

# BVAR in Russia

я и Боря

13 августа 2015 г.

## 1 Обзор литературы

Построение точных макроэкономических прогнозов является ключевым условием проведения верной политики центральными банками. Хорошо известно, что центральные банки развитых государств опираются на большое число макроиндикаторов при проведении политики ((**stock\_watson\_2002\_forecasting** или (**bernanke\_al\_2005**) для США., какая-нибудь аналогичная работа для ЕЗ). Однако обычная векторно-авторегрессионная модель, ставшая наиболее часто встречающимся инструментом для построения прогнозов, не может учесть большое количество переменных, так как количество параметров, подлежащих оценке, растет нелинейно с увеличением числа уравнений. При этом неучтенная при построении VAR информация может приводить к смещенным оценкам и неверным выводам как относительно прогнозируемых значений, так и виде функций импульсных откликов. Основные способы учета большого числа переменных – это использование DF и Байесовских VAR. Динамические факторы были предложены в работах **forni\_hallin\_lippi\_generalized** и **stock\_watson\_2002\_forecasting** В указанных работах предполагается, что дисперсия большого количества временных рядов может быть описана с помощью нескольких, искусственно построенных (common factors) с помощью метода главных компонент. Распирением метода **stock\_watson\_2002\_forecasting** служит FAVAR, предложенная в статье (**bernanke\_al\_2005**). В рамках FAVAR несколько динамических факторов добавляются как дополнительные переменные в обычную VAR. Цель данной работы состоит в построении прогноза основных макроиндикаторов (выпуска, инфляции и др.) для российской экономики. Задача осложняется отсутствием большого количества длинных временных рядов, что не позволяет провести построение DFM. Байесовские модели зарекомендовали себя как хороший инструмент построения прогнозов. В ряде работ было показано, что они обеспечивают более низкую ошибку прогноза, чем, например, обычные VAR и VECM. (?) две стратегии – использовать shrinkage, который становится уже с ростом числа переменных По Bloor and Matheson надо бы переписать.

Модификация Litterman prior была предложена в работах **doan\_litterman\_sims\_1984\_forecast** (sum of coefficients prior) и **sims\_1993\_nine** (co-persistence prior). Комбинация этих трех априорных распределений была использована в работе **robertson\_tallman\_1999\_vector** для предсказания безработицы, темпа роста и инфляции. В работе показано, что

смешанное априорное распределение получает получить более точные прогнозы, чем априорные распределения Litterman 1986 и **sims\_zha\_1998\_bayesian** (про эту работу пока ничего не писала).

Ключевую роль в развитии подхода сыграла статья (**de\_mol\_al\_2008\_forecasting**). В этой работе на основе асимптотического анализа было показано, что если данные характеризуются высокой мультиколлинеарностью (что характерно для выборок макрорядов большой размерности) сужение априорного распределения при увеличении количества переменных дает больший вес нескольким первым главным компонентам. Это означает, что если данные характеризуются факторной структурой, то наложение более узких априорных распределений с увеличением размерности модели не приводит к потере важной информации, т.к. для описания данных достаточно небольшого количества первых факторов. Эта точка зрения была подтверждена и развита в статье **banbura\_al\_2010\_large** в которой авторы строят VAR модели для 3, 7, 20 и 131 переменных и показывают, что модели с большей размерностью демонстрируют лучшие прогнозные способности, чем модели малой размерности и даже FAVAR (? Проверить, об этом пишет Бошман). Интересно отметить, что хорошая прогнозная способность достигается уже в модели с 20 переменными, поэтому как для прогнозирования, так и для структурного анализа достаточно сконцентрироваться на агрегированных данных. Аналогичная модель для Новой Зеландии была построена в работе **bloor\_matheson\_2010\_analysing** в которой они использовали метод условного прогнозного оценивания (**waggoner\_zha\_1999\_conditional**), что позволило им сравнить сценарии, основанные на различной условной информации. Строят три модели (с 9, 13 и 35 переменными), делают вывод, что BVAR обладает более высокой предсказательной способностью, чем AR и обычная VAR модель. Хотя результаты варьируют по разным переменным, в общем и целом, BVAR с большим числом переменных характеризуется более высокой точностью прогноза. Из Beauchemin Koop (2010) расширил результаты **banbura\_al\_2010\_large** и показал, что BVAR с большой размерностью обладают лучшей прогнозной способностью даже по отношению к более сложным моделям (???) Тот же метод построения априорного распределения (естественно-сопряженная версия Миннесоты-распределения (**kadiyala\_karlsson\_1997\_numerical** **sims\_zha\_1998\_bayesian**- проверить, они ли предложили? ), что в работах **banbura\_al\_2010\_large**, **bloor\_matheson\_2010\_analysing** и Кооп (2010) был применен в работе **beauchemin\_zaman\_2011\_medium** Они показывают, что BVAR с 16 переменными может быть с успехом использована как для прогнозов, так и для структурного анализа (трансмиссии монетарного шока (?)/структурного анализа монетарной политики. Аналогичное построение априорного распределения используется в работе **alessandri\_mumtaz\_2014** где с помощью линейной и нелинейной BVAR (?) модели показано, что учет финансовых индикаторов позволяет улучшить прогноз выпуска и инфляции, в т. ч. в кризисные периоды.

Во всех работах гиперпараметр, контролирующий жесткость (?), выбирается таким образом, чтобы максимизировать функцию правдоподобия данных (это максимизирует точность вневыборочного прогноза?). В работе **geweke\_whiteman\_2006\_bayesian**

было показано, что такой выбор гиперпараметра минимизирует ошибки прогноза на один период.

Сам Литтерман в своей работе показал, что использование априорного распределения (Bayesian shrinkage) в BVAR с не менее чем шестью переменными улучшает прогнозную силу модели. Однако до последнего времени считалось, что при использовании достаточно большого числа временных рядов уточнения правдоподобия только с помощью априорного распределения недостаточно. Это приводило к необходимости задавать дополнительные ограничения.

