

ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ ДЛЯ РЕЗУЛЬТАТОВ ЕДИНЫХ ГОСУДАРСТВЕННЫХ ЭКЗАМЕНОВ

В. А. Силаева, А. М. Силаев

Национальный исследовательский университет «Высшая школа экономики»,
Нижний Новгород

1. Введение

Единый государственный экзамен (ЕГЭ) в нашей стране с 2009 г. является основной формой обязательной итоговой аттестации программ среднего общего образования, а баллы ЕГЭ используются вузами для конкурсного отбора абитуриентов на различные образовательные программы. До 2016 г. официальный информационный портал ЕГЭ (ege.edu.ru) и Федеральный институт педагогических измерений (www.fipi.ru) публиковали некоторые статистические данные по результатам ЕГЭ в Российской Федерации. Например, на сайте информационного портала ЕГЭ (ege.edu.ru) публиковались средние тестовые баллы ЕГЭ для разных предметов. Кроме того, для отдельных лет в открытом доступе можно было найти количество участников ЕГЭ в разбивке по регионам, по предметам и по типам населенных пунктов, проценты участников ЕГЭ, набравших тот или иной тестовый балл по различным предметам, проценты участников ЕГЭ, не приступавших к выполнению заданий с развернутым ответом (часть С), количество 100 балльников по разным предметам и в разбивке по регионам, информацию об участии выпускников в ЕГЭ по нескольким предметам, а также некоторые другие статистические данные. По результатам ЕГЭ 2013 г. на сайте Федерального центра тестирования (www.rustest.ru) приведены гистограммы для баллов ЕГЭ по математике и русскому языку в различных субъектах Российской Федерации. К сожалению, после 2013 г. статистика баллов ЕГЭ стала труднодоступна, а с 2016 г. перестали публиковать в открытом доступе даже средние тестовые баллы ЕГЭ по различным предметам. Вместе с тем статистическая информация о результатах ЕГЭ может быть использована для оценки уровня образования выпускников школ по разным предметам, корректных выводов о привлекательности различных образовательных программ для абитуриентов и эффективности деятельности вузов. Анализ статистических характеристик баллов ЕГЭ актуален также потому, что баллы ЕГЭ используются в эконометрических исследованиях проблем экономики образования в нашей стране.

В статьях (Польдин, 2011; Пересецкий, Давтян, 2011; Замков, Пересецкий, 2013; Хавенсон, Соловьева, 2014) изучается связь ЕГЭ и успеваемость студентов на первых

курсах бакалавриата. В статье (Андрушак и др., 2012) исследуется проявление эффектов сообучения в студенческой группе, при этом результаты ЕГЭ по русскому языку и математике используются для оценки способностей студентов. В работах (Прахов, Юдкевич, 2012; Прахов, 2015) исследуется отдача от инвестиций в дополнительную подготовку школьников к поступлению в вуз в баллах ЕГЭ. Оценке результатов приема на образовательные программы в вузах по результатам ЕГЭ посвящены статьи (Польдин, Силаев, 2011; Польдин и др., 2014).

В настоящей работе рассматриваются эконометрические модели зависимости от различных факторов количества участников ЕГЭ, набравших различные баллы по отдельным предметам. По данным для отдельных регионов России исследуются зависимости количества хороших и плохих баллов ЕГЭ от числа учащихся N , принимавших участие в экзамене в различных регионах, а также от некоторых других влияющих переменных. Показывается, что, чем больше число участников экзамена в каждом регионе N , тем больше число как хороших, так и плохих оценок в регионе. Но, как правило, количество низких оценок в регионах растет немного медленнее, чем N , а количество хороших оценок в регионах растет быстрее, чем N .

