Population forecasting and the importance of being uncertain

Rob J Hyndman

MONASH University

2007 Knibbs Lecture

Sir George Handley Knibbs (1858–1929)

- President of the Institute of Surveyors, 1892–1901
- President of the British Astronomical Society (NSW), 1897–1898
- President of the Royal Society of NSW, 1898–1899
- Lecturer in Geodesy, Astronomy and Hydraulics, Uni of Sydney, 1889–1905
- Director of Technical Education, NSW, 1905
- Professor of Physics, Uni of Sydney, 1905.
- Commonwealth Statistician, 1906–1921
- Director of Institute of Science and Industry, 1921–1926

Outline

Outline

What is it?

What is it?

Clay model of sheep's liver

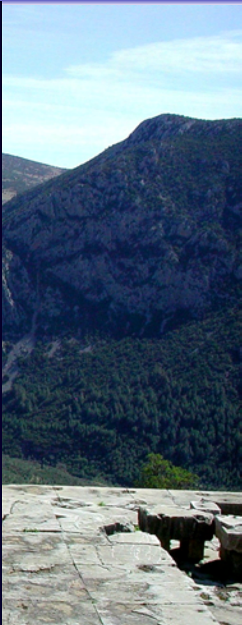
Used by Babylonian forecasters approximately 600 B.C.

Now in British Museum.

Delphic oracle

Delphic oracle

Delphic oracle



Delphic oracle

Vagrant forecasters

The British Vagrancy Act (1736) made it an offence to defraud by charging money for predictions.

Vagrant forecasters

The British Vagrancy Act (1736) made it an offence to defraud by charging money for predictions.

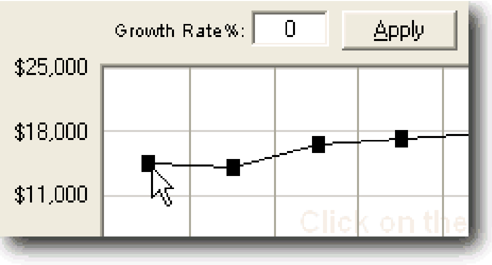
Punishment: a fine or three months' imprisonment with hard labour.

Outline

Standard business practice today

Graphic Forecaster

Create forecasts visually with a "drag and drop" graphic forecaster. The Graphic Forecaster is a simple and powerful tool to streamline the forecasting process. You can change your sales and expenses estimates by simply clicking your mouse button to move the line on your forecast chart or apply a specific growth rate to the whole year. Build forecasts using visual common sense.



View More Detail

Standard business practice today

Budget Maestro by Centage

Budget Maestro by Centage

[Click here for a free demo](#) FREE!

Application: Business Intelligence and Analytics

Price Range: Solutions start at $5K

A huge advance over spreadsheet-based systems, Budget Maestro is a complete solution for budgeting, forecasting, what-if scenario planning, reporting and analysis. Budget Maestro takes the pain out of the budgeting process (no tedious data entry and formula verification) while providing you a tool to more accurately analyze and measure business performance and profitability. Budget Maestro's capabilities include:

Budgeting and Forecasting: Budget Maestro utilizes database technology for real-time data collection and reporting. A common interface for all users fosters collaboration and increases the accuracy of data entry. There are no formulas or macros to create, no tedious re-keying of data and no mystery links to chase down and fix. Budget Maestro's built-in "financial intelligence and business rules" builds the formulas for you ensuring 100% accuracy.

Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.

Standard business practice today

- "What-if scenarios" based on assumed and fixed future conditions.
- Highly subjective.

Standard business practice today

- "What-if scenarios" based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.

Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.

Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.

Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.
- Largely guesswork.

Standard business practice today

- "What-if scenarios" based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.
- Largely guesswork.

Standard business practice today

- “What-if scenarios” based on assumed and fixed future conditions.
- Highly subjective.
- Not replicable or testable.
- No possible way of quantifying probabilistic uncertainty.
- Lack of uncertainty statements leads to false sense of accuracy.
- Largely guesswork.

Is this any better than a sheep’s liver or hallucinogens?

ABS population projections

The Australian Bureau of Statistics provide population "projections".

"The projections are not intended as predictions or forecasts, but are illustrations of growth and change in the population that would occur if assumptions made about future demographic trends were to prevail over the projection period.
While the assumptions are formulated on the basis of an assessment of past demographic trends, both in Australia and overseas, there is no certainty that any of the assumptions will be realised. In addition, no assessment has been made of changes in non-demographic conditions."

ABS 3222.0 - Population Projections, Australia, 2004 to 2101

ABS population projections

The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates

ABS population projections

The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates
- No objectivity.

ABS population projections

The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed

ABS population projections

The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- Arbitrary adjustments to some rates

ABS population projections

The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- Arbitrary adjustments to some rates
- No probabilistic basis.

ABS population projections

The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

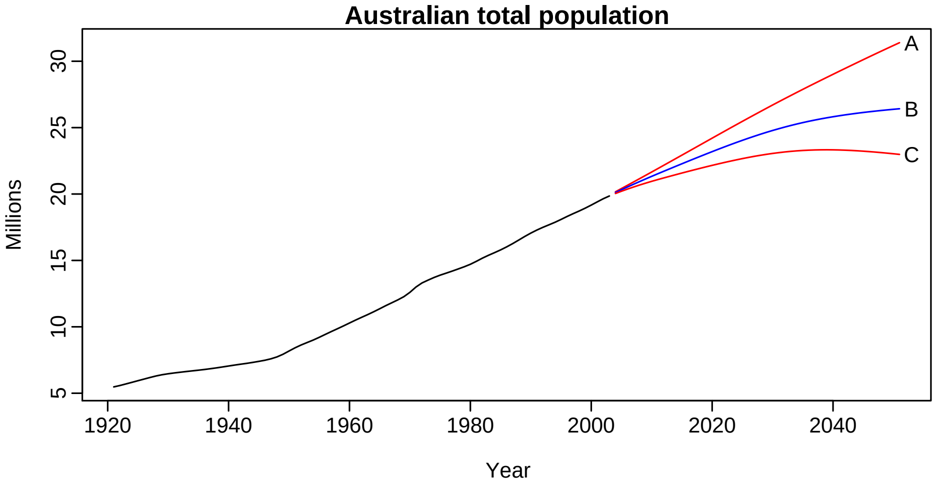
- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- Arbitrary adjustments to some rates
- No probabilistic basis.
- Not prediction intervals.

ABS population projections

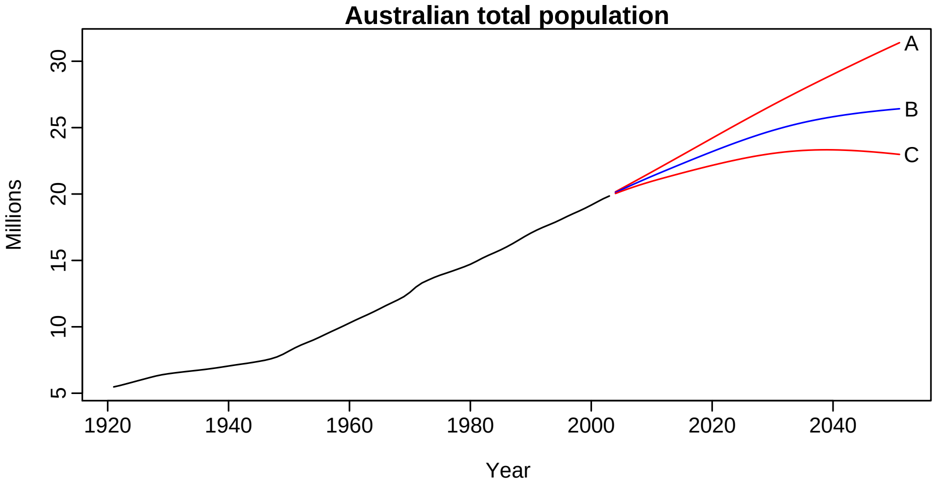
The ABS provides three projection scenarios labelled “High”, “Medium” and “Low”.

- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- Arbitrary adjustments to some rates
- No probabilistic basis.
- Not prediction intervals.
- Most users use the “Medium” projection, but it is unrelated to the mean, median or mode of the future distribution.

ABS population projections



ABS population projections



What do these projections mean?

Advantages of stochastic models

- Based on empirical data

Advantages of stochastic models

- Based on empirical data
- Computable

Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable

Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable
- Testable

Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable
- Testable
- Objective measure of uncertainty

Advantages of stochastic models

- Based on empirical data
- Computable
- Replicable
- Testable
- Objective measure of uncertainty
- Able to compute prediction intervals

Outline

Exponential smoothing

Exponential smoothing is extremely popular, simple to implement, and performs well in forecasting competitions.

Exponential smoothing

Exponential smoothing is extremely popular, simple to implement, and performs well in forecasting competitions.

"Unfortunately, exponential smoothing methods do not allow easy calculation of prediction intervals."

Makridakis, Wheelwright and Hyndman, p.177.
(Wiley, 3rd ed., 1998)

Exponential smoothing

Since 2002. . .

- a general class of state space models proposed underlying all the common exponential smoothing methods.
- analytical results for prediction intervals.
- likelihood calculation for estimation.
- AIC for model selection.
- an algorithm for automatic forecasting using the new class of models.
- new results on the admissible parameter space.

Exponential smoothing

- enables easy calculation of the likelihood

Exponential smoothing

- enables easy calculation of the likelihood
- provides facilities to compute (analytical) prediction intervals

Exponential smoothing

- enables easy calculation of the likelihood
- provides facilities to compute (analytical) prediction intervals
- two possible state space models for each method (additive error and multiplicative error). Equivalent point forecasts, different prediction intervals.

Exponential smoothing

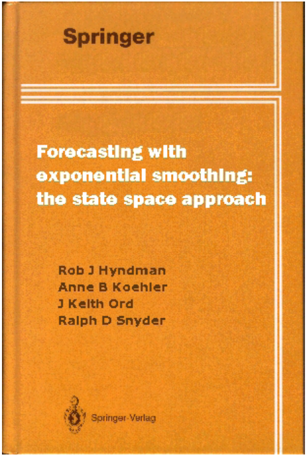
- enables easy calculation of the likelihood
- provides facilities to compute (analytical) prediction intervals
- two possible state space models for each method (additive error and multiplicative error). Equivalent point forecasts, different prediction intervals.

