Stochastic population forecasts using functional data models

Rob J Hyndman

Department of Econometrics and Business Statistics
MONASH University

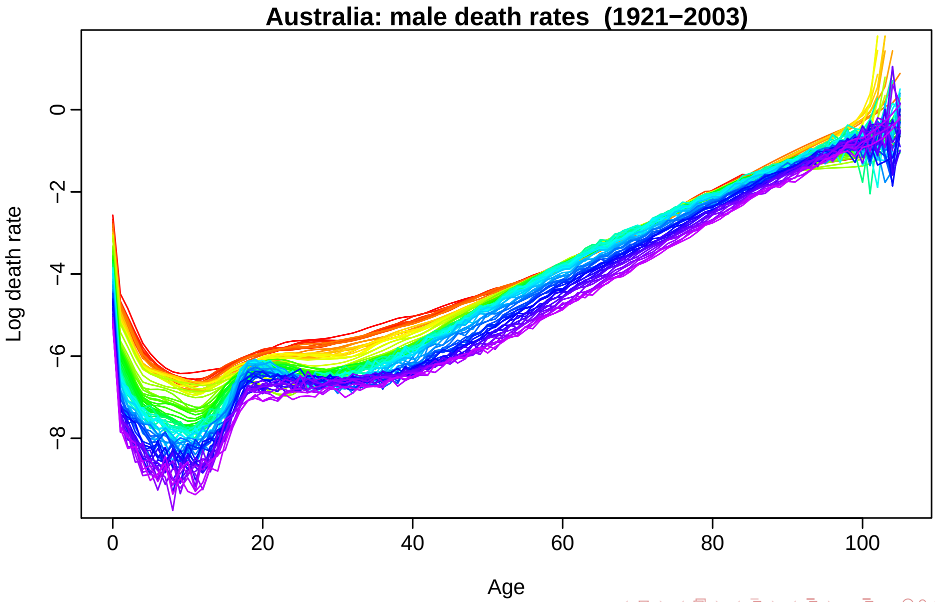
Outline

1. Functional time series
2. Current state of Australian population forecasting
3. Stochastic population forecasting

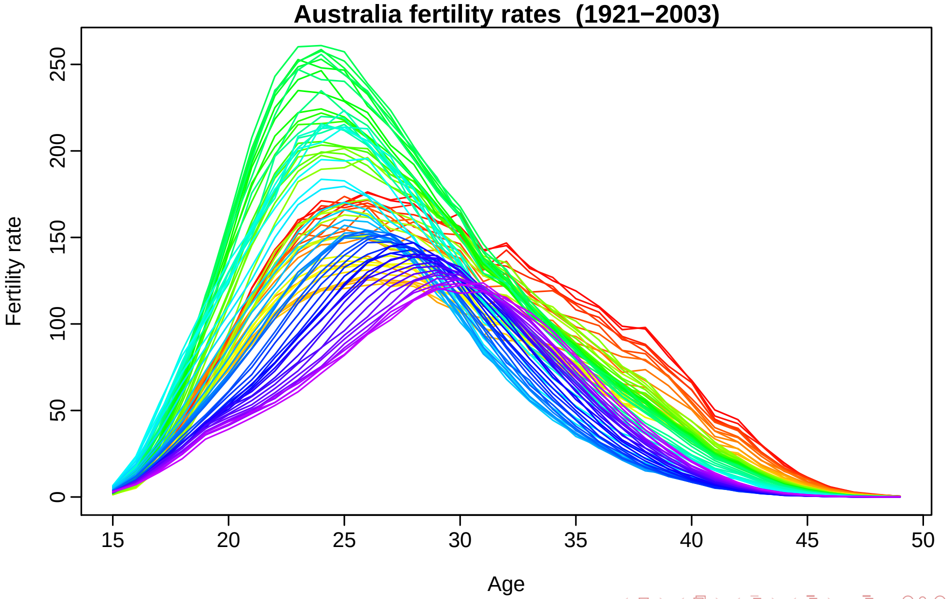
Outline

1. **Functional time series**
2. Current state of Australian population forecasting
3. Stochastic population forecasting

Mortality rates



Fertility rates



Some notation

Let $y_t(x_i)$ be the observed data in period t at location x_i, $i = 1, \dots, p$, $t = 1, \dots, n$. We assume

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

where $\varepsilon_{t,i}$ is an iid standard normal random variable and $\sigma_t(x_i)$ allows the amount of noise to vary with x.

Some notation

Let $y_t(x_i)$ be the observed data in period t at location x_i, $i = 1, \dots, p$, $t = 1, \dots, n$. We assume

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

where $\varepsilon_{t,i}$ is an iid standard normal random variable and $\sigma_t(x_i)$ allows the amount of noise to vary with x.

- We assume $s_t(x)$ is a smooth function of x.

Some notation

Let $y_t(x_i)$ be the observed data in period t at location x_i, $i = 1, \dots, p$, $t = 1, \dots, n$. We assume

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

where $\varepsilon_{t,i}$ is an iid standard normal random variable and $\sigma_t(x_i)$ allows the amount of noise to vary with x.

- We assume $s_t(x)$ is a smooth function 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 in period t at location x_i, $i = 1, \dots, p$, $t = 1, \dots, n$. We assume

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

where $\varepsilon_{t,i}$ is an iid standard normal random variable and $\sigma_t(x_i)$ allows the amount of noise to vary with x.

- We assume $s_t(x)$ is a smooth function 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 $\varepsilon_{t,x} \sim N(0,1)$ and $e_t(x) \sim \mathrm{N}(0, v(x))$.

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

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

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

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

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.

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

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 time series models.

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

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 time series models.
5. Put it all together to get forecasts of $y_t(x)$.

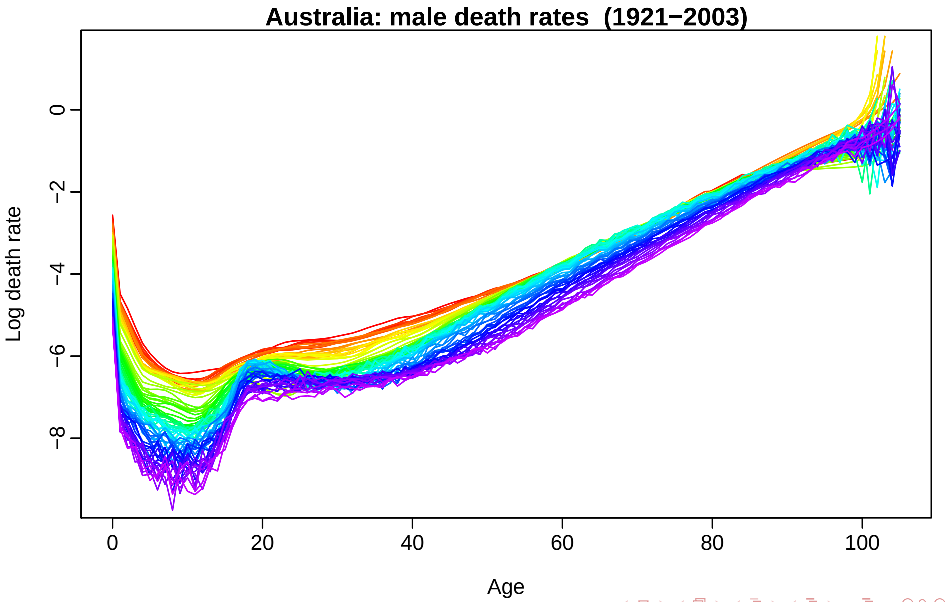
Functional time series model

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

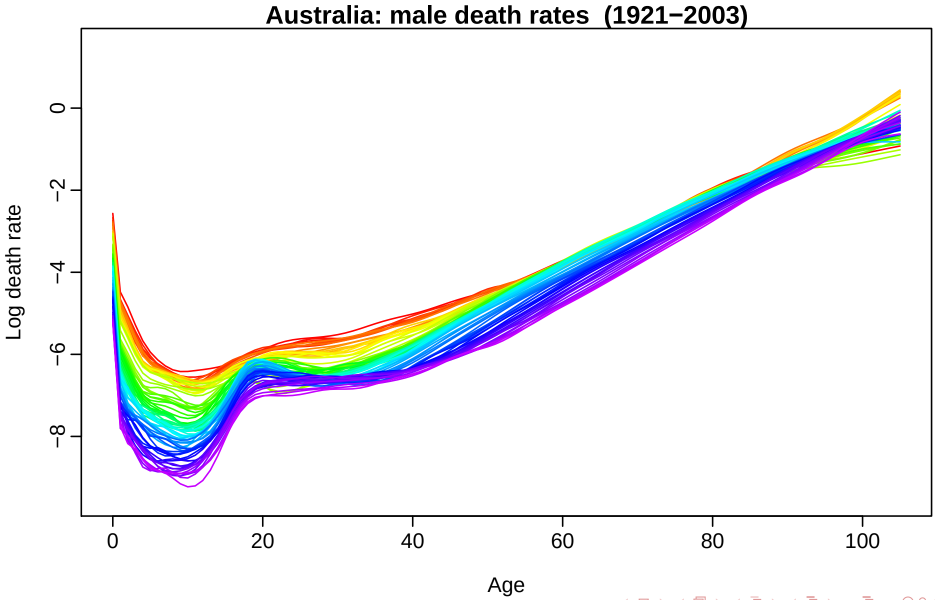
where $\varepsilon_{t,x} \sim N(0,1)$ and $e_t(x) \sim \mathrm{N}(0, v(x))$.

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 time series models.
5. Put it all together to get forecasts of $y_t(x)$.

1. Creating functional time series



1. Creating functional time series



1. Creating functional time series

Monotonic regression splines

- Fit penalized regression spline with a large number of knots.

1. Creating functional time series

Monotonic regression splines

- Fit penalized regression spline with a large number of knots.
- For mortality data, constrain curve to be monotonic for $x > b$.

1. Creating functional time series

Monotonic regression splines

- Fit penalized regression spline with a large number of knots.
- For mortality data, constrain curve to be monotonic for $x > b$.
- Choosing $b = 50$ seems to work quite well in practice for mortality data.

1. Creating functional time series

Monotonic regression splines

- Fit penalized regression spline with a large number of knots.
- For mortality data, constrain curve to be monotonic for $x > b$.
- Choosing $b = 50$ seems to work quite well in practice for mortality data.
- Fit is weighted with weights $w_t(x_i) = \sigma_t^{-2}(x_i)$ (based on Poisson deaths).

1. Creating functional time series

Monotonic regression splines

- Fit penalized regression spline with a large number of knots.
- For mortality data, constrain curve to be monotonic for $x > b$.
- Choosing $b = 50$ seems to work quite well in practice for mortality data.
- Fit is weighted with weights $w_t(x_i) = \sigma_t^{-2}(x_i)$ (based on Poisson deaths).
- This can be done using a modification of the gam function in the mgcv package in **R**.

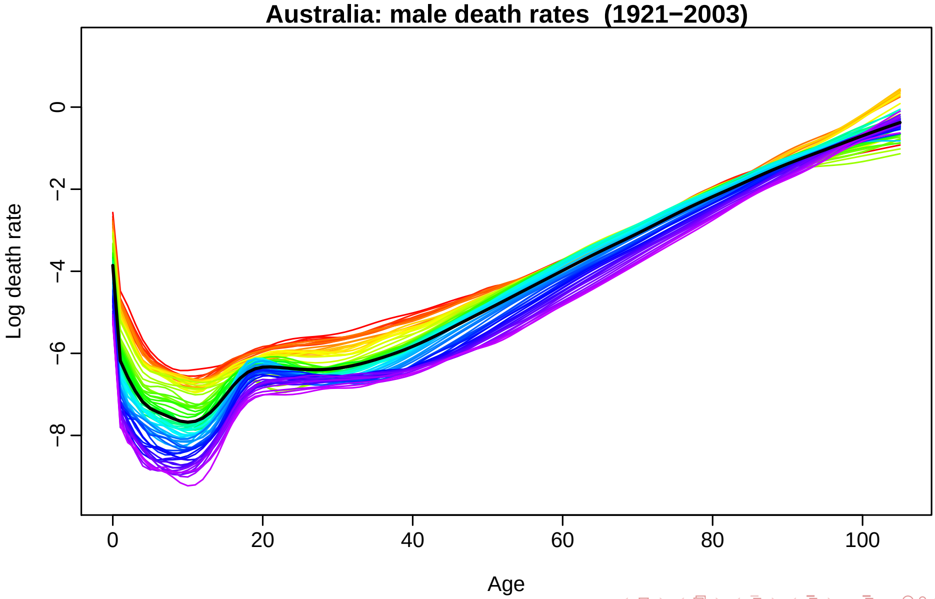
Functional time series model

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

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

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 time series models.
5. Put it all together to get forecasts of $y_t(x)$.

