MONASH University

MONASH BUSINESS SCHOOL

Forecasting the age structure of the scientific workforce in Australia

Rob J Hyndman & Kelly Nguyen

2 December 2025

Australian Academy of Science

Labour force model

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + G_{i,x,t} + E_{i,x,t}$$

i = Discipline x = Age t = Year

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

$D_{i,x,t}$ = number of deaths $\sim$ Binomial$(P_{i,x,t}, q_{x,t})$

$R_{i,x,t}$ = number of retirements $\sim$ Binomial$(P_{i,x,t} - D_{i,x,t}, r_x)$

$G_{i,x,t} = g_x G_{i,t}$ = number of graduates who work in discipline i

$E_{i,x,t}$ = other changes (career changes, migration, etc.)

Labour force model

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + G_{i,x,t} + E_{i,x,t}$$

i = Discipline x = Age t = Year

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

$D_{i,x,t}$ = number of deaths $\sim$ Binomial$(P_{i,x,t}, q_{x,t})$

$R_{i,x,t}$ = number of retirements $\sim$ Binomial$(P_{i,x,t} - D_{i,x,t}, r_x)$

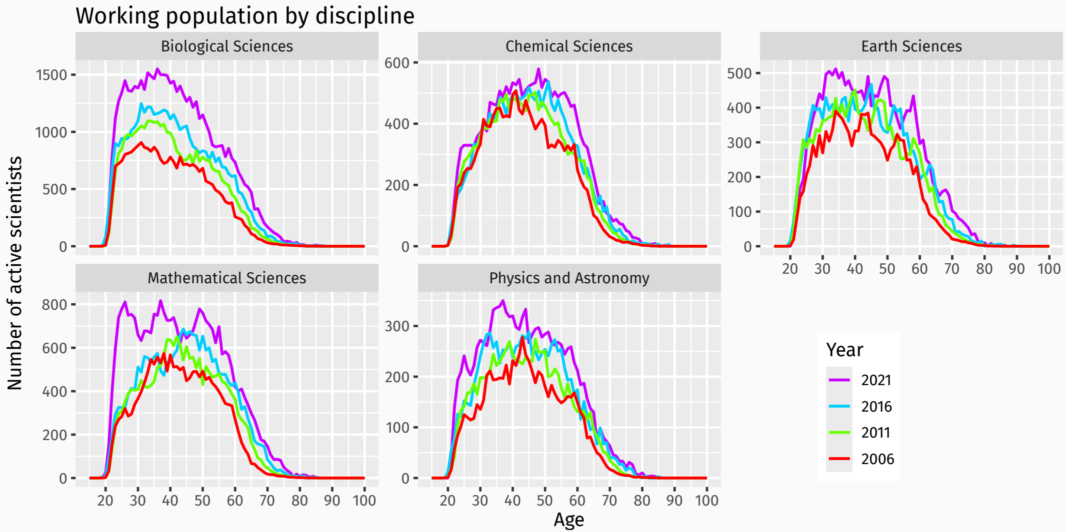
$G_{i,x,t} = g_x G_{i,t}$ = number of graduates who work in discipline i

$E_{i,x,t}$ = other changes (career changes, migration, etc.)

Simulate future sample paths of $P_{i,x,t}$ by simulating future $q_{x,t}$, $D_{i,x,t}$, $R_{i,x,t}$, $G_{i,t}$ and $E_{i,x,t}$

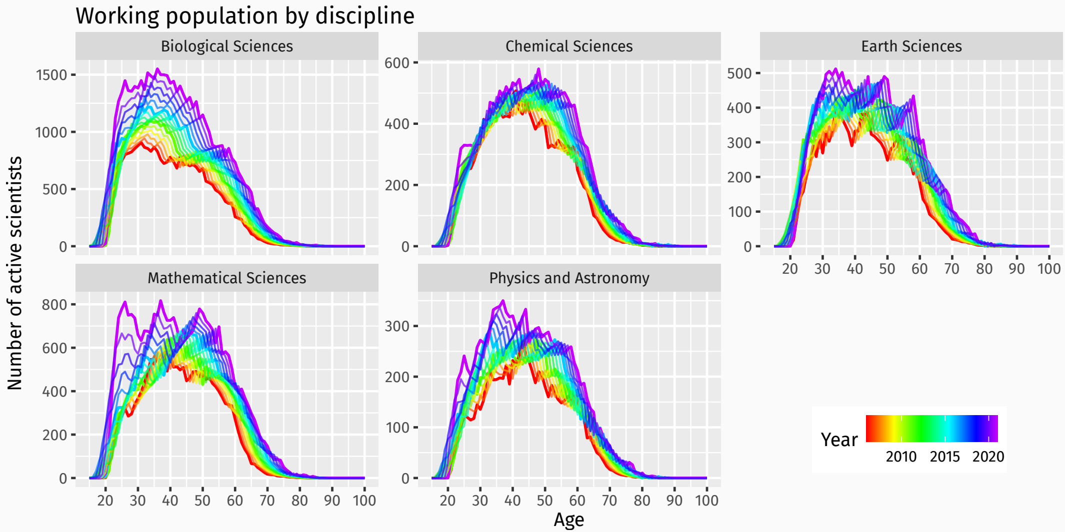
Working population: $P_{i,x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Working population: $P_{i,x,t}$

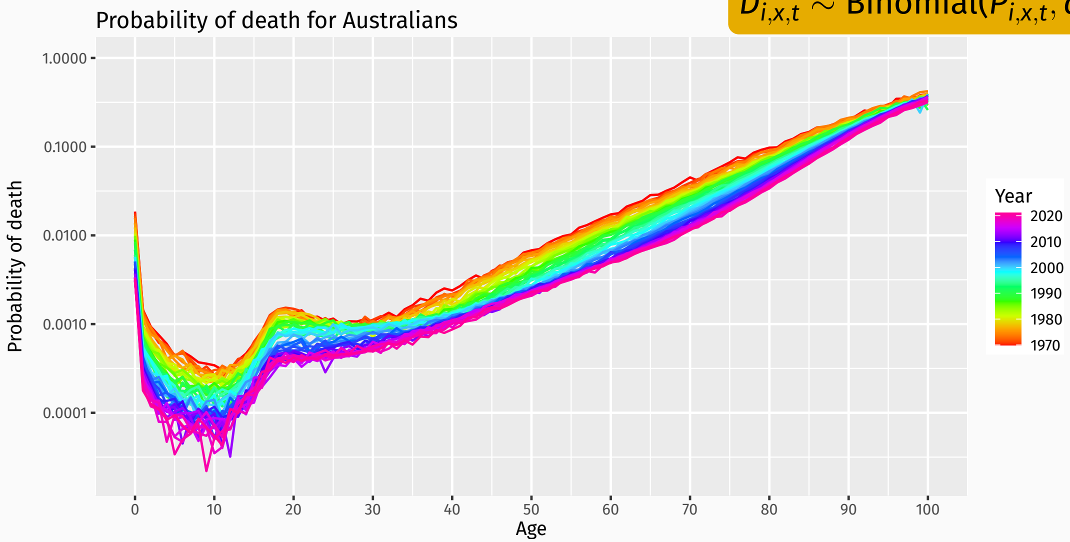
$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

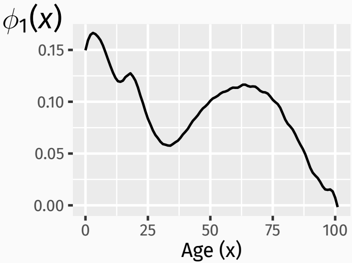
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$

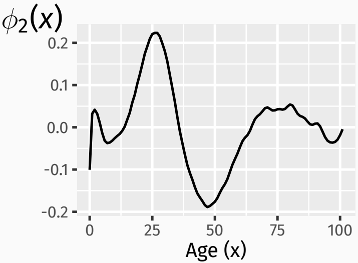


Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

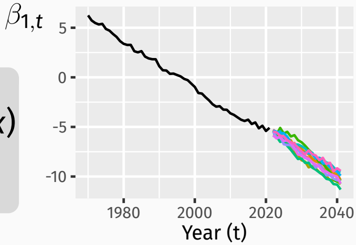
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$

