A Necessary and Sufficient Condition for Minimising the risk of predictive inference in methods of data perturbation

Cédric Wanko & Jean-Marie Berton

TRIS−UM1

50emes Journées de Statistique SFdS
28 Mai -1 Juin 2018 Paris, France
13emes Journées de Méthodologie Statistique de l'Insee
12-14 Juin 2018 Paris, France
S^{4}D Research Summer School on Statistics for Data Science
18-22 Juin 2018 Caen, France

Plan

Topic

In this paper, we work on predictive inference problems through data **Masking/Coding** Method in a Gaussian setting with the 3 following models:

1. Data perturbation by means of random noise.
2. Copula Theory.
3. Gram-Schmidt (GS) orthonormalization algorithm.

Literature: Problems

1. [DRECHSLER AND REITER (2010) (J.OF AM. ST. ASS.) MIRANDA AND VILHUBER (2016) (PRIV. ST. DATABASES)] methods of de-identification of respondents by the creation of *synthetic microdata*. **Problem:** Important *loss of information* because there is too much difference between the confidential data and the synthesized data in terms of similarity and doesn't take into account *non-confidential data*.

2. [BURRIDGE (2003) (STAT. & COMP.) (MURALIDHAR AND SARATHY, 2008) (TRANS. DATA PRIV.) CALVIÑO (2017) (J. OF. STAT.)] For the purpose of combating this loss of information we focus on methods of data perturbation by means of random noise and a *'similarity' coefficient*: SBNA method. **Problem: (inference)** *Non-confidential data* still remain *sensitive* because they are even so present in the coding process.

Literature: Results

1. [Burridge (2003) (Stat. & Comp.)] **Result:** Information Preserving Statistical Obfuscation (IPSO) method allow to generate synthetic variables that are totally independent of the confidential data and take into account *non-confidential data* ⇒ very *high level of security*.

2. [(Muralidhar and Sarathy) (2008) (Trans. Data Priv.) Domingo-Ferrer and González-Nicolás (2010) (Information Sciences) Calviño (2017) (J. Of. Stat.)] **Result:** SBNA method takes into account *non-confidential data* and allows to adjust degree of similarity between confidential and masked data ⇒ *avoids loss of information*.

Focus

Object of this paper:

We propose to *minimise the influence of non-confidential variables*, while still retaining them within the model. The non-confidential variables must therefore be only *slightly*, or *not at all significant* in the model's specification, while preserving its the basic theoretical properties.

Motivation

Motivation

Since *it is not possible to mask everything*, it is necessary to be aware of the existence of non-confidential public variables during the coding process and cognisant of the risks they entail.

How?

We define a necessary and sufficient condition that enables us to reinforce the security by *minimising risks of predictive divulgation* in the perturbation methods.

contribution

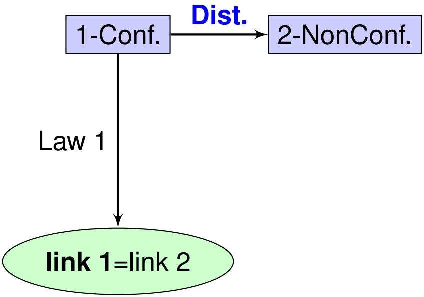
contribution

contribution in this paper: We show that when the *non-confidential data are orthonormalized and perturbed* in their structure, they then satisfy the *'predictive inference risk requirements'* that minimise predictive divulgation risk, while preserving both the 'data utility requirements' and the 'disclosure risk requirements'.

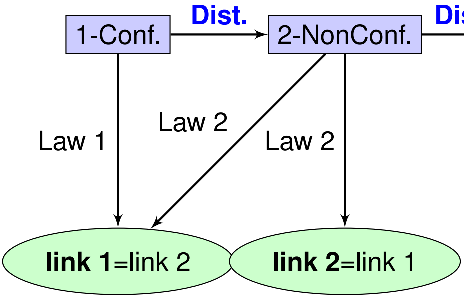
The Masking/Coding work flow 1

1-Conf.

