СТАТЬИ

УДК 338.27

VAR-МОДЕЛИРОВАНИЕ В ПРОГРАММНОЙ СРЕДЕ *R*

Бабешко Л.О.

ФГОБУ ВО Финансовый университет при Правительстве РФ, Москва, e-mail: LBabeshko@fa.ru

Данная статья посвящена вопросам реализации алгоритмов моделей векторной авторегрессии (VAR) в программной среде R. Достоинствами моделей векторной авторегрессии являются: простота их использования, точность прогнозов, сопоставимая с точностью сложных макроэкономических моделей, отсутствие каких-либо структурных или идентификационных ограничений на параметры. Однако последняя их особенность приводит к проблеме перепараметризации, поскольку количество оцениваемых параметров быстро возрастает с увеличением числа переменных и максимальной величины лага. Одним из следствий перепараметризации является рост дисперсий оценок параметров. Следует учесть и требование к увеличению выборки, обусловленное корреляционной зависимостью между временными рядами, включенными в спецификацию модели. Для решения этой проблемы рекомендуется контролировать количество параметров наложением на них общей структуры при помощи байесовских или эмпирических байесовских методов. Процедура байесовского оценивания применима к широкому спектру эконометрических моделей, в том числе и моделей векторной авторегрессии. Такая модификация VAR-моделей получила название BVAR (Bayesian vector autoregression) – байесовские модели векторной авторегрессии. Модели BVAR – модели с ограничениями. В качестве программной среды для оценки и исследования моделей векторной авторегрессии в работе выбран язык R, поддерживающий тысячи специализированных модулей и являющийся бесплатной альтернативой эконометрическим пакетам.

Ключевые слова: модели векторной авторегрессии, байесовские методы, проблема перепараметризации, оценки параметров, дисперсии оценок

VAR-SIMULATION IN SOFTWARE ENVIRONMENT *R*

Babeshko L.O.

The Financial University under the Government of the Russian Federation, Moscow, e-mail: LBabeshko@fa.ru

This article is devoted to the implementation of algorithms for vector autoregression (VAR) models in the software environment R. The advantages of vector autoregression models are: ease of use, forecast accuracy comparable to the accuracy of complex macroeconomic models, absence of any structural or identification constraints by parameters. However, their last feature leads to the problem of reparameterization, since the number of estimated parameters increases rapidly with an increase in the number of variables and the maximum value of the lag. One of the consequences of reparameterization is an increase in the variances of parameter estimates. The requirement to increase the sample due to the correlation dependence between the time series included in the model specification should also be taken into account. To solve this problem, it is recommended to control the number of parameters by imposing a common structure on them using Bayesian or empirical Bayesian methods. Bayesian estimation is applicable to a wide range of econometric models, including vector autoregression models. This modification of VAR models is called BVAR (Bayesian vector autoregression) – Bayesian vector autoregression models. BVAR models are limited models. As a software environment for evaluating and researching vector autoregression models, we have chosen the R language, which supports thousands of specialized modules and is a free alternative to econometric packages.

Keywords: vector autoregression models, Bayesian methods, reparameterization problem, parameter estimates, variance of estimates

Векторные авторегрессионные модели (*VAR*), предложенные К. Симсом в 1980 г., являются классическим инструментом анализа динамики нескольких связанных друг с другом временных рядов. Преимущество классической модели *VAR* по сравнению с системами одновременных уравнений (СОУ) в структурной форме состоит в отсутствии необходимости априорных ограничений, гарантирующих идентификацию. Спецификация модели *VAR*(*p*) для *m*-мерного вектора переменных модели Y_t имеет вид [1]:

$$Y_t = \delta + A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + \\ + A_k Y_{t-k} + \ldots + A_p Y_{t-p} + u_t, \quad (1)$$

где A_k – ($m\times m$)-матрица, $k = 1,\ldots, p$ – величина лага, u_t – *m*-мерный вектор с ($m\times m$)-ковариационной матрицей Ω и нулевым математическим ожиданием, $E\{u_t\} = 0$, δ – *m*-мерный вектор свободных членов. Система (1), с одной стороны, является обобщением авторегрессионных моделей для многомерных временных рядов (временные ряды должны быть стационарными), а с другой – представляет собой систему одновременных уравнений в ее приведенной форме (эндогенные переменные выражены в явном виде через предопределенные). Особенность спецификации модели *VAR* состоит в том, что вектор предопределенных переменных не включает экзогенные переменные.