2. Estimate $\mu(x)$



Functional time series model

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

$$s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)$$

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

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 time series models.
5. Put it all together to get forecasts of $y_t(x)$.

3. Functional PC

The optimal basis functions

$$s_t(x) = \mu(x) + \sum_{i=0}^{K} \beta_{t,i}\phi_i(x) + e_t(x).$$

where $e_t(x) = \displaystyle\sum_{i=K+1}^{n-1} \beta_{t,i}\,\phi_i(x)$.

3. Functional PC

The optimal basis functions

$$s_t(x) = \mu(x) + \sum_{i=0}^{K} \beta_{t,i}\phi_i(x) + e_t(x).$$

where $e_t(x) = \sum_{i=K+1}^{n-1} \beta_{t,i}\,\phi_i(x)$.

- For a given K, the basis functions $\phi_i(x)$ which minimize

$$\text{MISE} = \frac{1}{n}\sum_{t=1}^{n}\int v_t^2(x)\,dx$$

are the principal components.

3. Functional PC

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_i(x)$.

3. Functional PC

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_i(x)$.
- The PC scores for each year are given by

$$z_{i,t} = \int \phi_i(x)\hat{s}_t(x)dx$$

3. Functional PC

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_i(x)$.
- The PC scores for each year are given by
$$z_{i,t} = \int \phi_i(x)\hat{s}_t(x)dx$$
- The aim is to:

3. Functional PC

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_i(x)$.
- The PC scores for each year are given by
$$z_{i,t} = \int \phi_i(x)\hat{s}_t(x)dx$$
- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $z_{1,t}$ subject to the constraint $\int \phi_i^2(x)dx = 1$.

3. Functional PC

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_i(x)$.
- The PC scores for each year are given by

$$z_{i,t} = \int \phi_i(x)\hat{s}_t(x)dx$$

- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $z_{1,t}$ subject to the constraint $\int \phi_i^2(x)dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $z_{2,t}$ such that $\int \phi_i^2(x)dx = 1$ and $\int \phi_1(x)\phi_2(x)dx = 0$.

3. Functional PC

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_i(x)$.
- The PC scores for each year are given by

$$z_{i,t} = \int \phi_i(x)\hat{s}_t(x)dx$$

- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $z_{1,t}$ subject to the constraint $\int \phi_i^2(x)dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $z_{2,t}$ such that $\int \phi_i^2(x)dx = 1$ and $\int \phi_1(x)\phi_2(x)dx = 0$.
  3. Find the weight function $\phi_3(x)$ that . . .

3. Functional PC

Calculating the optimal basis functions

- Define F to be $n \times m$ matrix with (t, j)th element $\hat{s}_t(x_j^*) - \mu(x_j^*)$ where $\{x_1^*, \dots, x_m^*\}$ are a dense grid on x.

3. Functional PC

Calculating the optimal basis functions

- Define F to be $n \times m$ matrix with (t, j)th element $\hat{s}_t(x_j^*) - \mu(x_j^*)$ where $\{x_1^*, \dots, x_m^*\}$ are a dense grid on x.
- Let $V = (n-1)^{-1}F'F$ be $m \times m$ sample covariance matrix of F.

3. Functional PC

Calculating the optimal basis functions

- Define F to be $n \times m$ matrix with (t, j)th element $\hat{s}_t(x_j^*) - \mu(x_j^*)$ where $\{x_1^*, \dots, x_m^*\}$ are a dense grid on x.
- Let $V = (n-1)^{-1}F'F$ be $m \times m$ sample covariance matrix of F.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of V where $K \le n-1$. The (i, j)th element of Φ_K is $\phi_i(x_j^*)$.

3. Functional PC

Calculating the optimal basis functions

- Define F to be $n \times m$ matrix with (t, j)th element $\hat{s}_t(x_j^*) - \mu(x_j^*)$ where $\{x_1^*, \dots, x_m^*\}$ are a dense grid on x.
- Let $V = (n-1)^{-1}F'F$ be $m \times m$ sample covariance matrix of F.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of V where $K \le n-1$. The (i, j)th element of Φ_K is $\phi_i(x_j^*)$.
- The coefficients are given by $B_K = F\Phi_K$ which has (t, i)th element $\hat{\beta}_{t,i}$.

3. Functional PC

Calculating the optimal basis functions

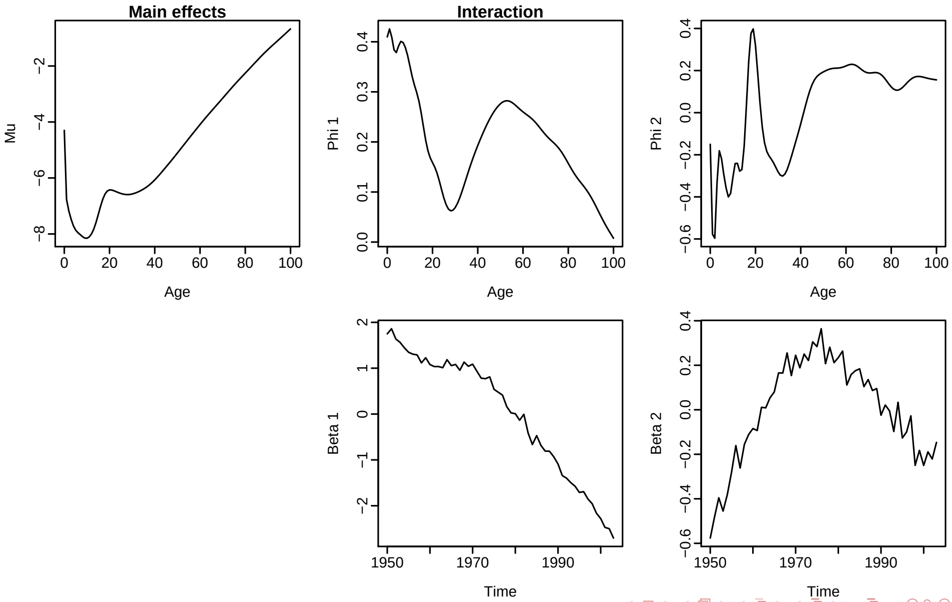
- Define F to be $n \times m$ matrix with (t, j)th element $\hat{s}_t(x_j^*) - \mu(x_j^*)$ where $\{x_1^*, \dots, x_m^*\}$ are a dense grid on x.
- Let $V = (n-1)^{-1} F'F$ be $m \times m$ sample covariance matrix of F.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of V where $K \le n-1$. The (i, j)th element of Φ_K is $\phi_i(x_j^*)$.
- The coefficients are given by $B_K = F\Phi_K$ which has (t, i)th element $\hat{\beta}_{t,i}$.
- The basis functions are orthogonal.

3. Functional PC

Calculating the optimal basis functions

- Define F to be $n \times m$ matrix with (t, j)th element $\hat{s}_t(x_j^*) - \mu(x_j^*)$ where $\{x_1^*, \dots, x_m^*\}$ are a dense grid on x.
- Let $V = (n-1)^{-1}F'F$ be $m \times m$ sample covariance matrix of F.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of V where $K \le n-1$. The (i, j)th element of Φ_K is $\phi_i(x_j^*)$.
- The coefficients are given by $B_K = F\Phi_K$ which has (t, i)th element $\hat{\beta}_{t,i}$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

3. Functional PC



3. Functional PC

Recap

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

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

- Pure time term excluded as it would make $\{\beta_{t,k}\}$ correlated.

3. Functional PC

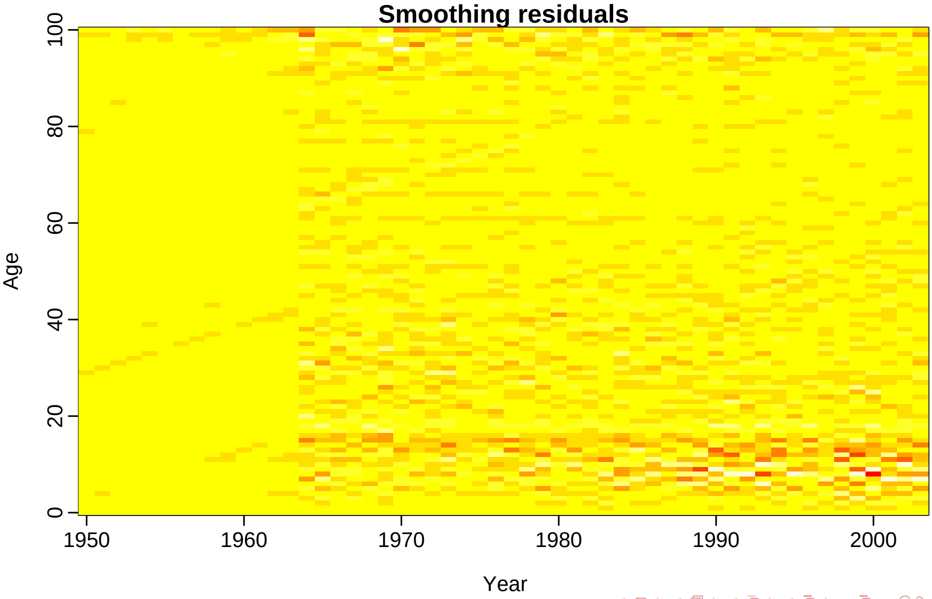
Recap

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

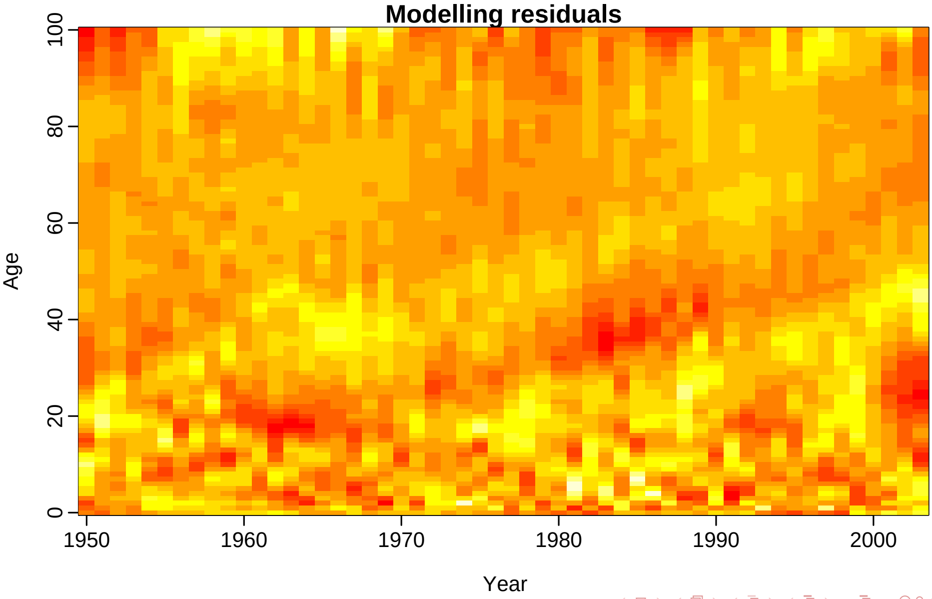
where $\varepsilon_{t,x} \sim N(0,1)$ and $e_t(x) \sim \mathrm{N}(0, v(x))$.

- Pure time term excluded as it would make $\{\beta_{t,k}\}$ correlated.
- We can check if any structure is left in the residuals $\varepsilon_{t,x}$ (smoothing problem) and $e_t(x)$ (modelling problem).

