UNIVERSITÀ DEGLI STUDI DI PALERMO
Dottorato in Scienze Economiche, Statistiche, Psicologiche e Sociali
Dipartimento di Scienze Economiche, Aziendali e Statistiche
&
Istituto di Biomedicina ed Immunologia Molecolare "Alberto Monroy"
SECS-S/01 – Statistica

Induced smoothing in LASSO regression

IL DOTTORE
Giovanna Cilluffo

IL COORDINATORE
Prof. Vito M. R. Muggeo

IL TUTOR
Prof. Vito M.R. Muggeo

IL CO TUTOR
Prof.ssa Stefania La Grutta

CICLO XXX
ANNO CONSEGUIMENTO TITOLO 2018

Università degli Studi di Palermo

Abstract

Induced smoothing in LASSO regression

by Giovanna CILLUFFO

The thesis is being carried out with the National research Council at the Institute of Biomedicine and Molecular Immunology "Alberto Monroy" of Palermo, where I am a fellow, under the supervision of MD Stefania La Grutta. Our research unit is focused on clinical research in allergic respiratory problems in children. In particular, we are interested in to assess the determinants of impaired lung function in a sample of outpatient asthmatic children aged between 5 and 17 years enrolled from 2011 to 2017. Our dataset is composed by $n = 529$ children and several covariates regarding host and environmental factors.

This thesis focuses on hypothesis testing in lasso regression, when one is interested in judging statistical significance for the parameters involved in the regression equation. To get reliable p-values we propose a new lasso-type estimator relying on the recent idea of induced smoothing which allows to obtain appropriate covariance matrix and Wald statistic relatively easily. In addition, we discuss the score statistic to carry out interval estimation on the regression coefficients in LASSO regression. Some simulation experiments reveal our approaches exhibits better performance when contrasted with the recent inferential tools in the lasso framework. Finally, we analysed data regarding asthmatic out-patient children which motivated our project.

Acknowledgements

I would like to express my special appreciation and thanks to Prof. Stefania La Grutta. I would like to thank her for believing in me, for encouraging my research, for allowing me to grow as a research scientist, and for investing and continuing to invest in me.

I want to thank my supervisor Prof. Vito Muggeo for his patience, motivation, and immense knowledge, thanks for the research experiences and possibilities he has given to me.

A special thank to my big family. No word can not express how grateful I am to my family for their support and love.

At the end I would like express appreciation to my mate, friend, colleague and "nerd" husband Gianluca. *Thanks for your collaboration in my research, for bearing my nervous states, for your comprehension and for our work weekends. I could not wish for a better life partner!*

Contents

List of Figures

List of Tables

Chapter 1

Motivating problem

Asthma is a common and potentially serious chronic disease that imposes a substantial burden on patients and their families. It causes respiratory symptoms, limitation of activities and exacerbations that sometimes require urgent health care (GINA, 2017). Objective assessment of asthma is related to the degree of airways obstruction measured by spirometry. Spirometry is needed to monitor children with asthma for signs of increasing airway obstruction (Strunk et al., 2006). The long-term goals of asthma management are symptom control and risk reduction. Impaired lung function risk can be minimized by optimizing asthma medications and by identifying and treating modifiable risk factors. In this framework we are interested in to assess numerous risk factors of impaired lung functions. Classical tools could not be easy applicable when the number of variables increases and if applied can cause regression coefficients and p-values to be misleading.

This thesis focuses on hypothesis testing and interval estimation in lasso regression, when one is interested in judging statistical significance for the parameters involved in the regression equation. The proposed method will be applied in order to assess the determinants of impaired lung function in a sample of outpatient asthmatic children aged between 5 and 17 years enrolled from 2011 to 2017. Our dataset is composed by $n = 529$ children and several covariates regarding host and environmental factors.

Our methodology enjoys many advantages. Firstly, it quantifies the uncertainty of the estimates; secondly, p-values and confidence intervals are easily derived with usual inferential tools. In medical research, a common practice consists in to adjust regression models for possible confounders since they can influence the magnitude of the relationship between the independent variable and outcome. Our proposal, allows to consider all measured information simultaneously even if the number of variables overcomes the sample size. A possible disadvantage of our

method is that it does not return exactly zero estimates, but actually the resulting p-value can be used to discard or not the covariates. Another drawback of the method is that estimates are biased, however, when the sample size is greater than the number of variables the bias can be perfectly computed and estimates can be corrected.

Existing methods are reported in the next sections of this chapter.

1.1 The Least Absolutes Shrinkage and Selection Operator

Regression models are widely used and well-established statistical tools in many fields of applied research. A linear regression model assumes that

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{y}$ is the response vector, $\boldsymbol{X}$ is $n \times p$ matrix of covariates, $\boldsymbol{\beta}$ is the vector of unknown parameters and $\boldsymbol{\epsilon}$ is the erratic component normally distributed.

The ordinary least squares estimator (Dismuke and Lindrooth, 2006; McCullagh, 1984) is obtained minimizing the least square objective function:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 \tag{1.1}$$

which leads to

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Typically all of the least-squares estimates will be nonzero and unique, however, when the sample size is smaller than the number of parameters ($n \leq p$) solutions are not unique and these solutions almost surely overfit the data.

Tibshirani (1996) proposed the Least Absolutes Shrinkage and Selection Operator (LASSO) which is a very elegant and relatively widespread solution to carry out variable selection and parameter estimation simultaneously when $n \leq p$. The LASSO objective function to be minimized at fixed λ is

$$\frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{\beta}||_1$$

where $\lambda||\boldsymbol{\beta}||_1$ is the L_1 penalty, λ parameter controls the amount of penalization. The larger λ, the more shrunken the estimate. To illustrate, Figure 1.1 shows LASSO estimation from geometrical point of view. The

ellipses represent the contours of the objective functions, the diamond (blue square) is the LASSO constraint and the dot is the point where contours is “tangent” to the constraint, i.e., the penalized estimate. The LASSO performs L_1 shrinkage, so that there are "corners" in the constraint, if the sum of squares "hits" one of these corners, then the coefficient corresponding to the axis is zero.

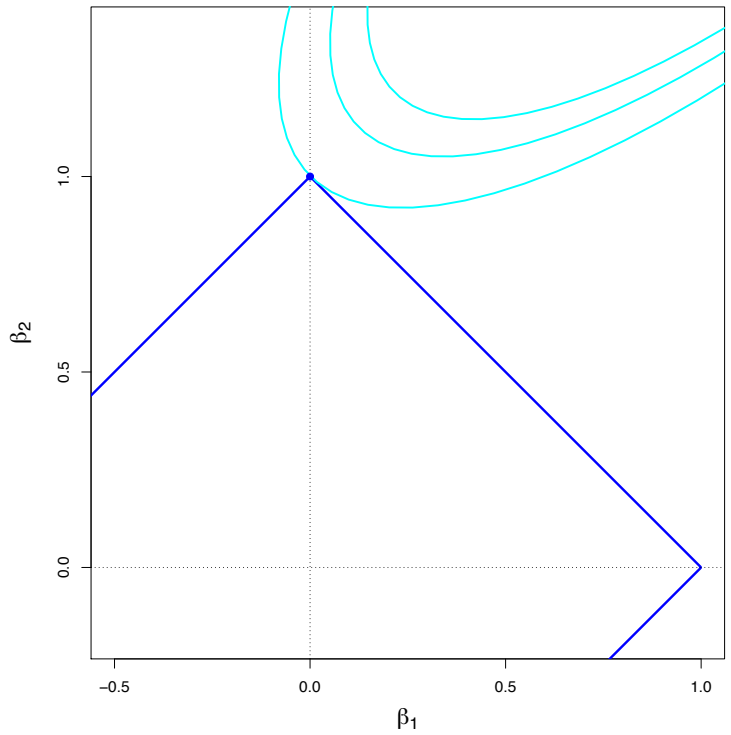

FIGURE 1.1: LASSO estimation, the blue square represents the constraints area $|\beta_1| + |\beta_2| \leq c$, while light blue ellipses are the contours of the least square errors function.

Typically the LASSO estimator is biased and its distribution is highly non-normal, as shown in Figure 1.2, posing issues to valid inference (Knight and Fu, 2000; Pötscher and Leeb, 2009; Kyung et al., 2010; Jagannath and Upadhye, 2016). Point estimation can be performed quite efficiently with current algorithms. Efron et al. (2004) proposed an efficient algorithm for computing LASSO estimate based on the entire regularization path, i. e. the entire path of the coefficient estimates as λ varies. Friedman et al. (2010) develop fast algorithms for estimation of generalized linear

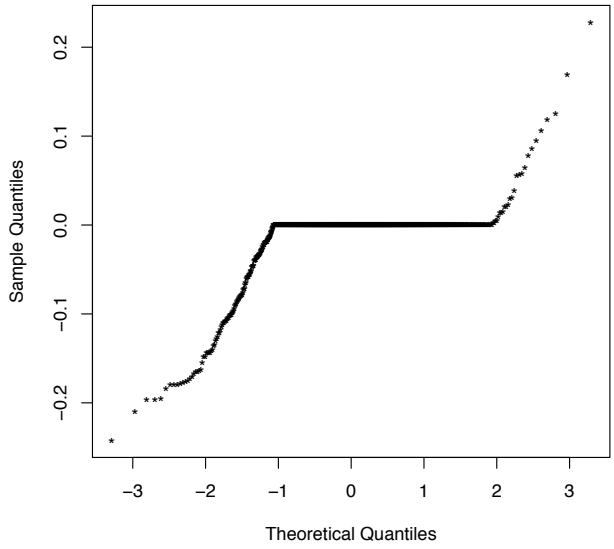

FIGURE 1.2: Sampling distribution of a zero coefficient in a simulation study

models with L_1 penalty using cyclical coordinate descent along a regularization path. Augugliaro et al. (2013) proposed a method based on the geometrical structure underlying the generalized linear model which allows to define a generalization of the equiangularity condition. Other methods have been proposed yet for the LASSO during the last decade (Meinshausen and Bühlmann, 2006; Zhang and Huang, 2008; Beck and Teboulle, 2009; Candes and Plan, 2009; Tutz and Gertheiss, 2016). A possible actual limitation of LASSO method is computation of standard errors and consequently inference, a literature review on uncertainty estimation and on inference is presented in section 1.1.2.

1.1.1 Theoretical conditions

In the last years, it has become clear that some theoretical conditions play a central role to guarantee model consistency (Zhao and Yu, 2006), defined as a correct amount of regularization that selects the true model, and for sparsity pattern recovery (Wainwright, 2009) of the LASSO. In particular, let $\boldsymbol{\beta} = (\beta_1, \dots \beta_q, \beta_{q+1}, \dots, \beta_p)$ where $\beta_j \neq 0$ for $j = 1, \dots q$ and $\beta_j = 0$ for $j = q+1, \dots p$, $\boldsymbol{X}_1$ the first q column of the $\boldsymbol{X}$ matrix and

$\boldsymbol{X}_2$ the last $p-q$; we define a block matrix $C = \boldsymbol{X}^T\boldsymbol{X}/n$ obtained by setting $C_{11} = \boldsymbol{X}_1^T\boldsymbol{X}_1/n$, $C_{12} = \boldsymbol{X}_1^T\boldsymbol{X}_2/n$, $C_{21} = \boldsymbol{X}_2^T\boldsymbol{X}_1/n$, $C_{22} = \boldsymbol{X}_2^T\boldsymbol{X}_2/n$. The strong *irrepresentable* condition states that there exists a positive constant vector η:

$$|C_{21}C_{11}^{-1}sign(\beta)| < \mathbf{1} - \eta$$

where $\mathbf{1}$ is a $p-q$ vector of one's and the inequality holds element-wise. In other word, LASSO selects the true model consistently if and (almost) only if the predictors that are not in the true model are "irrepresentable" by predictors that are in the true model. Regarding the sparsity pattern recovery the number of nonzero coefficients has to be $\beta_1, \ldots, \beta_k$ (out of p covariates) such that $k \leq n/(2\log p)$ having magnitudes at least $\beta_{min} = c\sigma\sqrt{2\log p}$ for some unspecified numerical constant c.

1.1.2 Inference in LASSO regression

Tibshirani (2011) reports *'we need better tools for inference with the LASSO and related methods. More basically, we need reliable ways to assess the sampling variability of the LASSO estimates'*. In fact just few proposals to compute standard errors, have been discussed in literature. Tibshirani (1996) proposed to use $\sum|\beta_j| \approx \sum\beta_j^2/|\beta_j|$ which is a sort of ridge-like approximation of the absolute value function. Fan and Li (2001) exploited, instead, the sandwich formula in more general likelihood settings. Unfortunately such approximations are unsatisfactory in practice, since they lead to a zero standard error for a zero point estimate which prevents to quantify uncertainty for the variables left out of the model. Osborne et al. (2000) derived a covariance matrix ensuring positive standard errors for all coefficient estimates, but the proposal does not quantify appropriately variability of the estimators (Kyung et al., 2010). A bootstrap approach was also discussed, but it has been shown to be inconsistent (Beran, 1982; Kyung et al., 2010). Several are the proposal to make inference in high dimensional setting, recently (Bühlmann et al., 2013) proposed a method for constructing p-values for general hypotheses in a high-dimensional linear model, based on Ridge estimation with an additional correction term. Wasserman and Roeder (2009) suggested to randomly split the data into three parts, the first part is used to estimate different models for each λ, the second part is used to select one model by cross-validation and the third part is used to find least squere estimate and to eliminate some variables using hypothesis testing. Zhang and Cheng (2017) and Dezeure et al. (2016) proposed a method based on bootstrap, Lan et al.

(2016) developed a new testing procedure introducing the correlated predictors screening method to control for predictors that are highly correlated with the target covariate. Meinshausen and Bühlmann (2009) used a multisplit method, related to Wasserman and Roeder (2009), for assigning statistical significance and constructing conservative p-values for high-dimensional problems. Minnier et al. (2012) proposed a perturbation resampling based procedures to approximate the distribution of a general class of penalized parameter estimators along with their covariance matrix. Regarding interval estimation, Meinshausen (2015) showed that a "group-bound" confidence interval can be derived without making any assumptions on the design matrix.

Desparsified or debiased LASSO has been proposed by Van de Geer et al. (2014), Zhang and Zhang (2014), and Javanmard and Montanari (2014) aimed at desparsifying the regularized solution, namely to correct the estimates returned by LASSO. These authors use a debiased version of the LASSO estimator

$$\hat{\boldsymbol{\beta}}^d = \hat{\boldsymbol{\beta}}_\lambda + 1/n\boldsymbol{\Theta}\boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_\lambda)$$

where $\hat{\beta}_\lambda$ is the LASSO estimate at λ, and $\boldsymbol{\Theta}$ is an approximate inverse of $\boldsymbol{\Sigma} = \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}/n$. When $n > p$ then $\boldsymbol{\Theta}$ is invertible and $\hat{\boldsymbol{\beta}}^d$ would be exactly unbiased, however when $n < p$, $\boldsymbol{\Theta}$ is not invertible and it needs to find an approximate inverse. The authors suggest different approximation of $\boldsymbol{\Theta}$, in particular Van de Geer et al. (2014) use a neighborhood-based methods to impose sparsity on the components and Javanmard and Montanari (2014) solve a convex program, Boot and Nibbering (2017) use diagonally scaled Moore-Penrose pseudoinverse. Related works can be found in Belloni et al. (2014) which consider a single main regressor and multidimensional control covariates. Janková et al. (2015) proposed a desparsified estimator based on the graphical LASSO to build confidence intervals. The desparsified LASSO methods (Hdi) return both single testing p-values as well as multiple testing corrected p-values, and the confidence intervals for individual parameters (Dezeure et al., 2015).

More specifically for hypothesis testing problems, Lockhart et al. (2014) have discussed the covariance test (covTest) for a newly added coefficient along the LASSO regularization path, based on the difference between the fitted values of the models with and without the relevant covariate entering the active set at proper lambda value. While p-values are returned for each covariate, the procedure assumes that all signal variables enter the LASSO solution path first; moreover covTest does not allow to obtain results at the same given value of the tuning parameter λ, such as

the 'optimal' one as returned by cross validation or any other criterion. Formally, let the path $\hat{\boldsymbol{\beta}}_\lambda$ is a continuous and piecewise linear function of λ, with knots at values $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_k \geq 0$, the covariance test is defined:

$$T_k = \left(\boldsymbol{y}^T \hat{\boldsymbol{\mu}}(\lambda_{k+1}) - \boldsymbol{y}^T \tilde{\boldsymbol{\mu}}_A(\lambda_{k+1})\right) / \sigma^2$$

where A indicates the active set just before λ_k; $\hat{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}_A$ are the fitted values at λ_{k+1} given by including and leaving out the j-th predictor into the current active set, respectively, and σ^2 is the square of dispersion parameter. Under the null hypothesis, T_k is asymptotically distributed as a standard exponential random variable, i.e.

$$T_k \xrightarrow{d} \text{Exp}(1)$$

Finally, yet another approach to inference in LASSO regression is the so-called selective inference (hereafter postSel) discussed in Lee and Taylor (2014); Lee et al. (2016); Tibshirani et al. (2016). The authors consider post-selection inference, namely inference given the selected model, which use the truncated Normal distribution for the parameter estimators with a fixed λ value. This approach, however, critically depends on ability of LASSO to screen, i.e. to pick up a model including all non-noisy variables. If the interest variable does not enter the model, no corresponding inference measure can be obtained. Formally, starting from the usual least squares problem with an additional L_1 penalty on the coefficients, the authors define an F statistic, given by the polyhedral set $\{A\boldsymbol{y} \geq b\}$. Assuming $\boldsymbol{y} \sim N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$, with $\boldsymbol{\theta} \in \mathbb{R}^n$ unknown and $\boldsymbol{\Sigma} \in \mathbb{R}^n$ known, for a fixed $\boldsymbol{\eta} \in \mathbb{R}^n$ the aim is to make inference for $\boldsymbol{\eta}^T\boldsymbol{\theta}$. The distribution of any linear function $\boldsymbol{\eta}^T\boldsymbol{y}$, given the polyhedral set $\{A\boldsymbol{y} \geq b\}$, can be written as the conditional distribution:

$$\boldsymbol{\eta}^T\boldsymbol{y} \mid V^{low} \leq \boldsymbol{\eta}^T\boldsymbol{y} \leq V^{up},\ V^0 \leq 0$$

where V^{low} and V^{up} are bound functions independent of $\boldsymbol{\eta}^T\boldsymbol{y}$. Since $\boldsymbol{\eta}^T\boldsymbol{y}$ has Gaussian distribution, the bounded quantity is a truncated Gaussian distribution. The Cumulative density function of a generic $N(\mu, \sigma^2)$ random variable truncated to lie in $[a; b]$, is

$$F^{[a,b]}_{\mu,\sigma^2}(x) = \frac{\Phi((x-\mu)/\sigma) - \Phi((a-\mu)/\sigma)}{\Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma)}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. The statistic $F^{[V^{low},V^{up}]}_{\eta^T\theta,\eta^T\Sigma\eta}(\eta^T y)$ is a pivotal quantity and leads to conditional p-values for hypothesis testing, and consequently, to conditional confidence intervals.

