Demographic Forecasting

Gary King
Harvard University

Joint work with Federico Girosi (RAND)
with contributions from Kevin Quinn and Gregory Wawro
What this Talk is About

- Mortality forecasts, which are studied in:
  - demography & sociology
  - public health & biostatistics
  - economics & social security and retirement planning
  - actuarial science & insurance companies
  - medical research & pharmaceutical companies
  - political science & public policy

- A better forecasting method
- A better *farcasting* method
- Other results we needed to achieve this original goal
Other Results (Needed to Develop Improved Forecasts)

A New Class of Statistical Models

- Output: same as linear regression
- Estimates a set of linear regressions together
- Allows different covariates in each regression
- We demonstrate that most hierarchical and spatial Bayesian models with covariates misrepresent prior information
- Better Bayesian priors
- Forecasts and farcasts based on much more information
Multidimensional Data Structures: 24 causes of death, 17 age groups, 2 sexes, 191 countries, all for 50 annual observations.

One time series analysis for each of 155,856 cross-sections: with 1 minute to analyze each, one run takes 108 days.

Every decision must be automated, systematized, and formalized: the same goal as including qualitative information in the model.

Explanatory variables:
- Available in many countries: tobacco consumption, GDP, human capital, trends, fat consumption, total fertility rates, etc.
- Numerous variables specific to a cause, age group, sex, and country

Most time series are very short. A majority of countries have only a few isolated annual observations; only 54 countries have at least 20 observations; Africa, AIDS, & Malaria are real problems.
Existing Method 1: Parameterize the Age Profile

- Gompertz (1825): log-mortality is linear in age after age 20
  - reduces 17 age-specific mortality rates to 2 parameters
  - forecast only these 2 parameters
  - Reduces variance, constrains forecasts

- Dozens of more general functional forms proposed
- But does it fit anything else?
Mortality Age Profile: The Same Pattern?

Cardiovascular Disease (m)

Age

ln(mortality)

France
USA
Brazil

Demographic Forecasting
Mortality Age Profile: The Same Pattern?

Breast Cancer (f)

Japan
Venezuela
New Zealand

Age
ln(mortality)
Mortality Age Profile: The Same Pattern?

Other Infectious Diseases (f)

- Thailand
- Sri Lanka
- Barbados
- Italy

In(mortality) vs. Age
Mortality Age Profile: The Same Pattern?

### Diagram

- **Suicide (m)**
  - Hungary
  - Canada
  - Colombia
  - Sri Lanka

- **Axes**:
  - **Y-axis**: In(mortality)
  - **X-axis**: Age

- **Data Points**: Sri Lanka, Colombia, Canada, Hungary
Parameterizing Age Profiles Does Not Work

- No mathematical form fits all or even most age profiles
- Out-of-sample age profiles often unrealistic
- The key empirical patterns are qualitative:
  - Adjacent age groups have similar mortality rates
  - Age profiles are more variable for younger ages
  - We don’t know much about levels or exact shapes
- Ignores covariate information
Random walk with drift; Lee-Carter; least squares on linear trend

Pros: simple, fast, works well in appropriate data

Cons: omits covariates; forecasts fan out; age profile becomes less smooth

Does it fit elsewhere?
The same pattern?
Random Walk with Drift $\approx$ Lee-Carter $\approx$ Least Squares

Data and Forecasts

Suicide (m) USA

Time

Age

1950 2060

15 20 25 30 35 40 45 50 55 60 65 70 75 80

-10.5 -10.0 -9.5 -9.0 -8.5 -8.0 -7.5 -7.0

Suicide (m) USA

Data and Forecasts
The same pattern?
Random Walk with Drift $\approx$ Lee-Carter $\approx$ Least Squares

Transportation Accidents (m) Portugal

Data and Forecasts

Time

Age

Transportation Accidents (m) Portugal

Data and Forecasts

1955 2060
Deterministic Projections Do Not Work

- Linearity does not fit most time series data
- Out-of-sample age profiles become unrealistic over time
Model mortality over countries ($c$) and ages ($a$) as:

$$m_{cat} = Z_{ca,t-\ell}\beta_{ca} + \epsilon_{cat}, \quad t = 1, \ldots, T$$

- $Z_{ca,t-\ell}$: covariates lagged $\ell$ years.
- $\beta_{ca}$: coefficients to be estimated
- Equation by equation estimation: huge variances
- Pool over countries: $\beta_{ca} \Rightarrow \beta_a$
  - Small variance (due to large $n$)
  - large biases (due to restrictive pooling over countries),
  - considerable information lost (due to no pooling over ages)
  - same covariates required in all cross-sections
Partial Pooling via a Bayesian Hierarchical Approach

