Forecasting with Bayesian Global Vector Autoregressive Models: A Comparison of Priors

Jesús Crespo Cuaresma, Martin Feldkircher and Florian Huber

Discussion by Marta Bańbura*

European Central Bank

8th ECB Workshop on Forecasting Techniques
Frankfurt am Main, 13-14 June 2014

* The views expressed are those of the author and do not necessarily reflect those of the ECB.

Summary of the paper

- Bayesian estimation in the GVAR framework:

$$x_{it} = a_{i0} + a_{i1}t + \Phi_i x_{i,t-1} + \Lambda_{i0} x_{it}^* + \Lambda_{i1} x_{i,t-1}^* + \pi_{i0} d_t + \pi_{i1} d_{t-1} + \epsilon_{it}$$

  i: refer to countries; $*$ refer to global variables

- Different priors on the coeffcients in the individual country models:

$$\Psi_i = (a_{i0}\ a_{i1}\ vec(\Phi_i)'\ vec(\Lambda_{i0})'\ vec(\Lambda_{i1})'\ vec(\pi_{i0})'\ vec(\pi_{i1})')'$$

  Minnesota prior, normal inverted Wishart, independent normal inverted Wishart, stochastic search variable selection (SSVS) prior

- Forecast evaluation over 2009-2012

$\Rightarrow$ Bayesian estimation helps, the best performing setup is the SSVS.

Shrinkage and forecasting

Shrinkage helps to improve forecast accuracy:

- Not a new result, Litterman (1986)
- Recently, increasingly popular in view of availability of large data sets

Different approaches to shrinkage:

- Factor models and principal components (data shrinkage, dimension reduction)
- Variable selection
- Regularisation methods
- Bayesian priors
- Statistical learning (bagging and boosting)
- ...

- GVARs - particular case of data shrinkage/dimension reduction

Shrinkage and forecasting

Importance of shrinkage have been established, however no "best strategy"

- Many applications find comparable performance of different strategies; otherwise the ordering could be dependent on the particular setup or a data set

  De Mol, Giannone and Reichlin (2008), Stock and Watson (2012), Kadiyala and Karlsson (1997), Carriero, Clark and Marcellino (2011), Koop (2013),

- Some methods are asymptotically equivalent under certain assumptions on the data generating process

  De Mol, Giannone and Reichlin (2008), Chudik and Pesaran (2011,2014)

- Recently, Kim and Swanson (2014) advocate "hybrid shrinkage" - combining factor extraction with other shrinkage techniques.

Shrinkage and forecasting in a GVAR

- A type of “hybrid” approach: data dimension reduction via the global variables + Bayesian priors

  Novel approach in the GVAR framework; most of the applications on shrinkage for single country models

- Stochastic search variable selection a clear winner among the priors

  $\Rightarrow$ Different from the results in Koop (2013) where there is no clear “ordering”.

  Could this be due to the specific design in this paper?

General remarks about the choice of priors

Some clarity would help on

- How do various priors differ in the degree of shrinkage and how is the latter chosen?
- What is the “underlying” prior model (and why)?

Some questions on the priors

1. Natural conjugate priors
   Why is the prior mean 0 (prior expectation model: white noise, different than in Minnesota type)? Why is the prior on the coefficients diffuse?

2. Minnesota prior
   How are the hyperparameters chosen? Why other variable lags are not shrunk more? Why is the prior on the coefficients for the global variables diffuse? Prior expectation model: random walk with intercept, deterministic trend and global variables?

3. Single unit root prior
   Perhaps better to call it “dummy initial observation” prior. Why is it governed by two hyperparameters? How are they chosen? Is this the only component of the prior (no further shrinkage is imposed)?

4. SSVS
   What is the “prior model”?

Some suggestions

- Choice of the degree of shrinkage
  - Elaborate on the selection of the degree of shrinkage e.g. use some criteria for the shrinkage selection (e.g. in-sample fit, marginal likelihood or hierarchical approach)
  - Try to make the shrinkage comparable, see e.g. Koop (2013) on the relation of shrinkage between the Minnesota type and SSVS priors; shrink also in the natural conjugate setp.
- Use comparable prior expectations
- Try dropping the deterministic trend and use the designs with random walk with drift “prior model”

This would help to understand which particular elements of the priors matter for forecast performance.

Further issues

- Number of lags
  Is it optimal to keep 1?
- What is exactly the “traditional cointegrated GVAR (Pesaran)”?
  Pesaran, Schuermann and Smith (2009) report substantial forecast accuracy gains when averaging over estimation windows and GVAR specification. Perhaps a worthwhile robustness check.
- Are there any patterns in the inclusion probabilities in the SSVS?
  E.g. What are typically inclusion probabilities for the global variables?
- What is the performance of “closed” country models with Bayesian shrinkage?
- Does cointegration help?

- Do we need the “hybrid” approach?

Conclusions

- Interesting new application of shrinkage via Bayesian priors
- The empirical work in the paper could be interesting for practitioners developing global models for forecasting and more generally for applied forecasting
- The design of the priors should be clarified and rationalised.