

О. В. Польшин, В. А. Силаева, А. М. Силаев

Сравнение приема на образовательные программы в вузе по результатам олимпиад и баллов ЕГЭ

В статье для сравнения качества приема в вузы на различные образовательные программы предлагается использовать скорректированные кривые спроса, полученные по результатам ЕГЭ зачисленных на бюджетные места студентов с учетом поступивших без вступительных экзаменов победителей и призеров олимпиад. Для устранения влияния структуры и количества предметов ЕГЭ на различные направления обучения предлагаются два метода коррекции баллов ЕГЭ. Первый метод предполагает линейную коррекцию суммарного балла ЕГЭ отдельного школьника с учетом средних значений и среднеквадратических отклонений баллов по отдельным предметам всех школьников в России. Во втором методе производится нелинейная коррекция баллов по каждому отдельному предмету в суммарном результате. При этом выбирается заданное стандартное распределение, по которому производится пересчет баллов ЕГЭ. В качестве такого стандартного распределения в статье используется бета-распределение.

Ключевые слова: единый государственный экзамен; кривые спроса; скорректированные баллы.

JEL classification: C18; I21.

1. Введение

С переходом на единый государственный экзамен (ЕГЭ) качество приема в вузы в последние годы принято оценивать по минимальным проходным (отсечным) и средним баллам студентов, зачисленных на первый курс по результатам ЕГЭ (Мониторинги НИУ ВШЭ, 2011–2013). Средний балл ЕГЭ зачисленных студентов выбран Министерством образования и науки РФ в качестве одного из важных показателей для оценки деятельности и составления списка неэффективных вузов (Мониторинг Министерства образования и науки РФ, 2012).

Однако информации только о среднем и проходном баллах недостаточно для того, чтобы иметь возможность сравнивать качество приема в вузы на программы с разным количеством бюджетных мест. В этом случае более подходящим инструментом оценки результатов приема могут быть кривые спроса, полученные в результате сортировки по убыванию суммы баллов ЕГЭ студентов, зачисленных на программу обучения, а также рассчитываемые на их основе количественные характеристики (Польшин, Силаев, 2011).

Инструменты анализа качества приема должны также учитывать число поступивших в вуз победителей и призеров олимпиад. Вузы имеют право принимать победителей и призеров престижных олимпиад без вступительных испытаний, поэтому их баллы не влияют на средний балл ЕГЭ поступивших абитуриентов. При этом число студентов, принятых

по результатам олимпиад, отражает степень привлекательности образовательной программы и всего вуза.

Еще одной существенной проблемой при сопоставлении результатов приема в вузы является разный набор или разное количество предметов, которые по требованию министерства или по решению вуза должны учитываться при поступлении на разные направления обучения. Кроме того, известно, что система начисления баллов ЕГЭ непостоянна, средние баллы ЕГЭ выпускников школ варьируются в зависимости не только от предмета, но и от года окончания школы.

Для иллюстрации в табл. 1 и 2 приведены средние значения и среднеквадратичные отклонения баллов по русскому языку (РЯ), математике (М), обществознанию (О), английскому языку (АЯ), информатике и ИКТ (ИИКТ), истории (И) и литературе (Л) по итогам ЕГЭ в 2009–2013 гг. Источник — статистические данные Федерального института педагогических измерений¹ и официального информационного портала ЕГЭ².

Таблица 1. Средние значения баллов по предметам по результатам ЕГЭ в России в 2009–2013 гг.

Год	РЯ	М	О	АЯ	ИИКТ	И	Л
2009	56.9	43.6	56.3	57.8	55.4	47.3	52.5
2010	58.0	43.7	56.0	55.1	62.3	48.7	54.6
2011	60.1	47.5	56.2	60.4	59.5	50.6	57.2
2012	61.1	44.6	55.2	60.8	60.3	51.1	56.3
2013	63.4	48.7	59.5	72.4	63.1	54.8	58.4

Таблица 2. Среднеквадратичные отклонения баллов по результатам ЕГЭ в России в 2009–2013 гг.

Год	РЯ	М	О	АЯ	ИИКТ	И	Л
2009	12.6	15.4	11.1	23.5	15.9	14.9	16.0
2010	12.4	15.6	10.6	21.3	15.5	15.1	16.2
2011	14.9	16.1	11.7	21.3	16.4	18.1	16.8
2012	14.6	16.0	12.7	21.6	19.8	18.9	17.5
2013	14.6	17.9	14.0	21.1	18.8	19.3	17.9

Можно констатировать, что происходит рост средних значений баллов ЕГЭ («инфляция оценок») и среднеквадратичных отклонений от года к году для большинства предметов. При этом в целом есть типичные интервалы значений показателей для отдельных предметов.

Поэтому возникает задача — при сохранении действия специфических факторов скомпенсировать в результатах приема в вузы влияние некоторых общих для всех выпускников факторов, к которым можно причислить разное количество и набор предметов в сумме баллов ЕГЭ и изменения в системе начисления баллов для выпускников разных лет.

¹ <http://www.fipi.ru/>.

² <http://www.ege.edu.ru/>.

Актуальность данной задачи связана также с тем, что в последние несколько лет баллы ЕГЭ часто используются в качестве инструментов статистического анализа в прикладных исследованиях различных вопросов образования в нашей стране. Так, например, в работах (Польдин, 2011; Пересецкий, Давтян, 2011; Замков, Пересецкий, 2013) изучается влияние ЕГЭ на успешность обучения студентов на первых курсах бакалавриата. В статье (Прахов, 2012) анализируется влияние дополнительной подготовки к поступлению в вуз на результаты ЕГЭ и оцениваются факторы, определяющие оценки ЕГЭ. В работе (Андрущак и др., 2012) исследуется проявление эффектов сообучения в студенческой группе, при этом результаты ЕГЭ по русскому языку и математике используются для оценки способностей студентов.

