Direct Semiparametric Estimation of Fixed Effects Panel Data Varying Coefficient Models

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

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

August 5, 2013

Abstract

In this paper we present a new technique to estimate varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in a unknown way. The estimator is based on 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. [1]

JEL codes: C14, C23
Key Words: Varying coefficients model, fixed effects, panel data, local linear regression, oracle efficient estimator.

[1]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.
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 rends 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 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 $\mu_i$ is arbitrarily correlated with the covariates $X_{it}$ and/or $Z_{it}$:

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

Furthermore, the function $m(z)$ is unknown and needs to be estimated and the $\nu_{it}$ are random errors. It is clear that, any attempt to estimate directly $m(\cdot)$ through standard nonparametric estimation techniques will render 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 $\mu_i$ from (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.

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

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 $E(\Delta Y_{it} | Z_{it}, Z_{it(t-1)}, X_{it}, X_{it(t-1)})$ in (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 $X^T_{it} m(Z_{it}) \equiv m(X_{it}, Z_{it})$, (2) becomes a fully nonparametric additive model

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

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 $X^T_{it} m(Z_{it}) \equiv g(Z_{it}) + \tilde{X}_{it}^T \beta_0$, where $X_{it} = \begin{pmatrix} 1, \tilde{X}_{it} \end{pmatrix}^T$ and $m(Z_{it}) = (g(Z_{it}), \beta_0)^T$ for some real valued $g(\cdot)$ and a vector $\beta_0$, the regression function in (2) becomes a semiparametric partially additive model

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

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) in a framework where $N \to \infty$ but $T$ remains fixed. It is based in applying a local approximation to the additive function $X^T_{it} m(Z_{it}) - X^T_{i(t-1)} m(Z_{i(t-1)})$. 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 lends 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), 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
where \( d_{ij} = 1 \) if \( i = j \) and 0 otherwise. Based in this model they propose a least-squared 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 term results from the unknown fixed effects and it is zero only in the case that they add the additional (strong) restriction that \( \sum \mu_i = 0 \). 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 (2) with \( d = q = 1 \). In this case, for any \( z \in A \), where \( A \) is a compact subsect in a nonempty interior of \( \mathbb{R} \), one has the following Taylor expansion

\[
X_{it} m(Z_{it}) - X_{it-1} m(Z_{it-1}) \approx m(z) \Delta X_{it} + m'(z) \left[ X_{it} (Z_{it} - z) - X_{i(t-1)} (Z_{i(t-1)} - z) \right] + m''(z) \left[ X_{it} (Z_{it} - z)^2 - X_{i(t-1)} (Z_{i(t-1)} - z)^2 \right] + \cdots + m^{(p)}(z) \left[ X_{it} (Z_{it} - z)^p - X_{i(t-1)} (Z_{i(t-1)} - z)^p \right],
\]

which suggests that one estimates \( m(z), m'(z), \ldots, m^{(p)}(z) \) by regressing the \( \Delta Y_{it} \)'s on the terms \( X_{it} (Z_{it} - z)^{\lambda} - X_{i(t-1)} (Z_{i(t-1)} - z)^{\lambda} \), \( \lambda = 0, 1, \cdots \), with kernel weights. Then, the quantities of interest can be estimated using a local linear regression estimator (see Fan and Gijbels (1995b), Ruppert and Wand (1994) or Zhan-Qian (1996)),

\[
\text{Min}_{\alpha, \delta} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it} - \alpha \Delta X_{it} - \delta \left[ X_{it} (Z_{it} - z) - X_{i(t-1)} (Z_{i(t-1)} - z) \right] \right\}^2 K_h (Z_{it} - z) K_h (Z_{i(t-1)} - z)
\]

(4)

where \( \alpha = m(z) \) and \( \delta = m'(z) \). \( K \) is a bivariate kernel such that \( K(u, v) = K(u)K(v) \), where for each \( u, v \),

\[
\int K(u) du = 1 \quad \text{and} \quad K_h(u) = \frac{1}{h} K \left( \frac{u}{h} \right),
\]
where $h$ is a bandwidth. For the case of a local constant approximation (i.e. Naradaya-Watson kernel regression estimator) $m(z)$ has the following closed form

$$
\hat{m}_h(z) = \sum_{i=1}^{N} \sum_{t=2}^{T} K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z) \Delta X_{it} \Delta Y_{it} \over \sum_{i=1}^{N} \sum_{t=2}^{T} K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z) \Delta X_{it}^2,
$$

and in the local linear regression case we have

$$
\begin{pmatrix} \hat{m}_h(z) \\ \hat{m}_h'(z) \end{pmatrix} = \left( \sum_{it} K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z) \tilde{Z}_{it} \tilde{Z}_{it}^T \right)^{-1} \sum_{it} K_h(Z_{it} - z) K_h(Z_{i(t-1)} - z) \tilde{Z}_{it} \Delta Y_{it}
$$

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

$$
\tilde{Z}_{it}^T = \begin{pmatrix} \Delta X_{it} & X_{it}(Z_{it} - z) - X_{i(t-1)}(Z_{i(t-1)} - z) \end{pmatrix}, \quad i = 1, \cdots, N; t = 2, \cdots, T.
$$

Note that in (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 (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 $O(\frac{1}{NTh})$. 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 $\sqrt{NTh}$.

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

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

By substituting (2) into (8) we get

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

As it can be realized from (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 (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

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

where

$$
\nu_{it} = \{ \hat{m}_h(Z_{i(t-1)}) - m(Z_{i(t-1)}) \} X_{i(t-1)} + \Delta v_{it}, \quad i = 1, \cdots, N; t = 2, \cdots, T.
$$
By doing so, we can estimate \( m(\cdot) \) through the following optimization problem:

\[
\text{Min}_\alpha, \delta \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it}^{(1)} - \alpha X_{it} - \delta X_{it} (Z_{it} - z) \right\}^2 K_H (Z_{it} - z). \tag{11}
\]

The solution for \( \alpha = m(z) \) and \( \delta = m'(z) \) in (11) are the oracle efficient local linear least squares estimators. This is shown in Theorem 3.3.

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). In this case, the local linear estimator of the function \( m(\cdot) \) and its partial derivatives is the solution to,

\[
\text{Min}_\beta \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it} - \tilde{Z}_{it}^T \beta \right\}^2 K_H (Z_{it} - z) K_H (Z_{i(t-1)} - z) \tag{12}
\]

where \( \beta = [\alpha^T, \delta^T]^T \) is now a \( d(1+q) \times 1 \) vector, \( \alpha = m(z) \) is a \( d \times 1 \) vector, \( \delta = \text{vec} \{ D_m(z) \} \) is a \( dq \times 1 \) vector, \( D_m(z) = \frac{\partial m(z)}{\partial z} \) is a \( d \times q \) matrix of first derivatives of \( m(\cdot) \) and

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

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

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

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

It is easy to verify that the solution to the minimization problem in (12) can be written in matrix form as

\[
\left[ \hat{\alpha} \ \hat{\delta} \right] = \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta Y, \tag{13}
\]

where

\[
W = \text{diag} \left\{ K_H (Z_{12} - z) K_H (Z_{11} - z), ..., K_H (Z_{NT} - z) K_H (Z_{N(T-1)} - z) \right\},
\]

\[
\Delta Y = \left[ \Delta Y_{12}, ..., \Delta Y_{NT} \right]^T,
\]

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

\[
\hat{m}(z; H) = e_1^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta Y, \tag{14}
\]

where \( e_1 = (I_d \otimes 0_{dq \times d}) \) 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 (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 ammount and direction of the weights. Often, instead of taking a matrix with 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 (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

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

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

(A.3) \( \mu_i \) can be arbitrarily correlated with both \( X_{it} \) and \( Z_{it} \) with unknown correlation structure.

(A.4) Let \( \|A\| = \sqrt{tr(A^T A)} \), then \( E\left\{\|X_{it}X_{it}^T\|^2 | Z_{it} = z_1, Z_{i(t-1)} = z_2\right\} \) is bounded and uniformly continuous in its support. Furthermore, let \( \mathcal{X}_{it} = \left(X_{it}^T \ X_{i(t-1)}^T\right)^T \) and \( \Delta \mathcal{X}_{it} = \left(\Delta X_{it}^T \Delta X_{i(t-1)}^T\right)^T \).
The matrix functions \( E \left( X_{it}'X_{it}^T \mid Z_{it} = z_1, Z_{it(t-1)} = z_2 \right) \), \( E \left( X_{it}'X_{it}^T \mid Z_{it} = z_1, Z_{it(t-1)} = z_2, Z_{it(t-3)} = z_3 \right) \), \( E \left( \Delta X_{it} \Delta X_{it}^T \mid Z_{it} = z_1, Z_{it(t-1)} = z_2 \right) \) and \( E \left( X_{it}' \Delta X_{it}^T \mid Z_{it} = z_1, Z_{it(t-1)} = z_2, Z_{it(t-2)} = z_3 \right) \) are bounded and uniformly continuous in their support.

