# Asymptotic confidence interval for R2 in multiple linear regression

J Dedecker, Odelia Guedj, Marie-Luce Taupin

# Asymptotic confidence interval for $R^2$ in multiple linear regression

J. Dedecker[1], O. Guedj[2], and M.L. Taupin[2]

[1]Laboratoire MAP5, UMR CNRS 8145, Université Paris Cité, France
[2]Laboratoire LaMME, UMR CNRS 8071, Université d'Evry Val d'Essonne, Université Paris Saclay, France

January 19, 2024

**Abstract**

Following White's approach of robust multiple linear regression [15], we give asymptotic confidence intervals for the multiple correlation coefficient $R^2$ under minimal moment conditions. We also give the asymptotic joint distribution of the empirical estimators of the individual $R^2$'s. Through different sets of simulations, we show that the procedure is indeed robust (contrary to the procedure involving the near exact distribution of the empirical estimator of $R^2$ is the multivariate Gaussian case) and can be also applied to count linear regression.

**Mathematics Subject Classification 2020 :** 62H20, 62J05
**Key Words :** Multiple correlation coefficient, Asymptotic distribution, Robustness, Heteroscedasticity

## 1 Introduction

Let $(Y, X^{(1)}, \dots X^{(p)})$ be a random vector with value in $\mathbb{R}^{p+1}$. The variable $Y$ is the response variable, and $(X^{(1)}, \dots, X^{(p)})$ is the vector of explanatory variables. Assume that all the variables are square integrable. The multiple correlation between $Y$ and $(X^{(1)}, \dots, X^{(p)})$ can be expressed using the orthogonal projection of $Y$ onto the $\mathbb{L}^2$-subspace $V$ generated by the variables $(X^{(0)}, X^{(1)}, \dots, X^{(p)})$, where we denote by $X^{(0)}$ the constant variable $X^{(0)} \equiv 1$. Let then $\mathrm{Proj}_V(Y)$ be the orthogonal projection of $Y$ onto $V$. The response variable $Y$ is uncorrelated to the vector $(X^{(1)}, \dots, X^{(p)})$ if and only if $\mathrm{Proj}_V(Y) = \mathbb{E}(Y)$.

To measure the strength of the correlation between $Y$ and $(X^{(1)}, \dots, X^{(p)})$, the usual measure is the coefficient $R^2$ defined by (assuming that $\mathrm{Var}(Y) > 0$):

$$R^2 = \frac{\mathrm{Var}(\mathrm{Proj}_V(Y))}{\mathrm{Var}(Y)}\,.$$

The coefficient $R^2$ is often referred as "the population variance-accounted-for effect size". As we see, it is defined as the proportion of the variance of $Y$ that is explained by the best linear predictor in $\mathbb{L}^2$ based on the variables $\mathbf{1}, X^{(1)}, \dots, X^{(p)}$.

In this paper, we consider the empirical estimator $\widehat{R}^2$ of the coefficient $R^2$ based on $n$ independent copies $(Y_i, X_i^{(1)}, \ldots, X_i^{(p)})_{1\leq i\leq n}$ of the vector $(Y, X^{(1)}, \ldots, X^{(p)})$. We prove its consistency, asymptotic normality, and we give an asymptotic confidence interval for $R^2$. Our approach is a robust one: we will make no additional assumptions on the distribution of the vector $(Y, X^{(1)}, \ldots, X^{(p)})$, except moment assumptions necessary for the Central Limit Theorem (CLT) to hold, and to estimate consistently the limiting variance.

Most of the existing literature on the distribution of $\widehat{R}^2$ deals with the case where the vector $(Y, X^{(1)}, \ldots X^{(p)})$ is normally distributed. In this context, Fisher (1928) [3] obtained the first exact expression of the density function of $\widehat{R}^2$, involving a Gauss hypergeometric series. Other exact expressions for the distribution and the cumulative distribution were given by Lee [8]. An exact expression of the cumulative distribution in terms of a series of Gamma distributions is given in Muirhead (1982, Theorem 5.2.5) [10]. The first two principal terms of the latter development were given earlier by Lee [8]. Some quantile tables were derived by Lee [9] from this second-order approximation of the cumulative distribution function. Lee's results were then implemented by Kelley [6] in his `R` MBESS package.

In the non-Gaussian case, there are relatively few results. Let us cite Muirhead [10] Theorem 5.1.6, who gave the asymptotic normality of $\widehat{R}^2$ assuming that $(Y, X^{(1)}, \ldots X^{(p)})$ has an elliptical distribution. The paper closest to ours is that of Ogasawara [11], who gave an Edgeworth expansion for $\sqrt{n}(\widehat{R}^2 - R^2)$, under some conditions on the distributions of the variables, assuming in particular that all the variables $Y, X^{(1)}, \ldots X^{(p)}$ have a moment of order 8. We will compare our result to that of Ogasawara in more details in Remark 3.1 of Section 3. Let us also mention the results of Ogasawara (2008) [12] for partial correlation.

The article is organised as follows: in Section 2, we recall White's results [15] concerning the least squares estimator of the coefficients of the orthogonal projection $\mathrm{Proj}_V(Y) = \alpha_0 + \alpha_1 X^{(1)} + \cdots + \alpha_p X^{(p)}$ (consistency, asymptotic normality and estimation of the limiting covariance matrix). This serves two purposes: it recalls the robust procedure described by White [15] in the linear model (without the assumption of normality or homoscedasticity), and it gives an initial idea of the proofs that will enable us to obtain an asymptotic confidence interval for $R^2$. In Section 3 we prove the consistency and asymptocic normality of $\widehat{R}^2$. We also give an estimator of the limiting variance, which together with asymptotic normality provide an asymptotic confidence interval for $R^2$. In Section 4, we give the asymptotic distribution of the joint law of the estimators of the individual $R^2$'s. In Section 5, we evaluate the performance of our confidence interval on different sets of simulations, and we compare this performance with that of the confidence interval that assumes that the vector $(Y, X^{(1)}, \ldots X^{(p)})$ is Gaussian. Finally, in Section 6, we give the extension of our procedure to the case of vector-valued random variables, and we briefly discuss the connections between our results and sensitivity analysis or screening methods.

## 2 Previous known results

Let $V_i$ be the $\mathbb{L}^2$-subspace generated by the variables $(X_i^{(0)}, X_i^{(1)}, \ldots, X_i^{(p)})$. Let $\varepsilon = Y - \mathrm{Proj}_V(Y)$ and $\varepsilon_i = Y_i - \mathrm{Proj}_{V_i}(Y_i)$. We can then write

$$Y_i = \mathrm{Proj}_{V_i}(Y_i) + \varepsilon_i.$$

where the variables $\varepsilon_i$ are such that $\mathbb{E}(X_i^{(j)}\varepsilon_i) = 0$ for any $j \in \{0, \ldots, p\}$.

Consider the following assumption:

**Assumption 2.1.** *Assume that* $\mathbf{1}, X^{(1)}, \dots, X^{(p)}$ *are linearly independent (as a family of random variables in* $\mathbb{L}^2$*).*

Under Assumption 2.1 one can uniquely write

$$Y_i = \alpha_0 + \alpha_1 X_i^{(1)} + \cdots + \alpha_p X_i^{(p)} + \varepsilon_i \,.$$

Let $\boldsymbol{\alpha}_{0:p} = (\alpha_0, \alpha_1, \dots, \alpha_p)^t \in \mathbb{R}^{p+1}$, and $\mathbf{X}$ be the matrix whose $j$th column is

$$\mathbf{X}^{(j)} = (X_1^{(j)}, \dots, X_n^{(j)})^t.$$

Let $\mathbf{Y} = (Y_1, \dots, Y_n)^t$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^t$. One can write

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha}_{0:p} + \boldsymbol{\varepsilon}.$$

Let $\mathbf{M}$ be the matrix such that $\mathbf{M}_{j,\ell} = \mathbb{E}(X^{(j)}X^{(\ell)})$. By Assumption (2.1) the matrix $\mathbf{M}$ is invertible and one can easily show that

$$\boldsymbol{\alpha}_{0:p} = \mathbf{M}^{-1}(\mathbb{E}(Y), \mathbb{E}(X^{(1)}Y) \dots, \mathbb{E}(X^{(p)}Y))^t \,.$$

## 2.1 Least square estimator of $\boldsymbol{\alpha}_{0:p}$

Let

$$\widehat{\mathbf{M}} = \frac{1}{n}\,\mathbf{X}^t\,\mathbf{X}\ .$$

By the strong law of large numbers

$$\lim_{n\to\infty} \widehat{\mathbf{M}} = \mathbf{M} \text{ almost surely.} \tag{2.1}$$

Under Assumption 2.1, $\mathbf{M}$ is invertible and for large enough $n$, $\mathbf{X}^t\,\mathbf{X}$ is also invertible, and one can define

$$\widehat{\boldsymbol{\alpha}}_{0:p} = (\widehat{\alpha}_0, \dots, \widehat{\alpha}_p)^t = (\mathbf{X}^t\,\mathbf{X})^{-1}\,\mathbf{X}^t\,\mathbf{Y}.$$

Let $\mathbb{V}_{p+1}$ be the sub-space of $\mathbb{R}^n$ generated by the $p+1$ columns $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(p)}$ of the matrix $\mathbf{X}$, and $\Pi_{\mathbb{V}_{p+1}}(\mathbf{Y})$ be the orthogonal projection of $\mathbf{Y}$ on $\mathbb{V}_{p+1}$ (with respect to the euclidean norm). We have

$$\widehat{\alpha}_0\,\mathbf{X}^{(0)} + \widehat{\alpha}_1\,\mathbf{X}^{(1)} + \cdots + \widehat{\alpha}_p\,\mathbf{X}^{(p)} = \Pi_{\mathbb{V}_{p+1}}(\mathbf{Y}),$$

and

$$\widehat{\boldsymbol{\alpha}}_{0:p} - \boldsymbol{\alpha}_{0:p} = (\widehat{\alpha}_0, \dots, \widehat{\alpha}_p)^t - (\alpha_0, \dots, \alpha_p)^t = (\mathbf{X}^t\,\mathbf{X})^{-1}\,\mathbf{X}^t\,\boldsymbol{\varepsilon} = \widehat{\mathbf{M}}^{-1}\frac{1}{n}\,\mathbf{X}^t\,\boldsymbol{\varepsilon}\,. \tag{2.2}$$

## 2.2 Consistency and asymptotic normality of $\widehat{\boldsymbol{\alpha}}_{0:p}$

Following White [15], we obtain the consistency and asymptotic normality of the least square estimators. The proofs of these results being simple and enlightening, we have chosen to recall them because they may make it easier to understand the proofs of Section 3.

**Proposition 2.1.** *Under Assumption 2.1,* $\widehat{\boldsymbol{\alpha}}_{0:p}$ *converges almost surely to* $\boldsymbol{\alpha}_{0:p}$ *as* $n \to \infty$*.*

**Proof of Proposition 2.1.** From (2.1), we get that

$$\lim_{n\to\infty} \widehat{\mathbf{M}}^{-1} = \mathbf{M}^{-1} \text{ almost surely.} \tag{2.3}$$

Starting from (2.2) and using (2.3), it suffices to show that

$$\lim_{n\to\infty} \frac{1}{n} \mathbf{X}^t \boldsymbol{\varepsilon} = 0 \text{ almost surely.} \tag{2.4}$$

By definition of $\varepsilon$, for any $j \in \{0, \ldots, p\}$, $\mathbb{E}(X^{(j)}\varepsilon) = 0$. Applying the strong law of large numbers to the $j$th coordinate of $n^{-1}\,\mathbf{X}^t\,\boldsymbol{\varepsilon}$, we get

$$\lim_{n\to\infty} \left(\frac{1}{n} \mathbf{X}^t \boldsymbol{\varepsilon}\right)_j = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i^{(j)} \varepsilon_i = 0 \text{ almost surely.} \qquad \square \tag{2.5}$$

To prove the asymptotic normality, we need moment assumptions.

**Assumption 2.2.** *Assume that* $\mathbb{E}({X^{(j)}}^2 \varepsilon^2) < \infty$ *for all* $j = 0, \ldots, p$.

**Proposition 2.2.** *Under Assumptions 2.1 and 2.2, let* $\mathbf{M}(\varepsilon)$ *be the matrix defined by*

$$\mathbf{M}(\varepsilon)_{j,\ell} = \mathbb{E}(X^{(j)} X^{(\ell)} \varepsilon^2) \quad \textit{for any } 0 \le j, \ell \le p. \tag{2.6}$$

*Then*

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}}_{0:p} - \boldsymbol{\alpha}_{0:p}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{p+1}(0, \mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1}).$$

**Remark 2.1.** *Assume that* $\varepsilon^2 - \mathbb{E}(\varepsilon^2)$ *is orthogonal to the space* $S$ *generated by the variables* $X^{(j)}X^{(\ell)}$ *for all* $0 \le j \le \ell \le p$ *(which is true, for instance, if* $\varepsilon$ *is independent of* $(X^{(1)}, \ldots, X^{(p)})$*). Then the limiting variance matrix in Proposition 2.2 writes*

$$\mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1} = \mathbb{E}(\varepsilon^2)\mathbf{M}^{-1}\,,$$

*and is the same as in the case where the conditional distribution of* $Y$ *given* $(X^{(1)}, \ldots, X^{(p)})$ *is Gaussian with* $\mathrm{Var}(Y|X^{(1)}, \ldots, X^{(p)}) = \sigma^2$ *(homoscedastic case; in that case* $\mathbb{E}(\varepsilon^2) = \sigma^2$*). This simple observation enabled White [15] to formulate his famous homoscedasticity test, which in fact consists of testing whether* $\varepsilon^2 - \mathbb{E}(\varepsilon^2)$ *is orthogonal to* $S$ *(*$H_0$ *hypothesis) or not.*

**Proof of Proposition 2.2.** Starting from (2.2) and using (2.3), it suffices to prove that

$$\frac{1}{\sqrt{n}} \mathbf{X}^t \boldsymbol{\varepsilon} \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{p+1}(0, \mathbf{M}(\varepsilon)). \tag{2.7}$$

Now (2.7) follows from a direct application of the central limit theorem in $\mathbb{R}^{p+1}$, since for any $0 \le j \le p$, $\mathbb{E}(X^{(j)}\varepsilon) = 0$ and since Assumption 2.2 holds.

