

# Efficient estimation of conditional covariance matrices for dimension reduction

JEAN-MICHEL LOUBES

*Institut de Mathématiques de Toulouse*

loubes@math.univ-toulouse.fr

CLÉMENT MARTEAU

*Institut de Mathématiques de Toulouse*

clement.marteau@insa-toulouse.fr

MICHAEL SOLÍS CHACÓN

*Institut de Mathématiques de Toulouse*

msolisch@math.univ-toulouse.fr

SEBASTIEN DA VEIGA

*Institut Français du Pétrole*

sebastien.da-veiga@ifp.fr

## Abstract

We consider the problem of estimating a conditional covariance matrix in an inverse regression setting. We show that this estimation can be achieved by estimating a quadratic functional extending the results of Da Veiga & Gamboa (2008). We prove that this method provides a new efficient estimator whose asymptotic properties are studied.

## 1 Introduction

Consider the nonparametric regression

$$Y = \varphi(X) + \epsilon,$$

where  $X \in \mathbb{R}^p$ ,  $Y \in \mathbb{R}$  and  $\mathbb{E}[\epsilon] = 0$ . The main difficulty with any regression method is that, as the dimension of  $X$  becomes larger, the number of observations needed for a good estimator increases exponentially. This phenomena is usually called the *curse of dimensionality*. All the “classical” methods could break down, as the dimension  $p$  increases, unless we have at hand a very huge sample.

For this reason, there have been along the past decades a very large number of methods to cope with this issue. Their aim is to reduce the dimensionality of the problem, using just to name a few, the generalized linear model in Brillinger (1983), the additive models in Hastie & Tibshirani (1990), sparsity constraint models as Li (2007) and references therein.

Alternatively, Li (1991a) proposed the procedure of Sliced Inverse Regression (SIR) considering the following semiparametric model,

$$Y = \phi(v_1^\top X, \dots, v_K^\top X, \epsilon)$$

where the  $v$ 's are unknown vectors in  $\mathbb{R}^p$ ,  $\epsilon$  is independent of  $X$  and  $\phi$  is an arbitrary function in  $\mathbb{R}^{K+1}$ . This model can gather all the relevant information about the variable  $Y$ , with only the projection of  $X$  onto the  $K \ll p$  dimensional subspace  $(v_1^\top X, \dots, v_K^\top X)$ . In the case when  $K$  is small, it is possible to reduce the dimension by estimating the  $v$ 's efficiently. This method is also used to search nonlinear structures in data and to estimate the projection directions  $v$ 's. For a review on SIR methods, we refer to Li (1991a,b); Duan & Li (1991); Hardle & Tsybakov (1991) and references therein. The  $v$ 's define the effective dimension reduction (e.d.r) direction and the eigenvectors of  $\mathbb{E}[\text{Cov}(X|Y)]$  are the e.d.r directions. Many estimators have been proposed in order to study the e.d.r directions in many different cases. For example, Zhu & Fang (1996) and Ferré & Yao (2005, 2003) use kernel estimators, Hsing (1999) combines nearest neighbor and SIR, Bura & Cook (2001) assume that  $\mathbb{E}[X|Y]$  has some parametric form, Setodji & Cook (2004) use k-means and Cook & Ni (2005) transform SIR to least square form.

In this paper, we propose an alternate estimation of the matrix

$$\text{Cov}(\mathbb{E}[X|Y]) = \mathbb{E}[\mathbb{E}[X|Y]\mathbb{E}[X|Y]^\top] - \mathbb{E}[X]\mathbb{E}[X]^\top,$$

using ideas developed by Da Veiga & Gamboa (2008), inspired by the prior work of Laurent (1996). More precisely since  $\mathbb{E}[X]\mathbb{E}[X]^\top$  can be easily estimated with many usual methods, we will focus on finding an estimator of  $\mathbb{E}[\mathbb{E}[X|Y]\mathbb{E}[X|Y]^\top]$ . For this we will show that this estimation implies an estimation of a quadratic functional rather than plugging non parametric estimate into this form as commonly used. This method has the advantage of getting an efficient estimator in a semi-parametric framework.

This paper is organized as follows. Section 2 is intended to motivate our investigation of  $\text{Cov}(\mathbb{E}[X|Y])$  using a Taylor approximation. In Section 3.1 we set up notation and hypothesis. Section 3.2 is devoted to demonstrate that each coordinate of  $\text{Cov}(\mathbb{E}[X|Y])$  converge efficiently. Also we find the normality asymptotic for the whole matrix. An asymptotic bound of the variance for the quadratic part for the Taylor's expansion of  $\text{Cov}(\mathbb{E}[X|Y])$  is found in Section 4. All technical Lemmas and their proofs are postponed to Sections 6 and 5 respectively.

## 2 Methodology

Our aim is to estimate  $\text{Cov}(\mathbb{E}[X|Y])$  efficiently when observing  $X \in \mathbb{R}^p$ , for  $p \geq 1$ , and  $Y \in \mathbb{R}$ . For this, write the matrix

$$\text{Cov}(\mathbb{E}[X|Y]) = \mathbb{E}[\mathbb{E}[X|Y]\mathbb{E}[X|Y]^\top] - \mathbb{E}[X]\mathbb{E}[X]^\top,$$

where  $A^\top$  means the transpose of  $A$ . If  $\mathbb{E}[X]$  can be easily estimated by classical methods, the remainder term

$$\mathbb{E}[\mathbb{E}[X|Y]\mathbb{E}[X|Y]^\top] = (T_{ij}^*)_{i,j} \quad i, j = 1, \dots, p;$$

is a non linear term whose estimation is the main topic of this paper. Each term of this matrix can be written as

$$T_{ij}^* = \int \left( \frac{\int x_i f(x_i, x_j, y) dx_i dx_j}{\int f(x_i, x_j, y) dx_i dx_j} \right) \left( \frac{\int x_j f(x_i, x_j, y) dx_i dx_j}{\int f(x_i, x_j, y) dx_i dx_j} \right) f(x_i, x_j, y) dx_i dx_j dy, \quad (1)$$

where  $f(x_i, x_j, y)$  for  $i$  and  $j$  fixed, is the joint density of  $(X_i, X_j, Y)$   $i, j = 1, \dots, p$ .

Hence, we focus on the efficient estimation of the corresponding non linear functional for  $f \in \mathbb{L}(dx_i, dx_j, dy)$

$$f \mapsto T_{ij}(f) = \int \left( \frac{\int x_i f(x_i, x_j, y) dx_i dx_j}{\int f(x_i, x_j, y) dx_i dx_j} \right) \left( \frac{\int x_j f(x_i, x_j, y) dx_i dx_j}{\int f(x_i, x_j, y) dx_i dx_j} \right) f(x_i, x_j, y) dx_i dx_j dy. \quad (2)$$

In the case  $i = j$ , this estimation has been considered in Da Veiga & Gamboa (2008); Laurent (1996). Here we extend their methodology to this case. Assume we have at hand an i.i.d sample  $(X_i^{(k)}, X_j^{(k)}, Y^{(k)})$ ,  $k = 1, \dots, n$  such that it is possible to build a preliminary estimator  $\hat{f}$  of  $f$  with a subsample of size  $n_1 < n$ . Now, the main idea is to make a Taylor's expansion of  $T_{ij}(f)$  in a neighborhood of  $\hat{f}$  which will play the role of a suitable approximation of  $f$ . More precisely, define an auxiliar function  $F : [0, 1] \rightarrow \mathbb{R}$ ;

$$F(u) = T_{ij}(uf + (1-u)\hat{f})$$

with  $u \in [0, 1]$ . The Taylor's expansion of  $F$  between 0 and 1 up to the third order is

$$F(1) = F(0) + F'(0) + \frac{1}{2}F''(0) + \frac{1}{6}F'''(\xi)(1-\xi)^3 \quad (3)$$

for some  $\xi \in [0, 1]$ . Moreover, we have

$$F(1) = T_{ij}(f)$$

$$F(0) = T_{ij}(\hat{f}) = \int \left( \frac{\int x_i \hat{f}(x_i, x_j, y) dx_i dx_j}{\int \hat{f}(x_i, x_j, y) dx_i dx_j} \right) \left( \frac{\int x_j \hat{f}(x_i, x_j, y) dx_i dx_j}{\int \hat{f}(x_i, x_j, y) dx_i dx_j} \right) \hat{f}(x_i, x_j, y) dx_i dx_j dy.$$

To simplify the notations, let

$$m_i(f_u, y) = \frac{\int x_i f_u(x_i, x_j, y) dx_i dx_j}{\int f_u(x_i, x_j, y) dx_i dx_j}$$

$$m_i(f_0, y) = m_i(\hat{f}, y) = \frac{\int x_i \hat{f}(x_i, x_j, y) dx_i dx_j}{\int \hat{f}(x_i, x_j, y) dx_i dx_j},$$

where  $f_u = uf + (1 - u)\hat{f}$ ,  $\forall u \in [0, 1]$ . Then, we can rewrite  $F(u)$  as

$$F(u) = \int m_i(f_u, y)m_j(f_u, y)f_u(x_i, x_j, y)dx_idx_jdy.$$

The Taylor's expansion of  $T_{ij}(f)$  is given in the next Proposition.

**Proposition 1** (Linearization of the operator  $T$ ). *For the functional  $T_{ij}(f)$  defined in (2), the following decomposition holds*

$$T_{ij}(f) = \int H_1(\hat{f}, x_i, x_j, y)f(x_i, x_j, y)dx_idx_jdy + \int H_2(\hat{f}, x_{i1}, x_{j2}, y)f(x_{i1}, x_{j1}, y)f(x_{i2}, x_{j2}, y)dx_{i1}dx_{j1}dx_{i2}dx_{j2}dy + \Gamma_n \quad (4)$$

where

$$H_1(\hat{f}, x_i, x_j, y) = x_im_j(\hat{f}, y) + x_jm_i(\hat{f}, y) - m_i(\hat{f}, y)m_j(\hat{f}, y) \quad (5)$$

$$H_2(\hat{f}, x_{i1}, x_{j2}, y) = \frac{1}{\int \hat{f}(x_i, x_j, y)dx_idx_j} (x_{i1} - m_i(\hat{f}, y))(x_{j2} - m_j(\hat{f}, y)) \quad (6)$$

$$\Gamma_n = \frac{1}{6}F'''(\xi)(1 - \xi)^3, \quad (7)$$

for some  $\xi \in ]0, 1[$ .

This decomposition has the main advantage of separating the terms to be estimated into a linear functional of  $f$ , which can be easily estimated and a second part which is a quadratic functional of  $f$ . In this case, Section 4 will be dedicated to estimate this kind of functionals and specifically to control its variance. This will enable to provide an efficient estimator of  $T_{ij}(f)$  using the decomposition of Proposition 1.

### 3 Main Results

In this section we build a procedure to estimate  $T_{ij}(f)$  efficiently. Since we used  $n_1 < n$  to build a preliminary approximation  $\hat{f}$ , we will use a sample of size  $n_2 = n - n_1$  to estimate (5) and (6). Since (5) is a linear functional of the density  $f$ , it can be estimated by its empirical counterpart

$$\frac{1}{n_2} \sum_{k=1}^{n_2} H_1(\hat{f}, X_i^{(k)}, X_j^{(k)}, Y^{(k)}). \quad (8)$$

Since (6) is a nonlinear functional of  $f$ , the estimation is harder. Its estimation will be a direct consequence of the technical results presented in Section 4, where we build an estimator for the general functional

$$\theta(f) = \int \eta(x_{i1}, x_{j2}, y)f(x_{i1}, x_{j1}, y)f(x_{i2}, x_{j2}, y)dx_{i1}dx_{j1}dx_{i2}dx_{j2}dy$$

where  $\eta : \mathbb{R}^3 \rightarrow \mathbb{R}$  is a bounded function. The estimator  $\hat{\theta}_n$  of  $\theta(f)$  is an extension of the method developed in Da Veiga & Gamboa (2008).

### 3.1 Hypothesis and Assumptions

The following notations will be used throughout the paper. Let  $d_s$  and  $b_s$  for  $s = 1, 2, 3$  be real numbers where  $d_s < b_s$ . Let, for  $i$  and  $j$  fixed,  $\mathbb{L}^2(dx_i dx_j dy)$  be the squared integrable functions in the cube  $[d_1, b_1] \times [d_2, b_2] \times [d_3, b_3]$ . Moreover, let  $(p_l(x_i, x_j, y))_{l \in D}$  be an orthonormal basis of  $\mathbb{L}^2(dx_i dx_j dy)$ , where  $D$  is a countable set. Let  $a_l = \int p_l f$  denote the scalar product of  $f$  with  $p_l$ .