(A.5) The function \( E \left[ \Delta X_{it} \Delta X_{it}^T \mid Z_{it} = z_1, Z_{it(t-1)} = z_2 \right] \) is positive definite for any interior point of \((z_1, z_2)\) in the support of \( f_{Z_{it},Z_{it(t-1)}} (z_1, z_2) \).

(A.6) The functions \( E \left\{ |X_{it} \Delta v_{it}|^{2+\delta} \mid Z_{it} = z_1, Z_{it(t-1)} = z_2 \right\} \), \( E \left\{ |X_{it(t-1)} \Delta v_{it}|^{2+\delta} \mid Z_{it} = z_1, Z_{it(t-1)} = z_2 \right\} \), and \( E \left\{ |\Delta X_{it} \Delta v_{it}|^{2+\delta} \mid Z_{it} = z_1, Z_{it(t-1)} = z_2 \right\} \), for some \( \delta > 0 \), are bounded and uniformly continuous in any point of their support.

(A.7) Let \( z \) be an interior point in the support of \( f_{Z_{it}} \). All second-order derivatives of \( m_1 (\cdot), m_2 (\cdot), \ldots, m_d (\cdot) \) are bounded and uniformly continuous.

(A.8) The Kernel functions \( K \) are compactly supported, bounded kernel such that \( \int uu^TK(u)du = \mu_2(K)I \) 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 \( \int u_1^i \cdots u_q^q K(u)du = 0 \), for all nonnegative integers \( i_1, \ldots, i_q \) such that their sum is odd.

(A.9) The bandwidth matrix \( H \) 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 \( N \|H\| \to \infty \).

As the reader may notice, all assumptions are rather standard in nonparametric regression analysis of panel data models. Assumption (A.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 cero at a faster rate. In our case this was no needed because our asymptotic analysis is performed for fixed \( T \).

On the other side, nonstationary proceses 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 (A.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 (A.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.
The variance is the estimator, we give a result for the case when obtain
Assume conditions (A.1)-(A.9) hold, then as Theorem 3.1 variance of the local linear least squares estimator. 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 (A.1)-(A.9) hold, then as \( N \) tends to infinity and \( T \) is fixed we obtain

\[
E \{ \hat{m}(z; H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} - m(z) = \frac{1}{2} \sum_{t=1}^{T} \left( \tilde{Z}^{T}W\tilde{Z} \right)^{-1} \tilde{Z}^{T}W \{ S_{m_{1}}(z) - S_{m_{2}}(z) \}
\]

\[
= \frac{1}{2} \sigma_{1}^{2} \left( \mu_{2}(K_{u})B_{\Delta XX}(z,z) - \mu_{2}(K_{v})B_{\Delta XX-1}(z,z) \right) \text{diag} \{ tr \{ H_{m_{r}}(z; H) \} \} i_{d} + o_{p}(tr \{ H \}),
\]

where for \( r = 1,\ldots,d \), \( H_{m_{r}}(z) \) is the Hessian matrix of the \( r \)th component of \( m(\cdot) \), while for \( \ell = 1,2 \) the \( \ell \)th element of \( S_{m_{d}} \) is

\[
\{ X_{i(t+1-\ell)} \otimes (Z_{i(t+1-\ell)} - z) \}^{T} H_{m}(z)(Z_{i(t+1-\ell)} - z).
\]

Furthermore, if \( \mu_{2}(K_{u}) = \mu_{2}(K_{v}) \) the bias term becomes

\[
E \{ \hat{m}(z; H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} - m(z) = \frac{1}{2} \mu_{2}(K_{u}) \text{diag} \{ tr \{ H_{m_{r}}(z; H) \} \} i_{d} + o_{p}(tr \{ H \}).
\]

The variance is

\[
\text{Var} \{ \hat{m}(z; H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} = \frac{2\sigma_{1}^{2}R(K_{u})R(K_{v})}{NT|H|}B_{\Delta XX}^{-1}(z,z)\{1 + o_{p}(1)\},
\]

where

\[
B_{\Delta XX}(z,z) = E \left[ \Delta X_{it}X_{it}^{T} | Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it},Z_{i(t-1)}}(z,z),
\]

\[
B_{\Delta XX-1}(z,z) = E \left[ \Delta X_{it}X_{it}^{T} | Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it},Z_{i(t-1)}}(z,z),
\]

\[
B_{\Delta XX}(z,z) = E \left[ \Delta X_{it}\Delta X_{it}^{T} | Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it},Z_{i(t-1)}}(z,z),
\]

\[
\text{diag} \{ tr \{ H_{m_{r}}(z; H) \} \} \text{ stands for a diagonal matrix of elements } tr \{ H_{m_{r}}(z; H) \}, \text{ for } r = 1,\ldots,d, \text{ and } i_{d} \text{ is a } d \times 1 \text{ unit vector.}
\]

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 (A.1)-(A.8) hold, then, if $h \to 0$ in such a way that $N h^2 \to \infty$ as $N$ tends to infinity and $T$ is fixed we get

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

where

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

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

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

The variance is

$$Var \{ \hat{m}(z; H) | X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \} = \frac{2 \sigma_v^2 R(K_u) R(K_v)}{N T h^2 f_{Z_{it}, Z_{i(t-1)}}(z, z) E \left[ \left( X_{it} - X_{i(t-1)} \right)^2 | Z_{it} = z, Z_{i(t-1)} = z \right]} \{ 1 + o_p(1) \}. \]$$

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 $\hat{m}(z; H)$:

**Theorem 3.2** Assume conditions (A.1)-(A.9) hold. Then,

$$\sqrt{NT |H|} \{ \hat{m}(z; H) - m(z) \} \to_d N(b(z), v(z)), \]$$

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

$$b(z) = \frac{1}{2} \mu_2(K_u) \text{diag} \left\{ \text{tr} \left\{ H_{md}(z) H \sqrt{NT |H|} \right\} \right\} \text{id}, \]$$

$$v(z) = 2 \sigma_v^2 R(K_u) R(K_v) B_{\Delta X \Delta X}^{-1}(z, z). \]$$

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 $\Delta Y_{it}$ as in (10), now for the multivariate case,

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

where

$$\nu_{it} = X_{i(t-1)}^T [\hat{m}(Z_{i(t-1)}; H) - m(Z_{i(t-1)})] + \Delta \nu_{it}, \quad i = 1, \cdots, N; \quad t = 2, \cdots, T. \quad (15)$$

In expression (15), $\hat{m}(Z_{i(t-1)}; H)$ is the first-step local linear estimator obtained in (14). Now we propose to estimate $m(Z_{it})$ as the solution to the following problem

$$\text{Min}_{\alpha, \beta} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta \tilde{Y}_{it}^{(1)} - (X_{it}^T \alpha + X_{it}^T \otimes (Z_{it} - z)^T \delta) \right\}^2 K_{\tilde{H}}(Z_{it} - z), \quad (16)$$

where $\beta = [\alpha^T, \delta^T]^T$ a $d(1 + q) \times 1$ vector. If we define $\tilde{Z}_{it}^{(1)} = [X_{it}^T, X_{it}^T \otimes (Z_{it} - z)^T]$, a $1 \times d(1 + q)$ vector, then (16) can be written as

$$\text{Min}_{\beta} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ \Delta Y_{it}^{(1)} - \tilde{Z}_{it}^{(1)} \beta \right\}^2 K_{\tilde{H}}(Z_{it} - z), \quad (17)$$

and the solution for $\alpha = m(z)$ is

$$\hat{m}^{(1)}(z; \tilde{H}) = e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \Delta \tilde{Y}^{(1)}, \quad (18)$$

where $\Delta \tilde{Y}^{(1)} = \left[ \Delta \tilde{Y}_{i2}^{(1)}, \cdots, \Delta \tilde{Y}_{NT}^{(1)} \right]^T$, $W^{(1)} = \text{diag}\{K_H(Z_{i2} - z), \ldots, K_H(Z_{NT} - z)\}$ and

$$\tilde{Z}^{(1)} = \begin{bmatrix} \Delta X_{i2}^T & X_{i2}^T \otimes (Z_{i2} - z)^T \\ \vdots & \vdots \\ \Delta X_{NT}^T & X_{NT}^T \otimes (Z_{NT} - z)^T \end{bmatrix}.$$

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 $H$:

(B.1) 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 $N |\tilde{H}| \to \infty$.