Implementation

This methodology is available as

- An R package (**forecast**) on CRAN
- An Excel add-in (**PhiCast**)

Exponential smoothing



New book due out in 2008

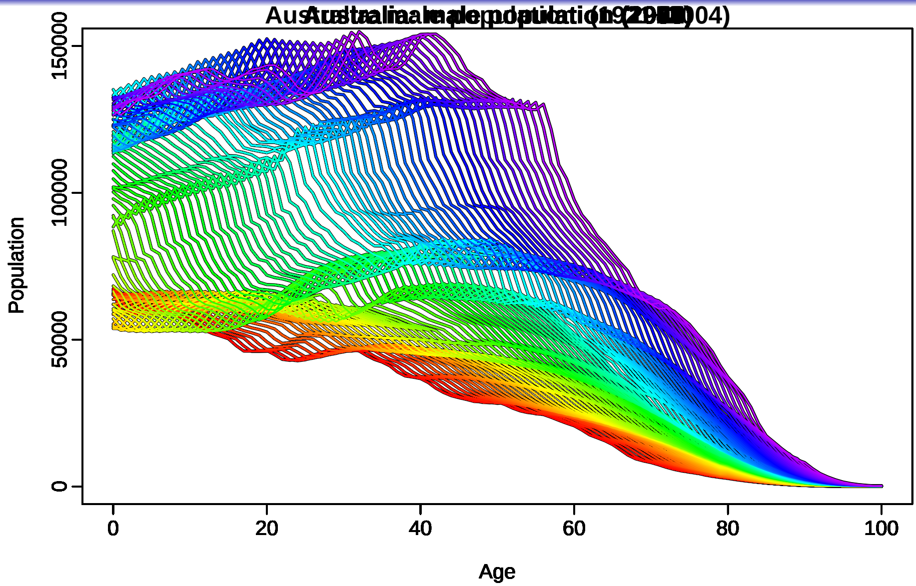
Forecasting with exponential smoothing: the state space approach

Rob J. Hyndman
Anne B. Koehler
J. Keith Ord
Ralph D. Snyder

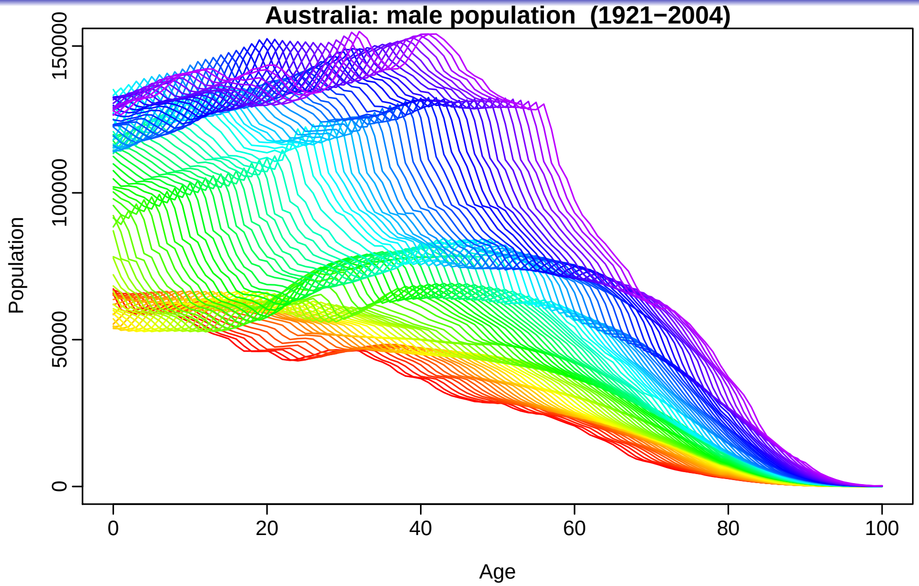
Published by Springer

Outline

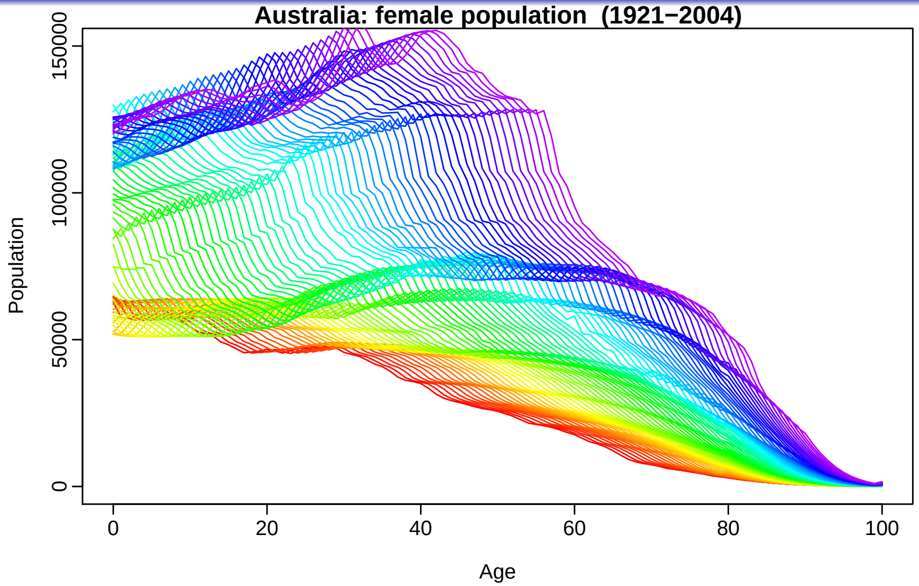
Annual age-specific population



Annual age-specific population



Annual age-specific population



Stochastic population forecasts

Key ideas

- Population is a function of **mortality**, **fertility** and **net migration**.

Stochastic population forecasts

Key ideas

- Population is a function of **mortality**, **fertility** and **net migration**.
- Build an age-sex-specific **stochastic model** for each of mortality, fertility & net migration.

Stochastic population forecasts

Key ideas

- Population is a function of **mortality**, **fertility** and **net migration**.
- Build an age-sex-specific **stochastic model** for each of mortality, fertility & net migration.
- Use the models to **simulate future sample paths** of all components.

Stochastic population forecasts

Key ideas

- Population is a function of **mortality**, **fertility** and **net migration**.
- Build an age-sex-specific **stochastic model** for each of mortality, fertility & net migration.
- Use the models to **simulate future sample paths** of all components.
- Compute future births, deaths, net migrants and populations from simulated rates.

Stochastic population forecasts

Key ideas

- Population is a function of **mortality**, **fertility** and **net migration**.
- Build an age-sex-specific **stochastic model** for each of mortality, fertility & net migration.
- Use the models to **simulate future sample paths** of all components.
- Compute future births, deaths, net migrants and populations from simulated rates.
- Combine the results to get **age-specific stochastic population forecasts**.

Demographic growth-balance equation

Demographic growth-balance equation

$$
\begin{aligned}
P_{t+1}(x+1) &= P_t(x) - D_t(x,x+1) + G_t(x,x+1) \\
P_{t+1}(0) &= B_t \qquad - D_t(B,0) \qquad + G_t(B,0)
\end{aligned}
$$

$x = 0, 1, 2, \ldots$.

Demographic growth-balance equation

Demographic growth-balance equation

$$P_{t+1}(x+1) = P_t(x) - D_t(x,x+1) + G_t(x,x+1)$$
$$P_{t+1}(0) = B_t \qquad - D_t(B,0) \qquad + G_t(B,0)$$

$x = 0, 1, 2, \dots.$

$P_t(x) =$ population of age x at 1 January, year t

$B_t =$ births in calendar year t

$D_t(x,x+1) =$ deaths in calendar year t of persons aged x at the beginning of year t

$D_t(B,0) =$ deaths in calendar year t of persons born in year t

$G_t(x,x+1) =$ net migrants in calendar year t of persons aged x at the beginning of year t

$G_t(B,0) =$ net migrants of infants born in calendar year t

The available data

The following data are available:

$P_t(x) =$ **population** of age x at 1 January, year t

$E_t(x) =$ **population** of age x at 30 June, year t

$B_t(x) =$ **births** in calendar year t to females of age x

$D_t(x) =$ **deaths** in calendar year t of persons of age x

The available data

The following data are available:

$P_t(x) =$ **population** of age x at 1 January, year t

$E_t(x) =$ **population** of age x at 30 June, year t

$B_t(x) =$ **births** in calendar year t to females of age x

$D_t(x) =$ **deaths** in calendar year t of persons of age x

From these, we can estimate:

- $m_t(x) = D_t(x)/E_t(x) =$ central death rate in calendar year t;

The available data

The following data are available:

$P_t(x) =$ **population** of age x at 1 January, year t

$E_t(x) =$ **population** of age x at 30 June, year t

$B_t(x) =$ **births** in calendar year t to females of age x

$D_t(x) =$ **deaths** in calendar year t of persons of age x

From these, we can estimate:

- $m_t(x) = D_t(x)/E_t(x) =$ central death rate in calendar year t;
- $f_t(x) = B_t(x)/E_t^F(x) =$ fertility rate for females of age x in calendar year t.

The available data

The following data are available:

$P_t(x) =$ **population** of age x at 1 January, year t

$E_t(x) =$ **population** of age x at 30 June, year t

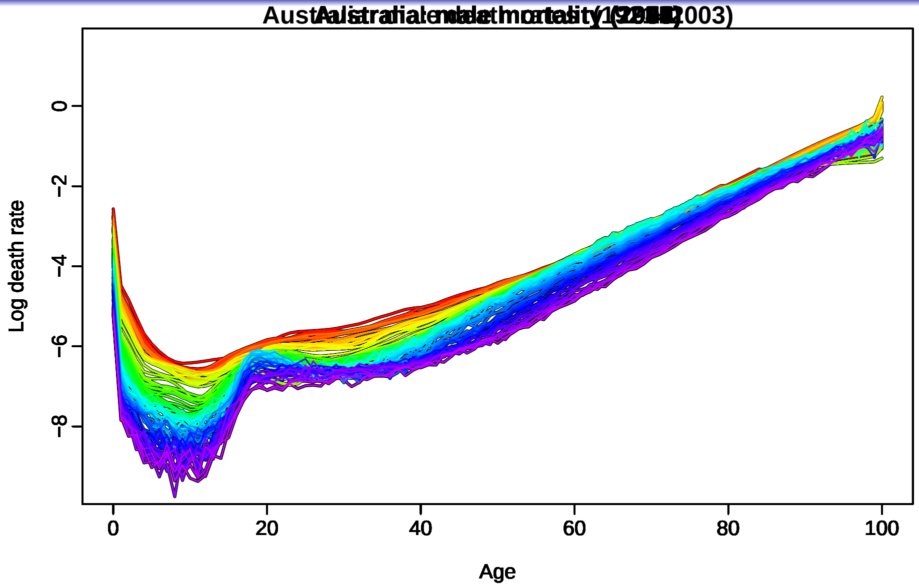
$B_t(x) =$ **births** in calendar year t to females of age x

