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)

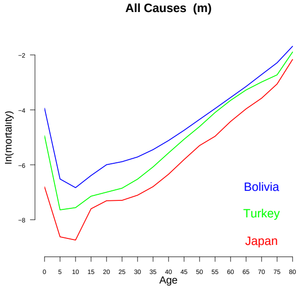
A New Class of Statistical Models

- 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

The Statistical Problem of Global Mortality Forecasting

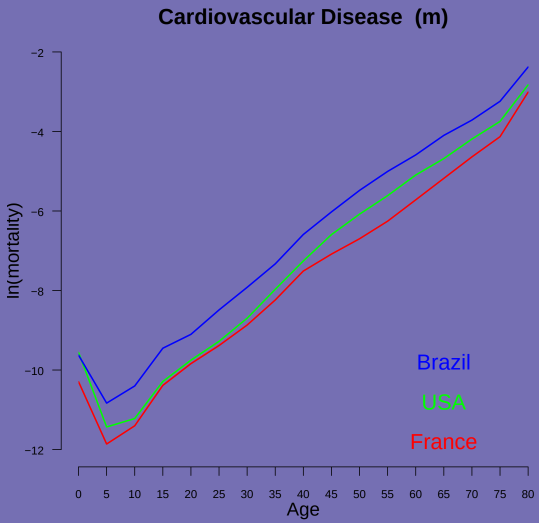
- Multidimensional Data Structures: 24 causes of death, 17 age groups, 2 sexes, 191 countries, all for 50 annual observations.
- One time series analysis for each of 155,856 cross-sections: with 1 minute to analyze each, one run takes 108 days
- Every decision must be automated, systematized, and formalized: 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

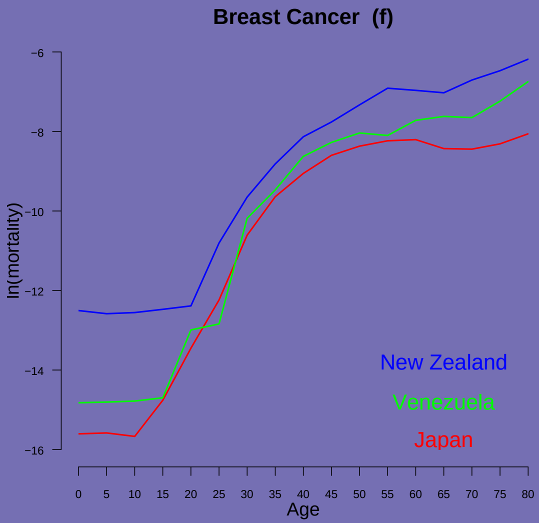


- Gompertz (1825): log-mortality is linear in age after age 20
  - reduces 17 age-specific mortality rates to 2 parameters
  - forecast only these 2 parameters
  - Reduces variance, constrains forecasts
- Dozens of more general functional forms proposed
- But does it fit anything else?

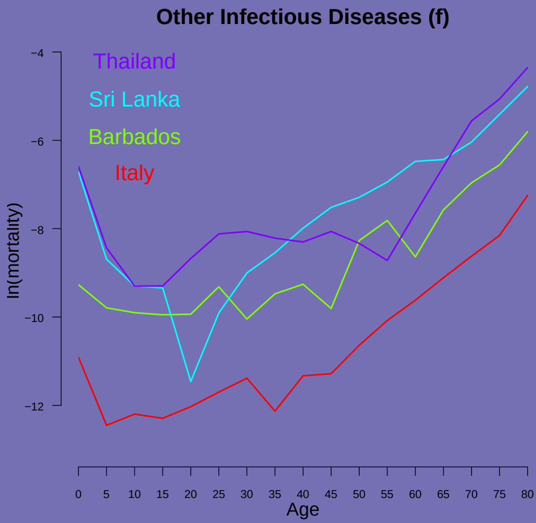
Mortality Age Profile: The Same Pattern?



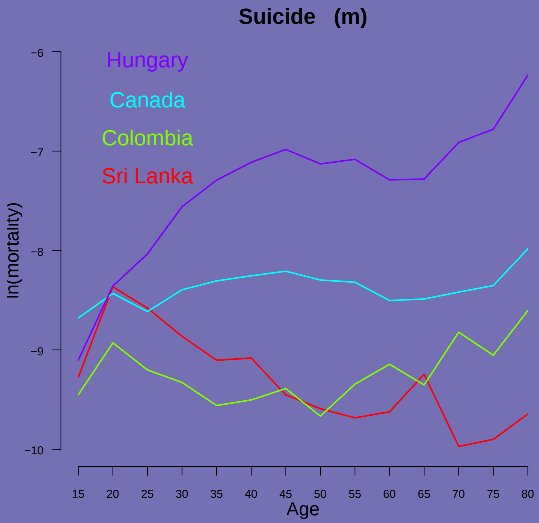
Mortality Age Profile: The Same Pattern?



Mortality Age Profile: The Same Pattern?



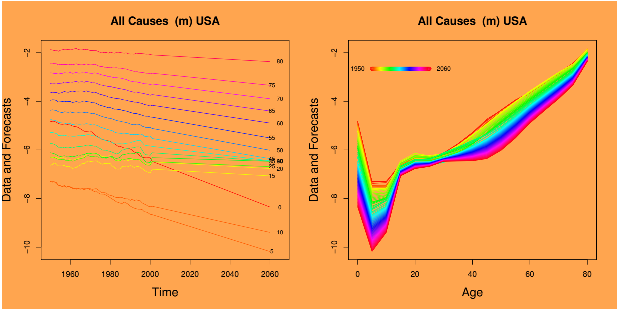
Mortality Age Profile: The Same Pattern?



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

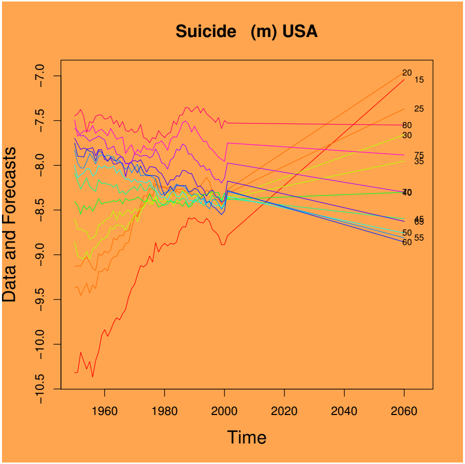
Existing Method 2: Deterministic Projections

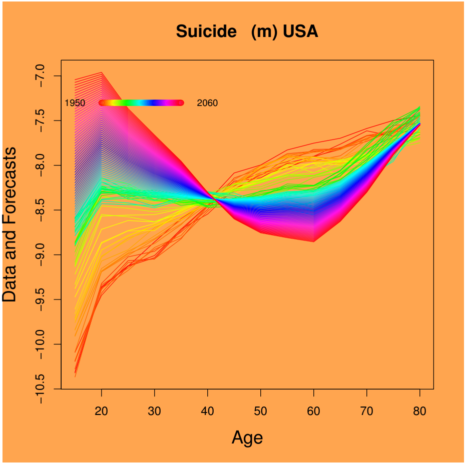


- Random walk with drift; Lee-Carter; least squares on linear trend
- Pros: simple, fast, works well in appropriate data
- Cons: omits covariates; forecasts fan out;
  age profile becomes less smooth
- Does it fit elsewhere?

The same pattern?

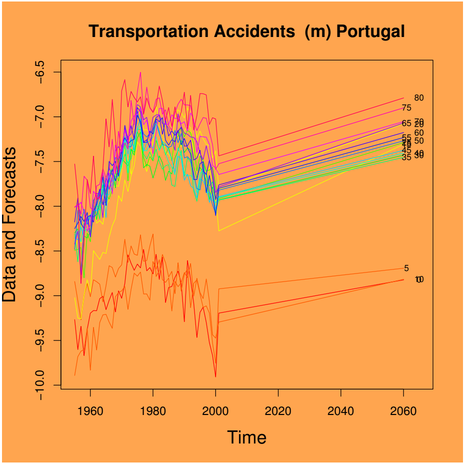
Random Walk with Drift ≈ Lee-Carter ≈ Least Squares

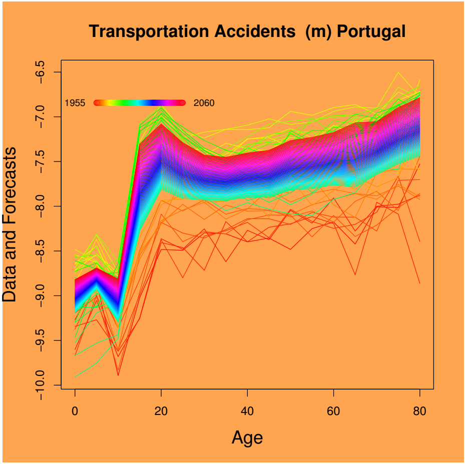




The same pattern?

Random Walk with Drift ≈ Lee-Carter ≈ Least Squares





Deterministic Projections Do Not Work

- Linearity does not fit most time series data
- Out-of-sample age profiles become unrealistic over time

Regression Approaches (Murray and Lopez, 1996)

- Model mortality over countries (c) and ages (a) as:

$$m_{cat} = \mathbf{Z}_{ca,t-\ell}\boldsymbol{\beta}_{ca} + \epsilon_{cat} \quad , \quad t = 1, \ldots, T$$

- $\mathbf{Z}_{ca,t-\ell}$: covariates lagged ℓ years.
- $\boldsymbol{\beta}_{ca}$: coefficients to be estimated
- Equation by equation estimation: huge variances
- Pool over countries: $\boldsymbol{\beta}_{ca} \Rightarrow \boldsymbol{\beta}_{a}$
  - 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:

$$\mathcal{P}(m \mid \boldsymbol{\beta}_i, \sigma_i) = \prod_t \mathcal{N}\left(m_{it} \mid \mathbf{Z}_{it}\boldsymbol{\beta}_i, \sigma_i^2\right)$$

- Add priors and form a posterior

$$\begin{aligned}\mathcal{P}(\boldsymbol{\beta}, \sigma, \theta \mid m) &\propto \mathcal{P}(m \mid \boldsymbol{\beta}, \sigma) \times \mathcal{P}(\boldsymbol{\beta} \mid \theta) \times \mathcal{P}(\theta)\mathcal{P}(\sigma) \\ &= (\text{Likelihood}) \times (\text{Key Prior}) \times (\text{Other priors})\end{aligned}$$

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

$$\boldsymbol{\beta}^{\text{Bayes}} \equiv \int \boldsymbol{\beta}\mathcal{P}(\boldsymbol{\beta}, \sigma, \theta \mid m)\, d\boldsymbol{\beta} d\theta d\sigma$$

- The hard part: specifying the prior for $\boldsymbol{\beta}$ 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 ($\boldsymbol{\beta}$'s).