(B.2) The bandwidth matrices $H$ and $\tilde{H}$ must fulfill that $N |H| |\tilde{H}| / \log(N) \to \infty$, and $\text{tr}(H)/\text{tr}(\tilde{H}) \to 0$ 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 (A.1) to (A.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 (A.1)-(A.8) and (B.1)-(B.2) holds, then, as $N$ tends to infinity and $T$ is fixed we get
\[
E\left\{ \hat{m}^{(1)}(z; \tilde{H}) | X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} - m(z) = \frac{1}{2} \mu_2(K_u) \text{diag}_r \left\{ \text{tr} \left\{ \mathcal{H}_{m_r}(z) \tilde{H} \right\} \right\} i_d + o_p \left( \text{tr} \left\{ \tilde{H} \right\} \right).
\]
and
\[
\text{Var} \left\{ \hat{m}^{(1)}(z; H) | X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} = \frac{2\sigma^2 R(K_u)}{NT} \left[ \frac{1}{2} \right] \mathcal{B}^{-1}_{XX}(z) \{1 + o_p(1)\},
\]
where $\text{diag}_r \{ \text{tr} \left\{ \mathcal{H}_{m_r}(z) H \right\} \}$ stands for a diagonal matrix of elements $\text{tr} \left\{ \mathcal{H}_{m_r}(z) H \right\}$, for $r = 1, \ldots, d$, and $i_d$ 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 $1/NT h^2$). This trade off should be solved by choosing a bandwidth matrix $H$ that minimizes simultaneously both quantities, 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) we propose the following measure of discrepancy,
\[
MSE(H) = E \left[ X^T (\hat{m}(Z; H) - m(Z)) \right]^2.
\]
In this mean squared error, the expectation is taken over $Z_1, \ldots, Z_q; X_1, \ldots, X_d$ and $\hat{m}(Z; H)$ is the estimator defined in (14). Therefore, for our problem, we can define the optimal bandwidth matrix $H_{opt}$ as the solution to the following minimization problem,
\[
H_{opt} = \arg \min_H \text{MSE}(H) = \arg \min_H \text{MSE}(H) = \arg \min_H \left[ X^T (\hat{m}(Z; H) - m(Z)) \right]^2.
\]
If \( Z_1, \ldots, Z_q; \ X_1, \ldots, X_d \) are random variables that are independent of the observed sample \( \mathcal{D} = (X_{11}, Z_{11}, \ldots, X_{NT}, Z_{NT})^T \), but they share the same distribution with \((X_{11}, Z_{11})\) it is straightforward to show that

\[
MSE(H) = \mathbb{E} \left[ b^T(Z) \Omega(Z)b(Z) + \text{tr} \{ \Omega(Z) V(Z) \} \right],
\]

where

\[
b(Z) = \mathbb{E} \{ \hat{m}(Z; H) \mid \mathcal{D}, Z \} - m(Z),
\]

\[
V(Z) = \text{Var} \{ \hat{m}(Z; H) \mid \mathcal{D}, Z \}, \quad \text{and}
\]

\[
\Omega(Z) = \mathbb{E} (XX^T \mid Z).
\]

As it can be realized from the expression above, it has been now formalized the idea of choosing a bandwith matrix \( H \) that simultaneously minimizes both (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. \( H(z_0) \). In this case, the local MSE criteria would be

\[
MSE(z_0; H) = \mathbb{E} \left[ X^T (\hat{m}(z_0; H) - m(z_0)) \right]^2,
\]

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 bandwith 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 (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 (19) bias and variance by their exact expressions. That is

\[
E \{ \hat{m}(Z; H) \mid \mathcal{D}, Z \} - m(Z) = \left| \mathbb{E} \{ \hat{m}(z; H) \mid \mathcal{D} \} - m(z) \right|_{z = Z} \quad (20)
\]

\[
\text{Var} \{ \hat{m}(Z; H) \mid \mathcal{D}, Z \} = \left| \text{Var} \{ \hat{m}(z; H) \mid \mathcal{D} \} \right|_{z = Z}, \quad (21)
\]

where clearly, according to Theorem 3.1

\[
E \{ \hat{m}(z; H) \mid \mathcal{D} \} - m(z) = e_1^T \left( \bar{Z}^T W \bar{Z} \right)^{-1} \bar{Z}^T W \tau
\]

\[
\text{Var} \{ \hat{m}(z; H) \mid \mathcal{D} \} = e_1^T \left( \bar{Z}^T W \bar{Z} \right)^{-1} \bar{Z}^T W \Sigma W \bar{Z} \left( \bar{Z}^T W \bar{Z} \right)^{-1} e_1.
\]
The corresponding estimator of the $MSE(H)$, according with (19) will be

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

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

$$\hat{H}_{opt} = \arg \min_H \tilde{MSE}(H).$$
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 \( \mu_i \) are fixed effects.

We consider the following varying-coefficient nonparametric models,

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

where \( X_{dit} \) and \( Z_{qit} \) are scalars random variables; \( v_{it} \) is an i.i.d. \( N(0,1) \) random variable; and \( \hat{m}(.) \) is a pre-specified function to be estimated. The observations follow a data generating process where \( Z_{qit} = w_{qit} + w_{qit(t-1)} \), being \( w_{qit} \) an i.i.d. uniformly distributed \([0, \Pi/2]\) random variable; and \( X_{dit} = 0.5X_{dit(t-1)} + \xi_{it} \), with \( \xi_{it} \) 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 \( m_2 (Z_{2it}) = \exp (-Z_{2it}^2) \); and we experiment with two specifications for the fixed effects,

(a) \( \mu_{1i} \) depends on \( Z_{1it} \), where the dependence is imposed by generating \( \mu_{1i} = c_0 Z_{1i} + u_i \) for \( i = 2, \ldots, N \) and \( Z_{1i} = T^{-1} \sum_t Z_{1it} \).

(b) \( \mu_{2i} \) depends on \( Z_{1it}Z_{2it} \) through the generating process \( \mu_{2i} = c_0 Z_i + u_i \) for \( i = 2, \ldots, N \) and \( Z_i = 1/2 (Z_{1i} + Z_{2i}) \).

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

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

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

\[
MSE \{\hat{m}_r(z; H)\} = \frac{1}{Q} \sum_{l=1}^{Q} E \left[ \left( \sum_{r=1}^{d} \left( \hat{m}_{lr}(z; H) - m_{lr}(z) \right) X_{it lr} \right)^2 \right]
\]
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 simulations results are summarized in Table 1, 2 and 3, respectively.

**Table 1:** AMSE for \(d=1\) and \(q=1\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>Local polynomial</th>
<th>Backfitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>1.2917346</td>
<td>1.2862042</td>
</tr>
<tr>
<td>300</td>
<td>0.9156186</td>
<td>0.915688</td>
</tr>
<tr>
<td>600</td>
<td>0.7275608</td>
<td>0.7136175</td>
</tr>
</tbody>
</table>

**Table 2:** AMSE for \(d=1\) and \(q=2\).

<table>
<thead>
<tr>
<th>(n)</th>
<th>Local polynomial</th>
<th>Backfitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>1.7953523</td>
<td>1.3868033</td>
</tr>
<tr>
<td>300</td>
<td>1.1077050</td>
<td>1.0171259</td>
</tr>
<tr>
<td>600</td>
<td>0.8181966</td>
<td>0.7658107</td>
</tr>
</tbody>
</table>

**Table 3:** AMSE for \(d=2\) and \(q=1\).

<table>
<thead>
<tr>
<th>(m_1(z))</th>
<th>(m_2(z))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local polynomial</td>
<td>Backfitting</td>
</tr>
<tr>
<td>150</td>
<td>1.3454139</td>
</tr>
<tr>
<td>300</td>
<td>0.9231963</td>
</tr>
<tr>
<td>600</td>
<td>0.7305442</td>
</tr>
</tbody>
</table>

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

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

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

This paper introduces a new technique that estimates varying coefficient models of unknown form in a panel data framework where individual effects are arbitrarily correlated with the explanatory variables in 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.

