Discussion on:
Dynamic sparse predictive regressions
Mauro Bernardi, Daniele Bianchi, Nicolas Bianco

Anna Simoni *(CREST, CNRS, École Polytechnique, ENSAE)*

12th ECB Conference on Forecasting Techniques - June 13, 2023

Summary. I

Very interesting paper.

I've enjoyed reading it.

Aim:

- Predicting the dynamics of economic variables (*e.g.* forecasting inflation, asset returns);
- large number of predictors;
- the relevance of the predictors may change over time, hence sparsity potentially varying over time.

Prediction is based on a Gaussian time-varying parameter regression model: for $t = 1, \dots, n$,

$$y_t = x_t' \widetilde{\beta}_t + \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, \sigma_t^2), \tag{1}$$

where x_t and $\widetilde{\beta}_t$ are p-dimensional, $\widetilde{\beta}_t$ is sparse, and p might be large compared with n (high-dimension).

High-dimension is dealt with by assuming sparsity and time-varying sparsity is induced through the following prior:

1) reparametrization:

$$\begin{aligned} \widetilde{\beta}_t &= \Gamma_t \beta_t \\ \Gamma_t &= diag(\gamma_{j,t}), \qquad \{\gamma_{j,t}\} \in \{0,1\}^{pn}, \quad \text{where } j = 1, \dots, p, \ t = 1, \dots, n; \end{aligned}$$

2) prior: Bernoulli-Gaussian (**BG**) dynamics

Summary. III

2.1 Random walk for $\beta_{j,t}$: for every $j = 1, \ldots, p$

$$\beta_{j,t} = \beta_{j,t-1} + v_{j,t}, \qquad v_{j,t} \sim \mathcal{N}(0, \eta_j^2)$$

and $\beta_{j,0} \sim \mathcal{N}(0, \kappa_0 \eta_j^2)$ (in vector form $\beta_j \sim \mathcal{N}_{n+1}(0, \eta^2, Q^{-1})$ with Q tridiagonal);

2.2 stochastic volatility: $h_t = \log(\sigma_t^2)$ and

$$h := (h_0, \ldots, h_n) \sim \mathcal{N}_{n+1}(0, \nu^2 Q^{-1}).$$

(Alternative: homoskedasticity (**BGH**)).

Summary. IV

2.3 Persistent stochastic process for $P(\gamma_{j,t} = 1)$:

$$\gamma_{j,t}|\omega_{j,t} \overset{ind.}{\sim} \mathcal{B}e(p_{j,t}), \qquad \omega_{j,t} = \frac{p_{j,t}}{1 - p_{j,t}}$$

and $\omega_j := (\omega_{j,0}, \ldots, \omega_{j,T}) \sim \mathcal{N}_{n+1}(0, \xi_j^2 Q^{-1})$. The components $(\gamma_{j,1}, \ldots, \gamma_{j,n})$ are correlated with respect to the marginal prior, given ξ_j^2, Q.

2.4 Priors on hyperparameters.

3) semi-parametric Variational Bayes algorithm based on two assumptions on the set of approximating densities:

- mean-field factorisation,
- parametric approximation for the density of h and the probability of $\gamma_{j,t}$ (to have smooth sequence of posterior inclusion probabilities).

The main novelty of this prior w.r.t. existing literature is the prior of $\gamma_{j,t}$ which allows persistency through correlation (in the marginal).

Question. Can $\gamma_{j,0}$ be zero? not clear from the text.

Remark. Probabilistic structure of this prior. Is it useful to write the BG prior as a spike-and-slab prior?

$$\widetilde{\beta}_{j,t}|\beta_{j,t-1},\gamma_{j,t} \sim \gamma_{j,t}\mathcal{N}(\beta_{j,t-1},\eta_j^2) + (1-\gamma_{j,t})\delta_0(\widetilde{\beta}_{j,t})$$

under the assumption $\widetilde{\beta}_{j,t} \perp\!\!\!\perp \gamma_{j,t-1}|\beta_{j,t-1},\gamma_{j,t}$.