2. Нормированные кривые спроса как инструмент оценки образовательных программ

Состав требуемых сертификатов ЕГЭ для разных направлений обучения может существенно отличаться, что составляет проблему при сопоставлении результатов приема в вузы. Для примера в табл. 3 указаны те наборы предметов, которые требовались для поступления в НИУ ВШЭ в Нижнем Новгороде в 2013 г. Полностью совпадали наборы дисциплин для следующих пар направлений: для Экономики и Менеджмента требовались сертификаты по четырем предметам — математике (М), русскому языку (РЯ), иностранному языку (ИЯ) и обществознанию (О); для направлений Бизнес-информатика (БИ) и Фундаментальная и прикладная лингвистика (ФПЛ) требовались сертификаты по трем предметам — математике, русскому языку и иностранному языку; для направлений Прикладная математика и информатика (ПМИ) и Программная инженерия (ПИ) также требовались сертификаты по трем предметам — математике, русскому языку, информатике и информационно-коммуникационным технологиям (ИИКТ). На направление Юриспруденция требуемый состав сертификатов ЕГЭ включает историю (И), русский язык, иностранный язык и обществознание, на направление Филология — русский язык, иностранный язык и литературу (Л).

Таблица 3. Перечень дисциплин ЕГЭ в 2013 г. по направлениям в НИУ ВШЭ в Нижнем Новгороде

Направление	Предметы ЕГЭ			
Экономика	М	РЯ	ИЯ	О
Менеджмент	М	РЯ	ИЯ	О
Юриспруденция	И	РЯ	ИЯ	О
Бизнес-информатика	М	РЯ	ИЯ	
Прикладная математика и информатика	М	РЯ	ИИКТ	
Программная инженерия	М	РЯ	ИИКТ	
Фундаментальная и прикладная лингвистика	М	РЯ	ИЯ	
Филология	Л	РЯ	ИЯ	

В таблице 4 приведены минимальные проходные и средние баллы в расчете на один предмет студентов, поступивших на бюджетные места по ЕГЭ в 2013 г. в НИУ ВШЭ в Нижнем Новгороде на восемь направлений обучения.

Таблица 4. Проходные и средние баллы (в расчете на один предмет) поступивших на бюджетные места по ЕГЭ в 2013 г. студентов в НИУ ВШЭ в Нижнем Новгороде

Направление	Проходной балл	Средний балл	Приняты по ЕГЭ, чел.	Приняты по олимпиадам, чел.
Экономика	83	86.3	68	9
Менеджмент	79.75	82.5	74	2
Юриспруденция	81.5	85.6	45	7
Бизнес-информатика	78.7	82.8	58	1
Прикладная математика и информатика	74	78.1	32	1
Программная инженерия	81.7	84.7	13	2
Фундаментальная и прикладная лингвистика	86.7	89.1	16	1
Филология	85.7	89.5	11	3

Из таблицы 4 видно, что наибольшие баллы (как проходные, так и средние) наблюдались при поступлении на направления Фундаментальная и прикладная лингвистика и Филология. Поэтому, принимая во внимание только эту информацию, можно сделать вывод, что качество приема на данные направления самое лучшее. Средний же балл на направления Менеджмент и Бизнес-информатика был меньше остальных (за исключением направления Прикладная математика и информатика), и, следовательно, качество приема на данные факультеты хуже, чем на другие направления. Однако корректно сравнивать между собой только результаты приема на направления с одинаковым набором предметов, т. к. статистические характеристики баллов ЕГЭ по разным предметам сильно отличаются. Кроме того, необходимо учитывать студентов, поступивших по результатам олимпиад, и общее число студентов, поступивших по баллам ЕГЭ.

Чтобы проиллюстрировать важность этих факторов, рассмотрим кривые спроса на разные направления обучения, полученные в результате сортировки по убыванию суммы баллов ЕГЭ студентов. Кривые спроса являются зеркальным отображением ненормированных эмпирических интегральных функций распределения сумм баллов ЕГЭ по выборке поступивших на бюджетные места студентов, и поэтому содержат в себе больше информации о поступлении, чем отдельные статистические характеристики, какими являются проходные и средние баллы. Информация о суммарных баллах по результатам ЕГЭ зачисленных в вузы абитуриентов, необходимая для построения кривых спроса, доступна на сайтах вузов³ в приказах о зачислении.

На рисунке 1 приведены кривые спроса, построенные по суммарным баллам ЕГЭ на разные направления обучения в НИУ ВШЭ в Нижнем Новгороде в 2013 г. На рисунке 2 по тем же данным построены нормированные кривые спроса в расчете на один предмет — это позволило устранить разницу в количестве предметов в суммарном балле ЕГЭ на разные образовательные программы.

Важным показателем качества образовательной программы является число поступивших на бюджетные места студентов, зачисленных в вуз по результатам олимпиад. В настоящее время олимпиады являются альтернативой ЕГЭ, победители и призеры олимпиад из утвержденного Перечня олимпиад школьников обладают приоритетным правом по-

³ <http://www.hse.ru/>, <http://www.nnov.hse.ru/>, <http://www.hse.spb.ru/>, <http://www.hse.perm.ru/>.

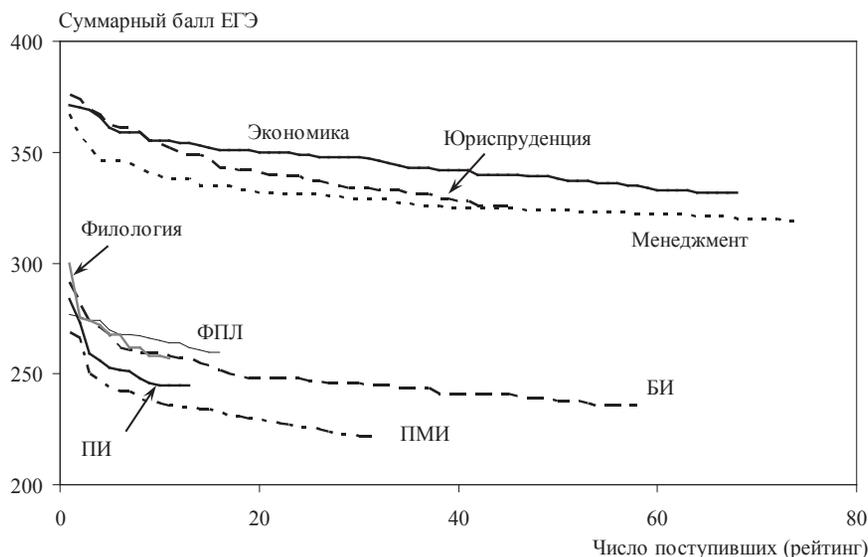


Рис. 1. Суммы баллов ЕГЭ в убывающем порядке (кривые спроса) на разные направления обучения в НИУ ВШЭ в Нижнем Новгороде в 2013 г.

