

ЕНТ: эмпирический анализ факторов

Серия «Экономические исследования» №2-2020

Жаныбек Айгазин, Амина Бауыржан,
Рустам Бекишев, Евгения Пак

Нур-Султан, 2020

Серия публикаций АЕРС (Центр исследований прикладной экономики) предназначена для распространения результатов научно-исследовательских работ работников АЕРС в целях стимулирования дискуссий в рамках соответствующих тем. Мнения, высказанные в документе, выражают личную позицию авторов и могут не совпадать с официальной позицией АЕРС.

ЕНТ: эмпирический анализ факторов

Авторы: Айгазин Ж.Ж., Бауыржан А.Б., Бекишев Р.А., Пак Е.А.

Апрель 2020 г.

AERC-WP-2020-2

Дисклеймер

Представленные материалы не несут в себе нарушение прав человека, нарушения тайны и конфиденциальности, не призван высмеять, оскорбить или унижить отдельных лиц или целые группы населения. Его использование любыми лицами при принятии решений не влечет за собой какой-либо ответственности АЕРС за возможные потери или убытки, явившиеся следствием таких решений.

© АЕРС

Любое воспроизводство представленных материалов допускается только с разрешения авторов

Казахстан, г. Нур-Султан, шоссе Қорғалжын, 23/1, ВП-25

тел.: +7 7172 97 20 36

info@aerc.org.kz

www.aerc.org.kz

ЕНТ: эмпирический анализ факторов

Жаныбек Айгазин¹
Амина Бауыржан²
Рустам Бекишев³
Евгения Пак⁴

Аннотация

С использованием инструментальных переменных получены статистически значимые результаты для объяснения среднего балла ЕНТ при сдаче на казахском и русском языках. Подтверждена гипотеза, что социально-экономический статус и бедность являются важными факторами объяснения различия в образовательных достижениях: высокий балл ЕНТ при сдаче на казахском и русском языках зависит от медианного дохода домохозяйств и уровня бедности соответственно, так как возможности родителей действуют как инвестиции в детей и, возможно, будут способствовать поступлению в престижный ВУЗ и являться хорошим стартом для будущей взрослой жизни. В остальном факторы различны: если балл ЕНТ на казахском языке зависит от навыков использования интернета в век информационных технологий, то на русском языке – от этнической разнородности, что связано с тем, что нетитульные этносы в основном сдают ЕНТ на русском языке. При увеличении на 1% реального медианного дохода домохозяйства средний балл ЕНТ при сдаче на казахском языке растет на 0.13%, рост на 1% доли пользователей интернета увеличивает средний балл ЕНТ на 0.35%. При снижении на 1% уровня бедности балл ЕНТ при сдаче на русском языке увеличивается на 0.04%, при росте на 1% этнического многообразия региона средний балл ЕНТ увеличивается на 0.035%. Итоги исследования могут быть полезны для государственных органов, осуществляющих руководство в сфере образования и науки, некоммерческим организациям, экспертному сообществу, а также широкому кругу общественности, кому интересны проблемы отечественного образования.

Ключевые слова: ЕНТ, Education Economics, pooled regression, fixed effects, random effects, инструментальная переменная, тест Бреуша-Пагана, тест Хаусмана, тест множителей Лагранжа Бреуша-Пагана, тест Песарана, тест Ву-Хаусмана, гетероскедастичность, робастная ковариационная матрица.

Классификация JEL: D41, C51, L62

¹ Директор AERC, к.э.н., aigazin@aerc.org.kz

² Старший аналитик Департамента прикладных исследований AERC, a.bauyrzhan@aerc.org.kz

³ Старший аналитик Департамента прикладных исследований AERC, MSc., rustam.bekishev@aerc.org.kz

⁴ Ведущий аналитик Департамента консалтинговых услуг AERC, zhenyapak@aerc.org.kz

Содержание

ВВЕДЕНИЕ	5
ОБЗОР ЛИТЕРАТУРЫ	7
ОПИСАНИЕ ДАННЫХ	12
МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ	16
ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ	19
ЗАКЛЮЧЕНИЕ	30
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	31
ПРИЛОЖЕНИЯ	35

ВВЕДЕНИЕ

В настоящее время структура занятости населения мира кардинально меняется: роль умственного труда возрастает, а потребность в малоквалифицированном труде снижается. Данные процессы имеют различную скорость для стран с разным уровнем экономического развития: для развитых стран эти процессы протекают быстрее, нежели чем для развивающихся стран. Это связано с предъявляемым развитыми странами высоким спросом на малоквалифицированный труд.

Экономика образования (Education Economics) занимается изучением экономических вопросов, связанных с образованием, включая спрос на образование, образование как инвестиции в человеческий капитал, финансирование и предоставление образовательных услуг, оценка эффективности разных образовательных программ и стратегий, различные прикладные вопросы. Начиная с ранних работ, посвященных взаимосвязи между образованием и рынком труда, экономика образования быстро расширялась, охватывая практически все области, связанные с образованием.

Одним из основных подходов в Education Economics является концепция производственной функции в сфере образования. Концепция связывает различные исходные данные, влияющие на обучение учащегося (школы, семьи, сверстники, кварталы и т.д.) с измеряемыми результатами, включая последующий успех на рынке труда, посещаемость колледжа и показатели успеваемости (результаты тестов).

В мировой экономической литературе образовательная сфера рассматривается уже как мощный дополнительный источник качественного экономического роста страны, становясь актуальной темой экономических исследований. Нобелевский лауреат по экономике 2018 г. П. Ромер в знаменитой теории эндогенного роста, выявил, что экономики, имеющие человеческий капитал и науку, имеют в долгосрочной перспективе большие темпы роста, чем экономики, лишенные этих ресурсов [Romer 1986]. Развитие человеческого капитала, несомненно, является следствием уровня образования любой экономики.

В Республике Казахстан измерение результатов успеваемости по единой методологии оценки начало применяться с 2004 г. с введением Единого национального тестирования (ЕНТ). ЕНТ проходит в один этап и оценивается по 140-балльной шкале. Ученики в один день сдают пять предметов: три обязательных (математическая грамотность, грамотность чтения, история Казахстана) – по 20 заданий на каждый и два профильных – по 30 заданий на каждый (10 заданий по профильным предметам имеют несколько правильных ответов, поэтому за правильный ответ на них дают по 2 балла).

Авторы настоящего исследования на основании имеющейся статистики и применения эконометрического моделирования попытались ответить на следующий вопрос: какие факторы могут объяснить успешность сдачи ЕНТ в разрезе языков, на котором сдается национальное тестирование. Другими словами, что является факторами, позволяющими успешно сдать ЕНТ и получить шанс для лучшей будущей жизни.

Итоги исследования могут быть полезны для государственных органов, осуществляющих руководство в сфере образования и науки, некоммерческим организациям, экспертному сообществу, а также широкому кругу общественности, кому интересны проблемы отечественного образования.

Опираясь на цель исследования, структура работы содержит раздел по освещению некоторых научных работ, посвященных изучению Education Economics и использованию инструментов статистической оценки данных в различных странах мира. Далее дается описание исходных данных, методология исследования, а также обсуждение полученных результатов и заключение.

ОБЗОР ЛИТЕРАТУРЫ

Одной из наиболее ранних работ современности, положивших начало оценке образовательной производственной функции, стало исследование Coleman (1968), в рамках которого автор доказал, что семья и сверстники оказывают большее влияние на школьные достижения ученика, нежели школа и учителя. Подтверждает его исследование работа Black, Devereux, Salvanes (2004), где авторы доказывают отрицательное влияние роста размера семьи на уровень образования детей в этой семье. Более того, авторы исследования отмечают, что каждый последующий ребенок в семье характеризуется более низким уровнем образования. Так, разница в образовании первого и пятого ребенка в семье равносильна разнице в образовании афроамериканца и белого.

Однако, Cameron и Heckman (1998) подчеркивают, что влияние семьи на достижения ребенка в учебе уменьшается с ростом уровня образования самого ребенка от начального школьного к высшему. И, в конечном итоге, влияние семьи при получении ребенком высшего образования практически нивелируется, возрастает роль других факторов.

С появлением возможностей исследования большего объема данных, у экономистов появилась возможность оценить влияние именно преподавателей, уровня их квалификации на достижения в учебе учеников, а влияние последнего – на экономический рост страны более детально. И уже сейчас многие исследования доказывают значимость для успеваемости учеников таких факторов, как школа и уровень квалификации преподавателей.

Сегодня большинство исследователей сходятся во мнении, что школы и учителя имеют значение [Greenwald, Hedges, Laine 1996] и [Hanushek 1996, 1997, 2002]. В работах [Murnane 1975], [Goldhaber, Brewer 1997], [Angrist, Lavy 2001], [Jepsen, Rivkin 2002], [Rivers, Sanders 2002], [Jacob, Lefgren 2004], [Rockoff 2004], [Kane, Staiger 2005], [Rivkin, Hanushek, Kain 2005] и [Kane, Rockoff, Staiger 2006] оценивается влияние качества учителей.

Исследование Aaronson, Barrow и Sander (2007), посвященное оценке влияния преподавателей в государственных средних школах Чикаго на успеваемость учеников, показывает, что в рамках одного стандартного отклонения преподавание математики в течение одного семестра более квалифицированным педагогом улучшает оценки учащихся по математике на 0.13 пунктов. При этом, полученные авторами результаты оценивания относительно стабильны во времени, достаточно невосприимчивы к различным условным переменным и не зависят от аудитории учеников. Авторами также была отмечена значимость квалификации преподавателя для успеваемости учеников с ограниченными возможностями.

К похожим выводам о значимости квалификации педагога для успеваемости учеников приходят Rivkin, Hanushek и Kain (2005). Авторы на основе анализа данных по штату Техас, утверждают о том, что повышение квалификации преподавателя положительно сказывается на успеваемости обучающихся по таким предметам, как математика и чтение. Более того, согласно проведенному исследованию, эффективность от уменьшения размеров класса до 10 человек меньше эффекта от повышения квалификации преподавателя.

В то же время, ряд экономистов изучал влияние нетипичных факторов на успеваемость учеников. К примеру, Aucejo и Romano (2016) оценивают влияние сокращения числа пропусков учениками и продолжительности учебного года на

результаты тестирований этих учеников, используя административные данные государственных школ штата Северная Каролина. Результаты исследования оказались следующими: продление школьного календаря на десять дней увеличивает результаты тестов по математике и чтению только на 1.7% и 0.8% от стандартного отклонения соответственно. При этом аналогичное сокращение числа прогулов школьных занятий учениками приводит к росту оценок учеников на 5.5% по математике и 2.9% по чтению. Таким образом, полученные результаты свидетельствуют о том, что в то время, как многие политические деятели сосредоточили свое внимание на продлении периода школьного обучения, на самом деле, борьба с прогулами может стать более эффективным вмешательством, поскольку, согласно полученной модели, эффект сокращения пропусков по сравнению с увеличением количества школьных дней существенно выше.

Тем не менее, два нобелевских лауреата по экономике Banerjee и Duflo (2007) в своем исследовании на примере Индии доказывают, что не всегда посещаемость является ключевым фактором успеха в повышении успеваемости учеников – многое зависит от качества самих школ. Авторы отмечают, что даже в крупных городах Индии, где прогулы учеников и учителей – нечастое явление, показатели успеваемости учеников всё равно очень низкие. Утверждая, что данные по Индии свидетельствуют о том, что многие государственные усилия, направленные на сокращение числа прогулов и удлинение периода обучения, не приводят к улучшению результатов теста среднего ученика, авторы делают акцент на вопросе повышения качества обучения в школах наименее затратным способом. В связи с поставленной задачей, авторы статьи оценивают влияние двух программ по улучшению качества школьного образования в Индии на результаты тестирований учеников. При этом, первая программа специально предназначена для самых слабых детей: это корректирующая образовательная программа, где молодая женщина («Бальсахи») работает над базовыми навыками с детьми, которые достигли 3 или 4 класса, не освоив их. Этих детей вывозят из обычного класса работать с этой женщиной по два часа в день (при том, что школьный день длится около четырех часов). Вторая программа также направлена на повышение успеваемости учеников, но адаптирована к текущему уровню каждого ребенка. Вторая программа основана на компьютерном обучении: дети по 2 часа в день играют в компьютерные игры, включающие решение математических задач, уровень сложности которых варьируется в зависимости от уровня знаний ученика. Оценки, проведенные в двух городах за два года, показывают, что обе программы эффективны. Обе программы сопряжены с невысокими издержками по внедрению и потому могут быть легко масштабируемыми на всю страну: программа коррекционного образования уже охватила десятки тысяч детей по всей Индии.

