
ПРАКТИКИ СБОРА И АНАЛИЗА ФОРМАЛИЗОВАННЫХ ДАННЫХ

Ю.В. Кузьмина
(Москва)

МЕТОД РАЗРЫВНОЙ РЕГРЕССИИ И МЕТОД ОТБОРА ПОДОБНОГО ПО ВЕРОЯТНОСТИ ДЛЯ ОЦЕНКИ ЭФФЕКТА ОДНОГО ГОДА ОБУЧЕНИЯ: ОПЫТ ПРИМЕНЕНИЯ НА ПРИМЕРЕ ДАННЫХ PISA 2009

В статье обсуждается понятие эффекта одного года обучения как возможного показателя эффективности национальных образовательных систем применительно к результатам международного исследования PISA 2009. Рассмотрены особенности оценки вклада одного года обучения с применением квазиэкспериментальных методов – нестрогого дизайна разрывной регрессии и метода отбора подобного по вероятности. Рассмотрены достоинства, ограничения и результаты применения каждого метода. Анализ проведен на данных PISA 2009 для десяти стран.

Ключевые слова: квазиэкспериментальные методы, эффект воздействия, разрывный дизайн регрессии, инструментальная переменная, отбор подобного по вероятности, эффект одного года обучения, чувствительный анализ, PISA

Введение

Одним из основных вопросов, который интересует исследователей в области социальных наук и образования, является оценка

Юлия Владимировна Кузьмина – НИУ Высшая школа экономики, Институт образования, Международная лаборатория анализа образовательной политики, научный сотрудник. E-mail: papushka7@gmail.com.

каузального эффекта каких-либо событий или воздействий. В области образовательных исследований, например, существует интерес к оценке эффективности образования или образовательных программ. Своеобразным «золотым стандартом» оценки эффекта воздействия служат рандомизированные эксперименты [1, 2]. Возможно, что наилучшим способом оценить эффект обучения может быть сравнение тех, кто посещает школу, с эквивалентной контрольной группой, то есть с теми, кто не посещает школу. Однако в реальных условиях такое сравнение трудно реализуемо. Существуют тем не менее многочисленные исследования, посвященные оценке эффекта образования или образовательных практик без обращения к экспериментальным данным, с использованием данных срезовых исследований [3, 4, 5]. Для этого применяют квазиэкспериментальные методы анализа, которые позволяют оценить эффект воздействия в нерандомизированных исследованиях, используя различные способы уменьшения ошибки отбора [2].

Здесь рассматривается понятие эффекта одного года обучения, как возможного показателя эффективности и возможности его оценки на срезовых данных с помощью квазиэкспериментальных методов: метода разрывной регрессии и метода отбора подобного по вероятности. Обсуждаются причины для выбора этих методов, их ограничения, а также способы проверки надежности полученных оценок.

Анализ проведен на данных международного исследования PISA 2009 (*Programme of International Student Assessment*).

Исследование PISA и эффект одного года

Международное исследование PISA проводится с целью оценки уровня развития функциональной грамотности школьников¹. Ре-

¹ Под функциональной грамотностью понимается умение школьников решать задачи повседневной жизни, используя теоретические знания, полученные ими в школе.

зультаты теста оцениваются по 1000-балльной шкале, где среднее международное равно 500 баллам, а стандартное отклонение – 100 баллам. В выборку включены 15-летние школьники, большая часть которых может учиться в 8–11 классах, в зависимости от особенностей национальных образовательных систем. Результаты PISA часто используются для оценки эффективности национальных образовательных систем, поскольку они позволяют сравнить достижения учащихся разных стран, измеряемые одним инструментом.

Однако данных о средних результатах PISA по странам не достаточно для того, чтобы делать выводы об эффективности школьных систем, поскольку различия в результатах между странами могут определяться факторами, не связанными прямо с качеством школьного образования (например, уровнем социально-экономического статуса, культурными особенностями и стереотипами). Более точным показателем эффективности образовательной системы может быть эффект (вклад) одного года обучения или разница в результатах между учащимися двух «последовательных» классов (например, между 9- и 10-классниками), поскольку этот показатель свидетельствует о «приросте» знаний за один год [6].

Сопоставление результатов 9- и 10-классников показывает, что во всех странах 10-классники имеют более высокие достижения по сравнению с 9-классниками [7]. Однако простая разница между результатами учащихся двух последовательных классов может включать в себя не только вклад школы в прирост результатов. Прежде всего, увеличение достижений 10-классников по сравнению с достижениями 9-классников может быть связано с эффектом взросления¹. Чтобы оценить вклад одного года обучения при контроле возможного эффекта взросления, следует

¹ В исследовании PISA принимают участие 15-летние школьники, однако максимальная разница в возрасте между учениками в выборке может достигать одного года (без одного дня). Во всех странах средний возраст 10-классников хоть незначительно, но больше среднего возраста 9-классников.

сравнивать результаты школьников одного возраста, либо учитывать эффект взросления, включая в регрессионную модель «возраст» в качестве независимой переменной. Но в этой модели не учитывается возможное влияние раннего школьного старта, который может быть связан, например, с характеристиками семьи ребенка и его дошкольной подготовкой. Впоследствии разница между достижениями 9- и 10-классников может быть связана с их подготовленностью, образованием, мотивацией родителей и проч., а не только с дополнительным годом обучения [8]. Таким образом, простое сравнение результатов 9- и 10-классников, или обычный регрессионный анализ, может привести к неверной оценке эффекта одного года обучения.

Чтобы учесть все названные выше факторы и уменьшить их влияние, для оценки эффекта одного года обучения чаще всего использовался метод разрывной регрессии [4; 5; 9].

Метод «разрывной регрессии» (Regression Discontinuity Design)

Метод разрывной регрессии (метод РР, или исследовательский план РР) известен уже более 50 лет, впервые он представлен в 1960 г. для оценки эффекта награждений на последующие академические достижения [10].

Этот метод используется для оценки эффекта какого-либо воздействия в случае, если существует и известен критерий отбора в группу, получающую воздействие. В дизайне разрывной регрессии отбор в экспериментальную группу определяется (полностью или частично) значениями критериальной переменной (*assignment*, или *forcing variable*) относительно известной пороговой точки (*cut-off point*). Например, участниками экспериментальной группы могут быть только те индивиды, значения критериальной переменной

которых выше, чем пороговые значения¹. Такой принцип отбора можно назвать «пороговое правило» (*cut-off rule*). От того, насколько строго соблюдается правило отбора участников, различают два вида исследовательского плана РР: строгий и нестрогий. При строгом типе дизайна вероятность получения воздействия в пороговой точке меняется с 0 до 1; некоторые исследователи отмечают, что допускается не больше 5% случаев нарушения порогового правила [4]. В этом случае эффект воздействия оценивается как разрыв между двумя регрессионными линиями (для экспериментальной и контрольной групп) в пороговой точке, зависимой переменной является результирующая переменная, а независимой – критериальная переменная.

