

УДК 303.7.023(075)
ББК 65в6
Р25

Рецензент:
доктор физико-математических наук, профессор,
заведующий кафедрой эконометрики и математических методов
экономики МШЭ МГУ им. М.В. Ломоносова *С.А. Айвазян*

Р25 **Ратникова, Т. А., Фурманов К. К.** Анализ панельных данных и данных о длительности состояний [Текст] : учеб. пособие / Т. А. Ратникова, К. К. Фурманов ; Нац. исслед. ун-т «Высшая школа экономики». — М.: Изд. дом Высшей школы экономики, 2014. — 376 с. — 1000 экз. — ISBN 978-5-7598-0745-2 (в обл.).

Учебное пособие охватывает темы эконометрики продвинутого уровня. В нем изложены теория и практика применения актуальных методов вероятностно-статистического анализа экономических и социологических данных, используемых для оценивания зависимостей по пролонгированным выборкам объектов, в роли которых могут выступать индивиды, семьи, фирмы, регионы, страны и т.п. Наличие последовательного ряда наблюдений позволяет учитывать индивидуальные особенности различных единиц наблюдения и их эволюции, а также изучать продолжительность пребывания объектов в том или ином состоянии (например, длительность периодов бедности для домохозяйств или периодов безработицы для индивидов).

Излагаются базовые теоретические концепции анализа панельных данных и данных о длительности состояний, а также принципы построения наиболее востребованных моделей. Примеры использования рассмотренных методов на практике строятся по реальным российским панельным данным РМЭЗ (Российского мониторинга экономического состояния и здоровья населения).

Применение изученных методов к реальным российским статистическим данным позволит глубже понять цели и задачи экономической политики государства (или фирмы), а также научиться оценивать результаты этой политики.

Пособие полезно магистрантам, аспирантам и исследователям, специализирующимся в областях математических методов анализа экономики, микро- и макроэкономического анализа, экономики фирм, анализа потребительского поведения населения, рынка труда, экономики здравоохранения, демографии.

УДК 303.7.023(075)
ББК 65в6

ISBN 978-5-7598-0745-2

© Ратникова Т.А., Фурманов К.К., 2014
© Оформление. Издательский дом
Высшей школы экономики, 2014

Содержание

От авторов.....	9
-----------------	---

Часть I. Методы анализа панельных данных

1. Введение.....	15
1.1. История создания микроэконометрики	15
1.2. Описание наиболее употребимых источников панельных данных	17
1.3. Преимущества использования панельных данных	19
1.4. Проблемы использования панельных данных.....	23
1.4.1. Гетерогенное смещение.....	23
1.4.2. Смещение самоотбора	25
2. Простейшие модели анализа панельных данных.....	28
2.1. Спецификация моделей.....	28
2.1.1. Модель сквозной регрессии	28
2.1.2. Модель регрессии с детерминированным индивидуальным эффектом (fixed effect model)	29
2.1.3. Модель регрессии со случайным индивидуальным эффектом (random effect model)	31
2.2. Оценивание моделей со случайным индивидуальным эффектом	32
2.2.1. Операторы BETWEEN (B) и WITHIN (W).....	32
2.2.2. Оценки «between» и «within».....	37
2.3. Ковариационная матрица случайного возмущения в модели со случайным индивидуальным эффектом.....	40
2.4. Интерпретация параметра θ^2	42
2.5. Оценивание параметра θ^2	43
2.6. Реализуемый GLS	44
2.7. Метод максимального правдоподобия	45
3. Сравнение оценок	48
3.1. Декомпозиция оценок	48
3.2. Асимптотические свойства оценок при $N \rightarrow \infty$ и $T \rightarrow \infty$	49

