

Forecasting the age structure of the scientific workforce in Australia

Rob J Hyndman & Kelly Nguyen

30 June 2025



Ideal labour force model

$$P_{x+1,t+1} = P_{x,t} - D_{x,t} - R_{x,t} + G_{x,t} - C_{x,t} + N_{x,t}$$

x = Age
 t = Year

$P_{x,t}$ = number of equivalent full-time workers

$D_{x,t}$ = number of deaths

$R_{x,t}$ = number of retirements

$G_{x,t}$ = number of graduates who work in science

$C_{x,t}$ = net number of people who have a career change

$N_{x,t}$ = net number of migrants

Pragmatic labour force model

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

x = Age

t = Year

$P_{x,t}$ = number of equivalent full-time workers

$q_{x,t}$ = probability of death

r_x = probability of retirement

g_x = proportion of graduates by age

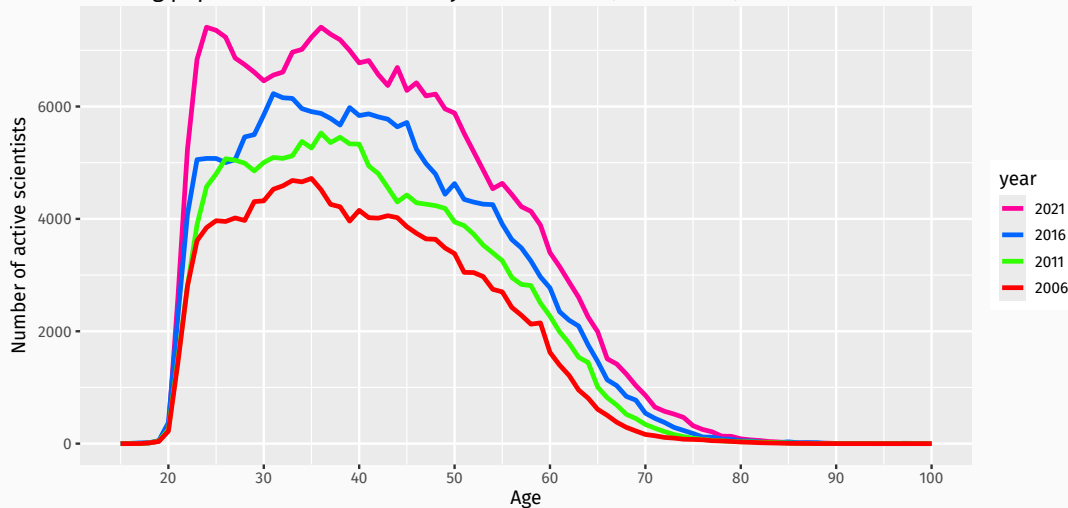
G_t = total number of graduates in science

$E_{x,t}$ = remainder

Working population: $P_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

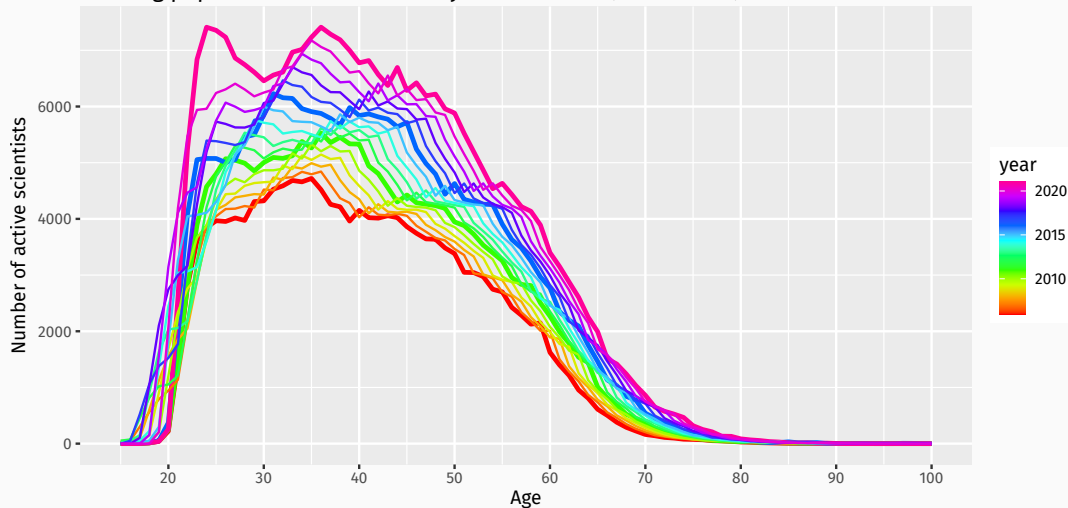
Working population: Natural and Physical Sciences (2006 – 2021)



Working population: $P_{x,t}$

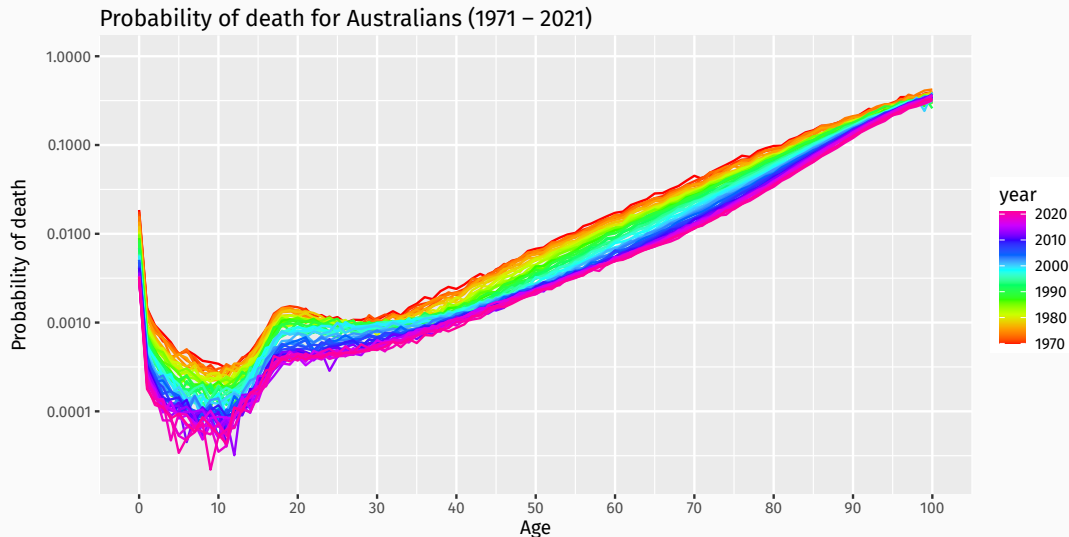
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Working population: Natural and Physical Sciences (2006 – 2021)