2. Модели для количества участников ЕГЭ в регионах России, получивших 100 баллов

Рассмотрим зависимости количества участников, получивших 100 баллов по русскому языку, по математике и другим дисциплинам в 2013 г. от числа учащихся, принимавших участие в экзамене в различных регионах России. Каждая серия состоит из 83 значений по числу субъектов Российской Федерации в 2013 г. Методом наименьших квадратов (МНК) оценим коэффициенты в линейных регрессиях вида

$$y_i = aN_i + bN_i^2 + \varepsilon_i, \quad i = \overline{1, 83}, \quad (1)$$

где зависимая переменная y_i – число 100 балльников, а N_i – количество учащихся в тыс. чел., сдававших экзамен по соответствующему предмету в i -м регионе. Дисперсии случайных величин ε_i считаем равными σ^2 . Результаты оценивания коэффициентов регрессий приведены в таблице 1. В последнем столбце таблицы также даны значения общего количества 100 балльных оценок для каждого предмета в 2013 г. Отметим, что, если в модели регрессий (1) включать константы, то их оценки будут незначимо отличаются от нуля при 5% уровне значимости. Уровни значимости в таблице 1 получены с учетом оценок стандартных ошибок коэффициентов регрессии в форме Уайта. Нет оценок для коэффициентов моделей по немецкому, французскому и испанскому языкам

из-за малого количества 100 балльных оценок в 2013 г. (4 оценки по немецкому языку, 5 оценок по французскому языку и нет оценок по испанскому языку).

Таблица 1. Оценки коэффициентов регрессий (1) для 100 балльных оценок.

Дисциплина	\hat{a}	\hat{b}	$\hat{\sigma}$	Число 100 балльных оценок
Русский язык	1.143**	0.096***	24.14	2559
Математика	0.167	0.0252***	8.72	538
Физика	0.681	0.330**	6.60	474
Химия	30.656**	1.034	47.27	3220
Информатика и ИКТ	10.703***	-0.360	5.26	563
Биология	-0.329	0.831	12.54	466
История	2.318**	0.163	7.33	500
География	15.581**	-6.988**	6.45	193
Английский язык	5.271***	0.458***	4.41	581
Обществознание	0.750***	0.020**	7.40	500
Литература	8.574***	0.980*	7.21	457

Примечание: *, **, *** — коэффициенты значимы на 10%, 5%, 1%-ном уровне соответственно.

Как видно из таблицы 1, большинство коэффициентов моделей МНК значимо отличаются от нуля. При этом, как и следовало ожидать, оценки коэффициентов \hat{a} получились положительными практически для всех предметов (для математики, физики и биологии эти оценки незначимо отличаются от нуля при 10% уровне значимости). Это означает, что, с увеличением числа участников экзаменов в каждом регионе N растет число 100 балльных оценок. Интересным представляется то, что оценки коэффициентов \hat{b} для регрессоров N^2 в моделях (1) также получают положительными для большинства

предметов (для химии, информатики и ИКТ, биологии, истории оценка \hat{b} незначимо отличаются от нуля при 10% уровне значимости, а для географии оценка \hat{b} отрицательна). Это можно объяснить тем, что, чем больше учащихся и соответственно число жителей в регионе, тем, вероятно, больше хороших школ и учителей, более развита сеть Интернет, есть больший доступ к средствам связи и информации, то есть имеются неучтенные в регрессии факторы, которые положительно влияют на успеваемость.

В качестве примера на рис. 1, 2 приведены диаграммы рассеяния для числа учащихся, получивших 100 баллов по русскому языку $R100$ и обществознанию $O100$, в зависимости от количества учащихся N (в тыс. чел.), сдававших экзамены по этим предметам в 2013 г. в 83 субъектах Российской Федерации.

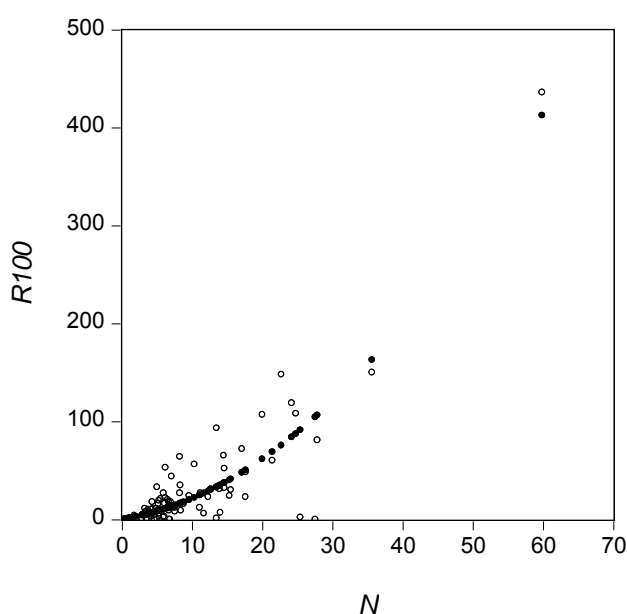


Рис. 1.

