

# Demographic Forecasting





# Demographic Forecasting



**Federico Girosi and Gary King**

with contributions from Kevin Quinn and Gregory Wawro

PRINCETON UNIVERSITY PRESS

Princeton and Oxford

Copyright © 2008 by Princeton University Press

Published by Princeton University Press, 41 William Street, Princeton, New Jersey 08540

In the United Kingdom: Princeton University Press, 6 Oxford Street,  
Woodstock, Oxfordshire OX20 1TW

All Rights Reserved

Library of Congress Cataloging-in-Publication Data

Girosi, Federico.

Demographic forecasting/Gary King and Federico Girosi.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-691-13094-1 (alk. paper)—978-0-691-13095-8 (pbk.: alk. paper)

1. Mortality—Forecasting—Methodology. 2. Mortality—Statistical methods.

3. Demography. I. King, Gary, 1958-. II. Title.

HB1321.G56 2008

304.6/40112 22

2008062102

British Library Cataloging-in-Publication Data is available

This book has been composed in Times

Printed on acid-free paper. ∞

press.princeton.edu

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

# Contents

<i>List of Figures</i>	xi
<i>List of Tables</i>	xiii
<i>Preface</i>	xv
<i>Acknowledgments</i>	xvii
<b>1 Qualitative Overview</b>	<b>1</b>
1.1 Introduction	1
1.2 Forecasting Mortality	3
1.2.1 The Data	3
1.2.2 The Patterns	5
1.2.3 Scientific versus Optimistic Forecasting Goals	8
1.3 Statistical Modeling	11
1.4 Implications for the Bayesian Modeling Literature	15
1.5 Incorporating Area Studies in Cross-National Comparative Research	16
1.6 Summary	18
<b>Part I Existing Methods for Forecasting Mortality</b>	<b>19</b>
<b>2 Methods without Covariates</b>	<b>21</b>
2.1 Patterns in Mortality Age Profiles	22
2.2 A Unified Statistical Framework	24
2.3 Population Extrapolation Approaches	25
2.4 Parametric Approaches	26
2.5 A Nonparametric Approach: Principal Components	28
2.5.1 Introduction	28
2.5.2 Estimation	32
2.6 The Lee-Carter Approach	34
2.6.1 The Model	34
2.6.2 Estimation	36
2.6.3 Forecasting	36
2.6.4 Properties	38
2.7 Summary	42
<b>3 Methods with Covariates</b>	<b>43</b>
3.1 Equation-by-Equation Maximum Likelihood	43
3.1.1 Poisson Regression	43
3.1.2 Least Squares	44
3.1.3 Computing Forecasts	46
3.1.4 Summary Evaluation	47
3.2 Time-Series, Cross-Sectional Pooling	48
3.2.1 The Model	48
3.2.2 Postestimation Intercept Correction	49
3.2.3 Summary Evaluation	49
3.3 Partially Pooling Cross Sections via Disturbance Correlations	50
3.4 Cause-Specific Methods with Microlevel Information	51

