Bayesian nonparametric sparse VAR models

Monica Billio[†] Roberto Casarin[†] Luca Rossini[†‡]

[†] University Ca' Foscari of Venice, Italy [‡] Free University of Bozen, Italy

High Dimensional Small Data Workshop, 19-20 October 2018

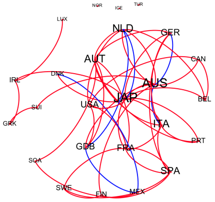
Forthcoming in Journal of Econometrics - Special Issue on Big Data in Dynamic Predictive Econometric Modeling

Our Contributions

- A new Bayesian nonparametric model for multivariate time series models;
- A nonparametric hierarchical prior distribution for coefficients:
  (i) Bayesian Lasso prior for shrinking;
  (ii) Dependent Dirichlet process (DPP) prior for clustering

Our Contributions

- A new Bayesian nonparametric model for multivariate time series models;
- A nonparametric hierarchical prior distribution for coefficients:
  (i) Bayesian Lasso prior for shrinking;
  (ii) Dependent Dirichlet process (DPP) prior for clustering
- Our model allows us to:
  (i) extract network of linkages between economic variables;
  (ii) study the shock transmission mechanism at different lags;
  (iii) identify the most relevant linkages (coefficients) between variables;
  (iv) cluster the linkages into groups via DPP (colored networks).



Panel VAR Models

We assume multi-unit (Canova and Ciccarelli, 2004, JoE) (Canova and Ciccarelli, 2009, IER)(Bassetti et al., 2014, JoE) (Koop and Korobilis, 2016, EER) VAR, where the $\mathbf{y}_{i,t}$ is a m_i vector of variables for the i-th unit in the panel with $i = 1, \ldots, N$

$$\mathbf{y}_{i,t} = \mathbf{b}_i + \sum_{j=1}^{N} \sum_{l=1}^{p} B_{ijl} \mathbf{y}_{j,t-l} + \varepsilon_{i,t}, \quad t = 1, \ldots, T \text{ and } i = 1, \ldots, N, \tag{1}$$

- $\mathbf{b}_i = (b_{i,1}, \ldots, b_{i,m})'$ the vector of intercepts
- B_{ijl} the $(m \times m)$ matrix of unit- and lag-specific coefficients.
- $\varepsilon_{i,t}$, are i.i.d. $\mathcal{N}_m(\mathbf{0}, \Sigma_i)$

Panel VAR Models

We assume multi-unit (Canova and Ciccarelli, 2004, JoE) (Canova and Ciccarelli, 2009, IER)(Bassetti et al., 2014, JoE) (Koop and Korobilis, 2016, EER) VAR, where the $\mathbf{y}_{i,t}$ is a m_i vector of variables for the i-th unit in the panel with $i = 1, \ldots, N$

$$\mathbf{y}_{i,t} = \mathbf{b}_i + \sum_{j=1}^{N}\sum_{l=1}^{p} B_{ijl}\mathbf{y}_{j,t-l} + \varepsilon_{i,t}, \quad t = 1, \ldots, T \text{ and } i = 1, \ldots, N, \tag{1}$$

- $\mathbf{b}_i = (b_{i,1}, \ldots, b_{i,m})'$ the vector of intercepts
- B_{ijl} the $(m \times m)$ matrix of unit- and lag-specific coefficients.
- $\varepsilon_{i,t}$, are i.i.d. $\mathcal{N}_m(\mathbf{0}, \Sigma_i)$

In a more compact form

$$\mathbf{y}_t = (I_{Nm} \otimes \mathbf{x}_t')\,\beta + \varepsilon_t \tag{2}$$

where $\beta = \text{vec}(B)$, $B = (B_1', \ldots, B_N')$, $\varepsilon_t' = (\varepsilon_{1t}', \ldots, \varepsilon_{Nt}')$,

Bayesian Setting

- In a high dimensional setting or when applied to large panel of time series, SUR/VAR models imply a large number of parameters and suffer of overparametrization and overfitting problems.

- Our Original Solution:

1. hierarchical Dirichlet process prior for VAR coefficients;
2. clustering and shrinkage of coefficients toward multiple locations.

$$\Downarrow \Downarrow$$

- A new two-stage hierarchical prior distribution:
  - (i) Bayesian Lasso prior of the Normal-Gamma family;
  - (ii) Random mixture distribution (DPP) for the Normal-Gamma hyperparameters.

Hierarchical Prior - 1st Stage

- Bayesian Lasso prior (Park and Casella, 2008, JASA) to incorporate the penalty

$$f(\boldsymbol{\beta}) = \prod_{j=1}^{n} \mathcal{NG}(\beta_j | 0, 1, \tau),$$

where $\mathcal{NG}(\beta|\mu, \gamma, \tau)$ denotes the normal-gamma distribution

$$\mathcal{NG}(\beta|\mu, \gamma, \tau) = \int_0^{+\infty} \mathcal{N}(\beta|\mu, \lambda)\mathcal{G}a(\lambda|\gamma, \tau/2)d\lambda$$

with location parameter μ, shape parameter $\gamma > 0$ and scale parameter $\tau > 0$.