$\square$

## 2.3 Estimation of the limiting covariance matrix

Under Assumption 2.2, by the strong law of large numbers, for any $j, \ell \in \{0, \ldots, p\}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i^{(j)} X_i^{(\ell)} \varepsilon_i^2 = \mathbf{M}(\varepsilon)_{j,\ell} \quad \text{almost surely.} \tag{2.8}$$

Of course, the quantity on right hand is not an estimator of $\mathbf{M}(\varepsilon)_{j,\ell}$ since the $\varepsilon_i$'s are not observed. White [15] proposed then to replace the $\varepsilon_i$'s by the residuals.

$$\widehat{\varepsilon}_i = Y_i - (\mathbf{X}\widehat{\boldsymbol{\alpha}}_{0:p})_i = Y_i - \widehat{\alpha}_0 X_i^{(0)} - \cdots - \widehat{\alpha}_p X_i^{(p)}.$$

Let then

$$\widehat{\mathbf{M}(\varepsilon)}_{j,\ell} = \frac{1}{n} \sum_{i=1}^{n} X_i^{(j)} X_i^{(\ell)} \widehat{\varepsilon}_i^2 \,.$$

To prove the consistency of $\widehat{\mathbf{M}(\varepsilon)}_{j,\ell}$, we shall assume that Assumption 2.2 holds and that:

**Assumption 2.3.** $\mathbb{E}({X^{(j)}}^4) < \infty$ *for all* $j \in \{0, \ldots, p\}$.

**Proposition 2.3.** *Under Assumptions 2.2 and 2.3, for any* $j, \ell \in \{0, \ldots, p\}$,

$$\lim_{n\to\infty} \widehat{\mathbf{M}(\varepsilon)}_{j,\ell} = \mathbf{M}(\varepsilon)_{j,\ell} \quad \textit{almost surely.}$$

*Consequently, under Assumptions 2.1-2.3,*

$$\lim_{n\to\infty} \widehat{\mathbf{M}}^{-1} \widehat{\mathbf{M}(\varepsilon)} \widehat{\mathbf{M}}^{-1} = \mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1} \quad \textit{almost surely.}$$

**Remark 2.2.** *The Assumptions 2.2 and 2.3 are equivalent to:* $\mathbb{E}({X^{(j)}}^2 Y^2) < \infty$ *and* $\mathbb{E}({X^{(j)}}^4) < \infty$ *for any* $j \in \{0, \ldots, p\}$.

**Remark 2.3.** *As usual, Propositions 2.2 and 2.3 enable to obtain confidence regions or test procedures for the coefficients* $\alpha_i$. *For instance, if one wants to test*

$$H_0 : \alpha_{i_1} = \cdots = \alpha_{i_k} = 0 \quad \textit{for some } 0 \le i_1 < i_2 < \cdots < i_k \le p$$

*one can proceed as follows. Let* $\mathbf{C}_k$ *be the matrix with* $k$ *rows and* $p$ *columns, such that all coordinates of the* $j$*th row are 0 except the coordinate* $i_j$ *which is equal to 1. Then, by Proposition 2.2,*

$$\sqrt{n}((\widehat{\alpha}_{i_1}, \ldots, \widehat{\alpha}_{i_k})^t - (\alpha_{i_1}, \ldots, \alpha_{i_k})^t) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_k(0, \mathbf{C}_k\mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1}\mathbf{C}_k^t).$$

*Now, under Assumptions 2.1-2.3, by applying Proposition 2.3*

$$\widehat{\mathbf{V}}_k = \mathbf{C}_k \widehat{\mathbf{M}}^{-1} \widehat{\mathbf{M}(\varepsilon)} \widehat{\mathbf{M}}^{-1} \mathbf{C}_k^t \quad \textit{converges almost surely to} \quad \mathbf{C}_k\mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1}\mathbf{C}_k^t \,.$$

*If* $\mathbf{M}(\varepsilon)$ *is invertible, then the matrix* $\mathbf{C}_k\mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1}\mathbf{C}_k^t$ *is also invertible and* $\widehat{\mathbf{V}}_k^{-1/2}$ *is well defined (for* $n$ *large enough). Let then* $(\xi_1, \ldots, \xi_k)^t = \sqrt{n}\widehat{\mathbf{V}}_k^{-1/2}(\widehat{\alpha}_{i_1}, \ldots, \widehat{\alpha}_{i_k})^t$. *Under* $H_0$, *we see that the test statistic* $\sum_{i=1}^k \xi_i^2$ *is such that*

$$\sum_{i=1}^{k} \xi_i^2 \xrightarrow[n\to\infty]{\mathcal{L}} \chi^2(k) \,.$$

**Proof of Proposition 2.3.** Write $\widehat{\varepsilon}_i^2 = \varepsilon_i^2 + R_i$, with $R_i = 2\varepsilon_i(\widehat{\varepsilon}_i - \varepsilon_i) + (\widehat{\varepsilon}_i - \varepsilon_i)^2$. The following upper bound holds

$$|R_i| \leq 2\sum_{j=0}^{p} |(\widehat{\alpha}_j - \alpha_j) X_i^{(j)} \varepsilon_i| + (p+1)\sum_{j=0}^{p} (\widehat{\alpha}_j - \alpha_j)^2 {X_i^{(j)}}^2 .$$

Tacking into account (2.8), to prove Proposition 2.3, it suffices to show that, for $1 \leq j, k \leq p$ and $1 \leq i \leq n$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |X_i^{(j)} X_i^{(k)} R_i| = 0 \quad \text{almost surely.} \tag{2.9}$$

Since $|\widehat{\alpha}_i - \alpha_i|$ converges almost surely to 0, (2.9) will be satisfied provided, for $1 \leq j, k, \ell \leq p$,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |X_i^{(j)} X_i^{(k)} X_i^{(\ell)} \varepsilon_i| < \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |X_i^{(j)} X_i^{(k)} {X_i^{(\ell)}}^2| < \infty \quad \text{almost surely,}$$

which is true by applying once again the strong law of large numbers under Assumptions 2.2 and 2.3. □

# 3 Confidence interval for $R^2$

Assuming that $\mathrm{Var}(Y) > 0$, the mutliple correlation coefficient $R^2$ is defined by:

$$R^2 = \frac{\mathrm{Var}(\mathrm{Proj}_V(Y))}{\mathrm{Var}(Y)} .$$

For $\varepsilon_i = Y_i - \mathrm{Proj}_{V_i}(Y_i)$, recall that, under Assumption 2.1,

$$Y_i = \mathrm{Proj}_{V_i}(Y_i) + \varepsilon_i = \alpha_0 \mathbf{1} + \alpha_1 X_i^{(1)} + \cdots + \alpha_p X_i^{(p)} + \varepsilon_i.$$

For any vector $\mathbf{z} = (z_1, \ldots, z_n)^t$ in $\mathbb{R}^n$, the empirical variance of $\mathbf{z}$ is given by

$$\mathrm{Var}_n(\mathbf{z}) = \frac{1}{n} \sum_{i=1}^{n} (z_i - \overline{z})^2.$$

Let $\mathbf{1}_n = (1, \ldots, 1)^t$ and let $\Pi_E$ be the orthogonal projection (with respect to the the euclidean norm) on a subspace $E$ of $\mathbb{R}^n$. Let $\mathbb{V}_{p+1}$ be the subspace of $\mathbb{R}^n$ generated by $(\mathbf{1}_n, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(p)})$. The empirical estimator $\widehat{R}^2$ of $R^2$ is then

$$\widehat{R}^2 = \frac{\mathrm{Var}_n(\Pi_{\mathbb{V}_{p+1}}(\mathbf{Y}))}{\mathrm{Var}_n(\mathbf{Y})} . \tag{3.1}$$

Our first goal is to give a simple expression of this estimator. Let $\boldsymbol{\alpha}_{1:p} = (\alpha_1, \ldots, \alpha_p)^t$ and $\widehat{\boldsymbol{\alpha}}_{1:p} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p)^t$; for the sake of simplicity, we shall omit the indexes and write $\boldsymbol{\alpha} = \boldsymbol{\alpha}_{1:p}$ and $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}_{1:p}$. Let also $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^t$ and $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_p)^t$, where

$$\theta_k = \frac{\mathrm{Cov}(Y, X^{(k)})}{\mathrm{Var}(Y)} \quad \text{and} \quad \widehat{\theta}_k = \frac{\sum_{i=1}^{n} (Y_i - \overline{\mathbf{Y}})(X_i^{(k)} - \overline{\mathbf{X}}^{(k)})}{\sum_{i=1}^{n} (Y_i - \overline{\mathbf{Y}})^2} . \tag{3.2}$$

The following lemma gives the expression of $R^2$ and $\widehat{R}^2$ in terms of $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\alpha}}$.

**Lemma 3.1.** *Under Assumption 2.1,* $R^2 = \boldsymbol{\theta}^t\boldsymbol{\alpha}$ *and* $\widehat{R}^2 = \widehat{\boldsymbol{\theta}}^t\widehat{\boldsymbol{\alpha}}$.

**Proof of Lemma 3.1.** We prove the second point, the first point being proved in the same way. Write

$$\Pi_{\mathbb{V}_{p+1}}(\mathbf{Y}) - \overline{\Pi_{\mathbb{V}_{p+1}}}(\mathbf{Y})\mathbf{1}_n = \widehat{\alpha}_1(\mathbf{X}^{(1)} - \overline{\mathbf{X}}^{(1)}\mathbf{1}_n) + \cdots + \widehat{\alpha}_p(\mathbf{X}^{(p)} - \overline{\mathbf{X}}^{(p)}\mathbf{1}_n)\,.$$

Now, if $S_p$ is the subspace of $\mathbb{R}^n$ generated by $((\mathbf{X}^{(1)} - \overline{\mathbf{X}}^{(1)}\mathbf{1}_n)), \ldots, (\mathbf{X}^{(p)} - \overline{\mathbf{X}}^{(p)}\mathbf{1}_n)))$ it is easy to check that

$$\widehat{\alpha}_1(\mathbf{X}^{(1)} - \overline{\mathbf{X}}^{(1)}\mathbf{1}_n)) + \cdots + \widehat{\alpha}_p(\mathbf{X}^{(p)} - \overline{\mathbf{X}}^{(p)}\mathbf{1}_n)) = \Pi_{S_p}(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n).$$

Since by Assumption 2.1 the variables $\mathbf{1}, X^{(1)}, \ldots, X^{(p)}$ are linearly independent in $\mathbb{L}^2$, the variance matrix $\mathbf{M}_0$ of $(X^{(1)}, \ldots, X^{(p)})^t$ defined by

$$(\mathbf{M}_0)_{j,\ell} = \mathrm{Cov}(X^{(j)}, X^{(\ell)}) \quad j, \ell \in \{1, \ldots, p\} \tag{3.3}$$

is invertible. Denote by $\mathbb{X}_0$ the matrix defined by

$$\mathbb{X}_0 = ((X_i^{(j)} - \overline{\mathbf{X}}^{(j)}))_{1 \le i \le n, 1 \le j \le p}\,. \tag{3.4}$$

By the strong law of large numbers

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{X}_0^t\mathbb{X}_0 = \mathbf{M}_0 \quad \text{almost surely.}$$

Since $\mathbf{M}_0$ defined in (3.3) is invertible, it follows that $\mathbb{X}_0^t\mathbb{X}_0$ is also invertible for $n$ large enough, in such a way that

$$\widehat{\boldsymbol{\alpha}} = (\mathbb{X}_0^t\mathbb{X}_0)^{-1}\mathbb{X}_0^t(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n), \tag{3.5}$$

and

$$\Pi_{S_p}((\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)) = \mathbb{X}_0(\mathbb{X}_0^t\mathbb{X}_0)^{-1}\mathbb{X}_0^t(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n).$$

Let $\|\cdot\|_{\mathrm{eu}}$ be the euclidean norm on $\mathbb{R}^n$. The following equality hold:

$$\begin{aligned} n\mathrm{Var}_n(\Pi_{\mathbb{V}_{p+1}}(\mathbf{Y})) &= \left\|\Pi_{S_p}(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)\right\|_{\mathrm{eu}}^2 \\ &= (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t\mathbb{X}_0(\mathbb{X}_0^t\mathbb{X}_0)^{-1}\mathbb{X}_0^t\mathbb{X}_0(\mathbb{X}_0^t\mathbb{X}_0)^{-1}\mathbb{X}_0^t(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n) \\ &= (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t\mathbb{X}_0(\mathbb{X}_0^t\mathbb{X}_0)^{-1}\mathbb{X}_0^t(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n) \\ &= (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t\mathbb{X}_0\widehat{\boldsymbol{\alpha}}\,. \end{aligned} \tag{3.6}$$

The result follows from (3.1), (3.6) and (3.7) since

$$\frac{(\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t\mathbb{X}_0}{n\mathrm{Var}_n(\mathbf{Y})} = \widehat{\boldsymbol{\theta}}^t\,. \tag{3.7}$$

□

Now, to build a confidence interval for $R^2$, we need to describe the asymptotic distribution of

$$\sqrt{n}\Big([\widehat{\boldsymbol{\alpha}} : \widehat{\boldsymbol{\theta}}] - [\boldsymbol{\alpha} : \boldsymbol{\theta}]\Big) := \sqrt{n}\left((\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_p)^t - (\alpha_1, \ldots, \alpha_p, \theta_1, \ldots, \theta_p)^t\right)\,.$$

Let us then define the two matrices involved in this asymptotic distribution. For $k \in \{1, \dots, p\}$, let

$$e^{(k)} = X^{(k)} - \mathrm{Proj}_W(X^{(k)}) \quad \text{and} \quad e_i^{(k)} = X_i^{(k)} - \mathrm{Proj}_{W_i}(X_i^{(k)}), \tag{3.8}$$

where $W$ is the subspace of $\mathbb{L}^2$ generated by $\mathbf{1}$ and $Y$, and $W_i$ is the subspace of $\mathbb{L}^2$ generated by $\mathbf{1}$ and $Y_i$. Let also $\boldsymbol{e}^{(k)} = (e_1^{(k)}, \dots, e_n^{(k)})^t$.

