Discussion: “Illusions of Sparsity by Giorgio Primiceri”

Pablo Guerron-Quintana

Boston College

June, 2018

Executive Summary

Motivation: Is regular coke (dense model) better than diet coke (sparse model)?

Figure: Sparse vs Dense

Executive Summary

Motivation: Should we use dense or sparse models?

Giorgio and coauthors look for answers to this key question.

To this end,

- Propose a flexible Bayesian model encompassing competing alternatives.
- Use macro, finance, and micro data.

Main message: Sparsity should not be taken for granted.

Only in one application (out of 6) sparsity emerges from data under uninformative priors.

Into the woods I

- Consider regression:

$$y_t = u_t'\phi + x_t'\beta + \epsilon_t,$$

where parameter of interest is vector β.

- Impose prior:

$$\beta = \begin{cases} \mathcal{N}(0, \sigma^2\gamma^2) & \text{with prob. } q \\ 0 & \text{with prob. } 1-q \end{cases}$$

- q determines whether you are in a Ridge world or a Lasso world.
- If in Ridge world, γ controls degree of shrinkage.
- Operationally, $q \sim Beta(a, b)$ and $R^2(\gamma^2, q) \sim Beta(A, B)$.

Into the woods I

Eyeball econometrics points to these modes for sparsity and shrinkage:

	q	γ
Macro I	$0.2 \sim 0.3$	0.135
Macro II	$0.9 \sim 1.0$	0.174
Financ I	$0.5 \sim 1.0$	0.174
Financ II	0.6	0.007
Micro I	0.0	$0.37 \sim 1.0$
Micro II	$0.5 \sim 0.6$	0.37

Take away:

- Only Micro I (decline in crime rates) clearly shows sparsity.
- Other applications prefer mixtures $q \in (0, 1]$
- But with significant shrinkage $\gamma >> 0$

Into the woods II

Consider simple regression:[1]

$$Y = \mu + v, \quad v \sim N(0, \sigma^2)$$

- Lasso's shrinkage function: $d^\ell(y) = \max(|y| - \frac{\lambda^\ell}{2}, 0)sign(y)$
- Ridge's shrinkage function: $d^r(y) = \frac{y}{1+\lambda^r}$

[1] inspired by Chernozhurov et al. Annals of Statistics, 2015

Into the woods II

Consider simple regression:[1]

$$Y = \mu + v, \quad v \sim N(0, \sigma^2)$$

- Lasso's shrinkage function: $d^\ell(y) = \max(|y| - \frac{\lambda^\ell}{2}, 0) sign(y)$
- Ridge's shrinkage function: $d^r(y) = \frac{y}{1+\lambda^r}$
- A poor man's shrinkage function for Giorgio's paper is