7 Appendix

Proof of Theorem 3.1

Taking assumption (A.2) conditional expectations in (14) and noting that

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

then

\[ E\{\hat{m}(z;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} = e_r^T(\tilde{Z}^TW\tilde{Z})^{-1}\tilde{Z}^TWM \]

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

Taylor’s Theorem implies that

\[ M = \tilde{Z}\begin{bmatrix} m(z) \\ \text{vec}\{D_m(z)\} \end{bmatrix} + \frac{1}{2}Q_m(z) + R(z), \quad (27) \]

where

\[ Q_m(z) = S_{m1}(z) - S_{m2}(z), \quad (28) \]

\[ S_{m1}(z) = [S_{m1,12}(z),\cdots,S_{m1,NT}(z)]^T, \]

\[ S_{m2}(z) = [S_{m2,11}(z),\cdots,S_{m2,N(T-1)}(z)]^T \]

and

\[ S_{m1,it}(z) = \left[\left\{X_{it} \otimes (Z_{it} - z)\right\}^T H_m(z)(Z_{it} - z)\right], \]

\[ S_{m2,i(t-1)}(z) = \left[\left\{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\right\}^T H_m(z)(Z_{i(t-1)} - z)\right]. \]

We denote by

\[ H_m(z) = \begin{pmatrix} H_{m1}(z) \\ H_{m2}(z) \\ \vdots \\ H_{md}(z) \end{pmatrix}, \]

17
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

$$R(z) = R_1(z) - R_2(z),$$

$$R_1(z) = [R_{1,12}^T(z), \ldots, R_{1,NT}^T(z)]^T,$$

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

and

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

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

We denote by

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

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

and

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

$$\mathcal{R}_d(Z_{i(t-1); z}) = \int_0^1 \left[ \frac{\partial^2 m_d}{\partial z \partial z^T}(z + \omega (Z_{i(t-1)} - z)) - \frac{\partial^2 m_d}{\partial z \partial z^T}(z) \right] (1 - \omega) \, d\omega. \quad (30)$$

We first analyze the bias term. In order to do this, note that substituting (27) into (26) and noting that $\text{vec} \{ D_m(z) \}$ in (27) vanishes because

$$e_1^T \bar{Z}^T W \bar{Z}^{-1} \bar{Z}^T W \mathcal{R} \left( m(z) \right)_{\text{vec} \{ D_m(z) \}} = e_1^T \mathcal{R} \left( m(z) \right)_{\text{vec} \{ D_m(z) \}} = m(z), \quad (31)$$

then,

$$E \{ \hat{m}(z; H) | X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \} - m(z) = \frac{1}{2} e_1^T (\bar{Z}^T W \bar{Z})^{-1} \bar{Z}^T W Q_m(z) + e_1^T (\bar{Z}^T W \bar{Z})^{-1} \bar{Z}^T W R(z). \quad (32)$$

We will first analyze the asymptotic behavior of $\frac{1}{2} e_1^T (\bar{Z}^T W \bar{Z})^{-1} \bar{Z}^T W Q_m(z)$. 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

$$(NT)^{-1} \bar{Z}^T W \bar{Z} = \begin{pmatrix} \mathcal{A}_{NT}^{11} & \mathcal{A}_{NT}^{12} \\ \mathcal{A}_{NT}^{12} & \mathcal{A}_{NT}^{22} \end{pmatrix} \quad (33)$$
where,

\[ A_{NT}^{11} = (NT)^{-1} \sum_{it} \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)}, \]

\[ A_{NT}^{12} = (NT)^{-1} \sum_{it} \Delta X_{it} \left\{ X_{it}^T \otimes (Z_{it} - z)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T \right\} K_{it} K_{i(t-1)}, \]

\[ A_{NT}^{21} = (NT)^{-1} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\} \Delta X_{it}^T K_{it} K_{i(t-1)}, \]

\[ A_{NT}^{22} = (NT)^{-1} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\} \times \]

\[ \left\{ X_{it}^T \otimes (Z_{it} - z)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T \right\} K_{it} K_{i(t-1)}. \]

We first show that as \( N \) tends to infinity

\[ A_{NT}^{11} = B_{\Delta X \Delta X}(z, z) + o_p(1). \] (34)

where,

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

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

\[ E \left( A_{NT}^{11} \right) = \int \int E \left[ \Delta X_{it} \Delta X_{it}^T | Z_{it} = z + H^{1/2} u, Z_{i(t-1)} = z + H^{1/2} v \right] \]

\[ \times f_{Z_{it}, Z_{i(t-1)}}(z = z + H^{1/2} u, Z_{i(t-1)} = z + H^{1/2} v) K(u) K(v) ddudv. \]

Furthermore, under assumptions (A.1) and (A.4) and a Taylor expansion, as \( N \) tends to infinity, (34) holds. All what we need to close the proof is show that \( \text{Var} \left( A_{NT}^{11} \right) \to 0 \), as the sample size tends to infinity. Note that under the assumption (A.1),

\[ \text{Var} \left( A_{NT}^{11} \right) = \frac{1}{NT} \text{Var} \left( \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)} \right) \]

\[ + \frac{1}{NT^2} \sum_{t=3} (T - t) \text{Cov} \left( \Delta X_{it2} \Delta X_{it2}^T K_{i2} K_{i1}, \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)} \right). \]

Under assumptions (A.4) to (A.6)

\[ \text{Var} \left( \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)} \right) \leq \frac{C}{NT|H|}. \]

and

\[ \text{Cov} \left( \Delta X_{it2} \Delta X_{it2}^T K_{i2} K_{i1}, \Delta X_{it} \Delta X_{it}^T K_{it} K_{i(t-1)} \right) \leq \frac{C'}{N|H|}. \]

Then if both \( NT|H| \) and \( N|H| \) tend to infinity the variance tends to cero and (34) holds.

Similarly one can show that

\[ A_{NT}^{12} = DB_{\Delta X \Delta X}(z, z) \left( I_d \otimes \mu_2 \left( K_u \right) H \right) - DB_{\Delta X \Delta X_{-1}}(z, z) \left( I_d \otimes \mu_2 \left( K_v \right) H \right) + o_p(H). \] (35)

\( DB_{\Delta X \Delta X}(Z_1, Z_2) \) and \( DB_{\Delta X \Delta X_{-1}}(Z_1, Z_2) \) are respectively \( d \times dq \) gradient matrices defined as

\[ DB_{\Delta X \Delta X}(Z_1, Z_2) = \left( \begin{array}{ccc}
\frac{\partial \Delta x_{i1}^{XX}(Z_1, Z_2)}{\partial Z_1^1} & \cdots & \frac{\partial \Delta x_{id}^{XX}(Z_1, Z_2)}{\partial Z_1^1} \\
\frac{\partial \Delta x_{i1}^{XX}(Z_1, Z_2)}{\partial Z_1^2} & \cdots & \frac{\partial \Delta x_{id}^{XX}(Z_1, Z_2)}{\partial Z_1^2} \\
\vdots & & \vdots \\
\frac{\partial \Delta x_{i1}^{XX}(Z_1, Z_2)}{\partial Z_1^d} & \cdots & \frac{\partial \Delta x_{id}^{XX}(Z_1, Z_2)}{\partial Z_1^d} 
\end{array} \right), \]

19
and
\[ b_{dd}^{\Delta X X}(Z_1, Z_2) = E \left[ \Delta X_{dit} X_{dit}' \mid Z_{it} = Z_1, Z_{i(t-1)} = Z_2 \right] f_{Z_{it}, Z_{i(t-1)}}(Z_1, Z_2). \]

The other gradient matrix is
\[
DB_{\Delta X X^{-1}}(Z_1, Z_2) = \begin{pmatrix}
\frac{\partial b_{dd}^{\Delta X X^{-1}}(Z_1, Z_2)}{\partial Z_1^1} & \cdots & \frac{\partial b_{dd}^{\Delta X X^{-1}}(Z_1, Z_2)}{\partial Z_{1}^d} \\
\vdots & \ddots & \vdots \\
\frac{\partial b_{dd}^{\Delta X X^{-1}}(Z_1, Z_2)}{\partial Z_{1}^d} & \cdots & \frac{\partial b_{dd}^{\Delta X X^{-1}}(Z_1, Z_2)}{\partial Z_{1}^d}
\end{pmatrix},
\]

and
\[ b_{dd}^{\Delta X X^{-1}}(Z_1, Z_2) = E \left[ \Delta X_{dit} X_{dit}'(t-1) \mid Z_{it} = Z_1, Z_{i(t-1)} = Z_2 \right] f_{Z_{it}, Z_{i(t-1)}}(Z_1, Z_2). \]

Finally,
\[ A_{NT}^{2,1} = B_{XX}(z, z) \otimes \mu_2(K_u) H + B_{X_{-1}X_{-1}}(z, z) \otimes \mu_2(K_v) H + o_p(H). \] (36)

where
\[ B_{XX}(z, z) = E \left[ X_{it} X_{it}' \mid Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it}, Z_{i(t-1)}}(z, z), \]
\[ B_{X_{-1}X_{-1}}(z, z) = E \left[ X_{i(t-1)} X_{i(t-1)}' \mid Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it}, Z_{i(t-1)}}(z, z). \]

