

ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

А.Л. Лукьянова

**ГЕНДЕРНЫЙ РАЗРЫВ В ОПЛАТЕ ТРУДА:
ЗА РАМКАМИ ДЕКОМПОЗИЦИИ
ОАКСАКИ — БЛАЙНДЕРА**

Препринт WP3/2022/08

Серия WP3

Проблемы рынка труда

Москва
2022

УДК 331.215

ББК 65.245

Л84

Редактор серии WP3
«Проблемы рынка труда»
В.Е. Гимпельсон

Лукьянова, А. Л.

Л84 Гендерный разрыв в оплате труда: за рамками декомпозиции Оаксаки — Блайндера [Электронный ресурс] : препринт WP3/2022/08 / А. Л. Лукьянова ; Нац. исслед. ун-т «Высшая школа экономики». — Электрон. текст. дан. (1,7 Мб). — М. : Изд. дом Высшей школы экономики, 2022. — 65 с. — (Серия WP3 «Проблемы рынка труда»).

В работе изучаются перспективы использования новейших методологических подходов к анализу гендерного разрыва в оплате труда на российских данных. Ключевым параметром при оценивании являлась средняя величина необъясненного гендерного разрыва в оплате труда. Используя данные РМЭЗ ВШЭ за 2005–2020 гг., мы сравнивали между собой линейную регрессию, декомпозицию Оаксаки — Блайндера и ряд методов, основанных на индексах склонности и алгоритмах машинного обучения. Последняя группа методов позволяет более гибко и робастно моделировать влияние пола, в том числе во взаимодействии с другими значимыми характеристиками женской и мужской занятости, такими как гендерная сегрегация и гендерные роли. Мы выяснили, что выборки мужчин и женщин действительно крайне гетерогенны и гетерогенность увеличивалась в течение рассматриваемого периода по мере усиления гендерной сегрегации. Гетерогенность выборок создает серьезные проблемы как для традиционных, так и для новых методов оценивания, поскольку ведет к нарушению условия сопоставимости характеристик (common support) и увеличивает риски неверной спецификации моделей. Обе проблемы могут быть сглажены при помощи методов с двойной робастностью. Вместе с тем небольшой объем годовых выборок в РМЭЗ ВШЭ создает ограничения для использования новых методов оценивания. Применительно к данным РМЭЗ ВШЭ нам не удалось добиться заметного снижения необъясненного гендерного разрыва по сравнению с декомпозицией Оаксаки — Блайндера. Однако методы, основанные на индексах склонности, имеют более сильную каузальную интерпретацию, поэтому совпадение оценок укрепляет обоснованность результатов, полученных с помощью стандартной регрессионной методологии.

УДК 332. 215
ББК 65.245

Ключевые слова: гендерный разрыв в оплате труда, декомпозиция Оаксаки — Блайндера, гетерогенность, оценка воздействия, индекс склонности, LASSO

Лукьянова Анна Львовна — кандидат экономических наук, старший научный сотрудник Центра трудовых исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ); Россия, 101000, Москва, ул. Мясницкая, д. 20; E-mail: alukyanova@hse.ru

Благодарности

Работа выполнена в рамках Программы фундаментальных исследований НИУ ВШЭ.

Препринты Национального исследовательского университета
«Высшая школа экономики» размещаются по адресу: <http://www.hse.ru/org/hse/wp>

© Лукьянова А.Л., 2022

1. Введение

Многочисленные работы по оценке гендерного разрыва в России основаны на использовании декомпозиции Оаксаки — Блайндера (Blinder, 1973; Оаксака, 1973) и ее модификаций. Методологический подход, положенный в основу декомпозиции Оаксаки — Блайндера, включает в себя разделение разрыва на объясненную и необъясненную части и выделение вклада отдельных ковариат в каждую из этих частей. Результаты, полученные на российских данных, показывают, что значительная доля гендерного разрыва в оплате труда не поддается объяснению (Oshcherkov, 2021; Рошин, Емелина, 2022). Неспособность популярных методов объяснить гендерный разрыв может быть связана с гетерогенностью двух выборок. Косвенным образом на важность гетерогенности указывает тот факт, что даже в объясненной части ключевым объясняющим фактором является гендерная сегрегация. Гетерогенность выборок ставит под вопрос адекватность использования классических методов декомпозиции, которые требуют соблюдения условия одинакового распределения характеристик в обеих группах (Fortin et al., 2011). Более того, при наличии гетерогенности необъясненная часть не может интерпретироваться как эффект дискриминации, и сама объясненная часть может быть сильно занижена.

Занижение объясненной части может быть обусловлено как эконометрическими, так и содержательными причинами. В прикладных исследованиях обе группы причин, скорее всего, действуют совместно. Эконометрические проблемы связаны с жесткостью предпосылок, накладываемых популярными методами декомпозиции. Так, методология Оаксаки — Блайндера базируется на фиксированной функциональной форме: зависимость между ковариатами и заработной платой предполагается линейной. Все методы декомпозиции, завязанные на линейных регрессиях, сталкиваются с проблемой обеспечения сопоставимости сравниваемых мужчин и женщин. Технически она возникает из-за отсутствия пересечения множеств характеристик мужчин и женщин, что происходит, если совместные распределения характеристик мужчин и женщин сильно различаются между собой. В результате в используемых выборках сложно найти мужчин и женщин с одинаковыми значениями по всем важным производительным характеристикам одновременно. В англоязычной литературе эта проблема известна как *problem of common support (overlap)*. Наибольшие опасения вызывают те характеристики и их комбинации, по которым наблюдается сильная гендерная сегрегация, — отраслевая и профессиональная принадлежность. С содержательной точки зрения при декомпозиции разрыва важно учесть, что женщины чаще, чем мужчины, находятся под одновременным воздействием нескольких факторов уязвимости (например, занятость в бюджетном секторе, неполная занятость, семейные обязанности). Стандартные методы декомпозиции этого не делают.

В последние годы в качестве альтернативы декомпозиции Оаксаки — Блайндера все чаще применяются методы, основанные на методологии оценки воздействия и расчете индексов склонности (*propensity scores*). Среди этих методов получили распространение точный мэтчинг, мэтчинг по индексу склонности, обратное вероятностное взвешивание (IPW) и его разновидности. В недавних работах (Meera et al., 2020) для оценки разрыва применяются последние достижения в области методологии оценки воздействия. В частности, модели с несколькими уровнями воздействия за счет пересечения гендера с другими важными характеристиками мужской и женской занятости. Еще одна методологиче-

ская инновация последних лет связана с использованием алгоритмов машинного обучения для отбора объясняющих переменных (см., например, (Bach et al., 2018; Strittmatter, Wunsch, 2021; Böheim, Stöllinger, 2021)). В этой статье используются все перечисленные подходы.

2. Данные и ключевые переменные

Мы используем данные РМЭЗ ВШЭ за 2005–2020 гг. из репрезентативной части выборки. «Расширенная» выборка ограничена индивидами в возрасте от 18 до 59 лет с полной информацией по полу и уровню образования. В эту выборку включены как занятые, так и незанятые индивиды. К занятым отнесены респонденты, указавшие наличие постоянной работы, а также те, кто сообщил о наличии приносящих доход занятий, которым они посвятили не менее 20 часов на протяжении последних 30 дней. Находящиеся в декретном отпуске считались незанятыми. Расширенная выборка насчитывает 117 225 наблюдений¹, из которых 51,8% относятся к женщинам. Данная выборка используется в дескриптивной части исследования.

При анализе гендерного разрыва в заработных платах рассматриваются лишь респонденты, занятые по найму по основному месту работы за исключением военнослужащих. Самозанятые исключаются из выборки. При этом самозанятыми считались респонденты, попавшие в одну из трех категорий: (1) назвавшие своим основным занятием фермерство или предпринимательство (вопрос J90); (2) указавшие, что занимаются на основной работе предпринимательской деятельностью (вопрос J29) и являются владельцами или совладельцами предприятия, на котором работают (вопрос J26); (3) не имеющие постоянной работы, но имеющие приработки, которыми они занимались не менее 20 часов в течение последних 30 дней (вопросы J1 и J56). Остальные категории занятых отнесены к занятым по найму, которые представляют основной объект исследования в данной работе. Эта выборка включает 76 778 наблюдений², из которых 49% относятся к женщинам.

Основной зависимой переменной в нашем исследовании является логарифм часовой заработной платы, который был рассчитан как отношение среднемесячной заработной платы за последние 12 месяцев к средней продолжительности рабочей недели, умноженной на 4,33 ($52/12 = 4,33$ — средняя продолжительность месяца в неделях). Для респондентов, у которых отсутствовали данные о средней заработной плате или продолжительности рабочей недели, использовались данные о сумме заработной платы и фактически отработанном рабочем времени за последние 30 дней.

Ключевая переменная, характеризующая воздействие, — это гендер (0 — мужчина, 1 женщина). Еще несколько переменных использовались как в качестве ковариат, так и — на отдельных этапах исследования — в качестве дополнительных переменных воздействия:

- parent: родительство (1 — при наличии собственных или приемных детей младше 18 лет, 0 — отсутствие детей младше 18 лет);

¹ Расширенная выборка за 2005, 2010, 2015, 2020 гг., по которой ведутся основные эконометрические расчеты, включает 28 797 наблюдений.

² Выборка наемных работников за 2005, 2010, 2015, 2020 гг., по которой ведутся эконометрические расчеты, включает 18 739 наблюдений.

- sector: сектор занятости (0 — частный, 1 — государственный);
- female_occupation: занят(а) в женской профессии (1 — да, 0 — нет);
- male_industry: занят(а) в мужской отрасли (1 — да, 0 — нет).

Респондент считался занятым в государственном секторе, если он(-а) ответил(-а) положительно на вопрос о том, является ли государство владельцем предприятия, на котором он(-а) работает, и одновременно ответил(-а) отрицательно на вопросы о том, есть ли среди (со-)владельцев российские и иностранные физические лица или компании. Все остальные респонденты отнесены к занятым в частном секторе. Таким образом, занятые на предприятиях смешанной формы собственности также классифицировались как занятые в частном секторе.

Для данных РМЭЗ ВШЭ оптимальным является выделение профессий на уровне второго разряда классификации ISCO-08. С углублением детализации увеличиваются риски ошибок измерения и неверной кодировки профессий. В небольших обследованиях могут возникать дополнительные смещения за счет небольшого и, соответственно, нерепрезентативного числа респондентов во многих узких группах. Например, на четвертом уровне кодировки в нашей выборке в разные годы выделяется от 297 до 342 групп, то есть от года к году заметно меняется даже количество выделенных профессий. На наиболее агрегированном первом уровне кодировки, наоборот, выделяется всего 10 укрупненных групп, которые отражают скорее общий уровень образования, сложность работы и сферу деятельности, нежели имеют отношение к конкретным профессиям. Поэтому при анализе сегментации более целесообразно ориентироваться на второй или третий уровень классификации. Для двухразрядной кодировки в нашей выборке выделяется 38–39 групп, для трехразрядной — 115–120 групп. При этом с содержательной точки зрения более предпочтительна трехразрядная классификация. Например, на втором уровне выделяются специалисты в области образования (группа 23), на третьем уровне идет детализация по уровням системы образования: профессорско-преподавательский персонал университетов и других организаций высшего образования (231), преподаватели средних профессиональных образовательных организаций (232), педагогические работники в средней школе (233), педагогические работники в начальном и дошкольном образовании (234), другие специалисты в области образования (235). Таким образом, на третьем уровне классификации группы уже довольно близки к тем представлениям о профессиях, которыми мы оперируем в обычной жизни. Однако в связи с относительно небольшим объемом годовых выборок РМЭЗ ВШЭ на третьем уровне классификации приемлемый уровень наполняемости обнаруживается лишь в самых массовых профессиях. В результате мы выделили 24 профессиональных группы, приближенных ко второму уровню классификации ISCO-08 (табл. 1).

Используемая в РМЭЗ ВШЭ классификация отраслей (вопрос J4.1) не соответствует ни одной из общепринятых, а варианты ответов выбирали сами респонденты, поэтому распределение по отраслям заметно отличается от росстатовского. Мы объединили имеющиеся варианты ответов в 14 отраслей³.

³ Выделенные отрасли: (1) сельское и лесное хозяйство; (2) добыча полезных ископаемых; (3) легкая и пищевая промышленность; (4) тяжелая промышленность; (5) строительство; (6) транспорт и связь; (7) государственное управление, армия, полиция; (8) образование; (9) наука, искусство, спорт; (10) здравоохранение, социальное

Отрасли и профессиональные группы можно условно разделить на мужские и женские в зависимости от того, какой пол доминирует в той или иной отрасли/профессии. Поскольку в России наемная рабочая сила наполовину состоит из женщин и участие женщин оставалось относительно стабильным в течение всего рассматриваемого периода, то мы можем зафиксировать «постоянные» критерии разделения на мужские и женские отрасли/профессии. К женским (мужским) отнесены те отрасли/профессии, в которых доля женщин (мужчин) равна или превышает 70%. Соответственно, отрасли/профессии, в которых доли женщин и мужчин находятся в промежутке между 30 и 70%, считаются смешанными или гендерно-нейтральными⁴. Определенное количество респондентов попадает в дополнительную категорию неклассифицированных отраслей/профессий. Это те отрасли/профессии, в которых в рассматриваемом году обследования занято менее 20 респондентов (независимо от пола).

Следующие переменные использовались только в качестве ковариат:

- age: возраст (3 категории: 18–25 лет / 26–45 лет / 46 лет и старше);
- married: семейное положение (3 категории: не состоит в браке / состоит в зарегистрированном браке / состоит в незарегистрированном браке);
- education: образование (4 категории: ниже среднего / среднее / среднее профессиональное / высшее);
- est_expert: ожидаемый общий стаж работы;
- tenure: специфический стаж (стаж работы у текущего работодателя);
- продолжительность рабочего времени (3 категории: до 30 часов в неделю / 30–48 часов в неделю / более 48 часов в неделю);
- тип поселения (4 категории: Москва и СПб / областной центр / город — не областной центр / пгт и село);
- дамми-переменные для регионов (7 категорий: на уровне федеральных округов);
- дамми-переменные для отраслей;
- дамми-переменные для профессий;
- дамми-переменные для размера предприятия (5 категорий: микро- и малые предприятия (1–99 чел.) / средние (100–249 чел.) / крупные (250–499 чел.) / очень крупные (500+ чел.) / нет ответа).

обслуживание; (11) торговля, общепит, бытовые услуги; (12) финансы, деловые услуги; (13) ЖКХ; (14) прочие отрасли.

⁴ В литературе не существует общепринятого деления профессий на мужские и женские. В опубликованных работах пороговые значения фиксировались на уровне 70% (Наким, 1993) или 80% (Anker, 1998). При анализе данных в условиях быстрого роста участия женщин в рабочей силе имеет смысл использовать динамические пороги, привязанные к доле представителей определенного пола в занятости: например, профессия считается мужской (женской), если доля занятых в ней мужчин (женщин) превышает их долю в общей численности занятых на 5 п. п. (Beller, 1982) или 10 п. п. (Blau et al., 1998). В действительности разные профессии, скорее всего, имеют различные критические значения, по достижении которых внутри профессии начинают происходить значимые изменения под влиянием притока лиц «нетипичного» пола. Так, по расчетам Рап (2015), в 1940–1990 гг. на американском рынке труда прирост занятости мужчин в профессии резко сокращался, когда доля женщин в этой профессии превышала 25–40% для беловоротничковых профессий и 13–30% для синеворотничковых профессий.

Для расчета ожидаемого опыта работы мы оценили отдельно для каждого возрастного интервала (18–22, 23–25, 26–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59 лет) пробит-модели следующего вида:

$$P(empl = 1|x) = \alpha_0 + \alpha_1 age + \alpha_2 parent + \alpha_3 n_{kids} + \beta married + \gamma education + \delta D + u, \quad (1)$$

где n_{kids} — количество детей до 18 лет, D — вектор дамми-переменных регионов и типов поселений. Модели оценивались отдельно для мужчин и женщин. Для повышения точности расчетов использовались данные за все годы с 2005 по 2020 г.

Далее для каждого индивида была рассчитана вероятность занятости (включая самозанятость и занятость по найму) в течение каждого отдельного года в возрасте от 18 до 59 лет. Сумма этих вероятностей к текущему возрасту индивида и составляет ожидаемый опыт работы. Результаты расчетов представлены на рис. 1. Мы видим, что у женщин ожидаемый стаж действительно в среднем короче в возрастах моложе 42 лет, однако в более старших возрастах разрыв меняет знак, и женщины начинают опережать мужчин по общей продолжительности стажа. При этом пик разрыва в пользу мужчин достигается в возрасте 33–34 года и составляет порядка 2,1–2,2 года, что, очевидно, связано с рождением детей. В дальнейшем гендерный разрыв сокращается: женщины возвращаются на рынок труда, в то время как среди мужчин более распространен досрочный выход на пенсию и ослабление связи с рынком труда. В результате к 53–54 годам преимущество в продолжительности стажа в пользу женщин достигает 2,1 года. В более старших возрастах гендерный разрыв вновь начинает сокращаться, что обусловлено более низким пенсионным возрастом для женщин.

Настоящее исследование преследует главным образом методологические цели, однако значительная часть используемых методов исходит из предпосылки о независимости и одинаковом распределении наблюдений (i.i.d.), что исключает использование массива из объединенных данных за все годы. Поэтому эконометрические оценки были сделаны лишь для 2005, 2010, 2015 и 2020 гг. Большинство дескриптивных оценок приводятся на графиках для всех лет рассматриваемого периода.

3. Дескриптивный анализ

Начнем с описания состава выборки. В табл. 2а и 2б показано распределение мужчин и женщин по статусу занятости. Мужчины имели более высокий уровень занятости, чем женщины, на всем протяжении рассматриваемого периода. Разрыв в уровнях составлял порядка 10 п. п., что существенно выше, чем в данных Росстата. Расхождения с официальными оценками отчасти связаны с нашей классификацией респондентов, находящихся в декрете, как незанятых, что сказывается прежде всего на уровне занятости женщин и практически не затрагивает мужчин.

Пик женской занятости пришелся на 2009 г., когда были заняты 69% женщин в возрасте 18–59 лет. В последующие годы уровень занятости снижался и в 2020 г. составил 64%. Это снижение, скорее всего, связано со старением населением и достижением пенсионного возраста представителями многочисленных послевоенных поколений. Пенси-

онная реформа 2018 г. не смогла переломить негативного тренда. Очевидно, свою роль могло сыграть и повышение рождаемости в начале 2010-х годов, а также последствия пандемии коронавируса. Среди мужчин колебания уровня занятости внутри рассматриваемого периода не были столь резкими и пик пришелся на более поздний период — 2010–2013 гг., когда уровень занятости достиг 78%. К 2020 г. он снизился до 75%. Пятилетний интервал между пиками мужской и женской занятости заставляет предположить, что основным драйвером снижения занятости действительно являлся выход на пенсию поколений бэби-бумеров.

Как у мужчин, так и у женщин самый высокий уровень занятости наблюдается в средних возрастах, в 26–45 лет. Мужчины значительно чаще отдают предпочтение самозанятости, особенно в средних возрастах. Среди мужчин самозанятыми являются 10% всех занятых, тогда как среди женщин лишь 6%. Таким образом, мы не наблюдаем какого-либо вытеснения женщин в сегменты вторичной занятости. Вряд ли женщины чаще, чем мужчины, сталкиваются с необходимостью браться за малоодоходную самозанятость по причине низких шансов найти занятость по найму.

В табл. 3 представлены описательные статистики по основным переменным для подвыборки занятых по найму. Мы видим существенные гендерные различия в продолжительности рабочего времени, доле занятых неполное рабочее время и в заработных платах. Женщины в среднем более образованны, чаще заняты в государственном секторе и на предприятиях меньшего размера⁵. Среди женщин больше тех, кто не состоит в браке и не имеет детей до 18 лет, что отчасти связано с более высоким средним возрастом женской части выборки. Различия в структурах занятости по отраслям и профессиям подробно обсуждаются в следующем разделе.

На рис. 2 показана динамика гендерного разрыва в средней продолжительности рабочего времени и средней часовой заработной плате, выраженного в процентах от значения соответствующего показателя у мужчин. На протяжении рассматриваемого периода произошло значительное снижение гендерного разрыва в заработных платах на фоне стабильных различий в продолжительности рабочего времени. Разрыв в заработных платах сократился за 2005–2020 гг. с 26 до 13%. Это снижение не было монотонным: после резкого сокращения разрыва в середине 2000-х годов произошел откат. Однако после 2012 г. сокращение разрыва возобновилось и имело устойчивый характер. Мы не склонны связывать наблюдаемую динамику с каким-то одним фактором (например, с экономическим циклом или демографическими тенденциями) или с конкретными изменениями в политике (например, с реформой оплаты труда в бюджетном секторе или повышением МРОТ). Скорее всего, все перечисленные факторы вносили вклад в снижение гендерного разрыва в течение рассматриваемого периода.

4. Гендерная сегрегация

Женщины явно не являются меньшинством среди занятых по найму: их доля даже в младшей возрастной группе составляет около 47%, а в старших возрастах доходит до

⁵ Чтобы избежать потери большого числа наблюдений, для переменной размера предприятия мы сохранили опцию «Нет ответа», объединив в нее тех, кто ответил «Не знаю», и тех, кто отказался от ответа.

52%. При этом рабочие места и условия занятости женщин и мужчин сильно отличаются друг от друга, что отражается и на заработных платах. Как и в других странах, наиболее отчетливо эти различия проявляются в распределении мужчин и женщин по профессиям и отраслям.

