Analiza kanoniczna dla danych funkcjonalnych

Waldemar Wołyński

Wydział Matematyki i Informatyki UAM Poznań

Poznań, 18 października 2017

WSTĘP

Problem

Dane pochodzą z EUROSTATU i dotyczą spożycia indywidualnego według celu w 27 wybranych krajach europejskich w latach 2000-2010.
Klasyfikacja Spożycia Indywidualnego według Celu (COICOP) obejmuje 12 grup wydatków:

1. Artykuły żywnościowe i napoje bezalkoholowe,
2. Napoje alkoholowe i tytoń,
3. Odzież i obuwie,
4. Mieszkanie, woda, elektryczność, gaz i inne paliwa,
5. Wyposażenie wnętrza, sprzęty domowe i bieżące utrzymanie budynku,
6. Opieka zdrowotna,
7. Transport,
8. Łączność,
9. Wypoczynek i kultura,
10. Szkolnictwo,
11. Hotele, kawiarnie i restauracje,
12. Różne towary i usługi.

Problem

Wydatki uwzględnione w grupie 2 zostały podzielone na wydatki na napoje alkoholowe (cecha Y_1) oraz wydatki na wyroby tytoniowe (cecha Y_2). Jedenaście pozostałych grup wydatków przyjęto jako cechy $X_1, X_2, \ldots, X_{11}$.

Powstały w ten sposób dwie nowe grupy cech:

- Grupa I: cechy Y_1, Y_2,
- Grupa II: cechy $X_1, \ldots, X_{11}$.

Dysponujemy watościami każdej z cech dla 27 jednostek (kraje) i 11 momentów czasowych (lata 2000-2010).

Jak zbadać korelację pomiędzy badanymi grupami cech?

CZĘŚĆ 1

Korelacja

Niech $X \in \mathbf{R}$ i $Y \in \mathbf{R}$ będą zmiennymi losowymi. Ponadto załóżmy, że $\mathrm{E}(X) = \mathrm{E}(Y) = 0$.

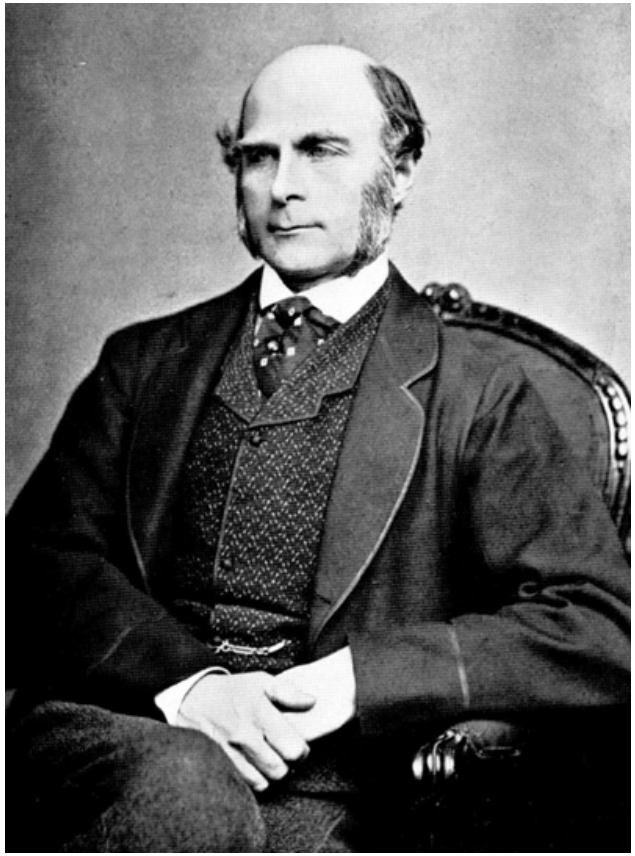
Współczynnikiem korelacji (liniowej Pearsona) nazywamy liczbę:

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Własności:

1. $-1 \leqslant \rho_{X,Y} \leqslant 1$,
2. Jeżeli $\rho_{X,Y} = 0$, to brak korelacji liniowej pomiędzy zmiennymi X i Y,
3. Jeżeli $|\rho_{X,Y}| = 1$, to istnieje liniowa zależność pomiędzy zmiennymi X i Y.

Karl Pearson
1857 - 1936



Francis Galton
1822 - 1911

Dane

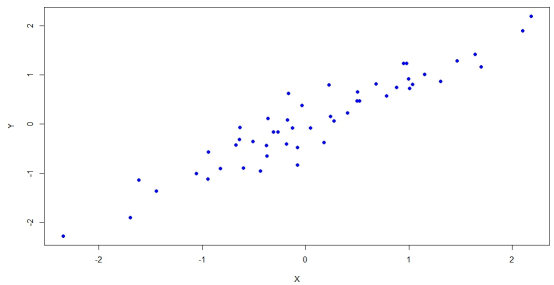
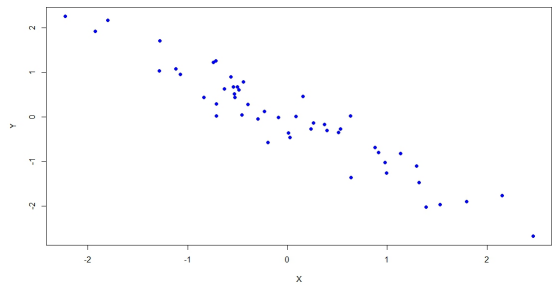
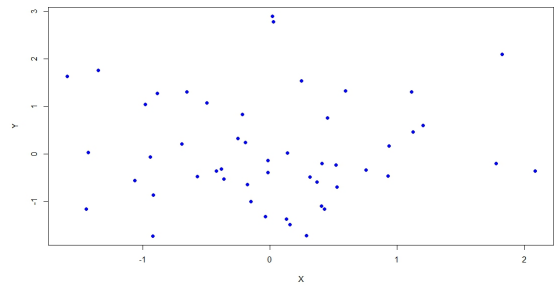
Niech

$\boldsymbol{x} = (x_1, \ldots, x_n)'$ - obserwacje (wycentrowane) zmiennej X,

$\boldsymbol{y} = (y_1, \ldots, y_n)'$ - obserwacje (wycentrowane) zmiennej Y.

