Bayesian Analysis of VARs, Etc.

Frank Schorfheide
University of Pennsylvania

July 22, 2024

Statistical Inference

- Frequentist:
  - pre-experimental perspective;
  - condition on "true" but unknown θ_0;
  - treat data Y as random;
  - study behavior of estimators and decision rules under repeated sampling.
- Bayesian:
  - post-experimental perspective;
  - condition on observed sample Y;
  - treat parameter θ as unknown and random;
  - derive estimators and decision rules that minimize expected loss (averaging over θ) conditional on observed Y.

Why Bayesian Inference

- John Geweke (ca. 1998-2000): **Why not?**
- Advantages:
  - parameter uncertainty and uncertainty about future shocks is treated symmetrically;
  - a lot of tools for working with models that have many parameters relative to the number of observations available for estimation;
  - availability of powerful computational tools.

Bayes Theorem

- **Bayes Theorem:**

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)} \propto p(Y|\theta)p(\theta) \tag{1}$$

- Marginal data density (MDD) or marginal likelihood:

$$p(Y) = \int p(Y|\theta)p(\theta)d\theta. \tag{2}$$

- Understanding how to use the proportionality $\propto$ will keep you sane, when doing derivations.

- In practical work, you will generate draws (sometimes independent, sometimes serially correlated) from the posterior on Day 1:

$$\theta^i \sim p(\theta|Y), \quad i = 1, \dots, N. \tag{3}$$

- On Day 2, you will post-process the draws and do something interesting.

One Calculation to Remember: AR(p) Model

- AR(p) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + u_t, \quad u_t \sim iid\mathcal{N}(0, \sigma^2). \tag{4}$$

- Let $k = p + 1$ and define

$$\phi = [\phi_0, \phi_1, \ldots, \phi_p]', \quad x_t = [1, y_{t-1}, \ldots, y_{t-p}]' \tag{5}$$

- Write the AR(p) model as linear regression

$$y_t = x_t'\phi + u_t, \quad u_t \sim iid\mathcal{N}(0, \sigma^2). \tag{6}$$

- Let Y be the $T \times 1$ vector with elements y_t, X be the $T \times k$ matrix with rows x_t'. Write likelihood as

$$p(Y|\phi, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp\left\{ -\frac{1}{2\sigma^2}(Y - X\phi)'(Y - X\phi) \right\}. \tag{7}$$

One Calculation to Remember: Known Variance

- Assume that prior takes the form

$$\phi|\sigma^2 \sim N\big(\underline{\phi}, \sigma^2 \underline{V}_\phi\big). \tag{8}$$

- Prior density as a function of ϕ up to a constant of proportionality is

$$p(\phi|\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(\phi-\underline{\phi})'\underline{V}_\phi^{-1}(\phi-\underline{\phi})\right\}. \tag{9}$$

- Combine prior and likelihood, condition on σ^2, and ignore terms that do not depend on σ^2:

$$p(\phi|Y,\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\left[(Y-X\phi)'(Y-X\phi)+(\phi-\underline{\phi})'\underline{V}_\phi^{-1}(\phi-\underline{\phi})\right]\right\}. \tag{10}$$

One Calculation to Remember: Guess and Verify

- **Educated guess:** Posterior distribution is Gaussian with mean $\bar{\phi}$ and covariance matrix $\bar{V}_\phi$:

$$\phi|(Y,\sigma^2) \sim N(\bar{\phi}, \sigma^2 \bar{V}_\phi). \tag{11}$$

- Then posterior density will have the form:

$$p(\phi|Y,\sigma^2) \propto \exp\left\{ -\frac{1}{2\sigma^2}(\phi - \bar{\phi})' \bar{V}_\phi^{-1} (\phi - \bar{\phi}) \right\}. \tag{12}$$

- Manipulate quadratic function in exponential term of (10) to verify that:

$$\bar{\phi} = \big(X'X + \underline{V}_\phi^{-1}\big)^{-1}\big(X'Y + \underline{V}_\phi^{-1}\underline{\phi}\big), \quad \bar{V}_\phi = \big(X'X + \underline{V}_\phi^{-1}\big)^{-1}. \tag{13}$$

Posterior Sampling

Algorithm (Direct Sampling)

For $i = 1$ to N:

1. *draw* $(\sigma^2)^i$ *from* $\mathcal{IG}(\bar{\nu}, \bar{s}^2)$;
2. *draw* $\phi^i | (\sigma^2)^i$ *from* $N(\bar{\phi}, (\sigma^2)^i \bar{V}_\phi)$;
3. *let* $\theta^i = [\phi^i, (\sigma^2)^i]$.

Note: To sample a σ^2 from the $\mathcal{IG}(\nu, s^2)$ distribution, generate ν *iid* draws Z_s from a $N\big(0, (s^2)^{-1}\big)$ and then set $(\sigma^2)^i = \left[\sum_{s=1}^{\nu} Z_s^2\right]^{-1}$.