$D_t(x) =$ **deaths** in calendar year t of persons of age x

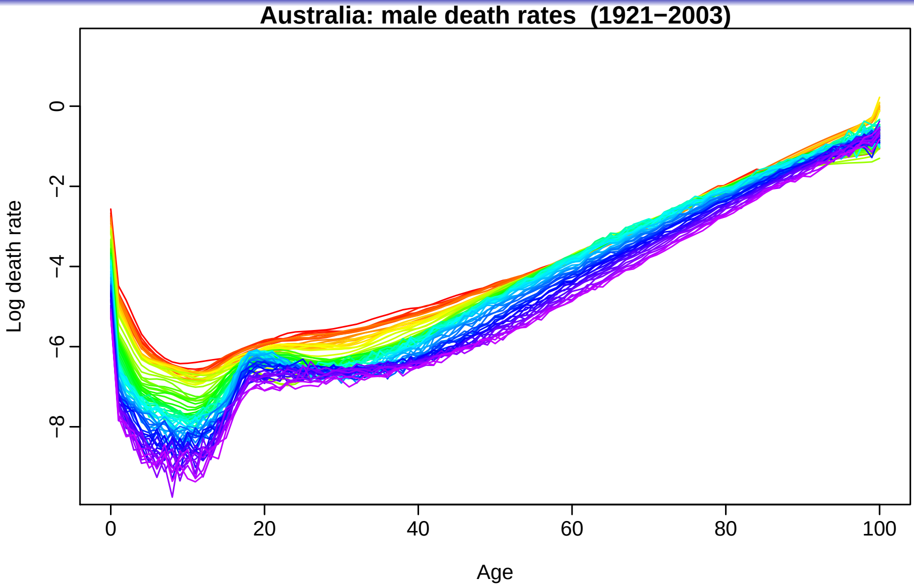
From these, we can estimate:

- $m_t(x) = D_t(x)/E_t(x) =$ central death rate in calendar year t;
- $f_t(x) = B_t(x)/E_t^F(x) =$ fertility rate for females of age x in calendar year t.
- $D_t(x, x+1)$ and $D_t(B, 0)$.

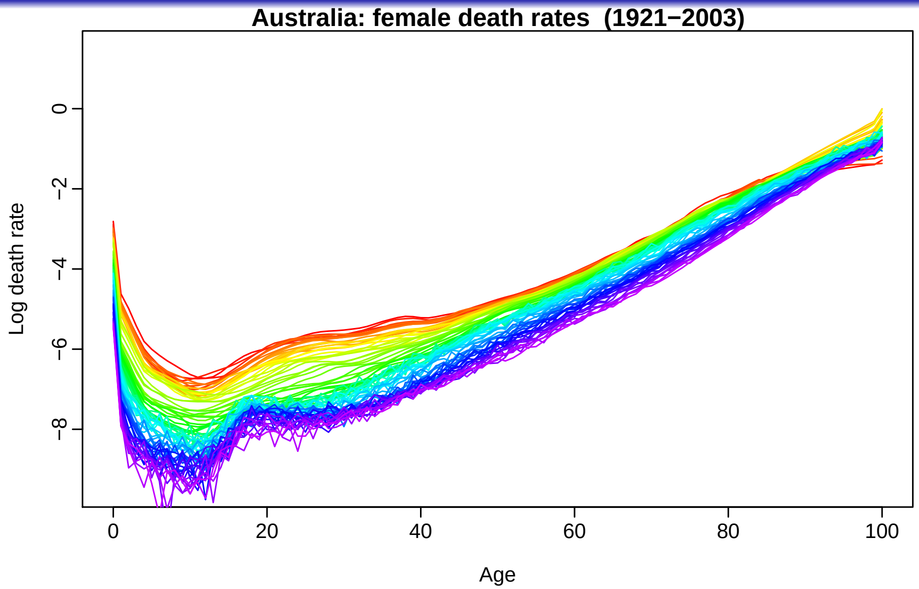
Mortality rates



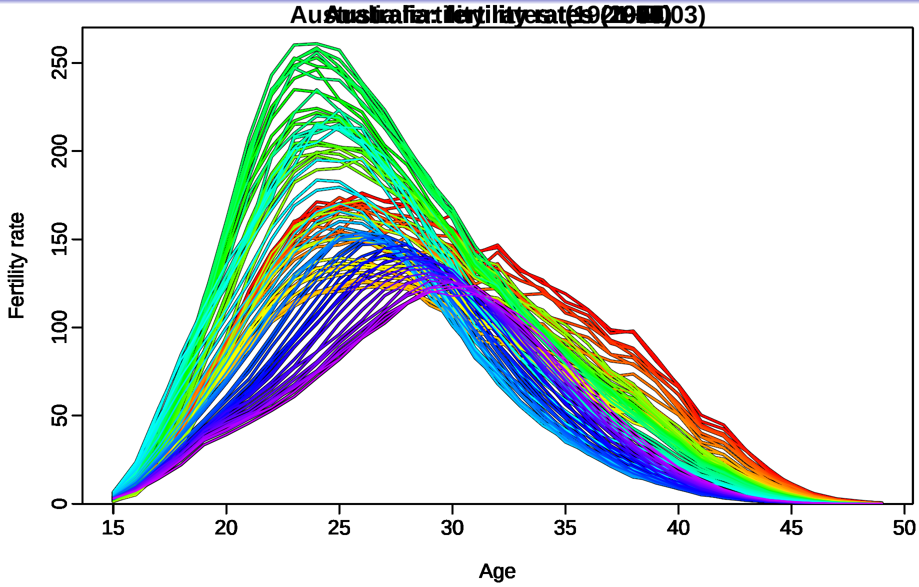
Mortality rates



Mortality rates



Fertility rates



Net migration

We need to *estimate* **migration** data based on difference in population numbers after adjusting for births and deaths.

Net migration

We need to *estimate* **migration** data based on difference in population numbers after adjusting for births and deaths.

Demographic growth-balance equation

$$G_t(x,x+1) = P_{t+1}(x+1) - P_t(x) + D_t(x,x+1)$$
$$G_t(B,0) = P_{t+1}(0) \qquad - B_t \qquad + D_t(B,0)$$

$x = 0, 1, 2, \ldots.$

Net migration

We need to *estimate* **migration** data based on difference in population numbers after adjusting for births and deaths.

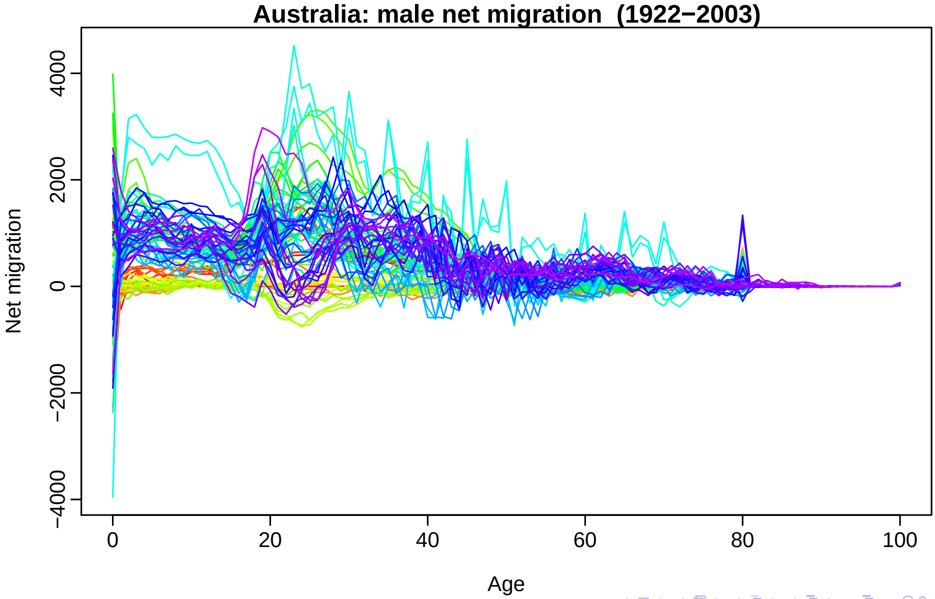
Demographic growth-balance equation

$$G_t(x, x+1) = P_{t+1}(x+1) - P_t(x) + D_t(x, x+1)$$
$$G_t(B, 0) = P_{t+1}(0) \qquad - B_t \qquad + D_t(B, 0)$$

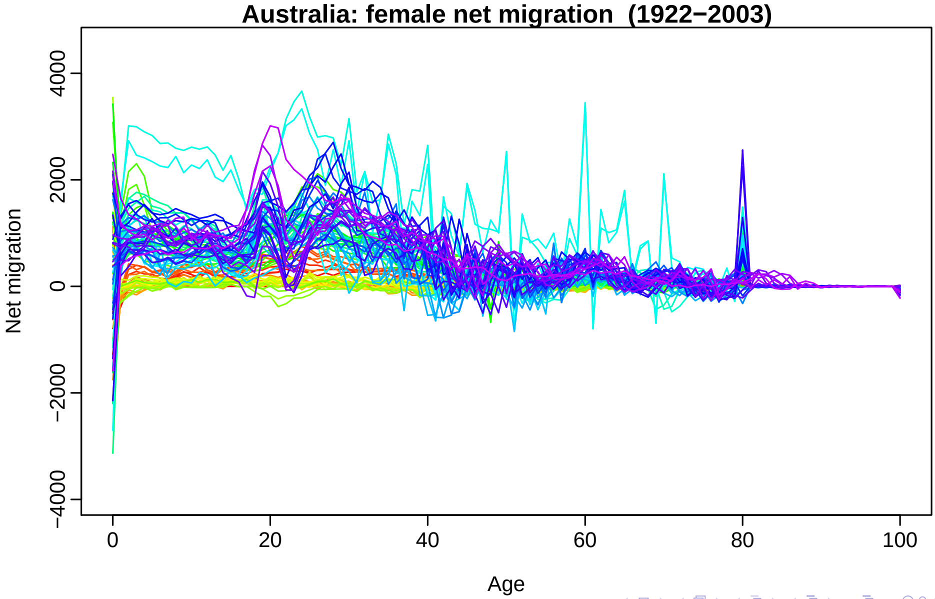
$x = 0, 1, 2, \dots.$

Note: "net migration" numbers also include **errors** associated with all estimates. i.e., a "residual".