Współczynnik korelacji (z próby):

$$r_{X,Y} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}.$$

Wykres rozrzutu (diagram korelacyjny)

Korelacje kanoniczne

Niech $\boldsymbol{X} \in \mathbf{R}^p$ i $\boldsymbol{Y} \in \mathbf{R}^q$ będą wektorami losowymi losowymi. Ponadto załóżmy, że $\mathrm{E}(\boldsymbol{X}) = \mathrm{E}(\boldsymbol{Y}) = \mathbf{0}$.

Założenie to nie powoduje utraty ogólności rozważań, ponieważ korelacje kanoniczne wyznaczane są jedynie na podstawie macierzy kowariancji wektorów $\boldsymbol{X}$ i $\boldsymbol{Y}$ postaci

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{XX} & \boldsymbol{\Sigma}_{XY} \\ \boldsymbol{\Sigma}_{YX} & \boldsymbol{\Sigma}_{YY} \end{bmatrix}.$$

Korelacje kanoniczne

Zmienne kanoniczne wektorów losowych $\boldsymbol{X}$ i $\boldsymbol{Y}$ definiujemy następująco:

$$U =< \boldsymbol{u}, \boldsymbol{X} >= \sum_{i=1}^{p} u_i X_i,$$

$$V =< \boldsymbol{v}, \boldsymbol{Y} >= \sum_{j=1}^{q} v_j Y_j,$$

gdzie $\boldsymbol{u}$ i $\boldsymbol{v}$ są wektorami wagowymi dobranymi tak, aby zmaksymalizować współczynnik korelacji $\rho_{U,V}$ przy dodatkowym warunku ograniczającym $\mathrm{Var}(U) = \mathrm{Var}(V) = 1$.

Maksymalny współczynnik korelacji $\rho_{U,V}$ nazywamy współczynnikiem korelacji kanonicznej.

Korelacje kanoniczne

Zagadnienie maksymalizacji współczynnika korelacji $\rho_{U,V}$ sprowadza się do rozwiązania równań:

$$|\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY}\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{YX} - \rho^2 \boldsymbol{I}_p|,$$
$$|\mathbf{\Sigma}_{YY}^{-1}\mathbf{\Sigma}_{YX}\mathbf{\Sigma}_{XX}^{-1}\mathbf{\Sigma}_{XY} - \rho^2 \boldsymbol{I}_q|.$$

Liczba niezerowych pierwiastków tych równań jest równa $s = \text{rząd}(\mathbf{\Sigma}_{XY})$.

Otrzymujemy s par nieskorelowanych zmiennych kanonicznych (U_i, V_i), $i = 1, \ldots, s$.
Pierwiastki kwadratowe z $\rho_1^2 \geqslant \rho_2^2 \geqslant \cdots \geqslant \rho_s^2$ są współczynnikami korelacji kanonicznej.
Wektory własne odpowiadające odpowiadające pierwiastkom równań są wektorami wagowymi.

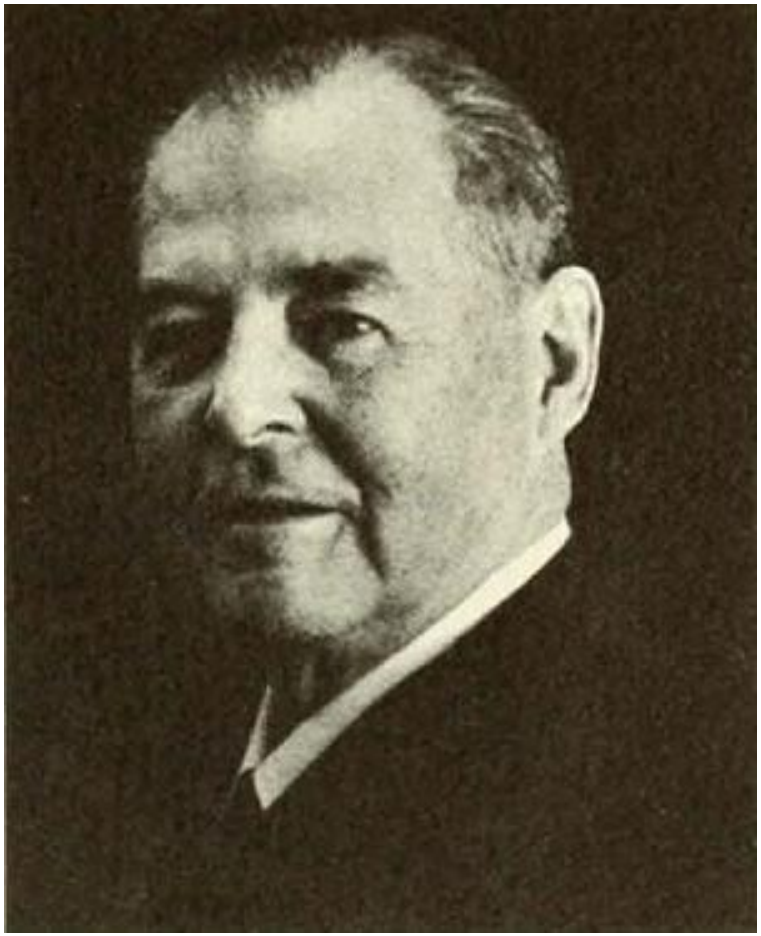
Korelacje kanoniczne

O wkładzie poszczególnych składowych wektorowów losowych $\boldsymbol{X}$ oraz $\boldsymbol{Y}$ w budowę zmiennych kanonicznych można wnioskować na podstawie wektorów wagowych $\boldsymbol{u}$ oraz $\boldsymbol{v}$.

Wkład j-tej składowej wektora losowego $\boldsymbol{X}$ w budowę zmiennej kanonicznej U określamy wzorem:

$$\frac{|u_j|}{\sum_{i=1}^{p}|u_i|} \times 100\%, \ j = 1, \ldots, p.$$

Analogicznie wnioskujemy dla zmiennej kanonicznej V.



