Combining variable selection and shrinkage to derive a multivariable regression model

Willi Sauerbrei[1], Edwin Kipruto[1], Georg Heinze[2] for TG2 of the STRATOS initiative

[1] Institute for Medical Biometry and Statistics, Faculty of Medicine and Medical Center University of Freiburg, Freiburg, Germany

[2] Medical University of Vienna,Austria

Overview

- TG2 - Selection of variables and functional forms
  - ➢ 7 methodological issues identified
- Variable selection strategies
  1. Traditional strategies
  2. Further strategies
  3. Penalized likelihood
- Bias and the role of shrinkage
  1. Nonnegative Garotte
  2. Lasso and extensions
- Conclusions

General assumption – sample size is ‘acceptable‘

TG2: Overview paper

BioMed Central
The Open Access Publisher

Diagnostic and Prognostic Research

Diagn Progn Res. 2020; 4: 3.
Published online 2020 Apr 2. doi: 10.1186/s41512-020-00074-3

PMCID: PMC7114804
PMID: 32266321

State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues

Willi Sauerbrei,[1] Aris Perperoglou,[2] Matthias Schmid,[3] Michal Abrahamowicz,[4] Heiko Becher,[5] Harald Binder,[1] Daniela Dunkler,[6] Frank E. Harrell, Jr,[7] Patrick Royston,[8] Georg Heinze,[6] and for TG2 of the STRATOS initiative

- 7 methodological issues identified

Towards state of the art– research required!

1. Investigation and comparison of the properties of **variable selection strategies**
2. Comparison of **spline procedures** in both univariable and multivariable contexts
3. How to model one or more variables with a ‚**spike-at-zero**‘?
4. Comparison of **multivariable procedures for model and function selection**
5. Role **of shrinkage** to correct for bias introduced by data-dependent modelling
6. Evaluation of new approaches for **post-selection inference**
7. Adaptation of procedures for **very large sample sizes** needed?

Selection of variables

- Central issues:
  - Model with focus on prediction (TG6) or description (TG2)?
  - To select or not to select (full model)?
  - Which variables to include?
- A large number of methods proposed (for many decades)
- High-dimensional data (HDD) triggered the development of further proposals
  - HDD - prediction is the main aim (TG9)
- Many critical issues, state of the art?

Traditional variable selection strategies

- Full model
  - Variance inflation in case of multicollinearity
- Stepwise procedures
  - Forward Selection (FS)
  - Stepwise Selection (StS)
  - Backward Elimination (BE)
  - Which stopping criteria (AIC, BIC, p-value)?
    - ➢ Has a severe influence on complexity of model selected
- All subset selection
  - which criteria (AIC, BIC)? Or variants of it?

Other procedures

- Procedures based on ‘change-in-estimate’
- Resampling-based variable selection procedures
- Bayesian approaches
- Modern variable selection strategies
  - Boosting
  - Penalised likelihood
    - Nonnegative garrote
    - Lasso (Extensions: Adaptive Lasso, Relaxed Lasso, etc.)
    - Elastic net
    - Smoothly Clipped Absolute Deviation (SCAD)

