MONASH University

Rob J Hyndman

Functional time series

with applications in demography

1. Tools for functional time series analysis

Mortality rates

Fertility rates

Outline

1 **Functional time series**

2 Functional principal components

3 Data visualization

4 References

Functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log[z_t(x_i)] & \text{if } \lambda = 0; \\ \lambda^{-1}\left[z_t^\lambda(x_i) - 1\right] & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $z_t(x_i)$ is observed data for age x_i in year t, $i = 1, \dots, N, \quad t = 1, \dots, T$.
- λ chosen so that $\varepsilon_{t,i} \sim \text{NID}(0, 1)$.
- We assume $s_t(x)$ is a smooth function of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_N$.
- We want to forecast **whole curve** $z_t(x)$ for $t = T + 1, \dots, T + h$.

Functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log[z_t(x_i)] & \text{if } \lambda = 0; \\ \lambda^{-1}\left[z_t^\lambda(x_i) - 1\right] & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $z_t(x_i)$ is observed data for age x_i in year t, $i = 1, \dots, N, \quad t = 1, \dots, T$.
- λ chosen so that $\varepsilon_{t,i} \sim \text{NID}(0, 1)$.
- We assume $s_t(x)$ is a smooth function of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_N$.
- We want to forecast **whole curve** $z_t(x)$ for $t = T+1, \dots, T+h$.

Functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log[z_t(x_i)] & \text{if } \lambda = 0; \\ \lambda^{-1}\left[z_t^\lambda(x_i) - 1\right] & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $z_t(x_i)$ is observed data for age x_i in year t, $i = 1, \dots, N, \quad t = 1, \dots, T$.
- λ chosen so that $\varepsilon_{t,i} \sim \text{NID}(0, 1)$.
- We assume $s_t(x)$ is a smooth function of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_N$.
- We want to forecast **whole curve** $z_t(x)$ for $t = T + 1, \dots, T + h$.

Functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log[z_t(x_i)] & \text{if } \lambda = 0; \\ \lambda^{-1}\left[z_t^\lambda(x_i) - 1\right] & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $z_t(x_i)$ is observed data for age x_i in year t, $i = 1, \dots, N, \quad t = 1, \dots, T$.
- λ chosen so that $\varepsilon_{t,i} \sim \text{NID}(0, 1)$.
- We assume $s_t(x)$ is a smooth function of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_N$.
- We want to forecast **whole curve** $z_t(x)$ for $t = T+1, \dots, T+h$.

Functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log[z_t(x_i)] & \text{if } \lambda = 0; \\ \lambda^{-1}\left[z_t^\lambda(x_i) - 1\right] & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- $z_t(x_i)$ is observed data for age x_i in year t, $i = 1, \dots, N, \quad t = 1, \dots, T$.
- λ chosen so that $\varepsilon_{t,i} \sim \text{NID}(0, 1)$.
- We assume $s_t(x)$ is a smooth function of x.
- We need to estimate $s_t(x)$ from the data for $x_1 < x < x_N$.
- We want to forecast **whole curve** $z_t(x)$ for $t = T + 1, \dots, T + h$.

Smoothing functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log(z_t(x_i)) & \text{if } \lambda = 0; \\ \lambda^{-1}\left(z_t^\lambda(x_i) - 1\right) & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- Estimate $s_t(x)$ using penalized regression spline with a large number of knots.
- For mortality data, use $\lambda = 0$ and constrain $s_t(x)$ to be monotonic for $x > 50$.
- For fertility data, use $\lambda = 0.4$ and constrain $s_t(x)$ to be concave.
- Fit is weighted with $w_t(x_i) = \sigma_t^{-2}(x_i)$.

Smoothing functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log(z_t(x_i)) & \text{if } \lambda = 0; \\ \lambda^{-1}\left(z_t^\lambda(x_i) - 1\right) & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- Estimate $s_t(x)$ using penalized regression spline with a large number of knots.
- For mortality data, use $\lambda = 0$ and constrain $s_t(x)$ to be monotonic for $x > 50$.
- For fertility data, use $\lambda = 0.4$ and constrain $s_t(x)$ to be concave.
- Fit is weighted with $w_t(x_i) = \sigma_t^{-2}(x_i)$.

Smoothing functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log(z_t(x_i)) & \text{if } \lambda = 0; \\ \lambda^{-1}\left(z_t^\lambda(x_i) - 1\right) & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- Estimate $s_t(x)$ using penalized regression spline with a large number of knots.
- For mortality data, use $\lambda = 0$ and constrain $s_t(x)$ to be monotonic for $x > 50$.
- For fertility data, use $\lambda = 0.4$ and constrain $s_t(x)$ to be concave.
- Fit is weighted with $w_t(x_i) = \sigma_t^{-2}(x_i)$.

Smoothing functional time series

$$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} \log(z_t(x_i)) & \text{if } \lambda = 0; \\ \lambda^{-1}\left(z_t^\lambda(x_i) - 1\right) & \text{otherwise.} \end{cases}$$
$$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$$

- Estimate $s_t(x)$ using penalized regression spline with a large number of knots.
- For mortality data, use $\lambda = 0$ and constrain $s_t(x)$ to be monotonic for $x > 50$.
- For fertility data, use $\lambda = 0.4$ and constrain $s_t(x)$ to be concave.
- Fit is weighted with $w_t(x_i) = \sigma_t^{-2}(x_i)$.

Smoothing functional time series

Mortality

$D_t(x_i)$ = number of deaths at age x_i in year t.

$E_t(x_i)$ = total population aged x_i on June 30 in year t.

$m_t(x_i) = D_t(x_i)/E_t(x_i)$ = observed mortality rate.

$\mu_t(x_i)$ = "true" mortality rate.

$$D_t(x_i) \sim \text{Poisson}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[m_t(x_i)] = \mu_t(x)$ and $\text{V}[m_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Smoothing functional time series

Mortality

$D_t(x_i) =$ number of deaths at age x_i in year t.

$E_t(x_i) =$ total population aged x_i on June 30 in year t.

$m_t(x_i) = D_t(x_i)/E_t(x_i) =$ observed mortality rate.

$\mu_t(x_i) =$ "true" mortality rate.

$$D_t(x_i) \sim \text{Poisson}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[m_t(x_i)] = \mu_t(x)$ and $\text{V}[m_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Smoothing functional time series

Mortality

$D_t(x_i) =$ number of deaths at age x_i in year t.

$E_t(x_i) =$ total population aged x_i on June 30 in year t.

$m_t(x_i) = D_t(x_i)/E_t(x_i) =$ observed mortality rate.

$\mu_t(x_i) =$ “true” mortality rate.

$$D_t(x_i) \sim \text{Poisson}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[m_t(x_i)] = \mu_t(x)$ and $\text{V}[m_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Smoothing functional time series

Mortality

$D_t(x_i) =$ number of deaths at age x_i in year t.

$E_t(x_i) =$ total population aged x_i on June 30 in year t.

$m_t(x_i) = D_t(x_i)/E_t(x_i) =$ observed mortality rate.

$\mu_t(x_i) =$ "true" mortality rate.

$$D_t(x_i) \sim \text{Poisson}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[m_t(x_i)] = \mu_t(x)$ and $\text{V}[m_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Taylor series approx

$$\text{V}[g_\lambda(X)] = \sigma_X^2[g'_\lambda(\mu_X)]^2$$
$$\text{V}[\log(X)] = \sigma_X^2/\mu_X^2$$

Smoothing functional time series

Mortality

$D_t(x_i)$ = number of deaths at age x_i in year t.

$E_t(x_i)$ = total population aged x_i on June 30 in year t.

$m_t(x_i) = D_t(x_i)/E_t(x_i)$ = observed mortality rate.

$\mu_t(x_i)$ = "true" mortality rate.

$$D_t(x_i) \sim \text{Poisson}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[m_t(x_i)] = \mu_t(x)$ and $\text{V}[m_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

$$\begin{aligned}\sigma^2(x_i) &= \text{V}(\log[m_t(x_i)]) \\ &\approx \left[\mu_t(x_i)E_t^{-1}(x_i)\right]\mu_t(x)^{-2} \\ &= \mu_t(x_i)^{-1}E_t(x_i)^{-1}\end{aligned}$$

Taylor series approx

$$\text{V}[g_\lambda(X)] = \sigma_X^2[g'_\lambda(\mu_X)]^2$$
$$\text{V}[\log(X)] = \sigma_X^2/\mu_X^2$$

Smoothing functional time series

Fertility

$B_t(x_i)$ = number of births to women aged x_i in year t.

$E_t(x_i)$ = total population aged x_i on June 30 in year t.

$f_t(x_i) = B_t(x_i)/E_t(x_i)$ = observed fertility rate.

$\mu_t(x_i)$ = "true" fertility rate.

$$B_t(x_i) \sim \text{Pn}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[f_t(x_i)] = \mu_t(x)$ and $\text{V}[f_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Smoothing functional time series

Fertility

$B_t(x_i)$ = number of births to women aged x_i in year t.

$E_t(x_i)$ = total population aged x_i on June 30 in year t.

$f_t(x_i) = B_t(x_i)/E_t(x_i)$ = observed fertility rate.

$\mu_t(x_i)$ = "true" fertility rate.

$$B_t(x_i) \sim \text{Pn}(E_t(x_i)\mu_t(x_i))$$

$\text{E}[f_t(x_i)] = \mu_t(x)$ and $\text{V}[f_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Fertility

$B_t(x_i)$ = number of births to women aged x_i in year t.
$E_t(x_i)$ = total population aged x_i on June 30 in year t.
$f_t(x_i) = B_t(x_i)/E_t(x_i)$ = observed fertility rate.
$\mu_t(x_i)$ = "true" fertility rate.

$$B_t(x_i) \sim \mathrm{Pn}(E_t(x_i)\mu_t(x_i))$$

$\mathrm{E}[f_t(x_i)] = \mu_t(x)$ and $\mathrm{V}[f_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Smoothing functional time series

Fertility

$B_t(x_i)$ = number of births to women aged x_i in year t.
$E_t(x_i)$ = total population aged x_i on June 30 in year t.
$f_t(x_i) = B_t(x_i)/E_t(x_i)$ = observed fertility rate.
$\mu_t(x_i)$ = "true" fertility rate.

$$B_t(x_i) \sim \mathrm{Pn}(E_t(x_i)\mu_t(x_i))$$

$\mathrm{E}[f_t(x_i)] = \mu_t(x)$ and $\mathrm{V}[f_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

Taylor series approx

$$\begin{aligned}\mathrm{V}[g_\lambda(X)] &= \sigma_X^2[g'_\lambda(\mu_X)]^2 \\ &= \sigma_X^2\mu_X^{2\lambda-2}\end{aligned}$$

Smoothing functional time series

Fertility

$B_t(x_i)$ = number of births to women aged x_i in year t.

$E_t(x_i)$ = total population aged x_i on June 30 in year t.

$f_t(x_i) = B_t(x_i)/E_t(x_i)$ = observed fertility rate.

