MATH 680 - Assignment 2 - March 30, 2015

Sahir Bhatnagar & Maxime Turgeon

Department of Epidemiology, Biostatistics and Occupational Health
McGill University

March 28, 2015

1 Numerical Results

1.1 Introduction

1.2 Nonparametric Bootstrap Smoothing

model	m	Cp	Bootstrap
Linear	2	1151	20
Quadratic	3	1434	13
Cubic	4	636	35
Quartic	5	1579	8
Quintic	6	1776	19
Sextic	7	2758	5

Table 1: C_p model selection for the Cholesterol data. $\sigma = 22$ was used in all bootstrap replications. Last column shows percentage each model was selected as the C_p minimizer, among $B = 4000$ bootstrap replications

	m1	m2	m3	m4	m5	m6
Mean	-13.87	-3.51	5.13	-1.68	-4.64	-12.01
St.dev.	3.45	3.40	5.87	5.80	11.90	45.35

Table 2: Mean and standard deviation of $\hat{\mu}_1^*$ as a function of the selected model, 4000 nonparametric boostrap replications.

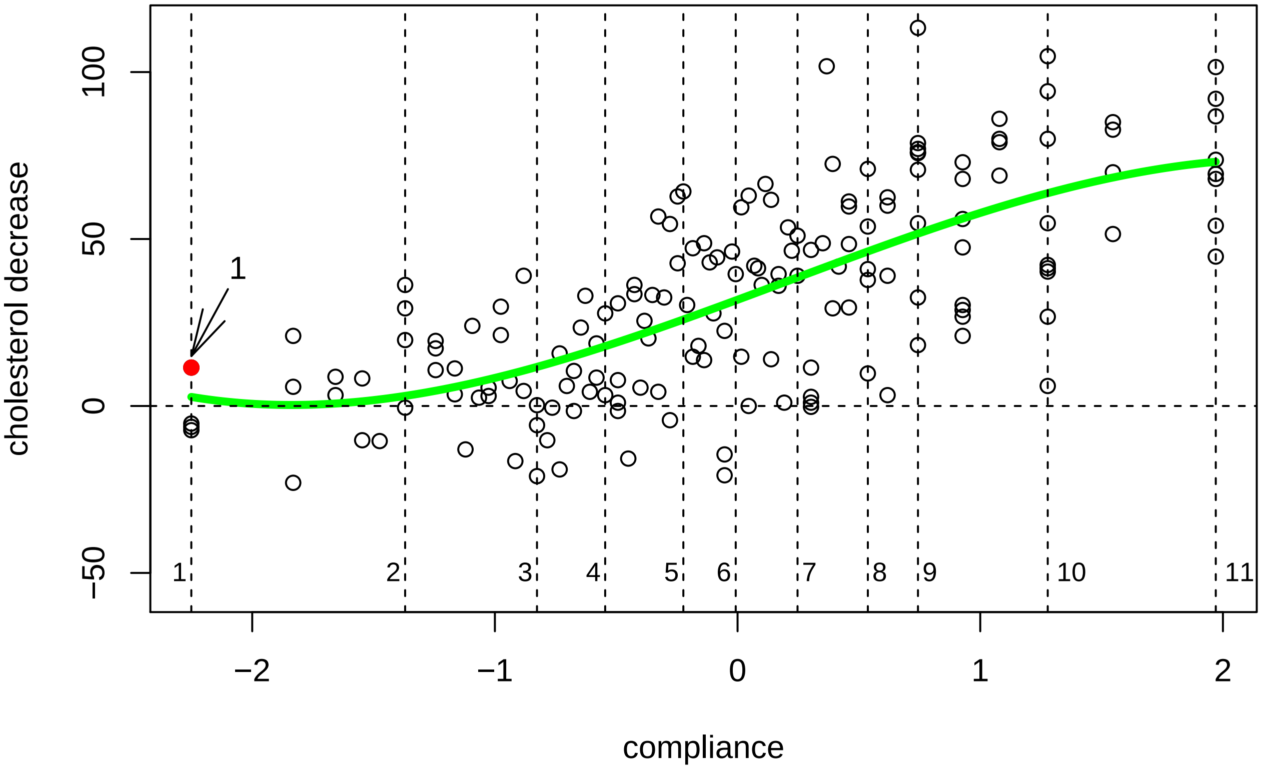


Figure 1: Choleterol decrease vs. compliance with cubic regression fitted curve

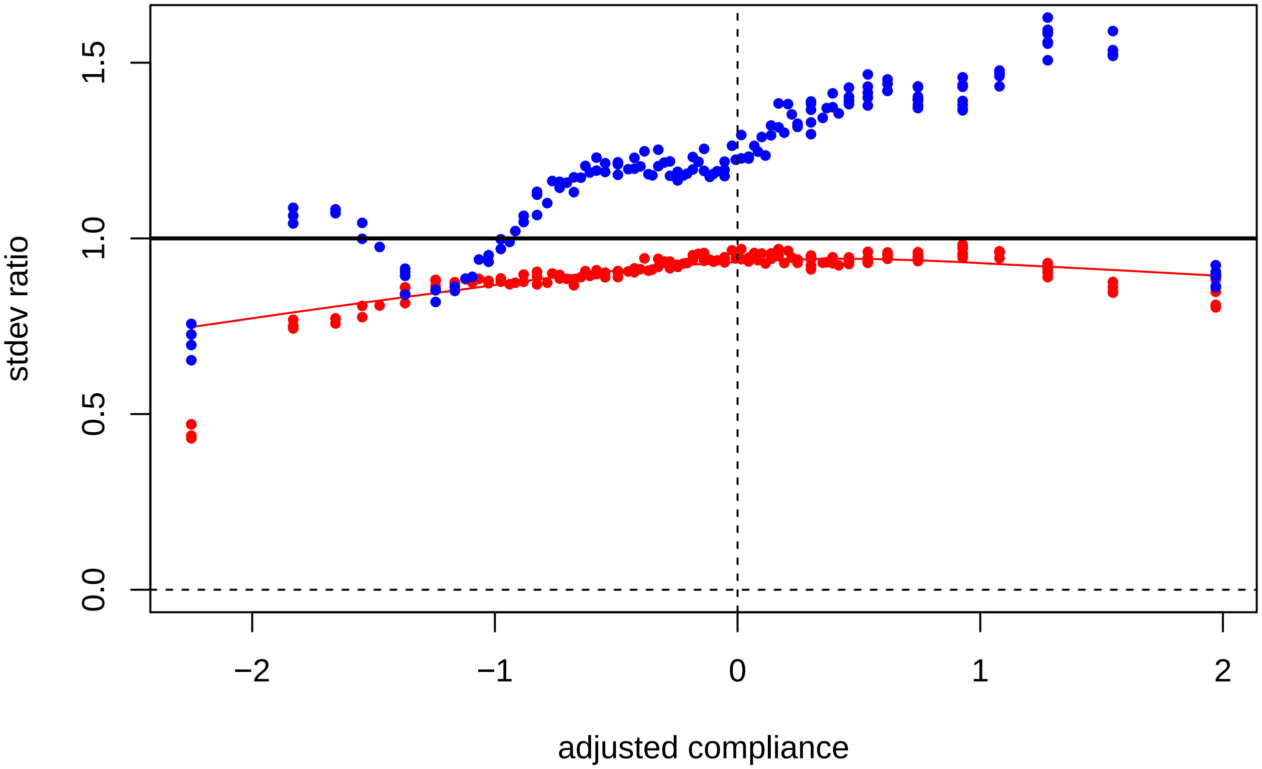


Figure 2: Solid points: ratio of standard deviations, taking account of model selection or not, for the 164 values $\hat{\mu}_j$ from the regression curve in Figure 1. Dashed line: ratio of $\tilde{sd}_B$ to $\hat{sd}_B$