$$d^{gp}(Y) = q \times d^r(y) + (1-q) \times d^\ell(y)$$

- Or in terms of ℓ_1 and ℓ_2 penalizations, Giorgio's proposal is

$$|\mu|^{gp} = q \times |\mu| + (1-q) \times \gamma^2 \times |\mu|^2$$

[1] inspired by Chernozhurov et al. Annals of Statistics, 2015

Into the woods III

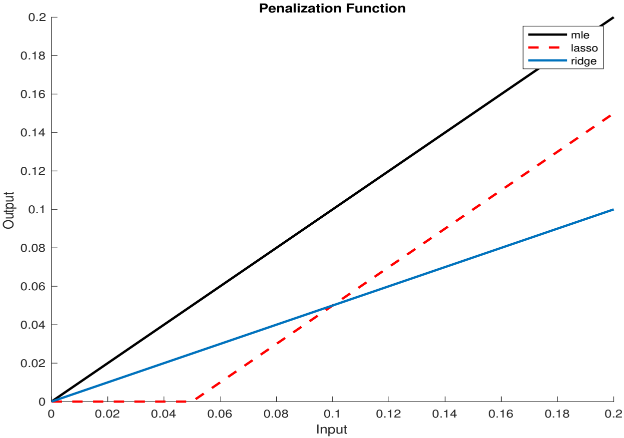

Figure: Lasso and Ridge Shrinkage Functions

Into the woods IV

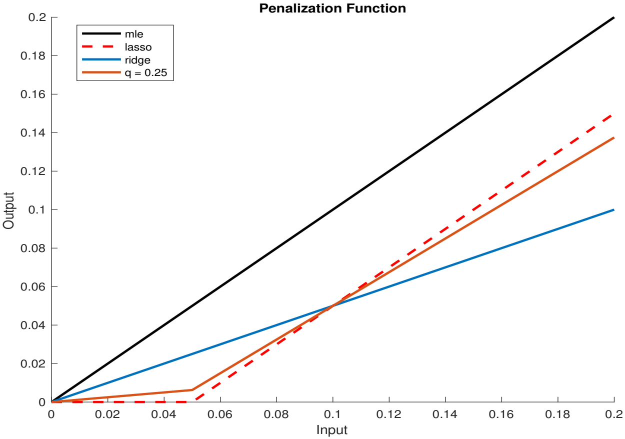

Figure: Lasso, Ridge, GP Shrinkage Functions

Into the woods IV

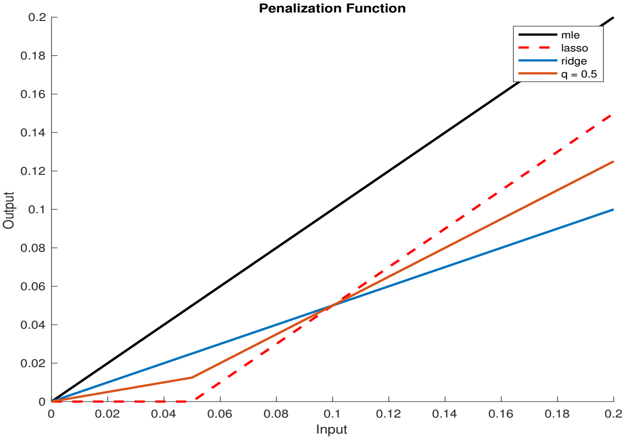

Figure: Lasso, Ridge, GP Shrinkage Functions

Into the woods IV

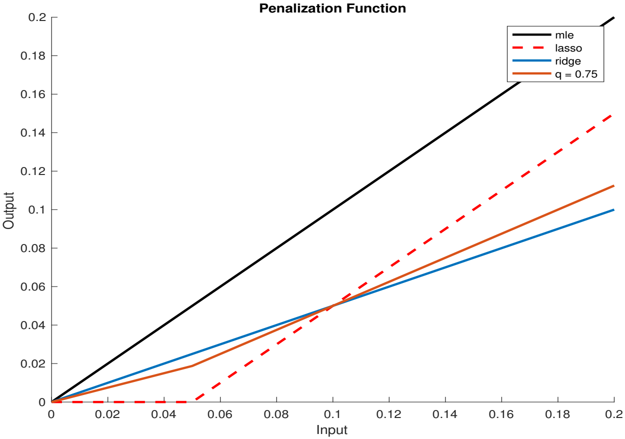

Figure: Lasso, Ridge, GP Shrinkage Functions

Into the woods IV

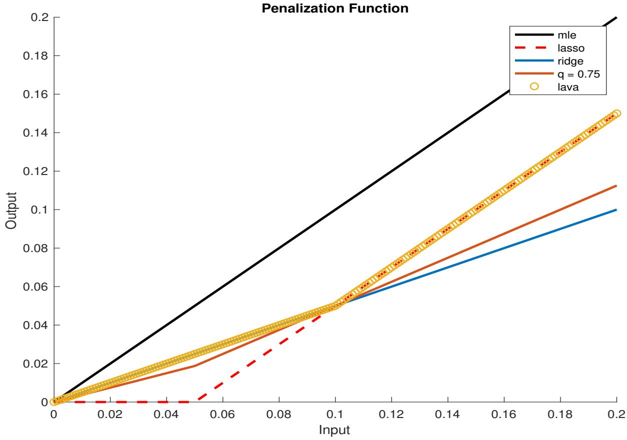

Figure: Lasso, Ridge, GP, and Lava Shrinkage Functions

Insights/Comments I

- Let's take U.S. macro forecasting and U.S. firms stock returns applications.
- Samples cover
  - → **Macro :** $1960:2 - 2014:12$
  - → **Finance:** $1963:7 - 2015:6$
- Sample covers a lot of Great episodes!
  1. Great Moderation,
  2. Great Recession (aka Financial crisis),
  3. Secular Stagnation (post-crisis era).

Insights/Comments II

- Why is this a concerned?

  1. changes in volatility,
  2. emergence of new factor post financial crisis,
