ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

А.Ю. Ощепков

ПРОБЛЕМЫ ИЗМЕРЕНИЯ СПЕЦИАЛЬНОГО СТАЖА НА ДАННЫХ РМЭЗ – ВШЭ

Препринт WP3/2016/04
Серия WP3
Проблемы рынка труда

Москва
2016

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

Ощепков, А. Ю.
Проблемы измерения специального стажа на данных РМЭЗ – ВШЭ [Электронный ресурс]: препринт WP3/2016/04 / А. Ю. Ощепков ; Нац. исслед. ун-т «Высшая школа экономики». – Электрон. текст. дан. (1,5 Мб). – М. : Изд. дом Высшей школы экономики, 2016. – (Серия WP3 «Проблемы рынка труда»). – 36 с.

В работе анализируются проблемы измерения специального стажа работников на данных РМЭЗ – ВШЭ. Подробно обсуждается потенциально возможный перечень таких проблем, основное внимание фокусируется на несостоятельности ответов на вопрос о дате начала работы у тех же самых респондентов в двух последовательных волнах обследования. Прежде всего мы определяем условия, при которых несовпадение этих дат указывает на наличие ошибки измерения специального стажа, а затем анализируем распространенность и величину этих ошибок, их связь с различными характеристиками индивидов, а также их влияние на результаты дескриптивного и регрессионного анализа.

Во-первых, мы показываем, что распространенность такого рода ошибок в данных РМЭЗ – ВШЭ высока: в среднем за период 1995–2014 гг. им были подвержены около четверти всех наблюдений по специальному стажу. Во-вторых, размер этих ошибок был существенным. В среднем за весь рассматриваемый период величина ошибки в абсолютном выражении составляла почти три года, и при этом примерно у четверти всех ошибок она превышала пять лет, тогда как в относительном выражении средняя величина ошибки составляла более 50% от длительности индивидуального стажа. В-третьих, как вероятность, так и размер этих ошибок не являются абсолютно случайными, а значимо коррелируют с некоторыми характеристиками индивидов. В частности, они связаны с размером получаемой заработной платы, что указывает на то, что оценивание стандартного уравнения заработной платы минцеровского типа на данных РМЭЗ – ВШЭ дает оценку отдачи от специального стажа, смещенную «вниз». Наконец, в-четвертых, мы приводим примеры того, как существующие ошибки измерения могут влиять на результаты анализа. Мы показываем, что любые выводы о динамике среднего специального стажа в России, сделанные на основе данных РМЭЗ – ВШЭ, имеют определенную долю условности, так как колебания этого показателя могли быть вызваны исключительно ошибками измерения. Кроме того, чувствительными к ошибкам измерения оказываются оценки отдачи от специального стажа, получаемые, прежде всего, на панельных данных, так как эти ошибки в большей степени влияют на вариацию «внутри индивидов во времени», чем на вариацию «между индивидами». В целом наша работа показывает, что исследователям следует с большой аккуратностью работать с данными по специальному стажу в обследовании РМЭЗ – ВШЭ и стараться учитывать возможные ошибки измерения в своих расчетах и выводах.

Ключевые слова: специальный стаж, ошибки измерения, РМЭЗ – ВШЭ, Россия

JEL codes: J62, J24, J31

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

© Ощепков А. Ю., 2016
© Оформление. Издательский дом Высшей школы экономики, 2016

1. Введение[1]

Специальный стаж работника – длительность занятости на текущем рабочем месте – одно из ключевых понятий в экономике труда. Оно занимает важное место в теории человеческого капитала, отражая запас *специфического* человеческого капитала – тех знаний, навыков и умений, которые повышают производительность труда работника на текущем рабочем месте, но которые становятся практически бесполезными при переходе на другую работу. В связи с этим сотни эмпирических исследований оценивают отдачу от специального стажа в разных странах. Специальный стаж также часто используется при анализе *трудовой мобильности*. Предполагается, что чем он короче, тем уровень мобильности выше. Различные показатели мобильности, основанные на информации о специальном стаже работников, широко используются в сравнительном анализе уровней и факторов трудовой мобильности по разным странам [Bassanini, Garnero, 2013; Burgess, 1998; OCED, 2009]. Кроме того, длительность трудовых отношений с работодателем имеет большое значение при анализе вероятности увольнения: обычно работники с более длительным стажем защищены от увольнения лучше, чем работники с коротким стажем.

Работая с данными обследований населения, исследователи используют информацию о специальном стаже, полученную на основе ответов респондентов. Это могут быть ответы на прямой вопрос о длительности трудовых отношений с работодателем или на вопрос о дате начала текущей работы. Естественно ожидать, что эти данные содержат в себе ошибки, так как респонденты часто либо помнят дату начала работы лишь приблизительно, либо не придают большого значения точности ответа и округляют свой стаж, либо по каким-то причинам умышленно дают ответ, не соответствующий действительности. Ряд западных исследований показывает, что распространенность и размер такого рода ошибок измерения могут быть весьма существенными, что оказывает заметное влияние на результаты анализа, в котором используются показатели специального стажа (например, [Brown, Light, 1992])[2].

Одной из основных проблем измерения специального стажа является несоответствие ответов одного и того же респондента в разные раунды обследования. Так как исследователю, работающему с данными, не дано знать, в каком интервью респондент

[1] Автор выражает благодарность В. Гимпельсону и Р. Капелюшникову за полезные комментарии и замечания. Работа подготовлена при поддержке Программы фундаментальных исследований НИУ ВШЭ в 2015 г.

[2] Длительность специального стажа является одним из типичных примеров показателей, получаемых на основе ответов респондентов, которые содержат в себе ошибки измерения (см. обзор [Bound et al., 2001]).

указал правильный стаж (или дату начала текущей работы), а в каком – неправильный, эта проблема обычно не имеет полноценного решения. В этом случае можно лишь ввести определенные предположения о природе расхождений в ответах и в дальнейшем выполнять анализ на основе этих предположений. Однако практика показывает, что разные предположения могут приводить к разным результатам.

В данной работе мы анализируем, с какими проблемами сталкиваются исследователи при измерении специального стажа с помощью данных Российского мониторинга экономики и здоровья (РМЭЗ – ВШЭ)[3]. Это обследование очень часто используется в эмпирических исследованиях российского рынка труда. Практически все работы, в которых изучался специальный стаж в России, включая работы, оценивающие отдачу от специального стажа [Lehmann, Wadsworth, 2000; Cheidvasser, Benítez-Silva, 2007; Мальцева, 2009; Капелюшников, 2011] или вероятность увольнения и анализирующие уровень трудовой мобильности и стабильности (например, [Карабчук, 2009]), выполнены с использованием этих данных. Однако до сих пор вопрос о том, какие ошибки могут содержать эти данные и как эти ошибки могут влиять на результаты и выводы анализа, практически не рассматривался. Между тем ошибки измерения потенциально могут иметь отношение к некоторым известным результатам, получаемым на данных РМЭЗ – ВШЭ. Например, это получаемая при оценивании стандартного минцеровского уравнения заработной платы незначимая или даже отрицательная отдача от специального стажа или относительно низкая доля занятых с длительным стажем в сравнении с другими источниками данных[4].

Мы восполняем этот пробел, анализируя данные РМЭЗ – ВШЭ за период с 1994 по 2014 г. Мы описываем возможные проблемы измерения специального стажа, измеряем распространенность и величину ошибок для каждого раунда обследования, предлагаем несколько способов корректировки этих ошибок, а также анализируем, как эти ошибки и способы их корректировки могут влиять на результаты дескриптивного и регрессионного анализа.

Работа построена следующим образом. Во втором разделе подробно описываются возможные проблемы измерения специального стажа на данных обследования РМЭЗ – ВШЭ. В третьем разделе приводятся результаты эмпирического анализа этих проблем. В четвертом разделе мы суммируем основные результаты и делаем общие выводы.

---

[3] Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ (RLMS – HSE), проводимый Национальным исследовательским университетом «Высшая школа экономики» и ЗАО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии РАН (сайты обследования RLMS – HSE: http://www.cpc.unc.edu/projects/rlms и http://www.hse.ru/rlms).

[4] Прежде всего, в сравнении с Обследованием населения по проблемам занятости (ОНПЗ).

2. Измерение специального стажа на данных РМЭЗ – ВШЭ

РМЭЗ – ВШЭ не содержит прямого вопроса о длительности трудовых отношений занятого индивида с текущим работодателем, однако содержит вопрос о дате начала текущей работы: *«Скажите, пожалуйста, с какого года и месяца Вы работаете на этом предприятии, на этой работе? Если Вы увольнялись и снова возвращались на это же предприятие, на эту работу, то назовите дату последнего возвращения».* На основе этой информации можно рассчитать длительность специального стажа как разницу между датой проведения обследования и датой начала работы, указанной респондентом. Это операция выглядит достаточно простой, однако при ее реализации на практике может возникнуть ряд вопросов. Основные из них мы рассмотрим ниже.

Отсутствие информации о дате начала работы

Первая проблема, с которой можно столкнуться при анализе данных по специальному стажу, состоит в том, что информация о дате начала работы отсутствует. Другим словами, неизвестен год начала текущей работы. В целом за период 1995–2014 гг. таких респондентов было примерно 7,5% от всех занятых и эта доля могла меняться от 3,5 до 11% год от года. Такие респонденты автоматически исключаются из любого анализа, который затрагивает показатели специального стажа (вопрос о том, в какой мере это выбытие является случайным и связано ли оно с какими-либо характеристиками респондентов, затрагивается в третьем разделе).

Оставшихся респондентов – у которых известен хотя бы год начала работы (см. строку 2 табл. 1) – можно разбить на две группы. Первая группа – это те, для которых известен год начала работы, но неизвестен месяц ее начала, а вторая – те, для которых известны и год и месяц начала работы. Очевидно, что потенциально наиболее точные показатели специального стажа будут доступны для второй группы. Размер этой группы в общем за период 1995–2014 гг. составлял 87,3% от всех занятых и мог меняться от 84 до 93,3% год от года (см. строку 3 табл. 1). Возникает вопрос, учитывать ли при анализе респондентов первой группы, и если учитывать, то как?

Аргументом против того, чтобы учитывать таких респондентов, является неточность их показателей стажа. Стаж для них можно рассчитать только в целых годах, и это может дать ошибочное представление о его длительности. Приведем простой пример. Допустим, что некоторого индивида опросили в сентябре 2008 г. (рис. 1). По состоянию на эту дату индивид имел работу, на вопрос о дате начала этой работы он ответил, что на данном месте он работает с 2007 г., однако по каким-то причинам не указал месяц начала работы. В этом случае мы будем считать, что специальный стаж этого индивида будет

равен одному году (2008 – 2007 = 1). Если же он фактически начал работу в декабре, то стаж этого индивида на самом деле будет равен десяти месяцам, что соответствует нулю лет, если мы измеряем стаж в целых годах. Таким образом, мы можем переоценивать длительность специального стажа для таких респондентов, измеряя ее в целых годах.

Аргументом *за* использование наблюдений с неизвестным месяцем начала работы (но известным годом) в анализе является увеличение размера «рабочей» выборки. По нашим оценкам, добавка таких наблюдений увеличивает выборку в среднем на 5,5 п.п. (увеличение выборки может составить от 2 до 10,5 п.п. в зависимости от года обследования).

