МНОГОФАКТОРНАЯ МОДЕЛЬ ИНДЕКСА ВВП С НЕЛИНЕЙНО РАСПРЕДЕЛЕННЫМИ ЛАГАМИ

Н. В. Концевая

Финансовый университет при Правительстве Российской Федерации, Москва
E-mail: NVKontsevaya@fa.ru

Рассматривается использование полиномиальной структуры лага при построении многофакторной динамической модели. Показан выбор степени полинома и количества лаговых переменных на примере моделирования и прогнозирования динамики индекса ВВП РФ, показаны преимущества многофакторной динамической модели с полиномиальным лагом в сравнении с классическими многофакторными моделями.

MULTIVARIATE MODEL OF THE GDP INDEX WITH NON-LINEARLY DISTRIBUTED LAGS

N. V. Kontsevaya

The use of the polynomial lag structure in the construction of a multifactor dynamic model is considered. The choice of the degree of the polynomial and the number of lag variables is shown by the example of modeling and forecasting the dynamics of the Russian GDP index, the advantages of a multifactor dynamic model with a polynomial lag in comparison with classical multifactor models are shown.

Идея использования распределенных лагов, предложенная Ирвингом Фишером (1930), обрела популярность в виде полиномиального обобщения лага Фишера, предложенного Ширли Алмон (1965). Подход Алмон можно интерпретировать как наложение набора линейных ограничений на оценку МНК [1, с. 180]. Общий вид модели с одним регрессором с распределенным лагом:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + b_p x_{t-p} + e_t \qquad (1)$$

Спецификация предполагает, что если в момент времени t происходит изменение независимой переменной x, то это изменение будет влиять на значения переменной y в течение p последующих моментов времени. Коэффициент регрессии b_0 при x_t характеризует среднее абсолютное изменение y, при изменении x_t на 1 единицу в текущий момент времени t, без последействия. Иными словами, это краткосрочный мультипликатор. Тогда величину b можно считать долгосрочным мультипликатором:

$$b_0 + b_1 + b_2 + \cdots + b_p = b. \qquad (2)$$

Существуют три основных проблемы при использовании моделей с распределенными лагами на практике. Во-первых, низкая вероятность построить статистически значимую модель с параметрами одинаковых знаков. Во-вторых, текущие и лаговые значения независимой переменной, обычно тесно коррелируют, и оценка параметров модели затруднена в условиях мультиколлинеарности регрессоров. В-третьих, – это проблема автокорреляции остатков. Все вместе делает оценивание модели ненадежной, а иногда и невозможной процедурой.

В качестве борьбы с недостатками выдвигаются априорные предположения относительно определенных ограничений на коэффициенты регрессии в условиях выбранной структуры лага. Ограничения связаны с предполагаемо различным по силе воздействии лаговых переменных на эндогенный показатель и учитываются с помощью коэффициентов регрессии при факторных переменных.

В общем виде, для полинома $k -$ ой степени структура параметров:

$$b_j = c_0 + c_1 j + c_2 j^2 + c_3 j^3 + \dots + c_k j^k . \qquad (3)$$

Подставляя в (1) найденные соотношения для b_j, получаем:

$$y_t = a + c_0 x_t + (c_0 + c_1 + \dots + c_k) x_{t-1} + (c_0 + 2c_1 + 4c_2 + \dots + 2^k c_k) x_{t-2} + (c_0 + 3c_1 + 9c_2 + \dots + 3^k c_k) x_{t-3} + (c_0 + pc_1 + p^2 c_k \dots + p^k c_{k)} x_{t-p} + e_t \qquad (4)$$

Перегруппировав слагаемые, получим модель вида:

$$y_t = a + c_0 z_0 + c_1 z_1 + c_2 z_2 + \dots + c_k z_k + e_t . \qquad (5)$$

где

$$z_k = x_{t-1} + 2^k x_{t-2} + 3^k x_{t-3} + \dots + p^k x_{t-p} . \qquad (6)$$

Тогда параметры многофакторной модели легко оценить обычным МНК:

$$\widetilde{\alpha} = (Z^T Z)^{-1} Z^T Y \qquad (7)$$

Причем мультиколлинеарнность для моделей с полиномиальной структурой лагов является меньшим злом, в сравнении с обычными распределенными лагами. Приблизительное впечатление о надежности оценочных коэффициентов регрессии можно получить, изучив, насколько хорошо линия регрессии объясняет данные, существует ли серийная корреляция в остатках и, среди прочего, является ли общая модель значимой [2, с. 78]. Ковариационная матрица тогда:

$$C_{\tilde{a}\tilde{a}} = \sigma^2 (Z^T Z)^{-1} \qquad (8)$$

Полиномиальная структура распределенных лагов определяет матрицу перехода H таким образом, чтобы выполнялось:

$$\tilde{\beta} = H\tilde{a} \qquad (9)$$

где $\tilde{\beta}$ – оцененные параметры исходной спецификации, причем ковариационная матрица для них:

$$C_{\tilde{\beta}\tilde{\beta}} = Cov(H\tilde{a}, H\tilde{a}) = HC_{\tilde{a}\tilde{a}} H^T. \qquad (10)$$

На практике лаги Алмон могут вызвать следующие трудности. Во-первых, выбор меньшего лага, чем его реальное значение, приведет к неверной спецификации модели, что проявится в остатках в виде невыполнения предпосылок МНК. Выбор слишком большей величины лага будет означать включение статистически не значимых факторов и снижение эффективности оценок, при этом, оценки будут несмещенными. В нескольких исследованиях изучались последствия неправильного указания длины лага или степени полинома в модели лага Алмона: Frost (1975), Harper (1977), Schmidt & Sickles (1975) и д.р. [3, С. 205.]

