Nowcasting with large, international data sets: do sparse priors help?

Philipp Hauber
Kiel Institute for the World Economy

Christian Schumacher
Deutsche Bundesbank

WORK IN PROGRESS

Motivation & Contribution

- The question to what extent **international variables** are useful in nowcasting domestic economic conditions has not received much attention in the literature.
- In this context, the notion of **sparsity** arises naturally due to possibly irrelevant variables included in large, international data sets (e.g. Kaufmann and Schumacher forthc., JAppEc).
- Our **contribution** to the literature:
  - We estimate a Bayesian factor model using the sparse prior proposed in Bhattarchaya and Dunson (2011, Biometrika).
  - We evaluate GDP nowcasts for the Euro area and the US and asses the relative forecast performance of models estimated **1)** only with the respective national variables and **2)** those based on the combined large, international data set.

Econometric framework

- **Factor model**: N variables, R (unobserved) factors $\rightarrow N \gg R$

$$x_t = \lambda F_t + e_t$$

- Factors evolve according to a VAR(P)

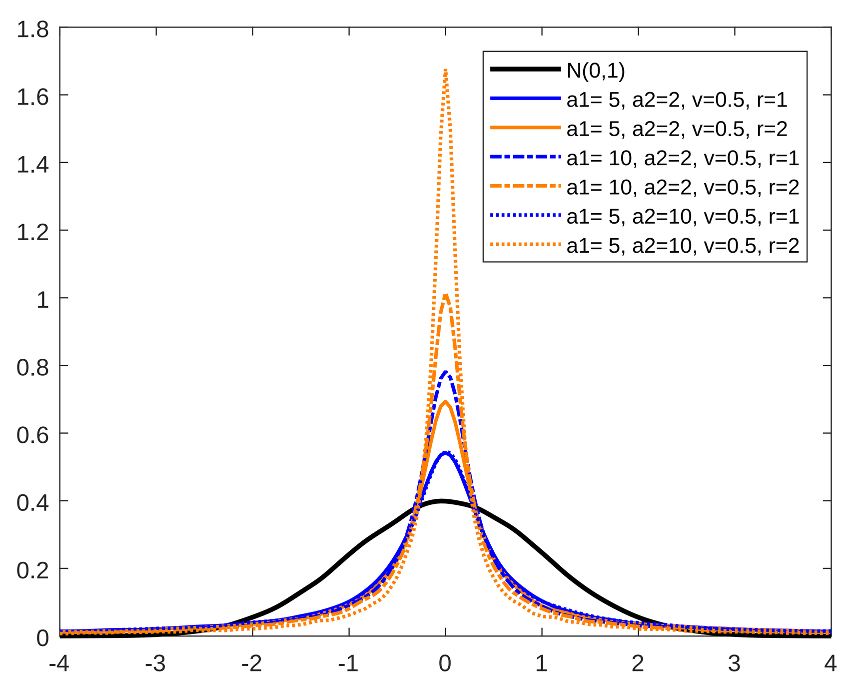
$$F_t = \phi_1^F F_{t-1} + ... + \phi_P^F F_{t-P} + \epsilon_t$$

- **Sparsity**: zero entries in the $N \times R$ loadings matrix $\lambda \rightarrow$ irrelevant variables, block structure, ...

Modeling sparsity

- A **sparse prior** on the loadings should have...
  ... a lot of mass around 0 to capture sparsity $\rightarrow$ global shrinkage
  ... fat tails to pick up non-zero loadings $\rightarrow$ local shrinkage

Figure 1 - Bhattarchaya and Dunson (2011)-prior



- Bhattarchaya and Dunson (2011, Biometrika): **multiplicative Gamma shrinkage prior**

$$p(\lambda_{i,r}|\psi,\tau) \sim N(0, \tau_r^{-1}\psi_{i,r}^{-1}) \; \forall \; r = 1 : R$$

- **Global** shrinkage:

$$\tau_r = \prod_{l=1}^{r} \delta_r \text{ where } \delta_1 \sim Ga(a_1, 1) \text{ and } \delta_l \sim Ga(a_2, 1) \; \forall \; l > 1$$

  – factor-specific: increasing shrinkage across columns of λ for $a_2 > 1$
  – $\uparrow a_1$: lowers prior variance for all r
  – $\uparrow a_2$: more shrinkage for higher r
  – works similar to Minnesota-type shrinkage but across factors rather than lags

