Bayesian nonparametric vector autoregressions

Maria Kalli

Joint work with Jim Griffin

Bayesian Nonparametrics Networking Workshop 2023
Monash Business School

- Vector Autoregressive (VAR) models (Sims, 1980) represent a multivariate time series y_t as

$$y_t = B_1 y_{t-1} + \ldots + B_L y_{t-L} + e_t \tag{1}$$

where,

- $y_t = (y_{1,t}, \ldots, y_{p,t})'$ is a p-dimensional vector of timeseries,
- $\{B_l\}_{l=1}^{L}$ are $(p \times p)$-dimensional matrices of unknown coefficients, with L the number of lags, and
- $e_t = (e_{1,t}, \ldots, e_{p,t})'$ is a p-dimensional innovation vector with distribution $\mathrm{N}(0, \Sigma)$

- VAR models are used by macro-economists to:
  - capture the structure and joint dynamic behaviour of macroeconomic series,
  - produce forecasts of key indicators such as inflation, and output growth,
  - carry out structural inference and policy analysis
- This leads to informed monetary and fiscal policy decisions.

- The assumptions of the 'vanilla' VAR in Eq(1),
    1. linear relationships between y_t and lags of y_t,
    2. Gaussian innovations,
    3. constant innovation variance,
    4. constant coefficients for the conditional mean, and
    5. stationarity,

  are considered unrealistic. See Weise (1999), Giavazzi et al. (2000), Sörensen et al. (2001), Ravn and Sola (2004), and Gambacorta et al. (2014)

This has motivated widely-used extensions to time varying parameters VAR model with stochastic volatility (TVP-SV-VAR), (Primiceri, 2005), allowing time variation in the B_l's and Σ

- The main issues with TVP-SV-VAR models are:
  - The random walk drift of the coefficients $\{B_l\}_{l=1}^{L}$.

    It may not reflect reality.

  - The large number of parameters, leading to poor forecasting out-of-sample performance.

    More variables and more lags (if needed) lead to overfitting.

Bitto and Frühwirth-Schnatter (2019), Cadonna et al. (2020), and Huber et al. (2021) are some papers trying to address overfitting by using shrinkage priors.

Alternatively, Kalli and Griffin (2018) follow a Bayesian nonparametric approach and directly model the stationary and transition densities as infinite mixtures leading to the Bayesian nonparametric vector autoregressive model (Bayes-NP-VAR)

- **Bayesian parametrics** restrict inference to a family of models that can be indexed by a finite number of parameters.
- **Bayesian nonparametrics** move away from this framework by allowing for a richer and larger class of models.
- Achieved by considering infinite dimensional families of probability models.

- Priors on such families are known as **Bayesian nonparametric priors**.
- Bayesian nonparametric methods place a prior on an infinite dimensional parameter space and adapt their complexity to the data.
- Flexible, data driven methods applicable to a wide range of problems that can be used to exploit large data sets.

  See Hjort et al. (2010)
- This leads to better out-of-sample forecasts. See Norets and Pati (2017).

The Bayesian nonparametric VAR (Bayes NP VAR) model's transition density of y_t and y_t^L (which represents the L lags of y_t) has the form:

$$p(y_t \mid y_t^L) = \sum_{j=1}^{\infty} \omega_j(y_t^L) k(y_t | y_t^L, \theta_j) \tag{2}$$

- $k(y_t|y_t^L, \theta_j)$ is the transition density of the jth component.
- The transition weights are:

$$\omega_j(y_t^L) = \frac{w_j \, k(y_t^L \mid \theta_j)}{\sum_{k=1}^{\infty} w_k \, k(y_t^L \mid \theta_k)},$$

where $k(y_t^L \mid \theta_k)$ is the stationary distribution of $k(y_t \mid y_t^L, \theta_k)$ and the mixture weights w_j are defined as

$w_1 = v_1, \; w_j = v_j \prod_{m<j}(1 - v_m)$, and $v_j \overset{iid}{\sim} \mathrm{Be}(1, M)$.