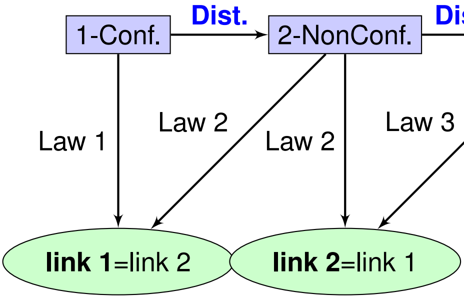
The Masking/Coding work flow 1



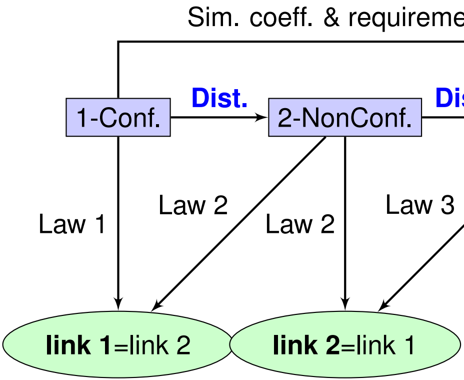
The Masking/Coding work flow 1



The Masking/Coding work flow 1



The Masking/Coding work flow 1



Plan

1. Introduction
   - Topic
   - Literature
   - Contribution
2. **Model**
   - Framework
   - Results
3. Perspective
   - Future research

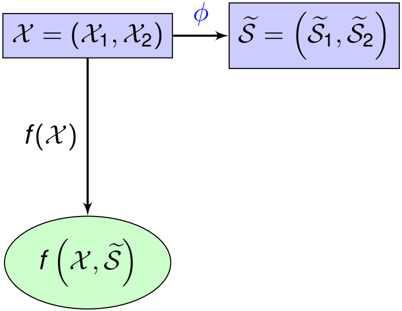
Notation

- f and F are respectively the density and distribution function of the bivariate gaussian distribution with Pearson correlation parameter $\rho \in [-1, +1]$.
- C is a distribution function on the unit cube $[0, 1]^K$ in $\mathbb{R}^K$ with uniform marginal distributions. C is called a *copula* such that: $F_\ell(\mathcal{S}) = C_\ell(F_{\mathrm{S}})$. $\ell \in [-1, +1]$ is the measure of dependence during the transformation.
- $r = \phi\rho$ is a biased estimate of ρ and $\widetilde{\rho}$ is a disturbed expression of ρ.
- $\widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{S}}_2\right)$, the orthonormalized and transformed vector of non-confidential data, $\mathcal{S} = (\mathcal{S}_1, \mathcal{S}_2)$ the non-confidential data and $\mathrm{S} = (\mathcal{S}_1/\mathcal{S}_2 = m_{\varsigma_2}, \mathcal{S}_2/\mathcal{S}_1 = m_{\varsigma_1})$ the vector of transformation with m_{ς_i} as the theoretical mean of the series $\mathcal{S}$ for the observation i.
- The confidential variables are $\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$ and the masked and publicly disclosed variables are $\mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2)$.

