

BVAR in Russia

я и Боря

10 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	$x_t = [y'_{t-1} \dots y'_{t-p} z'_t]'$
ε_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	$E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$
Φ_1, \dots	$m \times m$	коэффициенты VAR	$y_t = \Phi_1 y_{t-1} + \dots + \Phi_{ex} z_t$
Φ_{ex}	$m \times d$	коэффициенты при экзогенных переменных	
Φ	$k \times m$	упаковка матриц Φ_1, \dots	$\Phi = [\Phi_1 \dots \Phi_p \Phi_{ex}]'$
ϕ	$km \times 1$	вектор из матрицы Φ	$\text{vec } \Phi$
Ξ	$km \times km$	Априорная ковариационная матрица Φ	
$\underline{\Phi}$	$k \times m$	априорное математическое ожидание Φ	
$\underline{\Omega}$	$k \times k$	Матрица априорных масштабирующих коэффициентов ковариационной матрицы Φ	
Σ	$m \times m$	Ковариационная матрица ошибок	$\mathbb{E} \varepsilon_t \varepsilon_t' = \Sigma$
$\underline{\nu}$	скаляр		

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:

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

где $\Phi_{ex} = (c_1, \dots, c_m)'$ — вектор констант размерности m , Φ_l — авторегрессионные матрицы размерности $m \times m$ при $l = 1, \dots, p$. Вектор ε_t — m -мерный вектор ошибок с ковариационной матрицей $\mathbb{E} \varepsilon_t \varepsilon_t' = \Sigma$, некоррелированный с объясняющими пере-

менными. 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:

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

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:

$$p(\Phi, \Sigma|Y) = \frac{p(\Phi, \Sigma)L(Y|\Phi, \Sigma)}{p(Y)} \text{Т.к. } p(Y) \text{ не зависит от } \Phi \text{ и } \Sigma, \text{ то можно записать:} \quad (4)$$

$$p(\Phi, \Sigma|Y) \propto p(\Phi, \Sigma)L(Y|\Phi, \Sigma) \quad (5)$$

Так как $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, то функция правдоподобия задается как:¹

$$L(Y|\Phi, \Sigma) \propto |\Sigma|^{-T/2} \text{etr} \left\{ -\frac{1}{2} [\Sigma^{-1}(Y - X\Phi)'(Y - X\Phi)] \right\} \quad (6)$$

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

$$\begin{cases} \phi \sim \mathcal{N}(\phi; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \nu) \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (7)$$

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

Частными случаями независимого нормального-обратного Уишарта являются:

¹ 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 $\text{etr}() = \exp(\text{tr}())$.

(a) Распределение Миннесоты

$$\begin{cases} \phi \sim \mathcal{N}(\phi; \Xi) \\ \Sigma = \text{const} \end{cases} \quad (8)$$

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 Ξ имеет структуру кронекерова произведения $\Xi = \Sigma \otimes \underline{\Omega}$. В этом случае распределение Миннесоты становится частным случаем сопряжённого нормального-обратного Уишарта.

(b) Нормальное-Джеффри

$$\begin{cases} \phi \sim \mathcal{N}(\phi; \Xi) \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (9)$$

Получается из независимого нормального-обратного Уишарта при $\underline{\nu} \rightarrow 0$ и $\underline{S} \rightarrow 0$???????? http://www.tc.umn.edu/~nydic001/docs/unpubs/Wishart_Distribution.pdf

Для получения выборки из апостериорного распределения можно использовать схему Гиббса.

Распределение Миннесоты и нормальное-Джеффри являются противоположными крайностями независимого-обратного Уишарта. В распределении Миннесоты матрица Σ предполагается известной, а в нормальном-Джеффри матрица Σ имеет «размытое» неинформативное распределение.

Частным случаем нормального-Джеффри распределения является:

i. Неинформативное-Джеффри

$$\begin{cases} \phi \sim 1 \\ \Sigma \sim |\Sigma|^{-(m+1)/2} \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (10)$$

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

2. Сопряженное нормальное-обратное Уишарта распределение

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \phi | \Sigma \sim \mathcal{N}(\underline{\phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \quad (11)$$

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:

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

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».

