

Efficient semiparametric estimation in time-varying regression models

Lionel Truquet ^{*†}

Abstract

We study semiparametric inference in some linear regression models with time-varying coefficients, dependent regressors and dependent errors. This problem, which has been considered recently by [Zhang and Wu \(2012\)](#) under the functional dependence measure, is interesting for parsimony reasons or for testing stability of some coefficients in a linear regression model. In this paper, we propose a different procedure for estimating non time-varying parameters at the rate \sqrt{n} , in the spirit of the method introduced by [Robinson \(1988\)](#) for partially linear models. When the errors in the model are martingale differences, this approach can lead to more efficient estimates than the method considered in [Zhang and Wu \(2012\)](#). For a time-varying AR process with exogenous covariates and conditionally Gaussian errors, we derive a notion of efficient information matrix from a convolution theorem adapted to triangular arrays. For independent but non identically distributed Gaussian errors, we construct an asymptotically efficient estimator in a semiparametric sense.

1 Introduction

Regression models with time-varying parameters have been widely studied in the literature. This type of models includes standard regression models, some linear time series models such as the locally stationary AR process studied by [Dahlhaus \(1997\)](#) or some time series models with covariates considered for instance in [Chen and Hong \(2012\)](#). In [Zhang and Wu \(2012\)](#), the authors study a general linear regression model with time-varying coefficients and dependent errors. This model writes

$$y_i = x_i' \beta(i/n) + e_i, \quad 1 \leq i \leq n, \quad \mathbb{E}(e_i | x_i) = 0, \quad (1)$$

where $(x_i)_{1 \leq i \leq n}$ and $(e_i)_{1 \leq i \leq n}$ are triangular arrays of dependent random vectors and β is a smooth function which defines the time-varying coefficients in the regression model. Model (1) is quite general under the assumptions used in this paper. For instance, the regressors and the noise component can be stationary processes or locally stationary processes in the sense of [Dahlhaus \(1997\)](#). An interesting and natural question for model (1) is to reduce model complexity by assuming that some covariates are associated to a constant parameter. In [Zhang and Wu \(2012\)](#), a testing procedure is given for detecting some stable (or non time-varying) coefficients. Under the null hypothesis, one faces to a semiparametric model of type

$$y_i = x_{1,i}' \beta_1 + x_{2,i}' \beta_2(i/n) + e_i, \quad 1 \leq i \leq n, \quad (2)$$

with $x_i = (x_{i,1}', x_{i,2}')'$. Semiparametric models are often an interesting alternative when the number of regressors is important or when the sample size is moderate. Moreover, semiparametric models are often a

^{*}UMR 6625 CNRS IRMAR, University of Rennes 1, Campus de Beaulieu, F-35042 Rennes Cedex, France and

[†]ENSAI, Campus de Ker-Lann, rue Blaise Pascal, BP 37203, 35172 Bruz cedex, France. *Email: lionel.truquet@ensai.fr.*

first step to modify a parametric model by taking in account of a particular inhomogeneity in the data. For instance, [Gao and Hawthorne \(2006\)](#) investigated a regression model in which the time-varying trend is the single nonparametric component. In the same spirit, in [Truquet \(2016\)](#), we studied semiparametric versions of locally stationary ARCH models with a particular emphasis on the case of a time-varying unconditional variance and non time-varying lag parameters. The square of the ARCH process being a weak AR process, estimation of the ARCH coefficients is a particular case of statistical inference in model (1).

As shown in [Zhang and Wu \(2012\)](#), under some general assumptions, non time-varying parameters in model (1) can be estimated at the rate \sqrt{n} . The technique used in their paper consists in averaging the nonparametric estimator of the full vector $\beta(\cdot)$ for the coordinates corresponding to the stable parameters.

In this paper, we generalize the approach studied in [Truquet \(2016\)](#) for estimating stable parameters in ARCH models to model (1). Our method is based on the idea originally used by [Robinson \(1988\)](#) or [Speckman \(1988\)](#) for partially linear models and consists in removing the nonparametric part of the regression function by using partial regressions. This approach is also known as profile least squares method in the literature. Under some general assumptions, we show that β_1 can be estimated at the rate \sqrt{n} . But the main contribution of this paper addresses the question of efficiency. When the noise process $(e_i)_{1 \leq i \leq n}$ is a martingale difference, we show that both methods requires the estimation of the conditional variance of the noise to improve efficiency. However, the lower bound for the asymptotic variance obtained by the averaging method of [Zhang and Wu \(2012\)](#) is larger than the method based on partial regressions. Moreover if the noise is conditionally Gaussian and the regressors exogenous (except possible lag values of y_i), the lower bound obtained with our approach can be interpreted as a lower bound in semiparametric estimation. Estimating the conditional variance and using a plug-in method to get an efficient estimate of β_1 is a very difficult task in general. This is why we only consider the case of independent errors with a time-varying unconditional variance to construct an asymptotically optimal estimator which is also efficient in the Gaussian case.

Let us mention that our conclusions for efficiency were already known for the case of a linear regression model with semi-varying coefficients, a model in which some of the coefficients depend on a covariate and the others are constant. See [Fan and Zhang \(2008\)](#) for a survey of regression models with varying coefficients and for a discussion of the non efficiency of the averaging method. The case of time-varying coefficients could be seen as a particular case of a semi-varying regression models with the time as a covariate. However, this particular case requires a special attention due to the dependence between the observations and the asymptotic with triangular arrays which lead to more technical arguments. Moreover, it is not possible to use some available results for the semiparametric efficiency such as that of [Chamberlain \(1992\)](#). Indeed, working with triangular arrays and the infill asymptotic inherent to the local stationarity assumptions requires a special attention.

In this paper, we will not study the nonparametric estimation of $\beta_2(\cdot)$ because using a plug-in of β_1 with a \sqrt{n} -consistent estimator in (2) is asymptotically equivalent to the nonparametric estimation of parameter β_2 with known coefficients β_1 . The latter problem has been extensively studied in the literature. Then the most challenging problem is the estimation of the parametric part of the regression function, a problem we focus on in this paper.

The paper is organized as follows. In Section 2, we introduce our model assumptions and we study the asymptotic properties of our estimators. In Section 3, we consider an autoregressive process with exogenous covariates and we give a discussion about asymptotic semiparametric efficiency when the errors are conditionally Gaussian. Proofs of our two main results are postponed to the last section of the paper.

2 Semiparametric estimation in time-varying regression models

Let us introduce the model considered in this paper which is quite similar to the regression model studied in [Zhang and Wu \(2012\)](#). We assume that

$$y_i = x'_{1,i}\beta_1 + x'_{2,i}\beta_2(i/n) + e_i, \quad 1 \leq i \leq n, \quad (3)$$

where e_i is a random variable taking values in \mathbb{R} , $x_{1,i}$ and $x_{2,i}$ are two random vectors of size p_1 and p_2 respectively and $\beta_1 \in \mathbb{R}^{p_1}$, β_2 is a function defined on the unit interval $[0, 1]$ and taking values in \mathbb{R}^{p_2} . In the sequel, we set $p = p_1 + p_2$. Now we introduce the dependence structure of $(x'_{1,i}, x'_{2,i}, e_i)'$ as well as the functional dependence measure used in [Zhang and Wu \(2012\)](#). Let $(\xi_k)_{k \in \mathbb{Z}}$ a sequence of i.i.d random variables taking values in an arbitrary measurable space (E, \mathcal{E}) . For $k \in \mathbb{Z}$, we set $\mathcal{F}_k = \sigma(\xi_j : j \leq k)$. We also assume that there exist measurable functions $G_{n,i} : (\mathbb{R}^p)^{\mathbb{N}} \rightarrow \mathbb{R}^p$ and $H_{n,i} : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ such that

$$x_i = \begin{pmatrix} x_{1,i} \\ x_{2,i} \end{pmatrix} = G_{n,i}(\mathcal{F}_i), \quad e_i = H_{n,i}(\mathcal{F}_i).$$

If $\mathcal{F}_{0,k} = \sigma(\xi_0, \dots, \xi_{-k+1}, \xi'_{-k}, \xi_{-k-1}, \dots)$ for $k \in \mathbb{N}$, we set

$$\delta_{k,q}(G) = \sup_{n \geq 1} \max_{1 \leq i \leq n} \|G_{n,i}(\mathcal{F}_0) - G_{n,i}(\mathcal{F}_{0,k})\|_q,$$

$$\Theta_{m,q}(G) = \sum_{k \geq m} \delta_{k,q}(G).$$

In the sequel, we will often use some conditions of type $\Theta_{0,q}(G) < \infty$. In this case, it will be assumed implicitly that $\max_{n \geq 1, 1 \leq i \leq n} \|G_{n,i}\|_q < \infty$. Note that model (3) is quite general because it contains standard regression models, the autoregressive processes studied in [Zhang and Wu \(2012\)](#) and more generally autoregressive processes with exogenous covariates. The latter case will be studied in Section 3. The assumptions we will use in the sequel allow to consider locally stationary processes for the covariates and the noise component. This includes the moving averages considered in [Dahlhaus \(1997\)](#), the ARCH processes considered in [Dahlhaus and Subba Rao \(2006\)](#) or other Markov processes studied in [Subba Rao \(2006\)](#) and more recently in [Dahlhaus et al. \(2017\)](#). In particular the noise component e_i can be conditionally heteroscedastic.

2.1 Estimation of β_1

Throughout this paper, K will denote a probability density supported on $[-1, 1]$ and Lipschitz continuous. We set for $1 \leq i, j \leq n$,

$$k_{i,j} = \frac{\frac{1}{nb} K\left(\frac{i-j}{nb}\right)}{\frac{1}{nb} \sum_{\ell=1}^n K\left(\frac{i-\ell}{nb}\right)}.$$

We set

$$s_{1,i} = \mathbb{E}(x_{2,i}y_i), \quad s_{2,i} = \mathbb{E}(x_{2,i}x'_{1,i}), \quad s_{3,i} = \mathbb{E}(x_{2,i}x'_{2,i}),$$

$$q_{1,i} = s_{3,i}^{-1}s_{1,i}, \quad q_{2,i} = s_{3,i}^{-1}s_{2,i},$$

$$\widetilde{y}_i = y_i - x'_{2,i}q_{1,i}, \quad \widetilde{x}_{1,i} = x_{1,i} - q'_{2,i}x_{2,i}.$$

We have

$$s_{1,i} = s_{2,i}\beta_1 + s_{3,i}\beta_2(i/n).$$

This yields to the relations

$$q_{1,i} = q_{2,i}\beta_1 + \beta_2(i/n)$$

and

$$\widetilde{y}_i = \widetilde{x}'_{1,i}\beta_1 + e_i.$$

The quantities \widetilde{y}_i and $\widetilde{x}_{1,i}$ will be estimated by

$$\widehat{y}_i = y_i - x'_{2,i}\widehat{q}_{1,i}, \quad \widehat{x}_{1,i} = x_{1,i} - \widehat{q}'_{2,i}x_{2,i},$$

where

$$\widehat{s}_{1,i} = \sum_{j=1}^n k_{i,j}x_{2,j}y_j, \quad \widehat{s}_{2,i} = \sum_{j=1}^n k_{i,j}x_{2,j}x'_{1,j}, \quad \widehat{s}_{3,i} = \sum_{j=1}^n k_{i,j}x_{2,i}x'_{2,i},$$

$$\widehat{q}_{1,i} = \widehat{s}_{3,i}^{-1}\widehat{s}_{1,i}, \quad \widehat{q}_{2,i} = \widehat{s}_{3,i}^{-1}\widehat{s}_{2,i}.$$

We now define the estimator of parameter β_1 .

$$\widehat{\beta}_1 = \left(\sum_{i=1}^n \widehat{x}_{1,i}\widehat{x}'_{1,i} \right)^{-1} \sum_{i=1}^n \widehat{x}_{1,i}\widehat{y}_i.$$

Our estimator can be interpreted in two ways.