## **2 Методология**

### **2.1 Удобная табличка**

Буква	Размер	Описание	Формула
$p$	скаляр	количество лагов	
$m$	скаляр	количество эндогенных переменных	
$d$	скаляр	количество экзогенных переменных	
$k$	скаляр	количество параметров в одном уравнении	$k = mp + d$
$T$	скаляр	количество наблюдений	
$z_t$	$d \times 1$	вектор экзогенных переменных (считая константу)	
$y_t$	$m \times 1$	вектор эндогенных переменных	$y_t = \Phi' x_t + \varepsilon_t$
$x_t$	$k \times 1$	вектор всех регрессоров	$x_t = [y'_{t-1} \dots y'_{t-p} z'_t]'$
$\varepsilon_t$	$m \times 1$	вектор случайных ошибок	
$Y$	$T \times m$	все эндогенные переменные	$Y = [y_1, y_2, \dots, y_T]'$
$X$	$T \times k$	матрица регрессоров	$X = [x_1, x_2, \dots, x_T]'$
$E$	$T \times m$	матрица ошибок	$E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$
$y$	$mT \times 1$	векторизация $Y$	$y = \text{vec}(Y)$
$\varepsilon$	$mT \times 1$	векторизация $E$	$\varepsilon = \text{vec}(E)$
$\Phi_1, \dots$	$m \times m$	коэффициенты VAR	$y_t = \Phi_1 y_{t-1} + \dots + \Phi_{ex} z_t + \varepsilon_t$
$\Phi_{ex}$	$m \times d$	коэффициенты при экзогенных переменных	
$\Phi$	$k \times m$	упаковка матриц $\Phi_1, \dots$	$\Phi = [\Phi_1 \dots \Phi_p \Phi_{ex}]'$
$\phi$	$km \times 1$	вектор из матрицы $\Phi$	$\phi = \text{vec } \Phi$
$\Xi$	$km \times km$	Априорная ковариационная матрица $\Phi$	
$\underline{\Phi}$	$k \times m$	априорное математическое ожидание $\Phi$	
$\underline{\phi}$	$km \times 1$	вектор из матрицы $\underline{\Phi}$	$\underline{\phi} = \text{vec } \underline{\Phi}$
$\bar{\Xi}$	$km \times km$	Апостериорная ковариационная матрица $\Phi$	
$\bar{\Phi}$	$k \times m$	апостериорное математическое ожидание $\Phi$	
$\bar{\phi}$	$km \times 1$	вектор из матрицы $\bar{\Phi}$	$\bar{\phi} = \text{vec } \bar{\Phi}$
$\underline{\nu}$	скаляр		
$\bar{\nu}$	скаляр		$\bar{\nu} = T + \underline{\nu}$
$\underline{\Omega}$	$k \times k$	Матрица априорных масштабирующих коэффициентов ковариационной матрицы $\Phi$	$\Xi = \Sigma \otimes \underline{\Omega}$
$\bar{\Omega}$	$k \times k$	Матрица апостериорных масштабирующих коэффициентов ковариационной матрицы $\Phi$	$\bar{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1}, \bar{\Xi} = \Sigma \otimes \bar{\Omega}$
$\Sigma$	$m \times m$	Ковариационная матрица ошибок	$\mathbb{E} \varepsilon_t \varepsilon_t' = \Sigma$

## 2.2 Байесовская VAR

Рассмотрим переменные  $y_{it}$ , объединенные в вектор  $y_t = (y_{1t}, y_{2t}, \dots, y_{mt})'$  размерности  $m$ . Векторная авторегрессия в сокращенной форме записывается в виде:

$$y_t = \Phi_{ex} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \quad (1)$$

где  $\Phi_{ex} = (c_1, \dots, c_m)'$  — вектор констант размерности  $m$ ,  $\Phi_l$  — авторегрессионные матрицы размерности  $m \times m$  при  $l = 1, \dots, p$ . Вектор  $\varepsilon_t$  —  $m$ -мерный вектор ошибок с ковариационной матрицей  $\mathbb{E} \varepsilon_t \varepsilon_t' = \Sigma$ , некоррелированный с объясняющими переменными. Группируя матрицы параметров в общую матрицу  $\Phi = [\Phi_1 \dots \Phi_p \Phi_{ex}]'$  и определяя новый вектор  $x_t = [y_{t-1}' \dots y_{t-p}' 1]'$ , получаем VAR записанную в более компактном виде:

$$y_t = \Phi' x_t + \varepsilon_t \quad (2)$$

Если же сгруппировать переменные и шоки следующим образом:  $Y = [y_1, y_2, \dots, y_T]'$ ,  $X = [x_1, x_2, \dots, x_T]'$ ,  $E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$  то VAR можно записать как:

$$Y = X\Phi + E \quad (3)$$

Эта же модель может быть записано в векторизованном виде<sup>1</sup>:

$$Y = X\Phi + E \quad (4)$$

$$\text{vec}(Y) = \text{vec}(X\Phi I) + \text{vec}(E) \Leftrightarrow \quad (5)$$

$$y = (I_M \otimes X)\phi + \varepsilon \quad (6)$$

где  $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T)$  и вектор  $\phi = \text{vec} \Phi$  имеет размерность  $km \times 1$ .

Задача байесовского оценивания заключается в поиске апостериорных распределений параметров  $p(\Phi, \Sigma|Y)$  с использованием функции максимального правдоподобия,  $L(Y|\Phi, \Sigma)$ , и заданного априорного распределения,  $p(\Phi, \Sigma|Y)$ . Для этого используется правило Байеса:

$$p(\Phi, \Sigma|Y) = \frac{p(\Phi, \Sigma)L(Y|\Phi, \Sigma)}{p(Y)} \quad (7)$$

Т.к.  $p(Y)$  не зависит от  $\Phi$  и  $\Sigma$ , то можно записать:

$$p(\Phi, \Sigma|Y) \propto p(\Phi, \Sigma)L(Y|\Phi, \Sigma) \quad (8)$$

Так как  $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ , то функция правдоподобия задается как:<sup>2</sup>

$$L(Y|\Phi, \Sigma) \propto |\Sigma|^{-T/2} \text{etr} \left\{ -\frac{1}{2} [\Sigma^{-1}(Y - X\Phi)'(Y - X\Phi)] \right\} \quad (9)$$

В двух следующих разделах будут разобраны наиболее известные априорные и построенные на их основе апостериорные распределения.

<sup>1</sup>Третье уравнение системы следует из тождества:  $\text{vec}(ABC) = (C \otimes A) \text{vec}(B)$

<sup>2</sup> Другая форма записи функции правдоподобия:  $L(Y|\Phi, \Sigma) \propto |\Sigma|^{-T/2} \text{etr} \left\{ -\frac{1}{2} [\Sigma^{-1} \hat{E}' \hat{E}] \right\} \times \text{etr} \left\{ -\frac{1}{2} [\Sigma^{-1}(\Phi - \hat{\Phi})' X' X (\Phi - \hat{\Phi})] \right\}$ , где  $\hat{E} = Y - X\hat{\Phi}$  и  $\hat{\Phi} = (X'X)^{-1} X'Y$ . Здесь  $\text{etr}() = \exp(\text{tr}())$ .

## 2.3 Классификация популярных априорных распределений

Детально каждое из априорных распределений будет описано отдельно, а в этом разделе будет дана общая схема связи априорных распределений. Среди самых популярных априорных распределений можно назвать:

1. Независимое нормальное-обратное Уишарта распределение

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \underline{\nu}) \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (10)$$

В общем случае выборку из апостериорного распределения можно получить по схеме Гиббса.

Частными случаями независимого нормального-обратного Уишарта являются:

- (a) Распределение Миннесоты

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma = \text{const} \end{cases} \quad (11)$$

Получается из независимого нормального-обратного Уишарта при  $\underline{S} = (\underline{\nu} - m - 1) \cdot \Sigma$  и  $\underline{\nu} \rightarrow \infty$ . Апостериорное распределение выписывается в явной форме. Можно использовать алгоритм Монте-Карло, сразу генерирующий случайную выборку из апостериорного распределения, без необходимости периода «прожига».

Более того, алгоритм симуляции упрощается если матрица  $\Xi$  имеет структуру кронекерова произведения  $\Xi = \Sigma \otimes \underline{\Omega}$ . В этом случае распределение Миннесоты становится частным случаем сопряжённого нормального-обратного Уишарта.

- (b) Независимое нормальное-Джефффри

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (12)$$

Получается из независимого нормального-обратного Уишарта при  $\underline{S} = \underline{\nu}^{1/m} \cdot I$  и  $\underline{\nu} \rightarrow 0$ .

как тут с плотностью растущей к необратимым матрицам ??????????

