Common functional principal component models for mortality forecasting

Rob J Hyndman and Farah Yasmeen

Outline

1 **Functional time series**

2 Functional time series models

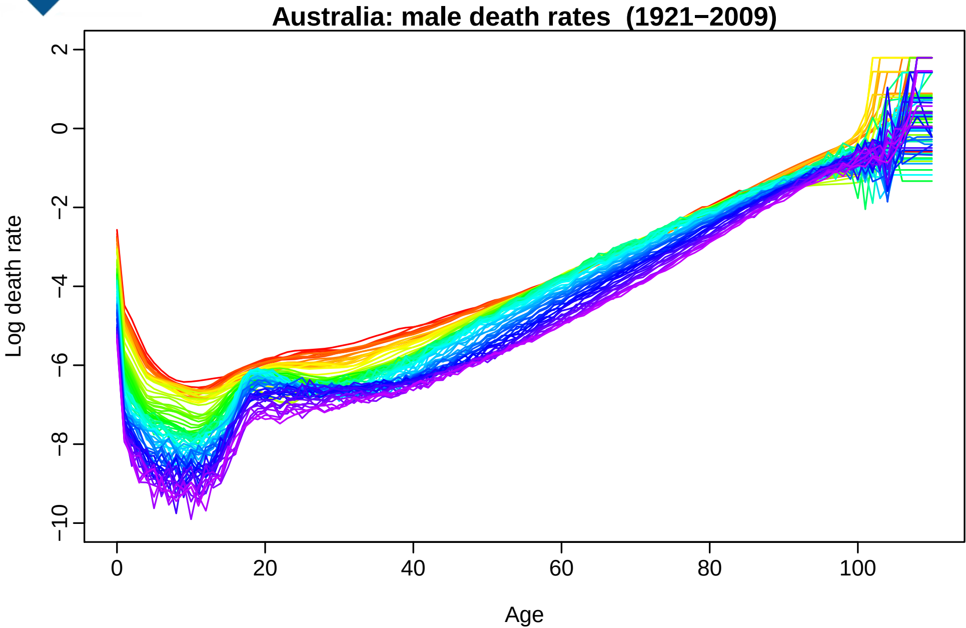
3 Common functional principal components