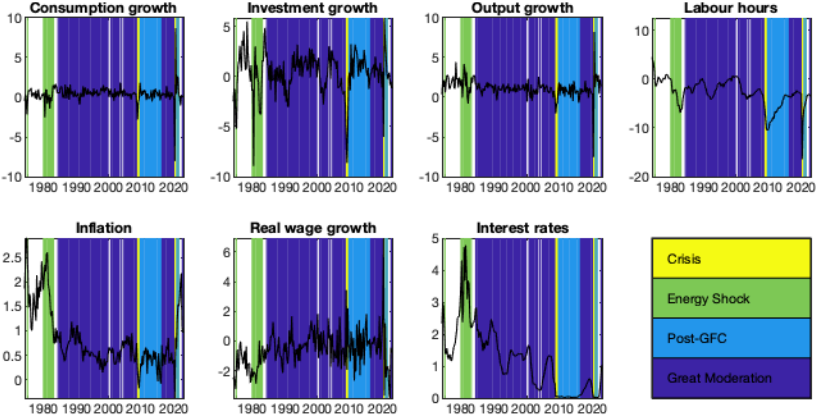
$\theta_j \overset{iid}{\sim} H$

- Kalli and Griffin (2018) assume a stationary VAR for $k(y_t|y_t^L, \theta_j) = \mathsf{N}(y_t \mid A_j + B_j y_t^L, \Sigma_j)$ with stationary covariance Ψ_j.
- Leads to a mixture of VAR models, where each component can interpreted as an economic regime.
- The mixing weights, $\omega_j(y_t^L)$ of the transition density depend on previous lags, allowing for different component transition densities to be favoured at different periods.
- Importantly, $\omega_j(y_t^L)$ acts like a discriminant function, which we can use to understand the variables that drive the changes in regimes.

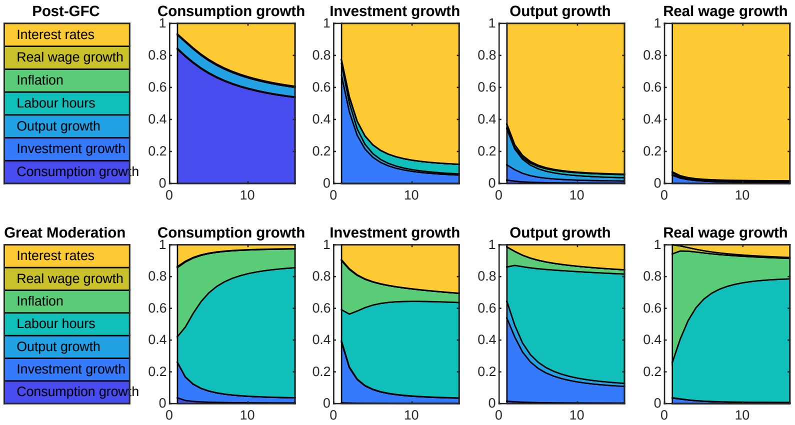
To illustrate this approach of identifying which variables are driving the allocation to regimes (components) we created a dataset based on Smets and Wouters (2007). Table 1 describes this dataset.

Variable	Construction
Output growth	Differenced, $100 \times \left(\ln\left(\frac{\text{RGDP}}{\text{LNSindex}}\right)\right)$
Consumption growth	Differenced, $100 \times \left(\ln\left(\frac{\text{RPCEC}}{\text{LNSindex}}\right)\right)$
Investment growth	Differenced, $100 \times \left(\ln\left(\frac{\text{RFPI}}{\text{LNSindex}}\right)\right)$
Inflation	$100 \times \left(\ln\left(\frac{\text{GDPDEF}_t}{\text{GDPDEF}_{t-1}}\right)\right)$
Labour hours	$100 \times \left(\ln\left(\frac{(\text{PRS85006023*LNS12000000})/100}{\text{LNSindex}}\right)\right)$
Real wage growth	Differenced, $100 \times \left(\ln\left(\frac{\text{PR85006103}}{\text{GDPDEF}}\right)\right)$
Interest rates	Federal Funds Rate/4

Table: Quarterly, seasonally adjusted, from Q2 1958 to Q4 2022. Sources FRED, Bureau of Labour Statistics.


Consumption growth
Investment growth
Output growth
Labour hours
Inflation
Real wage growth
Interest rates
Crisis
Energy Shock
Post-GFC
Great Moderation
1980 1990 2000 2010 2020

Comparing the Great Moderation and Post-Crisis periods



Comparing the Great Moderation and Post-Crisis periods - Means

Great Moderation

Consump. growth	Invest. growth	Output growth	Lab. hrs	Inflation	Wage growth	Interest rates
0.54	0.15	1.04	-1.33	0.64	-0.22	1.59
(0.10)	(-0.15)	(-0.15)	(-0.23)	(-0.29)	(0.23)	(0.44)

