Fighting Multicollinearity in Double Selection: A Bayesian Approach

Research practice 2: Proposal presentation

Mateo Graciano-Londoño
Mathematical Engineering
Student

Andrés Ramírez-Hassan
Department of Economics
Tutor

Universidad EAFIT, Medellín
Colombia

April 8^{th}, 2016

UNIVERSIDAD EAFIT®

Intuition on what we want to do

"Many empirical analyses focus on estimating the structural, causal, or treatment effect of some variable on an outcome of interest. For example, we might be interested in estimating the causal effect of some government policy on an economic outcome such as employment.(...) A problem empirical researchers face when relying on a conditional-on-observables identification strategy for estimating a structural effect is knowing which controls to include." [Belloni et al., 2014, pp. 608-609].

Problem statement

Consider the following structure [Belloni et al., 2014]:

$$y_i = \alpha d_i + x_i^{'} \beta_g + \epsilon_i \tag{1}$$

$$d_i = x_i^{'} \beta_m + \zeta_i \tag{2}$$

where y_i is the response, β_g, β_m are the structural and treatments effects of variables x_i respectively, d_i is the treatment, α is the treatment effect and ϵ_i, ζ_i are stochastic errors such that

$$E\left[\epsilon_i \mid x_i, d_i\right] = E\left[\zeta_i \mid x_i\right] = 0$$

Previous works

Belloni et al. [2014] showed that assumptions over the distribution of $\sqrt{n}(\alpha - \hat{\alpha})$ are not always true via simulation:

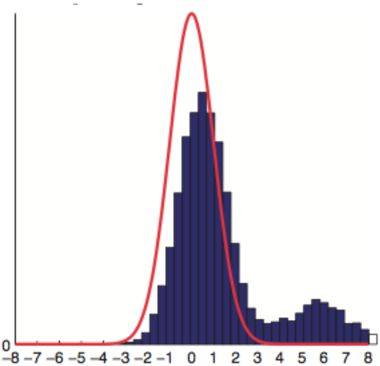


Figure: Theorical and simulated distribution, taken from Belloni et al. [2014].

Previous works: LASSO

The Lasso estimator as introduced in Tibshirani [1996] is an optimization problem which solves the following:

$$\beta^* = \min_{\beta \in R^p} \sum_{i=1}^{n} [d_i - x_i^{'} \beta_m]^2 + \lambda \sum_{j=1}^{p} \mid \beta_j \mid \qquad (3)$$

where λ is a penalization coefficient.

Previous works: Post double LASSO

Post double LASSO estimator is a three stages procedure:

1. Proceed with LASSO estimator on the treatment effect.
2. Proceed with LASSO estimator on the structural equation but without including the treatment.
3. Proceed with a linear regression on the structural equation using the treatment and the union of variables that were selected on previews stages.

MC^3

Markov chain Monte Carlo model composition (MC^3) is a Bayesian methodology which uses a stochastic search comparing different models by its posterior model probability.

As in Simmons et al. [2010], let $M = \{M_1, M_2, ...M_m\}$ the set of models under consideration, and d the observed data as in (2).

MC^3

The posterior model probability for model M_j is defined as

$$P(M_j \mid d, M) = \frac{P(d \mid M_j)\pi(M_j)}{\sum_{i=1}^{m} P(d \mid M_i)\pi(M_i)} \quad \forall j = 1, 2, ...m$$

where $P(d \mid M_j)$ is the integrated likelihood of the model M_j and $\pi(M_j)$ is the prior probability that M_j is the true model.

MC^3 with nonlocal(NL) priors

The idea of a nonlocal (to 0) prior is to effectively eliminate models with unnecessary explanatory variables, for instance consider the following nonlocal prior proposed by Johnson and Rossell [2012]:

$$\pi(\beta \mid \tau,\ \sigma^2,\ r,\ A_p) = d_p(2\pi)^{-p/2}(\tau\sigma^2)^{-rp-p/2} \mid A_p \mid^{1/2} exp\left\{-\frac{1}{2\tau\sigma^2}\beta' A_p \beta\right\}\prod_{i=1}^{p}\beta_i^{2r} \quad (4)$$

where τ, r, A_p are hyper-parameters for the prior.