Furthermore, denote by  $\mathbb{L}^2(dx_i dx_j)$  (resp.  $\mathbb{L}^2(dy)$ ) the set of squared integrable functions in  $[d_1, b_1] \times [d_2, b_2]$  (resp.  $[d_3, b_3]$ ). If  $(\alpha_{l_\alpha}(x_i, x_j))_{l_\alpha \in D_1}$  (resp.  $(\beta_{l_\beta}(y))_{l_\beta \in D_2}$ ) is an orthonormal basis of  $\mathbb{L}^2(dx_i dx_j)$  (resp.  $\mathbb{L}^2(dy)$ ) then  $p_l(x_i, x_j, y) = \alpha_{l_\alpha}(x_i, x_j) \beta_{l_\beta}(y)$  with  $l = (l_\alpha, l_\beta) \in D_1 \times D_2$ .

We also use the following subset of  $\mathbb{L}^2(dx_i dx_j dy)$

$$\mathcal{E} = \left\{ \sum_{l \in D} e_l p_l : (e_l)_{l \in D} \text{ is such that } \sum_{l \in D} \left| \frac{e_l}{c_l} \right|^2 < 1 \right\}$$

where  $(c_l)_{l \in D}$  is a given fixed sequence.

Moreover assume that  $(X_i, X_j, Y)$  have a bounded joint density  $f$  on  $[d_1, b_1] \times [d_2, b_2] \times [d_3, b_3]$  which lies in the ellipsoid  $\mathcal{E}$ .

In what follows,  $X_n \xrightarrow{\mathcal{D}} X$  (resp.  $X_n \xrightarrow{\mathcal{P}} X$ ) denotes the convergence in distribution or weak convergence (resp. convergence in probability) of  $X_n$  to  $X$ . Additionally, the support of  $f$  will be denoted by  $\text{supp } f$ .

Let  $(M_n)_{n \geq 1}$  denote a sequence of subsets  $D$ . For each  $n$  there exists  $M_n$  such that  $M_n \subset D$ . Let us denote by  $|M_n|$  the cardinal of  $M_n$ .

We shall make three main assumptions:

**Assumption 1.** For all  $n \geq 1$  there is a subset  $M_n \subset D$  such that  $(\sup_{l \notin M_n} |c_l|^2)^2 \approx |M_n|/n^2$  ( $A_n \approx B$  means  $\lambda_1 \leq A_n/B \leq \lambda_2$  for some positives constants  $\lambda_1$  and  $\lambda_2$ ). Moreover,  $\forall f \in \mathbb{L}^2(dx dy dz)$ ,  $\int (S_{M_n} f - f)^2 dx dy dz \rightarrow 0$  when  $n \rightarrow \infty$ , where  $S_{M_n} f = \sum_{l \in M_n} a_l p_l$

**Assumption 2.**  $\text{supp } f \subset [d_1, b_1] \times [d_2, b_2] \times [d_3, b_3]$  and  $\forall (x, y, z) \in \text{supp } f$ ,  $0 < \alpha \leq f(x, y, z) \leq \beta$  with  $\alpha, \beta \in \mathbb{R}$ .

**Assumption 3.** It is possible to find an estimator  $\hat{f}$  of  $f$  built with  $n_1 \approx n/\log(n)$  observations, such that for  $\epsilon > 0$ ,

$$\forall (x, y, z) \in \text{supp } f, 0 < \alpha - \epsilon \leq \hat{f}(x, y, z) \leq \beta + \epsilon$$

and,

$$\forall 2 \leq q \leq +\infty, \forall l \in \mathbb{N}^*, \mathbb{E}_f \| \hat{f} - f \|_q^l \leq C(q, l) n_1^{-l\lambda}$$

for some  $\lambda > 1/6$  and some constant  $C(q, l)$  not depending on  $f$  belonging to the ellipsoid  $\mathcal{E}$ .

Assumption 1 is necessary to bound the bias and variance of  $\hat{\theta}_n$ . Assumption 2 and 3 allow to establish that the remainder term in the Taylor expansion is negligible, i.e  $\Gamma_n = O(1/n)$ . Assumption 3 depends on the regularity of the density function. For instance for  $x \in \mathbb{R}^p$ ,  $s > 0$  and  $L > 0$ , consider the class

$\mathcal{H}_q(s, L)$  of Nikol'skii of functions  $f \in \mathbb{L}^q(dx)$  with partials derivatives up to order  $r = \lfloor s \rfloor$  inclusive, and for each of these derivatives  $g^{(r)}$

$$\|f^{(r)}(\cdot + h) - f^{(r)}(\cdot)\|_q \leq L |h|^{s-r} \quad \forall h \in \mathbb{R}.$$

Then, Assumption 3 is satisfied for  $f \in \mathcal{H}_q(s, L)$  with  $s > \frac{p}{4}$ .

### 3.2 Efficient Estimation of $T_{ij}(f)$

As seen in Section 2,  $T_{ij}(f)$  can be decomposed as (4). Hence, using (8) and (14) we consider the following estimate

$$\begin{aligned} \widehat{T}_{ij}^{(n)} &= \frac{1}{n_2} \sum_{k=1}^{n_2} H_1(\hat{f}, X_i^{(k)}, X_j^{(k)}, Y^{(k)}) \\ &+ \frac{1}{n_2(n_2 - 1)} \sum_{l \in M} \sum_{k \neq k'=1}^{n_2} p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) \int p_l(x_i, x_j, Y^{(k')}) H_3(\hat{f}, x_i, x_j, X_i^{(k')}, X_j^{(k')}, Y^{(k')}) dx_i dx_j \\ &- \frac{1}{n_2(n_2 - 1)} \sum_{l, l' \in M} \sum_{k \neq k'=1}^{n_2} p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) p_{l'}(X_i^{(k')}, X_j^{(k')}, Y^{(k')}) \\ &\quad \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) H_2(\hat{f}, x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \end{aligned}$$

where  $H_3(f, x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) = H_2(f, x_{i1}, x_{j2}, y) + H_2(f, x_{i2}, x_{j1}, y)$  and  $n_2 = n - n_1$ . The remainder  $\Gamma_n$  does not appear because we will prove that it is negligible when compared to the other error terms.

The asymptotic behavior of  $\widehat{T}_{ij}^{(n)}$  for  $i$  and  $j$  fixed is given in the next Theorem.

**Theorem 1.** *Let Assumptions 1-3 hold and  $|M_n|/n \rightarrow 0$  when  $n \rightarrow \infty$ . Then:*

$$\sqrt{n}(\widehat{T}_{ij}^{(n)} - T_{ij}(f)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, C_{ij}(f)), \quad (9)$$

and

$$\lim_{n \rightarrow \infty} n \mathbb{E}[\widehat{T}_{ij}^{(n)} - T_{ij}(f)]^2 = C_{ij}(f), \quad (10)$$

where

$$C_{ij}(f) = \text{Var}(H_1(f, X_i, X_j, Y))$$

Note that, in Theorem 1, it appears that the asymptotic variance of  $T_{ij}(f)$  depends only on  $H_1(f, X_i, X_j, Y)$ . Hence the asymptotic variance of  $\widehat{T}_{ij}^{(n)}$  is explained only by the linear part of (4). This will entail that the estimator is naturally efficient as proved in the following.

Indeed, the semi-parametric Cramér-Rao bound is given in the next theorem.

**Theorem 2** (Semi-parametric Cramér-Rao bound.). *Consider the estimation of*

$$T_{ij}(f) = \int \left( \frac{\int x_i f(x_i, x_j, y) dx_i dx_j}{\int f(x_i, x_j, y) dx_i dx_j} \right) \left( \frac{\int x_j f(x_i, x_j, y) dx_i dx_j}{\int f(x_i, x_j, y) dx_i dx_j} \right) f(x_i, x_j, y) dx_i dx_j dy$$

for a random vector  $(X_i, X_j, Y)$  with joint density  $f \in \mathcal{E}$ . Let  $f_0 \in \mathcal{E}$  be a density verifying the assumptions of Theorem 1. Then, for all estimator  $\widehat{T}_{ij}^{(n)}$  of  $T_{ij}(f)$  and every family  $\{\mathcal{V}_r(f_0)\}_{r>0}$  of neighborhoods of  $f_0$  we have

$$\inf_{\{\mathcal{V}_r(f_0)\}_{r>0}} \liminf_{n \rightarrow \infty} \sup_{f \in \mathcal{V}_r(f_0)} n\mathbb{E}[\widehat{T}_{ij}^{(n)} - T_{ij}(f_0)]^2 \geq C_{ij}(f_0)$$

where  $\mathcal{V}_r(f_0) = \{f : \|f - f_0\|_2 < r\}$  for  $r > 0$ .

Consequently, the estimator  $\widehat{T}_{ij}^{(n)}$  is efficient.

In the case of our estimate, its variance is  $C_{ij}(f)$ , which proves its asymptotically efficiency.

Remark that Theorem 1 proves asymptotic normality entry by entry of the matrix  $\mathbf{T}(f) = (T_{ij}(f))_{p \times p}$ . To extend the result for the whole matrix it is necessary to introduce the half-vectorization operator  $\text{vech}$ . This operator, stacks only the columns from the principal diagonal of a square matrix downwards in a column vector, that is, for an  $p \times p$  matrix  $\mathbf{A} = (a_{ij})$ ,

$$\text{vech}(\mathbf{A}) = [a_{11}, \dots, a_{p1}, a_{22}, \dots, a_{p2}, \dots, a_{33}, \dots, a_{pp}]^\top.$$

Let define the estimator matrix  $\widehat{\mathbf{T}}^{(n)} = (\widehat{T}_{ij}^{(n)})$  and  $\mathbf{H}_1(f)$  denote the matrix with entries  $(H_1(f, x_i, x_j, y))_{i,j}$ . Now we are able to state the following

**Corollary 1.** *Let Assumptions 1-3 hold and  $|M_n|/n \rightarrow 0$  when  $n \rightarrow \infty$ . Then  $\widehat{\mathbf{T}}^{(n)}$  has the following properties:*

$$\sqrt{n} \text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{C}(f)), \quad (11)$$

$$\lim_{n \rightarrow \infty} n\mathbb{E}[\text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f)) \text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f))^\top] = \mathbf{C}(f) \quad (12)$$

where

$$\mathbf{C}(f) = \text{Cov}(\text{vech}(\mathbf{H}_1(f)))$$

Previous results depend on the accurate estimation of the quadratic part of the estimator of  $T_{ij}^{(n)}$ , which is the issue of the following section.

## 4 Estimation of quadratic functionals

As pointed out in Section 2 the decomposition (4) has a quadratic part (6) that we want to estimate. To achieve this we will construct a general estimator of the form:

$$\theta = \int \eta(x_{i1}, x_{j2}, y) f(x_{i1}, x_{j1}, y) f(x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy,$$

for  $f \in \mathcal{E}$  and  $\eta : \mathbb{R}^3 \rightarrow \mathbb{R}$  a bounded function.

Given  $M_n$  a subset of  $D$ , consider the estimator

$$\begin{aligned} \hat{\theta}_n &= \frac{1}{n(n-1)} \sum_{l \in M} \sum_{k \neq k'=1}^n p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) \\ &\quad \int p_l(x_i, x_j, Y^{(k')}) \left( \eta(x_i, X_j^{(k')}, Y^{(k')}) + \eta(X_i^{(k')}, x_j, Y^{(k')}) \right) dx_i dx_j \\ &\quad - \frac{1}{n(n-1)} \sum_{l, l' \in M} \sum_{k \neq k'=1}^n p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) p_{l'}(X_i^{(k')}, X_j^{(k')}, Y^{(k')}) \\ &\quad \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \quad (13) \end{aligned}$$

In order to simplify the presentation of the main Theorem, let  $\psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) = \eta(x_{i1}, x_{j2}, y) + \eta(x_{i2}, x_{j1}, y)$  verifying

$$\int \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy = \int \psi(x_{i2}, x_{j2}, x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy.$$

With this notation we can simplify (13) in

$$\begin{aligned} \hat{\theta}_n &= \frac{1}{n(n-1)} \sum_{l \in M} \sum_{k \neq k'=1}^n p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) \int p_l(x_i, x_j, Y^{(k')}) \psi(x_i, x_j, X_i^{(k')}, X_j^{(k')}, Y^{(k')}) dx_i dx_j \\ &\quad - \frac{1}{n(n-1)} \sum_{l, l' \in M} \sum_{k \neq k'=1}^n p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) p_{l'}(X_i^{(k')}, X_j^{(k')}, Y^{(k')}) \\ &\quad \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \quad (14) \end{aligned}$$

Using simple algebra, it is possible to prove that this estimator has bias equal to

$$\begin{aligned} & - \int (S_M f(x_{i1}, x_{j1}, y) - f(x_{i1}, x_{j1}, y)) (S_M f(x_{i2}, x_{j2}, y) - f(x_{i2}, x_{j2}, y)) \\ & \quad \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \quad (15) \end{aligned}$$

The following Theorem gives an explicit bound for the variance of  $\hat{\theta}_n$ .