[http://www.tc.umn.edu/~nydic001/docs/unpubs/Wishart\\_Distribution.pdf](http://www.tc.umn.edu/~nydic001/docs/unpubs/Wishart_Distribution.pdf) — тут пишут, что при  $\nu < \dots$  получается необратимая матрица, а тут вроде как обратимая???

Для получения выборки из апостериорного распределения можно использовать схему Гиббса. Необходимые формулы для гиперпараметров апостериорного распределения получаются из общего случая просто подстановкой  $\underline{S} = 0$ ,  $\underline{\nu} = 0$ .

Распределение Миннесоты и независимое нормальное-Джефффри являются противоположными крайностями независимого-обратного Уишарта. В распределении Миннесоты матрица  $\Sigma$  предполагается известной, а в нормальном-Джефффри матрица  $\Sigma$  имеет «размытое» неинформативное распределение.

Частным случаем независимого нормального-Джефффри распределения является:

i. Неинформативное-Джефффри

$$\begin{cases} \phi \sim 1 \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (13)$$

Априорное распределение для ленивых. Не нужно указывать ни одного гиперпараметра!

Получается из независимого нормального-Джефффри при  $\underline{\phi} = 0$  и  $\underline{\Xi} = a \cdot I$  и  $a \rightarrow \infty$ . Является также частным случаем сопряженного нормального-Джефффри априорного распределения.

Для получения выборки из апостериорного распределения можно использовать прямое моделирование по схеме Монте-Карло без алгоритма Гиббса. распределения получаются из общего случая просто подстановкой  $\underline{S} = 0$ ,  $\underline{\nu} = 0$ ,  $\underline{\Xi}^{-1} = 0$ ,  $\underline{\phi} = 0$ . При этом формулы существенно упрощаются, в частности исчезает необходимость обращать матрицу размера  $km \times km$ .

## 2. Сопряженное нормальное-обратное Уишарта распределение

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \phi | \Sigma \sim \mathcal{N}(\underline{\phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \quad (14)$$

Для сопряженного нормального-обратного Уишарта распределения нет необходимости использовать алгоритм Гиббса, так как есть явные формулы для апостериорных распределений. Можно использовать алгоритм Монте-Карло, сразу генерирующий случайную выборку из апостериорного распределения. Исчезает период «прожига», необходимый для сходимости алгоритма Гиббса.

Частным случаем сопряжённого нормального-обратного Уишарта распределения оказывается:

(a) Распределение Миннесоты при  $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$ .

При  $\Xi = \Sigma \otimes \underline{\Omega}$  распределение Миннесоты является и частным случаем независимого нормального-обратного Уишарта, и сопряженного нормального-обратного Уишарта. Для получения выборки из апостериорного распределения можно использовать и схему Гиббса для независимого нормального-обратного Уишарта, и алгоритм Монте-Карло без «прожига».

(b) Сопряжённое нормальное-Джеффри

$$\begin{cases} \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi|\Sigma \sim \mathcal{N}(\underline{\phi}; \Sigma \otimes \underline{\Omega}) \end{cases} \quad (15)$$

Получается из сопряженного нормального-обратного Уишарта при  $\underline{S} = \underline{\nu}^{1/m} \cdot I$  и  $\underline{\nu} \rightarrow 0$ . Формулы для гиперпараметров апостериорного распределения получаются подстановкой  $\underline{\nu} = 0$ ,  $\underline{S} = 0$ .

Частным случаем сопряженного нормального-Джеффри является

i. Неинформативное-Джеффри

Получается из сопряженного нормального-Джеффри при  $\underline{\phi} = 0$  и  $\underline{\Omega} = a \cdot I$  и  $a \rightarrow \infty$ . Формулы для гиперпараметров апостериорного распределения получаются подстановкой  $\underline{\phi} = 0$ ,  $\underline{\Omega}^{-1} = 0$ ,  $\underline{\nu} = 0$ ,  $\underline{S} = 0$ .

????? Где в этой схеме Березовский априорное распределение Sim-Zha?

В нашей работе мы всегда явно указываем, идёт ли речь о независимом или сопряженном априорном распределении. Однако при чтении многих других работ надо быть внимательным, зачастую авторы говорят о «нормальном-обратном Уишарта» распределении, не уточняя, какое имеется ввиду.

## 2.4 Априорное распределение Миннесоты

Использование векторных авторегрессий в макроэкономическом анализе явилось следствием критики активно использовавшихся прежде традиционных эконометрических моделей. В частности, Sims(1980) делал акцент на необоснованности ограничений, вводимых в рамках традиционных моделей, и предложил использовать более простую по построению динамическую модель, основанную на разложении Вольда и не требующую введения никаких ограничений на взаимную динамику переменных - модель векторной авторегрессии. Однако для того чтобы правильно отражать динамику фактических временных рядов, VAR часто требуется большое количество лагов, что, в свою очередь, может снизить эффективность оценивания(?) и привести к высоким ошибкам прогноза. Одно из решений этой проблемы было предложено в работах **litterman\_1979** и **litterman\_1980** автор которых показал, что введение ограничений в форме априорных распределений параметров увеличивает точность оценок и прогнозов. Априорное распределение, позже(?) получившее название «априорное распределение Миннесоты» было предложено в работе **litterman\_1986\_forecasting** и (с некоторыми модификациями) в **doan\_litterman\_sims\_1984\_forecasting**



Априорное распределение параметров предполагается многомерным нормальным, зависящим от нескольких гиперпараметров. Параметры предполагаются независимыми, следовательно, их ковариационная матрица диагональна. Ковариационная матрица вектора  $\varepsilon_t$  также предполагается диагональной и постоянной. Тогда вектор  $\Phi$  не зависит от  $\Sigma$ :

$$\phi \sim \mathcal{N}(\phi, \Xi) \quad (16)$$

Априорная плотность распределения  $\phi$  также не зависит от  $\Sigma$  и может быть записана как:

$$p(\phi) = \frac{1}{(2\pi)^{km/2} |\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} (\phi - \underline{\phi})' \Xi^{-1} (\phi - \underline{\phi}) \right\}. \quad (17)$$

Комбинируя её с функцией правдоподобия (9), получаем, что апостериорное распределение параметров задаются в следующем виде:

$$\phi|Y \sim \mathcal{N}(\bar{\phi}, \bar{\Xi}) \quad (18)$$

где

$$\begin{aligned} \bar{\Xi} &= [\Xi^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} \\ \bar{\phi} &= \bar{\Xi} [\Xi^{-1} \underline{\phi} + (\Sigma^{-1} \otimes X')y]. \end{aligned}$$

Если  $\Xi$  имеет структуру кронекерова произведения,  $\Xi = \Sigma \otimes \underline{\Omega}$ , то формулы можно существенно упростить и обойтись обращением матриц меньшей размерности:

$$\begin{aligned} \bar{\Xi} &= [\Xi^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = [(\Sigma \otimes \underline{\Omega})^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = \\ &= [\Sigma^{-1} \otimes \underline{\Omega}^{-1} + \Sigma^{-1} \otimes (X'X)]^{-1} = \Sigma \otimes (\underline{\Omega}^{-1} + X'X)^{-1} = \Sigma \otimes \bar{\Omega} \end{aligned} \quad (19)$$

В результате получаем

$$\Phi|Y \sim \mathcal{N}(\bar{\Phi}, \Sigma \otimes \bar{\Omega})$$

На практике в качестве матрицы  $\Sigma$  используют её оценку  $\hat{\Sigma}$ , диагональные элементы которой равны:  $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \dots, \hat{\sigma}_m^2$ , где  $\hat{\sigma}_i^2$  — оценка дисперсии случайной составляющей в  $AR(p)$  модели для ряда  $i$ .<sup>3</sup>

Математическое ожидание априорного распределения параметров может быть записано с помощью матрицы  $\underline{\Phi} = \mathbb{E}(\Phi)$  размерности  $k \times m$ , где  $\underline{\Phi} = [\underline{\Phi}_1 \dots \underline{\Phi}_p \ \underline{\Phi}_{ex}]'$  и  $\underline{\phi} = \text{vec } \underline{\Phi}$ .

$$(\underline{\Phi}_l)_{ij} = \begin{cases} \delta_i & i = j, l = 1; \\ 0, & \text{в остальных случаях} \end{cases} \quad (20)$$

---

<sup>3</sup>Некоторые авторы для подсчета оценки дисперсии используют AR(1) модель, даже если сама VAR имеет большее количество лагов.

