Stata Mexican Conference

# Introduction to Bayesian VAR models in Stata

Gustavo Sánchez

StataCorp LLC

October 14, 2021
Puebla, México

# Outline

1. Bayesian analysis: Brief overview
   - The general idea
   - The method
2. The Stata tools for Bayesian analysis
3. Bayesian linear regression
4. VAR model
5. Bayesian VAR
   - Estimation
   - Postestimation

# The general idea

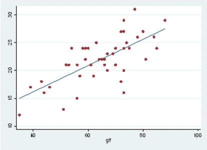
## Frequentist

Theoretical Model →

```
. list ln_wage union hours wks_work tenure race grade ttl_exp in 1/15, noobs
```

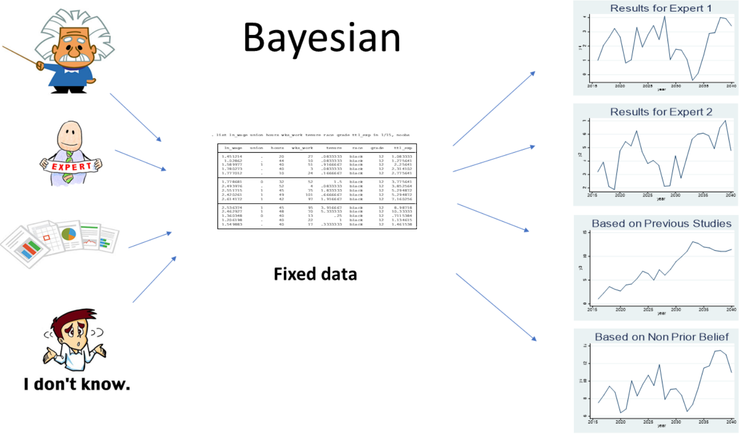
| ln_wage | union | hours | wks_work | tenure | race | grade | ttl_exp |
|---|---|---|---|---|---|---|---|
| 1.451214 | . | 20 | 27 | .0833333 | black | 12 | 1.083333 |
| 1.02862 | . | 44 | 10 | .0833333 | black | 12 | 1.275641 |
| 1.589977 | 1 | 40 | 51 | .9166667 | black | 12 | 2.25641 |
| 1.780273 | . | 40 | 3 | .0833333 | black | 12 | 2.314102 |
| 1.777012 | . | 10 | 24 | .1666667 | black | 12 | 2.775641 |
| 1.778681 | 0 | 32 | 52 | 1.5 | black | 12 | 3.775641 |
| 2.493976 | . | 52 | 4 | .0833333 | black | 12 | 3.852564 |
| 2.551715 | 1 | 45 | 75 | 1.833333 | black | 12 | 5.294872 |
| 2.420261 | 1 | 49 | 101 | .6666667 | black | 12 | 5.294872 |
| 2.614172 | 1 | 42 | 97 | 1.916667 | black | 12 | 7.160256 |
| 2.536374 | 1 | 45 | 95 | 3.916667 | black | 12 | 8.98718 |
| 2.462927 | 1 | 48 | 70 | 5.333333 | black | 12 | 10.33333 |
| 1.360348 | 0 | 40 | 13 | .25 | black | 12 | .7115384 |
| 1.206198 | . | 40 | 22 | 1 | black | 12 | 1.134615 |
| 1.549883 | . | 40 | 17 | .3333333 | black | 12 | 1.461538 |

→




**Random sample**

# The general idea



# The method

# The method

- Inverse law of probability (Bayes' Theorem):

$$p(\theta|y) = \frac{p(y|\theta)\,p(\theta)}{p(y)} = \frac{f(y;\theta)\,\pi(\theta)}{f(y)}$$

Where:

$f(y;\theta)$: probability density function for y given $\theta$.

$\pi(\theta)$: prior distribution for $\theta$

- The marginal distribution of y, f(y), does not depend on $\theta$; then we can write the fundamental equation for Bayesian analysis:

$$p(\theta|y) \propto L(\theta;y)\,\pi(\theta)$$

Where:

$L(\theta;y)$: likelihood function of the parameters given the data.

# The method

- Inverse law of probability (Bayes' Theorem):

$$p(\theta|y) = \frac{p(y|\theta)\,p(\theta)}{p(y)} = \frac{f(y;\theta)\,\pi(\theta)}{f(y)}$$

Where:

$f(y;\theta)$: probability density function for y given $\theta$.

$\pi(\theta)$: prior distribution for $\theta$

- The marginal distribution of y, f(y), does not depend on $\theta$; then we can write the fundamental equation for Bayesian analysis:

$$p(\theta|y) \propto L(\theta;y)\,\pi(\theta)$$

Where:

$L(\theta;y)$: likelihood function of the parameters given the data.

# The method

- Inverse law of probability (Bayes' Theorem):

$$p(\theta|y) = \frac{p(y|\theta)\,p(\theta)}{p(y)} = \frac{f(y;\theta)\,\pi(\theta)}{f(y)}$$

Where:

$f(y;\theta)$: probability density function for y given $\theta$.

$\pi(\theta)$: prior distribution for $\theta$

- The marginal distribution of y, f(y), does not depend on $\theta$; then we can write the fundamental equation for Bayesian analysis:

$$p(\theta|y) \propto L(\theta;y)\,\pi(\theta)$$

Where:

$L(\theta;y)$: likelihood function of the parameters given the data.

## The method

- Some prior-likelihood combinations have closed form solution.

- What about the cases with non-closed solutions, or more complex distributions?
  - Integration is performed via simulation.
  - We need to use intensive computational simulation tools to find the posterior distribution in most cases.

  - Markov chain Monte Carlo (MCMC) methods are the current standard in most software. Stata implements two alternatives:
    - Metropolis–Hastings (MH) algorithm
    - Gibbs sampling

# The method

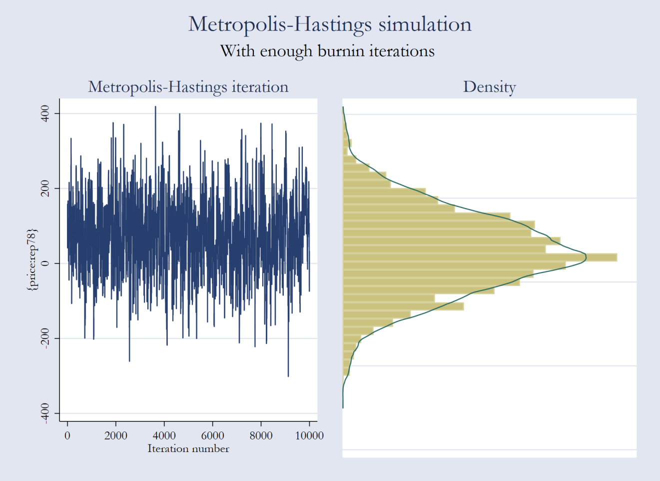
- Links for Bayesian analysis and MCMC on our YouTube channel:
  - Introduction to Bayesian statistics, part 1: The basic concepts

    https://www.youtube.com/watch?v=0F0QoMCSKJ4&feature=youtu.be
  - Introduction to Bayesian statistics, part 2: MCMC and the Metropolis–Hastings algorithm.