Harold Hotelling

1895 - 1973

Dane

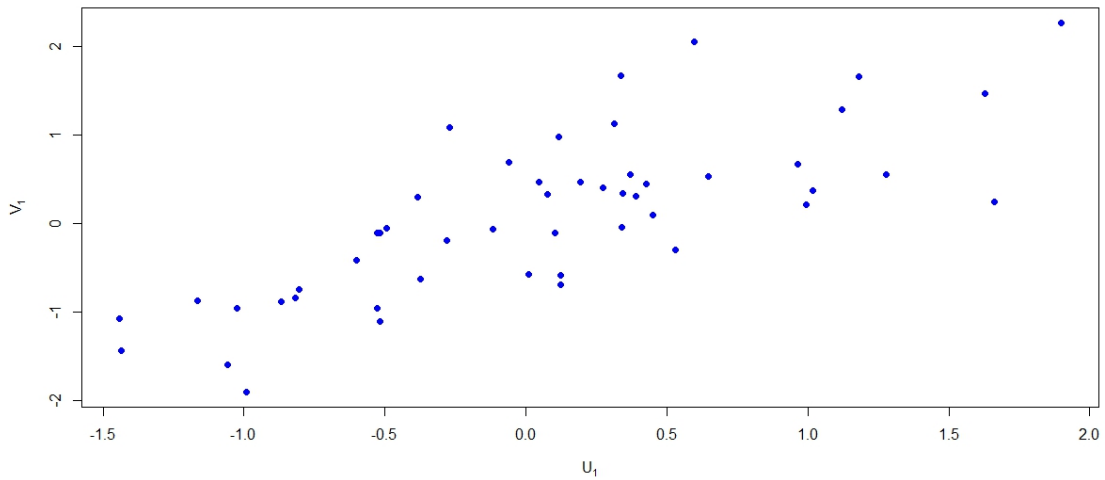
Niech

$\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$ - obserwacje (wycentrowane) wektora X,

$\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$ - obserwacje (wycentrowane) wektora Y.

Wtedy

$$\hat{\boldsymbol{\Sigma}}_{XX} = \frac{1}{n}\boldsymbol{x}'\boldsymbol{x},\ \hat{\boldsymbol{\Sigma}}_{YY} = \frac{1}{n}\boldsymbol{y}'\boldsymbol{y},\ \hat{\boldsymbol{\Sigma}}_{XY} = \frac{1}{n}\boldsymbol{x}'\boldsymbol{y} = \hat{\boldsymbol{\Sigma}}'_{YX}.$$



Rzut na pierwszą parę zmiennych kanonicznych

Uogólnienia

1. Więcej niż dwie grupy zmiennych.
2. Nieliniowe zmienne kanoniczne.
3. Dane geograficznie ważone.
4. Dane czasowo-przestrzenne.

Funkcjonalne korelacje kanoniczne

Niech $\boldsymbol{X} \in L_2^p(I)$ i $\boldsymbol{Y} \in L_2^q(I)$ są wektorowymi procesami losowymi, gdzie $L_2(I)$ jest przestrzenią Hilberta funkcji całkowalnych z kwadratem na przedziale I. Ponadto załóżmy, że $\mathrm{E}(\boldsymbol{X}(t)) = \mathrm{E}(\boldsymbol{Y}(t)) = \boldsymbol{0}, \ t \in I$.

Założenie to nie powoduje utraty ogólności rozważań, ponieważ funkcjonalne korelacje kanoniczne wyznaczane są jedynie na podstawie macierzy kowariancji procesów $\boldsymbol{X}$ i $\boldsymbol{Y}$ postaci

$$\boldsymbol{\Sigma}(s,t) = \begin{bmatrix} \boldsymbol{\Sigma}_{XX}(s,t) & \boldsymbol{\Sigma}_{XY}(s,t) \\ \boldsymbol{\Sigma}_{YX}(s,t) & \boldsymbol{\Sigma}_{YY}(s,t) \end{bmatrix}, \ s,t \in I.$$

Funkcjonalne korelacje kanoniczne

Funkcjonalne zmienne kanoniczne procesów losowych $\boldsymbol{X}$ i $\boldsymbol{Y}$ definiujemy następująco:

$$U =< \boldsymbol{u}, \boldsymbol{X} >= \int_I \boldsymbol{u}'(t)\boldsymbol{X}(t)dt,$$

$$V =< \boldsymbol{v}, \boldsymbol{Y} >= \int_I \boldsymbol{v}'(t)\boldsymbol{Y}(t)dt,$$

gdzie $\boldsymbol{u} \in L_2^p(I)$ i $\boldsymbol{v} \in L_2^q(I)$ są funkcjami wagowymi dobranymi tak, aby zmaksymalizować współczynnik korelacji $\rho_{U,V}$ przy dodatkowym warunku ograniczającym $\mathrm{Var}(U) = \mathrm{Var}(V) = 1$.

Maksymalny współczynnik korelacji $\rho_{U,V}$ nazywamy funkcjonalnym współczynnikiem korelacji kanonicznej.

Funkcjonalne korelacje kanoniczne - problem!!!

Maksymalizacja funkcjonalnego współczynnika korelacji kanonicznej nie daje zadawalających wyników. Dowolnie wybrając funkcję wagową $\boldsymbol{u}$ możemy znaleźć funkcję wagową $\boldsymbol{v}$ taką, aby funkcjonalny współczynnik korelacji kanonicznej $\rho_{\boldsymbol{X},\boldsymbol{Y}}$ był równy 1.
Rozwiązaniem jest modyfikacja warunków ograniczających poprzez wprowadzenie dodatkowej funkcji kary:

$$PEN_2(\boldsymbol{u}) = \int_I \left(\frac{\partial^2 \boldsymbol{u}(t)}{\partial t^2} \right)' \frac{\partial^2 \boldsymbol{u}(t)}{\partial t^2} dt$$

oraz

$$PEN_2(\boldsymbol{v}) = \int_I \left(\frac{\partial^2 \boldsymbol{v}(t)}{\partial t^2} \right)' \frac{\partial^2 \boldsymbol{v}(t)}{\partial t^2} dt.$$

Funkcjonalne korelacje kanoniczne - problem!!!

Współczynnik kary PEN_2 słuzy do oszacowania gładkości funkcji. Kwadrat drugiej pochodnej funkcji w momencie t jest nazywany jej krzywizną w punkcie t. Naturalną miarą krzywizny funkcji jest zatem scałkowany kwadrat drugiej pochodnej tej funkcji.

