

На правах рукописи

Фурманов Кирилл Константинович

**Моделирование длительности безработицы
по панельным данным опросов населения**

Специальность: 08.00.13
«Математические и инструментальные методы экономики»

АВТОРЕФЕРАТ
диссертации на соискание учёной степени
кандидата экономических наук

Москва – 2009

Диссертация выполнена в государственном образовательном бюджетном учреждении высшего профессионального образования «Государственный университет – Высшая школа экономики».

Научный руководитель: кандидат физико-математических наук
Ратникова Татьяна Анатольевна

Официальные оппоненты: доктор экономических наук
Коровкин Андрей Германович

кандидат физико-математических наук
Катышев Павел Константинович

Ведущая организация: Вычислительный центр им. А.А. Дородницына
РАН.

Защита состоится 3 декабря 2009 г. в 14 часов на заседании диссертационного совета Д 212.048.02 в Государственном университете – Высшей школе экономики по адресу 101000, Москва, ул. Мясницкая, д.20, ауд.311.

С диссертацией можно ознакомиться в библиотеке Государственного университета – Высшей школы экономики.

Автореферат разослан «3» ноября 2009 г.

Учёный секретарь
диссертационного совета,
д.э.н.

Смирнов С.Н.

I. Общая характеристика диссертации

Актуальность темы исследования. Методы анализа длительности различных состояний (безработицы, бедности и т.д.), вошедшие в эконометрику в конце 1970-х годов, быстро нашли широкое применение прежде всего в экономике труда, а также в демографии, экономике предприятий и экономике здравоохранения. Одной из основных тем исследований, посвящённых эконометрическому анализу рынка труда, является продолжительность безработицы.

Известно, что продолжительная безработица имеет ряд негативных последствий для экономико-социологического состояния трудового населения. Пребывая в состоянии безработицы долгое время, индивидум теряет трудовые навыки, постепенно разочаровывается в поиске работы и прекращает её поиск, что приводит к уменьшению численности рабочей силы. Существует связь средней продолжительности безработицы с уровнем преступности – если ожидаемые выгоды от поиска работы невелики, индивидум может предпочесть улучшить своё благосостояние незаконным образом. Длительная безработица приводит индивида в подавленное, стрессовое состояние, которое может привести к употреблению наркотиков или алкоголизму, к деформации личности безработного. У длительно безработных разрушаются семьи, они не могут позволить себе иметь детей, если, конечно, сохраняют способность нести ответственность за судьбу своих близких. Долговременная безработица способствует углублению экономического неравенства, погружая людей в бедность и лишения, из которых трудно выбраться.

Оценка влияния различных факторов, определяющих длительность безработицы и вероятность нахождения работы, может помочь в разработке политики помощи безработным.

Кроме различных индивидуальных и региональных характеристик, на вероятность трудоустройства влияет продолжительность пребывания в безработном состоянии. Зависимость вероятности выхода из некоторого состояния от продолжительности пребывания в нём называется временной

зависимостью (duration или temporal dependence), и выявление характера временной зависимости также входит в цели исследования длительности безработицы.

Российский Мониторинг Экономики и Здоровья (РМЭЗ) является одним из двух основных источников данных для эконометрического анализа длительности безработицы в России, наряду с данными государственных служб занятости (ГСЗ). При этом данные РМЭЗ содержат информацию о незарегистрированных безработных, которую не предоставляют ГСЗ, а дополнительная привлекательность РМЭЗ заключается в доступности широкому кругу исследователей. Однако использование данных мониторинга сопряжено со значительными трудностями, так как они не получены путём постоянного наблюдения за индивидами, а представляют последовательность опросов населения. Опрос респондентов проводится не чаще раза в год, что затрудняет расчёт продолжительности состояний безработицы и приводит к смещению отбора. Смещение возникает, так как моменты опроса с большей вероятностью попадают на продолжительные состояния безработицы, в то время как короткие состояния чаще остаются незамеченными, «проскакивая» между опросами. Это явление до сих пор игнорировалось исследователями, занимавшимися анализом длительности безработицы по данным РМЭЗ, однако оно приводит к серьёзному завышению средней длительности безработицы и к сильно искажённым оценкам временной зависимости. Таким образом, для получения надёжных результатов требуется метод анализа, ориентированный на применение к данным последовательных опросов населения, какими и являются данные РМЭЗ.

Целью исследования является разработка инструментария, позволяющего проводить анализ длительности безработицы по панельным данным опросов населения. Для достижения этой цели решались следующие **задачи исследования:**

- анализ последствий игнорирования особой структуры панельных данных опросов населения и обращения с ними как с данными постоянного наблюдения за индивидами,

- построение модели длительности безработицы, описывающей наблюдаемые в последовательности опросов изменения состояний индивидов (например, «занятость»-«безработица»-«вне рабочей силы»), и подлежащей оцениванию методами математической статистики,

- программная реализация процедуры оценивания параметров модели,

- апробация модели на данных РМЭЗ.

Объект и предмет исследования. Объектом исследования является продолжительность состояний безработицы, фиксируемых при проведении опросов населения. Предметом исследования являются статистические методы анализа длительности состояний.

Теоретической и методологической базой исследования послужили научные работы в области статистического анализа длительностей и экономической теории поиска работы, проведённые такими учёными как С.Дж. Никель, Т. Ланкастер, Н.М. Кифер, Р.Л Прентис, Л.А. Глоклер, С.Дж. Салан, П. Шмидт, А.Д. Витте, Дж.Дж. Хекман, У. Гренандер, К. Ямагучи, Т.Маньяк, Ж.-М. Робин.

Методы исследования. В диссертационной работе использовались методы

- эконометрического моделирования дискретного выбора и длительности состояний для построения модели длительности, ориентированной на использование данных последовательности опросов населения,

- статистической теории оценивания и проверки гипотез для оценивания параметров предложенной модели и выявления факторов, оказывающих значимое влияние на продолжительность безработицы,

- статистического эксперимента (методы Монте-Карло) для выяснения последствий игнорирования особой структуры данных последовательности опросов населения.

Информационную базу исследования составили:

- данные V-XIV раундов РМЭЗ (1994-2005 гг.),
- публикуемые Федеральной Службой Статистики РФ результаты выборочного Обследования Населения по Проблемам Занятости (ОНПЗ) за 2001-2005 гг.
- база данных Национального Обследования Благополучия домохозяйств и Участия в Социальных программах (НОБУС), проводившегося в 2003 г.

