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

Other results we needed to achieve this original goal

Approach: Formalizing qualitative insights in quantitative models
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
What this Talk is About

- Mortality forecasts, which are studied in:
  - demography & sociology
Mortality forecasts, which are studied in:
- demography & sociology
- public health & biostatistics

A better forecasting method

Other results we needed to achieve this original goal

Approach: Formalizing qualitative insights in quantitative models
What this Talk is About

- Mortality forecasts, which are studied in:
  - demography & sociology
  - public health & biostatistics
  - economics & social security and retirement planning
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
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
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
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
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
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
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

- Approach: Formalizing qualitative insights in quantitative models
Other Results (Needed to Develop Improved Forecasts)

Output: same as linear regression
Estimates a set of linear regressions together (over countries, age groups, years, etc.)
Can include different covariates in each regression

We demonstrate that most hierarchical and spatial Bayesian models with covariates misrepresent prior information

Better ways of creating Bayesian priors

Produces forecasts and farcasts using considerably more information

Demographic Forecasting
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 (over countries, age groups, years, etc.)

Can include different covariates in each regression

We demonstrate that most hierarchical and spatial Bayesian models with covariates misrepresent prior information

Better ways of creating Bayesian priors

Produces forecasts and farcasts using considerably more information
Other Results (Needed to Develop Improved Forecasts)

A New Class of Statistical Models

- Output: same as linear regression
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 (over countries, age groups, years, etc.)
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 (over countries, age groups, years, etc.)
- Can include different covariates in each regression
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 (over countries, age groups, years, etc.)
- Can include different covariates in each regression
- We demonstrate that most hierarchical and spatial Bayesian models with covariates misrepresent prior information
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 (over countries, age groups, years, etc.)
- Can include *different covariates* in each regression
- We demonstrate that most hierarchical and spatial Bayesian models with covariates misrepresent prior information
- Better ways of creating Bayesian priors
Output: same as linear regression

Estimates a set of linear regressions together (over countries, age groups, years, etc.)

Can include different covariates in each regression

We demonstrate that most hierarchical and spatial Bayesian models with covariates misrepresent prior information

Better ways of creating Bayesian priors

Produces forecasts and farcasts using considerably more information
The Statistical Problem of Global Mortality Forecasting

779,799,281 deaths, in annual mortality rates

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
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates

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.
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- Multidimensional Data Structures: 24 causes of death, 17 age groups, 2 sexes, 191 countries, all for 50 annual observations.

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.
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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:
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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
779,799,281 deaths, in annual mortality rates

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 Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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:
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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.
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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.
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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;
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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;
The Statistical Problem of Global Mortality Forecasting

- 779,799,281 deaths, in annual mortality rates
- 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 (intercept and slope)
  - then forecast only these 2 parameters
- Reduces variance, constrains forecasts
- Dozens of more general functional forms proposed
- But does it fit anything else?
Gompertz (1825): log-mortality is linear in age after age 20
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 (intercept and slope)
Gompertz (1825): log-mortality is linear in age after age 20
- reduces 17 age-specific mortality rates to 2 parameters (intercept and slope)
- Then forecast only these 2 parameters
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 (intercept and slope)
  - Then forecast only these 2 parameters
  - Reduces variance, constrains forecasts
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 (intercept and slope)
  - Then forecast only these 2 parameters
  - Reduces variance, constrains forecasts
- Dozens of more general functional forms proposed
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 (intercept and slope)
  - Then 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
Mortality Age Profile: The Same Pattern?

Breast Cancer (f)

Japan
Venezuela
New Zealand

In(mortality)

Age

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
Mortality Age Profile: The Same Pattern?

Suicide (m)

Hungary
Canada
Colombia
Sri Lanka

Age
In(mortality)

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

Key question: how to include this qualitative information
Also: Method ignores covariate information; the leading current method (McNown-Rogers) not replicable
No mathematical form fits all or even most age profiles
Parameterizing Age Profiles Does Not Work

- No mathematical form fits all or even most age profiles
- Out-of-sample age profiles often unrealistic

Key question: how to include this qualitative information

Also: Method ignores covariate information; the leading current method (McNown-Rogers) not replicable
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:
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
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
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
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
- Key question: how to include this qualitative information
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

Key question: how to include this qualitative information

Also: Method ignores covariate information; the leading current method (McNown-Rogers) not replicable
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?

Demographic Forecasting
Random walk with drift; Lee-Carter; least squares on linear trend
Existing Method 2: Deterministic Projections

- Random walk with drift; Lee-Carter; least squares on linear trend
- Pros: simple, fast, works well in appropriate data
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
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
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
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?
The same pattern?
Random Walk with Drift ≈ Lee-Carter ≈ Least Squares
The same pattern?
Random Walk with Drift \approx Lee-Carter \approx Least Squares

(Data and Forecasts)

Suicide (m) USA

Time

1960 1980 2000 2020 2040 2060

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

Suicide   (m) USA

Time

Data and Forecasts

1950 2060

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

Demographic Forecasting

12 / 100
The same pattern?
Random Walk with Drift \approx \text{Lee-Carter} \approx \text{Least Squares}
The same pattern?
The same pattern?
Random Walk with Drift $\approx$ Lee-Carter $\approx$ Least Squares
The same pattern?
Random Walk with Drift $\approx$ Lee-Carter $\approx$ Least Squares

Transportation Accidents (m) Portugal

Data and Forecasts

1955 2060

Demographic Forecasting
The same pattern? 
Random Walk with Drift $\approx$ Lee-Carter $\approx$ Least Squares
Deterministic Projections Do Not Work

Linearity does not fit most time series data.
Out-of-sample age profiles become unrealistic over time.
Deterministic Projections Do Not Work

- Linearity does not fit most time series data
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:

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

$Z_{ca, t-\ell} \in \mathbb{R}^{d_{ca}}$: covariates (GDP, tobacco . . . ) lagged $\ell$ years.