General objective

Propose a double post MC3 estimators based on local and non local prior distributions, and compare its performance with the frequentist counterpart under different multicollinearity degrees.

Specific objectives

- Implement the post double selection and MC^3 on simulations exercises. ✓
- Gather real information as in Donohue III and Levitt [2001], and use both methodologies.
- Compare both methodologies and analyse how they perform based on simulation and real cases.

Model specification

Considering (2) and (1) we define $dim(x_i) = 40$, $\alpha = 0$, β_g such that there are only 8 non zero coefficients and β_m with only 4 non zero coefficients.
We also define:

$$x_i1 = N_{10}(0, \Sigma)$$
$$x_i2 = N_5(0, I)$$
$$x_i3 = x_{i,j} = f_j(x_i1, x_i2) \qquad \forall j \in \{1, 2, ..., 25\}$$

where f_j is a non linear function and define.

$$x_i = (x_i1, x_i2, x_i3)$$

Defining different levels of Multillinearity

We define three different types of experiments with different Σ to generate x_i1,

1. Σ so that $\sigma_{ij} \in (0.5, 0.9)$
2. Σ so that $\sigma_{ij} \in (0, 0.5)$
3. $\Sigma = I_{10}$

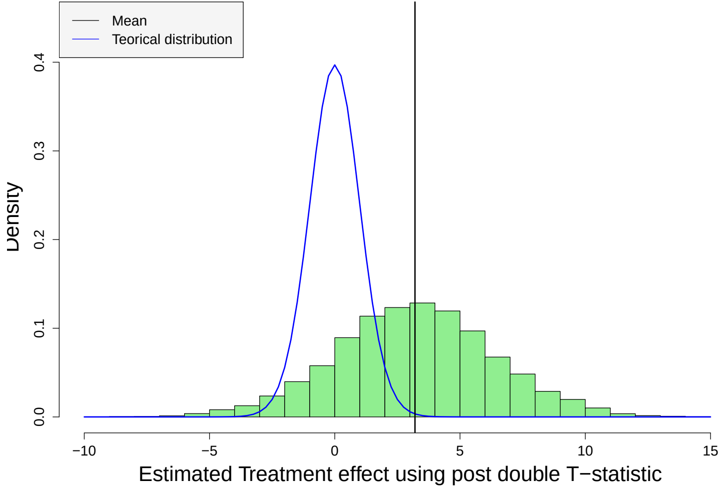
we also set the signal to noise ratio equals to 1 in both, the structural and the treatment equation.

Defining different levels of multillinearity

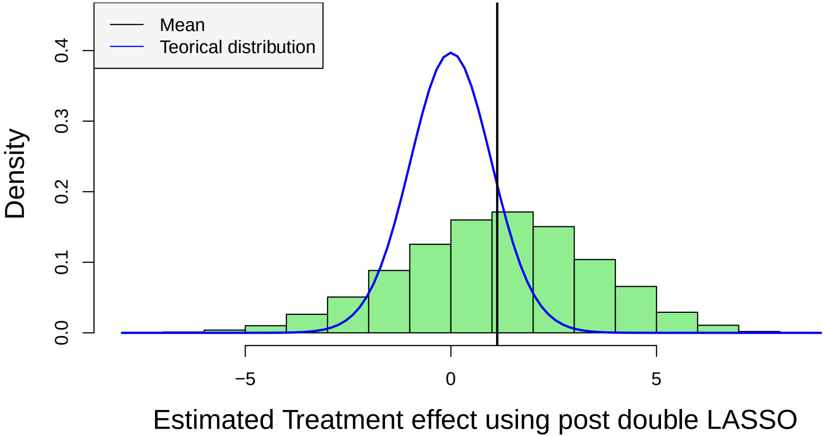
	Multicollinearity level		
Measure	Type 1	Type 2	Type 3
VIF	99.40	4.26	2.86
Condition number	111.84	20.71	12.72

As expected the condition number and the variance inflation factor (VIF) for the first case is clearly higher than the others due to its higher multicollinearity given by the definition of Σ in that case.

