

НЕКОРРЕКТНЫЕ ЗАИМСТВОВАНИЯ В РОССИЙСКИХ ДОКТОРСКИХ ДИССЕРТАЦИЯХ: СКОЛЬКО, ГДЕ И У КОГО?

Александра Макеева, Анжелика Цивинская,
Михаил Соколов, Надежда Соколова, Катерина Губа

Центр институционального анализа науки и образования
Европейский университет в Санкт-Петербурге

Препринт ЦИАНО ННС-С (5) 2019
ciase.ru

Макеева А., Цивинская А., Соколов М., Соколова Н., Губа К. Некорректные заимствования в российских докторских диссертациях: сколько, где и у кого? (Серия препринтов ННС-С (5)). СПб: ЦИАНО ЕУСПб, 2019. — 37 с.

Макеева Александра Алексеевна — лаборант Центра институционального анализа науки и образования, Европейский университет в Санкт-Петербурге.

Цивинская Анжелика Олеговна — младший научный сотрудник Центра институционального анализа науки и образования, Европейский университет в Санкт-Петербурге.

Соколов Михаил Михайлович — научный руководитель Центра институционального анализа науки и образования, профессор Европейского университета в Санкт-Петербурге.

Соколова Надежда Александровна — младший научный сотрудник Центра институционального анализа науки и образования, Европейский университет в Санкт-Петербурге.

Губа Катерина Сергеевна — директор Центра институционального анализа науки и образования, Европейский университет в Санкт-Петербурге.

Центр институционального анализа науки и образования открыт Европейским университетом в Санкт-Петербурге в 2017 году. Центр специализируется на изучении институциональных форм, в которых протекает академическая жизнь — организаций (университетов, исследовательских институтов), арен научной коммуникации (журналов, конференций), форм дисциплинарной самоорганизации (профессиональных ассоциаций). Центр исследует, под воздействием каких факторов эти формы возникают (государственная политика управления наукой и образованием, рыночные запросы, культурные, технологические и демографические сдвиги) и как они влияют на природу производимого знания.

Контакты:

**191187, Санкт-Петербург,
Гагаринская ул., д. 6/1, литера А,
Научно-исследовательский центр
«Центр институционального анализа науки и образования»
www.ciase.ru**

Аннотация

Некорректные заимствования в диссертациях политиков и иных публичных фигур часто становятся предметом громких скандалов.¹ Однако активистские расследования не позволяют составить общей картины распространенности плагиата в России. Мы попробовали восполнить этот пробел, проанализировав случайную выборку из 2468 докторских диссертаций, защищенных в 2006-2015 годах (8,8% от генеральной совокупности) с помощью скриптов программы «Антиплагиат». Наши данные позволяют не только оценить общее распространение плагиата, но и проверить несколько гипотез о том, какие факторы стоят за практикой заимствования. Мы рассматриваем три объяснения, основанные на теоретических моделях альтернативных конвенций (группы ученых разделяют разные представления о допустимых заимствованиях), аномии (за плагиатом стоят сбои в академической социализации) и рационального выбора (за заимствованием скрывается рациональное взвешивание издержек и выгод нарушения нормы). Затем мы анализируем влияние на вероятность появления плагиата ряда независимых переменных, соотносимых с этими гипотезами – дисциплины, региона и типа организации, где прошла защита, года защиты и наукометрических показателей диссертанта. Средняя доля обнаруженных заимствований составила 19,1%, медианная – 13,9%. Только четверть диссертаций имеет долю заимствований менее 7% текста, при этом, однако, лишь 6,1% авторов заимствовали более 50%. В целом, гипотеза альтернативных конвенций в свете полученных результатов выглядит наиболее убедительно: переменные, которые являются вероятными коррелятами аномии (экспансия дисциплины, высокие показатели имитативности научной деятельности) не оказывают на вероятность появления плагиата значимого воздействия. В то время как переменные, существенные с точки зрения модели рационального выбора (индивидуальная сила ученого, сила институции, распространение «Антиплагиата»), оказывают лишь слабое влияние.

1 Сбор данных и разработка инструментария проводились при финансовой поддержке АНО «Институт независимых исследований в области образования» в рамках договора на выполнение научно-исследовательской работы по теме «Где искать академическую нечестность?». Мы благодарны также компании «Антиплагиат» за содействие в проведении исследования.

Некорректные заимствования в российских докторских диссертациях: сколько, где и у кого?

Случаи обнаружения некорректных заимствований в диссертациях публичных фигур часто становятся причинами громких скандалов. Только за последние годы подобные скандалы, в которые были вовлечены известные политики, разразились в Германии, США и Финляндии². Разоблачения группы «Диссернет» коснулись и российских политиков, чиновников, многочисленных преподавателей и администраторов от науки и образования (ректоры вузов, директора школ). Вместе с тем отсутствуют данные, которые позволили бы оценить ситуацию с плагиатом в российской науке в целом, проследить ее динамику и ответить на вопрос о том, какие факторы стоят за появлением некорректных заимствований в диссертациях. «Диссернет» по результатам своих исследований публиковал распределения выявленных текстов с признаками обширного плагиата по дисциплинам, однако, в отсутствие систематической процедуры отбора диссертаций для проверок и четких критериев того, что заслуживает публичного осуждения как плагиат, а что – нет, возможность делать выводы на основании этих данных ограничена [Вольное 2018]. Нисколько не умаляя значения отдельных разоблачений научной недобросовестности, мы считаем, что для получения общей картины необходим более систематический подход.

В этой статье мы делаем попытку реализовать такой подход, представляя результаты проверки случайной выборки докторских диссертаций, защищенных в 2006-2015 годах, используя компьютерные алгоритмы системы «Антиплагиат». Это позволяет оценить общие координаты распространения некорректных заимствований, узнать, как изменяется их объем, и выявить, какие факторы приводят индивидов в «группу риска». Помимо всего прочего, подобный анализ позволяет сказать, насколько ситуация с некорректными заимствованиями в российских диссертациях вписывается в модели объяснения девиантного поведения, существующие в мировой литературе (рационального выбора, аномии, альтернативных конвенций).

В следующем параграфе мы рассмотрим общие подходы к пониманию плагиата и академической недобросовестности, затем опишем процедуры нашего исследования и представим его результаты. В заключении мы вернемся к вопросу о том, какие объяснения академической недобросовестности подходят к ситуации с российскими диссертациями.

² В 2011 году министр обороны ФРГ Карл-Теодор цу Гуттенберг вынужден был оставить свой пост после того, как в его диссертации было обнаружено порядка 75% плагиата. В США в 2017 году в центре скандала оказалась Моника Коули, которую президент Трамп хотел видеть своим представителем в Национальном совете безопасности (назначение не состоялось). В 2018 в Финляндии в плагиате в магистерской диссертации была обвинена Лаура Хухтасаари, кандидат в президенты от правой партии «Истинные финны» (степень не была отозвана, поскольку скопирована была лишь часть литературного обзора, но скандал мог поспособствовать поражению Хухтасаари на выборах).

1 Социологические подходы к объяснению некорректных заимствований

1.1 Проблема определения «некорректного заимствования»

Представления о том, какие права ученый имеет на плоды своего труда, изменялось со временем. Еще в середине XVII-го века английские суды разбирались с тем, сохраняет ли ученый, продавший свою рукопись книгоиздателю, хоть какие-то права на нее – например, право решать, будет ли его имя фигурировать на обложке изданной книги [Johns 2003]³. В то же время среди естествоиспытателей круга Роберта Бойля преобладало мнение, что, даже если такое право и может быть защищено законом, настоящему ученому безусловно следует публиковать свои открытия анонимно, поскольку, во-первых, это обеспечит большее доверие публики (автор, не ставящий на рукописи свое имя, вероятно, движим любовью к истине, а не тщеславием), а, во-вторых, естествоиспытатели в любом случае лишь пересказывают то, что подсмотрели в книге, написанной Создателем [Shapin 1995]. Тем не менее, именно в поколении Бойля в Англии распространяется наше современное представление о научном авторстве и практик, которые отражают его. Это распространение совпадает с возникновением современной экспериментальной науки, и совпадение кажется слишком точным, чтобы быть случайным. Связь собственнического отношения к открытиям с духом современной науки интересовала, среди прочих, Роберта Мертона [Merton 1973]. Мертон утверждал, что интеллектуальная собственность автора на свои открытия гарантирует индивиду вознаграждение за внесение вклада в копилку общего знания. Собственности в данном случае соответствует не право на информацию, позволяющее предотвращать ее распространение, как в случае с более древними производственными секретами, но право требовать, чтобы каждый распространитель распространял ее только с упоминанием оригинального производителя. Только это, предполагал Мертон, может служить условием, на котором ученые будут соглашаться производить информацию и помещать ее в открытый доступ. Поэтому рождение современной науки сопровождают споры о приоритете и авторстве (иногда довольно уродливые). Та же идея встречается в институциональной экономике науки, возводящей нашу концепцию авторства к потребности в интернализации экстерналий производства коллективных благ общедоступной информации [Stephans 1996].

По Мертону, плагиат представляет собой симптом распада научного этики, на котором зиждутся институты науки. Плагиатор, с этой точки зрения, совершает сразу два преступления: кражу и мошенничество, присутствующие в разных пропорциях в конкретном акте некорректного заимствования. Во-первых, он крадет у подлинного автора, отнимая у того право на заслуженное признание. Во-вторых, он мошеннически вводит в заблуждение общество, принимающему за «настоящего» доктора человека, который получил степень, не пройдя всех необходимых испытаний. Жертвой тут являются будущие работодатели, клиенты или студенты фальшивого ученого, а также коллеги – другие обладатели той же степени, – которые могут пострадать в результате того, что репутация дисциплины ухудшится. Поэтому плагиат остается проступком, даже если автор присвоенного текста не возражает и даже сам предла-

³ Удивительным образом первые иски такого рода были поданы учеными, которые не хотели видеть свое имя на обложке написанных ими книг [Johns 2003]. Если книга попадала под обвинения в атеизме и материализме, робкие авторы часто предпочитали, чтобы об их связи с ней забыли, а издатели, напротив, надеялись, что имя скандального вольнодумца привлечет интерес читателей.

гает использовать его таким образом. Элемент кражи в этом случае отсутствует, зато оба оказываются замешаны в мошенничестве.

Единодушная поддержка нормы, запрещающей плагиат, не означает, однако, единства ее толкования. Действительно, что составляет авторство и что может пониматься как корректное заимствование, а что – нет, – является предметом разных трактовок. Каждый автор заимствует что-то у своих предшественников. Мы используем шрифты и системы цитирования, вообще говоря, тоже кем-то разработанные, оформляем таблицы и диаграммы по изобретенным кем-то шаблонам, пишем статьи согласно предложенным кем-то композиционным конвенциям – и не цитируем этих людей⁴. Большинство ученых скажут, что мы и не должны, поскольку все вышеперечисленное составляет просто форму научного сообщения. Однако по поводу того, что составляет форму, а что – содержание научной работы нет общего согласия. Является ли формой или содержанием определенный протокол эксперимента (и если он является просто формой, можно ли целиком заимствовать его из другой статьи)? Статистический метод? Список литературы?

В случае с научным текстом обвинения в некорректных заимствованиях могут принять три формы. А может обвинить В в том, что тот заимствовал его идеи, данные или слова – текст как таковой. Разумеется, это не взаимоисключающие вещи. Если В украл текст целиком, заменив только первую страницу с именем, то украдено и первое, и второе, и третье. Однако до определенных пределов возможна и кража чего-то одного – скажем, А украл идею революционного эксперимента, но поставил его сам и написал текст самостоятельно; или взял большие куски текста с обзором литературы и описанием процедур исследования, но использовал данные этого исследования, чтобы развить собственные мысли. Под большинство определений академического плагиата, которыми пользуются соответствующие агенты (включая российский ВАК [*Шахрай и др.* 2015]), подпадает и то, и другое, и третье, однако из повседневной жизни известно, что им приписывается разная степень непростительности. Заимствование и использование чужой идеи, особенно таким способом, что ей уже не сможет воспользоваться законный автор, интуитивно кажется более важным и тяжелым прегрешением, а заимствование слов и предложений, но не идей – относительно мелким. К несчастью, кражу идей почти невозможно доказать в судебном порядке⁵. Напротив, благодаря программному обеспечению, распознавание текстуальных заимствований является разрешимой задачей, хотя само по себе оно и наименее важно. Только оно, однако, может быть использовано активистами в качестве судебного доказательства или включено в статистический анализ как зависимая переменная.

В статье мы обратимся к изучению случаев следования или нарушений именно этого аспекта нормы, который можно назвать *нормой текстуальной аутентичности*. Эта норма требует оригинального сочинения любого академического текста, написанного с использованием естественного языка. Иными словами, в случае с диссертантами она требует написать дис-

4 Один из авторов этой статьи в свое время потратил значительное время на то, чтобы идентифицировать автора одной из самых важных жанровых форм в российской науке – автореферата (неуспешно).

5 Исторические расследования научных преступлений обычно могут доказать, как максимум, вероятность заимствования. Скажем, теоретически, однажды может всплыть переписка, показывающая, что Лобачевский и Бойяи знали о неопубликованных разработках Гаусса в области неевклидовой геометрии. Тогда на оригинальность их работы падет тень подозрения. Судебных перспектив такое подозрение, однако, заведомо не будет иметь – невозможно доказать, что озарение не посетило их до того, как работа Гаусса попала в поле их зрения.