- Likelihood for equation-by-equation least squares:

\[ P(m \mid \beta_i, \sigma_i) = \prod_t \mathcal{N}(m_{it} \mid Z_{it}\beta_i, \sigma_i^2) \]

- Add priors and form a posterior

\[ P(\beta, \sigma, \theta \mid m) \propto P(m \mid \beta, \sigma) \times P(\beta \mid \theta) \times P(\theta)P(\sigma) = (\text{Likelihood}) \times (\text{Key Prior}) \times (\text{Other priors}) \]

- Calculate point estimate for \( \beta \) (for \( \hat{y} \)) as the mean posterior:

\[ \beta^{\text{Bayes}} \equiv \int \beta P(\beta, \sigma, \theta \mid m) \, d\beta d\theta d\sigma \]

- The hard part: specifying the prior for \( \beta \) and, as always, \( Z \)

- The easy part: easy-to-use software to implement everything we discuss today.
The (Problematic) Classical Bayesian Approach

Assumption: similarities among cross-sections imply similarities among coefficients ($\beta$'s).

Requirements: Comparing $\beta_i$ and $\beta_j$
- Similarity: $s_{ij}$
- Distance: $(\beta_i - \beta_j)' \Phi (\beta_i - \beta_j) \equiv \| \beta_i - \beta_j \|_\Phi^2$

Natural choice for the prior:

$$
\mathcal{P}(\beta \mid \Phi) \propto \exp \left( - \frac{1}{2} \sum_{ij} s_{ij} \| \beta_i - \beta_j \|_\Phi^2 \right)
$$
The (Problematic) Classical Bayesian Approach

- Requires the **same** covariates, **with the same meaning**, in every cross-section.

- Prior knowledge about $\beta$ exists for causal effects, not for control variables, or forecasting

- Everything depends on $\Phi$, the normalization factor:
  - $\Phi$ can’t be estimated, and must be set.
  - An **uninformative prior** for it would make Bayes irrelevant,
  - An **informative prior** can’t be used since we don’t have information
  - Common practice: make some **wild guesses**.

- The classical approach can be harmful: Making $\beta_i$ more smooth may make $\mu$ less smooth ($\mu = Z\beta$):

- Extensive trial-and-error runs: no useful parameter values.
Our Alternative Strategy: Priors on $\mu$

Three steps:

1. Specify a prior for $\mu$:

$$P(\mu | \theta) \propto \exp \left( -\frac{1}{2} H[\mu, \theta] \right), \text{ e.g., } H[\mu, \theta] \equiv \frac{\theta}{T} \sum_{t=1}^{T} \sum_{a=1}^{A-1} (\mu_{at} - \mu_{a+1,t})^2$$

- To do Bayes, we need a prior on $\beta$
- Problem: How to translate a prior on $\mu$ into a prior on $\beta$ (a few-to-many transformation)?

2. Constrain the prior on $\mu$ to the subspace spanned by the covariates:

$$\mu = Z\beta$$

3. In the subspace, we can invert $\mu = Z\beta$ as $\beta = (Z'Z)^{-1}Z'\mu$, giving:

$$P(\beta | \theta) \propto \exp \left( -\frac{1}{2} H[\mu, \theta] \right) = \exp \left( -\frac{1}{2} H[Z\beta, \theta] \right)$$

the same prior on $\mu$, expressed as a function of $\beta$ (with constant Jacobian).
Say that again?

In other words

Any prior information about $\mu$ (the expected value of the dependent variable) is “translated” into information about the coefficients $\beta$ via

$$\mu_{\text{cat}} = Z_{\text{cat}}\beta_{\text{ca}}$$

A Simple Analogy

- Suppose $\delta = \beta_1 - \beta_2$ and $P(\delta) = N(\delta|0, \sigma^2)$
- What is $P(\beta_1, \beta_2)$?
- It's a singular bivariate Normal
- It's defined over $\beta_1, \beta_2$ and constant in all directions but $(\beta_1 - \beta_2)$.
- We start with one-dimensional $P(\mu_{\text{cat}})$, and treat it as the multidimensional $P(\beta_{\text{ca}})$, constant in all directions but $Z_{\text{cat}}\beta_{\text{ca}}$. 
Advantages of the resulting prior over $\beta$, created from prior over $\mu$