Postprocessing Draws

- We previously defined $\theta = [\phi, (\sigma^2)]$, but on the general Bayesian inference slides, you can think of θ as generic parameter vector.
- Let $\mathbb{E}_\pi[h]$ be posterior expectation of $h(\theta)$:

$$\mathbb{E}_\pi[h] = \int h(\theta)p(\theta|Y)d\theta. \tag{14}$$

- Use Monte Carlo average to approximate posterior mean:

$$\bar{h}_N = \frac{1}{N}\sum_{i=1}^{N} h(\theta^i). \tag{15}$$

- Accuracy of Monte Carlo approximation: LLN/CLT implies ($\mathbb{V}_\pi[h] < \infty$):

$$\sqrt{N}(\bar{h}_N - \mathbb{E}_\pi[h]) \Longrightarrow N\big(0, \mathbb{V}_\pi[h]\big), \tag{16}$$

Forecasting: Combine Parameter and Shock Uncertainty

- **AR(1) Example:**

$$y_{T+h} = \phi^h y_T + \sum_{s=0}^{h-1} \phi^s u_{T+h-s}, \quad u_t \sim N(0, \sigma^2). \tag{17}$$

- *h*-step ahead conditional distribution:

$$y_{T+h}|(Y_{1:T}, \phi, \sigma^2) \sim N\left(\phi^h y_T, \sigma^2 \frac{1-\phi^h}{1-\phi}\right). \tag{18}$$

- Posterior predictive distribution:

$$p(y_{T+h}|Y_{1:T}) = \int p(y_{T+h}|y_T, \phi, \sigma^2) p(\phi, \sigma^2|Y_{1:T}) d(\phi, \sigma^2). \tag{19}$$

- **Sampling from posterior pred. distribution:** For each draw $(\phi^i, (\sigma^2)^i)$ from the posterior distribution $p(\phi, \sigma^2|Y_{1:T})$ sample a sequence of innovations $u^i_{T+1}, \ldots, u^i_{T+h}$ from a $N(0, (\sigma^2)^i)$ and compute y^i_{T+h} as a function of ϕ^i, $(\sigma^2)^i$, $u^i_{T+1}, \ldots, u^i_{T+h}$, and $Y_{1:T}$.

Adding Some Bells and Whistles

- Add prior $p(\sigma^2)$; work out posterior $p(\sigma^2|Y)$; combine pieces:

$$p(\phi, \sigma^2|Y) = p(\phi|Y, \sigma^2)p(\sigma^2|Y) \tag{20}$$

If prior is in Inverse-Gamma family, then so is the posterior $\Longrightarrow$ conjugacy.

- VAR extension:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p}, \quad u_t \sim iid\mathcal{N}(0, \Sigma). \tag{21}$$

Need Inverse Wishart for Σ and matrix Normal distribution for $\Phi = [\Phi_0, \ldots, \Phi_p]'$.

Prior Distributions

- **Three views:**
  1. Keep them "neutral" so that posterior reflects shape of likelihood function.
  2. Use them to "regularize" an estimation problem, e.g., add more information in settings with low observation-to-parameter ratio; make likelihood more elliptical is irregular settings (weak identification, etc.)
  3. Combine sample information in Y with additional information (subjective or objective).

- The popular Minnesota prior is used for regularization; the DSGE model prior of Del Negro and Schorfheide (2004, International Economic Review) was used to add information from economic theory.

- **Minnesota prior:** (i) centered at univariate random walk representations; (ii) coefficients on higher order lags are on average closer to zero; (iii) the sum of coefficients on a variable's own lags is close to one; (iv) if all variables are stable at some initial level, they tend to persist at that level.

Model Selection / Averaging

- Thus far, we have not discussed how to determine the number of lags in an autoregression.
- More generally, there is a question about which regressors to include in a regression model.
- In practice, it will also be useful to have a procedure to determine the scaling of the prior variance
- $\Longrightarrow$ **Bayesian model selection or model averaging**

What Do We Mean By Model?

- In the Bayesian world, a model is a joint distribution over data and parameters, so it comprises the likelihood function $p(Y|\theta)$ **and** the prior distribution $p(\theta)$.
- Thus, either changing the likelihood function, e.g., by including more lags in an AR model, or changing the prior distribution, e.g., by increasing its variance, **changes the model**!

Posterior Model Probabilities

- The posterior model probabilities are given by (we can have more than two models)

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_{j=1}^{M} \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad p(Y|\mathcal{M}_i) = \int p(Y|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}|\mathcal{M}_i) d\theta_{(i)}. \tag{22}$$

- Recall Bayes Theorem: (no subscripts to make notation easier)

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}. \tag{23}$$

- Model probabilities are updated based on the marginal likelihood / marginal data density (MDD) / denominator of Bayes Theorem. Can be interpreted as one-step-ahead pseudo-out-of-sample predictive score or as penalized likelihood function, e.g., Schwarz criterion.

