

BVAR in Russia

я и Боря

10 июля 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 and Watson или BBE 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 et al (2000) и Stock and Watson (2002). 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. Расширением метода Stock and Watson (2002) служит FAVAR, предложенная в статье (Bernanke et 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 et al(1984) (sum of coefficients prior) и Sims (1993) (co-persistence prior). Комбинация этих трех априорных распределений была использована в работе Robertson and Tallman(1999) 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 and Zha(1998) (про эту работу пока ничего не писала).

Ключевую роль в развитии подхода сыграла статья De Mol, et al 2008. В этой работе на основе асимптотического анализа было показано, что если данные характеризуются высокой мультиколлинеарностью (что характерно для выборок макро-рядов 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 et al.(2010), 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 and Matheson (2010), в которой они использовали метод условного прогнозного оценивания (Waggoner and Zha, 1999), 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 et al 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 and Karlsson, 1997; Sims and Zha, 1998- проверить, они ли предложили?), что в работах Banbura et al (2010), Bloor and Matheson (2010) и Koop (2010) был применен в работе Beauchemin and Zaman (2011). 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 and 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 and Whiteman, 2006) 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

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

2.2.1 Априорное распределение Миннесоты

Наиболее простым для применения является априорное распределение Миннесоты, предложенное в работе Litterman (1986) и Doan, et al (1984).

Изначально Litterman (1986) 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 (??) в векторизованном виде²:

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

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

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

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

Тогда:

$$\phi \sim \mathcal{N}(\underline{\phi}, \Xi) \quad (7)$$

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 . Априорная плотность распределения ϕ может быть записана как:

$$p(\phi) = \frac{1}{(2\pi)^{km/2} |\Xi|^{1/2}} \exp \left\{ -\frac{1}{2} (\phi - \underline{\phi})' \Xi^{-1} (\phi - \underline{\phi}) \right\}. \quad (8)$$

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

$$\phi|Y \sim \mathcal{N}(\bar{\phi}, \bar{\Xi}) \quad (9)$$

$$\text{где } \bar{\Xi} = [\Xi^{-1} + (\Sigma^{-1} \otimes (X'X))]^{-1} \quad (10)$$

$$\text{и } \bar{\phi} = \bar{\Xi}[\Xi^{-1} \underline{\phi} + (\Sigma^{-1} \otimes X)'y]. \quad (11)$$

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.:

$$\underline{\Phi} = \mathbb{E}[(\Phi_l)_{ij}|\Sigma] = \begin{cases} 1, & i = j, l = 1; \\ 0, & \text{в остальных случаях} \end{cases} \quad (12)$$

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:

$$\text{Var}((\Phi_l)_{ij}) = \begin{cases} \frac{\lambda_1^2}{l^2}, & \text{для переменной } i \text{ и лага } l, \\ \frac{\lambda_1^2 \lambda_2^2 \sigma_i^2}{l^2 \sigma_j^2}, & \text{для переменной } j \neq i \text{ и лага } l, \\ \lambda_3^2 \sigma_i^2, & \text{для коэффициентов при экзогенных переменных} \end{cases} \quad (13)$$

Как можно видеть из приведенной выше формулы (??) 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 .

2.2.2 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:

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

Можно показать, что с учетом функции правдоподобия (??) апостериорное рас-

пределение принадлежит к тому же классу (см, например, Zellner, 1971):

$$\Phi|\Sigma, Y \sim \mathcal{N}(\bar{\Phi}, \Sigma \otimes \bar{\Omega}), \quad \Sigma \sim IW(\bar{S}, \bar{\nu}), \quad (15)$$

$$\text{где } \bar{\Phi} = (\underline{\Omega}^{-1} + X'X)^{-1}(\underline{\Omega}^{-1}\underline{\Phi} + X'Y), \quad (16)$$

$$\bar{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1} \quad (17)$$

$$\bar{\nu} = \underline{\nu} + T \quad (18)$$

$$\bar{S} = \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}X'X\hat{\Phi} + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \bar{\Phi}'\bar{\Omega}^{-1}\bar{\Phi} \quad (19)$$

$$\hat{\Phi} = (X'X)^{-1}X'Y \quad (20)$$

$$\hat{E} = Y - X\hat{\Phi} \quad (21)$$

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

В работах Doan et al. (1984) and Sims (1993) 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 et al, 2015).