- The first interpretation concerns partial regressions. The terms $x'_{2,i}q_{1,i}$ and $x'_{2,i}q_{2,i}$ can be interpreted as the orthogonal projection of y_i and $x_{1,i}$ respectively onto the linear \mathbb{L}^2 space generated by the components of $x_{2,i}$. Then \widetilde{y}_i and $\widetilde{x}_{1,i}$ are the residuals components of these two regressions and a new regression is made for estimating β_1 . In practice, it is necessary to estimate non parametrically these quantities by \widehat{y}_i and $\widehat{x}_{1,i}$ respectively. Then our estimator has the same interpretation than in the partially linear models studied in [Robinson \(1988\)](#) or [Speckman \(1988\)](#), except that the response y_i and the regressors $x_{1,i}$ are projected onto the subspace of linear functions in $x_{2,i}$.
- The second interpretation is the profile least squares method studied for instance in [Fan and Huang \(2005\)](#) for semi-varying coefficients models. The quantity $\widehat{q}_{1,i} - \widehat{q}_{1,i}\beta_1$ is the Nadaraya-Watson estimator of the function β_2 for the model

$$y_i - x'_{1,i}\beta_1 = x'_{2,i}\beta_2(i/n) + e_i.$$

Then $\widehat{\beta}_1$ is defined as a minimizer of

$$\alpha \mapsto \sum_{i=1}^n \left(y_i - x'_{1,i}\beta_1 - x'_{2,i} [\widehat{q}_{1,i} - \widehat{q}_{1,i}\beta_1] \right)^2.$$

On can also use local polynomials in the first step if more regularity is assumed for β_2 , see the remark after [Theorem 1](#).

Let d be any positive integer. For a vector v of \mathbb{R}^d or a square matrix A of size $d \times d$, we will denote by $\|v\|$ the euclidean norm of v and by $\|A\|$ the corresponding operator norm of the matrix A . For a random vector U or arbitrary dimension, we write $U \in \mathbb{L}^r$, $r > 0$ if $\|U\|_r = \mathbb{E}^{1/r}(\|U\|^r) < \infty$. We will use the following assumptions.

A1 β is a Lipschitz function.

A2 There exist two measurable functions $G : [0, 1] \times (\mathbb{R}^d)^{\mathbb{N}} \rightarrow \mathbb{R}^p$, $H : [0, 1] \times \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ and a positive constant C such that, if $x_i(u) = G(u, \mathcal{F}_i)$ and $e_i(u) = H(u, \mathcal{F}_i)$,

$$\max_{1 \leq i \leq n} [\|x_i - x_i(i/n)\|_2 + \|x_i e_i - x_i(i/n) e_i(i/n)\|_2] = O\left(\frac{1}{n}\right),$$

and

$$\|x_i(u) - x_i(v)\|_2 + \|x_i(u) e_i(u) - x_i(v) e_i(v)\|_2 \leq C|u - v|.$$

A3 There exist $q > 8/7$ and $q_1 > 16/7$, $q_2 > 2$ such that $\frac{1}{q} = \frac{1}{q_1} + \frac{1}{q_2}$ and $\Theta_{0,2q_1}(G) + \Theta_{0,q_2}(L) < \infty$, with $L_{n,i} = H_{n,i} G_{n,i}$.

A4 For each $u \in [0, 1]$, we have $\det \mathbb{E} [G(u, \mathcal{F}_k) G(u, \mathcal{F}_k)'] \neq 0$.

Notes

1. Without additional assumptions, the quantity $\hat{s}_{3,i}$ can be degenerate for finite samples (e.g discrete covariates with the value 0 in the support of the distribution). Assumption **A4** guarantees it is not the case asymptotically. However one can always replace $\hat{s}_{3,i}$ by $\hat{s}_{3,i} + \mu_n I_{p_2}$ where I_{p_2} denotes the identity matrix of size p_2 and $(\mu_n)_n$ is a sequence of positive real numbers and decreasing to 0 at a rate which is sufficiently fast to not modify our asymptotic results (e.g $\mu_n = 1/n$). The new matrix is now positive definite and then invertible. Since this modification is only technical and complicates our notations, we will not use it.
2. Assumption **A2** is a local stationarity property. This assumption is satisfied for time-varying autoregressive processes or moving averages with time-varying coefficients. With respect to [Zhang and Wu \(2012\)](#), we directly use the functional dependence coefficient for the triangular arrays $G_{n,i}$, $H_{n,i}$ instead of their local stationary approximations $G(i/n, \mathcal{F}_i)$ and $H(i/n, \mathcal{F}_i)$. With this choice, we avoid to state a second result ensuring that the triangular arrays can be replaced in our estimators without changing their asymptotic behavior. Note that when $(G_{n,i}, H_{n,i}) = (G(i/n, \cdot), H(i/n, \cdot))$, we recover the functional dependence coefficients used in [Zhang and Wu \(2012\)](#).
3. The dependence/moment conditions given in assumption **A3** is slightly more restrictive than that of [Zhang and Wu \(2012\)](#) which is guaranteed if there exists $q > 1$ such that $\Theta_{0,4}(G) + \Theta_{0,2q}(L) < \infty$. For instance, taking $q_1 = q_2 = 2q$, our conditions are satisfied provided $\Theta_{0,4q}(G) + \Theta_{0,2q}(L) < \infty$ for $q > \frac{8}{7}$.
4. Sufficient conditions for **A2** – **A3** are given by

$$\|x_i - x_i(u)\|_4 + \|e_i - e_i(u)\|_4 \leq C \left[|u - \frac{i}{n}| + \frac{1}{n} \right],$$

$$\|x_i(u) - x_i(v)\|_4 + \|e_i(u) - e_i(v)\|_4 \leq C|u - v|$$

and $\Theta_{0,4q}(G) + \Theta_{0,4q}(H) < \infty$.

For $\ell = 1, 2$, let us also set $x_{\ell,i}(u) = G_\ell(u, \mathcal{F}_i)$, $e_i(u) = H(u, \mathcal{F}_i)$ and $y_i(u) = x_{1,i}(u)' \beta_1 + x_{2,i}(u)' \beta_2(u) + e_i(u)$. We also set

$$q_1(u) = \mathbb{E}^{-1} [x_{2,i}(u)x_{2,i}(u)'] \mathbb{E} [x_{2,i}(u)y_i(u)], \quad q_2(u) = \mathbb{E}^{-1} [x_{2,i}(u)x_{2,i}(u)'] \mathbb{E} [x_{2,i}(u)x_{1,i}(u)'],$$

$$\tilde{x}_{1,i}(u) = x_{1,i}(u) - q_2(u)' x_{2,i}(u).$$

Theorem 1. Assume **A1** – **A4**, $\Theta_{n,2q_1}(G) = O(n^{-\nu_1})$ for $\nu_1 > \frac{1}{2} - \frac{1}{q_1}$ and $\Theta_{n,q_2}(L) = O(n^{-\nu_2})$ for $\nu_2 > \frac{1}{2} - \frac{1}{q_2}$. Then if $b \asymp n^{-c}$ with $1/4 < c < \min \left\{ \frac{3}{4} - q_1^{-1}, \frac{3}{4} - q_2^{-1}, 2 - \frac{2}{q \wedge (4/3)}, 2 - \frac{4}{q_1 \wedge (8/3)}, \frac{1}{2} \right\}$, we have

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \Rightarrow \mathcal{N}(0, \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1}),$$

with

$$\Sigma_1 = \int_0^1 \mathbb{E} [\tilde{x}_{1,i}(u) \tilde{x}_{1,i}(u)'] du, \quad \Sigma_2 = \int_0^1 \sum_{\ell \in \mathbb{Z}} \text{Cov} [\tilde{x}_{1,0}(u) e_0(u), \tilde{x}_{1,\ell}(u) e_\ell(u)] du.$$

Notes

1. The optimal bandwidth for estimating Lipschitz time-varying parameters is $b \asymp n^{-1/3}$. When $q_1 = q_2 = 2q$ and $q > 6/5$, this choice is allowed for applying Theorem 1.
2. It is also possible to use local polynomials of order 1 for removing the nuisance term $x'_{i,2} \beta_2(i/n)$. To this end, let us define for $\ell = 0, 1, 2$,

$$S_{\ell,i} = \sum_{j=1}^n k_{i,j} x_{2,j} x'_{2,j} \left(\frac{j-i}{nb} \right)^\ell.$$

Then we define $\hat{q}_{1,i}$ and $\hat{q}_{2,i}$ as the first components of the two block matrices

$$\begin{pmatrix} S_{0,i} & S_{1,i} \\ S_{1,i} & S_{2,i} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n k_{i,j} x_{2,j} y_j \\ \sum_{j=1}^n k_{i,j} x_{2,j} y_j \frac{j-i}{nb} \end{pmatrix} \text{ and } \begin{pmatrix} S_{0,i} & S_{1,i} \\ S_{1,i} & S_{2,i} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n k_{i,j} x_{2,j} x'_{1,j} \\ \sum_{j=1}^n k_{i,j} x_{2,j} x'_{1,j} \frac{j-i}{nb} \end{pmatrix}$$

respectively. Moreover, let us denote by $\tilde{q}_{1,i}$ and $\tilde{q}_{2,i}$ the first components of the two block matrices

$$\begin{pmatrix} \mathbb{E} S_{0,i} & \mathbb{E} S_{1,i} \\ \mathbb{E} S_{1,i} & \mathbb{E} S_{2,i} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n k_{i,j} \mathbb{E} [x_{2,j} y_j] \\ \sum_{j=1}^n k_{i,j} \mathbb{E} [x_{2,j} y_j \frac{j-i}{nb}] \end{pmatrix}, \quad \begin{pmatrix} \mathbb{E} S_{0,i} & \mathbb{E} S_{1,i} \\ \mathbb{E} S_{1,i} & \mathbb{E} S_{2,i} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^n k_{i,j} \mathbb{E} [x_{2,j} x'_{1,j}] \\ \sum_{j=1}^n k_{i,j} \mathbb{E} [x_{2,j} x'_{1,j} \frac{j-i}{nb}] \end{pmatrix}$$

respectively. Then, under the assumptions of Theorem 1, one can show that

$$\max_{1 \leq i \leq n} |\tilde{q}_{1,i} - q_{1,i}| + \max_{1 \leq i \leq n} |\tilde{q}_{2,i} - q_{2,i}| = O\left(b + \frac{1}{n}\right).$$

The term $1/n$ corresponds to the deviation with respect to stationarity, i.e the order of the differences $q_{1,i} - q_1(i/n)$ and $q_{2,i} - q_2(i/n)$. Moreover we have $\beta_2(i/n) = q_{1,i} - q_{2,i} \beta_1$ and if the second derivative of β_2 exists and is continuous, we get a smaller bias for β_2 :

$$\max_{1 \leq i \leq n} |\tilde{q}_{1,i} - q_{1,i} - (\tilde{q}_{2,i} - q_{2,i}) \beta_1| = O\left(b^2 + \frac{1}{n}\right).$$

If we define an estimator of β_1 with these new expressions for $\hat{q}_{1,i}$ and $\hat{q}_{2,i}$, we obtain the profile least squares estimators studied in [Fan and Huang \(2005\)](#) for semi-varying coefficients regression models. Under the assumptions of [Theorem 1](#), the estimator $\hat{\beta}_1$ is \sqrt{n} -consistent but some assumptions can be relaxed. Inspection of the proof of [Theorem 1](#) shows that the only change due to this smaller bias concerns the term N_2 and the bandwidth condition $nb^4 \rightarrow 0$ can be replaced by $nb^6 \rightarrow 0$ (the bias is of order b for $\hat{q}_{2,i}$ and of order b^2 for \hat{r}_i). This leads to the condition $c > 1/6$ for the bandwidth. Moreover, one can choose $q > 12/11$ and $q_1 > 24/11$ in assumption [A3](#) to get a compatible upper bound for c in [Theorem 1](#).

