#### Springer Series in Statistics

Advisors: P. Bickel, P. Diggle, S. Fienberg, U. Gather, I. Olkin, S. Zeger Juha M. Alho and Bruce D. Spencer

# Statistical Demography and Forecasting

With 33 Illustrations



Juha Alho Department of Statistics University of Joensuu Joensuu, Finland Bruce Spencer Department of Statistics Northwestern University Evanston, IL 60208 USA

Library of Congress Control Number: 2005926699 (hard cover) Library of Congress Control Number: 2005927649 (soft cover)

ISBN 10: 0-387-23530-2 (hard cover) ISBN 13: 978-0387-23530-1 (hard cover) ISBN 10: 0-387-22538-2 (soft cover) ISBN 13: 978-0387-22538-8 (soft cover) Printed on acid-free paper.

© 2005 Springer Science+Business Media, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, Inc. 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America. (TB/MVY)

9 8 7 6 5 4 3 2 1 SPIN 11011019 (hard cover)

SPIN 11013662 (soft cover)

springeronline.com

To Irja and Donna

#### Preface

Statistics and demography share important common roots, yet as academic disciplines they have grown apart. Even a casual survey of leading journals shows that cross-references are rare. This is unfortunate, because many social problems call for a multi-disciplinary approach. Both statistics and demography are necessary ingredients in any serious analysis of the sustainability of pension or health care systems in the aging societies, in the assessment of potential inequities of formula-based allocations to local governments, in the estimation of the size of elusive populations such as drug users, in the investigation of the consequences of social ills such as unemployment, and so forth. This book was written to bring together much of the basic statistical theory and methodology for estimating and forecasting population growth and its components of births, deaths, and migration. Although relatively simple mathematical methods have traditionally been used to assess demographic trends and their role in the society, use of modern statistical methods offers significant advantages for more accurately measuring population and vital rates, for forecasting the future, and for assessing the uncertainty of the demographic estimates and forecasts.

For statisticians the book provides a unique introduction to demographic problems in a familiar language. For demographers, actuaries, epidemiologists, and professionals in related fields the book presents a unified statistical outlook on both classical methods of demography and recent developments. The book provides a self-contained introduction to the statistical theory of demographic rates (births, deaths, migration) in a multi-state setting. The book has a dual character. On the one hand, it is a monograph that can be consumed by a lone reader. There are many results that have appeared in journals or working papers only. Some appear here for the first time. The book is also useful as a classroom text, and includes exercises and complements to explore special topics in detail without interrupting the flow of the text. More than half of the book is readily accessible to undergraduates, but to fully benefit from the complete text may require more maturity.

Joensuu, Finland Evanston, Illinois, USA Juha M. Alho Bruce D. Spencer

#### Acknowledgments

This book was some 15 years in the making. We are grateful to many colleagues and students for advice, encouragement and helpful comments, both specific and general. We thank Bill Bell, Katie Bench, Henry Bienen, Petra Can, Tom Espenshade, Steve Fienberg, Marty Frankel, Olavi Haimi, Joan Hill, Jan Hoem, Jeff Jenkins, Jay Kadane, Anne Kearney, Nico Keilman, Nathan Keyfitz, Donna Kostanich, Bill Kruskal, Esa Läärä, Jukka Lassila, Ron Lee, Risto Lehtonen, Chijien Lin, Lincoln Moses, Fred Mosteller, Tom Mule, Jukka Nyblom, Erkki Pahkinen, Päivi Partanen, Rita Petroni, Jiahe Qian, Dave Raglin, Chris Rhoads, Gregg Robinson, Mikko A. Salo, the late I. Richard Savage, Eric Schindler, Tom Severini, Eric Song, Richard Suzman, Shripad Tuljapurkar, Tarmo Valkonen, Jim Vaupel, Nic van de Walle, Larry Wu, Sandy Zabell. Shelby Haberman and Mary Mulry went above and beyond the call in close reading and advice. Responsibility for remaining errors, of course, remains with the authors.

During preparation of the book we received financial support from U.S. National Institute on Aging grant R01 AG10156-01A1 to Northwestern University; The Searle Fund grant on Limits of Empirical Social Science for Policy Analysis, to Northwestern University; U.S. Census Bureau contract 50-YABC-7-66020 with Abt, Associates; Academy of Finland Grants 8684, 41495, and 201408, Statistics Finland Grant 5012, and European Commission Grant HPSE-CT-2001-00095 to University of Joensuu; and European Commission Grant QLRT-2001-02500 to the Research Institute of the Finnish Economy.