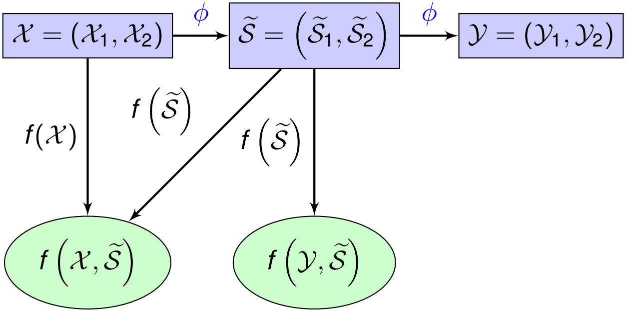
The Masking/Coding work flow 2

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$$

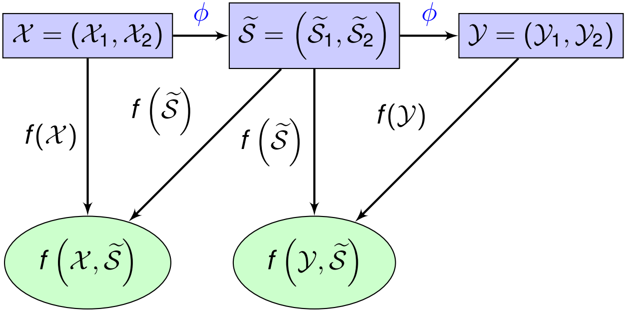
The Masking/Coding work flow 2



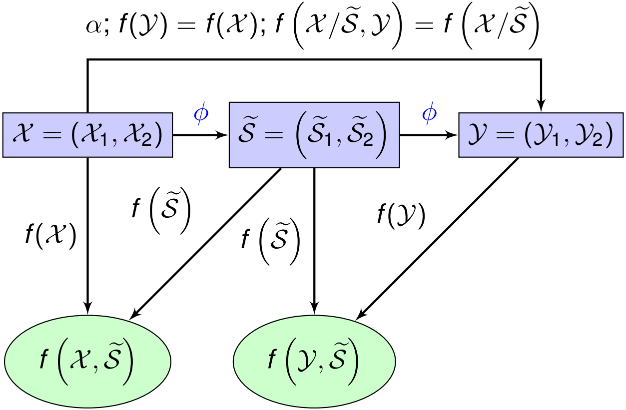
The Masking/Coding work flow 2



The Masking/Coding work flow 2



The Masking/Coding work flow 2



The model: Minimising Predictive Inference Risk

A THEORETICAL BASIS

- The *'data utility or accuracy requirements'*

For $y_i \sim f\left(\mathcal{X}/\widetilde{\mathcal{S}} = \widetilde{\varsigma}_i\right)$, the necessary conditions are

$$f(\mathcal{Y}) = f(\mathcal{X}) \text{ and } f\left(\mathcal{Y}, \widetilde{\mathcal{S}}\right) = f\left(\mathcal{X}, \widetilde{\mathcal{S}}\right).$$

- The *'disclosure risk requirements'*

Knowing $(\mathcal{Y}, \widetilde{\mathcal{S}})$ do not increase the risk of divulgation:

$$f\left(\mathcal{X}/\widetilde{\mathcal{S}}, \mathcal{Y}\right) = f\left(\mathcal{X}/\widetilde{\mathcal{S}}\right)$$

- The *'predictive inference risk requirements'* imply that if, according to a Gram-Schmidt orthonormalization:

$$\widetilde{\mathcal{S}}_1 = \mathcal{S}_1 \text{ and } \widetilde{\mathcal{S}}_2 = \mathcal{S}_2 - \frac{f(\mathcal{S}_1, \mathcal{S}_2)}{f(\mathcal{S}_1, \mathcal{S}_1)}\mathcal{S}_1$$

$$\text{and } f\left(\widetilde{\mathcal{S}}\right) = f(\widetilde{\mathcal{S}}_1) f(\widetilde{\mathcal{S}}_2)$$

$$\text{then } f(\mathcal{Y}) = f(\mathcal{X}) \text{ and } f(\mathcal{Y}, \mathcal{S}) = f(\mathcal{X}, \mathcal{S}).$$

The model: Minimising Predictive Inference Risk

A COMPUTATIONAL APPROACH

The estimation of the log-likelihood

$$\log \prod_{i=1}^{n} f_{\ell}(\mathcal{S}) = \log \prod_{i=1}^{n} \left\{ c_{\ell}(F_{S}) f_{\mathcal{S}_1/\mathcal{S}_2=m_{\varsigma_2}}(\varsigma_1) f_{\mathcal{S}_2/\mathcal{S}_1=m_{\varsigma_1}}(\varsigma_2) \right\}$$

or the estimation of the Gaussian *copula* associated with $f_{\ell}(\mathcal{S})$

$$\log \prod_{i=1}^{n} c_{\ell}(F_{S}) = \log \prod_{i=1}^{n} \left\{ f_{\ell}(\mathcal{S}) f^{-1}_{\mathcal{S}_1/\mathcal{S}_2=m_{\varsigma_2}}(\varsigma_1) f^{-1}_{\mathcal{S}_2/\mathcal{S}_1=m_{\varsigma_1}}(\varsigma_2) \right\}$$

give us the *measure dependence of the perturbation*:

$$\widetilde{\rho} = \phi\rho \left((1-\ell^2) \left(1 - (\phi\rho)^2 \right) + \phi\rho\ell \right).$$

