

Econometrics Journal (2013), volume 10, pp. 1–33.
Article No. ectj??????

Direct Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models

JUAN M. RODRIGUEZ-POO[†] AND ALEXANDRA SOBERON[‡]

[†]*Departament of Economics, Universidad de Cantabria,
Avda. Los Castros s/n, E-39005, Santander, Spain*

E-mail: rodrigjm@unican.es, soberonap@unican.es

Received: July 2012

Summary In this paper we present a new technique to estimate varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in an unknown way. The estimator is based in first differences and then a local linear regression is applied to estimate the unknown coefficients. To avoid a non-negligible asymptotic bias, we need to introduce a higher dimensional kernel weight. This enables us to remove the bias at the price of enlarging the variance term and hence, achieving a slower rate of convergence. To overcome this problem we propose a one step backfitting algorithm that enables the resulting estimator to achieve optimal rates of convergence for this type of problems. It exhibits also the so called oracle efficiency property. We also obtain the asymptotic distribution. Since the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. Monte Carlo results indicates good performance of the estimator in finite samples.

Keywords: *Varying coefficients model, fixed effects, panel data, local linear regression, oracle efficient estimator.*

1. INTRODUCTION

This paper is concerned with the estimation of varying coefficient panel data models. This type of specification consists in a linear regression model where regression coefficients are assumed to be varying depending on some exogenous continuous variables proposed by economic theory. For example, in the so called problem of returns to education, when estimating elasticity of wages to changes in education, it has been pointed out (see Schultz (2003)) that marginal returns to education might vary with the level of working experience. Therefore, conditionally on a level of education, the elasticity is going to change according with the level of working experience.

Within this context, the issue of potential misspecification of the functional form of the varying coefficients has motivated in empirical studies the use of nonparametric estimation techniques. In most part of cases, the estimation of the functional form of the coefficients has been performed through standard techniques such as spline smoothers, series estimators or local polynomial regression estimators (see Su and Ullah (2011)). Although in most cases, a direct application of the previous techniques renders correct inference results, it is true that not much attention has been paid to the asymptotic behavior of these estimators under non-standard settings. Unfortunately, some of these settings are rather relevant in empirical analysis of panel data models. One clear

example is presence in the econometric model of some unobserved explanatory variables that, although not varying along time, can be statistically correlated with other explanatory variables in the model (fixed effects). The presence of heterogeneity of unknown form that is correlated with some explanatory variables is not an easy problem to deal with. In fact, under this type of heterogeneity any estimation technique suffers of the so called *incidental parameters problem* (Neyman and Scott (1948)).

In order to obtain a consistent estimator of the parameters of interest one possible solution is to transform the model in order to remove the heterogeneity of unknown form. To be more precise, consider a linear panel data model where the heterogeneity μ_i is arbitrarily correlated with the covariates X_{it} and/or Z_{it} :

$$Y_{it} = X_{it}^T m(Z_{it}) + \mu_i + v_{it}, \quad i = 1, \dots, N; t = 1, \dots, T. \quad (1.1)$$

Furthermore, the function $m(z)$ is unknown and needs to be estimated and the v_{it} are random errors. It is clear that, any attempt to estimate directly $m(\cdot)$ through standard nonparametric estimation techniques will rend inconsistent estimators of the underlying curve. The reason is that $E(\mu_i | X_{it} = x, Z_{it} = z) \neq 0$. A standard solution to this problem is to remove μ_i from (1.1) by taking a transformation and then, estimate the unknown curve through the use of a nonparametric smoother. There exists several approaches to remove these effects. The simplest one is probably to take first differences, i. e.

$$\Delta Y_{it} = X_{it}^T m(Z_{it}) - X_{i(t-1)}^T m(Z_{i(t-1)}) + \Delta v_{it}, \quad i = 2, \dots, N; t = 2, \dots, T. \quad (1.2)$$

Direct nonparametric estimation of $m(\cdot)$ has been till now considered as rather cumbersome (see Su and Ullah (2011)). The reason is that, for each i , the conditional expectation $E(\Delta Y_{it} | Z_{it}, Z_{i(t-1)}, X_{it}, X_{i(t-1)})$ in (1.2) contains a linear combination of $X_{it}^T m(Z_{it})$ for different t . This can be considered as an additive function with the same functional form at a different times.

In some special cases consistent estimation of the quantities of interest have been provided in the literature. For the unrestricted model $X_{it}^T m(Z_{it}) \equiv m(X_{it}, Z_{it})$, (1.2) becomes a fully nonparametric additive model

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

In this case, Henderson et al. (2008) propose an iterative procedure based in a profile likelihood approach whereas in Mammen et al. (2009) is considered the consistent estimation of nonparametric additive panel data models with both time and individual effects via a smoothed backfitting algorithm. Furthermore, for $X_{it}^T m(Z_{it}) \equiv g(Z_{it}) + \tilde{X}_{it}^T \beta_0$, where $X_{it} = (1, \tilde{X}_{it}^T)^T$ and $m(Z_{it}) = (g(Z_{it}), \beta_0)^T$ for some real valued $g(\cdot)$ and a vector β_0 , the regression function in (1.2) becomes a semiparametric partially additive model

$$\Delta Y_{it} = \beta_0^T \Delta \tilde{X}_{it} + g(Z_{it}) - g(Z_{i(t-1)}) + \Delta v_{it}, \quad i = 2, \dots, N; t = 2, \dots, T.$$

Qian and Wang (2012) consider the marginal integration estimator of the nonparametric additive component resulting from the first differencing step, i.e. $G(Z_{it}, Z_{i(t-1)}) = g(Z_{it}) - g(Z_{i(t-1)})$.

The estimation procedure that we introduce in our paper mainly generalizes the previous results to a rather general varying coefficient model as the one specified in (1.1) in a framework

where $N \rightarrow \infty$ but T remains fixed. It is based in applying a local approximation to the additive function $X_{it}^T m(Z_{it}) - X_{i(t-1)}^T m(Z_{i(t-1)})$. The same idea was proposed in a completely different context in Yang (2002). Since the estimator is based in local approximation properties, we investigate the behavior of the bias remainder term under fairly general conditions. This term, which is negligible in standard local linear regression techniques (see Fan and Gijbels (1995b)), requires much more attention when dealing with first difference estimators. In fact, as it has been already pointed out in Lee and Mukherjee (2008), direct application of local linear regression techniques to first differencing transformations in panel data models tends to biased estimators and the bias does not degenerate even with large samples. Using a higher dimensional kernel weight, our estimation technique overcomes the problem of non-vanishing bias although, as expected, the variance term becomes larger. The same phenomena appears also in Henderson et al. (2008) where their final estimator already shows even a larger variance.

In order to obtain the standard rates of convergence for this type of problems, that is, to reduce the variance holding the bias constant, we propose to use the developments introduced in Fan and Zhang (1999). Their core idea was that the variance can be reduced by further smoothing, but bias can not be reduced by any kind of smoothing. We apply these ideas to our problem by using a one step backfitting algorithm. Since it has the form of an additive model we also show that our estimator is oracle efficient, that is, the variance-covariance matrix of any of the components of our estimator is the same asymptotically as if we would know the other component. Finally, we also propose a data driven method to select the bandwidth parameter.

As it was already pointed out before, to remove the heterogeneous effects other transformations are available in the literature. To our knowledge, for model (1.1), estimation of $m(\cdot)$ has been proposed in Sun et al. (2009) by the use of the so called Least Squares Dummy Variable Approach. They estimate $m(\cdot)$ through the following alternative specification

$$Y_{it} = X_{it}^T m(Z_{it}) + \sum_{j=1}^N \mu_j d_{ij} + v_{it}, \quad i = 1, \dots, N; t = 1, \dots, T, \quad (1.3)$$

where $d_{ij} = 1$ if $i = j$ and 0 otherwise. Based in this model they propose a least-squares method combined with a local linear regression approach that produces a consistent estimator of the unknown smoothing coefficient curves. Compared to our method, their estimator exhibits a larger bias. In fact, their bias presents two terms. One results from the local approximation of $m(\cdot)$. It is also present in our estimator. The second term results from the unknown fixed effects and it is zero only in the case that they add the additional (strong) restriction that $\sum_i \mu_i = 0$. This type of restrictions is also used in Mammen et al. (2009).

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

2. STATISTICAL MODEL AND ESTIMATION PROCEDURE

To illustrate our technique we start by the univariate case and further we will present our results for the multivariate case. Then consider (1.2) with $d = q = 1$. In this case, for any $z \in A$, where A is a compact subset in a nonempty interior of \mathbb{R} , one has the following Taylor expansion

$$\begin{aligned} X_{it}m(Z_{it}) - X_{i(t-1)}m(Z_{i(t-1)}) &\approx m(z)\Delta X_{it} + m'(z)[X_{it}(Z_{it} - z) - X_{i(t-1)}(Z_{i(t-1)} - z)] \\ &+ \frac{1}{2}m''(z)[X_{it}(Z_{it} - z)^2 - X_{i(t-1)}(Z_{i(t-1)} - z)^2] + \cdots + \frac{1}{p!}m^{(p)}(z)[X_{it}(Z_{it} - z)^p - X_{i(t-1)}(Z_{i(t-1)} - z)^p] \\ &\equiv \sum_{\lambda=0}^p \beta_\lambda [X_{it}(Z_{it} - z)^\lambda - X_{i(t-1)}(Z_{i(t-1)} - z)^\lambda], \end{aligned}$$

which suggests that one estimates $m(z)$, $m'(z)$, \dots , $m^{(p)}(z)$ by regressing the ΔY_{it} 's on the terms $X_{it}(Z_{it} - z)^\lambda - X_{i(t-1)}(Z_{i(t-1)} - z)^\lambda$ with kernel weights. Then, the quantities of interest can be estimated using a locally weighted linear regression (see Fan and Gijbels (1995b), Ruppert and Wand (1994) or Zhan-Qian (1996)),

$$\sum_{i=1}^N \sum_{t=2}^T \{ \Delta Y_{it} - \beta_0 \Delta X_{it} - \beta_1 [X_{it}(Z_{it} - z) - X_{i(t-1)}(Z_{i(t-1)} - z)] \}^2 K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z), \quad (2.4)$$

where K is a bivariate kernel such that $K(u, v) = K(u)K(v)$, where for each u, v ,

$$\int K(u)du = 1 \quad \text{and} \quad K_h(u) = \frac{1}{h}K(u/h),$$

and h is a bandwidth. Denote by $\hat{\beta}_0$ and $\hat{\beta}_1$ the minimizers of (2.4). The above exposition suggests as estimators for $m(\cdot)$ and $m'(\cdot)$, $\hat{m}_h(z) = \hat{\beta}_0$ and $\hat{m}'_h(z) = \hat{\beta}_1$ respectively.

