\widetilde{m}(z;\widetilde{H})-E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]=e_1^T\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\ddot{\upsilon}+e_1^T\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\widehat{\upsilon}.$$

Remember that $\ddot{\upsilon}_i=Q_T\upsilon_i$ and it is straightforward to show that $Q_T\widetilde{Z}_i^b=\widetilde{Z}_i$. Thus, the previous equation can be rewritten as

$$\widetilde{m}(z;\widetilde{H})-E[\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z}]=e_1^T\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{\top}W^b\upsilon+e_1^T\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\widehat{\upsilon}.$$

Taking into account that let $E\left(\upsilon\upsilon^\top|\mathbb{X},\mathbb{Z}\right)=\mathcal{V}$ be a $NT\times NT$ matrix whose ijth have the form of (A.29), the variance term of $\widetilde{m}(z;\widetilde{H})$ has the form

$$\begin{aligned}\mathrm{Var}(\widetilde{m}(z;\widetilde{H})|\mathbb{X},\mathbb{Z})&=e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^\top W^b\mathcal{V}W^b\widetilde{Z}\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1\\&\quad+e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bE\left(\widehat{\upsilon}\widehat{\upsilon}^\top|\mathbb{X},\mathbb{Z}\right)W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1\\&\quad+2e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bE\left(\widehat{\upsilon}\upsilon^\top|\mathbb{X},\mathbb{Z}\right)W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1\\&=\mathbf{I}_1+\mathbf{I}_2+\mathbf{I}_3.\end{aligned}\tag{A.53}$$

Then, with the aim of obtaining the asymptotic order of the variance of $\widetilde{m}(z;\widetilde{H})$ we have to analyze each of these terms separately. Following the same procedure as in (A.32) to analyze the behavior of $\widetilde{Z}^\top W^b\mathcal{V}W^b\widetilde{Z}$. Under Assumptions 3.1–3.7 and 3.11–3.12, using the result (A.48) and the Cramer–Wold device it is straightforward to show that as $N\to\infty$

$$\mathbf{I}_1=\frac{\sigma_\upsilon^2R(K)}{NT|\widetilde{H}|^{1/2}}\mathcal{B}_{X_tX_t}(z)^{-1}\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z)\,\mathcal{B}_{X_tX_t}(z)^{-1}(1+o_p(1)),\tag{A.54}$$

while

$$\mathbf{I}_2=o_p\left(\frac{\ln NT}{NT\,|H|^{T/2}\,|\widetilde{H}|^{1/2}}\right).\tag{A.55}$$

In order to prove this latter result we have to analyze the behavior of the following expression

$$(NT)^{-1}\widetilde{Z}^{b\top}W^bE\left(\widehat{\upsilon}\widehat{\upsilon}^\top|\mathbb{X},\mathbb{Z}\right)W^b\widetilde{Z}^b.\tag{A.56}$$

Thus, denote by $\widehat{r}(Z_{is};H)=\widehat{m}(Z_{is};H)-E\left[\widehat{m}(Z_{is};H)|\mathbb{X},\mathbb{Z}\right]$, then the upper left entry is

$$(NT^3|\widetilde{H}|)^{-1}\sum_i\sum_{tt'}\sum_{ss'}X_{it}X_{is}^\top E\left(\widehat{r}(Z_{is};H)\widehat{r}(Z_{is'};H)^\top|\mathbb{X},\mathbb{Z}\right)X_{is'}X_{it'}^\top\widetilde{\lambda}_{it}\widetilde{\lambda}_{it'}\tag{A.57}$$

and by the Cauchy–Schwarz inequality for variance–covariance matrices (A.57) is bounded by

$$\begin{aligned}&(NT^3|\widetilde{H}|)^{-1}\sum_i\sum_{tt'}\sum_{ss'}X_{it}X_{is}^\top \mathrm{vec}^{1/2}\left(\mathrm{diag}\left(E\left(\widehat{r}(Z_{is};H)\widehat{r}(Z_{is'};H)^\top|\mathbb{X},\mathbb{Z}\right)\right)\right)\\&\quad\times\mathrm{vec}^{1/2}\left(\mathrm{diag}\left(E\left(\widehat{r}(Z_{is};H)\widehat{r}(Z_{is};H)^\top|\mathbb{X},\mathbb{Z}\right)\right)\right)\widetilde{\lambda}_{it}\widetilde{\lambda}_{it'}\\&\quad=O_p\left(\frac{\ln NT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}\right),\end{aligned}\tag{A.58}$$

given that under the conditions of Theorem 3.1 and following [11],

$$\mathrm{vec}\left(\mathrm{diag}\left(E\left(\widehat{r}(z;H)\widehat{r}(z;H)^\top|\mathbb{X},\mathbb{Z}\right)\right)\right)=O_p\left(\frac{\ln NT}{NT|H|^{T/2}}\right),$$

uniformly in z.