Joensuu, Finland Evanston, Illinois, USA Juha M. Alho Bruce D. Spencer

### Contents

| Preface   | vii           |
|---|---------------|
| L ist of Examples                                 |               |
| List of Examples                                  |               |
| List of Figures                                   |               |
| Chapter 1. Introduction                           | 1             |
| 1. Role of Statistical Demography                 | 1             |
| 2. Guide for the Reader                           | 4             |
| 3. Statistical Notation and Preliminaries         | 4             |
| Chapter 2. Sources of Demographic Data            | 9             |
| 1. Populations: Open and Closed                   | 9             |
| 2. <i>De Facto</i> and <i>De Jure</i> Populations | 11            |
| 3. Censuses and Population Registers              | 15            |
| 4. Lexis Diagram and Classification of Events     | 16            |
| 5. Register Data and Epidemiologic Studies        | 19            |
| 5.1. Event Histories from Registers               | 19            |
| 5.2. Cohort and Case-Control Studies              | 19            |
| 5.3. Advantages and Disadvantages                 | 20            |
| 5.4. Confounding                                  | 22            |
| 6. Sampling in Censuses and Dual System Estima    | tion 24       |
| Exercises and Complements                         | 27            |
| Chapter 3. Sampling Designs and Inference         | 31            |
| 1. Simple Random Sampling                         | 32            |
| 2. Subgroups and Ratios                           | 35            |
| 3. Stratified Sampling                            | 36            |
| 3.1. Introduction                                 | 36            |
| 3.2. Stratified Simple Random Sampling            | 37            |
| 3.3. Design Effect for Stratified Simple Random   | m Sampling 38 |
| 3.4. Poststratification                           | 39            |
| 4. Sampling Weights                               | 40            |
| 4.1. Why Weight?                                  | 40            |

|     | 4.2. Forming Weights  | 41  |
|-----|---|-----|
|     | 4.3. Non-Response Adjustments                                 | 43  |
|     | 4.4. Effect of Weighting on Precision                         | 45  |
| 5.  | Cluster Sampling  | 46  |
|     | 5.1. Introduction   | 46  |
|     | 5.2. Single Stage Sampling with Replacement                   | 47  |
|     | 5.3. Single Stage Sampling without Replacement                | 47  |
|     | 5.4. Multi-Stage Sampling                                     | 49  |
|     | 5.5. Stratified Samples                                       | 50  |
| 6.  | Systematic Sampling   | 52  |
| 7.  | Distribution Theory for Sampling                              | 53  |
|     | 7.1. Central Limit Theorems                                   | 53  |
|     | 7.2. The Delta Method   | 55  |
|     | 7.3. Estimating Equations                                     | 56  |
| 8.  | Replication Estimates of Variance                             | 61  |
|     | 8.1. Jackknife Estimates                                      | 61  |
|     | 8.2. Bootstrap Estimates                                      | 62  |
|     | 8.3. Replication Weights                                      | 63  |
|     | Exercises and Complements                                     | 64  |
| Cha | pter 4. Waiting Times and Their Statistical                   |     |
|     | Estimation  | 71  |
| 1.  | Exponential Distribution                                      | 71  |
| 2.  | General Waiting Time  | 76  |
|     | 2.1. Hazards and Survival Probabilities                       | 76  |
|     | 2.2. Life Expectancies and Stable Populations                 | 79  |
|     | 2.2.1. Life Expectancy  | 79  |
|     | 2.2.2. Life Table Populations and Stable Populations          | 81  |
|     | 2.2.3. Changing Mortality                                     | 82  |
|     | 2.2.4. Basics of Pension Funding                              | 84  |
|     | 2.2.5. Effect of Heterogeneity                                | 85  |
|     | 2.3. Kaplan-Meier and Nelson-Aalen Estimators                 | 85  |
|     | 2.4. Estimation Based on Occurrence-Exposure Rates            | 88  |
| 3.  | Estimating Survival Proportions                               | 91  |
| 4.  | Childbearing as a Repeatable Event                            | 93  |
|     | 4.1. Poisson Process Model of Childbearing                    | 93  |
|     | 4.2. Summary Measures of Fertility and Reproduction           | 96  |
|     | 4.3. Period and Cohort Fertility                              | 101 |
|     | 4.3.1. Cohort Fertility is Smoother                           | 101 |
|     | 4.3.2. Adjusting for Timing                                   | 103 |
|     | 4.3.3. Effect of Parity on Pure Period Measures               | 104 |
|     | 4.4. Multiple Births and Effect of Pregnancy on Exposure Time | 106 |
| 5.  | Poisson Character of Demographic Events                       | 107 |
| 6.  | Simulation of Waiting Times and Counts                        | 109 |
|     | Exercises and Complements                                     | 110 |