In particular, for the case of a local constant approximation ($p = 0$) (i.e. Nadaraya-Watson kernel regression estimator) the estimator for $m(z)$ has the following closed form

$$\hat{\beta}_0 = \frac{\sum_{i=1}^N \sum_{t=2}^T K_h(Z_{it} - z)K_h(Z_{i(t-1)} - z)\Delta X_{it}\Delta Y_{it}}{\sum_{i=1}^N \sum_{t=2}^T K_h(Z_{it} - z)K_h(Z_{i(t-1)} - z)(\Delta X_{it})^2}, \quad (2.5)$$

and in the local linear regression case ($p = 1$) we have

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left(\sum_{it} K_h(Z_{it} - z)K_h(Z_{i(t-1)} - z)\tilde{Z}_{it}\tilde{Z}_{it}^T \right)^{-1} \sum_{it} K_h(Z_{it} - z)K_h(Z_{i(t-1)} - z)\tilde{Z}_{it}\Delta Y_{it} \quad (2.6)$$

where \tilde{Z}_{it} is a 2×1 vector such that

$$\tilde{Z}_{it}^T = \left(\Delta X_{it}, \quad X_{it}(Z_{it} - z) - X_{i(t-1)}(Z_{i(t-1)} - z) \right), \quad i = 1, \dots, N; t = 2, \dots, T. \quad (2.7)$$

Note that in (2.4) we propose a bivariate kernel that also contains $Z_{i(t-1)}$ instead of considering only Z_{it} . The reason is that, if we consider only a kernel around Z_{it} , the transformed regression equation (1.2) would be originally localized around Z_{it} without considering all other values. Consequently, the distance between Z_{is} (for $s \neq t$) and z cannot be controlled by the fixed bandwidth parameter and so that the transformed remainder terms cannot be negligible. The consequence of all that would be a nondegenerated bias in this type of local linear estimators that it is removed by considering a local approximation around the pair $(Z_{it}, Z_{i(t-1)})$. In Theorem

3.1 we will show that the local linear estimator with the bivariate kernel shows the same rate as standard local linear smoothers estimators. That is, with a bias of order $O(h^2)$. Unfortunately, the well known trade-off between bias and variance term appears and, although the introduction of this bivariate kernel drops the bias out, it enlarges the variance term that becomes of order $O(\frac{1}{NT h^2})$. This is also emphasized in Theorem 3.1. Of course this affects the achievable rate of convergence for this type of problems that slows down at the rate $\sqrt{NT}h$.

In order to recover the desirable rate of $\sqrt{NT}h$ we propose a transformation that is basically a one step backfitting algorithm. Let us denote by $\Delta Y_{it}^{(1)}$ the following expression

$$\Delta Y_{it}^{(1)} = \Delta Y_{it} + m(Z_{i(t-1)})X_{i(t-1)}, \quad i = 1, \dots, N; t = 2, \dots, T. \quad (2.8)$$

By substituting (1.2) into (2.8) we get

$$\Delta Y_{it}^{(1)} = m(Z_{it})X_{it} + \Delta v_{it}, \quad i = 1, \dots, N; t = 2, \dots, T. \quad (2.9)$$

As it can be realized from (2.9), estimation of $m(\cdot)$ is now a one dimensional problem, and therefore we can use again a local linear least squares estimation procedure with univariate kernel weights. However, there is still a problem that needs to be solved. In (2.8) the term $m(Z_{i(t-1)})$ is unknown. Then, we replace it by the initial local linear regression estimator, i.e. $\Delta \tilde{Y}_{it}^{(1)} = \Delta Y_{it} + \hat{m}_h(Z_{i(t-1)})X_{i(t-1)}$, having the following regression model

$$\Delta \tilde{Y}_{it}^{(1)} = m(Z_{it})X_{it} + \nu_{it}, \quad i = 1, \dots, N; t = 2, \dots, T, \quad (2.10)$$

where

$$\nu_{it} = \{\hat{m}_h(Z_{i(t-1)}) - m(Z_{i(t-1)})\}X_{i(t-1)} + \Delta v_{it}, \quad i = 1, \dots, N; t = 2, \dots, T.$$

By doing so, we can estimate $m(\cdot)$ through the following weighted local linear regression:

$$\sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta \tilde{Y}_{it}^{(1)} - \gamma_0 X_{it} - \gamma_1 X_{it}(Z_{it} - z) \right\}^2 K_{\tilde{h}}(Z_{it} - z). \quad (2.11)$$

Let $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ be the minimizers of (2.11). Then, as before, we propose as estimators for $m(\cdot)$ and $m'(\cdot)$, $\tilde{m}_{\tilde{h}}(z) = \tilde{\gamma}_0$ and $\tilde{m}'_{\tilde{h}}(z) = \tilde{\gamma}_1$ respectively.

Now, once our estimation procedure has been fully explained for the univariate case, we proceed to extend our results for the multivariate case. That is for $d \neq q \neq 1$ in (1.1). In this case, the quantities of interest can be estimated using a multivariate locally weighted linear regression,

$$\sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta Y_{it} - \tilde{Z}_{it}^T \beta \right\}^2 K_H(Z_{it} - z) K_H(Z_{i(t-1)} - z), \quad (2.12)$$

where we denote by

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

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

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

where H is a $q \times q$ symmetric positive definite bandwidth matrix and finally, we denote by

$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0^T & \hat{\beta}_1^T \end{pmatrix}^T$ a $d(1+q)$ -vector that minimizes (2.12). Again, the above exposition suggests as estimators for $m(z)$ and $D_m(z) = \frac{\partial m(z)}{\partial z}$, $\hat{m}(z; H) = \hat{\beta}_0$ and $\text{vec}(\hat{D}_m(z; H)) = \hat{\beta}_1$ respectively. $D_m(z)$ is a $d \times q$ -matrix of partial derivatives of the d -function $m(z)$ with respect the elements of the the $q \times 1$ vector z .

It is easy to verify that the solution to the minimization problem in (2.12) can be written in matrix form as

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta Y, \quad (2.13)$$

where

$$W = \text{diag} \{ K_H(Z_{12} - z)K_H(Z_{11} - z), \dots, K_H(Z_{NT} - z)K_H(Z_{N(T-1)} - z) \},$$

$$\Delta Y = [\Delta Y_{12}, \dots, \Delta Y_{NT}]^T,$$

and

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

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

$$\hat{m}(z; H) = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta Y, \quad (2.14)$$

where $e_1 = (I_d; 0_{dq \times d})$ is a $d(1+q) \times d$ selection matrix, I_d is a $d \times d$ identity matrix and $0_{dq \times d}$ a $dq \times d$ matrix of zeros. Note that the dimensions of W and \tilde{Z} are respectively $N(T-1) \times N(T-1)$ and $N(T-1) \times d(1+q)$.

Finally, there are several reasons to chose local linear least squares estimators against other candidates. First, the form in (2.14) suggests that this estimator is found by fitting a plane to the data using weighted least squares. The weights are chosen according to the kernel and the bandwidth matrix H . As it is already discussed in Ruppert and Wand (1994) if a Gaussian kernel with (possibly) compact support is chosen, then the weight given to Z_{it} is the value of the Gaussian density with mean $Z_{it} - z$ which has an ellipsoidal contour of the form $(Z_{it} - z)^T H^{-1} (Z_{it} - z) = c$, for $c > 0$. Clearly, the farther from z is Z_{it} the less weight it receives. However, H controls both the size and orientation of the ellipsoids at a given density level and therefore it controls also the amount and direction of the weights. Often, instead of taking a matrix H , we adopt a simpler form $H = \text{diag}\{h_1^2, \dots, h_q^2\}$. If we have a diagonal bandwidth matrix, this means that the ellipsoids have their axes in the same direction as the coordinate axes, whereas for a general H matrix they will correspond to the eigenvectors of H and, depending on the shape of $m(\cdot)$, there are situations where having a full bandwidth matrix is advantageous. Another important advantage of local linear least squares kernel estimators is that the asymptotic bias and variance expressions are particularly appealing and appear to be superior to those of Naradaya-Watson or other nonparametric estimators. In particular, Fan (1993) shows that the local linear least squares estimator has an important asymptotic minimax property. Furthermore, unlike the Naradaya-

Watson or other nonparametric estimators, the bias and variance of (2.14) near the boundary of the density of Z are of the same order of magnitude as in the interior. That is a very interesting property because, in applications, the boundary region might comprise a large proportion of the data.

3. ASYMPTOTIC PROPERTIES AND THE ORACLE EFFICIENT ESTIMATOR

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

ASSUMPTION 3.1. *Let $(Y_{it}, X_{it}, Z_{it})_{i=1, \dots, N; t=1, \dots, T}$ be a set of independent and identically distributed \mathbb{R}^{d+q+1} -random variables in the subscript i for each fixed t and strictly stationary over t for fixed i . They are a sample following (1.1). Furthermore, let $f_{Z_{1t}}(\cdot)$, $f_{Z_{1t}, Z_{1(t-1)}}(\cdot, \cdot)$, $f_{Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)}}(\cdot, \cdot, \cdot)$ be respectively the probability density functions of Z_{1t} , $(Z_{1t}, Z_{1(t-1)})$ and $(Z_{1t}, Z_{1(t-1)}, Z_{1(t-2)})$. All density functions are continuously differentiable in all their arguments and they are bounded from above and below in any point of their support.*

ASSUMPTION 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 . Furthermore, $E|v_{it}|^{2+\delta} < \infty$, for some $\delta > 0$.*

ASSUMPTION 3.3. *μ_i can be arbitrarily correlated with both X_{it} and Z_{it} with unknown correlation structure.*

ASSUMPTION 3.4. *Let $\|A\| = \sqrt{\text{tr}(A^T A)}$, then $E\{\|X_{it}X_{it}^T\|^2 | Z_{it} = z_1, Z_{i(t-1)} = z_2\}$ is bounded and uniformly continuous in its support. Furthermore, let $\mathcal{X}_{it} = \begin{pmatrix} X_{it}^T & X_{i(t-1)}^T \end{pmatrix}^T$ and $\Delta\mathcal{X}_{it} = \begin{pmatrix} \Delta X_{it}^T & \Delta X_{i(t-1)}^T \end{pmatrix}^T$. Also, the following matrix functions $E(\mathcal{X}_{it}\mathcal{X}_{it}^T | Z_{it} = z_1, Z_{i(t-1)} = z_2)$, $E(\mathcal{X}_{it}\mathcal{X}_{it}^T | Z_{it} = z_1, Z_{i(t-1)} = z_2, Z_{i(t-2)} = z_3)$, $E(\Delta\mathcal{X}_{it}\Delta\mathcal{X}_{it}^T | Z_{it} = z_1, Z_{i(t-1)} = z_2)$ and $E(\mathcal{X}_{it}\Delta\mathcal{X}_{it}^T | Z_{it} = z_1, Z_{i(t-1)} = z_2, Z_{i(t-2)} = z_3)$ are bounded and uniformly continuous in their support.*

ASSUMPTION 3.5. *The function $E[\Delta\mathcal{X}_{it}\Delta\mathcal{X}_{it}^T | Z_{it} = z_1, Z_{i(t-1)} = z_2]$ is positive definite for any interior point of (z_1, z_2) in the support of $f_{Z_{it}, Z_{i(t-1)}}(z_1, z_2)$.*

ASSUMPTION 3.6. *For some $\delta > 0$, the functions $E\{|X_{it}\Delta v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{i(t-1)} = z_2\}$, $E\{|X_{i(t-1)}\Delta v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{i(t-1)} = z_2\}$ and $E\{|\Delta X_{it}\Delta v_{it}|^{2+\delta} | Z_{it} = z_1, Z_{i(t-1)} = z_2\}$ are bounded and uniformly continuous in any point of their support.*

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

ASSUMPTION 3.8. *The Kernel functions K are compactly supported, bounded kernel such that*

$\int uu^T K(u)du = \mu_2(K)I$ 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 ι_1, \dots, ι_q such that their sum is odd.

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

As the reader may notice, all assumptions are rather standard in nonparametric regression analysis of panel data models. Assumption 3.1 establishes standard features about the sample and data generating process. The individuals are independent and for a fixed individual we allow for correlation along time. Also other possible time series structures might be considered such as strong mixing conditions (see Cai and Li (2008)) or nonstationary time series data (see Cai et al. (2009)). Mixing conditions are usually taken into account to make the covariances of the estimator tend to zero at a faster rate. In our case this was not needed because our asymptotic analysis is performed for fixed T . On the other side, nonstationary processes we believe they exceed the scope of this paper. Note also that marginal densities are assumed to be bounded from above and below. This assumption can be relaxed at the price of increasing the mathematical complexity of the proofs.