Вышеуказанные трудности практического использования метода Алмон делают практически невероятным построение спецификации модели, которая бы при оценке оказывалась качественной, адекватной и с непротиворечивой интерпретацией параметров. Видимо, поэтому на практике данный подход не завоевал популярность у эконометристов. Среди альтернативных моделей распределенного лага предложена многомерная теоретическая схема общей модели распределенного лага Грейс Вахбой (1969). Идея в использовании матриц для отслеживания многомерных эффектов различных лаговых переменных. При этом, многомерной модели распределенного лага, предложенной Вахбой, не хватает визуализации и примеров практического применения [4, С. 400].

Целью работы явилось обобщение метода Алмон на случай нескольких регрессоров, причем структура их лаги может включать как разные степени полиномов, так и разного размера максимальные лаги. Например, для случая двух объясняющих внешних факторов:

$$y_t = a + b_0 x1_t + b_1 x1_{t-1} + b_2 x1_{t-2} + \cdots + b_p x1_{t-p} + d_0 x2_t + \\ + d_1 x2_{t-1} + d_2 x2_{t-2} + \cdots + d_p x2_{t-q} + e_t \quad (11)$$

Тогда, предполагая полиномиальную структуру лагов каждого регрессора:

$$b_j = c_0 + c_1 j + c_2 j^2 + c_3 j^3 + \cdots + c_k j^k . \quad (12)$$

$$d_j = m_0 + m_1 j + m_2 j^2 + m_3 j^3 + \cdots + m_l j^l . \quad (13)$$

Сведем модель (11) к виду:

$$y_t = a + c_0 z_0 + c_1 z_1 + c_2 z_2 + \cdots + c_k z_k + m_0 v_0 + m_1 v_1 + m_2 v_2 + \cdots + \\ m_l v_l + e_t. \quad (14)$$

Параметры модели и ковариационную матрицу получим аналогично одномерной модели.

Покажем на примере. Построим динамическую модель индекса ВВП России (GDPEA_Q_DIRI, 2003.01 = 100, данные с сайта http://sophist.hse.ru/, дата обращения 06.0.2023), период наблюдения с 2009.01 по 2022.03, всего 54 квартала. В качестве объясняющих лаговых переменных выбраны: индекс инвестиций в основной капитал (INVFC_Q_DIRI, 1993.01 = 100, данные с сайта http://sophist.hse.ru/, дата обращения 06.0.2023) и индекс выпуска по базовым видам экономической деятельности (BBR_EA_Q_I, 2003.01 = 100), за те же периоды наблюдения. Построим двухфакторную динамическую модель, добавив влияние на текущие квартальные значения индекса ВВП регрессоров за кварталы предыдущего года. Часто бывает так, что между регрессорами существует высокая степень мультиколлинеарности, так что большинство или все оценочные коэффициенты регрессии статистически незначимы, и сделать убедительные выводы об истинных весах невозможно.

Эту проблему можно обойти, введя априорную информацию в процедуру оценки, как правило, наложив ограничения на истинные веса. Если ограничения действительны, оценки весов будут согласованными и более эффективными. Техника запаздывания Алмон вводит априорную информацию путем оценки модели распределенного запаздывания при условии, что веса лежат в полиноме степени p [5]. Результаты оценивания приведены в табл. 1.

Таблица 1

Оценка параметров динамической модели
Модель 2: МНК, использованы наблюдения 2010:1-2022:2 (T = 50)

	Коэффициент	Ст. ошибка	t-статистика	p-значение	
const	38,020	6,085	6,248	<0,0001	***
Z0	-0,058	0,025	-2,348	0,024	**
Z1	-0,005	0,022	-0,240	0,811	
Z2	0,017	0,005	3,338	0,002	***
V0	0,551	0,077	7,139	<0,0001	***
V1	-0,291	0,069	-4,219	0,000	***
V2	0,025	0,017	1,467	0,150	

Приведенная спецификация содержит два незначимых параметра, но все результаты тестирований этой модели на выполнение предпосылок Гаусса-Маркова, на нормальность и точность успешны. Таким образом, незначимость оценок двух параметров вызвана присутствием мультиколинеарности в лаговых переменных, что завышает дисперсии ошибок при оценке параметров, не вызывая смещения самих оценок.