**Assumption 3.1.** $\mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))^2 \varepsilon^2) < \infty$ *and* $\mathbb{E}((Y - \mathbb{E}(Y))^2 {e^{(j)}}^2) < \infty$ *for all* $j \in \{1, \dots, p\}$.

Let then $\mathbf{A}$ be the $2p \times 2p$ symmetric matrix defined as follows:

- If $(j,k) \in \{1, \dots, p\}^2$ then
$$\mathbf{A}_{j,k} = \mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))(X^{(k)} - \mathbb{E}(X^{(k)}))\varepsilon^2);$$
- If $(j,k) \in \{p+1, \dots, 2p\}^2$ then
$$\mathbf{A}_{j,k} = \mathbb{E}((Y - \mathbb{E}(Y))^2 e^{(j-p)} e^{(k-p)});$$
- $(j,k) \in \{1, \dots, p\} \times \{p+1, \dots, 2p\}$ then
$$\mathbf{A}_{j,k} = \mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))(Y - \mathbb{E}(Y))\varepsilon e^{(k-p)}).$$

Let $\delta_{j,k} = 0$ if $j \neq k$ and $\delta_{j,j} = 1$ and let $\mathbf{B}$ be the $2p \times 2p$ symetric matrix defined as follows:

- If $(j,k) \in \{1, \dots, p\}^2$ then $\mathbf{B}_{j,k} = (\mathbf{M}_0^{-1})_{j,k}$ (see (3.3) for the definition of $\mathbf{M}_0$);
- If $(j,k) \in \{p+1, \dots, 2p\}^2$ then $\mathbf{B}_{j,k} = (\mathrm{Var}(Y))^{-1}\delta_{j,k}$;
- If $(j,k) \in \{1, \dots, p\} \times \{p+1, \dots, 2p\}$ then $\mathbf{B}_{j,k} = 0$.

**Proposition 3.1.** *Under Assumptions 2.1 and 3.1*

$$\sqrt{n}\Big([\widehat{\boldsymbol{\alpha}} : \widehat{\boldsymbol{\theta}}] - [\boldsymbol{\alpha} : \boldsymbol{\theta}]\Big) \quad \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{2p}(0, \mathbf{BAB}).$$

**Proof of Proposition 3.1.** Let $\widehat{\mathbf{B}}$ be the $2p \times 2p$ symetric matrix defined as follows:

- If $(j,k) \in \{1, \dots, p\}^2$ then $\widehat{\mathbf{B}}_{j,k} = n(\mathbb{X}_0^t \mathbb{X}_0)^{-1}_{j,k}$;
- If $(j,k) \in \{p+1, \dots, 2p\}^2$ then $\widehat{\mathbf{B}}_{j,k} = (\mathrm{Var}_n(\mathbf{Y}))^{-1}\delta_{j,k}$;
- If $(j,k) \in \{1, \dots, p\} \times \{p+1, \dots, 2p\}$ then $\widehat{\mathbf{B}}_{j,k} = 0$.

Starting from (3.2) and (3.5), and noting that

$$\begin{aligned} Y_i - \overline{\mathbf{Y}} &= \alpha_1(X_i^{(1)} - \overline{\mathbf{X}}^{(1)}) + \cdots + \alpha_p(X_i^{(p)} - \overline{\mathbf{X}}^{(p)}) + (\varepsilon_i - \overline{\varepsilon}), \\ (X_i^{(k)} - \overline{\mathbf{X}}^{(k)}) &= \theta_k(Y_i - \overline{\mathbf{Y}}) + (e_i^{(k)} - \overline{\boldsymbol{e}}^{(k)}) \quad \text{for } k \in \{1, \dots, p\}, \end{aligned}$$

we see that

$$\sqrt{n}\left([\widehat{\boldsymbol{\alpha}}:\widehat{\boldsymbol{\theta}}]-[\boldsymbol{\alpha}:\boldsymbol{\theta}]\right)=\widehat{\mathbf{B}}\frac{1}{\sqrt{n}}(\boldsymbol{\varepsilon}^t\mathbb{X}_0,(\mathbf{Y}-\overline{\mathbf{Y}}\mathbf{1}_n)^t\mathbf{E})^t\,, \tag{3.9}$$

where $\mathbf{E}$ is the $n\times p$ matrix such that, for $(i,j)\in\{1,\dots,n\}\times\{1,\dots p\}$, $(\mathbf{E}_{i,j})=e_i^{(j)}$.

Denote by $\widetilde{\mathbb{X}}_0$ the matrix defined by

$$\tilde{\mathbb{X}}_0=(X_i^{(j)}-\mathbb{E}(X^{(j)}))_{1\le i\le n,1\le j\le p}\,. \tag{3.10}$$

One can easily check that

$$\frac{1}{\sqrt{n}}\left((\boldsymbol{\varepsilon}^t\mathbb{X}_0,(Y_1-\overline{\mathbf{Y}},\dots,Y_n-\overline{\mathbf{Y}})\mathbf{E})^t-(\boldsymbol{\varepsilon}^t\tilde{\mathbb{X}}_0,(Y_1-\mathbb{E}(Y),\dots,Y_n-\mathbb{E}(Y))\mathbf{E})^t\right)\xrightarrow[n\to\infty]{\mathbb{P}}0\,. \tag{3.11}$$

Now, by the multivariate central limit theorem,

$$\frac{1}{\sqrt{n}}(\boldsymbol{\varepsilon}^t\tilde{\mathbb{X}}_0,(Y_1-\mathbb{E}(Y),\dots,Y_n-\mathbb{E}(Y))\mathbf{E})^t\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}_{2p}(0,\mathbf{A}). \tag{3.12}$$

The result follows from (3.9), (3.11) and (3.12), since by the strong law of large numbers,

$$\lim_{n\to\infty}\widehat{\mathbf{B}}=\mathbf{B}\quad\text{almost surely.} \tag{3.13}$$

□

As a consequence, we get

**Proposition 3.2.** *Under Assumptions 2.1 and 3.1*

$$\sqrt{n}\left(\widehat{R}^2-R^2\right)\xrightarrow[n\to\infty]{\mathcal{L}}\mathcal{N}(0,V).$$

*where*

$$V=(\theta_1,\dots,\theta_p,\alpha_1,\dots,\alpha_p)\mathbf{BAB}(\theta_1,\dots,\theta_p,\alpha_1,\dots,\alpha_p)^t\,. \tag{3.14}$$

**Remark 3.1.** *Ogasawara [11] has given an Edgeworth expansion of $\sqrt{n}(\widehat{R}^2-R^2)$, which is a more precise result than Proposition 3.2, but requires some conditions on the distribution of $(Y,X^{(1)},\cdots,X^{(p)})$. In particular, he required that the all the variables have a moment of order 8. Note that, if we do not look for an Edgeworth expansion, the method of Ogasawara consists in expressing $R^2$ as a differentiable function of $S=((\mathrm{Cov}(X^{(i)},X^{(j)}))_{1\le i\le j\le p},(\mathrm{Cov}(Y,X^{(i)}))_{1\le i\le p},\mathrm{Var}(Y))$, proving the CLT for $\sqrt{n}(\widehat{S}-S)$ ($\widehat{S}$ being the empirical estimator of $S$), and applying the delta method. The differences with our approach are the following :*

- *To prove the CLT for $\sqrt{n}(\widehat{S}-S)$, one needs moments of order 4 for all variables, which is a more restrictive condition than the moment conditions of our Proposition 3.2.*
- *Applying the delta method to a function of $S$ implies that the limiting variance $V$ will be expressed as a function of a $q\times q$ matrix, where $q=(p+1)(p+2)/2$ (the matrix $\boldsymbol{\Omega}$ in [11]), while we can express $V$ as a function of a $2p\times 2p$ matrix (the matrix $\mathbf{BAB}$, see (3.14)). Note that, with our expression of $V$ we are able to give simple sufficient conditions ensuring that $V>0$ (see Lemma 3.2 below).*

- *Finally our approach can be easily extended to the the case where the variables $Y, X^{(1)}, \cdots, X^{(p)}$ are vector-valued (see Section 6.1).*

**Proof of Proposition 3.2.** Starting from Lemma 3.1 and Proposition 3.1, it suffices to apply the delta-method to the function $\phi : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ defined by

$$\phi(x, y) = x^t y \,.$$

The proof Proposition 3.2 is complete by evaluating the differential $D\phi_{x,y}$ of $\phi$ at point $(x, y)$:

$$D\phi_{x,y}(h_1, h_2) = y^t h_1 + x^t h_2 = (y, x)^t (h_1, h_2) \,. \qquad \square$$

To build a confidence interval for $R^2$, it remains to find a consistent estimator of $V$. We shall simply replace each element in the definition of $V$ by its empirical counterpart.

Let then $\widehat{\mathbf{A}}$ be the $2p \times 2p$ symmetric matrix defined as follows:

- If $(j, k) \in \{1, \ldots, p\}^2$ then

$$\widehat{\mathbf{A}}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} (X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(X_i^{(k)} - \overline{\mathbf{X}}^{(k)}) \widehat{\varepsilon}_i^2;$$

- If $(j, k) \in \{p+1, \ldots, 2p\}^2$ then

$$\widehat{\mathbf{A}}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{\mathbf{Y}})^2 \widehat{e}_i^{(j-p)} \widehat{e}_i^{(k-p)},$$

  where $\widehat{e}_i^{(k-p)} = (X_i^{(k-p)} - \overline{\mathbf{X}}^{(k-p)}) - \widehat{\theta}_{k-p}(Y_i - \overline{\mathbf{Y}})$;

- If $(j, k) \in \{1, \ldots, p\} \times \{p+1, \ldots, 2p\}$ then

$$\widehat{\mathbf{A}}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} (X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(Y_i - \overline{\mathbf{Y}}) \widehat{\varepsilon}_i \widehat{e}_i^{(k-p)}.$$

**Proposition 3.3.** *Let*

$$\widehat{V}_n = (\widehat{\theta}_1, \ldots, \widehat{\theta}_p, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_p) \widehat{\mathbf{B}} \widehat{\mathbf{A}} \widehat{\mathbf{B}} (\widehat{\theta}_1, \ldots, \widehat{\theta}_p, \widehat{\alpha}_1, \ldots, \widehat{\alpha}_p)^t \,. \tag{3.15}$$

*Under Assumptions 2.1 and 2.3 and $\mathbb{E}(Y^4) < \infty$, we have*

$$\lim_{n \to \infty} \widehat{V}_n = V \quad \textit{almost surely.}$$

**Proof of Proposition 3.3.** Recall that, by the strong law of large numbers,

$$\lim_{n \to \infty} (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_p)^t = (\alpha_1, \ldots, \alpha_p, \theta_1, \ldots, \theta_p)^t \quad \text{almost surely,}$$

and

$$\lim_{n \to \infty} \widehat{\mathbf{B}} = \mathbf{B} \quad \text{almost surely.}$$

Hence, it remains to prove that $\widehat{\mathbf{A}}$ converges to $\mathbf{A}$ almost surely.