В нашей работе мы всегда явно указываем, идёт ли речь о независимом или сопряжённом 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 Независимое нормальное-обратное Уишарта сопряжение

Априорно предполагается, что:

$$\begin{cases} \phi \sim \mathcal{N}(\underline{\phi}; \Xi) \\ \Sigma \sim \mathcal{IW}(\underline{S}; \underline{\nu}) \\ \phi \text{ и } \Sigma \text{ независимы} \end{cases} \quad (12)$$

В этом случае можно показать (ссылка), что условные апостериорные распределения имеют вид:

$$\begin{cases} \phi | \Sigma, Y \sim \mathcal{N}(\bar{\phi}; \bar{\Xi}) \\ \Sigma | \phi, Y \sim \mathcal{IW}(\bar{S}; \bar{\nu}) \end{cases} \quad (13)$$

где

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

Существует достаточно популярный альтернативный подход для подсчёта гиперпараметров апостериорного распределения.

Мы обнуляем матрицы \underline{S} и Ξ^{-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. Можно их добавить в конец матриц X и Y , можно в начало, можно в крапинку посередине.

Получим новые формулы для апостериорных гиперпараметров:

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

Фокус заключается в том, что наблюдения добавляются так, что гиперпараметры апостериорных наблюдений не изменяются. Заметим, что $X^{*'}X^* = X^{+'}X^+ + X'X$, $E^{*'}E^* = E^{+'}E^+ + E'E$.

Значит, чтобы новые формулы совпали со старыми необходимо, чтобы:

$$\left\{ E^{+'}E^+ = \underline{S} \right. \quad (14)$$

Гиперпараметры априорного распределения традиционно выбираются следующим образом:

....

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

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

$$\phi^{[j]} \sim \mathcal{N}(\bar{\phi}^{[j-1]}; \bar{\Xi}^{[j-1]}), \text{ где } \bar{\phi}^{[j-1]} \text{ и } \bar{\Xi}^{[j-1]} \text{ рассчитываются через } \Sigma^{[j-1]} \quad (15)$$

$$\Sigma^{[j]} \sim \mathcal{IW}(\bar{S}^{[j]}; \bar{\nu}), \text{ где } \bar{S}^{[j]} \text{ рассчитываются через } \phi^{[j]} \quad (16)$$

3. Увеличить j на единицу и перейти к пункту два

2.5 Априорное распределение Миннесоты

Хотя априорное распределение Миннесоты и является частным случаем независимого нормального-обратного Уишарта распределения, имеет смысл рассмотреть его отдельно. Оно было предложено в работе [litterman_1986_forecasting](#) и [doan_litterman_sims_1984_forecasting](#)

Изначально [litterman_1986_forecasting](#) предложил использовать априорное распределение Миннесоты, при котором в явном виде учитывается предпосылка о нестационарности большинства макрорядов.

Априорное распределение параметров предполагается многомерным нормальным, зависящим от нескольких гиперпараметров. Параметры предполагаются независимыми, следовательно, их ковариационная матрица диагональна. Ковариационная матрица вектора ε_t также предполагается диагональной. Запишем задачу (3) в векторизованном виде²:

$$\begin{aligned} Y &= X\Phi + E \\ \text{vec}(Y) &= \text{vec}(X\Phi I) + \text{vec}(E) \Leftrightarrow \\ y &= (I_M \otimes X)\phi + \varepsilon \end{aligned}$$

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

Тогда:

$$\phi \sim \mathcal{N}(\underline{\phi}, \Xi) \quad (17)$$

Априорная плотность распределения ϕ может быть записана как:

$$p(\phi) = \frac{1}{(2\pi)^{km/2} |\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} (\phi - \underline{\phi})' \Xi^{-1} (\phi - \underline{\phi}) \right\}. \quad (18)$$

Комбинируя её с функцией правдоподобия (6), получаем, что апостериорное распределение параметров задаются в следующем виде:

$$\phi|Y \sim \mathcal{N}(\bar{\phi}, \bar{\Xi}) \quad (19)$$

$$\text{где } \bar{\Xi} = [\Xi^{-1} + (\Sigma^{-1} \otimes (X'X))]^{-1} \quad (20)$$

$$\text{и } \bar{\phi} = \bar{\Xi}[\Xi^{-1}\underline{\phi} + (\Sigma^{-1} \otimes X)'y]. \quad (21)$$

На практике в качестве матрицы Σ 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 .