На наш взгляд, наиболее оптимальным вариантом является все же включение таких индивидов в анализ, но с вменением им определенного месяца начала работы. Если мы измеряем стаж таких индивидов в целых годах, то это означает, что мы имплицитно предполагаем, что они начали работать в январе. Однако при условии, что работа могла начаться с равной вероятностью в любой из двенадцати календарных месяцев (что, скорее всего, не так, так как доля рабочих дней различна в разные месяцы), имеет смысл вменять месяц начала работы, находящийся в середине года, т.е. июнь или июль. В этом случае ожидаемый размер ошибки будет минимальным и при этом он будет значительно меньше, чем в случае, когда месяц начала работы совсем не учитывается (т.е. принимается за январь).

Следует отметить, что в нашей работе, тем не менее, мы используем только те наблюдения, для которых известен как год, так и месяц начала текущей работы. Это объясняется целями нашего исследования: мы анализируем несостоятельность ответов на вопрос о дате начала работы у тех же самых респондентов во времени и поэтому должны иметь информацию не только о годе, но и о месяце начала работы для четкого сопоставления дат, указанных в разных интервью.

Дата начала текущей работы идет позже даты интервью

Еще одной естественной проблемой может быть то, что дата начала текущей работы, названная респондентом, идет *позже*, чем дата самого интервью. Например, интервью с респондентом было проведено в октябре 2012 г., а в качестве даты начала *текущей* работы он назвал ноябрь 2012 г. Это несоответствие приводит к тому, что стаж для такого индивида, рассчитываемый как разница между датой интервью и датой начала работы, будет отрицательным. Безусловно, это свидетельствует о наличии ошибки либо

в дате проведения интервью, либо в дате начала работы[5]. Забегая вперед, отметим, что данные РМЭЗ – ВШЭ содержат весьма незначительное число несоответствий такого рода, поэтому те наблюдения, в которых такое несоответствие присутствует, мы просто исключаем из выборки без каких-либо предположений о природе такой ошибки и способах ее корректировки.

Несоответствие между длительностью специального стажа и длительностью общего трудового стажа

Естественным ограничителем «сверху» для длительности специального стажа является возраст респондента. Если длительность специального стажа превышает возраст, то это будет свидетельствовать о наличии ошибки в дате начала работы. При этом, если это возможно, разумно использовать еще более жесткое ограничение: длительность специального стажа работника не должна превышать длительность его общего трудового стажа.

Вопрос об общем трудовом стаже в РМЭЗ – ВШЭ начал задаваться в 2001 г. Он звучит следующим образом: «Сколько полных лет и месяцев составляет Ваш официально оформленный общий трудовой стаж, не считая времени учебы на дневном отделении вуза или техникума? Сколько лет и сколько месяцев?». Такая формулировка предполагает, что респондент при ответе не должен учитывать стаж, который не оформлен официально (в трудовой книжке). Таким образом, если у респондента были периоды неофициальной работы, запись о которой в трудовой книжке отсутствует, то общий трудовой стаж для такого респондента будет недооцениваться при использовании этого вопроса. (Отметим, что вопрос о дате текущей работы не содержит в себе условия, что на этой работе респондент должен быть оформлен официально, что теоретически может приводить к тому, что длительность специального стажа будет превышать длительность общего стажа.) Кроме того, сменив несколько мест работы, респондент обычно лишь приблизительно может помнить (или вычислить в момент опроса) свой суммарный трудовой стаж. Все этого указывает на то, что для многих респондентов общий трудовой стаж будет измеряться с ошибками, и потому этот показатель является плохим кандидатом для сопоставления со специальным трудовым стажем. По этой причине в нашей работе в качестве верхней границы длительности специального стажа респондента мы принимаем максимальный потенциально возможный общий трудовой стаж, который для всех респондентов мы рассчитываем как разницу между возрастом и числом 15 (нижняя

[5] Нельзя исключать и такой случай: респондент имеет четкую договоренность о начале работы в ближайшее время, поэтому в момент опроса он сам себя считает занятым и указывает дату начала своей (будущей) работы. Этот случай можно классифицировать как ошибку в дате начала работы.

граница возраста экономической активности в России, используемая в обследованиях Росстата). Наши расчеты показывают, что в целом за период 1995–2014 гг. лишь менее 0,5% наблюдений специального стажа выходит за эту границу. Эти наблюдения мы исключаем из анализа без каких-либо корректировок.

Стаж более одного года у работников, незанятых в прошлом раунде

Каждый отдельный раунд обследования РМЭЗ – ВШЭ по естественным причинам не укладывается в один определенный календарный месяц, а проводится в течение нескольких месяцев. Обычно б*о*льшая часть всех интервью проводится в октябре или ноябре, но некоторые интервью могут проводиться в сентябре, декабре или даже в январе *следующего* года. В связи с этим месяц проведения интервью с одним и тем же индивидом (если он не выбывает из выборки) в разных раундах обследования может быть разным. Одним из следствий этого является то, что у индивидов, которые не были заняты на момент интервью в предыдущем раунде, может наблюдаться стаж работы один год и более в текущем раунде. Таким образом, априори нет оснований считать такие наблюдения ошибками, и поэтому общая рекомендация состоит в том, чтобы учитывать эти наблюдения в анализе.

Несоответствие между показателями стажа у одних и тех же индивидов в разные моменты времени

При использовании кросс-секционных данных исследователи имеют информацию о специальном стаже работника только на какой-либо один момент времени. В этом случае возможности каким-либо образом проверить корректность этой информации у них нет. Преимущество панельных данных состоит в том, что информацию о стаже можно наблюдать в разные моменты времени. Это позволяет проверить ее непротиворечивость. Согласованность показателей специального стажа в разные моменты времени, разумеется, еще не гарантирует полного отсутствия ошибок в данных о стаже. Однако их несогласованность с большой вероятностью указывает на наличие ошибок и требует внимания.

Рисунок 2 кратко отражает все возможные случаи (не)соответствия между датами начала текущей работы, указанными занятым респондентом в двух последовательных раундах опроса. Если один и тот же индивид был опрошен в ходе двух последовательных раундов обследования и при этом в обоих раундах находился в состоянии занятости, то его ответы о дате начала текущей работы могут как совпадать, так и не совпадать. Если его ответ о дате начала текущей работы при опросе в раунде $t - 1$ совпадает с ответом о дате начала работы в раунде t (такой случай мы будем называть Случай 1), то прирост

специального стажа для этого индивида будет равен периоду календарного времени между опросами (рис. 3). В этом случае можно сделать вывод, что смены места работы между раундами обследования не происходило.

Ситуация заметно осложняется, если ответы о дате начала текущей работы не совпадают. Здесь возможны два случая, которые мы будем называть Случай 2 и Случай 3. В Случае 2 дата начала работы, указанная во втором интервью, идет раньше, чем дата, указанная в первом интервью (рис. 4). Тогда прирост специального стажа для этого индивида окажется больше, чем период календарного времени между опросами. Почему это может происходить?

Здесь теоретически возможны два варианта. Первый вариант: индивид вернулся к прошлому работодателю, у которого он работал еще до даты начала работы, указанной в первом интервью, но затем ушел от него. Второй вариант: есть ошибка либо в дате начала работы, указанной в первом интервью, либо в дате, указанной во втором интервью.

В свою очередь, Случай 3 распадается на две возможные ситуации. В Случае 3.1 дата начала текущей работы, указанная во втором интервью, идет позже, чем дата, указанная в первом интервью, и при этом она идет до момента первого интервью (рис. 5). В этом случае прирост специального стажа окажется также больше, чем период календарного времени между опросами. Однако в отличие от Случая 2, это свидетельствует о присутствии ошибки в данных, так как в раунде $t - 1$ респондент должен был бы указать дату начала текущей работы, которая соответствует той, что он указал в раунде t. Ошибка содержится либо в дате, указанной в интервью в периоде $t - 1$, либо в дате, указанной в интервью в периоде t.

Наконец, Случай 3.2 возникает тогда, когда дата начала текущей работы, указанная в интервью в раунде t, идет позже, чем дата, указанная в интервью в раунде $t - 1$, и при этом она идет в периоде между первым и вторым интервью (рис. 6). В этом случае длительность стажа на момент интервью в периоде t однозначно короче периода между опросами, что говорит о том, что индивид менял место работы.

Выполненный анализ возможных случаев (не)соответствия ответов одного и того же респондента во времени показывает, что основания надеяться на то, что ошибок в данных по специальному стажу нет, есть в двух случаях. Это Случай 1, в котором можно предположить, что место работы не менялось, и Случай 3.2, который, наоборот, говорит о смене места работы. Случай 2 может быть вызван либо ошибкой данных, либо тем, что индивид перешел к бывшему работодателю. Но так как формулировка вопроса в РМЭЗ – ВШЭ включает в себя просьбу указать именно дату последнего *возвращения* к бывшему

работодателю, а не дату начала работы у него, то, видимо, Случай 2 связан только с ошибками в данных. Наконец, Случай 3.1. однозначно указывает на наличие ошибки.

3. Результаты анализа данных

Ограничение выборки

Наша выборка включает в себя индивидов, которые присутствовали в выборке и при этом были занятыми хотя бы в один из раундов обследования РМЭЗ – ВШЭ в период с 1994 по 2014 г. Статус занятости респондентов мы определяем традиционно на основе их ответов на следующий вопрос (j1): *Давайте поговорим о вашем основном занятии в настоящее время. Вы сейчас: 1) работаете; 2) находитесь в отпуске – декретном или по уходу за ребенком до 3-х лет; 3) находитесь в любом другом оплачиваемом отпуске; 4) находитесь в неоплачиваемом отпуске; 5) не имеете работы*. К занятым мы относим тех индивидов, которые выбрали варианты ответа 1, 3 или 4. Хотя женщины, находящиеся в декретном отпуске или отпуске по уходу за ребенком, являются (формально) занятыми и для них потенциально доступна информация о стаже, мы исключаем эту группу из анализа в целях сопоставимости с предыдущими исследованиями (эта группа обычно исключается из анализа при оценивании отдачи от специального стажа). Дополнительно к группе занятых принято относить индивидов, имевших в отчетный период случайный / непостоянный заработок (например, подвоз на машине за деньги). Однако такие индивиды также исключаются из нашего анализа, так как вопрос о дате начала работы им не задается.

Сопоставление дат начала текущей работы и дат проведения интервью показывает, что несоответствие между ними наблюдается лишь в восьми наблюдениях на всем массиве данных с 1994 по 2014 г. Эти наблюдения мы просто исключаем из выборки. Следующим ограничением нашей выборки является то, что мы используем только те наблюдения, для которых известны как год, так и месяц начала текущей работы[6]. В результате всех этих ограничений из анализа в среднем за разные годы исключаются примерно 13,5% занятых индивидов (табл. 1).

[6] Оценка бинарной пробит-модели показывает, что вероятность неответа на вопрос о дате начала текущей работы значимо связана с социально-демографическими характеристиками респондентов: с возрастом (положительно), полом (женщины чаще не отвечают, чем мужчины), образованием респондентов (менее образованные не дают ответ чаще), а также типом населенного пункта (вероятность неответа в сельской местности выше, чем в городской). Таким образом, отбор в выборку респондентов, имеющих информацию о специальном стаже, не является абсолютно случайным, что следует иметь в виду при работе с этой выборкой.

