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
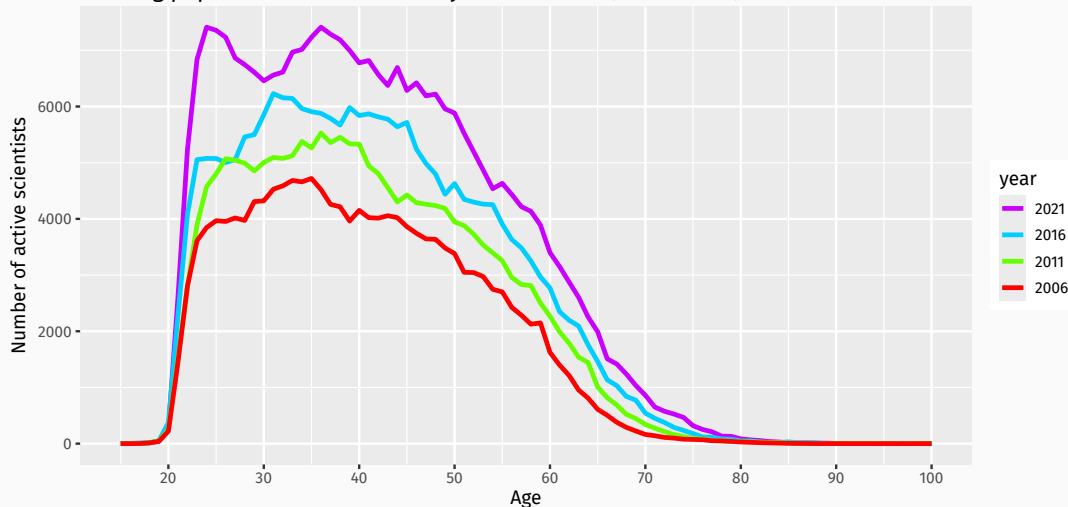
g_x = proportion of graduates by age

G_t = total number of graduates in science

$E_{x,t}$ = remainder

Working population: $P_{x,t}$

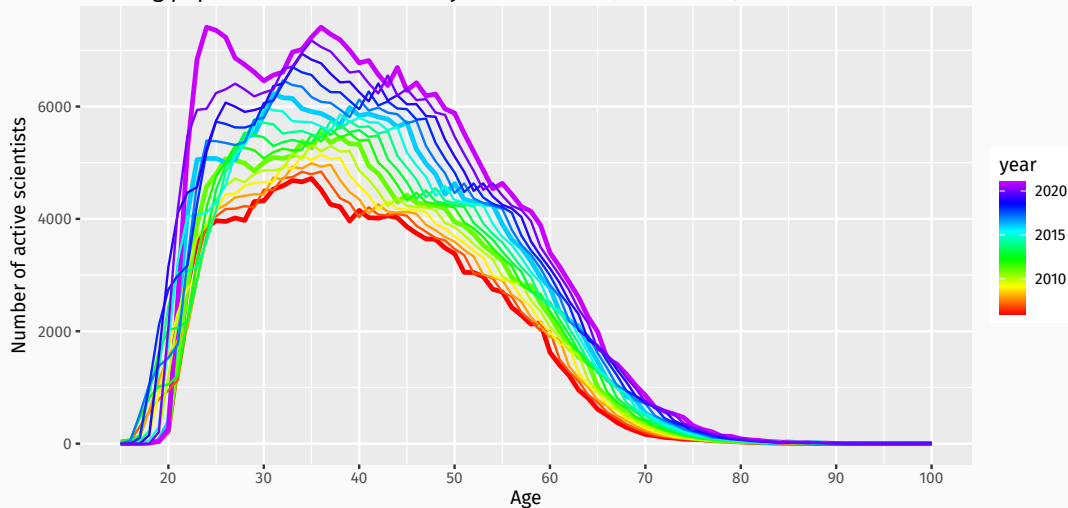$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Working population: Natural and Physical Sciences (2006 – 2021)
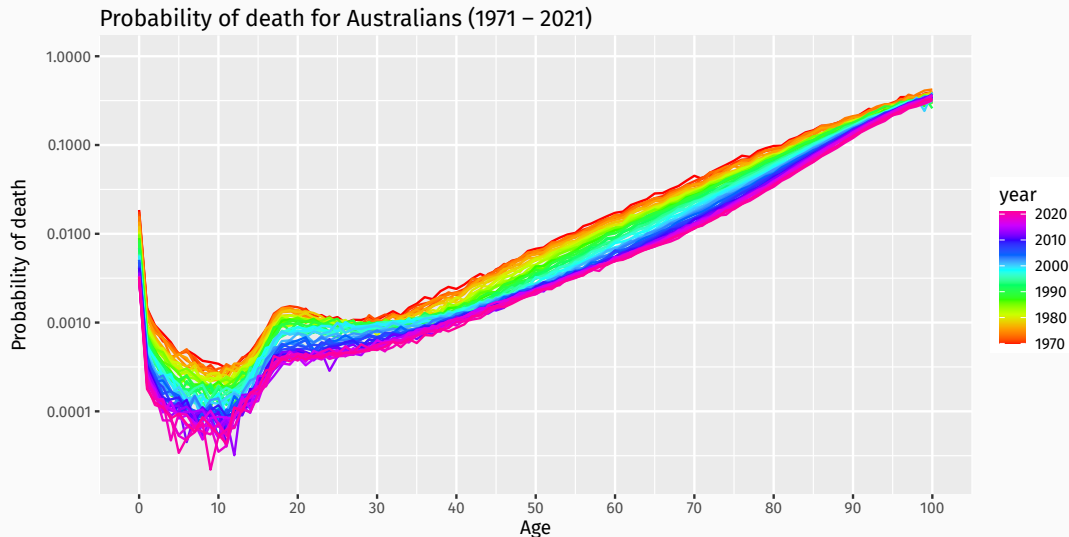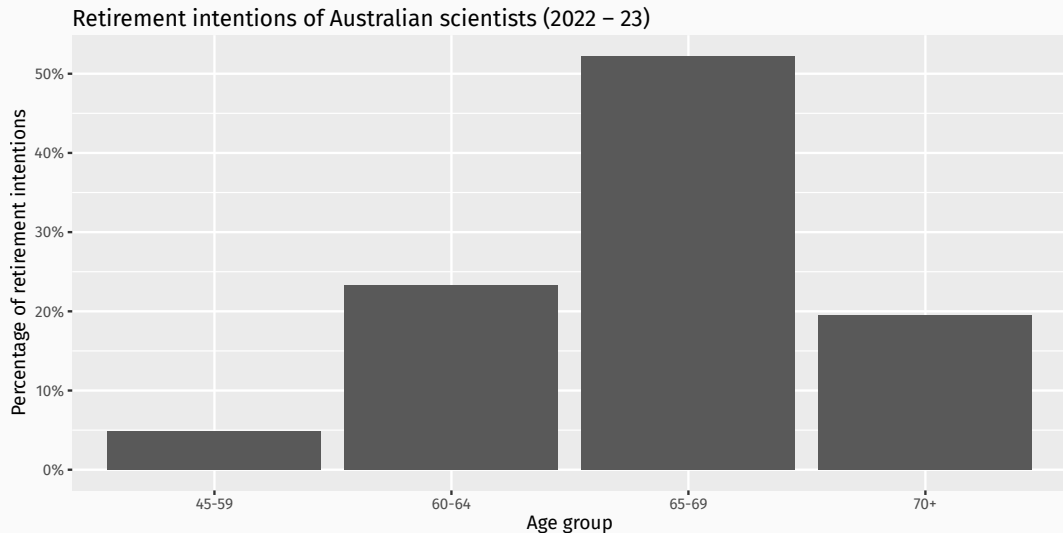
Working population: $P_{x,t}$

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

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

Death probability: $q_{x,t}$

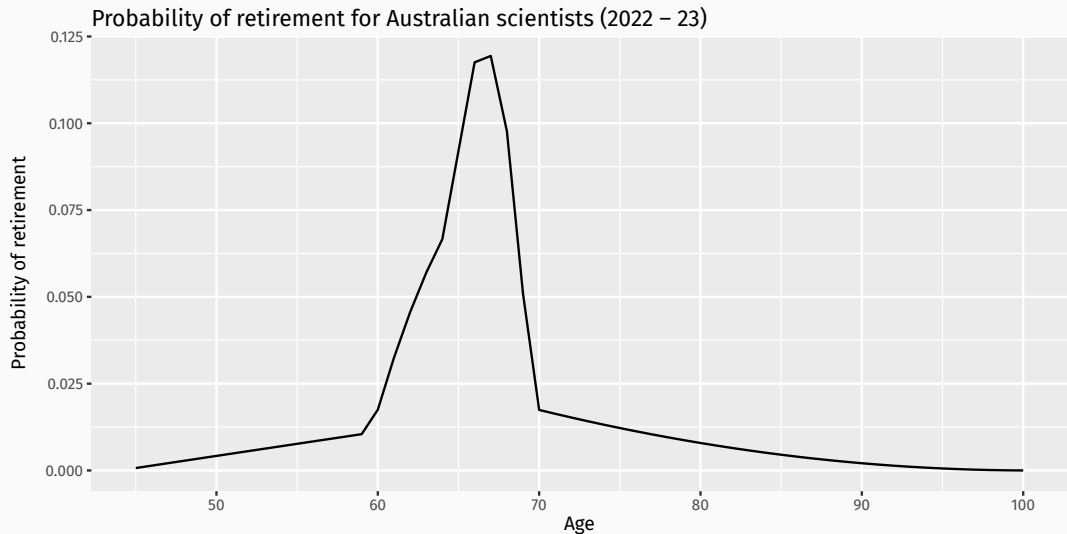$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Probability of death for Australians (1971 – 2021)

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 intentions of Australian scientists (2022 – 23)

Retirement rates: r_x

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

Probability of retirement for Australian scientists (2022 – 23)

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: (2006 – 2023)
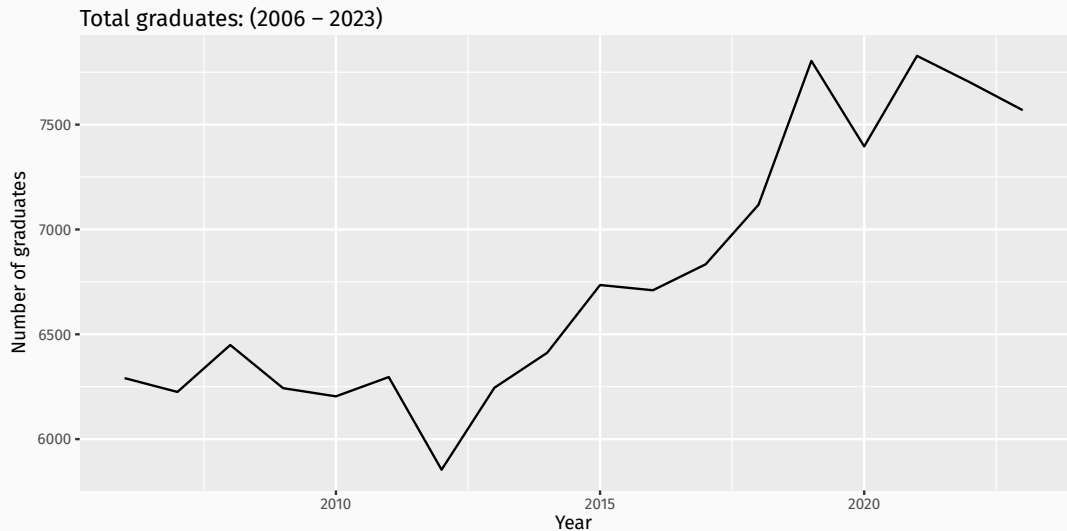
Graduate completions: g_x

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

Graduate completions by year and age (2006 – 2023)
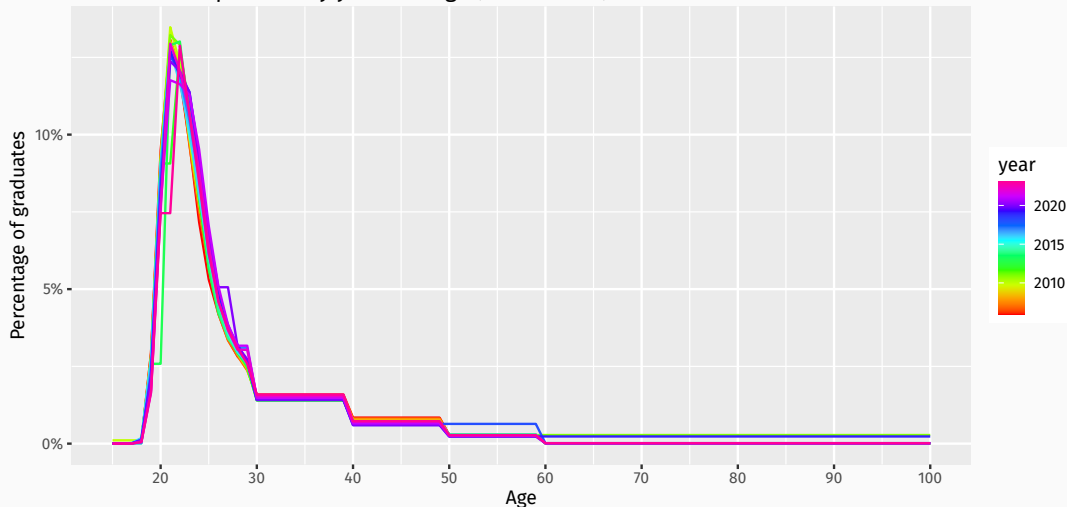
Graduate completions: g_x