Requirements: Comparing $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}_j$

- Similarity: s_{ij}
- Distance: $(\boldsymbol{\beta}_i - \boldsymbol{\beta}_j)'\Phi(\boldsymbol{\beta}_i - \boldsymbol{\beta}_j) \equiv \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2_\Phi$

Natural choice for the prior:

$$\mathcal{P}(\boldsymbol{\beta} \mid \Phi) \propto \exp\left(-\frac{1}{2}\sum_{ij} s_{ij}\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2_\Phi\right)$$

The (Problematic) Classical Bayesian Approach

- Requires the same covariates, with the same meaning, in every cross-section.
- Prior knowledge about β exists for causal effects, not for control variables, or forecasting
- Everything depends on Φ, the normalization factor:
    - Φ 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 β_i more smooth may make μ less smooth ($\mu = \mathbf{Z}\beta$):
- Extensive trial-and-error runs: no useful parameter values.

Our Alternative Strategy: Priors on μ

Three steps:

1. Specify a prior for μ:

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

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

2. Constrain the prior on μ to the subspace spanned by the covariates: $\mu = \mathbf{Z}\boldsymbol{\beta}$
3. In the subspace, we can invert $\mu = \mathbf{Z}\boldsymbol{\beta}$ as $\boldsymbol{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mu$, giving:

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

   the same prior on μ, expressed as a function of $\boldsymbol{\beta}$ (with constant Jacobian).

Say that again?

In other words

Any prior information about μ (the expected value of the dependent variable) is "translated" into information about the coefficients $\boldsymbol{\beta}$ via

$$\mu_{cat} = Z_{cat}\boldsymbol{\beta}_{ca}$$

A Simple Analogy

- Suppose $\delta = \beta_1 - \beta_2$ and $P(\delta) = N(\delta|0, \sigma^2)$
- What is $P(\beta_1, \beta_2)$?
- Its a singular bivariate Normal
- Its defined over β_1, β_2 and constant in all directions but $(\beta_1 - \beta_2)$.
- We start with one-dimensional $P(\mu_{cat})$, and treat it as the multidimensional $P(\boldsymbol{\beta}_{ca})$, constant in all directions but $Z_{cat}\boldsymbol{\beta}_{ca}$.

Advantages of the resulting prior over β, created from prior over μ

- Fully Bayesian: The same theory of inference applies
- μ_i and μ_j can always be compared, even with different covariates.
- The normalization matrix Φ is unnecessary (normalization is performed by $\mathbf{Z}$, which is known)

Basic Prior: Smoothness over Age Groups

- Prior knowledge: log-mortality age profile are smooth variations of a "typical" age profile $\bar{\mu}(a)$:

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

- Discretize age and time:

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

- where W^n is a matrix uniquely determined by n and θ

From a prior on μ to a prior on β

Add the specification $\mu_{at} = \bar{\mu}_a + \mathbf{Z}_{at}\boldsymbol{\beta}_a$:

$$\mathcal{P}(\boldsymbol{\beta} \mid \theta) = \exp\left(-\frac{\theta}{T}\sum_{aa't} W^n_{aa'}(\mathbf{Z}_{at}\boldsymbol{\beta}_a)(\mathbf{Z}_{a't}\boldsymbol{\beta}_{a'})\right)$$

$$= \exp\left(-\theta\sum_{aa'} W^n_{aa'}\boldsymbol{\beta}'_a\mathbf{C}_{aa'}\boldsymbol{\beta}_{a'}\right)$$

where we have defined:

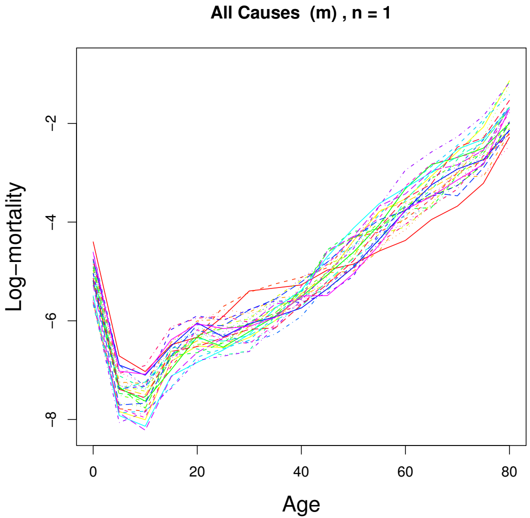
$\mathbf{C}_{aa'} \equiv \frac{1}{T}\mathbf{Z}'_a\mathbf{Z}_{a'}$ $\quad$ $\mathbf{Z}_a$ is a $T \times d_a$ data matrix for age group a

The Prior on the Coefficients $\boldsymbol{\beta}$

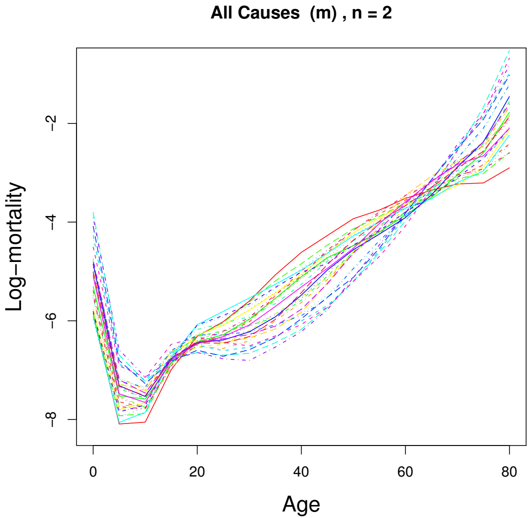
$$\mathcal{P}(\boldsymbol{\beta} \mid \theta) \propto \exp\left(-\theta \sum_{aa'} W^{n}_{aa'} \boldsymbol{\beta}'_a \mathbf{C}_{aa'} \boldsymbol{\beta}_{a'} \right)$$

- The prior is normal (and improper)
- n: determines by the prior through the "interaction" matrix W^n.
- θ: the "strength" of the prior
- Different age groups can have different covariates: the matrices $\mathbf{C}_{aa'}$ are rectangular $(d_a \times d_{a'})$.

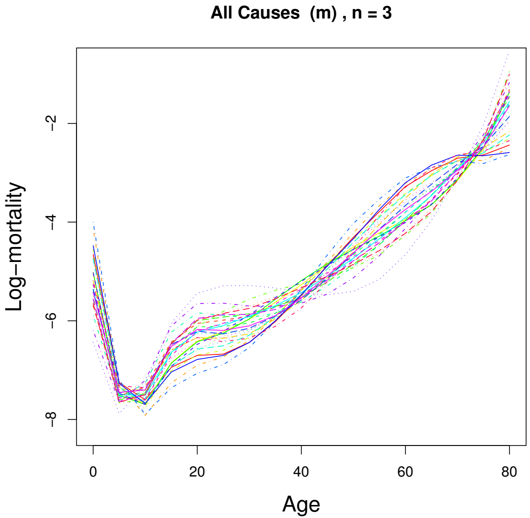
Samples From Age Prior



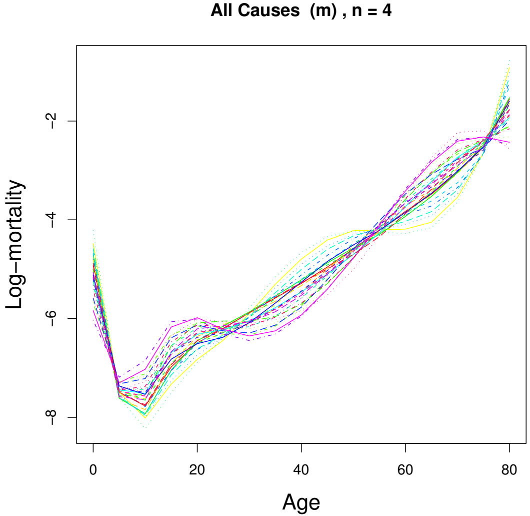
Samples From Age Prior



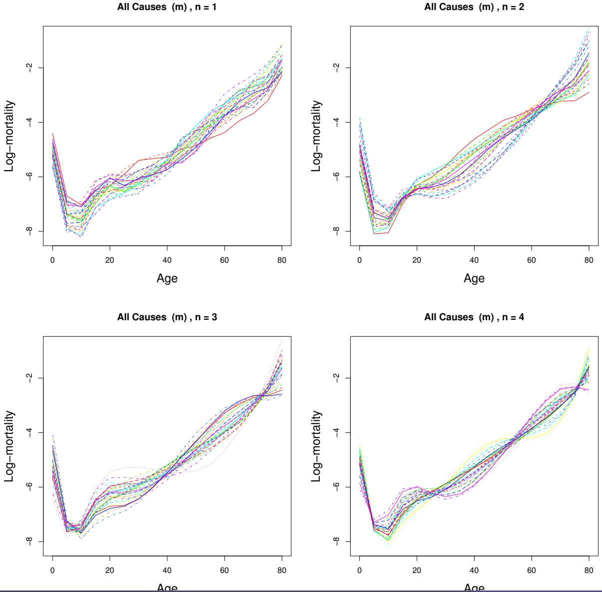
Samples From Age Prior



Samples From Age Prior



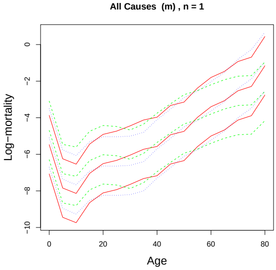
Samples From Age Prior



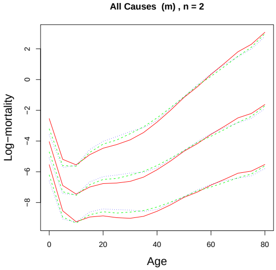
Formalizing (Prior) Indifference

equal color = equal probability

Level indifference



Level and slope indifference



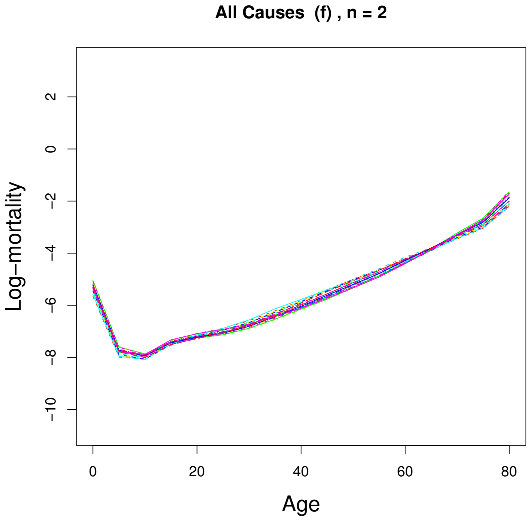
Smoothness Parameter

- The prior:

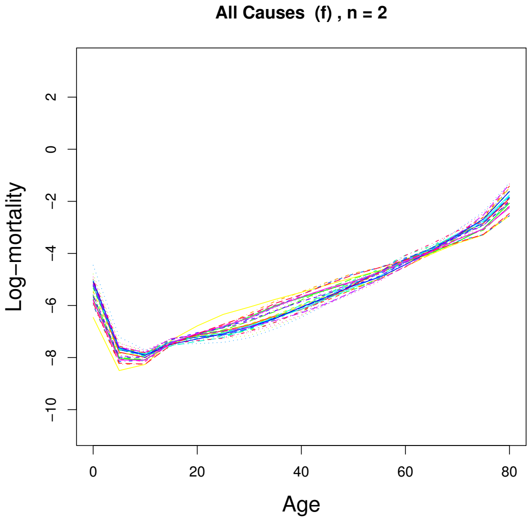
$$\mathcal{P}(\boldsymbol{\beta} \mid \theta) \propto \exp\left(-\theta \sum_{aa'} W^{n}_{aa'} \boldsymbol{\beta}'_{a} \mathbf{C}_{aa'} \boldsymbol{\beta}_{a'}\right)$$