**vi • Contents**

3.4.1	Direct Decomposition Methods	51
	Modeling	51
3.4.2	Microsimulation Methods	52
3.4.3	Interpretation	53
3.5	Summary	53
 <b>Part II Statistical Modeling</b>		 55
<b>4</b>	<b>The Model</b>	<b>57</b>
4.1	Overview	57
4.2	Priors on Coefficients	59
4.3	Problems with Priors on Coefficients	60
4.3.1	Little Direct Prior Knowledge Exists about Coefficients	61
4.3.2	Normalization Factors Cannot Be Estimated	62
4.3.3	We Know about the Dependent Variable, Not the Coefficients	64
4.3.4	Difficulties with Incomparable Covariates	65
4.4	Priors on the Expected Value of the Dependent Variable	65
4.4.1	Step 1: Specify a Prior for the Dependent Variable	66
4.4.2	Step 2: Translate to a Prior on the Coefficients	67
4.4.3	Interpretation	68
4.5	A Basic Prior for Smoothing over Age Groups	69
4.5.1	Step 1: A Prior for $\mu$	69
4.5.2	Step 2: From the Prior on $\mu$ to the Prior on $\beta$	71
4.5.3	Interpretation	71
4.6	Concluding Remark	73
<b>5</b>	<b>Priors over Grouped Continuous Variables</b>	<b>74</b>
5.1	Definition and Analysis of Prior Indifference	74
5.1.1	A Simple Special Case	76
5.1.2	General Expressions for Prior Indifference	76
5.1.3	Interpretation	77
5.2	Step 1: A Prior for $\mu$	80
5.2.1	Measuring Smoothness	81
5.2.2	Varying the Degree of Smoothness over Age Groups	83
5.2.3	Null Space and Prior Indifference	83
5.2.4	Nonzero Mean Smoothness Functional	85
5.2.5	Discretizing: From Age to Age Groups	85
5.2.6	Interpretation	86
5.3	Step 2: From the Prior on $\mu$ to the Prior on $\beta$	92
5.3.1	Analysis	92
5.3.2	Interpretation	92
<b>6</b>	<b>Model Selection</b>	<b>94</b>
6.1	Choosing the Smoothness Functional	94
6.2	Choosing a Prior for the Smoothing Parameter	97
6.2.1	Smoothness Parameter for a Nonparametric Prior	98
6.2.2	Smoothness Parameter for the Prior over the Coefficients	100
6.3	Choosing Where to Smooth	104

6.4	Choosing Covariates	108
6.4.1	Size of the Null Space	109
6.4.2	Content of the Null Space	110
6.5	Choosing a Likelihood and Variance Function	112
6.5.1	Deriving the Normal Specification	112
6.5.2	Accuracy of the Log-Normal Approximation to the Poisson	114
6.5.3	Variance Specification	120
<b>7</b>	<b>Adding Priors over Time and Space</b>	<b>124</b>
7.1	Smoothing over Time	124
7.1.1	Prior Indifference and the Null Space	125
7.2	Smoothing over Countries	127
7.2.1	Null Space and Prior Indifference	128
7.2.2	Interpretation	130
7.3	Smoothing Simultaneously over Age, Country, and Time	131
7.4	Smoothing Time Trend Interactions	132
7.4.1	Smoothing Trends over Age Groups	133
7.4.2	Smoothing Trends over Countries	133
7.5	Smoothing with General Interactions	134
7.6	Choosing a Prior for Multiple Smoothing Parameters	136
7.6.1	Example	139
7.6.2	Estimating the Expected Value of the Summary Measures	141
7.7	Summary	144
<b>8</b>	<b>Comparisons and Extensions</b>	<b>145</b>
8.1	Priors on Coefficients versus Dependent Variables	145
8.1.1	Defining Distances	145
8.1.2	Conditional Densities	147
8.1.3	Connections to “Virtual Examples” in Pattern Recognition	147
8.2	Extensions to Hierarchical Models and Empirical Bayes	148
8.2.1	The Advantages of Empirical Bayes without Empirical Bayes	149
8.2.2	Hierarchical Models as Special Cases of Spatial Models	151
8.3	Smoothing Data without Forecasting	151
8.4	Priors When the Dependent Variable Changes Meaning	153
<b>Part III</b>	<b>Estimation</b>	<b>159</b>
<b>9</b>	<b>Markov Chain Monte Carlo Estimation</b>	<b>161</b>
9.1	Complete Model Summary	161
9.1.1	Likelihood	162
9.1.2	Prior for $\beta$	162
9.1.3	Prior for $\sigma_i$	162
9.1.4	Prior for $\theta$	163
9.1.5	The Posterior Density	164
9.2	The Gibbs Sampling Algorithm	164
9.2.1	Sampling $\sigma$	165
	The Conditional Density	165
	Interpretation	165