Net migration



Net migration



Some notation

Let $y_t(x_i)$ be the observed data (mortality, fertility or net migration) in period t at age x_i, $i = 1, \dots, p$, $t = 1, \dots, n$.

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $y_t(x)$ transformed first if necessary

Some notation

Let $y_t(x_i)$ be the observed data (mortality, fertility or net migration) in period t at age x_i, $i = 1, \dots, p$, $t = 1, \dots, n$.

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $y_t(x)$ transformed first if necessary
- $\varepsilon_{t,i} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$

Some notation

Let $y_t(x_i)$ be the observed data (mortality, fertility or net migration) in period t at age x_i, $i = 1, \dots, p$, $t = 1, \dots, n$.

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $y_t(x)$ transformed first if necessary
- $\varepsilon_{t,i} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$
- $s_t(x)$ and $\sigma_t(x)$ are smooth functions of x.

Some notation

Let $y_t(x_i)$ be the observed data (mortality, fertility or net migration) in period t at age x_i, $i = 1, \dots, p$, $t = 1, \dots, n$.

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $y_t(x)$ transformed first if necessary
- $\varepsilon_{t,i} \overset{\text{iid}}{\sim} \mathrm{N}(0, 1)$
- $s_t(x)$ and $\sigma_t(x)$ are smooth functions of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_p$.

Some notation

Let $y_t(x_i)$ be the observed data (mortality, fertility or net migration) in period t at age x_i, $i = 1, \dots, p$, $t = 1, \dots, n$.

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $y_t(x)$ transformed first if necessary
- $\varepsilon_{t,i} \overset{\text{iid}}{\sim} \text{N}(0, 1)$
- $s_t(x)$ and $\sigma_t(x)$ are smooth functions of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_p$.
- We want to forecast **whole curve** $y_t(x)$ for $t = n + 1, \dots, n + h$.

Functional time series model

$$\begin{aligned} y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

Functional time series model

$$\begin{aligned} y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.

Functional time series model

$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $s_t(x)$ across years.

Functional time series model

$$\begin{aligned} y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $s_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.

Functional time series model

$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $s_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.
4. Forecast $\beta_{t,k}$ using exponential smoothing.

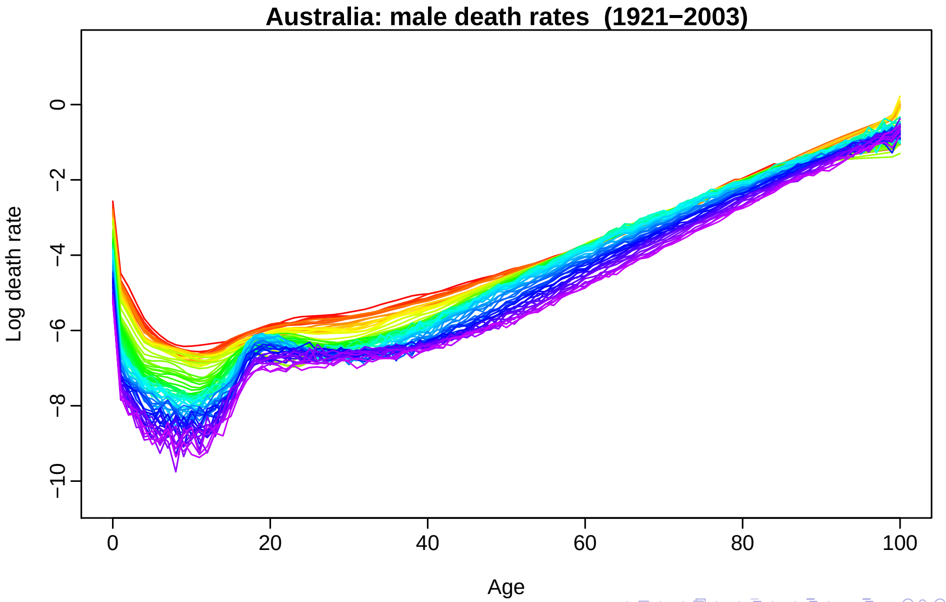
Functional time series model

$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

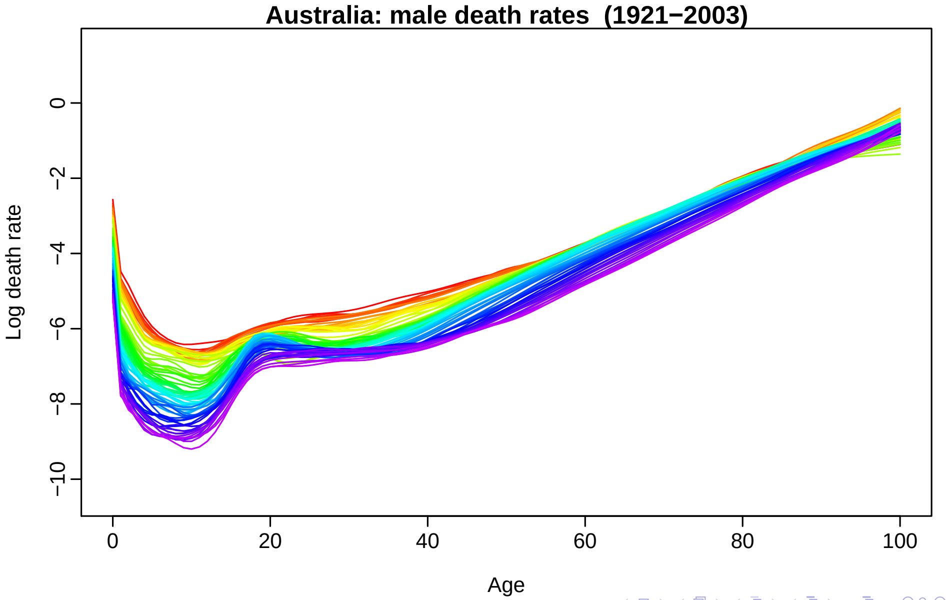
where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \text{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \text{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $s_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.
4. Forecast $\beta_{t,k}$ using exponential smoothing.
5. Put it all together to get forecasts of $y_t(x)$.