To prove this point, we first introduce the matrix $\widetilde{\mathbf{A}}$ defined as follows:

- If $(j,k) \in \{1,\dots,p\}^2$ then

$$\widetilde{\mathbf{A}}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(X_i^{(k)} - \overline{\mathbf{X}}^{(k)})\varepsilon_i^2;$$

- If $(j,k) \in \{p+1,\dots,2p\}^2$ then

$$\widetilde{\mathbf{A}}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{\mathbf{Y}})^2 e_i^{(j-p)} e_i^{(k-p)};$$

- If $(j,k) \in \{1,\dots,p\} \times \{p+1,\dots,2p\}$ then

$$\widetilde{\mathbf{A}}_{j,k} = \frac{1}{n}\sum_{i=1}^{n}(X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(Y_i - \overline{\mathbf{Y}})\varepsilon_i e_i^{(k-p)}.$$

Since $\mathbb{E}(Y^4) < \infty$ and $\mathbb{E}(X^{(j)^4}) < \infty$, we deduce from the strong law of large numbers that

$$\lim_{n\to\infty} \widetilde{\mathbf{A}} = \mathbf{A} \quad \text{almost surely.}$$

Let us now prove that, for $(j,k) \in \{1,\dots,p\}^2$,

$$\lim_{n\to\infty} |\widehat{\mathbf{A}}_{j,k} - \widetilde{\mathbf{A}}_{j,k}| = 0 \quad \text{almost surely.}$$

We write $\widehat{\varepsilon}_i^2 = \varepsilon_i^2 + R_i$, with $R_i = 2\varepsilon_i(\widehat{\varepsilon}_i - \varepsilon_i) + (\widehat{\varepsilon}_i - \varepsilon_i)^2$. One can easily see that

$$|R_i| \le 2\sum_{\ell=0}^{p} |(\widehat{\alpha}_\ell - \alpha_\ell) X_i^{(\ell)} \varepsilon_i| + (p+1)\sum_{\ell=0}^{p} (\widehat{\alpha}_\ell - \alpha_\ell)^2 {X_i^{(\ell)}}^2 . \tag{3.16}$$

Moreover

$$|\widehat{\mathbf{A}}_{j,k} - \widetilde{\mathbf{A}}_{j,k}| \le \frac{1}{n}\sum_{i=1}^{n} |(X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(X_i^{(k)} - \overline{\mathbf{X}}^{(k)})R_i| . \tag{3.17}$$

Combining (3.16) and (3.17), and using that $|\widehat{\alpha}_j - \alpha_j|$ converges almost surely to 0, we infer that $\widehat{\mathbf{A}}_{j,k} - \widetilde{\mathbf{A}}_{j,k}$ converges almost surely to 0 as soon as

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |X_i^{(j)} X_i^{(k)} X_i^{(\ell)} \varepsilon_i| < \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |X_i^{(j)} X_i^{(k)} {X_i^{(\ell)}}^2| < \infty \quad \text{almost surely,}$$

which is true (applying once again the strong law of large numbers) since $\mathbb{E}(Y^4) < \infty$ and $\mathbb{E}(X^{(j)^4}) < \infty$ for any $j \in \{1,\dots,p\}$.

Let us now prove that, for $(j,k) \in \{p+1,\dots,2p\}^2$, $\widehat{\mathbf{A}}_{j,k} - \widetilde{\mathbf{A}}_{j,k}$ converges almost surely to 0 as $n \to \infty$. We write $\widehat{e}_i^{(j-p)}\widehat{e}_i^{(k-p)} = e_i^{(j-p)}e_i^{(k-p)} + T_i$, with

$$|T_i| \le |e_i^{(j-p)}(\widehat{e}_i^{(k-p)} - e_i^{(k-p)})| + |e_i^{(k-p)}(\widehat{e}_i^{(j-p)} - e_i^{(j-p)})| + \frac{1}{2}(\widehat{e}_i^{(k-p)} - e_i^{(k-p)})^2 + \frac{1}{2}(\widehat{e}_i^{(j-p)} - e_i^{(j-p)})^2 . \tag{3.18}$$

For $\ell \in \{1, \dots, p\}$, let $\beta_\ell = \mathbb{E}(X_\ell) - \theta_\ell \mathbb{E}(Y)$ and $\widehat{\beta}_\ell = \overline{\mathbf{X}}^{(\ell)} - \widehat{\theta}_\ell \overline{\mathbf{Y}}$. With these notations

$$X_k^{(\ell)} = \beta_\ell + \theta_\ell Y_k + e_k^{(\ell)} \quad \text{and} \quad X_k^{(\ell)} = \widehat{\beta}_\ell + \widehat{\theta}_\ell Y_k + \widehat{e}_k^{(\ell)}. \tag{3.19}$$

From (3.18) and (3.19), we infer that

$$\begin{aligned} |T_i| \leq |e_i^{(j-p)}(\widehat{\beta}_{k-p} - \beta_{k-p})| + |e_i^{(j-p)} Y_i (\widehat{\theta}_{k-p} - \theta_{k-p})| \\ + |e_i^{(k-p)}(\widehat{\beta}_{j-p} - \beta_{j-p})| + |e_i^{(k-p)} Y_i (\widehat{\theta}_{j-p} - \theta_{j-p})| \\ + (\widehat{\beta}_{k-p} - \beta_{k-p})^2 + (\widehat{\beta}_{j-p} - \beta_{j-p})^2 \\ + (\widehat{\theta}_{k-p} - \theta_{k-p})^2 Y_i^2 + (\widehat{\theta}_{j-p} - \theta_{j-p})^2 Y_i^2 \,. \end{aligned} \tag{3.20}$$

Moreover

$$|\widehat{\mathbf{A}}_{j,k} - \widetilde{\mathbf{A}}_{j,k}| \leq \frac{1}{n} \sum_{i=1}^n |(Y_i - \overline{\mathbf{Y}})^2 T_i| \,. \tag{3.21}$$

Combining (3.20) and (3.21), and using that $|\widehat{\beta}_\ell - \beta_\ell|$ and $|\widehat{\theta}_\ell - \theta_\ell|$ converge almost surely to 0, we infer that $\widehat{A}_{j,k} - \tilde{A}_{j,k}$ converges almost surely to 0 as soon as, almost surely

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n |Y_i^2 e_i^{(j-p)}| < \infty, \ \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n |Y_i^3 e_i^{(j-p)}| < \infty, \text{ and } \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^n Y_i^4 < \infty,$$

which is true (applying once again the strong law of large numbers) since $\mathbb{E}(Y^4) < \infty$ and $\mathbb{E}({X^{(j)}}^4) < \infty$ for any $j \in \{1, \dots, p\}$.

The fact that, for $(j,k) \in \{1, \dots, p\} \times \{p+1, \dots, 2p\}$, $\widehat{\mathbf{A}}_{j,k} - \widetilde{\mathbf{A}}_{j,k}$ converges almost surely to 0 as $n \to \infty$ may be proved exactly in the same way. □

As an immediate Corollary, we get

**Corollary 3.1.** *Assume that Assumptions 2.1 and 2.3 hold, and that $\mathbb{E}(Y^4) < \infty$. Assume also that $V > 0$, where $V$ is defined by (3.14). For $\delta \in (0,1)$, let $c_{1-(\delta/2)}$ be the quantile of order $1-\delta$ of the $\mathcal{N}(0,1)$-distribution, and let $\widehat{V}_n$ be defined by (3.15). Then*

$$\left[ \widehat{R}^2 - \frac{c_{1-(\delta/2)} \sqrt{\widehat{V}_n}}{\sqrt{n}}, \widehat{R}^2 + \frac{c_{1-(\delta/2)} \sqrt{\widehat{V}_n}}{\sqrt{n}} \right]$$

*is an asymptotic confidence interval for $R^2$ of level $1-\delta$.*

**Remark 3.2.** *As is sensitivity analysis, one can also define*

$$R_{T_i}^2 = 1 - \frac{\mathrm{Var}(\mathrm{Proj}_{V_{(-i)}}(Y))}{\mathrm{Var}(Y)} \,,$$

*where $V_{(-i)}$ is the linear space of $\mathbb{L}^2$ generated by the variables $\mathbf{1}, X^{(1)}, \dots, X^{(i-1)}, X^{(i+1)}, \dots, X^{(p)}$. Then $R_{T_i}^2$ is the proportion of the variance of $Y$ that is not explained by the best linear predictor based on the*

*variables* $\mathbf{1}, X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(p)}$. *Denoting by* $R^2_{(-i)}$ *the proportion of the variance of* $Y$ *that is explained by the best linear predictor based on* $\mathbf{1}, X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(p)}$, *we immediately see that* $R^2_{T_i} = 1 - R^2_{(-i)}$. *Consequently, a confidence interval for* $R^2_{T_i}$ *is directly obtained from a confidence interval for* $R^2_{(-i)}$.

To be complete, we discuss the positivity of the variance term $V$.

**Lemma 3.2.** *Assume that* $\mathbf{M}_0$ *is invertible, that* $\mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)})^2\varepsilon^2) < \infty$ *and Assumption 3.1 holds.*

1. *If* $R^2 = 1$ *then* $\widehat{R}^2 = 1$ *almost surely, and* $V = 0$.
2. *If* $R^2 = 0$ *then* $V = 0$, *and* $n\widehat{R}^2$ *converges in distribution to* $(G_1, \ldots, G_p)(G_{p+1}, \ldots, G_{2p})^t$, *where* $(G_1, \ldots, G_{2p})$ *is the Gaussian random vector with covariance matrix* $\mathbf{BAB}$.
3. *If* $R^2 > 0$ *and if the family*

$$((Y-\mathbb{E}(Y))^2, ((Y-\mathbb{E}(Y))(X^{(j)}-\mathbb{E}(X^{(j)})))_{1\le j\le p}, ((X^{(j)}-\mathbb{E}(X^{(j)}))(X^{(j)}-\mathbb{E}(X^{(j)})))_{1\le i\le j\le p}) \quad (3.22)$$

*is linearly independent, then* $V > 0$.

**Remark 3.3.** *The fact that the family* (3.22) *is linearly independent implies that* $(Y, X^{(1)}, \ldots, X^{(p)})$ *is also linearly independent, so that* $R^2 < 1$.

**Remark 3.4.** *Note that one can easily check that the variable* $(G_1, \ldots, G_p)(G_{p+1}, \ldots, G_{2p})^t$ *of Item 2 is non negative. Indeed, in that particular case,* $(G_1, \ldots, G_p)^t = \mathbf{M}_0^{-1}(Z_1, \ldots, Z_p)^t$ *and* $(G_{p+1}, \ldots, G_{2p})^t = (\mathrm{Var}(Y))^{-1}(Z_1, \ldots, Z_p)^t$, *where* $(Z_1, \ldots, Z_p)^t$ *is a Gaussian vector with covariance matrix* $(\mathbf{A}_{i,j})_{1\le i,j\le p}$. *The asymptotic distribution of Item 2 may be used to test* $H_0 : R^2 = 0$ *against* $H_1 : R^2 \neq 0$, *but it is simpler to test the equivalent hypothesis* $H_0 : \alpha_1 = \cdots = \alpha_p = 0$ *against* $H_1 : \alpha_i \neq 0$ *for some* $i \in \{1, \ldots, p\}$ *(see Remark 2.3).*

**Proof of Lemma 3.2.** Item 1 is clear: $R^2 = 1$ if and only if $Y = \alpha_0 + \alpha_1 X^{(1)} + \cdots + \alpha_p X^{(p)}$ almost surely. In that case $\widehat{R}^2 = 1$ almost surely. Consequently $\widehat{R}^2 - R^2 = 0$ almost surely, and $V = 0$. We now prove Item 2. $R^2 = 0$ if and only if $Y - \mathbb{E}(Y)$ is orthogonal to the space generated by $(X^{(1)} - \mathbb{E}(X^{(1)}), \ldots, (X^{(p)} - \mathbb{E}(X^{(p)}))$. This is equivalent to $\alpha_1 = \cdots = \alpha_p = 0$, which is also equivalent to $\theta_1 = \cdots = \theta_p = 0$. This implies that $V = 0$. Applying Proposition 3.1, we see that

$$\sqrt{n}(\widehat{\alpha}_1, \ldots, \widehat{\alpha}_p, \widehat{\theta}_1, \ldots, \widehat{\theta}_p)^t \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{2p}(0, \mathbf{B}\,\mathbf{A}\,\mathbf{B}).$$

The last assertion of Item 2 follows by noting that

$$n\widehat{R}^2 = \left(\sqrt{n}(\widehat{\theta}_2, \ldots, \widehat{\theta}_p)\right)\left(\sqrt{n}(\widehat{\alpha}_2, \ldots, \widehat{\alpha}_p)\right)^t.$$

We now prove Item 3. Since $R^2 > 0$, it follows that the two vectors $(\alpha_1, \ldots, \alpha_p)^t$ and $(\theta_1, \ldots, \theta_p)^t$ are not equal to $(0, \ldots, 0)^t$. Let

$$(a_1, \ldots, a_p, b_1, \ldots, b_p)^t = \mathbf{B}(\theta_1, \ldots, \theta_p, \alpha_1, \ldots, \alpha_p)^t.$$

By definition of the matrix $\mathbf{B}$, and since $\mathbf{M}_0$ is invertible, we infer that the two vectors $(a_1,\dots,a_p)^t$ and $(b_1,\dots,b_p)^t$ are not equal to $(0,\dots,0)^t$.

