Forecasting old-age dependency

Rob J Hyndman

MONASH University

24 February 2022

Outline

Outline

1. Demographic change and old-age dependency
2. Forecasting population age structure
3. Sustainable pension age schemes

Ageing population

Ageing population

1. advances in medical care leading to longer life
2. high fertility rates after World War II
3. large decline in fertility rates over past thirty years.

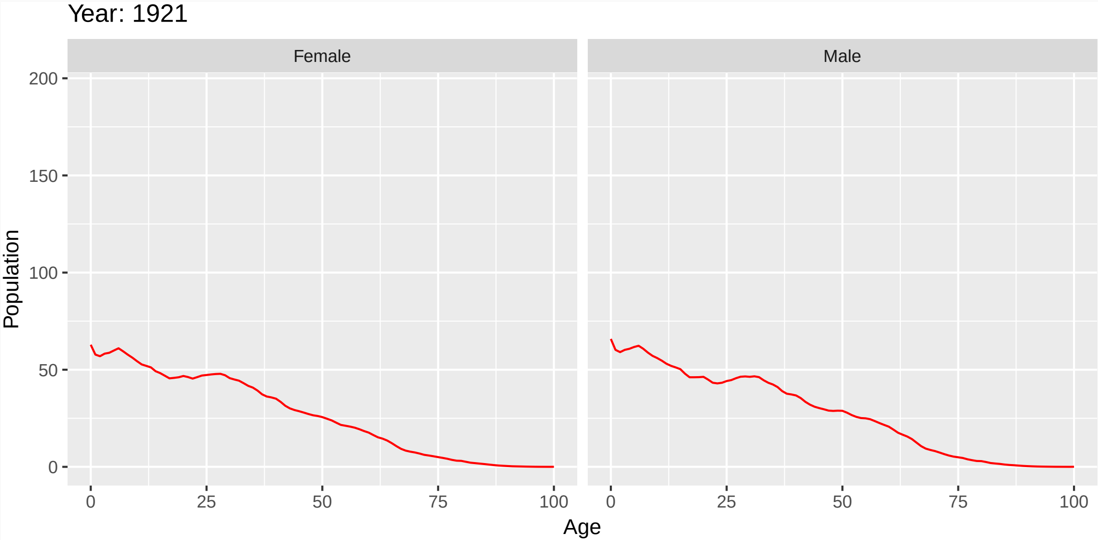
Ageing population

1. advances in medical care leading to longer life
2. high fertility rates after World War II
3. large decline in fertility rates over past thirty years.

Consequently:

- older people leaving workforce, fewer workers replacing them
- increasing old-age dependency ratio (OADR)

Population data



Old-age Dependency Ratio (OADR)

$$\text{OADR} = \frac{\text{number of people aged over pension age}}{\text{number of people aged 15 to pension age}}$$

Old-age Dependency Ratio (OADR)

$$\text{OADR} = \frac{\text{number of people aged over pension age}}{\text{number of people aged 15 to pension age}}$$

OADR in year t

$$\text{OADR}_t = \frac{\sum_{x \geq \lfloor a_t \rfloor} P_t(x) - (a_t - \lfloor a_t \rfloor) P_t(\lfloor a_t \rfloor)}{\sum_{x=15}^{\lfloor a_t \rfloor - 1} P_t(x) + (a_t - \lfloor a_t \rfloor) P_t(\lfloor a_t \rfloor)}$$

$P_t(x)$ = population age x in year t

a_t = pension age in year t

Old-age Dependency Ratio (OADR)

$$\text{OADR} = \frac{\text{number of people aged over pension age}}{\text{number of people aged 15 to pension age}}$$

OADR in year t

$$\text{OADR}_t = \frac{\sum_{x \geq \lfloor a_t \rfloor} P_t(x) - (a_t - \lfloor a_t \rfloor) P_t(\lfloor a_t \rfloor)}{\sum_{x=15}^{\lfloor a_t \rfloor - 1} P_t(x) + (a_t - \lfloor a_t \rfloor) P_t(\lfloor a_t \rfloor)}$$