Априорное распределение Миннесоты было задумано таким образом, чтобы учесть нестационарность многих макроэкономических временных рядов. В этом случае  $\delta_i$ , принимают значение единица.<sup>4</sup>

Априорное распределение Миннесоты предполагает, что априорная ковариационная матрица параметров  $\Xi$  диагональна. Пусть  $\Xi_i$  обозначает блок  $\Xi$ , размера  $k \times k$ , связанный с коэффициентами уравнения  $i$ , т.е.

$$\Xi = \begin{pmatrix} \Xi_1 & 0_{k \times k} & \cdots & 0_{k \times k} & 0_{k \times k} \\ 0_{k \times k} & \Xi_2 & \cdots & 0_{k \times k} & 0_{k \times k} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{k \times k} & 0_{k \times k} & \cdots & \Xi_{m-1} & 0_{k \times k} \\ 0_{k \times k} & 0_{k \times k} & \cdots & 0_{k \times k} & \Xi_m \end{pmatrix} \quad \Xi_i = \begin{pmatrix} \Xi_{i,lag=1} & 0_{m \times m} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ 0_{m \times m} & \Xi_{i,lag=2} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m \times m} & 0_{m \times m} & \cdots & \Xi_{i,lag=p} & 0_{m \times 1} \\ 0_{1 \times m} & 0_{1 \times m} & \cdots & 0_{1 \times m} & \Xi_{i,const} \end{pmatrix}$$

Тогда диагональные элементы  $\Xi_{i,lag=l}$  определяются по формулам:

$$(\Xi_{i,lag=l})_{jj} = \begin{cases} \left( \frac{\lambda_{tight}}{l^{\lambda_{lag}}} \right)^2, & j = i \\ \left( \frac{\lambda_{tight} \cdot \lambda_{kron} \sigma_i}{l^{\lambda_{lag} \sigma_j}} \right)^2, & j \neq i \end{cases} \quad \Xi_{i,const} = \lambda_{const}^2 \sigma_i^2, \quad (21)$$

Как можно видеть из приведенной выше формулы (21) априорная дисперсия параметров зависит от нескольких гиперпараметров, задаваемых исследователем. Гиперпараметры имеют следующую интерпретацию:  $\lambda_{tight}$  (параметр регуляризации) отражает общую «жесткость» априорного распределения. Если  $\lambda_{tight} \rightarrow 0$ , то априорное распределение полностью определяет апостериорное распределение, и данные не играют никакой роли при оценке параметров. Наоборот, если  $\lambda_{tight} \rightarrow \infty$ , то априорное распределение перестает влиять и оценка параметров сходится к обычной оценке МНК. Параметр  $\lambda_{kron}$  (параметр кросс-регуляризации) добавляет дополнительную жесткость лагам других переменных по сравнению с лагами зависимой переменной. Если  $\lambda_{kron} < 1$ , то собственные лаги зависимой переменной помогают предсказывать значение переменной лучше, чем лаги других переменных, поэтому коэффициенты при лагах других переменных оказываются жестче регуляризованы к нулю. Параметр  $\lambda_{const}$  отражает относительную жесткость распределения константы.

При  $\lambda_{kron} = 1$  матрица  $\Xi$  имеет структуру кронекерова произведения и представима в виде:

$$\Xi = \Sigma \otimes \underline{\Omega},$$

где  $\underline{\Omega}$  — матрица размера  $k \times k$ , соответствующая отдельному уравнению. Кронекерово домножение слева на матрицу  $\Sigma$  для  $i$ -го уравнения означает домножение дисперсий, указанных в матрице  $\underline{\Omega}$ , на коэффициент  $\sigma_i^2$ . Сама  $\underline{\Omega}$  представима в виде:

---

<sup>4</sup>В настоящее время широкое распространение получила практика назначать  $\delta_i = 1$  для нестационарных рядов и  $\delta_i < 1$  для стационарных.

$$\underline{\Omega} = \begin{pmatrix} \underline{\Omega}_{lag=1} & 0_{m \times m} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ 0_{m \times m} & \underline{\Omega}_{lag=2} & \cdots & 0_{m \times m} & 0_{m \times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m \times m} & 0_{m \times m} & \cdots & \underline{\Omega}_{lag=p} & 0_{m \times 1} \\ 0_{1 \times m} & 0_{1 \times m} & \cdots & 0_{1 \times m} & \underline{\Omega}_{const} \end{pmatrix} \quad (22)$$

При этом матрица  $\underline{\Omega}_{lag=l}$  имеет размерность  $m \times m$ , и ее диагональные элементы определяются по формулам:

$$(\underline{\Omega}_{lag=l})_{jj} = \left( \frac{\lambda_{tight}}{l^{\lambda_{lag}} \sigma_j} \right)^2 \quad \underline{\Omega}_{const} = \lambda_{const}^2 \quad (23)$$

При использовании априорного распределения Миннесоты, нет необходимости применять алгоритм Гиббса для получения апостериорного распределения. Алгоритм генерации случайной выборки непосредственно из апостериорного распределения происходит методом Монте-Карло:

1. На  $s$ -ом шаге сгенерировать очередную итерацию согласно:

$$\phi^{[s]} \sim \mathcal{N}(\bar{\phi}; \Xi) \quad (24)$$

2. Увеличить  $s$  на единицу и перейти к пункту один

Если  $\Xi$  имеет структуру кронекерова произведения,  $\Xi = \Sigma \otimes \Omega$ , то вместо вектора  $\phi^{[s]}$  можно генерировать матрицу  $\Phi^{[s]}$  численно более простым алгоритмом:

1. Генерируют матрицу  $V$  размера  $k \times m$  из независимых стандартных нормальных величин
2. Считают матрицу  $\Phi^{[s]}$  по формуле:

$$\Phi^{[s]} = \bar{\Phi} + \text{chol}(\bar{\Omega}) \cdot V \cdot \text{chol}(\Sigma^{[s]})'$$

Преимущества априорного распределения Миннесоты хорошо известны. Прежде всего, оно просто задается. Кроме того, оно успешно применялось в литературе для решения различных задач. И наконец, получившееся апостериорное распределение является нормальным, а значит, легко можно получить значение любой функции параметров с помощью методов Монте-Карло. Однако существенным недостатком этого распределения является то, что оно не предполагает использования байесовской процедуры для оценки  $\Sigma$ .