В работах Banbura et al (2010), Berg and Henzel (2013) соответствующее априорное распределение вводится путем добавления искусственных наблюдений:

$$y^+ = \begin{bmatrix} \text{diag}(\delta_1\sigma_1, \dots, \delta_n\sigma_n) \\ 0_{n(p-1) \times n} \\ \text{diag}(\sigma_1, \dots, \sigma_n) \\ 0_{1 \times n} \end{bmatrix} \quad x^+ = \begin{bmatrix} \text{diag}(1, 2, \dots, p) \otimes \text{diag}(\sigma_1, \dots, \sigma_n)/\lambda & 0_{np \times 1} \\ 0_{n \times np} & 0_{n \times 1} \\ 0_{1 \times np} & \epsilon \end{bmatrix} \quad (22)$$

Модификация априорных распределений

Априорное распределение суммы коэффициентов³ было предложено в работе Doan et al(1984). 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:

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

³sum-of-coefficients prior

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):

$$y^{++} = [\text{diag}(\mu_1, \dots, \mu_n)/\tau] \quad x^{++} = [(1, 2 \dots p) \otimes \text{diag}(\mu_1, \dots, \mu_n)/\tau \quad 0_{n \times 1}] \quad (23)$$

Априорное распределение изначального наблюдения⁴, предложенное в работе Sims(1993) 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.

2.3 Выбор гиперпараметров и числа лагов

Как было показано в работе De Mol et al(2008) 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 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(2010) 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 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:

⁴dummy initial observation prior

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^∞ :

$$FIT^\infty = \frac{1}{2} \cdot \frac{MSFE_y^\infty}{MSFE_y^0} + \frac{1}{2} \cdot \frac{MSFE_\pi^\infty}{MSFE_\pi^0} \quad (24)$$

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 $FIT^{\lambda,m}$:

$$FIT^{\lambda,m} = \frac{1}{2} \cdot \frac{MSFE_y^{\lambda,m}}{MSFE_y^0} + \frac{1}{2} \cdot \frac{MSFE_\pi^{\lambda,m}}{MSFE_\pi^0} \quad (25)$$

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

$$\lambda^* = \operatorname{argmin} |FIT^{\lambda,m} - FIT^\infty| \quad (26)$$

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

Второй алгоритм предложен в работе (Carriero et al, 2002) и представляет собой выбор такого параметра λ_1 , который бы максимизировал функцию предельной плотности:

$$\lambda_{1t}^* = \operatorname{argmax}_{\lambda_1} \ln p(Y) \quad (27)$$

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

$$p(Y) = \int p(Y|\phi)p(\phi)d\phi \quad (28)$$

Если априорное распределение является нормальным - обратным Уишарта, то предельная плотность $p(Y)$ может быть посчитана аналитически (Zellner,1971; Bauwens

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.

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 .

et al,1999; Carriero et al,2012):

$$p(Y) = \pi^{-\frac{Tm}{2}} \times + |(I + X\underline{\Omega}X')^{-1}|^{\frac{N}{2}} \times |\underline{S}|^{\frac{\nu}{2}} \times \frac{\Gamma_N(\frac{\nu+T}{2})}{\Gamma_N(\frac{\nu}{2})} \times \\ \times |\underline{S} + (Y - X\underline{\Phi})'(I + X\underline{\Omega}X')^{-1}(Y - X\underline{\Phi})|^{-\frac{\nu+T}{2}}, \quad (29)$$

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

$$p^* = \operatorname{argmax}_p \ln p(Y) \quad (30)$$