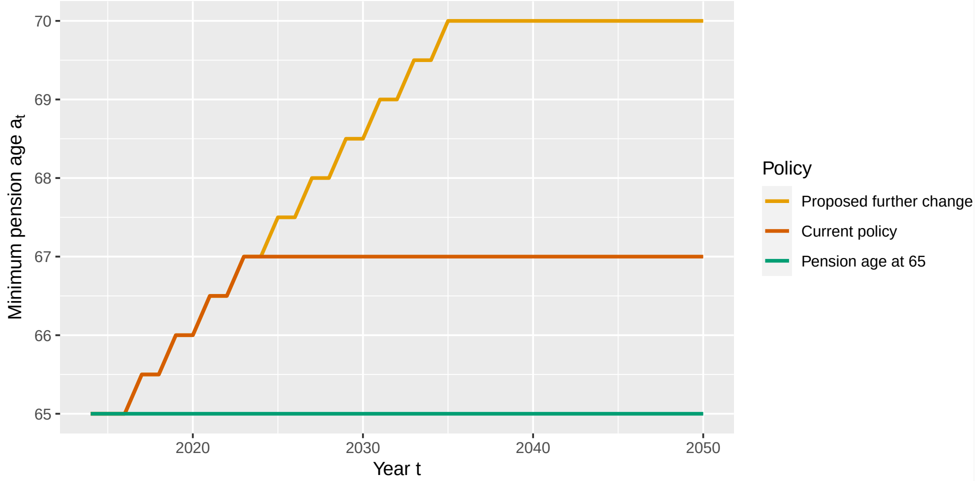
$P_t(x)$ = population age x in year t

a_t = pension age in year t

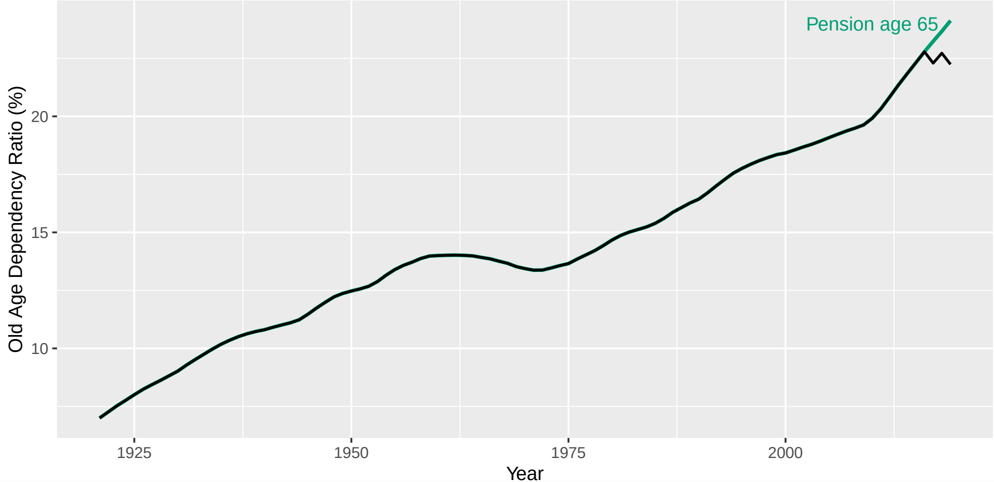
Assumptions:

- everyone retires at pension age
- pension age can only be adjusted at beginning of year
- birthdays are uniformly distributed over year

The age pension in Australia



Old Age Dependency Ratio in Australia



Outline

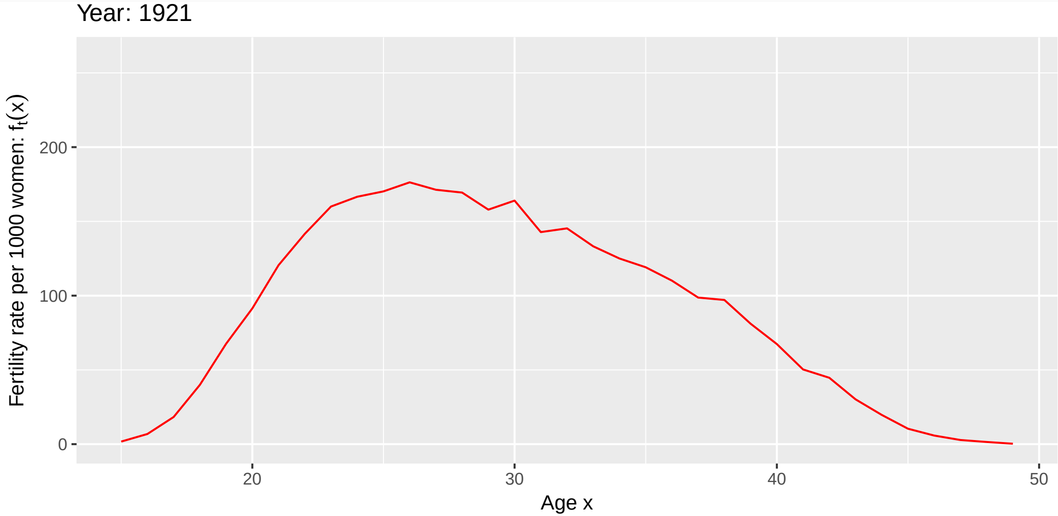
Available data

$$B_t(x) = \text{Births in calendar year } t \text{ to females age } x$$
$$D_t(x) = \text{Deaths in calendar year } t \text{ of persons age } x$$
$$P_t(x) = \text{Population age } x \text{ at 1 January year } t$$
$$E_t(x) = \text{Population age } x \text{ at 30 June year } t$$

- Ages: $x = 0, 1, 2, \ldots, 99, 100+$
- Year: $t = 1, 2, \ldots, T$
- $m_t^F(x) = D_t^F(x)/E_t^F(x)$ = female central death rates, year t
- $m_t^M(x) = D_t^M(x)/E_t^M(x)$ = male central death rates, year t
- $f_t(x) = B_t(x)/E_t^F(x)$ = fertility rates, year t