Data dependent model-building introduces biases

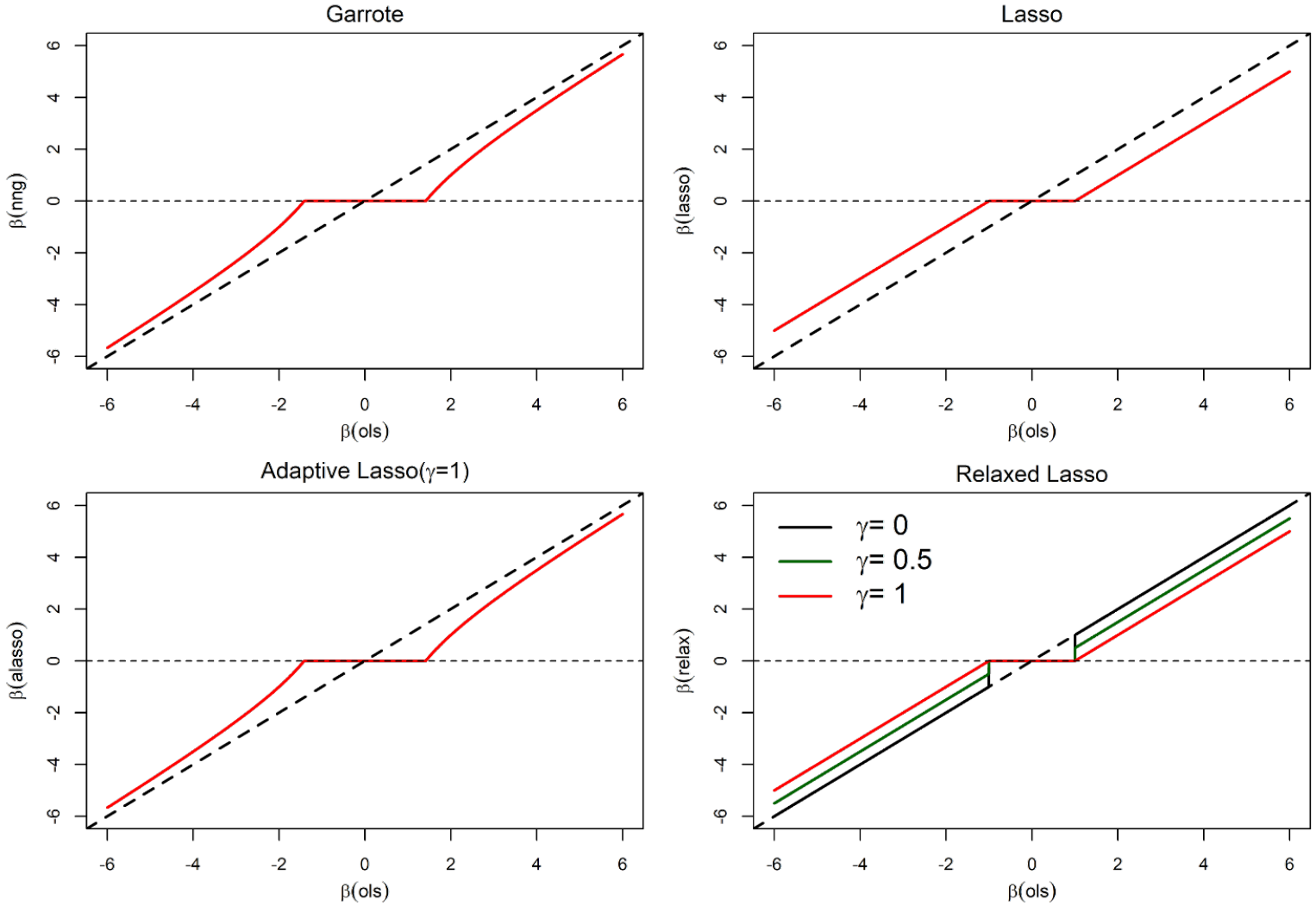
- Bias and the role of shrinkage methods
  - Several modern selection procedures combine variable selection and shrinkage to correct for the bias.
  - Post-estimation shrinkage (2 step approach) can be used for many types of models.

    Step 1: Select a model

    Step 2: Use leave-one-out (or other resampling technique) to estimate parameterwise shrinkage factors

Data dependent model-building introduces biases

- Combine variable selection and shrinkage



Method	Effects	
	Large	Small
NNG	Hardly	Severe
Lasso	Equal amount	Equal amount
Alasso	Hardly	Severe
Rlasso	Equal amount	Equal amount
	$\gamma = 1$ Lasso $\gamma = 0$ No shrinkage $\gamma = 0.5$ Less shrinkage than lasso	

Amount of shrinkage

Nonnegative garrote - initial estimates

- Breiman (1995) proposed **OLS** as initial estimates
- Problematic for **strongly correlated data** and not usable in **high dimensional data**
- Yuan and Lin (2007) proposed **ridge**, **lasso** and other initial estimates
- For ridge or lasso - which penalty parameter λ? Optimal or larger?

