Rob J Hyndman

Functional time series
with applications in demography

1. Tools for functional time series analysis
Fertility rates

Australia fertility rates (1921)

Age

Fertility rate

Functional time series with applications in demography

1. Tools for functional time series analysis
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 \( t \), \( i = 1, \ldots, N, \quad t = 1, \ldots, T \).
- \( \lambda \) 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 \( t = T + 1, \ldots, T + h \).
\[ y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log[z_t(x_i)] & \text{if } \lambda = 0; \\
\lambda^{-1} [z_t^\lambda(x_i) - 1] & \text{otherwise.}
\end{cases} \]

\[ = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \]

- \( z_t(x_i) \) is observed data for age \( x_i \) in year \( t \), \( i = 1, \ldots, N, \quad t = 1, \ldots, T \).
- \( \lambda \) 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 \textbf{whole curve} \( z_t(x) \) for \( t = T + 1, \ldots, T + h \).
\[
y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log[z_t(x_i)] & \text{if } \lambda = 0; \\
\lambda^{-1} \left[ z_t^\lambda(x_i) - 1 \right] & \text{otherwise.}
\end{cases}
\]

\[
y_t(x_i) = s_t(x_i) + \sigma_t(x_i) \varepsilon_{t,i}
\]

- \(z_t(x_i)\) is observed data for age \(x_i\) in year \(t\), \(i = 1, \ldots, N, \quad t = 1, \ldots, T\).
- \(\lambda\) 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 \(t = T + 1, \ldots, T + h\).
$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log[z_t(x_i)] & \text{if } \lambda = 0; \\
\lambda^{-1} [z_t^\lambda(x_i) - 1] & \text{otherwise.}
\end{cases}$

$= s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i}$

- $z_t(x_i)$ is observed data for age $x_i$ in year $t$, $i = 1, \ldots, N$, $t = 1, \ldots, T$.
- $\lambda$ 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 $t = T + 1, \ldots, T + h$. 

Functional time series with applications in demography

1. Tools for functional time series analysis
$y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log[z_t(x_i)] & \text{if } \lambda = 0; \\
\lambda^{-1} [z_t^\lambda(x_i) - 1] & \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 $t$, $i = 1, \ldots, N$, $t = 1, \ldots, T$.
- $\lambda$ 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 $t = T + 1, \ldots, T + h$. 
Smoothing functional time series

\[ y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log(z_t(x_i)) & \text{if } \lambda = 0; \\
\lambda^{-1} (z_t^\lambda(x_i) - 1) & \text{otherwise.} 
\end{cases} \]

\[ = s_t(x_i) + \sigma_t(x_i) \varepsilon_{t,i} \]

- 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

\[ y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log(z_t(x_i)) & \text{if } \lambda = 0; \\
\lambda^{-1}(z_t^\lambda(x_i) - 1) & \text{otherwise.}
\end{cases} \]

\[ = s_t(x_i) + \sigma_t(x_i)\varepsilon_{t,i} \]

- 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

\[ y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log(z_t(x_i)) & \text{if } \lambda = 0; \\
\lambda^{-1} (z_t^\lambda(x_i) - 1) & \text{otherwise.}
\end{cases} \]

\[ = s_t(x_i) + \sigma_t(x_i) \varepsilon_{t,i} \]

- 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 \( w_t(x_i) = \sigma_t^{-2}(x_i) \).
\[ y_t(x_i) = g_\lambda(z_t(x_i)) = \begin{cases} 
\log(z_t(x_i)) & \text{if } \lambda = 0; \\
\lambda^{-1} (z_\lambda^t(x_i) - 1) & \text{otherwise.}
\end{cases} \]

= \( s_t(x_i) + \sigma_t(x_i) \varepsilon_{t,i} \)

- 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 \( w_t(x_i) = \sigma_t^{-2}(x_i) \).
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.