| Cha | pter 5. Regression Models for Counts and Survival     | 117 |
|-----|---|-----|
| 1.  | Generalized Linear Models                             | 118 |
|     | 1.1. Exponential Family                               | 118 |
|     | 1.2. Use of Explanatory Variables                     | 119 |
|     | 1.3. Maximum Likelihood Estimation                    | 119 |
|     | 1.4. Numerical Solution                               | 120 |
|     | 1.5. Inferences                                       | 121 |
|     | 1.6. Diagnostic Checks                                | 122 |
| 2.  | Binary Regression                                     | 123 |
|     | 2.1. Interpretation of Parameters and Goodness        |     |
|     | of Fit  | 123 |
|     | 2.2. Examples of Logistic Regression                  | 124 |
|     | 2.3. Applicability in Case-Control Studies            | 129 |
| 3.  | Poisson Regression                                    | 130 |
|     | 3.1. Interpretation of Parameters                     | 130 |
|     | 3.2. Examples of Poisson Regression                   | 131 |
|     | 3.3. Standardization                                  | 133 |
|     | 3.4. Loglinear Models for Capture-Recapture Data      | 136 |
| 4.  | Overdispersion and Random Effects                     | 138 |
|     | 4.1. Direct Estimation of Overdispersion              | 139 |
|     | 4.2. Marginal Models for Overdispersion               | 139 |
|     | 4.3. Random Effect Models                             | 140 |
| 5.  | Observable Heterogeneity in Capture-Recapture Studies | 143 |
| 6.  | Bilinear Models                                       | 146 |
| 7.  | Proportional Hazards Models for Survival              | 150 |
| 8.  | Heterogeneity and Selection by Survival               | 154 |
| 9.  | Estimation of Population Density                      | 156 |
| 10. | Simulation of the Regression Models                   | 158 |
|     | Exercises and Complements                             | 159 |
| Cha | pter 6. Multistate Models and Cohort-Component        |     |
|     | Book-Keeping  | 166 |
| 1.  | Multistate Life-Tables                                | 167 |
|     | 1.1. Numerical Solution Using Runge-Kutta Algorithm   | 167 |
|     | 1.2. Extension to Multistate Case                     | 168 |
|     | 1.3. Duration-Dependent Life-Tables                   | 172 |
|     | 1.3.1. Heterogeneity Attributable to Duration         | 172 |
|     | 1.3.2. Forms of Duration-Dependence                   | 173 |
|     | 1.3.3. Aspects of Computer Implementation             | 174 |
|     | 1.3.4. Policy Significance of Duration-Dependence     | 175 |
|     | 1.4. Nonparametric Intensity Estimation               | 175 |
|     | 1.5. Analysis of Nuptiality                           | 177 |
|     | 1.6. A Model for Disability Insurance                 | 179 |
| 2.  | Linear Growth Model                                   | 180 |
|     | 2.1. Matrix Formulation                               | 180 |

|     | 2.2. Stable Populations                                    | 183 |
|-----|--|-----|
|     | 2.3. Weak Ergodicity                                       | 185 |
| 3.  | Open Populations and Parametrization of Migration          | 186 |
|     | 3.1. Open Population Systems                               | 186 |
|     | 3.2. Parametric Models                                     | 186 |
|     | 3.2.1. Migrant Pool Model                                  | 187 |
|     | 3.2.2. Bilinear Models                                     | 187 |
| 4.  | Demographic Functionals                                    | 189 |
| 5.  | Elementwise Aspects of the Matrix Formulation              | 191 |
| 6.  | Markov Chain Models  | 191 |
|     | Exercises and Complements                                  | 193 |
| Cha | apter 7. Approaches to Forecasting Demographic Rates       | 198 |
| 1.  | Trends, Random Walks, and Volatility                       | 198 |
| 2.  | Linear Stationary Processes                                | 201 |
|     | 2.1. Properties and Modeling                               | 202 |
|     | 2.1.1. Definition and Basic Properties                     | 202 |
|     | 2.1.2. ARIMA Models  | 203 |
|     | 2.1.3. Practical Modeling                                  | 206 |
|     | 2.2. Characterization of Predictions and Prediction Errors | 210 |
|     | 2.2.1. Stationary Processes                                | 210 |
|     | 2.2.2. Integrated Processes                                | 211 |
|     | 2.2.3. Cross-Correlations                                  | 216 |
| 3.  | Handling of Nonconstant Mean                               | 216 |
|     | 3.1. Differencing  | 216 |
|     | 3.2. Regression  | 218 |
|     | 3.3. Structural Models                                     | 219 |
| 4.  | Heteroscedastic Innovations                                | 220 |
|     | 4.1. Deterministic Models of Volatility                    | 221 |
|     | 4.2. Stochastic Volatility                                 | 222 |
|     | Exercises and Complements                                  | 223 |
| Cha | apter 8. Uncertainty in Demographic Forecasts: Concepts,   |     |
|     | Issues, and Evidence                                       | 226 |
| 1.  | Historical Aspects of Cohort-Component Forecasting         | 228 |
|     | 1.1. Adoption of the Cohort-Component Approach             | 228 |
|     | 1.2. Whelpton's Legacy                                     | 228 |
|     | 1.3. Do We Know Better Now?                                | 231 |
| 2.  | Dimensionality Reduction for Mortality                     | 234 |
|     | 2.1. Age-Specific Mortality                                | 234 |
|     | 2.2. Cause-Specific Mortality                              | 236 |
| 3.  | Conceptual Aspects of Error Analysis                       | 238 |
|     | 3.1. Expected Error and Empirical Error                    | 238 |
|     | 3.2. Decomposing Errors                                    | 238 |
|     | 3.2.1. Error Classifications                               | 238 |
|     | 3.2.2. Alternative Decompositions                          | 240 |