2.2 Improving efficiency when the noise process is a martingale difference

In this subsection, we assume the following martingale difference structure for the noise component e_i :

$$\mathbb{E}(e_i | \mathcal{F}_{i-1}) = 0, \quad \text{Var}(e_i | \mathcal{F}_{i-1}) = \sigma^2(i/n, x_i),$$

where $\sigma : [0, 1] \times \mathbb{R}^p \rightarrow \mathbb{R}_+^*$ is an unknown function. We also assume that $G_{n,i}(\mathcal{F}_i) = G_{n,i}(\mathcal{F}_{i-1})$. For simplicity of notations, we set $\sigma_i = \sigma(i/n, x_i)$. Now let $p_i = p(i/n, x_i)$ a weight function. One can write a second regression model by multiplying the first one by the weights p_i ,

$$p_i y_i = p_i x'_{1,i} \beta_1 + p_i x'_{2,i} \beta_2(i/n) + p_i e_i, \quad 1 \leq i \leq n. \quad (4)$$

In the sequel, we assume that the assumptions [A1](#) – [A4](#) are satisfied for this new regression model. The asymptotic variance given in [Theorem 1](#) has a simpler form. In particular,

$$\Sigma_2 = \mathbb{E} \int_0^1 p_i^2(u) \sigma_i^2(u) \widetilde{p} x_{1,i}(u) \widetilde{p} x_{1,i}(u)' du,$$

where $p_i(u) = p(u, x_i(u))$, $\sigma_i(u) = \sigma(u, x_i(u))$ and px denotes the vector $(p_i x_i)_{1 \leq i \leq n}$. Now, if we set

$$q_2(u) = \mathbb{E}^{-1} \left[p_i(u)^2 x_{2,i}(u) x_{2,i}(u)' \right] \cdot \mathbb{E} \left[p_i(u)^2 x_{2,i}(u) x_{1,i}(u)' \right],$$

$$q_2^*(u) = \mathbb{E}^{-1} \left[\sigma_i(u)^{-2} x_{2,i}(u) x_{2,i}(u)' \right] \cdot \mathbb{E} \left[\sigma_i(u)^{-2} x_{2,i}(u) x_{1,i}(u)' \right]$$

and changing our notations with

$$\widetilde{x}_{1,i}(u) = x_{1,i}(u) - q_2(u)' x_{2,i}(u), \quad \widetilde{x}_{1,i}^*(u) = x_{1,i}(u) - q_2^*(u)' x_{2,i}(u),$$

we have

$$\Sigma_1 = \mathbb{E} \int_0^1 p_i(u)^2 \widetilde{x}_{1,i}(u) \widetilde{x}_{1,i}(u)' du = \mathbb{E} \int_0^1 p_i(u)^2 \widetilde{x}_{1,i}(u) \widetilde{x}_{1,i}^*(u)' du = \mathbb{E} \int_0^1 p_i(u)^2 \sigma_i(u) \widetilde{x}_{1,i}(u) \frac{\widetilde{x}_{1,i}^*(u)'}{\sigma_i(u)} du.$$

From the Cauchy-Schwarz inequality, we have for $x, y \in \mathbb{R}^{p_1}$,

$$(x' \Sigma_1 y)^2 \leq x' \Sigma_2 x \cdot y' \Sigma^* y,$$

where

$$\Sigma^* = \mathbb{E} \int_0^1 \frac{\widetilde{x}_{1,i}^*(u) \widetilde{x}_{1,i}^*(u)'}{\sigma_i(u)^2} du. \quad (5)$$

Now setting $x = \Sigma_1^{-1}z$ and $y = (\Sigma^*)^{-1}z$ for $z \in \mathbb{R}^{p_1}$, we get

$$z' (\Sigma^*)^{-1} z \leq z' \Sigma_1^{-1} \Sigma_2 \Sigma_1^{-1} z.$$

Note that $(\Sigma^*)^{-1}$ is the asymptotic variance corresponding to the choice $p_i = \sigma_i^{-1}$. This means that weighted least squares can improve efficiency of our estimator, provided to first estimate the conditional variance of the noise. In the subsection 2.4, we will construct an efficient estimator when the errors are independent but have a time-varying unconditional variance.

2.3 Comparison with a related method

Zhang and Wu (2012) have proposed another method for the estimation of parameter β_1 . Their method consists in first estimating the full vector $\beta(\cdot)$ of coefficients using local polynomials and then averaging the first component of this estimator, i.e $\hat{\beta}_1 = \int_0^1 \hat{\beta}_1(u) du$. They have shown that under some assumptions, $\hat{\beta}_1$ is root- n consistent. Comparing efficiency of both methods seems difficult for a general noise process. Nevertheless, one can get interesting properties when the noise is a martingale difference. To this end, we keep the notations of the previous subsection. Application of the method of Zhang and Wu (2012) in the regression model (4) leads to an asymptotic variance S given by (see Theorem 3.1 of that paper):

$$S = \int_0^1 AS_1(u)^{-1} S_2(u) S_1(u)^{-1} A' du,$$

where

$$S_1(u) = \mathbb{E} \left[p_i(u)^2 x_i(u) x_i(u)' \right], \quad S_2(u) = \mathbb{E} \left[p_i(u)^4 \sigma_i(u)^2 x_i(u) x_i(u)' \right]$$

and A is a matrix of size $p_1 \times p$ and which associates the p_1 first components of a vector of \mathbb{R}^p . Note that when $p_i = \sigma_i^{-1}$, the asymptotic variance reduces to $\int_0^1 AS^*(u)^{-1} A' du$ where $S^*(u) = \mathbb{E} \left[\frac{x_i(u) x_i(u)'}{\sigma_i(u)^2} \right]$. Using the formula for the inverse of a matrix defined by blocs, one can check that $AS^*(u)^{-1} A' = \mathbb{E}^{-1} \left[\frac{h_i^*(u) h_i^*(u)'}{\sigma_i(u)^2} \right]$. Using Lemma 6, we have

$$\int_0^1 AS^*(u)^{-1} A' \geq (\Sigma^*)^{-1}. \quad (6)$$

Moreover, the matrix $\int_0^1 AS^*(u)^{-1} A'$ is optimal for the estimator of Zhang and Wu (2012). Indeed we have from the Cauchy-Schwarz inequality

$$(x' S_1(u) x)^2 \leq x' S_2(u) x \cdot y' S^*(u) y$$

and the result follows by taking $x = S_1(u)^{-1}z$ and $y = S^*(u)^{-1}z$. Then we obtain the following important conclusion for a martingale difference noise: removing the nonparametric component with partial regressions can lead to a more efficient estimator than averaging directly the nonparametric estimator. In the next section, we show that an AR process with a conditionally Gaussian noise and some strongly exogenous covariates, the matrix $(\Sigma^*)^{-1}$ is a lower bound in semiparametric estimation.

2.4 Weighted least squares and asymptotic efficiency improvement

In this subsection, we consider the simple case $e_i = \sigma(i/n)f(\xi_i)$ with $\mathbb{E}f(\xi_i) = 0$ and $\text{Var}f(\xi_i) = 1$. In this case, we first derive an estimator of $\sigma^2(i/n)$. A natural candidate is

$$\hat{\sigma}^2(i/n) = \sum_{j=1}^n k_{i,j} [y_j - x'_j \hat{\beta}(i/n)]^2,$$

where $\hat{\beta}$ is an estimate of β . We will use the simple Nadaraya-Watson estimator of β , i.e

$$\hat{\beta}(i/n) = \left(\sum_{j=1}^n k_{i,j} x_j x'_j \right)^{-1} \sum_{j=1}^n k_{i,j} x_j y_j.$$

Another solution is to take the semiparametric estimator $\hat{\beta}_1$ of β_1 which is \sqrt{n} -consistent under the assumptions of Theorem 1 and to define an estimator of $\beta_2(i/n)$ as follows:

$$\hat{\beta}_2(i/n) = \arg \min_{\alpha \in \mathbb{R}^{p_2}} \sum_{j=1}^n k_{i,j} [y_i - x'_{i,1} \hat{\beta}_1 - x'_{i,2} \alpha]^2.$$

But this second choice will not improve our asymptotic results and for simplicity we only consider the Nadaraya-Watson estimator for the plug-in. Our plug-in estimator is defined as follows. For simplicity we will simply note σ_i (resp. $\hat{\sigma}_i$) for $\sigma(i/n)$ (resp. $\hat{\sigma}(i/n)$). We set

$$\hat{s}_{1,i}^* = \sum_{j=1}^n k_{i,j} \frac{x_{2,j} y_j}{\hat{\sigma}_j^2}, \quad \hat{s}_{2,i}^* = \sum_{j=1}^n k_{i,j} \frac{x_{2,j} x'_{1,j}}{\hat{\sigma}_j^2}, \quad \hat{s}_{3,i}^* = \sum_{j=1}^n k_{i,j} \frac{x_{2,j} x'_{2,j}}{\hat{\sigma}_j^2},$$

$$\hat{q}_{1,i}^* = (\hat{s}_{3,i}^*)^{-1} \hat{s}_{1,i}^*, \quad \hat{q}_{2,i}^* = (\hat{s}_{3,i}^*)^{-1} \hat{s}_{2,i}^*$$

and

$$\hat{y}_i^* = y_i - x'_{2,i} \hat{q}_{1,i}^*, \quad \hat{x}_{1,i}^* = x_{1,i} - (\hat{q}_{2,i}^*)' x_{2,i}.$$

The optimal estimator of β_1 is defined by

$$\hat{\beta}_1^* = \left(\sum_{i=1}^n \frac{\hat{x}_{1,i}^* (\hat{x}_{1,i}^*)'}{\hat{\sigma}_i^2} \right)^{-1} \sum_{i=1}^n \frac{\hat{x}_{1,i}^* \hat{y}_i^*}{\hat{\sigma}_i^2}.$$

We have the following result.

Theorem 2. *Assume that all the assumptions A1-A4 hold true with $q_2 > 4$ and σ is positive and Lipschitz over $[0, 1]$. Then*

$$\sqrt{n}(\hat{\beta}_1^* - \beta_1) \Rightarrow \mathcal{N}(0, (\Sigma^*)^{-1}),$$

where Σ^* is defined in (5).

3 Asymptotic semiparametric efficiency for an autoregressive process with exogenous covariates

3.1 An autoregressive process with covariates

We consider a process $(z_i)_{1 \leq i \leq n}$ such that $z_i = G_{n,i}^{(2)}(\mathcal{F}_i)$ and an error process $(e_i)_{1 \leq i \leq n}$ defined as in Section 2. We define a process $(y_i)_{1 \leq i \leq n}$ recursively by

$$y_i = \sum_{j=1}^q a_j(i/n)y_{i-j} + z_i' \gamma(i/n) + e_i$$

and we set $x_i = (y_{i-1}, \dots, y_{i-q}, z_i)'$ and $p = q + d$. Finally, we set $\beta(u) = (a_1(u), \dots, a_q(u), \gamma(u)')'$ for the parameter of interest. The following assumptions will be needed.

B1 The function $u \mapsto \beta(u)$ is Lipschitz. Moreover, for each $u \in [0, 1]$, the roots of the polynomial $\mathcal{P}_u(z) = 1 - \sum_{j=1}^q a_j(u)z^j$ are outside the unit disc.