3.3. Асимптотические свойства оценок при $N \rightarrow \infty$ и конечных T	51
3.4. Свойства оценок при конечных значениях N и T	51
3.4.1. Сравнительная эффективность оценок.....	51
3.4.2. Сравнение оценок при конечных значениях N и T в зависимости от структуры дисперсии наблюдений.....	52
4. Тестирование спецификации	54
4.1. Критика Мундлаком спецификации модели со случайным индивидуальным эффектом.....	54
4.2. Тесты Хаусмана на ошибки спецификации.....	58
4.2.1. Принцип тестов Хаусмана.....	58
4.2.2. Применение теста Хаусмана к модели со случайным индивидуальным эффектом.....	59
4.3. Тесты на существование и независимость индивидуального эффекта.....	60
4.4. О применимости теста Хаусмана.....	62
5. Классификация моделей анализа панельных данных	63
5.1. Схема используемых моделей.....	63
5.2. Модель анализа ковариаций.....	65
6. Пример: оценивание уравнения заработной платы по данным РМЭЗ	69
6.1. Постановка задачи.....	69
6.2. Модель с индивидуальными эффектами.....	70
6.3. Качество подгонки и выбор наиболее адекватной модели.....	73
6.4. Модель с индивидуальными и временными эффектами.....	75
6.5. Ковариационный анализ (тестирование возможности объединения данных в панель).....	79
7. Особенности оценивания моделей с панельными данными в условиях гетероскедастичности и автокорреляции случайных возмущений	82
7.1. Оценивание ковариационных матриц ошибок в условиях гетероскедастичности и автокорреляции.....	82
7.2. Тестирование гетероскедастичности и автокорреляции.....	85

8. Оценивание коэффициентов панельных регрессий в условиях коррелированности регрессоров и случайной ошибки	90
8.1. Метод Хаусмана — Тейлора	90
8.1.1. Идея и преимущества метода	90
8.1.2. Основные допущения	92
8.1.3. Состоятельное, но неэффективное оценивание	93
8.1.4. Состоятельное и эффективное оценивание	95
8.1.5. Тестирование априорных ограничений	98
8.1.6. Пример: использование метода Хаусмана — Тейлора для оценивания эффекта от образования по данным РМЭЗ	100
8.2. Ошибки измерения в панельных данных	104
8.2.1. Основные источники ошибок измерений	104
8.2.2. Методы оценивания регрессий по панельным данным при наличии ошибок измерений	105
8.3. Оценивание динамических моделей	109
8.3.1. Авторегрессионные модели с детерминированным эффектом. Обобщенный метод моментов	109
8.3.2. Авторегрессионные динамические модели с экзогенными переменными и детерминированным эффектом. Обобщенный метод моментов	116
8.3.3. Классификация и сравнительный анализ оценок линейных динамических регрессий	117
8.3.4. Метод максимального правдоподобия для оценивания динамических регрессий со случайным индивидуальным эффектом	119
8.3.5. Проблема стационарности и коинтеграция	122
8.3.6. Тест на единичные корни для панельных данных	125
8.3.7. Тесты на панельную коинтеграцию	130
9. Модели с дискретными и ограниченными зависимыми переменными	135
9.1. Модели бинарного выбора	135
9.1.1. Оценивание моделей с детерминированным индивидуальным эффектом	136
9.1.2. Оценивание моделей со случайным индивидуальным эффектом	141
9.1.3. Пример: выявление детерминант задолженности по заработной плате в 1990-е годы по данным РМЭЗ	142
9.2. Модель тобит	146
9.3. Оценивание динамических моделей бинарного выбора	147

10. Методы борьбы с истощением выборки	152
10.1. Анализ несбалансированных панелей.....	152
10.1.1. Модель со случайным индивидуальным эффектом с несбалансированными данными	152
10.1.2. ANOVA-методы оценки ковариационных матриц	155
10.2. Панели с замещением	158
10.3. Псевдопанели	161
10.3.1. Оценивание по данным о когортах.....	162
10.3.2. Влияние выбора когорт на величину смещения	165
10.3.3. Влияние выбора когорт на дисперсию.....	167
10.3.4. Пример: оценивание кривой Энгеля	169
10.4. Смещение самоотбора в неполных панелях	172
10.4.1. Оценивание при наличии случайно пропущенных данных	173
10.4.2. Тестирование смещения самоотбора	174
10.4.3. Оценивание при наличии неслучайно пропущенных данных	176
11. Оценивание многоуровневых (или иерархических) моделей со случайными коэффициентами	178
11.1. Линейные иерархические модели	180
11.2. Оценивание иерархических моделей	183
11.3. Пример: оценивание бинарной иерархической модели присутствия ПИИ в предприятиях пищевой промышленности России.....	185
12. Практикум: анализ панельных данных в пакете STATA	191
12.1. Пакет статистической обработки данных STATA	191
12.1.1. Краткая характеристика пакета STATA.....	191
12.1.2. Организация данных в пакете STATA	192
12.2. Примерная схема анализа панельных данных для решения некоторой частной прикладной задачи	195
12.2.1. Постановка задачи	195
12.2.2. Изучение основных описательных статистик и визуальный анализ данных	199
12.2.3. Построение линейной регрессионной модели.....	208
12.2.4. Оценивание «between»-регрессии.....	214
12.2.5. Оценивание «within»-регрессии или модели с детерминированными эффектами	216