Using the the results shown in (34), (35) and (36) we obtain
\[ NT \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} = \begin{pmatrix}
C_{11} & C_{12} \\
C_{21} & C_{22}
\end{pmatrix}, \] (37)

where
\[ C_{11} = \frac{B_{\Delta X X}(z, z)}{A_{XX}(z, z)} + o_p(1), \]
\[ C_{12} = -B_{\Delta X X}(z, z) \left( DB_{\Delta X X}(z, z) (I_d \otimes \mu_2(K_u) H) - DB_{\Delta X X^{-1}}(z, z) (I_d \otimes \mu_2(K_v) H) \right) \]
\[ \times \left( \left[ B_{XX}(z, z) \otimes \mu_2(K_u) H + B_{XX^{-1}}(z, z) \otimes \mu_2(K_v) H \right]^{-1} + o_p(1), \right. \]
\[ C_{21} = \left. \left( \left[ B_{XX}(z, z) \otimes \mu_2(K_u) H + B_{XX^{-1}}(z, z) \otimes \mu_2(K_v) H \right]^{-1} \right. \right. \]
\[ \times \left. \left. \left( DB_{\Delta X X}(z, z) (I_d \otimes \mu_2(K_u) H) - DB_{\Delta X X^{-1}}(z, z) (I_d \otimes \mu_2(K_v) H) \right)^T B_{\Delta X X}(z, z) \right) \right] + o_p(1), \]
\[ C_{22} = \left( \left[ B_{XX}(z, z) \otimes \mu_2(K_u) H + B_{XX^{-1}}(z, z) \otimes \mu_2(K_v) H \right]^{-1} + o_p(H^{-1}). \right. \]

Also it is straightforward to show that the terms in
\[ (NT)^{-1} \tilde{Z}^T W S_{m1}(z) \]
are asymptotically equal to
\[ (NT)^{-1} \sum_{it} \Delta X_{it} \{ X_{it}' \otimes (Z_{it} - z)^T \} H_m(z)(Z_{it} - z)K_{it}K_{i(t-1)} \]
\[ (NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \} \{ X_{it}' \otimes (Z_{it} - z)^T \} H_m(z)(Z_{it} - z)K_{it}K_{i(t-1)} \]
\[ (NT)^{-1} \sum_{it} \Delta X_{it} \{ X_{it}' \otimes (Z_{it} - z)^T \} H_m(z)(Z_{it} - z)K_{it}K_{i(t-1)} \] (38)
\[ = \mu_2(K_u) E \left[ \Delta X_{it} X_{it}' \mid Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it}, Z_{i(t-1)}}(z, z) \times \text{diag} \{ \text{tr} \{ H_m(z) H \} \} i_d + o_p(\text{tr} \{ H \}), \]
where \( \text{diag} \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \} \) stands for a diagonal matrix of elements \( \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \), for \( r = 1, \cdots, d \), and \( i_d \) is a \( d \times 1 \) unit vector.

\[
(NT)^{-1} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z) \right\} \left\{ X_{it}^T \otimes (Z_{it} - z)^T \right\} \mathcal{H}_m(z)(Z_{it} - z)K_{it}K_{i(l-1)}
\]

\[
= \int B_{XX}(z, z) \otimes \left( H^{1/2}u \right) \left( H^{1/2}u \right)^T \mathcal{H}_m(z) \left( H^{1/2}u \right) K(u) K(v) dudv
\]

\[
- \int B_{X^{-1}X}(z, z) \otimes \left( H^{1/2}v \right) \left( H^{1/2}v \right)^T \mathcal{H}_m(z) \left( H^{1/2}u \right) K(u) K(v) dudv + o_p \left( H^{3/2} \right)
\]

Finally, the terms in

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

\[
\left( NT \right)^{-1} \sum_{it} \Delta X_{it} \left\{ X_{i(l-1)}^T \otimes (Z_{i(l-1)} - z)^T \right\} \mathcal{H}_m(z)(Z_{i(l-1)} - z)K_{it}K_{i(l-1)}
\]

are of order

\[
(NT)^{-1} \sum_{it} \Delta X_{it} \left\{ X_{i(l-1)}^T \otimes (Z_{i(l-1)} - z)^T \right\} \mathcal{H}_m(z)(Z_{i(l-1)} - z)K_{it}K_{i(l-1)}
\]

\[
= \mu_2(K_v) E \left[ \Delta X_{it} X_{i(l-1)}^T \right] Z_{it} = z, Z_{i(l-1)} = z \int f_{Z_{it},Z_{i(l-1)}}(z, z) \times \text{diag} \{ \text{tr} \{ \mathcal{H}_{m_r}(z) H \} \} i_d + o_p \left( \text{tr} \{ H \} \right),
\]

and

\[
\left( NT \right)^{-1} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z) \right\} \left\{ X_{i(l-1)}^T \otimes (Z_{i(l-1)} - z)^T \right\} \mathcal{H}_m(z)(Z_{i(l-1)} - z)K_{it}K_{i(l-1)}
\]

\[
= \int B_{XX}(z, z) \otimes \left( H^{1/2}u \right) \left( H^{1/2}v \right)^T \mathcal{H}_m(z) \left( H^{1/2}u \right) K(u) K(v) dudv
\]

\[
- \int B_{X^{-1}X}(z, z) \otimes \left( H^{1/2}v \right) \left( H^{1/2}v \right)^T \mathcal{H}_m(z) \left( H^{1/2}u \right) K(u) K(v) dudv + o_p \left( H^{3/2} \right)
\]

\[
= O_p \left( H^{3/2} \right).
\]

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 (29) and (30) note that

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

where

\[
\mathcal{E}_1(z) = \frac{1}{NT} \sum_{it} \Delta X_{it}
\]

\[
\times \left\{ X_{it} \otimes (Z_{it} - z)^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z)^T \mathcal{R}(Z_{i(l-1)}; z)(Z_{i(l-1)} - z) \right\} K_{it}K_{i(l-1)}
\]

and

\[
\mathcal{E}_2(z) = \frac{1}{NT} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z) \right\} \times
\]

\[
\left\{ X_{it} \otimes (Z_{it} - z)^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z)^T \mathcal{R}(Z_{i(l-1)}; z)(Z_{i(l-1)} - z) \right\} K_{it}K_{i(l-1)}.
\]
\[ \mathcal{E}_1(z) = \mathcal{E}_{11}(z) + \mathcal{E}_{12}(z), \]

where

\[ \mathcal{E}_{11}(z) = \frac{1}{NT} \sum_{it} K_{it} K_{i(t-1)} \Delta X_{it} \times \]

\[ \left[ \{X_{it} \otimes (Z_{it} - z)\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \mathcal{R}(Z_{i(t-1)}; z)(Z_{i(t-1)} - z) \right]. \]

and

\[ \mathcal{E}_{12}(z) = \frac{1}{NT} \sum_{it} K_{it} K_{i(t-1)} \Delta X_{it} \times \]

\[ \left[ \{X_{i(t-1)} \otimes (Z_{i(t-1)} - z)\}^T \{\mathcal{R}(Z_{i(t-1)}; z) - \mathcal{R}(Z_{i(t-1)}; z)\} (Z_{i(t-1)} - z) \right]. \]

We will show now that, as \( N \) tends to infinity

\[ E \{ \mathcal{E}_1(z) \} = o_p \left( \text{tr} \{ H \} \right). \]

In order to prove this note that

\[ E \{ \mathcal{E}_{11}(z) \} = \int \int K(u) K(v) \left\{ \mathcal{B}_{XX}(z + H^{1/2}u, z + H^{1/2}v) \otimes \left( H^{1/2}u \right)^T \right\} \mathcal{R} \left( z + H^{1/2}u; z \right) \left( H^{1/2}u \right) dudv \]

\[ - \int \int K(u) K(v) \left\{ \mathcal{B}_{XX-1}(z + H^{1/2}u, z + H^{1/2}v) \otimes \left( H^{1/2}v \right)^T \right\} \mathcal{R} \left( z + H^{1/2}u; z \right) \left( H^{1/2}v \right) dudv. \]

By (30) and assumption (A.7)

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

where \( \varsigma (\eta) \) is the modulus of continuity of \( \frac{\partial^2 m_4}{\partial z_i \partial z_j} (z) \). Hence by boundedness of \( f \), \( \mathcal{B}_{XX} \) and \( \mathcal{B}_{XX-1} \)

\[ |E \{ \mathcal{E}_{11}(z) \} | \leq C_1 \int \int \int_0^1 \left| \left( H^{1/2}u \right)^T \right| \varsigma \left( \omega \| H^{1/2}u \| \right) \|H^{1/2}u|K(u)K(v)\| \omega \| H^{1/2}u \| \|H^{1/2}u|K(u)K(v)\| d\omega dudv \]

\[ +C_2 \int \int \int_0^1 \left| \left( H^{1/2}v \right)^T \right| \varsigma \left( \omega \| H^{1/2}u \| \right) \|H^{1/2}v|K(u)K(v)\| \omega \| H^{1/2}u \| \|H^{1/2}v|K(u)K(v)\| d\omega dudv. \]