$\mu_t(x_i)$ = "true" fertility rate.

$$B_t(x_i) \sim \text{Pn}(E_t(x_i)\mu_t(x_i))$$

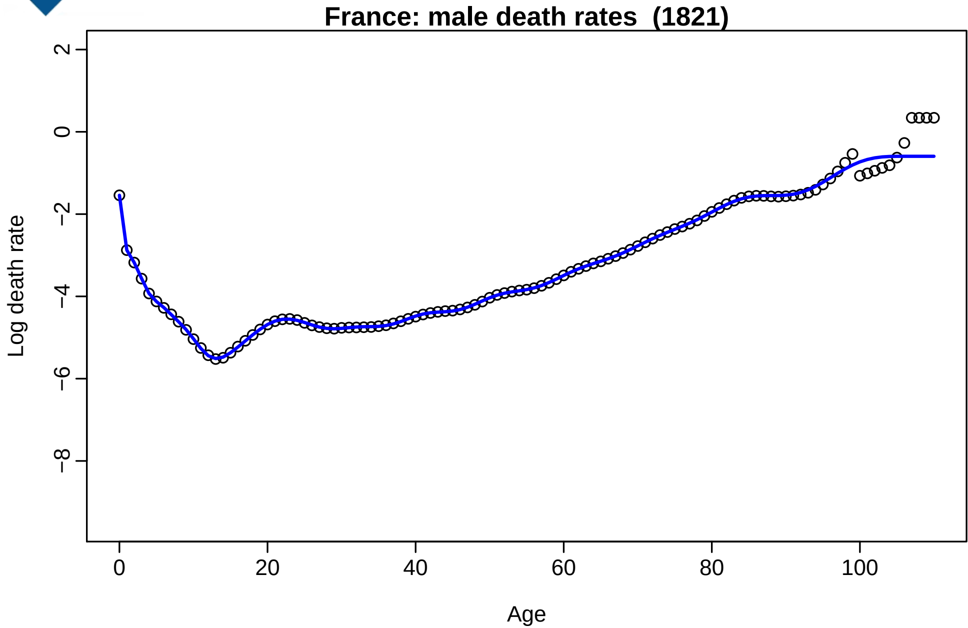
$\text{E}[f_t(x_i)] = \mu_t(x)$ and $\text{V}[f_t(x_i)] = \mu_t(x_i)E_t^{-1}(x_i)$.

$$\begin{aligned}
\sigma^2(x_i) &= \text{V}(g_\lambda[f_t(x_i)]) \\
&= \left[\mu_t(x_i)E_t^{-1}(x_i)\right]\mu_t(x)^{2\lambda-2} \\
&= \mu_t(x_i)^{2\lambda-1}E_t(x_i)^{-1} \\
&\approx f_t(x_i)^{2\lambda-1}E_t(x_i)^{-1}
\end{aligned}$$

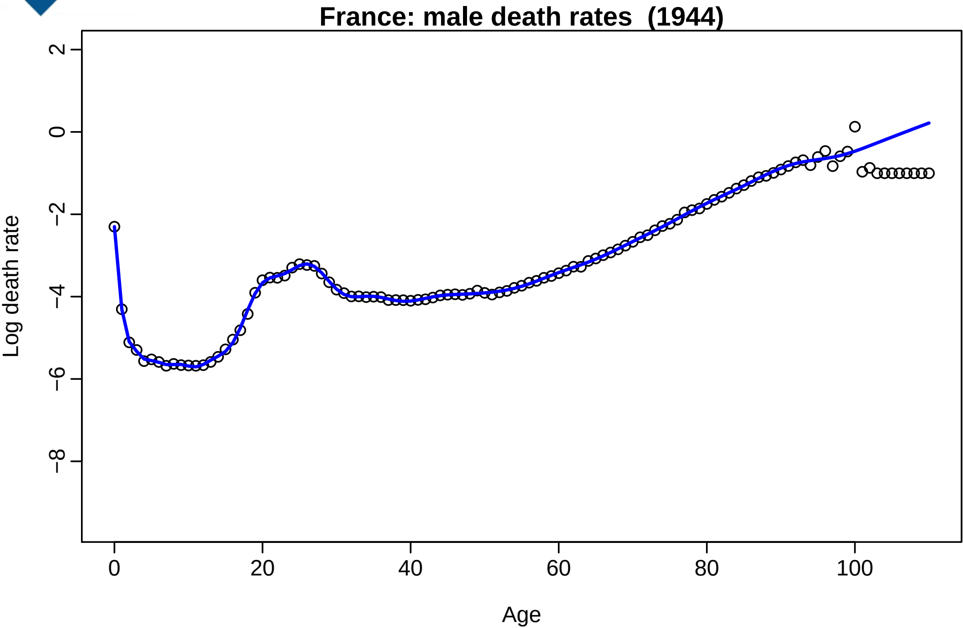
Taylor series approx

$$\begin{aligned}
\text{V}[g_\lambda(X)] &= \sigma_X^2[g'_\lambda(\mu_X)]^2 \\
&= \sigma_X^2\mu_X^{2\lambda-2}
\end{aligned}$$