- **Local** shrinkage:

$$\psi_{i,r} \sim Ga(\upsilon/2, \upsilon/2)$$

  – factor- **and** variable-specific
  – $\downarrow \upsilon$: concentrates mass of $\psi_{i,r}$ around 0

Simulation: Normal vs. sparse prior

- simulate 2-factor model, randomly setting a share of λ to 0
- compare RMSFE over all common components (forecast horizon = 3 periods)
- sparse prior forecasts as well or better than Normal prior

Table 1 - Relative root mean squared forecast errors ($\mathbf{RMSFE_{sparse}/RMSFE_{Normal}}$)

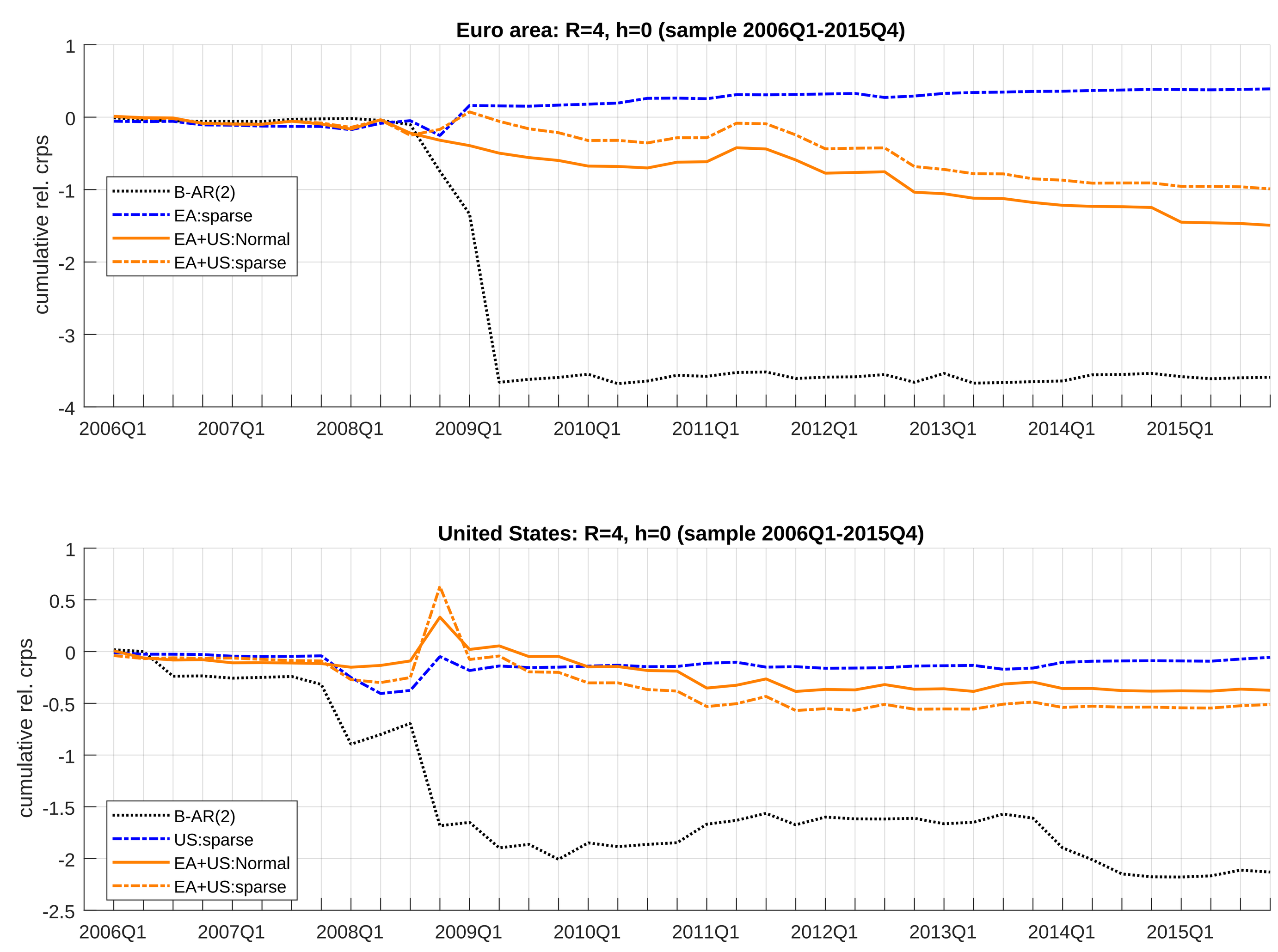
share of non-zero loadings	prior specification ($\mathbf{a_1,a_2,\upsilon}$) (5,2,3)	(5,2,10)	(5,2,0.1)	(10,2,0.1)
1	0.99 (0.97-1.02)	1.00 (0.97-1.02)	1.00 (0.98-1.03)	1.00 (0.98-1.03)
0.66	0.99 (0.96-1.01)	1.00 (0.97-1.02)	1.00 (0.97-1.03)	1.00 (0.97-1.03)
0.5	0.99 (0.96-1.02)	0.99 (0.97-1.02)	0.99 (0.96-1.02)	0.98 (0.96-1.01)
0.33	0.98 (0.95-1.01)	0.99 (0.96-1.01)	0.97 (0.93-1.00)	0.97 (0.92-1.00)
0.2	0.94 (0.90-0.98)	0.98 (0.95-1.00)	0.93 (0.88-0.98)	0.92 (0.87-0.98)
0.1	0.86 (0.79-0.92)	0.92 (0.82-0.98)	0.85 (0.77-0.94)	0.84 (0.77-0.93)