    https://www.youtube.com/watch?v=OTO1DygELpY&feature=youtu.be

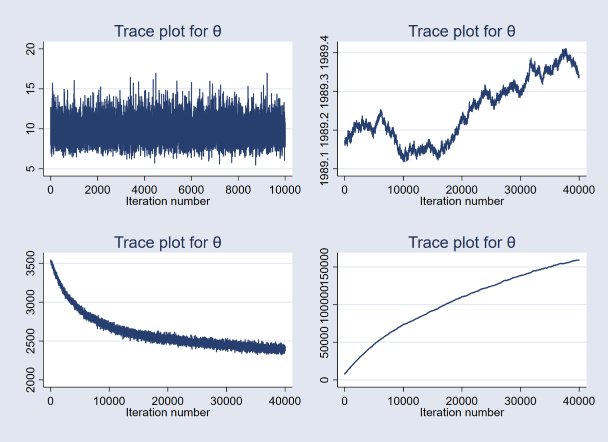
# The method

- Metropolis–Hastings simulation
  - The trace plot illustrates the sequence of accepted proposal states for a simulation with enough burnin iterations.



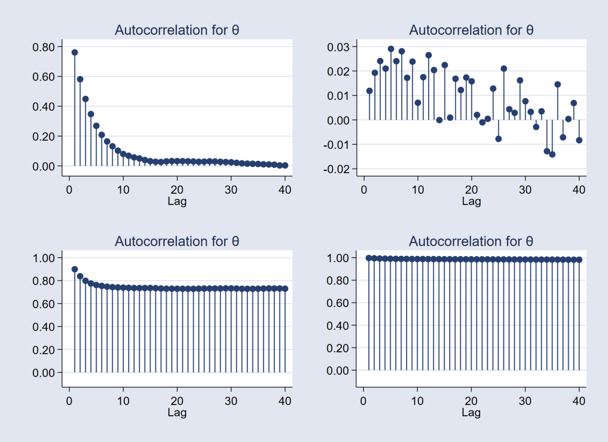
# The method

- We expect to obtain a stationary sequence when convergence is achieved.



# The method

- An efficient MCMC should have small autocorrelation.
- We expect autocorrelation to become negligible after a few lags.



# The Stata tools for Bayesian regression

## Stata's convenient syntax: `bayes:`

```
regress y x1 x2 x3

bayes: regress y x1 x2 x3
```

```
mixed y x1 x2 x3 || region:

bayes: mixed y x1 x2 x3 || region:
```

```
var y1 y2 y3, lags(1/4)

bayes: var y1 y2 y3, lags(1/4)
```

# Stata's Bayesian suite consists of the following commands

| *Command* | *Description* |
|---|---|
| **Estimation** | |
| `bayes:` | Bayesian regression models using the `bayes` prefix |
| `bayesmh` | General Bayesian models using MH |
| `bayesmh` *evaluators* | User-defined Bayesian models using MH |
| **Postestimation** | |
| `bayesgraph` | Graphical convergence diagnostics |
| `bayesstats ess` | Effective sample sizes and more |
| `bayesstats grubin` | Gelman–Rubin convergence diagnostics |
| `bayesstats summary` | Summary statistics |
| `bayesstats ic` | Information criteria and Bayes factors |
| `bayestest model` | Model posterior probabilities |
| `bayestest interval` | Interval hypothesis testing |
| `bayespredict` | Bayesian predictions (available only after `bayesmh`) |
| `bayesstats ppvalues` | Bayesian predictive *p*-values (available only after `bayesmh`) |
| **Added in latest version** | |
| `bayes:var` | Bayesian VAR models |
| `bayes:dsge` | Bayesian DSGE models |
| `bayes:xt` | Bayesian panel data models |

# Example 1: Infant mortality in México

- Let's work with a simple linear regression for the infant mortality in México as a function of a couple of macroeconomic variables:
  - We used `import fred` to get data from the Federal Reserve Economic Data (FRED) for México on infant mortality, GDP per capita, inflation.
  - Let's consider the following model specification:

$$mortality = \alpha_1 + \beta_{gdp\_cap} * gdp\_cap + \beta_{inflation} * inflation + \epsilon_1$$

Where:

| | |
|---|---|
| mortality | : Infant mortality rate for México |
| gdp_cap | : Constant GDP per capita for México. |
| inflation | : Inflation for México. |

## Example 1: Linear Regression

- Linear regression with the `bayes:` prefix

```
. bayes, rseed(123): regress mortality gdp_cap inflation
```

- Equivalent model with `bayesmh`

```
. bayesmh mortality gdp_cap inflation, rseed(123)   ///
>     likelihood(normal({sigma2}))                  ///
>     prior({mortality:gdp_cap}, normal(0,10000))   ///
>     prior({mortality:inflation}, normal(0,10000)) ///
>     prior({mortality:_cons}, normal(0,10000))     ///
>     prior({sigma2}, igamma(.01,.01))              ///
>     block({mortality:gdp_cap inflation _cons})    ///
>     block({sigma2})
```

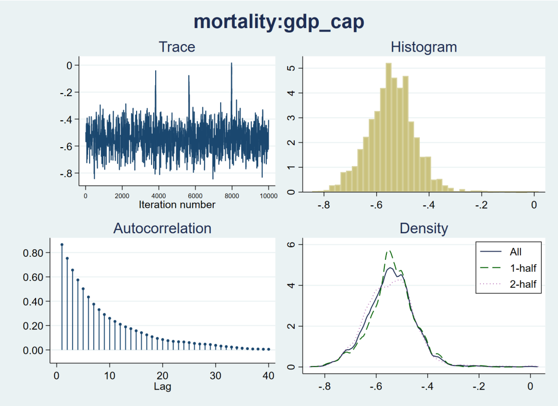
# Example 1: `bayes:` prefix

```
. bayes, rseed(123) saving(mortality,replace): ///
>         regress mortality_mx gdp_cap_mx inf_mx

Burn-in ...
Simulation ...
file mortality.dta saved.
Model summary
------------------------------------------------------------------------------
Likelihood:
  mortality_mx ~ regress(xb_mortality_mx,{sigma2})

Priors:
  {mortality_mx:gdp_cap_mx inf_mx _cons} ~ normal(0,10000)                 (1)
                                {sigma2} ~ igamma(.01,.01)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_mortality_mx.

Bayesian linear regression                       MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =         20
                                                 Acceptance rate  =      .3287
                                                 Efficiency:  min =     .06442
                                                              avg =     .07066
Log marginal-likelihood = -59.016584                          max =     .08461

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
mortality_mx |
  gdp_cap_mx | -.5415947   .0937332   .003693  -.5424703  -.7232312  -.3585976
      inf_mx |   .767075   .2742269   .009428    .765858   .2590686    1.28898
       _cons |  65.44467   9.555959   .373734   65.52571   46.82975   84.16327
-------------+----------------------------------------------------------------
      sigma2 |  2.857122   1.184845   .045357   2.579777    1.38332   5.930879
------------------------------------------------------------------------------
Note: Default priors are used for model parameters.
```

