

УДК 519.862.7

## Численное моделирование динамики потока научных и научно-педагогических кадров на основе статистических данных по МГУ им. Н.П.Огарева

© П. А. Шаманаев<sup>1</sup>, В.А. Атряхин<sup>2</sup>

**Аннотация.** В работе описывается процесс численного моделирования динамики потока научных и научно-педагогических кадров. В качестве математической модели используется система одновременных эконометрических уравнений. Для устранения мультиколлинеарности используется метод последовательного присоединения регрессоров. Для отыскания оценок неизвестных параметров системы используется двухшаговый метод наименьших квадратов.

**Ключевые слова:** система одновременных эконометрических уравнений, двухшаговый метод наименьших квадратов, матрица корреляции, устранение мультиколлинеарности, метод последовательного присоединения регрессоров.

### 1. Введение

Воспроизводство научных и научно-педагогических кадров по приоритетным отраслям развития науки, техники и технологий является одной из актуальных задач системы высшего профессионального образования. В связи с этим появляется необходимость в разработке и апробации моделей и методик среднесрочного и долгосрочного прогнозирования потока кадров высшей квалификации для воспроизводства научно-педагогического потенциала высшей школы в субъектах Российской Федерации, и, в частности, в Приволжском федеральном округе.

В работе для исследования задачи прогнозирования динамики потока научных и научно-педагогических кадров высшей квалификации применяется математическая модель в виде одновременных эконометрических уравнений [1].

### 2. Постановка задачи

Обозначим через  $x_j(k)$  - число выпускников высшего профессионального образования в  $k$  - ом году по  $j$  - ой укрупненной группе специальностей высшего профессионального образования,  $j = \overline{1, m}$ . Обозначим через  $y_i(k)$  - число поступивших в аспирантуру в  $k$  - ом году по  $i$  - ой укрупненной группе специальностей высшей научной классификации,  $i = \overline{1, n}$ . В качестве математической модели, прогнозирующей поток научных и научно-педагогических кадров, возьмем систему эконометрических одновременных уравнений вида [2]:

$$y(k) = By(k) + Ax(k) + \varepsilon, \quad (2.1)$$

где  $y(k) = \text{column}(y_1(k), y_2(k), \dots, y_n(k))$ ,  $x(k) \equiv \text{column}(x_0(k), x_1(k), x_2(k), \dots, x_m(k))$ ,  $x_0(k) = 1$ ,  $k = 1, 2, \dots$ ;  $B \equiv [b_{ij}]$  - постоянная  $(n \times n)$  - матрица с нулевыми диагональными

<sup>1</sup>Заведующий кафедрой прикладной математики, Мордовский государственный университет им. Н. П. Огарева, г. Саранск; korspra@yandex.ru.

<sup>2</sup>Аспирант кафедры прикладной математики, Мордовский государственный университет им. Н. П. Огарева, г. Саранск; atrvol@rambler.ru.

элементами;  $A \equiv [a_{ij}]$  - постоянная  $(n \times (m+1))$  - матрица;  $\varepsilon = \text{column}(\varepsilon(1), \varepsilon(2), \dots, \varepsilon(n))$  - вектор случайных возмущений. Предположим, что  $\varepsilon$  имеет многомерное нормальное распределение с математическим ожиданием равным нулю и матрицей ковариации  $\Omega = \sigma^2 E_n$ ,  $E_n$  - единичная матрица размерности  $n$ .

Ставится задача нахождения статистических оценок неизвестных элементов матриц  $A$  и  $B$ . Так как система уравнений (2.1) представляет собой сверхидентифицированную систему эконометрических взаимосвязанных уравнений, то для нахождения оценок неизвестных параметров применяется двухшаговый метод наименьших квадратов [3], [4].