Распространенность ошибок

Насколько распространены ошибки в данных о специальном стаже? Для ответа на этот вопрос мы разделили все наблюдения на четыре группы, соответствующие четырем возможным случаям, описанным в методологическом разделе. Получившееся распределение за отдельные годы и за период 1995–2014 гг. в целом также представлено в табл. 2. Напомним, что в Случае 1 дата начала текущей работы, указанная в интервью в раунде t, соответствует дате, указанной в интервью в раунде $t - 1$. Это говорит о том, что между интервью индивид не менял работу. Во всех других случаях (2, 3.1 и 3.2) дата начала текущей работы, указанная в интервью в раунде t, *не* соответствует дате, указанной в интервью в раунде $t - 1$. Если в Случаях 2 и 3.1 это несоответствие свидетельствует о наличии ошибки, то в Случае 3.2 оно указывает на то, что респондент сменил место работы в период между интервью.

Суммарная доля Случаев 2 и 3.1 показывает распространенность ошибок. Эта доля оказывается достаточно большой: примерно четверть всех наблюдений, для которых известны год и месяц начала текущей работы, содержат в себе ошибку. Кроме четырех описанных случаев, есть достаточно большая доля индивидов, непротиворечивость данных по стажу для которых проверить нельзя (Случай 4). Эти индивиды либо не присутствовали в выборке в прошлый раунд обследования, либо присутствовали, но не были занятыми. Если предположить, что вероятность ошибки среди этих наблюдений такая же, как и среди тех, кто наблюдался и был занятым в прошлом периоде, то доля ошибок в целом за период возрастет почти до 34%, а в начале периода она превышала 50%.

Размер ошибок

Безусловно, значение имеет не только распространенность ошибок, но и их размер. Если размер большинства ошибок небольшой, то можно ожидать, что они будут слабо искажать результаты анализа, даже несмотря на их большую распространенность. В табл. 2 представлена информация о размерах абсолютной (верхняя панель таблицы) и относительной (нижняя панель таблицы) ошибки отдельно за каждый год и за весь период в целом. В целом за период средняя величина абсолютной ошибки составила 35 месяцев. Другими словами, разница между датами начала той же самой текущей работы, указанными респондентами в два последовательных раунда обследования, в среднем составляла почти три года.

Однако, как известно, среднее очень подвержено влиянию наблюдений, находящихся на краях распределения. В нашем случае такая высокая средняя ошибка

вызвана наличием относительно небольшого количества наблюдений с относительно большой ошибкой. Если «срезать» верхний 1% распределения, то величина средней ошибки, например в 2014 г., сократится с 31,1 месяца до 27,7 месяцев. Если же «срезать» верхние 5% наблюдений, то среднее сократится до 21,3 месяцев, т.е. составит уже менее двух лет. В соответствии с этими наблюдениями медианное значение средней абсолютной ошибки было гораздо меньше среднего и составляло за весь период 12 месяцев. Это значит, что размер половины всех ошибок не превышал одного года. В табл. 2 представлено также более подробное распределение ошибок по интервалам. Доля ошибок размером меньше одного месяца составляла 17–25% от числа всех ошибок. Примерно на таком же уровне находилась и доля ошибок величиной пять лет и более.

Относительная ошибка (соотношение между абсолютной ошибкой и длительностью специального стажа) в среднем за период составляла примерно 55%. Другим словами, ненаблюдаемая «реальная» длительность специального стажа у респондентов, подпадающих под Случаи 2 и 3.1, могла в среднем быть примерно наполовину больше или наполовину меньше, чем наблюдаемая длительность стажа, рассчитанная на основе ответов респондентов о дате начала работы. Медианное значение относительной ошибки опять было во все годы гораздо меньше среднего значения, что говорит о существовании относительно небольшого количества наблюдений с относительно высокой ошибкой.

Динамика ошибок с течением времени

Самостоятельный интерес представляет вопрос о том, сокращались ли ошибки с течением времени. Улучшалось ли качество измерения длительности специального стажа в обследовании РМЭЗ – ВШЭ от волны к волне? Рисунок 7 представляет основные индикаторы распространенности и величины ошибок: долю наблюдений, содержащих ошибку, среднюю абсолютную ошибку, долю ошибок более пяти лет и среднюю относительную ошибку – в динамике. *Доля ошибок* была максимальной в 1995 г. – 40% от всех наблюдений специального стажа, но уже к 1998 г. эта доля сократилась примерно до 27%. В последующий период эта доля находилась примерно на этом же уровне. Заметно сокращение данного показателя в 2010 г., что связано с увеличением выборки РМЭЗ – ВШЭ в этот год, за счет чего «скакнула» доля тех, чей стаж проверить нельзя, а доли остальных групп «просели». С 2011 г. доля ошибок находится на уровне около 25%.

Средняя ошибка в период с 1995 по 2001 г. находилась на уровне 45 месяцев, в 2002 г. она достаточно резко сократилась до 32,5 месяцев и весь последующий период находилась примерно на том же уровне. *Доля ошибок размером более пяти лет* (60 месяцев) полностью повторяла динамику средней ошибки. В 1995–2001 гг. эта доля

составляла примерно 25%, но с 2002 г. она находилась в границах 16–18%. Отметим, что эти показатели никак не отреагировали на увеличение выборки в 2010 г. (что вполне естественно, так как выборка расширялась за счет добавления новых респондентов, информацию о начале работы которых проверить нельзя).

Относительная ошибка в 1995 г. составляла в среднем почти 80%, но уже к 1998 г. сократилась до 52%. Весь последующий период она находилась в диапазоне 45–55%, однако в 2011 г. произошел ее резкий рост почти до уровня 1995 г. Весь последующий период после 2011 г. она сокращалась. (Отметим, что высокое значение средней относительной ошибки в 2011–2012 гг. вызвано наличием всего лишь нескольких наблюдений с коротким стажем, но большой абсолютной ошибкой.)

Таким образом, все индикаторы согласуются в том, что в 1990-е и в начале 2000-х годов качество измерения стажа было ниже, чем в последующий период. Однако никакой явной тенденции сокращения доли и размера ошибок от волны к волне не прослеживается, последние 13 лет ошибки находятся примерно на одинаковом уровне.

У кого ошибки чаще и больше?

В данном разделе анализируется, с какими характеристиками работников или рабочих мест связаны вероятность и размер ошибки в данных по специальному стажу. Этот вопрос имеет принципиальное значение для регрессионного анализа, где используется информация о специальном стаже – например, при оценивании отдачи от специального стажа или оценивании вероятности увольнения работника. Если вероятность и размер ошибки носят абсолютно случайный характер, то это не будет оказывать никакого влияния на результаты эконометрического оценивания. Если же ошибки неслучайны и связаны с какими-либо характеристиками респондентов, которые не учитываются в оцениваемых уравнениях, то это может приводить к смещению оценок.

В первом столбце табл. 3 приведены результаты анализа факторов, которые увеличивают вероятность ошибки. В таблице приведены предельные эффекты, полученные из оценивания ММП простой пробит-модели, где зависимая переменная принимает значение 1, если ошибка есть, и 0, если ошибки нет. Второй столбец табл. 3 показывает, как те же самые факторы связаны с размером ошибки. В этом случае зависимой переменной был (абсолютный) размер ошибки в месяцах, и уравнение оценивалось простым МНК. Результаты оценивания обоих уравнений качественно согласуются между собой – факторы, увеличивающие вероятность ошибки в данных по специальному стажу, увеличивают также и ее размер.

Прежде всего, вероятность и размер ошибки растут вместе с заработной платой работника. Это указывает на то, что оценки отдачи от специального стажа на данных РМЭЗ – ВШЭ с помощью стандартного уравнения заработной платы минцеровского типа подвержены смещению «вниз», т.е. *недооценивают* фактический уровень отдачи. Возникает ситуация, когда ошибка измерения независимой переменной коррелирует с ошибками модели. При этом сила смещения будет зависеть от того, какую долю занимает вариация ошибки в вариации наблюдаемого специального стажа (в английском языке используется термин *noise to signal ratio*) (см., например, [Bound et al., 2001]).

Как и следовало ожидать, вероятность и размер ошибки растут вместе с длительностью специального стажа. Чем дольше работник не менял работу, тем дальше от момента опроса находится дата начала работы и, следовательно, тем сильнее она забывается. Кроме того, независимо от стажа, вероятность и размер ошибок растут с возрастом, что может объясняться как ухудшением памяти, так и накоплением разнообразного жизненного опыта (включая возможные многочисленные переходы с одной работы на другую), на фоне которого дата начала текущей работы выглядит не такой важной датой, которую следует хорошо помнить.

Интересно отметить, что мужчины чаще дают несостоятельные ответы, чем женщины. Вероятно, это связано с психологическими различиями между полами: женщины считаются в среднем более аккуратными и внимательными к деталям, чем мужчины.

У работников с высшим и средним профессиональным образованием вероятность и размер ошибок меньше, чем у работников, имеющих только среднее образование. На наш взгляд, это связано с тем, что у более образованных работников в среднем более развиты когнитивные способности и память.

Среди профессионально-квалификационных групп различий практически нет. Заметно выделяется лишь группа руководителей, где вероятность и размер ошибки выше, чем у базовой группы – специалистов высшего уровня квалификации. Это вряд ли можно объяснить тем, что работники этой группы не держат в голове или не придают значения дате начала своей работы из-за того, что управляют большим штатом сотрудников. Результаты показывают, что наличие подчиненных, наоборот, уменьшает вероятность и размер ошибок в ответах на вопрос о дате начала работы, что указывает на то, что менеджеры в среднем более (само)организованы и лучше хранят и структурируют информацию. Возможно, большая забывчивость руководителей связана с тем, что в силу специфики деятельности дата начала их работы носит более условный характер, чем в случае наемных работников.

Вероятность и размер ошибки выше у работников, состоящих в браке. На наш взгляд, это опять же связано с тем, что у людей, состоящих в браке, память занята более важными вещами, чем датой начала работы (например, им нужно помнить дату вступления в брак). Кроме того, эти люди с большей вероятностью имеют детей, поэтому лично себе и своим жизненным ситуациям они уделяют гораздо меньше внимания.

Ответы работников средних предприятий с большей вероятностью подвержены ошибкам, чем ответы работников малых и крупных предприятий. Ответы работников государственных предприятий и организаций в большей мере подвержены ошибкам, чем ответы работников частных компаний. Кроме того, у занятых в сельской местности вероятность и размер ошибок выше, чем у занятых в городской местности. Наконец, самооценка здоровья никак не связана с вероятностью и размером ошибки.

Таким образом, несмотря на сложность интерпретаций некоторых корреляций, в общем можно сказать, что несостоятельность в данных по специальному стажу явно носит неслучайный характер. Как вероятность, так и размер ошибки значимо коррелируют с рядом индивидуальных характеристик, включая величину заработной платы, что необходимо учитывать при использовании показателей специального стажа в регрессионном анализе.

Варианты корректировки

К сожалению, не существует универсального и однозначно правильного способа исправления несоответствий в ответах респондентов в разные годы обследования. Исследователям не дано знать, когда респондент сообщил достоверную информацию, а когда умышленно или неумышленно (по забывчивости) сказал неправду. Если любая дополнительная информация недоступна, остается только строить различные предположения по поводу того, какой из ответов является правильным.

