

# 1 Байесовские VAR

На английском языке существует большое количество статей, посвященных теории и практике оценивания байесовских VAR-моделей (далее BVAR). На русском же языке ситуация с BVAR-моделями просто ужасна. Поэтому в данном разделе мы попытаемся одновременно изложить и теорию, и практику оценивания.

Практику оценивания BVAR моделей мы изложим в статистической среде R. Здесь нам будут полезны два пакета:

1. Пакет MSBVAR, выложенный на официальном репозитории, устанавливается командой:

```
install_packages("MSBVAR")
```

2. Пакет bvarr, выложенные на github, устанавливается почти также легко:

```
library("devtools")  
install_github("bdemeshev/bvarr")
```

Загружаем установленные пакеты и делаем доступным набор данных

```
library("MSBVAR")  
library("bvarr")  
library("ggplot2")  
data(Yraw)
```

Рассмотрим стандартную VAR(p) модель

$$y_t = a_0 + \sum_{j=1}^p A_j y_{t-j} + \varepsilon_t \quad (1)$$

Векторы  $y_t$ ,  $a_0$  и  $\varepsilon_t$  имеют размер  $M \times 1$ , матрицы  $A_i$  — размер  $M \times M$ .

Вектор  $\varepsilon_t$  имеет многомерное нормальное распределение,  $\mathcal{N}(0, \Sigma)$ .

В байесовском подходе параметры  $\Sigma$ ,  $a_0$ ,  $A_1, \dots, A_p$  предполагаются случайными. При наличии априорных предположений об их распределении можно вывести апостериорные распределения.

Для описания априорных распределений нам будут удобны дополнительные обозначения. А именно, все параметры модели, кроме  $\Sigma$ , мы «запишем» в матрицу  $A$  и в вектор  $\alpha$ :

1. Матрица  $A$ . Справа от столбца  $a_0$  выпишем друг за другом матрицы  $A_1, A_2, \dots, A_p$ , а затем транспонируем полученный результат:

$$A = (a_0 \ A_1 \ A_2 \ \dots \ A_p)'$$

Матрица  $A$  имеет размер  $(1 + pM) \times M$ . Величину  $1 + pM$  назовём буквой  $K$ .

2. Вектор  $\alpha$ . Расположим столбцы матрицы  $A$  друг под другом:

$$\alpha = \text{vec}(A)$$

Вектор  $\alpha$  состоит из  $M + pM^2$  элементов.

Поместим все переменные правой части в вектор-строку  $X_t$ , т.е. это вектор, который начинается с единицы, за ней находится вектор  $y_{t-1}$ , за ним — вектор  $y_{t-2}$  и так далее,  $X_t = (1, y'_{t-1}, y'_{t-2}, \dots, y'_{t-p})$ . Вектор  $X_t$  имеет размеры  $1 \times K$ , где  $K = 1 + pM$ .

Тогда VAR(p) модель можно также записать в виде:

$$y'_t = X_t A + \varepsilon'_t \quad (2)$$

или

$$y_t = (I_M \otimes X_t) \alpha + \varepsilon_t \quad (3)$$

где  $I_M$  — единичная матрица размера  $M \times M$ .

Также можно записать нашу модель без индекса  $t$ . Расположим вектор-строки  $y'_1, y'_2, \dots, y'_T$  одну под другой, полученную матрицу назовем буквой  $Y$ . Аналогично, с помощью  $E$  обозначим вектор-строки  $\varepsilon_1, \dots, \varepsilon_T$ , записанные одна под другой. В матрицу  $X$  поместим строки  $X_1, X_2, \dots, X_T$  одну под одной.

$$Y = XA + E \quad (4)$$

Вектор  $y = \text{vec}(Y)$  — это вектор-столбец зависимых переменных, в котором сначала идут все наблюдения за первой эндогенной переменной, потом все наблюдения за второй эндогенной переменной и т.д. Аналогично,  $\varepsilon = \text{vec}(E)$ .

$$y = (I_M \otimes X) \alpha + \varepsilon \quad (5)$$

где вектор ошибок  $\varepsilon$  имеет многомерное нормальное распределение,  $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T)$ .

Для удобства введем обозначение  $Z_t = (I_M \otimes X_t)$ , а также обозначения для МНК-оценок вектора  $\alpha$ , матрицы  $A$  и матрицы  $\Sigma$ . Здесь мы фактически предполагаем, что оцениваются отдельные модели для каждой компоненты  $y_t$ :

$$\hat{\alpha} = \left( \sum_t Z'_t Z_t \right)^{-1} \left( \sum_t Z'_t y_t \right) \quad (6)$$

$$\hat{A} = (X'X)^{-1} X'Y \quad (7)$$

Суммы квадратов остатков:

$$\hat{S} = (Y - X\hat{A})'(Y - X\hat{A}) \quad (8)$$

$$\hat{\Sigma} = \frac{\hat{S}}{T - K} \quad (9)$$

Существует несколько популярных вариантов априорного распределение неизвестных параметров:

- Априорное распределение Джефффри:

Параметры  $\alpha$  и  $\Sigma$  предполагаются априорно независимыми,

$$\begin{aligned} p(\alpha) &\propto 1 \\ p(\Sigma) &\propto |\Sigma|^{-(M+1)/2} \end{aligned} \quad (10)$$

При таких предположениях апостериорное распределение имеет вид:

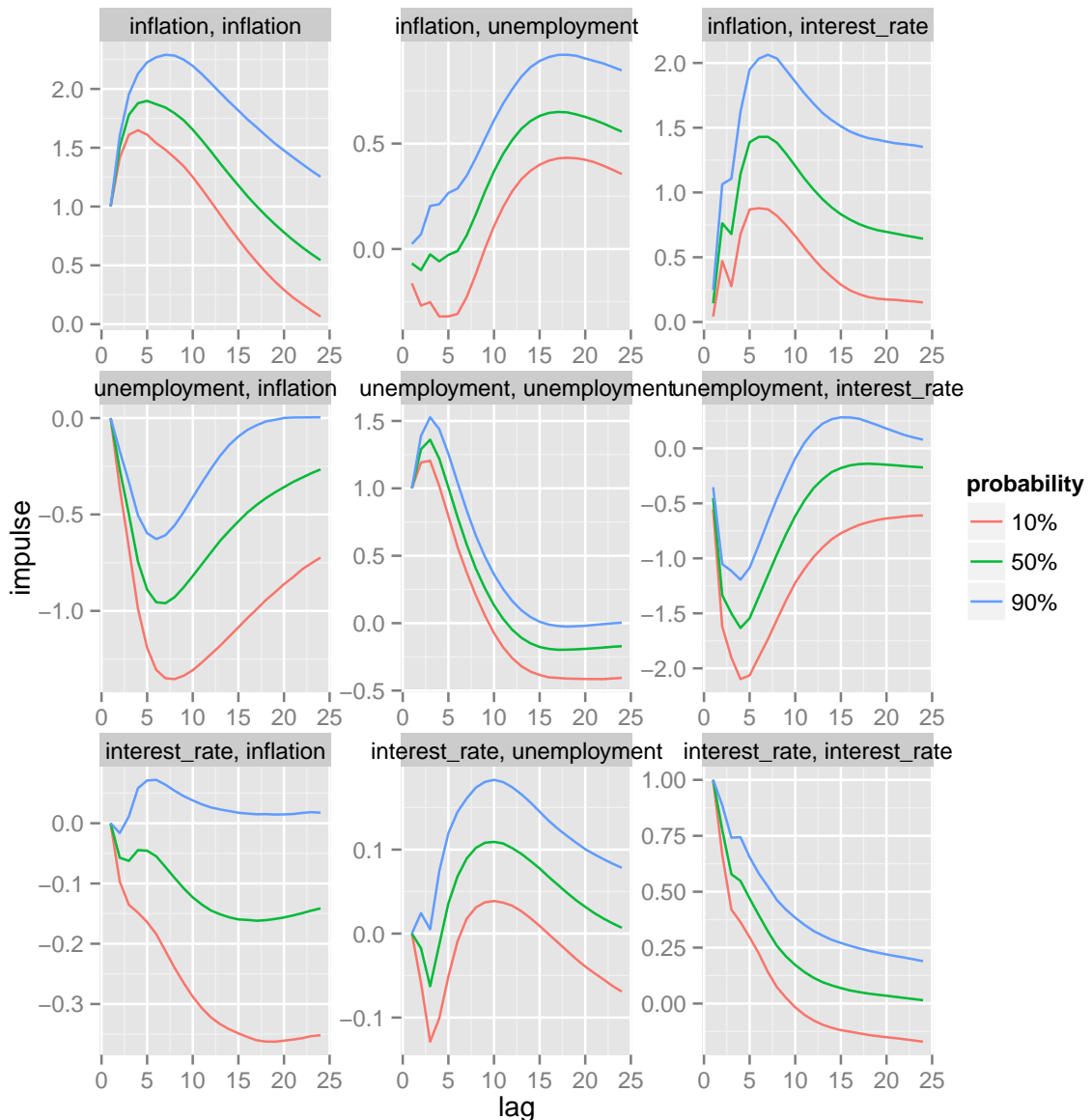
$$\begin{aligned}\alpha|\Sigma, y &\sim \mathcal{N}(\hat{a}, \Sigma) \\ \Sigma|y &\sim \mathcal{W}^{-1}(\hat{S}, T - K)\end{aligned}\tag{11}$$

В данном подходе исследователю вообще не требуется подбирать гиперпараметры априорного распределения. С одной стороны — это плохо, нет возможности посмотреть, как меняются апостериорные выводы при смене априорного распределения. С другой стороны — хорошо, так как подбирать ничего не нужно.

```
model <- bvar(Yraw, prior="diffuse")
```

```
## Iteration 2000 out of 10000
## Iteration 4000 out of 10000
## Iteration 6000 out of 10000
## Iteration 8000 out of 10000
## Iteration 10000 out of 10000
```

```
bvar.imp.plot(model)
```



Можно рассматривать как предельный частный случай сопряженного нормального-Уишарта априорного распределения, при  $\underline{\alpha} = 0$ ,  $\underline{\nu} = 0$  уточнить

- Сопряженное нормальное-Уишарта априорное распределение

Здесь параметры  $\alpha$  и  $\Sigma$  предполагаются априорно зависимыми:

$$\begin{aligned}\Sigma &\sim \mathcal{W}^{-1}(\underline{S}, \underline{\nu}) \\ \alpha|\Sigma &\sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})\end{aligned}\tag{12}$$

В этом случае условные апостериорные распределения будут иметь вид

$$\begin{aligned}\alpha|\Sigma, y &\sim \mathcal{N}(\bar{\alpha}, \Sigma \otimes \bar{V}) \\ \Sigma|y &\sim \mathcal{W}^{-1}(\bar{S}, \bar{\nu})\end{aligned}\tag{13}$$

