



HAL
open science

Robust Lasso-Zero for sparse corruption and model selection with missing covariates

Pascaline Descloux, Claire Boyer, Julie Josse, Aude Sportisse, Sylvain Sardy

► **To cite this version:**

Pascaline Descloux, Claire Boyer, Julie Josse, Aude Sportisse, Sylvain Sardy. Robust Lasso-Zero for sparse corruption and model selection with missing covariates. *Scandinavian Journal of Statistics*, 2022. hal-02569696v2

HAL Id: hal-02569696

<https://hal.science/hal-02569696v2>

Submitted on 22 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust Lasso-Zero for sparse corruption and model selection with missing covariates

Pascaline Descloux⁽¹⁾, Claire Boyer^(2,3), Julie Josse^(4,5), Aude Sportisse^(6,7), and Sylvain Sardy⁽¹⁾

⁽¹⁾ *Department of Mathematics, University of Geneva, Switzerland*

⁽²⁾ *Sorbonne Université, CNRS, Laboratoire de Probabilités, Statistique et Modélisation (LPSM), F-75005 Paris, France,*

⁽³⁾ *MOKAPLAN, INRIA Paris*

⁽⁴⁾ *Inria centre at Université Côte d'Azur, branch of Montpellier, France*

⁽⁵⁾ *Institut Desbrest d'Épidémiologie et de Santé Publique, Montpellier, France*

⁽⁶⁾ *Inria centre at Université Côte d'Azur, Maasai, Valbonne, France*

⁽⁷⁾ *Interdisciplinary Institute for Artificial Intelligence, 3iA Côte d'Azur, France*

March 22, 2022

Abstract

We propose Robust Lasso-Zero, an extension of the Lasso-Zero methodology, initially introduced for sparse linear models, to the sparse corruptions problem. We give theoretical guarantees on the

sign recovery of the parameters for a slightly simplified version of the estimator, called Thresholded Justice Pursuit. The use of Robust Lasso-Zero is showcased for variable selection with missing values in the covariates. In addition to not requiring the specification of a model for the covariates, nor estimating their covariance matrix or the noise variance, the method has the great advantage of handling missing not-at random values without specifying a parametric model. Numerical experiments and a medical application underline the relevance of Robust Lasso-Zero in such a context with few available competitors. The method is easy to use and implemented in the R library `lass0`.

Keywords: incomplete data, informative missing values, Lasso-Zero, sparse corruptions, support recovery

1 Introduction

Consider the framework of sparse linear models in high dimension,

$$y = X\beta^0 + \epsilon, \quad (1)$$

where $y \in \mathbb{R}^n$ is the vector of observations, $X \in \mathbb{R}^{n \times p}$ is the design matrix, and $\beta^0 \in \mathbb{R}^p$ is the parameter of interest, assumed to be s -sparse (only s out of its p entries are different from zero). The additive noise $\epsilon \in \mathbb{R}^n$ is a classical noise vector, qualified of “dense noise” and assumed to be Gaussian with covariance matrix $\sigma^2 I_n$. In case of additional occasional corruptions, the sparse corruption problem can be then written as

$$y = X\beta^0 + \sqrt{n}\omega^0 + \epsilon, \quad (2)$$

where $\omega^0 \in \mathbb{R}^n$ is a k -sparse corruption vector; see for instance [Chen et al. \[2013a\]](#). Noting that (2) can be rewritten as

$$y = \begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix} \begin{bmatrix} \beta^0 \\ \omega^0 \end{bmatrix} + \epsilon,$$

the sparse corruption model can be seen as a sparse linear model with an augmented design matrix and an augmented sparse vector of parameters. We are interested in theoretical guarantees of support recovery for β^0 in (2), with beneficial consequences for variable selection with missing covariates.

Related literature. To recover β^0 when there is no dense noise (when $\epsilon = 0$ or equivalently when $\sigma^2 = 0$), several authors proposed the *Justice Pursuit* (JP) program, name coined by [Laska et al. \[2009\]](#), by solving

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p, \omega \in \mathbb{R}^n} \quad & \|\beta\|_1 + \|\omega\|_1 \\ \text{s.t.} \quad & y = X\beta + \omega, \end{aligned} \quad (3)$$

which is nothing else than the *Basis Pursuit* (BP) problem, with the augmented matrix $\begin{bmatrix} X & I_n \end{bmatrix}$ [Wright et al., 2009]. Wright and Ma [2010] analyzed JP for Gaussian measurements, providing support recovery results when $n \simeq p$ using cross-polytope arguments. Besides, Laska et al. [2009] and Li et al. [2010] proved that if the entries of X are i.i.d. standard Gaussian as well, then the matrix $\begin{bmatrix} X & I_n \end{bmatrix}$ satisfies some restricted isometry property with high probability, implying exact recovery of both β^0 and ω^0 , provided that $n \gtrsim (s + k) \log(p)$. However, in these works, the sparsity level k of ω^0 cannot be fixed to a proportion of the sample size n . Therefore, Li [2013] and Nguyen and Tran [2013b] introduced a tuning parameter $\lambda > 0$ and solve

$$\min_{\beta \in \mathbb{R}^p, \omega \in \mathbb{R}^n} \|\beta\|_1 + \lambda \|\omega\|_1 \quad \text{s.t.} \quad y = X\beta + \omega. \quad (4)$$

In a sub-orthogonal or Gaussian design, they both proved exact recovery, even for a large proportion of corruption.

In the case of sparse and dense noise (i.e. when $\omega^0 \neq 0$ and $\sigma^2 > 0$), Nguyen and Tran [2013a] proposed to jointly estimate β^0 and ω^0 by solving

$$\min_{\beta \in \mathbb{R}^p, \omega \in \mathbb{R}^n} \frac{1}{2} \|y - X\beta - \omega\|_2^2 + \lambda_\beta \|\beta\|_1 + \lambda_\omega \|\omega\|_1. \quad (5)$$

In the special case where $\lambda_\beta = \lambda_\omega$, problem (5) boils down to the Lasso [Tibshirani, 1996] applied to the response y and the design matrix $\begin{bmatrix} X & I_n \end{bmatrix}$. Assuming a standard Gaussian design and the invertibility and incoherence properties for the covariance matrix, they obtained sign recovery guarantee for an arbitrarily large fraction of corruption, provided that $n \geq Ck \log(p) \log(n)$. In addition, the required number of samples is proven to be optimal. More recently in the case of a Gaussian design with an invertible covariance matrix, Dalalyan and Thompson [2019] obtained an

optimal rate of estimation of β^0 when considering an ℓ_1 -penalized Huber's M -estimator, which is actually equivalent to (5) [Sardy et al., 2001].

Contributions. To estimate the support of the parameter vector β^0 in the sparse corruption problem, we study an extension of the Lasso-Zero methodology [Descloux and Sardy, 2021], initially introduced for standard sparse linear models, to the sparse corruptions problem. We provide theoretical guarantees on the sign recovery of β^0 for a simplified version of the Robust Lasso-Zero, that we call the Thresholded Justice Pursuit (TJP). These guarantees are extensions of recent results on the Thresholded Basis Pursuit. The first one extends a result of Tardivel and Bogdan [2019], providing a necessary and sufficient condition for consistent recovery in a setting where the design matrix is fixed but the nonzero absolute coefficients tend to infinity. The second one extends a result of Descloux and Sardy [2021], proving sign consistency for correlated Gaussian designs when p , s and k grow with n , allowing a positive fraction of corruptions.

In this paper, we also consider the case where the matrix of covariates X contains missing values, which can be due to manual errors, poor calibration, insufficient resolution, etc. In the high-dimensional setting, note that the naive complete case analysis [Rubin, 1976], which discards all incomplete rows, is not an option, because the missingness of a single entry causes the loss of an entire row, which contains a lot of information when p is large [Zhu and Marcotte, 1996]. Showing that missing values in the covariates can be reformulated into a sparse corruption problem, we recommend the Robust Lasso-Zero for dealing with missing data. For support recovery, this approach requires neither to specify a model for the covariates or the missing data mechanism, nor an estimation of the co-

covariates covariance matrix or of the noise variance, and hence provides a simple (quasi hyperparameter-free) method for the user. Numerical experiments and a medical application also underline the effectiveness of Robust Lasso-Zero with respect to few available competitors, especially when the missing-data mechanism is MNAR (for Missing Not At Random), i.e. when data are said informative, since their own values can cause the missingness.

Organization. After defining the Robust Lasso-Zero methodology in Section 2, we analyse the sign recovery properties of the Thresholded Justice Pursuit in Section 2.3. Section 3.1 is dedicated to variable selection with missing values and the selection of tuning parameters is discussed in Section 3.2. Numerical experiments are presented in Section 4 and an application in Section 5.

Notation. Define $[p] := \{1, \dots, p\}$, and the complement of a subset $S \subset [p]$ is denoted \bar{S} . For a matrix A of size $u \times v$ and a set $T \subset [v]$, we use A_T to denote the submatrix of size $u \times |T|$ with columns indexed by T . We define the missing value indicator matrix $M \in \mathbb{R}^{n \times p}$ by $M_{ij} = \mathbf{1}_{\{X_{ij}^{\text{NA}} = \text{NA}\}}$, and the set of incomplete rows by $\mathcal{M} := \{i \in [n] \mid M_{ij} = 1 \text{ for some } j \in [p]\}$.

2 Robust Lasso-Zero

2.1 Lasso-Zero in a nutshell

Under the linear model (1), the Thresholded Basis Pursuit (TBP) estimates the parameter β^0 by setting the small coefficients of the BP solution to zero. Since BP fits the observations y exactly, noise is generally over-

fitted. The Lasso-Zero [Descloux and Sardy, 2021] alleviates this issue by solving repeated BP problems, respectively fed with the augmented matrices $[X|G^{(k)}]$, where $G^{(k)} \in \mathbb{R}^{n \times n}$, $k = 1, \dots, M$, are different i.i.d. Gaussian noise dictionaries meaning that the columns of $G^{(k)}$ form a set of random vectors used to approximate the additive noise term in the observations. Note that each entry of $G^{(k)}$ is drawn according to a standard Gaussian distribution $\mathcal{N}(0, 1)$. Once the noise dictionaries generated, the corresponding obtained estimates $\hat{\beta}^{(1)}, \dots, \hat{\beta}^{(M)}$ are then aggregated by taking the component-wise medians, further thresholded at level $\tau > 0$. Descloux and Sardy [2021] show that the Lasso-Zero tuned by Quantile Universal Thresholding [Giacobino et al., 2017] achieves a very good trade-off between high power and low false discovery rate compared to competitors.

2.2 Definition of Robust Lasso-Zero

The Robust Lasso-Zero approach is an extension of the Lasso-Zero methodology for the sparse corruptions problem. Consider the sparse corruption model (2), for which S^0 and T^0 denote the respective supports of β^0 and ω^0 , with $s := |S^0|$ and $k := |T^0|$ their respective sparsity degrees.

To fix the notation, we then consider the following parameterization of Justice Pursuit (JP): To fix the notation, we then consider the following parameterization of Justice Pursuit (JP):

$$(\hat{\beta}_\lambda^{\text{JP}}, \hat{\omega}_\lambda^{\text{JP}}) \in \arg \min_{\beta \in \mathbb{R}^p, \omega \in \mathbb{R}^n} \|\beta\|_1 + \lambda \|\omega\|_1 \quad \text{s.t.} \quad y = X\beta + \sqrt{n}\omega. \quad (6)$$

Renormalization by \sqrt{n} balances the augmented design matrix $\begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix}$: in practice the columns of X are often standardized so that $\|X_j\|_2^2 = n$ for every $j \in [p]$, (or at least this is true in expectation) and in this way, the norms of all columns of $\begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix}$ scale on the same order.

While the original Lasso-Zero algorithm was developed to solve standard sparse regression, and built upon solving a sequence of Basis Pursuit programs involving noise dictionaries, the Robust Lasso-Zero is designed to address the sparse corruptions setting, by solving a sequence of Justice Pursuit programs (6) involving noise dictionaries (see (7)) and a median aggregation (see stage 2) in Algorithm 1). The Robust Lasso-Zero approach is fully described in Algorithm 1. In the latter, attention has been paid to the estimation of the support of β^0 . However the estimation of the corruption support is also possible by computing the corresponding vectors $\hat{\omega}_\lambda^{\text{med}}$ and $\hat{\omega}_{(\lambda,\tau)}^{\text{Rlass0}}$, at stages 2) and 3).

Algorithm 1 Robust Lasso-Zero

Given data (y, X) , for given hyper-parameters $\lambda > 0, \tau \geq 0$ and $M \in \mathbb{N}^*$:

1) For $k = 1, \dots, M$:

- i) generate a matrix $G^{(k)}$ of size $n \times n$ with i.i.d. $\mathcal{N}(0, 1)$ entries
- ii) compute the solution $(\hat{\beta}_\lambda^{(k)}, \hat{\omega}_\lambda^{(k)}, \hat{\gamma}_\lambda^{(k)})$ to the augmented JP problem

$$\begin{aligned}
 (\hat{\beta}_\lambda^{(k)}, \hat{\omega}_\lambda^{(k)}, \hat{\gamma}_\lambda^{(k)}) \in & \arg \min_{\beta \in \mathbb{R}^p, \omega \in \mathbb{R}^n, \gamma \in \mathbb{R}^n} \|\beta\|_1 + \lambda \|\omega\|_1 + \|\gamma\|_1 \\
 \text{s.t.} & y = X\beta + \sqrt{n}\omega + G^{(k)}\gamma.
 \end{aligned} \tag{7}$$

2) Define the vector $\hat{\beta}_\lambda^{\text{med}}$ by

$$\hat{\beta}_{\lambda,j}^{\text{med}} := \text{median}\{\hat{\beta}_{\lambda,j}^{(k)}, k = 1, \dots, M\} \quad \text{for every } j \in [p].$$

3) Calculate the estimate $\hat{\beta}_{(\lambda,\tau)}^{\text{Rlass0}} := \eta_\tau(\hat{\beta}_\lambda^{\text{med}})$, where $\eta_\tau(x) = x\mathbf{1}_{(\tau,+\infty)}(|x|)$ hard-thresholds component-wise.

Since the minimization problem (7) in Algorithm 1 can be recast as a linear program, any relevant solver can be used (e.g., proximal methods). Algorithm 1 includes three apparent hyper-parameters: the number M of samplings, the regularization parameter λ of (6), the thresholding parameter τ of the Robust Lasso-Zero methodology. Their choice in practice is discussed in Section 3.2, making the methodology quasi hyperparameter-free.

