

SECOND EDITION

ROUTLEDGE ADVANCED TEXTS IN ECONOMICS AND FINANCE

# APPLIED HEALTH ECONOMICS

ANDREW M. JONES, NIGEL RICE,  
TERESA BAGO D'UVA AND SILVIA BALIA

ROUTLEDGE



# Applied Health Economics

The first edition of *Applied Health Economics* did an expert job of showing how the availability of large-scale datasets and the rapid advancement of advanced econometric techniques can help health economists and health professionals make sense of information better than ever before.

This second edition has been revised and updated throughout, and includes new chapters on the description and modelling of individual health care costs, thus broadening the book's readership to those working on risk adjustment and health technology appraisal. The text also fully reflects the very latest advances in the health economics field and the key journal literature.

Practical applications of the methods are illustrated using data on health from the British Health and Lifestyle Survey (HALS), the British Household Panel Survey (BHPS), the European Community Household Panel (ECHP), the US Medical Expenditure Panel Survey (MEPS) and the Survey of Health, Ageing and Retirement in Europe (SHARE). There is a strong emphasis on applied work, illustrating the use of relevant computer software with code provided for Stata. Familiarity with the basic syntax and structure of Stata is assumed. The Stata code and extracts from the statistical output are embedded directly in the main text and explained at regular intervals.

The book is built around empirical case studies rather than general theory, and the emphasis is on learning by example. It presents a detailed dissection of the methods and results of recent research papers written by the authors and their colleagues. Relevant methods are presented alongside the Stata code that can be used to implement them, and the empirical results are discussed at each stage.

This text brings together the theory and application of health economics and econometrics, and will be a valuable reference for applied economists and students of health economics and applied econometrics.

**Andrew M. Jones** is Professor of Economics at the University of York, UK, where he is also Head of the Department of Economics and Related Studies.

**Nigel Rice** is Professor of Health Economics and Director of the Health, Econometrics and Data Group (HEDG) at the Centre for Health Economics, University of York, UK.

**Teresa Bago d'Uva** is an Associate Professor in the Department of Applied Economics at the Erasmus School of Economics, The Netherlands.

**Silvia Balia** is an Assistant Professor at the Department of Economics and Business and a researcher at the Centre for North-South Economic Research (CRENoS) at the University of Cagliari, Italy.

## Routledge advanced texts in economics and finance

- 1 Financial Econometrics**  
*Peijie Wang*
- 2 Macroeconomics for Developing Countries**  
Second edition  
*Raghendra Jha*
- 3 Advanced Mathematical Economics**  
*Rakesh Vohra*
- 4 Advanced Econometric Theory**  
*John S. Chipman*
- 5 Understanding Macroeconomic Theory**  
*John M. Barron, Bradley T. Ewing and Gerald J. Lynch*
- 6 Regional Economics**  
*Roberta Capello*
- 7 Mathematical Finance**  
Core theory, problems and statistical algorithms  
*Nikolai Dokuchaev*
- 8 Applied Health Economics**  
*Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva and Silvia Balia*
- 9 Information Economics**  
*Urs Birchler and Monika Büttler*
- 10 Financial Econometrics**  
Second edition  
*Peijie Wang*
- 11 Development Finance**  
Debates, dogmas and new directions  
*Stephen Spratt*
- 12 Culture and Economics**  
On values, economics and international business  
*Eelke de Jong*
- 13 Modern Public Economics**  
Second edition  
*Raghendra Jha*
- 14 Introduction to Estimating Economic Models**  
*Atsushi Maki*
- 15 Advanced Econometric Theory**  
*John Chipman*
- 16 Behavioral Economics**  
*Edward Cartwright*
- 17 Essentials of Advanced Macroeconomic Theory**  
*Ola Olsson*
- 18 Behavioural Economics and Finance**  
*Michelle Baddeley*
- 19 Applied Health Economics**  
Second edition  
*Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva and Silvia Balia*

# **Applied Health Economics**

Second edition

**Andrew M. Jones, Nigel Rice,  
Teresa Bago d'Uva and  
Silvia Balia**

First published 2013  
by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

Simultaneously published in the USA and Canada  
by Routledge  
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2013 Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva and  
Silvia Balia

The right of Andrew M. Jones, Nigel Rice, Teresa Bago d'Uva and  
Silvia Balia to be identified as authors of this work has been asserted  
by them in accordance with the Copyright, Designs and Patent Act  
1988.