**Theorem 3.** *Let Assumption 1 hold. Then if  $|M_n|/n \rightarrow 0$  when  $n \rightarrow 0$ , then  $\hat{\theta}_n$  has the following property*

$$\left| n\mathbb{E}[(\hat{\theta}_n - \theta)^2] - \Lambda(f, \eta) \right| \leq \gamma \left[ \frac{|M_n|}{n} + \|S_{M_n} f - f\|_2 + \|S_{M_n} g - g\|_2 \right], \quad (16)$$

where  $g(x_i, x_j, y) = \int f(x_{i2}, x_{j2}, y) \psi(x_i, x_j, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2}$  and

$$\Lambda(f, \eta) = \int g(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int g(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2,$$

where  $\gamma$  is constant depending only on  $\|f\|_\infty, \|\eta\|_\infty$ , and  $\Delta_{x_i x_j} = (b_1 - a_1) \times (b_2 - a_2)$ . Moreover, this constant is an increasing function of these quantities.

Note that equation (16) implies that

$$\lim_{n \rightarrow \infty} n\mathbb{E}[\hat{\theta}_n - \theta]^2 = \Lambda(f, \eta).$$

These results will be stated in order to control the term

$$Q = \int H_2(\hat{f}, x_{i1}, x_{j2}, y) f(x_{i1}, x_{j1}, y) f(x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy$$

which has the form of the quadratic functional  $\theta$  with the particular choice  $\eta(x_{i1}, x_{j2}, y) = H_2(\hat{f}, x_{i1}, x_{j2}, y)$ . We point out that we also show that in this particular frame, we get  $\Lambda(f, \eta) = 0$ . This is the reason why the asymptotic variance of the estimate  $\hat{T}_{ij}^{(n)}$  built in the previous section, is only governed by its linear part, yielding asymptotic efficiency.

## 5 Proofs

### *Proof of Proposition 1.*

We need to calculate the three first derivatives of  $F(u)$ . In order to facilitate the calculation, we are going to differentiate  $m_i(f_u, y)$ :

$$\begin{aligned} \frac{d}{du} (m_i(f_u, y)) &= \frac{d}{du} \left( \frac{\int x_i f_u(x_i, x_j, y) dx_i dx_j}{\int f_u(x_i, x_j, y) dx_i dx_j} \right) \\ &= \frac{\int x_i (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j}{\int f_u(x_i, x_j, y) dx_i dx_j} \\ &\quad - \frac{\int x_i f_u(x_i, x_j, y) dx_i dx_j \int (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j}{\left( \int f_u(x_i, x_j, y) dx_i dx_j \right)^2}, \\ &= \frac{\int x_i (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j}{\int f_u(x_i, x_j, y) dx_i dx_j} \\ &\quad - \frac{m_i(f_u, y) \int (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j}{\int f_u(x_i, x_j, y) dx_i dx_j}, \\ &= \frac{\int (x_i - m_i(f_u, y)) (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j}{\int f_u(x_i, x_j, y) dx_i dx_j}. \end{aligned} \quad (17)$$

Now, using (17) we first compute  $F'(u)$ ,

$$\begin{aligned} &\int \frac{d}{du} (m_i(f_u, y)) m_j(f_u, y) f_u(x_i, x_j, y) + m_i(f_u, y) \frac{d}{du} (m_j(f_u, y)) f_u(x_i, x_j, y) \\ &+ m_i(f_u, y) m_j(f_u, y) \frac{d}{du} (f_u(x_i, x_j, y)) dx_i dx_j dy, \\ &= \int [x_i m_j(f_u, y) + x_j m_i(f_u, y) - m_i(f_u, y) m_j(f_u, y)] (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j dy. \end{aligned}$$

Taking  $u = 0$  we have

$$F'(0) = \int [x_i m_j(\hat{f}, y) + x_j m_i(\hat{f}, y) - m_i(\hat{f}, y) m_j(\hat{f}, y)] (f(x_i, x_j, y) - \hat{f}(x_i, x_j, y)) dx_i dx_j dy. \quad (18)$$

We derive now  $m_i(f_u, y)m_j(f_u, y)$  to obtain

$$\begin{aligned} \frac{d}{du} (m_i(f_u, y)m_j(f_u, y)) &= \frac{d}{du} (m_i(f_u, y)) m_j(f_u, y) + m_i(f_u, y) \frac{d}{du} (m_j(f_u, y)) \\ &= m_j(f_u, y) \frac{\int (x_i - m_i(f_u, y))(f(x_i, x_j, y) - \hat{f}(x_i, x_j, y))dx_i dx_j}{\int f_u(x_i, x_j, y)dx_i dx_j} \\ &\quad + m_i(f_u, y) \frac{\int (x_j - m_j(f_u, y))(f(x_i, x_j, y) - \hat{f}(x_i, x_j, y))dx_i dx_j}{\int f_u(x_i, x_j, y)dx_i dx_j}. \end{aligned} \quad (19)$$

Following with  $F''(u)$  and using (17) and (19) we get,

$$\begin{aligned} F''(u) &= \int \left[ x_{i1} \frac{\int (x_{j2} - m_j(f_u, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_{i2} dx_{j2}}{\int f_u(x_i, x_j, y)dx_i dx_j} \right. \\ &\quad + x_{j1} \frac{\int (x_{i2} - m_i(f_u, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_{i2} dx_{j2}}{\int f_u(x_i, x_j, y)dx_i dx_j} \\ &\quad - m_j(f_u, y) \frac{\int (x_{i2} - m_i(f_u, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_i dx_j}{\int f_u(x_i, x_j, y)dx_i dx_j} \\ &\quad \left. - m_i(f_u, y) \frac{\int (x_{j2} - m_j(f_u, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_{i2} dx_{j2}}{\int f_u(x_i, x_j, y)dx_i dx_j} \right] \\ &\quad (f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y))dx_{i1} dx_{j1} dy. \end{aligned}$$

Simplifying the last expression we obtain

$$\begin{aligned} F''(u) &= \\ &\int \frac{1}{\int f_u(x_i, x_j, y)dx_i dx_j} \left\{ (x_{i1} - m_i(f_u, y))(x_{j2} - m_j(f_u, y)) + (x_{i2} - m_i(f_u, y))(x_{j1} - m_j(f_u, y)) \right\} \\ &\quad (f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \end{aligned}$$

Besides, when  $u = 0$

$$\begin{aligned} F''(0) &= \quad (20) \\ &\int \frac{1}{\int \hat{f}(x_i, x_j, y)dx_i dx_j} \left\{ (x_{i1} - m_i(\hat{f}, y))(x_{j2} - m_j(\hat{f}, y)) + (x_{i2} - m_i(\hat{f}, y))(x_{j1} - m_j(\hat{f}, y)) \right\} \\ &\quad (f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\ &= \int \frac{2}{\int \hat{f}(x_i, x_j, y)dx_i dx_j} (x_{i1} - m_i(\hat{f}, y))(x_{j2} - m_j(\hat{f}, y)) \\ &\quad (f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y))dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \end{aligned} \quad (21)$$

Using the previous arguments we can finally find  $F'''(u)$ :

$$F'''(u) = \int \frac{-6}{\int f_u(x_i, x_j, y) dx_i dx_j} (x_{i1} - m_j(f_u, y))(x_{j2} - m_j(f_u, y)) (f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y))(f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y)) (f(x_{i3}, x_{j3}, y) - \hat{f}(x_{i3}, x_{j3}, y)) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy \quad (22)$$

Replacing (18), (21) and (22) into (3) we get the desired decomposition.  $\square$

**Proof of Theorem 1.**

We will first control the remaining term (7),

$$\Gamma_n = \frac{1}{6} F'''(\xi)(1 - \xi)^3.$$

Remember that

$$F'''(\xi) = -6 \int \frac{(x_{i1} - m_i(f_\xi, y))(x_{j2} - m_j(f_\xi, y))}{(\int f_\xi(x_i, x_j, y) dx_i dx_j)^2} \left( f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y) \right) \left( f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y) \right) \left( f(x_{i3}, x_{j3}, y) - \hat{f}(x_{i3}, x_{j3}, y) \right) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy,$$

Assumptions 1 and 2 ensure that the first part of the integrand is bounded by a constant  $\mu$ . Furthermore,

$$\begin{aligned} |\Gamma_n| &\leq \mu \int \left| f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y) \right| \left| f(x_{i2}, x_{j2}, y) - \hat{f}(x_{i2}, x_{j2}, y) \right| \\ &\quad \left| f(x_{i3}, x_{j3}, y) - \hat{f}(x_{i3}, x_{j3}, y) \right| dx_{i1} dx_{j1} dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy \\ &= \mu \int \left( \int \left| f(x_i, x_j, y) - \hat{f}(x_i, x_j, y) \right| dx_i dx_j \right)^3 dy \\ &\leq \mu \Delta_{x_i x_j}^3 \int \left| f(x_{i1}, x_{j1}, y) - \hat{f}(x_{i1}, x_{j1}, y) \right|^3 dx_i dx_j dy \end{aligned}$$

by the Hölder inequality. Then  $\mathbb{E}[\Gamma_n^2] = O(\mathbb{E}(f |f - \hat{f}|^3)^2) = O(\mathbb{E}\|f - \hat{f}\|_3^6)$ . Since  $\hat{f}$  verifies Assumption 3, this quantity is of order  $O(n_1^{-6\lambda})$ . Since we also assume  $n_1 \approx n/\log(n)$  and  $\lambda > 1/6$ , then  $n_1^{-6\lambda} = o(\frac{1}{n})$ . Therefore, we get  $\mathbb{E}[\Gamma_n^2] = o(1/n)$  which implies that the remaining term  $\Gamma_n$  is negligible.

To prove the asymptotic normality of  $\hat{T}_{ij}^{(n)}$ , we shall show that  $\sqrt{n} \left( \hat{T}_{ij}^{(n)} - T_{ij}(f) \right)$  and define

$$Z_{ij}^{(n)} = \frac{1}{n_2} \sum_{k=1}^{n_2} H_1(f, X_i^{(k)}, X_j^{(k)}, Y^{(k)}) - \int H_1(f, x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \quad (23)$$

have the same asymptotic behavior. We can get for  $Z_{ij}^{(n)}$  a classic central limit theorem with variance

$$\begin{aligned} C_{ij}(f) &= \text{Var}(H_1(f, x_i, x_j, y)) \\ &= \int H_1(f, x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int H_1(f, x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 \end{aligned}$$

which implies (9) and (10). In order to establish our claim, we will show that

$$R_{ij}^{(n)} = \sqrt{n} \left[ \widehat{T}_{ij}^{(n)} - T_{ij}(f) - Z_{ij}^{(n)} \right] \quad (24)$$

has second-order moment converging to 0.

