Fast Inference in Panel Vector Autoregressive Models using Integrated Rotated Gaussian Approximations

Martin Feldkircher (Oesterreichische Nationalbank)
Florian Huber (University of Salzburg)
Gary Koop (University of Strathclyde)
Michael Pfarrhofer (University of Salzburg)

Motivation

- Much evidence that working with multi-country time series models is beneficial for macro forecasting (spillovers, interlinkages)
- Relevant data sets can be enormous
- Our empirical work has up to 18 variables for each of 39 countries
- Unrestricted multi-country VAR involves approx. 500 equations, each of which has approx. 1000 RHS variables (lag length = 2)
- Over-parameterization problems can be addressed by:
- Restricting model/compressing data (e.g. global VARs, factor or other compression methods)
- Global VARs: Crespo Cuaresma et al. (2016); Huber (2016); Chudik et al. (2016);
- Shrinkage priors (subjective or global-local shrinkage priors): Koop & Korobilis (2016; 2018), Canova & Ciccarelli (2009)

Motivation

- But with both approaches issues arise:
  - Global VAR models introduce strong restrictions on the coefficients associated with the variables from "other" countries $\rightarrow$ can be deleterious for forecasting
  - PVAR models with suitable shrinkage priors can handle such issues but become computationally burdensome (Bayesian MCMC methods not scaleable)
- In this paper, we estimate a huge unrestricted panel VAR
- Bayesian shrinkage = Horseshoe prior
- Computation hurdle overcome using Integrated Rotated Gaussian Approximation (IRGA) techniques
- IRGA proposed in van den Boom, Reeves & Dunson (2020, Biometrika): Approximating posteriors with high-dimensional nuisance parameters via integrated rotated Gaussian approximation
- Machine learning tool which vastly reduces computational burden

The Panel VAR

- Want to model M-dimensional vector of macroeconomic and financial variables $\boldsymbol{y}_{it}$ that is specific to country $i = 1, \ldots, N$
- Stacking $\boldsymbol{y}_{it}$ for all countries yields an $n = (MN)-$ dim. vector $\boldsymbol{y}_t = (\boldsymbol{y}'_{1t}, \ldots, \boldsymbol{y}'_{Nt})'$
- $\boldsymbol{y}_{it}$ evolves according a VAR model:

$$\boldsymbol{y}_{it} = \boldsymbol{A}_{i1}\boldsymbol{y}_{it-1} + \cdots + \boldsymbol{A}_{iP}\boldsymbol{y}_{it-P} + \boldsymbol{B}_i\boldsymbol{z}_t + \varepsilon_{it},$$

- $\boldsymbol{A}_{ij}$ are $M \times M$ coefficient matrices
- $\boldsymbol{z}_t = (\boldsymbol{y}'_{-i,t-1}, \ldots, \boldsymbol{y}'_{-i,t-P})'$ with $\boldsymbol{y}_{-i,t} = (\boldsymbol{y}'_{1t}, \ldots, \boldsymbol{y}'_{i-1,t}, \boldsymbol{y}_{i+1,t} \ldots, \boldsymbol{y}'_{Nt})'$
- $\boldsymbol{B}_i$ is an $M \times K_{other}$ matrix with $K_{other} = (n - M)P$
- Errors are i.i.d. Gaussian with covariance Σ unrestricted p.d. matrix
- Computation greatly simplified by writing in "equation-by-equation" form:

$$y_{ij,t} = \boldsymbol{A}'_{ij,\bullet}\boldsymbol{x}_{it} + \boldsymbol{B}'_{ij,\bullet}\boldsymbol{z}_{it} + \sum_{s=1}^{j-1} q_{is}y_{is,t} + \sum_{v<i} \boldsymbol{q}'_{iv}\boldsymbol{y}_{vt} + \varepsilon_{ij,t}$$

  where $A_{ij,\bullet}$ and $\boldsymbol{B}_{ij,\bullet}$ denote the j^{th} rows of $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$
- $\boldsymbol{q}_{ij,\bullet} = (q_{i1}, \ldots, q_{i,j-1}, \boldsymbol{q}'_{i1}, \ldots, \boldsymbol{q}'_{ii-1})'$ are the covariance parameters

Brief Introduction to Global-local shrinkage priors

- Popular in range of fields
- Can be represented as scale mixtures of Normals
- For the j^{th} coefficient in a model ϕ_j

$$\phi_j \sim \mathcal{N}(0, \psi_j \lambda), \quad \psi_j \sim f, \quad \lambda \sim g,$$

- λ controls global shrinkage (common to all coefficients)
- Global shrinkage commonly used in VARs (e.g. Minnesota prior) to reduce over-fitting concerns
- ψ_j does local shrinkage (specific to j^{th} coefficient)
- I.e. if ψ_j is estimate to be close to zero then ψ_j is shrunk to be close to zero
- f and g are mixing densities and a large range of choices have been proposed
- IRGA methods work for priors in this class (Gaussian invariant to rotations)
- We use Horseshoe but many others possible (e.g. the Lasso, Dirichlet-Laplace, etc.)