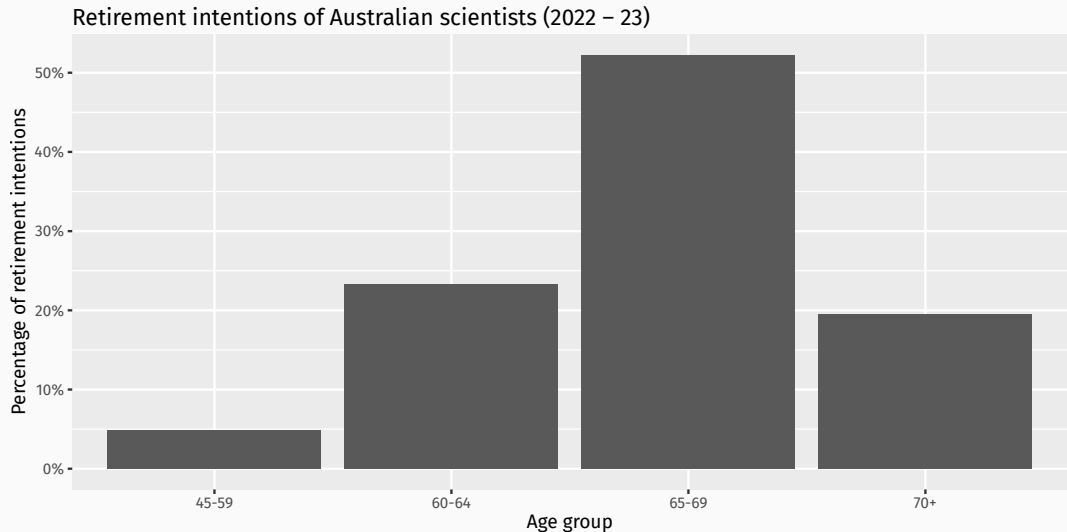
Death rates: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



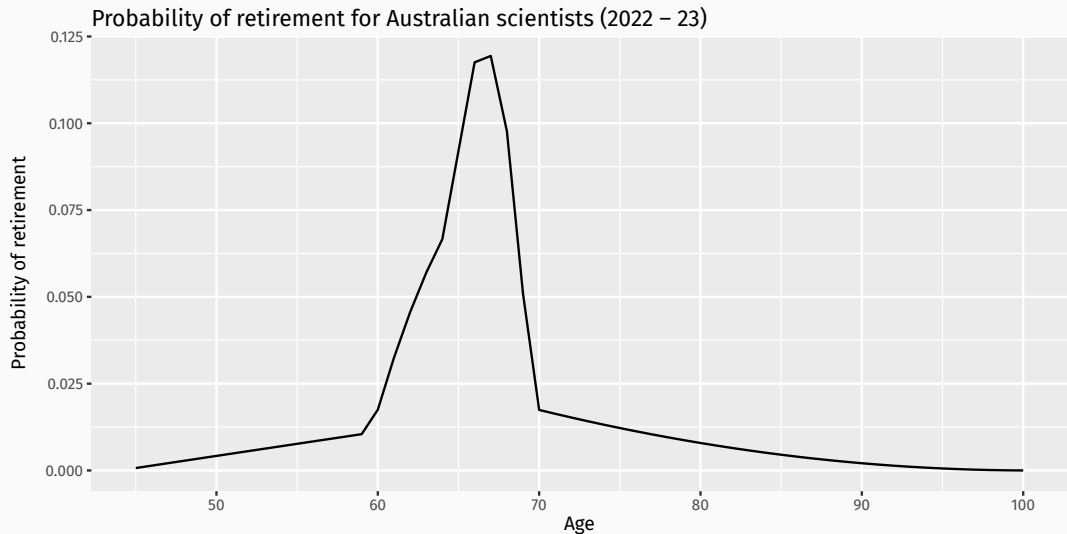
Retirement rates: r_x

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Retirement rates: r_x

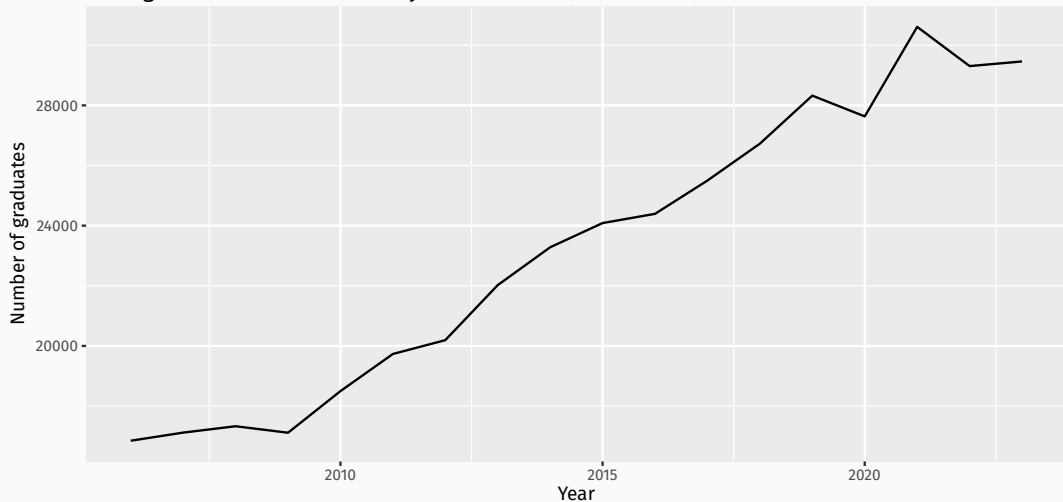
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Graduate completions: G_t

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

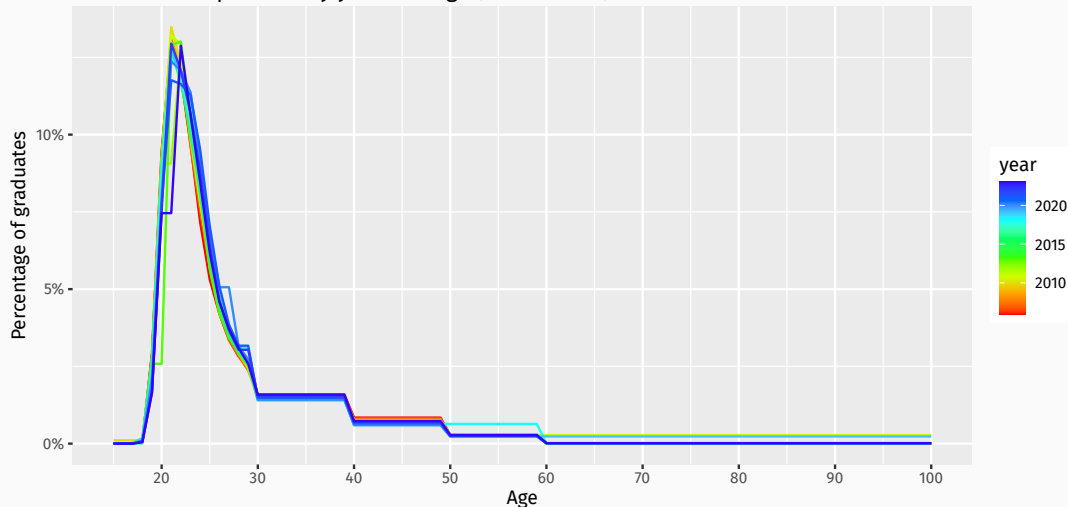
Total graduates: Natural and Physical Sciences (2006 – 2023)