**viii • Contents**

9.2.2 Sampling $\theta$	166
The Conditional Density	166
Interpretation	166
9.2.3 Sampling $\beta$	167
The Conditional Density	167
Interpretation	168
9.2.4 Uncertainty Estimates	169
9.3 Summary	169
<b>10 Fast Estimation without Markov Chains</b>	<b>170</b>
10.1 Maximum A Posteriori Estimator	170
10.2 Marginal Maximum A Posteriori Estimator	171
10.3 Conditional Maximum A Posteriori Estimator	172
10.4 Summary	173
<b>Part IV Empirical Evidence</b>	<b>175</b>
<b>11 Illustrative Analyses</b>	<b>177</b>
11.1 Forecasts without Covariates: Linear Trends	178
11.1.1 Smoothing over Age Groups Only	178
11.1.2 Smoothing over Age and Time	181
11.2 Forecasts without Covariates: Nonlinear Trends	182
11.3 Forecasts with Covariates: Smoothing over Age and Time	187
11.4 Smoothing over Countries	189
<b>12 Comparative Analyses</b>	<b>196</b>
12.1 All Causes in Males	197
12.2 Lung Disease in Males	200
12.2.1 Comparison with Least Squares	202
12.2.2 Country-by-Country Analysis	203
12.3 Breast Cancer in Females	205
12.3.1 Comparison with Least Squares	205
12.3.2 Country-by-country Analysis	205
12.4 Comparison on OECD Countries	206
12.4.1 Transportation Accidents in Males	208
12.4.2 Cardiovascular Disease in Males	210
<b>13 Concluding Remarks</b>	<b>211</b>
<b>Appendixes</b>	<b>213</b>
<b>A Notation</b>	<b>215</b>
A.1 Principles	215
A.2 Glossary	216
<b>B Mathematical Refresher</b>	<b>219</b>
B.1 Real Analysis	219
B.1.1 Vector Space	219



B.1.2	Metric Space	220
B.1.3	Normed Space	221
B.1.4	Scalar Product Space	222
B.1.5	Functions, Mappings, and Operators	223
B.1.6	Functional	224
B.1.7	Span	224
B.1.8	Basis and Dimension	224
B.1.9	Orthonormality	225
B.1.10	Subspace	225
B.1.11	Orthogonal Complement	226
B.1.12	Direct Sum	226
B.1.13	Projection Operators	227
B.2	Linear Algebra	229
B.2.1	Range, Null Space, Rank, and Nullity	229
B.2.2	Eigenvalues and Eigenvectors for Symmetric Matrices	232
B.2.3	Definiteness	234
B.2.4	Singular Values Decomposition	234
Definition		234
For Approximation		235
B.2.5	Generalized Inverse	236
B.2.6	Quadratic Form Identity	238
B.3	Probability Densities	239
B.3.1	The Normal Distribution	239
B.3.2	The Gamma Distribution	239
B.3.3	The Log-Normal Distribution	240
<b>C</b>	<b>Improper Normal Priors</b>	241
C.1	Definitions	241
C.2	An Intuitive Special Case	242
C.3	The General Case	243
C.4	Drawing Random Samples	246
<b>D</b>	<b>Discretization of the Derivative Operator</b>	247
<b>E</b>	<b>Smoothness over Graphs</b>	249
	<i>Bibliography</i>	251
	<i>Index</i>	259