При нестрогом типе РР анализа вероятность получения воздействия не изменяется с 0 до 1, таким образом, значения критериальной переменной лишь частично определяют получение воздействия. В этом случае эффект воздействия эквивалентен эффекту воздействия при анализе с помощью инструментальной переменной, в качестве которой выступает критериальная переменная [11; 12]. Для расчета эффекта воздействия при использовании инструментальной переменной применяется двухшаговый метод наименьших квадратов. На первом шаге строится регрессионная модель, где зависимой переменной выступает вероятность получения воздействия, а инструментальная переменная становится независимой переменной. На втором шаге рассчитанная вероятность получения воздействия включается как независимая переменная в следующее регрессионное уравнение. Полученный коэффициент регрессии переменной «предсказанная вероятность» во втором уравнении интерпретируется как эффект воздействия.

Однако даже в случае нестрогого дизайна разрывной регрессии критериальная (инструментальная) переменная должна отвечать необходимым условиям.

¹ Или меньше, в зависимости от правила.

1. Наличие сильной связи между инструментальной переменной и вероятностью получения воздействия. Чтобы оценить силу связи существует ряд тестов. Наиболее распространены показатели F -статистики, полученные на первой стадии 2-шаговой регрессии. Популярным считается практическое правило: если значения F -статистики меньше 10, то инструментальная переменная считается слабой, и оценки, полученные с помощью инструментального анализа, нельзя считать достоверными [13; 14]. В то же время некоторые авторы указывают, что даже если показатели F -статистики больше 10, этого недостаточно для вывода о релевантности инструментальной переменной [13].

Также для оценки «силы» выбранной инструментальной переменной и оценки связи между инструментальной переменной и переменной воздействия могут быть использованы статистика Андерсона-Рубина (*Anderson-Rubin Wald test*) и Клейбергена-Паапа (*Kleibergen-Paap LM statistic*). Релевантность выбранной инструментальной переменной может быть также подтверждена проверкой эндогенности переменной воздействия (в первом уравнении при двухшаговой регрессии). Для этого используется тестовая статистика эндогенности. Если полученные показатели этих тестов статистически незначимы, это может свидетельствовать о нерелевантности выбранной инструментальной переменной и ненадежности полученных оценок эффекта воздействия [14; 15].

2. Отсутствие прямой связи инструментальной переменной с результирующей. Связь инструментальной переменной с результатами может осуществляться только через воздействие. Иногда это условие имеет расширенное толкование: инструментальная переменная не должна иметь ни прямого эффекта на результирующую переменную, ни непрямого, через другие факторы, кроме как через воздействие. Однако в эмпирических исследованиях встречаются случаи, когда инструментальная переменная оказывает небольшой прямой эффект на результат [16]. Даже в случае нестрогого дизайна разрывной регрессии необходимо исключить возможность

влияния критериальной переменной на результирующую иначе как только через воздействие [17].

Нестрогий метод разрывной регрессии и анализ с помощью инструментальной переменной позволяют получить оценки локального среднего эффекта воздействия (*Local Average Treatment Effect*) (ЛСЭВ). Надо отметить, что ЛСЭВ рассчитывается не для всей совокупности, а только для тех индивидов, чье участие в экспериментальной группе определяется инструментальной переменной или соответствует «пороговому правилу»¹ [18]. В отдельных случаях оценки ЛСЭВ могут быть экстраполированы на всю популяцию. В частности, если разница в средних результатах между «соответствующими индивидами» и группой в целом невелика, то оценки ЛСЭВ могут сближаться с оценками для других подсовокупностей [19; 20].

В случае использования дизайна РР необходимо привлечь случаи, когда значение критериальной переменной близки к пороговым. Такое сужение позволяет сделать вывод: все основные «базовые» характеристики, способные изменить оценку воздействия, будут иметь одинаковое распределение в небольшом промежутке до и после пороговой точки [21].

При оценке эффекта одного года обучения методом разрывной регрессии используется тот факт, что во многих странах достижение ребенком школьного возраста к «пороговой дате» служит основным критерием начала школьного обучения. В случае соблюдения правила отбора ребенок, родившийся до пороговой даты, должен учиться на год дольше, чем тот, кто родился позже. Учеников, которые пошли в школу с соблюдением правила «пороговая дата», можно назвать «соответствующие индивиды». Таким образом, при оценке вклада одного года обучения по данным PISA критериальной переменной выступает возраст ребенка (относи-

¹ В англоязычной литературе такие индивиды называются compliers («соответствующие индивиды»).

тельно пороговой даты), зависимой переменной – тестовые баллы, эффект одного года обучения рассчитывается как разрыв между двумя регрессионными линиями, построенными для двух последовательных классов (например, для 9 и 10), в пороговой дате [4].

Но в выборке PISA 2009 практически нет стран, где «пороговое правило» выполнялось бы строго, и было бы возможным применение строгого метода разрывной регрессии. Поэтому необходим расчет эффекта одного года обучения с использованием возраста как инструментальной переменной в рамках нестрогого дизайна разрывной регрессии. Кроме того, необходимо проверить, является ли относительный возраст ребенка «сильной» инструментальной переменной. Если в какой-либо стране связь между возрастом ребенка (относительно пороговой даты) и началом школьного обучения недостаточно сильна, эта переменная не может быть признана критерием получения воздействия, а значит – метод разрывной регрессии (даже нестрогого дизайна) не может быть применен. В таком случае для оценки эффекта одного года обучения может быть использован другой квазиэкспериментальный метод – отбора подобного по вероятности.

Метод отбора подобного по вероятности (Propensity Score Matching)

Метод отбора подобного по вероятности для оценки эффекта воздействия был описан Д. Рубиным и П. Розенбаумом в 1983 г. [22] Этот метод позволяет оценить средний эффект воздействия для экспериментальной группы (СЭВЭ), который равен разнице в потенциальных результатах индивидов в случае получения воздействия и в случае его отсутствия. В качестве потенциального результата для индивидов, не получивших воздействие, используется соответствующий средний результат для контрольной группы.

Базовая идея этого метода такова: найти для каждого участника, получившего воздействие, идентичного участника, не по-

лучившего воздействия, а затем – среднюю разницу между ними в показателях результирующей переменной. Чтобы определить идентичных (схожих по основным характеристикам) участников из двух групп, для каждого индивида рассчитывается вероятность получения воздействия (*propensity score*) с учетом всех возможных факторов, которые могут повлиять на результирующую переменную и вероятность получения воздействия. После того как для каждого наблюдения рассчитана вероятность получения воздействия, используются разные алгоритмы для сравнения результатов индивидов, получивших воздействие, с индивидами, не получившими воздействие, с одинаковыми (или близкими) показателями рассчитанной вероятности [23]. Таким образом, участники двух групп «выравниваются» по важным характеристикам, что позволяет учесть возможную ошибку отбора при оценке различий между контрольной и экспериментальной группами.