12.2.6. Оценивание модели со случайными эффектами.....	218
12.2.7. Выбор наиболее адекватной модели.....	220
12.2.8. Использование фиктивных переменных в регрессионных моделях.....	225
12.3. Оценивание полной эконометрической модели преступности с эндогенными регрессорами	229
12.3.1. Оценивание модели со случайными эффектами методом инструментальных переменных	230
12.3.2. Двухшаговая процедура оценивания регрессии с детерминированными эффектами	231
12.4. Оценивание динамической модели преступности	235
12.5. Самостоятельное упражнение: проверка возможности объединения данных в панель	240

Часть II. Моделирование длительности состояний

1. Вероятностная модель длительности.....	247
1.1. Распределение длительностей: способы описания	247
1.1.1. Функция дожития	247
1.1.2. Функция риска.....	251
1.1.3. Интегральная функция риска	256
1.1.4. Функция квантилей	259
1.2. Геометрическая интерпретация математического ожидания.....	260
1.3. Часто используемые распределения длительностей.....	261
1.4. Несобственные распределения	269
1.5. Условные распределения. Остаточное время жизни	270
1.6. Характеристики дискретных распределений.....	274
1.7. Практикум: генерирование случайных выборок.....	277
2. Основы статистического анализа данных о длительности.....	282
2.1. Неполнота данных	282
2.1.1. Цензурирование.....	284
2.1.2. Усечение	286
2.2. Оценивание распределения длительностей.....	287
2.2.1. Непараметрические методы	287
2.2.2. Параметрическое оценивание.....	293
2.3. Описательная статистика	296
2.4. Сравнение функций дожития в нескольких выборках.....	298

2.5. Пример: оценка силы смертности по данным РМЭЗ.....	300
2.6. Практикум: исследование досрочного расторжения договоров страхования жизни.....	305
3. Регрессионные модели длительности	314
3.1. Модель пропорциональных рисков	314
3.1.1. Формулировка модели. Интерпретация коэффициентов	314
3.1.2. Метод частичного правдоподобия.....	317
3.1.3. Совпадающие моменты прекращения.....	318
3.1.4. Оценка опорного распределения. Остатки Кокса — Снелла	321
3.2. Модель ускоренного времени.....	323
3.2.1. Формулировка модели	323
3.2.2. Линейная форма модели ускоренного времени.....	324
3.3. Обзор параметрических моделей	325
3.4. Прогнозирование в моделях длительности.....	331
3.5. Практикум: регрессионная модель досрочного расторжения договоров страхования жизни (1)	332
4. Ненаблюдаемая разнородность.....	345
4.1. Распределение смеси	345
4.2. Ненаблюдаемая разнородность в модели пропорциональных рисков	349
4.3. Ненаблюдаемая разнородность в модели ускоренного времени.....	352
4.4. Модели <i>never-stayer</i>	354
4.5. Проблема выявления ненаблюдаемой разнородности.....	357
4.6. Практикум: регрессионная модель досрочного расторжения договоров страхования жизни (2)	358
Заключение	363
Библиография	364

От авторов

Это учебное пособие — переработанный и значительно дополненный вариант книги Т.А. Ратниковой «Введение в эконометрический анализ панельных данных», изданной в 2010 г.

В то время лишь в нескольких учебниках по эконометрике, изданных на русском языке, существовали разделы, более или менее подробные, посвященные анализу панельных данных («Эконометрика» И.И. Елисеевой, «Эконометрика. Начальный курс» Я. Магнуса, П.К. Катышева, А.А. Пересецкого, перевод учебника М. Вербика «Путеводитель по современной эконометрике» под редакцией С.А. Айвазяна, «Эконометрика» В.П. Носко). Обстоятельные иностранные монографии [Hsiao, 1986; Baltagi, 1995; Matyas, Sevestre, 1996] и сейчас еще не переведены на русский язык и доступны студентам далеко не повсеместно. Кроме того, методический арсенал эконометриста, занимающегося анализом панелей, значительно расширился за последние десятилетия и продолжает расти; новые методы нашли свое место как в академических журналах, так и в программном обеспечении — их описание заслуживает отдельной книги, и настоящее пособие — попытка заполнить этот пробел.