# Example 1: `bayesgraph`

- We can use `bayesgraph` to look at the trace, the correlation, and the density. For example:

```
.   bayesgraph diagnostic {gdp_cap}
```

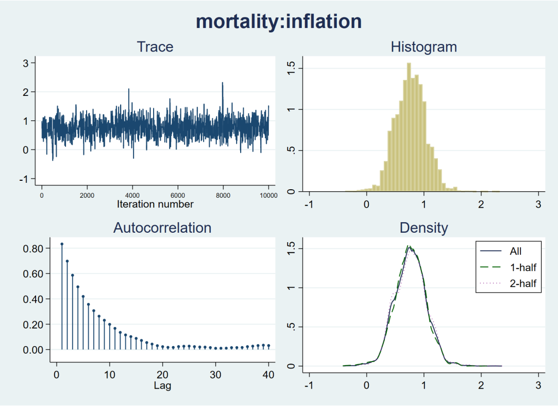



- The trace indicates that convergence was achieved
- Correlation becomes negligible after 15 periods

# Example 1: `bayesgraph`

- We can use `bayesgraph` to look at the trace, the correlation, and the density. For example: inflation

```
. bayesgraph diagnostic {inflation}
```

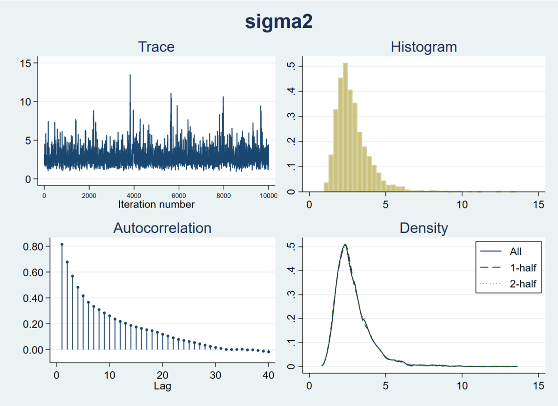



- The trace indicates that convergence was achieved
- Correlation becomes negligible after 15 periods

# Example 1: `bayesgraph`

- We can use `bayesgraph` to look at the trace, the correlation, and the density. For example:

```
. bayesgraph diagnostic {sigma2}
```




- The trace indicates that convergence was achieved
- Correlation becomes negligible after 15 periods

# Vector Autoregressive (VAR) Models

# Probabilities available with Bayesian VARs

- Can we compute probabilities for events associated to multiple equation forecasts?

```
. collect preview
```

| | 2010Q1 | 2010Q1 | 2010Q1-Q2 |
|---|---|---|---|
| Event | | | |
| inflation_over_5=1 | .6778 | | |
| infl_over_5__exchrate_chg_over_3=1 | | .4554 | |
| exchrate_change_over_3=1 | | | .4303 |

- Probability forecasting (Garrat et al. (2006)) allows defining events for forecasted variables conditional in the estimation sample.
  - Forecasts are based on econometric models subject to uncertainty on the future, on the parameters, on the model, and also on the policies.
  - See an example in Sanchez and Zavarce (2013) for probability forecast accounting for future uncertainty.
- The Bayesian approach allows obtaining probabilities for events based on parameters and future uncertainty.

# Probabilities available with Bayesian VARs

- Can we compute probabilities for events associated to multiple equation forecasts?

```
. collect preview
```

| | 2010Q1 | 2010Q1 | 2010Q1-Q2 |
|---|---|---|---|
| Event | | | |
| inflation_over_5=1 | .6778 | | |
| infl_over_5__exchrate_chg_over_3=1 | | .4554 | |
| exchrate_change_over_3=1 | | | .4303 |

- Probability forecasting (Garrat et al. (2006)) allows defining events for forecasted variables conditional in the estimation sample.
  - Forecasts are based on econometric models subject to uncertainty on the future, on the parameters, on the model, and also on the policies.
  - See an example in Sanchez and Zavarce (2013) for probability forecast accounting for future uncertainty.
- The Bayesian approach allows obtaining probabilities for events based on parameters and future uncertainty.

# Vector Autoregressive Models VAR

- VARs are extensions of AR(p) models for vector valued dependent variables with no structural form.
- A VAR model can be written as:

$$\mathbf{y_t} = \mathbf{v} + \mathbf{A_1 y_{t-1}} + \ldots + \mathbf{A_p y_{t-p}} + \mathbf{C_0 x_t} + \mathbf{C_1 x_{t-1}} + \ldots + \mathbf{C_s x_{t-s}} + \mathbf{u_t}$$

Where:

$\mathbf{y}_t = (y_{1t}, ..., y_{Kt})$ is a Kx1 random vector
$\mathbf{A_1}$ through $\mathbf{A_p}$ are KxK matrices of parameters.
$\mathbf{x_t}$ is an Mx1 vector of exogenous variables
$\mathbf{C_0}$ through $\mathbf{C_s}$ are KxM matrices of parameters.
$\mathbf{v}$ is a Kx1 vector of parameters
$\mathbf{u_t}$ is a vector assumed to be white noise:

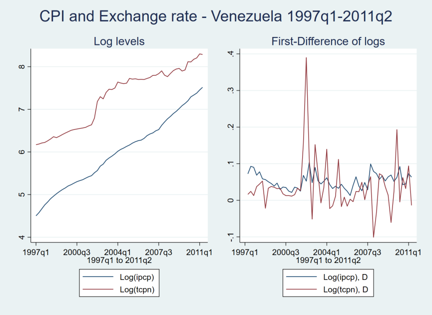
$$\begin{aligned} E(\mathbf{u_t}) &= \mathbf{0} \\ E(\mathbf{u_t u_t}') &= \mathbf{\Sigma} \\ E(\mathbf{u_t u_s}') &= \mathbf{0} \;;\quad t \neq s \end{aligned}$$

- The number of coefficients is quadratic to the number of dependent variables and proportional the number of lags.