В случае несоответствия между датами начала текущей работы, указанными в разных интервью, можно сделать несколько альтернативных предположений. Пожалуй, наиболее естественное предположение состоит в том, что дата, указанная в интервью, которое находится ближе всего по времени к началу работы, является с большей вероятностью правильной. Это предположение основано на естественном свойстве человеческой памяти: чем дальше некоторое событие, случившееся в прошлом, отстоит от момента разговора, тем сильнее оно забывается. Например, если в опросе в 2009 г. респондент указал дату начала работы февраль 2005 г., а в опросе 2010 г. указан апрель 2004 г., то верить следует дате, указанной в опросе 2009 г. Такой способ корректировки мы будем называть «верим ближайшему интервью».

Другое возможное предположение состоит в том, что из двух дат верной является та, которая идет раньше. Это позволяет оценить верхнюю границу для стажа, выше которой стаж быть не может при любых вариантах ошибок. Такой способ корректировки мы будем называть «верим более ранней дате». В свою очередь, предположение о том, что верной является всегда та дата, которая идет позже, позволяет оценить нижнюю границу, ниже которой стаж быть не может при любых вариантах ошибок. Такой способ корректировки мы будем называть «верим более поздней дате».

Далее, используя эти три альтернативных предположения, мы рассчитываем три соответствующие переменные специального стажа, сравниваем их с переменной стажа, рассчитываемой на данных «как есть», а также сравниваем их между собой. Кроме того, мы сравниваем, как эти переменные ведут себя в разных эмпирических приложениях, где используется специальный стаж.

Однако прежде чем перейти к описанию полученных результатов, следует обсудить еще один способ корректировки ошибок в специальном стаже, доступный на данных РМЭЗ – ВШЭ, – это использование самостоятельного вопроса о смене места работы. Вопрос формулируется следующим образом: *«Пожалуйста, вспомните и скажите, Вы сменили место работы или профессию по сравнению с ноябрем предыдущего года или всё осталось по-прежнему?»* К вопросу предлагается несколько вариантов ответа: « *1) Профессия и место работы остались прежними; 2) Сменили профессию, но не сменили место работы; 3) Сменили место работы, но не сменили профессию; 4) Сменили и место работы, и профессию; 5) В ноябре предыдущего года не работал(а)».* Группу «мобильных» индивидов в данном случае будут составлять те, кто выбрал варианты ответов 3 или 4[7].

Потенциально ответы на этот вопрос могли бы помочь выбрать корректную дату начала работы в противоречивых случаях – Случай 2 и Случай 3.1. Однако принципиальный момент здесь состоит в том, что «помощь» возможна только в том случае, *если месяц ноябрь находится между двумя возможными датами начала работы*, как в Случае 2, так и в Случае 3.1. В качестве примера рассмотрим Случай 2 (рис. 4). Если месяц ноябрь находится строго после даты начала работы, указанной в интервью 1, то

[7] При использовании этого вопроса следует сделать ряд замечаний. Во-первых, данный вопрос не задавался в 1994 и 1995 гг., а в 1996 г. формулировка ответов не позволяет отделить тех, кто сменил работу, от тех, кто сменил профессию. Во-вторых, в 1996 и 1998 гг. респондента просили сравнить свое положение не с ноябрем, а с декабрем прошлого года. В-третьих, следует иметь в виду, что в 1998 г. был другой порядок ответов. В-четвертых, несмотря на то, что вариант ответа «в предыдущем году не работал» присутствует в вопросниках, в распространяемом файле микроданных соответствующая категория присутствует не за все годы. Наконец, в-пятых, вопрос не задается индивидам, занятость которых носит нерегулярный характер (положительно ответившим на вопрос j56).

либо обе даты начала работы допустимы (в случае, если индивид не менял место работы), либо обе даты оказываются в противоречии с ответом на вопрос о смене места работы (в случае, если он менял место работы). То же самое происходит в случае, когда ноябрь идет раньше обеих дат. Но когда ноябрь находится между датами, то если индивид не менял место работы с ноября, есть основания верить той дате, которая идет раньше или совпадает с ноябрем. В случае 2 это дата, указанная в интервью 2. Если же индивид указал, что он менял место работы, то следует верить той дате, которая идет позже. Это дата, указанная в интервью 1. Аналогично происходит принятие решения в Случае 3.1 (рис. 5), только дата 1 и дата 2 меняются местами.

Сколько противоречивых наблюдений в специальном стаже можно разрешить, используя вопрос о смене места работы? Расчеты показывают, что в целом за период 1995–2014 гг. в Случае 2 можно верифицировать только 1245 наблюдений из 15230 (примерно 8%), а в случае 3.1 только 322 наблюдения из 13655 (примерно 2,5%). В целом можно верифицировать лишь немногим более 5% противоречивых наблюдений по специальному стажу. Таким образом, использование вопроса о смене места работы практически никак не помогает скорректировать ошибки в специальном стаже[8]. Отметим, что Случай 1 (сохранение прежнего места работы) и Случай 3.2 (смена работы), в которых предполагается отсутствие ошибок в стаже, достаточно хорошо согласуются с ответами на вопрос о смене места работы. Для Случая 1 доля соответствия составляет 99%, а для случая 3.2 уровень соответствия составляет почти 90%.

Влияние ошибок в измерении стажа на результаты анализа

В данном разделе мы рассматриваем, как меняются результаты дескриптивного и регрессионного анализа в зависимости от способа корректировки несостоятельных во времени ответов респондентов.

Рисунок 8 представляет динамику среднего стажа на данных РМЭЗ – ВШЭ в рассматриваемый период, рассчитанного с использованием переменной стажа «как есть», а также с использованием трех альтернативных вариантов корректировки этой переменной, описанных выше. В 1994 г. все линии выходят из одной точки примерно 100 месяцев, так как для данных 1994 г. корректировки не применялись. Для большинства лет первый способ корректировки («верим ближайшему интервью») дает значения среднего стажа ниже, чем значения, существующие без какой-либо корректировки. Лишь для четырех

[8] Отметим, что, к сожалению, вопрос о смене места работы с трудом может быть полезным и в проверке тех случаев, которые нельзя проверить с помощью сопоставления дат начала работы во времени (это респонденты, которые в прошлом раунде не работали или вообще не присутствовали в выборке). Согласно нашим расчетам, почти три четверти «новых» наблюдений по стажу не согласуются с ответами на вопрос о смене работы. Этот момент требует дальнейшего изучения.

волн обследования из 19 средний стаж после этой корректировки оказывается выше. Однако разница между средними практически всегда оказывается несущественной. В среднем за весь период она составляет менее одного месяца и лишь в трех случаях (2001, 2003 и 2004 гг.) превышает один месяц.

Второй способ корректировки («верим ранней дате») задает верхнюю границу для среднего стажа. Здесь разница с вариантом «как есть» уже заметно больше, превышение в среднем за период составляло 4,5 месяца, а в 1990-е – начале 2000-х годов составляло 6–9 месяцев. Наконец, третий способ корректировки («верим поздней дате») задает нижнюю границу для среднего стажа. Нижняя граница среднего стажа отставала от варианта «как есть» в среднем на пять месяцев, в 1990-е – начале 2000-х годов – на 6–9 месяцев, а в последующий период на 4–5 месяцев. Верхняя (зеленая) и нижняя (темно синяя) линии на рис. 8 задают границы, в рамках которых средний стаж мог принимать любые значения в каждой из рассматриваемых волн РМЭЗ – ВШЭ. Этот диапазон был достаточно широким: в среднем за период его ширина составляла почти 10 месяцев, в 1990-е годы и в 2000–2001 гг. она превышала один год, а в последующий период была на уровне 8–9 месяцев. Наиболее выдающимся примером является 1995 г., когда средний стаж на данных «как есть» составлял примерно 93 месяца (7,75 лет), верхняя граница была 102,2 месяца (8,5 лет), а нижняя граница – 84,4 месяца (7 лет). В последний доступный на данный момент год обследования, 2014-й, средний стаж на данных «как есть» составлял 90,2 месяца (7,5 лет), его верхняя граница была равна 94 месяцам (7,8 лет), а нижняя – 86,2 месяца (7,2 года).

Если говорить о динамике среднего стажа, то данные «как есть» и все варианты корректировки задают схожий профиль: средний стаж сокращался с 1994 г. примерно до 2006–2007 гг., но затем стал расти и к 2013 г. достиг уровня примерно 2000 г. Тем не менее большой интервал между верхней и нижней границей допускает любую другую возможную динамику. Например, это наличие почти монотонного повышательного или понижательного тренда или же постоянство среднего стажа в рассматриваемый период. Таким образом, следует помнить, что любые тенденции в средней продолжительности специального стажа, которые можно построить на данных РМЭЗ – ВШЭ, имеют определенную долю условности.

Далее мы рассмотрим, как меняется распределение занятых в данных РМЭЗ – ВШЭ по группам с различной продолжительностью стажа в зависимости от различных способов корректировки ошибок (рис. 9). Доля занятых, имеющих специальный стаж менее одного года, нечувствительна к ошибкам измерения. На данных «как есть» эта доля составляет 21,6% и примерно столько же получается, если мы «верим ближайшему интервью»,

«верим ранней дате» или «верим поздней дате». При расчете доли занятых со специальным стажем от одного года до трех лет отличия между способами корректировки гораздо более заметны. Разница между расчетом на данных «как есть» и если мы «верим ближайшему интервью» оказывается незначительной (21% против 21,6% соответственно), но интервал между верхней (19,3%) и нижней границей (23,3%) составляет уже 4 п.п. Разница в доле занятых в интервале от трех до пяти лет опять же, как и в интервале до одного года, оказывается несущественной – все способы расчета дают примерно 13%. Несколько сильнее отличия в доле занятых от 5 до 10 лет – разница между верхней и нижней границей составляет 1 п.п. Гораздо более явно различия между разными способами корректировки проявляются для группы занятых с максимальным уровнем специального стажа – 10 лет и более. Если на данных «как есть» эта доля составляет примерно 25,9%, что примерно соответствует доле, если мы «верим ближайшему интервью» (25,7%), то верхняя граница этой доли может достигать 27,6%, а нижняя – опускаться до 24,2%. Таким образом, диапазон значений в данном случае составляет более 3 п.п.

Как отмечалось в предшествующих исследованиях, распределение занятых по группам с различной длительностью специального стажа на данных РМЭЗ – ВШЭ отличается от распределений, которые можно построить с использованием двух других источников микроданных – Обследования населения по проблемам занятости и Обследования заработных плат по профессиям (например, [Капелюшников, 2011]). Доля занятых с более коротким стажем на данных РМЭЗ – ВШЭ гораздо больше, чем на данных ОНПЗ или ОЗПП, а доля занятых с более длинным стажем гораздо меньше. В какой мере это отличие данных РМЭЗ – ВШЭ может объяснить наличие в них ошибок измерения стажа?