3. Functional PC



3. Functional PC



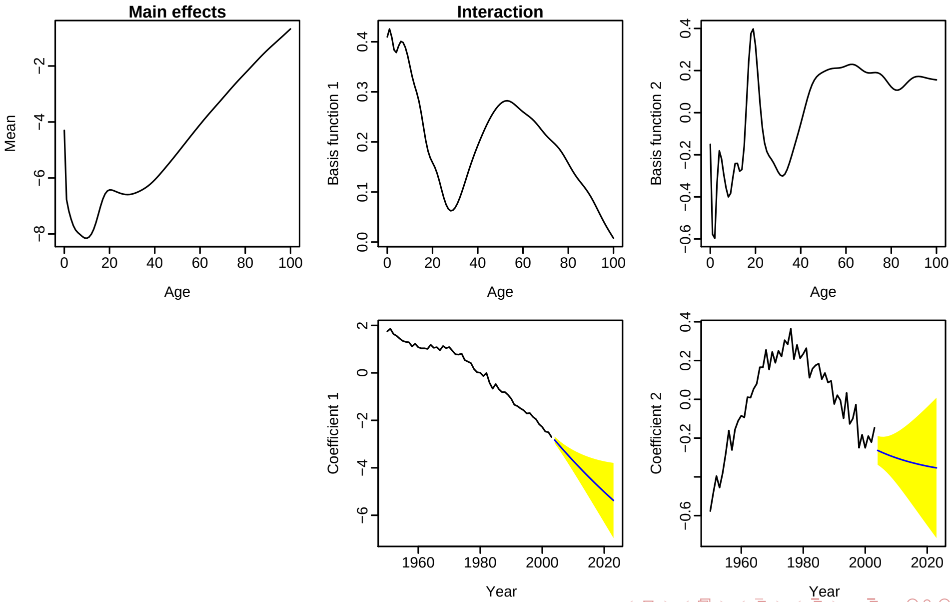
Functional time series model

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

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

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 time series models.**
5. Put it all together to get forecasts of $y_t(x)$.

4. Forecasting the coefficients



4. Forecasting the coefficients

- I use exponential smoothing state space models based on Ord, Koehler & Snyder (JASA, 1997) and Hyndman, Koehler, Snyder & Grose (IJF, 2002).

4. Forecasting the coefficients

- I use exponential smoothing state space models based on Ord, Koehler & Snyder (JASA, 1997) and Hyndman, Koehler, Snyder & Grose (IJF, 2002).
- These provide a stochastic framework for exponential smoothing forecasts.

4. Forecasting the coefficients

- I use exponential smoothing state space models based on Ord, Koehler & Snyder (JASA, 1997) and Hyndman, Koehler, Snyder & Grose (IJF, 2002).
- These provide a stochastic framework for exponential smoothing forecasts.
- The models shown are equivalent to

4. Forecasting the coefficients

- I use exponential smoothing state space models based on Ord, Koehler & Snyder (JASA, 1997) and Hyndman, Koehler, Snyder & Grose (IJF, 2002).
- These provide a stochastic framework for exponential smoothing forecasts.
- The models shown are equivalent to
  - damped Holt's method

4. Forecasting the coefficients

- I use exponential smoothing state space models based on Ord, Koehler & Snyder (JASA, 1997) and Hyndman, Koehler, Snyder & Grose (IJF, 2002).
- These provide a stochastic framework for exponential smoothing forecasts.
- The models shown are equivalent to
  - damped Holt's method
  - ARIMA(0,1,2)

4. Forecasting the coefficients

- I use exponential smoothing state space models based on Ord, Koehler & Snyder (JASA, 1997) and Hyndman, Koehler, Snyder & Grose (IJF, 2002).
- These provide a stochastic framework for exponential smoothing forecasts.
- The models shown are equivalent to
  - damped Holt's method
  - ARIMA(0,1,2)
- Univariate models are ok because the series are uncorrelated.

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

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 time series models.
5. **Put it all together to get forecasts of $y_t(x)$.**

5. Forecasts of $y_t(x)$

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

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

5. Forecasts of $y_t(x)$

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

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

Let $\mathcal{I} = \{y_t(x_i); t = 1,\dots,n; i = 1,\dots,p\}$.

- $\mathrm{E}[y_{n+h}(x) \mid \mathcal{I}, \boldsymbol{\Phi}] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n+h|n,k}\,\hat{\phi}_k(x)$.

5. Forecasts of $y_t(x)$

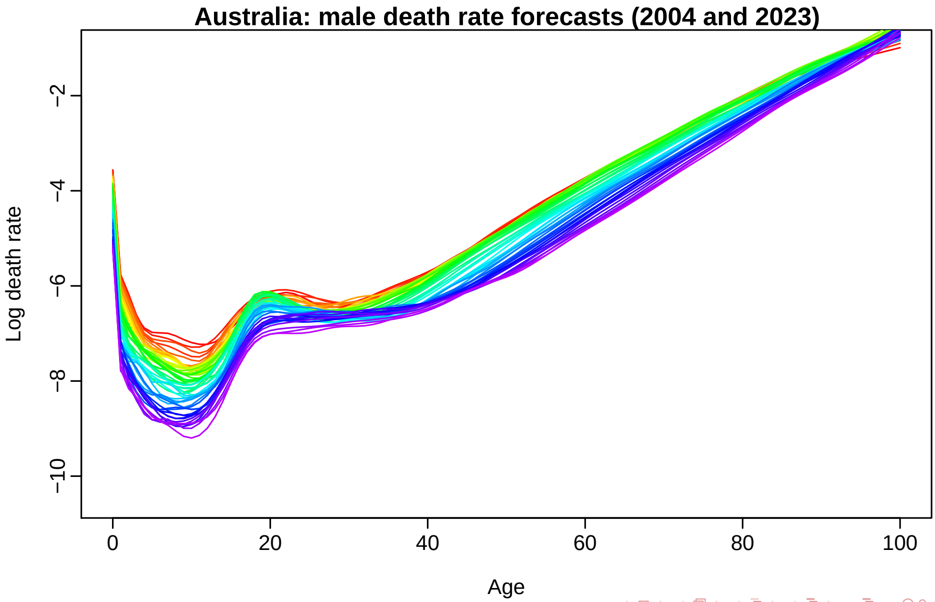
$$
\begin{aligned}
y_t(x) &= s_t(x) + \sigma_t(x)\varepsilon_{t,x} \\
s_t(x) &= \mu(x) + \sum_{k=1}^{K} \beta_{t,k}\,\phi_k(x) + e_t(x)
\end{aligned}
$$

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

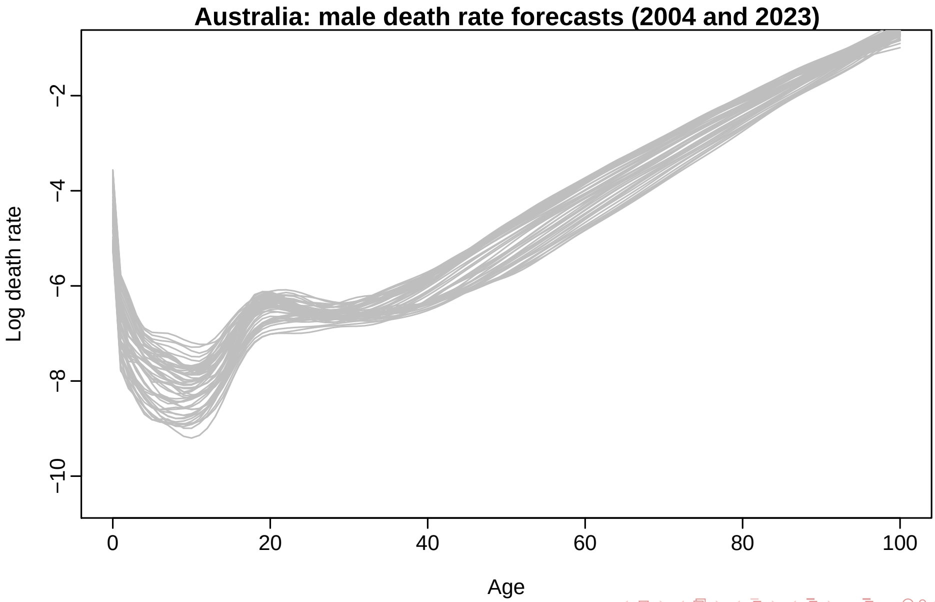
Let $\mathcal{I} = \{y_t(x_i); t = 1,\dots,n; i = 1,\dots,p\}$.

- $\mathrm{E}[y_{n+h}(x) \mid \mathcal{I}, \Phi] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\beta}_{n+h|n,k}\,\hat{\phi}_k(x)$.
- $\mathrm{Var}[y_{n+h}(x) \mid \mathcal{I}, \Phi] =$ $\sigma^2_{n+h}(x) + \hat{\sigma}^2_{\mu}(x) + \sum_{k=1}^{K} v_{n+h|n,k}\,\hat{\phi}^2_k(x) + v(x)$ where $v_{n+h|n,k} = \mathrm{Var}(\beta_{n+h,k} \mid \beta_{1,k},\dots,\beta_{n,k})$.

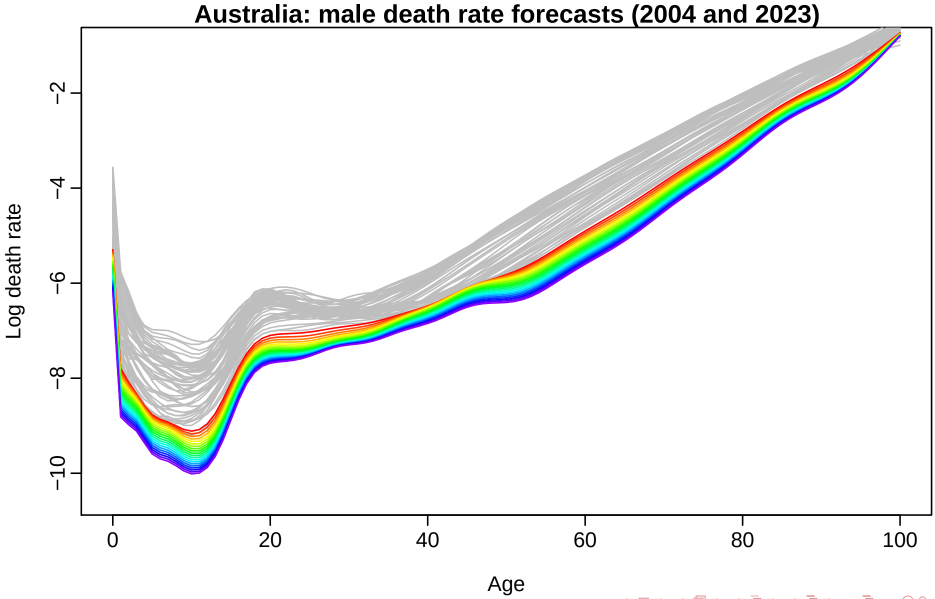
5. Forecasts of $y_t(x)$



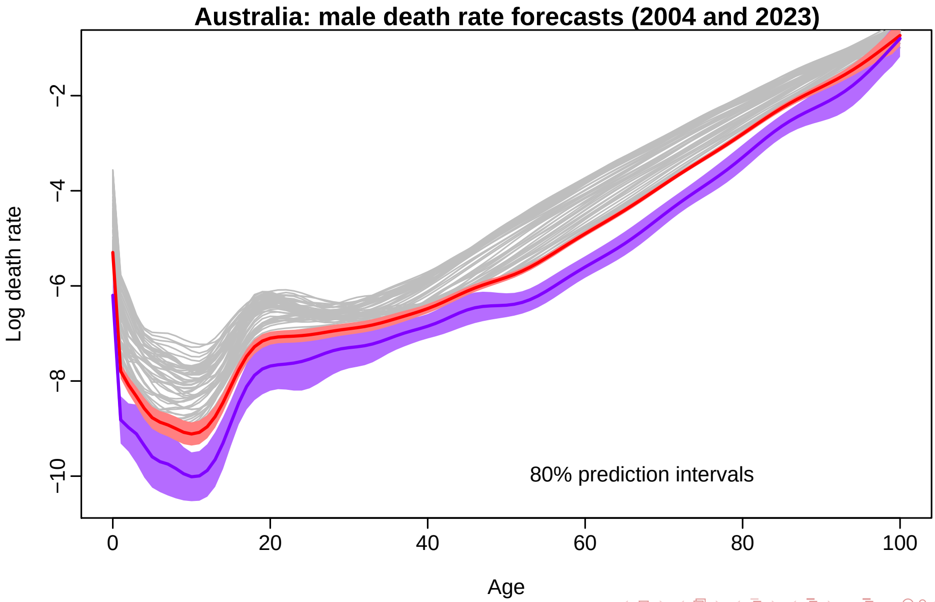
5. Forecasts of $y_t(x)$



5. Forecasts of $y_t(x)$



5. Forecasts of $y_t(x)$



Some references

- Erbas, Hyndman and Gertig (2007) Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, **26**(2), 458–470.