Входными данными для проведения вычислений является следующая статистическая информация:

1. Статистическая информация по количеству выпускников поступивших в аспирантуру в разрезе укрупненных групп специальностей высшей научной классификации [5].
2. Статистическая информация по количеству выпускников в разрезе укрупненных групп специальностей высшего профессионального образования [6].
3. Таблица соответствия укрупнённых групп специальностей высшего профессионального образования укрупнённым группам специальностей высшей научной классификации. Она состоит из 28 столбцов (по числу укрупнённых групп специальностей высшего профессионального образования) и 18 строк (по числу укрупнённых групп специальностей высшей научной классификации). На пересечении каждого столбца и каждой строки либо «0», либо «1». Данные значения проставляются на основании анализа статистической информации за годы, предшествующие прогнозируемому периоду. Если в течение этих лет были студенты, окончившие высшее учебное заведение по специальности, относящейся к  $i$ -ой укрупнённой группе специальностей высшего профессионального образования, и поступившие в аспирантуру по специальности, относящейся  $j$ -ой укрупнённой группе специальностей высшей научной классификации, то на пересечении  $i$ -того столбца и  $j$ -ой строки будет стоять «1». В противном случае - «0».
4. Таблица соответствия укрупнённых групп специальностей высшей научной классификации укрупнённым группам специальностей высшей научной классификации. Она состоит из 18 строк и 18 столбцов (по числу укрупнённых групп специальностей высшей научной классификации). На пересечении каждого столбца и каждой строки либо «1», либо «0». Данные значения проставляются на основании экспертных оценок и анализа статистических данных за годы, предшествующие прогнозируемому периоду. Если студенты какой-либо специальности высшего профессионального образования за эти годы поступали в аспирантуру на специальности, относящиеся к  $i$ -ой и к  $j$ -ой укрупнённой группе специальностей высшей научной классификации, то на пересечении  $i$ -того столбца и  $j$ -ой строки будет стоять «1».

### 3. Апробация математической модели на основании статистических данных по МГУ им. Н. П. Огарёва

Согласно статистическим данным, представленным учебно-методическим управлением МГУ им. Н. П. Огарёва, на сегодняшний день обучение в ВУЗе ведется по 18 укрупненным группам специальностей высшего профессионального образования и по 17 укрупнен-

ным группам специальностей высшей научной классификации. Как следствие, выбранная математическая модель (2.1) характеризуется большим числом экзогенных переменных. Следствием этого стало появление мультиколлинеарности. Для устранения мультиколлинеарности использовался метод последовательного присоединения регрессоров [3].

Проиллюстрируем использование метода последовательного присоединения регрессоров на примере второго этапа построения прогноза двухшаговым методом наименьших квадратов по первой укрупненной группе специальностей высшей научной классификации «01.00.00 Физико-математические науки».

На основании матриц соответствия выделяем экзогенные переменные, от которых зависит исследуемый фактор. Статистическая информация, необходимая для расчетов, представлена на рисунке 3.1.

	Физико-математические науки	Химические науки	Биологические науки	Технические науки	Экономические науки	Педагогические науки	Науки о Земле	Физико-математические науки	Электронная техника радиотехника и связь	Информатика и вычислительная техника
Годы	Y1	Y2	Y3	Y4	Y7	Y11	Y18	X1	X21	X23
2005	8	1	3	34	9	2	3	148	227	34
2006	7	1	3	30	16	1	3	181	272	41
2007	7	1	3	22	10	3	4	180	195	34
2008	13	1	5	22	23	1	3	164	212	57
2009	21	5	8	26	23	0	7	127	162	47

Р и с у н о к 3.1

Численность выпускников в разрезе УГС ВПО ( $x_1, x_{21}, x_{23}$ ) и аспирантов в разрезе УГС ВНК ( $y_1, y_2, y_3, y_4, y_7, y_{11}, y_{18}$ )

Первый регрессор выбирается на основе корреляционной матрицы. Корреляционная матрица представлена на рисунке 3.2.

