C. R. Math. Rep. Acad. Sci. Canada Vol. **46** (4) 2024, pp. 119–144

On Variable Selection in Partially Linear Regression Model

Jean Roland Ebende Penda, Emmanuel de Dieu Nkou, Stéphane Bouka, and Guy Martial Nkiet

Presented by Christian Genest, FRSC

Abstract. This paper deals with variable selection in a partially linear regression model. After transforming the latter into a linear regression model by using an appropriate conditional expectation, the authors apply a known method to the resulting model, which reduces the problem to an estimation issue for two specified parameters. The authors then propose estimators for these parameters based on nonparametric estimators of a density and regression functions. They establish consistency, under specified assumptions, of the method thus proposed. A simulation study, made in order to assess the finite-sample behaviour of the proposed method with comparison to existing ones, is presented.

Résumé. Cet article considère le problème de sélection des variables dans un modèle de régression partiellement linéaire. Après une transformation de ce modèle en un modèle de régression linéaire, par utilisation d'une espérance conditionnelle appropriée, les auteurs appliquent au modèle ainsi obtenu une méthode connue ramenant le problème abordé à un problème d'estimation de deux paramètres spécifiés. Ils proposent alors des estimateurs de ces paramètres basés sur des estimateurs non-paramétriques d'une densité et de fonctions de régression. Ils établissent la convergence, sous des postulats spécifiés, de la méthode ainsi proposée. Des simulations leur permettent d'évaluer les performances, à taille d'échantillon finie, de la méthode proposée et de les comparer à celles de méthodes existantes.

1. **Introduction** Variable selection is one of the most relevant problems in statistical modeling using regression models. It has been intensively studied in the literature, but mainly in the linear case. However, partially linear regression models, introduced about twenty years ago, are semiparametric models which are flexible and easily interpretable, since they contain both parametric and nonparametric components. They allow easier interpretation of the effect of each variable and may be preferred to a completely nonparametric regression model.

In spite of the obvious value of partially linear regression models, limited work has been done on variable selection for them as noted in [5]. Bunea [2]

Received by the editors on July 24, 2024; revised November 21, 2024.
AMS Subject Classification: Primary: 62G05, 62G20; secondary: 62J99.
Keywords: Partially linear regression model, variable selection, nonparametric estimation, asymptotic behaviour.

proposed a procedure which estimates the parametric and nonparametric components simultaneously via penalized least squares with a L_0 penalty, whereas a two-stage approach that first estimates the nonparametric component, and then selects the significant variables in the parametric component was introduced in [3]. Later, Xie and Huang [20] considered the SCAD-penalized regression for partially linear models for high dimensional data, an approach based on penalized least squares was investigated in [9] in the context of sampling with measurement errors, and a double-penalized least squares approach was introduced in [12]. More recent methods are based on the Dantzig selector, which received a considerable amount of attention, and on partial correlation between the partial residuals of the response and the predictors; see [7, 10]. Note that the partially linear regression model has also been considered for variable selection purposes in specific situations such as the case of right-censored data and, more recently, the case where one considers that some of the explanatory variables may have varying coefficients while the remained explanatory variables possess constant ones; see [8, 21].

Most of the methods mentioned above are based on penalization approaches, in fact inheriting this from the methods for variable selection in linear regression where these approaches have become very popular following [19] and [4], who introduced the LASSO and the SCAD-penalty. However, there are other approaches which led to methods with very good performances in selecting variables in linear regression. This is for example the case of [11], where the variable selection problem was treated as a problem of estimating a suitable subset of explanarory variables, and a criterion which characterizes this subset as depending on two parameters was introduced, thereby reducing the variable selection method to the estimation of these parameters by a method which was then proposed.

In this paper, we extend the approach of [11] to the case of a partially linear regression (PLR) model. After a transformation of this model allowing to reduce it to a linear regression model involving transformed explanatory and response variables, we apply the aforementioned method to the resulting linear model.

The rest of the paper is organized as follows. In Section 2, the PLR model with which we deal is specified, the aforementioned criterion is introduced as well as the resulting characterization of the subset of significant variables as depending on two parameters: a permutation and a dimension. In Section 3, an estimator of the previous criterion based on nonparametric estimators of regression functions and a density is introduced, leading to the estimation of the aforementioned parameters, which achieves the proposed variable selection procedure. Section 4 is devoted to consistency results for our proposal under prespecified conditions. A simulation study aiming to assess the finite-sample behaviour of the proposed method with comparison to existing ones is presented in Section 5. All the proofs are postponed to Section 6.

2. The Variable Selection Problem in PLR Model In this section, we first specify the partially linear model that we tackle for variable selection purposes. After an appropriate transformation, this model reduces to a linear regression model with multiple predictor and a scalar response, and we show how the aforementioned variable selection problem boils down to a problem of estimating a suitable subset of the set of explanatory variables. Then, applying an approach proposed in [11], we introduce a criterion that allows to characterize the previous subset.

2.1. The model and its transformation We consider the partially linear model defined as

$$Y = X^\top \beta + \Phi(Z) + \varepsilon = \sum_{j=1}^{p} \beta_j X_j + \Phi(Z) + \varepsilon, \tag{2.1}$$

where $X = (X_1, \ldots, X_p)^\top$ is the random vector of predictors, $\beta = (\beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^p$ is the vector of parameters, Z is a scalar random covariable, Φ is an arbitrary and unknown measurable function from $\mathbb{R}$ to itself, ε is a real random variable independent of (X, Z) and such that $\mathbb{E}(\varepsilon) = 0$, and Y is the scalar response.

We are interested in selecting variables in the parametric part of Model (2.1), that is determining the predictors X_j's that are significant in explaining the response Y, on the basis of a sample of (X, Y, Z). This comes down to the estimation of the subset I_1 of $I := \{1, \ldots, p\}$ made up of indices of the coefficients that are non-zero, namely $I_1 = \{j \in I \,:\, \beta_j \neq 0\}$.

To achieve this, we will first transform (2.1) so as to reduce it to a linear regression model with multiple predictors and a scalar response. More specifically, assuming that $\mathbb{E}(\|X\|) < \infty$ and $\mathbb{E}(|Y|) < \infty$, where $\|\cdot\|$ denotes the Euclidean norm, and applying to (2.1) the conditional expectation given Z, we obtain:

$$\mathbb{E}(Y|Z) = \mathbb{E}(X|Z)^\top \beta + \Phi(Z) + \mathbb{E}(\varepsilon|Z) = \mathbb{E}(X|Z)^\top \beta + \Phi(Z), \tag{2.2}$$

because $\mathbb{E}(\varepsilon|Z) = \mathbb{E}(\varepsilon) = 0$. Taking the difference between (2.1) and (2.2) yields

$$Y - r(Z) = (X - s(Z))^\top \beta + \varepsilon = \sum_{j=1}^{p} \beta_j (X_j - s_j(Z)) + \varepsilon, \tag{2.3}$$

where

$$r(z) = \mathbb{E}(Y|Z = z), \quad s(z) = \mathbb{E}(X|Z = z) = (s_1(z), \ldots, s_p(z))^\top$$

with $s_j(z) = \mathbb{E}(X_j|Z = z)$. We have, in this way, obtained a linear regression model to which we can apply an appropriate variable selection method. Our approach consists in applying the method of [11] to the transformed model; to this end, we will first consider a criterion introduced in this latter work and adapt it to the case at hand.

2.2. A criterion for characterizing I_1 Assuming that $\mathbb{E}\left(\|X\|^2\right) < \infty$ and $\mathbb{E}\left(Y^2\right) < \infty$, we consider the covariance matrix and the cross-covariance vector given by

$$V_1 = \mathbb{E}\big((X - s(Z))(X - s(Z))^\top\big) \tag{2.4}$$

and

$$V_{12} = \mathbb{E}\big((Y - r(Z))(X - s(Z))\big), \tag{2.5}$$

and we assume that V_1 is invertible. For any subset $J := \{i_1, \dots, i_k\}$ of I, consider the $k \times p$ matrix defined by

$$A_J = \begin{pmatrix} a_{11}^{(J)} & a_{12}^{(J)} & \cdots & a_{1p}^{(J)} \\ a_{21}^{(J)} & a_{22}^{(J)} & \cdots & a_{2p}^{(J)} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1}^{(J)} & a_{k2}^{(J)} & \cdots & a_{kp}^{(J)} \end{pmatrix},$$

where

$$a_{\ell j}^{(J)} = \begin{cases} 1 & \text{if } j = i_\ell \\ 0 & \text{if } j \neq i_\ell \end{cases}, \ 1 \leqslant \ell \leq k, \ 1 \leqslant j \leq p.$$

This matrix transforms any vector $x = (x_1, \dots, x_p)^\top$ to the vector $A_J x = (x_{i_1}, \dots, x_{i_k})^\top$ of lower dimension whose components are selected from the initial vector x by taking only the components x_i such that $i \in J$. A criterion allowing to measure the degree of relevance of variables whose indices belong to J was introduced in[13]; it is in fact a distance between the vector β of coefficients of Model (2.3) with the whole predictors and that of coefficients of the same model with only the predictors whose indices belong to J. Therefore, it measures the loss incurred by selecting these latter variables; see [11]. It is given by

$$\xi_J = \|V_{12} - V_1 \Pi_J V_{12}\|, \tag{2.6}$$

where $\Pi_J = A_J^\top \left(A_J V_1 A_J^\top\right)^{-1} A_J$.

Rather than characterizing I_1 as a subset that minimizes ξ_J over all subsets J of I, we prefer to use another strategy, leading to a faster procedure, which leads to an expression of this subset as depending on two parameters, a permutation of I and a dimension, determined by using the above criterion. Let us recall that a permutation of I is a bijective map $\tau : I \to I$; for any integer $j \in I$, we denote by $\tau(j)$ the integer of I which is mapped to by j. Putting $I^{[j]} = I \setminus \{j\}$ and letting τ be the permutation of I satisfying

$$\xi_{I^{[\tau(1)]}} \geqslant \xi_{I^{[\tau(2)]}} \geqslant \cdots \geqslant \xi_{I^{[\tau(p)]}}$$

with $\tau(\ell) < \tau(j)$ if $\xi_{I[\tau(\ell)]} = \xi_{I[\tau(j)]}$ and $\ell < j$, there exists an integer d in $\{1, \dots, p-1\}$, called dimension, such that

$$\xi_{I[\tau(1)]} \geqslant \xi_{I[\tau(2)]} \geqslant \cdots \geqslant \xi_{I[\tau(d)]} > 0 = \xi_{I[\tau(d+1)]} = \cdots = \xi_{I[\tau(p)]},$$

and, therefore (see [11, 13]),

$$I_1 = \{\tau(j) \,:\, 1 \leqslant j \leqslant d\}. \tag{2.7}$$

Equation (2.7) provides an expression of I_1 as depending on the permutation τ and the dimension d, and shows that a plug-in estimate of this set can be obtained from estimators of these two parameters.