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

Average graduate completions by age (2006 – 2023)
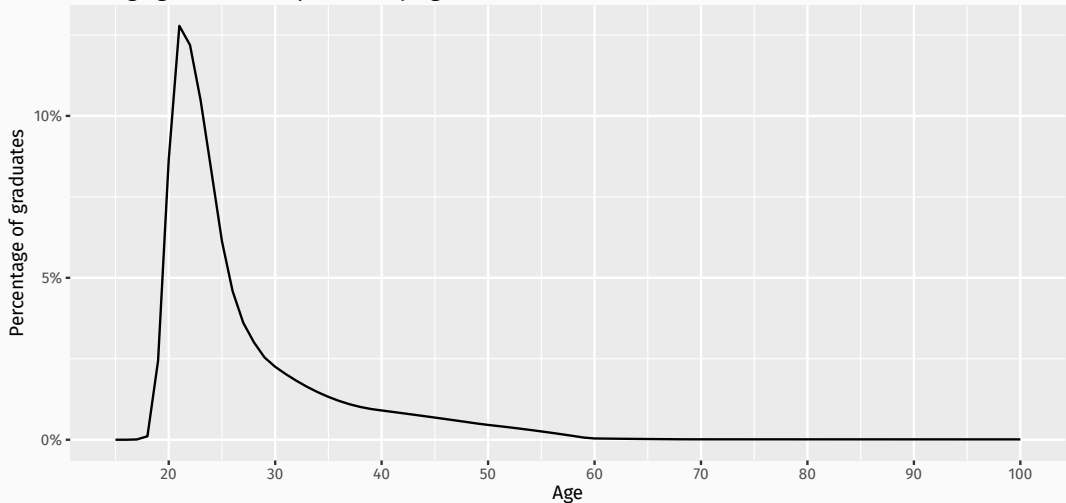
$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

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

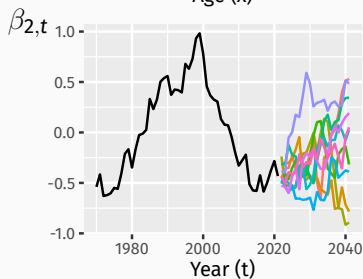# Remainder: $E_{x,t}$

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

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

Remainder: Natural and Physical Sciences (2006 – 2021)
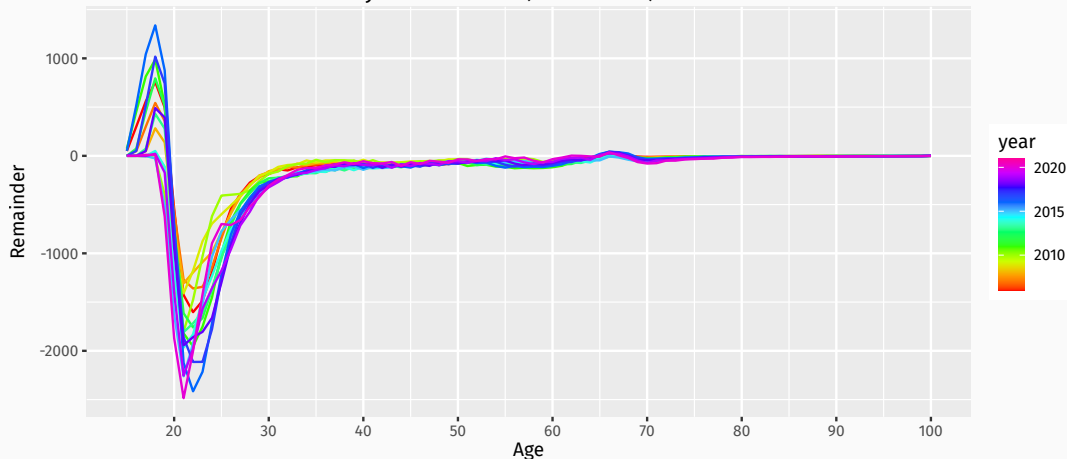
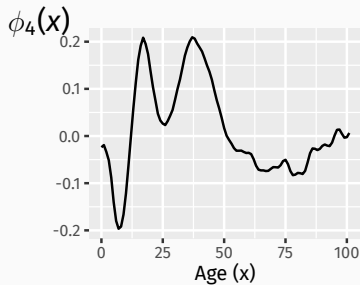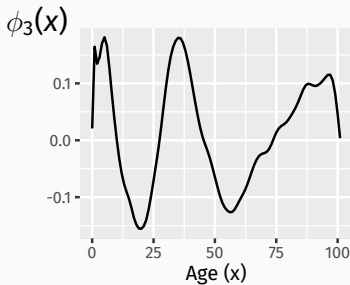$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

G_t ARIMA model of total graduates by year

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

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

Forecasting models

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

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

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

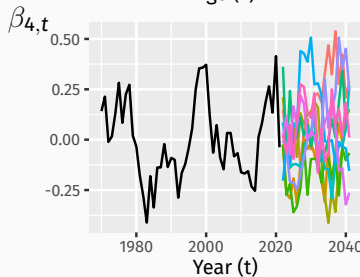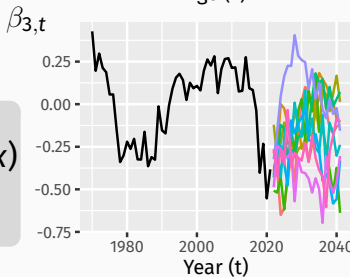Total Science Graduates: Australia

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

15

Forecasting models: $q_{x,t}$

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

Probability of death for Australians (1971 – 2021)

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: $q_{x,t}$

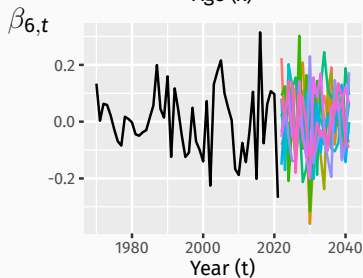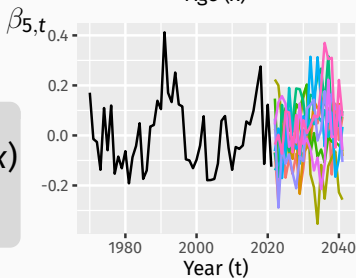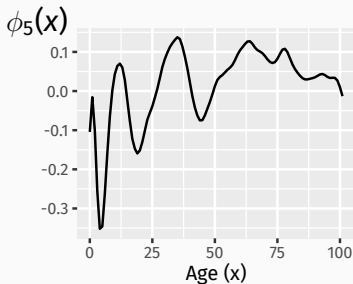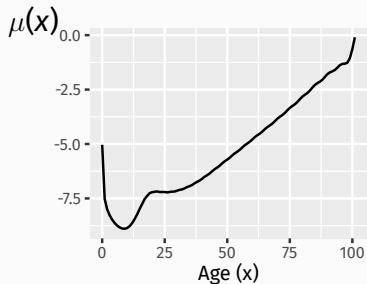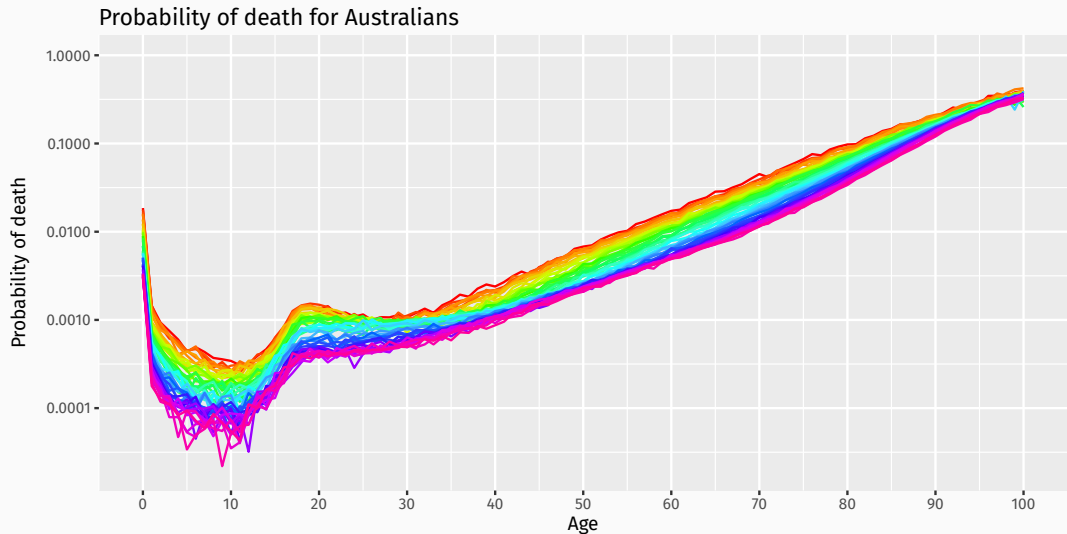$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

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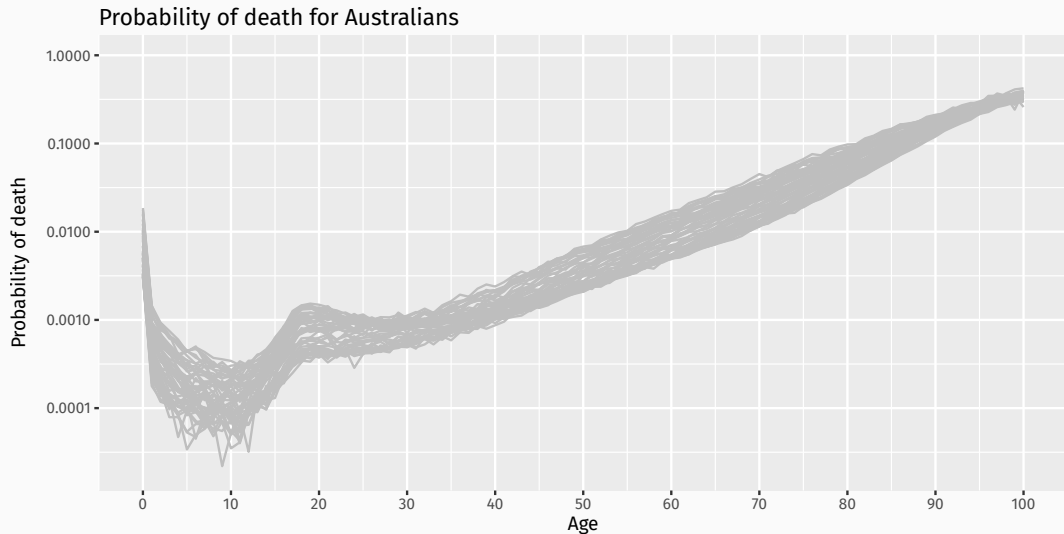
$$\beta_{k,t} \sim \text{ARIMA}$$

Forecasting models: $q_{x,t}$

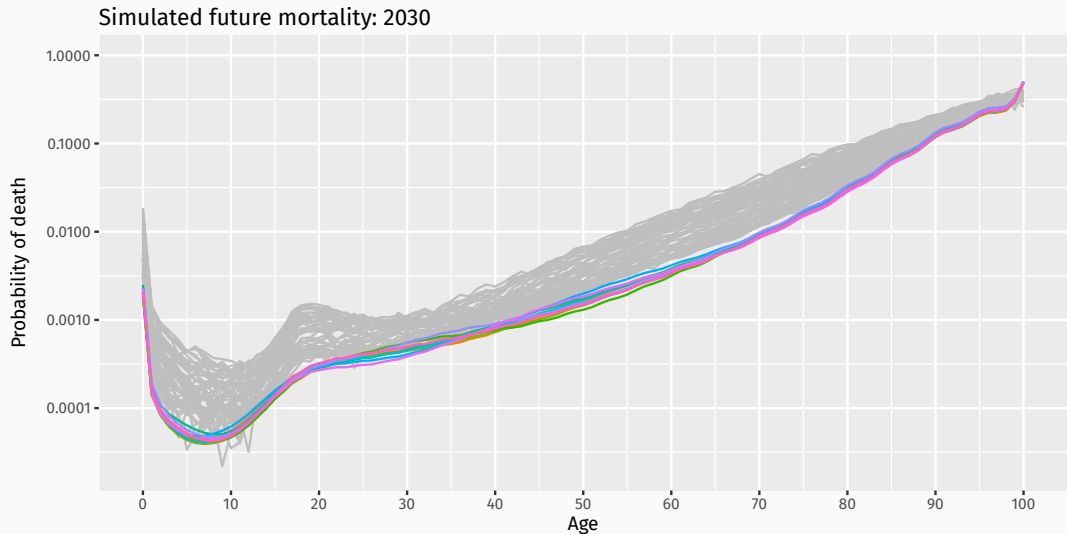$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