Hyperparameter Selection or Averaging for VAR

- "Performance" of Bayesian VAR is sensitive to prior variance.
- Choose prior variance in a data-driven way.
- Hierarchical model

$$p(Y|\Phi,\Sigma)p(\Phi,\Sigma|\lambda)p(\lambda), \tag{24}$$

where λ controls features of the prior.

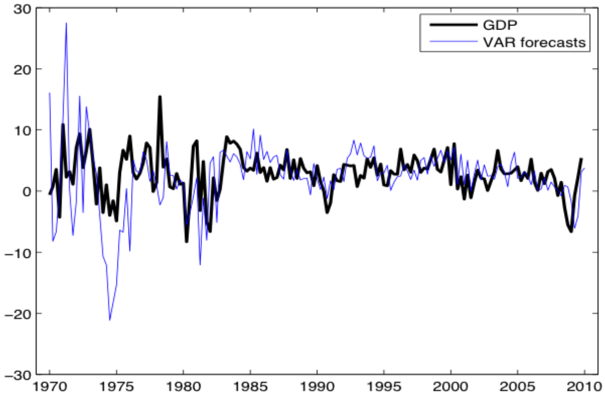
Selection vs. Averaging

- **Selection:**
  - Let
$$\hat{\lambda} = \text{argmax } p(Y|\lambda), \quad p(Y|\lambda) = \int p(Y|\Phi, \Sigma)p(\Phi, \Sigma|\lambda)d(\Phi, \Sigma). \tag{25}$$
  - Continue the analysis with $p(\Phi, \Sigma|Y, \hat{\lambda})$.
- **Averaging:**
  - Specify a prior $p(\lambda)$; factorize posterior as
$$p(\Phi, \Sigma, \lambda|Y) = p(\Phi, \Sigma|Y, \lambda)p(\lambda|Y), \quad p(\lambda|Y) \propto p(Y|\lambda)p(\lambda). \tag{26}$$
  - Analysis is based on marginal posterior $p(\Phi, \Sigma|Y)$.

(The following pages are taken from a WP version of Giannone, Lenza, and Primiceri (2015, REStat))

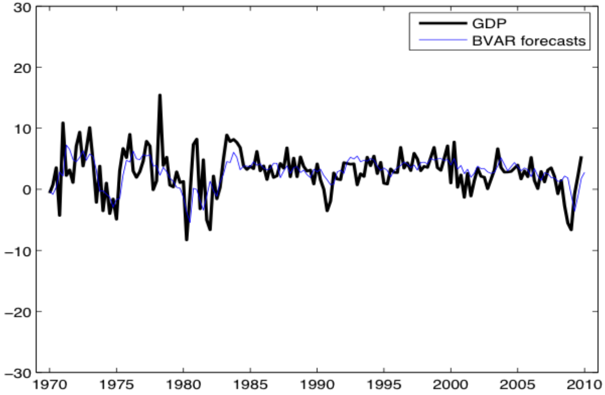
US GDP growth and VAR forecast (1-step ahead)

Flat-prior VAR



US GDP growth and BVAR forecast (1-step ahead)

BVAR (MN+SOC+DIO priors + hyperparameter selection)



BVARs: Forecasting performance

Mean Squared Forecast Errors

		7-variable VAR		
		Flat-prior	BVAR with MN prior (λ=0.2)	BVAR with MN+SOC+DIO
1 Quarter Ahead	Real GDP	19.18	9.61	7.97
	GDP Deflator	2.27	1.53	1.35
	Federal Funds Rate	1.83	1.08	1.03
1 Year Ahead	Real GDP	11.90	5.48	3.42
	GDP Deflator	2.22	1.85	1.58
	Federal Funds Rate	0.56	0.40	0.31

Modifications to the Prior

- The conjugate MNIW prior has the property that prior and posterior variance have a Kronecker structure: $\Sigma \otimes V_\phi$.
  - (+) posteriors can be computed equation by equation;
  - (+) computations require only inversion of $n \times n$ and $k \times k$ matrices; but not $nk \times nk$.
  - (−) strong symmetry restrictions on prior and posterior.
- Many generalizations require inversion of $nk \times nk$ matrices; not good for high-dimensional inference.
- Chan (2021, QE): write system as

$$Ay_t = B_1 y_{t-1} + \ldots + B_p y_{t-p} + e_t, \quad e_t \sim iidN(0, D), \tag{27}$$

where A is lower triangular with ones on the diagonal such that $|A| = 1$; D is diagonal.