Для иллюстрации этих различий был рассчитан индекс диссимилиации Дункана (Duncan, Duncan, 1955). Для профессиональной структуры индекс рассчитывается по следующей формуле:

$$D_t = 0,5 * \sum_j |m_{jt} - f_{jt}|, \quad (2)$$

где m_{jt} (f_{jt}) — доля мужчин (женщин), занятых в j -й профессиональной группе, в общей численности занятых мужчин (женщин) в период t . Индекс Дункана принимает значения от 0 до 1 и показывает сколько мужчин или женщин — в долях от единицы — должны сменить профессию, чтобы добиться равномерного распределения полов по профессиям. Нулевое значение наблюдается при полном отсутствии сегрегации (во всех профессиях мужчины и женщины составляют ровно половину занятых), единица — при абсолютной сегрегации (во всех профессиях заняты либо только мужчины, либо только женщины). Сходным образом индекс Дункана может быть рассчитан для отраслей, секторов, типов предприятий и т. д.

Профессиональная сегрегация. На рис. 3 представлены результаты расчетов индекса Дункана для профессиональной структуры. Для принятой в этом исследовании кодировки профессий уровень гендерной сегрегации за 2005–2020 гг. увеличился с 0,57 до 0,61. Другими словами, в 2020 г. 61% женщин должны были сменить профессию, чтобы иметь одинаковое с мужчинами распределение по профессиям. Согласно Massey, Denton (1993), индекс Дункана свыше 0,6 свидетельствует о высокой сегрегации, значения от 0,3 до 0,6 соответствуют умеренной сегрегации и меньше 0,3 — слабой сегрегации. Таким образом, российский рынок труда характеризуется высокой степенью гендерной сегрегации по профессиональному признаку.

Об усилении профессиональной поляризации по признаку полу говорят данные об изменении распределения работников по «мужским» и «женским» профессиям (табл. 4). Мы видим, что за 2005–2020 гг. уменьшилась доля женщин, занятых в смешанных профессиях, и перетоки были равномерно перераспределены между мужскими и женскими профессиями. Среди мужчин также происходило интенсивное сокращение доли занятых в смешанных профессиях, но с явным преобладанием потоков в мужские профессии. Причиной могло быть как перемещение мужчин из смешанных профессий в мужские, так и изменение статуса отдельных профессий со смешанного на мужской. Представленность мужчин в женских профессиях изменилась слабо. Следует отметить, что женские профессии в целом являются более «открытыми» по сравнению с мужскими: в 2020 г. в женских профессиях трудились 18% от всех мужчин, занятых по найму, в то время как в мужских профессиях было занято всего 9% от всех женщин, занятых по найму. Закрытости мужских профессий могут способствовать как законодательные запреты и тяжелые условия труда, так и традиционная сексистская субкультура соответствующих профессиональных сообществ. При этом сокращение занятости в промышленности и, соответственно, сни-

жение потребности в рабочих кадрах не привело к сокращению занятости мужчин в мужских профессиях: продолжают расширяться «новые» мужские профессии в сферах охраны и ИКТ.

Отраслевая сегрегация. В табл. 5 показано изменение отраслевой структуры занятости мужчин и женщин между 2005 и 2020 гг. Паттерны изменения отраслевой структуры схожи для обоих полов: увеличилась доля торговли и деловых услуг на фоне продолжающегося сокращения занятости в сельском хозяйстве, легкой и пищевой промышленности, а также в строительстве. Ощутимо выросла доля занятых в образовании и государственном управлении, особенно среди женщин. Среди мужчин заметен относительный рост занятости в тяжелой промышленности и в силовых структурах.

Эти изменения отразились в увеличении отраслевой сегрегации по полу (рис. 3). Пиковые значения индекса Дункана были достигнуты в 2015 г.: индекс Дункана вырос с 0,33 в 2005 г. до 0,42 в 2015 г. В дальнейшем сегрегация начинает сокращаться и выходит на плато в последние годы. При этом достигнутый к 2020 г. уровень (0,38) все равно существенно выше исходного. Согласно Massey, Denton (1993), такие значения гендерной сегрегации можно отнести к умеренному уровню.

Усиление отраслевой поляризации по полу отчетливо проявляется в изменении доли работников, занятых в мужских и женских отраслях (табл. 6). Мы видим резкое увеличение доли занятых в мужских отраслях, особенно среди мужчин, на фоне сокращения доли занятых в смешанных отраслях. В основе этого тренда лежит не рост занятости в «старых» мужских отраслях, а расширение «списка» мужских отраслей. Если считать женскими те отрасли, в которых женщины составляют свыше 70% работников, то в 2020 г. таковыми следует — в нашей классификации — считать здравоохранение и образование. Этот список не изменился по сравнению с 2005 г. Мужскими отраслями в 2020 г. являлись строительство, сельское хозяйство, добывающая и тяжелая промышленность, транспорт и связь. При этом в 2005 г. доля мужчин превышала 70% лишь в строительстве. Впрочем, для перехода столь значительного числа отраслей в категорию мужских потребовались небольшие изменения в их гендерном составе. Во всех «новых» мужских отраслях доля мужчин составляла 63–69% и в 2005 г. В результате в 2020 г. около половины мужчин, занятых по найму, трудились в мужских отраслях, в то время как в 2005 г. таковых было всего 14%, а 78% мужчин были заняты в смешанных отраслях (в 2020 г. — 47%).

Одновременно несколько снизился удельный вес занятых в женских отраслях. Доля женщин, занятых в женских отраслях, практически не изменилась. Занятость женщин в смешанных отраслях снизилась, но в меньшей степени, чем среди мужчин. Среди женщин также наблюдался рост доли занятых в мужских отраслях, что связано с резким увеличением числа мужских отраслей.

Гендерная сегрегация по другим характеристикам

На рис. 4 представлены индексы Дункана для других характеристик занятости: сектора (частный vs государственный), уровня образования и размера предприятия. Мы видим заметный рост гендерной сегрегации по секторам и размерам предприятия на фоне небольшого снижения образовательной сегрегации. При этом по всем трем характеристикам индекс Дункана на всем протяжении рассматриваемого периода находился ниже 0,3,

то есть, в отличие от отраслевой и профессиональной сегрегации по полу, в отношении этих характеристик сегрегация являлась слабой, что неудивительно, учитывая, количество выделенных групп существенно меньше по этим переменным.

Рост гендерной сегрегации по секторам тесным образом связан с увеличением отраслевой сегрегации и отражает описанное выше расширение занятости женщин в преимущественно бюджетных отраслях образования и государственного управления на фоне увеличения занятости мужчин в производственных отраслях, где доминирует частный сектор. Гендерная сегрегация по размерам предприятия подпитывалась более быстрой реструктуризацией женской занятости в пользу малых предприятий.

5. Методология

5.1. Модели с одной переменной воздействия

Обозначим дамми-переменную для пола респондента как F_i , где $F_i = 1$ для женщин, занятых по найму, и $F_i = 0$ для мужчин, занятых по найму⁶. Зависимую переменную — логарифм часовой заработной платы — обозначим через Y_i . Тогда средний (безусловный) гендерный разрыв в оплате труда составит⁷:

$$\Delta = E[Y_i | F_i = 1] - E[Y_i | F_i = 0]. \quad (3)$$

Гендерный разрыв в средних заработных платах можно разделить на две части: объясненную и необъясненную. Пусть X_i — вектор наблюдаемых характеристик работников и рабочих мест, которые влияют на размер заработной платы. Предположим, что предсказанная заработная плата мужчин, имеющих такие же наблюдаемые характеристики, как женщины, равна $E_{X|F=1}[\mu_0(x)]$, где $\mu_0(x) = E[Y_i | F_i = 0, X_i = x]$. Прибавляя и вычитая это выражение в уравнение (3), получаем:

$$\Delta = \underbrace{E[Y_i | F_i = 1] - E_{X|F=1}[\mu_0(x)]}_{\text{Необъясненная часть } \delta} + \underbrace{E_{X|F=1}[\mu_0(x)] - E[Y_i | F_i = 0]}_{\text{Объясненная часть } \eta}. \quad (4)$$

Объясненная часть отражает «объективные» различия в заработных платах мужчин и женщин, которые связаны с различиями в «запасах» наблюдаемых производительных характеристик, например, с тем, что мужчины и женщины имеют разный уровень образования и трудятся на разных рабочих местах. Необъясненная часть гендерного разрыва — это та часть различий в заработных платах мужчин и женщин, которая не связана с различиями в наблюдаемых характеристиках. Это различия в заработных платах мужчин и женщин с одинаковыми наблюдаемыми характеристиками. Необъясненная часть традиционно связывается с дискриминацией, но она также отражает воздействие различий в ненаблюдаемых производительных характеристиках и гендерные особенности отбора в занятость. Именно оценка среднего необъясненного гендерного разрыва, то есть ожидае-

⁶ Для упрощения обозначений индекс времени опущен во всех формулах.

⁷ Часть обозначений, формул и некоторые методологические подходы заимствованы из работы Strittmatter, Wunsch (2021).

мого разрыва в средних заработных платах «идентичных» мужчин и женщин, является основной задачей в большинстве исследований гендерного неравенства на рынке труда.

В качестве исходного пункта анализа мы оцениваем простую линейную регрессию (LRM — linear regression model) с включением дамми-переменной для пола:

$$Y_i = \alpha + F_i \delta_{LRM} + X_i \beta + \varepsilon_i, \quad (5)$$

где вектор X_i включает ковариаты, перечисленные в разделе 2. Оценка коэффициента $\hat{\delta}_{LRM}$ при переменной пола может интерпретироваться как средний необъясненный гендерный разрыв в оплате труда. Если интерпретировать полученный коэффициент в терминах эффектов воздействия, то $\hat{\delta}_{LRM}$ может рассматриваться как оценка эффекта воздействия для подвергшихся воздействию (*Average Treatment Effect on the Treated, ATET*)⁸. При этом следует помнить, что линейная регрессия исходит из очень жестких предположений о механизмах формирования заработных плат мужчин и женщин и о размере δ_{LRM} . В частности, предполагается, что (1) зависимость заработной платы от наблюдаемых характеристик носит линейный характер, (2) все наблюдаемые характеристики имеют одинаковые отдачи на рынке труда для обоих полов, а (3) величина необъясненного разрыва одинакова для всех групп работников.

Два последних условия ослабляются при использовании следующего метода оценивания — декомпозиции Оаксаки — Блайндера (Blinder, 1973; Оакаса, 1973). Этот метод исходит из того, что вследствие дискриминации и действия других искажающих факторов рынок труда распадается на два сегмента — рынок труда мужчин и рынок труда женщин с разными структурами отдач от наблюдаемых характеристик, то есть с разными коэффициентами в уравнениях заработной платы. Сама процедура декомпозиции зависит от конкретных предположений исследователей относительно того, какой вид имеет структура отдач в отсутствие искажений и дискриминации. Мы используем вариант, при котором предполагается, что структура отдач у мужчин не подвержена влиянию дискриминации и других искажений. У женщин, наоборот, фактические отдачи искажены, и в отсутствие дискриминации женщины получали бы от своих характеристик те же отдачи, что фактически получают мужчины.

При использовании метода Оаксаки — Блайндера расчеты производятся в два шага. На первом шаге оценивается уравнение заработной платы отдельно для мужчин:

$$Y_i = \alpha_0 + X_i \beta_0 + u_i. \quad (6)$$

Коэффициенты β_0 описывают структуру отдач от наблюдаемых характеристик для мужчин. На втором шаге оценки коэффициентов этого уравнения используются для расчета контрафактической заработной платы женщин: $\hat{\mu}_0(x) = \hat{\alpha}_0 + X_i \hat{\beta}_0$ и среднего необъясненного разрыва для женщин, который равен среднему остатку из «мужского» уравнения заработной платы, рассчитанному по подвыборке женщин:

⁸ В случае линейной регрессии АТЕ (*Average Treatment Effect*) и АТЕТ (*Average Treatment Effect on the Treated*) равны между собой, что вытекает из предположений данной модели.

$$\hat{\delta}_{\text{ОВ}} = \frac{1}{N_1} \sum_{i=1}^N F_i (Y_i - \hat{\mu}_0(x)), \quad (7)$$

где $N_1 = \sum_{i=1}^N F_i$ — количество женщин в выборке.

Во всех методах, основанных на регрессиях, неявным образом присутствует еще одна проблема, связанная с обеспечением сопоставимости сравниваемых мужчин и женщин. Технически она возникает из-за отсутствия пересечения множеств характеристик мужчин и женщин, что происходит, если совместные распределения характеристик у мужчин и женщин сильно различаются между собой. В результате невозможно или сложно найти мужчин и женщин с одинаковыми значениями одновременно по всем важным производительным характеристикам. Если для некоторых представителей одной группы не удастся найти идентичных по характеристикам представителей другой группы, то такая проблема в англоязычной литературе называется *lack of common support*, или *lack of overlap*. Эта проблема может привести к серьезному смещению оценок рассматриваемых эффектов. Если представителей с идентичными характеристиками очень мало, то возникает проблема *thin common support*, последствия которой сказываются прежде всего на эффективности оценок. При использовании регрессионных методов проблема *common support* «преодолевается» на чисто техническом уровне: исследователи имплицитно экстраполируют зависимости, наблюдаемые для значений характеристик одной группы, на значения характеристик другой группы. Возможность подобной экстраполяции базируется на очень сильных предположениях о линейности взаимосвязи и ее соблюдении за пределами области *common support*. В контексте гендерного разрыва наибольшие опасения вызывают те характеристики и их комбинации, по которым наблюдается значительная гендерная сегрегация.

Следующий используемый нами метод — обратное вероятностное взвешивание (*inverse probability weighting, IPW*) (Hirano et al., 2003). При использовании метода IPW расчеты также ведутся в два шага. На первом шаге оценивается индекс склонности (*propensity score*), характеризующий условную вероятность подвергнуться воздействию — в нашем случае это вероятность быть женщиной при имеющихся у индивида наблюдаемых характеристиках. Для оценивания индекса склонности используется бинарная логит- или пробит-функция:

$$p(X_i) = \Pr(F_i = 1|X_i) = F(X_i\gamma).$$

Мы используем пробит-функцию.

На втором шаге рассчитываются ожидаемые значения индекса склонности ($\hat{p}(X_i)$) для каждого респондента. Все $\hat{p}(X_i)$ принимают значения от 0 до 1, поскольку строятся на базе бинарной модели.

Далее необъясненный гендерный разрыв рассчитывается по следующей формуле:

$$\hat{\delta}_{IPW} = \frac{1}{N_1} \sum_{i=1}^N F_i Y_i - \sum_{i=1}^N \hat{W}_i^0 Y_i. \quad (8)$$

Веса индивидуальных наблюдений для расчета АТЕТ равны 1 для группы воздействия (женщины) и $\hat{p}(X_i)/(1 - \hat{p}(X_i))$ для контрольной группы (мужчины). Таким образом, мужчины с нетипичными для своего пола характеристиками (имеющие высокие расчетные значения индекса склонности) получают более высокие значения весов. В результате взвешивания создается «искусственная выборка», в которой наблюдаемые характеристики распределены одинаково среди мужчин и женщин. Различия в заработных платах во взвешенных группах мужчин и женщин не могут быть «списаны» на объективные различия в наблюдаемых характеристиках, а представляют собой необъясненный гендерный разрыв. Важно, однако, отметить, что индекс склонности учитывает только наблюдаемые характеристики, ненаблюдаемые характеристики не учитываются — точнее учитываются, но лишь в той мере, в какой они коррелированы с наблюдаемыми характеристиками.

В отличие от линейной регрессии и декомпозиции Оаксаки — Блайндера метод IPW не требует линейной формы зависимости между ковариатами и заработной платой и не ограничивает вариацию необъясненного разрыва по группам работников.

Основную сложность при реализации методов, основанных на использовании индекса склонности, представляет выбор спецификации для индекса склонности. В этом разделе мы приведем общие рекомендации по отбору ковариат, а их применение в отношении гендера обсудим в следующем разделе. Во-первых, в уравнение индекса склонности следует включать только те ковариаты, которые имели место до начала воздействия. Не следует включать ковариаты, возникшие после воздействия или являющиеся результатом воздействия. Во-вторых, для минимизации смещения и сокращения дисперсии в оценке эффекта следует отдавать приоритет включению ковариат, тесно связанных с исходом (Y). В-третьих, следует избегать включения ковариат, ведущих себя как инструментальные переменные, то есть оказывающих влияние на исход только через переменную воздействия и не имеющих самостоятельного влияния на исход. В-четвертых, хорошими кандидатами для включения в модель являются полиномы включенных ковариат и пересечения между ковариатами, особенно если таковые имеют обоснование в теории изучаемого явления.

Метод IPW отличается от «классических» методов мэтчинга, в которых индекс склонности напрямую используется для подбора контрольной группы. К таким «классическим» методам мэтчинга относится метод ближайшего соседа или N ближайших соседей, радиусный мэтчинг, мэтчинг с использованием метрики Махаланобиса, мэтчинг по ядрел-функции и др. Их основным достоинством является возможность преодолеть «проклятие размерности» (*curse of dimensionality*), поскольку вместо мэтчинга по множеству переменных осуществляется мэтчинг по одной переменной — индексу склонности⁹.

⁹ «Проклятие размерности» возникает при точном мэтчинге. Его суть состоит в том, что при увеличении числа переменных для мэтчинга в группе воздействия увеличивается число наблюдений, для которых ока-

Различные варианты мэтчинга по индексу склонности длительное время использовались при оценивании, однако в последние годы все чаще звучит критика этого подхода. King, Nielsen (2019) показывают, что использование индекса склонности для прямого мэтчинга чревато произвольностью в выборе модели, несопоставимостью между группами в наблюдаемых характеристиках (дисбаланс ковариат) и, как следствие, усилением смещения в оценках эффектов. Перечисленные проблемы наиболее вероятны в выборках сравнительно небольшого размера, к которым относятся и годовые выборки РМЭЗ ВШЭ. Метод IPW позволяет ослабить некоторые проблемы, поскольку индекс склонности используется в нем не для прямого мэтчинга, а для перевзвешивания выборки. Однако риск неверной спецификации индекса склонности по-прежнему сохраняется. Поэтому при использовании IPW требуется внимательно подходить к предварительной диагностике моделей, в частности, анализировать баланс ковариат в исходных и взвешенных данных. С этой целью производится расчет стандартизированных различий и соотношения дисперсий между группой воздействия и контрольной группой по всем наблюдаемым характеристикам до и после взвешивания. По результатам диагностики может быть принято решение об изменении спецификации модели для индекса склонности: изменении функциональной формы, добавлении переменных или пересечений между переменными.

Еще одним потенциальным недостатком обратного вероятностного взвешивания является нестабильность оценок при наличии наблюдений, для которых вероятность подвергнуться воздействию близка к единице (Khan, Tamer, 2010). В этом случае резко увеличиваются стандартные ошибки полученных оценок эффекта воздействия. Для решения этой проблемы Lechner, Strittmatter (2019) предлагают цензурировать веса, устанавливая верхнюю границу для значений индексов склонности, используемых для расчета размера эффекта.

Еще один метод из семейства IPW повышает шансы на получение робастных оценок разрыва. Это метод корректировки регрессии с использованием обратного вероятностного взвешивания (*inverse probability weighted regression adjustment, IPWRA*). IPWRA представляет собой смесь из двух методов: линейной регрессии и обратного вероятностного взвешивания. Линейная регрессия используется для моделирования исходов (в нашем случае — заработных плат), а модели бинарного выбора — для моделирования индекса склонности.

Расчет необъясненного гендерного разрыва методом IPWRA ($\hat{\delta}_{IPWRA}$) ведется следующим образом (три первых шага совпадают с методом IPW):

1. Сначала оценивается логит- или пробит-регрессия для индекса склонности (условной вероятности быть женщиной с учетом наблюдаемых характеристик).
2. На основе этой модели рассчитываются ожидаемые значения индекса склонности для каждого респондента (и для мужчин, и для женщин).
3. Рассчитываются IPW-веса, которые для расчета АТЕТ равны 1 для наблюдений из группы воздействия (женщины) и $\hat{p}(X_i)/(1 - \hat{p}(X_i))$ для наблюдений из контрольной группы (мужчины).

зывается невозможным найти сопоставимые контрольные (не подвергшиеся воздействию) пары с точными совпадениями по всем переменным. Иллюстрация проклятия размерности представлена в следующем разделе при обсуждении условия *common support*.

4. Оценивается линейная регрессия для исходов (заработных плат) с взвешиванием наблюдений на IPW-веса.

Основным достоинством метода IPWRA считается так называемая двойная робастность по отношению к неверной спецификации уравнения заработной платы и по отношению к неверной спецификации индекса склонности. Он позволяет получить состоятельные оценки в случае неверной спецификации одного из уравнений, если второе уравнение специфицировано правильно. Однако если оба уравнения специфицированы неверно, то оценки эффекта все равно окажутся смещенными¹⁰.

5.2. Модели с двумя переменными воздействия

Meera et al. (2020) предлагают рассматривать формирование гендерного разрыва как результат нескольких одновременных воздействий и использовать для идентификации разрыва модели с несколькими уровнями воздействия (Cattaneo, 2010; Cattaneo et al., 2013). При использовании гендера как единственной переменной воздействия нам придется включить в спецификацию в качестве ковариат все остальные переменные, влияющие на заработную плату. Поэтому итоговый результат показывает «чистое» влияние гендера. Подобный способ построения контрольной группы является одним из достоинств методов, основанных на индексах склонности, но он не позволяет учесть более сложные объяснения причин формирования гендерного разрыва в оплате труда. Например, бюджетники получают более низкие заработные платы из-за особенностей формирования заработной платы в бюджетном секторе, а среди женщин значительно выше доля занятых в этом секторе. Аналогичным образом уровень заработной платы в профессии нередко зависит от доли женщин в этой профессии. При моделировании воздействия с одной переменной не удастся в полной мере учесть все важные аспекты уязвимого положения женщин на рынке труда. Использование двух переменных воздействия дает возможность проанализировать пересечения между гендером и другими значимыми характеристиками женской занятости, такими как гендерная сегрегация и гендерные роли.