B2 There exists a measurable function $M : [0, 1] \times E^{\mathbb{N}} \rightarrow \mathbb{R}$ such that if $z_i(u) = M(u, \mathcal{F}_i)$,

$$\max_{1 \leq i \leq n} [\|z_i - z_i(i/n)\|_4 + \|e_i - e_i(i/n)\|_4] = O\left(\frac{1}{n}\right),$$

and

$$\|z_i(u) - z_i(v)\|_4 + \|e_i(u) - e_i(v)\|_4 \leq C|u - v|.$$

B3 There exist $q > \frac{8}{7}$ such that $\Theta_{0,4q}(M) + \Theta_{0,4q}(H) < \infty$.

Under the previous assumptions, we show that $(y_i)_{1 \leq i \leq n}$ can be approximated locally by the stationary process

$$y_i(u) = \sum_{j=1}^q a_j(u)y_{i-j}(u) + z_i(u)' \gamma(u) + e_i(u), \quad i \in \mathbb{Z}, u \in [0, 1].$$

For initialization, we assume that for $i \leq 0$, $y_i = y_i(0)$. Note that for integers $n \geq 1$ and $1 \leq i \leq n$, there exists an application $N_{n,i} : (\mathbb{R}^d)^{\mathbb{N}} \rightarrow \mathbb{R}$ such that $y_i = N_{n,i}(\mathcal{F}_i)$. Moreover, there exists also an application $N : [0, 1] \times (\mathbb{R}^d)^{\mathbb{N}} \rightarrow \mathbb{R}$ such that $y_i(u) = N(u, \mathcal{F}_i)$.

Proposition 1. *Under the assumptions B1-B3, there exists a constant $C > 0$ such that for all $(u, v, i) \in [0, 1]^2 \times \{1, 2, \dots, n\}$,*

$$\|y_i - y_i(u)\|_4 \leq C \left[|u - \frac{i}{n}| + \frac{1}{n} \right], \quad \|y_i(u) - y_i(v)\|_4 \leq C|u - v|.$$

Moreover, we have $\Theta_{0,4q}(N) < \infty$.

Note. Assumptions **B1-B3** entail assumptions **A1-A3**. There is no general arguments for checking **A4**. However, if **A4** is not valid then we have $\sum_{j=1}^q \lambda_j y_{i-j}(u) + \mu' z_i(u) = 0$ a.s for some $u \in [0, 1]$ and some constants λ_j and μ . This type of equality cannot hold in general. For instance, assumption **A4** is automatic for a non degenerate noise such that the e'_i s are independent and independent from the z'_i s.

Proof of Proposition 1 Let us introduce the following notations. For $i \leq n$, we set

$$Y_i = (y_i, y_{i-1}, \dots, y_{i-q+1})', \quad B_i = (z_i' \gamma(i/n) + e_i, \mathbf{0}'_{1,q-1})',$$

with the convention that $y_i = y_i(0)$ and $(\gamma(i/n)', e_i, z_i') = (\gamma(0)', e_i(0), z_i(0))$ when $i \leq 0$. Moreover we denote by $A(i/n)$ the companion matrix associated to the polynomial $\mathcal{P}_{i/n}$. Then, for $q+1 \leq i \leq n$, we have $Y_i = A(i/n)Y_{i-1} + B_i$. From the assumption **B2**, for each $u \in [0, 1]$, the spectral radius of the matrix $A(u)$ is smaller than 1. Then, one can consider

$$m_u = \inf \{k \geq 1 : |A(u)^k| < 1\} < \infty.$$

Now, let $\mathcal{O}_u = \{v \in [0, 1] : |A(v)^{m_u}| < 1\}$. The set \mathcal{O}_u is an open subset of $[0, 1]$. From the compactness of the unit interval, we have $[0, 1] = \cup_{i=1}^h \mathcal{O}_{u_i}$ for suitable points u_1, \dots, u_h in $[0, 1]$. Then if m denotes the lowest common multiple of the integers m_{u_1}, \dots, m_{u_h} , we have $\max_{u \in [0, 1]} |A(u)^m| < 1$. From the uniform continuity of the application $(u_1, \dots, u_h) \mapsto A(u_1) \cdots A(u_h)$, we have, if n is large enough,

$$\rho = \max_{m-1 \leq k \leq n} \left| A\left(\frac{k}{n}\right) A\left(\frac{k-1}{n}\right) \cdots A\left(\frac{k-m+1}{n}\right) \right| < 1.$$

Moreover, we have $\sup_{u \in [0, 1]} \|Y_i(u)\|_4 < \infty$. From our assumptions, we deduce that there exists a positive real number D such that,

$$\begin{aligned} \|Y_i - Y_i(u)\|_4 &\leq \rho \|Y_{i-m} - Y_{i-m}(u)\|_4 + D \left[\left| \frac{i}{n} - u \right| + \frac{1}{n} \right] \\ &\leq D \sum_{h=0}^{\infty} \rho^h \left[\left| u - \frac{i-h\ell}{n} \right| + \frac{1}{n} \right], \end{aligned}$$

which leads to the first part of the proposition, the bound for $\|Y_i(u) - Y_i(v)\|_4$ can be obtained in the same way. For the second part, we use the expansion

$$Y_i = B_i + \sum_{h=1}^{\infty} A\left(\frac{i}{n}\right) \cdots A\left(\frac{i-h+1}{n}\right) B_{i-h}.$$

Then we obtain $\Theta_{0,4q}(N) < \infty$ by using the contraction coefficient ρ and the assumption $\Theta_{0,4q}(M) + \Theta_{0,4q}(H) < \infty$. \square

3.2 Gaussian inputs and efficient information matrix

In this subsection, we assume that $e_i = \sigma(i/n, z_i) f(\xi_i)$ with $f(\xi_i) \sim \mathcal{N}(0, 1)$ and the two processes $(z_i)_{1 \leq i \leq n}$ and $(f(\xi_i))_{i \in \mathbb{Z}}$ are independent. This independence property occurs for example when $\xi_i = (\xi_i^{(1)}, \xi_i^{(2)})$, with $\xi_i^{(1)}$ and $\xi_i^{(2)}$ independent, $f(\xi_i) = \xi_i^{(1)}$ and z_i only depending on $(\xi_j^{(2)})_{j \leq i}$. This type of condition on the covariates is called strong exogeneity (sometimes Granger or Sims causality in the literature) and is very useful to get a simple expression of the conditional likelihood function. We will assume that the conditional standard deviation $\sigma_i = \sigma(i/n, z_i)$ satisfies $\inf_{n \geq 1, 1 \leq i \leq n} \sigma_i \geq \sigma_- > 0$ a.s. for a constant σ_- . One could also allow σ_i to depend on past values of y_i but it is necessary to introduce additional regularity conditions to define the model. This is why we prefer to consider conditional heteroscedasticity only with respect to the covariates.

We assume that the parameter $\beta(u)$ writes as $(\beta_1, \beta_2(u))$ with $\beta_1 \in \mathbb{R}^{p_1}$ is constant. The coordinates of β_1 can contain some of the a'_j 's or some of the coordinates of γ . For instance, β_1 can contains all the parameters except a time-varying trend β_2 . One can also consider the case of constant lag coefficients a_j and a covariate z_i with time-varying coefficients.

We will first prove a LAN expansion of the likelihood ratio conditionally to some initial values $y_0, y_1, \dots, y_{-q+1}$ not depending on n . Let μ_n be the probability distribution of the vector $(z'_1, z'_2, \dots, z'_n)'$. Note that μ_n does not depend on a parameter of interest in this problem. We also denote by $L_{n,\beta}(\cdot | y_0^-, z_{1:n})$ the density of (y_1, y_2, \dots, y_n) conditionally to z_i and y_j for $1 \leq i \leq n$ and $-q+1 \leq j \leq 0$. Using the exogeneity property of the covariates, we have the expression

$$L_{n,\beta}(y_{1:n} | y_0^-, z_{1:n}) = \prod_{i=1}^n (2\pi\sigma_i^2)^{-1/2} \exp\left(-\frac{|y_i - x'_i\beta(i/n)|^2}{2\sigma_i^2}\right).$$

Then we set

$$dP_{n,\beta}(y_{1:n}, z_{1:n} | y_0^-) = L_{n,\beta}(y_{1:n} | y_0^-, z_{1:n}) d\mu_n(z_{1:n}) d\lambda_n(y_{1:n}),$$

where λ_n denotes the Lebesgue measure on \mathbb{R}^n . We introduce the following local parametrization. Let us denote by \mathbb{H} the Hilbert space $\mathbb{R}^{p_1} \times L^2([0, 1])$ endowed with the scalar product

$$(hg)_{\mathbb{H}} = \int_0^1 h(u)' M(u) h(u) du, \quad M(u) = \mathbb{E} \left[\frac{x_0(u) x_0(u)'}{\sigma^2(u, x_0(u))} \right].$$

We set $h(u), g(u)$ for simplicity of notations. It is intended that the p_1 -first components of $h(u)$ and $g(u)$ do not depend on u . Note that, in a parametric problem, $M(u)$ corresponds to the Fisher information matrix for the estimation of the parameter $\beta(u)$ in the model

$$y_i(u) = x'_i(u)\beta(u) + \sigma(u, z_i(u)) f(\xi_i).$$

We also set $\mathcal{H} = \mathbb{R}^{p_1} \times \mathcal{L}^{p_2}$ where \mathcal{L} denotes the set of Lipschitz functions defined over $[0, 1]$. Then if $\beta, h \in \mathcal{L} \times \mathcal{L}$, the log likelihood ratio obtained after shifting parameter β is given by

$$\log \frac{dP_{n,\beta+h/\sqrt{n}}}{dP_{n,\beta}}(y_{1:n}, z_{1:n} | y_0^-) = \sum_{i=1}^n \left[\frac{|y_i - x'_i\beta|^2}{2\sigma_i^2} - \frac{|y_i - x'_i(\beta+h)|^2}{2\sigma_i^2} \right].$$

Let us first derive a LAN expansion for this likelihood ratio.

Proposition 2. *Under the assumptions B1 – B3, we have*

$$\log \frac{dP_{n,\beta+h/\sqrt{n}}}{dP_{n,\beta}}(y_{1:n}, z_{1:n} | y_0^-) = \Delta_{n,\beta,h} - \frac{1}{2} \|h\|_{\mathbb{H}}^2 + o_{\mathbb{P}}(1),$$

where

$$\Delta_{n,\beta,h} := \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{f(\xi_i) x'_i h(i/n)}{\sigma_i} \Rightarrow \mathcal{N}(0, \|h\|_{\mathbb{H}}^2).$$

Proof of Proposition 2 We have

$$\log \frac{dP_{n,\beta+h/\sqrt{n}}}{dP_{n,\beta}}(y_{1:n}, z_{1:n}|y_0^-) = \Delta_{n,\beta,h} - \frac{1}{2}\Delta'_{n,h},$$

where

$$\Delta'_{n,h} = \frac{1}{n} \sum_{i=1}^n \frac{h'(i/n)x_i x'_i h(i/n)}{\sigma_i^2}.$$

First, one can note that the initial values $y_0, y_{-1}, \dots, y_{-q+1}$ are forgotten exponential fast. More precisely if y_i is initialized with the random variables $y_0(0), y_{-1}(0), \dots, y_{-q+1}(0)$, the \mathbb{L}^4 -norm of the difference is bounded by $C\rho^i$ with $\rho \in (0, 1)$. Hence, one can assume that $(y_i)_{1 \leq i \leq n}$ is initialized with $(y_i(0))_{-q+1 \leq i \leq 0}$. Using the assumptions **B1** – **B3** and Proposition 1, there exists a positive constant C such that for all $n \geq 1$, $1 \leq i \leq n$ and $u \in [0, 1]$,

$$\left| \mathbb{E}(\sigma_i^{-2} x_i x'_i) - \mathbb{E}(\sigma_i(u)^{-2} x_i(u) x'_i(u)) \right| \leq C [|u - i/n| + 1/n].$$

Moreover, the application $u \mapsto \mathbb{E}(\sigma_i(u)^{-2} x_i(u) x'_i(u))$ is Lipschitz and $\max_{1 \leq i \leq n} \mathbb{E} \sigma_i^{-4} |x_i|^4 = O(1)$. Then the convergence of $\Delta_{n,\beta,h}$ follows from the central limit theorem for triangular arrays of martingale differences. Moreover, using Lemma A1 of [Zhang and Wu \(2012\)](#) with $q = 2$, we have

$$\Delta'_{n,h} - \frac{1}{n} \sum_{i=1}^n h(i/n)' \mathbb{E}[\sigma_i^{-2} x_i x'_i] h(i/n) = O_{\mathbb{P}}(1/\sqrt{n}).$$

Then we deduce from the properties stated above that $\Delta'_{n,h} \rightarrow \|h\|_{\mathbb{H}}^2$. This completes the proof of Proposition 2. \square

From this LAN expansion, we deduce the following result.

