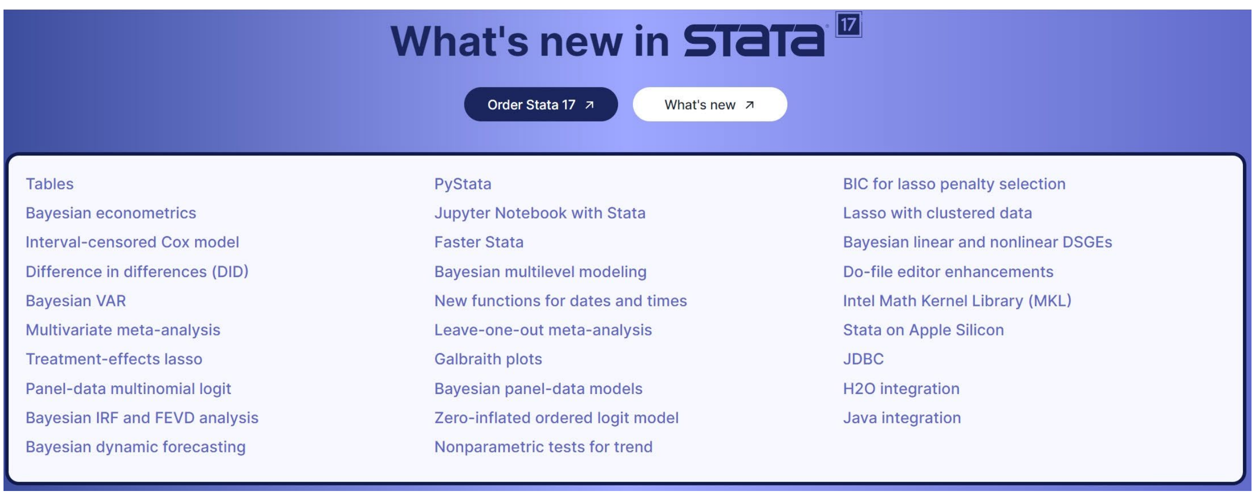
贝叶斯 VAR 模型

王群勇 (南开大学数量经济研究所, QunyongWang@outlook.com)

Contents

贝叶斯方法



基于模拟的方法和贝叶斯分层模型是 50 年以来最重要的统计思想之一(Gelman & Vehtari,2021)

Stata 15 引入 bayes, bayesmh

贝叶斯方法

对于计量模型，数据为y，参数为θ。贝叶斯定理表示为

$$\pi(\theta|y)=\frac{f(y|\theta)\pi(\theta)}{f(y)}\propto f(y|\theta)\pi(\theta).$$

其中，$f(y|\theta)$为似然函数(概率密度函数)，$\pi(\theta)$为先验分布，$\pi(\theta|y)$为后验分布。

$f(y|\theta)\pi(\theta)$叫做贝叶斯核(kernel)。$f(y)$叫做边际似然函数。

$$\text{后验分布}\propto\text{似然函数}\times\text{先验分布}$$

贝叶斯方法

频数方法中，参数是确定的，参数估计量是随机的。贝叶斯方法中，参数是随机的。

频数方法中原假设是否成立是一个确定性事件，无法判断原假设成立的概率。贝叶斯方法是原假设为随机事件。

贝叶斯方法将先验分布与似然函数得到后验分布（适于参数个数较多的模型）。

极大似然估计是数值计算，贝叶斯估计是抽取随机数（适于复杂模型）。

贝叶斯的抽样分布属于精确分布（适于小样本量）。

贝叶斯推断将多种不确定性，包括数据、模型和参数，融合到统一的框架内（适于模型的比较与平均）。

贝叶斯方法

共轭先验：后验分布与先验分布属于同一分布族

非共轭先验：后验分布没有明确的表达式 —> MCMC 抽样

- Metropolis-Hastings 抽样
- Gibbs 抽样

MH 抽样

数值例子：$y_i \sim binomial(10,4,\theta)$, prior: $\theta \sim Beta(1,1)$

设初始值$\theta_1=0.517$, 令$\theta_{new}=0.380$?