|     | 3.3. Acknowledging Model Error                           | 240 |
|-----|--|-----|
|     | 3.3.1. Classes of Parametric Models                      | 240 |
|     | 3.3.2. Data Period Bias                                  | 241 |
|     | 3.4. Feedback Effects of Forecasts                       | 242 |
|     | 3.5. Interpretation of Prediction Intervals              | 244 |
|     | 3.5.1. Uncertainty in Terms of Subjective Probabilities  | 244 |
|     | 3.5.2. Frequency Properties of Prediction Intervals      | 248 |
|     | 3.6. Role of Judgment                                    | 249 |
|     | 3.6.1. Expert Arguments                                  | 249 |
|     | 3.6.2. Scenarios   | 250 |
|     | 3.6.3. Conditional Forecasts                             | 251 |
| 4.  | Practical Error Assessment                               | 251 |
|     | 4.1. Error Measures                                      | 252 |
|     | 4.2. Baseline Forecasts                                  | 253 |
|     | 4.3. Modeling Errors in World Forecasts                  | 256 |
|     | 4.3.1. An Error Model for Growth Rates                   | 256 |
|     | 4.3.2. Second Moments                                    | 257 |
|     | 4.3.3. Predictive Distributions for Countries and the    |     |
|     | World  | 259 |
|     | 4.4. Random Jump-off Values                              | 261 |
|     | 4.4.1. Jump-off Population                               | 262 |
|     | 4.4.2. Mortality   | 263 |
| 5.  | Measuring Correlatedness                                 | 264 |
|     | Exercises and Complements                                | 267 |
| Cha | apter 9. Statistical Propagation of Error in Forecasting | 269 |
| 1.  | Törnqvist's Contribution                                 | 269 |
| 2.  | Predictive Distributions                                 | 271 |
|     | 2.1. Regression with a Known Covariance Structure        | 271 |
|     | 2.2. Random Walks  | 274 |
|     | 2.3. ARIMA(1,1,0) Models                                 | 276 |
| 3.  | Forecast as a Database and Its Uses                      | 277 |
| 4.  | Parametrizations of Covariance Structure                 | 278 |
|     | 4.1. Effect of Correlations on the Variance of a Sum     | 279 |
|     | 4.2. Scaled Model for Error                              | 280 |
|     | 4.3. Structure of Error in Migration Forecasts           | 283 |
| 5.  | Analytical Propagation of Error                          | 284 |
|     | 5.1. Births  | 284 |
|     | 5.2. General Linear Growth                               | 285 |
| 6.  | Simulation Approach and Computer Implementation          | 287 |
| 7.  | Post Processing  | 289 |
|     | 7.1. Altering a Distributional Form                      | 289 |
|     | 7.2. Creating Correlated Populations                     | 292 |
|     | 7.2.1. Use of Seeds                                      | 292 |
|     | 7.2.2. Sorting Techniques                                | 293 |
|     | Exercises and Complements                                | 294 |

| Cha | pter 10. Errors in Census Numbers                                | 296 |
|-----|--|-----|
| 1.  | Introduction   | 296 |
| 2.  | Effects of Errors on Estimates and Forecasts                     | 297 |
|     | 2.1. Effects on Mortality Rates                                  | 297 |
|     | 2.2. Effects on Forecasts  | 298 |
|     | 2.3. Effects on Evaluation of Past Population Forecasts          | 298 |
| 3.  | Use of Demographic Analysis to Assess Error in U.S. Censuses     | 299 |
| 4.  | Assessment of Dual System Estimates of Population Size           | 300 |
| 5.  | Decomposition of Error in the Dual System Estimator              | 303 |
|     | 5.1. A Probability Model for the Census                          | 303 |
|     | 5.2. Poststratification  | 304 |
|     | 5.3. Overview of Error Components                                | 305 |
|     | 5.4. Data Error Bias   | 308 |
|     | 5.5. Decomposition of Model Bias                                 | 309 |
|     | 5.5.1. Synthetic Estimation Bias and Correlation Bias            | 309 |
|     | 5.5.2. Poststratified Estimator                                  | 310 |
|     | 5.6. Estimation of Correlation Bias in a Poststratified Dual     |     |
|     | System Estimator   | 312 |
|     | 5.7. Estimation of Synthetic Estimation Bias in a Poststratified |     |
|     | Dual System Estimator  | 314 |
| 6.  | Assessment of Error in Functions of Dual System Estimators       |     |
|     | and Functions of Census Counts                                   | 316 |
|     | 6.1. Overview  | 316 |
|     | 6.2. Computation   | 317 |
|     | Exercises and Complements  | 319 |
| Cha | pter 11. Financial Applications                                  | 327 |
| 1.  | Predictive Distribution of Adjustment for Life Expectancy        |     |
|     | Change   | 327 |
|     | 1.1. Adjustment Factor for Mortality Change                      | 327 |
|     | 1.2. Sampling Variation in Pension Adjustment Factors            | 329 |
|     | 1.3. The Predictive Distribution of the Pension                  |     |
|     | Adjustment Factor  | 330 |
| 2.  | Fertility Dependent Pension Benefits                             | 332 |
| 3.  | Measuring Sustainability   | 335 |
| 4.  | State Aid to Municipalities                                      | 337 |
| 5.  | Public Liabilities   | 339 |
|     | 5.1. Economic Series   | 340 |
|     | 5.2. Wealth in Terms of Random Returns and Discounting           | 340 |
|     | 5.3. Random Public Liability                                     | 341 |
|     | Exercises and Complements  | 342 |
| Cha | pter 12. Decision Analysis and Small Area Estimates              | 344 |
| 1.  | Introduction   | 344 |
| 2.  | Small Area Analysis  | 345 |
|     | 5  |     |