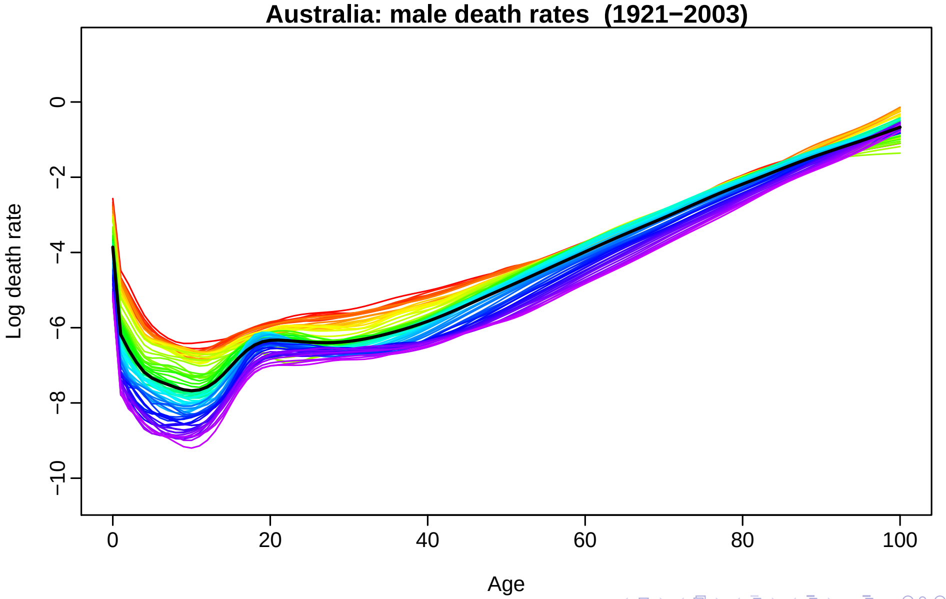
Functional time series model



Functional time series model



Functional time series model



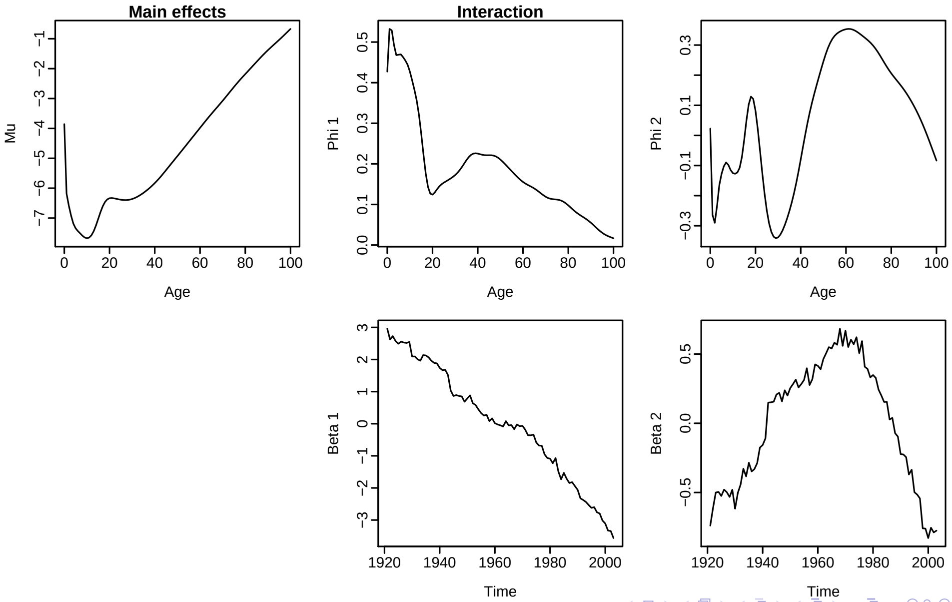
Functional time series model

$$\begin{aligned} y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

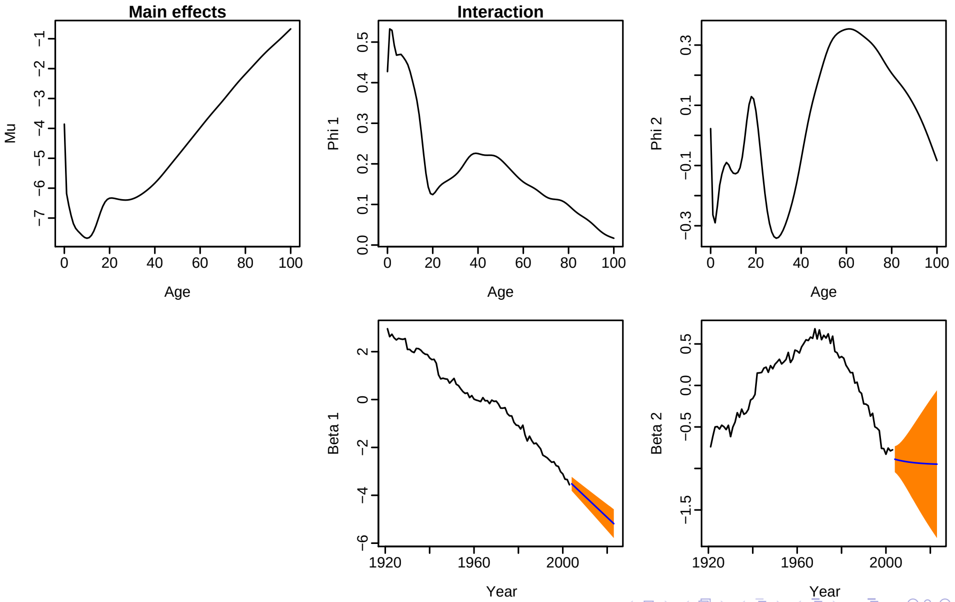
where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $s_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.
4. Forecast $\beta_{t,k}$ using exponential smoothing.
5. Put it all together to get forecasts of $y_t(x)$.

Functional time series model



Functional time series model



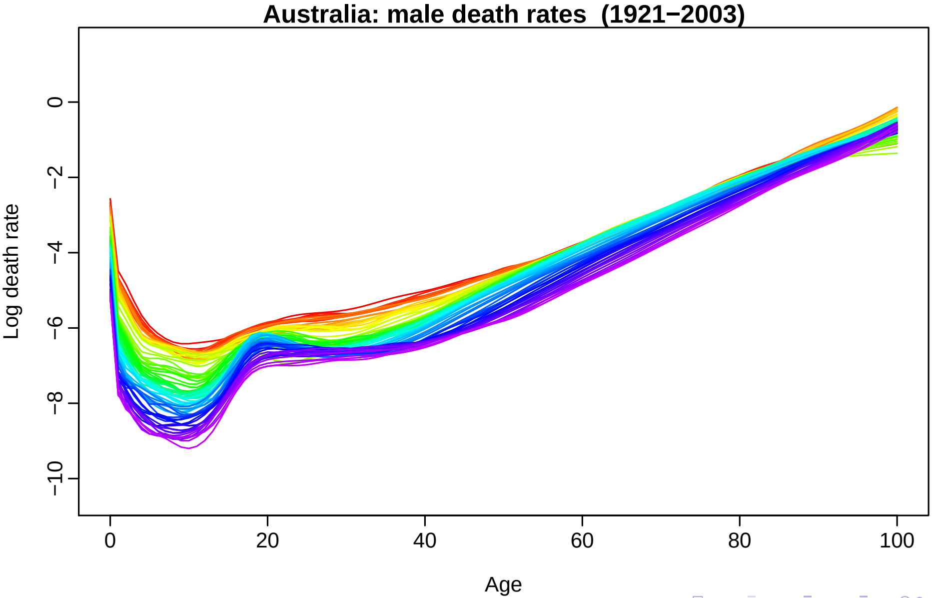
Functional time series model

$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

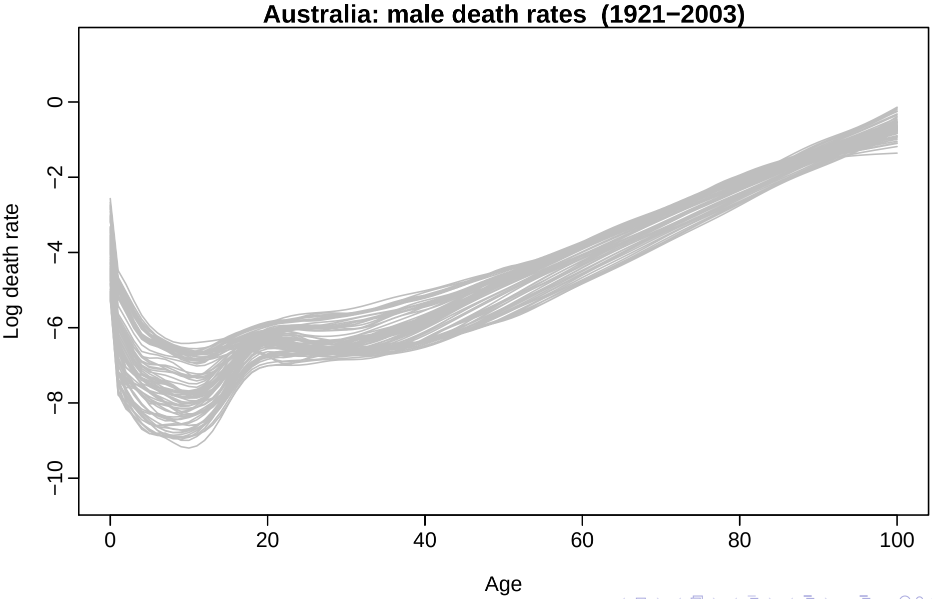
where $\varepsilon_{t,x} \overset{\text{iid}}{\sim} \mathrm{N}(0,1)$ and $e_t(x) \overset{\text{iid}}{\sim} \mathrm{N}(0, v(x))$.

1. Estimate smooth functions $s_t(x)$ using nonparametric regression.
2. Estimate $\mu(x)$ as mean $s_t(x)$ across years.
3. Estimate $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.
4. Forecast $\beta_{t,k}$ using exponential smoothing.
5. Put it all together to get forecasts of $y_t(x)$.