Научная новизна исследования:

- Разработана модель длительности, пригодная для анализа длительности состояний по панельным данным опросов населения (таким как данные РМЭЗ). Модель учитывает различные варианты выхода из состояния безработицы, допускает наличие изменяющихся во времени объясняющих переменных и гибкую спецификацию временной зависимости, её параметры интерпретируемы. Процедура оценивания параметров модели реализована в программном модуле для статистического пакета Stata.
- Объяснена причина специфического характера временной зависимости вероятности выхода из безработицы, выявленная различными исследователями на основании данных РМЭЗ,
- Получены оценки средней продолжительности завершённых состояний безработицы на основе данных РМЭЗ, НОБУС и ОНПЗ, согласующиеся друг с другом и объяснены различия между оценками средней продолжительности безработицы, предоставляемыми Росстатом с одной стороны и полученными различными исследователями на основании данных РМЭЗ с другой стороны
- Получены оценки эконометрической модели длительности безработицы по данным РМЭЗ, не подверженные смещению отбора и разделяющие выход в

занятость и выход из состава рабочей силы и выявлены категории безработных с наибольшим риском попадания в долгосрочную безработицу.

Теоретическая и практическая значимость диссертационного исследования. Практическая значимость исследования состоит в разработке модели длительности состояний, пригодной для анализа панельных данных опросов населения и позволяющей оценивать влияние экзогенных переменных и временной зависимости на вероятности разных типов выхода из изучаемого состояния. Так как основным источником индивидуальных данных для анализа российской экономики является РМЭЗ, являющийся панельным опросом домохозяйств, разработка такого метода представляется важным для эконометрического анализа российской действительности.

Полученные оценки влияния различных факторов на длительность состояния безработицы могут быть использованы при разработке программ оказания помощи безработным.

Предложенная модель может быть использована для анализа данных в различных областях науки, помимо исследований продолжительности безработицы – например, при моделировании стабильности браков, продолжительности бедности, процесса смены рабочих мест и т.д.

Результаты исследования могут быть использованы в курсах «математическая статистика», «эконометрика панельных и качественных данных», «эконометрический анализ рынка труда».

Апробация результатов. Основные результаты диссертационного исследования были представлены на 32-м заседании международной научной школы-семинара «Системное моделирование социально-экономических процессов» имени академика С.С. Шаталина (2009 г.), семинаре лаборатории исследования рынка труда ГУ-ВШЭ (2008 г.) и трёх семинарах лаборатории макроструктурного моделирования экономики России ГУ-ВШЭ (2008-2009 гг.).

Разработанный в ходе исследования метод анализа данных и написанное для его реализации программное обеспечение прошли апробацию в ходе

выполнения исследовательского проекта «Анализ поведения российского населения по отношению к рынку труда с использованием модифицированных моделей времени жизни». Проект был поддержан грантом факультета экономики ГУ-ВШЭ 2008 года.

Результаты исследования использовались при преподавании курса «Эконометрика-2».

Публикации. Результаты диссертационного исследования опубликованы в четырёх работах общим объёмом 3.5 п.л. (вклад автора – 3.3 п.л.), в том числе в трёх статьях в ведущих научных журналах, рекомендованных ВАК МОиН РФ.

Структура диссертации. Диссертация состоит из введения, трёх глав, заключения, списка использованной литературы, который включает в себя 96 отечественных и зарубежных источников, и восьми приложений. Общий объём диссертации - 137 страниц.

II. Основное содержание диссертации

Во **введении** обосновывается актуальность диссертационного исследования. Формулируются цель, задачи, объект и предмет исследования, описывается его теоретическая, методологическая и информационная база; объясняется научная новизна и содержание полученных результатов, теоретическая и практическая значимость работы.

В **первой главе** («Эконометрические модели длительности состояний») приводится обзор основных понятий, моделей и методов анализа длительностей. В **параграфе 2.1** обосновывается необходимость в использовании особых методов анализа, отличных от классических методов эконометрики, обусловленная следующими особенностями данных о длительности состояний:

- Длительности состояний часто имеют распределения, сильно отличающиеся от нормального.

- Точное определение длительности состояний затруднено, так как связано с продолжительным постоянным наблюдениям за исследуемыми объектами. Это приводит к таким явлениям как цензурирование и урезание данных.
- Длительность состояния является результатом некоторого случайного процесса смены состояний, текущего во времени. За время наблюдения одного состояния характеристики исследуемого объекта, описываемые объясняющими переменными модели, могут изменяться.

В *параграфе 1.2* даётся определение основным понятиям анализа длительностей. Функцией дожития случайной длительности T называется функция, описывающая вероятность того, что состояние продлится не менее, чем t единиц времени:

$$S_T(t) = P(T \geq t)$$

Функция риска для длительности, задаваемой непрерывной случайной величиной, опреляется следующим образом:

$$\lambda(t) = \lim_{h \rightarrow 0} P(t \leq T < t + h | T \geq t) / h$$

Можно сказать, что функция риска выражает интенсивность, с которой состояние, длящееся уже в течение времени t , стремится к прекращению, или вероятность, с которой это состояние прекратится в ближайшее время. Для дискретных случайных величин функция риска определяется как как условная вероятность завершения состояния в момент t при условии, что состояние уже «дожило» до этого момента:

$$\lambda(t) = P(T = t | T \geq t)$$

Функция риска представляет для нас большой интерес - в частности потому, что она отражает характер *временной зависимости* (duration или temporal dependence). Если вероятность прекращения состояния увеличивается по мере роста продолжительности пребывания в этом состоянии (функция риска растёт), то говорят о положительной временной зависимости. При падении функции риска говорят об отрицательной временной зависимости, а при постоянстве риска – об отсутствии временной зависимости. Определение

характера временной зависимости наряду с оценкой влияния объясняющих переменных является одной из основных задач анализа длительностей. Рассматриваются различные спецификации функций дожития риска, описывающие разный характер временной зависимости.

Именно на примере оценивания временной зависимости наиболее ярко видны расхождения результатов исследований по данным РМЭЗ от полученных на основании данных о зарегистрированных безработных.