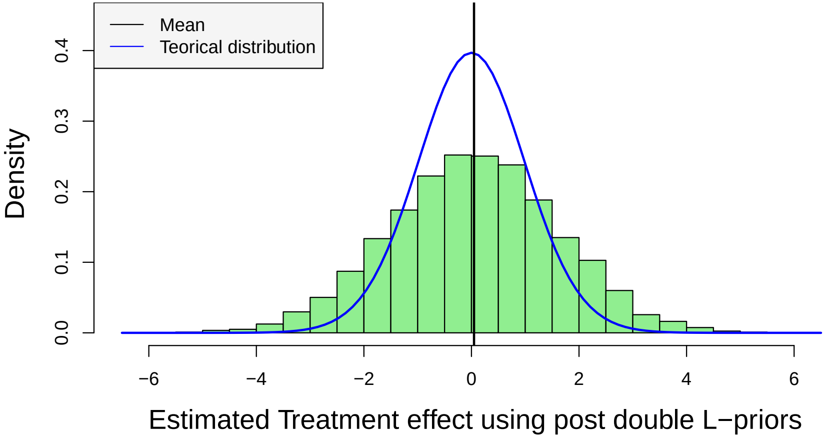
Type 1 results



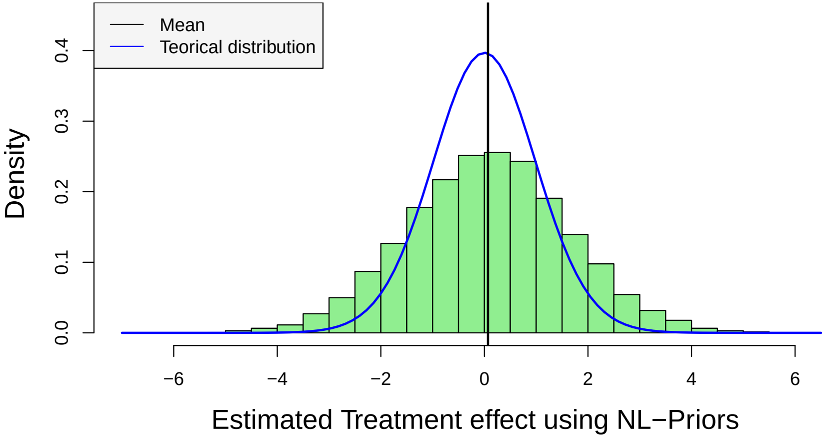
Type 1 results



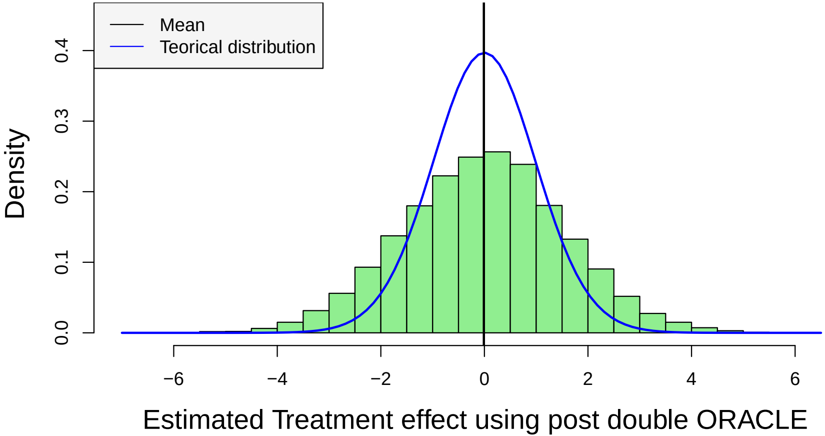
Type 1 results



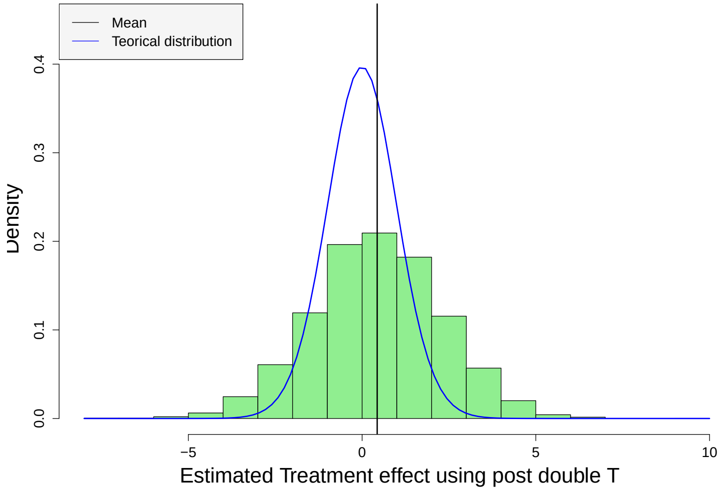
Type 1 results



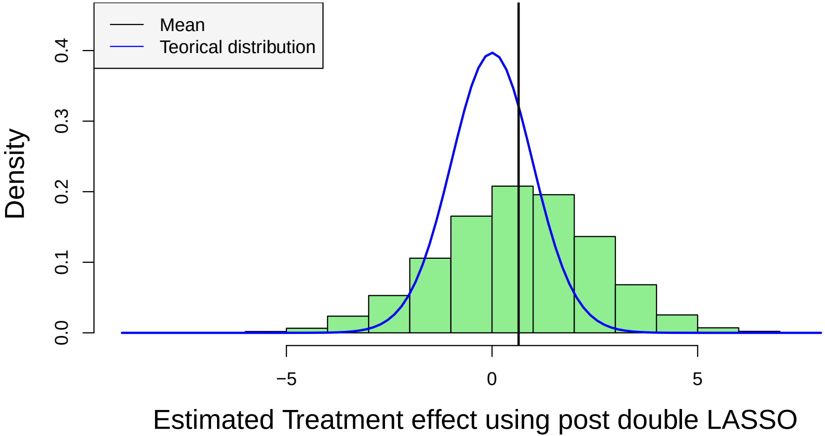
Type 1 results



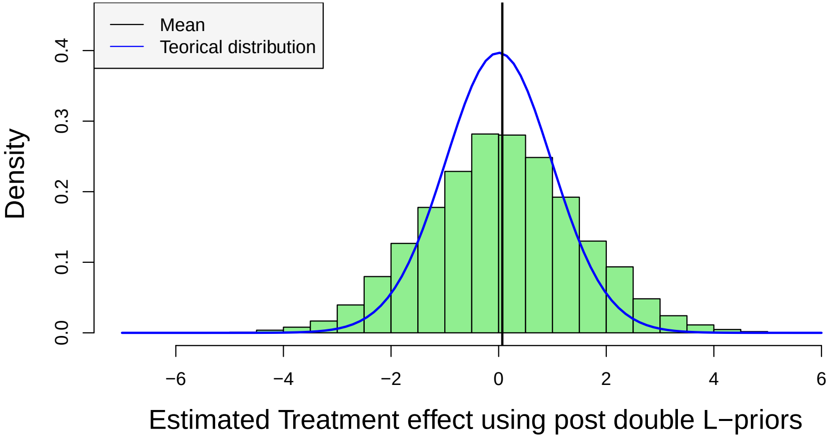
Type 2 results



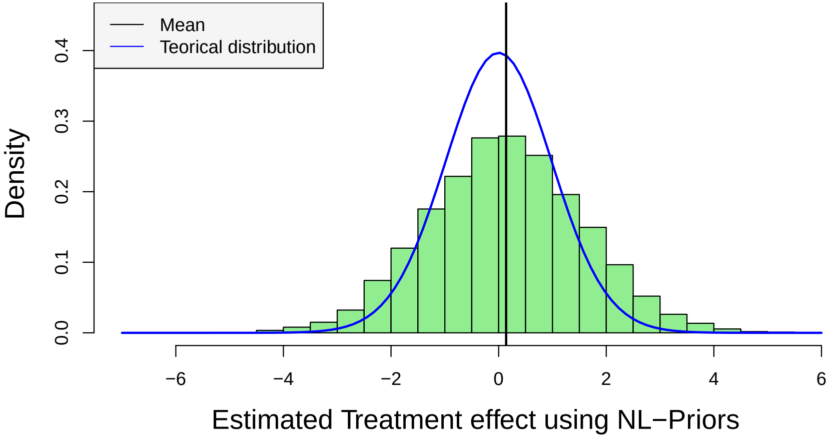
Type 2 results



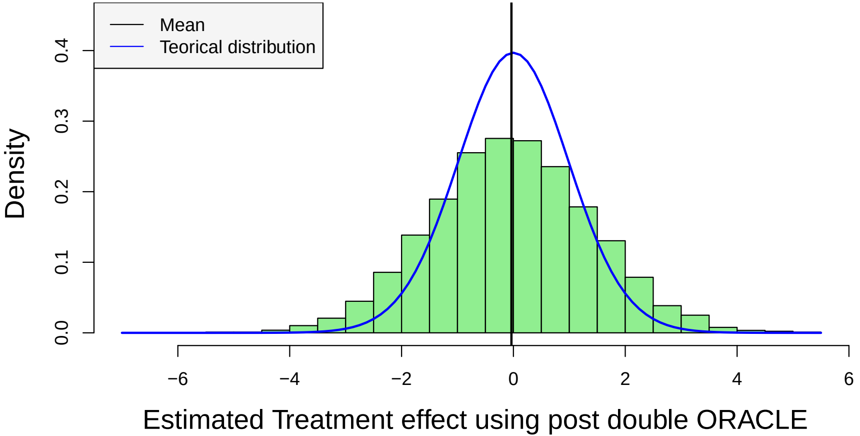
Type 2 results



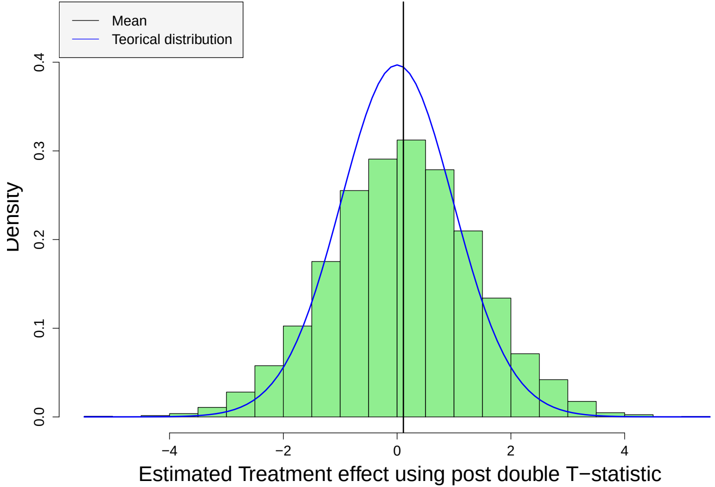