1.1.3 Tuning parameter selection

Another important topic, which we not will explore in depth, in penalized methods is the choice of the tuning parameter. Different approaches have been proposed in literature, the most famous being cross-validation and generalized cross-validation (Craven and Wahba, 1978), Akaike Information Criteria (AIC) (Akaike, 1998), Bayesian Information Criteria (BIC) (Zou et al., 2007) and Generalized Information Criteria (GIC) (Zhang et al., 2010). Their formulation are given as follows:

$$
\begin{aligned}
\text{CV}(\lambda) &= \sum_{s=1}^{k} \sum_{(x_k,y_k)\in T^{-s}} (y_k - x_k^T \hat{\boldsymbol{\beta}}_\lambda)^2 \\
\text{GCV}(\lambda) &= \frac{||\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\lambda||_2^2}{n(1-df/n)^2} \\
\text{AIC}(\lambda) &= \frac{\log(||\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\lambda||_2^2)}{n} + \frac{2\text{df}}{n} \\
\text{BIC}(\lambda) &= \frac{\log(||\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\lambda||_2^2)}{n} + \frac{\log(n)\text{df}}{n} \\
\text{GIC}(\lambda) &= \log(||\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_\lambda||_2^2) + \text{c}_n \log(p)\text{df}n^{-1}
\end{aligned}
$$

almost all criteria require the degrees of freedom in theri formulation. df are often used to quantify the model complexity of a statistical modeling, in LASSO regression the number of nonzero coefficients is an unbiased estimate of the df (Zou et al., 2007; Tibshirani and Taylor, 2012). How to choose the criterion is not our topic, however, Zou et al. (2007) suggested to use BIC as the model selection criterion when the sparsity of the model is the primary concern, Fan and Tang (2013) proposed to select the tuning parameter by optimizing the GIC, many authors suggested to use CV (Park and Hastie, 2007; Friedman et al., 2007; Breheny and Huang, 2011).

1.1.4 The adaptive LASSO

Zou (2006) proposed a variant of the LASSO, called the adaptive LASSO, where adaptive weights are used for penalizing different coefficients in

the L_1 penalty. In this context the objective function becomes:

$$\frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||_2^2 + \lambda||\boldsymbol{w} \odot \boldsymbol{\beta}||_1$$

where$||\boldsymbol{w}\odot\boldsymbol{\beta}|| = \sum_{j=1}^{p} |\beta_j w_j|$ and typically $\boldsymbol{w} = 1/\hat{\boldsymbol{\beta}}_{\text{OLS}}$. It is well known the LASSO penalty shrinks all coefficients towards zero causing important bias for the nonzero estimates, and the adaptive LASSO tries to attenuate such bias by introducing the $\boldsymbol{w}$s understood to weight the importance of the coefficients. Also in this case, the finite-sample distribution of the adaptive LASSO estimates in not normal, in particular Pötscher and Schneider (2009) showed that it is a mixture of a singular normal distribution and an absolutely continuous nonnormal distribution .

The thesis is structured as follows: Chapter 2 describes our proposal into detail, Chapter 3 reports results of some simulation studies, Chapter 4 focuses on data analysis and finally some conclusions are reported in Chapter 5.

Chapter 2

The induced smoothing in LASSO

2.1 The induced smoothing

The idea of 'natural' or induced smoothing (hereafter IS) has been introduced by Brown and Wang (2005). The authors focus on the estimation of standard errors or covariance matrices to deal with non-smooth estimating equation which prevents the usual estimating algorithms and asymptotics to be applied, the method can be employed to any statistical estimation. We focus on linear regression, considering the following linear model $\boldsymbol{y} \sim \boldsymbol{X}\beta + \epsilon$ with $\epsilon \sim N(0, \sigma)$, where $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ denote the vector of observed response, $\boldsymbol{X}$ be the $n \times p$ design matrix with regression coefficients $\boldsymbol{\beta}$ and let $U(\boldsymbol{\beta})$ be a vector of estimating equations for $\boldsymbol{\beta}$. The resulting estimator $\hat{\beta}$ is obtained by solving $U(\boldsymbol{\beta}) = 0$. When $U(\boldsymbol{\beta})$ is smooth standard errors can be obtained by using the usual sandwich formula $\boldsymbol{V} = \boldsymbol{\tilde{U}'^{-1}\mathcal{I}\tilde{U}'^{-1}}$, with $\mathcal{I}$ indicating the Fisher information matrix. However, the estimating function $U(\boldsymbol{\beta})$ is often non-smooth, as a consequence, standard error computation for $\hat{\beta}$ represents a challenging issue. The idea behind the elegant IS approach allows to overcome this issue, assuming a limit multi-normal distribution for $\hat{\boldsymbol{\beta}}$ with covariance matrix $\boldsymbol{V}$ that allows to write $\boldsymbol{V}^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \boldsymbol{z}$ or equivalently $\hat{\boldsymbol{\beta}} \sim \boldsymbol{\beta} + \boldsymbol{V}^{1/2}\boldsymbol{z}$, where $\boldsymbol{V}^{1/2}$ is the 'square root' matrix of $\boldsymbol{V}$, and $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I}_p)$ are p-dimensional multi-normal standard realizations. Roughly speaking, because of the relationship between $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$, the idea is to consider the estimating equation in terms of perturbations, i.e. $U(\boldsymbol{\beta} + \boldsymbol{V}^{1/2}\boldsymbol{z})$. Formally, the IS estimating equation is defined as expectation over multi-normal random perturbations, namely

$$\widetilde{U}(\boldsymbol{\beta}) = E_z[U(\boldsymbol{\beta} + \boldsymbol{V}^{1/2}\boldsymbol{z})], \tag{2.1}$$

where $E_z[\,\cdot\,]$ represents expectation over $\boldsymbol{z} \sim N(0, I_p)$, standard multi-normal random perturbations. In this way, $\widetilde{U}(\boldsymbol{\beta})$ is smooth, thus the

slope matrix $\widetilde{U}'(\boldsymbol{\beta})$ exists and the usual sandwich formula applies to compute the covariance matrix of estimator $\hat{\beta}$. Induced smoothing can be applied in several contexts for example in quantile regression in order to smooth the first derivative of the quantile objective function, or in weighted rank regression for the accelerated failure time model (Brown and Wang, 2007).

2.2 The induced smoothing in LASSO

It is well known that LASSO problem in equation (1.1) can be expressed using the sub-gradient definition (Tibshirani and Taylor, 2012). Conversely, we express the estimating equation in LASSO, using the Heaviside step function (Bracewell and Bracewell, 1986) as follows:

$$U(\boldsymbol{\beta}) = \boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda\{2\,\mathrm{I}(\boldsymbol{\beta} > 0) - \mathbf{1}_p\}, \tag{2.2}$$

where $\mathrm{I}(\cdot)$ is the usual indicator function equal to one when its argument is true. $U(\boldsymbol{\beta})$ is clearly non-smooth thus preventing application of usual asymptotic theory for computing covariance matrix and carrying out inference. The IS paradigm replaces (2.2) with the naturally smoothed counterpart: to define it, as previously proposed by Knight and Fu (2000) and Pötscher and Leeb (2009), we recognize that the distribution of $\hat{\boldsymbol{\beta}}$ can be seen as a mixture of a standard Normal and a singular Normal distribution with a pointmass at zero. We assume a standard Normal for the continuous part and a Normal distribution with negligible variance for the zero mass for the other, it is possible to write for the LASSO estimator

$$v^{-1/2}(\hat{\beta} - \beta) \sim \nu \qquad \text{where} \qquad f(\nu) \approx c\,\phi(\nu) + (1-c)\,\phi_\epsilon(\nu). \tag{2.3}$$

$\phi(\cdot)$ is the pdf of a standard Normal, $\phi_\epsilon(\cdot)$ the pdf of a zero-mean Normal with very small variance ($\epsilon = 10^{-6}$, say), and c is the (unknown) mixture weight. Application of the IS exploits perturbations from the aforementioned mixture distribution,

$$\begin{aligned}
\tilde{U}(\beta) = E_\nu[U(\beta + v^{1/2}\,\nu)] &= \int U(\beta + v^{1/2}\,\nu) f(\nu) d\nu \\
&= \int U(\beta + v^{1/2}\,\nu)\,\{c\phi(\nu) + (1-c)\phi_\epsilon(\nu)\}\,d\nu \\
&= c\int U(\beta + v^{1/2}\,\nu)\phi(\nu)\mathrm{d}\nu + (1-c)\int U(\beta + v^{1/2}\nu)\phi_\epsilon(\nu)\mathrm{d}\nu.
\end{aligned}$$

Thus, assuming the aforementioned 2-components mixture, the overall smooth penalty turns out to be

$$\mathcal{P}(\beta, v; c) = c\{2\Phi(\beta/v^{1/2}) - 1\} + (1-c)\{2\Phi_\epsilon(\beta/v^{1/2}) - 1\},$$

where $\Phi(\cdot)$ and $\Phi_\epsilon(\cdot)$ are the Normal cumulative distribution functions. The mixture weight c depends on several factors, including the true signal, the error variance and covariate scale. However its value is unknown in practice and to account for that, we propose to consider the penalty averaged over the range of c according to its distribution function $F(c)$,

$$\mathcal{P}(\beta, v) = \int_0^1 \mathcal{P}(\beta, v; c)\, \mathrm{d}F(c). \tag{2.4}$$

To reflect uncertainty in c, we assume $c \sim \text{unif}(0,1)$, and therefore an empirical version of resulting average penalty (2.4) is simply

$$\overline{\mathcal{P}}(\beta, v) = \sum_{k=1}^{K} \mathcal{P}(\beta, v; c_k)/K \tag{2.5}$$

where $c_1, c_2, \ldots, c_K$ are K equispaced values in (0,1). Note the averaged penalty is independent of c but still depending on β and v. $\overline{\mathcal{P}}$ seems to be a rational trade-off which balances the 2-component mixtures. Simulation study, in next chapter, shows that the choice of c is negligible.

The IS estimating equation turns out to be

$$\widetilde{\boldsymbol{U}}(\boldsymbol{\beta}) = \boldsymbol{X}^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda \overline{\mathcal{P}}(\boldsymbol{\beta}, \boldsymbol{v}) \tag{2.6}$$

where $\overline{\mathcal{P}}(\boldsymbol{\beta}, \boldsymbol{v}) = K^{-1}\sum_k \mathcal{P}(\boldsymbol{\beta}, \boldsymbol{v}; c_k)$ and $\mathcal{P}(\boldsymbol{\beta}, \boldsymbol{v}; c_k) = c_k\{(2\Phi(\boldsymbol{\beta}/\boldsymbol{v}^{1/2}) - \mathbf{1}_p)\} + (1-c_k)\{(2\Phi_\epsilon(\boldsymbol{\beta}/\boldsymbol{v}^{1/2}) - \mathbf{1}_p)\}$. In the penalties, $\boldsymbol{v}$ is the main diagonal of $\boldsymbol{V} = \text{var}(\hat{\boldsymbol{\beta}})$, and $\boldsymbol{a}/\boldsymbol{b}$ means the element-wise ratio of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$; Figure 2.1 compares the estimating equation in LASSO, OLS and in IS-Lasso for a nonzero coefficient (left panel) and for a zero coefficient (right panel). For a non zero coefficient, IS-Lasso and LASSO are overlapped, conversely for a zero coefficient LASSO is exactly zero while IS-Lasso is different from zero even if very close to the LASSO estimate.

$\widetilde{\boldsymbol{U}}(\boldsymbol{\beta})$ is smooth, thus the slope matrix $\widetilde{\boldsymbol{U}}'(\boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}}\widetilde{U}(\boldsymbol{\beta})$ exists and it is found to be

$$\widetilde{\boldsymbol{H}}(\boldsymbol{\beta}) = \widetilde{\boldsymbol{U}}'(\boldsymbol{\beta}) = \boldsymbol{X}^T\boldsymbol{X} + \lambda\, \overline{\mathcal{P}'}(\boldsymbol{\beta}, \boldsymbol{v}) \tag{2.7}$$

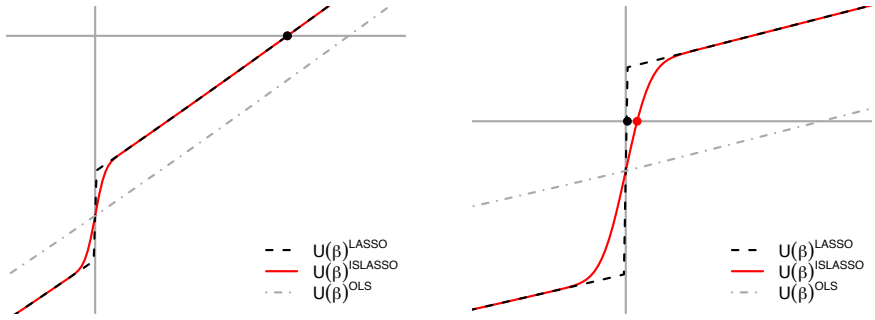

FIGURE 2.1: Comparison of estimating equation for IS-Lasso (red solid line), OLS (gray dotted line) and LASSO (black dashed line). Left panel reports an example for a nonzero coefficients, right panel reports an example for a zero coefficient.

where the penalty derivative $\overline{\mathcal{P}'}(\boldsymbol{\beta}, \boldsymbol{v}) = \frac{\partial}{\partial \boldsymbol{\beta}} \overline{\mathcal{P}}(\boldsymbol{\beta}, \boldsymbol{v})$ is

$$\overline{\mathcal{P}'}(\boldsymbol{\beta}, \boldsymbol{v}) = \frac{1}{K} \sum_k \{c_k \mathrm{diag}(2\phi(\boldsymbol{\beta}/\boldsymbol{v}^{1/2})/\boldsymbol{v}^{1/2}) + (1-c_k)\mathrm{diag}(2\phi_\epsilon(\boldsymbol{\beta}/\boldsymbol{v}^{1/2})/\boldsymbol{v}^{1/2})\},$$

and $\mathrm{diag}(\cdot)$ means a diagonal matrix.

Existence of $\widetilde{\boldsymbol{H}}(\boldsymbol{\beta})$ allows to apply the sandwich formula to compute the covariance matrix (e.g. Royall, 1986) of $\hat{\boldsymbol{\beta}}$, i.e.

$$\boldsymbol{V} = \widetilde{\boldsymbol{H}}(\hat{\boldsymbol{\beta}})^{-1} \, \boldsymbol{\mathcal{I}} \, \widetilde{\boldsymbol{H}}(\hat{\boldsymbol{\beta}})^{-1}, \tag{2.8}$$

where $\hat{\boldsymbol{\beta}}$ is the final value at convergence, and $\boldsymbol{\mathcal{I}} = \boldsymbol{X}^T \boldsymbol{X}/\sigma^2$ is the usual Information matrix independent of $\hat{\boldsymbol{\beta}}$.

Clearly $\widetilde{\boldsymbol{U}}$ requires $\boldsymbol{V}$ (via the main diagonal $\boldsymbol{v}$, see (2.6)), and in turn $\boldsymbol{V}$ needs $\widetilde{\boldsymbol{U}}$ (via the first derivative $\widetilde{\boldsymbol{H}}$, see (2.8)). Hence an iterative procedure is called for, alternating computation of $\widetilde{\boldsymbol{U}}$ and $\boldsymbol{V}$. More specifically:

0. Initialize: fix initial guesses for $\boldsymbol{V}$ and $\boldsymbol{\beta}$; in particular, we set $\boldsymbol{V}^{(0)} = \boldsymbol{I}_p/n$;

1. compute $\widetilde{\boldsymbol{U}}(\boldsymbol{\beta})$ according to (2.6) and solve $\widetilde{\boldsymbol{U}}(\boldsymbol{\beta}) = \boldsymbol{0}$ to get a new update of $\boldsymbol{\beta}$;

2. compute $\widetilde{\boldsymbol{H}}$ at the current $\boldsymbol{\beta}$ value, and then update $\boldsymbol{V}$ according to (2.8);

3. update the guesses for $\boldsymbol{V}$ and $\boldsymbol{\beta}$ and repeat steps 1. and 2. till convergence.

The IS-Lasso algorithm appears quite straightforward, it just needs a few Newton-Raphson steps:

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\mathbf{0}} - \widetilde{\boldsymbol{H}}(\hat{\boldsymbol{\beta}}^{\mathbf{0}})^{-1}\widetilde{\boldsymbol{U}}(\hat{\boldsymbol{\beta}}^{\mathbf{0}})$$

where $\hat{\boldsymbol{\beta}}^{\mathbf{0}}$ is the initial guess for $\boldsymbol{\beta}$. Moreover, as discussed in Brown and Wang (2005), convergence is reliable and rapid; in the proposed IS-Lasso framework, we have experienced convergence in less than 10 iterations when $n > p$, and somewhat slower when $n \leq p$. IS-Lasso substantially replaces the non-smooth absolute value function with a smooth approximation depending on the estimate standard error. As an example, Figure 2.2 portrays the effect of the induced smoothing on the LASSO penalty $\sum_{j=1}^{2} |\beta_j|$ for two different standard error estimates.