The model: Minimising Predictive Inference Risk

A SIMULATION BASIS

$$\gamma = \mu_{\mathcal{X}} - \Sigma_{\mathcal{X}\widetilde{\mathcal{S}}}\Sigma^{-1}_{\widetilde{\mathcal{S}}\widetilde{\mathcal{S}}}\mu_{\widetilde{\mathcal{S}}}$$
$$\beta = \Sigma_{\mathcal{X}\widetilde{\mathcal{S}}}\Sigma^{-1}_{\widetilde{\mathcal{S}}\widetilde{\mathcal{S}}}$$
$$\Sigma_{\epsilon} = \left(\Sigma_{\mathcal{X}\mathcal{X}} - \Sigma_{\mathcal{X}\widetilde{\mathcal{S}}}\Sigma^{-1}_{\widetilde{\mathcal{S}}\widetilde{\mathcal{S}}}\Sigma_{\widetilde{\mathcal{S}}\mathcal{X}}\right)$$

with the constant γ and μ the mean

β the parameter with respect to $\widetilde{\mathcal{S}}$

and Σ_{ϵ} the covariance error matrix.

Taking into account the *similarity coefficient α*:

$$\gamma = (I - \alpha)\,\mu_{\mathcal{X}} - \beta\mu_{\widetilde{\mathcal{S}}}$$
$$\beta^{T} = \Sigma^{-1}_{\widetilde{\mathcal{S}}\widetilde{\mathcal{S}}}\Sigma_{\widetilde{\mathcal{S}}\mathcal{X}}\left(I - \alpha^{T}\right)$$
$$\Sigma_{\epsilon} = \left(\Sigma_{\mathcal{X}\mathcal{X}} - \Sigma_{\mathcal{X}\widetilde{\mathcal{S}}}\Sigma^{-1}_{\widetilde{\mathcal{S}}\widetilde{\mathcal{S}}}\Sigma_{\widetilde{\mathcal{S}}\mathcal{X}}\right) - \alpha\left(\Sigma_{\mathcal{X}\mathcal{X}} - \Sigma_{\mathcal{X}\widetilde{\mathcal{S}}}\Sigma^{-1}_{\widetilde{\mathcal{S}}\widetilde{\mathcal{S}}}\Sigma_{\widetilde{\mathcal{S}}\mathcal{X}}\right)\alpha^{T}.$$

Main Results

Condition 1:

$\widetilde{\rho} = 0$ is a necessary and sufficient condition for that $\phi = 0$ knowing the given ρ and fixed ℓ.

consequences:

If $\widetilde{\rho} = 0$ then $\phi = 0$ is always a *solution* ("dummy" perturbation).

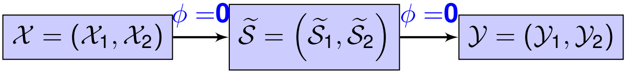
Main Results

$$\boxed{\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)}$$

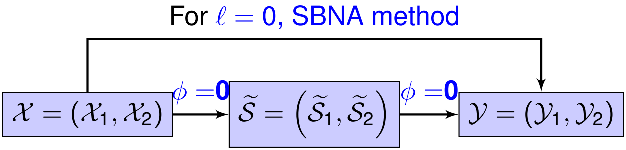
Main Results

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2) \xrightarrow{\phi = \mathbf{0}} \widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{S}}_2\right)$$

Main Results



Main Results



Main Results

Condition 2:

$\widetilde{\rho} = 0$ and $\phi \neq 0$ is a necessary and sufficient condition for minimising the risk of predictive inference in data coding models by perturbation.

consequences:

If $\phi \neq 0$ then $\lim_{\phi \neq 0} \widetilde{S} \to \pm\infty$ and $\sigma^2_{\widetilde{S}} \to +\infty$.