2.3 Theoretical guarantees on the Thresholded Justice Pursuit

Discarding the noise dictionaries in Algorithm 1 amounts to thresholding the solution $(\hat{\beta}_\lambda^{\text{JP}}, \hat{\omega}_\lambda^{\text{JP}})$ of the Justice Pursuit problem (6). The Robust Lasso-Zero can therefore be regarded as an extension of this simpler estimator, via *Thresholded Justice Pursuit* (TJP):

$$\hat{\beta}_{(\lambda,\tau)}^{\text{TJP}} = \eta_\tau(\hat{\beta}_\lambda^{\text{JP}}) \quad \text{and} \quad \hat{\omega}_{(\lambda,\tau)}^{\text{TJP}} = \eta_\tau(\hat{\omega}_\lambda^{\text{JP}}). \quad (8)$$

We present two results about sign consistency of TJP. Note that the statistical analysis of Robust Lasso-Zero methodology would be more mathematically involved, and is actually beyond the scope of this paper.

2.3.1 Identifiability as a necessary and sufficient condition for consistent sign recovery

First introduced in [Tardivel and Bogdan \[2019\]](#) for the TBP, we propose the following extension of the identifiability notion for the TJP characterized by the uniqueness of the solution to JP.

Definition 1. The pair $(\beta^0, \omega^0) \in \mathbb{R}^p \times \mathbb{R}^n$ is said to be *identifiable* with respect to $X \in \mathbb{R}^{n \times p}$ and the parameter $\lambda > 0$ if it is the unique solution to

JP (6) when $y = X\beta^0 + \sqrt{n}\omega^0$.

In the case of JP, we provide the following characterization of identifiability, which depends on the sign vectors $\text{sign}(\beta^0)$ and $\text{sign}(\omega^0)$.

Lemma 1. *The pair $(\beta^0, \omega^0) \in \mathbb{R}^p \times \mathbb{R}^n$ is identifiable with respect to $X \in \mathbb{R}^{n \times p}$ and the parameter $\lambda > 0$ if and only if for every pair $(\beta, \omega) \neq (0, 0)$ such that $X\beta + \sqrt{n}\lambda^{-1}\omega = 0$,*

$$|\text{sign}(\beta^0)^T \beta + \text{sign}(\omega^0)^T \omega| < \|\beta_{\overline{S^0}}\|_1 + \|\omega_{\overline{T^0}}\|_1.$$

Proof. See Appendix A. □

Note that recently, [Schneider and Tardivel \[2020\]](#) provide a necessary and sufficient condition for the uniqueness of the solution to BP and show that the set of design matrices which do not satisfy this condition is negligible with respect to the Lebesgue measure on $\mathbb{R}^{n \times p}$.

Remark 1 (Identifiability of JP vs. BP). Let us observe that (β_0, ω_0) is identifiable for JP w.r.t. X and λ then β_0 is identifiable for BP w.r.t. X (i.e. in the noiseless case and without corruptions, β_0 can be recovered by solving BP). Moreover, when the observations are not corrupted (i.e. when $y = X\beta_0 + \epsilon$), the identifiability of β_0 for BP w.r.t. X is actually a necessary and sufficient condition for sign recovery by thresholded BP (see [Tardivel and Bogdan \[2019\]](#)). Consequently, this corroborates the intuitive idea that sign recovery is a more difficult task via thresholded JP in the corrupted case than via thresholded BP in the non-corrupted case.

In order to show that identifiability is necessary and sufficient for TJP to consistently recover $\text{sign}(\beta^0)$ and $\text{sign}(\omega^0)$, we consider a fixed matrix $X \in \mathbb{R}^{n \times p}$ and a sequence of parameters $\{(\beta^{(r)}, \omega^{(r)})\}_{r \in \mathbb{N}^*}$ such that the following holds:

- (i) there exist sign vectors $\theta \in \{1, -1, 0\}^p$ and $\tilde{\theta} \in \{1, -1, 0\}^n$ such that $\text{sign}(\beta^{(r)}) = \theta$ and $\text{sign}(\omega^{(r)}) = \tilde{\theta}$ for every $r \in \mathbb{N}^*$,
- (ii) $\lim_{r \rightarrow +\infty} \min\{\beta_{\min}^{(r)}, \omega_{\min}^{(r)}\} = +\infty$, where $\beta_{\min}^{(r)} := \min_{j \in \text{supp}(\beta)} |\beta_j^{(r)}|$, and $\omega_{\min}^{(r)} := \min_{j \in \text{supp}(\omega)} |\omega_j^{(r)}|$.
- (iii) there exists $q > 0$ such that $\frac{\min\{\beta_{\min}^{(r)}, \omega_{\min}^{(r)}\}}{\max\{\|\beta^{(r)}\|_{\infty}, \|\omega^{(r)}\|_{\infty}\}} \geq q$.

Note that these assumptions are similar to the ones of [Tardivel and Bogdan \[2019\]](#). Assumption (i) requires that each parameter of the sequence $\{(\beta^{(r)}, \omega^{(r)})\}_{r \in \mathbb{N}^*}$ have invariant sign vectors. For a fixed design matrix and a fixed λ , Assumptions (ii) and (iii) are asymptotic statements, meaning that the nonzero absolute components of the sequence parameters $\{(\beta^{(r)}, \omega^{(r)})\}_{r \in \mathbb{N}^*}$ tend to infinity in a certain sense. We use the notation $S^0 := \text{supp}(\theta) = \text{supp}(\beta^{(r)})$ and $T^0 := \text{supp}(\tilde{\theta}) = \text{supp}(\omega^{(r)})$. For each parameter $(\beta^{(r)}, \omega^{(r)})$, $r \in \mathbb{N}^*$, let us consider the sparse corruption problem $y^{(r)} := X\beta^{(r)} + \sqrt{n}\omega^{(r)} + \epsilon$. We denote by $(\hat{\beta}_{\lambda}^{\text{JP}(r)}, \hat{\omega}_{\lambda}^{\text{JP}(r)})$ the JP solution when $y = y^{(r)}$ and $(\hat{\beta}_{(\lambda, \tau)}^{\text{TJP}(r)}, \hat{\omega}_{(\lambda, \tau)}^{\text{TJP}(r)})$ the corresponding TJP estimates.

Theorem 1. *Let $\lambda > 0$ and let X be a matrix of size $n \times p$ such that for any $y \in \mathbb{R}^n$, the solution to JP (6) is unique.*

Necessary condition: *If there exists $\tau \geq 0$ such that*

$$\text{sign}(\hat{\beta}_{\lambda, \tau}^{\text{TJP}}) = \text{sign}(\beta_0) \quad \text{and} \quad \text{sign}(\hat{\omega}_{\lambda, \tau}^{\text{TJP}}) = \text{sign}(\omega_0)$$

then (β_0, ω_0) is identifiable with respect to X and λ .

Sufficient condition (asymptotically): *Let $\{(\beta^{(r)}, \omega^{(r)})\}_{r \in \mathbb{N}^*}$ be a sequence satisfying Assumptions (i)-(iii) above. If the pair of sign vectors (θ, θ') is identifiable with respect to X and λ , then for every $\epsilon \in \mathbb{R}^n$, there exists $R = R(\epsilon) > 0$ such that for every $r \geq R$ there is a threshold $\tau = \tau(r) > 0$ for which*

$$\text{sign}(\hat{\beta}_{\lambda, \tau}^{\text{TJP}}) = \theta \quad \text{and} \quad \text{sign}(\hat{\omega}_{\lambda, \tau}^{\text{TJP}}) = \theta'. \quad (9)$$

Proof. See Appendix A. □

Remark 2. One might be interested in recovering the signs of the sparse corruption. If $\omega^{(r)}$ is considered as noise, then only the recovery of $\text{sign}(\beta^{(r)})$ matters. In this case one could weaken Assumptions (ii) and (iii) above by replacing $\min\{\beta_{\min}^{(r)}, \omega_{\min}^{(r)}\}$ by $\beta_{\min}^{(r)}$, and identifiability of $(\theta, \tilde{\theta})$ would be sufficient for recovering $\text{sign}(\beta^0)$. However, recovery of both $\text{sign}(\beta^{(r)})$ and $\text{sign}(\omega^{(r)})$ is needed for proving necessity of identifiability.

Identifiability of sign vectors is necessary and sufficient for sign recovery when the nonzero coefficients are large. In the next section, we particularize these results to Gaussian designs, and prove that sign consistency holds, allowing p, s and k to grow with the sample size n , in a classical way.

2.3.2 Sign consistency of TJP for correlated Gaussian designs

We make the following assumptions:

- (iv) the rows of $X \in \mathbb{R}^{n \times p}$ (with $n < p$) are random and i.i.d. $\mathcal{N}(0, \Sigma)$;
- (v) The smallest eigenvalue of the covariance matrix Σ is assumed to be positive: $\lambda_{\min}(\Sigma) > 0$,
- (vi) the variance of the covariates is equal to one: $\Sigma_{ii} = 1$ for every $i \in [p]$;
- (vii) the noise is assumed to be Gaussian $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ and independent of the columns of X .

Assumptions (iv) and (v) imply that almost surely $\text{rank } X = n$.

Proposition 2. *Under Assumptions (iv)-(vii), choosing $\lambda = \frac{1}{\sqrt{\log p}}$ ensures with probability greater than $1 - ce^{-c'n} - 1.14^{-n} - 2e^{-\frac{1}{8}(\sqrt{p}-\sqrt{n})^2}$, that there exists a value of $\tau > 0$ such that*

$$\text{sign}(\hat{\beta}_{(\lambda, \tau)}^{\text{TJP}}) = \text{sign}(\beta^0),$$

provided that

$$n \geq C \frac{\kappa(\Sigma)}{\lambda_{\min}(\Sigma)} s \log p, \quad (10)$$

$$\frac{n}{k} \geq \max \left\{ \frac{1}{C'}, \frac{\kappa(\Sigma)}{C''} \right\}, \quad (11)$$

$$\beta_{\min}^0 > \frac{10\sqrt{2} \max\{1, \lambda\} \sigma \sqrt{p+n}}{\left(\frac{\lambda_{\min}(\Sigma)}{4} (\sqrt{p/n} - 1)^2 + 1 \right)^{1/2}}, \quad (12)$$

where $\kappa(\Sigma) := \frac{\lambda_{\max}(\Sigma)}{\lambda_{\min}(\Sigma)}$ is the conditioning number of Σ , and C, C', C'' are some numerical constants with $C \geq 144^2$.

Proof. See Appendix B. □

Note that the lower-bound required on β_{\min}^0 in Proposition 2 is strong but complies with the one required for TBP in [Descloux and Sardy, 2021]. We believe that this limitation, which is not observed in simulations, is a proof artefact that could be relaxed. Besides, the advantage of noise dictionaries used in the Lasso-Zero in [Descloux and Sardy, 2021] (and thus for the Robust Lasso-Zero) is twofold: first it allows to better separate signal from noise, second it allows a form of resampling which, in the spirit of stability selection, allows to better identify needles in the haystack.

Note also that the conditioning number of Σ comes into play in the lower-bounds required on n and k . This quantity indeed seems natural to arise in the sparse corruption problem helping discriminating design instability from corruptions.

Corollary 3 trivially follows from Proposition 2 and ensures that, for correlated Gaussian designs and signal-to-noise ratios high enough, if Σ is well-conditioned, TJP successfully recovers $\text{sign}(\beta^0)$ with high probability, even with a positive fraction of corruptions.

Corollary 3. *Let $X \in \mathbb{R}^{n \times p}$ be a Gaussian matrix satisfying Assumptions (iv)-(vii). Assume also that the eigenvalues of covariance matrix Σ are bounded $0 < \gamma_1 \leq \lambda_{\min}(\Sigma) \leq \lambda_{\max}(\Sigma) \leq \gamma_2$.*

Sufficient condition: *If $p/n \rightarrow \delta > 1$, one has*

$$\mathbb{P} \left(\exists \tau > 0, \text{sign}(\hat{\beta}_{(\lambda, \tau)}^{\text{TJP}}) = \text{sign}(\beta^0) \right) \geq 1 - (c + 2e^{-\frac{1}{8}p})e^{-c'n},$$

with c and c' some numerical constants and provided that $n = \Omega(s \log p)$, $k = \mathcal{O}(n)$ and $\beta_{\min}^0 = \Omega(\sqrt{n})$ ¹.

3 Model selection with missing covariates

In practice the matrix of covariates X is often partially known and one only observes an incomplete matrix, denoted X^{NA} .² High dimensional variable selection with missing values turns out to be a challenging problem and very few solutions are available, not to mention implementations. Available solutions either require strong assumptions on the missing value mechanism, a lot of parameters tuning or strong assumption on the covariates distribution which is hard in high dimensions. They include the Expectation-Maximization algorithm [Dempster et al., 1977] for sparse linear regression [Garcia et al., 2010] and regression imputation methods [Van Buuren, 2018].

¹Recall that $f(n) = \Omega(g(n))$ (resp. $f(n) = \mathcal{O}(g(n))$) means that there exists $\kappa > 0$ (resp. $\kappa' > 0$) such that $|f(n)| \geq \kappa|g(n)|$ (resp. $|f(n)| \leq \kappa'|g(n)|$) for large n enough.

²NA, for Not Available, is a usual symbol for missing values.

A method combining penalized regression techniques with multiple imputation and stability selection has been developed [Liu et al., 2016]. Yet, aggregating different models for the resulting multiple imputed data sets becomes increasingly complex as the number of data grows. Rosenbaum et al. [2013] modified the Dantzig selector by using a consistent estimation of the design covariance matrix. Following the same idea, Loh and Wainwright [2012] and Datta and Zou [2017] reformulated the Lasso also using an estimate of the design covariance matrix, possibly resulting in a non-convex problem. Chen and Caramanis [2013] presented a variant of orthogonal matching pursuit which recovers the support and achieves the minimax optimal rate. Jiang et al. [2019] proposed Adaptive Bayesian SLOPE, combining SLOPE and Spike-and-Slab Lasso. While some of these methods have interesting theoretical guarantees, they all require an estimation of the design covariance matrix, which is often obtained under the restrictive MCAR (Missing Completely At Random) assumption, when the missingness does not depend on the data.