На рис. 10 представлены распределения занятых по интервалам специального стажа «как есть», а также верхняя и нижняя граница, рассчитанные на данных РМЭЗ – ВШЭ, в сравнении с распределениями на данных ОНПЗ и ОЗПП в 2013 г. (2013 г. выбран для сравнения неслучайно – это последний год, на который доступны данные всех трех обследований, так как ОЗПП проводится один раз в два года и в 2014 г. оно не проводилось). Прежде всего, мы подтверждаем, что доля работников с коротким стажем на данных РМЭЗ – ВШЭ заметно выше, а доля работников с длительным стажем заметно ниже, чем на данных ОНПЗ или ОЗПП. Так, в 2013 г. на данных ОНПЗ (по состоянию на октябрь) доля работников с коротким стажем была 11,5%, на данных ОЗПП – 12,5%, а на данных РМЭЗ – ВШЭ – 20%. В то же время доля работников со стажем 10 лет и более на данных ОНПЗ составляла 35%, на данных ОЗПП – 33%, а на данных РМЭЗ – ВШЭ – лишь

26,5%. Как мы уже отмечали выше, ошибки измерения в данных РМЭЗ – ВШЭ практически не влияют на долю занятых с коротким стажем, однако они меняют доли занятых со стажем 10 лет и более, а также со стажем от одного до трех лет. Результаты, наиболее близкие к тем, которые получаются на данных ОНПЗ и ОЗПП, дает вариант корректировки «верим ранней дате». Доля занятых со стажем 10 лет и более вырастает в этом случае с 26,4 до 29%. Однако существенная разница все же остается: в сравнении с данными ОНПЗ она составляет 7 п.п. (вместо 9,6% без корректировки), а в сравнении с данными ОЗПП – 5 п.п. (вместо 7,6%). Таким образом, ошибки измерения могут объяснить примерно треть отставания данных РМЭЗ – ВШЭ от данных ОНПЗ или ОЗПП по доле занятых со стажем более 10 лет.

Далее мы переходим к вопросу о том, как ошибки измерения и их возможные способы корректировки могут влиять на результаты оценивания отдачи от специального стажа. Для исследования этого вопроса мы оцениваем уравнение заработной платы следующего вида [Mincer, 1974]:

$$\ln Wage_i = \beta_0 + \beta_1 Educ_i + \beta_2 Age_i + \beta_3 Age_sq_i + \beta_4 Ten_i + \beta_5 Ten_sq_i + \beta_6 \ln(\text{worked hours}) + \varepsilon_i, \quad (3)$$

где $\ln Wage_i$ – логарифм заработной платы работника i за месяц; $Educ$ – его уровень образования, Age и Age_sq – возраст и возраст в квадрате; Ten и Ten_sq – специфический стаж работы и специфический стаж работы в квадрате соответственно. Экономическая теория предсказывает, что коэффициент β_4 должен быть положительным, а β_5 – отрицательным, так как накопление человеческого капитала замедляется с течением времени и, начиная с какого-то момента, его выбытие начинает превышать его прирост.

Данное уравнение мы оцениваем, используя четыре разные переменные специального стажа, полученные с использованием альтернативных предположений о природе ошибок, на пуле данных за все годы с 1995 по 2014 г. без учета и с учетом индивидуальных фиксированных эффектов работников. Результаты оценивания отдельно для мужчин и женщин приведены в табл. 4.

Прежде всего отметим, что практически во всех спецификациях переменная «стаж» имеет значимый положительный коэффициент, а переменная «стаж в квадрате» имеет значимый отрицательный коэффициент, что согласуется с предсказываемой теорией человеческого капитала положительной, но убывающей отдачей от специального стажа. Согласно нашим оценкам, для мужчин на данных «как есть» отдача от дополнительного года специального стажа составляет примерно 1,2%. Эта величина оказывается практически нечувствительной к трем рассматриваемым нами способам корректировки ошибок в переменной специального стажа. Однако в спецификации, учитывающей все

ненаблюдаемые постоянные во времени индивидуальные характеристики работников (модель FE), чувствительность коэффициента при специальном стаже оказывается выше[9]. Оценка на данных «как есть» составляет 2,1%, но если мы «верим ранней дате», то она составляет 2,4%, а если мы «верим поздней дате», то 1,9%. Другими словами, отдача от специального стажа может меняться в диапазоне от 1,9 до 2,4%.

Качественно схожие выводы можно сделать в случае женщин. На данных «как есть» отдача от дополнительного года составляет 1%, и она может лежать в границах от 0,9 до 1,1%. Однако чувствительность отдач, полученных в модели с фиксированными эффектами, к разным способам корректировки ошибок оказывается еще выше, чем у мужчин. На данных «как есть» отдача составляет 0,8%, но на самом деле она может быть в 2 раза выше и составлять 1,6% (в случае, если мы «верим ближайшему интервью»). Отметим, что б*о*льшая чувствительность оценок отдачи от специального стажа к разным способам корректировки ошибок в случае, когда оценка происходит на панельных данных, согласуется с качественными выводами западных исследований и объясняется тем, что придание б*о*льшей состоятельности данным по стажу во времени влияет главным образом на вариацию «внутри индивидов во времени» (вариация “within”) и достаточно слабо влияет на вариацию «между индивидами» (вариацию “between”) (см. [Brown, Light, 1992]).

4. Основные выводы и заключение

В данной работе анализируются проблемы измерения специального стажа работников на данных РМЭЗ – ВШЭ. Эти данные широко используются в различных эмпирических исследованиях, в том числе для оценивания отдачи от специального стажа и анализа трудовой мобильности в России, однако до сих пор вопрос об ошибках измерения стажа не изучался и даже не поднимался. Полученные нами результаты можно кратко сформулировать следующим образом.

Во-первых, точная дата начала текущей работы (год и месяц) неизвестна примерно для 13% занятых респондентов. Это изначально ограничивает рабочую выборку в любых исследованиях, использующих информацию о специальном стаже, и потенциально может влиять на получаемые результаты, так как выбытие этих наблюдений происходит неслучайно.

[9] Отметим, что учет ненаблюдаемых постоянных индивидуальных характеристик работников с помощью модели с фиксированными эффектами (трансформация данных “within”) практически всегда увеличивает отдачу от специального стажа. Возможно, это связано с тем, что более производительные работники имеют в среднем более высокий специальный стаж. Более тщательное изучение этого вопроса выходит за рамки данной работы.

Во-вторых, достаточно распространена несостоятельность ответов респондентов во времени. Ошибкам такого рода подвержены около 25% всех наблюдений по специальному стажу. В первых волнах обследования РМЭЗ – ВШЭ, проводившихся в 1990-е годы, распространенность ошибок достигала 40%, но в 2000-е годы доля ошибок сократилась до уровня 25–30%. Однако тенденции ежегодного сокращения доли ошибок нам обнаружить не удалось. Если к этому добавить тех респондентов, несостоятельность данных во времени по которым проверить по разным причинам нельзя, и предположить, что вероятность ошибки среди них примерно такая же, то доля ошибок в среднем за весь период возрастет примерно до 34%, а в начале периода она превысит 50%.

В-третьих, средний размер ошибки был весьма существенным. В целом за весь период размер *абсолютной* ошибки составлял почти три года. Так же как и распространенность ошибок, этот показатель в 1990-е годы был выше, чем в 2000-е, но, опять же, явной тенденции к ежегодному сокращению нам обнаружить не удалось. Медианное значение абсолютной ошибки составляло один год, т.е. примерно половина всех ошибок была по размеру меньше одного года, а половина – больше одного года. Доля наблюдений с величиной ошибки пять лет и более находилась весь период на уровне 20–25%. Средний размер *относительной* ошибки составлял примерно 55% (в процентах от длительности индивидуального стажа).

В-четвертых, результаты эконометрического анализа говорят о том, что как вероятность, так и размер ошибки не являются абсолютно случайными, а значимо коррелируют с некоторыми характеристиками индивидов. Прежде всего они связаны с размером заработной платы, что указывает на возможность смещения оценок отдачи от специального стажа, получаемых при стандартном оценивании уравнения заработной платы минцеровского типа. По всей видимости, происходит недооценивание коэффициента при специальном стаже (attenuation bias), что в какой-то мере ответственно за низкие и даже отрицательные отдачи от специального стажа, особенно в 1990-е и в начале 2000-х годов, когда вероятность и размер ошибок были выше. Кроме того, вероятность и размер ошибок растут с возрастом и выше у мужчин, чем у женщин.

В-пятых, мы применили разные способы корректировки существующих ошибок и посмотрели, как это влияет на получаемые результаты дескриптивного и регрессионного анализа. Для каждого года мы также оценили верхнюю и нижнюю границы специального стажа, выше и ниже которой стаж быть не может при любой ошибке измерения. Согласно нашим оценкам, в 1990-е годы величина среднего стажа может меняться на 12–17 месяцев в зависимости от способа корректировки, а в 2000-е годы – на 7–9 месяцев. В целом этот анализ свидетельствует о том, что из-за наличия ошибок измерения любая тенденция

в изменении среднего уровня специального стажа, наблюдаемая на данных РМЭЗ – ВШЭ, имеет определенную долю условности. Ошибки измерения специального стажа на данных РМЭЗ – ВШЭ также могут объяснить примерно треть отставания доли занятых с длительным стажем от показателей на данных ОНПЗ или ОЗПП.

Оценки отдачи от специального стажа, получаемые на годовых кросс-секциях и на пуле данных, практически не меняются при разных способах корректировки. Однако оценки, получаемые методом фиксированных эффектов, меняются достаточно сильно. Особенно сильно они меняются у женщин: если без какой-либо корректировки данных отдача от специального стажа у женщин составляет примерно 0,8%, то после корректировки она может достигать 1,6%.

В целом наша работа показывает, что исследователям следует с большой аккуратностью работать с данными по специальному стажу из обследования РМЭЗ – ВШЭ и стараться учитывать возможные ошибки измерения в своих расчетах и выводах.

Рисунки

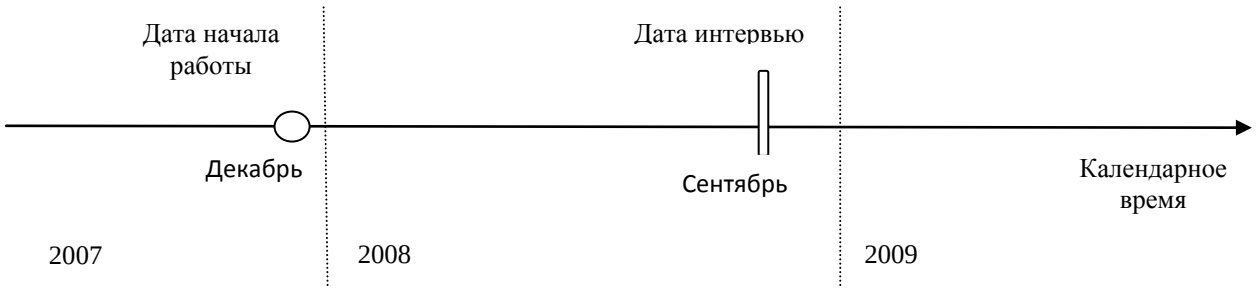

Рис. 1. Пример ситуации, когда измерение стажа без учета информации о месяце начала работы дает ошибочный результат