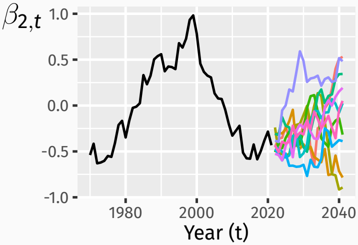




$$q_{x,t} = \mu(x) + \sum_{k=1}^{6} \beta_{k,t}\phi_k(x) + \varepsilon_t(x)$$

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

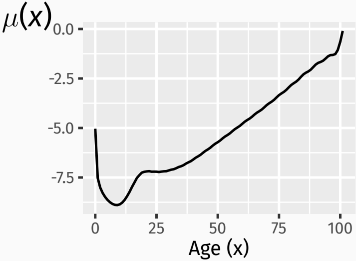


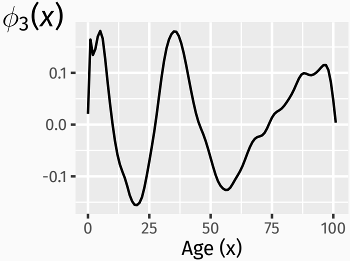


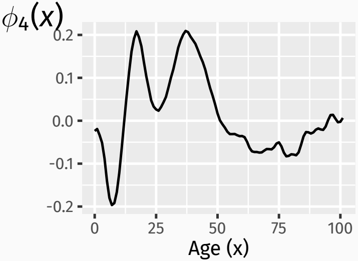
Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$

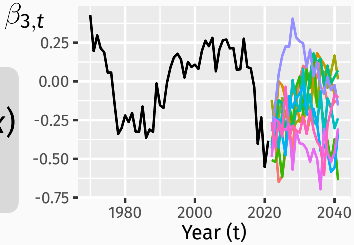


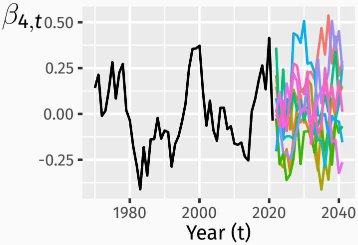




$$q_{x,t} = \mu(x) + \sum_{k=1}^{6} \beta_{k,t}\phi_k(x) + \varepsilon_t(x)$$

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

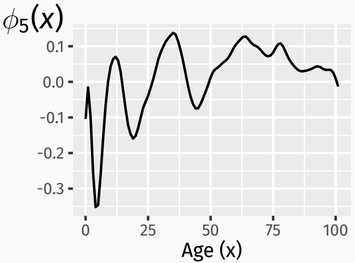


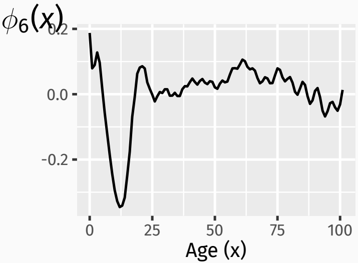


Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

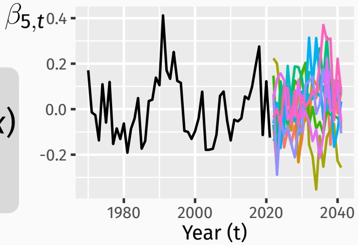
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$

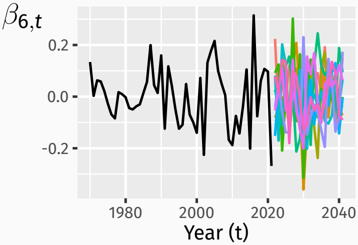




$$q_{x,t} = \mu(x) + \sum_{k=1}^{6} \beta_{k,t}\phi_k(x) + \varepsilon_t(x)$$

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

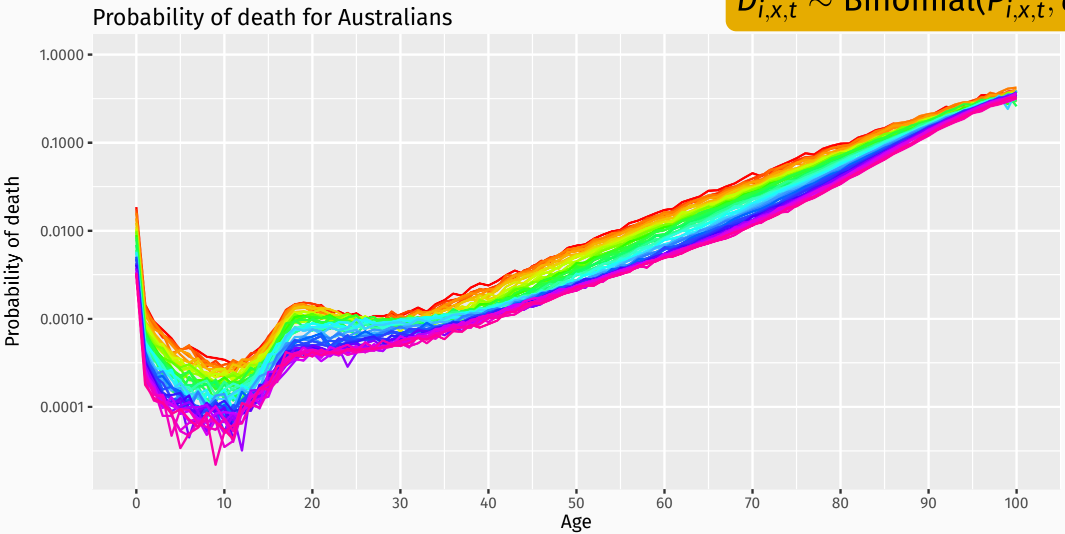




Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

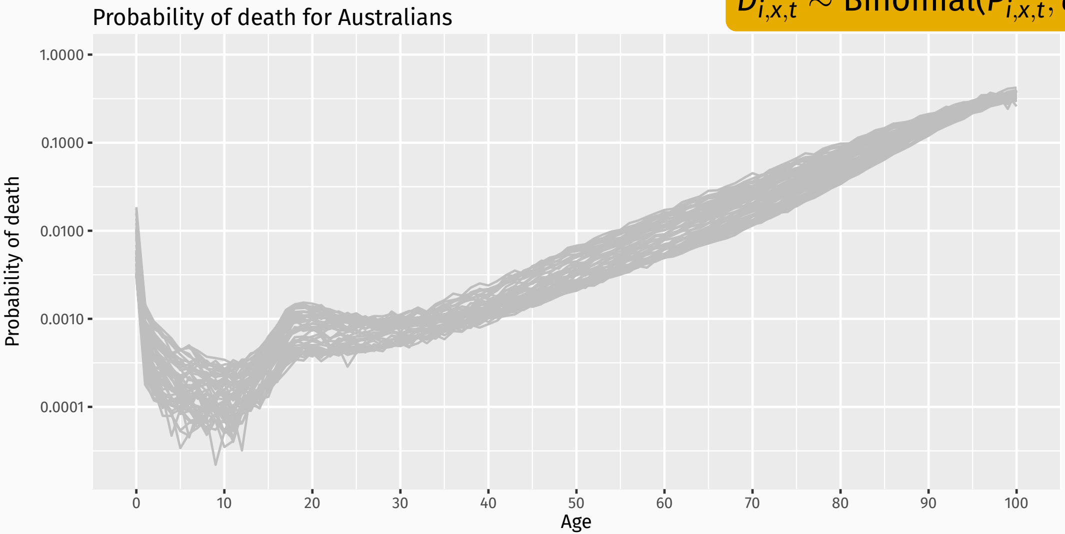
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$



Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

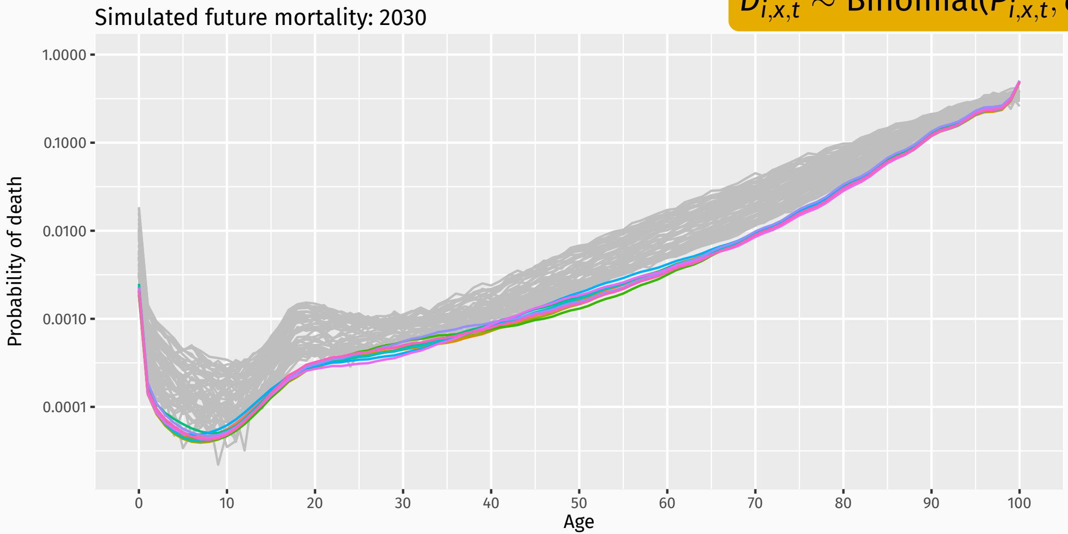
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$



Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

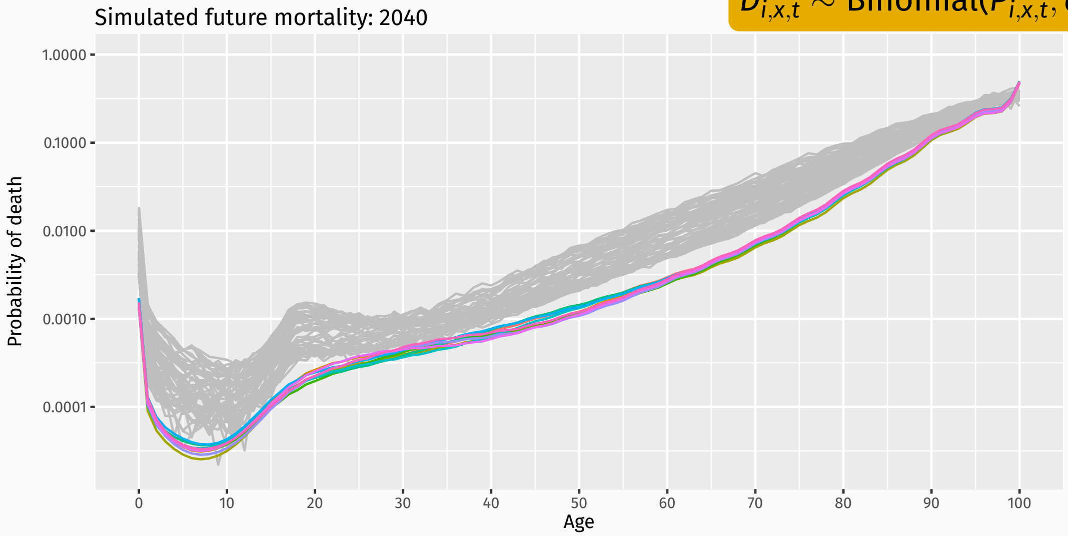
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$



Death probability: $q_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

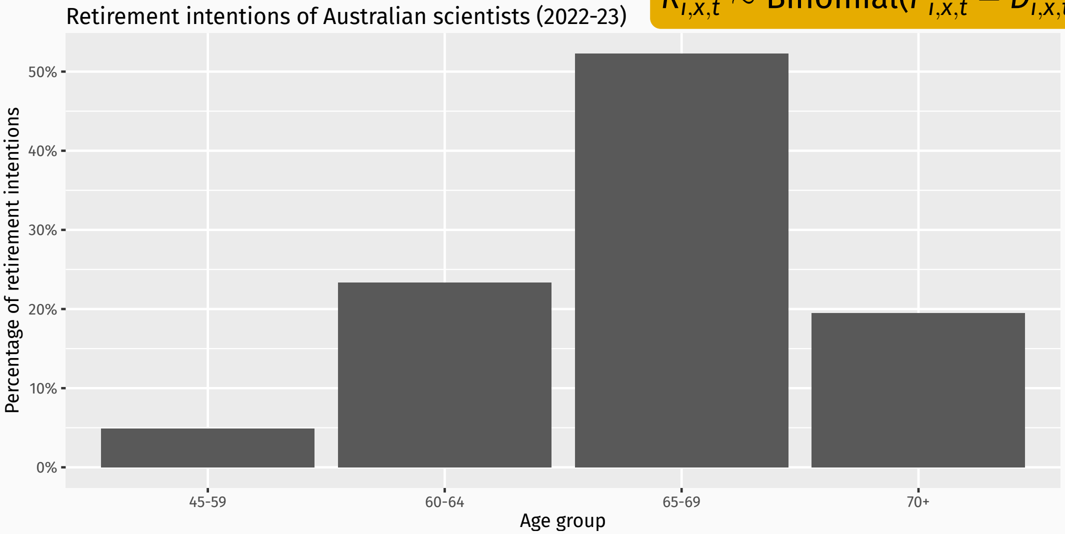
$$D_{i,x,t} \sim \text{Binomial}(P_{i,x,t}, q_{x,t})$$



Retirement rates: r_x

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

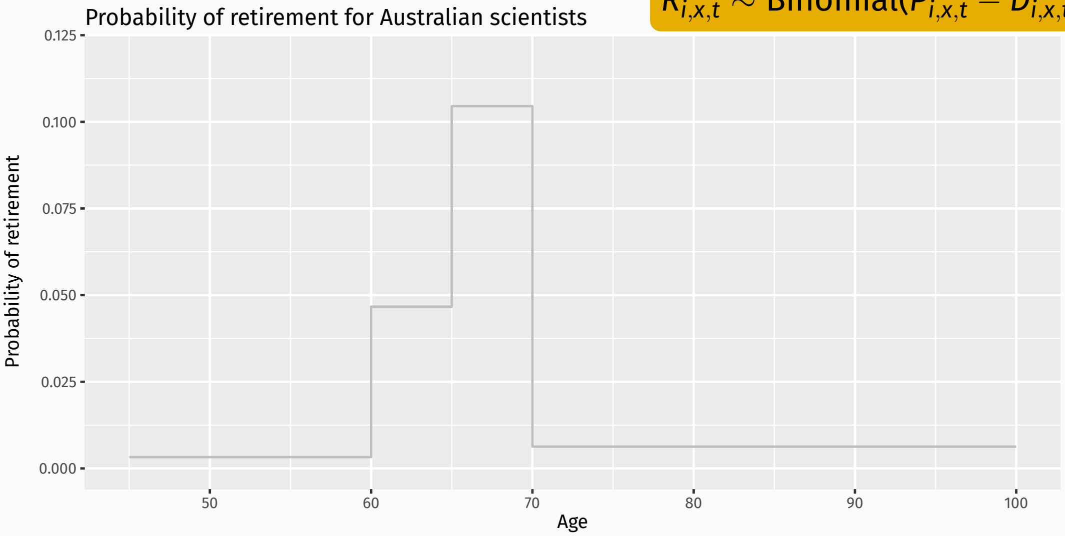
$$R_{i,x,t} \sim \text{Binomial}(P_{i,x,t} - D_{i,x,t}, r_x)$$



Retirement rates: r_x

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

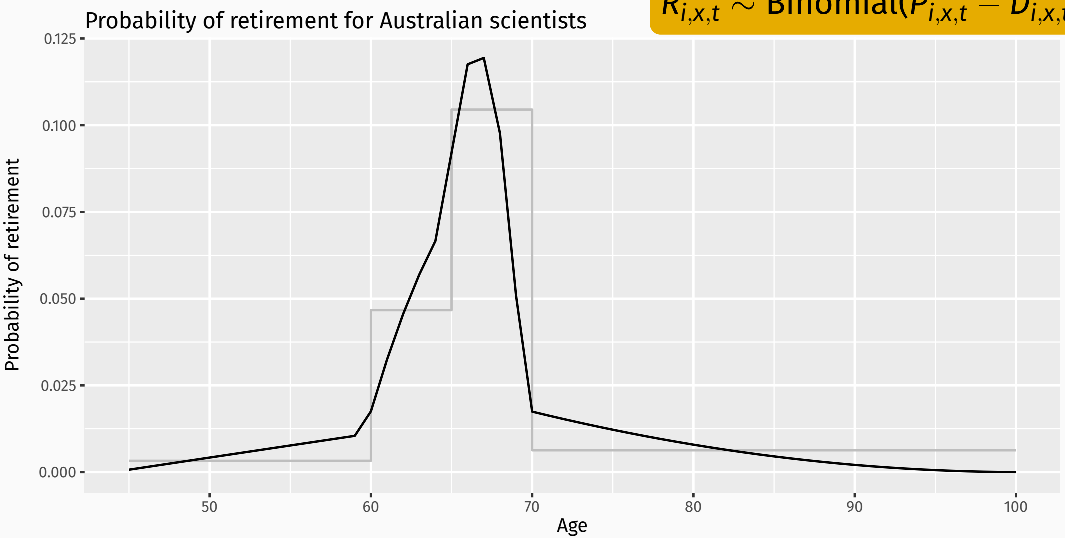
$$R_{i,x,t} \sim \text{Binomial}(P_{i,x,t} - D_{i,x,t}, r_x)$$