Now, by definition, $V = (a_1,\dots,a_p,b_1,\dots,b_p)\,\mathbf{A}(a_1,\dots,a_p,b_1,\dots,b_p)^t$. Since $\mathbf{A}$ is the covariance matrix of the vector

$$\Big((X^{(1)}-\mathbb{E}(X^{(1)}))\varepsilon,\dots,(X^{(p)}-\mathbb{E}(X^{(p)}))\varepsilon,(Y-\mathbb{E}(Y))e^{(1)},\dots,(Y-\mathbb{E}(Y))e^{(p)}\Big)^t\,,$$

we can write

$$V=\mathrm{Var}\left(a_1(X^{(1)}-\mathbb{E}(X^{(1)}))\varepsilon+\cdots+a_p(X^{(p)}-\mathbb{E}(X^{(p)}))\varepsilon+b_1(Y-\mathbb{E}(Y))e^{(1)}+\cdots+b_p(Y-\mathbb{E}(Y))e^{(p)}\right).$$

Hence $V=0$ if and only if

$$a_1(X^{(1)}-\mathbb{E}(X^{(1)}))\varepsilon+\cdots+a_p(X^{(p)}-\mathbb{E}(X^{(p)}))\varepsilon=-(b_1(Y-\mathbb{E}(Y))e^{(1)}+\cdots+b_p(Y-\mathbb{E}(Y))e^{(p)})\quad\text{almost surely.}\tag{3.23}$$

Recall now that

$$\begin{aligned}\varepsilon&=(Y-\mathbb{E}(Y))-\alpha_1(X^{(1)}-\mathbb{E}(X^{(1)}))-\cdots-\alpha_p(X^{(p)}-\mathbb{E}(X^{(p)}))\\ e^{(k)}&=(X^{(k)}-\mathbb{E}(X^{(k)}))-\theta_k(Y-\mathbb{E}(Y))\quad\text{for } k\in\{1,\dots,p\}.\end{aligned}$$

Without loss of generality, assume that $\alpha_1\neq 0$. Then, on the left side of (3.23) the terms

$$\begin{aligned}-a_1\alpha_1(X^{(1)}-\mathbb{E}(X^{(1)}))^2,-a_2\alpha_1(X^{(1)}-\mathbb{E}(X^{(1)}))(X^{(2)}-\mathbb{E}(X^{(2)})),\\ \dots,-a_p\alpha_1(X^{(1)}-\mathbb{E}(X^{(1)}))(X^{(p)}-\mathbb{E}(X^{(p)}))\end{aligned}$$

appear, but they do not appear in the right side of (3.23). Since we assumed that the family (3.22) is linearly independent, we infer that $V=0$ implies $a_1=\cdots=a_p=0$. Since we know that $(a_1,\dots,a_p)^t$ is not equal to $(0,\dots,0)^t$, we conclude that $V>0$. □

# 4 Asymptotic joint distribution of individual $R^2$'s

For $k\in\{1,\dots,p\}$, let $U^{(k)}$ be the $\mathbb{L}^2$-subspace generated by $\mathbf{1}$ and $X^{(k)}$, and let (assuming that $\mathrm{Var}(Y)>0$)

$$R^2_{(k)}=\frac{\mathrm{Var}(\mathrm{Proj}_{U^{(k)}}(Y))}{\mathrm{Var}(Y)}\,.$$

From Lemma 3.1, we have $R^2_{(k)}=\tau_k\theta_k$, where $\theta_k$ is defined by (3.2) and

$$\tau_k=\frac{\mathrm{Cov}(Y,X^{(k)})}{\mathrm{Var}(X^{(k)})}.\tag{4.1}$$

The empirical estimator of $R^2_{(k)}$ is then

$$\widehat{R}^2_{(k)}=\widehat{\tau}_k\widehat{\theta}_k,\quad\text{where}\quad\hat{\tau}_k=\frac{\sum_{i=1}^n(Y_i-\overline{\mathbf{Y}})(X_i^{(k)}-\overline{\mathbf{X}}^{(k)})}{\sum_{i=1}^n(X_i^{(k)}-\overline{\mathbf{X}}^{(k)})^2}\,.\tag{4.2}$$

The asymptotic distribution of $\sqrt{n}(\widehat{R}^2_{(k)} - R^2_{(k)})$ is given in Proposition 3.2. In this section, we give the asymptotic joint distribution of

$$\sqrt{n}\left((\widehat{R}^2_{(1)} \ldots, \widehat{R}^2_{(p)})^t - (R^2_{(1)}, \ldots, R^2_{(p)})^t\right).$$

Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^t$ and $\widehat{\boldsymbol{\tau}} = (\widehat{\tau}_1, \ldots, \widehat{\tau}_p)^t$. The first step is to identify the limit distribution of

$$\sqrt{n}\left([\hat{\boldsymbol{\tau}} : \hat{\boldsymbol{\theta}}] - [\boldsymbol{\tau} : \boldsymbol{\theta}]\right).$$

Let us then define the two matrices involved in this asymptotic distribution. For $k \in \{1, \ldots, p\}$, let

$$\varepsilon^{(k)} = Y - \mathrm{Proj}_{U^{(k)}}(Y) \quad \text{and} \quad \varepsilon_i^{(k)} = Y_i - \mathrm{Proj}_{U_i^{(k)}}(Y_i),$$

where $U_i^{(k)}$ is the subspace of $\mathbb{L}^2$ generated by $\mathbf{1}$ and $X_i^{(k)}$. Let also $\boldsymbol{\varepsilon}^{(k)} = (\varepsilon_1^{(k)}, \ldots, \varepsilon_n^{(k)})^t$. Recall that $e^{(k)}$ and $e_i^{(k)}$ have been defined in (3.8).

**Assumption 4.1.** $\mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))^2 {\varepsilon^{(j)}}^2) < \infty$ *and* $\mathbb{E}((Y - \mathbb{E}(Y))^2 {e^{(j)}}^2) < \infty$ *for all* $j \in \{1, \ldots, p\}$.

Let then $\mathbf{C}$ be the $2p \times 2p$ symmetric matrix defined as follows:

- If $(j,k) \in \{1, \ldots, p\}^2$ then

$$\mathbf{C}_{j,k} = \mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))(X^{(k)} - \mathbb{E}(X^{(k)}))\varepsilon^{(j)}\varepsilon^{(k)}); \tag{4.3}$$

- If $(j,k) \in \{p+1, \ldots, 2p\}^2$ then

$$\mathbf{C}_{j,k} = \mathbb{E}((Y - \mathbb{E}(Y))^2 e^{(j-p)} e^{(k-p)});$$

- $(j,k) \in \{1, \ldots, p\} \times \{p+1, \ldots, 2p\}$ then

$$\mathbf{C}_{j,k} = \mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))(Y - \mathbb{E}(Y))\varepsilon^{(j)} e^{(k-p)}).$$

Let $\mathbf{D}$ be the $2p \times 2p$ diagonal matrix defined as follows :

- If $j \in \{1, \ldots, p\}$ then $\mathbf{D}_{j,j} = (\mathrm{Var}(X^{(j)}))^{-1}$.
- If $j \in \{p+1, \ldots, 2p\}$ then $\mathbf{D}_{j,j} = (\mathrm{Var}(Y))^{-1}$.

**Proposition 4.1.** *Under Assumption 4.1 and assuming that* $\mathrm{Var}(X^{(j)}) > 0$ *for all* $j \in \{1, \ldots, p\}$, *we have*

$$\sqrt{n}\Big([\widehat{\boldsymbol{\tau}} : \widehat{\boldsymbol{\theta}}] - [\boldsymbol{\tau} : \boldsymbol{\theta}]\Big) \quad \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{2p}(0, \mathbf{DCD}).$$

**Proof of Proposition 4.1.** Let $\widehat{\mathbf{D}}$ be the $2p \times 2p$ diagonal matrix defined as follows :

- If $j \in \{1, \ldots, p\}$ then $\widehat{\mathbf{D}}_{j,j} = (\mathrm{Var}_n(\mathbf{X}^{(j)}))^{-1}$.
- If $j \in \{p+1, \ldots, 2p\}$ then $\widehat{\mathbf{D}}_{j,j} = (\mathrm{Var}_n(\mathbf{Y}))^{-1}$.

Starting from (3.2) and (3.5), and noting that, for $k \in \{1, \dots, p\}$,

$$Y_i - \overline{\mathbf{Y}} = \tau_k (X_i^{(k)} - \overline{\mathbf{X}}^{(k)}) + (\varepsilon_i^{(k)} - \overline{\boldsymbol{\varepsilon}}^{(k)}),$$
$$(X_i^{(k)} - \overline{\mathbf{X}}^{(k)}) = \theta_k (Y_i - \overline{\mathbf{Y}}) + (e_i^{(k)} - \overline{\boldsymbol{e}}^{(k)}),$$

we see that (recall that $\mathbb{X}_0$ and $\tilde{\mathbb{X}}_0$ have been defined in (3.4) and (3.10))

$$\sqrt{n}\left([\widehat{\boldsymbol{\tau}} : \widehat{\boldsymbol{\theta}}] - [\boldsymbol{\tau} : \boldsymbol{\theta}]\right) = \widehat{\mathbf{D}} \frac{1}{\sqrt{n}} (\boldsymbol{\varepsilon}^{(1)t} \mathbb{X}_0^{(1)}, \dots, \boldsymbol{\varepsilon}^{(p)t} \mathbb{X}_0^{(p)}, (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t \boldsymbol{e}^{(1)}, \dots, (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t \boldsymbol{e}^{(p)})^t. \tag{4.4}$$

One can easily check that

$$\frac{1}{\sqrt{n}}\Big( (\boldsymbol{\varepsilon}^{(1)t} \mathbb{X}_0^{(1)}, \dots, \boldsymbol{\varepsilon}^{(p)t} \mathbb{X}_0^{(p)}, (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t \boldsymbol{e}^{(1)}, \dots, (\mathbf{Y} - \overline{\mathbf{Y}}\mathbf{1}_n)^t \boldsymbol{e}^{(p)})^t$$
$$- (\boldsymbol{\varepsilon}^{(1)t} \tilde{\mathbb{X}}_0^{(1)}, \dots, \boldsymbol{\varepsilon}^{(p)t} \tilde{\mathbb{X}}_0^{(p)}, (\mathbf{Y} - \mathbb{E}(Y)\mathbf{1}_n)^t \boldsymbol{e}^{(1)}, \dots, (\mathbf{Y} - \mathbb{E}(Y)\mathbf{1}_n)^t \boldsymbol{e}^{(p)})^t \Big) \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{4.5}$$

Now, by the multivariate central limit theorem,

$$\frac{1}{\sqrt{n}} (\boldsymbol{\varepsilon}^{(1)t} \tilde{\mathbb{X}}_0^{(1)}, \dots, \boldsymbol{\varepsilon}^{(p)t} \tilde{\mathbb{X}}_0^{(p)}, (\mathbf{Y} - \mathbb{E}(Y)\mathbf{1}_n)^t \boldsymbol{e}^{(1)}, \dots, (\mathbf{Y} - \mathbb{E}(Y)\mathbf{1}_n)^t \boldsymbol{e}^{(p)})^t \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_{2p}(0, \mathbf{C}). \tag{4.6}$$

Moreover, by the strong law of large numbers,

$$\lim_{n\to\infty} \widehat{\mathbf{D}} = \mathbf{D} \quad \text{almost surely} \tag{4.7}$$

The result follows from (4.4), (4.5), (4.6) and (4.7). □

Let $\mathbf{H}$ be the matrix with $p$ rows and $2p$ column such that

- If $(i, j) \in \{1, \dots, p\}^2$ then $\mathbf{H}_{i,j} = \theta_i \delta_{i,i}$.
- If $i \in \{1, \dots, p\}$ and $j \in \{p+1, \dots, 2p\}$ then $\mathbf{H}_{i,j} = \tau_i \delta_{i,p+i}$.

As a consequence of Proposition 4.1, we have

**Proposition 4.2.** *Under Assumption 4.1 and assuming that* $\mathrm{Var}(X^{(j)}) > 0$ *for all* $j \in \{1, \dots, p\}$, *we have*

$$\sqrt{n}\left( (\widehat{R}^2_{(1)} \dots, \widehat{R}^2_{(p)})^t - (R^2_{(1)}, \dots, R^2_{(p)})^t \right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, \mathbf{HDCDH}^t).$$

**Remark 4.1.** *Other quantities of interest are*

$$S^2_{(k)} = \frac{\mathrm{Var}(\mathrm{Proj}_{U^{(k)}}(Y))}{\mathrm{Var}(\mathrm{Proj}_V(Y))} \quad \textit{for } k \in \{1, \dots, p\},$$

*which are the proportions of the variance of* $\mathrm{Proj}_V(Y)$ *that are explained by the best linear predictors in* $\mathbb{L}^2$ *based on the variables* $\mathbf{1}, X^{(k)}$. *Note that*

$$S^2_{(k)} = \frac{R^2_{(k)}}{R^2} = \frac{\tau_k \theta_k}{\boldsymbol{\theta}^t \boldsymbol{\alpha}}.$$

*Then, the asymptotic distribution of* $\sqrt{n}\left( (\widehat{S}^2_{(1)} \dots, \widehat{S}^2_{(p)})^t - (S^2_{(1)}, \dots, S^2_{(p)})^t \right)$ *may be derived from the asymptotic distribution of* $\sqrt{n}\left( [\hat{\boldsymbol{\tau}} : \hat{\boldsymbol{\theta}} : \hat{\boldsymbol{\alpha}}] - [\boldsymbol{\tau} : \boldsymbol{\theta} : \boldsymbol{\alpha}] \right)$ *via the delta method.*

**Proof of Proposition 4.2.** Let $\phi : \mathbb{R}^{2p} \to \mathbb{R}^p$ be such that $\phi(x_1, \ldots, x_p, y_1, \ldots, y_p) = (x_1 y_1, \ldots, x_p y_p)^t$. Since

$$\left( (\widehat{R}^2_{(1)} \ \ldots, \widehat{R}^2_{(p)})^t - (R^2_{(1)}, \ldots, R^2_{(p)})^t \right) = \phi([\widehat{\boldsymbol{\tau}} : \widehat{\boldsymbol{\theta}}]) - \phi([\boldsymbol{\tau} : \boldsymbol{\theta}]) \,,$$

it suffices to apply the delta method to the function $\phi$. The results follows from Proposition 4.1 and the fact that the differential $D\phi_{\boldsymbol{\tau},\boldsymbol{\theta}}$ of $\phi$ at point $(\boldsymbol{\tau}, \boldsymbol{\theta})$ is given by

$$D\phi_{\boldsymbol{\tau},\boldsymbol{\theta}}(h) = \mathbf{H}h \,. \qquad \square$$

To conclude this section, let us a give a consistent estimator of the matrix $\mathbf{HDCDH}^t$. We shall simply replace each matrix by its empirical counterpart.