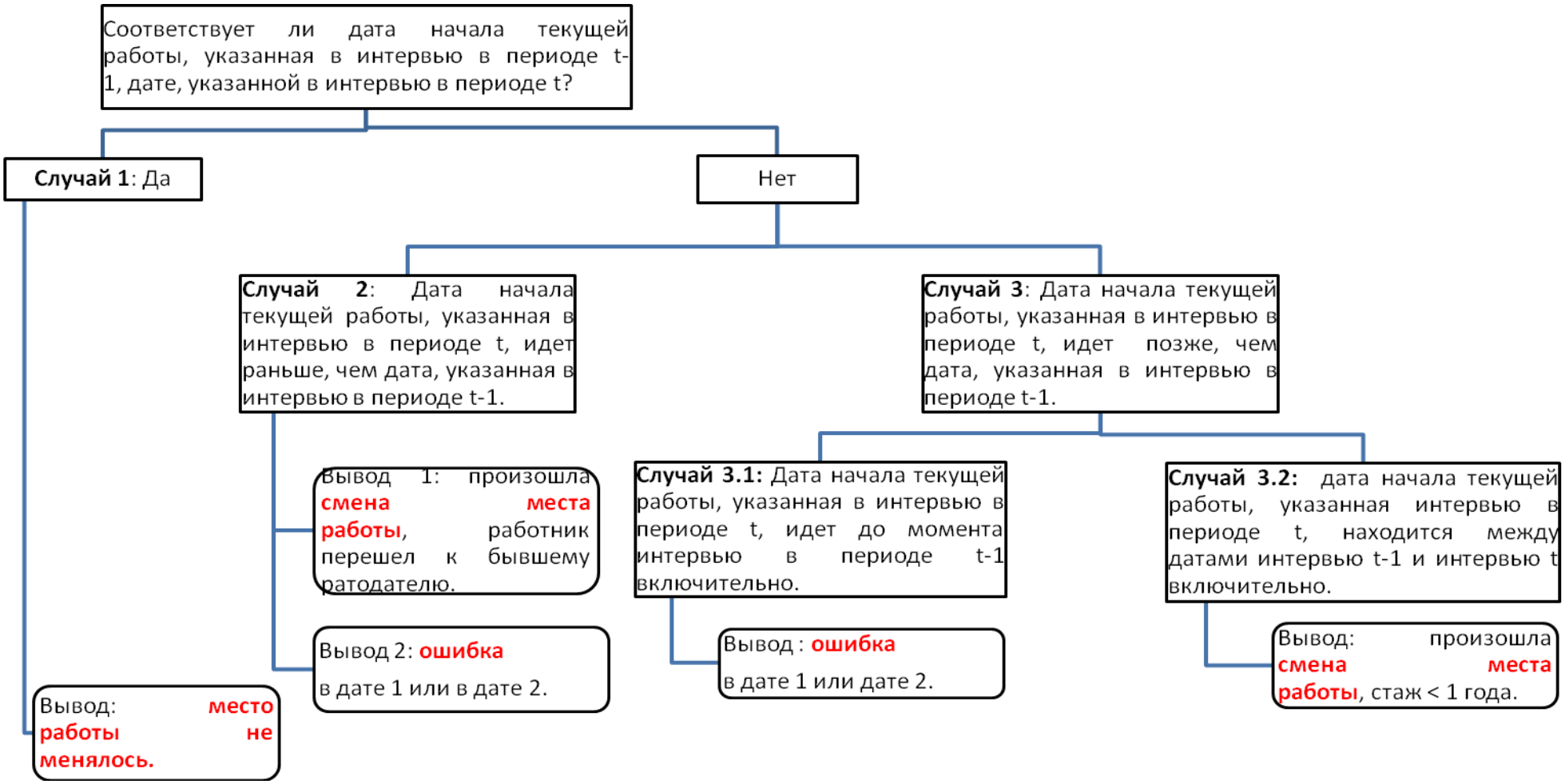

Рис. 2. Возможные случаи (не)соответствия между датами начала текущей работы, указанными респондентом в двух последовательных раундах опроса

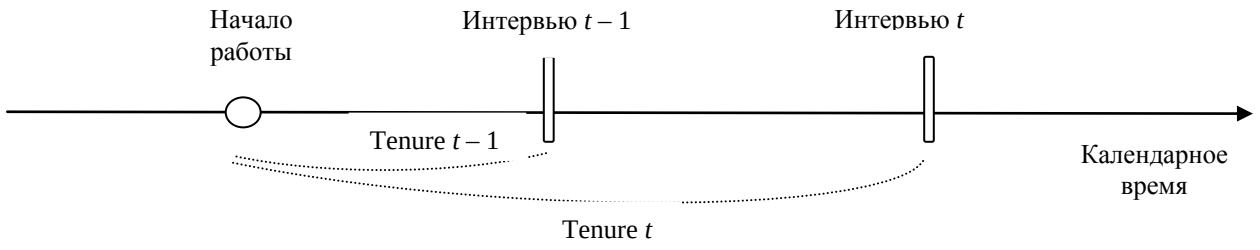

Рис. 3. Случай 1: дата начала текущей работы совпадает в двух интервью

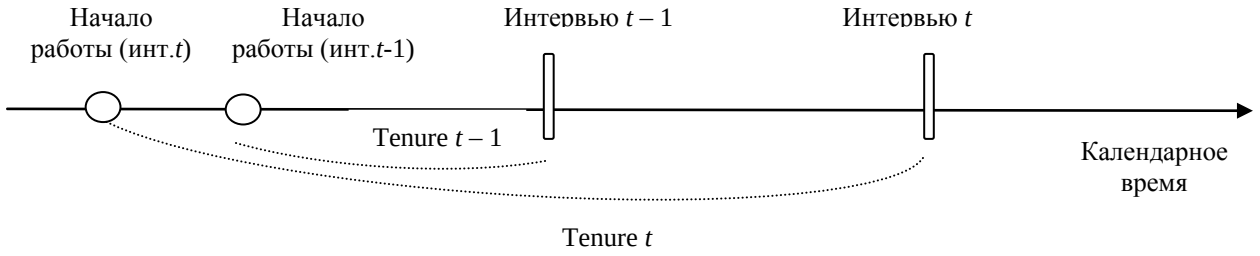

Рис. 4. Случай 2: дата начала текущей работы, указанная в интервью t, идет раньше, чем дата, указанная в интервью t – 1

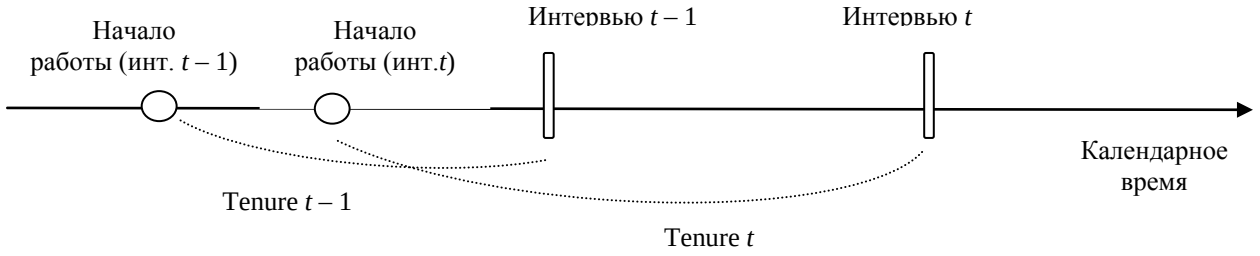

Рис. 5. Случай 3.1: дата начала текущей работы, указанная в интервью t, идет позже, чем дата, указанная в интервью t – 1, и при этом она идет до момента интервью t – 1 включительно

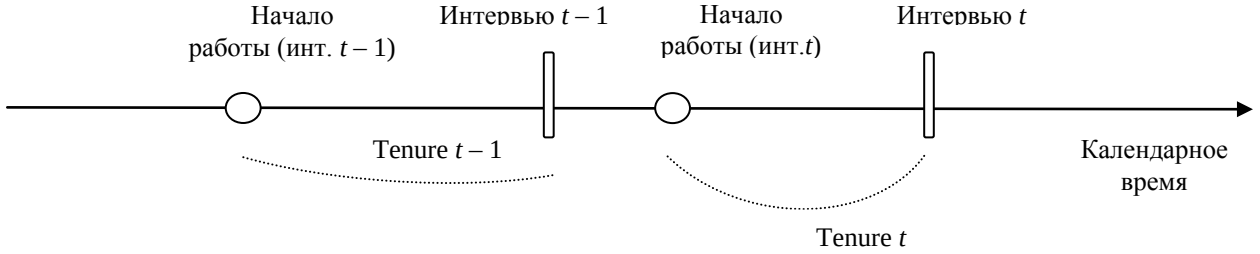

Рис. 6. Случай 3.2: дата начала текущей работы, указанная в интервью t, идет позже, чем дата, указанная в интервью t – 1, и при этом она идет в периоде между интервью t – 1 и интервью t включительно

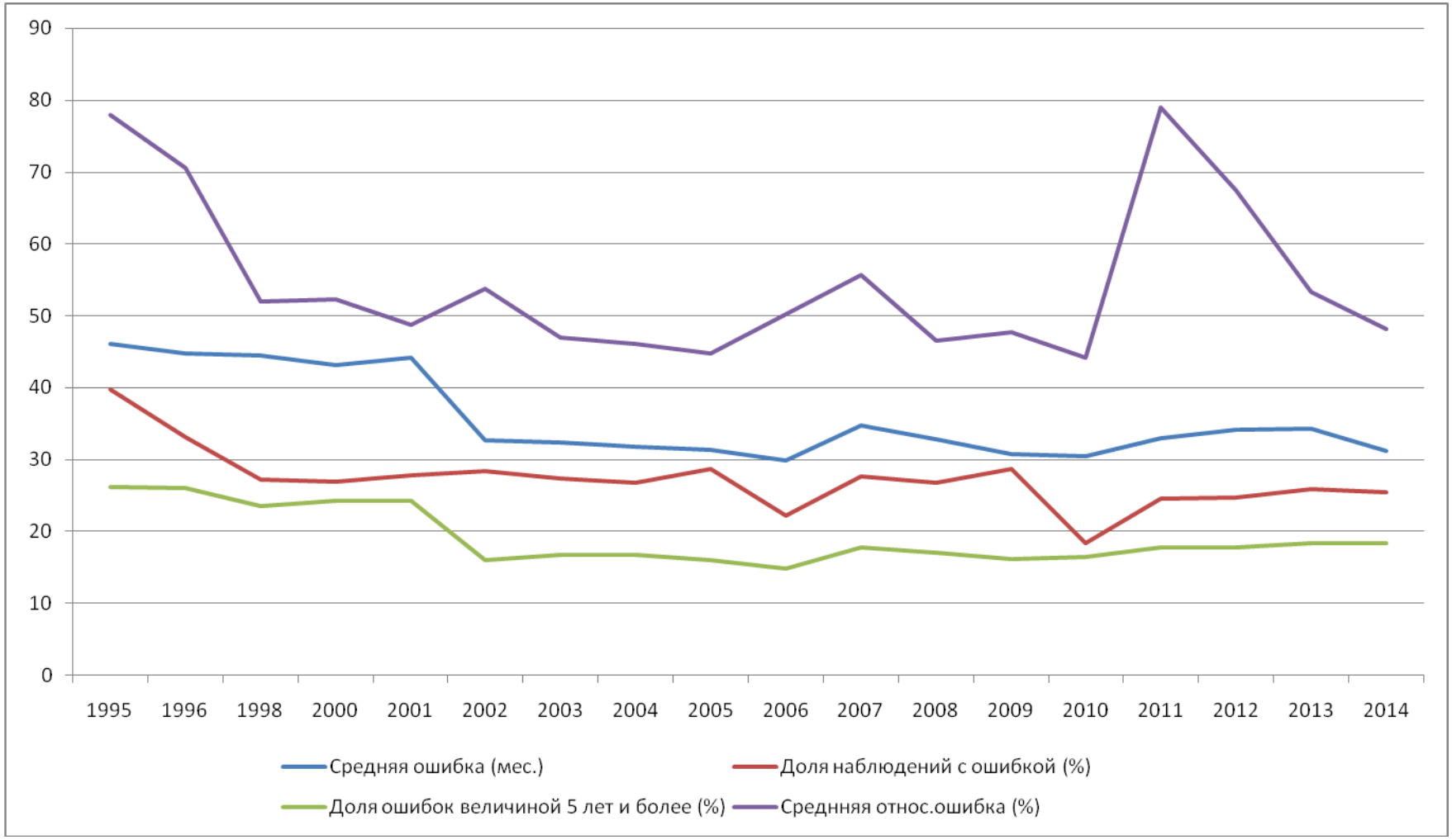

Рис. 7. Динамика распространенности и размера ошибок в данных по специальному стажу во времени, РМЭЗ – ВШЭ, 1995–2014 гг.