Таблица 2

Оценка качества динамической модели

Среднее завис. перемен	164,3120		Ст. откл. завис. перем	14,16468
Сумма кв. остатков	248,4678		Ст. ошибка модели	2,403814
R-квадрат	0,974727		Исправ. R-квадрат	0,971200
F(6, 43)	276,4010		Р-значение (F)	1,14e-32
параметр rho	0,001216		Стат. Дарбина-Уотсона	1,924285

Тест на нормальное распределение ошибок -

Нулевая гипотеза: ошибки распределены по нормальному закону.

Тест Дурника-Хансена (Doornik-Hansen) = 0,855, p-значение 0,652

Тест Жарка-Бера (Jarque-Bera) = 0,750, p-значение 0,687

Тест Вайта на гетероскедастичность - Нулевая гипотеза: гетероскедастичность отсутствует. Тестовая статистика: LM = 32,0218

p-значение = P(Хи-квадрат(27) > 32,0218) = 0,231277

LM тест на автокорреляцию до порядка 8 - Нулевая гипотеза: автокорреляция отсутствует. Тестовая статистика: LMF = 0,225562

p-значение = P(F(8, 35) > 0,225562) = 0,983748

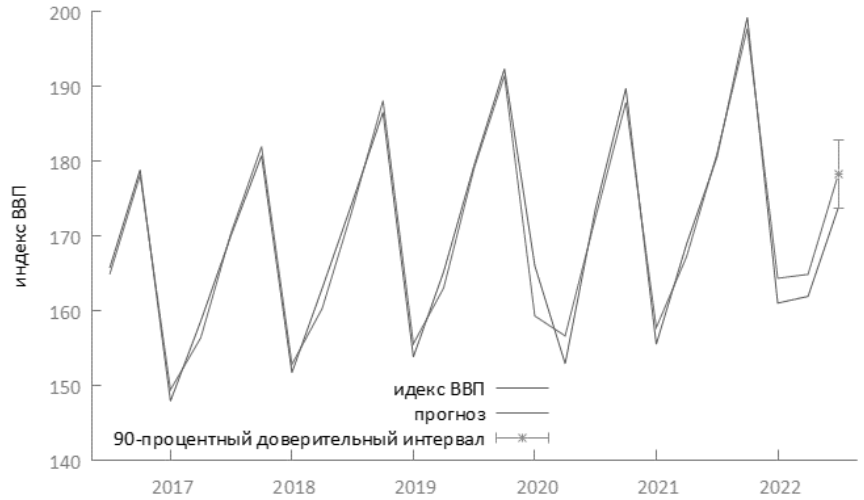
Далее необходимо восстановить исходную спецификацию динамической модели с двумя лаговыми регрессорами, см. табл. 3.

Таблица 3

Оценка параметров восстановленной динамической модели

Пара-метры	X1 - индекс инвестиций	X2 - индекс выпуска	Ст. ошибка	Ст. ошибк а	t-стати-стика	t-стати-стика
alf	38,020		6,085		6,248	
lag 0	-0,058	0,551	0,025	0,081	-2,273	6,839
lag 1	-0,046	0,284	0,017	0,049	-2,687	5,833
lag 2	0,001	0,068	0,022	0,036	0,049	1,857
lag 3	0,078	-0,100	0,034	0,042	2,314	-2,356
lag 4	0,190	-0,218	0,048	0,094	3,973	-2,313
					t-критич	**1,681**

Модель при проверке на контрольном квартале (3.2022) оказалась адекватна с P=0,9 (см. рисунок), не смотря на сложный 2022 год.



Модель индекса ВВП с распределенными лагами

Используя несколько регрессов, структура влияния которых во времени распределена нелинейно, можно получать модели хорошего качества, согласующиеся с теорией в плане интерпретации оцененных параметров, демонстрирующие адекватные прогностические способности и приемлемые для прогнозирования. Направлением улучшения спецификации модели может служить отбор лагов, при которых параметры модели положительны, причем значения максимальных лагов и порядок их следования могут не совпадать для разных объясняющих переменных.

СПИСОК ЛИТЕРАТУРЫ

1. *Almon S.* The distributed lag between capital appropriations and expenditures // Econometrica : Jan. 1965. Vol. 33. Is. 1. P. 178-196.

2. *Shrestha M. B., Bhatta G. R.,* Selecting appropriate methodological framework for time series data analysis // The Journal of Finance and Data Science. 2018. Vol. 4. I. 2. P. 71-89.

3. *Harper C. P.* Testing for the Existence of a Lagged Relationship within Almon's Method

// The Review of Economics and Statistics. 1977. Vol. 59 (2). P. 204-210.

4. *Wahba G.* Estimation of the Coefficients in a Multidimensional Distributed Lag Model // Econometrica. 1969. Vol. 37. No. 3. P. 398-407.

5. *Waud R. N.* Almon Lag. In: The New Palgrave Dictionary of Economics // The New Palgrave Dictionary of Economics. 2018. P. 259-260. [Electronic resource]. URL: https://doi.org/10.1057/978-1-349-95189-5_385 (date of application: 13.09.2023).