$\beta_{ca} \in \mathbb{R}^{d_{ca}}$: coefficients to be estimated.

Cannot estimate equation by equation (variance is too large);

Pool over countries: $\beta_{ca} \Rightarrow \beta_{a}$

Properties:
- Small variance (due to large $n$)
- Large biases (due to restrictive pooling over countries), considerable information lost (due to no pooling over ages)

Demographic Forecasting
Model mortality over countries \((c)\) and ages \((a)\) as:

\[
m_{cat} = Z_{ca,t-\ell} \beta_{ca} + \epsilon_{cat} , \quad t = 1, \ldots, T
\]
Model mortality over countries \( (c) \) and ages \( (a) \) as:

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

\( Z_{ca,t-\ell} \in \mathbb{R}^{dca} \): covariates (GDP, tobacco \ldots) lagged \( \ell \) years.
Model mortality over countries \((c)\) and ages \((a)\) as:

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

- \(Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}}\) : covariates (GDP, tobacco \ldots) lagged \(\ell\) years.
- \(\beta_{ca} \in \mathbb{R}^{d_{ca}}\) : coefficients to be estimated
Model mortality over countries ($c$) and ages ($a$) as:

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

- $Z_{ca,t-\ell} \in \mathbb{R}^{dca}$ : covariates (GDP, tobacco ... ) lagged $\ell$ years.
- $\beta_{ca} \in \mathbb{R}^{dca}$ : coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
Regression Approaches (Murray and Lopez, 1996)

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

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

- \(Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}}\): covariates (GDP, tobacco \ldots) lagged \(\ell\) years.
- \(\beta_{ca} \in \mathbb{R}^{d_{ca}}\): coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
- Pool over countries: \(\beta_{ca} \Rightarrow \beta_{a}\)
Model mortality over countries \((c)\) and ages \((a)\) as:

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

- \(Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}}\): covariates (GDP, tobacco \ldots) lagged \(\ell\) years.
- \(\beta_{ca} \in \mathbb{R}^{d_{ca}}\): coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
- Pool over countries: \(\beta_{ca} \Rightarrow \beta_{a}\)
- Properties:
Regression Approaches (Murray and Lopez, 1996)

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

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

- \( Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}} \) : covariates (GDP, tobacco . . . ) lagged \( \ell \) years.
- \( \beta_{ca} \in \mathbb{R}^{d_{ca}} \) : coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
- Pool over countries: \( \beta_{ca} \Rightarrow \beta_{a} \)
- Properties:
  - Small variance (due to large \( n \))
Regression Approaches (Murray and Lopez, 1996)

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

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

- \(Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}}\) : covariates (GDP, tobacco . . . ) lagged \(\ell\) years.
- \(\beta_{ca} \in \mathbb{R}^{d_{ca}}\) : coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
- Pool over countries: \(\beta_{ca} \Rightarrow \beta_a\)
- Properties:
  - Small variance (due to large \(n\))
  - large biases (due to restrictive pooling over countries),
Model mortality over countries \((c)\) and ages \((a)\) as:

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

- \(Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}}\): covariates (GDP, tobacco . . . ) lagged \(\ell\) years.
- \(\beta_{ca} \in \mathbb{R}^{d_{ca}}\): coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
- Pool over countries: \(\beta_{ca} \Rightarrow \beta_{a}\)
- Properties:
  - Small variance (due to large \(n\))
  - large biases (due to restrictive pooling over countries),
  - considerable information lost (due to no pooling over ages)
Model mortality over countries \((c)\) and ages \((a)\) as:

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

- \( Z_{ca,t-\ell} \in \mathbb{R}^{d_{ca}} \): covariates (GDP, tobacco \ldots) lagged \( \ell \) years.
- \( \beta_{ca} \in \mathbb{R}^{d_{ca}} \): coefficients to be estimated
- Cannot estimate equation by equation (variance is too large);
- Pool over countries: \( \beta_{ca} \Rightarrow \beta_{a} \)

Properties:
- 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:

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

\[ P(\beta, \sigma, \theta \mid m) \propto P(m \mid \beta, \sigma) \times P(\beta \mid \theta) \times P(\theta)P(\sigma) = (\text{Likelihood}) \times (\text{Key Prior}) \times (\text{Other priors}) \]
Partial Pooling via a Bayesian Hierarchical Approach

- Likelihood for equation-by-equation least squares:
  \[ P(m | \beta_i, \sigma_i) = \prod_t N(m_{it} | Z_{it}\beta_i, \sigma_i^2) \]

- Add priors and form a posterior
  \[ P(\beta, \sigma, \theta | m) \propto P(m | \beta, \sigma) \times P(\beta | \theta) \times P(\theta)P(\sigma) \]
  \[ = (\text{Likelihood}) \times (\text{Key Prior}) \times (\text{Other priors}) \]

- Calculate point estimate for \( \beta \) (for \( \hat{y} \)) as the mean posterior:
  \[ \beta^{\text{Bayes}} \equiv \int \beta P(\beta, \sigma, \theta | m) d\beta d\theta d\sigma \]
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

\[ \mathcal{P}(\beta, \sigma, \theta \mid m) \propto \mathcal{P}(m \mid \beta, \sigma) \times \mathcal{P}(\beta \mid \theta) \times \mathcal{P}(\theta) \mathcal{P}(\sigma) \]

\[ = (\text{Likelihood}) \times (\text{Key Prior}) \times (\text{Other priors}) \]