3.1 Relation to the sparse corruption model

To tackle the problem of estimating the sparse model parameter when the design matrix is incomplete, we suggest an easy-to-implement solution for the user, which consists in imputing the missing entries in X^{NA} with the imputation of his choice to get a completed matrix \tilde{X} , such as the mean imputation, and to take into account the impact of the occasional poor imputation as follows. Given the matrix \tilde{X} , the linear model (1) can be rewritten in the form of the sparse corruption model (2), where $\omega^0 := \frac{1}{\sqrt{n}}(X - \tilde{X})\beta^0$ is the (unknown) corruption due to imputations. In classical (i.e. non-sparse) regression, one could not say much about

ω^0 without any prior knowledge of the distribution of the covariates or the missing data mechanism. Since the key point here is that when β^0 is sparse, then so is ω^0 , even if all rows of the design matrix contain missing entries. Indeed, for every $i \in [n]$,

$$\omega_i^0 = \frac{1}{\sqrt{n}} \sum_{j=1}^p (X_{ij} - \tilde{X}_{ij}) \beta_j^0 = \frac{1}{\sqrt{n}} \sum_{j \in S^0} (X_{ij} - \tilde{X}_{ij}) \beta_j^0, \quad (13)$$

so ω_i^0 is nonzero only if the i^{th} row of X^{NA} contains missing value(s) on the support S^0 , since $(X_{ij} - \tilde{X}_{ij}) = 0$ if X_{ij} is observed. So the problem of missing covariates can be rephrased as a sparse corruption problem, as already pointed out in [Chen et al. \[2013b\]](#). We propose to use the Robust Lasso-Zero methodology presented in Section 2.2, specifically designed to solve the sparse corruption problem, see Algorithm 2.

Note that if the i^{th} row of X is fully observed, then $\omega_i^0 = 0$ by (13). Thus the dimension of ω^0 can be reduced by restricting it to the incomplete rows of X^{NA} . The corruption vector ω^0 is now of size $|\mathcal{M}|$ and (2) becomes

$$y = X\beta^0 + \sqrt{n}I_{\mathcal{M}}\omega^0 + \epsilon, \quad (14)$$

where $I_{\mathcal{M}} \in \mathbb{R}^{n \times |\mathcal{M}|}$ is the submatrix of the identity matrix $I_n \in \mathbb{R}^{n \times n}$ with columns indexed by $i \in \mathcal{M}$.

Algorithm 2 Robust Lasso-Zero for missing data

Given data (y, X^{NA}) , for given hyper-parameters $\lambda > 0, \tau \geq 0$ and $M \in \mathbb{N}^*$:

- 1) Impute X^{NA} and rescale the imputed matrix X such that all columns have Euclidean norm equal to \sqrt{n} .
 - 2) Run Algorithm 1 with the design matrix X .
-

3.2 Selection of tuning parameters

Formally, Algorithm 2 requires three apparent hyper-parameters: M , λ and τ .

The parameter M is the number of noise dictionaries also used to perform the median estimation. It controls the precision in estimating the marginal medians and does not play an important role: it can be seen as an analogue of the number of bootstrap samples in a bootstrap procedure. As a matter of fact, as it exists rules of thumb for bootstrap samples, we recommend $M = 20$ -as we used in our experiments- providing satisfactory precision. However, it is important to take M larger than one (see Figure 3 in [Descloux and Sardy, 2021] for instance).

The parameter λ in (6) tunes the balance between the corruption regularization and the model parameter one. For a fair isotropic penalty on β, ω and γ , we fix $\lambda = 1$ (even if Proposition 2 and Corollary 3 do not provide theoretical guarantees for $\lambda = 1$).

Finally, the thresholding parameter τ can be selected through the Quantile Universal Threshold (QUT) [Giacobino et al., 2017] methodology, generally driven by model selection rather than prediction. Indeed, under the null model, no sparse corruption exists: indeed if $\beta^0 = 0$, so is ω^0 since $\omega^0 = \frac{1}{\sqrt{n}}(X - \tilde{X})\beta^0 = 0$. QUT selects the tuning parameter so that under the null model ($\beta^0 = 0$), the null vector $\hat{\beta} = 0$ is recovered with probability $1 - \alpha$. Under the null model, $y = \epsilon$ whatever the missing data pattern is. Then given a fixed value of λ and a fixed imputed matrix \tilde{X} , the corresponding QUT value of τ is the upper α -quantile of $\|\hat{\beta}_\lambda^{\text{med}}(\epsilon)\|_\infty$, where $\hat{\beta}_\lambda^{\text{med}}(\epsilon)$ is the vectors of medians obtained at stage 2) of Algorithm 1 applied to \tilde{X} and $y = \epsilon$. To free ourselves from preliminary estimation of the noise level σ , we exploit the noise coefficients $\hat{\gamma}^{(k)}$ of Robust Lasso-Zero

to pivotize the statistic $\|\hat{\beta}_\lambda^{\text{med}}(\epsilon)\|_\infty$, as explained in [Descloux and Sardy \[2021\]](#). Thus, the parameter τ , through the QUT methodology, is depending on λ (that we choose equal to 1) and the FDR α under the null hypothesis. As usual, one can choose a priori a FDR of $\alpha = 0.05$, or smaller depending on the application.

Despite all these apparent hyperparameters, the method can be still considered as relying on mild tuning as the only important choice is the choice of τ by the QUT methodology.

4 Numerical experiments

We evaluate the performance of the Robust-Lasso Zero when missing data affect the design matrix. The code reproducing these experiments is available at <https://github.com/pascalinedescloux/robust-lasso-zero-NA>.

4.1 Simulation settings

Simulation scenarios. We generate data according to model (1) with the covariates matrix obtained by drawing $n = 100$ observations from a Gaussian distribution $\mathcal{N}(0, \Sigma)$, where $\Sigma \in \mathbb{R}^{200 \times 200}$ is a Toeplitz matrix, such that $\Sigma_{ij} = \rho^{|i-j|}$; the variance of the noise $\sigma = 0.5$ and the coefficient β^0 are drawn uniformly from $\{\pm 1\}$. We make the following parameters vary:

- Correlation structures indexed by ρ with $\rho = 0$ (uncorrelated) and $\rho = 0.75$ (correlated);
- Sparsity degrees indexed by s with $s \in \{3, 10\}$.

Before generating the response vector y , all columns of X are mean-centered and standardized. Missing data are then introduced in X according to two different mechanisms, MCAR or MNAR, and in two different proportions. Recall that the data are said (i) MCAR if the missingness does not depend on the data values and (ii) MNAR if the missingness depends on the data values. More precisely, any entry of X is missing according to the following logistic model

$$\mathbb{P}(X_{ij}^{\text{NA}} = \text{NA} \mid X_{ij} = x) = \frac{1}{1 + e^{-a|x|-b}},$$

where $a \geq 0$ and $b \in \mathbb{R}$. Choosing $a = 0$ yields MCAR data, whereas $a = 5$ leads to MNAR setting in which high absolute entries are more likely to be missing. For a fixed a , the value of b is chosen so that the overall average proportion of missing values is π , with $\pi = 5\%$ and $\pi = 20\%$.

Two sets of simulations are run. The first one is “ s -oracle”, meaning that the tuning parameters of the different methods are chosen so that the estimated support has correct support size s . In the second set, no knowledge of s, β^0 or σ is provided.

Performance evaluation. The performance of each coming estimator is assessed in terms of the following criteria, averaged over 100 replications:

- the Probability of Sign Recovery (PSR), $\text{PSR} = \mathbb{P}(\text{sign}(\hat{\beta}) = \text{sign}(\beta^0))$,
- the signed True Positive Rate (sTPR), where $\text{s-TPR} = \mathbb{E}(\text{s-TPP})$ with

$$\text{s-TPP} := \frac{|\{j \mid \beta_j^0 > 0, \hat{\beta}_j > 0\}| + |\{j \mid \beta_j^0 < 0, \hat{\beta}_j < 0\}|}{|S^0|}, \quad (15)$$

which is the proportion of nonzero coefficients whose sign is correctly identified;

- the signed False Discovery Rate (sFDR): $\text{s-FDR} = \mathbb{E}(\text{s-FDP})$ with

$$\text{s-FDP} := \frac{|\hat{S}| - |\{j \mid \beta_j^0 > 0, \hat{\beta}_j > 0\}| - |\{j \mid \beta_j^0 < 0, \hat{\beta}_j < 0\}|}{\max\{1, |\hat{S}|\}}, \quad (16)$$

which is the proportion of incorrect signs among all discoveries.

Estimators considered. We compare the following estimators:

- **Rlass0:** the Robust Lasso-Zero described in Algorithm 2 using the mean imputation and M equal to 30. The tuning parameters are obtained using arbitrarily $\lambda = 1$ and selecting τ by quantile universal threshold (QUT) at level $\alpha = 0.05$.
- **lass0:** the Lasso-Zero proposed in Descloux and Sardy [2021]. The automatic tuning is performed by QUT, at level $\alpha = 0.05$.
- **lasso:** the Lasso [Tibshirani, 1996] performed on the mean-imputed matrix where the regularization parameter is tuned by cross-validation.
- **NClasso:** the nonconvex ℓ_1 estimator of Loh and Wainwright [2012]. It is only included under the s -oracle setting, as selection of the tuning parameter in practice is not discussed in their work.
- **ABSLOPE:** Adaptive Bayesian SLOPE of Jiang et al. [2019].

To ease the readability of the numerical results, we also provide in Appendix C.2 additional comparisons to other estimators: the thresholded Robust Lasso proposed in Nguyen and Tran [2013a] and the thresholded lasso in Pokarowski et al. [2019]. Rlass0 still remains competitive in terms of sign recovery, in particular in difficult cases, i.e. when the percentage of missing values increases, when the missing data are informative and when the covariates are correlated. For the cases where Rlass0 has a lower

probability of sign recovery than the other methods, it still remains competitive in terms of s-FDR and outperforms the thresholded methods in terms of s-TPR.

In the following simulations, we impute each missing variable by its empirical mean computed with the observed individuals (as practitioners would tend to do). We refer to Appendix C.1 in which we provide extra numerical experiments when different strategies of imputation are considered, without changing the conclusions observed in this section.

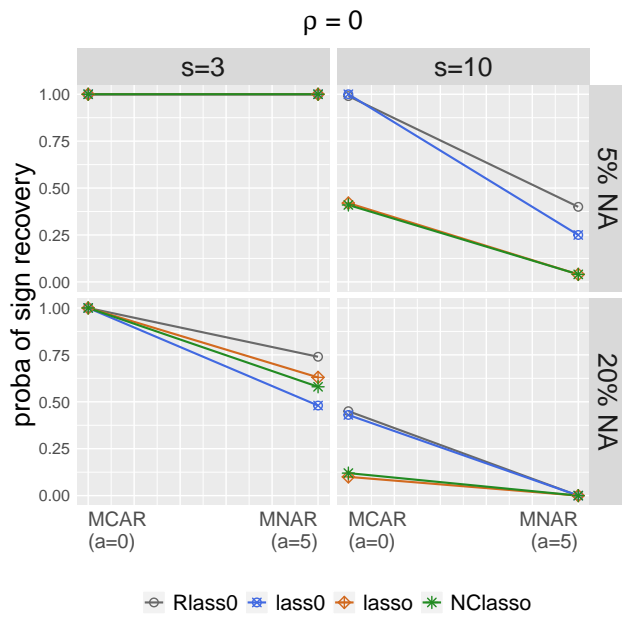
4.2 Results

4.2.1 With s -oracle hyperparameter tuning

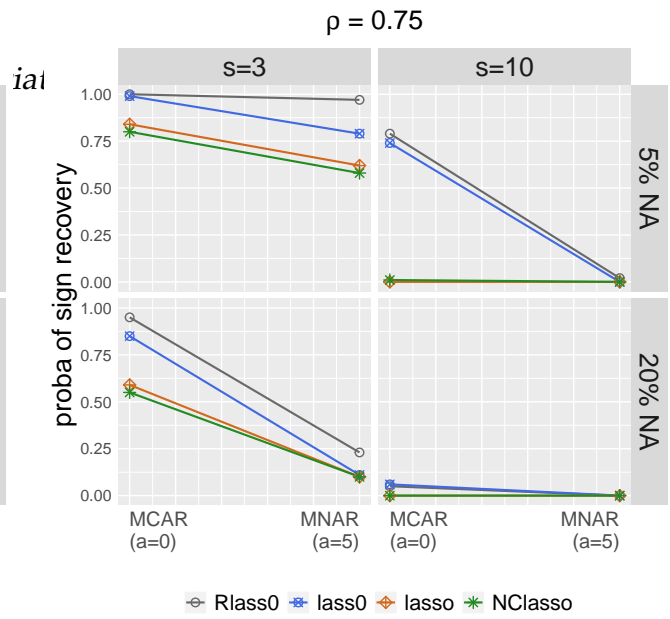
Under the s -oracle tuning, an s-TPP (15) of one means that the signs of β^0 are exactly recovered, and the s-TPP is related to the s-FDP (16) through $\text{s-FDP} = 1 - \text{s-TPP}$. That is why, in Figure 1, only the average s-TPP and the estimated probability of sign recovery are reported.

Small missingness – High sparsity (5% of NA and $s = 3$). In the non-correlated case, in Figure 1 (a) and (c), MCAR and MNAR results are similar across methods. With correlation, in Figure 1 (b) and (d), Rlass0 improves PSR and sTPR, specially with MNAR data.

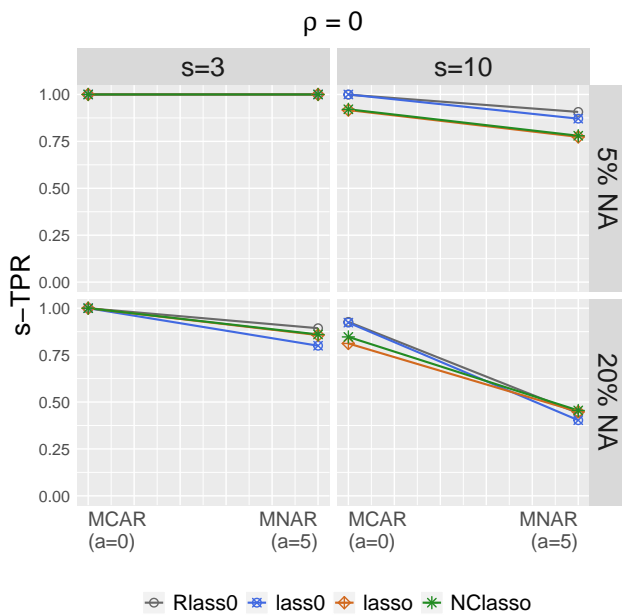
Increasing missingness – High sparsity (20% of NA and $s = 3$). The benefit of Rlass0 is noticeable when increasing the percentage of missing data to 20%, for both performance indicators. Indeed, with no correlation (Figure 1 (a)(c)(bottom left)), the improvement is clear when dealing with MNAR. With correlation (Figure 1 (b)(d)(bottom left)), Rlass0 outperforms the other methods: while the improvement can be marginal when



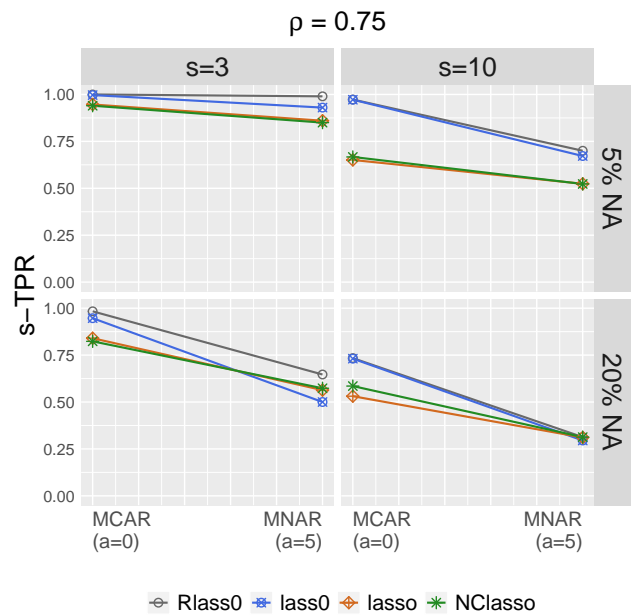
(a) PSR in the non-correlated case



(b) PSR in the correlated case



(c) s-TPR in the non-correlated case



(d) s-TPR in the correlated case

Figure 1: PSR and s-TPR with an s -oracle tuning, for sparsity levels, $s = 3$ and $s = 10$ (subplots columns), proportions of missing values 5% or 20% (subplots rows), and two missing data mechanisms (MCAR vs MNAR).

compared to `lass0` for MCAR, it becomes significant for MNAR.