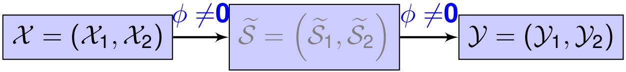
Main Results

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$$

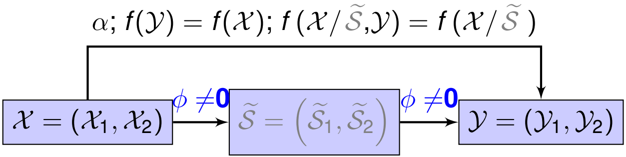
Main Results

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2) \xrightarrow{\phi \neq \mathbf{0}} \widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{S}}_2\right)$$

Main Results



Main Results



Main Results

Condition 3:

Condition 2 is robust in the multivariate case.

consequences:

We can *Cypher/Mask a random vector* $\mathcal{X} = (\mathcal{X}_1, ..., \mathcal{X}_K)$ for a random vector $\mathcal{S} = (\mathcal{S}_1, ..., \mathcal{S}_K)$.

Main Results

$$\mathcal{X} = (\mathcal{X}_1, ..., \mathcal{X}_K)$$

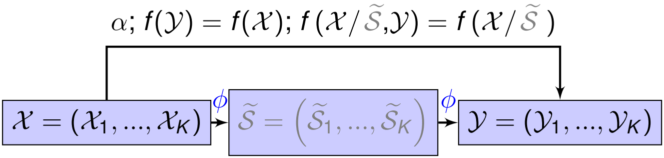
Main Results

$$\mathcal{X} = (\mathcal{X}_1, ..., \mathcal{X}_K) \xrightarrow{\phi} \widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, ..., \widetilde{\mathcal{S}}_K\right)$$

Main Results

$$\mathcal{X} = (\mathcal{X}_1, ..., \mathcal{X}_K) \xrightarrow{\phi} \widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, ..., \widetilde{\mathcal{S}}_K\right) \xrightarrow{\phi} \mathcal{Y} = (\mathcal{Y}_1, ..., \mathcal{Y}_K)$$

Main Results



Main Results

Condition 4:

Condition 2 and condition 3 are robust with respect to forecasts.

consequences:

It's possible to *forecast* a value with masked data then decoding the result. The decoding result is the result of a forecast with confidential data.

Main Results

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2)$$

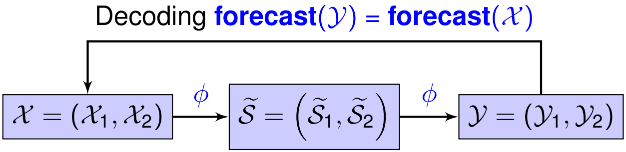
Main Results

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2) \xrightarrow{\phi} \widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{S}}_2\right)$$

Main Results

$$\mathcal{X} = (\mathcal{X}_1, \mathcal{X}_2) \xrightarrow{\phi} \widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, \widetilde{\mathcal{S}}_2\right) \xrightarrow{\phi} \mathcal{Y} = (\mathcal{Y}_1, \mathcal{Y}_2)$$

Main Results



Empirical Example: Results of a simulation basis (GADP, EGADP and IPSO linear and non linear)

• On one hand if $\ell = 0$ and the given ρ is then either $\phi = 0$, in which case γ, β and Σ_ϵ remain unchanged because $\widetilde{\rho}(\phi\rho)^{-1} = 1$, or if $\phi \neq 0$ and $\widetilde{\rho}(\phi\rho)^{-1} \approx 0$, in which case

$$\Sigma_{\widetilde{S}\widetilde{S}} \to \pm\infty$$
$$\gamma \to \mu_{\mathcal{X}}$$
$$\beta \to 0$$
$$\Sigma_\epsilon \to \Sigma_{\mathcal{X}\mathcal{X}}$$