Assumption 3.2 is also standard for first difference estimators (see Wooldridge (2002) for the fully parametric case). Furthermore, independence between the v_{it} -errors and the X_{it} -variables and/or the Z_{it} -variables is assumed without loss of generality. We could relax this assumption by assuming some dependence based on second order moments. For example, heteroskedasticity of unknown form can be allowed and in fact, under more complex structures in the variance-covariance matrix a transformation of the estimator proposed in You et al. (2010) can be developed in our setting. This type of assumption also rules out the existence of endogenous explanatory variables and imposes strict exogeneity conditions. If this would be the case then an instrumental variable approach such as the one proposed in Cai and Li (2008) or Cai and Xiong (2012) is needed. Assumption 3.3 imposes the so called fixed effects. Note that this assumption is much weaker than the one introduced in (Sun et al. (2009)) for their Least Squares Dummy Variable approach to work. Basically they impose a smooth relationship between heterogeneity and explanatory variables and, to avoid an additional bias term they need $\sum_i \mu_i = 0$.

Assumptions 3.4 and 3.5 are some smoothness conditions on moment functionals. Assumption 3.6 is the equivalent to a standard rank condition for identification of this type of models. Assumptions 3.7 to 3.9 are standard in local linear regression estimators (see Ruppert and Wand (1994)). Finally, all our results hold straightforwardly for the random coefficient setting.

Under these assumptions we now establish some results on the conditional mean and the conditional variance of the local linear least squares estimator.

THEOREM 3.1. Assume conditions 3.1-3.9 hold, then as N tends to infinity and T is fixed we obtain

$$\begin{aligned} & E\{\widehat{m}(z; H)|X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}\} - m(z) \\ &= \frac{1}{2}e_1^T \left(\widetilde{Z}^T W \widetilde{Z}\right)^{-1} \widetilde{Z}^T W \{S_{m_1}(z) - S_{m_2}(z)\} \\ &= \frac{1}{2}\mathcal{B}_{\Delta X \Delta X}^{-1}(z, z) [\mu_2(K_u) \mathcal{B}_{\Delta X X}(z, z) - \mu_2(K_v) \mathcal{B}_{\Delta X X_{-1}}(z, z)] \text{diag}_r \{tr\{\mathcal{H}_{m_r}(z) H\}\} i_d + o_p(tr\{H\}), \end{aligned}$$

where for $r = 1, \dots, d$, $\mathcal{H}_{m_r}(z)$ is the Hessian matrix of the r th component of $m(\cdot)$, while for $\ell = 1, 2$ the i th element of S_{m_ℓ} is

$$\{X_{i(t+1-\ell)} \otimes (Z_{i(t+1-\ell)} - z)\}^T \mathcal{H}_m(z) (Z_{i(t+1-\ell)} - z).$$

Furthermore, if $\mu_2(K_u) = \mu_2(K_v)$ the bias term becomes

$$E\{\widehat{m}(z; H)|X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}\} - m(z) = \frac{1}{2}\mu_2(K_u) \text{diag}_r \{tr\{\mathcal{H}_{m_r}(z) H\}\} i_d + o_p(tr\{H\}).$$

The variance is

$$\text{Var}\{\widehat{m}(z; H)|X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}\} = \frac{2\sigma_v^2 R(K_u) R(K_v)}{NT|H|} \mathcal{B}_{\Delta X \Delta X}^{-1}(z, z) \{1 + o_p(1)\},$$

where

$$\begin{aligned} \mathcal{B}_{\Delta X X}(z, z) &= E[\Delta X_{it} X_{it}^T | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it}, Z_{i(t-1)}}(z, z), \\ \mathcal{B}_{\Delta X X_{-1}}(z, z) &= E[\Delta X_{it} X_{i(t-1)}^T | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it}, Z_{i(t-1)}}(z, z), \\ \mathcal{B}_{\Delta X \Delta X}(z, z) &= E[\Delta X_{it} \Delta X_{it}^T | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it}, Z_{i(t-1)}}(z, z), \end{aligned}$$

$\text{diag}_r \{tr\{\mathcal{H}_{m_r}(z) H\}\}$ stands for a diagonal matrix of elements $tr\{\mathcal{H}_{m_r}(z) H\}$, for $r = 1, \dots, d$, and i_d is a $d \times 1$ unit vector.

The proof of this result is done in the Appendix. Just to illustrate the asymptotic behavior of our estimator, we give a result for the case when $d = q = 1$ and $H = h^2 I$. In this case, the above result can be written as

COROLLARY 3.1. Assume conditions 3.1-3.8 hold, then, if $h \rightarrow 0$ in such a way that $Nh^2 \rightarrow \infty$ as N tends to infinity and T is fixed we get

$$\begin{aligned} & E\{\widehat{m}(z; H)|X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}\} - m(z) \\ &= \frac{1}{2}c(z, z)m''(z)h^2 + o_p(h^2). \end{aligned}$$

where

$$c(z, z) = \frac{\mu_2(K_u) E[\Delta X_{it} X_{it} | Z_{it} = z, Z_{i(t-1)} = z] - \mu_2(K_v) E[\Delta X_{it} X_{i(t-1)} | Z_{it} = z, Z_{i(t-1)} = z]}{E[(X_{it} - X_{i(t-1)})^2 | Z_{it} = z, Z_{i(t-1)} = z]}.$$

Furthermore, if $\mu_2(K_u) = \mu_2(K_v)$ then the bias term has the following expression

$$E\{\widehat{m}(z; H)|X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}\} - m(z) = \frac{1}{2}m''(z)h^2 + o_p(h^2).$$

The variance is

$$\begin{aligned} & \text{Var} \{ \widehat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} \\ &= \frac{2\sigma_v^2 R(K_u) R(K_v)}{NT h^2 f_{Z_{it}, Z_{i(t-1)}}(z, z) E \left[(X_{it} - X_{i(t-1)})^2 \middle| Z_{it} = z, Z_{i(t-1)} = z \right]} \{1 + o_p(1)\}. \end{aligned}$$

Note that in the standard case that $\mu_2(K_u) = \mu_2(K_v)$ then we obtain a nice result for the bias. In fact, the resulting asymptotic bias has the same expression as in the standard local linear estimator.

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

THEOREM 3.2. *Assume conditions 3.1-3.9 hold. Then,*

$$\sqrt{NT|H|} \{ \widehat{m}(z; H) - m(z) \} \rightarrow_d \mathcal{N}(b(z), v(z)),$$

as N tends to infinity and T is fixed, where

$$\begin{aligned} b(z) &= \frac{1}{2} \mu_2(K_u) \text{diag}_r \left\{ \text{tr} \left\{ \mathcal{H}_{md}(z) H \sqrt{NT|H|} \right\} \right\} \iota_d, \\ v(z) &= 2\sigma_v^2 R(K_u) R(K_v) \mathcal{B}_{\Delta X \Delta X}^{-1}(z, z). \end{aligned}$$

The proof of this result is shown in Appendix.

Note that the rate at which our estimator converges is $NT|H|$. Under the conditions established in the propositions, our estimator is both consistent and asymptotically normal. However, its rate of convergence is sub-optimal since the lower rate of convergence for this type of estimators is $NT|H|^{1/2}$. As we already indicated in the previous section, in order to achieve optimality we propose to reduce the dimensionality of the problem by redefining ΔY_{it} as in (2.10), now for the multivariate case,

$$\Delta \widetilde{Y}_{it}^{(1)} = X_{it}^T m(Z_{it}) + \nu_{it}, \quad i = 1, \dots, N; t = 2, \dots, T,$$

where

$$\nu_{it} = X_{i(t-1)}^T \{ \widehat{m}(Z_{i(t-1)}; H) - m(Z_{i(t-1)}) \} + \Delta \nu_{it}, \quad i = 1, \dots, N; t = 2, \dots, T. \quad (3.15)$$

In expression (3.15), $\widehat{m}(Z_{i(t-1)}; H)$ is the first-step local linear estimator obtained in (2.14). Now we propose to estimate $m(Z_{it})$ using a multivariate locally weighted linear regression,

$$\sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta \widetilde{Y}_{it}^{(1)} - (X_{it}^T \gamma_0 + X_{it}^T \otimes (Z_{it} - z)^T \gamma_1) \right\}^2 K_{\widetilde{H}}(Z_{it} - z), \quad (3.16)$$

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

If we define $\tilde{Z}_{it}^{(1)T} = [X_{it}^T \quad X_{it}^T \otimes (Z_{it} - z)^T]$ as a $1 \times d(1+q)$ vector, (3.16) can be written as

$$\sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta Y_{it}^{(1)} - \tilde{Z}_{it}^{(1)T} \gamma \right\}^2 K_{\tilde{H}}(Z_{it} - z), \quad (3.17)$$

where we denote by $\tilde{\gamma} = (\tilde{\gamma}_0^T \quad \tilde{\gamma}_1^T)^T$ the $d(1+q)$ -vector that minimizes (3.17). Following the same reasoning as before we may write

$$\tilde{m}(z; \tilde{H}) \equiv \tilde{\gamma}_0 = e_1^T \left(\tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \Delta \tilde{Y}^{(1)}, \quad (3.18)$$

where $\Delta \tilde{Y}^{(1)} = [\Delta \tilde{Y}_{12}^{(1)}, \dots, \Delta \tilde{Y}_{NT}^{(1)}]^T$, $W^{(1)} = \text{diag}\{K_H(Z_{12} - z), \dots, K_H(Z_{NT} - z)\}$ and

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

In order to show the asymptotic properties of this estimator we need to assume the following about the bandwidth \tilde{H} and its relationship with H :

ASSUMPTION 3.10. *The bandwidth matrix \tilde{H} is symmetric and strictly definite positive. Furthermore, each entry of the matrix tends to zero as N tends to infinity in such a way that $N |\tilde{H}| \rightarrow \infty$.*

ASSUMPTION 3.11. *The bandwidth matrices H and \tilde{H} must fulfill that $N |H| |\tilde{H}| / \log(N) \rightarrow \infty$, and $\text{tr}(H) / \text{tr}(\tilde{H}) \rightarrow 0$ as N tends to infinity.*

In general for the kernel function and conditional moments and densities we need both smoothness and boundedness conditions already established in assumptions 3.1 to 3.8. They are required to use uniform convergence results as the ones established in Masry (1996). It is then possible to show the following result

THEOREM 3.3. *Assume conditions 3.1-3.8 and 3.10-3.11 holds, then, as N tends to infinity and T is fixed we get*

$$E \left\{ \tilde{m}(z; \tilde{H}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} - m(z) = \frac{1}{2} \mu_2(K_u) \text{diag}_r \left\{ \text{tr} \left\{ \mathcal{H}_{m_r}(z) \tilde{H} \right\} \right\} \iota_{d+o_p} \left(\text{tr} \left\{ \tilde{H} \right\} \right)$$

and

$$\text{Var} \left\{ \tilde{m}(z; \tilde{H}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} = \frac{2\sigma_v^2 R(K_u)}{NT |\tilde{H}|^{1/2}} \mathcal{B}_{XX}^{-1}(z) \{1 + o_p(1)\},$$

where $\text{diag}_r \left\{ \text{tr} \left\{ \mathcal{H}_{m_r}(z) \tilde{H} \right\} \right\}$ stands for a diagonal matrix of elements $\text{tr} \left\{ \mathcal{H}_{m_r}(z) \tilde{H} \right\}$, for $r = 1, \dots, d$, and ι_d is a $d \times 1$ unit vector.

The proof of the Theorem 3.3 is done in the Appendix.