Integrated Rotated Gaussian Approximations in PVARs

- MCMC methods for posterior and predictive inference in PVAR are standard
- Problem: MCMC is too slow
- $\boldsymbol{B}_i$ is huge ($M \times K_{other}$)
- K_{other} is number of coeffs from other countries
- Manipulating (inverting etc.) posterior covariance matrices of dimensions in tens of thousands is tough (very slow and liable to crash)
- $\boldsymbol{B}_i$ is likely very sparse (most elements equal to zero)
- $\boldsymbol{A}_i$ is much smaller
- $\boldsymbol{A}_i$ likely non-sparse (own country effects more important than other country effects)
- Idea of IRGA: Use MCMC methods on important parameters (here $\boldsymbol{A}_i$ and Σ_i)
- Other parameters (like $\boldsymbol{B}_i$) are approximated using some fast algorithm
- We use Vector Approximate Message Passing (VAMP)

Integrated Rotated Gaussian Approximations in PVARs

- PVAR equations are:

$$\boldsymbol{y}_{ij} = \tilde{\boldsymbol{x}}_{ij}\alpha_{ij} + \boldsymbol{z}_i\boldsymbol{B}_{ij,\bullet} + \varepsilon_{ij} \Leftrightarrow \boldsymbol{y}_{ij} \sim \mathcal{N}(\tilde{\boldsymbol{x}}_i\alpha_{ij} + \boldsymbol{z}_i\boldsymbol{B}_{ij,\bullet}, \sigma^2_{\varepsilon,ij}\boldsymbol{I}_T) \tag{1}$$

  where $\tilde{\boldsymbol{x}}_{ij}$ is a $T \times K_i(= Mp + j - 1 + (i-1)M)$ dimensional matrix with typical t^{th} row $\tilde{\boldsymbol{x}}_{it} = (y_{i1,t}, \ldots, y_{ij-1,t}, \boldsymbol{y}'_{1t}, \ldots, \boldsymbol{y}'_{i-1t})'$, $\alpha_{ij} = (q_{i1}, \ldots, q_{ij-1}, \boldsymbol{q}'_{i1}, \ldots, \boldsymbol{q}'_{ii-1}, \boldsymbol{A}'_{ij,\bullet})'$
- Let $\boldsymbol{Q}$ be $T \times T$ rotation matrix obtained from QR-decomposition of $\tilde{\boldsymbol{x}}_i$ and decompose $\boldsymbol{Q} = (\boldsymbol{Q}_1, \boldsymbol{Q}_2)$ with $\boldsymbol{Q}_1$ being $T \times K_i$ and $\boldsymbol{Q}_2$ is $T \times (T - K_i)$
- Multiplying by $\boldsymbol{Q}$ and exploiting rotation invariance of the Gaussian yields equivalent representation of (1):

$$\boldsymbol{Q}'_1\boldsymbol{y}_{ij} \sim \mathcal{N}(\boldsymbol{Q}'_1\tilde{\boldsymbol{x}}_{ij}\alpha_{ij} + \boldsymbol{Q}'_1\boldsymbol{z}_i\boldsymbol{B}_{ij,\bullet}, \sigma^2_{\varepsilon,ij}\boldsymbol{I}_{K_i}) \tag{2}$$

$$\boldsymbol{Q}'_2\boldsymbol{y}_{ij} \sim \mathcal{N}(\boldsymbol{Q}'_2\boldsymbol{z}_i\boldsymbol{B}_{ij,\bullet}, \sigma^2_{\varepsilon,ij}\boldsymbol{I}_{T-K_i}) \tag{3}$$