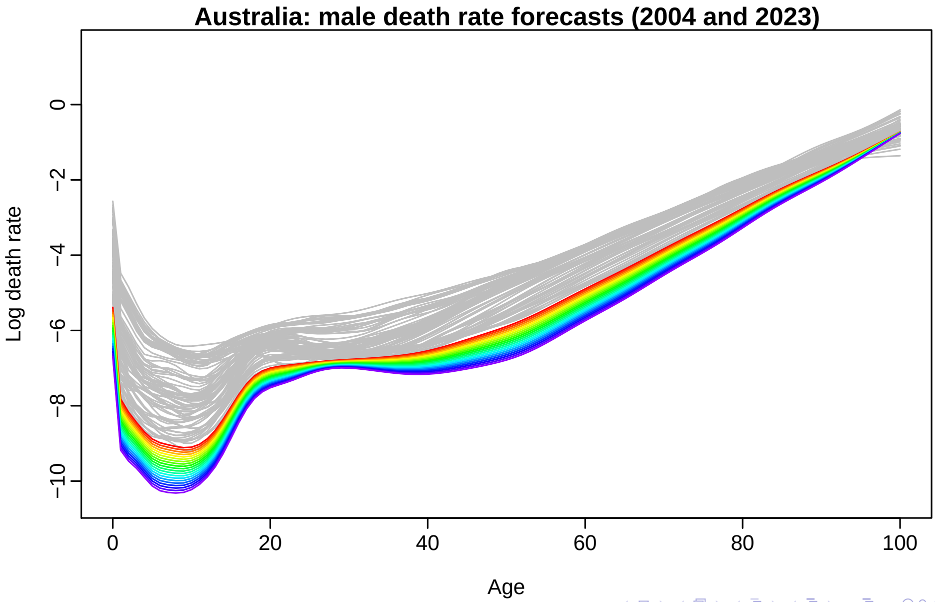
Functional time series model



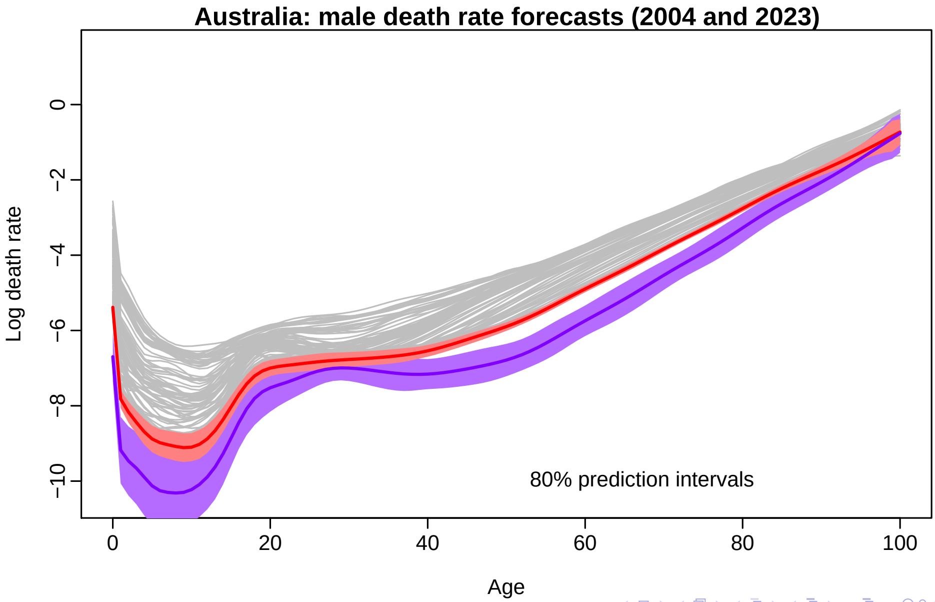
Functional time series model



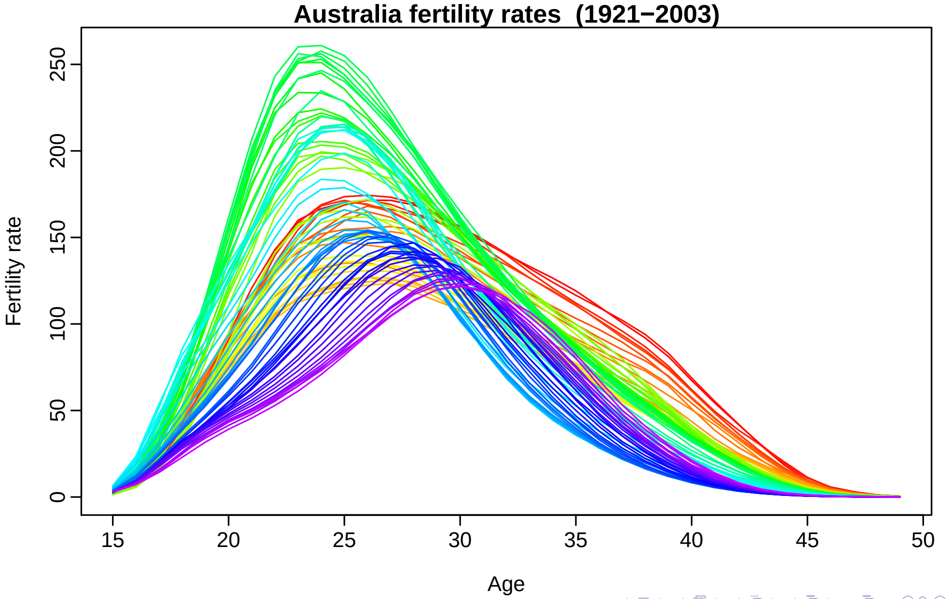
Functional time series model



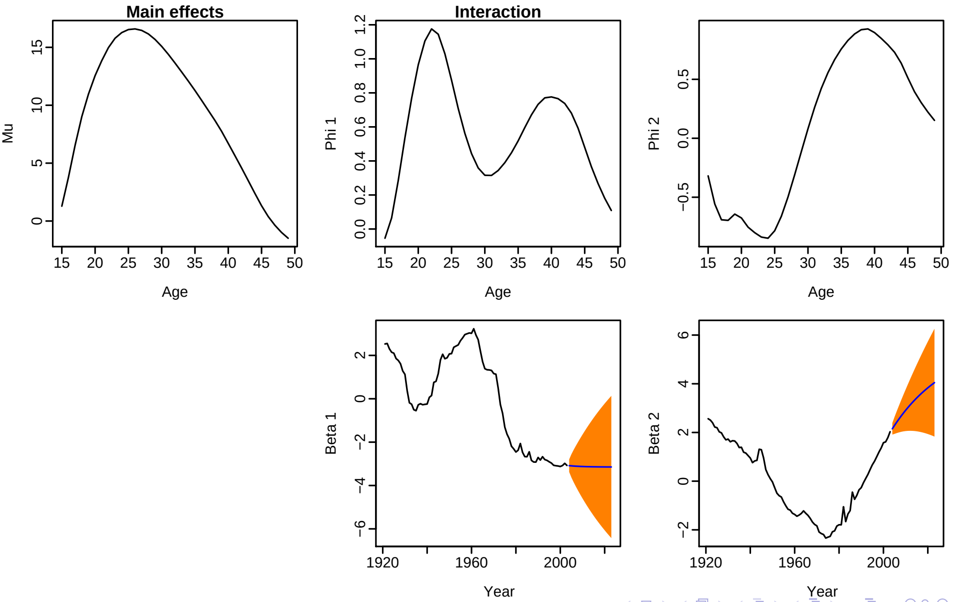
Functional time series model



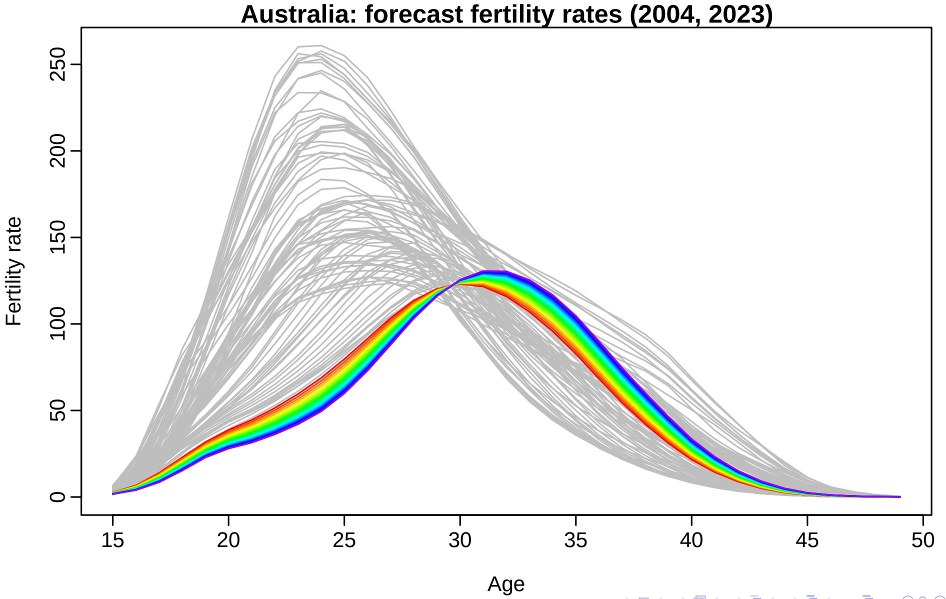
Fertility


Australia fertility rates (1921−2003)
Fertility rate
Age
0
50
100
150
200
250
15
20
25
30
35
40
45
50

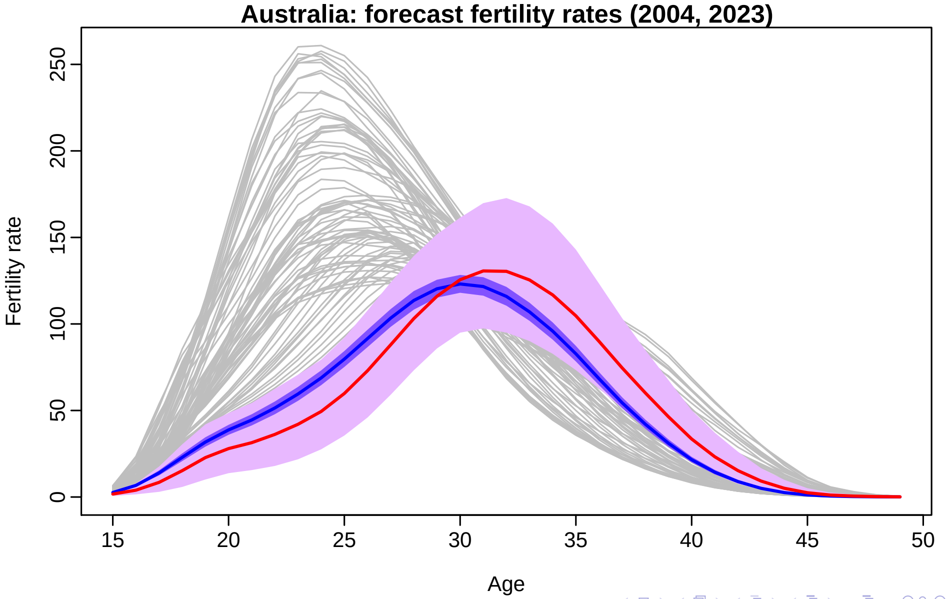
Fertility


Main effects
Mu
Age
Interaction
Phi 1
Age
Phi 2
Age
Beta 1
Year
Beta 2
Year

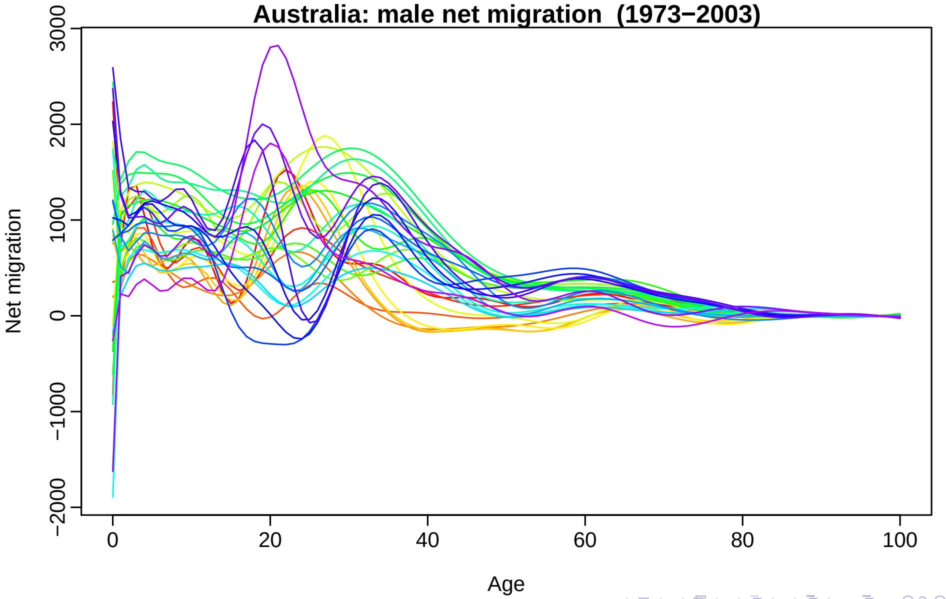
Fertility


Australia: forecast fertility rates (2004, 2023)
Fertility rate
0
50
100
150
200
250
Age
15
20
25
30
35
40
45
50

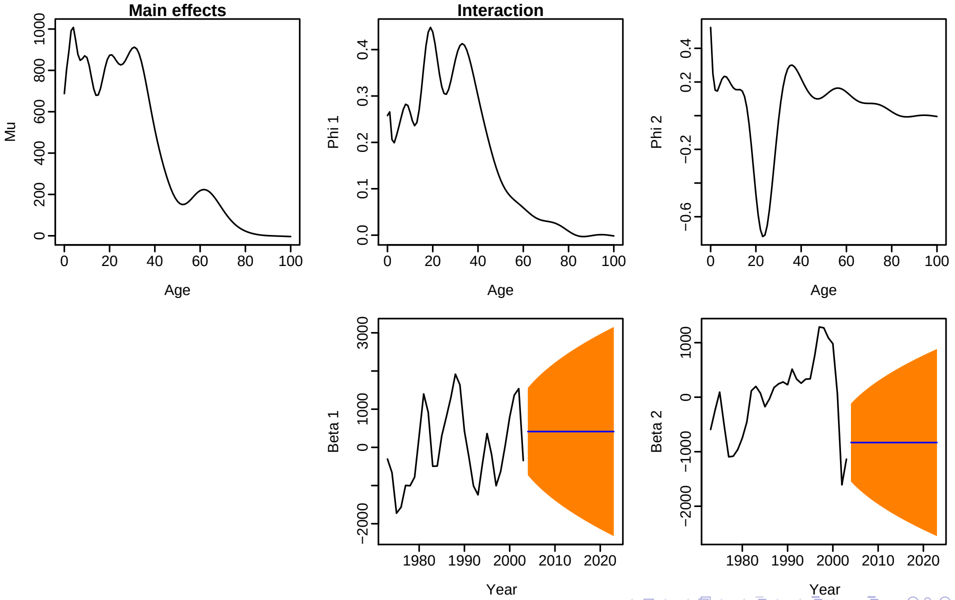
Fertility


Australia: forecast fertility rates (2004, 2023)
Fertility rate
0
50
100
150
200
250
Age
15
20
25
30
35
40
45
50