сертификацию собственноручно от первого до последнего слова. Именно случаи нарушения этой нормы, благодаря развитию машинного анализа текстов, могут быть выявлены со значительной точностью на больших массивах данных, и мы можем проанализировать, кто следует или не следует ей.

1.2 Теоретические объяснения академических девиаций и исследовательские гипотезы

С чем связано следование или неследование академической норме? В социальных науках есть несколько готовых объяснений, которые обычно применяются для случаев поведения, определяемого как нарушение нормы. Во-первых, нарушители могут не разделять эту норму в принципе – норма действительна в одной группе и не принимается в другой; если эти группы соприкасаются и если одна из них обладает большей властью, то слабая группа помечается ярлыком девианта (напр., [Becker 2000]). Применительно к плагиату мы можем предполагать, что разные группы ученых по-разному понимают, что такое оригинальность и где проходит граница между формой и содержанием.

Во-вторых, даже в группе, в которой есть консенсус по поводу нормы, находятся индивиды, нарушающие ее в корыстных целях. Объяснения девиации распадаются на два больших класса – нормативные или социализационные, с одной стороны, и рационально-структурные – с другой. В первом случае девиантным индивидам известно о норме, но у них нет моральных ограничений, которые должны сдерживать от пренебрежения ею [Thomas 1923]. В отношении плагиата такое объяснение подразумевает, что его появление является результатом сбоя в академической социализации или в уровне морального контроля над индивидом со стороны сообщества – по каким-то причинам некоторые не воспринимают этос, который должен уберечь их от подобных проступков. Основное эмпирическое ожидание, следующее из этой теории, будет состоять в том, что нечестные индивиды более склонны и к другим способам «срезать углы» – фейковым публикациям, публикациям в мусорных журналах, созданию «цитатных» картелей и применению иных способов накачивания наукометрических показателей. Из этого следует Гипотеза 1.

Гипотеза 1. *Чем интенсивнее нарушения нормы текстовой аутентичности, тем выше вероятность встретить другие формы академической нечестности⁶.*

Далее, многие из тех, кто писал о распространении диссертационного плагиата в России, исходили из того, что он типичен для дисциплин, действие академических норм в которых было слабым в силу их недавней экспансии. Скажем, под воздействием взрывного студенческого спроса, программы по экономике и менеджменту множились, создавая параллельный спрос университетов на преподавателей с соответствующими степенями. Этот спрос множил ряды диссертантов, которые оказались в академическом мире случайно и не интернализировали научный этос; соответственно, они легко нарушали все мыслимые нормы научной жизни. Эти рассуждения подводят нас к Гипотезе 2.

⁶ Теории, выводящие девиацию из альтернативных норм, и теории, выводящие ее из индивидуального отклонения, допускают множество переходных степеней: девиантные индивиды и особенно группы таковых могут выдумывать собственные нормы, оправдывающие их поведение (как в теориях нейтрализации [Sykes, Matza 1957]). Кроме того, как указывает Беккер, индивиды и группы, получающие ярлык девиантных, часто вынуждены воспринимать другие практики, имеющие то же клеймо, что создает своего рода самоисполняющееся пророчество. Тем не менее аналитически различие в предполагаемом механизме ясно.

Гипотеза 2. *Чем сильнее была экспансия дисциплины после 1991 года, тем выше доля некорректных заимствований в ней.*

Рационально-структурные теории (напр., [Becker 1968]) утверждают, что все индивиды одинаково оппортунистичны и мало отличаются друг от друга в отношении испытываемых ими моральных ограничений, но сильно отличаются в отношении относительных затрат и выгод, ожидаемых от нарушения нормы. Если социализационная теория предполагает, что индивиды следуют норме, поскольку считают ее правильной и легитимной, то рационально-структурная теория считает, что норме с большей вероятностью следуют те, для кого затраты на следование ниже, а риски нарушения – наоборот, выше. Применительно к плагиату из рационально-структурной теории следует, что к некорректным заимствованиям будут склонны те, кому (а) сложнее всего написать оригинальный текст (потому что выгода копирования чужой работы для них больше, чем для тех, для кого это не проблема) и (б) кто по каким-то причинам меньше других боится разоблачения. Первое предполагает, что распространение плагиата будет характеризовать «слабую» науку. Тогда следует ожидать, что наличие некорректно цитируемого текста будет сопровождаться низкими значениями показателей цитирования и прочих характеризующих интеллектуальное влияние метрик. Кроме того, мы предполагаем, что заимствования будут присутствовать в работах диссертантов, представляющих скорее слабые в исследовательском плане институции, чем сильные, а также в работах диссертантов из академически более слабых локаций. Это дает нам Гипотезу 3.

Гипотеза 3. *Чем выше доля заимствования, тем ниже показатели, характеризующие научное влияние и признание индивида.*

Отметим отличие этой гипотезы от Гипотезы 1. Согласно Гипотезе 1, в публикационной биографии плагиаторов найдутся факты, свидетельствующие о попытке «накачивать» показатели, но это не исключает того, что эти показатели будут высокими. Напротив, Гипотеза 3 утверждает, что мы найдем среди плагиаторов индивидов с низким влиянием, хотя не обязательно использующих мусорные журналы или иные способы обмана наукометрии.

Важное замечание, которое необходимо сделать в этом месте. Наше исследование не было направлено на то, чтобы определить, правда ли строгое соблюдение нормы текстуальной аутентичности – пример «хорошего института» в науке, который способствует ее развитию. Тем не менее его данные позволяют проверить валидность по крайней мере одного из рациональных обоснований этой нормы. В духе теории рыночных сигналов [Spence 1973; Stiglitz 2000] норма текстуальной аутентичности может быть обоснована тем, что для более сильных ученых выигрыши, связанные с заимствованием чужого текста, ниже, чем для более слабых авторов. Попросту, хороший ученый потратит на то, чтобы корректно описать другими словами тот же эксперимент, несколько часов. Плохой ученый, который не понимает, что это за эксперимент, потратит на это месяцы и, возможно, вовсе откажется от замысла дописывать диссертацию. Если это так, насаждение нормы текстуальной аутентичности приведет к повышению сортирующей способности диссертационного сита. Результаты проверки нашей Гипотезы 3 позволят сказать, испытывают ли сильные ученые особое отвращение к плагиату и, соответственно, валиден ли этот аргумент.

Далее, мы можем предполагать, что риски разоблачения увеличивались в течение изучаемого периода – «Антиплагиат» получал все более широкое распространение, «Диссернет» публиковал результаты своих исследований, а ВАК стремился закрутить гайки. Риски разоблачения, можем предполагать мы, непрерывно возрастали; соответственно, наши рацию-

нальные оппортунисты должны быть отказываться от копипэйста в пользу оригинального написания текста. Отсюда следует Гипотеза 4:

Гипотеза 4. *Чем позже период, к которому относится анализируемая диссертация, тем меньше в ней доля заимствований.*

Последняя из наших трех теорий, теория альтернативных конвенций, предполагает, что ответственные за то, что кажется «Антиплагиату» некорректными заимствованиями, на самом деле просто следуют альтернативным нормам. Они не считают, что ведут себя противно научному этосу, и не ожидают, что столкнутся с отрицательными санкциями. Они и их окружение не полагают, что копирование фрагментов чужих текстов обязательно подпадает под категорию нарушения прав интеллектуальной собственности или демонстрации интеллектуальной слабости. Возможно, например, что в их восприятии плагиат – допустимая демонстрация пренебрежения к процессу защиты (или ее составляющим, таким как написание раздела «актуальность работы»⁷) как «пустой формальности» или они разделяют сильную идеологию, отвергающую идею индивидуального авторства в принципе (в этом случае особенно вероятно заимствование у своих учеников или учителей). При этом такие представления могут разделяться не только соискателями, но и всеми возможными участниками процесса защиты, включая экспертов ВАК по какой-то специальности.

Из этой теории сложно сделать какие-то позитивные гипотезы о том, кто будет замечен в нарушении норм текстуальной аутентичности. Скорее, она может объяснить, почему предсказания, следующие из других теорий, не осуществляются. Самое большее – мы можем связать распространение более строгих конвенций, характерных для англо-американской науки, с большей глобализованностью академической среды. Так, в случае с плагиатом, доля заимствований будет ниже в более интернационализированных институциях, скажем, в университетах Программы «5/100», по сравнению с обычными вузами. Это предсказание, впрочем, не противоречит ожиданиям, основанным на двух других теориях. Мы можем предполагать, что в лучших университетах будут сконцентрированы также индивиды, впитавшие ценности этоса и не имеющие сложностей с тем, чтобы написать статью самостоятельно. Скорее, мы можем интерпретировать как аргумент в пользу модели альтернативных конвенций то, что влияние глобализованности институций на объем заимствований будет единственным наблюдаемым эффектом в условиях, когда все остальные окажутся незначимыми. Продолжая эту линию рассуждений, мы можем предполагать также, что вузы в «мировых городах» – прежде всего Москве и Петербурге – первыми воспримут импортированные нормы.

Гипотеза 5. *Чем выше глобализация академической среды, тем ниже уровень некорректных заимствований.*

⁷ Интересно, что это понимание разделяет сам «Диссернет», который не проверяет вводные части работ на оригинальность – видимо, предполагая, что ни один человек в здравом уме не будет писать раздел про актуальность и практическую значимость своей работы самостоятельно.

2 Эмпирические исследования академической недобросовестности

Академическая нечестность – ни в коем случае не специфически российская проблема. Международные сравнения показывают, что большинство студентов в мире списывали хотя бы раз за время учебы [McCabe 1992; McCabe, Bowers 1994; Stearns 1997]. В настоящее время большинство исследований академической нечестности сконцентрировано на нечестности студентов [Fields 2003; Ward, Beck 1990; Macfarlane et al. 2014], восприятию ими этических кодексов [Gundersen et al. 2008; McKay et al. 2007; Rezaee et al. 2001], а также кодексов как таковых [Kelley et al. 2006; Lind 2005; Schoenherr, Williams-Jones 2011]. Именно студенты чаще всего становятся объектом исследования в работах по изучению академической нечестности. Количество работ, посвященных эмпирическому изучению распространенности нечестности среди ученых или аспирантов – на порядок меньше [Wajda-Johnston et al. 2001]. Однако фабрикация данных, подлог и плагиат распространены не только в среде студентов, о чем в том числе свидетельствуют регулярные случаи отзыва статей редакциями журналов. То, что студенты оказываются предпочитаемым объектом исследования, связано, вероятно, с тем, что состоявшиеся исследователи (имеющие степень, публикации и т.д.) обычно вызывают меньше подозрений, нежели студенты, с другой – с тем, что изучать студентов проще. Чаще всего в исследованиях академической нечестности применяются опросы, а организовать представительный опрос студентов значительно легче, чем «взрослых» ученых. Кроме того, есть подозрение, что опрос вообще плохо подходит для изучения девиантного поведения состоявшихся ученых, так как они несут большие репутационные риски и потому менее склонны сообщать о сомнительном поведении.

К счастью, существуют неопросные методы, которые могут пролить свет на масштабы академической нечестности и ее детерминанты. Так, для исследователей может быть интересна информация об отзывах статей, а также данные, содержащиеся в закрытых делах ORI (Office of research integrity). Однако такие источники данных имеют свои ограничения, связанные с тем, что у исследователей есть информация только об обнаруженных случаях нарушения академических норм, в то время как оставшиеся «нераскрытыми» случаи остаются неисследованными. Опора на такую информационную базу может привести к искаженным выводам, если не все ученые имеют равные шансы подвергнуться проверке. Проблема хорошо известна в криминологии: статистика о количестве наркотиков, изъятых у разных категорий населения, не позволяет судить, насколько они вовлечены в наркобизнес и наркопотребление, если нет сведений о том, с какой частотой они проверялись. Представим себе город, населенный двумя этническими группами, А и Б, среди которых наркотики имеют равное распространение. Среди полицейских, однако, существует предрассудок, что А куда более склонны к расширению сознания неодобряемыми законом средствами, чем Б (такое легко себе представить, если полицейские в основном рекрутируются из Б). Поэтому А они проверяют в 10 раз чаще и, закономерно, в статистике Б оказываются вовлечены в наркопреступления в 10 раз реже. Такую же ошибку можно предполагать в исследованиях, опирающихся на данные ретракций и закрытых дел. Так, согласно модели, построенной на основе теории игр, ученые, уличенные в нечестности, на самом деле могут нарушать академические нормы реже, нежели те, кто в этом уличен не был [Lacetera, Zirulia 2009; Hoover 2006]. В исследованиях, опирающихся на выявленные случаи, происходит неизбежное смещение, и выводы зачастую не могут быть экстраполированы на всю популяцию ученых. К несчастью, такая же проблема существует с результатами проверок «Диссернетом», которые направлялись подо-

зрениями участников группы в отношении того, где может быть плагиат, и соображениями политической важности изобличения той или иной крупной фигуры⁸.

Таким образом, единственным достоверным способом оценить масштабы и детерминанты нечестного поведения (например, некорректного заимствования) представляется проверка случайных выборок ученых, а не только тех, чье сомнительное поведение было «раскрыто». В настоящее время существуют технические средства, позволяющие проверять тексты работ на наличие некорректных текстовых заимствований. Использование такого программного обеспечения сопряжено с определенными техническими и порой юридическими сложностями: начиная от необходимости отсматривать отчеты программы вручную и заканчивая вопросом авторского права и открытого доступа к публикациям. Несмотря на это, в последние годы появляются статьи, использующие такое ПО (преимущественно «Turnitin») для оценки некорректных заимствований, в том числе самоплагиата [Honig, Bedi 2012; Horbach, Halffman 2017]. Однако такие работы в основном посвящены отдельным группам ученых (например, участникам конференции Академии менеджмента [Honig, Bedi 2012] и голландским ученым из четырех областей знаний [Horbach, Halffman 2017]). Насколько нам известно, до сих пор не предпринималось попыток оценить масштабы заимствований во всей популяции ученых той или иной страны. В настоящей работе мы стремимся заполнить этот пробел на примере российских докторов наук.