$$\rho=\frac{\text{posterior}(\theta_{new})}{\text{posterior}_{\theta_1}}=\frac{Beta(1,1,0.380)\times Binomial(10,4,0.380)}{Beta(1,1,0.517)\times Binomial(10,4,0.517)}=1.307$$

$\theta_2=0.380$, 令$\theta_{new}=0.286$?

$$\rho = \frac{\text{posterior}(\theta_{new})}{\text{posterior}_{\theta_1}} = \frac{Beta(1,1,0.286) \times Binomial(10,4,0.286)}{Beta(1,1,0.380) \times Binomial(10,4,0.380)} = 0.747$$

θ_3以 0.747 的概率取 0.747，以 0.253 的概率取θ_2.

……

问题：如何生成θ_{new}?

MCMC

MH 抽样: 设后验分布为$f(x)$，

(1) 给定x_t, 定义工具分布$q(y|x_t)$ (proposed distribution), 从$q(y|x_t)$抽取随机数x_{new}。其中，工具分布是更容易抽样的分布。

(2) 定义接受概率(acceptance probability): $\rho(x_t, x_{new}) = \min\left[\frac{f(x_{new})}{f(x_t)}\frac{q(x_t|x_{new})}{q(x_{new}|x_t)}, 1\right]$,

$$x_{t+1} = \begin{cases} x_{new} & \rho(x_t, x_{new}) \\ x_t & 1-\rho(x_t, x_{new}) \end{cases}$$

对称情形：$q(x_t|x_{new}) = q(x_{new}|x_t), \rho(x_t, x_{new}) = \min\left[\frac{f(x_{new})}{f(x_t)}, 1\right]$,

独立情形：$q(x_t|x_{new}) = q(x_t), q(x_{new}|x_t) = q(x_{new}), \rho(x_t, x_{new}) = \min\left[\frac{f(x_{new})}{f(x_t)}\frac{q(x_t)}{q(x_{new})}, 1\right]$。

计算接受概率时，只有分布的核起作用，其它常数都被略掉。

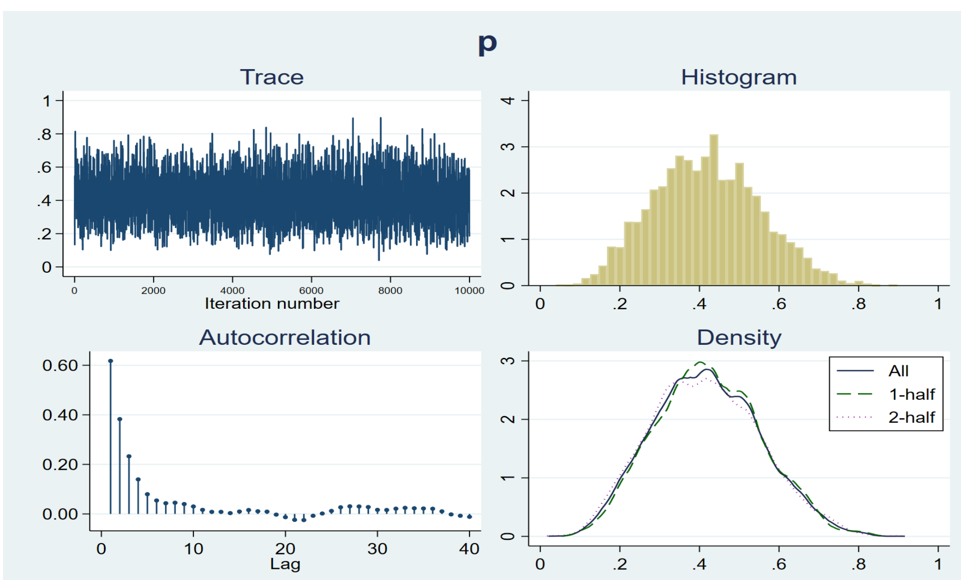
MH 抽样

```
. use bintrial, clear
(Federal Reserve Economic Data - St. Louis Fed)

. bayesmh y, likelihood(dbernoulli({p})) prior({p},beta(1,1))

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  y ~ bernoulli({p})

Prior:
```

```
  {p} ~ beta(1,1)
------------------------------------------------------------------------------

Bayesian Bernoulli model                         MCMC iterations  =     12,500
Random-walk Metropolis-Hastings sampling         Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =         10
                                                 Acceptance rate  =      .4823
Log marginal-likelihood = -7.8194591             Efficiency       =      .2291

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
           p |  .4187117   .1342192   .002804   .4152274   .1746616   .6876875
------------------------------------------------------------------------------
```