Хотя оценка эффективности компьютерной программы в США была противоположной результатам, полученным в Индии, Carter, Greenberg и Walker (2017) изучили влияние разрешения/запрета использования компьютеров во время занятий по курсу «экономика» в Военной Академии США и выявили, что в классах с разрешенным доступом к компьютеру результаты итогового тестирования по предмету оказывались на 0.18 ниже результатов аналогичного теста в классах с запрещенным доступом к компьютеру. Авторы связывают данный эффект с тем, что в классах с разрешенным доступом к компьютеру не контролируются цели его использования учениками. Как следствие, ученики в таких классах могут отвлекаться на компьютер не в учебных целях, и компьютер в этом случае – фактор

рассеивания внимания учеников от лекций. Немного подробнее данный вопрос исследуют Comi, Argentin, Gui, Origo и Pagani (2017), где авторы рассматривают влияние использования ИКТ преподавателями на результаты межпредметного тестирования учеников 10-х классов. Было доказано, что компьютерные методы обучения повышают успеваемость учащихся, когда используются в целях повышения осведомленности учеников, в целях налаживания коммуникационного канала между учениками и учителем. Однако негативное влияние на успеваемость оказывает практика активного использования компьютеров в классах во время занятий.

Довольно необычным представляется исследование Angrist и Krueger (1991), в рамках которого авторы рассмотрели влияние сезона рождения ребенка на уровень его образования. Политика в отношении возраста поступления в школу и законы об обязательном посещении школы приводят к значимости такого фактора как сезон рождения ребенка для его успеваемости: дети, родившиеся в начале года вступают в школу в более старшем возрасте в сравнении с детьми, рожденными ближе к концу года, что делает неудивительным более высокую в среднем успеваемость первых. При этом рожденные в начале года дети, как следствие, достигают возраста, в котором по закону они имеют право бросить обучение в школе раньше детей, родившихся в конце года. По оценкам авторов статьи, порядка 25% учеников продолжают посещать школу в старших классах в силу закона об обязательном посещении и не могут бросить ее из-за того, что не достигли соответствующего возраста, в котором им дозволено делать выбор. Авторы статьи затем оценивают, как закон об обязательном школьном образовании через такой инструмент, как квартал рождения ребенка, влияет на дальнейшие заработки бывших учеников школ США. Согласно полученным результатам, ученики, которые вынуждены дольше посещать школу в соответствии с законами об обязательном школьном обучении в итоге получают более высокую заработную плату. Кроме того, исследователи выяснили, что законы об обязательном школьном обучении имеют положительный внешний эффект. Тем не менее, авторы отмечают, что данные результаты еще не означают, что законодательство об обязательном школьном посещении всегда выгодно – есть и другие факторы, которые могут изменить влияние закона об обязательном школьном посещении.

Наболевшим вопросом при изучении сферы образования является также значимость уровня полученного образования для будущего карьерного роста и заработка людей. Этой теме посвящено не одно исследование. Наиболее цитируемым в данном классе работ является исследование Ashenfelter и Krueger (1994), в рамках которого изучались денежные эффекты от дополнительного года обучения близнецов, растущих в одинаковых условиях, чтобы выявить чистый эффект уровня образования на заработок. Было выяснено, что дополнительный год обучения увеличивает заработную плату на 12-16%.

Card и Krueger (1992) утверждают, что дополнительный год образования значимо влияет на дальнейшие заработки учеников только в случае, если речь идет об образовании в качественных школах. А качество школы оценивается по количеству высококвалифицированных преподавателей.

Но уже в более позднем исследовании тот же Card (2001) приходит к выводу, что результат оценивания влияния уровня образования на заработок зависит просто от выбранной эконометрической модели: оценка влияния с помощью МНК обычно ниже по сравнению с моделями инструментальных переменных. В то же время чаще всего оценка влияния образования на заработок с помощью

инструментальных переменных является относительно неточной, так как ни одна из эмпирических стратегий не основана на истинной рандомизации. Таким образом, ни оценка МНК, ни оценка через инструментальные переменные не могут стать решающими в дискуссии о влиянии уровня образования на заработок учеников в будущем.

Не менее важным все еще остается вопрос влияния сферы образования на экономический рост страны. Так, Barro (1991), Benhabib и Spiegel (1994), Barro и Sala-i-Martin (1995) считают, что школьное обучение положительно коррелирует с темпами роста ВВП на душу населения по странам. Однако в противовес данным исследованиям Bils и Klenow (2000) опровергают значимость увеличения охвата образования для экономического роста страны. Используя данные ЮНЕСКО за 1977-1983 гг. авторы построили модель, демонстрирующую слабую связь между охватом населения начальным школьным образованием и экономическим ростом.

Однако в ранней работе Argow (1973), в рамках которой он доказывает невозможность корректной оценки указанного влияния, поскольку во многих работах понятие роста человеческого капитала подменяется ростом расходов на образование, что не одно и то же. В некоторых исследованиях была попытка решить данный вопрос, оценивая человеческий капитал результатами тестирования учеников. Однако, как утверждает Argow, данные тесты оценивают скорее интеллект, а не способности человека производить что-либо. Следовательно, не имея инструментов измерения человеческого капитала, то есть человеческих способностей, мы не можем говорить о корректной оценке его влияния на экономический рост страны.

Тем не менее, Hanushek и Kimko (2000) настаивают на положительном влиянии роста человеческого капитала на экономический рост страны, если рассматривать качественный, а не количественный рост человеческого капитала. И в рамках своего исследования авторы оценивают качественный рост человеческого капитала как рост результатов тестирования учеников только по математике и точным наукам. Оценивая человеческий капитал таким образом, его влияние на экономический рост строго и стабильно положительное.

Одинаковую позицию занимают исследователи в вопросе влияния уровня образования на социально-демографические показатели страны. Так, Kenkel (1991) и Lleras-Muney (2005) провели исследования независимо друг от друга и пришли к одинаковому выводу, что повышение уровня образованности населения приводит к росту здоровья нации, так как с ростом образованности люди начинают следовать более здоровым привычкам, вести здоровый образ жизни.

Положительна корреляция между уровнем образования матерей и результатами родов, выявленная в исследовании Currie и Moretti (2003). Авторы оценили влияние уровня образования матерей на результаты родов по 4 каналам: дородовой уход, брачность, курение, возраст матери. Было доказано, что с повышением уровня образования матерей уменьшается число детей, рожденных вне брака, снижается младенческая смертность, доля детей, рожденных с врожденными заболеваниями (как результат ведения здорового образа жизни образованными матерями во время беременности).

Социально-экономический статус и последствия бедности являются важными факторами в объяснении различия в образовательных достижениях детей [Brooks-Gunn и Duncan, 1997; Mayer 1997; Brooks-Gunn et al., 1994, 1995, 2000]. Автор экономического бестселлера «Фрикономика» Levitt совместно с Fryer пришли к выводу, что существуют расовые различия достижений учеников: данные

показывают, что чернокожие дети учатся плохо в школе не из-за цвета кожи, а потому что они чаще всего выходцами являются из малообразованных и бедных семей; кроме того, даже с учетом фактора дохода и образования родителей разрыв между чернокожими и белыми детьми появляется уже на третий год обучения в школе [Fryer, Levitt, 1997].

Статистическое подтверждение, казалось бы, очевидной положительной зависимости между уровнем образования и снижением преступности приводится в работе Lochner и Moretti (2004). Используя данные переписи населения и данные ФБР, авторы пришли к выводу, что обучение значительно снижает вероятность лишения свободы и ареста. Причем такой результат был получен и при оценивании влияния уровня образования на вероятность ареста в рамках оценивания модели МНК, и при использовании моделей с инструментальной переменной.

В то время как влияние образовательной сферы на экономический рост, а уровня образования – на дальнейшие заработки не столь очевидно, исследователи единодушны в том, что рост человеческого капитала оказывает положительное влияние на социальную сферу жизни общества.

Таким образом, в Education Economics накоплен значительный объем исследований влияния образования на экономику, а также влияния социально-экономических факторов на успеваемость. Однако в казахстанской экономической литературе пока не было замечено подобных попыток.

ОПИСАНИЕ ДАННЫХ

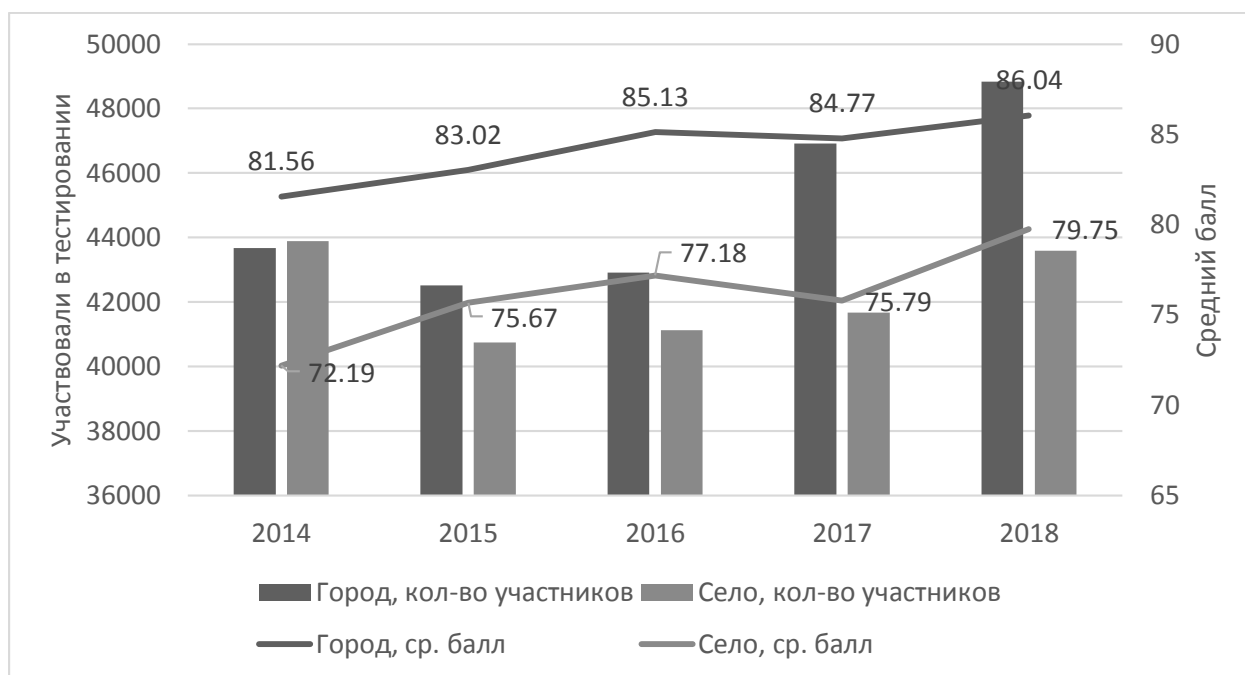
Необходимые статистические данные для исследования были взяты из официальных источников. Данные по результатам ЕНТ были получены из открытых источников РГКП «Национальный центр тестирования» Министерства образования и науки РК⁵. В связи с тем, что сведения по набранным баллам ЕНТ в разрезе регионов доступны лишь с 2014 г., в данном исследовании все остальные показатели были применены, начиная с данного периода времени.