Extensions of the Reduced-Form VAR Framework

1. Mixed frequency observations
2. Priors for long-run behavior
3. Spike-and-slab priors: sparse versus dense models
4. Time-varying coefficients
5. Stochastic volatility
6. Outliers
7. Censoring

MF-VARs

Reference: Schorfheide and Song (2015): "Real-Time Forecasting with a Mixed-Frequency VAR," *Journal of Business Economics & Economic Statistics.*

- In macroeconomic applications, vector autoregressions (VARs) are typically estimated either exclusively based on
  - quarterly observations
    $\Longrightarrow$ large set of macroeconomic series is available
  - monthly information
    $\Longrightarrow$ VAR is able to track the economy more closely in real time
- The paper...
  - develops a mixed-frequency VAR (MF-VAR) to exploit the respective advantages of both monthly and quarterly VARs;
  - generates, evaluates forecasts and documents how within-quarter information alters the forecasts in real time.

State-Space Representation of MF-VAR

- State-Transition Equation
  - Economy evolves at monthly frequency according to the following VAR(p) dynamics:

$$x_t = \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + \Phi_c + u_t, \quad u_t \sim iidN(0, \Sigma)$$

  - Partition: $x_t' = [x_{m,t}', \; x_{q,t}']$
- Measurement Equation
  - Actual observations are denoted by y_t and subscript indicates the observation frequency

$$\begin{aligned} y_{m,t} &= x_{m,t} \\ y_{q,t} &= \frac{1}{3}(x_{q,t} + x_{q,t-1} + x_{q,t-2}) \quad \text{if observed in period } t \end{aligned}$$

- Estimate model with Bayesian state-space techniques...

Priors For Long-Run Dynamics

Reference: Giannone, Lenza, Primiceri (2019): "Priors for the Long-Run," *Journal of the American Statistical Association*.

- Estimation is typically based on conditional likelihood functions that ignore the likelihood of the initial observations.
- **Example:**

$$y_t = c + \phi y_{t-1} + u_t = \underbrace{\phi^{t-1} y_1 + c \sum_{s=0}^{t-2} \phi^s}_{DC_t} + \underbrace{\sum_{s=0}^{t-2} \phi^s u_{t-j}}_{SC_t} \tag{28}$$

- Write

$$DC_t = \begin{cases} y_1 + (t-1)c & \text{if } \phi = 1 \\ \frac{c}{1-\phi} + \phi^{t-1}\left(y_1 - \frac{c}{1-\phi}\right) & \text{if} |\phi| < 1 \end{cases} \tag{29}$$

- Deterministic component may absorb too much low frequency variation of the time series.

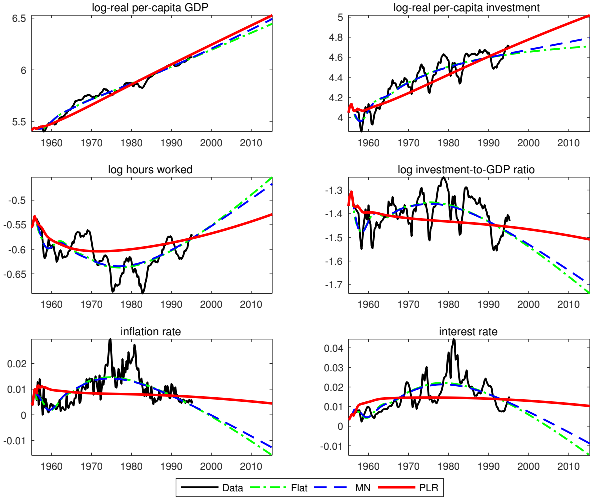


FIGURE 2.1. Deterministic component for selected variables implied by various 7-variable VARs. Flat: BVAR with a flat prior; MN: BVAR with the Minnesota prior; PLR: BVAR with the prior for the long run.

Priors For Long-Run Dynamics – The Basic Idea

- Write VAR in VECM form:

$$\Delta y_t = \Pi_0 + \Pi_* y_{t-1} + \sum_{j=1}^{p-1} \Pi_i \Delta y_{t-j} + u_t \tag{30}$$

where $\Pi_* = \alpha\beta'$.

- Reasonable prior for columns of α will depend on the rows of β':
  - if i'th row of β' corresponds to a linear combination that is stationary, then it makes sense to choose a prior for i'th column of α with mass away form zero.
  - if i'th row of β' corresponds to a linear combination that is non-stationary, then it makes sense to choose a prior for i'th column of α with mass away form zero.
- See paper for details on how to implement this.

Spike-and-Slab Priors: Sparse versus Dense Models

Reference: Giannone, Lenza, Primiceri (2021): "Economic Prediction With Big Data: The Illusion of Sparsity," *Econometrica*.

- **Sparse models:** only a few predictors are relevant.
- **Dense models:** many predictors are relevant but only have small individual effects.
- **Model:**

$$y_t = x_t'\phi + z_t'\beta + u_t. \tag{31}$$

Here x_t's are included in all specifications (low dimensional), z_t's are optional (high dimensional).