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

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

- The hard part: specifying the prior for \( \beta \) and, as always, \( Z \)
Partial Pooling via a Bayesian Hierarchical Approach

- Likelihood for equation-by-equation least squares:
  \[ \mathcal{P}(m \mid \beta_i, \sigma_i) = \prod_t \mathcal{N}(m_{it} \mid Z_{it} \beta_i, \sigma_i^2) \]

- Add priors and form a posterior
  \[ \mathcal{P}(\beta, \sigma, \theta \mid m) \propto \mathcal{P}(m \mid \beta, \sigma) \times \mathcal{P}(\beta \mid \theta) \times \mathcal{P}(\theta)\mathcal{P}(\sigma) = \text{(Likelihood)} \times \text{(Key Prior)} \times \text{(Other priors)} \]

- Calculate point estimate for \( \beta \) (for \( \hat{y} \)) as the mean posterior:
  \[ \beta^{\text{Bayes}} \equiv \int \beta \mathcal{P}(\beta, \sigma, \theta \mid m) \, d\beta d\theta d\sigma \]

- 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:
$s_{ij}$ measures the similarity between cross-section $i$ and $j$.

$\sum_{ij} s_{ij} \parallel \beta_i - \beta_j \parallel^2 \Phi \equiv \parallel \beta_i - \beta_j \parallel^2 \Phi$

measures the distance between $\beta_i$ and $\beta_j$.

Natural choice for the prior:
$P(\beta | \Phi) \propto \exp \left( -\frac{1}{2} \sum_{ij} s_{ij} \parallel \beta_i - \beta_j \parallel^2 \Phi \right)$
The (Problematic) Classical Bayesian Approach

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

Requirements:

$s_{ij}$ measures the similarity between cross-section $i$ and $j$.

$(\beta_i - \beta_j)' \Phi (\beta_i - \beta_j) \equiv \| \beta_i - \beta_j \|_2^2 \Phi$

measures the distance between $\beta_i$ and $\beta_j$.

Natural choice for the prior:

$P(\beta | \Phi) \propto \exp \left( -\frac{1}{2} \sum_{ij} s_{ij} \| \beta_i - \beta_j \|_2^2 \Phi \right)$
The (Problematic) Classical Bayesian Approach

Assumption: similarities among cross-sections imply similarities among coefficients (β’s).

Requirements:
The (Problematic) Classical Bayesian Approach

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

Requirements:

- $s_{ij}$ measures the similarity between cross-section $i$ and $j$. 

$$ |\beta_i - \beta_j| \Phi = \parallel\beta_i - \beta_j\parallel_2 \Phi $$ 

Natural choice for the prior:

$$ P(\beta|\Phi) \propto \exp \left( -\frac{1}{2} \sum_{ij} s_{ij} \parallel\beta_i - \beta_j\parallel_2 \Phi \right) $$
The (Problematic) Classical Bayesian Approach

Assumption: similarities among cross-sections imply similarities among coefficients (β’s).

Requirements:

- \( s_{ij} \) measures the similarity between cross-section \( i \) and \( j \).
- \( (\beta_i - \beta_j)'\Phi(\beta_i - \beta_j) \equiv \|\beta_i - \beta_j\|^2_\Phi \) measures the distance between \( \beta_i \) and \( \beta_j \).
The (Problematic) Classical Bayesian Approach

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

**Requirements:**
- $s_{ij}$ measures the similarity between cross-section $i$ and $j$.
- $(\beta_i - \beta_j)' \Phi (\beta_i - \beta_j) \equiv \|\beta_i - \beta_j\|^2_\Phi$ measures the distance between $\beta_i$ and $\beta_j$.

**Natural choice for the prior:**

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

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

Prior knowledge about $\beta$ 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 $\beta$ more smooth may make $\mu$ less smooth ($\mu = Z\beta$):

$\mu - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j$

Coefficient variation + Covariate variation

Extensive trial-and-error runs, yielded no useful parameter values.
The (Problematic) Classical Bayesian Approach

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

Additionally, the parameter $\Phi$, the normalization factor, cannot be estimated and must be set. An uninformative prior for it would make Bayes irrelevant, while an informative prior cannot be used since we do not have information. Common practice involves making something work.

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

$$
\mu - \mu_{jt} = \mathbb{Z} \mu_{it} (\beta_i - \beta_j) + (\mathbb{Z} \mu_{it} - \mathbb{Z} \mu_{jt}) \beta_j = \text{Coefficient variation} + \text{Covariate variation}
$$

Extensive trial-and-error runs yielded no useful parameter values.
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:

$$\mu - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j$$

Coefficient variation + Covariate variation.

Extensive trial-and-error runs, yielded no useful parameter values.
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.