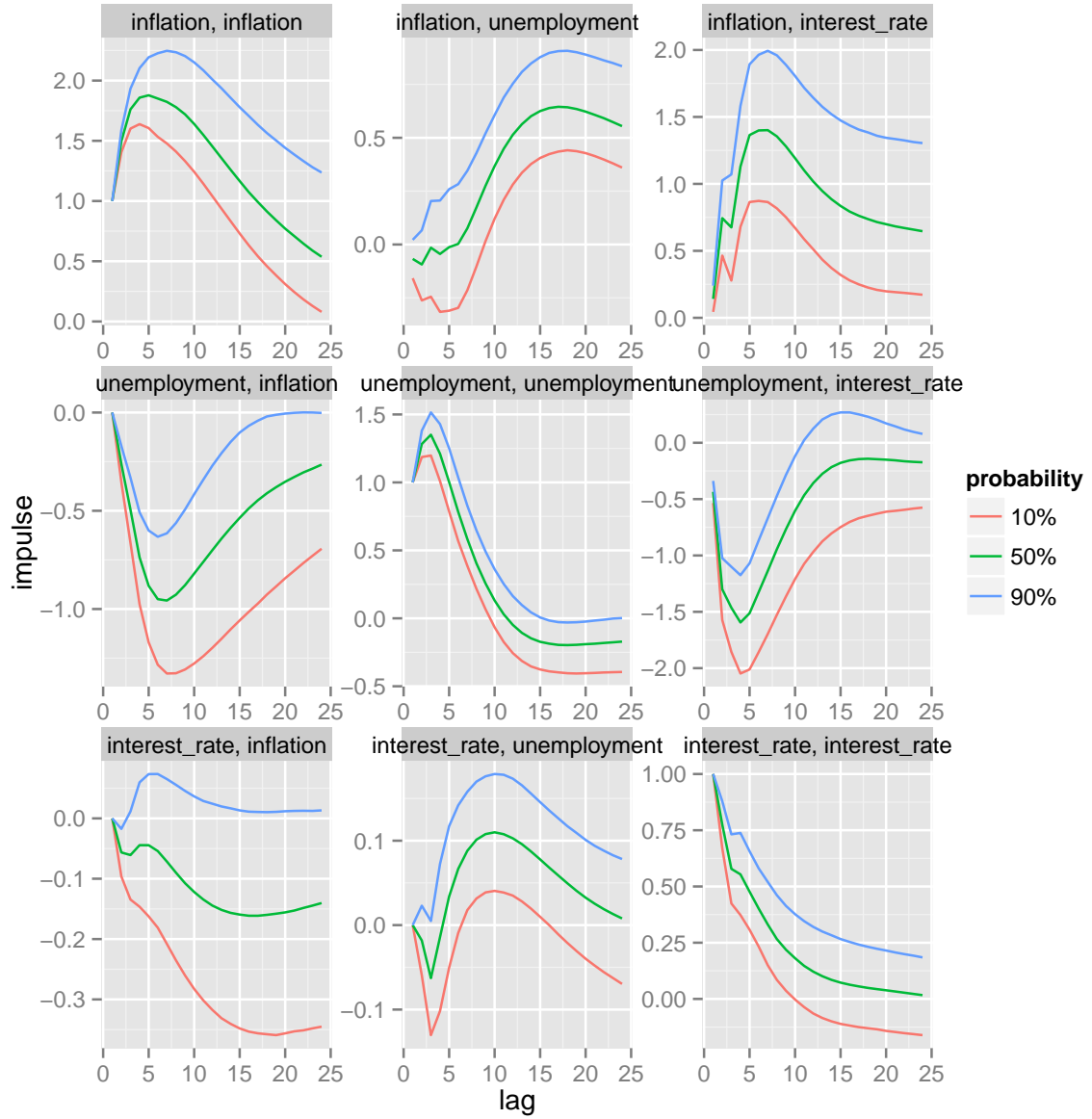
где  $\bar{V} = (\underline{V}^{-1} + X'X)^{-1}$ ,  $\bar{\alpha} = \text{vec}(\bar{A})$ ,  $\bar{A} = \bar{V}(\underline{V}^{-1}\underline{A} + X'X\hat{A})$ ,  $\bar{\nu} = T + \underline{\nu}$ ,  $\bar{S} = S + \underline{S} + \hat{A}'X'X\hat{A} + \underline{A}'\underline{V}^{-1}\underline{A} - \bar{A}'(\underline{V}^{-1} + X'X)\bar{A}$

Мы видим, что априорное распределение описывается огромным количеством гипер-параметров,  $\underline{S}$ ,  $\underline{\nu}$ ,  $\underline{\alpha}$  и  $\underline{V}$ . Обычно выбирают  $\underline{\alpha} = 0$ ,  $\underline{\nu} = M + 1$ ,  $\underline{S} = I_M$  и матрицу  $\underline{V}$  пропорциональной единичной матрице, например,  $\underline{V} = 10I_K$

```
model <- bvar(Yraw, prior="conjugate")
```

```
## Iteration 2000 out of 10000
## Iteration 4000 out of 10000
## Iteration 6000 out of 10000
## Iteration 8000 out of 10000
## Iteration 10000 out of 10000
```

```
bvar.imp.plot(model)
```



- Minnesota (Litterman) Априорное распределение Литтермана (или априорное распределение Миннесоты)

Данная модель является гремучей смесью байесовского и частотного подходов. От байесовского подхода взято предположение об априорном распределении параметра  $\alpha$ :

$$\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{U}) \quad (14)$$

От частотного подхода взята трактовка матрицы  $\Sigma$  как матрицы неизвестных констант, и поэтому используется некоторая оценка  $\hat{\Sigma}$ . На практике используют либо  $\hat{\Sigma} = \hat{S}/T$ , либо  $\hat{\Sigma} = (s_1^2, s_2^2, \dots, s_n^2)$ , где  $s_i^2$  — оценка дисперсии ошибки  $i$ -го уравнения, полученная с помощью обычного МНК.