- The second equation follows since $\boldsymbol{Q}'_2\tilde{\boldsymbol{x}}_{ij} = \boldsymbol{0}$, α_{ij} does not appear in it!
- The two sub-models in (2) and (3) motivate first approximating $\boldsymbol{B}_{ij,\bullet}$ and $\sigma^2_{\varepsilon,ij}$ and then, conditional on the approximate posterior $\hat{p}(\boldsymbol{B}_{ij,\bullet}|\boldsymbol{Q}'_2\boldsymbol{y}_{ij})$, estimate posterior distribution of α_{ij} through MCMC

Computational Aspects

- Any Gaussian approximation can be used for $\hat{p}(\boldsymbol{B}_{ij,\bullet}|\boldsymbol{Q}_2'\boldsymbol{y}_{ij})$, we use VAMP
- The sub-model in (2) can then be estimated conditional on $\hat{p}(\boldsymbol{Q}_1'\boldsymbol{B}_{ij,\bullet}|\boldsymbol{Q}_2'\boldsymbol{y}) = \mathcal{N}(\overline{\boldsymbol{B}}_{ij,\bullet}, \overline{\boldsymbol{V}}_{ij,\bullet})$
- Rewriting (2) and plugging in the approximate moments of $\hat{p}(\boldsymbol{B}_{ij,\bullet}|\boldsymbol{Q}_2'\boldsymbol{y})$ yields:

$$(\boldsymbol{y}_{ij} - \boldsymbol{Q}_1'\boldsymbol{z}_i\overline{\boldsymbol{B}}_{ij,\bullet}) \sim \mathcal{N}(\boldsymbol{Q}_1'\tilde{\boldsymbol{x}}_{ij}\alpha_{ij}, \boldsymbol{Q}_1'\boldsymbol{z}_i\overline{\boldsymbol{V}}_{ij,\bullet}\boldsymbol{z}_i'\boldsymbol{Q}_1 + \sigma^2_{\varepsilon,ij}\boldsymbol{I}_{K_i})$$

- This gives us Gaussian likelihood (simply a regression model)
- Can be combined with any (conditionally) Gaussian prior on α_{ij} leading to a textbook form of the posterior of α_{ij} (we use a Horseshoe prior)

Data

- Data from 2001m2 to 2019m12 (OECD's short-term indicator data base)
- Up to 18 variables for each of 39 countries
- Macroeconomic data
  - Industrial production and output gap
  - Exports and imports of goods
  - Unemployment rate
  - Consumer prices
  - etc.
- Financial data
  - Short-term interest rates
  - Local currency per US dollar
  - Share prices
  - etc.
- Leading indicators
  - OECD leading indicator, amplitude adjusted
  - Production of total construction and manufacturing
  - Volume of retail trade
  - Consumer and manufacturing confidence indicators
  - etc.
- Variables transformed to stationarity

Empirical Results

- In interests of time, I will focus on real time forecasting exercise, but note:
- Evidence based on Diebold-Yilmaz spillover measure
- Heatmaps indicate great deal of sparsity (spillovers from country i to country j mostly are zero)
- But there are a few cases of strong spillovers
- Many other cases of weak spillovers

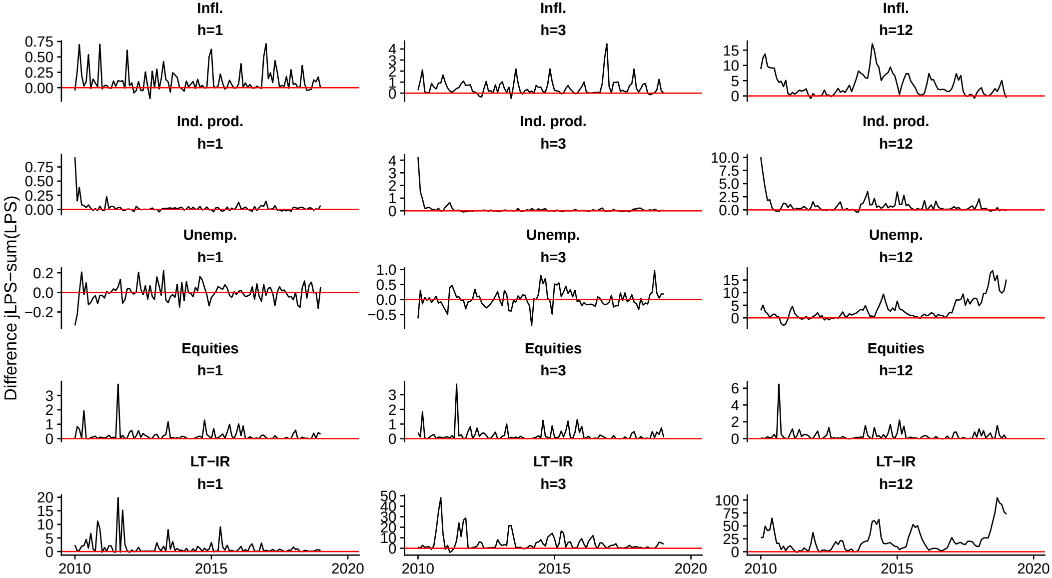
Design of Real Time Forecasting Exercise