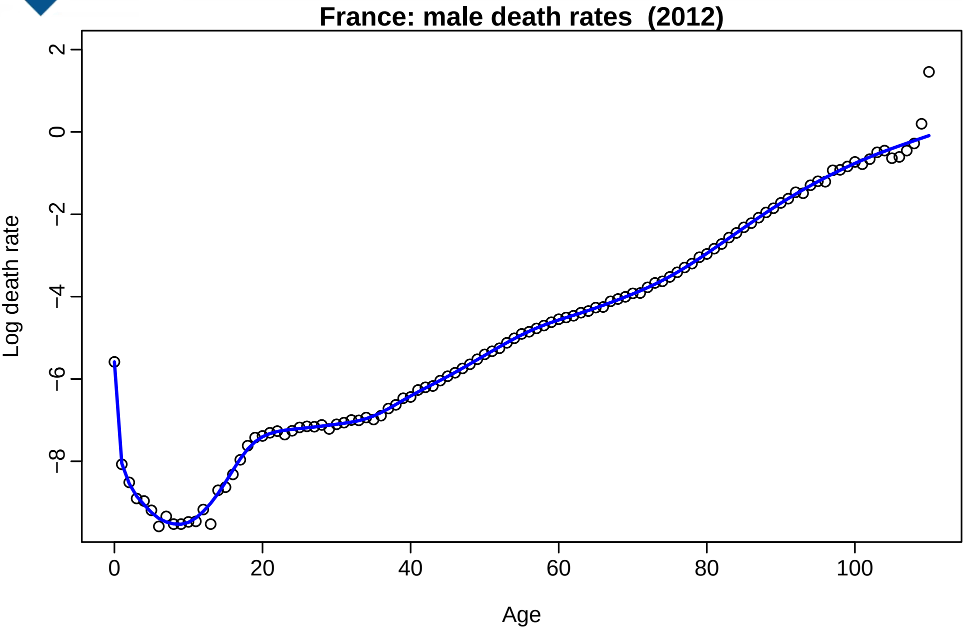
Smoothing functional time series



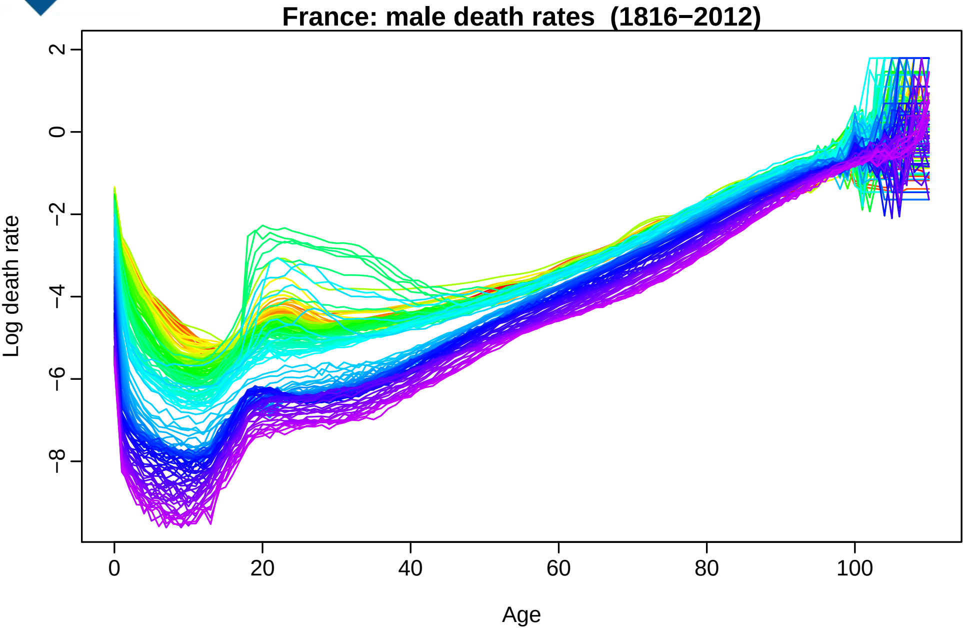
Smoothing functional time series



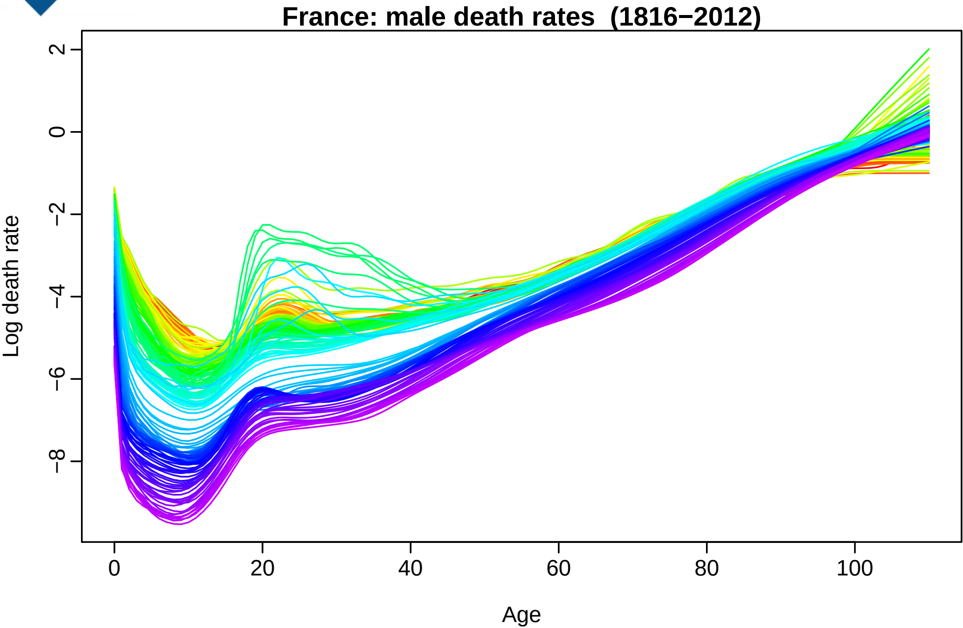
Smoothing functional time series



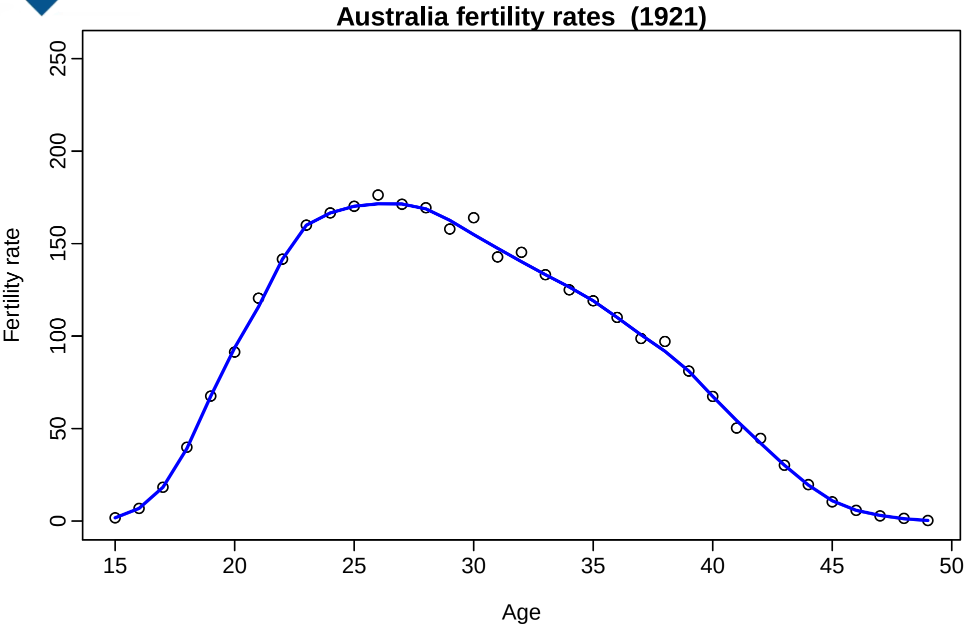
Smoothing functional time series



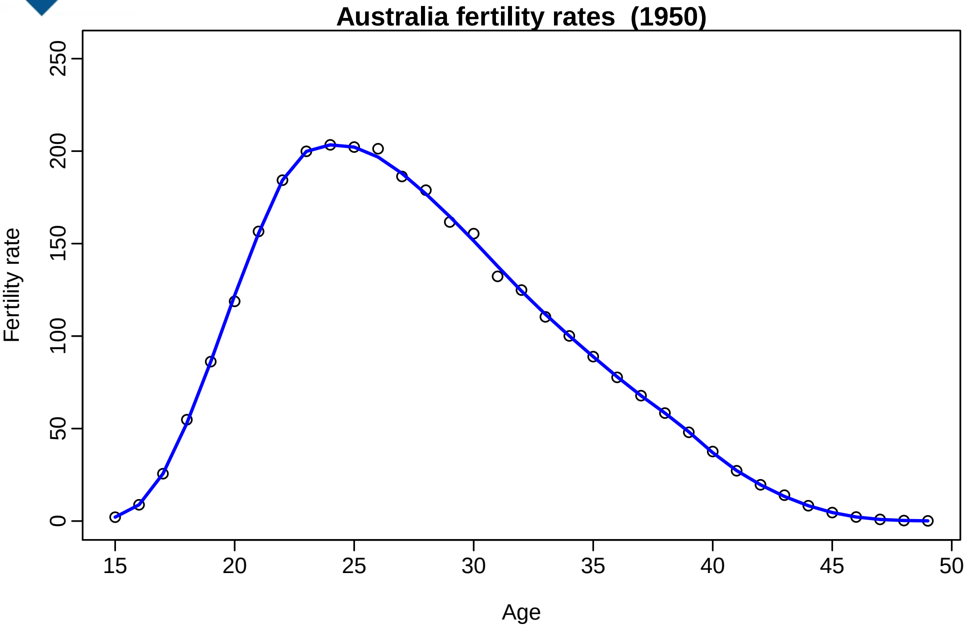
Smoothing functional time series



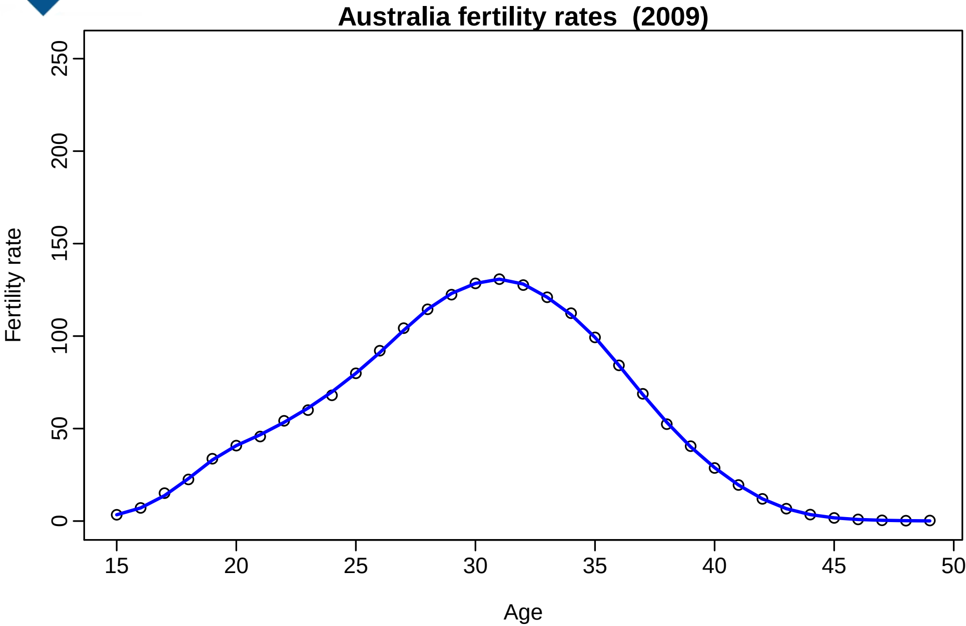
Smoothing functional time series



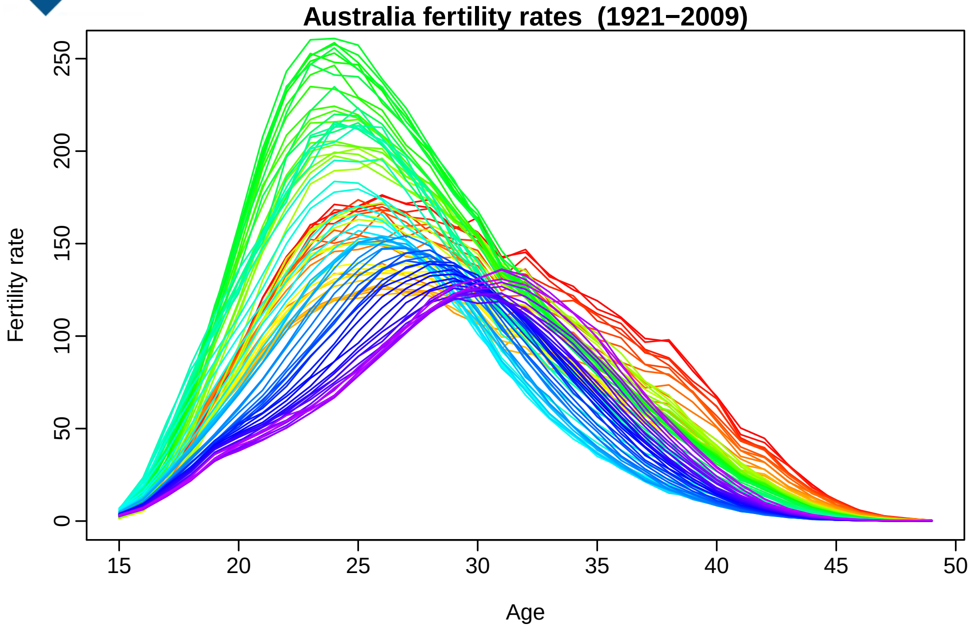
Smoothing functional time series



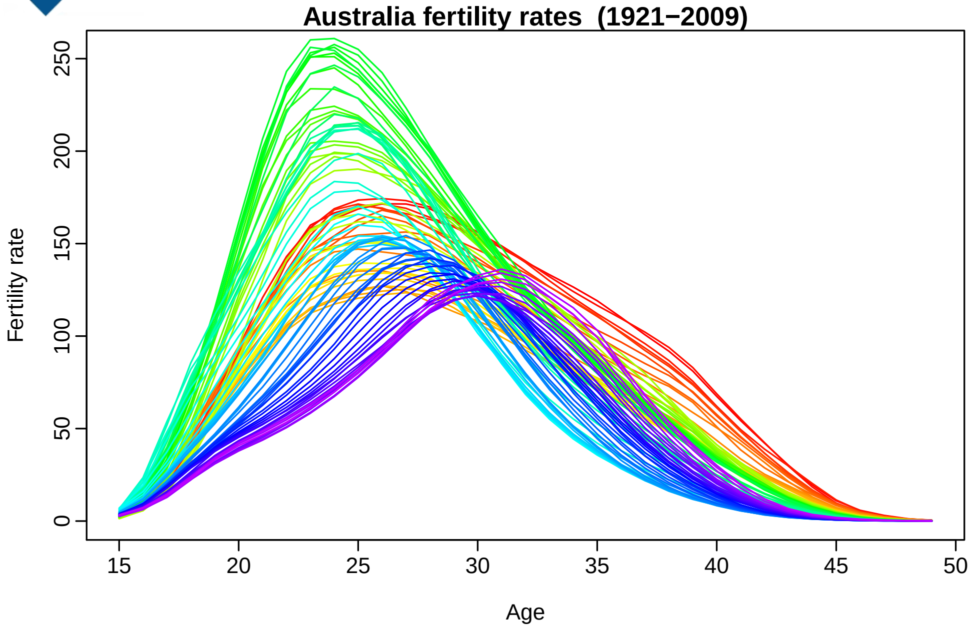
Smoothing functional time series



Smoothing functional time series



Smoothing functional time series



Outline

1 Functional time series

2 Functional principal components

3 Data visualization

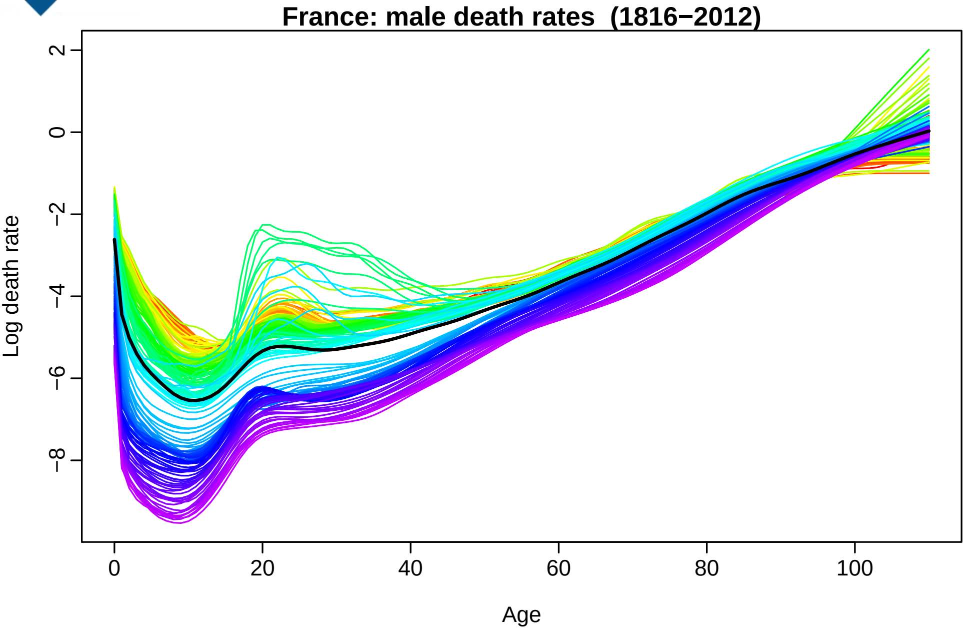
4 References

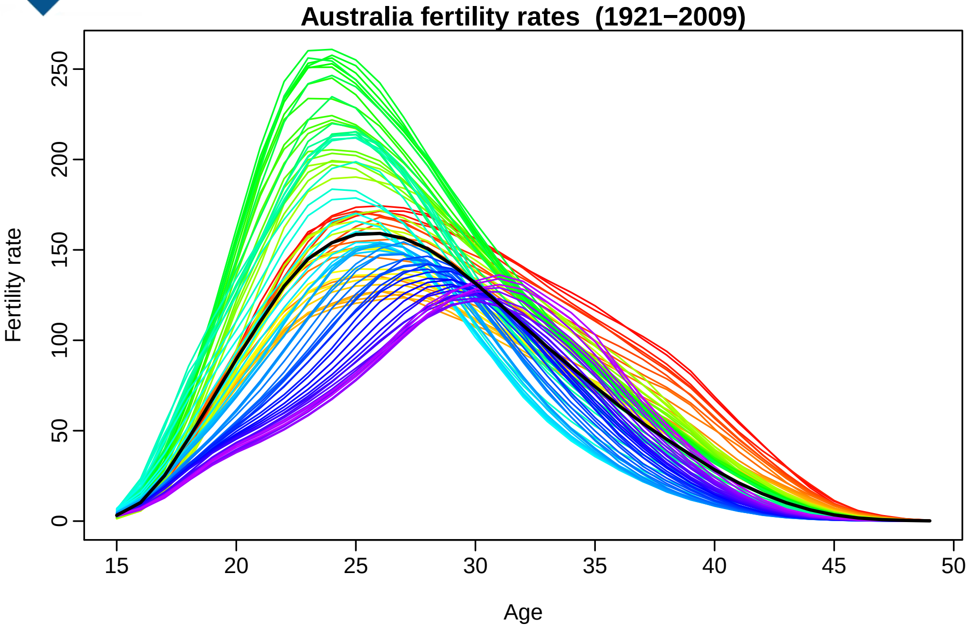
Functional principal components

$$y_t(x_i) = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i},$$
$$s_t(x) = \mu(x) + \sum_{k=1}^{T-1} \beta_{t,k}\,\phi_k(x)$$

1. Estimate smooth functions $s_t(x)$ using weighted penalized regression splines.
2. Compute $\mu(x)$ as $\bar{s}(x)$ across years.
3. Compute $\beta_{t,k}$ and $\phi_k(x)$ using functional principal components.

$\mu(x)$


France: male death rates (1816−2012)
Log death rate
2
0
−2
−4
−6
−8
0
20
40
60
80
100
Age


Australia fertility rates (1921−2009)
Fertility rate
0
50
100
150
200
250
15
20
25
30
35
40
45
50
Age

Functional principal components

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_k(x)$.
- The PC scores for each year are given by

$$\beta_{k,t} = \int \phi_k(x) \left[\hat{s}_t(x) - \bar{s}(x)\right] dx$$

- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $\beta_{1,t}$ subject to the constraint $\int \phi_1^2(x)dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_2^2(x)dx = 1$ and $\int \phi_1(x)\phi_2(x)dx = 0$.
  3. Find the weight function $\phi_3(x)$ that ...

Functional principal components

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_k(x)$.
- The PC scores for each year are given by

$$\beta_{k,t} = \int \phi_k(x)\left[\hat{s}_t(x) - \bar{s}(x)\right] dx$$

- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $\beta_{1,t}$ subject to the constraint $\int \phi_1^2(x)dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_i^2(x)dx = 1$ and $\int \phi_1(x)\phi_2(x)dx = 0$.
  3. Find the weight function $\phi_3(x)$ that

Functional principal components

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_k(x)$.
- The PC scores for each year are given by

$$\beta_{k,t} = \int \phi_k(x) \left[\hat{s}_t(x) - \bar{s}(x)\right] dx$$

- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $\beta_{1,t}$ subject to the constraint $\int \phi_i^2(x) dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_i^2(x) dx = 1$ and $\int \phi_1(x)\phi_2(x) dx = 0$.
  3. Find the weight function $\phi_3(x)$ that ...

Functional principal components

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_k(x)$.
- The PC scores for each year are given by
$$\beta_{k,t} = \int \phi_k(x) \left[\hat{s}_t(x) - \bar{s}(x)\right] dx$$
- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $\beta_{1,t}$ subject to the constraint $\int \phi_i^2(x)dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_i^2(x)dx = 1$ and $\int \phi_1(x)\phi_2(x)dx = 0$.
  3. Find the weight function $\phi_3(x)$ that ...

Functional principal components

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_k(x)$.
- The PC scores for each year are given by
