

УДК 338.27

МЕТОДИЧЕСКИЕ АСПЕКТЫ ОБОБЩЕННЫХ МЕТОДОВ ДИАГНОСТИКИ В МОДЕЛЯХ ВЕКТОРНОЙ АВТОРЕГРЕССИИ

Бабешко Л.О.

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

Моделям векторной авторегрессии (VAR), предложенным Симсом в 1980 г., свободным от проблем идентификации переменных и нашедшим широкое применение благодаря их гибкости, релевантности, сопоставимости по точности со сложными структурными макроэкономическими моделями, простоте реализации и экономической интерпретации результатов, посвящено множество публикаций прикладного характера. При построении эконометрических моделей любых типов, в том числе и моделей векторной авторегрессии, одним из важнейших этапов является этап проверки выполнения их предпосылок – диагностика моделей. Для упрощения анализа m -мерный VAR(p)-процесс представляют в виде mp -мерного VAR(1)-процесса. Такой подход позволяет применять для оценки и прогнозирования векторной авторегрессии аппарат, разработанный для моделей авторегрессии одномерных временных рядов, а к диагностике моделей VAR(p) применять тесты, включающие обобщенные статистики одномерных аналогов. Для тестирования предпосылок моделей векторной авторегрессии разработаны обобщенные тесты – многомерные аналоги одномерных тестов на автокорреляцию (Portmanteau Test, PT.adjusted), гетероскедастичность (Breusch-Godfrey, Edgerton-Shukur), нормальность распределения возмущений (Jarque-Bera test), тестирование причинности (Granger). Поскольку в последние десятилетия в качестве инструментальной поддержки эконометрических методов оценки, диагностики и анализа моделей в большинстве экономических вузов используется бесплатная альтернатива профессиональным специализированным пакетам – программная среда R, в статье обсуждаются вопросы реализации многомерных тестов диагностики VAR-моделей на языке R.

Ключевые слова: модель векторной авторегрессии, диагностические тесты, тест Харке – Бера, обобщенный тест Харке – Бера, тесты Портманто, коэффициент асимметрии, куртозис

METHODICAL ASPECTS OF GENERALIZED DIAGNOSTICS METHODS IN VECTOR AUTOREGRESSIVE MODELS

Babeshko L.O.

Financial University under the Government of the Russian Federation, Moscow,

e-mail: LBabeshko@fa.ru

The vector autoregression (VAR) models proposed by Sims in 1980, free of variable identification problems and widely used due to their flexibility, relevance, comparability in accuracy with complex structural macroeconomic models, simplicity of implementation and economic interpretation of the results, are devoted many publications of an applied nature. When constructing econometric models of any type, including vector autoregressive models, one of the most important stages is the stage of verifying that their premises are fulfilled – model diagnostics. To simplify the analysis, the m -dimensional VAR (p) -process is represented as an mp -dimensional VAR (1) -process. This approach makes it possible to use the apparatus developed for autoregressive models of one-dimensional time series for estimating and predicting vector autoregression, and apply tests including generalized statistics of one-dimensional analogues to the diagnosis of VAR (p) models. To test the prerequisites of the vector autoregressive models, generalized tests have been developed – multidimensional analogues of one-dimensional autocorrelation tests (Portmanteau Test, PT.adjusted), heteroscedastic behavior (Breusch-Godfrey, Edgerton-Shukur), normal distribution of disturbances (Jarque-Bera test), causality testing (Granger). Since in recent decades, as an instrumental support for econometric methods of model assessment, diagnostics, and analysis, most economic universities use a free alternative to professional specialized packages – the software environment R, the article discusses the implementation of multivariate diagnostic tests of VAR models in the R language.

Keywords: vector autoregressive models, diagnostic tests, Jarque-Bera test, generalized Jarque-Bera test, Portmanteau tests, skewness, kurtosis

Модели векторной авторегрессии (VAR, *vector autoregressive model*), предложенные Кристофером Симсом в 1980 г. и предназначенные для описания нескольких динамических процессов на основе их общей истории, широко используются как альтернатива сложным структурным макроэкономическим моделям благодаря их гибкости, релевантности, сопоставимости по точности, простоте реализации и экономической интерпретации результатов [1–3]. Данный факт особенно важен для повышения мотивации студентов экономических направлений вузов к изучению многомерных моде-

лей временных рядов в эконометрике и их применению при исследовании экономических процессов.