# Figures

1.1	Distribution of Number of Observations	5
1.2	Leading Causes of Deaths Worldwide by Sex	6
1.3	World Age Profiles for 23 Causes of Death in Females	7
1.4	World Age Profiles for 20 Causes of Death in Males	8
2.1	All-Cause Mortality Age Profiles	22
2.2	Cause-Specific Mortality Age Profiles	23
2.3	Decomposing Age Profiles with Principal Components	30
2.4	Principal Components of Log-Mortality: 1st, 2nd, and 17th	32
2.5	First Four Time Series of $\gamma_{it}$	33
2.6	Data and Lee-Carter Forecasts by Age and Time, Part I	40
2.7	Data and Lee-Carter Forecasts by Age and Time, Part II	41
4.1	Age Profile Samples from a Simple Smoothness Prior	73
5.1	The Sin Function and Its 2nd Derivative, for Different Frequencies	82
5.2	Age Profile Samples from Smoothness Priors with Added Arbitrary Elements of the Null Space	88
5.3	Age Profile Samples from Smoothness Priors with Varying Degrees of Smoothness	89
5.4	Age Profile Samples from Smoothness Priors with Varying Degrees of Smoothness and Nonzero Mean	91
6.1	Age Profile Samples from “Mixed” Smoothness Priors	96
6.2	Samples from the Prior over Age Groups	99
6.3	Summary Measures of the Prior as a Function of the Standard Deviation of the Prior	103
6.4	Samples from Age Group Prior with Different Measure and Zero Mean	105
6.5	Smoothed Age Profiles of Respiratory Infectious Disease in Sri Lankan Males, Prior with Zero Mean	106
6.6	Smoothed Age Profiles of Respiratory Infectious Disease in Sri Lankan Males, Prior with Nonzero Mean	107
6.7	Samples from Age Group Prior with Different Measure and Nonzero Mean	108
6.8	Log-Normal Approximation to the Poisson	114
6.9	Estimation Error for $\lambda$ When $\lambda$ Is Small, with Large $\lambda$ Approximation	117
6.10	Estimation Error for $\lambda$ When $\lambda$ Is Small, with No Large $\lambda$ Approximation	117
6.11	Approximating the Variance of the Logarithm of a Poisson Variable	118
6.12	The Log-Normal Variance Approximation for Cardiovascular Disease in Men	121
6.13	The Log-Normal Variance Approximation for Breast Cancer	122
7.1	Scatterplots of Summary Measures by Prior Parameters	140
7.2	Result of the Empirical Bayes-like Procedure for Setting Summary Measure Target Values	143
8.1	The Effects of Changes in ICD Codes	154
8.2	Modeling the Effects of Changes in ICD Codes	155
8.3	The Null Space for Models of Changes in ICD Codes	156
11.1	Respiratory Infections in Males, Belize: Lee-Carter and Bayes: Advantages of Smoothing over Age Groups	179
11.2	Respiratory Infections in Males, Bulgaria: Lee-Carter and Bayesian Method: When Smoothing over Age Groups Is Insufficient	182
11.3	Respiratory Infections in Males, Bulgaria: Smoothing over Age, and Age Trends	183
11.4	Least-Squares Forecast of Lung Cancer Log-Mortality in Males: Peru	184

**xii • Figures**

11.5	Least-Squares Forecast of Lung Cancer Log-Mortality in Males: Ukraine and Trinidad and Tobago	185
11.6	Scatterplot of Summary Measures against Prior Parameters	188
11.7	Forecasts of Lung Cancer Log-Mortality in Males for Peru Obtained Using the Bayesian Method	189
11.8	Forecasts of Lung Cancer Log-Mortality in Males for Ukraine and Trinidad and Tobago Obtained Using the Bayesian Method	190
11.9	Forecasts of Breast Cancer Log-Mortality in Females for Croatia Obtained using Least-Squares and the Bayesian method	191
11.10	Forecasts of Breast Cancer Log-Mortality in Females for Four Countries	192
11.11	Least-Squares Forecasts of Log-Mortality by Transportation Accidents for Argentina and Chile	193
11.12	Bayesian Forecasts of Log-Mortality by Transportation Accidents for Argentina, with No Smoothing over Countries	194
11.13	Bayesian Forecasts of Log-Mortality by Transportation Accidents for Argentina, with Smoothing over Countries	194
12.1	Percentage Improvement of the MAP Method over the Lee-Carter Method for Lung Disease in Males	201
12.2	Percentage Improvement of the MAP Method over the Lee-Carter Method for Lung Disease in Males	202
12.3	Percentage Improvement of the MAP Method over the Lee-Carter Method for Breast Cancer in Females	206

## Tables

7.1	Summary Measures and Parameter Values: Using Four Target Summary Measures	141
7.2	Summary Measures and Parameter Values: Using Three Target Summary Measures	142
12.1	Percentage Improvement of Four Methods over the Lee-Carter Method for All Causes in Males.	199
12.2	Percentage Improvement over Lee-Carter, as a Function of $\sigma_{time}$ .	203
12.3	Percentage Improvement of Four Methods over the Lee-Carter Method for Lung Disease in Males.	204
12.4	Percentage Improvement of Four Methods over the Lee-Carter Method for Breast Cancer in Females	207
12.5	Percentage Improvement of Four Methods over the Lee-Carter Method for Transportation Accidents in Males: Most OECD Countries	208
12.6	Percentage Improvement of Four Methods over the Lee-Carter Method for Cardiovascular Disease in Males	209



## Preface

We introduce a framework for forecasting age-sex-country-cause-specific mortality rates that can incorporate more information, and thus has the potential to forecast better, than existing approaches. Mortality forecasts are used in a wide variety of academic fields and for global and national health policy making, medical and pharmaceutical research, and social security and retirement planning.