Graduate completions: g_x

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

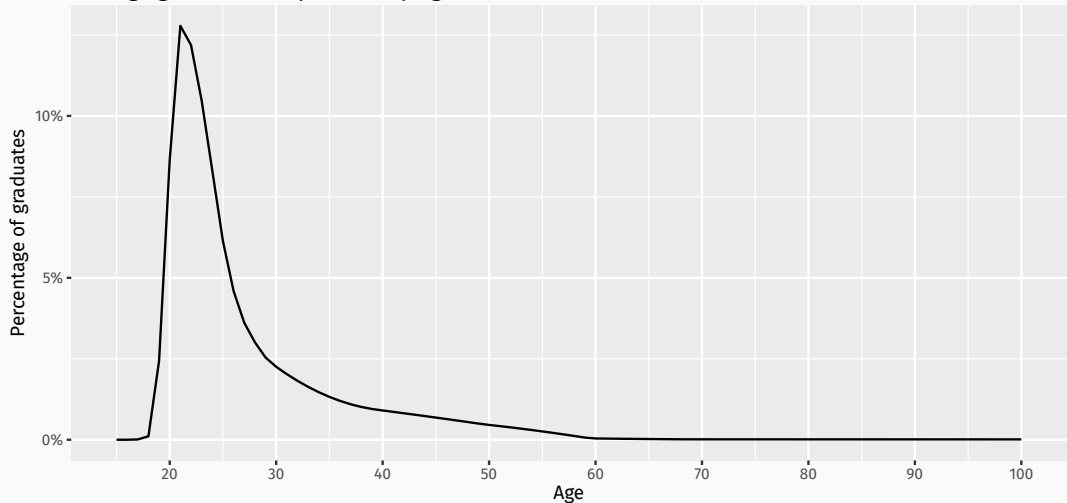
Graduate completions by year and age (2006 – 2023)



Graduate completions: g_x

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Average graduate completions by age (2006 – 2023)



Remainder: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

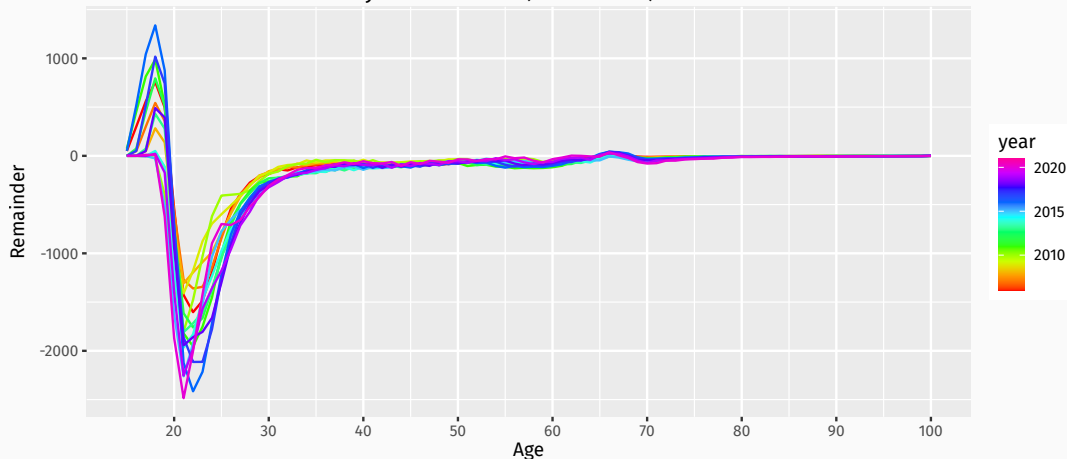
$$E_{x,t} = P_{x+1,t+1} - P_{x,t}(1 - q_{x,t} - r_x) - g_x G_t$$

Remainder: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

$$E_{x,t} = P_{x+1,t+1} - P_{x,t}(1 - q_{x,t} - r_x) - g_x G_t$$

Remainder: Natural and Physical Sciences (2006 – 2021)



$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

- G_t ARIMA model of total graduates by year
- $q_{x,t}$ functional time series model
- $E_{x,t}$ functional time series model

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

G_t ARIMA model of total graduates by year

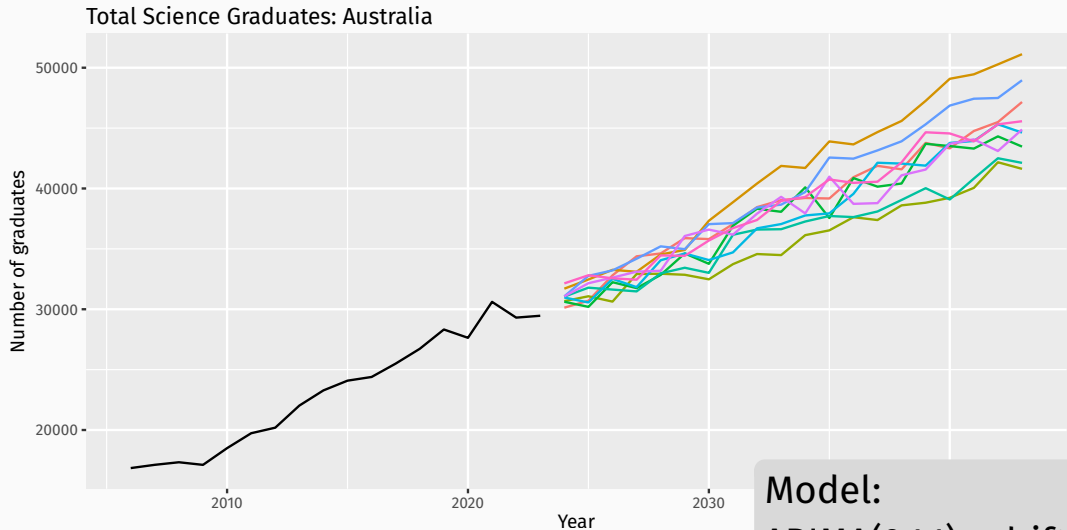
$q_{x,t}$ functional time series model

$E_{x,t}$ functional time series model

- Future sample paths of all components simulated to obtain probabilistic forecasts of $P_{x,t}$