Многоуровневое воздействие может моделироваться как комбинация из двух бинарных воздействий. Например, если в качестве одного воздействия рассматривается гендер, а в качестве второго — сектор занятости, то итоговое воздействие будет включать четыре уровня:

- уровень воздействия 0: female = 0, sector = 0 (мужчины, занятые в частном секторе);
- уровень воздействия 1: female = 0, sector = 1 (мужчины, занятые в государственном секторе);
- уровень воздействия 2: female = 1, sector = 0 (женщины, занятые в частном секторе);
- уровень воздействия 3: female = 1, sector = 1 (женщины, занятые в государственном секторе).

Для идентификации эффектов многоуровневых воздействий применяется метод IPWRA, описанный в предыдущем разделе. IPWRA позволяет получить эффекты воздействия для каждой из переменных воздействия и от их совместного воздействия. Это свой-

¹⁰ Для расчетов использовались команды `teffects ipw` и `teffects ipwra` для STATA.

ство очень важно для оценки воздействия пола в комбинации с родительством, сектором занятости, занятостью в женских профессиях и в мужских отраслях.

5.3. Оценки методом LASSO

В последние годы при эконометрическом оценивании все чаще используются методы машинного обучения. Эти методы позволяют выбирать спецификацию и параметры модели исходя из самих данных, а не из заданных извне предпосылок. В машинном обучении выбор спецификации чаще всего делается при помощи методов регуляризации, и одним из популярных подходов к понижению размерности вектора ковариат является LASSO (Least Absolute Shrinkage and Selection Operator). LASSO применяется для борьбы с «избыточностью» данных, когда ковариаты коррелируют между собой, что ведет к мультиколлинеарности и, как следствие, к нестабильности оценок коэффициентов многомерной линейной регрессии. Малозначимые и дублирующие друг друга переменные могут быть удалены без ухудшения объясняющей способности модели. Более того, упрощение модели может улучшить ее интерпретируемость.

В методе LASSO на коэффициенты накладывается дополнительное условие, которое штрафует оцениваемую модель за избыточную сложность. Если в классическом МНК для расчета оценок коэффициентов минимизируется сумма квадратов ошибок:

$$\frac{1}{2N} (y - X\beta')'(y - X\beta'),$$

то в LASSO минимизируется следующее выражение:

$$\frac{1}{2N} (y - X\beta')'(y - X\beta') + \lambda \sum_{j=1}^K |\beta_j|,$$

где λ — параметр регуляризации, имеющий смысл штрафа за сложность.

Введением второго слагаемого в целевую функцию достигается некоторый компромисс между ошибкой регрессии и размерностью вектора ковариат. В ходе минимизации некоторые коэффициенты становятся равными нулю. Если $\lambda = 0$, задача сводится к МНК и ни один коэффициент не равен в точности нулю. При увеличении λ «обнуляется» все большее число коэффициентов, соответствующие переменные исключаются из модели, и модель становится более лаконичной. В статистической теории было разработано несколько разных способов для определения оптимальной величины λ . В данной работе, за исключением декомпозиции Оаксаки —Блайндера, используется метод подстановки (*plug-in*). Это самый простой и быстрый способ расчета. Поскольку в нашей работе методы машинного обучения играют роль теста на спецификацию по отношению к другим используемым методам, то мы не будем фокусироваться на подборе методов машинного обучения, максимально подходящих для наших данных. Это задача отдельного исследования.

Беллони, Черножуков и Хансен разработали методы LASSO, которые можно использовать для статистического вывода и причинно-следственного анализа (Belloni et al., 2013; 2014). При помощи этих методов может быть рассчитана оценка коэффициента и стандартные ошибки в той ситуации, когда известна только одна или несколько переменных интереса, принадлежащих истинной модели. При этом истинный процесс, генерирующий данные, в точности неизвестен, однако в имеющемся наборе данных есть множество других переменных, которые были сгенерированы тем же самым процессом. Эта информация может быть использована для идентификации. Процесс оценивания идет

итеративно и включает оценивание множества LASSO-регрессий, поэтому финальные оценки коэффициентов при переменных интереса оказываются робастны по отношению к неизбежным ошибкам в спецификации модели.

Для идентификации эффекта гендера в линейной регрессии ($\hat{\delta}_{LRM,lasso}$) в данной работе используется процедура двойного отбора (*double selection*), предложенная в упомянутых выше работах Беллони, Черножукова и Хансена. Процедура двойного отбора включает в себя три этапа. На первом этапе при помощи описанного выше обычного метода LASSO оценивается регрессия для переменной Y с максимально полным набором ковариат, но без переменной воздействия. В нашем случае оценивается LASSO-регрессия для логарифма заработной платы с включением всех ковариат, которые входили в наши расширенные спецификации (с отраслями и профессиями), плюс все пересечения между ковариатами, но без переменной гендера. На этом этапе у большинства переменных окажутся нулевые коэффициенты. Ненулевые коэффициенты останутся лишь при тех переменных, которые имеют наиболее сильную корреляцию с заработной платой. На втором этапе оценивается LASSO-регрессия с полным набором ковариат и их пересечений, где зависимой переменной является переменная воздействия (гендер). Опять же большинство коэффициентов в этой регрессии окажутся нулями. Ненулевые коэффициенты останутся лишь при тех переменных, которые имеют наиболее сильную корреляцию с гендером. На третьем этапе объединяются множества переменных с ненулевыми коэффициентами на первом и втором этапах и методом наименьших квадратов оценивается финальная регрессия для логарифма заработной платы с включением переменной гендера и объединенного множества переменных (\tilde{X}_i):

$$Y_i = \alpha_{lasso} + F_i \delta_{LRM,lasso} + \tilde{X}_i \beta + \varepsilon_i. \quad (9)$$

В работе Belloni et al. (2013) показано, что оценка $\hat{\delta}_{LRM,lasso}$ является состоятельной и асимптотически нормальной. Процедура двойного отбора обладает свойством двойной робастности, то есть состоятельные оценки параметров интереса могут быть получены, если правильно специфицировано хотя бы одно из двух уравнений (для заработной платы или для гендера).

Для декомпозиции Оаксаки — Блайндера мы оценили классическую LASSO-регрессию для заработных плат по подвыборке мужчин. Поиск оптимальной величины λ осуществлялся при помощи метода кросс-валидации. На основе полученного уравнения были рассчитаны прогнозные значения заработных плат для женщин. Эти значения показывают, сколько получали бы женщины, если бы работодатели оценивали производительные характеристики женщин так же, как производительные характеристики мужчин. Усредненный остаток — разность между фактическими и прогнозными заработными платами женщин — и представляет собой необъясненный гендерный разрыв в оплате труда ($\delta_{OB,lasso}$).

Для методов, использующих индекс склонности, мы оцениваем один вариант LASSO. Этот вариант базируется на методе расширенного обратного вероятностного

взвешивания (*augmented inverse probability weighting, AIPW*)¹¹. Данный метод по смыслу похож на IPWRA. Он также совмещает корректировку регрессии и обратное вероятностное взвешивание, поэтому также обладает свойством двойной робастности. При использовании данного метода также оценивается уравнение для индекса склонности и рассчитываются IPW-веса. Однако далее схема несколько меняется: оцениваются уравнения регрессии для исходов (заработных плат мужчин и женщин), рассчитываются ожидаемые значения заработных плат для каждого наблюдения по каждому из исходов. На финальном этапе находятся взвешенные значения заработных плат с использованием IPW-весов, на базе которых рассчитываются эффекты воздействия. Для LASSO-AIPW необходимы дополнительные поправки на ошибки регуляризации. Расчет АТЕТ для LASSO-AIPW реализован в пакете STATA в соответствии с алгоритмом, описанным в Chernozhukov et al. (2018).

6. Результаты

6.1. Анализ сопоставимости характеристик мужчин и женщин

Как отмечалось в предыдущем разделе, важной предпосылкой многих методов идентификации гендерного разрыва является условие сопоставимости характеристик мужчин и женщин (*common support*). Поэтому в качестве предварительного шага проведем анализ выполнимости этого условия на данных РМЭЗ ВШЭ.

Выводы относительно *common support* зависят от той последовательности, в которой добавляются характеристики. Необходимость выбора последовательности связана с «проклятием размерности»: чем больше характеристик, по которым мы хотим обеспечить *common support*, тем выше вероятность получить пустые ячейки. Выборка сопоставимых наблюдений (мужчин и женщин, имеющих «двойников» с идентичными наборами характеристик среди лиц противоположного пола) будет сокращаться по мере добавления дополнительных переменных. Поэтому сначала мы определим «рейтинг» важности наблюдаемых характеристик с точки зрения влияния на заработные платы, а затем будет оценивать их воздействие на *common support*, включая в порядке приоритетности.

«Рейтинг» важности переменных строится по выборке мужчин как группы не подверженной дискриминации и прочим искажениям на рынке труда, влияющим на гендерный разрыв в заработных платах. Мы оцениваем для них уравнение заработной платы $\mu_0(x)$ и определяем вклад различных переменных в объяснение вариации заработных плат мужчин. Этот вклад измеряется изменением скорректированного R^2 при последовательном исключении того или иного блока переменных, например, всех переменных образования. Таким образом сначала мы определяем наиболее важные переменные с точки зрения их влияния на заработные платы мужчин, а затем двигаемся к менее важным. На этом шаге анализ ведется на выборке данных, объединенных за все годы с 2005 по 2020 г., поскольку цель данного упражнения — выявить переменные, которые сильнее всего коррелированы с заработными платами мужчин. Годовые дамми-переменные остаются во всех уравнениях. Уравнения оценивались методом наименьших квадратов.

¹¹ Данный метод используется в команде `telasso` в STATA.

В исходную спецификацию уравнения заработной платы включены образование, ожидаемый общий стаж работы и квадрат этой переменной, специфический стаж и квадрат этой переменной, отраслевая и профессиональная принадлежность, продолжительность рабочего времени, семейное положение, наличие детей до 18 лет, возраст, сектор занятости, профессия, отрасль, размер предприятия, тип населенного пункта, регион, годовые дамми-переменные (см. описание переменных в разделе 2).

Итоговая последовательность исключения переменных и соответствующие им значения скорректированного R^2 представлены в табл. 7. В нашем исследовании эти расчеты носят вспомогательный и чисто технический характер, однако они демонстрируют, какие переменные сильнее всего коррелируют с заработными платами мужчин. На первом месте оказывается отраслевая принадлежность, после нее следуют географические переменные и лишь затем переменные, относящиеся к человеческому капиталу и характеристикам рабочих мест.

Последовательность исключения переменных из МНК-уравнений определяет порядок добавления переменных при проверке условия *common support*. Для проверки *common support* мы переходим к данным за 2005, 2010, 2015 и 2020 гг., по которым ведутся все дальнейшие расчеты необъясненного гендерного разрыва. Поскольку среди объясняющих переменных есть как категориальные, так и непрерывные переменные (две переменные стажа), то мы не можем применить точный мэтчинг, как в статье Норо (2008). Поэтому мы используем метод приближенно-точного мэтчинга (*coarsened exact matching*), с помощью которого можно найти точные соответствия по категориальным переменным и приближенные соответствия по непрерывным переменным. Алгоритм такого мэтчинга был предложен в работе Iacus et al. (2012) и реализован для пакета STATA в виде пользовательской команды *sem* (Blackwell et al., 2009). С помощью этой команды мы рассчитали количество мужчин и женщин, имеющих сопоставимые характеристики, и долю женщин, для которых в выборке нашлись мужчины с сопоставимыми характеристиками. Далее были рассчитаны нескорректированный¹² и скорректированный гендерные разрывы по наблюдениям, удовлетворяющим условию *common support*. Результаты расчетов представлены в табл. 8 и на рис. 5.

Предварим содержательный анализ *common support* несколькими техническими комментариями к табл. 8. В этой таблице показано, как изменяется число наблюдений, удовлетворяющих условию *common support* при последовательном добавлении характеристик (см. колонки “Matched”). На некоторых шагах сокращается общее число наблюдений по мужчинам и женщинам (см. колонки “All”), что связано с пропущенными значениями по наблюдаемым характеристикам. Мы видим, что пропущенные значения не ведут к существенному истощению выборки. В последней колонке показано суммарное количество мужчин и женщин, имеющих сопоставимых «двойников» среди лиц противоположного пола. Именно такое количество наблюдений участвовало в расчете нескорректированного и скорректированного гендерного разрывов. На последней итерации выборка сократилась до 18 наблюдений с исходных 15 558 наблюдений. Для расчета нескорректированного разрыва строилась регрессия, в которой зависимой переменной был логарифм

¹² «Нескорректированный» гендерный разрыв скорректирован на годовые эффекты, чтобы устранить влияние тренда (см. пояснения ниже по тексту).

часовой заработной платы, а независимыми переменными пол и годовые дамми-переменные. Для расчета скорректированного разрыва эта регрессия расширялась за счет добавления всех ковариат, включенных в анализ *common support* на соответствующем шаге. Кроме того, при оценивании регрессий для скорректированного разрыва использовались веса, рассчитанные на основе алгоритма приближенно-точного мэтчинга. Совместное использование этих весов и включение в регрессию ковариат, по которым велся мэтчинг, позволяет добиться баланса характеристик мужчин и женщин в выборке наблюдений, удовлетворяющих условию *common support*.

Наши расчеты показывают, что *common support* соблюдается в полном объеме по двум важнейшим детерминантам заработной платы: отрасль и тип поселения. Добавление к этим переменным федерального округа также почти не сказывается на результатах: для 95% женщин находятся «двойники» среди мужчин в обследованиях за соответствующие годы. Сопоставимость выборок мужчин и женщин резко сокращается — до 44% — при добавлении профессий. При добавлении образования в выборке остается лишь около четверти сопоставимых наблюдений. На последнем шаге «двойники» находятся лишь для 0,1% женщин.

Показательно, что оценки как нескорректированного, так и скорректированного гендерного разрыва мало чувствительны к изменению определения *common support*. Обе оценки резко снижаются и становятся статистически незначимыми лишь на последних шагах анализа. При этом скорректированный разрыв лишь ненамного меньше по своей абсолютной величине нескорректированного разрыва. Более того, оба разрыва имеют схожие траектории изменения при добавлении дополнительных характеристик в *common support*. Некоторые видимые различия в траекториях наблюдаются лишь при добавлении отрасли, типа поселения и размера предприятия. Таким образом, даже в узких группах предельно сопоставимых работников мы видим один и тот же гендерный разрыв в оплате труда.

6.2. Результаты для моделей с одной переменной воздействия (гендером)

В табл. 9 представлены результаты оценивания гендерного разрыва с помощью регрессионных методов. Чтобы результаты не зависели от пропущенных значений в некоторых переменных, все модели оцениваются по выборке, в которой есть полная информация по всем переменным, включенным в максимально расширенную спецификацию (с переменными для отраслей и профессий). Поэтому полученные значения нескорректированных разрывов (различия в средних заработных платах) несколько отличаются от показанных на рис. 2, но общий тренд на сокращение гендерного разрыва в оплате труда остается неизменным и в рассматриваемой выборке.

Учитывая важность отраслевой и профессиональной сегрегации, результаты показаны для спецификаций, включающих и не включающих переменные отраслевой и профессиональной принадлежности. Их включение действительно существенно влияет на результаты оценивания и даже на качественные выводы о природе гендерного разрыва в оплате труда. Так, согласно оценкам линейной регрессии до включения переменных для отраслей и профессий скорректированный гендерный разрыв заметно больше по абсолютной величине, чем простая разница в средних заработных платах для всех рассматриваемых лет. Если бы эти результаты были верны, то это бы означало, что с учетом

наблюдаемых характеристик и при условии равенства отдач от этих характеристик женщинам недоплачивают на 4–6 п. п. больше, чем показывает сравнение средних заработных плат. Добавление в спецификацию переменных для отраслей и профессий ведет к тому, что скорректированный разрыв, во-первых, заметно снижается (на 2–7 п. п.) и, во-вторых, становится ниже нескорректированного для первой половины рассматриваемого периода. Таким образом, гендерная сегрегация оказывается важным фактором даже для моделей с наиболее жесткими предпосылками.

Оценки разрыва для линейной регрессии, полученные методом LASSO ($\hat{\delta}_{LRM,lasso}$), меньше по абсолютной величине, чем в базовой и расширенной спецификациях, и — для большинства рассматриваемых лет — меньше, чем разница в средних заработных платах. Этот результат говорит о том, что одной из причин завышенных оценок в МНК-регрессиях являются ошибки спецификации. Можно подобрать спецификацию, при которой даже в предположении о равенстве отдач наблюдаемые характеристики могут частично объяснить наблюдаемый гендерный разрыв.

В декомпозиции Оаксаки — Блайндера ослабляется условие о равенстве отдач от наблюдаемых характеристик. Это практически всегда ведет к снижению абсолютной величины разрыва, что видно, если сравнить между собой $\delta_{LRM,1}$ и $\hat{\delta}_{OB,1}$ либо $\delta_{LRM,2}$ и $\hat{\delta}_{OB,2}$. Мы видим, что важность гибкости при моделировании отдач, по крайней мере, не сокращается с 2005 по 2020 г., а скорее всего, даже увеличивается, особенно для расширенной спецификации. Как в модели линейной регрессии, в декомпозиции Оаксаки — Блайндера оценки разрыва сокращаются при добавлении в спецификацию переменных для отраслей и профессий. Без этих переменных скорректированный (необъясненный) разрыв был выше исходной нескорректированной величины на 1,5–2,5 п. п., после их добавления скорректированный разрыв сокращается на 1,5–2 п. п. по сравнению с исходной величиной, а в 2020 г. эта разница увеличивается до 5,3 п. п. Из табл. 10, в которой представлены детализированные результаты декомпозиции Оаксаки — Блайндера, видно, что сокращение необъясненной части разрыва в 2020 г. было обусловлено ростом относительных отдач женщин по блоку «Характеристики рабочих мест» и «Сегрегация». При этом «Сегрегация» по-прежнему негативно — и с усилением негативного воздействия — влияла на заработки женщин через распределение по отраслям и профессиям, отраженное в объясненной части разрыва. В совокупности это означает, что женщины все больше концентрируются в низкооплачиваемых отраслях и профессиях, при этом внутри этих отраслей и профессий заработки мужчин и женщин становятся все более схожими. Еще одно интересное наблюдение можно извлечь из табл. 10. Это резкий рост в 2020 г. негативного влияния на заработки женщин демографического фактора, прежде всего, за счет необъясненной части. Возможно, этот рост связан с более сильным воздействием пандемии на женщин с детьми. В любом случае это обстоятельство указывает на важность изучения пересечений между гендером и родительством и между гендером и возрастом, что будет сделано в следующем разделе.

В отличие от линейной регрессии использование методов LASSO не всегда позволяет сократить разрыв (см. панель для $\hat{\delta}_{OB,lasso}$). Поэтому мы делаем вывод о том, что в случае декомпозиции Оаксаки — Блайндера расширенная спецификация демонстрирует хорошие результаты и, по-видимому, неплохо аппроксимирует ту информацию о процес-

се генерации данных (то есть о процессе формирования заработных плат), которая отражается в имеющихся наблюдаемых переменных.

Теперь переходим к методам, базирующимся на индексе склонности: методу обратного вероятностного взвешивания (IPW) и методу корректировки регрессии с использованием обратного вероятностного взвешивания (IPWRA). Выбор переменных для включения в уравнение для оценки индекса склонности всегда является непростой задачей, но в случае, если гендер рассматривается как воздействие, возникает ряд специфических проблем. В частности, не ясно как быть с рекомендацией включать в модель только те ковариаты, которые имели место *до начала* воздействия. Важную роль, на наш взгляд, играет разграничение между полом и гендером. Биологический пол как совокупность генетических, анатомических и физиологических характеристик начинает формироваться еще до рождения и фиксируется при рождении человека. Гендер — это социальное понятие, относящееся к социальным и культурным нормам, которые общество предписывает выполнять людям в зависимости от их биологического пола. Рождаясь представителями женского и мужского пола, люди учатся быть девочками и мальчиками, затем женщинами и мужчинами. Поведенческие черты и характеристики, приобретенные в процессе гендерной социализации, определяют гендерные роли. Гендерные роли оказывают значительное влияние на распределение домашних обязанностей внутри семьи, решение о выходе на рынок труда, желаемую продолжительность рабочего времени, выбор профессии и проч.

Используя методы, базирующиеся на индексе склонности, мы пытаемся частично деконструировать гендерные роли. Мы ставим задачу найти мужчин и женщин максимально схожих с точки зрения поведенческих характеристик, включая те, которые относятся к выбору профессии, отрасли, продолжительности рабочего времени, то есть те, которые характеризуют реализацию гендерных ролей на рынке труда. Оставшаяся часть различий будет отнесена к эффекту пола (включая влияние биологических различий, дискриминацию и отбор) и ненаблюдаемым аспектам гендерных ролей, которые влияют на заработки, но не были учтены в нашей модели. В такой постановке задачи акцент смещается на обеспечение баланса ковариат при построении индекса склонности. При этом мы готовы принять критику, что, например, женщина, выбирающая работу воспитателя в детском саду, действует в соответствии со своей традиционной гендерной ролью, в то время как мужчина, выбирающий ту же работу, действует вопреки традиционной гендерной роли. Возможно, следование норме и отклонение от нее вознаграждаются по-разному, что мы игнорируем. Отчасти этот вопрос поднимается в следующих разделах, где рассматриваются сложные воздействия из двух переменных.