У данного распределения огромное количество гиперпараметров! Чтобы исследователю не приходилось специфицировать их все, используются следующие ограничения.

Во-первых, предполагается, что матрица  $\underline{A}'$  имеет вид:

$$\underline{A}' = [0_{M \times 1}, \rho \cdot I_M, 0_{M \times M}, 0_{M \times M}, \dots, 0_{M \times M}] \quad (15)$$

то есть в математических ожиданиях мы априорно предполагаем, что каждый ряд является  $AR(1)$  процессом с коэффициентом авторегрессии равным  $\rho$ . Например, можно взять  $\rho = 0.9$ , что означает, что исследуемые ряды похожи на ряды с единичным корнем.

Во-вторых, описание матрицы  $\underline{U}$  сокращается до нескольких гиперпараметров. Матрица  $\underline{U}$  предполагается диагональной. Рассмотрим её диагональный элемент  $u$ , соответствующий уравнению для  $y_{ti}$ .

$$u = \begin{cases} a_1/r^2, & \text{для коэффициента при } y_{(t-r),i} \\ \frac{a_2\sigma_i^2}{r^2\sigma_j^2}, & \text{для коэффициента при } y_{(t-r),j} \text{ и } j \neq i \\ a_3\sigma_i^2 & \text{для коэффициента при константе и экзогенных переменных} \end{cases} \quad (16)$$

Скаляры  $a_1, a_2, a_3$  выбирает исследователь. Убывание  $u$  с ростом  $r$  отражает идею, что ближайшие по времени прошлые значения переменных оказывают более сильное влияние, чем отдаленные во времени. Выбор  $a_1 > a_2$  будет означать, что наиболее сильное влияние на эндогенную переменную оказывают её собственные лаговые значения, а не лаговые значения других эндогенных переменных. На практике при выборе констант  $\sigma_i^2$  снова примешивают частотную философию и полагают  $\sigma_i^2 = s_i^2$ .

Таким образом, задача исследователя сводится к выбору четырех априорных гиперпараметров,  $a_1, a_2, a_3, \rho$ . Величина  $a_1$  ограничивает величину коэффициентов при собственных лагах эндогенной переменной, величина  $a_2$  — коэффициента при лагах других эндогенных переменных,  $a_3$  — коэффициента при экзогенных переменных. Обычно значения  $a_1, a_2$  небольшие, от нуля до единицы, а  $a_3$  может быть довольно большим, т.к. априорно коэффициент при экзогенной переменной, например, при константе, может быть довольно большим. Выбор  $a_1 = a_2 = 0.5$  и  $a_3 = 100$  может оказаться разумным. Параметр  $\rho$  отвечает за априорное мнение исследователя о близости процессов к процессам с единичным корнем.

Существует много вариантов спецификации формулы для диагональных элементов матрицы  $\underline{U}$ . Например, (Kadiyala and Karlsson, 1997) делят на  $r$ , а не на  $r^2$ . А Canova, 2007 предлагает более общий вид формулы,