Рис. 8. Средний стаж на данных РМЭЗ – ВШЭ с использованием разных способов корректировки данных, 1994–2014 гг.

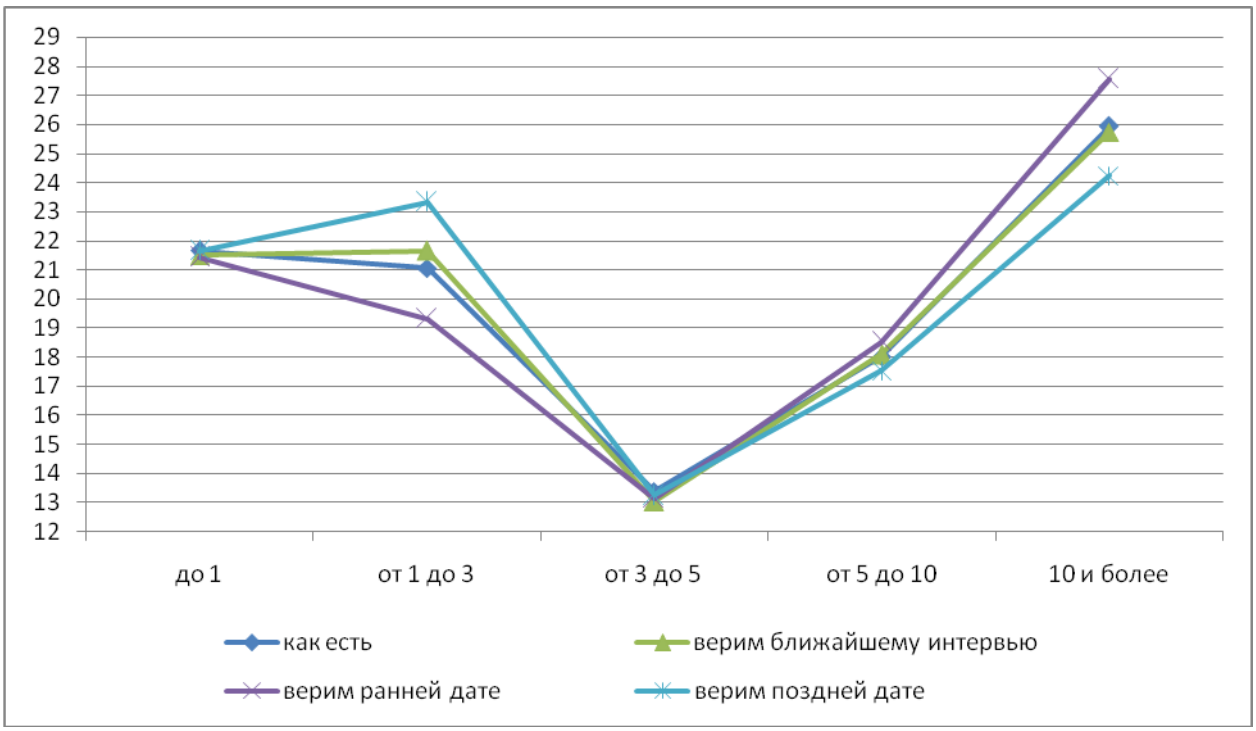

Рис. 9. Распределение занятых по группам с разной продолжительностью специального стажа на данных РМЭЗ – ВШЭ при использовании разных способов корректировки ошибок, 1995–2013 гг.

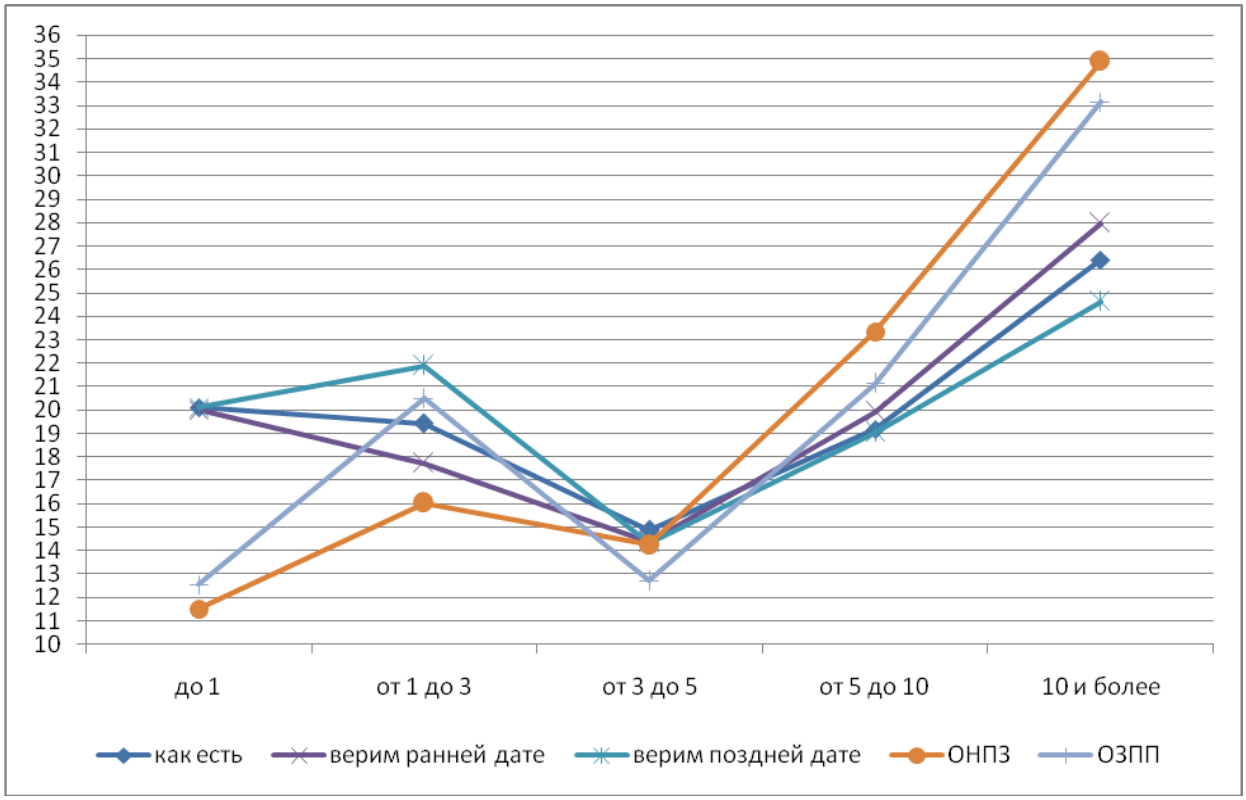

Рис. 10. Распределение занятых по группам с разной продолжительностью специального стажа на данных РМЭЗ – ВШЭ при использовании разных способов корректировки ошибок в сравнении с данными ОНПЗ и ОЗПП, 2013 г.

Таблицы

Таблица 1. Размер выборки и распространенность ошибок в данных по специальному стажу

	1995	**1996**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**1995–2014**
Общее число занятых	4715	4326	3997	3944	4143	4116	4123	4109	3804	5141	5004	4809	4630	8131	8089	8099	7592	6009	94781
% занятых с известным спец. стажем (известен хотя бы год начала работы)	93,9	93,3	92,0	89,8	88,9	89,8	90,0	90,3	92,1	91,9	93,2	94,5	93,5	89,9	93,8	94,5	94,8	94,7	92,5
% занятых с известным спец. стажем (известен как год, так и месяц начала работы)	86,7	84,9	86,0	84,7	83,9	84,7	84,0	87,1	88,0	87,2	88,1	92,4	89,7	85,6	89,2	89,2	87,6	88,2	87,3
Распределение по случаям (%)																			
Случай 1 (соответствие дат, работа не менялась)	24,4	27,7	21,3	21,9	27,6	31,3	33,2	33,1	34,9	27,8	34,1	36,3	39,5	27,4	36,0	36,0	37,5	40,1	31,4
Случай 2 (несоответствие дат, ошибка)	19,3	16,9	14,1	14,3	16,9	15,8	14,8	15,3	15,1	11,5	14,4	14,5	13,9	9,7	12,3	13,1	13,7	13,4	13,4
Случай 3.1 (несоответствие дат, ошибка)	20,5	16,1	13,1	12,6	10,9	12,6	12,6	11,5	13,6	10,7	13,3	12,3	14,8	8,7	12,2	11,7	12,1	12,1	12,0
Случай 3.2 (несоответствие дат, смена работы)	9,6	9,9	14,3	14,2	10,3	11,4	12,3	11,6	13,1	9,9	12,3	11,0	11,2	8,1	9,4	10,2	10,7	10,6	10,4
Случай 4 (спец.стаж нельзя проверить)	26,2	29,3	37,2	37,0	34,2	28,9	27,1	28,5	23,2	40,0	26,0	25,9	20,6	46,1	30,0	29,0	26,0	23,8	32,8
Доля ошибок, %																			
Ошибки всего, % (случай 2+ случай 3.1)	39,7	33,1	27,2	26,9	27,8	28,4	27,4	26,9	28,8	22,3	27,7	26,9	28,7	18,4	24,6	24,8	25,8	25,5	25,4
Оценка доли ошибок в случае 4	10,4	9,7	10,1	10,0	9,5	8,2	7,4	7,6	6,7	8,9	7,2	7,0	5,9	8,5	7,4	7,2	6,7	6,1	8,3
Итого % ошибок	50,2	42,7	37,4	36,8	37,4	36,6	34,8	34,5	35,5	31,2	34,9	33,8	34,6	26,9	31,9	32,0	32,5	31,6	33,7

Таблица 2. Анализ размера ошибок в данных РМЭЗ – ВШЭ по специальному стажу, 1995–2014 гг.

