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 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. [1]

JEL codes: C14, C23

Key Words: Varying coefficients model, fixed effects, panel data, local linear regression, oracle 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 Associate Editor and the referees. 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 $Z_{it}$:

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

$$\Delta Y_{it} = X_{it}^T m(Z_{it}) - X_{it(t-1)}^T m(Z_{i(t-1)}) + \Delta \nu_{it}, \quad i = 2, \cdots, N; t = 2, \cdots, T.$$ (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 (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 \( X_{it} \equiv 1 \) (2) becomes a fully nonparametric additive model

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

In this case, Mammen et al. (2009) consider the consistent estimation of nonparametric additive panel data models with both time and individual effects via backfitting. Furthermore, for \( X_{T_{it}} \equiv g(Z_{it}) + \tilde{X}_{T_{it}}^{T} \beta_0 \), 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 \( \beta_0 \), the regression function in (2) becomes a semiparametric partially additive model

\[
\Delta Y_{it} = \beta_{0}^{T} \Delta \tilde{X}_{it} + g(Z_{it}) - g(Z_{i(t-1)}) + \Delta v_{it}, \quad i = 2, \ldots, N; t = 2, \ldots, T.
\]

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). 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 (1995)), 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. 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 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 \( m(\cdot) \) through the following alternative specification

\[
Y_{it} = X_{it}^{T} m(Z_{it}) + \sum_{j=1}^{n} \mu_{i}d_{ij} + v_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T,
\]

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 \( \sum_i \mu_i = 0 \).

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 \( m(z) \), \( m'(z) \), \( \cdots \), \( m^{(p)}(z) \) 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 (1995), Ruppert and Wand (1994) or Zhan-Qian (1996)),

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

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

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

where \( h \) is a bandwidth. For the case of a local constant approximation (i.e. Nadaraya-Watson kernel regression estimator) \( m(z) \) has the following closed form

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

and in the local linear regression case we have

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

(6)
where $\tilde{Z}_{it}$ is a $2 \times 1$ vector such that
\[ \tilde{Z}_{it}^T = \left( \Delta X_{it} \quad X_{it}(Z_{it} - z) - X_{i(t-1)}(Z_{i(t-1)} - z) \right), \quad i = 1, \ldots, N; t = 2, \ldots, T. \tag{7} \]

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 smother 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
\[ \Delta Y_{it}^{(1)} = \Delta Y_{it} + m(Z_{i(t-1)})X_{i(t-1)}, \quad i = 1, \ldots, N; t = 2, \ldots, T. \tag{8} \]

By substituting (2) into (8) we get
\[ \Delta Y_{it}^{(1)} = m(Z_{it})X_{it} + \Delta v_{it}, \quad i = 1, \ldots, N; t = 2, \ldots, T. \tag{9} \]

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. Once we have done this $\Delta Y_{it}^{(1)}$ becomes
\[ \Delta \tilde{Y}_{it}^{(1)} = m(Z_{it})X_{it} + \nu_{it}, \quad i = 1, \ldots, N; t = 2, \ldots, T, \tag{10} \]

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

By doing so, the optimization problem to solve is now the following:
\[ \min_{\alpha, \delta} \sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta \tilde{Y}_{it}^{(1)} - \alpha X_{it} - \delta X_{it}(Z_{it} - z) \right\}^2 K_h(Z_{it} - z). \tag{11} \]

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 \( m(\cdot) \) ans its partial derivatives is the solution to,

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

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, \( \delta = D_m(z) \) is a \( dq \times 1 \) vector, \( D_m(z) = \text{vec} \left( \frac{\partial m(z)}{\partial z} \right) \) is a \( dq \times 1 \) vector of first derivatives of \( m(\cdot) \) and

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

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

\[
\begin{bmatrix} \hat{\alpha} \\ \hat{\delta} \end{bmatrix} = (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W \Delta Y,
\]

where

\[
\begin{align*}
W &= \text{diag} \left\{ K_H(Z_{i2} - z)K_H(Z_{i11} - z), ..., K_H(Z_{iNT} - z)K_H(Z_{i(N(T-1))} - z) \right\}, \\
\Delta Y &= [\Delta Y_{i2}, ..., \Delta Y_{iNT}]^T,
\end{align*}
\]

and

\[
\tilde{Z} = \begin{bmatrix} \Delta X_{i2}^T & X_{i2}^T \otimes (Z_{i2} - z)^T - X_{i11}^T \otimes (Z_{i11} - z)^T \\
& \vdots & \ddots & \vdots \\
\Delta X_{iNT}^T & X_{iNT}^T \otimes (Z_{iNT} - z)^T - X_{i(N(T-1))}^T \otimes (Z_{i(N(T-1))} - z)^T 
\end{bmatrix}.
\]

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

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

where \( e_1 = (I_d^T 0_{dq \times d}) \) 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 choosen 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 \( H = \text{diag}\{h_1^2, \cdots, h_q^2\} \). 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 (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 \((Y_{it}, X_{it}, Z_{it})_{i=1,...,N_i=1,...,T}\) 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) \( \mu_i \) has zero mean and finite variance. Furthermore, \( \mu_i \) can be arbitrarily correlated with both \( X_{it} \) and \( Z_{it} \) with unknown correlation structure.

(A.4) Let \( X_{it} = \left( X_{it}^T \ X_{i(t-1)}^T \right)^T \) and \( \Delta X_{it} = \left( \Delta X_{it}^T \ \Delta X_{i(t-1)}^T \right)^T \). The matrix functions \( E \left( X_{it}X_{it}^T | z_1, z_2 \right) \), \( E \left( X_{it}X_{it}^T | z_1, z_2, z_3 \right) \), \( E \left( \Delta X_{it}\Delta X_{it}^T | z_1, z_2 \right) \) and \( E \left( \Delta X_{it}\Delta X_{it}^T | z_1, z_2, z_3 \right) \) are bounded and uniformly continuous in their support. Furthermore, let \( \|A\| = \sqrt{tr \left(A^T A \right)} \), 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.

(A.5) The function \( E \left[ \Delta X_{it}\Delta X_{it}^T | Z_{it} = z_1, Z_{i(t-1)} = z_2 \right] \) is positive definite for any interior point of \((z_1, z_2)\) in the support of \( f_{Z_{it},Z_{i(t-1)}}(z_1, z_2) \).
(A.6) The functions $E \{ |X_{it} \Delta v_{it}|^{2+\delta} \mid Z_{it} = z_1, Z_{i(t-1)} = z_2 \}$, $E \{ |X_{i(t-1)} \Delta v_{it}|^{2+\delta} \mid Z_{it} = z_1, Z_{i(t-1)} = z_2 \}$ and $E \{ |\Delta X_{it} \Delta v_{it}|^{2+\delta} \mid Z_{it} = z_1, Z_{i(t-1)} = z_2 \}$, 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 $m_1 (\cdot), m_2 (\cdot), \cdots, m_d (\cdot)$ are bounded and uniformly continuous.

(A.8) The Kernel functions $K$ are compactly supported, bounded kernel such that $\int uu^T K(u) du = \mu_2 (K) I$, where $\mu_2 (K) \neq 0$ is scalar and $I$ is the $q \times q$ identity matrix. In addition, all odd-order moments of $K$ vanish, that is $\int u^{\iota_1}_1 \cdots u^{\iota_q}_q K(u) du = 0$, for all nonnegative integers $\iota_1, \cdots, \iota_q$ 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 $N \mid H \mid \rightarrow \infty$.

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 zero at a faster rate. In our case this was no needed because of the assumptions imposed on $v_{it}$. 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 (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 relationship between heterogeneity and explanatory variables and, to avoid an additional bias term they need $\sum_i \mu_i = 0$.