| 3.  | Formula-Based Allocations                                   | 346 |
|-----|---|-----|
|     | 3.1. Theoretical Construction                               | 346 |
|     | 3.1.1. Apportionment of the U.S. House of                   |     |
|     | Representatives   | 347 |
|     | 3.1.2. Rationale Behind Allocation Formulas                 | 348 |
|     | 3.2. Effect of Inaccurate Demographic Statistics            | 349 |
|     | 3.3. Beyond Accuracy  | 350 |
| 4.  | Decision Theory and Loss Functions                          | 351 |
|     | 4.1. Introduction   | 351 |
|     | 4.2. Decision Theory for Statistical Agencies               | 353 |
|     | 4.3. Loss Functions for Small Area Estimates                | 357 |
|     | 4.4. Loss Functions for Apportionment and Redistricting     | 359 |
|     | 4.1.1. Apportionment  | 359 |
|     | 4.1.2. Redistricting  | 360 |
|     | 4.5. Loss Functions and Allocation of Funds                 | 361 |
|     | 4.5.1. Effects of Over- and Under-Allocation                | 361 |
|     | 4.5.2. Formula Nonoptimality                                | 362 |
|     | 4.5.3. Optimal Data Quality with Multiple Statistics        |     |
|     | and Uses  | 363 |
| 5.  | Comparing Risks of Adjusted and Unadjusted Census           |     |
|     | Estimates   | 363 |
|     | 5.1. Accounting for Variances of Bias Estimates             | 364 |
|     | 5.2. Effect of Unmeasured Biases on Comparisons of Accuracy | 365 |
| 6.  | Decision Analysis of Adjustment for Census Undercount       | 365 |
| 7.  | Cost-Benefit Analysis of Demographic Data                   | 367 |
|     | Exercises and Complements                                   | 368 |
| Ref | erences   | 371 |
| Aut | hor Index   | 397 |
| Sub | ject Index  | 405 |

# List of Examples

| Chap | ter 2. Sources of Demographic Data                          | 9  |
|------|---|----|
| 1.1. | Who Counts in the U.S. Census?                              | 10 |
| 1.2. | Who Belongs to the Sami Population?                         | 10 |
| 2.1. | Accident Rates in Nordic Countries.                         | 12 |
| 2.2. | Undercount in U.S. Censuses.                                | 12 |
| 2.3. | What Is a Household?  | 14 |
| 2.4. | Corporate Demography.                                       | 14 |
| 3.1. | Nigerian Censuses.  | 15 |
| 5.1. | British Doctors' Study.                                     | 20 |
| 5.2. | Doll and Hill Study.  | 21 |
| 6.1. | Underreporting of Occupational Diseases.                    | 26 |
| 6.2. | Numbers of Drug Users.                                      | 26 |
| Chap | ter 3. Sampling Designs and Inference                       | 31 |
| 1.1. | The 1970 Draft Lottery in the U.S.                          | 33 |
| 1.2. | Child Stunting.   | 34 |
| 3.1. | NELS:88 Base-Year School Sample.                            | 37 |
| 3.2. | Design Effects for NELS:88.                                 | 39 |
| 3.3. | Poststratification in the 1990 U.S. Post Enumeration Survey |    |
|      | (PES).  | 40 |
| 4.1. | NELS:88 First Followup Schools.                             | 41 |
| 4.2. | Extreme Weights in the 1990 U.S. PES.                       | 42 |
| 4.3. | Nonparticipation in a Survey in an STD Clinic.              | 43 |
| 4.4. | The Dual System Estimator as a Propensity-Weighted Census.  | 44 |
| 4.5. | Extreme Weights in the Survey of Consumer Finance.          | 46 |
| 5.1. | Survey of the Homeless in Chicago.                          | 48 |
| 5.2. | NELS:88 Sample of Students.                                 | 51 |
| 5.3. | The U.S. Current Population Survey.                         | 51 |
| 6.1. | Systematic Sampling of Private Schools in the National      |    |
|      | Assessment of Educational Progress.                         | 53 |
| 7.1. | Model-Based Variance of the Dual System Estimator (DSE).    | 56 |
| 7.2. | Design-Based Variance of the Dual System Estimator (DSE).   | 58 |
|      |   |    |