All rights reserved. No part of this book may be reprinted or  
reproduced or utilised in any form or by any electronic, mechanical,  
or other means, now known or hereafter invented, including  
photocopying and recording, or in any information storage or retrieval  
system, without permission in writing from the publishers.

*Trademark notice:* Product or corporate names may be trademarks or  
registered trademarks, and are used only for identification and  
explanation without intent to infringe.

*British Library Cataloguing in Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging in Publication Data*

Applied health economics/by Andrew M. Jones ... [et al.]. – 2nd ed.  
p. cm.

Includes bibliographical references and index.

I. Jones, Andrew M., 1960–

[DNLM: 1. Economics, Medical. 2. Health Care Costs. 3. Statistics  
as Topic. W 74.1]

338.4'73621–dc23

2012006760

ISBN: 978-0-415-67681-6 (hbk)

ISBN: 978-0-415-67682-3 (pbk)

ISBN: 978-0-203-10241-1 (ebk)

Typeset in Times New Roman  
by Florence Production Ltd, Stoodleigh, Devon

# Contents

<i>List of illustrations</i>	viii
<i>Preface</i>	xiii
<i>Acknowledgements</i>	xv
<i>Introduction</i>	xvii
<b>PART I</b>	
<b>Describing and summarising data</b>	<b>1</b>
<b>1 Data and survey design</b>	<b>3</b>
1.1 <i>The Health and Lifestyle Survey</i>	<i>3</i>
1.2 <i>The British Household Panel Survey</i>	<i>5</i>
1.3 <i>The European Community Household Panel</i>	<i>7</i>
1.4 <i>The US Medical Expenditure Panel Survey</i>	<i>8</i>
1.5 <i>Survey of Health, Ageing and Retirement in Europe</i>	<i>10</i>
1.6 <i>Overview</i>	<i>11</i>
<b>2 Describing the dynamics of health</b>	<b>12</b>
2.1 <i>Introduction</i>	<i>12</i>
2.2 <i>Graphical analysis</i>	<i>14</i>
2.3 <i>Tabulating the data</i>	<i>21</i>
2.4 <i>Overview</i>	<i>29</i>
<b>3 Describing health care costs</b>	<b>30</b>
3.1 <i>Introduction</i>	<i>30</i>
3.2 <i>Data description</i>	<i>31</i>
3.3 <i>Modelling health care cost data</i>	<i>40</i>
3.4 <i>Linear regression models</i>	<i>44</i>
3.5 <i>Overview</i>	<i>65</i>
<b>PART II</b>	
<b>Categorical data</b>	<b>67</b>
<b>4 Reporting heterogeneity in health</b>	<b>69</b>
4.1 <i>Introduction</i>	<i>69</i>
4.2 <i>Data</i>	<i>71</i>