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

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

Forecasting models: $q_{x,t}$

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

Probability of death for Australians
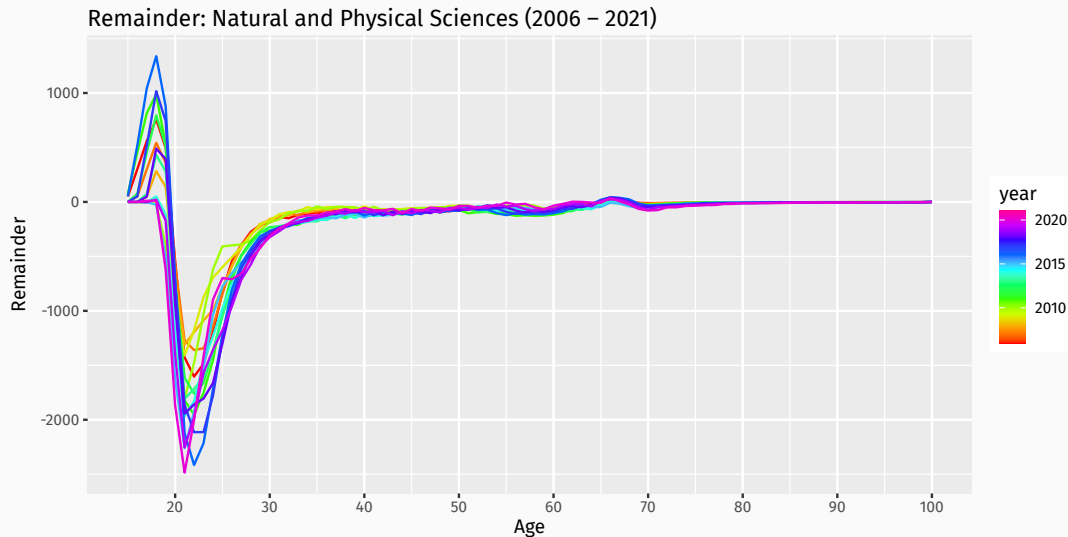
Forecasting models: $q_{x,t}$

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

Probability of death for Australians

Forecasting models: $q_{x,t}$

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

Simulated future mortality: 2030

Forecasting models: $q_{x,t}$

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

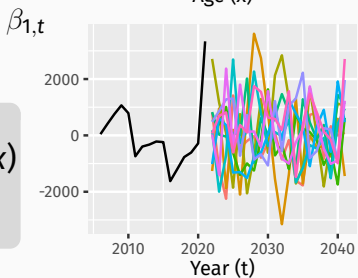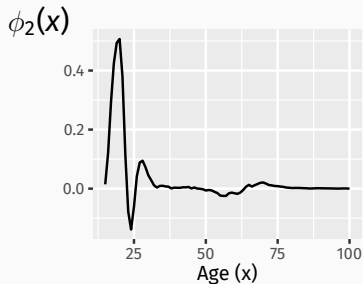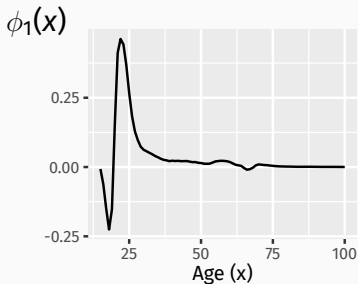Simulated future mortality: 2040

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: $E_{x,t}$

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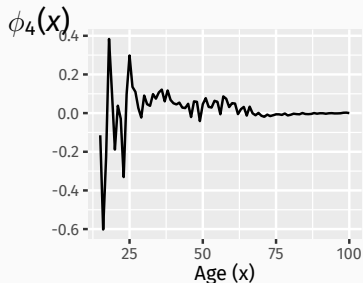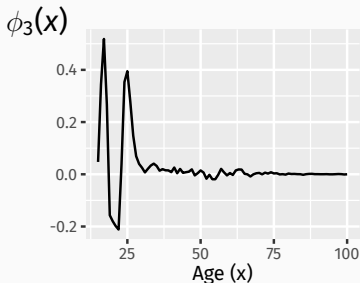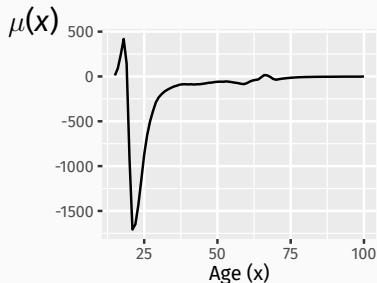

$$E_{x,t} = \mu(x) + \sum_{k=1}^{6} \beta_{k,t} \phi_k(x) + \varepsilon_t(x)$$
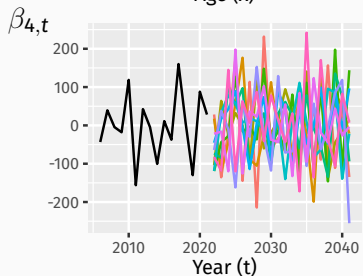
$$\beta_{k,t} \sim \text{ARIMA}$$

Forecasting models: $E_{x,t}$

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

$\mu(x)$

$\phi_3(x)$

$\phi_4(x)$

$\beta_{3,t}$

$\beta_{4,t}$

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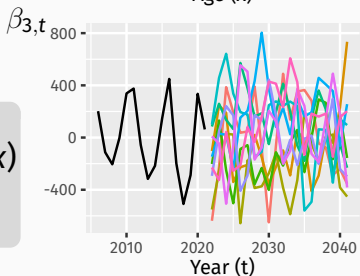
$$\beta_{k,t} \sim \text{ARIMA}$$

Forecasting models: $E_{x,t}$

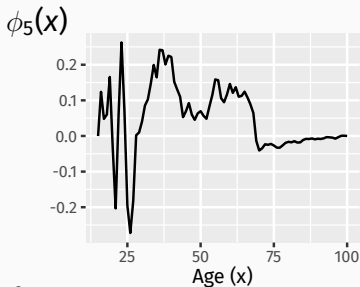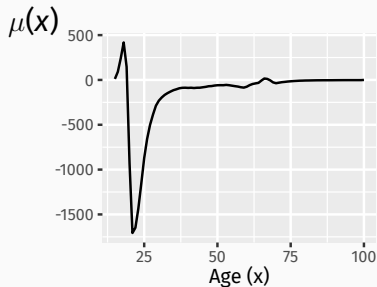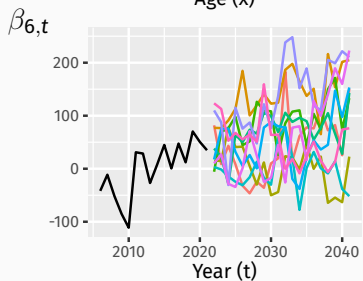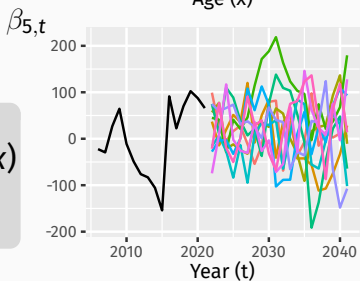$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

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

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

Forecasting models: $E_{x,t}$

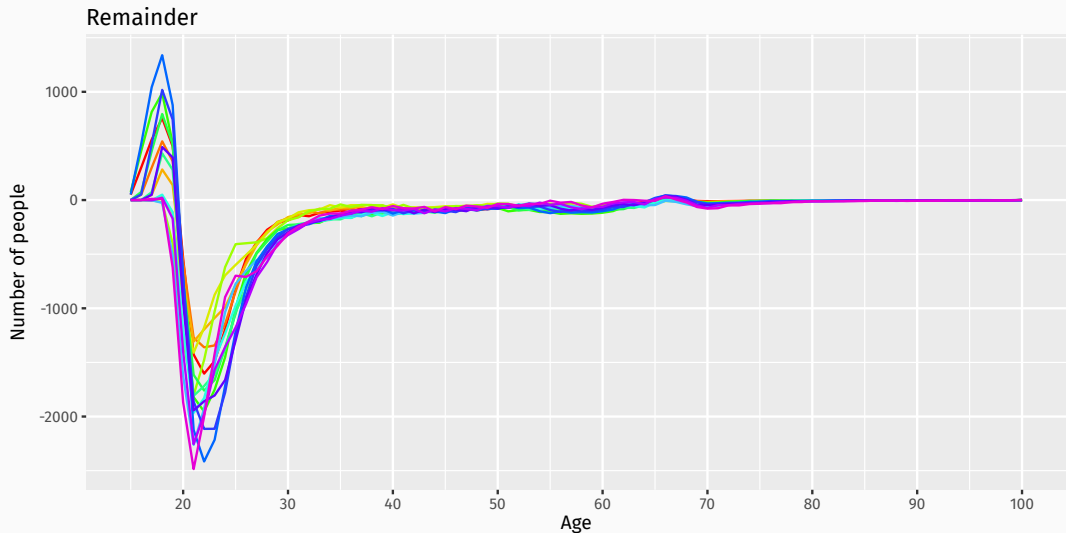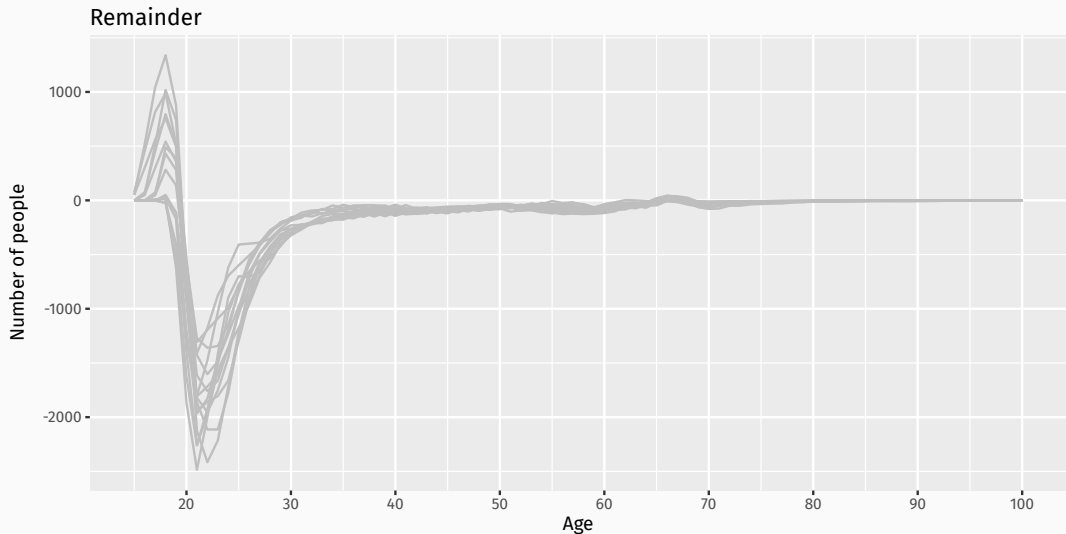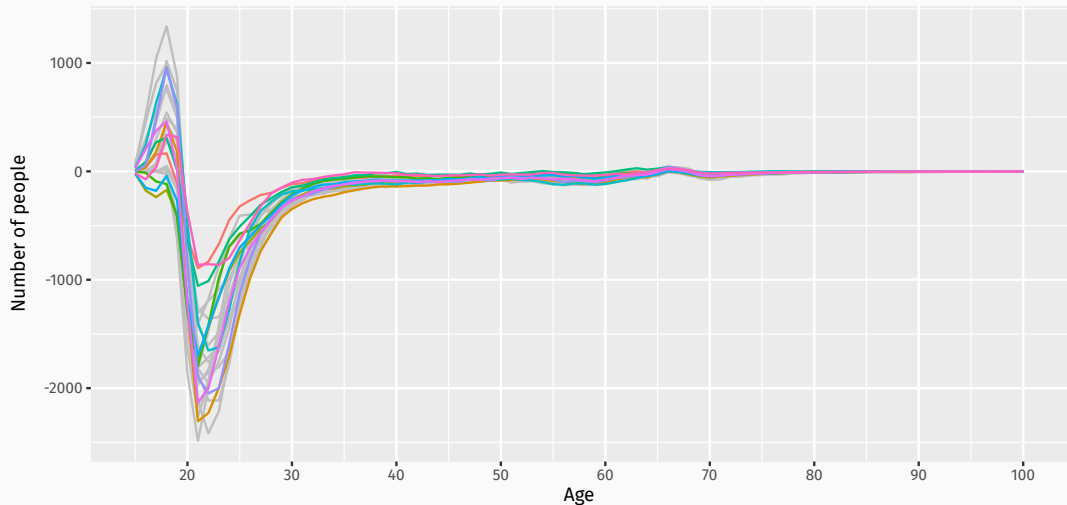$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