Retirement rates: r_x

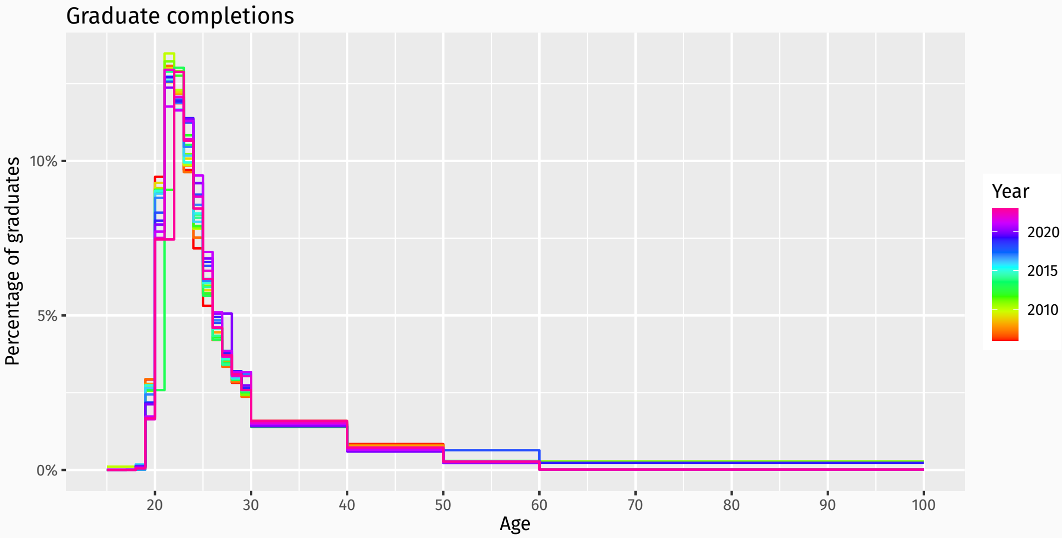
$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

$$R_{i,x,t} \sim \text{Binomial}(P_{i,x,t} - D_{i,x,t}, r_x)$$



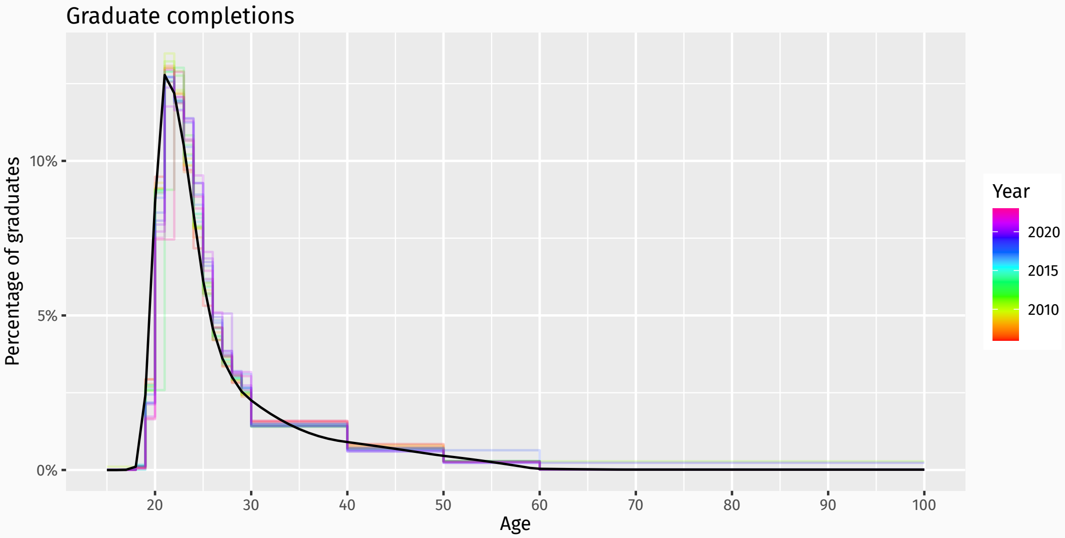
Graduate completions: g_x

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



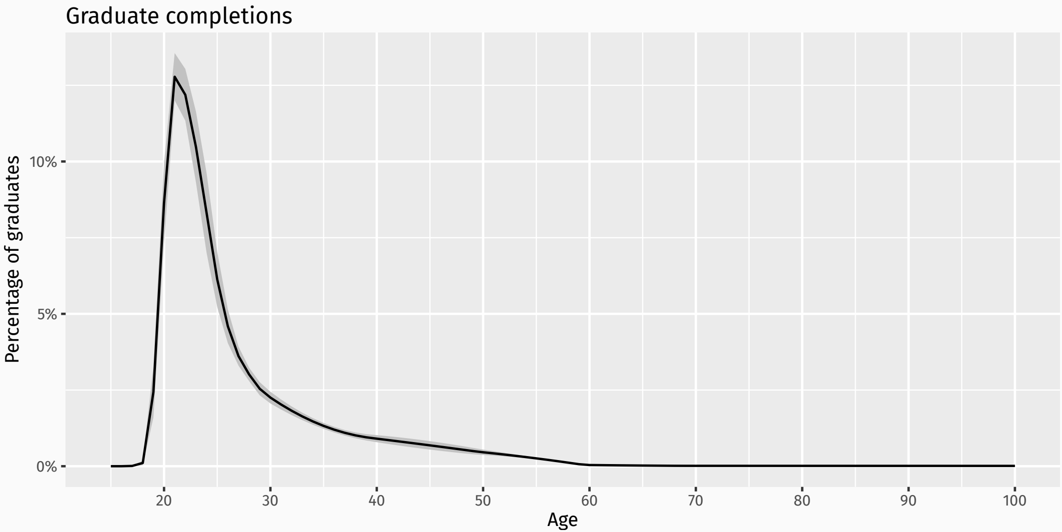
Graduate completions: g_x

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



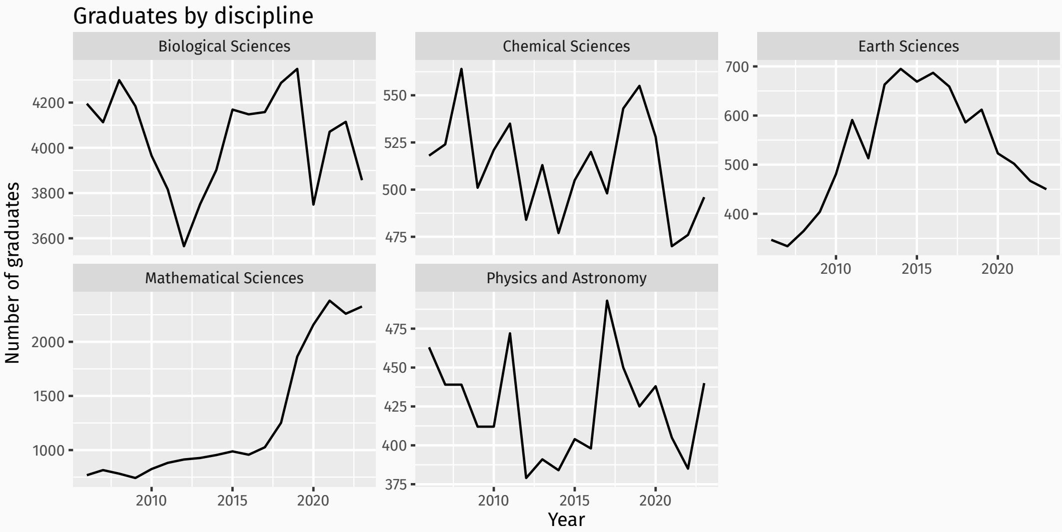
Graduate completions: g_x

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



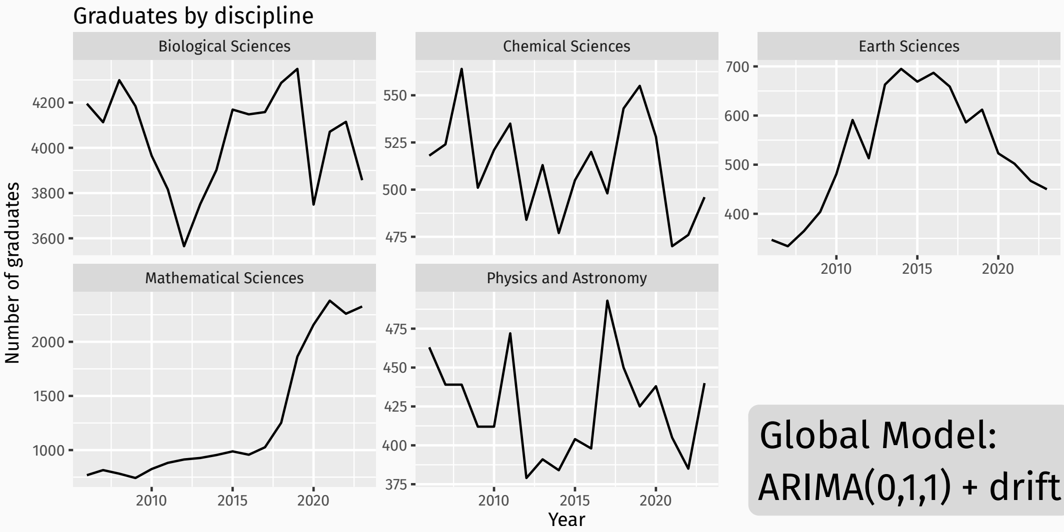
Graduate completions: $G_{i,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Graduate completions: $G_{i,t}$