MH 抽样

```
. bayesgraph diag {p}
```



Gibbs 抽样

```
. bayesmh y, likelihood(dbernoulli({p})) prior({p},beta(1,1)) block({p}, gibbs)

Burn-in ...
Simulation ...

Model summary
------------------------------------------------------------------------------
Likelihood:
  y ~ bernoulli({p})

Prior:
  {p} ~ beta(1,1)
------------------------------------------------------------------------------

Bayesian Bernoulli model                         MCMC iterations  =     12,500
Gibbs sampling                                   Burn-in          =      2,500
                                                 MCMC sample size =     10,000
                                                 Number of obs    =         10
                                                 Acceptance rate  =          1
Log marginal-likelihood = -7.8006434             Efficiency       =      .9783

------------------------------------------------------------------------------
             |                                                Equal-tailed
             |      Mean   Std. dev.     MCSE     Median  [95% cred. interval]
-------------+----------------------------------------------------------------
           p |  .4157295   .1367685   .001383   .4123873    .164642   .6885919
------------------------------------------------------------------------------
```

Contents

VAR

令$y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$, VAR(p) model: $y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + C x_t + u_t, u_t \sim N(0, \Sigma)$.

例：3 个变量构成的 VAR(1)模型

$$\begin{aligned} y_{1t} &= c_1 + a_{11}y_{1t-1} + a_{12}y_{2t-1} + a_{13}y_{3t-1} + u_{1t}, \\ y_{2t} &= c_2 + a_{21}y_{1t-1} + a_{22}y_{2t-1} + a_{23}y_{3t-1} + u_{2t}, \\ y_{3t} &= c_3 + a_{31}y_{1t-1} + a_{32}y_{2t-1} + a_{33}y_{3t-1} + u_{3t}, \end{aligned}$$

$$(1 - a_1L - \cdots - a_pL^p)y_t = C + u_t.$$

其中，$E(u_{it}u_{js}) = 0(i \neq j, t \neq s)$。$Var(u_t) = \Sigma.$

参数个数为(包括常数项和协方差矩阵)：$m(mp+1) + m(m+1)/2$。

VAR

VAR 模型: $y_t = C + A_1y_{t-1} + A_2y_{t-2} + \cdots + A_py_{t-p} + u_t$，$y_t$为$(m \times 1)$。$A_j$为$m \times m$矩阵，所有参数$A = \left(\mathrm{vech}(A_1) \setminus \mathrm{vech}(A_2) \cdots \setminus \mathrm{vech}(A_p) \setminus C\right)$.

把$(1, \ldots, T)$期观测值叠加起来，$Y = XB + U$，$Y = (y_1{}' \setminus y_2{}' \setminus \ldots \setminus y_m{}')$，为$(T \times m)$,每一列为一个变量。矩阵；$X$为$Y$的滞后项和常数项构成的矩阵，为$T \times (mp+1)$矩阵。$B$为$(mp+1) \times m$矩阵。$U$为$(T \times m)$。

逐列叠加起来，$\mathrm{vech}(Y) = y, A = \mathrm{vech}(B)$，其中

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ \cdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \cdots \\ y_{2T} \\ \cdots \end{bmatrix} \rightarrow \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{m1} \\ y_{12} & y_{22} & \cdots & y_{m2} \\ & & & \cdots \\ y_{1T} & y_{2T} & \cdots & y_{mT} \end{bmatrix}$$