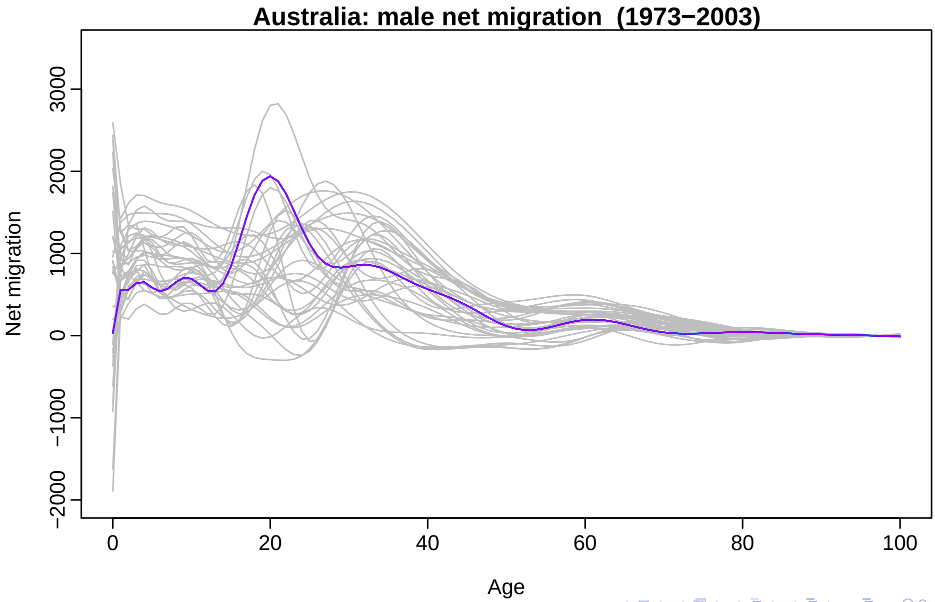
Migration: male



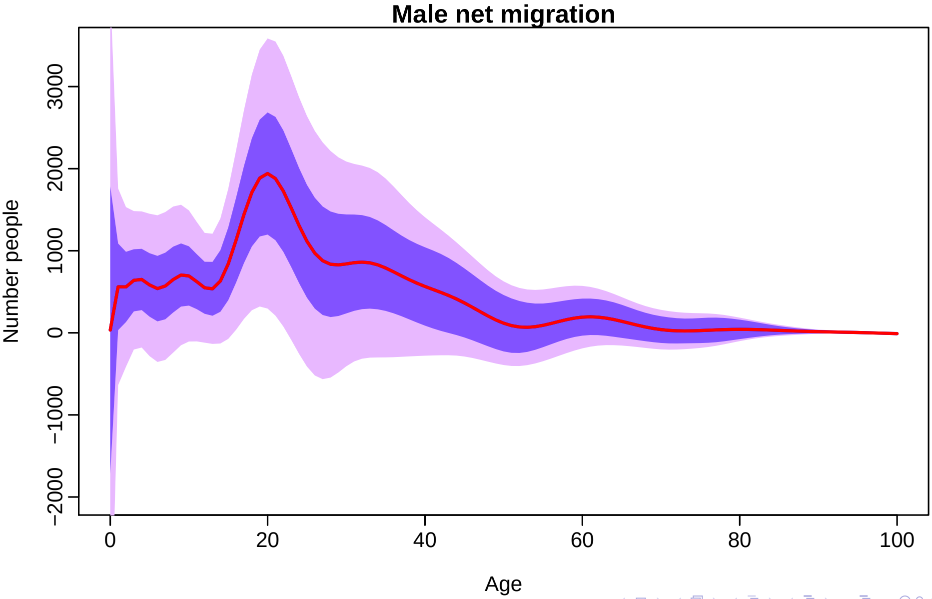
Migration: male



Migration: male



Migration: male



Simulation

$$\begin{aligned} y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

- For each of $m_t^F(x)$, $m_t^M(x)$, $f_t(x)$, $G_t^F(x, x+1)$, and $G_t^M(x, x+1)$:

Simulation

$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

- For each of $m_t^F(x)$, $m_t^M(x)$, $f_t(x)$, $G_t^F(x,x+1)$, and $G_t^M(x,x+1)$:
  - Generate random sample paths of $\beta_{t,k}$ for $t = n+1,\dots,n+h$ conditional on $\beta_{1,k},\dots,\beta_{n,k}$.

Simulation

$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

- For each of $m_t^F(x)$, $m_t^M(x)$, $f_t(x)$, $G_t^F(x,x+1)$, and $G_t^M(x,x+1)$:
  - Generate random sample paths of $\beta_{t,k}$ for $t = n+1,\dots,n+h$ conditional on $\beta_{1,k},\dots,\beta_{n,k}$.
  - Generate random values for $e_t(x)$ and $\varepsilon_{t,x}$.

Simulation

$$\begin{aligned} y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\ s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x) \end{aligned}$$

- For each of $m_t^F(x)$, $m_t^M(x)$, $f_t(x)$, $G_t^F(x, x+1)$, and $G_t^M(x, x+1)$:
  - Generate random sample paths of $\beta_{t,k}$ for $t = n+1, \dots, n+h$ conditional on $\beta_{1,k}, \dots, \beta_{n,k}$.
  - Generate random values for $e_t(x)$ and $\varepsilon_{t,x}$.
- Use simulated rates to generate $B_t(x)$, $D_t^F(x, x+1)$, $D_t^M(x, x+1)$ for $t = n+1, \dots, n+h$, assuming deaths and births are Poisson.

Simulation

Demographic growth-balance equation used to get population sample paths.

Demographic growth-balance equation

$$
\begin{aligned}
P_{t+1}(x+1) &= P_t(x) - D_t(x,x+1) + G_t(x,x+1) \\
P_{t+1}(0) &= B_t \qquad - D_t(B,0) \qquad\quad + G_t(B,0)
\end{aligned}
$$

$x = 0, 1, 2, \ldots.$

Simulation

Demographic growth-balance equation used to get population sample paths.

Demographic growth-balance equation

$$P_{t+1}(x+1) = P_t(x) - D_t(x, x+1) + G_t(x, x+1)$$
$$P_{t+1}(0) = B_t \quad - D_t(B, 0) \quad + G_t(B, 0)$$

$x = 0, 1, 2, \ldots.$

- 10000 sample paths of population $P_t(x)$, deaths $D_t(x)$ and births $B_t(x)$ generated for $t = 2005, \ldots, 2024$ and $x = 0, 1, 2, \ldots,$.

Simulation

Demographic growth-balance equation used to get population sample paths.

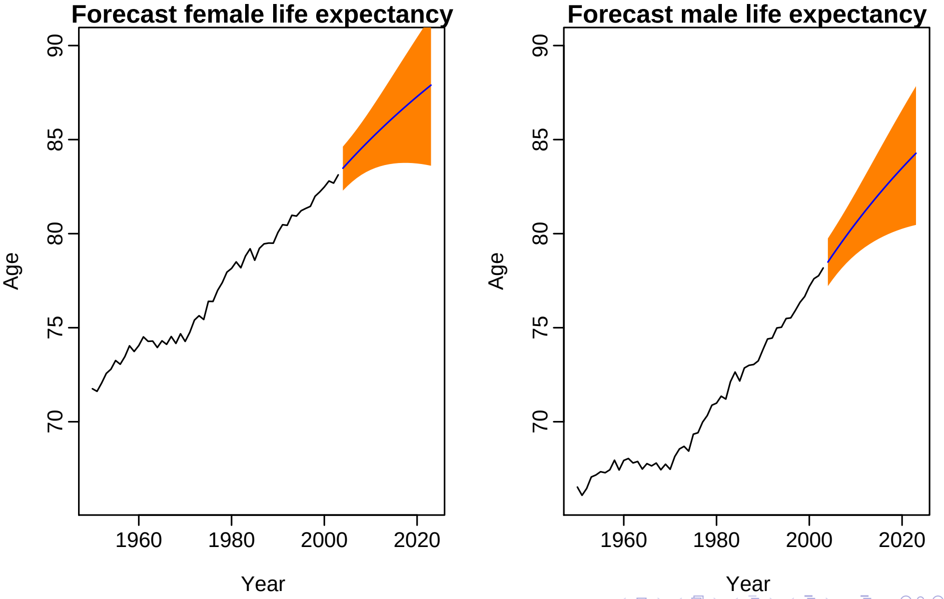
Demographic growth-balance equation

$$
\begin{aligned}
P_{t+1}(x+1) &= P_t(x) - D_t(x,x+1) + G_t(x,x+1) \\
P_{t+1}(0) &= B_t \qquad - D_t(B,0) \qquad + G_t(B,0)
\end{aligned}
$$

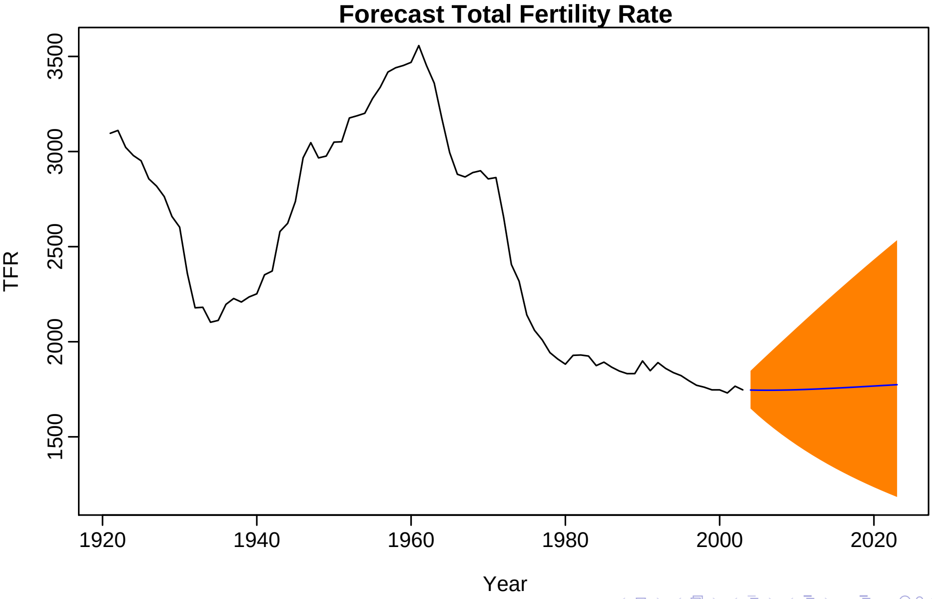
$x = 0, 1, 2, \dots$.