3. The Variable Selection Procedure Our variable selection procedure relies on estimating the above parameters τ and d. More specifically, having estimated these parameters by $\widehat{\tau}$ and $\widehat{d}$, respectively, we take as our set of indices of predictors to be selected the subset $\widehat{I}_1$ of I given by

$$\widehat{I}_1^{(n)} = \left\{\widehat{\tau}(j) \,:\, 1 \leqslant j \leqslant \widehat{d}\right\}. \tag{3.1}$$

In this section, we introduce an estimator of ξ_J obtained from an i.i.d. sample $\left\{(X^{(i)}, Y_i, Z_i)\right\}_{1\leqslant i\leqslant n}$ of (X, Y, Z). Then, we define estimators $\widehat{\tau}$ and $\widehat{d}$, thereby achieving our proposal for variable selection in the parametric part of Model (2.1).

3.1. Estimation of the criterion For estimating ξ_J, we first need to estimate the covariance matrix and the cross-covariance vector given in (2.4) and (2.5). For this purpose, we will use Nadaraya-Watson estimators, which will be modified as done in [22], and repeated in [16]. Specifically, denoting by f the density of Z, and by ν and η_j the functions defined as $\nu(z) = f(z)r(z)$ and $\eta_j(z) = f(z)s_j(z)$, we consider the kernel estimators of f, ν, and η_j given by

$$\widehat{f}_n(z) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{z - Z_i}{h_n}\right),\ \widehat{\nu}_n(z) = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K\left(\frac{z - Z_i}{h_n}\right)$$

and

$$\widehat{\eta}_{j,n}(z) = \frac{1}{nh_n}\sum_{i=1}^{n} X_{ij} K\left(\frac{z - Z_i}{h_n}\right),$$

where $K : \mathbb{R} \to \mathbb{R}$ is a kernel, h_n is a sequence of positive numbers, and X_{ij} is the j-th coordinate of $X^{(i)}$, i.e., $X^{(i)} = (X_{i1}, \dots, X_{ip})^\top$. Considering a sequence $(b_n)_{n\in\mathbb{N}^*}$ of positive real numbers converging to 0 as $n \to \infty$, we define

$$\widehat{f}_{b_n}(z) = \max(\widehat{f}_n(z), b_n),$$

and estimate r and s by $\widehat{r}_{b_n}$ and $\widehat{s}_{b_n}$, respectively, given by

$$\widehat{r}_{b_n}(z) = \frac{\widehat{\nu}_n(z)}{\widehat{f}_{b_n}(z)}, \quad \widehat{s}_{b_n}(z) = \left(\widehat{s}_{b_n,1}(z), \dots, \widehat{s}_{b_n,p}(z)\right)^\top,$$

where

$$\widehat{s}_{b_n,j}(z) = \frac{\widehat{\eta}_{j,n}(z)}{\widehat{f}_{b_n}(z)}.$$

Remark 3.1. *To overcome technical difficulties due to small values in the denominator, the preceding modified estimator $\widehat{f}_{b_n}$ of the density was introduced in [22]. In order to guarantee good accuracy for this estimator, the sequence b_n can be chosen as $b_n = \min(\epsilon, n^{-c_2})$, where ϵ is a sufficiently small positive number; see [16]. Indeed, in this case b_n still satisfies Assumption* 4.7 *and is such that* $\sup_{x\in\mathbb{R}} |\widehat{f}_{b_n}(x) - \widehat{f}_n(x)| \leqslant \epsilon$.

Using the preceding estimators, we obtain the estimators $\widehat{V}_1^{(n)}$ and $\widehat{V}_{12}^{(n)}$ of V_1 and V_{12}, respectively, defined as

$$\widehat{V}_1^{(n)} = \frac{1}{n}\sum_{i=1}^{n}(X^{(i)} - \widehat{s}_{b_n}(Z_i))(X^{(i)} - \widehat{s}_{b_n}(Z_i))^\top$$

and

$$\widehat{V}_{12}^{(n)} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{r}_{b_n}(Z_i))(X^{(i)} - \widehat{s}_{b_n}(Z_i)).$$

Then, putting

$$\widehat{\Pi}_J^{(n)} = A_J^T(A_J\widehat{V}_1^{(n)}A_J^T)^{-1}A_J,$$

we propose the following estimator of ξ_J:

$$\widehat{\xi}_J^{(n)} = \|\widehat{V}_{12}^{(n)} - \widehat{V}_1^{(n)}\widehat{\Pi}_J^{(n)}\widehat{V}_{12}^{(n)}\|. \tag{3.2}$$

3.2. Estimation of the permutation To estimate τ, it would be natural to sort the $\widehat{\xi}_{I[j]}^{(n)}$'s in decreasing order as it was done for defining τ from the $\xi_{I[j]}$'s , but such an approach does not guarantee the consistency of the resulting estimator because of possible ties. We rather use estimates of the $\xi_{I[j]}$'s obtained from appropriate penalizations of the $\widehat{\xi}_{I[j]}^{(n)}$'s which circumvent the issue of ties. More specifically, we consider

$$\widehat{\phi}_j^{(n)} = \widehat{\xi}_{I[j]}^{(n)} + \frac{v\,(j)}{n^\alpha},$$

where $0 < \alpha < 1/4$ and $v : I \to \mathbb{R}_+$ is a strictly decreasing function, and we estimate τ by sorting the $\widehat{\phi}_i^{(n)}$'s in decreasing order. The resulting estimator is the permutation $\widehat{\tau}$ of I satisfying

$$\widehat{\phi}_{\widehat{\tau}(1)}^{(n)} > \widehat{\phi}_{\widehat{\tau}(2)}^{(n)} > \cdots > \widehat{\phi}_{\widehat{\tau}(p)}^{(n)}.$$

Remark 3.2. *The introduction of the above penalization terms* $v(j)/n^{\alpha}$ *is motivated by the fact that they allow to circumvent the issue of ties in the values of the estimates of the used criterion. Indeed, even if one has* $\widehat{\xi}^{(n)}_{I[j]} = \widehat{\xi}^{(n)}_{I[\ell]}$ *for* $j \neq \ell$, *we will have* $\widehat{\phi}^{(n)}_j \neq \widehat{\phi}^{(n)}_\ell$. *This property is necessary to prove consistency of the proposed estimators. The above upper bound for the tuning parameter* α *is lower than the one given in [11], where it was assumed that* $\alpha < 1/2$, *which does not guarantee consistency of the introduced method in the present case, undoubtedly because of the non-parametric nature of the used estimators. Indeed, in the proofs of the consistency results it is rather* $\alpha < 1/4$ *which is required.*

3.3. Estimation of the dimension For the same reasons as above, and following [11], we will use a penalized estimator of $\xi_{\mathcal{J}_j}$, where $\mathcal{J}_j = \{\tau(\ell) \,:\, 1 \leqslant \ell \leqslant j\}$, in order to estimate d. More precisely, letting

$$\widehat{\mathcal{J}}^{(n)}_j = \left\{\widehat{\tau}(\ell) \,:\, 1 \leqslant \ell \leqslant j\right\},$$

we consider

$$\widehat{\psi}^{(n)}_j = \widehat{\xi}^{(n)}_{\widehat{\mathcal{J}}^{(n)}_j} + \frac{w\left(\widehat{\tau}(j)\right)}{n^{\gamma}},$$

where $0 < \gamma < 1/4$ and $w \,:\, I \to \mathbb{R}_+$ is a strictly increasing function. Then, we take as estimator of d the statistic

$$\widehat{d} = \arg\min_{j \in I}\left(\widehat{\psi}^{(n)}_j\right).$$

Remark 3.3. *Here too, the penalization terms* $w\left(\widehat{\tau}(j)\right)/n^{\gamma}$ *are introduced in order to circumvent the issue of ties in the values of the estimates of the used criterion, which is necessary to prove the consistency of* $\widehat{d}$. *For the tuning parameter* γ, *it is* $\gamma < 1/4$ *which is required rather than* $\gamma < 1/2$ *as in [11]. The reasons for this are the same as those mentioned in the previous remark.*

4. **Consistency Results** In this section, we first introduce the assumptions needed to obtain the main results of the paper, and then we state convergence theorems that give consistency of our variable selection method.

4.1. Assumptions We make the following assumptions:

Assumption 4.1. $\mathbb{E}\left(\|X\|^4\right) < \infty$ *and* $\mathbb{E}\left(Y^4\right) < \infty$.

Assumption 4.2. *There exists a sequence* $(M_n)_{n \geqslant 1}$ *such that* $M_n \sim \sqrt{\ln(n)}$ *as* $n \to \infty$, *and*

$$\max_{1 \leqslant i \leqslant n} \|X^{(i)}\| \leqslant M_n, \quad \max_{1 \leqslant i \leqslant n} |Y_i| \leqslant M_n.$$

Assumption 4.3. *The density* f *of* Z *is bounded.*

ASSUMPTION 4.4. *The η_j's, ν and f are 3-times differentiable and their third derivatives satisfy the Lipschitz conditions*

$$\left|f^{(3)}\left(y+u\right)-f^{(3)}\left(y\right)\right|\leqslant c|u|,\ \ \left|\nu^{(3)}\left(y+u\right)-\nu^{(3)}\left(y\right)\right|\leqslant c|u|$$

and $\left|\eta_j^{(3)}\left(y+u\right)-\eta_j^{(3)}\left(y\right)\right|\leqslant c|u|$, *for* $j\in\{1,\dots,p\}$.

ASSUMPTION 4.5. *There exists a constant $c>0$ such that, for any $(y,u)\in\mathbb{R}^2$ and any $(k,j)\in\{1,\dots,p\}^2$,*

$$|s_k\left(y+u\right)s_j\left(y+u\right)-s_k\left(y\right)s_j\left(y\right)|\leqslant c|u|,\ \ |s_k\left(y+u\right)r\left(y+u\right)-s_k\left(y\right)r\left(y\right)|\leqslant c|u|,$$

and $\left|r^2\left(y+u\right)-r^2\left(y\right)\right|\leqslant c|u|$.

ASSUMPTION 4.6. *Denoting by $\mathbb{I}_{\{.\}}$ the indicator function. Then*

(i) *For any pair (k,j) of $\{1,\dots,p\}^2$, $\sqrt{n}\,\mathbb{E}\left(s_k\left(Z\right)s_j\left(Z\right)\mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)=o(1)$ and $\sqrt{n}\,\mathbb{E}\left(s_k\left(Z\right)r\left(Z\right)\mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)=o(1)$;*
(ii) $\sqrt{n}\,\mathbb{E}\left(r^2\left(Z\right)\mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)=o(1)$.