Assumptions (A.4) and (A.5) are some smoothness conditions on moment functionals. Assumption (A.6) is the equivalent to a standard rank condition for identification of this type of models. Assumptions (A.7) to (A.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 (A.1)-(A.9) hold, then as \( N \) tends to infinity and \( T \) is fixed we obtain

\[
E \{ \hat{m}(z;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} - m(z)
\]

\[
= \frac{1}{2} c_1^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \{ S_{m_1}(z) - S_{m_2}(z) \}
\]

\[
= \frac{1}{2} \mathcal{B}_{\Delta X \Delta X}(z,z) \left[ \mu_2(K_u) \mathcal{B}_{\Delta XX}(z,z) - \mu_2(K_v) \mathcal{B}_{\Delta XX_{-1}}(z,z) \right] \text{diag} \{ \text{tr} \{ \mathcal{H}_{mr}(z) H \} \} i_d + o_p(\text{tr} \{ H \}).
\]

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

\[
E \{ \hat{m}(z;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} - m(z) = \frac{1}{2} \mu_2(K_u) \text{diag} \{ \text{tr} \{ \mathcal{H}_{mr}(z) H \} \} i_d + o_p(\text{tr} \{ H \}).
\]

The variance is

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

where

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

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

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

\[
\text{diag} \{ \text{tr} \{ \mathcal{H}_{mr}(z) H \} \} \text{ stands for a diagonal matrix of elements } \text{tr} \{ \mathcal{H}_{mr}(z) H \}, \text{ for } r = 1,\ldots,d, \text{ and } i_d \text{ is a } d \times 1 \text{ unit vector.}
\]

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 \( Nh^2 \to \infty \) as \( N \) tends to infinity and \( T \) is fixed we get

\[
E \{ \hat{m}(z;H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} - m(z)
\]

\[
= \frac{1}{2} c(z,z) m''(z) h^2 + o_p(h^2).
\]

where

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

(1)
Furthermore, if \( \mu_2(K_u) = \mu_2(K_v) \) then the bias term has the following expression

\[
E \{ \tilde{m}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \} - m(z) = \frac{1}{2} m''(z) h^2 + o_p(h^2).
\]

The variance is

\[
\text{Var} \{ \tilde{m}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \} = \frac{2\sigma_u^2 R(K_u) R(K_v)}{NT h^2 f_{Z_{it}, Z_{i(t-1)}}(z, z) E \left\{ \left( X_{it} - X_{i(t-1)} \right)^2 \mid Z_{it} = z, Z_{i(t-1)} = z \right\}} \{1 + o_p(1)\}.
\]

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 \( \tilde{m}(z; H) \):

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

\[
\sqrt{NT |H|} \{ \tilde{m}(z; H) - m(z) \} \rightarrow_d \mathcal{N}(b(z), v(z)),
\]

as \( N \) tends to infinity and \( T \) is fixed, where

\[
b(z) = \frac{1}{2} \mu_u(K_u) \text{diag} \left\{ \text{tr} \left\{ \mathcal{H}_{md}(z) H \sqrt{NT |H|} \right\} \right\} \nu_{it},
\]

\[
v(z) = 2\sigma_v^2 R(K_u) R(K_v) B^{-1}_{\tilde{\Delta}X\tilde{\Delta}X}(z, z).
\]

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

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

where

\[
\nu_{it} = X_{i(t-1)}^T \left[ \tilde{m}(Z_{i(t-1)}; H) - m(Z_{i(t-1)}) \right] + \Delta v_{it}, \quad i = 1, \ldots, N; t = 2, \ldots, T.
\]

In expression (15), \( \tilde{m}(Z_{i(t-1)}; H) \) 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

\[
\min_{\alpha, \delta} \sum_{i=1}^N \sum_{t=2}^T \left\{ \Delta \tilde{Y}_{it}^{(1)} - (X_{it}^T \alpha + X_{it}^T \otimes (Z_{it} - z)^T \delta) \right\}^2 K_H(Z_{it} - z),
\]
where $\beta = [\alpha^T, \delta^T]^T$ a $d(1 + q) \times 1$ vector. If we define $\tilde{Z}_{it}^{(1)T} = [X_{it}^T, X_{it}^T \otimes (Z_{it} - z)^T]$, a $1 \times d(1 + q)$ vector, then (16) can be written as

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

(17)

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

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

(18)

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

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

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 $N \left| \tilde{H} \right| \to \infty$.

(B.2) The bandwidth matrices $H$ and $\tilde{H}$ must fulfill that: (i) $N \left| H \right| / \log (N) \to \infty$, (ii) $\text{tr}(H) / \text{tr}(\tilde{H}) \to 0$, 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

$$E \left\{ \hat{m}^{(1)}(z; \tilde{H}) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \right\} - m(z) = \frac{1}{2} \mu_2 (K_u) \text{diag}_r \left\{ \text{tr} \left\{ \mathcal{H}_{mr} (z) \tilde{H} \right\} \right\} \text{id} + o_p \left( \text{tr} \left\{ \tilde{H} \right\} \right).$$

and

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

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

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

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

where

$$MSE \{ \hat{m}(\cdot; H) \} = E \left\{ \sum_{r=1}^d \{ \hat{m}_r(Z_{it}) - m_r(Z_{it}) \} X_{rit} \right\}^2.$$

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

Let $X, Z$ be vectors of random variables such that $X = (X_1, \ldots, X_d)$ and $Z = (Z_1, \ldots, Z_q)$, and $\mathcal{D}$ the observed covariates vector, $\mathcal{D} = (Z_{111}, \ldots, Z_{1NT}, \ldots, Z_{q11}, \ldots, Z_{qNT}, X_{111}, \ldots, X_{1NT}, \ldots, X_{d11}, \ldots, X_{dNT})^T$. Let $bias(\hat{m}_r(z; H)|\mathcal{D})$ be the conditional bias of a local linear estimator of order $(p = 1)$, $\hat{m}_r(z; H)$, given $\mathcal{D}$ and denoting $b(z; H) = bias(\hat{m}(z; H)|\mathcal{D})$, the proposed MSE is

$$MSE \{ \hat{m}(\cdot) \} = E \left( trE \left[ \{ \hat{m}(Z_{it}; H) - m(Z_{it}) \}^T X_{it}X_{it}^T \{ \hat{m}(Z_{it}; H) - m(Z_{it}) \} \right] \right| Z)$$

$$= E \left( b^T(Z; H)\Omega(Z)b(Z; H) + tr \{ \Omega(Z)Var \{ \hat{m}(Z; H)|Z, \mathcal{D} \} \} \right),$$

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

$$Var \{ \hat{m}(Z; H)|Z, \mathcal{D} \} = Var \{ \hat{m}(z; H)|\mathcal{D} \}|_{Z=z} \quad and \quad MSE \{ \hat{m}(\cdot) \} = E \{ E \{ MSE (\hat{m}(z; H)|\mathcal{D}) \}|_{Z=z} | \mathcal{D} \},$$

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

$$MSE \{ \hat{m}(z; H)|\mathcal{D} \} = b^T(z; H)\Omega(z)b(z; H) + tr \{ \Omega(z)Var \{ \hat{m}(z; H)|\mathcal{D} \} \}.$$
However, given that $MSE\{\hat{m}(z; H)\} \{ D\}$ also depends on some unknown quantities, to obtain $\hat{H}$ we need to estimate both the conditional bias and variance previously. Then, taking the idea of Zhang and Lee (2000) as benchmark we need an additional assumption:

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

Thus, in the bandwidth selector procedure that we propose we first calculate the conditional bias based on a Taylor expansion of order $(p+g)$. Combining this approximation with (14)

$$E\{\hat{m}(z; H)|\mathcal{D}\} - m(z) = e_1^T \left( \bar{Z}^T W \bar{Z} \right)^{-1} \bar{Z}^T W \left( \frac{1}{2} Q_b(z) D_{b2}(z) + \frac{1}{3!} Q_b^{(p+g)}(z) D_{b(p+g)}(z) + R_b(z) \right),$$

where $D_{b2}(z) = vec\left( \mathcal{H}_{m_1}^{(2)}(z) \right)$ and $D_{b(p+g)}(z) = vec\left( \mathcal{H}_{m_1}^{(p+g)}(z) \right)$ are $dq^2$ and $dq^{(p+g)}$ vectors that contains the vec operator of the Hessian and the $(p+g)$-th derivative matrix of the $r$-th component of $m_1(\cdot)$, respectively, and