Corollary 1. *The limit distribution of any regular estimator of parameter β_1 equals the distribution of $L_1 + L_2$ of independent random vectors such that*

$$L_1 \sim \mathcal{N}_{p_1}(0, (\Sigma^*)^{-1}), \quad \Sigma^* = \int_0^1 \mathbb{E}(\sigma_i(u)^{-2} \tilde{x}_{1,i}^*(u) \tilde{x}_{1,i}^{*\prime}(u)) du,$$

with

$$\tilde{x}_{1,i}^*(u) = x_{1,i}(u) - q_2^*(u)' x_{2,i}(u), \quad q_2^*(u) = \mathbb{E}^{-1} \left[\sigma_i(u)^{-2} x_{2,i}(u) x_{2,i}(u)' \right] \mathbb{E} \left[\sigma_i(u)^{-2} x_{2,i}(u) x_{1,i}(u)' \right].$$

Note. The matrix Σ^* plays the rule of an efficient information matrix. Its definition involves the whole family of stationary approximations $\{(y_i(u), z_i(u), e_i(u)) : u \in [0, 1]\}$ of our triangular array.

Proof of Corollary 1 We use the general convolution theorem given in [van der Vaart and Wellner \(1996\)](#), theorem 3.11.2, which is compatible with triangular arrays. To this end, we consider the sequence of parameters $\{\kappa_n(h) : h \in H\}$ defined by $\kappa_n(h) = \beta_1 + h_1/\sqrt{n}$ where h_1 denotes the vector of constant parameters. This sequence of parameters is regular, we have

$$\sqrt{n}(\kappa_n(h) - \kappa_n(0)) = h_1 := \dot{\kappa}(h).$$

The adjoint operator $\dot{\kappa}^* : \mathbb{R}^{p_1} \rightarrow \mathbb{H}$ of the projection operator $\dot{\kappa}$ is defined by the equalities,

$$\dot{\kappa}(h)' v = (h, \dot{\kappa}^* v)_{\mathbb{H}}, \quad v \in \mathbb{R}^{p_1}, h \in H.$$

First, we replace the two random vectors $x_{1,i}(u)/\sigma_i(u)$ and $x_{2,i}(u)/\sigma_i(u)$ by orthogonal vectors in \mathbb{L}^2 . We have

$$\frac{x_{1,i}(u)}{\sigma_i(u)} = \frac{\tilde{x}_{1,i}^*(u)}{\sigma_i(u)} + q_2^*(u)' \frac{x_{2,i}(u)}{\sigma_i(u)}.$$

Then we have

$$M(u) = \begin{pmatrix} I_{p_1} & q_2^*(u)' \\ \mathbf{0}_{p_2,p_1} & I_{p_2} \end{pmatrix} \cdot \begin{pmatrix} \mathbb{E} \left[\frac{\tilde{x}_{1,i}^*(u)\tilde{x}_{1,i}^*(u)'}{\sigma_i(u)^2} \right] & \mathbf{0}_{p_1,p_2} \\ \mathbf{0}_{p_2,p_1} & \mathbb{E} \left[\frac{x_{2,i}(u)x_{2,i}(u)'}{\sigma_i(u)^2} \right] \end{pmatrix} \cdot \begin{pmatrix} I_{p_1} & \mathbf{0}_{p_1,p_2} \\ q_2^*(u) & I_{p_2} \end{pmatrix}.$$

Then if $k^*v = (g_1', g_2(u)')'$, we get

$$h_1'v = h_1'\Sigma g_1 + \int_0^1 (h_1'q_2^*(u)' + h_2(u)') \mathbb{E} \left[\frac{x_{2,i}(u)x_{2,i}(u)'}{\sigma_i(u)^2} \right] (q_2^*(u)g_1 + g_2(u)) du.$$

Then we get $g_1 = \Sigma^{-1}v$ and $g_2(u) = -q_2^*(u)\Sigma^{-1}v$. We deduce that

$$\|k^*v\|_{\mathbb{H}} = v'\Sigma^{-1}v$$

and the convolution theorem gives the result, from the LAN expansion given in Proposition 2. \square

Note. Under the assumptions of Theorem 2 and when $\sigma_i = \sigma(i/n)$, we deduce that the estimator $\hat{\beta}_1^*$ is asymptotically efficient in a semiparametric sense. Rigorously, it is necessary to show that this estimator is regular in the sense of [van der Vaart and Wellner \(1996\)](#). This step does not present any difficulty, provided to follow carefully the proof of Theorem 2. Details are omitted.

4 Proofs of Theorem 1 and Theorem 2

4.1 Additional results for the proof of Theorem 1

In the sequel, we set $\tilde{q}_{\ell,i} = \mathbb{E}^{-1}(\hat{s}_{3,i}) \mathbb{E}(\hat{s}_{\ell,i})$ for $\ell = 1, 2$.

Lemma 1. Assume that the assumptions of Theorem 1 hold true. Setting $m_i = \mathbb{E}\hat{s}_{3,i}$, we have

1. $\max_{1 \leq i \leq n} \|m_i^{-1}\| = O(1)$, $\max_{1 \leq i \leq n} \|\hat{s}_{3,i}^{-1}\| = O_{\mathbb{P}}(1)$,

2. for $\ell = 1, 2$,

$$\max_{1 \leq i \leq n} \left| \hat{q}_{\ell,i} - \tilde{q}_{\ell,i} - m_i^{-1} [\hat{s}_{\ell,1} - \mathbb{E}\hat{s}_{\ell,j}] + m_i^{-1} [\hat{s}_{3,i} - \mathbb{E}\hat{s}_{3,i}] m_i^{-1} \mathbb{E}\hat{s}_{\ell,i} \right| = o_{\mathbb{P}} \left(\frac{1}{\sqrt{n}} \right).$$

3. for $\ell = 1, 2$, $\max_{1 \leq i \leq n} |\hat{q}_{\ell,i} - \tilde{q}_{\ell,i}| = o_{\mathbb{P}}(n^{-1/4})$.

4. For $\ell = 1, 2$, we have

$$\max_{1 \leq i \leq n} |\tilde{q}_{\ell,i} - q_{\ell,i}| = o(n^{-1/4}), \quad \max_{1 \leq i \leq n} |q_{\ell,i} - q_{\ell}(i/n)| = O(1/n).$$

Proof of Lemma 1

1. Using Lemma A1 in [Zhang and Wu \(2012\)](#), there exists $C > 0$ such that

$$\| \max_{1 \leq i \leq n} |\hat{\delta}_{3,i} - \mathbb{E}\hat{\delta}_{3,i}| \|_2 \leq C \sqrt{na_n} \Theta_{0,4}(G),$$

where

$$a_n = \max_{1 \leq i \leq n} \left[k_{i,1} + \sum_{j=1}^{n-1} |k_{i,j+1} - k_{i,j}| \right] = O\left(\frac{1}{nb}\right).$$

Using our assumptions, we deduce that

$$\max_{1 \leq i \leq n} |\hat{\delta}_{3,i} - \mathbb{E}\hat{\delta}_{3,i}| = o_{\mathbb{P}}(1). \quad (7)$$

Moreover, setting $G = (G_1, G_2)$ and $g_2(u) = \mathbb{E}G_2(u, \mathcal{F}_i) G_2(u, \mathcal{F}_i)'$, we have from assumption **A2**,

$$\max_{1 \leq i \leq n} |\mathbb{E}\hat{\delta}_{3,i} - g_2(i/n)| = O(b + 1/n). \quad (8)$$

From (7) and (8), we conclude that $\max_{1 \leq i \leq n} |\hat{\delta}_{3,i} - g_2(i/n)| = o_{\mathbb{P}}(1)$. Then the result follows from assumptions **A2** and **A4** which guarantee that $\max_{1 \leq i \leq n} |g_2(i/n)^{-1}| = O(1)$.

2. One can check that the quantity inside the norm is equal to

$$\begin{aligned} d_i &= \hat{\delta}_{3,i}^{-1} (\hat{\delta}_{3,i} - \mathbb{E}\hat{\delta}_{3,i}) m_i^{-1} (\hat{\delta}_{3,i} - \mathbb{E}\hat{\delta}_{3,i}) m_i^{-1} \hat{\delta}_{\ell,i} \\ &+ m_i^{-1} (\mathbb{E}\hat{\delta}_{3,i} - \hat{\delta}_{3,i}) m_i^{-1} (\hat{\delta}_{\ell,i} - \mathbb{E}\hat{\delta}_{\ell,i}). \end{aligned}$$

Using Lemma 3 and the previous point of this lemma, we get

$$\max_{1 \leq i \leq n} |d_i| = O_{\mathbb{P}}\left(\frac{v_n^2}{(nb)^2}\right), \text{ where } v_n = n^{1/q_1} + n^{1/q_2} + \sqrt{nb \log n}$$

and the result follows from our bandwidth conditions.

3. The result follows from the previous point, Lemma 3 and our bandwidth conditions.
 4. As in the proof of point 1., for $\ell = 1, 2, 3$, we have from assumption **A2**

$$\max_{1 \leq i \leq n} \left| \mathbb{E}\hat{\delta}_{\ell,i} - \sum_{j=1}^n k_{i,j} s_{\ell}(j/n) \right| = O(1/n), \quad \max_{1 \leq i \leq n} \left| \sum_{j=1}^n k_{i,j} s_{\ell}(j/n) - s_{\ell}(i/n) \right| = O(b).$$

Then the result follows from assumption **A4** and the condition $b = o(n^{-1/4})$. \square

4.2 Proof of Theorem 1

We set $D_n = \sum_{i=1}^n \hat{x}_{1,i} \hat{x}'_{1,i}$. Using the decomposition

$$\hat{y}_i = \hat{x}'_{1,i} \beta_1 + x'_{2,i} \hat{r}_i + e_i, \quad \hat{r}_i = (\hat{q}_{2,i} - q_{2,i}) \beta_1 - (\hat{q}_{1,i} - q_{1,i}),$$

we get

$$\hat{\beta}_1 - \beta_1 = D_n^{-1} \sum_{i=1}^n \hat{x}_{1,i} [x'_{2,i} \hat{r}_i + e_i].$$

Using now the decomposition $\hat{x}_{1,i} = \tilde{x}_{1,i} - (\hat{q}_{2,i} - q_{2,i})' x_{2,i}$, we have

$$\hat{\beta}_1 - \beta_1 = D_n^{-1} [N_1 - N_2 + N_3 - N_4],$$

where

$$\begin{aligned} N_1 &= \sum_{i=1}^n \tilde{x}_{1,i} x'_{2,i} \hat{r}_i, & N_2 &= \sum_{i=1}^n (\hat{q}_{2,i} - q_{2,i})' x_{2,i} x'_{2,i} \hat{r}_i, \\ N_3 &= \sum_{i=1}^n \tilde{x}_{1,i} e_i, & N_4 &= \sum_{i=1}^n (\hat{q}_{2,i} - q_{2,i})' x_{2,i} e_i. \end{aligned}$$

The proof of Theorem 1 will follow from the following points.