4 Australian mortality

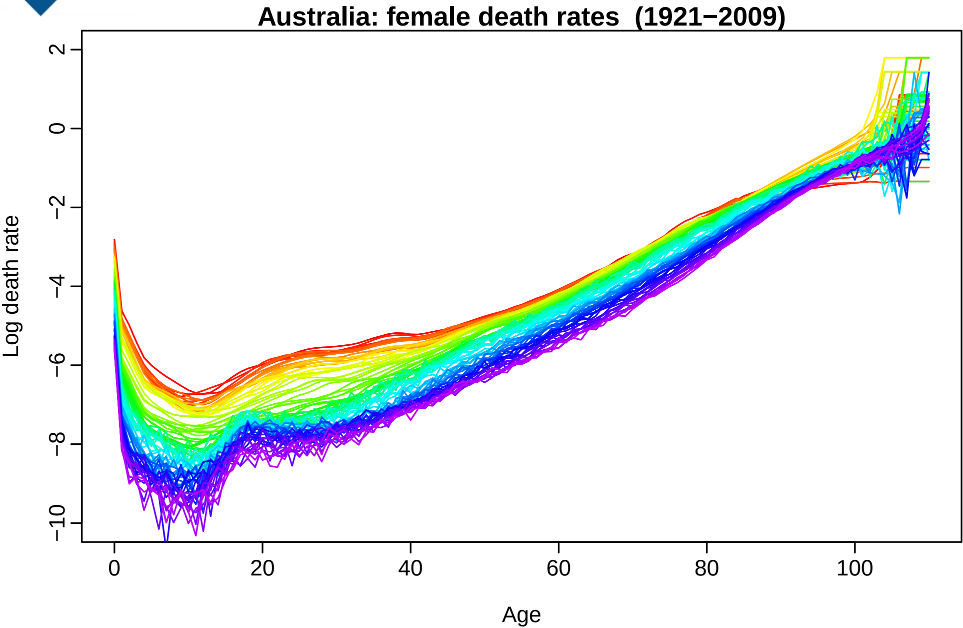
5 References

Functional time series

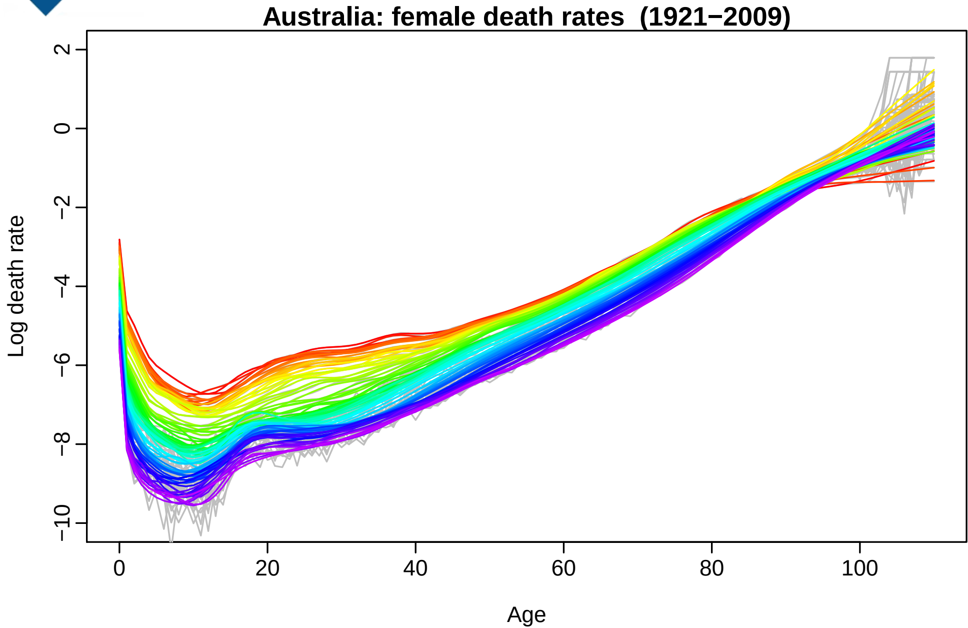
Functional time series



Functional time series



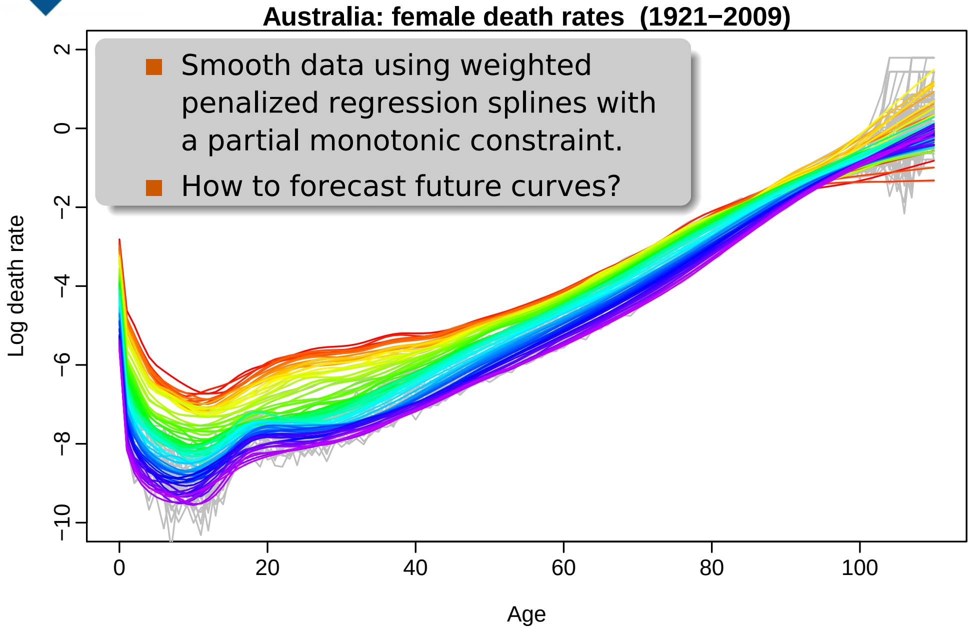
Functional time series



Functional time series

- Smooth data using weighted penalized regression splines with a partial monotonic constraint.
- How to forecast future curves?

Functional time series



- Smooth data using weighted penalized regression splines with a partial monotonic constraint.
- How to forecast future curves?