Итак, в качестве исходной спецификации для индексов склонности мы рассматриваем набор ковариат, использовавшихся в расширенном уравнении заработной платы. Результаты диагностики модели представлены в табл. 11, где показаны значения стандартизированных разностей и соотношения дисперсий в исходной и взвешенной выборках. В идеале стандартизированные разности должны быть равны нулю, а соотношение дисперсий — близко к единице, что указывает на достижение хорошего баланса по соответствующей ковариате. Для подавляющего большинства ковариат взвешивание наблюдений привело к значительному улучшению баланса, то есть распределения характеристик

мужчин и женщин во взвешенных выборках обладают большим сходством, чем до взвешивания. Улучшение баланса заметно и в тех ситуациях, где наблюдалась значительная гендерная сегрегация (см., например, строительство или образование).

В табл. 11 приводятся значения статистики χ^2 и *p*-value для теста на сверхидентификацию, предложенного в работе (Imai, Ratkovic, 2014). Нулевая гипотеза теста состоит в том, что ковариаты сбалансированы. Для всех рассматриваемых лет мы не можем отвергнуть эту гипотезу: *p*-value не опускается ниже 0,429. Формальный тест проходят все рассматриваемые спецификации.

Графики распределения расчетных индексов склонности позволяют наглядно проверить, соблюдается ли условие *common support (overlap)* для индексов склонности в контрольной группе и группе подвергшихся воздействию. На рис. 6 представлены кернел-диаграммы плотности распределения расчетных индексов склонности для всех рассматриваемых лет. Для каждого респондента была рассчитана вероятность того, что он/она имеет женский пол исходя из его/ее наблюдаемых характеристик и коэффициентов пробит-модели. Мы видим, что на всех графиках для женщин значительное число предсказаний сосредоточено около единицы, а для мужчин — около нуля. Это плохой знак, который может свидетельствовать о проблемах с соблюдением условия *common support (overlap)*. Мы видим, что пересечение есть практически для всех значений индекса склонности, то есть наблюдаемая ситуация не означает полного отсутствия сопоставимости (*lack of common support*). Однако на концах распределения много значений из одной группы и мало значений из другой группы. Это ситуация слабой сопоставимости (*weak common support*). При расчете АТЕТ (а нас интересует именно этот эффект) опасность представляют значения с высокими расчетными значениями индекса склонности: соответствующим наблюдениям из контрольной группы будет придаваться очень большой вес, что приведет к нестабильности оценок и выразится в увеличении стандартных ошибок. Наблюдения по женщинам при расчете АТЕТ получают единичный вес, поэтому наличие большого количества наблюдений по мужчинам из левого хвоста распределения с маленькими весами не создает проблем.

Как обсуждалось в методологическом разделе, лучшим выходом из такой ситуации является цензурирование индекса склонности, другими словами, «обрезание» хвоста высоких значений. Чем больше наблюдений цензурируется, тем более эффективными становятся оценки эффекта воздействия. Однако «расплатой» за снижение стандартных ошибок является ограничение по обобщению результатов на всю генеральную совокупность. Без цензурирования результаты можно полностью обобщать на генеральную совокупность — в нашем случае на всех женщин, занятых по найму. Чем жестче цензурирование, тем больше требуется оговорок при обобщении результатов и формулировке выводов. Кроме того, цензурирование может привести к нарушению баланса ковариат. Для того чтобы продемонстрировать последствия цензурирования, мы последовательно установим потолок для расчетного индекса склонности на уровне 1 (без цензурирования); 0,99; 0,95 и 0,90.

Результаты оценивания представлены в табл. 12. Поскольку зависимой переменной на этом этапе анализа является часовая заработная плата, а не ее логарифм, то разрывы выражены сразу в процентах. Результаты по методу IPW без цензурирования дают мак-

симильные или очень близкие к максимальным оценки гендерных разрывов из всех ранее рассмотренных методов. Ближе всего они к результатам линейной регрессии в спецификации без отраслей и профессий, но при существенно больших стандартных ошибках. Высокие значения разрывов, скорее всего, указывают на то, что модель для индекса склонности была специфицирована неверно, несмотря на удовлетворительные значения статистик в тестах на баланс ковариат. В данном случае стоит доверять кернел-диаграммам на рис. 6. Большие стандартные ошибки свидетельствуют о нестабильности оценок, источником которой являются наблюдения с высокими значениями индекса склонности. Цензурирование (со всеми упомянутыми выше оговорками) потенциально может помочь в решении этих проблем.

Действительно, цензурирование ведет к снижению разрывов и уменьшению стандартных ошибок. Однако снижение разрывов происходит немонотонно: в некоторых случаях разрыв увеличивается и при более жестком цензурировании. Стандартные ошибки уменьшаются, но также немонотонно: при цензурировании выборка становится более однородной с точки зрения индексов склонности, но одновременно сокращается число наблюдений. Так, при ограничении индекса склонности потолком в 0,90 исключается около 20% наблюдений. В целом цензурирование не ведет к качественному улучшению результатов, а с учетом дополнительных ограничений на интерпретацию не выглядит как предпочтительная стратегия идентификации. Мы принимаем вывод, что модель для индекса склонности была исходно специфицирована неверно.

Методы с двойной робастностью и LASSO могут исправить эту ошибку. Мы начнем с IPWRA (см. табл. 12). Напомним, что данный метод позволяет получить состоятельные оценки, если правильно специфицировано хотя бы одно из уравнений: либо для заработной платы, либо индекса склонности. Результаты декомпозиции Оаксаки — Блайндера для спецификации с отраслями и профессиями позволяют верить, что эта спецификация уравнения заработной платы выглядит достаточно убедительно. Действительно, оценки необъясненного разрыва сокращаются по сравнению с IPW до цензурирования, особенно для 2020 г., что подтверждает гипотезу о неверной спецификации индекса склонности. В целом оценки разрыва довольно близки к расширенной версии декомпозиции Оаксаки — Блайндера, что неудивительно, поскольку в обоих случаях использовалась одна и та же спецификация для уравнений заработной платы, а индекс склонности специфицирован неверно и поэтому мало влиял на устранение смещений.

Наш финальный тест на спецификацию — LASSO-AIPW. В методе LASSO-AIPW спецификация всех трех уравнений — уравнение для индекса склонности и два уравнения заработной платы (для мужчин и для женщин) — подбирается исходя из данных. В результате мы можем сравнить спецификации, выбранные методом машинного обучения, с теми, что мы сформулировали самостоятельно исходя из теории и опыта предшествующих исследований. Список отобранных методом LASSO-AIPW переменных представлен в табл. 13. Включение переменной отмечено крестиком безотносительно от знака коэффициента. Для удобства мы подсветили светло-синим цветом те переменные, которые были отобраны как минимум три раза, и желтым цветом — переменные, отобранные два раза.

С точки зрения данных ключевую роль в определении «мужских» и «женских» рабочих

мест играют профессия и отрасль. Семейное положение, сектор занятости и общий стаж работы оказывают влияние во взаимодействии с другими переменными. Некоторую роль играют продолжительность рабочей недели и размеры предприятия. При этом образование является слабым дифференцирующим признаком, в отдельные годы в спецификациях возникает только среднее профессиональное образование. Отобранные LASSO-AIPW-спецификации успешно проходят тесты на баланс ковариат (см. *p-value* для теста *overid* в табл. 12), хотя для 2020 г. значение *p-value* снижается до 0,07.

На рис. 7 показано, в какой мере соблюдается условие *common support* для индексов склонности, рассчитанных методом LASSO-AIPW. Несмотря на то что на графиках заметно явное увеличение количества наблюдений в средней части шкалы, остаются значительные массы наблюдений на концах шкалы. Это означает, что сопоставимость группы воздействия и контрольной группы действительно улучшилась, но риски нестабильности результатов сохраняются. Именно это наблюдается в табл. 12. Мы видим резкое сокращение оценок разрыва по сравнению с IPWRA для 2005 г., потом их рост для 2010 г. и сближение с оценками по IPWRA в 2015 и 2020 гг. Для 2005 и 2010 гг. стандартные ошибки в LASSO-AIPW существенно выше, чем в IPWRA. Таким образом, использование LASSO дает полезные «подсказки» по формированию моделей, позволяет тестировать спецификации, но не может гарантировать достоверности и точности оценок. Возможно также, что объем годовых выборок РМЭЗ ВШЭ недостаточен для полноценного использования методов машинного обучения.

С точки зрения дальнейшего изложения важным выводом является то, что использование IPWRA позволяет получить устойчивые результаты.

6.3. Результаты для моделей с двумя переменными воздействия

В данном разделе мы рассматриваем формирование гендерного разрыва как результат двух одновременных воздействий и используем для идентификации разрыва модели с несколькими уровнями воздействия. Подобный подход позволяет учесть то обстоятельство, что женщины чаще, чем мужчины, оказываются в ситуации с множеством источников уязвимости. Кроме того, одни и те же факторы, например, родительство, могут совершенно по-разному воздействовать на мужчин и женщин.

Гендер и родительство

Родительство имеет разные последствия для мужчин и женщин. Если до рождения детей поведение мужчин и женщин на рынке труда складывается по схожим траекториям, то после рождения детей траектории расходятся. Заработные платы и уровень занятости мужчин при этом практически не меняются, когда мужчины становятся отцами, в то время как трудовые доходы женщин резко сокращаются при рождении первого ребенка и не возвращаются на прежний уровень даже в средне- и долгосрочной перспективе (Angelov et al., 2016; Kleven et al., 2019a; 2019b). Штраф за материнство складывается из сокращения предложения труда и снижения ставок заработной платы, в том числе из-за отказа женщин от карьеры и перехода на рабочие места с более низким уровнем оплаты, но с более удобным графиком работы и большим числом семейных льгот. Результаты предшествующих исследований по России неоднозначны. Lebedinski et al. (2020), используя метод событийного анализа (*event study*), показали, что рождение ребенка не сказывается

на предложении труда и заработках мужчин, но приводит к сокращению уровня занятости женщин, не затрагивая — в среднесрочной перспективе — ставок заработной платы. Другие авторы, используя обычные регрессии и метод инструментальных переменных, обнаруживают значительный штраф за материнство для женщин (Бирюкова, Макаренцева, 2017) и небольшую премию за отцовство среди мужчин (Аистов, 2013; Ощепков, 2020).

В этом разделе мы рассматриваем гендер и родительство как два одновременных воздействия. Родительство определяется как наличие хотя бы одного ребенка в возрасте 0–18 лет. Это очень широкий возрастной интервал, и фактически мы усредняем эффекты для родителей, имеющих разное число детей очень разного возраста. Поэтому результаты следует интерпретировать как относящиеся к средне- и долгосрочным эффектам родительства. Для более детального анализа хорошо было бы выделить группы по количеству и возрастам детей. Эта операция не представляет принципиальной технической сложности, однако упирается в ограничения, связанные с объемом годовых выборок РМЭЗ ВШЭ.

Комбинируя переменные гендера и родительства, мы получаем воздействие с четырьмя уровнями:

- уровень воздействия 0: female = 0, parent = 0 (мужчины без детей младше 18 лет);
- уровень воздействия 1: female = 0, parent = 1 (мужчины с детьми младше 18 лет);
- уровень воздействия 2: female = 1, parent = 0 (женщины без детей младше 18 лет);
- уровень воздействия 3: female = 1, parent = 1 (женщины с детьми младше 18 лет).

Далее оцениваются эффекты двух видов: абсолютные и относительные. Абсолютные эффекты рассчитываются по отношению к базовой группе с нулевым уровнем воздействия, то есть по отношению к мужчинам без детей. Относительные эффекты рассчитываются по сравнению с другими группами воздействия: женщины с детьми по сравнению с мужчинами с детьми (группа 3 vs группа 1); женщины с детьми по сравнению с женщинами без детей (группа 3 vs группа 2) и женщины без детей по сравнению с мужчинами с детьми (группа 2 vs группа 1). Оценивание производилось по методу IPWRA, спецификации уравнений индекса склонности и заработной платы такие же, как в разделе 6.2 за вычетом переменной наличия детей, поскольку теперь она выступает как переменная воздействия.

Результаты оценивания представлены в табл. 14. В таблице показаны различия в средних значениях часовых заработных плат и необъясненный гендерный разрыв, рассчитанный как АТЕТ по методу IPWRA. Необъясненный гендерный разрыв выражен в уровнях (в рублях) и в процентах. Поскольку данные по заработным платам дефлировались, то размеры разрывов можно сопоставлять между раундами, но удобнее сравнивать АТЕТ, выраженные в процентах. Стандартные ошибки полученных оценок достаточно велики, что, скорее всего, связано с небольшим объемом годовых выборок РМЭЗ ВШЭ. При моделировании эффектов множественных воздействий проблема недостаточности объема выборки встает острее, чем в моделях с одной переменной воздействия, поскольку здесь мы имеем дело с более дробными группами. Не исключено и влияние проблем, связанных со спецификацией уравнений, хотя, как мы видели в предыдущем разделе,

благодаря свойству двойной робастности IPWRA демонстрирует довольно устойчивые результаты на данных РМЭЗ ВШЭ.

Наши оценки указывают на то, что родительство — по крайней мере, в том виде как мы его определили, — не является значимым детерминантом гендерного разрыва. Во-первых, нам не удалось подтвердить существование штрафа за материнство: различия в заработках сопоставимых женщин с детьми и без детей не являются статистически значимыми ни в одном из раундов. Вместе с тем к интерпретации этого результата нужно подходить осторожно и помнить, что в нашей выборке присутствуют только женщины, занятые по найму. Возможно, эффект отбора ведет к самоотбору в занятость наиболее производительных женщин с детьми, тогда как менее производительные женщины с детьми остаются незанятыми и переходят в самозанятость. Во-вторых, наличие премии за отцовство также имеет слабые основания: различия в заработных платах сопоставимых мужчин с детьми и мужчин без детей значимы на 5%-м уровне лишь для 2010 г.

В то же время наблюдаются устойчиво большие различия в заработках при сравнении похожих групп: женщин, не имеющих детей, с мужчинами, не имеющими детей, и женщин, имеющих детей, с мужчинами, имеющими детей. При этом разрыв между группами без детей меньше, чем между группами с детьми, кроме 2005 г. Эти различия могут означать, что родительство все же ведет к ухудшению структуры занятости женщин, например, провоцируя перемещение в отрасли и виды деятельности с более низкими заработками, но с более благоприятными для семейной жизни условиями труда. Используя методы, основанные на мэтчинге и индексах склонности, мы сравниваем сопоставимых мужчин и женщин, включая сопоставимость по структуре занятости. Однако эти методы не «схватывают» структурные изменения, которые могут быть следствием рождения детей. Впрочем, МНК-регрессии также не в состоянии справиться с этой проблемой. Вопросы влияния рождения детей на отбор женщин в занятость и структуру занятости женщин нуждаются в дополнительном изучении.

Гендер и сектор занятости

Сектор занятости является еще одним важным источником гендерной сегрегации и, как было показано в разделе 4, усиливающим свое структурное влияние на занятость. Исследования межсекторного разрыва в оплате труда показывают, что работники государственного сектора по-прежнему получают меньше, чем сопоставимые по характеристикам работники частного сектора, однако разрыв существенно сократился после майских указов 2012 г., положивших начало реформе оплаты труда в бюджетном секторе (Gimpelson, Lukyanova, 2009; Gimpelson et al., 2019). Лукьянова (2021) указывает на высокую значимость самоотбора в бюджетный сектор. Важно понимать, что среди мужчин и женщин отбор может работать по-разному.

По аналогии с предыдущим разделом мы выделяем четыре уровня воздействия в зависимости от пола индивида и сектора занятости:

- уровень воздействия 0: female = 0, sector = 0 (мужчины, занятые в частном секторе);
- уровень воздействия 1: female = 0, sector = 1 (мужчины, занятые в государственном секторе);

- уровень воздействия 2: female = 1, sector = 0 (женщины, занятые в частном секторе);
- уровень воздействия 3: female = 1, sector = 1 (женщины, занятые в государственном секторе).

Оценивание производилось по методу IPWRA. Спецификации уравнений индекса склонности и заработной платы такие же, как в разделе 6.2 за вычетом переменной сектора занятости, поскольку теперь она выступает как переменная воздействия. Результаты расчетов АТЕТ представлены в табл. 15.

Сектор занятости действительно является важным детерминантом заработных плат как среди мужчин, так и среди женщин, но его влияние заметно снижалось на протяжении рассматриваемого периода. Сокращение межсекторных разрывов внутри групп мужчин и женщин видно уже на уровне простых различий в средних часовых заработках. К 2020 г. они незначимы даже на 10%-ном уровне. Однако рабочие места в государственном секторе в среднем требуют более высокого уровня образования и квалификации, поэтому информативнее анализировать условные разрывы, скорректированные на различия в характеристиках (АТЕТ). Эти разрывы также сократились. Если в 2005 г. мужчины, занятые в госсекторе, получали на 21% меньше, чем мужчины, занятые в частном секторе, то к 2020 г. этот разрыв уменьшился до 10%. Еще более радикальные изменения наблюдались среди женщин. Различия в заработках между женщинами, занятыми в частном и государственном секторах, сократились с 29% в 2005 г. до статистически незначимых 4% в 2020 г. Если в начале периода женщины имели более высокие штрафы за занятость в госсекторе, то в конце периода ситуация сменилась на обратную: в государственном секторе недоплачивали только мужчинам, и то существенно меньше, чем в середине 2000-х годов. Нельзя исключить, что этот разворот является временным и обусловлен пандемией, однако и в предшествующие годы тенденции развивались именно в этом направлении. По состоянию на 2020 г. мы констатируем, что сектор занятости из фактора, усиливающего гендерный разрыв, превратился в фактор, его сглаживающий.

Внутри секторов относительное положение женщин за рассматриваемый период улучшилось только в государственном секторе. В 2005 г. женщины, занятые в государственном секторе, получали в среднем на 32% меньше сопоставимых коллег-мужчин, к 2020 г. разрыв уменьшился до 18%. В частном секторе женщины получали на 30% меньше мужчин в 2005 г. и на 26% меньше в 2020 г. Одновременно увеличивался разрыв между женщинами, занятыми в частном секторе, и мужчинами, занятыми в государственном секторе: с 11 до 19%. Таким образом, государственный сектор по всем параметрам становился все более привлекательной опцией для женщин, что могло стать причиной наблюдавшегося усиления гендерной сегрегации по сектору занятости (см. рис. 8).

Гендер и занятость в женских профессиях

В разделе 4 было показано, что на протяжении рассматриваемого периода произошел заметный рост гендерной сегрегации по профессиям, прежде всего за счет усиления концентрации женщин в женских профессиях. В разделе 6.2 при рассмотрении спецификаций, отобранных методом LASSO, также отмечалось, что профессиональная принадлежность является одним из наиболее сильных предикторов разделения рабочих мест на мужские и женские. В данном разделе мы обращаемся к вопросу о том, является ли заня-

тость в женских профессиях дополнительным фактором, усиливающим гендерный разрыв в оплате труда. В обзоре литературы мы подробно обсуждали возможные механизмы, посредством которых профессиональная сегрегация может влиять на заработные платы. Это механизмы, связанные с инвестициями в человеческий капитал, компенсирующими различия и дискриминацию.

В данном разделе мы выделяем занятость в профессиях с доминирующей женской занятостью как отдельный вид воздействия. Женские профессии определяются как профессии, в которых доля женщин превышает 70%. Профессии, в которых доля женщин составляет 70% и ниже, определяются как неженские профессии. В эту группу входят смешанные профессии, профессии с доминирующей мужской занятостью, а также небольшое число профессий, которые мы не смогли классифицировать из-за малого числа наблюдений (см. подробнее раздел 4). К сожалению, объем годовых выборок РМЭЗ ВШЭ недостаточен, чтобы рассматривать отнесение к мужским и смешанным профессиям как отдельные воздействия. Отнесение профессий к группе женских профессий довольно стабильно на протяжении всех рассматриваемых раундов, однако отдельные профессии все же меняли свой статус на протяжении периода, что следует принимать во внимание при интерпретации результатов.

Комбинируя гендер и тип профессии по гендерному составу, получаем четыре уровня воздействия:

- уровень воздействия 0: $female = 0, female_occupation = 0$ (мужчины, занятые в неженских профессиях);
- уровень воздействия 1: $female = 0, female_occupation = 1$ (мужчины, занятые в женских профессиях);
- уровень воздействия 2: $female = 1, female_occupation = 0$ (женщины, занятые в неженских профессиях);
- уровень воздействия 3: $female = 1, female_occupation = 1$ (женщины, занятые в женских профессиях).

Оценивание производилось по методу IPWRA. Спецификации уравнений индекса склонности и заработной платы такие же, как в разделе 6.2, за вычетом переменных профессиональной принадлежности, поскольку теперь они учтены при формировании переменной воздействия. Результаты расчетов АТЕТ представлены в табл. 16.

Оценки имеют большие стандартные ошибки, поэтому не исключено, что какие-то существенные различия мы все-таки упускаем, поэтому мы упомянем и незначимые, но устойчивые различия. Мужчины, занятые в женских профессиях, стабильно получают более низкие заработные платы, чем сопоставимые мужчины, занятые в неженских профессиях. Однако значимые различия наблюдаются лишь для 2015 г., и то лишь на 10%-ном уровне. Аналогичным образом женщины, занятые в женских профессиях, стабильно получают более низкие заработные платы, чем сопоставимые женщины, занятые неженских профессиях. Различия значимы на 10%-ном уровне для 2015 и 2020 гг. Таким образом, мы наблюдаем некоторые признаки штрафа за занятость в женских профессиях. Заниженная оценка важности и престижа женских профессий может «работать» на усиление гендерного разрыва в оплате труда.