Following the same lines, the upper right entry of (A.56) is

$$(NT^2|\widetilde{H}|)^{-1}\sum_i\sum_{tt'}\sum_{ss'}X_{it}X_{is}^\top E\left(\widehat{r}(Z_{is};H)\widehat{r}(Z_{is'};H)^\top|\mathbb{X},\mathbb{Z}\right)(X_{it'}\otimes(Z_{it'}-z))^\top\widetilde{\lambda}_{it}\widetilde{\lambda}_{it'}=o_p\left(\frac{\ln NT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}\right)\tag{A.59}$$

and the lower right entry of (A.56) is

$$\begin{aligned}&(NT^2|\widetilde{H}|)^{-1}\sum_i\sum_{tt'}\sum_{ss'}(X_{it}\otimes(Z_{it}-z))X_{is}^\top E\left(\widehat{r}(Z_{is};H)\widehat{r}(Z_{is'};H)^\top|\mathbb{X},\mathbb{Z}\right)(X_{it'}\otimes(Z_{it'}-z))^\top\widetilde{\lambda}_{it}\widetilde{\lambda}_{it'}\\&\quad=o_p\left(\frac{\ln NT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}\right).\end{aligned}\tag{A.60}$$

Then, combining the results (A.58)–(A.60) with (A.48) and by the Cramer–Wold device the proof of (A.55) is done. Finally, focus on $\mathbf{I}_3$ the Cauchy–Schwarz inequality is enough to show that

$$\mathbf{I}_3=o_p\left(\sqrt{\frac{\ln NT}{NT|H|^{T/2}|\widetilde{H}|^{1/2}}}\right)\tag{A.61}$$

and the proof is done. ■

Proof of Corollary 3.2. The proof of this corollary relies on the proof of Theorem 3.1.

Taking the expression (3.3) for the univariate case, the conditional bias and variance of $\widehat{m}_h(z)$ for the case when $d=q=1$ and $H=h^2I$ are given as follows

$$E[\widehat{m}_h(z)|\mathbb{X},\mathbb{Z}]-m(z)=\frac{1}{2}e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top W\left(\Pi(z)-\overline{\Pi}(z)\right),\tag{A.62}$$

$$\mathrm{Var}(\widehat{m}_h(z)|\mathbb{X},\mathbb{Z})=e_1^\top\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}\widetilde{Z}^\top WVW\widetilde{Z}\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1}e_1,\tag{A.63}$$

where, for any ξ_{it} between Z_{it} and z and ξ_{is} between Z_{is} and z, the corresponding entries of the vectors $\Pi(z)$ and $\overline{\Pi}(z)$ are

$$\Pi_{it}(z)=X_{it}r(\xi_{it};z)\quad\text{and}\quad\overline{\Pi}_{it}(z)=T^{-1}\sum_{s=1}^T X_{is}r(\xi_{is};z),$$

where $r(\xi_{it};z)=(Z_{it}-z)^2\frac{\partial^2 m(\xi_{it})}{\partial z^2}$ and $r(\xi_{is};z)$ is defined in a similar way.

Starting from the conditional bias standpoint, as N tends to infinity the elements of the matrix (A.8) are

$$\mathcal{A}_{NT}^{11}=(Th^T)^{-1}\sum_{t=1}^T E\left[\ddot{X}_{it}^2\lambda_i\right]=\frac{1}{T}\sum_{t=1}^T\mathcal{B}_{\ddot{X}_t\ddot{X}_t}(z,\ldots,z)+o_p(1),\tag{A.64}$$

$$\begin{aligned}\mathcal{A}_{NT}^{12}&=(Th^T)^{-1}\sum_{t=1}^T E\left[\ddot{X}_{it}\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^T X_{is}(Z_{is}-z)\right)\lambda_i\right]\\&=\frac{h^2}{T}\sum_{t=1}^T\left(\mathcal{D}\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)\mu_2(K_{u_\tau})-\frac{1}{T}\sum_{s=1}^T\mathcal{D}\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\mu_2(K_{u_s})\right)+o_p(h^2)\end{aligned}\tag{A.65}$$