Априорное распределение Миннесоты было задумано таким образом, чтобы учесть нестационарность многих макроэкономических временных рядов. В этом априорном распределении предполагается, что диагональные элементы матрицы первого лага Φ_1 имеют матожидание единица, а остальные элементы матрицы первого лага и все элементы остальных матриц равны нулю, т.е.:

²Третье уравнение системы следует из тождества: $\text{vec}(ABC) = (C \otimes A) \text{vec}(B)$

$$\underline{\Phi} = \mathbb{E}[(\Phi_l)_{ij}|\Sigma] = \begin{cases} 1, & i = j, l = 1; \\ 0, & \text{в остальных случаях} \end{cases} \quad (22)$$

Однако это априорное распределение может быть обобщено до случая, когда все диагональные элементы матрицы первого лага имеют матожидание ϕ_{ii} , где ϕ_{ii} принимает значение единица для нестационарных рядов и меньше единицы для стационарных (рядов с высокой степенью персистентности).

Априорное распределение Миннесоты предполагает, что априорная ковариационная матрица Ξ диагональна. Пусть Ξ_i обозначает блок Ξ , размера $k \times k$, связанный с коэффициентами уравнения i . Тогда диагональные элементы Ξ_i определяются по формулам:

$$\text{Var}((\Phi_l)_{ij}) = \begin{cases} \frac{\lambda_1^2}{l^2}, & \text{для переменной } i \text{ и лага } l, \\ \frac{\lambda_1^2 \lambda_2^2 \sigma_i^2}{l^2 \sigma_j^2}, & \text{для переменной } j \neq i \text{ и лага } l, \\ \lambda_3^2 \sigma_i^2, & \text{для коэффициентов при экзогенных переменных} \end{cases} \quad (23)$$

Как можно видеть из приведенной выше формулы (23) априорная дисперсия параметров зависит от нескольких гиперпараметров, задаваемых исследователем. Гиперпараметры имеют следующую интерпретацию: λ_1 (параметр регуляризации) отражает общую «жесткость» априорного распределения. Если $\lambda_1 \rightarrow 0$, то априорное распределение полностью определяет апостериорное распределение, и данные не играют никакой роли при оценке параметров. Наоборот, если $\lambda_1 \rightarrow \infty$, то априорное распределение перестает влиять и оценка параметров сходится к обычной оценке МНК. Параметр λ_2 (параметр кросс-регуляризации) добавляет дополнительную жесткость лагам других переменных по сравнению с лагами зависимой переменной. Если $\lambda_2 < 1$, то собственные лаги зависимой переменной помогают предсказывать значение переменной лучше, чем лаги других переменных. Параметр λ_3 отражает относительную жесткость распределения константы.

При $\lambda_2 = 1$ матрица Ξ 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 Σ для i -го уравнения означает домножение дисперсий, указанных в матрице $\underline{\Omega}$, на коэффициент σ_i^2 . Начало диагонали матрицы $\underline{\Omega}$ соответствует коэффициентам для лага равного 1, затем коэффициентам для лага равного 2 и т.д. Конец диагонали матрицы $\underline{\Omega}$ соответствует экзогенным переменным. Таким образом диагональ матрицы $\underline{\Omega}$ состоит из вектора $\left(\frac{\lambda_1^2}{l^2 \sigma_1^2}, \frac{\lambda_1^2}{l^2 \sigma_2^2}, \dots, \frac{\lambda_1^2}{l^2 \sigma_m^2} \right)$, повторенного p раз, и в конце стоит число λ_3^2 .

Преимущества априорного распределения Миннесоты хорошо известны. Во-первых, оно просто задается, во-вторых, оно успешно применялось в литературе для решения различных задач. В-третьих, получившееся апостериорное распределение является

нормальным, и 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 Σ .