Статистические данные, описывающие состояние системы образования, такие, как качественный и количественный состав учителей в разрезе разных предметов, контингент организаций общего среднего образования по ведомственной принадлежности, численность учащихся в разрезе языков обучения и наполняемость классов, были собраны из ежегодных национальных сборников, публикуемых АО «Информационно-аналитический центр» Министерства образования и науки РК. Статистика по социально-экономическому развитию регионов взята из бюллетеней, публикуемых Комитетом по статистике Министерства национальной экономики РК.

Статистические данные представлены в разрезе 14 регионов и двух городов республиканского значения Республики Казахстан за 2014-2018 гг.

На рисунке 1 представлена сравнительная информация о количестве абитуриентов и результатах сдачи ЕНТ среди учащихся городской и сельской местностей. Несмотря на то, что учащиеся в сельской местности всегда набирали относительно меньше, средний балл среди тестируемых в сельской местности увеличился на 9,7% в 2018 г. по сравнению с 2014 г. За аналогичный период времени, средний уровень в городской местности увеличился лишь на 5,5%.

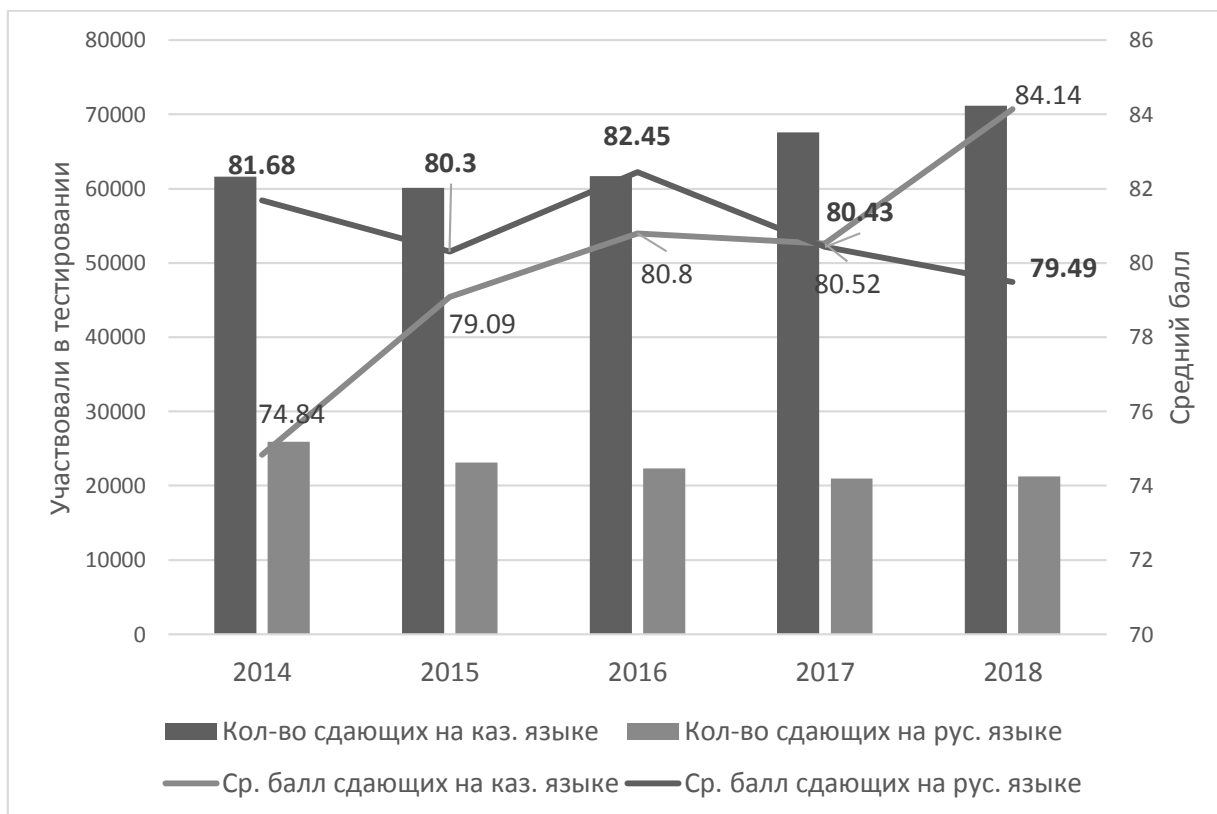
Рисунок 1. Сравнительная статистика результатов ЕНТ в разрезе типа местности



⁵ <http://testcenter.kz/ru/rezultaty-testirovaniya-i-konkursa/>

Стоит отметить, что количество участников, сдающих экзамен на казахском языке, ежегодно значительно превышает количество сдающих на русском языке. Более того, среди тех, кто сдает ЕНТ на казахском языке, отмечаются более высокие результаты по сравнению с теми, кто сдает на русском языке.

Рисунок 2. Сравнительная статистика результатов ЕНТ в разрезе языка сдачи



Примечательно то, что за последние пять лет ежегодно абитуриенты из г. Алматы в среднем набирают значительно больше по сравнению с учащимися из других регионов. Напротив, самые низкие результаты ежегодно показывают тестируемые из Атырауской области.

Таблица 1. Средний балл участников ЕНТ в разрезе регионов

Регион	2014	2015	2016	2017	2018
г. Алматы	93.4	91.8	96.0	95.9	101.1
Павлодарская	79.9	80.2	88.2	91.9	91.0
Жамбылская	72.3	78.6	78.3	78.0	89.4
Костанайская	79.9	78.8	83.9	81.9	86.7
Кызылординская	74.2	81.7	82.6	82.4	86.1
Актюбинская	84.0	82.7	82.8	82.4	85.6
Восточно-Казахстанская	78.0	80.2	83.3	82.3	82.3
Западно-Казахстанская	81.3	83.6	82.1	80.3	81.9

Регион	2014	2015	2016	2017	2018
Алматинская	72.6	76.4	79.6	77.5	81.0
г. Нур-Султан	87.6	86.7	89.2	85.0	79.9
Акмолинская	80.2	78.2	77.7	80.0	79.0
Мангистауская	74.5	75.8	78.5	77.7	78.8
Карагандинская	76.9	79.6	81.8	78.3	77.9
Южно-Казахстанская	74.5	78.7	77.7	77.0	77.8
Северо-Казахстанская	77.9	77.8	78.8	75.1	77.6
Атырауская	62.1	61.0	70.0	72.2	73.8

Статистические данные представлены в разрезе 16 регионов Республики Казахстан за 2014-2018 гг⁶. Временные ряды с номинальными значениями были переведены в цены базового 2014 г. путем дефлирования на индекс потребительских цен конкретного региона. Имеющаяся выборка незначительная: 80 наблюдений для 79 переменных. Панель несбалансированная, так как имеются пропуски некоторых данных. Небольшое число наблюдений связано с тем, что данные по ЕНТ есть в свободном доступе только с 2014 г., хотя ЕНТ стартовало в Республике Казахстана с 2004 г.

Перечень переменных (79 переменных) представлены в приложении 1.

В приложении 2 представлена корреляционная матрица, которая позволяет не применять в одном уравнении сильно коррелированные экзогенные переменные в целях исключения мультиколлинеарности. Для удобства в корреляционной матрице остались только те переменные, которые участвовали в моделировании.

Все переменные датасета будут проверены на стационарность путем поиска наличия единичных корней с помощью модифицированных тестов ADF (Augmented Dickey-Fuller) [Croissant, Millo 2019]. За последние годы объем литературы по тестированию корней и коинтеграции панелей значительно вырос и в настоящее время проводится различие между тестами первого поколения [Maddala, Wu 1999; Levin, Lin and Chu, 2002; Im, Pesaran and Shin, 2003], разработанными с учетом допущения о перекрестной независимости панельных данных, и тестами второго поколения [Bai, Ng 2004; Smith et al., 2004; Moon, Perron 2004; Choi 2006; Pesaran 2007], учитывающими различные зависимости между различными переменными и групповыми тестами единичных корней, которые позволяют учесть структурные разрывы [Im, Lee 2001]. Кроме того, в последние годы стало более широко признаваться, что преимущества методов панельных данных в настройках макрорядов включают использование данных, для которых диапазоны данных отдельных временных рядов недостаточны для изучения многих интересных гипотез [Arouri, Rault 2014].

В настоящем исследовании датасет является гетерогенным, так как панель представляет собой статистические данные регионов страны, которые по своей природе неоднородны [Hsiao 2003; Klevermarken 1989]. Все переменные были проверены на наличие единичных корней (нестационарность) тестами первого поколения (Приложение 3).

⁶ В 2018 г. городу Шымкент был придан статус города республиканского значения и он стал отдельным регионом страны. Однако временные ряды за ранние периоды пока не пересчитаны. В исследовании используется Южно-Казахстанская область как сумма Туркестанской области и города Шымкент.

Тесты на наличие единичных корней проверялись с учетом тренда и с максимальным числом лагов – 2. В случае стационарной переменной – в уравнении будет использоваться ее первая разность.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

В настоящем исследовании будут проанализированы факторы, объясняющие средний балл ЕНТ при сдаче на казахском и русском языках, т.е. будут построены модели для двух эндогенных переменных ENT_kz_score и ENT_ru_score.

Для оценки уравнений регрессии применялся стандартный метод наименьших квадратов (МНК) и двухшаговый метод наименьших квадратов (2МНК). База данных, представляющая собой панельные данные, позволила построить следующие виды моделей:

1. Модель по объединенной выборке (pooling regression);
2. Модель с фиксированными эффектами (fixed effects): индивидуальные, временные и совместные эффекты;
3. Модель со случайными эффектами (random effects): индивидуальные, временные и совместные эффекты.
4. Модель инструментальных переменных (instrumental variables).

Указанные модели являются классическими при работе с панельными данными, поэтому в настоящей работе будет приведен их общий вид. В общем виде **модель по объединенной выборке** имеет вид [Baltagi 2005]:

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \quad i = 1, \dots, N; t = 1, \dots, T \quad (1)$$

где y_{it} – вектор i зависимой переменной в период времени t , X'_{it} – вектор независимой переменной u_{it} – соответствующая ошибка.

Приводим (1) в матричную форму:

$$y = X \times \beta + u \quad (2)$$

где β – неизвестный вектор размера $k \times 1$.

Модель с фиксированными эффектами имеет общий вид:

$$y = \alpha i_{NT} + X\beta + Z_{\mu}\mu + \vartheta = Z\delta + Z_{\mu}\mu + \vartheta \quad (3)$$

Уравнение (3) оценивается с помощью МНК, чтобы получить параметры α , β , μ . Вектор Z имеет размерность $NT \times (K+1)$; Z_{μ} является матрицей индивидуальных фиктивных переменных с размерностью $NT \times N$. Если N бесконечно большое, то уравнение (3) будет включать большое множество индивидуальных фиктивных переменных, а матрица, которая должна быть инвертирована МНК, является большой и имеет размерность $(N+K)$. Фактически, поскольку α и β являются искомыми параметрами, которые можно оценить с помощью МНК из (3), предварительно умножив модель на Q и выполнив оценку МНК на полученной преобразованной модели:

$$Qy = QX\beta + Q\vartheta \quad (4)$$

Из (1) **модель со случайными эффектами** имеет общий вид:

$$y_{it} = \mu + X'_{it}\beta + w_{it} \quad \beta \in R^k, \mu \in R \quad (5)$$

где ошибка $w_{it} = u_i + u_{it}$, μ – константа, u_i – случайная ошибка, инвариантная по времени для каждого объекта. Параметры μ и β оцениваются с помощью МНК.