- We first show that the leading term $\frac{N_3}{\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{x}_{1,i} e_i$ converges in distribution. Using assumption **A2** and Lemma 1 (4.), we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \|\tilde{x}_{1,i} e_i - \tilde{x}_{1,i}(i/n) e_i(i/n)\|_2 = O\left(\frac{1}{\sqrt{n}}\right).$$

As in the proof of Theorem 1 given in [Zhang and Wu \(2012\)](#), using m -dependent approximations, it is possible to show that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{x}_{1,i}(i/n) e_i(i/n) \Rightarrow \mathcal{N}(0, \Sigma_2).$$

- Using Lemma 1, we have

$$\frac{1}{\sqrt{n}} |N_2| \leq C \frac{1}{\sqrt{n}} \sum_{i=1}^n |x_{2,i} x'_{2,i}| \cdot \max_{\ell=1,2} \max_{1 \leq i \leq n} |\hat{q}_{\ell,i} - q_{\ell,i}|^2 = o_{\mathbb{P}}(1).$$

- Next, we show that $\frac{N_4}{\sqrt{n}} = o_{\mathbb{P}}(1)$. With the same arguments, one can also show that $\frac{N_1}{\sqrt{n}} = o_{\mathbb{P}}(1)$. Using Lemma 1, it is only necessary to prove that $\frac{1}{\sqrt{n}} \sum_{i=1}^n A'_i x_{2,i} e_i = o_{\mathbb{P}}(1)$ when

$$A_i = \tilde{q}_{2,i} - q_{2,i}, \quad A_i = m_i^{-1} (\hat{\delta}_{2,i} - \mathbb{E} \hat{\delta}_{2,i}), \quad A_i = m_i^{-1} (\hat{\delta}_{3,i} - \mathbb{E} \hat{\delta}_{3,i}) m_i^{-1} \mathbb{E} \hat{\delta}_{2,i}.$$

The first case follows from Lemma 1, Lemma A.1 in [Zhang and Wu \(2012\)](#) and our bandwidth conditions. The second and third case follows from Lemma 5, applied with $r_1 = q_1$, $r_2 = q_2$ and using our bandwidth conditions. One can also note that for the proof of $N_1 = o(\sqrt{n})$, we have to show that $\frac{1}{\sqrt{n}} \sum_{i,j=1}^n k_{i,j} z_i w_j = o_{\mathbb{P}}(1)$ when z_i and w_j are elements of the matrices $x_i x'_i$ and $x_j x'_j$ respectively. One can apply Lemma 5 with $r_1 = r_2 = q_1$ and the result follows from our bandwidth conditions.

- To end the proof of Theorem 1, it remains to show that $\frac{D_n}{n} = \Sigma_1 + o_{\mathbb{P}}(1)$. We first note that

$$\begin{aligned} D_n &= \sum_{i=1}^n \tilde{x}_{1,i} \tilde{x}'_{1,i} + \sum_{i=1}^n (\hat{q}_{2,i} - q_{2,i}) x_{2,i} x'_{2,i} (\hat{q}_{2,i} - q_{2,i})' \\ &\quad + \sum_{i=1}^n (\hat{q}_{2,i} - q_{2,i})' x_{2,i} \tilde{x}'_{1,i} + \sum_{i=1}^n \tilde{x}_{1,i} x'_{2,i} (\hat{q}_{2,i} - q_{2,i}). \end{aligned}$$

It has already been shown in the previous points that

$$\frac{D_n}{n} = \frac{1}{n} \sum_{i=1}^n \tilde{x}_{1,i} \tilde{x}'_{1,i} + o_{\mathbb{P}}(1).$$

Moreover, using the moment bound given by Lemma A.1 in Zhang and Wu (2012), we have

$$\frac{1}{n} \sum_{i=1}^n [\tilde{x}_{1,i} \tilde{x}'_{1,i} - \mathbb{E}(\tilde{x}_{1,i} \tilde{x}'_{1,i})] = o_{\mathbb{P}}(1).$$

Then using assumption A2 and the convergence of Riemann sums, we get

$$\begin{aligned} \frac{1}{n} D_n &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}[\tilde{x}_{1,0}(i/n) \tilde{x}'_{1,0}(i/n)] + o_{\mathbb{P}}(1) \\ &= \Sigma_1 + o_{\mathbb{P}}(1). \square \end{aligned}$$

4.3 Additional results for the proof of Theorem 2

Lemma 2. Assume that the assumptions of Theorem 2 hold true. Then the following properties are valid.

1. $\max_{1 \leq i \leq n} |\hat{\beta}(i/n) - \beta(i/n)| = O_{\mathbb{P}}\left(b + \frac{\sqrt{\log(n)}}{\sqrt{nb}} + \frac{1}{n^{1-\frac{1}{q_2}b}}\right) = o_{\mathbb{P}}(n^{-1/4})$.
2. $\max_{1 \leq i \leq n} \sum_{j=1}^n k_{i,j} |x_j|^2 = O_{\mathbb{P}}(1)$ and $\max_{1 \leq i \leq n} \sum_{j=1}^n k_{i,j} |x_j e_j| = O_{\mathbb{P}}(1)$.
3. $\max_{1 \leq i \leq n} \left| \hat{\sigma}_i^2 - \sum_{j=1}^n k_{i,j} e_j^2 \right| = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$ and $\max_{1 \leq i \leq n} \left| \sum_{j=1}^n k_{i,j} e_j^2 - \sigma_i^2 \right| = o_{\mathbb{P}}(n^{-1/4})$.
4. We have

$$\max_{1 \leq i \leq n} \left| \frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2} \right| = o_{\mathbb{P}}(n^{-1/4}), \quad \max_{1 \leq i \leq n} \left| \frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2} - \frac{\sigma_i^2 - \sum_{j=1}^n k_{i,j} e_j^2}{\sigma_i^4} \right| = o_{\mathbb{P}}(n^{-1/2}).$$

5. Define

$$\check{s}_{1,i} = \sum_{j=1}^n k_{i,j} \frac{x_{1,j} y_j}{\sigma_j^2}, \quad \bar{s}_{1,i} = \sum_{j=1}^n k_{i,j} \left[\sigma_j^2 - \sum_{t=1}^n k_{j,t} e_t^2 \right] \frac{x_{2,j} y_j}{\sigma_j^4},$$

with similar definition for $\check{s}_{2,i}$, $\check{s}_{3,i}$ and $\bar{s}_{2,i}$, $\bar{s}_{3,i}$. Then we have for $\ell = 1, 2, 3$,

$$\max_{1 \leq i \leq n} |\hat{s}_{\ell,i}^* - \check{s}_{\ell,i} - \bar{s}_{\ell,i}| = o_{\mathbb{P}}(n^{-1/2}).$$

Moreover $\max_{1 \leq i \leq n} |\bar{s}_{\ell,i}| = o_{\mathbb{P}}(n^{-1/4})$ and $\max_{1 \leq i \leq n} \left| (\hat{s}_{3,i}^*)^{-1} \right| = O_{\mathbb{P}}(1)$.

6. We set $m_i^* = \mathbb{E}(\check{s}_{3,i})$ and for $\ell = 1, 2$, $\bar{q}_{\ell,i}^* = (m_i^*)^{-1} \mathbb{E}(\check{s}_{\ell,i})$. Then,

$$\max_{1 \leq i \leq n} \left| \hat{q}_{\ell,i}^* - \bar{q}_{\ell,i}^* - (m_i^*)^{-1} [\mathbb{E}\check{s}_{3,i} - \hat{s}_{3,i}^*] (m_i^*)^{-1} - (m_i^*)^{-1} [\hat{s}_{\ell,i}^* - \mathbb{E}\check{s}_{\ell,i}] \right| = o_{\mathbb{P}}(n^{-1/2}).$$

Moreover, if $s_{\ell,i}^*$ and $q_{\ell,i}^*$ are defined as $s_{\ell,i}$ and $q_{\ell,i}$ but with a division by σ_i^2 inside all the mathematical expectations, we have also

$$\max_{1 \leq i \leq n} \left[|\hat{q}_{\ell,i}^* - q_{\ell,i}^*| + |\bar{q}_{\ell,i}^* - q_{\ell,i}^*| \right] = o_{\mathbb{P}}(n^{-1/4}).$$

Proof of Lemma 2

1. For the first point, we set $d_i = \sum_{j=1}^n k_{i,j} x_j x'_j$ and we use the decomposition

$$\hat{\beta}(i/n) - \beta(i/n) = d_i^{-1} \sum_{j=1}^n k_{i,j} x_j e_j + d_i^{-1} \sum_{j=1}^n k_{i,j} x_j x'_j [\beta(j/n) - \beta(i/n)].$$

As in the proof of Lemma 1, point 1, we have $\max_{1 \leq i \leq n} |d_i^{-1}| = O_{\mathbb{P}}(1)$. Then the result follows from the Lipschitz properties of β , Lemma 3, the second point of the present lemma and our bandwidth conditions.

2. For the second point, we use the fact that $\sum_{n \geq 1, 1 \leq i \leq n} \mathbb{E}|x_i|^2 < \infty$ and from Lemma A.1 in [Zhang and Wu \(2012\)](#),

$$\max_{1 \leq i \leq n} \left| \sum_{j=1}^n k_{i,j} [|x_j|^2 - \mathbb{E}|x_j|^2] \right| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nb}}\right).$$

3. We have

$$\begin{aligned} & \hat{\sigma}^2(i/n) - \sum_{j=1}^n k_{i,j} e_j^2 \\ &= \sum_{j=1}^n k_{i,j} \left| x'_j (\beta(j/n) - \hat{\beta}(i/n)) \right|^2 - 2 \sum_{j=1}^n k_{i,j} e_j x'_j (\hat{\beta}(i/n) - \beta(j/n)) \\ &= A_i - 2B_i. \end{aligned}$$

We first deal with A_i . We have

$$|A_i| \leq 2 \sum_{j=1}^n k_{i,j} |x_j|^2 \left[|\beta(j/n) - \beta(i/n)|^2 + |\hat{\beta}(i/n) - \beta(i/n)|^2 \right].$$

From the two first points and using our bandwidth conditions, we have

$$\max_{1 \leq i \leq n} |A_i| = O_{\mathbb{P}}(b^2) + o_{\mathbb{P}}(n^{-1/2}) = o_{\mathbb{P}}(n^{-1/2}).$$

Next, we have $B_i = B_{i1} + B_{i2}$ with

$$B_{i1} = \sum_{j=1}^n k_{i,j} e_j x'_j [\hat{\beta}(i/n) - \beta(i/n)], \quad B_{i2} = \sum_{j=1}^n k_{i,j} e_j x'_j [\beta(i/n) - \beta(j/n)].$$

Under our bandwidth condition $c < \frac{3}{4} - \frac{1}{q_2}$ and the assumption on $\Theta_{n,q_2}(L)$, we get from Lemma 3 $\max_{1 \leq i \leq n} \left| \sum_{j=1}^n k_{i,j} x_j e_j \right| = o_{\mathbb{P}}(n^{-1/4})$. From point 1., we deduce that $\max_{1 \leq i \leq n} |B_{i1}| = o_{\mathbb{P}}(n^{-1/2})$. Finally, we have $B_{i2} = b \sum_{j=1}^n e_j x'_j \widetilde{k}_{i,j}$ with $\widetilde{k}_{i,j} = k_{i,j} b^{-1} (\beta(i/n) - \beta(j/n))$. One can note that the weights $\widetilde{k}_{i,j}$ satisfies similar properties as $k_{i,j}$. Indeed, we have $\max_{1 \leq i, j \leq n} |\widetilde{k}_{i,j}| = O\left(\frac{1}{nb}\right)$ and

$$\max_{1 \leq i \leq n} \max_{1 \leq j \leq n-1} |\widetilde{k}_{i,j} - \widetilde{k}_{i,j+1}| = O\left(\frac{1}{n^2 b^2}\right).$$

Then, we claim that Lemma 3 can be applied by replacing the weights $k_{i,j}$ by the weights $\widetilde{k}_{i,j}$ because the two previous properties on the weights are sufficient to get this result (see also the proof of Lemma A.3 in Zhang and Wu (2012)). Using our bandwidth conditions, we easily get $\max_{1 \leq i \leq n} |B_{i2}| = o_{\mathbb{P}}(n^{-1/2})$. This completes the proof of the first assertion. The second assertions easily follows Lemma 3, the Lipschitz property of σ and our bandwidth conditions.