Let then $\widehat{\mathbf{C}}$ be the $2p \times 2p$ symmetric matrix defined as follows:

- If $(i, j) \in \{1, \ldots, p\}^2$ then

$$\widehat{\mathbf{C}}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} (X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(X_i^{(k)} - \overline{\mathbf{X}}^{(k)}) \widehat{\varepsilon}_i^{(j)} \widehat{\varepsilon}_i^{(k)},$$

  where $\widehat{\varepsilon}_i^{(k)} = (Y_i - \overline{\mathbf{Y}}) - \widehat{\tau}_k (X_i^{(k)} - \overline{\mathbf{X}}^{(k)})$;

- If $(j, k) \in \{p+1, \ldots, 2p\}^2$ then

$$\widehat{\mathbf{C}}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \overline{\mathbf{Y}})^2 \widehat{e}_i^{(j-p)} \widehat{e}_i^{(k-p)};$$

- If $(j, k) \in \{1, \ldots, p\} \times \{p+1, \ldots, 2p\}$ then

$$\widehat{\mathbf{C}}_{j,k} = \frac{1}{n} \sum_{i=1}^{n} (X_i^{(j)} - \overline{\mathbf{X}}^{(j)})(Y_i - \overline{\mathbf{Y}}) \widehat{\varepsilon}_i^{(j)} \widehat{e}_i^{(k-p)}.$$

**Proposition 4.3.** *Assume that* $\mathrm{Var}(X^{(j)}) > 0$, *that* $\mathbb{E}(Y^4) < \infty$ *and that* $\mathbb{E}({X^{(j)}}^4) < \infty$ *for any* $j \in \{1, \ldots, p\}$. *Then*

$$\lim_{n \to \infty} \widehat{\mathbf{H}} \widehat{\mathbf{D}} \widehat{\mathbf{C}} \widehat{\mathbf{D}} \widehat{\mathbf{H}}^t = \mathbf{HDCDH}^t \quad \textit{almost surely.}$$

**Proof of Proposition 4.3.** By the strong law of large numbers, $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{D}}$ converge almost surely to $\mathbf{H}$ and $\mathbf{D}$ respectively. Hence, it remains to prove that $\widehat{\mathbf{C}}$ converges to $\mathbf{C}$ almost surely, which can be done as in the proof of Proposition 3.3.

# 5 Simulations

We shall consider different models. For each model, we shall estimate the coverage level of two confidence intervals of level 95% for the $R^2$, for $n = 200$ to $n = 1000$, via a basic Monte-Carlo procedure (with $N = 3000$ repetitions).

| $n$ | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| CI1 | 0.942 | 0.945 | 0.952 | 0.949 | 0.951 | 0.947 | 0.947 | 0.952 | 0.95 |
| CI2 | 0.949 | 0.952 | 0.954 | 0.95 | 0.953 | 0.951 | 0.949 | 0.951 | 0.951 |

Table 1: Estimated coverage levels of CI1 and CI2 at level 95% for model (5.1) with $\varepsilon \sim \mathcal{N}(0,1)$ and $\mathrm{Proj}_V(Y) = 0.5 + 0.5X^{(1)} + X^{(2)}$.

| $n$ | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| CI1 | 0.934 | 0.937 | 0.941 | 0.942 | 0.947 | 0.949 | 0.95 | 0.945 | 0.95 |
| CI2 | 0.936 | 0.932 | 0.933 | 0.936 | 0.937 | 0.936 | 0.939 | 0.935 | 0.934 |

Table 2: Estimated coverage levels of CI1 and CI2 at level 95% for model (5.1) with $\varepsilon \sim \mathrm{St}(10)$ and $\mathrm{Proj}_V(Y) = 0.5 + 0.5X^{(1)} + X^{(2)}$.

The first confidence interval (CI1) is the one described in Corollary 3.1 (with the small change that we take the quantile $t_{n,0.975}$ of the Student distribution $\mathrm{St}(n)$ instead of the quantile $c_{0.975}$; this has no theoretical justification, but can improve a bit the coverage level for small $n$). The second confidence interval (CI2) is the non-asymptotic confidence interval ci.R2 of the R package MBESS, as described by Kelley [6], and based on a precise approximation of the distribution of $\widehat{R}^2$ given by Lee [8], [9], when the vector $(Y, X^{(1)}, \ldots, X^{(p)})$ is Gaussian. It is clearly indicated in Kelley [7] (Discussion, page 552-553) that CI2 is not robust to the non-normality of $(Y, X^{(1)}, \ldots, X^{(p)})$, which will be confirmed by the simulations.

### 5.1 An example where $(Y, X^{(1)}, X^{(2)})$ is Gaussian

We consider here the model

$$Y = 0.5 + 0.5X^{(1)} + X^{(2)} + \varepsilon\,, \tag{5.1}$$

where $X_1, X_2, \varepsilon$ are i.i.d. with $\mathcal{N}(0,1)$ distribution. Let $V$ be the sub-space of $\mathbb{L}^2$ generated by $\mathbf{1}, X^{(1)}, X^{(2)}$, so that $\mathrm{Proj}_V(Y) = 0.5 + 0.5X^{(1)} + X^{(2)}$. One can then easily check that $R^2 = 5/9$.

The estimated coverage levels for CI1 and CI2 are given in Table 1 below.

We see that the estimated coverage level of CI2 is always close to 0.95, which is not a surprise since CI2 is based on a precise approximation of the distribution of $\widehat{R}^2$ in the case where $(Y, X^{(1)}, X^{(2)})$ is Gaussian. We see that CI1 also gives very good results, with estimated coverage levels between 0.945 and 0.952 as soon as $n \geq 300$.

### 5.2 An example where the error term has a Student distribution

We consider here a slight modification of the model (5.1), where $\varepsilon \sim \mathrm{St}(10)$. Again, $V$ is the sub-space of $\mathbb{L}^2$ generated by $\mathbf{1}, X^{(1)}, X^{(2)}$, so that $\mathrm{Proj}_V(Y) = 0.5 + 0.5X^{(1)} + X^{(2)}$. One can then easily check that $R^2 = 0.5$.

The estimated coverage levels for CI1 and CI2 are given in Table 2 below.

We see that the estimated coverage level of CI2 is always between 0.932 and 0.94, and is not getting closer to 0.95 as $n$ increases. This confirms that CI2 is not robust to non normality, even when the distribution of the error is symmetric. We see that CI1 is always better than CI2 as soon as $n \geq 300$, with a coverage level between 0.94 and 0.95 when $n \geq 400$.

| $n$ | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| CI1 | 0.91 | 0.926 | 0.928 | 0.935 | 0.937 | 0.939 | 0.944 | 0.944 | 0.946 |
| CI2 | 0.878 | 0.878 | 0.878 | 0.877 | 0.877 | 0.887 | 0.884 | 0.885 | 0.887 |

Table 3: Estimated coverage levels of CI1 and CI2 at level 95% for the heteroscedastic model (5.2) with $\text{Proj}_V(Y) = 0.5 + 0.5X^{(1)} + X^{(2)}$.

| $n$ | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| CI1 | 0.935 | 0.941 | 0.943 | 0.943 | 0.943 | 0.948 | 0.947 | 0.945 | 0.947 |
| CI2 | 0.875 | 0.871 | 0.885 | 0.868 | 0.861 | 0.87 | 0.866 | 0.874 | 0.878 |

Table 4: Estimated coverage levels of CI1 and CI2 at level 95% for model (5.3) with $\text{Proj}_V(Y) \simeq -0.752 + 2.196|X|$.

### 5.3 An heteroscedastic example

We continue with a modification of the model (5.1). We consider the model

$$Y = 0.5 + 0.5X^{(1)} + X^{(2)} + \left(\sqrt{0.2 + 0.8{X^{(1)}}^2}\right) e\,, \tag{5.2}$$

where $X^{(1)}, X^{(2)}, e$ are i.i.d. with $\mathcal{N}(0,1)$ distribution. Let $V$ be the sub-space of $\mathbb{L}^2$ generated by $\mathbf{1}, X^{(1)}, X^{(2)}$, so that $\text{Proj}_V(Y) = 0.5 + 0.5X^{(1)} + X^{(2)}$. Again, one can easily check that $R^2 = 5/9$.

The estimated coverage levels for CI1 and CI2 are given in Table 3 below.

We see that the estimated coverage level of CI2 is always around 0.88. Again, this confirms that CI2 is not robust to non normality. We see that CI1 is always better than CI2, with a coverage level greater than 0.935 for $n \geq 500$, and greater than 0.94 for $n \geq 800$.

### 5.4 An example where the model is misspecified

We consider the model

$$Y = X^2 + e\,, \tag{5.3}$$

where $X \sim \mathcal{N}(0,1)$, $e \sim \mathcal{N}(0,1)$ and $e$ is independent of $X$. Let $V$ be the sub-space of $\mathbb{L}^2$ generated by $\mathbf{1}, X, |X|$. Elementary computations show that $\text{Proj}_V(Y) = \alpha_0 + \alpha_1 X + \alpha_2|X|$ with $\alpha_0 \simeq -0.752, \alpha_1 = 0, \alpha_2 \simeq 2.196$, and $R^2 \simeq 0.584$. The model is misspecified in the sense that $X^2$ does not belong to $V$. It follows that the error term $\varepsilon = Y - \text{Proj}_V(Y)$ is such that

$$\mathbb{E}(\varepsilon|X) = X^2 - \alpha_0 - \alpha_2|X| \neq 0\,.$$

The estimated coverage levels for CI1 and CI2 are given in Table 4 below.

We see that the estimated coverage level of CI2 is always around 0.87. Again, this confirms that CI2 is not robust to non normality. We see that CI1 is always better than CI2, with a coverage level greater than 0.94 for $n \geq 300$.

| $n$ | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| CI1 | 0.914 | 0.926 | 0.927 | 0.933 | 0.934 | 0.939 | 0.94 | 0.94 | 0.942 |
| CI2 | 0.75 | 0.755 | 0.752 | 0.751 | 0.754 | 0.759 | 0.74 | 0.755 | 0.746 |

Table 5: Estimated coverage levels of CI1 and CI2 at level 95% for model (5.3) with $\mathrm{Proj}_V(Y) = X^2$.

| $n$ | 200 | 300 | 400 | 500 | 600 | 700 | 800 | 900 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| CI1 | 0.922 | 0.932 | 0.933 | 0.936 | 0.94 | 0.941 | 0.946 | 0.943 | 0.943 |
| CI2 | 0.872 | 0.863 | 0.857 | 0.861 | 0.863 | 0.853 | 0.862 | 0.858 | 0.861 |

Table 6: Estimated coverage levels of CI1 and CI2 at level 95% for The Poisson regression model with $\mathrm{Proj}_V(Y) = 0.5 + X^{(1)} + X^{(2)}$.

### 5.5 An example of polynomial regression

We consider again the model (5.3), but now $V$ is the sub-space of $\mathbb{L}^2$ generated by $\mathbf{1}, X, X^2$. It is then obvious that $\mathrm{Proj}_V(Y) = \alpha_0 + \alpha_1 X + \alpha_2 X^2$ with $\alpha_0 = 0, \alpha_1 = 0, \alpha_2 = 1$, and $R^2 = 2/3$.

The estimated coverage levels for CI1 and CI2 are given in Table 5 below.

We see that the estimated coverage level of CI2 is always around 0.75. Again, this confirms that CI2 is not robust to non normality. We see that CI1 is always much better than CI2, with a coverage level greater than 0.93 for $n \geq 500$ and greater than 0.94 for $n \geq 800$.

### 5.6 An example of Poisson regression

In this example, $V$ is the subspace of $\mathbb{L}^2$ generated by $\mathbf{1}, X^{(1)}, X^{(2)}$, where $X^{(1)}, X^{(2)}$ are independent, $X^{(1)}$ is uniformly distributed over $[0,1]$, and $X^{(2)}$ is exponentially distributed with parameter 1. The response variable $Y$ is a count variable, whose conditional distribution given $X^{(1)}, X^{(2)}$ is a Poisson distribution with parameter $0.5 + X^{(1)} + X^{(2)}$. In that case, we have $\mathrm{Proj}_V(Y) = \mathbb{E}(Y|X^{(1)}, X^{(2)}) = \alpha_0 + \alpha_1 X^{(1)} + \alpha_2 X^{(2)}$ with $\alpha_0 = 0.5, \alpha_1 = 2, \alpha_2 = 1$, and $R^2 = 13/37$. Note that this is again an heteroscedastic model, since

$$\mathrm{Var}(\varepsilon|X^{(1)}, X^{(2)}) = \mathrm{Var}(Y - \mathrm{Proj}_V(Y)|X^{(1)}, X^{(2)}) = 0.5 + X^{(1)} + X^{(2)}\,.$$

The estimated coverage levels for CI1 and CI2 are given in Table 6 below.