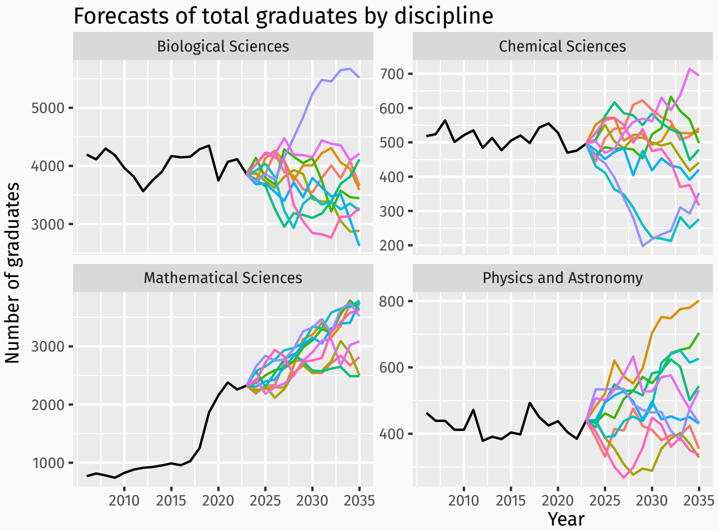
$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



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

Graduate completions: $G_{i,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



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

Remainder: $E_{x,t}$

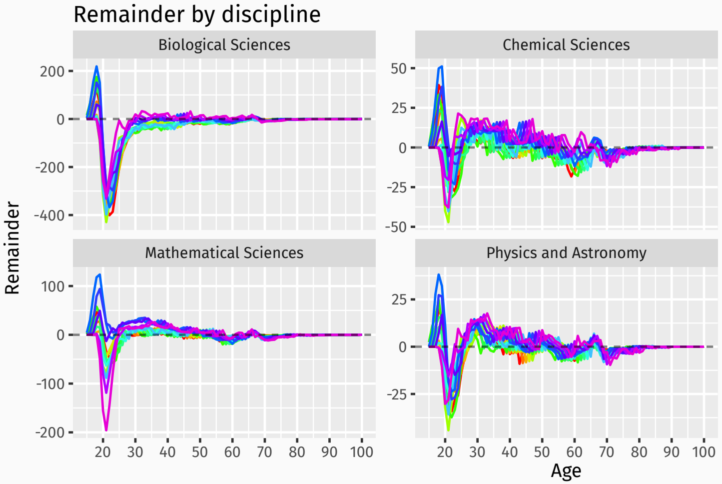
$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

$$E_{i,x,t} = P_{i,x+1,t+1} - P_{i,x,t} + D_{i,x,t} + R_{i,x,t} - g_x G_{i,t}$$

Remainder: $E_{x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

$$E_{i,x,t} = P_{i,x+1,t+1} - P_{i,x,t} + D_{i,x,t} + R_{i,x,t} - g_x G_{i,t}$$



Simulate using functional time series models, as for $q_{x,t}$.

Labour force model

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + G_{i,x,t} + E_{i,x,t}$$

i = Discipline x = Age t = Year

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

$D_{i,x,t}$ = number of deaths $\sim$ Binomial($P_{i,x,t}, q_{x,t}$)

$R_{i,x,t}$ = number of retirements $\sim$ Binomial($P_{i,x,t} - D_{i,x,t}, r_x$)

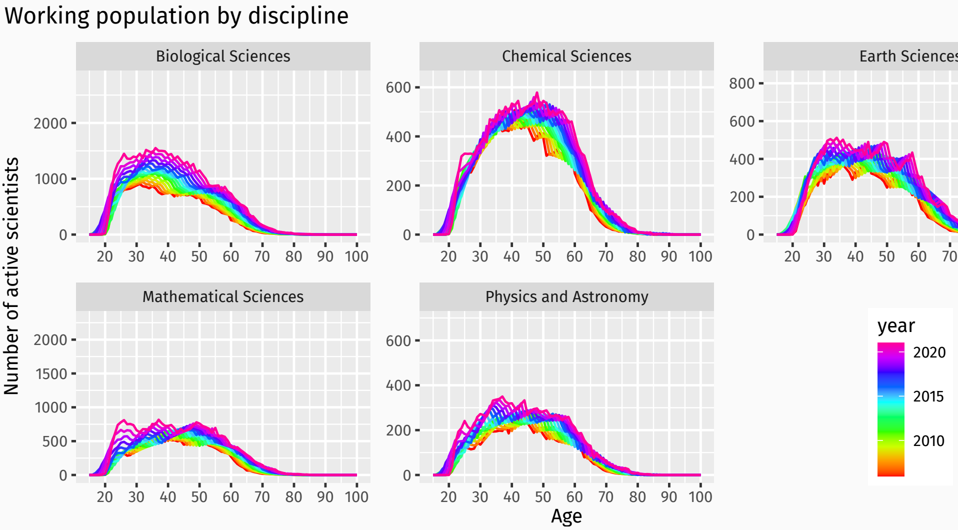
$G_{i,x,t} = g_x G_{i,t}$ = number of graduates who work in discipline i

$E_{i,x,t}$ = other changes (career changes, migration, etc.)

Simulate future sample paths of $P_{i,x,t}$ by simulating future $q_{x,t}$, $D_{i,x,t}$, $R_{i,x,t}$, $G_{i,t}$ and $E_{i,x,t}$

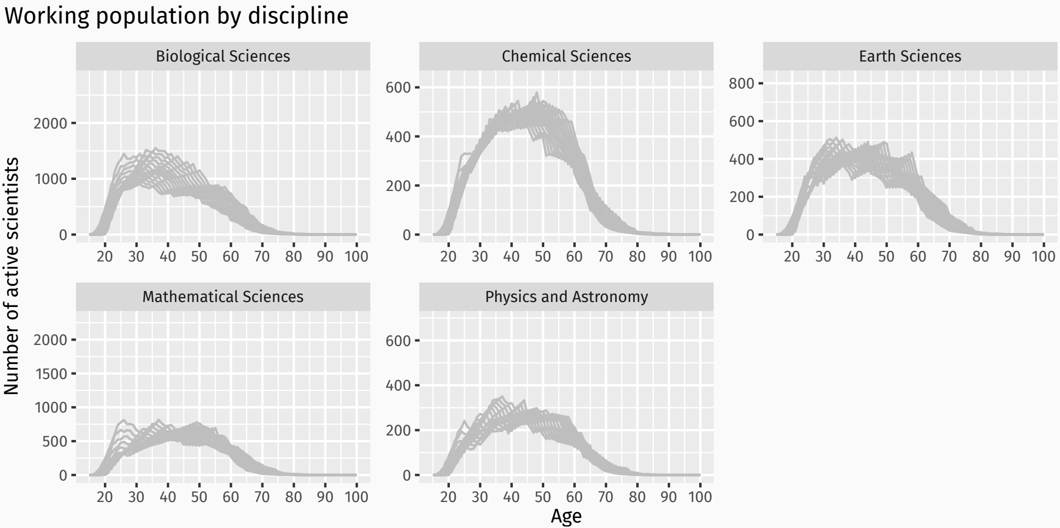
Population: $P_{i,x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



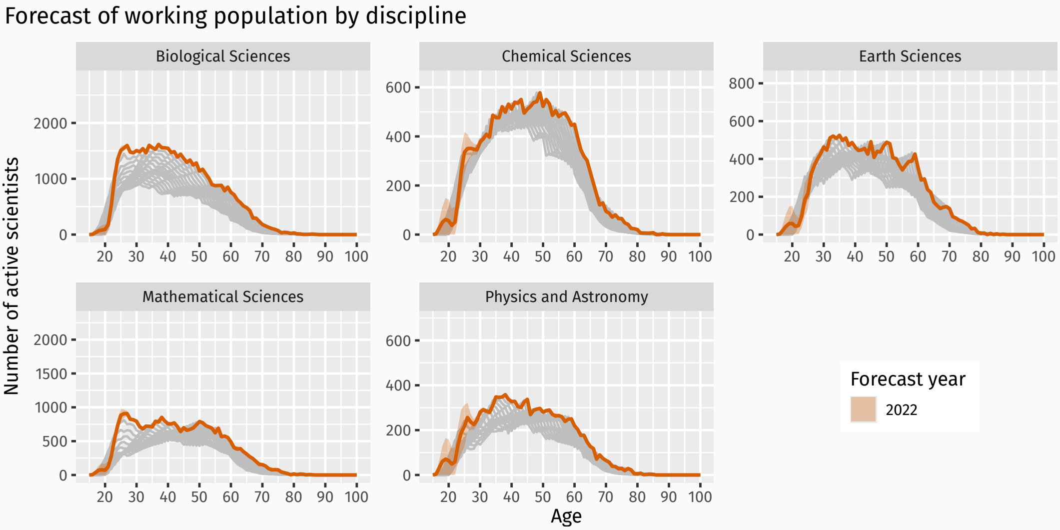
Population: $P_{i,x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