Forecasting models: G_t

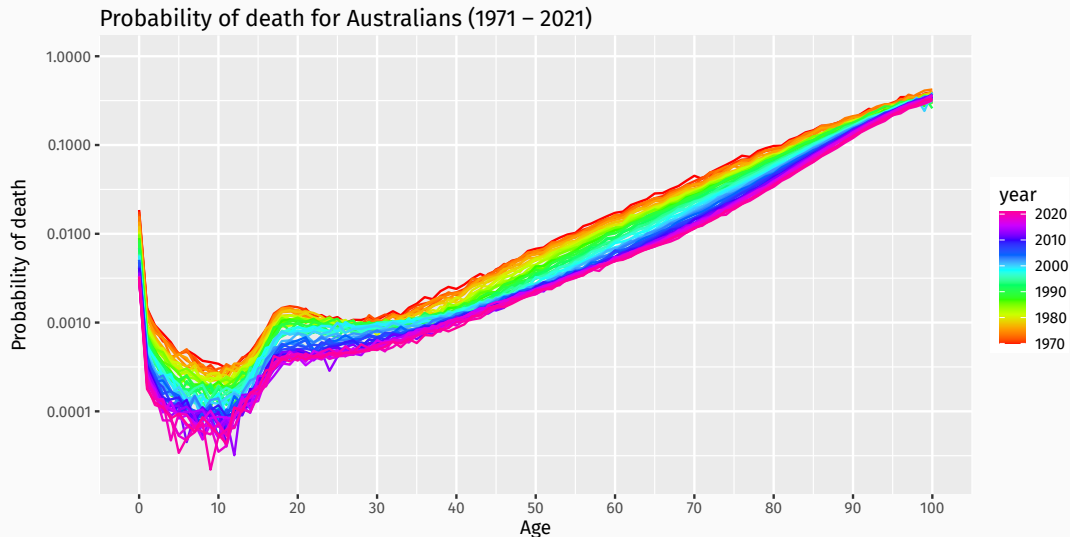
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Model:
ARIMA(0,1,1) + drift

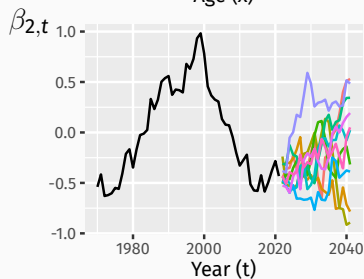
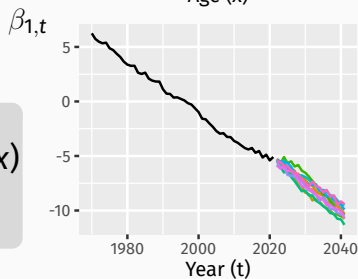
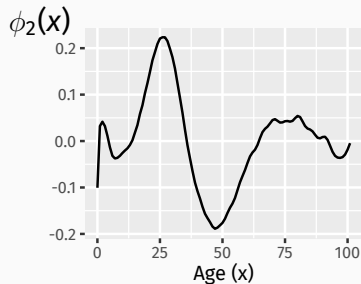
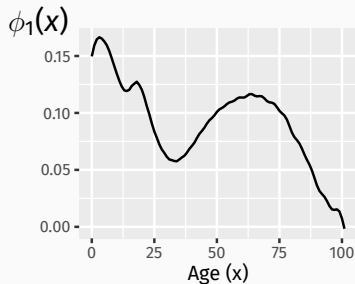
Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

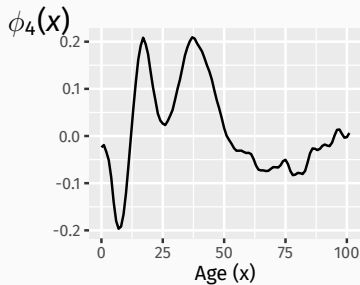
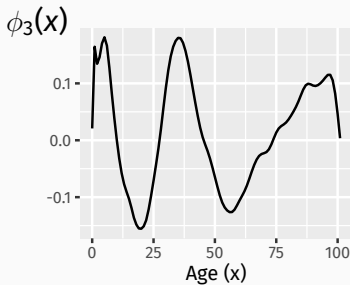
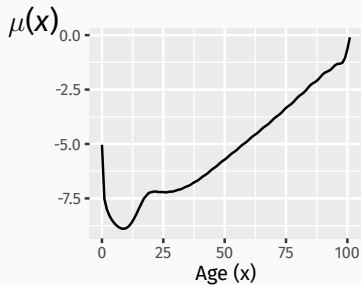


$$q_{x,t} = \mu(x) + \sum_{k=1}^6 \beta_{k,t} \phi_k(x) + \varepsilon_t(x)$$

$\beta_{k,t} \sim \text{ARIMA}$

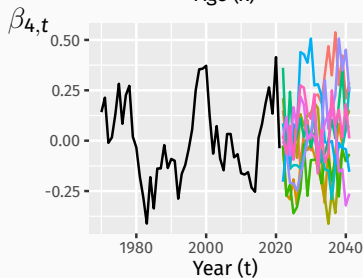
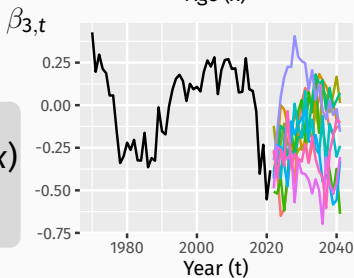
Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



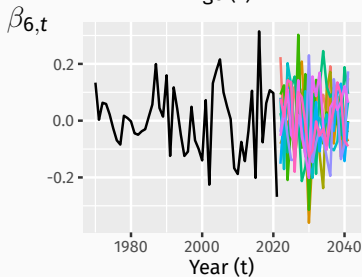
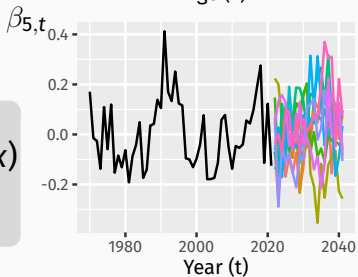
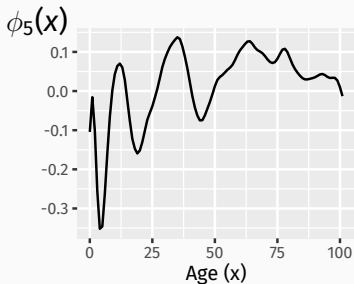
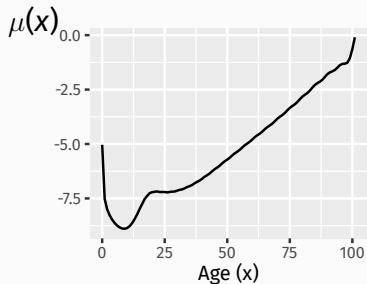
$$q_{x,t} = \mu(x) + \sum_{k=1}^6 \beta_{k,t} \phi_k(x) + \varepsilon_t(x)$$