Гендерные разрывы внутри групп женских и неженских профессий имели схожие траектории: резкое снижение в 2010 г. по сравнению с 2005 г. и последующая стабилизация на этом уровне. Значения разрывов между группами также очень близки между собой в каждом конкретном году. Подобная динамика, скорее всего, отражает реальное снижение гендерных разрывов внутри профессиональных групп, а не изменение списка женских профессий. Всего одна профессия четко перестала быть женской за этот период (43 — Служащие в сфере обработки числовой информации и учета материальных ценностей), но переход случился в 2015 г., еще две группы, где доля женщин очень близка к 70%, в отдельные годы были отнесены к женским профессиям, в другие — к неженским, но без общего паттерна по времени. Резкое сокращение гендерного разрыва внутри профессиональных групп после 2005 г. могло быть связано как с общим сокращением (например, вследствие повышения минимальных заработных плат) либо с сокращением внутри крупных профессиональных групп. Этот вопрос требует дополнительного изучения. Однако даже после сокращения разрыв остается на уровне порядка 20%.

Гендер и занятость в мужских отраслях

Усиление поляризации мужской и женской занятости по отраслевому признаку является еще одной тенденцией последних десятилетий. Можно предположить, что изменение гендерного состава отдельных отраслей провоцировалось или сопровождалось изменениями в относительных заработных платах. В этом разделе мы выделяем в качестве воздействия занятость в мужских отраслях, поскольку в отраслевом разрезе маскулинизация доминировала над феминизацией в течение рассматриваемого периода. Из-за ограниченного объема выборки мы не можем выделить занятость в женских и смешанных отраслях как отдельные воздействия и рассматриваем их совместно. Отрасль считалась мужской, если доля мужчин в общей численности занятых превышала 70% (см. дескриптивные результаты в разделе 4).

Комбинируя гендер и занятость в мужских отраслях, получаем четыре уровня воздействия:

- уровень воздействия 0: $female = 0, male_industry = 0$ (мужчины, занятые в немужских отраслях);
- уровень воздействия 1: $female = 0, male_industry = 1$ (мужчины, занятые в мужских отраслях);
- уровень воздействия 2: $female = 1, male_industry = 0$ (женщины, занятые в немужских отраслях);
- уровень воздействия 3: $female = 1, male_industry = 1$ (женщины, занятые в мужских отраслях).

Оценивание производилось по методу IPWRA. Спецификации уравнений индекса склонности и заработной платы такие же, как в разделе 6.2, за вычетом переменных отраслевой принадлежности, поскольку теперь они учтены при формировании переменной воздействия. Результаты расчетов АТЕТ представлены в табл. 17.

Занятость в мужских отраслях действительно дает дополнительные денежные выгоды как мужчинам, так и женщинам. Для мужчин премия за занятость в мужской отрасли увеличивалась на протяжении рассматриваемого периода с 3% в 2005 г. до 19% в 2020 г. Для женщин оценки размера премии неустойчивы, в первую очередь потому, что очень

мало женщин в мужских отраслях, особенно в начале периода. К концу периода доля женщин, занятых в мужских отраслях, увеличивается, прежде всего за счет перехода части смешанных отраслей в разряд мужских, что не только сказывается на размере оценок, но и ведет к снижению стандартных ошибок. Несмотря на неустойчивость оценок прослеживается явная тенденцию к понижению премий за занятость в мужских отраслях для женщин: с 26% в 2005 г. до статистически незначимого уровня в 2020 г. Таким образом, маскулинизация занятости внутри отраслей сопровождалась увеличением премий для мужчин и ухудшением относительных заработков женщин.

Сходные процессы происходили в разрезе групп отраслей. В группе мужских отраслей наблюдалось увеличение гендерного разрыва: с 5% в 2005 г. до 25% в 2020 г. В немужских отраслях шел обратный процесс на выравнивание заработных плат мужчин и женщин: разрыв с исходных высоких значений в 32% в 2005 г. снизился до 21% в 2020 г. Подобная динамика должна была способствовать дальнейшей поляризации занятости по отраслевому признаку. Относительное положение женщин ухудшилось в мужских отраслях и улучшилось в немужских отраслях.

7. Заключение

В этой работе на данных РМЭЗ ВШЭ за 2005–2020 гг. изучались перспективы использования новых методологических подходов к анализу гендерного разрыва в оплате труда. Мы сравнивали между собой линейную регрессию, декомпозицию Оаксаки — Блайндера и ряд методов, основанных на расчете индексов склонности. При анализе рассматривались различные модификации указанных методов, включая полученные с использованием методов машинного обучения. Ключевым параметром при оценивании являлась общая величина необъясненного гендерного разрыва в оплате труда, в терминологии методов оценки воздействия — средний эффект воздействия для подвергшихся воздействию, где под «воздействием» понималась принадлежность к женскому полу. Поэтому мы не пытались оценить вклад конкретных наблюдаемых характеристик в формирование гендерного разрыва, как это делается в большинстве работ по российским данным. Основное внимание при сравнении различных методологических подходов уделялось тому, каким образом заложенные в них предпосылки могут влиять на результаты и в какой мере они способны учитывать гетерогенность выборок мужчин и женщин.

Мы выяснили, что выборки мужчин и женщин действительно крайне гетерогенны в том смысле, что распределения по целому ряду характеристик сильно различаются между собой. Наибольшие различия наблюдаются по тем характеристикам, по которым наблюдается сильная гендерная сегрегация: профессиональная и отраслевая принадлежность. При этом проблема становилась все более серьезной на протяжении рассматриваемого периода, поскольку оба вида сегрегации углублялись так же, как и гендерная сегрегация по признаку сектора и размеров предприятия. Гетерогенность выборок ведет к нарушению условия сопоставимости характеристик (*common support*), которое является явным или имплицитным требованием для адекватного использования различных методов оценивания.

Существующая гетерогенность делает невозможным использование методов точного (*exact matching*) и даже приближенно-точного (*coarsened exact matching*) мэтчинга либо

сильно ограничивает набор ковариат, по которым возможно проведение мэтчинга. Однако в последнем случае из анализа пришлось бы исключать многие значимые детерминанты заработных плат. Гетерогенность выборок создает проблемы и при использовании методов, основанных на индексах склонности, в частности метода обратного вероятностного взвешивания (*IPW*). Формальное достижение баланса ковариат при помощи перевзвешивания выборок и цензурирование больших значений индекса склонности не решает проблем, возникающих вследствие гетерогенности. Полученные оценки разрывов имеют большие стандартные ошибки, а цензурирование ограничивает обобщение результатов на всю генеральную совокупность. Проблема может быть отчасти сглажена при помощи методов с двойной робастностью и подключения методов машинного обучения. В данной работе показаны результаты использования корректировки регрессии с использованием обратного вероятностного взвешивания (*IPWRA*), расширенного обратного вероятностного взвешивания (*AIPW*) и методов *LASSO*. Эти модификации приводят к заметному повышению эффективности оценок, по крайней мере, в моделях с одной переменной воздействия. Вместе с тем мы столкнулись и с серьезными ограничениями использования новых методов оценивания применительно к данным РМЭЗ ВШЭ. Предпосылка о независимости и одинаковом распределении наблюдений (*i.i.d.*) не позволяет использовать объединенный массив данных по разным раундам этого панельного обследования, поэтому оценивание приходится вести по годовым выборкам. Методы, основанные на расчете индексов склонности, требуют больших объемов данных, что особенно остро проявляется при оценивании моделей с двумя переменными воздействия.

Неожиданным «открытием», сделанным в ходе сравнения различных методологических подходов применительно к данным РМЭЗ ВШЭ, стало неплохое поведение декомпозиции Оаксаки — Блайндера при условии верной спецификации уравнения заработной платы. Верная спецификация предполагает включение в уравнение переменных отраслевой и профессиональной принадлежности. Это не решает проблему гетерогенности выборок по существу, но в отношении данных РМЭЗ ВШЭ экстраполяция структурных зависимостей, идентифицированных для одной гендерной группы, на другую гендерную группу оказывается вполне приемлемой, особенно для 2010–2020 гг., когда в российской экономике устанавливается некоторая структурная стабильность.

Интересные результаты, хотя и с большими стандартными ошибками, были получены при моделировании воздействия с двум переменными. Применительно к гендерному разрыву в оплате труда этот подход позволяет включить в анализ пересечения между гендером и другими значимыми характеристиками женской и мужской занятости, такими как гендерная сегрегация и гендерные роли. Сегрегация и гендерные роли могут служить факторами, усиливающими гендерный разрыв, поскольку работа в «женской» профессии, «мужской» отрасли, брак и родительство могут оказывать разное влияние на мужчин и женщин. Содержательные аспекты этих взаимодействий обсуждаются ниже, с точки зрения методологии этот подход является более гибким и реалистичным по сравнению с моделированием гендера как единственного воздействия, когда вторая переменная используется в качестве ковариаты и де-факто предполагается гендерно-нейтральной. На наш взгляд, модели с двумя воздействиями имеют хорошие перспективы на больших выборках, например, на данных социально-экономических обследований, проводимых Росста-

том. Наряду с моделями с множественными воздействиями следует обратить внимание на активно развивающееся направление методов медиации, в которых оценивается прямое и косвенное воздействие гендера и которые имеют более сильную каузальную интерпретацию (см. (Huber, 2015)).

Проводя методологические изыскания, мы выяснили несколько интересных содержательных фактов о динамике гендерного разрыва на российском рынке труда. В течение рассматриваемого периода наблюдалось существенное сокращение общей величины гендерного разрыва и его необъясненной части. При этом необъясненная часть сокращалась быстрее, чем общий гендерный разрыв, что говорит об увеличении вклада наблюдаемых переменных в формирование гендерного разрыва. Среди наблюдаемых переменных, действующих в направлении сокращения разрыва, наиболее существенными оказываются гендерная сегрегация и демографические переменные. Несмотря на позитивные тенденции, по самым оптимистичным оценкам (IPWRA), на конец периода необъясненная часть составляла не менее 60% от общей величины гендерного разрыва. Примечательно, что необъясненный гендерный разрыв остается крайне стабильным даже внутри очень узких подгрупп населения, что было показано при анализе *common support* и при использовании моделей с двумя переменными воздействиями. Подобная стабильность указывает на то, что причины гендерного разрыва не связаны с действием механизмов отбора. Наиболее вероятной причиной гендерного разрыва остается дискриминация, что стыкуется с историческими сведениями и результатами других исследований по российскому рынку труда. Определенную роль могут играть гендерные различия в неденежных компонентах компенсации и в целом различия в оценках сравнительной важности денежных и неденежных мотиваций, а также различия в предпочтениях относительно содержательных аспектов работы.

Рассматривая модели с двумя воздействиями, мы выяснили, что родительство, определяемое как наличие детей в возрасте 0–18 лет, не является значимым детерминантом гендерного разрыва: нам не удалось обнаружить значимого штрафа за материнство или устойчивой премии за отцовство. Вклад сектора занятости в формирование гендерного разрыва радикально изменился на протяжении рассматриваемого периода: из фактора, усиливавшего гендерные различия в середине 2000-х годов, сектор занятости превратился в фактор сглаживания гендерного неравенства. В то же время были получены свидетельства в пользу заниженной оценки женского труда, которая выражается в штрафах за работу в женских профессиях и премий за работу в мужских отраслях. Из негативных трендов мы зафиксировали ухудшение относительного положения женщин в мужских отраслях.

Источники

- Аистов А.* Супружеская премия // Прикладная эконометрика. 2013. Т. 31. № 3. С. 99–114.
- Бирюкова С., Макаренцева А.* Оценки «штрафа за материнство» в России // Население и экономика. 2017. № 1(1). С. 50–70.
- Лукьянова А.* Что держит бюджетников на низкооплачиваемых рабочих местах? Роль отбора и некогнитивных факторов в объяснении межсекторных различий в оплате труда // Прикладная эконометрика. 2021. Т. 62. С. 32–53.
- Ощепков А.* Отцы и дети: «премия» за отцовство на российском рынке труда // Экономический журнал Высшей школы экономики. 2020. Т. 24. № 2. С. 157–190.
- Роцин С.Ю., Емелина Н.К.* Мета-анализ гендерного разрыва в оплате труда в России // Экономический журнал ВШЭ. 2022. Т. 26(2). С. 213–239.
- Angelov N., Johansson P., Lindahl E.* Parenthood and the gender gap in pay // Journal of Labor Economics. 2016. Vol. 34(3). P. 545–579.
- Anker R.* Gender and jobs: Sex segregation of occupations in the world. Geneva: International Labour Office, 1998.
- Bach P., Chernozhukov V., Spindler M.* Closing the US gender wage gap requires understanding its heterogeneity (arXiv preprint arXiv:1812. 04345). 2018. URL: <https://arxiv.org/abs/1812.04345>.
- Beller A.* Occupational segregation by sex: Determinants and changes // Journal of Human Resources. 1982. Vol. 20. P. 235–250.
- Belloni A., Chernozhukov V., Hansen C.* High-dimensional methods and inference on structural and treatment effects // Journal of Economic Perspectives. 2014. Vol. 28(1). P. 29–50.
- Belloni A., Chernozhukov V., Hansen C.* Inference on treatment effects after selection amongst high-dimensional controls (with an application to abortion and crime) // Review of Economic Studies. 2013. Vol. 81(2). P. 608–650.
- Blackwell M., Iacus S., King G., Porro G.* CEM: Coarsened exact matching in Stata // Stata Journal. 2009. Vol. 9(4). P. 524–546.
- Blau F., Simpson P., Anderson D.* Continuing progress? Trends in occupational segregation in the United States over the 1970s and 1980s // Feminist Economics. 1998. Vol. 4(3). P. 29–71.

- Blinder A.* Wage discrimination: reduced form and structural estimates // *Journal of Human Resources*. 1973. Vol. 8(4). P. 436–455.
- Böheim R., Stöllinger P.* (2021) Decomposition of the gender wage gap using the LASSO estimator // *Applied Economics Letters*. Vol. 28(10). P. 817–828.
- Cattaneo M.* Efficient semiparametric estimation of multi-valued treatment effects under ignorability // *Journal of Econometrics*. 2010. Vol. 155(2). P. 138–154.
- Cattaneo M., Drukker D., Holland A.* (2013) Estimation of multivalued treatment effects under conditional independence // *Stata Journal*. 2013. Vol. 13(3). P. 407–450.
- Chernozhukov V., Chetverikov D., Demirer M., Duflo E., Hansen C., Newey W., Robins J.* Double/debiased machine learning for treatment and structural parameters // *The Econometrics Journal*. 2018. Vol. 21(1). P. C1–C68.
- Duncan O., Duncan B.* A methodological analysis of segregation indexes // *American Sociological Review*. 1955. Vol. 20. P. 210–217.
- Fortin N., Lemieux T., Firpo S.* Decomposition methods // *Handbook of Labor Economics* / O. Ashenfelter, D. Card (eds). Vol. 4A. Amsterdam: Elsevier, 2011. P. 1–102.
- Gimpelson V.E., Lukyanova A.* Are public sector workers underpaid in Russia? Estimating the public-private wage gap. IZA Discussion Paper No. 3941. 2009. URL: <https://docs.iza.org/dp3941.pdf>.
- Gimpelson V., Sharunina A., Lukyanova A.* Economics and politics of the public-private wage gap (the case of Russia). IZA Discussion Paper No. 12247. 2019. URL: <https://docs.iza.org/dp12247.pdf>.
- Hakim C.* Segregated and integrated occupations: A new approach to analyzing social change // *European Sociological Review*. 1993. Vol. 9(3). P. 289–314.
- Hirano K., Imbens G., Ridder G.* Efficient estimation of average treatment effects using the estimated propensity score // *Econometrica*. 2003. Vol. 71(4). P. 1161–1189.
- Huber M.* Causal pitfalls in the decomposition of wage gaps // *Journal of Business & Economic Statistics*. 2015. Vol. 33(2). P. 179–191.
- Iacus S., King G., Porro G.* Causal inference without balance checking: Coarsened exact matching // *Political Analysis*. 2012. Vol. 20(1). P. 1–24.
- Imai K., Ratkovic M.* Covariate balancing propensity score // *Journal of the Royal Statistical Society, Series B*. 2014. Vol. 76. P. 243–263.
- Khan S., Tamer E.* Irregular identification, support conditions and inverse weight estimation // *Econometrica*. 2010. Vol. 78(6). P. 2021–2042.

- King G., Nielsen R.* Why propensity scores should not be used for matching? // *Political Analysis*. 2019. Vol. 27(4). P. 435–454.
- Kleven H., Landais C., Posch J., Steinhauer A., Zweimuller J.* Child penalties across countries: Evidence and explanations // *AEA Papers and Proceedings*. 2019. Vol. 109. P. 122–126.
- Kleven H., Landais C., Søgaard J.* Children and gender inequality: Evidence from Denmark // *American Economic Journal: Applied Economics*. 2019. Vol. 11(4). P. 181–209.
- Lebedinski L., Perugini C., Vladisavljevic M.* Child penalty in Russia: Evidence from an event study. IZA DP No. 13928. 2020. URL: <https://docs.iza.org/dp13928.pdf>.
- Lechner M., Strittmatter A.* Practical procedures to deal with common support problems in matching estimation // *Econometric Reviews*. 2019. Vol. 38(2). P. 193–207.
- Massey D., Denton N.* *American apartheid: Segregation and the making of the underclass*. Cambridge: Harvard University Press, 1993.
- Meara K., Pastore F., Webster A.* The gender pay gap in the USA: a matching study // *Journal Population Economics*. 2020. Vol. 33. P. 271–305.
- Ñopo H.* Matching as a tool to decompose wage gaps // *Review of Economics and Statistics*. 2008. Vol. 90(2). P. 290–299.
- Oaxaca R.* Male-female wage differentials in urban labor markets // *International Economic Review*. 1973. Vol. 14(2). P. 693–709.
- Oshchepkov A.* Gender pay gap in Russia: Literature review and new decomposition results // *Gendering Post-Soviet Space* / T. Karabchuk, K. Kumo, K. Gatskova, E. Skoglund (eds). Springer, 2021. P. 211–233.
- Pan J.* Gender segregation in occupations: The role of tipping and social interactions // *Journal of Labor Economics*. 2015. Vol. 33(2). P. 365–408.
- Strittmatter A., Wunsch C.* The gender pay gap revisited with big data: Do methodological choices matter? CESifo Working Paper No. 8912. 2021. URL: <http://dx.doi.org/10.2139/ssrn.3798933>.