ASSUMPTION 4.7. *When n is large enough $h_n\sim n^{-c_1}$ and $b_n\sim n^{-c_2}$, where c_1 and c_2 are numbers satisfying $c_1>1/5$, $0<c_2<1/26$ and $1/8+c_2/4<c_1<1/4-3c_2$.*

ASSUMPTION 4.8. (i) *The kernel K is bounded, i.e., $\sup_{u\in\mathbb{R}}|K(u)|=D<\infty$.*
(ii) *The kernel K is continuous and its support is the interval $[-1,\,1]$;*
(iii) *K is symmetric about 0;*
(iv) *The kernel K is of order k, i.e.,*

$$\int K(u)\,du=1\quad\text{and}\quad\int u^kK(u)\,du=0\ \text{for}\ k\in\{1,2,3\}.$$

(v)

$$\int|K(u)|\,du<\infty,\ \int|u|^4\,K(u)\,du<\infty\ .$$

REMARK 4.1. *Assumption 4.1, according to which the moments of order four are finite, a classical assumption in the statistical literature was made, for instance, in [22]. This latter work also considered assumptions on regression functions of the types of Assumptions 4.4, 4.5 and 4.6. Assumption 4.2 is weaker than Assumption 3.1 of [16] where it is supposed that X is bounded. For instance, it has been considered in [15]. Note that it can be proved that the results of [16] still hold under such an assumption. Assumptions 4.3, 4.7 and 4.8 were introduced in [16], the last two being slightly different from the conditions 4 and 5 of [22].*

4.2. Results Here, we give the main results of the paper. The objective is to prove that the introduced method for selecting variables in the model (2.1) is consistent; but to achieve this, a convergence result for the estimator of the introduced criterion is needed. This result relies on the consistency of the introduced estimators of the covariance matrix and the cross-covariance vector. This latter consistency result is given in the following theorem.

Theorem 4.1. *Under the assumptions* 4.1 *to* 4.8*, the sequences* $\widehat{V}_1^{(n)}$ *and* $\widehat{V}_{12}^{(n)}$ *almost surely converge, as* $n \to \infty$*, to* V_1 *and* V_{12}*, respectively.*

As a consequence of this theorem, we obtain the following corollary which gives the strong consistency of the proposed estimator, given in (3.2), of the criterion introduced in (2.6). This result is obviously obtained from Theorem 4.1 by using the continuous mapping theorem.

Corollary 4.1. *Under the assumptions* 4.1 *to* 4.8*, for any subset* J *of* I*, the sequence* $\widehat{\xi}_J^{(n)}$ *almost surely converges to* ξ_J *as* $n \to \infty$*.*

We will now state the main result of this paper. It establishes the consistency of our proposal for variable selection in the partially linear regression model given in (2.1). Since our method consists in estimating the subset I_1 by the one given in (3.1), we just have to derive the asymptotic value of the probability that $\widehat{I}_1^{(n)}$ equals I_1. We have:

Theorem 4.2. *Under the assumptions* 4.1 *to* 4.8*,*

$$\lim_{n\to\infty} \mathbb{P}\left(\widehat{I}_1^{(n)} = I_1\right) = 1.$$

5. **Simulations** This section presents the results of simulations made in order to assess the finite-sample behaviour of the proposed variable selection method and to compare it with existing ones.

5.1. The simulated model We generated each vector $X^{(i)} = (X_{i1}, \dots, X_{ip})^\top$, $i \in \{1, \dots, n\}$, $p \in \{9, 30, 50\}$, according to the multivariate normal distribution $\mathcal{N}(0, \Gamma)$, where $\Gamma = (\Gamma_{kj})_{1 \leqslant k,j \leqslant p}$ with $\Gamma_{kj} = 0.5^{|k-j|}$. The responses Y_i were then generated according to

$$Y_i = X^{(i)\top}\beta + \sin(2\pi Z_i) + \varepsilon_i,$$

where $\beta = (\beta_j)_{1 \leqslant j \leqslant p}$ with $\beta_1 = 0.5$, $\beta_2 = 1$, $\beta_3 = 1.5$, $\beta_4 = 2$, $\beta_6 = 4.5$, and $\beta_j = 0$ for all j in $\{1, \dots, p\} \setminus \{1, 2, 3, 4, 6\}$. The Z_i's are generated independently from the uniform distibution on $[0, 1]$, and the ε_i's from the standard normal distribution. For this example, the true set of relevant variables is $I_1 = \{1, 2, 3, 4, 6\}$ and the nonparametric terms are known, thereby allowing a fair comparison of our method with other existing methods.

5.2. Choosing bandwith and tuning parameters Our procedure depends on two tuning parameters α and γ which may influence its performance. We chose these parameters and the bandwith h_n from the data by using a generalized cross validation procedure (GCV). It consists in minimizing the GCV criterion defined on $J = \{j_1, \ldots, j_k\}$ as

$$\mathrm{GCV}_J = \frac{\|\mathbb{Y} - \widehat{\mathbb{Y}}_J\|^2}{n - \mathrm{Tr}\left(\mathbb{X}_J(\mathbb{X}_J^\top \mathbb{X}_J)^{-1}\mathbb{X}_J^\top\right)},$$

where $\mathbb{Y} = (Y_1 - \widehat{r}_{b_n}(Z_1), \ldots, Y_n - \widehat{r}_{b_n}(Z_n))^\top$,

$$\mathbb{X}_J = \begin{pmatrix} X_{1j_1} - \widehat{s}_{b_n,j_1}(Z_1) & X_{1j_2} - \widehat{s}_{b_n,j_2}(Z_1) & \cdots & X_{1j_k} - \widehat{s}_{b_n,j_k}(Z_1) \\ X_{2j_1} - \widehat{s}_{b_n,j_1}(Z_2) & X_{2j_2} - \widehat{s}_{b_n,j_2}(Z_2) & \cdots & X_{2j_k} - \widehat{s}_{b_n,j_k}(Z_2) \\ \vdots & \vdots & \cdots & \vdots \\ X_{nj_1} - \widehat{s}_{b_n,j_1}(Z_n) & X_{nj_2} - \widehat{s}_{b_n,j_2}(Z_n) & \cdots & X_{nj_k} - \widehat{s}_{b_n,j_k}(Z_n) \end{pmatrix},$$

and $\widehat{\mathbb{Y}}_J = \mathbb{X}_J(\mathbb{X}_J^\top \mathbb{X}_J)^{-1}\mathbb{X}_J^\top \mathbb{Y}$. More specifically, we generated a training sample according to the aforementioned model, and for each triple (h_n, α, γ) in a fine grid of $]0,1[\times]0,1/4[\times]0,1/4[$, we apply our method for selecting variables on the previous sample. We then get $\widehat{I}_1(h_n, \alpha, \gamma)$ as estimate of I_1 and compute $\varphi(h_n, \alpha, \gamma) = \mathrm{GCV}_J$ with $J = \widehat{I}_1(h_n, \alpha, \gamma)$. Then, we chose for the bandwith and the tuning parameters the triple $(\widehat{h}_n, \widehat{\alpha}, \widehat{\gamma})$ that minimizes $\varphi(h_n, \alpha, \gamma)$ over the previous grid of $]0,1[\times]0,1/4[\times]0,1/4[$.

5.3. Simulation strategy and results We simulated 400 independent replicates of samples generated as indicated above. Over these 400 replicates, the following five measurements are computed in order to assess the performance of the methods:

(1) average model size, i.e., MSIZE$= 400^{-1}\sum_{k=1}^{400}|\widehat{I}_{1,k}|$, where $|\widehat{I}_{1,k}|$ is the cardinality the subset $\widehat{I}_{1,k}$ of selected variables at the kth replication;
(2) coverage probability, i.e., CVP$= 400^{-1}\sum_{k=1}^{400}\mathbb{I}_{\{I_1 \subset \widehat{I}_{1,k}\}}$;
(3) equality probability, i.e., EQP$= 400^{-1}\sum_{k=1}^{400}\mathbb{I}_{\{\widehat{I}_{1,k} = I_1\}}$;
(4) average false discovery rate, i.e., FDR$= 400^{-1}\sum_{k=1}^{400} N_k/|\widehat{I}_{1,k}|$, where N_k is the number of false discovery variables for the kth replication;
(5) average mean-squared errors of prediction (MSE) after variable selection, i.e., MSE $= 400^{-1}\sum_{k=1}^{400}\mathrm{MSE}_k$, where $\mathrm{MSE}_k = (1/n)\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$, $\widehat{Y}_i$ being the predictor computed according to (2.3) with only the variables selected at the kth replication;
(6) average computational time (TIME) in seconds.

We compared our method (OM) to four known methods, as reported in [4]. They consist in minimizing penalized least squares, each being related to a particular penalty function: SCAD penalty, Hard thresholding penalty (Hard), L^1 penalty

(LASSO) and Ridge regression (Ridge). Our method was performed by using the Gaussian kernel $K(t) = (2\pi)^{-1/2}e^{-t^2/2}$, the penalty functions $v(x) = \ln(x)^{-0.5}$, $w(x) = \ln(x)^{0.5}$ and by taking $b_n = \min(0.001, n^{-1/27}) = 0.001$. The obtained results are reported in Tables 1 to 3. It can be seen that our method outperforms the four others in most cases. Indeed, regarding MSIZE, this method obtained the closest values to the true number of relevant variables (which equals 5), while the other methods get values that are quite far from it. All the methods had very high values for CVP, but our method got the best results for EQP for which the four other methods gave very bad results, thereby showing that they select the right variables but also several irrelevant variables. It is also seen through the values obtained for FDR, where the four other methods obtained higher values than ours. The latter seems to select the right variables but not additional irrelevant variables, especially for $n = 200$. Concerning MSE, all the five methods gave comparable results. However, our method presents, in all cases, a higher computational time than that of the four other methods.

Table 1: Measurements for different methods over 400 replicates, with sample size $n = 100$, 200 and dimension $p = 9$.

Sample size	Method	Measurement					
		MSIZE	CVP	EQP	FDR	MSE	TIME
$n = 100$	OM	4.960	0.980	0.970	0.000	0.250	8.025
	SCAD	6.335	1.000	0.335	0.185	0.245	4.545
	Hard	7.000	1.000	0.335	0.245	0.245	3.940
	LASSO	8.000	1.000	0.000	0.370	0.245	4.275
	Ridge	9.000	1.000	0.000	0.445	0.245	4.240
$n = 200$	OM	4.995	0.935	0.995	0.000	0.185	33.95
	SCAD	6.335	1.000	0.4000	0.175	0.185	27.05
	Hard	7.465	1.000	0.215	0.295	0.190	25.85
	LASSO	8.735	1.000	0.000	0.425	0.185	26.60
	Ridge	9.000	1.000	0.000	0.445	0.185	26.55

Table 2: Measurements for different methods over 400 replicates, with sample size $n = 100$, 200 and dimension $p = 30$.