vi *Contents*

4.3 *Standard analysis* 72

4.4 *Using vignettes* 75

4.5 *Overview* 98

*Appendix* 101

**5 Health and lifestyles** 106

5.1 *Introduction* 106

5.2 *HALS data and sample* 107

5.3 *Descriptive analysis* 113

5.4 *Estimation strategy and results* 121

5.5 *Overview* 138

**PART III**

**Duration data** 139

**6 Smoking and mortality** 141

6.1 *Introduction* 141

6.2 *Basic concepts of duration analysis* 141

6.3 *The HALS data* 143

6.4 *Duration data in HALS* 145

6.5 *Descriptive statistics* 149

6.6 *Duration models* 150

6.7 *Overview* 181

**7 Health and retirement** 182

7.1 *Introduction* 182

7.2 *Preparing and summarising the data* 182

7.3 *Dealing with self-reported health* 193

7.4 *Empirical approach to duration modelling* 196

7.5 *Stock sampling and discrete-time hazard analysis* 201

7.6 *Overview* 217

**PART IV**

**Panel data** 219

**8 Health and wages** 221

8.1 *Introduction* 221

8.2 *BHPS sample and variables* 222

8.3 *Empirical model and estimation* 225

8.4 *Overview* 241

*Appendix* 243

**9 Modelling the dynamics of health** 244

9.1 *Introduction* 244

9.2 *Static models* 251

9.3	<i>Dynamic models</i>	268
9.4	<i>Overview</i>	276
<b>10</b>	<b>Non-response and attrition bias</b>	<b>277</b>
10.1	<i>Introduction</i>	277
10.2	<i>Testing for non-response bias</i>	280
10.3	<i>Estimation</i>	283
10.4	<i>Overview</i>	290
	<i>Appendix</i>	290
<b>PART V</b>		
	<b>Health care data</b>	<b>293</b>
<b>11</b>	<b>Models for count data</b>	<b>295</b>
11.1	<i>Introduction</i>	295
11.2	<i>The Poisson model</i>	296
11.3	<i>The negative binomial model</i>	299
11.4	<i>Zero-inflated models</i>	302
11.5	<i>Hurdle models</i>	308
11.6	<i>Finite mixture/latent class models</i>	311
11.7	<i>Latent class models for panel data</i>	318
11.8	<i>Overview</i>	337
	<i>Appendix</i>	337
<b>12</b>	<b>Modelling health care costs</b>	<b>342</b>
12.1	<i>Introduction</i>	342
12.2	<i>Exponential conditional mean models</i>	343
12.3	<i>Generalised linear models</i>	359
12.4	<i>Finite mixture models</i>	370
12.5	<i>Comparing model performance</i>	377
12.6	<i>Overview</i>	380
	<i>Bibliography</i>	381
	<i>Index</i>	388

# Illustrations

## FIGURES

2.1	Bar chart for SAH, men	15
2.2	Bar chart for SAH by wave, men	16
2.3	Bar chart for SAH by age group, wave 1, men	17
2.4	Bar chart for SAH by quintile of <code>meaninc</code> , men	18
2.5	Empirical CDFs of <code>meaninc</code> , men	19
2.6	Bar chart for SAH by education, men	20
2.7	Bar chart for SAH by previous SAH, wave 2, men	21
3.1	Distribution of positive costs	35
3.2	Distribution of transformed costs	36
3.3	Box plots of health care costs	37
3.4	Normal probability plots of health care costs	37
3.5	Mean of cost against variance – level of cost (LHS): log of costs (RHS)	40
3.6	Empirical CDFs by insurance status: all expenditures and expenditures < \$15,000	43
3.7	Normal probability plot for Box–Cox residuals	64
6.1	Non-parametric functions for smoking initiation	153
6.2	Cox–Snell residuals test – smoking initiation	156
6.3	Log-logistic functions for smoking initiation	160
6.4	Cox–Snell residuals test – smoking initiation for starters	161
6.5	Log-logistic functions for smoking initiation	164
6.6	Non-parametric functions for smoking cessation	167
6.7	Cox–Snell residuals test – smoking cessation	170
6.8	Weibull estimated functions for smoking cessation	171
6.9	Normal probability plot for lifespan	174
6.10	Non-parametric functions for lifespan	176
6.11	Cox–Snell residuals test – lifespan	177
6.12	Gompertz estimated functions for lifespan	181
7.1	Life table estimates of the proportion not retired by health limitations	203
12.1	Normal probability plot for the exponential conditional mean model	345

12.2	Hosmer–Lemeshow plot: nonlinear least squares	348
12.3	Hosmer–Lemeshow plot: Poisson regression	350
12.4	Weibull density functions	351
12.5	Normal pp plot for generalised linear model	364