Outline

1 Functional time series

2 Functional time series models

3 Common functional principal components

4 Australian mortality

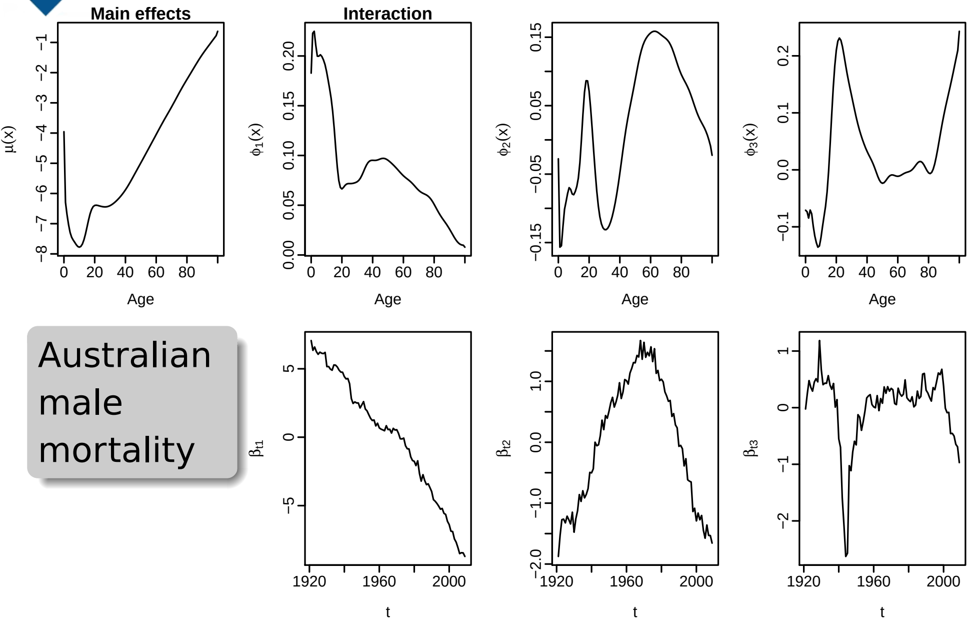
5 References

Functional time series model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,j,k}\,\phi_{j,k}(x) + r_{t,j}(x)$$

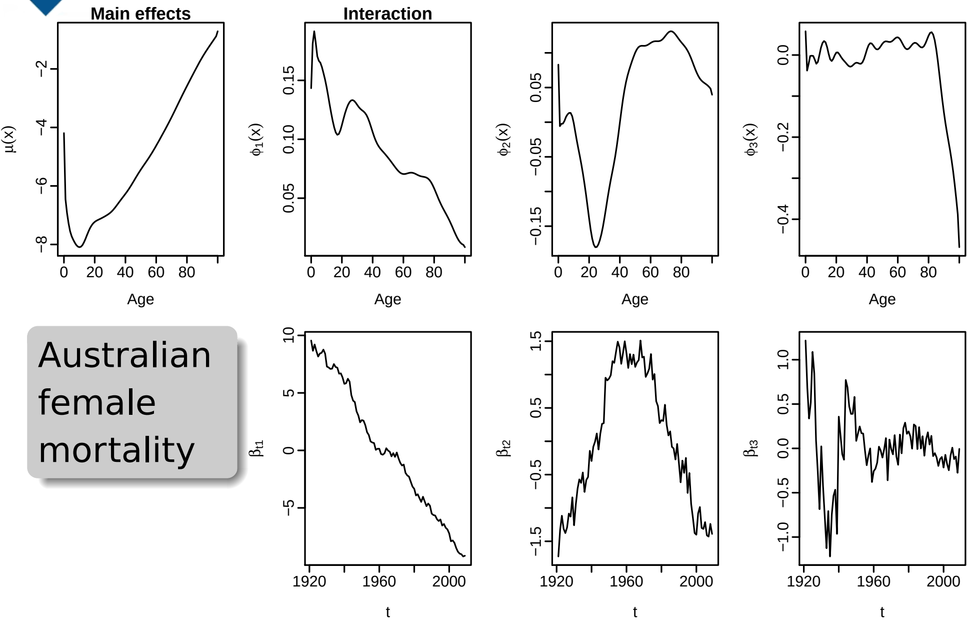
1. $f_{t,j}(x) =$ smoothed log mortality rate for age x in group j in year t.
2. Compute $\mu_j(x)$ as $\bar{f}_j(x)$ across years.
3. Compute $\beta_{t,j,k}$ and $\phi_{j,k}(x)$ using functional principal components.
4. Forecast $\{\beta_{t,j,k}\}$ using univariate time series models (e.g., ETS, ARIMA, ARFIMA, . . .)

Functional time series model



Australian male mortality

Functional time series model



Australian female mortality

The problem

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,j,k}\, \phi_{j,k}(x) + r_{t,j}(x)$$

- Groups may be males and females, or states within a country.
- Expected that groups will behave similarly.
- Fitting separate models to the groups leads to divergent forecasts when the coefficients are non-stationary.
- We require “coherent” forecasts:

The problem

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,j,k}\,\phi_{j,k}(x) + r_{t,j}(x)$$

- Groups may be males and females, or states within a country.
- Expected that groups will behave similarly.
- Fitting separate models to the groups leads to divergent forecasts when the coefficients are non-stationary.
- We require “coherent” forecasts:

$$\lim_{t\to\infty} E|f_{t,i} - f_{t,j}| < \infty \text{ for all } i \text{ and } j$$

The problem

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,j,k}\,\phi_{j,k}(x) + r_{t,j}(x)$$

- Groups may be males and females, or states within a country.
- Expected that groups will behave similarly.
- Fitting separate models to the groups leads to divergent forecasts when the coefficients are non-stationary.
- We require “coherent” forecasts:

$$\lim_{t\to\infty} \mathrm{E}\|f_{t,i} - f_{t,j}\| < \infty \text{ for all } i \text{ and } j$$