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

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

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

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

\[ D_t(x_i) = \text{number of deaths at age } x_i \text{ in year } t. \]

\[ E_t(x_i) = \text{total population aged } x_i \text{ on June 30 in year } t. \]

\[ m_t(x_i) = \frac{D_t(x_i)}{E_t(x_i)} = \text{observed mortality rate}. \]

\[ \mu_t(x_i) = \text{“true” mortality rate}. \]

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

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

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

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

Taylor series approx

\[ V[g_\lambda(X)] = \sigma_X^2[g'_\lambda(\mu_X)]^2 \]
\[ V[\log(X)] = \sigma_X^2/\mu_X^2 \]
Mortality

\[ D_t(x_i) = \text{number of deaths at age } x_i \text{ in year } t. \]
\[ E_t(x_i) = \text{total population aged } x_i \text{ on June 30 in year } t. \]
\[ m_t(x_i) = D_t(x_i)/E_t(x_i) = \text{observed mortality rate.} \]
\[ \mu_t(x_i) = \text{“true” mortality rate.} \]

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

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

\[ \sigma^2(x_i) = \mathbb{V}(\log[m_t(x_i)]) \]
\[ \approx \left[ \mu_t(x_i)E_t^{-1}(x_i) \right] \mu_t(x)^{-2} \]
\[ = \mu_t(x_i)^{-1}E_t(x_i)^{-1} \]

**Taylor series approx**

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

\( B_t(x_i) \sim Pn(E_t(x_i)\mu_t(x_i)) \)

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

\( f_t(x_i) = B_t(x_i)/E_t(x_i) = \text{observed fertility rate.} \)

\( \mu_t(x_i) = \text{“true” fertility rate.} \)

\[ B_t(x_i) \sim P_n(E_t(x_i)\mu_t(x_i)) \]

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

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

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

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

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

Taylor series approx

\[ V[g_\lambda(X)] = \sigma_X^2[g'_\lambda(\mu_X)]^2 \]
\[ = \sigma_X^2 \mu_X^{2\lambda - 2} \]
Fertility

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

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

$f_t(x_i) = B_t(x_i)/E_t(x_i) = \text{observed fertility rate.}$

$\mu_t(x_i) = \text{“true” fertility rate.}$

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

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

$\sigma^2(x_i) = V(g_\lambda[f_t(x_i)])$

$= \left[\mu_t(x_i)E_t^{-1}(x_i)\right] \mu_t(x)^{2\lambda-2}$

$= \mu_t(x_i)^{2\lambda-1}E_t(x_i)^{-1}$

$\approx f_t(x_i)^{2\lambda-1}E_t(x_i)^{-1}$

Taylor series approx

$V[g_\lambda(X)] = \sigma_X^2[g'_\lambda(\mu_X)]^2$

$= \sigma_X^2\mu_X^{2\lambda-2}$
Smoothing functional time series

France: male death rates (1821)

Log death rate vs. Age
Smoothing functional time series

France: male death rates (1816–2012)

Log death rate

Age

Functional time series with applications in demography
1. Tools for functional time series analysis
Smoothing functional time series

France: male death rates (1816–2012)
Smoothing functional time series

Australia fertility rates (1921)
Australia fertility rates (1950)
Australia fertility rates (2009)
Australia fertility rates (1921–2009)
Functional time series with applications in demography

1. Tools for functional time series analysis
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.
France: male death rates (1816–2012)
Australia fertility rates (1921–2009)
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 $\int \phi_1(x) dx = 1$.
2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_{2,t}$ such that $\int \phi_2(x) dx = 1$ and $\int \phi_1(x) \phi_2(x) dx = 0$.
3. Find the weight function $\phi_3(x)$ that...
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 $\beta_1$ subject to the constraint $\int \phi_1^2(x) dx = 1$.
2. Find the weight function $\phi_2(x)$ that maximizes the variance of $\beta_2$ 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 . . .
In FDA, each principal component is specified by a weight function $\phi_k(x)$. The PC scores for each year are given by

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

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 $\int \phi_1^2(x) dx = 1$.
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...
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...
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 $MISE = \frac{1}{T} \sum_{t=1}^{T} \int r_t^2 dx$ is the $k$th 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

$$\text{MISE} = \frac{1}{T} \sum_{t=1}^{T} \int r_t^2 dx$$

is the $k$th principal component (computed recursively).
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 $G$.

SVD of $G = \Phi \Lambda \Psi'$ where $\phi_k(x)$ is $k$th column of $\Phi$.

$\beta_{t,k}$ is $(t,k)$th element of $G \Phi$.

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 $i \neq j$. 
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 \( G \).

SVD of \( G = \Phi \Lambda \Psi' \) where \( \phi_k(x) \) is \( k \)th column of \( \Phi \).