• On the other hand, if $\ell \neq 0$ and the given ρ be then either $\phi = 0$, and in which case γ, β and Σ_ϵ vary because $\widetilde{\rho}(\phi\rho)^{-1} = (1 - \ell^2)$, or if $\phi \neq 0$ and $\widetilde{\rho}(\phi\rho)^{-1} \approx 0$, and in which case

$$\Sigma_{\widetilde{S}\widetilde{S}} \to \pm\infty$$
$$\gamma \to \mu_{\mathcal{X}}$$
$$\beta \to 0 \text{ and } \Sigma_\epsilon \to \Sigma_{\mathcal{X}\mathcal{X}}$$

Empirical Example: Results of a simulation basis (SBNA, MicroHybrid and PCA linear and non linear models)

On one hand, $\forall \ell \in [-1;1]$ and for the given ρ, if $\phi = 0$, then we again come upon the same results as the preceding ones, or nearly so by $(I - \alpha)$. On the other, if $\phi \neq 0$, we then obtain

$$\begin{array}{rcl} \gamma & \to & (I - \alpha)\,\mu_{\mathcal{X}} \\ \beta^T & \to & 0 \\ \Sigma_\epsilon & \to & \Sigma_{\mathcal{X}\mathcal{X}} - \alpha\Sigma_{\mathcal{X}\mathcal{X}}\alpha^T \end{array}$$

Empirical Example: Results of a simulation basis (Multivariate case)

Let $\ell = -2,00E - 10$, then $\ell = 0.6$ with a given $\rho = 0,87038828$ when $\alpha = \begin{pmatrix} 0,7 & 0 \\ 0 & 0,7 \end{pmatrix}$

For a *first-order* perturbation, we have:

$$y_i = \gamma + \alpha x_i + \underbrace{\beta_2}_{\#N/A} \phi_2 \rho \widetilde{\rho}^{-1} \varsigma_i + \beta_1 \phi_1 \rho \widetilde{\rho}^{-1} \varsigma_i + \epsilon_i, \forall i \in \{1...n\} \qquad \text{(r9)}$$

For a *second-order, 'dummy'* perturbation:

$$y_i = \gamma + \alpha x_i + \beta_2 \phi_2 \rho \widetilde{\rho}^{-1} \varsigma_i + \epsilon_i, \forall i \in \{1...n\} \qquad \text{(r10)}$$

and for a *third-order* perturbation:

$$y_i = \gamma + \alpha x_i + \underbrace{\beta_2}_{\#N/A} \phi_2 \rho \widetilde{\rho}^{-1} \varsigma_i + \beta_3 \phi_3 \rho \widetilde{\rho}^{-1} \varsigma_i + \epsilon_i, \forall i \in \{1...n\} \qquad \text{(r11)}$$

Empirical Example: Results of a simulation basis

★

Table 1. Covariances between Original, Masked and Perturbed non confidential data.