# Example 2: VAR model for CPI and exchange rate

- Data for consumer price index and exchange rate:

```
 . describe

Contains data from C:\Users\gas\Documents\mexico\mexico21\data_bvar21_modified.dta
 Observations:           58                  Data for Venezuela
    Variables:            5                  13 Oct 2021 00:56
-------------------------------------------------------------------------------------
Variable      Storage   Display    Value
    name         type    format    label      Variable label
-------------------------------------------------------------------------------------
ipcp            float   %9.0g                 Consumer price index for Venezuela
tcpn            float   %9.0g                 Exchange rate Bs/US$ for Venezuela
quarter         float   %tq                   1997q1 to 2011q2
lipcp           double  %10.0g                Log(ipcp)
ltcpn           double  %10.0g                Log(tcpn)
-------------------------------------------------------------------------------------
Sorted by: quarter
```



# Example 2: Estimation with the `var` command

```
. var D.ltcpn D.lipcp if tin(1999q1,2009Q4),lags(1/2) vsquish
Vector autoregression
Sample:  1999q1 thru 2009q4                    Number of obs     =         44
Log likelihood =   174.0957                    AIC               =  -7.458896
FPE            =   1.98e-06                    HQIC              =  -7.308518
Det(Sigma_ml)  =   1.25e-06                    SBIC              =  -7.053398
Equation           Parms      RMSE     R-sq      chi2     P>chi2
----------------------------------------------------------------
D_ltcpn               5     .073623   0.1294   6.541196   0.1622
D_lipcp               5     .017428   0.3239   21.07584   0.0003
----------------------------------------------------------------

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
D_ltcpn      |
       ltcpn |
         LD. |   .3345647   .1480213     2.26   0.024     .0444482    .6246811
        L2D. |  -.2544125   .1560179    -1.63   0.103    -.5602019    .0513769
       lipcp |
         LD. |  -.2390203   .5925201    -0.40   0.687    -1.400338    .9222976
        L2D. |   .3130268   .5830365     0.54   0.591    -.8297037    1.455757
       _cons |   .0304274   .0329316     0.92   0.356    -.0341174    .0949722
-------------+----------------------------------------------------------------
D_lipcp      |
       ltcpn |
         LD. |   .0843611   .0350404     2.41   0.016     .0156832    .1530391
        L2D. |  -.0259031   .0369334    -0.70   0.483    -.0982913    .0464851
       lipcp |
         LD. |   .1827309   .1402647     1.30   0.193    -.0921828    .4576445
        L2D. |   .4045934   .1380196     2.93   0.003     .1340798    .6751069
       _cons |   .0179177   .0077958     2.30   0.022     .0026383    .0331971
------------------------------------------------------------------------------
```

## Bayesian VAR models with `bayes:var`

- Overparameterization in VAR models is particular problematic with small samples.
- Bayesian VAR allows shrinking the vector of regression coefficients by controlling the effective number of lags through the priors.
- The Minnesota family of priors represent a flexible specification that allows the expert's knowledge to be incorporated in the estimation.
- Bayes factors can be used to select the number of lags, and also the exogenous variables.

# Bayesian VAR models with `bayes:var`

- The Bayesian approach to fit VAR models assigns prior distributions to all the regression parameters:
    - The likelihood is derived from the linear specification

      $$\mathbf{y_t} = \mathbf{A_1 y_{t-1}} + \ldots + \mathbf{A_p y_{t-p}} + \mathbf{C x_t} + \mathbf{u_t} \quad ; \quad \mathbf{u_t} \sim N(0, \Sigma)$$

    - For the regression coefficients $\beta = vec(\mathbf{C}, \mathbf{A_1}, ..., \mathbf{A_p})$ the prior corresponds to a multivariate normal:

      $$\beta | \boldsymbol{y} \sim N\left(\boldsymbol{\beta_0}, \boldsymbol{\Omega}\right)$$

    - For the regression covariance matrix $\boldsymbol{\Sigma}$ the prior distribution would be either inverse Wishart or Jeffreys.

# Minnesota priors

- `bayes:var` has four prior families alternatives:
  - Minnesota prior with fixed covariance $\boldsymbol{\Sigma}$

    `bayes, minnfixedcovprior... : var...`

  - Conjugate Minnesota prior (The default)

    `bayes, minnconjprior... : var...`

  - Minnesota prior for $\beta$ and inverse-Wishart prior for $\boldsymbol{\Sigma}$

    `bayes, minnwishprior... : var...`

  - Minnesota prior for $\beta$ and Jeffreys prior for $\boldsymbol{\Sigma}$

    `bayes, minnjeffprior... : var...`

# Original Minnesota prior with fixed covariance

- Doan, Litterman, and Sims 1984 and Litterman 1986, assumed a known fixed-error covariance matrix.
- Prior for vector of coefficients:

$$\beta \sim N(\boldsymbol{\beta_0}, \boldsymbol{\Omega_0})$$

- $\boldsymbol{\Omega_0}$ is diagonal (i.e. no correlation between the coefficients in $\beta$).

- The covariance matrix of the error is known ($\boldsymbol{\Sigma} = \boldsymbol{\Sigma_0}$):

$$\mathbf{u} \sim N(\mathbf{0}, \boldsymbol{\Sigma_0} \otimes \boldsymbol{I_T})$$

# Original Minnesota prior with fixed covariance

- Example: $\mathbf{y_{1t}} = a_{11} y_{1t-1} + a_{12} y_{2t-1} + c1 + u_{1t}$
  $\mathbf{y_{2t}} = a_{21} y_{1t-1} + a_{22} y_{2t-1} + c2 + u_{1t}$

- Independent error terms with known fixed variances

$$u_1 \sim N\left(0, \hat{\sigma}_1^2\right),\ u_2 \sim N\left(0, \hat{\sigma}_2^2\right)$$

- Prior expectations: $E[a_{11}] = E[a_{22}] = 1$
  $E[a_{12}] = E[a_{21}] = E[c_1] = E[c_2] = 0$

- Prior variances:
  $Var[a_{11}] = Var[a_{22}] = \lambda_1^2$
  $Var[a_{12}] = Var[a_{21}] \sim \lambda_1^2 \lambda_2^2$
  $Var[c_1] = Var[c_2] \sim \lambda_1^2 \lambda_4^2$

  Assuming $\lambda_3 = 1$

# `bayes:var` with fixed covariance

- Default estimation with `minnfixedcovprior`. Original Minnesota prior with $\lambda_1 = 0.1$ , $\lambda_2 = 0.5$ , $\lambda_4 = 100$

```
bayes, minnfixedcovprior:var y1 y2,lags(1)
```

- Increase the self-tightness, for example, set $\lambda_1 = 1$

```
bayes, minnfixedcovprior(selftight(1)):    ///
      var y1 y2,lags(1)
```

- Specify zero-mean priors for all the coefficients:

```
bayes, minnfixedcovprior(mean(0,0)):      ///
      var y1 y2,lags(1)
```

- Reduce the exogenous-variables tightness parameter, for example, set it to $\lambda_4 = 50$

```
bayes, minnfixedcovprior(exogtight(50)):  ///
      var y1 y2,lags(1)
```