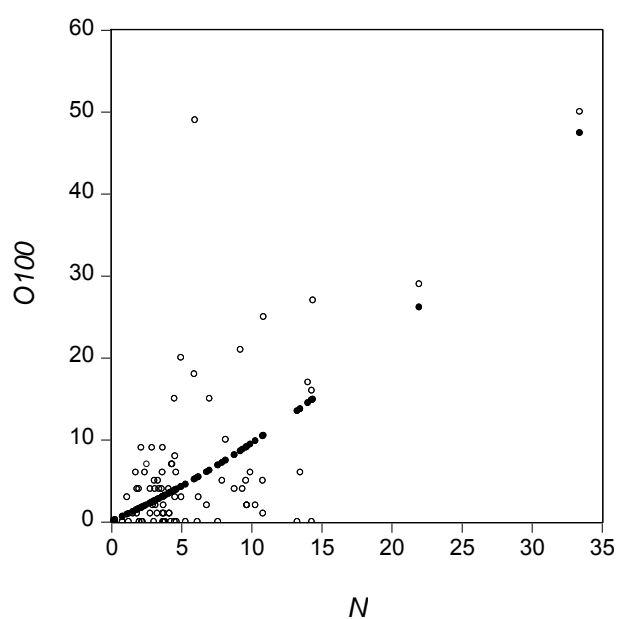


Рис. 2.

На этих же рисунках показаны точки, соответствующие зависимостям:

$$R100 = 1.144 N + 0.096 N^2; \quad O100 = 0.75 N + 0.02 N^2; \quad (2)$$

(0.479) (0.010) ; (0.226) (0.08) ;

найденным для моделей линейных регрессий с помощью метода наименьших квадратов (в скобках приведены значения стандартных ошибок для коэффициентов регрессий).

3. Модели для количества участников ЕГЭ в регионах России,

не приступавших к выполнению заданий с развернутым ответом (часть С)

Рассмотрим модели линейных регрессий вида (1), где зависимая переменная y_i – число плохих работ, а N_i – количество учащихся в тыс. чел., сдававших экзамен по соответствующему предмету в i -м регионе. К плохим отнесем работы, в которых участники ЕГЭ не приступали к выполнению заданий части С. Результаты МНК

оценивания коэффициентов регрессий приведены в таблице 2. В последнем столбце таблицы даны значения общего числа плохих работ для каждого предмета в 2013 г.

Таблица 2. Оценки коэффициентов регрессий (1) для плохих оценок.

Дисциплина	\hat{a}	\hat{b}	$\hat{\sigma}$	Число плохих работ
Русский язык	55.638***	-0.667***	748.90	33326
Математика	393.772***	-1.714***	813.98	289003
Физика	382.717***	-15.259***	293.22	64793
Химия	96.778***	-3.050	123.69	7791
Информатика и ИКТ	313.002***	-26.418***	107.68	15366
Биология	84.929***	-1.476	143.32	12300
История	77.454***	6.965	281.45	15965
География	-59.583	317.649***	40.68	3152
Английский язык	57.257***	-2.470***	47.58	3617
Немецкий язык	79.001***	-147.218***	1.842	178
Французский язык	39.547***	-58.115***	0.663	51
Обществознание	49.454***	-0.875**	372.63	17771
Литература	23.557***	-3.220**	19.56	795

Примечание: *, **, *** — коэффициенты значимы на 10%, 5%, 1%-ном уровне соответственно. Уровни значимости в таблице 2 получены с учетом оценок стандартных ошибок коэффициентов регрессии в форме Уайта. Нет оценок для коэффициентов модели по испанскому языкам из-за малого количества плохих работ, так как всего 4 участника не приступали к выполнению части С по испанскому языку в 2013 г.