Что предсказывает вероятность быть вовлеченным в академическую недобросовестность? В литературе исследуется влияние трех групп показателей – личных характеристик, характеристик организации и страновых характеристик. В исследованиях персональных характеристик ученых основное внимание уделяется некоторым их психологическим чертам [Antes et al. 2007], карьерной ступени и полу. Результаты таких исследований пока не складываются в какую-то общую картину. Некоторые из них показывают, что мужчины более склонны к академической нечестности [Fang et al. 2013]. Однако в исследованиях, в которых вводится контроль по наличию и развитию этических кодексов, статусу ученого и культурным переменным странового и организационного уровня, гендер не является статистически значимым предиктором [Fanelli et al. 2015]. Недавнее исследование Фанелли также не подтвердило связи пола с академической нечестностью [Fanelli et al. 2018]. Катц с соавторами также сомневаются в гендерных различиях среди нарушителей академических норм: во-первых, «преступления чаще приписываются мужчинам» и потому чаще обнаруживаются в их среде, а, во-вторых, мужчины получают больше исследовательского финансирования, и, соответственно, чаще оказываются в ситуации, стимулирующей нечестность [Kaatz et al. 2013]. Не обнаружено связи академической нечестности с полом ученого и в исследовании авторов, специализирующихся на изучении менеджмента [Honig, Bedi 2012].

Также неоднозначны выводы о связи нечестности с положением ученого на карьерной лестнице. В то время как одни исследователи обнаруживают, что молодые исследователи более склонны к нечестности [Lewellyn et al. 2017; Fanelli et al. 2015], другие указывают, что нечестность встречается с равной вероятностью на всех ступенях карьеры [Fang et al. 2013; Honig, Bedi 2012]. Тем не менее результаты Фанелли и др., указывающие на меньшую добро-

⁸ Обычным решением проблемы избирательных проверок является использование двухуровневых регрессий Хекмана, вводящих в качестве отдельной независимой переменной вероятность быть проверенным. К несчастью, для их использования надо обладать информацией о проверках, закончившихся и позитивным, и негативным результатом, а она обычно недоступна.

совестность младших когорт, выглядят на нынешний момент наиболее полными и убедительными: в их работе влияние ступени карьеры контролируется по наибольшему набору других переменных. Полученные результаты объясняются в первую очередь тем, что молодые ученые, в соответствии с ожиданиями, основанными на рационально-структурной модели, потеряют от разоблачения меньше (они еще не создали себе репутацию, которую можно уничтожить, и им легче переключиться на другую сферу деятельности, если все двери в науке для них закроются). Кроме того, в духе нормативной теории, мы можем предполагать, что молодые исследователи еще не до конца усвоили этос и нормы научной жизни. Однако, как показало другое исследование, в случае самоплагиата более «взрослые» ученые демонстрируют существенно бóльшую частоту сомнительного поведения, нежели «молодые» [Horbach, Halffman 2017]. Тут, впрочем, надо отметить, что у старших есть явно больше возможностей и соблазна в этом отношении (они опубликовали больше), а допустимость самоплагиата до сих пор является дискуссионным вопросом в академической среде и далеко не всегда он однозначно считается нарушением академических норм.

Другие работы исследовали влияние факторов странового и организационного уровня. Главный из таких факторов, отраженных в литературе, – культура «Publish or perish», наличием которой различаются как страны, так и организации. Существует широко распространенное мнение, что требование публиковаться в журналах высокого уровня может вызывать проявления академической нечестности (прежде всего в форме фабрикации результатов), особенно если такие публикации вознаграждаются финансово [Qui 2010]. Еще один фактор, связь которого с академической нечестностью [Louis et al. 1995; Anderson 1996] и попыткой обмануть показатели [Anderson et al. 2007] эмпирически доказана, – соревновательная среда. Тем не менее, некоторые исследования показали, что публикационное давление имеет положительную связь с академической нечестностью только в случае, если таковое сопровождается денежным вознаграждением за публикации на штучной основе, в то время как остальные меры политики стимулирования публикационной активности не повышают, а возможно, даже понижают ее проявления [Fanelli et al. 2018; Fanelli et al. 2015].

Страны также различаются исторически сложившейся академической культурой, более или менее способствующей взаимному контролю [Anderson et al. 2007]. Если коллеги не критикуют и не проверяют работу друг друга, то шансы публикации подлога возрастают [Wright et al. 2008]. Ли и Шранк показали, что академическая нечестность более распространена в развивающихся странах, принявших иерархическую кафедральную немецкую модель в противовес более коллегиальным системам англо-американских департаментов [Lee, Schrank 2010]. Эти предположения также были подтверждены в исследовании Фанелли и др. [Fanelli et al. 2015; Fanelli et al. 2018].

На уровне организаций значение также может иметь институциональная среда. Исследователи отмечают, что «политика в сфере академической нечестности может играть важнейшую роль в предупреждении случаев нечестности», и именно потому различные организации разрабатывают собственные этические кодексы [Resnik et al. 2015]. Действительно, наличие у университета соответствующих нормативных документов делает нарушения академической этики студентами [Zimmerman 1998; McCabe, Trevino 1993] и учеными [Fanelli et al. 2015] менее вероятным. Однако непонятно, какова в таких случаях причинно-следственная связь: возможно, кодекс – не причина интернализации норм, а следствие озабоченности вопросами академической этики и чувствительности к ее нарушениям [Fanelli et al. 2018].

Наше исследование затрагивает единственную страну, что исключает кросс-культурные сравнения. Кроме того, поскольку оно опирается на документарные данные, мы располагали лишь ограниченным набором характеризующих индивида переменных (мы могли идентифицировать пол по имени и отчеству, но не могли идентифицировать возраст и тем более социальный статус). Мы могли опираться, однако, на значительный набор переменных, характеризующих научные достижения и институциональную среду индивида.

3 Данные и методы

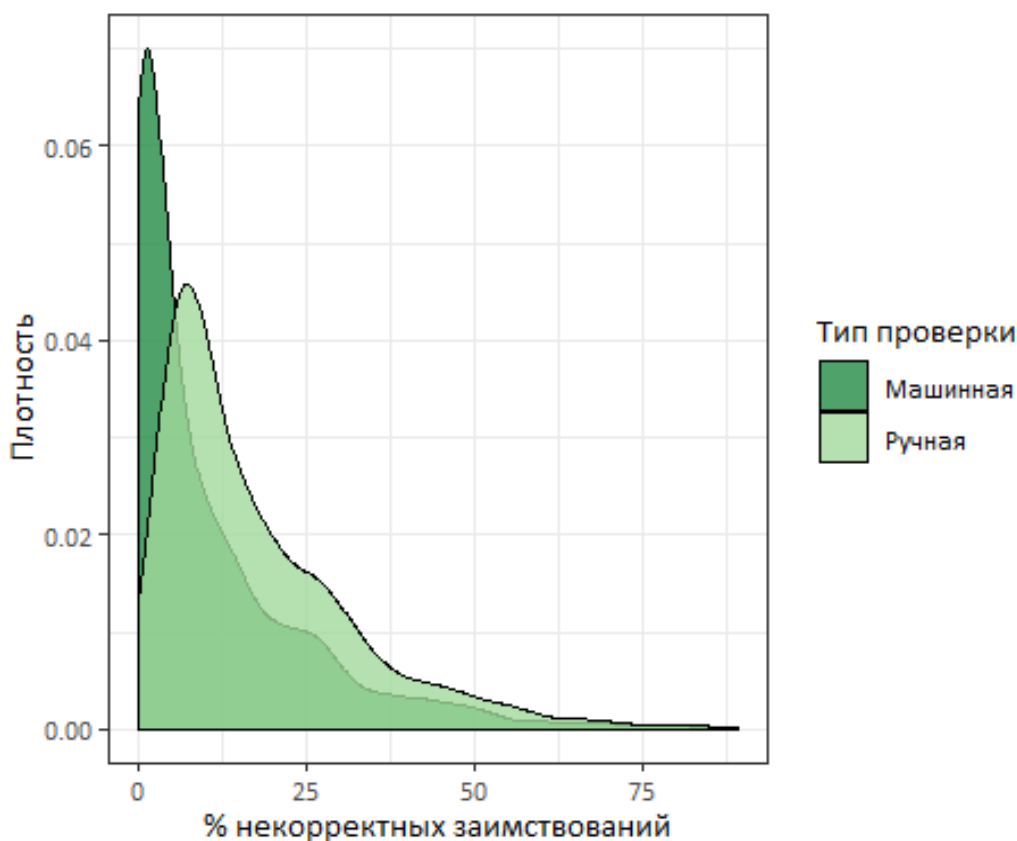
Исследование строится на проверке случайной выборки диссертаций программой «Антиплагиат»⁹. Была собрана случайная выборка 2600 докторских диссертаций за период 2006-2015 годы по всем специальностям. Выборка формировалась на основе списка защищенных диссертаций с сайта ВАК – с 2006 года всем защитам докторских диссертаций обязательно должны были предшествовать объявления на сайте. Однако, поскольку авторефераты начали регулярно размещаться только в конце 2006 года, он оказался покрыт лишь в незначительной степени. Мы также решили поставить верхнюю границу на начале 2016 года, предполагая, что в дальнейшем доля заимствований резко снизится вследствие распространения «Антиплагиата» и нам просто нечего будет анализировать. Как читатель увидит дальше, это предположение могло быть ошибочным. Всего за этот период было защищено 27735 докторских диссертаций (исключая, вероятно, небольшое число диссертаций по закрытой тематике, которые миновали публикацию в интернете); наша выборка, таким образом, составила 8,8% от генеральной совокупности. Затем диссертации были проверены в системе «Антиплагиат». С помощью алгоритма отсекались тексты, выпущенные позже того года, в котором была написана диссертация, а также тексты, среди авторов которых встречается данный диссертант. Некоторые диссертации системе не удалось загрузить, корректно были распознаны 2468, которые и использовались в основных подсчетах. Мы сравнивали тексты диссертаций с двумя модулями «Антиплагиата» – Российская научная библиотека и модуль eLIBRARY.RU – поскольку они содержат большую часть релевантных академических текстов и тексты в них имеют однозначную датировку, что позволяет установить направление заимствования. Первый дает возможность учесть заимствования в основном из диссертаций, второй – в основном из научных статей.

Для оценки надежности «Антиплагиата» мы сравнили результаты автоматической проверки с ручной проверкой для 860 диссертаций. Ручная проверка заключалась в просмотре всех источников, опубликованных позднее года защиты диссертации, и всех источников, авторы ко-

9 «Антиплагиат» – система для поиска текстовых заимствований. Сначала текст извлекается из загруженного документа, затем ищутся совпадения в границах подключенных модулей. Итогом становится отчет о проверке документа, в котором указывается процент заимствования, процент оригинальности и процент процитированных источников. В сыром виде отчет не идентифицирует направление заимствования: если мы проверяем диссертацию 2000 года, а в источнике за 2001 встречается фрагмент из нее, то «Антиплагиат» покажет совпадение. Другая важная проблема – это заимствование текста из собственной работы. Если автор вставил в диссертацию фрагменты текста из собственных ранее опубликованных статей, то программа также покажет совпадающие фрагменты. И в первом и во втором случае мы не можем сказать, что автор ответственен за некорректные заимствования, так как в одном случае текст взяли у него, а в другом случае он вставил свой собственный текст. Мы решали эти проблемы следующим образом: в программу «Антиплагиат» были включены специально разработанные алгоритмы, которые считали заимствованием только те совпадения, которые были сделаны у автора с другой фамилией и если источники были опубликованы ранее года защиты диссертации.

торых носили ту же фамилию. Кроме того, просматривался сам источник с тем, чтобы исключить мелкие совпадения в виде титульного листа или списка литературы. Тем самым в целом мы повторяли действия алгоритма «Антиплагиата». При ручной проверке средний процент некорректных заимствований на 7% ниже, чем при автоматической проверке (медианный – на 7,5%). При этом только четверть работ имеет разность в результатах проверок более 8 процентных пунктов. Распределение машинной и ручной проверки смотрите на Рис. 1. Надо отметить, однако, что ручная проверка была заведомо консервативной (мы не отслеживали мелкие заимствования размером 0,1% текста – порядка пяти-семи предложений, – которые иногда составляли приличную часть заимствований), и истинный процент заимствований находится между ними¹⁰. В дальнейшем анализе в качестве независимой переменной использовались данные машинной проверки. Расчеты на диссертациях, проверенных вручную, дают сходные коэффициенты (доступно по требованию).

Рисунок 1. Распределение процента некорректных заимствований в зависимости от типа проверки



Кроме того, мы использовали данные интервью с председателями, учеными секретарями и членами диссертационных советов в разных организациях. В интервью мы фокусировались

¹⁰ Более детальный анализ, проведенный на подвыборке диссертаций с заметным расхождением в проценте найденных заимствований, показал, что автоматическая идентификация заимствования допускала ошибки в части установления тождественности автора и неверного распознавания годов. Также в текстах могут не распознаваться легитимные цитаты (частая проблема для юридических наук). Совпадения также могут включать библиографические описания или приложения. Однако мы нашли всего два случая, в которых машинная проверка нашла значительные заимствования (свыше 60%), а ручная нашла их в пределах 10%. Эти случаи были удалены из дальнейших подсчетов.