As it turns out, the tools we developed in pursuit of this goal also have broader statistical implications, in addition to their use for forecasting mortality or other variables with similar statistical properties. First, our methods make it possible to include different explanatory variables in a time-series regression for each crosssection, while still borrowing strength from one regression to improve the estimation of all. Second, we show that many existing Bayesian (hierarchical and spatial) models with explanatory variables use prior densities that incorrectly formalize prior knowledge. Many demographers and public health researchers have fortuitously avoided this problem so prevalent in other fields by using prior knowledge only as an ex post check on empirical results, but this approach excludes considerable information from their models. We show how to incorporate this demographic knowledge into a model in a statistically appropriate way that also turns out to have the advantage of requiring many fewer adjustable parameters than classic Bayesian approaches. Finally, we develop a set of tools useful for developing models with Bayesian priors in the presence of partial prior ignorance. This approach also provides many of the attractive features claimed by the empirical Bayes approach but does so fully within the standard Bayesian theory of inference.

### Software and Data

Accompanying this book is a free and open source software package that implements all our suggestions (see <http://GKing.Harvard.edu/yourcast/> for a copy). The software is entitled “YourCast: Time Series Cross-Sectional Forecasting with Your Assumptions” to emphasize a key intended contribution of our approach: that the assumptions made by the statistical model you run are governed entirely by your choices and your assumptions, and the sophistication of those assumptions and the degree to which they match empirical reality are limited primarily by what you may know or are willing to assume rather than by arbitrary choices hidden behind or hard-coded into a complicated mathematical model. Although some of the tools we introduce require technical sophistication to implement, the ideas are conceptually straightforward. As such, the software and methods should be usable even by those who decide not to digest all of our detailed mathematical arguments.

YourCast is distributed as an R package and is part of the R Project for Statistical Computing (R Development Core Team, 2007). Included in that package are demonstrations

**xvi • Preface**

that include all the data and code necessary to replicate all the empirical analyses in this book (King, 1995). The complete data set from which examples are drawn is available in Giroso and King (2006).

## **Background**

The methods developed in this book rely on fields of statistics and mathematics, at least some of which are likely to be unfamiliar to many interested in mortality forecasting. Yet, given the highly important public policy issues at stake, the advantage to scholars and citizens of any forecasting improvement, even when achieved via unfamiliar mathematical techniques, should, in our view, outweigh higher costs to researchers in learning the methods. We have thus not shied away from introducing new methods but have tried to reduce the associated costs to researchers in a variety of ways. Most importantly, we explain our methodology in a way that should make all of our results accessible to those who are familiar only with linear regression analysis and a course in Bayesian inference. We also include in appendix A a detailed glossary of notation. In addition, because different aspects of the necessary mathematical background are likely unfamiliar to different audiences, as they were even to us when we started, we offer an extensive mathematical refresher in appendix B that should be relatively complete as is.

Although we have attempted to keep the book as readable as possible, we have also included, for more mathematically sophisticated readers, all necessary proofs and evidence so that the work would be relatively self-contained.

## **Publications**

In some of the fields with which the content of this book intersects, almost all new results appear first in articles. Book presses are left to print texts that only summarize prior research. In this project, we resisted the temptation to send preliminary or partial results to scholarly journals because we felt the whole of our book would be greater than the sum of the parts, sliced into smaller articles, and because our goal was to produce a relatively complete and usable forecasting method in practice. Thus, although we have presented preliminary results in talks over the years, and shared earlier versions of this manuscript, this book is the first complete account of our approach.