- Fully Bayesian: The same theory of inference applies
- $\mu_i$ and $\mu_j$ can always be compared, even with different covariates.
- The normalization matrix $\Phi$ is unnecessary (normalization is performed by $Z$, which is known)
Prior knowledge: log-mortality age profile are smooth variations of a "typical" age profile $\bar{\mu}(a)$:

$$H[\mu, \theta] \equiv \frac{\theta}{AT} \int_0^T dt \int_0^A da \left( \frac{dn}{da} \left[ \mu(a, t) - \bar{\mu}(a) \right] \right)^2$$

Discretize age and time:

$$P(\mu \mid \theta) \propto \exp \left( -\frac{1}{2} \theta \sum_{aa' t} (\mu_{at} - \bar{\mu}_a)' W^n_{aa'} (\mu_{a't} - \bar{\mu}_{a'}) \right)$$

where $W^n$ is a matrix uniquely determined by $n$ and $\theta$
From a prior on $\mu$ to a prior on $\beta$

Add the specification $\mu_{at} = \bar{\mu}_a + Z_{at} \beta_a$:

$$P(\beta \mid \theta) = \exp \left( -\frac{\theta}{T} \sum_{aa'} W^n_{a a'} (Z_{at} \beta_a)(Z_{a't} \beta_a') \right)$$

$$= \exp \left( -\theta \sum_{aa'} W^n_{a a'} \beta_a' C_{aa'} \beta_a' \right)$$

where we have defined:

$$C_{aa'} \equiv \frac{1}{T} Z'_a Z_{a'} \quad Z_a \text{ is a } T \times d_a \text{ data matrix for age group } a$$
The Prior on the Coefficients $\beta$

$$
\mathcal{P}(\beta | \theta) \propto \exp \left( -\theta \sum_{aa'} W_{aa'}^{n} \beta'_a C_{aa'} \beta'_a \right)
$$

- The prior is normal (and improper)
- $n$: determines by the prior through the “interaction” matrix $W^n$.
- $\theta$: the “strength” of the prior
- Different age groups can have different covariates: the matrices $C_{aa'}$ are rectangular ($d_a \times d_{a'}$).
All Causes \( (m), n = 1 \)
Samples From Age Prior

All Causes \((m), n = 2\)

![Log-mortality vs Age Graph](image-url)
Samples From Age Prior

All Causes (m), n = 3

![Graph showing log-mortality against age for all causes.](Image)
All Causes (m), n = 4

Age

Log-mortality

Age

0 20 40 60 80

−8 −6 −4 −2

Demographic Forecasting

28 / 96
Samples From Age Prior

All Causes (m), n = 1

All Causes (m), n = 2

All Causes (m), n = 3

All Causes (m), n = 4
Formalizing (Prior) Indifference

equal color = equal probability

Level indifference

Level and slope indifference
Smoothness Parameter

- The prior:

\[ \mathcal{P}(\beta \mid \theta) \propto \exp \left( -\theta \sum_{aa'} W_{aa'}^{n} \beta'_a C_{aa'} \beta_{a'} \right) \]

- We figured out what \( n \) is
- but what is the smoothness parameter, \( \theta \)?
- \( \theta \) controls the prior standard deviation
Samples from Age Prior

All Causes (f), n = 2

Log-mortality vs Age

Age

0 20 40 60 80

-10 -8 -6 -4 -2 0 2

All Causes (f), n = 2
Samples from Age Prior

All Causes (f), n = 2

Log–mortality

Age
Samples from Age Prior

All Causes (f), n = 2

Log–mortality vs Age

Demographic Forecasting
Samples from Age Prior

All Causes (f), n = 2

Age

Log-mortality

0 20 40 60 80

-10 -8 -6 -4 -2 0 2

Demographic Forecasting
Samples from Age Prior

All Causes (f), n = 2

Log-mortality vs Age for all causes.
Samples from Age Prior

All Causes \( (f) \), \( n = 2 \)

![Graph showing samples from Age Prior with Age on the x-axis and Log-mortality on the y-axis.](image)
Samples from Age Prior

All Causes (f), n = 2

![Graph showing log-mortality against age for all causes.](image-url)
Samples from Age Prior