Lower sparsity ($s = 10$). The performance of all estimators tends to deteriorate. One can identify two groups of estimators: `Rlass0` and `lass0` generally outperforms `lasso` and `NClasso`, except with a high proportion (20%) of MNAR missing data for which they all behave the same. While comparable when $s = 10$, `Rlass0` proves to be better than `lass0` in the case of a small proportion of MNAR missing data (5%).

4.2.2 With automatic hyperparameter tuning

Figures 2 and 3 point to the poor performance of `lasso` in terms of PSR for all experimental settings. The automatic tuning, being done by cross-validation, is known to lead to support overestimation. Indeed, its very good performance in `sTPR` is made at the cost of a very high `sFDR`.

Small missingness – High sparsity (5% of NA and $s = 3$). In Figures 2(a)(top left) and 3(a)(c)(top left), for the non-correlated case, `Rlass0`, `lass0` and `ABSlope` performs very well, providing a PSR and `s-TPR` of one, and a `s-FDR` of zero, either when dealing with MCAR or MNAR data (the `lasso` being already out of the game). In Figures 2(b)(top left) and 3(b)(d)(top left), adding correlation in the design matrix seems beneficial for `ABSlope`, at the price of high FDR, however.

Increasing missingness – High sparsity (20% of NA and $s = 3$). With no correlation, one sees in Figure 2(a)(bottom left) that `Rlass0` provides the best PSR, whatever the type of missing data is. One could also note that the performances in terms of PSR of either `lass0` or `ABSLOPE` are extremely

variable depending on the type of missing data (MCAR or MNAR) considered: the PSR of lass0 is comparable to the one of Rlass0 when facing MCAR data and is much lower than the one of Rlass0 when facing MNAR data; the converse is true for ABSLOPE.

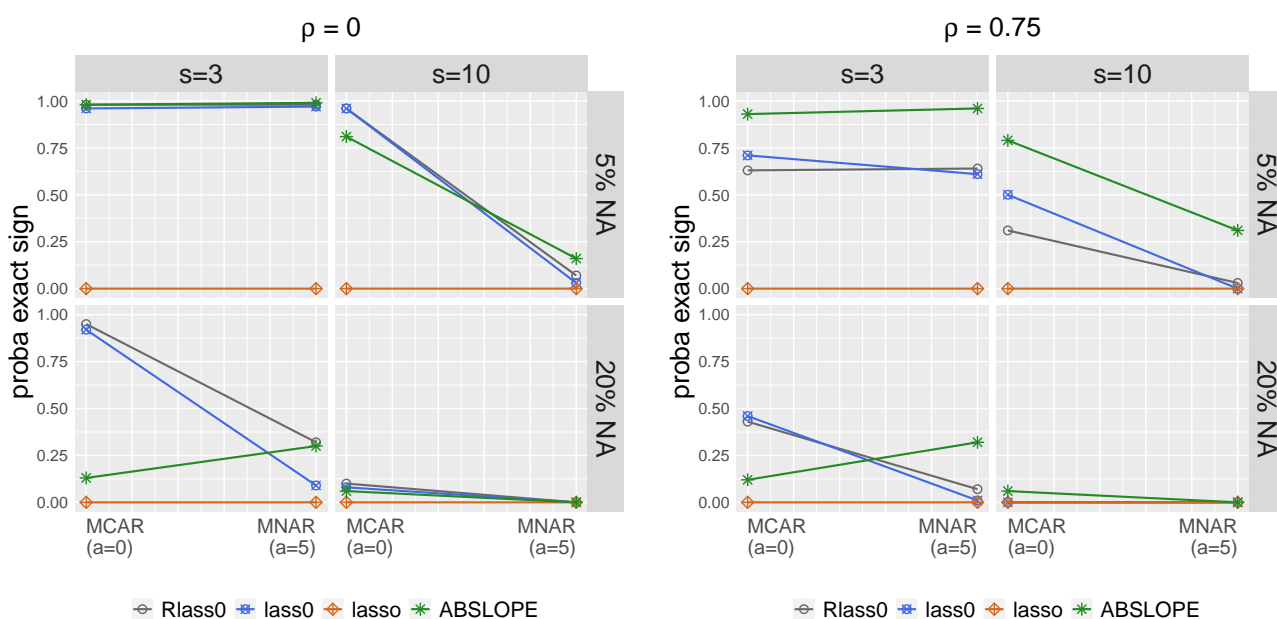
Regarding the s -TPR and s -FDR results in Figure 3 (a-d)(bottom left), the following observations hold in both correlated or non-correlated cases:

- (i) With MCAR data, all the methods behave similarly in terms of s -TPR, identifying correctly signs and coefficient locations in the support of β^0 , see Figure 3(a)(b)(bottom left);
- (ii) With MNAR data, lasso and ABSLOPE remain stable in terms of s -TPR, providing an s -TPR of one, whereas the s -TPR of Rlass0 deteriorates (to 0.6 and 0.5 respectively for the non-correlated and correlated cases), and even worse for lass0, see Figure 3(a)(b)(bottom left);
- (iii) Lasso and ABSLOPE lead to high s -FDR, while lass0 and Rlass0 always give the best s -FDR, see Figure 3(c)(d)(bottom left).

Lower sparsity ($s = 10$). For low missingness (5%), see Figure 2 (a)(b) (top right), ABSLOPE gives high PSR. In terms of s -TPR, lasso and ABSLOPE have high TPR. Moreover Rlass0 improves s -TPR compared to lass0 specially for a small proportion of MNAR missing data. In terms of s -FDR, lass0 and Rlass0 bring very low s -FDR, proving their FDR stability with respect to MCAR/MNAR data, and correlation.

4.2.3 Summary and discussion

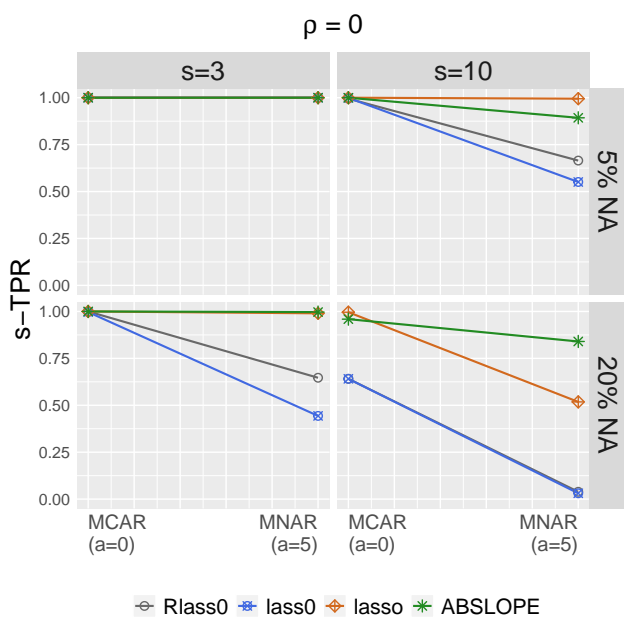
The results of experiments with s -oracle tuning (Section 4.2.1) show that the Robust Lasso-Zero performs better than competitors for sign re-



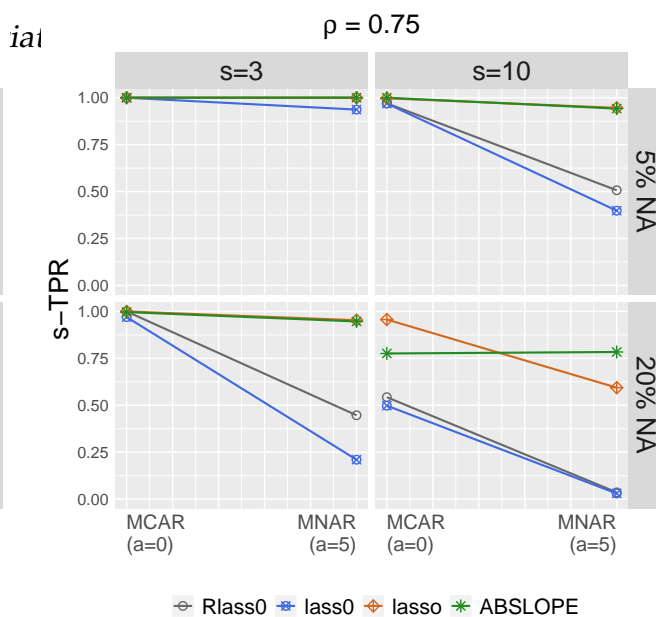
(a) PSR in the non-correlated case

(b) PSR in the correlated case

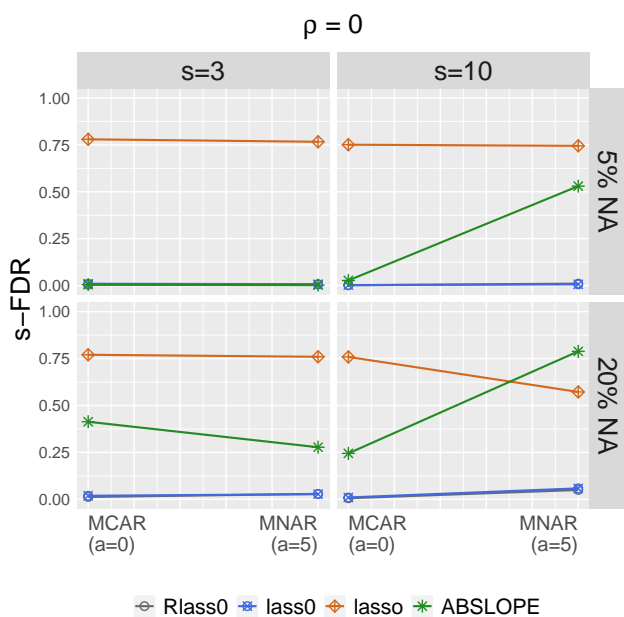
Figure 2: PSR with automatic tuning, for sparsity levels $s = 3$ and $s = 10$ (subplots columns), proportions of missing values 5% or 20% (subplots rows), and two missing data mechanisms (MCAR vs MNAR).



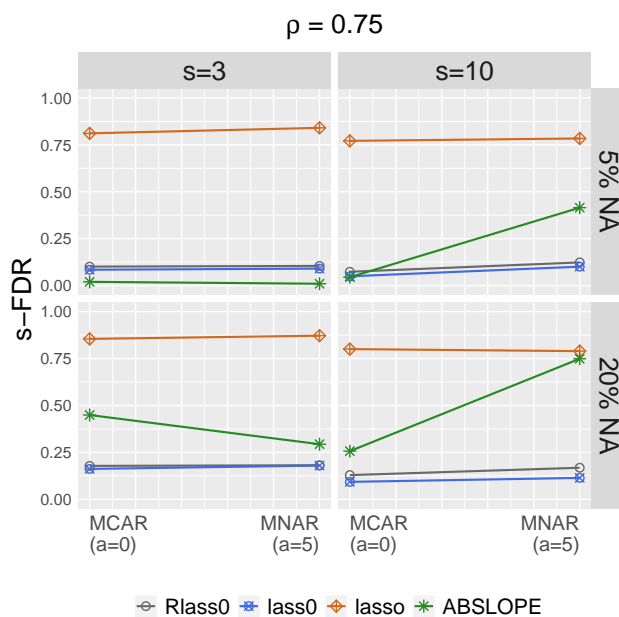
(a) s-TPR in the non-correlated case



(b) s-TPR in the correlated case



(c) s-FDR in the non-correlated case



(d) s-FDR in the correlated case

Figure 3: s-FDR and s-TPR with automatic tuning, for sparsity levels $s = 3$ and $s = 10$ (subplots columns), proportions of missing values 5% or 20% (subplots rows), and two missing data mechanisms (MCAR vs MNAR).

covery, and is more robust to MNAR data compared to its nonrobust counterpart when the sparsity index and/or proportion of missing entries is low. In particular, the Robust Lasso-Zero performs better than NClasso, one of the rare existing ℓ_1 -estimator designed to handle missing values.

While not designed to handle MNAR data, ABSLOPE appears to be a valid competitor in terms of s-TPR or PSR when the model complexity increases, and when dealing with MNAR data. Its poor performance in FDR in such settings reveals its tendency to overestimate the support of β^0 , under higher sparsity degrees, and with informative MNAR missing data.

With automatic tuning (Section 4.2.2), Robust Lasso-Zero is the best method overall. Moreover, our results show that the choice of Robust Lasso-Zero tuned by QUT, with its low s-FDR, is particularly appropriate in cases where one wants to maintain a low proportion of false discoveries.

5 Application to the Traumabase dataset

We illustrate our approach on the public health APHP (Assistance Publique Hopitaux de Paris) TraumaBase[®] Group for traumatized patients. Effective and timely management of trauma is crucial to improve outcomes, as delays or errors entail high risks for the patient.