В *параграфе 1.3* рассматриваются распространённые подходы к моделированию наблюдаемой разнородности – различию изучаемых объектов, которое может быть учтено введением в модель объясняющих переменных. Описываются непрерывные модели пропорциональных рисков, ускоренного времени и пропорционального отношения выбытия (proportional odds) и дискретные модели, основанные на logit и cloglog-спецификациях для функции риска. Демонстрируется связь разнородности и временной зависимости.

Параграф 1.4 посвящён включению в модель ненаблюдаемой разнородности – тех различий изучаемых объектов, которые не удаётся учесть включением в модель объясняющих переменных. Как правило, суммарное воздействие ненаблюдаемых факторов моделируется с помощью случайной величины, про которую обычно предполагается, что она является независимой от объясняющих переменных.

Акцент делается на моделировании латентных классов объектов, в частности, на модели разделённой совокупности (split population model), в которой предполагается наличие класса «безнадёжных» объектов, для которых вероятность прекращения состояния (т.е. нахождения работы) равна нулю.

Рассматривается проблема идентификации ненаблюдаемой разнородности и временной зависимости.

В *параграфе 1.5* рассматриваются особенности данных о длительностях состояний – наличие цензурированных наблюдений и урезания выборки. Наблюдение называется цензурированным, если оно не даёт точной информации о длительности, а лишь отмечает, что длительность принадлежит

некоторому множеству значений. Урезание – свойство отбора наблюдений, приводящее к недостаточной доле каких-либо наблюдений в выборке (например, урезание возникает, если из выборки исключить все наблюдения со слишком малыми длительностями).

Обзор механизмов цензурирования и урезания приводится в расширенном относительно обычных работ по эконометрике виде, так как наблюдение путём проведения последовательности опросов может рассматриваться как особая комбинация этих механизмов.

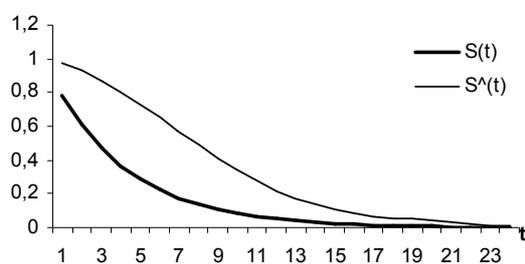
Рассматриваются непараметрические методы Каплана-Мейера и Нельсона-Аалена и особенности применения метода максимального правдоподобия для анализа цензурированных и урезанных выборок. Указывается на проблему начальных условий, возникающую при оценивании моделей с ненаблюдаемой разнородностью и временной зависимостью по урезанным слева данным. Если в выборке имеются наблюдения с «задержанным входом» (delayed entry), которые на момент начала наблюдения уже продолжались в течение некоторого времени, то индивидуальный эффект, описывающий ненаблюдаемую разнородность, уже нельзя считать независимым от экзогенных переменных модели. Дело в том, что до начала наблюдения более склонны «доживать» состояния, которым свойственно меньшее значение случайного эффекта и, соответственно, меньший уровень риска. Таким образом, индивидуальный эффект оказывается коррелирован с продолжительностью пребывания в состоянии безработицы на момент начала наблюдения.

В *параграфе 1.6* описывается подход конкурентных, или соревнующихся, рисков (competing risks) для моделирования различных типов выхода из изучаемого состояния. При таком подходе рассматриваются латентные независимые случайные величины T_1, T_2, \dots, T_S , описывающие продолжительность пребывания в изучаемом состоянии до выхода в состояния 1, 2, ..., S. Минимальное значение этих величин задаёт наблюдаемую

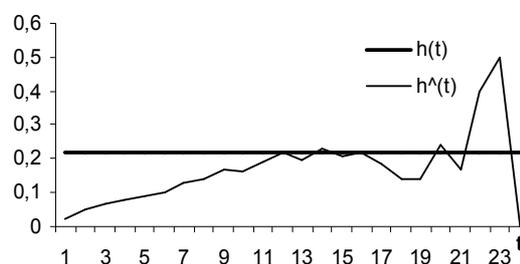
длительность: $T = \min(T_1, T_2, \dots, T_S)$. Кроме наблюдаемой длительности исследователю также известно состояние, в которое произошёл выход.

При отсутствии цензурирования или наличии только цензурирования справа задача оценивания модели длительностей T_i сводится к оцениванию модели с одним состоянием выхода. Оценивание модели с различными типами выхода по интервально цензурированным данным затруднительно. Возможные подходы к решению этой задачи разбираются в следующей главе диссертации.

Во **второй главе** («Оценивание моделей длительности состояний по данным опросов населения») рассматривается задача анализа длительности состояний в случае, когда имеющиеся данные получены в результате нескольких дискретных моментов наблюдения (раундов опроса). В **параграфе 2.1** описываются проблемы, связанные с моделированием длительностей по таким данным. Указывается на смещение отбора, возникающее вследствие того, что опросы с большей вероятностью фиксируют продолжительные состояния, в то время как короткие состояния «проскакивают» между опросами, оставаясь незамеченными. Описывается статистический эксперимент, в котором моделировался отбор состояний, происходящий при последовательности опросов. Результаты эксперимента демонстрируют последствия смещения отбора. На приведённом ниже рисунке показаны «истинные» функции риска и дожития, а также их оценки полученные методом Каплана-Мейера по сгенерированным в ходе эксперимента данным:



Функция дожития $S(t)$ и её оценка $S^{\wedge}(t)$



Функция риска $h(t)$ и её оценка $h^{\wedge}(t)$

Рис. 1. Результаты статистического эксперимента

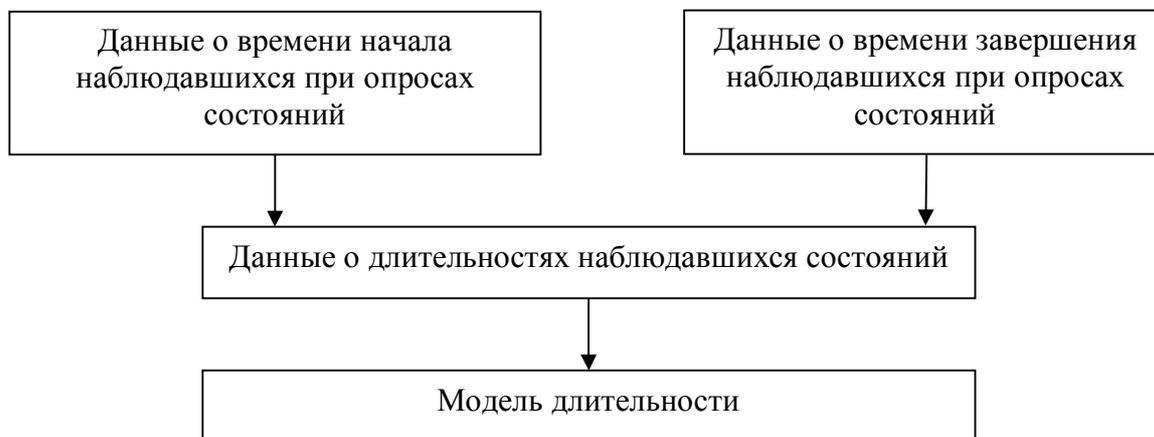
Как показывает график, процедура отбора приводит к искажению временной зависимости и сильному завышению функции дожития.

Параграф 2.2 посвящён анализу длительностей в случае, когда имеются данные всего одного опроса и, соответственно, невозможно отследить смену состояний изучаемых объектов. В диссертационном исследовании эти методы необходимы для сравнения результатов, полученных по данным РМЭЗ с оценками, полученными на основании других данных. Изложенные подходы к решению этой проблемы известны в эконометрике, хотя не нашли широкого применения. Причина малой распространённости заключается, видимо, в том, что в основной части исследований изучается экономика стран, относительно которых имеются более подробные данные, позволяющие применять более точные методы анализа.

Описывается метод синтетической когорты (см. например, Kiefer, Lundberg, Neumann, 1984), который используется в следующей главе для сравнения результатов, полученных по данным РМЭЗ, с оценками, которые можно получить из других источников данных. Метод предполагает неизменность функции дожития для состояний, начавшихся в разные моменты времени, предшествующие опросу. Подробно рассматривается частный случай, в котором приток объектов в изучаемое состояние постоянен. Подобная предпосылка делается нами в дальнейшем при анализе эмпирических данных, так как мы не располагаем информацией о притоке.

В **параграфе 2.3** рассматриваются данные последовательности опросов населения. Показывается, что особенностью этих данных является наличие урезания слева, а также интервального цензурирования и цензурирования справа. Цензурирование справа широко распространено в данных о длительности состояний и не представляет особой проблемы для исследователей. Интервальное цензурирование затрудняет оценивание моделей с различными типами выхода. Урезание приводит к смещению отбора, последствия которого рассматриваются в диссертации далее. Учёт урезания и цензурирования приводит к подходу, существенно отличающемуся от использованного в предыдущих исследованиях по данным РМЭЗ – схема такого подхода изображена на рис.2. В его основе лежит тот факт, что хотя

длительности состояний, попадающих на моменты опросов населения, подвержены смещению отбора, но вероятность завершения некоторого состояния в промежуток времени между опросами может быть состоятельно оценена. Зная вероятность прекращения можно восстановить закон распределения длительностей.



а) подход, использованный в работах (Grogan, van den Berg, 1999), Карцева (2002)



б) предлагаемый подход.

Рис. 2. Сравнение подходов к анализу длительностей по данным последовательности опросов населения

Рассматривается случай наличия двух раундов опроса. На момент первого раунда i -й объект находится в изучаемом состоянии в течение времени t_i . Между двумя наблюдениями за i -м объектом проходит время l_i . За это время состояние может завершиться одним из S возможных типов выхода. Во втором раунде выясняется, прекратилось ли состояние, наблюдавшееся в первом раунде, и, если прекратилось, то какой тип выхода произошёл. Наступивший выход или его отсутствие обозначается величиной Y_i , равной нулю, если изучаемое состояние всё ещё продолжается, иначе – номеру наступившего типа

выхода $(1, \dots, S)$. В дальнейшем, в соответствии с кодировкой величины Y_i , мы будем называть изучаемое состояние «состоянием 0», а тип выхода $s = 1, \dots, S$ – выходом в состояние s .

Сначала решается задача анализа длительностей в случае одного типа выхода. Особое внимание уделяется оцениванию модели экспоненциальной регрессии, в которой предполагается, что наблюдаемые длительности T_i имеют показательное распределение с параметром $\lambda_i = \exp(x_i' \beta)$, где x_i – набор объясняющих переменных в i -м наблюдении, а β – вектор коэффициентов при этих переменных. Показывается, что задача оценивания этой модели сводится к задаче оценивания cloglog-регрессии – известной и реализованной в современных статистических пакетах процедуре:

$$P(Y_i = 1) = 1 - \exp(-\exp(x_i' \beta + \ln l_i)).$$

При оценивании cloglog-регрессии в качестве объясняющих величин фигурируют не только переменные x_i , но и логарифм времени, прошедшего между двумя раундами опроса $\ln l_i$, причём коэффициент при этой дополнительной переменной равен единице, так что при оценивании на него должно быть наложено ограничение.

После этого допускается наличие двух типов выхода, для их моделирования используется модель конкурентных рисков (competing risks). То есть, длительность состояния 0 у i -го объекта задаётся с помощью двух латентных величин $T_{1,i}$ и $T_{2,i}$ – длительностей до выхода в состояние 1 и в состояние 2, соответственно. Величина T_i – наблюдаемая длительность состояния 0 – задаётся следующим образом: $T_i = \min(T_{1,i}, T_{2,i})$. Показывается, что в случае показательного распределения величин $T_{1,i}$ и $T_{2,i}$ вероятности выходов имеют следующее выражение:

$$P(Y_i = 0) = \exp(-(\lambda_{1,i} + \lambda_{2,i}) \cdot l_i),$$

$$P(Y_i = 1) = \frac{\lambda_{1,i} + \lambda_{1,i} \exp(-(\lambda_{1,i} + \lambda_{2,i})l_i) - (\lambda_{1,i} + \lambda_{2,i}) \exp(-\lambda_{2,i}l_i)}{(\lambda_{1,i} + \lambda_{2,i})} + (1 - \exp(-\lambda_{1,i}l_i)) \exp(-\lambda_{2,i}l_i),$$

$$P(Y_i = 2) = \frac{\lambda_{2,i} + \lambda_{2,i} \exp(-(\lambda_{1,i} + \lambda_{2,i})l_i) - (\lambda_{1,i} + \lambda_{2,i}) \exp(-\lambda_{1,i}l_i)}{(\lambda_{1,i} + \lambda_{2,i})} + (1 - \exp(-\lambda_{2,i}l_i)) \exp(-\lambda_{1,i}l_i),$$

где $\lambda_{1,i}$ и $\lambda_{2,i}$ – параметры распределения величин $T_{1,i}$ и $T_{2,i}$.