Some references

- Erbas, Hyndman and Gertig (2007) Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, **26**(2), 458–470.
- Hyndman and Ullah (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, to appear.

Some references

- Erbas, Hyndman and Gertig (2007) Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, **26**(2), 458–470.
- Hyndman and Ullah (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, to appear.
- Booth, Hyndman, Tickle and De Jong (2006) Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, **15**(9), 289–310.

Some references

- Erbas, Hyndman and Gertig (2007) Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, **26**(2), 458–470.
- Hyndman and Ullah (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, to appear.
- Booth, Hyndman, Tickle and De Jong (2006) Lee-Carter mortality forecasting: a multi-country comparison of variants and extensions. *Demographic Research*, **15**(9), 289–310.
- Hyndman (2006) demography: Forecasting mortality and fertility data. **R package** v0.98. **www.robhyndman.info/Rlibrary/demography**

Outline

1. Functional time series
2. Current state of Australian population forecasting
3. Stochastic population forecasting

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
- No variation allowed across ages.

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
- No variation allowed across ages.
- No probabilistic basis.

ABS population projections

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

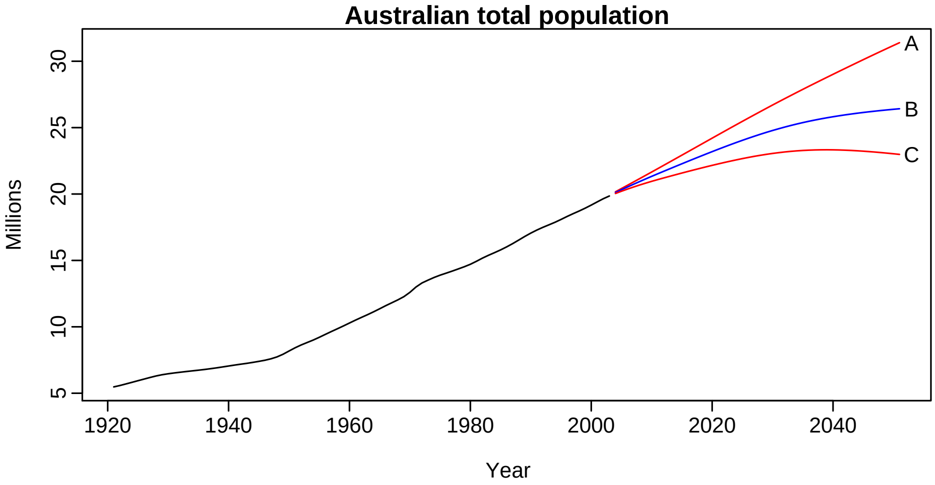
- Based on assumed mortality, fertility and migration rates
- No objectivity.
- No dynamic changes in rates allowed
- No variation allowed across ages.
- No probabilistic basis.
- Not prediction intervals.

ABS population projections

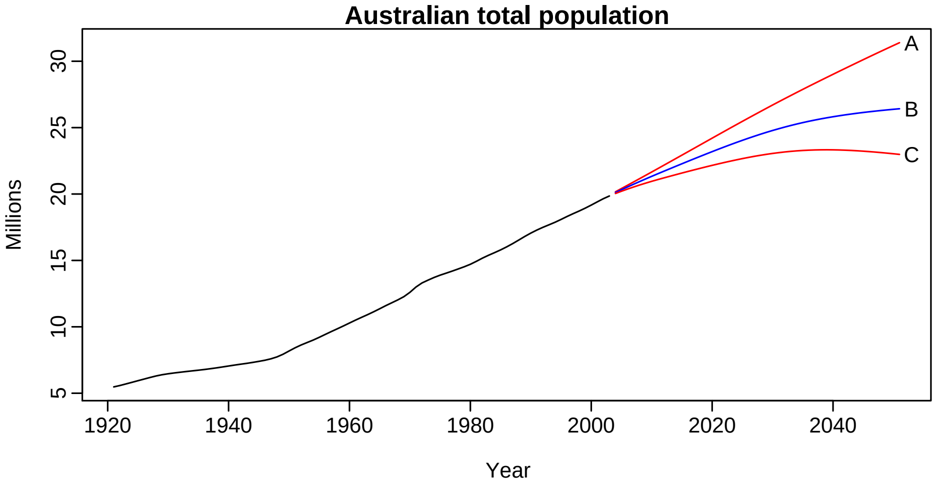
The ABS provides three projection scenarios labelled "High", "Medium" and "Low".

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

Outline

1. Functional time series
2. Current state of Australian population forecasting
3. Stochastic population forecasting

Stochastic population forecasts

- Forecasts represent the median of the future distribution.

Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty

Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)

Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)
- The probability of future events can be estimated.

Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)
- The probability of future events can be estimated.
- Economic planning is better based on prediction intervals rather than mean or median forecasts.

Stochastic population forecasts

- Forecasts represent the median of the future distribution.
- Percentiles of distribution allow information about uncertainty
- Prediction intervals with specified probability coverage for population size and all derived variables (total fertility rate, life expectancy, old-age dependencies, etc.)
- The probability of future events can be estimated.
- Economic planning is better based on prediction intervals rather than mean or median forecasts.
- Stochastic models allow true policy analysis to be carried out.

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

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

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

$P_t(x) =$ 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

$D_t(B, 0) =$ infant deaths in calendar 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

Key ideas

- Build a *stochastic functional model* for each of mortality, fertility and net migration.

Key ideas

- Build a *stochastic functional model* for each of mortality, fertility and net migration.
- Treat all observed data as *functional* (i.e., smooth curves of age) rather than discrete values.

Key ideas

- Build a *stochastic functional model* for each of mortality, fertility and net migration.
- Treat all observed data as *functional* (i.e., smooth curves of age) rather than discrete values.
- Use the models to *simulate future sample paths* of all components giving the entire age distribution at every year into the future.

Key ideas

- Build a *stochastic functional model* for each of mortality, fertility and net migration.
- Treat all observed data as *functional* (i.e., smooth curves of age) rather than discrete values.
- Use the models to *simulate future sample paths* of all components giving the entire age distribution at every year into the future.
- Compute future births, deaths, net migrants and populations from simulated rates.

Key ideas

- Build a *stochastic functional model* for each of mortality, fertility and net migration.
- Treat all observed data as *functional* (i.e., smooth curves of age) rather than discrete values.
- Use the models to *simulate future sample paths* of all components giving the entire age distribution at every year into the future.
- Compute future births, deaths, net migrants and populations from simulated rates.
- Combine the results to get *age-specific stochastic population forecasts*.

The available data

In most countries, 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

In most countries, 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

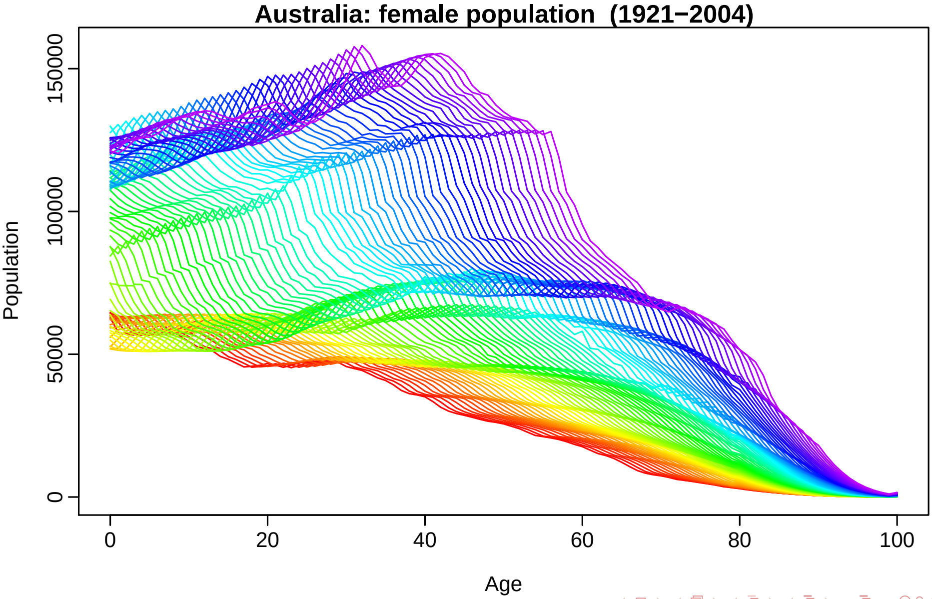
In most countries, the following data are available:

$P_t(x) =$ **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

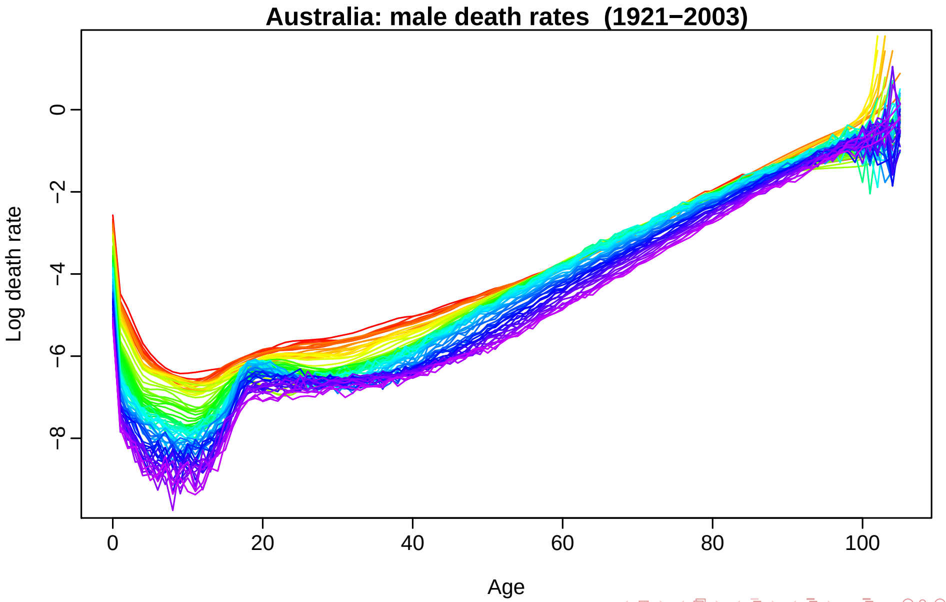
From these, we can estimate:

- $m_t(x) = D_t(x)/E_t(x) =$ 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.