In our analysis, we focus on one specific challenge: selecting a sparse model from data containing missing covariates in order to explain the level of platelet. This model can aid creating an innovative response to the public health challenge of major trauma. Explanatory variables for the level of platelet consist in fifteen quantitative variables containing missing values, which have been selected by doctors. They give clinical measurements on 490 patients. In Figure 4, one sees the percentage of missing

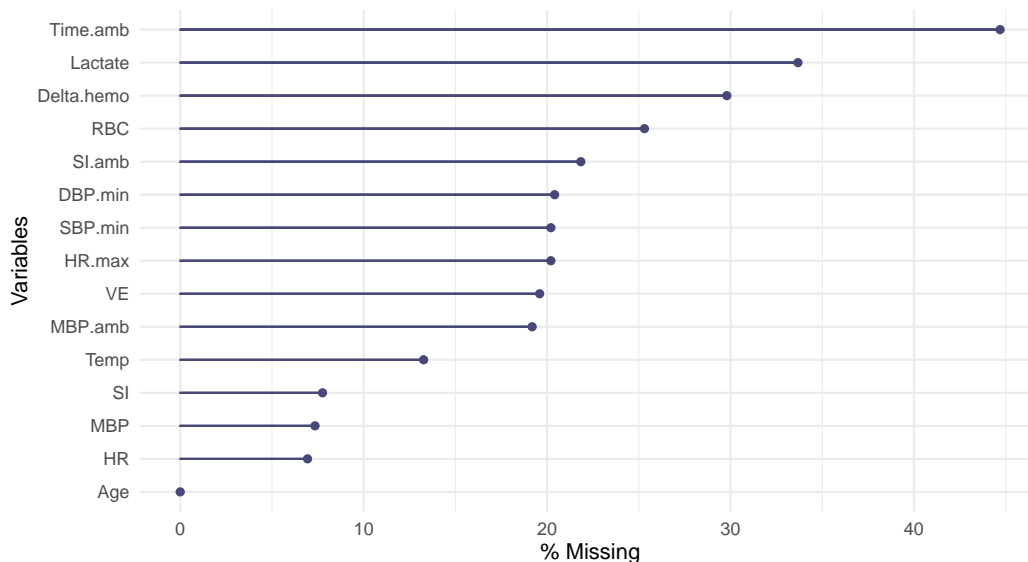


Figure 4: Percentage of missing values in the Traumbase dataset.

Variable	Rlass0	lass0	lasso	ABSLOPE
Age	-	0	-	-
SI	0	0	0	-
Delta.hemo	0	0	0	+
Lactates	0	0	0	+
Temperature	0	0	0	+
VE	-	0	-	0
RBC	-	0	0	-
DBP.min	0	0	-	+
HR.max	0	0	-	0
SI.amb	0	0	0	+

Table 1: Sign of estimated effects on the platelet for Rlass0, lass0, lasso or ABSLOPE. Variables not shown here are not selected by any method.

values in each variable, varying from 0 to 45% and leading to 20% is the whole dataset. Based on discussions with doctors, some variables may have informative missingness (M(N)AR variables). Both percentage and nature of missing data demonstrate the importance of taking appropriate account of missing data. More information can be found in Appendix D.

We compare the Robust Lasso-Zero to the Lasso-Zero, Lasso and ABSLOPE estimators. The signs of the coefficients are shown in Table 1. Lasso does not select any variable, whereas its robust counterpart selects three. According to doctors, the results given by the Robust Lasso-Zero methodology are the most coherent. Indeed, a negative effect of age (*Age*), vascular filling (*VE*) and blood transfusion (*RBC*) was expected, as they all result in low platelet levels and therefore a higher risk of severe bleeding. Lasso similarly selects *Age* and *VE*, but also minimum value of diastolic blood pressure *DBP.min* and the maximum heart rate *HR.max*. The effect of *DBP.min* is not what doctors expected. For ABSLOPE, the effects on platelets of delta Hemocue (*Delta.Hemocue*), the lactates (*Lactates*), the temperature (*Temperature*) and the shock index measured on ambulance (*SI.amb*), at odds with the effect of the shock index at hospital (*SI*), are not in agreement with the doctors opinion either.

6 Funding

Aude Sportisse was supported by the French government, through the 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002.

References

- Yudong Chen and Constantine Caramanis. Noisy and missing data regression: Distribution-oblivious support recovery. In International Conference on Machine Learning, pages 383–391, 2013.
- Yudong Chen, Constantine Caramanis, and Shie Mannor. Robust Sparse Regression under Adversarial Corruption. In International Conference on Machine Learning, pages 774–782, 2013a. URL <http://proceedings.mlr.press/v28/chen13h.html>.
- Yudong Chen, Constantine Caramanis, and Shie Mannor. Robust sparse regression under adversarial corruption. In International Conference on Machine Learning, pages 774–782, 2013b.
- Arnak S Dalalyan and Philip Thompson. Outlier-robust estimation of a sparse linear model using ℓ_1 -penalized huber’s m-estimator. arXiv preprint arXiv:1904.06288, 2019.
- Abhirup Datta and Hui Zou. CoCoLasso for high-dimensional error-in-variables regression. The Annals of Statistics, 45(6):2400–2426, 2017. ISSN 0090-5364, 2168-8966. doi: 10.1214/16-AOS1527. URL <https://projecteuclid.org/euclid.aos/1513328577>.
- Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C. Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. Communications on Pure and Applied Mathematics, 63(1): 1–38, 2010. ISSN 1097-0312. doi: 10.1002/cpa.20303. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.20303>.
- Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likeli-

- hood from incomplete data via the em algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1):1–22, 1977.
- Pascaline Descloux and Sylvain Sardy. Model selection with lasso-zero: adding straw to the haystack to better find needles. Journal of Computational and Graphical Statistics, pages 1–14, 2021.
- Simon Foucart and Holger Rauhut. A mathematical introduction to compressive sensing. Number 3. Birkhäuser Basel, 2013. URL <http://www.ams.org/bull/2017-54-01/S0273-0979-2016-01546-1/>.
- Ramon I Garcia, Joseph G Ibrahim, and Hongtu Zhu. Variable selection for regression models with missing data. Statistica Sinica, 20(1):149, 2010.
- Caroline Giacobino, Sylvain Sardy, Jairo Diaz-Rodriguez, and Nick Hengartner. Quantile universal threshold. Electronic Journal of Statistics, 11(2):4701–4722, 2017. ISSN 1935-7524. doi: 10.1214/17-EJS1366. URL <https://projecteuclid.org/euclid.ejs/1511492459>.
- Wei Jiang, Malgorzata Bogdan, Julie Josse, Blazej Miasojedow, Veronika Rockova, and TraumaBase Group. Adaptive Bayesian SLOPE – High-dimensional Model Selection with Missing Values. arXiv e-prints, art. arXiv:1909.06631, Sep 2019.
- J. N. Laska, M. A. Davenport, and R. G. Baraniuk. Exact signal recovery from sparsely corrupted measurements through the Pursuit of Justice. In 2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers, pages 1556–1560, 2009. doi: 10.1109/ACSSC.2009.5470141.
- B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. The Annals of Statistics, 28(5):1302–1338,

2000. ISSN 0090-5364, 2168-8966. doi: 10.1214/aos/1015957395. URL <https://projecteuclid.org/euclid.aos/1015957395>.
- Xiaodong Li. Compressed Sensing and Matrix Completion with Constant Proportion of Corruptions. Constructive Approximation, 37(1):73–99, 2013. ISSN 1432-0940. doi: 10.1007/s00365-012-9176-9. URL <https://doi.org/10.1007/s00365-012-9176-9>.
- Zhi Li, Feng Wu, and John Wright. On the systematic measurement matrix for compressed sensing in the presence of gross errors. In 2010 Data Compression Conference, pages 356–365. IEEE, 2010.
- Ying Liu, Yuanjia Wang, Yang Feng, and Melanie M Wall. Variable selection and prediction with incomplete high-dimensional data. The annals of applied statistics, 10(1):418, 2016.
- Po-Ling Loh and Martin J. Wainwright. High-Dimensional Regression with Noisy and Missing Data: Provable Guarantees with Nonconvexity. The Annals of Statistics, 40(3):1637–1664, 2012. ISSN 0090-5364. URL <http://www.jstor.org/stable/41713688>.
- N. H. Nguyen and T. D. Tran. Robust Lasso With Missing and Grossly Corrupted Observations. IEEE Transactions on Information Theory, 59(4):2036–2058, 2013a. ISSN 0018-9448. doi: 10.1109/TIT.2012.2232347.
- N. H. Nguyen and T. D. Tran. Exact Recoverability From Dense Corrupted Observations via ℓ_1 -Minimization. IEEE Transactions on Information Theory, 59(4):2017–2035, 2013b. ISSN 0018-9448. doi: 10.1109/TIT.2013.2240435.
- Piotr Pokarowski, Wojciech Rejchel, Agnieszka Soltys, Michal Frej, and Jan

- Mielniczuk. Improving lasso for model selection and prediction. arXiv preprint arXiv:1907.03025, 2019.
- Mathieu Rosenbaum, Alexandre B Tsybakov, et al. Improved matrix uncertainty selector. In From Probability to Statistics and Back: High-Dimensional Models and Processes—A Festschrift in Honor of Jon A. Wellner, pages 276–290. Institute of Mathematical Statistics, 2013.
- Donald B Rubin. Inference and missing data. Biometrika, 63(3):581–592, 1976.
- Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. arXiv:1003.2990 [math], 2010. URL <http://arxiv.org/abs/1003.2990>. arXiv: 1003.2990.
- Sylvain Sardy, Paul Tseng, and A. G. Bruce. Robust wavelet denoising. IEEE Transactions on Signal Processing, 49:1146–1152, 2001.
- Ulrike Schneider and Patrick Tardivel. The geometry of uniqueness and model selection of penalized estimators including slope, lasso, and basis pursuit. arXiv preprint arXiv:2004.09106, 2020.
- Patrick Tardivel and Malgorzata Bogdan. On the sign recovery by lasso, thresholded lasso and thresholded basis pursuit denoising. 2019.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1):267–288, 1996.
- Stef Van Buuren. Flexible imputation of missing data. Chapman and Hall/CRC, 2018.

J. Wright and Y. Ma. Dense Error Correction Via ℓ_1 -Minimization. IEEE Transactions on Information Theory, 56(7):3540–3560, 2010. ISSN 0018-9448. doi: 10.1109/TIT.2010.2048473.

J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust Face Recognition via Sparse Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31(2):210–227, 2009. ISSN 0162-8828. doi: 10.1109/TPAMI.2008.79.

Dao Li Zhu and Patrice Marcotte. Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. SIAM Journal on Optimization, 6(3):714–726, 1996.

Corresponding author The corresponding author is Aude Sportisse (email: aude.sportisse@inria.fr).

A Proof of Theorem 1

Lemma 1 implies that under the sign invariance assumption (i), identifiability of $(\beta^{(r)}, \omega^{(r)})$ is equivalent to identifiability of $(\theta, \tilde{\theta})$.

Proof of Lemma 1. Note that $(\hat{\beta}_\lambda^{\text{JP}}, \hat{\omega}_\lambda^{\text{JP}})$ is a solution to JP (6) if and only if $(\hat{\beta}_\lambda^{\text{JP}}, \hat{\omega}_\lambda^{\text{JP}}) = (\tilde{\beta}, \lambda^{-1}\tilde{\omega})$, where $(\tilde{\beta}, \tilde{\omega})$ is a solution to

$$\min_{(\beta, \omega) \in \mathbb{R}^p \times \mathbb{R}^n} \|\beta\|_1 + \|\omega\|_1 \quad \text{s.t.} \quad y = X\beta + \sqrt{n}\lambda^{-1}\omega. \quad (17)$$

So (β^0, ω^0) is identifiable with respect to X and $\lambda > 0$ if and only if the pair $(\beta^0, \lambda\omega^0)$ is the unique solution of (17) when $y = X\beta^0 + \sqrt{n}\omega^0$. But (17) is just Basis Pursuit with response vector $y \in \mathbb{R}^n$ and augmented matrix $\begin{bmatrix} X & \sqrt{n}\lambda^{-1}I_n \end{bmatrix}$, so by a result of Daubechies et al. [2010] this is the case if

and only if for every $(\beta, \omega) \neq (0, 0)$ such that $X\beta + \sqrt{n}\lambda^{-1}\omega = 0$, we have $|\text{sign}(\beta^0)^T\beta + \text{sign}(\omega^0)^T\omega| < \|\beta_{\overline{S^0}}\|_1 + \|\omega_{\overline{T^0}}\|_1$, which proves our statement. \square

We will need the following auxiliary lemma.

Lemma 2. *Under assumptions (i) and (ii), if the pair $(\theta, \tilde{\theta})$ is identifiable with respect to X and λ , then for any $\epsilon \in \mathbb{R}^n$,*

$$\lim_{r \rightarrow +\infty} \frac{1}{u_r} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)} \\ \hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

where $u_r := \|\beta^{(r)}\|_1 + \lambda\|\omega^{(r)}\|_1$.

Proof. First note that by assumption (ii), $\lim_{r \rightarrow +\infty} u_r = +\infty$. Now let $\epsilon \in \mathbb{R}^n$ and denote by $(\hat{\beta}_\lambda^{\text{JP}}(\epsilon), \hat{\omega}_\lambda^{\text{JP}}(\epsilon))$ the JP solution when $y = \epsilon$. In particular, one has $\epsilon = X\hat{\beta}_\lambda^{\text{JP}}(\epsilon) + \sqrt{n}\hat{\omega}_\lambda^{\text{JP}}(\epsilon)$, so for every $r \in \mathbb{N}^*$, let us consider $y^{(r)}$ as follows

$$y^{(r)} = X(\beta^{(r)} + \hat{\beta}_\lambda^{\text{JP}}(\epsilon)) + \sqrt{n}(\omega^{(r)} + \hat{\omega}_\lambda^{\text{JP}}(\epsilon)).$$

Hence $(\beta^{(r)} + \hat{\beta}_\lambda^{\text{JP}}(\epsilon), \omega^{(r)} + \hat{\omega}_\lambda^{\text{JP}}(\epsilon))$ is feasible for JP when $y = y^{(r)}$, so

$$\begin{aligned} & \frac{\|\hat{\beta}_\lambda^{\text{JP}(r)}\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}(r)}\|_1}{u_r} \\ & \leq \frac{\|\beta^{(r)} + \hat{\beta}_\lambda^{\text{JP}}(\epsilon)\|_1 + \lambda\|\omega^{(r)} + \hat{\omega}_\lambda^{\text{JP}}(\epsilon)\|_1}{u_r} \\ & \leq \frac{(\|\beta^{(r)}\|_1 + \lambda\|\omega^{(r)}\|_1) + (\|\hat{\beta}_\lambda^{\text{JP}}(\epsilon)\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}}(\epsilon)\|_1)}{u_r} \\ & = 1 + \frac{\|\hat{\beta}_\lambda^{\text{JP}}(\epsilon)\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}}(\epsilon)\|_1}{u_r}. \end{aligned} \tag{18}$$