- We extend Bayesian Lasso by allowing
  - $\gamma \neq 1$ (more flexible shrinking);
  - $\mu \neq 0$ (shrinking to location different from zero)
  - multiple location, shape and scale parameters $(\mu_j, \gamma_j, \tau_j)$.

Hierarchical Prior - 2nd Stage

- Random mixture distribution, i.e. DPP, for Normal-Gamma hyperparameters $(\mu_j, \gamma_j, \tau_j)$, which allows for parameter parsimony through two components:

  1) First component is a random Dirac distribution, which induces sparsity in the VAR coefficients by shrinking μ_j toward zero;

Hierarchical Prior - 2nd Stage

- Random mixture distribution, i.e. DPP, for Normal-Gamma hyperparameters $(\mu_j, \gamma_j, \tau_j)$, which allows for parameter parsimony through two components:

  1) First component is a random Dirac distribution, which induces sparsity in the VAR coefficients by shrinking μ_j toward zero;

  2) Second component is a vector of Dirichlet process priors (DPP), which allows for clustering of the VAR coefficients.

Hierarchical Prior with random measures

- We define the parameters $\boldsymbol{\theta}^* = (\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, \boldsymbol{\tau}^*)$ and assume DP prior $\mathbb{Q}_i$ for $\boldsymbol{\theta}^*_{ij}$.

$$\beta_{ij} \overset{ind}{\sim} \mathcal{NG}(\beta_{ij}|\boldsymbol{\theta}^*_{ij}),$$
$$\boldsymbol{\theta}^*_{ij}|\mathbb{Q}_i \overset{i.i.d.}{\sim} \mathbb{Q}_i,$$

for $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$.

Hierarchical Prior with random measures

- We define the parameters $\boldsymbol{\theta}^* = (\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, \boldsymbol{\tau}^*)$ and assume DP prior $\mathbb{Q}_i$ for $\boldsymbol{\theta}^*_{ij}$.

$$\beta_{ij} \overset{ind}{\sim} \mathcal{NG}(\beta_{ij}|\boldsymbol{\theta}^*_{ij}),$$
$$\boldsymbol{\theta}^*_{ij}|\mathbb{Q}_i \overset{i.i.d.}{\sim} \mathbb{Q}_i,$$

for $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$.

- Following a construction of the hierarchical prior similar to (Hatjispyros et al., 2011, CSDA), we define the vector of DP priors:

$$\mathbb{Q}_1(d\boldsymbol{\theta}_1) = \pi_1 \mathbb{P}_0(d\boldsymbol{\theta}_1) + (1-\pi_1)\mathbb{P}_1(d\boldsymbol{\theta}_1)$$
$$\vdots$$
$$\mathbb{Q}_N(d\boldsymbol{\theta}_N) = \pi_N \mathbb{P}_0(d\boldsymbol{\theta}_N) + (1-\pi_N)\mathbb{P}_N(d\boldsymbol{\theta}_N)$$

Definition of random measures

$$\mathbb{Q}_i(d\boldsymbol{\theta}_i) = \pi_i \mathbb{P}_0(d\boldsymbol{\theta}_i) + (1-\pi_i)\mathbb{P}_i(d\boldsymbol{\theta}_i)$$

- $\mathbb{P}_0(d\boldsymbol{\theta}) \sim \delta_{\{(0,\gamma_0,\tau_0)\}}(d(\mu,\gamma,\tau))$, Dirac's mass;
- $(\gamma_0,\tau_0) \sim g(\gamma_0,\tau_0|\nu_0,p_0,s_0,n_0)$ moves as a two-parameters gamma (see Miller, 1980, Techn); (2-par. Gamma)

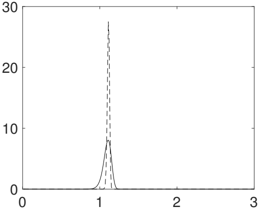
Definition of random measures

$$\mathbb{Q}_i(d\boldsymbol{\theta}_i) = \pi_i \mathbb{P}_0(d\boldsymbol{\theta}_i) + (1-\pi_i)\mathbb{P}_i(d\boldsymbol{\theta}_i)$$

- $\mathbb{P}_0(d\boldsymbol{\theta}) \sim \delta_{\{(0,\gamma_0,\tau_0)\}}(d(\mu,\gamma,\tau))$, Dirac's mass;
- $(\gamma_0,\tau_0) \sim g(\gamma_0,\tau_0|\nu_0,p_0,s_0,n_0)$ moves as a two-parameters gamma (see Miller, 1980, Techn); (▸ 2-par. Gamma)
- $\mathbb{P}_i(d\boldsymbol{\theta}) \overset{i.i.d.}{\sim} \text{DPP}(d\boldsymbol{\theta}, G)$, with $\sum_{k=1}^{\infty} w_k \delta_{(\mu_k,\gamma_k,\tau_k)}(d\boldsymbol{\theta})$
- $G \sim \mathcal{N}(\mu|c,d) \times g(\gamma,\tau|\nu_1,p_1,s_1,n_1)$;
- $\pi_i \sim \mathcal{B}e(\pi_i|1,\alpha_i)$ - is the sparse component probability