на эффектах государственного регулирования присуждения ученых степеней, пытаюсь понять, что изменилось в поведении ученых и в науке под воздействием борьбы за повышение качества диссертаций. Мы также интересовались способами использования «Антиплагиата» и пониманием того, что составляет «некорректное заимствование». Всего было собрано 20 интервью с информантами из 15 разных советов, в том числе советов при региональных университетах.

4 Результаты

4.1 Общее распределение

Рисунок 1 дает общее представление о том, как распределены доли некорректных заимствований. Про значительное большинство диссертаций мы можем сказать, что в них присутствует нечто, что может квалифицироваться как нарушение нормы текстуальной оригинальности. Таким образом, мы можем констатировать, что заимствования частей чужого текста является широко распространенной, едва ли не универсальной, практикой. С другой стороны, ее радикальные нарушения встречаются также сравнительно редко. Заимствования свыше 40% встречаются примерно в 10% диссертаций, а заимствования свыше 50% мы обнаружили в 151 случае (6,1%), из них в двух доля была завышена из-за неверной автоматической идентификации источника. В этом смысле один из выводов нашего исследования может состоять в том, что норма текстуальной аутентичности, видимо, повсеместно нарушается. Однако это не значит, что нарушаются все нормы, регулирующие оригинальность вклада автора; фактически тотальный плагиат сравнительно редко проходит через жернова советов. Вообще говоря, эта картина сама по себе лучше вписывается в теорию альтернативной конвенции – нормы действуют, но не совсем те, которые предписывает ВАК и иные контролирующие органы. Если бы консенсус по поводу того, что текстуальные заимствования неприемлемы, действительно существовал, мы могли бы ожидать встретить бимодальное распределение с одним пиком около нуля (честные ученые) и вторым ближе к 100% заимствования (плагиаторы). Действительно кража 10% текста, когда можешь украсть 100%, сродни ограблению, в ходе которого грабитель забирает у жертвы 1/10 денег из кошелька: учитывая, что квалификация преступления от этого не изменится и на наказание это слабо повлияет, рациональнее брать все, если уж вообще взялся грабить.

4.2 Deskриптивная статистика и простые статистические тесты

На первой фазе нашего анализа мы попытались связать количество некорректных заимствований со следующими характеристиками защиты и диссертанта: год защиты¹¹, федеральный округ и организация (институт РАН, ведущий университет¹², другой университет, прочие организации), где проходила защита, специальность и пол диссертанта. Тип организации может считаться переменной, характеризующей академическую силу среды, из которой происходит претендент на степень (Гипотеза З), а также глобализированность этой среды (Гипотеза

11 Для года защиты характеристики распределения не приводятся для 2006 года, который был исключен из последующего анализа по годам в связи с малым числом наблюдений – 7 штук.

12 В эту группу вошли все университеты программы «5-100», МГУ и СПбГУ.

5). Дисциплины, в соответствии с широко распространенным предположением, различаются уровнем интернализации норм научного этики их представителями (Гипотеза 2). Пол также характеризует вероятность девиации – рассмотренная выше литература предполагает, что мужчины в целом более склонны к нарушению социальных норм в достижительных целях (Гипотеза 1). Федеральный округ может характеризовать влияние локальной академической культуры и глобализованность норм (Центральный и Северо-Западный округа в которых свыше 90% защит приходится на Москву и Петербург, предположительно более глобализованные, чем остальные – Гипотеза 5). Наконец, год характеризует эффекты контроля: распространение «Антиплагиата» и активность «Диссернета», по идее, должны были оказать сдерживающий эффект на активность рациональных злоумышленников (Гипотеза 4).

Количество наблюдений в каждой группе, средние и медианные значения, а также характеристики распределения числа некорректных заимствований представлены в Таблице 1.

Таблица 1. Описательные статистики категориальных переменных анализа

	Среднее	Медиана	Минимум	Максимум	Ср. откл.	Асимм-я	N
Специальность							
Экономические	22,9	18,3	2,3	85,8	16,7	1,3	433
Технические	19,7	14,3	0,7	74,2	17,2	1,1	363
Медицинские	19,1	15,1	2	88,8	13,9	1,6	330
Педагогические	20,3	15,6	2,5	89,2	16	1,5	202
Физико-математические	12,6	6,2	0,5	94,4	15,3	2,3	166
Исторические	14,9	10,3	1,4	91,8	14,9	3,1	129
Биологические	19,1	11,9	1,9	80,8	16,5	1,3	121
Филологические	9,2	6,9	0,8	61	8,5	3,5	117
Юридические	28,1	25,6	4,5	67,6	13,3	0,6	106
Сельск. и ветеринарные	31,3	29,8	3	84,5	18,7	0,7	99
Философские	12,1	8,4	1,9	47,9	10,1	1,7	91
Химические	26,2	20,8	2,5	81,1	19	0,9	66
Геологические	11,5	8,7	0,9	57,9	11,1	2,6	57
Психологические	17	12,5	3,9	62,9	14,1	1,6	53
Социологические	14,5	12,3	2,2	41,5	9,7	1,2	45
Политические	16,1	11,9	3,1	45,5	11	1,1	40
Искусствоведческие	4,6	4,1	1,4	9,9	2,3	0,8	20
Культурологические	9,6	7	2,9	32,9	7,7	1,7	16
Тип организации							
Университет	20,6	16,2	0,7	94,4	16,4	1,3	1466
Ведущий университет	14,6	10,1	0,6	66	13	1,6	342
Другие организации	20,3	14,8	0,8	89,2	17,1	1,5	334
Институты РАН	15,5	10,1	0,5	89,1	15	2	311
Федеральный округ							
Центральный	18,6	13,3	0,5	91,8	15,9	1,5	1170

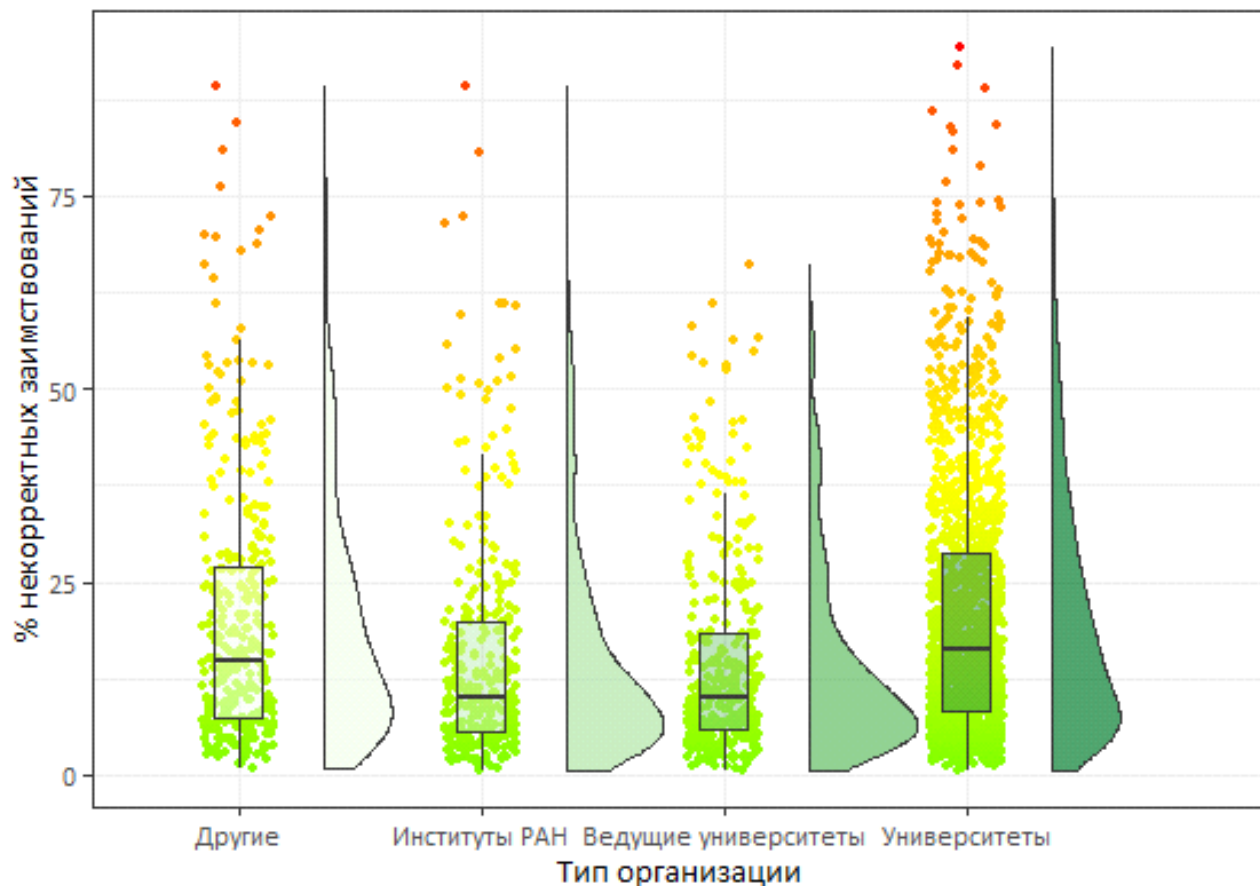
Приволжский	22,6	17,3	1,3	94,4	18,8	1,2	322
Северо-Западный	16,3	12,2	1,8	73,6	13,6	1,6	313
Сибирский	18,7	13,2	0,9	89,1	15,5	1,4	266
Южный	22	18,9	2,1	80,8	16,5	1,2	177
Уральский	16,8	11,9	1,9	65,3	14,2	1,4	112
Северо-Кавказский	23,7	18,4	0,8	73,7	17,6	0,9	59
Дальневосточный	15,5	12	2	54,3	12,2	1,3	34
Год защиты							
2007	19,6	16,9	1,4	89,1	15,9	1,8	68
2008	19,4	13,9	2	74	15,8	1,2	329
2009	20	13,9	0,5	88,8	17,2	1,4	534
2010	19,7	14,5	0,6	81,1	16,6	1,2	314
2011	19,7	13,9	0,8	84,5	17	1,3	371
2012	19,7	15,7	1,3	89,2	15,4	1,6	254
2013	19,4	15,6	2	94,4	15,5	1,5	184
2014	17,3	13,3	1,6	80,7	13,8	1,9	252
2015	13,7	9,6	0,9	91,8	12,5	3	124
Пол диссертанта							
Мужчины	19,2	14,1	0,5	91,8	16,8	1,4	1358
Женщины	18,9	13,8	0,7	94,4	15,1	1,5	1095

Видно, что распределения количества некорректных заимствований не подчиняются нормальному закону распределения, медианы и средние не совпадают, асимметрия зачастую превышает пороговые значения, характерные для данных из нормального распределения. В связи с неприменимостью для данного случая параметрических методов, мы использовали тест Краскела-Уоллиса с поправкой Бонферрони, являющийся консервативной оценкой. Так как одним из требований для критерия Краскелла-Уоллиса является однородность распределений, для всех категориальных переменных был проведен тест Флингера-Килеена, во всех случаях показавший наличие неоднородности. При интерпретации результатов теста Краскелла-Уоллиса стоит учитывать нарушение этого требования. Мы предлагаем ориентироваться при интерпретации на наиболее выдающиеся и значимые различия между группами. Результаты теста представлены в Таблицах 1-4 в Приложении.

Нам удалось обнаружить значимые различия в количестве некорректных заимствований для организаций (Рис. 2), федеральных округов (Рис. 3) и специальностей (Рис. 5), частично показана разница по годам (Рис. 4). Между мужчинами и женщинами статистически значимых различий по количеству некорректных заимствований обнаружено не было.

Так, рядовые университеты и другие организации имеют медианное значение количества некорректных заимствований 16,2% и 14,8% соответственно и статистически значимо между собой не различаются. В то же время медианное значение для институтов РАН и ведущих университетов совпадает и составляет 10,1%. Эти институции отличаются от рядовых университетов и других организаций на уровне значимости $<0,001$, не различаясь между собой.

Рисунок 2. Распределение % некорректных заимствований по типам организации



Если говорить о федеральных округах, здесь картина сложнее. По большей части, статистически значимых различий между ними нет. Исключение – значимая разница между Центральным округом против Приволжского и Южного (в последних медианное количество некорректных заимствований больше на 4% и 5,6% соответственно, значимо на уровне $<0,05$). Мы также нашли разницу для Северо-Западного против Северо-Кавказского (в последнем количество некорректных заимствований выше на 6,2%, значимо на уровне $<0,05$), Приволжского и Южного (в обоих количество некорректных заимствований выше на 5,1% и 6,7% соответственно, значимо на уровне $<0,01$). В то время как явных лидеров по количеству некорректных заимствований мы выделить не можем, стоит отметить, что Южный, Северо-Кавказский и Приволжский округа демонстрируют высокие медианные значения по сравнению с остальными округами.

Рисунок 3. Распределение процента некорректных заимствований в зависимости от федерального округа, в котором проходила защита



По годам некорректные заимствования практически не различаются, и хотя может показаться, что произошло существенное снижение к 2015 году, статистически значимые различия обнаруживаются только между, с одной стороны, 2015 и, с другой – 2008, 2009, 2012 и 2013 годами, причем разница между 2015 и 2008–2012 меньше, чем между 2013 и 2015. Мы не видим никакой однозначной тенденции к снижению доли заимствований, которую можно было бы объяснить появлением «Антиплагиата» (получившего широкое распространение уже к 2010 году) или деятельностью «Диссернета», начавшейся в 2013 году.