Define  $\widehat{Z}_{ij}^{(n)}$  as  $Z_{ij}^{(n)}$  with  $f$  replaced by  $\hat{f}$ . Let us note that  $R_{ij}^{(n)} = R_1 + R_2$  where

$$R_1 = \sqrt{n} \left[ \widehat{T}_{ij}^{(n)} - T_{ij}(f) - \widehat{Z}_{ij}^{(n)} \right]$$

$$R_2 = \sqrt{n} \left[ \widehat{Z}_{ij}^{(n)} - Z_{ij}^{(n)} \right].$$

It only remains to state that  $\mathbb{E}[R_1^2]$  and  $\mathbb{E}[R_2^2]$  converges to 0. We can rewrite  $R_1$  as

$$R_1 = -\sqrt{n} \left[ \widehat{Q} - Q + \Gamma_n \right]$$

where we note that

$$Q = \int H_2(\hat{f}, x_{i1}, x_{j2}, y) f(x_{i1}, x_{j1}, y) f(x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy$$

$$H_2(\hat{f}, x_{i1}, x_{j2}, y) = \frac{1}{\int \hat{f}(x_i, x_j, y) dx_i dx_j} \left( x_{i1} - m_i(\hat{f}, y) \right) \left( x_{j2} - m_j(\hat{f}, y) \right)$$

has the form of a quadratic functional studied in Section 4 with  $\eta(x_{i1}, x_{j2}, y) = H_2(\hat{f}, x_{i1}, x_{j2}, y)$ . Hence such functional can be estimated as done in Section 4 and let  $\widehat{Q}$  be its corresponding estimator. Since  $\mathbb{E}[\Gamma_n^2] = o(1/n)$ , we only have to control the term  $\sqrt{n}(\widehat{Q} - Q)$  which is such that  $\lim_{n \rightarrow \infty} n \mathbb{E}[\widehat{Q} - Q]^2 = 0$  by Lemma 7. This Lemma implies that  $\mathbb{E}[R_1^2] \rightarrow 0$  as  $n \rightarrow \infty$ . For  $R_2$  we have

$$\mathbb{E}[R_2^2] = \frac{n}{n_2} \left[ \int \left( H_1(f, x_i, x_j, y) - H_1(\hat{f}, x_i, x_j, y) \right)^2 f(x_i, x_j, y) dx_i dx_j dy \right]$$

$$- \frac{n}{n_2} \left[ \int H_1(f, x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy - \int H_1(\hat{f}, x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy \right]^2.$$

The same arguments as the ones of Lemma 7 (mean value and Assumptions 2 and 3) show that  $\mathbb{E}[R_2^2] \rightarrow 0$ .  $\square$

**Proof of Theorem 2.** To prove the inequality we will use the usual framework described in Ibragimov & Khas'minskii (1991). The first step is to calculate the Fréchet derivative of  $T_{ij}(f)$  at some point  $f_0 \in \mathcal{E}$ . Assumptions 2 and 3 and equation (4), imply that

$$T_{ij}(f) - T_{ij}(f_0) = \int \left( x_i m_j(f_0, y) + x_j m_i(f_0, y) - m_i(f_0, y) m_j(f_0, y) \right)$$

$$\left( f(x_i, x_j, y) - f_0(x_i, x_j, y) \right) dx_i dx_j dy + O \left( \int (f - f_0)^2 \right)$$

where  $m_i(f_0, y) = \int x_i f_0(x_i, x_j, y) dx_i dx_j dy / \int f_0(x_i, x_j, y) dx_i dx_j dy$ . Therefore, the Fréchet derivative of  $T_{ij}(f)$  at  $f_0$  is  $T'_{ij}(f_0) \cdot h = \langle H_1(f_0, \cdot), h \rangle$  with

$$H_1(f_0, x_i, x_j, y) = x_i m_j(f_0, y) + x_j m_i(f_0, y) - m_i(f_0, y) m_j(f_0, y).$$

Using the results of Ibragimov & Khas'minskii (1991), denote  $H(f_0) = \{u \in \mathbb{L}^2(dx_i dx_j dy), \int u(x_i, x_j, y) \sqrt{f_0(x_i, x_j, y)} dx_i dx_j dy = 0\}$  the set of functions in  $\mathbb{L}^2(dx_i dx_j dy)$  orthogonal to  $\sqrt{f_0}$ ,  $\text{Pr}_{H(f_0)}$  the projection onto  $H(f_0)$ ,  $A_n(t) = (\sqrt{f_0})t/\sqrt{n}$  and  $P_{f_0}^{(n)}$  the joint distribution of  $(X_i^{(k)}, X_j^{(k)})$   $k = 1, \dots, n$  under  $f_0$ . Since  $(X_i^{(k)}, X_j^{(k)})$   $k = 1, \dots, n$  are i.i.d., the family  $\{P_{f_0}^{(n)}, f \in \mathcal{E}\}$  is differentiable in quadratic mean at  $f_0$  and therefore locally asymptotically normal at all points  $f_0 \in \mathcal{E}$  in the direction  $H(f_0)$  with normalizing factor  $A_n(f_0)$  (see the details in Van der Vaart (2000)). Then, by the results of Ibragimov & Khas'minskii (1991) say that under these conditions, denoting  $K_n = B_n \theta'(f_0) A_n \text{Pr}_{H(f_0)}$  with  $B_n = \sqrt{n}u$ , if  $K_n \xrightarrow{\mathcal{D}} K$  and if  $K(u) = \langle t, u \rangle$ , then for every estimator  $\widehat{T}_{ij}^{(n)}$  of  $T_{ij}(f)$  and every family  $\mathcal{V}(f_0)$  of vicinities of  $f_0$ , we have

$$\inf_{\{\mathcal{V}(f_0)\}} \liminf_{n \rightarrow \infty} \sup_{f \in \mathcal{V}(f_0)} n \mathbb{E} [\widehat{T}_{ij}^{(n)} - T_{ij}(f_0)]^2 \geq \|t\|_{\mathbb{L}^2(dx_i dx_j dy)}^2.$$

Here,

$$K_n(u) = \sqrt{n} T'(f_0) \cdot \frac{\sqrt{f_0}}{\sqrt{n}} \text{Pr}_{H(f_0)}(u) = T'(f_0) \left( \sqrt{f_0} \left( u - \sqrt{f_0} \int u \sqrt{f_0} \right) \right),$$

since for any  $u \in \mathbb{L}^2(dx_i dx_j dy)$  we can write it as  $u = \sqrt{f_0} \langle \sqrt{f_0}, u \rangle + \text{Pr}_{H(f_0)}(u)$ . In this case  $K_n(u)$  does not depend on  $n$  and

$$\begin{aligned} K(h) &= T'(f_0) \cdot \left( \sqrt{f_0} \left( u - \sqrt{f_0} \int h \sqrt{f_0} \right) \right) \\ &= \int H_1(f_0, \cdot) \sqrt{f_0} u - \int H_1(f_0, \cdot) \sqrt{f_0} \int u \sqrt{f_0} \\ &= \langle t, u \rangle \end{aligned}$$

with

$$t(x_i, x_j, y) = H_1(f_0, x_i, x_j, y) \sqrt{f_0} - \left( \int H_1(f_0, x_i, x_j, y) f_0 \right) \sqrt{f_0}.$$

The semi-parametric Cramér-Rao bound for this problem is thus

$$\|t\|_{\mathbb{L}^2(dx_i, dx_j, dy)}^2 = \int H_1(f_0, x_i, x_j, y)^2 f_0 dx_i dx_j dy - \left( \int H_1(f_0, x_i, x_j, y) f_0 dx_i dx_j dy \right)^2 = C_{ij}(f_0)$$

and we recognize the expression  $C_{ij}(f_0)$  found in Theorem 1.  $\square$

**Proof of Corollary 1.** The proof is based in the following observation. Employing equation (24) we have

$$\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f) = \mathbf{Z}^{(n)}(f) + \frac{\mathbf{R}^{(n)}}{\sqrt{n}}$$

where  $\mathbf{Z}^{(n)}(f)$  and  $\mathbf{R}^{(n)}$  are matrices with elements  $Z_{ij}^{(n)}$  and  $R_{ij}^{(n)}$ , defined in (23) and (24), respectively.

Hence we have,

$$n\mathbb{E}[\|\text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f) - \mathbf{Z}^{(n)}(f))\|^2] = \mathbb{E}[\|\text{vech}(\mathbf{R}^{(n)})\|^2] = \sum_{i \leq j} \mathbb{E}[(R_{ij}^{(n)})^2].$$

We see by Lemma 7 that  $\mathbb{E}[R_{ij}^2] \rightarrow 0$  as  $n \rightarrow 0$ . It follows that

$$n\mathbb{E}[\|\text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f) - \mathbf{Z}^{(n)}(f))\|^2] \rightarrow 0 \text{ as } n \rightarrow 0.$$

We know that if  $X_n$ ,  $X$  and  $Y_n$  are random variables, then if  $X_n \xrightarrow{\mathcal{D}} X$  and  $(X_n - Y_n) \xrightarrow{\mathcal{P}} 0$ , follows that  $Y_n \xrightarrow{\mathcal{D}} X$ .

Remember also that convergence in  $\mathbb{L}^2$  implies convergence in probability, therefore

$$\sqrt{n} \text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f) - \mathbf{Z}^{(n)}(f)) \xrightarrow{\mathcal{P}} 0.$$

By the multivariate central limit theorem we have that  $\sqrt{n} \text{vech}(\mathbf{Z}^{(n)}(f)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{C}(f))$ . Therefore,  $\sqrt{n} \text{vech}(\widehat{\mathbf{T}}^{(n)} - \mathbf{T}(f)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathbf{C}(f))$ .  $\square$

**Proof of Theorem 3.** For abbreviation, we write  $M$  instead of  $M_n$  and set  $m = |M_n|$ . We first compute the mean squared error of  $\hat{\theta}_n$  as

$$\mathbb{E}[\hat{\theta}_n - \theta]^2 = \text{Bias}^2(\hat{\theta}_n) + \text{Var}(\hat{\theta}_n)$$

where  $\text{Bias}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n] - \theta$ .

We begin the proof by bounding  $\text{Var}(\hat{\theta}_n)$ . Let  $A$  and  $B$  be  $m \times 1$  vectors with components

$$\begin{aligned} a_l &= \int p_l(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \quad l = 1, \dots, m, \\ b_l &= \int p_l(x_{i1}, x_{j1}, y) f(x_{i2}, x_{j2}, y) \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\ &= \int p_l(x_i, x_j, y) g(x_i, x_j, y) dx_i dx_j dy \quad l = 1, \dots, m \end{aligned}$$

where  $g(x_i, x_j, y) = \int f(x_{i2}, x_{j2}, y) \psi(x_i, x_j, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2}$ . Let  $Q$  and  $R$  be  $m \times 1$  vectors of centered functions

$$\begin{aligned} q_l(x_i, x_j, y) &= p_l(x_i, x_j, y) - a_l \\ r_l(x_i, x_j, y) &= \int p_l(x_{i2}, x_{j2}, y) \psi(x_i, x_j, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} - b_l \end{aligned}$$

for  $l = 1, \dots, m$ . Let  $C$  a  $m \times m$  matrix of constants

$$c_{ll'} = \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \quad l, l' = 1, \dots, m.$$

Let us denote by  $U_n$  the process

$$U_n h = \frac{1}{n(n-1)} \sum_{k \neq k'=1}^n h(X_i^{(k)}, X_j^{(k)}, Y^{(k)}, X_i^{(k')}, X_j^{(k')}, Y^{(k')})$$

and  $P_n$  the empirical measure

$$P_n h = \frac{1}{n} \sum_{k=1}^n h(X_i^{(k)}, X_j^{(k)}, Y^{(k)})$$

for some  $h$  in  $\mathbb{L}^2(dx_i, dx_j, dy)$ . With these notations,  $\hat{\theta}_n$  has the Hoeffding's decomposition

$$\begin{aligned} \hat{\theta}_n &= \frac{1}{n(n-1)} \sum_{l \in M} \sum_{k \neq k'=1}^n (q_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) + a_l) (r_l(X_i^{(k')}, X_j^{(k')}, Y^{(k')}) + b_l) \\ &\quad - \frac{1}{n(n-1)} \sum_{l, l' \in M} \sum_{k \neq k'=1}^n (q_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) + a_l) (q_{l'}(X_i^{(k')}, X_j^{(k')}, Y^{(k')}) + a_{l'}) c_{ll'} \\ &= U_n K + P_n L + A^\top B - A^\top C A \end{aligned}$$

where

$$\begin{aligned} K(x_{i1}, x_{j1}, y_1, x_{i2}, x_{j2}, y_2) &= Q^\top(x_{i1}, x_{j1}, y_1) R(x_{i2}, x_{j2}, y_2) - Q^\top(x_{i1}, x_{j1}, y_1) C Q(x_{i2}, x_{j2}, y_2) \\ L(x_i, x_j, y) &= A^\top R(x_i, x_j, y) + B Q(x_i, x_j, y) - 2A^\top C Q(x_i, x_j, y). \end{aligned}$$

Therefore  $\text{Var}(\hat{\theta}_n) = \text{Var}(U_n K) + \text{Var}(P_n L) - 2 \text{Cov}(U_n K, P_n L)$ . These three terms are bounded in Lemmas 2 - 4, which gives

$$\text{Var}(\hat{\theta}_n) \leq \frac{20}{n(n-1)} \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 (m+1) + \frac{12}{n} \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2.$$

For  $n$  enough large and a constant  $\gamma \in \mathbb{R}$ ,

$$\text{Var}(\hat{\theta}_n) \leq \gamma \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 \left( \frac{m}{n^2} + \frac{1}{n} \right).$$

The term Bias  $(\hat{\theta}_n)$  is easily computed, as proven in Lemma 5, is equal to

$$\begin{aligned} & - \int (S_M f(x_{i1}, x_{j1}, y) - f(x_{i1}, x_{j1}, y)) (S_M f(x_{i2}, x_{j2}, y) - f(x_{i2}, x_{j2}, y)) \\ & \quad \eta(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \end{aligned}$$

From Lemma 5, the bias of  $\hat{\theta}_n$  is bounded by

$$\left| \text{Bias}(\hat{\theta}_n) \right| \leq \Delta_{x_i x_j} \|\eta\|_\infty \sup_{l \notin M} |c_l|^2.$$

The assumption of  $\left( \sup_{l \notin M} |c_l|^2 \right)^2 \approx m/n^2$  and since  $m/n \rightarrow 0$ , we deduce that  $\mathbb{E}[\hat{\theta}_n - \theta]^2$  has a parametric rate of convergence  $O(1/n)$ .