Модели векторной авторегрессии, с одной стороны, являются обобщением авторегрессионных моделей для многомерных временных рядов, с другой – частным случаем систем одновременных уравнений, сохраняя преемственность в методах оценивания, диагностики и исследования, изучаемых студентами бакалавриата в рамках базовой дисциплины «Эконометрика». При построении эконометрических моделей любых типов (моделей временных рядов, регрессионных

моделей с одним уравнением, систем одновременных уравнений) одним из важнейших этапов является этап проверки выполнения их предпосылок – диагностика моделей. Большинство диагностических тестов (*Diagnostic Tests*) базируется на предпосылке нормальности распределения вектора возмущений. Нарушение этой предпосылки приводит к ошибочным выводам и интерпретации результатов оценивания [4, 5].

Для проверки нормальности распределения возмущений, по крайней мере в эконометрике, широкое применение получил тест Харке – Бера (*Jarque-Bera*), реализация которого выполнена практически во всех эконометрических пакетах [6–8]. Обобщенный вариант теста используется для проверки нормальности распределения вектора возмущений модели векторной авторегрессии. Для проверки остатков модели на наличие автокорреляции применяется обобщенная процедура тестов Бокса – Пирса (*Box-Pierce Q-statistic*) и Бокса – Льюинга (*Ljung-Box Q-statistic*). Для исследования *гетероскедастичности* в многомерных моделях временных рядов используются обобщенные тесты *ARCH*.

Поскольку в последние десятилетия в качестве инструментальной поддержки эконометрических методов оценки, диагностики и анализа моделей, в большинстве экономических вузов используется бесплатная альтернатива профессиональным специализированным пакетам – программная среда *R* [9], реализация многомерных тестов будет рассмотрена на языке *R*.

*Обобщенный тест Харке – Бера
на нормальность распределения
ошибок модели VAR*

Одним из достоинств моделей векторной авторегрессии является отсутствие необходимости априорных ограничений на параметры модели, гарантирующих идентификацию. Включенные в спецификацию переменные – эндогенные переменные модели, зависят от своих прошлых значений и прошлых значений других переменных.

Для упрощения анализа m -мерный $VAR(p)$ -процесс представляют в виде mp -мерного $VAR(1)$ -процесса (m – число переменных модели, p – максимальная величина лага) [10]. Такой подход реализован в функции $VAR()$ пакета *vars* программной среды *R* и позволяет применять для оценки и прогнозирования векторной авторегрессии аппарат, разработанный для моделей авторегрессии одномерных временных рядов, а к диагностике моделей $VAR(p)$ применять тесты, включающие обобщенные статистики одномерных аналогов.

Статистика одномерного (*univariate*) теста Харке – Бера (*Jarque-Bera test*) [6, 10] на нормальность распределения остатков эконометрической модели основана на сравнении центральных нормированных моментов третьего (коэффициент асимметрии, *Skewness*), $S = \mu_3 / \sigma^3$, и четвертого (коэффициент эксцесса, *Kurtosis*) $K = \mu_4 / \sigma^4$ порядков с соответствующими характеристиками нормального распределения, для которого, как известно, $S = 0$, $K = 3$, и имеет вид

$$JB = n \left(\frac{\hat{S}^2}{6} + \frac{(\hat{K} - 3)^2}{24} \right). \quad (1)$$

Оценки характеристик, включенных в формулу (1), вычисляются через остатки (e) регрессионной модели:

$$\hat{S} = \sum_{t=1}^n e_t^3 / n \hat{\sigma}_{ML}^3, \quad \hat{K} = \sum_{t=1}^n e_t^4 / n \hat{\sigma}_{ML}^4,$$

$$\hat{\sigma}_{ML}^2 = \sum_{t=1}^n e_t^2 / n.$$

Нулевая и альтернативная гипотезы теста Харке – Бера формулируются следующим образом: $H_0: S = 0, K = 3, H_1: S \neq 0, K \neq 3$.