Remainder

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

Remainder

Forecasting models: $E_{x,t}$

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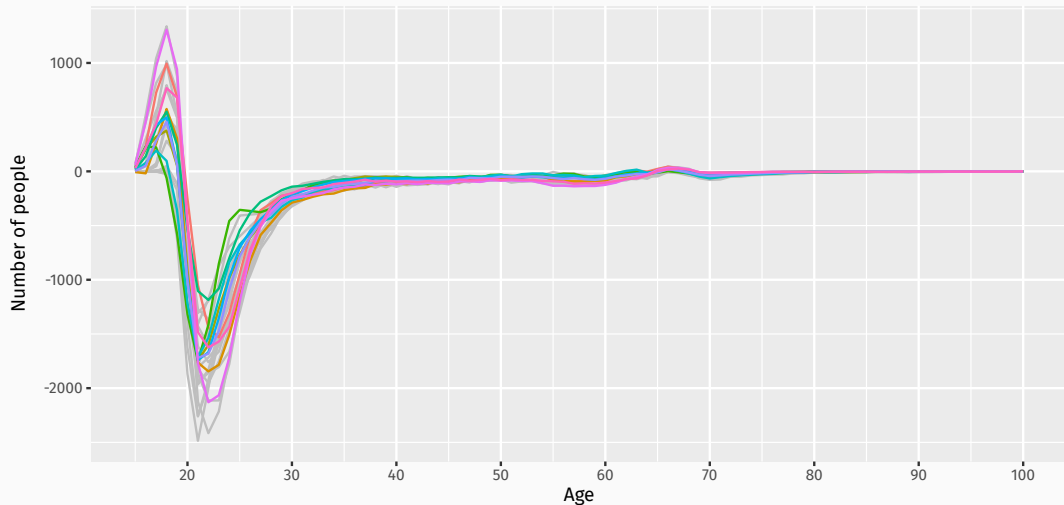

Simulated future remainder: 2030

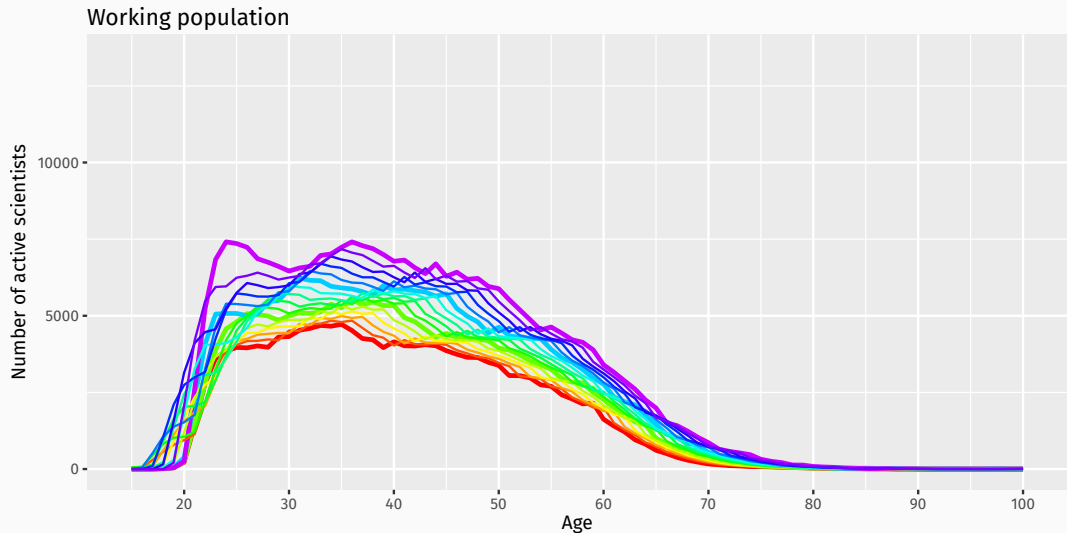# Forecasting models: $E_{x,t}$

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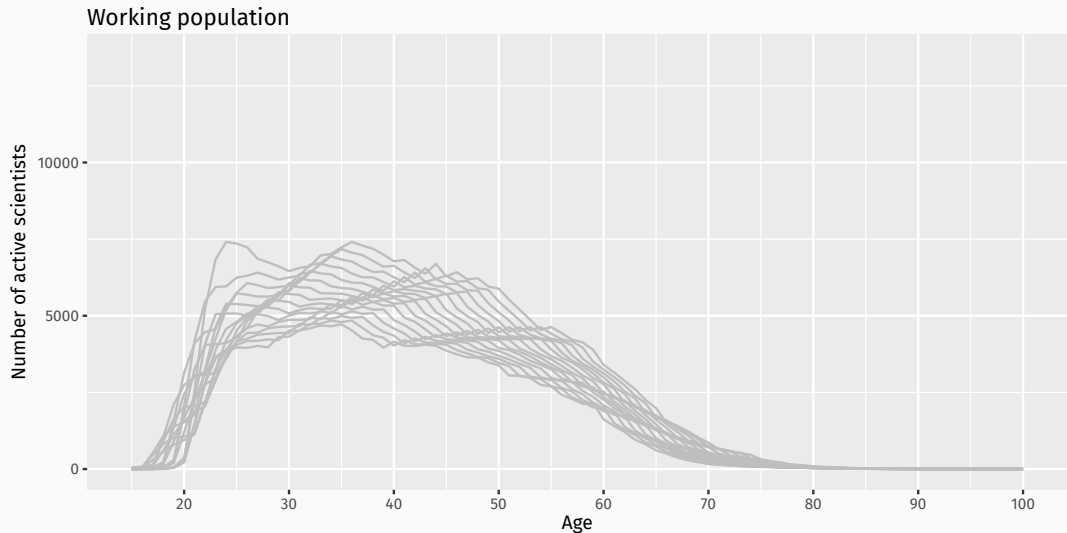

Simulated future remainder: 2040

Final forecasts: $P_{x,t}$

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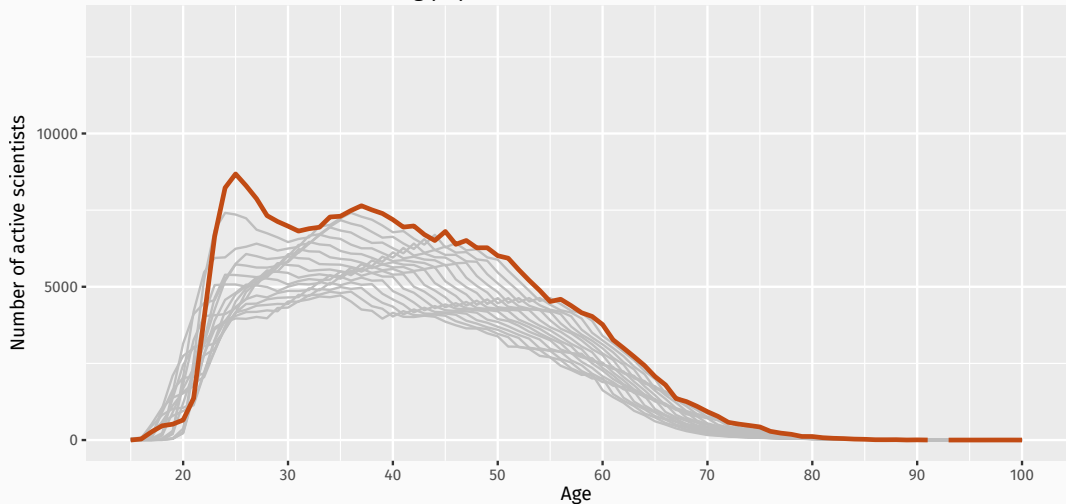

Working population

Final forecasts: $P_{x,t}$

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

Working population

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

90% forecast interval for working population: 2022

Forecasts: $P_{x,t}$

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

90% forecast interval for working population: 2023

90% forecast interval for working population: 2024

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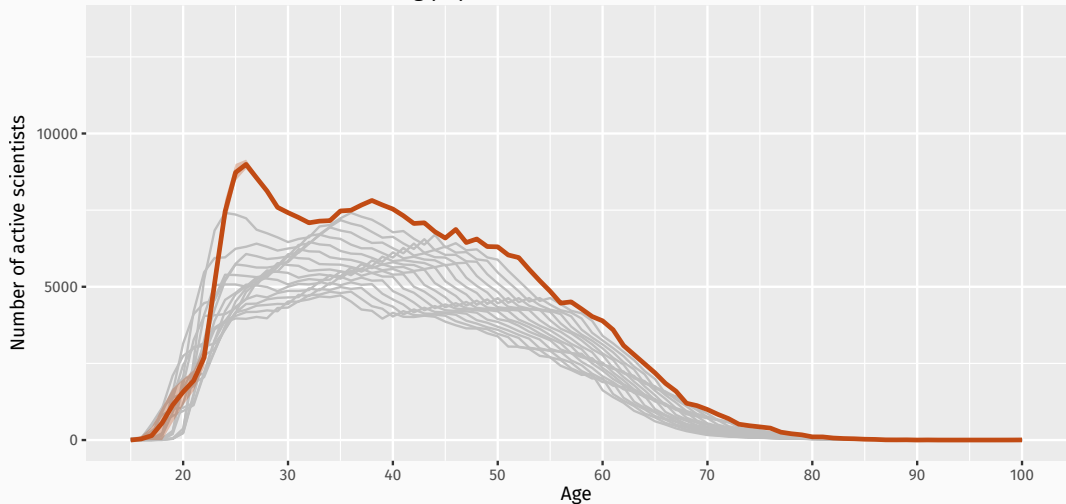

90% forecast interval for working population: 2025

Forecasts: $P_{x,t}$

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

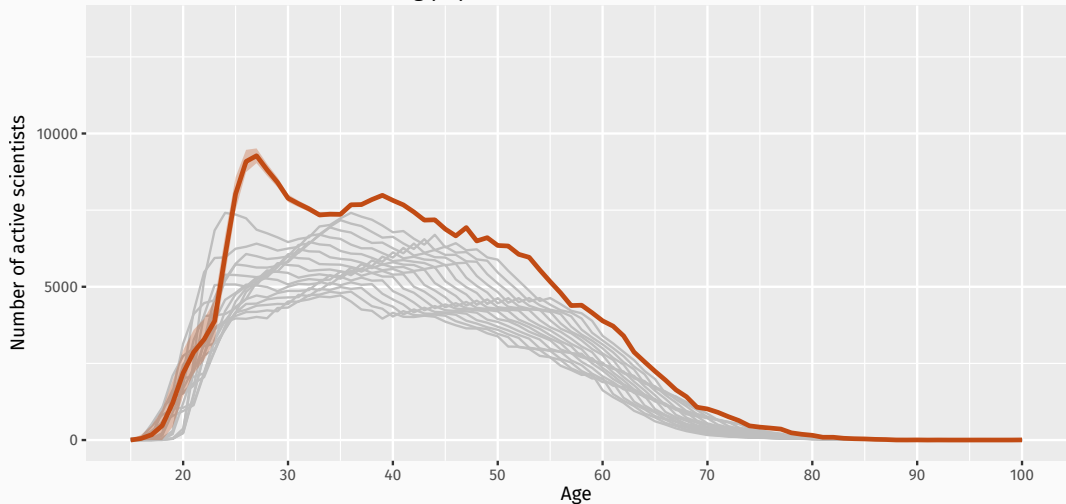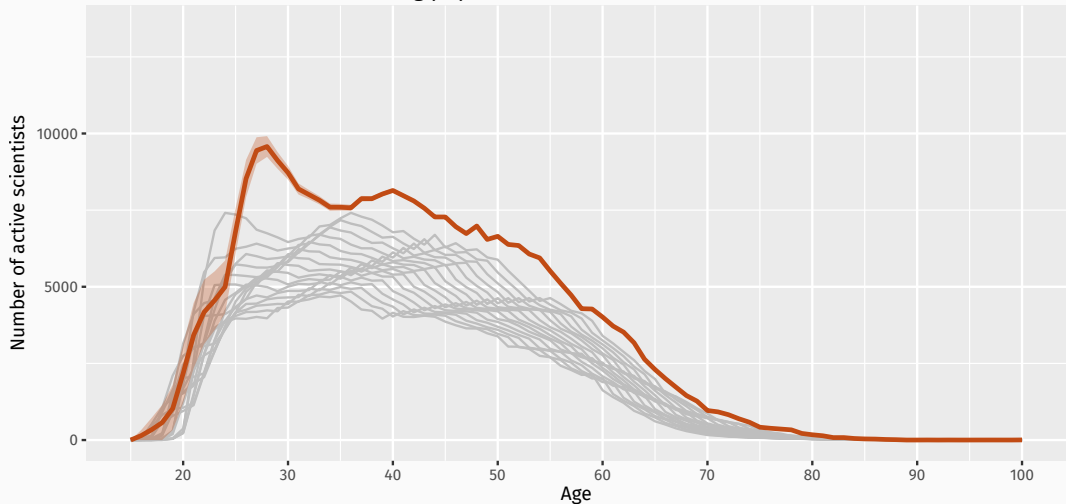90% forecast interval for working population: 2026