## 2.5 Сопряженное нормальное-обратное Уишарта априорное распределение

Указанного недостатка априорного распределения Миннесоты можно избежать, если рассматривать сопряженное априорное распределение, т.е. распределение, при котором априорное распределение, функция правдоподобия и апостериорное распределение принадлежат одному классу. Т.к. функция правдоподобия может быть разбита на две части, одна из которых соответствует нормальному распределению (при условии известной ковариационной матрицы остатков), а другая — обратному распределению Уишарта, то и сопряженным априорным распределением для рассматриваемой модели будет также нормальное-обратное Уишарта распределение.

Априорное нормальное-обратное Уишарта распределение может быть записано как:

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \phi | \Sigma \sim \mathcal{N}(\underline{\phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \quad (25)$$

Здесь нужно отметить, что система (25) записана для случая, когда ковариационная матрица параметров имеет кронекерову структуру, что означает, что  $\lambda_{kron}$  полагается равной единице. Хотя строго, говоря, это предположение не обязательно, оно существенно ускоряет расчеты, поэтому, как правило, используется в макроэкономических приложениях.

Гиперпараметры вектора математического ожидания ( $\underline{\phi}$ ) и ковариационной матрицы  $\underline{\Omega}$  условного априорного распределения могут быть заданы точно так же, как и в случае распределения Миннесоты для случая  $\lambda_{kron} = 1$  (см. (20), (22) и (23)).  $\underline{S}$  выбирается так, чтобы среднее  $\Sigma$  совпадало с фиксированной ковариационной матрицей в априорном распределении Миннесоты. Т.к. безусловное распределение параметров имеет вид:

$$\mathbb{E}(\phi) = \underline{\phi} \quad \text{Var}(\phi) = (\underline{\nu} - m - 1)^{-1}(\underline{S} \otimes \underline{\Omega}), \quad (26)$$

то диагональные элементы  $\underline{S}$  выбираются следующим образом:

$$(\underline{S})_{ii} = (\underline{\nu} - m - 1)\hat{\sigma}_i^2 \quad (27)$$

Выбор степеней свободы обратного Уишарта распределения  $\underline{\nu}$  в соответствии с:

$$\underline{\nu} \geq \max\{m + 2, m + 2h - T\} \quad (28)$$

обеспечивает существование как априорной дисперсии параметров, так и апостериорной дисперсии прогнозов на горизонте  $h$  (см. [kadiaa\\_karlsson\\_1997\\_estimation](#)).

Можно показать, что с учетом функции правдоподобия (9) апостериорное распределение принадлежит тому же классу (см, например, [zellner\\_1996\\_introduction](#)):

$$\begin{cases} \Sigma | Y \sim \mathcal{IW}(\bar{S}, \bar{\nu}) \\ \Phi | \Sigma, Y \sim \mathcal{N}(\bar{\Phi}, \Sigma \otimes \bar{\Omega}) \end{cases} \quad (29)$$

где

$$\begin{aligned}
\bar{\nu} &= \underline{\nu} + T \\
\bar{\Omega} &= (\underline{\Omega}^{-1} + X'X)^{-1} \\
\bar{\Phi} &= \bar{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y) \\
\bar{S} &= \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}X'X\hat{\Phi} + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \bar{\Phi}'\bar{\Omega}^{-1}\bar{\Phi} \\
\hat{\Phi} &= (X'X)^{-1}X'Y \\
\hat{E} &= Y - X\hat{\Phi}
\end{aligned}$$

Существует достаточно популярный альтернативный подход для подсчёта гиперпараметров апостериорного распределения.

Мы обнуляем матрицы  $\underline{S}$  и  $\underline{\Omega}^{-1}$ , а чтобы компенсировать разницу добавляем дополнительные наблюдения в матрицу  $X$  и в матрицу  $Y$ :

$$X^* = \begin{bmatrix} X^+ \\ X \end{bmatrix}, Y^* = \begin{bmatrix} Y^+ \\ Y \end{bmatrix}$$

Отметим, что матрицы  $X$  и  $Y$  входят в гиперпараметры апостериорного распределения только в составе матриц  $X'X$ ,  $X'Y$  и  $Y'Y$ , поэтому абсолютно не важно, в каком порядке добавлять искусственные наблюдения и каким образом по отношению к матрицам  $X$  и  $Y$ . Можно их добавить в конец матриц  $X$  и  $Y$ , можно в начало, можно посередине.

Получим новые формулы для апостериорных гиперпараметров: ВЫВЕРИТЬ!

$$\begin{aligned}
\bar{\nu} &= \underline{\nu} + T \\
\bar{\Omega} &= (X^{*'}X^*)^{-1} = (X^{+'}X^+ + X'X)^{-1} \\
\bar{\Phi} &= \bar{\Omega} \cdot (X^{*'}Y^*) = \bar{\Omega} \cdot (X^{+'}Y^+ + X'Y) \\
\bar{S} &= \hat{E}'\hat{E} + \hat{\Phi}X'X\hat{\Phi} - \bar{\Phi}'\bar{\Omega}^{-1}\bar{\Phi} \\
\hat{\Phi} &= (X'X)^{-1}X'Y \\
\hat{E}^* &= Y^* - X^*\hat{\Phi}
\end{aligned}$$

Другими словами, наблюдения добавляются так, что гиперпараметры апостериорных наблюдений не изменяются. Заметим, что  $X^{*'}X^* = X^{+'}X^+ + X'X$ ,  $E^{*'}E^* = E^{+'}E^+ + E'E$ .

Чтобы новые формулы совпадали со старыми необходимо, чтобы

$$\begin{cases} X^{+'}X^+ = \underline{\Omega}^{-1} \\ X^{+'}Y^+ = \underline{\Omega}^{-1}\underline{\Phi} \\ \dots \end{cases} \quad (30)$$

Заметим, что новые формулы позволяют трактовать:

- $\bar{\Phi}$  — как результат построения регрессий  $Y^*$  на  $X^*$ :

$$\bar{\Phi} = (X^{*'} X^*)^{-1} \cdot (X^{*'} Y^*)$$

- $\bar{S}$  — как ...

Эти условия будут выполнены если добавить наблюдения по схеме:

По аналогии с работами **banbura\_al\_2010\_large** **berg\_henzel\_2013\_point** соответствующее априорное распределение вводится путем добавления искусственных наблюдений<sup>5</sup>:

$$Y^{NIW} = \begin{bmatrix} \frac{\text{diag}(\delta_1 \sigma_1, \dots, \delta_m \sigma_m)}{\lambda_{tight}} \\ 0_{m(p-1) \times m} \\ \text{diag}(\sigma_1, \dots, \sigma_m) \\ 0_{1 \times m} \end{bmatrix} \quad X^{NIW} = \begin{bmatrix} \frac{\text{diag}(1, 2^{\lambda_{lag}}, \dots, p^{\lambda_{lag}}) \otimes \text{diag}(\sigma_1, \dots, \sigma_m)}{\lambda_{tight}} & 0_{mp \times 1} \\ 0_{m \times mp} & 0_{m \times 1} \\ 0_{1 \times mp} & \frac{1}{\lambda_{const}} \end{bmatrix} \quad (31)$$

В работах **doan\_litterman\_sims\_1984\_forecasting** and **sims\_1993\_nine** было предложено добавить к этим априорным распределениям дополнительную характеристику, введение которой обуславливается возможным наличием во временных рядах единичных корней и коинтеграционных соотношений. Это позволяет исключить появление неправдоподобно большой доли внутривыборочной дисперсии, объясняемой экзогенными переменными **carriero\_al\_2015\_bayesian**

### Модификации априорного распределения

Априорное распределение суммы коэффициентов<sup>6</sup> было предложено в работе **doan\_litterman\_sims\_1984\_forecasting** Это распределение отражает следующую идею: если переменные в VAR имеют единичный корень, то можно учесть эту информацию, задав априорное распределение, в котором сумма коэффициентов при всех лагах зависимой переменной равна единице (см. **robertson\_tallman\_1999\_vector** и **blake\_mumtaz\_2012\_applied** Другими словами, когда среднее значение лагированных значений какой либо переменной находится на некотором уровне, то это же самое значение, является хорошим прогнозом для будущих значений этой переменной.