Таблица 1. Используемая классификация занятий

Наша классификация	ISCO-08
10	11. Руководители высшего звена, высшие должностные лица и законодатели
	12. Управляющие в корпоративном секторе и в других организациях
	13. Руководители подразделений в сфере производства и специализированных сервисных услуг
	14. Руководители в гостиничном и ресторанном бизнесе, розничной и оптовой торговле и родственных сферах обслуживания
21	21. Специалисты в области науки и техники
	25. Специалисты по ИКТ
22	22. Специалисты в области здравоохранения
23	23. Специалисты в области образования
24	24. Специалисты в сфере бизнеса и администрирования
26	26. Специалисты в области права, гуманитарных областей и культуры
31	31. Специалисты-техники в области науки и техники
	35. Специалисты-техники в области ИКТ
32	32. Средний медицинский персонал здравоохранения
33	33. Средний специальный персонал по экономической и административной деятельности
34	34. Средний специальный персонал в области правовой, социальной работы, культуры, спорта и родственных занятий
44	41. Служащие общего профиля и обслуживающие офисную технику
	42. Служащие сферы обслуживания населения
	44. Другие офисные служащие
43	43. Служащие в сфере обработки числовой информации и учета материальных ценностей
51	51. Работники сферы индивидуальных услуг
	53. Работники, оказывающие услуги по индивидуальному уходу
52	52. Продавцы
54	54. Работники служб, осуществляющих охрану граждан и собственности
60	61. Квалифицированные работники с/х, производящие товарную продукцию
	62. Товарные производители лесной и рыбной продукции и охотники
	92. Неквалифицированные рабочие с/х, рыбоводства и рыболовства
71	71. Рабочие, занятые в строительстве, и рабочие родственных занятий
72	72. Рабочие, занятые в металлообрабатывающем и машиностроительном производстве, механики и ремонтники
75	73. Рабочие, занятые изготовлением прецизионных инструментов и приборов, рабочие художественных промыслов и полиграфического производства
	75. Рабочие пищевой, деревообрабатывающей, текстильной и швейной промышленности и рабочие родственных занятий
74	74. Рабочие в области электротехники и электроники
81	81. Операторы промышленных установок и стационарного оборудования
	82. Сборщики
83	83. Водители и операторы подвижного оборудования
91	91. Уборщики и прислуга
	93. Неквалифицированные рабочие, занятые в горнодобывающей промышленности, строительстве, обрабатывающей промышленности и на транспорте
	94. Помощники в приготовлении пищи
	95. Уличные торговцы и другие неквалифицированные работники, оказывающие различные уличные услуги
	96. Неквалифицированные работники по сбору мусора и другие неквалифицированные работники

Таблица 2а. Распределение по статусу занятости: мужчины (в процентах от численности мужчин соответствующей возрастной группы)

		2005	2010	2015	2020	Среднее за все годы (2005–2020 гг.)
Все (18–59 лет)	Самозанятые	8,9	7,2	6,8	7,2	7,4
	Занятые по найму	66,7	70,4	69,2	68,3	69,0
	Незанятые	24,4	22,5	24,0	24,5	23,6
В возрасте 18–25 лет	Самозанятые	8,1	5,4	3,7	3,9	5,3
	Занятые по найму	50,1	59,1	50,9	44,4	50,9
	Незанятые	41,8	35,5	45,4	51,7	43,7
В возрасте 26–45 лет	Самозанятые	9,8	8,3	7,8	7,7	8,4
	Занятые по найму	72,8	76,7	77,5	79,9	77,1
	Незанятые	17,5	15,0	14,7	12,4	14,5
В возрасте 46–59 лет	Самозанятые	7,8	6,6	7,0	8,5	7,1
	Занятые по найму	68,5	68,5	67,1	66,0	68,2
	Незанятые	23,7	24,9	25,9	25,5	24,7

Таблица 2б. Распределение по статусу занятости: женщины (в процентах от численности женщин соответствующей возрастной группы)

		2005	2010	2015	2020	Среднее за все годы (2005–2020 гг.)
Все (18–59 лет)	Самозанятые	5,1	3,8	4,0	3,7	4,0
	Занятые по найму	60,4	62,7	60,7	60,3	61,9
	Незанятые	34,5	33,5	35,3	36,1	34,1
В возрасте 18–25 лет	Самозанятые	3,7	3,2	2,6	1,5	2,8
	Занятые по найму	39,5	43,3	40,2	34,2	41,8
	Незанятые	56,8	53,5	57,2	64,3	55,4
В возрасте 26–45 лет	Самозанятые	6,3	3,8	4,4	4,4	4,4
	Занятые по найму	67,7	69,9	66,2	68,1	68,9
	Незанятые	26,0	26,3	29,4	27,5	26,7
В возрасте 46–59 лет	Самозанятые	4,3	4,3	4,2	3,9	4,3
	Занятые по найму	62,9	65,4	63,4	63,5	63,7
	Незанятые	32,8	30,4	32,5	32,6	32,0

Таблица 3. Описание выборки

	2005		2020	
	Женщины	Мужчины	Женщины	Мужчины
Возраст, лет	39,1	37,9	40,4	38,8
Есть дети до 18 лет, %	47,0	46,6	40,4	43,1
Средняя продолжительность рабочей недели	41,7	46,8	41,1	46,2
Рабочая неделя меньше 30 часов, %	6,6	1,7	6,5	2,4
Часовая заработная плата, руб.	30,1	40,4	53,2	61,2
Занят(-а) в госсекторе, %	46,3	31,3	42,4	23,0
Семейное положение, %				
Не состоит в браке	37,3	25,7	37,4	28,6
Зарегистрированный брак	52,5	62,9	50,5	58,3
Незарегистрированный брак	10,2	11,5	12,1	13,1
Образование, %				
Ниже среднего	5,0	10,7	6,0	12,2
Среднее	32,7	48,4	22,6	34,0
Среднее профессиональное	34,2	20,9	30,1	26,5
Высшее	28,1	20,0	41,4	27,3
Размер предприятия, %				
Микро- и малые предприятия (1–99 чел.)	43,2	32,6	55,0	37,4
Средние (100–249 чел.)	12,9	11,2	9,6	10,8
Крупные (250–499 чел.)	6,9	7,0	3,7	4,7
Очень крупные (500 + чел.)	14,8	16,2	7,9	11,2
Нет ответа	22,3	33,0	23,9	35,8

Таблица 4. Распределение работников по «мужским» и «женским» профессиям

	Неклассифицированные профессии*	Женские профессии	Смешанные профессии	Мужские профессии
Оба пола				
2005	0,0	44,5	23,1	32,4
2010	0,5	43,4	26,5	29,6
2015	1,1	43,9	17,2	37,7
2020	0,9	46,1	15,6	37,4
Женщины				
2005	0,0	72,9	20,9	6,2
2010	0,2	71,3	23,9	4,6
2015	0,5	73,2	17,4	8,9
2020	0,6	75,2	15,7	8,6
Мужчины				
2005	0,1	16,9	25,3	57,8
2010	0,7	17,1	29,0	53,2
2015	1,7	16,3	17,1	65,0
2020	1,2	18,4	15,5	64,9

Примечание. *Не классифицированы профессии, в которых занято менее 20 респондентов в соответствующем году обследования.

Таблица 5. Изменение отраслевой структуры занятости

	Женщины			Мужчины		
	2005	2020	Изменение	2005	2020	Изменение
Сельское и лесное хозяйство	4,4	1,9	-2,4	8,0	6,1	-1,9
Добыча полезных ископаемых	1,9	1,1	-0,8	4,0	3,9	-0,2
Легкая и пищевая промышленность	11,6	6,7	-4,9	12,0	8,1	-3,9
Тяжелая промышленность	5,8	4,6	-1,2	9,5	12,1	2,6
Строительство	3,4	2,1	-1,3	14,3	10,7	-3,6
Транспорт и связь	7,2	4,9	-2,3	14,4	13,9	-0,5
Госуправление, армия, полиция	5,0	7,1	2,1	8,2	9,8	1,6
Образование	14,6	17,1	2,5	3,1	3,3	0,2
Наука, искусство, спорт	4,4	4,2	-0,2	2,7	2,0	-0,6
Здравоохранение, социальное обслуживание	13,1	12,6	-0,5	2,8	2,4	-0,4
Торговля, общепит, бытовые услуги	21,6	28,7	7,1	14,3	18,7	4,4
Финансы, деловые услуги	3,5	5,8	2,3	1,3	3,8	2,5
ЖКХ	3,4	2,9	-0,4	5,2	4,8	-0,4
Прочие отрасли	0,3	0,3	0,1	0,5	0,6	0,1

Таблица 6. Распределение работников по «мужским» и «женским» отраслям

	Неклассифицированные отрасли*	Женские отрасли	Смешанные отрасли	Мужские отрасли
Оба пола				
2005	0,4	19,0	71,7	8,9
2010	0,0	19,0	50,8	30,2
2015	0,0	17,2	55,7	27,1
2020	0,4	17,4	51,2	31,0
Женщины				
2005	0,3	31,3	65,0	3,4
2010	0,0	31,3	53,5	15,1
2015	0,0	30,1	58,3	11,6
2020	0,3	29,7	55,4	14,6
Мужчины				
2005	0,5	7,2	78,1	14,3
2010	0,0	7,4	48,3	44,3
2015	0,0	4,9	53,3	41,8
2020	0,6	5,7	47,2	46,6

Примечание. * Не классифицированы отрасли, в которых занято менее 20 респондентов в соответствующем году обследования.

Таблица 7. Порядок добавления переменных для тестирования условия common support

R ²	1 шаг	2 шаг	3 шаг	4 шаг	5 шаг	6 шаг	7 шаг	8 шаг	9 шаг	10 шаг	11 шаг	12 шаг
	0,469	0,427	0,375	0,301	0,244	0,175	0,133	0,105	0,087	0,075	0,067	0,060
Отрасль	-0,042	–	–	–	–	–	–	–	–	–	–	–
Тип поселения	-0,040	-0,052	–	–	–	–	–	–	–	–	–	–
Федеральный округ	-0,031	-0,033	-0,074	–	–	–	–	–	–	–	–	–
Профессия	-0,038	-0,044	-0,052	-0,057	–	–	–	–	–	–	–	–
Образование	-0,013	-0,016	-0,022	-0,021	-0,069	–	–	–	–	–	–	–
Рабочее время	-0,032	-0,029	-0,031	-0,032	-0,037	-0,042	–	–	–	–	–	–
Размер предприятия	-0,008	-0,010	-0,015	-0,018	-0,022	-0,027	-0,028	–	–	–	–	–
Возраст	0,000	-0,001	-0,002	-0,005	-0,005	-0,014	-0,015	-0,018	–	–	–	–
Сектор	-0,004	-0,008	-0,014	-0,015	-0,021	-0,016	-0,011	-0,011	-0,012	–	–	–
Наличие детей	-0,002	-0,002	-0,002	-0,002	-0,002	-0,003	-0,002	-0,002	-0,008	-0,008	–	–
Специфический стаж	-0,003	-0,003	-0,003	-0,002	-0,003	-0,004	-0,006	-0,007	-0,009	-0,006	-0,007	–
Семейное положение	-0,002	-0,003	-0,002	-0,001	-0,001	-0,002	-0,002	-0,002	-0,002	-0,002	-0,006	-0,008
Общий стаж	-0,001	-0,001	0,000	-0,001	0,000	-0,004	-0,005	-0,007	-0,001	-0,001	-0,004	-0,005

Примечание. Все регрессии включают годовые дамми-переменные. При оценивании регрессий использовались данные за все годы с 2005 по 2020 г.

Таблица 8. Анализ условия common support

Переменные	All		Matched		Unmatched		% matched females	Gender gap on the matched sample				
	male	female	male	female	male	female		Raw		Adjusted		N
								log points	%	log points	%	
Год	7152	8406	7152	8406	0	0	100,0	-0,213*	-19,2	-0,213*	-19,2	15558
+ отрасль	7152	8406	7152	8406	0	0	100,0	-0,213*	-19,2	-0,179*	-16,4	15558
+ тип поселения	7152	8406	7138	8393	14	13	99,8	-0,213*	-19,1	-0,157*	-14,5	15531
+ федеральный округ	7152	8406	6765	7998	387	408	95,1	-0,214*	-19,3	-0,162*	-14,9	14763
+ профессия	7125	8387	2533	3656	4592	4731	43,6	-0,239*	-21,3	-0,213*	-19,1	6189
+ образование	7125	8387	1523	2162	5602	6225	25,8	-0,204*	-18,4	-0,196*	-17,8	3685
+ рабочее время	7125	8387	1202	1663	5923	6724	19,8	-0,243*	-21,5	-0,226*	-20,2	2865
+ размер предприятия	7107	8355	607	840	6500	7515	10,1	-0,256*	-22,6	-0,185*	-16,9	1447
+ возраст (катег.)	7107	8355	354	447	6753	7908	5,4	-0,253*	-22,3	-0,184*	-16,8	801
+ сектор	7107	8355	313	397	6794	7958	4,8	-0,267*	-23,4	-0,196*	-17,8	710
+ наличие детей	7099	8351	220	273	6879	8078	3,3	-0,257*	-22,7	-0,199*	-18,0	493
+ специфический стаж	6647	7914	56	58	6591	7856	0,7	-0,283*	-24,6	-0,254*	-22,4	114
+ семейное положение	6646	7905	40	39	6606	7866	0,5	-0,222	-19,9	-0,191*	-17,4	79
+ общий стаж	6646	7905	9	9	6637	7896	0,1	-0,379	-31,5	-0,400	-33,0	18

Примечание. Расчеты по объединенным данным РМЭЗ ВШЭ за 2005, 2010, 2015 и 2020 гг. Метод — Coarsened Exact Matching.

* Различия статистически значимы на 5%-ном уровне.

Таблица 9. Гендерный разрыв по результатам оценивания регрессионных моделей

	2005	2010	2015	2020
Разница в средних				
коэффициент	-0,310***	-0,223***	-0,205***	-0,185***
ст. ошибка	[0,0281]	[0,0172]	[0,0196]	[0,0207]
в процентах	-26,7	-20,0	-18,5	-16,9
Линейная регрессия — без отраслей и профессий ($\delta_{LRM,1}$)				
коэффициент	-0,377***	-0,275***	-0,276***	-0,244***
ст. ошибка	[0,0247]	[0,0151]	[0,0168]	[0,0187]
в процентах	-31,4	-24,0	-24,1	-21,7
Линейная регрессия — с отраслями и профессиями ($\delta_{LRM,2}$)				
коэффициент	-0,275***	-0,221***	-0,244***	-0,203***
ст. ошибка	[0,0292]	[0,0174]	[0,0199]	[0,0220]
в процентах	-24,0	-19,8	-21,7	-18,4
Линейная регрессия — LASSO ($\hat{\delta}_{LRM,lasso}$)				
коэффициент	-0,250***	-0,208***	-0,228***	-0,160***
ст. ошибка	[0,0321]	[0,0179]	[0,0195]	[0,0225]
в процентах	-22,1	-18,8	-20,4	-14,8
Декомпозиция Оаксаки — Блайндера — без отраслей и профессий ($\hat{\delta}_{OB,1}$)				
коэффициент	-0,339***	-0,252***	-0,225***	-0,205***
ст. ошибка	[0,0289]	[0,0174]	[0,0196]	[0,0214]
в процентах	-28,8	-22,3	-20,1	-18,5
Декомпозиция Оаксаки — Блайндера — с отраслями и профессиями ($\hat{\delta}_{OB,2}$)				
коэффициент	-0,291***	-0,205***	-0,183***	-0,123***
ст. ошибка	[0,0426]	[0,0234]	[0,0274]	[0,0311]
в процентах	-25,2	-18,5	-16,7	-11,6
Декомпозиция Оаксаки — Блайндера — LASSO ($\hat{\delta}_{OB,lasso}$)				
коэффициент	-0,248***	-0,212***	-0,195***	-0,107***
ст. ошибка	[0,0159]	[0,0089]	[0,0106]	[0,0128]
в процентах	-22,0	-19,1	-17,7	-10,2
Число наблюдений	2637	5636	3725	2566

Примечание. Для декомпозиции Оаксаки — Блайндера приводится величина необъясненного гендерного разрыва. Детализированные результаты декомпозиции Оаксаки — Блайндера для спецификаций, включающих и не включающих отраслевые и профессиональные переменные, см. в табл. 10.

Таблица 10. Результаты оценивания декомпозиции Оаксаки — Блайндера

	Без отраслей и профессий				С отраслями и профессиями			
	2005	2010	2015	2020	2005	2010	2015	2020
Средняя ЗП (ln) — женщины	3,143*** [0,0189]	3,587*** [0,0116]	3,671*** [0,0135]	3,814*** [0,0145]	3,143*** [0,0190]	3,587*** [0,0116]	3,671*** [0,0136]	3,814*** [0,0146]
Средняя ЗП (ln) — мужчины	3,453*** [0,0213]	3,810*** [0,0129]	3,875*** [0,0144]	3,999*** [0,0150]	3,453*** [0,0215]	3,810*** [0,0129]	3,875*** [0,0145]	3,999*** [0,0151]
Разница в средних ЗП	-0,310*** [0,0285]	-0,223*** [0,0173]	-0,205*** [0,0197]	-0,185*** [0,0209]	-0,310*** [0,0287]	-0,223*** [0,0174]	-0,205*** [0,0198]	-0,185*** [0,0210]
Объясненная часть	0,0283 [0,0236]	0,0288* [0,0143]	0,0205 [0,0170]	0,0201 [0,0171]	-0,0189 [0,0413]	-0,0184 [0,0224]	-0,0216 [0,0266]	-0,0618* [0,0296]
Необъясненная часть	-0,339*** [0,0289]	-0,252*** [0,0174]	-0,225*** [0,0196]	-0,205*** [0,0214]	-0,291*** [0,0426]	-0,205*** [0,0234]	-0,183*** [0,0274]	-0,123*** [0,0311]
Объясненная часть								
Демография	-0,0087 [0,0080]	-0,0195*** [0,0057]	-0,0235*** [0,0060]	-0,0273*** [0,0073]	-0,0100 [0,0076]	-0,0144** [0,0052]	-0,0191*** [0,0055]	-0,0229*** [0,0068]
Человеческий капитал	0,0456*** [0,0113]	0,0476*** [0,00820]	0,0615*** [0,0099]	0,0580*** [0,0103]	0,0269* [0,0112]	0,0296*** [0,0076]	0,0424*** [0,0093]	0,0327*** [0,0095]
География	-0,0095 [0,0126]	-0,0077 [0,0068]	0,0128 [0,0083]	0,0042 [0,0073]	-0,0082 [0,0106]	-0,0071 [0,0063]	0,0112 [0,0077]	0,0037 [0,0072]
Характеристики рабочих мест	0,0008 [0,0136]	0,0084 [0,0076]	-0,0302** [0,0095]	-0,0147 [0,0108]	0,0143 [0,0129]	0,0303*** [0,0079]	-0,0095 [0,0098]	0,0091 [0,0116]
Сегрегация					-0,0420 [0,0381]	-0,0568** [0,0205]	-0,0466* [0,0237]	-0,0844** [0,0279]

	Без отраслей и профессий				С отраслями и профессиями			
	2005	2010	2015	2020	2005	2010	2015	2020
Необъясненная часть								
Демография	-0,0275 [0,0375]	-0,0295 [0,0194]	-0,0051 [0,0231]	-0,119** [0,0386]	-0,0320 [0,0356]	-0,0204 [0,0182]	0,0046 [0,0219]	-0,0981** [0,0368]
Человеческий капитал	0,0741 [0,144]	-0,0148 [0,0846]	-0,0192 [0,0960]	0,1520 [0,128]	0,0522 [0,137]	-0,0358 [0,0801]	-0,1030 [0,0916]	0,1100 [0,122]
География	-0,0215 [0,0172]	-0,0063 [0,0128]	-0,0095 [0,0138]	-0,0280 [0,0147]	-0,0104 [0,0165]	-0,0024 [0,0121]	-0,0026 [0,0132]	-0,0200 [0,0140]
Характеристики рабочих мест	-0,147** [0,0553]	-0,0526 [0,0310]	-0,0725 [0,0378]	0,166*** [0,0460]	-0,126* [0,0544]	-0,0717* [0,0307]	-0,131*** [0,0381]	0,111* [0,0472]
Сегрегация					-0,0188 [0,0467]	-0,0467 [0,0318]	0,0025 [0,0294]	0,0966* [0,0369]
Константа	-0,217 [0,140]	-0,146 [0,0826]	-0,119 [0,0944]	-0,376** [0,120]	-0,1940 [0,141]	-0,0280 [0,0853]	0,0517 [0,0951]	-0,323** [0,120]
N	2637	5636	3725	2566	2637	5636	3725	2566

Примечание. Блок «Демография» объединяет возраст, семейное положение, наличие детей; блок «Человеческий капитал» — образование, общий и специфический трудовой стаж; блок «География» — тип населенного пункта и регион проживания (федеральный округ); блок «Характеристики рабочих мест» — сектор, размер предприятия, рабочее время; блок «Сегрегация» — отраслевую и профессиональную принадлежность.

Таблица 11. Диагностика модели для индекса склонности: тест на баланс ковариат

	2005				2010				2015				2020			
	Raw		Weighted		Raw		Weighted		Raw		Weighted		Raw		Weighted	
N наблюдений	2637		2637		5634		5634.00		3719		3719		2561		2561	
N подвергшихся воздействию	1445		1321		3069		2973.80		2003		1958		1388		1246	
N контрольных	1192		1316		2565		2660.20		1716		1761		1173		1315	
Тест overid chi2(65) p-value	43,250 0,980				66,379 0,429				36,9875 0,998				29,002 1,000			
	Standardized differences		Variance ratio		Standardized differences		Variance ratio		Standardized differences		Variance ratio		Standardized differences		Variance ratio	
	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted	Raw	Weighted
Наличие детей	-0,019	0,020	0,996	1,004	-0,049	-0,053	0,981	0,971	-0,055	-0,065	0,986	0,972	-0,152	0,093	0,981	1,021
Семейное положение (не женат/не замужем)																
в зарег. браке	-0,252	0,014	1,106	0,994	-0,271	0,036	1,100	0,987	-0,231	0,086	1,083	0,961	-0,254	0,075	1,144	0,966
в незарег. браке	-0,038	-0,075	0,906	0,831	-0,043	-0,013	0,910	0,973	-0,026	-0,084	0,945	0,830	0,024	-0,221	1,061	0,625
Возраст (18–25 лет)																
26–45 лет	-0,040	0,109	1,002	1,009	0,018	-0,083	1,000	0,988	-0,088	-0,016	1,019	1,004	-0,143	0,089	1,049	0,968
46–59 лет	0,097	-0,122	1,061	0,939	0,081	0,126	1,051	1,065	0,154	0,010	1,112	1,008	0,183	-0,075	1,092	0,965
Образование (неполное среднее)																
среднее	-0,301	0,049	0,877	1,028	-0,309	0,098	0,800	1,056	-0,305	-0,041	0,765	0,962	-0,276	-0,007	0,760	0,991
среднее проф.	0,310	-0,036	1,373	0,967	0,228	-0,057	1,245	0,937	0,169	-0,034	1,185	0,961	0,078	0,035	1,084	1,044
высшее	0,190	-0,010	1,256	0,990	0,253	-0,077	1,270	0,927	0,296	-0,056	1,240	0,965	0,292	-0,003	1,169	0,999
Госсектор	0,308	0,022	1,087	1,005	0,368	-0,081	1,171	0,958	0,467	0,017	1,284	1,008	0,476	-0,011	1,280	0,996
Размер предприятия (1–99 человек)																
100–249 человек	0,029	-0,051	1,066	0,889	-0,017	0,027	0,961	1,066	0,021	0,008	1,056	1,023	-0,037	0,020	0,913	1,054
250–499 человек	-0,015	-0,007	0,952	0,976	-0,029	-0,031	0,896	0,890	-0,029	0,101	0,882	1,484	-0,049	0,010	0,813	1,046
500 человек	-0,048	0,073	0,914	1,140	-0,097	-0,034	0,798	0,921	-0,124	0,001	0,752	1,001	-0,111	-0,023	0,754	0,945
Неизвестен	-0,240	0,041	0,771	1,043	-0,252	0,176	0,760	1,150	-0,248	0,033	0,766	1,030	-0,273	0,078	0,741	1,083