$\beta_{k,t} \sim \text{ARIMA}$



Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

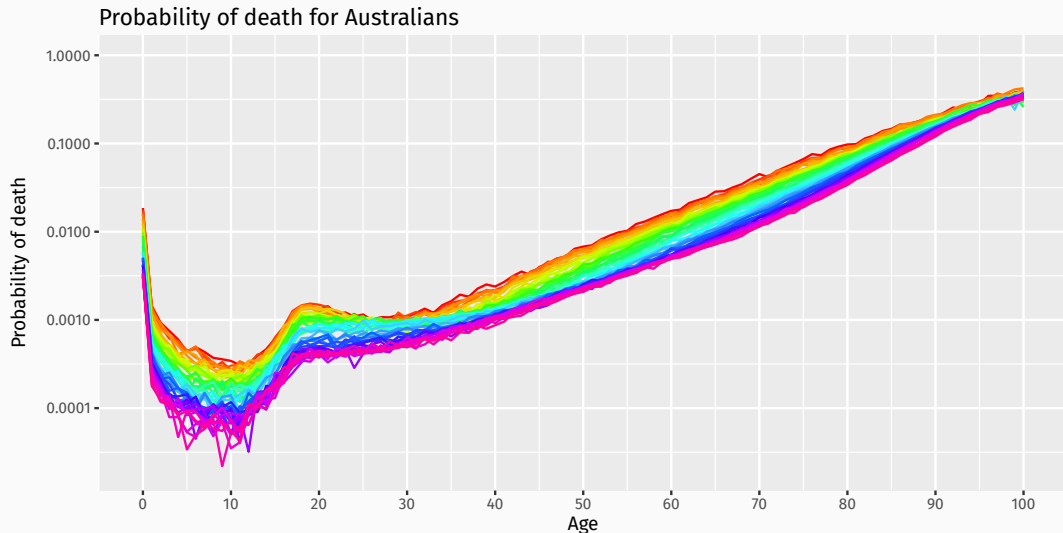


$$q_{x,t} = \mu(x) + \sum_{k=1}^6 \beta_{k,t} \phi_k(x) + \varepsilon_{t,x}$$

$\beta_{k,t} \sim \text{ARIMA}$

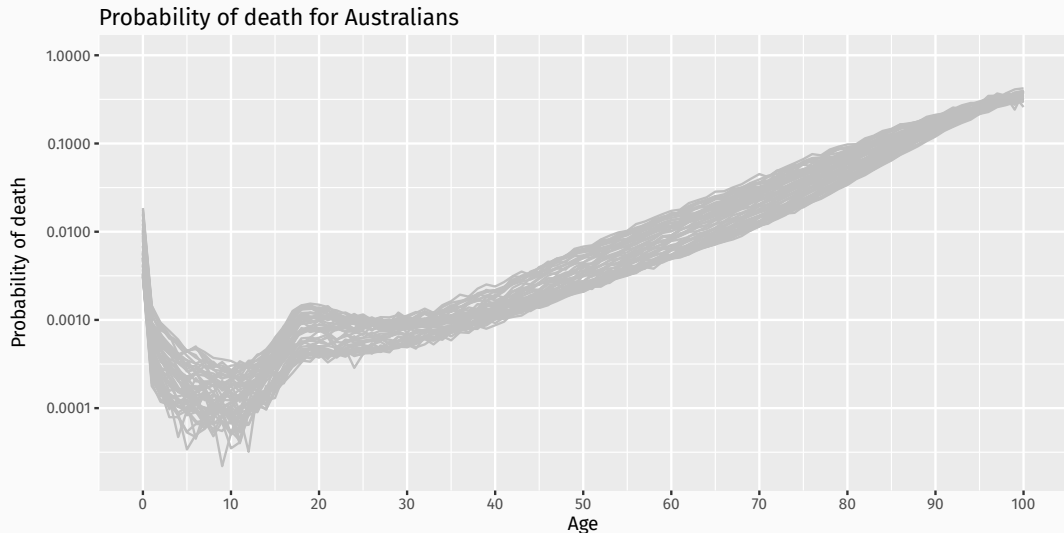
Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



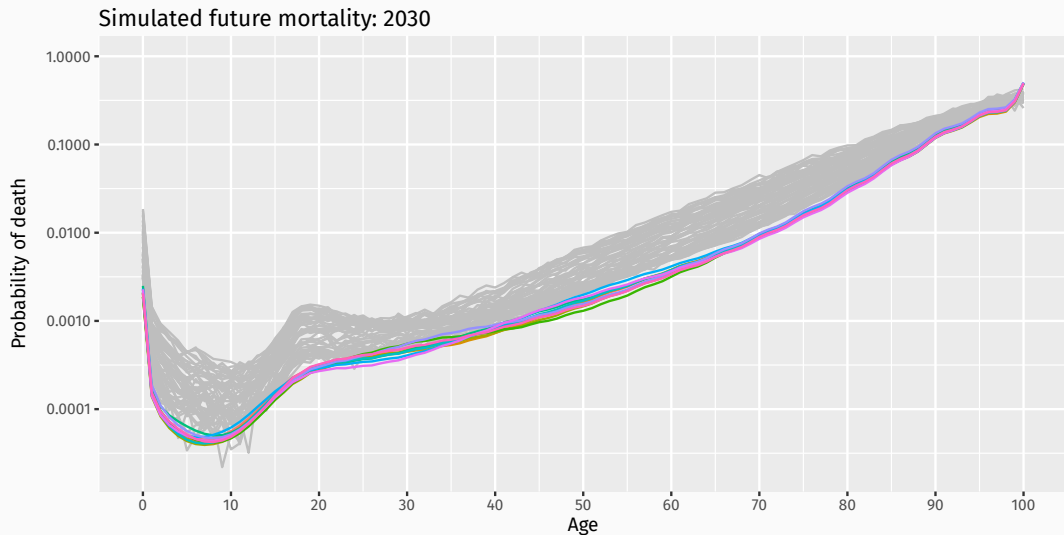
Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



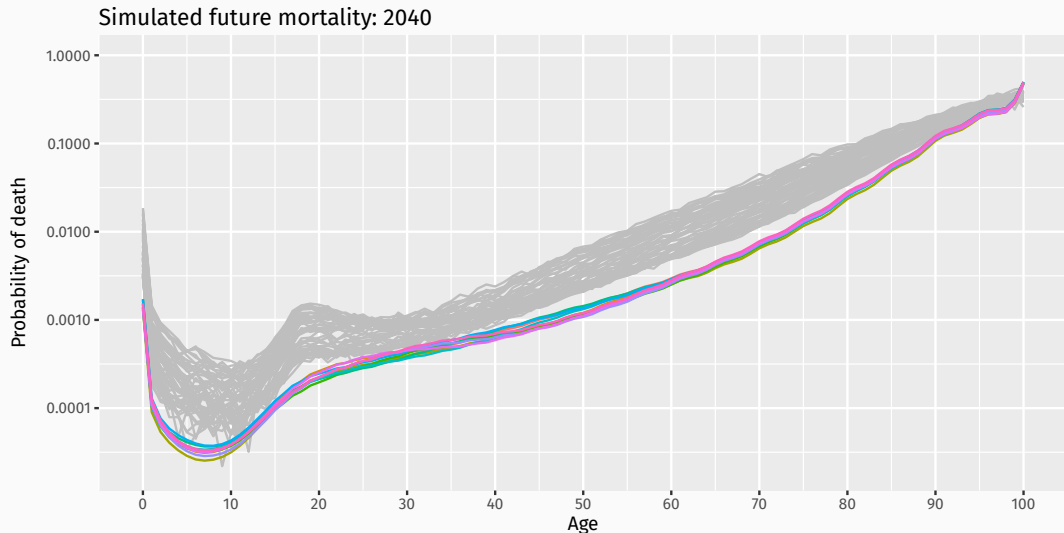
Forecasting models: $q_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Forecasting models: $q_{x,t}$