We see that the estimated coverage level of CI2 is always around 0.86. Again, this confirms that CI2 is not robust to non normality. We see that CI1 is always much better than CI2, with a coverage level greater than 0.93 for $n \geq 300$ and greater than 0.94 for $n \geq 600$.

## 6 Extension to higher dimension and discussions

### 6.1 Extension to vector-valued random variables

In this section, we assume that $(Y, X^{(1)}, \ldots X^{(p)})$ is a random vector with value in $\mathbb{H}^{p+1}$, where $\mathbb{H}$ is a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|_{\mathbb{H}}$. Let $\mathbb{L}^2(\mathbb{H})$ be the space of $\mathbb{H}$-valued random variables $Z$ such that $\mathbb{E}(\|Z\|_{\mathbb{H}}^2) < \infty$, and recall that $\mathbb{L}^2(\mathbb{H})$ is also a separable Hilbert space with inner product $\mathbb{E}(\langle \cdot, \cdot \rangle)$.

We assume in the following that all the variables $Y, X^{(1)}, \dots X^{(p)}$ belong to $\mathbb{L}^2(\mathbb{H})$. Let $V$ be the subspace of $\mathbb{L}^2(\mathbb{H})$ generated by the variables $(X^{(1)}, \dots, X^{(p)})$, and let $\mathrm{Proj}_V(Y)$ be the orthogonal projection of $Y$ onto $V$. To measure the quality of approximation of $Y$ by $\mathrm{Proj}_V(Y)$, one can consider the coefficient $R^2$ defined by (assuming that $\mathbb{E}(\|Y\|_{\mathbb{H}}^2) > 0$):

$$R^2 = \frac{\mathbb{E}\left(\|\mathrm{Proj}_V(Y)\|_{\mathbb{H}}^2\right)}{\mathbb{E}\left(\|Y\|_{\mathbb{H}}^2\right)}\,.$$

Let $(Y_i, X_i^{(1)}, \dots X_i^{(p)})_{1\le i\le n}$ be $n$ independent copies of the vector $(Y, X^{(1)}, \dots X^{(p)})$, and let $V_i$ be the $\mathbb{L}^2(\mathbb{H})$-subspace generated by the variables $(X_i^{(1)}, \dots, X_i^{(p)})$. Let $\varepsilon = Y - \mathrm{Proj}_V(Y)$ and $\varepsilon_i = Y_i - \mathrm{Proj}_{V_i}(Y_i)$. We can then write

$$Y_i = \mathrm{Proj}_{V_i}(Y_i) + \varepsilon_i.$$

where the variables $\varepsilon_i$ are such that $\mathbb{E}(\langle X_i^{(j)}, \varepsilon_i\rangle) = 0$ for any $j \in \{0, \dots, p\}$.

Assume that $X^{(1)}, \dots, X^{(p)}$ are linearly independent (as a family of random variables in $\mathbb{L}^2(\mathbb{H})$). Then one can uniquely write

$$Y_i = \alpha_1 X_i^{(1)} + \cdots + \alpha_p X_i^{(p)} + \varepsilon_i\,.$$

Let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^t \in \mathbb{R}^p$, and $\mathbf{M}$ be the matrix such that

$$\mathbf{M}_{j,\ell} = \mathbb{E}(\langle X^{(j)}, X^{(\ell)}\rangle)\,. \tag{6.1}$$

Since $X^{(1)}, \dots, X^{(p)}$ are linearly independent, $\mathbf{M}$ is invertible and one can easily show that

$$\boldsymbol{\alpha} = \mathbf{M}^{-1}(\mathbb{E}(\langle X^{(1)}, Y\rangle) \dots, \mathbb{E}(\langle X^{(p)}, Y\rangle))^t\,.$$

Let now $\widehat{\mathbf{M}}$ be the matrix such that $\widehat{\mathbf{M}}_{j,\ell} = n^{-1}\sum_{i=1}^n \langle X_i^{(j)}, X_i^{(\ell)}\rangle$. By the strong law of large numbers

$$\lim_{n\to\infty} \widehat{\mathbf{M}} = \mathbf{M} \text{ almost surely.} \tag{6.2}$$

Since $\mathbf{M}$ is invertible, for $n$ large enough, $\widehat{\mathbf{M}}$ is invertible. Consequently, one can define

$$\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_p)^t = \widehat{\mathbf{M}}^{-1}\left(\frac{1}{n}\sum_{i=1}^n \langle X_i^{(1)}, Y_i\rangle, \dots, \frac{1}{n}\sum_{i=1}^n \langle X_i^{(p)}, Y_i\rangle\right)^t\,.$$

It follows that

$$\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} = (\widehat{\alpha}_1, \dots, \widehat{\alpha}_p)^t - (\alpha_1, \dots, \alpha_p)^t = \widehat{\mathbf{M}}^{-1}\left(\frac{1}{n}\sum_{i=1}^n \langle X_i^{(1)}, \varepsilon_i\rangle, \dots, \frac{1}{n}\sum_{i=1}^n \langle X_i^{(p)}, \varepsilon_i\rangle\right)^t\,. \tag{6.3}$$

Then, proceeding as in Section 2.2, we see that $\widehat{\boldsymbol{\alpha}}$ converges almost surely to $\boldsymbol{\alpha}$, and that: if for any $j \in \{0, \dots, p\}$, one has $\mathbb{E}(\langle X^{(j)}, \varepsilon\rangle^2) < \infty$, then

$$\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}_p(0, \mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1})\,,$$

where $\mathbf{M}(\varepsilon)$ is the matrix defined by $\mathbf{M}(\varepsilon)_{j,\ell} = \mathbb{E}(\langle X^{(j)}, \varepsilon\rangle\langle X^{(\ell)}, \varepsilon\rangle)$ for any $0 \le j, \ell \le p$.

One can then estimate the matrix $\mathbf{M}(\varepsilon)$ as in Section 2.3. Define the residuals $\widehat{\varepsilon}_i$ by:

$$\widehat{\varepsilon}_i = Y_i - \widehat{\alpha}_1 X_i^{(1)} - \cdots - \widehat{\alpha}_p X_i^{(p)}.$$

Let then

$$\widehat{\mathbf{M}(\varepsilon)}_{j,\ell} = \frac{1}{n}\sum_{i=1}^{n} \langle X_i^{(j)}, \widehat{\varepsilon}_i\rangle \langle X_i^{(\ell)}, \widehat{\varepsilon}_i\rangle \,.$$

As in Proposition 2.3, the following consistency result holds: assume that $\mathbb{E}(\|X^{(j)}\|_{\mathbb{H}}^2 \|\varepsilon\|_{\mathbb{H}}^2) < \infty$ and $\mathbb{E}(\|X^{(j)}\|_{\mathbb{H}}^4) < \infty$ for any $j \in \{0,\dots,p\}$. Then, for any $j,\ell \in \{0,\dots,p\}$,

$$\lim_{n\to\infty} \widehat{\mathbf{M}(\varepsilon)}_{j,\ell} = \mathbf{M}(\varepsilon)_{j,\ell} \quad \text{almost surely.}$$

Consequently, since $\mathbf{M}$ is invertible,

$$\lim_{n\to\infty} \widehat{\mathbf{M}}^{-1}\widehat{\mathbf{M}(\varepsilon)}\widehat{\mathbf{M}}^{-1} = \mathbf{M}^{-1}\mathbf{M}(\varepsilon)\mathbf{M}^{-1} \quad \text{almost surely.}$$

The empirical estimator of $R^2$ is

$$\widehat{R}^2 = \frac{\sum_{i=1}^n \|\mathrm{Proj}_{V_i}(Y_i)\|_{\mathbb{H}}^2}{\sum_{i=1}^n \|Y_i\|_{\mathbb{H}}^2}\,.$$

Let $\boldsymbol{\theta} = (\theta_1,\dots,\theta_p)^t$, with $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1,\dots,\widehat{\theta}_p)^t$

$$\theta_j = \frac{\mathbb{E}(\langle X^{(j)}, Y\rangle)}{\mathbb{E}(\|Y\|_{\mathbb{H}}^2)} \quad \text{and} \quad \widehat{\theta}_j = \frac{\sum_{i=1}^n \langle X_i^{(j)}, Y_i\rangle}{\sum_{i=1}^n \|Y_i\|_{\mathbb{H}}^2} \quad \text{for } j \in \{1,\dots,p\}.$$

Then, as in Lemma 3.1 one can prove that $R^2 = \boldsymbol{\theta}^t\boldsymbol{\alpha}$ and $\widehat{R}^2 = \widehat{\boldsymbol{\theta}}^t\widehat{\boldsymbol{\alpha}}$. Let

$$e^{(k)} = X^{(k)} - \mathrm{Proj}_W(X^{(k)}) \quad \text{and} \quad e_i^{(k)} = X_i^{(k)} - \mathrm{Proj}_{W_i}(X_i^{(k)}), \tag{6.4}$$

where $W$ is the subspace of $\mathbb{L}^2(\mathbb{H})$ generated by $Y$, and $W_i$ is the subspace of $\mathbb{L}^2(\mathbb{H})$ generated by $Y_i$. Let also $\boldsymbol{e}^{(k)} = (e_1^{(k)},\dots,e_n^{(k)})^t$. As in Proposition 3.2, one can prove the following result: if $\mathbb{E}(\langle X^{(j)},\varepsilon\rangle^2) < \infty$ and $\mathbb{E}(\langle Y, e^{(j)}\rangle^2) < \infty$ for all $j \in \{1,\dots,p\}$. Then

$$\sqrt{n}\left(\widehat{R}^2 - R^2\right) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,V).$$

where

$$V = (\theta_1,\dots,\theta_p,\alpha_1,\dots,\alpha_p)\mathbf{BAB}(\theta_1,\dots,\theta_p,\alpha_1,\dots,\alpha_p)^t\,.$$

Here, the matrices $\mathbf{A}$ and $\mathbf{B}$ are as follows:

- If $(j,k) \in \{1,\dots,p\}^2$ then
$$\mathbf{A}_{j,k} = \mathbb{E}(\langle X^{(j)},\varepsilon\rangle\langle X^{(k)},\varepsilon\rangle);$$
- If $(j,k) \in \{p+1,\dots,2p\}^2$ then
$$\mathbf{A}_{j,k} = \mathbb{E}(\langle Y, e^{(j-p)}\rangle\langle Y, e^{(k-p)}\rangle);$$

- $(j,k) \in \{1,\dots,p\} \times \{p+1,\dots,2p\}$ then

$$\mathbf{A}_{j,k} = \mathbb{E}(\langle X^{(j)}, \varepsilon \rangle \langle Y, e^{(k-p)} \rangle).$$

Let $\delta_{j,k} = 0$ if $j \neq k$ and $\delta_{j,j} = 1$ and let $\mathbf{B}$ be the $2p \times 2p$ matrix defined as follows:

- If $(j,k) \in \{1,\dots,p\}^2$ then $\mathbf{B}_{j,k} = (\mathbf{M}^{-1})_{j,k}$ (see (6.1) for the definition of $\mathbf{M}$);
- If $(j,k) \in \{p+1,\dots,2p\}^2$ then $\mathbf{B}_{j,k} = (\mathbb{E}(\|Y\|_{\mathbb{H}}^2))^{-1}\delta_{j,k}$;
- If $(j,k) \in \{1,\dots,p\} \times \{p+1,\dots,2p\}$ then $\mathbf{B}_{j,k} = 0$.

Now the matrices $\mathbf{A}$ and $\mathbf{B}$ can be consistently estimated by their empirical counterparts based on the residuals, provided $\mathbb{E}(\|Y\|_{\mathbb{H}}^4) < \infty$ and $\mathbb{E}(\|X^{(j)}\|_{\mathbb{H}}^4) < \infty$ for any $j \in \{1,\dots,p\}$ (as done in the proof Proposition 3.3). As in Proposition 3.3 this provides a consistent estimator of the limiting variance $V$.