При моделировании *VAR*(*p*)-процессов спецификацию модели (1) обобщают следующим образом [2]:

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + ... + A_p Y_{t-p} + CD_t + u_t \, , \quad (2)$$

где A_k – ($m \times m$)-матрица коэффициентов, $k = 1,..., p$ – величина лага, u_t – m-мерный вектор процесса белого шума с ($m \times m$)-ковариационной матрицей Ω, D_t – ($M \times 1$)-вектор столбец, включающий детерминированные регрессоры (которыми могут быть: константа, тренд, фиктивные переменные); C – ($m \times M$)-матрица коэффициентов детерминированных регрессоров. Отсутствие автокорреляции вектора возмущений в модели *VAR*(*p*) значительно упрощает процедуру ее оценивания. Так как случайные возмущения уравнений представляют собой белый шум, который не коррелирует с прошлыми значениями уровней ряда, для оценки параметров модели *VAR* используется обычный метод наименьших квадратов (МНК), который последовательно применяется к каждому уравнению системы отдельно. Состоятельная оценка ковариационной матрицы случайных возмущений вычисляется через МНК-остатки отдельных уравнений, входящих в систему. В случае, если возмущение модели имеет нормальное распределение, для оценки параметров можно использовать условный метод максимального правдоподобия (ММП), алгоритм которого, как и алгоритм МНК, включен в современные эконометрические пакеты.

VAR(p)-модели в программной среде R

При построении моделей *VAR*(*p*) необходимо: 1) проверить временные ряды, включенные в вектор эндогенных переменных модели, на стационарность; 2) выбрать максимальную величину лага, учитываемую в спецификации. Для проверки на стационарность временного ряда широкое применение получил расширенный тест Дики–Фуллера *ADF* (*Augmented Dickey–Fuller Test*), реализованный во всех эконометрических пакетах. В тесте *ADF* предполагается, что нестационарность экономических данных объясняется их генерацией процессами случайного блуждания, случайного блуждания с дрейфом или комбинацией случайного блуждания с дрейфом и линейным временным трендом. В программной среде *R* расширенный тест Дики–Фуллера выполняется при помощи функции *adf.test*(*x*) пакета *tseries* с основным аргументом *x* – объект «временной ряд» [1].

Максимальная величина лага модели *VAR*(*p*) подбирается при помощи информационных критериев. Модель оценивается при различных значениях лага, и выбирается лаг, минимизирующий значение информационного критерия. Функция, реализующая процедуру выбора максимальной величины лага в программной среде, *R* – *VARselect*() пакета *vars*. Функция возвращает значения информационных критериев и ошибку прогноза для последовательного увеличения лага.

Для оценки модели *VAR*(*p*) и построения прогнозов в программной среде *R* используется функция *VAR* () пакета *vars* с основными параметрами: *M* – матрица, включающая значения эндогенных переменных модели; *p* – величина лага (по умолчанию равная единице), *type* – тип включения детерминированных регрессоров («*const*» – включение свободного члена, «*trend*» – включение тренда, «*both*» – включение свободного члена и тренда, «*none*» – модель без свободного члена и тренда), *exogen* – включение экзогенных переменных в спецификацию модели, *ic* – выбор информационного критерия для определения оптимальной величины максимального значения лага («*AIC*» – Акайке, «*HQ*» – Хеннона–Куина, «*SC*» – Шварца).

Прогнозирование *VAR*-модели выполняется после оценки ее параметров и диагностики предпосылок при помощи функции *predict* для объектов класса *varest* для периода упреждения, задаваемого параметром *n.ahead*, с вычислением доверительных интервалов, границы которых вычисляются для заданной доверительной вероятности (по умолчанию 0,95). Функция возвращает список из трех элементов – объектов класса *varprd* [2]:

fcst – список матриц, содержащих прогнозные значения, нижнюю и верхнюю границы в соответствии с выбранным доверительным интервалом, *ci* и его среднее значение; *endog* – матричный объект, содержащий эндогенные переменные; *varest* – объект, формируемый функцией *VAR*().

Графическое представление объектов класса *varprd* осуществляется при помощи функции *fanchart*(), которая позволяет пользователям задавать цвета и критические значения для построения доверительных интервалов.