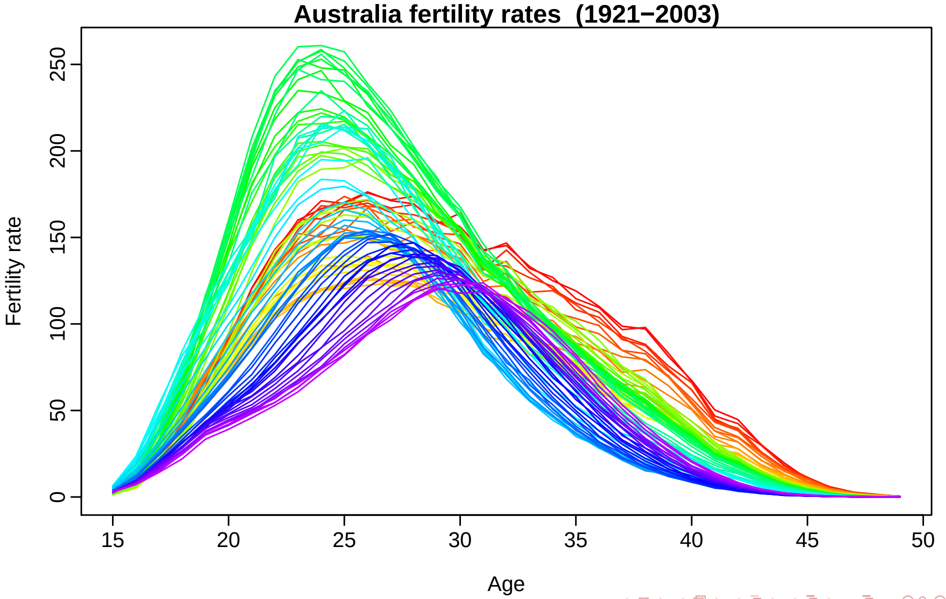
Australia's start-of-year population



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.

Net migration

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

Demographic growth-balance equation

$$G_t(x, x+1) = P_{t+1}(x+1) - P_t(x) + D_t(x, x+1)$$

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

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

Net migration

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

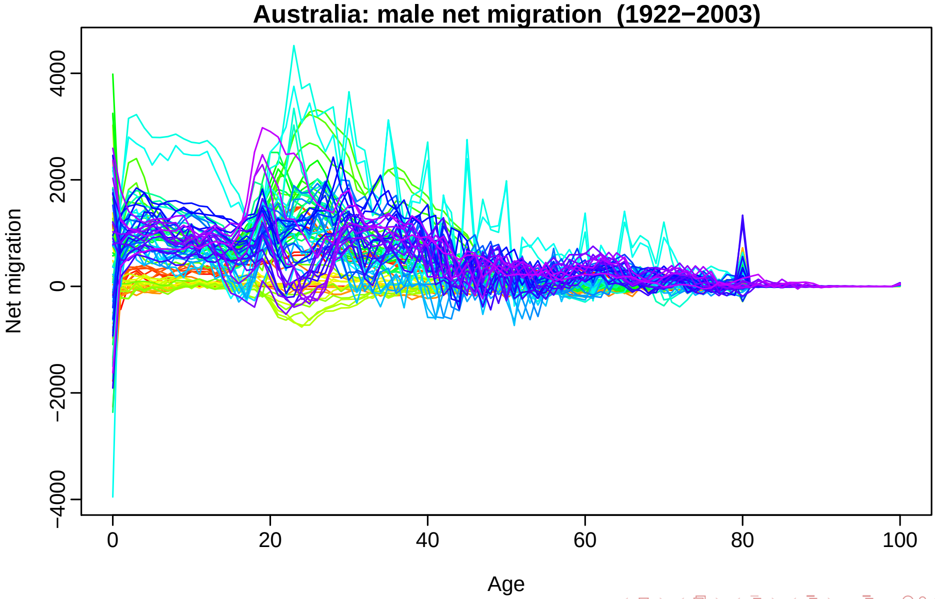
Demographic growth-balance equation

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

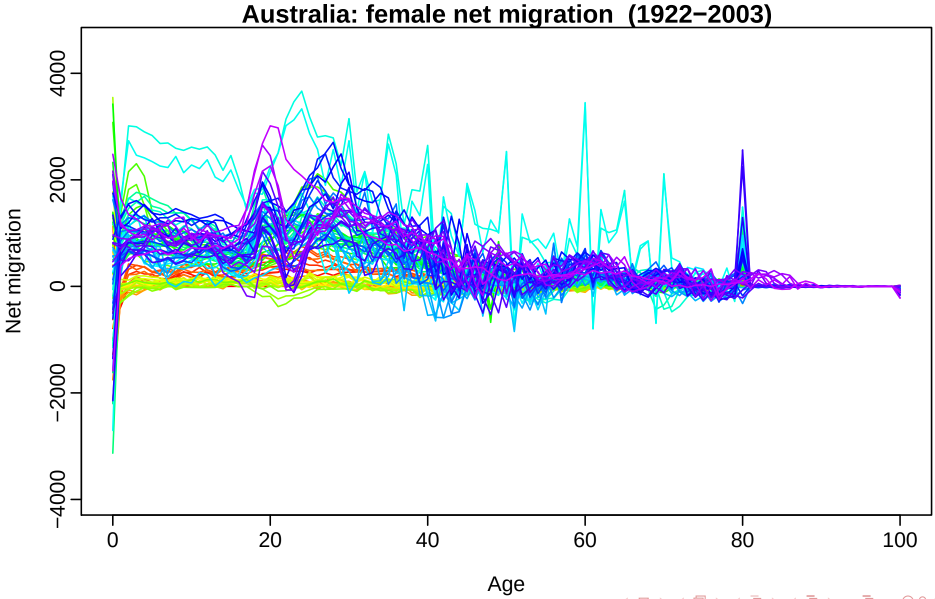
$x = 0, 1, 2, \dots$.

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

Net migration



Net migration



Stochastic population forecasts

Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration.

Stochastic population forecasts

Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration.
- Models: Five functional time series models for mortality (M/F), fertility and net migration (M/F) assuming independence between components.

Stochastic population forecasts

Component models

- Data: age/sex-specific mortality rates, fertility rates and net migration.
- Models: Five functional time series models for mortality (M/F), fertility and net migration (M/F) assuming independence between components.
- For each component:

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

Functional time series

Let $g_\lambda(u) = \begin{cases} \log(u) & \lambda = 0; \\ \frac{x^\lambda - 1}{\lambda} & \lambda > 0. \end{cases}$

- **Mortality rates:**
  $y_t(x_i) = g_0(m_t(x_i))$ where $m_t(x_i) =$ empirical mortality rate at age x_i.

Functional time series

Let $g_\lambda(u) = \begin{cases} \log(u) & \lambda = 0; \\ \frac{x^\lambda - 1}{\lambda} & \lambda > 0. \end{cases}$

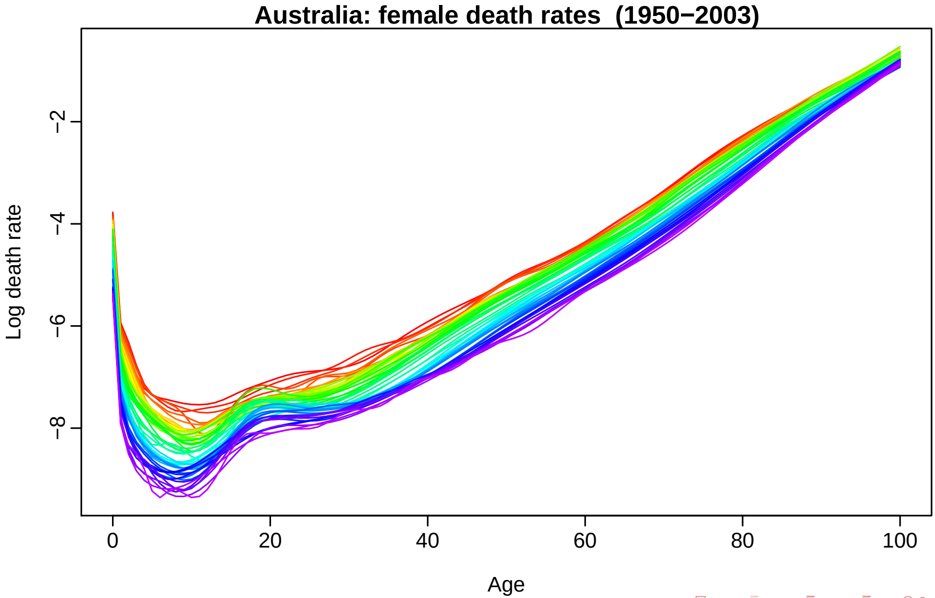
- **Mortality rates:** $y_t(x_i) = g_0(m_t(x_i))$ where $m_t(x_i) =$ empirical mortality rate at age x_i.
- **Fertility rates:** $y_t(x_i) = g_{0.45}(p_t(x_i))$ where $p_t(x_i) =$ empirical fertility rate at age x_i.

Functional time series

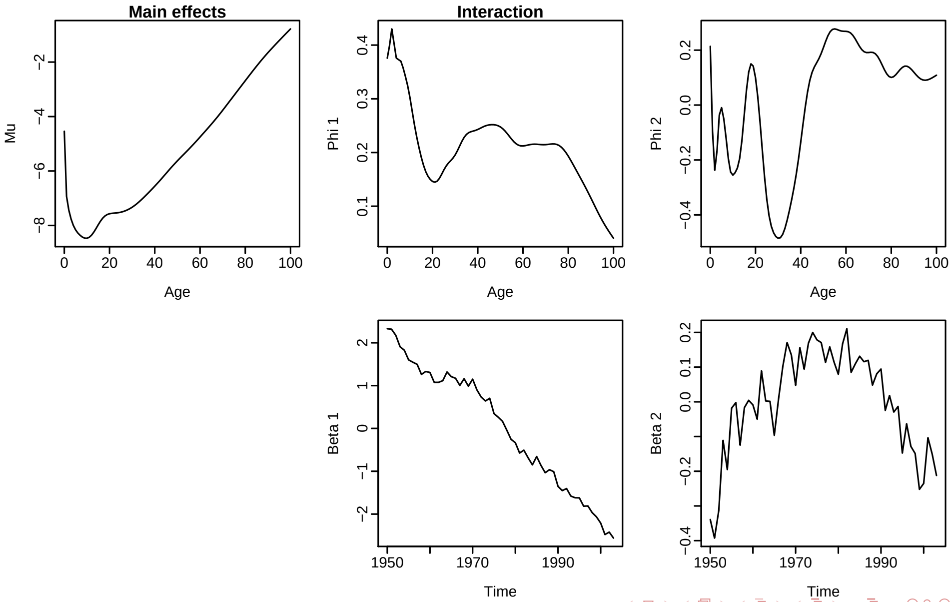
Let $g_\lambda(u) = \begin{cases} \log(u) & \lambda = 0; \\ \frac{x^\lambda - 1}{\lambda} & \lambda > 0. \end{cases}$

- **Mortality rates:**
  $y_t(x_i) = g_0(m_t(x_i))$ where $m_t(x_i) =$ empirical mortality rate at age x_i.
- **Fertility rates:**
  $y_t(x_i) = g_{0.45}(p_t(x_i))$ where $p_t(x_i) =$ empirical fertility rate at age x_i.
- **Net migration:**
  $y_t(x_i) =$ empirical net migration at age x_i.

