Discussion of

Carriero, Clark, Marcellino Real Time Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility

and

Kim and Swanson, Mining Big Data Using Parsimonious Factor and Shrinkage Methods

By **Marek Jarociński**

European Central Bank, DGR/MPR

ECB Big Data Workshop, Frankfurt, 7 April 2014

What these papers do

- These papers forecast Y using a regression model

$$Y(t+h) = Z(t)\ b + e(t+h)$$

- Z is a (function of) potentially large set of predictors

- These papers experiment with imaginative ways to estimate the above equation in order to obtain good forecasts of Y

A summary of Carriero, Clark, Marcellino

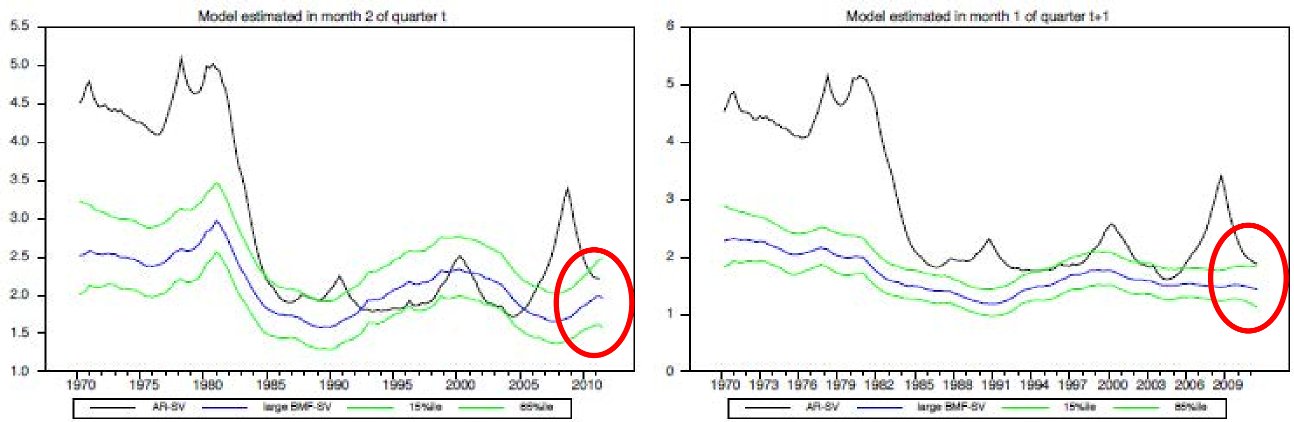
$$Y(t+h) = Z(t)\, b + e(t+h)$$

- Challenge: **Ragged edge -** the set of available monthly indicators is <u>different in each month </u>of the quarter.
- The usual solution: specifying a model that handles missing values, but this is cumbersome.
- What they do: They use three separate regressions:
  1. for the *first* month of the quarter (Jan,Apr,Jul,Oct)
  2. for the *second* month of the quarter (Feb,May,Aug,Nov)
  3. for the *third* month of the quarter (Mar,Jun,Sep,Dec)
  - Bayesian shrinkage of the regression coefficients, stochastic volatility in e

Comments on Carriero, Clark, Marcellino

- Their idea of dealing with the ragged edge is **simple**
  – simple solutions often work best in forecasting!
  – With the ragged edge out of the way, they can focus on more important things, like stochastic volatility
- **Price** of the simplicity:
  – No consistency imposed between the model used e.g. in January and the model used e.g. in February – although these are models of the same quantity (GDP growth in the first quarter) and using the same type of indicators.

Example of the consistency issue: different stochastic volatility processes in January and in February



- The `February' model (left panel) thinks the volatility of GDP *increased* in 2010.
- The `January' model (right panel) thinks it *slightly decreased* in 2010.
- Scope for imposing consistency? Bayesian shrinkage also *across* equations?

Summary of Carriero, Clark, Marcellino

- The paper uses a simple, practical approach to deal with the ragged edge
- The price of simplicity: their approach imposes no consistency between the GDP model in January and in February.
- Their predictive performance evaluations suggest that this price is worth paying

Summary of Kim and Swanson

$$Y(t+h) = Z(t)\, b + e(t+h)$$

- Z = [W, F] where W – observable and F – unobservable
- Estimation of F from a large dataset X
  - Using PCA, ICA, SPCA
- Estimation of the parameters b
  - Using OLS, ridge, BMA, …