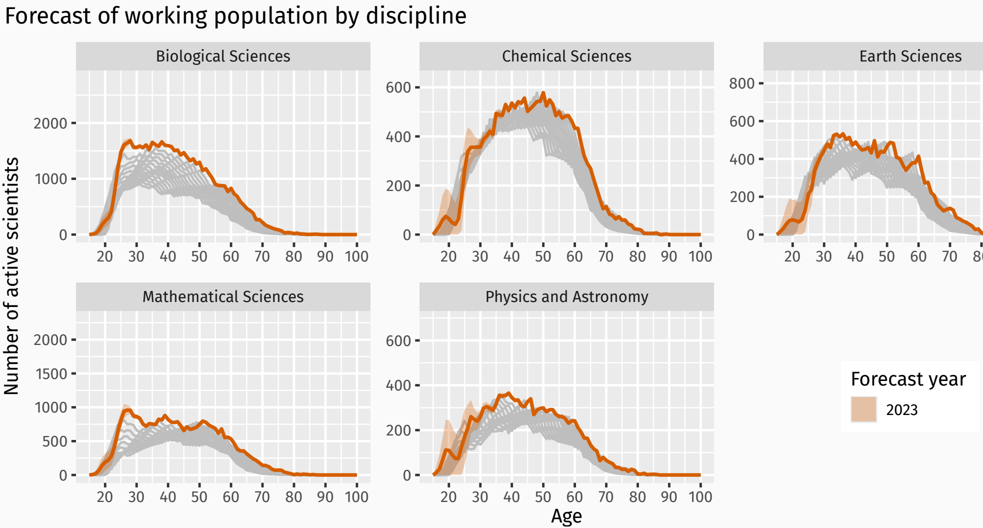
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



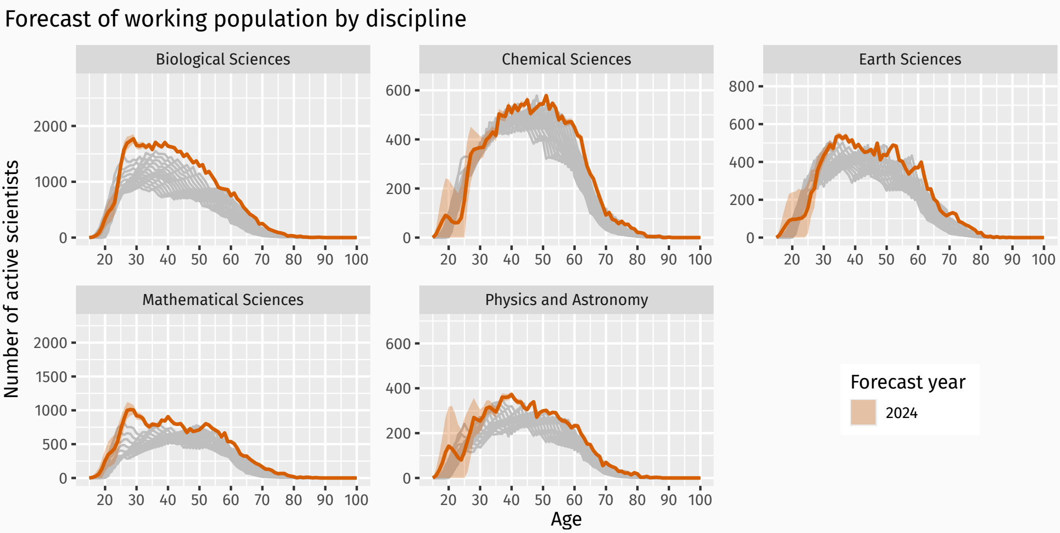
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



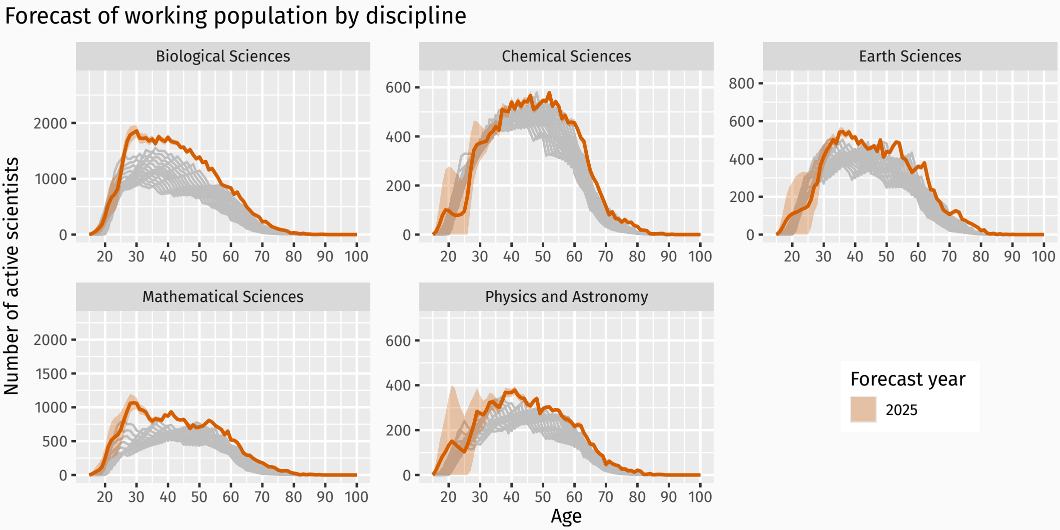
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



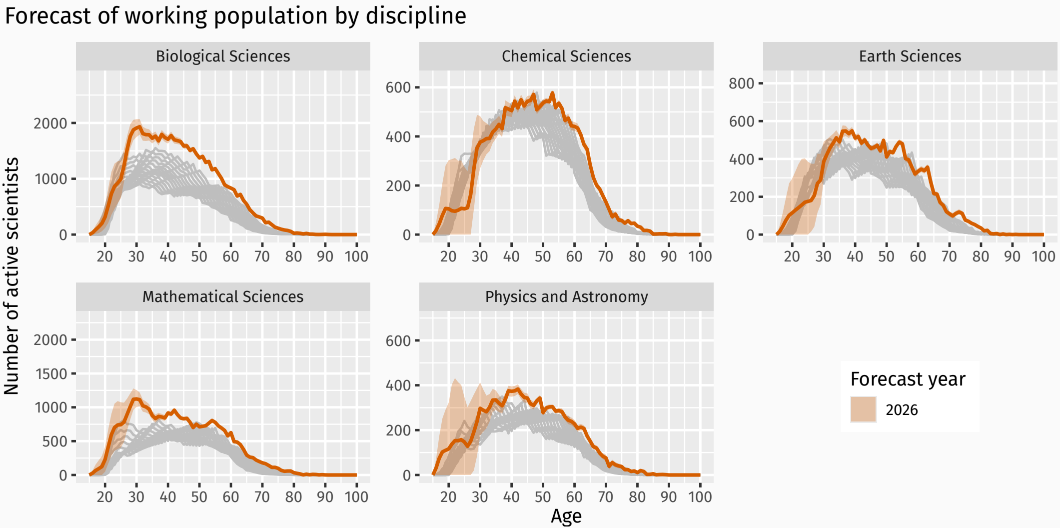
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



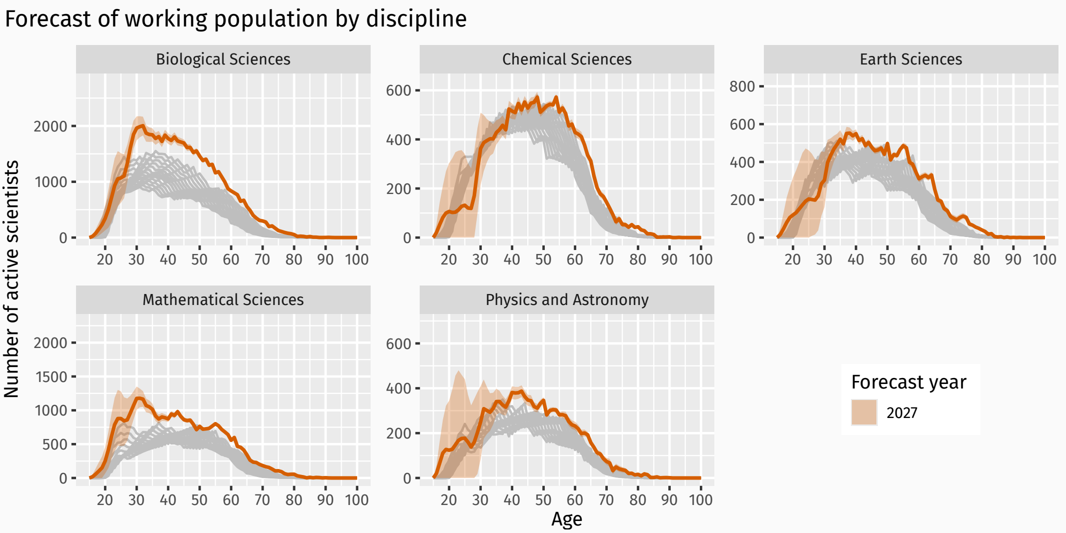
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