$$\begin{aligned}\mathcal{A}_{NT}^{22} &= (Th^T)^{-1}\sum_{t=1}^{T} E\left[\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)^2\lambda_i\right]\\ &= \left(1-\frac{1}{T}\right)\frac{1}{T}\sum_{t=1}^{T}\mathcal{B}_{X_tX_t}(z,\ldots,z)\mu_2(K_{u_\tau})\end{aligned} \tag{A.66}$$

so the inverse term of (A.62) can be written as

$$(NT)\left(\widetilde{Z}^\top W\widetilde{Z}\right)^{-1} = \begin{pmatrix}\mathcal{C}_{NT}^{11} & \mathcal{C}_{NT}^{12}\\ \mathcal{C}_{NT}^{21} & \mathcal{C}_{NT}^{22}\end{pmatrix}, \tag{A.67}$$

where now

$$\begin{aligned}\mathcal{C}_{NT}^{11} &= \left((Th^T)^{-1}\sum_{t=1}^{T}E\left[\ddot{X}_{it}^2\lambda_i\right]\right)^{-1} + o_p\left(h^T\right),\\ \mathcal{C}_{NT}^{12} &= -\left(T^{-1}\sum_{t=1}^{T}E\left[\ddot{X}_{it}^2\lambda_i\right]\right)^{-1}T^{-1}\sum_{t=1}^{T}E\left[\ddot{X}_{it}\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)\lambda_i\right]\\ &\quad\times\left((Th^T)^{-1}\sum_{t=1}^{T}E\left[\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)^2\lambda_i\right]\right)^{-1} + o_p\left(h^T\right),\\ \mathcal{C}_{NT}^{22} &= \left((Th^T)^{-1}\sum_{t=1}^{T}E\left[\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)^2\lambda_i\right]\right)^{-1} + o_p\left(h^T\right).\end{aligned}$$

Focus now on the numerator of (A.62), as $N\to\infty$ it can be written such as

$$\begin{aligned}&(NT)^{-1}\widetilde{Z}^\top W\left(\Pi(z)-\overline{\Pi}(z)\right)\\ &= \begin{pmatrix}(Th^T)^{-1}\sum_{t=1}^{T}E\left[\ddot{X}_{it}\left(X_{it}r_h(\xi_{it};z)-T^{-1}\sum_{s=1}^{T}X_{is}r_h(\xi_{is};z)\right)\lambda_i\right]\\ (Th^T)^{-1}\sum_{t=1}^{T}E\left[\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)\left(X_{it}r_h(\xi_{it};z)-T^{-1}\sum_{s=1}^{T}X_{is}r_h(\xi_{is};z)\right)\lambda_i\right]\end{pmatrix},\end{aligned} \tag{A.68}$$

where we can show

$$\begin{aligned}&\frac{1}{Th^T}\sum_{t=1}^{T}E\left[\ddot{X}_{it}X_{it}r_h(\xi_{it};z)\lambda_i\right]-\frac{1}{T^2h^T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[\ddot{X}_{it}X_{is}r_h(\xi_{is};z)\lambda_i\right]\\ &= \frac{h^2}{T}\sum_{t=1}^{T}\left(\mathcal{B}_{\ddot{X}_tX_t}(z,\ldots,z)\mu_2(K_{u_t})-\frac{1}{T}\sum_{s=1}^{T}\mathcal{B}_{\ddot{X}_tX_s}(z,\ldots,z)\mu_2(K_{u_s})\right)+o_p(h^2)\end{aligned} \tag{A.69}$$

and

$$\begin{aligned}&\frac{1}{Th^T}\sum_{t=1}^{T}E\left[\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)X_{it}r_h(\xi_{it};z)\lambda_i\right]\\ &\quad-\frac{1}{T^2h^T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[\left(X_{it}(Z_{it}-z)-T^{-1}\sum_{s=1}^{T}X_{is}(Z_{is}-z)\right)X_{is}r_h(\xi_{is};z)\lambda_i\right]\\ &= \frac{h^3}{T}\sum_{t=1}^{T}\left(\mathcal{B}_{X_tX_t}(z,\ldots,z)u_t^3-\frac{2}{T}\sum_{s=1}^{T}\mathcal{B}_{X_tX_s}(z,\ldots,z)u_su_t^2+\frac{1}{T^2}\sum_{s=1}^{T}\sum_{s'=1}^{T}\mathcal{B}_{X_sX_{s'}}(z,\ldots,z)u_{s'}u_s^2\right)\prod_{\ell=1}^{T}K_{u_\ell}du_\ell\\ &= O_p(h^3).\end{aligned} \tag{A.70}$$