Показывается, что анализ данных более чем двух опросов сводится к случаю двух опросов искусственным увеличением числа наблюдений.

Хотя функция правдоподобия может быть выписана в случае зависимости параметров $\lambda_{1,i}$ и $\lambda_{2,i}$ от объясняющих переменных, при нахождении оценок параметров возникает проблема множественных максимумов, так что при анализе данных РМЭЗ такой подход использовался только для определения средней продолжительности безработицы и вероятностей выхода в занятость и из рабочей силы.

В *параграфе 2.4* даётся альтернативное решение задачи, поставленной в предыдущем параграфе. Рассматривается дискретная модель, в которой весь ненаблюдаемый промежуток между раундами опроса разбивается на конечное число отрезков одинаковой продолжительности (в условную единицу времени). Различные варианты смены состояний изучаемого объекта, приводящие к разным значениям объясняемой переменной Y_i (тип, или состояние, выхода) изображены на рис. 3 (проиллюстрирован случай двух типов выхода):

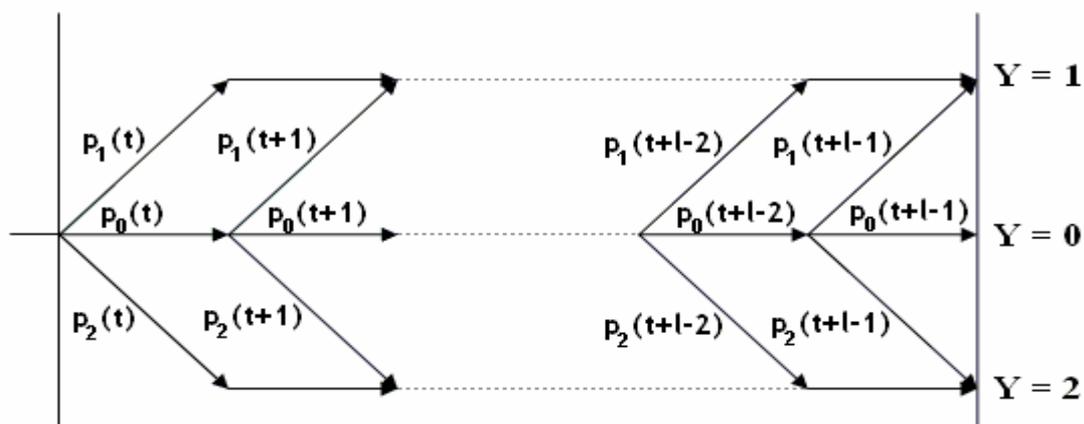


Рис. 3. Возможные выходы из изучаемого состояния в течение ненаблюдаемого промежутка времени между двумя опросами.

На рис. 3 за $p_s(j)$ обозначается вероятность перехода в состояние s в течение одной единицы времени для состояния 0, продлившегося в течение j временных единиц. Предлагается моделировать эти вероятности с помощью логистической модели множественного выбора:

$$p_0(t_i | \alpha, \beta, x_i) = \frac{1}{1 + \left(\sum_{k=1}^S \exp(x_i' \beta_k + \varphi_k(t_i | \alpha_k)) \right)}$$

$$p_s(t_i | \alpha, \beta, x_i) = \frac{\exp(x_i' \beta_s + \varphi_s(t_i | \alpha_s))}{1 + \left(\sum_{k=1}^S \exp(x_i' \beta_k + \varphi_k(t_i | \alpha_k)) \right)}, s = 1, \dots, S$$

где β_s - вектор коэффициентов при объясняющих переменных x_i , которые описывают направление и степень влияния этих переменных на вероятность выхода в состояние s , β - вектор всех коэффициентов модели, α_s - вектор параметров временной зависимости для вероятности выхода в состояние s , α - вектор всех параметров временной зависимости, $\varphi_s(t | \alpha_s)$ – некоторые функции, задающие форму временной зависимости для вероятности выхода в состояние s .

Такой подход гарантирует нахождение вероятностей в допустимых пределах и даёт возможность интерпретации параметров β . Значения элементов вектора $\exp(\beta_s)$ отражают коэффициенты пропорциональности прироста относительного риска выхода в состояние s при увеличении соответствующих объясняющих переменных на единицу. Под относительным риском понимается величина $p_s(j) / p_0(j)$.

Вероятности возможных выходов или продолжения состояния 0 в период между двумя раундами опроса имеют следующее выражение (для краткости опущены параметры модели):

$$P(Y_i = 0 | t_i, l_i) = \prod_{j=t_i}^{t_i+l_i-1} p_0(j)$$

$$P(Y_i = s | t_i, l_i) = \sum_{j=t_i}^{t_i+l_i-1} \left(p_s(j) \prod_{k=t_i}^{j-1} p_0(k) \right), s = 1, \dots, S$$

Отсюда выводится выражение для функции правдоподобия, максимизация которой даёт оценки параметров модели α и β . Показывается, что введение ненаблюдаемой разнородности в модель приводит к проблеме начальных условий, кроме того случая, когда временная зависимость отсутствует. Выводится функция правдоподобия для простейшей модели с

ненаблюдаемой неоднородностью и отсутствием временной зависимости – модели разделённой совокупности (split population model).

Третья глава диссертации («Моделирование длительности безработицы в России») посвящена оцениванию средней продолжительности безработицы по данным РМЭЗ, НОБУС и ОНПЗ и построению эконометрической модели длительности безработицы по данным РМЭЗ.

В *параграфе 3.1* даётся обзор исследований, посвящённых проблеме длительности безработицы. Основной акцент делается на использованных различными исследователями методов анализа и полученных результатов в отношении временной зависимости.

В *параграфе 3.2* приводится теоретическое основание эконометрических моделей длительности безработицы – модель поиска работы. Также рассматриваются различные причины временной зависимости, в особенности предпосылка о падении резервной заработной платы как причины роста вероятности нахождения работы.