### 6.2 An alternative to the usual sensitivity analysis

In sensitivity analysis, one usually assumes that

$$Y = m(X^{(1)},\dots,X^{(p)}) \tag{6.5}$$

(or that $Y = m(X^{(1)},\dots,X^{(p)}) + \varepsilon$ as in [4]) where $(X^{(1)},\dots,X^{(p)})$ is a $p$ random vector with a known distribution $= P_1 \otimes \cdots \otimes P_p$ on a compact subset $\mathcal{X}$ of $\mathbb{R}^p$. The number of variables $p$ may be large, and the function $m$ is an unknown function from $\mathbb{R}^p$ to $\mathbb{R}$, which may present strong non-linearities and high order interaction effects between its coordinates. On the basis of a $n$-sample $(Y_i, X_i^{(1)},\dots,X_i^{(p)})$, for $i = 1,\dots,n$, we want to measure the impact of the main effects or interactions, i.e. to determine the influence of each variable or group of variables on the output variable $Y$. The usual sensitivity analysis is mainly based on Hoeffding's decomposition, strongly related to the independency between the inputs variables $X^{(1)},\dots,X^{(p)}$. In that latter case, if $m$ is square integrable, one may consider the classical Hoeffding-Sobol decomposition [14] that leads to write $m$ according to its ANOVA functionnal expansion:

$$m(X^{(1)},\dots,X^{(p)}) = m_0 + \sum_{\substack{i_1 < i_2 < \cdots < i_k, \\ i_i,\dots,i_k \in \{1,\dots,p\}}} m_{i_1,\dots,i_k}(X^{(i_1)},\dots,X^{(i_k)}) \tag{6.6}$$

where the functions $m_{i_1,\dots,i_k}$ are centered and orthogonal in $\mathbb{L}^2$ involving conditional expectation that is

$$\begin{aligned} m_0 &= \mathbb{E}(m(X^{(1)},\dots,X^{(p)})), \quad m_j(X^{(j)}) = \mathbb{E}(m(X^{(1)},\dots,X^{(p)})|X^{(j)}) - m_0 \\ m_{j,k}(X^{(j)},X^{(k)}) &= \mathbb{E}(m(X^{(1)},\dots,X^{(p)})|X^{(j)},X^{(k)}) - m_j(X^{(j)}) - m_k(X^{(k)}) - m_0, \quad \dots \end{aligned}$$

Under the assumption of independence on the $X^{(j)}$'s this decomposition is unique, with

$$\mathbb{E}(m_{i_1,\dots,i_k}(X^{(i_1)},\dots,X^{(i_k)})) = 0$$

and for all $(i_1,\dots,i_k) \neq (i'_1,\dots,i'_\ell)$,

$$\mathbb{E}\left(m_{i_1,\dots,i_k}(X^{(i_1)},\dots,X^{(i_k)}) m_{i'_1,\dots,i'_\ell}(X^{(i'_1)},\dots,X^{(i'_\ell)})\right) = 0\,.$$

This leads to the following variance decomposition:

$$\mathrm{Var}(Y) = \sum_{\substack{i_1 < i_2 < \cdots < i_k, \\ i_i, \ldots, i_k \in \{1, \ldots, p\}}} \mathrm{Var}\left( m_{i_1,\ldots,i_k}(X^{(i_1)}, \ldots, X^{(i_k)}) \right).$$

The Sobol sensitivity indices introduced by Sobol [14] are defined for any group $(X^{(i_1)}, \ldots, X^{(i_k)})$ by

$$S_{i_1,\ldots,i_k} = \frac{\mathrm{Var}\left( m_{i_1,\ldots,i_k}(X^{(i_1)}, \ldots, X^{(i_k)}) \right)}{\mathrm{Var}\left(Y\right)}.$$

They quantify the contribution of a subset of variables $(X^{(1)}, \ldots, X^{(p)})$ to the output $Y$. Several approaches are available for estimating these sensitivity indices, see for example Iooss and Lemaître [5] for a recent review.

An alternative approach, which is much less computationally expensive, is as follows. Recall that the $X^{(i)}$'s are bounded and let $\tilde{X}^{(i)} = X^{(i)} - \mathbb{E}(X^{(i)})$ (this centering may be done, since the distribution of $(X^{(1)}, \ldots, X^{(p)})$ is known). Since the variables $X^{(j)}$'s are independent, the family

$$\mathcal{F} = \left\{ \tilde{X}^{(i_1)} \tilde{X}^{(i_2)} \ldots \tilde{X}^{(i_k)}, i_1 < \cdots < i_k \right\}$$

is an orthogonal family of $\mathbb{L}^2$. We can then compute the individual $R^2$'s for this family, that is

$$R^2_{i_1,\ldots,i_k} = \frac{\mathrm{Var}\left( \mathrm{Proj}_{V^{(i_1,\ldots,i_k)}}(Y) \right)}{\mathrm{Var}\left(Y\right)}.$$

where $V^{(i_1,\ldots,i_k)}$ is the subspace of $\mathbb{L}^2$ generated by the product $\tilde{X}^{(i_1)} \tilde{X}^{(i_2)} \ldots \tilde{X}^{(i_k)}$.

The main difference between the two approaches is that the sum of all the indices $R_{i_1,\ldots,i_k}$ is not equal to 1, unless $Y$ belongs to the subspace $V_{\mathcal{F}}$ of $\mathbb{L}^2$ generated by the family $\mathcal{F}$. More precisely, we have that

$$\sum_{\substack{i_1 < i_2 < \cdots < i_k, \\ i_i, \ldots, i_k \in \{1, \ldots, p\}}} R^2_{i_1,\ldots,i_k} = \frac{\mathrm{Var}\left( \mathrm{Proj}_{V_{\mathcal{F}}}(Y) \right)}{\mathrm{Var}\left(Y\right)}.$$

The advantage of using these $R^2$'s is that they are very easy to estimate (based on $n$ independent observations $(Y_i, X_i^{(1)}, \ldots, X_i^{(p)})_{1 \leq i \leq n}$). In addition, as we have seen, we can give confidence intervals for these quantities.

### 6.3 A first step toward robust screening

In this section, we outline a strategy for robust screening based on the results of Section 4.

**Quick context** The basic idea of screening is to find, among a set of covariates $X^{(1)}, \ldots, X^{(p)}$, the ones having an association with an outcome $Y$. One usual way which is often cited consists in computing the $p$ correlations:

$$\mathrm{cor}(Y, X^{(j)}) \text{ for } 1 \leq j \leq p.$$

Then, given a threshold $\gamma_n$ we keep the covariates such that:

$$|\text{cor}(Y, X^{(j)})| \geq \gamma_n.$$

To assess the performance of a screening method, whether it is based on computing $p$ correlations as we mentioned or not, one has to ask two questions:

1. Is the method able to retrieve all the covariates that are associated with the outcome $Y$? This question is known in the literature as the Sure Screening Property (see Fan and Lv [1]).
2. Is the method specific enough, i.e. can we ensure that not too many covariates are wrongly selected? This question directly relates to the control of the False Discovery Rate (FDR).

The questions of the sure screening property and the control of the FDR in a model-free environment is a growing field of research (see for instance Fan and Lv [1], Zhao and Li [16], Fan and Song [2], W. Pan, X. Wang, W. Xiao, H. Zhu [13]) To answer both questions, the choice of the threshold $\gamma_n$ is crucial. One paper which is often cited is the paper by Fan and Lv [1]. Their context is the following: they suppose a true joint linear model:

$$Y = \sum_{j=1}^{p} X^{(j)} \beta_j^\star + \varepsilon$$

where $\varepsilon$ is a standard Gaussian error. Moreover, they suppose that for each $j = 1, \ldots, p$, the $X_i^{(j)}$ are i.i.d (respectively to $i = 1, \ldots, n$). They also suppose that $X^{(j)}$ are standardized with mean 0 and standard deviation 1. They consider the screening set:

$$\widehat{\mathcal{M}} = \left\{ j : |\text{cor}(Y, X^{(j)})| \geq \frac{n}{\log(n)} \right\}.$$

Ensuring the Sure Screening Property in that specific case consists in showing that the support of $\beta^\star$ is *at least* included in $\widehat{\mathcal{M}}$ with a probability close to one. However, the threshold chosen $\gamma_n = n/\log(n)$ does not limit the number of false positives, i.e., the covariates that are not in the support of $\beta^\star$ but are selected marginally. This threshold, although not having any theoretical support for controlling the FDR, is widely used in the screening literature.

The two questions 1. and 2. above are addressed for instance in Zhao and Li [16] in the context of the Cox model. In this paper the authors propose to choose

$$\widehat{\mathcal{M}} = \left\{ j : \frac{\sqrt{n}|\hat{\alpha}_j|}{\sqrt{\hat{v}_j}} \geq \gamma \right\},$$

where $\hat{\alpha}_j$ corresponds to the estimator in the marginal (and possibly misspecified) model, and $\hat{v}_j$ is a consistent estimator of the asymptotic variance of $\sqrt{n}(\hat{\alpha}_j - \alpha_j)$. According to Zhao and Li [16], denoting by $\Phi$ the distribution function of a standard Gaussian and by $q$ the expected false positive rate, the idea is to choose

$$\gamma = \Phi^{-1}\left(1 - \frac{q}{2}\right).$$

It is common in papers dealing with screening to consider that independence beetwen $X^{(j)}$ being truly associated with $Y$ in the joint model is equivalent to being associated with $Y$ marginally (see Condition 3 in [1] and Assumption 8 in Zhao and Li [16])). However, if we are working in a model-free context as in Section 4, these assumptions are unnecessary and it seems possible to define a threshold $\gamma$ using the distribution function $\Phi$.

**Our idea.** First, we highlight the fact that even in a model-free context, the question of screening the covariates is still relevant as a lowering dimension step before trying to fit a joint association between the remaining covariates and the outcome.

Recall from Section 4 the expression of $\tau_j$ and $\hat{\tau}_j$ given in Equation (4.1) and (4.2)

$$\tau_j = \frac{\mathrm{Cov}(Y, X^{(j)})}{\mathrm{Var}(X^{(j)})} \quad \text{and} \quad \widehat{\tau}_j = \frac{\sum_{i=1}^n (Y_i - \overline{\mathbf{Y}})(X_i^{(j)} - \overline{\mathbf{X}}^{(j)})}{\sum_{i=1}^n (X_i^{(j)} - \overline{\mathbf{X}}^{(j)})^2}.$$

Then, Proposition 4.1 gives

$$\sqrt{n}(\widehat{\tau}_j - \tau_j) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0, v_j)$$

where

$$v_j = \frac{\mathbb{E}((X^{(j)} - \mathbb{E}(X^{(j)}))^2 {\varepsilon^{(j)}}^2)}{(\mathrm{Var}(X^{(j)}))^2}.$$

Moreover, it has been shown in Section 4 that if we let

$$\widehat{v_j} = \frac{1}{n} \sum_{i=1}^n \frac{(X_i^{(j)} - \overline{\mathbf{X}}^{(j)})^2}{(\mathrm{Var}_n(X^{(j)}))^2}$$

then, for all $j = 1, \ldots, p$,

$$\lim_{n\to\infty} \widehat{v}_j = v_j \text{ almost surely.}$$

Similarly to Zhao and Li [16], a reasonnable screening set rule should be

$$\widehat{\mathcal{M}} = \left\{ j : \frac{\sqrt{n}|\widehat{\tau}_j|}{\sqrt{\widehat{v_j}}} \geq \gamma \right\} \quad \text{with } \gamma = \Phi^{-1}\left(1 - \frac{q}{2}\right).$$

The next step is to show that this rule does indeed allow us to obtain the sure screening property, as well as satisfactory control of the FDR. We believe that moment assumptions on the variables should be sufficient to show these two properties, and we plan to study these issues in future work.

# References


[1] J. Fan and Lv, (2008). Sure independence screening for ultrahigh dimensional feature space. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **70**, 849–911.

[2] J. Fan and R. Song, (2010). Sure independence screening in generalized linear models with NP-dimensionality. Annals of Statistics, **38**, 3567-3604.

[3] R. A. Fisher, (1928). The general sampling distribution of the multiple correlation coefficient. Proc. Roy. Soc. A, **121**, pp 654-673.

[4] S. Huet and M.-L. Taupin, (2017). Metamodel construction for sensitivity analysis ESAIM: Procs **60** pp 27-69.

[5] B. Iooss, and P. Lemaître, (2015). A review on global sensitivity analysis methods. *Uncertainty management in Simulation-Optimization of Complex Systems: Algorithms and Applications*, Springer.

[6] K. Kelley, (2007). Confidence Intervals for Standardized Effect Sizes: Theory, Application, and Implementation Journal of Statistical Software **20**, no 8, 24 pages.

[7] K. Kelley, (2008). Sample Size Planning for the Squared Multiple Correlation Coefficient: Accuracy in Parameter Estimation via Narrow Confidence Intervals Multivariate Behavioral Research **43**, pp 524-555.

[8] Y.-S. Lee, (1971). Some results on the sampling distribution of the multiple correlation coefficient. J. Roy. Statist. Soc. Ser. B, **33**, 117-130.

[9] Y.-S. Lee, (1972). Tables of upper percentage points of the multiple correlation coefficient. Biometrika, **59**, pp 175-189.

[10] R. J. Muirhead, (1982). Aspects of multivariate statistical theory. *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, Inc., New York, xix+673 pp

[11] H. Ogasawara, (2006). Asymptotic expansion and conditional robustness for the sample multiple correlation coefficient under nonnormality. Comm. Statist. Simulation Comput. **35**, no. 1, pp 177-199.

[12] H. Ogasawara, (2008). Asymptotic expansions of the distribution of the estimator for the generalized partial correlation under non-normality. Behaviormetrika **35**, no. 1, pp. 15–33.

[13] W. Pan, X. Wang, W. Xiao, H. Zhu, (2019). A Generic Sure Independence Screening Procedure. J. Am. Stat Assoc., **114** (526). pp: 928-937.

[14] I. M. Sobol, (1993). Sensitivity estimates for nonlinear mathematical models. Journal Math. Model. Comput. Exp., **1**, 407-414.

[15] H. White, (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica, **48**, no. 4. pp 817-838.

[16] S. D. Zhao and Y. Li, (2012) Principled sure independence screening for Cox models with ultra-high-dimensional covariates. Journal of Multivariate Analysis, **105**, 397–411.