\( \beta_{t,k} \) is \((t,k)\)th element of \( G\Phi \).

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 \( i \neq j \).
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 $G$.

SVD of $G = \Phi \Lambda \Psi'$ where $\phi_k(x)$ is $k$th column of $\Phi$.

$\beta_{t,k}$ is $(t,k)$th element of $G\Phi$.

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 $i \neq j$. 
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 \( G \).

SVD of \( G = \Phi \Lambda \Psi' \) where \( \phi_k(x) \) is \( k \)th column of \( \Phi \).

\( \beta_{t,k} \) is \((t, k)\)th element of \( G\Phi \).

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 \( G \).
- SVD of \( G = \Phi \Lambda \Psi' \) where \( \phi_k(x) \) is \( k \)th column of \( \Phi \).
- \( \beta_{t,k} \) is \((t, k)\)th element of \( G\Phi \).
- 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 $G$.
- SVD of $G = \Phi \Lambda \Psi'$ where $\phi_k(x)$ is $k$th column of $\Phi$.
- $\beta_{t,k}$ is $(t, k)$th element of $G \Phi$.
- 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 time series with applications in demography
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 \( G \).

SVD of \( G = \Phi \Lambda \Psi' \) where \( \phi_k(x) \) is \( k \)th column of \( \Phi \).

\( \beta_{t,k} \) is \((t, k)\)th element of \( G\Phi \).

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 \( i \neq j \).
Eigenvector approach

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

Let $V = (T - 1)^{-1}G'G$ be $m \times m$ sample covariance matrix of $G$.

Let $\Phi_K = [\phi_1, \ldots, \phi_K]$ consist of the first $K$ eigenvectors of $V$ where $K \leq T - 1$. The $(i,j)$th element of $\Phi_K$ is $\phi_i(x_j^*)$.

Robust versions possible using robust “covariance” estimation.
Eigenvector approach

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

Main effects

Interaction

French male mortality

Robust PCA

Functional time series with applications in demography

1. Tools for functional time series analysis
Functional principal components

Main effects

Interaction

Australian fertility

Functional time series with applications in demography

1. Tools for functional time series analysis
Functional principal components

Australian fertility

Robust PCA
French male mortality rates

France: male death rates (1816–2012)
French male mortality rates

France: male death rates (1816–2012)

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.
Febrerò, Galeano and González-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.
**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.
**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.
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

Scatterplot of first two PC scores

PC score 1

PC score 2
Robust PC scores

Scatterplot of first two PC scores

- WW1
- WW2
- 1800s
- 2000s

PC score 2

PC score 1

Functional time series with applications in demography

1. Tools for functional time series analysis
Robust PC scores

Functional time series with applications in demography

1. Tools for functional time series analysis

- PC scores

- Time

- Comp.1

- Comp.2
French male mortality rates

France: differenced male death rates (1816–2012)
Robust PC scores

Scatterplot of first two PC scores on differences

PC score 1
PC score 2
Robust PC scores

Scatterplot of first two PC scores on differences

- WW1
- WW2
- 1800s
- 2000s

PC score 1 vs. PC score 2
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.
The halfspace depth of a point \( q \):

(Due to Hotelling, 1929; Tukey, 1975)

For each closed halfspace that contains \( q \), count the number of observations not in the 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 the convex hull of the data has depth zero.

Fig. 5: Depth of certain points with respect to a data set.

\[
depth(A) = 3, \quad depth(B) = 1, \quad depth(C) = 2.
\]
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

France: male death rates (1816–2012)
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 time series with applications in demography

1. Tools for functional time series analysis

---

**Age**

Log death rate

- 1850
- 1854
- 1870
- 1871
- 1872
- 1914
- 1918
- 1919
- 1920
- 1940
- 1941
- 1943
- 1944
- 1945
- 1946

---

**Functional time series with applications in demography 1. Tools for functional time series analysis**
Functional bagplot

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

1850
1854
1870
1871
1872
1914
1918
1919
1920
1940
1941
1943
1944
1945
1946
Kernel density estimate

Scatterplot of first two PC scores on differences

PC score 1
PC score 2
Kernel density estimate
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.
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

Functional time series with applications in demography

1. Tools for functional time series analysis
1 Functional time series

2 Functional principal components

3 Data visualization

4 References


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