Абсолютная ошибка (случай 2 + 3.1)	**1995**	**1996**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Доля ошибок величиной:																		
1 месяц	17,2	17,2	19,6	18,9	19,6	22,9	24,9	21,2	21,2	22,6	20,8	21,4	20,5	20,5	18,8	18,5	20,9	20,1
2–5 месяцев	11,6	11,3	12,0	11,7	10,5	12,0	12,9	11,6	12,3	14,1	13,3	13,8	14,7	14,8	13,8	15,8	14,6	15,1
6–11 месяцев	11,7	11,2	10,7	11,4	11,4	11,3	10,6	11,9	11,9	12,1	12,2	11,6	12,7	13,7	13,7	14,3	12,7	14,2
12–23 месяца	13,6	17,2	17,2	16,8	19,6	23,1	19,6	21,0	19,5	19,9	18,5	17,3	18,3	17,4	16,6	14,2	14,7	15,7
24–59 месяцев	19,7	17,1	17,1	17,0	14,7	14,8	15,4	17,6	19,1	16,4	17,5	18,9	17,7	17,3	19,5	19,5	18,9	16,6
60–119 месяцев	12,3	11,0	8,1	9,5	10,6	7,0	8,4	8,2	8,1	7,2	9,1	8,4	8,8	9,0	9,6	9,2	10,0	10,5
более 120 месяцев	13,9	15,0	15,4	14,8	13,7	8,9	8,3	8,6	7,9	7,7	8,7	8,7	7,4	7,4	8,1	8,5	8,3	7,8
Среднее	46,1	44,8	44,4	43,1	44,1	32,6	32,4	31,7	31,3	29,8	34,7	32,8	30,7	30,3	32,9	34,1	34,3	31,1
Медиана	13	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
Мин	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Макс	472	522	517	460	503	491	505	447	492	449	591	493	512	512	477	541	541	475
Доля ошибок величиной 5 лет и более	26,2	26,0	23,5	24,3	24,3	16,0	16,7	16,8	16,1	14,8	17,7	17,1	16,2	16,4	17,8	17,8	18,3	18,3
Относительная ошибка (среди 2 + 3.1)	**1995**	**1996**	**1998**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
Среднее	78,0	70,5	52,0	52,3	48,8	53,8	47,0	46,1	44,8	50,3	55,7	46,6	47,8	44,2	79,0	67,6	53,3	48,2
Медиана	23,5	22,2	17,6	18,6	20,6	17,6	15,0	19,4	20,0	17,6	20,4	19,0	18,0	19,1	21,4	20,7	19,0	18,8
Мин	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
Макс	2247,6	2871,4	1914,8	1908,7	1688,2	3816,7	2125,0	2981,8	1725,0	2300,0	3892,3	2606,7	1690,9	1682,4	34400	8420,0	2878,6	3316,7

Таблица 3. Факторы, связанные с вероятностью и размером ошибки в специальном стаже, оценки на данных РМЭЗ – ВШЭ, 1995–2014 гг.

Факторы	**Вероятность ошибки**		**Величина ошибки**	
	coef	**se**	**coef**	**se**
ЗП (лог)	0,008***	0,003	0,024***	0,009
Спец.стаж (месяцы)	0,000***	0,000	0,001***	0,000
Возраст (лет)	0,002***	0,000	0,006***	0,001
Пол (мужчины)	0,018***	0,004	0,057***	0,013
Образование (база: полное среднее)				
ниже неполн. ср.	–0,032	0,028	–0,101	0,092
неполн. ср.	0,014*	0,008	0,043*	0,023
неполн. ср.+ПТУ	0,016	0,010	0,048	0,030
полн. ср.+ПТУ	0,011**	0,006	0,034**	0,017
ср. проф.	–0,011**	0,005	–0,034**	0,016
высшее	–0,014**	0,006	–0,043**	0,018
Занятие (база: специалисты высшего уровня квалификации)				
Военнослужащие	–0,013	0,021	–0,039	0,067
Руководители	0,021**	0,009	0,065**	0,027
Специалисты среднего уровня	–0,007	0,006	–0,022	0,017
Служащие и клерки	–0,001	0,008	–0,004	0,025
Работники сферы услуг	0,007	0,007	0,023	0,022
Квал. рабочие с\х	–0,001	0,030	–0,004	0,093
Квал. рабочие промышленности	0,010	0,007	0,031	0,022
Операторы, аппаратчики, машинисты	0,012*	0,007	0,037*	0,022
Неквал. рабочие	–0,010	0,007	–0,030	0,023
Состоит в браке	0,025***	0,004	0,078***	0,012
Размер предприятия (база: более 1000 чел.)				
меньше 10	0,003	0,006	0,010	0,019
от 11 до 50	0,021***	0,005	0,064***	0,014
от 51 до 100	0,017***	0,006	0,050***	0,018
от 101 до 500	0,015***	0,005	0,046***	0,015
от 501 до 1000	0,008	0,008	0,023	0,024
Форма собственности предприятия (государственная)	0,012***	0,004	0,036***	0,011
Часы работы (лог)	–0,001	0,005	–0,002	0,014
Самооценка здоровья (база: среднее)				
Очень хорошее	–0,026**	0,013	–0,084**	0,042
Хорошее	–0,011***	0,004	–0,034***	0,012
Плохое	–0,008	0,008	–0,026	0,024
Очень плохое	–0,058*	0,035	–0,191	0,123
Есть кто-то в подчинении	–0,018***	0,005	–0,057***	0,014
Село	0,025***	0,007	0,077***	0,021
Число наблюдений	74 903			

***Значимость на 1%-м уровне, ** – значимость на 5%-м уровне, * – значимость на 10%-м уровне. В столбце «Вероятность ошибки» приведены предельные эффекты (рассчитываются в «средней» точке). В обоих уравнениях включаются годовые и региональные (на уровне psu) дамми.

Таблица 4. Отдача от дополнительного года специального стажа при разных способах корректировки ошибок измерения стажа, данные РМЭЗ – ВШЭ, 1995–2014 гг.

	Мужчины		**Женщины**	
	Pool 1995–2014	**FE 1995–2014**	**Pool 1995–2014**	**FE 1995–2014**
	«как есть»			
Стаж	0,012***	0,021***	0,010***	0,008*
	(0,002)	(0,005)	(0,001)	(0,005)
Стаж в квадрате/100	–0,016***	–0,055***	–0,010***	–0,019
	(0,005)	(0,016)	(0,004)	(0,014)
	«верим ближайшему интервью»			
Стаж	0,013***	0,022***	0,010***	0,016***
	(0,002)	(0,005)	(0,001)	(0,005)
Стаж в квадрате/100	–0,015***	–0,042**	–0,012***	–0,035**
	(0,005)	(0,017)	(0,004)	(0,014)
	«верим ранней дате»			
Стаж	0,012***	0,024***	0,009***	0,013***
	(0,002)	(0,005)	(0,001)	(0,005)
Стаж в квадрате/100	–0,015***	–0,055***	–0,009**	–0,031**
	(0,005)	(0,017)	(0,004)	(0,014)
	«верим поздней дате»			
Стаж	0,013***	0,019***	0,011***	0,011**
	(0,002)	(0,006)	(0,001)	(0,005)
Стаж в квадрате/100	–0,016***	–0,043**	–0,013***	–0,024
	(0,006)	(0,018)	(0,004)	(0,015)

*** Значимость на 1%-м уровне, ** – значимость на 5%-м уровне, * – значимость на 10%-м уровне. Во всех спецификациях включаются региональные дамми (на уровне psu). В скобках приведены стандартные ошибки, рассчитанные методом, робастным к гетероскедастичности с корректировкой на кластеризацию, кластером выступает индивид.

Литература

Капелюшников Р. (2011) Эволюция человеческого капитала в России: 1990–2000-е годы // Российский работник: образование, профессия, квалификация / под ред. В.Е. Гимпельсона, Р.К. Капелюшникова. М.: Изд. дом. ВШЭ.

Карабчук Т. (2009) Детерминанты стабильности занятости в России и Восточной Германии: сравнительный анализ // Экономическая социология. Т. 10. № 2. С. 12–53.

Мальцева И. (2009) Трудовая мобильность и стабильность: насколько высока отдача от специфического человеческого капитала в России? // Экономический журнал Высшей школы экономики. Т. 13. № 2. С. 243–278.

Bassanini A., Garnero A. (2013) Dismissal protection and worker flows in OECD countries: Evidence from cross-country/cross-industry data // Labour Economics. Vol. 21. P. 25–41.

Bound J., Brown C., Mathiwetz N. (2001) Measurement Error in Survey Data // Handbook of Econometrics. Ch. 59 / J. Heckman, E. Leamer (eds). Vol. 5. P. 3707–3843.

Brown J., Light A. (1992) Interpreting Panel Data on Job Tenure // Journal of Labor Economics. Vol. 10. No. 3. P. 219–257.

Burgess S. (1998) The reallocation of labour: an international comparison using job tenure data, Mimeo, CEPR, LSE.

Cheidvasser S., Benítez-Silva H. (2007) The Educated Russian's Curse: Returns to Education in the Russian Federation during the 1990s // Labour. Vol. 21. No. 1. P. 1–41.

Lehmann H., Wadsworth J. (2000) Tenures That Shook the World: Worker Turnover in Russia, Poland, and Britain // Journal of Comparative Economics. Vol. 28. No. 4. P. 639–664.

Mincer J. (1974) Schooling, Experience and Earnings. N.Y.: Columbia University Press.

OECD (2009) Employment Outlook. OECD, Paris.

Oshchepkov, A.

The pitfalls of measuring employment tenure with the RLMS – HSE data [Electronic resource] : Working paper WP3/2016/04 / A. Oshchepkov ; National Research University Higher School of Economics. – Electronic text data (1,5 Mb). – Moscow : Higher School of Economics Publ. House, 2016. – (Series WP3 "Labour Markets in Transition"). – 36 p. (In Russian.)

In this study, we analyze some pitfalls of measuring employment tenure in the RLMS – HSE data, focusing on the inconsistency of responses to the "when you started your current job" question in two consecutive rounds of the survey. If job start dates differ between two interviews, this signals either that the job has been changed or that at least one of those two dates is incorrect. We determine conditions when the inconsistency between dates means measurement errors in tenure and analyze the incidence and magnitude of these errors as well as their correlation with individual characteristics and impact on regression analysis. Our findings may be summarized as follows.

Firstly, the prevalence of such errors in the RLMS – HSE data is high: in the period 1995–2014 about a quarter of all tenure observations were contaminated. Secondly, the magnitude of these errors is substantial. The mean absolute error (a difference between job start dates in two consecutive surveys) was about 3 years, while about a quarter of all errors exceeded 5 years. The mean relative error exceeded 50% of individual's tenure length. Thirdly, both the probability and magnitude of measurement errors in tenure are not random and correlated with some individuals' characteristics including wages. This suggests that the standard mincer-type wage equation underestimates returns to tenure in the RLMS – HSE data. Finally, we provide examples of how measurement errors in tenure may affect results of descriptive and econometric analysis. We show that in 1995-2014 due to these errors the average tenure varied in fairy wide limits, which challenges any conclusions about the rise or fall in the average tenure in Russia built on RLMS – HSE data. We also show that panel data estimates of returns to tenure are very sensitive to measurement errors as these errors affect variation in tenure length within individuals more than that variation between individuals. All in all, our results suggest that researchers using RLMS-HSE data should work with tenure data with great care and try to take into account measurement errors in their analyses.

Key words: tenure, measurement errors, RLMS – HSE, Russia

JEL codes: J62, J24, J31

Препринт WP3/2016/04
Серия WP3
Проблемы рынка труда

Ощепков Алексей Юрьевич

Проблемы измерения специального стажа на данных РМЭЗ – ВШЭ

Изд. № 1967

Национальный исследовательский университет
«Высшая школа экономики»