$$Q_b(z) = S_{b1}(z) - S_{b2}(z), \quad Q_b^{(p+g)}(z) = S_{b1}^{(p+g)}(z) - S_{b2}^{(p+g)}(z)$$
$$S_{b1}(z) = \left[ S_{b1,12}(z), \ldots, S_{b1,NT}(z) \right]^T, \quad S_{b1}^{(p+g)}(z) = \left[ S_{b1,12}^{(p+g)}(z), \ldots, S_{b1,NT}^{(p+g)}(z) \right]^T,$$
$$S_{b2}(z) = \left[ S_{b2,11}(z), \ldots, S_{b2,N(T-1)}(z) \right]^T, \quad S_{b2}^{(p+g)}(z) = \left[ S_{b2,11}^{(p+g)}(z), \ldots, S_{b2,N(T-1)}^{(p+g)}(z) \right]^T,$$

with

$$S_{b1,lt}(z) = \left[ \{ X_{it} \otimes (Z_{lt} - z) \otimes (Z_{it} - z) \} \right]^T,$$
$$S_{b2,lt-1}(z) = \left[ \{ X_{i(t-1)} \otimes (Z_{lt-1} - z) \otimes (Z_{i(t-1)} - z) \} \right]^T,$$
$$S_{b1}^{(p+g)}(z) = \left[ \{ X_{it} \otimes (Z_{lt} - z)^{(p+g)} \} \right]^T,$$
$$S_{b2}^{(p+g)}(z) = \left[ \{ X_{i(t-1)} \otimes (Z_{lt-1} - z)^{(p+g)} \} \right]^T.$$

Thereby, the conditional bias to estimate is

$$b(z; H) = e_1^T \left( \bar{Z}^T W \bar{Z} \right)^{-1} \bar{Z}^T W \bar{Z}_{(p+g)} D,$$

where

$$\bar{Z}_{(p+g)} = \begin{bmatrix}
S_{b1,12}(z) - S_{b2,11}(z) & S_{b1,12}^3(z) - S_{b2,11}^3(z) & \cdots & S_{b1,12}^{(p+g)}(z) - S_{b2,11}^{(p+g)}(z) \\
S_{b1,12}(z) & S_{b1,11}(z) & \cdots & S_{b1,12}^{(p+g)}(z) \\
\vdots & \vdots & \ddots & \vdots \\
S_{b1,NT}(z) - S_{b2,N(T-1)}(z) & S_{b1,NT}^3(z) - S_{b2,N(T-1)}^3(z) & \cdots & S_{b1,NT}^{(p+g)}(z) - S_{b2,N(T-1)}^{(p+g)}(z)
\end{bmatrix},$$

and

$$D = \begin{bmatrix} D_{b2}^T(z), & D_{b3}^T(z), & \cdots, & D_{b(p+g)}^T(z) \end{bmatrix}^T.$$
is a vector of unknown functions that need to be estimated. For convenience we take \( g = 2 \) so \( \tilde{Z}_3 \) is a \( N(T-1) \times dq^2(1+q) \) matrix and \( D_3 \) a \( dq^2(1+q) \times 1 \) vector.

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

\[
\hat{D}_{bk} = e_k^T \left( \tilde{Z}^T_{(p+g)} W_* \tilde{Z}_{(p+g)} \right)^{-1} \tilde{Z}^T_{(p+g)} W_* \Delta Y, \quad k = 2, \ldots , 4
\]  

(20)

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

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

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

\[
\text{Var}\{ \tilde{m}(z; H) \mid \mathcal{D} \} = e_1^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \nu W \tilde{Z} \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} e_1
\]

where \( \nu = E(\Delta v \Delta v^T | \mathcal{D}) \) is a \( N(T-1) \times N(T-1) \) matrix that contains the \( V_{ij} \)'s matrices described in (50). Using the information of (50), \( \nu = \sigma_v^2 A \) where \( A \) is a \( N(T-1) \times N(T-1) \) matrix of constant, the variance can be rewritten as

\[
\text{Var}\{ \tilde{m}(z; H) | \mathcal{D} \} = e_1^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W A W \tilde{Z} \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} e_1 \sigma_v^2.
\]  

(21)

To calculate (21) we need to estimate the unknown quantity \( \sigma_v^2 \) for which we use the following normalized weighted residual sum of squares from a \((p+g)\)th-order polynomial fit

\[
\hat{\sigma}_v^2 = \frac{1}{\text{tr}(W_*) - \text{tr}\left\{ (\tilde{Z}^T_{(p+g)} W_* \tilde{Z}_{(p+g)})^{-1} \tilde{Z}^T_{(p+g)} W_* A W_* \tilde{Z}_{(p+g)} \right\} } \sum_{i=1}^{N} \sum_{t=2}^{T} \left( \Delta Y_{it} - \Delta \tilde{Y}_{it} \right)^2 K_{H_*}(Z_{it} - z)K_{H_*}(Z_{i(t-1)} - z),
\]  

(22)

where \( \Delta Y_{it} = \tilde{Z}^T_{(p+g)it} D + \Delta v_{sit}, \) \( g = 3, \) \( \Delta v_{sit} \) is the idiosyncratic error term and

\[
\Delta \tilde{Y}_* = \begin{bmatrix} \Delta \tilde{Y}_{*12}, \ldots , \Delta \tilde{Y}_{*NT} \end{bmatrix}^T = \tilde{Z}_{(p+g)} \begin{bmatrix} \tilde{Z}^T_{(p+g)} W_* \tilde{Z}_{(p+g)} \end{bmatrix}^{-1} \tilde{Z}^T_{(p+g)} W_* \Delta Y.
\]

Once known \( \sigma_v^2 \) the conditional variance can be already calculated.

To get the \( \Omega(z) \) matrix it is necessary to estimate the \( a_{rr'}(z) \) element of this matrix, where \( r, r' = 1, \ldots , d \) and \( \Omega(z) = E(X_{rit}X_{r't}^T | Z = z) \), through a fully nonparametric model. For that, we can use some standard nonparametric techniques such as the Nadaraya-Watson estimator (see H"ardle (1990)).

Finally, using (19)-(22) we obtain the following estimate of \( \sqrt{\text{MSE}\{ \tilde{m}(z; H) | \mathcal{D} \}} \),

\[
\sqrt{\text{MSE}\{ \tilde{m}(z; H) | \mathcal{D} \}} = \hat{D}^T \tilde{Z}^T_{(p+g)} W \tilde{Z} \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \left( \hat{\Omega} \otimes e_{1+q} e_{1+q}^T \right) \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \tilde{Z}_{(p+g)} \hat{D} + \\
\text{tr}\left\{ \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W A W \tilde{Z} \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \left( \hat{\Omega} \otimes e_{(1+q)} e_{(1+q)}^T \right) \hat{\sigma}_v^2 \right\}
\]  

(23)

for \((p + g)\) and the minimizer of this expression is \( \hat{H} \), the estimated variable bandwidth matrix for the local lineal estimator with kernel weights.
Meanwhile, to get the estimated variable bandwidth matrix for the “oracle” estimator we adapt the prior procedure to the particularities of this estimator so now \( \hat{H} \) is selected solving

\[
\hat{H} = \arg\min_{H} \text{MSE} \left\{ \hat{m}(z; H) | \mathcal{D} \right\},
\]

where for \( (g = 2) \),

\[
\text{MSE} \left\{ \hat{m}(z; H) | \mathcal{D} \right\} = \mathbb{E} \left[ b^{(1)T}(z; H)\Omega(z)b^{(1)}(z; H) + tr \left\{ \Omega(z)\text{Var}\{\hat{m}(z; H)|\mathcal{D}\} \right\} \right] = \nabla^{(1)T} \tilde{Z}^{(1)} \left[ \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right]^{-1} \left( \hat{\Omega} \otimes e_{1+q} e_{T}^{T} \right) \left[ \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right]^{-1} \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \nabla^{(1)}
\]

\[
+ \, \text{tr} \left\{ \left[ \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right]^{-1} \tilde{Z}^{(1)T} W^{(1)} \left( A \hat{\sigma}_{v}^{2} + \hat{\nu} \right) W^{(1)} \tilde{Z}^{(1)} \left[ \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right]^{-1} \left( \hat{\Omega} \otimes e_{1+q} e_{T}^{T} \right) \right\}
\]

being \( \hat{\nu} = \text{diag} \{ E(\Delta \hat{v}_{1}\Delta \hat{v}_{1}^{T} | \mathcal{D} \}, \cdots , E(\Delta \hat{v}_{N}\Delta \hat{v}_{N}^{T} | \mathcal{D} \} \}, \Delta \hat{v}_{it} = X_{it}^{T} \{ \hat{m}(Z_{it}) \} - E(\hat{m}(Z_{it}) | \mathcal{D} \} \) and \( b^{(1)}(z; H) = \text{bias} (\hat{m}(z; H)|\mathcal{D}) \).