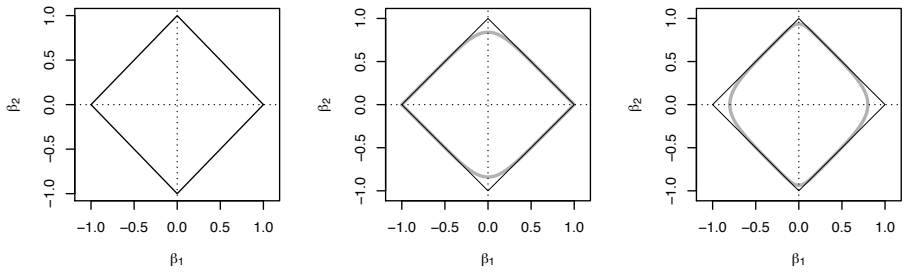

FIGURE 2.2: Contrasting the plain LASSO penalty (diamond, black thin lines) with the induced smoothed counterpart (thick gray lines). The amount of smoothing at kink depends on the variance of the corresponding estimator and it is determined automatically by data.

The smaller the standard error, the closer the approximation. For sample size with n independent units, the elements of $\boldsymbol{V}$ decrease at $n^{-1/2}$ rate indicating that the IS-Lasso is asymptotically equivalent to the LASSO; however in finite sample the estimating functions (2.6) will be smoothed enough to compute the derivative (2.7) and then the covariance matrix via the sandwich formula (2.8). The IS-Lasso estimator is biased, it is possible to obtain an unbiased version of the estimator, as depicted in Figure 2.3, through the light blue triangle we derive a measure of the bias as follows $\text{bias}_\lambda = \hat{\beta}_0 - \hat{\beta}_\lambda = (X^T X)^{-1}[U(\hat{\beta}_\lambda) - \widetilde{U}(\hat{\beta}_\lambda)]$ where $\widetilde{U}(\hat{\beta}_\lambda) = 0$.

When $n > p$ we perfectly quantify the bias, for $n < p$ decomposition techniques can be applied in order to obtain $(X^T X)^{-1}$ and approximate the bias.

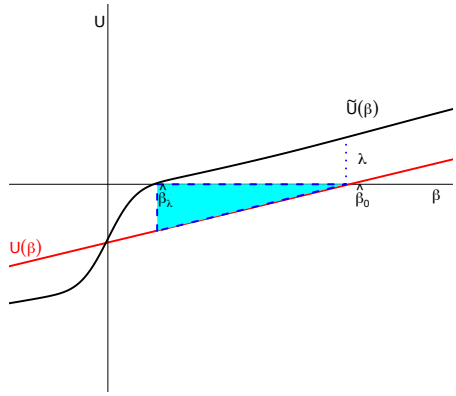

FIGURE 2.3: Bias of the estimator from the geometrical point of view. $\hat{\beta}_\lambda$ is the biased estimator, $\hat{\beta}_0$ is the value of the estimator when $\lambda = 0$, using the light blue triangle is possible to obtain the unbiased estimator.

2.2.1 The IS-Lasso Wald statistic

For the hypothesis of main interest $H_0 : \beta = 0$, given the IS-Lasso point estimate $\hat{\beta}$ and corresponding standard error $\mathrm{SE}(\hat{\beta})$ coming from (2.8), the Wald statistic under H_0 is defined as

$$W_0 = \frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})}. \tag{2.9}$$

In usual inferential problems, the p-value is obtained assuming that $W_0 \xrightarrow{d} \mathcal{N}(0, 1)$, and thus good performance of W_0 depends on how much reliable the Normal approximation is for W_0. For the LASSO estimate, the Wald statistic is useless as the sampling distribution has a positive probability mass at zero (Knight and Fu, 2000; Pötscher and Leeb, 2009), and no appropriate measure of standard errors are available as discussed in Introduction. However, as it is intuitive from the replacement of the in-

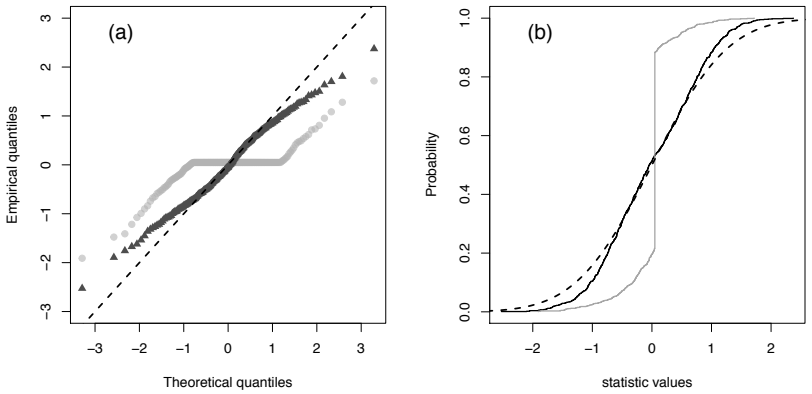

FIGURE 2.4: Contrasting plain LASSO (gray circles) and IS-Lasso (black triangles) Wald statistics. Panel (a) refers to QQ-plot and panel (b) to the cumulative distribution functions. The dashed lines correspond to the $\mathcal{N}(0,1)$ distribution. At each replicate the optimal lambda has been obtained via cross validation.

dicator function with the normal cumulative distribution function, the IS-Lasso estimator has no probability mass, but an obvious smooth peak around zero, as showed in Figure 2.4. Moreover the returned sandwich formula appropriately quantifies the estimator variability. More specifically, the studentized form (2.9) is adequately close to the standard Normal -noticeably with variance less than one- making it a valid tool for hypothesis testing.

However, the regression coefficient estimator is still biased for nonzero coefficient, that prevents the proposed Wald statistic to be used for interval estimation. In the next section we discuss an approach based on Score statistic to build confidence intervals.

2.2.2 The IS-Lasso Score statistic

The quantity $\widetilde{\boldsymbol{U}}(\boldsymbol{\beta})$ and $\widetilde{\boldsymbol{H}}(\boldsymbol{\beta})$ depend on $\boldsymbol{\beta}$, for simplicity from now we indicate $\widetilde{\boldsymbol{U}}(\boldsymbol{\beta}) = \widetilde{\boldsymbol{U}}$ and $\widetilde{\boldsymbol{H}}(\boldsymbol{\beta}) = \widetilde{\boldsymbol{H}}$. We partition the regression vector of coefficients $\boldsymbol{\beta}$ in $\boldsymbol{\beta}_1 \in \mathbb{R}_1^p$ and $\boldsymbol{\beta}_2 \in \mathbb{R}_2^p$ as the interest and nuisance parameters, respectively and indicate with $\widetilde{\boldsymbol{U}}_j$, $\widetilde{\boldsymbol{H}}_{jk}$ and $\boldsymbol{\mathcal{I}_{jk}}$ $(j,k = 1,2)$ the corresponding blocks of the Score vector, and of the Hessian and Information matrices. It is well know that score inference on β_1 relies on the profiled score, obtained through Taylor expansion

$$\widetilde{\boldsymbol{U}}_{1|2} = \widetilde{\boldsymbol{U}}_1 - \widetilde{\boldsymbol{H}}_{12}\widetilde{\boldsymbol{H}}_{22}^{-1}\widetilde{\boldsymbol{U}}_2, \tag{2.10}$$

with:

$$\widetilde{\boldsymbol{U}} = \begin{bmatrix} \widetilde{\boldsymbol{U}}_1 \\ \widetilde{\boldsymbol{U}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_1}^T(\boldsymbol{y} - \boldsymbol{X_1}\boldsymbol{\beta_1}) + \lambda\overline{\mathcal{P}}_1(\boldsymbol{\beta}_1, \boldsymbol{v}) \\ \boldsymbol{X_2}^T(\boldsymbol{y} - \boldsymbol{X_2}\boldsymbol{\beta_2}) + \lambda\overline{\mathcal{P}}_2(\boldsymbol{\beta}_2, \boldsymbol{v}) \end{bmatrix}$$

a standardize form of $\widetilde{\boldsymbol{U}}_{1|2}$ can be used to make inference on β_1 (Boos, 1992; Hu and Kalbfleisch, 2000). Equation (2.10) can be expressed in matrix notation via

$$\widetilde{\boldsymbol{U}}_{1|2} = \boldsymbol{A}\widetilde{\boldsymbol{U}} = [\, I, -\boldsymbol{b}][\widetilde{\boldsymbol{U}}_1^T, \widetilde{\boldsymbol{U}}_2^T]^T \tag{2.11}$$

where $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{b} = \widetilde{\boldsymbol{H}}_{12}\widetilde{\boldsymbol{H}}_{22}^{-1}$. Unlike the usual inferential contexts where the regularity conditions are met, in our case, $\mathbb{E}[\widetilde{\boldsymbol{U}}] \neq 0$ since $\mathbb{E}[\widetilde{\boldsymbol{U}_1}] = \lambda\overline{\mathcal{P}}_1$ and $\mathbb{E}[\widetilde{\boldsymbol{U}_2}] = \lambda\overline{\mathcal{P}}_2$ and the expected value of (2.10) can be written as:

$$\mathbb{E}[\widetilde{\boldsymbol{U}}_{1|2}] = \mathbb{E}[\widetilde{\boldsymbol{U}}_1] + \boldsymbol{b}(\widetilde{\boldsymbol{U}}_2 - \mathbb{E}[\widetilde{\boldsymbol{U}}_2]) = \lambda\overline{\mathcal{P}}_1 - \boldsymbol{b}\lambda\overline{\mathcal{P}}_2 \tag{2.12}$$

Starting from equation (2.11), the variance is easily obtained as:

$$\mathbb{V}(\widetilde{\boldsymbol{U}}_{1|2}) = \boldsymbol{A}\mathbb{V}(\widetilde{\boldsymbol{U}})\boldsymbol{A}^T = A\boldsymbol{\mathcal{I}}A^T = \sigma^2 A(X^T X)A^T \tag{2.13}$$

Consequently, the studentized Score statistic has to centred in order to be used for inference and it takes the form:

$$\widetilde{\boldsymbol{S}}_{1|2} = [\widetilde{\boldsymbol{U}}_{1|2} - \mathbb{E}[\widetilde{\boldsymbol{U}}_{1|2}]]^T\mathbb{V}(\widetilde{\boldsymbol{U}}_{1|2})^{-1}[\widetilde{\boldsymbol{U}}_{1|2} - \mathbb{E}[\widetilde{\boldsymbol{U}}_{1|2}]] \xrightarrow{d} \chi^2_{p_1}, \tag{2.14}$$

where p_1 is the dimension of the interest parameter $\boldsymbol{\beta}_1$. The proposed studentized Score statistic $\widetilde{\boldsymbol{S}}_{1|2}$ can be employed both for hypothesis testing and interval estimation. In hypothesis testing problem, we refer to the usual aforementioned hypotheses $H_0 : \beta_j = 0$ versus $H_1 : \beta_j \neq 0$. For interval estimation a $(1-\alpha)$ confidence interval for β_1 is given by:

$$\mathrm{CI}_{1-\alpha} = \{\bar{\beta}_1 : \widetilde{\boldsymbol{S}}_{1|2}(\bar{\beta}_1) \leq \chi^2_{p_1;1-\alpha}\}.$$

From a practical point of view, once profiled score is computed CIs are obtained through inversion. To illustrate, Figure 2.5 portrays an example of profile score for two coefficients in a toy dataset.

2.2.3 Some extensions

The extension to non-normal responses is very easy, the strategy is to replace the Newton-Raphson step with a IWLS step wherein the continuous response is the working response and proper weights depending

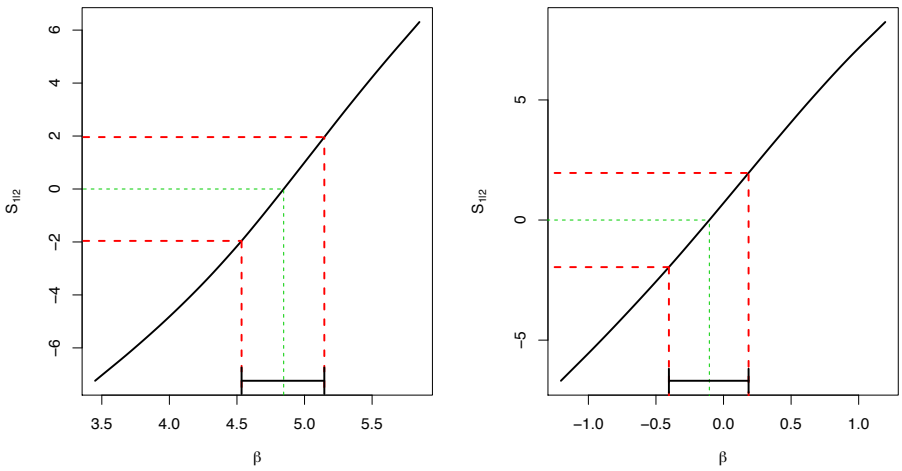

FIGURE 2.5: Illustrating the profile score with corresponding point estimate and 95% confidence intervals in a simulated dataset. The left and right panels refer to a non-zero and zero coefficient respectively. In the right panel the functions have been shifted to guarantee $S_{1|2}(\hat{\beta}) = 0$ (thus the dashed horizontal lines do not correspond to quantiles $z_{.025}$ and $z_{.975}$).

on family and link function have to be accounted for. In addition, the IS-Lasso allows to define a pseudo hat matrix at convergence, namely $\boldsymbol{A} = \boldsymbol{X}(\widetilde{\boldsymbol{H}})^{-1}\boldsymbol{X}^T$. The $j = 1, \ldots, p$ leading elements of $\boldsymbol{A}$, h_j say, could be used to quantify the amount of penalization of each estimate such that $h_j = 1$ for unpenalized coefficient, $h_j \lesssim 1$ for weakly penalized coefficients, and $h_j \approx 0$ for strongly penalized (almost null) coefficients. The coefficient-specific h_j could be also used to set weights in the adaptive LASSO penalty introduced in section (1.1.4). Since to use the estimates from a preliminary unpenalized fit is not always available, an alternative within the IS-Lasso would be to use $w_j = 1/h_j$ which are always computable even when $n < p$.

Chapter 3

Simulation studies

We present some simulations carried out to assess the finite sample behavior of our proposed framework.

3.1 Point estimation

In this section, we investigate the finite sample behavior of our lasso-type estimator in terms of bias and variability. We generate 500 replicates from a linear regression model $y \sim \mathcal{N}(X\beta, I_n)$ for different scenarios: two sample sizes $n = 100$ or $n = 200$ and number of parameters p with ratio $p/n = (0.5, 0.8, 1.2, 2)$ for each n; the fixed p covariates come from a multi-normal distribution with identity covariance matrix, and the 20 coefficients different from zero are a sequence from -1.5 to 1.5 by 0.16. Table 3.1 reports summary of sampling distributions for 10 coefficients in comparison with lasso, IS-lasso and lasso report estimates very close. In addition, IS-lasso estimators are biased if the corresponding coefficient is nonzero and SE always tend to slightly underestimate uncertainty for non-zero coefficients.

Table 3.2 reports the means and standard deviations of the sampling distributions at different values of c from a simulation study. We generate 500 replicates from a linear regression model $y \sim \mathcal{N}(X\beta, I_n)$ with $n = 100$ and $p = 30, 150$ and 5 non zero coefficients. Estimates and standard deviations are very similar both for $p = 30$ and $p = 150$, the choice of c, therefore, does not affect the estimates and its variability.

TABLE 3.1: Mean (M) and standard deviation (SD) of the sampling distributions in the simulation study of IS-Lasso. SE is the average of the standard errors. Lasso is the average of the lasso estimates. The tuning parameter is chosen by cross validation.