Sample size	Method	Measurement					
		MSIZE	CVP	EQP	FDR	MSE	TIME
$n = 100$	OM	5.125	0.635	0.655	0.045	0.825	58.45
	SCAD	10.65	1.000	0.080	0.455	0.825	13.50
	Hard	9.920	0.880	0.140	0.375	0.825	12.45
	LASSO	17.65	1.000	0.000	0.705	0.825	13.20
	Ridge	30.00	1.000	0.000	0.835	0.825	13.25
$n = 200$	OM	5.025	0.955	0.825	0.025	0.570	190.5
	SCAD	10.35	1.000	0.025	0.475	0.570	89.55
	Hard	15.20	1.000	0.050	0.610	0.570	87.35
	LASSO	20.80	1.000	0.000	0.745	0.570	89.85
	Ridge	30.00	1.000	0.000	0.835	0.570	90.75

Table 3: Measurements for different methods over 400 replicates, with sample size $n = 100$, 200 and dimension $p = 50$.

Sample size	Method	Measurement					
		MSIZE	CVP	EQP	FDR	MSE	TIME
$n = 100$	OM	5.150	0.775	0.565	0.045	0.745	140.5
	SCAD	22.50	1.000	0.050	0.700	0.745	24.65
	Hard	16.10	0.850	0.025	0.605	0.755	23.45
	LASSO	23.20	1.000	0.000	0.775	0.745	24.40
	Ridge	49.95	1.000	0.000	0.895	0.745	24.35
$n = 200$	OM	5.475	0.625	0.575	0.035	0.795	296.5
	SCAD	25.15	1.000	0.025	0.745	0.805	150.5
	Hard	14.95	0.975	0.050	0.535	0.805	147.5
	LASSO	25.15	1.000	0.000	0.795	0.795	148.5
	Ridge	50.00	1.000	0.000	0.900	0.795	150.5

6. Proofs

6.1. Proof of Theorem 4.1 We will just give the proof of the convergence of the $\widehat{V}_1^{(n)}$ since that of $\widehat{V}_{12}^{(n)}$ is obtained from a similar reasoning. We have

$$V_1 = \mathbb{E}\big(XX^\top\big) - \Sigma - \Sigma^\top + \Lambda,$$

where $\Sigma = \mathbb{E}\big(Xs(Z)^\top\big)$ and $\Lambda = \mathbb{E}\big(s(Z)s(Z)^\top\big)$. Similarly,

$$\widehat{V}_1^{(n)} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)}(X^{(i)})^\top - \widehat{\Sigma}_n - \widehat{\Sigma}_n^\top + \widehat{\Lambda}_n$$

where $\widehat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^{n} X^{(i)}\widehat{s}_{b_n}(Z_i)^\top$ and $\widehat{\Lambda}_n = \frac{1}{n}\sum_{i=1}^{n} \widehat{s}_{b_n}(Z_i)\widehat{s}_{b_n}(Z_i)^\top$. It is then sufficient to obtain the following almost sure convergences:

$$\frac{1}{n}\sum_{i=1}^{n} X^{(i)}(X^{(i)})^\top \to \mathbb{E}\big(XX^\top\big), \tag{6.1}$$

$$\widehat{\Sigma}_n \to \Sigma, \tag{6.2}$$

$$\widehat{\Lambda}_n \to \Lambda, \tag{6.3}$$

as $n \to \infty$. The convergence in (6.1) is achieved through the law of large numbers. That of (6.3) is established in Theorem 3.3 of [16]. It remains to obtain (6.2). For this purpose, noticing that

$$\widehat{\Sigma} = \big(\sigma_{kj}\big)_{1\leqslant k,j\leqslant p} \quad \text{and} \quad \widehat{\Sigma}_n = \big(\widehat{\sigma}_{kj}^{(n)}\big)_{1\leqslant k,j\leqslant p},$$

where

$$\sigma_{kj} = \mathbb{E}\big(X_k s_j(Z)\big) = \mathbb{E}\left(\frac{X_k\eta_j(Z)}{f(Z)}\right)$$

and

$$\widehat{\sigma}_{kj}^{(n)} = \frac{1}{n}\sum_{i=1}^{n} X_{ik}\widehat{s}_{b_n,j}(Z_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{X_{ik}\widehat{\eta}_{j,n}(Z_i)}{\widehat{f}_{b_n}(Z_i)},$$

it is enough to prove that for all $(k,j) \in \{1,\dots,p\}^2$, $\widehat{\sigma}_{kj}^{(n)}$ converges almost surely to σ_{kj} as $n \to \infty$, what is obtained from the following properties:

$$\widehat{\sigma}_{kj}^{(n)} - \mathbb{E}\left(\widehat{\sigma}_{kj}^{(n)}\right) = O_{a.s.}\left(\frac{(\ln(n))^{1-c_1-c_2}}{n^{\frac{1}{2}-(c_1+c_2)}}\right), \tag{6.4}$$

and

$$\lim_{n\to\infty} \mathbb{E}\left(\widehat{\sigma}_{kj}^{(n)}\right) = \sigma_{kj}. \tag{6.5}$$

Proof of (6.4): From the class of functions

$$\mathcal{H}_n = \left\{ g_{(k,j)} : (x = (x_1, x_2, \ldots, x_p)^\top, z) \longmapsto g_{k,j}(x,z) = \frac{1}{n} \frac{x_k \widehat{\eta}_{j,n}(z)}{\widehat{f}_{b_n}(z)}, 1 \leqslant k, j \leqslant p \right\},$$

we use a similar reasoning than in the proof of Theorem 3.1 of [16] (see p. 1299). Since $|\widehat{\eta}_{j,n}(z)| \leqslant \frac{1}{nh_n} \sum_{i=1}^n |X_{ij} K(\frac{z-Z_i}{h_n})| \leqslant \frac{DM_n}{h_n}$ and $\widehat{f}_{b_n}(z) \geq b_n$, we have for any $g \in \mathcal{H}_n$,

$$\mathbb{E}\left(g(X,Z)\right) \leqslant \frac{D}{h_n n^2 b_n} \sum_{i=1}^n \mathbb{E}\left(|X_k||X_{ij}|\right) \leqslant \frac{DM_n \mathbb{E}\left(\|X\|\right)}{h_n n b_n} \leqslant \frac{DM_n \mathbb{E}\left(\|X\|^2\right)^{1/2}}{h_n n b_n} =: \mu_n$$

and

$$\mathbb{E}\left(g^2(X,Z)\right) \leqslant \frac{D^2}{h_n^2 n^4 b_n^2} \mathbb{E}\left(\left(\sum_{i=1}^n |X_k||X_{ij}| \right)^2 \right) \leqslant \frac{D^2 M_n^2 \mathbb{E}\left(\|X\|^2\right)}{h_n^2 n^2 b_n^2} =: \sigma_n^2.$$

Then, we can apply Talagrand's inequality (see [18] and Proposition 2.2 of [6]): there exist constants $A > 0$, $K_1 > 0$ and $K_2 > 0$ such that for all scalar t satisfying the inequality

$$t \geqslant K_1 \left[\mu_n \ln\left(\frac{A\mu_n}{\sigma_n}\right) + \sqrt{n}\sigma_n \sqrt{\ln\left(\frac{A\mu_n}{\sigma_n}\right)} \right]$$

one has

$$\sum_{n=1}^{\infty} P\left\{ \sup_{g \in \mathcal{H}_n} \left| \sum_{i=1}^n \left\{ g\left(X_i, Z_i\right) - \mathbb{E}\Big(h\left(X, Z\right) \Big) \right\} \right| > t \right\}$$

$$\leqslant K_2 \exp\left\{ -\frac{1}{K_2} \frac{t}{\mu_n} \log\left(1 + \frac{t\mu_n}{K_2 \left(\sqrt{n}\sigma_n + \mu_n \sqrt{\log \frac{A\mu_n}{\sigma_n}} \right)^2} \right) \right\},$$

i.e.,

$$\sum_{n=1}^{\infty} P\left\{ \sup_{1 \leqslant k,j \leqslant p} \left| \frac{1}{n} \sum_{i=1}^n \frac{X_{ik} \widehat{\eta}_{j,n}\left(Z_i\right)}{\widehat{f}_{b_n}\left(Z_i\right)} - \mathbb{E}\left(\frac{X_k \widehat{\eta}_{j,n}\left(Z\right)}{\widehat{f}_{b_n}\left(Z\right)} \right) \right| > t \right\}$$

$$\leqslant K_2 \exp\left\{ -\frac{1}{K_2} \frac{n h_n b_n t}{DCM_n} \log\left(1 + \frac{h_n b_n t}{K_2\, DCM_n \left(1 + \sqrt{\ln(A)}\right)^2} \right) \right\},$$

where $C = \mathbb{E}\left(\|X\|^2\right)$. Since $h_n \sim n^{-c_1}$ and $b_n \sim n^{-c_2}$, we have

$$\lim_{n\to\infty}\left(n^{c_1}h_n\right) = \lim_{n\to\infty}\left(n^{c_2}b_n\right) = 1.$$

Thus, for n large enough, we have $h_nb_n > n^{-c_1-c_2}/4$. Hence

$$P\left\{\sup_{1\leq k,j\leq p}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ik}\widehat{\eta}_{j,n}\left(Z_i\right)}{\widehat{f}_{b_n}\left(Z_i\right)} - \mathbb{E}\left(\frac{X_k\widehat{\eta}_{j,n}\left(Z\right)}{\widehat{f}_{b_n}\left(Z\right)}\right)\right| > t\right\}$$

$$\leqslant K_2\exp\left\{-\frac{1}{K_2}\frac{n^{1-c_1-c_2}t}{4DCM_n}\log\left(1+\frac{n^{-c_1-c_2}t}{4K_2\,DCM_n\left(1+\sqrt{\ln(A)}\right)^2}\right)\right\}. \tag{6.6}$$

Let us put $t_n = (\ln(n))^{1-c_1-c_2}/n^{\frac{1}{2}-(c_1+c_2)}$; Assumption 4.7 implies that $\theta-\frac{1}{2} > 0$ where $\theta = 1-c_1-c_2$. Then,

$$\lim_{n\to\infty}\frac{(\ln(n))^{\theta}}{M_n} = \lim_{n\to\infty}\left(\ln(n)\right)^{\theta-\frac{1}{2}} = \infty$$

and, therefore, we have for n large enough

$$\frac{(\ln(n))^{\theta}}{M_n} \geqslant 4K_1DC\sqrt{\ln(A)} \geqslant \frac{2K_1DC}{\sqrt{n}}\sqrt{\ln(A)}\left(\sqrt{n}+\sqrt{\ln(A)}\right). \tag{6.7}$$