В *параграфе 3.3* даётся описание данных РМЭЗ, на которых основывается дальнейший анализ. Подробно описывается составление анализируемой подвыборки безработных. Рассматривается проблема определения понятия «безработица». В настоящем исследовании при определении безработицы опирается на методологию Международной Организации Труда (МОТ), согласно которой безработным считается индивид

- не имеющий работы,
- желающий найти её и активно ищущий,
- готовый приступить к работе в ближайшее время.

Впрочем, при анализе использовались дополнительные ограничения. В анализируемую подвыборку попали только индивиды в возрасте от 18 до 60 лет включительно независимо от пола. Хотя российская служба статистики включает в состав безработных индивидов от 15 до 72 лет, крайние возрастные группы были намеренно исключены из-за опасения, что для них закон распределения длительности безработицы может сильно отличаться. По тем же

причинам в выборку не включались индивиды, трудоустройству которых мешает плохое состояние здоровья.

К сожалению, данные РМЭЗ до 1998 года не позволяли определить, готовы ли респондент приступить к работе незамедлительно, так что при анализе ранних данных этот критерий не учитывался. Как отмечают К.В. Маркова и С.Ю. Рошин (Маркова, Рошин, 2004, стр. 86), это не должно приводить к значительному искажению результатов, т. к. более 90% индивидов, отнесённых к безработным по остальным критериям, удовлетворяют и критерию готовности приступить к работе (по данным 1998-2001 гг.).

Год	Средняя длительность безработицы, мес.	Медианная длительность безработицы, мес.
1994-1995*	6.3	4.4
1995-1996*	8.6	5.9
1996-1998*	13.6	9.4
1998-2000*	12.3	8.6
2000-2001	7.6	5.3
2001-2002	7.2	5.0
2002-2003	6.2	4.3
2003-2004	6.5	4.5
2004-2005	6.9	4.8

* ПРИМЕЧАНИЕ. При анализе первых четырёх панелей для определения безработицы не использовался критерий готовности приступить к работе.

Табл. 1. Оценки средней и медианной продолжительности безработицы по данным РМЭЗ.

В **параграфе 3.4** приводятся результаты оценивания функции дожития для длительности безработицы по данным РМЭЗ 1994-2005 гг., полученные с помощью непрерывной модели конкурентных рисков (см. параграф 2.3) и дискретной модели длительности (см. параграф 2.4). Оценки медианной длительности безработицы оказываются значительно ниже приводимых в предыдущих исследованиях. Так, дескриптивные статистики и графики оценок функций дожития, приводимые в работах Фолей, Гроган и ван ден Берга и Карцевой, дают следующие оценки медианной продолжительности безработицы:

Гроган и ван ден Берг (данные 1994-1996 гг.) – 7 мес.,

Фолей (1992-1994 гг.) – 14 мес.,

Карцева (1994-2000 гг.) – 17,9 мес.

Как видно, оценки расходятся весьма сильно. Оценивание непрерывной модели с отсутствием временной зависимости даёт результат, отражённый в табл. 1. Практически к тем же числам приводит применение дискретной модели. При учёте временной зависимости оценки средней и медианной продолжительности оказываются ещё ниже.

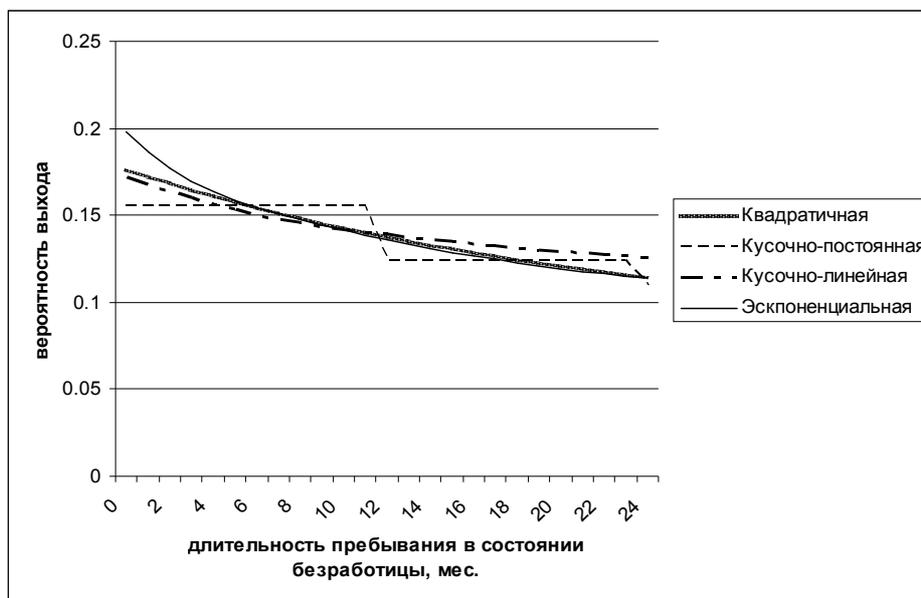


Рис. 4. Оценки функции риска выхода из безработицы, 1994-1996 гг. при разных спецификациях временной зависимости

Другое расхождение с результатами предыдущих исследований касается формы функции риска. Оценивание функции риска без учёта индивидуальных различий безработных для различных спецификаций временной зависимости по данным 1994-1996 гг. даёт результат, изображённый на рис. 4, который явно свидетельствует о падении вероятности выхода из безработицы по мере пребывания в этом состоянии. В то же время из графиков функций риска, приводимых в работах Гроган и ван ден Берга и Фолея (рис. 5), следует, что в начале периода безработицы риск выхода почти равен нулю. Это расходится с нашими оценками функции риска, но похоже на результаты случайного эксперимента (см. рис. 1), что даёт основание считать найденную в ранних работах временную зависимость следствием смещения отбора. Отметим, что некоторые из полученных нами оценок свидетельствуют о существовании интервала роста функции риска для данных 90-х годов (что совпадает с

выводами Фолея и Гроган и ван ден Берга), однако все наши оценки риска выхода значительно превышают те, что приводятся в указанных работах.