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

4. BANDWIDTH SELECTION

As the reader can realize from previous sections, the bandwidth matrix H plays a crucial role in the estimation of the unknown quantity $m(\cdot)$. In fact, as we have learned from the asymptotic expressions, when choosing H there exists a trade off between the bias and the variance of our estimator. Consider the simplest case, $H = h^2 I$. If we choose h very small, then according to Corollary 3.1 the bias of our estimator will be reduced (it is of order h^2) but at the price of enlarging the variance (the order of this term is $1/NTh^2$). This trade off should be solved by choosing a bandwidth matrix H that minimizes the Mean Square Error (MSE), that is the sum of the squared bias and variance. There exists many different measures of discrepancy between the estimator $\hat{m}(\cdot; H)$ and the function $m(\cdot; H)$. In Härdle (1990), Chapter 5, it can be found a comprehensive discussion of these measures. For the sake of simplicity, and taking into account the data generating process in (1.1) we propose the following measure of discrepancy,

$$MSE(H) = E [X^T (\hat{m}(Z; H) - m(Z))]^2.$$

In this MSE, the expectation is taken over $Z_1, \dots, Z_q; X_1, \dots, X_d$ and $\hat{m}(Z; H)$ is the estimator defined in (2.14). Therefore, for our problem, we can define the optimal bandwidth matrix H_{opt} as the solution to the following minimization problem,

$$H_{opt} = \arg \min_H MSE(H) = \arg \min_H E [X^T (\hat{m}(Z; H) - m(Z))]^2.$$

If $Z_1, \dots, Z_q; X_1, \dots, X_d$ are random variables that are independent of the observed sample $\mathcal{D} = (X_{11}, Z_{11}, \dots, X_{NT}, Z_{NT})^T$, but they share the same distribution with (X_{11}, Z_{11}) it is straightforward to show that

$$MSE(H) = E [b^T(Z) \Omega(Z) b(Z) + \text{tr} \{ \Omega(Z) V(Z) \}], \quad (4.19)$$

where

$$\begin{aligned} b(Z) &= E \{ \hat{m}(Z; H) | \mathcal{D}, Z \} - m(Z), \\ V(Z) &= \text{Var} \{ \hat{m}(Z; H) | \mathcal{D}, Z \}, \quad \text{and} \\ \Omega(Z) &= E (X X^T | Z). \end{aligned}$$

As it can be realized from the expression above, it has been now formalized the idea of choosing a bandwidth matrix H that minimizes the MSE, that is the sum of the squared bias and variance.

Note that, the way we have defined the measure of discrepancy determines, in our case, the choice of a global bandwidth. That is, we will choose a bandwidth that remains constant with the location point. Of course, another possibility would be to choose a bandwidth that varies locally according to this location point. i.e. $H(z)$. In this case, the local MSE criteria would be

$$MSE(z; H) = E [X^T (\hat{m}(z; H) - m(z))]^2,$$

where now the expectation is taken over X . Müller and Stadtmüller (1987) discussed the issue of local variable bandwidth for convolution-type regression estimators. Furthermore, Fan and Gijbels (1992) propose a variable bandwidth for the estimation of local polynomial regression. In our case, we propose to choose a global bandwidth. The reason is twofold. First, all components in our model have been assumed to have the same degree of smoothness and second, the use of local bandwidths, except for the case where the curve presents a rather complicated structure, increases the computational burden without much improvement in the final results. This is probably due to the local adaptation property that already exhibits local linear regression smoothers.

Unfortunately, the selection of H_{opt} does not solve all problems in bandwidth selection. In fact, as it can be realized, the MSE depends on some unknown quantities and therefore, our optimal bandwidth matrix can not be estimated from data. There are several alternative solutions to approximate the unknown quantities in the MSE. One alternative is to replace in (4.19) both bias and variance terms by their respective first order asymptotic expressions that were obtained in Theorem 3.1. This is the so called ‘plug-in’ method (see for details Ruppert et al. (1995)). Another possibility is, as suggested in Fan and Gijbels (1995a), to replace directly in (4.19) bias and variance by their exact expressions. That is

$$E \{ \hat{m}(Z; H) | \mathcal{D}, Z \} - m(z) = \{ E \{ \hat{m}(z; H) | \mathcal{D} \} - m(z) \}_{z=Z} \tag{4.20}$$

$$\text{Var} \{ \hat{m}(Z; H) | \mathcal{D}, Z \} = \text{Var} \{ \hat{m}(z; H) | \mathcal{D} \}_{z=Z}, \tag{4.21}$$

where clearly, according to Theorem 3.1

$$E \{ \hat{m}(z; H) | \mathcal{D} \} - m(z) = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \tau \tag{4.22}$$

$$\text{Var} \{ \hat{m}(z; H) | \mathcal{D} \} = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \mathcal{V} W \tilde{Z} \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} e_1,$$

τ is a $N(T-1)$ vector such that, for $i = 1, \dots, N, t = 2, \dots, T$,

$$\begin{aligned} \tau_{it} &= X_{it}^T m(Z_{it}) - X_{i(t-1)}^T m(Z_{i(t-1)}) \\ &\quad - \left\{ X_{it}^T D_m(z) (Z_{it} - z) - X_{i(t-1)}^T D_m(z) (Z_{i(t-1)} - z) \right\} \end{aligned}$$

and \mathcal{V} is a $N(T-1) \times N(T-1)$ matrix that contains the V_{ij} ’s matrices,

$$V_{ij} = E(\Delta v_i \Delta v_j^T | X_{i1}, \dots, X_{iT}, Z_{i1}, \dots, Z_{iT}) = \begin{cases} 2\sigma_v^2, & \text{for } i = j, \quad t = s, \\ -\sigma_v^2, & \text{for } i = j, \quad |t - s| < 2, \\ 0, & \text{for } i = j, \quad |t - s| \geq 2. \end{cases} \tag{4.23}$$

In order to estimate both bias and variance we need to calculate τ and \mathcal{V} . Note that for τ , developing a fifth order Taylor expansion of both $m(Z_{it})$ and $m(Z_{i(t-1)})$ around z a local polynomial regression of order five would guarantee that the proposed bandwidth selection procedure will be

\sqrt{N} -consistent for the local linear fit (see Hall et al. (1991) for details). However, for the sake of simplicity a local cubic polynomial regression would be close to a \sqrt{N} -consistent selection rule and it will lead to a nice reduction in the computational effort. In this case (for $d = q = 1$), the vector $\hat{\tau}$ will contain the (estimated) expressions for the second and third order derivatives of the local cubic polynomial regression of the terms ΔY_{it} onto $X_{it} (Z_{it} - z)^\lambda - X_{i(t-1)} (Z_{i(t-1)} - z)^\lambda$, $\lambda = 0, 1, \dots, 3$.

On the other side, in order to estimate \mathcal{V} , note that, because of assumption 3.2, estimation of \mathcal{V} is tantamount to the estimation of σ_v^2 . In order to estimate this last quantity note that under the homoskedastic assumption we can consistently estimate this by

$$\hat{\sigma}_v^2 = \frac{1}{2N(T-1)} \sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta Y_{it} - \Delta X_{it}^T \hat{m}^{-i}(Z_{it}; H) + \Delta X_{i(t-1)}^T \hat{m}^{-i}(Z_{i(t-1)}; H) \right\}^2. \quad (4.24)$$

Note that both $\hat{\tau}$ and $\hat{\sigma}_v^2$ depend on a bandwidth matrix H that needs to be determined from data. A suitable pilot bandwidth matrix H^* that can be used for these computations can be obtained using the global RSC procedure proposed in Fan and Gijbels (1995a). Furthermore, we denote by $\hat{m}^{-i}(Z_{it}; H)$ the leave-one-out estimator of $m(Z_{it})$. That is, when estimating $m(Z_{it})$ using (2.14) we use all data except those that belong to the i -th subject. Note that once we have estimated τ and σ_v^2 we can now provide an estimator for $b(H)$, $V(H)$ and $\Omega(H)$. Mainly,

$$\begin{aligned} \hat{b}(Z_{it}) &= E \{ \hat{m}^{-i}(Z_{it}; H) | \mathfrak{D} \} - m(Z_{it}) = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \hat{\tau}, \\ \hat{V}(Z_{it}) &= \text{Var} \{ \hat{m}^{-i}(Z_{it}; H) | \mathfrak{D} \} = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \hat{V} W \tilde{Z} \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} e_1, \\ \hat{\Omega}(Z_{it}) &= \frac{\sum_{j \neq i, t} X_{jt} X_{jt}^T K_H(Z_{jt} - Z_{it}) K_H(Z_{j(t-1)} - Z_{i(t-1)})}{\sum_{j \neq i, t} K_H(Z_{jt} - Z_{it}) K_H(Z_{j(t-1)} - Z_{i(t-1)})}. \end{aligned}$$

The corresponding estimator of the $MSE(H)$, according with (4.19) will be

$$\widehat{MSE}(H) = \frac{1}{N(T-1)} \sum_{it} \left[\hat{b}^T(Z_{it}) \hat{\Omega}(Z_{it}) \hat{b}(Z_{it}) + \text{tr} \left\{ \hat{\Omega}(Z_{it}) \hat{V}(Z_{it}) \right\} \right]. \quad (4.25)$$

Then, we define the estimator of H_{opt} , \hat{H}_{opt} as the solution to the following problem,

$$\hat{H}_{opt} = \arg \min_H \widehat{MSE}(H).$$

Although we do not provide theoretical properties of this bandwidth, in Zhang and Lee (2000) they have been studied in detail for the local MSE case and we believe it is straightforward to analyze them for the global MSE case that we present here. Finally, we propose to use the same procedure to estimate the bandwidth matrix H when estimating the oracle efficient estimator.

5. MONTE CARLO EXPERIMENT

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

We consider the following varying-coefficient nonparametric models,

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

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

We consider three different cases of study,

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

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

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

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

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

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

$$MSE\{\hat{m}_l(z; H)\} = \frac{1}{Q} \sum_{l=1}^Q E \left[\left\{ \sum_{r=1}^d (\hat{m}_{lr}(z; H) - m_{lr}(z)) X_{it,lr} \right\}^2 \right]$$

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

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

The simulation results are summarized in the following figures.

We further carried out a simulation study to analyze the behavior in finite samples of the multivariate locally estimator with kernel weights, $\hat{m}(z; H)$, and the oracle estimator, $\tilde{m}(z; \tilde{H})$, proposed in Sections 2 and 3. Looking at Figures 1, 2 and 3 we can highlight the following.

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

On the other hand, from Figures 1, 2 and 3 we can see that both estimators carry out quite

	Local polynomial	Backfitting
n=150	1.374397	1.275632
n=300	1.283836	1.185536
n=600	1.232254	1.176898

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

	Local polynomial	Backfitting
n=150	5.765032	1.6335596
n=300	3.259442	1.1884004
n=600	3.050246	0.9952557

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

	Local polynomial	Backfitting
n=150	2.08248	1.721583
n=300	1.755733	1.490219
n=600	1.549600	1.375290

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

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

6. CONCLUSION

This paper introduces a new technique that estimates varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in an unknown way. The resulting estimator is robust to misspecification in the functional form of the varying parameters and we have shown that it is consistent and asymptotically normal. Furthermore we have shown that it achieves the optimal rate of convergence for

this type of problems and it exhibits the so called oracle efficiency property. Since the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. Monte Carlo results indicates good performance of the estimator in finite samples.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Programa Nacional de Formación de Profesorado Universitario/ Spanish Ministry of Education. Ref. AP-2007-02209. We would like to thank also Wenceslao Gonzalez-Manteiga, Juan A. Cuesta-Albertos and Stefan Sperlich for their very helpful comments and suggestions. We acknowledge also the comments made by the Co-Editor, Oliver Linton, and one referee. They really helped us to improve the quality of the paper. Of course, all possible errors and mistakes are ours.