#### xx List of Examples

| 7.3.<br>7.4.             | Parameter Interpretation Under An Erroneous Model.<br>Fieller Intervals for a Ratio Estimator. | 59<br>60 |
|--------------------------|--|----------|
| Chan                     | ter 4 Waiting Times and Their Statistical Estimation   | 71       |
| 1 1                      | Memorylessness of Exponential Waiting Time   | 71       |
| 1.1.                     | Independent Causes of Death  | 72       |
| 1.2.                     | Cross-Sectional Heterogeneity of Constant Hazard Rates   | 72       |
| 1.5.                     | Gamma Distribution for Frailty   | 75       |
| 1. <del>4</del> .<br>2.1 | Weibull Distribution   | 75<br>77 |
| 2.1.                     | Linear Survival Functions  | 77       |
| 2.2.                     | Balducci Model for Survival Function   | 77       |
| 2.5.                     | Competing Risks  | 77       |
| 2.1.                     | Mortality and Marital Status in Finland  | 78       |
| 2.6                      | Effect of Changes in Hazards on Life Expectancy  | 83       |
| 2.7                      | Life Expectancy Calculation from Kaplan-Meier Estimates  | 86       |
| 2.8.                     | Survival Probabilities for Habsburgs.  | 86       |
| 2.9.                     | Actuarial Estimator.   | 89       |
| 2.10.                    | Distribution of Death During First Year.   | 90       |
| 2.11.                    | Proportion of Deaths During First Days.  | 90       |
| 4.1.                     | Age-Specific Fertility Rates for Italy and the U.S.  | 95       |
| 4.2.                     | Finnish Fertility, 1776–1999.  | 97       |
| 4.3.                     | Time Trends in Sex Ratios in Finland.  | 98       |
| 4.4.                     | Alternative Measures of Mean Age at Childbearing, Finland                                      |          |
|                          | 2000.  | 100      |
| 4.5.                     | Parity Progression Ratios.   | 105      |
| 6.1.                     | Simulation of Weibull Random Variates.   | 109      |
| Chap                     | ter 5. Regression Models for Counts and Survival   | 117      |
| 1.1.                     | Exponential Distribution.  | 118      |
| 1.2.                     | Bernoulli Distribution.  | 118      |
| 1.3.                     | Leverage in Simple Generalized Linear Model.   | 122      |
| 2.1.                     | Sex Ratios of the Habsburgs.   | 124      |
| 2.2.                     | Child Mortality among the Habsburgs.   | 125      |
| 2.3.                     | Testing Effects of Exposure on Illness.  | 125      |
| 2.4.                     | Detecting Confounding.   | 127      |
| 2.5.                     | Choosing the Sword.  | 127      |
| 3.1.                     | Poisson Models for Births.   | 131      |
| 3.2.                     | Mortality of Young Widows.   | 132      |
| 3.3.                     | Age-Period-Cohort Problem.   | 132      |
| 3.4.                     | Number of the Habsburg Offspring.  | 132      |
| 3.5.                     | Regression Models for Rates of Small Areas.  | 132      |
| 3.6.                     | Relative Risk of Mortality for Unemployed.   | 135      |
| 3.7.                     | Triple Systems Estimates of Numbers of Drug Users.   | 138      |
| 4.1.                     | Overdispersion in Habsburg Cohort Sizes.   | 142      |
| 5.1.                     | Heterogeneity in Reporting of Occupational Disease.  | 145      |
| 5.2.                     | Heterogeneity in Census Enumeration Probabilities.   | 145      |

| 6.1. | Lee-Carter Model for Mortality.                            | 147 |
|------|--|-----|
| 6.2. | Mortality among Elderly.                                   | 149 |
| 7.1. | A Simple Example of Cox Regression.                        | 150 |
| 7.2. | A Simple Example of Cox Regression with Censoring.         | 151 |
| 7.3. | Changes in Mortality of the Habsburgs.                     | 153 |
| 7.4. | Time-Varying Covariates.                                   | 154 |
| 7.5. | Likelihood for Matched Studies.                            | 154 |
| Char | oter 6. Multistate Models and Cohort-Component             |     |
| 1    | Book-Keeping   | 166 |
| 1.1. | Runge-Kutta Illustration.                                  | 167 |
| 1.2. | A Three-State Labor Force Model.                           | 169 |
| 1.3. | Hazards Producing a Linear Solution.                       | 170 |
| 1.4. | Remarriage Probability Varies with Time Spent Non-married. | 172 |
| 2.1. | Two-Sex Problem.   | 183 |
| 4.1. | Marriage Prevalence as a Functional.                       | 190 |
| 4.2. | Life Expectancy as a Functional.                           | 190 |
| 4.3. | Age Dependency Ratio.                                      | 190 |
| 4.4. | A Relation between Prevalence and Incidence.               | 190 |
| 6.1. | Metapopulation of Butterflies.                             | 192 |
| Chap | oter 7. Approaches to Forecasting Demographic Rates        | 198 |
| 1.1. | Cohort Fertility Is Smoother.                              | 199 |
| 1.2. | Cholesky Decomposition.                                    | 201 |
| 2.1. | MA(q) Processes.   | 203 |
| 2.2. | AR(1) Processes.   | 203 |
| 2.3. | EWMA Processes.  | 205 |
| 2.4. | Vital Processes Appear Nonstationary.                      | 207 |
| 2.5. | Standard Error Under AR(1) Residuals.                      | 211 |
| 2.6. | Correlations of Forecast Errors For AR(1) Processes.       | 211 |
| 2.7. | Correlations of Forecast Errors for Integrated AR(1)       |     |
|      | Processes.   | 212 |
| 2.8. | Standard Error and Random Error.                           | 212 |
| 3.1. | Forecasting a Random Walk with a Drift.                    | 217 |
| 3.2. | Trend in Finnish Fertility up to 1930.                     | 217 |
| 3.3. | Alternative Time Series Forecasts of the U.S. Growth Rate. | 219 |
| 3.4. | Stochastic Local Level Process.                            | 220 |
| 3.5. | Stochastic Linear Trend Process.                           | 220 |
| 4.1. | A Heteroscedastic Process with Time Invariant              |     |
|      | Autocorrelations.  | 222 |
| Chap | oter 8. Uncertainty in Demographic Forecasts: Concepts,    |     |
|      | Issues, and Evidence                                       | 226 |
| 1.1. | Cohort Approach to Fertility Forecasting.                  | 231 |
| 1.2. | Effect of Marriage Duration on Fertility.                  | 232 |
| 1.3. | Was the Baby-Boom a Unique Phenomenon?                     | 232 |