Важным допущением в панельных моделях является некоррелированность объясняющих переменных и случайной ошибки, т.е. регрессоры должны быть экзогенными. Невыполнение этого условия может привести к смещенным и несостоятельным оценкам при применении стандартных методов оценки регрессии (МНК), следовательно, к ошибочным выводам. Во многих случаях данное условие не выполняется и регрессоры называются эндогенными. Зависимость регрессоров и ошибок может возникать по разным причинам: (1) пропущенные существенные переменные, (2) ошибки измерения регрессоров, (3) самоотбор, (4) одновременность, (5) серийная корреляция ошибок при наличии лагированной зависимой переменной среди регрессоров. Ruud (2000) показывает, что проблемы (2)-(5) можно рассматривать как частные случаи (1) [Ebbes 2007]. Мы не будем приводить алгебраическое выражение каждой причины эндогенности, что легко можно найти в теоретических материалах, а для решения вопросов эндогенности применим поиск **инструментальных переменных** с помощью оценки двухшаговым методом наименьших квадратов (2МНК). В инструментальной переменной главная проблема – большое количество инструментов. Поэтому мы не будем использовать множество инструментов, а ограничимся одним. Нет необходимости приводить все условия и проблемы эндогенности, однако, для решения этого вопроса мы должны найти инструментальную переменную, которая была бы хорошо коррелирована с зависимой переменной и плохо – с ошибкой уравнения.

На **применимость модели объединенной выборки** проводится тест, с помощью которого проверяется гипотеза, что одинаковые коэффициенты применимы к каждой переменной [Croissant, Millo 2008]. Применяется стандартный F-тест, основанный на сравнении модели объединенной выборки, и модели на основе оценки каждой переменной. Нулевая гипотеза говорит, что набор данных может быть пулируемым, т.е. переменные имеют одинаковые коэффициенты наклона. Поэтому, если $p < 0.05$, то нулевая гипотеза отклоняется, и осуществляется переход к оценке моделей с фиксированными эффектами.

В моделях фиксированных эффектов необходимо сделать **выбор между временными и индивидуальными эффектами** с помощью тестов F-тест для индивидуальных эффектов и тестов временных эффектов множителей Лагранжа Бреуша-Пагана [Breusch, Pagan 1980]. Для данных тестов нулевая гипотеза означает не использование временных фиксированных эффектов, а использование индивидуальных или двусторонних эффектов, альтернативная гипотеза – наоборот. Поэтому, если $p < 0.05$, то отвергается нулевая гипотеза, и необходимо использование временных фиксированных эффектов.

Для выявления более состоятельных эффектов был проведен **тест на спецификацию** Хаусмана [Hausman 1978], где тестировались модели с фиксированными и случайными эффектами. Гипотеза H_0 предполагает состоятельность оценок модели с фиксированными эффектами и состоятельность и эффективность оценок модели со случайными эффектами; гипотеза H_1 говорит о состоятельности оценок модели с фиксированными эффектами и несостоятельности оценок модели со случайными эффектами.

Результаты **моделей случайных эффектов сопоставляются с результатами модели объединенной выборки** с использованием теста множителей Лагранжа Бреуша-Пагана [Breusch, Pagan 1980]. Нулевая гипотеза гласит, что разница между переменными равна нулю, т.е. нет необходимости применять панель, а достаточна обычная МНК регрессия – модель объединенной выборки. Не имеются существенных различий между переменными, то есть нет эффекта панели. Альтернативная гипотеза состоит в том, что есть панельные эффекты и применимы случайные эффекты.

Далее «наилучшая» модель, полученная после применения тестов Хаусмана и Бреуша-Пагана должна быть проверена **на перекрестную зависимость** с помощью теста множителей Лагранжа Бреуша-Пагана на независимость и Песаран тест на перекрестную зависимость [Pesaran 2004]. Нулевая гипотеза в этих тестах говорит, что остатки по всей совокупности некоррелированы. Тесты используются для проверки корреляции остатков по всей совокупности. Перекрестная зависимость может привести к байесу в результатах тестов (также известным сейчас в качестве современной корреляции). Однако, согласно Балтаджи [Baltagi 2005], проблема перекрестной зависимости – проблема макропанелей с длинными временными рядами (межстрановая модель), тогда как в нашем случае используется микропанель (данные внутри страны). Следовательно, нет необходимости проверки на **серийную корреляцию** теста Бреуша-Годфри [Wooldridge 2002].

Модель необходимо проверить на наличие **гетероскедастичности** с помощью теста Бреуша-Пагана [Breusch, Pagan 1979]. Нулевая гипотеза говорит о наличии гомоскедастичности данных, альтернативная свидетельствует о гетероскедастичности. При наличии гетероскедастичности необходимо использование робастной ковариационной матрицы [Kleiber, Zeileis 2008] или обобщенного МНК с неизвестной формой дисперсии [Yobero 2016].

Для переменных в моделях была проведена логарифмическая трансформация, что является частным случаем преобразования Бокса-Кокса. Многие переменные не нашли применения в моделировании, потому что они не показали значительные связи, либо при наличии связи были отвергнуты по статистическим или теоретическим причинам.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Модель баллов ЕНТ на казахском языке

Зависимой переменной в уравнении объяснения среднего балла ЕНТ с казахским языком сдачи является переменная *ENT_kz_score*. Для построения модели объединенной выборки проверено множество переменных из нашего набора данных, и в итоге получена следующая модель (таблица 3).

Таблица 3. Оценка модели среднего балла ЕНТ на казахском языке по объединенной выборке

	Зависимая переменная <i>ENT_kz_score</i>
Константа	1.357861 (0.541446)*
log(real_medianInc)	0.144974 (0.044189)**
log(internet)	0.309606 (0.083698)***
log(altyn_belgy)	0.038199 (0.011214)**
Общая сумма квадратов	0.71167
Сумма квадратов остатков	0.45352
R ²	0.36274
Скоп. R ²	0.33758
F-statistic	14.4201
p-value	1.5846e-07

., *, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

В модели объединенной выборки были получены знаки коэффициентов регрессии, соответствующие теоретическим и логическим предпосылкам. На рост показателя успешности сдачи ЕНТ на казахском языке влияют теоретические предпосылки: реальный медианный доход домохозяйств, число пользователей интернетом и число выпускников школ со знаком отличия «Алтын белгі». При росте на 1% реального медианного дохода семьи средний балл ЕНТ увеличивается на 0.14%; увеличение на 1% числа пользователей интернета приводит к росту среднего балла ЕНТ на 0.31%; при росте на 1% числа выпускников средний балл ЕНТ увеличивается на 0.04%.

Следующим шагом явилось тестирование на пулируемость статистических данных.

F statistic

$data: \log(ENT_kz_score) \sim \log(real_medianInc) + \log(internet) + \log(altyn_belgy)$
 $F = 7.0831, df1 = 60, df2 = 16, p\text{-value} = 4.507e-05$
alternative hypothesis: instability

Результаты F-теста говорят, что отвергается нулевая гипотеза о пулируемости данных и необходимо перейти к исследованию фиксированных эффектов.

Таблица 4. Оценка модели среднего балла ЕНТ на казахском языке с фиксированными эффектами

	Fixed (individual)	Fixed (time)	Fixed (two ways)
log(real_medianInc)	-0.225145 (0.077434)**	0.177641 (0.046289)***	-0.083006 (0.115011)
log(internet)	0.145965 (0.059011)*	0.214057 (0.096692)*	0.061928 (0.060783)
log(altyn_belgy)	0.033760 (0.010266)**	0.036375 (0.012418)**	0.012761 (0.014323)
Общая сумма квадратов	0.19218	0.60943	0.089935
Сумма квадратов остатков	0.10891	0.41889	0.086901
R ²	0.43332	0.31265	0.03373
Скоп. R ²	0.2661	0.24583	-0.33922
F-statistic	15.548	10.9168	0.663249
p-value	1.2704e-07	5.4121e-06	0.57807

*, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

Из таблицы 4 следует, что наиболее статистически значимыми и соответствующими теоретическим знакам явились оценки фиксированных временных эффектов. Двусторонние фиксированные эффекты показали отсутствие значимости уравнения в целом, а также оценок регрессоров. Индивидуальные временные эффекты показали статистическую значимость, однако переменная реального медианного дохода домохозяйства не соответствует теоретическим предпосылкам, так как рост дохода, наоборот, должен стимулировать увеличение «инвестирования» родителями в образование своих детей.

Далее необходимо протестировать значимость эффектов с помощью тестов: F-тест для индивидуальных эффектов и тест временных эффектов множителей Лагранжа Бреуша-Пагана [Breusch, Pagan 1980].

F test for individual effects

data: log(ENT_kz_score) ~ log(real_medianInc) + log(internet) + log(altyn_belgy)
 F = 15.784, df1 = 11, df2 = 61, p-value = 5.217e-14
 alternative hypothesis: significant effects

Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced panels

data: log(ENT_kz_score) ~ log(real_medianInc) + log(internet) + log(altyn_belgy)
 chisq = 0.0019136, df = 1, p-value = 0.9651
 alternative hypothesis: significant effects

Результаты тестов говорят, что фиксированные временные эффекты не значимы и необходимо выбрать фиксированные индивидуальные эффекты или перейти к моделированию случайных эффектов.

Таблица 5. Оценка модели среднего балла ЕНТ на казахском языке со случайными эффектами

	Random (individual)	Random (time)	Random (two ways)
Константа	3.816259 (0.733537)***	1.357861 (0.541446)*	4.002602 (0.743829)***
log(real_medianInc)	-0.047045 (0.062428)	0.144974 (0.044189)**	-0.062784 (0.063339)
log(internet)	0.204443 (0.059531)***	0.309606 (0.083698)***	0.199044 (0.058847)***
log(altyn_belgy)	0.034295 (0.010255)***	0.038199 (0.011214)***	0.034209 (0.010167)***
Общая сумма квадратов	0.23107	0.71167	0.22539
Сумма квадратов остатков	0.15374	0.45352	0.14848
R ²	0.33466	0.36274	0.34124
Скорр. R ²	0.3084	0.33758	0.31524
χ ²	38.228	43.2604	39.3687
p-value	2.5292e-08	2.1668e-09	1.45e-08

., *, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

Результаты применения случайных эффектов для определения факторов ЕНТ на казахском языке показывают, что все случайные эффекты статистически значимы. Однако в индивидуальных и двусторонних случайных эффектах коэффициент регрессии реального медианного дохода домохозяйства статистически незначим, и его знак не соответствует теоретическим предпосылкам, так как рост дохода, наоборот, должен стимулировать увеличение «инвестирования» родителями в образование своих детей. В уравнении случайных временных эффектов коэффициенты регрессии значимы и соответствуют теоретическим предпосылкам: при увеличении на 1% реального медианного дохода домохозяйства средний балл ЕНТ растет на 0.14%; рост на 1% пользователей интернета увеличивает средний балл ЕНТ на 0.31%; число выпускников с «Алтын белгі» повышает средний балл ЕНТ на 0.04%.

Для выявления более состоятельных эффектов был проведен тест на спецификацию Хаусмана [Hausman 1978], где тестировались модели с фиксированными индивидуальными и случайными временными эффектами.

Hausman Test

$data: \log(ENT_kz_score) \sim \log(real_medianInc) + \log(internet) + \log(altyn_belgy)$
 $chisq = 32.3, df = 3, p\text{-value} = 4.524e-07$
alternative hypothesis: one model is inconsistent

Результат теста Хаусмана показывает, что отвергается нулевая гипотеза и утверждается состоятельность оценок модели с фиксированными индивидуальными эффектами и несостоятельность оценок модели со случайными временными эффектами.

С помощью теста множителей Лагранжа Бреуша-Пагана [Breusch, Pagan 1980] необходимо определить, какая из двух моделей подходит: объединенной выборки или с фиксированными индивидуальными эффектами. Понятно, что

модель согласно F-тесту, не может быть пулируема (объединена). Но необходимо соблюсти все тесты.

Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels
data: $\log(ENT_kz_score) \sim \log(real_medianInc) + \log(internet) + \log(altyn_belgy)$
chisq = 52.787, df = 1, p-value = 3.717e-13
alternative hypothesis: significant effects

Результат теста показывает, что принимается альтернативная гипотеза и должна быть использована модель случайных эффектов.