\section{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,

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

where \( X_{dit} \) and \( Z_{qit} \) are scalars random variables; \( v_{it} \) is an i.i.d.N(0,1) random variable; and \( m(.) \) is a pre-specified function to be estimated. The observations follow a data generating process where \( Z_{qit} = w_{qit} + w_{qit(t-1)} \), 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,

\[
\begin{align*}
(1) \quad Y_{it} &= X_{1it}m_{1}(Z_{1it}) + \mu_{1i} + v_{it} \\
(2) \quad Y_{it} &= X_{1it}m_{1}(Z_{1it}, Z_{2it}) + \mu_{2i} + v_{it} \\
(3) \quad Y_{it} &= X_{1it}m_{1}(Z_{1it}) + X_{2it}m_{2}(Z_{2it}) + \mu_{1i} + v_{it},
\end{align*}
\]

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) \( \mu_{1i} \) depends on \( Z_{1it} \), where the dependence is imposed by generating \( \mu_{1i} = c_{0}Z_{1i}. + u_{i} \) for \( i = 2, \cdots , N \) 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 \( i = 2, \cdots , N \) and \( Z_{i.} = \frac{1}{2}(Z_{1i.} + Z_{2i.}) \).
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 \( \{Z_{qit}\}_{i=1}^{N},t=2 \).

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

\[
ISE \{ \hat{m}_t (z; H) \} = \frac{1}{Q} \sum_{l=1}^{Q} \int \left\{ \sum_{r=1}^{d} (\hat{m}_{lr} (z; H) - m_{lr} (z)) X_{it,lr} \right\}^2 f(z)dz
\]

which can me 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 \( q=1 \).

| \( n \) | \( \hat{m}_1(z) \) | \( m_1(z) \) |
|-------|-----------------|-------------|
|       | Local polynomial | Backfitting |
| 150   | 1.2917536       | 1.2862342   |
| 300   | 0.9158168       | 0.915898    |
| 600   | 0.7278588       | 0.7136175   |

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

| \( n \) | \( \hat{m}_1(z) \) | \( m_1(z) \) |
|-------|-----------------|-------------|
|       | Local polynomial | Backfitting |
| 150   | 1.7953523       | 1.5848331   |
| 300   | 1.1097905       | 1.0171259   |
| 600   | 0.8181966       | 0.7656107   |

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

| \( n \) | \( \hat{m}_2(z) \) | \( m_2(z) \) |
|-------|-----------------|-------------|
|       | Local polynomial | Backfitting |
| 150   | 1.34543139      | 1.3072503   |
| 300   | 0.9231841       | 0.8952494   |
| 600   | 0.7305442       | 0.7217711   |

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 a unknown way. The resulting estimator is robust to misspecification in the functional form of the varying parameters and we have shown that it is consistent and asymptotically normal. Furthermore we have shown that it achieves the optimal rate of convergence for this type of problems and it exhibits the so called oracle efficiency property. Since the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. Monte Carlo results indicates good performance of the estimator in finite samples.

7 Appendix

Proof of Theorem 3.1

Taking 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 
\[ E\{\hat{m}(z; H)|X_{11},...,X_{NT},Z_{11},...,Z_{NT}\} = e_T^T (\tilde{Z}^T W \tilde{Z})^{-1} \tilde{Z}^T W M \] (24)
where 
\[ M = \left[ X_{12}^T m(Z_{12}) - X_{11}^T m(Z_{11}),...,X_{NT}^T m(Z_{NT}) - X_{N(T-1)}^T m(Z_{N(T-1)}) \right]^T. \]

Taylor’s Theorem implies that 
\[ M = \tilde{Z} \left[ \begin{array}{c} m(z) \\ D_m(z) \end{array} \right] + \frac{1}{2} Q_m(z) + R(z), \] (25)
where
\[ Q_m(z) = S_{m1}(z) - S_{m2}(z), \]
\[ S_{m1}(z) = [S_{m1,12}(z), \ldots, S_{m1,NT}(z)]^T, \]
\[ S_{m2}(z) = [S_{m2,11}(z), \ldots, S_{m2,N(T-1)}(z)]^T \]
and
\[ S_{m1, it}(z) = \left\{ X_{it} \otimes (Z_{it} - z) \right\}^T H_m(z)(Z_{it} - z), \]
\[ S_{m2, i(t-1)}(z) = \left\{ X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\}^T H_m(z)(Z_{i(t-1)} - z). \]

We denote by
\[ H_m(z) = \begin{pmatrix} H_{m1}(z) \\ H_{m2}(z) \\ \vdots \\ H_{md}(z) \end{pmatrix}, \]
a \(dq \times q\) matrix such that \(H_{md}(z)\) is the Hessian matrix of the \(d\)-th component of \(m(\cdot)\). The remainder term can be written as
\[ R(z) = R_1(z) - R_2(z), \]
\[ R_1(z) = [R_{1,12}(z), \ldots, R_{1,NT}(z)]^T, \]
\[ R_2(z) = [R_{2,11}(z), \ldots, R_{2,N(T-1)}(z)]^T \]
and
\[ R_{1, it}(z) = \left\{ X_{it} \otimes (Z_{it} - z) \right\}^T R(Z_{it}; z) (Z_{it} - z), \]
\[ R_{2, i(t-1)}(z) = \left\{ X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\}^T R(Z_{i(t-1)}; z). \]

We denote by
\[ R(Z_{it}; z) = \begin{pmatrix} R_1(Z_{it}; z) \\ R_2(Z_{it}; z) \\ \vdots \\ R_d(Z_{it}; z) \end{pmatrix}, \]
\[ R(Z_{i(t-1)}; z) = \begin{pmatrix} R_1(Z_{i(t-1)}; z) \\ R_2(Z_{i(t-1)}; z) \\ \vdots \\ R_d(Z_{i(t-1)}; z) \end{pmatrix}, \]
and
\[ R_d(Z_{it}; z) = \int_0^1 \left[ \frac{\partial^2 m_d}{\partial z \partial T} (z + \omega (Z_{it} - z)) - \frac{\partial^2 m_d}{\partial z \partial T} (z) \right] (1 - \omega) \, d\omega, \]
\[ R_d(Z_{i(t-1)}; z) = \int_0^1 \left[ \frac{\partial^2 m_d}{\partial z \partial T} (z + \omega (Z_{i(t-1)} - z)) - \frac{\partial^2 m_d}{\partial z \partial T} (z) \right] (1 - \omega) \, d\omega. \]
We first analyze the bias term. In order to do this, note that substituting (25) into (24) and noting that \( D_m(z) \) in (25) vanishes because
\[
e_T \left[ \tilde{Z}^T W \tilde{Z} \right]^{-1} \tilde{Z}^T W \tilde{Z} \begin{bmatrix} m(z) \\ D_m(z) \end{bmatrix} = e_T \begin{bmatrix} m(z) \\ D_m(z) \end{bmatrix} = m(z),
\] (29)
then,
\[
E \{ \hat{m}(z; H) | X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} - m(z) = \frac{1}{2} \tilde{e}_T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W Q_m(z) + e_T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W R(z). \tag{30}
\]
We will first analyze the asymptotic behavior of \( \tilde{e}_T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W Q_m(z) \). 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
\[
(NT)^{-1} \tilde{Z}^T W \tilde{Z} = \begin{pmatrix} A_{NT}^{11} & A_{NT}^{12} \\ A_{NT}^{21} & A_{NT}^{22} \end{pmatrix}
\] (31)
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
\[
A_{NT}^{11} = B_{\Delta X \Delta X}(z, z) + o_p(1).
\] (32)
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
\[
E(A_{NT}^{11}) = \int \int E \left[ \Delta X_{it} \Delta X_{it}^T | Z_{it} = z + H^{1/2} u, Z_{i(t-1)} = z + H^{1/2} v \right] 
\times f_{Z_{it}, Z_{i(t-1)}}(Z_{it} = z + H^{1/2} u, Z_{i(t-1)} = z + H^{1/2} v) K(u) K(v) dudv.
\]
Furthermore, under assumptions (A.1) and (A.4) and a Taylor expansion, as $N$ tends to infinity, (32) holds. All what we need to close the proof is show that $\text{Var}(A_{NT}^{11}) \to 0$, as the sample size tends to infinity. Note that under the assumption (A.1),

$$\text{Var}(A_{NT}^{11}) = \frac{1}{NT^2} \text{Var}(\Delta X_{it} \Delta X_{it}^T K_{i(t-1)})$$