Статистика (1) имеет распределение хи-квадрат с двумя степенями свободы, и, если вычисленное значение больше критического, нулевая гипотеза о нормальном распределении возмущений регрессионной модели не отклоняется. Тест Харке – Бера является асимптотическим тестом и применим к большим выборкам.

В обобщенной процедуре (*multivariate*) многомерного теста Харке – Бера, нормирование моментов выполняется при помощи матрицы факторизации P для оценки автоковариационной матрицы остатков модели $VAR(p)$, представляющей собой *нижнюю* треугольную матрицу с положительными диагональными элементами (полученную при помощи разложения Холецкого), $PP' = \hat{\Omega}$. Разложение можно также записать через верхнюю треугольную матрицу $U'U = \hat{\Omega}$, где $U = P'$. Статистика обобщенного теста вычисляется по формуле [10]:

$$JB_m = s_3^2 + s_4^2 \sim \chi_{2m}^2, \quad (2)$$

где

$$s_3^2 = Tb_1'b_1/6 \sim \chi_m^2,$$

$$s_4^2 = T(b_2 - 3m)'(b_2 - 3m)/24 \sim \chi_m^2, \quad (3)$$

$$\omega_t = (\omega_{1t}, \omega_{2t}, \dots, \omega_{mt})' = P^{-1}e_t,$$

$$b_{i1} = \frac{1}{T} \sum_{t=1}^T \omega_{it}^3, \quad b_{i2} = \frac{1}{T} \sum_{t=1}^T \omega_{it}^4.$$

Статистики (2), (3) реализованы в R при помощи функции *normality.test()* пакета `{vars}`: *normality.test(x, multivariate.only = TRUE)*, с основными параметрами: x – объект класса "Varest", генерируемого *VAR()* или объект класса *vec2var*, генерируемого *vec2var()*; *multivariate.only* – логическая переменная, если *TRUE* (по умолчанию), вычисляется только статистика многомерного теста (2), если *FALSE*, вычисляются статистики для многомерного теста (2) и многомерных коэффициентов асимметрии и эксцесса (3). Объекты, созданные функциями с расширением *test*, поддерживают графическое представление эмпирического распределения остатков модели *VAR(p)* (и их квадратов), а также их автокорреляционной функции (*ACF*) и функции частных автокорреляций (*PACF*) [8].

Обобщенные тесты на автокорреляцию остатков

Для проверки адекватности моделей векторной авторегрессии, так же как и для моделей одномерных временных рядов, применяются тесты на отсутствие автокорреляции остатков, в частности обобщение

$$Q_h = T \sum_{i=1}^h \text{tr}(\hat{C}_i' \cdot \hat{C}_0^{-1} \cdot \hat{C}_i \cdot \hat{C}_0^{-1}) \sim \chi^2(m^2(h-p)), \tag{5}$$

где

$$\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T e_t \cdot e_{t-i}', \quad i = 0, 1, \dots, h \tag{6}$$

– оценка автоковариационной матрицы остатков модели *VAR(p)*, p – порядок модели векторной авторегрессии, h – порядок автокорреляции, для которого выполнена нулевая предпосылка, m – число эндогенных переменных модели. Статистика (5) используется для проверки нулевой гипотезы:

$$H_0 : R_1 = \dots = R_i = \dots = R_h = 0, \tag{7}$$

где R_i – автокорреляционная матрица, оценка которой вычисляется по формуле

$$\hat{R}_i = \hat{D}^{-1} \cdot \hat{C}_i \cdot \hat{D}^{-1}, \tag{8}$$

где \hat{D} – диагональная матрица с диагональными элементами, равными стандартным ошибкам возмущений *VAR(p)*-процесса

$$[\hat{D}]_{diag} = [\hat{C}_0]_{diag}^{0,5}.$$

тестов Бокса – Пирса (*Box-Pierce Q-statistic*) и Бокса – Льюинга (*Ljung-Box Q-statistic*), которые реализованы в эконометрических пакетах.