# Example 3: Default values for fixed-error covariance

```
 . matrix b0 = J(1,2,0)
. bayes,minnfixedcovprior(mean(b0)) rseed(123) dryrun: ///
>          var D.ltcpn D.lipcp if tin(1999q1,2009Q4),lags(1/2)
Model summary
------------------------------------------------------------------------------
Likelihood:
  D_ltcpn D_lipcp ~ mvnormal(2,xb_D_ltcpn,xb_D_lipcp,_Sigma0)
Priors:
  {D_ltcpn:L( 2D).ltcpn}                                                   (1)
  {D_ltcpn:L( 2D).lipcp}                                                   (1)
         {D_ltcpn:_cons}                                                   (1)
  {D_lipcp:L( 2D).ltcpn}                                                   (2)
  {D_lipcp:L( 2D).lipcp}                                                   (2)
         {D_lipcp:_cons} ~ minnesota(2,2,1,b0,_Sigma0,.1,.5,1,100)         (2)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_D_ltcpn.
(2) Parameters are elements of the linear form xb_D_lipcp.
```

# Example 3: Default values for fixed-error covariance

```
. matrix b0 = J(1,2,0)
. bayes,minnfixedcovprior(mean(b0)) rseed(321) ///
>       noheader nomodelsummary:              ///
>       var D.ltcpn D.lipcp if tin(1999q1,2009Q4),lags(1/2)

Burn-in ...
Simulation ...
```

| | Mean | Std. dev. | MCSE | Median | Equal-tailed [95% cred. | interval] |
|---|---|---|---|---|---|---|
| **D_ltcpn** | | | | | | |
| ltcpn | | | | | | |
| LD. | .0891295 | .0814777 | .000815 | .0886701 | -.0707556 | .2498772 |
| L2D. | -.0208036 | .047453 | .000475 | -.021048 | -.1137706 | .0725059 |
| lipcp | | | | | | |
| LD. | -.0281904 | .1875232 | .00185 | -.0292024 | -.3915603 | .3362476 |
| L2D. | -.0013577 | .0979527 | .00098 | -.0009393 | -.1959376 | .1905186 |
| _cons | .0351268 | .0146977 | .000147 | .0351596 | .0056313 | .0636382 |
| **D_lipcp** | | | | | | |
| ltcpn | | | | | | |
| LD. | .0086841 | .0119654 | .00012 | .0086399 | -.0148593 | .0319786 |
| L2D. | .0007114 | .0062214 | .000062 | .0006591 | -.0113796 | .0128985 |
| lipcp | | | | | | |
| LD. | .1245865 | .0794112 | .000794 | .124164 | -.0309254 | .2821711 |
| L2D. | .0483971 | .0471263 | .000471 | .0488689 | -.0434823 | .1397729 |
| _cons | .0394916 | .005068 | .000051 | .0395036 | .029676 | .0495442 |

# Conjugate Minnesota prior (Default for `bayes:var`)

- Example: $$\mathbf{y_{1t}} = a_{11} y_{1t-1} + a_{12} y_{2t-1} + c1 + u_{1t}$$ $$\mathbf{y_{2t}} = a_{21} y_{1t-1} + a_{22} y_{2t-1} + c2 + u_{1t}$$

- Error terms have a multivariate normal distribution

$$(\mathbf{u_1}, \mathbf{u_2}) \sim MVN\left((0,0), \boldsymbol{\Sigma}\right)$$

- Prior expectations: $E[a_{11}] = E[a_{22}] = 1$ $$E[a_{12}] = E[a_{21}] = E[c_1] = E[c_2] = 0$$

- Prior variances: $Var[a_{11}, a_{12}, a_{21}, a_{22}, c_1, c_2] = \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi_0}$

$$\boldsymbol{\Sigma} \sim InvWishart\left(\alpha_0, \boldsymbol{S_0}\right) \quad ; \quad \alpha_0 = K + 2$$
$$\boldsymbol{S_0} = (\alpha_0 - K - 1)\boldsymbol{\Sigma_0}$$

$$\boldsymbol{\Phi_0} = diag\left(\frac{\lambda_1^2}{\hat{\sigma}_1^2}, \frac{\lambda_1^2}{\hat{\sigma}_2^2}, \lambda_1^2 \lambda_4^2\right)$$

## `bayes:var` with conjugate Minnesota prior

- Default estimation with `minnconjprior`. Original Minnesota prior with $\lambda_1 = 0.1$ and $\lambda_4 = 100$

```
bayes, minnconjprior:var y1 y2,lags(1)
```

- Increase the self-tightness, for example, set $\lambda_1 = 1$

```
bayes, minnconjprior(selftight(1)):     ///
        var y1 y2,lags(1)
```

- Specify zero-mean priors for all the coefficients:

```
bayes, minnconjprior(mean(0,0)):        ///
        var y1 y2,lags(1)
```

- Reduce the exogenous-variables tightness parameter, for example, set it to $\lambda_4 = 50$

```
bayes, minnconjprior(exogtight(50)):   ///
        var y1 y2,lags(1)
```

# Example 4: Conjugate prior with self-tightness equal to 1

```
. matrix b0 = J(1,2,0)
. bayes,minnconjprior(mean(b0) selftight(1)) rseed(123) dryrun: ///
>         var D.ltcpn D.lipcp if tin(1999q1,2009Q4)
Model summary
------------------------------------------------------------------------------
Likelihood:
  D_ltcpn D_lipcp ~ mvnormal(2,xb_D_ltcpn,xb_D_lipcp,{Sigma,m})

Priors:
  {D_ltcpn:L( 2D).ltcpn}                                                  (1)
  {D_ltcpn:L( 2D).lipcp}                                                  (1)
         {D_ltcpn:_cons}                                                  (1)
  {D_lipcp:L( 2D).ltcpn}                                                  (2)
  {D_lipcp:L( 2D).lipcp}                                                  (2)
         {D_lipcp:_cons} ~ varconjugate(2,2,1,b0,{Sigma,m},_Phi0)         (2)
               {Sigma,m} ~ iwishart(2,4,_Sigma0)
------------------------------------------------------------------------------
(1) Parameters are elements of the linear form xb_D_ltcpn.
(2) Parameters are elements of the linear form xb_D_lipcp.
```