VAR

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \cdots \\ y_{1T} \\ y_{21} \\ y_{22} \\ \cdots \\ y_{2T} \\ \cdots \end{bmatrix} \rightarrow Var(u) = \Sigma \otimes I_{mp+1}. \begin{bmatrix} y_{11} \\ y_{21} \\ \cdots \\ y_{m1} \\ y_{21} \\ y_{22} \\ \cdots \\ y_{m2} \\ \cdots \end{bmatrix} \rightarrow Var(u_t) = E(u_tu_t{}') = \Sigma, Var(u) = I_{mp+1} \otimes \Sigma.$$

贝叶斯 VAR (BVAR)

BVAR 模型是用贝叶斯方法来估计 VAR 模型参数。

贝叶斯 VAR 由 Doan, Litterman, and Sims (1984)提出，Kadiyala and Karlsson (1997), Ba´nbura, Giannone, and Reichlin (2008), and Dieppe, Legrand, and van Roye (2016)详细介绍了贝叶斯 VAR 的优势。

1. 避免参数过多的问题。通过对参数的先验约束降低参数维度，达到收缩(shrinkage)。
2. 通过分层先验(hierarchical prior)更方便地处理异质性问题，不论是截面、面板还是时间序列数据。
3. 贝叶斯因子给出了选择滞后阶数和变量的统一框架。

BVAR 模型在参数较多、样本量较少时表现出较好的拟合效果。

Minniesota 先验

Litterman Minnesota (original): A为正态先验，Σ固定

Normal-flat: A为正态先验，独立于Σ的先验

conjugate Minnesota: $A|\Sigma$为正态先验，Σ为 Inverse-Wishart 先验

normal-iwishart 先验: $A|\Sigma$为正态先验， Σ为 Inverse-Wishart 先验

independent normal-iwishart 先验: $A|\Sigma$为正态先验， Σ为 Inverse-Wishart 先验，不同方程的系数的先验是相互独立的。

Jeffreys 先验: A为正态先验，Σ非信息先验

Sims-Zha normal-flat: 结构 VAR 模型的 Normal-flat 先验

Sims-Zha normal-flat: 结构 VAR 模型的 Normal-iwishart 先验

Giannone, Lenza, and Primiceri: 超参数通过优化程序自动选择。

Stata

option	suboption	note
minnfixedcovprior[(*subopts*)]	mean(*vector*) mean(*numlist*)	β_0中β_{ii}^1 (m)
minnconjprior[(*subopts*)]	mean(*vector*) mean(*numlist*)	β_0中β_{ii}^1 (m)；默认值为 b_{ii}^1=1，其它为 0
	phi(*matrix*)	Φ_0 $((mp+1)\times(mp+1))$
	df(#)	α_0 (default: $m+2$)
	scale(*matrix*)	S_0 $(m \times m)$
minniwishprior[(*subopts*)]	mean(*vector*) mean(*numlist*)	β_0中β_{ii}^1 (m)
	cov(*matrix*)	Σ_0
	df(#)	α_0
	scale(*matrix*)	S_0
minnjeffprior[(*subopts*)]	mean(*vector*) mean(*numlist*)	β_0中β_{ii}^1 (m)
	cov(*matrix*)	Σ_0 $(m \times m)$

m个变量构成的 VAR(p)，那么 *vector* 应该为 m 维向量

Stata

option	note	默认值
selftight(#)	因变量自身滞后的紧度	$\lambda_1=1$
crosstight(#)	其它因变量滞后的紧度	$\lambda_2=0.5$
lagdecay(#)	衰减速度	$\lambda_3=1$
exogtight(#)	外生变量的紧度	$\lambda_4=100$
arcov	对每个方程单独估计 AR 模型估计协方差矩阵	
varcov	对所有方程估计 VAR 模型估计协方差矩阵	

Stata

bayes , [*bayesopts*] : `var` *varlist*, [*varoptions*]

Gibbs 抽样，100%接受率，避免了 MH 抽样的有效性不足问题。

默认先验为: `minnconjprior`

例： Minnesota prior

Link: original minnesoto prior

Link: original minnesoto prior(user-defined tightness)

Link: original minnesoto prior(var lags)

例： Conjugate Minnesota prior

Link: conjugate minnesoto prior

Link: conjugate minnesoto prior (user-defined prior)