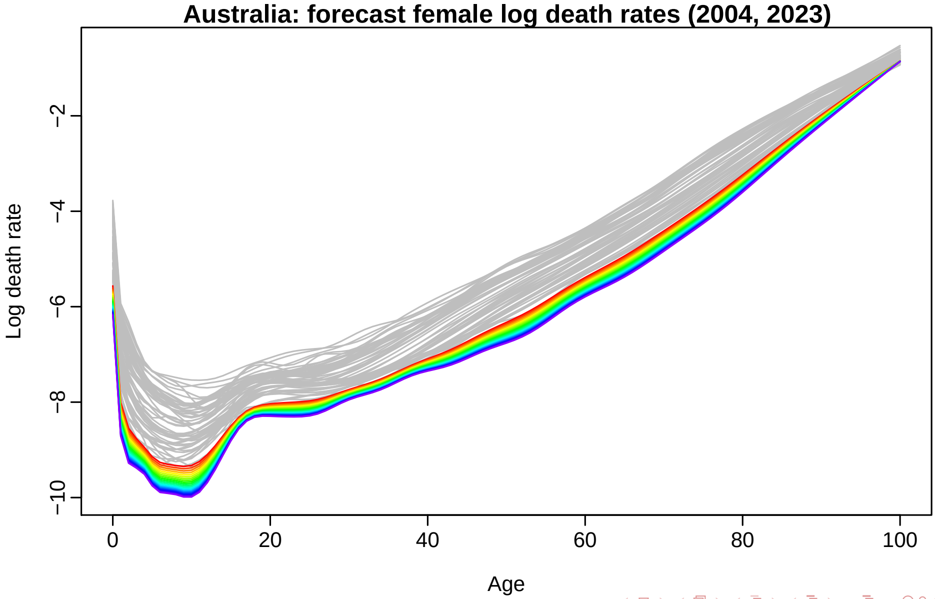
Mortality: female



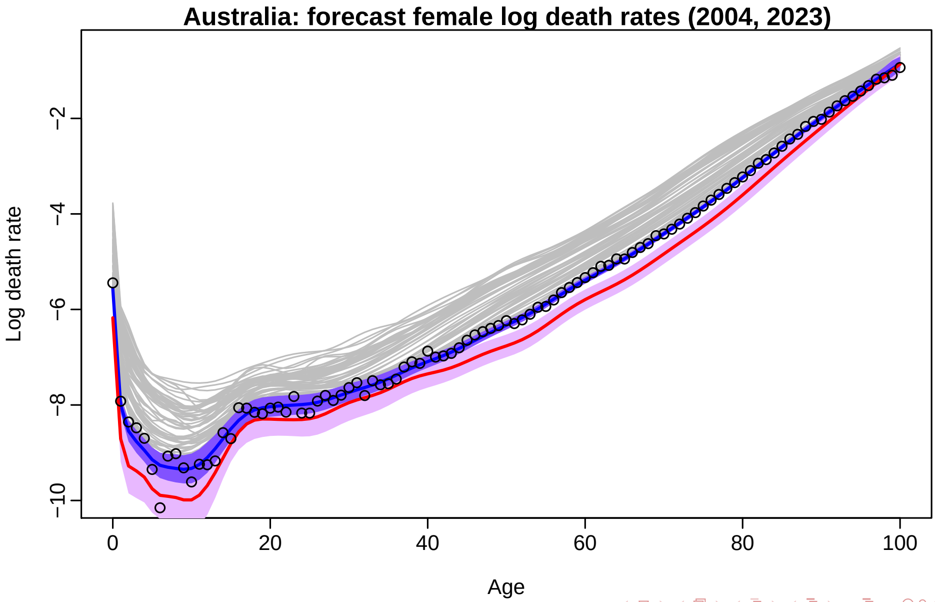
Mortality: female



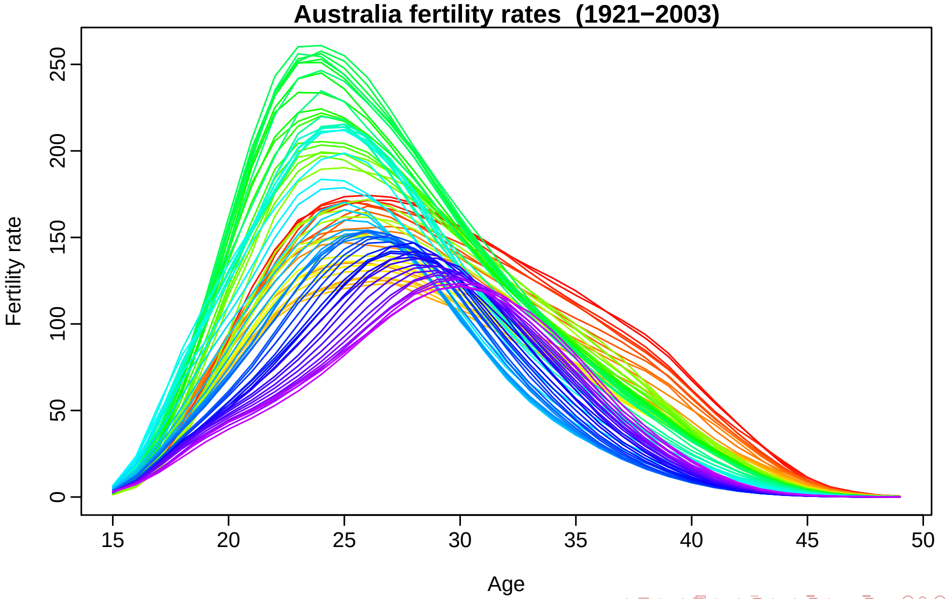
Mortality: female



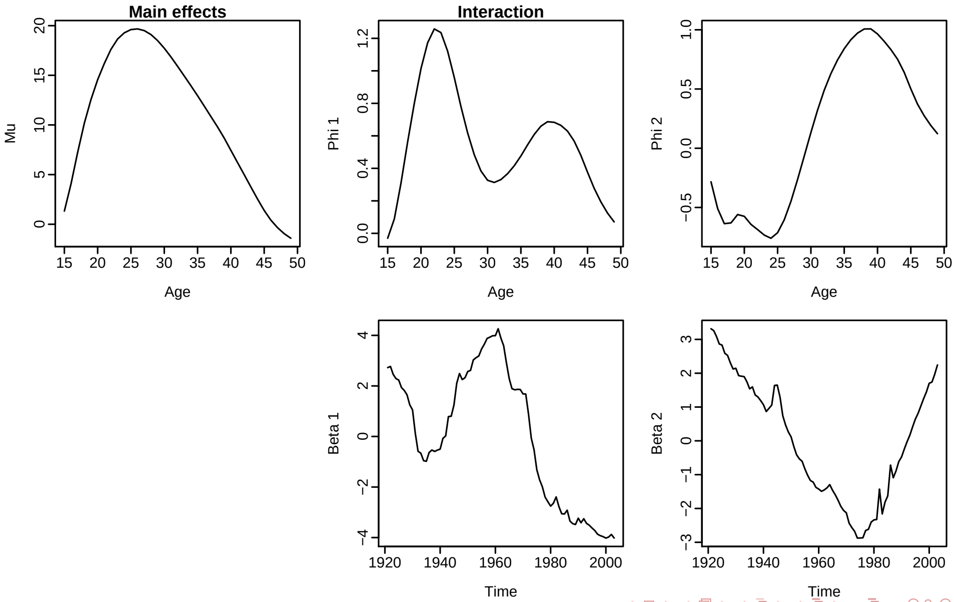
Mortality: female



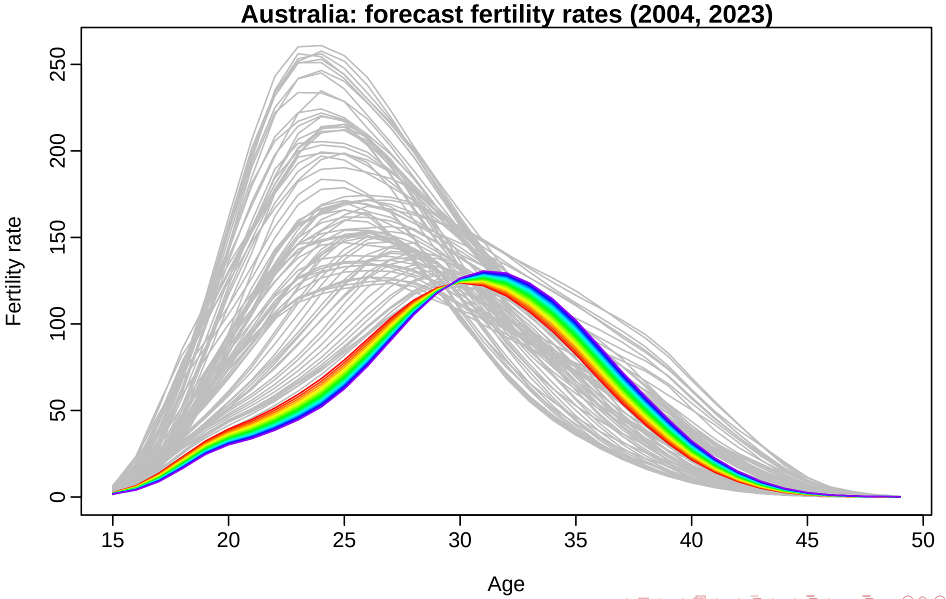
Fertility


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

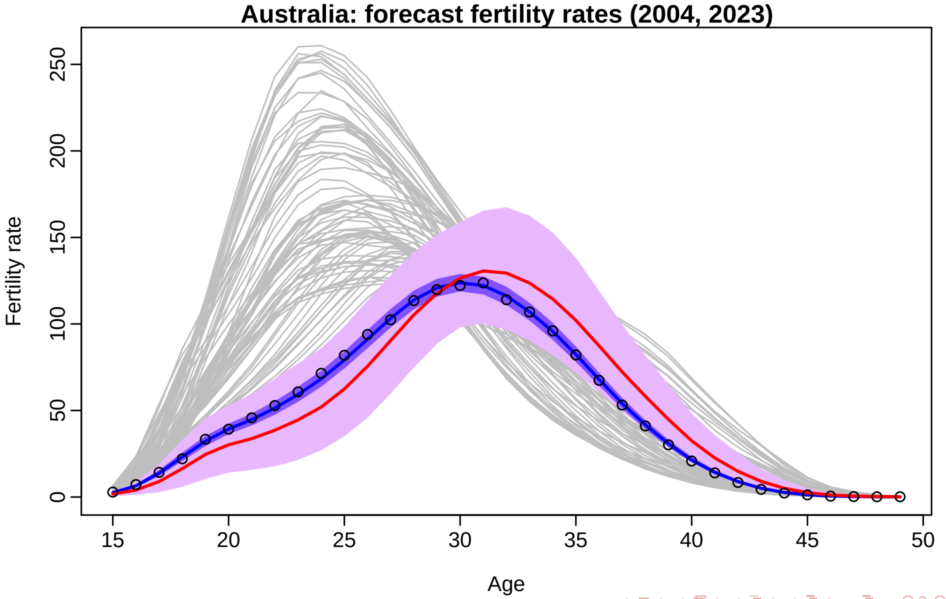
Fertility


Main effects
Mu
0
5
10
15
20
15 20 25 30 35 40 45 50
Age
Interaction
Phi 1
0.0
0.4
0.8
1.2
15 20 25 30 35 40 45 50
Age
Phi 2
−0.5
0.0
0.5
1.0
15 20 25 30 35 40 45 50
Age
Beta 1
−4
−2
0
2
4
1920 1940 1960 1980 2000
Time
Beta 2
−3
−2
−1
0
1
2
3
1920 1940 1960 1980 2000
Time