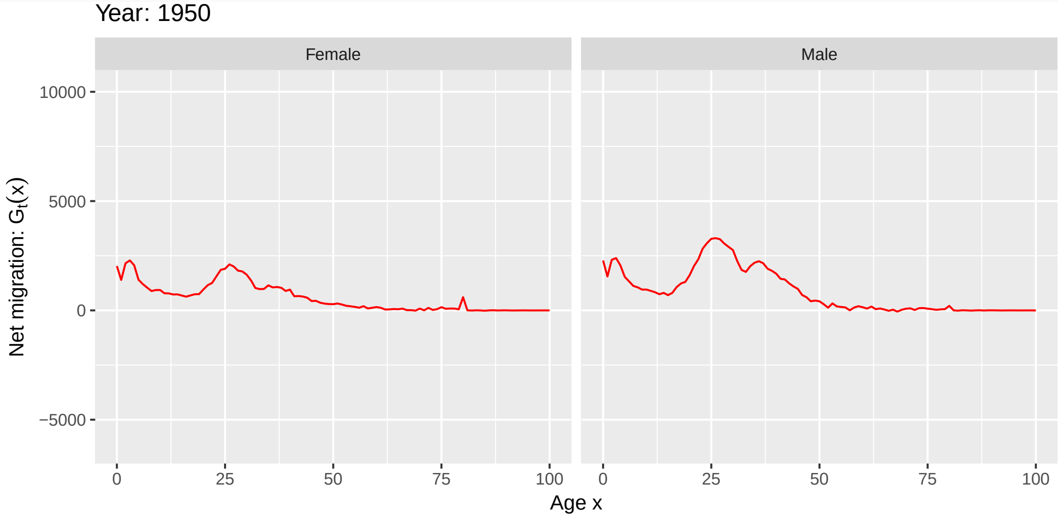
Mortality rates

Fertility rates



Migration rates

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



Functional linear model for fertility

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

$$s_t(x) = \mu(x) + \sum_{j=1}^{J} \beta_{t,j}\phi_j(x) + e_t(x)$$

- $y_t(x) = \log(f_t(x))$, year t, age x
- $s_t(x)$ = smoothed version of $y_t(x)$
- $\varepsilon_{t,x} \sim$ iid N(0,1)
- $\mu(x)$ = mean $s_t(x)$ across years.
- $\phi_j(x)$ and $\beta_{t,j}$ are principal components & scores of $s_t(x) - \mu(x)$.
- $e_t(x)$ serially uncorrelated with mean zero

Functional linear model for fertility

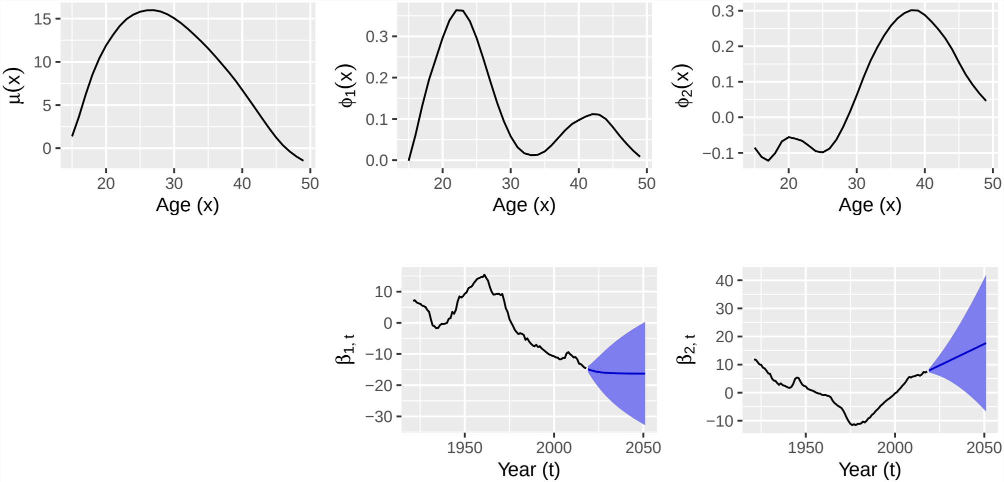
$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_{t,x}$$