На этапе расчета вероятности большое внимание должно быть уделено отбору переменных, включаемых в модель. Игнорирование важной переменной приводит к неверным оценкам, а включение нерелевантных переменных может привести к увеличению дисперсии полученных оценок [24]. Некоторые авторы считают, что в модель должны быть включены ковариаты, связанные как с результирующей переменной, так и с воздействием [25]. Другие авторы считают, что основным критерием выбора переменных для включения в модель должна быть их сильная связь с результатами, независимо от того, связаны они с воздействием или нет [26]. Более того, включение в модель переменных, которые связаны с воздействием, но не связаны с результатами, уменьшает достоверность полученных данных. Например, в модель не может быть включена инструментальная переменная, поскольку она связана только с воздействием [26; 27].

Не существует также единства в ответе на вопрос, какое количество переменных должно быть включено в модель при расчете вероятности. Некоторые авторы полагают, что стоит избегать чрезмерно сложных моделей с большим количеством переменных,

достаточно включить несколько переменных, имеющих сильную связь с результирующей переменной и с воздействием [26]. Другие считают лучшей стратегией включение большого количества переменных, даже если анализ показывает, что их связь с результирующей переменной недостаточно сильна [25, 28].

Независимо от того, какая стратегия при выборе переменных используется, должны выполняться несколько условий.

1. Условная независимость (*Conditional Independence Assumption*). Это требование означает, что существуют значимые наблюдаемые факторы, связанные с результатами, которые могут быть статистически проконтролированы. Если все значимые факторы взяты под контроль, потенциальный результат становится независимым от полученного воздействия [23; 28].

2. Наличие «пересечения», или общего основания (*Common Support Assumption and Overlap*). Это условие означает, что наблюдаемые факторы, включенные в модель, полностью не предсказывают получение воздействия, а следовательно, почти для всех значений ковариат существуют наблюдения в контрольной и экспериментальной группах. Таким образом, распределение всех переменных в двух группах должно «перекрываться». Чтобы проверить соблюдение этого условия, после расчета вероятности используется процедура проверки сбалансированности (*balancing property*). Один из способов проверки включает в себя разделение всей выборки на несколько групп по показателям рассчитанной вероятности получения воздействия. Затем в каждом интервале сравниваются показатели ковариат в контрольной и экспериментальной группах. Условие считается выполненным, если в каждом интервале между участниками контрольной и экспериментальной групп нет статистически значимой разницы по всем включенным ковариатам. Если условие сбалансированности не выполняется со всеми включенными в модель переменными, необходимо заново рассчитать вероятность получения воздействия с новым набором переменных [29].

После того как рассчитана вероятность и проверено соблюдение условия сбалансированности, необходимо выбрать алгоритм расчёта разницы в результатах между контрольной и экспериментальной группами. Существует несколько алгоритмов, остановимся кратко на описании двух:

– метод ближайшего соседа (*nearest neighbor matching*). Это наиболее «прямой» метод оценки. Для каждого наблюдения в экспериментальной группе подбирается наблюдение из контрольной группы, наиболее близкое ему по значению рассчитанной вероятности. Для каждой пары вычисляется разница в результатах. Средняя разница по всем найденным парам интерпретируется как средний эффект воздействия для экспериментальной группы. Чаще всего наблюдение из контрольной группы используется один раз, поэтому, если число участников контрольной группы значительно больше числа участников экспериментальной группы, часть данных о контрольной группе игнорируется.

– «радиальное сравнение» (*Radius matching*). Здесь для расчёта разницы между контрольной и экспериментальной группой используется не один (ближайший по значениям рассчитанной вероятности), а нескольких участников контрольной группы, рассчитанные вероятности которых находятся в пределах установленного допустимого максимального расстояния (*caliper*). Для этой группы рассчитывается средний показатель результирующей переменной, который и сравнивается с показателем соответствующего участника, получившего воздействие. Таким образом, учитывается большее количество оценок участников контрольной группы [23].

Основные различия между способами расчета эффекта воздействия заключаются в алгоритме отбора сравниваемых участников из контрольной группы. И хотя достаточно часто оценки СЭВЭ, рассчитанные с помощью разных методов, существенно не отличаются, рекомендуется использовать несколько алгоритмов и сравнить полученные результаты [30].

Для проверки того, насколько результаты метода отбора подобного по вероятности чувствительны к наличию неучтенных важных факторов, некоторые исследователи рекомендуют в качестве последнего шага проводить анализ чувствительности, то есть устойчивости эффекта воздействия к присутствию неучтенных смешивающих переменных [23, 30]. Один из способов чувствительного анализа – симуляция переменной (анализ, основанный на имитационном моделировании) [31]. Допускается, что существует важный фактор (с бинарной структурой), который имеет сильную связь с воздействием и с результирующей переменной. Один из способов смоделировать неучтенный фактор – предположить, что он имеет такое же распределение, как какая-либо наблюдаемая переменная (имитирует ее). Задача исследователя – выбрать, какую переменную из имеющихся может имитировать воображаемый фактор. После этого «симулированный» фактор включается в модель для расчета вероятности наряду с другими переменными и рассчитываются новые показатели СЭВЭ. Сравнивая рассчитанный эффект с «базовым», необходимо оценить, остается ли эффект воздействия значимым и насколько он уменьшается по сравнению с базовой оценкой [31]. Таким образом, на основе чувствительного анализа можно косвенно оценить, насколько учтены все значимые факторы при расчете вероятности получить воздействие и выполняется требование условной независимости.

Применение метода отбора по вероятности для оценки эффекта одного года позволяет «выровнять» 9- и 10-классников по важным характеристикам, уменьшив тем самым ошибку отбора. Это позволит оценить разницу в результатах между 9- и 10-классниками с учетом возможных искажающих факторов.

Переменные

Для расчета эффекта одного года обучения и сравнения оценок, полученных двумя методами, использованы результаты PISA

2009 по математике для 10 стран: Австрии, Албании, Бразилии, Венгрии, Германии, Ирландии, Люксембурга, России, Хорватии и Чехии. Эти страны выбраны потому, что в них большая часть учеников, включенных в выборку, учатся в 9-м и 10-м классах (см. табл. 1). Кроме того, эти страны различаются по ситуации с выполнением правила «пороговая дата», что позволяет оценить разные условия применения дизайна разрывной регрессии.

Для оценки эффекта одного года обучения в выборке оставлены только учащиеся 9 и 10 классов, ни разу не остававшиеся на второй год.

Результирующая переменная – баллы PISA 2009 по математике.

Переменная «воздействия» – один год обучения (дихотомическая переменная, принимает значение «1», если школьник учится в 10 классе, «0» – в 9-м классе).

Критериальная переменная – возраст ученика (относительно пороговой даты). В отобранных странах все ученики родились в один год (1993), а данные о дате рождения представлены только месяцем. Для включения в анализ образована дихотомическая переменная: «1» – родился до пороговой даты, «0» – после пороговой даты. Использование дихотомической переменной представляется приемлемым, поскольку для принятия решения о приеме в школу критерием служит рождение до или после пороговой даты, независимо от того, насколько раньше или позже этой даты ученик родился.