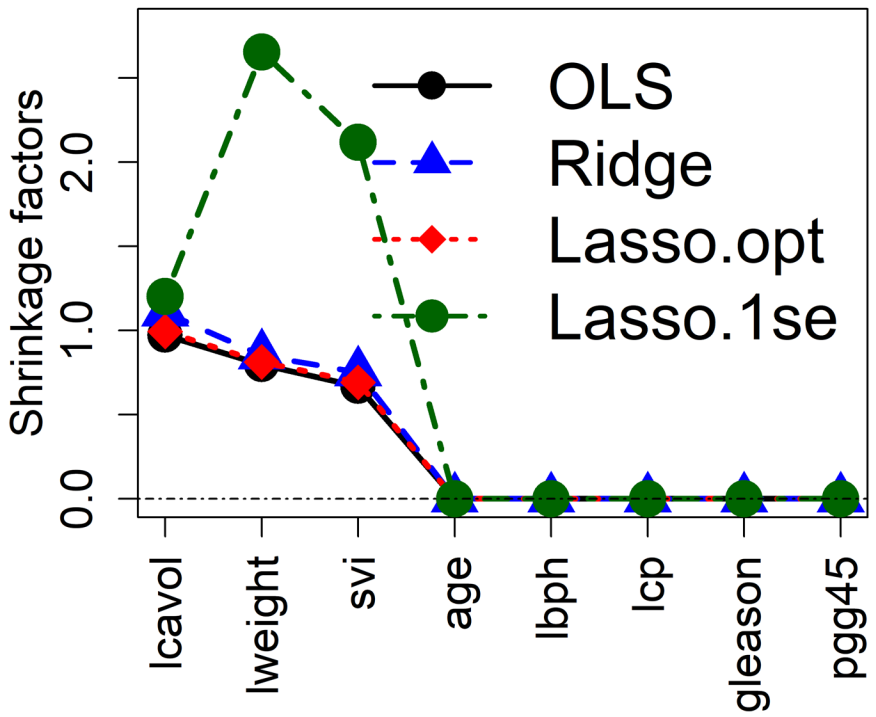
Prostate data (n = 97, p = 8 variables), linear regression model

Predictor	**OLS**	**Ridge(opt)**	**Lasso(opt)**	**Lasso(λ_{1se})**
Lcavol	0.662	0.577	0.647	**0.517**
lweight	0.265	0.257	0.260	**0.104**
svi	0.314	0.282	0.299	**0.126**
age	-0.157	-0.124	-0.143	0.000
lbph	0.140	0.124	0.132	0.000
lcp	-0.148	-0.055	-0.113	0.000
gleason	0.035	0.046	0.030	0.000
Pgg45	0.125	0.096	0.112	0.000
#Variables	8	8	8	3
R^2	0.663	0.659	0.663	**0.561**