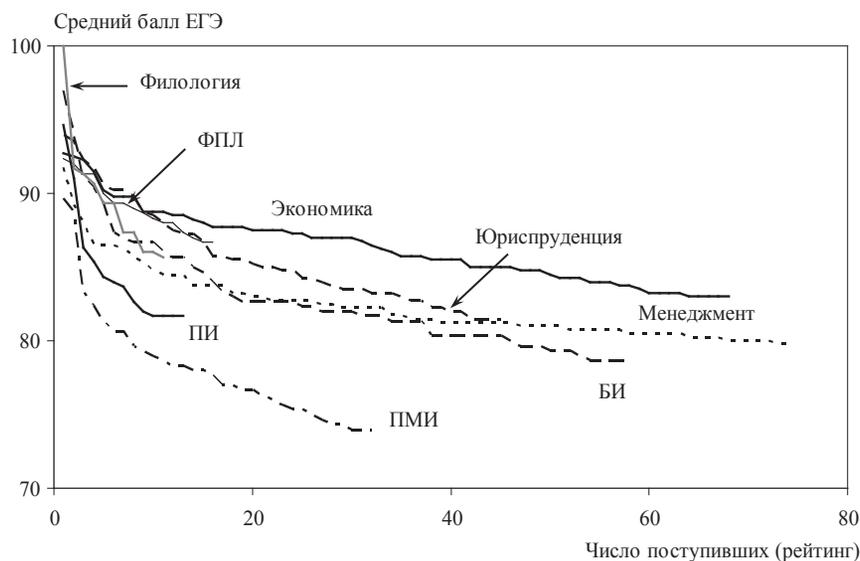


Рис. 2. Нормированные кривые спроса поступивших по ЕГЭ студентов на бюджетные места (в пересчете на один предмет) на разные направления обучения в НИУ ВШЭ в Нижнем Новгороде в 2013 г.

ступления в вузы⁴. Отметим, что абитуриенты, принимаемые по результатам олимпиад без вступительных экзаменов, зачисляются в вуз на основании отдельных приказов, при этом набранная ими сумма баллов ЕГЭ не имеет значения и не является публичной информации.

⁴ Победители и призеры некоторых олимпиад имеют льготу зачесть результат олимпиады как 100 баллов по профилирующему предмету. Этот высший балл суммируется с собственными баллами ЕГЭ по остальным предметам, и с получившимся суммарным результатом абитуриент участвует в общем конкурсе.

ей. Поскольку победители олимпиад поступают на бюджетные места без вступительных экзаменов, можно считать, что их рейтинг выше, чем у поступающих по результатам ЕГЭ.

Поэтому, если учитывать прошедших по результатам олимпиад студентов вместе с теми, которые поступили по баллам ЕГЭ, то кривые спроса на рис. 2 должны сдвинуться вправо на число принятых по олимпиадам. Прделав в соответствии с данными табл. 1 необходимые преобразования сдвига кривых спроса на рис. 2, получим новые кривые спроса, представленные на рис. 3. При этом, например, кривая спроса на направление Экономика на рис. 3 сдвинута вправо на 9 единиц по сравнению с кривой спроса на рис. 2, т. к. на это направление в 2013 г. поступило по олимпиадам 9 человек. Аналогично построены и другие кривые спроса на рис. 3.

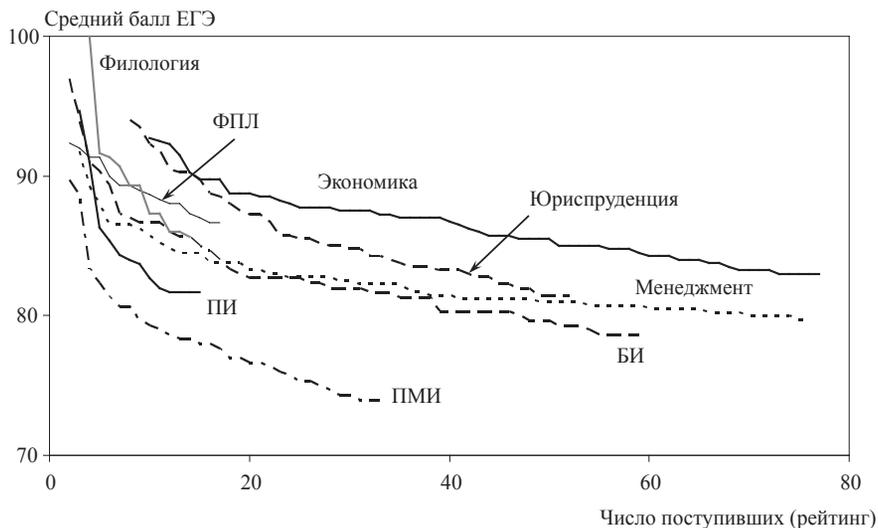


Рис. 3. Нормированные кривые спроса с учетом поступивших без вступительных испытаний победителей и призеров олимпиад в НИУ ВШЭ в Нижнем Новгороде в 2013 г.