- We figured out what n is
- but what is the smoothness parameter, θ?
- θ controls the prior standard deviation

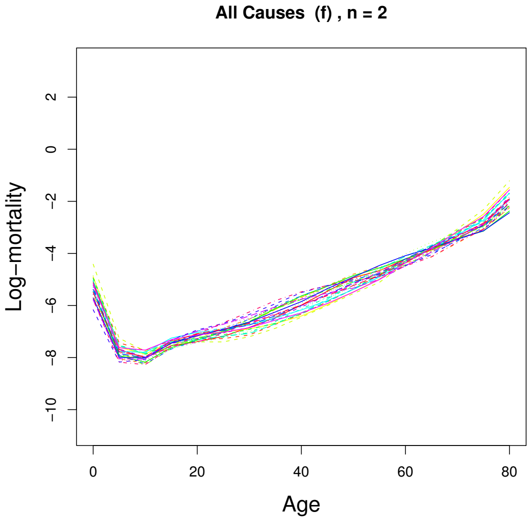
Samples from Age Prior



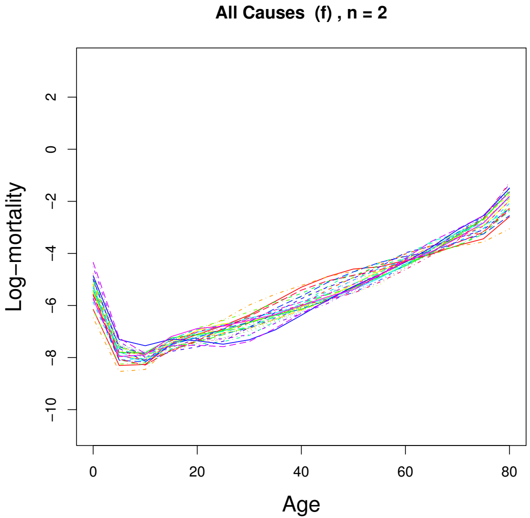
Samples from Age Prior



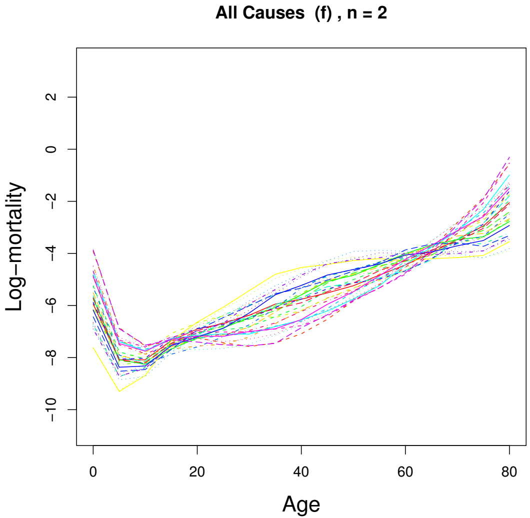
Samples from Age Prior



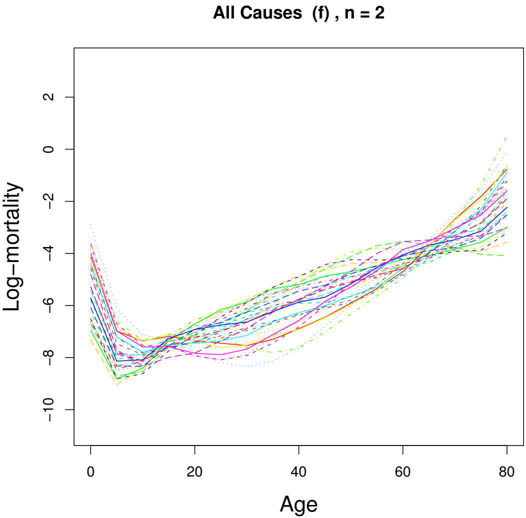
Samples from Age Prior



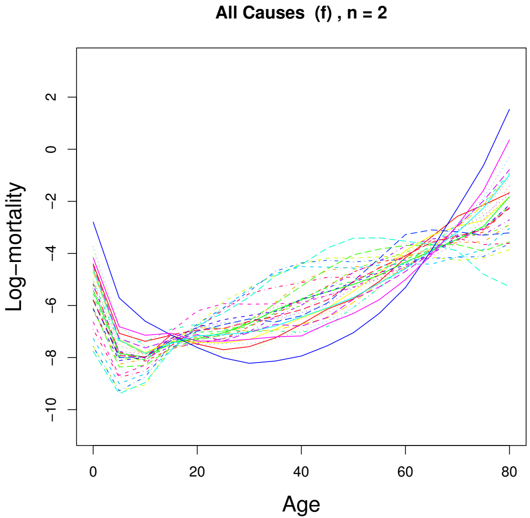
Samples from Age Prior



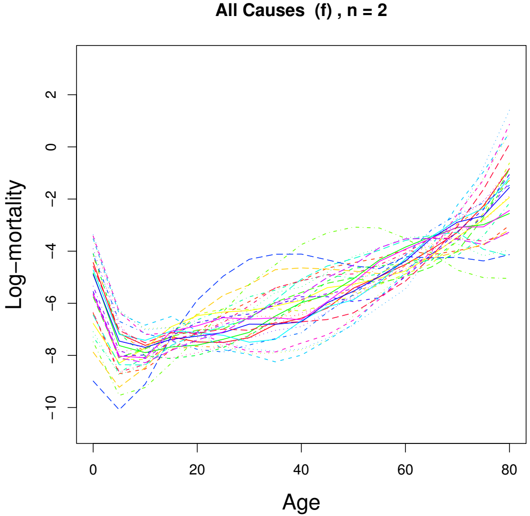
Samples from Age Prior



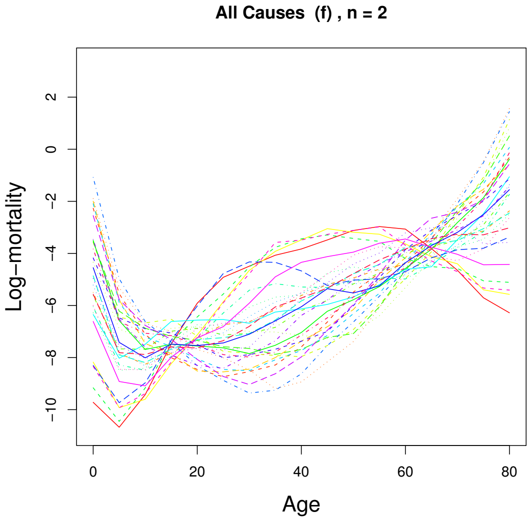
Samples from Age Prior



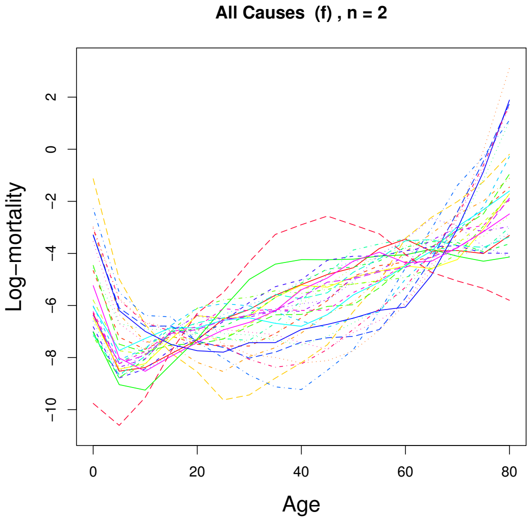
Samples from Age Prior



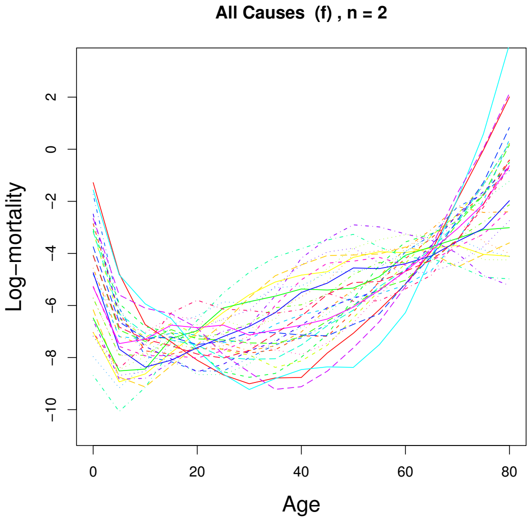
Samples from Age Prior



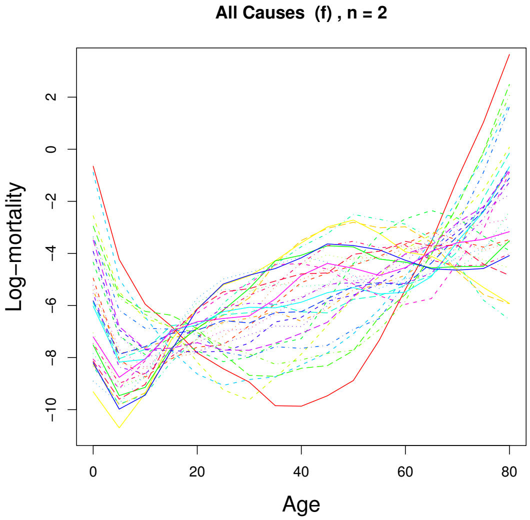
Samples from Age Prior



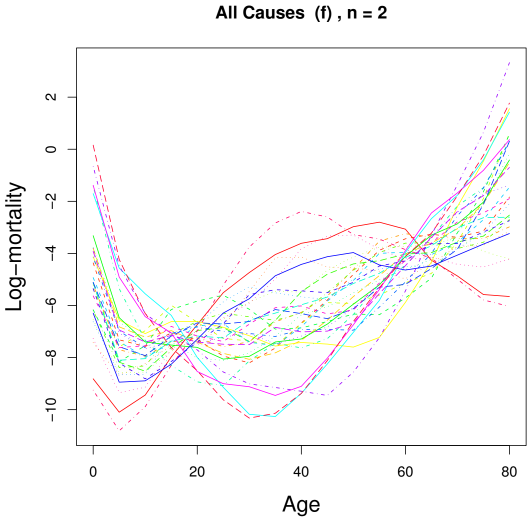
Samples from Age Prior



Samples from Age Prior



Samples from Age Prior



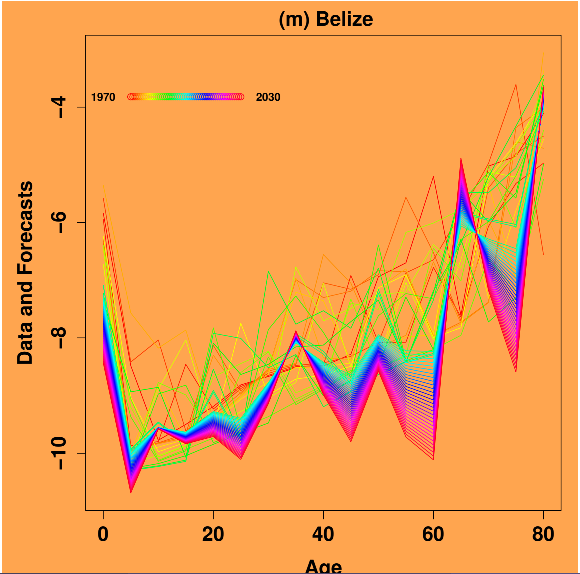
Generalizations

- 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:

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

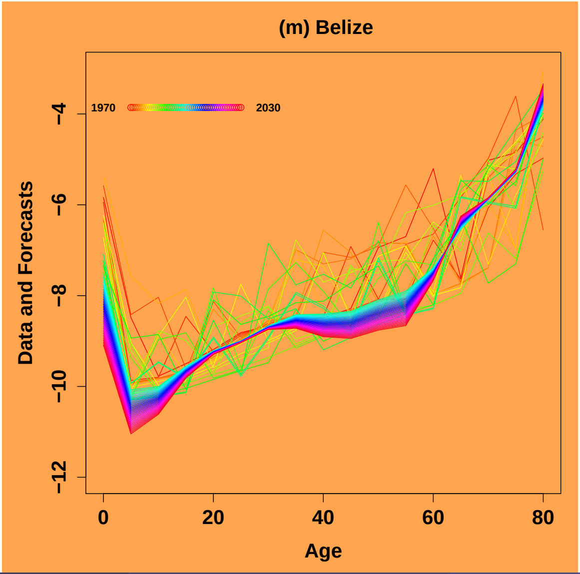
Mortality from Respiratory Infections, Males

Least Squares



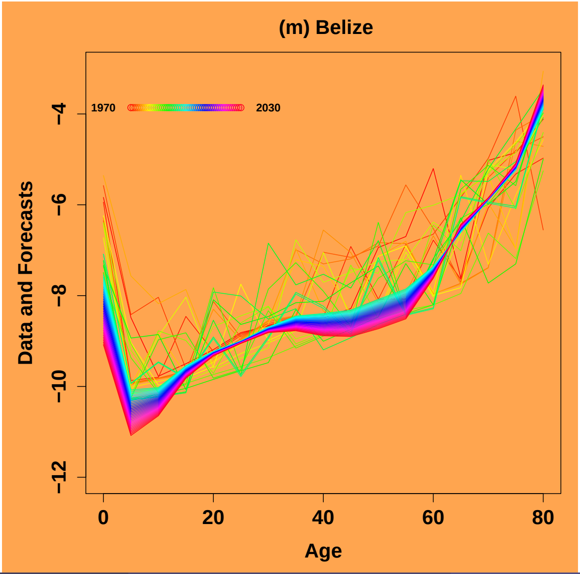
Mortality from Respiratory Infections, males, $\sigma = 2.00$

Smoothing over Age Groups



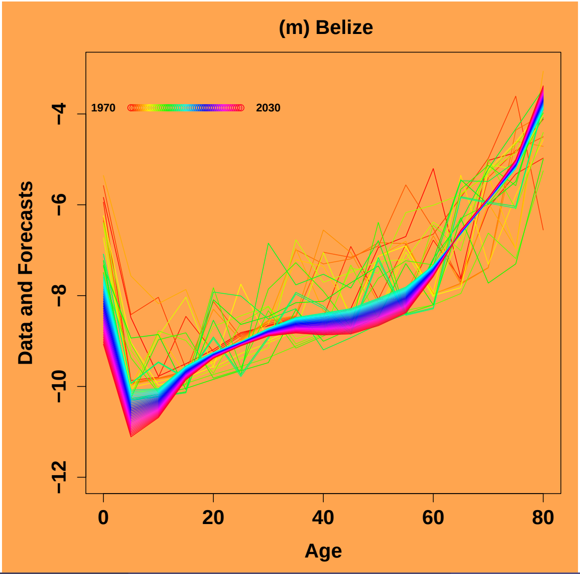
Mortality from Respiratory Infections, males, $\sigma = 1.51$

Smoothing over Age Groups



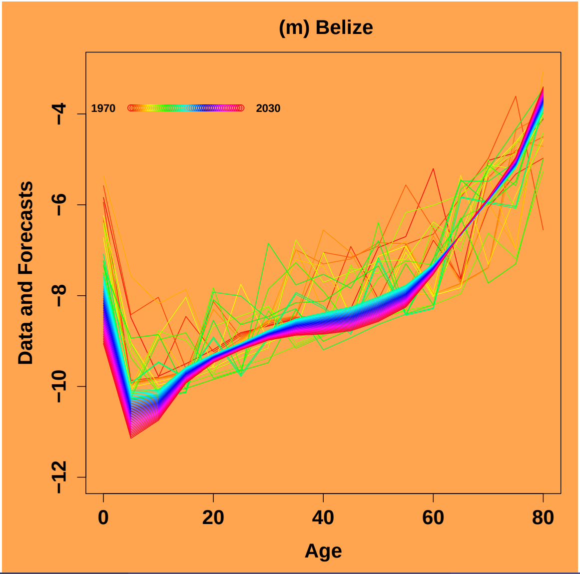
Mortality from Respiratory Infections, males, $\sigma = 1.15$

Smoothing over Age Groups



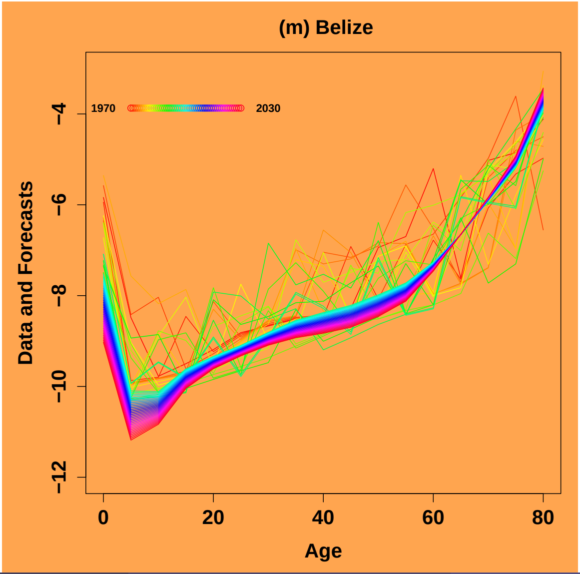
Mortality from Respiratory Infections, males, $\sigma = 0.87$

Smoothing over Age Groups



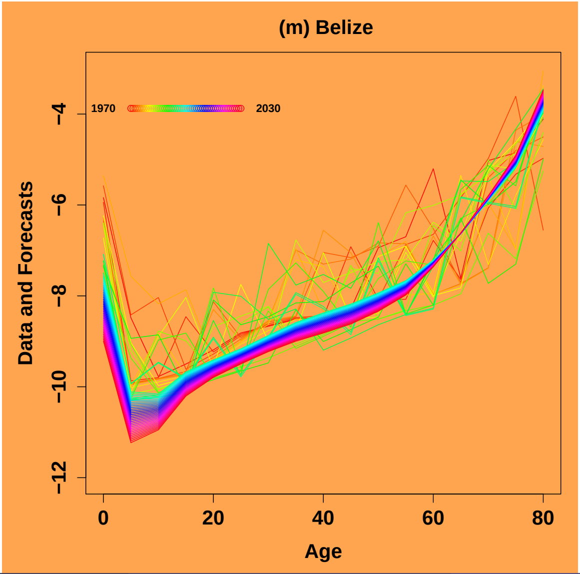
Mortality from Respiratory Infections, males, $\sigma = 0.66$

Smoothing over Age Groups



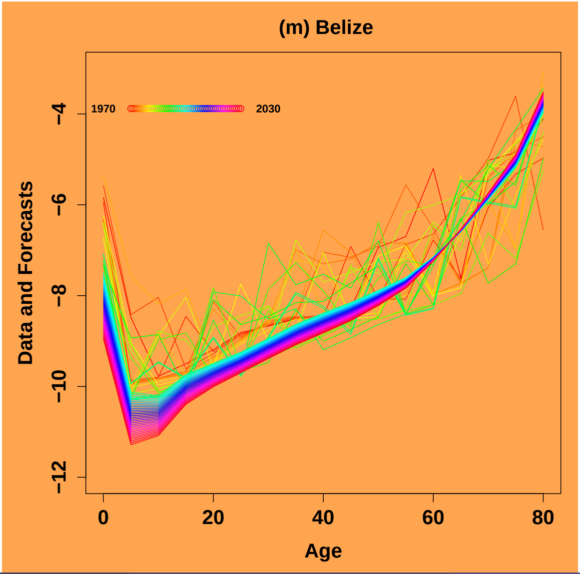
Mortality from Respiratory Infections, males, $\sigma = 0.50$

Smoothing over Age Groups



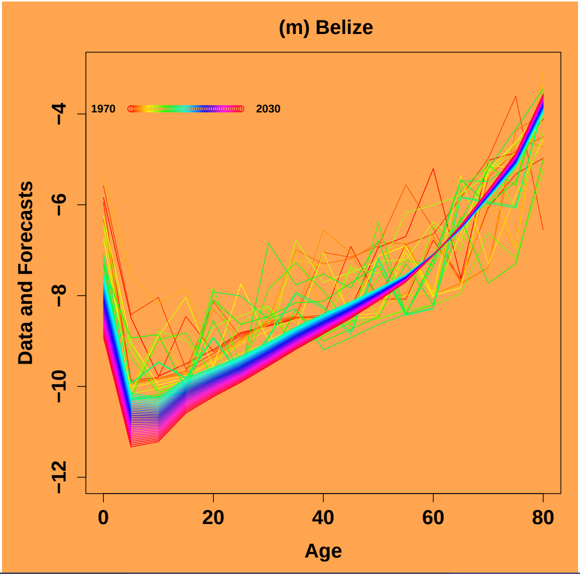
Mortality from Respiratory Infections, males, $\sigma = 0.38$

Smoothing over Age Groups



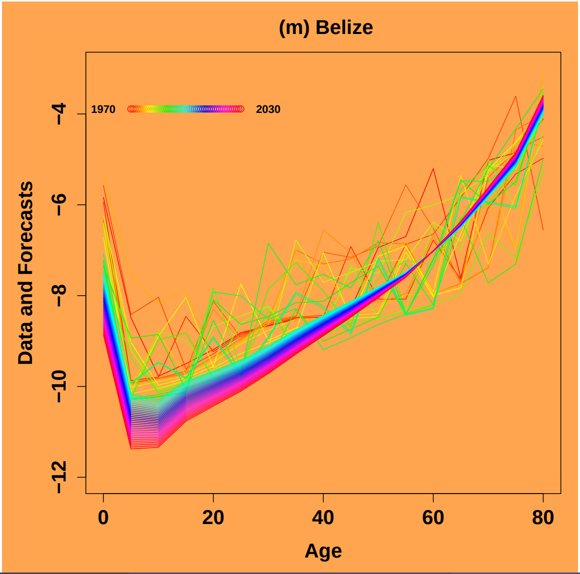
Mortality from Respiratory Infections, males, $\sigma = 0.28$