Therefore, substituting (A.67) and (A.68) in (A.62), the conditional bias is

$$\begin{aligned}E[\widehat{m}_h(z)|\mathbb{X},\mathbb{Z}]-m(z) &= \left(\frac{1}{Th^T}\sum_{t=1}^{T}E[\ddot{X}_{it}^2\lambda_i]\right)^{-1}\frac{1}{2Th^T}\\ &\quad\times\sum_{t=1}^{T}E\left[\ddot{X}_{it}\left(X_{it}r_h(\xi_{it};z)-\frac{1}{T}\sum_{s=1}^{T}X_{is}r_h(\xi_{is};z)\right)\lambda_i\right]+o_p(h^2).\end{aligned} \tag{A.71}$$

From the point of view of the variance we focus on the behavior of the middle term of (A.63). Following the same procedure as in (A.32) but assuming strict stationarity is not allowed, under conditions 3.2, 3.4 and 3.13 we can show that as $N \to \infty$,

$$\mathrm{Var}(\widehat{m}_h(z)|\mathbb{X}, \mathbb{Z}) = \frac{\sigma_v^2}{NTh^T} \frac{1}{Th^T} \sum_{t=1}^{T} E[\ddot{X}_{it}^2 \lambda_i^2] \left(\frac{1}{Th^T} \sum_{t=1}^{T} E[\ddot{X}_{it}^2 \lambda_i] \right)^{-2} + o_p\left(\frac{1}{NTh^T}\right). \quad \blacksquare \tag{A.72}$$

Proof of Corollary 3.3. The proof of this corollary relies on the proof of Theorem 3.2.

If we start from (A.45) and take a standard Taylor expansion around $m(\cdot)$ we obtain

$$E\left[\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X}, \mathbb{Z}\right] - m(z) = \frac{1}{2} e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b \Pi(z) + e_1^\top \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} \widetilde{Z}^{b\top} W^b E\left(M^{(2)}|\mathbb{X}, \mathbb{Z}\right) + o_p(\widetilde{h}^2) \tag{A.73}$$

where

$$M^{(2)} = \left[\left(T^{-1} \sum_{s=1}^{T} X_{1s} \left(\widehat{m}_h(Z_{1s}) - m(Z_{1s})\right) \right), \ldots, \left(T^{-1} \sum_{s=1}^{T} X_{Ns} \left(\widehat{m}_h(Z_{Ns}) - m(Z_{Ns})\right) \right) \right]^\top \otimes \iota_T^\top$$

and each entry of $\Pi(z)$ is $X_{it} r(\xi_{it}, z)$.

Following a similar proof scheme as previously, if we analyze each of these terms separately we obtain that (A.73) can be written as

$$(NT)^{-1} \widetilde{Z}^{b\top} W^b \widetilde{Z}^b = \begin{pmatrix} (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 \widetilde{\lambda}_{it}\right] & (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z) \widetilde{\lambda}_{it}\right] \\ (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z) \widetilde{\lambda}_{it}\right] & (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z)^2 \widetilde{\lambda}_{it}\right] \end{pmatrix}$$

so the inverse term is

$$(NT) \left(\widetilde{Z}^{b\top} W^b \widetilde{Z}^b\right)^{-1} = \begin{pmatrix} \mathcal{C}_{NT}^{b11} & \mathcal{C}_{NT}^{b12} \\ \mathcal{C}_{NT}^{b21} & \mathcal{C}_{NT}^{b22} \end{pmatrix}, \tag{A.74}$$

where

$$\mathcal{C}_{NT}^{b11} = \left((T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 \widetilde{\lambda}_{it}\right] \right)^{-1} + o_p(\widetilde{h}),$$

$$\mathcal{C}_{NT}^{b12} = \left((T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 \widetilde{\lambda}_{it}\right] \right)^{-1} (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z) \widetilde{\lambda}_{it}\right] \left((T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 \widetilde{\lambda}_{it}\right] \right)^{-1} + o_p(\widetilde{h}),$$

$$\mathcal{C}_{NT}^{b22} = \left((T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z)^2 \widetilde{\lambda}_{it}\right] \right)^{-1} + o_p(\widetilde{h}).$$