Необходимости применять алгоритм Гиббса для априорного распределения Миннесоты нет. Алгоритм генерации случайной выборки непосредственно из апостериорного распределения прост:

1. На j -ом шаге сгенерировать очередную итерацию согласно:

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

2. Увеличить j на единицу и перейти к пункту один

2.6 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:

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

Можно показать, что с учетом функции правдоподобия (6) апостериорное распределение принадлежит к тому же классу (см, например, [zellner_1996_introduction](#)):

$$\begin{cases} \Sigma | Y \sim \mathcal{IW}(\bar{S}, \bar{\nu}) \\ \Phi | \Sigma, Y \sim \mathcal{N}(\bar{\Phi}, \Sigma \otimes \bar{\Omega}) \end{cases} \quad (26)$$

где

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

Как и ранее, гиперпараметры апостериорного распределения можно получить занулив матрицы $\underline{\Omega}^{-1}$ и \underline{S} , но добавив дополнительные наблюдения X^+ и Y^+ в матрицы X и Y . При этом формулы для апостериорных гиперпараметров принимают вид:

ВЫВЕРИТЬ!

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

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 ...

Чтобы новые формулы совпадали со старыми необходимо, чтобы

$$\begin{cases} X^{+'} X^+ = \underline{\Omega}^{-1} \\ X^{+'} Y^+ = \underline{\Omega}^{-1} \underline{\Phi} \end{cases} \quad (27)$$

Эти условия будут выполнены если добавить наблюдения по схеме:

В работах **banbura_al_2010_large berg_henzel_2013_point**) соответствующее априорное распределение вводится путем добавления искусственных наблюдений:

$$y^+ = \begin{bmatrix} \text{diag}(\delta_1 \sigma_1, \dots, \delta_n \sigma_n) \\ 0_{n(p-1) \times n} \\ \text{diag}(\sigma_1, \dots, \sigma_n) \\ 0_{1 \times n} \end{bmatrix} \quad x^+ = \begin{bmatrix} \text{diag}(1, 2, \dots, p) \otimes \text{diag}(\sigma_1, \dots, \sigma_n) / \lambda & 0_{np \times 1} \\ 0_{n \times np} & 0_{n \times 1} \\ 0_{1 \times np} & \epsilon \end{bmatrix} \quad (28)$$

При этом математические ожидания и дисперсии априорного распределения параметров могут быть заданы по тому же принципу, что и в априорном распределении Миннесоты (см. (22)-(23)).

...

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**

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 j -ом шаге сгенерировать очередную итерацию согласно:

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

2. Увеличить j на единицу и перейти к пункту один

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

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

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

2.7 Соответствие гиперпараметров в разных работах

A	CCM15	BGR10, BH13	Формула?
λ_{tight}	λ_1	λ	
λ_{kron}	$\lambda_2 = 1$	$\vartheta = 1$	
λ_{power}	1	1	
λ_{const}	λ_0	∞	
λ_{exo}	NA	NA	
λ_{sc}	λ_3	τ	
λ_{io}	λ_4	NA	

Доступные реализации кода: (или в предыдущую): список реализаций и что какое распределение они реализуют

Модификация априорных распределений

Априорное распределение суммы коэффициентов³ было предложено в работе **doan_litterman_sims_1984_forecasting** Это распределение отражает следующую идею: когда среднее значение лагированных значений какой либо переменной находится на некотором уровне \bar{y}_{0i} , то это же самое значение \bar{y}_{0i} , является хорошим прогнозом для будущих значений этой переменной. В качестве \bar{y}_{0i} мы используем среднее значение переменной \bar{y}_i по первым p наблюдениям. Внедрение этого априорного распределения производится путем добавления искусственных дамми-наблюдений по следующей схеме:

$$y_d(i, j) = \begin{cases} \bar{y}_{0i}/\lambda_3, & \text{если } i = j \\ 0 & \text{в обратном случае} \end{cases} \quad x_d(i, s) = \begin{cases} \bar{y}_{0i}/\lambda_3, & \text{если } i = j, s < km \\ 0 & \text{в обратном случае,} \end{cases}$$

где $i, j = 1, \dots, m$, $s = 1, \dots, km$. Когда $\lambda_3 \rightarrow 0$, модель стремится к виду, предполагающему запись в разностях, т.е. единичных корней становится столько же, сколько переменных, и нет коинтеграции.