The problem

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,j,k}\,\phi_{j,k}(x) + r_{t,j}(x)$$

- Groups may be males and females, or states within a country.
- Expected that groups will behave similarly.
- Fitting separate models to the groups leads to divergent forecasts when the coefficients are non-stationary.
- We require "coherent" forecasts:

$$\lim_{t\to\infty} \mathrm{E}\big\|f_{t,j} - f_{t,i}\big\| < \infty \text{ for all } i \text{ and } j$$

The problem

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,j,k}\, \phi_{j,k}(x) + r_{t,j}(x)$$

- Groups may be males and females, or states within a country.
- Expected that groups will behave similarly.
- Fitting separate models to the groups leads to divergent forecasts when the coefficients are non-stationary.
- We require “coherent” forecasts:

$$\lim_{t\to\infty} \mathrm{E}\|f_{t,j} - f_{t,i}\| < \infty \text{ for all } i \text{ and } j$$

Outline

1 Functional time series

2 Functional time series models

3 Common functional principal components

4 Australian mortality

5 References

Partial Common Functional Principal Components

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- Coherence when $\{\gamma_{t,j,\ell} - \gamma_{t,i,\ell}\}$ is stationary for each combination of i, j and ℓ so that

$$\lim_{t\to\infty} \mathrm{E}\big\|f_{t,j} - f_{t,i}\big\| < \infty \qquad \text{for all } i \text{ and } j.$$

- Can impose coherence by requiring either cointegrated scores, or stationary scores.

Partial Common Functional Principal Components

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- Coherence when $\{\gamma_{t,j,\ell} - \gamma_{t,i,\ell}\}$ is stationary for each combination of i, j and ℓ so that

$$\lim_{t\to\infty} \mathrm{E}\|f_{t,j} - f_{t,i}\| < \infty \qquad \text{for all } i \text{ and } j.$$

- Can impose coherence by requiring either cointegrated scores, or stationary scores.

Partial Common Functional Principal Components

PCFPC(K,L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- Coherence when $\{\gamma_{t,j,\ell} - \gamma_{t,i,\ell}\}$ is stationary for each combination of i, j and ℓ so that

$$\lim_{t\to\infty} \mathrm{E}\|f_{t,j} - f_{t,i}\| < \infty \qquad \text{for all } i \text{ and } j.$$

- Can impose coherence by requiring either cointegrated scores, or stationary scores.

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- **Model 1: PCFPC($K, 0$)**. No idiosyncratic principal components in the model.
- **Model 2: PCFPC(K, L)** with a coherence constraint. For each ℓ, $\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}$ is stationary for all i, j.
- **Model 3: PCFPC(K, L)** with a coherence constraint. For each ℓ and j, $\{\gamma_{t,\ell,j}\}$ is stationary.
- **Model 4: PCFPC($0, L$)**. All principal components and scores are idiosyncratic.

Partial Common Functional Principal Components

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- **Model 1: PCFPC($K, 0$)**. No idiosyncratic principal components in the model.
- **Model 2: PCFPC(K, L)** with a coherence constraint. For each ℓ, $\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}$ is stationary for all i, j.
- **Model 3: PCFPC(K, L)** with a coherence constraint. For each ℓ and j, $\{\gamma_{t,\ell,j}\}$ is stationary.
- **Model 4: PCFPC($0, L$)**. All principal components and scores are idiosyncratic.

Partial Common Functional Principal Components

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- **Model 1: PCFPC($K, 0$).** No idiosyncratic principal components in the model.
- **Model 2: PCFPC(K, L)** with a coherence constraint. For each ℓ, $\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}$ is stationary for all i, j.
- **Model 3: PCFPC(K, L)** with a coherence constraint. For each ℓ and j, $\{\gamma_{t,\ell,j}\}$ is stationary.
- **Model 4: PCFPC($0, L$).** All principal components and scores are idiosyncratic.

Partial Common Functional Principal Components

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- **Model 1: PCFPC($K, 0$)**. No idiosyncratic principal components in the model.
- **Model 2: PCFPC(K, L)** with a coherence constraint. For each ℓ, $\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}$ is stationary for all i, j.
- **Model 3: PCFPC(K, L)** with a coherence constraint. For each ℓ and j, $\{\gamma_{t,\ell,j}\}$ is stationary.
- **Model 4: PCFPC($0, L$)**. All principal components and scores are idiosyncratic.