Therefore

$$\begin{aligned}
 & \frac{1}{u_r} (\|\hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)}\|_1 + \lambda \|\hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)}\|_1) \\
 & \leq \frac{1}{u_r} ((\|\beta^{(r)}\|_1 + \lambda \|\omega^{(r)}\|_1) + (\|\hat{\beta}_\lambda^{\text{JP}(r)}\|_1 + \lambda \|\hat{\omega}_\lambda^{\text{JP}(r)}\|_1)) \\
 & = 1 + \frac{\|\hat{\beta}_\lambda^{\text{JP}(r)}\|_1 + \lambda \|\hat{\omega}_\lambda^{\text{JP}(r)}\|_1}{u_r} \\
 & \leq 2 + \frac{\|\hat{\beta}_\lambda^{\text{JP}}(\epsilon)\|_1 + \lambda \|\hat{\omega}_\lambda^{\text{JP}}(\epsilon)\|_1}{u_r},
 \end{aligned} \tag{19}$$

using (18) for last inequality. Since $\lim_{r \rightarrow +\infty} \frac{\|\hat{\beta}_\lambda^{\text{JP}}(\epsilon)\|_1 + \lambda \|\hat{\omega}_\lambda^{\text{JP}}(\epsilon)\|_1}{u_r} = 0$, and

since $\begin{bmatrix} \beta \\ \omega \end{bmatrix} \mapsto \|\beta\|_1 + \lambda \|\omega\|_1$ defines a norm on \mathbb{R}^{p+n} , one deduces that the

sequence $\frac{1}{u_r} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)} \\ \hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)} \end{bmatrix}$ is bounded. Therefore we need to check that every convergent subsequence converges to zero. Let

$$\frac{1}{u_{\phi(r)}} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(\phi(r))} - \beta^{(\phi(r))} \\ \hat{\omega}_\lambda^{\text{JP}(\phi(r))} - \omega^{(\phi(r))} \end{bmatrix}$$

(with $\phi : \mathbb{N}^* \rightarrow \mathbb{N}^*$ strictly increasing) be an arbitrary convergent subsequence. Since

$$\frac{\|\beta^{(r)}\|_1 + \lambda \|\omega^{(r)}\|_1}{u_r} = 1 \tag{20}$$

for every r , and by (18), the sequences $\frac{1}{u_r} \begin{bmatrix} \beta^{(r)} \\ \omega^{(r)} \end{bmatrix}$ and $\frac{1}{u_r} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(r)} \\ \hat{\omega}_\lambda^{\text{JP}(r)} \end{bmatrix}$ are bounded as well. Hence without loss of generality (otherwise, reduce the subsequence),

$$\lim_{r \rightarrow +\infty} \frac{1}{u_{\phi(r)}} \begin{bmatrix} \beta^{(\phi(r))} \\ \omega^{(\phi(r))} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \tag{21}$$

and

$$\lim_{r \rightarrow +\infty} \frac{1}{u_{\phi(r)}} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(\phi(r))} \\ \hat{\omega}_\lambda^{\text{JP}(\phi(r))} \end{bmatrix} = \begin{bmatrix} \nu'_1 \\ \nu'_2 \end{bmatrix} \tag{22}$$

for some $\begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}, \begin{bmatrix} \nu'_1 \\ \nu'_2 \end{bmatrix} \in \mathbb{R}^{p+n}$. By (20), one necessarily has

$$\|\nu_1\|_1 + \lambda\|\nu_2\|_1 = 1, \quad (23)$$

and (18) implies that

$$\|\nu'_1\|_1 + \lambda\|\nu'_2\|_1 \leq 1. \quad (24)$$

Now

$$\begin{aligned} \lim_{r \rightarrow +\infty} \frac{X(\hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)}) + \sqrt{n}(\hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)})}{u_r} &= \lim_{r \rightarrow +\infty} \frac{y^{(r)} - (X\beta^{(r)} + \sqrt{n}\omega^{(r)})}{u_r} \\ &= \lim_{r \rightarrow +\infty} \frac{\epsilon}{u_r} = 0, \end{aligned}$$

so one deduces that

$$\lim_{r \rightarrow +\infty} \begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(\phi(r))}/u_{\phi(r)} \\ \hat{\omega}_\lambda^{\text{JP}(\phi(r))}/u_{\phi(r)} \end{bmatrix} = \lim_{r \rightarrow +\infty} \begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix} \begin{bmatrix} \beta^{(\phi(r))}/u_{\phi(r)} \\ \omega^{(\phi(r))}/u_{\phi(r)} \end{bmatrix},$$

so by (21) and (22),

$$\begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix} \begin{bmatrix} \nu'_1 \\ \nu'_2 \end{bmatrix} = \begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}. \quad (25)$$

Assuming for now that (ν_1, ν_2) is identifiable with respect to X and λ , equality (25) together with (23) and (24) imply that $\begin{bmatrix} \nu'_1 \\ \nu'_2 \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix}$, hence

$$\lim_{r \rightarrow +\infty} \frac{1}{u_{\phi(r)}} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(\phi(r))} - \beta^{(\phi(r))} \\ \hat{\omega}_\lambda^{\text{JP}(\phi(r))} - \omega^{(\phi(r))} \end{bmatrix} = \begin{bmatrix} \nu'_1 \\ \nu'_2 \end{bmatrix} - \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

It remains to check that (ν_1, ν_2) is identifiable with respect to X and λ , which we will do using Lemma 1. Note that (21) and assumption (i) imply

$$\text{sign}(\nu_1) = \theta - \theta', \quad (26)$$

$$\text{sign}(\nu_2) = \tilde{\theta} - \tilde{\theta}', \quad (27)$$

where $\theta'_j := \theta_j \mathbf{1}_{\{\nu_{1,j}=0, \theta_j \neq 0\}}$, and $\tilde{\theta}'_j = \tilde{\theta}_j \mathbf{1}_{\{\nu_{2,j}=0, \tilde{\theta}_j \neq 0\}}$, and hence

$$\overline{\text{supp}(\nu_1)} = \overline{\text{supp}(\theta)} \sqcup \text{supp}(\theta') = \overline{S^0} \sqcup \text{supp}(\theta'), \quad (28)$$

$$\overline{\text{supp}(\nu_2)} = \overline{\text{supp}(\tilde{\theta})} \sqcup \text{supp}(\tilde{\theta}') = \overline{T^0} \sqcup \text{supp}(\tilde{\theta}'). \quad (29)$$

Consider a pair $(\beta, \omega) \neq (0, 0)$ such that $X\beta + \sqrt{n}\lambda^{-1}\omega = 0$. By (26) and (27),

$$\begin{aligned} |\text{sign}(\nu_1)^T \beta + \text{sign}(\nu_2)^T \omega| &= |(\theta - \theta')^T \beta + (\tilde{\theta} - \tilde{\theta}')^T \omega| \\ &\leq |\theta^T \beta + \tilde{\theta}^T \omega| + |(\theta')^T \beta| + |(\tilde{\theta}')^T \omega|. \end{aligned} \quad (30)$$

But since $(\theta, \tilde{\theta})$ is identifiable with respect to X and λ , Lemma 1 implies $|\theta^T \beta + \tilde{\theta}^T \omega| < \|\beta_{\overline{S^0}}\|_1 + \|\omega_{\overline{T^0}}\|_1$. Plugging this into (30) gives

$$\begin{aligned} |\text{sign}(\nu_1)^T \beta + \text{sign}(\nu_2)^T \omega| &< \|\beta_{\overline{S^0}}\|_1 + \|\omega_{\overline{T^0}}\|_1 + |(\theta')^T \beta| + |(\tilde{\theta}')^T \omega| \\ &\leq \|\beta_{\overline{S^0}}\|_1 + \|\beta_{\text{supp}(\theta')}\|_1 + \|\omega_{\overline{T^0}}\|_1 + \|\omega_{\text{supp}(\tilde{\theta}')}\|_1 \\ &= \|\beta_{\overline{\text{supp}(\nu_1)}}\|_1 + \|\omega_{\overline{\text{supp}(\nu_2)}}\|_1, \end{aligned}$$

where the equality comes from (28) and (29). By Lemma 1, one concludes that (ν_1, ν_2) is identifiable with respect to X and λ . \square

Proof of Theorem 1. Let us assume that $(\theta, \tilde{\theta})$ is identifiable with respect to X and λ , and let $\epsilon \in \mathbb{R}^n$. By Lemma 2,

$$\lim_{r \rightarrow +\infty} \frac{1}{u_r} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)} \\ \hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (31)$$

Since

$$\min\{1, \lambda\} \max\{\|\beta^{(r)}\|_\infty, \|\omega^{(r)}\|_\infty\} \leq u_r \leq (|S^0| + \lambda|T^0|) \max\{\|\beta^{(r)}\|_\infty, \|\omega^{(r)}\|_\infty\},$$

(31) is equivalent to $\lim_{r \rightarrow +\infty} \frac{1}{\max\{\|\beta^{(r)}\|_\infty, \|\omega^{(r)}\|_\infty\}} \begin{bmatrix} \hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)} \\ \hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Therefore there exists $R > 0$ such that for every $r \geq R$,

$$\|\hat{\beta}_\lambda^{\text{JP}(r)} - \beta^{(r)}\|_\infty < \frac{q}{2} \max\{\|\beta^{(r)}\|_\infty, \|\omega^{(r)}\|_\infty\} \quad (32)$$

and

$$\|\hat{\omega}_\lambda^{\text{JP}(r)} - \omega^{(r)}\|_\infty < \frac{q}{2} \max\{\|\beta^{(r)}\|_\infty, \|\omega^{(r)}\|_\infty\}. \quad (33)$$

Setting $\tau := \frac{q}{2} \max\{\|\beta^{(r)}\|_\infty, \|\omega^{(r)}\|_\infty\}$, (32) implies that $|\hat{\beta}_{\lambda,j}^{\text{JP}(r)}| < \tau$ for every $j \notin S^0$, hence $\hat{\beta}_{(\lambda,\tau),j}^{\text{TJP}(r)} = 0$. If $j \in S^0$, assumption (iii) implies

$$|\beta_j^{(r)}| \geq \beta_{\min}^{(r)} \geq 2\tau, \quad (34)$$

and by (32), we have

$$|\hat{\beta}_{\lambda,j}^{\text{JP}(r)} - \beta_j^{(r)}| < \tau, \quad (35)$$

so (34) and (35) together imply $|\hat{\beta}_{\lambda,j}^{\text{JP}(r)}| > \tau$ and $\text{sign}(\hat{\beta}_{\lambda,j}^{\text{JP}(r)}) = \text{sign}(\beta_j^{(r)})$. So we conclude that $\text{sign}(\hat{\beta}_{(\lambda,\tau)}^{\text{TJP}(r)}) = \text{sign}(\beta^{(r)})$. Analogously, (33) implies $\text{sign}(\hat{\omega}_{(\lambda,\tau)}^{\text{TJP}(r)}) = \text{sign}(\omega^{(r)})$.

Conversely, let us assume that for some $\epsilon \in \mathbb{R}^n$, $r \in \mathbb{N}^*$ and $\tau > 0$,

$$\text{sign}(\hat{\beta}_{(\lambda,\tau)}^{\text{TJP}(r)}) = \theta, \quad \text{sign}(\hat{\omega}_{(\lambda,\tau)}^{\text{TJP}(r)}) = \tilde{\theta}. \quad (36)$$

Note that the JP solution $(\hat{\beta}_\lambda^{\text{JP}(r)}, \hat{\omega}_\lambda^{\text{JP}(r)})$ is unique by assumption, hence $(\hat{\beta}_\lambda^{\text{JP}(r)}, \hat{\omega}_\lambda^{\text{JP}(r)})$ is identifiable with respect to X and λ . Now by (36), all nonzero components of θ and $\tilde{\theta}$ must have the same sign as the corresponding entries of $\hat{\beta}_\lambda^{\text{JP}(r)}$ and $\hat{\omega}_\lambda^{\text{JP}(r)}$ respectively. Hence

$$\begin{aligned} \theta &= \text{sign}(\theta) = \text{sign}(\hat{\beta}_\lambda^{\text{JP}(r)}) - \delta, \\ \tilde{\theta} &= \text{sign}(\tilde{\theta}) = \text{sign}(\hat{\omega}_\lambda^{\text{JP}(r)}) - \tilde{\delta}, \end{aligned} \quad (37)$$

where $\delta_j = \text{sign}(\hat{\beta}_{\lambda,j}^{\text{JP}(r)}) \mathbf{1}_{\{\hat{\beta}_{\lambda,j}^{\text{JP}(r)} \neq 0, \theta_j = 0\}}$ and $\tilde{\delta}_i = \text{sign}(\hat{\omega}_{\lambda,i}^{\text{JP}(r)}) \mathbf{1}_{\{\hat{\omega}_{\lambda,i}^{\text{JP}(r)} \neq 0, \tilde{\theta}_i = 0\}}$, and

$$\begin{aligned} \overline{S^0} &= \overline{\text{supp}(\theta)} = \overline{\text{supp}(\hat{\beta}_\lambda^{\text{JP}(r)})} \sqcup \text{supp}(\delta) \\ \overline{T^0} &= \overline{\text{supp}(\tilde{\theta})} = \overline{\text{supp}(\hat{\omega}_\lambda^{\text{JP}(r)})} \sqcup \text{supp}(\tilde{\delta}). \end{aligned} \quad (38)$$

In order to apply Lemma 1, let us consider a pair $(\beta, \omega) \neq (0, 0)$ such that $X\beta + \sqrt{n}\lambda^{-1}\omega = 0$. By (37), one has

$$\begin{aligned} |\theta^T \beta + \tilde{\theta}^T \omega| &= |\text{sign}(\hat{\beta}_\lambda^{\text{JP}(r)})^T \beta - \delta^T \beta + \text{sign}(\hat{\omega}_\lambda^{\text{JP}(r)})^T \omega - \tilde{\delta}^T \omega| \\ &\leq |\text{sign}(\hat{\beta}_\lambda^{\text{JP}(r)})^T \beta + \text{sign}(\hat{\omega}_\lambda^{\text{JP}(r)})^T \omega| + |\delta^T \beta| + |\tilde{\delta}^T \omega| \\ &\leq \|\beta_{\text{supp}(\hat{\beta}_\lambda^{\text{JP}(r)})}\|_1 + \|\omega_{\text{supp}(\hat{\omega}_\lambda^{\text{JP}(r)})}\|_1 + \|\beta_{\text{supp}(\delta)}\|_1 + \|\omega_{\text{supp}(\tilde{\delta})}\|_1 \\ &= \|\beta_{S^0}\|_1 + \|\omega_{T^0}\|_1, \end{aligned}$$

where we have used Lemma 1 and the fact that $(\hat{\beta}_\lambda^{\text{JP}(r)}, \hat{\omega}_\lambda^{\text{JP}(r)})$ is identifiable with respect to X and λ in the last inequality, and (38) for the last equality. Lemma 1 concludes our proof. \square

B Proof of Proposition 2

Proof of Proposition 2. We define $\tilde{X} := \begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix}$, and $\tilde{\nu} = \begin{bmatrix} \tilde{\beta} \\ \tilde{\omega} \end{bmatrix} := \tilde{X}^T(\tilde{X}\tilde{X}^T)^{-1}\epsilon$.