  3. zero lower bound,
  4. a recent break in growth rates (?)
  5. more generally, some evidence that economic complexity has increased since the 1970s.

- It will be informative to discuss whether sparsity has changed (and if so, why).

Insights/Comments III

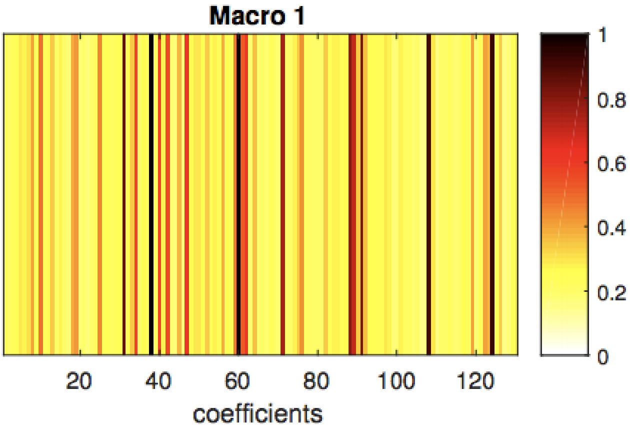

Figure: Heat map probability of inclusion of each predictor Macro I application

Insights/Comments IV

- There are between 5 and 8 “dominant” regressors.

Insights/Comments III

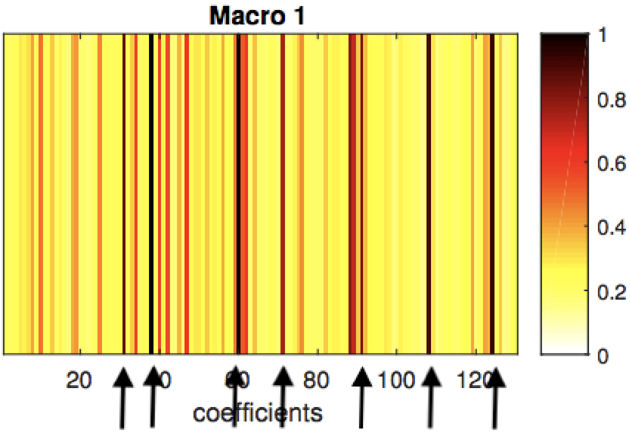

Figure: Heat map probability of inclusion of each predictor Macro I application

Insights/Comments IV

- There are between 5 and 8 “dominant” regressors.
- How well do these regressors forecast **industrial production**?
- Valuable to know if in hurry to get quick prediction.

Insights/Comments IV

- There are between 5 and 8 “dominant” regressors.
- How well do these regressors forecast **industrial production**?
- Valuable to know if in hurry to get quick prediction.
- Related, part of Lasso literature is about efficient algorithms.
- How computationally expensive is the proposed approach?
- Treatment of industrial production data
  - Timing of release: IP March 23 while NIPA March 29.
  - Revised versus real time data.