Smoothing over Age Groups



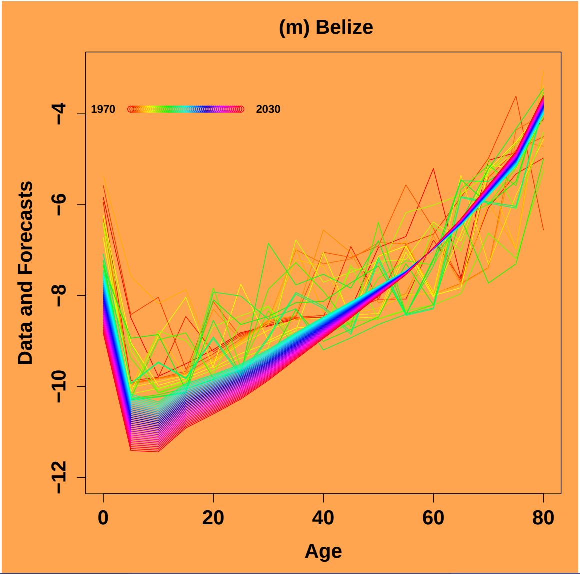
Mortality from Respiratory Infections, males, $\sigma = 0.21$

Smoothing over Age Groups



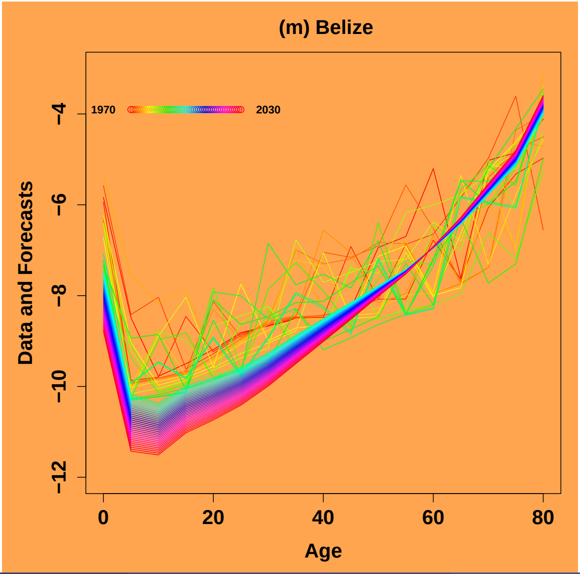
Mortality from Respiratory Infections, males, $\sigma = 0.16$

Smoothing over Age Groups



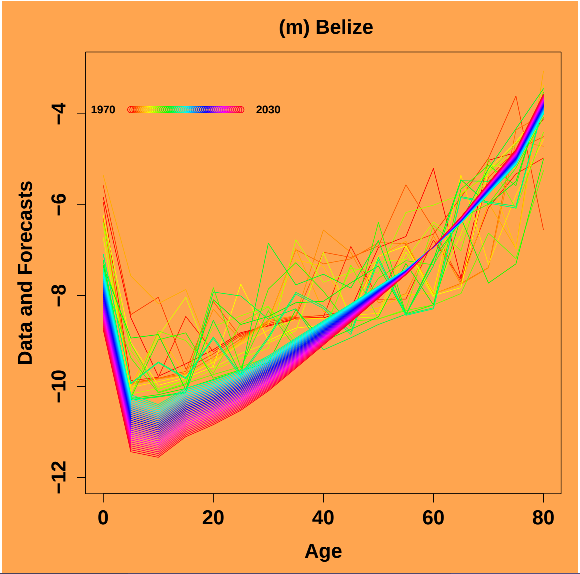
Mortality from Respiratory Infections, males, $\sigma = 0.12$

Smoothing over Age Groups



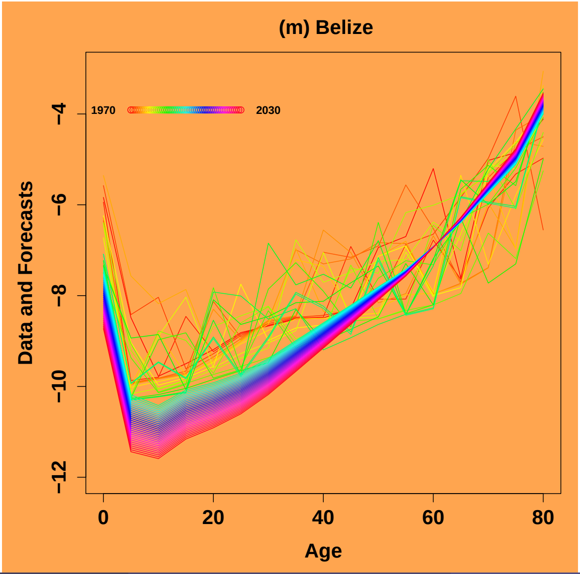
Mortality from Respiratory Infections, males, $\sigma = 0.09$

Smoothing over Age Groups



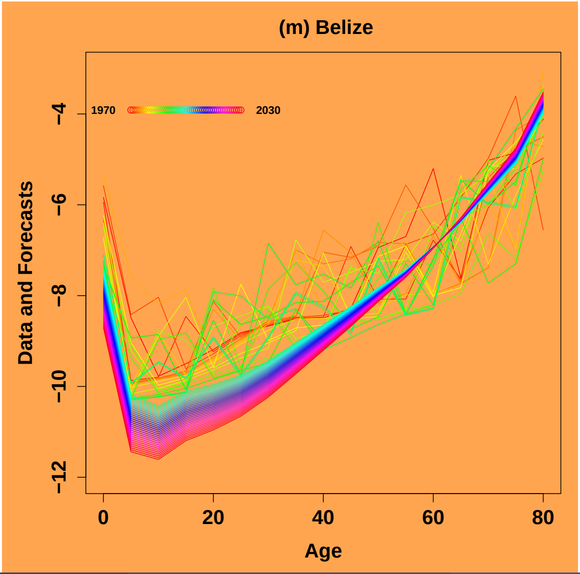
Mortality from Respiratory Infections, males, $\sigma = 0.07$

Smoothing over Age Groups



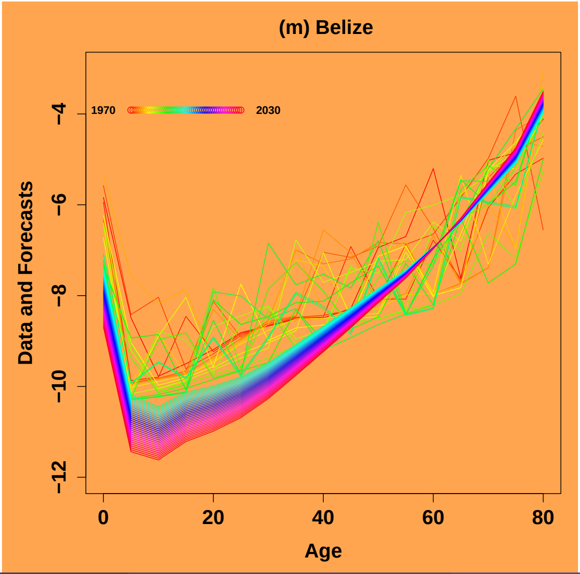
Mortality from Respiratory Infections, males, $\sigma = 0.05$

Smoothing over Age Groups



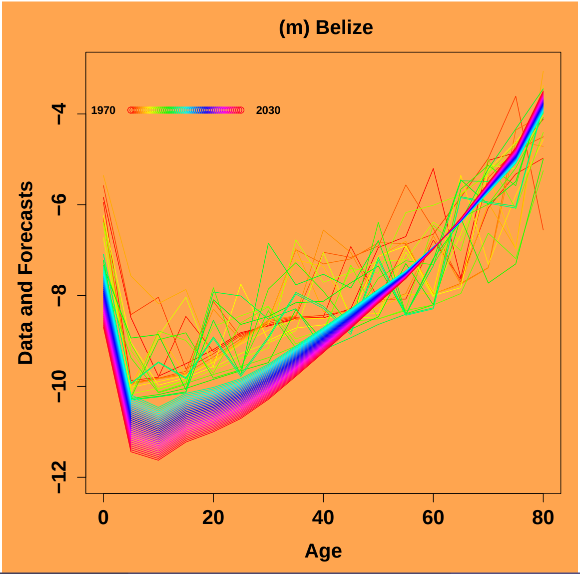
Mortality from Respiratory Infections, males, $\sigma = 0.04$

Smoothing over Age Groups



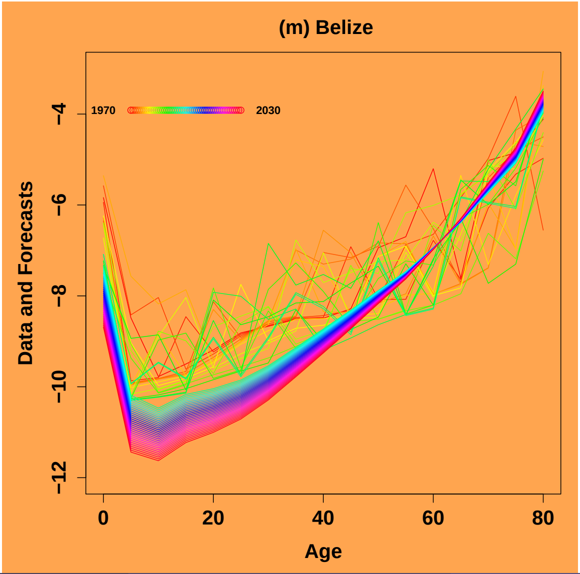
Mortality from Respiratory Infections, males, $\sigma = 0.03$

Smoothing over Age Groups



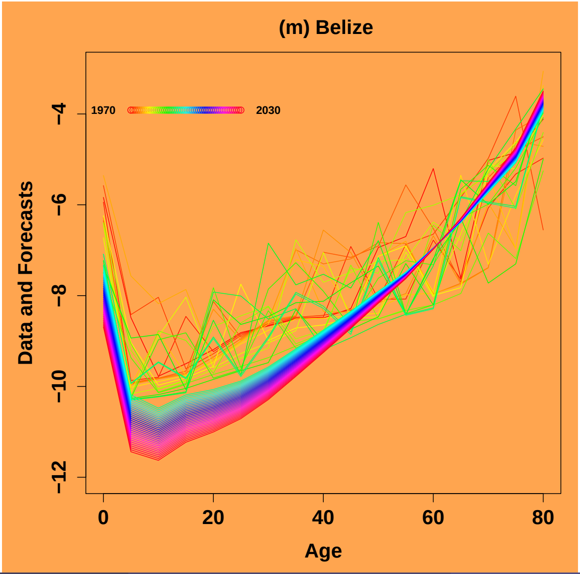
Mortality from Respiratory Infections, males, $\sigma = 0.02$

Smoothing over Age Groups



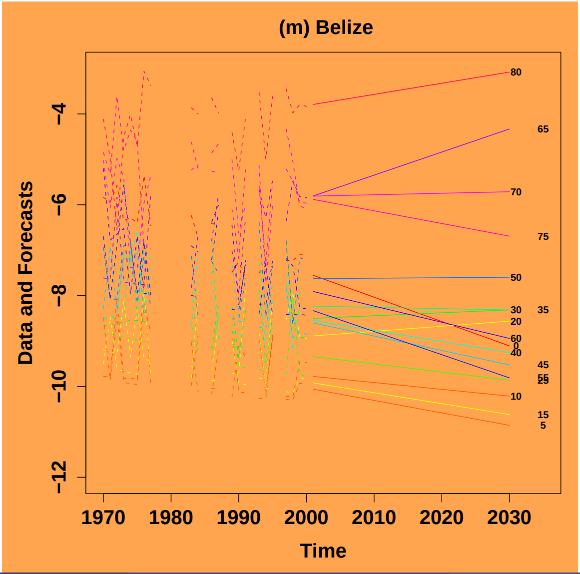
Mortality from Respiratory Infections, males, $\sigma = 0.01$