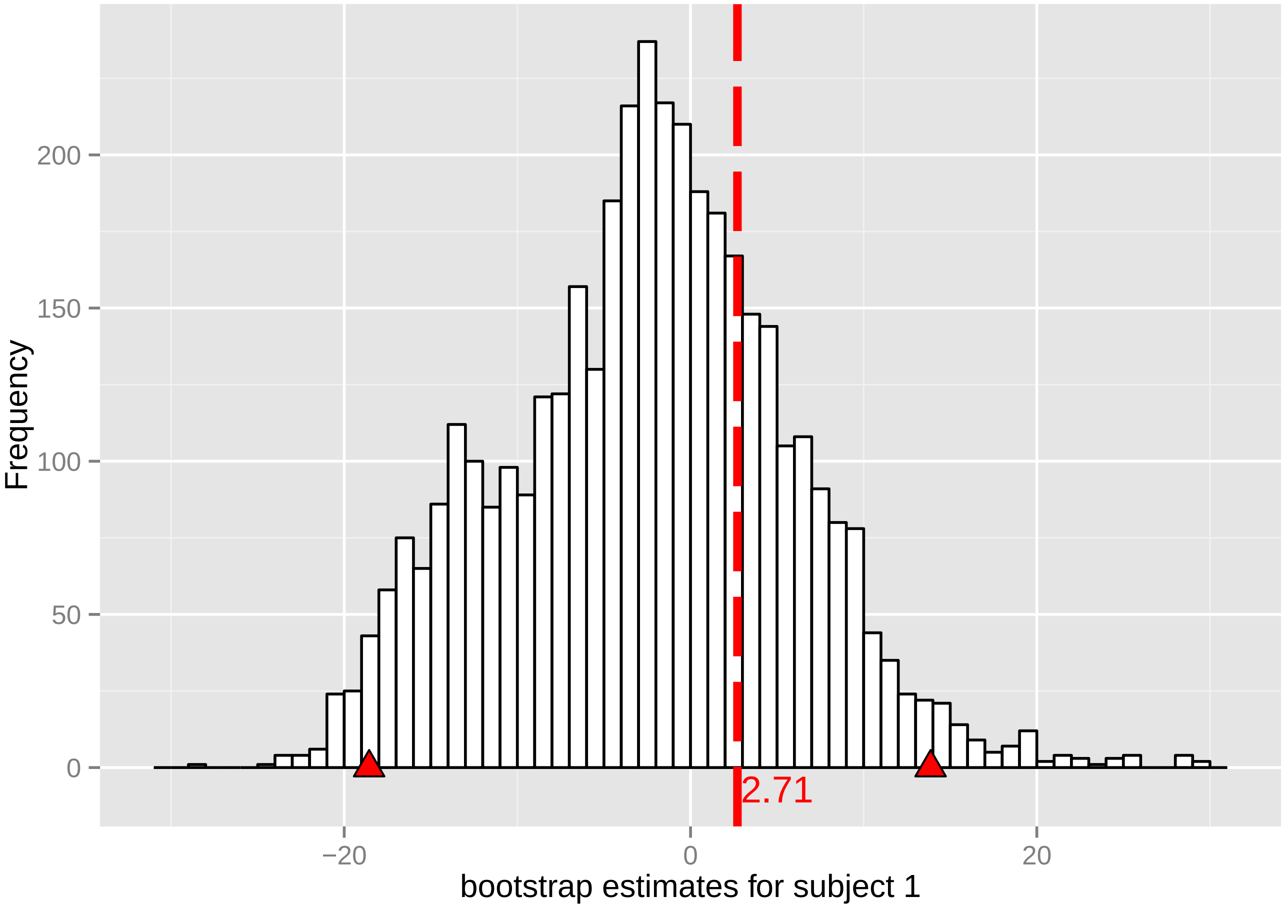


Figure 3: B=4000 bootstrap replications of the Cp-OLS regression estimate for Subject 1. Triangles indicate 2.5th and 97.5th percentiles of the histogram