Post-Crisis

Consump. growth	Invest. growth	Output growth	Lab. hrs	Inflation	Wage growth	Interest rates
0.64	0.94	1.01	-5.68	0.50	-0.31	0.03
(0.19)	(0.22)	(-0.17)	(-1.02)	(-0.53)	(0.15)	(-1.26)

- We define the log odds of being in component j over component m as $\delta_{j,m}(y_t^L) = \log \frac{\omega_j(y_t^L)}{\omega_m(y_t^L)}$
- For the BayesNP-VAR model

$$\delta_{j,m}(y_t^L) = \alpha_{j,m} + \beta_{j,m}(y_t^L)$$

where

$\beta_{j,m}(y_t^L) = V_t^{(j,m)} \Lambda^{(j,m)} {V_t^{(j,m)}}^T$ and $V_t^{(j,m)} = U_{j,m}^T (y_t^L - \mu^\star)$

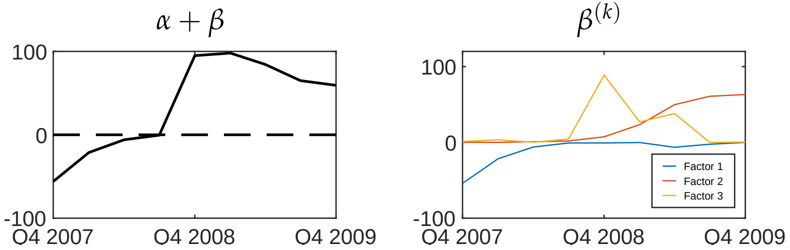
> $U_{j,m}$ is a square matrix whose columns are the eigenvectors of $\Psi_m^{-1} - \Psi_j^{-1}$,
> $\Lambda^{(j,m)}$ is a diagonal matrix whose non-zero entries are the corresponding eigenvalues, and
> $\mu^\star$ is a function of μ_j, μ_m, Ψ_j and Ψ_m

The t^{th} observation is allocated to the jth regime if $\delta_{j,m}(y_t^L) > 0$.

This rule only depends on the observed values through $\beta_{j,m}$ which can decomposed be as

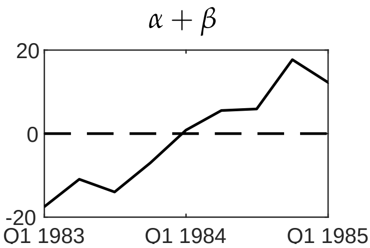
$$\beta_{j,m} = \sum_{k=1}^{L} \beta_{j,m}^{(k)}$$

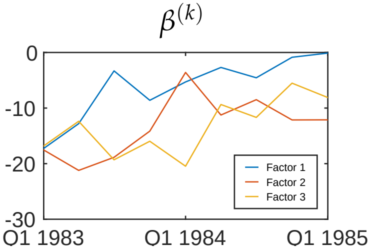
Into the Global Financial Crisis



Factor	Consump. growth	Invest. growth	Output growth	Lab. hrs	Inflation	Wage	Interest rates
1	0	0	0	0	0.29	0	-0.95
2	0	0	0	-0.99	0	0	0
3	-0.41	-0.54	-0.51	0	-0.24	0.46	0

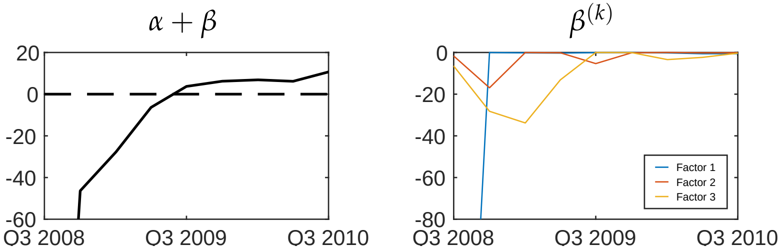
Recovery from Energy shock





Factor	Consump. growth	Invest. growth	Output growth	Lab. hrs	Inflation	Interest rates
1	0	0	-0.42	0.55	0.65	-0.26
2	0.32	0.24	-0.55	-0.70	0	0
3	-0.30	0	0.52	-0.38	0.66	0

Recovery from GFC



Factor	Consump. growth	Invest. growth	Output growth	Inflation	Real Wage	Interest rates
1	0	0	0	0	0	0.99
2	-0.78	0.42	0.26	0.28	0.24	0
3	-0.41	-0.54	-0.51	-0.24	0.46	0

Summary

- We use Bayes NP VAR to find different regimes.
- Standard techniques for analysing VAR model can be used in each component.
- The log odds can be used to understand movements between different components and use a decomposition to understand variables which drive the change.