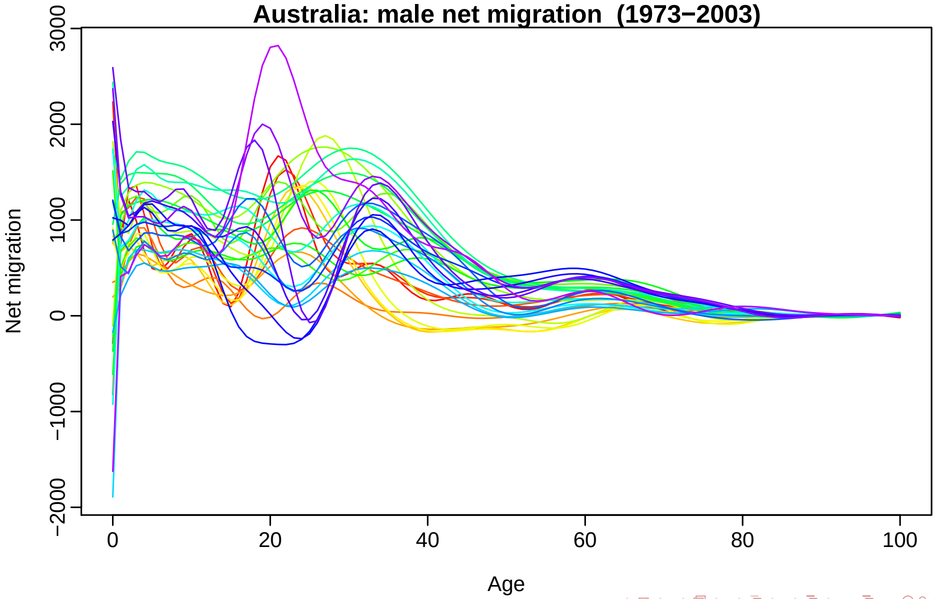
Fertility


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

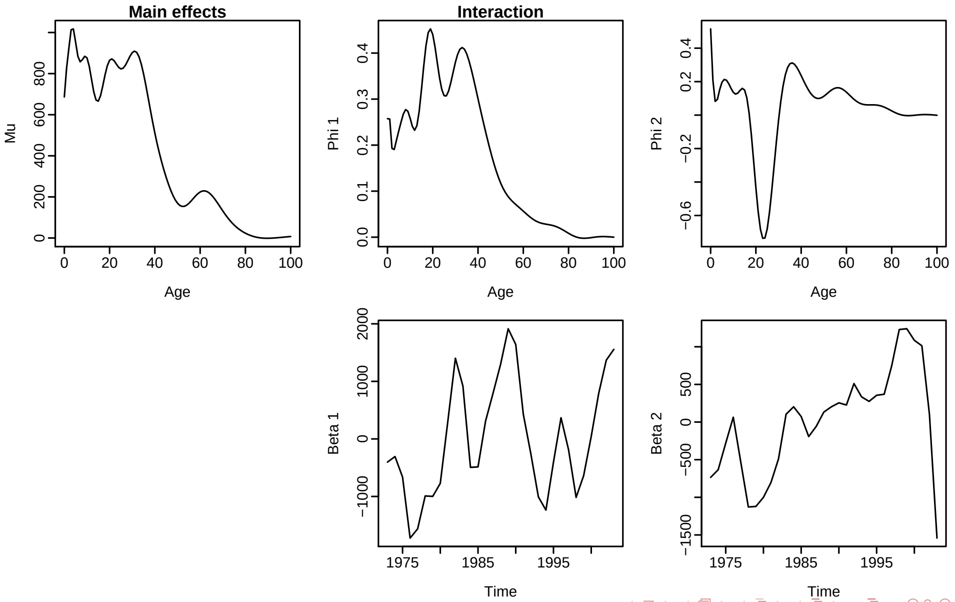
Fertility


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

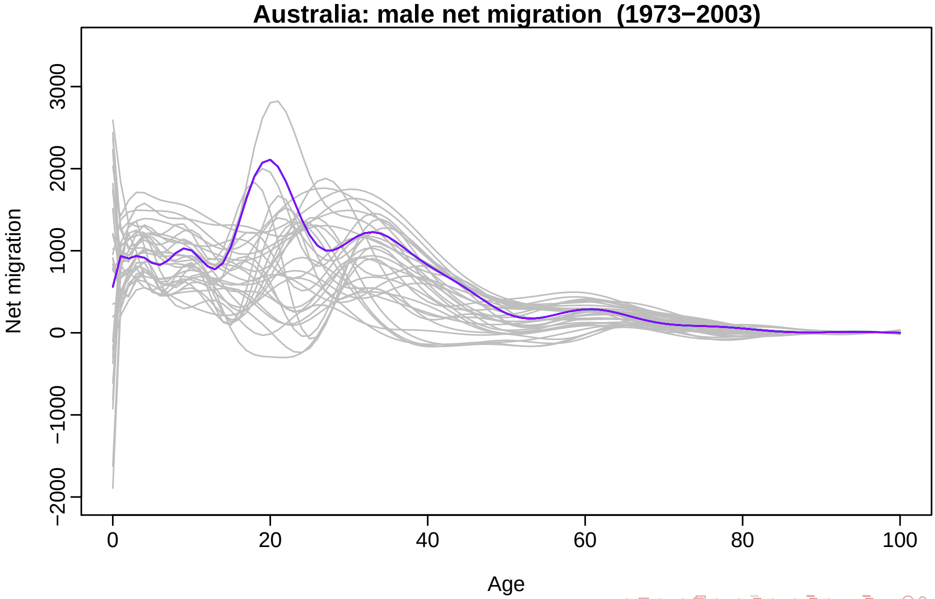
Migration: male



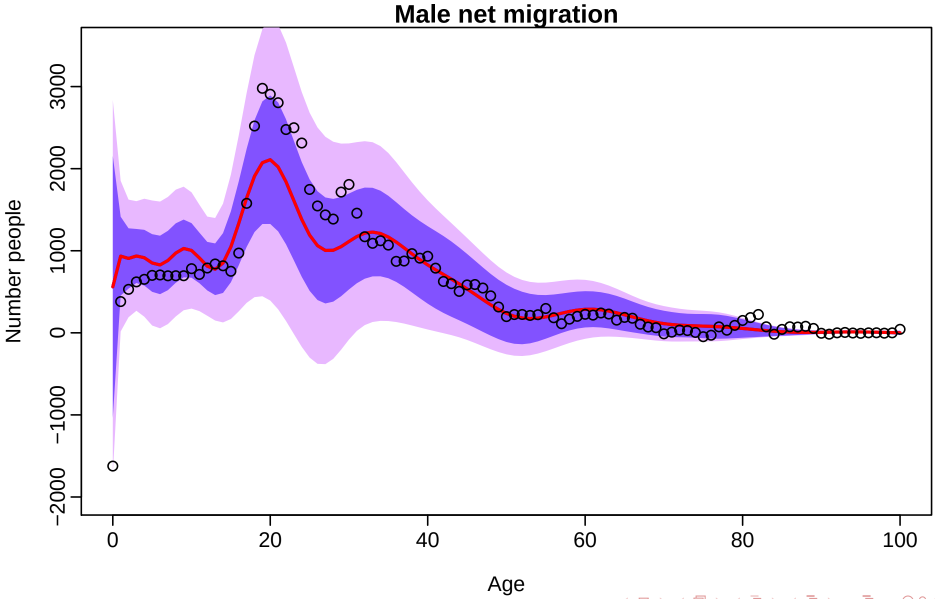
Migration: male



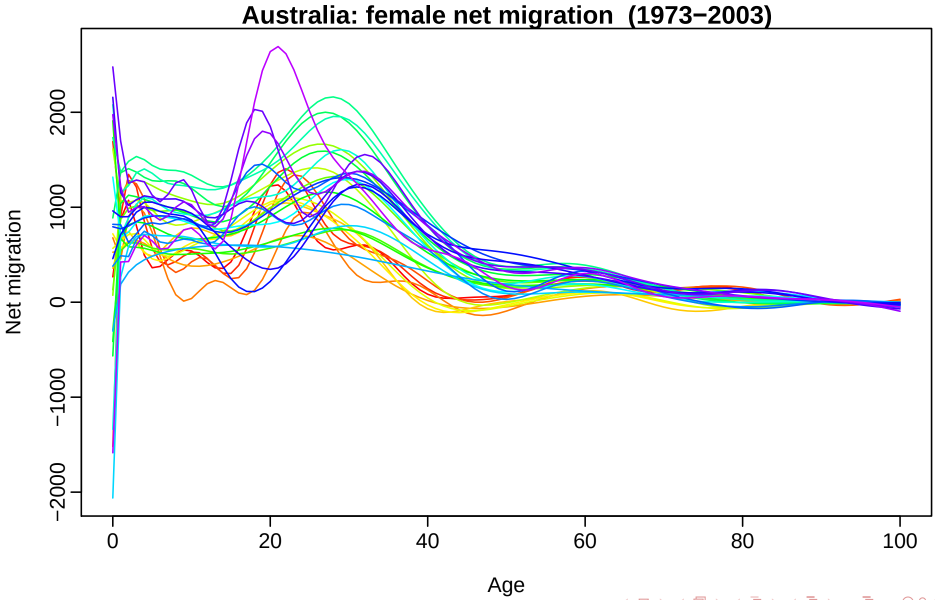
Migration: male



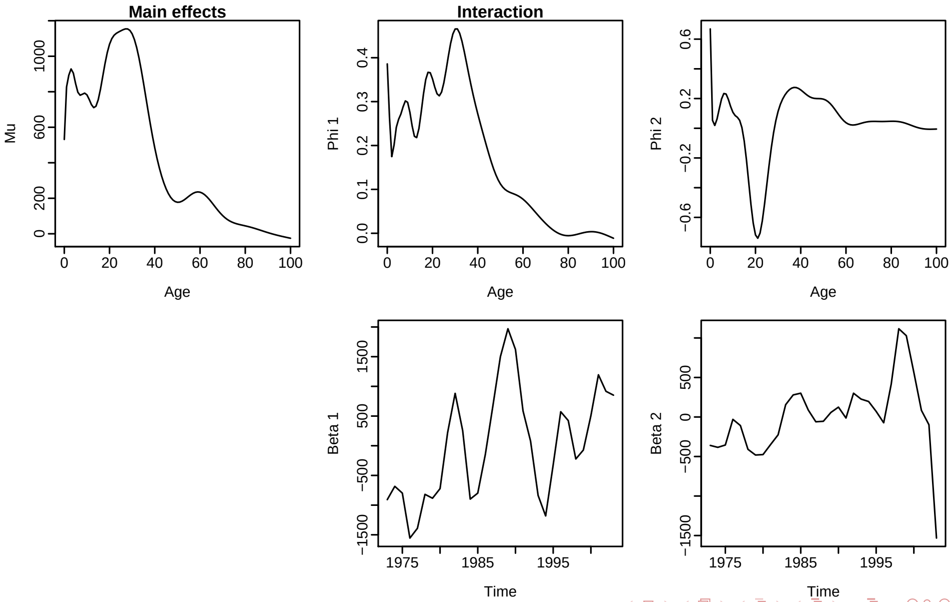
Migration: male



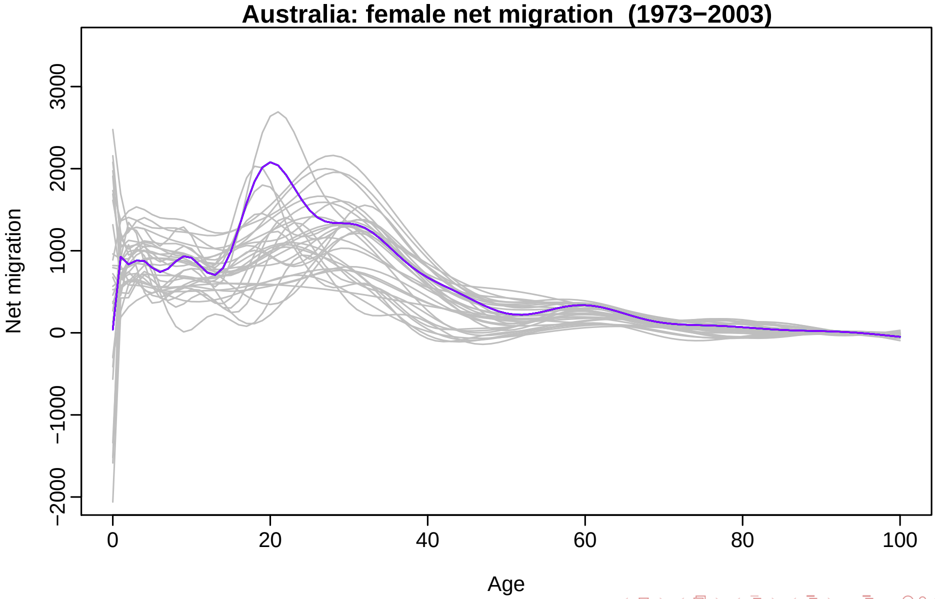
Migration: female



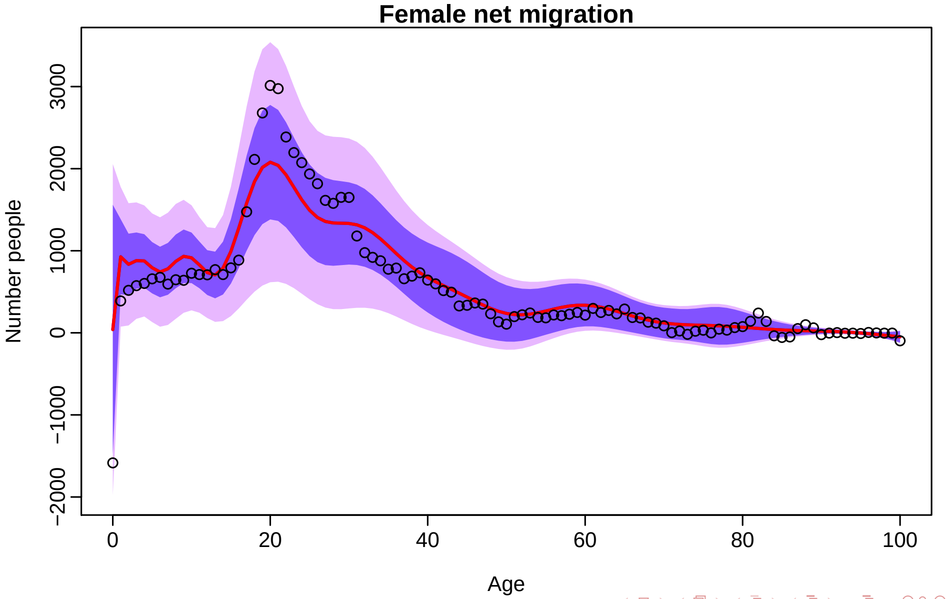
Migration: female



Migration: female



Migration: female



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

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

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

Simulation

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

Demographic growth-balance equation

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

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

- 10000 sample paths of population $P_t(x)$, deaths $D_t(x)$ and births $B_t(x)$ generated for $t = 2004, \dots, 2023$ and $x = 0, 1, 2, \dots,$.