Другие переменные: пол, семейные факторы (социально-экономический статус – СЭС, индексы, например, профессионального статуса родителей, образовательных семейных ресурсов), расположение школы, вид образовательной программы (общий или профессиональный).

Таблица 1
 ХАРАКТЕРИСТИКИ ВЫБОРКИ, ПОРОГОВЫЕ ДАТЫ И КОЛИЧЕСТВО УЧЕНИКОВ,
 ПОШЕДШИХ В ШКОЛУ С НАРУШЕНИЕМ «ПОРОГОВОГО ПРАВИЛА»

Страны	Случаи, включенные в анализ			Пороговая дата	Ученики, пошедшие в школу с нарушением «порогового правила», в %		
	Всего	Класс			Всего	Класс	
		9, в %	10, в %				
Австрия	5752	45	55	1 сентября	13	26	2
Албания	3781	43	57	1 октября	26	54	5
Бразилия	9159	36	64	1 сентября	30	38	24
Венгрия	4205	75	25	1 июня	20	25	6
Германия	3181	57	43	1 июля	12	11	12
Ирландия	2774	69	31	1 июня	17	13	27
Люксембург	2935	41	59	1 сентября	8	13	5
Россия	4530	65	35	1 октября	52	74	11
Хорватия	4823	78	22	1 апреля	9	8	14
Чехия	5535	52	48	1 сентября	20	38	1

Последовательность анализа

На первом этапе использован метод разрывной регрессии, поскольку ранее он применялся для оценки эффекта одного года на данных PISA.

Пороговая дата выбирается для каждой страны с учетом нескольких критериев: 1) согласно законодательству этой страны, в школу принимаются дети, достигшие школьного возраста к этой дате; 2) для учеников, рожденных после этой даты, резко снижается вероятность обучения в 10 классе.

Для того чтобы оценить величину «скачка» вероятности в пороговой точке, построены графики изменения вероятности учиться в 10 классе в зависимости от возраста (возраст в месяцах, центрированный относительно пороговой даты).

Для проверки наличия связи возраста с результатами, проведен регрессионный анализ, в первой модели переменная «возраст» включена как единственная независимая; во вторую модель добавлены другие переменные: продолжительность обучения, семейные характеристики, пол, местоположение школы.

Расчет эффекта с помощью инструментальной переменной проведен также с использованием двух моделей: с ковариатами и без. В модели с ковариатами другие переменные включены в качестве независимых переменных во втором уравнении двухшаговой регрессии. Таким образом, модель с ковариатами показывает эффект одного года обучения при контроле некоторых факторов.

Для проверки релевантности переменной «возраст» в качестве инструментальной при оценке эффекта одного года обучения для каждой страны проверена гипотеза о «силе» инструментальной переменной и эндогенности переменной воздействия.

Так как метод разрывной регрессии предполагает отбор участников, родившихся в небольшом интервале от пороговой даты, отобраны ученики, рожденные в промежутке плюс-минус 3 месяца от пороговой даты. Также проведен анализ на полной выборке.

Сравнение результатов показало, что нет значимых различий в оцениваемом эффекте на полной и ограниченной выборке, однако для ограниченной выборки величина стандартной ошибки больше, поэтому в дальнейшем используются данные для полной выборки.

На втором этапе во всех странах эффект одного года обучения оценен с помощью отбора подобного по вероятности. Для всех учеников, включенных в анализ, рассчитана вероятность их обучения в 10 классе (получить воздействие). Для расчета вероятности обучения в 10-м классе в модель отобраны переменные, показавшие в предыдущих исследованиях значимую связь с результатами PISA и имеющие значимую связь с результатами по математике по итогам регрессионного анализа, независимо от того, как они связаны с вероятностью обучения в 10 классе. Возраст не включен ни в одну модель, поскольку он может быть рассмотрен как инструментальная переменная для получения воздействия. Как уже говорилось выше, включение такой переменной в модель для расчета вероятности уменьшает достоверность полученных данных [27]. Расчет выполнялся с помощью модуля `rscope` в программе STATA 11.1. После расчета вероятности проверено условие сбалансированности и рассчитан СЭВЭ с использованием двух алгоритмов: метода ближайшего соседства и радиального сравнения. Условие сбалансированности проверено для каждой страны сравнением средних показателей каждой переменной, включенной в модель, в каждом выделенном блоке с использованием t -критерия Стьюдента.

Затем проведен чувствительный анализ, причем в качестве переменной, которую может имитировать воображаемый фактор, выбрана переменная «социально-экономический статус (СЭС) выше среднего». Предыдущие исследования факторов, связанных с результатами PISA, показали, что СЭС – это один из наиболее сильных предикторов [7]. Поэтому мы допускаем, что есть такой неучтенный фактор, также сильно связанный с результатами, как СЭС. Поскольку имитационный чувствительный анализ работает

только с дихотомическими переменными, для каждой страны была создана фиктивная переменная «СЭС выше среднего», которая принимает значение «1», если индивидуальный показатель СЭС больше среднего по стране, и значение «0» в другом случае. Для сенситивного имитационного анализа использован метод ближайшего соседа.

Далее сравниваются результаты, полученные двумя методами.

Результаты

Пороговые даты и количество «несоответствующих участников» среди 9-ти и 10-классников отражены в *табл. 1*. Можно отметить, что в большинстве стран правило отбора соблюдается строже для 10-классников, то есть среди учащихся 10-классов доля тех, кто пошел в школу, не соблюдая правила, меньше, чем в 9 классе.

По итогам регрессионного анализа во всех странах (кроме Албании) зафиксирован положительный и значимый эффект одного года обучения, который находится в промежутке от 17 до 64 баллов PISA (*табл. 2*).

Результаты оценки эффекта одного года обучения, полученные с помощью метода инструментальной переменной, отличаются от оценок, полученных с помощью регрессионного анализа. Вклад дополнительного года обучения, полученный в результате инструментального анализа, оценивается от 12 до 30 баллов PISA, то есть ниже, чем по итогам регрессионного анализа. В России и Албании эффект одного года при анализе с помощью инструментальной переменной оценен как незначимый. Для двух стран (Ирландии и Бразилии) оценки эффекта, полученные с помощью инструментальной переменной, практически совпадают с результатами регрессионного анализа.