Отметим, что кривые на рис. 3 по сравнению с рис. 2 проходят с меньшим числом взаимных пересечений, что позволяет лучше оценивать качество приема на разные направления. Для сравнения качества приема на разные направления обучения можно воспользоваться критериями доминирования первой и второй степени, введенными в статью (Польдин, Силаев, 2011) для кривых спроса по аналогии с критериями стохастического доминирования для случайных величин. Имеет место *доминирование первой степени* спроса на Программу 1 над спросом на Программу 2, если соответствующая кривая спроса на Программу 1 проходит не ниже кривой спроса на Программу 2 и хотя бы часть кривой спроса 1 расположена выше кривой спроса 2. В этом случае для каждого студента Программы 2 сумма баллов ЕГЭ не меньше, чем у студента с таким же порядковым номером (рейтингом) из Программы 1, и хотя бы у одного строго больше. Если число мест на программах различается, то рассматриваем только те участки кривых спроса, для которых есть пары студентов с одинаковыми рангами. Будем считать, что на участке от 1 до Q имеет место *доминирование второй степени* спроса на Программу 1 над спросом на Программу 2, если для всех $n \in [1, Q]$ площадь под кривой спроса 1 на участке от 1 до n не меньше, чем аналогичная площадь под кривой спроса 2, и хотя бы для одного значения n строго больше.

Применение критериев доминирования для кривых спроса на рис. 3 (с учетом 100-балльных результатов студентов, принятых по олимпиадам) позволяет сделать более обоснованные выводы об успешности приема в 2013 г. в НИУ ВШЭ в Нижнем Новгороде на ту или иную образовательную программу. По убыванию качества приема выстраивается такой порядок: на первом месте — направление Экономика, на втором — Юриспруденция, затем — другие направления. Отметим, что этот порядок отличается от того, который следует из сравнения проходных или средних баллов ЕГЭ, представленных в табл. 4.

3. Линейный алгоритм коррекции суммарных баллов ЕГЭ

Результаты освоения различных дисциплин зависят от множества факторов. На оценки влияют врожденные способности и черты характера, воспитание, качество и объем обучения и т. п. По аналогии с предельными теоремами, можно ожидать, что распределение результатов обучения по каждой дисциплине в популяции описывается неким колоколообразным распределением. Среднему уровню освоения предмета в популяции будем ставить в соответствие средний результат теста в принятой шкале, например 50 баллов из 100. Тем самым исключается ситуация, когда, например, среднему уровню освоения математики соответствуют 40 баллов, а среднему уровню освоения иностранного языка — 70 баллов. Также можно ожидать, что в смежные периоды времени не происходит существенных изменений в факторах, определяющих результаты обучения, так что параметры распределения этих результатов сохраняются. Перед выводом алгоритмов коррекции баллов сделаем следующие предположения относительно распределений баллов ЕГЭ.

1. Отличия статистических характеристик распределений баллов ЕГЭ выпускников школ по отдельным предметам связаны с различиями системы начисления баллов по этим предметам, а не с различиями в уровне освоения школьниками этих дисциплин.

2. Отличия статистических характеристик распределений баллов ЕГЭ выпускников школ в различные годы связаны с изменением порядка начисления баллов ЕГЭ, а не с общей подготовкой выпускников. Другими словами, предполагается, что год окончания школы не влияет на способности отдельно взятого школьника к успешной сдаче ЕГЭ.

Для упрощения расчетов будем в качестве результатов ЕГЭ по иностранному языку учитывать результаты ЕГЭ только по английскому языку. Это оправдано тем фактом, что, судя по статистическим данным, число учащихся, сдававших немецкий, французский и испанский языки в 2009–2013 гг. суммарно составляло менее 10% от общего числа всех выпускников, сдававших ЕГЭ по иностранному языку. Соответственно доля выбравших английский язык в качестве ЕГЭ по иностранному языку превышала 90%.

Поскольку при приеме в вузы требуется предъявлять сертификаты ЕГЭ по трем или четырем предметам, интерес представляет распределение случайной величины V — суммы баллов ЕГЭ по дисциплинам (в расчете на один предмет), набор которых зависит от выбранной программы обучения:

$$V = \frac{1}{n} \sum_{i=1}^n \xi_i, \quad (1)$$

где n — число требуемых сертификатов ЕГЭ ($n = 3$ или $n = 4$), ξ_i — баллы ЕГЭ по предмету i .

К сожалению, недостаток информации о результатах ЕГЭ выпускников школ не дает возможности оценить совместное многомерное распределение вероятности баллов ЕГЭ для различных предметов $\Pr(\xi_1, \xi_2, \dots, \xi_n)$ и затем найти одномерное распределение для величины V . Но составить представление о плотности вероятности суммы баллов вполне возможно, построив гистограмму величины V в предположении статистической независимости баллов ЕГЭ по отдельным дисциплинам. На рисунке 4 приведены гистограммы для нормированной суммы баллов V , вычисленные по одномерным распределениям величин ξ_i в предположении их статистической независимости в случае набора из четырех требуемых дисциплин (математика, русский язык, обществознание и английский язык).

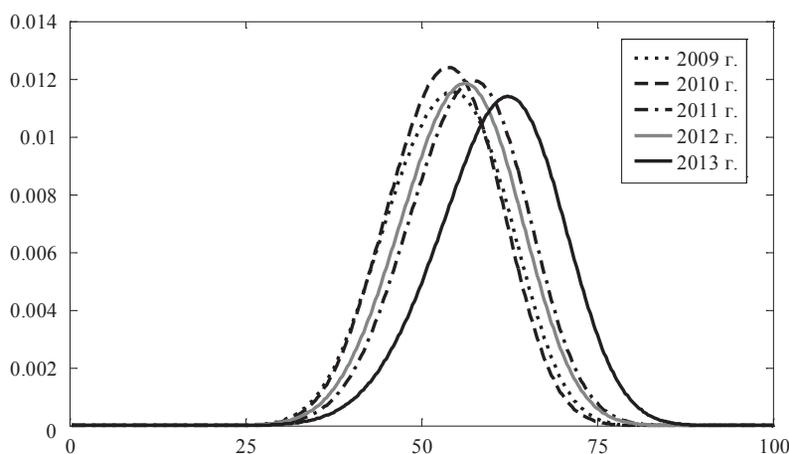


Рис. 4. Гистограммы суммы баллов ЕГЭ по четырем предметам (математика, русский язык, обществознание и английский язык) в 2009–2013 гг.