$$s_t(x) = \mu(x) + \sum_{j=1}^{J} \beta_{t,j}\phi_j(x) + e_t(x)$$

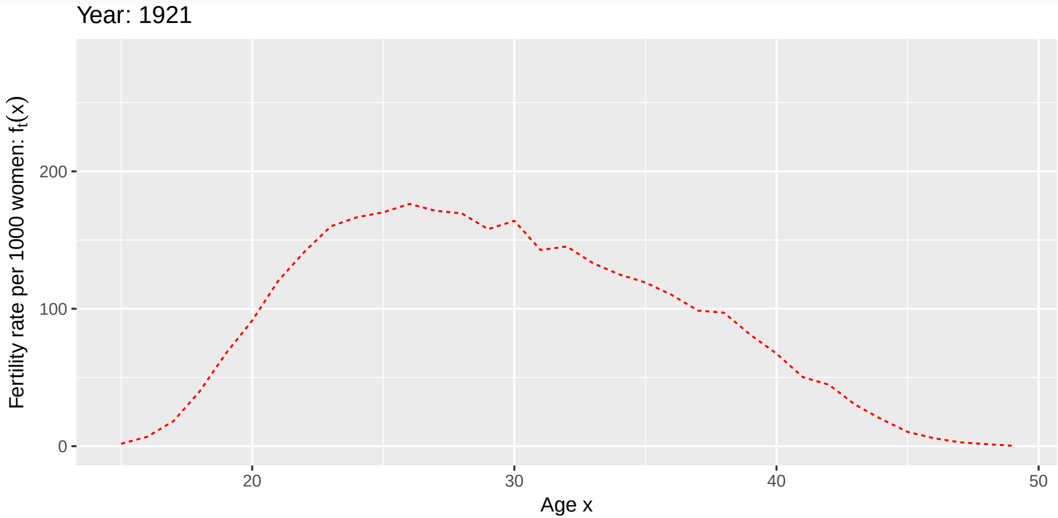
- $\{\beta_{1,j}, \dots, \beta_{T,j}\}$ is automatically modelled with ARIMA process
- $\widehat{\beta}_{T+h|T,j}$ is an h-step forecast from the ARIMA model

$$\widehat{y}_{T+h|T}(x) = \widehat{\mu}(x) + \sum_{j=1}^{J} \phi_j(x)\widehat{\beta}_{T+h|T,j}$$