REFERENCES

- Cai, Z. and Q. Li (2008). “Nonparametric estimation of varying coefficient dynamic panel data models”. *Econometric Theory* 24, 1321–1342.
- Cai, Z., Q. Li, and J. P. Park (2009). “Functional-coefficient models for nonstationary time series data”. *Journal of Econometrics* 148, 101–113.
- Cai, Z. and H. Xiong (2012). “Partially varying coefficient instrumental variables models”. *Statistica Neerlandica* 66, 85–110.
- Fan, J. (1993). “Local linear regression smoothers and their minimax efficiencies”. *Annals of Statistics* 21, 196–216.
- Fan, J. and I. Gijbels (1992). “Variable bandwidth and local linear regression smoothers”. *Annals of Statistics* 20, 2008–2036.
- Fan, J. and I. Gijbels (1995a). “Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation”. *Journal of the Royal Statistical Society, Ser. B* 57, 371–394.
- Fan, J. and I. Gijbels (1995b). *Local polynomial modelling and its applications*. Chapman & Hall.
- Fan, J. and W. Zhang (1999). “Statistical estimation in varying coefficient models”. *Annals of Statistics* 27, 1491–1518.
- Hall, P., S. J. Sheather, M. C. Jones, and J. S. Marron (1991). “On optimal data-based bandwidth selection in kernel density estimation”. *Biometrika* 78, 263–271.
- Härdle, W. (1990). *Applied nonparametric regression*. Cambridge University Press.
- Henderson, D. J., R. J. Carroll, and Q. Li (2008). “Nonparametric estimation and testing of fixed effects panel data model”. *Journal of Econometrics* 144, 257–275.
- Lee, Y. and D. Mukherjee (2008). “New nonparametric estimation of the marginal effects in fixed-effects panel models: An application on the environmental Kuznets Curve”. Working Paper. University of Michigan.
- Mammen, E., B. Støve, and D. Tjøstheim (2009). “Nonparametric additive models for panels of time series”. *Econometric Theory* 25, 442–481.

- Masry, E. (1996). “Multivariate local polynomial regression for time series: Uniform strong consistency rate”. *Journal of Time Series Analysis* 17, 571–599.
- Müller, H.-G. and U. Stadtmüller (1987). “Variable bandwidth kernel estimators of regression curves”. *Annals of Statistics* 15, 182–201.
- Neyman, J. and E. Scott (1948). “Consistent estimates based on partially consistent observations”. *Econometrica* 16, 1–32.
- Qian, J. and L. Wang (2012). “Estimating semiparametric panel data models by marginal integration”. *Journal of Econometrics* 167, 483–493.
- Ruppert, D., S. J. Sheather, and M. P. Wand (1995). “An effective bandwidth selector for local least squares regression”. *Journal of the American Statistical Association* 90, 1257–1270.
- Ruppert, D. and M. P. Wand (1994). “Multivariate locally weighted least squares regression”. *The Annals of Statistics* 22(3), 1346–1370.
- Schultz, T. P. (2003). “Human capital, schooling and health”. *Economics and Human Biology* 1, 207–221.
- Su, L. and A. Ullah (2011). “Nonparametric and semiparametric panel econometric models: estimation and testing”. in *Handbook of Empirical Economics and Finance*. A. Ullah and D.E.A. Giles (eds), New York: Taylor and Francis Group.
- Sun, Y., R. J. Carroll, and D. Li (2009). “Semiparametric estimation of fixed effects panel data varying coefficient models”. *Advances in Econometrics* 25, 101–130.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data models*. The MIT Press, Cambridge, Massachusetts.
- Yang, L. (2002). “Direct estimation in an additive model when the components are proportional”. *Statistica Sinica* 12, 801–821.
- You, J., X. Zhou, and Y. Zhou (2010). “Statistical inference for panel data semiparametric partially linear regression models with heteroscedastic errors”. *Journal of Multivariate Analysis* 101, 1079–1101.
- Zhan-Qian, L. (1996). “Multivariate locally weighted polynomial fitting and partial derivative estimation”. *Journal of Multivariate Analysis* 59, 187–205.
- Zhang, W. and S. Lee (2000). “Variable bandwidth selection in varying-coefficient models”. *Journal of Multivariate Analysis* 74, 116–134.

APPENDIX

Proof of Theorem 3.1

Taking assumption 3.2, conditional expectations in (2.14) and noting that

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

then

$$E\{\widehat{m}(z; H)|X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}\} = e_1^T (\widetilde{Z}^T W \widetilde{Z})^{-1} \widetilde{Z}^T W M \quad (\text{A. 1})$$

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

Taylor's Theorem implies that

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

where

$$Q_m(z) = S_{m1}(z) - S_{m2}(z), \quad (\text{A. 3})$$

$$S_{m1}(z) = [S_{m1,12}^T(z), \dots, S_{m1,NT}^T(z)]^T, \\ S_{m2}(z) = [S_{m2,11}^T(z), \dots, S_{m2,N(T-1)}^T(z)]^T$$

and

$$S_{m1,it}(z) = [\{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{H}_m(z)(Z_{it} - z)], \\ S_{m2,i(t-1)}(z) = [\{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{H}_m(z)(Z_{i(t-1)} - z)].$$

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

The remainder term can be written as

$$R(z) = R_1(z) - R_2(z), \quad (\text{A. 4})$$

$$R_1(z) = [R_{1,12}^T(z), \dots, R_{1,NT}^T(z)]^T, \\ R_2(z) = [R_{2,11}^T(z), \dots, R_{2,N(T-1)}^T(z)]^T$$

and

$$R_{1,it}(z) = [\{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z)], \\ R_{2,i(t-1)}(z) = [\{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{R}(Z_{i(t-1)}; z)].$$

We denote by

$$\mathcal{R}(Z_{it}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{it}; z) \\ \mathcal{R}_2(Z_{it}; z) \\ \vdots \\ \mathcal{R}_d(Z_{it}; z) \end{pmatrix}, \mathcal{R}(Z_{i(t-1)}; z) = \begin{pmatrix} \mathcal{R}_1(Z_{i(t-1)}; z) \\ \mathcal{R}_2(Z_{i(t-1)}; z) \\ \vdots \\ \mathcal{R}_d(Z_{i(t-1)}; z) \end{pmatrix},$$

and

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

We first analyze the bias term. In order to do this, note that substituting (A. 2) into (A. 1) and noting that $\text{vec}\{D_m(z)\}$ in (A. 2) vanishes because

$$e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \tilde{Z} \begin{bmatrix} m(z) \\ \text{vec}\{D_m(z)\} \end{bmatrix} = e_1^T \begin{bmatrix} m(z) \\ \text{vec}\{D_m(z)\} \end{bmatrix} = m(z), \quad (\text{A. 6})$$

then,

$$E \{ \hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(z) = \frac{1}{2} e_1^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W Q_m(z) + e_1^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W R(z). \quad (\text{A. 7})$$

We will first analyze the asymptotic behavior of $\frac{1}{2} e_1^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W Q_m(z)$. Later in the paper we will do the same with the second term. For the sake of simplicity let us denote

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

now, define the symmetric block matrix

$$(NT)^{-1} \tilde{Z}^T W \tilde{Z} = \begin{pmatrix} \mathcal{A}_{NT}^{11} & \mathcal{A}_{NT}^{12} \\ \mathcal{A}_{NT}^{21} & \mathcal{A}_{NT}^{22} \end{pmatrix} \quad (\text{A. 8})$$

where,

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

We first show that as N tends to infinity

$$\mathcal{A}_{NT}^{11} = \mathcal{B}_{\Delta X \Delta X}(z, z) + o_p(1), \quad (\text{A. 9})$$

where

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

In order to do so, note that under the stationarity assumption and using iterated expectations

$$\begin{aligned} E(\mathcal{A}_{NT}^{11}) &= \int \int E \left[\Delta X_{it} \Delta X_{it}^T | Z_{it} = z + H^{1/2}u, Z_{i(t-1)} = z + H^{1/2}v \right] \\ &\quad \times f_{Z_{it}, Z_{i(t-1)}}(Z_{it} = z + H^{1/2}u, Z_{i(t-1)} = z + H^{1/2}v) K(u) K(v) dudv. \end{aligned}$$

Furthermore, under assumptions 3.1 and 3.4 and a Taylor expansion, as N tends to infinity, (A. 9) holds. All what we need to close the proof is show that $\text{Var}(\mathcal{A}_{NT}^{11}) \rightarrow 0$, as the sample size

tends to infinity. Note that under the assumption 3.1,

$$\begin{aligned} \text{Var}(\mathcal{A}_{NT}^{11}) &= \frac{1}{NT} \text{Var}(\Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)}) \\ &+ \frac{1}{NT^2} \sum_{t=3} (T-t) \text{Cov}(\Delta X_{i2} \Delta X_{i2}^T K_{i2} K_{i1}, \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)}). \end{aligned}$$

Under assumptions 3.4 to 3.6

$$\text{Var}(\Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)}) \leq \frac{C}{NT|H|},$$

and

$$\text{Cov}(\Delta X_{i2} \Delta X_{i2}^T K_{i2} K_{i1}, \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)}) \leq \frac{C'}{N|H|}.$$

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

Similarly one can show that

$$\mathcal{A}_{NT}^{12} = \mathcal{DB}_{\Delta XX}(z, z)(I_d \otimes \mu_2(K_u)H) - \mathcal{DB}_{\Delta XX_{-1}}(z, z)(I_d \otimes \mu_2(K_v)H) + o_p(H). \quad (\text{A. 10})$$

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

$$\mathcal{DB}_{\Delta XX}(Z_1, Z_2) = \begin{pmatrix} \frac{\partial b_{11}^{\Delta XX}(Z_1, Z_2)}{\partial Z_1^T} & \dots & \frac{\partial b_{1d}^{\Delta XX}(Z_1, Z_2)}{\partial Z_1^T} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{d1}^{\Delta XX}(Z_1, Z_2)}{\partial Z_1^T} & \dots & \frac{\partial b_{dd}^{\Delta XX}(Z_1, Z_2)}{\partial Z_1^T} \end{pmatrix},$$

and

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

The other gradient matrix is

$$\mathcal{DB}_{\Delta XX_{-1}}(Z_1, Z_2) = \begin{pmatrix} \frac{\partial b_{11}^{\Delta XX_{-1}}(Z_1, Z_2)}{\partial Z_1^T} & \dots & \frac{\partial b_{1d}^{\Delta XX_{-1}}(Z_1, Z_2)}{\partial Z_1^T} \\ \vdots & \ddots & \vdots \\ \frac{\partial b_{d1}^{\Delta XX_{-1}}(Z_1, Z_2)}{\partial Z_1^T} & \dots & \frac{\partial b_{dd}^{\Delta XX_{-1}}(Z_1, Z_2)}{\partial Z_1^T} \end{pmatrix},$$

and

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

Finally,

$$\mathcal{A}_{NT}^{22} = \mathcal{B}_{XX}(z, z) \otimes \mu_2(K_u)H + \mathcal{B}_{X_{-1}X_{-1}}(z, z) \otimes \mu_2(K_v)H + o_p(H). \quad (\text{A. 11})$$

where

$$\begin{aligned} \mathcal{B}_{XX}(z, z) &= E[X_{it} X_{it}^T | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it}, Z_{i(t-1)}}(z, z), \\ \mathcal{B}_{X_{-1}X_{-1}}(z, z) &= E[X_{i(t-1)} X_{i(t-1)}^T | Z_{it} = z, Z_{i(t-1)} = z] f_{Z_{it}, Z_{i(t-1)}}(z, z). \end{aligned}$$