Forecasts: $P_{x,t}$

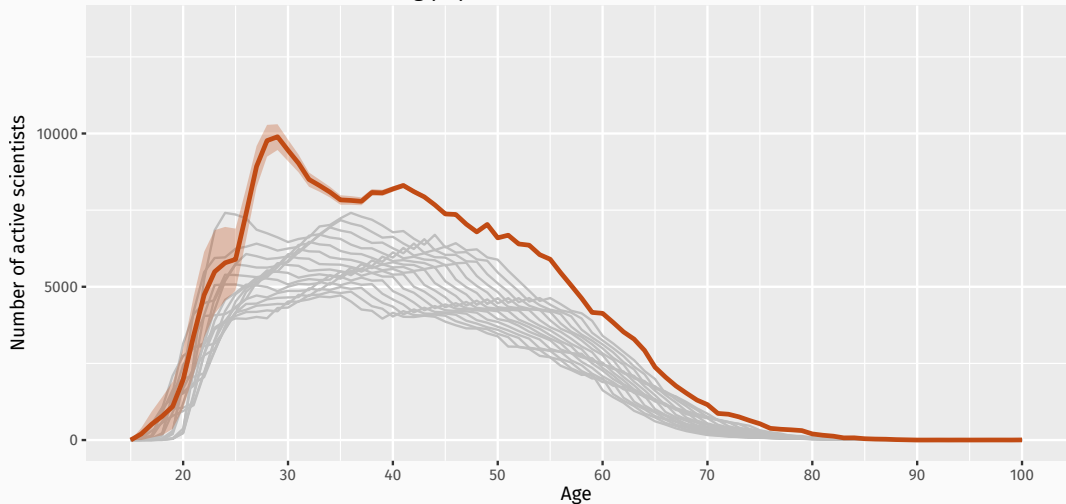$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

90% forecast interval for working population: 2027

Forecasts: $P_{x,t}$

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

90% forecast interval for working population: 2028

Forecasts: $P_{x,t}$

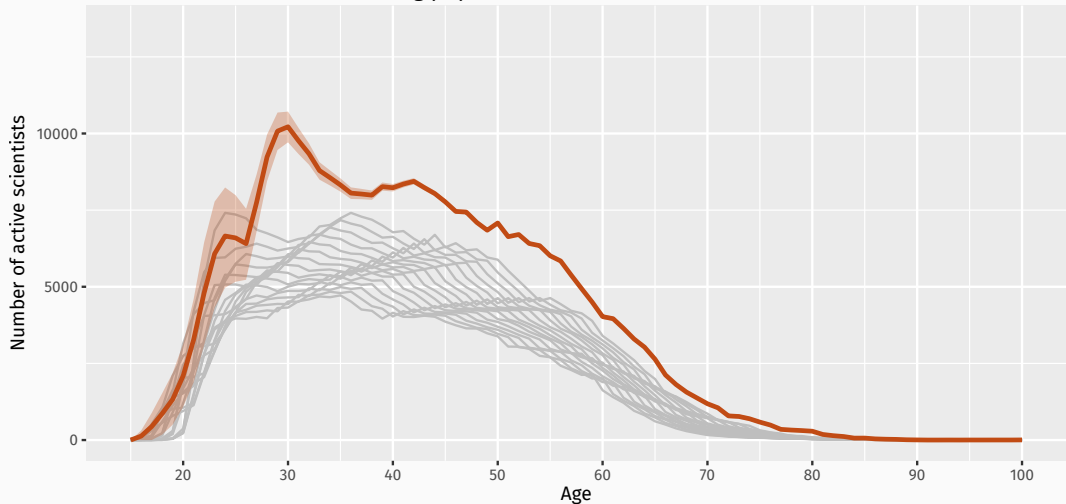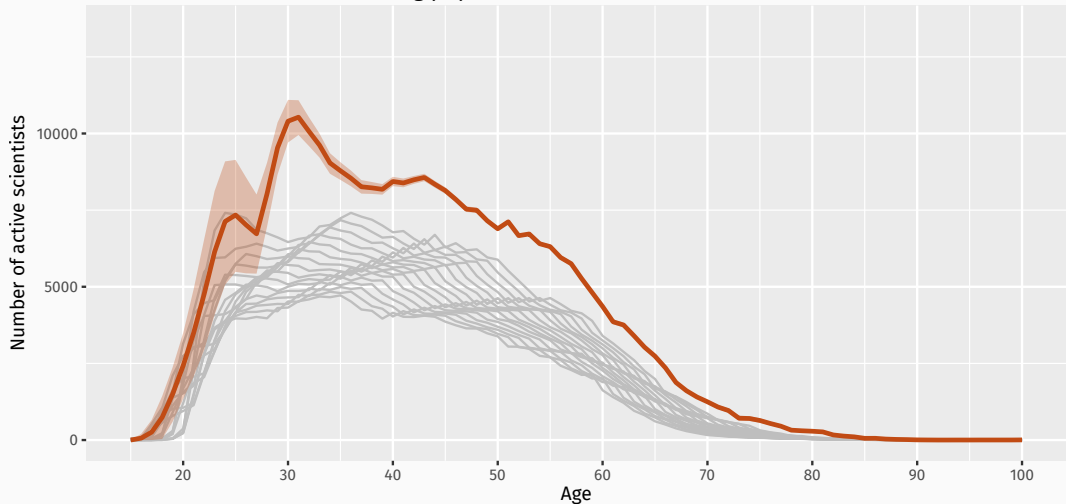$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$

90% forecast interval for working population: 2029

Forecasts: $P_{x,t}$

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

90% forecast interval for working population: 2030

Forecasts: $P_{x,t}$

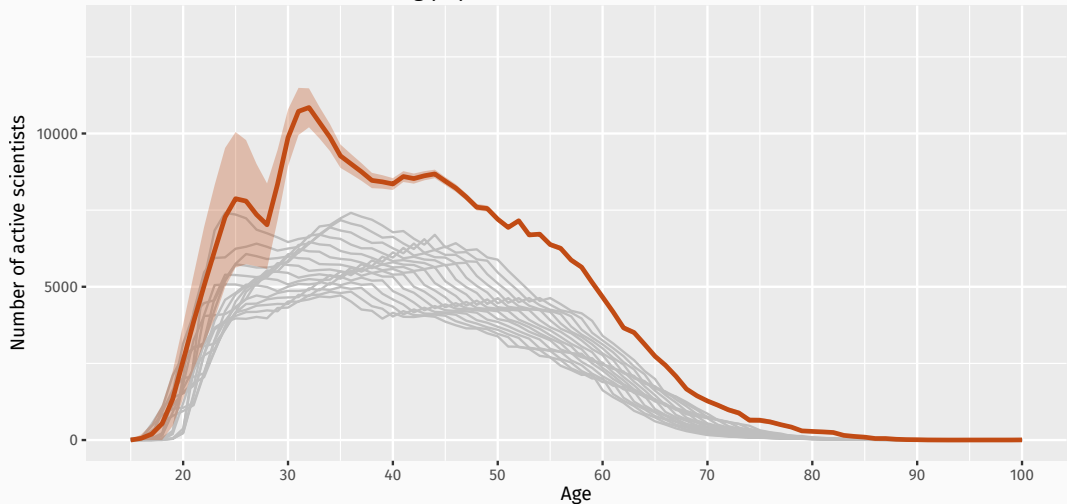$$P_{x+1,t+1} = P_{x,t}(1 - q_{x,t} - r_x) + g_x G_t + E_{x,t}$$
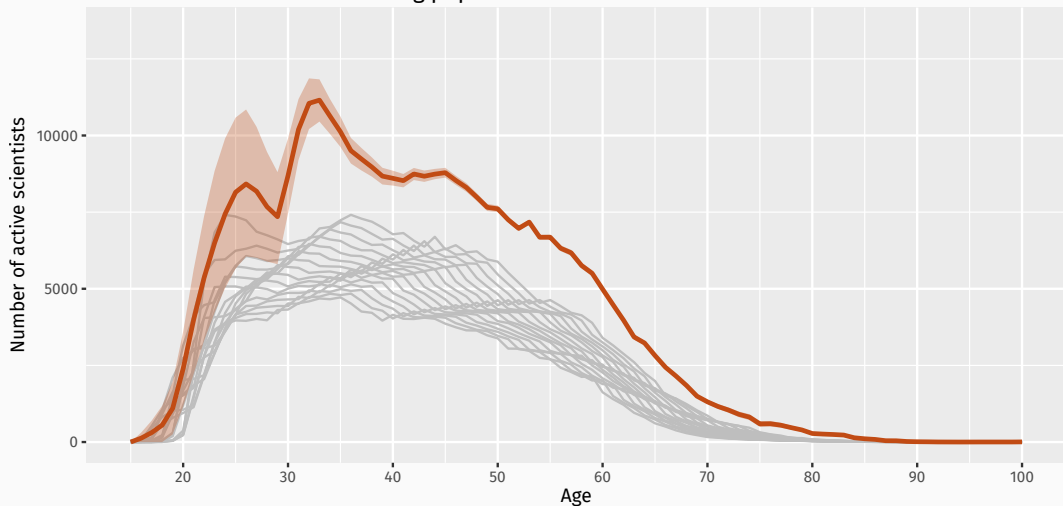

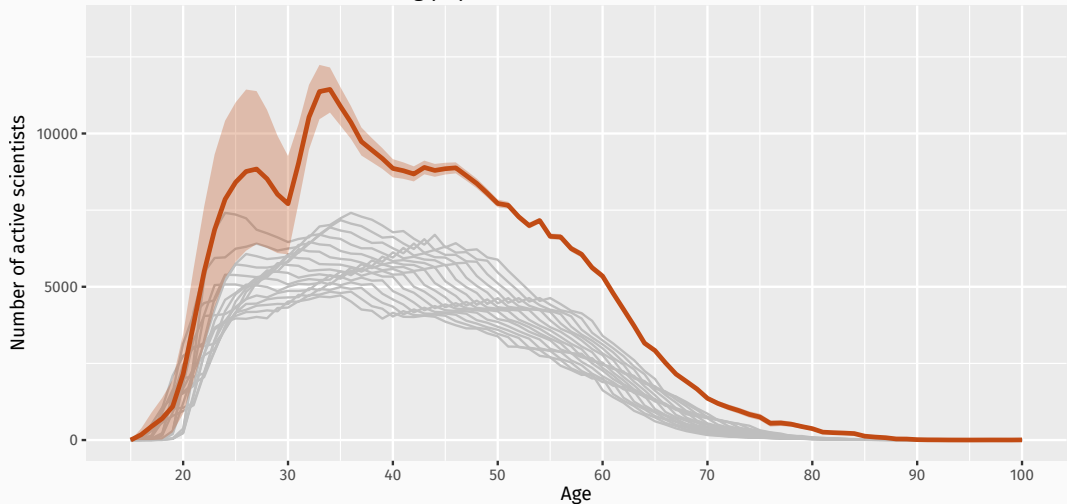90% forecast interval for working population: 2031

Forecasts: $P_{x,t}$

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

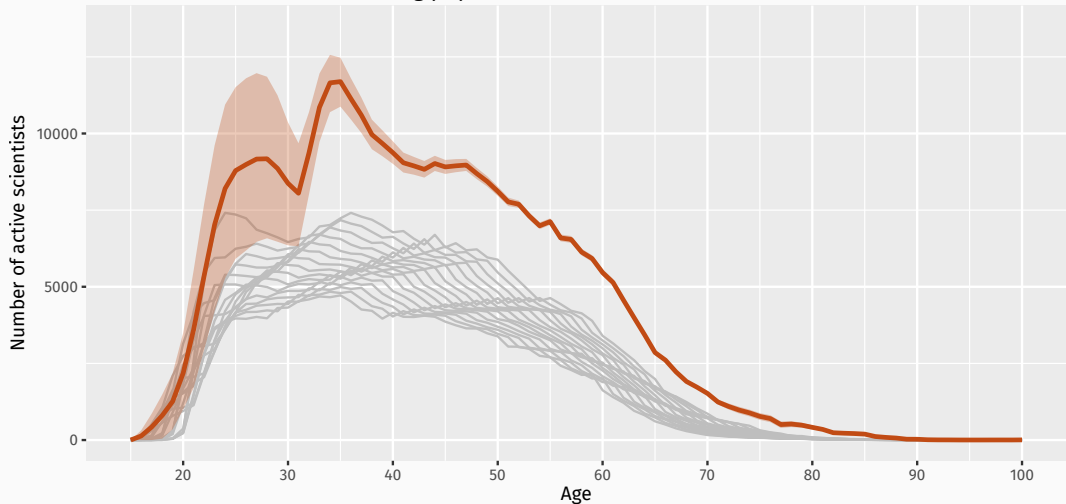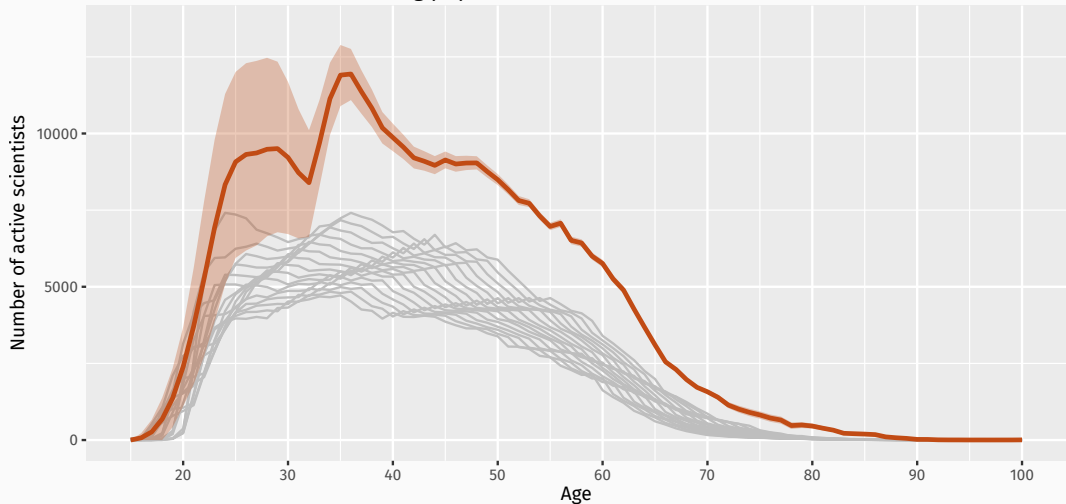90% forecast interval for working population: 2032