Можно обратить внимание, что в результате суммирования баллов ЕГЭ и деления суммы на $n = 4$ происходит «нормализация» величины V : суммарная величина визуально хорошо описывается нормальным распределением вероятности, хотя гистограммы по каждому предмету далеко не гауссовы.

Поскольку распределение величины V близко к гауссовскому (отличия от нормального распределения сильнее проявляются на краях интервала оценок при значениях, близких к 0 или 100), то, как представляется, достаточно устранить разницу в средних значениях и дисперсиях для двух разных систем начисления баллов, чтобы скомпенсировать основную часть их отличий друг от друга. Это можно сделать с помощью линейного преобразования величины V , которое приводит среднее значение и дисперсию наблюдаемой суммы баллов ЕГЭ к необходимым значениям в базовом периоде для некоторого базового набора предметов.

Предположим, что наблюдаемая сумма баллов V и сумма баллов базового периода W связаны между собой линейным уравнением

$$W = kV + b. \quad (2)$$

Тогда коэффициенты k и b выражаются через средние значения и среднеквадратичные отклонения случайных величин V и W :

$$k = \sigma_w / \sigma_v, \quad b = E(W) - kE(V). \quad (3)$$

На практике вместо истинного среднего значения $E(V)$ и среднеквадратичного отклонения σ_v величины V будем использовать их оценки, вычисленные по известным статистическим данным о результатах ЕГЭ по всей России и по приближенным значениям коэффициентов корреляции баллов ЕГЭ по отдельным предметам. А для суммы баллов базового периода (в расчете на один предмет) W среднее значение и среднеквадратичное отклонение можно выбирать заранее заданными (стандартными), например, $E(W) = 50$, $\sigma_w = 10$.

Далее считаем, что фактическая S и скорректированная S_c сумма баллов отдельного выпускника связаны такой же линейной зависимостью

$$S_c = kS + b. \quad (4)$$

Эта формула позволяет получить для каждого значения S в интервале от 0 до 100 (в случае нормированной суммы баллов в расчете на один предмет) соответствующую скорректированную оценку S_c .

Чтобы найти оценку среднеквадратичного отклонения σ_v , необходимо знать коэффициенты корреляции баллов ЕГЭ по отдельным предметам российских школьников, которые входят в сумму (1). К сожалению, в официальной статистике результатов ЕГЭ данные о коэффициентах корреляции не приводятся, но их можно приближенно оценить по баллам абитуриентов, поступающих в вузы на различные факультеты. Оценки коэффициентов корреляции для разных предметов ЕГЭ различаются для разных пар предметов, но в целом попадают в интервал от 0.2 до 0.5. Для упрощения расчетов в дальнейшем будем предполагать, что коэффициенты корреляции между баллами отдельных дисциплин одинаковы и равны 0.3. Кроме того, допустимо считать, что все поступающие в вуз студенты сдают ЕГЭ по всем предметам в год поступления.

Полученные при этих предположениях скорректированные кривые спроса на различные направления обучения в НИУ ВШЭ в Нижнем Новгороде в 2013 г. представлены на рис. 5. Сравнивая кривые спроса на рис. 5 и 3, можно увидеть, что коррекция баллов привела в итоге к небольшому изменению относительного расположения кривых, в частности, произошел рост кривых спроса на направления Экономика и Менеджмент относительно кривых спроса на другие направления обучения.

Алгоритм коррекции суммарных баллов ЕГЭ позволяет учитывать изменения в 2009–2013 гг. в порядке начисления баллов по различным предметам. По скорректированным баллам можно исследовать динамику изменений кривых спроса во времени, сравнивать между собой качество поступивших студентов и выявлять факторы, повлиявшие на эти изменения. Кроме того, можно строить суммарные скорректированные кривые спроса, объединяя результаты ЕГЭ для разных групп абитуриентов. Для примера на рис. 6 приведены скорректированные суммарные кривые спроса при поступлении в НИУ ВШЭ в Нижнем Новгороде по результатам ЕГЭ студентов для каждого года поступления с 2009 по 2013 г. При этом проведено суммирование и поступивших без вступительных экзаменов победителей и призеров олимпиад.

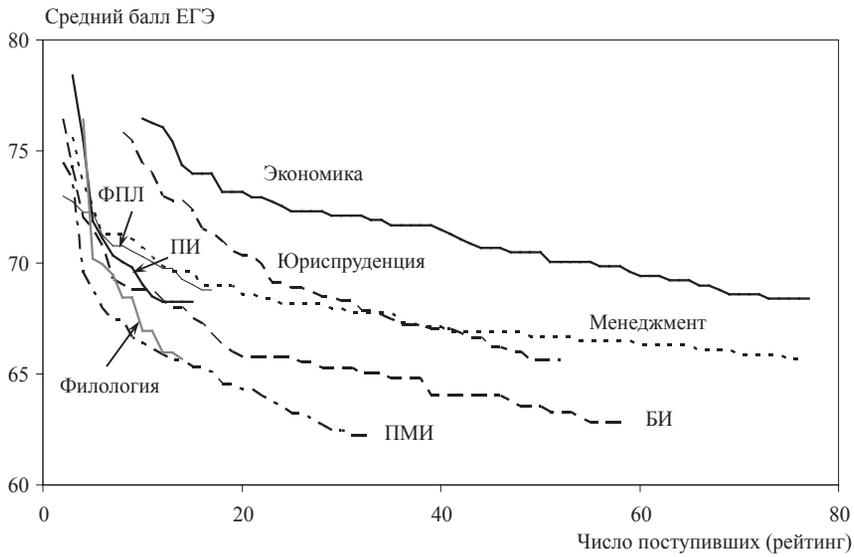


Рис. 5. Скорректированные нормированные кривые спроса в НИУ ВШЭ в Нижнем Новгороде в 2013 г.