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

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

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

and

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

Then if both $NT|H|$ and $N|H|$ tend to infinity the variance tends to cero and (32) holds. Similarly one can show that

$$A_{NT}^{12} = DB_{\Delta XX}(z, z) (I_d \otimes \mu_2(K_u)H) - DB_{\Delta XX-1}(z, z) (I_d \otimes \mu_2(K_v)H) + o_p(H).$$ 

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

$$DB_{\Delta XX}(Z_1, Z_2) = \left( \begin{array}{ccc} \frac{\partial \Delta_{z1}X(X)(Z_1, Z_2)}{\partial Z_1^1} & \ldots & \frac{\partial \Delta_{z1}X(X)(Z_1, Z_2)}{\partial Z_1^{dq}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Delta_{z1}X(X)(Z_1, Z_2)}{\partial Z_1^1} & \ldots & \frac{\partial \Delta_{z1}X(X)(Z_1, Z_2)}{\partial Z_1^{dq}} \end{array} \right),$$

and

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

The other gradient matrix is

$$DB_{\Delta XX-1}(Z_1, Z_2) = \left( \begin{array}{ccc} \frac{\partial \Delta_{z1}^{XX-1}(X)(Z_1, Z_2)}{\partial Z_1^1} & \ldots & \frac{\partial \Delta_{z1}^{XX-1}(X)(Z_1, Z_2)}{\partial Z_1^{dq}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \Delta_{z1}^{XX-1}(X)(Z_1, Z_2)}{\partial Z_1^1} & \ldots & \frac{\partial \Delta_{z1}^{XX-1}(X)(Z_1, Z_2)}{\partial Z_1^{dq}} \end{array} \right),$$

and

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

Finally,

$$A_{NT}^{22} = B_{XX}(z, z) \otimes \mu_2(K_u)H + B_{XX-1}(z, z) \otimes \mu_2(K_v)H + o_p(H).$$ 

where

$$B_{XX}(z, z) = E \left[ X_{it} X_{it}^T | Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it},Z_{i(t-1)}}(z, z), \quad B_{XX-1}(z, z) = E \left[ X_{i(t-1)} X_{i(t-1)}^T | Z_{it} = z, Z_{i(t-1)} = z \right] f_{Z_{it},Z_{i(t-1)}}(z, z).$$
Using the results shown in (32), (33) and (34) we obtain

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

where

\[
C_{11} = \mathcal{B}^{-1}_X(z, z) + o_p(1),
\]
\[
C_{12} = -\mathcal{B}^{-1}_X(z, z) \left[ \mathcal{D} \mathcal{B} \mathcal{X} \mathcal{X} (z, z) (I_d \otimes \mu_2 (K_u) H) - \mathcal{D} \mathcal{B} \mathcal{X} \mathcal{X} \mathcal{X}^{-1} (z, z) (I_d \otimes \mu_2 (K_v) H) \right] \\
\times \left( [\mathcal{B} \mathcal{X} \mathcal{X} (z, z) \otimes \mu_2 (K_u) H + \mathcal{B} \mathcal{X} \mathcal{X}^{-1} (z, z) \otimes \mu_2 (K_v) H]^{-1} + o_p(1), \right)
\]
\[
C_{21} = \left( [\mathcal{B} \mathcal{X} \mathcal{X} (z, z) \otimes \mu_2 (K_u) H + \mathcal{B} \mathcal{X} \mathcal{X} \mathcal{X}^{-1} (z, z) \otimes \mu_2 (K_v) H]^{-1} \\
\times \left[ \mathcal{D} \mathcal{B} \mathcal{X} \mathcal{X} \mathcal{X} (z, z) (I_d \otimes \mu_2 (K_u) H) - \mathcal{D} \mathcal{B} \mathcal{X} \mathcal{X} \mathcal{X} \mathcal{X} \mathcal{X}^{-1} (z, z) (I_d \otimes \mu_2 (K_v) H) \right]^{T} \mathcal{B}^{-1}_X(z, z) \right) + o_p(1),
\]
\[
C_{22} = \left( [\mathcal{B} \mathcal{X} \mathcal{X} (z, z) \otimes \mu_2 (K_u) H + \mathcal{B} \mathcal{X} \mathcal{X} \mathcal{X}^{-1} (z, z) \otimes \mu_2 (K_v) H]^{-1} + o_p (H^{-1}). \right)
\]

Also it is straightforward to show that the terms in

\[
(NT)^{-1} \hat{Z}^T W S_{m1}(z)
\]

are asymptotically equal to

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

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

where \( diag_r \{ \mathcal{H}_m (z) H \} \) stands for a diagonal matrix of elements \( \mathcal{H}_m (z) H \), for \( r = 1, \cdots, d \), and \( i_d \) is a \( d \times 1 \) unit vector.

Finally, the terms in

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

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

21
are of order

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

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

and

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

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

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

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

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 (27) and (28) note that

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

where

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

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

and

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

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

\[\mathcal{E}_1(z) = \mathcal{E}_{11}(z) + \mathcal{E}_{12}(z), \]

where

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

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

and

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

\[\left\{ X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\}^T \left\{ \mathcal{R}(Z_{it}; z) - \mathcal{R}(Z_{i(t-1); z}) \right\} (Z_{i(t-1)} - z) \].
We will show now that, as $N$ tends to infinity

$$E \{ \mathcal{E}_1(z) \} = o_p(\text{tr} \{ H \}) \tag{45}. $$

In order to prove this note that

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

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

By (28) 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 $\frac{\partial^2 m}{\partial z_i \partial z_j}(z)$. Hence by boundedness of $f$, $\mathcal{B}_{\Delta XX}$ and $\mathcal{B}_{\Delta XX^{-1}}$

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

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

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

Similarly,

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

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

Therefore,

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

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

$$\varsigma \left( \omega \| H^{1/2}v \| \right) \| H^{1/2}v \| K(u)K(v) d\omega dudv.$$ 

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

$$\mathcal{E}_2(z) = \mathcal{E}_{21}(z) + \mathcal{E}_{22}(z), \tag{46}$$

where

$$\mathcal{E}_{21}(z) = \frac{1}{NT} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\} \times$$

$$\left[ \left\{ X_{it} \otimes (Z_{it} - z) \right\}^T \mathcal{R}(Z_{it}; z)(Z_{it} - z) - \left\{ X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\}^T \mathcal{R}(Z_{it}; z)(Z_{i(t-1)} - z) \right]. \tag{47}$$
and
\[
\mathcal{E}_{22}(z) = \frac{1}{NT} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\} \times \left[ \left\{ X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \right\}^T \{ R(Z_{i(t-1)}; z) - R(Z_{i(t-1)}; z) \} \right] (Z_{i(t-1)} - z) .
\]

Following the same lines as for the proof of (45) it is easy to show that
\[
E \{ \mathcal{E}_2(z) \} = o_p (H^{3/2} ) .
\]

Substituting (35), (36), (38) and (45) into (30), the asymptotic bias can be written as
\[
E \{ \hat{m}(z; H) \mid X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} - m(z)
\]
\[
= \frac{1}{2} e_1^T \left( \bar{Z}^T W \bar{Z} \right)^{-1} \bar{Z}^T W \{ S_{m_1} (z) - S_{m_2} (z) \}
\]
\[
= \frac{1}{2} B_{\Delta X \Delta X}^{-1} (z, z) \left[ \mu_2 (K_u) B_{\Delta X \Delta X} (z, z) - \mu_2 (K_v) B_{\Delta X \Delta X} (z, z) \right] \text{diag} \{ \text{tr} \{ H_{mr} (z) H \} \} i_d + o_p (\text{tr} \{ H \}) .
\]