Using the results shown in (A. 9), (A. 10) and (A. 11) we obtain

$$NT(\tilde{Z}^T W \tilde{Z})^{-1} = \begin{pmatrix} \mathcal{C}_{11} & \mathcal{C}_{12} \\ \mathcal{C}_{21} & \mathcal{C}_{22} \end{pmatrix}, \quad (\text{A. 12})$$

where

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

Also it is straightforward to show that the terms in

$$(NT)^{-1} \tilde{Z}^T W S_{m1}(z)$$

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

are asymptotically equal to

$$\begin{aligned}
&(NT)^{-1} \sum_{it} \Delta X_{it} \{X_{it}^T \otimes (Z_{it} - z)^T\} \mathcal{H}_m(z) (Z_{it} - z) K_{it} K_{i(t-1)} \tag{A.13} \\
&= \mu_2(K_u) E \left[\Delta X_{it} X_{it}^T \middle| Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it}, Z_{i(t-1)}}(z, z) \times \text{diag}_r \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \} i_d + o_p(\text{tr} \{ H \}),
\end{aligned}$$

where $\text{diag}_r \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \}$ stands for a diagonal matrix of elements $\text{tr} \{ \mathcal{H}_{m_r}(z) H \}$, for $r = 1, \dots, d$, and i_d is a $d \times 1$ unit vector.

$$\begin{aligned}
&(NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \{X_{it}^T \otimes (Z_{it} - z)^T\} \mathcal{H}_m(z) (Z_{it} - z) K_{it} K_{i(t-1)} \\
&= \int \mathcal{B}_{XX}(z, z) \otimes (H^{1/2}u) (H^{1/2}u)^T \mathcal{H}_m(z) (H^{1/2}u) K(u) K(v) dudv \tag{A.14} \\
&\quad - \int \mathcal{B}_{X_{-1}X}(z, z) \otimes (H^{1/2}v) (H^{1/2}u)^T \mathcal{H}_m(z) (H^{1/2}u) K(u) K(v) dudv + o_p(H^{3/2}) \\
&= O_p(H^{3/2}).
\end{aligned}$$

Finally, the terms in

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

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

are of order

$$\begin{aligned}
&(NT)^{-1} \sum_{it} \Delta X_{it} \{X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T\} \mathcal{H}_m(z) (Z_{i(t-1)} - z) K_{it} K_{i(t-1)} \tag{A.15} \\
&= \mu_2(K_v) E \left[\Delta X_{it} X_{i(t-1)}^T \middle| Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it}, Z_{i(t-1)}}(z, z) \times \text{diag}_r \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \} i_d + o_p(\text{tr} \{ H \}),
\end{aligned}$$

and

$$\begin{aligned}
& (NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \{X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T\} \mathcal{H}_m(z)(Z_{i(t-1)} - z) K_{it} K_{i(t-1)} \\
&= \int \mathcal{B}_{XX_{-1}}(z, z) \otimes (H^{1/2}u) (H^{1/2}v)^T \mathcal{H}_m(z) (H^{1/2}v) K(u) K(v) dudv \\
&- \int \mathcal{B}_{X_{-1}X_{-1}}(z, z) \otimes (H^{1/2}v) (H^{1/2}v)^T \mathcal{H}_m(z) (H^{1/2}v) K(u) K(v) dudv + o_p(H^{3/2}) \\
&= O_p(H^{3/2}).
\end{aligned} \tag{A.16}$$

The second term for the bias expression is $e_1^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W R(z)$. We already know what is the asymptotic expression for $(\tilde{Z}^T W \tilde{Z})^{-1}$ so now we proceed to analyze the asymptotic behavior of $\tilde{Z}^T W R(z)$. According to (A. 4) and (A. 5) note that

$$(NT)^{-1} \tilde{Z}^T W R(z) = \begin{pmatrix} \mathcal{E}_1(z) \\ \mathcal{E}_2(z) \end{pmatrix},$$

where

$$\begin{aligned}
\mathcal{E}_1(z) &= \frac{1}{NT} \sum_{it} \Delta X_{it} \\
&\times \left[\{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{R}(Z_{i(t-1)}; z)(Z_{i(t-1)} - z) \right] K_{it} K_{i(t-1)}
\end{aligned} \tag{A.17}$$

and

$$\begin{aligned}
\mathcal{E}_2(z) &= \frac{1}{NT} \sum_{it} \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \times \\
&\left[\{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{R}(Z_{i(t-1)}; z)(Z_{i(t-1)} - z) \right] K_{it} K_{i(t-1)}.
\end{aligned} \tag{A.18}$$

$$\mathcal{E}_1(z) = \mathcal{E}_{11}(z) + \mathcal{E}_{12}(z), \tag{A.19}$$

where

$$\begin{aligned}
\mathcal{E}_{11}(z) &= \frac{1}{NT} \sum_{it} K_{it} K_{i(t-1)} \Delta X_{it} \times \\
&\left[\{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{i(t-1)} - z) \right]
\end{aligned} \tag{A.20}$$

and

$$\begin{aligned}
\mathcal{E}_{12}(z) &= \frac{1}{NT} \sum_{it} K_{it} K_{i(t-1)} \Delta X_{it} \times \\
&\left[\{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \{\mathcal{R}(Z_{it}; z) - \mathcal{R}(Z_{i(t-1)}; z)\} (Z_{i(t-1)} - z) \right].
\end{aligned} \tag{A.21}$$

We will show now that, as N tends to infinity

$$E \{\mathcal{E}_1(z)\} = o_p(\text{tr}\{H\}). \tag{A.22}$$

In order to prove this note that

$$\begin{aligned}
E \{\mathcal{E}_{11}(z)\} &= \int \int K(u) K(v) \left\{ \mathcal{B}_{\Delta XX}(z + H^{1/2}u, z + H^{1/2}v) \otimes (H^{1/2}u)^T \right\} \mathcal{R}(z + H^{1/2}u; z) (H^{1/2}u) dudv \\
&- \int \int K(u) K(v) \left\{ \mathcal{B}_{\Delta XX_{-1}}(z + H^{1/2}u, z + H^{1/2}v) \otimes (H^{1/2}v)^T \right\} \mathcal{R}(z + H^{1/2}u; z) (H^{1/2}v) dudv.
\end{aligned}$$

By (A. 5) and assumption 3.7

$$|\mathcal{R}_d(z + H^{1/2}u; z)| \leq \int_0^1 \varsigma(\omega \|H^{1/2}u\|) (1 - \omega) d\omega, \forall d,$$

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 , $\mathcal{B}_{\Delta XX}$ and $\mathcal{B}_{\Delta XX_{-1}}$

$$\begin{aligned} |E\{\mathcal{E}_{11}(z)\}| &\leq C_1 \int \int \int_0^1 |(H^{1/2}u)^T| \|\varsigma(\omega \|H^{1/2}u\|)\| \|H^{1/2}u\| K(u)K(v) d\omega dudv \\ &+ C_2 \int \int \int_0^1 |(H^{1/2}v)^T| \|\varsigma(\omega \|H^{1/2}u\|)\| \|H^{1/2}v\| K(u)K(v) d\omega dudv. \end{aligned}$$

And $E\{\mathcal{E}_{11}(z)\} = o_p(\text{tr}\{H\})$ follows by dominated convergence.

Similarly,

$$\begin{aligned} E\{\mathcal{E}_{12}(z)\} &= \int \int K(u)K(v) \left\{ \mathcal{B}_{\Delta XX_{-1}}(z + H^{1/2}u, z + H^{1/2}v) \otimes (H^{1/2}v)^T \right\} \times \\ &\left\{ \mathcal{R}(z + H^{1/2}u; z) - \mathcal{R}(z + H^{1/2}v; z) \right\} (H^{1/2}v) dudv. \end{aligned}$$

Therefore,

$$\begin{aligned} |E\{\mathcal{E}_{12}(z)\}| &\leq C_3 \int \int \int_0^1 |(H^{1/2}v)^T| \times \\ &|\varsigma(\omega \|H^{1/2}u\|)| \|H^{1/2}v\| K(u)K(v) d\omega dudv + C_4 \int \int \int_0^1 |(H^{1/2}v)^T| \times \\ &|\varsigma(\omega \|H^{1/2}v\|)| \|H^{1/2}v\| K(u)K(v) d\omega dudv. \end{aligned}$$

Then, proceeding as in the proof of the previous result we get also that $E\{\mathcal{E}_{12}(z)\} = o_p(\text{tr}\{H\})$.

Now, for $\mathcal{E}_2(z)$ note that

$$\mathcal{E}_2(z) = \mathcal{E}_{21}(z) + \mathcal{E}_{22}(z), \quad (\text{A. 23})$$

where

$$\begin{aligned} \mathcal{E}_{21}(z) &= \frac{1}{NT} \sum_{it} \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \times \\ &\left[\{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{i(t-1)} - z) \right] \end{aligned} \quad (\text{A. 24})$$

and

$$\begin{aligned} \mathcal{E}_{22}(z) &= \frac{1}{NT} \sum_{it} \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \times \\ &\left[\{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \{ \mathcal{R}(Z_{it}; z) - \mathcal{R}(Z_{i(t-1)}; z) \} (Z_{i(t-1)} - z) \right]. \end{aligned} \quad (\text{A. 25})$$

Following the same lines as for the proof of (A. 22) it is easy to show that

$$E\{\mathcal{E}_2(z)\} = o_p(H^{3/2}). \quad (\text{A. 26})$$

Substituting (A. 12), (A. 13), (A. 15) and (A. 22) into (A. 7), the asymptotic bias can be written

as

$$\begin{aligned}
& E \{ \hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(z) \\
&= \frac{1}{2} e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \{ S_{m_1}(z) - S_{m_2}(z) \} \\
&= \frac{1}{2} \mathcal{B}_{\Delta X \Delta X}^{-1}(z, z) [\mu_2(K_u) \mathcal{B}_{\Delta X X}(z, z) - \mu_2(K_v) \mathcal{B}_{\Delta X X_{-1}}(z, z)] \text{diag}_r \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \} i_d + o_p(\text{tr} \{ H \}).
\end{aligned}$$

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

$$V_{ij} = E(\Delta v_i \Delta v_j^T | X_{i1}, \dots, X_{iT}, Z_{i1}, \dots, Z_{iT}) = \begin{cases} 2\sigma_v^2, & \text{for } i = j, \quad t = s, \\ -\sigma_v^2, & \text{for } i = j, \quad |t - s| < 2, \\ 0, & \text{for } i = j, \quad |t - s| \geq 2. \end{cases} \quad (\text{A. 27})$$

Then, taking into account that

$$\hat{m}(z; H) - E \{ \hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta v, \quad (\text{A. 28})$$

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

$$\begin{aligned}
& \text{Var} \{ \hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} \\
&= e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \mathcal{V} W^T \tilde{Z} \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} e_1.
\end{aligned} \quad (\text{A. 29})$$

Based on assumption 3.2 and the fact that the v_{it} are i.i.d. then, the upper left entry of $\frac{1}{NT} \tilde{Z}^T W \mathcal{V} W^T \tilde{Z}$ is

$$\begin{aligned}
& \frac{2\sigma_v^2}{NT} \sum_{it} \Delta X_{it} \Delta X_{it}^T K_{it}^2 K_{i(t-1)}^2 - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \Delta X_{i(t-1)}^T K_{it} K_{i(t-1)}^2 K_{i(t-2)} \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \Delta X_{i(t-2)}^T K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} \\
&= \frac{2\sigma_v^2 R(K_u) R(K_v)}{|H|} \mathcal{B}_{\Delta X \Delta X}(z, z) \{1 + o_p(1)\},
\end{aligned} \quad (\text{A. 30})$$

because

$$\frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \Delta X_{i(t-1)}^T K_{it} K_{i(t-1)}^2 K_{i(t-2)} = \frac{\sigma_v^2 R(K_v)}{|H|^{1/2}} \mathcal{B}_{\Delta X \Delta X_{-1}}(z, z, z) \{1 + o_p(1)\} \quad (\text{A. 31})$$

and

$$\frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \Delta X_{i(t-2)}^T K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} = \sigma_v^2 R(K_v) \mathcal{B}_{\Delta X \Delta X_{-2}}(z, z, z, z) \{1 + o_p(1)\}. \quad (\text{A. 32})$$