All Causes (f), n = 2

Log-mortality vs Age
Samples from Age Prior

All Causes (f), n = 2
Samples from Age Prior

All Causes (f), n = 2
Samples from Age Prior

All Causes (f), n = 2
Samples from Age Prior

All Causes (f), n = 2

Log-mortality vs Age
Samples from Age Prior

All Causes (f), n = 2
Generalizations

- The above tools: smooth over a (possibly discretized) continuous variable — age or age groups.
- We can also smooth over time (also a discretized continuous variable).
- Can smooth when cross-sectional unit $i$ is a label, such as country.
- Can smooth simultaneously over different types of variables (age, country, and time).
- We can smooth interactions:
  - Smoothing *trends* over age groups.
  - Smoothing trends over age groups as they vary across countries, etc.
- The mathematical form for *all* these (separately or together) turns out to be the same:

$$P(\beta \mid \theta) \propto \exp \left( -\frac{\theta}{2} \sum_{ij} W_{ij} \beta_i \mathbf{C}_{ij} \beta_j \right), \quad \mathbf{C}_{aa'} \equiv \frac{1}{T} \mathbf{Z}_a \mathbf{Z}_{a'}$$
Mortality from Respiratory Infections, Males

Least Squares

Data and Forecasts

1970 2030

Age

(m) Belize

Data and Forecasts

Age
Mortality from Respiratory Infections, males, $\sigma = 2.00$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 1.51$

Smoothing over Age Groups

Data and Forecasts

1970 2030

Demographic Forecasting
Mortality from Respiratory Infections, males, $\sigma = 1.15$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.87$

Smoothing over Age Groups

Data and Forecasts

(m) Belize

Age

[Age distribution graph with data and forecasts for Belize from 1970 to 2030]
Mortality from Respiratory Infections, males, $\sigma = 0.66$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.50$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.38$

Smoothing over Age Groups

![Graph showing mortality data and forecasts for Belize](image)
Mortality from Respiratory Infections, males, $\sigma = 0.28$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.21$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.16$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.12$

Smoothing over Age Groups

Data and Forecasts

(m) Belize

Age

1970 2030

Demographic Forecasting
Mortality from Respiratory Infections, males, $\sigma = 0.09$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.07$

Smoothing over Age Groups

![Graph showing data and forecasts for mortality from respiratory infections in Belize. The x-axis represents age, and the y-axis represents data and forecasts. The graph includes data from 1970 and forecasts into 2030.]
Mortality from Respiratory Infections, males, \( \sigma = 0.05 \)

Smoothing over Age Groups

![Graph showing data and forecasts for mortality from respiratory infections in Belize, with age on the x-axis and data for different years on the y-axis, including 1970 and 2030.](image-url)
Mortality from Respiratory Infections, males, $\sigma = 0.04$

Smoothing over Age Groups

Data and Forecasts

1970 2030

(m) Belize

Age

Data and Forecasts

0 20 40 60 80

−12 −10 −8 −6 −4

(m) Belize

Age

Data and Forecasts

0 20 40 60 80

−12 −10 −8 −6 −4
Mortality from Respiratory Infections, males, $\sigma = 0.03$

Smoothing over Age Groups

Demographic Forecasting
Mortality from Respiratory Infections, males, $\sigma = 0.02$

Smoothing over Age Groups

Data and Forecasts

Belize

1970 2030

(m) Belize

Age

Data and Forecasts

0 20 40 60 80

−12 −10 −8 −6 −4

Age

0 20 40 60 60 80

1970 2030
Mortality from Respiratory Infections, males, $\sigma = 0.01$

Smoothing over Age Groups
Mortality from Respiratory Infections, males

Least Squares

Data and Forecasts

(m) Belize

Time


-12 -10 -8 -6 -4

Belize

0
5
10
15
20
25
30 35
40
45
50
55
60
65
70
75
80

Demographic Forecasting
Mortality from Respiratory Infections, males, $\sigma = 2.00$

Smoothing over Age Groups

![Graph showing mortality over time and age groups in Belize](image)
Mortality from Respiratory Infections, males, $\sigma = 1.51$

Smoothing over Age Groups

Data and Forecasts

(m) Belize

Time


(−12 −10 −8 −6 −4)
Mortality from Respiratory Infections, males, $\sigma = 1.15$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.87$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.66$

Smoothing over Age Groups

![Graph showing mortality data and forecasts over time for Belize.](image)
Mortality from Respiratory Infections, males, $\sigma = 0.50$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.38$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.28$

Smoothing over Age Groups

![Graph showing mortality data and forecasts for Belize from 1970 to 2030.](image)
Mortality from Respiratory Infections, males, $\sigma = 0.21$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.16$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.12$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.09$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.07$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.05$

Smoothing over Age Groups

Demographic Forecasting
Mortality from Respiratory Infections, males, $\sigma = 0.04$

Smoothing over Age Groups

![Diagram showing mortality data and forecasts for Belize from 1970 to 2030. The X-axis represents time from 1970 to 2030, and the Y-axis represents data and forecasts. The graph includes trend lines for different age groups.]
Mortality from Respiratory Infections, males, $\sigma = 0.03$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.02$