- 10000 sample paths of population $P_t(x)$, deaths $D_t(x)$ and births $B_t(x)$ generated for $t = 2005, \dots, 2024$ and $x = 0, 1, 2, \dots,$.
- This allows the computation of the empirical forecast distribution of any demographic quantity that is based on births, deaths and population numbers.

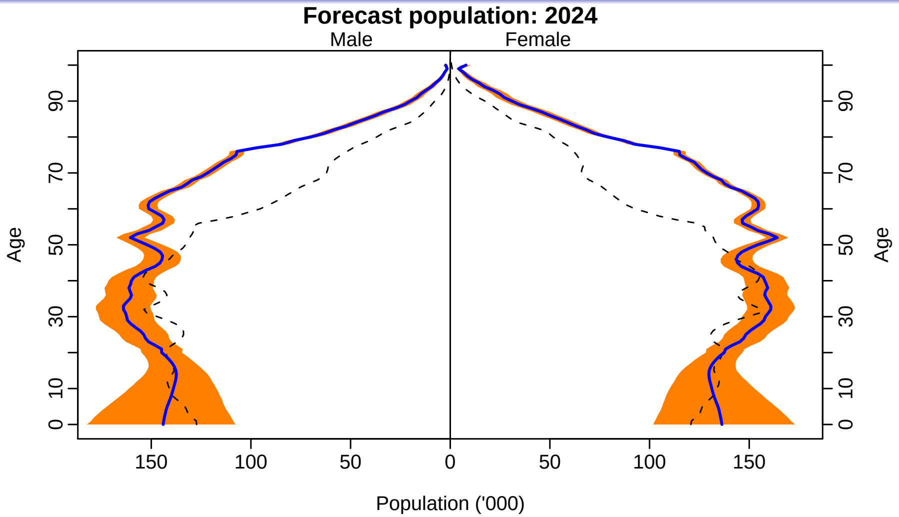
Forecasts of life expectancy at age 0



Forecasts of TFR

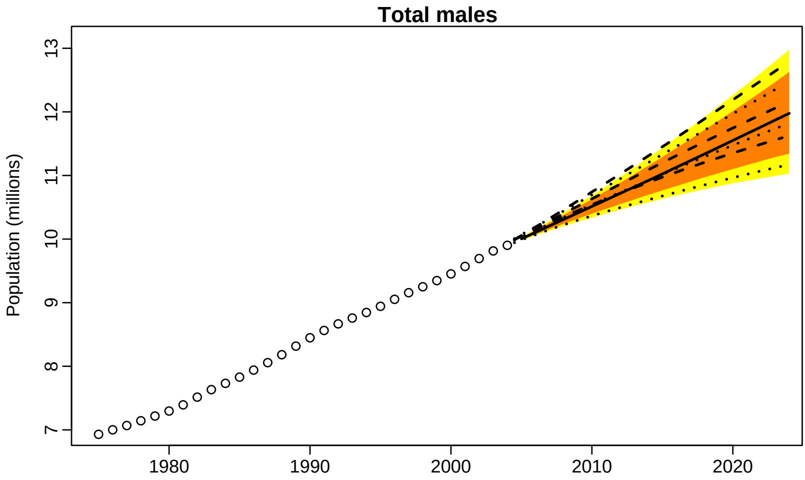


Population forecasts



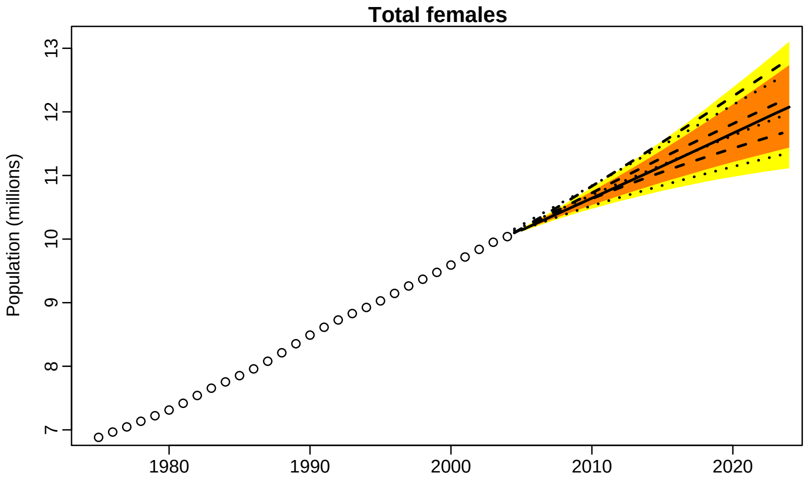
Forecast population pyramid for 2024, along with 80% prediction intervals. Dashed: actual population pyramid for 2004.

Population forecasts



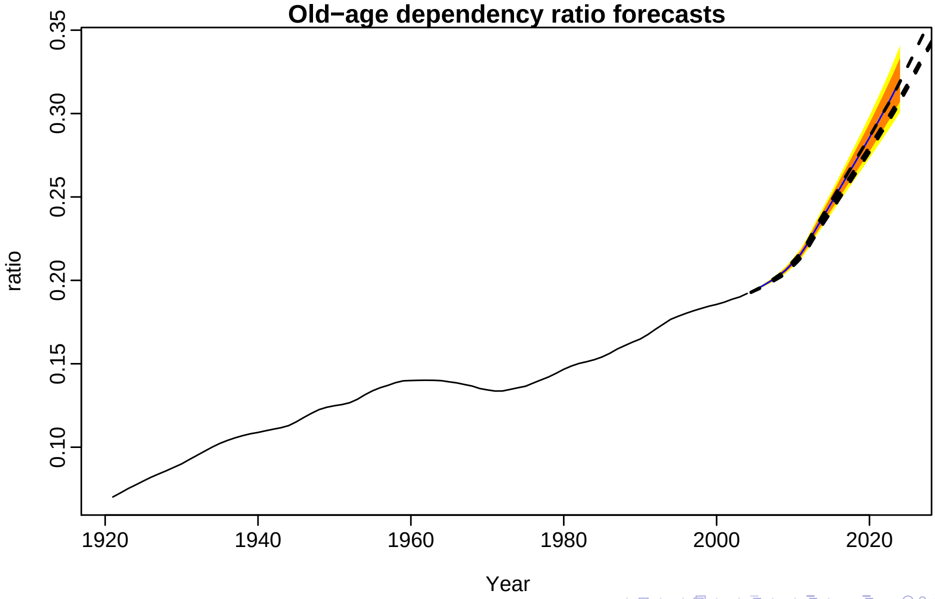
Twenty-year forecasts of total population along with 80% and 95% prediction intervals. Dashed: ABS (2006) projections. Dotted: ABS (2003) projections.

Population forecasts



Twenty-year forecasts of total population along with 80% and 95% prediction intervals. Dashed: ABS (2006) projections. Dotted: ABS (2003) projections.

Old-age dependency ratio



Advantages of stochastic simulation approach

- Functional data analysis provides a way of forecasting age-specific mortality, fertility and net migration.

Advantages of stochastic simulation approach

- Functional data analysis provides a way of forecasting age-specific mortality, fertility and net migration.
- Stochastic age-specific cohort-component simulation provides a way of forecasting many demographic quantities with prediction intervals.

Advantages of stochastic simulation approach

- Functional data analysis provides a way of forecasting age-specific mortality, fertility and net migration.
- Stochastic age-specific cohort-component simulation provides a way of forecasting many demographic quantities with prediction intervals.
- No need to select combinations of assumed rates.

Advantages of stochastic simulation approach

- Functional data analysis provides a way of forecasting age-specific mortality, fertility and net migration.
- Stochastic age-specific cohort-component simulation provides a way of forecasting many demographic quantities with prediction intervals.
- No need to select combinations of assumed rates.
- True prediction intervals with specified coverage for population and all derived variables (TFR, life expectancy, old-age dependencies, etc.)

Advantages of stochastic simulation approach

- Functional data analysis provides a way of forecasting age-specific mortality, fertility and net migration.
- Stochastic age-specific cohort-component simulation provides a way of forecasting many demographic quantities with prediction intervals.
- No need to select combinations of assumed rates.
- True prediction intervals with specified coverage for population and all derived variables (TFR, life expectancy, old-age dependencies, etc.)
- Economic planning is better based on prediction intervals rather than mean or median forecasts.

Advantages of stochastic simulation approach

- Functional data analysis provides a way of forecasting age-specific mortality, fertility and net migration.
- Stochastic age-specific cohort-component simulation provides a way of forecasting many demographic quantities with prediction intervals.
- No need to select combinations of assumed rates.
- True prediction intervals with specified coverage for population and all derived variables (TFR, life expectancy, old-age dependencies, etc.)
- Economic planning is better based on prediction intervals rather than mean or median forecasts.
- Stochastic models allow true policy analysis to be carried out.

Extensions

- We intend extending this to Australian states, and to other countries.
- We intend extending the individual models to allow dependencies between sexes and allow cohort effects.

Extensions

- We intend extending this to Australian states, and to other countries.
- We intend extending the individual models to allow dependencies between sexes and allow cohort effects.

Software and papers:

- Booth (2004) On the importance of being uncertain: forecasting population futures for Australia. *People and Place*, **12**(2), 1–11.
- Hyndman and Booth (2007) Stochastic population forecasts using functional data models for mortality, fertility and migration. **www.robhyndman.info**
- Hyndman, Booth, Tickle and Maindonald (2006) demography **R package** v0.97. **www.robhyndman.info/Rlibrary/demography**

Outline

Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.

Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.

Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.

Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.

> A good forecaster is not smarter than everyone else, he merely has his ignorance better organised.
>
> *(Anonymous)*

Conclusions

- Uncertainty statements are **essential** when making predictions and should be provided whether they are asked for or not.
- Uncertainty statements can take the form of prediction intervals or prediction densities.

> A good forecaster is not smarter than everyone else, he merely has his ignorance better organised.
>
> *(Anonymous)*

Slides available from
http://www.robhyndman.info/