				SIMULATION A		SIMULATION B		SIMULATION C	
				$\phi_1 = 1,148912529$		$\phi_2 = 2,29782E-10$		$\phi_3 = -1,148912529$	
$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{Y}_1$	$\mathcal{Y}_2$	$\tilde{\mathcal{S}}_1$	$\tilde{\mathcal{S}}_2$	$\tilde{\mathcal{S}}_1$	$\tilde{\mathcal{S}}_2$	$\tilde{\mathcal{S}}_1$	$\tilde{\mathcal{S}}_2$
13, 1336116	3, 405629582	12, 99177362	0, 798138311	−1476822990	−1611185857	−0, 495433694	−0, 540508759	−673626014, 9	−734913199, 2
12, 21902051	0, 381739203	12, 38209476	2, 185487635	−1148640103	153446272, 1	−0, 385337318	0, 051477025	−523931344, 9	69991733, 26
12, 45489329	3, 918321291	12, 8913032	5, 311166729	−820457216, 5	652146656, 4	−0, 275240941	0, 218777355	−374236674, 9	297464866, 3
12, 65588765	2, 607636941	12, 29699151	0, 828069752	−492274329, 9	2416778785	−0, 165144565	0, 810763138	−224542005	1102369799
13, 1550851	1, 343879314	13, 48421775	4, 103615276	−164091443, 3	383615680, 2	−0, 055048188	0, 128692562	−74847334, 99	174979333, 1
12, 91117111	4, 329667476	12, 65936243	4, 422694477	164091443, 3	−1649547425	0, 055048188	−0, 553378015	74847334, 99	−752411132, 5
12, 40024436	−0, 743247498	12, 15790086	−1, 139169272	492274329, 9	−1150847041	0, 165144565	−0, 386077685	224542005	−524937999, 4
12, 66972201	3, 28571431	12, 16437319	2, 288800287	820457216, 5	613785088, 3	0, 275240941	0, 205908099	374236674, 9	279966933
12, 72553848	4, 881391268	12, 89529343	3, 857115875	1148640103	1112485473	0, 385337318	0, 373208429	523931344, 9	507440066, 1
11, 50835506	−2, 030608038	11, 91021843	−1, 27579522	1476822990	−920677632, 5	0, 495433694	−0, 308862148	673626014, 9	−419950399, 5
$\Sigma_{\mathcal{XX}}$		$\Sigma_{\mathcal{YY}}$				$\Sigma_{\mathcal{X}\tilde{\mathcal{S}}} = \Sigma_{\mathcal{Y}\tilde{\mathcal{S}}}$			
0, 210845383	0, 690522006	0, 210845383	0, 690522006	−180802242, 3	23927306, 2	−0, 060654204	0, 008026956	−82469662, 81	10914006, 65
0, 690522006	4, 803779739	0, 690522006	4, 803779739	−453851894, 9	697451625, 4	−1, 94289E − 17	0, 174892348	−207016308, 3	318129906, 1

Empirical Example: Results of a simulation basis

★★

Table 2. Covariances between Original, Masked and Perturbed non confidential data.

				SIMULATION D		SIMULATION E		SIMULATION F	
				$\phi_1 = 1,807425894$		$\phi_2 = 1,79518E-09$		$\phi_3 = -0,7303204$	
$\mathcal{X}_1$	$\mathcal{X}_2$	$\mathcal{Y}_1$	$\mathcal{Y}_2$	$\tilde{\mathcal{S}}_1$	$\tilde{\mathcal{S}}_2$	$\tilde{\mathcal{S}}_1$	$\tilde{\mathcal{S}}_2$	$\tilde{\mathcal{S}}_1$	$\tilde{\mathcal{S}}_2$
13, 1336116	3, 405629582	12, 99177362	0, 798138311	−701235298, 5	−765034403, 6	−0, 774115146	−0, 844544934	344089442, 6	375395052, 3
12, 21902051	0, 381739203	12, 38209476	2, 185487635	−545405232, 1	72860419, 39	−0, 602089558	0, 080432851	267625122	−35751909, 74
12, 45489329	3, 918321291	12, 8913032	5, 311166729	−389575165, 8	309656782, 4	−0, 43006397	0, 341839616	191160801, 4	−151945616, 4
12, 65588765	2, 607636941	12, 29699151	0, 828069752	−233745099, 5	1147551605	−0, 258038382	1, 266817402	114696480, 9	−563092578, 4
13, 1550851	1, 343879314	13, 48421775	4, 103615276	−77915033, 16	182151048, 5	−0, 086012794	0, 201082127	38232160, 29	−89379774, 35
12, 91117111	4, 329667476	12, 65936243	4, 422694477	77915033, 16	−783249508, 5	0, 086012794	−0, 864653147	−38232160, 29	384333029, 7
12, 40024436	−0, 743247498	12, 15790086	−1, 139169272	233745099, 5	−546453145, 5	0, 258038382	−0, 603246382	−114696480, 9	268139323
12, 66972201	3, 28571431	12, 16437319	2, 288800287	389575165, 8	291441677, 6	0, 43006397	0, 321731404	−191160801, 4	−143007639
12, 72553848	4, 881391268	12, 89529343	3, 857115875	545405232, 1	528238040, 6	0, 602089558	0, 583138169	−267625122	−259201345, 6
11, 50835506	−2, 030608038	11, 91021843	−1, 27579522	701235298, 5	−437162516, 4	0, 774115146	−0, 482597105	−344089442, 6	214511458, 4
$\Sigma_{\mathcal{X}\mathcal{X}}$		$\Sigma_{\mathcal{Y}\mathcal{Y}}$				$\Sigma_{\mathcal{X}\tilde{\mathcal{S}}} = \Sigma_{\mathcal{Y}\tilde{\mathcal{S}}}$			
0, 210845383	0, 690522006	0, 210845383	0, 690522006	−85849770, 25	11361328, 9	−0, 094772194	0, 012542119	42125659, 76	−5574895, 239
0, 690522006	4, 803779739	0, 690522006	4, 803779739	−215501093, 3	331168800, 9	−0, 237898264	0, 365587393	105744321, 8	−162501357, 8