Модель среднего балла ЕНТ на казахском языке не может быть представлена моделью объединенной регрессии (результат F-теста на пулируемость статистических данных), тест множителей Лагранжа Бреуша-Пагана отдает предпочтение модели со случайными временными эффектами по сравнению с моделью объединенной выборки, Хаусман-тест показывает значимость оценок фиксированных индивидуальных эффектов по сравнению с оценками случайных временных эффектов. Таким образом, тесты указывают на модель фиксированных индивидуальных эффектов, где переменная *real_medianInc*, дает знак коэффициента регрессии не в соответствии с теоретическими предположениями, так как реальный доход домохозяйств должен приводить к росту балла ЕНТ, а не к его снижению.

Далее, модель фиксированных эффектов необходимо проверить на наличие гетероскедастичности с помощью теста Бреуша-Пагана [Breusch, Pagan 1979].

Breusch-Pagan test
data: *fixed_ENT_kz_score_IE*
BP = 11.914, df = 3, p-value = 0.007682

Мы отвергаем нулевую гипотезу о гомоскедастичности случайных остатков и принимаем альтернативную гипотезу о наличии гетероскедастичности.

Кроме того, зависимая переменная (*ENT_kz_score*) и одна из объясняющих переменных (*altyn_belgy*) были взяты из одного источника, что подразумевает пропуск существенных переменных между ними, а значит - наличие эндогенности в модели случайных временных эффектов.

Переходим к применению 2МНК для построения уравнения регрессии.

Таблица 6. Оценка модели среднего балла ЕНТ на казахском языке с инструментальной переменной

	Зависимая переменная <i>ENT_kz_score</i>
Константа	1.45973 (0.60287)*
log(real_medianInc)	0.13153 (0.04618)**
log(internet)	0.34629 (0.07775)***
log(altyn_belgy)	0.01343 (0.01250)
Weak instruments	< 2e-16 ***
Wu-Hausman	0.00512 **
Остатки стандартной ошибки	0.07969
Множественный R2	0.3218
Скорректированный R2	0.2951
Тест Вальда	29.7
p-value	1.593e-06

., *, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

Как видно из таблицы 6, в модели с инструментальной переменной выбранный инструмент *exc_student* для эндогенной переменной *altyn_belgy* показал приемлемость, и была отвергнута нулевая гипотеза, что инструмент слабый [Stock, Yogo 2005]. Данная переменная в качестве инструмента имеет достаточно сильную природу поддержки: *exc_student* представляет собой количество учащихся, имеющих успеваемость на «отлично», а *altyn_belgy* – это отличники, подтвердившие данный статус результатами ЕНТ. Согласно тесту Ву-Хаусмана [Hausman 1978], принимается нулевая гипотеза, что все переменные экзогенны.

Поскольку предыдущая модель случайных временных эффектов имела гетероскедастичность остатков, то, очевидно, что полученная модель инструментальной переменной имеет гетероскедастичность.

В эконометрике для исправления гетероскедастичности применяется коррекция стандартных ошибок. Hausman, Newey, Woutersen, Chao, Swanson (2012) предлагают корректировать стандартные ошибки в моделях с инструментальными переменными, т.е. нахождение робастных стандартных ошибок.

Оценка ковариационной матрицы, согласованная с гетероскедастичностью, является распространенным инструментом, данную идею первоначально представили Huber (1967), Eicker (1963) и White (1980). Таким образом, для искомой модели среднего балла ЕНТ на казахском языке с инструментальной переменной будут подобраны робастные стандартные ошибки с помощью робастной ковариационной матрицы или «сэндвич оценки» [Kleiberg, Zeileis 2008].

Таблица 7. Стандартные ошибки переменных модели с инструментальной переменной скорректированные с помощью робастной ковариационной матрицы

Условия гетероскедастичности	Константа	$\log(\text{real_medianInc})$	$\log(\text{internet})$	$\log(\text{altyn_belgy})$
HC0	0.6028728	0.04617759	0.07774768	0.01250072
HC1	0.6185345	0.04737721	0.07976744	0.01282547
HC2	0.6295914	0.04800731	0.08222176	0.01292625
HC3	0.6592208	0.04994636	0.08740996	0.01339149
HC4	0.6813528	0.04972279	0.09615903	0.01355858

Все типы условий гетероскедастичности предполагают отсутствие внутригрупповой (последовательной) корреляции между ошибками и учитывают гетероскедастичность по группам (временные периоды). Что касается ковариационной матрицы ошибок для каждой отдельной группы наблюдений, то метод Arellano (1987) допускает полностью общую структуру с гетероскедастичностью и серийной корреляцией.

Таблица 8. Скорректированные стандартные ошибки модели с инструментальной переменной с помощью Arellano (1987)

	Зависимая переменная ENT_kz_score
Константа	1.459727 (0.659221)*
$\log(\text{real_medianInc})$	0.131527 (0.049946)**
$\log(\text{internet})$	0.346291 (0.087410)***
$\log(\text{altyn_belgy})$	0.0134328 (0.01391)

., *, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

С помощью метода Arellano (1987) были скорректированы стандартные ошибки. Если сравнить результаты стандартных ошибок с моделью в таблице 6, то видно, что ошибки были скорректированы для всех переменных модели.

В итоге, при увеличении на 1% реального медианного дохода домохозяйства средний балл ЕНТ при сдаче на казахском языке растет на 0.13%, при росте на 1% доли пользователей интернета средний балл ЕНТ увеличивается на 0.35%. Число учащихся, подтвердивших знак отличия «Алтын белгі», оказалось статистически незначимым после применения инструментальной переменной, что можно объяснить тем, что подтверждение знака отличия «Алтын белгі» происходит только по результатам самого ЕНТ, а не до него. То есть ЕНТ является следствием, а не причиной. В связи с этим, переменные должны поменяться местами, однако такой анализ выходит за рамки данного исследования.

Очевидно, что построенная модель не охватывает все возможные факторы, которые могут объяснить возможность получения высокого балла ЕНТ, так как официальная статистика не может полностью охватить генеральную совокупность сплошным учетом. В век информационных технологий фактор пользования

интернетом и уровень дохода родителей являются базовыми условиями успешности сдачи выпускного экзамена, что показал вышеизложенный эмпирический материал.

Модель баллов ЕНТ на русском языке

Зависимой переменной в уравнении объяснения среднего балла ЕНТ с русским языком сдачи является переменная *ENT_ru_score*. Для построения модели объединенной выборки было проверено множество переменных из нашего набора данных, в результате чего получена следующая модель (таблица 9).

Таблица 9. Оценка модели среднего балла ЕНТ на русском языке по обобщенной регрессии

	Зависимая переменная <i>ENT_ru_score</i>
Константа	4.1426296 (0.1213962)***
log(poverty)	-0.0653084 (0.0106363)***
log(fract)	0.0321194 (0.0096083)**
log(rooms)	0.0359581 (0.0135939)**
Общая сумма квадратов	0.32318
Сумма квадратов остатков	0.19868
R ²	0.38523
Скоп. R ²	0.36096
F-statistic	15.8742
p-value	4.1619e-08

., *, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

В модели объединенной выборки были получены знаки коэффициентов регрессии, соответствующие теоретическим и логическим предпосылкам. На рост показателя успешности сдачи ЕНТ на русском языке влияют теоретические предпосылки: уровень бедности, коэффициент фракционализации и количество учебных кабинетов.

В литературе этническая неоднородность измеряется с помощью ряда показателей, включающих индекс фракционализации, энтропию и индекс поляризации. Коэффициент фракционализации показывает этническую неоднородность страны/региона. Он рассчитывается следующим образом [Буфетова, Коломак, Михалёва 2017]:

$$F = 1 - \sum_{i=1}^N (n_i)^2 \quad (6)$$

где n_i – доля населения i -й этнической группы в общей численности населения, N – число таких групп. Индекс принимает значения от 0 до $1-1/N$ и показывает вероятность того, что два случайно выбранных человека принадлежат к разным национальностям. Другими словами, показатель моноэтнического региона стремится к 0, а полиэтничных – к 1. Более моноэтническими в Республике Казахстан являются

южные и западные регионы, тогда как северные, восточные и центральные – полиэтничными.

Согласно модели объединенной выборки при росте на 1% уровня бедности средний балл ЕНТ снижается на 0.07%; увеличение на 1% коэффициента фракционализации увеличивает средний балл ЕНТ на 0.03%; рост на 1% количества учебных кабинетов приводит к росту среднего балла ЕНТ на 0.036%.

Следующим шагом явилось тестирование на пулируемость статистических данных.

F statistic

data: log(ENT_ru_score) ~ log(poverty) + log(fract) + log(rooms)

F = 3.351, df1 = 60, df2 = 16, p-value = 0.004847

alternative hypothesis: unstability

Результаты F-теста говорят, что отвергается нулевая гипотеза о пулируемости данных и необходимо перейти к исследованию фиксированных эффектов.

Таблица 10. Оценка модели среднего балла ЕНТ на русском языке с фиксированными эффектами

	Fixed (individual)	Fixed (time)	Fixed (two ways)
log(poverty)	-0.023911 (0.020500)	-0.0670899 (0.0114463)***	-0.013708 (0.030609)
log(fract)	0.561697 (0.270939)*	0.0321678 (0.0096953)**	0.453293 (0.347773)
log(rooms)	0.082888 (0.090533)	0.0366437 (0.0137930)**	0.044016 (0.099931)
Общая сумма квадратов	0.11404	0.30787	0.098735
Сумма квадратов остатков	0.10099	0.19127	0.095721
R ²	0.11449	0.37874	0.03052
Скоп. R ²	-0.14681	0.31834	-0.34367
F-statistic	2.62882	14.6313	0.598139
p-value	0.058165	1.5537e-07	0.61884

., * , ** , *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

В таблице 10 видно, что наиболее статистически значимыми и соответствующими теоретическим знакам явились оценки фиксированных временных эффектов. Двусторонние фиксированные эффекты показали отсутствие статистической значимости.

Необходимо протестировать значимость эффектов с помощью тестов: F-тест для индивидуальных эффектов и тест временных эффектов множителей Лагранжа Бреуша-Пагана [Breusch, Pagan 1980].

F test for individual effects

data: log(ENT_ru_score) ~ log(poverty) + log(fract) + log(rooms)

F = 0.78364, df1 = 4, df2 = 57, p-value = 0.5405

alternative hypothesis: significant effects

Lagrange Multiplier Test - time effects (Breusch-Pagan) for balanced panels
data: log(ENT_ru_score) ~ log(poverty) + log(fract) + log(rooms)
chisq = 0.44426, df = 1, p-value = 0.5051
alternative hypothesis: significant effects

Результаты тестов говорят о принятии нулевой гипотезы, что фиксированные временные эффекты незначимы и необходимо выбрать фиксированные индивидуальные эффекты или перейти к моделированию случайных эффектов.

Таблица 11. Оценка модели среднего балла ЕНТ на русском языке со случайными эффектами

	Random (individual)	Random (time)	Random (two ways)
Константа	4.186128 (0.197058)***	4.1423296 (0.1212787)***	4.184060 (0.194784)***
log(poverty)	-0.052665 (0.013812)***	-0.0654217 (0.0106691)***	-0.052853 (0.014582)***
log(fract)	0.032819 (0.016099)*	0.0321215 (0.0095962)***	0.032599 (0.015846)*
log(rooms)	0.030010 (0.021921)	0.0360025 (0.0135817)**	0.030233 (0.021698)
Общая сумма квадратов	0.16273	0.32208	0.1571
Сумма квадратов остатков	0.13081	0.19815	0.12828
R ²	0.19614	0.38477	0.18347
Скорр. R ²	0.16441	0.36048	0.15124
χ ²	18.5437	47.5302	17.0771
p-value	0.00033969	2.6806e-10	0.0006814

., * , ** , *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

Результаты применения случайных эффектов для определения факторов ЕНТ на русском языке показывают, что все случайные эффекты статистически значимы (таблица 11). Однако, в индивидуальных и двусторонних случайных эффектах коэффициент регрессии количества учебных кабинетов статистически незначим. В уравнении случайных временных эффектов коэффициенты регрессии значимы и соответствуют теоретическим предпосылкам: при снижении на 1% уровня бедности средний балл ЕНТ растет на 0.07%; рост на 1% этнического многообразия (*fract*) приводит к росту среднего балла ЕНТ на 0.03%; увеличение на 1% количества учебных кабинетов увеличивает балл ЕНТ на 0.04%.