Отметим близкое друг к другу расположение скорректированных кривых спроса в 2009–2011 гг., рост качества приема с 2009 по 2010 г. и некоторый спад спроса в 2012 и 2013 гг. Основной причиной снижения спроса, по-видимому, стало дополнительное требование, начиная с 2012 г., сертификата ЕГЭ по иностранному языку при поступлении на все программы обучения за исключением направлений Программная инженерия и Прикладная математика и информатика. Абитуриенты, вовремя не сдавшие ЕГЭ по иностранному языку, вынуждены были поступать в другие вузы или на другие программы обучения.

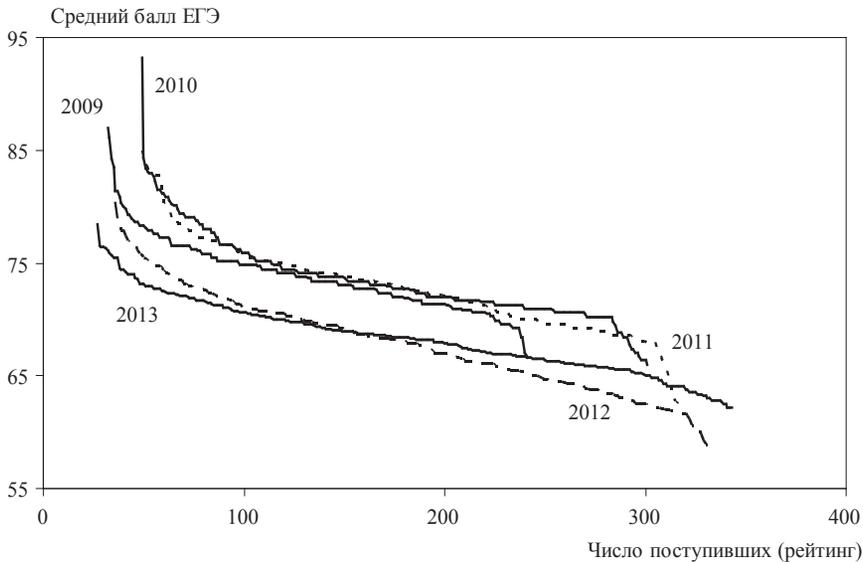


Рис. 6. Скорректированные нормированные кривые спроса в НИУ ВШЭ в Нижнем Новгороде в 2009–2013 гг.

4. Алгоритм нелинейной коррекции баллов ЕГЭ

Достоинством метода коррекции баллов ЕГЭ с помощью линейного преобразования (4) является то, что не надо знать баллы ЕГЭ по каждому предмету в суммарном результате. Часто баллы по отдельным предметам неизвестны, а есть информация только о сумме баллов ЕГЭ поступающих абитуриентов (например из приказов о зачислении). Другим преимуществом линейного алгоритма коррекции является независимость от вероятностного распределения суммы баллов ЕГЭ, а также возможность использовать для пересчета скорректированные баллы при изменении характеристик суммы баллов базового периода. С другой стороны, можно еще раз отметить, что алгоритм требует оценочных значений коэффициентов корреляции баллов ЕГЭ по различным предметам. Другой проблемой является снижение качества коррекции баллов при использовании линейного алгоритма в случае, если сумма баллов ЕГЭ слишком мала (близка к нулю) или слишком велика (близка к 100 в расчете на один предмет). В частности, желательным свойством алгоритма преобразования баллов было бы отсутствие какой-либо коррекции на краях интервала возможных значений, т. е. при нулевом суммарном балле и при максимальном результате в 100 баллов. Очевидно, что линейное преобразование баллов (4) этому требованию не удовлетворяет. При этом многие студенты, поступающие по ЕГЭ в престижные вузы, имеют результаты как раз в зоне больших баллов. Для таких данных можно предложить нелинейную коррекцию баллов ЕГЭ.

Дадим описание алгоритма нелинейной коррекции оценок в предположении, что известны результаты абитуриентов по каждому отдельному предмету в итоговой сумме баллов ЕГЭ. Метод коррекции баллов для каждой отдельной дисциплины основан на следующем известном свойстве нелинейных преобразований случайных величин (Sriboonchitta et al., 2009). Если две случайные величины X и Y имеют непрерывные функции распределения вероятностей $F(x) = \Pr(X \leq x)$ и $G(y) = \Pr(Y \leq y)$, то третья случайная величина Z , полученная с помощью нелинейного преобразования $Z = G^{-1}(F(X))$, будет распределена как Y .

Предположим, что случайная величина ξ — балл ЕГЭ по какой-либо отдельной дисциплине — принимает значения x_1, x_2, \dots, x_m , упорядоченные по возрастанию из интервала $[0; 100]$, с вероятностями p_1, p_2, \dots, p_m . Введем вспомогательную случайную величину X , которая имеет непрерывную функцию распределения, совпадающую в точках x_1, x_2, \dots, x_m с функцией распределения величины ξ :

$$F(x_k) = \Pr(X \leq x_k) = \Pr(\xi \leq x_k) = \sum_{j=1}^k p_j, \quad i = 1, \dots, m. \quad (5)$$

Распределение случайной величины Y можно выбрать произвольным с непрерывной функцией распределения. Однако, учитывая, что баллы по предмету ограничены интервалом $[0; 100]$, удобно использовать бета-распределение в интервале $[0; 100]$ с плотностью вероятности

$$g(y) = \frac{y^{\alpha-1} (100-y)^{\beta-1}}{100^{\alpha+\beta-1} B(\alpha, \beta)}, \quad 0 \leq y \leq 100, \quad (6)$$