	TRUE VALUE									
	-1.50	-1.34	-1.18	-1.03	-0.87	0	0	0	0	0
	$n = 100, p = 50$									
Lasso	-1.434	-1.275	-1.121	-0.955	-0.834	-0.008	-0.025	-0.008	-0.014	0.001
M	-1.431	-1.268	-1.114	-0.952	-0.828	-0.014	-0.032	-0.009	-0.018	0.003
SD	0.111	0.133	0.111	0.130	0.123	0.066	0.067	0.069	0.064	0.068
SE	0.114	0.116	0.116	0.122	0.119	0.069	0.080	0.080	0.075	0.089
	$n = 100, p = 80$									
Lasso	-1.432	-1.284	-1.162	-0.867	-0.752	-0.007	-0.004	0.004	0.001	0.023
M	-1.426	-1.278	-1.157	-0.841	-0.749	-0.012	-0.005	0.002	0.004	0.027
SD	0.123	0.115	0.116	0.129	0.135	0.039	0.043	0.044	0.040	0.046
SE	0.118	0.119	0.127	0.125	0.122	0.057	0.058	0.060	0.056	0.064
	$n = 100, p = 120$									
Lasso	-1.359	-1.246	-1.062	-0.902	-0.746	0.010	-0.005	-0.001	0.012	-0.037
M	-1.308	-1.222	-1.047	-0.872	-0.730	0.015	-0.010	-0.002	0.018	-0.055
SD	0.122	0.126	0.137	0.132	0.122	0.028	0.023	0.030	0.030	0.046
SE	0.114	0.118	0.120	0.117	0.111	0.044	0.036	0.042	0.057	0.069
	$n = 100, p = 200$									
Lasso	-1.137	-1.214	-1.020	-0.965	-0.908	0.003	0.007	-0.018	0.015	-0.003
M	-1.033	-1.180	-0.966	-0.940	-0.928	0.005	0.015	-0.026	0.028	-0.002
SD	0.131	0.125	0.131	0.117	0.115	0.020	0.023	0.024	0.028	0.016
SE	0.132	0.121	0.119	0.123	0.127	0.034	0.033	0.042	0.041	0.025
	$n = 200, p = 100$									
Lasso	-1.394	-1.287	-1.070	-0.971	-0.782	-0.009	0.006	-0.004	0.004	-0.008
M	-1.390	-1.282	-1.069	-0.960	-0.775	-0.016	0.012	-0.007	0.009	-0.011
SD	0.083	0.071	0.078	0.080	0.084	0.027	0.028	0.028	0.028	0.026
SE	0.076	0.076	0.078	0.078	0.078	0.034	0.038	0.038	0.037	0.035
	$n = 200, p = 160$									
Lasso	-1.469	-1.264	-1.079	-0.928	-0.826	-0.013	0.003	-0.001	-0.005	-0.001
M	-1.468	-1.251	-1.065	-0.916	-0.823	-0.022	0.003	-0.006	-0.007	-0.002
SD	0.070	0.082	0.088	0.084	0.082	0.023	0.013	0.016	0.015	0.018
SE	0.076	0.077	0.078	0.077	0.077	0.036	0.020	0.025	0.025	0.033
	$n = 200, p = 240$									
Lasso	-1.406	-1.217	-1.094	-0.936	-0.755	0.002	-0.001	0.004	0.000	0.001
M	-1.386	-1.203	-1.078	-0.929	-0.739	0.004	0.002	0.006	0.001	0.002
SD	0.077	0.081	0.082	0.080	0.080	0.013	0.009	0.010	0.011	0.010
SE	0.077	0.076	0.078	0.077	0.081	0.027	0.022	0.024	0.026	0.021
	$n = 200, p = 400$									
Lasso	-1.378	-1.250	-1.071	-0.904	-0.741	0.002	0.000	-0.003	-0.003	-0.006
M	-1.353	-1.232	-1.049	-0.882	-0.719	0.003	0.002	-0.005	-0.006	-0.011
SD	0.080	0.079	0.080	0.080	0.082	0.006	0.005	0.006	0.006	0.009
SE	0.075	0.076	0.078	0.077	0.078	0.015	0.014	0.017	0.016	0.021

TABLE 3.2: Summary of sampling distributions: Mean (M) and the standard deviation (SD) for different value of c. The tuning parameter is chosen by cross validation. Only 4 non zero and 4 null coefficients are reported.

			TRUE VALUE							
			3	3	3	3	0	0	0	0
$p = 30$	M	$c = 0.12.854$	2.905	2.896	2.913	-0.005	-0.001	-0.013		0.007
		$c = 0.9$	2.845	2.904	2.892	2.908	-0.007	-0.003	-0.017	0.008
		$c = \bar{c}$	.849	2.905	2.894	2.911	-0.007	-0.002	-0.015	0.008
	SD	$c = 0.1$	0.123	0.110	0.111	0.108	0.048	0.046	0.055	0.049
		$c = 0.9$	0.128	0.112	0.111	0.110	0.055	0.052	0.059	0.054
		$c = \bar{c}\,0.126$	0.110	0.109	0.109	0.053	0.051	0.058	0.052	
$p = 150$	M	$c = 0.1$	2.759	2.765	2.834	2.814	-0.004	0.003	0.001	0.002
		$c = 0.9$	2.735	2.738	2.821	2.797	-0.007	0.003	0.001	0.002
		$c = \bar{c}$	2.736	2.737	2.822	2.797	-0.007	0.003	0.001	0.001
	SD	$c = 0.1$	0.113	0.111	0.105	0.113	0.022	0.020	0.020	0.020
		$c = 0.9$	0.115	0.112	0.105	0.111	0.027	0.021	0.023	0.022
		$c = \bar{c}$	0.114	0.113	0.106	0.113	0.025	0.020	0.022	0.022

3.2 Hypothesis testing

In the second batch of simulations we assessed performance of the IS-lasso Wald statistic in actual hypothesis testing problems maintaining the same scenarios of section 3.1. The hypothesis of interest is $H_0 : \beta_j = 0$ for any coefficient β_j in the regression equation. We compare the Wald statistic (2.9) with two competitors covTest and postSel discussed previously and implemented in the R packages covTest (Lockhart et al., 2013) and selectiveInference (Tibshirani et al., 2017). covTest does not require a fixed λ value as it returns results across the path, and it uses a standard Exponential distribution to get p-values. postSel carries out selective inference, namely it returns p-value using a reference truncated normal distribution only for the covariates in the selected active set. When the covariate is not included in the model no inference measure is given; however the traditional null hypothesis, $H_0 : \beta_j = 0$, would have not been rejected, and thus we fix p-value=1. That appears to be a sound rule, when in practice one is interested in returning a p-value for each candidate covariate. At each replicate, the tuning parameter λ was optimized through 5-fold cross validation for postSel and IS-lasso.

3.2.1 Power function under violation of the theoretical conditions

Comparisons in terms of power functions are reported in Figure 3.1 for $n = 100, 200$ for different number p of covariates. The empirical rejection rates under the null hypothesis are averaged among all zero coefficients. When the true coefficient is non-null, some differences are noteworthy: the Wald statistic based on IS-lasso appears to provide the most powerful test in all scenarios, with a quite negligible difference when $n = 100$, $p = 200$ and true $\beta = \pm 0.08$. Overall covTest appears to provide somewhat wiggly patterns at $n = 100$, especially when the number of covariates is large. All tests exhibit correct sizes, with values lower than 0.05, to better understand the differences Figure 3.2 zoom the Figure before in the interval from -0.08 to 0.08. CovTest is the most conservative test, Post-sel is the closest to the nominal level and IS-wald represents a trade-off between the two test.

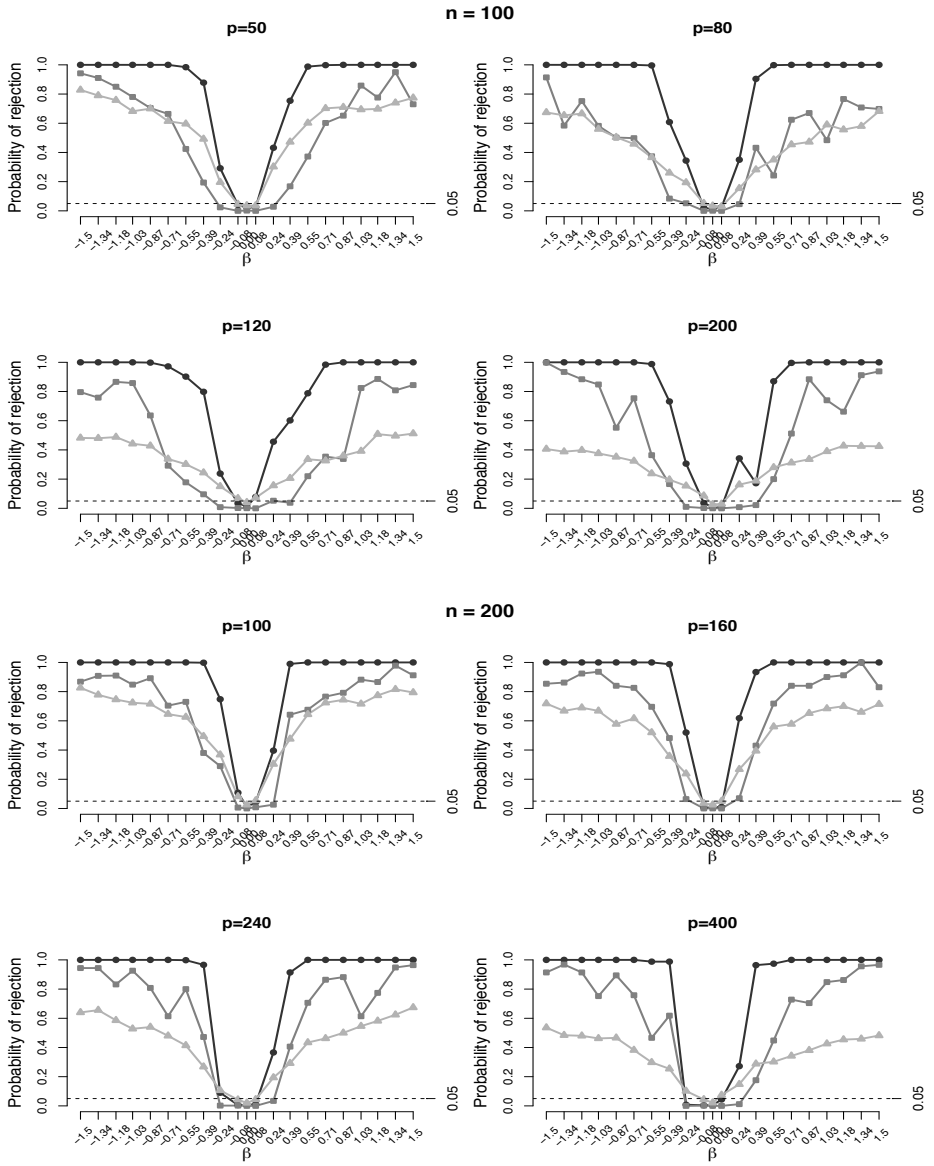

FIGURE 3.1: Power functions (at 5% level) of different tests, $n = 100$ (upper panels), $n = 200$ (lower panels) and number p of covariates on the top: IS-Wald (black circles), covTest (dark gray squares) and postSel (light gray triangles). At each replicate the optimal lambda employed by IS and postSel has been obtained via cross validation.

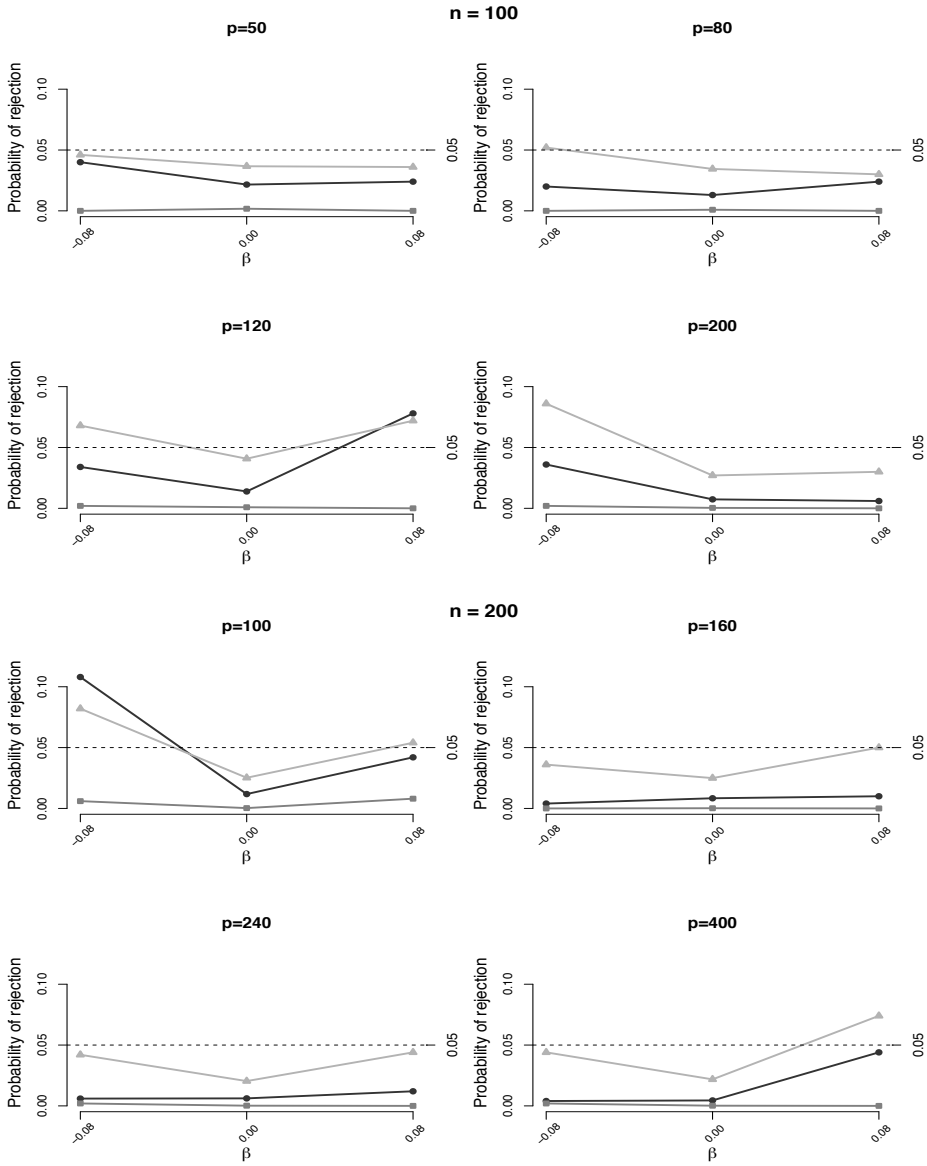

FIGURE 3.2: Size of different tests, $n = 100$ (upper panels), $n = 200$ (lower panels) and number p of covariates on the top: IS-Wald (black circles), covTest (dark gray squares) and postSel (light gray triangles). At each replicate the optimal lambda employed by IS and postSel has been obtained via cross validation.

3.2.2 Power function with correlated covariates

Generating $X \sim \mathcal{N}_p(0, \Sigma)$ with covariance matrix following Toeplitz structure $\Sigma_{j,k} = 0.5^{|j-k|}$ (Figure 3.3), again IS-Wald test shows the highest power and adequate size, CovTest presents a wiggle trend and postSel has the lowest power, in addition, the latter two tests gain in power with respect the previous scenario.

3.2.3 Power function conditionally to the selected model

Inferences by postSel are conditional to the the selected model, namely postSel returns p-values only for the selected covariates. To make (unconditional) comparisons with IS-lasso and covTest, we have set p-value=1 for the unselected variables, which appears pretty sound from a practical viewpoint. However for the sake of completeness, we have also compared the three tests conditionally, namely computing the rejection rates for the hypothesis $H_0 : \beta_j = 0$ only when the corresponding variable X_j entered the selected model. Comparative performances of the three tests was unchanged with respect to the unconditional context.

3.2.4 Power function under theoretical conditions

In the third batch of simulations we assessed performance of the IS-lasso Wald statistic under theoretical conditions namely (Section 1.1.1). We generate 500 replicates from a linear regression model $y \sim \mathcal{N}(X\beta, I_n)$ for different scenarios: two sample sizes $n = 100$ or $n = 200$ and number of parameters p with ratio $p/n = (0.5, 0.8, 1.2, 2)$ for each n and the number of active set $k \leq n/(2 \log p) = 8$ with magnitude at least of $|4|$; the fixed p covariates come from a multi-normal distribution with identity covariance matrix, and the 8 coefficients different from zero take values in a sequence from -5 to -4 and from 4 to 5 by 0.33.

All tests exhibit very good performance (Figure 3.4) with a power $\geq 95\%$ and a size close to the nominal level.

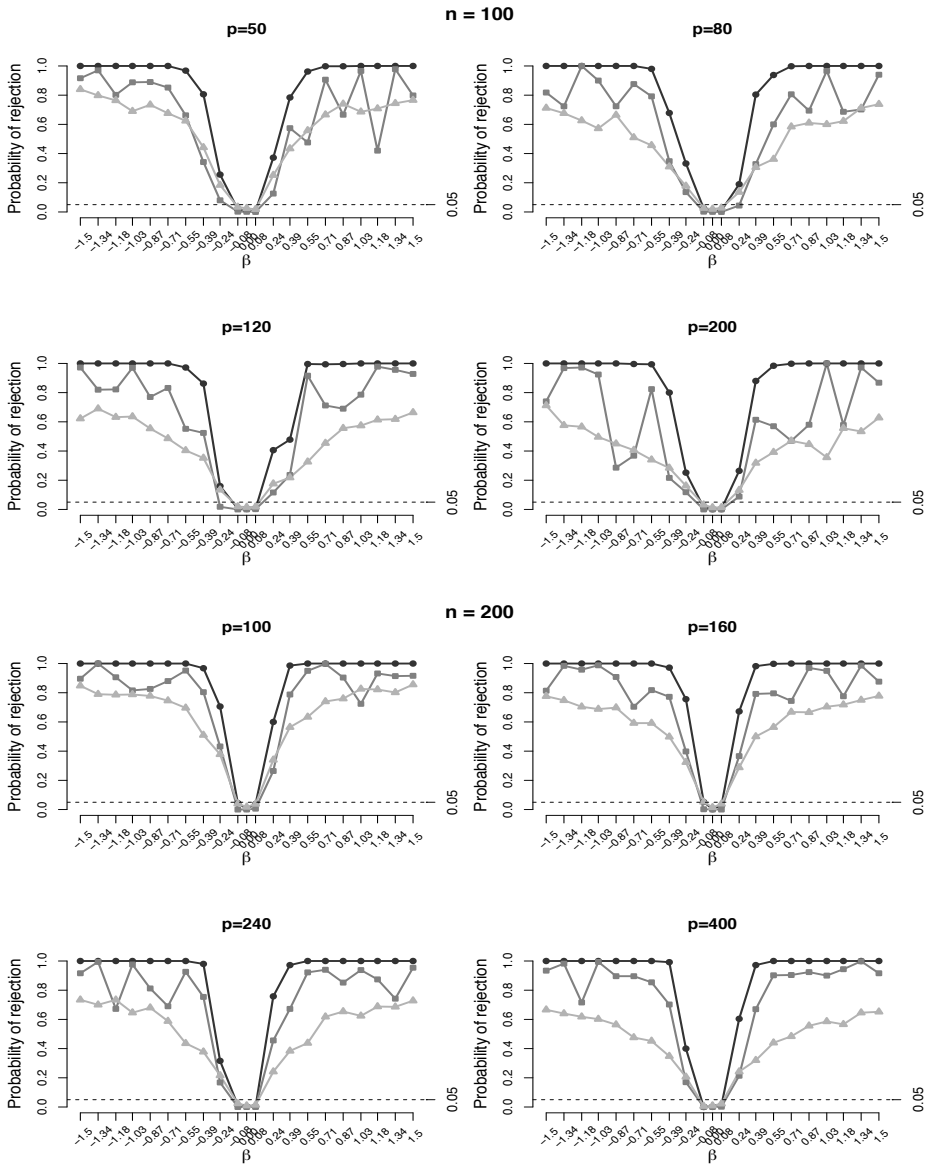

FIGURE 3.3: Power functions (at 5% level) of different tests, $n = 100$ (upper panels), $n = 200$ (lower panels) and number p of covariates on the top: IS-Wald (black circles), covTest (dark gray squares) and postSel (light gray triangles). At each replicate the optimal lambda employed by IS and postSel has been obtained via cross validation, and $X \sim \mathcal{N}_p(0, \Sigma)$ with covariance matrix following Toeplitz structure $\Sigma_{j,k} = 0.5^{|j-k|}$.