Для проверки нулевой гипотезы $H_0: \rho_1 = \rho_2 = \dots = \rho_h = 0$, против альтернативной

$H_1: \sum_{i=1}^h \rho^2(i) > 0$, в тесте Бокса – Пирса используется статистика Q :

$$Q = T \sum_{i=1}^h r^2(i) \sim \chi^2(h), \tag{4}$$

где $\rho(i)$ – теоретические значения коэффициентов корреляции, $r(i)$ – выборочные коэффициенты корреляции с числовыми характеристиками: $E\{r\} = \rho$, $Var\{r\} = (1 - \rho^2)/T$, h – порядок автокорреляции, для которого выполнена нулевая предпосылка. Если уровни временного ряда – *белый шум*, то коэффициенты автокорреляции $r(i), i = 1, \dots, h$ асимптотически независимы и при большом объеме наблюдений $r(i) \sim N(0, 1/T)$. Вычисленное по формуле (4) значение статистики сравнивается с критическим, и если $Q > \chi_\alpha^2(h)$, то нулевая гипотеза о белом шуме (отсутствии автокорреляции) отклоняется на уровне значимости α (правосторонняя критическая область). Обобщением статистики (4) для случая моделей *VAR(p)* является статистика теста Портманто (*Portmanteau Test, (PT.asymptotic)*):

На малых выборках используют тест Бокса – Льюинга со статистикой

$$Q_{LB} = T(T+2) \sum_{i=1}^h \frac{r^2(i)}{(T-i)} \sim \chi^2(h), \quad (9)$$

в которую каждое слагаемое входит со своим весом $w_i = (T-i)/\{T(T+2)\}$.

Обобщенный (многомерный) вариант статистики (9) для моделей $VAR(p)$ с учетом (5) принимает вид (*PT.adjusted*):

$$Q_{LB(h)} = T^2 \sum_{i=1}^h \frac{1}{T-i} \text{tr} \left(\hat{C}_i' \cdot \hat{C}_0^{-1} \cdot \hat{C}_i \cdot \hat{C}_0^{-1} \right) \sim \chi^2(m^2(h-p)). \quad (10)$$

Для тестирования автокорреляции используется также обобщенный вариант теста Бреуша – Годфри. Тест является асимптотическим, то есть для достоверности выводов требуется большой объем выборки. В тесте рассматривается авторегрессия остатков на их лаговые значения и оцениваются две модели: исходная модель $VAR(p)$ (усеченная, с ограничениями, *restricted*)

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + CD_t + \varepsilon_t, \quad (11)$$

и вспомогательная, общая (неусеченная, без ограничений, *unrestricted*) модель

$$e_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + CD_t + B_1 e_{t-1} + B_2 e_{t-2} + \dots + B_h e_{t-h} + u_t, \quad (12)$$

где e_t – остатки, полученные в результате МНК оценивания основной модели (11), $e_{t-j} = 0$ при $t < j+1$. Если нулевая гипотеза

$$H_0 : B_1 = B_2 = \dots = B_h = 0$$

(против альтернативной $H_1 : \exists B_i \neq 0$ для $i = 1, 2, \dots, h$) верна, то при большом значении объема выборки T статистика теста определяется по формуле [8]:

$$LM_h = T \cdot \left(m - \text{tr} \left(\Omega_R^{-1} \cdot \Omega_{UR} \right) \right) \sim \chi^2(hm^2), \quad (13)$$

где Ω_R – автоковариационная матрица вектора остатков исходной (усеченной) модели $VAR(p)$ (11), Ω_{UR} – автоковариационная матрица вектора остатков неусеченной модели (12), m – число эндогенных переменных модели, имеет хи-квадрат распределение с параметром hm^2 , и гипотеза не отклоняется при уровне значимости α , если выполняется неравенство

$$BG = LM_h < \chi_{\alpha}^2(hm^2).$$

Тесты на проверку наличия автокорреляции реализованы в программной среде R при помощи функции *serial.test()* пакета `{vars}`:

$$\text{serial.test}(x, \text{lags.pt}, \text{lags.bg}, \text{type} = c("PT.asymptotic", "PT.adjusted", "BG")),$$

с основными аргументами: x – объект, генерируемый функцией $VAR()$; *lags.pt* – величина лага, используемая в статистике теста Портманто (5), *lags.bg* – величина лага, используемая в статистике теста Бреуша – Годфри (13); *type* – тип теста: *PT.asymptotic* – (5); *PT.adjusted* – (10); *BG* – *Breusch-Godfrey* – (13). По умолчанию используется асимптотический тест Портманто.