Estimation, data & evaluation set-up

- standard multi-block **Gibbs Sampler**, restrict $\text{Var}(\epsilon) = I_R$ to identify scale of factors
- model specification: $P = 3$, $R = 4$, $\mathbf{a_1 = 5}$, $\mathbf{a_2 = 2}$ and $\boldsymbol{\upsilon = 0.1}$
- combine US and Euro area data sets ($\mathbf{N_{EA} = 73}$, $\mathbf{N_{US} = 106}$): monthly indicators (hard, soft and financial) plus target variable: quarterly GDP growth
- estimation sample starts in 1996Q1, **pseudo real-time**, evaluation sample: 2006Q1-2015Q4
- evaluate density (**CRPS**) and point (**RMSFE**) nowcasts
- Factor models with different data sets
  1. **national** data sets $\rightarrow$ model with EA (US) data to nowcast EA (US) GDP growth
  2. **international** data set $\rightarrow$ model with EA&US data ($\boldsymbol{N = 179}$!) to nowcast EA and US GDP growth
- Both models estimated with the sparse and the Normal prior
- Naive benchmark: B-AR(2)

Results

Figure 2 - cumulative CRPS relative to national data, Normal prior model ($\mathbf{R = 4}$)



Notes: Differences between the cumulated CRPS of respective models and the national data, Normal prior model

Table 2 - CRPS and RMSFE ($\mathbf{R = 4}$)

Sample	2006Q1-2015Q4 CRPS	2006Q1-2015Q4 RMSFE	2006Q1-2009Q4 CRPS	2006Q1-2009Q4 RMSFE	2010Q1-2015Q4 CRPS	2010Q1-2015Q4 RMSFE
Euro area GDP						
B-AR(2)	-0.34	1.00	-0.52	1.00	-0.22	1.00
EA data: Normal	-0.25	0.58	-0.29	0.52	-0.22	0.89
sparse	**-0.24**	**0.54**	**-0.28**	**0.49**	**-0.21**	**0.83**
EA+US data: Normal	-0.29	0.71	-0.33	0.60	-0.26	1.23
sparse	-0.27	0.65	-0.30	0.53	-0.25	1.18
US GDP						
B-AR(2)	-0.36	1.00	-0.48	1.00	-0.28	1.00
US data: Normal	**-0.31**	0.78	**-0.36**	**0.72**	-0.28	0.93
sparse	-0.31	**0.78**	-0.37	0.72	**-0.27**	**0.91**
EA+US data: Normal	-0.32	0.82	-0.36	0.72	-0.29	1.02
sparse	-0.32	0.84	-0.37	0.76	-0.29	1.01

Notes: average CRPS (positive orientation); RMSFE relative to B-AR(2). In bold: best performing model.

Conclusions

- In terms of RMSFE, the **factor models outperform the naive, B-AR(2)** model for the Euro area and the US across all subsamples.
- **Mixed results for CRPS**: in the post-crisis subsample, B-AR(2) only marginally worse.
- **Sparse models perform better** than the ones estimated with a Normal prior for the Euro area, for the US only in the post-crisis subsample.
- **Adding international data does not improve nowcast performance** as the national data seem to capture the relevant information sufficiently.

References

Bhattacharya, A., and Dunson, D. B. Sparse bayesian infinite factor models. *Biometrika 98*, 2 (2011), 291–306.

Kaufmann, S., and Schumacher, C. Identifying relevant and irrelevant variables in sparse factor models. *Journal of Applied Econometrics* (forthc.).