Что касается данных о длительности незавершённых состояний безработицы, публикуемых Росстатом, то на протяжении 90-х годов средняя длительность постепенно росла с 4.4 мес. (1992 г.) до 9.7 мес. (1999 г.), а в период 2000-2006 гг. колебалась в пределах от 8.2 до 9.1 мес. Однако связь длительностей завершённых и незавершённых состояний неочевидна, для их сравнения требуются дополнительный анализ, который требует либо привлечения информации о месячном притоке в состояние безработицы, либо наличия периода стабильности рынка труда, в котором приток в безработицу был бы постоянен.

В *параграфе 3.5* для проверки полученных результатов привлекаются данные обследований НОБУС и ОНПЗ. Эти данные содержат информацию о незавершённых состояниях безработицы, и для их анализа используется метод синтетической когорты, описанный в параграфе 2.2.

Функция риска, (Foley, 1997).

Функция риска, (Grogan, van den Berg, 1999)

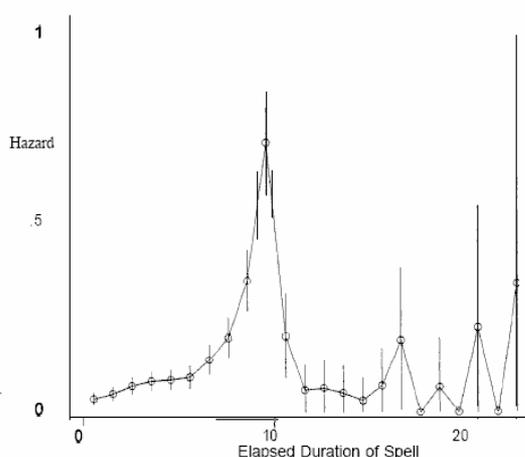
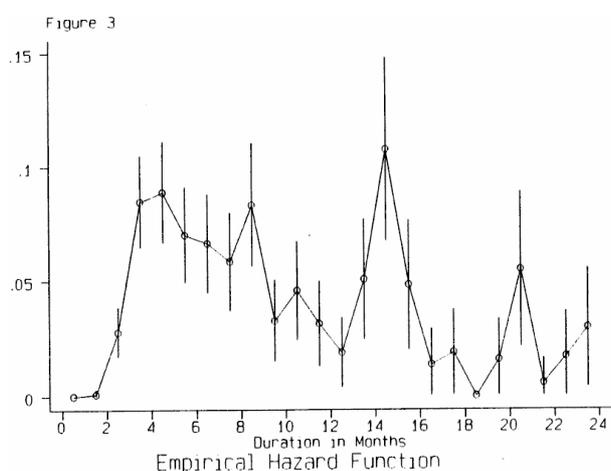


Рис. 5. Оценки функции риска в работах (Foley, 1997) и (Grogan, van den Berg, 1999).

Так как Росстат включает в категорию безработных людей в возрасте от 15 до 72 лет, то соответствующие возрастные группы были включены и в анализируемую нами выборку из данных РМЭЗ, чтобы обеспечить сопоставимость результатов. На рис.6 приведены графики оценок средней продолжительности безработицы за период с 2000 по 2005 год, полученных на

основании данных РМЭЗ и ОНПЗ в предпосылке об отсутствии временной зависимости. Видно, что хотя расхождение существует, в целом оба источника данных свидетельствуют о средней длительности порядка 6-7 месяцев.

Обследование НОБУС проводилось только в 2003 году, так что на его основании можно оценить среднюю продолжительность только для периода 2002-2003 гг. Хотя полученные оценки оказались неустойчивы к выбору распределения, описывающего длительность безработицы, наиболее сильное свидетельство имеется в пользу оценки в 5.9 месяца.

В *параграфе 3.6* описываются результаты оценивания эконометрической модели длительности безработицы по данным РМЭЗ за период 2000-2005 гг. В качестве объясняющих переменных использовались следующие индивидуальные и региональные характеристики:

Пол и возраст. С возрастом вероятность нахождения работы падает, а вероятность выхода из рабочей силы растёт. Дамми-переменная пола не оказывает значимого влияния на вероятности трудоустройства и выхода из рабочей силы (хотя значимым является её произведение с переменной опыта работы).

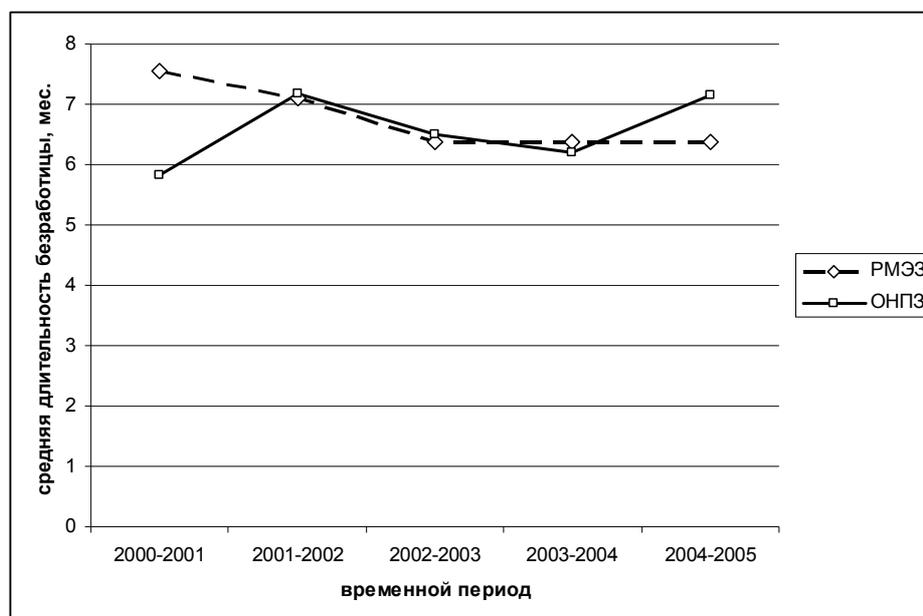


Рис. 6. Оценки средней продолжительности завершённых состояний безработицы по данным РМЭЗ и ОНПЗ

Образование. Значимого влияния уровня образования на длительность безработицы не обнаружено.

Положение на рынке труда. Отсутствие опыта работы сильно уменьшает вероятность трудоустройства, причём это влияние сильнее для мужчин, чем для женщин. Регистрация в государственной службе занятости не увеличивает шансы найти работу, но зарегистрированные безработные меньше склонны покидать состав рабочей силы.