- The BayesNP VAR is a stationary model,
  - $\omega(y_t^L)$ depends on previous lagged values
  - but it is not indexed by time, t
- Would the introduction of $\omega_t(y_t^L)$, lead to better forecasts ?

- The Dirichlet process can be derived via the normalisation of a Gamma process.
- The jump sizes of a Gamma process can be constructed by:

  Multiplying each jump size of a Beta process with independent Gamma random variables see Griffin and Leisen (2017)
- We use this result to introduce time dependence in $\omega(y_t^L)$ and build a time varying (and thus non-stationary) version of our BayesNP-VAR.

- The transition density will be :

$$p_t(y_t|y_t^L) = \sum_{j=1}^{\infty} \omega_{j,t}(y_t^L) k_j(y_t|y_t^L)$$

where

  - $\omega_{j,t}(y_t^L) = \frac{w_j x_{j,t}\, k(y_t, y_t^L|\theta_j)}{\sum_{k=1}^{\infty} w_k x_{k,t}\, k(y_t^L|\theta_k)}$
  - $w_1, w_2, w_3, \ldots$ are the weights of a Beta process.
  - $x_{j,t}$ follows a Cox-Ingersoll-Ross (CIR) process Cox et al. (1985)

- In practice, we use a finite approximation

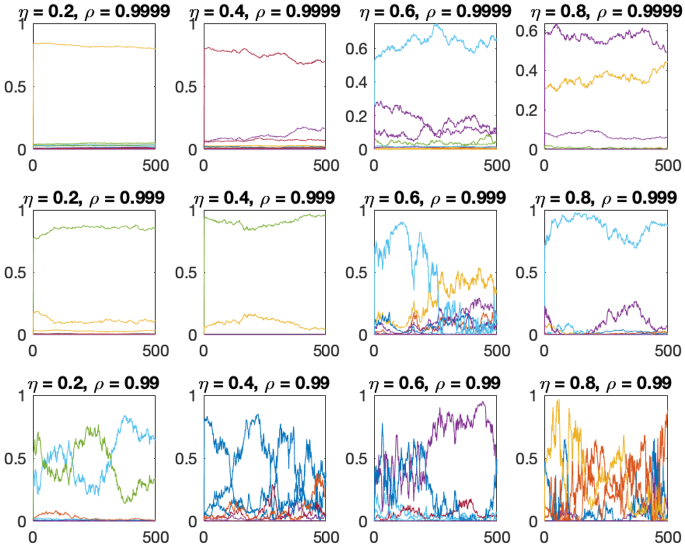
$$p_t(y_t|y_t^L) = \sum_{j=1}^{K} \omega_{j,t}(y_t^L) k_j(y_t|y_t^L)$$

where

- $\omega_{j,t}(y_t^L) = \frac{w_j x_{j,t}\, k(y_t, y_t^L|\theta_j)}{\sum_{k=1}^{K} w_k x_{k,t}\, k(y_t^L|\theta_k)}$
- $w_j \sim \text{Be}(\frac{M}{K}, \phi\frac{M}{K})$.
- $x_{j,t}$ follows a CIR process which has a Gamma stationary distribution with shape parameter $M/K(1+\phi)$, scale parameter 1 and autoregressive parameter ρ.

- Let $\eta = \dfrac{M/K + 1}{(M/K(1+\phi) + 1)}$ be the proportion of variation within each unnormalised weight – the variation of $\omega_{t,i}(y_t^L)$ around its stationary mean.
- ρ controls the level of time dependence of the CIR process.

So, it should be interesting to see how different values η and ρ affect the behaviour of $\omega_{t,i}(y_t^L)$.



η = 0.2, ρ = 0.9999
η = 0.4, ρ = 0.9999
η = 0.6, ρ = 0.9999
η = 0.8, ρ = 0.9999
η = 0.2, ρ = 0.999
η = 0.4, ρ = 0.999
η = 0.6, ρ = 0.999
η = 0.8, ρ = 0.999
η = 0.2, ρ = 0.99
η = 0.4, ρ = 0.99
η = 0.6, ρ = 0.99
η = 0.8, ρ = 0.99

- We place a uniform prior on η and for ρ we select a $Beta(200 \times 0.999, 200 \times 0.001)$ prior.
- Inference is made using a Metropolis-within-Gibbs scheme with proposals adapted during the MCMC run.

- We use log-predictive scores to compare the out-of-sample predictive performance of competing models.

