Common functional principal component models for mortality forecasting

Rob J Hyndman and Farah Yasmeen
1 Functional time series

2 Functional time series models

3 Common functional principal components

4 Australian mortality

5 References
Common functional principal component models for mortality
Australia: male death rates (1921–2009)
Australia: female death rates (1921–2009)

- Smooth data using weighted penalized regression splines with a partial monotonic constraint.

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

How to forecast future curves?
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, ...).
Common functional principal component models for mortality forecasting

Australian male mortality
Functional time series model

Main effects

Interaction

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:
  \[ \lim_{t \to \infty} \| f_{t,j} - f_{t,i} \| < \infty \text{ for all } i \text{ and } j \]
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:

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

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

\[ \lim_{t \to \infty} E \| f_{t,j} - f_{t,i} \| < \infty \text{ for all } i \text{ and } j \]
Common functional principal components
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 \(\ell\) so that
  \[
  \lim_{t \to \infty} E \|f_{t,j} - f_{t,i}\| < \infty \quad \text{for all } i \text{ and } j.
  \]
- 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 \(\ell\) so that
  \[
  \lim_{t \to \infty} E\|f_{t,j} - f_{t,i}\| < \infty \quad \text{for all } i \text{ and } j.
  \]

- 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 \(\ell\) so that
  \[
  \lim_{t \to \infty} \mathbb{E} \| f_{t,j} - f_{t,i} \| < \infty \quad \text{for all } i \text{ and } j.
  \]

- Can impose coherence by requiring either cointegrated scores, or stationary scores.
Partial Common Functional Principal Components

**PCFPC**(\(K, L\)) model

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

- **Model 1**: PCFPC\((K, 0)\). No idiosyncratic principal components in the model.
- **Model 2**: PCFPC\((K, L)\) with a coherence constraint. For each \(\ell\), \(\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}\) is stationary for all \(i, j\).
- **Model 3**: PCFPC\((K, L)\) with a coherence constraint. For each \(\ell\) 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 \(\ell\), \(\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}\) is stationary for all \(i,j\).
- **Model 3: PCFPC\((K,L)\)** with a coherence constraint. For each \(\ell\) and \(j\), \(\{\gamma_{t,\ell,j}\}\) is stationary.
- **Model 4: PCFPC\((0,L)\)**. All principal components and scores are idiosyncratic.
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**: \(\text{PCFPC}(K, 0)\). No idiosyncratic principal components in the model.
- **Model 2**: \(\text{PCFPC}(K, L)\) with a coherence constraint. For each \(\ell\), \(\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}\) is stationary for all \(i, j\).
- **Model 3**: \(\text{PCFPC}(K, L)\) with a coherence constraint. For each \(\ell\) and \(j\), \(\{\gamma_{t,\ell,j}\}\) is stationary.
- **Model 4**: \(\text{PCFPC}(0, L)\). 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 $\ell$, $\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}$ is stationary for all $i, j$.
- **Model 3**: **PCFPC**($K, L$) with a coherence constraint. For each $\ell$ 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 \(\ell\), \(\{\gamma_{t,i,\ell} - \gamma_{t,j,\ell}\}\) is stationary for all \(i, j\).
- **Model 3**: PCFPC\((K, L)\) with a coherence constraint. For each \(\ell\) and \(j\), \(\{\gamma_{t,\ell,j}\}\) is stationary.
- **Model 4**: PCFPC\((0, L)\). All principal components and scores are idiosyncratic.
1 Functional time series
2 Functional time series models
3 Common functional principal components
4 **Australian mortality**
5 References
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.
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.
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.
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.
Life expectancy forecasts

Year

Age

1960 1980 2000 2020 2040
60 65 70 75 80 85 90

(b): Life expectancy difference: F−M

Year

Number of years

1960 2000 2040
0 2 4 6 8

Female−PCFPC(6, 6)
Male−PCFPC(6, 6)
Female−Independent
Male−Independent

Common functional principal component models for mortality

Australian mortality
Experimental set up

Training data

Test data

$\rightarrow$ time
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

Common functional principal component models for mortality forecasting

Australian mortality
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

$h = 4$
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

$h = 6$
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

Common functional principal component models for mortality forecasting

Australian mortality
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

<table>
<thead>
<tr>
<th>Forecast horizon</th>
<th>Groups</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>PCFPC(6,0)</td>
<td>PCFPC(6,6)</td>
<td>PCFPC(6,6)</td>
<td>PCFPC(0,6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(All common)</td>
<td>(Cointegrated)</td>
<td>(Stationary)</td>
<td>(Divergent)</td>
</tr>
<tr>
<td>$h = 5$</td>
<td>Combined (F &amp; M)</td>
<td>2.59</td>
<td>2.60</td>
<td>2.50</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td>Female (F)</td>
<td>2.81</td>
<td>2.75</td>
<td>2.70</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>Male (M)</td>
<td>2.38</td>
<td>2.45</td>
<td><strong>2.29</strong></td>
<td>2.42</td>
</tr>
<tr>
<td>$h = 10$</td>
<td>Combined (F &amp; M)</td>
<td><strong>4.57</strong></td>
<td>4.66</td>
<td>4.60</td>
<td>4.65</td>
</tr>
<tr>
<td></td>
<td>Female (F)</td>
<td>4.67</td>
<td>4.43</td>
<td>4.63</td>
<td><strong>4.23</strong></td>
</tr>
<tr>
<td></td>
<td>Male (M)</td>
<td><strong>4.48</strong></td>
<td>4.89</td>
<td>4.57</td>
<td>5.06</td>
</tr>
<tr>
<td>$h = 15$</td>
<td>Combined (F &amp; M)</td>
<td><strong>7.72</strong></td>
<td>8.00</td>
<td>7.84</td>
<td>8.15</td>
</tr>
<tr>
<td></td>
<td>Female (F)</td>
<td>7.31</td>
<td>6.64</td>
<td>7.23</td>
<td><strong>6.47</strong></td>
</tr>
<tr>
<td></td>
<td>Male (M)</td>
<td><strong>8.14</strong></td>
<td>9.36</td>
<td>8.44</td>
<td>9.82</td>
</tr>
<tr>
<td>$h = 20$</td>
<td>Combined (F &amp; M)</td>
<td><strong>12.97</strong></td>
<td>13.56</td>
<td>13.35</td>
<td>14.10</td>
</tr>
<tr>
<td></td>
<td>Female (F)</td>
<td>12.26</td>
<td>10.41</td>
<td>12.08</td>
<td><strong>10.35</strong></td>
</tr>
<tr>
<td></td>
<td>Male (M)</td>
<td><strong>13.69</strong></td>
<td>16.70</td>
<td>14.63</td>
<td>17.86</td>
</tr>
</tbody>
</table>
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.
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.
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.
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.
1 Functional time series
2 Functional time series models
3 Common functional principal components
4 Australian mortality
5 References

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