4. Using the previous point, we have

$$\max_{1 \leq i \leq n} \left| \frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2} - \frac{\sigma_i^2 - \hat{\sigma}_i^2}{\sigma_i^4} - \frac{(\hat{\sigma}_i^2 - \sigma_i^2)^2}{\sigma_i^4 \hat{\sigma}_i^2} \right| = o_{\mathbb{P}}(n^{-1/2}).$$

Using point 3 and the positivity of σ , we get the result.

5. Setting $z_{1,i} = x_{2,i} y_i$, $z_{2,i} = x_{2,i} x'_{1,i}$ and $z_{3,i} = x_{2,i} x'_{2,i}$, we have

$$\hat{s}_{\ell,i}^* = \check{s}_{\ell,i} + \sum_{j=1}^n k_{i,j} z_{\ell,i} \left[\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2} \right].$$

Then the first assertion follows from the previous point of the present lemma. For the second assertion, we note that

$$\max_{1 \leq i \leq n} |\bar{s}_{\ell,i}| \leq \max_{1 \leq i \leq n} \left| \sigma_i^2 - \sum_{j=1}^n k_{i,j} e_j^2 \right|$$

and the result follows from the point 3. Finally, the last assertion follows from the first one, by using

$$\max_{1 \leq i \leq n} \left| \check{s}_{3,i} - \mathbb{E}(\sigma_i^{-2} x_{2,i} x'_{2,i}) \right| = o_{\mathbb{P}}(1), \quad \max_{1 \leq i \leq n} \left| \mathbb{E}(\sigma_i^{-2} x_{2,i} x'_{2,i}) - \mathbb{E}(\sigma_i^{-2} x_2(i/n) x_2(i/n)') \right| = O(1/n)$$

and assumption A4.

6. The first assertion follows from the previous point, using the same decomposition as in Lemma 1 for the difference $\hat{q}_{\ell,i}^* - \widetilde{q}_{\ell,i}^*$. The other assertions can be shown as in Lemma 1, points 3 and 4. \square

4.4 Proof of Theorem 2

As in the proof of Theorem 1, we set $D_n = \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} \hat{x}_{1,i}^* (\hat{x}_{1,i}^*)'$ and we use the decomposition

$$\hat{y}_i^* = (\hat{x}_{1,i}^*)' \beta_1 + x'_{2,i} \hat{r}_i^* + e_i, \quad \hat{r}_i^* = (\hat{q}_{2,i}^* - q_{2,i}^*) \beta_1 - (\hat{q}_{1,i}^* - q_{1,i}^*).$$

We get

$$\hat{\beta}_1^* - \beta_1 = (D_n^*)^{-1} \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} \hat{x}_{1,i}^* [x_{2,i}' \hat{r}_i^* + e_i].$$

Using now the decomposition $\hat{x}_{1,i}^* = \tilde{x}_{1,i}^* - (\hat{q}_{2,i}^* - q_{2,i}^*)' x_{2,i}$, with $\tilde{x}_{1,i}^* = x_{1,i} - (q_{2,i}^*)' x_{2,i}$, we have

$$\hat{\beta}_1^* - \beta_1 = (D_n^*)^{-1} [N_1^* - N_2^* + N_3^* - N_4^*],$$

where

$$\begin{aligned} N_1^* &= \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} \tilde{x}_{1,i}^* x_{2,i}' \hat{r}_i^*, & N_2^* &= \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} (\hat{q}_{2,i}^* - q_{2,i}^*)' x_{2,i} x_{2,i}' \hat{r}_i^*, \\ N_3^* &= \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} \tilde{x}_{1,i}^* e_i, & N_4^* &= \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} (\hat{q}_{2,i}^* - q_{2,i}^*)' x_{2,i} e_i. \end{aligned}$$

- We first consider the leading term $\frac{1}{\sqrt{n}} N_3^*$. Using Lemma 2, we have

$$\frac{1}{\sqrt{n}} N_3^* = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\tilde{x}_{1,i}^* e_i}{\sigma_i^2} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\sigma_i^2 - \sum_{j=1}^n k_{i,j} e_j^2}{\sigma_i^4} \tilde{x}_{1,i}^* e_i + o_{\mathbb{P}}(1).$$

As in the proof of Theorem 1, one can show that the first term is asymptotically Gaussian with mean 0 and variance Σ^* . Note that one can use directly the central limit theorem for martingale differences for this step. For the second term, we decompose

$$\sum_{i=1}^n \frac{\sigma_i^2 - \sum_{j=1}^n k_{i,j} e_j^2}{\sigma_i^4} \tilde{x}_{1,i}^* e_i = S_1 + S_2,$$

with $S_1 = \sum_{i,j=1}^n k_{i,j} (\sigma_j^2 - \sigma_i^2) \frac{\tilde{x}_{1,i}^* e_i}{\sigma_i^4} = o_{\mathbb{P}}(\sqrt{n})$ by using the moment inequality given in Zhang and Wu (2012), Lemma A.1 and $S_2 = \sum_{i,j=1}^n k_{i,j} (e_j^2 - \sigma_j^2) \frac{\tilde{x}_{1,i}^* e_i}{\sigma_i^4} = o_{\mathbb{P}}(\sqrt{n})$ by using Lemma 5 with $r = 4/3$, $r_1 = 2$ and $r_2 = 4$.

- From Lemma 2, we easily get $N_2^* = o_{\mathbb{P}}(\sqrt{n})$.
- Next, we show that $N_4^* = o_{\mathbb{P}}(\sqrt{n})$. Using Lemma 2, it is easily seen that

$$N_4^* = \sum_{i=1}^n (\hat{q}_{2,i}^* - q_{2,i}^*)' \frac{x_{2,i} e_i}{\sigma_i^2} + o_{\mathbb{P}}(\sqrt{n}).$$

Moreover, using Lemma A.1 in Zhang and Wu (2012), we have

$$N_4^* = \sum_{i=1}^n (\hat{q}_{2,i}^* - \bar{q}_{2,i}^*)' \frac{x_{2,i} e_i}{\sigma_i^2} + o_{\mathbb{P}}(\sqrt{n}).$$

Using Lemma 2(6), it remains to show that $\sum_{i=1}^n z_i' \frac{x_{2,i} e_i}{\sigma_i^2} = o_{\mathbb{P}}(\sqrt{n})$ when

$$z_i = (m_i^*)^{-1} [m_i - \hat{s}_{3,i}^*] (m_i^*)^{-1} \text{ or } z_i = (m_i^*)^{-1} [\hat{s}_{2,i}^* - \mathbb{E}\check{s}_{2,i}].$$

Let us detail the first case. We have $m_i^* z_i m_i^* = z_{1,i} + z_{2,i}$ with $z_{1,i} = \check{s}_{3,i} - \mathbb{E}\check{s}_{3,i}$ and $z_{2,i} = \bar{s}_{3,i}$. For $z_{1,i}$, one can use Lemma 5 with $r_1 = q_1$ and $r_2 = q_2$. For the $z_{2,i}$, we have

$$\left| \sum_{i=1}^n (m_i^*)^{-1} z'_{2,i} (m_i^*)^{-1} \frac{x_{2,i} e_i}{\sigma_i^2} \right| \leq Cn \max_{1 \leq j \leq n} \left| \sum_{i=1}^n k_{i,j} \frac{x'_{2,i} e_i}{\sigma_i^2} \right| \cdot \max_{1 \leq j \leq n} \left| \sum_{i=1}^n k_{j,i} e_i^2 - \sigma_j^2 \right|.$$

Then the result follows from Lemma 3. Note that the first maximum involves the weights $k_{i,j}$ instead of $k_{j,i}$ but it does not change the result and this maximum is a $o_{\mathbb{P}}(n^{-1/4})$. Moreover, the second maximum is $o_{\mathbb{P}}(n^{-1/4})$ by applying Lemma 2(3). The case $z_i = (m_i^*)^{-1} [\hat{s}_{2,i}^* - \mathbb{E}\hat{s}_{2,i}]$ is similar and then omitted. Using analog arguments, one can also show that $N_1^* = o_{\mathbb{P}}(\sqrt{n})$. Details are omitted.

- Finally, we show that $D_n^*/n \rightarrow \Sigma^*$ in probability. Using the decomposition

$$\begin{aligned} D_n^* &= \sum_{i=1}^n \frac{1}{\hat{\sigma}_i^2} \tilde{x}_{1,i}^* (\tilde{x}_{1,i}^*)' + \sum_{i=1}^n (\hat{q}_{2,i}^* - q_{2,i}^*) \frac{x_{2,i} x'_{2,i}}{\hat{\sigma}_i^2} (\hat{q}_{2,i}^* - q_{2,i}^*)' \\ &+ \sum_{i=1}^n (\hat{q}_{2,i}^* - q_{2,i}^*)' \frac{x_{2,i} \tilde{x}_{1,i}^*}{\hat{\sigma}_i^2} + \sum_{i=1}^n \frac{\tilde{x}_{1,i}^* x'_{2,i}}{\hat{\sigma}_i^2} (\hat{q}_{2,i}^* - q_{2,i}^*). \end{aligned}$$

Using the previous points and Lemma 2(4), we get

$$\frac{D_n^*}{n} = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{x}_{1,i}^* (\tilde{x}_{1,i}^*)'}{\hat{\sigma}_i^2} + o_{\mathbb{P}}(1) = \frac{1}{n} \sum_{i=1}^n \frac{\tilde{x}_{1,i}^* (\tilde{x}_{1,i}^*)'}{\sigma_i^2} = o_{\mathbb{P}}(1).$$

As for the end of the proof of Theorem 1, we get $n^{-1} D_n^* = \Sigma^* + o_{\mathbb{P}}(1)$. \square

4.5 Auxiliary results

We first recall a lemma given in Zhang and Wu (2012) for uniform convergence of kernel smoothers.

Lemma 3. *Let $z_{n,i} = J_{n,i}(\mathcal{F}_i)$ be a centered triangular array. We assume that there exists $r > 2$ such that $\sup_{1 \leq i \leq n, n \geq 1} \|z_{n,i}\|_r < \infty$ and $\Theta_{n,r}(J) = O(n^{-\nu})$ for $\nu > \frac{1}{2} - \frac{1}{r}$. Then,*

$$\sup_{u \in [0,1]} \left| \sum_{i=1}^n \frac{1}{nb} K\left(\frac{u - i/n}{b}\right) z_{n,i} \right| = \frac{O_{\mathbb{P}}(v_n)}{nb}, \text{ where } v_n = n^{1/r} + \sqrt{nb \log n}.$$

The second lemma is useful for getting convergence rates of quadratic forms.