Также можно отметить, что не только снизилось оценки вклада дополнительного года обучения, но и стерлись различия между странами. По итогам инструментального анализа, вклад одного

Таблица 2
ЭФФЕКТ ВОЗРАСТА И ОДНОГО ГОДА ОБУЧЕНИЯ ПО ИТОГАМ РЕГРЕССИОННОГО И
ИНСТРУМЕНТАЛЬНОГО АНАЛИЗА (НЕСТАНДАРТИЗИРОВАННЫЕ КОЭФФИЦИЕНТЫ
РЕГРЕССИИ, В БАЛЛАХ PISA)

Страна	Эффект					
	возраста			одного года обучения		
	Регрессионный анализ			Инструментальный анализ		
	Модель					
	без ковариат	с ковариатами	с ковариатами	без ковариат	с ковариатами	с ковариатами
Австрия	20,1*** (2,4)	-13,9*** (3,3)	44,3*** (4,6)	24,5*** (3,9)	26,6*** (3,3)	
Албания	0,86 (3,1)	-5,1* (2,9)	9,9 (6,2)	6,37 (6,7)	4,3 (6,4)	
Бразилия	8,9*** (1,6)	-1,8 (1,4)	29,4*** (2,3)	28,1*** (5,6)	30,04*** (4,9)	
Венгрия	6,7*** (2,5)	-8,8*** (2,7)	27,7*** (3,1)	12,9** (4,6)	12,1*** (3,7)	
Германия	29,3*** (2,98)	-24,8*** (4,08)	64,1*** (4,5)	38,5*** (3,82)	31,6*** (3,4)	
Ирландия	12,3*** (3,2)	-0,5 (3,8)	17,6*** (4,2)	23,3*** (5,3)	17,5*** (4,8)	
Люксембург	24,7*** (2,8)	-22,6*** (4,4)	54,2*** (4,4)	29,2*** (3,2)	27,3*** (2,9)	
Россия	3,2 (2,9)	-1,6 (2,8)	19,9*** (5,1)	12,6 (14,1)	9,01 (13,4)	
Хорватия	19,5*** (2,7)	-11,9*** (3,1)	44,4*** (3,3)	25,1*** (3,8)	26,7*** (3,02)	
Чехия	14,7*** (2,6)	-21,1*** (2,8)	48,04*** (6,2)	27,1*** (3,7)	21,8*** (3,5)	

Примечания: * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

В скобках указана стандартная ошибка.

года изменяется от страны к стране обучения в меньшей степени, чем можно было бы судить по итогам регрессионного анализа.

Обобщая полученные на первом этапе результаты, можно условно разделить все страны на три группы.

1. Количество «несоответствующих индивидов» не превышает 20%, при этом в контрольной группе их число не превышает 40, а в экспериментальной группе – не больше 14%. Относительный возраст имеет положительную связь с результатами в модели без ковариат и отрицательно связан с результатами при учете продолжительности обучения. Эффект одного года обучения оценен как значимый и положительный, при этом оценки, полученные с помощью инструментальной переменной, ниже, чем в простой регрессионной модели. Это относится к таким странам, как Чехия, Венгрия, Германия, Австрия, Хорватия, Люксембург.

2. Правило отбора соблюдается слабо, число «несоответствующих индивидов» в контрольной группе превышает 50%. В регрессионном анализе возраст не имеет значимой связи с результатами в обеих моделях. Оценки эффекта, полученные с помощью инструментальной переменной, незначимы. К этой группе относятся Россия и Албания.

3. В выборке число «несоответствующих индивидов» не превышает 30%, но в экспериментальной группе их больше 20%. В регрессионном анализе коэффициент переменной «возраст» значим в модели без ковариат и становится незначимым при включении в модель переменной «продолжительность обучения». Оценки эффекта одного года обучения, полученные в ходе инструментального анализа, значимы и не отличаются от оценок, полученных в ходе регрессионного анализа. К этой группе относятся Бразилия и Ирландия.

Оценки релевантности инструментальной переменной (табл. 3) показывают, что возраст выступает сильной инструментальной переменной для оценки эффекта одного года обучения только для первой группы стран. Для России и Албании оценки теста

Андерсона-Рубина свидетельствуют об отсутствии сильной связи между инструментальной переменной и переменной воздействия. Для всех стран, относящихся ко второй и третьей группам, оценки теста на эндогенность незначимы, таким образом, переменную «продолжительность обучения» нельзя считать эндогенной. Разумеется, для Бразилии и Ирландии полученные в ходе инструментального анализа оценки эффекта одного года обучения значимы, однако существует вероятность, что они неточны.

Таблица 3

ОЦЕНКИ РЕЛЕВАНТНОСТИ ИНСТРУМЕНТАЛЬНОЙ ПЕРЕМЕННОЙ

Страна	F-test	Тест Андерсона-Рубина	Тестовая статистика эндогенности
Австрия	6053	38,01 (0,00)	21,7 (0,00)
Албания	1358	0,91 (0,34)	0,28 (0,6)
Бразилия	794	23,8 (0,00)	2,1 (0,15)
Венгрия	1855	7,66 (0,005)	13,9 (0,00)
Германия	3986	93,07 (0,00)	49,99 (0,00)
Ирландия	1196	19,02 (0,00)	0,02 (0,89)
Люксембург	6899	77,8 (0,00)	56,7 (0,00)
Россия	198	0,79 (0,37)	1,48 (0,22)
Хорватия	3072	41,77 (0,00)	25,01 (0,00)
Чехия	6211	53,3 (0,00)	25,6 (0,00)

Примечание: В скобках указаны значения *p-value*.

Отсутствие значимых оценок эффекта, полученных в ходе инструментального анализа, может означать как то, что эффект действительно отсутствует, так и то, что выбранная инструментальная переменная «возраст относительно пороговой даты» нерелевантна. И наоборот: значимость полученного эффекта не гарантирует релевантности выбранной инструментальной переменной.

Применение метода отбора по вероятности позволяет получить альтернативные оценки эффекта воздействия и уточнить полученные на первом этапе результаты.

В табл. 4 указаны переменные, включенные в модель при расчете вероятности обучения школьника в 10 классе, для которых соблюдено условие сбалансированности.

Таблица 4

ПЕРЕМЕННЫЕ, ВКЛЮЧЕННЫЕ В МОДЕЛЬ ДЛЯ РАСЧЕТА
 ВЕРОЯТНОСТИ ОБУЧЕНИЯ В 10-М КЛАССЕ (МЕТОД ОТБОРА
 ПОДОБНОГО ПО ВЕРОЯТНОСТИ), ДЛЯ КОТОРЫХ СОБЛЮДЕНО
 УСЛОВИЕ СБАЛАНСИРОВАННОСТИ

	Австрия	Албания	Бразилия	Венгрия	Германия	Ирландия	Люксембург	Россия	Хорватия	Чехия
СЭС	+	+	+	+	+	+	+	+	+	+
Пол	+	+	+	+	+	+	+	+	+	+
Местоположение школы	+		+			+	+	+	+	+
Профессиональный статус родителей		+	+	+	+	+		+	+	+
Образование родителей		+		+		+	+	+	+	+
Образовательные ресурсы	+				+			+	+	+
Индекс домашнего имущества		+	+	+	+	+	+	+		+
Культурные ресурсы	+	+	+	+	+			+	+	+
Иммигрант. статус		+	+		+	+		+	+	+
Индекс ИКТ		+		+	+		+	+		+
Материал.ресурсы	+	+		+	+	+	+	+	+	+
Вид программы	+			+					+	
Число полученных блоков	7	8	10	9	9	6	10	8	5	8

В табл. 5 отражены оценки эффекта одного года обучения, полученные посредством двух алгоритмов сравнения: ближайшего

соседства и радиального сравнения¹. Также в табл. 5 отображено число наблюдений в каждой группе, которые используются для расчета СЭВЭ.