## TABLES

3.1	Description of dataset	32
3.2	Summary statistics of variables	41
3.3	OLS on total expenditures	45
3.4	Maximum likelihood estimation of log-normal model	51
3.5	Heteroskedasticity on log-scale	56
3.6	Square root transformed linear regression	58
3.7	Comparison of approaches for a binary regressor	60
3.8	Comparison of predictions following retransformation	61
3.9	Box–Cox transformed regression	63
3.10	Summary of residuals from Box–Cox transformed regression	64
4.1	Ordered probit for self-reported health in the domain of breathing	73
4.2	Partial effects on the probability of reporting no problems with breathing for reference individual, with homogeneous reporting	74
4.3	Ordered probit for vignette ratings in the domain of <i>breathing</i>	80
4.4	Generalised ordered probit for vignette ratings in the domain of <i>breathing</i>	82
4.5	Interval regression for self-reported health in the domain of breathing with parallel cut-point shift	88
4.6	Partial effects on the probability of having no problems with breathing for reference individual, with parallel cut-point shift	89
4.7	Interval regression for self-reported health in the domain of breathing with non-parallel cut-point shift	91
4.8	Partial effects on the probability of having no problems with breathing for reference individual, with non-parallel cut-point shift – two-step estimation	91
4.9	HOPIT for self-reported health in the domain of breathing with cut-point shift – one-step estimation	93
4.10	Partial effects on the probability of having no problems with breathing for reference individual in model with non-parallel cut-point shift – one-step HOPIT estimation	98
4.11	Estimated coefficients in latent health index in all models	99
4.12	Partial effects on the probability of reporting no problems with breathing for reference individual in all models	100
5.1	Probit model for mortality – without exclusion restrictions	124
5.2	Probit model for mortality – with exclusion restrictions	125
5.3	Multivariate probit model – without exclusion restrictions	128

5.4	Multivariate probit model – with exclusion restrictions	131
5.5	Average partial effects from alternative models for mortality	137
6.1	Information criteria – smoking initiation	156
6.2	Smoking initiation – coefficients from log-logistic model (AFT)	159
6.3	Information criteria – smoking initiation for starters	161
6.4	Smoking initiation for starters – coefficients from log-logistic model (AFT)	163
6.5	Information criteria – smoking cessation	169
6.6	Smoking cessation – coefficients from Weibull model (AFT)	172
6.7	Information criteria – lifespan	177
6.8	Lifespan – coefficients from Gompertz model	179
6.9	Lifespan – hazard ratio from Gompertz model	180
7.1	Variable names and definitions	187
7.2	Labour market status by wave	191
7.3	Descriptive statistics	192
7.4	Ordered probits for self-assessed health	195
7.5	Life-table for retirement by health limitations	202
7.6	Discrete-time hazard model – no heterogeneity	207
7.7	Complementary log-log model with frailty	210
7.8	Discrete-time duration model with gamma distributed frailty	212
7.9	Discrete-time duration model with Heckman–Singer frailty	214
7.10	Discrete-time duration model with Heckman–Singer frailty	215
7.11	Discrete-time duration models with latent self-assessed health	216
8.1	Variable labels and definitions	226
8.2	Summary statistics for full sample of observations	227
8.3	OLS on full sample of observations	229
8.4	RE on full sample of observations	231
8.5	FE on full sample of observations	233
8.6	Hausman and Taylor IV estimator on full sample of observations	237
8.7	Men – Amemiya and MaCurdy IV estimator on full sample of observations	240
8.8	Men – comparison across estimators	242
9.1	Pooled probit model, unbalanced panel	253
9.2	Pooled probit model, average partial effects	254
9.3	Pooled probit model, balanced panel	255
9.4	Mundlak specification of pooled probit model, unbalanced panel	257
9.5	Random effects probit model, unbalanced panel	259
9.6	Random effects probit model, partial effects	260
9.7	Random effects probit model, balanced panel	264
9.8	Mundlak specification of random effects probit model, unbalanced panel	266
9.9	Conditional logit model, unbalanced panel	268
9.10	Dynamic pooled probit model, unbalanced panel	271
9.11	Dynamic pooled probit model with initial conditions, unbalanced panel	272