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

Similarly,

\[ E \{ \mathcal{E}_{12}(z) \} = \int \int K(u) K(v) \left\{ \mathcal{B}_{XX-1}(z + H^{1/2}u, z + H^{1/2}v) \otimes \left( H^{1/2}v \right)^T \right\} \times \]

\[ \left\{ \mathcal{R} \left( z + H^{1/2}u; z \right) - \mathcal{R} \left( z + H^{1/2}v; z \right) \right\} \left( H^{1/2}v \right) dudv. \]

Therefore,

\[ |E \{ \mathcal{E}_{12}(z) \} | \leq C_3 \int \int \int_0^1 \left| \left( H^{1/2}u \right)^T \right| \times \]

\[ \varsigma \left( \omega \| H^{1/2}u \| \right) \|H^{1/2}v|K(u)K(v)\| \omega \| H^{1/2}u \| \|H^{1/2}v|K(u)K(v)\| d\omega dudv + C_4 \int \int \int_0^1 \left| \left( H^{1/2}v \right)^T \right| \times \]

\[ \varsigma \left( \omega \| H^{1/2}v \| \right) \|H^{1/2}v|K(u)K(v)\| \omega \| H^{1/2}u \| \|H^{1/2}v|K(u)K(v)\| d\omega dudv. \]
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
\[
\mathcal{E}_2(z) = \mathcal{E}_{21}(z) + \mathcal{E}_{22}(z),
\]
where
\[
\mathcal{E}_{21}(z) = \frac{1}{NT} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\} \times \left[ \{ X_{it} \otimes (Z_{it} - z) \}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \{ X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \}^T \mathcal{R}(Z_{i(t-1)}; Z_{i(t-1)} - z) \right].
\]

Following the same lines as for the proof of (47) it is easy to show that
\[
E \{ \mathcal{E}_2(z) \} = o_p \left( H^{3/2} \right).
\]
Substituting (37), (38), (40) and (47) into (32), the asymptotic bias can be written as
\[
E \{ \hat{m}(z; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} - m(z) = \frac{1}{2} e_1^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \{ S_m (z) - S_{m_2} (z) \}
= \frac{1}{2} B_{X \Delta X X}^{-1}(z, z) [\mu_2 (K_u) B_{\Delta XX} (z, z) - \mu_2 (K_v) B_{\Delta XX_{-1}} (z, z)] \text{diag} \{ \text{tr} \{ \mathcal{H}_{mv} (z) H \} \} i_d + o_p(\text{tr} \{ H \}).
\]

To obtain an asymptotic expression for the variance let us first define the \((NT - 1) \times 1\)-vector \( \Delta v = (\Delta v_1, \cdots, \Delta v_N)^T \) where \( \Delta v_i = (\Delta v_{i1}, ..., \Delta v_{iT})^T \) and let \( E (\Delta v \Delta v^T) = V \) be a \((NT - 1) \times (NT - 1)\) matrix that contains the \( V_{ij} \)'s matrices
\[
V_{ij} = E(\Delta v_i \Delta v_j^T) | X_{i1}, ..., X_{iT}, Z_{i1}, ..., Z_{iT}) = \begin{cases} 2\sigma^2_v, & \text{for } i = j, \quad t = s \\ -\sigma^2_v, & \text{for } i = j, \quad |t - s| < 2 \\ 0, & \text{for } i = j, \quad |t - s| \geq 2. \end{cases}
\]
Then, taking into account that
\[
\hat{m}(z; H) - E \{ \hat{m}(z; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} = e_1^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta v,
\]
the variance of \( \hat{m}(z; H) \) can be written as
\[
\text{Var} \{ \hat{m}(z; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} = e_1^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W \text{Var} W^T \tilde{Z} (\tilde{Z}^T W \tilde{Z})^{-1} e_1.
\]
Based on assumption (A.2) and the fact that the \( v_{it} \) are i.i.d. then, the upper left entry of \( \frac{1}{N T} \tilde{Z}^T W Y W^T \tilde{Z} \) is

\[
\begin{align*}
\frac{2\sigma_v^2}{NT} & \sum_{it} \Delta X_{it} \Delta X_{it}^T K_{ii(t-1)}^2 - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \Delta X_{i(t-1)}^T K_{it} K_{i(t-1)}^2 K_{i(t-2)} \\
& \quad - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=2}^T \Delta X_{it} \Delta X_{i(t-2)}^T K_{it} K_{i(t-1)} K_{i(t-2)}^2 K_{i(t-3)} \\
& = \frac{2\sigma_v^2 R(K_u) R(K_v)}{|H|} B_{\Delta X \Delta X} (z, z) \{1 + o_p(1)\}, \quad (55)
\end{align*}
\]

because

\[
\frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \Delta X_{i(t-1)}^T K_{it} K_{i(t-1)}^2 K_{i(t-2)} = \frac{\sigma_v^2 R(K_v)}{|H|^{1/2}} B_{\Delta X \Delta X_{-1}} (z, z, z) \{1 + o_p(1)\}, \quad (56)
\]

and

\[
\frac{\sigma_v^2}{NT} \sum_i \sum_{t=2}^T \Delta X_{it} \Delta X_{i(t-2)}^T K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} = \sigma_v^2 R(K_v) B_{\Delta X \Delta X_{-2}} (z, z, z, z) \{1 + o_p(1)\}. \quad (57)
\]

The upper right block is

\[
\begin{align*}
\frac{2\sigma_v^2}{NT} & \sum_{it} \Delta X_{it} \left\{ X_{it}^T \otimes (Z_{it} - z)^T - X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T \right\} K_{it}^2 K_{i(t-1)}^2 \\
& \quad - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=3}^T \Delta X_{it} \left\{ X_{i(t-1)}^T \otimes (Z_{i(t-1)} - z)^T - X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z)^T \right\} K_{it}^2 K_{i(t-1)}^2 K_{i(t-2)} \\
& \quad - \frac{\sigma_v^2}{NT} \sum_i \sum_{t=4}^T \Delta X_{it} \left\{ X_{i(t-2)}^T \otimes (Z_{i(t-2)} - z)^T - X_{i(t-3)}^T \otimes (Z_{i(t-3)} - z)^T \right\} K_{it} K_{i(t-1)} K_{i(t-2)}^2 K_{i(t-3)} \\
& = I_1 - I_2 - I_3. \quad (58)
\end{align*}
\]

\[
I_1 = \frac{\sigma_v^2}{|H|} \int \left\{ B_{\Delta X X} \left( z + H^{1/2} u, z + H^{1/2} v \right) \otimes \left( H^{1/2} u \right)^T \\
- B_{\Delta X X_{-1}} \left( z + H^{1/2} u, z + H^{1/2} v \right) \otimes \left( H^{1/2} v \right)^T \right\} K^2(u) K^2(v) dudv \{1 + o_p(1)\}
\]

\[
= O_p \left(|H|\right), \quad (59)
\]

\[
I_2 = \frac{\sigma_v^2}{|H|^{1/2}} \int \left\{ B_{\Delta X X_{-1}} \left( z + H^{1/2} u, z + H^{1/2} v, z + H^{1/2} w \right) \otimes \left( H^{1/2} v \right)^T \\
- B_{\Delta X X_{-2}} \left( z + H^{1/2} u, z + H^{1/2} v, z + H^{1/2} w \right) \otimes \left( H^{1/2} w \right)^T \right\} K(u) K^2(v) K(w) dudvdw \{1 + o_p(1)\}
\]

\[
= O_p \left(|H|^{1/2}\right),
\]

\[24\]
and

\[ I_3 = \sigma_v^2 \int \left\{ B_{\Delta XX} (z + H^{1/2}u, z + H^{1/2}v, z + H^{1/2}w, z + H^{1/2}w) \otimes (H^{1/2}w)^T \right. \]

\[-B_{\Delta XX-3} (z + H^{1/2}u, z + H^{1/2}v, z + H^{1/2}w, z + H^{1/2}w) \otimes (H^{1/2}s)^T \left\} \times K(u)K(v)K(w)K(s)dudvdwds \{1 + o_p(1)\} \]

\[ = o_p(1), \quad (60) \]

where

\[ \frac{B_{\Delta XX}(z, z)}{f_{Z_{it}, Z_{it}(t-1)}(z, z)} = E \left[ \Delta X_{it}X_{it}^T | Z_{it} = z, Z_{it(t-1)} = z \right], \]

\[ \frac{B_{\Delta XX-1}(z, z)}{f_{Z_{it}, Z_{it}(t-1)}(z, z)} = E \left[ \Delta X_{it}X_{it(1-1)}^T | Z_{it} = z, Z_{it(t-1)} = z \right], \]

\[ \frac{B_{\Delta XX-1}(z, z, z)}{f_{Z_{it}, Z_{it}(t-1), Z_{it}(t-2)}(z, z, z)} = E \left[ \Delta X_{it}X_{it(1-2)}^T | Z_{it} = z, Z_{it(t-1)} = z, Z_{it(t-2)} = z \right], \]

\[ \frac{B_{\Delta XX-3}(z, z, z)}{f_{Z_{it}, Z_{it}(t-1), Z_{it}(t-2), Z_{it}(t-3)}(z, z, z, z)} = E \left[ \Delta X_{it}X_{it(1-3)}^T | Z_{it} = z, Z_{it(t-1)} = z, Z_{it(t-2)} = z, Z_{it(t-3)} = z \right]. \]