Po wprowadzeniu kary warunki ograniczające przyjmują postać:

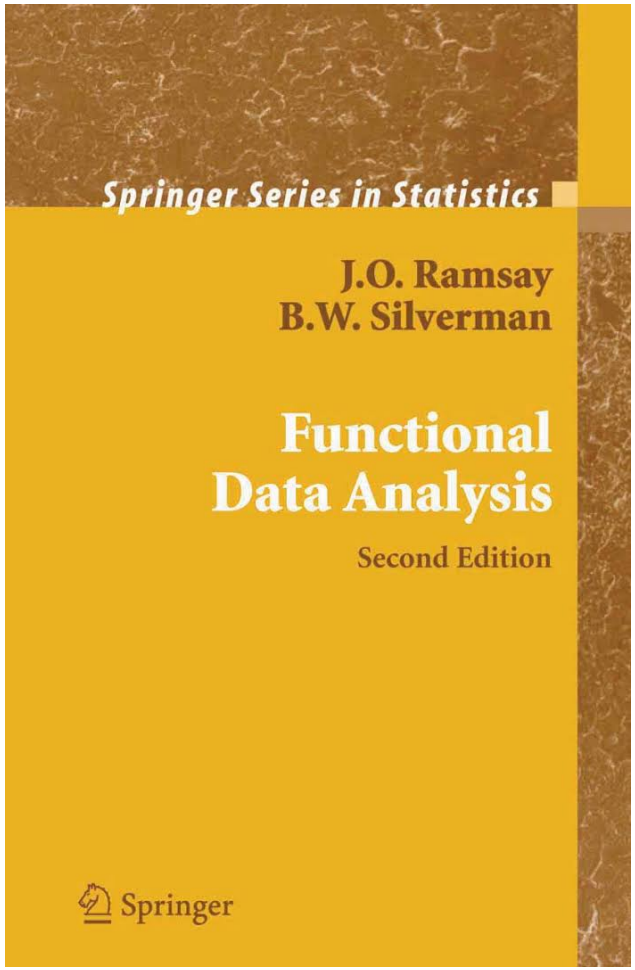
$$\begin{aligned} \mathrm{Var}(U) + \lambda PEN_2(\boldsymbol{u}) &= 1, \\ \mathrm{Var}(V) + \lambda PEN_2(\boldsymbol{v}) &= 1. \end{aligned}$$

Dane funkcjonalne

Besse (1979);

Ramsay & Silverman (1997, 2005);

Horváth & Kokoszka (2012).



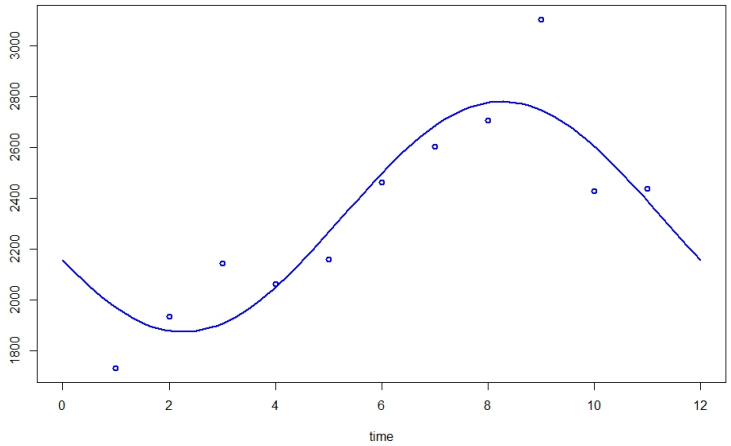
Model

W dalszym ciągu zakładamy, że każda składowa X_g procesu $\boldsymbol{X}$ i każda składowa Y_h procesu $\boldsymbol{Y}$ może być reprezentowana przez skończoną liczbę funkcji bazowych $\{\varphi_e\}$ i $\{\varphi_f\}$:

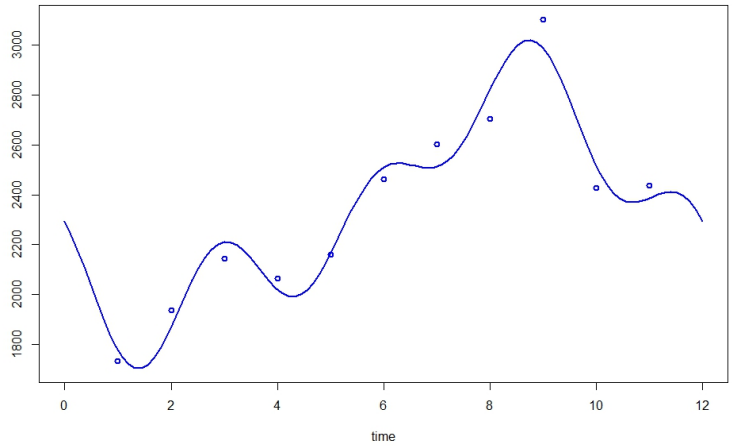
$$X_g(t) = \sum_{e=0}^{E_g} \alpha_{ge}\varphi_e(t), t \in I, g = 1, 2, ..., p,$$

$$Y_h(t) = \sum_{f=0}^{F_h} \beta_{hf}\varphi_f(t), t \in I, h = 1, 2, ..., q.$$

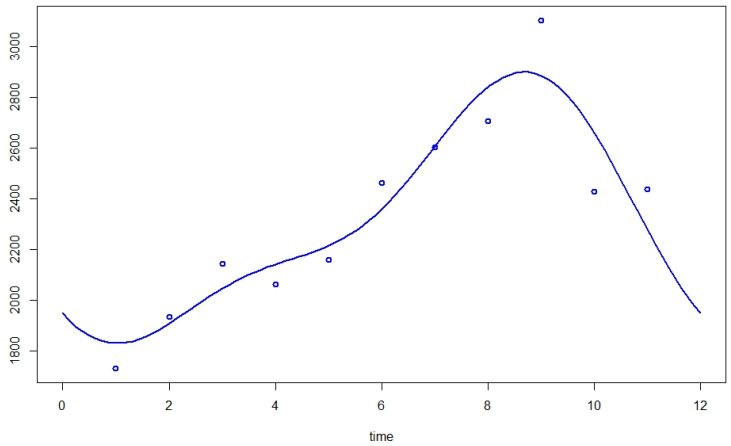
Stopień gładkości funkcji X_g i Y_h zależy od wartości E_g i F_h (małe wartości dają mniejsze wygładzenie funkcji).