- Prior – part 1:

$$p(\sigma^2) \propto \frac{1}{\sigma^2}, \quad \phi \propto c. \tag{32}$$

Spike-and-Slab Priors: Sparse versus Dense Models

- Prior – part 2: "spike and slab":

$$\beta_i|(\sigma^2,\gamma^2,q) \sim \begin{cases} N(0,\sigma^2\gamma^2) & \text{with prob. } q \\ 0 & \text{with prob. } 1-q \end{cases} \tag{33}$$

- For $q=1$ we obtain our "standard" prior ("Ridge Regression").

- Rewrite prior as

$$\beta_i|(\sigma^2,\gamma^2,\nu_i) \sim N(0,\sigma^2\gamma^2\nu_i), \quad \nu_i \sim \text{Bernoulli}(q). \tag{34}$$

- By changing the mixing distribution, we can generate a wide variety of priors, including a Bayesian version of LASSO.

Spike-and-Slab Priors: Sparse versus Dense Models

- Standardize and orthogonalize the regressors x_t prior to the estimation.
- To specify a prior on the hyperparameters (q, γ^2) they suggest to define

$$R^2(\gamma^2, q) = \frac{qk\gamma^2\bar{\sigma}_z^2}{qk\gamma^2\bar{\sigma}_z^2 + 1} \tag{35}$$

  where k is the number of regressors z and $\bar{\sigma}_z^2$ is the average sample variance of the z_j's.

- The prior takes the form

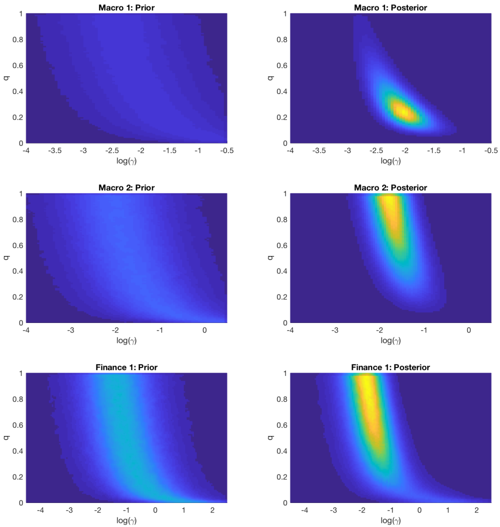
$$q \sim \text{Beta}(a, b), \quad R^2 \sim \text{Beta}(A, B). \tag{36}$$

- The paper works out the posterior.

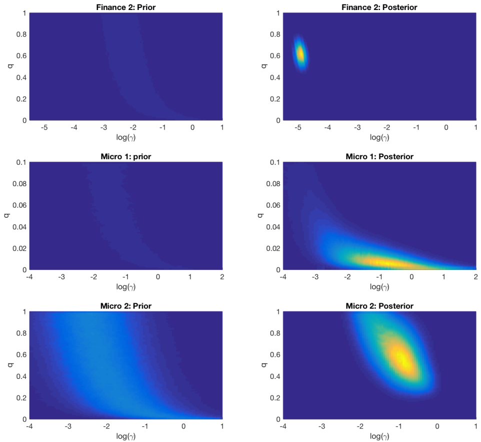
Table 1. Description of the datasets.

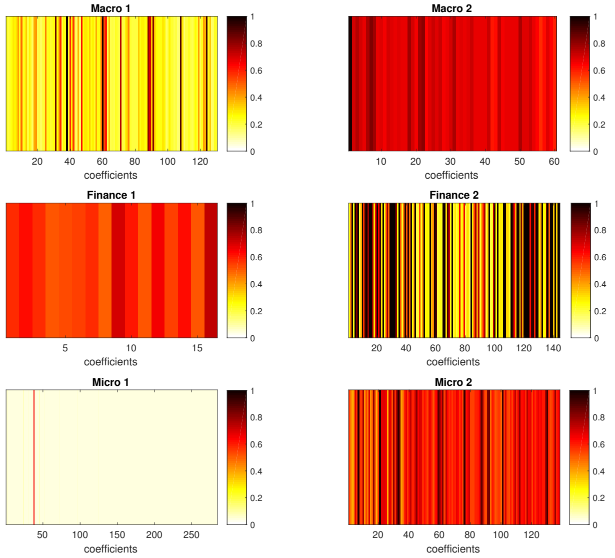
	Dependent variable	**Possible predictors**	**Sample**
Macro 1	Monthly growth rate of US industrial production	130 lagged macroeconomic indicators	659 monthly time-series observations, from February 1960 to December 2014
Macro 2	Average growth rate of GDP over the sample 1960-1985	60 socio-economic, institutional and geographical characteristics, measured at pre-60s value	90 cross-sectional country observations
Finance 1	US equity premium (S&P 500)	16 lagged financial and macroeconomic indicators	58 annual time-series observations, from 1948 to 2015
Finance 2	Stock returns of US firms	144 dummies classifying stock as very low, low, high or very high in terms of 36 lagged characteristics	1400k panel observations for an average of 2250 stocks over a span of 624 months, from July 1963 to June 2015
Micro 1	Per-capita crime (murder) rates	Effective abortion rate and 284 controls including possible covariate of crime and their transformations	576 panel observations for 48 US states over a span of 144 months, from January 1986 to December 1997
Micro 2	Number of pro-plaintiff eminent domain decisions in a specific circuit and in a specific year	Characteristics of judicial panels capturing aspects related to gender, race, religion, political affiliation, education and professional history of the judges, together with some interactions among the latter, for a total of 138 regressors	312 panel circuit/year observations, from 1975 to 2008