$$\beta_{k,t} = \int \phi_k(x)\left[\hat{s}_t(x) - \bar{s}(x)\right] dx$$
- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $\beta_{1,t}$ subject to the constraint $\int \phi_i^2(x)dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_i^2(x)dx = 1$ and $\int \phi_1(x)\phi_2(x)dx = 0$.
  3. Find the weight function $\phi_3(x)$ that ...

Functional principal components

(Ramsay and Silverman, 1997,2002).

- In FDA, each principal component is specified by a weight function $\phi_k(x)$.
- The PC scores for each year are given by

$$\beta_{k,t} = \int \phi_k(x) \left[\hat{s}_t(x) - \bar{s}(x)\right] dx$$

- The aim is to:
  1. Find the weight function $\phi_1(x)$ that maximizes the variance of $\beta_{1,t}$ subject to the constraint $\int \phi_i^2(x) dx = 1$.
  2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_i^2(x) dx = 1$ and $\int \phi_1(x)\phi_2(x) dx = 0$.
  3. Find the weight function $\phi_3(x)$ that ...

Functional principal components

The optimal basis functions

Approximate $s_t(x)$ using

$$s_t(x) = \bar{s}(x) + \sum_{k=0}^{K} \beta_{t,k}\phi_k(x) + r_t(x)$$

The basis function $\phi_k(x)$ which minimizes $\text{MISE} = \frac{1}{T}\sum_{t=1}^{T}\int r_t^2 dx$ is the kth principal component (computed recursively).

Functional principal components

The optimal basis functions

Approximate $s_t(x)$ using

$$s_t(x) = \bar{s}(x) + \sum_{k=0}^{K} \beta_{t,k}\phi_k(x) + r_t(x)$$

The basis function $\phi_k(x)$ which minimizes

$\text{MISE} = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} \int r_t^2 dx$ is the kth principal component

(computed recursively).

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Computationally equivalent approach

- Let $s_t^*(x) = s_t(x) - \bar{s}(x)$.
- Discretize $s_t^*(x)$ on a dense grid of q equally spaced points.
- Denote discretized $s_t^*(x)$ as $T \times q$ matrix $\boldsymbol{G}$.
- SVD of $\boldsymbol{G} = \Phi\Lambda\Psi'$ where $\phi_k(x)$ is kth column of Φ.
- $\beta_{t,k}$ is (t,k)th element of $\boldsymbol{G}\Phi$.
- The basis functions are orthogonal.
- This means the coefficients series are also uncorrelated with each other. i.e., $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{t,j}) = 0$ for $i \neq j$. However, $\text{Corr}(\hat{\beta}_{t,i}, \hat{\beta}_{s,j}) \neq 0$ in general for $t \neq s$ and $i \neq j$.

Functional principal components

Eigenvector approach

- Let $\boldsymbol{V} = (T-1)^{-1}\boldsymbol{G}'\boldsymbol{G}$ be $m \times m$ sample covariance matrix of $\boldsymbol{G}$.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of $\boldsymbol{V}$ where $K \leq T-1$. The (i,j)th element of Φ_K is $\phi_i(x_j^*)$.
- Robust versions possible using robust "covariance" estimation.

Functional principal components

Eigenvector approach

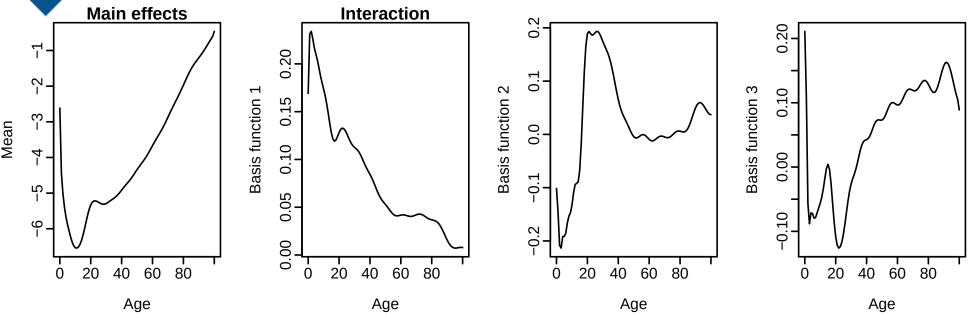
- Let $\boldsymbol{V} = (T-1)^{-1}\boldsymbol{G}'\boldsymbol{G}$ be $m \times m$ sample covariance matrix of $\boldsymbol{G}$.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of $\boldsymbol{V}$ where $K \leq T-1$. The (i,j)th element of Φ_K is $\phi_i(x_j^*)$.
- Robust versions possible using robust "covariance" estimation.

Functional principal components

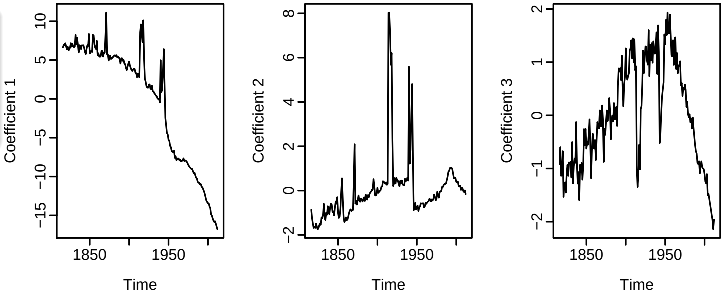
Eigenvector approach

- Let $\boldsymbol{V} = (T-1)^{-1}\boldsymbol{G}'\boldsymbol{G}$ be $m \times m$ sample covariance matrix of $\boldsymbol{G}$.
- Let $\Phi_K = [\phi_1, \dots, \phi_K]$ consist of the first K eigenvectors of $\boldsymbol{V}$ where $K \le T-1$. The (i,j)th element of Φ_K is $\phi_i(x_j^*)$.
- Robust versions possible using robust “covariance” estimation.