Prior Hyperparameter setting

- Sparse case: larger values of (γ_j, τ_j) and strong shrinkage of β_j to 0 ($\nu_0 = 30 \quad s_0 = 1/30 \quad p_0 = 0.5 \quad n_0 = 18$, dashed line the τ_j prior)
- Nonsparse case: smaller values of (γ_j, τ_j) and weak shrinkage of β_j to multiple locations ($\nu_1 = 3 \quad s_1 = 1/3 \quad p_1 = 0.5 \quad n_1 = 10$, solid line the τ_j prior)



- Prior for Σ: Hyper Inverse Wishart $\Sigma \sim \mathcal{HIW}_{\mathcal{G}}(b, L)$ (HIW).
- Prior for $\mathcal{G}$: $p(\mathcal{G}) \propto \psi^{|E|}(1-\psi)^{T-|E|}$

Hierarchical Prior - DAG

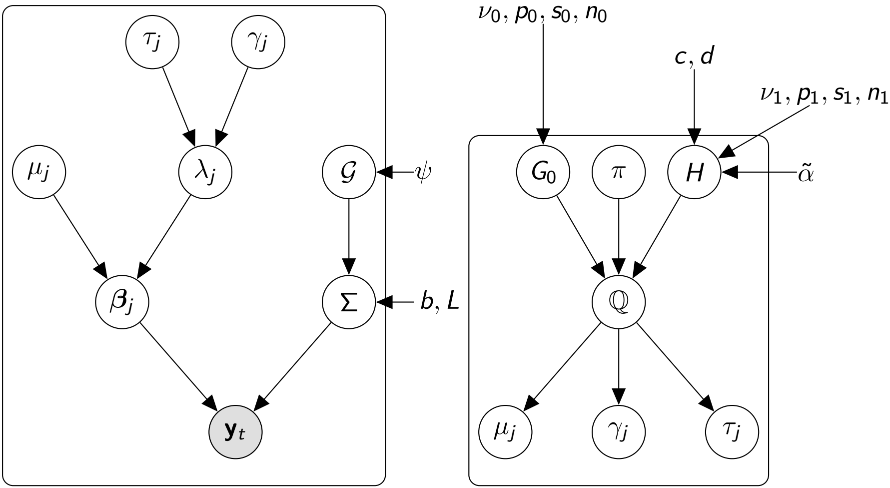


Figure : Left: first stage. Right: second stage.

Data augmentation

Data augmentation and Gibbs sampler for posterior approximation is used (see (Walker, 2007, ComSta) and (Kalli et al., 2011, StaCo)): (More details)

- latent variable u_{ij} reduces problem from infinite to finite mixture, through use of finite sets

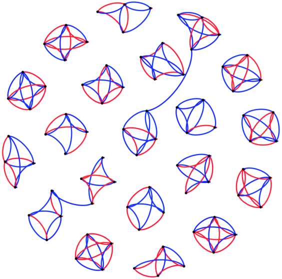
$$A_{w_i}(u_{ij}) = \{k : u_{ij} < w_{ik}\}, \qquad j = 1, \ldots, n_i$$

  where w_{ik} are the stick-breaking mixture weights.

- allocation variable d_{ij} allocates each coefficient to one component of the mixture model $\mathbb{P}_i$;
- latent variable δ_{ij} selects the random measures $\mathbb{P}_0$ (sparse) and $\mathbb{P}_i$ (non-sparse).

Weighted networks for simulated experiment

- We consider different dataset with sample size $T = 100$ from a VAR(1) where $\mathbf{y}_t$ takes:
  - $m = 20$ (small); $m = 40$ (medium); $m = 80$ (large). Other Results



(a) $m = 80$

i

- The blue edges represent negative weights, while red ones represent positive weights;
- nodes represent m variables of VAR model;
- clockwise-oriented edge between two nodes i and j represents non-null coefficient for variable $y_{i,t-1}$ in i-th VAR equation.

Is it good our method?

Our Hierarchical Bayesian nonparametric prior compared with:

- Bayesian Lasso (Park and Casella, 2008, JASA);
- Elastic-Net (Zou and Hastie, 2005, JRSSB);
- Stochastic Search Variable Selection (SSVS) (George et al., 2008, JoE);
- OLS, unrestriced estimator, equivalent to diffuse prior.

	BNP-Lasso	B-Lasso	E-net	SSVS	OLS
$m = 20$	0.228	0.2513	0.2582	0.2938	0.3382
$m = 40$	0.2663	0.3145	0.3143	0.401	0.4835
$m = 80$ (random)	0.2294	0.3011	0.2951	0.5413	0.7048
$m = 80$ (block)	0.2916	0.3773	0.3743	0.5633	0.7290

Table : Mean square deviation statistics for different m.

Is it good our method?