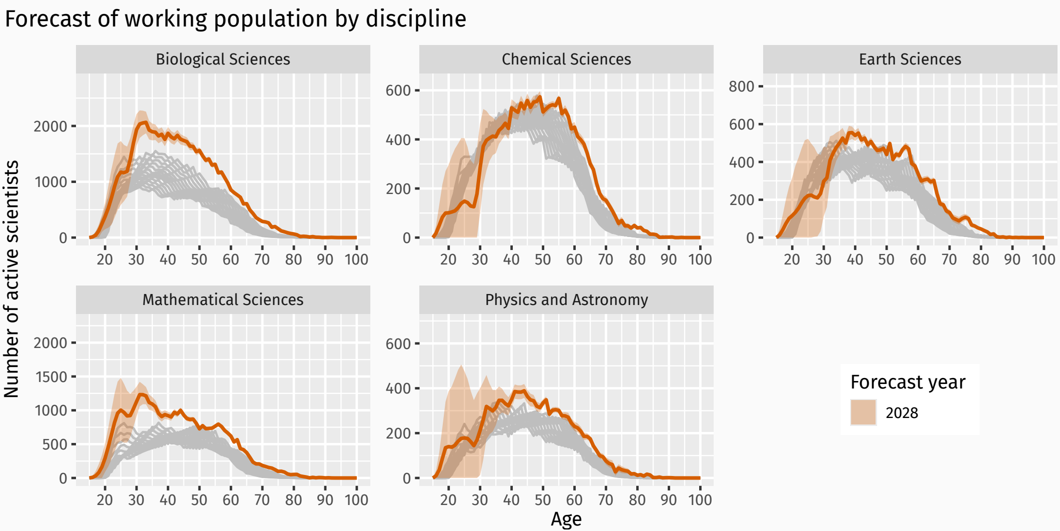
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



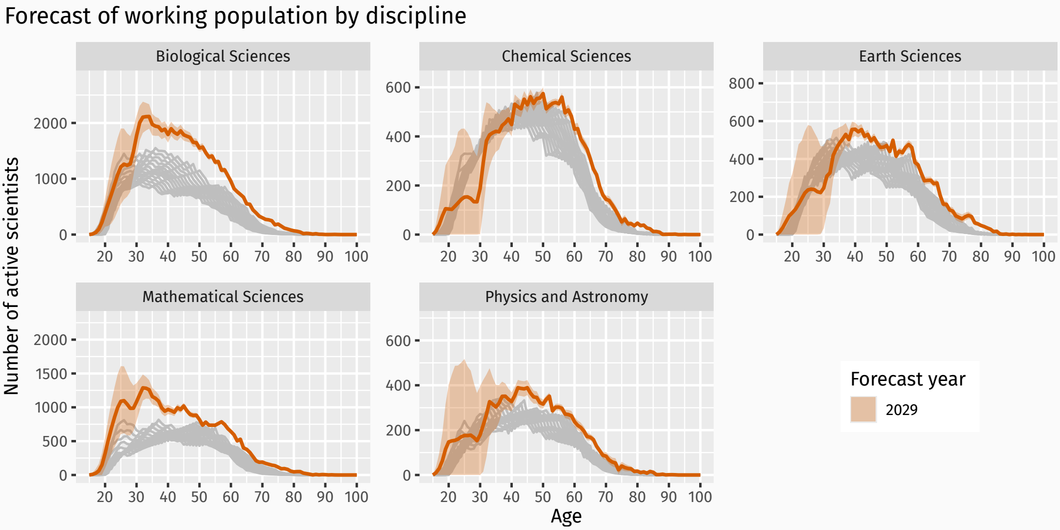
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Forecasts by discipline

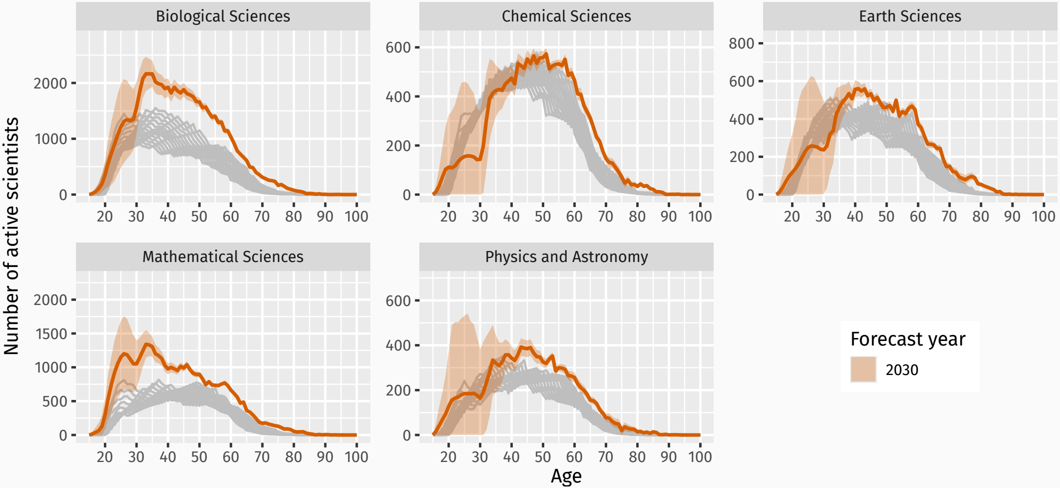
$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Forecasts by discipline

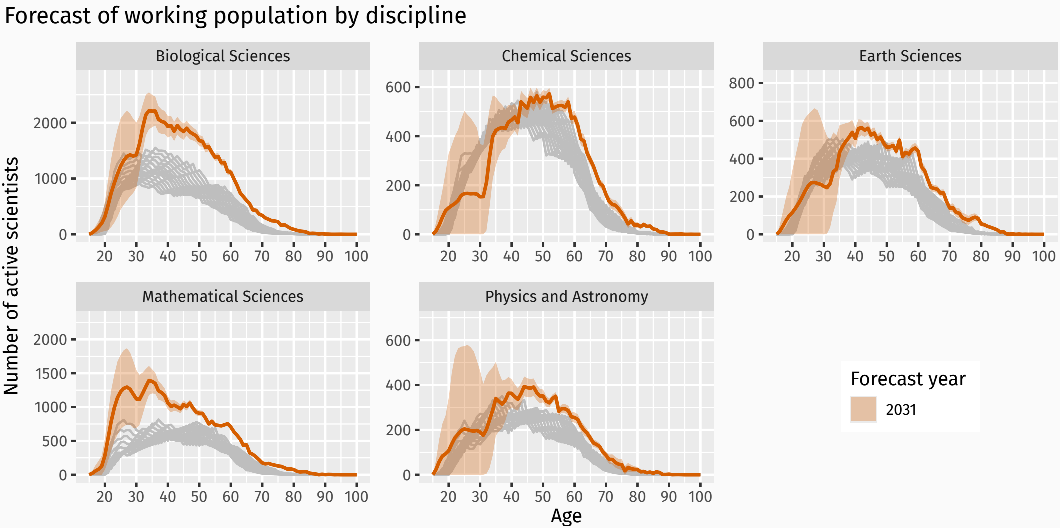
$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$