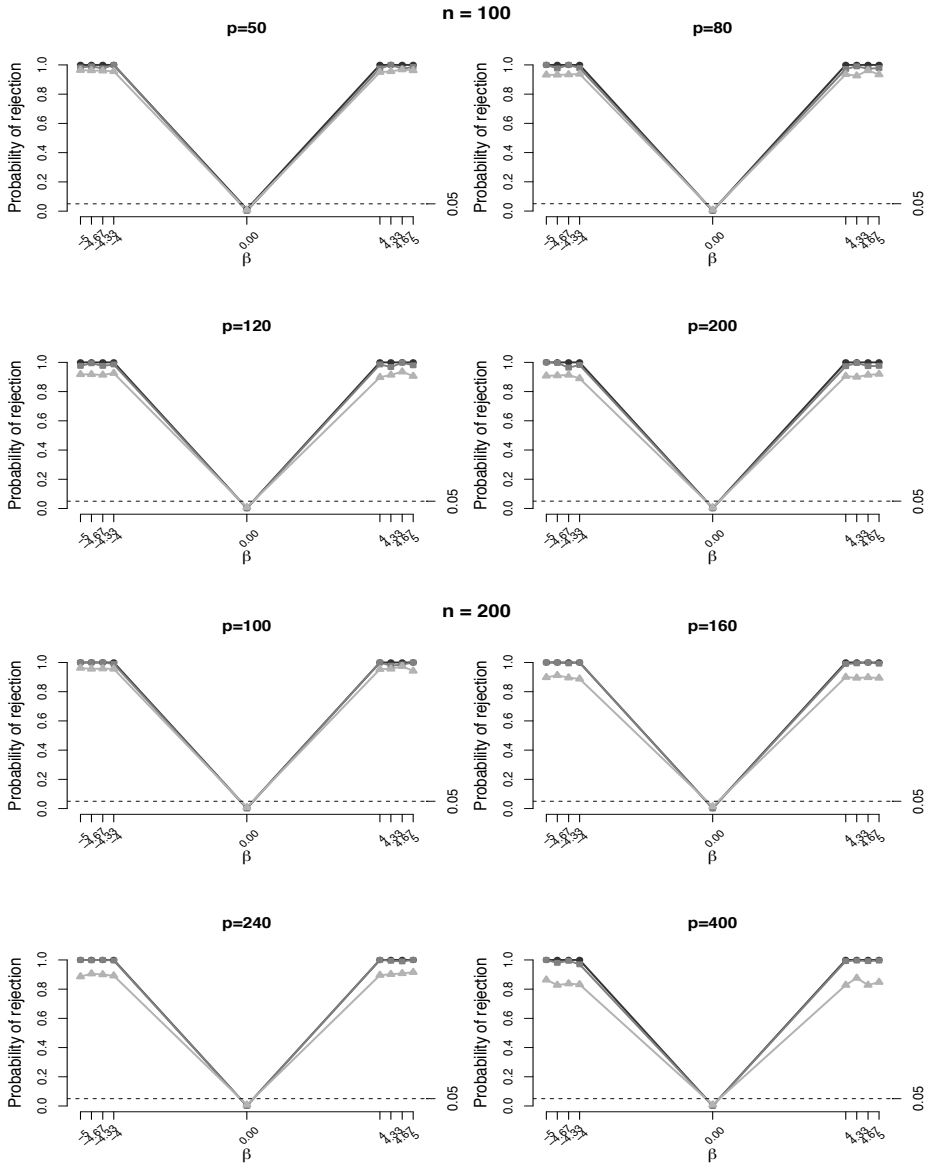

FIGURE 3.4: Power functions (at 5% level) of different tests under theoretical conditions, $n = 100$ (upper panels), $n = 200$ (lower panels) and number p of covariates on the top: IS-Wald (black circles), covTest (dark gray squares) and postSel (light gray triangles). At each replicate the optimal lambda employed by IS and postSel has been obtained via cross validation.

3.2.5 Power function of Score statistic

As discussed in Section 2.2.2 the Score statistic can be applied for hypothesis testing problem, Figure 3.5 shows the power function of two simulation studies adding Score test. IS-Score is overlapped to IS-Wald in all scenarios. Focus on size, as depicted in Figure 3.6 , we observe that IS-Score is closer to nominal level than IS-Wald, and the two test come closer when p increases.

3.2.6 Conclusions

Simulation studies on hypothesis testing show very good performance of both IS-Wald and IS-Score test. The size is adequate and under the nominal value for all tests. When theoretical conditions are fulfilled covTest and postSel gain and achieve a good power level, unfortunately in real case theoretical conditions hardly are satisfied.

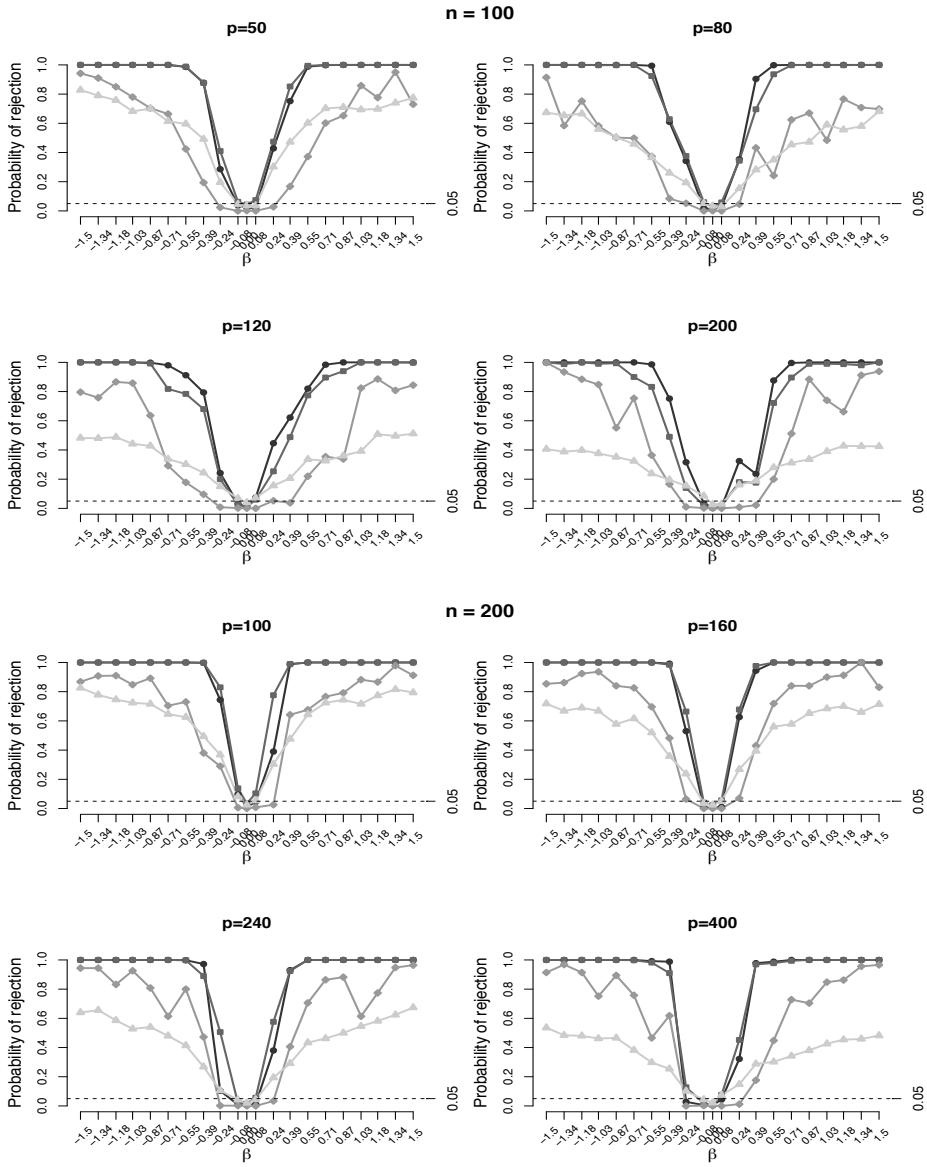

FIGURE 3.5: Power functions (at 5% level) of different tests under theoretical conditions, $n = 100$ (upper panels), $n = 200$ (lower panels) and number p of covariates on the top: IS-Wald (black circles), IS-Score (medium gray square), covTest (dark gray squares) and postSel (light gray triangles). At each replicate the optimal lambda employed by IS and postSel has been obtained via cross validation.

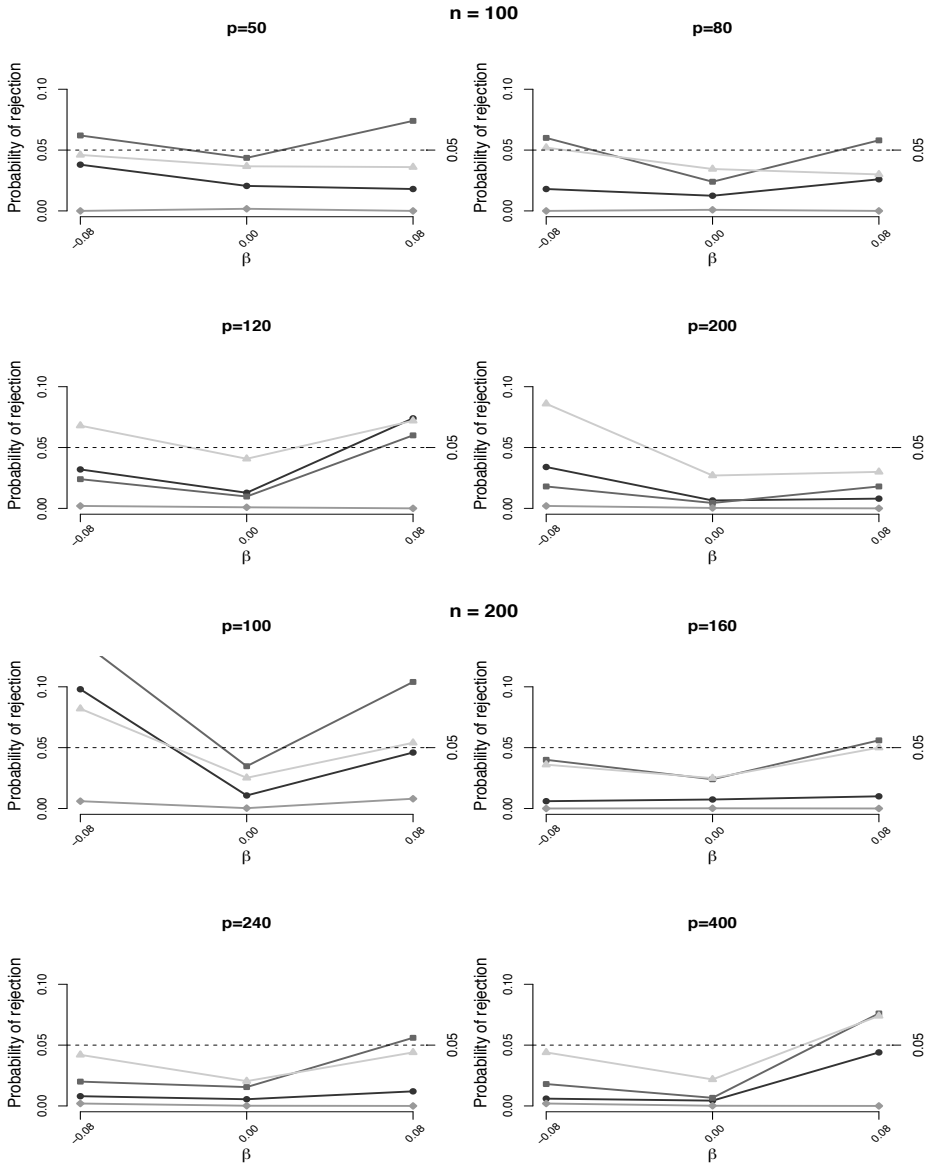

FIGURE 3.6: Size of different tests under theoretical conditions, $n = 100$ (upper panels), $n = 200$ (lower panels) and number p of covariates on the top: IS-Wald (black circles), IS-Score (medium gray square), covTest (dark gray squares) and postSel (light gray triangles). At each replicate the optimal lambda employed by IS and postSel has been obtained via cross validation.

3.3 Interval estimation

We compare confidence intervals of the IS-Score statistic (2.14) with two competitors Hdi and postSel discussed previously and implemented in the R packages hdi (Dezeure et al., 2015) and selectiveInference (Tibshirani et al., 2017). Hdi for the choice of tuning parameter uses an internal procedure denoted by Zhang and Zhang (2014). postSel returns confidence intervals only for the covariates in the selected active set for a fixed λ. Again, when the covariate is not included in the model no CI is given; however we fix CI=0 which means not significant interval estimation. At each replicate, the tuning parameter λ was optimized through 5-fold cross validation for postSel and IS-lasso.

We compare CIs in terms of coverage and median width of the Score, postSel and Hdi. The reason of the median width , and not average for example, is linked to postSel, because most of the time it returns infinity upper or/and lower bound.

3.3.1 Confidence intervals under theoretical conditions

Table 3.3 shows results based on 300 runs, respectively in low and high dimensional settings, wherein the tuning parameter λ has been selected via 5-fold cross validation at each replicate. The Score test shows good performance with CIs coverage $\geq 95\%$ in all scenarios, postSel is very close to 95%, while Hdi has the lowest performance specially for non-zero coefficients. Figure 3.7 depicts the median upper and lower bound of the CIs of the first 20 coefficients of simulation studies reported in Table 3.3, Hdi (green line) seems to have smaller CIs than Score, conversely postSel has larger CIs than Score and Hdi, indeed from Table 3.3 we know that even if Hdi has the lowest width its coverage never achieves the nominal level. Moreover, since postSel makes inference conditionally to the selected model and most of the time zero coefficients are not included in, CIs show a zero width.

3.3.2 Confidence intervals under violation of the theoretical conditions

We also consider scenarios in which the $\boldsymbol{\beta}_{min}$ condition is not met, namely with same combinations of n and p and $\boldsymbol{\beta} = (-1, -0.5, 0.65, 0.88, 1, 0, \ldots, 0)^T$. Table 3.4 reports results of the simulations in which the $\boldsymbol{\beta}_{min}$ condition is violate. Score shows again the best coverage, postSel achieves a good coverage at expense of a large median width and Hdi

TABLE 3.3: Coverage levels and median widths (in italic) of 95% CIs from Score, Hdi and postSel for 10 selected parameters. Results are based on 300 replicates in low and high dimensional setting, σ is equal to 1 and the optimal λ has been obtained via cross validation.