Таблица 1

Годовые значения макроэкономических показателей

№	Год	*Y*	*X*	№	Год	*Y*	*X*
1	1991	0,9	0,5	16	2006	17809,7	5698,8
2	1992	9,2	6,6	17	2007	21968,6	8034,1
3	1993	106,8	46,3	18	2008	27543,5	10526,1
4	1994	422,1	156	19	2009	29269,6	7344,8
5	1995	1016,6	363,4	20	2010	32514,6	10472,7
6	1996	1435,9	475,2	21	2011	40883,8	14584,1
7	1997	1776,1	514,8	22	2012	47273,4	16721,9
8	1998	2003,8	393,5	23	2013	52433,6	16985
9	1999	3285,7	715,3	24	2014	56735,9	17695,5
10	2000	4476,8	1365,7	25	2015	58531,1	18402,8
11	2001	5886,8	1963,1	26	2016	61398,5	19773,4
12	2002	7484,1	2169,3	27	2017	65289,5	21681,2
13	2003	9058,7	2755,1	28	2018	70147,5	22996,2
14	2004	11477,9	3558,9	29	2019	75578,5	25427,6
15	2005	14438,2	4338,7				

VAR(p)-модель: оценка на основании эмпирических данных

Для построения модели *VAR* были использованы два макроэкономических показателя: расходы на конечное потребление (*Y*) и валовое накопление (*X*) в РФ. В табл. 1 приводятся годовые данные показателей за период с 1991 по 2019 гг. включительно (в млрд руб.) [3].

Для проверки временных рядов на стационарность использовался расширенный тест Дики–Фуллера в программной среде *R*. Как и следовало ожидать, макроэкономические ряды *Y* и *X* нестационарны. Поэтому в качестве первичной обработки они были подвергнуты логарифмированию и вычислению разностей первого порядка. Преобразованные переменные *y* и *x*, практически показывающие процентное изменение динамики исходных показателей, являются стационарными и используются в качестве исходных переменных при построении модели *VAR*. Выбор величины лага модели *VAR*(*p*) выполнялся одновременно с тестированием предпосылок модели относительно случайного возмущения при помощи обобщенных тестов, применяемых для диагностики моделей многомерных временных рядов: теста Портманто (*Portmanteau Test*) на автокорреляцию, теста Бреуша–Годфри (*Breusch–Godfrey*) на гетероскедастичность, теста Харке–Бера (*Jarque–Bera test*) на нормальность распределения возмущений. Остатки модели *VAR*(1) оказались автокоррелированными, и к тому же стандартные ошибки модели *VAR*(1): $s_y = 0{,}198$, $s_x = 0{,}248$ больше стандартных ошибок модели *VAR*(2) $s_y = 0{,}098$, $s_x = 0{,}204$. Запишем спецификацию модели авторегрессии второго порядка *VAR*(2) с двумя переменными:

$$\begin{cases} y_t = a_{10} + a_{11}y_{t-1} + a_{12}x_{t-1} + b_{11}y_{t-2} + b_{12}x_{t-2} + v_t \\ x_t = a_{20} + a_{21}y_{t-1} + a_{22}x_{t-1} + b_{21}y_{t-2} + b_{22}x_{t-2} + u_t \end{cases} \quad (3)$$

и ее оцененную форму:

$$\begin{cases} \hat{y}_t = \underset{(0,024)}{0{,}076} + \underset{(0,150)}{0{,}706} \cdot y_{t-1} - \underset{(0,120)}{0{,}164} \cdot x_{t-1} + \underset{(0,109)}{0{,}105} \cdot y_{t-2} - \underset{(0,110)}{0{,}145} \cdot x_{t-2}, R^2 = 0{,}901, \\ \hat{x}_t = \underset{(0,049)}{0{,}090} + \underset{(0,311)}{0{,}944} \cdot y_{t-1} - \underset{(0,248)}{0{,}234} \cdot x_{t-1} + \underset{(0,227)}{0{,}287} \cdot y_{t-2} - \underset{(0,228)}{0{,}541} \cdot x_{t-2}, R^2 = 0{,}660. \end{cases} \quad (4)$$

BVAR(p)-модель: оценка на основании эмпирических данных

Модели векторной авторегрессии являются моделями без ограничений на параметры, поэтому их основным недостатком служит перепараметризация, поскольку количество оцениваемых параметров быстро возрастает с увеличением числа переменных и величины лага, а как известно, для правильного отражения динамики фактических временных рядов требуется включение большого числа лагов [4]. Следует учесть и требование к увеличению выборки для оценки автоковариационной матрицы параметров [5]. Одним из следствий перепараметризации является рост дисперсий оценок параметров. Для решения этой проблемы рекомендуется контролировать количество параметров наложением на них общей структуры при помощи байесовских или эмпирических байесовских методов (*Bayesian vector autoregression*, *BVAR*).