## Acknowledgments

We are especially grateful for the contributions of Kevin Quinn and Greg Wawro. Although they should not be held responsible for our errors or arguments, we could not have written this book without what we learned from them. Wawro and Quinn were postdoctoral fellows at what is now the Institute for Quantitative Social Science (in the 1998–1999 and 1999–2000 academic years, respectively), and each contributed a great deal. Greg Wawro replicated Murray and Lopez’s (1996) forecasts, uncovered some surprising errors (traced to a remarkably basic error in the SAS Statistics package), and managed to improve substantially the quality of the World Health Organization’s mortality database. He also demonstrated the failure of a whole range of classic econometric tests and techniques when applied to our data: paradoxically, we found that standard tests frequently confirmed the presence of problems such as cointegration, nonstationarity, parameter heterogeneity, nonrandom missingness, dynamic processes, and omitted variable bias, but following the textbook econometric correction for each using a variety of off-the-shelf approaches degraded or only slightly improved the forecasts. We now accept new model features only if they demonstrate their worth by improving out-of-sample forecasting performance, specification tests for the various econometric problems being largely beside the point.

Kevin Quinn had the vision to see the connection between what we were trying to accomplish and the literature on Markov random fields. He worked out how to extend the Markov random field approach, designed for modeling physical space, to processes that varied over conceptual space (such as age). He also designed clever ways to implement these ideas via Gibbs sampling in Gauss code. These contributions were invaluable.

We also thank Chris Murray, who originally challenged us to develop a method of forecasting mortality that outperformed existing approaches—including his own—and his offices at the World Health Organization supplied us with data and research support. Chris, along with David Evans, Majid Ezzati, Emmanuela Gakidou, Alan Lopez, Colin Mathers, and Josh Salomon, taught us a great deal about mortality data and patterns. They were all especially helpful during the sessions we had pouring over thousands of forecasts from successive versions of our model. To say they were good sports every time we showed up in Geneva lugging two linear feet of printouts in a tiny font is no small understatement.

This project is also an outgrowth of a National Institute of Aging grant on the “Global Burden of Disease in Aging Populations” (P01 AG17625-01). The frequent meetings of this distinguished group of scholars have provided invaluable feedback on our work and, through site visits around the world, have greatly informed our analyses. Our thanks to the other project leaders on this grant, David Cutler, Ken Hill, Alan Lopez, and Chris Murray, to Richard Suzman at NIA, and to the many other participants in our meetings.

Our thanks to Emmanuela Gakidou, James Honaker, Catherine Michaud, Ken Scheve, Ajay Tandon, and Neils Tomijiman for logistical help with and insightful comments on our work. Thanks also to Sam Abrams, Chris Adolph, Marcus Augustine, Anders Corr, Alexis Diamond, Suzanne Globetti, Josh Goldstein, Mie Inoue, Ethan Katz, Ryan Moore,

**xviii • Acknowledgments**

Claudia Pedroza, Nirmala Ravishankar, Heather Stoll, and Liz Stuart for superb research assistance, and Anita Goldpergel and Elena Villalon for expert programming assistance. Our appreciation goes to Gadi Geiger for his mechanical engineering contributions. For helpful discussions on previous drafts or talks, we thank Marc Alexander, Barbara Anderson, Jim Berger, Heather Booth, Brad Carlin, Majid Ezzati, Andrew Gelman, Ed George, Ken Hill, John Maindonald, Xiao-Li Meng, Jamie Robins, Don Rubin, Samir Soneji, Herbert Smith, Len Smith, Dalene Stangl, Liz Stuart, Leonie Tickle, John Wilmoth, and Chris Zorn. Our special thanks to Ron Lee and Nan Li for their generous help with the demographic literature and insights into their models and perspectives, and Sander Greenland for patiently teaching us about the epidemiology literature. Micah Altman, Shawn Bunn, Matt Cox, and William Wei at the Harvard-MIT Data Center provided a series of first rate computational environments that always matched our quickly evolving needs. Thanks also to Jaronica Fuller for consistently cheering up everyone within miles of our offices, and Marie Cole, Beverly MacMillen, and Kim Schader for organizing the team we assembled and making difficult administrative issues transparent. For research support, in addition to the National Institutes of Aging, we are grateful to the National Science Foundation (SES-0112072, IIS-9874747) and the World Health Organization and, at Harvard, the Weatherhead Initiative and the Institute for Quantitative Social Science.

# Demographic Forecasting