Let us now analyze the numerator of (A.73), as $N \to \infty$ we get

$$(NT)^{-1} \widetilde{Z}^{b\top} W^b \left(\Pi(z) + E\left(M^{(2)}|\mathbb{X}, \mathbb{Z}\right)\right)$$
$$= \begin{pmatrix} (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 r(\xi_{it}, z) \widetilde{\lambda}_{it}\right] + (T^2\widetilde{h})^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it} X_{is} \left(E(\widehat{m}_h(Z_{is}) - m(Z_{is}))\right) \widetilde{\lambda}_{it}\right] \\ (T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z) r(\xi_{it}, z) \widetilde{\lambda}_{it}\right] + (T^2\widetilde{h})^{-1} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left[X_{it} (Z_{it} - z) X_{is} \left(E(\widehat{m}_h(Z_{is}) - m(Z_{is}))\right) \widetilde{\lambda}_{it}\right] \end{pmatrix}. \tag{A.75}$$

Using standard properties of the kernel density estimators and assuming strict stationarity is not allowed, we can show

$$(T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 r(\xi_{it}, z) \widetilde{\lambda}_{it}\right] = \frac{\widetilde{h}^2}{T} \sum_{t=1}^{T} \mathcal{B}_{X_t X_t}(z) m''(z) \mu_2(K_{u_\tau}) + o_p(\widetilde{h}),$$

$$(T\widetilde{h})^{-1} \sum_{t=1}^{T} E\left[X_{it}^2 (Z_{it} - z) r(\xi_{it}, z) \widetilde{\lambda}_{it}\right] = O_p(\widetilde{h}^3)$$

whereas following what it is established in (A.51) we can prove

$$(T^2\widetilde{h})^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}X_{is}\left(E(\widehat{m}_h(Z_{is})-m(Z_{is}))\right)\widetilde{\lambda}_{it}\right]=o_p(\widetilde{h}^2),$$

$$(T^2\widetilde{h})^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}E\left[X_{it}(Z_{it}-z)X_{is}\left(E(\widehat{m}_h(Z_{is})-m(Z_{is}))\right)\widetilde{\lambda}_{it}\right]=o_p(\widetilde{h}^2h^2).$$

Therefore, replacing the results of (A.74) and (A.76) in (A.73) the conditional bias expression of the one-step backfitting estimator is

$$E\left[\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X},\mathbb{Z}\right]-m(z)=\frac{1}{2T\widetilde{h}}\sum_{t=1}^{T}E\left[X_{it}^2r(\xi_{it},z)\widetilde{\lambda}_{it}\right]\left(\frac{1}{T\widetilde{h}}\sum_{t=1}^{T}E\left[X_{it}^2\widetilde{\lambda}_{it}\right]\right)^{-1}+o_p(\widetilde{h}^2). \tag{A.76}$$

Finally, as in the multivariate case the variance term of the one-step backfitting estimator has the form

$$\begin{aligned}\operatorname{Var}\left(\widetilde{m}_{\widetilde{h}}(z)|\mathbb{X},\mathbb{Z}\right)&=e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^b\mathcal{V}W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1\\&\quad+e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bE(\widehat{v}\widehat{v}^\top|\mathbb{X},\mathbb{Z})W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1\\&\quad+e_1^\top\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}\widetilde{Z}^{b\top}W^bE(\widehat{v}v^\top|\mathbb{X},\mathbb{Z})W^b\widetilde{Z}^b\left(\widetilde{Z}^{b\top}W^b\widetilde{Z}^b\right)^{-1}e_1\\&=\mathbf{I}_1+\mathbf{I}_2+\mathbf{I}_3,\end{aligned} \tag{A.77}$$

where $\mathcal{V}$ is a $NT\times NT$ matrix of $E(vv^\top|\mathbb{X},\mathbb{Z})$ whose ijth have the form of (A.29), $\ddot{v}_i=Q_Tv_i$ and

$$\widehat{v}_i=\left(T^{-1}\sum_{s=1}^{T}\left(X_{is}\left(\widehat{m}_h(Z_{is})-E[\widehat{m}_h(Z_{is})|\mathbb{X},\mathbb{Z}]\right)\right),\ldots,T^{-1}\sum_{s=1}^{T}\left(X_{is}\left(\widehat{m}_h(Z_{is})-E[\widehat{m}_h(Z_{is})|\mathbb{X},\mathbb{Z}]\right)\right)\right)^\top.$$