1800
2000
2200
2400
2600
2800
3000
0
2
4
6
8
10
12
time

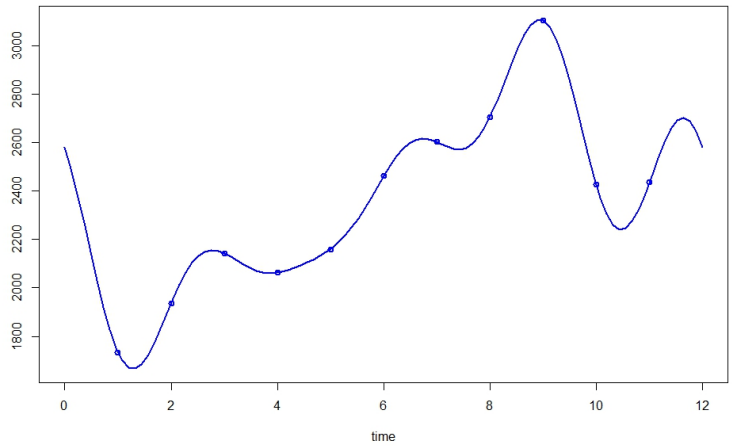
$E = 2$


1800
2000
2200
2400
2600
2800
3000
0
2
4
6
8
10
12
time

$E = 8$


1800
2000
2200
2400
2600
2800
3000
0
2
4
6
8
10
12
time

$E = 4$


1800
2000
2200
2400
2600
2800
3000
0
2
4
6
8
10
12
time

$E = 10$

Model

Wprowadźmy następujące oznaczenia:

$$\boldsymbol{\alpha} = (\alpha_{10}, ..., \alpha_{1E_1}, ..., \alpha_{p0}, ..., \alpha_{pE_p})',$$

$$\boldsymbol{\beta} = (\beta_{10}, ..., \beta_{1F_1}, ..., \beta_{q0}, ..., \beta_{qF_q})',$$

$$\boldsymbol{\Phi}_1(t) = \begin{bmatrix} \varphi'_{E_1}(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \varphi'_{E_2}(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \varphi'_{E_p}(t) \end{bmatrix},$$

$$\boldsymbol{\Phi}_2(t) = \begin{bmatrix} \varphi'_{F_1}(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \varphi'_{F_2}(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \varphi'_{F_q}(t) \end{bmatrix},$$

gdzie $\varphi_{E_1}, ..., \varphi_{E_p}$ i $\varphi_{F_1}, ..., \varphi_{F_q}$ są wektorami ortonormalnych funkcji bazowych przestrzeni $L_2(I)$.

Model

Używając notacji macierzowej procesy $\boldsymbol{X}$ i $\boldsymbol{Y}$ mają następującą reprezentację

$$\boldsymbol{X}(t) = \boldsymbol{\Phi}_1(t)\boldsymbol{\alpha}, \quad \boldsymbol{Y}(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{\beta}.$$

Oznacza to, że realizacje procesów $\boldsymbol{X}$ i $\boldsymbol{Y}$ zawarte są w skończenie wymiarowych podprzestrzeniach przestrzeni $L_2^p(I)$ i $L_2^q(I)$ odpowiednio. Te podprzestrzenie będziemy oznaczali przez $\mathcal{L}_2^p(I)$ i $\mathcal{L}_2^q(I)$.

Ponadto, dla wektorów $\boldsymbol{\alpha}$ i $\boldsymbol{\beta}$ mamy:

$$\mathrm{E}(\boldsymbol{\alpha}) = \mathbf{0}, \ \mathrm{E}(\boldsymbol{\beta}) = \mathbf{0}$$

oraz

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\alpha\alpha} & \boldsymbol{\Sigma}_{\alpha\beta} \\ \boldsymbol{\Sigma}_{\beta\alpha} & \boldsymbol{\Sigma}_{\beta\beta} \end{bmatrix}.$$

Bazy w $L_2(I)$

- Wielomianowa;
- Fouriera;
- Funkcji sklejanych (spline);
- Funkcji falkowych (wavelet).

Baza Fouriera zdefiniowana jest w następujący sposób:

$$\varphi_0(t) = \frac{1}{\sqrt{T}},$$

$$\varphi_{2k-1}(t) = \sqrt{\frac{2}{T}} \sin \frac{2\pi kt}{T},$$

$$\varphi_{2k}(t) = \sqrt{\frac{2}{T}} \cos \frac{2\pi kt}{T},$$

gdzie $t \in I = [0, T], k = 1, 2, \ldots$.

Idea

Podstawowa zależność

$$\boldsymbol{X} \quad \longleftrightarrow \quad \boldsymbol{\alpha}$$

$$\text{proces} \quad \longleftrightarrow \quad \text{wektor}$$

Jeżeli $\boldsymbol{X} \in \mathcal{L}_2^p(I)$ i $\boldsymbol{Y} \in \mathcal{L}_2^q(I)$, to

$$\boldsymbol{\Sigma}_{XX}(s,t) = \boldsymbol{\Phi}_1(s)\boldsymbol{\Sigma}_{\alpha\alpha}\boldsymbol{\Phi}_1'(t),$$

$$\boldsymbol{\Sigma}_{XY}(s,t) = \boldsymbol{\Phi}_1(s)\boldsymbol{\Sigma}_{\alpha\beta}\boldsymbol{\Phi}_2'(t),$$

$$\boldsymbol{\Sigma}_{YY}(s,t) = \boldsymbol{\Phi}_2(s)\boldsymbol{\Sigma}_{\beta\beta}\boldsymbol{\Phi}_2'(t).$$

Funkcjonalne zmienne kanoniczne

Niech $\boldsymbol{u} \in \mathcal{L}_2^p(I)$ i $\boldsymbol{v} \in \mathcal{L}_2^q(I)$.
Wtedy

$$\boldsymbol{u}(t) = \boldsymbol{\Phi}_1(t)\boldsymbol{u}, \;\; \boldsymbol{v}(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{v},$$

gdzie $t \in I$.
Zatem

$$\mathrm{Cov}(U, V) = \boldsymbol{u}'\boldsymbol{\Sigma}_{\alpha\beta}\boldsymbol{v},$$

$$\mathrm{Var}(U) = \boldsymbol{u}'\boldsymbol{\Sigma}_{\alpha\alpha}\boldsymbol{u}, \quad \mathrm{Var}(U) = \boldsymbol{v}'\boldsymbol{\Sigma}_{\beta\beta}\boldsymbol{v},$$

$$PEN_2(u) = \boldsymbol{u}'\boldsymbol{R}_1\boldsymbol{u}, \quad PEN_2(v) = \boldsymbol{v}'\boldsymbol{R}_2\boldsymbol{v},$$

gdzie

$$\boldsymbol{R}_i = \int_I \left(\frac{\partial^2 \boldsymbol{\Phi}_i(t)}{\partial t^2}\right)' \frac{\partial^2 \boldsymbol{\Phi}_i(t)}{\partial t^2} dt, \;\; i = 1, 2.$$