Finally to prove (16), note that

$$\begin{aligned} n\mathbb{E}[\hat{\theta}_n - \theta]^2 &= n \text{Bias}^2(\hat{\theta}_n) + n \text{Var}(\hat{\theta}_n) \\ &= n \text{Bias}^2(\hat{\theta}_n) + n \text{Var}(U_n K) + n \text{Var}(P_n L). \end{aligned}$$

We previously proved that for some  $\lambda_1, \lambda_2 \in \mathbb{R}$

$$\begin{aligned} n \text{Bias}^2(\hat{\theta}_n) &\leq \lambda_1 \Delta_{x_i x_j}^2 \|\eta\|_\infty^2 \frac{m}{n} \\ n \text{Var}(U_n K) &\leq \lambda_2 \Delta_{x_i x_j}^2 \|f\|_\infty^2 \|\eta\|_\infty^2 \frac{m}{n}. \end{aligned}$$

Thus, Lemma 6 implies

$$|n \text{Var}(P_n L) - \Lambda(f, \eta)| \leq \lambda [\|S_M f - f\|_2 + \|S_M g - g\|_2],$$

where  $\lambda$  is an increasing function of  $\|f\|_\infty^2$ ,  $\|\eta\|_\infty^2$  and  $\Delta_{x_i x_j}$ . From all this we deduce (16) which ends the proof of Theorem 3.  $\square$

## 6 Technical Results

**Lemma 1** (Bias of  $\hat{\theta}_n$ ). *The estimator  $\hat{\theta}_n$  defined in (14) estimates  $\theta$  with bias equal to*

$$-\int (S_M f(x_{i1}, x_{j1}, y) - f(x_{i1}, x_{j1}, y)) (S_M f(x_{i2}, x_{j2}, y) - f(x_{i2}, x_{j2}, y)) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy.$$

*Proof.* Let  $\hat{\theta}_n = \hat{\theta}_n^1 - \hat{\theta}_n^2$  where

$$\begin{aligned} \hat{\theta}_n^1 &= \frac{1}{n(n-1)} \sum_{l \in M} \sum_{k \neq k'=1} p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) \int p_l(x_i, x_j, Y^{(k')}) \psi(x_i, x_j, X_i^{(k')}, X_j^{(k')}, Y^{(k')}) dx_i dx_j \\ \hat{\theta}_n^2 &= -\frac{1}{n(n-1)} \sum_{l, l' \in M} \sum_{k \neq k'=1}^n p_l(X_i^{(k)}, X_j^{(k)}, Y^{(k)}) p_{l'}(X_i^{(k')}, X_j^{(k')}, Y^{(k')}) \\ &\quad \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy. \end{aligned}$$

Let us first compute  $\mathbb{E}[\hat{\theta}_n^1]$ .

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_n^1] &= \sum_{l \in M} \int p_l(x_{i1}, x_{j1}, y) f(x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dy \\
&\quad \int p_l(x_{i1}, x_{j1}, y) \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) f(x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&= \sum_{l \in M} a_l \int p_l(x_{i1}, x_{j1}, y) \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) f(x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&= \int \left( \sum_{l \in M} a_l p_l(x_{i2}, x_{j2}, y) \right) \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) f(x_{i2}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&= \int S_M f(x_{i1}, x_{j1}, y) f(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&\quad + \int S_M f(x_{i2}, x_{j2}, y) f(x_{i1}, x_{j1}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy
\end{aligned}$$

Now for  $\hat{\theta}_n^2$ , we get

$$\begin{aligned}
\mathbb{E}[\hat{\theta}_n^2] &= \sum_{l, l' \in M} \int p_l(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \int p_{l'}(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \\
&\quad \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&= \sum_{l, l' \in M} a_l a_{l'} \int p_l(x_{i1}, x_{j1}, y) p_{l'}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&= \int \left( \sum_{l \in M} a_l p_l(x_{i1}, x_{j1}, y) \right) \left( \sum_{l' \in M} a_{l'} p_{l'}(x_{i2}, x_{j2}, y) \right) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&= \int S_M f(x_{i1}, x_{j1}, y) S_M f(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy.
\end{aligned}$$

Arranging these terms and using

$$\text{Bias}(\hat{\theta}_n) = \mathbb{E}[\hat{\theta}_n] - \theta = \mathbb{E}[\hat{\theta}_n^1] - \mathbb{E}[\hat{\theta}_n^2] - \theta$$

we obtain the desire bias.  $\square$

**Lemma 2** (Bound of  $\text{Var}(U_n K)$ ). *Under the assumptions of Theorem 3, we have*

$$\text{Var}(U_n K) \leq \frac{20}{n(n-1)} \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 (m+1)$$

*Proof.* Note that  $U_n K$  is centered because  $Q$  and  $R$  are centered and  $(X_i^{(k)}, X_j^{(k)}, Y^{(k)})$ ,  $k =$

$1, \dots, n$  is an independent sample. So  $\text{Var}(U_n K)$  is equal to

$$\begin{aligned} \mathbb{E}[U_n K]^2 &= \mathbb{E}\left(\frac{1}{(n(n-1))^2} \sum_{k_1 \neq k'_1=1}^n \sum_{k_2 \neq k'_2=1}^n K(X_i^{(k_1)}, X_j^{(k_1)}, Y^{(k_1)}, X_i^{(k'_1)}, X_j^{(k'_1)}, Y^{(k'_1)}) \right. \\ &\quad \left. K(X_i^{(k_2)}, X_j^{(k_2)}, Y^{(k_2)}, X_i^{(k'_2)}, X_j^{(k'_2)}, Y^{(k'_2)})\right) \\ &= \frac{1}{n(n-1)} \mathbb{E}\left(K^2(X_i^{(1)}, X_j^{(1)}, Y^{(1)}, X_i^{(2)}, X_j^{(2)}, Y^{(2)}) \right. \\ &\quad \left. + K(X_i^{(1)}, X_j^{(1)}, Y^{(1)}, X_i^{(2)}, X_j^{(2)}, Y^{(2)}) K(X_i^{(2)}, X_j^{(2)}, Y^{(2)}, X_i^{(1)}, X_j^{(1)}, Y^{(1)})\right) \end{aligned}$$

By the Cauchy-Schwarz inequality, we get

$$\text{Var}(U_n K) \leq \frac{2}{n(n-1)} \mathbb{E}[K^2(X_i^{(1)}, X_j^{(1)}, Y^{(1)}, X_i^{(2)}, X_j^{(2)}, Y^{(2)})].$$

Moreover, using the fact that  $2|\mathbb{E}[XY]| \leq \mathbb{E}[X^2] + \mathbb{E}[Y^2]$ , we obtain

$$\begin{aligned} \mathbb{E}[K^2(X_i^{(1)}, X_j^{(1)}, Y^{(1)}, X_i^{(2)}, X_j^{(2)}, Y^{(2)})] &\leq 2 \left[ \mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})R(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))^2] \right. \\ &\quad \left. + \mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})CQ(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))^2] \right]. \end{aligned}$$

We will bound these two terms. The first one is

$$\begin{aligned} &\mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})R(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))^2] \\ &= \sum_{l, l' \in M} \left( \int p_l(x_i, x_j, y) p_{l'}(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy - a_l a_{l'} \right) \\ &\quad \left( \int p_l(x_{i2}, x_{j2}, y) p_{l'}(x_{i3}, x_{j3}, y) \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) \right. \\ &\quad \left. \psi(x_{i1}, x_{j1}, x_{i3}, x_{j3}, y) f(x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy - b_l b_{l'} \right) \\ &= W_1 - W_2 - W_3 + W_4 \end{aligned}$$

where

$$\begin{aligned}
W_1 &= \int \sum_{l,l' \in M} p_l(x_{i_1}, x_{j_1}, y) p_{l'}(x_{i_1}, x_{j_1}, y) p_l(x_{i_2}, x_{j_2}, y') p_{l'}(x_{i_3}, x_{j_3}, y') \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y') \\
&\quad \psi(x_{i_4}, x_{j_4}, x_{i_3}, x_{j_3}, y') f(x_{i_1}, x_{j_1}, y) f(x_{i_4}, x_{j_4}, y') dx_{i_1} dx_{j_1} dx_{i_2} dx_{j_2} dx_{i_3} dx_{j_3} dx_{i_4} dx_{j_4} dy dy' \\
W_2 &= \int \sum_{l,l' \in M} b_l b_{l'} p_l(x_{i_1}, x_{j_1}, y) p_{l'}(x_{i_1}, x_{j_1}, y) f(x_{i_1}, x_{j_1}, y) dx_{i_1} dx_{j_1} dy \\
W_3 &= \int \sum_{l,l' \in M} a_l a_{l'} p_l(x_{i_2}, x_{j_2}, y') p_{l'}(x_{i_3}, x_{j_3}, y') \\
&\quad \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y') \psi(x_{i_4}, x_{j_4}, x_{i_3}, x_{j_3}, y') f(x_{i_4}, x_{j_4}, y') dx_{i_2} dx_{j_2} dx_{i_3} dx_{j_3} dx_{i_4} dx_{j_4} dy' \\
W_4 &= \sum_{l,l' \in M} a_l a_{l'} b_l b_{l'}.
\end{aligned}$$

$W_2$  and  $W_3$  are positive, hence

$$\mathbb{E} \left[ \left( 2Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) R(X_i^{(2)}, X_j^{(2)}, Y^{(2)}) \right)^2 \right] \leq W_1 + W_4.$$

$$\begin{aligned}
W_1 &= \int \sum_{l,l' \in M} p_l(x_{i_1}, x_{j_1}, y) p_{l'}(x_{i_1}, x_{j_1}, y) \left( \int p_l(x_{i_2}, x_{j_2}, y') \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y') dx_{i_2} dx_{j_2} \right) \\
&\quad \left( \int p_{l'}(x_{i_3}, x_{j_3}, y') \psi(x_{i_4}, x_{j_4}, x_{i_3}, x_{j_3}, y') dx_{i_3} dx_{j_3} \right) f(x_{i_1}, x_{j_1}, y) f(x_{i_4}, x_{j_4}, y') dx_{i_1} dx_{j_1} dx_{i_4} dx_{j_4} dy dy' \\
&\leq \|f\|_\infty^2 \sum_{l,l' \in M} \int p_l(x_{i_1}, x_{j_1}, y) p_{l'}(x_{i_1}, x_{j_1}, y) dx_{i_1} dx_{j_1} dy \\
&\quad \int \left( \int p_l(x_{i_2}, x_{j_2}, y') \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y') dx_{i_2} dx_{j_2} \right) \\
&\quad \left( \int p_{l'}(x_{i_3}, x_{j_3}, y') \psi(x_{i_4}, x_{j_4}, x_{i_3}, x_{j_3}, y') dx_{i_3} dx_{j_3} \right) dx_{i_2} dx_{j_2} dx_{i_4} dx_{j_4} dy'
\end{aligned}$$