| 1.4.          | Trend Extrapolation Versus Judgment.                       | 232 |
|---------------|--|-----|
| 1.5.          | Counterintuitive Data on Economic Shocks and               |     |
|               | Demographics.  | 233 |
| 2.1.          | Rates of Mortality Decline in Europe.                      | 235 |
| 2.2.          | Emerging Cause of Death.                                   | 237 |
| 3.1.          | Sensitivity to Assumptions.                                | 239 |
| 3.2.          | Planning Optimism.   | 244 |
| 3.3.          | Achieving Approximate Consensus on Probabilities.          | 246 |
| 3.4.          | Elicitation of Probabilities via Betting.                  | 247 |
| 3.5.          | Assessing Prediction Intervals for ARIMA Forecasts.        | 248 |
| 3.6.          | Mortality Differences Across Countries.                    | 250 |
| 3.7.          | Fertility in the Mediterranean Countries.                  | 250 |
| 3.8.          | Migration to Germany.                                      | 250 |
| 4.1.          | Error Estimates for Fertility Forecasts in Europe.         | 254 |
| 4.2.          | Error Estimates for Mortality Forecasts in Europe.         | 255 |
| 5.1.          | Constant Correlations Across Ages.                         | 265 |
| 5.2.          | Constant Correlations Across Causes of Death.              | 265 |
| 5.3.          | Uncorrelated Errors for Different Vital Rates.             | 265 |
| 5.4.          | Constant Correlations Across Countries within a            |     |
|               | Region.  | 266 |
| Char          | oter 9. Statistical Propagation of Error in Forecasting    | 269 |
| 2.1.          | Posterior of an AR(1) Process With Known Autocorrelations. | 274 |
| 2.2           | Conditional Likelihood of an $AR(1)$ Process.              | 274 |
| 2.3.          | Predictive Distribution of a Random Walk.                  | 275 |
| 2.4.          | Predictive Distribution of a Random Walk With a Drift.     | 275 |
| 4.1.          | Independence, AR(1), and Perfect Dependence.               | 279 |
| 4.2.          | Error in a Cohort Survival Setting.                        | 279 |
| 4.3.          | Autoregressive Model for Correlations Across Age.          | 281 |
| 4.4.          | Specifying a Linear Process to Match Judgment.             | 282 |
| 5.1.          | Representation of a Closed Female Population.              | 285 |
| 61            | Storage Space Required by the Database                     | 288 |
| 7.1.          | Stochastic Forecast Database for Finland.                  | 290 |
| Char          | star 10 Errors in Consus Numbers                           | 206 |
|               | Doot Enumeration Surveys in the 1000 and 2000 U.S.         | 290 |
| 4.1.          | Consusses  | 200 |
| 12            | Cellsuses.   | 202 |
| 4. <i>2</i> . | Post Enumeration Survey in the U.K. in 2001.               | 204 |
| 5.1.          | Error Components in the 1000 LLS DES                       | 207 |
| 5.2.          | Error Components in the 2000 LLS. A C.E.                   | 207 |
| J.J.          | Error Components in the 2000 U.S. A.C.E.                   | 307 |
| 5.4.          | Estimates of Correlation Bias Based on DA Totals.          | 512 |
| 5.5.          | Estimates of Correlation Bias Based on DA Sex Ratios.      | 313 |
| 5.6.          | Surrogate Variables for Undercount and Overcount in the    | 215 |
|               | 2000 U.S. Census.  | 315 |

| Chapt | ter 12. Decision Analysis and Small Area Estimates         | 344 |
|-------|--|-----|
| 4.1.  | Asymmetric Consequences of Forecast Error.                 | 351 |
| 4.2.  | Posterior Risk Under Linear Loss.                          | 353 |
| 4.3.  | When Policy Makers Prefer Error to Accuracy.               | 354 |
| 4.4.  | Non-Adjustment of Undercount Estimates for Correlation     |     |
|       | Bias.  | 356 |
| 4.5.  | Adjustment for Correlation Bias for Hispanics in the 2000  |     |
|       | U.S. Census.   | 356 |
| 4.6.  | Alternative Estimates of Population.                       | 357 |
| 4.7.  | Value Judgements in Sample Allocation.                     | 358 |
| 4.8.  | Expected Loss of Adjusted and Unadjusted 2000 U.S. Census  |     |
|       | for Redistricting.   | 360 |
| 6.1.  | Expected Loss of Adjusted and Unadjusted 1990 U.S. Census. | 365 |
| 6.2.  | Expected Loss of Adjusted and Unadjusted 2000 U.S. Census, |     |
|       | A.C.E. Revision II.  | 366 |
| 7.1.  | Decennial Census.  | 367 |
| 7.2.  | Mid-Decade Census.   | 368 |