$\mu - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j$

Coefficient variation + Covariate variation

Extensive trial-and-error runs, yielded no useful parameter values.
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,
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 on it.
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**.

\[
\mu - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j
\]

Coefficient variation + Covariate variation

Extensive trial-and-error runs, yielded no useful parameter values.
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 ($\mu = Z\beta$):
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 ($\mu = Z\beta$):

  $$\mu_{it} - \mu_{jt}$$
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 ($\mu = Z\beta$):
  $$\mu_{it} - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j$$
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 ($\mu = Z\beta$):

\[
\mu_{it} - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j
\]

= Coefficient variation + Covariate variation

Extensive trial-and-error runs yielded no useful parameter values.
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 ($\mu = Z\beta$):
  \[
  \mu_{it} - \mu_{jt} = Z_{it}(\beta_i - \beta_j) + (Z_{it} - Z_{jt})\beta_j
  \]
  \[
  = \text{Coefficient variation} + \text{Covariate variation}
  \]
- Extensive trial-and-error runs, yielded no useful parameter values.
Our Alternative Strategy: Priors on $\mu$

Three steps:

1. Specify a prior for $\mu$:
   
   $$P(\mu | \theta) \propto \exp\left(-\frac{1}{2} H[\mu, \theta]\right),$$
   
   e.g.,

   $$H[\mu, \theta] \equiv \theta^T \sum_{t=1} A - \sum_{a=1} (\mu a t - \mu a + 1, t)^2$$

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)?

3. Constrain the prior on $\mu$ to the subspace spanned by the covariates:
   
   $$\mu = Z \beta$$

4. In the subspace, we can invert $\mu = Z \beta$ as
   
   $$\beta = (Z^T Z)^{-1} Z^T \mu,$$

   giving:

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

   the same prior on $\mu$, expressed as a function of $\beta$ (with constant Jacobian).
Our Alternative Strategy: Priors on $\mu$

Three steps:

1. Specify a prior for $\mu$:

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

Three steps:

1. Specify a prior for $\mu$:

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

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

2. To do Bayes, we need a prior on $\beta$

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

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

   the same prior on $\mu$, expressed as a function of $\beta$ (with constant Jacobian).
Our Alternative Strategy: Priors on $\mu$

Three steps:

1. Specify a prior for $\mu$:

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

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

Three steps:

1. **Specify a prior for $\mu$:**

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

   - To do Bayes, we need a prior on $\beta$
   - 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:**

$$\mu = Z\beta$$
Our Alternative Strategy: Priors on $\mu$

Three steps:

1. Specify a prior for $\mu$:

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

- To do Bayes, we need a prior on $\beta$
- 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:

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

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:

\[ \mu = Z \beta \]

**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 \( P(\mu) \), and treat it as the multidimensional \( P(\beta) \), constant in all directions but \( Z \beta \).

Demographic Forecasting

20 / 100
In other words

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

$$\mu_{cat} = Z_{cat}\beta_{ca}$$
In other words

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

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

A Simple Analogy
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

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

A Simple Analogy

- Suppose $\delta = \beta_1 - \beta_2$ and $P(\delta) = N(\delta|0, \sigma^2)$
In other words

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

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

A Simple Analogy

- Suppose $\delta = \beta_1 - \beta_2$ and $P(\delta) = N(\delta|0,\sigma^2)$
- What is $P(\beta_1, \beta_2)$?
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
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 $(\beta_1 - \beta_2)$. 
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

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

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 $\beta_1, \beta_2$ 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 $P(\beta_{ca})$, constant in all directions but $Z_{\text{cat}}\beta_{ca}$. 
Advantages of the resulting prior over $\beta$, created from prior over $\mu$
Advantages of the resulting prior over $\beta$, created from prior over $\mu$

- Fully Bayesian: The same theory of inference applies.

Priors are based on knowledge rather than guesses.

The normalization matrix $\Phi$ is unnecessary (task is performed by $Z$, which is known).
Advantages of the resulting prior over $\beta$, created from prior over $\mu$

- Fully Bayesian: The same theory of inference applies
- Can use standard Bayesian machinery for estimation.
Advantages of the resulting prior over $\beta$, created from prior over $\mu$

- Fully Bayesian: The same theory of inference applies
- Can use standard Bayesian machinery for estimation.
- $\mu_i$ and $\mu_j$ can always be compared, even with different covariates.
Advantages of the resulting prior over $\beta$, created from prior over $\mu$

- Fully Bayesian: The same theory of inference applies
- Can use standard Bayesian machinery for estimation.
- $\mu_i$ and $\mu_j$ can always be compared, even with different covariates.
- The normalization matrix $\Phi$ is unnecessary (task is performed by $Z$, which is known)
Advantages of the resulting prior over $\beta$, created from prior over $\mu$

- Fully Bayesian: The same theory of inference applies
- Can use standard Bayesian machinery for estimation.
- $\mu_i$ and $\mu_j$ can always be compared, even with different covariates.
- The normalization matrix $\Phi$ is unnecessary (task is performed by $Z$, which is known)
- Priors are based on knowledge rather than guesses.
An Age Prior

The prior is normal (and improper); Adjustable parameters: $n$ and $\theta$.

The choice of $n$ uniquely determines the "interaction" matrix $W_n$.

The variance of the prior is inversely proportional to $\theta$, which controls the "strength" of the prior.

Different age groups can have different covariates: the matrices $C_{aa'} \equiv 1^T Z_a Z_a'$ are rectangular ($d_a \times d_{a'}$).
The prior is normal (and improper); adjustable parameters:

\[ P(\mu \mid \theta) \sim P(\beta \mid \theta) \propto \exp \left( -\theta \sum_{aa'} W_{aa'}^n \beta_a^' C_{aa'} \beta_a^' \right) \]

The variance of the prior is inversely proportional to \( \theta \), which controls the "strength" of the prior.