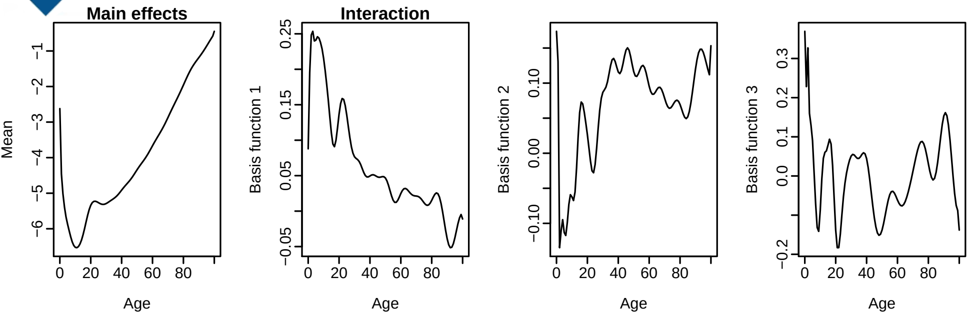
Functional principal components



French male mortality

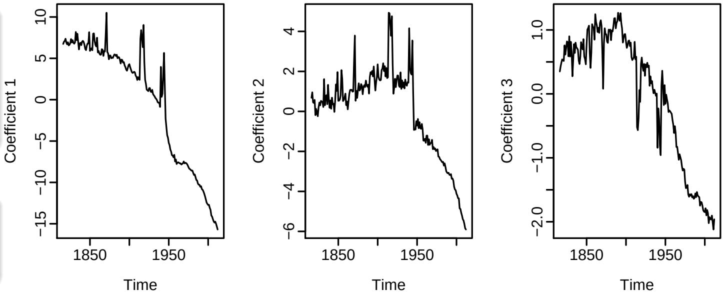


Functional principal components

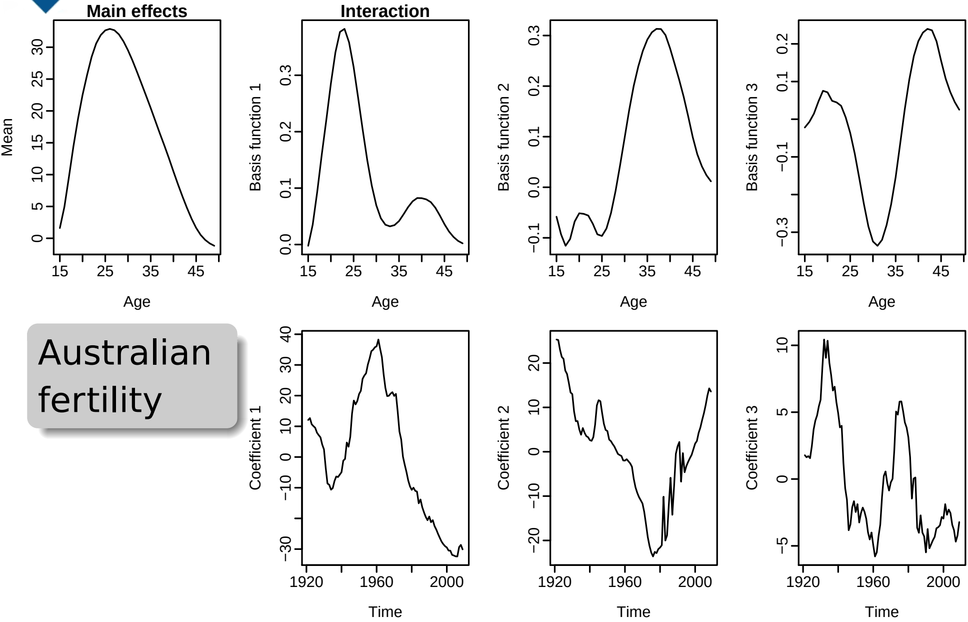


French male mortality

Robust PCA

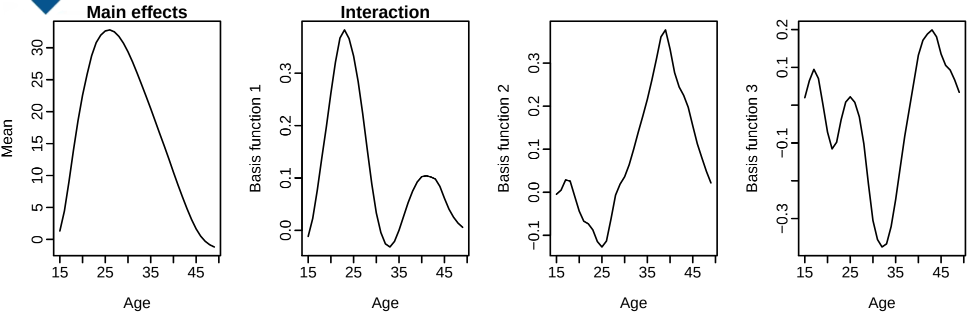


Functional principal components



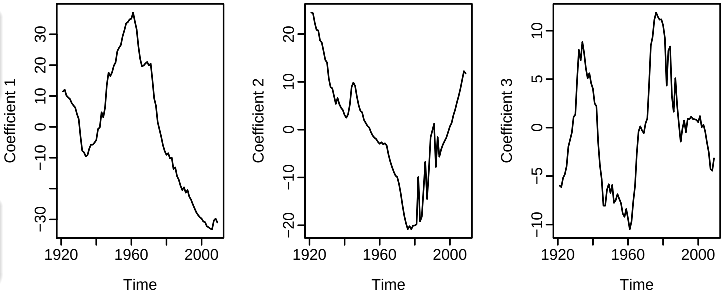
Australian fertility

Functional principal components



Australian fertility

Robust PCA



Outline

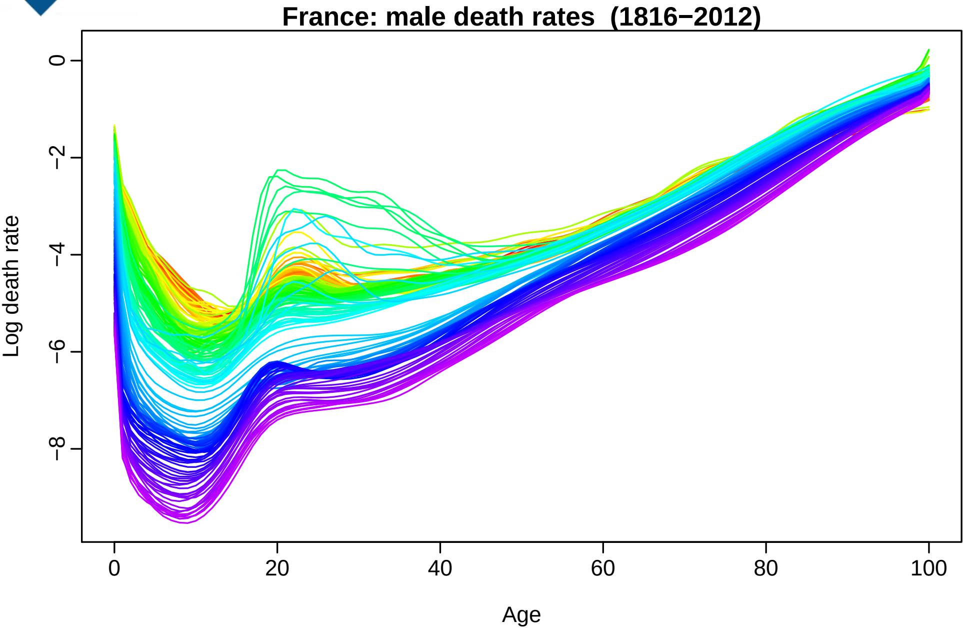
1 Functional time series

2 Functional principal components

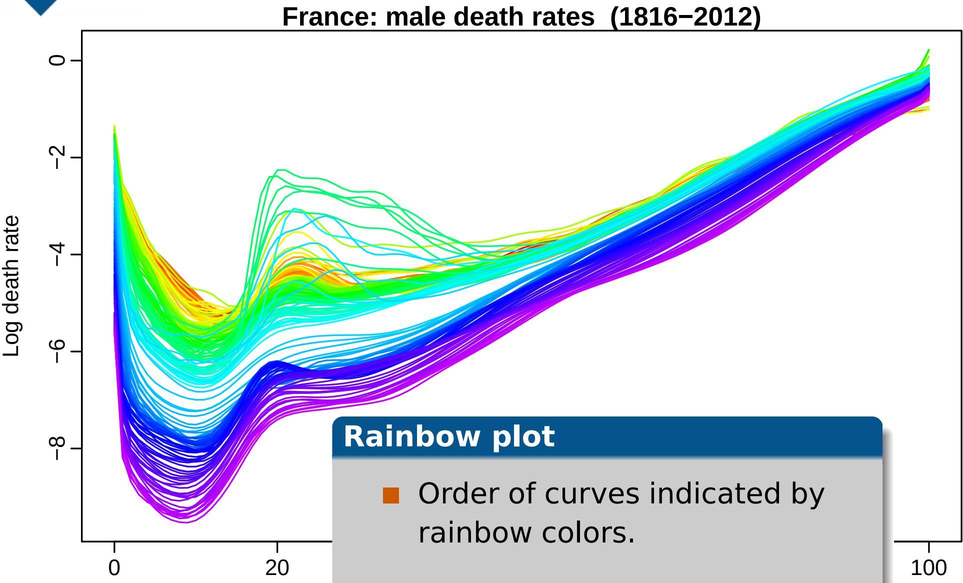
3 Data visualization

4 References

French male mortality rates



French male mortality rates



Rainbow plot

- Order of curves indicated by rainbow colors.
- Other orderings are possible.

Order by functional depth

Febrero, Galeano and Gonzalez-Manteiga (2007) proposed:

$$o_t = \int D(y_t(x))\,dx$$

where $D(y_t(x))$ is a univariate depth measure for each x.

- o_t provides an ordering of curves by “functional depth”.
- Problem: may not detect shape outliers.

Order by functional depth

Febrero, Galeano and Gonzalez-Manteiga (2007) proposed:

$$o_t = \int D(y_t(x))\, dx$$

where $D(y_t(x))$ is a univariate depth measure for each x.

- o_t provides an ordering of curves by “functional depth”.
- Problem: may not detect shape outliers.

Bivariate functional depth

Alternative: Apply bivariate depth measures to first two PC scores.

Plot $\beta_{t,2}$ vs $\beta_{t,1}$

- Each point in scatterplot represents one curve.
- Outliers show up in bivariate score space.
- Curves can be ordered by bivariate depth.

Bivariate functional depth

Alternative: Apply bivariate depth measures to first two PC scores.

Plot $\beta_{t,2}$ vs $\beta_{t,1}$

- Each point in scatterplot represents one curve.
- Outliers show up in bivariate score space.
- Curves can be ordered by bivariate depth.