Funkcjonalne zmienne kanoniczne

Funkcjonalne zmienne kanoniczne U i V dobieramy tak, zmaksymalizować

$$\mathrm{Cov}(U, V) = \boldsymbol{u}'\boldsymbol{\Sigma}_{\alpha\beta}\boldsymbol{v},$$

przy warunkach

$$\boldsymbol{u}'(\boldsymbol{\Sigma}_{\alpha\alpha} + \lambda\boldsymbol{R}_1)\boldsymbol{u} = \boldsymbol{v}'(\boldsymbol{\Sigma}_{\beta\beta} + \lambda\boldsymbol{R}_2)\boldsymbol{v} = 1.$$

Dane

Procesy $\boldsymbol{X}$ i $\boldsymbol{Y}$ są obserwowane w skończonej liczbie momentów czasowych. Proces transformacji danych dyskretnych do danych funkcjonalnych jest wykonywany oddzielnie dla każdej realizacji każdej składowej procesu. Niech x_{gj} oznacza obserwowaną wartość składowej X_g, $g = 1, 2, \ldots p$ w j-tym momencie czasowym t_j, gdzie $j = 1, 2, ..., J$. Podobnie, niech y_{hj} oznacza obserwowaną wartość składowej Y_h, $h = 1, 2, \ldots q$ w j-tym momencie czasowym t_j, gdzie $j = 1, 2, ..., J$. Wtedy nasze dane składają się z pJ par (t_j, x_{gj}) oraz z qJ par (t_j, y_{hj}).
Współczynniki $\boldsymbol{\alpha}_i$ i $\boldsymbol{\beta}_i$ są estymowane metodą najmniejszych kwadratów.
Oznaczmy te estymatory przez $\boldsymbol{a}_i$ i $\boldsymbol{b}_i$, $i = 1, 2, \ldots, n$.
W rezultacie procesu transformacji otrzymujemy dane funkcjonalne postaci:

$$\boldsymbol{x}_i(t) = \boldsymbol{\Phi}_1(t)\boldsymbol{a}_i, \ \boldsymbol{y}_i(t) = \boldsymbol{\Phi}_2(t)\boldsymbol{b}_i,$$

gdzie $t \in I$, $i = 1, 2, \ldots, n$

Dane

Niech

$$\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)' \quad \boldsymbol{B} = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)'.$$

Wtedy

$$\hat{\boldsymbol{\Sigma}}_{\alpha\alpha} = \frac{1}{n}\boldsymbol{A}'\boldsymbol{A}, \quad \hat{\boldsymbol{\Sigma}}_{\beta\beta} = \frac{1}{n}\boldsymbol{B}'\boldsymbol{B}, \quad \hat{\boldsymbol{\Sigma}}_{\alpha\beta} = \frac{1}{n}\boldsymbol{A}'\boldsymbol{B}.$$

Zatem dla funkcjonalnego współczynnika korelacji kanonicznej mamy:

$$r_{\boldsymbol{X},\boldsymbol{Y}} = \max_{\boldsymbol{u},\boldsymbol{v}} \boldsymbol{u}'\hat{\boldsymbol{\Sigma}}_{\alpha\beta}\boldsymbol{v},$$

przy warunku

$$\boldsymbol{u}'(\hat{\boldsymbol{\Sigma}}_{\alpha\alpha} + \lambda\boldsymbol{R}_1)\boldsymbol{u} = \boldsymbol{v}'(\hat{\boldsymbol{\Sigma}}_{\beta\beta} + \lambda\boldsymbol{R}_2)\boldsymbol{v} = 1.$$

Wkład zmiennych pierwotnych

O wkładzie poszczególnych składowych wektorowych procesów losowych $\boldsymbol{X}$ oraz $\boldsymbol{Y}$ w budowę zmiennych kanonicznych można wnioskować na podstawie wektorowych funkcji wagowych $\boldsymbol{u}$ oraz $\boldsymbol{v}$.

Niech P_j będzie polem pod modułem funkcji u_j na przedziale I. Wkład j-tej składowej procesu losowego $\boldsymbol{X}$ w budowę zmiennej kanonicznej U określamy wzorem:

$$P_j^\star = \frac{P_j}{\sum_{i=1}^{p} P_i} \times 100\%,\ j = 1, \ldots, p.$$

Analogicznie wnioskujemy dla zmiennej kanonicznej V.

CZĘŚĆ 2

Przykład

Dwie grupy cech:

- Grupa I: cechy Y_1, Y_2 - wydatki na napoje alkoholowe i wyroby tytoniowe,
- Grupa II: cechy $X_1, \ldots, X_{11}$ - wydatki na pozostałe artykuły konsumpcyjne.

Dysponujemy watościami każdej z cech dla 27 jednostek (kraje) i 11 momentów czasowych (lata 2000-2010).

Przykład

W celu ujednolicenia wartości rozpatrywanych cech, które mają różne przedziały zmienności, przeprowadzono ich unitaryzację.
Niech x_{kij} będzie wartością cechy X_k zaobserwowaną w i-tym kraju oraz j-tym roku.
Wówczas zunitaryzowana wartość z_{kij} wartości x_{kij} ma postać:

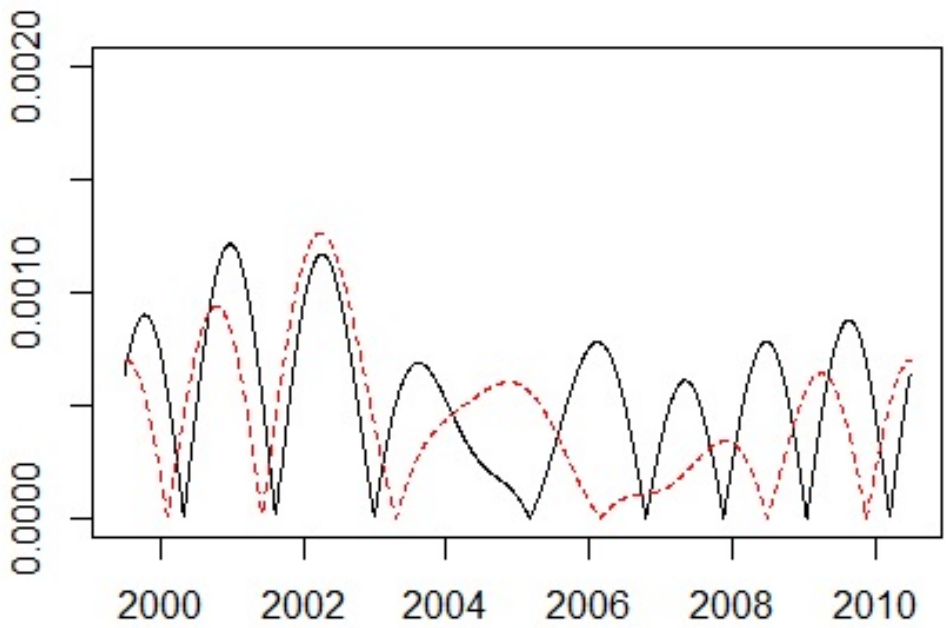
$$z_{kij} = \frac{x_{kij} - \min_i x_{kij}}{r_{kj}}.$$