The upper right block is

$$\begin{aligned}
& \frac{2\sigma_v^2}{NT} \sum_{it} \Delta X_{it} \left\{ X_{it}^T \otimes (Z_{it} - z)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T \right\} K_{it}^2 K_{i(t-1)}^2 \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \left\{ X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T - X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z)^T \right\} K_{it} K_{i(t-1)}^2 K_{i(t-2)} \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \left\{ X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z)^T - X_{i(t-3)}^T \otimes (Z_{i(t-3)} - z)^T \right\} K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} \\
& = \mathbf{I}_1 - \mathbf{I}_2 - \mathbf{I}_3.
\end{aligned} \tag{A. 33}$$

$$\begin{aligned}
\mathbf{I}_1 &= \frac{\sigma_v^2}{|H|} \int \left\{ \mathcal{B}_{\Delta XX} \left(z + H^{1/2}u, z + H^{1/2}v \right) \otimes \left(H^{1/2}u \right)^T \right. \\
& \quad \left. - \mathcal{B}_{\Delta XX-1} \left(z + H^{1/2}u, z + H^{1/2}v \right) \otimes \left(H^{1/2}v \right)^T \right\} K^2(u) K^2(v) dudv \{1 + o_p(1)\} \\
&= O_p(|H|),
\end{aligned} \tag{A. 34}$$

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

and

$$\begin{aligned}
\mathbf{I}_3 &= \sigma_v^2 \int \left\{ \mathcal{B}_{\Delta XX-2} \left(z + H^{1/2}u, z + H^{1/2}v, z + H^{1/2}w, z + H^{1/2}s \right) \otimes \left(H^{1/2}w \right)^T \right. \\
& \quad \left. - \mathcal{B}_{\Delta XX-3} \left(z + H^{1/2}u, z + H^{1/2}v, z + H^{1/2}w, z + H^{1/2}s \right) \otimes \left(H^{1/2}s \right)^T \right\} \\
& \quad \times K(u) K(v) K(w) K(s) dudvdwds \{1 + o_p(1)\} \\
&= O_p(1),
\end{aligned} \tag{A. 35}$$

where

$$\begin{aligned}
\frac{\mathcal{B}_{\Delta XX}(z, z)}{f_{Z_{it}, Z_{i(t-1)}}(z, z)} &= E \left[\Delta X_{it} X_{it}^T \mid Z_{it} = z, Z_{i(t-1)} = z \right], \\
\frac{\mathcal{B}_{\Delta XX-1}(z, z)}{f_{Z_{it}, Z_{i(t-1)}}(z, z)} &= E \left[\Delta X_{it} X_{i(t-1)}^T \mid Z_{it} = z, Z_{i(t-1)} = z \right], \\
\frac{\mathcal{B}_{\Delta XX-1}(z, z, z)}{f_{Z_{it}, Z_{i(t-1)}, Z_{i(t-2)}}(z, z, z)} &= E \left[\Delta X_{it} X_{i(t-1)}^T \mid Z_{it} = z, Z_{i(t-1)} = z, Z_{i(t-2)} = z \right], \\
\frac{\mathcal{B}_{\Delta XX-2}(z, z, z)}{f_{Z_{it}, Z_{i(t-1)}, Z_{i(t-2)}}(z, z, z)} &= E \left[\Delta X_{it} X_{i(t-2)}^T \mid Z_{it} = z, Z_{i(t-1)} = z, Z_{i(t-2)} = z \right], \\
\frac{\mathcal{B}_{\Delta XX-2}(z, z, z, z)}{f_{Z_{it}, Z_{i(t-1)}, Z_{i(t-2)}, Z_{i(t-3)}}(z, z, z, z)} &= E \left[\Delta X_{it} X_{i(t-2)}^T \mid Z_{it} = z, Z_{i(t-1)} = z, Z_{i(t-2)} = z, Z_{i(t-3)} = z \right], \\
\frac{\mathcal{B}_{\Delta XX-3}(z, z, z, z)}{f_{Z_{it}, Z_{i(t-1)}, Z_{i(t-2)}, Z_{i(t-3)}, Z_{i(t-4)}}(z, z, z, z)} &= E \left[\Delta X_{it} X_{i(t-3)}^T \mid Z_{it} = z, Z_{i(t-1)} = z, Z_{i(t-2)} = z, Z_{i(t-3)} = z, Z_{i(t-4)} = z \right].
\end{aligned}$$

Finally, the lower-right block is

$$\begin{aligned}
& \frac{2\sigma_v^2}{NT} \sum_{it} \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \{X_{it}^T \otimes (Z_{it} - z)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T\} K_{it}^2 K_{i(t-1)}^2 \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \\
& \times \{X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T - X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z)^T\} K_{it} K_{i(t-1)}^2 K_{i(t-2)} \\
& - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} \\
& \times \{X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z)^T - X_{i(t-3)}^T \otimes (Z_{i(t-3)} - z)^T\} K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} \\
& = \mathbf{I}_1 - \mathbf{I}_2 - \mathbf{I}_3,
\end{aligned} \tag{A. 36}$$

where

$$\begin{aligned}
\mathbf{I}_1 &= \frac{2\sigma_v^2 \mu_2 (K^2) R(K_v)}{|H|} \mathcal{B}_{XX}(z, z) \otimes H + \frac{2\sigma_v^2 \mu_2 (K^2) R(K_u)}{|H|} \mathcal{B}_{X_{-1}X_{-1}}(z, z) \otimes H + o_p(|H|^{-1} H), \\
\mathbf{I}_2 &= \frac{\sigma_v^2 \mu_2 (K^2)}{|H|^{1/2}} \mathcal{B}_{X_{-1}X_{-1}}(z, z, z) \otimes H + o_p(|H|^{-1/2} H), \\
\mathbf{I}_3 &= o_p(H).
\end{aligned}$$

So now, substituting (A. 12), (A. 30), (A. 33) and (A. 36) into (A. 29) we obtain

$$\text{Var} \{ \hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} = \frac{2\sigma_v^2 R(K_u) R(K_v)}{NT |H|} \mathcal{B}_{\Delta X \Delta X}^{-1}(z, z) \{1 + o_p(1)\}.$$

■

Proof of Theorem 3.2

Let

$$\begin{aligned}
\hat{m}(z; H) - m(z) &= \{ \hat{m}(z; H) - E[\hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}] \} \\
&+ \{ E[\hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}] - m(z) \} \equiv \mathbf{I}_1 + \mathbf{I}_2.
\end{aligned}$$

Now we show that

$$\sqrt{NT|H|} \mathbf{I}_1 \rightarrow \mathcal{N}(0, 2\sigma_v^2 R(K_u) R(K_v) \mathcal{B}_{\Delta X \Delta X}^{-1}(z, z)), \tag{A. 37}$$

as N tends to infinity.

In order to show this let

$$\hat{m}(z; H) - E\{ \hat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} = e_1^T \left(\tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta v, \tag{A. 38}$$

where $\Delta v = [\Delta v_{11}, \dots, \Delta v_{NT}]^T$. We are going to show the asymptotic normality of

$$\frac{1}{\sqrt{NT}} \tilde{Z}^T W \Delta v. \tag{A. 39}$$

Since (A. 39) is a multivariate vector we define a unit vector $d \in \mathbb{R}^{d(1+q)}$ in such a way that

$$\frac{1}{\sqrt{NT}} d^T \tilde{Z}^T W \Delta v = \frac{1}{\sqrt{NT}} \sum_i \sum_t \lambda_{it}, \tag{A. 40}$$

where

$$\lambda_{it} = |H|^{1/2} d^T \tilde{Z}_{it} K_{it} K_{i(t-1)} \Delta v_{it}, \quad i = 1, \dots, N \quad t = 2, \dots, T$$

and

$$\tilde{Z}_{it} = \begin{pmatrix} \Delta X_{it} \\ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \end{pmatrix}. \quad (\text{A. 41})$$

By Theorem 3.1 and conditions thereof we have that

$$\text{Var}(\lambda_{it}) = \quad (\text{A. 42})$$

$$2\sigma_v^2 d^T \begin{pmatrix} R(K_u) R(K_v) \mathcal{B}_{\Delta X \Delta X}(z, z) & 0 \\ 0 & \mu_2(K^2) R(K_v) \mathcal{B}_{XX}(z, z) \otimes H + \mu_2(K^2) R(K_u) \mathcal{B}_{X_{-1} X_{-1}}(z, z) \otimes H \end{pmatrix} \\ \times d \{1 + o_p(1)\}$$

and

$$\sum_t |\text{Cov}(\lambda_{i1}, \lambda_{it})| = o_p(1).$$

Define now, $\lambda_{n,i}^* = T^{-1/2} \sum_{t=1}^T \lambda_{it}$. For fixed T , the $\{\lambda_{n,i}^*\}$ are independent random variables. Therefore, to show (A. 37) it suffices to check Liapunov's condition. By Minkowski's inequality

$$E |\lambda_{n,i}^*|^{2+\delta} \leq CT^{\frac{2+\delta}{2}} E |\lambda_{it}|^{2+\delta}.$$

Because of (A. 41) we split λ_{it} into two components λ_{1it} and λ_{2it} and we analyze them by separate.

$$\begin{aligned} E |\lambda_{1it}|^{2+\delta} &\leq |H|^{\frac{2+\delta}{2}} E |d^T \Delta X_{it} K_{it} K_{i(t-1)} \Delta v_{it}|^{2+\delta} \\ &= |H|^{\frac{2+\delta}{2}} E \left[E \left\{ |d^T \Delta X_{it} \Delta v_{it}|^{2+\delta} \middle| Z_{it}, Z_{i(t-1)} \right\} K_{it}^{2+\delta} K_{i(t-1)}^{2+\delta} \right] \\ &= |H|^{-\delta/2} \int E \left\{ |d^T \Delta X_{it} \Delta v_{it}|^{2+\delta} \middle| Z_{it} = z + H^{1/2}u, Z_{i(t-1)} = z + H^{1/2}v \right\} \\ &\quad \times f_{Z_{it}, Z_{i(t-1)}}(z + H^{1/2}u, z + H^{1/2}v) K^{2+\delta}(u) K^{2+\delta}(v) dudv \\ &= |H|^{-\delta/2} E \left\{ |d^T \Delta X_{it} \Delta v_{it}|^{2+\delta} \middle| Z_{it} = z, Z_{i(t-1)} = z \right\} f_{Z_{it}, Z_{i(t-1)}}(z, z) \int K^{2+\delta}(u) K^{2+\delta}(v) dudv \\ &\quad + o_p(|H|^{-\delta/2}). \end{aligned}$$

$$\begin{aligned}
E |\lambda_{2it}|^{2+\delta} &\leq |H|^{\frac{2+\delta}{2}} E |d^T \{X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\} K_{it} K_{i(t-1)} \Delta v_{it}|^{2+\delta} \\
&\leq |H|^{\frac{2+\delta}{2}} E |d^T X_{it} \otimes (Z_{it} - z) K_{it} K_{i(t-1)} \Delta v_{it}|^{2+\delta} + |H|^{\frac{2+\delta}{2}} E |d^T X_{i(t-1)} \otimes (Z_{i(t-1)} - z) K_{it} K_{i(t-1)} \Delta v_{it}|^{2+\delta} \\
&= |H|^{\frac{2+\delta}{2}} E \left[E \left\{ |d^T X_{it} \Delta v_{it}|^{2+\delta} \middle| Z_{it}, Z_{i(t-1)} \right\} \otimes |Z_{it} - z|^{2+\delta} K_{it}^{2+\delta} K_{i(t-1)}^{2+\delta} \right] \\
&\quad + |H|^{\frac{2+\delta}{2}} E \left[E \left\{ |d^T X_{i(t-1)} \Delta v_{it}|^{2+\delta} \middle| Z_{it}, Z_{i(t-1)} \right\} \otimes |Z_{i(t-1)} - z|^{2+\delta} K_{it}^{2+\delta} K_{i(t-1)}^{2+\delta} \right] \\
&= |H| E \left\{ |d^T X_{it} \Delta v_{it}|^{2+\delta} \middle| Z_{it} = z, Z_{i(t-1)} = z \right\} f_{Z_{it}, Z_{i(t-1)}}(z, z) \otimes \int |u|^{2+\delta} K^{2+\delta}(u) K^{2+\delta}(v) dudv \\
&\quad + |H| E \left\{ |d^T X_{i(t-1)} \Delta v_{it}|^{2+\delta} \middle| Z_{it}, Z_{i(t-1)} \right\} f_{Z_{it}, Z_{i(t-1)}}(z, z) \otimes \int |v|^{2+\delta} K^{2+\delta}(u) K^{2+\delta}(v) dudv + o_p(|H|).
\end{aligned}$$

Therefore, $(NT)^{-\frac{2+\delta}{2}} \sum_{i=1}^N E |\lambda_{n,i}^*|^{2+\delta} \leq C (N|H|)^{-\delta/2}$. This indeed tends to zero when $N|H| \rightarrow \infty$ and therefore Lyapunov's condition holds and (A. 37) follows. \mathcal{D} has been already defined in (A. 42). Finally, using (A. 12) and applying the Cramer-Wold device the proof is done.