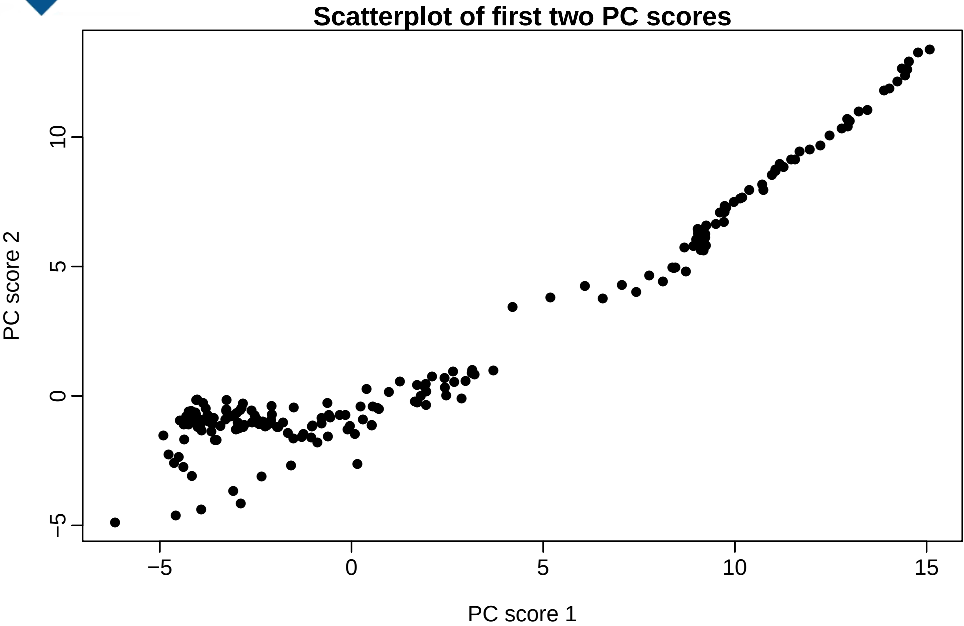
Bivariate functional depth

Alternative: Apply bivariate depth measures to first two PC scores.

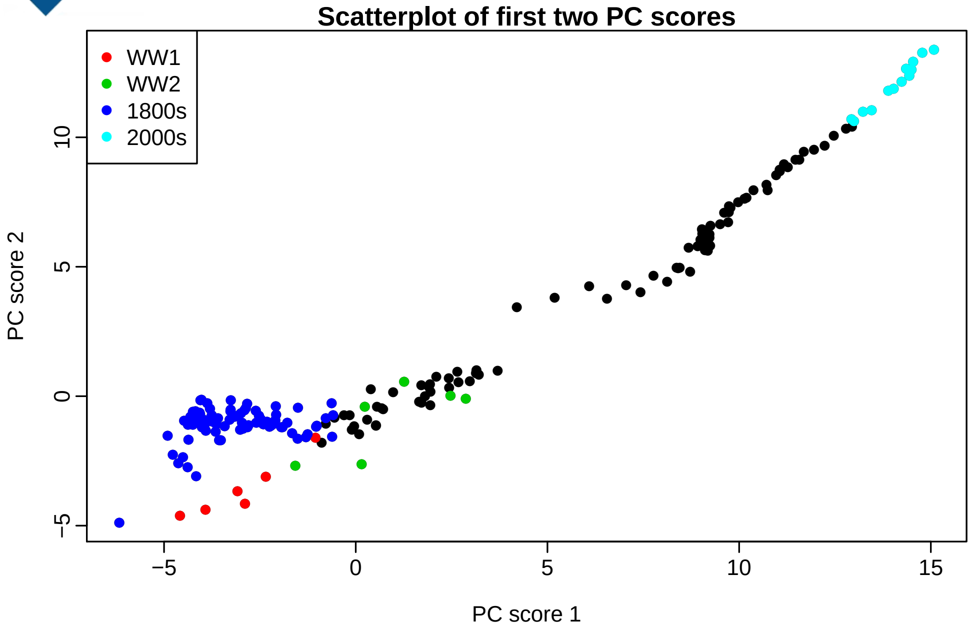
Plot $\beta_{t,2}$ vs $\beta_{t,1}$

- Each point in scatterplot represents one curve.
- Outliers show up in bivariate score space.
- Curves can be ordered by bivariate depth.