Обобщенные тесты на гетероскедастичность остатков

Для исследования гетероскедастичности в многомерных моделях временных рядов используются тесты *ARCH*, основанные на оценивании вспомогательной модели вида [10]:

$$\text{vec}(e_t e_t') = \beta_0 + B_1 \text{vec}(e_{t-1} e_{t-1}') + B_2 \text{vec}(e_{t-2} e_{t-2}') + \dots + B_q \text{vec}(e_{t-q} e_{t-q}') + v_t, \quad (14)$$

где v_t – процесс белого шума, $\text{vec}(\cdot)$ – оператор формирования вектора из нижней треугольной матрицы с элементами на главной диагонали и вертикальным выстраиванием её столб-

цов. Для реализации оператора $vec(\cdot)$ в программной среде используется функция $lower.tri(x,diag=TRUE)$, пакета $\{base\}$, с основными параметрами: x – матрица, $diag$ – логическая переменная: $TRUE$ – включать элементы на главной диагонали, $FALSE$ – не включать элементы на главной диагонали. Если размерность матрицы x – $(m \times m)$, то сформированный вектор имеет размерность – $m(m + 1)/2$.

Таким образом, размерность параметров в модели (14):

$$\beta_0 - m(m + 1)/2, B_i - (m(m + 1)/2) \times (m(m + 1)/2), i = 1, \dots, q.$$

Нулевая и альтернативная гипотезы формулируются следующим образом:

$$H_0 : B_1 = B_2 = \dots = B_q = 0, H_1 : B_1 \neq 0 \cap B_2 \neq 0 \cap \dots \cap B_q \neq 0,$$

и проверяются при помощи статистики

$$VARCH_{LM(q)} = Tm(m + 1)R_m^2/2 \sim \chi^2(qm^2(m + 1)^2/4), \tag{15}$$

с коэффициентом детерминации:

$$R_m^2 = 1 - \frac{2}{m(m + 1)} tr(\hat{\Sigma}\hat{\Sigma}_0^{-1}),$$

где $\hat{\Sigma}$ – автоковариационная матрица вектора возмущений модели без ограничений на параметры, $\hat{\Sigma}_0$ – с ограничениями на параметры в рамках гипотезы H_0 . Статистика (15) реализована в программной среде R в виде функции $arch.test()$ пакета $vars$, которая вычисляет одномерные и многомерные тесты $ARCH-LM$ для моделей $VAR(p)$. Основные параметры функции:

$$arch.test(x, lags.single = 16, lags.multi = 5, multivariate.only = TRUE),$$

где x – объект класса "Varest", генерируется $VAR()$ или $vec2var()$; $lags.single$ – целое число, указывающее величину лага для одномерной статистики $ARCH$; $lags.multi$ – целое число, указывающее величину лага для многомерной статистики $ARCH$; $multivariate.only$ – Если $TRUE$ (по умолчанию), вычисляется только статистика многомерного теста.

В таблице приведены скрипт и фрагмент протокола результатов оценивания и диагностики модели векторной авторегрессии $VAR(1)$, оцененной для двух эндогенных переменных: объём инвестиций и величина капитала, по годовым данным фирмы «General Electric» за 20 лет, встроенным в пакет "AER".