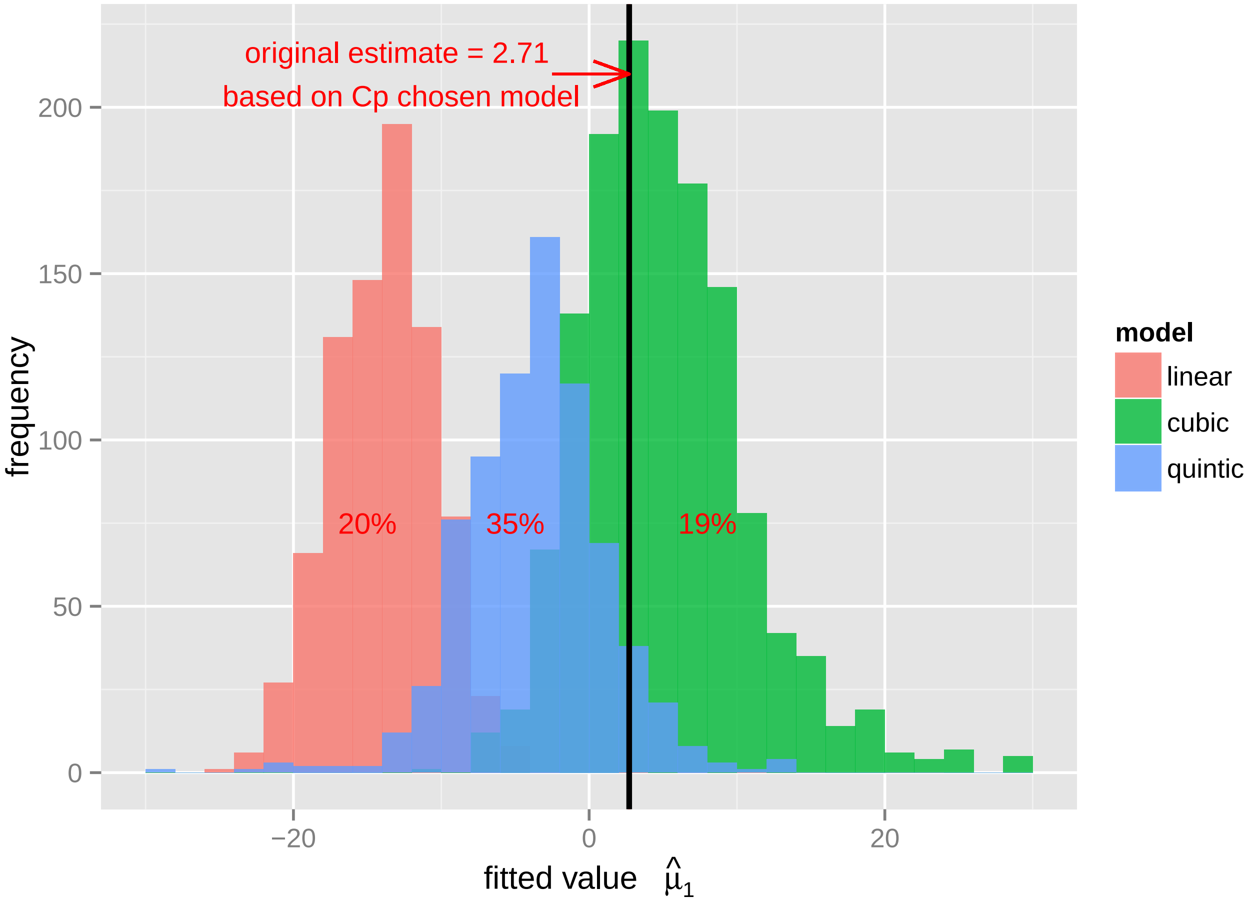


Figure 4: Fitted values for subject 1, from B=4000 nonparametric bootsrap replications separated by three most frequently chosen models by Cp