gdzie

$$r_{kj} = \max_i x_{kij} - \min_i x_{kij}$$

jest rozstępem k-tej cechy w j-tym roku.
Po unitaryzacji rozstęp wszystkich cech we wszystkich latach jest stały i równy 1, natomiast wariancje i kowariancje w danym roku są proporcjonalne do wariancji i kowariancji cech bez unitaryzacji w tym roku.

Przykład



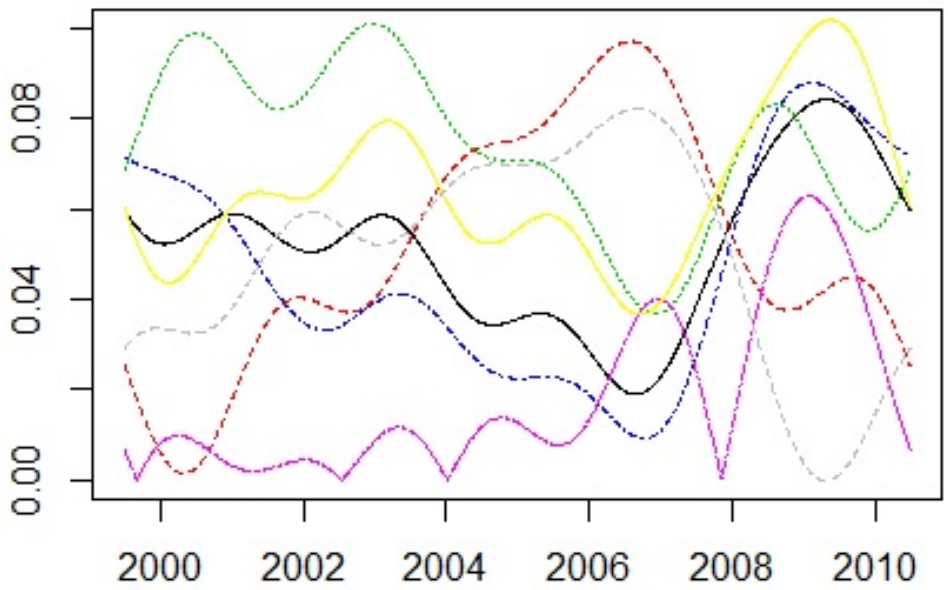
Wykres funkcji wagowych pierwszej zmiennej kanonicznej dla procesu $\boldsymbol{Y}$

Przykład

j	$P_j^\star$
1 - napoje alkoholowe	54,7
2 - wyroby tytoniowe	45,3

Wskaźniki $P_j^\star$ odpowiadające zmiennym procesu $\boldsymbol{Y}$.

Przykład



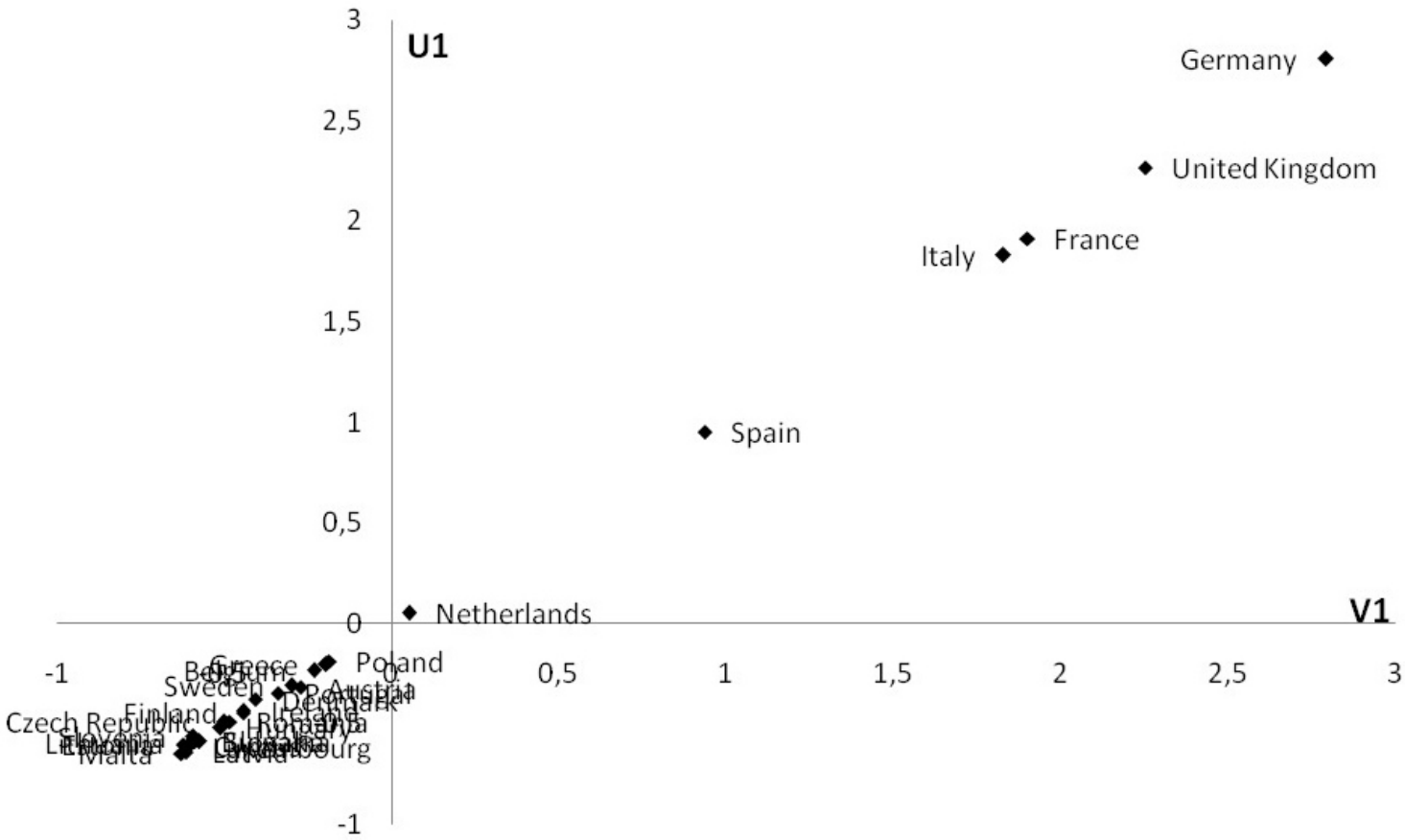
Wykres funkcji wagowych pierwszej zmiennej kanonicznej dla procesu $\boldsymbol{X}$