# List of Figures

| Ch | apter 2. Sources of Demographic Data                           |     |
|----|--|-----|
| 1. | Lexis Diagram.   | 17  |
| 2. | Example of Confounding.  | 24  |
| Ch | apter 4. Waiting Times and Their Statistical Estimation        |     |
| 1. | Log of Mortality Hazard for the Married, Widowed, and Single   |     |
|    | and Divorced Women in Finland, in 1998.                        | 78  |
| 2. | Log of the Hazard Increment of Mortality in Finland in         |     |
|    | 1881–1890 and 1986–1990, for Females and Males.                | 82  |
| 3. | Survival Probabilities for Females and Males among the         |     |
|    | Members of the Main Line of the Family of Habsburgs.           | 87  |
| 4. | The Distribution of Life Times of Those Born in 1994, Who      |     |
|    | Died in Age Zero, in Finland.                                  | 90  |
| 5. | Total Fertility Rate in Finland in 1776–1999 and in the United |     |
|    | States in 1920–1999.   | 97  |
| 6. | Sex Ratio at Birth (Actual and Smoothed) in Finland in         |     |
|    | 1751–2000.   | 98  |
| 7. | Approximate Completed Fertility for Birth Cohorts Born in      |     |
|    | Finland in 1905–1965.  | 102 |
| Ch | napter 6. Multistate Models and Cohort-Component               |     |
|    | Book-Keeping   |     |
| 1. | Average Relative Risk of Remarriage Among Widowed and          |     |
|    | Divorced as a Function of the Duration of Widowhood and        |     |
|    | Divorce, Respectively.   | 173 |
| 2. | Possible State Transitions in Nuptiality Processes.            | 177 |
| 3. | Relative Risk of Death Among Married as a Function of the      |     |
|    | Duration of Marriage: Average, in Age 30, in Age 40, and in    |     |
|    | Age 50.  | 178 |
| 4. | Distribution of Time Spent in the Divorced State, if Ever      |     |
|    | Divorced, for a Single at Age 17.                              | 179 |

| 5.       | Average Density of Male Migration in Finland, Across Three Regions, During 1987–1997.   | 188 |
|----------|---|-----|
| 6.       | Two Most Important Patterns of Deviation from Average Age   |     |
|          | Distribution of Migration Intensity.  | 188 |
| 7.       | Coefficients of Deviations from the Mean for the Six Flows,<br>During 1987–1997.  | 189 |
| Ch       | apter 7. Approaches to Forecasting Demographic Rates  |     |
| 1.       | Hypothetical Cohort and Period Fertility Under a Pure Period<br>Random Walk Model.  | 199 |
| 2.       | Hypothetical Mortality Rates and a Moving Average Estimate of their Level.  | 204 |
| 3.       | The Growth Rate of the U.S. Population in 1900–1999, and<br>Three Forecasts: AR(1) and ARIMA(2,1,0) with and without a            |     |
| 4.       | Constant Term.<br>Total Fertility Rate of Finland in 1920–1996, and its Forecast for  | 208 |
|          | 1997–2021 with 50% Prediction Intervals.  | 214 |
| 5.       | (A) Lag-Plot of the First Differences Y(t) at Lag 1.  | 215 |
| ~        | (B) Lag-Plot of the First Differences Y(t) at Lag 2.  | 215 |
| 6.       | Absolute First Differences of the U.S. Growth Rate in 1900–1999, and an Exponentially Smoothed Trend Estimate.                    | 221 |
| Ch       | apter 8. Uncertainty in Demographic Forecasts: Concepts,<br>Issues, and Evidence  |     |
| 1.       | Smoothed Rate of Decline in Age-Specific Mortality for<br>Females and Males and its Median Across 11 European                     |     |
|          | Countries, for Females, and for Males.  | 235 |
| 2.       | Distribution of Absolute Errors of Decline in Growth Rate.  | 243 |
| 3.       | Change in the Expected Value for the Probability of Heads in a<br>Sequence of Coin Tossing Experiments for an Individual with a   |     |
|          | Prior Expectation of 0.9 and an Individual with a Prior   |     |
| 4.       | Expectation of 0.1.<br>Median Relative Error of Fertility Forecast as a Function of   | 247 |
| 5        | Average, and a Random Walk Approximation.   | 254 |
| 5.       | Lead Time for Nine Countries with Long Data Series, their<br>Average and a Random Walk Approximation                              | 256 |
|          | Average, and a Random wark Approximation.   | 250 |
| Ch<br>1. | apter 9. Statistical Propagation of Error in Forecasting<br>Predictive Distribution of a Fertility Measure and its Modified       |     |
|          | Distribution.   | 291 |
| Ch       | apter 11. Financial Applications  |     |
| 1.       | Predictive Distribution of the Adjustment Factor in 2010–2060:<br>Median, First and Third Quartiles, and First and Ninth Deciles. | 332 |

| 2. | Predictive Didtribution of Old-Age Dependency Ratio (Ages       |     |
|----|---|-----|
|    | 60+/Ages 20-59) in Finland in 2010, 2030, and 2050.             | 334 |
| 3. | Pension Contributions, as % of the Total Wages of the Covered   |     |
|    | Employees, in Finland in 1995–2070, Under Current Rules and     |     |
|    | Under a Fertility Dependent Rule, if the Population Follows the |     |
|    | High Old-Age Dependency Ratio Variant.                          | 335 |
| 4. | Replacement Rate and Contribution Rate Under Full Wages         |     |
|    | Indexation and Full Wage-Bill Indexation, and an Example of     |     |
|    | Potential Viable Region $\{(c, r)   c \le 0.38, r \ge 0.28\}$ . | 336 |
| 5. | Relative Burden of Social and Health Care Allocations in        |     |
|    | 1940–1997 in Finland, and the Median, Quartiles, and First and  |     |
|    | Ninth Deciles of its Predictive Distribution in 1998–2050.      | 339 |