We will assume for now that the following properties hold.

a) Every pair (β, ω) such that $X\beta + \sqrt{n}\omega = 0$ satisfies

$$\|\beta_{S^0}\|_1 + \lambda\|\omega_{T^0}\|_1 \leq \frac{1}{3}(\|\beta_{S^0}\|_1 + \lambda\|\omega_{T^0}\|_1),$$

b) $\|\tilde{\nu}\|_2 \leq \frac{\sqrt{2}\sigma}{\left(\frac{\lambda_{\min}(\Sigma)}{4}(\sqrt{p/n-1})^2+1\right)^{1/2}}$.

Since $\tilde{X}\tilde{\nu} = X\tilde{\beta} + \sqrt{n}\tilde{\omega} = \epsilon$, one can rewrite model (2) as

$$y = X(\beta^0 + \tilde{\beta}) + \sqrt{n}(\omega^0 + \tilde{\omega}).$$

By property a) and Lemma 3 below, one has

$$\|\hat{\beta}_\lambda^{\text{JP}} - (\beta^0 + \tilde{\beta})\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}} - (\omega^0 + \tilde{\omega})\|_1 \leq 4(\|\tilde{\beta}_{S^0}\|_1 + \lambda\|\tilde{\omega}_{T^0}\|_1), \quad (39)$$

and therefore $\|\hat{\beta}_\lambda^{\text{JP}} - (\beta^0 + \tilde{\beta})\|_1 \leq 4(\|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1)$. Consequently, for any $j \in [p]$ one has

$$\begin{aligned} |\hat{\beta}_{\lambda,j}^{\text{JP}} - \beta_j^0| &\leq |\hat{\beta}_{\lambda,j}^{\text{JP}} - (\beta_j^0 + \tilde{\beta}_j)| + |\tilde{\beta}_j| \leq \|\hat{\beta}_\lambda^{\text{JP}} - (\beta^0 + \tilde{\beta})\|_1 + \|\tilde{\beta}\|_1 \\ &\leq 4(\|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1) + \|\tilde{\beta}\|_1 \leq 5(\|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1) \\ &\leq 5 \max\{1, \lambda\}(\|\tilde{\beta}\|_1 + \|\tilde{\omega}\|_1) = 5 \max\{1, \lambda\}\|\tilde{\nu}\|_1 \\ &\leq 5 \max\{1, \lambda\}\sqrt{p+n}\|\tilde{\nu}\|_2 \leq \frac{5\sqrt{2} \max\{1, \lambda\}\sigma\sqrt{p+n}}{\left(\frac{\lambda_{\min}(\Sigma)}{4}\right)(\sqrt{p/n} - 1)^2 + 1}^{1/2} \end{aligned}$$

where we have used property **b)** in the last inequality. Now setting

$$\tau := \frac{5\sqrt{2} \max\{1, \lambda\}\sigma\sqrt{p+n}}{\left(\frac{\lambda_{\min}(\Sigma)}{4}\right)(\sqrt{p/n} - 1)^2 + 1}^{1/2},$$

one gets

$$|\hat{\beta}_{\lambda,j}^{\text{JP}} - \beta_j^0| \leq \tau \quad (40)$$

for every $j \in [p]$. If $j \in \overline{S^0}$, we have $|\hat{\beta}_{\lambda,j}^{\text{JP}}| \leq \tau$, hence $\hat{\beta}_{(\lambda,\tau),j}^{\text{TJP}} = 0$. If $j \in S^0$, assumption (12) implies $|\beta_j^0| > 2\tau$, which together with (40) gives $\text{sign}(\hat{\beta}_{(\lambda,\tau),j}^{\text{TJP}}) = \text{sign}(\beta_j^0)$.

It remains to prove that properties **a)** and **b)** hold with high probability. First, Lemma 1 in [Nguyen and Tran \[2013a\]](#), implies that with probability greater than $1 - ce^{-c'n}$ the matrix X satisfies the extended restricted eigenvalue property

$$\begin{aligned} \|\beta_{\overline{S^0}}\|_1 + \lambda\|\omega_{\overline{T^0}}\|_1 &\leq 3(\|\beta_{S^0}\|_1 + \lambda\|\omega_{T^0}\|_1) \\ &\downarrow \\ \frac{1}{n}\|X\beta + \sqrt{n}\omega\|_2^2 &\geq \gamma^2(\|\beta\|_2^2 + \|\omega\|_2^2), \end{aligned} \quad (41)$$

with $\gamma^2 = \frac{\min\{\lambda_{\min}(\Sigma), 1\}}{16^2}$. Property (41) clearly implies **a)**. Finally, Lemma 4 below proves that **b)** holds with probability at least $1 - 1.14^{-n} - 2e^{-\frac{1}{8}(\sqrt{p}-\sqrt{n})^2}$, which concludes our proof. \square

Lemma 3. *Assume that for some sets $S^0 \subset [p]$ and $T^0 \subset [n]$, and some constant $\rho \in (0, 1)$, the matrix $X \in \mathbb{R}^{n \times p}$ satisfies*

$$\|\beta_{S^0}\|_1 + \lambda\|\omega_{T^0}\|_1 \leq \rho(\|\beta_{\overline{S^0}}\|_1 + \lambda\|\omega_{\overline{T^0}}\|_1), \quad (42)$$

for every pair $(\beta, \omega) \in \mathbb{R}^p \times \mathbb{R}^n$ such that $X\beta + \sqrt{n}\omega = 0$. Then for every pair $(\tilde{\beta}, \tilde{\omega}) \in \mathbb{R}^p \times \mathbb{R}^n$, the solution $(\hat{\beta}_\lambda^{\text{JP}}, \hat{\omega}_\lambda^{\text{JP}})$ to JP (6) with $y = X\tilde{\beta} + \sqrt{n}\tilde{\omega}$ satisfies

$$\|\hat{\beta}_\lambda^{\text{JP}} - \tilde{\beta}\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}} - \tilde{\omega}\|_1 \leq \frac{2(1+\rho)}{1-\rho}(\|\tilde{\beta}_{\overline{S^0}}\|_1 + \lambda\|\tilde{\omega}_{\overline{T^0}}\|_1).$$

Proof. This proof is a simple extension of the one of Theorem 4.14 in Foucart and Rauhut [2013]. Let us consider $y = X\tilde{\beta} + \sqrt{n}\tilde{\omega}$ for an arbitrary pair $(\tilde{\beta}, \tilde{\omega})$, and define $\beta' := \hat{\beta}_\lambda^{\text{JP}} - \tilde{\beta}$ and $\omega' := \hat{\omega}_\lambda^{\text{JP}} - \tilde{\omega}$. Clearly $X\beta' + \sqrt{n}\omega' = 0$, so by (42),

$$\|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1 \leq \rho(\|\beta'_{\overline{S^0}}\|_1 + \lambda\|\omega'_{\overline{T^0}}\|_1). \quad (43)$$

We also have

$$\begin{aligned} \|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1 &= \|\tilde{\beta}_{S^0}\|_1 + \|\tilde{\beta}_{\overline{S^0}}\|_1 + \lambda(\|\tilde{\omega}_{T^0}\|_1 + \|\tilde{\omega}_{\overline{T^0}}\|_1) \\ &= \|\hat{\beta}_{\lambda, S^0}^{\text{JP}} - \beta'_{S^0}\|_1 + \|\tilde{\beta}_{\overline{S^0}}\|_1 + \lambda(\|\hat{\omega}_{\lambda, T^0}^{\text{JP}} - \omega'_{T^0}\|_1 + \|\tilde{\omega}_{\overline{T^0}}\|_1) \\ &\leq \|\hat{\beta}_{\lambda, S^0}^{\text{JP}}\|_1 + \|\beta'_{S^0}\|_1 + \|\tilde{\beta}_{\overline{S^0}}\|_1 + \lambda(\|\hat{\omega}_{\lambda, T^0}^{\text{JP}}\|_1 + \|\omega'_{T^0}\|_1 + \|\tilde{\omega}_{\overline{T^0}}\|_1), \end{aligned}$$

and

$$\|\beta'_{\overline{S^0}}\|_1 + \lambda\|\omega'_{\overline{T^0}}\|_1 \leq (\|\hat{\beta}_{\lambda, S^0}^{\text{JP}}\|_1 + \|\tilde{\beta}_{\overline{S^0}}\|_1) + \lambda(\|\hat{\omega}_{\lambda, T^0}^{\text{JP}}\|_1 + \|\tilde{\omega}_{\overline{T^0}}\|_1).$$

Adding the last two inequalities yields

$$\begin{aligned} \|\beta'_{\overline{S^0}}\|_1 + \lambda\|\omega'_{\overline{T^0}}\|_1 + \|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1 &\leq \|\hat{\beta}_\lambda^{\text{JP}}\|_1 + \|\beta'_{S^0}\|_1 + 2\|\tilde{\beta}_{\overline{S^0}}\|_1 \\ &\quad + \lambda(\|\hat{\omega}_\lambda^{\text{JP}}\|_1 + \|\omega'_{T^0}\|_1 + 2\|\tilde{\omega}_{\overline{T^0}}\|_1), \end{aligned}$$

and rearranging terms gives

$$\begin{aligned} \|\beta'_{\overline{S^0}}\|_1 + \lambda\|\omega'_{\overline{T^0}}\|_1 &\leq (\|\hat{\beta}_\lambda^{\text{JP}}\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}}\|_1) - (\|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1) \\ &\quad + (\|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1) + 2(\|\tilde{\beta}_{\overline{S^0}}\|_1 + \lambda\|\tilde{\omega}_{\overline{T^0}}\|_1). \end{aligned}$$

Using (43) and the fact that $\|\hat{\beta}_\lambda^{\text{JP}}\|_1 + \lambda\|\hat{\omega}_\lambda^{\text{JP}}\|_1 \leq \|\tilde{\beta}\|_1 + \lambda\|\tilde{\omega}\|_1$ by minimality of the JP solution, we get

$$\|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1 \leq \rho(\|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1) + 2(\|\tilde{\beta}_{S^0}\|_1 + \lambda\|\tilde{\omega}_{T^0}\|_1),$$

hence

$$\|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1 \leq \frac{2}{1-\rho}(\|\tilde{\beta}_{S^0}\|_1 + \lambda\|\tilde{\omega}_{T^0}\|_1). \quad (44)$$

Now inequality (43) also implies

$$\begin{aligned} \|\beta'\|_1 + \lambda\|\omega'\|_1 &= \|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1 + \|\beta'_{S^c}\|_1 + \lambda\|\omega'_{T^c}\|_1 \\ &\leq (1+\rho)(\|\beta'_{S^0}\|_1 + \lambda\|\omega'_{T^0}\|_1) \end{aligned} \quad (45)$$

and continuing (45) with (44) gives the desired inequality. \square

Lemma 4. Let $\tilde{X} := \begin{bmatrix} X & \sqrt{n}I_n \end{bmatrix}$. Under assumptions *iv)*, *v)*, *vi)* and *vii)*,

$$\|\tilde{X}^T(\tilde{X}\tilde{X}^T)^{-1}\epsilon\|_2 \leq \frac{\sqrt{2}\sigma}{(\frac{\lambda_{\min}(\Sigma)}{4})(\sqrt{p/n}-1)^2+1)^{1/2}},$$

with probability at least $1 - 1.14^{-n} - 2e^{-\frac{1}{8}(\sqrt{p}-\sqrt{n})^2}$.

Proof. We have

$$\|\tilde{X}^T(\tilde{X}\tilde{X}^T)^{-1}\epsilon\|_2^2 = \epsilon^T(\tilde{X}\tilde{X}^T)^{-1}\epsilon \leq \frac{\|\epsilon\|_2^2}{\lambda_{\min}(\tilde{X}\tilde{X}^T)} = \frac{\sigma\|\frac{1}{\sigma}\epsilon\|_2^2}{\lambda_{\min}(\tilde{X}\tilde{X}^T)}.$$

Since $\|\frac{1}{\sigma}\epsilon\|_2^2 \sim \chi_n^2$, it is upper bounded by $2n$ with probability larger than $1 - 1.14^{-n}$ (a corollary of Lemma 1 in Laurent and Massart [2000]). So

$$\mathbb{P}\left(\|\tilde{\nu}\|_2 \leq \frac{\sqrt{2n}\sigma}{\sigma_{\min}(\tilde{X})}\right) \geq 1 - 1.14^{-n}. \quad (46)$$

Let us now bound $\sigma_{\min}(\tilde{X})$. One has

$$\sigma_{\min}^2(\tilde{X}) = \lambda_{\min}(\tilde{X}\tilde{X}^T) = \lambda_{\min}(XX^T + nI_n) = \sigma_{\min}^2(X) + n. \quad (47)$$

One can write $X = G\Sigma^{1/2}$ where $G \in \mathbb{R}^{n \times p}$ with $G_{ij} \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, thus

$$\sigma_{\min}(X) \geq \sigma_{\min}(G)\sigma_{\min}(\Sigma^{1/2}) = \sigma_{\min}(G)\sqrt{\lambda_{\min}(\Sigma)}. \quad (48)$$

Now it is known (see [Rudelson and Vershynin \[2010\]](#), eq. (2.3)) that

$$\sigma_{\min}(G) \geq \frac{1}{2}(\sqrt{p} - \sqrt{n}) = \frac{\sqrt{n}}{2}(\sqrt{p/n} - 1)$$

with probability at least $1 - 2e^{-\frac{1}{8}(\sqrt{p} - \sqrt{n})^2}$. Together with (47) and (48) this gives

$$\mathbb{P} \left(\sigma_{\min}(\tilde{X}) \geq \left(\frac{n\lambda_{\min}(\Sigma)}{4}(\sqrt{p/n} - 1)^2 + n \right)^{1/2} \right) \geq 1 - 2e^{-\frac{1}{8}(\sqrt{p} - \sqrt{n})^2}.$$

With (46), this implies

$$\mathbb{P} \left(\|\tilde{\nu}\|_2 \leq \frac{\sqrt{2}\sigma}{\left(\frac{\lambda_{\min}(\Sigma)}{4}(\sqrt{p/n} - 1)^2 + 1\right)^{1/2}} \right) \geq 1 - 1.14^{-n} - 2e^{-\frac{1}{8}(\sqrt{p} - \sqrt{n})^2}.$$

□

C Additional numerical experiments

C.1 Comparison of imputation methods

We perform other numerical experiments for Robust Lasso-Zero changing the initial naive imputation in order to show if it affects its performances. We compare two different ways of imputing beforehand of the Robust Lasso methodology when covariates are centered at $(1, \dots, 1)$: either an imputation by the empirical mean of the observed entries is considered (and thus should be around 1), or an arbitrary imputation by zero. The results are shown in [Figure 5](#) for non-correlated features in (a), and correlated ones in (b): no clear conclusion can be actually drawn.