Для выявления более состоятельных эффектов был проведен тест на спецификацию Хаусмана [Hausman 1978], где тестировались модели с фиксированными индивидуальными и случайными временными эффектами.

Hausman Test
data: log(ENT_ru_score) ~ log(poverty) + log(fract) + log(rooms)
chisq = 8.1524, df = 3, p-value = 0.04296
alternative hypothesis: one model is inconsistent

Результат теста Хаусмана показывает, что отвергается нулевая гипотеза и утверждается состоятельность оценок модели с фиксированными индивидуальными эффектами и несостоятельность оценок модели со случайными временными эффектами.

С помощью теста множителей Лагранжа Бреуша-Пагана [Breusch, Pagan 1980] необходимо определить какая из двух моделей подходит: объединенной выборки или фиксированные индивидуальные эффекты. Понятно, что модель согласно F-тесту, не может быть пулируема (объединена). Но необходимо соблюсти все тесты.

Lagrange Multiplier Test - (Breusch-Pagan) for balanced panels

data: log(ENT_ru_score) ~ log(poverty) + log(fract) + log(rooms)

chisq = 13.889, df = 1, p-value = 0.000194

alternative hypothesis: significant effects

Результат теста показывает, что принимается нулевая гипотеза и используется модель случайных эффектов.

Модель среднего балла ЕНТ на русском языке не может быть представлена моделью объединенной регрессии (результат F-теста на пулируемость статистических данных), тест множителей Лагранжа Бреуша-Пагана отдает предпочтение модели случайных временных эффектов по сравнению с моделью объединенной выборки, Хаусман-тест показывает значимость оценок фиксированных индивидуальных эффектов по сравнению с оценками случайных временных эффектов. Таким образом, тесты указывают на «лучшую модель» – с фиксированными индивидуальными эффектами.

Модель фиксированных эффектов необходимо проверить на наличие гетероскедастичности с помощью теста Бреуша-Пагана [Breusch, Pagan 1979].

Breusch-Pagan test

data: fixed_ENT_ru_score_IE

BP = 1.2623, df = 3, p-value = 0.7381

Принимается нулевая гипотеза о гомоскедастичности случайных остатков. Однако, зависимая переменная (*ENT_ru_score*) и одна из объясняющих переменных (*rooms*) были взяты из одного источника, что подразумевает пропуск существенных переменных между ними, а значит – наличие эндогенности в модели случайных временных эффектов.

Переходим к применению 2МНК для построения уравнения регрессии.

Таблица 12. Оценка модели среднего балла ЕНТ на русском языке с инструментальной переменной

	Зависимая переменная <i>ENT_ru_score</i>
Константа	4.81037 (0.38662)***
log(poverty)	-0.04213 (0.01765)*
log(fract)	0.03542 (0.01229)**
log(rooms)	-0.03934 (0.04345)
Weak instruments	0.00121**
Wu-Hausman	0.01322*
Остатки стандартной ошибки	0.06058
Множественный R2	0.1371
Скорректированный R2	0.103
Тест Вальда	26.13
p-value	8.955e-06

., *, **, *** — значимость на 10%, 5%, 1% и 0.1% уровне соответственно

Как видно из таблицы 12, в модели с инструментальной переменной выбранный инструмент *newroom_school* для эндогенной переменной *rooms* показал свою приемлемость, и была отвергнута нулевая гипотеза при 99% уровне значимости, что инструмент слабый [Stock, Yogo 2005]. Данная переменная в качестве инструмента имеет достаточно сильную природу поддержки: *newroom_school* представляет собой количество дневных государственных общеобразовательных школ, имеющих кабинеты новой модификации (химии, биологии, физики, лингафонных мультимедийных кабинетов) с сервисным обслуживанием, а *rooms* – количество учебных кабинетов, т.е. переменные имеют гомогенную природу. Согласно тесту Ву-Хаусмана [Hausman 1978], при 95% уровне значимости принимается нулевая гипотеза, что все переменные экзогенны.

Таким образом, моделирование с помощью инструментальной переменной привело к следующим результатам: при снижении на 1% уровня бедности балл ЕНТ при сдаче на русском языке увеличивается на 0.04%, при росте на 1% этнического многообразия региона средний балл ЕНТ увеличивается на 0.035%. Переменная *rooms* не показала свою статистическую значимость.

Достижение высокого балла ЕНТ при сдаче на казахском и русском языках зависит от фактора благосостояния населения – медианный доход и уровень бедности. В остальном факторы различны: если балл ЕНТ на казахском языке зависит от навыков использования интернета, то на русском языке – от этнической разнородности, что связано с тем, что нетитульные этносы в основном сдают ЕНТ на русском языке.

ЗАКЛЮЧЕНИЕ

1. Данные по образованию были получены из одной базы данных, однако результаты моделирования показали различность факторов при сдаче ЕНТ на казахском и русском языках. Кроме того, в модели ЕНТ на казахском языке наблюдалась гетероскедастичность остатков, тогда как в модели ЕНТ на русском языке – гомоскедастичность остатков.

2. При увеличении на 1% реального медианного дохода домохозяйства средний балл ЕНТ при сдаче на казахском языке растет на 0.13%, рост на 1% доли пользователей интернета увеличивает средний балл ЕНТ на 0.35%.

3. При снижении на 1% уровня бедности балл ЕНТ при сдаче на русском языке увеличивается на 0.04%, при росте на 1% этнического многообразия региона средний балл ЕНТ увеличивается на 0.035%.

4. Подтверждена гипотеза, что социально-экономический статус и бедность являются важными факторами объяснения различия в образовательных достижениях: высокий балл ЕНТ при сдаче на казахском и русском языках зависит от медианного дохода домохозяйств и уровня бедности соответственно, так как возможности родителей действуют как инвестиции в детей и, возможно, будут способствовать поступлению в престижный ВУЗ и являться хорошим стартом для будущей взрослой жизни. В остальном факторы различны: если балл ЕНТ на казахском языке зависит от навыков использования интернета в век информационных технологий, то на русском языке – от этнической разнородности, что связано с тем, что нетитульные этносы в основном сдают ЕНТ на русском языке.

5. Очевидно, что модели не охватывают все возможные факторы, которые могут объяснить возможность получения высокого балла ЕНТ, так как официальная статистика не может полностью охватить генеральную совокупность сплошным учетом. При наличии данных с высокой частотностью (месяц, квартал) возможно получение других результатов.

6. Рекомендации для будущего продолжения исследования:

а) внести в модели переменные нового содержания и трехязычия и рассчитать влияния их эффектов на возможность получения хороших результатов по ЕНТ. Только с 2019/2020 учебного года общее среднее образование полностью перешло на новое содержание и трехязычие;

б) внести в модели переменные подушевого финансирования. Оно действует только в режиме пилотной апробации.

Список литературы и использованных источников

1. Буфетова А., Коломак Е., Михалёва М. (2017). Национальное разнообразие и экономическое развитие регионов России // Мир экономики и управления. Т. 17, №3. С. 143–157
2. Aaronson D., Barrow L., Sander W. (2007). Teachers and Student Achievement in the Chicago Public High Schools // Journal of Labor Economics, 2007, vol. 25, no.1
3. Angrist J., Lavy V. (2001). Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools // Journal of Labor Economics 19, no. 2:343–69
4. Angrist J., Krueger A. (1991). Does Compulsory School attendance affect Schooling and Earnings? // The Quarterly Journal of Economics, November, V.106, Issue 4: 979-1014
5. Arouri M., Rault C. (2014). An Econometric Analysis of the Impact of Oil Prices on Stock Markets in Gulf Cooperation Countries // Emerging Markets and the Global Economy, 161–178
6. Arellano M. (1987). Computing robust standard errors for within-group estimators // Oxford Bulletin of Economics and Statistics, 49(4), pp. 431-434
7. Arrow K. (1973). Higher Education as a Filter // Journal of Public Economics 2: 193-216
8. Ashenfelter O., Krueger A. (1994) Estimates of the Economic Return to Schooling from a New Sample of Twins // The American Economic Review, Vol. 84, No. 5: 1157-73
9. Aucejo M., Romano T. (2016). Assessing the effect of school days and absences on test score performance // Economics of Education Review 55: 70–87
10. Bai J., Ng S. (2004). A PANIC attack on unit roots and cointegration // Econometrica 72, 1127–1177
11. Baltagi B. (2005). Econometric Analysis of Panel Data. John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England
12. Banerjee A., Cole S., Duflo E., Linden L. (2007). Remedying Education: Evidence from two Randomized Experiments in India // The Quarterly Journal of Economics, August, V.122, Issue 3: 1235–64
13. Barro R. (1991). Economic Growth in a Cross Section of Countries // The Quarterly Journal of Economics, Vol. 106, No. 2. May: 407-43
14. Barro R., Sala-i-Martin X. (1995). Economic Growth. McGraw-Hill, 539 p
15. Benhabib J., Spiegel M. (1994). The role of human capital in economic development Evidence from aggregate cross-country data // Journal of Monetary Economics 34: 143-73
16. Bilal M., Klenow P. (2000). Does Schooling Cause Growth? // The American Economic Review 90(5): 1160-83
17. Black S., Devereux P., Salvanes K. (2004). The More the Merrier? The Effect of Family Composition on Children's Education // IZA Discussion Papers, No. 1269, Institute for the Study of Labor (IZA), Bonn
18. Breusch T., Pagan A. (1979). A Simple Test for Heteroskedasticity and Random Coefficient Variation // Econometrica. 47(5): 1287–1294
19. Breusch T., Pagan A. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics // Review of Economic Studies, 47, 239–253
20. Brooks-Gunn J., Duncan G. (1997). The Consequences of Growing Up Poor. New York: Russell Sage
21. Brooks-Gunn J., Duncan G. (2000). Family Poverty, Welfare Reform and Child Development // Child Development 71:1, 188–196
22. Brooks-Gunn J., Duncan G., Klebanov P. (1994). Economic Deprivation and Early-Childhood Development // Child Development 65:2, 296–318
23. Brooks-Gunn J., Duncan G., Klebanov P. (1996). Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment and Maternal Characteristics // Child Development 67, 396–408