Different age groups can have different covariates: the matrices \( C_{aa'} \equiv Z_a Z_a^' \) are rectangular (\( d_a \times d_a^' \)).
An Age Prior

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

- The prior is normal (and improper);
The prior is normal (and improper);

- Adjustable parameters: \( n \) and \( \theta \).
An Age Prior

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

- The prior is normal (and improper);
- Adjustable parameters: \( n \) and \( \theta \).
- The choice of \( n \) uniquely determines the “interaction” matrix \( W^n \).
An Age Prior

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

- The prior is normal (and improper);
- Adjustable parameters: \( n \) and \( \theta \).
- The choice of \( n \) uniquely determines the “interaction” matrix \( W^n \).
- The variance of the prior is inversely proportional to \( \theta \), which controls the “strength” of the prior.
The prior is normal (and improper);

Adjustable parameters: \( n \) and \( \theta \).

The choice of \( n \) uniquely determines the “interaction” matrix \( W^n \).

The variance of the prior is inversely proportional to \( \theta \), which controls the “strength” of the prior.

Different age groups can have different covariates: the matrices \( C_{aa'} \equiv \frac{1}{T}Z'_a Z_{a'} \) are rectangular \((d_a \times d_{a'})\).
All Causes (m), n = 1

Log-mortality

Age

Samples From Age Prior
All Causes (m), n = 2
Samples From Age Prior

All Causes (m), n = 3

Age

Log-mortality

All Causes (m), n = 3

Age

Log-mortality
Samples From Age Prior

All Causes \( (m), n = 4 \)

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

All Causes (m), n = 1

All Causes (m), n = 2

All Causes (m), n = 3

All Causes (m), n = 4
Prior Indifference

These priors are “indifferent” to transformations:

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

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.
These priors are “indifferent” to transformations:

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

equal color = equal probability
Formalizing (Prior) Indifference

equal color = equal probability

Level indifference
Formalizing (Prior) Indifference

equal color = equal probability

Level indifference

Level and slope indifference
The prior: \( P(\beta | \theta) \propto \exp(-\theta \sum a a' W_n a a' \beta' a C a a' \beta' a) \)

We figured out what \( n \) is, but what is the smoothness parameter, \( \theta \)?

\( \theta \) controls the prior standard deviation.
The prior:

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

The prior:

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

We figured out what $n$ is
The prior:

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

- We figured out what \( n \) is
- but what is the smoothness parameter, \( \theta \)?
Smoothness Parameter

- The prior:

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

- 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

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
Samples from Age Prior

All Causes (f), n = 2

Age
Log-mortality

Age
0 20 40 60 80
−10 −8 −6 −4 −2 0 2

Log-mortality
−10 −8 −6 −4 −2 0 2

Demographic Forecasting 33 / 100
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

Log-mortality vs Age

Demographic Forecasting
Samples from Age Prior

All Causes (f), n = 2

Log–mortality vs 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

![Graph showing log-mortality against age for different causes.](image)
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:

$$P(\beta | \theta) \propto \exp \left[ -\theta^2 \sum_{ij} W_{ij} \beta_i' C_{ij} \beta_j \right], \quad C_{aa} \equiv 1^T Z_a Z_a'$$
Generalizations

- The above tools: smooth over a (possibly discretized) continuous variable — age or age groups.

The mathematical form for all these (separately or together) turns out to be the same:

$$P(\beta | \theta) \propto \exp \left( -\theta^2 \sum_{ij} W_{ij} \beta_i' C_{ij} \beta_j \right), \quad C_{aa'} \equiv \frac{1}{T} \mathbf{Z}_a \mathbf{Z}_a'$$
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).

The mathematical form for all these (separately or together) turns out to be the same:

\[ P(\beta | \theta) \propto \exp \left( -\theta \sum_{ij} W_{ij} \beta_i' \mathbb{C}_{ij} \beta_j \right), \]

\[ \mathbb{C}_{aa}' \equiv \mathbb{1}_{T} \mathbb{Z}_a \mathbb{Z}_a' \]
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.

The mathematical form for all these (separately or together) turns out to be the same:

$$P(\beta | \theta) \propto \exp\left(-\theta^2 \sum_{ij} W_{ij} \beta_i C_{ij} \beta_j\right), \quad C_{aa}^t \equiv 1$$
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).

$$P(\beta | \theta) \propto \exp \left[ -\theta \sum_{ij} W_{ij} \beta'_i C_{ij} \beta_j \right],$$
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:

$$P(\beta | \theta) \propto \exp \left( -\theta^2 \sum_{ij} W_{ij} \beta_i' C_{ij} \beta_j \right),$$

$C_{aa} \equiv 1_T Z_a Z_a'$
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.
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.
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:

\[
P(\beta \mid \theta) \propto \exp \left( -\frac{\theta}{2} \sum_{ij} W_{ij}\beta'_i C_{ij}\beta_j \right), \quad C_{aa'} \equiv \frac{1}{T} Z_a Z_{a'}
\]
Mortality from Respiratory Infections, Males

Least Squares

Data and Forecasts

(m) Belize

1970 - 2030

Age

Data and Forecasts

(m) Belize

1970 - 2030

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

Smoothing over Age Groups

Data and Forecasts

(m) Belize
Mortality from Respiratory Infections, males, $\sigma = 1.51$

Smoothing over Age Groups

Data and Forecasts

Age

Demographic Forecasting

(m) Belize
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

![Graph showing data and forecasts for mortality from respiratory infections in Belize from 1970 to 2030, with age groups and smoothing parameters.]
Mortality from Respiratory Infections, males, $\sigma = 0.28$

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

Smoothing over Age Groups

![Graph showing data and forecasts for mortality from Respiratory Infections in Belize over age groups from 1970 to 2030.](Image)
Mortality from Respiratory Infections, males, $\sigma = 0.16$