例： Minnesota iwishart prior

Link: minnesoto inv-wishart prior

Link: minnesoto inv-wishart prior (user-defined prior)

例： Minnesota Jeffreys prior

Link: minnesoto Jeffreys prior

Link: minnesoto Jeffreys prior (user-defined prior)

Contents

确定滞后阶数

```
. use bvus, clear
(Federal Reserve Economic Data - St. Louis Fed)

. varsoc inflation ogap fedfunds, maxlag(12)

Lag-order selection criteria

   Sample: 1958q3 thru 2010q4                         Number of obs = 210
```

Lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-1488.6				296.509	14.2057	14.225	14.2535
1	-723.715	1529.8	9	0.000	.221616	7.00681	7.08413	7.19807
2	-689.089	69.252	9	0.000	.173634	6.76275	6.89806	7.09746*
3	-673.171	31.836	9	0.000	.162585	6.69686	6.89017	7.17502
4	-661.806	22.729	9	0.007	.159006	6.67434	6.92564	7.29595
5	-639.015	45.583	9	0.000	.139492	6.543	6.85228	7.30805
6	-619.85	38.329	9	0.000	.126698	6.44619	6.81346*	7.35469
7	-615.967	7.7663	9	0.558	.133135	6.49492	6.92019	7.54687
8	-610.886	10.161	9	0.338	.138349	6.53225	7.0155	7.72765
9	-587.182	47.409	9	0.000	.120437	6.39221	6.93345	7.73105
10	-581.902	10.559	9	0.307	.124996	6.42764	7.02688	7.90993
11	-567.442	28.921*	9	0.001	.118912*	6.37564*	7.03286	8.00137
12	-565.064	4.7562	9	0.855	.126973	6.4387	7.15392	8.20789

```
   * optimal lag
   Endogenous: inflation ogap fedfunds
    Exogenous: _cons
```

确定滞后阶数

VAR 模型根据信息准则容易出现过度拟合或者选择过高的滞后阶数。

```
forvalues i=1/6 {
    qui bayes, rseed(17) saving(bvarsim`i', replace): var inflation ogap fedfunds
 if date < tq(2004q1), lags(1/`i')
    est store bvar`i'
    local mods "`mods' bvar`i'"
}
bayestest model `mods'
```

注：在 est store 之前，必须在 bayes 指令中通过 saving()选项保存模拟结果。

codes for lag selection

VAR 模型的平稳性

根据每次模拟系数计算特征根，得到特征根的分布。

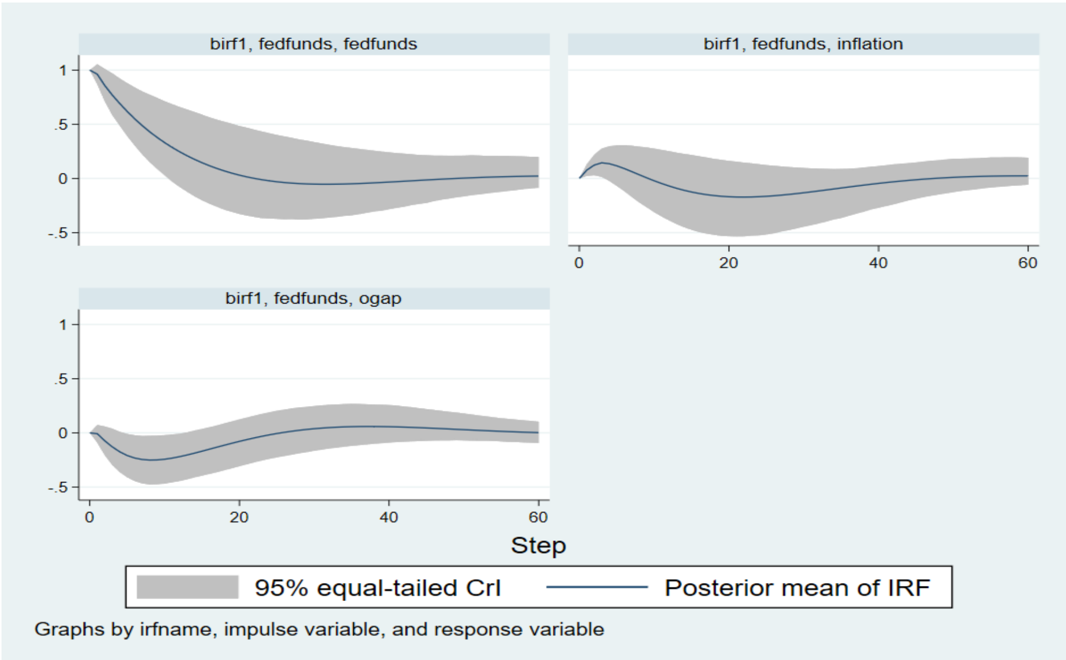
```
. qui bayes, rseed(17) saving(bvarsim, replace): var inflation ogap fedfunds if d
ate < tq(2004q1), lags(1/4)

. estimates store bvar

. bayesvarstable

Eigenvalue stability condition                  Companion matrix size =     12
                                                MCMC sample size      =  10000
```