Further, since $\lim_{n\to\infty}\left(nh_nb_n/\sqrt{n}\,n^{\frac{1}{2}-c_1-c_2}\right) = 1$, then for n large enough $1/\sqrt{n} \geqslant n^{\frac{1}{2}-c_1-c_2}/\left(2nh_nb_n\right)$. Pluging this latter inequality in (6.7) yields

$$t_n \geqslant \frac{K_1DCM_n}{nh_nb_n}\sqrt{\ln(A)}\left(\sqrt{n}+\sqrt{\ln(A)}\right),$$

what is equivalent to

$$t_n \geqslant K_1\left[\mu_n\ln\left(\frac{A\mu_n}{\sigma_n}\right)+\sqrt{n}\sigma_n\sqrt{\ln\left(\frac{A\mu_n}{\sigma_n}\right)}\right].$$

Then, (6.6) can be applied to t_n and we obtain

$$P\left\{\sup_{1\leqslant k,j\leqslant p}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ik}\widehat{\eta}_{j,n}\left(Z_i\right)}{\widehat{f}_{b_n}\left(Z_i\right)} - \mathbb{E}\left(\frac{X_k\widehat{\eta}_{j,n}\left(Z\right)}{\widehat{f}_{b_n}\left(Z\right)}\right)\right| > t_n\right\} \leqslant u_n, \tag{6.8}$$

where

$$u_n = K_2\exp\left\{-\frac{1}{K_2}\frac{\sqrt{n}(\ln(n))^{\theta}}{4DCM_n}\log\left(1+\frac{(\ln(n))^{\theta}}{4K_2DC\sqrt{n}M_n\left(1+\sqrt{\ln(A)}\right)^2}\right)\right\}.$$

Clearly, $u_n \sim z_n$ as $n \to \infty$, where

$$z_n = K_2 \exp\left\{ -\frac{(\ln(n))^{2\theta}}{16K_2^2 D^2 C^2 M_n^2 \left(1+\sqrt{\ln(A)}\right)^2} \right\},$$

and since $\sum_{n=1}^{\infty} z_n < \infty$ because $\lim_{n\to\infty}\left(\frac{(\ln(n))^{\theta}}{M_n}\right) = \infty$, we deduce that $\sum_{n=1}^{\infty} u_n < \infty$. Then from (6.8) it follows that

$$\sum_{n=1}^{\infty} P\left\{ \sup_{1\leqslant k,j\leqslant p} \left| \frac{1}{n}\sum_{i=1}^{n} \frac{X_{ik}\widehat{\eta}_{j,n}(Z_i)}{\widehat{f}_{b_n}(Z_i)} - \mathbb{E}\left(\frac{X_k\widehat{\eta}_{j,n}(Z)}{\widehat{f}_{b_n}(Z)}\right) \right| > t_n \right\} < \infty,$$

and the required result is obtained from the Borel-Cantelli's lemma.

Proof of (6.5): Considering

$$R_{b_n,j}(y) = \frac{\eta_j(y)}{f_{b_n}^{1/2}(y)},\ \ \widetilde{R}_{b_n,j}(y) = \frac{X_{ij}}{f_{b_n}^{1/2}(y)},\ \ I_{kj,n}^{(1)}(y) = \frac{X_{ik}\eta_j(y)}{f_{b_n}(y)} = \widetilde{R}_{b_n,ik}(y)R_{b_n,j}(y),$$

$$I_{kj,n}^{(2)}(y) = \frac{X_{ik}\widehat{\eta}_{j,n}(y)+\eta_j(y)X_{ik}}{f_{b_n}(y)} = \frac{\widetilde{R}_{b_n,k}(y)\widehat{\eta}_{j,n}(y)}{f_{b_n}^{1/2}(y)} + R_{b_n,j}(y)\widetilde{R}_{b_n,k}(y),$$

and

$$I_{kj,n}^{(3)}(y) = 2R_{b_n,j}(y)\widetilde{R}_{b_n,k}(y)\frac{\widehat{f}_{b_n}^{1/2}(y)}{f_{b_n}^{1/2}(y)},$$

we have

$$\widehat{\sigma}_{kj}^{(n)} = \frac{1}{n}\sum_{i=1}^{n}\left\{ I_{kj,n}^{(1)}(Z_i) + I_{kj,n}^{(2)}(Z_i) - I_{kj,n}^{(3)}(Z_i) \right\} - A_n + B_n - C_n,$$

where

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\left\{ X_{ik}\left(\widehat{\eta}_{j,n}(Z_i) - \eta_j(Z_i)\right)\right\}\left(\frac{\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)}{\widehat{f}_{b_n}(Z_i)\, f_{b_n}(Z_i)}\right),$$

$$B_n = \frac{1}{n}\sum_{i=1}^{n}\widetilde{R}_{b_n,k}(Z_i)\, R_{b_n,j}(Z_i)\,\frac{\left(\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)\right)^2}{\widehat{f}_{b_n}(Z_i)\, f_{b_n}(Z_i)},$$

and

$$C_n = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{f}_{b_n}^{1/2}(Z_i) - f_{b_n}^{1/2}(Z_i)\right)^2 \frac{\widetilde{R}_{b_n,k}(Z_i)\, R_{b_n,j}(Z_i)}{f_{b_n}(Z_i)}.$$

From Theorem 2.1.8 in [17], we have

$$\sup_{z\in\mathbb{R}}\left|\widehat{f}_n(z)-f(z)\right| = O_{a.s.}(\rho_n)\,, \tag{6.9}$$

where $\rho_n = h_n^4 + n^{-1/2}h_n^{-1}\ln(n)$, and since $\left|\widehat{f}_{b_n}(Z)-f_{b_n}(Z)\right| \leqslant \left|\widehat{f}_n(Z)-f(Z)\right|$ (see Eq. (4.4) in [22]), we deduce by using the Cauchy-Schwarz inequality and Lemma 4.6 in [16] that

$$\begin{aligned}
|\mathbb{E}(A_n)| &\leqslant \frac{K_3\rho_n}{nb_n^2}\sum_{i=1}^n \mathbb{E}\left(\left|X_{ik}\left(\widehat{\eta}_{j,n}(Z_i)-\eta_j(Z_i)\right)\right|\right)\\
&\leqslant \frac{K_3\rho_n}{nb_n^2}\sum_{i=1}^n \mathbb{E}\left(X_{ik}^2\right)^{1/2}\mathbb{E}\left(\left(\widehat{\eta}_{j,n}(Z_i)-\eta_j(Z_i)\right)^2\right)^{1/2}\\
&\leqslant \frac{K_4\rho_n\lambda_n^{1/2}M_n}{b_n^2}, \qquad (6.10)
\end{aligned}$$

where $\lambda_n = n^{-1}h_n^{-1}$ and K_3 and K_4 are positive constants. Further,

$$|\mathbb{E}(B_n)| \leqslant \frac{K_3^2\rho_n^2}{nb_n^3}\sum_{i=1}^n \mathbb{E}\left(|X_{ik}\eta_j(Z_i)|\right) \leqslant \frac{K_5\rho_n^2M_n}{b_n^3}, \tag{6.11}$$

where $K_5 > 0$, and

$$\begin{aligned}
|\mathbb{E}(C_n)| &\leqslant \frac{1}{nb_n^2}\sum_{i=1}^n \mathbb{E}\left(|X_{ik}\eta_j(Z_i)|\left(\widehat{f}_{b_n}^{1/2}(Z_i)-f_{b_n}^{1/2}(Z_i)\right)^2\right)\\
&= \frac{1}{nb_n^2}\sum_{i=1}^n \mathbb{E}\left(|X_{ik}\eta_j(Z_i)|\left(\frac{\widehat{f}_{b_n}(Z_i)-f_{b_n}(Z_i)}{\widehat{f}_{b_n}^{1/2}(Z_i)+f_{b_n}^{1/2}(Z_i)}\right)^2\right)\\
&\leqslant \frac{K_3^2\rho_n^2}{4nb_n^3}\sum_{i=1}^n \mathbb{E}\left(|X_{ik}\eta_j(Z_i)|\right)\\
&\leqslant \frac{K_5\rho_n^2M_n}{b_n^3}. \qquad (6.12)
\end{aligned}$$

Since $\lambda_n^{1/2} \sim n^{(c_1-1)/2} \to 0$ as $n\to\infty$ because $c_1 < 1$, it follows from (6.10), (6.11) and (6.12) that $\mathbb{E}(-A_n+B_n-C_n) = O\left(b_n^{-3}\rho_nM_n\right)$ and, therefore,

$$\mathbb{E}\left(\widehat{\sigma}_{kj}^{(n)}\right) = \mathbb{E}\left(I_{kj,n}^{(1)}(Z)\right) + \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(I_{kj,n}^{(2)}(Z_i)-I_{kj,n}^{(3)}(Z_i)\right) + O\left(\frac{\rho_nM_n}{b_n^3}\right). \tag{6.13}$$