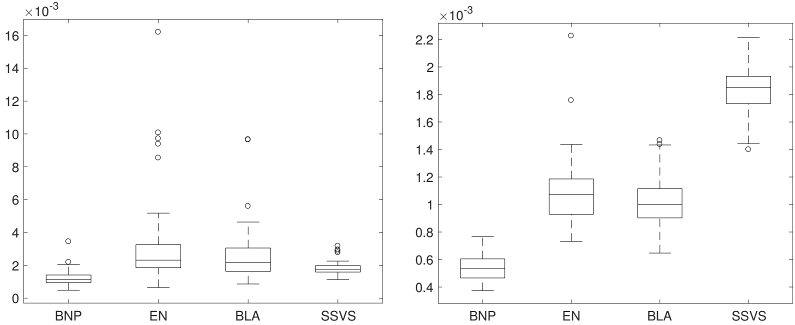


Figure : Boxplots of MSD statistics in the Monte Carlo exercises with $m = 20$ (left) and $m = 80$ (right) over 50 different simulations.

Empirical Applications

- Business Cycle Shock Transmission (1961:Q1-2015:Q2)

- GDP growth rates for most industrialized countries:
    i) Australia, Canada, Japan, Mexico, South Africa, Turkey, US;
    ii) Europe:
        - Core countries - Austria, Belgium, Finland, France, Germany, Netherlands;
        - Periphery countries - Greece, Ireland, Italy, Portugal, Spain;
        - Other EU Countries.

Empirical Applications

- Business Cycle Shock Transmission (1961:Q1-2015:Q2)

- GDP growth rates for most industrialized countries:
  - i) Australia, Canada, Japan, Mexico, South Africa, Turkey, US;
  - ii) Europe:
    - Core countries - Austria, Belgium, Finland, France, Germany, Netherlands;
    - Periphery countries - Greece, Ireland, Italy, Portugal, Spain;
    - Other EU Countries.

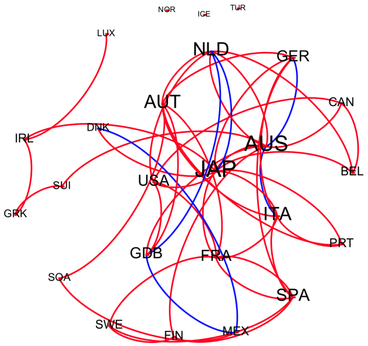
- European Financial Markets (2005-2014)

- Daily realized volatilities using intraday high-low-close price indexes;
- 118 institutions of the financial sector of Euro Stoxx 600:
  - 42 Banks, 31 Financial Services, 31 Insurance companies and 22 Real Estates for European Countries.

Business Cycle Analysis

- Extract network structures to investigate shock transmission effects between variables and cluster the linkages into groups;
- Study shock transmission mechanism at different lags;
- Analyse the network topology;
- Identify most relevant linkages between variables (Billio et al., 2012, JFE),(Diebold and Yilmaz, 2014, JoE),(Bianchi et al., 2018, JoE).

Shock transmission at different lags ▸ lag t-3 ▸ lag t-4

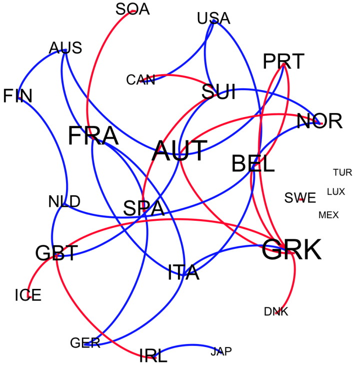


lag $t-1$

- core EU transmit highest percentage of shocks;
- periphery EU receive more shocks;
- Japan is exposed at risk of receiving highest percentage of shocks from other countries, followed by Spain and Australia;
- Australia can transmit the highest percentage of shocks, followed by France, Germany and UK;
- EU countries have positive linkages.

Shock transmission at different lags

lag t-3 lag t-4



lag $t-2$

- core EU transmit highest percentage of shocks;
- periphery EU receive more shocks;
- in the other lags, core EU receive and transmit highest percentage of shocks;
- a shock to Greece, France and Austria turns out to have a bigger effect on the economy.

Some network Statistics - Back

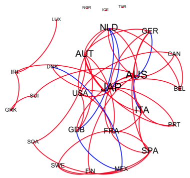
- average path length = average graph distance between all pair of nodes (connected nodes have graph distance 1), reaches minimum value at lag $t-3 \longrightarrow$ faster shock transmission;
- lag $t-1$ is dense graph, density of the graph (0.122), highest number of links (73) $\longrightarrow$ highest presence of contagion effects between countries;
- average degree shows presence of connectedness and transmission of shocks between countries. At lag $t-1$, average degree is 2.92 $\longrightarrow$ network with higher shocks transmission.

	Links	Avg Degree	Density	Avg Path length
$t-1$	73	2.92	0.122	3.423
$t-2$	45	1.80	0.075	3.211
$t-3$	41	1.64	0.068	2.479
$t-4$	52	2.08	0.087	2.718
multilayer	143	5.72	0.238	2.377

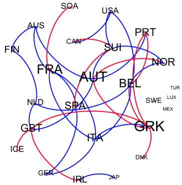
Table : The network statistics for single layers and multilayers.