# Example 4: Conjugate prior / self-tightness equal to 1

```
 . matrix b0 = J(1,2,0)
. bayes,minnconjprior(mean(b0) selftight(1))    ///
>     noheader nomodelsummary rseed(123):        ///
>     var D.ltcpn D.lipcp if tin(1999q1,2009Q4)

Burn-in ...
Simulation ...
```

| | Mean | Std. dev. | MCSE | Median | Equal-tailed [95% cred. | interval] |
|---|---|---|---|---|---|---|
| **D_ltcpn** | | | | | | |
| ltcpn | | | | | | |
| LD. | .3198622 | .1480341 | .001463 | .3206658 | .0270658 | .6109393 |
| L2D. | -.2291803 | .1507129 | .001507 | -.2297776 | -.5236664 | .0669669 |
| lipcp | | | | | | |
| LD. | -.2301976 | .5885461 | .005778 | -.237586 | -1.400547 | .9419769 |
| L2D. | .2699294 | .5640323 | .00564 | .2749025 | -.8348297 | 1.38465 |
| _cons | .031954 | .0329246 | .000329 | .0315799 | -.03232 | .0974368 |
| **D_lipcp** | | | | | | |
| ltcpn | | | | | | |
| LD. | .0807677 | .0353689 | .000354 | .081093 | .0097579 | .1504552 |
| L2D. | -.0214587 | .0355133 | .000355 | -.0210783 | -.0924263 | .0460883 |
| lipcp | | | | | | |
| LD. | .1919819 | .1411463 | .001411 | .1910727 | -.0855342 | .4686913 |
| L2D. | .3709372 | .1343597 | .001342 | .3709623 | .1049788 | .6377188 |
| _cons | .019075 | .0079081 | .000079 | .0191754 | .0034905 | .034522 |
| Sigma_1_1 | .0049435 | .0010658 | .00001 | .0047932 | .0033016 | .0074403 |
| Sigma_2_1 | .0002021 | .0001791 | 1.8e-06 | .0001944 | -.0001319 | .0005796 |
| Sigma_2_2 | .0002814 | .0000593 | 5.9e-07 | .0002736 | .000188 | .0004213 |

# MVN inverse-Wishart and MVN Jeffreys priors

- Example: $\mathbf{y_{1t}} = a_{11}y_{1t-1} + a_{12}y_{2t-1} + c1 + u_{1t}$
  $\mathbf{y_{2t}} = a_{21}y_{1t-1} + a_{22}y_{2t-1} + c2 + u_{1t}$

- Error terms have a multivariate normal distribution

$$(\mathbf{u_1}, \mathbf{u_2}) \sim MVN\left((0,0), \boldsymbol{\Sigma}\right)$$

- Prior for coefficients:

  $\beta \sim N(\boldsymbol{\beta_0}, \boldsymbol{\Omega_0})$ where: $\boldsymbol{\beta_0}$ and $\boldsymbol{\Omega_0}$ are those from the original Minnesota prior.

- Prior variances:
  - For MVN inverse Wishart (`minninvwishprior`):

$$\boldsymbol{\Sigma} \sim InvWishart(\alpha_0, \boldsymbol{S_0}) \quad ; \quad \alpha_0 = K + 2, \quad \boldsymbol{S_0} = (\alpha_0 - K - 1)\boldsymbol{\Sigma_0}$$

  - For MVN Jeffreys (`minnjeffprior`): $\boldsymbol{\Sigma} \sim Jeffreys(K)$

# Lag selection

- In the classical estimation we can select the optimal number of lags by using a few different information criteria like the ones implemented with `varsoc` (AIC, BIC, FPE, HQIC)

- In the Bayesian approach, we can perform the selection using posterior probabilities (with `bayestest model` and Bayes factors with `bayesstats ic`).

# Lag selection

- Fit the competing models, saving the mcmc simulation and storing the results:

```
 . matrix b0 = J(1,2,0)
. bayes,minnconjprior(mean(b0) selftight(1))   ///
>      rseed(123) saving(bvarsim,replace):     ///
>      var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/1)
. estimates store bvar1
. bayes,minnconjprior(mean(b0) selftight(1))  ///
>     rseed(123) saving(bvarsim,replace):     ///
>     var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/2)
. estimates store bvar2
. bayes,minnconjprior(mean(b0) selftight(1))  ///
>     rseed(123) saving(bvarsim,replace):     ///
>     var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/3)
. estimates store bvar3
. bayes,minnconjprior(mean(b0) selftight(1))  ///
>     rseed(123) saving(bvarsim,replace):     ///
>     var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/4)
. estimates store bvar4
. bayes,minnconjprior(mean(b0) selftight(1))  ///
>     rseed(123) saving(bvarsim,replace):     ///
>     var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/5)
. estimates store bvar5
. bayestest model bvar1 bvar2 bvar3 bvar4 bvar5
. bayesstats ic bvar1 bvar2 bvar3 bvar4 bvar5, basemodel(bvar5)
```

# Lag selection

- Selection based on posterior probabilities and bayes factors:

```
. bayestest model bvar1 bvar2 bvar3 bvar4 bvar5
Bayesian model tests

           |     log(ML)         P(M)       P(M|y)
-----------+--------------------------------------
     bvar1 |    142.3473       0.2000       0.5300
     bvar2 |    141.8501       0.2000       0.3224
     bvar3 |    140.9117       0.2000       0.1261
     bvar4 |    138.8614       0.2000       0.0162
     bvar5 |    137.7435       0.2000       0.0053

Note: Marginal likelihood (ML) is computed using
      Laplace-Metropolis approximation.

.
.
. bayesstats ic bvar1 bvar2 bvar3 bvar4 bvar5, basemodel(bvar5)
Bayesian information criteria

           |        DIC      log(ML)      log(BF)
-----------+--------------------------------------
     bvar1 |  -319.8079     142.3473     4.603722
     bvar2 |  -322.1719     141.8501     4.106574
     bvar3 |  -322.5092     140.9117     3.168116
     bvar4 |  -319.0206     138.8614     1.117898
     bvar5 |  -318.4444     137.7435            .

Note: Marginal likelihood (ML) is computed
      using Laplace-Metropolis approximation.
```

# Lag selection

- Change prior probabilities:

```
. bayestest model bvar1 bvar2 bvar3 bvar4 bvar5,   ///
>          prior(.1,.25,.25,.35,.05)
Bayesian model tests

             |   log(ML)       P(M)      P(M|y)
-------------+------------------------------------
       bvar1 |  142.3473     0.1000     0.3098
       bvar2 |  141.8501     0.2500     0.4711
       bvar3 |  140.9117     0.2500     0.1843
       bvar4 |  138.8614     0.3500     0.0332
       bvar5 |  137.7435     0.0500     0.0016
--------------------------------------------------
Note: Marginal likelihood (ML) is computed using
      Laplace-Metropolis approximation.
```

- Compare bayes factors for the first two models

```
. bayesstats ic bvar1 bvar2, basemodel(bvar2)
Bayesian information criteria

             |       DIC    log(ML)    log(BF)
-------------+---------------------------------
       bvar1 | -319.8079   142.3473   .4971489
       bvar2 | -322.1719   141.8501          .
-----------------------------------------------
Note: Marginal likelihood (ML) is computed
      using Laplace-Metropolis approximation.
```

# Postestimation

# Stability condition and Impulse-response functions (IRF)

- If the VAR is stable, we can derive the alternative moving average representation

$$\mathbf{y_t} = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{D_i x_{t-i}} + \sum_{i=0}^{\infty} \boldsymbol{\Phi}_i \mathbf{u_{t-1}}$$

Where:

$\mathbf{y_t}$ : $I(1)$

$\boldsymbol{\mu}$ : Kx1 time-invariant mean of the process

$\boldsymbol{\Phi}_i$ : KxK matrices of parameters (MA coefficients - IRF)

$\mathbf{u_{t-1}}, \mathbf{u_{t-2}}, \ldots$ : i.i.d shocks

- The **stability condition** is satisfied if all eigenvalues for a companion matrix are less than 1.
- The companion matrix can be derived from the moving average representation.