Using the bias expression computed in Theorem 3.1 we can then write

$$\begin{aligned}
&E \{ \widehat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(z) \\
&= \frac{1}{2} \mu_2 (K_u) \text{diag}_r \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \} i_d + O_p \left(H^{3/2} \right) + o_p(\text{tr} \{ H \}). \tag{A. 43}
\end{aligned}$$

Note that by the law of iterated expectations,

$$E \{ \widehat{m}(z; H) \} = \int E \{ \widehat{m}(z; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} dF(X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT}).$$

The leading term in (A. 43) does not depend on the sample and then, the proof is closed. ■

Proof of Theorem 3.3

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

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

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

$$E \left\{ \widetilde{m}(z; \widetilde{H}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} = e_1^T \left(\widetilde{Z}^{(1)T} W^{(1)} \widetilde{Z}^{(1)} \right)^{-1} \widetilde{Z}^{(1)T} W^{(1)} \left[M^{(1)} + M^{(2)} \right], \tag{A. 45}$$

where

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

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

$$M^{(1)} = \tilde{Z}^{(1)} \begin{bmatrix} m(z) \\ \text{vec}\{D_m(z)\} \end{bmatrix} + \frac{1}{2}Q_m^{(1)}(z) + R(z),$$

where

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

and

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

Using assumption 3.1 we get

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

and therefore,

$$\begin{aligned} & E \left\{ \tilde{m}(z; \tilde{H}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} - m(z) \\ &= e_1^T \left(\tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \left\{ \frac{1}{2} Q_m^{(1)}(z) + M^{(2)} \right\} + o_p \left(\text{tr} \{ \tilde{H} \} \right). \end{aligned} \quad (\text{A. 46})$$

To obtain an asymptotic expression for the bias we first calculate

$$\begin{aligned} & \frac{1}{NT} \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} = \\ & \left(\begin{array}{cc} (NT)^{-1} \sum_{it} X_{it} X_{it}^T K_{it} & (NT)^{-1} \sum_{it} X_{it} \{X_{it}^T \otimes (Z_{it} - z)^T\} K_{it} \\ (NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z)\} X_{it}^T K_{it} & (NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z)\} \{X_{it}^T \otimes (Z_{it} - z)^T\} K_{it} \end{array} \right). \end{aligned}$$

Using standard properties of kernel density estimators, under conditions 3.1 to 3.9 and as N tends to infinity,

$$\begin{aligned} & (NT)^{-1} \sum_{it} X_{it} X_{it}^T K_{it} = \mathcal{B}_{XX}(z) + o_p(1), \\ & (NT)^{-1} \sum_{it} X_{it} \{X_{it}^T \otimes (Z_{it} - z)^T\} K_{it} = \mathcal{DB}_{XX}(z) \left(I_d \otimes \mu_2(K_u) \tilde{H} \right) + o_p(\tilde{H}), \\ & (NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z)\} \{X_{it}^T \otimes (Z_{it} - z)^T\} K_{it} = \mathcal{B}_{XX}(z) \otimes \mu_2(K_u) \tilde{H} + o_p(\tilde{H}). \end{aligned}$$

Note that $\mathcal{B}_{XX}(z)$ and $\mathcal{DB}_{XX}(z)$ are defined as in the proof of Theorem 3.1 but the moment functions now are taken conditionally only to $Z_{it} = z$.

Using the previous results,

$$NT \left(\tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} = \begin{pmatrix} \mathcal{C}_{11}^{(1)} & \mathcal{C}_{12}^{(1)} \\ \mathcal{C}_{21}^{(1)} & \mathcal{C}_{22}^{(1)} \end{pmatrix}, \quad (\text{A. 47})$$

where

$$\begin{aligned} \mathcal{C}_{11}^{(1)} &= \mathcal{B}_{XX}^{-1}(z) + o_p(1), \\ \mathcal{C}_{12}^{(1)} &= -\mathcal{B}_{XX}^{-1}(z) [\mathcal{DB}_{XX}(z)] (\mathcal{B}_{XX}^{-1}(z) \otimes I_q) + o_p(1), \\ \mathcal{C}_{22}^{(1)} &= \left(\mathcal{B}_{XX}(z) \otimes \mu_2(K_u) \tilde{H} \right)^{-1} + o_p(\tilde{H}^{-1}). \end{aligned}$$

Furthermore the terms in

$$(NT)^{-1} \tilde{Z}^{(1)T} W^{(1)} Q_m^{(1)}(z) = \quad (\text{A. 48})$$

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

are of order

$$\mu_2(K_u) E[X_{it} X_{it}^T | Z_t = z] f_{Z_t}(z) \times \text{diag}_r \{ \text{tr} \{ \mathcal{H}_{md}(z) \tilde{H} \} \} i_d + o_p \left(\text{tr} \{ \tilde{H} \} \right)$$

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

$$(NT)^{-1} \tilde{Z}^{(1)T} W^{(1)} M^{(2)} = \quad (\text{A. 49})$$

$$\begin{pmatrix} (NT)^{-1} \sum_{it} X_{it} X_{i(t-1)}^T (E \{ \hat{m}(Z_{i(t-1)}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(Z_{i(t-1)})) K_{it} \\ (NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z)\} X_{i(t-1)}^T (E \{ \hat{m}(Z_{i(t-1)}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(Z_{i(t-1)})) K_{it} \end{pmatrix}.$$

It is straightforward to show that

$$(NT)^{-1} \sum_{it} X_{it} X_{i(t-1)}^T (E \{ \hat{m}(Z_{i(t-1)}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(Z_{i(t-1)})) K_{it} = o_p \left(\text{tr} \{ \tilde{H} \} \right),$$

as N tends to infinity, and

$$\begin{aligned} (NT)^{-1} \sum_{it} \{X_{it} \otimes (Z_{it} - z)\} X_{i(t-1)}^T (E \{ \hat{m}(Z_{i(t-1)}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \} - m(Z_{i(t-1)})) K_{it} \\ = o_p \left(\text{tr} \{ H \} \text{tr} \{ \tilde{H} \} \right), \end{aligned}$$

as N tends to infinity. Under assumptions 3.1 to 3.9, the bias is $o_p(\text{tr}\{H\})$, and the rate is uniform in z (see Masry (1996) for details).

Now substitute the asymptotic expressions for (A. 47), (A. 48) and (A. 49) into (A. 46) apply that $\text{tr}\{H\} \rightarrow 0$ $\text{tr}\{\tilde{H}\} \rightarrow 0$ in such a way that $N|H| \rightarrow \infty$, $N|\tilde{H}| \rightarrow \infty$ and we have shown that the asymptotic bias in $\tilde{m}(z; \tilde{H})$ is of the same order as it was in the first step.

For the variance term, recall that

$$\begin{aligned} \tilde{m}(z; \tilde{H}) - E \left\{ \tilde{m}(z; \tilde{H}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} = e_1^T \left(\tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \Delta v \\ + e_1^T \left(\tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \hat{v}, \end{aligned}$$

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

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

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

$$r(Z_{i(t-1)}; H) = \hat{m}(Z_{i(t-1)}; H) - E \{ \hat{m}(Z_{i(t-1)}; H) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \},$$

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

Then, the variance of $\tilde{m}(z; \tilde{H})$ takes the form

$$\text{Var} \left\{ \tilde{m}(z; \tilde{H}) | X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} \quad (\text{A. 50})$$

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

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

$$\mathbf{I}_1 = \frac{2\sigma_v^2 R(K_u)}{NT \left| \tilde{H} \right|^{1/2}} \mathcal{B}_{XX}^{-1}(z) \{1 + o_p(1)\}. \quad (\text{A. 51})$$

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

$$\frac{1}{NT} \tilde{Z}^{(1)T} W^{(1)} E \left\{ \widehat{v} \widehat{v}^T \mid X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} W^{(1)T} \tilde{Z}^{(1)}. \quad (\text{A. 52})$$

The upper left entry is

$$(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T E \left\{ r(Z_{i(t-1)}; H) r(Z_{i(s-1)}; H)^T \mid X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} X_{i(s-1)} X_{is}^T K_{it} K_{is}. \quad (\text{A. 53})$$

Applying the Cauchy-Schwarz inequality for covariance matrices then (A. 53) is bounded by

$$\begin{aligned}
&(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T \text{vec}^{1/2} \left\{ \text{diag} \left(E \left\{ r(Z_{i(t-1)}; H) r(Z_{i(t-1)}; H)^T \mid X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} \right) \right\} \\
&\times \text{vec}^{1/2} \left\{ \text{diag} \left(E \left\{ r(Z_{i(s-1)}; H) r(Z_{i(s-1)}; H)^T \mid X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} \right) \right\}^T X_{i(s-1)} X_{is}^T K_{it} K_{is}.
\end{aligned}$$

Now, note that under the conditions of the Theorem 3.1

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

uniformly in z , and therefore (A. 53) is of order $O_p \left(\frac{\log NT}{NT |H| |\tilde{H}|^{1/2}} \right)$.

Following the same lines, it is easy to show that the upper right entry of (A. 52) is

$$\begin{aligned}
&(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T E \left\{ r(Z_{i(t-1)}; H) r(Z_{i(s-1)}; H)^T \mid X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} \\
&\times X_{i(s-1)} \{X_{is} \otimes (Z_{is} - z)\}^T K_{it} K_{is} = o_p \left(\frac{\log NT}{NT |H| |\tilde{H}|^{1/2}} \right),
\end{aligned}$$

and finally the lower right entry of (A. 52) is

$$\begin{aligned}
&(NT)^{-1} \sum_i \sum_{ts} \{X_{it} \otimes (Z_{it} - z)\} X_{i(t-1)}^T E \left\{ r(Z_{i(t-1)}; H) r(Z_{i(s-1)}; H)^T \mid X_{11}, \dots, X_{NT}, Z_{11}, \dots, Z_{NT} \right\} \\
&\times X_{i(s-1)} \{X_{is} \otimes (Z_{is} - z)\}^T K_{it} K_{is} = O_p \left(\frac{\log NT}{NT |H| |\tilde{H}|^{1/2}} \right).
\end{aligned}$$

Now, combining results in (A. 47) and (A. 52) we show that $\mathbf{I}_2 = o_p \left(\frac{\log NT}{NT |H| |\tilde{H}|^{1/2}} \right)$. Finally

a standard Cauchy-Schwarz inequality is enough to show that $\mathbf{I}_3 = o_p\left(\sqrt{\frac{\log NT}{NT|H|\tilde{H}}}\right)$ and then the proof of the result is closed. ■