To obtain an asymptotic expression for the variance let us first define the \((N(T - 1) \times 1)\)-vector \(\Delta v = (\Delta v_1, ..., \Delta v_N)^T \) where \(\Delta v_i = (\Delta v_{i2}, ..., \Delta v_{iT})^T \) and let \(E (\Delta v \Delta v^T) = V \) be a \(N(T - 1) \times N(T - 1)\) matrix that contains the \(V_{ij}\)'s matrices
\[
V_{ij} = E (\Delta v_i \Delta v_j^T) \mid X_{11}, ..., X_{iT}, Z_{11}, ..., Z_{iT} = \begin{cases} 2 \sigma_v^2, & \text{for } i = j, \ t = s \\ -\sigma_v^2, & \text{for } i = j, \ |t = s| < 2. \\ 0, & \text{for } i = j, \ |t = s| \geq 2. \end{cases}
\]

Then, taking into account that
\[
\hat{m}(z; H) - E \{ \hat{m}(z; H) \mid X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \} = e_1^T \left( \bar{Z}^T W \bar{Z} \right)^{-1} \bar{Z}^T W \Delta v ,
\]
the variance of \(\hat{m}(z; H)\) can be written as
\[
\text{Var} \{ \hat{m}(z; H) \mid X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \}
\]
\[
= e_1^T (\bar{Z}^T W \bar{Z})^{-1} \bar{Z}^T W \mu W^T \bar{Z} (\bar{Z}^T W \bar{Z})^{-1} e_1.
\]

Based on assumption \((A.2)\) and the fact that the \(v_{it}\) are i.i.d. then, the upper left entry of \(\frac{1}{N^T} \bar{Z}^T W \mu W^T \bar{Z} \)

is
\[
= \frac{2 \sigma_v^2}{N T} \sum_{i t} \Delta X_{i t} \Delta X_{i t}^T K_{i t}^2 K_{i(t-1)}^2 - \frac{\sigma_v^2}{N T} \sum_{i t} \Delta X_{i t} \Delta X_{i(t-1)}^T K_{i t} K_{i(t-1)} K_{i(t-2)}
\]
\[
- \frac{\sigma_v^2}{N T} \sum_{i t} \Delta X_{i t} \Delta X_{i(t-2)}^T K_{i t} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)}
\]
\[
= \frac{2 \sigma_v^2 R(K_u) R(K_v)}{|H|} B_{\Delta X \Delta X} (z, z) \{ 1 + o_p (1) \} ,
\]

24
because
\[
\frac{\sigma^2}{NT} \sum_{i} \sum_{t=3}^{T} \Delta X_{it} \Delta X_{i(t-1)}^T K_{it} K_{i(t-1)}^2 K_{i(t-2)} = \frac{\sigma^2 R(K_v)}{|H|^{1/2}} B_{\Delta X \Delta X_{-1}} (z, z, z) \{ 1 + o_p(1) \},
\] (54)

and
\[
\frac{\sigma^2}{NT} \sum_{i} \sum_{t=4}^{T} \Delta X_{it} \Delta X_{i(t-1)}^T K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)} = \frac{\sigma^2 R(K_v)}{|H|^{1/2}} B_{\Delta X \Delta X_{-2}} (z, z, z, z) \{ 1 + o_p(1) \}.
\] (55)

The upper right block is
\[
\frac{2\sigma^2}{NT} \sum_{it} \Delta X_{it} \left\{ X_{it}^T (Z_{it} - z)^T - X_{i(t-1)}^T (Z_{i(t-1)} - z)^T \right\} K_{it}^2 K_{i(t-1)}^2
- \frac{\sigma^2}{NT} \sum_{i} \sum_{t=3}^{T} \Delta X_{it} \left\{ X_{i(t-1)}^T (Z_{i(t-1)} - z)^T - X_{i(t-2)}^T (Z_{i(t-2)} - z)^T \right\} K_{it} K_{i(t-1)} K_{i(t-2)}
- \frac{\sigma^2}{NT} \sum_{i} \sum_{t=4}^{T} \Delta X_{it} \left\{ X_{i(t-2)}^T (Z_{i(t-2)} - z)^T - X_{i(t-3)}^T (Z_{i(t-3)} - z)^T \right\} K_{it} K_{i(t-1)} K_{i(t-2)} K_{i(t-3)}
= I_1 - I_2 - I_3.
\] (56)

\[
I_1 = \frac{\sigma^2}{|H|} \int \left\{ B_{\Delta XX} \left( z + H^{1/2}u, z + H^{1/2}v \right) \otimes \left( H^{1/2}u \right)^T
- B_{\Delta XX_{-1}} \left( z + H^{1/2}u, z + H^{1/2}v \right) \otimes \left( H^{1/2}v \right)^T \right\} K^2(u) K^2(v) dudv \{ 1 + o_p(1) \}
= O_p \left( |H| \right),
\] (57)

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

and
\[
I_3 = \frac{\sigma^2}{H} \int \left\{ B_{\Delta XX_{-2}} \left( z + H^{1/2}u, z + H^{1/2}v, z + H^{1/2}w, z + H^{1/2}s \right) \otimes \left( H^{1/2}w \right)^T
- B_{\Delta XX_{-3}} \left( z + H^{1/2}u, z + H^{1/2}v, z + H^{1/2}w, z + H^{1/2}s \right) \otimes \left( H^{1/2}s \right)^T \right\} K(u) K(v) K(w) K(s) dudvdwds \{ 1 + o_p(1) \}
= O_p \left( 1 \right).
\] (58)
where
\[
\frac{B_{\Delta XX}(z, z)}{f_{Z_{il}, Z_{i(l-1)}}(z, z)} = E \left[ \Delta X_{il} X_{it}^T \mid Z_{it} = z, Z_{i(l-1)} = z \right],
\]
\[
\frac{B_{\Delta XX-1}(z, z)}{f_{Z_{il}, Z_{i(l-1)}}(z, z)} = E \left[ \Delta X_{il} X_{i(l-1)}^T \mid Z_{il} = z, Z_{i(l-1)} = z \right],
\]
\[
\frac{B_{\Delta XX-2}(z, z)}{f_{Z_{il}, Z_{i(l-1)}, Z_{i(l-2)}}(z, z)} = E \left[ \Delta X_{il} X_{i(l-2)}^T \mid Z_{il} = z, Z_{i(l-1)} = z, Z_{i(l-2)} = z \right],
\]
\[
\frac{B_{\Delta XX-3}(z, z)}{f_{Z_{il}, Z_{i(l-1)}, Z_{i(l-2)}, Z_{i(l-3)}}(z, z)} = E \left[ \Delta X_{il} X_{i(l-3)}^T \mid Z_{il} = z, Z_{i(l-1)} = z, Z_{i(l-2)} = z, Z_{i(l-3)} = z \right].
\]

Finally, the lower-right block is
\[
\frac{2\sigma^2}{NT} \sum_{it} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z) \right\} \left\{ X_{i(l-1)}^T \otimes (Z_{i(l-1)} - z)^T - X_{i(l-2)}^T \otimes (Z_{i(l-2)} - z)^T \right\} K_{il}^2 K_{il}^{-1}
\]
\[
- \frac{\sigma^2}{NT} \sum_{i} \sum_{t=3}^{T} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z) \right\}
\]
\[
\times \left\{ X_{i(l-1)}^T \otimes (Z_{i(l-1)} - z)^T - X_{i(l-2)}^T \otimes (Z_{i(l-2)} - z)^T \right\} K_{il}^2 K_{i(l-1)}^2 K_{i(l-2)}
\]
\[
- \frac{\sigma^2}{NT} \sum_{i} \sum_{t=4}^{T} \left\{ X_{it} \otimes (Z_{it} - z) - X_{i(l-1)} \otimes (Z_{i(l-1)} - z) \right\}
\]
\[
\times \left\{ X_{i(l-2)}^T \otimes (Z_{i(l-2)} - z)^T - X_{i(l-3)}^T \otimes (Z_{i(l-3)} - z)^T \right\} K_{it} K_{i(l-1)} K_{i(l-2)} K_{i(l-3)}
\]
\[\equiv I_1 - I_2 - I_3, \tag{60}\]

where
\[
I_1 = \frac{2\sigma^2 \mu_2(K^2)}{|H|} B_{XX}(z, z) \otimes H + \frac{2\sigma^2 \mu_2(K^2)}{|H|} B_{XX^{-1}X^{-1}}(z, z) \otimes H + o_p \left( |H|^{-1} H \right),
\]
\[
I_2 = \frac{\sigma^2 \mu_2(K^2)}{|H|^{1/2}} B_{XX^{-1}X^{-1}}(z, z) \otimes H + o_p \left( |H|^{-1/2} H \right),
\]
\[
I_3 = o_p(H).
\]