- The spike part does not depend on $\beta_{j,t-1}$;
- The slab part is persistent and depends on $\beta_{j,t}$ not on $\widetilde{\beta}_{j,t}$, so the past sparsity pattern affects the value of $\widetilde{\beta}_{j,t}$ only through $\gamma_{j,t}$;

- the conditional marginal is

$$\pi(\widetilde{\beta}_{j,t}|\beta_{j,t-1}, p_{j,t}) \sim p_{j,t}\mathcal{N}(\beta_{j,t-1}, \eta_j^2) + (1 - p_{j,t})\delta_0(\widetilde{\beta}_{j,t})$$

and then we can integrate out $p_{j,t-1}$.

Comments and questions. III

This priors competes with the following state-of-the-art priors:

- Koop & Korobilis (2022):
  - soft-spike-and-slab prior with two normals, one of them with variance $\to 0$
  - the variance of $\widetilde{\beta}_{j,t}$ vary over time.

- Rockova & McAlinn (2021): soft-spike-and-slab prior

$$\widetilde{\beta}_{j,t}|\widetilde{\beta}_{j,t-1}, \gamma_{j,t} \sim \gamma_{j,t}\psi_1(\mu_{j,t}, \eta_j^2) + (1-\gamma_{j,t})\psi_0(\lambda_0),$$

  where
  - $\mu_t = \phi_0 + \phi_1(\widetilde{\beta}_{t-1} - \phi_0)$ with $|\phi_1| < 1 \Rightarrow$ motivation to take $\phi_1 = 1$;
  - ψ_0 could be Laplace density.
  - The probability of $\gamma_{j,t}$ depends on $\widetilde{\beta}_{j,t-1}$ explicitly.

  Question. Comparison of the persistency of the sparsity through time induced by the two priors?

Comments and questions. IV

Other questions:

- $\gamma_{jt} \sim Bern(p_{jt})$, $\frac{p_{jt}}{1-p_{jt}} = \omega_{jt}$ and $\omega_j \sim \mathcal{N}_{n+1}(0, \xi_j^2 Q^{-1})$. Motivation for this prior? Could you for instance consider $p_{jt} = p_j$ with $p_j \sim \mathcal{B}eta$? This would also give correlated components.

- Small number of hyperparameters compared to competitor priors but persistency. What if persistency is not satisfied by the true $\widetilde{\beta}_t$?

- Simulations: try $AR(1)$ with less persistency ($\phi_1 = 0.98$ currently) for the active coefficients.

- How large n can be? interesting to see the effect of n in the simulations.

- Correlation between predictors?

We can re-interpret this model in terms of groups, where

- the components of each group show dependency structure
- sparsity among groups and within group $\Rightarrow$ bi-level sparsity.
- In the paper: sparsity at one level.

Every covariates j defines a group:

$$\widetilde{\beta}_j := (\widetilde{\beta}_{j,0}, \ldots, \widetilde{\beta}_{j,n}) = \Gamma_j \beta_j$$

with $\Gamma_j = diag(\gamma_{j,t})$, $t = 0, \ldots, n$ and β_j is $(n+1)$-vector. Here $\gamma_{j,t}$ are standard deviations, not binary variables.

- There are p (potentially active) groups;

- each group has $n + 1$ (potentially active) components.

Reinterpretation. II

The group structure is useful:

- to reduce dimension;
- if one believes there are predictors that are never relevant.

Then, one can for instance extend Mogliani & Simoni (2023, wp) to allow for temporal dependence inside each group.

- Mogliani & Simoni (2023, wp) consider a double spike-and-slab prior.
- Comparison of the two priors would be interesting.

In practice: extend your GMRF prior for β_j to a hard-spike-and-slab (with a Dirac at 0). That is, there is a non-zero probability that a group is inactive.

Question. suppose some predictors are never relevant (as in your simulation), what is the computational cost?