Finally, the lower-right block is

\[ 2\sigma_v^2 \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{it(t-1)} \otimes (Z_{it(t-1)} - z) \right\} \left\{ X_{it}^T \otimes (Z_{it} - z)^T - X_{it(t-1)}^T \otimes (Z_{it(t-1)} - z)^T \right\} K_{it}^2 K_{it(t-1)}^2 \]

\[ -\sigma_v^2 \sum_{i=1}^{n} \sum_{t=3}^{T} \left\{ X_{it} \otimes (Z_{it} - z) - X_{it(t-1)} \otimes (Z_{it(t-1)} - z) \right\} \]

\[ \times \left\{ X_{it(t-1)}^T \otimes (Z_{it(t-1)} - z)^T - X_{it(t-2)}^T \otimes (Z_{it(t-2)} - z)^T \right\} K_{it}^2 K_{it(t-1)}^2 K_{it(t-2)} \]

\[ -\sigma_v^2 \sum_{i=1}^{n} \sum_{t=4}^{T} \left\{ X_{it} \otimes (Z_{it} - z) - X_{it(t-1)} \otimes (Z_{it(t-1)} - z) \right\} \]

\[ \times \left\{ X_{it(t-2)}^T \otimes (Z_{it(t-2)} - z)^T - X_{it(t-3)}^T \otimes (Z_{it(t-3)} - z)^T \right\} K_{it}^2 K_{it(t-1)}^2 K_{it(t-2)} K_{it(t-3)} \]

\[ = I_1 - I_2 - I_3, \quad (62) \]

where

\[ I_1 = \frac{2\sigma_v^2 \mu_2(K^2) R(K_v)}{|H|} B_{XX}(z, z) \otimes H + \frac{2\sigma_v^2 \mu_2(K^2) R(K_u)}{|H|} B_{XX-1}(z, z) \otimes H + o_p \left(|H|^{-1/2} H \right), \]

\[ I_2 = \frac{\sigma_v^2 \mu_2(K^2)}{|H|^{1/2}} B_{XX-1}(z, z) \otimes H + o_p \left(|H|^{-1/2} H \right), \]

\[ I_3 = o_p(H). \]

So now, substituting (37), (55), (58) and (62) into (54) we obtain

\[ Var \{ \hat{m}(z; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} = \frac{2\sigma_v^2 R(K_u) R(K_v)}{NT|H|} B_{\Delta XX}^{-1}(z, z) \{1 + o_p(1)\}. \]
Proof of Theorem 3.2

Let
\[
\hat{m} (z; H) - m (z) = \{ \hat{m} (z; H) - E [\hat{m} (z; H)] | X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \} \\
+ \{ E [\hat{m} (z; H)] | X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \} - m (z) \equiv I_1 + I_2.
\]

Now we show that
\[
\sqrt{NT} | H | I_1 \to N \left( 0, 2\sigma_v^2 R (K_u) R (K_v) \mathcal{B}_{\Delta X \Delta X}^{-1} (z, z) \right),
\]
(63)
as \( N \) tends to infinity.

In order to show this let
\[
\hat{m} (z; H) - E \{ \hat{m} (z; H) | X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \} = e_1^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W \Delta v,
\]
(64)
where \( \Delta v = [\Delta v_{11}, \cdots, \Delta v_{NT}]^T \). We are going to show the asymptotic normality of
\[
\frac{1}{\sqrt{NT}} \tilde{Z}^T W \Delta v.
\]
(65)

Since (65) is a multivariate vector we define a unit vector \( d \in \mathbb{R}^{d(1+q)} \) in such a way that
\[
\frac{1}{\sqrt{NT}} d^T \tilde{Z}^T W \Delta v = \frac{1}{\sqrt{NT}} \sum_i \sum_t \lambda_{it},
\]
(66)
where
\[
\lambda_{it} = |H|^{1/2} d^T \tilde{Z}_{it} K_{it} K_{i(t-1)} \Delta v_{it}, \quad i = 1, \cdots, N \quad t = 2, \ldots, T.
\]
and
\[
\tilde{Z}_{it} = \left( \begin{array}{c} \Delta X_{it} \\ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \end{array} \right).
\]
(67)

By Theorem 3.1 and conditions thereof we have that
\[
\text{Var} (\lambda_{it}) = 2\sigma_v^2 d^T \left( R (K_u) R (K_v) \mathcal{B}_{\Delta X \Delta X} (z, z) 0 \\
0 \mu_2 (K^2) R (K_v) \mathcal{B}_{XX} (z, z) \otimes H + \mu_2 (K^2) R (K_u) \mathcal{B}_{X^{-1}X^{-1}} (z, z) \otimes H \right) d \{ 1 + o_p (1) \}
\]
(68)
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 (63) it suffices to check Liapunov’s condition. By Minkowski’s inequality
\[
E |\lambda_{n,i}^*|^{2+\delta} \leq CT^{2+\delta} E |\lambda_{it}|^{2+\delta}.
\]
Because of (67) we split \( \lambda_{it} \) into two components \( \lambda_{1it} \) and \( \lambda_{2it} \) and we analyze them by separate.

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

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

Therefore, \((NT)^{-\frac{2+\delta}{2}} \sum_{i=1}^{N} E \left[ |\lambda_{n,i}|^{2+\delta} \right] \leq C (N |H|)^{-\delta/2}\). This indeed tends to zero when \(N|H| \to \infty\) and therefore Lyapunov’s condition holds and (63) follows. \(D\) has been already defined in (68). Finally, using (37) and applying the Cramer-Wold device the proof is done.

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

\[
E \left\{ \hat{m} (z; H) \right\} X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \right\} - m(z) \\
= \text{diag}_r \left\{ \text{tr} \{ H_{m_r} (z) H \} \right\} i_d + O_p \left( H^{3/2} \right) + o_p \left( \text{tr} \{ H \} \right). \tag{69}
\]

Note that by the law of iterated expectations,

\[
E \left\{ \hat{m} (z; H) \right\} = \int E \left\{ \hat{m} (z; H) \right\} X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \right\} dF (X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT}).
\]

The leading term in (69) does not depend on the sample and then, the proof is closed. and the proof is done.
Proof of Theorem 3.3

The proof of this result follows the same lines as in the proof of Theorem 3.1. Let
\[
\hat{m}^{(1)}(z; \bar{H}) = e_{1}^{T} \left( \bar{Z}^{(1)T} W^{(1)} \hat{Z}^{(1)} \right)^{-1} \bar{Z}^{(1)T} W^{(1)} \Delta \bar{Y}^{(1)}.
\] (70)

Then proceeding as before in the proof of Theorem 3.1 we get
\[
E \left\{ \hat{m}^{(1)}(z; \bar{H}) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} = e_{1}^{T} \left( \bar{Z}^{(1)T} W^{(1)} \hat{Z}^{(1)} \right)^{-1} \bar{Z}^{(1)T} W^{(1)} \left[ M^{(1)} + M^{(2)} \right],
\] (71)

where
\[
M^{(1)} = \left[ \{ X_{12}^{T} m(Z_{12}) \}^{T}, \ldots, \{ X_{NT}^{T} m(Z_{NT}) \}^{T} \right],
\]
\[
M^{(2)} = \left[ \{ X_{11}^{T} \{ E \{ \hat{m}(Z_{11}; H) \mid X_{11}, \ldots, Z_{NT} \} - m(Z_{11}) \} \}^{T}, \ldots,
\right.
\]
\[
\left. \{ X_{N(T-1)}^{T} \{ E \{ \hat{m}(Z_{N(T-1); H) \mid X_{11}, \ldots, Z_{NT} \} - m(Z_{N(T-1)}) \} \}^{T} \right]
\]
are \(N(T-1) \times 1\) vectors. We can approximate \(M^{(1)}\) through a Taylor’s expansion, i.e.
\[
M^{(1)} = \bar{Z}^{(1)} \left[ \frac{m(z)}{\text{vec} \{ D_{m}(z) \}} \right] + \frac{1}{2} Q_{m}^{(1)}(z) + R(z),
\]
where,
\[
Q_{m}^{(1)}(z) = \left[ S_{m,12}^{(1)}(z), \ldots, S_{m,NT}^{(1)}(z) \right]^{T},
\]
and
\[
S_{m,zt}^{(1)}(z) = \{ X_{zt}^{T} \otimes (Z_{zt} - z)^{T} \} H_{m}(Z_{zt} - z).
\]
Using (A.1) we get
\[
e_{1}^{T} \left( \bar{Z}^{(1)T} W \bar{Z} \right)^{-1} \bar{Z}^{(1)T} W^{(1)} R(z) = o_p \left( \text{tr} \left\{ \bar{H} \right\} \right),
\]
and therefore,
\[
E \left\{ \hat{m}^{(1)}(z; \bar{H}) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} - m(z)
\]
\[
= e_{1}^{T} \left( \bar{Z}^{(1)T} W^{(1)} \hat{Z}^{(1)} \right)^{-1} \bar{Z}^{(1)T} W^{(1)} \left\{ \frac{1}{2} Q_{m}^{(1)}(z) + M^{(2)} \right\} + o_p \left( \text{tr} \left\{ \bar{H} \right\} \right).
\] (72)