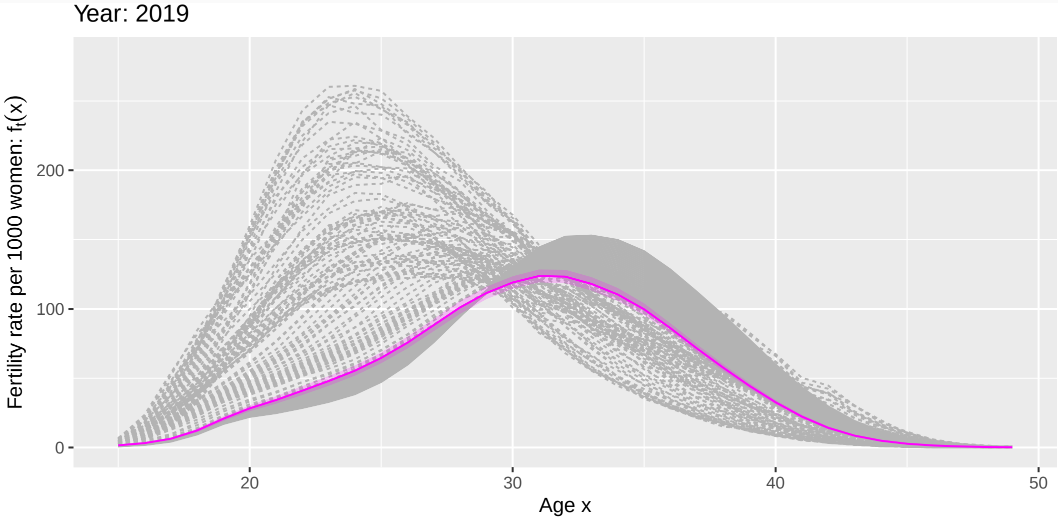
Fertility forecasts



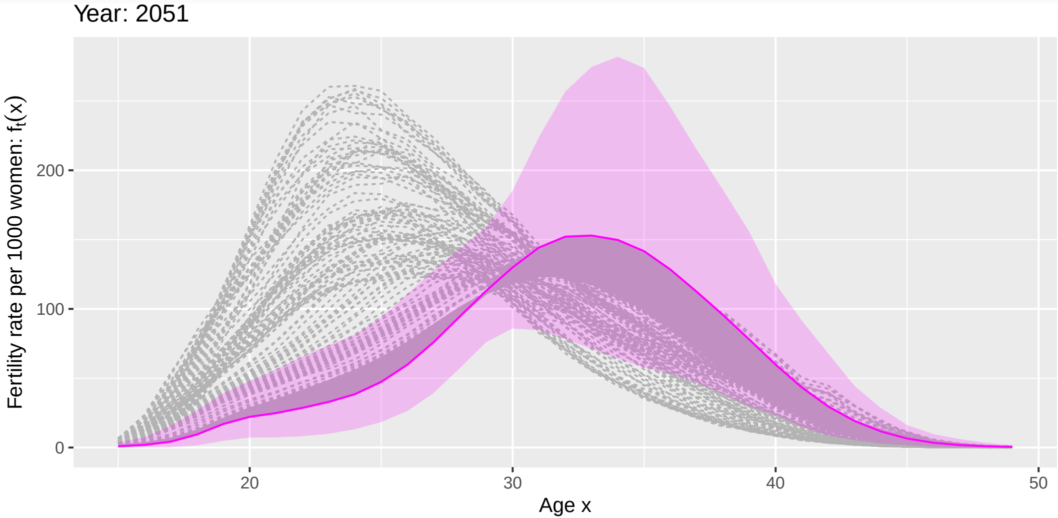
Fertility forecasts



Fertility forecasts



Fertility forecasts



Functional linear model for fertility

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

$$s_t(x) = \mu(x) + \sum_{j=1}^{J} \beta_{t,j}\phi_j(x) + e_t(x)$$

- $y_t(x) = \log(f_t(x))$
- $s_t(x)$ = smoothed version of $y_t(x)$
- $\varepsilon_{t,x} \sim$ iid N(0,1). $e_t(x)$ serially uncorrelated with mean zero
- $\mu(x)$ = mean $s_t(x)$ across years.
- $\phi_j(x)$ and $\beta_{t,j}$ are principal components & scores of $s_t(x) - \mu(x)$.
- $\{\beta_{1,j}, \dots, \beta_{T,j}\}$ is modelled with ARIMA processes

Functional linear models for mortality

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

$$s_t(x) = \mu(x) + \sum_{j=1}^{J} \beta_{t,j}\phi_j(x) + e_t(x)$$

- $y_t(x) = \log(m_t^M(x) m_t^F(x))$ and $\log(m_t^M(x)/m_t^F(x))$
- $s_t(x)$ = smoothed version of $y_t(x)$
- $\varepsilon_{t,x} \sim$ iid N(0,1). $e_t(x)$ serially uncorrelated with mean zero
- $\mu(x)$ = mean $s_t(x)$ across years.
- $\phi_j(x)$ and $\beta_{t,j}$ are principal components & scores of $s_t(x) - \mu(x)$.
- $\{\beta_{1,j}, \dots, \beta_{T,j}\}$ is modelled with ARIMA or ARFIMA processes

Functional linear models for migration

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

$$s_t(x) = \mu(x) + \sum_{j=1}^{J} \beta_{t,j}\phi_j(x) + e_t(x)$$

- $y_t(x) = (G_t^M(x) + G_t^F(x))$ and $(G_t^M(x) - G_t^F(x))$
- $s_t(x)$ = smoothed version of $y_t(x)$
- $\varepsilon_{t,x} \sim$ iid N(0,1). $e_t(x)$ serially uncorrelated with mean zero
- $\mu(x)$ = mean $s_t(x)$ across years.
- $\phi_j(x)$ and $\beta_{t,j}$ are principal components & scores of $s_t(x) - \mu(x)$.
- $\{\beta_{1,j}, \dots, \beta_{T,j}\}$ is modelled with ARIMA or ARFIMA processes

Population forecasts

- Fit similar models to fertility rates, mortality rates, migration numbers
- Constrained so sex differences do not diverge
- Assume mortality, fertility and migration are independent processes
- Simulate births, deaths and migrants for future years
- Assume births and deaths are Poisson, migrant distribution is bootstrapped
- Compute populations by age and sex for future years

Population forecasts

Slightly over-simplified

For each year of each simulated population

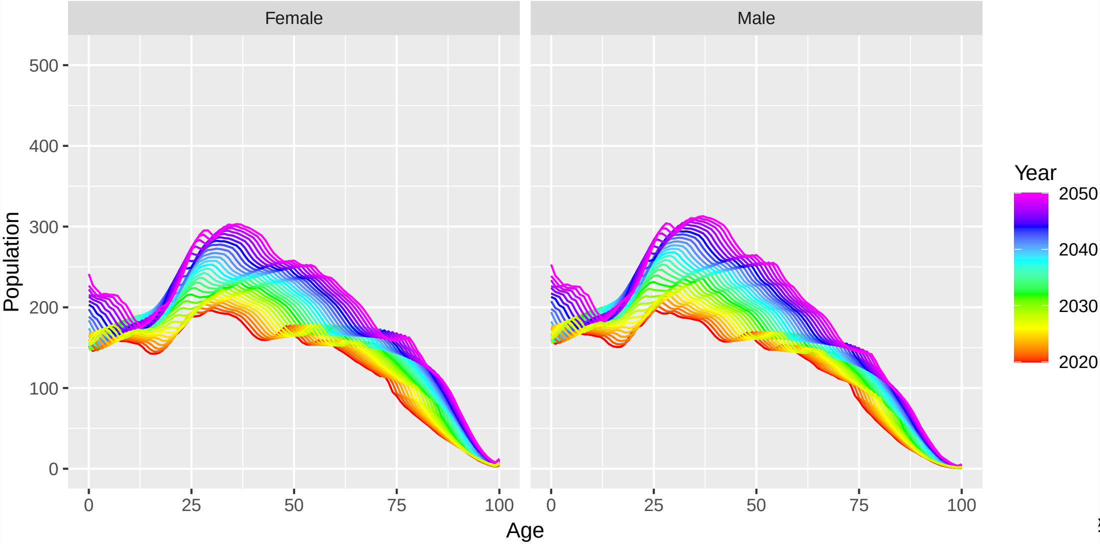
- Draw from forecast distribution of fertility rates: $f_t(x)$
- Generate number of births from Poisson population: B_t
- Draw from forecast distribution of mortality rates: $m_t^F(x)$, $m_t^M(x)$
- Generate number of deaths by age and sex:

  $D_t(x) \sim \text{Poisson}(P_t(x)m_t(x))$
- Draw from forecast distribution of migration: $G_t^F(x)$, $G_t^M(x)$
- Compute population for next year from

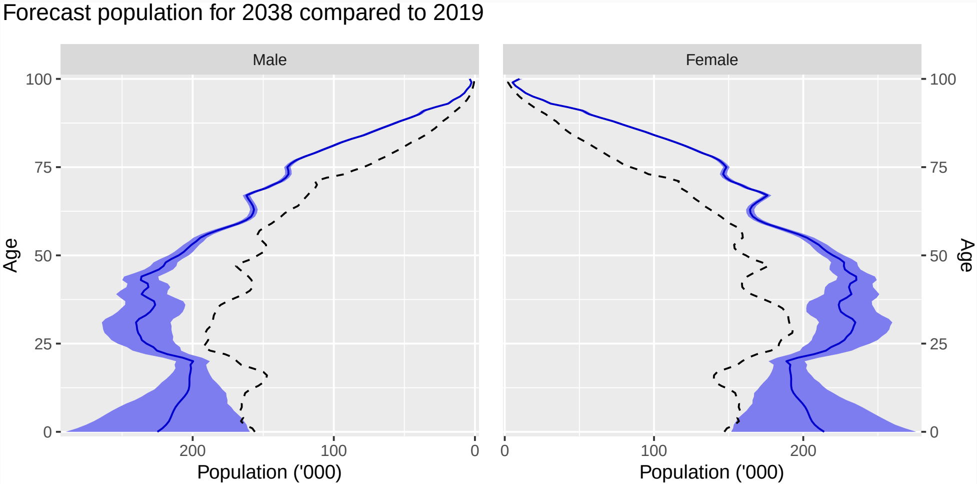
$$P_{t+1}(x) = P_t(x) + B_t 1(x = 0) - D_t(x) + G_t(x)$$

Repeat for all future years. Repeat for multiple simulated populations.

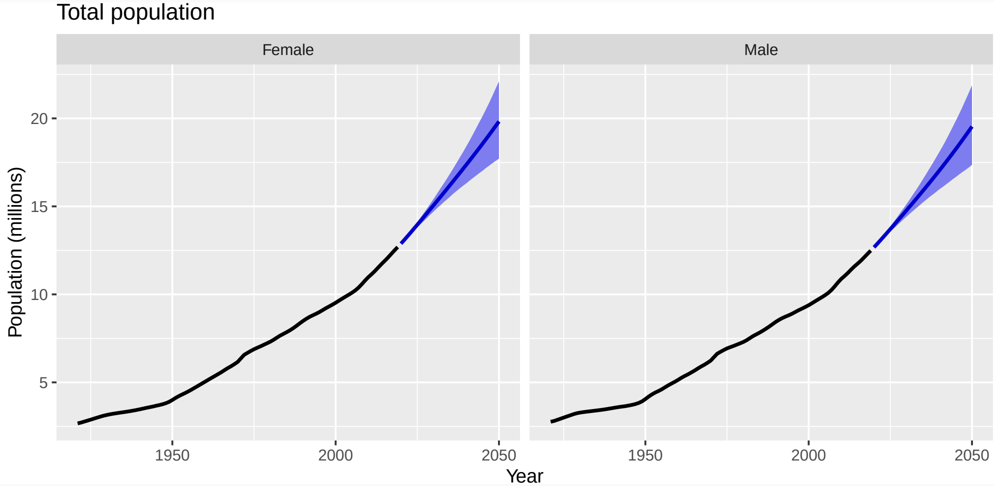
Population forecasts



Population forecasts

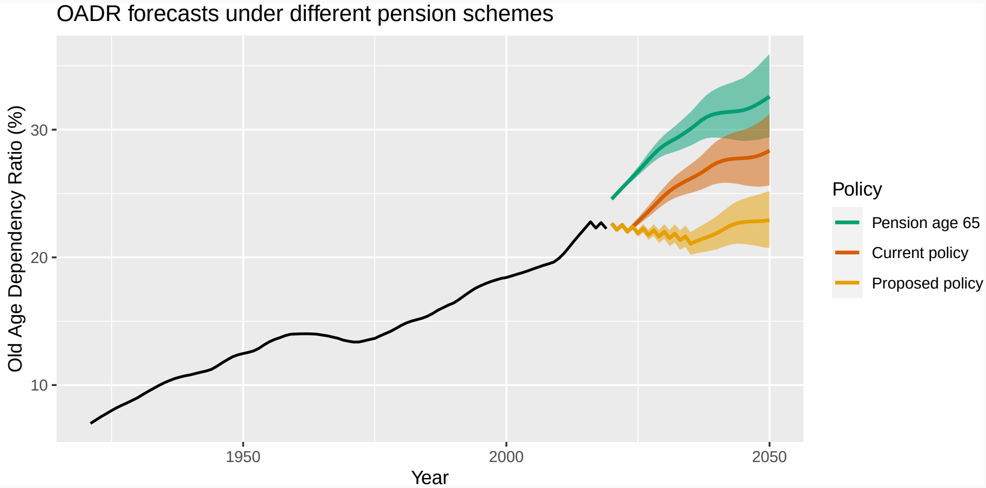


Population forecasts



Fifty-year forecasts with 80% prediction intervals.