9.12	Dynamic random effects probit model with initial conditions, unbalanced panel	273
9.13	Dynamic random effects probit model with initial conditions, partial effects	274
10.1	Dynamic pooled probit with IPW, unbalanced panel	288
10.2	Dynamic pooled probit with IPW, balanced panel	289
11.1	Poisson regression	297
11.2	Average marginal effects from Poisson regression	298
11.3	Poisson regression with robust standard errors	299
11.4	Negative binomial model	301
11.5	Average marginal effects from negative binomial model	301
11.6	Generalised negative binomial model	302
11.7	Zero-inflated Poisson model with constant zero-inflation probability	304
11.8	Average marginal effects from zero-inflated Poisson model with constant zero-inflation probability	304
11.9	Zero-inflated Poisson model with variable zero-inflation probability	305
11.10	Zero-inflated NB model for number of specialist visits with variable zero-inflation probability	306
11.11	Average marginal effects from zero-inflated NB model with variable zero-inflation probability	307
11.12	Logit-NB2 hurdle model	309
11.13	Average marginal effects on probability of receiving some health care from first part of hurdle model (Logit)	310
11.14	Average marginal effects on conditionally positive number of visits from second part of hurdle model (truncated NB2)	311
11.15	LCNB model with two latent classes	313
11.16	Summary statistics of fitted values by latent class (LCNB)	314
11.17	Average marginal effects of continuous regressors from LCNB model – latent class 1	315
11.18	Average marginal effects of continuous regressors from LCNB model – latent class 2	316
11.19	Average marginal effects of continuous regressors from LCNB model	316
11.20	Average incremental effects of male from LCNB model	317
11.21	AIC and BIC of alternative models	317
11.22	LCNB-Pan with two latent classes	323
11.23	Summary statistics of fitted values by latent class (LCNB-Pan)	323
11.24	AIC and BIC of NB, LCNB and LCNB-Pan (with two latent classes)	325
11.25	LCH-Pan (with two latent classes), with constant class membership	329
11.26	AIC and BIC of LCNB-Pan, hurdle and LCH-Pan	330
11.27	LCH-Pan (with two latent classes), with variable class membership	332

xii *Illustrations*

11.28	Summary statistics for individual p in LCH-Pan, with variable class membership	333
11.29	Summary statistics of fitted values by latent class in LCH-Pan, with variable class membership	335
11.30	AIC and BIC of LCNB-Pan and LCH-Pan (with two latent classes) with constant and variable class memberships	336
12.1	Nonlinear regression: exponential conditional mean model	344
12.2	Average marginal effects for exponential conditional mean model	346
12.3	Average marginal effects from Poisson regression	349
12.4	Exponential hazard function	353
12.5	Marginal effects from exponential hazard function	354
12.6	Weibull hazard function	356
12.7	Marginal effects from Weibull hazard function	356
12.8	Link test Weibull hazard function	357
12.9	Generalised gamma	358
12.10	Generalised gamma	360
12.11	Generalised gamma with heteroskedasticity	361
12.12	Generalised linear model; square root link, gamma distribution	363
12.13	Link test for generalised linear model	366
12.14	Summary results for alternative GLMs	367
12.15	Extended estimating equations	369
12.16	Finite mixture model (mixture of gammas)	372
12.17	Finite mixture model information criteria	375
12.18	Finite mixture model (three-component mixture of gammas)	376
12.19	Comparison of model specification and predictive performance	379

# Preface

Large-scale survey datasets, in particular complex survey designs such as panel data and surveys linked to administrative data, provide a rich source of information for health economists. Panel data offer the scope to control for individual heterogeneity and to model the dynamics of individual behaviour. The detailed information on individual morbidity characteristics and use of health care services contained in administrative data allows the analyst to model observed heterogeneity. However, the measures of outcome used in health economics do not always lend themselves to straightforward analysis and are often qualitative or categorical or display non-symmetrical and heteroskedastic distributions. These features of outcomes create challenges for estimating econometric models. The dramatic growth in computing power over recent years has been accompanied by the development of methods that help to overcome these challenges. The purpose of this book is to provide a practical guide to the skills necessary to put the required methods into practice.