24. Brooks-Gunn J., Duncan G., Klebanov P., Sealand N., (1993). Do Neighborhoods Influence Child and Adolescent Development? // *American Journal of Sociology* 99:2, 353–395
25. Brooks-Gunn J., Klebanov P., Duncan G. (1995). Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment, and Maternal Characteristics // *Child Development* 67:2, 396–408
26. Brooks-Gunn J., Klebanov P., Liaw F., Spiker D. (1993). Enhancing the Development of Low Birth-Weight, Premature Infants: Changes in Cognition and Behavior over the First Three Years // *Child Development* 63, 736–753
27. Cameron S., Heckman J. (1998). Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males // *Journal of Political Economy*, vol. 106, no. 2: 262-333
28. Card D., Krueger A. (1992). Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States // *The Journal of Political Economy*, Vol. 100, No. 1. Feb.: 1-40
29. Card D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems // *Econometrica*, vol. 69, No. 5. Sep.: 1127-60
30. Carter S., Greenberg K., Walker M. (2017). The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy // *Economics of Education Review* 56: 118–32
31. Choi I. (2006). Combination unit root tests for cross-sectionally correlated panels // In Corbae D., Durlauf S., Hansen B. (Eds.). *Econometric Theory and Practice: Frontiers of Analysis and Applied Research, Essays in Honor of Peter C.B. Phillips*. Cambridge University Press, Cambridge
32. Coleman J. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office
33. Comi S., Argentin G., Gui M., Origo F., Pagani L. (2017). Is it the way they use it? Teachers, ICT and student achievement // *Economics of Education Review*, vol. 56: 24-39
34. Croissant Y., Millo G. (2008). Panel Data Econometrics in R: The plm Package // *Journal of Statistical Software*, July, vol. 27, Issue 2: 2-43
35. Croissant Y., Millo G. (2019). *Panel Data Econometrics in R*. John Wiley & Sons Ltd
36. Currie J., Moretti E. (2003). Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings // *The Quarterly Journal of Economics*, November, Vol. 118, Issue 4: 1495–1532
37. Ebbes P. (2007). A non-technical guide to instrumental variables and regressor-error dependencies (in Russian) // *Quantile*, issue 2, 3-20
38. Eicker F. (1963). Asymptotic Normality and Consistency of the Least Squares Estimator for Families of Linear Regression // *Annals of Mathematical Statistics*, 34, 447-456
39. Fryer R., Levitt S. (2004). Understanding the black-white test score gap in the first two years of school // *The Review of Economics and Statistics*, vol. LXXXVI, May, number 2: 447-464
40. Goldhaber D., Brewer D. (1997). Why don't school and teachers seem to matter? // *Journal of Human Resources* 32, no. 3:505–23
41. Greenwald R., Hedges L., Laine R. (1996). The effect of school resources on student achievement // *Review of Educational Research* 66:361–96
42. Hanushek E. (1996). Measuring investment in education // *Journal of Economic Perspectives* 10, no. 4:9–30
43. Hanushek E. (1997). Assessing the effects of school resources on student performance: An update // *Education Evaluation and Policy Analysis* 19: 141–64
44. Hanushek E., Kimko D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations // *The American Economic Review* Vol. 90, No. 5, Dec.: 1184-1208
45. Hanushek E. (2002). Publicly provided education // In *Handbook of public finance*, vol. 4, ed. Alan Auerbach and Martin Feldstein. Amsterdam: North-Holland Press

46. Hausman J. (1978). Specification tests in econometrics // *Econometrica*, 46:1251 – 1271
47. Hausman J., Newey W., Woutersen T., Chao J., Swanson N. (2012). Instrumental Variable Estimation With Heteroskedasticity And Many Instruments // *Quantitative Economics*, Volume 3, Issue 2 (July)
48. Hsiao C. (2003). *Analysis of Panel Data*. Cambridge University Press, Cambridge
49. Huber P. (1967). The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions // in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, eds. L. M. LeCam and J. Neyman, Berkeley: University of California Press, pp. 221-233
50. Im K., Lee J. (2001). Panel LM unit root test with level shifts // Discussion paper, Department of Economics, University of Central Florida
51. Im K., Pesaran M., Shin Y. (2003). Testing for unit roots in heterogeneous panels // *Journal of Econometrics*, 115(1):53–74
52. Jacob B., Lefgren L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago // *Journal of Human Resources* 39, no. 1: 50–79
53. Jepsen C., Rivkin S. (2002). What is the tradeoff between smaller classes and teacher quality? // Working paper no. 9205, National Bureau of Economic Research, Cambridge, MA
54. Kane T., Rockoff J., Staiger D. (2006). What does certification tell us about teacher effectiveness? Evidence from New York City // Working paper no. 12155, National Bureau of Economic Research, Cambridge, MA
55. Kane T., Staiger D. (2005). Using imperfect information to identify effective teachers // Working paper, Department of Economics, Dartmouth University
56. Kenkel D. (1991). Health Behavior, Health Knowledge, and Schooling // *Journal of Political Economy*, vol. 99, no. 2: 287-305
57. Kleiber C., Zeileis A. (2008). *Applied Econometrics with R*. Springer
58. Klevmarcken N. (1989). Panel studies: What can we learn from them? // *Introduction*, *European Economic Review* 33, 523–529
59. Levin A., Lin C., Chu C. (2002). Unit root tests in panel data: Asymptotic and finite sample properties // *Journal of Econometrics*, 108:1–24
60. Lochner L., Moretti E. (2004). The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports // *The American Economic Review*, Vol. 94, No. 1, Mar.: 155-189
61. Lleras-Muney A. (2005). The Relationship Between Education and Adult Mortality in the United States // *Review of Economic Studies* 72: 189–221
62. Maddala G., Wu S. (1999). A comparative study of unit root tests with panel data and a new simple test // *Oxford Bulletin of Economics and Statistics*, 61: 631–52
63. Mayer S. (1997). *What Money Can't Buy: Family Income and Children's Life Chances*. Harvard University Press
64. Moon H., Perron B. (2004). Testing for unit root in panels with dynamic factors // *Journal of Econometrics* 122, 81–126
65. Murnane R. (1975). *The impact of school resources on the learning of inner city children*. Cambridge, MA: Ballinger
66. Pesaran M. (2004). *General Diagnostic Tests for Cross Section Dependence in Panels* // Technical Report 1229, CESifo Working Paper Series
67. Pesaran M. (2007). A simple panel unit root test in the presence of cross-section dependence // *Journal of Applied Econometrics*, vol. 22(2), pages 265-312
68. Rivers J., Sanders W. (2002). Teacher quality and equity in educational opportunity: Findings and policy implications. In *Teacher quality*, ed. Lance T. Izumi and Williamson M. Evers. Stanford, CA: Hoover Institution Press
69. Rivkin S., Hanushek E., Kain J. (2005). Teachers, schools, and academic achievement // *Econometrica* 73, no. 2:417–58

70. Rockoff J. (2004). The impact of individual teachers on student achievement: Evidence from panel data // *American Economic Review* 94, no. 2:247–52
71. Romer P. (1986). Increasing Returns and Long-Run Growth // *The Journal of Political Economy*. October, 1002-1037
72. Ruud P. (2000). *An Introduction to Classical Econometric Theory*. New York: Oxford University Press
73. Smith L., Leybourne S., Kim T., Newbold P. (2004). More powerful panel data unit root tests with an application to mean reversion in real exchange rates // *Journal of Applied Econometrics* 19, 147–170
74. Stock J., Yogo M. (2005). Testing for Weak Instruments in Linear IV Regression // In: Andrews DWK *Identification and Inference for Econometric Models*. New York: Cambridge University Press; pp. 80-108
75. White H. (1980). A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity // *Econometrica*, 48, 817-838
76. Wooldridge J. (2002). *Econometric Analysis of Cross-Section and Panel Data*. MIT Press
77. Yobero C. (2016). Methods for Detecting and Resolving Heteroskedasticity // <https://rpubs.com/cyobero/187387>

Приложение 1. Переменные (факторы) модели

№	Переменная	Пояснение
1.	<i>ENT_kz_student</i>	Количество сдающих ЕНТ на казахском языке
2.	<i>ENT_kz_score</i>	Средний балл ЕНТ на казахском языке
3.	<i>ENT_ru_student</i>	Количество сдающих ЕНТ на русском языке
4.	<i>ENT_ru_score</i>	Средний балл ЕНТ на русском языке
5.	<i>school</i>	Количество общеобразовательных школ в стране
6.	<i>student</i>	Численность учащихся в общеобразовательных школах, человек
7.	<i>girl_student</i>	Доля девочек среди учащихся, процентов
8.	<i>teacher</i>	Численность педагогических работников в общеобразовательных школах, человек
9.	<i>girl_teacher</i>	Доля женщин среди педагогических работников в общеобразовательных школах, процентов
10.	<i>rooms</i>	Количество учебных кабинетов, единиц
11.	<i>real_revenues</i>	Объем услуг организаций образования в реальных ценах, млн. тенге
12.	<i>real_budget</i>	Расходы местного бюджета на образование в реальных ценах, млн.тенге
13.	<i>rep_school</i>	Сеть организаций республиканских школ общего среднего образования, единиц
14.	<i>MZ_school</i>	Сеть организаций общего среднего образования по ведомственной принадлежности МЗ РК, единиц
15.	<i>MO_school</i>	Сеть организаций общего среднего образования по ведомственной принадлежности МО РК, единиц
16.	<i>sport_school</i>	Сеть организаций общего среднего образования по ведомственной принадлежности Комитет по делам спорта и физической культуры, единиц
17.	<i>NISH</i>	Сеть Назарбаев интеллектуальные школы
18.	<i>rep_student</i>	Контингент организаций республиканских школ общего среднего образования, человек
19.	<i>MZ_student</i>	Контингент организаций общего среднего образования по ведомственной принадлежности МЗ РК, человек
20.	<i>MO_student</i>	Контингент организаций общего среднего образования по ведомственной принадлежности МО РК, человек
21.	<i>sport_student</i>	Контингент организаций общего среднего образования по ведомственной принадлежности Комитет по делам спорта и физической культуры, человек
22.	<i>NISH_student</i>	Контингент Назарбаев интеллектуальные школы, человек
23.	<i>depth_school</i>	Сеть гимназий, лицеев и школ с классами углубленного изучения предметов (школы с классами углубленного изучения предметов), единиц
24.	<i>gymnasium</i>	Сеть гимназий, лицеев и школ с классами углубленного изучения предметов (гимназии), единиц
25.	<i>lyceum</i>	Сеть гимназий, лицеев и школ с классами углубленного изучения предметов (лицеи), единиц
26.	<i>student_depth_school</i>	Контингент гимназий, лицеев и школ с классами углубленного изучения предметов (школы с классами углубленного изучения предметов), человек
27.	<i>student_gymnasium</i>	Контингент гимназий, лицеев и школ с классами углубленного изучения предметов (гимназии), человек
28.	<i>student_lyceum</i>	Контингент гимназий, лицеев и школ с классами углубленного изучения предметов (лицеи), человек

29.	<i>net_kz_school</i>	Сеть дневных государственных общеобразовательных школ по языкам обучения (с казахским языком обучения), единиц
30.	<i>net_ru_school</i>	Сеть дневных государственных общеобразовательных школ по языкам обучения (с русским языком обучения), единиц
31.	<i>net_mix_school</i>	Сеть дневных государственных общеобразовательных школ по языкам обучения (со смешанными языками обучения), единиц
32.	<i>kz_student</i>	Контингент дневных государственных общеобразовательных школ по языкам обучения (с казахским языком обучения), человек
33.	<i>ru_student</i>	Контингент дневных государственных общеобразовательных школ по языкам обучения (с русским языком обучения), человек
34.	<i>mix_student</i>	Контингент дневных государственных общеобразовательных школ по языкам обучения (со смешанными языками обучения), человек
35.	<i>disable</i>	Всего детей с ограниченными возможностями в развитии от 7 до 18 лет по региону, человек
36.	<i>inclus_student</i>	Численность учащихся, охваченных инклюзивным образованием, в общеобразовательных организациях, человек
37.	<i>share_inclus</i>	Доля учащихся, охваченных инклюзивным образованием, %
38.	<i>occupancy_student</i>	Средняя наполняемость 10–11 (12) классов дневных государственных общеобразовательных школ (численность учащихся), человек
39.	<i>occupancy_room</i>	Средняя наполняемость 10–11 (12) классов дневных государственных общеобразовательных школ (число классов), человек
40.	<i>aver_occupancy_student</i>	Средняя наполняемость 10–11 (12) классов дневных государственных общеобразовательных школ (наполняемость), человек
41.	<i>exc_student</i>	Успеваемость учащихся дневных государственных общеобразовательных школ (отличники), человек
42.	<i>good_student</i>	Успеваемость учащихся дневных государственных общеобразовательных школ (хорошисты), человек
43.	<i>second_student</i>	Успеваемость учащихся дневных государственных общеобразовательных школ (ПОВТОРНЫЙ КУРС), человек
44.	<i>altyn_belgy</i>	Численность претендентов дневных государственных общеобразовательных школ, подведомственных МИО, подтвердивших знак «Алтын белгі», человек
45.	<i>bus</i>	Сведения об автотранспорте для организации подвоза учащихся дневных государственных общеобразовательных школ, подведомственных МИО (количество автобусов), единиц
46.	<i>bought_bus</i>	Сведения об автотранспорте для организации подвоза учащихся дневных государственных общеобразовательных школ, подведомственных МИО (количество приобретенного автотранспорта в отчетном году), единиц
47.	<i>need_bus</i>	Сведения об автотранспорте для организации подвоза учащихся дневных государственных общеобразовательных школ, подведомственных МИО (всего потребность в спецавтобусах), единиц
48.	<i>exc_teacher</i>	Педагогические работники дневных государственных общеобразовательных школ, подведомственных МИО (имеющие высшую категорию), человек
49.	<i>i_teacher</i>	Педагогические работники дневных государственных общеобразовательных школ, подведомственных МИО (имеющие первую категорию), человек

