# Regularizing Bayesian Predictive Regressions

Guanhao Feng

Booth School of Business, University of Chicago

R/Finance 2017

(Joint work with Nicholas Polson)

## What do we study?

- A **Bayesian predictive regression** requires prior information, motivated by economic theory or constraints.

- **Regularization methods** address the bias-variance tradeoff in high dimensional regression problems of the model complexity.

- A **duality** between Bayesian predictive regressions and regularization, which leads to a framework of **prior sensitivity analysis** through the regularization path of the predictors.

## Motivation: VAR(1) regularization

- A VAR(1) model with a demeaned vector $Z_t$

$$Z_t = \beta Z_{t-1} + \epsilon_t, \text{ where } B = \text{Vec}(\beta)$$

$$\text{Cov}(\epsilon_t) = \Sigma$$

- Regularization requires a measure of model goodness of fit $l(B, \Sigma)$ and a penalty function $\phi(B, \Sigma)$.

- A regularization estimation leads to such an optimization problem

$$\min_{B, \Sigma \in \Re^d} l(B, \Sigma) + \phi(B, \Sigma).$$

## Bayesian MAP (Maximum a Posteriori)

- Probabilistically, $l(B, \Sigma)$ and $\phi(B, \Sigma)$ correspond to the negative logarithms of the likelihood and a prior distribution.
- A probabilistic approach leads to a Bayesian hierarchical model

$$p(y \mid B, \Sigma) \propto \exp\{-l(B, \Sigma)\},$$

$$p(B, \Sigma) \propto \exp\{-\phi(B, \Sigma)\}.$$

- The solution to the minimization problem corresponds to maximizing the posterior density,

$$p(B, \Sigma \mid y) \propto p(y \mid B, \Sigma) \times p(B, \Sigma) = \exp\{-l(B, \Sigma) - \phi(B, \Sigma)\}$$

$$(\hat{B}, \hat{\Sigma}) = \operatorname*{argmax}_{B,\Sigma} \; p(B, \Sigma \mid y)$$

## Intuition: Prior Sensitivity Analysis

- If $\phi(B,\Sigma) = \lambda\phi_1(B) + \gamma\phi_2(\Sigma)$, then $\lambda$ and $\gamma$ are hyper-parameters of a prior distribution and tuning parameters in a regularization problem.
- A Bayesian study requires a full prior specification with $(\lambda,\gamma)$, but regularization problem uses $(\lambda,\gamma)$ to control the bias-variance tradeoff by maximizing out-of-sample forecasting power.
- The regularized estimates $(\hat{B},\hat{\Sigma})_{(\lambda,\gamma)}$ provide a regularization path, which can be interpreted as a prior sensitivity analysis for the MAP forecast.

## Data-Driven Selection

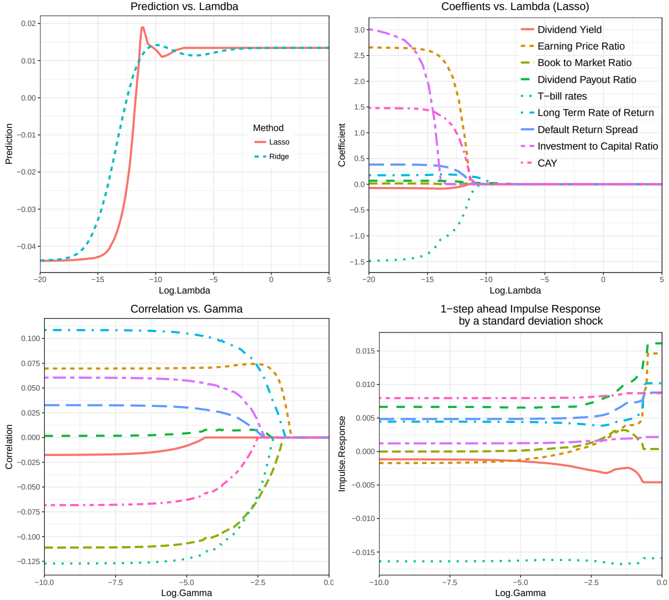
- We use cross-validation, AIC and BIC for model selection criteria.
- Search the optimal pair of $(\lambda, \gamma)$ for the prior specification to maximize the out-of-sample predictive performance.
- Graphically display the prior sensitivity analysis through the regularization path of a wide range of $(\lambda, \gamma)$.
- Additional insights using shrinkage priors include high-dimensional predictor selection and a interpretable sparse variance-covariance matrix.

# Quarterly Predictors from Goyal and Welch (2008)

We use quarterly data from 1952 to 2015 to forecast excess return of S&P 500 index in the regularized VAR(1) in a 10-dimensional model.

| Predictor | Description |
|---|---|
| Dividend Yield | Difference between the log of dividends and the log of lagged prices. |
| Earning Price Ratio | Ratio of earnings to prices. |
| Book to Market Ratio | Ratio of book value to market value for the Dow Jones Industrial Average. |
| Dividend Payout Ratio | Ratio of dividends to earnings. |
| T-bill rates | 3- Month Treasury Bill |
| Long Term Rate of Return | Long term yield on government bonds. |
| Default Return Spread | Difference between long-term corporate bond and long-term government bond returns. |
| Investment to Capital Ratio | Ratio of aggregate investment to aggregate capital for the whole economy. |
| Consumption, wealth, income ratio | CAY, see Lettau and Ludvigson (2001). |

# Regularizing Equity Premium Predictors



# Market Timing Strategy