q is Slab Probability and γ is Width



q is Slab Probability and γ is Width





Macro 1
20 40 60 80 100 120
coefficients
Macro 2
10 20 30 40 50 60
coefficients
Finance 1
5 10 15
coefficients
Finance 2
20 40 60 80 100 120 140
coefficients
Micro 1
50 100 150 200 250
coefficients
Micro 2
20 40 60 80 100 120
coefficients
0 0.2 0.4 0.6 0.8 1

Time-Varying Coefficients

- Most Common Versions of TVP Models:
  - Parameters follow AR law of motion.
  - Parameters follow regime switching process.

Example: A Time-Varying Inflation Target in an AR Model

- Suppose inflation evolves according to:

$$\pi_t = \pi_t^* + \tilde{\pi}_t \tag{37}$$

where

$$\tilde{\pi}_t = \rho\tilde{\pi}_{t-1} + \sigma_\epsilon \epsilon_t, \quad \pi_t^* = \pi_{t-1}^* + \sigma_\eta \eta_t. \tag{38}$$

- This looks like a state-space model:

$$y_t = \begin{bmatrix} 1 & 1 \end{bmatrix} s_t \tag{39}$$

$$s_t = \begin{bmatrix} \pi_t^* \\ \tilde{\pi}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix} s_{t-1} + \begin{bmatrix} \sigma_\eta & 0 \\ 0 & \sigma_\epsilon \end{bmatrix} \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}. \tag{40}$$

Estimation: Gibbs-Sampling Algorithm

- Generate draws from $p(\theta, S_{1:T}|Y_{1:T})$ using Carter and Kohn (1994)'s approach.
- Gibbs-sampling algorithm iterates over the conditional posteriors of θ and $S_{1:T}$.
- Recall the linear Gaussian state space representation

$$\begin{aligned} y_t &= A + Bs_t + u_t, \quad u_t \sim N(0, H) \\ s_t &= \Phi s_{t-1} + e_t, \quad e_t \sim N(0, Q) \end{aligned} \tag{41}$$

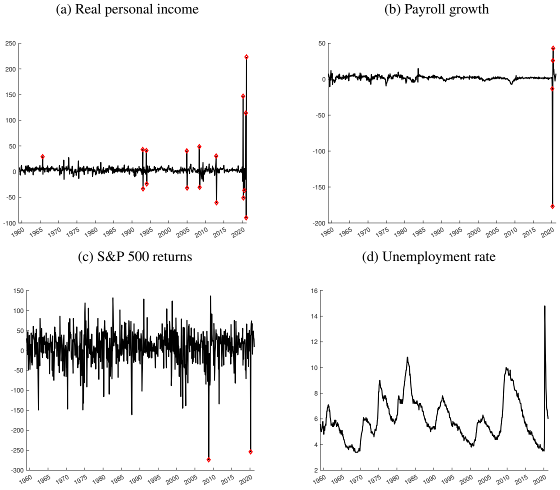
  with $\theta = (A, B, H, \Phi, Q)$
- For $i = 1, ..., n_{sim}$
  (a) Draw $\theta^{(i)}$ from $p\left(\theta \mid Y_{1:T}, S_{1:T}^{(i-1)}\right)$
    - Conditional on $S_{1:T}^{(i-1)}$, drawing θ is a standard linear regression
    - (Measurement) $y_t = A + Bs_t + u_t$
    - (Transition) $s_t = \Phi s_{t-1} + e_t$

  (b) Draw $S_{1:T}^{(i)}$ from $p\left(S_{1:T} \mid Y_{1:T}, \theta^{(i)}\right)$
    - Kalman / simulation smoother

COVID-19 Outliers

Reference: Carriero, Clark, Marcellino, and Mertens (2022): "Addressing COVID-19 Outliers in BVARs with Stochastic Volatilities," *Review of Economics and Statistics*.

- Main goal: adjust BVAR-SV so that it can handle outliers.
- Problem: once an outlier is observed, estimate of SV σ_t rises and is expected to stay high for a while.