Smoothing over Age Groups
Mortality from Respiratory Infections, males, $\sigma = 0.01$

Smoothing over Age Groups
Smoothing Trends over Age Groups

Log-mortality in Belize males from respiratory infections

Least Squares

Smoothing Age Groups
Smoothing Trends over Age Groups and Time
Log-Mortality in Bulgarian males from respiratory infections

Least Squares

Smoothing Age and Time
Using Covariates (GDP, tobacco, trend, log trend)

Lung cancer in Korean Males

Least Squares

Smooth over age, time, age/time
Using Covariates (GDP, tobacco, trend, log trend)

Lung cancer in Males, Singapore

Least Squares

Smooth over age, time, age/time
What about ICD Changes?

Other Infectious Diseases: USA, age 0 (m)

Other Infectious Diseases: France, age 0 (m)

Other Infectious Diseases: Australia, age 0 (m)

Other Infectious Diseases: United Kingdom, age 0 (m)
Fixing ICD Changes

Other Infectious Diseases: USA, age 0 (m)

Other Infectious Diseases: France, age 0 (m)

Other Infectious Diseases: Australia, age 0 (m)

Other Infectious Diseases: United Kingdom, age 0 (m)
A book manuscript, YourCast software, etc.

http://GKing.Harvard.edu
With Country Smoothing

Transportation Accidents (males)
Sri Lanka
BAYES

Demographic Forecasting
Coverage of WHO data base (age specific, all causes)

- % of world countries
- % of world population

# Observations

% of world countries
% of world population
Prior Indifference

- These priors are “indifferent” to transformations:

\[ \mu(a, t) \sim \mu(a, t) + p(a, t) \]

- where \( p(a, t) \) is a polynomial in \( a \) (whose degree is the degree of the derivative in the prior)

- Prior information is about relative (not absolute) levels of log-mortality
## Preview of Results: Out-of-Sample Evaluation

**Mean Absolute Error in Males (over age and country)**

<table>
<thead>
<tr>
<th>Cause</th>
<th>% Improvement Over Best Previous</th>
<th>% Improvement to Best Conceivable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular</td>
<td>22</td>
<td>49</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>24</td>
<td>47</td>
</tr>
<tr>
<td>Transportation</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>Respiratory Chronic</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td>Other Infectious</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>Stomach Cancer</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>All-Cause</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>Suicide</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Respiratory Infectious</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

- Each row averages 6,800 forecast errors (17 age groups, 40 countries, and 10 out-of-sample years).
- % to best conceivable = % of the way our method takes us from the best existing to the best conceivable forecast.
- The new method out-performs with the same covariates, for most countries, causes, sexes, and age groups.
- Does considerably better with more informative covariates.
## Preview of Results: Out-of-Sample Evaluation

### Mean Absolute Error in Males (over age and country)

<table>
<thead>
<tr>
<th>Cause</th>
<th>Best Previous</th>
<th>Our Method</th>
<th>Best Conceivable</th>
<th>% Improvement Over Best Previous</th>
<th>% Improvement to Best Conceivable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular</td>
<td>0.34</td>
<td>0.27</td>
<td>0.19</td>
<td>22</td>
<td>49</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>0.36</td>
<td>0.27</td>
<td>0.17</td>
<td>24</td>
<td>47</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.37</td>
<td>0.31</td>
<td>0.18</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>Respiratory Chronic</td>
<td>0.45</td>
<td>0.39</td>
<td>0.26</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td>Other Infectious</td>
<td>0.55</td>
<td>0.48</td>
<td>0.32</td>
<td>12</td>
<td>30</td>
</tr>
<tr>
<td>Stomach Cancer</td>
<td>0.30</td>
<td>0.27</td>
<td>0.20</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>All-Cause</td>
<td>0.17</td>
<td>0.15</td>
<td>0.08</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>Suicide</td>
<td>0.31</td>
<td>0.29</td>
<td>0.18</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>Respiratory Infectious</td>
<td>0.49</td>
<td>0.47</td>
<td>0.28</td>
<td>3</td>
<td>7</td>
</tr>
</tbody>
</table>

- Each row averages 6,800 forecast errors (17 age groups, 40 countries, and 10 out-of-sample years).
- % to best conceivable = % of the way our method takes us from the best existing to the best conceivable forecast.
- The new method out-performs with the same covariates, for most countries, causes, sexes, and age groups.
- Does much better with better covariates.