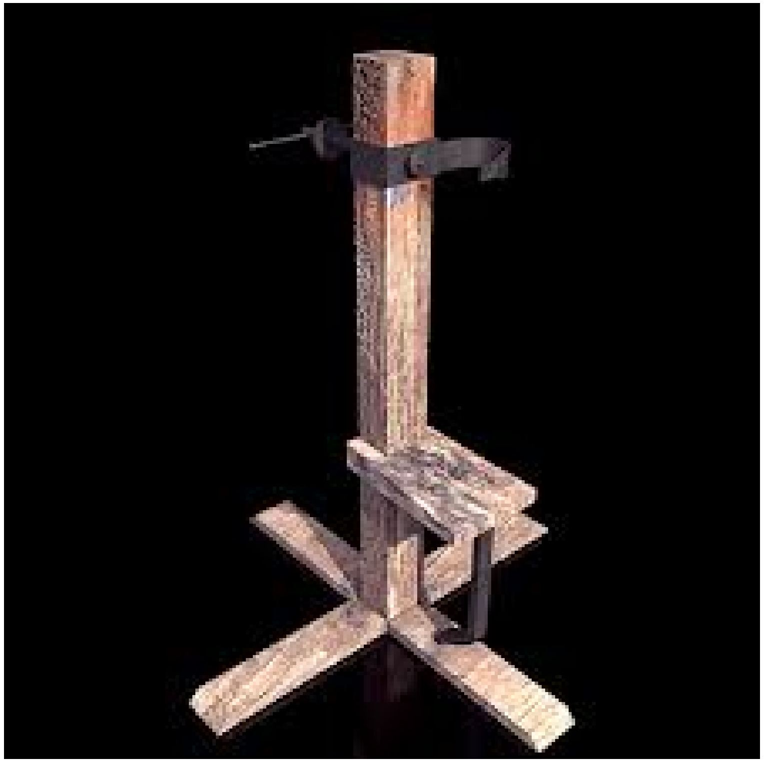
boosting, …

bagging, …

lasso (least angle regression), ...

elastic net, …

and non-negative garrote.



$$\geq 0$$

Findings of Kim and Swanson

- Simple averaging (1/N) does not win.
- Different models win for different variables and horizons. THERE IS NO PATTERN.

For example:

- To forecast GDP 1 period ahead:
  - **Boost** the **ICA** using the **rolling sample**
- To forecast GDP 3 periods ahead:
  - **LASSO** the **PCA** using the **recursive sample**

Comments on Kim and Swanson

- Simple averaging (1/N) does not win.
- Different models win for different variables and horizons. THERE IS NO PATTERN.

  – **Perhaps different models win by chance?**

    -> check if the difference between best and e.g. 1/N is **statistically significant**

  – **Perhaps there is a pattern after all?**

    -> **Organize** the models along some meaningful dimensions and uncover the pattern

A Bayesian perspective on all these procedures

$$Y(t+h) = X(t)\, b + e(t+h)$$

X - large dataset

Shrinkage in b comes from two sources

- Extraction of factors from X
- Nonstandard estimation of the coefficients b

Two questions

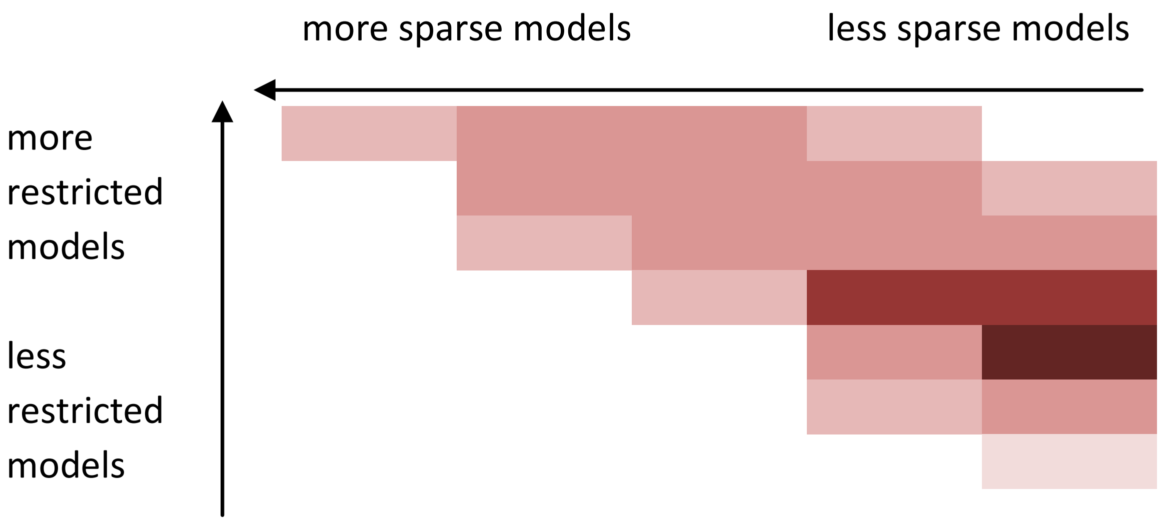
1. **How much to restrict b?**
   – More restricted b – worse fit in-sample, but maybe better forecast out-of-sample
2. **Few variables with large coefficients or many variables with small coefficients?**
   – few variables – more volatile forecasts, many variables – useful signals may get lost

Two questions

1. **How much to restrict b? ←** prior **variance**
   – More restricted b – worse fit in-sample, but maybe better forecast out-of-sample
2. **Few variables with large coefficients or many variables with small coefficients? ←** prior **kurtosis**
   – few variables – more volatile forecasts, many variables – useful signals may get lost

These two questions can serve to organize models

- Example from my work*



* Jarociński, Cross-country growth regressions with Bayesian shrinkage (forthcoming in EER)

Summary

- Both papers are very useful, though in different ways
- Carriero, Clark, Marcellino propose a practical way of dealing with the ragged edge problem.
- Kim and Swanson carry out a large-scale forecasting horse-race and find thought-provoking results