# Orthogonal shocks

- Alternative representation in terms of orthogonal shocks for the IRFs.
- Let's have a matrix $\mathbf{P}$ such that $\mathbf{\Sigma} = \mathbf{PP}'$

$$
\begin{aligned}
\mathbf{y_t} &= \mu + \sum_{s=0}^{\infty} \mathbf{\Phi}_s \mathbf{PP}' \mathbf{u_{t-s}} \\
&= \mu + \sum_{s=0}^{\infty} \mathbf{\Theta}_s \mathbf{P}^{-1} \mathbf{u_{t-s}} \\
&= \mu + \sum_{s=0}^{\infty} \mathbf{\Theta}_s \mathbf{w_{t-s}}
\end{aligned}
$$

Where: $E[\mathbf{P}^{-1}\mathbf{u_{t-s}}] = 0$

$E[\mathbf{P}^{-1}\mathbf{u_t}(\mathbf{P}^{-1}\mathbf{u_t})'] = \mathbf{I_K}$

# Check stability condition

```
 . quietly bayes, minnconjprior(mean(b0) selftight(1)) ///
>           rseed(123) saving(bvarsim,replace):       ///
>           var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/2)

. bayesvarstable

Eigenvalue stability condition                 Companion matrix size =     4
                                               MCMC sample size      = 10000
```

| Eigenvalue modulus | Mean | Std. dev. | MCSE | Median | Equal-tailed [95% cred. | interval] |
|---|---|---|---|---|---|---|
| 1 | .724855 | .1133179 | .001133 | .726628 | .4961755 | .9460597 |
| 2 | .555972 | .1096657 | .001097 | .5613761 | .3230319 | .75542 |
| 3 | .4702513 | .1310388 | .00131 | .4800829 | .1898752 | .6959185 |
| 4 | .3814394 | .1461639 | .001462 | .4034342 | .0473722 | .62082 |

```
Pr(eigenvalues lie inside the unit circle) = 0.9925
```

- The probability that all the eigenvalues are less than one supports the stability of the selected model.

# Impulse-response functions - Table

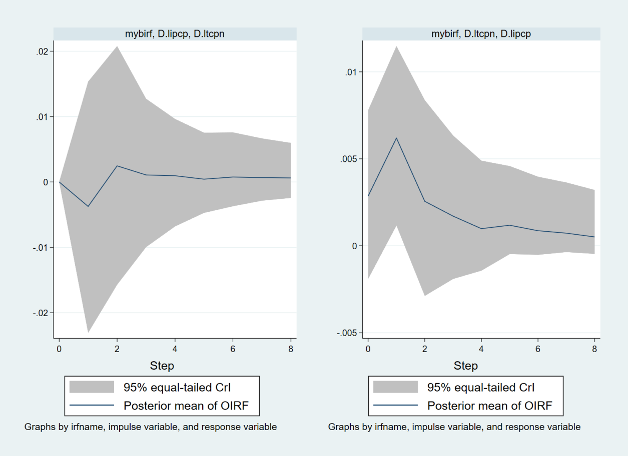
- Impulse-response functions considering the order ltcpn->lipcp.

```
 . quietly bayes, minnconjprior(mean(b0) selftight(1)) ///
>         rseed(123) saving(bvarsim,replace):          ///
>         var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/2)
. estimates store ex_irf
. bayesirf create mybirf, step(8) set(mybirf)
(file mybirf.irf created)
(file mybirf.irf now active)
(file mybirf.irf updated)
. bayesirf table oirf,nocri
Results from mybirf
```

| Step | (1) oirf | (2) oirf | (3) oirf | (4) oirf |
|---|---|---|---|---|
| 0 | .069921 | .002857 | 0 | .016263 |
| 1 | .021745 | .006194 | -.00374 | .003127 |
| 2 | -.008252 | .002557 | .002466 | .006641 |
| 3 | -.006352 | .001705 | .001075 | .002657 |
| 4 | .001069 | .000986 | .000956 | .003497 |
| 5 | .002663 | .00118 | .000429 | .001828 |
| 6 | .000771 | .000866 | .00076 | .002067 |
| 7 | -.000276 | .000725 | .000665 | .0013 |
| 8 | .000018 | .000509 | .000622 | .001374 |

```
Posterior means reported.
(1) irfname = mybirf, impulse = D.ltcpn, and response = D.ltcpn.
(2) irfname = mybirf, impulse = D.ltcpn, and response = D.lipcp.
(3) irfname = mybirf, impulse = D.lipcp, and response = D.ltcpn.
(4) irfname = mybirf, impulse = D.lipcp, and response = D.lipcp.
```

# Impulse-response functions - Graph



# Bayesian forecasting

- Posterior predictive distribution of the replicated data:

$$\Pr(\mathbf{y}_{T+1:T+h}|\mathbf{y}_T) = \int f(\mathbf{y}_{T+1:T+h}|\mathbf{y}_T;\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$$

Where:

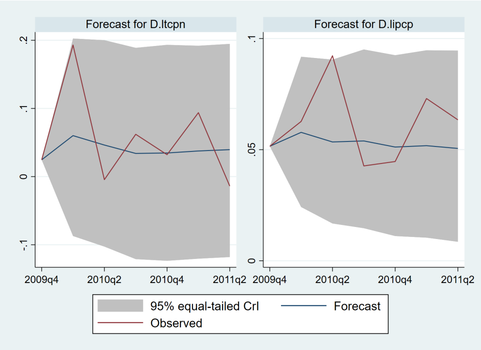
$$\begin{aligned}
\tilde{\mathbf{y}}^s_{T+1} &= \mathbf{y}_T + \mathbf{u}^1 && ;\ \mathbf{u}^1 \sim N(0,\boldsymbol{\Sigma}^s) \\
\tilde{\mathbf{y}}^s_{T+2} &= \tilde{\mathbf{y}}^s_{T+1} + \mathbf{u}^2 && ;\ \mathbf{u}^2 \sim N(0,\boldsymbol{\Sigma}^s) \\
&\dots \\
\tilde{\mathbf{y}}^s_{T+h} &= \tilde{\mathbf{y}}^s_{T+h-1} + \mathbf{u}^h && ;\ \mathbf{u}^h \sim N(0,\boldsymbol{\Sigma}^s)
\end{aligned}$$