Partial Common Functional Principal Components

PCFPC(K, L) model

$$f_{t,j}(x) = \mu_j(x) + \sum_{k=1}^{K} \beta_{t,k}\phi_k(x) + \sum_{\ell=1}^{L} \gamma_{t,j,\ell}\psi_{j,\ell}(x) + \varepsilon_{t,j}(x)$$

- **Model 1: PCFPC($K, 0$)**. No idiosyncratic principal components in the model.
- **Model 2: PCFPC(K, L)** with a coherence constraint. For each ℓ, $\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}$ is stationary for all i, j.
- **Model 3: PCFPC(K, L)** with a coherence constraint. For each ℓ and j, $\{\gamma_{t,\ell,j}\}$ is stationary.
- **Model 4: PCFPC($0, L$)**. All principal components and scores are idiosyncratic.

Outline

1 Functional time series

2 Functional time series models

3 Common functional principal components

4 Australian mortality

5 References

Australian mortality

- Data obtained from Human Mortality Database.
- All data smoothed (independently for each year) using penalized regression splines with monotonicity constraint above age 65.
- $K = L = 6$.
- ARIMA models for common PC scores.
- ARFIMA models for stationary PC scores with $0 < d < 0.5$.
- VECM using the Johansen procedure for cointegrated PC scores.

Australian mortality

- Data obtained from Human Mortality Database.
- All data smoothed (independently for each year) using penalized regression splines with monotonicity constraint above age 65.
- $K = L = 6$.
- ARIMA models for common PC scores.
- ARFIMA models for stationary PC scores with $0 < d < 0.5$.
- VECM using the Johansen procedure for cointegrated PC scores.

Australian mortality

- Data obtained from Human Mortality Database.
- All data smoothed (independently for each year) using penalized regression splines with monotonicity constraint above age 65.
- $K = L = 6$.
- ARIMA models for common PC scores.
- ARFIMA models for stationary PC scores with $0 < d < 0.5$.
- VECM using the Johansen procedure for cointegrated PC scores.

Australian mortality

- Data obtained from Human Mortality Database.
- All data smoothed (independently for each year) using penalized regression splines with monotonicity constraint above age 65.
- $K = L = 6$.
- ARIMA models for common PC scores.
- ARFIMA models for stationary PC scores with $0 < d < 0.5$.
- VECM using the Johansen procedure for cointegrated PC scores.

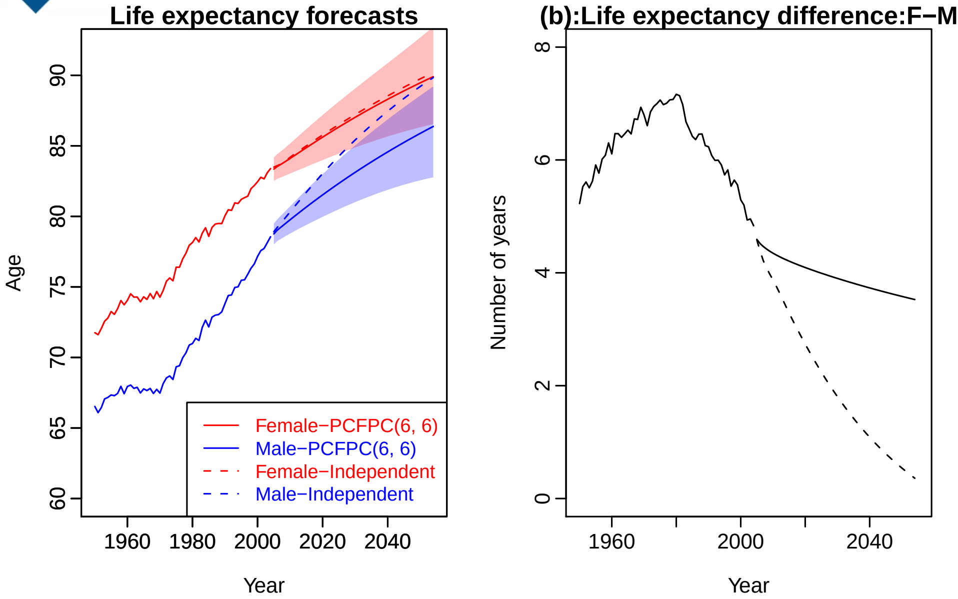
Australian mortality

- Data obtained from Human Mortality Database.
- All data smoothed (independently for each year) using penalized regression splines with monotonicity constraint above age 65.
- $K = L = 6$.
- ARIMA models for common PC scores.
- ARFIMA models for stationary PC scores with $0 < d < 0.5$.
- VECM using the Johansen procedure for cointegrated PC scores.

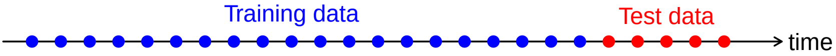
Australian mortality

- Data obtained from Human Mortality Database.
- All data smoothed (independently for each year) using penalized regression splines with monotonicity constraint above age 65.
- $K = L = 6$.
- ARIMA models for common PC scores.
- ARFIMA models for stationary PC scores with $0 < d < 0.5$.
- VECM using the Johansen procedure for cointegrated PC scores.