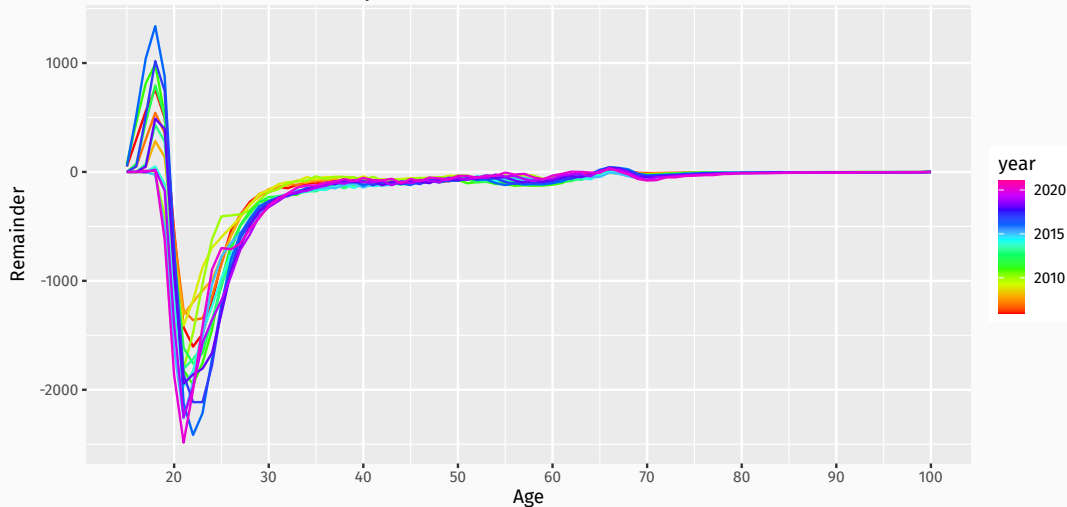
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Remainder: $E_{x,t}$

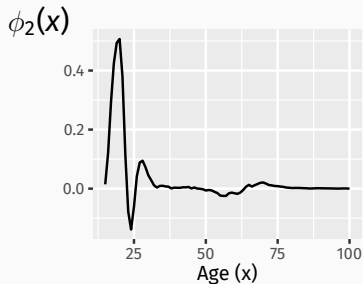
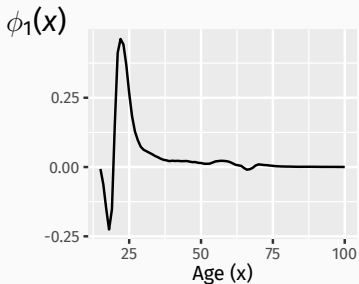
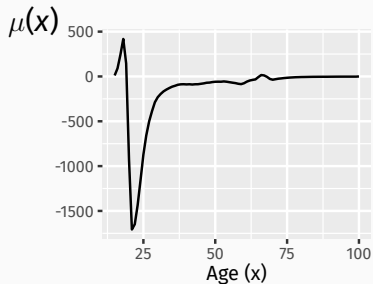
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Remainder: Natural and Physical Sciences (2006 – 2021)



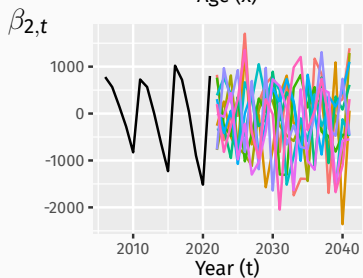
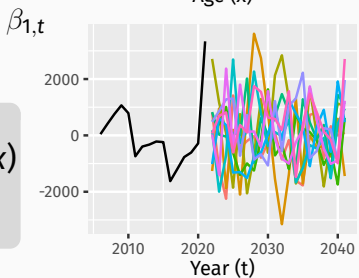
Forecasting models: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



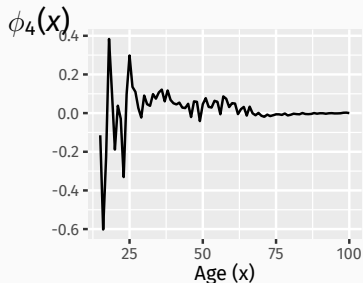
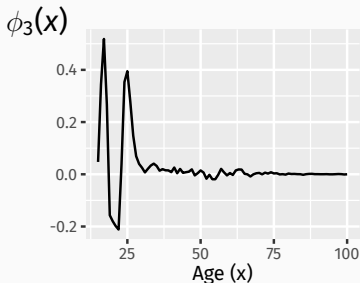
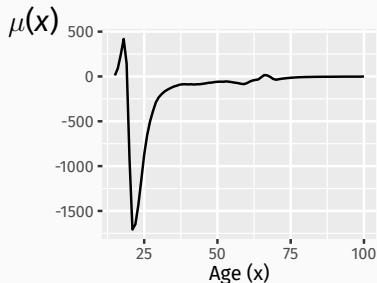
$$E_{x,t} = \mu(x) + \sum_{k=1}^6 \beta_{k,t} \phi_k(x) + \varepsilon_t(x)$$

$\beta_{k,t} \sim \text{ARIMA}$



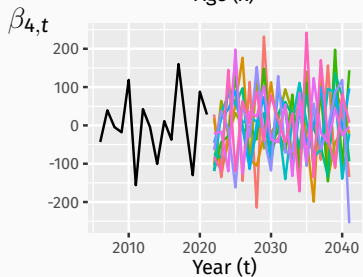
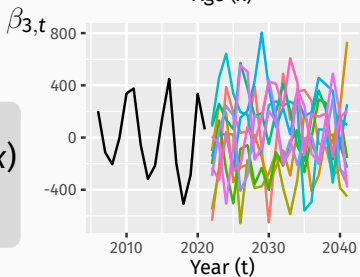
Forecasting models: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



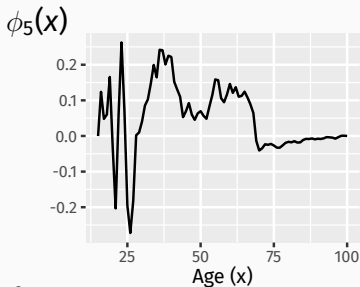
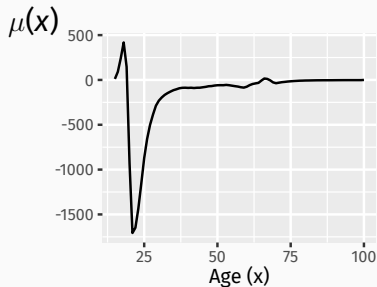
$$E_{x,t} = \mu(x) + \sum_{k=1}^6 \beta_{k,t} \phi_k(x) + \varepsilon_t(x)$$

$\beta_{k,t} \sim \text{ARIMA}$



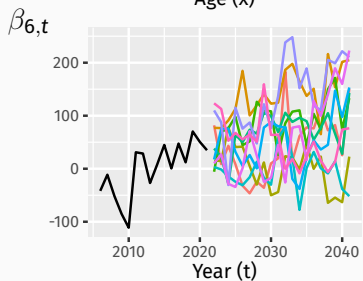
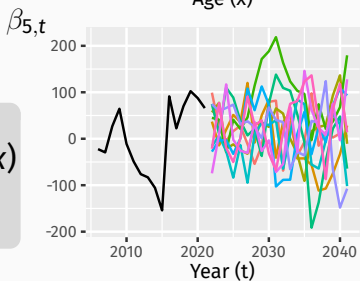
Forecasting models: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