Eigenvalue modulus	Mean	Std. dev.	MCSE	Median	Equal-tailed [95% cred. interval]	
1	.9473457	.0199198	.000199	.9481282	.9057116	.9838371
2	.9417123	.0257058	.000257	.9453142	.877582	.9811621
3	.8184194	.0716288	.000716	.8274233	.6763741	.9322606
4	.5930213	.0930861	.000931	.5836551	.4256008	.7733104
5	.4859573	.0896516	.000897	.4866775	.330644	.6554575
6	.3659255	.0417669	.000418	.3635287	.291461	.459251
7	.3499339	.0365851	.000366	.3496959	.2767796	.4214287
8	.3155561	.0383687	.000384	.3173136	.2348504	.3856269
9	.3014183	.0396995	.000397	.3038818	.2177103	.3736035
10	.2670156	.0479518	.00048	.2717858	.1582521	.3475958
11	.2361436	.0556598	.000557	.2414199	.1135724	.329785
12	.1887299	.0805818	.000806	.2036124	.0151749	.3102756

```
Pr(eigenvalues lie inside the unit circle) = 0.9977
```

贝叶斯脉冲响应

```
. . bayesirf create birf1, step(60) set(birfex, replace)
(file birfex.irf created)
(file birfex.irf now active)
(file birfex.irf updated)

. . bayesirf graph irf, impulse(fedfunds)
```



贝叶斯脉冲响应: 表格

```
. bayesirf table irf, response(ogap) impulse(fedfunds) step(12)

Results from birf1
```

Step	(1) irf	(1) Lower	(1) Upper
0	0	0	0
1	-.008015	-.089753	.072662
2	-.072428	-.205354	.059264
3	-.128667	-.296316	.039592
4	-.174391	-.361456	.009988
5	-.208873	-.409928	-.009742
6	-.232076	-.444489	-.021792
7	-.245563	-.466458	-.028681
8	-.251	-.475859	-.026581
9	-.249854	-.474231	-.025333
10	-.243505	-.468124	-.021195
11	-.233139	-.456024	-.014813
12	-.219759	-.444103	-.008206

```
Posterior means reported.
95% equal-tailed credible lower and upper bounds reported.
(1) irfname = birf1, impulse = fedfunds, and response = ogap.
```