Таблица 5

ЭФФЕКТ ОДНОГО ГОДА ОБУЧЕНИЯ, РАССЧИТАННЫЙ МЕТОДОМ
ОТБОРА ПОДОБНОГО ПО ВЕРОЯТНОСТИ (В БАЛЛАХ PISA)

Страна	Алгоритм			
	ближайшего соседа		радиального сравнения	
	Эффект одного года	Число наблю- дений в контрольной группе, ис- пользуемых для сравнения	Эффект одного года	Число наблю- дений в контрольной группе, ис- пользуемых для сравнения
Австрия	36,9*** (2,9)	1567	36,4*** (2,2)	2571
Албания	5,7 (3,6)	969	12,6*** (2,8)	1623
Бразилия	28,3*** (2,1)	2416	33,1*** (1,6)	3289
Венгрия	21,8*** (3,9)	880	21,5*** (2,8)	3127
Германия	45,7*** (3,9)	824	49,6*** (2,9)	1787
Ирландия	13,7*** (4,1)	668	16,9*** (3,1)	1895
Люксембург	34,9*** (3,7)	775	40,8*** (2,8)	1204
Россия	21,5*** (3,4)	1069	25,9*** (2,5)	2930
Хорватия	34,3*** (3,9)	881	38,6*** (2,7)	3752
Чехия	37,9*** (3,2)	1468	34,8*** (2,4)	2867

Примечания: * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$

В скобках указана стандартная ошибка.

Как видно из результатов, оценки СЭВЭ, полученные двумя алгоритмами сравнения, значимо не отличаются друг от друга. В дальнейшем при интерпретации результатов мы будем пользоваться оценками, полученными с помощью алгоритма ближайшего соседа.

¹ Получены также оценки с использованием других алгоритмов (*kernel matching u stratification matching*). Они значимо не отличаются от оценок, приведенных в табл. 5.

При оценке методом отбора по вероятности во всех странах, кроме Албании, зафиксирован значимый эффект одного года обучения. В целом, оценки вклада одного года обучения в результаты по математике находятся в промежутке от 13 до 46 баллов PISA.

Результаты чувствительного имитационного анализа (табл. 6) показывают, что при включении в модель для расчета вероятности другого значимого неучтенного фактора эффект одного года остается положительным и значимым во всех странах, где зафиксирован значимый эффект одного года. Во всех странах, кроме России и Германии, разница между базовой и «имитационной» оценкой СЭВЭ не превышает размера стандартной ошибки, то есть статистически незначима. Можно сказать, что полученные с помощью метода отбора подобного по вероятности оценки одного года обучения устойчивы к наличию возможных неучтенных факторов.

Таблица 6

РЕЗУЛЬТАТЫ ИМИТАЦИОННОГО ЧУВСТВИТЕЛЬНОГО АНАЛИЗА

Страна	Оценки эффекта (метод ближайшего соседа)		Разница между двумя оценками (в баллах PISA)
	базовые	имитационные	
Австрия	36,9*** (2,9)	36,3*** (3,5)	0,6
Албания	5,7 (3,6)	7,6 (4,1)	-2,1
Бразилия	28,3*** (2,1)	27,4*** (2,4)	0,9
Венгрия	21,8*** (3,9)	18,01*** (4,4)	3,8
Германия	45,7*** (3,9)	40,4*** (4,6)	5,3
Ирландия	13,7*** (4,1)	14,1*** (5,01)	-0,4
Люксембург	34,9*** (3,7)	36,2*** (4,4)	-1,3
Россия	21,5*** (3,4)	17,6*** (3,9)	3,9
Хорватия	34,3*** (3,9)	33,8*** (4,7)	0,5
Чехия	37,9*** (3,2)	35,04*** (3,7)	2,86

Примечания: *** $p < 0,01$

В скобках указана стандартная ошибка.

Таким образом, метод отбора подобного по вероятности позволяет получить оценки среднего эффекта одного года для экспериментальной группы во всех выбранных странах.

Обсуждение результатов

В целом, оценки эффекта одного года обучения, полученные с помощью инструментальной переменной в рамках нестроого дизайна разрывной регрессии, ниже, чем те, что получены с помощью метода отбора подобного по вероятности, для стран, где возраст выступает релевантной инструментальной переменной. Эти различия могут быть связаны с тем, что эти методы оценивают эффект воздействия для разных подгрупп. Метод инструментальной переменной оценивает ЛСЭВ – эффект воздействия для участников, соблюдающих правило отбора. Метод отбора подобного по вероятности оценивает СЭВЭ – средний эффект воздействия для экспериментальной группы. Оценки ЛСЭВ могут быть близкими к СЭВЭ в том случае, если разница в средних результатах между «соответствующими участниками» и экспериментальной группой в целом невелика. Однако в странах, которые мы ранее отнесли к первой группе (Чехии, Венгрии, Германии, Австрии, Хорватии), 10-классники, рожденные до пороговой даты, имеют более низкие в среднем баллы по математике, чем те, кто родился после пороговой даты. Таким образом, существуют различия в результатах между «соответствующими участниками» и экспериментальной группой в целом, что означает невозможность экстраполяции ЛСЭВ на всю выборку или экспериментальную группу.

Иную картину можно наблюдать для тех стран, где при контроле продолжительности обучения результаты учеников, рожденных до и после пороговой даты, не отличаются (Бразилия, Ирландия). В этих странах оценки ЛСЭВ могут быть экстраполированы на другие субпопуляции в связи с однородностью генеральной совокупности. Надо сказать, что невозможность распространения оценок ЛСЭВ на всю совокупность – одно из «слабых мест» метода анализа с

помощью инструментальной переменной. Ряд авторов ставят под сомнение полезность оценок ЛСЭВ для интерпретации эффекта каких-то программ [19].

В целом можно говорить, что применение метода разрывной регрессии, даже нестроого дизайна, для оценки эффекта одного года обучения, возможно только в тех странах, где количество учеников, пошедших в школу с нарушением правила, не превышает 20%. При этом важно распределение «несоответствующих индивидов» в контрольной и экспериментальной группах. Наш анализ показал, что возраст является сильной инструментальной переменной в тех странах, где число «несоответствующих участников» в контрольной группе не превышает 40%, а в экспериментальной – не более 14%. Это существенно ограничивает число стран, где возможно применение метода разрывной регрессии для оценки вклада школы в результаты PISA.

Для метода отбора подобного по вероятности ограничением служит наличие ненаблюдаемых факторов, которые могут повлиять на результаты (например, уровень интеллекта). Поэтому нельзя исключать вероятность переоценки эффекта одного года обучения по итогам метода отбора подобного по вероятности. Однако сенситивный анализ показывает, что полученные оценки в целом устойчивы к наличию сильно действующих факторов (таких как социально-экономический статус). Таким образом, метод отбора подобного по вероятности представляется приемлемым способом для оценки эффекта одного года обучения для данных PISA, так как устойчив к наличию неучтенных сильных факторов и позволяет сравнивать результаты в странах, независимо от соблюдения правила пороговой точки.