Forecasts: $P_{x,t}$

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

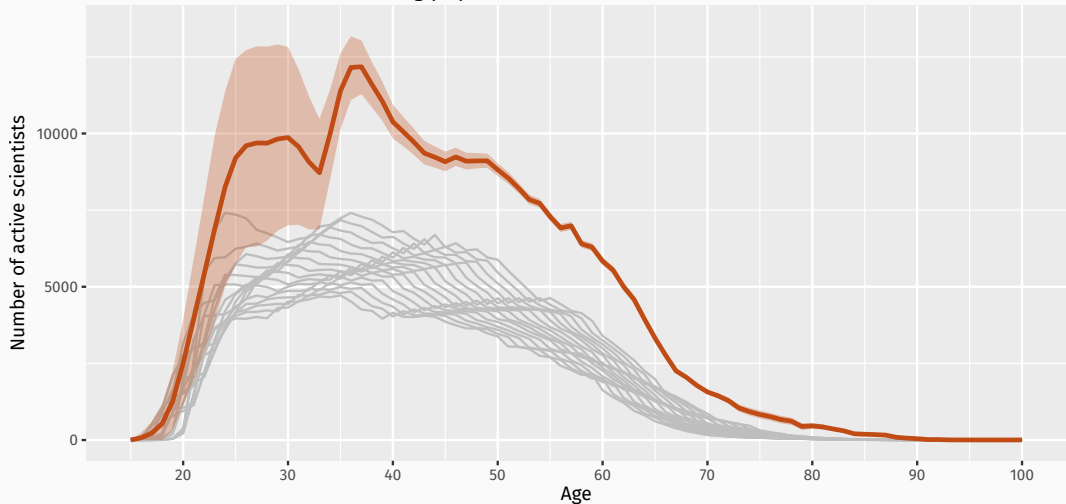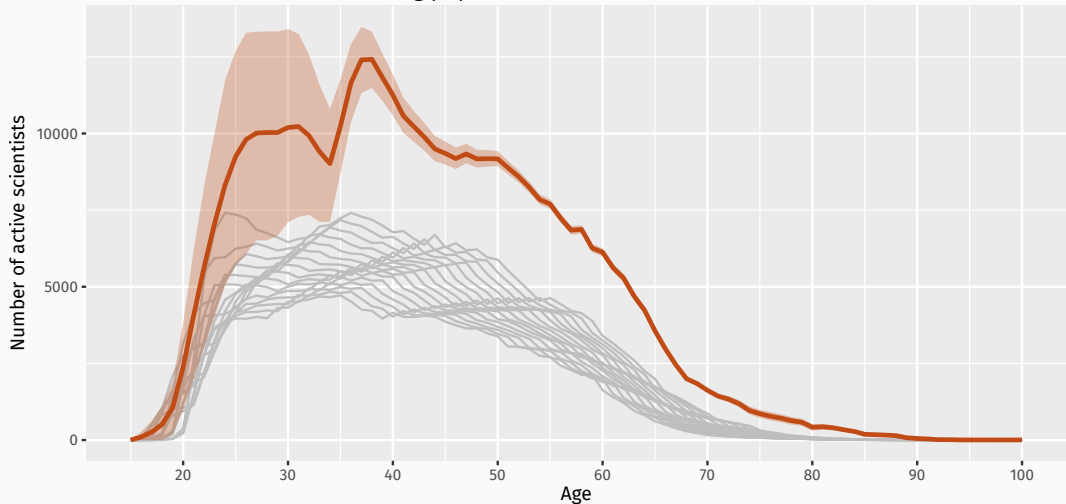90% forecast interval for working population: 2033

Forecasts: $P_{x,t}$

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

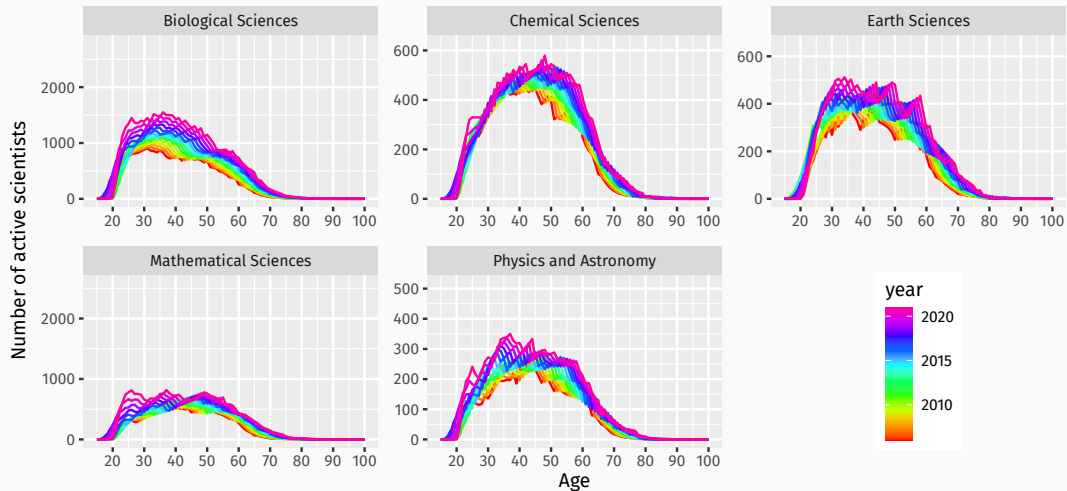
90% forecast interval for working population: 2034

90% forecast interval for working population: 2035

Population by discipline

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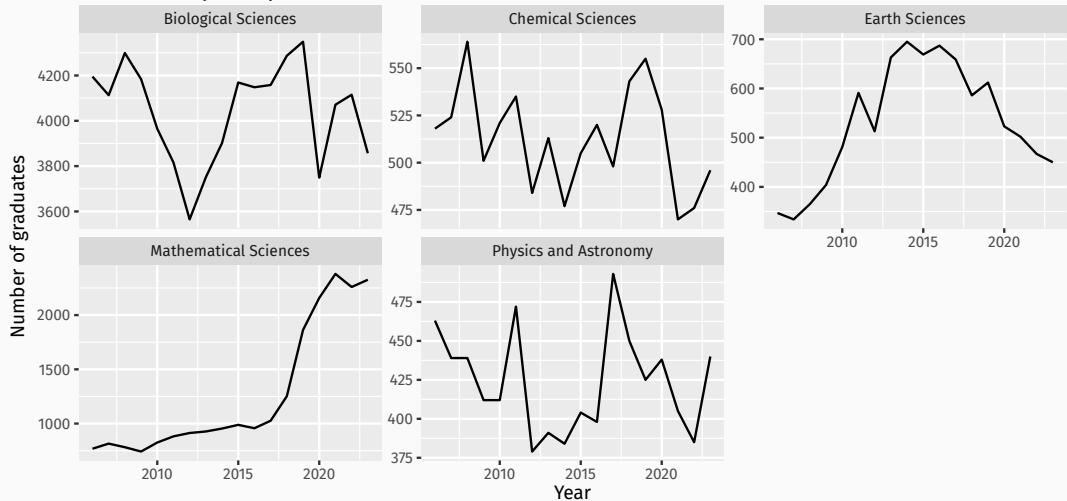

Working population by discipline

Graduates by discipline

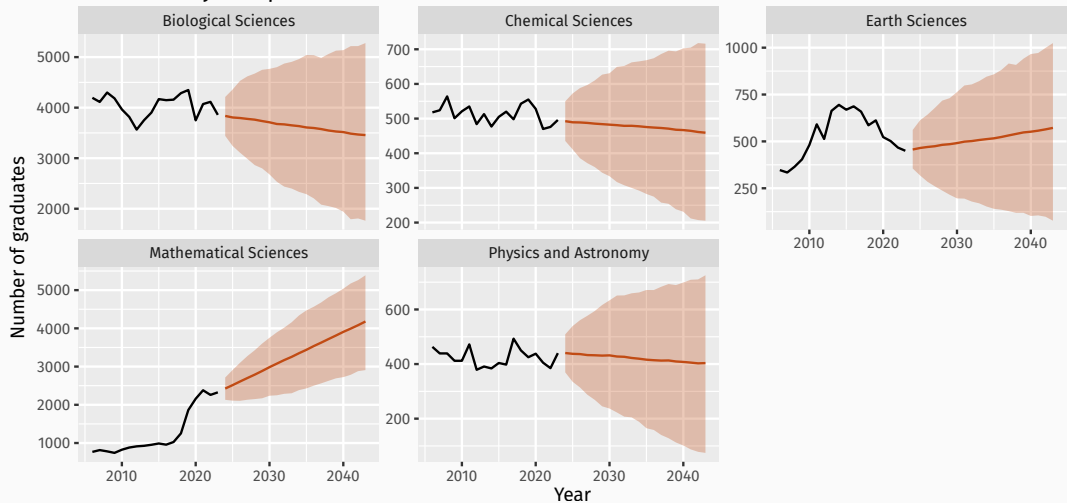$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$

Graduates by discipline: (2006 – 2023)

Graduate forecasts

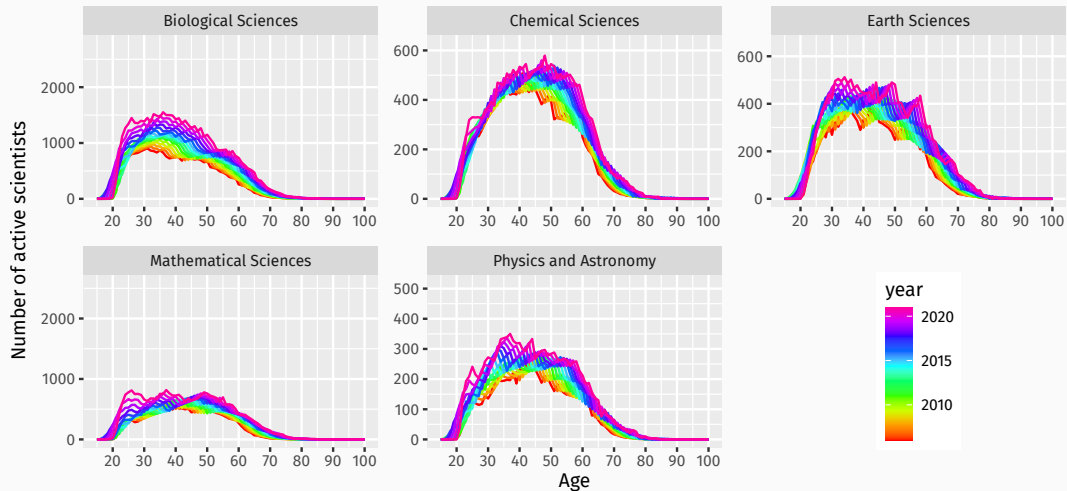$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$