Внедрение этого априорного распределения производится путем добавления искусственных дамми-наблюдений по следующей схеме:

$$Y^{SC} = \frac{1}{\lambda_{sc}} [\text{diag}(\delta_1 \mu_1, \dots, \delta_m \mu_m)] \quad (32)$$

$$X^{SC} = \frac{1}{\lambda_{sc}} [(1_{1 \times p}) \otimes \text{diag}(\delta_1 \mu_1, \dots, \delta_m \mu_m) \quad 0_{m \times 1}], \quad (33)$$

<sup>5</sup>Формулы, приведенные в самих работах **banbura\_al\_2010\_large** **berg\_henzel\_2013\_point** и отражающие введение новых наблюдений, являются частным случаем (31) для  $\lambda_{lag} = 1$  и  $\lambda_{const} \rightarrow \infty$ .

<sup>6</sup>sum-of-coefficients prior

где  $(1_{1 \times p})$  - вектор-строка из единиц длиной  $p$ ,  $\mu_i$  есть  $i$ -ая компонента вектора  $\mu$ , который состоит из средних начальных значений всех переменных<sup>7</sup>:  $\mu = \frac{1}{p} \sum_{t=1}^p y_t$

Априорное распределение начального наблюдения<sup>8</sup>, предложенное в работе **sims\_1993\_nine** отражает априорную веру в то, что переменные имеют общий стохастический тренд. Для этого вводится единственное дамми-наблюдение такое, что все значения всех переменных равны соответствующему среднему начальных значений  $\mu_i$  с точностью до коэффициента масштаба:  $\lambda_{io}$ . Это происходит путем добавления в систему дамми-наблюдения следующего вида:

$$Y^{IO} = \frac{1}{\lambda_{io}} [\delta_1 \mu_1, \dots, \delta_m \mu_m] \quad (34)$$

$$X^{IO} = \frac{1}{\lambda_{io}} [(1_{1 \times p}) \otimes (\delta_1 \mu_1, \dots, \delta_m \mu_m) \quad 1], \quad (35)$$

(?) Это априорное распределение приводит к тому, что среднее по каждой переменной есть линейная комбинация всех остальных средних.

Гиперпараметр  $\lambda_{io}$  отражает жесткость указанного априорного распределения. Когда  $\lambda_{io} \rightarrow 0$ , модель принимает вид, в котором либо все переменные стационарны со средним, равным выборочному среднему начальных условий, либо нестационарны (без дрейфа) и коинтегрированы.

Как и в случае распределения Миннесоты, необходимости использовать алгоритм Гиббса нет, можно генерировать случайную выборку непосредственно из апостериорного распределения. Например, можно применять такой алгоритм:

1. На  $s$ -ом шаге сгенерировать очередную итерацию согласно:

$$\begin{aligned} \Sigma^{[s]} &\sim \mathcal{IW}(\bar{S}, \bar{\nu}) \\ \phi^{[s]} &\sim \mathcal{N}(\bar{\phi}; \Sigma^{[s]} \otimes \bar{\Omega}) \end{aligned}$$

2. Увеличить  $j$  на единицу и перейти к пункту один

На практике вместо генерирования вектора  $\phi^{[s]}$  генерируют сразу матрицу  $\Phi^{[s]}$  в два шага:

1. Генерируют матрицу  $V$  размера  $k \times m$  из независимых стандартных нормальных величин
2. Считают матрицу  $\Phi^{[s]}$  по формуле:

$$\Phi^{[s]} = \bar{\Phi} + \text{chol}(\bar{\Omega}) \cdot V \cdot \text{chol}(\Sigma^{[s]})'$$

<sup>7</sup>Некоторые авторы для расчета  $mu$  рассчитывают среднее по всем наблюдениям, т. е.  $\mu = \frac{1}{T} \sum_{t=1}^T y_t$  (см. **banbura\_al\_2010\_large** и **carriero\_al\_2015\_bayesian**). Однако, в соответствии с работой **sims\_zha\_1998\_bayesian** для расчета среднего используются только первые  $p$  наблюдений.

<sup>8</sup>dummy initial observation prior

Два ключевых недостатка априорного распределения Миннесоты (отсутствие байесовского оценивания  $\Sigma$  обязательная апостериорная независимость отдельных уравнений системы) решаются при использовании сопряженного нормального - обратного Уишарта распределения. Однако в этом случае параметры априорного распределения для разных уравнений выбираются симметрично. В частности, все коэффициенты при первом лаге зависимой переменной априорно имеют одну и ту же дисперсию  $\lambda_{tight}^2$ . Хотя обычно эта предпосылка не является слишком ограничивающей, в реальности легко встретиться с задачами, в которых ковариационная матрица априорного распределения не должна быть симметрично сформирована для разных уравнений. Например, довольно известным в литературе является следующий пример (см, **kadiyala\_karlsson\_1997\_numerical**). Положим, исследователь хочет учесть в VAR наличие нейтральности денег. При построении модели эта предпосылка может быть учтена наложением такого априорного распределения, в котором все коэффициенты при лагах денег в уравнении для выпуска имеют нулевое матожидание и низкую дисперсию. Однако это означает, что и в других уравнениях дисперсия коэффициентов в априорном распределении будет относительно низкой. Это характеристика может быть нежелательной, и, чтобы этого избежать, априорное распределение можно задать как независимое - обратное Уишарта.

## 2.6 Независимое нормальное-обратное Уишарта распределение

Этот тип распределений предполагает, что ковариационная матрица параметров может быть произвольной формы:

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \underline{\nu}) \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (36)$$

В этом случае можно показать (ссылка karlsson? - но тут нет доказательства), что условные апостериорные распределения имеют вид:

$$\begin{cases} \phi | \Sigma, Y \sim \mathcal{N}(\bar{\phi}; \bar{\Xi}) \\ \Sigma | \phi, Y \sim \mathcal{IW}(\bar{S}; \bar{\nu}) \end{cases} \quad (37)$$

где<sup>9</sup>

$$\begin{aligned} \bar{\nu} &= \underline{\nu} + T \\ \bar{S} &= \underline{S} + E' E, \text{ где } E = Y - X\Phi \\ \bar{\Xi} &= (\underline{\Xi}^{-1} + \Sigma^{-1} \otimes X' X)^{-1} \\ \bar{\phi} &= \bar{\Xi} \cdot (\underline{\Xi}^{-1} \underline{\phi} + \text{vec}(X' Y \Sigma^{-1})) \end{aligned}$$

Гиперпараметры априорного распределения могут быть выбраны точно так же, как и в априорном распределении Миннесоты (см. (20) и (21)). При необходимости

---

<sup>9</sup> Другой вариант записи для  $\bar{\phi}$  имеет вид:  $\bar{\phi} = \bar{\Xi} \cdot (\underline{\Xi}^{-1} \underline{\phi} + (\Sigma^{-1} \otimes (X' X)) \text{vec } \hat{\Phi})$



неинформативное априорное распределение для коэффициентов при детерминированных переменных переменных можно задать, обнулив соответствующее значение в матрице  $\Xi^{-1}$ . Использование произвольной ковариационной матрицы приводит к тому, что исследователю оказываются известны только условные апостериорные распределения для  $\phi$  и  $\Sigma$ . Это обуславливает необходимость использования алгоритма Гиббса для получения реализаций из совместного апостериорного распределения.