Smoothing over Age Groups



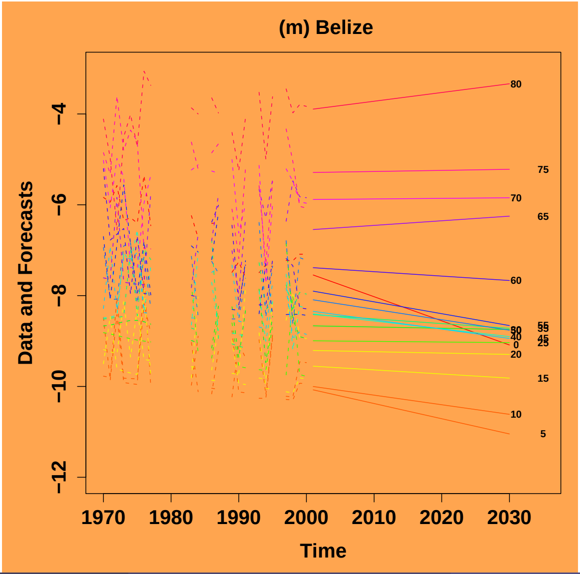
Mortality from Respiratory Infections, males

Least Squares



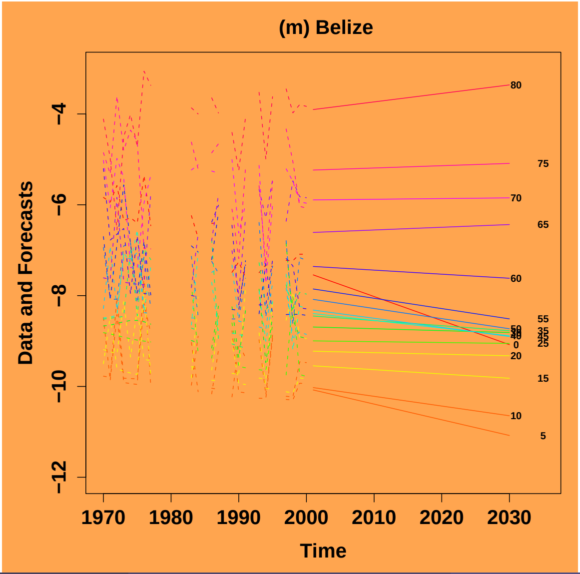
Mortality from Respiratory Infections, males, $\sigma = 2.00$

Smoothing over Age Groups



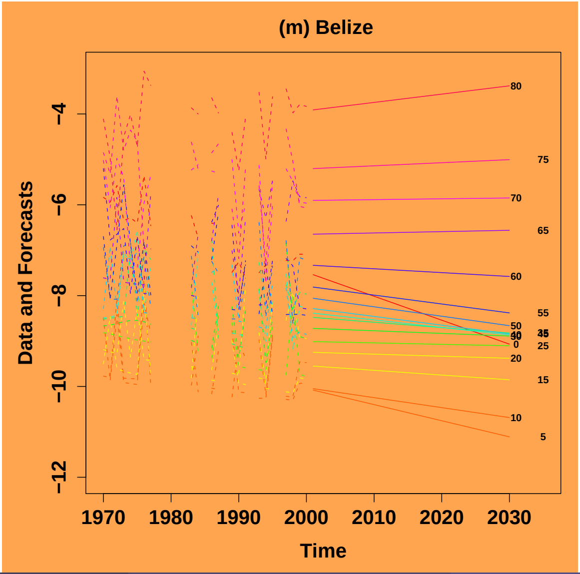
Mortality from Respiratory Infections, males, $\sigma = 1.51$

Smoothing over Age Groups



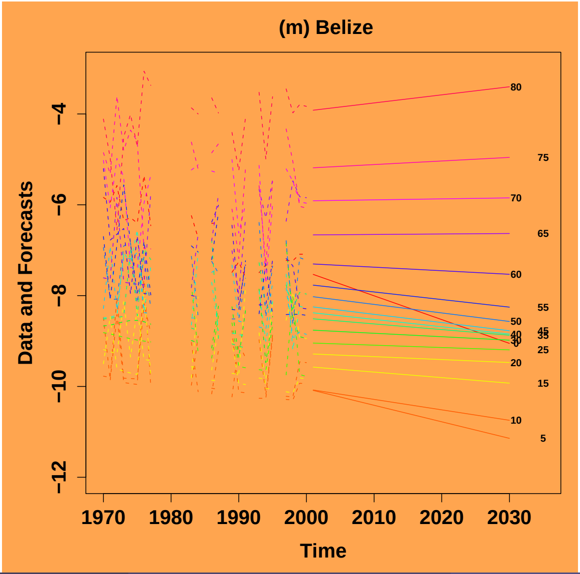
Mortality from Respiratory Infections, males, $\sigma = 1.15$

Smoothing over Age Groups



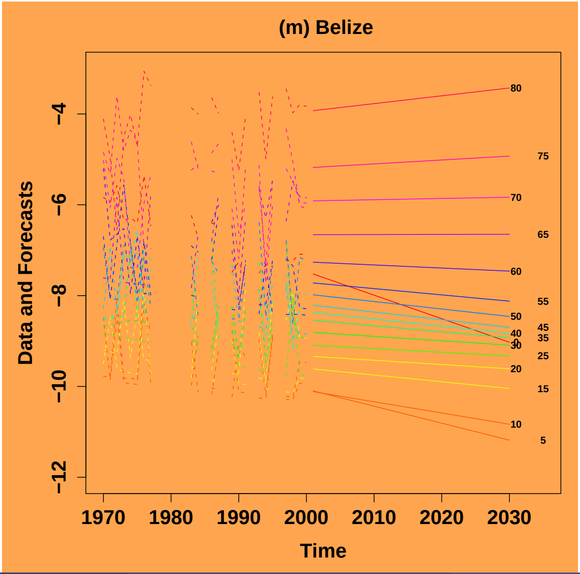
Mortality from Respiratory Infections, males, $\sigma = 0.87$

Smoothing over Age Groups



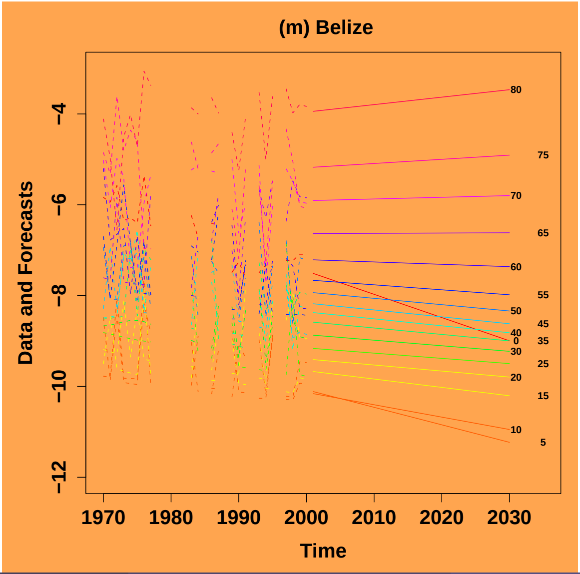
Mortality from Respiratory Infections, males, $\sigma = 0.66$

Smoothing over Age Groups



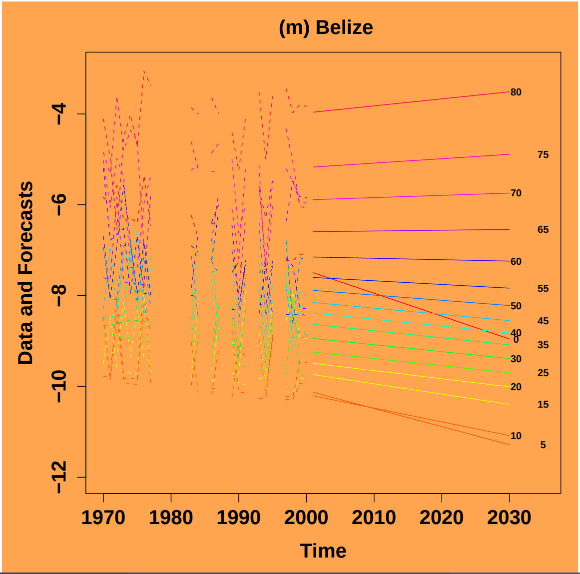
Mortality from Respiratory Infections, males, $\sigma = 0.50$

Smoothing over Age Groups



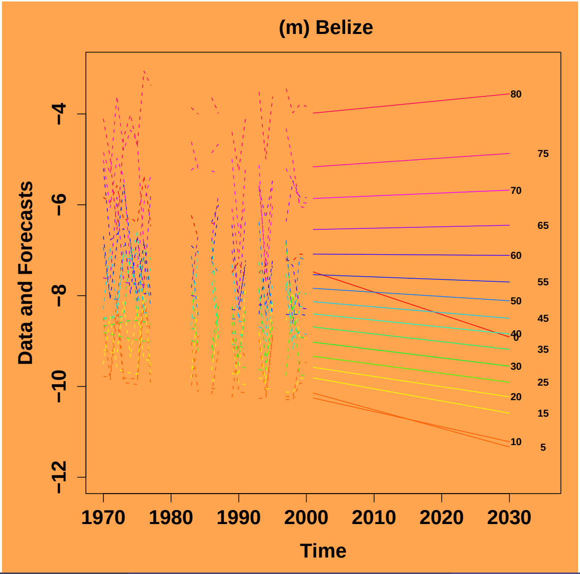
Mortality from Respiratory Infections, males, $\sigma = 0.38$

Smoothing over Age Groups



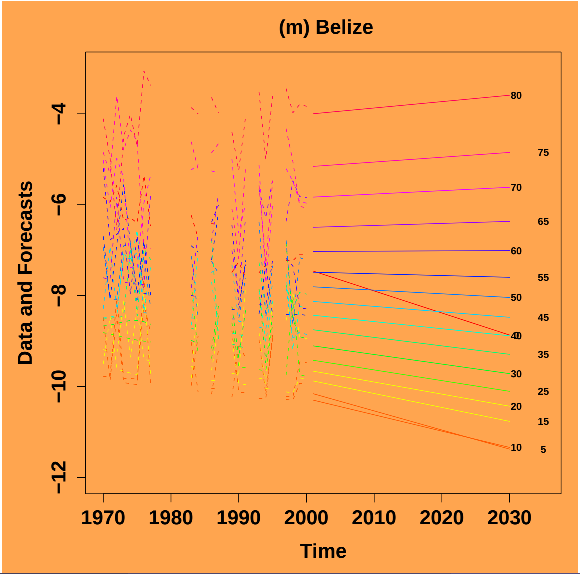
Mortality from Respiratory Infections, males, $\sigma = 0.28$