Simulation

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

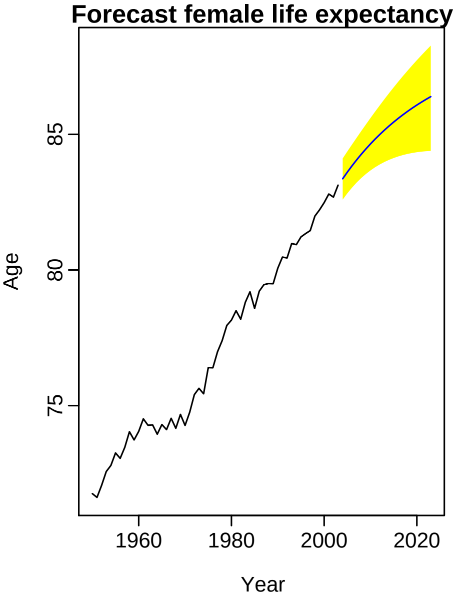
Demographic growth-balance equation

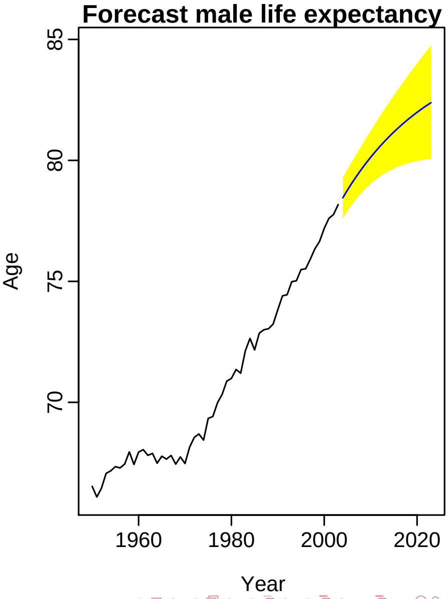
$$P_{t+1}(x+1) = P_t(x) - D_t(x, x+1) + G_t(x, x+1)$$
$$P_{t+1}(0) = B_t \qquad - D_t(B, 0) \qquad + G_t(B, 0)$$

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

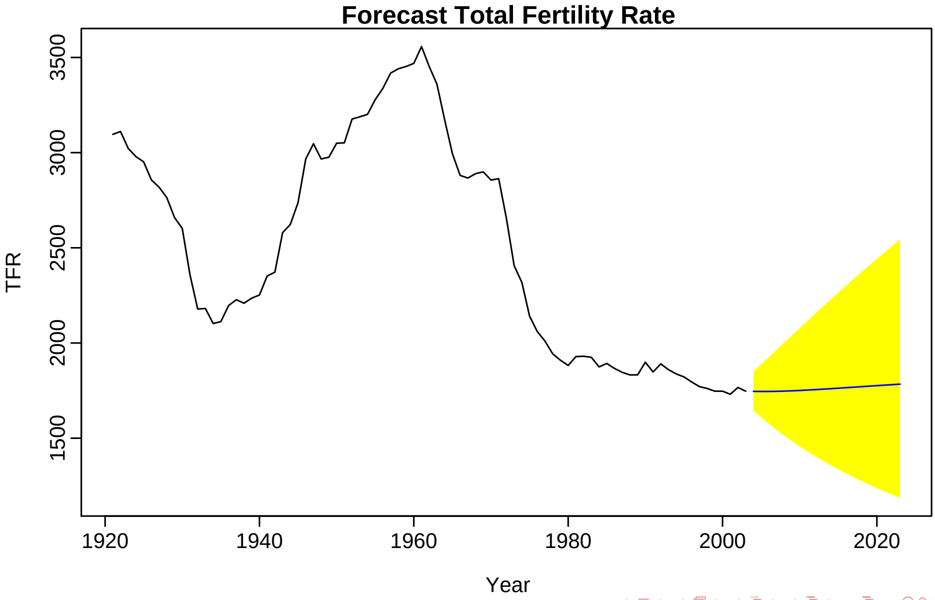
- 10000 sample paths of population $P_t(x)$, deaths $D_t(x)$ and births $B_t(x)$ generated for $t = 2004, \dots, 2023$ and $x = 0, 1, 2, \dots,$.
- This allows the computation of the empirical forecast distribution of any demographic quantity that is based on births, deaths and population numbers.

Forecasts of life expectancy at age 0

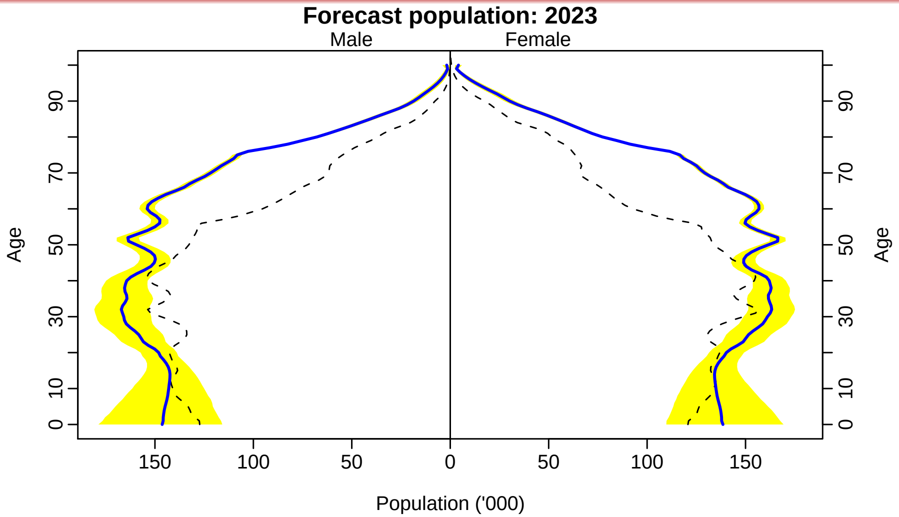




Forecasts of TFR

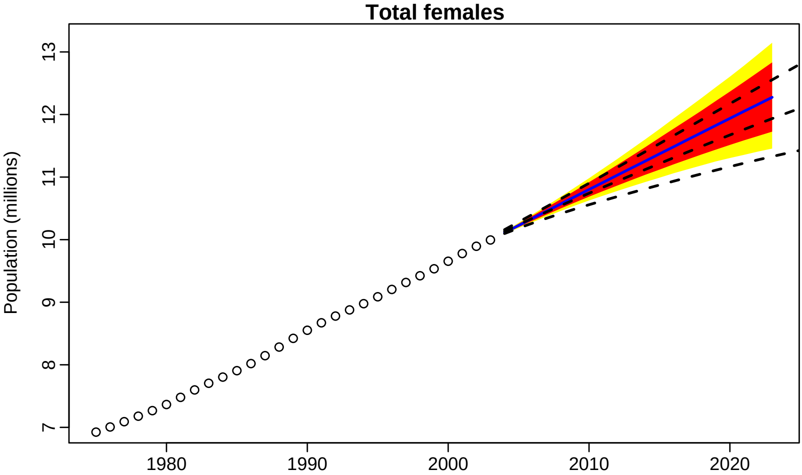


Population forecasts



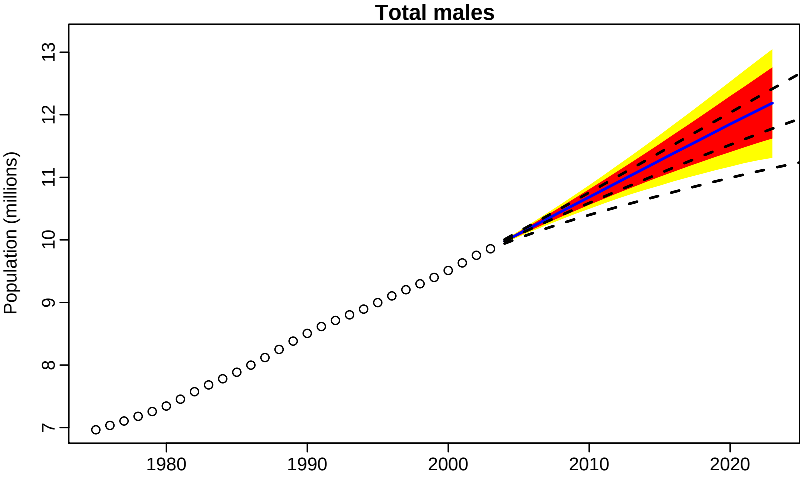
Forecast population pyramid for 2023, along with 80% prediction intervals. Dashed: actual population pyramid for 2003.

Population forecasts



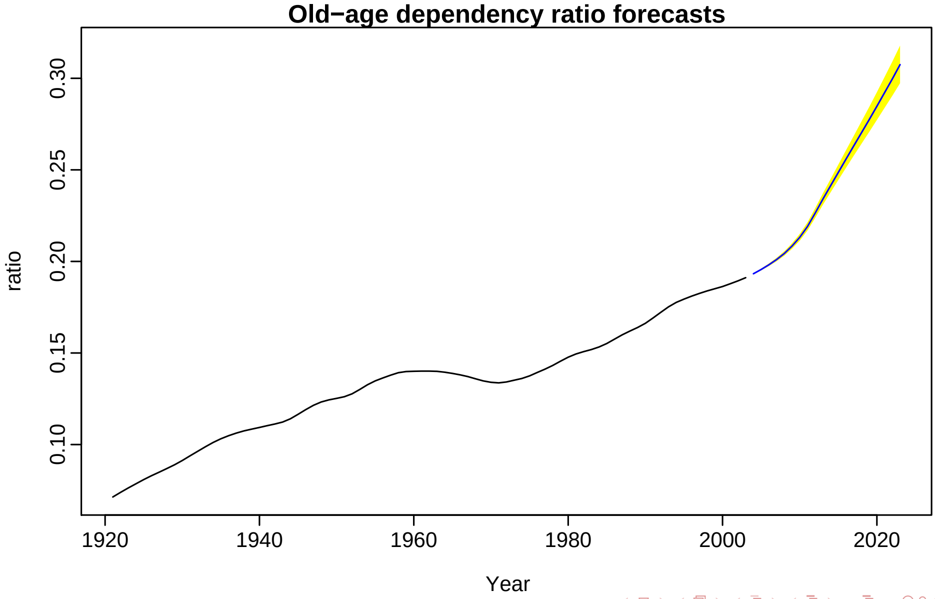
Twenty-year forecasts of total population along with 80% and 95% prediction intervals. Dashed lines show the ABS (2003) projections, series A, B and C.

Population forecasts



Twenty-year forecasts of total population along with 80% and 95% prediction intervals. Dashed lines show the ABS (2003) projections, series A, B and C.

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.)

Extensions

- We intend extending this to Australian states, and to other countries.

Extensions

- We intend extending this to Australian states, and to other countries.
- We intend extending the individual models to:

Extensions

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

Extensions

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

Extensions

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

Extensions

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

Software and papers:

Hyndman and Booth (2006). Working paper: “Stochastic population forecasts using functional data models for mortality, fertility and migration”.

www.robhyndman.info