Save dynamic forecasts $(\tilde{\mathbf{y}}^s_{T+1}, \tilde{\mathbf{y}}^s_{T+2}, \dots, \tilde{\mathbf{y}}^s_{T+h})$

# Forecasting - `bayesfcast compute`

- Let's get dynamic forecasts and plot them (also save the simulated outcomes).

```
. quietly bayes, minnconjprior(mean(b0) selftight(1))    ///
>         rseed(123) saving(bvar_mcmc,replace):          ///
>         var D.ltcpn D.lipcp if tin(1999q1,2009Q4), lags(1/2)
.
. bayesfcast compute bvar_,dynamic(tq(2010q1)) step(6)   ///
>         mcmcsaving(fcast_mcmc,replace) rseed(123)
.
. bayesfcast graph bvar_D_ltcpn bvar_D_lipcp,observed
```



# Forecasting with `bayesfcast compute`

- How about obtaining probabilities for events associated to different levels of the endogenous variables?

```
. /* Use mcmc simulations for the predicted outcome variables */
. use fcast_mcmc,clear
.
. /* Rename to identify quarter predictions */
. foreach name in D_ltcpn D_lipcp {
.    rename `name´_200 `name´_10Q1
.    rename `name´_201 `name´_10Q2
.    rename `name´_202 `name´_10Q3
.    rename `name´_203 `name´_10Q4
.    rename `name´_204 `name´_11Q1
.    rename `name´_205 `name´_11Q2
. }
. /* Event for t+1: inflation >.05 */
. generate inflation_over_5=cond(D_lipcp_10Q1>.05,1,0)
. /* Event for t+1: inflation>.05 and exchrate_change>.03 */
. generate infl_over_5__exchrate_chg_over_3 = ///
>          cond(D_lipcp_10Q1>.05 & D_ltcpn_10Q1>.03,1,0)
. /* Event for t+1 & t+2: exchrate_change>.03 */
. generate exchrate_change_over_3 =          ///
>          cond(D_ltcpn_10Q1>.03 & D_ltcpn_10Q2>.03,1,0)
```

# Forecasting with `bayesfcast compute`

- How about obtaining probabilities for events associated to different levels of the endogenous variables?

```
 . /* Use mcmc simulations for the predicted outcome variables */
. use fcast_mcmc,clear
.
. /* Rename to identify quarter predictions */
. foreach name in D_ltcpn D_lipcp {
.    rename `name´_200 `name´_10Q1
.    rename `name´_201 `name´_10Q2
.    rename `name´_202 `name´_10Q3
.    rename `name´_203 `name´_10Q4
.    rename `name´_204 `name´_11Q1
.    rename `name´_205 `name´_11Q2
. }
. /* Event for t+1: inflation >.05 */
. generate inflation_over_5=cond(D_lipcp_10Q1>.05,1,0)
. /* Event for t+1: inflation>.05 and exchrate_change>.03 */
. generate infl_over_5__exchrate_chg_over_3 = ///
>           cond(D_lipcp_10Q1>.05 & D_ltcpn_10Q1>.03,1,0)
. /* Event for t+1 & t+2: exchrate_change>.03 */
. generate exchrate_change_over_3 =           ///
>           cond(D_ltcpn_10Q1>.03 & D_ltcpn_10Q2>.03,1,0)
```

# Forecasting with bayesfcast compute

- We can now combine `proportion` with `collect` to report probabilities for events associated to our forecasts:

```
. collect : proportion inflation_over_5
. collect : proportion infl_over_5__exchrate_chg_over_3
. collect : proportion exchrate_change_over_3
. quietly collect layout (colname[1.inflation_over_5         ///
>                          1.infl_over_5__exchrate_chg_over_3 ///
>                          1.exchrate_change_over_3])         ///
>                (cmdset#result[_r_b])
.
. collect style header result,level(hide)
. collect label values cmdset 1 "2010Q1" 2 "2010Q1" 3 "2010Q1-Q2"
. collect label dim colname "Event", modify
. collect style header colname, level(value) title(label)
. collect preview
```

| | 2010Q1 | 2010Q1 | 2010Q1-Q2 |
|---|---|---|---|
| Event | | | |
| inflation_over_5=1 | .6778 | | |
| infl_over_5__exchrate_chg_over_3=1 | | .4554 | |
| exchrate_change_over_3=1 | | | .4303 |

# Forecasting with `bayesfcast compute`

- We can now combine `proportion` with `collect` to report probabilities for events associated to our forecasts:

```
. collect :  proportion inflation_over_5
. collect :  proportion infl_over_5__exchrate_chg_over_3
. collect :  proportion exchrate_change_over_3
. quietly collect layout (colname[1.inflation_over_5         ///
>                           1.infl_over_5__exchrate_chg_over_3 ///
>                           1.exchrate_change_over_3])         ///
>                   (cmdset#result[_r_b])
.
. collect style header result,level(hide)
. collect label values cmdset 1 "2010Q1" 2 "2010Q1" 3 "2010Q1-Q2"
. collect label dim colname "Event", modify
. collect style header colname, level(value) title(label)
. collect preview
```

| | 2010Q1 | 2010Q1 | 2010Q1-Q2 |
|---|---|---|---|
| Event | | | |
| inflation_over_5=1 | .6778 | | |
| infl_over_5__exchrate_chg_over_3=1 | | .4554 | |
| exchrate_change_over_3=1 | | | .4303 |

# Summing up

- Frequentist analysis base the conclusions on the distributions of statistics derived from random samples, assuming unknown fixed parameters.
- Bayesian analysis answers questions based on the distribution of parameters conditional on the observed sample.
- Bayesian VAR models are particularly convenient when working with small samples. Shrinking the parameter space with the priors allows controlling the number of lags more effectively.
- Impulse-response analysis and Forecasting are based on the full probability distributions for the parameters and the predictions.
- The posterior predictive distribution can be used to define events that can be evaluated for policy analysis.

# References


- Cameron, A. and Trivedi, P. 2005. *Microeconometric Methods and Applications*. Cambridge University Press, Section 13.2.2, 422–423.
- Doan, T., R. B. Litterman, and C. Sims. 1984. *Forecasting and conditional projection using realistic prior distributions.* Econometric Reviews 1: 1–100.
- Garratt, A., Lee, K., M.H. Pesaran and Y. Shin. 2006, *Global and National Macroeconometric Modeling: A long-run structural approach.* Oxford University Press, Oxford.
- Litterman, R. B. 1986. *Forecasting with Bayesian vector autoregressions-Five years of experience.* Journal of Business and Economic Statistics 4: 25–38.
- Sanchez, G. and Zavarce 2013, *Prospectos para la economía: Una aplicación con modelos VAR cointegrados y proyecciones probabilísticas.* in "Global and National Macroeconometric Modeling" Stata Press.