This book highlights practical applications of econometric methods, illustrated using data on health from the British Health and Lifestyle Survey (HALS), the British Household Panel Survey (BHPS), the European Community Household Panel (ECHP), the US Medical Expenditure Panel Survey (MEPS) and the Survey of Health, Ageing and Retirement in Europe (SHARE). Throughout the book, there is a strong emphasis on applied work, illustrating the use of relevant computer software with code provided for Stata version 12 ([www.stata.com](http://www.stata.com)). Familiarity with the basic syntax and structure of Stata is assumed. The Stata code and extracts from the statistical output are embedded directly in the main text (using the font Courier New) and explained as we go along, for example:

```
use "c:\stata\data\bhps.dta" , clear
```

The corresponding Stata output appears alongside in a smaller font. The code presented in this book can be downloaded from the web pages of the Health, Econometrics and Data Group (HEDG): [www.york.ac.uk/res/herc/research/hedg/](http://www.york.ac.uk/res/herc/research/hedg/).

We do not attempt to provide a review of the extensive health economics literature that makes use of econometric methods (for surveys of the literature see Jones, 2000, 2009). Instead, the book is built around empirical case studies, rather than general theory, and the emphasis is on learning by example. We present a

detailed dissection of methods and results of some recent research papers written by the authors and our colleagues. Relevant methods are presented alongside the Stata code that can be used to implement them and the empirical results are discussed as we go along. To our knowledge, no comparable text exists. There are health economics texts and there are econometrics texts but these tend to focus on theory rather than application and tend not to bring the two disciplines together for the benefit of applied economists. The emphasis is on hands-on empirical analysis: the area that econometric texts tend to neglect. The closest in spirit is Angus Deaton's (1997) excellent book on the analysis of household surveys, but that emphasises issues in the economics of development, poverty and welfare rather than health. A general knowledge of microeconomic methods is assumed. For more details, readers can refer to texts such as Baltagi (2005), Cameron and Trivedi (2005), Greene (2003) and Wooldridge (2002b). Cameron and Trivedi (2009) provide a comprehensive source for the use of microeconometrics based on Stata and we have adopted one of their datasets in the analysis presented in [Chapters 3](#) and [12](#).

As the book is built around case studies, and these reflect the particular interests of the authors, we do not claim to cover the full diversity of topics within applied health economics. However, we hope that these examples provide guidance and inspiration for those working on other topics within the field who want to make use of econometric methods. The book is primarily aimed at advanced undergraduates and postgraduates in health economics, along with health economics researchers in academic, government and private sector organisations who want to learn more about empirical research methods. In addition, the book may be used by other applied economists, in areas such as labour and environmental economics, and by health and social statisticians.

The second edition has been revised and updated throughout and two new chapters, on the description and modelling of individual health care costs, have been added.

# Acknowledgements

Data from the British Household Panel Survey (BHPS) were supplied by the UK Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analysis or interpretations presented here. The European Community Household Panel Users' Database (ECHP), version of December 2003, was supplied by Eurostat. Data from the Health and Lifestyle Survey (HALS) were supplied by the UK Data Archive. SHARE data are from release 2.5.0 of wave 1 (2004). The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th framework programme (projects SHARE-I3, RII-CT- 2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th framework programme (SHARE-PREP, 211909 and SHARE-LEAP, 227822). Additional funding from the US National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064, IAG BSR06-11, R21 AG025169) as well as from various national sources is gratefully acknowledged (see [www.share-project.org](http://www.share-project.org) for a full list of funding institutions). The subsample of MEPS data corresponds to that used in Cameron and Trivedi (2009). Instructions on how to download these data are available at the Stata Press website ([www.stata-press.com/books/musr.html](http://www.stata-press.com/books/musr.html)).