Почему исследователи обращаются к панельным данным? Например, потому что такие данные позволяют учесть не измеримые, не наблюдаемые статистикой различия между обследуемыми объектами (регионами, фирмами, индивидами). В настоящее время данные такого рода встречаются нередко, так что говорить об анализе панельных данных как об узкой или маловостребованной отрасли науки не приходится. Обращаясь к пространственной выборке, аналитик изучает различия между наблюдаемыми объектами, а исследуя временные ряды, — изменение состояния отдельного объекта с течением времени. Использование панельных данных позволяет подступить к решению обеих задач и построить модель, объясняющую динамику состояний множества объектов.

Первая часть настоящего пособия выросла из курса лекций, читаемого студентам-старшекурсникам бакалавриата и магистрантам факультета экономики НИУ ВШЭ с 2001 г. В 2004 г. материалы лекций, работа над которыми была поддержана грантом НФПК (Национального фонда подготовки кадров) в рамках программы

«Совершенствование преподавания социально-экономических дисциплин в вузах» инновационного проекта развития образования, появились в электронном виде на сайте университета. В 2006 г. они были существенно дополнены и опубликованы в разделе «Лекционные и методические материалы» в «Экономическом журнале ВШЭ». Еще через четыре года они переросли в пособие [Ратникова, 2010], о котором уже говорилось.

Вторая, полностью новая, часть настоящей книги — «Модели длительности состояний» посвящена теме, которая практически не рассматривается в учебной литературе по эконометрике. В известных отечественных учебниках [Айвазян, Мхитарян, 1998; Магнус, Катышев, Пересецкий, 2004] этой теме уделено всего несколько страниц, иностранные пособия [Greene, 2012; Cameron, Trivedi, 2005] содержат немногословные параграфы. И те и другие книги могут создать впечатление, будто модели длительности — это причудливые вариации на тему классической регрессионной модели, обвешанные мишурой из непривычных терминов. В действительности речь идет о существенно обособленном подходе к анализу, а необычные термины и понятия — способы адекватно передать суть исследуемых процессов, плохо укладывающуюся в рамки привычных для эконометристов средств описания данных. Поэтому, в отличие от первой части, где материал излагается в предположении, что читатель твердо усвоил базовый курс эконометрики, вторая часть содержит изложение почти с нуля. В значительной мере она доступна для студента, изучившего курс теории вероятностей и математической статистики и знакомого с регрессионным анализом.

При составлении второй части мы столкнулись с проблемой выбора терминов — одно и то же понятие исследователи трактуют по-разному. Для важнейшего понятия анализа длительностей нами использован термин «функция риска» вместо более распространенного — «функция опасности отказов». Последний вариант, применяемый специалистами по теории надежности, несет на себе слишком явный отпечаток конкретной области применения. По той же причине был отклонен и вариант «сила смертности», распространенный среди демографов и актуариев. В математической статистике есть и другая функция риска, используемая при анализе критериев для про-

верки гипотез, но вряд ли совпадение терминов приведет к путанице — слишком уж различна суть этих функций.

Мы признательны всем, кто оказывал содействие в подготовке и совершенствовании этой книги. Учебник не смог бы состояться без его идейного вдохновителя — Э.Б. Ершова, без Г.Г. Канторовича, который потратил немало своих времени и сил, помогая совершенствовать текст на ранних стадиях разработки; без участия французских коллег — Ф. Гарда, Б. Дормонт и М. Морель, которые охотно делились своим опытом; без ценных советов и рекомендаций С.А. Айвазяна, В.А. Бессонова, П.К. Катышева, Е.В. Коссовой, А.А. Пересецкого, И.Г. Поспелова; без весьма полезного опыта работы с реальными данными в Центре трудовых исследований под руководством В.Е. Гимпельсона и Р.И. Капелюшникова; без В.С. Автономова, который принял решение о выделении средств факультета на издание пособия в 2010 г.; без кропотливого и самоотверженного труда сотрудников «Экономического журнала» и Издательского дома НИУ ВШЭ, которым пришлось немало повозиться с многочисленными формулами. Отдельную благодарность мы выражаем своим ученикам, в особенности Анне Гладышевой, Ирине Чернышевой и Татьяне Барладян, чьи исследовательские работы легли в основу приведенных в этой книге примеров.

Вся ответственность за недостатки, содержащиеся в пособии, целиком лежит на авторах. Мы будем признательны, если читатель, обнаруживший ту или иную ошибку, будет не только внимателен, но и великодушен и сообщит нам о своей находке, дав возможность ее исправить.