So now, substituting (35), (53), (56) and (60) into (52) we obtain
\[
\text{Var} \{ \hat{m}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \} = \frac{2\sigma^2 R(K_{u}) R(K_{v})}{NT|H|} B_{\Delta XX^{-1}}^{-1}(z, z) \{ 1 + o_p(1) \}.
\]

**Proof of Theorem 3.2**

Let
\[
\hat{m}(z; H) - m(z) = \{ \hat{m}(z; H) - E[\hat{m}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT}] \}
\]
\[+ \{ E[\hat{m}(z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT}] - m(z) \} \equiv I_1 + I_2.
\]
Now we show that
\[ \sqrt{NT}|H|_1 \to N\left(0, 2\sigma_v^2 R(K_u) R(K_v) B_{\Delta X \Delta X}^{-1}(z, z)\right), \]
(61)
as $N$ tends to infinity.

In order to show this let
\[ \hat{m}(z; H) - E \{ \hat{m}(z; H)|X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT}\} = e_i^T \left( \tilde{Z}^T W \tilde{Z} \right)^{-1} \tilde{Z}^T W \Delta v, \]
(62)
where $\Delta v = [\Delta v_{11}, \cdots, \Delta v_{NT}]^T$. We are going to show the asymptotic normality of
\[ \frac{1}{\sqrt{NT}} \tilde{Z}^T W \Delta v. \]
(63)
Since (63) is a multivariate vector we define a unit vector $d \in \mathbb{R}^{d(1+q)}$ in such a way that
\[ \frac{1}{\sqrt{NT}} d^T \tilde{Z}^T W \Delta v = \frac{1}{\sqrt{NT}} \sum_i \sum_t \lambda_{it}, \]
(64)
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
\[ \tilde{Z}_{it} = \left( \begin{array}{c} \Delta X_{it} \\ X_{it} \otimes (Z_{it} - z) - X_{i(t-1)} \otimes (Z_{i(t-1)} - z) \end{array} \right). \]
(65)

By Theorem 3.1 and conditions thereof we have that
\[ \text{Var}(\lambda_{it}) = \]
(66)
\[ 2\sigma_v^2 d^T \begin{pmatrix} R(K_u) R(K_v) B_{\Delta X \Delta X}(z, z) & 0 \\ 0 & \mu_2(K^2) R(K_v) B_{XX}(z, z) \otimes H + \mu_2(K^2) R(K_u) B_{X_iX_i}(z, z) \otimes H \end{pmatrix} d \{1 + o_p(1)\} \]
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 (61) 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 (65) we split $\lambda_{it}$ into two components $\lambda_{1it}$ and $\lambda_{2it}$ and we analyze them by separate.

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

Therefore, \((NT)^{-\frac{2+\delta}{2}} \sum_{i=1}^{N} E \left| \lambda_{n,i}^* \right|^{2+\delta} \leq C (N|H|)^{-\delta/2}. \) This indeed tends to zero when \(N|H| \to \infty\) and therefore Lyapunov’s condition holds and (61) follows. \(D\) has been already defined in (66). Finally, using (35) and applying the Cramer-Wold device the proof is done.

Using the bias expression computed in Theorem 3.1 we can then write
\[
E \left\{ \hat{m} (z; H) \right\} X_{11}, \cdots, X_{NT}, Z_{11}, \cdots, Z_{NT} \right\} m(z) = \text{diag} \{ \text{tr} \{ \mathcal{H}_{mr} (z) H \} \} i_d + O_p \left( H^{3/2} \right) + o_p \left( \text{tr} \{ H \} \right).
\]

(67)

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 (67) does not depend on the sample and then, the proof is closed. and the proof is done.

\[ \blacksquare \]

**Proof of Theorem 3.3**

The proof of this result follows the same lines as in the proof of Theorem 3.1. Let
\[
\hat{m}^{(1)}(z; \tilde{H}) = e_1^T (\tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)})^{-1} \tilde{Z}^{(1)T} W^{(1)} \Delta \tilde{Y}^{(1)}.
\]

(68)

Then proceeding as before in the proof of Theorem 3.1 we get
\[
E \left\{ \hat{m}^{(1)}(z; \tilde{H}) \left| X_{11}, ..., X_{NT}, Z_{11}, ..., Z_{NT} \right\} = e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \left[ M^{(1)} + M^{(2)} \right],
\]

(69)

where
\[
M^{(1)} = \left[ \left\{ X_{12}^T m(Z_{12}) \right\}^T, ..., \left\{ X_{NT}^T m(Z_{NT}) \right\}^T \right]^T,
\]
\[
M^{(2)} = \left[ \left\{ X_{11}^T \left\{ E \left\{ \hat{m}(Z_{11}; H) \left| X_{11}, ..., Z_{NT} \right\} - m(Z_{11}) \right\} \right\}^T, ..., \right\}^T,
\]
\[
\left\{ X_{N(T-1)}^T \left\{ E \left\{ \hat{m}(Z_{N(T-1)}; H) \left| X_{11}, ..., Z_{NT} \right\} - m(Z_{N(T-1)}) \right\} \right\}^T \right]^T.
\]

28
are \( N(T - 1) \times 1 \) vectors. We can approximate \( M^{(1)} \) through a Taylor’s expansion, i.e.

\[
M^{(1)} = \tilde{Z}^{(1)} \left[ \begin{array}{c} m(z) \\ D_m(z) \end{array} \right] + \frac{1}{2} Q_m^{(1)}(z) + R(z),
\]

where,

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

and

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

Using (A.1) we get

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

and therefore,

\[
E \left\{ \tilde{m}^{(1)}(z; \tilde{H}) | X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} - m(z) = e_1^T \left( \tilde{Z}^{(1)} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)} W^{(1)} \left\{ \frac{1}{2} Q_m^{(1)}(z) + M^{(2)} \right\} + o_p \left( \text{tr} \left\{ \tilde{H} \right\} \right).
\]

To obtain an asymptotic expression for the bias we first calculate

\[
\frac{1}{NT} \tilde{Z}^{(1)} W^{(1)} \tilde{Z}^{(1)} =
\]

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

Using standard properties of kernel density estimators, under conditions (A.1) to (A.9) and as \( N \) tends to infinity,

\[
(NT)^{-1} \sum_{it} X_{it} X_{it}^T K_{it} = \mathcal{B}_{XX}(z) + o_p(1),
\]

\[
(NT)^{-1} \sum_{it} X_{it} \{ X_{it}^T \otimes (Z_{it} - z)^T \} K_{it} = \mathcal{D}\mathcal{B}_{XX}(z) \left( I_d \otimes \mu_2(K_u) \tilde{H} \right) + o_p \left( \tilde{H} \right),
\]

\[
(NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} \{ X_{it}^T \otimes (Z_{it} - z)^T \} K_{it} = \mathcal{B}_{XX}(z) \otimes \mu_2(K_u) \tilde{H} + o_p \left( \tilde{H} \right).
\]

Note that \( \mathcal{B}_{XX}(z) \) and \( \mathcal{D}\mathcal{B}_{XX}(z) \) 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,