где α, β — неотрицательные коэффициенты, $B(\alpha, \beta)$ — бета-функция Эйлера, и непрерывной функцией распределения

$$G(y) = \int_0^y g(t) dt, \quad 0 \leq y \leq 100. \quad (7)$$

Тогда случайная величина $Z = G^{-1}(F(X))$ будет иметь функцию распределения $G(z) = \Pr(Z \leq z)$, совпадающую с функцией распределения величины Y . В результате можно использовать нелинейное преобразование баллов, полученных по отдельным дисциплинам ЕГЭ:

$$M_i^c = G^{-1}(F_i(M_i)), \quad i = 1, \dots, n. \quad (8)$$

Эта формула дает для каждой фактической оценки ЕГЭ M_i (по i -му предмету) в интервале от 0 до 100 соответствующую скорректированную оценку M_i^c . Причем $M_i^c = 0$ при $M_i = 0$, и $M_i^c = 100$ при $M_i = 100$. В целом для всех школьников в России оценки пересчитываются так, что их новые вероятности по данному предмету описываются бета-распределением случайной величины Y . Пересчитав баллы по каждому предмету, найдем скорректированный суммарный балл ЕГЭ (в расчете на один предмет) абитуриента по формуле

$$S^c = \frac{1}{n} \sum_{i=1}^n M_i^c, \quad (9)$$

где n — число предметов, которые учитываются при поступлении на выбранную программу обучения. Далее, после сортировки по убыванию скорректированных суммарных баллов поступивших в вуз студентов, в итоге получим возможность построения соответствующих скорректированных кривых спроса на различные направления обучения.

Отметим, что данный алгоритм не требует знания коэффициентов корреляции баллов ЕГЭ по разным предметам и точно корректирует оценки на всем интервале возможных значений, в том числе близкие к 100 баллам. Но чтобы скорректированные баллы S^c можно было сравнивать друг с другом при поступлении на разные программы обучения, а также с баллами S_c линейного алгоритма (4), надо подбирать соответствующие значения параметров α и β .

Среднее значение и дисперсия бета-распределенной случайной величины Y с плотностью вероятности (6) выражаются через параметры следующим образом:

$$E(Y) = \frac{100\alpha}{\alpha + \beta}, \quad \sigma_Y^2 = \frac{10^4 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}. \quad (10)$$

Будем считать, что $\beta = \alpha$. Тогда получим $E(Y) = 50$, $\sigma_Y = \frac{50}{\sqrt{2\alpha + 1}}$. Значение параметра

α выберем равным $\alpha = 5.44$ при $n = 4$ и $\alpha = 6.17$ при $n = 3$. Варианту $n = 4$ будет соответствовать $\sigma_Y \approx 14.51$, при $n = 3$ получим $\sigma_Y \approx 13.69$. Если предположить, что коэффициент взаимной корреляции для всех слагаемых в сумме (9) снова примерно равен $\rho = 0.3$, то оценка среднеквадратичного отклонения для S^c будет равна 10. Это будет соответствовать аналогичному параметру для S_c в случае линейной коррекции баллов.

На рисунке 7 представлены скорректированные кривые спроса при поступлении на различные направления обучения в НИУ ВШЭ в Нижнем Новгороде в 2013 г., найденные с помощью нелинейного преобразования баллов отдельных дисциплин (8).

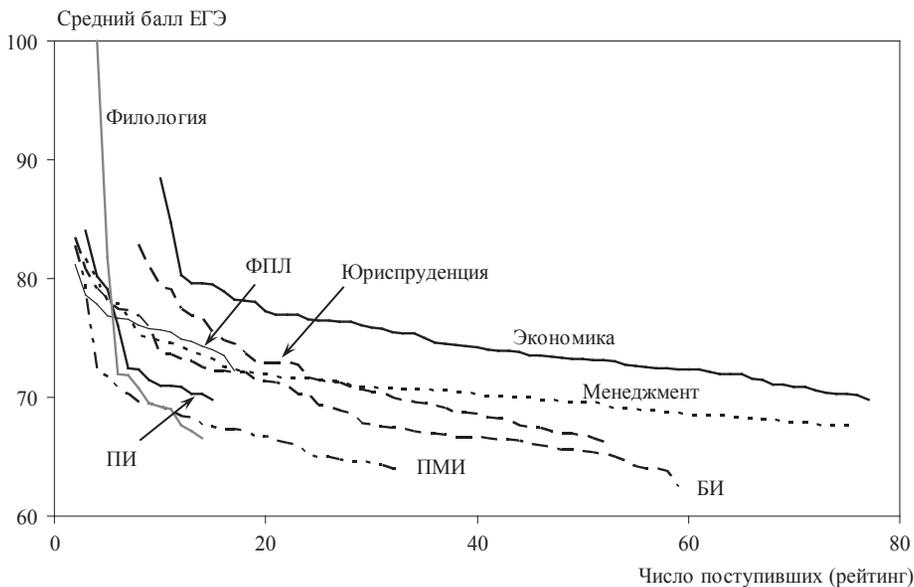


Рис. 7. Скорректированные с помощью нелинейного алгоритма кривые спроса при поступлении в НИУ ВШЭ в Нижнем Новгороде в 2013 г. с учетом поступивших победителей и призеров олимпиад

Сравнивая кривые спроса на рис. 7 с соответствующими кривыми спроса на рис. 5, можно отметить, что использование алгоритма нелинейной коррекции баллов приводит к росту участков кривых спроса с высокими баллами. Например, на рис. 7 возник характерный загиб вверх участка кривой спроса на направление Экономика с баллами выше 80 и проявился 100-балльный результат студентки, поступившей на направление Филология с максимальными оценками по всем предметам. При этом картины относительного расположения кривых спроса на разные направления обучения на рис. 5 и 7 похожи друг на друга. Поэтому можно рекомендовать использовать более сложный нелинейный алгоритм с бета-распределением, если имеется информация о баллах ЕГЭ отдельных абитуриентов по каждому предмету. В итоге можно получить более точные скорректированные результаты во всем диапазоне суммарных баллов ЕГЭ. Но в целом линейная коррекция баллов также приводит к вполне приемлемым результатам. Кроме того, линейный алгоритм проще и не требует информации об оценках по каждому предмету.

