ПРОГНОЗ ПОЯВЛЕНИЯ НОВЫХ СИЛЬНЫХ СЕКТОРОВ В ЭКОНОМИКЕ РЕГИОНА

Кудров А.В., Афанасьев М.Ю. (Москва)ⁱ

Рассматривается задача прогнозирования появления в регионе новых сильных секторов. В настоящее время достаточно широко используются алгоритмы формирования рекомендаций при выборе (например, см. (Glodberg, Roeder, Gupta, Perkins, 2001)). Весьма общая постановка этой задачи состоит в рекомендации тех вариантов выбора, которые соответствуют в наибольшей степени характеристикам объекта, для которого вырабатываются рекомендации. Для структуры сильных секторов экономики региона рекомендации выдают *список потенциально достижимых сильных секторов при заданных характеристиках региона*. Провидимые исследования ориентированы на развитие теорий локализованной специализации и диверсификации экономики.

Определим показатель RCA_{cp} выявленных сравнительных преимуществ:

$$RCA_{cp} = \frac{y_{cp}/\sum_p y_{cp}}{\sum_c y_{cp}/\sum_{c,p} y_{cp}}, \quad (1)$$

где y_{cp} - объем производства сектора p экономики региона c . Показатель RCA_{cp} представляет собой отношение доли производства от сектора p в общем объеме производства от всех секторов экономики региона c к доле производства сектора p по всем регионам в объеме производства от всех секторов экономики всех регионов. В соответствии с работой (Hausmann, Klinger, 2006), для выявления сравнительных преимуществ в экономиках используется показатель RCA_{cp} , для которого проверяется условие типа ограничения снизу:

$$a_{c,p} = \begin{cases} 1, & \text{если } RCA_{cp} \geq 1 \\ 0, & \text{если } RCA_{cp} < 1 \end{cases}$$

Матрица $A = (a_{c,p})$ содержит данные о секторах экономики, которые в разных регионах развиты на уровне выявленных сравнительных преимуществ, определенных при помощи выражения (1). Строки этой матрицы соответствуют регионам, столбцы — секторам экономики. Далее будем называть вектор $(a_{c,p_1}, \dots, a_{c,p_m})$ *структурой сильных секторов* экономики региона c .

Пусть $R_1 = (r_{c,p})$ – матрица с элементами $r_{c,p} = \frac{a_{c,p}}{k_{c,0}}$, где $k_{c,0} = \sum_p a_{c,p}$.

Пусть $R_2 = (r_{p,c}^*)$ – матрица с элементами $r_{p,c}^* = \frac{a_{c,p}}{k_{p,0}}$, где $k_{p,0} = \sum_c a_{c,p}$.

В соответствии с (Afanasiev, Kudrov, 2020; Hausmann et al., 2011) экономическая сложность региона определяется как собственный вектор матрицы $R_1 R_2$. Заметим, что

элемент $w_{i,j}$ на пересечении i -ой строки и j -го столбца матрицы $R_1 R_2$, то есть $(R_1 R_2)_{ij}$, дается формулой:

$$\frac{1}{k_{c_i,0}} \sum_t \frac{a_{c_i,p_t} a_{c_j,p_t}}{k_{p_t,0}}.$$

Интерпретация элементов матрицы $(w_{i,j})$ как характеристик *степени вложенности* множества сильных секторов региона c_i в множество сильных секторов для региона c_j , а также способ их оценки на основе данных о налоговых поступлениях по 82 секторам экономики, представлены в работе (Афанасьев, Кудров, 2021).

Пусть случайные величины

$$A_{c_i,p_k} = \begin{cases} 1, & \text{с вероятностью } \pi(f_{c_i,p_k}^1, f_{c_i,p_k}^2, \dots) \\ 0, & \text{с вероятностью } 1 - \pi(f_{c_i,p_k}^1, f_{c_i,p_k}^2, \dots) \end{cases}, \quad (2)$$

где $f_{c_i,p_k}^1, f_{c_i,p_k}^2, \dots$ – факторы, влияющие на возникновение событий $A_{c_i,p_k} = 1$ и $A_{c_i,p_k} = 0$. Далее мы будем предполагать, что зависимость $\pi(f_{c_i,p_k}^1, f_{c_i,p_k}^2, \dots)$ имеет вид:

$$\pi(f_{c_i,p_k}^1, f_{c_i,p_k}^2, \dots) = \frac{e^{\beta_0 + \sum \beta_h f_{c_i,p_k}^h}}{1 + e^{\beta_0 + \sum \beta_h f_{c_i,p_k}^h}}, \text{ где } \beta_0 - \text{константа}. \quad (3)$$