Выводы

1. Оценить эффективность одного года обучения для результатов PISA2009 можно двумя методами: методом разрывной ре-

грессии (нестрогого дизайна, рассматривая возраст относительно пороговой даты как инструментальную переменную) и методом отбора подобного по вероятности.

2. Метод разрывной регрессии при оценке эффекта одного года обучения применим в том случае, если для данных выполняется правило пороговой даты при приеме в школу и возраст служит «сильным» критерием для начала школьного обучения. В таком случае возможно получение оценок эффекта с учетом смешивающих факторов.

3. В случае, если возраст относительно пороговой даты не может служить сильным критерием для начала школьного обучения, метод отбора подобного по вероятности выступает надежной альтернативой для оценки эффекта одного года обучения на данных PISA. Кроме того, полученные оценки относятся ко всей экспериментальной группе, а не только к тем участникам, кто пошел в школу с соблюдением правила, как в случае с использованием инструментальной переменной. Оценки, полученные методом отбора по вероятности, устойчивы к наличию некоторых неучтенных факторов, которые могут быть связаны с баллами PISA по математике.

4. Таким образом, в ситуации, когда надо сравнить оценки эффекта одного года обучения в большом количестве стран, если в части из них не соблюдается правило пороговой точки, метод отбора подобного по вероятности представляется приемлемым и доступным для оценки эффекта одного года обучения для данных PISA, однако нам не удалось найти работы, где этот метод применялся для этих целей.

ЛИТЕРАТУРА

1. *Кэмпбелл Д.* Модели экспериментов в социальной психологии и прикладных исследованиях. СПб.: Социально-психол. центр, 1996.
2. *Schneider B., Carnoy M., Kilpatrick J., Schmidt W.H., Shavelson R.J.* Estimating Causal Effects Using Experimental and Observational Designs (Report from the

Governing Board of the American Educational Research Association Grants Program). Washington, DC: American Educational Research Association, 2007.

3. *Seidel T., Shavelson R.J.* Teaching Effectiveness Research in the Past Decade: the Role of Theory and Research Design in Disentangling Meta-analysis Results // Review of Educational Research. 2007. Vol. 77. No. 4. P. 454–499.

4. *Luyten H., Peschar J., Coe R.* Effects of Schooling on Reading Performance, Reading Engagement, and Reading Activities of 15-Year-Olds in England // American Education Research J. 2008. Vol. 45(2). P. 319–342.

5. *Luyten H.* An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95 // Oxford Review of Education. 2006. Vol. 32(3). P. 397–429.

6. *Тюменева Ю.А., Кузьмина Ю.В.* Что дает год обучения российскому школьнику. На материалах PISA-2009: грамотность чтения // Вопросы образования. 2013. № 1. P. 107–136.

7. *OECD.* PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I). Paris, 2010.

8. *Ayap A.* The Tough Choice at High School Door: An Investigation of the Factors That Lead Students to General or Vocational Schools // International Journal of Educational Development. 2003. Vol. 23 (5). P. 517–527.

9. *Suggate S.P.* School entry age and reading achievement in the 2006 Programme for International Student Assessment (PISA) // International Journal of Educational Research. 2009. Vol. 48 (3). P. 151–161.

10. *Thistlethwait D.L., Campbell D.T.* Regression-discontinuity Analysis: An Alternative to the Ex-post Facto Experiment. Beverly Hills: Sage Publications, 1960.

11. *Imbens G.W., Lemieux T.* Regression Discontinuity Designs: A Guide to Practice // Journal of Econometrics. 2008. Vol. 142 (2). P. 615–635.

12. *Angrist J.D., Imbens G.W., Rubin D. B.* Identification of Causal Effects Using Instrumental Variables // Journal of the American Statistical Association. 1996. Vol. 91(434). P. 444–455.

13. *Martens E.P., Pestman W.R., Boer A., Belitser S.V., Klungel O.H.* Instrumental Variables: Application and Limitations // Epidemiology. 2006. Vol. 17(3). P. 260–267.

14. *Bound J., Jaeger D.A., Baker R.* Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak // Journal of the American Statistical Association. 1995. Vol. 90(430). P. 443–450.

15. *Staiger D., Stock J.H.* Instrumental Variables Regression with Weak Instruments // Econometrica. 1997. Vol. 65(3). P. 557–586.

16. *Kolesár M., Chetty R., Friedman J.N., Glaeser E.L., Imbens G.W.* Identification And Inference With Many Invalid Instrument. Working Paper 17519, National Bureau Of Economic Research, 2011. URL: <http://www.nber.org/papers/w17519>.

17. *Van der Klaauw*. Regression-discontinuity Analysis: A Survey of Recent Developments in Economics // *Labour*. 2008. Vol. 22.
18. *Imbens G.W., Angrist J.D.* Identification and Estimation of Local Average Treatment Effects // *Econometrica*. 1994. Vol. 62(2). P. 467–475.
19. *Heckman J.J., Urzua S.* Comparing IV with Structural Models: What Simple IV Can and Cannot Identify // *Journal of Econometrics*. 2009. Vol. 156(1). P. 27–37.
20. *Imbens G.W., Wooldridge J.* What's New in Econometrics: Instrumental Variables with Treatment Effect Heterogeneity: Local Average Treatment Effects // *Lecture Notes*, 2007. URL: http://www.nber.org/WNE/lect_5_late_fig.pdf.
21. *Lee D.S., Lemieux T.* Regression Discontinuity Designs in Economics. NBER Working Paper 14723, 2009.
22. *Rosenbaum P., Rubin D.* The Central Role of the Propensity Score in Observational Studies for Causal Effects // *Biometrika*. 1983. Vol. 70. P. 41–55.
23. *Caliendo M., Kopeinig S.* Some Practical Guidance for the Implementation of Propensity Score Matching // *Journal of Economic Surveys*. 2008. Vol. 22(1). P. 31–72.
24. *Brookhart M.A., Schneeweiss S., Rothman K.J., Glynn R.J., Avorn J., Sturmer T.* Variable Selection for Propensity Score Models // *American Journal of Epidemiology*. 2006. Vol. 163 (12).
25. *D'Agostino R.B., Jr, D'Agostino R.B., Sr.* Estimating Treatment Effects Using Observational Data // *Journal of American Medical Association*. 2007. Vol. 297. P. 314–316.
26. *Augurzky, B., Schmidt C.* The Propensity Score: A Means to An End, Working Paper. University of Heidelberg, 2000
27. *Bhattacharya J., Vogt W.B.* Do Instrumental Variables Belong in Propensity Scores? Technical Working Paper 343. URL: <http://www.nber.org/papers/t0343>.
28. *Sen S.* Causal Inference: Propensity Scores, Instrumental Variables, and Sensitivity Analyses // *Epidemiology*. 2006. Vol. 219.
29. *Khandker S.R., Koolwal G.B., Samad H.A.* Handbook on Impact Evaluation: Quantitative Methods and Practices. World Bank Training Series, 2009.
30. *Becker S.O., Ichino A.* Estimation of Average Treatment Effects Based on Propensity Scores // *The Stata Journal*. 2002. Vol. 4. P. 358–377.
31. *Nannicini T.* Simulation-based Sensitivity Analysis for Matching Estimators // *The Stata Journal*. 2007. Vol. 7(3).