	Y1	Y2	Y3	Y4	Y7	Y11	Y18	X1	X21	X23
Y	1,00									
Y2	0,32	1,00								
Y3	0,91	0,44	1,00							
Y4	0,32	1,00	0,45	1,00						
Y7	0,79	0,16	0,85	0,16	1,00					
Y11	0,31	0,77	0,49	0,79	0,00	1,00				
Y18	0,66	0,22	0,76	0,24	0,39	0,68	1,00			
X1	0,86	0,52	0,79	0,53	0,42	0,69	0,82	1,00		
X21	0,77	0,69	0,67	0,68	0,63	0,30	0,19	0,64	1,00	
X23	0,62	0,44	0,57	0,44	0,80	0,38	0,35	0,21	0,19	1,00

Р и с у н о к 3.2

Корреляционная матрица

Наибольшую корреляцию эндогенная переменная  $y_1$  имеет с переменной  $y_3$ . Следовательно в модель необходимо включить регрессор  $y_3$ . Для выбора второго регрессора рассмотрим следующий набор моделей:

$$y_1 = a_{10} + b_{13}y_3 + c_{1i}z_i + \nu, \quad (3.1)$$

где  $i \in \{1, 2, 4, 7, 11, 18, 21, 23\}$ . Если  $i \in \{2, 4, 7, 11, 18\}$ , то  $c_{1i} = b_{1i}$  и  $z_i = y_i$ . Если  $i \in \{1, 21, 23\}$ , то  $c_{1i} = a_{1i}$  и  $z_i = x_i$ .

Для каждой из моделей вида (3.1) рассчитывается коэффициент детерминации  $R^2$ . Результаты вычислений приведены на рисунке 3.3.

	Y3, Y2	Y3, Y4	Y3, Y7	Y3, Y11	Y3, Y18	Y3, X1	Y3, X21	Y3, X23
R <sup>2</sup>	0,9785	0,9783	0,9794	0,9820	0,9807	0,9824	0,9885	0,9799

Р и с у н о к 3.3

Коэффициенты детерминации для соответствующих моделей вида (3.1)

Наибольший коэффициент детерминации получился в модели:

$$y_1 = a_{10} + b_{13}y_3 + a_{1,21}x_{21} + \nu, \quad (3.2)$$

Для выбора третьего регрессора рассмотрим следующий набор моделей:

$$y_1 = a_{10} + b_{13}y_3 + a_{1,21}x_{21} + c_{1i}z_i + \nu, \quad (3.3)$$

где  $i \in \{1, 2, 4, 7, 11, 18, 23\}$ . Если  $i \in \{2, 4, 7, 11, 18\}$ , то  $c_{1i} = b_{1i}$  и  $z_i = y_i$ . Если  $i \in \{1, 23\}$ , то  $c_{1i} = a_{1i}$  и  $z_i = x_i$ .

Для каждой из моделей вида (3.3) рассчитывается коэффициент детерминации  $R^2$ . Результаты вычислений приведены на рисунке 3.4.

	Y3, X21, Y2	Y3, X21, Y4	Y3, X21, Y7	Y3, X21, Y11	Y3, X21, Y18	Y3, X21, X1	Y3, X21, X23
R <sup>2</sup>	0,9965	0,9885	0,9888	0,9917	0,9891	0,9900	0,9940

Р и с у н о к 3.4

Коэффициенты детерминации для соответствующих моделей вида (3.3)

Наибольший коэффициент детерминации получился в модели:

$$y_1 = a_{10} + b_{12}y_2 + b_{13}y_3 + a_{1,21}x_{21} + \nu, \quad (3.4)$$

Берем её за основу и продолжаем присоединение регрессоров. Для выбора четвертого регрессора рассмотрим следующий набор моделей:

$$y_1 = a_{10} + b_{12}y_2 + b_{13}y_3 + a_{1,21}x_{21} + c_{1i}z_i + \nu, \quad (3.5)$$

где  $i \in \{1, 4, 7, 11, 18, 23\}$ . Если  $i \in \{4, 7, 11, 18\}$ , то  $c_{1i} = b_{1i}$  и  $z_i = y_i$ . Если  $i \in \{1, 23\}$ , то  $c_{1i} = a_{1i}$  и  $z_i = x_i$ .