Рисунок 4. Распределение процента некорректных заимствований в зависимости от года защиты

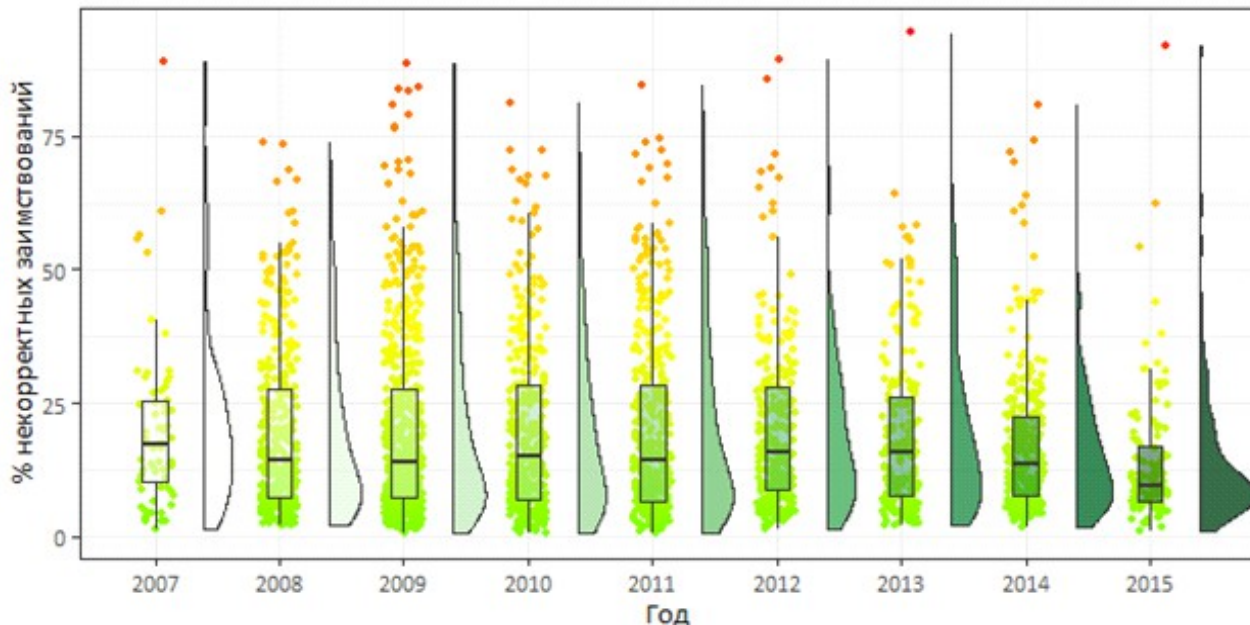
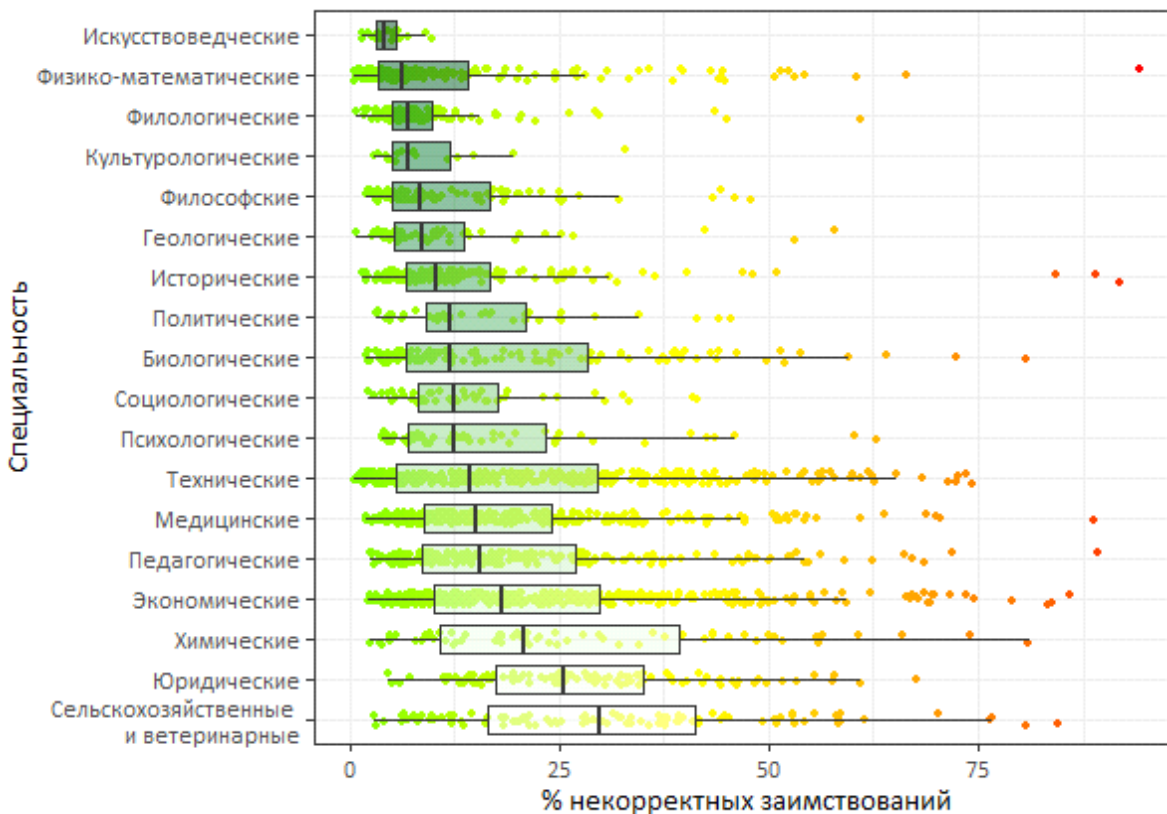


Рисунок 5. Распределение процента некорректных заимствований в зависимости от специальности, по которой проходила защита



Нам удалось обнаружить определенное количество специальностей, статистически значимо отличающихся от других по количеству некорректных заимствований. Наименьшее количество некорректных заимствований мы обнаруживаем в искусствоведческих, физико-математических, филологических и культурологических науках. Тем не менее, нужно отметить, что

единичные диссертации с высоким уровнем некорректных заимствований встречаются почти во всех специальностях, даже в тех, в которых общий уровень низок (физико-математические, филологические, философские, и т.д.). Не нашлось их только в культурологии и искусствоведении, но это может объясняться тем, что в нашей выборке их число не велико – 16 и 20 диссертаций соответственно.

Наибольшее количество некорректных заимствований обнаруживается в сельскохозяйственных науках – оно значимо выше по сравнению со всеми специальностями, за исключением юридических наук. Не отстают и химики, значимо отличаясь от физико-математических, исторических, философских, филологических, искусствоведческих, культурологических и геологических наук. Юристы также существенно отличаются ото всех, кроме химиков, физиков и математиков. Однако необходимо отметить, что юридические науки – это особый случай: «Антиплагиат» регулярно распознает выдержки из нормативно-правовых актов как некорректные заимствования, на что часто жаловались наши собеседники в интервью.

4.3 Регрессионный анализ

На следующем шаге мы включили рассмотренные выше переменные в регрессионную модель для предсказания процента некорректных заимствований. Регрессия позволяет оценить влияние отдельных переменных при контроле по всем остальным – важная предосторожность в случае, если между ними может существовать ложная корреляция (скажем, доля защит, приходящихся на разные дисциплины, изменялась по годам, причем число экономических защит после 2013 сокращалось значительно сильнее, чем естественнонаучных; таким образом, эффект года может быть связан с этим фактором, а не с тем, что во всех дисциплинах стали меньше заимствовать). Т. к. наши данные распределены ненормально, бета-регрессия является одним из возможных подходов. Альтернативой может служить использование линейной регрессии, но с преобразованием данных и взвешиванием, уменьшающим влияние наблюдений в «хвосте» распределения. В таблице 2 представлены предельные эффекты, рассчитанные с помощью бета-регрессии (она показала лучшие характеристики по сравнению с линейной регрессией по предсказанным значениям и при анализе остатков).

Таблица 2. Бета-регрессия: предельные эффекты¹³

Предикторы	% некорректных заимствований	
	DF/dx	std. error
Мужчины	-0,71	0,56
Другие организации	0,51	0,85
Институты РАН	-1,66	1,09
Ведущие университеты	-2,18*	1,02
2008	-1,63	1,63
2009	-1,57	1,6
2010	-1,02	1,67
2011	-0,98	1,65

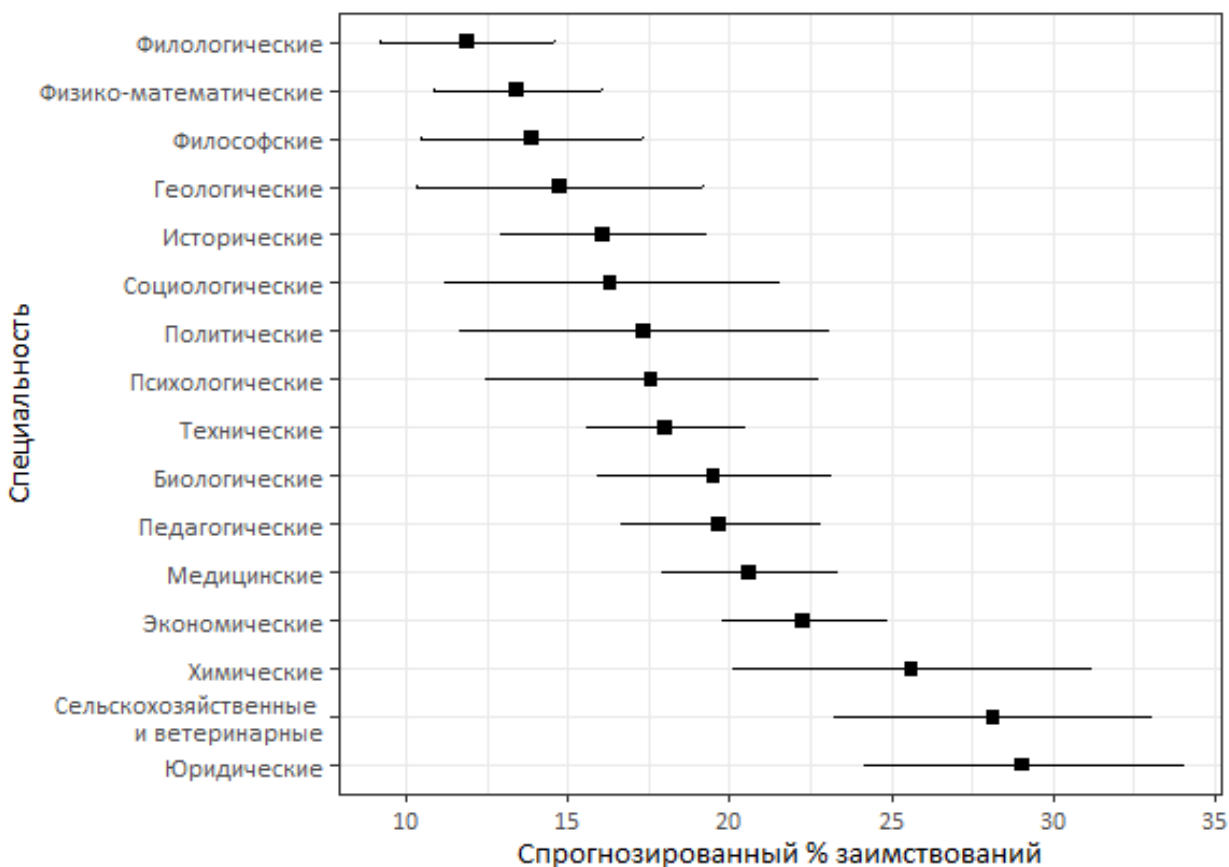
13 Референтные категории для пола – женский, для типа организаций – рядовой университет, для года – 2007, для округов – Центральный, для дисциплин – физико-математические науки.

2012	-0,74	1,72
2013	-1,24	1,74
2014	-2,44	1,63
2015	-5,58^{***}	1,55
Дальневосточный	-1,3	2,17
Приволжский	2,77^{**}	0,87
Северо-Кавказский	3,3	1,86
Сибирский	1,45	0,93
Северо-Западный	-0,98	0,82
Уральский	-0,85	1,27
Южный	2,37[*]	1,12
Химические	15,39^{***}	2,62
Биологические	7,88^{***}	2,02
Технические	5,89^{***}	1,53
Сельскохозяйственные и ветеринарные	18,23^{***}	2,49
Исторические	3,51	1,8
Экономические	10,8^{***}	1,6
Философские	0,61	1,86
Филологические	-2,13	1,61
Юридические	19,24^{***}	2,4
Педагогические	8,03^{***}	1,83
Медицинские	9,01^{***}	1,72
Психологические	5,46[*]	2,55
Социологические	3,84	2,57
Политические	5,17	2,74
Геологические	1,75	2,24
N		2393
Псевдо R ²		0,1595
* p<0,05 ** p<0,01 *** p<0,001		

Согласно представленной модели в ведущих университетах на 2,18% меньше некорректных заимствований по сравнению с обычными университетами при усреднении остальных характеристик. Если говорить о годах как группах, то мы видим отличие 2015 года от 2007 на 5,58%. Тем не менее, различия в других годах нам выявить не удалось. Для федеральных округов, в сравнении с Центральным, значимо отличаются Приволжский (2,77%) и Южный (2,37%). Если говорить о специальностях, то значимо большее количество некорректных заимствований по сравнению с физико-математическими науками демонстрируют юридические (19,24%), сельскохозяйственные и ветеринарные (18,23%), химические (15,39%), экономические (10,8%), медицинские (9,01%), педагогические (8,03%), биологические (7,88%), технические (5,89%) и психологические науки (5,46%).