We are very grateful to all of the co-authors of the joint work that we use as case studies: Paul Contoyannis, Martin Forster, Xander Koolman, Owen O'Donnell, Jennifer Roberts and Eddy van Doorslaer. The specific papers that are adapted for the case studies are the following:

- Bago d'Uva, T. (2006) 'Latent class models for health care utilisation', *Health Economics*, 15: 329–343.
- Bago d'Uva, T. and Jones, A. (2009) 'Health care utilisation in Europe: new evidence from the ECHP', *Journal of Health Economics*, 28: 265–279.
- Bago d'Uva, T., O'Donnell, O. and van Doorslaer, E. (2008) 'Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans', *International Journal of Epidemiology*, 37: 1375–1383.
- Balia, S. and Jones, A.M. (2008) 'Mortality, lifestyle and socio-economic status', *Journal of Health Economics*, 27: 1–26.

- Balia, S. and Jones, A.M. (2011) ‘Catching the habit: a study of inequality of opportunity in smoking-related mortality’, *Journal of the Royal Statistical Society Series A*, 174: 175–194.
- Contoyannis, P. and Rice, N. (2001) ‘The impact of health on wages: evidence from the British Household Panel Survey’, *Empirical Economics*, 26: 599–622.
- Contoyannis, P., Jones, A.M. and Rice, N. (2004b) ‘The dynamics of health in the British Household Panel Survey’, *Journal of Applied Econometrics*, 19: 473–503.
- Forster, M. and Jones, A.M. (2001) ‘The role of tobacco taxes in starting and quitting smoking: duration analysis of British data’, *Journal of the Royal Statistical Society Series A*, 164: 517–547.
- Jones, A.M. (2011) ‘Models for health care’, in Hendry, D. and Clements, M. (eds), *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press, pp. 625–654.
- Jones, A.M., Koolman, X. and Rice, N. (2006) ‘Health-related non-response in the BHPS and ECHP: using inverse probability weighted estimators in nonlinear models’, *Journal of the Royal Statistical Society Series A*, 169: 543–569.
- Jones, A.M., Rice, N. and Roberts, J. (2010) ‘Sick of work or too sick to work? Evidence on self-reported health shocks and early retirement from the BHPS’, *Economic Modelling*, 27: 866–880.

A draft of the first edition of the book was used as teaching material for a short course named ‘Applied Health Economics’, which was hosted by the Health, Econometrics and Data Group (HEDG) at the University of York, 19–30 June 2006. This course was part of the Marie Curie Training Programme in Applied Health Economics. We are grateful for the input from other members of HEDG who were involved with the course: Cristina Hernández Quevedo, Eugenio Zucchelli, Silvana Robone, Pedro Rosa Dias and Rodrigo Moreno Serra. Also we would like to thank the course participants for their valuable feedback on the material.

We are particularly grateful to Ranjeeta Thomas for her assistance in preparing the second edition, which included commenting on all of the chapters and the Stata code.

# Introduction

This book provides a practical guide to doing applied health economics. It is built around a series of case studies that are based on recent research. The first, which runs through the book, explores the dynamics of self-reported health in the British Household Panel Survey (BHPS). The aim is to investigate socioeconomic gradients in health, persistence of health problems and the difficulties created by sample attrition in panel data (Contoyannis *et al.*, 2004b; Jones *et al.*, 2006). The data for this and all the other case studies are introduced in [Chapter 1](#), which also introduces some general principles of survey design.

[Chapter 2](#) uses the BHPS sample to show how descriptive techniques, including graphs and tables, can be used to summarise and explore the raw data and provide an intuitive understanding of how variables are distributed and associated with each other. [Chapter 3](#) uses data on the medical expenditures of elderly individuals in the United States from the Medical Expenditure Panel Survey (MEPS) dataset. The chapter describes the challenges that modelling health care costs can present. Individual level data on medical expenditures typically feature a spike at zero, are heavily right skewed, have heavy tails and exhibit heteroskedasticity.

Subjective and self-reported measures of health raise questions of reliability. [Chapter 4](#) explores the issue of reporting bias using French data from the Survey of Health, Ageing and Retirement in Europe (SHARE). The standard ordered probit model is extended to include applications of the generalised ordered model and the ‘HOPIT’. These exploit hypothetical ‘vignettes’ to deal with reporting bias (Bago d’Uva *et al.*, 2008a).