Продолжение табл. 11

Тип населенного пункта (село и пгт)																
город	-0,017	0,029	0,984	1,024	-0,017	-0,016	0,984	0,985	-0,030	-0,080	0,971	0,922	-0,048	-0,091	0,960	0,933
областной центр	0,026	-0,013	1,024	0,988	0,013	-0,030	1,011	0,975	0,023	0,175	1,018	1,100	-0,004	0,068	0,997	1,055
Москва и СПб	-0,007	-0,010	0,987	0,983	-0,019	0,173	0,953	1,380	0,041	-0,044	1,106	0,896	0,012	0,052	1,030	1,131
Федеральный округ (ЦФО)																
СЗФО	0,018	0,065	1,046	1,167	0,027	0,017	1,073	1,046	0,073	0,017	1,205	1,049	0,027	0,075	1,070	1,216
ЮФО	0,034	0,012	1,080	1,031	-0,006	-0,063	0,986	0,870	-0,052	-0,042	0,904	0,916	-0,034	0,027	0,939	1,055
ПФО	0,008	0,012	1,011	1,018	0,014	-0,002	1,019	0,997	0,009	-0,001	1,012	0,998	-0,014	0,045	0,981	1,065
УФО	-0,040	0,081	0,896	1,248	0,003	0,001	1,008	1,003	-0,002	0,012	0,996	1,035	0,007	0,034	1,018	1,097
СФО	0,007	-0,019	1,018	0,958	-0,013	-0,043	0,967	0,890	-0,001	0,117	0,998	1,325	-0,030	0,029	0,923	1,077
ДВФО	-0,022	-0,020	0,916	0,915	-0,044	-0,059	0,823	0,738	-0,048	-0,088	0,794	0,662	-0,029	-0,107	0,881	0,629
Рабочее время (менее 30 часов)																
30–48 часов	0,149	0,042	0,813	0,933	0,174	0,131	0,783	0,811	0,148	0,144	0,790	0,773	0,131	0,187	0,785	0,736
более 48 часов	-0,307	-0,048	0,580	0,906	-0,310	-0,098	0,562	0,812	-0,289	-0,091	0,556	0,807	-0,263	-0,013	0,546	0,968
Профессии																
21	-0,024	-0,001	0,893	0,996	-0,088	-0,005	0,681	0,977	-0,126	-0,012	0,599	0,953	-0,171	0,027	0,477	1,117
22	0,135	0,015	2,532	1,100	0,134	-0,003	3,050	0,975	0,196	0,015	6,921	1,125	0,206	0,046	8,434	1,446
23	0,370	-0,088	6,564	0,738	0,341	-0,016	4,111	0,945	0,454	-0,133	8,630	0,672	0,434	-0,187	6,846	0,611
24	0,214	-0,019	11,240	0,871	0,208	0,007	5,840	1,053	0,149	0,004	6,075	1,038	0,222	0,020	6,954	1,144
26	0,200	0,021	3,225	1,118	0,217	0,004	3,220	1,021	0,217	-0,006	3,227	0,970	0,193	0,006	2,623	1,027
30	-0,054	0,002	0,814	1,010	-0,073	-0,026	0,726	0,887	-0,211	-0,034	0,410	0,868	-0,229	-0,005	0,399	0,982
32	0,287	0,010	7,306	1,055	0,333	0,000	12,193	1,000	0,322	0,064	10,400	1,445	0,322	0,032	7,762	1,161
33	0,279	-0,024	2,695	0,929	0,309	-0,041	2,596	0,892	0,314	-0,011	2,381	0,973	0,311	0,075	2,414	1,231
34	0,078	0,012	2,090	1,115	0,085	-0,032	2,004	0,789	0,107	-0,029	2,147	0,827	0,063	-0,008	1,620	0,946
43	0,231	-0,025	6,064	0,866	0,153	0,009	2,805	1,061	0,068	0,004	1,520	1,026	0,035	0,016	1,244	1,104
44	0,294	0,029	14,340	1,185	0,297	-0,072	11,845	0,691	0,297	-0,036	11,088	0,829	0,304	0,050	19,024	1,357
51	0,239	-0,015	2,712	0,946	0,175	-0,018	2,086	0,932	0,213	0,010	2,791	1,045	0,256	0,004	3,804	1,018
52	0,340	-0,025	6,560	0,904	0,347	-0,023	5,590	0,915	0,322	-0,004	3,888	0,986	0,302	0,014	3,668	1,051
54	-0,281	0,051	0,159	1,287	-0,328	0,002	0,122	1,007	-0,326	0,057	0,103	1,299	-0,334	0,047	0,140	1,226
60	-0,024	-0,009	0,827	0,931	0,002	-0,009	1,025	0,907	0,005	-0,010	1,070	0,871	-0,023	0,000	0,677	1,005

71	-0,317	0,073	0,179	1,350	-0,314	0,006	0,140	1,032	-0,390	0,162	0,070	1,815	-0,335	0,042	0,099	1,222
72	-0,485	-0,097	0,052	0,653	-0,484	0,191	0,029	1,851	-0,455	-0,022	0,052	0,915	-0,466	-0,023	0,023	0,904
74	-0,320	0,093	0,111	1,532	-0,284	0,013	0,129	1,073	-0,263	-0,028	0,137	0,841	-0,340	-0,117	0,026	0,413
75	0,064	-0,002	1,587	0,984	0,076	-0,009	1,870	0,927	0,031	-0,001	1,249	0,993	0,042	0,001	1,354	1,011
81	-0,077	0,007	0,747	1,028	-0,175	-0,035	0,488	0,863	-0,176	-0,023	0,407	0,889	-0,149	0,027	0,489	1,131
83	-0,640	-0,020	0,099	0,949	-0,633	0,040	0,057	1,114	-0,635	0,042	0,043	1,122	-0,573	-0,033	0,076	0,899
93	-0,185	-0,011	0,264	0,931	-0,147	-0,022	0,358	0,858	-0,122	-0,018	0,436	0,886	-0,170	-0,008	0,282	0,947
96	0,287	0,066	4,750	1,377	0,230	-0,017	3,517	0,921	0,237	0,028	3,693	1,156	0,233	0,046	3,678	1,263
Отрасли (с/х)																
добыча	-0,157	0,005	0,400	1,031	-0,150	0,049	0,447	1,261	-0,232	0,096	0,243	1,551	-0,242	-0,016	0,194	0,906
легкая пром.	-0,028	-0,014	0,935	0,966	-0,024	-0,044	0,936	0,880	-0,090	-0,021	0,778	0,942	-0,057	0,073	0,845	1,214
тяжелая пром.	-0,154	0,030	0,626	1,087	-0,237	0,073	0,454	1,249	-0,220	-0,014	0,510	0,960	-0,313	0,064	0,413	1,178
строительство	-0,358	0,057	0,295	1,190	-0,413	0,188	0,222	1,607	-0,406	0,130	0,215	1,426	-0,312	-0,047	0,245	0,813
транспорт и связь	-0,238	-0,005	0,542	0,986	-0,258	-0,045	0,487	0,886	-0,323	-0,051	0,380	0,868	-0,275	0,018	0,444	1,052
госуправление,																
силовой блок	-0,133	0,090	0,633	1,324	-0,120	-0,006	0,692	0,982	-0,087	0,053	0,753	1,181	-0,052	0,092	0,859	1,327
образование	0,440	-0,072	4,033	0,839	0,416	-0,052	3,388	0,877	0,507	-0,121	4,989	0,753	0,534	-0,140	5,040	0,755
наука, искусство, спорт	0,105	0,038	1,661	1,219	0,149	-0,024	2,180	0,885	0,163	-0,014	2,551	0,929	0,130	0,027	2,063	1,161
здравоохранение	0,414	0,037	4,393	1,122	0,402	0,021	4,000	1,068	0,440	0,084	4,918	1,330	0,441	0,039	6,247	1,134
торговля, общепит	0,176	-0,082	1,449	0,845	0,226	-0,026	1,542	0,953	0,228	-0,045	1,475	0,922	0,154	-0,023	1,272	0,963
финансы, бизнес-услуги	0,160	0,035	3,222	1,269	0,137	0,013	2,721	1,098	0,133	0,002	1,842	1,010	0,102	0,006	1,590	1,029
ЖКХ	-0,121	0,028	0,592	1,120	-0,111	-0,101	0,608	0,609	-0,115	-0,075	0,563	0,670	-0,150	-0,042	0,499	0,817
прочие	-0,041	-0,006	0,413	0,885	0,005	-0,017	1,030	0,894	-0,052	0,132	0,516	3,281	-0,009	-0,008	0,845	0,860
Стаж	0,116	-0,105	1,136	0,999	0,163	0,114	1,169	0,975	0,201	-0,017	1,292	1,017	0,188	-0,054	1,316	0,885
Стаж^2	0,154	-0,107	1,257	0,929	0,205	0,110	1,269	0,950	0,261	-0,013	1,396	0,969	0,244	-0,078	1,373	0,797
Спецстаж	0,231	-0,119	1,225	0,798	0,224	-0,042	1,361	0,920	0,216	0,015	1,489	1,006	0,266	-0,077	1,432	0,925
Спецстаж^2	0,177	-0,128	1,262	0,673	0,198	-0,043	1,485	0,816	0,221	0,010	1,827	1,125	0,258	-0,070	1,765	0,957

Примечание. Нумерация профессиональных групп не в полной мере соответствует ISCO-08. Отдельные небольшие группы на уровне двухзначной кодировки были объединены с более крупными группами на том уровне кодировки. Исключение составили работники из 6-й профессиональной группы. В группу 60 были добавлены 92 «Неквалифицированные рабочие сельского хозяйства».

Таблица 12. Результаты оценивания с использованием индекса склонности, одна переменная воздействия

	2005	2010	2015	2020
IPW — без цензурирования				
необъясненный разрыв (АТЕТ), %	–30,4***	–24,0***	–23,7***	–26,2***
ст. ошибка	[7,0]	[4,7]	[5,1]	[7,7]
N	2637	5636	3725	2566
IPW — цензурировано на 0,99				
необъясненный разрыв (АТЕТ), %	–30,2***	–23,9***	–23,9***	–17,9***
ст. ошибка	[7,0]	[4,7]	[5,1]	[5,8]
N	2598	5573	3634	2522
IPW — цензурировано на 0,95				
необъясненный разрыв (АТЕТ), %	–28,6***	–21,6***	–21,4***	–14,1***
ст. ошибка	[3,9]	[2,5]	[3,0]	[3,8]
N	2337	5202	3347	2307
IPW — цензурировано на 0,90				
необъясненный разрыв (АТЕТ), %	–25,3***	–19,8***	–19,2***	–18,3***
ст. ошибка	[4,2]	[2,7]	[3,0]	[3,5]
N	2063	4670	2980	2071
IPWRA				
необъясненный разрыв (АТЕТ), %	–26,0***	–20,7***	–18,5***	–10,7***
ст. ошибка	[2,9]	[2,1]	[2,3]	[2,8]
N	2637	5636	3725	2566
LASSO-AIPW				
необъясненный разрыв (АТЕТ), %	–17,3**	–25,4***	–19,6***	–13,6***
ст. ошибка	[8,2]	[5,7]	[2,4]	[4,5]
N	2637	5636	3725	2566
p-value для теста overid	0,212	0,670	0,809	0,071

Примечание. **, *** Статистически значимо на 5%-ном и 1%-ном уровне.

Таблица 13. Переменные, отобранные в спецификацию индекса склонности методом LASSO-AIPW

	2005	2010	2015	2020
Среднее профессиональное образование	x			
Отрасль				
Строительство	x	x	x	
Образование	x	x	x	x
Здравоохранение	x	x	x	x
Профессия (isco)				
23			x	
30			x	x
32		x		
33			x	
44	x	x	x	x
52	x	x	x	x
54	x	x	x	x
71	x	x	x	x
72	x	x	x	x
74	x	x	x	x
81		x	x	
83	x	x	x	x
93	x	x		
Рабочая неделя более 48 часов	x	x		
Пересечения				
Нет детей # isco-93	x			
Есть дети # Не состоит в браке		x	x	
Возраст (46–60) # Среднее проф. обр.			x	
Возраст (26–45) # isco-21			x	x
Возраст (18–25) # Опыт работы		x		
Состоит в зарег. браке # Рабочая неделя более 48 ч.	x	x		x
Состоит в зарег. браке # isco-10		x		
Состоит в зарег. браке # isco-30			x	x
Не состоит в браке # Опыт работы	x	x	x	x
Состоит в зарег. браке # Частный сектор				x
Среднее проф. обр. # Рабочая неделя 30–48 ч.	x			
Среднее проф. обр. # Здравоохранение	x	x		
Среднее проф. обр. # Госсектор		x		
Рабочая неделя 30–48 ч. # isco-33	x			x
Рабочая неделя 30–48 ч. # isco-24		x		
Рабочая неделя 30–48 ч. # isco-33		x	x	
Рабочая неделя 30–48 ч. # Спецстаж	x			
Госсектор # Рабочая неделя 30–48 ч.			x	
Госсектор # Малый бизнес (1–99 чел.)	x	x	x	x
Частный сектор # isco-10		x		
Госсектор # Здравоохранение			x	
Госсектор # Опыт работы			x	
Госсектор # Спецстаж		x	x	x
Малый бизнес (1–99 чел.) # isco-52		x		
Малый бизнес (1–99 чел.) # Спецстаж		x		
Торговля # Опыт работы			x	
isco-26 # Опыт работы		x		
isco-33 # Опыт работы			x	
isco-52 # Опыт работы		x	x	x
Константа	x	x	x	x

Таблица 14. Модель с двумя воздействиями: гендер и родительство

	Мужчины без детей	Мужчины с детьми	Женщины без детей	Женщины с детьми
Уровень воздействия (Т)	0 (0 0)	1 (0 1)	2 (1 0)	3 (1 1)
2005				
Средние часовые ЗП	37,2	43,1	28,6	31,1
По отношению к Т=0 (0 0)				
Разница в средних		5,9***	-8,5***	-6,1***
IPWRA, АТЕТ		2,9	-12,1***	-9,8***
IPWRA, %		7	-30	-24
По отношению к Т=1 (0 1)				
Разница в средних			-14,5***	-12,0***
IPWRA, АТЕТ			-7,4***	-14,7***
IPWRA, %			-20	-32
По отношению к Т=2 (1 0)				
Разница в средних				2,5*
IPWRA, АТЕТ				2,2
IPWRA, %				8
2010				
Средние часовые ЗП	50,9	59,6	43,5	45,9
По отношению к Т=0 (0 0)				
Разница в средних		8,7***	-7,4***	-5,0***
IPWRA, АТЕТ		5,0**	-10,9***	-19,2***
IPWRA, %		9	-20	-30
По отношению к Т=1 (0 1)				
Разница в средних			-16,1***	-13,7***
IPWRA, АТЕТ			-23,8***	-18,8***
IPWRA, %			-35	-29
По отношению к Т=2 (1 0)				
Разница в средних				2,4*
IPWRA, АТЕТ				1,2
IPWRA, %				3
2015				
Средние часовые ЗП	52,0	63,9	47,1	47,9
По отношению к Т=0 (0 0)				
Разница в средних		11,8***	-4,9***	-4,1**
IPWRA, АТЕТ		4,6	-11,6***	-11,6***
IPWRA, %		8	-20	-19
По отношению к Т=1 (0 1)				
Разница в средних			-16,7***	-15,9***
IPWRA, АТЕТ			-27,9***	-23,2***
IPWRA, %			-37	-33
По отношению к Т=2 (1 0)				
Разница в средних				0,8
IPWRA, АТЕТ				-2,7
IPWRA, %				-5

2020				
Средние часовые ЗП	57,8	68,0	53,9	50,8
По отношению к Т=0 (0 0)				
Разница в средних		10,2***	-4,0**	-7,0***
IPWRA, АТЕТ		6,2**	-11,6***	-10,6***
IPWRA, %		10	-18	-17
По отношению к Т=1 (0 1)				
Разница в средних			-14,1***	-17,2***
IPWRA, АТЕТ			-17,0***	-20,8***
IPWRA, %			-24	-29
По отношению к Т=2 (1 0)				
Разница в средних				-3,1*
IPWRA, АТЕТ				-1,6
IPWRA, %				-3

Примечание *, **, *** — статистически значимо на 10-, 5- и 1%-ном уровне.

Таблица 15. Модель с двумя воздействиями: гендер и сектор занятости

	Мужчины/ частный сектор	Мужчины/ госсектор	Женщины/ частный сектор	Женщины/ госсектор
Уровень воздействия (Т)	0 (0 0)	1 (0 1)	2 (1 0)	3 (1 1)
2005				
Средние часовые ЗП	42,8	34,6	34,1	25,5
По отношению к Т=0 (0 0)				
Разница в средних		-8,1***	-8,7***	-17,3***
IPWRA, АТЕТ		-9,2***	-14,9***	-26,0***
IPWRA, %		-21	-30	-50
По отношению к Т=1 (0 1)				
Разница в средних			-0,5	-9,1***
IPWRA, АТЕТ			-4,3*	-12,0***
IPWRA, %			-11	-32
По отношению к Т=2 (1 0)				
Разница в средних				-8,6***
IPWRA, АТЕТ				-10,4***
IPWRA, %				-29
2010				
Средние часовые ЗП	56,9	49,4	48,2	40,4
По отношению к Т=0 (0 0)				
Разница в средних		-7,5***	-8,7***	-16,4***
IPWRA, АТЕТ		-8,8***	-15,7***	-24,3***
IPWRA, %		-15	-25	-38
По отношению к Т=1 (0 1)				
Разница в средних			-1,1	-8,9***
IPWRA, АТЕТ			-4,7***	-14,1***
IPWRA, %			-9	-26
По отношению к Т=2 (1 0)				
Разница в средних				-7,8***
IPWRA, АТЕТ				-15,4***
IPWRA, %				-28

2015				
Средние часовые ЗП	58,4	54,5	49,0	45,9
По отношению к Т=0 (0 0)				
Разница в средних		-3,9**	-9,4***	-12,5***
IPWRA, АТЕТ		-7,7***	-18,6***	-23,7***
IPWRA, %		-12	-28	-34
По отношению к Т=1 (0 1)				
Разница в средних			-5,5***	-8,6***
IPWRA, АТЕТ			-7,9***	-13,1***
IPWRA, %			-14	-22
По отношению к Т=2 (1 0)				
Разница в средних				-3,1*
IPWRA, АТЕТ				-9,4***
IPWRA, %				-17
2020				
Средние часовые ЗП	63,5	61,5	51,8	53,4
По отношению к Т=0 (0 0)				
Разница в средних		-2,0	-11,7***	-10,1***
IPWRA, АТЕТ		-6,4***	-18,7***	-22,3***
IPWRA, %		-9	-26	-29
По отношению к Т=1 (0 1)				
Разница в средних			-9,7***	-8,2***
IPWRA, АТЕТ			-12,7***	-12,2***
IPWRA, %			-20	-19
По отношению к Т=2 (1 0)				
Разница в средних				1,5
IPWRA, АТЕТ				-2,2
IPWRA, %				-4

Примечание. *, **, *** Статистически значимо на 10-, 5- и 1%-ном уровне.