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

where

\[
C_{11}^{(1)} = \mathcal{B}_{XX}^{-1}(z) + o_p(1),
\]

\[
C_{12}^{(1)} = -\mathcal{B}_{XX}^{-1}(z) \left[ \mathcal{D}\mathcal{B}_{XX}^{X}(z) \right] \left( \mathcal{B}_{XX}^{-1}(z) \otimes I_d \right) + o_p(1),
\]

\[
C_{22}^{(1)} = \left( \mathcal{B}_{XX}(z) \otimes \mu_2(K_u) \tilde{H} \right)^{-1} + o_p \left( \tilde{H}^{-1} \right).
\]
Furthermore the terms in
\[(NT)^{-1} \tilde{Z}^{(1)T} W^{(1)} Q^{(1)}_m(z) = \]
\[
\left( (NT)^{-1} \sum_{it} X_{it} \left\{ X_{iT}^T \otimes (Z_{it} - z)^T \right\} \mathcal{H}_m(z)(Z_{it} - z)K_{it} \\
(NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \right\} X_{iT}^T \otimes (Z_{it} - z)^T \mathcal{H}_m(z)(Z_{it} - z)K_{it} \right) \]
are of order
\[
\mu_2(K_u) E \left[ X_{it}X_{it}^T \right]_{Z_t = z} f_{Z_t}(z) \times \text{diag} \left\{ \tilde{H}_{md}(z) \right\} i_d + o_p \left( \text{tr} \left\{ \tilde{H} \right\} \right)
\]
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
\[
(NT)^{-1} \tilde{Z}^{(1)T} W^{(1)} M^{(2)} = \]
\[
\left( (NT)^{-1} \sum_{it} X_{it}X_{it}^T \left( E \left\{ \hat{m}(Z_{it(t-1)}) \right\} X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} - m(Z_{it(t-1)}) \right) K_{it} \\
(NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} X_{iT}^T \left( E \left\{ \hat{m}(Z_{it(t-1)}) \right\} X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} - m(Z_{it(t-1)}) \right) K_{it} \right).
\]
It is straightforward to show that
\[
(NT)^{-1} \sum_{it} X_{it}X_{it}^T \left( E \left\{ \hat{m}(Z_{it(t-1)}) \right\} X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} - m(Z_{it(t-1)}) \right) K_{it} = o_p \left( \text{tr} \left\{ H \right\} \right),
\]
as \(N\) tends to infinity, and
\[
(NT)^{-1} \sum_{it} \{ X_{it} \otimes (Z_{it} - z) \} X_{iT}^T \left( E \left\{ \hat{m}(Z_{it(t-1)}) \right\} X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} - m(Z_{it(t-1)}) \right) K_{it} = o_p \left( \text{tr} \left\{ H \right\} \text{tr} \left\{ \tilde{H} \right\} \right),
\]
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 (71), (72) and (73) into (70) apply that \(\text{tr} \left\{ H \right\} \to 0\) \(\text{tr} \left\{ \tilde{H} \right\} \to 0\) in such a way that \(N|H| \to \infty, N|\tilde{H}| \to \infty\) 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
\[
\hat{m}^{(1)}(z; \tilde{H}) - E \left\{ \hat{m}^{(1)}(z; H) \right| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} = e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \Delta v
+ e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \tilde{v},
\]
where \(\tilde{v} = (\tilde{v}_1, \ldots, \tilde{v}_N)^T\) is a \((N(T - 1) \times 1)\)-vector, such that
\[
\tilde{v}_i = \left\{ X_{it}^T r(Z_{it}; H) \right\}^T, \left\{ X_{it}^T r(Z_{it(t-1)}; H) \right\}^T, i = 1, \ldots, N, \text{ and } r(Z_{it(t-1)}; H) = \hat{m}(Z_{it(t-1)}; H) - E \left\{ \hat{m}(Z_{it(t-1)}; H) \right| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\}, i = 1, \ldots, N; t = 2, \ldots, T.$
Then, the variance of $\hat{m}^{(1)} (z; \hat{H})$ takes the form
\[
\text{Var} \left\{ \hat{m}^{(1)} (z; H) \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\}
\]  
(74)
\[
eq e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} (1) \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} \left( \tilde{Z}^{(1)T} W^{(1)} (1) \tilde{Z}^{(1)} \right)^{-1} e_1
\]
+ $e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} (1) \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} E \left\{ \tilde{v} \tilde{v}^T \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} W^{(1) T} \tilde{Z}^{(1)} \left( \tilde{Z}^{(1)T} W^{(1)} (1) \tilde{Z}^{(1)} \right)^{-1} e_1$
\[
+ 2 e_1^T \left( \tilde{Z}^{(1)T} W^{(1)} (1) \tilde{Z}^{(1)} \right)^{-1} \tilde{Z}^{(1)T} W^{(1)} E \left\{ \tilde{\Delta} \tilde{v}^T \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} W^{(1) T} \tilde{Z}^{(1)} \left( \tilde{Z}^{(1)T} W^{(1)} (1) \tilde{Z}^{(1)} \right)^{-1} e_1.
\]
\[
\equiv I_1 + I_2 + I_3.
\]
Following exactly the same lines as in the proof of the variance term in Theorem 3.1 we get, as $N$ tends to infinity,
\[
I_1 = \frac{2\sigma_\varepsilon^2 R_1 (K_\varepsilon)}{NT \left| \tilde{H} \right|^{1/2}} B_X^{-1} (z) \{ 1 + o_p(1) \}.
\]  
(75)
In order to calculate the asymptotic order of $I_2$, we just need to calculate
\[
\frac{1}{NT} \tilde{Z}^{(1) T} W^{(1)} E \left\{ \tilde{v} \tilde{v}^T \mid X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} W^{(1) T} \tilde{Z}^{(1)}.
\]  
(76)
The upper left entry is
\[
(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T E \left\{ r \left( Z_{i(t-1)} ; H \right) r \left( Z_{i(s-1)} ; H \right)^T \left| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} X_{i(s-1)} X_{is}^T K_{it} K_{is}.
\]
(77)
Applying the Cauchy-Schwarz inequality for covariance matrices then (77) is bounded by
\[
(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T \text{vec}^{1/2} \left\{ \text{diag} \left( E \left\{ r \left( Z_{i(t-1)} ; H \right) r \left( Z_{i(s-1)} ; H \right)^T \left| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} \right) \right\} \right\}
\times \text{vec}^{1/2} \left\{ \text{diag} \left( E \left\{ r \left( Z_{i(s-1)} ; H \right) r \left( Z_{i(s-1)} ; H \right)^T \left| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} \right) \right\}^T X_{i(s-1)} X_{is}^T K_{it} K_{is}.
\]
Now, note that under the conditions of the Theorem 3.1
\[
\text{vec} \left\{ \text{diag} \left( E \left\{ r \left( z ; H \right) r \left( z ; H \right)^T \left| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\} \right) \right\} = O_p \left( \log \frac{NT}{|H|} \right),
\]
uniformly in $z$, and therefore (77) is of order $O_p \left( \log \frac{NT}{|H|} \right)$.
Following the same lines, it is easy to show that the upper right entry of (76) is
\[
(NT)^{-1} \sum_i \sum_{ts} X_{it} X_{i(t-1)}^T E \left\{ r \left( Z_{i(t-1)} ; H \right) r \left( Z_{i(s-1)} ; H \right)^T \left| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\}
\times X_{i(s-1)} \left\{ Z_{is} \otimes (Z_{is} - z) \right\}^T K_{it} K_{is} = o_p \left( \log \frac{NT}{|H|} \right),
\]
and finally the lower right entry of (76) is
\[
(NT)^{-1} \sum_i \sum_{ts} \left\{ X_{it} \otimes (Z_{it} - z) \right\} X_{i(t-1)}^T E \left\{ r \left( Z_{i(t-1)} ; H \right) r \left( Z_{i(s-1)} ; H \right)^T \left| X_{11}, \ldots, X_{NT}, Z_{11}, \ldots, Z_{NT} \right\}
\times X_{i(s-1)} \left\{ Z_{is} \otimes (Z_{is} - z) \right\}^T K_{it} K_{is} = O_p \left( \log \frac{NT}{|H|} \right).
\]
Now, combining results in (71) and (76) 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

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. Y. 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, Y. and I. Gijbels (1995). “Local polynomial modelling and its applications”. Chapman & Hall.

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

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

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. Stove, 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.

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. and M. Wand (1994). “Multivariate locally weighted least squares regression”. *The Annals of Statistics* 22(3), 1346–1370.
Schultz, T. (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. 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, Massachusets.

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.