В моделях *BVAR* ограничения на параметры накладываются с учетом априорной информации о характере их распределения. Математическое ожидание параметра при эндогенной переменной с единичным лагом полагается равным единице. В соответствии со схемой назначения априорной информации для параметров модели *VAR*, предложенной Литтерманом, [6], [4]:

1) параметры первого лага эндогенной переменной соответствуют параметрам авторегрессии первого порядка $AR(1)$, параметры при остальных лаговых переменных приравниваются к нулю;

2) стандартные ошибки параметров определяются через постоянные коэффициенты – гиперпараметры, назначаемые исследователем:

λ_1 – параметр регуляризации: при $\lambda_1 \to 0$ априорное распределение совпадает с апостериорным, исходные данные не оказывают влияния на оценку параметров; при $\lambda_1 \to \infty$ апостериорное математическое ожидание параметров сходится к МНК-оценке;

λ_2 – параметр кросс-регуляризации, добавляющий дополнительную жесткость лагам других переменных, включенных в уравнение. Значение $\lambda_2 < 1$ показывает, что влияние лагов других переменных на прогноз эндогенной переменной меньше, чем ее собственных лагов;

λ_3 – отвечает за скорость убывания априорной дисперсии с увеличением лага. Стандартные ошибки параметров вычисляются по следующему правилу [6]:

$$\begin{cases} \lambda_1 / l^{\lambda_3} & \text{если} \quad i = j \\ (\sigma_i \cdot \lambda_1 \cdot \lambda_2) / (\sigma_j \cdot l^{\lambda_3}) & \text{если} \quad i \neq j \end{cases}, \quad (5)$$

где i – номер i-ой переменной в i-ом уравнении модели векторной авторегрессии, j – номер j-ой переменной в i-ом уравнении модели векторной авторегрессии, σ_i, σ_j – стандартные ошибки возмущений модели авторегрессии первого порядка $AR(1)$, для i-ой и j-ой переменных соответственно. Продемонстрируем алгоритм задания априорных параметров распределения для модели векторной авторегрессии второго порядка $VAR(2)$ с двумя переменными (3), с вектором параметров модели $VAR(2)$, включающим следующие элементы:

$$\beta = \left(a_{10}, a_{11}, a_{12}, b_{11}, b_{12}, a_{20}, a_{21}, a_{22}, b_{21}, b_{22}\right)^T. \quad (6)$$

Из обозначений (3) следует, что коэффициентом авторегрессионной модели первого порядка для переменной y_t является параметр a_{11}, для переменной x_t является параметр a_{22}, и, следовательно, в соответствии со схемой назначения априорной информации для параметров модели *VAR*, предложенной Литтерманом, вектор априорных математических ожиданий параметров модели будет иметь вид:

$$\tilde{\beta}_0 = \left(0,1,0,0,0,0,0,1,0,0\right)^T. \quad (7)$$

Диагональные элементы матрицы H априорных дисперсий параметров (6) определяются по правилу (5). В табл. 2 приведены формулы для вычисления средних квадратических отклонений параметров (ско).

Таблица 2

Средние квадратические отклонения параметров модели

Параметры	a_{10}	a_{11}	a_{12}	b_{11}	b_{12}
ско	$\sigma_1\lambda_4$	λ_1	$\sigma_1\lambda_1\lambda_2/\sigma_2$	$\lambda_1/2^{\lambda^3}$	$\sigma_1\lambda_1\lambda_2/\sigma_2 2^{\lambda_3}$
Параметры	a_{20}	a_{21}	a_{22}	b_{21}	b_{22}
ско	$\sigma_2\lambda_4$	$\sigma_2\lambda_1\lambda_2/\sigma_1$	λ_1	$\sigma_2\lambda_1\lambda_2/\sigma_1 2^{\lambda_3}$	$\lambda_1/2^{\lambda_3}$

Применение байесовского подхода к оцениванию параметров модели векторной авторегрессии приводит к следующим формулам для вычисления параметров апостериорного распределения [6]:

$$\beta^* = \left(H^{-1} + \Sigma^{-1} \otimes X_t^T X_t\right)^{-1} \left(H^{-1}\tilde{\beta}_0 + \Sigma^{-1} \otimes X_t^T X_t\hat{\beta}\right), \quad (8)$$

$$V\left(\beta^*\right) = \left(H^{-1} + \Sigma^{-1} \otimes X_t^T X_t\right)^{-1}, \quad (9)$$

где β^* – вектор апостериорных математических ожиданий параметров модели *VAR*(2) с двумя переменными, X_t – матрица исходных временных рядов, $\hat{\beta}$ – вектор МНК-оценок параметров, $\tilde{\beta}_0$ – вектор априорных математических ожиданий параметров модели, Σ – матрица с единицами на главной диагонали, H – диагональная матрица с априорными дисперсиями параметров модели. Ниже приводятся результаты оценивания модели алгоритма (8), (9) с гиперпараметрами: $\lambda_1 = 0,7$, $\lambda_2 = 1$, $\lambda_3 = 2$.