Из таблицы 2 видно, что оценки \hat{a} снова получились положительными для всех предметов за исключением географии, для которой оценка \hat{a} незначимо отличается от

нуля при 10% уровне значимости. При этом оценки \hat{b} для коэффициентов при N^2 в моделях (1) получились отрицательными для большинства предметов (только для химии, биологии и истории оценки \hat{b} незначимо отличаются от нуля при 10% уровне значимости, а для географии оценка \hat{b} положительна). Положительность коэффициента a при N и отрицательность коэффициента b при N^2 в моделях (1) приводит к вогнутому виду зависимости количества плохих работ от числа участников экзамена в каждом регионе.

Аналогичное поведение прослеживается для результатов ЕГЭ за 2012 г. Для иллюстрации на рис. 3, 4 приведены диаграммы рассеяния для числа участников ЕГЭ по математике (переменная M_b в тыс. чел.) и по физике (переменная Ph_b в тыс. чел.), не приступавших к выполнению заданий с развернутым ответом (часть С), в зависимости от количества учащихся N (в тыс. чел.), сдававших экзамены по этим предметам в 2012 г. в 83 субъектах Российской Федерации. На этих же рисунках показаны точки, соответствующие вогнутым зависимостям, полученным с помощью МНК:

$$M_b = 0.372N - 0.001N^2; \quad Ph_b = 0.358N - 0.015N^2; \quad (3)$$

(0.016) (0.0003); \quad (0.020) (0.003)

в скобках приведены значения стандартных ошибок для коэффициентов регрессий.

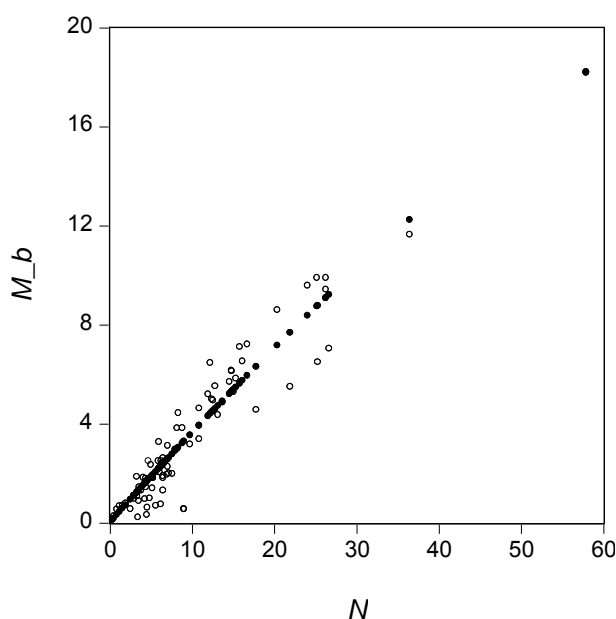


Рис. 3.

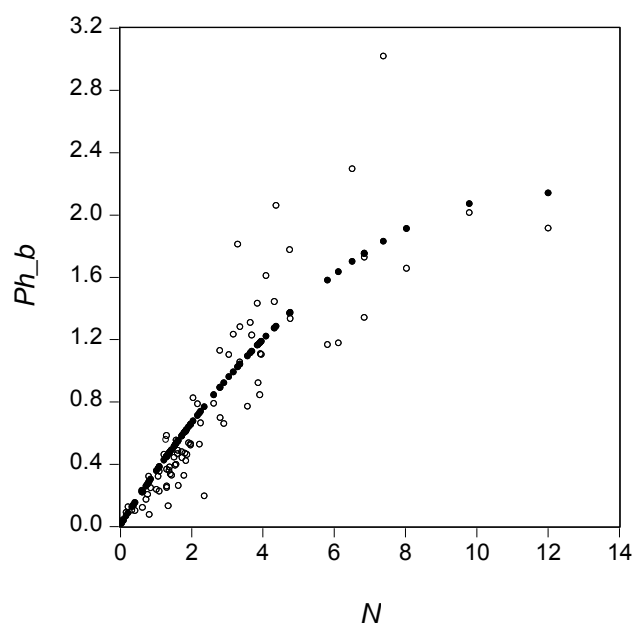


Рис. 4.