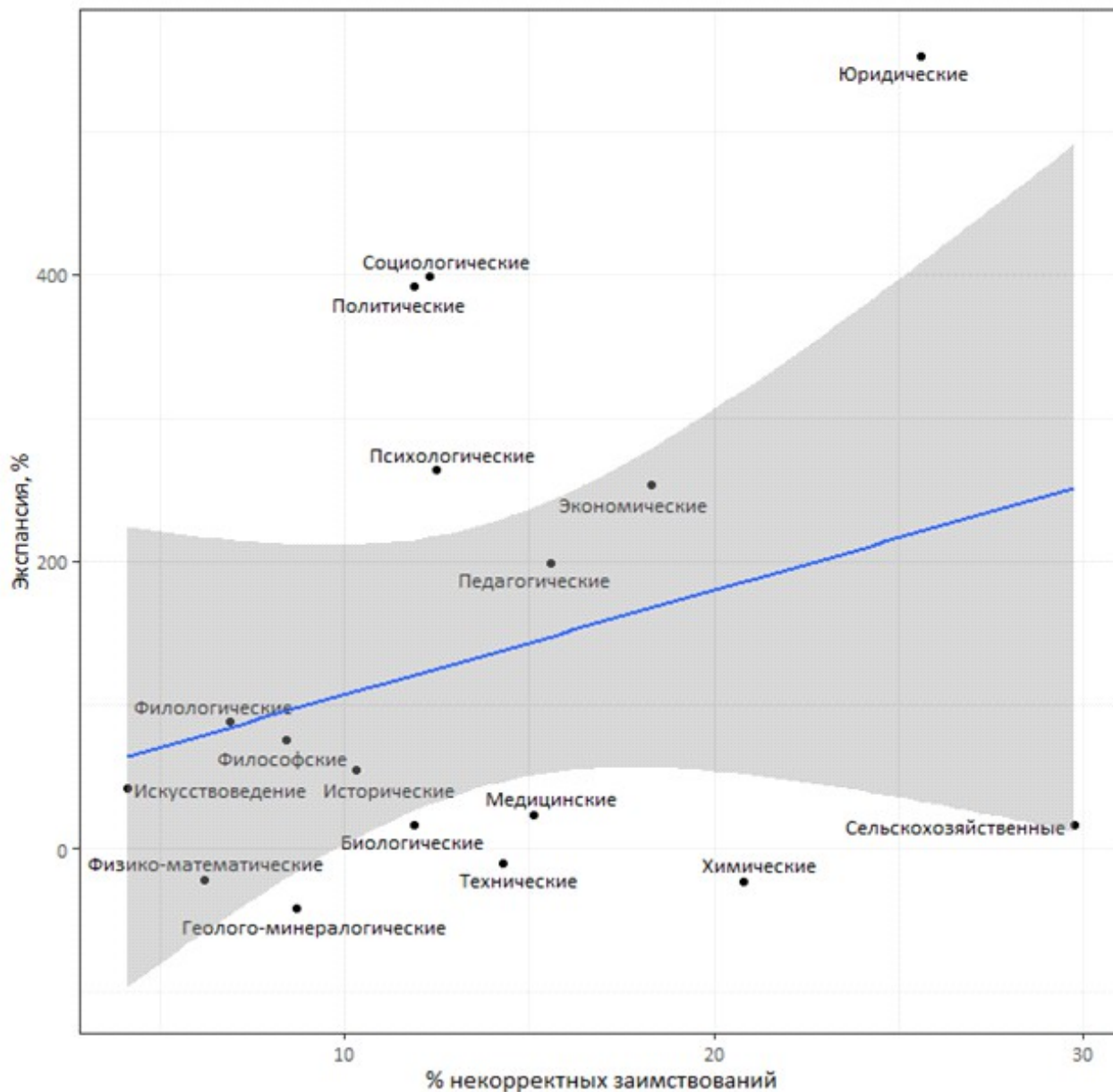
На Рисунке 6 представлены предсказанные значения % заимствования по дисциплинам, усредненных по остальным характеристикам.

Рисунок 6. Спрогнозированный процент некорректных заимствований в зависимости от специальности, по которой проходила защита



Возвращаясь к нашим исходным теоретическим соображениям, мы видим, что идея альтернативных конвенций получает наибольшую поддержку. Существует значимая разница между более глобализованными институциями и остальными и более глобализованными столичными регионами и Южным, Северо-Кавказским и Приволжским округами. При этом нет различий между мужчинами и женщинами (которые предполагаются нормативно-социализационными теориями), а эффект года (предполагаемый рационально-структурными гипотезами) слаб и прослеживается только в 2015 году. Наконец, различия между дисциплинами существуют, но совсем не те, которые можно ожидать, основываясь на Гипотезе 2. Доля некорректных заимствований слабо связана с уровнем экспансии. Иллюстрацией этого является Рисунок 7, отражающий связь между средней долей заимствований и уровнем экспансии дисциплины, рассчитанный как отношение числа защит в 1995 (минимальное число защит в постсоветский период) и 2006 (максимальное).

Рисунок 7. Процент некорректных заимствований (ось X) и экспансия дисциплины (ось Y), 1993-2006



Связь, если она и прослеживается, очень условна. Самые большие доли заимствований обнаруживаются в юриспруденции (но надо помнить о неточности оценок заимствований у юристов), сельскохозяйственных науках и химии. Из них юриспруденция выросла сильнее всего, а химия была одной из переживших наибольшее сжатие.

4.4 Влияние известности и научной репутации

Гипотеза 3 гласит, что соблазн заимствовать чужой текст сильнее для тех, кто не способен написать собственный в силу общей интеллектуальной слабости. Напротив, человеку, который пишет много и хорошо, может быть проще писать самому, чем использовать копипэйт. Кроме того, в духе Гипотезы 1, мы можем предполагать, что если заимствование воспринимается как разновидность девиации, то те, кто не усвоил нормы научного этоса, первыми от-

реагируют на возросшее публикационное давление попытками обмануть систему: будут печататься в «мусорных» журналах, станут цитировать собственные статьи и т.д. К счастью для нашего исследования, Российский индекс научного цитирования (РИНЦ) предоставляет широкий репертуар метрик, которые можно оценить как указание не только на количество, но и на качество цитирований. К качественным показателям мы отнесли публикации и цитирования в ядре РИНЦ (состоящем из журналов, за которые «проголосовали» в ходе опроса высокоцитируемые российские ученые), импакт-фактор журналов, где были опубликованы и процитированы статьи, долю зарубежных статей и цитирований в них, а также долю статей в журналах ВАК. К имитационным показателям мы отнесли долю публикаций и цитирований, а также долю Хирш-индекса в изданиях eLIBRARY.RU, не вошедших в РИНЦ (категория, в которую входят в том числе исключенные из РИНЦ «мусорные» издания и сборники), долю самоцитирования и долю цитирования соавторами. Имитационные показатели были, таким образом, связаны с практиками, к которым могли бы прибегнуть беспринципные ученые, чтобы быстро улучшить свой наукометрический профиль – начать публиковаться за деньги, цитировать самого себя или обмениваться ссылками с кем-то, с кем сам ученый тесно связан.

Из 2468 докторов 2028 имеют профиль в платформе eLIBRARY.RU, что составляет 82,1% от всей выборки. При этом статистически значимой связи между отсутствием профиля и специальностью, и годом защиты диссертации не было выявлено. Из всех имеющих профиль, самостоятельно зарегистрировались на сайте 82,79% докторов, профиль остальных автоматически сгенерирован системой. Для доли некорректных заимствований и показателей публикационной активности была рассчитана корреляция Спирмена (см. Таблицу 5 в Приложении). В целом сила связи между долей заимствования и показателями публикационной активности невелика. Для показателей публикаций и цитирований по ядру РИНЦ, так же как в переводных изданиях и доли статей в зарубежных журналах, высокие значения ассоциируются с меньшей долей некорректных заимствований в тексте диссертации, но даже для них абсолютная величина коэффициента корреляции не превышает 0,2. У нас нет, таким образом, оснований заявить, что заимствования являются маркером «слабой науки».

На наблюдаемую связь могли, однако, повлиять различия в публикационных и цитатных практиках дисциплин. Чтобы нивелировать их, мы создали две полярные выборки докторов. В выборку вошли 10% от представителей каждой дисциплины, у которых обнаружился минимум заимствований, и 10% тех, у кого обнаружился максимум (число представителей каждой дисциплины, таким образом, было строго пропорционально ее доле в общем числе защит). В выборках, таким образом, было равное количество экономистов, психологов, химиков и т.д. Мы обозначили эти группы «плагиат» и «не плагиат».

Далее мы сфокусировались на публикационных показателях, которые позволяют противопоставить наши группы и при этом могут быть интерпретированы как показатели преимущественно качественной и/или имитационной деятельности ученого. Чтобы определить, существует ли разница в публикационной активности или же текст докторской диссертации не является определяющим в выбранной стратегии публикаций, мы сравнили поляризованные выборки докторов по указанным выше показателям (см. Таблицу 5). Статистическая значимость различий была рассчитана с помощью перестановочного теста при случайном назначении в группы «плагиат» / «не плагиат».

Таблица 3. Модельные оценки

Переменная	Плагиат (среднее)	Не плагиат (среднее)	Плагиат (медиана)	Не плагиат (медиана)	Средний эффект воздействия (Δ)	Эффект воздействия по медиане (Δ)
Качественные						
Публикации ядро, %	19,96	24,99	11,21	14,19	-5,03*	-2,98
Цитирования ядро, %	17,79	24,29	9,09	9,68	-6,50**	-0,59
Импакт-фактор опубликованные	0,37	0,45	0,28	0,32	-0,07*	-0,04
Импакт-фактор процитированные	0,42	0,52	0,32	0,32	-0,10*	0
Статьи зарубежные, %	3,74	6,84	0	1,4	-3,10***	-1,4
Цитирования зарубежные, %	6,94	10,4	2,8	2,9	-3,45**	-0,1
Статьи ВАК, %	50,39	46,51	50	47,1	3,88	2,9
Имитационные						
Публикации eLIBRARY.RU, %	5,40	4,35	2,26	2,13	1,05	0,13
Цитирования eLIBRARY.RU, %	8,35	7,70	4,15	4,62	0,65	-0,47
Индекс Хирша eLIBRARY.RU, %	4,7	4,55	0	0	0,15	0
Самоцитирования, %	20,94	28,13	15,6	20,8	-7,19***	-5,2*
Цитирования соавторами, %	36,7	40	36,7	40	-3,3	-3,3
Пояснение: *p<0,05, **p<0,01, ***p<0,001						

Статистически значимыми оказались различия между поляризованными выборками докторов по всем качественным показателям, кроме доли статей в журналах ВАК, при этом статистически значимой разницы по имитационным показателям, за исключением большего самоцитирования в группе «не плагиат», обнаружено не было. Отсутствие различий в числе публикаций в журналах из списка ВАК, разумеется, может быть поводом усомниться в их качестве. Определенное количество статей в таких изданиях является обязательным для защиты докторской диссертации, и в этом смысле разница между более сильными и более слабыми учеными в нашей выборке может быть размыта. Группа имитационных показателей (за исключением большего самоцитирования в группе «не плагиат») в целом не показывает различий между докторами с высокой и низкой долей плагиата в диссертации. То, что более добросовестные ученые чаще цитируют сами себя, на первый взгляд выглядит парадоксом, но на второй может показаться вполне логичным: возможно, что они в принципе относятся к цитированию, включая цитирование собственных работ, аккуратнее. Это отража-

ется и на числе ссылок на себя, и на доле некорректных заимствований¹⁴. При этом по медианным показателям значимо различие только по показателю самоцитирования. Это, скорее всего, свидетельствует о факте, что среди добросовестно написавших диссертацию чаще встречаются ученые с исключительными высокими качественными показателями, но наличие плагиата в диссертации не влияет значимо на «среднего» ученого. Т.е. среди группы не плагиат чаще встречаются «звезды» науки по публикационным показателям.

В целом эти результаты можно считать опровержением Гипотезы 1 и частичным подтверждением Гипотезы 3. До некоторой степени, отсутствие заимствований объективно является сигналом, указывающим на способности индивида. В этом смысле норма текстуальной аутентичности может быть рационально обоснована со ссылкой на этот эффект – требование писать самостоятельно может способствовать отбору индивидов, которые затем будут публиковаться в лучших журналах. Эффекты, предсказанные Гипотезой 3, однако, оказываются очень слабыми. Контраст между полярными группами, заимствующими больше и меньше всего, достигает уровня значимости в 0,001 по единственному показателю – числу публикаций в иностранных журналах (по средним). Более того, те же наблюдения можно объяснить и с помощью альтернативных конвенций: ученые, знакомые с международными публикационными нормами с их более строгими правилами в отношении заимствований, следуют им и когда публикуются на русском.