Przykład

j	$P_j^\star$
1 - artykuły żywnościowe i napoje bezalkoholowe	10,29
2 - odzież i obuwie	10,12
3 - mieszkanie, woda, elektryczność, gaz i inne paliwa	9,28
4 - wyposażenie wnętrza, sprzęty domowe i bieżące utrzymanie budynku	9,36
5 - opieka zdrowotna	7,23
6 - transport	9,96
7 - łączność	9,19
8 - wypoczynek i kultura	9,27
9 - szkolnictwo	8,16
10 - hotele, kawiarnie i restauracje	9,26
11 - rózne towary i usługi	7,54

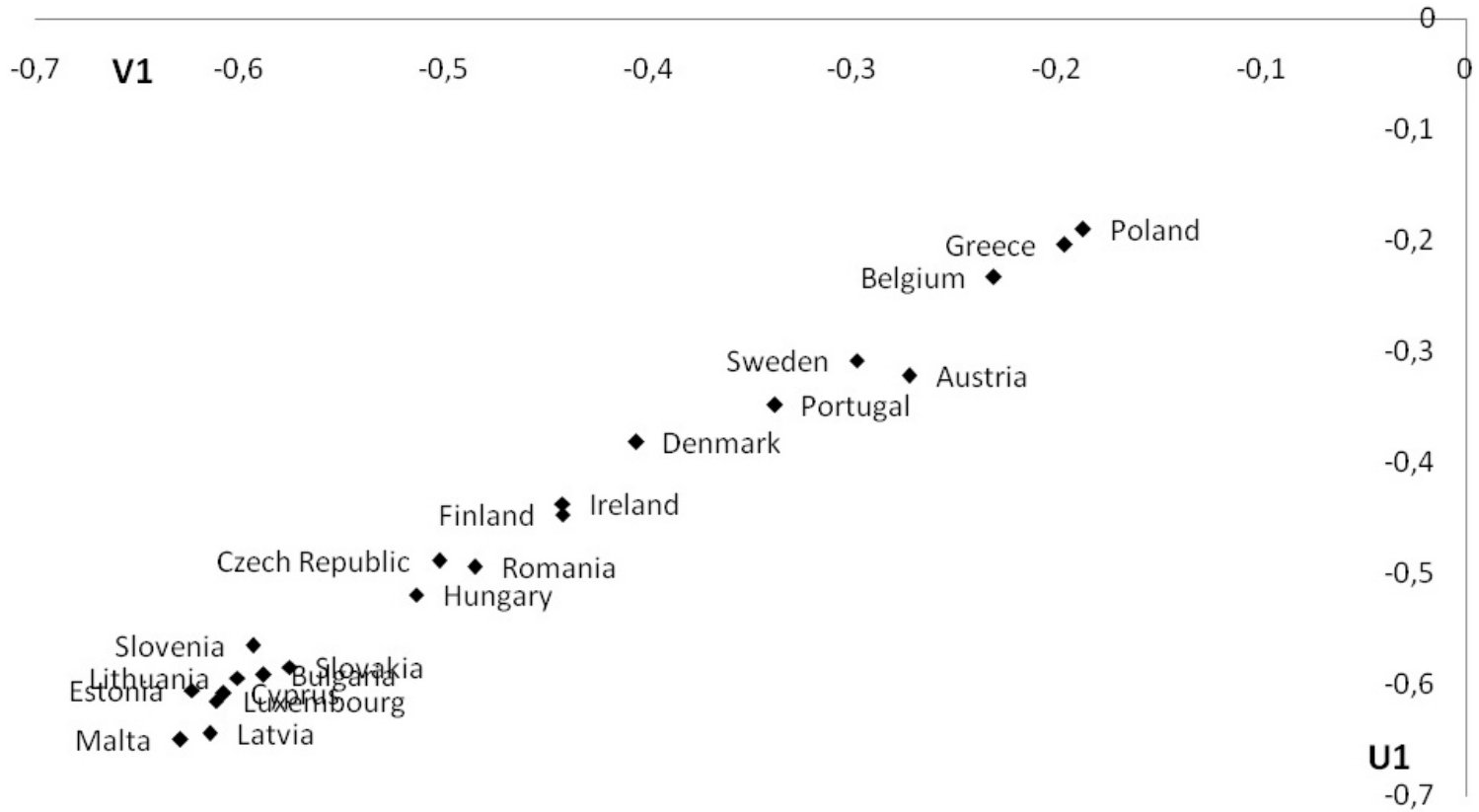
Wskaźniki $P_j^\star$ odpowiadające zmiennym procesu $\boldsymbol{X}$.

Przykład



Położenie 27 krajów w układzie pierwszych zmiennych kanonicznych

Przykład



Powiększona lewa dolna ćwiartka wykresu

Literatura

1. Górecki, T., Krzyśko, M., Waszak, Ł., Wołyński, W. (2016): Selected statistical methods of data analysis for multivariate functional data, *Statistical Papers*, Published online.
2. Horváth, L., Kokoszka, P. (2012): *Inference for Functional Data with Applications*, Springer.
3. Ramsay, J.O., Silverman, B.W. (2005): *Functional Data Analysis, Second Edition*, Springer.
4. Deręgowski, K., Krzyśko, M., Waszak, Ł., Wołyński, W. (2017): Zastosowanie funkcjonalnej analizy kanonicznej w badaniu zależności między wydatkami konsumpcyjnymi w europejskich gospodarstwach domowych, *Wiadomości statystyczne* **672**, 19-37.

KONIEC