4. Учет факторов, положительно влияющих на оценки учащихся

Для поиска факторов, положительно влияющих на оценки учащихся, использовались данные из статистического сборника Росстата «Регионы России. Социально-экономические показатели. 2016», в котором приведены сведения о субъектах РФ в 2005 – 2015 гг. Некоторые переменные оказываются значимыми, если их включать в модели

линейной регрессии для количества 100 балльных работ или плохих работ, в которых участники ЕГЭ не приступали к выполнению заданий части С. В таблице 3 приведены МНК оценки коэффициентов линейных регрессий для ЕГЭ по русскому языку и математике в 2013 г. для моделей, когда в число регрессоров добавлены константы, количество общеобразовательных организаций (без вечерних общеобразовательных организаций), а также число образовательных организаций высшего образования и филиалов организаций высшего образования.

Таблица 3. Оценки коэффициентов регрессий для ЕГЭ по русскому языку и математике.

Переменная	100 баллов по русскому языку	Плохие работы по русскому языку	100 баллов по математике	Плохие работы по математике
Константа	4.49	-24.87	1.720	-199.94
Число участников экзамена в регионе (тыс. чел.), N	-0.873	144.31***	-0.197	462.46***
Квадрат числа участников экзамена в регионе (тыс. чел.), N^2	0.073***	0.296	0.0162*	-3.093***
Число общеобразовательных организаций в регионе	-0.009	-0.003	-0.0064	-0.984
Число вузов и филиалов вузов в регионе	0.842**	-33.637**	0.233*	8.456
Стандартная ошибка, $\hat{\sigma}$	21.92	636.24	8.23	791.17
Коэффициент детерминации, R^2	0.853	0.413	0.672	0.931

Примечание: *, **, *** — коэффициенты значимы на 10%, 5%, 1%-ном уровне соответственно. Уровни значимости получены с учетом оценок стандартных ошибок коэффициентов регрессии в форме Уайта.

Из таблицы 3 видно, что увеличение числа высших учебных заведений в регионах сопровождается ростом числа 100 балльных оценок как по русскому языку, так и по математике и снижением количества плохих работ по русскому языку. При этом число

общеобразовательных организаций оказывается незначимым фактором в рассматриваемых моделях.

5. Заключение

В работе рассмотрены эконометрические модели для количества 100 балльных работ ЕГЭ и плохих работ, в которых участники ЕГЭ не приступали к выполнению заданий с развернутым ответом (заданий части С), от различных влияющих переменных. Выявлено, что, для большинства предметов количество плохих работ в регионах растет немного медленнее, чем число участников экзамена в каждом регионе N , а количество 100 балльных работ в регионах растет быстрее, чем N . Дополнительным фактором, положительно влияющим на оценки учащихся, является число вузов и филиалов вузов в регионах.

Литература

1. Андрущак Г. В., Польшин О. В., Юдкевич М. М. (2012). Эффекты сообучения в административно формируемых студенческих группах. Прикладная эконометрика, 26 (2), 3–16.
2. Замков О. О., Пересецкий А. А. (2013). ЕГЭ и академические успехи студентов бакалавриата МИЭФ НИУ ВШЭ. Прикладная эконометрика, 30 (2), 93–114.
3. Пересецкий А. А., Давтян М. А. (2011). Эффективность ЕГЭ и олимпиад как инструмента отбора абитуриентов. Прикладная эконометрика, 23 (3), 41–56.
4. Польшин О. В. (2011). Прогнозирование успеваемости в вузе по результатам ЕГЭ. Прикладная эконометрика, № 1 (21), 56–69.
5. Польшин О. В., Силаев А. М. (2011). Сравнение образовательных программ по результатам ЕГЭ зачисленных студентов. Вопросы образования, 3, 192–209.
6. Польшин О. В., Силаева В. А., Силаев А. М. (2014). Сравнение приема на образовательные программы в вузе по результатам олимпиад и баллов ЕГЭ. Прикладная эконометрика, 36 (4), 118–132.
7. Прахов И. А., Юдкевич М. М. (2012). Влияние дохода на результаты ЕГЭ и выбор вуза. Вопросы образования, 1, 126–147.
8. Прахов И. А. (2015). Динамика инвестиций и отдача от дополнительной подготовки к поступлению в вуз. Прикладная эконометрика, 37 (1), 107–124.
9. Регионы России. Социально-экономические показатели. 2016: стат. сб. М.: Росстат, 2016. 1326 с.
10. Хавенсон Т. Е., Соловьева А. А. (2014). Связь результатов Единого государственного экзамена и успеваемости в вузе. Вопросы образования, 1, 176–199.