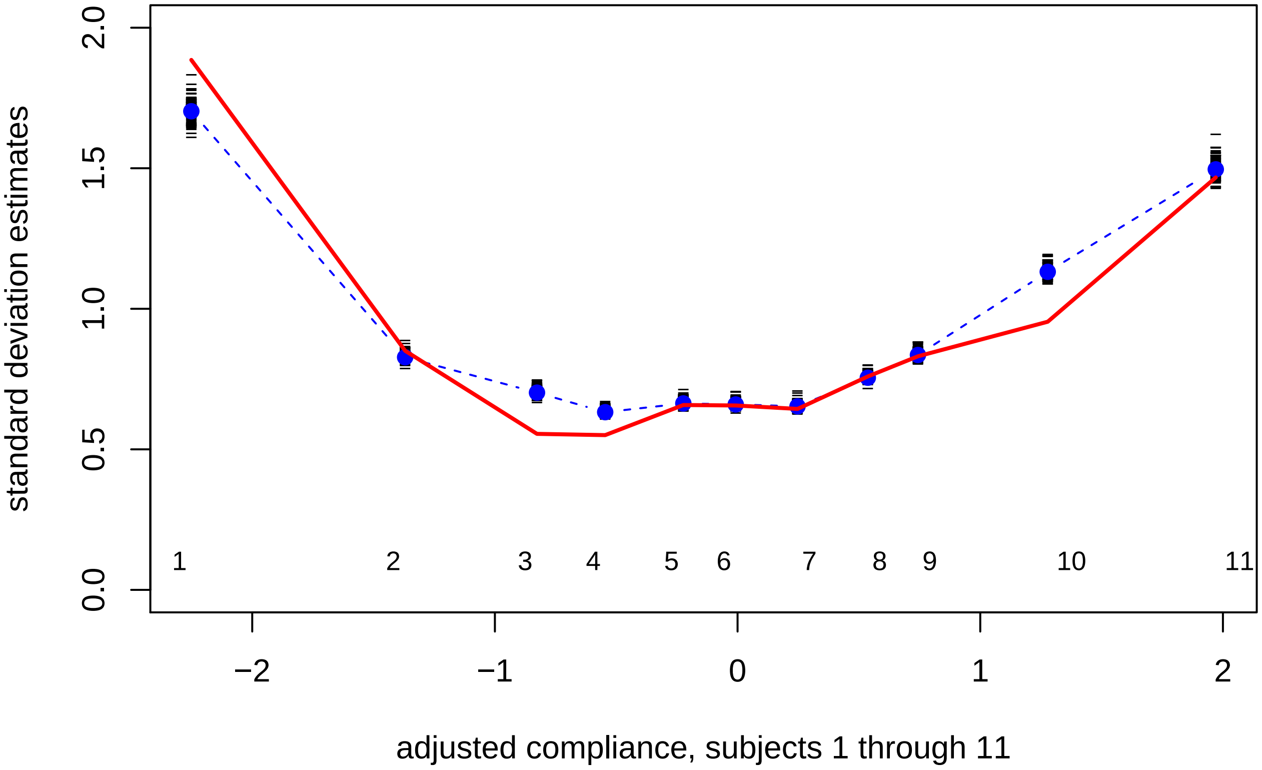


Figure 5: Simulation test of Theorem 2. Cholesterol data; 100 simulations, 1000 parametric bootstraps each, for the 11 subjects indicated at the bottom of Figure 1

	m1	m2	m3	m4	m5	m6
Mean	-13.87	-3.51	5.13	-1.68	-3.83	-3.54
St.dev.	3.45	3.40	5.87	5.80	5.61	8.83

Table 3: Mean and standard deviation of $\hat{\mu}_1^*$ as a function of the selected model. Some bootstrap samples that led to the quintic and sextic model being selected give a bad fit. Removing them and recomputing table 2 shows that the discrepancy between our table 2 and Efron's is due to bootstrap variability.

type	Interval	Length	Center.point
Standard	(-23.52, 28.94)	52.47	2.71
Percentile	(-18.33, 13.93)	32.26	-2.20
Smoothed	(-15.13, 9.29)	24.42	-2.92

Table 4: Three approximate 95% bootstrap confidence intervals for μ_1, the response for Subject 1, Cholesterol data