Graduates by discipline

Population by discipline

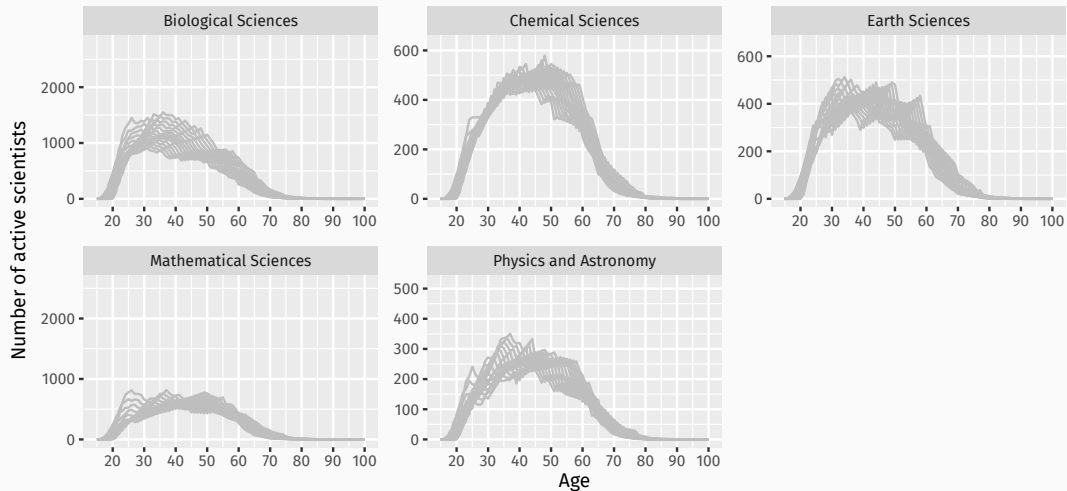$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$

Working population by discipline

Population by discipline

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

Working population by discipline

Forecasts by discipline

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

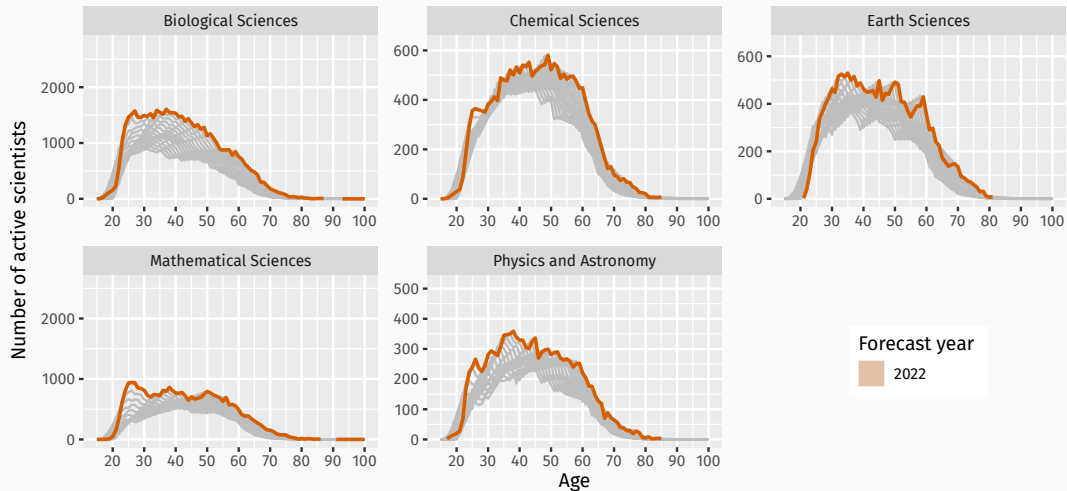90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
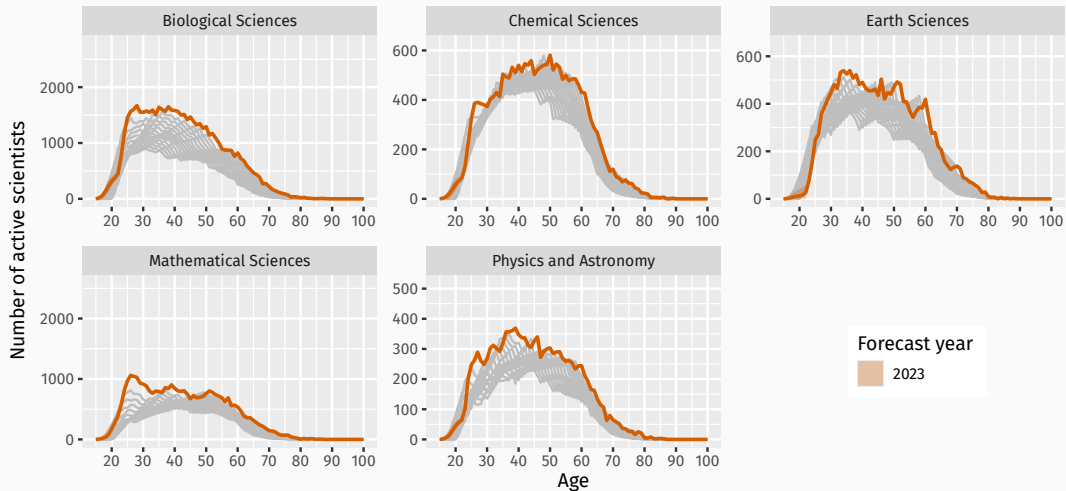
90% forecast interval for working population by discipline

Forecasts by discipline

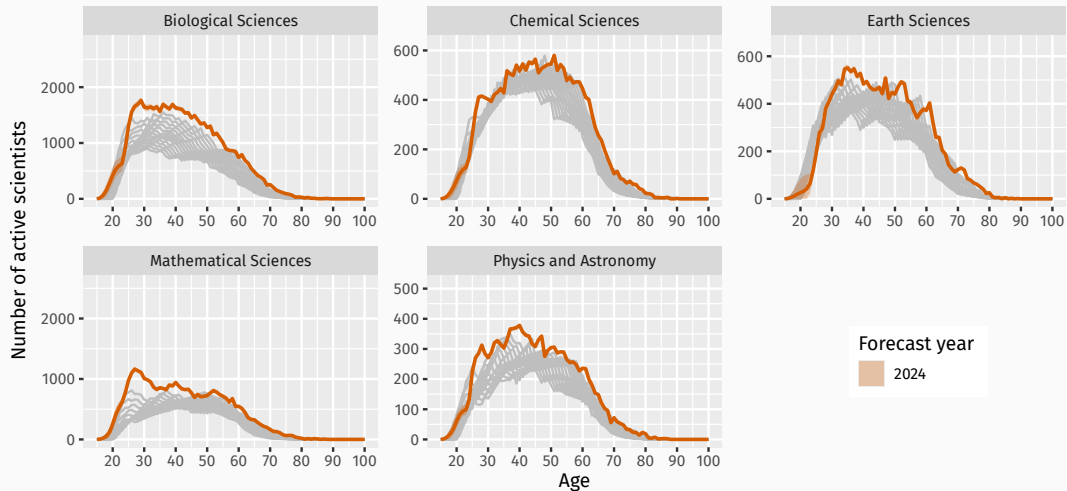$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$