- Forecasts evaluated in terms of RMSE and log predictive scores
- Forecast evaluation period begins in 2010m1
- Models in Forecasting Exercise
  - PVAR-IRGA: unrestricted PVAR with Horseshoe prior estimated using IRGA methods
  - BVAR-SC: Single-country BVARs with hierarchical Minnesota prior
  - FAVAR-10: BVAR-SC with individual country models augmented by 10 factors extracted from the remaining (“foreign") series
  - GVAR: Global VAR estimated using Bayesian methods (see cran package of Böck and Feldkircher and Huber)
  - Dynamic factor model (not shown, it forecasts poorly)
- Computation time of PVAR-IRGA less than all of these alternatives

Does Joint Modelling Pay Off?

- Dovern, Feldkircher and Huber (2016, JEDC) "Does joint modelling of the world economy payoff? Evaluating global forecasts from a Bayesian GVAR"
- Idea: Difference between log of joint predictive density (across all countries) for a variable and sum of log of marginals good measure of usefulness of going with a multi-country VAR relative to each country separate
- This is plotted on next slide for four key variables and three forecast horizons
- Evidence mixed but overall evidence for usefulness of joint modelling for
- Inflation and long term interest rates in most periods
- Industrial production at beginning of sample
- Stock returns occasionally
- Not for unemployment rate except for $h = 12$ at end of sample

Difference Between Joint and Marginals



Results of Forecasting Exercise

- Next table gives RMSEs and LPLs for five core variabls (data available for all countries)
- PVAR-IRGA rows (in red) are actual values
- Other rows are benchmarked relative to this
- Most important finding: PVAR-IRGA works!
- Bayesian analysis of huge dimensional unrestricted PVARs (and VARs) made possible through IRGA methods
- Second most important finding: PVAR-IRGA works well!
- Density forecast performance substantially better than other models (see LPLs)
- Point forecasts a bit better than other models (see RMSFEs)
- Some exceptions to previous two statements exist

Forecast results

Model	RMSE					LPS				
	Infl.	Ind. prod.	Unemp.	Equities	LT-IR	Infl.	Ind. prod.	Unemp.	Equities	LT-IR
h=1										
PVAR-IRGA	0.377	0.488	0.137	0.760	0.143	-54.677	-21.233	21.902	-45.523	14.940
BVAR-SC	1.008	1.005	0.916	0.999	0.925	-4.393	-0.301	6.167	-1.472	0.525
FAVAR-10	1.010	1.008	0.918	1.003	0.926	-4.277	-0.347	6.207	-1.627	0.596
GVAR	1.159	1.107	1.176	1.084	1.346	-29.304	-3.842	-9.514	-3.409	-28.079
h=3										
PVAR-IRGA	0.518	0.591	0.243	0.770	0.288	-108.127	-27.948	3.153	-46.868	-7.912
BVAR-SC	0.998	1.012	0.892	0.986	0.943	-4.950	-1.329	4.565	-1.256	2.208
FAVAR-10	0.999	1.015	0.886	0.988	0.945	-4.731	-1.331	5.121	-1.306	2.396
GVAR	1.118	1.088	0.949	1.048	1.068	-17.970	-2.317	2.420	-1.134	-11.967
h=12										
PVAR-IRGA	0.777	0.700	0.526	0.738	0.699	-174.003	-35.570	-23.228	-47.238	-41.437
BVAR-SC	1.003	1.029	0.929	0.988	0.984	-2.334	-2.308	0.984	-0.370	7.607
FAVAR-10	1.007	1.034	0.906	0.993	0.999	-3.164	-2.445	2.123	-0.717	8.052
GVAR	1.236	1.296	1.048	1.125	1.068	-28.957	-5.799	-0.180	-2.399	-6.960

Summary

- Goal: Posterior and predictive inference in a PVAR of huge dimension using Bayesian shrinkage (Horseshoe prior)
- Impossible to achieve this goal with MCMC (too slow)
- IRGA methods are approximate but offer vast reduction in computational burden
- Forecasting success relative to competitors