Другими словами/ другой вариант введения (оставить нужно тот, который будет понятнее):

$$y^{++} = [\text{diag}(\mu_1, \dots, \mu_n)/\tau] \quad x^{++} = [(1, 2 \dots p) \otimes \text{diag}(\mu_1, \dots, \mu_n)/\tau \quad 0_{n \times 1}] \quad (29)$$

Априорное распределение изначального наблюдения⁴, предложенное в работе **sims_1993_nine** означает, что исследователь вводит единственное дамми-наблюдение, такое, что все значения всех переменных равны соответствующему среднему начальных условий с точностью до коэффициента масштаба $1/\lambda_4$. Это происходит путем добавления в систему дамми-наблюдений следующего вида:

$$y_d(j) = \begin{cases} \bar{y}_{0i}/\lambda_4 & \bar{y}_{0i}/\lambda_4 \text{ если } s < km \\ 1/\lambda_4 & \text{в обратном случае,} \end{cases}$$

где $j = 1, \dots, m$, $s = 1, \dots, km$. Когда $\lambda_4 \rightarrow 0$ модель принимает вид, в котором либо все переменные стационарны со средним, равным выборочному среднему начальных условий, либо существуют коинтегрированные ряды с единичным корнем но без дрейфа.

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 λ_1 с увеличением размерности выборки, что означает наложение более жесткого априорного распределения. На

³sum-of-coefficients prior

⁴dummy initial observation prior

данный момент в литературе используется два подхода к определению оптимальной величины λ_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 λ состоит в том, 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 ϕ 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 λ 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 $\lambda \rightarrow \infty$ ⁵ и рассчитываются среднеквадратичная ошибка прогноза выпуска ($MSFE_y^\infty$) и инфляции ($MSFE_\pi^\infty$) и показатель FIT^∞ :

$$FIT^\infty = \frac{1}{2} \cdot \frac{MSFE_y^\infty}{MSFE_y^0} + \frac{1}{2} \cdot \frac{MSFE_\pi^\infty}{MSFE_\pi^0} \quad (30)$$

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

$$FIT^{\lambda,m} = \frac{1}{2} \cdot \frac{MSFE_y^{\lambda,m}}{MSFE_y^0} + \frac{1}{2} \cdot \frac{MSFE_\pi^{\lambda,m}}{MSFE_\pi^0} \quad (31)$$

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

$$\lambda^* = \arg \min |FIT^{\lambda,m} - FIT^\infty| \quad (32)$$

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

⁵При $\lambda \rightarrow \infty$ 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.

Второй алгоритм предложен в работе (Carriero et al, 2012) и представляет собой выбор такого параметра λ_1 , который бы максимизировал функцию предельной плотности:

$$\lambda_{1t}^* = \arg \max_{\lambda_1} \ln p(Y) \quad (33)$$

При этом функция предельной плотности может быть получена путем интегрирования коэффициентов модели: ⁶

$$p(Y) = \int p(Y|\phi)p(\phi)d\phi \quad (34)$$

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):

$$p(Y) = \pi^{-\frac{Tm}{2}} \times \left| (I + X\underline{\Omega}X')^{-1} \right|^{\frac{N}{2}} \times |\underline{S}|^{\frac{\nu}{2}} \times \frac{\Gamma_N(\frac{\nu+T}{2})}{\Gamma_N(\frac{\nu}{2})} \times \left| \underline{S} + (Y - X\underline{\Phi})'(I + X\underline{\Omega}X')^{-1}(Y - X\underline{\Phi}) \right|^{-\frac{\nu+T}{2}}, \quad (35)$$

где $\Gamma_N(\cdot)$ обозначает N -мерную гамма функцию. Выбор числа лагов происходит аналогично путем максимизации по p функции предельной плотности (35):

$$p^* = \arg \max_p \ln p(Y) \quad (36)$$

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 Φ_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 24⁷. Модель

⁶Т.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 .

⁷Обычная VAR для сравнения качества прогноза строится для 3 и 6 переменных

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. (2006) необходимо выделить период, на котором происходит определение λ . Мы определяем λ 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 λ и по p , и выбираются такие значения, для которых (35) максимально.

Приложения

Данные

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

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.