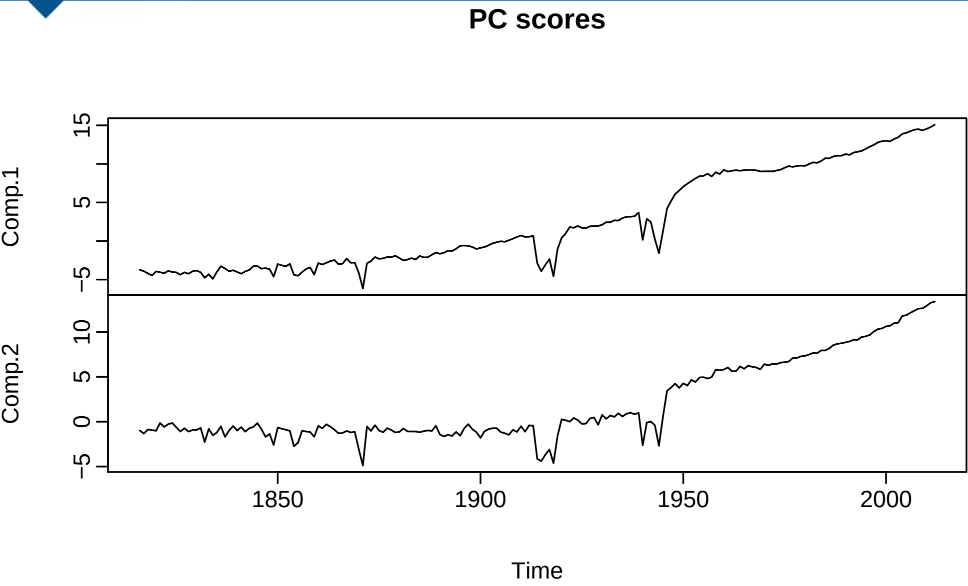
Robust PC scores



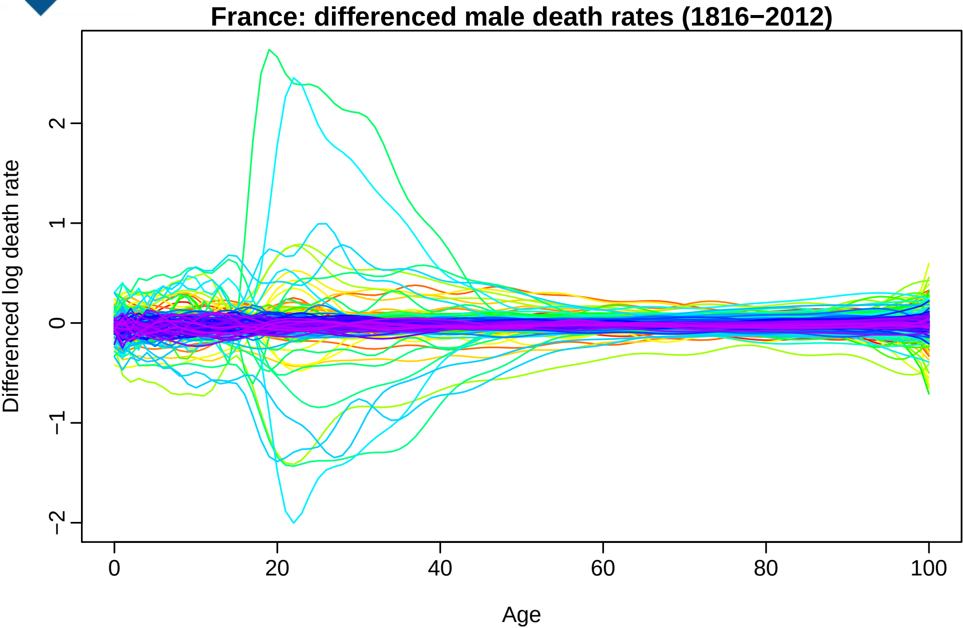
Robust PC scores



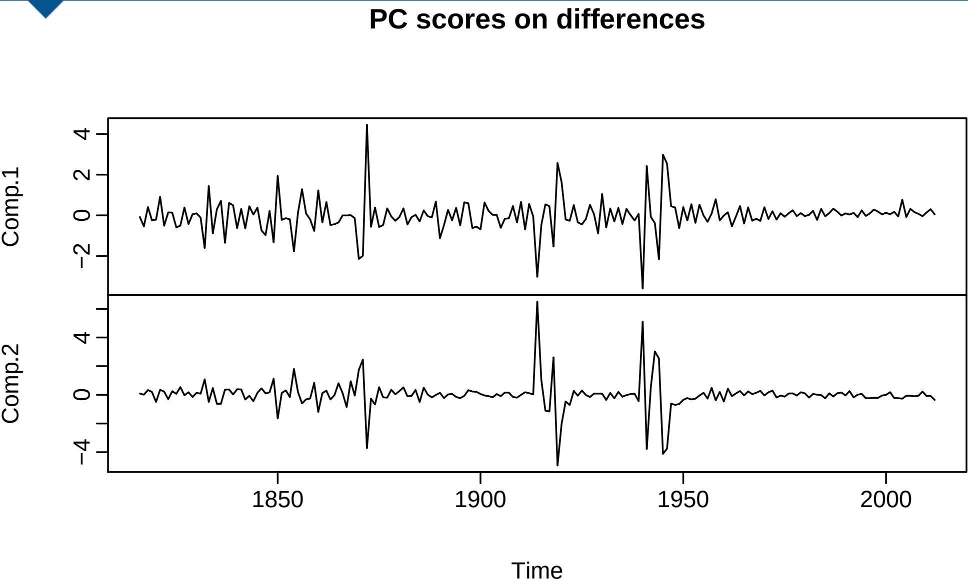
Robust PC scores



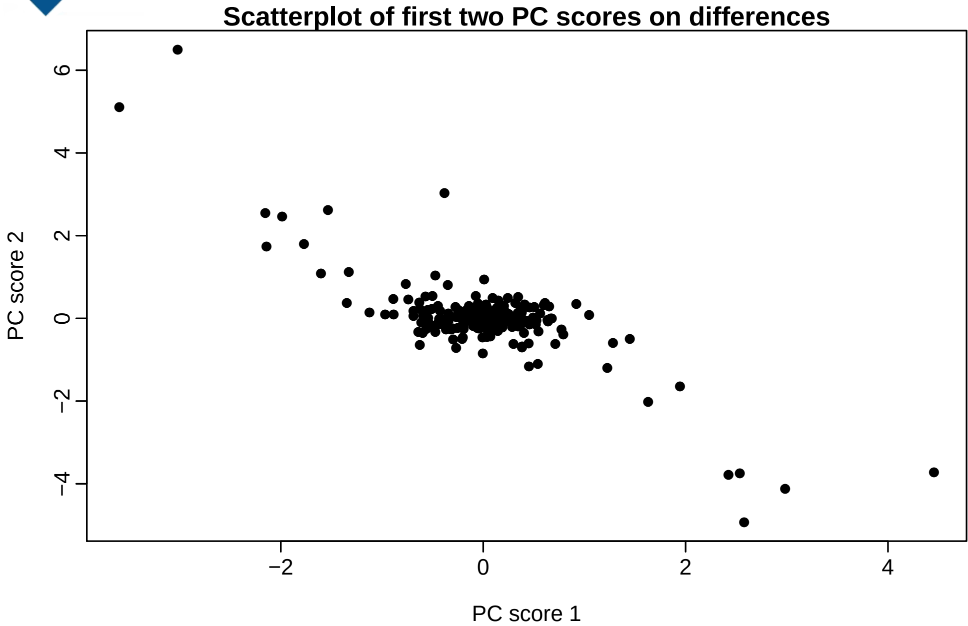
French male mortality rates



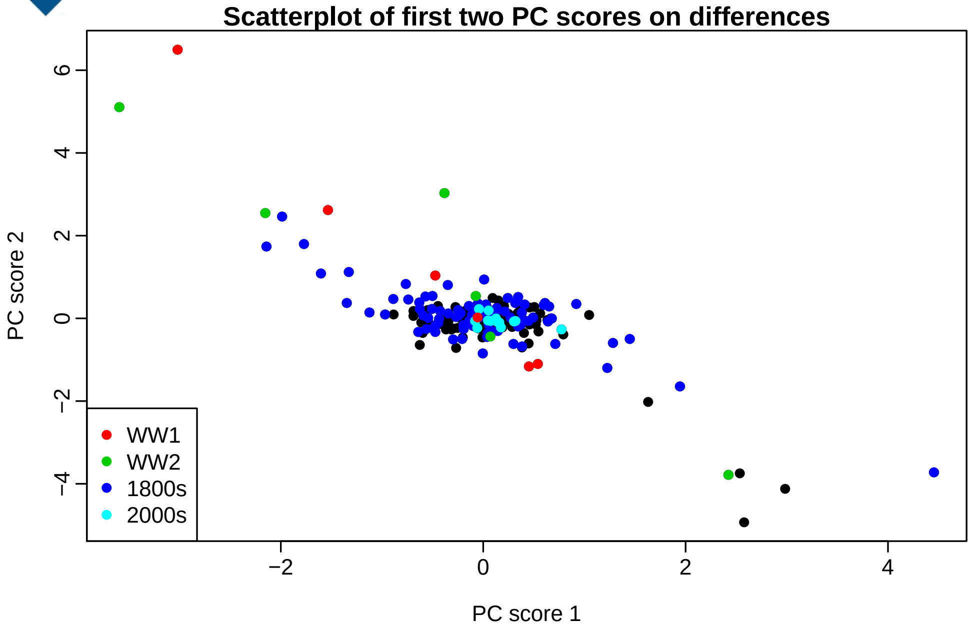
Robust PC scores



Robust PC scores



Robust PC scores



Halfspace location depth

The halfspace depth of a point q:

(Due to Hotelling, 1929; Tukey, 1975)

- For each closed halfspace that contains q, count number of observations not in halfspace. The minimum over all halfspaces is the depth of that point.
- The median is the point with maximum depth (not generally unique).
- Any point outside convex hull of the data has depth zero.

Halfspace location depth

The

(Du

ce.

h

h

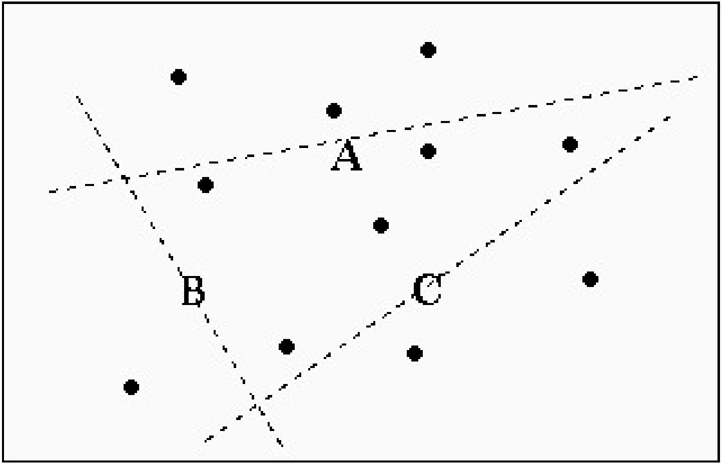


Fig.5 : Depth of certain points with respect to a data set.
depth(A) = 3 , depth(B) = 1 , depth(C) = 2.

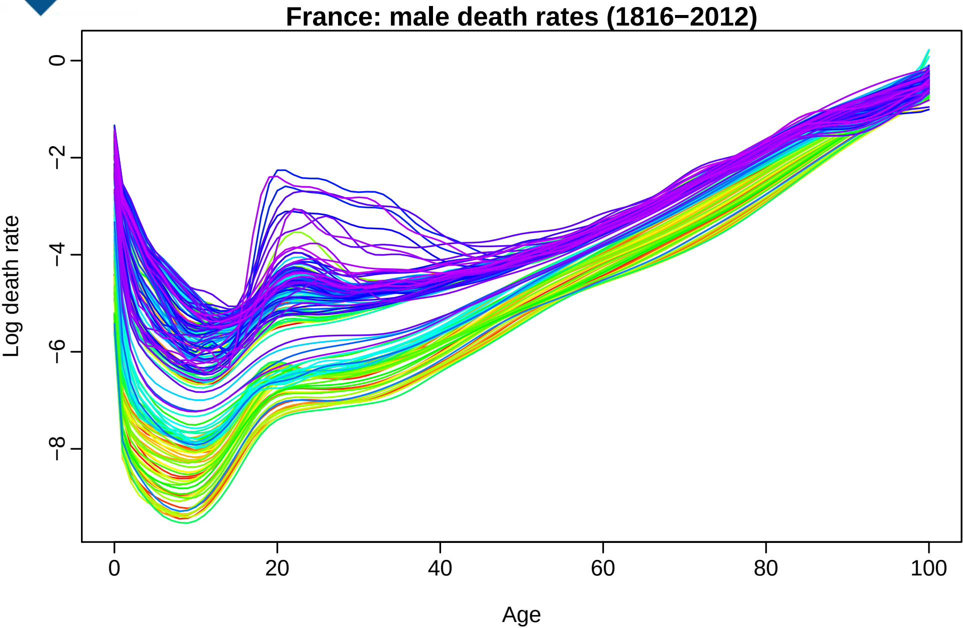
Halfspace location depth

The halfspace depth of a point q:

(Due to Hotelling, 1929; Tukey, 1975)

- For each closed halfspace that contains q, count number of observations not in halfspace. The minimum over all halfspaces is the depth of that point.
- The median is the point with maximum depth (not generally unique).
- Any point outside convex hull of the data has depth zero.

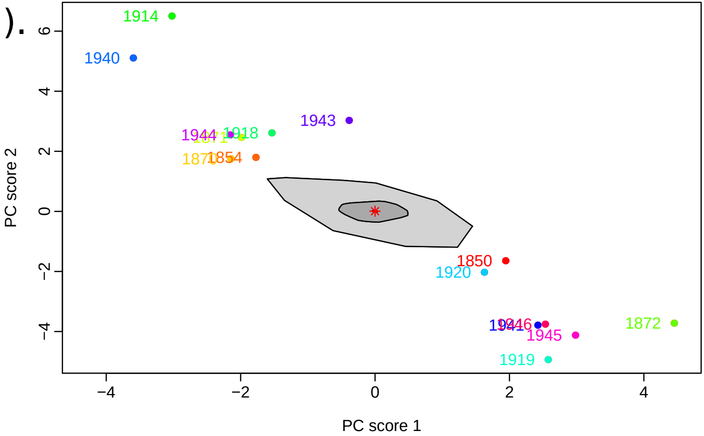
Ordering by halfspace depth



Bivariate bagplot

Due to Rousseeuw, Ruts & Tukey (Am.Stat. 1999).

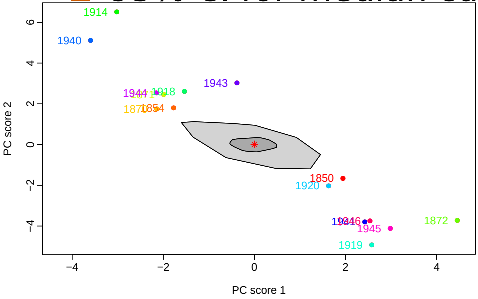
- Rank points by halfspace location depth.
- Display median, 50% convex hull and outer convex hull (with 99% coverage if bivariate normal).

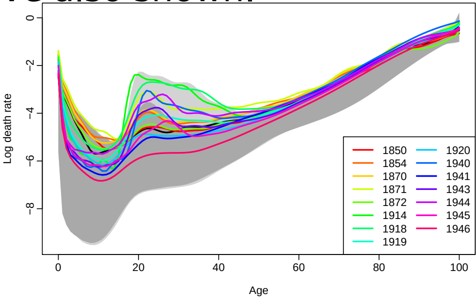


Bivariate bagplot

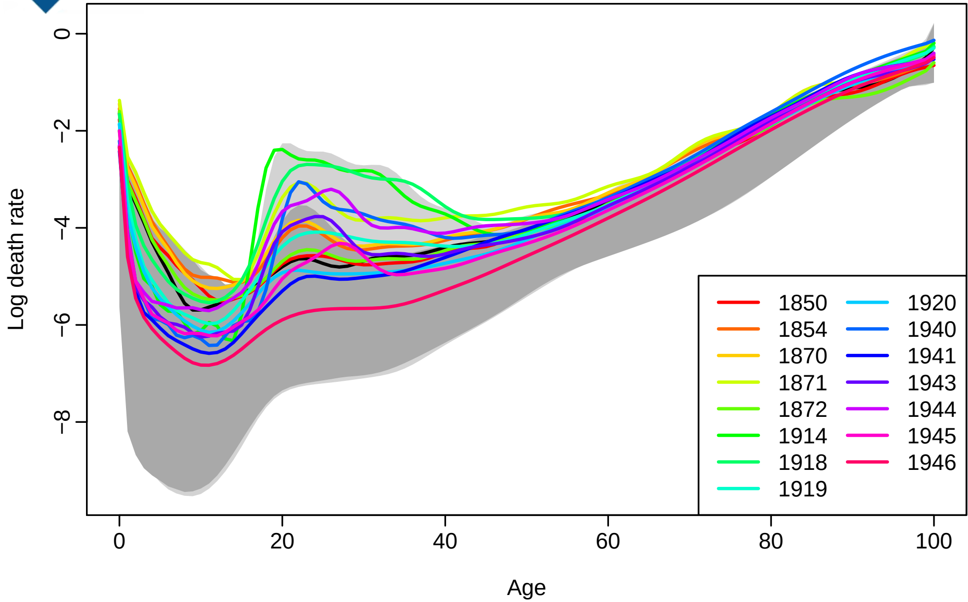
Due to Rousseeuw, Ruts & Tukey (Am.Stat. 1999).

- Rank points by halfspace location depth.
- Display median, 50% convex hull and outer convex hull (with 99% coverage if bivariate normal).
- Boundaries contain all curves inside bags.
- 95% CI for median curve also shown.