Since  $p_l$ 's are orthonormal we have

$$W_1 \leq \|f\|_\infty^2 \sum_{l \in M} \int \left( \int p_l(x_{i_2}, x_{j_2}, y') \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y') dx_{i_2} dx_{j_2} \right)^2 dx_{i_4} dx_{j_4} dy'.$$

Moreover by the Cauchy-Schwarz inequality and  $\|\psi\|_\infty \leq 2\|\eta\|_\infty$

$$\begin{aligned}
\left( \int p_l(x_{i_2}, x_{j_2}, y') \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y') dx_{i_2} dx_{j_2} \right)^2 &\leq \int p_l(x_{i_2}, x_{j_2}, y')^2 dx_{i_2} dx_{j_2} \\
&\quad \int \psi(x_{i_4}, x_{j_4}, x_{i_2}, x_{j_2}, y')^2 dx_{i_2} dx_{j_2} \\
&\leq \|\psi\|_\infty^2 \Delta_{x_i x_j} \int p_l(x_{i_2}, x_{j_2}, y')^2 dx_{i_2} dx_{j_2} \\
&\leq 4\|\eta\|_\infty^2 \Delta_{x_i x_j} \int p_l(x_{i_2}, x_{j_2}, y')^2 dx_{i_2} dx_{j_2},
\end{aligned}$$

and then

$$\begin{aligned} & \int \left( \int p_l(x_{i2}, x_{j2}, y') \psi(x_{i4}, x_{j4}, x_{i2}, x_{j2}, y') dx_{i2} dx_{j2} \right)^2 dx_{i4} dx_{j4} dy' \\ & \leq 4 \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 \int p_l(x_{i2}, x_{j2}, y')^2 dx_{i2} dx_{j2} dy' \\ & = 4 \|\eta\|_\infty^2 \Delta_{x_i x_j}^2. \end{aligned}$$

Finally,

$$W_1 \leq 4 \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 m.$$

For the term  $W_4$  using the facts that  $S_M f$  and  $S_M g$  are projection and that  $\int f = 1$ , we have

$$W_4 = \left( \sum_{l \in M} a_l b_l \right)^2 \leq \sum_{l \in M} a_l^2 \sum_{l \in M} b_l^2 \leq \|f\|_2^2 \|g\|_2^2 \leq \|f\|_\infty \|g\|_2^2.$$

By the Cauchy-Schwartz inequality we have  $\|g\|_2^2 \leq 4 \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2$  and then

$$W_4 \leq 4 \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2$$

which leads to

$$\mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})R(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))^2] \leq 4 \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 (m+1). \quad (25)$$

The second term  $\mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})CQ(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))] = W_5 - 2W_6 + W_7$  where

$$W_5 = \int \sum_{l_1, l'_1} \sum_{l_2, l'_2} c_{l_1 l'_1} c_{l_2 l'_2} p_{l_1}(x_{i1}, x_{j1}, y) p_{l_2}(x_{i1}, x_{j1}, y) p_{l'_1}(x_{i2}, x_{j2}, y') p_{l'_2}(x_{i2}, x_{j2}, y') \\ f(x_{i1}, x_{j1}, y) f(x_{i2}, x_{j2}, y') dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy' dy$$

$$W_6 = \int \sum_{l_1, l'_1} \sum_{l_2, l'_2} c_{l_1 l'_1} c_{l_2 l'_2} a_{l_1} a_{l_2} p_{l'_1}(x_i, x_j, y) p_{l'_2}(x_i, x_j, y) dx_i dx_j dy$$

$$W_7 = \sum_{l_1, l'_1} \sum_{l_2, l'_2} c_{l_1 l'_1} c_{l_2 l'_2} a_{l_1} a_{l'_1} a_{l_2} a_{l'_2}.$$

Using the previous manipulation, we show that  $W_6 \geq 0$ . Thus

$$\mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})CQ(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))] \leq W_5 + W_7.$$

First, observe that

$$\begin{aligned}
W_5 &= \sum_{l_1, l'_1} \sum_{l_2, l'_2} c_{l_1 l'_1} c_{l_2 l'_2} \left( \int p_{l_1}(x_{i1}, x_{j1}, y) p_{l_2}(x_{i1}, x_{j1}, y) f(x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dy \right) \\
&\quad \left( \int p_{l'_1}(x_{i2}, x_{j2}, y') p_{l'_2}(x_{i2}, x_{j2}, y') f(x_{i2}, x_{j2}, y') dx_{i2} dx_{j2} dy' \right) \\
&\leq \|f\|_\infty^2 \sum_{l_1, l'_1} \sum_{l_2, l'_2} c_{l_1 l'_1} c_{l_2 l'_2} \left( \int p_{l_1}(x_{i1}, x_{j1}, y) p_{l_2}(x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dy \right) \\
&\quad \left( \int p_{l'_1}(x_{i2}, x_{j2}, y') p_{l'_2}(x_{i2}, x_{j2}, y') dx_{i2} dx_{j2} dy' \right) \\
&= \|f\|_\infty^2 \sum_{l, l'} c_{ll'}^2
\end{aligned}$$

again using the orthonormality of the the  $p_l$ 's. Besides given the decomposition  $p_l(x_i, x_j, y) = \alpha_{l_\alpha}(x_i, x_j) \beta_{l_\beta}(y)$ ,

$$\begin{aligned}
\sum_{l, l'} c_{ll'}^2 &= \int \sum_{l_\beta, l'_\beta} \beta_{l_\beta}(y) \beta_{l'_\beta}(y) \beta_{l_\beta}(y') \beta_{l'_\beta}(y') \\
&\quad \sum_{l_\alpha, l'_\alpha} \left( \int \alpha_{l_\alpha}(x_{i1}, x_{j1}) \alpha_{l'_\alpha}(x_{i2}, x_{j2}) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} \right) \\
&\quad \left( \int \alpha_{l_\alpha}(x_{i3}, x_{j3}) \alpha_{l'_\alpha}(x_{i4}, x_{j4}) \eta(x_{i3}, x_{j4}, y') dx_{i3} dx_{j3} dx_{i4} dx_{j4} \right) dy dy'
\end{aligned}$$

But

$$\begin{aligned}
&\sum_{l_\alpha, l'_\alpha} \left( \int \alpha_{l_\alpha}(x_{i1}, x_{j1}) \alpha_{l'_\alpha}(x_{i2}, x_{j2}) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} \right) \\
&\quad \left( \int \alpha_{l_\alpha}(x_{i3}, x_{j3}) \alpha_{l'_\alpha}(x_{i4}, x_{j4}) \eta(x_{i3}, x_{j4}, y') dx_{i3} dx_{j3} dx_{i4} dx_{j4} \right) \\
&= \sum_{l_\alpha, l'_\alpha} \int \alpha_{l_\alpha}(x_{i1}, x_{j1}) \alpha_{l'_\alpha}(x_{i2}, x_{j2}) \eta(x_{i1}, x_{j2}, y) \alpha_{l_\alpha}(x_{i3}, x_{j3}) \\
&\quad \alpha_{l'_\alpha}(x_{i4}, x_{j4}) \eta(x_{i3}, x_{j4}, y') dx_{i1} dx_{j1} dx_{i2} dx_{j2} dx_{i3} dx_{j3} dx_{i4} dx_{j4} \\
&= \int \sum_{l_\alpha} \left( \int \alpha_{l_\alpha}(x_{i1}, x_{j1}) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} \right) \alpha_{l_\alpha}(x_{i3}, x_{j3}) \\
&\quad \sum_{l'_\alpha} \left( \int \alpha_{l'_\alpha}(x_{i4}, x_{j4}) \eta(x_{i3}, x_{j4}, y') dx_{i4} dx_{j4} \right) \alpha_{l'_\alpha}(x_{i2}, x_{j2}) dx_{i2} dx_{j2} dx_{i3} dx_{j3} \\
&\leq \int \eta(x_{i3}, x_{j3}, x_{i2}, x_{j2}, y) \eta(x_{i3}, x_{j2}, y') dx_{i2} dx_{j2} dx_{i3} dx_{j3} \\
&\leq \Delta_{x_i x_j}^2 \|\eta\|_\infty^2
\end{aligned}$$

using the orthonormality of the basis  $\alpha_{l_\alpha}$ . Then we get

$$\begin{aligned}
\sum_{l,l'} c_{ll'}^2 &\leq \Delta_{x_i x_j}^2 \|\eta\|_\infty^2 \left( \int \sum_{l_\beta, l'_\beta} \beta_{l_\beta}(y) \beta_{l'_\beta}(y) \beta_{l_\beta}(y') \beta_{l'_\beta}(y') dy dy' \right) \\
&= \Delta_{x_i x_j}^2 \|\eta\|_\infty^2 \sum_{l_\beta, l'_\beta} \left( \int \beta_{l_\beta}(y) \beta_{l'_\beta}(y) dy \right)^2 \\
&\leq \Delta_{x_i x_j}^2 \|\eta\|_\infty^2 \sum_{l_\beta} \left( \int \beta_{l_\beta}^2(y) dy \right)^2 \\
&\leq \Delta_{x_i x_j}^2 \|\eta\|_\infty^2 m
\end{aligned}$$

since the  $\beta_{l_\beta}$  are orthonormal. Finally

$$W_5 \leq \|f\|_\infty^2 \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 m.$$

Now for  $W_7$  we first will bound,

$$\begin{aligned}
\left| \sum_{l,l'} c_{ll'} a_l a_{l'} \right| &= \left| \int \sum_{l,l' \in M} a_l a_{l'} p_{l_2}(x_{i1}, x_{j1}, y) p_{l'_1}(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \right| \\
&\leq \int |S_M(x_{i1}, x_{j1}, y) S_M(x_{i2}, x_{j2}, y) \eta(x_{i1}, x_{j2}, y)| dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\
&\leq \|\eta\|_\infty \int \left( \int |S_M(x_{i1}, x_{j1}, y) S_M(x_{i2}, x_{j2}, y)| dy \right) dx_{i1} dx_{j1} dx_{i2} dx_{j2}.
\end{aligned}$$

Taking squares in both sides and using the Cauchy-Schwartz inequality twice, we get

$$\begin{aligned}
\left( \sum_{l,l'} c_{ll'} a_l a_{l'} \right)^2 &= \|\eta\|_\infty^2 \left( \int \left( \int |S_M(x_{i1}, x_{j1}, y) S_M(x_{i2}, x_{j2}, y)| dy \right) dx_{i1} dx_{j1} dx_{i2} dx_{j2} \right)^2 \\
&\leq \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 \int \left( \int |S_M(x_{i1}, x_{j1}, y) S_M(x_{i2}, x_{j2}, y)| dy \right)^2 dx_{i1} dx_{j1} dx_{i2} dx_{j2} \\
&\leq \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 \int \left( \int S_M(x_{i1}, x_{j1}, y)^2 dy \right) \left( \int S_M(x_{i2}, x_{j2}, y')^2 dy' \right) dx_{i1} dx_{j1} dx_{i2} dx_{j2} \\
&= \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 \int S_M(x_{i1}, x_{j1}, y)^2 S_M(x_{i1}, x_{j1}, y')^2 dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy dy' \\
&= \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 \left( \int S_M(x_i, x_j, y)^2 dx_i dx_j dy \right) \\
&\leq \|\eta\|_\infty^2 \Delta_{x_i x_j}^2 \|f\|_\infty^2.
\end{aligned}$$

Finally,

$$\mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})CQ(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))^2] \leq \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 (m+1). \quad (26)$$

Collecting (25) and (26), we obtain

$$\text{Var}(U_n K) \leq \frac{20}{n(n-1)} \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2 (m+1)$$

which concludes the proof of Lemma 2.  $\square$

**Lemma 3** (Bound for  $\text{Var}(P_n L)$ ). *Under the assumptions of Theorem 3, we have*

$$\text{Var}(P_n L) \leq \frac{12}{n} \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2.$$