4.5 Существуют ли различия в практиках заимствования между дисциплинами?

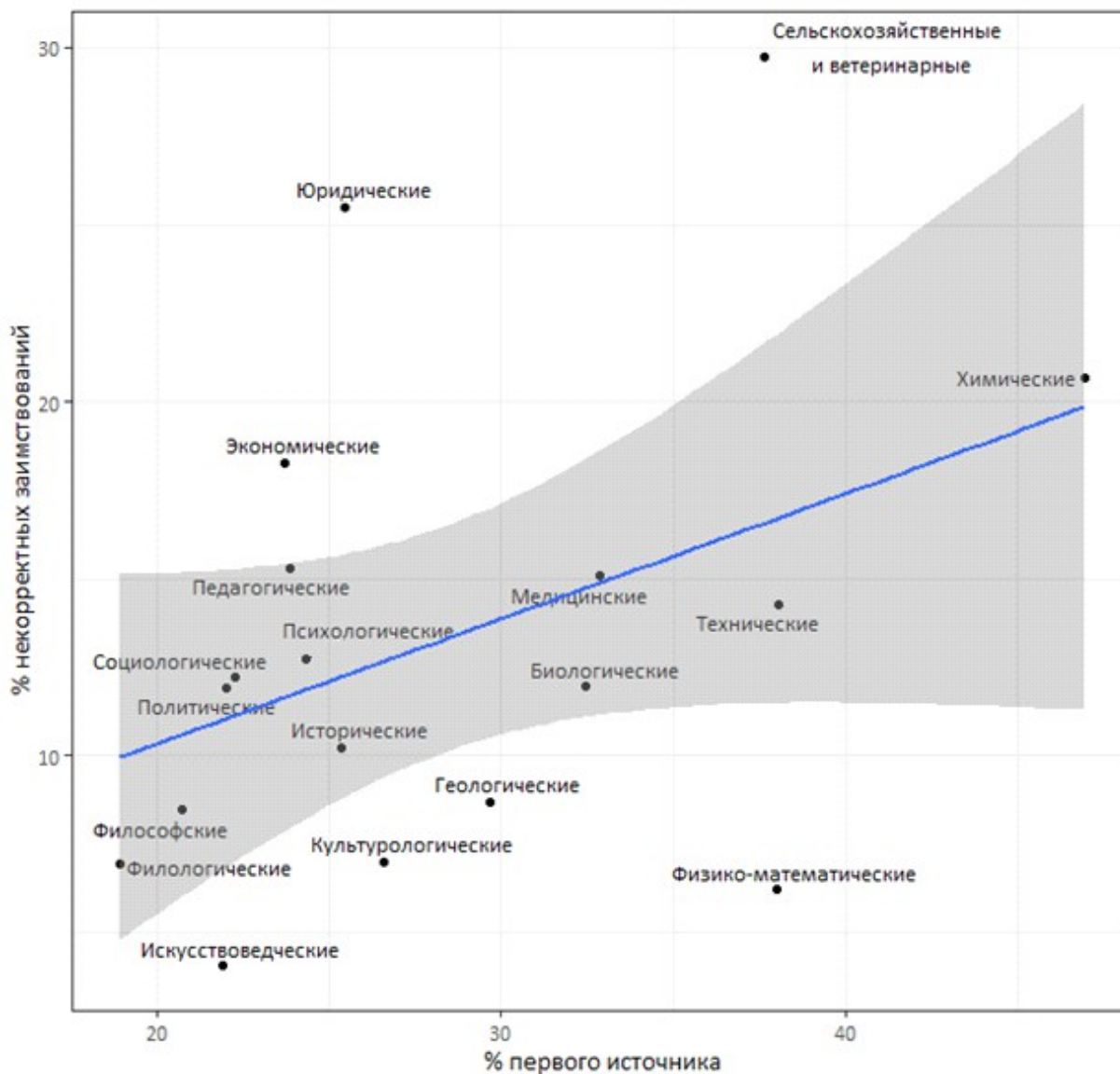
В качестве последнего шага мы попробовали понять, используя наличные данные, существуют ли принципиальные различия в том, что заимствуется. В частных дискуссиях по поводу полученных результатов критики утверждали, что сам характер некорректных заимствований может отличаться – естественники в основном заимствуют целиком описания методов или процедур коллективных исследований или вставляют в свою работу тексты аспирантов, которыми они руководили¹⁵. Мы попробовали проверить это предположение, взяв сочетание двух простейших показателей, предлагаемых «Антиплагиатом», – общей доли выявлен-

14 Мы использовали в анализе еще один тип метрик – длину текста, измеренную в символах, словах и предложениях. Между ними и долей некорректных заимствований наблюдается невысокая, но устойчивая отрицательная корреляция: чем больше в работе символов, слов и предложений, тем меньше в ней заимствований. Корреляция между плагиатом и количеством слов составляет -0,101, с количеством предложений -0,107. Объемные диссертации в целом имеют существенно меньший процент плагиата, чем короткие. Эффект длины остается значимым и при контроле по дисциплине, это значит, что готовность много писать сокращает вероятность использования других текстов. В этом смысле плагиат может быть сигналом, свидетельствующим о недостатках навыков письма.

15 Мы благодарны Михаилу Гельфанду и Андрею Ростовцеву за эти соображения. Эти же рассуждения дублировались в интервью: *Тут вопрос довольно скользкий, я с вами соглашусь в том, что не надо уж совсем копипастить, потому что я тоже, например... нормальная ситуация: у меня есть аспирант, она работает, она получает какие-то результаты, мы публикуем совместную статью. Сначала на ней защищается она, потом на ней защищаюсь я, она как человек, который это делал под моим руководством, потом я, потому что, извините, доктор наук – это не человек, который умеет читать интегралы, – это я доказал, когда защищал кандидатскую, диплом, – а человек, который умеет ставить задачи и может предложить новые научные направления и показать какие-то результаты. Никто же не ждет от доктора наук, что он сам будет считать все интегралы, но я с вами согласен, что действительно даже в такой ситуации копипаста – это, наверное, уже перебор. Надо все это творчески менять, я понимаю, что человек уже сделал, но хаметь тоже не нужно, хотя бывает, конечно (м., доктор ф-м. наук).*

ных заимствований и доли первого источника в этих заимствованиях, которые вместе должны были дать нам представление о том, имеем мы дело с компилятивной работой (заимствовал по строчке из ряда источников) или с заимствованием целых фрагментов. Рисунок 8 отражает средние показатели для разных дисциплин. Мы действительно наблюдаем здесь четкое различие между дисциплинами по тому, какая доля общего заимствования приходится на первый источник. Естественные науки (справа) более склонны заимствовать крупными кусками; социальные и гуманитарные – компилировать из большого числа текстов.

Рисунок 8. Процент первого источника в доле некорректных заимствований (ось X) и процент некорректных заимствований (ось Y)



Выводы

Подводя итоги нашего исследования, мы можем сделать следующие выводы.

Во-первых, нормы, требующие собственноручного написания всего текста диссертации как условия претензий на оригинальность, видимо, не имеют широкого распространения в российском академическом мире. Определенный объем заимствований (в пределах 15-20%), широко распространен, а заимствования порядка 5-7% встречаются в большинстве диссертаций¹⁶. Однако заимствования сверх этой доли текста встречаются редко; диссертации с ними, видимо, не доходят до защиты.

В целом эти наблюдения лучше укладываются в представление об альтернативных нормах, определяющих заимствования, не превышающие определенного предела, как приемлемые, а превышающие его – как девиацию¹⁷. Диссертации, грубо нарушающие эту норму, встречаются, однако составляют незначительное меньшинство в общем объеме. Возможно, что в отношении этого небольшого числа заимствованных диссертаций действуют иные закономерности, отличающиеся от «нормального» плагиата, и этот факт объясняет несовпадение наших выводов с выводами «Диссернета» (например, в отношении распределения по дисциплинам). Генерализируя наши результаты, мы можем утверждать, однако, что совокупность заимствующих значительную часть (свыше половины) текста диссертаций сравнительно невелика и возможно даже, что большинство подобных текстов уже обнаружено активистами.

Во-вторых, нормы, допускающие текстуальные заимствования, не изменились в результате внедрения «Антиплагиата». Это говорит против предположения о том, что заимствование чужого текста есть предмет рационального выбора стратегии нарушения норм. Если бы нарушение было делом утилитарного расчета, появление инструмента идентификации нарушителя должно было бы сократить число соответствующих случаев. Очевидно, что этого не произошло. Более вероятно, что имеет место широкий консенсус о допустимом заимствовании (видимо, свой для каждой дисциплины).

В-третьих, объем заимствований отчетливо зависит от дисциплины, но, вопреки гипотезе о роли экспансии в разрушении норм научного этоса, предрасположенность к текстовым заимствованиям слабо коррелирует с тенденцией к расширению или сжатию дисциплины. Она также слабо коррелирует с положением дисциплины на шкале естественнонаучный – гуманитарный (hard-soft). Дисциплины, занимающие близкое положение в этом спектре (физ.-мат. науки и химия, экономика и социология) могут разительно отличаться друг от друга. Скорее, мы имеем дело со специфическими дисциплинарными традициями с их представле-

16 Исключением является филология, которая, возможно, демонстрирует нулевую толерантность к плагиату. Эта особенность филологии может быть объяснена в духе теории сигналов тем, что для филологов самостоятельное написание текстов является лучшим сигналом профессиональной пригодности, чем для других дисциплин, меньше зависящих от дискурсивных средств коммуникации. Случай филологии, однако, заслуживает самостоятельного изучения, поскольку иные факторы могут определять низкое распространение плагиата. Например, для филологов издержки собственноручного написания текста могут быть меньше, как и издержки, связанные с выявлением плагиата, поскольку филологи могут быть более чувствительными к внедрению иноязычного текста. В результате привлекательность стратегии заимствования чужого текста меньше по сравнению с другими дисциплинами.

17 Цифра в 20% неоригинальности по «Антиплагиату» неоднократно появлялась в интервью как тот объем, который диссовет считал безопасным. При этом, кажется, часто не проводилось различие между заимствованиями из собственных более ранних текстов и текстов других людей.

ниями о допустимом и нормальном, которые отчасти могут объясняться особенностями интеллектуального труда в соответствующей области знания (скажем, отношения филологии с языком).

В-четвертых, на установки по отношению к заимствованиям, вероятно, влиял тип и статус институции (ведущие университеты, РАН, обычные вузы), возможно, за счет контакта с «мировой культурой» с ее более жесткими стандартами. В целом, однако, связь является слабой. Также на установки влиял регион, с Северо-Западным и Центральным (в основном за счет Санкт-Петербурга и Москвы), демонстрирующими наименьший уровень заимствований, а Южный и Приволжский – наибольший. Предельные эффекты структурных переменных, однако, не превышают 5%.

В-пятых, мы нашли слабую, но значимую отрицательную связь уровня заимствований с некоторыми показателями академической силы (публикации в международных изданиях), которая сохраняется при контроле по типу институции. Эта связь может быть интерпретирована двояко: или в духе рационального выбора (те, кто может писать сам, не заимствует у других), или в духе диффузии норм (те, кто соприкасается с более строгими глобальными нормами, транслирует их дальше). Нет связи между долей заимствований и показателями имитативности, такими как массированные публикации в «плохих» журналах, что противоречит нормативно-социализационному объяснению.

В целом гипотезы, связанные с нормативно-социализационной и рационально-структурными теориями не получают поддержки в ходе анализа. Это может пониматься как аргумент в пользу теории альтернативных норм; было бы лучше, разумеется, если бы доводы в ее пользу не были преимущественно негативными (т.е. мы нашли бы что-то, что подтверждает именно ее, а не опровергает альтернативные построения), однако на данный момент, кажется, просто нет другого объяснения паттернам, которые мы наблюдаем. Отношение к заимствованиям частей чужого текста в российской науке, видимо, похоже на отношение к превышению скорости: большинство водителей считают, что есть допустимые превышения (в пределах 10-15 километров в час) и недопустимые (скажем, свыше 40 километров в час). Нормы, запрещающие плагиат, могут распространяться по каналам наиболее глобализированных учреждений (сильные университеты, институты РАН) в «мировых городах», которые транслируют их далее, однако этот процесс разворачивается лишь очень медленно.

Список литературы

Вольное сетевое сообщество Диссернет. (2018) Структура, объем и ключевые игроки диссеродельной индустрии в России [Электронный ресурс] Электронные текстовые данные. Режим доступа: https://www.dissernet.org/publications/struktura_disserodelnoy_industrii.htm

Шахрай С.М., Аристер Н.И., Тедеев А.А. (2015) О плагиате в диссертациях на соискание ученой степени. Москва: МИИ.

Anderson M.S., Ronning E.A., De Vries R., Martinson B.C. (2007) The perverse effects of competition on scientists' work and relationships // *Science and Engineering Ethics*, vol. 13, no 4, pp. 437–461.

Anderson M.S. (1996) Misconduct and departmental context: Evidence from the Acadia Institute's graduate education project // *Journal of Information Ethics*, vol. 5, no 1, pp. 15–33.

Antes A.L., Brown R.P., Murphy S.T., Waples E.P., Mumford M.D., Connelly S., Devenport L.D. (2007) Personality and ethical decision-making in research: The role of perceptions of self and others // *Journal of Empirical Research on Human Research Ethics*. vol. 2, no 4, pp. 15–34.

Becker G.S. (1968) Crime and punishment: An economic approach // *The economic dimensions of crime* / N. Fielding, A. Clarke, R. Witt (eds.) London: Palgrave Macmillan, pp. 13–58.