Таблица 16. Модель с двумя воздействиями: гендер и занятость в женских профессиях

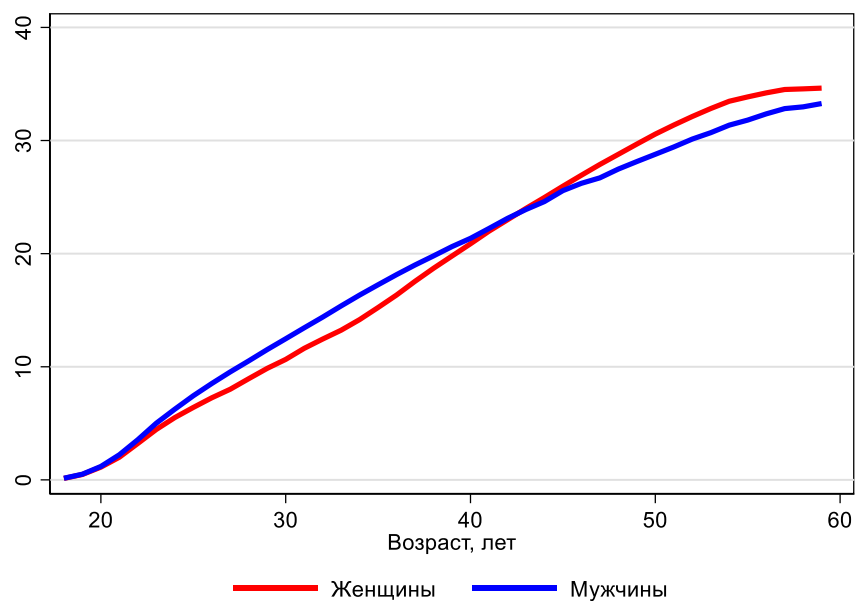
	Мужчины в неженских профессиях	Мужчины в женских профессиях	Женщины в неженских профессиях	Женщины в женских про- фессиях
Уровень воздействия (Т)	0 (0 0)	1 (0 1)	2 (1 0)	3 (1 1)
2005				
Средние часовые ЗП	39,8	40,1	31,1	29,2
По отношению к Т=0 (0 0)				
Разница в средних		0,3	-8,7***	-10,6***
IPWRA, АТЕТ		-2,6	-12,1***	-13,1***
IPWRA, %		-6	-28	-31
По отношению к Т=1 (0 1)				
Разница в средних			-9,0***	-10,9***
IPWRA, АТЕТ			-8,2**	-11,9***
IPWRA, %			-21	-29
По отношению к Т=2 (1 0)				
Разница в средних				-1,9
IPWRA, АТЕТ				-2,9
IPWRA, %				-9
2010				
Средние часовые ЗП	53,6	58,8	44,9	44,2
По отношению к Т=0 (0 0)				
Разница в средних		5,1**	-8,7***	-9,4***
IPWRA, АТЕТ		-1,8	-12,1***	-13,1***
IPWRA, %		-3	-21	-23
По отношению к Т=1 (0 1)				
Разница в средних			-13,8***	-14,5***
IPWRA, АТЕТ			-6,2***	-11,7***
IPWRA, %			-12	-21
По отношению к Т=2 (1 0)				
Разница в средних				0,7
IPWRA, АТЕТ				-1,7
IPWRA, %				-4
2015				
Средние часовые ЗП	57,1	58,9	49,3	46,8
По отношению к Т=0 (0 0)				
Разница в средних		1,8	-7,8***	-10,3***
IPWRA, АТЕТ		-4,8*	-13,2***	-14,1***
IPWRA, %		-7	-21	-23
По отношению к Т=1 (0 1)				
Разница в средних			-9,6***	-12,1***
IPWRA, АТЕТ			-9,2**	-10,9***
IPWRA, %			-16	-19
По отношению к Т=2 (1 0)				
Разница в средних				-2,5
IPWRA, АТЕТ				-4,0*
IPWRA, %				-8

2020				
Средние часовые ЗП	63,3	61,3	54,5	51,9
По отношению к Т=0 (0 0)				
Разница в средних		-2,0	-8,9***	-11,4***
IPWRA, АТЕТ		-2,5	-11,6***	-12,1***
IPWRA, %		-4	-18	-19
По отношению к Т=1 (0 1)				
Разница в средних			-6,8**	-9,4***
IPWRA, АТЕТ			-9,4***	-14,7***
IPWRA, %			-15	-22
По отношению к Т=2 (1 0)				
Разница в средних				-2,5
IPWRA, АТЕТ				-4,2*
IPWRA, %				-7

Таблица 17. Модель с двумя воздействиями: гендер и занятость в мужских отраслях

	Мужчины в немужских отраслях	Мужчины в мужских отраслях	Женщины в немужских отраслях	Женщины в мужских отраслях
Уровень воздействия (Т)	0 (0 0)	1 (0 1)	2 (1 0)	3 (1 1)
2005				
Средние часовые ЗП	39,9	39,5	29,2	42,8
По отношению к Т=0 (0 0)				
Разница в средних		-0,5	-10,7***	2,8
IPWRA, АТЕТ		1,2	-13,7***	-3,8
IPWRA, %		3,2	-32	-8
По отношению к Т=1 (0 1)				
Разница в средних			-10,2***	3,3
IPWRA, АТЕТ			-19,0***	-2,0
IPWRA, %			-39	-5
По отношению к Т=2 (1 0)				
Разница в средних				13,5***
IPWRA, АТЕТ				8,8***
IPWRA, %				26
2010				
Средние часовые ЗП	51,4	58,6	43,0	52,0
По отношению к Т=0 (0 0)				
Разница в средних		7,3***	-8,4***	-0,7
IPWRA, АТЕТ		9,7***	-13,4***	-6,6***
IPWRA, %		20	-24	-11
По отношению к Т=1 (0 1)				
Разница в средних			-15,6***	-6,6***
IPWRA, АТЕТ			-23,3***	-14,8***
IPWRA, %			-35	-22
По отношению к Т=2 (1 0)				
Разница в средних				9,0***
IPWRA, АТЕТ				7,8***
IPWRA, %				18
2015				
Средние часовые ЗП	56,8	58,3	46,7	53,2
По отношению к Т=0 (0 0)				
Разница в средних		1,5	-10,0***	-3,6
IPWRA, АТЕТ		6,5***	-14,9***	-9,2***
IPWRA, %		13	-24	-15
По отношению к Т=1 (0 1)				
Разница в средних			-11,5***	-5,1
IPWRA, АТЕТ			-22,5***	-16,2***
IPWRA, %			-32	-23
По отношению к Т=2 (1 0)				
Разница в средних				6,5**
IPWRA, АТЕТ				6,3**
IPWRA, %				13
2020				
Средние часовые ЗП	60,9	65,2	52,0	55,6
По отношению к Т=0 (0 0)				
Разница в средних		4,3*	-8,9***	-5,3**
IPWRA, АТЕТ		10,3***	-13,7***	-11,7***
IPWRA, %		19	-21	-17
По отношению к Т=1 (0 1)				
Разница в средних			-13,2***	-9,7***
IPWRA, АТЕТ			-19,8***	-18,1***
IPWRA, %			-28	-25
По отношению к Т=2 (1 0)				
Разница в средних				3,6
IPWRA, АТЕТ				3,5
IPWRA, %				7

Примечание. *, **, *** Статистически значимо на 10-, 5- и 1%-ном уровне.



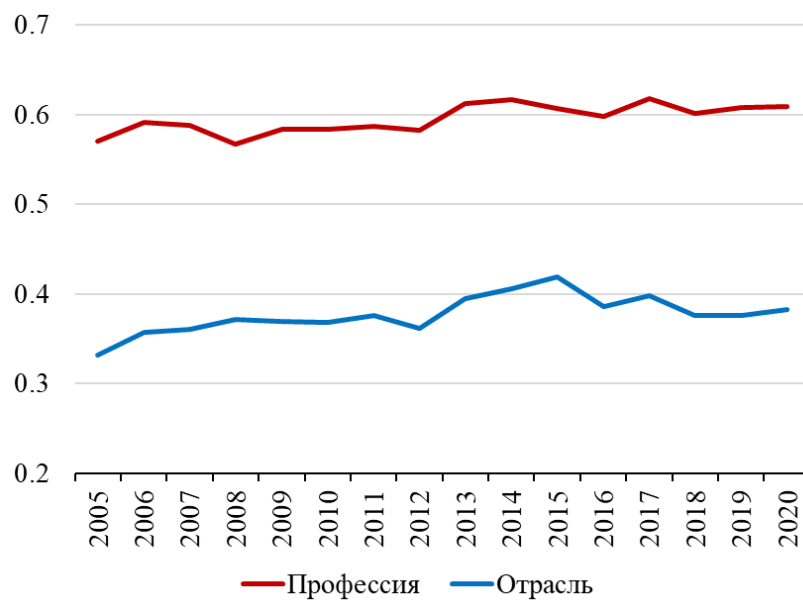
Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 1. Ожидаемый стаж



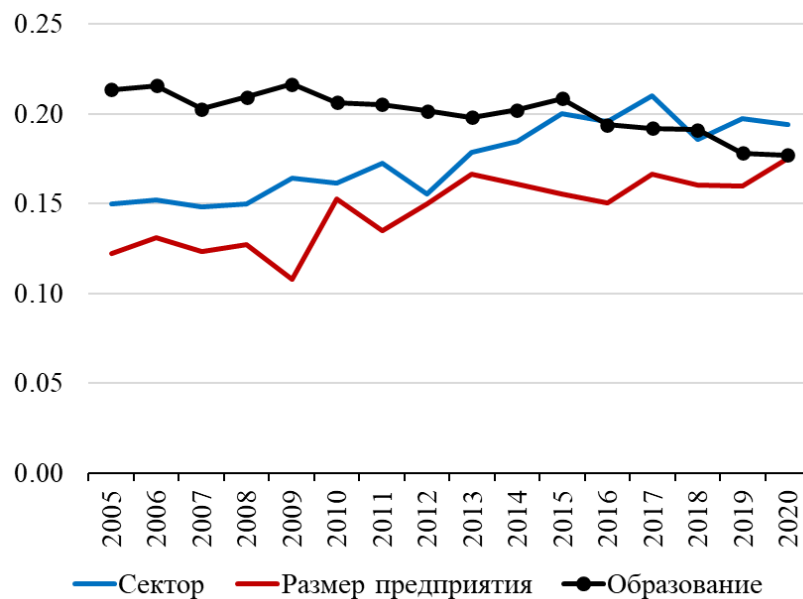
Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 2. Гендерные разрывы в продолжительности рабочего времени и заработных платах, %



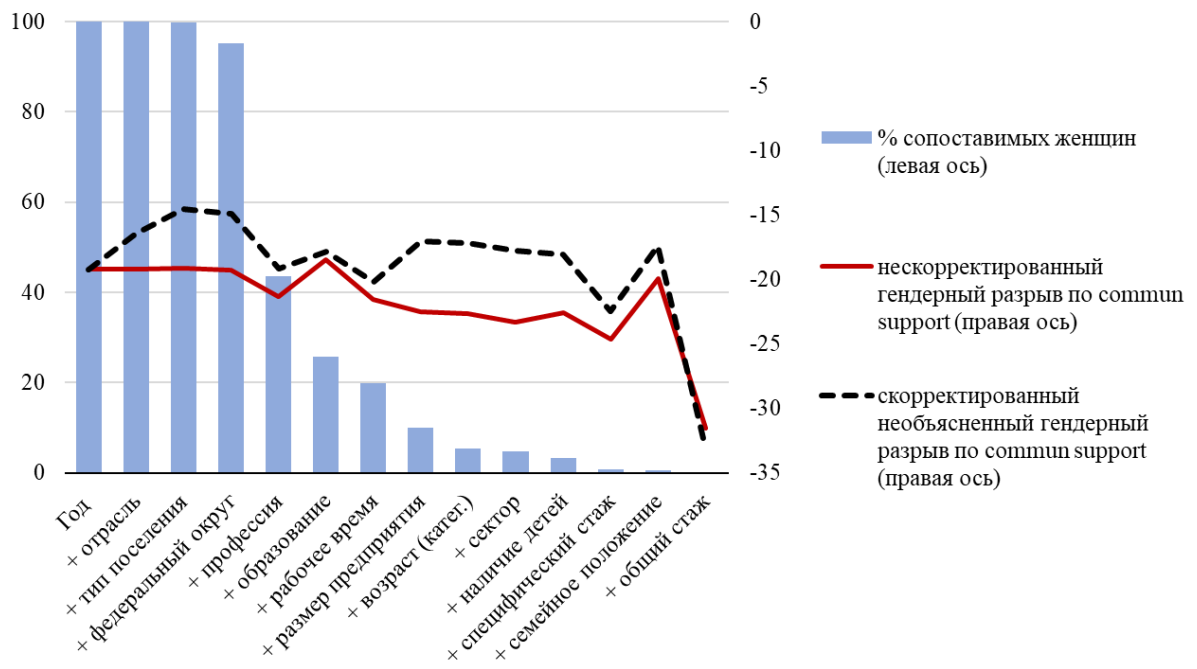
Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 3. Профессиональная и отраслевая сегрегация по полу



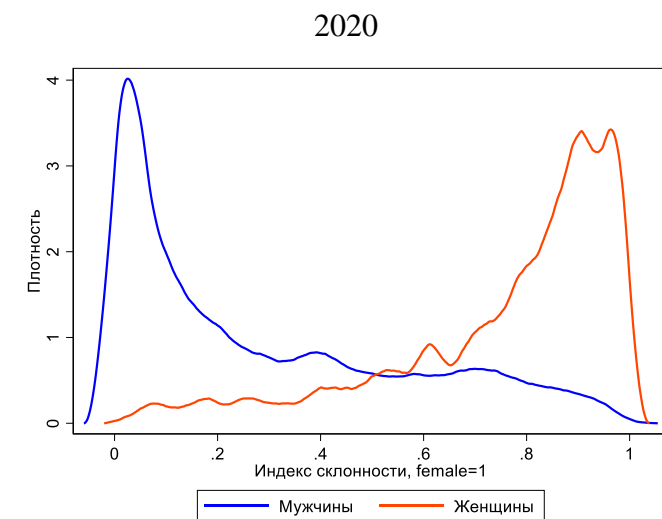
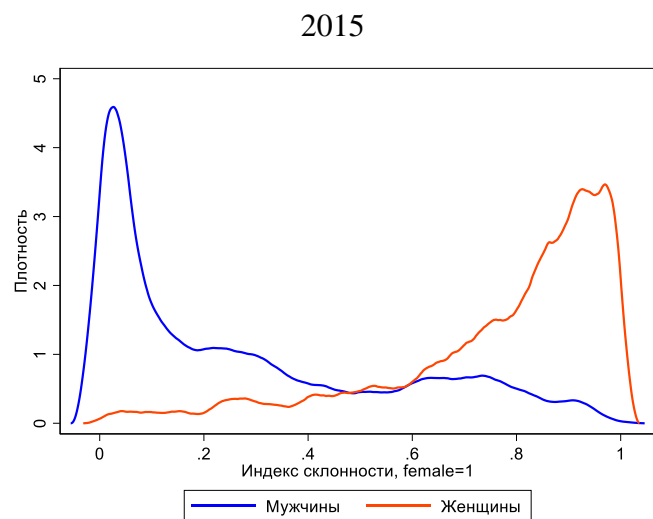
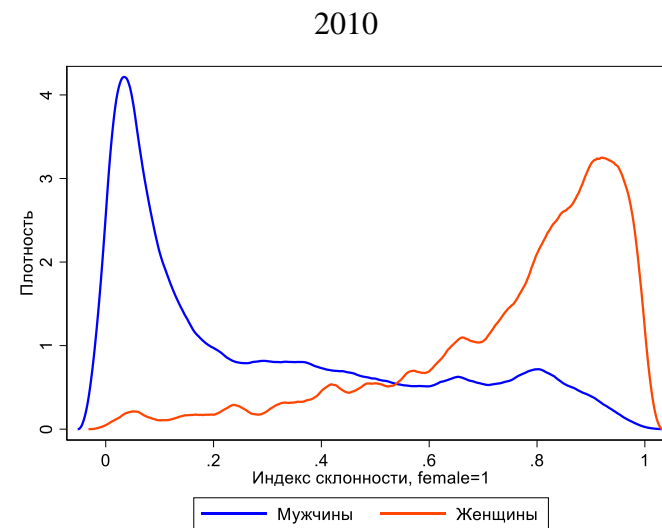
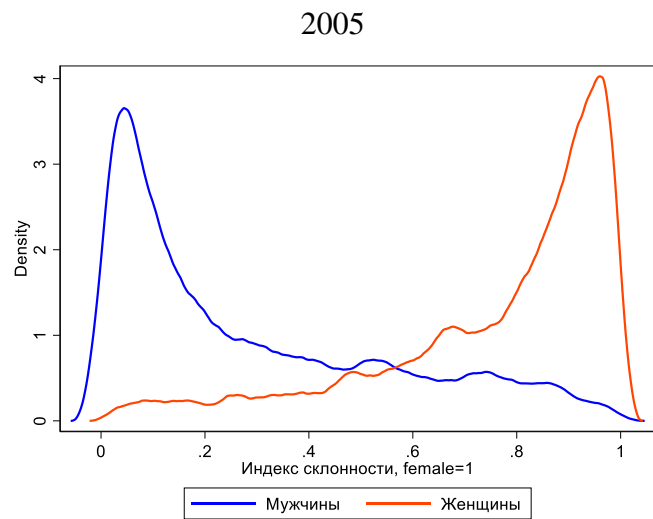
Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 4. Гендерная сегрегация по другим характеристикам



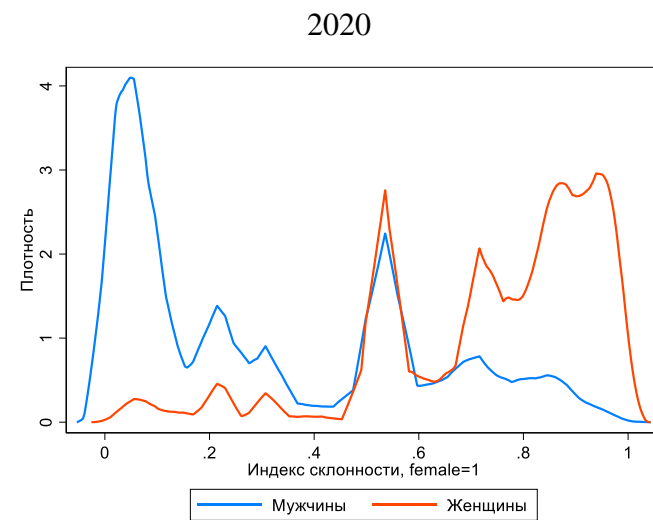
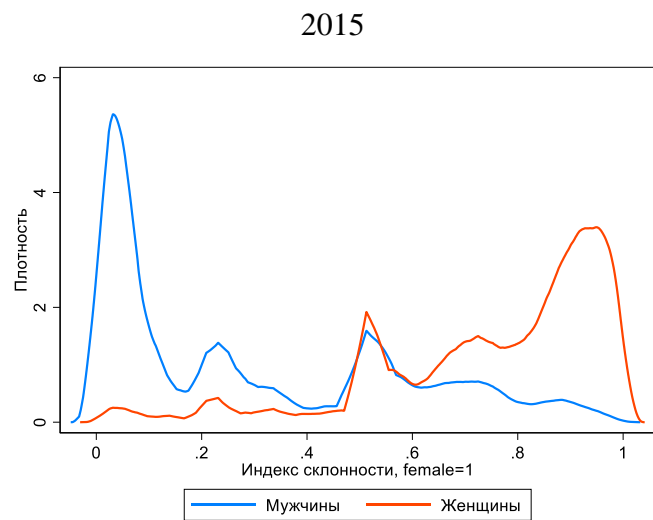
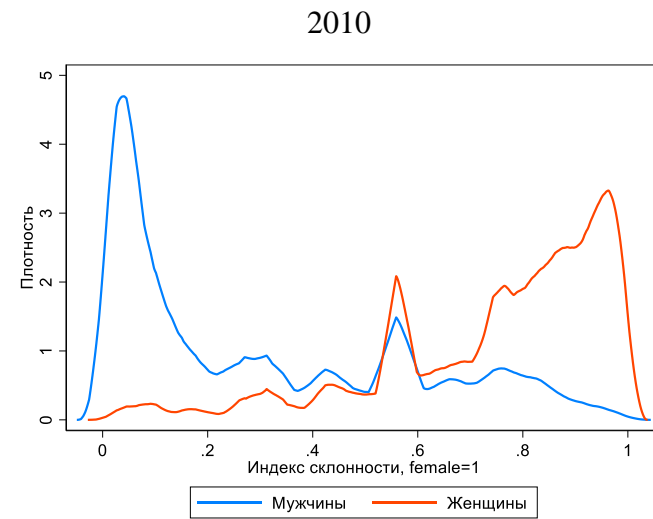
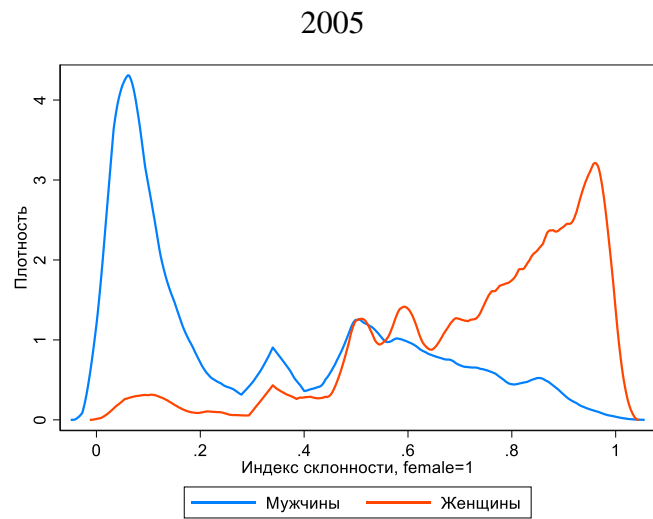
Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 5. Анализ условия common support



Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 6. Анализ условия common support для расчетных значений индекса склонности



Источник: Расчеты автора по РМЭЗ ВШЭ.

Рис. 7. Анализ условия common support для расчетных значений индекса склонности — LASSO-AIPW

Lukyanova, Anna.

Gender pay gap: Looking beyond the Oaxaca-Blinder decomposition [Electronic Resource] : Working paper WP3/2022/08 / A. Lukyanova ; National Research University Higher School of Economics. – Electronic text data (1,7 Mb). – Moscow : HSE Publ. House, 2022. – 65 p. – (Series WP3 “Labour Markets in Transition”). (In Russian)

The paper investigates how recent advances in the treatment effects literature can be applied to study gender pay inequality in Russia. The key parameter of interest is the mean unexplained gender wage gap. Using the RLMS HSE data for 2005–2020, we consider linear regression, Oaxaca-Blinder (OB) decomposition, and estimators based on propensity scores (PS) and machine learning algorithms. The latter group of methods provides more flexible and robust ways to model the impact of gender. PS-based estimators address the difficulties of comparing like with like. LASSO is used to select data-driven specifications of treatment and wage equations for all methods under consideration. In addition to the models with gender as a single treatment, we estimate the models with multivalued treatments interacting gender with relevant characteristics of female and male employment that reflect gender segregation and gender roles. We show that the samples of men and women are indeed extremely heterogeneous, and that heterogeneity has risen since 2005, mirroring an increase in gender segregation. Heterogeneity causes common support violations and aggravates the risks of model misspecification. Double robust estimators can tackle both problems. Unfortunately, a relatively small size of annual samples in the HSE RLMS imposes serious restrictions on the use of treatment-effects methods. With the HSE RLMS data, we did not obtain significant improvements in the estimates of the unexplained gender gap compared to the OB decomposition. However, since PS-based estimators have stronger causal interpretations, the similarity of the results enhances the credibility of the OB-estimates.

Keywords: gender pay gap, Oaxaca-Blinder decomposition, common support, treatment effects, propensity score, LASSO

JEL-codes: C21, J16, J31

Lukyanova Anna – Cand. Sci. (Econ.), Senior Researcher, Center for Labor Studies, National Research University Higher School of Economics; 20 Myasnitskaya Ulitsa, Moscow, 101000, Russian Federation; E-mail: alukyanova@hse.ru

Acknowledgements

This article is an output of a research project implemented as a part of the Basic Research Program at the National Research University Higher School of Economics (HSE University).

Препринт WP3/2022/08
Серия WP3
Проблемы рынка труда

Лукьянова Анна Львовна

**Гендерный разрыв в оплате труда:
за рамками декомпозиции Оаксаки — Блайндера**

Изд. № 2691