Furthermore,

$$\begin{aligned}
&\left|\mathbb{E}\left(I^{(2)}_{kj,n}\left(Z_i\right) - I^{(3)}_{kj,n}\left(Z_i\right)\right)\right| \\
&\quad \leqslant \quad \mathbb{E}\left(\left|\frac{\widetilde{R}_{b_n,k}(Z_i)\widehat{\eta}_{j,n}(Z_i)}{f^{1/2}_{b_n}(Z_i)} - \frac{\widetilde{R}_{b_n,k}(Z)R_{b_n,j}(Z_i)\widehat{f}^{1/2}_{b_n}(Z_i)}{f^{1/2}_{b_n}(Z_i)}\right|\right) \\
&\quad + \quad \mathbb{E}\left(\left|\widetilde{R}_{b_n,k}(Z_i)R_{b_n,j}(Z_i) - \frac{\widetilde{R}_{b_n,k}(Z_i)R_{b_n,j}(Z_i)\widehat{f}^{1/2}_{b_n}(Z_i)}{f^{1/2}_{b_n}(Z_i)}\right|\right) \\
&\quad =: \quad E_{1n} + E_{2n}
\end{aligned}$$

with

$$\begin{aligned}
E_{2n} &= \mathbb{E}\left(\left|\widetilde{R}_{b_n,k}(Z_i)R_{b_n,j}(Z_i)\right|\left|\frac{\widehat{f}^{1/2}_{b_n}(Z_i) - f^{1/2}_{b_n}(Z_i)}{f^{1/2}_{b_n}(Z_i)}\right|\right) \\
&= \mathbb{E}\left(\left|\widetilde{R}_{b_n,k}(Z_i)R_{b_n,j}(Z_i)\right|\frac{\left|\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)\right|}{f^{1/2}_{b_n}(Z_i)\left(\widehat{f}^{1/2}_{b_n}(Z_i) + f^{1/2}_{b_n}(Z_i)\right)}\right) \\
&\leqslant \frac{K_3\rho_n}{b_n^2}\mathbb{E}\left(\left|X_{ik}\eta_j\left(Z_i\right)\right|\right) \\
&\leqslant \frac{K_6\rho_n M_n}{b_n^2}, \qquad (6.14)
\end{aligned}$$

where $K_6 > 0$. By using the Cauchy-Schwarz inequality we get

$$\begin{aligned}
E_{1n} &\leqslant \mathbb{E}\left(\frac{\left(\widetilde{R}_{b_n,k}(Z_i)\right)^2}{f_{b_n}(Z_i)}\right)^{1/2}\mathbb{E}\left(\left(\widehat{\eta}_{j,n}\left(Z_i\right) - R_{b_n,j}(Z_i)\widehat{f}^{1/2}_{b_n}(Z_i)\right)^2\right)^{1/2} \\
&= \mathbb{E}\left(\frac{X_{ik}^2}{f^2_{b_n}(Z_i)}\right)^{1/2}\mathbb{E}\left(\left(\widehat{\eta}_{j,n}\left(Z_i\right) - \eta_j(Z_i)\frac{\widehat{f}^{1/2}_{b_n}(Z_i)}{f^{1/2}_{b_n}(Z_i)}\right)^2\right)^{1/2} \\
&\leqslant \frac{M_n}{b_n}\left\{2\mathbb{E}\left(\left(\widehat{\eta}_{j,n}\left(Z_i\right) - \eta_j(Z_i)\right)^2\right) + 2\mathbb{E}\left(\eta_j^2(Z_i)\left(1 - \frac{\widehat{f}^{1/2}_{b_n}(Z_i)}{f^{1/2}_{b_n}(Z_i)}\right)^2\right)\right\}^{1/2} \\
&= \frac{\sqrt{2}M_n}{b_n}\left\{\mathbb{E}\left(\left(\widehat{\eta}_{j,n}\left(Z_i\right) - \eta_j(Z_i)\right)^2\right) + \mathbb{E}\left(\eta_j^2(Z_i)\frac{\left(\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)\right)^2}{f_{b_n}(Z_i)\left(\widehat{f}^{1/2}_{b_n}(Z_i) + f^{1/2}_{b_n}(Z_i)\right)^2}\right)\right\}^{1/2}.
\end{aligned}$$

By using (6.9) and Lemma 4.6 in [16] we then obtain

$$\begin{aligned} E_{1n} &\leqslant \frac{\sqrt{2}M_n}{b_n}\left\{K_7\lambda_n + K_8\frac{\rho_n^2}{b_n^2}\right\}^{1/2} \\ &= \frac{\sqrt{2}\rho_n M_n}{b_n^2}\left\{K_7\frac{b_n^2\lambda_n}{\rho_n^2} + K_8\right\}^{1/2}, \end{aligned} \tag{6.15}$$

where $\lambda_n = n^{-1}h_n^{-1}$, and K_7 and K_8 are positive constants. It is easy to verify that $\rho_n \sim n^{-1/2}h_n^{-1}\ln(n)$ and, consequently, that

$$\frac{b_n^2\lambda_n}{\rho_n^2} \sim n^{-2c_2}h_n\ln^2(n) \sim n^{-c_1-2c_2}\ln^{-2}(n),$$

which yields $\lim_{n\to\infty}\left(\rho_n^{-2}b_n^2\lambda_n\right) = 0$. Then, from (6.15) one deduces that $E_{1n} = O\left(b_n^{-2}\rho_n M_n\right)$, which, together with (6.14), allows to conclude that

$$n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(I_{kj,n}^{(2)}(Z_i) - I_{kj,n}^{(3)}(Z_i)\right) = O\left(b_n^{-2}\rho_n M_n\right).$$

Since

$$b_n^{-3}\rho_n M_n \sim n^{c_1+3c_2-1/2}\ln^{-1/2}(n),$$

it follows that $\lim_{n\to\infty} b_n^{-3}\rho_n M_n = 0$ and, consequently, that $\lim_{n\to\infty} b_n^{-2}\rho_n M_n = 0$. Then, (6.13) implies that $\lim_{n\to\infty}\mathbb{E}\left(\widehat{\sigma}_{k,l}^{(n)}\right) = \lim_{n\to\infty}\mathbb{E}\left(I_{kj,n}^{(1)}(Z)\right)$. Moreover, since $\left|\mathbb{E}\left(I_{kj,n}^{(1)}(Z)\right)\right| \leqslant \left|\frac{X_k\eta_j(Z)}{f(Z)}\right|$, and

$$\mathbb{E}\left(\left|\frac{X_k\eta_j(Z)}{f(Z)}\right|\right) \leqslant \mathbb{E}\left(X_k^2\right)^{1/2}\mathbb{E}\left(s_j^2(Z)\right)^{1/2} \leqslant \mathbb{E}\left(X_k^2\right)^{1/2}\mathbb{E}\left(X_j^2\right)^{1/2} < \infty,$$

we can apply the dominated convergence theorem that gives

$$\lim_{n\to\infty}\mathbb{E}\left(I_{kj,n}^{(1)}(Z)\right) = \mathbb{E}\left(\frac{X_k\eta_j(Z)}{f(Z)}\right) = \sigma_{kj}.$$

6.2. A technical lemma In this section, we give a lemma that is useful for proving Theorem 4.2. Let us introduce the $\mathbb{R}^{p+1}$-valued random vectors

$$U = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} U_1 \\ \vdots \\ U_{p+1} \end{pmatrix},\ U^{(i)} = \begin{pmatrix} X^{(i)} \\ Y^{(i)} \end{pmatrix} = \begin{pmatrix} U_{i1} \\ \vdots \\ U_{ip+1} \end{pmatrix},$$

where

$$U_j = \begin{cases} X_j & \text{if } 1 \leqslant j \leqslant p \\ Y & \text{if } j = p+1 \end{cases} \quad \text{and} \quad U_{ij} = \begin{cases} X_{ij} & \text{if } 1 \leqslant j \leqslant p \\ Y_i & \text{if } j = p+1 \end{cases},$$

and the functions

$$\mathscr{T}(z) = \begin{pmatrix} s(z) \\ r(z) \end{pmatrix} = \begin{pmatrix} \mathscr{T}_1(z) \\ \vdots \\ \mathscr{T}_{p+1}(z) \end{pmatrix}, \quad \widehat{\mathscr{T}_{b_n}}(z) = \begin{pmatrix} \widehat{s}_{b_n}(z) \\ \widehat{r}_{b_n}(z) \end{pmatrix} = \begin{pmatrix} \widehat{\mathscr{T}_{b_n,1}}(z) \\ \vdots \\ \widehat{\mathscr{T}_{b_n,p+1}}(z) \end{pmatrix},$$

where

$$\mathscr{T}_j(z) = \frac{\mathfrak{X}_j(z)}{f(z)}, \quad \widehat{\mathscr{T}_{b_n,j}}(z) = \frac{\widehat{\mathfrak{X}}_{j,n}(z)}{\widehat{f}_{b_n}(z)}$$

with

$$\mathfrak{X}_j(z) = \begin{cases} \eta_j(z) & \text{if } 1 \leqslant j \leqslant p \\ \nu(z) & \text{if } j = p+1 \end{cases} \quad \text{and} \quad \widehat{\mathfrak{X}}_{j,n}(z) = \begin{cases} \widehat{\eta}_{j,n}(z) & \text{if } 1 \leqslant j \leqslant p \\ \widehat{\nu}(z) & \text{if } j = p+1 \end{cases}.$$

We then have the following result:

LEMMA 6.1. *Under Assumptions* 4.1 *to* 4.8, *for all* $(k, j) \in \{1, \ldots, p+1\}^2$,

$$n^{\alpha-1/2} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(U_{ik} \widehat{\mathscr{T}_{b_n,j}}(Z_i) - \mathbb{E}\left(U_k \mathscr{T}_j(Z)\right) \right) \right| = o_p(1).$$

Proof. Since

$$\begin{aligned} U_{ik} \widehat{\mathscr{T}_{b_n,j}}(Z_i) &= -\frac{U_{ik}\left(\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)\right)\left(\widehat{\mathfrak{X}}_{j,n}(Z_i) - \mathfrak{X}_j(Z_i)\right)}{\widehat{f}_{b_n}(Z_i) f_{b_n}(Z_i)} \\ &\quad -\frac{U_{ik}\left(\widehat{\mathfrak{X}}_{j,n}(Z_i) - \mathfrak{X}_j(Z_i)\right)}{f_{b_n}(Z_i)} - \frac{U_{ik}\mathfrak{X}_j(Z_i)\left(\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)\right)}{\widehat{f}_{b_n}(Z_i) f_{b_n}(Z_i)} \\ &\quad +\frac{U_{ik}\mathscr{T}_j(Z_i)\left(f(Z_i) - f_{b_n}(Z_i)\right)}{f_{b_n}(Z_i)} + U_{ik}\mathscr{T}_j(Z_i), \end{aligned}$$

it follows that

$$n^{\alpha-1/2} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(U_{ik} \widehat{\mathscr{T}_{b_n,j}}(Z_i) - \mathbb{E}\left(U_k \mathscr{T}_j(Z)\right) \right) \right| \leqslant F_{n,1}+F_{n,2}+F_{n,3}+F_{n,4}+F_{n,5},$$

where

$$F_{n,1} = n^{\alpha-1}\sum_{i=1}^{n}\frac{|U_{ik}|\left|\widehat{f}_{b_n}(Z_i)-f_{b_n}(Z_i)\right|\left|\widehat{\mathcal{X}}_{j,n}(Z_i)-\mathcal{X}_j(Z_i)\right|}{\widehat{f}_{b_n}(Z_i)f_{b_n}(Z_i)},$$

$$F_{n,2} = n^{\alpha-1}\sum_{i=1}^{n}\frac{|U_{ik}|\left|\widehat{\mathcal{X}}_{j,n}(Z_i)-\mathcal{X}_j(Z_i)\right|}{f_{b_n}(Z_i)},$$

$$F_{n,3} = n^{\alpha-1}\sum_{i=1}^{n}\frac{|U_{ik}|\,|\mathcal{X}_j(Z_i)|\left|\widehat{f}_{b_n}(Z_i)-f_{b_n}(Z_i)\right|}{\widehat{f}_{b_n}(Z_i)f_{b_n}(Z_i)},$$