Smoothing over Age Groups



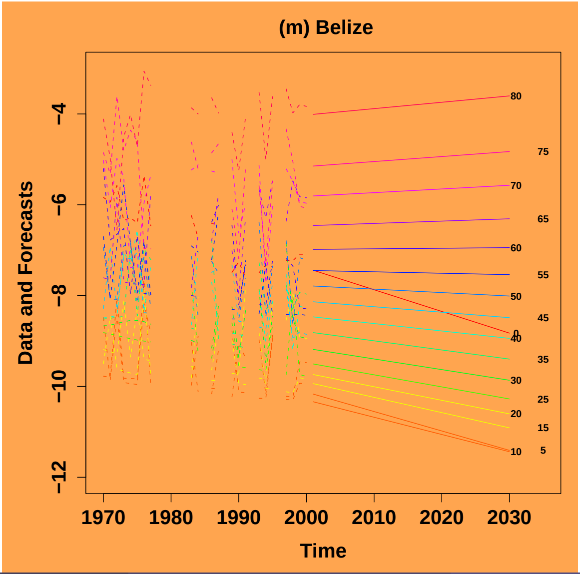
Mortality from Respiratory Infections, males, $\sigma = 0.21$

Smoothing over Age Groups



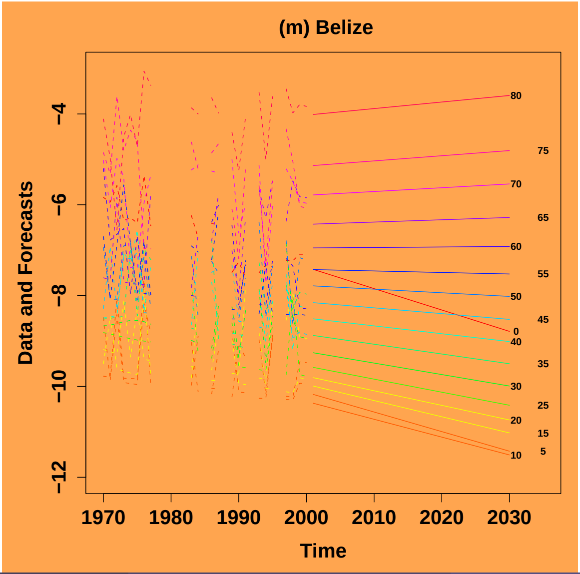
Mortality from Respiratory Infections, males, $\sigma = 0.16$

Smoothing over Age Groups



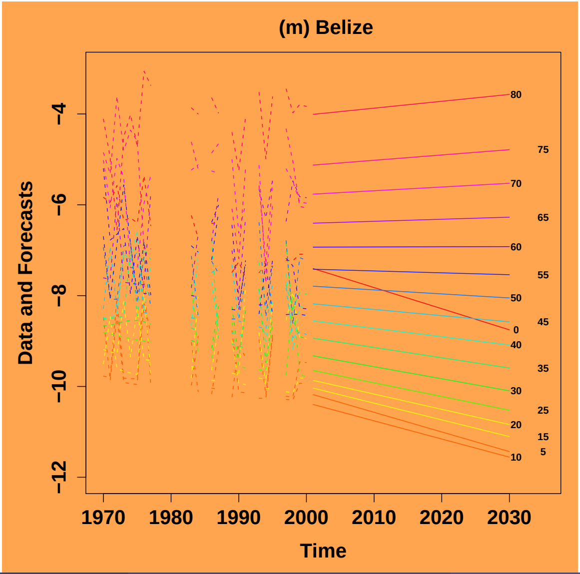
Mortality from Respiratory Infections, males, $\sigma = 0.12$

Smoothing over Age Groups



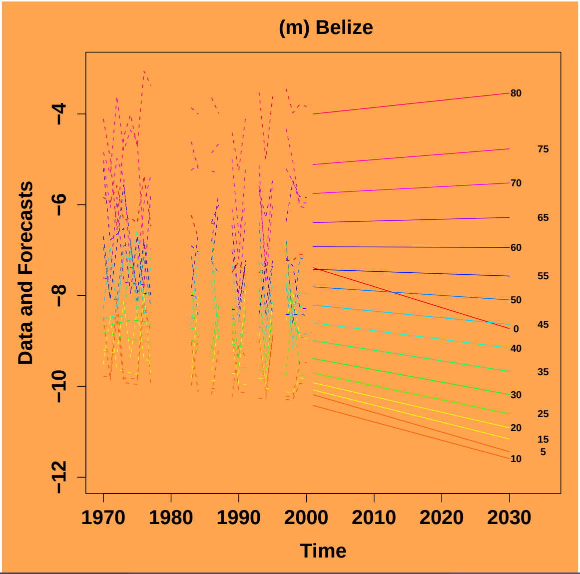
Mortality from Respiratory Infections, males, $\sigma = 0.09$

Smoothing over Age Groups



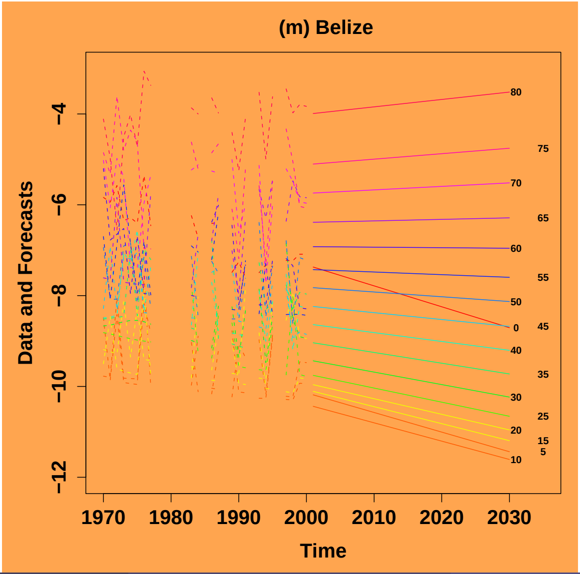
Mortality from Respiratory Infections, males, $\sigma = 0.07$

Smoothing over Age Groups



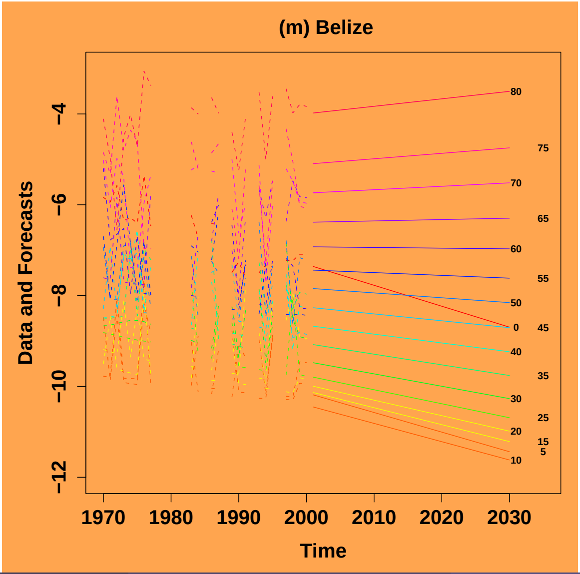
Mortality from Respiratory Infections, males, $\sigma = 0.05$

Smoothing over Age Groups



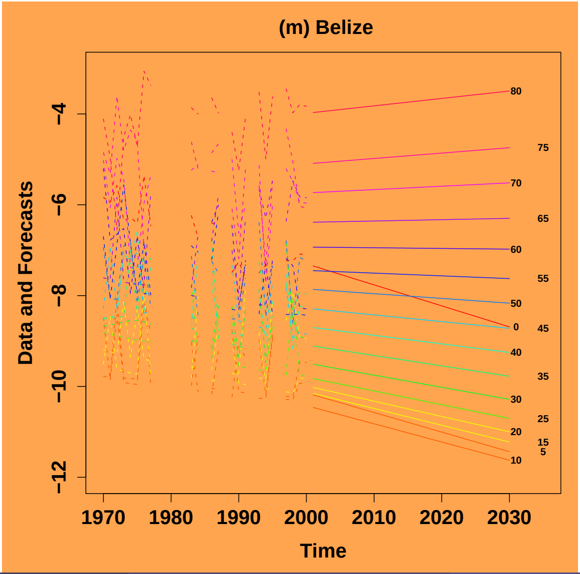
Mortality from Respiratory Infections, males, $\sigma = 0.04$

Smoothing over Age Groups



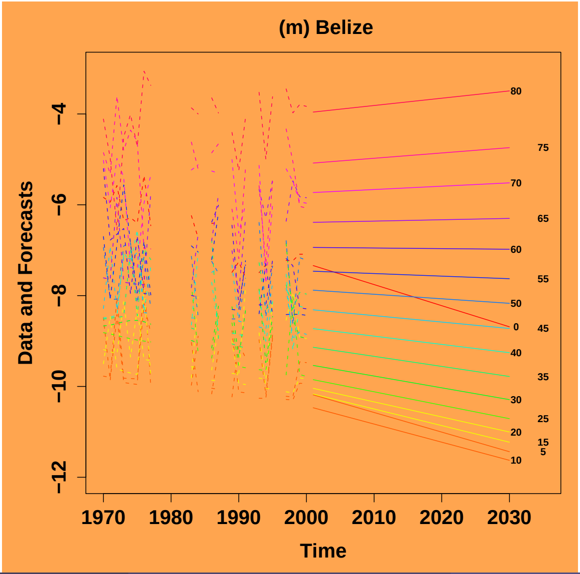
Mortality from Respiratory Infections, males, $\sigma = 0.03$

Smoothing over Age Groups



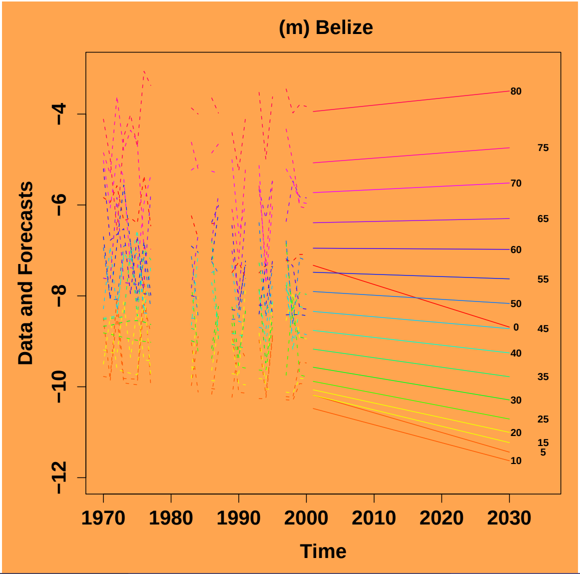
Mortality from Respiratory Infections, males, $\sigma = 0.02$

Smoothing over Age Groups



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

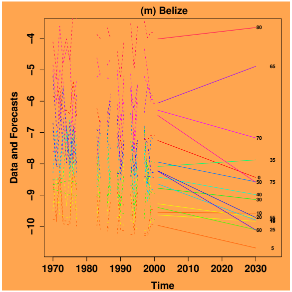
Smoothing over Age Groups

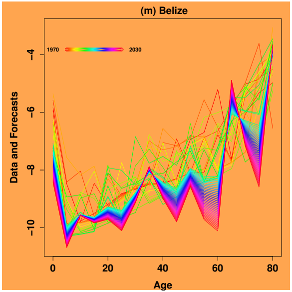


Smoothing Trends over Age Groups

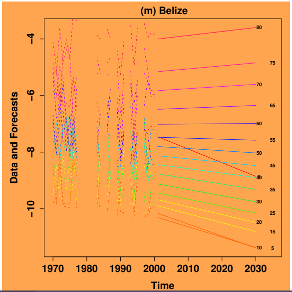
Log-mortality in Belize males from respiratory infections