		TRUE VALUE									
(n, p)		3.0	-3.1	4.0	3.5	-5.0	0.0	0.0	0.0	0.0	0.0
	Score	0.96	0.98	0.97	0.96	0.98	0.98	0.96	0.96	0.97	0.97
		0.68	*0.75*	*0.75*	*0.78*	*0.69*	*0.85*	*0.80*	*0.74*	*0.81*	*0.71*
(50, 40)	Hdi	0.95	0.87	0.91	0.86	0.90	0.90	0.87	0.88	0.96	0.91
		0.63	*0.57*	*0.53*	*0.59*	*0.53*	*0.55*	*0.54*	*0.56*	*0.53*	*0.56*
	postSel	0.94	0.94	0.94	0.93	0.95	0.99	0.99	0.99	1.00	0.98
		0.92	*1.14*	*1.14*	*1.20*	*0.99*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	Score	0.98	0.98	0.99	0.98	0.99	0.99	0.98	0.98	0.98	0.98
		0.79	*0.74*	*0.83*	*0.87*	*1.00*	*0.95*	*0.86*	*0.78*	*1.02*	*0.83*
(50, 60)	Hdi	0.82	0.87	0.91	0.88	0.70	0.93	0.95	0.95	0.88	0.95
		0.54	*0.63*	*0.56*	*0.58*	*0.57*	*0.51*	*0.56*	*0.56*	*0.54*	*0.56*
	postSel	0.92	0.93	0.93	0.94	0.93	0.99	0.98	0.99	1.00	0.99
		1.37	*1.30*	*1.51*	*1.75*	*1.74*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	Score	0.98	0.96	0.98	0.99	0.97	0.98	0.99	0.97	0.97	0.98
		0.57	*0.55*	*0.60*	*0.62*	*0.56*	*0.53*	*0.56*	*0.57*	*0.58*	*0.54*
(100, 80)	Hdi	0.82	0.90	0.86	0.74	0.95	0.85	0.96	0.96	0.73	0.89
		0.42	*0.42*	*0.43*	*0.42*	*0.43*	*0.41*	*0.41*	*0.41*	*0.42*	*0.40*
	postSel	0.95	0.94	0.95	0.95	0.96	0.98	1.00	0.99	0.98	0.99
		1.28	*1.18*	*1.37*	*1.53*	*1.18*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	Score	0.99	1.00	0.98	0.99	0.99	0.97	0.99	1.00	0.98	0.97
		0.57	*0.67*	*0.61*	*0.57*	*0.51*	*0.56*	*0.60*	*0.70*	*0.56*	*0.66*
(100, 120)	Hdi	0.97	0.96	0.93	0.91	0.97	0.97	0.97	0.95	0.91	0.97
		0.46	*0.44*	*0.42*	*0.43*	*0.42*	*0.44*	*0.42*	*0.41*	*0.42*	*0.44*
	postSel	0.91	0.94	0.93	0.94	0.93	0.99	0.99	0.99	0.99	1.00
		1.45	*1.67*	*1.58*	*1.47*	*1.13*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

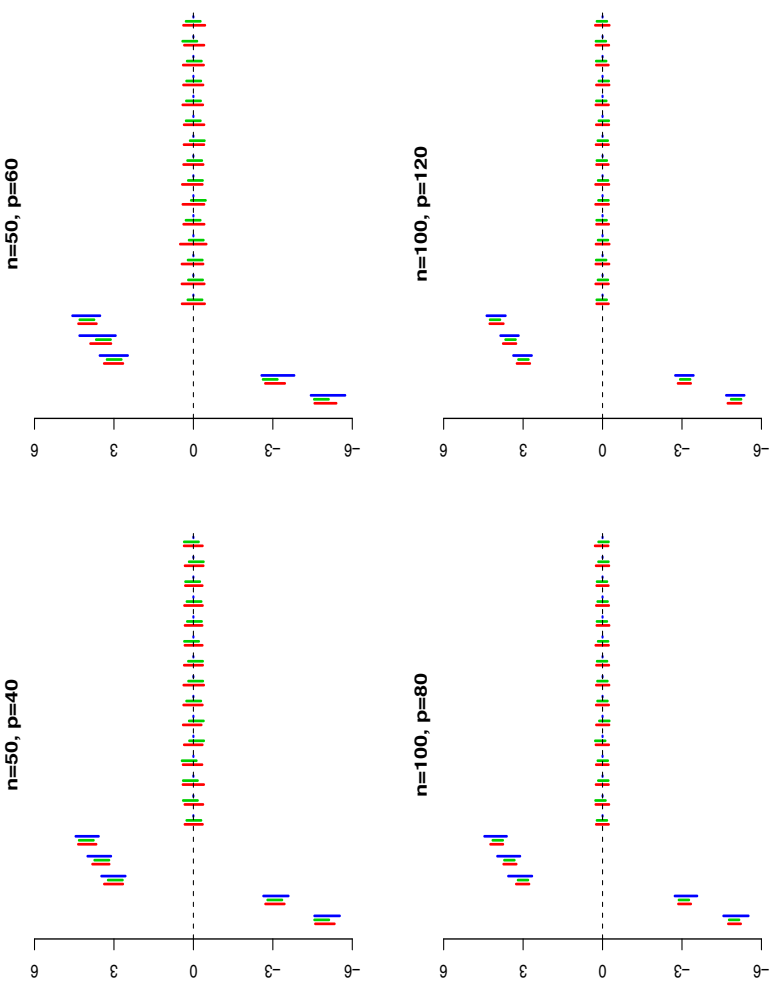

FIGURE 3.7: Interval estimation performance of different approaches as reported in Table 3.3. Each bar represents the medians (across the 300 replicates) of the 95% CI limits. Medians rather than means have been computed because most of the times postSel returned infinity values. Sample size n and number p of covariates on the top: Score (red line), Hdi (green line) and postSel (blue line).

has the lowest coverage. Figure 3.8 graphically reports the median width for the first 20 coefficients of the considered method, again, postSel has the largest CIs, Hdi the shortest and Score is near to Hdi when n increases.

TABLE 3.4: Coverage levels and median widths of 95% CIs from Score, Hdi and postSel for 10 selected parameters. Results are based on 300 replicates in low and high dimensional setting, σ is equal to 1 and the optimal λ has been obtained via cross validation.

		TRUE VALUE									
(n, p)		-1.0	-0.5	0.65	0.88	1.0	0.0	0.0	0.0	0.0	0.0
(50, 40)	Score	0.97	1.00	0.96	0.99	0.98	0.98	0.98	0.97	1.00	0.97
		0.80	*0.97*	*0.86*	*1.02*	*0.79*	*1.06*	*0.99*	*0.88*	*1.12*	*0.80*
	Hdi	0.89	0.91	0.91	0.90	0.95	0.90	0.92	0.92	0.93	0.92
		0.56	*0.58*	*0.56*	*0.56*	*0.63*	*0.53*	*0.61*	*0.56*	*0.59*	*0.61*
	postSel	0.95	0.95	0.93	0.95	0.95	1.00	0.99	0.98	0.99	0.99
		1.92	*2.94*	*2.56*	*2.64*	*2.05*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
(50, 60)	Score	0.99	0.98	1.00	1.00	0.99	1.00	0.99	0.99	0.99	0.99
		1.02	*0.85*	*1.09*	*1.11*	*1.43*	*1.23*	*1.03*	*0.95*	*1.38*	*0.97*
	Hdi	0.88	0.90	0.90	0.86	0.95	0.96	0.96	0.96	0.95	0.91
		0.58	*0.56*	*0.56*	*0.57*	*0.60*	*0.55*	*0.57*	*0.64*	*0.60*	*0.57*
	postSel	0.96	0.94	0.96	0.95	0.96	1.00	0.99	0.99	0.99	0.98
		2.73	*2.62*	*2.68*	*3.60*	*3.47*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
(100, 80)	Score	0.98	0.98	0.99	0.99	0.98	0.98	0.99	0.98	0.99	0.98
		0.57	*0.54*	*0.60*	*0.62*	*0.55*	*0.53*	*0.56*	*0.57*	*0.58*	*0.54*
	Hdi	0.89	0.92	0.94	0.89	0.93	0.96	0.90	0.96	0.93	0.96
		0.40	*0.43*	*0.47*	*0.43*	*0.41*	*0.40*	*0.40*	*0.42*	*0.40*	*0.40*
	postSel	0.94	0.94	0.96	0.94	0.93	0.98	0.99	0.98	0.99	0.99
		1.41	*1.46*	*1.69*	*1.87*	*1.46*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
(100, 120)	Score	0.99	1.00	0.99	0.99	0.99	0.98	0.99	1.00	1.00	0.98
		0.63	*0.78*	*0.67*	*0.63*	*0.54*	*0.61*	*0.65*	*0.78*	*0.62*	*0.73*
	Hdi	0.93	0.92	0.85	0.95	0.91	0.97	0.91	0.96	0.97	0.95
		0.42	*0.41*	*0.43*	*0.42*	*0.44*	*0.42*	*0.40*	*0.42*	*0.42*	*0.42*
	postSel	0.93	0.91	0.90	0.90	0.94	0.99	0.97	0.99	0.98	0.99
		2.32	*2.80*	*2.40*	*2.12*	*1.76*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*

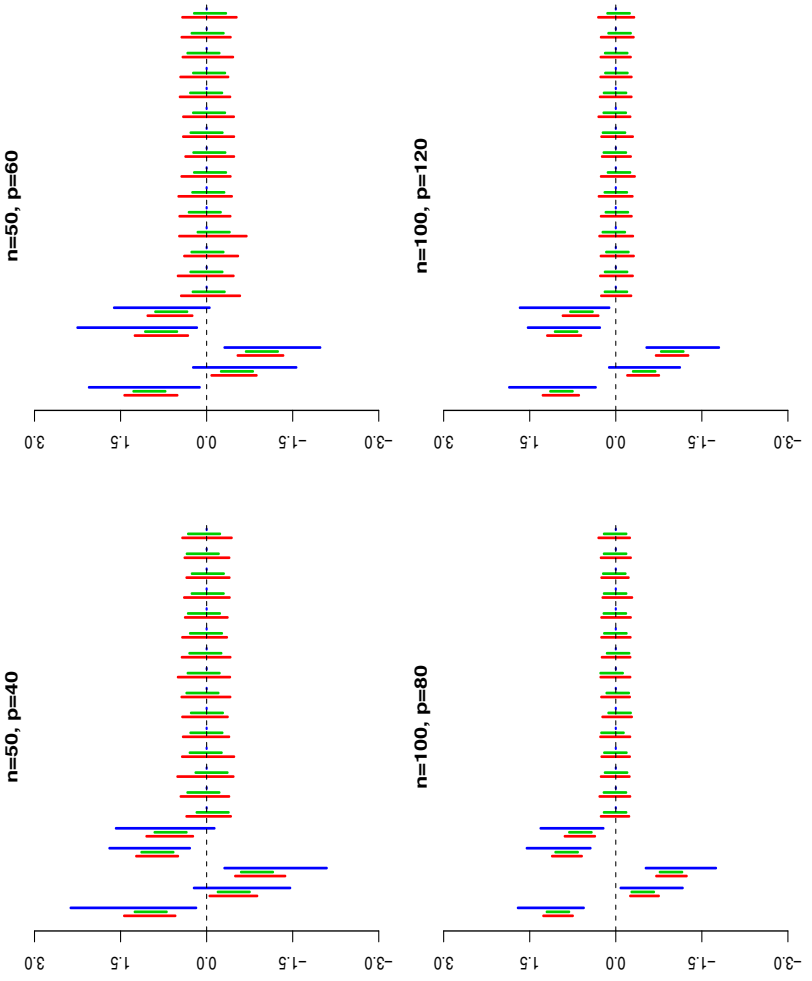

FIGURE 3.8: Interval estimation performance of different approaches as reported in Table 3.4. Each bar represents the medians (across the 300 replicates) of the 95% CI limits. Medians rather than means have been computed because most of the times postSel returned infinity values. Sample size n and number p of covariates on the top: Score (red line), Hdi (green line) and postSel (blue line).

3.3.3 Conclusions

Simulation studies show good performance of Score, which achieves in all scenarios a very high coverage, Hdi has not very well performance, postSel shows not bad performance under theoretical condition. All tests are not comparable in terms of width, however, Hdi has the lowest width but also the lowest coverage than other tests, postSel reports the largest width, Score seems to be a good trade-off between the tests.

Chapter 4

Modelling lung function in asthmatic children

4.1 Motivating data

Between September 2011 and 2017, $n = 529$ asthmatic children, aged 5–17 years, were recruited as a part of the 'Childhood Asthma and Environment Study' (CHASER, clinicaltrials.gov NCT02433275) an ongoing cross-sectional study at the outpatient clinic of research unit of *Pediatric Allergology & Pulmonology*. The parents or legal guardians were interviewed by means of a modified version of the SIDRIA (Italian Studies on Respiratory Disorders in Children and the Environment) questionnaire (Simoni et al., 2005; Migliore et al., 2009), including questions regarding socio-demographic characteristics, parental history of asthma, early and current outdoor and indoor environmental exposures, child's history of wheeze and presence of co-morbidities. Asthma severity level (Intermittent Asthma, IA; Mild Persistent Asthma, MPA; Moderate/Severe Persistent Asthma, MSPA) and asthma control status (Well Controlled, WC; Partially Controlled, PC; Uncontrolled Asthma, UA) were retrospectively performed according to GINA (Global Initiative for Asthma) (GINA, 2017). Pulmonary function tests were performed through a portable spirometer (Pony FX, Cosmed, Rome, Italy). FVC, FEV_1, and $FEF_{25-75\%}$ were measured according to ATS/ERS guidelines (Miller et al., 2005). The observed spirometric values were transformed in z-score and in percent of predicted value, according with the Global Lungs Initiative (GLI) (Quanjer et al., 2012). Allergic sensitization were defined upon a positive skin response after 15 min (i.e., a wheal ≥ 3 *mm* larger than the negative control test) (Bernstein et al., 2008) to any of the following allergens: indoor (dermatophagoides mix, dog and cat dander), outdoor (grass mix, parietaria judaica, cupressus, olive) and alternaria (ALK SQ extracts). Positive Skin Prick Tests (SPT+) were defined as at least one positive SPT. None of the

patients were on drug treatment at the time of the enrollment. All children were caucasic.

4.2 Exploratory analysis

Table 4.1 shows demographic characteristics of the 529 subjects. No difference is found between the asthmatic groups, except for pre-term born, GINA and maternal education. MSPA are more frequently pre-term born and uncontrolled than the other two groups, in addition MSPA mothers are less educated than intermittent and mild asthmatic group.

TABLE 4.1: Subject characteristics by asthma severity level (Intermittent Asthma, IA; Mild Persistent Asthma, MPA; Moderate/Severe Persistent Asthma, MSPA)

	IA	MPA	MSPA	p-value**
	n=229	n=212	n=88	
Gender				0.583
Female	91 (40%)	75 (35%)	31 (35%)	
Male	138 (60%)	137 (65%)	57 (65%)	
Age, *(years)*	8.97 (2.83)	8.51 (2.91)	9.02 (2.84)	0.174
Weight, *(kg)*	35.65 (14.64)	33.96 (14.43)	35.83 (16.13)	0.416
Height, *(cm)*	132.97 (16.75)	130.46 (16.75)	132.62 (17.21)	0.269
Birth weight, *(kg)*	3.23 (0.48)	3.24 (0.51)	3.09 (0.61)	0.057
Preterm born *(<37 years)*	13 (6%)	27 (13%)	14 (17%)	**0.006**
Caesarean section	124 (55%)	123 (59%)	56 (66%)	0.2448
Breast feeding ($>$ 3 *months*)	152 (70%)	140 (71%)	56 (68%)	0.922
Age at weaning *(years)*	5.18 (1.02)	5.21 (1.11)	5.36 (1.36)	0.464
GINA Uncontrolled, n(%)	49 (21%)	110 (52%)	60 (68%)	**<0.001**
Maternal education				**0.014**
$<$8 years	39 (17%)	59 (28%)	25 (29%)	
$\geq$8 years	186 (83%)	152 (72%)	62 (71%)	
Paternal education				0.742
$<$8 years	54 (24%)	56 (27%)	20 (24%)	
$\geq$8 years	170 (76%)	152 (73%)	65 (76%)	
Paternal history of asthma	43 (19%)	36 (17%)	25 (29%)	0.075
Maternal history of asthma	47 (21%)	37 (18%)	17 (20%)	0.689

Data are expressed as n (%) or mean (SD); **p-value comes from a X^2 test for categorical variables and form ANOVA for continuous variables.

Subjects are equally exposed to traffic, smoke, pet and mold (Table 4.2), however MSPA children live $<$ 500 meters to an industry more frequently than the others.

MSPA has more frequently food allergy than the other two groups and IA has less frequently history of wheezing than persistent (Table 4.3).

For all spirometric indices both in absolute values, Z-score and predicted values, MSPA shows lung functions significantly lower than IA and MPA (Table 4.4).

TABLE 4.2: Indoor and outdoor exposure by asthma severity level (Intermittent Asthma, IA; Mild Persistent Asthma, MPA; Moderate/Severe Persistent Asthma, MSPA)

	IA	MPA	MSPA	*p*-value**
	n=229	n=212	n=88	
Maternal smoke in pregnancy	23 (10%)	20 (10%)	13 (15%)	0.349
Paternal smoke in pregnancy	79 (35%)	67 (32%)	36 (42%)	0.275
Early Exposure*				
Maternal smoke early exposure	27 (12%)	19 (9%)	15 (18%)	0.114
Paternal smoke early exposure	76 (34%)	70 (33%)	33 (38%)	0.699
Early mold exposure	55 (25%)	72 (35%)	25 (29%)	0.056
Early dog exposure	16 (7%)	18 (9%)	9 (10%)	0.605
Early cat exposure	11 (5%)	5 (2%)	5 (6%)	0.277
Current Exposure				
Passive smoke exposure	66 (29%)	70 (33%)	33 (38%)	0.318
Current mold exposure	52 (23%)	47 (22%)	21 (24%)	0.953
Current dog exposure	33 (15%)	30 (14%)	18 (20%)	0.361
Current cat exposure	15 (7%)	10 (5%)	4 (5%)	0.634
Intense traffic level	190 (84%)	179 (84%)	70 (80%)	0.573
Proximity to landfill	4 (2%)	4 (2%)	4 (5%)	0.302
Proximity to industry	10 (4%)	11 (5%)	11 (12%)	**0.023**
Proximity to continuously vehicular traffic	192 (85%)	176 (83%)	74 (85%)	0.869
Proximity to high traffic road(<50m)	84 (37%)	67 (32%)	22 (25%)	0.109

*Early exposure means during the first year of life, data are expressed as n (%); **p-value comes from a X^2 test for categorical variables and form ANOVA for continuous variables.

TABLE 4.3: Co-morbidities distribution by asthma severity level (Intermittent Asthma, IA; Mild Persistent Asthma, MPA; Moderate/Severe Persistent Asthma, MSPA)

	IA	MPA	MSPA	p-value**
	n=229	n=212	n=88	
Eczema	54 (24%)	50 (24%)	30 (34%)	0.120
Rhinitis	142 (62%)	123 (58%)	56 (64%)	0.550
Conjunctivitis	56 (25%)	66 (31%)	26 (30%)	0.333
Oral syndrome	7 (3%)	7 (3%)	1 (1%)	0.567
Acute urticaria	47 (21%)	29 (14%)	13 (15%)	0.129
Angioedema	13 (6%)	8 (4%)	6 (7%)	0.476
Anaphylaxis	6 (3%)	7 (3%)	4 (5%)	0.690
Food allergy	20 (9%)	20 (9%)	19 (22%)	**0.003**
Otitis	38 (17%)	34 (16%)	17 (19%)	0.784
Laryngospasm	73 (32%)	75 (36%)	27 (31%)	0.666
Upper respiratory infection	75 (33%)	65 (31%)	28 (32%)	0.867
Sinusitis	29 (13%)	32 (15%)	11 (12%)	0.716
Snoring	102 (45%)	93 (44%)	42 (48%)	0.823
Wheezing	157 (69%)	171 (81%)	72 (82%)	**0.004**

Data are expressed as n (%); **p-value comes from a X^2 test for categorical variables and form ANOVA for continuous variables.