Для каждой из моделей вида (3.5) рассчитывается коэффициент детерминации  $R^2$ . Результаты вычислений приведены на рисунке 3.5.

	Y2, Y3, X21, Y4	Y2, Y3, X21, Y7	Y2, Y3, X21, Y11	Y2, Y3, X21, Y18	Y2, Y3, X21, X1	Y2, Y3, X21, X23
R <sup>2</sup>	0,9964	0,9993	0,9939	0,9990	0,9999	0,9986

Р и с у н о к 3.5

Коэффициенты детерминации для соответствующих моделей вида (3.5)

Наибольший коэффициент детерминации получился в модели:

$$y_1 = a_{10} + b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{1,21}x_{21} + \nu, \quad (3.6)$$

Присоединение регрессоров продолжается до тех пор, пока наибольший коэффициент детерминации на очередном шаге не окажется меньше, чем коэффициент детерминации у выбранной на предыдущем шаге модели, или пока количество регрессоров не достигнет количества лет, за которые доступна статистическая информация.

Для построения прогноза по первой укрупненной группе специальностей высшей научной классификации «01.00.00 Физико-математические науки» итоговый набор регрессоров представлен на рисунке 3.6.

	Химические науки	Биологические науки	Физико-математические науки	Электронная техника радиотехника и связь
Годы	Y2	Y3	X1	X21
2005	1,800	3,047	148	227
2006	1,000	3,387	181	272
2007	2,600	2,775	180	195
2008	0,200	4,742	164	212
2009	3,400	8,049	127	162

Р и с у н о к 3.6

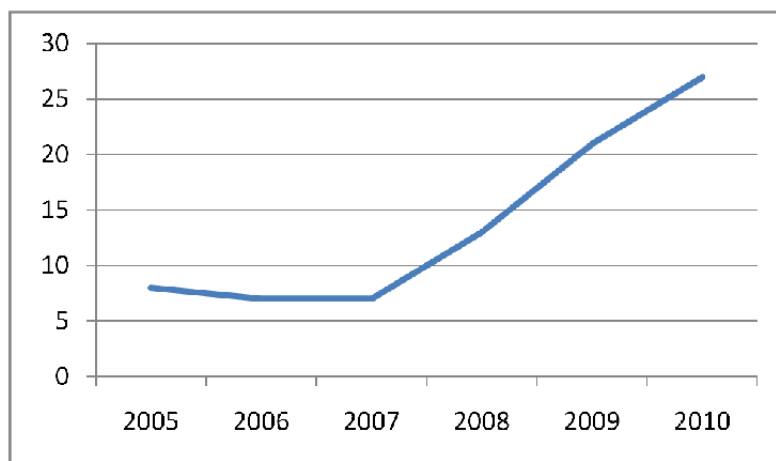
Итоговый набор регрессоров

После применения второго шага ДМНК к уравнению (3.6) получены следующие оценки неизвестных параметров, входящих в это уравнение:

$$\hat{b}_{12} = -0,67, \quad \hat{b}_{13} = 2,29, \quad \hat{a}_{10} = 13,42, \quad \hat{a}_{11} = -0,03, \quad \hat{a}_{1,21} = -0,03. \quad (3.7)$$

Эти оценки берутся в качестве значений соответствующих элементов первых строк матриц  $B$  и  $A$ . Остальные элементы первых строк матриц  $B$  и  $A$  полагаем равными нулю.

Согласно предложенной модели прогнозируемое значение на 2010 год  $y_1 = 27$ . Динамика численности поступающих в аспирантуру по укрупненной группе специальностей высшей научной классификации «01.00.00 Физико-математические науки» с 2005 по 2010 год приведена на рисунке 3.7.



Р и с у н о к 3.7

Динамика численности поступающих в аспирантуру по УГС ВНК «01.00.00 Физико-математические науки»

Для каждого уравнения системы (2.1) оценки неизвестных параметров находятся аналогичным образом.