*Proof.* First note that given the independence of  $(X_i^{(k)}, X_j^{(k)}, Y^{(k)})$   $k = 1, \dots, n$  we have

$$\text{Var}(P_n L) = \frac{1}{n} \text{Var}(L(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))$$

we can write  $L(X_i^{(1)}, X_j^{(1)}, Y^{(1)})$  as

$$\begin{aligned} & A^\top R \left( X_i^{(1)}, X_j^{(1)}, Y^{(1)} \right) + B^\top Q \left( X_i^{(1)}, X_j^{(1)}, Y^{(1)} \right) - 2A^\top C Q \left( X_i^{(1)}, X_j^{(1)}, Y^{(1)} \right) \\ &= \sum_{l \in M} a_l \left( \int p_l(x_i, x_j, Y^{(1)}) \psi(x_i, x_j, X_i^{(1)}, X_j^{(1)}, Y^{(1)}) dx_i dx_j - b_l \right) \\ & \quad + \sum_{l \in M} b_l \left( p_l(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - a_l \right) - 2 \sum_{l, l' \in M} c_{ll'} a_{l'} \left( p_l(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - a_l \right) \\ &= \int \sum_{l \in M} a_l p_l(x_i, x_j, Y^{(1)}) \psi(x_i, x_j, X_i^{(1)}, X_j^{(1)}, Y^{(1)}) dx_i dx_j \\ & \quad + \sum_{l \in M} b_l p_l(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - 2 \sum_{l, l' \in M} c_{ll'} a_{l'} p_l(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - 2A^\top B - 2A^\top C A. \\ &= \int S_M f(x_i, x_j, Y^{(1)}) \psi(x_i, x_j, X_i^{(1)}, X_j^{(1)}, Y^{(1)}) dx_i dx_j + S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) \\ & \quad - 2 \sum_{l, l' \in M} c_{ll'} a_{l'} p_l(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - 2A^\top B - 2A^\top C A. \end{aligned}$$

Let  $h(x_i, x_j, y) = \int S_M f(x_{i2}, x_{j2}, y) \psi(x_i, x_j, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2}$ , we have

$$\begin{aligned} & S_M h(x_i, x_j, y) \\ &= \sum_{l \in M} \left( \int h(x_{i2}, x_{j2}, y) p_l(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} dy \right) p_l(x_i, x_j, y) \\ &= \sum_{l \in M} \left( \int S_M f(x_{i3}, x_{j3}, y) \psi(x_{i2}, x_{j2}, x_{i3}, x_{j3}, y) p_l(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy \right) p_l(x_i, x_j, y) \\ &= \sum_{l, l' \in M} \left( \int a_{l'} p_{l'}(x_{i3}, x_{j3}, y) \psi(x_{i2}, x_{j2}, x_{i3}, x_{j3}, y) p_l(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy \right) p_l(x_i, x_j, y) \\ &= 2 \sum_{l, l' \in M} \left( \int a_{l'} p_{l'}(x_{i3}, x_{j3}, y) \eta(x_{i2}, x_{j3}, y) p_l(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} dx_{i3} dx_{j3} dy \right) p_l(x_i, x_j, y) \\ &= 2 \sum_{l, l' \in M} a_{l'} c_{ll'} p_l(x_i, x_j, y) \end{aligned}$$

and we can write

$$L(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) = h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - S_M h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - 2A^\top B - 2A^\top C A.$$

Thus,

$$\begin{aligned} & \text{Var}(L(X_i^{(1)}, X_j^{(1)}, Y^{(1)})) \\ &= \text{Var}(h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M h(X_i^{(1)}, X_j^{(1)}, Y^{(1)})) \\ &\leq \mathbb{E}[(h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2] \\ &\leq \mathbb{E}[(h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2 + (S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2 + (S_M h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2]. \end{aligned}$$

Each of these terms can be bounded

$$\begin{aligned} & \mathbb{E}[(h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2] \\ &= \int \left( \int S_M f(x_{i2}, x_{j2}, y) \psi(x_{i1} x_{j2}, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} \right)^2 f(x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dy \\ &\leq \Delta_{x_i x_j} \int S_M f(x_{i2}, x_{j2}, y)^2 \psi(x_{i1} x_{j2}, x_{i2}, x_{j2}, y)^2 f(x_{i1}, x_{j1}, y) dx_{i1} dx_{j1} dx_{i2} dx_{j2} dy \\ &\leq 4\Delta_{x_i x_j}^2 \|f\|_\infty \|\eta\|_\infty^2 \int S_M f(x_i, x_j, y)^2 dx_i dx_j dy \\ &= 4\Delta_{x_i x_j}^2 \|f\|_\infty \|\eta\|_\infty^2 \|S_M f\|_2^2 \\ &\leq 4\Delta_{x_i x_j}^2 \|f\|_\infty \|\eta\|_\infty^2 \|f\|_2^2 \\ &\leq 4\Delta_{x_i x_j}^2 \|f\|_\infty^2 \|\eta\|_\infty^2 \end{aligned}$$

and similar calculations are valid for the others two terms,

$$\begin{aligned} \mathbb{E}[(S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2] &\leq \|f\|_\infty \|S_M g\|_2^2 \leq \|f\|_\infty \|g\|_2^2 \leq 4\Delta_{x_i x_j}^2 \|f\|_\infty \|\eta\|_\infty^2 \\ \mathbb{E}[(S_M h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))^2] &\leq \|f\|_\infty \|S_M h\|_2^2 \leq \|f\|_\infty \|h\|_2^2 \leq 4\Delta_{x_i x_j}^2 \|f\|_\infty \|\eta\|_\infty^2. \end{aligned}$$

Finally we get,

$$\text{Var}(P_n L) \leq \frac{12}{n} \|\eta\|_\infty^2 \|f\|_\infty^2 \Delta_{x_i x_j}^2.$$

□

**Lemma 4** (Computation of  $\text{Cov}(U_n K, P_n L)$ ). *Under the assumptions of Theorem 3, we have*

$$\text{Cov}(U_n K, P_n L) = 0.$$

**Proof of Lemma 4.** Since  $U_n K$  and  $P_n L$  are centered, we have

$$\begin{aligned}
& \text{Cov}(U_n K, P_n L) \\
&= \mathbb{E}[U_n K P_n L] \\
&= \mathbb{E}\left[\frac{1}{n^2(n-1)} \sum_{k \neq k'=1}^n K(X_i^{(k)}, X_j^{(k)}, Y^{(k)}, X_i^{(k')}, X_j^{(k')}, Y^{(k')}) \sum_{k=1}^n L(X_i^{(k)}, X_j^{(k)}, Y^{(k)})\right] \\
&= \frac{1}{n} \mathbb{E}[K(X_i^{(1)}, X_j^{(1)}, Y^{(1)}, X_i^{(2)}, X_j^{(2)}, Y^{(2)}) (L(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + L(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))] \\
&= \frac{1}{n} \mathbb{E}[(Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})R(X_i^{(2)}, X_j^{(2)}, Y^{(2)}) - Q^\top(X_i^{(1)}, X_j^{(1)}, Y^{(1)})CQ(X_i^{(2)}, X_j^{(2)}, Y^{(2)})) \\
&\quad (A^\top R(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + B^\top Q(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) - 2A^\top CQ(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) \\
&\quad + A^\top R(X_i^{(2)}, X_j^{(2)}, Y^{(2)}) + B^\top Q(X_i^{(2)}, X_j^{(2)}, Y^{(2)}) - 2A^\top CQ(X_i^{(2)}, X_j^{(2)}, Y^{(2)}))] \\
&= 0.
\end{aligned}$$

Since  $K$ ,  $L$ ,  $Q$  and  $R$  are centered. □

**Lemma 5** (Bound of Bias  $(\hat{\theta}_n)$ ). *Under the assumptions of Theorem 3, we have*

$$|\text{Bias}(\hat{\theta}_n)| \leq \Delta_{x_i x_j} \|\eta\|_\infty \sup_{l \notin M} |c_l|^2.$$

*Proof.*

$$\begin{aligned}
|\text{Bias}(\hat{\theta}_n)| &\leq \|\eta\|_\infty \int \left( \int |S_M f(x_{i1}, x_{j1}, y) - f(x_{i1}, x_{j1}, y)| dx_{i1} dx_{j1} \right) \\
&\quad \left( \int |S_M f(x_{i2}, x_{j2}, y) - f(x_{i2}, x_{j2}, y)| dx_{i2} dx_{j2} \right) dy \\
&= \|\eta\|_\infty \int \left( \int |S_M f(x_i, x_j, y) - f(x_i, x_j, y)| dx_i dx_j \right)^2 dy \\
&\leq \Delta_{x_i x_j} \|\eta\|_\infty \int (S_M f(x_i, x_j, y) - f(x_i, x_j, y))^2 dx_i dx_j dy \\
&= \Delta_{x_i x_j} \|\eta\|_\infty \sum_{l, l' \notin M} a_l a_{l'} \int p_l(x_i, x_j, y) p_{l'}(x_i, x_j, y) dx_i dx_j dy \\
&= \Delta_{x_i x_j} \|\eta\|_\infty \sum_{l \notin M} |a_l|^2 \leq \Delta_{x_i x_j} \|\eta\|_\infty \sup_{l \notin M} |c_l|^2.
\end{aligned}$$

We use the Hölder's inequality and the fact that  $f \in \mathcal{E}$  then  $\sum_{l \notin M} |a_l|^2 \leq \sup_{l \notin M} |c_l|^2$ . □

**Lemma 6** (Asymptotic variance of  $\sqrt{n}(P_n L)$ ). *Under the assumptions of Theorem 3, we have*

$$n \text{Var}(P_n L) \rightarrow \Lambda(f, \eta)$$

where

$$\Lambda(f, \eta) = \int g(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int g(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2.$$

*Proof.* We proved in Lemma 3 that

$$\begin{aligned}
& \text{Var}(L(X_i^{(1)}, X_j^{(1)}, Y^{(1)})) \\
&= \text{Var}(h(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M g(X_i^{(1)}, X_j^{(1)}, Y^{(1)}) + S_M h(X_i^{(1)}, X_j^{(1)}, Y^{(1)})) \\
&= \text{Var}(A_1 + A_2 + A_3) \\
&= \sum_{k,l=1}^3 \text{Cov}(A_k, A_l).
\end{aligned}$$

We claim that  $\forall k, l \in \{1, 2, 3\}^2$ , we have

$$\left| \text{Cov}(A_k, A_l) - \epsilon_{kl} \left[ \int g(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int g(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 \right] \right| \leq \lambda [\|S_M f - f\|_2 + \|S_M g - g\|_2] \quad (27)$$

where

$$\epsilon_{kl} = \begin{cases} -1 & \text{if } k = 3 \text{ or } l = 3 \text{ and } k \neq l \\ 1 & \text{otherwise} \end{cases},$$

and where  $\lambda$  depends only on  $\|f\|_\infty$ ,  $\|\eta\|_\infty$  and  $\Delta_{x_i x_j}$ . We will do the details only for the case  $k = l = 3$  since the calculations are similar for others configurations.

$$\text{Var}(A_3) = \int S_M^2 h(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy - \left( \int S_M h(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2.$$

The computation will be done in two steps. We first bound the quantity by the Cauchy-Schwartz inequality

$$\begin{aligned}
& \left| \int S_M^2 h(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy - \int g(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy \right| \\
& \leq \int |S_M^2 h(x_i, x_j, y) f(x_i, x_j, y) - S_M^2 g(x_i, x_j, y) f(x_i, x_j, y)| dx_i dx_j dy \\
& \quad + \int |S_M^2 g(x_i, x_j, y) f(x_i, x_j, y) - g(x_i, x_j, y)^2 f(x_i, x_j, y)| dx_i dx_j dy \\
& \leq \|f\|_\infty \|S_M h + S_M g\|_2 \|S_M h - S_M g\|_2 + \|f\|_\infty \|S_M g + g\|_2 \|S_M g - g\|_2.
\end{aligned}$$

Using several times the fact that since  $S_M$  is a projection,  $\|S_M g\|_2 \leq \|g\|_2$ , the sum is bounded by

$$\begin{aligned}
& \|f\|_\infty \|h + g\|_2 \|h - g\|_2 + 2\|f\|_\infty \|g\|_2 \|S_M g - g\|_2 \\
& \leq \|f\|_\infty (\|h\|_2 + \|g\|_2) \|h - g\|_2 + 2\|f\|_\infty \|g\|_2 \|S_M g - g\|_2.
\end{aligned}$$

We saw previously that  $\|g\|_2 \leq 2\Delta_{x_i x_j} \|f\|_\infty^{1/2} \|\eta\|_\infty$  and  $\|h\|_2 \leq 2\Delta_{x_i x_j} \|f\|_\infty^{1/2} \|\eta\|_\infty$ . The sum is then bound by

$$4\Delta_{x_i x_j} \|f\|_\infty^{3/2} \|\eta\|_\infty \|h - g\|_2 + 4\Delta_{x_i x_j} \|f\|_\infty^{3/2} \|\eta\|_\infty \|S_M g - g\|_2.$$