TABLE 4.4: Spirometric indices by severity level (Intermittent Asthma, IA; Mild Persistent Asthma, MPA; Moderate/Severe Persistent Asthma, MSPA)

	IA	MPA	MSPA	p-value**
	n=229	n=212	n=88	
FEV_1	1.83 (0.68)	1.68 (0.62)	1.49 (0.54)	**<0.001**
Z-FEV_1	0.06 (1.14)	-0.2 (1.07)	-1.37 (1.06)	**<0.001**
$FEV_1\%$	100.6 (13.55)	97.56 (12.94)	83.48 (12.7)	**<0.001**
FVC	2.1 (0.83)	1.99 (0.79)	1.85 (0.72)	**0.041**
Z-FVC	0.17 (1.28)	0.18 (1.15)	-0.75 (1.18)	**<0.001**
FVC %	102.27 (15.94)	102.48 (14.49)	91.17 (14.19)	**<0.001**
$FEF_{25-75\%}$	2.12 (0.84)	1.79 (0.69)	1.47 (0.62)	**<0.001**
Z-$FEF_{25-75\%}$	-0.34 (1.26)	-0.84 (0.83)	-1.63 (0.99)	**<0.001**
$FEF_{25-75\%}\%$	93.08 (32.28)	81.4 (18.15)	65.41 (19.56)	**<0.001**
FEV_1/FVC	87.84 (6.16)	85.4 (7.05)	81.9 (8.83)	**<0.001**
Z-FEV_1/FVC	-0.18 (1.05)	-0.62 (1.06)	-1.01 (1.22)	**<0.001**
FEV_1/FVC %	98.23 (7.09)	95 (7.69)	91.65 (9.89)	**<0.001**

Data are presented as mean (SD). FVC, forced vital capacity; FEV_1, forced expiratory volume in 1 second; FEF_{25-75}, forced inspiratory flow 25-75%; **p-value comes from a X^2 test for categorical variables and form ANOVA for continuous variables.

4.2.1 Regression analysis

Table 4.5 shows IS-lasso model for $FEV_1\%$ with Gamma family and identity link. The model includes $p = 82$ covariates and the response variable is $FEV_1\%$. Male have higher $FEV_1\%$ than females, children with persistent asthma have a lower $FEV_1\%$ predicted than intermittent. Age could be interpreted as a proxy of the asthma duration, indeed, $FEV_1\%$ predicted decreases by 0.51 if age increases by one unit. Environmental factors result to be risk factors for impaired lung function, in particular a main adverse role on $FEV_1\%$ is recorded for proximity to landfill, industry and high traffic road. 95% CIs obtained through Score, are represented in Figure 4.1, in appendix is possible to find a table which reveal the names of the variables associated to the number reported in the figure.

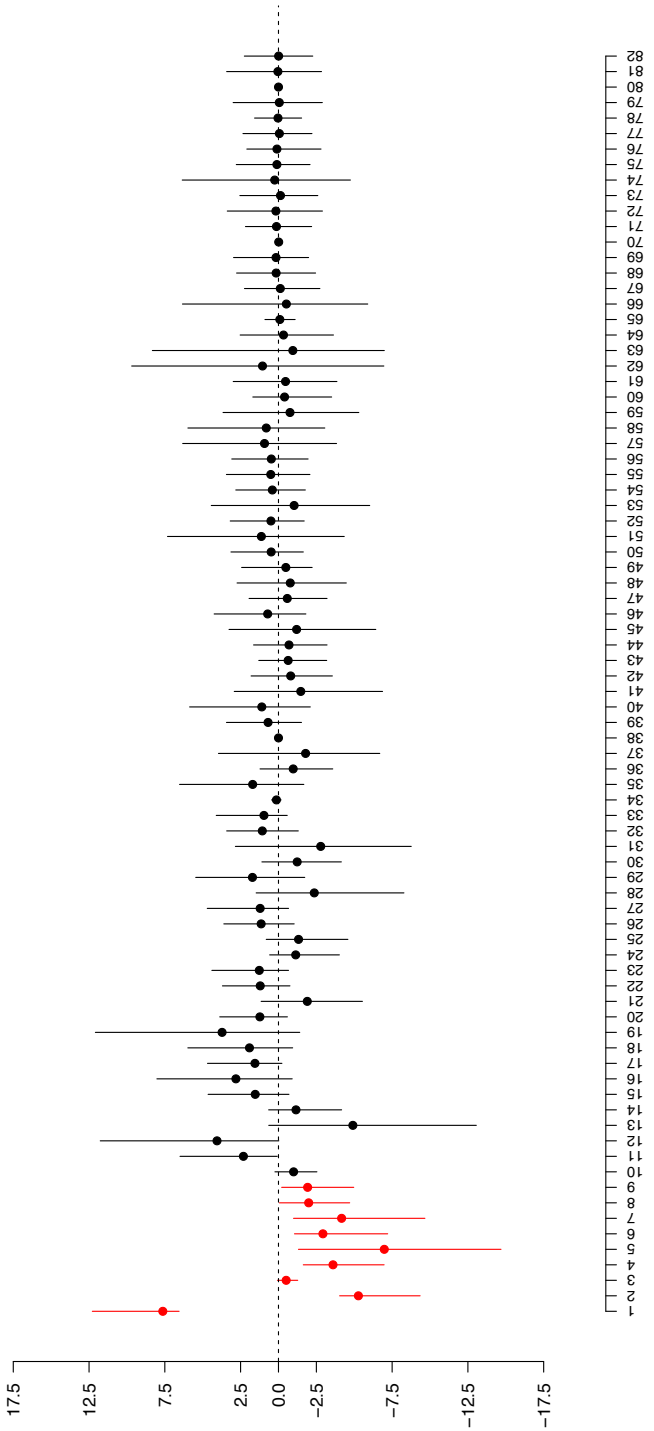

FIGURE 4.1: Estimates and 95% confidence interval of all covariates of the model in table 4.5. Names of the variables are in table A.1. Red line indicates significant IC.

TABLE 4.5: IS-lasso model for $FEV_1\%$ with Gamma family and identity link. The model is estimated including 82 covariates, table shows only significant covariates, $\lambda = 0.94$ is chosen by AIC.

	$\hat{\beta}$	SE	p-value
Male	7.634	0.853	0.000
Persistent	-5.273	0.861	0.000
Age	-0.507	0.137	0.000
Paternal history of asthma	-3.598	1.046	0.001
Proximity to landfill	-6.982	2.701	0.010
Proximity to industry	-4.166	1.717	0.015
Proximity to continuously vehicular traffic	-2.931	1.192	0.014
Proximity to high traffic road(<50m)	-1.923	0.903	0.033
Cough (>4 days/week)	-1.923	0.898	0.026

Table 4.6 reports IS-lasso model for FVC% in logarithmic scale. Male have higher FVC% than females, children with persistent asthma have a lower FVC% predicted than intermittent, FVC% predicted decreases if age increases by one unit and finally current cat exposure has a protective effect on FVC%. 95% CIs obtained through Score, are represented in Figure 4.2, in appendix is possible to find a table which reveal the names of the variables associated to the number reported in the figure.

TABLE 4.6: IS-lasso model for FVC% with Gamma family and logarithmic link. The model is estimated including 82 covariates, table shows only significant covariates, $\lambda = 5$ is chosen by AIC.

	$\hat{\beta}$	SE	p-value
Male	0.130	0.012	0.000
Persistent	-0.022	0.010	0.020
Current cat exposure	0.054	0.024	0.024
Age	-0.003	0.002	0.027
Number of cohabitant	-0.008	0.004	0.045

Results on $FEF_{25-75}\%$ are reported in Table 4.7. Covariates with a significant effect on $FEF_{25-75}\%$ are Paternal hisory of asthma, asthma severity and proximity to continuously vehicular traffic, these variables have a negative effect on the lung function. 95% CIs obtained through Score, are represented in Figure 4.3, in appendix is reported a table which reveal the names of the variables associated to the number reported in the figure.

TABLE 4.7: IS-lasso model for $FEF_{25-75}\%$ with Gamma family and logarithmic link. The model is estimated including 82 covariates, table shows only significant covariates, $\lambda = 2.2$ is chosen by AIC.

	$\hat{\beta}$	SE	p-value
Paternal history of asthma	-0.082	0.033	0.012
Persistent	-0.152	0.027	0.000
Proximity to continuously vehicular traffic	-0.103	0.041	0.011

4.2.2 Comparisons with other proposal

Since both covTest and postSel don't allow to use a Gamma family, in this section we provide a model on FEV_1 with Gaussian family in order to compare the IS-Lasso p-values with the other two competitors. Interestingly, the third and fourth variable are significant with IS-Wald and not significant for the other two tests (Figure 4.4).

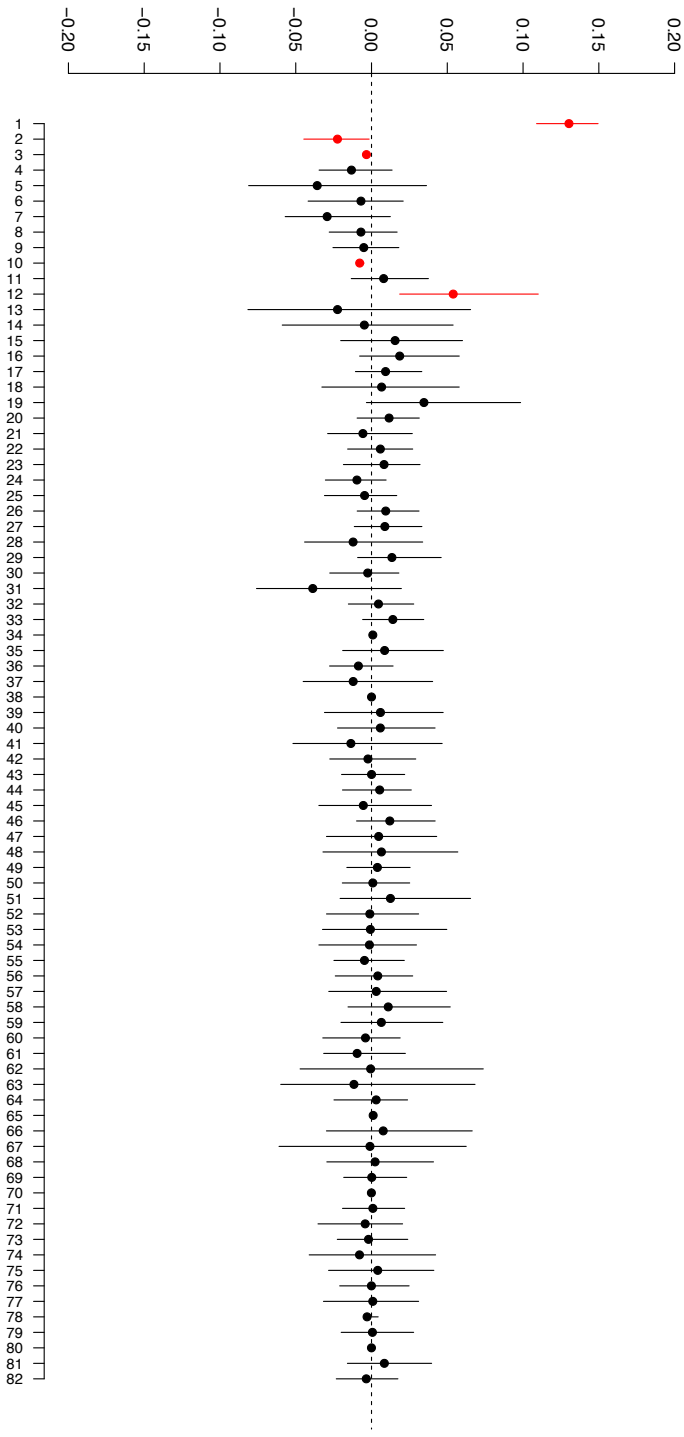

FIGURE 4.2: Estimates and 95% confidence interval of all covariates of the model in table 4.6. Names of the variables are in table A.1. Red line indicates significant IC.

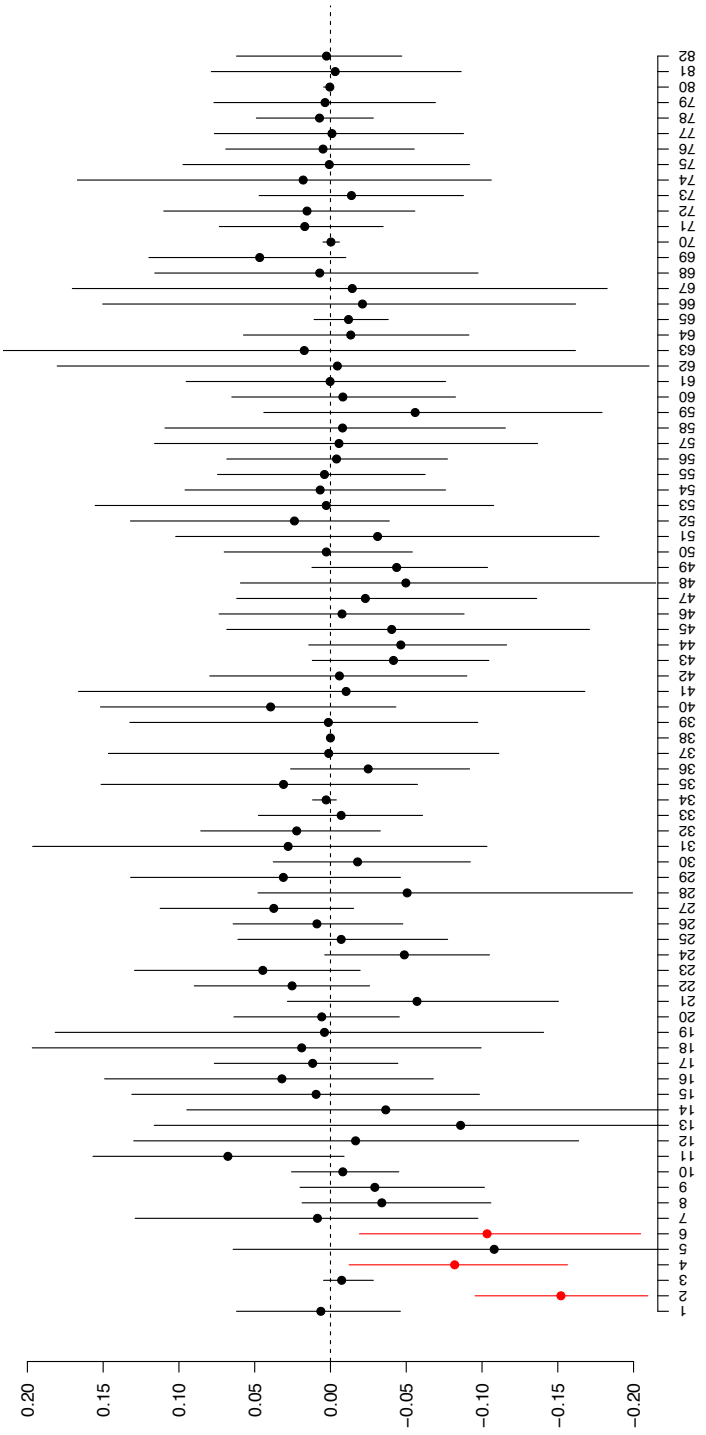

FIGURE 4.3: Estimates and 95% confidence interval of all covariates of the model in table 4.7. Names of the variables are in table A.1. Red line indicates significant IC.

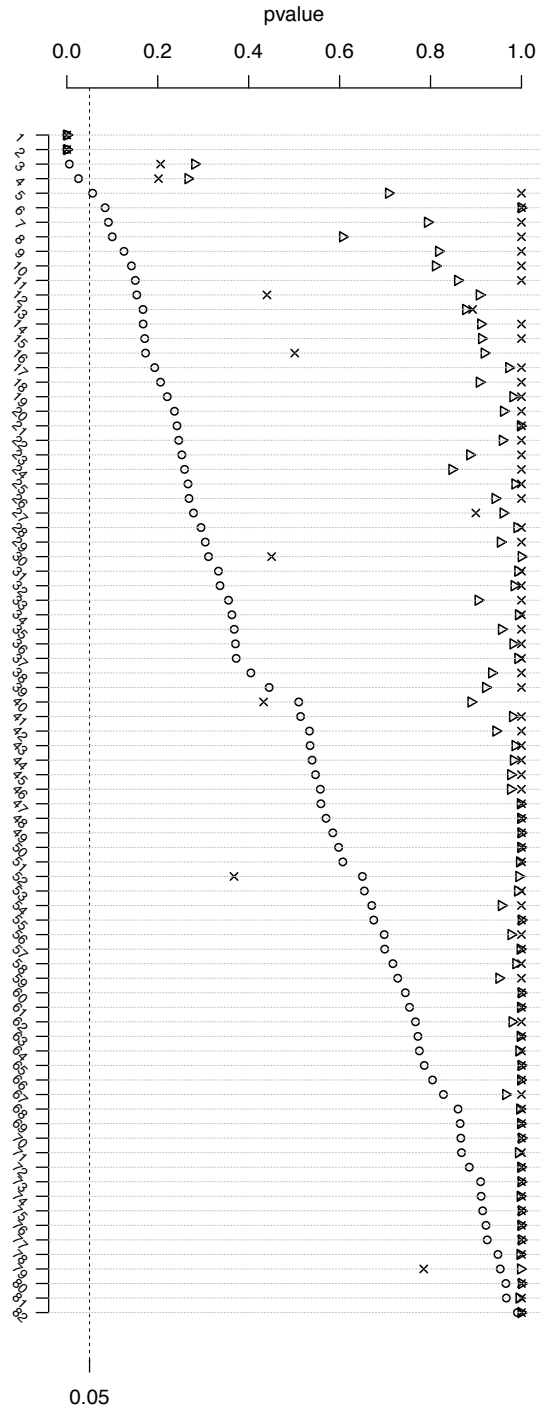

FIGURE 4.4: P-values on real data of IS-Wald (circles), covTest (triangles) and postSel (crosses).