Note: Data for selected time series, with data transformations as listed in Table 1. Red dots denote observations that are more than five times the inter-quartile range away from the series median.

COVID-19 Outliers

- Consider the following VAR-SV:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + A^{-1}\Sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, I). \tag{42}$$

- Extreme outliers:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + A^{-1} O_t \Sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, I). \tag{43}$$

where O_t is diagonal with elements

$$o_{it} = \begin{cases} 1 & \text{with prob. } 1 - p_j \\ U[2, 20] & \text{with prob. } p_j \end{cases} \tag{44}$$

- Student t distribution:

$$o_{it} \sim IG \tag{45}$$

- And you can combine the two...

VARs with Censored Variables

References: Aruoba, Mlikota, Schorfheide, and Villalvazo (2022, Journal of Econometrics): "SVARs With Occasionally-Binding Constraints." Mavroeidis (2021, Econometrica): "Identification at the Zero Lower Bound."

- **Example:** nominal interest rates are constrained by an effective lower bound. Seems easy:

$$R_t = \max\left\{R_t^*, 0\right\}. \tag{46}$$

- **Preliminaries: two equivalent VAR representations:** $x_t' = [y_{t-1}', \ldots, y_{t-p}', 1]$
  - Φ Representation:

$$\begin{aligned} y_t &= \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_c + \Phi_\epsilon \epsilon_t && (47) \\ y_t' &= x_t'\Phi + \epsilon_t'\Phi^\epsilon, \quad . && (48) \end{aligned}$$

  - A Representation:

$$y_t'A = x_t'B + \epsilon_t', \quad A = (\Phi^\epsilon)^{-1}, \quad B = \Phi A. \tag{49}$$

ELB SVAR Specification

- We use a hybrid $A\Phi$ representation:
  - Monetary policy in the A representation (contemporary variables interact)
  - Private sector equations in the Φ representation, allowed to differ at and away from ELB. Resembles a DSGE model with regime-dependent effect of state variables.
- Partition $y_t' = [y_{1,t}, y_{2,t}']$ and $\epsilon_t' = [\epsilon_{1,t}, \epsilon_{2,t}']$;
  $y_{1,t}$ is interest rate., $\epsilon_{1,t}$ is the monetary policy shock.
- ELB/Censoring:

$$y_{1,t} = \max\{y_{1,t}^*, -\mu_1\} \text{ where } y_{1,t}^* \text{ is the shadow interest rate.} \tag{50}$$

- Assume that both central bank and private sector react to lagged $y_{1,t}$ instead of $y_{1,t}^*$, and private sector reacts to $\epsilon_{1,t}$ even at the ELB.

ELB SVAR Specification

- Monetary policy rule:

$$y_{1,t}^{*}A_{11} + y_{2,t}'A_{21} = x_t'B_{\cdot 1} + \epsilon_{1,t} \text{ with } y_{1,t} = \max\left\{y_{1,t}^{*}, -\mu_1\right\}. \tag{51}$$

- Private-sector behavior / decision rules:

$$y_{2,t}' = x_t'\Phi_{\cdot 2}(s_t) + \underbrace{\epsilon_{1,t}\Phi_{12}^{\epsilon}(s_t) + \epsilon_{2,t}'\Phi_{22}^{\epsilon}(s_t)}_{u_{2,t}'(s_t)}, \quad s_t = \mathbb{I}\{y_{1,t} > -\mu\}. \tag{52}$$

$\Longrightarrow$ Incorporates: (a) censoring of policy rate (b) regime-dependent private-sector behavior.

- Derive Φ representation of monetary policy rule:

$$y_{1,t}^{*} = x_t'\Phi_{\cdot 1}(s_t) + u_{1,t}(s_t). \tag{53}$$

$\Longrightarrow$ Internal consistency requires:

1. If $y_{1,t} > -\mu_1$ ($y_{1,t}^{*} > -\mu_1$), the "1" decision rules have to be active.
2. If $y_{1,t} = -\mu_1$ ($y_{1,t}^{*} \leq -\mu_1$), the "0" decision rules have to be active.

VARs with Censored Variables

- Once the uniqueness restrictions have been imposed, the estimation is relatively straightforward.

- In the AMSV paper we use sequential Monte Carlo for the estimation and apply the framework to a monetary VAR.

What Else???

1. It's a bit like a box with Lego pieces: you can put them together to build models for your empirical analysis.

2. Or you can develop new pieces... much of the literature focuses on: setting up models in a way such that they can be estimated in high-dimensional settings; allowing for more nonlinearities.