Smoothing over Age Groups

Data and Forecasts

(m) Belize
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

![Graph showing data and forecasts for mortality from respiratory infections in Belize.](image)
Mortality from Respiratory Infections, males, \( \sigma = 0.04 \)

Smoothing over Age Groups

Data and Forecasts

(m) Belize

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

Smoothing over Age Groups

Data and Forecasts
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

Demographic Forecasting
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

Data and Forecasts
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

Data and Forecasts

(m) Belize

Time


-12 -10 -8 -6 -4


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

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

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

Data and Forecasts

(m) Belize

Time


Time

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

Data and Forecasts

(m) Belize

Time

Demographic Forecasting
Smoothing Trends over Age Groups

Demographic Forecasting
Smoothing Trends over Age Groups

Log-mortality in Belize males from respiratory infections

<table>
<thead>
<tr>
<th>Time (m)</th>
<th>Belize</th>
<th>Data and Forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2030</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age (m)</th>
<th>Belize</th>
<th>Data and Forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Demographic Forecasting
Smoothing Trends over Age Groups

Log-mortality in Belize males from respiratory infections

Least Squares
Smoothing Trends over Age Groups

Log-mortality in Belize males from respiratory infections

Least Squares
Smoothing Trends over Age Groups

Log-mortality in Belize males from respiratory infections

Least Squares
Smoothing Trends over Age Groups

Log-mortality in Belize males from respiratory infections

Least Squares

Smoothing 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
Log-mortality in Belize males from respiratory infections

Least Squares

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

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

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

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

Least Squares

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

Least Squares

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

Lung cancer in Korean Males

Data and Forecasts
Using Covariates (GDP, tobacco, trend, log trend)
Lung cancer in Korean Males

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

Lung cancer in Korean Males

Least Squares
Using Covariates (GDP, tobacco, trend, log trend)
Lung cancer in Korean Males

Least Squares
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 Korean Males

Least Squares

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

Lung cancer in Males, Singapore
Using Covariates (GDP, tobacco, trend, log trend)

Lung cancer in Males, Singapore

Least Squares
Using Covariates (GDP, tobacco, trend, log trend)
Lung cancer in Males, Singapore

Least Squares
Using Covariates (GDP, tobacco, trend, log trend)
Lung cancer in Males, Singapore

Least Squares
Using Covariates (GDP, tobacco, trend, log trend)
Lung cancer in Males, Singapore

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

Data and Forecasts
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)
http://GKing.Harvard.edu
<table>
<thead>
<tr>
<th>Category</th>
<th>% Improvement Over Best</th>
<th>Previous 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.

Demographic Forecasting 90 / 100
<table>
<thead>
<tr>
<th>Category</th>
<th>Improvement Over Best to Best Previous Conceivable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiovascular</td>
<td>22%</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>49%</td>
</tr>
<tr>
<td>Transportation</td>
<td>24%</td>
</tr>
<tr>
<td>Respiratory Chronic</td>
<td>16%</td>
</tr>
<tr>
<td>Other Infectious</td>
<td>31%</td>
</tr>
<tr>
<td>Stomach Cancer</td>
<td>30%</td>
</tr>
<tr>
<td>All-Cause</td>
<td>22%</td>
</tr>
<tr>
<td>Suicide</td>
<td>24%</td>
</tr>
<tr>
<td>Respiratory Infectious</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>% Improvement</th>
<th>Over Best Previous</th>
<th>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>% 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).
## Preview of Results: Out-of-Sample Evaluation

Mean Absolute Error in Males (over age and country)

<table>
<thead>
<tr>
<th>% Improvement</th>
<th>Over Best Previous</th>
<th>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.
## Preview of Results: Out-of-Sample Evaluation

Mean Absolute Error in Males (over age and country)

<table>
<thead>
<tr>
<th>% Improvement</th>
<th>Over Best Previous</th>
<th>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.
## Preview of Results: Out-of-Sample Evaluation

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

<table>
<thead>
<tr>
<th>Category</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**
Basic Prior: Smoothness over Age Groups

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

$$\mathcal{H}[\mu, \theta] \equiv \theta \int T_0 dt \int A_0 da \left( \frac{d}{da}n \right) \left( \mu(a, t) - \bar{\mu}(a) \right)^2$$

Discretize age and time:

$$P(\mu | \theta) \propto \exp \left( -\frac{1}{2} \theta \sum_{a} \sum_{a'} (\mu_{at} - \bar{\mu}_{a}) \left( \frac{d}{da}W_{n} \right) (\mu_{a't} - \bar{\mu}_{a'}) \right)$$

where $W_{n}$ is a matrix uniquely determined by $n$ and $\theta$. 
Prior knowledge: log-mortality age profile are smooth variations of a "typical" age profile $\bar{\mu}(a)$:

$$H[\mu, \theta] \equiv$$
Basic Prior: Smoothness over Age Groups

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

$$H[\mu, \theta] \equiv$$
Prior knowledge: log-mortality age profile are smooth variations of a “typical” age profile $\bar{\mu}(a)$:

$$H[\mu, \theta] \equiv [\mu(a, t) - \bar{\mu}(a)]$$
Prior knowledge: log-mortality age profile are smooth variations of a "typical" age profile $\bar{\mu}(a)$:

$$H[\mu, \theta] \equiv \frac{d^n}{da^n} [\mu(a, t) - \bar{\mu}(a)]$$
Prior knowledge: log-mortality age profile are smooth variations of a “typical” age profile $\bar{\mu}(a)$:

\[
H[\mu, \theta] \equiv \left( \frac{d^n}{da^n} [\mu(a, t) - \bar{\mu}(a)] \right)^2
\]
Prior knowledge: log-mortality age profile are smooth variations of a “typical” age profile $\bar{\mu}(a)$:

$$H[\mu, \theta] \equiv \int_0^A da \left( \frac{d^n}{da^n} [\mu(a, t) - \bar{\mu}(a)] \right)^2$$
Basic Prior: Smoothness over Age Groups