$$\begin{aligned}F_{n,4} &= n^{\alpha-1}\sum_{i=1}^{n}\frac{|U_{ik}|\,|\mathscr{T}_j(Z_i)|\,|f(Z_i)-f_{b_n}(Z_i)|}{f_{b_n}(Z_i)}\\ &= n^{\alpha-1}\sum_{i=1}^{n}\frac{|U_{ik}|\,|\mathscr{T}_j(Z_i)|\,|f(Z_i)-f_{b_n}(Z_i)|\,\mathbb{I}_{\{f(Z_i)\leqslant b_n\}}}{f_{b_n}(Z_i)},\end{aligned}$$

$$F_{n,5} = n^{\alpha-1/2}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(U_{ik}\mathscr{T}_j(Z_i)-\mathbb{E}\left(U_k\mathscr{T}_j(Z)\right)\right)\right|, \tag{6.16}$$

and it remains to prove that $F_{n,\ell}=o_p(1)$ for $\ell\in\{1,\dots,5\}$. Using Lemma 3.3 of [22] and Theorem 2.1.8 of [17], we get $F_{n,1}\leqslant L_1 b_n^{-2}M_n n^{\alpha}\rho_n^2$, where $\rho_n = h_n^4+n^{-1/2}h_n^{-1}\ln(n)$ and L_1 is a positive constant. Since $\rho_n\sim n^{-1/2}h_n^{-1}\ln(n)$, it follows from Assumptions 4.2 and 4.7 that

$$b_n^{-2}M_n n^{\alpha}\rho_n^2 \sim n^{(2c_1+2c_2-1/2)+(\alpha-1/2)}\ln^{5/2}(n),$$

which yields $F_{n,1}=o_p(1)$ since $\alpha<1/2$ and, from Assumption 4.7, $c_1+c_2<1/4$. By a similar reasoning we get $F_{n,2}\leqslant L_2 b_n^{-1}M_n n^{\alpha-1/2}\rho_n$, which implies that $F_{n,2}=o_p(1)$ since $b_n^{-1}M_n n^{\alpha-1/2}\rho_n\sim n^{\alpha+c_1+c_2-1/2}\ln^{3/2}(n)$ and $\alpha+c_1+c_2-1/2<0$, because $\alpha<1/4$ and $c_1+c_2<1/4$. Furthermore,

$$F_{n,3}\leqslant L_3 b_n^{-2}M_n n^{\alpha}\rho_n\left(\frac{1}{n}\sum_{i=1}^{n}|\mathcal{X}_j(Z_i)|\right),$$

where L_3 is a positive constant. Since $\mathbb{E}\left(|\mathcal{X}_j(Z)|\right)<\infty$ (see Lemma 4.2 in [16]), we can apply the strong law of large numbers; it ensures that $\frac{1}{n}\sum_{i=1}^{n}|\mathcal{X}_j(Z_i)|$ converges almost surely to $\mathbb{E}\left(|\mathcal{X}_j(Z)|\right)$ as $n\to\infty$. Therefore, since $b_n^{-2}M_n n^{\alpha}\rho_n\sim n^{\alpha+c_1+2c_2-1/2}\ln^{3/2}(n)$ and $\alpha+c_1+2c_2-1/2<0$, because $\alpha<1/4$ and $c_1+2c_2<1/4$, it follows from the above inequality that $F_{n,3}=o_p(1)$. Furthermore, since

$$\left|\widehat{f}_{b_n}(z)-f(z)\right|\leqslant\left|\widehat{f}_{b_n}(z)-\widehat{f}_n(z)\right|+\left|\widehat{f}_n(z)-f(z)\right|\leqslant b_n+\left|\widehat{f}_n(z)-f(z)\right|\leqslant b_n+C\rho_n$$

almost surely, we get, by using Eq. (4.4) in [22] and (6.9),

$$|f(Z_i) - f_{b_n}(Z_i)| \leqslant \left|f(Z_i) - \widehat{f}_{b_n}(Z_i)\right| + \left|\widehat{f}_{b_n}(Z_i) - f_{b_n}(Z_i)\right| \leqslant b_n + 2C\rho_n \leqslant L_4 b_n$$

almost surely, where L_4 is a positive constant, the last inequality being due to the fact that $\lim_{n\to\infty}\left(b_n^{-1}\rho_n\right) = 0$. Thus, almost surely,

$$\begin{aligned} F_{n,4} &\leqslant L_4 n^{\alpha} M_n \left(\frac{1}{n}\sum_{i=1}^{n} |\mathscr{T}_j(Z_i)|\, \mathbb{I}_{\{f(Z_i)\leqslant b_n\}}\right) \\ &= L_4 n^{\alpha-1/4} M_n n^{-1/4}\left(s_n\zeta_n + \kappa_n\right), \end{aligned} \tag{6.17}$$

where

$$s_n^2 = n\ \mathrm{Var}\left(|\mathscr{T}_j(Z)|\, \mathbb{I}_{\{f(Z)\leqslant b_n\}}\right),\ \ \kappa_n = \sqrt{n}\mathbb{E}\left(|\mathscr{T}_j(Z)|\, \mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)$$

and

$$\zeta_n = s_n^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} |\mathscr{T}_j(Z_i)|\, \mathbb{I}_{\{f(Z_i)\leqslant b_n\}} - \mathbb{E}\left(|\mathscr{T}_j(Z)|\, \mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)\right).$$

Since

$$n^{-1/2}s_n^2 \leqslant \sqrt{n}\mathbb{E}\left(\mathscr{T}_j^2(Z)\mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)$$

and

$$n^{-1/4}\kappa_n \leqslant \left(\sqrt{n}\mathbb{E}\left(\mathscr{T}_j^2(Z)\mathbb{I}_{\{f(Z)\leqslant b_n\}}\right)\right)^{1/2},$$

we deduce from Assumption 4.6 that $n^{-1/4}s_n \to 0$ and $n^{-1/4}\kappa_n \to 0$ as $n \to \infty$. The central limit theorem for triangular arrays (see Theorem 2.2 in [1]) ensures that ζ_n converges in distribution, as $n \to \infty$, to the standard normal distribution. Then, from (6.17) and the fact that $n^{\alpha-1/4}M_n \to 0$ as $n \to \infty$, we deduce that $F_{n,4} = o_p(1)$. Finally, since

$$\mathbb{E}\left(U_k^2\mathscr{T}_j^2(Z)\right) \leqslant \mathbb{E}\left(U_k^4\right)^{1/2}\mathbb{E}\left(\mathscr{T}_j^4(Z)\right)^{1/2} \leqslant \mathbb{E}\left(U_k^4\right)^{1/2}\mathbb{E}\left(U_j^4\right)^{1/2} < \infty,$$

we can apply the central limit theorem; we then get the convergence in distribution of the sequence $\dfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(U_{ik}\mathscr{T}_j(Z_i) - \mathbb{E}\left(U_k\mathscr{T}_j(Z)\right)\right)$ to a normal distribution $\mathcal{N}(0,\, \vartheta_{kj}^2)$, where $\vartheta_{kj}^2 = Var\left(U_k\mathscr{T}_j(Z)\right)$. Then since $\alpha < 1/2$, we deduce from (6.16) that $F_{n,5} = o_p(1)$. The proof is complete.

6.3. Proof of Theorem 4.2 The proof is in every way similar to that of Theorem 3.1 in [14] provided that the convergences in probability to 0 described below in (6.18) and (6.19) are verified. It is therefore sufficient to establish these convergence properties. Let $r \in \mathbb{N}^*$ and $(m_1, \ldots, m_r) \in (\mathbb{N}^*)^r$ such that $\sum_{\ell=1}^r m_\ell = p$ and $\xi_{I[\tau(1)]} = \cdots = \xi_{I[\tau(m_1)]} > \xi_{I[\tau(m_1+1)]} = \cdots = \xi_{I[\tau(m_1+m_2)]} > \cdots > \xi_{I[\tau(m_1+m_2+\cdots+m_{r-1}+1)]} = \cdots = \xi_{I[\tau(m_1+m_2+\cdots+m_r)]}$. Then, putting

$$E = \{\ell \in \mathbb{N}^* \,/\, 1 \leqslant \ell \leq r,\ m_\ell \geqslant 2\}$$

and $F_\ell := \left\{\left(\sum_{k=0}^{\ell-1} m_k\right) + 1, \ldots, \left(\sum_{k=0}^{\ell} m_k\right) - 1\right\}$ with $m_0 = 0$, we just have to prove that for $\ell \in E$, $j \in F_\ell$ and J such that $I_1 \subset J \subset I$,

$$n^\alpha \left(\widehat{\xi}^{(n)}_{I[\tau(j)]} - \widehat{\xi}^{(n)}_{I[\tau(j+1)]}\right) \xrightarrow{P} 0 \quad \text{as } n \to \infty, \tag{6.18}$$

$$n^\gamma \widehat{\xi}^{(n)}_J \xrightarrow{P} 0 \quad \text{as } n \to \infty, \tag{6.19}$$

where $\xrightarrow{P}$ denotes convergence in probability. Using the notations of Section 6.2, we consider the covariance matrices

$$\widehat{V}^{(n)} = \frac{1}{n}\sum_{i=1}^n \left(U^{(i)} - \widehat{\mathscr{T}}_{b_n}(Z_i)\right)\left(U^{(i)} - \widehat{\mathscr{T}}_{b_n}(Z_i)\right)^\top,$$

$$V = \mathbb{E}\left(\left(U - \mathscr{T}(Z)\right)\left(U - \mathscr{T}(Z)\right)^\top\right),$$

and we put

$$\widehat{H}^{(n)} = \sqrt{n}\left(\widehat{V}^{(n)} - V\right).$$

Proof of (6.18): Note that if $j \in F_\ell$ then $\xi_{I[\tau(j)]} = \xi_{I[\tau(j+1)]} =: a_\ell$. It is shown in the proof of Lemma 3 of [11] that

$$\left|n^\alpha \left(\widehat{\xi}^{(n)}_{I[\tau(j)]} - \widehat{\xi}^{(n)}_{I[\tau(j+1)]}\right)\right| \ \leqslant \ n^{\alpha-\frac{1}{2}}\widehat{\Delta}^{(n)}_j \|\widehat{H}^{(n)}\|_2, \tag{6.20}$$

where $\|\cdot\|_2$ denotes the matrix norm defined as $\|A\|_2 = \sqrt{\mathrm{tr}\,(AA^\top)} = \left(\sum_{k,j} A_{kj}^2\right)^{1/2}$,

$$\widehat{\Delta}^{(n)}_j = \begin{cases} \|\widehat{\Psi}^{(n)}_{I[\tau(j)]} - \widehat{\Psi}^{(n)}_{I[\tau(j+1)]}\|_\infty & \text{if } a_\ell = 0 \\[2ex] \dfrac{n^{-1/2}\left(\|\widehat{\Psi}^{(n)}_{I[\tau(j)]}\|^2_\infty + \|\widehat{\Psi}^{(n)}_{I[\tau(j+1)]}\|^2_\infty\right)\|\widehat{H}^{(n)}\|_2 + 2a_\ell\left(\|\widehat{\Psi}^{(n)}_{I[\tau(j)]}\|_\infty + \|\widehat{\Psi}^{(n)}_{I[\tau(j+1)]}\|_\infty\right)}{\|n^{-\frac{1}{2}}\widehat{\Psi}^{(n)}_{I[\tau(j)]}(\widehat{H}^{(n)}) + \delta_{I[\tau(j)]}\| + \|n^{-\frac{1}{2}}\widehat{\Psi}^{(n)}_{I[\tau(j+1)]}(\widehat{H}^{(n)}) + \delta_{I[\tau(j+1)]}\|} & \text{if } a_\ell \neq 0 \end{cases},$$