Семейное положение. Замужние женщины более склонны выходить из рабочей силы. Впрочем, этот (вроде бы очевидный) вывод выявляется только при высоком уровне значимости (10%).

Субъективная оценка богатства. Индивиды, относящие себя к категории бедных, находят работу с меньшей вероятностью, чем остальные.

Географические характеристики. Наименьшую вероятность выхода в занятость имеют жители посёлков городского типа (ПГТ). Сельские жители проигрывают в этом отношении жителям областных центров, но заметно выигрывают у жителей ПГТ. Однако, в селе также выше вероятность выхода из рабочей силы. В регионах с более высоким уровнем безработицы меньше вероятность нахождения работы и больше вероятность выхода из состава рабочей силы.

Временная зависимость. Обнаружена сильная отрицательная временная зависимость для вероятности выхода в занятость и немонотонная (вначале отрицательная, затем положительная) зависимость для выхода из состава рабочей силы. Индивиды, не работавшие в течение последних трёх лет, имеют такую же вероятность найти работу или выйти из рабочей силы, как и безработные, вообще не имеющие опыта работы.

Временной эффект. В отличие от временной зависимости под временных эффектом понимается изменение вероятности выхода из безработицы в отдельные периоды времени (например, годы) вне связи с продолжительностью пребывания в безработице. Было обнаружено, что в 2003 и 2005 году вероятность выхода из рабочей силы была выше, чем в остальных. Для вероятности трудоустройства значимого временного эффекта выявлено не было.

Факторами, в наибольшей степени снижающими вероятность нахождения работы, оказались:

- отсутствие опыта работы,
- проживание в посёлке городского типа,
- продолжительность пребывания в состоянии безработицы.



Рис. 7. Вероятность нахождения работы в зависимости от пола, наличия опыта работы и времени пребывания в состоянии безработицы

На рис. 7 приведены модельные значения вероятности нахождения работы в зависимости от пола и наличия опыта работы. Остальные характеристики безработных предполагались следующими:

- индивиды живут в областном центре, но не в Москве или Санкт-Петербурге,
- уровень безработицы в регионе проживания равен 5%,
- индивиды имеют среднее, но не имеют высшего образования,
- принадлежат к средней возрастной категории (25-49 лет),
- не состоят в браке,
- относят себя к средней категории при оценке собственного богатства (ни бедные, ни богатые),
- временной эффект отсутствует (т.е. наблюдается закономерность, выявленная для 2001, 2002 и 2004 годов).

Линии, соответствующие индивидам без опыта работы, горизонтальны, так как для таких индивидов мы не можем установить начало периода безработицы и учесть временную зависимость.

В **заключении** приводятся выводы и рекомендации, сделанные на основании диссертационного исследования.

III. Выводы и рекомендации

Проведённое исследование позволяет сделать следующие выводы:

1. Проведённый статистический эксперимент показал, что игнорирование особого характера данных о длительности безработицы, полученных в результате последовательности опросов населения, приводит к значительному искажению функции риска и появлению кажущейся положительной временной зависимости. Несоответствие получаемых оценок функции риска истинным значениям проявляется в области малых длительностей (меньше временного промежутка между опросами).

2. Оценивание средней продолжительности безработицы по данным РМЭЗ 1994-2005 гг. показало, что при учёте смещения отбора получаемые оценки согласуются с данными ОНПЗ и НОБУС и дополняют их в отношении результативности поиска работы (вероятности трудоустройства в противоположность вероятности выхода из состава рабочей силы).

3. Результаты оценивания модели длительности безработицы по данным РМЭЗ 2000-2005 гг. показали, что тремя основными факторами, уменьшающими вероятность нахождения работы являются отсутствие опыта работы, проживание в посёлке городского типа и временная зависимость (падение вероятности трудоустройства с течением времени). Полученные оценки являются устойчивыми к изменениям спецификации и состава выборки, что свидетельствует об их надёжности.

4. Выявлена отрицательная временная зависимость для вероятности нахождения работы в России – по мере роста продолжительности пребывания в безработице вероятность трудоустройства падает. Этот вывод противоречит результатам предыдущих исследований по данным РМЭЗ, на которых, видимо,

сказалось смещение отбора. Оценки для вероятности выхода из состава рабочей силы свидетельствуют о немонотонной временной зависимости – эта вероятность сначала падает, а затем растёт.

На основании диссертационного исследования можно сделать рекомендации, относящиеся к моделированию длительности состояний и сбору данных о длительностях:

1. Разработанная модель рекомендуется к использованию для анализа данных о длительностях, полученных при значительных промежутках между моментами наблюдения за исследуемыми объектами и в тех случаях, когда промежутки между моментами наблюдения за объектами (опросы) имеют неравную длину.

2. Разработанная модель также может быть рекомендована к использованию при моделировании состояний с различными типами выхода по интервально-цензурированным данным (в которых длительности сгруппированы по интервалам).

3. В качестве альтернативы использованию разработанной модели можно рекомендовать осуществление непрерывного наблюдения за изучаемыми объектами либо сбор данных о переходах между различными состояниями в течение ненаблюдаемых промежутков на основании опросов респондентов.

Работы, опубликованные автором в ведущих рецензируемых научных журналах, рекомендованных ВАК Министерства образования и науки РФ

1. Фурманов К.К. Об измерении средней продолжительности безработицы по данным Российского мониторинга экономики и здоровья//Прикладная эконометрика, 2009, №2(14), С. 74-99 (1.2 п.л.)
2. Фурманов К.К. Сколь долг поиск работы в России: новый подход к оцениванию//Экономический журнал ВШЭ, Том 13, №2, 2009, С.279-294 (0.9 п.л.), (в соавторстве с Ратниковой Т.А., вклад автора – 0.7 п.л.)
3. Фурманов К.К. Моделирование длительности безработицы по данным Российского мониторинга экономики и здоровья//Экономический журнал ВШЭ, Том 13, №3, 2009, С. 403-427 (1.3 п.л.)

Лицензия ЛР № 020832 от 15 октября 1993 г.

Подписано в печать « » ноября 2009 г. Формат 60x84/16

Бумага офсетная. Печать офсетная.

Усл. печ. л. 1.

Тираж 100 экз. Заказ № 276 Типография издательства ГУ - ВШЭ, 125319,
г. Москва, Кочновский пр-д., д. 3