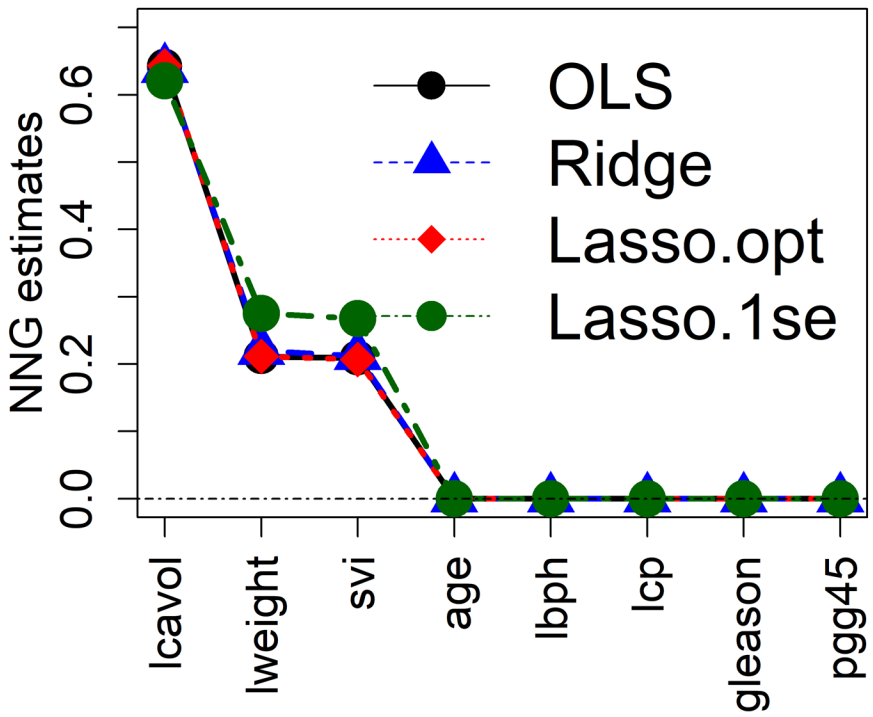
⇦ Overshrunken initial estimate

Variables eliminated
(Variable screening)

Nonnegative garrote - initial estimates and shrinkage factors



- NNG eliminates some variables and corrects for overshrinkage



- NNG results (selected variables, parameter estimates) are very similar for all initial estimates

Combine variable selection and shrinkage

- Tuning parameter play a key role
- Lasso is popular for high dimensional data but suffers from overshrinkage of large effects
- Adaptive lasso and relaxed lasso were proposed to correct for overshrinkage
- NNG can be used for correlated and high dimensional data
  - ➢ Further investigations in such data showed promising results

Conclusion

We are far away from ‘state of the art’ on selection of variables and functional forms

Many more comparisons are urgently needed!

> *‘Exact distributional results are virtually impossible to obtain, even for simplest of common subset selection algorithms’*
>
> *Picard & Cook, JASA, 1984*

➞ **Informative simulation studies are needed!**

… Conclusions

- Member of TG2 identified seven issues
- Other experts may have different experiences and preferences

  … and may raise further issues

- To help deriving evidence-supported guidance, more cooperative and comparative research is needed from experts

References

Breiman, L., 1995. Better subset regression using the nonnegative garrote. Technometrics, 37(4), pp.373-384.

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

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

Meinshausen, N., 2007. Relaxed lasso. Computational Statistics & Data Analysis, 52(1), pp.374-393.

Sauerbrei, W., Perperoglou, A., Schmid, M., Abrahamowicz, M., Becher, H., Binder, H., Dunkler, D., Harrell, F.E., Royston, P., Heinze, G. and TG2 of the STRATOS initiative, 2020. State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. Diagnostic and prognostic research, 4, pp.1-18.

Thanks to all members of TG2 !

- Georg Heinze (Austria)
- Aris Perperoglou (U.K.)
- Willi Sauerbrei (Germany)
- Michal Abrahamowicz (Canada)
- Heiko Becher (Germany)
- Harald Binder (Germany)
- Thomas Cowling (U.K.)
- Daniela Dunkler (Austria)
- Rolf Groenwold (Netherlands)
- Frank Harrell (U.S.A)
- Nadja Klein (Germany)
- Geraldine Rauch (Germany)
- Patrick Royston (U.K.)
- Matthias Schmid (Germany)

And the early career adjunct members

- Michael Kammer (Vienna, Austria)
- Edwin Kipruto (Freiburg, Germany)
- Christine Wallisch (Vienna, Austria)