Reduced-Form and Structural VARs

- So far, we considered reduced form VARs, e.g.,

$$y_t = \Phi_1 y_{t-1} + u_t, \quad \mathbb{E}[u_t u_t'] = \Sigma \tag{54}$$

- u_t has the interpretation of one-step ahead forecast error.
- Under stationarity we have the MA(∞) representation

$$y_t = (I - \Phi_1 L)^{-1} u_t = \sum_{j=0}^{\infty} \Phi_1^j u_{t-j} = \sum_{j=0}^{\infty} C_j u_{t-j}. \tag{55}$$

- Dynamic macroeconomic theory suggest that the one-step ahead forecast errors are functions of some fundamental shocks, such as technology shocks, preference shocks, or monetary policy shocks.

SVARs

- Let ϵ_t a vector of such fundamental shocks with $\mathbb{E}[\epsilon_t \epsilon_t'] = I$. Assume

$$u_t = \Phi_\epsilon \epsilon_t. \tag{56}$$

- Express the VAR in structural form as:

$$y_t = \Phi_1 y_{t-1} + \Phi_\epsilon \epsilon_t \tag{57}$$

or, equivalently,

$$A y_t = B_1 y_{t-1} + \epsilon_t. \tag{58}$$

- MA representation of y_t in terms of ϵ_t:

$$y_t = \sum_{j=0}^{\infty} \Phi_1^j \Phi_\epsilon \epsilon_{t-j} = \sum_{j=0}^{\infty} C_j \Phi_\epsilon \epsilon_{t-j}. \tag{59}$$

Impulse Responses

What is the effect of a shock (e.g., monetary policy) in period t on y_{t+j}?

$$\begin{aligned} IRF(j) &= \Big\{ \mathbb{E}[y_{t+j}|\epsilon_{t,1}=1,\mathcal{F}_{t-1}] - \mathbb{E}[y_{t+j}|\mathcal{F}_{t-1}], \ldots, \\ &\quad \mathbb{E}[y_{t+j}|\epsilon_{t,n}=1,\mathcal{F}_{t-1}] - \mathbb{E}[y_{t+j}|\mathcal{F}_{t-1}] \Big\} \\ &= \frac{\partial y_{t+j}}{\partial \epsilon_t'} \\ &= C_j \Phi_\epsilon \end{aligned} \tag{60}$$

(Lack of) Identification

- To equate (54) and (57), Φ_ϵ has to satisfy

$$\Phi_\epsilon \Phi_\epsilon' = \Sigma \tag{61}$$

- Φ_ϵ has n^2 elements but Σ only $n(n+1)/2 \Longrightarrow$ **underidentification**.
- Cholesky factorization (Σ_{tr} is lower triangular)

$$\Sigma = \Sigma_{tr}\Sigma_{tr}'. \tag{62}$$

- Let Ω be an orthogonal matrix, meaning that $\Omega\Omega' = \Omega'\Omega = I$. Then

$$u_t = \Sigma_{tr}\Omega\epsilon_t, \tag{63}$$

where Σ_{tr} is identifiable and Ω is not, because:

$$\mathbb{E}[u_t u_t'] = A\Omega\mathbb{E}[\epsilon_t\epsilon_t']\Omega'\Sigma_{tr}' = \Sigma_{tr}\Omega\Omega'\Sigma_{tr}' = \Sigma_{tr}\Sigma_{tr}' = \Sigma.$$

Consequences – Or, Bayesian Inference with Non-Identified Parameters

- The joint distribution of data and parameters is given by

$$p(Y, \Phi, \Sigma, \Omega) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma)p(\Omega|\Phi, \Sigma). \tag{64}$$

- Integrating the joint density with respect to Ω yields

$$p(Y, \Phi, \Sigma) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma). \tag{65}$$

$\Longrightarrow$ posterior of reduced form parameters (Φ, Σ) is not affected by the presence of non-identifiable matrix Ω.

- The conditional posterior density of Ω can be calculated as follows:

$$p(\Omega|Y, \Phi, \Sigma) = \frac{p(Y, \Phi, \Sigma)p(\Omega|\Phi, \Sigma)}{\int p(Y, \Phi, \Sigma)p(\Omega|\Phi, \Sigma)d\Omega} = p(\Omega|\Phi, \Sigma). \tag{66}$$

Algorithm

For $i = 1, \ldots, N$:

1. Draw (Φ^i, Σ^i) from the posterior $p(\Phi, \Sigma | Y)$.
2. Draw Ω^i from the conditional prior distribution $p(\Omega | \Phi^i, \Sigma^i)$. $\square$

Interpretation of SVAR Literature

- Hard or soft restrictions on $\Omega|(\Phi, \Sigma)$. Two surveys: Stock and Watson (2001, JEP) and Ramey (2016, Handbook of Macroeconomics).
- In practice, it is sometimes easier to impose restrictions directly on Φ_ϵ or on A in (58).
- Literature argues about how to impose priors on Ω. Results are sensitive to these priors because there is no updating.
- Inference on the parameter versus the identified set; robust Bayesian inference?