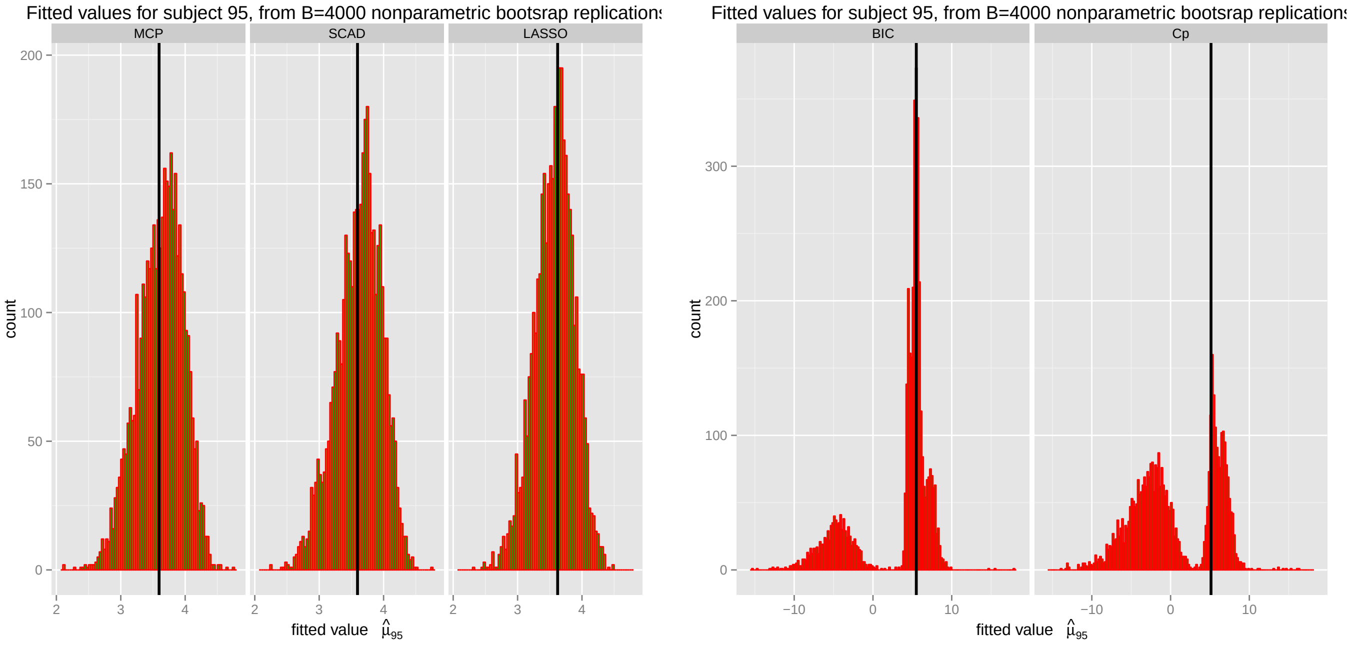
1.3 Accuracy of the Smoothed Bootstrap Estimates

1.4 Parametric Bootstrap Smoothing

2 Discussion

The main example in this paper was based on all subset selection in a linear regression context. The discussion by Wang, Sherwood and Li investigate the proposed method in a regularization procedure, a GLM, quantile and nonparametric regression. Overall, their numerical examples show that Efron's proposed smoothed estimator results in more accurate confidence intervals with good coverage probabilities. The rationale behind using L_1-norm type penalty functions for model selection as opposed to the all subset method used by Efron, is due to the well known result that all subset selection methods are unstable (Breiman, 1996a). Since these model selection procedures are driven by the data, it has been shown that even changing one point from the learning set, can result in a completely different chosen model (Breiman, 1996b).