Lemma 4. *Let $z_{n,i} = J_{1,n,i}(\mathcal{F}_i)$ and $w_{n,i} = J_{2,n,i}(\mathcal{F}_i)$ be two centered triangular arrays. Let $q \in (1, 2]$ and $q_1, q_2 > 2$ two real numbers such that $\frac{1}{q_1} + \frac{1}{q_2} = \frac{1}{q}$ and $\Theta_{0,q_1}(J_1) + \Theta_{0,q_2}(J_2) < \infty$ and $Q_n = \sum_{1 \leq i < j \leq n} \alpha_{i,j} z_{n,i} w_{n,j}$, where for $1 \leq i < j \leq n$, $\alpha_{i,j}$ is a real number. Then we have*

$$\|Q_n - \mathbb{E}Q_n\|_q \leq Cn^{1/q} \left[\Theta_{0,q_1}(J_1) + \Theta_{0,q_2}(J_2) \right]^2 a_n,$$

where

$$a_n^2 = \left(\max_{1 \leq i < n} \sum_{j=i+1}^n |\alpha_{i,j}|^2 \right) \vee \left(\max_{1 < j \leq n} \sum_{i=1}^{j-1} |\alpha_{i,j}|^2 \right).$$

Proof of Lemma 4 We follows the proof of Proposition 1 in [Liu and Wu \(2010\)](#) which gives a control of \mathbb{L}^q norms of quadratic forms when $q \geq 2$. See also Lemma A.2 in [Zhang and Wu \(2012\)](#) for a similar result. We set

$$u_{n,j} = \sum_{i=1}^{j-1} \alpha_{i,j} z_{n,i}, \quad v_{n,i}^{(k)} = \sum_{j=i+1}^n \alpha_{i,j} w_{n,j}^{(k)},$$

where for $k \in \mathbb{Z}$, $w_{n,j}^{(k)}$ is defined as $w_{n,j}$ but the innovation ξ_k is replaced with ξ'_k in \mathcal{F}_j (ξ' is a copy of ξ). Defining

$$\mathcal{P}_k = \mathbb{E}(\cdot | \mathcal{F}_k) - \mathbb{E}(\cdot | \mathcal{F}_{k-1}),$$

we have

$$\begin{aligned} \|\mathcal{P}_k Q_n\|_q &\leq \left\| \sum_{1 \leq i < j \leq n} \alpha_{i,j} [z_{n,i} w_{n,j} - z_{n,i}^{(k)} w_{n,j}^{(k)}] \right\|_q \\ &\leq \left\| \sum_{j=2}^n u_{n,j} [w_{n,j} - w_{n,j}^{(k)}] \right\|_q + \left\| \sum_{i=1}^{n-1} v_{n,i}^{(k)} [z_{n,i} - z_{n,i}^{(k)}] \right\|_q \\ &= I_k + II_k. \end{aligned}$$

Using the moment bound given in Lemma A.1 in [Zhang and Wu \(2012\)](#), we have

$$\begin{aligned} I_k &\leq \sum_{j=2}^n \|u_{n,j}\|_{q_1} \cdot \|w_{n,j} - w_{n,j}^{(k)}\|_{q_2} \\ &\leq C a_n \sum_{j=2}^n \delta_{j-k, q_2}(J_2) \Theta_{0, q_1}(J_1). \end{aligned}$$

Using similar arguments, we get

$$II_k \leq C a_n \sum_{i=1}^{n-1} \delta_{i-k, q_1}(J_1) \Theta_{0, q_2}(J_2).$$

Since

$$\sum_{j=2}^n \delta_{j-k, 2q}(J_2) \leq \Theta_{0, q_2}(J_2), \quad \sum_{i=1}^{n-1} \delta_{i-k, 2q}(J_1) \leq \Theta_{0, q_1}(J_1),$$

we get

$$\begin{aligned} \|\mathcal{P}_k Q_n\|_q^q &\leq C [I_k^q + II_k^q] \\ &\leq C a_n^q (\Theta_{0, q_1}(J_1) + \Theta_{0, q_2}(J_2))^{2q-1} \left[\sum_{j=2}^n \delta_{j-k, 2q}(J_2) + \sum_{i=1}^{n-1} \delta_{i-k, 2q}(J_1) \right]. \end{aligned}$$

Finally, using Burkholder's inequality, we get

$$\begin{aligned}
\|Q_n - \mathbb{E}Q_n\|_q &= \left\| \sum_{k=-\infty}^n \mathcal{P}_k Q_n \right\|_q \\
&\leq C \mathbb{E}^{1/q} \left[\left| \sum_{k=-\infty}^n [\mathcal{P}_k Q_n]^2 \right|^{q/2} \right] \\
&\leq C \left[\sum_{k=-\infty}^n \mathbb{E} |\mathcal{P}_k Q_n|^q \right]^{1/q} \\
&\leq C a_n n^{1/q} \left[\Theta_{0,q_1}(J_1) + \Theta_{0,q_2}(J_2) \right]^2.
\end{aligned}$$

This completes the proof of the lemma. \square

Lemma 5. Let $z_i = J_{1,n,i}(\mathcal{F}_i)$ and $w_i = J_{2,n,i}(\mathcal{F}_i)$ be two centered triangular arrays. Assume that there exists $r \in (1, 2]$ such that $\Theta_{0,r_1}(J_1) + \Theta_{0,r_2}(J_2) < \infty$ where $r_1, r_2 > 2$ are such that $\frac{1}{r} = \frac{1}{r_1} + \frac{1}{r_2}$. Then, if $n^d b \rightarrow \infty$ for $d = \min\left(2 - \frac{2}{r}, \frac{1}{2}\right)$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n z_i \cdot \sum_{j=1}^n k_{i,j} w_j = o_{\mathbb{P}}(1).$$

Proof of Lemma 5 Since $\max_{1 \leq i \leq n} \mathbb{E} |z_i w_i| = O(1)$, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n k_{i,i} z_i w_i = O_{\mathbb{P}}\left(\frac{1}{\sqrt{nb}}\right) = o_{\mathbb{P}}(1).$$

Moreover, we have

$$\max_{1 \leq i < n} \left(\sum_{j=i+1}^n k_{i,j}^2 \right) \vee \left(\max_{1 < j \leq n} \sum_{i=1}^{j-1} k_{i,j}^2 \right) = O\left(\frac{1}{nb}\right).$$

Using Lemma 4, we get under our bandwidth conditions

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i \neq j \leq n} k_{i,j} z_i w_j = O_{\mathbb{P}}\left(\frac{n^{1/r}}{n \sqrt{b}}\right) = o_{\mathbb{P}}(1).$$

Finally, we have

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i \neq j \leq n} k_{i,j} \mathbb{E}(z_i w_j) = O\left(\frac{1}{\sqrt{nb}}\right). \tag{9}$$

To show (9), we bound the covariances as follows. If $i \leq j$, we have

$$\begin{aligned}
\left| \mathbb{E}(z_i w_j) \right| &\leq \sum_{\ell \leq i} \left| \mathbb{E}(\mathcal{P}_{\ell} z_i \cdot \mathcal{P}_{\ell} w_j) \right| \\
&\leq \sum_{\ell \leq i} \delta_{i-\ell, r_1}(J_1) \delta_{j-\ell, r_2}(J_2).
\end{aligned}$$

Hence, we obtain

$$\frac{1}{\sqrt{n}} \sum_{1 \leq i, j \leq n} k_{i,j} \left| \mathbb{E}(z_i w_j) \right| \leq \sqrt{n} \max_{1 \leq i, j \leq n} k_{i,j} \Theta_{0,r_1}(J_1) \Theta_{0,r_2}(J_2),$$

which leads to (9) since $\max_{1 \leq i, j \leq n} k_{i,j} = O\left(\frac{1}{nb}\right)$. The proof of Lemma 5 is now complete. \square

Lemma 6. Let $\Sigma : [0, 1] \rightarrow \mathcal{S}_d^+$ be a continuous application taking values in the space \mathcal{S}_d^+ of square matrices of size d , symmetric and positive definite. Assume that for each $u \in [0, 1]$, $\Sigma(u)$ is invertible. Then we have the inequality

$$\left(\int_0^1 \Sigma(u) du \right)^{-1} \leq \int_0^1 \Sigma(u)^{-1} du,$$

where \leq denotes the order relation for nonnegative definite matrices. Moreover, the equality between these two matrices holds if and only if the application Σ is constant.

Proof of Lemma 6 For $x, y \in \mathbb{R}^d$, we have from the Cauchy-Schwarz inequality,

$$(x' \cdot y)^2 = \left| \int_0^1 x' \cdot \Sigma(u)^{-1/2} \Sigma(u)^{1/2} \cdot y du \right|^2 \leq x' \cdot \int_0^1 \Sigma(u)^{-1} du \cdot x \cdot y' \cdot \int_0^1 \Sigma(u) du \cdot y.$$

Then setting $y = \left(\int_0^1 \Sigma(u) du \right)^{-1} \cdot x$, we get

$$x' \cdot \left(\int_0^1 \Sigma(u) du \right)^{-1} \cdot x \leq x' \cdot \int_0^1 \Sigma(u)^{-1} du \cdot x,$$

which shows the inequality. Moreover, if the two matrices are equal, the equality condition in the Cauchy-Schwarz inequality entails that for each $x \in \mathbb{R}^d \setminus \{0\}$, there exists a real number λ_x such that $x = \lambda_x \Sigma(u) \left(\int_0^1 \Sigma(u) du \right)^{-1} x$. Integrating the previous relation with respect to u , we find that $\lambda_x = 1$ and then $\Sigma(v) = \int_0^1 \Sigma(u) du$, $v \in [0, 1]$. \square

References

- G. Chamberlain. Efficiency bounds for semiparametric regression. *Econometrica*, 60:567–596, 1992.
- B. Chen and Y. Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80:1157–1183, 2012.
- R. Dahlhaus. Fitting time series models to nonstationary processes. *Ann. Statist.*, 25:1–37, 1997.
- R. Dahlhaus and S. Subba Rao. Statistical inference for time-varying arch processes. *Ann. Statist.*, 34: 1075–1114, 2006.
- R. Dahlhaus, S. Richter, and W.B. Wu. Towards a general theory for non-linear locally stationary processes. *arXiv:1704.02860*, 2017.
- J. Fan and T. Huang. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11:1031–1057, 2005.

- J. Fan and W. Zhang. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1: 179–195, 2008.
- J. Gao and K. Hawthorne. Semiparametric estimation and testing of the trend of temperature series. *Econom. J.*, 9:332–355, 2006.
- W. Liu and W.B. Wu. Asymptotics of spectral density estimates. *Econometric Theory*, 26:1218–1245, 2010.
- P.M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56:931–954, 1988.
- P. Speckman. Kernel smoothing in partial linear models. *J. Roy. Statist. Soc. B*, 50:413–436, 1988.
- S. Subba Rao. On some nonstationary, nonlinear random processes and their stationary approximations. *Adv. in App. Probab.*, 38:1155–1172, 2006.
- L. Truquet. Parameter stability and semiparametric inference in time-varying arch models. *Forthcoming in J. R. Stat. Soc. Ser. B*, 2016.
- A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer Verlag, New-York, 1996.
- T. Zhang and W.B. Wu. Inference of time-varying regression models. *Ann. Statist.*, 40:1376–1402, 2012.