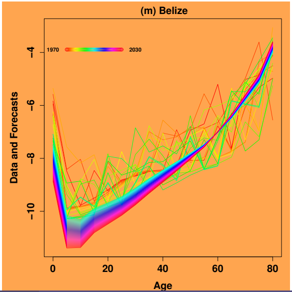
Least Squares





Smoothing
Age Groups

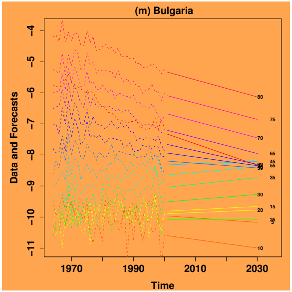


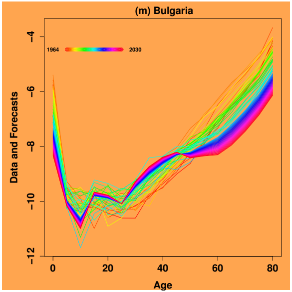


Smoothing Trends over Age Groups and Time

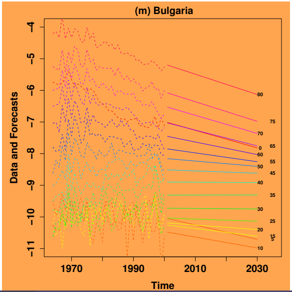
Log-Mortality in Bulgarian males from respiratory infections

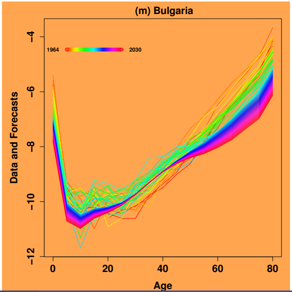
Least Squares





Smoothing
Age and Time

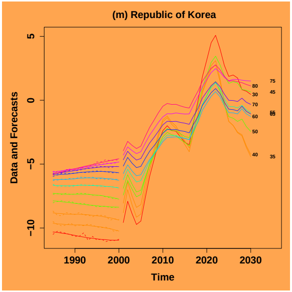


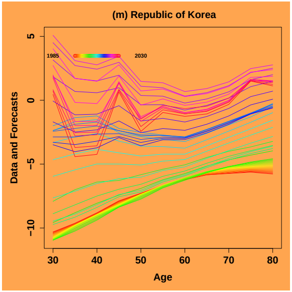


Using Covariates (GDP, tobacco, trend, log trend)

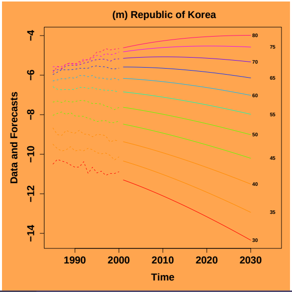
Lung cancer in Korean Males

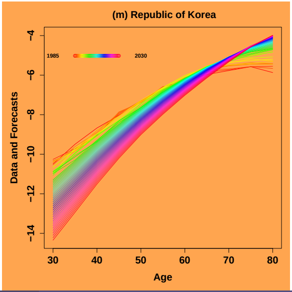
Least Squares





Smooth over age, time, age/time

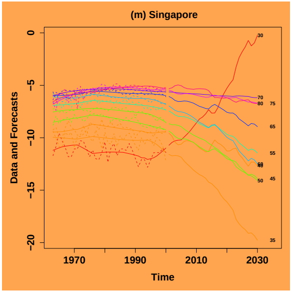


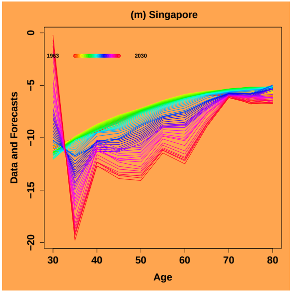


Using Covariates (GDP, tobacco, trend, log trend)

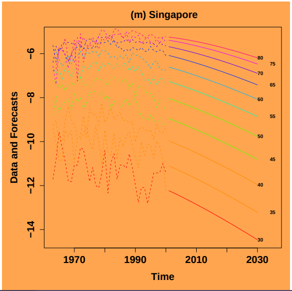
Lung cancer in Males, Singapore

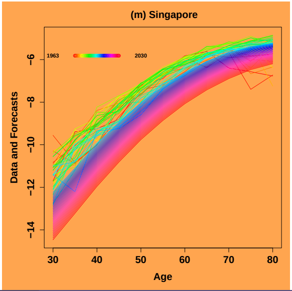
Least Squares



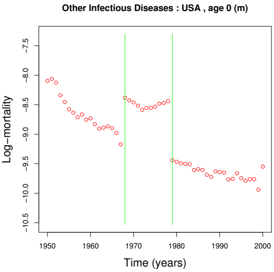


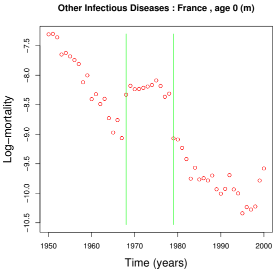
Smooth over age, time, age/time

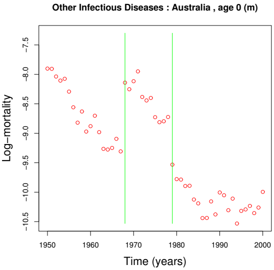


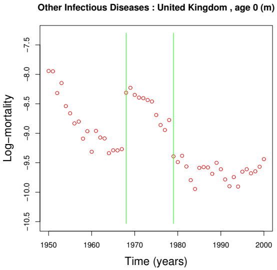


What about ICD Changes?

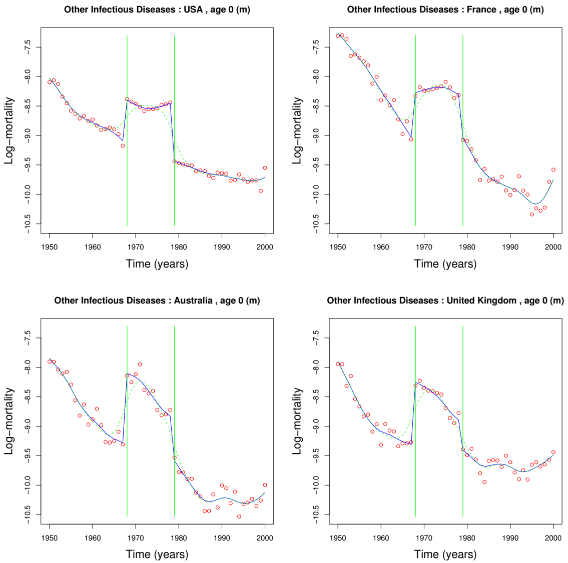








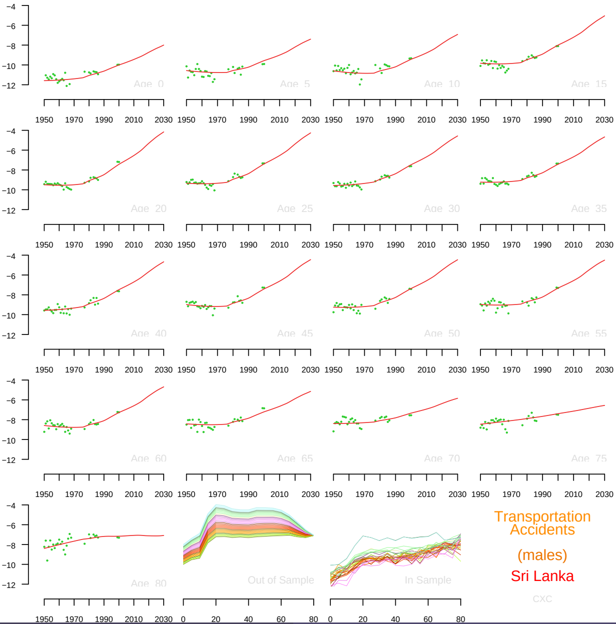
Fixing ICD Changes



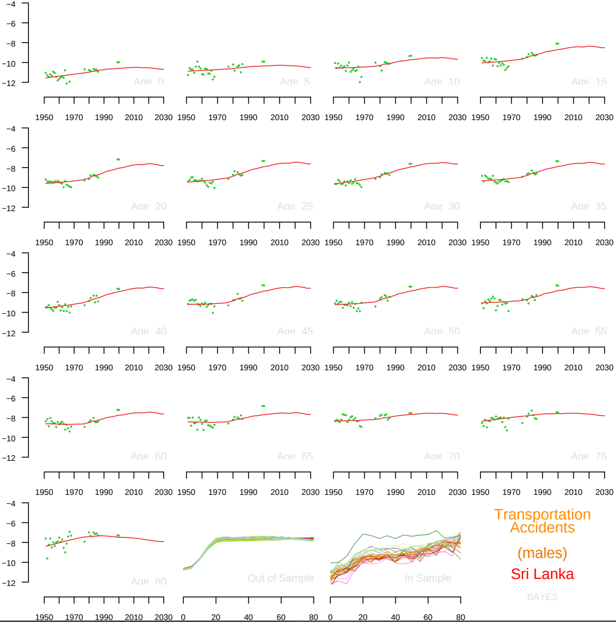
A book manuscript, YourCast software, etc.

http://GKing.Harvard.edu

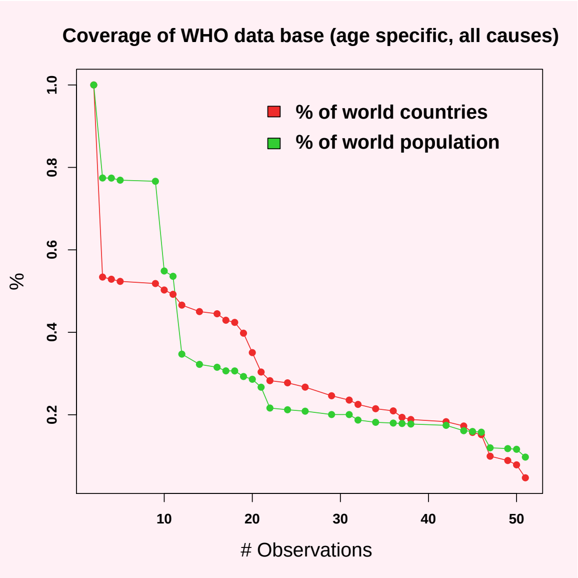
Without Country Smoothing



With Country Smoothing



Many Short Time Series



Prior Indifference

- These priors are "indifferent" to transformations:

$$\mu(a, t) \rightsquigarrow \mu(a, t) + p(a, t)$$

- 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)

	% Improvement	
	Over Best Previous	to Best Conceivable
Cardiovascular	22	49
Lung Cancer	24	47
Transportation	16	31
Respiratory Chronic	13	30
Other Infectious	12	30
Stomach Cancer	8	24
All-Cause	12	22
Suicide	7	17
Respiratory Infectious	3	7

- 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)

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

- 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