贝叶斯脉冲响应: 更改变量顺序

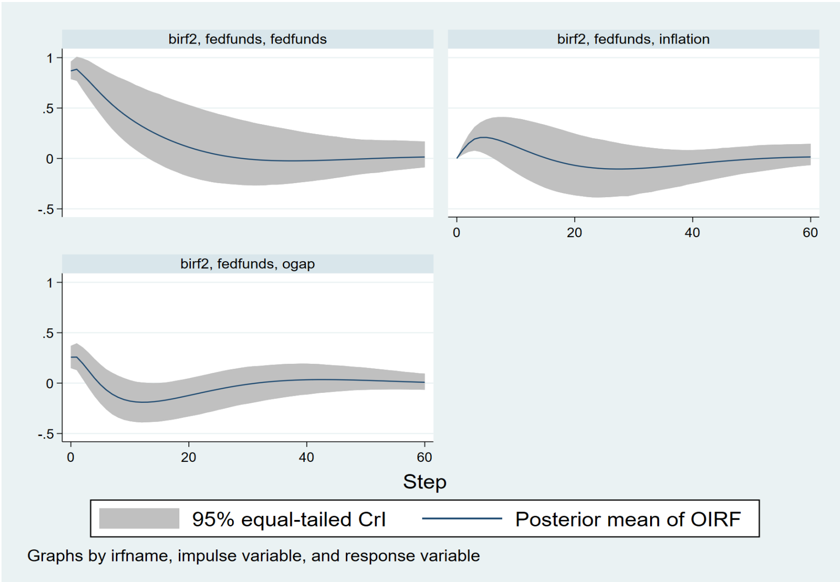
```
. bayesirf create birf2, step(60) set(birfex2, replace) order(inflation fedfunds
ogap)
(file birfex2.irf created)
(file birfex2.irf now active)
(file birfex2.irf updated)

. bayesirf table oirf, irf(birf2) response(ogap) impulse(fedfunds) step(10)

Results from birf2
```

Step	(1) oirf	(1) Lower	(1) Upper
0	.257325	.148406	.370283
1	.258308	.128963	.395361
2	.195725	.041816	.35614
3	.123306	-.047116	.30244
4	.050238	-.130222	.241047
5	-.014137	-.201699	.186256
6	-.067159	-.261451	.13775
7	-.109004	-.309117	.103761
8	-.140558	-.341167	.074475
9	-.163085	-.364218	.05027
10	-.177966	-.377733	.030365

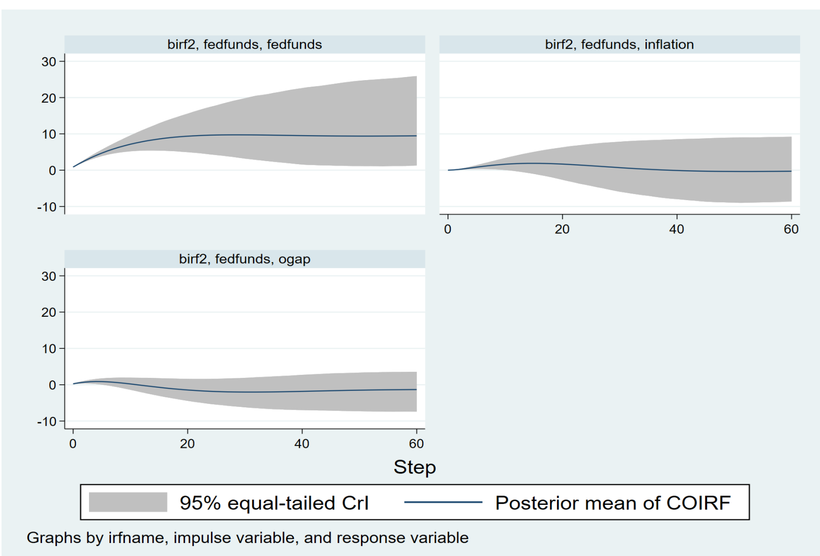
```
Posterior means reported.
95% equal-tailed credible lower and upper bounds reported.
(1) irfname = birf2, impulse = fedfunds, and response = ogap.
```

贝叶斯脉冲响应: (累积)正交脉冲响应

```
. bayesirf graph oirf, impulse(fedfunds)
```




```
. bayesirf graph coirf, impulse(fedfunds)
```