We performed a similar analysis on the `prostate` data set (Stamey et al., 1989); a study that examined the relation between level of prostate specific antigen (PSA) and eight clinical measures e.g. Gleason score, cancer volume, prostate weight in 97 men who were about to receive prostatectomy. We produced $B = 4000$ bootstrap samples of the `prostate` data set. For each of the bootstrap samples, we performed model selection using all subset selection as well as LASSO (Tibshirani, 1994), SCAD Fan and Li (2001) and MCP Zhang (2010). For each model selection procedure, we calculated the fitted value for observation 95 based on the chosen model (resulting in 4000 fitted values for each of the 5 procedures). All the analysis was performed in `R` (R Core Team, 2013). The LASSO was implemented using the `glmnet` package (Friedman et al., 2010). SCAD and MCP were implemented using the coordinate descent algorithm in the `ncvreg` package (Breheny and Huang, 2011). The tuning parameter λ was chosen using 5 fold cross validation. Note, for the SCAD and MCP penalties, which have an additional tuning parameter, we used the suggested value of 3.7 and 3.0, respectively. All subset selection via the BIC and Cp criterion were implemented using the `leaps` package (Lumley and Miller, 2009). The histograms for the fitted values are given in Figure 6.



(a) L_1-Norm Penalty Functions

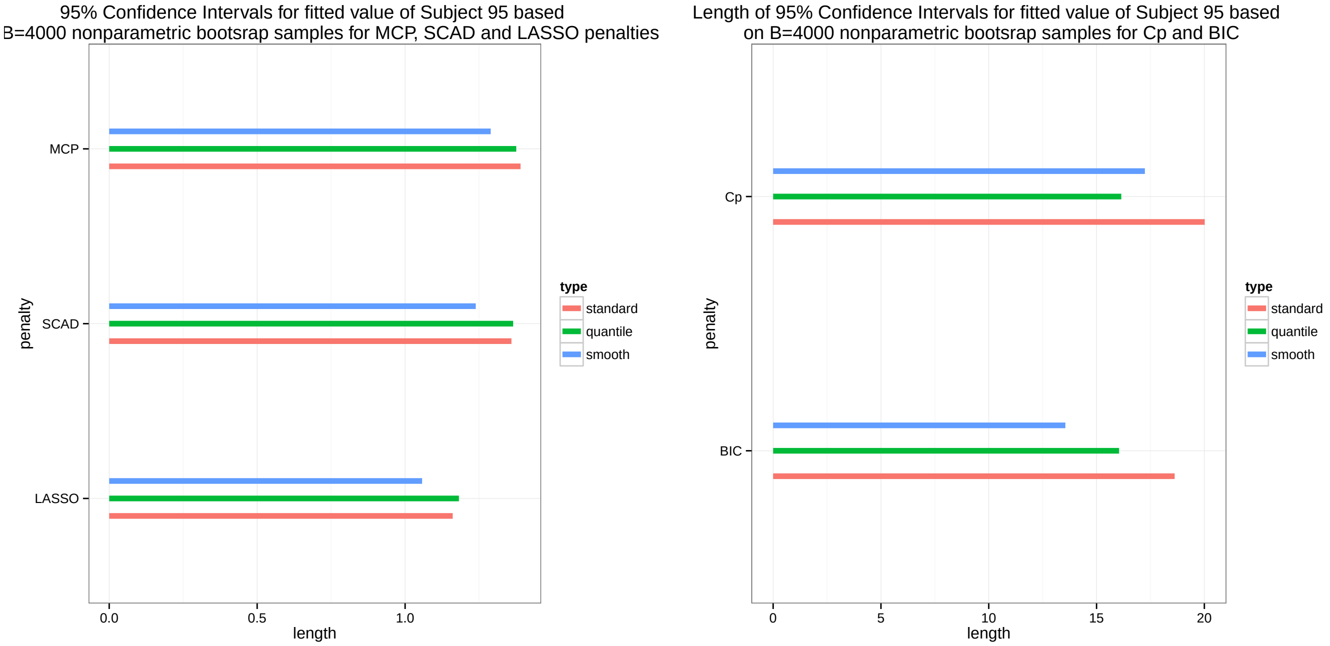
(b) All Subset Selection

Figure 6: Histogram of fitted values for subject 95 based on 4000 Bootstrap samples for the `prostate` data set

Figure 6b confirms the work of Breiman (1996a), i.e., the all subset selection procedure is very sensitive to changes in the data. The LASSO, SCAD and MCP all produce much more stable estimates, with the distributions of the fitted values looking normal and centred around their mean (Figure 6a).