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

$$H[\mu, \theta] \equiv \int_0^T dt \int_0^A da \left( \frac{d^n}{da^n} [\mu(a, t) - \bar{\mu}(a)] \right)^2$$
Prior knowledge: log-mortality age profile are smooth variations of a “typical” age profile $\bar{\mu}(a)$:

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

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

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

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

$\mu = \bar{\mu} + Z$ at $\beta$

$P(\beta | \theta) = \exp(-\theta^T \sum a a' W n a a' (Z \beta a) (Z a' \beta a'))$

$= \exp(-\theta \sum a a' W n a a' \beta a C a a' \beta a')$

where we have defined:

$C a a' \equiv 1^T Z a' Z a$

$Z a$ is a $T \times d$ a data matrix for age group $a$.
From a prior on $\mu$ to a prior on $\beta$

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

$$P(\beta | \theta) = \exp(-\theta^T \sum a a' W_n a a' (Z_{at} \beta_a (Z_{a'} t \beta_a'))$$

where we have defined:

$$C_{aa'} \equiv 1^T Z_a Z_a'$$

$Z_{at}$ is a $T \times d_a$ data matrix for age group $a$.
Add the specification $\mu_{at} = \bar{\mu}_a + Z_{at}\beta_a$:

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

$$
= \exp \left( -\theta \sum_{aa'} W_{aa'}^n \beta_a' C_{aa'} \beta_{a'} \right)
$$
From a prior on $\mu$ to a prior on $\beta$

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

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

$$= \exp \left( -\theta \sum_{aa'} W_{aa'}^n \beta'_a C_{aa'} \beta_{a'} \right)$$

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 $\beta$

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

The prior is normal (and improper); the prior is uniquely determined by the choice of $n$, through the "interaction" matrix $W_{aa'}^n$. Different age groups can have different covariates: the matrices $C_{aa'}$ are rectangular ($d_a \times d_{a'}$). The variance of the prior is inversely proportional to $\theta$, which controls the "strength" of the prior.
The Prior on the Coefficients $\beta$

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

- The prior is normal (and improper);
The Prior on the Coefficients $\beta$

$$\mathcal{P}(\beta \mid \theta) \propto \exp \left( -\theta \sum_{aa'} W^n_{aa'} \beta'_a C_{aa'} \beta_{a'} \right)$$

- The prior is normal (and improper);
- The prior is uniquely determined by the choice of $n$, through the “interaction” matrix $W^n$. 
The Prior on the Coefficients $\beta$

$$P(\beta \mid \theta) \propto \exp \left( -\theta \sum_{aa'} W^n_{aa'} \beta_a' C_{aa'} \beta_{a'} \right)$$

- The prior is normal (and improper);
- The prior is uniquely determined by the choice of $n$, through the “interaction” matrix $W^n$.
- Different age groups can have different covariates: the matrices $C_{aa'}$ are rectangular ($d_a \times d_{a'}$).
The Prior on the Coefficients $\beta$

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

- The prior is normal (and improper);
- The prior is uniquely determined by the choice of $n$, through the “interaction” matrix $W^n$.
- Different age groups can have different covariates: the matrices $C_{aa'}$ are rectangular ($d_a \times d_{a'}$).
- The variance of the prior is inversely proportional to $\theta$, which controls the “strength” of the prior.
Without Country Smoothing

Demographic Forecasting

Transportation Accidents
(males)
Sri Lanka
CXC
With Country Smoothing

Transportation Accidents (males) Sri Lanka

Demographic Forecasting
Formalizing Similarity

Standard Bayesian Approach

Assume coefficients are similar—but we know little about the coefficients.

Requires the same covariates in each cross-section.

- Why measure water quality in the U.S.?

Requires covariates with the same meaning in each cross-section.

- Does GDP mean the same thing in Botswana and the U.S.?

Imposes no assumptions on covariates or mortality—If covariates are dissimilar, then making coefficients similar makes mortality dissimilar (since $E(y_t) = X_t \beta$ in each cross-section).

Alternative Approach

Assume expected mortality is similar.

Coefficients are unobserved, mortality patterns are well known.

Different covariates allowed in each cross-section.

Covariates with the same name can have different meanings.