$(\widehat{\Psi}_J^{(n)})_{n\in\mathbb{N}^*}$ is a sequence of random operators, from the space of $(p+1)\times(p+1)$ matrices to $\mathbb{R}^p$, that converges almost surely uniformly to an appropriate linear map Ψ_J, and $\delta_J = V_{12} - V_1\Pi_J V_{12}$. In order to get (6.18) it is enough to prove that

$$n^{\alpha-\frac{1}{2}}\|\widehat{H}^{(n)}\|_2 \xrightarrow{P} 0 \quad \text{as } n\to\infty. \tag{6.21}$$

Indeed, if (6.21) holds then, since

$$\|n^{-\frac{1}{2}}\widehat{\Psi}^{(n)}_{I[\tau(j)]}(\widehat{H}^{(n)})\| \leqslant n^{-\alpha}\|\widehat{\Psi}^{(n)}_{I[\tau(j)]}\|_\infty \left(n^{\alpha-\frac{1}{2}}\|\widehat{H}^{(n)}\|_2\right),$$

we deduce that $\widehat{\Delta}_j^{(n)} \xrightarrow{P} \Delta_j$, as $n\to\infty$, where

$$\Delta_j = \begin{cases} \|\Psi_{I[\tau(j)]} - \Psi_{I[\tau(j+1)]}\|_\infty & \text{if } a_\ell = 0 \\ \frac{2a_\ell\left(\|\Psi_{I[\tau(j)]}\|_\infty + \|\Psi_{I[\tau(j+1)]}\|_\infty\right)}{\|\delta_{I[\tau(j)]}\| + \|\delta_{I[\tau(j+1)]}\|} & \text{if } a_\ell \neq 0 \end{cases}.$$

Then, (6.20) allows to conclude that (6.18) holds. It remains to prove (6.21). To this end, we first note that $\widehat{H}^{(n)} = \widehat{H}_1^{(n)} + \widehat{H}_2^{(n)} + (\widehat{H}_2^{(n)})^\top + \widehat{H}_3^{(n)}$, where

$$\widehat{H}_1^{(n)} = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n U^{(i)}(U^{(i)})^\top - \mathbb{E}\left(UU^\top\right)\right)$$

$$\widehat{H}_2^{(n)} = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n U^{(i)}\widehat{\mathscr{T}_{b_n}}(Z_i)^\top - \mathbb{E}\left(U\mathscr{T}(Z)^\top\right)\right)$$

$$\widehat{H}_3^{(n)} = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \widehat{\mathscr{T}_{b_n}}(Z_i)\widehat{\mathscr{T}_{b_n}}(Z_i)^\top - \mathbb{E}\left(\mathscr{T}(Z)\mathscr{T}(Z)^\top\right)\right),$$

and, consequently,

$$n^{\alpha-\frac{1}{2}}\|\widehat{H}^{(n)}\|_2 \leqslant n^{\alpha-\frac{1}{2}}\|\widehat{H}_1^{(n)}\|_2 + 2n^{\alpha-\frac{1}{2}}\|\widehat{H}_2^{(n)}\|_2 + n^{\alpha-\frac{1}{2}}\|\widehat{H}_3^{(n)}\|_2. \tag{6.22}$$

By the central limit theorem, we get the convergence in distribution, as $n\to\infty$, of $\widehat{H}_1^{(n)}$ to a random variable having a normal distribution in the space of $(p+1)\times(p+1)$ matrices. Thus $n^{\alpha-\frac{1}{2}}\|\widehat{H}_1^{(n)}\|_2 \xrightarrow{P} 0$ as $n\to\infty$. Furthermore, Theorem 2.1 of [22] ensures that $\widehat{H}_3^{(n)}$ converges in distribution, as $n\to\infty$, to a random variable having a normal distribution. We then deduce that $n^{\alpha-\frac{1}{2}}\|\widehat{H}_3^{(n)}\|_2 \xrightarrow{P} 0$ as $n\to\infty$. Furthermore,

$$\begin{aligned} n^{\alpha-\frac{1}{2}}\|\widehat{H}_2^{(n)}\|_2 &= n^{\alpha-1/2}\left(\sum_{k=1}^{p+1}\sum_{j=1}^{p+1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(U_{ik}\widehat{\mathscr{T}_{b_n,j}}(Z_i) - \mathbb{E}\left(U_k\mathscr{T}_j(Z)\right)\right)\right)^2\right)^{1/2} \\ &\leqslant (p+1)\sum_{k=1}^{p+1}\sum_{j=1}^{p+1} n^{\alpha-1/2}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(U_{ik}\widehat{\mathscr{T}_{b_n,j}}(Z_i) - \mathbb{E}\left(U_k\mathscr{T}_j(Z)\right)\right)\right|, \end{aligned}$$

and Lemma 6.1 allows to conclude that $n^{\alpha-\frac{1}{2}}\|\widehat{H}_2^{(n)}\|_2 \xrightarrow{P} 0$ as $n \to \infty$. Finally, using (6.22) we get (6.21).

Proof of (6.19): Since $I_1 \subset J$, we have $\sqrt{n}\widehat{\xi}_J^{(n)} = \|\widehat{\Psi}_J^{(n)}(\widehat{H}^{(n)})\|$ (see [11]). Letting γ_0 be a real number such that $\gamma < \gamma_0 < 1/4$, we have $n^{\gamma_0}\widehat{\xi}_J^{(n)} \leqslant \|\widehat{\Psi}_J^{(n)}\|_\infty \left(n^{\gamma_0-1/2}\|\widehat{H}^{(n)}\|_2\right)$; this inequality and the fact that, as in (6.21), $n^{\gamma_0-1/2}\|\widehat{H}^{(n)}\|_2 \xrightarrow{P} 0$ as $n \to \infty$ allow to deduce that $n^{\gamma_0}\widehat{\xi}_J^{(n)} \xrightarrow{P} 0$ as $n \to \infty$. Then, using the equality $n^{\gamma}\widehat{\xi}_J^{(n)} = n^{\gamma-\gamma_0}\left(n^{\gamma_0}\widehat{\xi}_J^{(n)}\right)$ we can conclude that (6.19) holds.

References

1. S.C. Bagui, D.K. Bhaumik, and K.L. Mehra, A few counter examples useful in teaching central limit theorems, *Amer. Statist.* **67**(2013), 49–56.
2. F. Bunea, Consistent covariate selection and post model selection inference in semiparametric regression, *Ann. Statist.* **32**(2004), 898–927.
3. F. Bunea and M.H. Wegkamp, Two-stage model selection procedures in partially linear regression, *Canad. J. Statist.* **32**(2004), 105–118.
4. J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.* **96**(2001), 1348–1360.
5. J. Fan and R. Li, New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, *J. Amer. Statist. Assoc.* **99**(2004), 710–723.
6. E. Giné and A. Guillou, On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals, *Ann. Inst. Henri Poincaré Probab. Stat.* **37**(2001), 503–522.
7. F. Li, L. Lin, and Y. Su, Variable selection and parameter estimation for partially linear models via Dantzig selector, *Metrika* **76**(2013), 225–238.
8. T. Li, Q. Yin, and J. Peng, Variable selection and parameter estimation for partially linear models via Dantzig selector, *Comm. Statist. Simulation Comput.* **90**(2020), 2681–2704.
9. H. Liang and R. Li, Variable selection for partially linear models with measurement errors, *J. Amer. Statist. Assoc.* **104**(2009), 234–248.
10. J. Liu, L. Lou, and R. Li, Variable selection for partially linear models via partial correlation, *J. Multivariate Anal.* **167**(2018), 418–434.
11. A. Mbina Mbina, G.M. Nkiet, and A. N'Guessan, Variable selection in multivariate linear regression with random predictors, *South African Statist. J.* **57**(2023), 27–44.
12. X. Ni, H.H. Zhang, and D. Zhang, Automatic model selection for partially linear models, *J. Multivariate Anal.* **100**(2009), 2100–2111.
13. G.M. Nkiet, Sélection des variables dans un modèle structurel de régression linéaire. *C. R. Acad. Sci. Paris Sér. I Math.* **333**(2001), 1105–1110.
14. G.M. Nkiet, Direct variable selection for discrimination among several groups. *J. Multivariate Anal.* **105**(2012), 151–163.
15. E.D.D. Nkou, Strong consistency of kernel method for sliced average variance estimation, *Comm. Statist. Theory Methods* **52**(2023), 7586–7600.
16. E.D.D. Nkou and G.M. Nkiet, Strong consistency of kernel estimator in a semiparametric regression model. *Statistics* **53**(2019), 1289–1305.
17. B.L.S. Prakasa Rao, *Nonparametric functional estimation*, Academic Press, Orlando, 1983.
18. M. Talagrand, Sharper bounds for Gaussian and empirical processes, *Ann. Probab.* **22**(1994), 28–76.
19. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. Roy. Statist. Soc. Ser. B* **58**(1996), 267–288.
20. H. Xie and J. Huang, SCAD-penalized regression in high-dimensional partially linear models, *Ann. Statist.* **37**(2009), 673–696.

21. Y. Yang, L. Xue, and W. Cheng, Variable selection for partially linear models with randomly censored data, *Comm. Statist. Simulation Comput.* **39**(2010), 1577–1589.
22. L.-X. Zhu and K.-T. Fang, Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.* **24**(1996), 1053–1068.

Laboratoire de Probabilités, Statistique et Informatique, Unité de Recherche en Mathématiques et Informatique, Université des Sciences et Techniques de Masuku, BP 813 Franceville, GABON
e-mail: ebendependa@gmail.com, emmanueldedieunkou@gmail.com
stephane.bouka@univ-masuku.org, guymartial.nkiet@univ-masuku.org