Lifestyle factors, such as smoking and drinking, are thought to have an influence on health. But these health-related behaviours are individual choices that are themselves influenced by, often unobservable, individual characteristics such as time preference rates. [Chapter 5](#) uses data from the Health and Lifestyle Survey (HALS) to show how the multivariate probit model can be used to model mortality, morbidity and lifestyles jointly, taking account of the problem of unobservables (Balia and Jones, 2008). This illustrates the kind of models that can be applied to categorical data in cross-section surveys.

[Part III](#) moves from cross-sectional data to longitudinal data, in particular to duration analysis. There are two types of duration data: continuous and discrete time. [Chapter 6](#) takes the analysis of HALS a step further by estimating continuous

time duration models for initiation and cessation of smoking and for the risk of death (this draws on work by Forster and Jones, 2001, and Balia and Jones, 2011). [Chapter 7](#) illustrates convenient methods for discrete-time duration analysis. The BHPS is used to investigate the extent to which ‘health shocks’ are a factor that leads to early retirement, following Roberts and co-workers (2010).

Longitudinal data is the focus of [Part IV](#), which presents linear and nonlinear panel data regression methods. Linear models are covered in [Chapter 8](#), where BHPS data are used to estimate classical Mincerian wage equations that are augmented by measures of self-reported health (Contoyannis and Rice, 2001). [Chapter 9](#) stays with the BHPS but moves to nonlinear dynamic specifications (Contoyannis *et al.*, 2004b). The outcome of interest is a binary measure of health problems and the focus is on socioeconomic gradients in health. [Chapter 10](#) continues this analysis but shifts the emphasis to the potential problems created by sample attrition in panel data (Jones *et al.*, 2006). The chapter shows how to test for attrition bias and illustrates how inverse probability weights provide one way of dealing with the problem.

[Part V](#) turns to methods that are suitable for modelling individual data on health care utilisation when that is measured by numbers of visits or by levels of expenditure. Health care utilisation is most frequently modelled using count data regressions. [Chapter 11](#) analyses data on specialist visits taken from the European Community Household Panel (ECHP). The chapter reviews and applies standard methods and also introduces recent developments of the literature that use a latent class specification (Bago d’Uva, 2006). [Chapter 12](#) builds on the earlier analysis of individual health care costs in [Chapter 3](#). It uses the MEPS data to show how nonlinear regression models can augment the standard linear model when the outcome of interest is highly non-normal and it compares the performance of different specifications.

The key methods and Stata commands that are covered in each chapter are summarised below.

<i>Chapter</i>	<i>Key methods</i>	<i>Key Stata commands</i>
2	Descriptive statistics: Summary statistics Bar charts Empirical distribution functions Frequency tables	summarize graph bar distplot tabulate xttrans
3	Descriptive statistics: Histograms Box-whisker plots Normal plots Linear regression: Predictions	tw histogram graph box pnorm regress predict preserve & restore

<i>Chapter</i>	<i>Key methods</i>	<i>Key Stata commands</i>
	Diagnostics	estat hetest estat ovtest linktest
	Box–Cox regression	boxcox
4	Ordered probit Generalised ordered probit Interval regression HOPIT	oprobit ml intreg
5	Binary choice: Probit Multivariate choice: Multivariate probit Average partial effects Information criteria	probit  mvprobit preserve & restore estimates store estimates stats
6	Continuous time survival models: Kaplan–Meier and Nelson–Aalen Parametric models (exponential, Weibull, log-logistic, log-normal, Gompertz) Cox–Snell residuals	stset sts graph streg stcurve  line, gr combine
7	Discrete time survival models: Life tables  Complementary log-log models Unobserved frailty  Heckman–Singer model	ltable sts graph  cloglog pgmhaz8 xtcloglog hshaz
8	Linear panel data models: Random effects Fixed effects Hausman tests Hausman–Taylor estimator	xtset xtreg, re xtreg, fe hausman xthtaylor
9	Panel probit models: Pooled Random effects Average partial effects Conditional logit Dynamic panel probit	xtsum probit, dprobit xtprobit margins clogit redpace