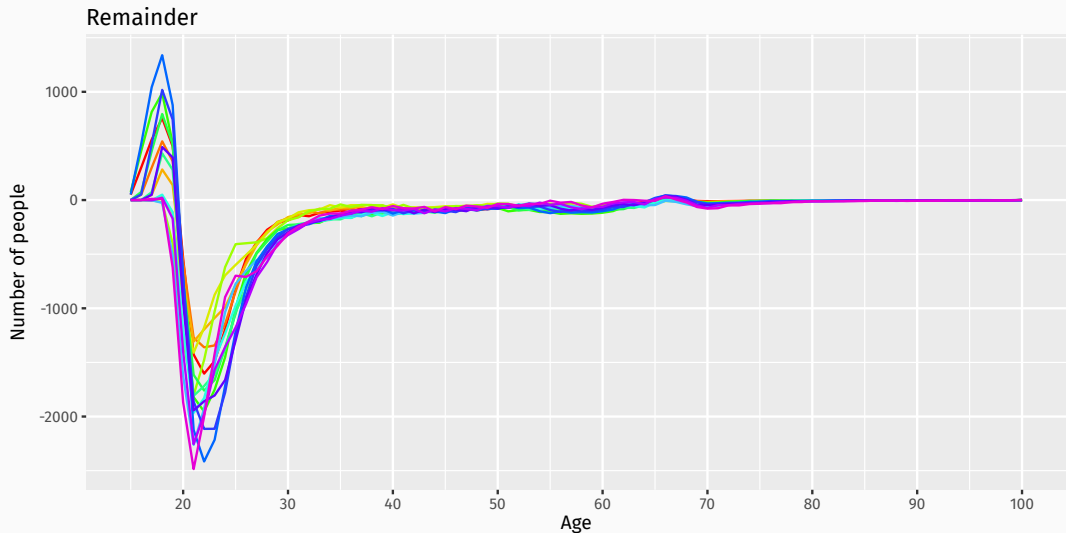
$$E_{x,t} = \mu(x) + \sum_{k=1}^6 \beta_{k,t} \phi_k(x) + \varepsilon_t(x)$$

$\beta_{k,t} \sim \text{ARIMA}$



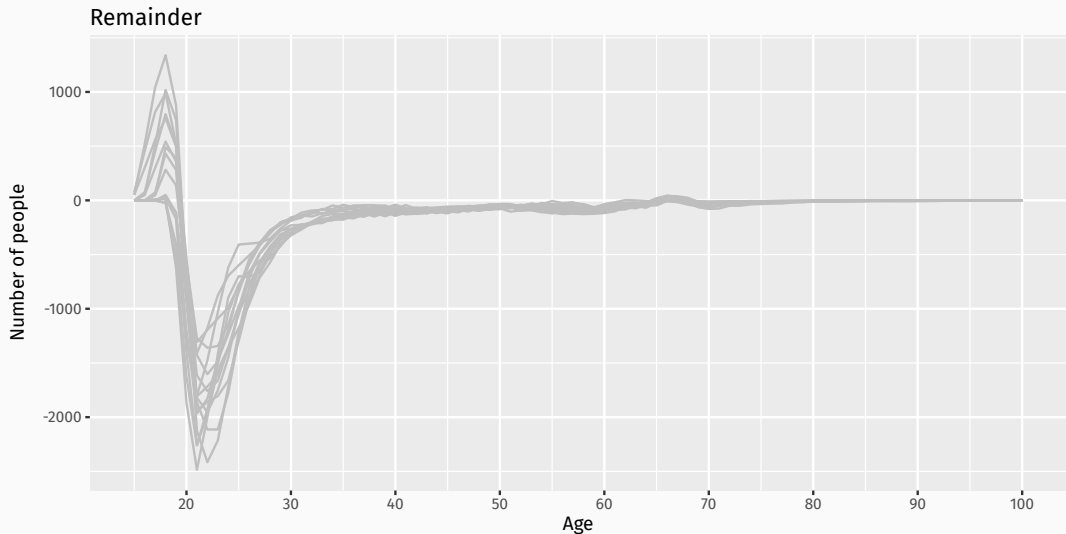
Forecasting models: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Forecasting models: $E_{x,t}$

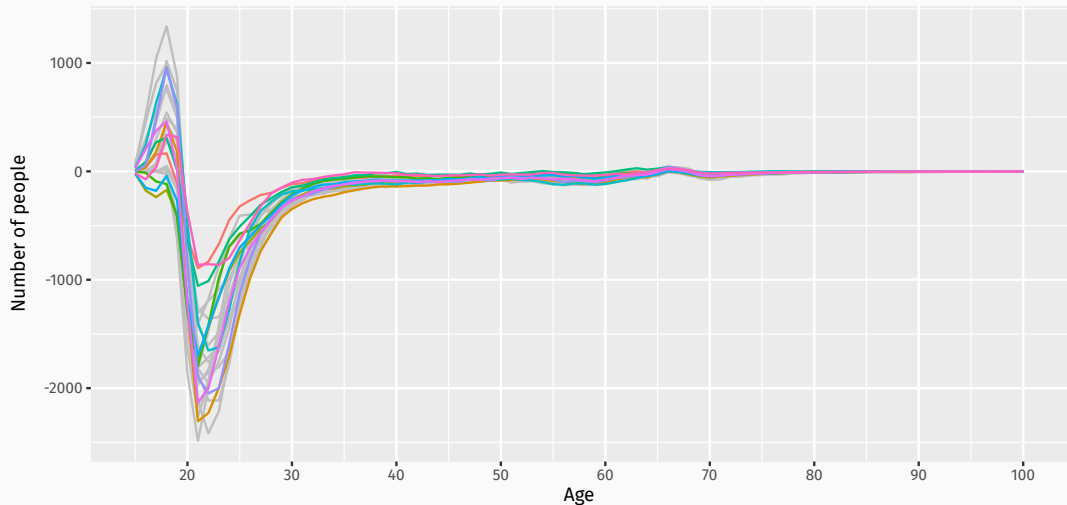
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Forecasting models: $E_{x,t}$