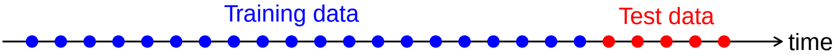
Australian mortality



Experimental set up

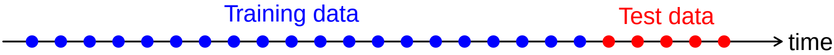


Experimental set up

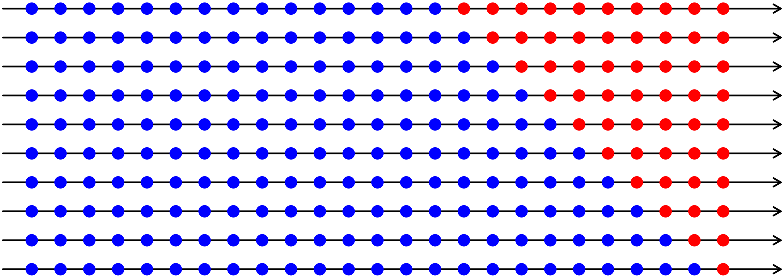


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

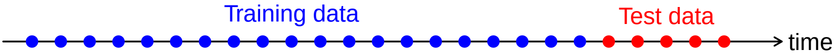
Experimental set up



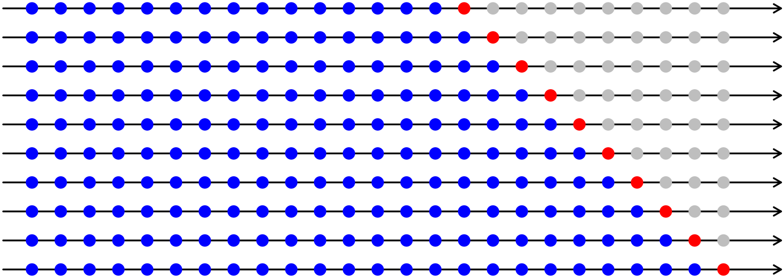
Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead



Experimental set up

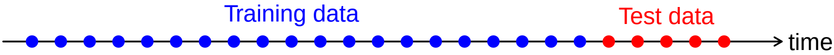


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

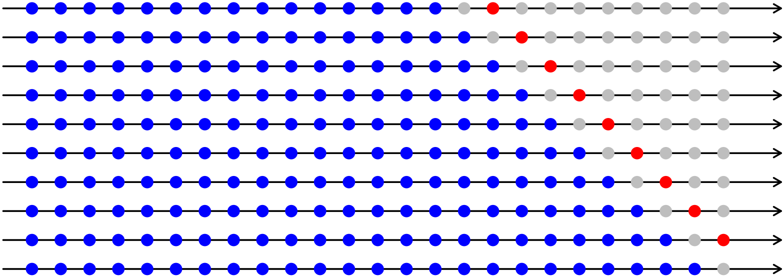


$h = 1$

Experimental set up

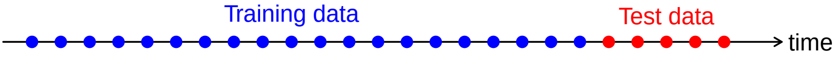


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

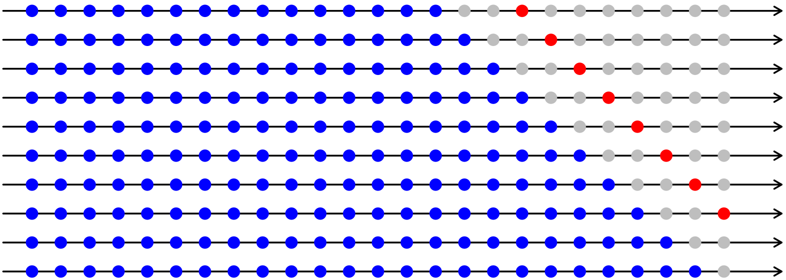


$h = 2$

Experimental set up

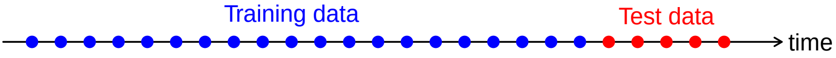


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

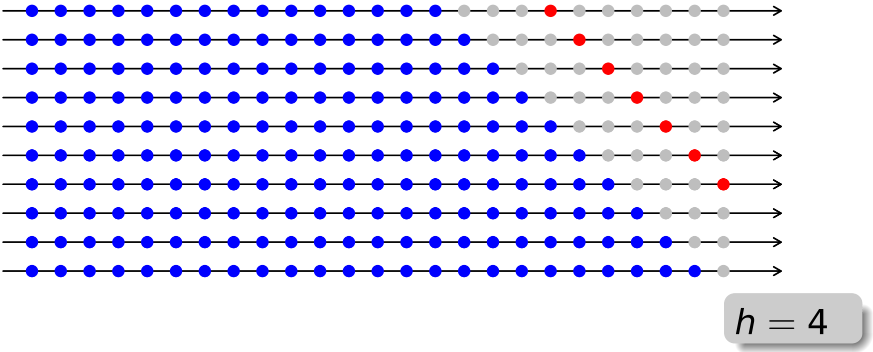


$h = 3$

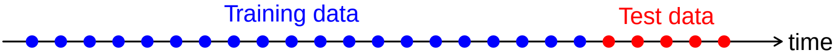
Experimental set up



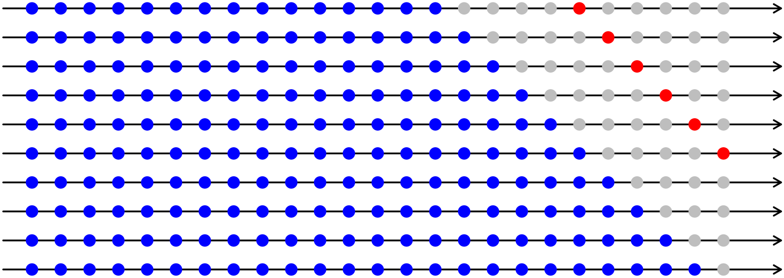
Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead



Experimental set up

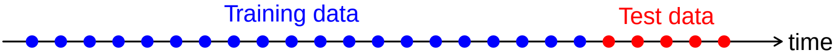


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

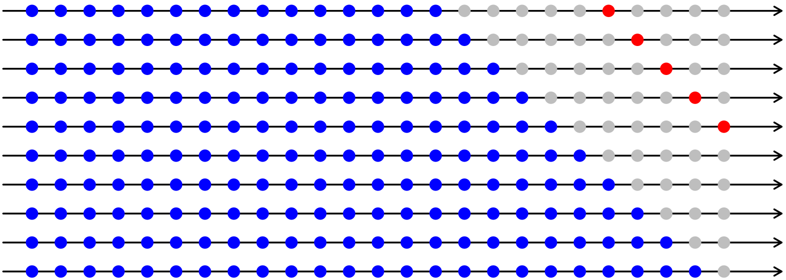


$h = 5$

Experimental set up

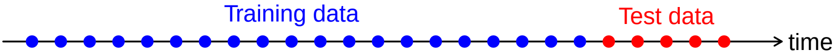


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

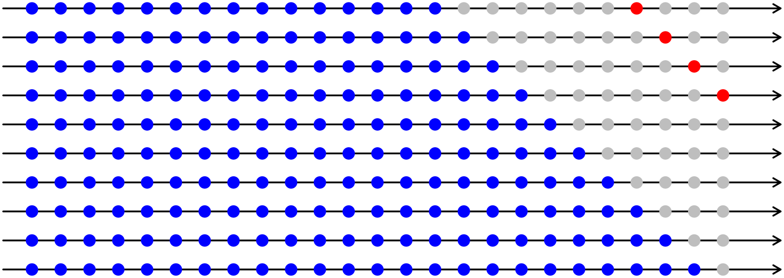


$h = 6$

Experimental set up

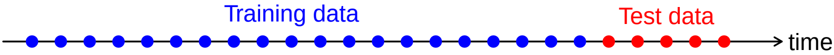


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

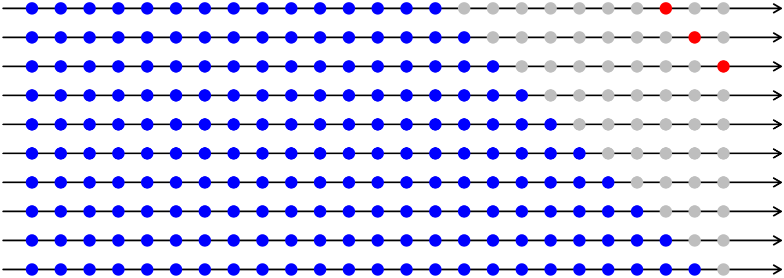


$h = 7$

Experimental set up

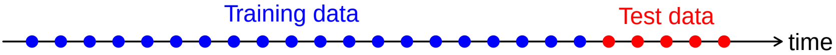


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

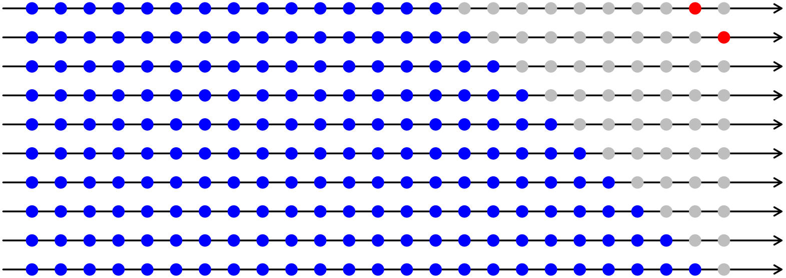


$h = 8$

Experimental set up

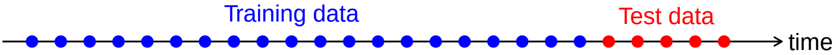


Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead

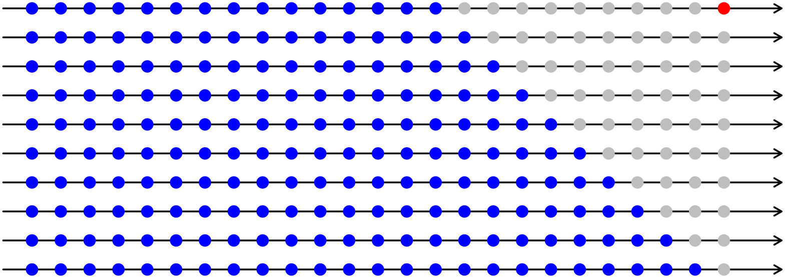


$h = 9$

Experimental set up



Rolling forecast origin: 1969–2008, forecasting up to 20 years ahead



$h = 10$

Out-of-sample MSE

Forecast horizon	Groups	**Model 1** PCFPC(6,0) (All common)	**Model 2** PCFPC(6,6) (Cointegrated)	**Model 3** PCFPC(6,6) (Stationary)	**Model 4** PCFPC(0,6) (Divergent)
$h = 5$	Combined (F & M)	2.59	2.60	**2.50**	2.52
	Female (F)	2.81	2.75	2.70	**2.63**
	Male (M)	2.38	2.45	**2.29**	2.42
$h = 10$	Combined (F & M)	**4.57**	4.66	4.60	4.65
	Female(F)	4.67	4.43	4.63	**4.23**
	Male (M)	**4.48**	4.89	4.57	5.06
$h = 15$	Combined (F & M)	**7.72**	8.00	7.84	8.15
	Female (F)	7.31	6.64	7.23	**6.47**
	Male(M)	**8.14**	9.36	8.44	9.82
$h = 20$	Combined (F & M)	**12.97**	13.56	13.35	14.10
	Female (F)	12.26	10.41	12.08	**10.35**
	Male (M)	**13.69**	16.70	14.63	17.86

Common functional PC

- The best coherent model has all principal components and scores in common. So the models differ only in mean.
- The independent models work better for female data – due to the hump in male mortality being captured in common components?
- PCFPC model more general, so poor performance a problem of model selection.
- PCFPC used $K = L = 6$. May be too many? How to do order selection?
- Maybe PCFPC (cointegrated) would be better if we had a good automated VECM procedure.

Common functional PC

- The best coherent model has all principal components and scores in common. So the models differ only in mean.
- The independent models work better for female data – due to the hump in male mortality being captured in common components?
- PCFPC model more general, so poor performance a problem of model selection.
- PCFPC used $K = L = 6$. May be too many? How to do order selection?
- Maybe PCFPC (cointegrated) would be better if we had a good automated VECM procedure.

Common functional PC

- The best coherent model has all principal components and scores in common. So the models differ only in mean.
- The independent models work better for female data – due to the hump in male mortality being captured in common components?
- PCFPC model more general, so poor performance a problem of model selection.
- PCFPC used $K = L = 6$. May be too many? How to do order selection?
- Maybe PCFPC (cointegrated) would be better if we had a good automated VECM procedure.

Common functional PC

- The best coherent model has all principal components and scores in common. So the models differ only in mean.
- The independent models work better for female data – due to the hump in male mortality being captured in common components?
- PCFPC model more general, so poor performance a problem of model selection.
- PCFPC used $K = L = 6$. May be too many? How to do order selection?
- Maybe PCFPC (cointegrated) would be better if we had a good automated VECM procedure.

Common functional PC

- The best coherent model has all principal components and scores in common. So the models differ only in mean.
- The independent models work better for female data – due to the hump in male mortality being captured in common components?
- PCFPC model more general, so poor performance a problem of model selection.
- PCFPC used $K = L = 6$. May be too many? How to do order selection?
- Maybe PCFPC (cointegrated) would be better if we had a good automated VECM procedure.

Outline

1 Functional time series

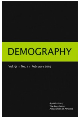
2 Functional time series models

3 Common functional principal components

4 Australian mortality

5 References

Selected references



Hyndman, Booth, Yasmeen (2013). "Coherent mortality forecasting: the product-ratio method with functional time series models". *Demography* **50**(1), 261–283.

Hyndman (2014). *demography: Forecasting mortality, fertility, migration and population data*. cran.r-project.org/package=demography