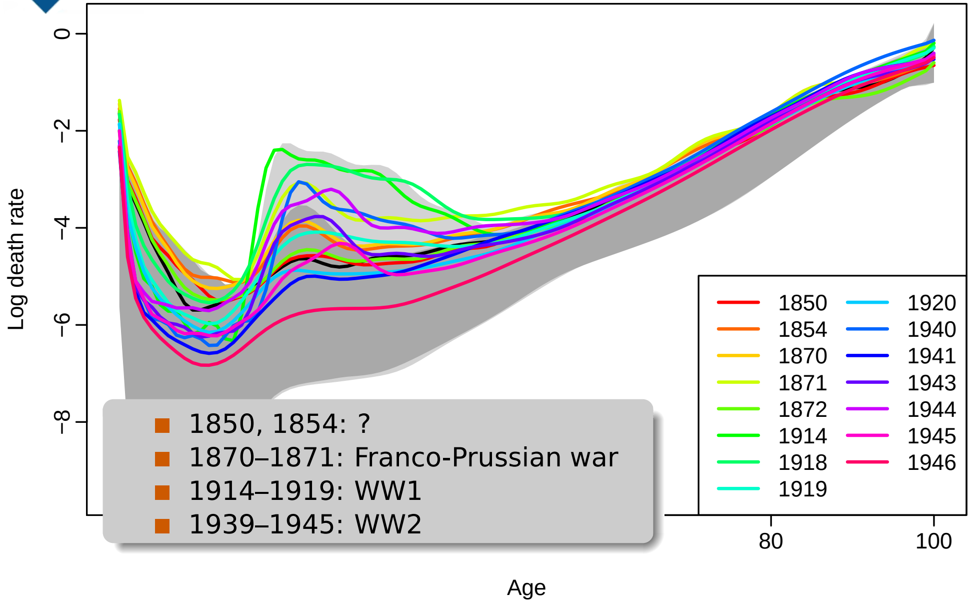




Functional bagplot

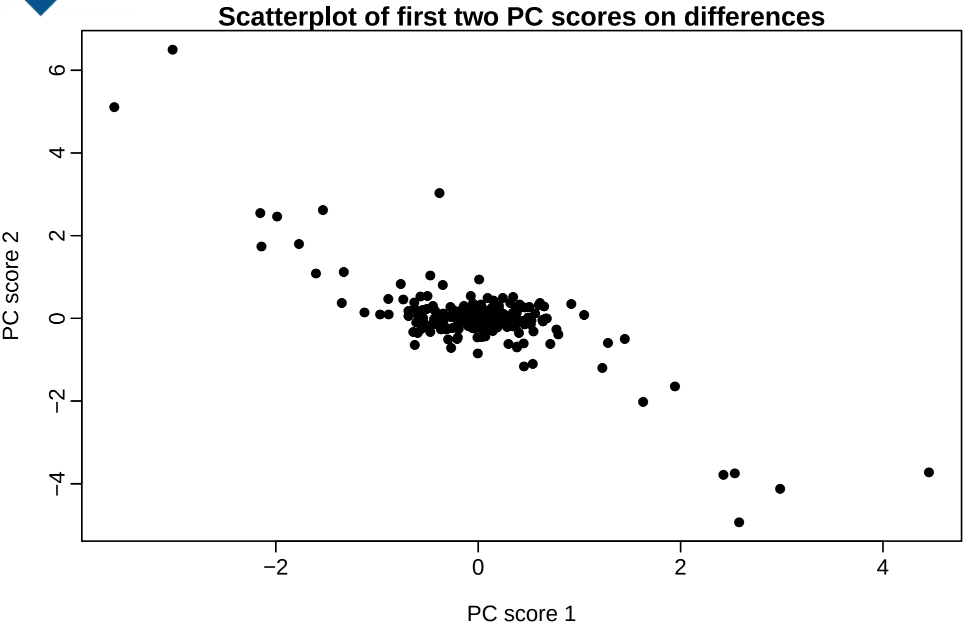


Functional bagplot

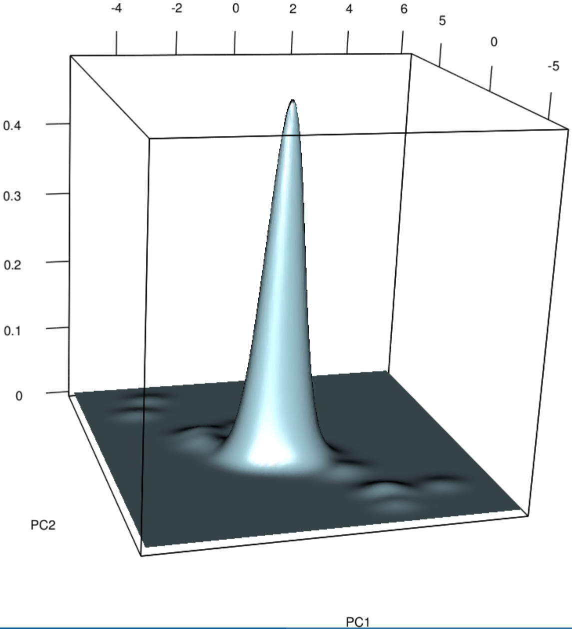


- 1850, 1854: ?
- 1870–1871: Franco-Prussian war
- 1914–1919: WW1
- 1939–1945: WW2

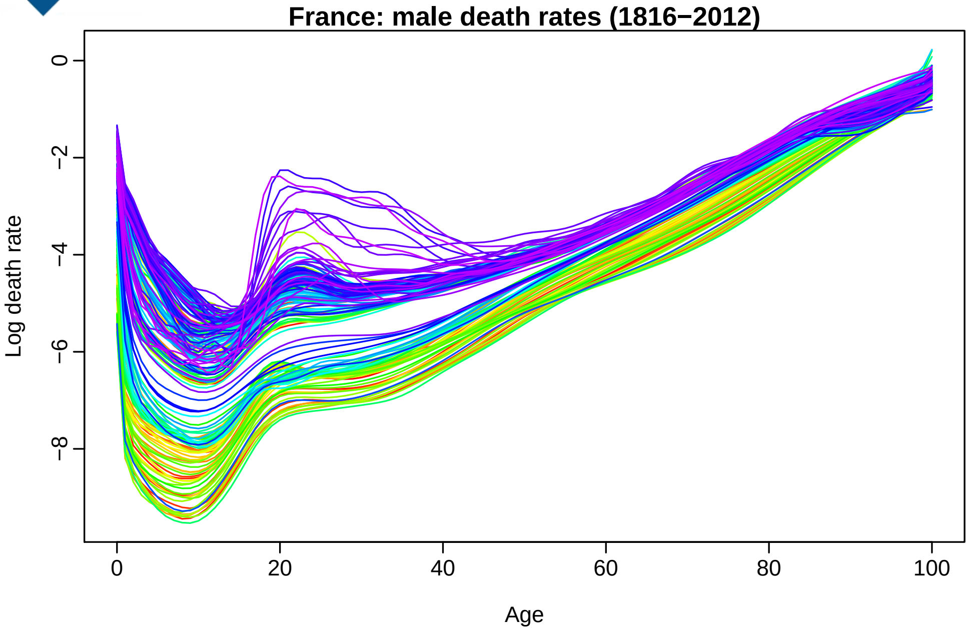
Kernel density estimate



Kernel density estimate

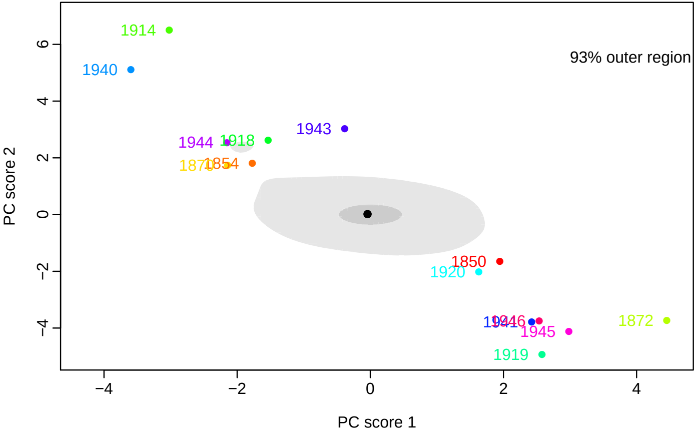


Ordering by bivariate density



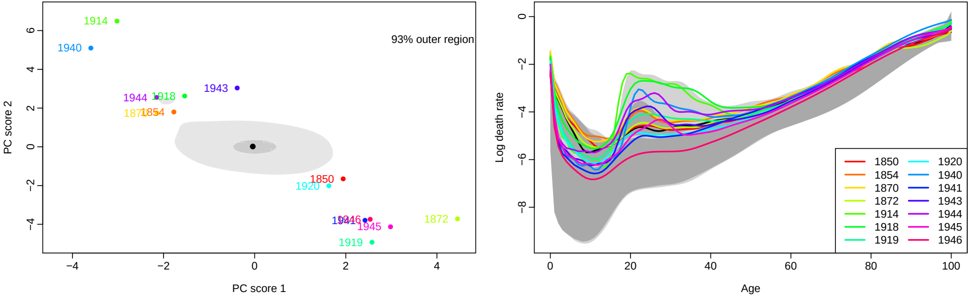
Functional HDR boxplot

- Bivariate HDR boxplot due to Hyndman (1996).
- Rank points by value of kernel density estimate.
- Display mode, 50% and (usually) 99% highest density regions (HDRs) and mode.

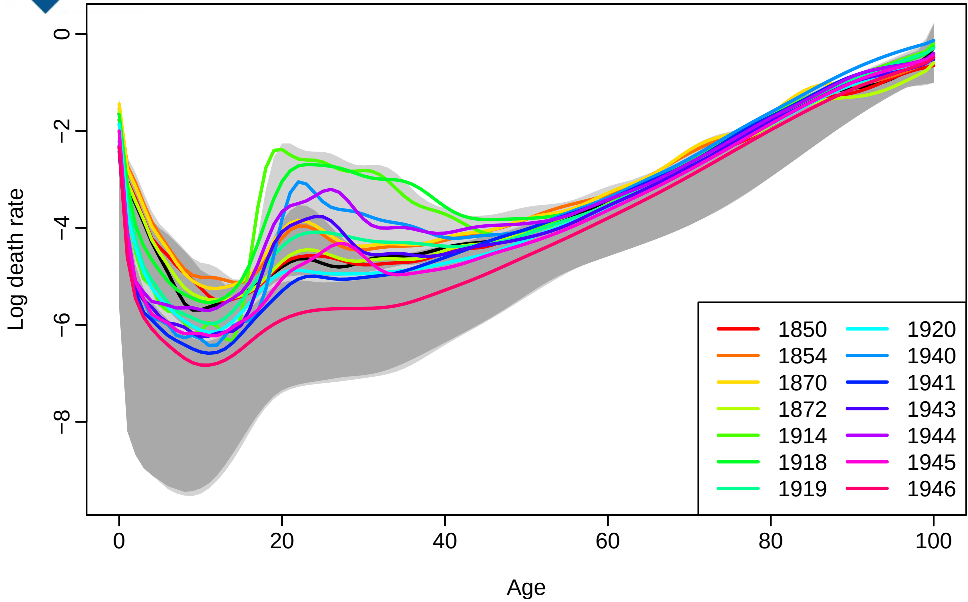


Functional HDR boxplot

- Bivariate HDR boxplot due to Hyndman (1996).
- Rank points by value of kernel density estimate.
- Display mode, 50% and (usually) 99% highest density regions (HDRs) and mode.
- Boundaries contain all curves inside HDRs.



Functional HDR boxplot



Outline

1 Functional time series

2 Functional principal components

3 Data visualization

4 References

Selected references

Hyndman, Shang (2010). “Rainbow plots, bagplots and boxplots for functional data”. *Journal of Computational and Graphical Statistics* **19**(1), 29–45

Hyndman, Ullah (2007). “Robust forecasting of mortality and fertility rates: A functional data approach”. *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

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