Demographic Forecasting
Formalizing Similarity

Standard Bayesian Approach

Assume coefficients are similar
— But we know little about the coefficients
Requires the same covariates in each cross-section
— Why measure water quality in the U.S.?
Requires covariates with the same meaning in each cross-section
— Does GDP mean the same thing in Botswana and the U.S.?
Imposes no assumptions on covariates or mortality
— If covariates are dissimilar, then making coefficients similar makes mortality dissimilar \( \text{since } E(y_t) = X_t \beta \text{ in each cross-section} \)

Alternative Approach

Assume expected mortality is similar
Coefficients are unobserved, mortality patterns are well known
Different covariates allowed in each cross-section
Covariates with the same name can have different meanings

Demographic Forecasting
Formalizing Similarity

Standard Bayesian Approach

- Assume *coefficients* are similar
Formalizing Similarity

Standard Bayesian Approach

- Assume coefficients are similar
  - But we know little about the coefficients

Alternative Approach

Assume expected mortality is similar
Coefficients are unobserved, mortality patterns are well known
Different covariates allowed in each cross-section
Covariates with the same name can have different meanings
Formalizing Similarity

Standard Bayesian Approach

- Assume **coefficients** are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section

**Alternative Approach**

- Assume expected mortality is similar
- Coefficients are unobserved, mortality patterns are well known
- Different covariates allowed in each cross-section
- Covariates with the same name can have different meanings
Formalizing Similarity

Standard Bayesian Approach

- Assume *coefficients* are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
Formalizing Similarity

Standard Bayesian Approach

- Assume *coefficients* are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
Formalizing Similarity

Standard Bayesian Approach

- **Assume coefficients** are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
Formalizing Similarity

Standard Bayesian Approach

- Assume *coefficients* are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
Formalizing Similarity

Standard Bayesian Approach

- Assume **coefficients** are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
  - If covariates are dissimilar, then making coefficients similar makes mortality dissimilar [since \( E(y_t) = X_t \beta \) in each cross-section]
Formalizing Similarity

**Standard Bayesian Approach**

- Assume *coefficients* are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
  - If covariates are dissimilar, then making coefficients similar makes mortality dissimilar [since $E(y_t) = X_t \beta$ in each cross-section]

**Alternative Approach**
Formalizing Similarity

Standard Bayesian Approach

- Assume coefficients are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
  - If covariates are dissimilar, then making coefficients similar makes mortality dissimilar [since \( E(y_t) = X_t\beta \) in each cross-section]

Alternative Approach

- Assume expected mortality is similar
Formalizing Similarity

Standard Bayesian Approach

- Assume **coefficients** are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
  - If covariates are dissimilar, then making coefficients similar makes mortality dissimilar [since $E(y_t) = X_t\beta$ in each cross-section]

Alternative Approach

- Assume **expected mortality** is similar
- Coefficients are unobserved, mortality patterns are well known
Formalizing Similarity

**Standard Bayesian Approach**
- Assume *coefficients* are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
  - If covariates are dissimilar, then making coefficients similar makes mortality dissimilar [since $E(y_t) = X_t\beta$ in each cross-section]

**Alternative Approach**
- Assume *expected mortality* is similar
- Coefficients are unobserved, mortality patterns are well known
- Different covariates allowed in each cross-section
Formalizing Similarity

Standard Bayesian Approach

- Assume *coefficients* are similar
  - But we know little about the coefficients
- Requires the same covariates in each cross-section
  - Why measure water quality in the U.S.?
- Requires covariates with the same meaning in each cross-section
  - Does GDP mean the same thing in Botswana and the U.S.?
- Imposes no assumptions on covariates or mortality
  - If covariates are dissimilar, then making coefficients similar makes mortality dissimilar [since \( E(y_t) = X_t \beta \) in each cross-section]

Alternative Approach

- Assume *expected mortality* is similar
- Coefficients are unobserved, mortality patterns are well known
- Different covariates allowed in each cross-section
- Covariates with the same name can have different meanings
Many Short Time Series

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

- % of world countries
- % of world population

# Observations

% of world countries

% of world population

Demographic Forecasting
## Preview of Results: Out-of-Sample Evaluation

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

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

| Category               | Best Previous | Our Method | Best Conceivable | % Improvement
|------------------------|---------------|------------|------------------|----------------
|                        |               |            |                  | 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.
<table>
<thead>
<tr>
<th>Category</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).
## Preview of Results: Out-of-Sample Evaluation

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

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean Absolute Error</th>
<th>% Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best Previous</td>
<td>Our Method</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>0.36</td>
<td>0.27</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>Respiratory Chronic</td>
<td>0.45</td>
<td>0.39</td>
</tr>
<tr>
<td>Other Infectious</td>
<td>0.55</td>
<td>0.48</td>
</tr>
<tr>
<td>Stomach Cancer</td>
<td>0.30</td>
<td>0.27</td>
</tr>
<tr>
<td>All-Cause</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td>Suicide</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>Respiratory Infectious</td>
<td>0.49</td>
<td>0.47</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.
### Preview of Results: Out-of-Sample Evaluation

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

<table>
<thead>
<tr>
<th></th>
<th>Mean Absolute Error</th>
<th>% Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best Previous</td>
<td>Our Method</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>0.34</td>
<td>0.27</td>
</tr>
<tr>
<td>Lung Cancer</td>
<td>0.36</td>
<td>0.27</td>
</tr>
<tr>
<td>Transportation</td>
<td>0.37</td>
<td>0.31</td>
</tr>
<tr>
<td>Respiratory Chronic</td>
<td>0.45</td>
<td>0.39</td>
</tr>
<tr>
<td>Other Infectious</td>
<td>0.55</td>
<td>0.48</td>
</tr>
<tr>
<td>Stomach Cancer</td>
<td>0.30</td>
<td>0.27</td>
</tr>
<tr>
<td>All-Cause</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td>Suicide</td>
<td>0.31</td>
<td>0.29</td>
</tr>
<tr>
<td>Respiratory Infectious</td>
<td>0.49</td>
<td>0.47</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.
### Preview of Results: Out-of-Sample Evaluation

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

<table>
<thead>
<tr>
<th></th>
<th>Mean Absolute Error</th>
<th>% Improvement</th>
</tr>
</thead>
</table>
|                          | 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.