To obtain an asymptotic expression for the bias we first calculate
\[
\frac{1}{NT} \bar{Z}^{(1)T} W^{(1)} \hat{Z}^{(1)} = \left( (NT)^{-1} \sum_{it} X_{it}^{T} X_{it} K_{it} \right) \left( (NT)^{-1} \sum_{it} X_{it} \left\{ X_{it}^{T} \otimes (Z_{it} - z)^{T} \right\} K_{it} \right)
\]
\[
\left( (NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} X_{it}^{T} K_{it} \right) \left( (NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} \left\{ X_{it}^{T} \otimes (Z_{it} - z)^{T} \right\} K_{it} \right)
\]
Using standard properties of kernel density estimators, under conditions (A.1) to (A.9) and as \(N\) tends to infinity,
\[
(NT)^{-1} \sum_{it} X_{it}^{T} X_{it} K_{it} = B_{XX}(z) + o_p(1),
\]
\[
(NT)^{-1} \sum_{it} X_{it} \left\{ X_{it}^{T} \otimes (Z_{it} - z)^{T} \right\} K_{it} = DB_{XX}(z) \left( I_d \otimes \mu_2(K_u) \bar{H} \right) + o_p \left( \bar{H} \right),
\]
\[
(NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} \left\{ X_{it}^{T} \otimes (Z_{it} - z)^{T} \right\} K_{it} = B_{XX}(z) \otimes \mu_2(K_u) \bar{H} + o_p \left( \bar{H} \right).
\]
Note that $\mathcal{B}_X(z)$ and $\mathcal{D}\mathcal{B}_X(z)$ 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,

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

where

$$C_{11}^{(1)} = B_{XX}^{-1}(z) + o_p(1),$$
$$C_{12}^{(1)} = -B_{XX}^{-1}(z) \left[ \mathcal{DB}^{XX}(z) \right] (B_{XX}^{-1}(z) \otimes I_q) + o_p(1),$$
$$C_{22}^{(1)} = B_{XX}(z) \otimes \mu_2(K_u \tilde{H})^{-1} + o_p \left( \tilde{H}^{-1} \right).$$

Furthermore the terms in

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

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

are of order

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

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

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

$$\begin{pmatrix} (NT)^{-1} \sum_{it} X_{it} X_{i(t-1)}^T \{ E \{ \tilde{m}(Z_{i(t-1)}) \} X_{i1}, \cdots, X_{iNT}, Z_{i1}, \cdots, Z_{iNT} - m(Z_{i(t-1)}) \} K_{it} \\ (NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} X_{i(t-1)}^T \{ E \{ \tilde{m}(Z_{i(t-1)}) \} X_{11}, \cdots, X_{iNT}, Z_{11}, \cdots, Z_{iNT} - m(Z_{i(t-1)}) \} K_{it} \end{pmatrix}.$$ 

It is straightforward to show that

$$(NT)^{-1} \sum_{it} X_{it} X_{i(t-1)}^T \{ E \{ \tilde{m}(Z_{i(t-1)}) \} X_{i1}, \cdots, X_{iNT}, Z_{i1}, \cdots, Z_{iNT} - m(Z_{i(t-1)}) \} K_{it} = o_p \left( \text{tr} \left\{ \mathcal{H} \right\} \right),$$

as $N$ tends to infinity, and

$$(NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} X_{i(t-1)}^T \{ E \{ \tilde{m}(Z_{i(t-1)}) \} X_{i1}, \cdots, X_{iNT}, Z_{i1}, \cdots, Z_{iNT} - m(Z_{i(t-1)}) \} K_{it} = o_p \left( \text{tr} \left\{ H \right\} \text{tr} \left\{ \mathcal{H} \right\} \right),$$

as $N$ tends to infinity. Under assumptions (A.1) to (A.9), the bias is $o_p \left( \text{tr} \left\{ H \right\} \right)$, and the rate is uniform in $z$ (see Masry (1996) for details).

Now substitute the asymptotic expressions for (73), (74) and (75) into (72) apply that $\text{tr} \{ H \} \to 0 \ \text{and} \ \text{tr} \{ \mathcal{H} \} \to 0$ in such a way that $N |H| \to \infty$, $N \left| \mathcal{H} \right| \to \infty$ and we have shown that the asymptotic bias
in \( \hat{m}^{(1)}(z; H) \) is of the same order as it was in the first step.

For the variance term, recall that

\[
\hat{m}^{(1)}(z; \tilde{H}) - E \left\{ \hat{m}^{(1)}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} = e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \Delta v + e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \tilde{v},
\]

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

\[
\hat{v}_i = \left( \left\{ X_i^{T} r(z_0; H) \right\}_T, \ldots, \left\{ X_i^{T} r(z_{T-1}; H) \right\}_T \right)^T,
\]

\( i = 1, \ldots, N, \) and

\[
r(Z_{i(t-1)}; H) = \hat{m} \left( Z_{i(t-1)}; H \right) - E \left\{ \hat{m} \left( Z_{i(t-1)}; H \right) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\},
\]

\( i = 1, \ldots, N; t = 2, \ldots, T. \)

Then, the variance of \( \hat{m}^{(1)}(z; \tilde{H}) \) takes the form

\[
\text{Var} \left\{ \hat{m}^{(1)}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} = I_1 + I_2 + I_3.
\]

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

\[
I_1 = \frac{2\sigma^2 R(K_u)}{NT |\tilde{H}|^{1/2}} B_{XX}^{-1}(z) \{ 1 + o_p(1) \}. \tag{77}
\]

In order to calculate the asymptotic order of \( I_2 \), we just need to calculate

\[
\frac{1}{NT} \tilde{Z}^{(1)T} W^{(1)} E \left\{ \tilde{v} \tilde{v}^T \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} W^{(1)T} \tilde{Z}^{(1)} \tag{78}
\]

The upper left entry is

\[
(NT)^{-1} \sum_{i} \sum_{ts} X_{it} X_{i(t-1)}^T E \left\{ r(Z_{i(t-1)}; H) r(Z_{i(s-1)}; H)^T \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} X_{i(s-1)} X_{is}^T K_{it} K_{is}.
\]

Applying the Cauchy-Schwarz inequality for covariance matrices then (79) is bounded by

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

30
Now, note that under the conditions of the Theorem 3.1

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

uniformly in \( z \), and therefore (79) is of order \( O_p\left( \frac{\log NT}{NT |H|} \right) \).

Following the same lines, it is easy to show that the upper right entry of (78) is

\[
(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T E\left\{ r(\{z_{i(s-1)};H\}) r(\{z_{i(s-1)};H\})^T \mid X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \right\}
\]

\[
\times X_{i(s-1)} \{ X_{is} \otimes (Z_{is} - z) \}^T K_{it} K_{is} = o_p\left( \frac{\log NT}{NT |H|} \frac{\log NT}{NT |H|} \right),
\]

and finally the lower right entry of (78) is

\[
(NT)^{-1} \sum_i \sum_{ts} \{ X_{it} \otimes (Z_{it} - z) \} X_{i(t-1)}^T E\left\{ r(\{z_{i(t-1)};H\}) r(\{z_{i(s-1)};H\})^T \mid X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \right\}
\]

\[
\times X_{i(s-1)} \{ X_{is} \otimes (Z_{is} - z) \}^T K_{it} K_{is} = O_p\left( \frac{\log NT}{NT |H|} \frac{\log NT}{NT |H|} \right).
\]

Now, combining results in (73) and (78) we show that \( I_2 = o_p\left( \frac{\log NT}{NT |H|} \right) \). Finally a standard Cauchy-Schwarz inequality is enough to show that \( I_3 = o_p\left( \frac{\log NT}{NT |H|} \right) \) and then the proof of the result is closed.

References