Получить Марковскую цепь, сходящуюся к апостериорному распределению можно, например, так:

1. Сгенерировать произвольно стартовую матрицу  $\Sigma^{[0]}$ , например, единичную
2. На  $s$ -ом шаге сгенерировать очередную итерацию согласно:

$$\phi^{[s]} \sim \mathcal{N}(\bar{\phi}^{[s-1]}; \bar{\Xi}^{[s-1]}), \text{ где } \bar{\phi}^{[s-1]} \text{ и } \bar{\Xi}^{[s-1]} \text{ рассчитываются через } \Sigma^{[s-1]} \quad (38)$$

$$\Sigma^{[s]} \sim \mathcal{IW}(\bar{S}^{[s]}; \bar{\nu}), \text{ где } \bar{S}^{[s]} \text{ рассчитываются через } \phi^{[s]} \quad (39)$$

3. Увеличить  $s$  на единицу и перейти к пункту два

## 2.7 Соответствие гиперпараметров в разных работах

DM16	CCM15	BGR10, BH13	KK97	Формула?
$\lambda_{tight}$	$\lambda_1$	$\lambda$	$\sqrt{\pi_1}$	
$\lambda_{kron}$	$\lambda_2 = 1$	$\vartheta = 1$	$\sqrt{\frac{\pi_2}{\pi_1}}$	
$\lambda_{lag}$	1	1	0.5	
$\lambda_{const}$	$\lambda_0$	$\infty$	$\sqrt{\pi_3}$	
$\lambda_{exo}$	NA	NA	NA	
$\lambda_{sc}$	$\lambda_3$	$\tau$		
$\lambda_{io}$	$\lambda_4$	NA		

## 2.8 Выбор гиперпараметров и числа лагов

Как было показано в работе **demol\_al\_2008\_forecasting** и подтверждено в других более поздних работах, использование сравнительно большого количества временных рядов требует уменьшения параметра  $\lambda_1$  с увеличением размерности выборки, что означает наложение более жесткого априорного распределения. На данный момент в литературе используется два подхода к определению оптимальной величины  $\lambda_1$ . В своей работе мы используем оба и сравниваем качество прогноза.

Первый алгоритм был предложен в работе **banbura\_al\_2010\_large** и он основан на идее о том, что регуляризация должна быть настолько жесткой, чтобы не исключить возможность избыточной параметризации модели, при этом предполагается, что трехмерная VAR - достаточно простая (parsimonious) модель, не содержащая слишком большого количества параметров. Процедура выбора  $\lambda$  состоит в том,

что что средний внутривыборочный прогноз для реального ВВП и индекса цен тот же самый, как на первой выборке (на которой происходит) оценивание. Т.е. каждая модель регуляризуется до размера простой VAR. При этом референтной моделью является та, для которой апостериорное распределение не зависит от функции правдоподобия, т. е. для которой  $\lambda = 0$ . Это означает, что дисперсии всех параметров  $\phi$  равны нулю, т. е. переменные описываются моделью случайного блуждания (RW) со смещением,  $y_{i,t} = c + y_{i,t-1} + \varepsilon_t, i = 1, \dots, m$ . Обозначим эту модель индексом 0, т.к.  $\lambda = 0$ . Схема выбора  $\lambda$  состоит из следующих этапов:

1. На первом этапе строятся внутривыборочные однопериодные прогнозы на обучающей выборке и рассчитывается среднеквадратичная ошибка прогноза выпуска ( $MSFE_y^0$ ) и инфляции ( $MSFE_\pi^0$ ).
2. Оценивается трехмерная VAR для  $\lambda \rightarrow \infty$ <sup>10</sup> и рассчитываются среднеквадратичная ошибка прогноза выпуска ( $MSFE_y^\infty$ ) и инфляции ( $MSFE_\pi^\infty$ ) и показатель  $FIT^\infty$ :

$$FIT^\infty = \frac{1}{2} \cdot \frac{MSFE_y^\infty}{MSFE_y^0} + \frac{1}{2} \cdot \frac{MSFE_\pi^\infty}{MSFE_\pi^0} \quad (40)$$

3. Оцениваются BVAR модели для  $m$  переменных и для большого числа различных  $\lambda$  рассчитываются среднеквадратичные ошибки прогноза для выпуска ( $MSFE_y^{\lambda,m}$ ) и инфляции ( $MSFE_\pi^{\lambda,m}$ ) и показатель  $FIT^{\lambda,m}$ :

$$FIT^{\lambda,m} = \frac{1}{2} \cdot \frac{MSFE_y^{\lambda,m}}{MSFE_y^0} + \frac{1}{2} \cdot \frac{MSFE_\pi^{\lambda,m}}{MSFE_\pi^0} \quad (41)$$

4. Оптимальное  $\lambda$  рассчитывается как значение, при котором минимизируется отклонение  $FIT^{\lambda,m}$  от  $FIT^\infty$ :

$$\lambda^* = \arg \min |FIT^{\lambda,m} - FIT^\infty| \quad (42)$$

После того как выбрано оптимальное  $\lambda$  для каждой модели, происходит построение вневыборочных прогнозов на оценивающей выборке.

Второй алгоритм предложен в работе (Carriero et al, 2012) и представляет собой выбор такого параметра  $\lambda_1$ , который бы максимизировал функцию предельной плотности:

$$\lambda_{1t}^* = \arg \max_{\lambda_1} \ln p(Y) \quad (43)$$

---

<sup>10</sup>При  $\lambda \rightarrow \infty$  оценки BVAR совпадают с оценками VAR методами OLS или ML, т. к. апостериорное распределение параметров в этом случае совпадает с функцией правдоподобия. Считается, что трехмерная VAR содержит достаточно маленькое число параметров, и байесовская регуляризация не требуется.