$$\begin{cases} \hat{y}_t = \underset{(0,234)}{0,051} + \underset{(0,480)}{0,757} \cdot y_{t-1} - \underset{(0,377)}{0,191} \cdot x_{t-1} - \underset{(0,109)}{0,01071}_{t-2} - \underset{(0,129)}{0,013} \cdot x_{t-2}, \\ \hat{x}_t = \underset{(0,236)}{0,067} + \underset{(0,582)}{0,009} \cdot y_{t-1} + \underset{(0,458)}{0,290} \cdot x_{t-1} + \underset{(0,225)}{0,017} \cdot y_{t-2} - \underset{(0,171)}{0,045} \cdot x_{t-2}. \end{cases} \quad (10)$$

Стандартные ошибки модели *BVAR*(2) ($s_y = 0,124$; $s_x = 0,272$) больше стандартных ошибок модели *VAR*(2) ($s_y = 0,098$; $s_x = 0,204$). Оценки параметров и их дисперсий существенно зависят от значений гиперпараметров.

Точечную и интервальную оценку параметров регрессионной модели байесовским методом в программной среде *R* можно выполнить при помощи функции *MCMCregress* пакета *MCMCpack*: MCMCregress(formula, data, b0=b0,B0=B0), с основными параметрами: *formula* – спецификация регрессионной модели; *data* – данные в форме data.frame; *b0* – априорное среднее значение вектора параметров; *B0* – априорное значение автоковариационной матрицы оценок параметров [7]. Ниже приводятся результаты оценивания модели (3) по данным табл. 1 при помощи функции *MCMCregress* с начальными значениями вектора параметров (7) и оценкой автоковариационной матрицы *H*:

$$\begin{cases} \hat{y}_t = \underset{(0,188)}{0,694} \cdot y_{t-1} - \underset{(0,148)}{0,112} \cdot x_{t-1} - \underset{(0,137)}{0,138} \cdot y_{t-2} - \underset{(0,137)}{0,147} \cdot x_{t-2}, \\ \hat{x}_t = \underset{(0,343)}{0,925} \cdot y_{t-1} - \underset{(0,270)}{0,167} \cdot x_{t-1} + \underset{(0,250)}{0,327} \cdot y_{t-2} - \underset{(0,251)}{0,543} \cdot x_{t-2}. \end{cases} \quad (11)$$

Выводы

Оценки параметров моделей (4) и (11) отличаются незначительно, однако стандартные ошибки модели *BVAR*(2), оцененные в рамках метода *MCMC*, $s_y = 0,005$, $s_x = 0,017$, значительно меньше стандартных ошибок *BVAR*(2), оцененных при помощи алгоритмов (8), (9) и модели *BVAR*(2). Этот результат подтверждается многочисленными исследованиями в области количественной оценки воздействия внутренних и внешних факторов при моделировании макроэкономических процессов [8].

Список литературы

1. Бабешко Л.О., Орлова И.В. Инструментарий современного эконометрического моделирования. М.: Центркаталог, 2020. 336 с.

2. Pfaff B. Analysis of Integrated and Cointegrated Time Series with R. Springer. 2008. 189 p.

3. Единый архив экономических и социологических данных (ЕФЭСД), НИУ ВШЭ [Электронный ресурс]. URL: http://sophist.hse.ru (дата обращения: 27.01.2021).

4. Демешев Б.Б., Малаховская О.А. Картографирование BVAR. Препринт / Высшая школа экономики. М., 2015. 37 с.

5. Пелипась И. Шиманович Г., Кирхнер Р. Международные связи и внешние шоки: опыт использования различных спецификаций глобальной VAR для Беларуси // Аналитические записки исследовательского центра ИПМ. 2016. № 2. 31 с.

6. Погосян К. Альтернативные модели прогнозирования основных макроэкономических показателей в Армении // Квантиль. 2015. № 13. С. 25–39.

7. Бабешко Л.О. Байесовский подход в эконометрике и его реализация в программной среде R // Современные проблемы науки и образования. 2019. № 3. [Электронный ресурс]. URL: http://science-education.ru/ru/article/view?id=28938 (дата обращения: 23.03.2021).

8. Шевелев А.А. Байесовский подход к оценке воздействия внешних шоков на макроэкономические показатели России // Мир экономики и управления. 2017. Т. 17. № 1. С. 26–40.