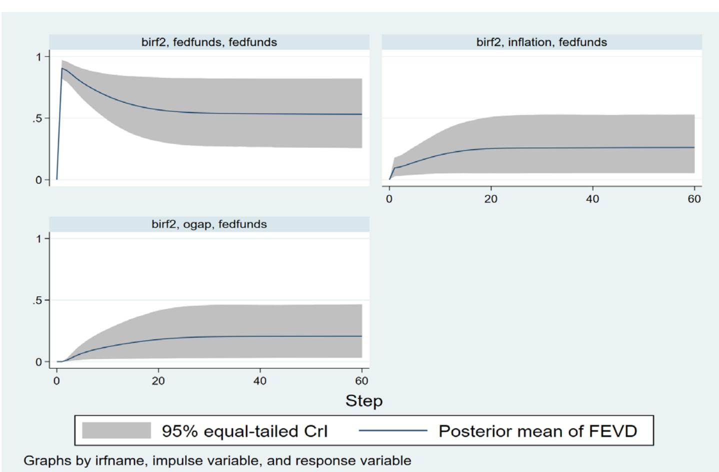
贝叶斯方差分解

```
. bayesirf table fevd, irf(birf2) response(fedfunds) step(7)

Results from birf2
```

Step	(1) fevd	(1) Lower	(1) Upper	(2) fevd	(2) Lower	(2) Upper
0	0	0	0	0	0	0
1	.095083	.027875	.180163	.904917	.819837	.972125
2	.102093	.029869	.192633	.885277	.794215	.957829
3	.114495	.034531	.216095	.852453	.751346	.936377
4	.128495	.038329	.242878	.818721	.702946	.917391
5	.142093	.041094	.268065	.789353	.659076	.902321
6	.155334	.043475	.293326	.763024	.619474	.890185
7	.16808	.045986	.318506	.739026	.582641	.879954

Step	(3) fevd	(3) Lower	(3) Upper
0	0	0	0
1	0	0	0
2	.01263	.002624	.027988
3	.033052	.00823	.069683
4	.052784	.013287	.111021
5	.068554	.016546	.144177
6	.081642	.018741	.174098
7	.092895	.020159	.199897

```
Posterior means reported.
95% equal-tailed credible lower and upper bounds reported.
(1) irfname = birf2, impulse = inflation, and response = fedfunds.
(2) irfname = birf2, impulse = fedfunds, and response = fedfunds.
(3) irfname = birf2, impulse = ogap, and response = fedfunds.

. bayesirf graph fevd, irf(birf2) response(fedfunds)
```



Contents

贝叶斯预测

贝叶斯预测利用后验分布对y_{t+h}进行预测:

$$f(y_{T+1:T+h}|y_{1:T}) = \int f(y_{T+1:T+h}|y_{1:T},\theta)f(\theta|D)d\theta.$$

其中，D表示数据信息。

$$f(y_{T+1:T+h}|y_{1:T},\theta) = f(y_{T+1}|y_{1:T},\theta)f(y_{T+2:T+h}|y_{1:T+1},\theta)\cdots f(y_{T+h}|y_{1:T+h-1},\theta)$$

贝叶斯预测

设θ的 MCMC 序列为$(\theta^1,\cdots,\theta^S)$,

1. foreach θ^s,
- 根据$f(y_{T+1}|y_{1:T},\theta^s)$，计算$y_{T+1}^s$
- 根据$f(y_{T+2}|y_{1:T},y_{T+1}^s,\theta^s)$，计算$y_{T+2}^s$
- ……
- 根据$f(y_{T+h}|y_{1:T},y_{T+1}^s,\cdots,y_{T+h-1}^s,\theta^s)$，计算$y_{T+h}^s$

2. 对每个y_{T+h}可以得到S个预测值，进而得到其预测的均值、中位数或置信区间。

贝叶斯预测

bayesfcast compute *prefix*, **dynamic**() *stat hpd*

预测指标包括：后验均值(`b1_*`)、后验标准差(`b1_*_sd`)、置信区间下界(`b1_*_lb`)、置信区间上界(`b1_*_ub`)。

bayesfcast graph *varlist*, **observed nocri**

贝叶斯预测

```
. bayesfcast compute b_, step(28)

. qui var inflation ogap fedfunds if date < tq(2004q1), lags(1/4)

. fcast compute f_, step(28) dynamic(tq(2004q1))

. bayesfcast graph f_inflation b_inflation f_ogap b_ogap f_fedfunds b_fedfunds, o
bserved byopts(rows(3) title("Freq
> uentist (left) vs. Bayesian (right)"))
```