При этом функция предельной плотности может быть получена путем интегрирования коэффициентов модели: <sup>11</sup>

$$p(Y) = \int p(Y|\phi)p(\phi)d\phi \quad (44)$$

Если априорное распределение является нормальным - обратным Уишарта, то предельная плотность  $p(Y)$  может быть посчитана аналитически (zellner\_1996\_introduction Bauwens et al,1999; Carriero et al,2012):

$$p(Y) = \pi^{-\frac{Tm}{2}} \times + |(I + X\Omega X')^{-1}|^{\frac{N}{2}} \times |\underline{S}|^{\frac{\nu}{2}} \times \frac{\Gamma_N(\frac{\nu+T}{2})}{\Gamma_N(\frac{\nu}{2})} \times \\ \times |\underline{S} + (Y - X\underline{\Phi})'(I + X\Omega X')^{-1}(Y - X\underline{\Phi})|^{-\frac{\nu+T}{2}}, \quad (45)$$

где  $\Gamma_N(\cdot)$  обозначает  $N$ -мерную гамма функцию. Выбор числа лагов происходит аналогично путем максимизации по  $p$  функции предельной плотности (45):

$$p^* = \arg \max_p \ln p(Y) \quad (46)$$

## 2.9 Особенности кодирования

При практической реализации алгоритма Гиббса или Монте-Карло часто приходится обращаться положительно определённые симметричные матрицы. Для этого можно использовать следующий способ:

1. Получить разложение Холецкого для заданной матрицы  $A$

$$A = U'U,$$

где  $U$  — верхнетреугольная матрица

2. Обратить матрицу  $U$ . Существуют специальные алгоритмы обращения верхнетреугольных матриц.
3. Получить  $A^{-1}$  по формуле

$$A^{-1} = U^{-1}U^{-1'}$$

Однако даже данный способ сопряжён с численными трудностями, если обрабатываемая матрица плохо обусловлена. Например, при большом количестве эндогенных переменных и большом количестве лагов матрица  $X'X$  является плохо обусловленной и для неё может быть численно трудно получить разложение Холецкого. В таком случае мы использовали псевдообратную матрицу Мура-Пенроуза.

---

<sup>11</sup>Т.к. интегрирование происходит по всем коэффициентам, но не по гиперпараметрам априорного распределения  $(\lambda_1, \lambda_2, \lambda_3)$  и не по числу лагов  $p$ , то предельная плотность является функцией  $\lambda_j, j = 1 \dots 3$  и  $p$ .

### 3 Данные

Для расчетов мы используем 24 временных ряда с января 1995г. по апрель 2015г. Границы выборки обусловлены доступностью данных, исходная выборка содержит 244 наблюдения. Полный список взятых временных рядов указан в Приложении 2. После устранения сезонности в рядах, демонстрирующих сезонные колебания, мы логарифмируем всех ряды кроме процентной ставки. Далее происходит проверка на стационарность, для чего используются ADF и KPSS тесты. Такая проверка необходима, для того чтобы определить матожидание априорного распределения для параметров  $\Phi_1$ . Следуя методологии других работ, посвященных прогнозированию с помощью BVAR (например ()), мы назначаем  $(\Phi_1)_{ii} = 1$  для нестационарных рядов и  $(\Phi_1)_{ii} = 0$  для стационарных. На втором этапе мы оцениваем три rolling VAR модели для разного набора переменных и строим по ним прогнозы. Период оценивания составляет всегда 120 месяцев, Прогноз строится на 1, 3, 6 и 12 месяцев. Мы строим VAR для разного количества переменных: для 3, 6 и 24<sup>12</sup>. Модель с тремя переменными рассматривается как наиболее простая модель, содержащая только самые важные переменные: показатель деловой активности (индекс промышленного производства), индекс цен (подсчитанной с помощью ИПЦ) и инструмента монетарной политики (в качестве прокси для которого мы берем процентную ставку межбанковского рынка). Модель с шестью переменными специфицируется по аналогии со многими монетарными моделями, использовавшимися для структурного анализа различных экономик (Sims, 1992; Kim and Roubini (2000); Bjornland (2008); Uhlig and Scholl (2008) и включает в дополнение к уже указанным переменным валютный курс, денежный агрегат M2 и цены на нефть (последний показатель включен для отражения экспортоориентированности российской экономики). В модель с 24 переменными строится с использованием всех доступных временных рядов. При определении параметра жесткости распределения по методу Banbura et al.() необходимо выделить период, на котором происходит определение  $\lambda$ . Мы определяем  $\lambda$  на самом раннем доступном промежутке: с января 1996г. по декабрь 2005г (первые 12 месяцев используются в качестве лаговых значений переменных для этой регрессии). Количество лагов определяется путем минимизации информационных критериев для обычной VAR на той же самой подвыборке. (что делать с var по 24 рядам??) Далее количество лагов и параметр жесткости фиксируются и используются для построения прогнозов на всех остальных подвыборках. При определении параметра жесткости путем максимизации предельной плотности данных (marginal data density) при оценке каждой BVAR происходит совместный поиск на сетке по  $\lambda$  и по  $p$ , и выбираются такие значения, для которых (45) максимально.

---

<sup>12</sup>Обычная VAR для сравнения качества прогноза строится для 3 и 6 переменных

## Список литературы

Litterman, Robert B (1986). “Forecasting with Bayesian vector autoregressions—five years of experience”. В: *Journal of Business & Economic Statistics* 4.1, с. 25—38.

## Приложения

### Данные

22

Название временного ряда	Тип данных	База (если есть)	Источник
Индекс промышленного производства	Базисный индекс	2010	IFS
Индекс потребительских цен	Базисный индекс	2010	IFS
Индекс занятости в промышленности	Базисный индекс	2010	IFS
Процентная ставка межбанковского рынка	В процентах годовых		IFS
Процентная ставка по кредитам	В процентах годовых		IFS
Индекс реальных денежных доходов	Базисный индекс	январь 1992	ФСГС
Уровень безработицы	В процентах		IFS
Индекс цен на нефть марки Brent	Базисный индекс	2010	IFS
Индекс цен производителей	Цепной индекс		IFS
Ввод в действие новых жилых домов	В тыс.кв.м.		ФСГС
Индекс реальных инвестиций в основной капитал	Базисный индекс	январь 1994	ЦАД
Индекс реальных зарплат	Базисный индекс	январь 1993	ФСГС
Денежный агрегат М2	в млрд. руб.		ЦБ
Реальный эффективный валютный курс	Базисный индекс	2010	IFS
Цена натурального газа	Долл. за млн. БТЕ	2010	IFS
Международные резервы за исключением золота	Млрд. долл.		IFS
Номинальный валютный курс	руб. за долл.		IFS
Заявленная потребность в работниках	Тыс. чел.		ЦАД
Индекс реального объема сельхозпроизводства	Базисный индекс	январь 1993	ЦАД
Индекс реального объема розничной торговли	Базисный индекс	январь 1994	ЦАД
Сальдо консолидированного бюджета			ЦАД
Экспорт товаров	млн. долл.		IFS
Импорт товаров	млн. долл.		IFS

## Доступные реализации кода

Источник	Min	Conj N-IW	Ind N-IW	SoC	IO	Comment
Коды в среде Матлаб						
Carriero <sup>13</sup>	-	?	-	+	+	Дамми-наблюдения вводятся как для conj NIW, при этом делает Gibbs sampling, при этом $\bar{\Phi}$ одно и то же, а $\Sigma$ пересчитывается на каждом шаге в зависимости от предыдущего $\Phi$
Blake Mumtaz <sup>14</sup>	-	+	+	+	+	Называет Миннесотой Ind NIW. Код для conj NIW построен так же, как у Carriero через Gibbs Sampling: разница - добавляет 2 един. строки для дамми при определении коэффициента при константе

<sup>13</sup><http://cremfi.econ.qmul.ac.uk/efp/info.php>

<sup>14</sup>[http://www.bankofengland.co.uk/education/Pages/ccbs/technical\\_handbooks/techbook4.aspx](http://www.bankofengland.co.uk/education/Pages/ccbs/technical_handbooks/techbook4.aspx)