Приемные комиссии НИУ ВШЭ в Москве и в трех филиалах (Санкт-Петербурге, Нижнем Новгороде и Перми) в течение последних нескольких лет формируют открытые базы данных о результатах ЕГЭ абитуриентов, подавших заявления для поступления на различные программы, с указанием результатов ЕГЭ по всем требуемым предметам. Поэтому можно сравнивать итоги приема на разные программы с использованием нелинейной коррекции баллов ЕГЭ. Для примера на рис. 8 приведены скорректированные с помощью нелинейного алгоритма кривые спроса при поступлении в 2013 г. на направление Экономика на факультеты экономики НИУ ВШЭ в Москве, Санкт-Петербурге, Нижнем Новгороде и Перми с учетом поступивших без вступительных испытаний победителей и призеров олимпиад.

Результаты поступления абитуриентов на рис. 8 демонстрируют сильное разделение (сегментацию) по баллам ЕГЭ студентов, поступивших на 1 курс факультетов экономики

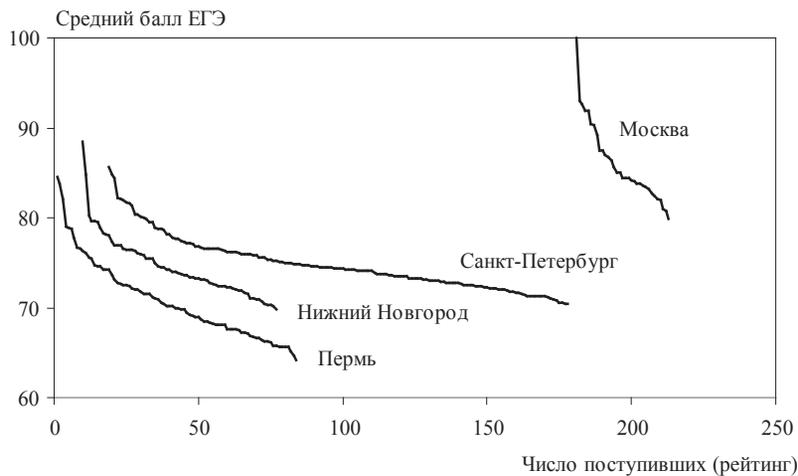


Рис. 8. Скорректированные с помощью нелинейного алгоритма кривые спроса при поступлении в 2013 г. на направление Экономика на факультеты экономики НИУ ВШЭ в Москве и трех филиалах с учетом поступивших победителей и призеров олимпиад

НИУ ВШЭ в Москве и в филиалах. Кроме того, различается и число принятых без экзаменов победителей и призеров олимпиад на факультеты экономики: в Москве — 180, в Санкт-Петербурге — 18, в Нижнем Новгороде — 9, в Перми — 0. Поскольку учебные планы на рассматриваемых факультетах по содержанию близки друг к другу, то разницу в кривых спроса можно объяснять территориальными причинами, качеством обучения, репутацией и другими факторами.

В общем случае, после того как построены скомпенсированные кривые спроса по баллам ЕГЭ, в которых устранены различия в количестве и наборе предметов на разные программы обучения, а также учтены результаты поступивших по олимпиадам студентов, можно более точно изучить влияние других факторов: степень конкуренции со стороны других вузов и других образовательных программ, различия в подготовке школьников, разные перспективы трудоустройства и развития карьеры после учебы, качество образовательных программ и др.

5. Заключение

В работе показано, что применение средних и проходных баллов для анализа результатов приема на различные направления обучения в вузе может приводить к ошибочным выводам. В качестве более точного инструмента анализа предлагаются скорректированные кривые спроса, которые можно получать с помощью линейного преобразования по данным о суммарных баллах ЕГЭ студентов, поступивших на бюджетные места, или с помощью нелинейного алгоритма, корректируя оценки по каждому отдельному предмету в суммарном результате. Поступление без вступительных экзаменов победителей и призеров олимпиад учитывается в кривых спроса с помощью соответствующего параллельного сдвига вправо.

По скорректированным кривым спроса можно делать более обоснованные выводы о привлекательности различных программ для абитуриентов и эффективности деятельности вуза. После коррекции можно рассчитывать средние и проходные баллы для наглядного рейтингования программ, при этом избегая искажений, связанных с различиями в структуре вступительных испытаний на различных направлениях обучения.

Список литературы

Андрущак Г. В., Польдин О. В., Юдкевич М. М. (2012). Эффекты сообучения в административно формируемых студенческих группах. *Прикладная эконометрика*, 26 (2), 3–16.

Замков О. О., Пересецкий А. А. (2013). ЕГЭ и академические успехи студентов бакалавриата МИЭФ НИУ ВШЭ. *Прикладная эконометрика*, 30 (2), 93–114.

Мониторинг Министерства образования и науки РФ деятельности федеральных образовательных учреждений высшего профессионального образования. (2012). http://минобрнауки.рф/пресс-центр/2774/файл/1265/12.10.31-Мониторинг_Результаты.pdf.

Мониторинги Национального исследовательского университета «Высшая школа экономики»: Рейтинг качества приема в российские вузы — 2010, Качество приема в вузы — 2011, 2012, 2013. <http://www.hse.ru/org/hse/ex/>; http://www.hse.ru/ege/second_section; http://www.hse.ru/ege/second_section2012; http://www.hse.ru/egesecond_section2013.

Пересецкий А. А., Давтян М. А. (2011). Эффективность ЕГЭ и олимпиад как инструмента отбора абитуриентов. *Прикладная эконометрика*, 23 (3), 41–56.

Польдин О. В. (2011). Прогнозирование успеваемости в вузе по результатам ЕГЭ. *Прикладная эконометрика*, 21 (1), 56–69.

Польдин О. В., Силаев А. М. (2011). Сравнение образовательных программ по результатам ЕГЭ зачисленных студентов. *Вопросы образования*, 3, 192–209.

Прахов И. А. (2012). Единый государственный экзамен и детерминанты результативности абитуриентов: роль инвестиций в подготовку к поступлению. *Прикладная эконометрика*, 27 (3), 86–108.

Sriboonchitta S., Wong W.-K., Dhompongsa S., Nguyen H. T. (2009). *Stochastic dominance and applications to finance, risk and economics*. CRC Press, Chapman & Hall Book.