#### 4. Проверка адекватности предложенной модели

Приведем анализ результатов для укрупненной группы специальностей высшей научной классификации «01.00.00 Физико-математические науки».

Проверка адекватности построенной регрессионной модели сводится к проверке гипотезы о значимости множественного коэффициента детерминации с помощью F-критерия Фишера. Коэффициент детерминации для регрессионной модели (3.6) примет максимальное значение равно 0,9999. Расчеты показывают, что при уровне значимости  $\alpha = 0,05$  уравнение множественной регрессии (3.6) значимо.

Гипотеза о значимости частных коэффициентов корреляции проверялась с помощью t-критерия Стьюдента. Расчеты показывают, что с вероятностью 0,95 основную гипотезу о незначимости частного коэффициента корреляции можно отклонить.

Остатки регрессионной модели для укрупненной группы специальностей высшей научной классификации «01.00.00 Физико-математические науки» приведены на рисунке 4.1.

e1	e2	e3	e4	e5
-0,000000794	0,000000819	0,000000066	-0,000000378	0,000000153

Р и с у н о к 4.1

Остатки регрессионной модели

Согласно критерию Дарбина-Уотсона автокорреляция остатков регрессионной модели отсутствует.

На основании корреляционной матрицы, построенной по исходным статистическим данным и остаткам регрессионной модели, можно сделать вывод о том, что корреляция между факторами и остатками  $e_i$  отсутствует.

Работа выполнена при поддержке аналитической ведущей целевой программы «Развитие научного потенциала высшей школы (2010-2011 года)»

## СПИСОК ЛИТЕРАТУРЫ

1. Haavelmo T. The statistical implications of a system of simultaneous equations // *Econometrica* 1943a. Vol. 11. N 1, pp. 1-12.
2. Шаманаев П.А., Атряхин В.А. Построение математической модели прогнозирования потока научных и научно-педагогических кадров - Саранск: Журнал Средневолжского математического общества, Том 12, №4, с. 234-238.
3. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики: Учебник для вузов. - М.: ЮНИТИ, 1998. - 1022 с.
4. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. - М.: Дело, 2004. - 576 с.
5. Общероссийский классификатор специальностей по образованию ОК 009-2003. Постановление Госстандарта России от 30.09.2003. № 276-ст. № ОК 009-2003.
6. Номенклатура специальностей научных работников [Электронный ресурс]// Приложение к Приказу Минобрнауки РФ от 25.02.2009 № 59 «Об утверждении номенклатуры специальностей научных работников» (в ред. Приказа Минобрнауки РФ от 11.08.2009 № 294). Режим доступа: - [http://vak.ed.gov.ru/common/img/uploaded/files/vak/norm\\_doc/2010/Prilozhenie\\_k\\_prikazu\\_N59\\_ot\\_25.02.2009.doc](http://vak.ed.gov.ru/common/img/uploaded/files/vak/norm_doc/2010/Prilozhenie_k_prikazu_N59_ot_25.02.2009.doc)

## Numerical simulation of flow dynamics scientific and teaching staff on the basis of statistical Data on Mordovian State University after N. P. Ogarev

© P.A. Shamanaev<sup>3</sup>, V.A. Atryahin<sup>4</sup>

**Abstract.** The paper describes the process of numerical simulation of flow dynamics of scientific and pedagogical staff. As a mathematical model uses a system of simultaneous econometric equations. To eliminate multicollinearity using the method of sequential addition of covariates. To find the estimates of unknown parameters the system uses a two-step method of least squares.

**Key Words:** system of simultaneous econometric equations, two-stage least squares method, correlation matrix, the elimination of multicollinearity, the method of sequential addition of covariates.

<sup>3</sup>Chief of Applied Mathematics Chair, Mordovian State University after N. P. Ogarev, Saransk; korspa@yandex.ru.

<sup>4</sup>Postgraduate student of Applied Mathematics Chair, Mordovian State University after N. P. Ogarev, Saransk; atrvol@rambler.ru.