Chapter 5

Conclusion

In this thesis we have introduced the Induced Smoothed lasso, a new framework for regression models with the lasso penalty. Induced smoothing is a relatively new idea that has been successfully employed in some contexts to cope with non-smooth estimating equations, for instance rank regression (Fu et al., 2010). We have applied IS to the lasso regression: in the end the L_1 penalty is replaced by its smooth counterparts wherein the amount of smoothing acting on each coefficient is tuned by the corresponding standard error computed by data. As the sample size increases and the standard error decreases, the IS-lasso gets closer to the original lasso, making the IS-lasso asymptotically equivalent to the lasso. However in finite samples the estimating equations will be always smooth permitting to compute the estimates covariance matrix via the sandwich formula. Reliable standard errors allow to build the usual Wald statistic to test for a non-zero regression coefficient. In addition to the Wald statistic who is useless in interval estimation, we have also assessed Score test. Both the proposed IS-lasso Wald and Score seem to be a good inferential tools in LASSO regression. Simulation experiments for different scenarios discussed in section 3 have showed good results when compared to the other competitors. The coverage levels of the interval estimators are pretty close to the nominal level in different scenarios, even when the theoretical conditions are not met. Implementation of the proposed IS-lasso methods is available in a R package `islasso` with some C++ source code making the R implementation pretty stable and fast. For instance, with $p = 250$ covariates our fitter function requires about 0.38 seconds for $n = 100$, and about 0.66 seconds if $n = 1000$ to achieve convergence (averages on 10 fits) on a McBook Pro Intel(R) Core(TM) i7 CPU at 2 GHz on a macOS Sierra 64 bit machine with 8 GB of RAM. In addition, `binomial`, `poisson` and `Gamma` families have already been implemented. The main function of the package is `islasso`:

```
islasso(formula, family = gaussian(), lambda, data,
```

```
        weights, subset, offset, unpenalized,
        control = list())
```

function is very familiar and requires a `formula` and `family` for specifying the model and family, the `lambda` parameter allows to set a value for the tuning parameter. In addition, `control` allows to choose several option already implemented, as for example the adaptive
method. Since asthma is a multi-factorial disease could be important consider several factor simultaneously, only one study (Pescatore et al., 2014) used the LASSO regression to predict asthma at school age in preschool children with wheeze or cough using 38 covariates. However, no inference measure were reported in Pescatore et al. (2014). We have assessed several environmental and host risk factor for impaired lung function in a sample of a asthmatic children, but in addition the development of IS-Lasso model allow us to provide standard errors, p-values and confidence intervals. Our results are in line with other studies, for instance see Table 4.5, Calderón-Garcidueñas et al. (2003) showed that a lifelong exposure to urban air pollution causes respiratory damage in children. Furthermore exposures to significant concentrations of air pollutants could be a risk factor for neurodegenerative diseases (Calderón-Garcidueñas et al., 2007). In addition, as suggested from other studies (Perzanowski et al., 2002; Gaffin et al., 2012), we found a protective effect of cat exposure.

Appendix A

Appendix

TABLE A.1: Variables included in the IS-Lasso model.

1	Male	42	Maternal smoke current exposure
2	Persistent	43	Kindergarten
3	Age	44	Breastfeeding
4	Paternal history of asthma	45	Angioedema
5	Proximity to landfill	46	Current dog exposure
6	Proximity to continuously vehicular traffic	47	Smoke in the car
7	Proximity to industry	48	Maternal smoke in pregnancy
8	Cough (>4 days/week)	49	Uncontrolled asthma
9	Proximity to high traffic road(<50m)	50	Early mold exposure
10	Number of cohabitants	51	Paternal food allergy
11	Sinusitis	52	Family current smoke exposure
12	Early cat exposure	53	SLIT
13	Tonsillectomy	54	Rhinitis diagnosis
14	Paternal smoke in pregnancy	55	Maternal history of asthma
15	Windows opened in summer	56	Paternal education ($\geq$ 8 years)
16	Maternal food allergy	57	Maternal history of angioedema
17	Conjunctivitis	58	Paternal history of allergy
18	Maternal smoke early exposure	59	Maternal history of eczema
19	Oral syndrome	60	Maternal education ($\geq$ 8 years)
20	House change	61	Swimming
21	Food allergy	62	Maternal history of bronchiolitis
22	Proximity to bus stop	63	Paternal history of bronchiolitis
23	Proximity to continuously truck traffic	64	No disease at born
24	Caesarean delivery	65	Weaning age (months)
25	Atopy	66	Paternal history of angioedema
26	Laryngospasm	67	Paternal early smoke exposure
27	Paternal history of rhinitis	68	Windows opened in winter
28	Current cat exposure	69	Eczema
29	Maternal history of allergy	70	Weight gain in pregnancy
30	Wheezing	71	Snoring
31	Anaphilaxis	72	Intense traffic
32	Upper respiratory infection	73	Current mold exposure
33	Exercise induced bronchoconstriction	74	Proximity to power plant
34	Pollution bother	75	Passive smoke exposure
35	Early dog exposure	76	Maternal history of rhinitis
36	Broncholitis	77	Rhinitis
37	Paternal history of eczema	78	Order of birth
38	Birth weight	79	Otitis
39	Paternal current smoke exposure	80	Weight
40	Pre-term born	81	Catarrh
41	Adenotonsillectomy	82	Physical activity ($\geq$ 3 times at week)

Bibliography

Akaike, H., 1998. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike. Springer, pp. 199–213.

Augugliaro, L., Mineo, A. M., Wit, E. C., 2013. Differential geometric least angle regression: a differential geometric approach to sparse generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75, 471–498.

Beck, A., Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences 2, 183–202.

Belloni, A., Chernozhukov, V., Hansen, C., 2014. Inference on treatment effects after selection among high-dimensional controls. The Review of Economic Studies 81, 608–650.

Beran, R., 1982. Estimated sampling distributions: the bootstrap and competitors. The Annals of Statistics, 212–225.

Bernstein, I. L., Li, J. T., Bernstein, D. I., Hamilton, R., Spector, S. L., Tan, R., Sicherer, S., Golden, D. B., Khan, D. A., Nicklas, R. A., et al., 2008. Allergy diagnostic testing: an updated practice parameter. Annals of Allergy, Asthma & Immunology 100 (3), S1–S148.

Boos, D. D., 1992. On generalized score tests. The American Statistician 46 (4), 327–333.

Boot, T., Nibbering, D., 2017. Confidence intervals in high-dimensional regressions based on regularized psuedoinverses. arXiv preprint arXiv:1703.03282.

Bracewell, R. N., Bracewell, R. N., 1986. The Fourier transform and its applications. Vol. 31999. McGraw-Hill New York.

Breheny, P., Huang, J., 2011. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. The annals of applied statistics 5 (1), 232.

Brown, B., Wang, Y.-G., 2007. Induced smoothing for rank regression with censored survival times. Statistics in medicine 26 (4), 828–836.

Brown, B. M., Wang, Y. G., 2005. Standard errors and covariance matrices for smoothed rank estimators. Biometrika, 149–158.

Bühlmann, P., et al., 2013. Statistical significance in high-dimensional linear models. Bernoulli 19 (4), 1212–1242.

Calderón-Garcidueñas, L., Franco-Lira, M., Torres-Jardón, R., Henriquez-Roldán, C., Barragán-Mejía, G., Valencia-Salazar, G., Gonzaléz-Maciel, A., Reynoso-Robles, R., Villarreal-Calderón, R., Reed, W., 2007. Pediatric respiratory and systemic effects of chronic air pollution exposure: nose, lung, heart, and brain pathology. Toxicologic Pathology 35 (1), 154–162.

Calderón-Garcidueñas, L., Mora-Tiscareño, A., Fordham, L. A., Valencia-Salazar, G., Chung, C. J., Rodriguez-Alcaraz, A., Paredes, R., Variakojis, D., Villarreal-Calderón, A., Flores-Camacho, L., et al., 2003. Respiratory damage in children exposed to urban pollution. Pediatric Pulmonology 36 (2), 148–161.

Candes, E. J., Plan, Y., 2009. Near ideal model selection by l1 minimization. The Annals of Statistics 37, 2145—2177.

Craven, P., Wahba, G., 1978. Smoothing noisy data with spline functions. Numerische mathematik 31 (4), 377–403.

Dezeure, R., Bühlmann, P., Meier, L., Meinshausen, N., 2015. High-dimensional inference: Confidence intervals, p-values and R-software hdi. Statistical Science 30 (4), 533–558.

Dezeure, R., Bühlmann, P., Zhang, C.-H., 2016. High-dimensional simultaneous inference with the bootstrap. arXiv preprint arXiv:1606.03940.

Dismuke, C., Lindrooth, R., 2006. Ordinary least squares. Methods and Designs for Outcomes Research 93, 93–104.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., et al., 2004. Least angle regression. The Annals of statistics 32 (2), 407–499.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.

Fan, Y., Tang, C. Y., 2013. Tuning parameter selection in high dimensional penalized likelihood. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75 (3), 531–552.

Friedman, J., Hastie, T., Höfling, H., Tibshirani, R., et al., 2007. Pathwise coordinate optimization. The Annals of Applied Statistics 1 (2), 302–332.

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software 33, 1–22.

Fu, L., Wang, Y.-G., Bai, Z., 2010. Rank regression for analysis of clustered data: A natural induced smoothing approach. Computational Statistics and Data Analysis 54, 1036–1050.

Gaffin, J. M., Spergel, J. M., Boguniewicz, M., Eichenfield, L. F., Paller, A. S., Fowler Jr, J. F., Dinulos, J. G., Tilles, S. A., Schneider, L. C., Phipatanakul, W., 2012. Effect of cat and daycare exposures on the risk of asthma in children with atopic dermatitis. In: Allergy and Asthma Proceedings. Vol. 33. OceanSide Publications, p. 282.

GINA, 2017. Global initiative for asthma; global strategy for asthma management and prevention. www.ginasthma.org.

Hu, F., Kalbfleisch, J. D., 2000. The estimating function bootstrap. Canadian Journal of Statistics 28 (3), 449–481.

Jagannath, R., Upadhye, N. S., 2016. The lasso estimator: Distributional properties. arXiv preprint arXiv:1605.03280.

Janková, J., Van De Geer, S., et al., 2015. Confidence intervals for high-dimensional inverse covariance estimation. Electronic Journal of Statistics 9 (1), 1205–1229.

Javanmard, A., Montanari, A., 2014. Confidence intervals and hypothesis testing for high-dimensional regression. Journal of Machine Learning Research 15, 2869–2909.

Knight, K., Fu, W., 2000. Asymptotics for lasso-type estimators. Annals of Statistics 28, 1356–1378.

Kyung, M., Gill, J., Ghosh, M., Casella, G., et al., 2010. Penalized regression, standard errors, and bayesian lassos. Bayesian Analysis 5 (2), 369–411.

Lan, W., Zhong, P.-S., Li, R., Wang, H., Tsai, C.-L., 2016. Testing a single regression coefficient in high dimensional linear models. Journal of Econometrics 195 (1), 154–168.

Lee, J., Taylor, J., 2014. Exact post model selection inference for marginal screening. Advances in Neural Information Processing Systems, 136–144.

Lee, J. D., Sun, D. L., Sun, Y., Taylor, J., 2016. Exact post-selection inference, with application to the lasso. The Annals of Statistics 44, 907–927.

Lockhart, R., Taylor, J., Tibshirani, R., Tibshirani, R., 2013. covTest: Computes covariance test for adaptive linear modelling. R package version 1.02.
URL https://CRAN.R-project.org/package=covTest

Lockhart, R., Taylor, J., Tibshirani, R. J., Tibshirani, R., 2014. A significance test for the lasso. The Annals of Statistics 42, 413–468.

McCullagh, P., 1984. Generalized linear models. European Journal of Operational Research 16 (3), 285–292.

Meinshausen, N., 2015. Group bound: confidence intervals for groups of variables in sparse high dimensional regression without assumptions on the design. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77 (5), 923–945.

Meinshausen, N., Bühlmann, P., 2006. High-dimensional graphs and variable selection with the lasso. The Annals of Statistics, 1436–1462.

Meinshausen, N., Bühlmann, P., 2009. P-values for high-dimensional regression. Journal of the American Statistical Association 104 (488), 1671–1681.

Migliore, E., Berti, G., Galassi, C., Pearce, N., Forastiere, F., Calabrese, R., Armenio, L., Biggeri, A., Bisanti, L., Bugiani, M., et al., 2009. Respiratory symptoms in children living near busy roads and their relationship to vehicular traffic: results of an italian multicenter study (sidria 2). Environmental Health 8 (1), 27.

Miller, M. R., Hankinson, J., Brusasco, V., Burgos, F., Casaburi, R., Coates, A., Crapo, R., Enright, P. v., Van Der Grinten, C., Gustafsson, P., et al., 2005. Standardisation of spirometry. European Respiratory Journal 26 (2), 319–338.

Minnier, J., Lu, T., Tianxi, C., 2012. A perturbation method for inference on regularized regression estimates. Journal of the American Statistical Association 106, 1371–1382.

Osborne, M. R., Presnell, B., Turlach, B. A., 2000. On the lasso and its dual. Journal of Computational Graphical Statistics 9, 319–337.

Park, M. Y., Hastie, T., 2007. L1-regularization path algorithm for generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69 (4), 659–677.

Perzanowski, M. S., Ronmark, E., Platts-Mills, T. A., Lundback, B., 2002. Effect of cat and dog ownership on sensitization and development of asthma among preteenage children. American Journal of Respiratory and Critical Care Medicine 166 (5), 696–702.

Pescatore, A. M., Dogaru, C. M., Duembgen, L., Silverman, M., Gaillard, E. A., Spycher, B. D., Kuehni, C. E., 2014. A simple asthma prediction tool for preschool children with wheeze or cough. Journal of Allergy and Clinical Immunology 133 (1), 111–118.

Pötscher, B. M., Leeb, H., 2009. On the distribution of penalized maximum likelihood estimators: The lasso, scad, and thresholding. Journal of Multivariate Analysis 10, 2065–2082.

Pötscher, B. M., Schneider, U., 2009. On the distribution of the adaptive lasso estimator. Journal of Statistical Planning and Inference 139 (8), 2775–2790.

Quanjer, P. H., Hall, G. L., Stanojevic, S., Cole, T. J., Stocks, J., 2012. Age- and height-based prediction bias in spirometry reference equations. European Respiratory Journal 40 (1), 190–197.

Royall, R. M., 1986. Model robust confidence intervals using maximum likelihood estimators. International Statistical Review 54, 221–226.

Simoni, M., Lombardi, E., Berti, G., Rusconi, F., La Grutta, S., Piffer, S., Petronio, M., Galassi, C., Forastiere, F., Viegi, G., 2005. Mould/dampness exposure at home is associated with respiratory disorders in italian children and adolescents: the sidria-2 study. Occupational and environmental medicine 62 (9), 616–622.

Strunk, R. C., Weiss, S. T., Yates, K. P., Tonascia, J., Zeiger, R. S., Szefler, S. J., Group, C. R., et al., 2006. Mild to moderate asthma affects lung

growth in children and adolescents. Journal of Allergy and Clinical Immunology 118 (5), 1040–1047.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.

Tibshirani, R., 2011. Regression shrinkage and selection via the lasso: a retrospective. Journal of the Royal Statistical Society. Series B (Methodological) 73, 273–282.

Tibshirani, R., Tibshirani, R., Taylor, J., Loftus, J., Reid, S., 2017. selectiveInference: Tools for Post-Selection Inference. R package version 1.2.2. URL https://CRAN.R-project.org/package=selectiveInference

Tibshirani, R. J., Taylor, J., 2012. Degrees of freedom in lasso problems. The Annals of Statistics 40 (2), 1198–1232.

Tibshirani, R. J., Taylor, J., Lockhart, R., Tibshirani, R., 2016. Exact post-selection inference for sequential regression procedures. Journal of the American Statistical Association 111, 600–620.

Tutz, G., Gertheiss, J., 2016. Regularized regression for categorical data (with discussion). Statistical Modelling 16, 161–200.

Van de Geer, S., Bühlmann, P., Ritov, Y. A., Dezeure, R., 2014. On asymptotically optimal confidence regions and tests for high-dimensional models. The Annals of Statistics 42, 1166–1202.

Wainwright, M. J., 2009. Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (lasso). IEEE transactions on information theory 55, 2183–2202.

Wasserman, L., Roeder, K., 2009. High dimensional variable selection. Annals of statistics 37 (5A), 2178.

Zhang, C., Huang, J., 2008. The sparsity and bias of the lasso selection in high-dimensional linear regression. The Annals of Statistics, 1567–1594.

Zhang, C., Zhang, S., 2014. Confidence intervals for low dimensional parameters in high dimensional linear models. Journal of the Royal Statistical Society: Series B 76, 217–242.

Zhang, X., Cheng, G., 2017. Simultaneous inference for high-dimensional linear models. Journal of the American Statistical Association, 1–12.

Zhang, Y., Li, R., Tsai, C.-L., 2010. Regularization parameter selections via generalized information criterion. Journal of the American Statistical Association 105 (489), 312–323.

Zhao, P., Yu, B., 2006. On model selection consistency of lasso. Journal of Machine learning research 7, 2541–2563.

Zou, H., 2006. The adaptive lasso and its oracle properties. Journal of the American statistical association 101 (476), 1418–1429.

Zou, H., Hastie, T., Tibshirani, R., et al., 2007. On the "degrees of freedom" of the lasso. The Annals of Statistics 35 (5), 2173–2192.