Becker H.S. (2008) *Outsiders*. NY: Simon and Schuster.

Biagioli M., Galison P. (2003) *Scientific authorship: Credit and intellectual property in science*. London and New York: Routledge.

Fanelli D., Costas R., Fang F.C., Casadevall A., Bik E.M. (2018) Why do scientists fabricate and falsify data? A matched-control analysis of papers containing problematic image duplications // *Science and Engineering Ethics*, pp. 1–19.

Fanelli D., Costas R., Larivière V. (2015) Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity // *PloS One*. vol. 10, no 6, e0127556.

Fang F.C., Bennett J.W., Casadevall A. (2013) Males are overrepresented among life science researchers committing scientific misconduct // *MBio*. vol. 4, no 1, e00640–12.

Fields M.U. (2004) The influence of psychological type, self-esteem, and gender on academic dishonesty of students in higher education // *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 64 (3-A), 737.

Gundersen D.E., Capozzoli E.A., Rajamma R.K. (2008) Learned ethical behavior: An academic perspective // *Journal of Education for Business*. vol. 83, no 6, pp. 315–324.

Honig B., Bedi A. (2012) The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management // *Academy of Management Learning and Education*. vol. 11, no 1, pp. 101–123.

- Horbach S.S., Halfman W.W. (2017) The extent and causes of academic text recycling or 'self-plagiarism' // *Research Policy*. vol. 48, no 1, pp. 492–502.
- Hoover G.A. (2006) A game-theoretic model of plagiarism // *Atlantic Economic Journal*. vol. 34, no 4, pp. 449–454.
- Johns A. (2003) The ambivalence of authorship in Early Modern natural philosophy // *Scientific authorship. Credit and intellectual property in science.* / M. Biagioli, P. Galison (eds.) New York and London: Routledge, pp. 91–112.
- Kaatz A., Vogelmann P.N., Carnes M. (2013) Are men more likely than women to commit scientific misconduct? Maybe, maybe not // *Mbio*. vol. 4, no 2, e00156-13.
- Kelley P.C., Agle B.R., DeMott J. (2005) Mapping our progress: Identifying, categorizing and comparing universities' ethics infrastructures // *Journal of Academic Ethics*. vol. 3, no 2-4, pp. 205–229.
- Lacetera N., Zirulia L. (2009) The economics of scientific misconduct // *The Journal of Law, Economics, and Organization*. vol. 27, no 3, pp. 568–603.
- Lewellyn K.B., Judge W.Q., Smith A. (2017) Exploring the questionable academic practice of conference paper double dipping // *Academy of Management Learning and Education*. vol. 16, no 2, pp. 217–236.
- Lind A. R. (2005) Evaluating research misconduct policies at major research universities: A pilot study // *Accountability in Research*. vol. 12, no 3, pp. 241–262.
- Louis K.S., Anderson M.S., Rosenberg L. (1995) Academic misconduct and values: The department's influence // *The Review of Higher Education*. vol. 18, no 4, pp. 393–422.
- Lee C.S., Schrank A. (2010) Incubating innovation or cultivating corruption? The developmental state and the life sciences in Asia // *Social Forces*. vol. 88, no 3, pp. 1231–1255.
- Macfarlane B., Zhang J., Pun A. (2014) Academic integrity: a review of the literature // *Studies in Higher Education*. vol. 39, no 2, pp. 339–358.
- McCabe D.L., Bowers W. J. (1994) Academic dishonesty among males in college: A thirty year perspective // *Journal of College Student Development*. vol. 35, no 1, pp. 5–10.
- McCabe D.L. (1992) The influence of situational ethics on cheating among college students // *Sociological Inquiry*. vol. 62, no 3, pp. 365–374.
- McCabe D.L., Trevino L.K. (1993) Academic dishonesty: Honor codes and other contextual influences // *The Journal of Higher Education*. vol. 64, no 5, pp. 522–538.
- Merton R. (1973) *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Qiu J. (2010) Publish or perish in China // *Nature*. vol. 463, no 7278, pp. 142–143.

- Resnik D.B., Rasmussen L.M., Kissling G.E. (2015) An international study of research misconduct policies // *Accountability in Research*. vol. 22, no 5, pp. 249–266.
- Rezaee Z., Elmore R.C., Szendi J.Z. (2001) Ethical behavior in higher educational institutions: The role of the code of conduct // *Journal of Business Ethics*. vol. 30, no 2, pp. 171–183.
- Schoenherr J., Williams-Jones B. (2011) Research integrity/misconduct policies of Canadian universities // *Canadian Journal of Higher Education*. vol. 41, no 1, pp. 1–17.
- Shapin S. (1994) *A social history of truth: civility and science in seventeenth-century England*. Chicago: University of Chicago Press.
- Spence M. (1973) Job market signaling // *The Quarterly Journal of Economics*. vol. 87, no 3, pp. 355–374.
- Stephan P. (1996) The economics of science // *Journal of Economic Literature*. vol. 34, no 3, pp. 1199–1235.
- Stiglitz J. (2000) The contributions of the economics of information to twentieth century economics // *The Quarterly Journal of Economics*. vol. 115, no 4, pp. 1441–1478.
- Stearns S.A. (1997) Administrative ramifications of student cheating // *Journal of the Association for Communication Administration (JACA)*. vol. 2, pp. 133–139.
- Sykes G.M., Matza D. (1957) Techniques of neutralization: A theory of delinquency // *American Sociological Review*. vol. 22, no 6, pp. 664–670.
- Thomas W. (1923) *The unadjusted girl: with cases and standpoint for behavior analysis*. Boston: Little, Brown and Company.
- Wajda-Johnston V.A., Handal P.J., Brawer P.A., Fabricatore A.N. (2001) Academic dishonesty at the graduate level // *Ethics and Behavior*. vol. 11, no 3, pp. 287–305.
- Walsh J.P., Lee Y.N., Tang L. (2019) Pathogenic organization in science: Division of labor and retractions // *Research Policy*. vol. 48, no 2, pp. 444–461.
- Ward D.A., Beck W.L. (1990) Gender and dishonesty // *The Journal of Social Psychology*. vol. 130, no 3, pp. 333–339.
- Wright D.E., Titus S.L., Cornelison J.B. (2008) Mentoring and research misconduct: An analysis of research mentoring in closed ORI cases // *Science and Engineering Ethics*. vol. 14, no 3, pp. 323–336.
- Zimmerman J. (1998) Academic dishonesty attitudes and self-reported behaviors in a university population // *Dissertation Abstracts International*, A 59, 4083.

Приложения

Таблица 1. Сравнение медиан и результаты теста Краскелла-Уоллиса по федеральным округам¹⁸

	Центр.	Дальнев.	Приволж.	Северо-К.	Сибир.	Северо-З.	Урал.
Дальневосточный	1,3						
Приволжский	-4*	-5,3					
Северо-Кавказский	-5,1	-6,4	-1,1				
Сибирский	0,1	-1,2	4,1	5,2			
Северо-Западный	1,1	-0,2	5,1**	6,2*	1		
Уральский	1,4	0,1	5,4	6,5	1,3	0,3	
Южный	-5,6*	-6,9	-1,6	-0,5	-5,7	-6,7**	0,3

*p<0,05, **p<0,01, ***p<0,001

Таблица 2. Сравнение медиан и результаты теста Краскелла-Уоллиса по типам организаций

	Другие	Институты РАН	Ведущие университеты
Институты РАН	4,7***		
Ведущие университеты	4,7***		0
Университеты	-1,4	-6,1***	-6,1***

*p<0,05, **p<0,01, ***p<0,001

Таблица 3. Сравнение медиан и результаты теста Краскелла-Уоллиса по годам

	2007	2008	2009	2010	2011	2012	2013	2014
2008	3							
2009	3	0						
2010	2,4	-0,6	-0,6					
2011	3	0	0	0,6				
2012	1,2	-1,8	-1,8	-1,2	-1,8			
2013	1,3	-1,7	-2	-1,1	-1,7	0,1		
2014	3,3	0,6	0,6	1,2	0,6	2,4	2,3	
2015	7,3	4,3*	4,3*	4,9	4,3	6,1***	6*	3,7

*p<0,05, **p<0,01, ***p<0,001

¹⁸ Таблицы 1-4 следует читать следующим образом: в них отображена значимость парных сравнений тестом Краскелла-Уоллиса, числами показана разность между медианами групп: из категории по столбцам вычитается медиана по строкам.

Таблица 4. Сравнение медиан и результаты теста Краскелла-Уоллиса по специальностям

	Физико-мат.	Химич.	Биолог.	Технич.	Сельск. и ветер.	Историр.
Химические	-14,6***					
Биологические	-5,7***	8,9				
Технические	-8,1***	6,5	-2,4			
Сельскохозяй. и ветеринарные	-23,6***	-9	-17,9***	-15,5***		
Исторические	-4,1**	10,5***	1,6	4	19,5***	
Экономические	-12,4***	2,2	-6,4	-4**	11,2**	-8***
Философские	-2,2	12,4***	3,5	5,9	21,4***	1,9
Филологические	-0,7	13,9***	5***	7,4***	22,9***	3,4***
Юридические	-19,4***	-4,8	-13,7***	-11,3***	4,2	-15,3***
Педагогические	-9,4***	5,2	-3,7	-1,3	14,2***	-5,3*
Медицинские	-8,9***	5,7	-3,2	-0,8	14,7***	-4,8**
Искусствоведческие	2,1	16,7***	7,8	9,9***	25,7***	6,2***
Психологические	-6,3*	8,3	-0,6	1,8	17,3***	-2,2
Социологические	-6,1	8,5	-0,4	2	17,5***	-2
Политические	-5,7	8,9	0	2,4	17,9***	-1,6
Культурологические	-0,8	13,8*	4,9	7,3	22,8***	3,3
Геологические	-2,5	12,1***	3,2	5,6	21,1***	1,6

	Эконом.	Философ.	Филолог.	Юридич.	Педагог.	Медицинские
Химические						
Биологические						
Технические						
Сельскохозяй. и ветеринарные						
Исторические						
Экономические						
Философские	9,9***					
Филологические	11,4***	1,5				
Юридические	-7,3**	-17,2***	-18,7***			
Педагогические	2,7	-7,2***	-8,7***	10***		
Медицинские	3,2	-6,7***	-8,2***	10,5***	0,5	
Искусствоведческие	14,2***	4,3**	2,8*	21,5***	11,5***	11***
Психологические	5,8	-4,1	-5,6**	13,1***	3,1	2,6
Социологические	6	-3,9	-5,4**	13,3***	3,3	2,8
Политические	6,4	-3,5	-5**	13,7***	3,7	3,2
Культурологические	11,3**	1,4	-0,1	18,6***	8,6	8,1*
Геологические	9,6***	-0,3	-1,8	16,9***	6,9***	6,4***

	Искусствовед.	Психолог.	Социологические	Политические	Культурол.
Химические					
Биологические					
Технические					

Сельскохозяй. и ветеринарные						
Исторические						
Экономические						
Философские						
Филологические						
Юридические						
Педагогические						
Медицинские						
Искусствоведческие						
Психологические	-8,4***					
Социологические	-8,2***	0,2				
Политические	-7,8***	0,6	0,4			
Культурологические	-2,9	5,5	5,3	4,9		
Геологические	-4,6*	3,8	3,6	3,2	-1,7	
*p<0,05, **p<0,01, ***p<0,001						

Таблица 5. Корреляция Спирмена между процентом некорректных заимствований и показателями публикационной активности

Название переменной	Коэффициент корреляции
Публикации eLIBRARY.RU	0,04
Публикации РИНЦ	0,04
Публикации ядро	-0,12
Цитирования eLIBRARY.RU	0,02
Цитирования РИНЦ	0,01
Цитирования ядро	-0,11
Индекс Хирша eLIBRARY.RU	0,05
Индекс Хирша РИНЦ	0,04
Индекс Хирша ядро	-0,09
N процитировавших работ	0,04
N ссылок на самую цитируемую	0
Среднее N цитирований	0,03
Индекс Хирша	0,08
Индекс Хирша статьи	0,01
Начало карьеры	0,12
Соавторы	0,08
Импакт-фактор опубликованные	-0,01
Импакт-фактор процитированные	0,02
Публикации процитированные	0,03
Публикации процитированные, %	-0,01
Самоцитирования	-0,09

Самоцитирования, %	-0,14
Цитирования соавторами	-0,02
Цитирования соавторами, %	-0,05
Статьи зарубежные	-0,10
Статьи зарубежные, %	-0,13
Статьи российские	0,08
Статьи российские, %	0,05
Статьи ВАК	0,07
Статьи ВАК, %	0,03
Статьи переводные	-0,16
Статьи переводные, %	-0,18
Статьи с импакт-фактором	0,04
Статьи с импакт-фактором, %	-0,02
Цитирования зарубежные	-0,05
Цитирования зарубежные, %	-0,09
Цитирования российские	0,02
Цитирования российские, %	-0,01
Цитирования ВАК	0,00
Цитирования ВАК, %	-0,02
Цитирования переводные	-0,12
Цитирования переводные, %	-0,12
Цитирования с импакт-фактором	-0,01
Цитирования с импакт-фактором, %	-0,10
Публикации 5 лет	0,04
Публикации 5 лет, %	0
Цитирования 5 лет	0,04
Цитирования 5 лет, %	0,05
Цитирования всех за 5 лет	0,04
Цитирования всех за 5 лет, %	0,07
Регистрация	-0,02