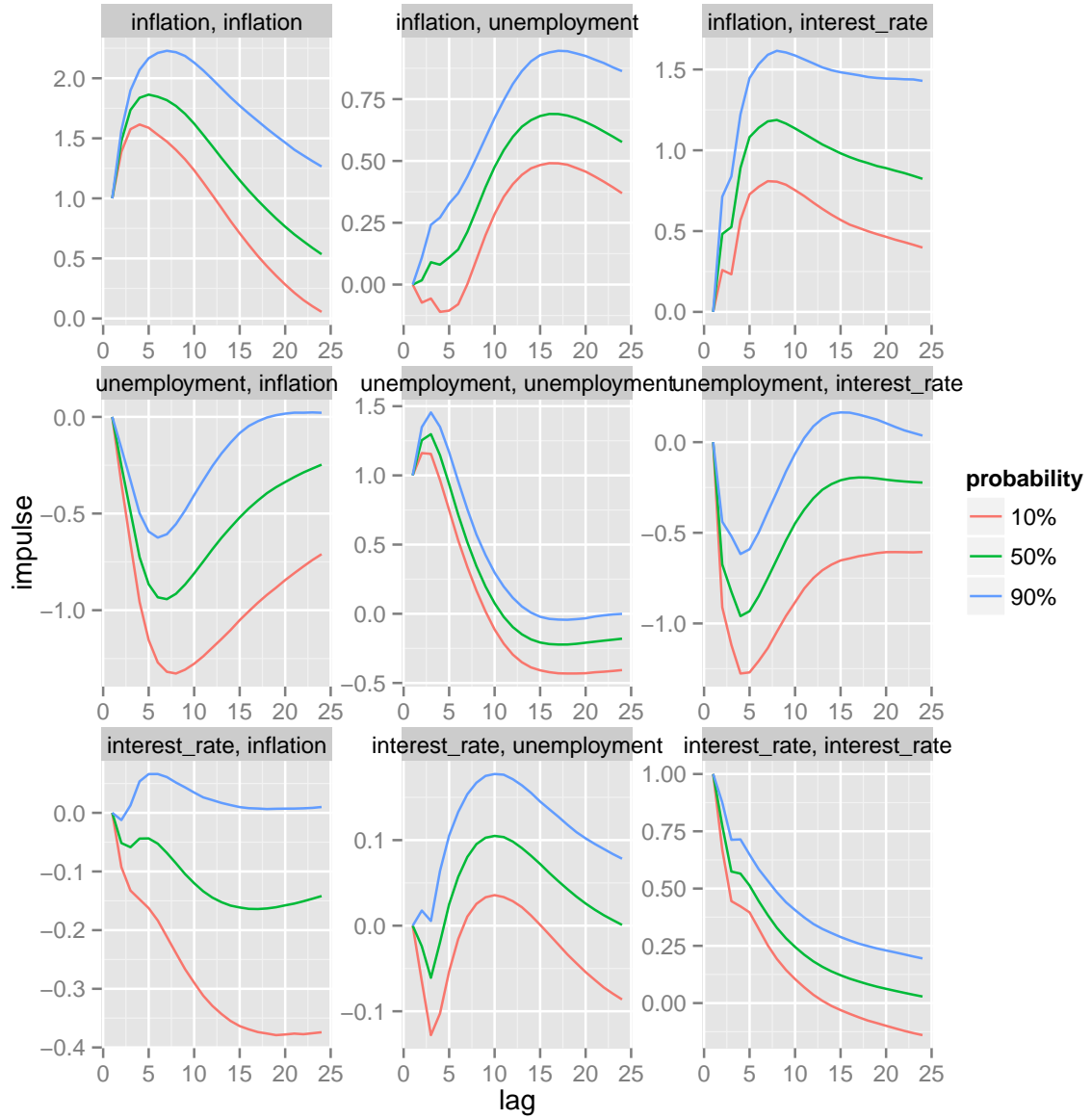
$$u = \begin{cases} a_1/d(r), & \text{для коэффициента при } y_{(t-r),i} \\ \frac{a_1a_2\sigma_j^2}{d(r)\sigma_i^2}, & \text{для коэффициента при } y_{(t-r),j} \text{ и } j \neq i \\ a_1a_3 & \text{для коэффициента при константе и экзогенных переменных} \end{cases} \quad (17)$$

где в качестве  $d(r)$  берут обычно либо степенную функцию  $d(r) = r^{a_4}$ , либо показательную,  $d(r) = r^{1-a_4}$ .

```
model <- bvar(Yraw, prior="minnesota")

## Iteration 2000 out of 10000
## Iteration 4000 out of 10000
## Iteration 6000 out of 10000
## Iteration 8000 out of 10000
## Iteration 10000 out of 10000

bvar.imp.plot(model)
```



- Независимое нормальное-Уишарта апприорное распределение

Параметры  $\alpha$  и  $\Sigma$  предполагаются независимыми,

$$\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{U}) \Sigma \sim \mathcal{W}^{-1}(\underline{S}, \underline{\nu}) \quad (18)$$

Условные апостериорные распределения будут иметь вид

$$\alpha|y, \Sigma \sim \mathcal{N}(\bar{\alpha}, \bar{U}) \Sigma|y, \alpha \sim \mathcal{W}^{-1}(\bar{S}, \bar{\nu}) \quad (19)$$

где  $\bar{\alpha} = \bar{U}(U\alpha + \sum_{i=1}^T Z_i' \Sigma^{-1} y_i)$  и  $\bar{U} = (\underline{U} + \sum_{i=1}^T Z_i' \Sigma^{-1} Z_i)^{-1}$ ,  $\bar{\nu} = T + \underline{\nu}$ ,  $\bar{S} = \underline{S} + \sum_{i=1}^T (y_i - Z_i \alpha)(y_i - Z_i \alpha)'$ .

Один из возможных вариантов, взять  $\underline{\alpha} = 0$ ,  $\underline{U}$  пропорционально единичной матрицы, так чтобы допускать значения коэффициентов из  $\alpha$  в широком диапазоне,  $\underline{\nu} = M + 1$  и  $\underline{S} = I_M$ .