We now have to deal with  $\|h - g\|_2$ :

$$\begin{aligned}
& \|h - g\|_2^2 \\
&= \int \left( \int (S_M f(x_{i2}, x_{j2}, y) - f(x_{i2}, x_{j2}, y)) \psi(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} \right)^2 dx_{i1} dx_{j1} dy \\
&\leq \int \left( \int (S_M f(x_{i2}, x_{j2}, y) - f(x_{i2}, x_{j2}, y))^2 dx_{i2} dx_{j2} \right) \left( \int \psi^2(x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} \right) dx_{i1} dx_{j1} dy \\
&\leq 4\Delta_{x_i x_j}^2 \|\eta\|_\infty^2 \|S_M f - f\|_2^2.
\end{aligned}$$

Finally this first part is bounded by

$$\begin{aligned}
& \left| \int S_M^2 h(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy - \int g(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy \right| \\
&\leq 4\Delta_{x_i x_j} \|f\|_\infty^{3/2} \|\eta\|_\infty \left( 2\Delta_{x_i x_j} \|\eta\|_\infty \|S_M f - f\|_2 + \|S_M g - g\|_2 \right).
\end{aligned}$$

Following with the second quantity

$$\begin{aligned}
& \left| \left( \int S_M h(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 - \left( \int g(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 \right| \\
&= \left| \left( \int (S_M h(x_i, x_j, y) - g(x_i, x_j, y)) f(x_i, x_j, y) dx_i dx_j dy \right) \right. \\
&\quad \left. \left( \int (S_M h(x_i, x_j, y) + g(x_i, x_j, y)) f(x_i, x_j, y) dx_i dx_j dy \right) \right|.
\end{aligned}$$

By using the Cauchy-Schwartz inequality, it is bounded by

$$\begin{aligned}
& \|f\|_2 \|S_M h - g\|_2 \|f\|_2 \|S_M h + g\|_2 \\
&\leq \|f\|_2^2 (\|h\|_2 + \|g\|_2) (\|S_M h - S_M g\|_2 + \|S_M g - g\|_2) \\
&\leq 4\Delta_{x_i x_j} \|f\|_\infty^{3/2} \|\eta\|_\infty (\|h - g\|_2 + \|S_M g - g\|_2) \\
&\leq 4\Delta_{x_i x_j} \|f\|_\infty^{3/2} \|\eta\|_\infty \left( 2\Delta_{x_i x_j} \|\eta\|_\infty \|S_M f - f\|_2 + \|S_M g - g\|_2 \right)
\end{aligned}$$

using the previous calculations. Collecting the two inequalities gives (27) for  $k = l = 3$ . Finally, since by assumption  $\forall t \in \mathbb{L}^2(d\mu)$ ,  $\|S_M t - t\|_2 \rightarrow 0$  when  $n \rightarrow \infty$  a direct consequence of (27) is

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \text{Var}(L(X_i^{(1)}, X_j^{(1)}, Y^{(1)})) \\
&= \int g^2(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy - \left( \int g(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 \\
&= \Lambda(f, \eta).
\end{aligned}$$

We conclude by noting that  $\text{Var}(\sqrt{n}(P_n L)) = \text{Var}(L(X_i^{(1)}, X_j^{(1)}, Y^{(1)}))$ .  $\square$

**Lemma 7** (Asymptotics for  $\sqrt{n}(\hat{Q} - Q)$ ). Under the assumptions of Theorem 1, we have

$$\lim_{n \rightarrow \infty} n\mathbb{E}[\hat{Q} - Q]^2 = 0.$$

*Proof.* The bound given in (16) states that if  $|M_n|/n \rightarrow 0$  we have

$$\begin{aligned} & \left| n\mathbb{E}[(\hat{Q} - Q)^2 | \hat{f}] - \left[ \int \hat{g}(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int \hat{g}(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 \right] \right| \\ & \leq \gamma(\|f\|_\infty, \|\eta\|_\infty, \Delta_{x_i x_j}) \left[ \frac{|M_n|}{n} + \|S_M f - f\|_2 + \|S_M \hat{g} - \hat{g}\|_2 \right] \end{aligned}$$

where  $\hat{g}(x_i, x_j, y) = \int H_3(\hat{f}, x_i, x_j, x_{i2}, x_{j2}, y) f(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2}$ , where we recall that  $H_3(f, x_{i1}, x_{j1}, x_{i2}, x_{j2}, y) = H_2(f, x_{i1}, x_{j2}, y) + H_2(f, x_{i2}, x_{j1}, y)$  with  $H_2(\hat{f}, x_{i1}, x_{j2}, y) = \frac{1}{\int \hat{f}(x_i, x_j, y) dx_i dx_j} (x_{i1} - m_i(\hat{f}, y))(x_{j2} - m_j(\hat{f}, y))$ . By deconditioning we get

$$\begin{aligned} & \left| n\mathbb{E}[(\hat{Q} - Q)^2] - \mathbb{E} \left[ \int \hat{g}(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int \hat{g}(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2 \right] \right| \\ & \leq \gamma(\|f\|_\infty, \|\eta\|_\infty, \Delta_{x_i x_j}) \left[ \frac{|M_n|}{n} + \|S_M f - f\|_2 + \mathbb{E}[\|S_M \hat{g} - \hat{g}\|_2] \right] \end{aligned}$$

Note that

$$\begin{aligned} \mathbb{E}[\|S_{M_n} \hat{g} - \hat{g}\|_2] & \leq \mathbb{E}[\|S_M \hat{g} - S_M g\|_2] + \mathbb{E}[\|\hat{g} - g\|_2] + \mathbb{E}[\|S_{M_n} g - g\|_2] \\ & \leq 2\mathbb{E}[\|\hat{g} - g\|_2] + \mathbb{E}[\|S_{M_n} g - g\|_2] \end{aligned}$$

where  $g(x_i, x_j, y) = \int H_3(f, x_i, x_j, x_{i2}, x_{j2}, y) f(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2}$ . The second term converges to 0 since  $g \in \mathbb{L}^2(dx dy dz)$  and  $\forall t \in \mathbb{L}^2(dx dy dz)$ ,  $\int (S_M t - t)^2 dx dy dz \rightarrow 0$ . Moreover

$$\begin{aligned} \|\hat{g} - g\|_2^2 & = \int [\hat{g}(x_i, x_j, y) - g(x_i, x_j, y)]^2 dx_i dx_j dy \\ & = \int \left[ \int \left( H_3(\hat{f}, x_i, x_j, x_{i2}, x_{j2}, y) - H_3(f, x_i, x_j, x_{i2}, x_{j2}, y) \right) f(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} \right]^2 dx_i dx_j dy \\ & \leq \int \left[ \int \left( H_3(\hat{f}, x_i, x_j, x_{i2}, x_{j2}, y) - H_3(f, x_i, x_j, x_{i2}, x_{j2}, y) \right)^2 dx_{i2} dx_{j2} \right] \\ & \quad \left[ \int f(x_{i2}, x_{j2}, y)^2 dx_{i2} dx_{j2} \right] dx_i dx_j dy \\ & \leq \Delta_{x_i x_j} \|f\|_\infty^2 \int \left( H_2(\hat{f}, x_i, x_j, x_{i2}, x_{j2}, y) - H_2(f, x_i, x_j, x_{i2}, x_{j2}, y) \right)^2 dx_i dx_j dx_{i2} dx_{j2} dy \\ & \leq \delta \Delta_{x_i x_j}^2 \|f\|_\infty^2 \int \left( \hat{f}(x_i, x_j, y) - f(x_i, x_j, y) \right)^2 dx_i dx_j dy \end{aligned}$$

for some constant  $\delta$  that comes out of applying the mean value theorem to  $H_3(\hat{f}, x_i, x_j, x_{i2}, x_{j2}, y) - H_3(f, x_i, x_j, x_{i2}, x_{j2}, y)$ . The constant  $\delta$  was taken under Assumptions 1-3. Since  $\mathbb{E}[\|f - \hat{f}\|_2] \rightarrow 0$  then  $\mathbb{E}[\|g - \hat{g}\|_2] \rightarrow 0$ . Now show that the expectation of

$$\int \hat{g}(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \left( \int \hat{g}(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j dy \right)^2$$

converges to 0. We develop the proof for only the first term. We get

$$\begin{aligned}
& \left| \int \hat{g}(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy - \int g(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j dy \right| \\
& \leq \int |\hat{g}(x_i, x_j, y)^2 - g(x_i, x_j, y)^2| f(x_i, x_j, y) dx_i dx_j dy \\
& \leq \lambda \int (\hat{g}(x_i, x_j, y) - g(x_i, x_j, y))^2 dx_i dx_j dy \\
& = \lambda \|\hat{g} - g\|_2^2
\end{aligned}$$

for some constant  $\lambda$ . By taking the expectation of both sides, we see it is enough to show that  $\mathbb{E}[\|\hat{g} - g\|_2^2] \rightarrow 0$ . Besides, we can verify

$$\begin{aligned}
g(x_i, x_j, y) &= \int H_3(f, x_i, x_j, x_{i2}, x_{j2}, y) f(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} \\
&= \frac{2}{\int f(x_i, x_j, y) dx_i dx_j} (x_i - \hat{m}_i(y)) \\
&\quad \left( \int x_{j2} f(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} - \hat{m}_j(y) \int f(x_{i2}, x_{j2}, y) dx_{i2} dx_{j2} \right) \\
&= 0
\end{aligned}$$

which proves that the expectation of  $\int \hat{g}(x_i, x_j, y)^2 f(x_i, x_j, y) dx_i dx_j$  converges to 0. Similar computations shows that the expectation of  $(\int \hat{g}(x_i, x_j, y) f(x_i, x_j, y) dx_i dx_j)^2$  also converges to 0. Finally we have

$$\lim_{n \rightarrow \infty} n \mathbb{E}[\hat{Q} - Q]^2 = 0.$$

□

## References

- Brillinger, D. R. (1983). A generalized linear model with " gaussian" regressor variables. *A Festschrift for Erich L. Lehmann in Honor of His Sixty-Fifth Birthday*, (pp. 97–114).
- Bura, E. & Cook, R. D. (2001). Estimating the structural dimension of regressions via parametric inverse regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 393–410.
- Cook, R. D. & Ni, L. (2005). Sufficient dimension reduction via inverse regression. *Journal of the American Statistical Association*, 100(470), 410–428.
- Da Veiga, S. & Gamboa, F. (2008). Efficient estimation of nonlinear conditional functionals of a density. *Submitted in The Annals of Statistics*. 42 pages 62G20, 62G06, 62G07, 62P30.

- Duan, N. & Li, K. C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, (pp. 505–530).
- Ferré, L. & Yao, A. F. (2003). Functional sliced inverse regression analysis. *Statistics*, 37(6), 475–488.
- Ferré, L. & Yao, A.-F. (2005). Smoothed functional inverse regression. *Statist. Sinica*, 15(3), 665–683.
- Hardle, W. & Tsybakov, A. B. (1991). Sliced inverse regression for dimension reduction: Comment. *Journal of the American Statistical Association*, 86(414), pp. 333–335.
- Hastie, T. J. & Tibshirani, R. J. (1990). *Generalized Additive Models*. Monographs on Statistics and Applied Probability ;; 43. Chapman & Hall/CRC, 1st ed. edition.
- Hsing, T. (1999). Nearest neighbor inverse regression. *The Annals of Statistics*, 27(2), 697–731.
- Ibragimov, I. A. & Khas'minskii, R. Z. (1991). Asymptotically normal families of distributions and efficient estimation. *The Annals of Statistics*, (pp. 1681–1724).
- Laurent, B. (1996). Efficient estimation of integral functionals of a density. *The Annals of Statistics*, 24(2), 659–681.
- Li, K. C. (1991a). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414), 316–327.
- Li, K. C. (1991b). Sliced inverse regression for dimension reduction: Rejoinder. *Journal of the American Statistical Association*, 86(414), pp. 337–342.
- Li, L. (2007). Sparse sufficient dimension reduction. *Biometrika*, 94(3), 603.
- Setodji, C. M. & Cook, R. D. (2004). K-means inverse regression. *Technometrics*, 46(4), 421–429.
- Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge Series on Statistical and Probabilistic Mathematics. Cambridge University Press.
- Zhu, L. X. & Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics*, 24(3), 1053–1068.