Empirical Example: Results of a simulation basis

⋆

	Dependent variable:					
	Y1	Y2	Y1	Y2	Y1	Y2
	(1)	(2)	(3)	(4)	(5)	(6)
X1	0.700* (0.292)		0.700* (0.292)		0.700* (0.292)	
X2		0.700* (0.292)		0.700* (0.292)		0.700* (0.292)
S1A	-0.000(0.000)	-0.000(0.000)				
S2A	0.000(0.000)	0.000(0.000)				
S1B			-0.182(0.423)	-0.457(1.954)		
S2B			0.014(0.291)	0.401(1.491)		
S1C					-0.000(0.000)	-0.000(0.000)
S2C					0.000(0.000)	0.000(0.000)
Constant	3.775(3.671)	0.641(0.866)	3.775(3.671)	0.641(0.866)	3.775(3.671)	0.641(0.866)
Observations	10	10	10	10	10	10
R^2	0.580	0.548	0.580	0.548	0.580	0.548
Adjusted R^2	0.370	0.322	0.370	0.322	0.370	0.322
Residual Std. Error (df = 6)	0.384	1.903	0.384	1.903	0.384	1.903
F Statistic (df = 3; 6)	2.761	2.423	2.761	2.423	2.761	2.423
Note:						*p<0.1; **p<0.05; ***p<0.01

Empirical Example: Results of a simulation basis

★★

	Dependent variable:					
	Y1	Y2	Y1	Y2	Y1	Y2
	(1)	(2)	(3)	(4)	(5)	(6)
X1	0.700* (0.292)		0.700* (0.292)		0.700* (0.292)	
X2		0.700* (0.292)		0.700* (0.292)		0.700* (0.292)
S1D	-0.000(0.000)	-0.000(0.000)				
S2D	0.000(0.000)	0.000(0.000)				
S1E			-0.116(0.271)	-0.292(1.250)		
S2E			0.009(0.186)	0.257(0.954)		
S1F					0.000(0.000)	0.000(0.000)
S2F					-0.000(0.000)	-0.000(0.000)
Constant	3.775(3.671)	0.641(0.866)	3.775(3.671)	0.641(0.866)	3.775(3.671)	0.641(0.866)
Observations	10	10	10	10	10	10
R^2	0.580	0.548	0.580	0.548	0.580	0.548
Adjusted R^2	0.370	0.322	0.370	0.322	0.370	0.322
Residual Std. Error (df = 6)	0.384	1.903	0.384	1.903	0.384	1.903
F Statistic (df = 3; 6)	2.761	2.423	2.761	2.423	2.761	2.423

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Plan

- Open Questions

$\star$ Open questions on predictive inference risk requirements for $\widetilde{\rho} \neq 0$.

$\star$ Open question on the possibility of re-identifying a target variable due to the fact that we can generate a random vector $\widetilde{\mathcal{S}} = \left(\widetilde{\mathcal{S}}_1, ..., \widetilde{\mathcal{S}}_K\right) \Longrightarrow$ a Complementary Unsupervised learning approach.

Thank U ! ! !

lab. TRIS-UM1 (Traitement et Recherche sur l'Information et la Statistique)
LAMPS-UPVD (LAboratoire de Mathématiques et de PhySique)
Scientific and Organizing Committee of JDS 2018
Scientific and Organizing Committee of the JMS 2018
Scientific and Organizing Committee of S^4D 2018.