Предположим, что последовательность $a_{c_1,p_1}, \dots, a_{c_1,p_m}, \dots, a_{c_n,p_1}, \dots, a_{c_n,p_m}$ является реализацией случайной последовательности $A_{c_1,p_1}, \dots, A_{c_1,p_m}, \dots, A_{c_n,p_1}, \dots, A_{c_n,p_m}$, элементы которой предполагаются независимыми при заданных $f_{c_1,p_1}^1, f_{c_1,p_2}^1, \dots$. Кроме того, предположим, что распределение $P(A_{c_i,p_k} | f_{c_1,p_1}^1, f_{c_1,p_2}^1, \dots) = P(A_{c_i,p_k} | f_{c_i,p_k}^1, \dots, f_{c_i,p_k}^h, \dots)$, то есть факторы, влияющие на распределение A_{c_i,p_k} , ограничиваются множеством $f_{c_i,p_k}^1, \dots, f_{c_i,p_k}^h, \dots$, у которых нижняя пара индексов (c_i, p_k) . Тогда функция правдоподобия имеет вид:

$$\begin{aligned} P(A_{c_1,p_1} = a_{c_1,p_1}, \dots, A_{c_1,p_m} = a_{c_1,p_m}, \dots, A_{c_n,p_1} = a_{c_n,p_1}, \dots, A_{c_n,p_m} = a_{c_n,p_m} | f_{c_1,p_1}^1, f_{c_1,p_2}^1, \dots) = \\ = P(A_{c_1,p_1} = a_{c_1,p_1} | f_{c_1,p_1}^1, \dots, f_{c_1,p_1}^h, \dots) \cdot \dots \\ \cdot P(A_{c_n,p_m} = a_{c_n,p_m} | f_{c_n,p_m}^1, \dots, f_{c_n,p_m}^h, \dots) = \\ = \frac{e^{a_{c_1,p_1}(\beta_0 + \sum \beta_h f_{c_1,p_1}^h)}}{(1 + e^{\beta_0 + \sum \beta_h f_{c_1,p_1}^h})} \cdot \dots \cdot \frac{e^{a_{c_n,p_m}(\beta_0 + \sum \beta_h f_{c_n,p_m}^h)}}{(1 + e^{\beta_0 + \sum \beta_h f_{c_n,p_m}^h})}. \end{aligned}$$

Максимизируя эту функцию правдоподобия, мы получаем оценки параметров $\beta_1, \dots, \beta_h, \dots$. Для оценки качества модели мы будем использовать критерий Акаике, а также величину:

$$G = \frac{1}{mn} \sum_{i,k} I(I(\pi(f_{c_i,p_k}^1, f_{c_i,p_k}^2, \dots) > 0.5) = a_{c_i,p_k}),$$

где $I(A)$ – индикатор события A . Величина G характеризует долю правильно идентифицированных моделью случаев.

Рассмотрим способы формирования объясняющих переменных $f_{c_i,p_k}^1, \dots, f_{c_i,p_k}^h, \dots$.

Способ 1: для данного способа $h = 2$. Определим факторы $f_{c_i,p_k}^1, f_{c_i,p_k}^2$ для всех $i = 1, \dots, n; k = 1, \dots, m$:

$$f_{c_i,p_k}^1 = \frac{\sum_{j=1, i \neq j}^n w_{i,j} a_{c_j,p_k}}{1 - w_{i,i}};$$

$$f_{c_i,p_k}^2 = \frac{\sum_{j=1, i \neq j}^n \frac{w_{j,i}}{w_{j,j}} a_{c_j,p_k}}{\sum_{j=1, i \neq j}^n \frac{w_{j,i}}{w_{j,j}}}.$$

Оценки параметров модели (2) - (3) при данном способе формирования объясняющих факторов представлены в таблице 1.

Таблица 1. Оценка параметров модели идентификации сильных секторов для способа 1

	<i>Estimate</i>	<i>Standard deviation</i>	<i>t-statistics</i>	<i>p-value</i>
Constant	-2,94	0,07	-43,42	0,00
f_{c_i,p_k}^1	2,91	0,83	3,52	0,00
f_{c_i,p_k}^2	18,15	0,91	20,12	0,00
G	0,8			
AIC	5681			

Как видно из таблицы 1, коэффициенты при f_{c_i,p_k}^1 и f_{c_i,p_k}^2 являются значимыми и положительными. Для этого способа выбора объясняющих переменных доля правильно идентифицированных моделью случаев составляет 80%.

Методом имитации проведена проверка гипотезы равенстве нулю показателя вложенности структуры сильных секторов региона $c_i, i = 1, \dots, n$, в структуру сильных секторов региона $c_j, j = 1, \dots, n, j \neq i$, а именно:

$$H_0: \frac{w_{i,j}}{w_{i,i}} = 0 \text{ против } H_1: \frac{w_{i,j}}{w_{i,i}} \neq 0.$$

Показано, что 95%-квантиль для функции распределения $\frac{w_{i,j}}{w_{i,i}}$ в условиях справедливости нулевой гипотезы равен 0,47 при заданном уровне значимости $\alpha = 0,95$.