90% forecast interval for working population by discipline

Forecasts by discipline

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

90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
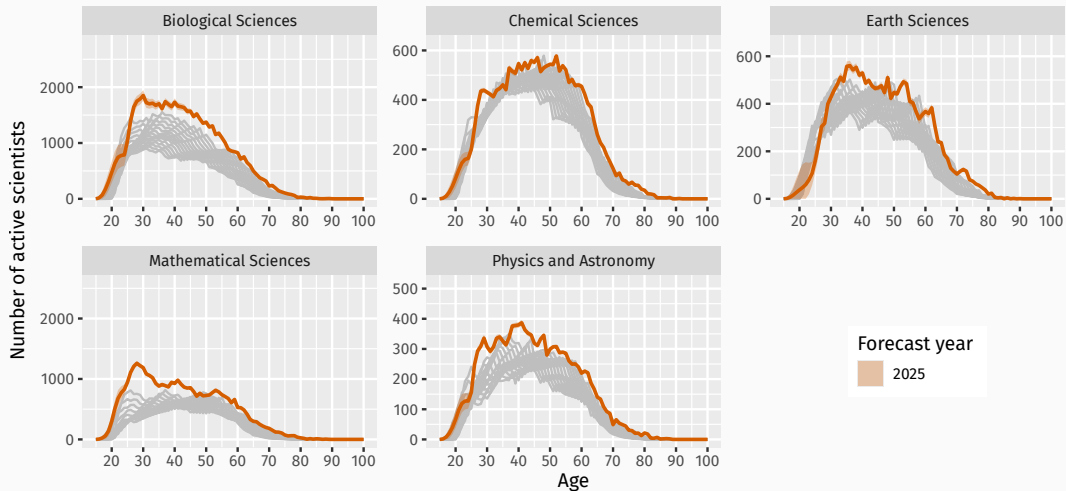
90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
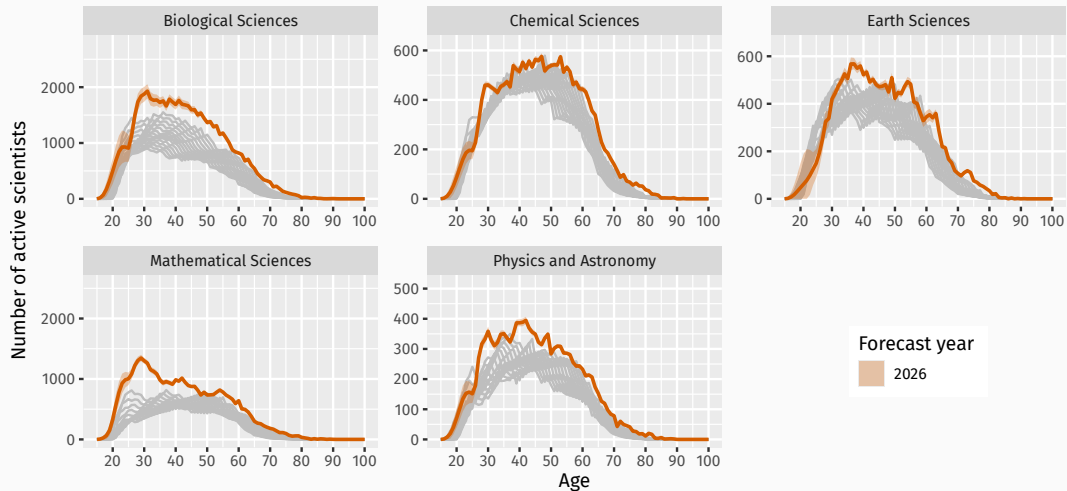
90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
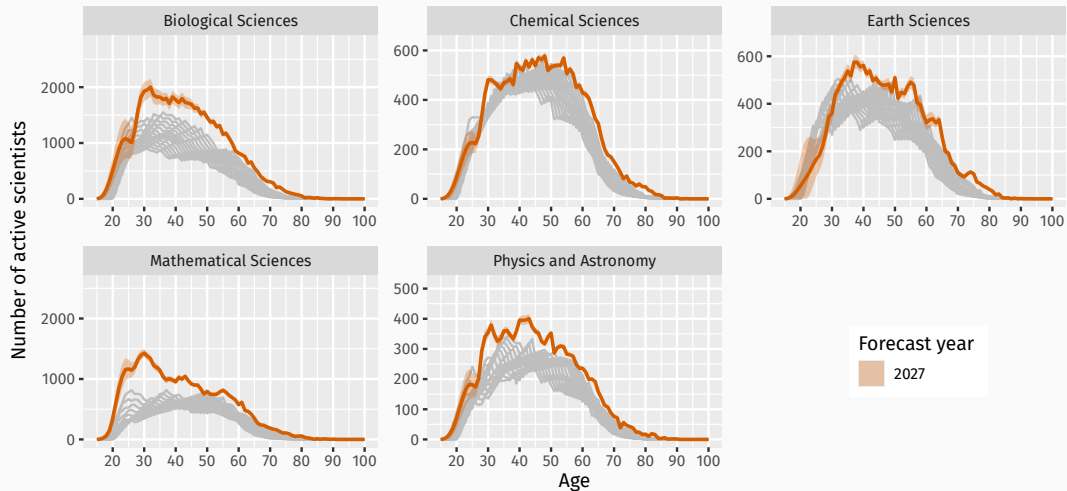
90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
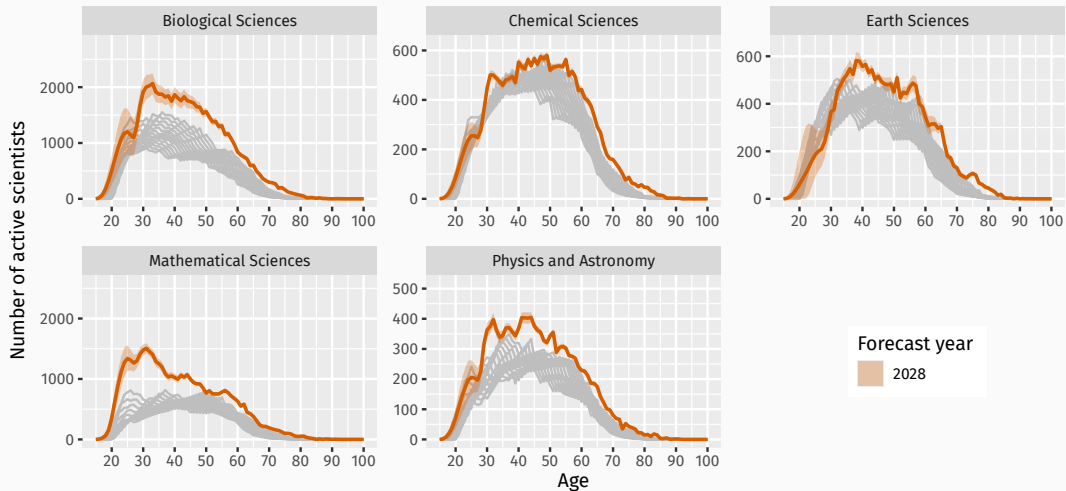
90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
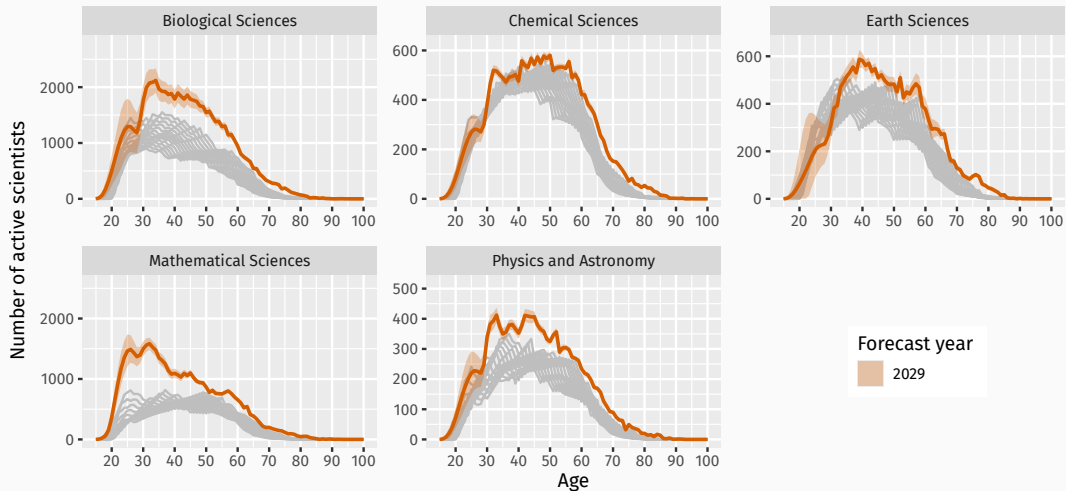
90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
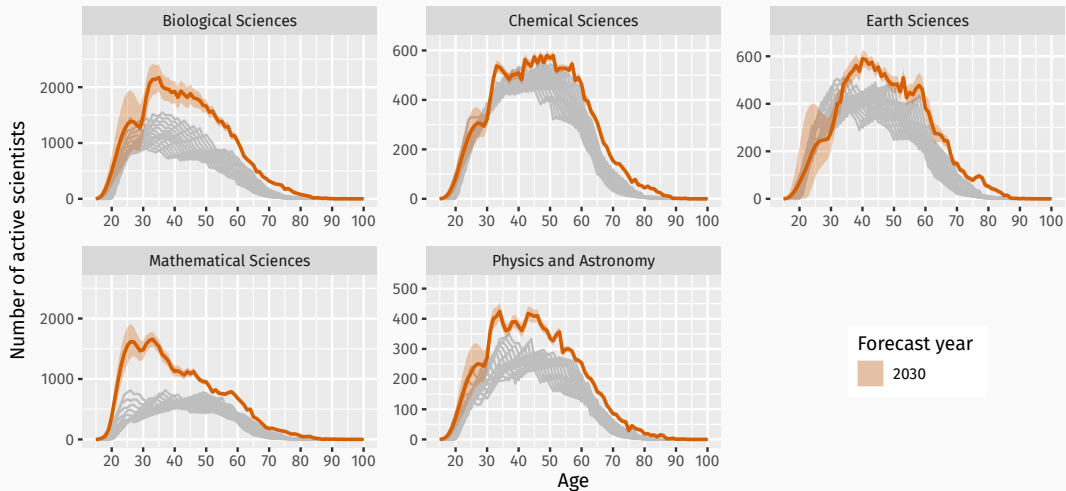
90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
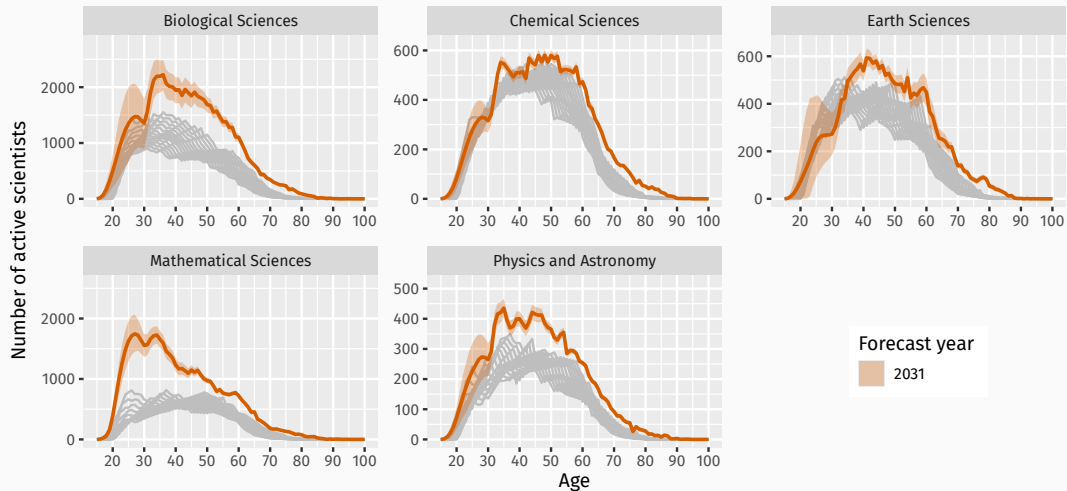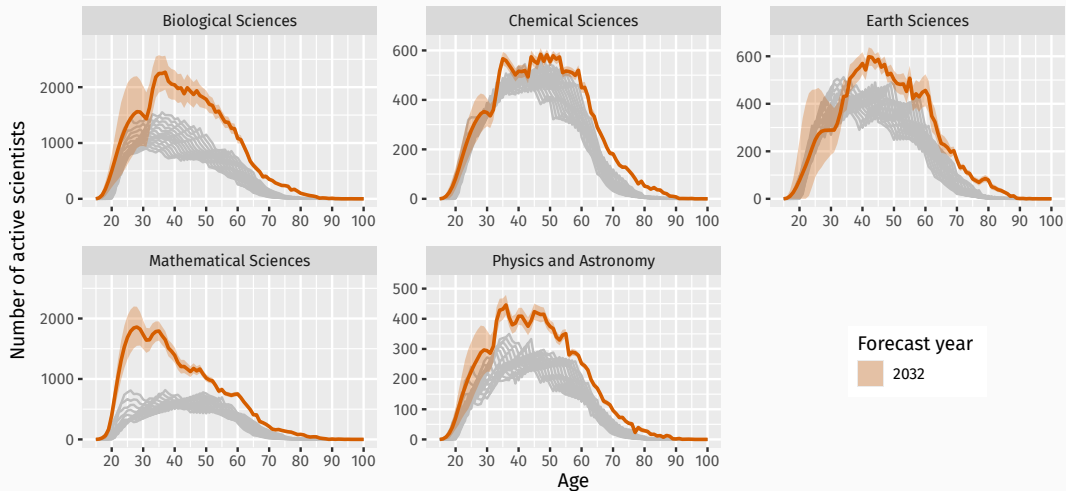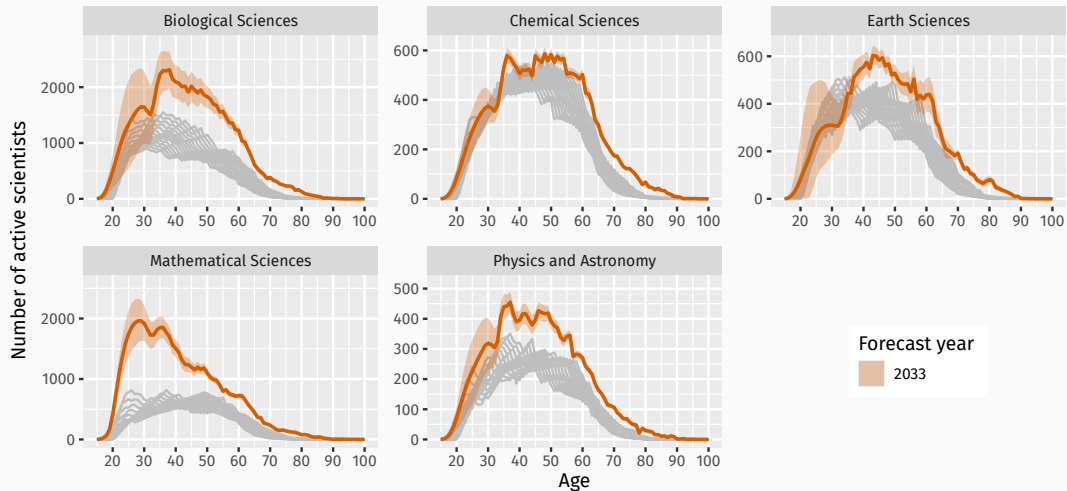
90% forecast interval for working population by discipline

Forecasts by discipline

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

90% forecast interval for working population by discipline

Forecasts by discipline

$$P_{x+1,t+1,i} = P_{x,t,i}(1 - q_{x,t} - r_x) + g_x G_{t,i} + E_{x,t,i}$$
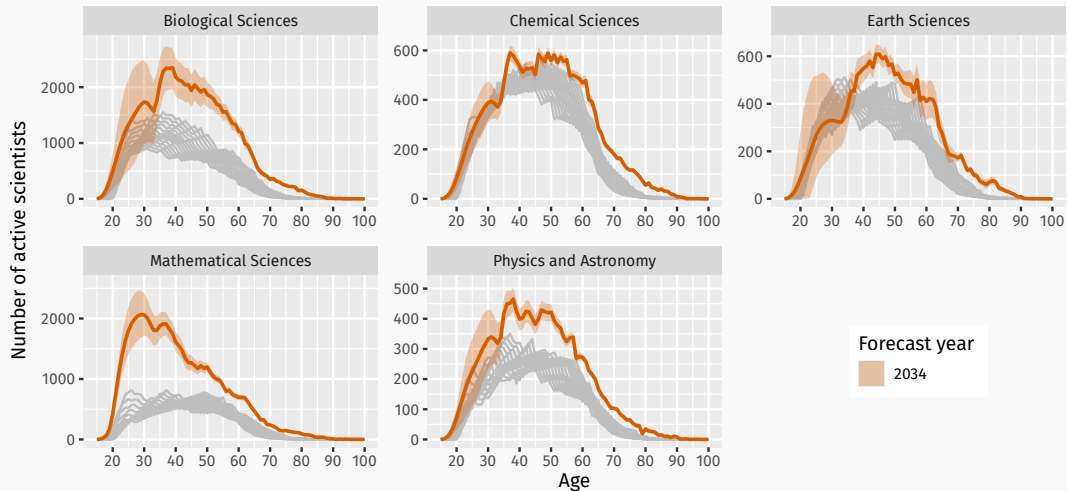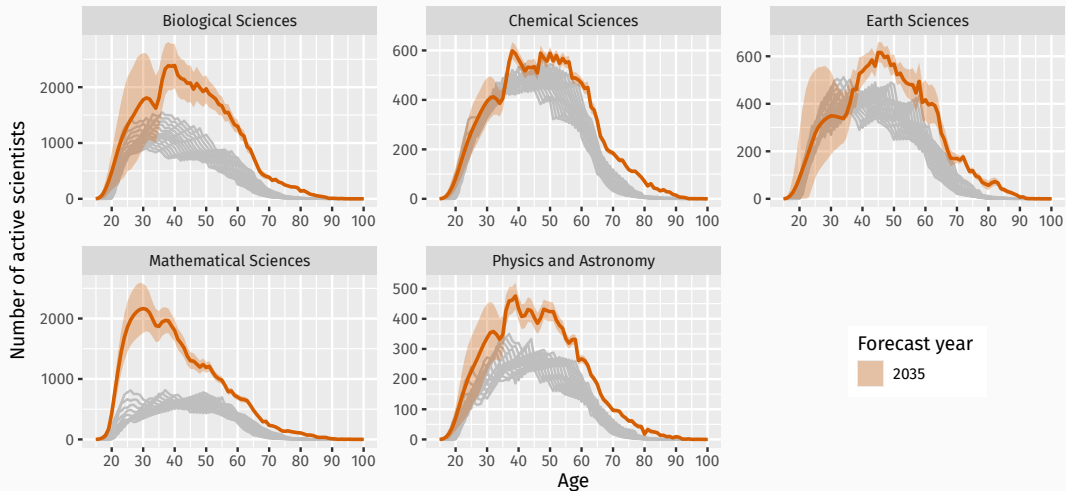
90% forecast interval for working population by discipline

Forecasts by discipline

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

90% forecast interval for working population by discipline

More information

robjhyndman.com/isf2025

robjhyndman.com/postdoc