Результаты оценки и диагностики модели $VAR(p)$

Фрагменты скрипта	Фрагменты протокола
#оценка модели VAR	Estimation results for equation Dy:
library(tseries)	=====
adf.test()	Dy = Dy.l1 + Dx.l1 + trend
library(vars)	Estimate Std.Error t value Pr(> t)
library("AER")	Dy.l1 -0.2328 0.2139 -1.089 0.29352
data("Grunfeld", package="AER")	Dx.l1 -0.8288 0.2300 -3.603 0.00261 **
Y<-GE\$invest	trend 3.8148 1.0155 3.757 0.00191 **
X<-GE\$capital	Residual standard error: 23.32 on 15 df
Dy<- diff(Y)	Multiple R-Squared: 0.5133, Adj R-squared: 0.416
Dx<- diff(X)	F-statistic: 5.274 on 3 and 15 DF, p-value: 0.01104
#adf.test(Dy)	Estimation results for equation Dx:
#adf.test(Dx)	=====
varmat <- as.matrix(cbind(Dy,Dx))	Dx = Dy.l1 + Dx.l1 + trend
varfit <- VAR(varmat,1,type="trend",ic="AIC")	Estimate Std. Error t value Pr(> t)
summary(varfit)	Dy.l1 0.95253 0.07516 12.674 2.04e-09 ***
	Dx.l1 0.90503 0.08083 11.197 1.11e-08 ***
	trend 0.07676 0.35684 0.215 0.833
	Residual standard error: 8.193 on 15 df
	Multiple R-Squared: 0.9809, Adj R-squared: 0.977
	F-statistic: 256.4 on 3 and 15 DF, p-value: 4.166e-13

Окончание таблицы	
Фрагменты скрипта	Фрагменты протокола
# диагностика:	
# проверка на гетероскедастичность arch.test(varfit, lags.single = 1, lags. multi=1,multivariate.only= TRUE)	ARCH (multivariate) data: Residuals of VAR object varfit Chi-squared = 10.866, df = 9, p-value = 0.285
# проверка на автокорреляцию serial.test(varfit, lags.pt = 16, type = "PT.ad- justed")	Portmanteau Test (adjusted) data: Residuals of VAR object varfit Chi-squared = 60.636, df = 60, p-value = 0.4527
# проверка на нормальность var.norm<-normality.test(varfit, multivariate. only = TRUE)	JB-Test (multivariate) data: Residuals of VAR object varfit Chi-squared = 9.5024, df = 4, p-value = 0.0497 Skewness only (multivariate) data: Residuals of VAR object varfit Chi-squared = 5.9443, df = 2, p-value = 0.05119 Kurtosis only (multivariate) data Residuals of VAR object varfit Chi-squared = 3.5581, df = 2, p-value = 0.1688

Выводы

Как следует из протокола, остатки модели имеют нормальное распределение. Включение лаговых переменных в спецификацию $Var(p)$ позволило преодолеть проблему автокорреляции, уточнить причинно-следственные связи и повысить качество модели.

С точки зрения образовательного процесса аппарат Var -моделей полностью соответствует принципу систематичности и последовательности обучения: новый учебный материал по построению и диагностике $VAR(p)$ моделей базируется на знаниях ранее изученных эконометрических методов, расширяет и дополняет их.

Список литературы

1. Маматова Н. Применение модели векторной авторегрессии для анализа потребления электроэнергии // Математические модели экономики. 2015. № 4. С. 15–19.
2. Паньков М.О. Анализ взаимосвязей торгов акций на фондовых биржах с применением методов векторной авторегрессии // Финансовый менеджмент. 2018 № 3. С. 39–43.
3. Суханова Е.И., Ширнаева С.Ю. Прогнозирование показателей стабилизационных процессов экономики России на основе моделей векторной авторегрессии // Фундаментальные исследования. 2014. № 9–7. С. 1590–1595.
4. Бабешко Л.О., Бич М.Г., Орлова И.В. Эконометрика и эконометрическое моделирование: учебник. М.: Вузский учебник: ИНФРА-М, 2017. 400 с.
5. Носко В.П. Эконометрика Кн. 2. Ч. 3, 4: учебник. М.: Издательский дом «Дело» РАНХиГС, 2011. 576 с.
6. Jarque C.M. and Bera A.K. A test for normality of observations and regression residuals, International Statistical Review. 1987. № 55. P. 163–172.
7. Kleiber C. Applied Econometrics with R. N.Y.. Springer-Verlag, 2008. P. 222.
8. Pfaff B. Analysis of Integrated and Cointegrated Time Series with R. Springer, 2008. 189 p.
9. Зададаев С.А. Математика на языке R: учебник. М.: Прометей, 2018. 324 с.
10. Айвазян С.А., Фантаццини Д. Эконометрика-2: Продвинутый курс с приложениями в финансах: учебник. М.: Магистр: Инфра-Б, 2014. 944 с.