Type 2 results



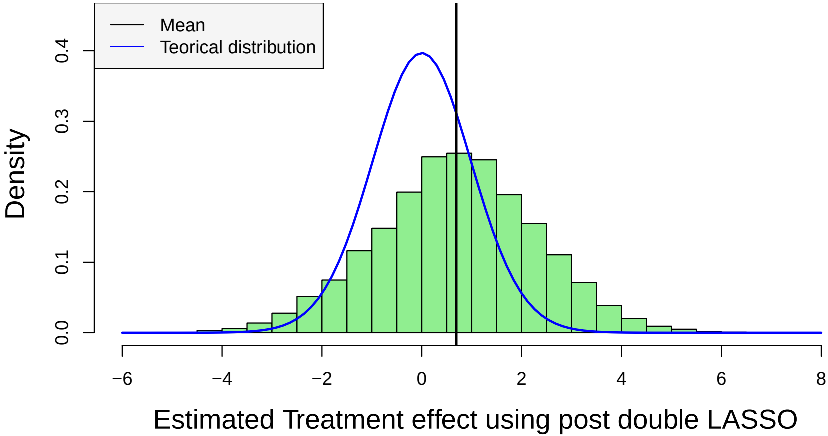
Type 2 results



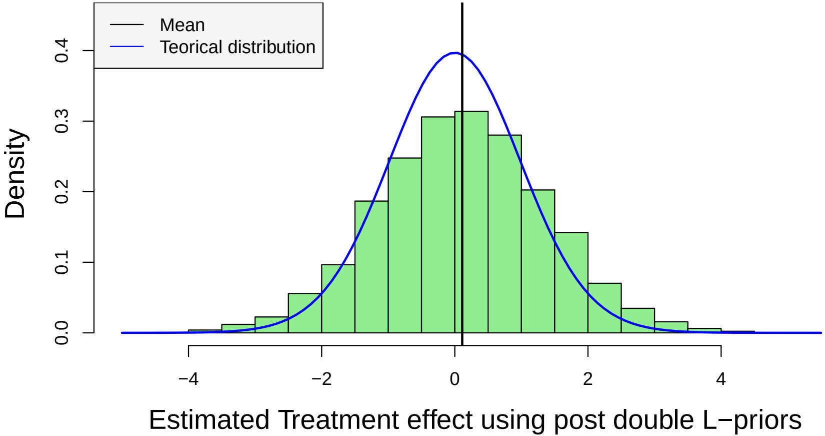
Type 3 results



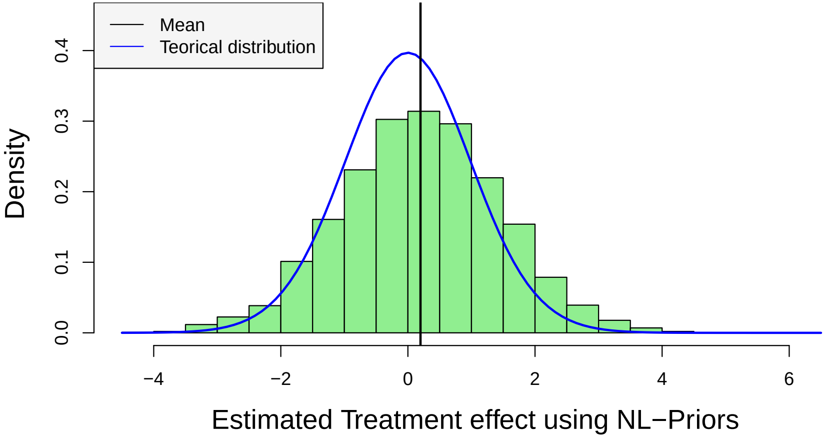
Type 3 results



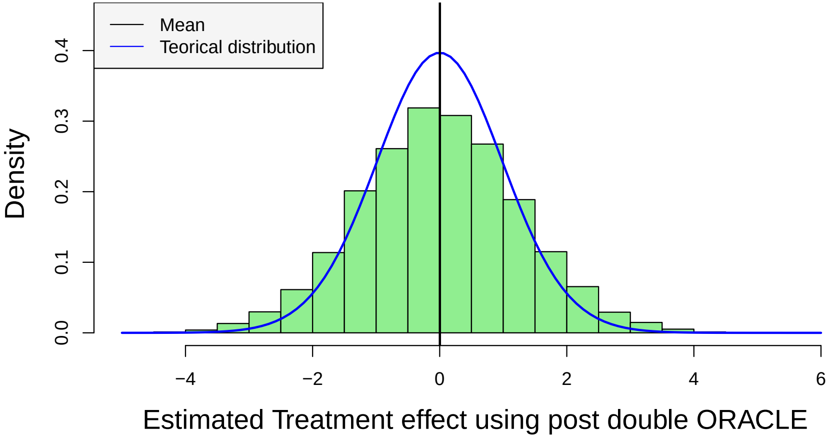
Type 3 results



Type 3 results



Type 3 results



Summary

	Multicollinearity level		
Procedure	Type 1	Type 2	Type 3
Post double T	0.146	0.0699	0.0495
Post double LASSO	0.1311	0.0732	0.0661
Post double L-prior	0.0511	0.0488	0.0513
Post double NL-prior	0.0551	0.0517	0.0560
Post double ORACLE	0.0551	0.0517	0.0560

Table: Rejection rates (at 0.05) for different set of data based on 8000 Monte Carlo simulation.

Summary

Note that even with a high multicollinearity level Bayesian procedures has 0.05 rejection rate which is the teorical expected value.

Also it is impresive that NL-prior selection leads to the same results as the non plausible procedure post double ORACLE.

References

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.

Donohue III, J. J. and Levitt, S. D. (2001). The impact of legalized abortion on crime. *Quarterly Journal of Economics*, 116(2):379–420.

Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107(498):649–660.

References

Simmons, S. J., Fang, F., Fang, Q., and Ricanek, K. (2010). Markov chain Monte Carlo model composition search strategy for quantitative trait loci in a Bayesian hierarchical model. *World Academy of Science, Engineering and Technology*, 63:58–61.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.

Any questions?