Figure 7 compares the lengths of the three confidence intervals types for each procedure. We see that the new smoothed confidence intervals, based on the proposed smoothed standard deviation $\widehat{sd}_B$, outperforms (is shorter) the standard and quantile confidence intervals across all model selection procedures, with the LASSO providing the smallest intervals. Table 5 provides the numerical values of our analysis. Again, we see that the L_1-Norm penalties all perform similarly, with good coverage probabilities. Although the coverage probabilities are reasonable for the all subset methods, the length of the interval is much wider.

Wang, Sherwood and Li also show that when the columns of the predictor matrix are orthogonal and C_p, AIC or BIC are used as model selection criteria, an analytical solution to the asymptotic variance of the smoothed estimator can be derived . With a numerical example, they show that Efron's estimator performs well in this setting. Professor Politis points out that the proposed estimator does not apply to



(a) L_1-Norm Penalty Functions

(b) All Subset Selection

Figure 7: Length of confidence intervals for $\hat{\mu}_{95}$ based on 4000 Bootstrap samples for the `prostate` data set

the residual bootstrap because this presupposes a choice of the model. He then gives a summary of his own work on model-free prediction as an alternative approach. Gupta and Lahiri give show two alternative methods for constructing confidence intervals; the Adaptive LASSO (Zou, 2006) and maximum frequency Bootstrap-t (MF). The Adaptive LASSO has the oracle property, i.e., it performs as well as if the true underlying model were given in advance and has been shown. The maximum frequency Bootstrap-t limits the calculation of the bagged estimator and its standard error to those resamples that led to the most chosen model. While Gupta and Lahiri's simulations show good performance of this approach, it does not do well in the cholesterol example; MF 95% CI [-5.93,15.35] compared to the smoothed interval [-13.3, 8.0]. This is because a substantial number of bootstrap resamples, which led to more negative predicted values are being ignored by the MF.

Table 5: Prostate data, B=4000, Observation 95

model	type	fitted value	sd	length	coverage
LASSO	standard	3.62	0.31	1.21	0.94
	quantile			1.20	0.95
	smooth	3.57	0.29	1.14	0.93
SCAD	standard	3.60	0.35	1.37	0.95
	quantile			1.33	0.95
	smooth	3.62	0.33	1.28	0.93
MCP	standard	3.60	0.35	1.38	0.96
	quantile			1.35	0.95
	smooth	3.61	0.33	1.29	0.94
BIC	standard	5.50	4.75	18.62	0.84
	quantile			16.05	0.95
	smooth	3.22	3.46	13.55	0.83
Cp	standard	5.13	5.11	20.02	0.86
	quantile			16.15	0.95
	smooth	0.64	4.40	17.24	0.97

References

P. Breheny and J. Huang. Coordinate descent algorithms for nonconvex penalized regression, with applications to biologival feature selection. *Ann Appl Stat*, 5(1):232–253, Jan 2011. 6

Leo Breiman. Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6):2350–2383, 12 1996a. doi: 10.1214/aos/1032181158. URL http://dx.doi.org/10.1214/aos/1032181158. 6, 7

Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996b. ISSN 0885-6125. doi: 10.1023/A:1018054314350. 6

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001. doi: 10.1198/016214501753382273. URL http://www.tandfonline.com/doi/abs/10.1198/016214501753382273. 6

J. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw*, 33(1):1–22, 2010. 6

Thomas Lumley and Alan Miller. leaps: regression subset selection. 2009. URL http://CRAN.R-project.org/package=leaps. R package version 2.9. 7

R Core Team. R: A language and environment for statistical computing. 2013. URL http://www.R-project.org/. 6

T. A. Stamey, J. N. Kabalin, J. E. McNeal, I. M. Johnstone, F. Freiha, E. A. Redwine, and N. Yang. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients. *J. Urol.*, 141(5):1076–1083, May 1989. 6

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288, 1994. 6

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 04 2010. doi: 10.1214/09-AOS729. URL http://dx.doi.org/10.1214/09-AOS729. 6

Hui Zou. The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101 (476):1418–1429, 2006. 8