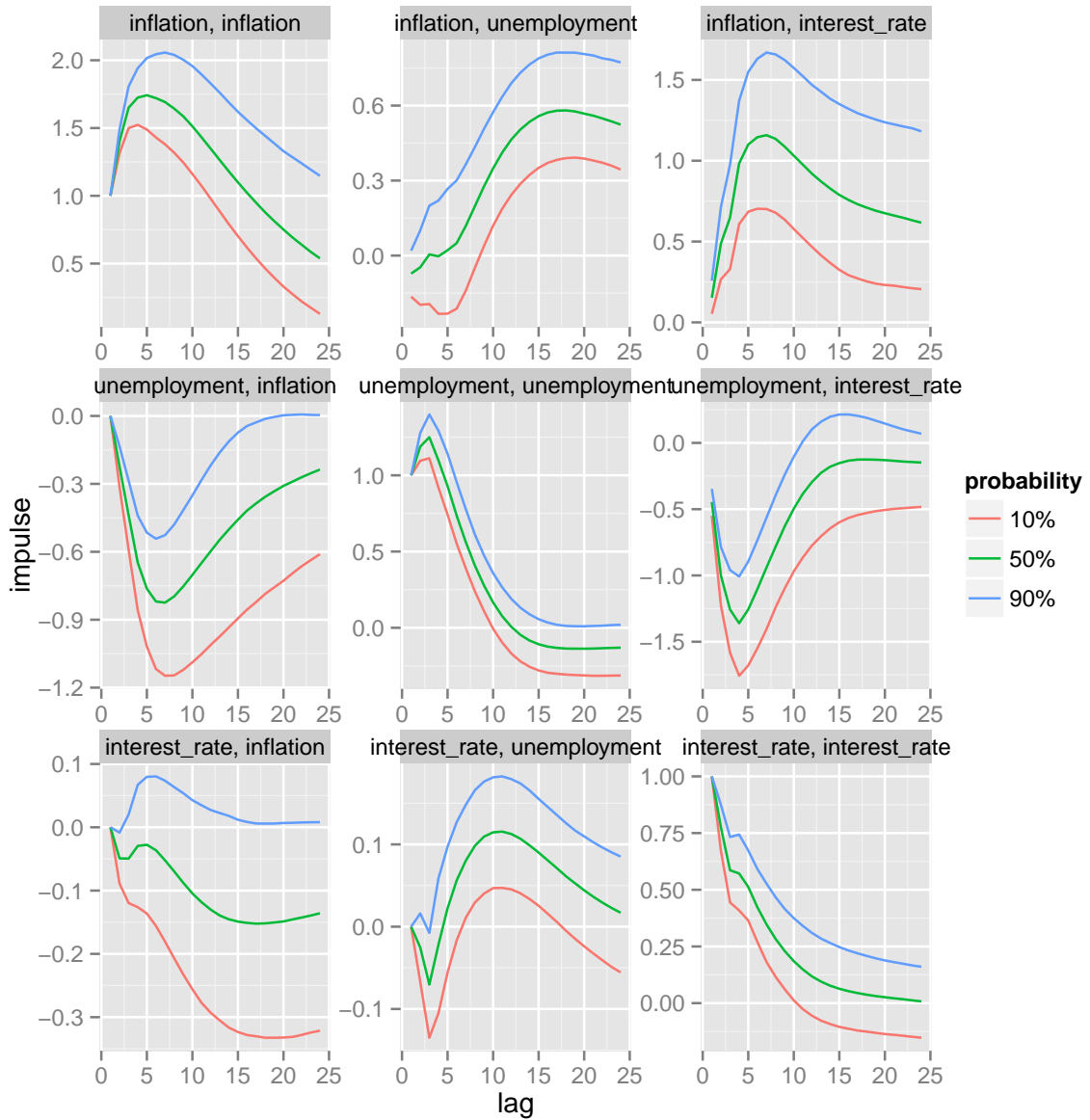
```
model <- bvar(Yraw, prior="independent")
```

```

## Iteration 2000 out of 12000
## Iteration 4000 out of 12000
## Iteration 6000 out of 12000
## Iteration 8000 out of 12000
## Iteration 10000 out of 12000
## Iteration 12000 out of 12000

```

```
bvar.imp.plot(model)
```



- Априорное нормальное распределение, спецификация Симс-Жа

Является еще более гремучей, чем модель Литтермана, смесью байесовского и частотного подхода. Помимо смеси двух философий здесь используется добавление искусственных наблюдений.

От байесовского подхода взято предположение об априорном распределении параметра  $\alpha$ :

$$\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{U}) \quad (20)$$



От частотного подхода взята трактовка матрицы  $\Sigma$  как матрицы неизвестных констант, её оценка  $\hat{\Sigma}$  получается путем построения оценивания обычной VAR(p) модели с константой.

Изначально имеется  $T$  наблюдений, из которых  $p$  стартовых наблюдений пропадает, так как у них невозможно посчитать лаговые значения порядка  $p$ . То есть к реальным  $T - p$  наблюдения Симс и Жа добавляют в модель  $m + 1$  искусственное наблюдение в начале временного ряда по следующей схеме:

$$Y_{dummy} = \begin{pmatrix} \mu_5 \bar{y}_1 & 0 & 0 & \dots & 0 \\ 0 & \mu_5 \bar{y}_2 & 0 & \dots & 0 \\ \dots & & & & \\ 0 & 0 & \dots & 0 & \mu_5 \bar{y}_m \\ \mu_6 \bar{y}_1 & \mu_6 \bar{y}_2 & \dots & \mu_6 \bar{y}_m \end{pmatrix}$$

А соответствующий вектор переменных правой части можно в виде блоков записать как  $X_{dummy} = (c \ Y_{dummy} \ Y_{dummy} \ \dots \ Y_{dummy})$ , где  $c$  — вектор-столбец,  $c = (0 \ 0 \ \dots \ 0 \ \mu_6)'$ .

Большое значение коэффициента  $\mu_5$  означает склонность исследователя считать, что при равенстве лаговых значений некоторой переменной её среднему  $\bar{y}_i$ , хорошим прогнозом будущего значения этой переменной также будет число  $\bar{y}_i$ . Используя  $\mu_5 \rightarrow \infty$  исследователь предполагает, что каждый ряд является процессом с единичным корнем, а коинтеграция отсутствует.

Большое значение коэффициента  $\mu_6$  означает склонность исследователя считать, что при равенстве лаговых значений всех переменных их средним  $\bar{y}_i$ , хорошим прогнозом будущего значения переменных также будут средние  $\bar{y}_i$ . Используя  $\mu_5 \rightarrow \infty$  исследователь предполагает, что либо каждый ряд является стационарным процессом, либо некоторые из рядов являются процессами с единичным корнем, коинтеграция при этом допускается.