Kuzmina Yulia

*National Research University Higher School of Economics (NRU HSE),
Moscow, papushka7@gmail.com*

Regression Discontinuity Design Method and Propensity Score Matching (PSM) in One Year of Schooling Effect Evaluation: An Experience of Applying to PISA 2009 Data

This paper reports the findings of an analysis of the effect of one year of schooling on math performance for 15-year-olds in ten different countries. The analysis was done on PISA 2009 data by applying two quasi-experimental methods – instrumental variable analysis (IV analysis) within fuzzy Regression Discontinuity Design (RDD) framework and Propensity Score Matching (PSM). We find that a year of schooling has positive impact on achievement in all countries except Albania. The advantages and restrictions of applying both methods are discussed.

Keywords: treatment effect, quasi-experimental methods, regression discontinuity design, instrumental variable, propensity score matching, the effect of one year's of schooling, sensitivity analysis, PISA

References

1. Campbell D. *Models of Experiments in Social Psychology and Applied Researches* (transl., in Russian), SPb.: Socialno-psikhol. Centr, 1996.
2. Schneider B., Carnoy M., Kilpatrick J., Schmidt W.H., Shavelson R.J. *Estimating Causal Effects Using Experimental and Observational Designs* (Report from the Governing Board of the American Educational Research Association Grants Program). Washington, DC: American Educational Research Association, 2007.
3. Seidel T., Shavelson R.J. "Teaching Effectiveness Research in the Past Decade: the Role of Theory and Research Design in Disentangling Meta-analysis Results", *Review of Educational Research*, 2007, 77 (4), 454–499.
4. Luyten H., Peschar J., Coe R. "Effects of Schooling on Reading Performance, Reading Engagement, and Reading Activities of 15-Year- Olds in England", *American Education Research J.*, 2008, 45(2), 319–342.
5. Luyten H. "An empirical assessment of the absolute effect of schooling: regression-discontinuity applied to TIMSS-95", *Oxford Review of Education*, 2006, 32 (3), 397–429.
6. Tyumeneva Y.A., Kuzmina Y.V. "Chto daet god obuchenija rossijskomu shkol'niku. Na materialah PISA-2009: gramotnost' chteniija" (What gives the year of study of Russian schoolchildren. On materials of the PISA-2009: The reading literacy), *Voprosy obrazovaniya*, 2013, 1, 107-136.

7. OECD. *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)*. Paris, 2010.
8. Аурай А. “The Tough Choice at High School Door: An Investigation of the Factors That Lead Students to General or Vocational Schools”, *International Journal of Educational Development*, 2003, 23 (5), 517–527.
9. Suggate S.P. “School entry age and reading achievement in the 2006 Programme for International Student Assessment (PISA)”, *International Journal of Educational Research*, 2009, 48 (3), 151–161.
10. Thistlethwait D.L., Campbell D.T. *Regression-discontinuity analysis: An alternative to the ex-post facto experiment*. Beverly Hills: Sage Publications, 1960.
11. Imbens G.W., Lemieux T. “Regression discontinuity designs: A guide to practice”, *Journal of Econometrics*, 2008, 142 (2), 615–635.
12. Angrist J.D., Imbens G.W., Rubin D.B. “Identification of Causal Effects Using Instrumental Variables”, *Journal of the American Statistical Association*, 1996, 91(434), 444–455.
13. Martens E.P., Pestman W.R., Boer A., Belitser S. V., Klungel O.H. “Instrumental Variables: Application and Limitations”, *Epidemiology*, 2006, 17(3), 260–267.
14. Bound J., Jaeger D.A., Baker R. “Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variables is Weak”, *Journal of the American Statistical Association*, 1995, 90 (430), 443–450.
15. Staiger D., Stock J.H. “Instrumental Variables Regression with Weak Instruments”, *Econometrica*, 1997, 65(3), 557–586.
16. Kolesár M., Chetty R., Friedman J.N., Glaeser E.L., Imbens G.W. *Identification And Inference With Many Invalid Instrument*. Working Paper 17519, National Bureau Of Economic Research, 2011. URL: <http://www.nber.org/papers/w17519>.
17. Van der Klaauw. “Regression-discontinuity analysis: A survey of recent developments in economics”, *Labour*, 2008, Vol. 22.
18. Imbens G.W., Angrist J.D. “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 1994, 62(2), 467–475.
19. Heckman J.J., Urzua S. “Comparing IV with structural models: What simple IV can and cannot identify”, *Journal of Econometrics*, 2009, 156 (1), 27–37.
20. Imbens G.W., Wooldridge J. “What’s New in Econometrics: Instrumental Variables with Treatment Effect Heterogeneity: Local Average Treatment Effects”, *Lecture Notes*, 2007. URL: http://www.nber.org/WNE/lect_5_late_fig.pdf.
21. Lee D.S., Lemieux T. *Regression discontinuity designs in economics*. NBER Working Paper 14723, 2009.
22. Rosenbaum P., Rubin D. “The central role of the propensity score in observational studies for causal effects”, *Biometrika*, 1983, 70, 41–55.
23. Caliendo M., Kopeinig S. “Some Practical Guidance for the Implementation of Propensity Score Matching”, *Journal of Economic Surveys*, 2008, 22(1), 31–72.

24. Brookhart M.A., Schneeweiss S., Rothman K. J., Glynn R. J., Avorn J., Sturmer T. “Variable Selection for Propensity Score Models”, *American Journal of Epidemiology*, 2006, 163 (12).
25. D’Agostino R.B., Jr, D’Agostino R.B., Sr. “Estimating Treatment Effects Using Observational Data”, *Journal of American Medical Association*, 2007, 297, 314–316.
26. Augurzky, B., Schmidt C. *The Propensity Score: A Means to An End*. Working Paper. University of Heidelberg, 2000
27. Bhattacharya J., Vogt W.B. *Do instrumental variables belong in propensity scores?* Technical Working Paper 343. URL: <http://www.nber.org/papers/t0343>.
28. Sen S. “Causal Inference: Propensity Scores, Instrumental Variables, and Sensitivity Analyses”, *Epidemiology*, 2006, 219.
29. Khandker S.R., Koolwal G.B., Samad H.A. *Handbook on Impact Evaluation: Quantitative Methods and Practices*. World Bank Training Series, 2009.
30. Becker S.O., Ichino A. “Estimation of average treatment effects based on propensity scores”, *The Stata Journal*, 2002, 4, 358–377.
31. Nannicini T. “Simulation-based sensitivity analysis for matching estimators”, *The Stata Journal*, 2007, 7(3).