50.	<i>ii_teacher</i>	Педагогические работники дневных государственных общеобразовательных школ, подведомственных МИО (имеющие вторую категорию), человек
51.	<i>non_teacher</i>	Педагогические работники дневных государственных общеобразовательных школ, подведомственных МИО (не имеющие категорию), человек
52.	<i>math_teacher</i>	Качественный и количественный состав учителей математики в общем среднем образовании, человек
53.	<i>physics_teacher</i>	Качественный и количественный состав учителей физики в общем среднем образовании, человек
54.	<i>chem_teacher</i>	Качественный и количественный состав учителей химии в общем среднем образовании, человек
55.	<i>it_teacher</i>	Качественный и количественный состав учителей информатики в общем среднем образовании, человек
56.	<i>kaz_teacher</i>	Качественный и количественный состав учителей казахского языка в общем среднем образовании, человек
57.	<i>rus_teacher</i>	Качественный и количественный состав учителей русского языка в общем среднем образовании, человек
58.	<i>biolog_teacher</i>	Качественный и количественный состав учителей биологии в общем среднем образовании, человек
59.	<i>history_teacher</i>	Качественный и количественный состав учителей истории в общем среднем образовании, человек
60.	<i>geogr_teacher</i>	Качественный и количественный состав учителей географии в общем среднем образовании, человек
61.	<i>eng_teacher</i>	Качественный и количественный состав учителей английского языка в общем среднем образовании, человек
62.	<i>newroom_school</i>	Количество дневных государственных общеобразовательных школ, имеющих кабинеты новой модификации (химии, биологии, физики, лингафонных мультимедийных кабинетов) с сервисным обслуживанием, единиц
63.	<i>unem</i>	Уровень безработицы, в %
64.	<i>popul</i>	Численность населения региона, человек
65.	<i>realAvgWpc</i>	Среднемесячная реальная заработная плата одного работника, тенге
66.	<i>realAvgIncpc</i>	Реальные денежные доходы, в среднем на душу в месяц, тенге
67.	<i>inflation14</i>	Уровень инфляции, в % (2014=100%)
68.	<i>realGRPpc</i>	Реальный ВРП на душу населения, тыс. тенге
69.	<i>cars</i>	Обеспеченность населения легковыми автомобилями в личной собственности на 100 человек постоянного населения
70.	<i>gini</i>	Коэффициент Джини
71.	<i>real_medianInc</i>	Реальный медианный доход населения, тенге
72.	<i>poverty</i>	Доля населения уменяющая доход ниже прожиточного минимума
73.	<i>real_depositpc</i>	Депозиты на душу населения в постоянных ценах, тенге
74.	<i>real_creditpc</i>	Кредиты банков на душу населения в постоянных ценах, тенге
75.	<i>comp</i>	Доля обычных и опытных пользователей компьютеров в структуре пользователей в возрасте 16-74 лет
76.	<i>internet</i>	Доля пользователей интернета в возрасте 16-74 лет
77.	<i>marriage</i>	Количество заключенных браков на 1000 человек населения
78.	<i>divorce</i>	Количество разводов на 1000 человек населения
79.	<i>fract</i>	Индекс фракционализации региона

Приложение 3. Результаты тестов на наличие единичного корня

№	Переменная	Levin-Lin-Chu, p-value	Maddala-Wu, p-value	Im-Pesaran-Shin, p-value	стационарный/нестационарный
1.	<i>ENT_kz_student</i>	6.426e-14	3.412e-09	6.387e-08	Стационарный
2.	<i>ENT_kz_score</i>	3.004e-08	2.734e-13	2.312e-10	Стационарный
3.	<i>ENT_ru_student</i>	< 2.2e-16	6.311e-15	9.084e-13	Стационарный
4.	<i>ENT_ru_score</i>	2.38e-07	2.072e-11	1.531e-08	Стационарный
5.	<i>school</i>	5.188e-10	1.342e-06	8.168e-06	Стационарный
6.	<i>student</i>	1.125e-14	6.781e-10	1.661e-08	Стационарный
7.	<i>girl_student</i>	1.078e-14	7.96e-12	6.836e-10	Стационарный
8.	<i>teacher</i>	8.835e-13	2.998e-08	3.939e-07	Стационарный
9.	<i>girl_teacher</i>	9.045e-12	2.417e-11	3.163e-09	Стационарный
10.	<i>rooms</i>	5.417e-11	3.152e-06	1.648e-05	Стационарный
11.	<i>real_revenues</i>	0.1085	7.88e-08	9.173e-06	Нестационарный
12.	<i>real_budget</i>	6.207e-12	7.858e-08	9.307e-07	Стационарный
13.	<i>rep_school</i>	1	2.728e-08	5.752e-06	Нестационарный
14.	<i>MZ_school</i>	9.949e-07	3.394e-12	4.634e-09	Стационарный
15.	<i>MO_school</i>	2.453e-06	2.785e-06	0.0003528	Стационарный
16.	<i>sport_school</i>	2.957e-11	1.536e-08	2.619e-07	Стационарный
17.	<i>NISH</i>	4.169e-07	< 2.2e-16	6.855e-14	Стационарный
18.	<i>rep_student</i>	1	1.065e-05	0.0005267	Нестационарный
19.	<i>MZ_student</i>	-	-	< 2.2e-16	Нестационарный
20.	<i>MO_student</i>	0.02361	0.0002998	0.005701	Стационарный
21.	<i>sport_student</i>	5.297e-07	8.051e-06	5.111e-05	Стационарный
22.	<i>NISH_student</i>	0.001356	8.985e-14	1.824e-09	Стационарный
23.	<i>depth_school</i>	7.464e-09	6.043e-15	1.563e-08	Стационарный
24.	<i>gymnasium</i>	1.104e-15	1.637e-14	2.23e-12	Стационарный
25.	<i>lyceum</i>	8.838e-13	2.836e-11	2.308e-08	Стационарный
26.	<i>student_depth_school</i>	0.0001026	9.289e-10	3.934e-07	Стационарный
27.	<i>student_gymnasium</i>	1.217e-13	5.993e-12	3.362e-10	Стационарный
28.	<i>student_lyceum</i>	3.45e-08	2.542e-14	1.106e-09	Стационарный
29.	<i>net_kz_school</i>	7.509e-10	9.433e-08	1.992e-06	Стационарный
30.	<i>net_ru_school</i>	2.122e-12	4.171e-15	1.103e-12	Стационарный
31.	<i>net_mix_school</i>	2.045e-08	2.604e-06	1.408e-05	Стационарный
32.	<i>kz_student</i>	1.38e-08	1.347e-09	1.037e-07	Стационарный
33.	<i>ru_student</i>	< 2.2e-16	< 2.2e-16	< 2.2e-16	Стационарный
34.	<i>mix_student</i>	< 2.2e-16	5.213e-12	3.717e-10	Стационарный
35.	<i>disable</i>	1.727e-07	2.335e-06	2.232e-05	Стационарный
36.	<i>inclus_student</i>	8.623e-07	2.174e-07	3.586e-05	Стационарный
37.	<i>share_inclus</i>	< 2.2e-16	< 2.2e-16	< 2.2e-16	Стационарный
38.	<i>occupancy_student</i>	2.662e-13	2.477e-09	4.866e-08	Стационарный
39.	<i>occupancy_room</i>	1.193e-12	6.1e-09	1.045e-07	Стационарный
40.	<i>aver_occupancy_student</i>	0.1604	7.47e-10	3.351e-06	Нестационарный
41.	<i>exc_student</i>	1.176e-12	6.92e-09	1.332e-07	Стационарный
42.	<i>good_student</i>	2.506e-14	1.996e-10	7.302e-09	Стационарный
43.	<i>second_student</i>	9.09e-11	1.297e-10	1.235e-08	Стационарный
44.	<i>altyn_belgy</i>	5.383e-14	6.179e-11	2.614e-09	Стационарный
45.	<i>bus</i>	1.445e-14	1.956e-13	1.686e-11	Стационарный
46.	<i>bought_bus</i>	1.601e-10	3.766e-15	1.605e-11	Стационарный
47.	<i>need_bus</i>	1.2e-09	5.381e-10	4.126e-08	Стационарный
48.	<i>exc_teacher</i>	< 2.2e-16	8.291e-15	3.412e-12	Стационарный
49.	<i>i_teacher</i>	4.11e-08	5.533e-06	5.047e-05	Стационарный

№	Переменная	Levin-Lin-Chu, p-value	Maddala-Wu, p-value	Im-Pesaran-Shin, p-value	стационарный/нестационарный
50.	<i>ii_teacher</i>	9.386e-09	1.692e-06	1.779e-05	Стационарный
51.	<i>non_teacher</i>	3.396e-10	1.563e-06	1.02e-05	Стационарный
52.	<i>math_teacher</i>	1.185e-09	1.005e-07	1.938e-06	Стационарный
53.	<i>physics_teacher</i>	1.553e-09	1.43e-07	2.67e-06	Стационарный
54.	<i>chem_teacher</i>	3.359e-10	6.031e-08	1.248e-06	Стационарный
55.	<i>it_teacher</i>	1.478e-08	3.84e-07	5.04e-06	Стационарный
56.	<i>kaz_teacher</i>	1.608e-10	2.985e-08	7.413e-07	Стационарный
57.	<i>rus_teacher</i>	2.943e-10	1.402e-07	2.506e-06	Стационарный
58.	<i>biolog_teacher</i>	3.962e-09	3.766e-07	5.683e-06	Стационарный
59.	<i>history_teacher</i>	2.712e-08	4.025e-06	3.85e-05	Стационарный
60.	<i>geogr_teacher</i>	1.308e-09	4.23e-07	5.804e-06	Стационарный
61.	<i>eng_teacher</i>	2.197e-09	2.806e-08	6.838e-07	Стационарный
62.	<i>newroom_school</i>	4.205e-07	1.352e-07	2.292e-06	Стационарный
63.	<i>unem</i>	6.358e-10	3.31e-15	4.521e-12	Стационарный
64.	<i>popul</i>	7.688e-16	1.286e-11	7.636e-10	Стационарный
65.	<i>realAvgWpc</i>	1.087e-07	1.28e-05	5.071e-05	Стационарный
66.	<i>realAvgIncpc</i>	0.0008457	0.0227	0.02323	Стационарный
67.	<i>inflation14</i>	-	NA	NA	Нестационарный
68.	<i>realGRPpc</i>	0.006506	0.07737	0.05769	Стационарный
69.	<i>cars</i>	0.0008156	0.3438	0.237	Нестационарный
70.	<i>gini</i>	< 2.2e-16	< 2.2e-16	< 2.2e-16	Стационарный
71.	<i>real_medianInc</i>	0.04669	0.06629	0.3244	Стационарный
72.	<i>poverty</i>	1.158e-06	0.000144	0.0003886	Стационарный
73.	<i>real_depositpc</i>	0.3116	0.9714	0.869	Нестационарный
74.	<i>real_creditpc</i>	1	0.8786	1	Нестационарный
75.	<i>comp</i>	1.131e-10	1.396e-12	4.277e-09	Стационарный
76.	<i>internet</i>	6.995e-12	5.73e-11	1.548e-08	Стационарный
77.	<i>marriage</i>	3.446e-12	1.92e-09	4.012e-08	Стационарный
78.	<i>divorce</i>	1.359e-13	1.605e-11	7.406e-10	Стационарный
79.	<i>fract</i>	1.237e-11	< 2.2e-16	3.183e-12	Стационарный