Multiplex vs single layer

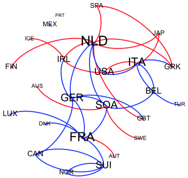
- Multiplex networks encode significantly more information than their single layers taken in isolation;
- Centrality measures require to be adjusted for the multiplex.



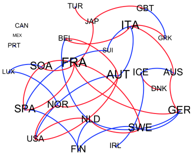
$t-1$



$t-2$



$t-3$

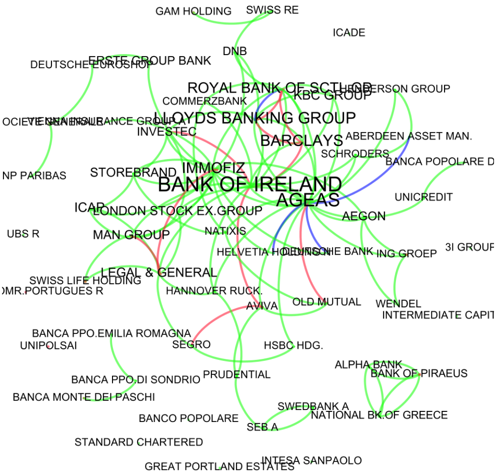


$t-4$

Comments

- Single Layer:
  - Betweenness centrality: Italy is country with more control over the network at all lags except $t-1$, while Japan is central at $t-1$ and France at $t-3$ and $t-4$;
  - Eigencentrality: Italy is one of the central country in the shock transmission over lags. Japan at $t-1$, $t-3$ and France at $t-2$ and $t-4$ are more central.
- Multiplex:
  - Japan, US and biggest European economies (Germany, Italy, France and Spain) appears to be countries that receive the highest percentage of shocks;
  - France, Netherlands, Italy and Austria are most important and central countries, while Turkey, Iceland, Mexico and Finland are less important due to their isolation.

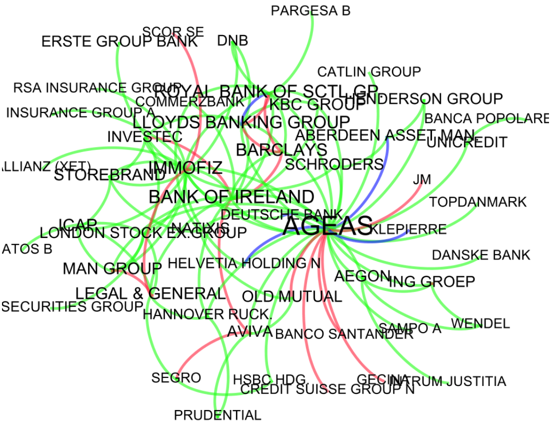
Realized volatility network - Colored network



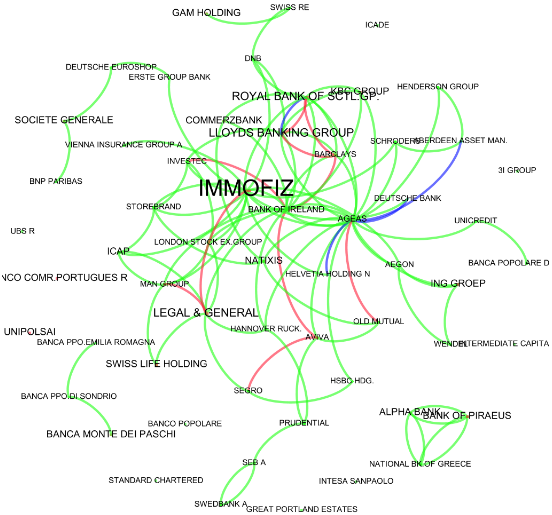
Levels of linkage strengths:

- "negative" (blue);
- "weak positive" (green);
- "positive" (orange);
- "strong positive" (red).

Realized volatility network - Insurance (AGEAS 2-steps ego network)



Realized volatility network - Real estate



Conclusions

- We propose novel Bayesian nonparametric prior, which allows for shrinking coefficients toward multiple locations and for identifying groups of coefficients;
- The proposed hierarchical prior allows for Bayesian nonparametrics in high-dimensional time series models;
- Our approach allows us to
  - extract network of linkages at different lags;
  - identify and cluster the most relevant linkages;

Multiplex network representation provides evidence of stronger shock transmission effects than in single layers and allows for a better identification of variables which are central in shocks transmission.

THANK YOU!!!!!

- The presentation is based on:
  - Billio, M., Casarin, R. and Rossini, L. (2018)."Bayesian nonparametric sparse VAR models". Forthcoming in *Journal of Econometrics*
  - Billio, M., Casarin, R. and Rossini, L. (201-)."Business shock transmission: a multilayer network perspective".
- email address: billio@unive.it.
- Other addresses: r.casarin@unive.it (R. Casarin) and luca.rossini@unibz.it (L. Rossini);

References I

Bassetti, F., Casarin, R., and Leisen, F. (2014). Pitman-Yor process prior for Bayesian inference. *Journal of Econometrics*, 140:49–72.

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2018). Modeling contagion and systemic risk. *Journal of Econometrics*.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Econometrics*, 104:535–559.