Indeed, in Figure 5 (a) (with non-correlated variables), for $s = 3$ (left), both imputations lead to similar results; but for $s = 10$ (right), the mean imputation outperforms the zero-imputation with 5% MNAR values (top right), and the contrary holds for 10% of MCAR values (bottom right).

In Figure 5 (b) (with correlated design), the imputation strategies may lead to the same results (see for instance when $s = 3$ with MCAR values (left), or for $s = 10$ with 5% MCAR values (top right), or even for $s = 10$ and 20% MNAR values (bottom right)), which is the most challenging setting in these simulations. In other cases (i.e. all the settings with MNAR data and 20% MCAR data (bottom right)), the mean imputation outperforms the zero imputation.

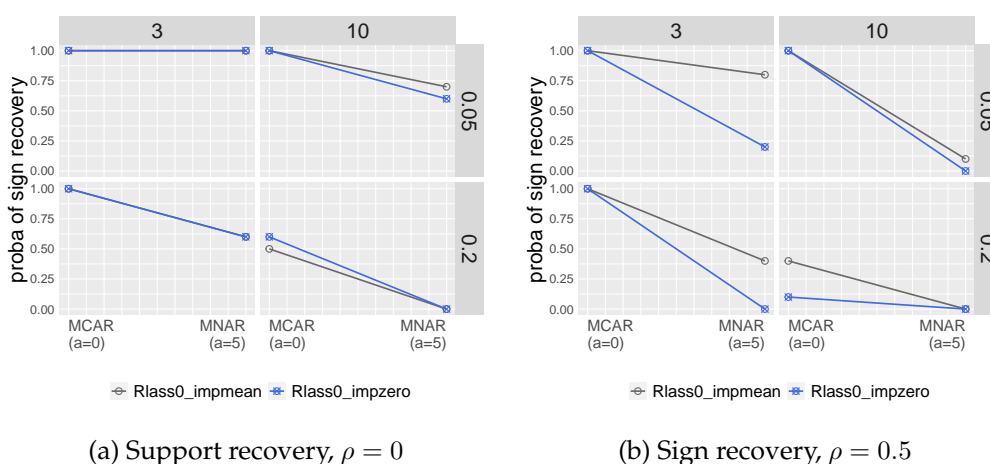


Figure 5: Support and sign recovery for Robust Lasso with mean imputation (Rlass0_impmean) and imputation by zeros (Rlass0_impzero), when $n = 100$, $p = 200$, $\sigma = 0.5$ and the covariates are centered at $(1, \dots, 1)$. An s -oracle tuning for the threshold is used for different sparsity levels $s = 3$ and $s = 10$ (subplots columns), and different proportions of missing values 5% or 20% (subplots rows), and different missing data mechanisms (MCAR vs MNAR).

C.2 Comparison of the Robust Lasso-Zero with other estimators

For a non-oracle hyperparameter tuning, we compare the Robust Lasso-Zero with the thresholded Robust Lasso proposed in [Nguyen and Tran \[2013a\]](#) and the thresholded lasso in [Pokarowski et al. \[2019\]](#).

For the Robust Lasso-Zero, we select the threshold τ by using the quantile universal threshold (QUT) at level $\alpha = 0.05$.

For the thresholded version of the Robust-Lasso and of the Lasso, we used the Generalized Information Criterion (GIC), proposed in the R package `DMRnet` [Pokarowski et al. \[2019\]](#). One could note that this criterion requires the estimation of the noise level, which is not needed for the Robust Lasso-Zero using quantile universal threshold instead. Besides, these methods require the choice of the parameter c controlling the amount of penalization in regard of the model complexity in the GIC, which means that these methods require one hyperparameter for the thresholding tuning, which is one more than needed for the Robust Lasso-Zero with QUT. It is recommended to use the default value $c = 2.5$ for linear regression. [Figure 7 \(a\)](#) shows that if we choose such a parameter, both thresholded versions of the Robust Lasso and the Lasso do not recover the sign. In [Figure 7 \(b\)](#) and (c), Robust Lasso-Zero with QUT outperforms the other methods in terms of s-TPR (the signed True Positive Rate) and s-FDR (the signed False Discovery Rate). The difficulty in choosing c may be due to the `DMRnet` code and not to the method, but this hyperparameter choice remains under the responsibility of the user, which shows the drawback of the information criteria involving more hyperparameters. In [Figure 6](#), we have also compared the Robust Lasso-Zero with the thresholded versions of the Robust Lasso and the Lasso, but by manually tuning an optimal

parameter controlling the penalization in the GIC.

Note here that the thresholded version of the Robust Lasso always outperforms the thresholded lasso, which was expected, because it allows to account for the corruptions.

In terms of sign recovery, when the covariates are correlated (Figure 6 (a)) or not (Figure 6 (b)), the Robust Lasso-Zero is at least equivalent to the two other strategies or outperforms them in the MNAR setting when $s = 10$ (right) and in the MCAR setting when $s = 10$ (right) (except in the correlated case for 5% of MCAR values). The Robust Lasso-Zero should be then the recommended strategy handling difficult cases in terms of sign recovery, i.e. when the percentage of missing values increases, when the mechanism is MNAR and when the covariates are correlated.

Besides, even if the Robust Lasso Zero may provide lower probabilities of sign recovery than the thresholded versions of the (Robust) Lasso in some particular cases (see for instance in Figure 6 (b) (left) for $s = 3$, 5% or with 20% MCAR values), the Robust Lasso Zero always remains competitive in terms of s-FDR (see Figure 6 (e) and (f)) and always give the best s-TPR (see Figure 6 (c) and (d)).

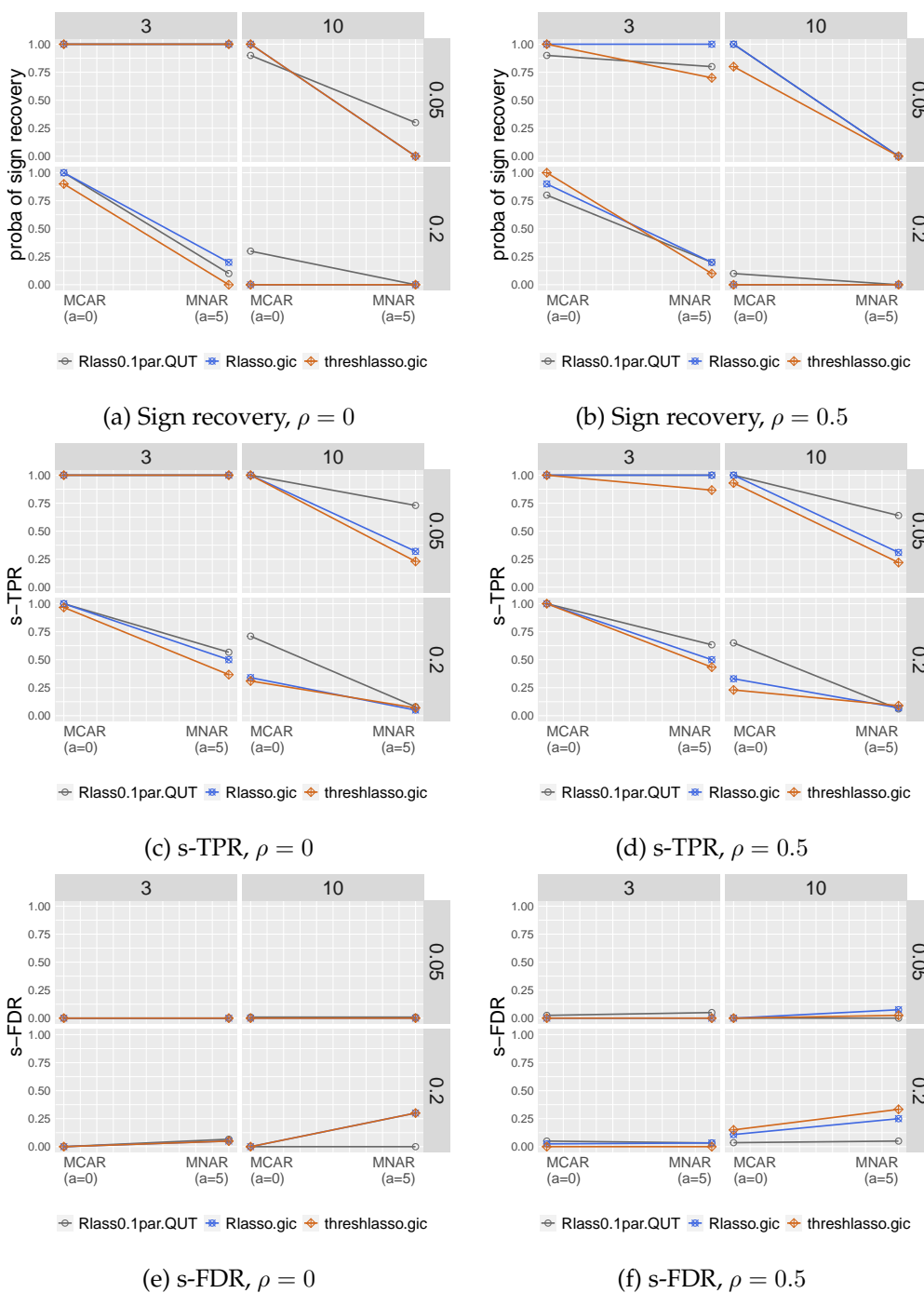


Figure 6: Sign recovery, s-TPR and s-FDR for the Robust Lasso-Zero, a thresholded version of the Robust Lasso and a thresholded version of the Lasso, when $n = 100$, $p = 200$, $\sigma = 0.5$ and the covariates are centered at $(1, \dots, 1)$. To tune the threshold, we use QUT for the Robust Lasso-Zero and GIC for the thresholded versions of both the Robust Lasso and the Lasso (the parameter c is chosen scaling as n). Different sparsity levels $s = 3$ and $s = 10$ (subplots columns), different proportions of missing values 5% or 20% (subplots rows), and different missing data mechanisms (MCAR vs MNAR) are considered.

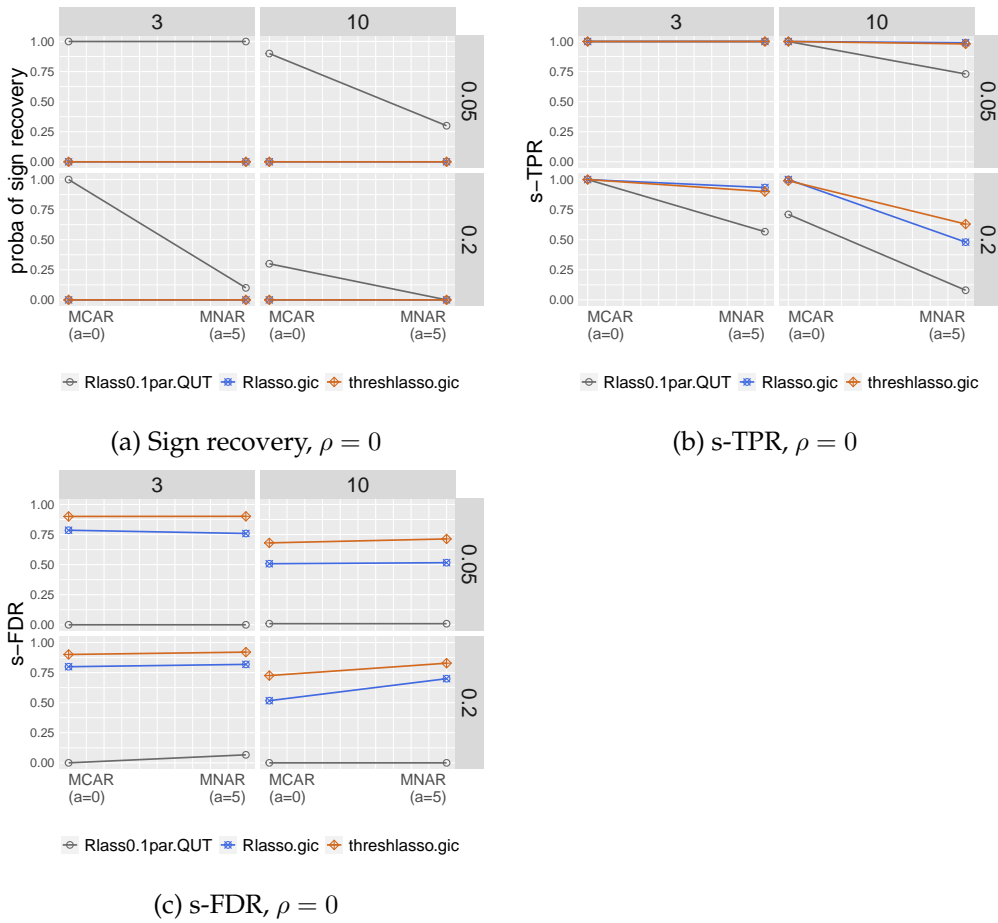


Figure 7: Sign recovery, s-TPR and s-FDR for the Robust Lasso-Zero, a thresholded version of the Robust Lasso and a thresholded version of the Lasso, when $n = 100$, $p = 200$, $\sigma = 0.5$ and the covariates are centered at $(1, \dots, 1)$. To tune the threshold, we use QUT for the Robust Lasso-Zero and GIC for the thresholded versions of both the Robust Lasso and the Lasso (we take for c the default value i.e. $c = 2.5$). Different sparsity levels $s = 3$ and $s = 10$ (subplots columns), different proportions of missing values 5% or 20% (subplots rows), and different missing data mechanisms (MCAR vs MNAR) are considered.

D Variables in the Traumabase dataset

The variables of the Traumabase dataset are:

- *Time.amb*: Time spent in the ambulance, *i.e.*, transportation time from accident site to hospital, in minutes.
- *Lactate*: The conjugate base of lactic acid.
- *Delta.Hemo*: The difference between the hemoglobin on arrival at hospital and that in the ambulance.
- *RBC*: A binary index which indicates whether the transfusion of Red Blood Cells Concentrates is performed.
- *SI.amb*: Shock index measured on ambulance.
- *DBP.min*: Minimum value of measured diastolic blood pressure in the ambulance.
- *SBP.min*: Minimum value of measured systolic blood pressure in the ambulance.
- *HR.max*: Maximum value of measured heart rate in the ambulance.
- *VE*: A volume expander is a type of intravenous therapy that has the function of providing volume for the circulatory system.
- *MBP.amb*: Mean arterial pressure measured in the ambulance.
- *Temp*: Patient's body temperature.
- *SI*: Shock index $SI = HR/SBP$ indicates level of occult shock based on heart rate and systolic blood pressure on arrival at hospital.
- *MBP*: Mean arterial pressure $MBP = (2DBP+SPB)/3$ is an average blood pressure in an individual during a single cardiac cycle.
- *HR*: Heart rate measured on arrival of hospital.

- *Age*: *Age*.