OADR forecasts



Outline

1. Demographic change and old-age dependency
2. Forecasting population age structure
3. Sustainable pension age schemes

Sustainable pension age schemes

$$
\begin{aligned}
a_t &= \text{pension age in year } t \\
O_t &= \text{OADR in year } t \\
O^* &= \text{target OADR}
\end{aligned}
$$

Find $a_{T+1}, \dots, a_{T+H}$ where

- a_t is minimum pension age such that $O_t \leq O^*$.
- $a_{t-1} \leq a_t < a_{t-1} + 1$ to prevent (a) years where no-one is able to retire; and (b) years where retired people become ineligible for the pension.
- a_t must be in increments of one month.

Sustainable pension age schemes

Starting with $h = 1$:

1. Set $a_{T+h} = a_{T+h-1}$.
2. Increment a_{T+h} by one-month intervals until either $\widehat{O}_{T+h|T} \leq O^*$ or $a_{T+h} - a_{T+h-1} = 11$ months, where $\widehat{O}_{T+h|T}$ denotes the mean of the simulated $O_{T+h|T}$ values.

Repeat for $h = 2, \dots, H$.

Sustainable pension age schemes

Starting with $h = 1$:

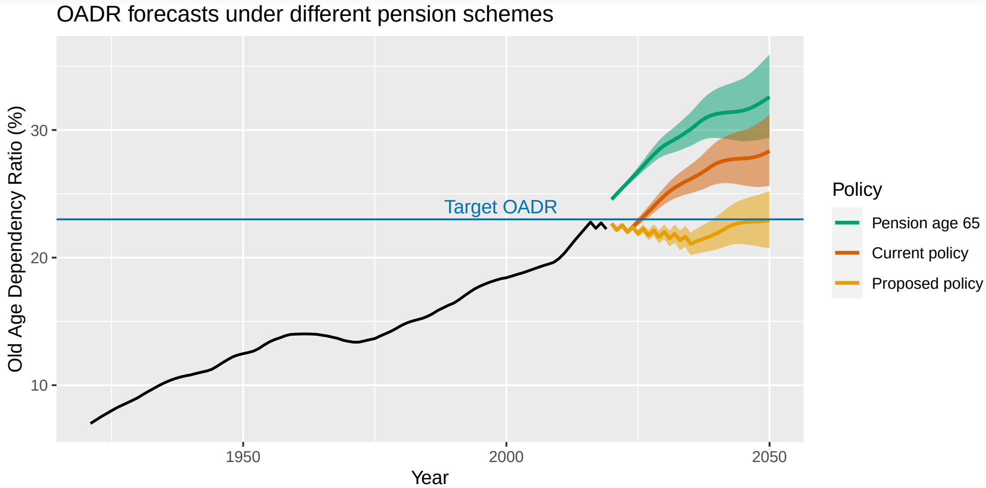
1. Set $a_{T+h} = a_{T+h-1}$.
2. Increment a_{T+h} by one-month intervals until either $\hat{O}_{T+h|T} \leq O^*$ or $a_{T+h} - a_{T+h-1} = 11$ months, where $\hat{O}_{T+h|T}$ denotes the mean of the simulated $O_{T+h|T}$ values.

Repeat for $h = 2, \dots, H$.

Confidence intervals

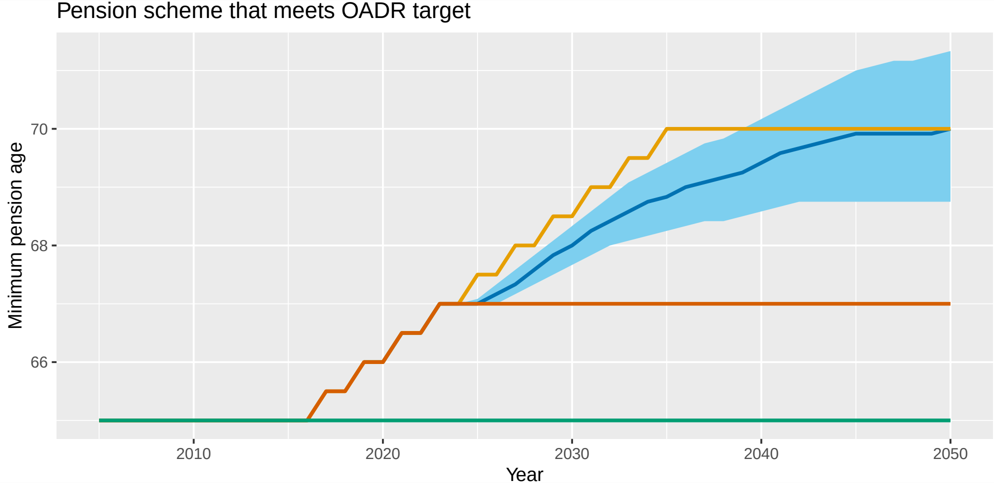
- Range of pension age schemes for which O^* is contained within 80% prediction intervals of simulated $O_{T+H|T}$ values.
- Same algorithm but $\hat{O}_{T+H|T}$ replaced by 10% and 90% quantiles.