Forecast of working population by discipline



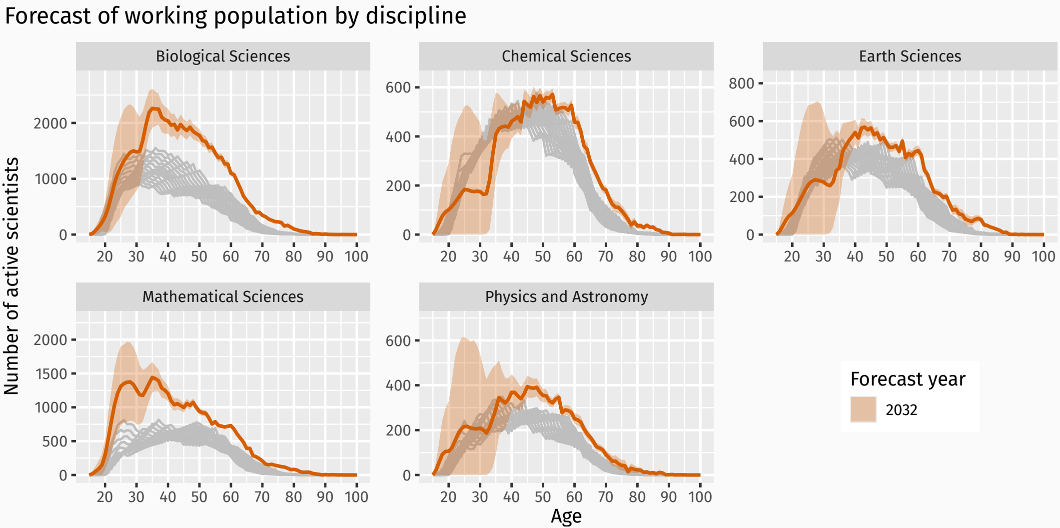
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



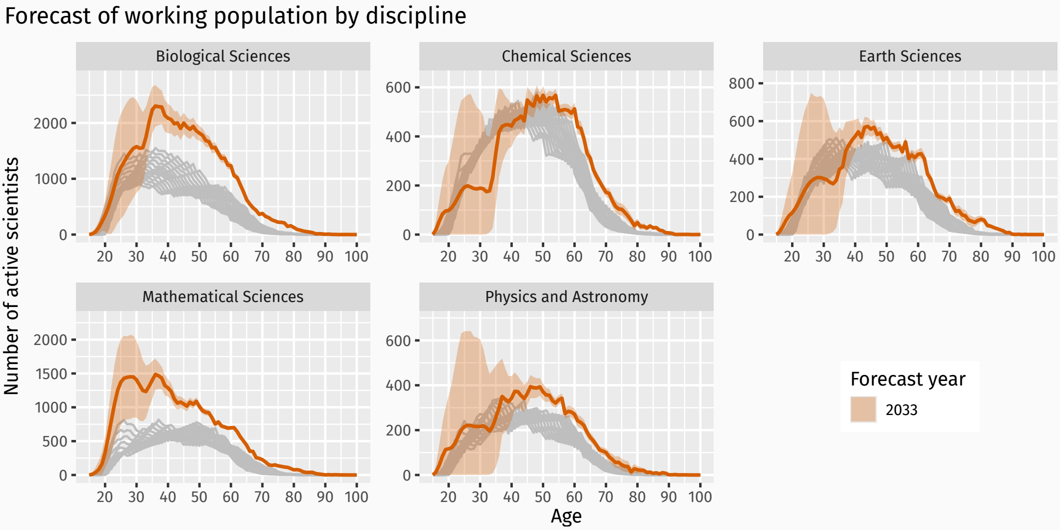
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



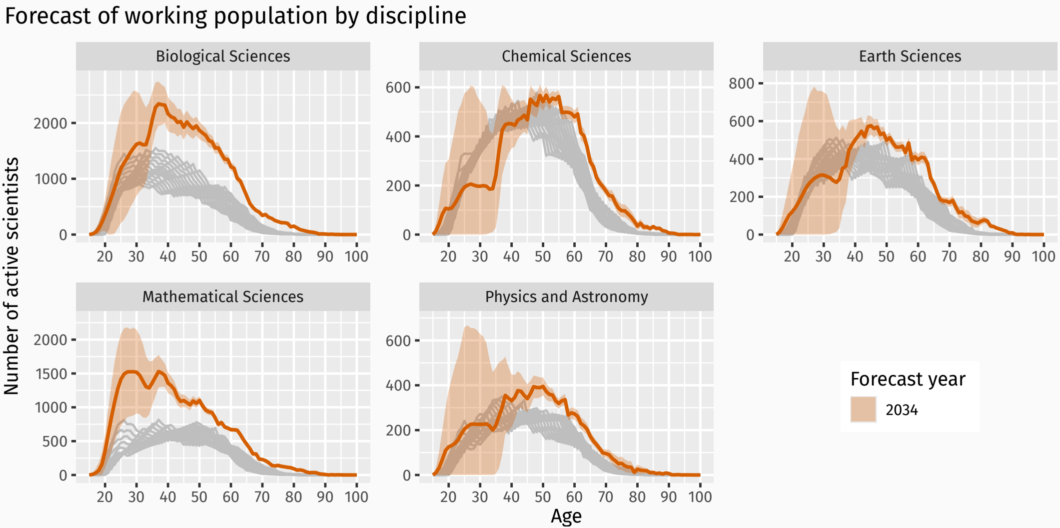
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



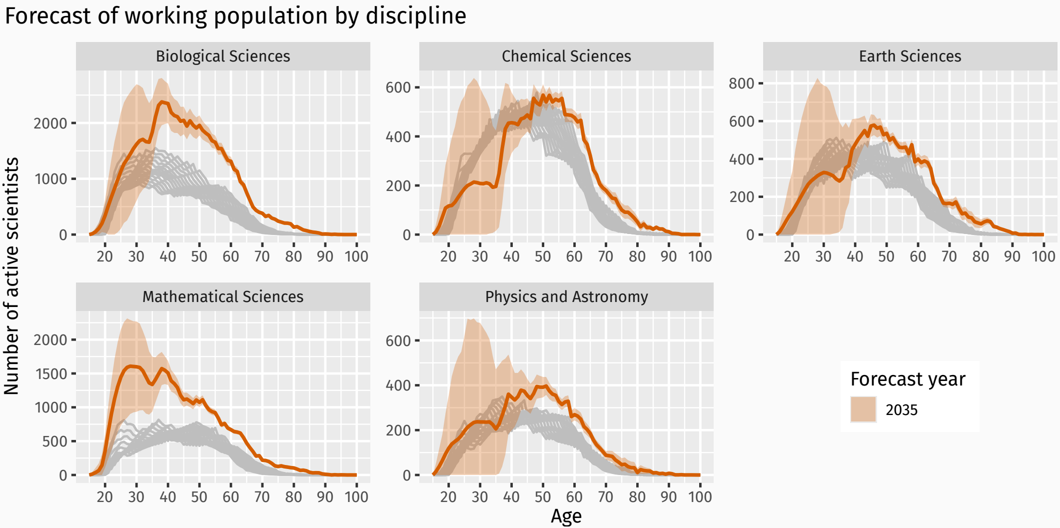
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



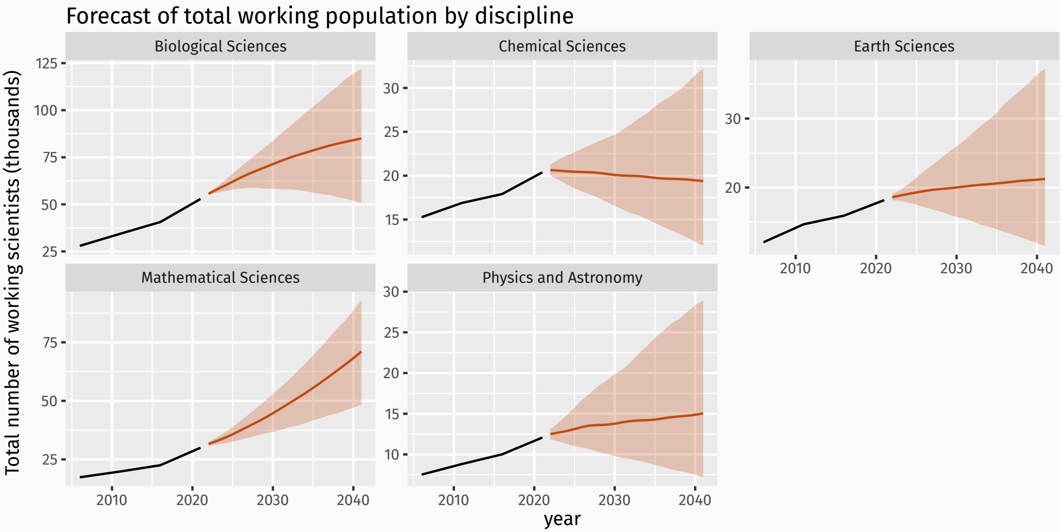
Forecasts by discipline

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Population: $\sum_x P_{i,x,t}$

$$P_{i,x+1,t+1} = P_{i,x,t} - D_{i,x,t} - R_{i,x,t} + g_x G_{i,t} + E_{i,x,t}$$



Comments

- Ignoring impact of AI, other emerging technologies, etc.
- Ignoring policy changes or exogenous global economic shifts.
- Forecasts designed to inform policy decisions, not just predict future, and so may render themselves invalid.
- Evaluation challenging due to long forecast horizon vs historical data.
- Better model possible with more detailed data
- Model applicable to other countries/sectors with similar data.

More information

robjhyndman.com/asc2025