Способ 2: для данного способа $h = 2$. Определим факторы $f_{c_i,p_k}^1, f_{c_i,p_k}^2$ для всех $i = 1, \dots, n; k = 1, \dots, m$:

$$f_{c_i,p_k}^1 = \frac{\sum_{j=1, i \neq j}^n I\left(\frac{w_{i,j}}{w_{i,i}} > 0.47\right) w_{i,j} a_{c_j,p_k}}{\sum_{j=1, i \neq j}^n I\left(\frac{w_{i,j}}{w_{i,i}} > 0.47\right) w_{i,j}};$$

$$f_{c_i, p_k}^2 = \frac{\sum_{j=1, i \neq j}^n I\left(\frac{w_{j,i}}{w_{j,j}} > 0.47\right) \frac{w_{j,i}}{w_{j,j}} a_{c_j, p_k}}{\sum_{j=1, i \neq j}^n I\left(\frac{w_{j,i}}{w_{j,j}} > 0.47\right) \frac{w_{j,i}}{w_{j,j}}}.$$

Оценки параметров модели (2) - (3) при данном способе формирования объясняющих факторов представлены в таблице 2.

Таблица 2. Оценка параметров модели идентификации сильных секторов для способа 2

	Оценка	Стандартное отклонение	t-статистика	p-значение
Константа	-3,03	0,07	-43,33	0,00
f_{c_i, p_k}^1	4,16	0,25	16,62	0,00
f_{c_i, p_k}^2	6,70	0,28	24,26	0,00
G	0,83			
AIC	4950			

Как видно из Табл. 2, коэффициенты при f_{c_i, p_k}^1 и f_{c_i, p_k}^2 являются значимыми и положительными. Для этого способа выбора объясняющих переменных доля правильно идентифицированных моделью случаев незначительно отличается от доли для случая 2 и равна 83% от общего числа случаев, но значение критерия Акаике меньше, чем для первого способа выбора объясняющих переменных. Способ 2 предпочтительнее.

В результате использования представленной выше модели были сформированы списки потенциально сильных секторов в регионах. Например, для ряда регионов, специализирующихся в сельском хозяйстве, требуется развитие сектора “Транспортировка и хранение”, что даст возможность большего территориального охвата в снабжении производимой продукцией. Вследствие инфраструктурных проблем рекомендации для развития сельского хозяйства состоят также в развитии сектора “Забор, очистка и распределение воды”, что необходимо для функционирования поливочных систем, а также сектора “Обеспечение электрической энергией, газом и паром; кондиционирование воздуха”, что также требуется для развития тепличных комплексов, требующих больших энергозатрат при создании микроклимата. Следует также заметить, что развитие сельского хозяйства способствует формированию спроса на развитие сектора сельхозмашиностроения, а также многих других секторов. В таблице 3 в качестве примера приведены оценки вероятности появления в регионах сектора «транспортировка и хранение» в качестве сильного. Спектр выявленных потенциальных сильных секторов в структурах региональных экономик по результатам использования модели является более широким.

Таблица 3. Оценки вероятности возникновения сектора «транспортировка и хранение» в качестве сильного

Регион	Оценка вероятности
Белгородская область	0,89
Воронежская область	0,63
Курская область	0,57
Московская область	0,67
г. Санкт-Петербург	0,92
Республика Адыгея	0,91

На основе упорядочения секторов по значению этих вероятностей и оценок их потенциального вклада в социально-экономическое развитие, может быть дана экспертная оценка целесообразности развития в регионе нового сильного сектора.

Литература

- Афанасьев М.Ю., Кудров А.В.* О свойствах вложенности структур сильных секторов региональных экономик. // Многомерный статистический анализ, эконометрика и моделирование реальных процессов. Труды X-й Международной школы-семинара. Под ред. В.Л. Макарова. – М.: ЦЭМИ РАН, 2021.
- Afanasiev, M.Yu., & Kudrov, A.V.* Estimates of economic complexity in the structure of the regional economy. *Montenegrin Journal of Economics*. 2020. 16(4), 43–53. doi: 10.14254/1800-5845/2020.16-4.4
- Glodberg, K., Roeder, T., Gupta, D., & Perkins, C.* Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*. 2001. 4. Pp. 133–151.
- Hausmann, R., Hidalgo, C., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A.* The Atlas of Economic Complexity. Mapping Paths to Prosperity. Cambridge, MA: Center for International Development, Harvard University, Harvard Kennedy School, Macro Connections, Massachusetts Institute of Technology. 2011.
- Hausmann, R., & Klinger, B.* Structural Transformation and Patterns of Comparative Advantage in the Product Space. *SSRN Electronic Journal*. 2006. doi: 10.2139/ssrn.939646.

¹ Кудров Александр Владимирович – ЦЭМИ РАН, kovlal@inbox.ru
Афанасьев Михаил Юрьевич – ЦЭМИ РАН, miafan@cemi.rssi.ru