Canova, F. and Ciccarelli, M. (2004). Forecasting and turning point prediction in a Bayesian panel VAR model. *Journal of Econometrics*, 120(2):327–359.

Canova, F. and Ciccarelli, M. (2009). Estimating multicountry var models*. *International Economic Review*, 50(3):929–959.

Diebold, F. X. and Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182:119–134.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for var model restrictions. *Journal of Econometrics*, 142:553–580.

Hatjispyros, S. J., Nicoleris, T. N., and Walker, S. G. (2011). Dependent mixtures of Dirichlet processes. *Computational Statistics & Data Analysis*, 55:2011–2025.

Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing*, 21:93–105.

Koop, G. and Korobilis, D. (2016). Model uncertainty in panel vector autoregressive models. *European Economic Review*.

Miller, R. (1980). Bayesian analysis of the two-parameter gamma distribution. *Technometrics*, 22(1).

References II

Park, T. and Casella, G. (2008). The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Sethuraman, J. (1994). A constructive definition of the Dirichlet process prior. *Statistica Sinica*, 2:639–650.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation*, 36:45–54.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society B*, 67:301–320.

Stick-Breaking representation and DPP

Definition - Stick-Breaking Representation (Sethuraman, 1994, StatSin)

The atoms $(\theta_j)_j$ and weights w_j satisfy following hypotheses:

- $(\theta_j)_j$ and $(w_j)_j$ are stochastically independent;
- $(\theta_j)_j$ is independent and identically distributed sequence of random elements with common probability distribution G_0;
- weights (w_j) are determined through stick-breaking construction:

$$w_j = v_j \prod_{k=1}^{j-1}(1 - v_k)$$

  with $w_1 = v_1$, $\prod_{k<1}(1 - v_k) = 1$ and v_j independent random variables distributed as $\mathcal{B}e(1, \alpha)$.

Then $\Phi := \sum w_j \delta_{\theta_j} \sim DP(\alpha, G_0)$, where $\alpha > 0$ is the concentration and G_0 is the base measure.

Back

Two-parameters Gamma

Two-parameters Gamma (Miller, 1980, Techn) has density

$$g(\gamma, \tau | \nu, p, s, n) \propto \tau^{\nu\gamma - 1} p^{\gamma - 1} \exp\{-s\tau\} \frac{1}{\Gamma(\gamma)^n}$$

with hyperparameters fixed such that: (ν, p, s, n) are all strictly positive and

$$g(\gamma, \tau) = g(\tau|\gamma) g(\gamma)$$

where

$$g(\gamma) = \int_0^\infty g(\gamma, \tau) d\tau = C \frac{\Gamma(\nu_0 \gamma)}{\Gamma(\gamma)^{n_0}} \frac{p_0^{\gamma - 1}}{s_0^{\nu_0 \gamma}} \quad (3)$$

$$g(\tau|\gamma) = \frac{g(\gamma, \tau)}{g(\gamma)} = \frac{\tau^{\nu_0 \gamma - 1} e^{-s_0 \tau}}{\Gamma(\nu_0 \gamma)} s_0^{\nu_0 \gamma} \quad (4)$$

with normalizing constant $C = \int_0^\infty g(\gamma) d\gamma$. Back

Hyper-Inverse Wishart (HIW) distribution

- HIW is unique hyper-Markov distribution for Σ with consistent clique-marginals that are inverse Wishart;
- The density of $\Sigma_P \sim \mathcal{IW}(b, D_P)$ for each $P \in \mathcal{P}$

$$p(\Sigma_P|b, D_P) \propto |\Sigma_P|^{-(b+2|P|)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(\Sigma_P^{-1} D_P)\right\}$$

- Full HIW density is:

$$p(\Sigma|b, D) = \frac{\prod_{P\in\mathcal{P}} p(\Sigma_P|b, D_P)}{\prod_{S\in\mathcal{S}} p(\Sigma_S|b, D_S)} \sim \mathcal{HIW}_G(b, D)$$

with b is degree-of-freedom parameter, D is location matrix. $\mathcal{P}, \mathcal{S}$ are prime and separator components.

Back

Gibbs Sampler definition

This distribution is not tractable thus apply Gibbs sampling to draw random numbers. The Gibbs sampler iterates over

1. Draw latent variables U, V given $[\Theta, G, \Sigma, \Lambda, \beta, D, \Delta, Y]$;
2. Draw latent variable Λ given $[\Theta, G, \Sigma, \beta, U, V, D, \Delta, Y]$;
3. Draw atoms Θ given $[G, \Sigma, \Lambda, \beta, U, V, D, \Delta, Y]$;
4. Draw coefficients β given $[\Theta, G, \Sigma, \Lambda, U, V, D, \Delta, Y]$;
5. Draw covariance and graph matrix Σ, G given $[\Theta, \Lambda, \beta, U, V, D, \Delta, Y]$;
6. Draw allocation variable D and sparse variable Δ given $[\Theta, G, \Sigma, \Lambda, \beta, U, V, Y]$
7. Draw probability π given $[\Theta, G, \Sigma, \Lambda, \beta, U, V, Y]$