- The log-predictive score is defined as:

$$-\sum_{i=s}^{T-h} \log p_t(y_{i+h}|y_1,\ldots,y_i),$$

  where T is the size of the time series, s is the time from where the prediction starts, and h is the predictive horizon. We looked at $h = 1, 2,$ and 4 quarters.

- Smaller scores indicate better predictive performance.

Joint and marginal log-predictive scores for Smets and Wouters (2007) US data									
Model	Horizon	Joint and marginal log-predictive scores							
		Overall	consumption	investment	output	lab.hrs	inf	r.wage	int.rate
BayesNP-TW-VAR(1)	1	53.93	10.68	38.10	16.10	11.78	0.83	34.92	25.23
	2	79.68	10.95	39.25	16.13	20.36	0.35	30.47	30.06
	4	104.22	10.04	37.75	15.22	27.58	0.19	28.35	29.37
BayesNP-VAR(1)	1	51.70	11.00	39.75	18.00	13.00	1.46	34.91	26.05
	2	89.50	11.82	41.23	17.30	22.00	1.00	30.77	29.32
	4	129.53	10.30	39.28	16.00	28.60	1.00	29.27	28.79
TVP-SV-VAR(1)	1	183.78	26.11	51.98	27.02	31.07	16.71	66.85	15.56
	2	243.81	42.98	67.42	44.72	59.89	34.27	76.85	34.04
	4	314.91	67.36	89.76	71.06	84.18	56.77	93.15	63.88

- Both BayesNP-VAR and BayesNP-TW-VAR outperform the TVP-SV-VAR especially at the longer horisons.
- BayesNP-TW-VAR performs better than BayesNP-VAR in all horizons for the joint log scores, and marginally better on the scores of some of the variables.

- Introduced two Bayesian nonparametric models for analysing macroeconomic time series focusing on the transition density of y_t and y_t^L.
- Both models allow for departures from linearity in the conditional mean, heteroskedasticity in the conditional variance, and non-Gaussianity. One is stationary and other is nonstationary.
- Discussed a method for understanding which variables are driving the transitions between regimes.
- The Bayesian nonparametric approach led to more robust out-of-sample predictive performance at all horizons.

Thank you so much for your attention!

References I

Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework, *JoE* **210**: 75–97.

Cadonna, A., Frühwirth-Schnatter, S. and Knaus, P. (2020). Triple the gamma—a unifying shrinkage prior for variance and variable selection in sparse state space and tvp models, *Econometrics* **8**(2).

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A Theory of the Term Structure of Interest Rates, *Econometrica* **53**: 385–407.

Gambacorta, L., Hofmann, B. and Peersman, G. (2014). The Effectiveness of Unconventional Monetary Policy at the Zero Lower Bound: A Cross-Country Analysis, *JMCB* **46**: 615–642.

Giavazzi, F., Jappelli, T. and Pargano, M. (2000). Searching for nonlinear effects of fiscal policy: Evidence from industrial and developing countries, *European Economic Review* **44**: 1259–1289.

Griffin, J. E. and Leisen, F. (2017). Compound random measures and their use in Bayesian non-parametrics, *JRSS B* **79**: 525–545.

References II

Hjort, N. L., Holmes, C., Müller, P. and Walker, S. G. (eds) (2010). *Bayesian Nonparametrics.*, Statistic and Probabilistic Mathematics, 1st edn, Cambridge University Press.

Huber, F., Koop, G. and Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models, *JBES* **39**: 669–683.

Kalli, M. and Griffin, J. E. (2018). Bayesian Nonparametric Vector Autoregressive Models, *JoE* **203**: 267–282.

Norets, A. and Pati, D. (2017). Adaptive Bayesian Estimation of Conditional Densities, *Econometric Theory* **33**: 980–1012.

Primiceri, G. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy, *RES* **72**: 821–852.

Ravn, M. O. and Sola, M. (2004). Asymmetric effects of monetary policy in the US, *Federal Reserve Bank of St. Louis Review* **86**: 41–60.

Sims, C. (1980). Macroeconomics and Reality, *Econometrica* **48**: 1–48.

Smets, F. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach, *AER* **97**: 586–606.

References III

Sörensen, B. E., Wu, L. and Yosha, O. (2001). Output Fluctuations and Fiscal Policy, *European Economic Review* **45**: 1271–1310.

Weise, C. L. (1999). The asymmetric effects of monetary policy: A non linear vector autoregression approach, *JMCB* **31**: 85–108.