Analyzing each of these terms separately, under conditions 3.2–3.9 and 3.11–3.13 and using the Cramer–Wold device and the result in (A.74) we can show that as $N\to\infty$

$$\mathbf{I}_1=\frac{\sigma_v^2}{NT\widetilde{h}}\left(\frac{1}{T\widetilde{h}}\sum_{t=1}^{T}E\left[\ddot{X}_{it}^2\widetilde{\lambda}_{it}^2\right]\right)\left(\frac{1}{T\widetilde{h}}\sum_{t=1}^{T}E\left[X_{it}^2\widetilde{\lambda}_{it}\right]\right)^{-2}, \tag{A.78}$$

$$\mathbf{I}_2=o_p\left(\frac{\ln NT}{NTh^T\widetilde{h}}\right) \tag{A.79}$$

whereas $\mathbf{I}_3=o_p\left(\sqrt{\frac{\ln NT}{NTh^T\widetilde{h}}}\right)$. Note that for the result (A.79) we follow the proof proposed for (A.55) and for $\mathbf{I}_3$ we follow (A.61). Then, the proof is done. ■

References

[1] Z. Cai, Q. Li, Nonparametric estimation of varying coefficient dynamic panel data models, Econometric Theory 24 (2008) 1321–1342.
[2] Z. Cai, Q. Li, J.Y. Park, Functional-coefficient models for nonstationary time series data, J. Econometrics 148 (2009) 101–113.
[3] J. Fan, I. Gijbels, Local Polynomial Modelling and its Applications, Chapman & Hall, 1995.
[4] J. Fan, W. Zhang, Statistical estimation in varying coefficient models, Ann. Statist. 27 (1999) 1491–1518.
[5] D.J. Henderson, R.J. Carroll, Q. Li, Nonparametric estimation and testing of fixed effects panel data model, J. Econometrics 144 (2008) 257–275.
[6] D.J. Henderson, A. Ullah, A nonparametric random effects estimator, Econom. Lett. 88 (2005) 403–407.
[7] C. Hsiao, Analysis of Panel Data, second ed., in: Econometric Society Monographs, 2003.
[8] Y. Lee, D. Mukherjee, New nonparametric estimation of the marginal effects in fixed-effects panel models: An application on the environmental Kuznets curve, 2008. Available at SSRN 1119315.
[9] E. Mammen, B. Støve, D. Tjøstheim, Nonparametric additive models for panels of time series, Econometric Theory 25 (2009) 442–481.
[10] C. Martins-Filho, F. Yao, Nonparametric regression estimation with general parametric error covariance, J. Multivariate Anal. 100 (2009) 309–333.
[11] E. Masry, Multivariate local polynomial regression for time series: uniform strong consistency rate, J. Time Ser. Anal. 17 (1996) 571–599.
[12] K. Mundra, Nonparametric slope estimation for fixed-effect panel data. Working Paper, Department of Economics, San Diego State University, 2005.
[13] J.M. Rodriguez-Poo, A. Soberon, Direct semi-parametric estimation of fixed effects panel data varying coefficient models, Econom. J. 17 (2014) 107–138.
[14] A.F. Ruckstuhl, A.H. Welsh, R.J. Carroll, Nonparametric function estimation of the relationship between two repeatedly measured variables, Statist. Sinica 10 (2000) 51–71.
[15] D. Ruppert, M.P. Wand, Multivariate locally weighted least squares regression, Ann. Statist. 22 (3) (1994) 1346–1370.
[16] L. Su, X. Lu, Nonparametric dynamic panel data models: Kernel estimation and specification testing, J. Econometrics 176 (2013) 112–133.
[17] L. Su, A. Ullah, Nonparametric and semiparametric panel econometric models: estimation and testing, in: A. Ullah, D.E. Giles (Eds.), Handbook of Empirical Economics and Finance, Taylor and Francis Group, New York, 2011, pp. 455–497.
[18] Y. Sun, R.J. Carroll, D. Li, Semiparametric estimation of fixed effects panel data varying coefficient models, Adv. Econom. 25 (2009) 101–130.
[19] J.M. Wooldridge, Econometric Analysis of Cross Section and Panel Data Models, The MIT Press, Cambridge, Massachusetts, 2003.
[20] J. You, X. Zhou, Y. Zhou, Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors, J. Multivariate Anal. 101 (2010) 1079–1101.
[21] L. Zhan-Qian, Multivariate locally weighted polynomial fitting and partial derivative estimation, J. Multivariate Anal. 59 (1996) 187–205.