Target OADR

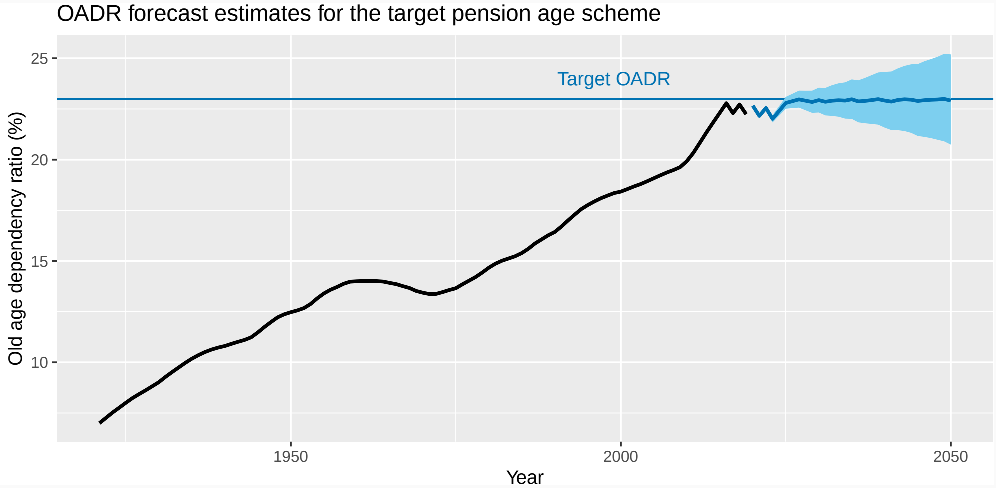


OADR forecasts under different pension schemes
Target OADR
Old Age Dependency Ratio (%)
30
20
10
1950
2000
2050
Year
Policy
Pension age 65
Current policy
Proposed policy

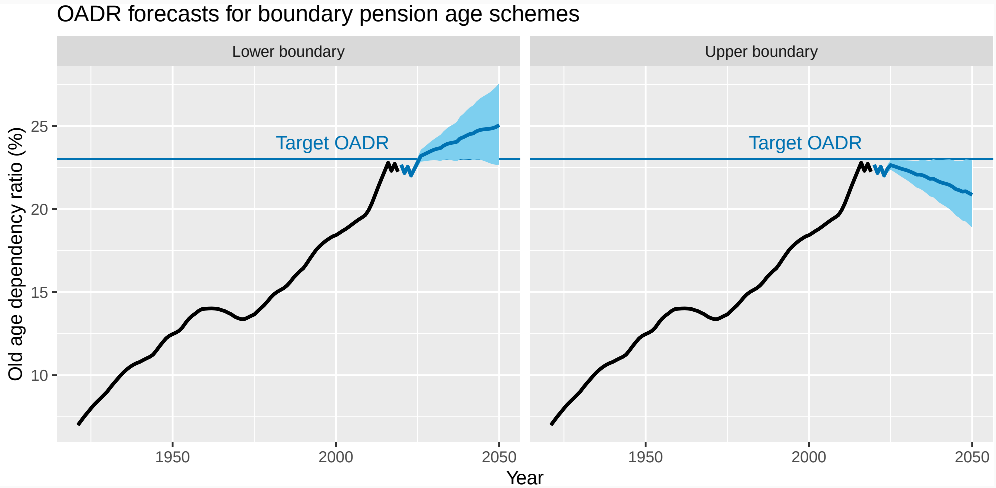
Target pension age scheme



Target pension age scheme



Target pension age scheme



OADR when pension age set at prediction interval boundaries.

Target pension age scheme

This analysis does not allow for:

- superannuation and other sources of income.
- changing GDP per capita
- COVID-19

Relevant papers

- Rob J Hyndman & Shahid Ullah (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis* **51**, 4942–4956.
- Rob J Hyndman & Heather Booth (2008) Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting* **24**(3), 323–342.
- Rob J Hyndman, Heather Booth & Farah Yasmeen (2013) Coherent mortality forecasting: the product-ratio method with functional time series models. *Demography* **50**(1), 261–283.
- Rob J Hyndman, Yijun Zeng & Han Lin Shang (2021) Forecasting the old-age dependency ratio to determine a sustainable pension age. *Australian & New Zealand Journal of Statistics*, **63**(2), 241–256.

More information

`robjhyndman.com/seminars/oadr/`

Find me at ...

robjhyndman.com

@robjhyndman

@robjhyndman

rob.hyndman@monash.edu