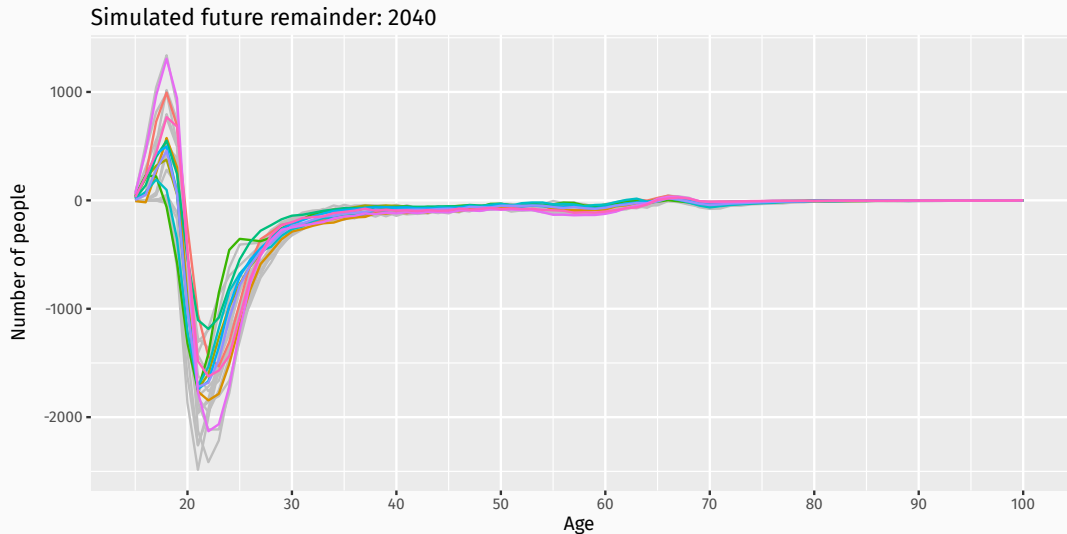
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Simulated future remainder: 2030



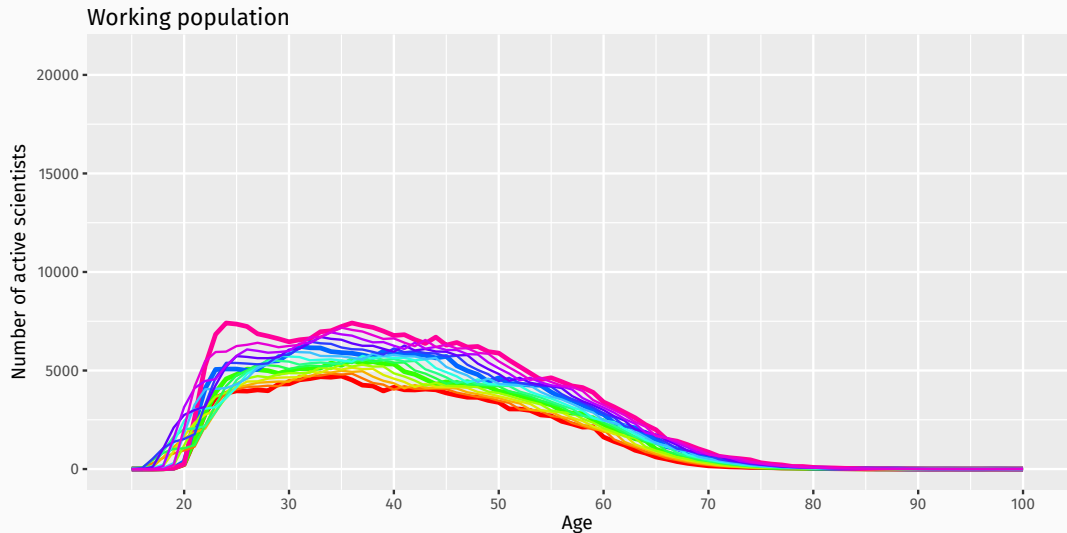
Forecasting models: $E_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



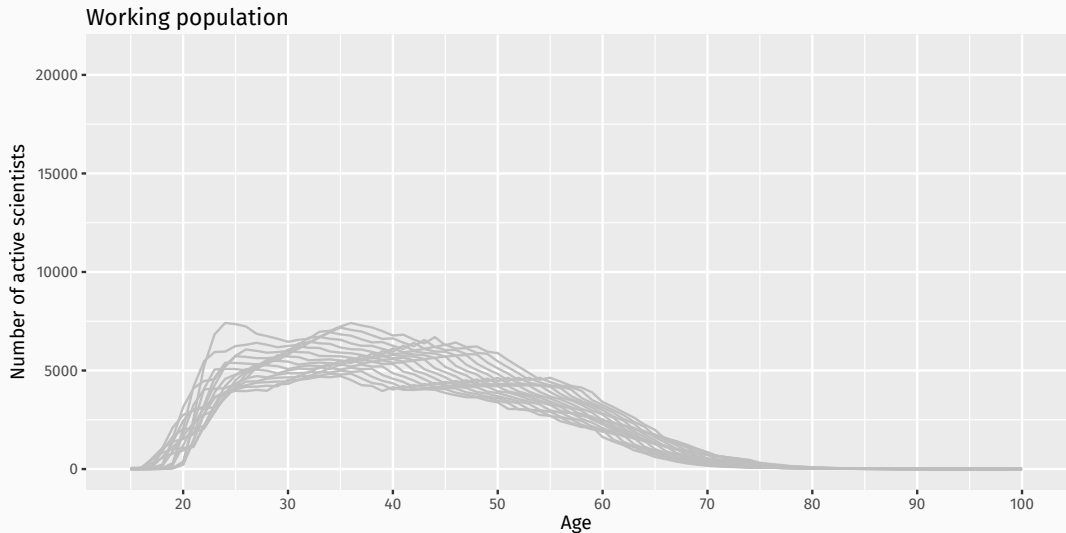
Final forecasts: $P_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Final forecasts: $P_{x,t}$

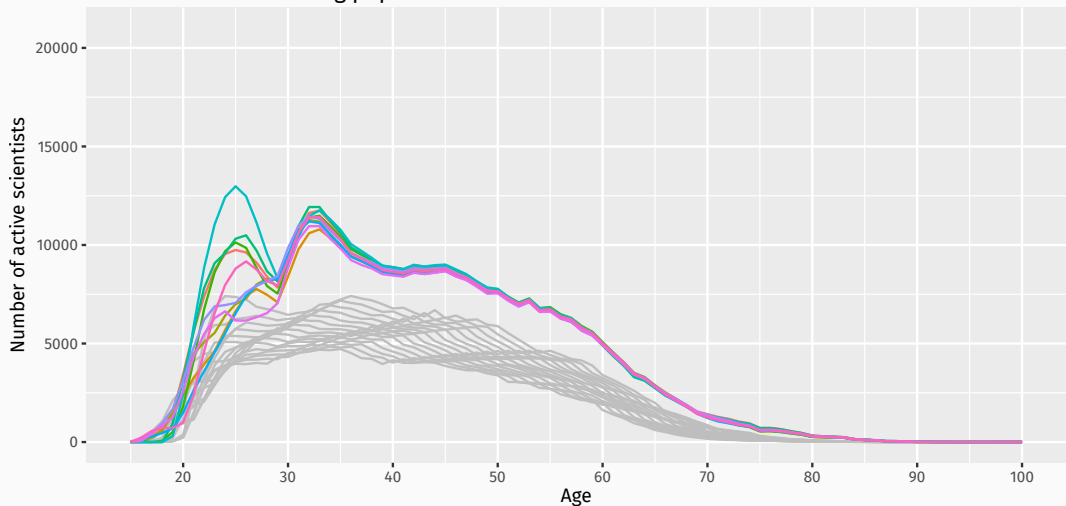
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$



Final forecasts: $P_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Simulated future working population: 2030



Final forecasts: $P_{x,t}$

$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Simulated future working population: 2040