Вектор  $\underline{a}$  определяется полностью аналогично модели Литтермана, то есть так чтобы матрица  $\underline{A}$  приняла вид

$$\underline{A}' = [0_{M \times 1}, \rho \cdot I_M, 0_{M \times M}, 0_{M \times M}, \dots, 0_{M \times M}] \quad (21)$$

Матрица  $\underline{U}$  предполагается диагональной, её диагональный элемент  $u$ , соответствующий уравнению для  $y_{ti}$  определяется как:

$$u = \begin{cases} \frac{\lambda_0 \lambda_1}{r \lambda_3 \sigma_j^2}, & \text{для коэффициента при } y_{(t-r),j} \text{ и любых } j \\ \lambda_0 \lambda_4 & \text{для коэффициента при константе} \\ \lambda_0 \lambda_5 & \text{для коэффициента при прочих экзогенных переменных} \end{cases} \quad (22)$$

где  $\lambda_0 \in [0; 1]$ , а остальные гиперпараметры  $\lambda_i$  и  $\mu_j$  больше неотрицательны.

Величины  $\lambda_2, \mu_1, \dots, \mu_4$  в модели Симс-Жа не используются. Такой стиль обозначений выбран для большей сопоставимостью с моделью Литтермана.

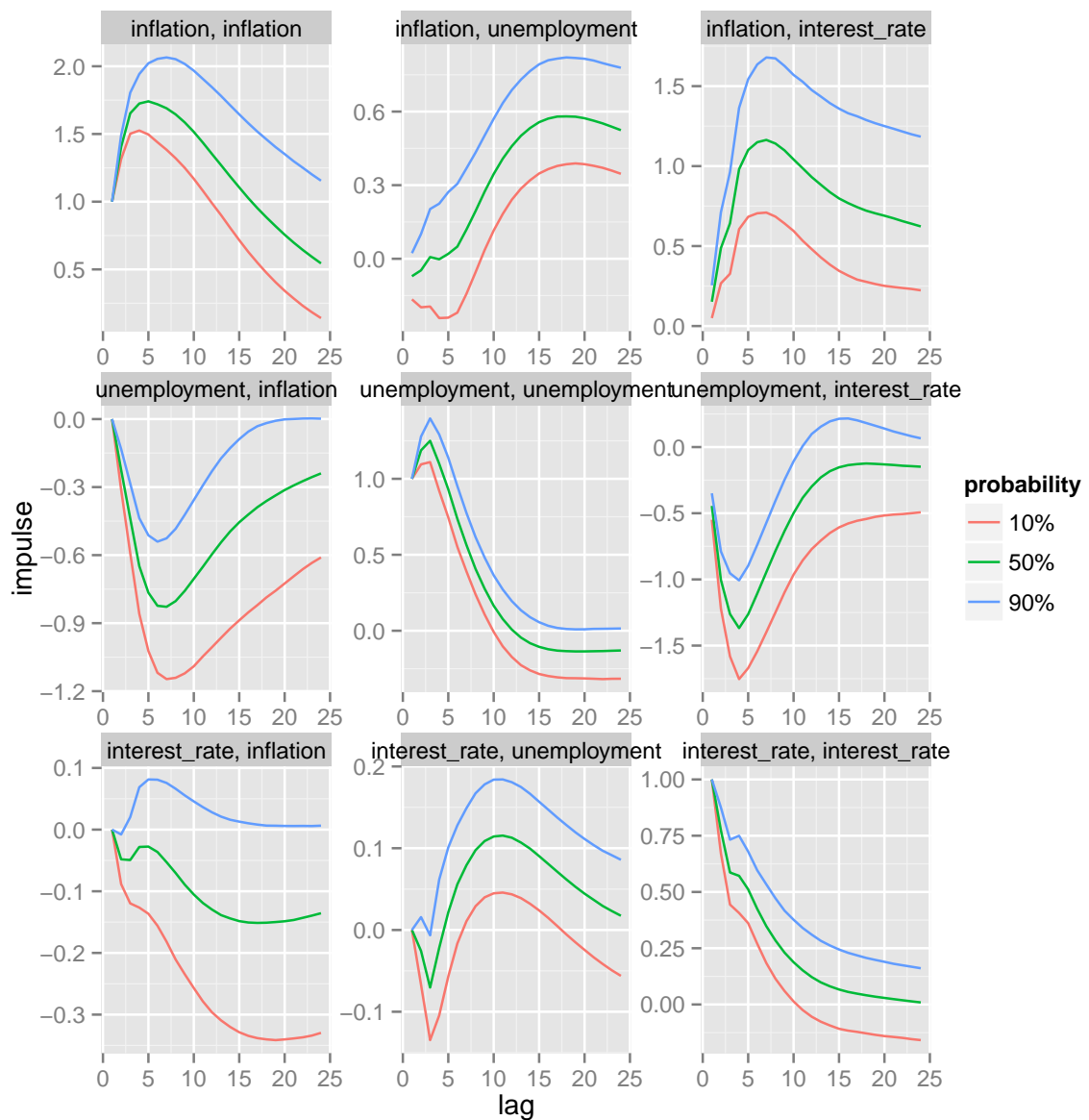
Кратко гиперпараметры модели Симс-Жа можно описать так:

Коэффициент	Диапазон	Смысл
$\lambda_0$	$[0; 1]$	Масштаб ковариационной матрицы
$\lambda_1$		Стандартное отклонение от $A_1$
$\lambda_3$		Скорость убывания дисперсии при росте лага
$\lambda_4$		Масштаб для констант
$\lambda_5$		Масштаб для коэффициентов при экзогенных переменных
$\mu_5$		Компонента суммы коэффициентов
$\mu_6$		Компонента исходного дамми-наблюдения

```
model <- bvar(Yraw, prior="independent")
```

```
## Iteration 2000 out of 12000
## Iteration 4000 out of 12000
## Iteration 6000 out of 12000
## Iteration 8000 out of 12000
## Iteration 10000 out of 12000
## Iteration 12000 out of 12000
```

```
bvar.imp.plot(model)
```

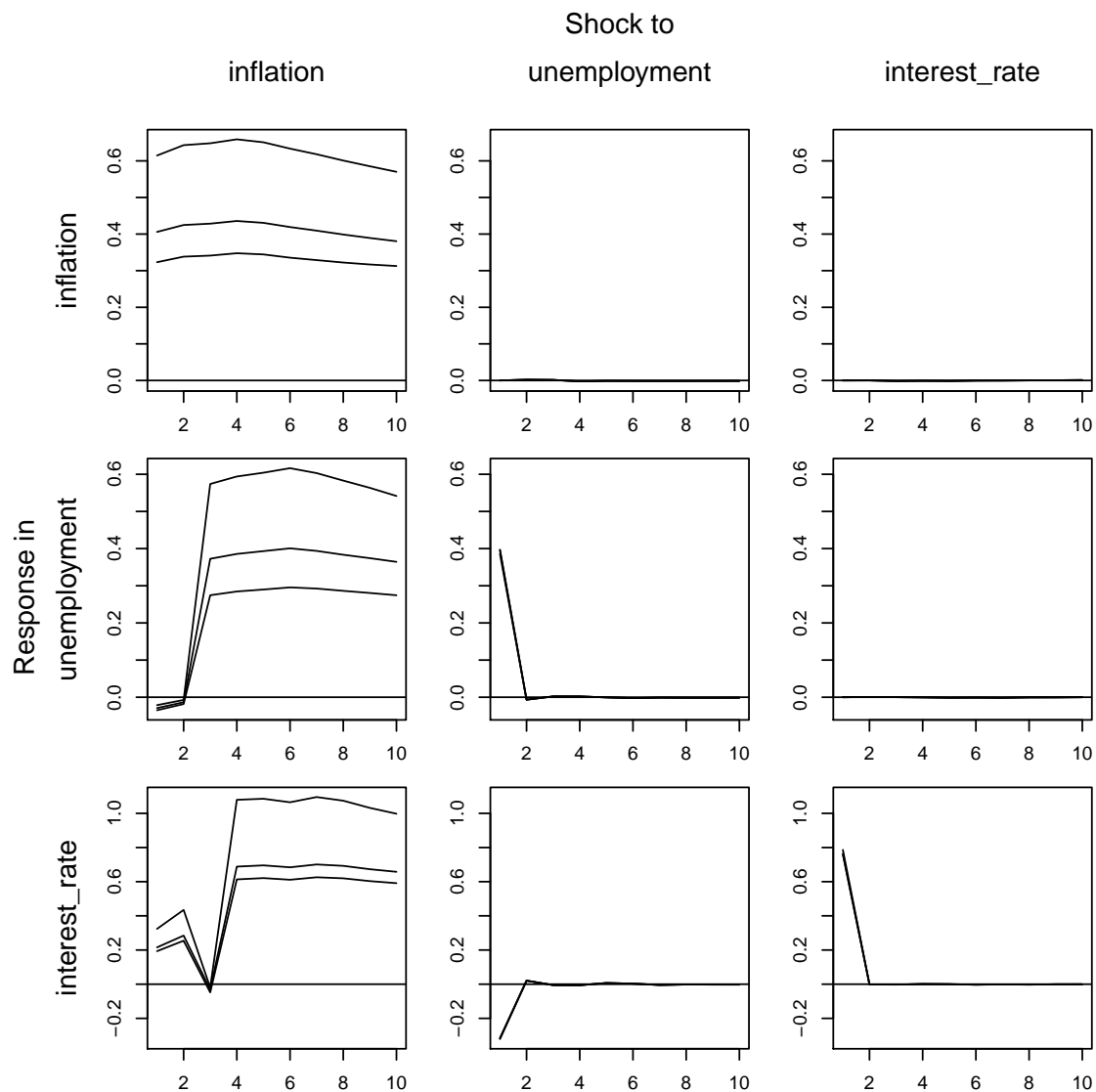


Спецификацией Симс-Жа можно воспользоваться и в независимом нормальном-Уишарта априорном распределении.

```
fit.BVAR <- szbvar(ts(Yraw), p=4,
                  lambda0=0.6, lambda1=0.1,
                  lambda3=2, lambda4=0.25, lambda5=0, mu5=0,
                  mu6=0, qm=4, posterior.fit=FALSE)
posterior.impulses <- mc.irf(fit.BVAR, nsteps=10, draws=5000)

## Monte Carlo IRF Iteration = 1000
## Monte Carlo IRF Iteration = 2000
## Monte Carlo IRF Iteration = 3000
## Monte Carlo IRF Iteration = 4000
## Monte Carlo IRF Iteration = 5000

plot(posterior.impulses)
```



За пределами данного обзора остались

- Модели VAR с параметрами меняющимися во времени
- Модели VAR со стохастическим отбором параметров