Back

Detailed Gibbs sampler - ◂ Back I

1. For each $i = 1, 2$ we draw from

$$\pi(v_{ij}| \dots) \propto \mathcal{B}e\left(1 + \sum_{j=1}^{n_i} \mathbb{I}(d_{ij} = d, \delta_{ij} = 1), \alpha + \sum_{j=1}^{n_i} \mathbb{I}(d_{ij} > d, \delta_{ij} = 1)\right)$$

For U we simulate from:

$$f(u_{ij}| \dots) \propto \begin{cases} \mathbb{I}(u_{ij} < w_{1d_{ij}})^{\delta_{ij}} & \text{if } \delta_{ij} = 1 \\ \mathbb{I}(u_{ij} < 1)^{1-\delta_{ij}} & \text{if } \delta_{ij} = 0 \end{cases}$$

Detailed Gibbs sampler - Back II

2. Full conditional posterior distribution of latent variables λ_{ij}:

$$f(\lambda_{ij}|\dots) \propto \mathcal{G}i\mathcal{G}(A_{ij}, B_{ij}, C_{ij})$$

where

$$A_{ij} = \left[(1-\delta_{ij})\tau_0 + \delta_{ij}\tau_{Id_{ij}}\right] \qquad B_{ij} = \left[(1-\delta_{ij})\beta_{ij}^2 + \delta_{ij}(\beta_{ij} - \mu_{id_{ij}})^2\right]$$

$$C_{ij} = \left[(1-\delta_{ij})\gamma_0 + \gamma_{id_{ij}}\delta_{ij} - \frac{1}{2}\right]$$

Detailed Gibbs sampler - Back III

3. For sparse case, posterior distribution of atoms:

$$\pi((\gamma_0, \tau_0)|\dots) \propto g(\gamma_0, \tau_0|\nu_0^*, p_0^*, s_0^*, n_0^*)$$

where $\nu_0^* = \nu_0 + r_{1,0} + r_{2,0}$, $p_0^* = p_0 \prod_{j|\delta_{1j}=0} \lambda_{1j} \prod_{j|\delta_{2j}=0} \lambda_{2j}$,
$s_0^* = s_0 + \frac{1}{2}\sum_{j|\delta_{1j}=0} \lambda_{1j} + \frac{1}{2}\sum_{j|\delta_{2j}=0} \lambda_{2j}$ and $n_0^* = n_0 + r_{1,0} + r_{2,0}$

4. For nonsparse case, posterior for atoms μ_{ik}:

$$\pi(\mu_{ik}|\dots) \propto \mathcal{N}(E_k, V_k)$$

and posterior for (γ_{ik}, τ_{ik}):

$$\pi((\gamma_{ik}, \tau_{ik})|\dots) \propto g(\gamma_{ik}, \tau_{ik}|\nu_{ik}^*, p_{ik}^*, s_{ik}^*, n_{ik}^*)$$

where $\nu_{ik}^* = \nu_1 + r_{i1k}$, $p_{ik}^* = p_1 \prod_{j|\delta_{ij}=1, d_{ij}=k} \lambda_{ij}$, $s_{ik}^* = s_1 + \frac{1}{2}\sum_{j|\delta_{ij}=1, d_{ij}=k} \lambda_{ij}$
and $n_{ik}^* = n_1 + r_{i,1k}$

Detailed Gibbs sampler - Back IV

5. Full conditional for coefficients β_i:

$$\pi(\beta_i | \dots) \propto \mathcal{N}_{n_i}(\tilde{v}_i, M_i)$$

where $M_i = \left(\sum_t \mathbf{x}_t \Sigma^{-1} \mathbf{x}_t' + \Lambda_i^{-1} \right)^{-1}$,

$\tilde{v}_i = M \left(\sum_t \mathbf{x}_t \Sigma^{-1} \mathbf{y}_t + \Lambda_i^{-1} (\boldsymbol{\mu}_i^* \odot \boldsymbol{\delta}_i) \right)$

6. Posterior for Covariance matrix Σ:

$$p(\Sigma | \dots) \propto \mathcal{HIW}_G \left(b + T, D + \sum_t (y_t - X_t \beta)'(y_t - X_t \beta) \right)$$

7. Full conditional for probability of being sparse π_l:

$$f(\pi_i | \dots) \propto \mathcal{B}e \left(n_i + 1 - \sum_{i=1}^{n_i} \mathbb{I}(\delta_{1i} = 1), \alpha_i + \sum_{i=1}^{n_i} \mathbb{I}(\delta_{1i} = 1) \right)$$

Simulated Experiment

Back

Consider different dataset with sample size $T = 100$ from a VAR(1)

$$\mathbf{y}_t = B\mathbf{y}_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{i.i.d}{\sim} \mathcal{N}_m(\mathbf{0}, \Sigma) \quad t = 1, \ldots, 100,$$

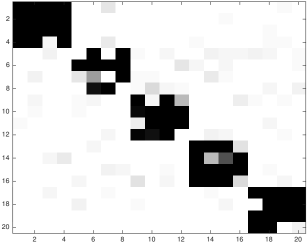
where dimension of $\mathbf{y}_t$ and of B takes different values:

- $m = 20$ (small); $m = 40$ (medium); $m = 80$ (large),

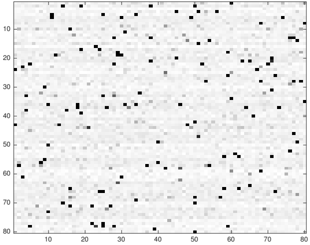
and different settings of matrix B:

- block-diagonal matrix $B = \text{diag}\{B_1, \ldots, B_{m/4}\} \in \mathcal{M}_{(m,m)}$ with blocks B_j $(j = 1, \ldots, m/4)$ of (4×4) matrices on main diagonal, where elements are randomly taken from $\mathcal{U}(-1.4, 1.4)$ and checked for weak stationarity condition;
- random matrix B is (80×80) matrix with 150 elements randomly chosen from $\mathcal{U}(-1.4, 1.4)$ and checked for weak stationarity condition.

Posterior Mean of allocation matrix - Back



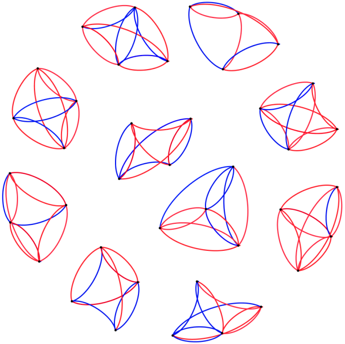
$m = 20$



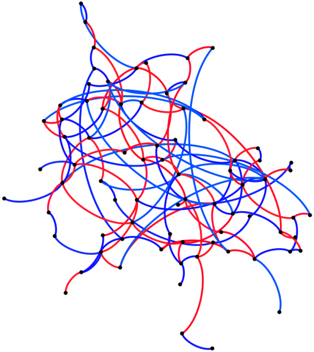
$m = 80$ with random entries

Figure : Posterior mean of Δ, which shows allocation of coefficients between two random measures $\mathbb{P}_0$ and $\mathbb{P}_l$. White color indicates if coefficient δ_{ij} is equal to zero (i.e. sparse component), while black one if δ_{ij} is equal to one (nonsparse components).

Weighted Networks Extraction - Back

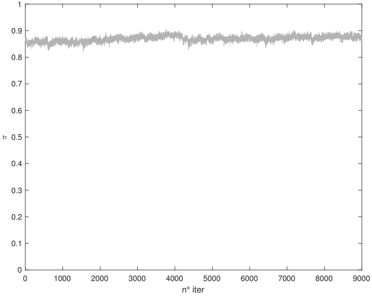


$m = 40$



$m = 80$ with random entries

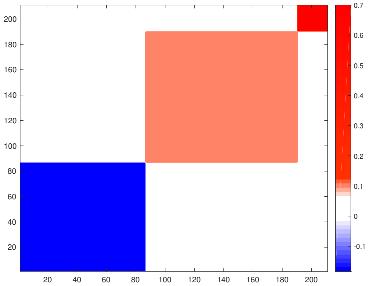
MCMC samples for probability of being sparse - Back



MCMC sample for π

- MCMC samples for the probability of being sparse, π;
- Posterior mean has value 0.87 providing evidence of high sparsity in the model;
- small proportion of coefficient is not null and then is responsible of the transmission of shocks between countries.

Choice of component for network construction - Back

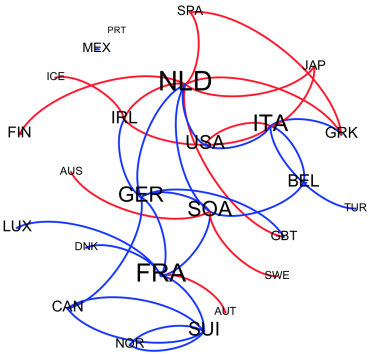


Co-clustering matrix

- Choice of nonsparse components and posterior pairwise probabilities of joint classification $P(D_i = D_j | Y, \delta = 1)$.
- Presence of different clusters from the co-clustering matrix based on the location atom, μ, generated at each MCMC iteration and build up from the least square marginal clustering.
- Presence of 3 types of relation: "negative", "positive" and "strong positive".

Contagion transmission

Back

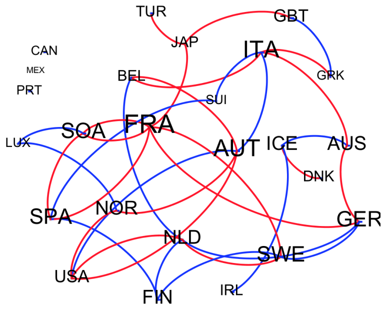


lag $t - 3$

- Germany and Italy have the highest in-degree and Netherlands have the highest out-degree;
- EU countries have positive effects between them;
- existence of negative effects between the rest-of-the-world countries and EU countries is shown.

Contagion transmission

Back



lag $t - 4$

- Germany and Italy have the highest in-degree and Netherlands have the highest out-degree;
- EU countries have positive effects between them;
- existence of negative effects between the rest-of-the-world countries and EU countries is shown.