AN APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO INTERNATIONAL COMPARISON OF ECONOMIC ACTIVITIES

Alexei Verenikin
Lomonosov Moscow State University
Anna Verenikina
Peoples' Friendship University of Russia

The 12th International Days of Statistics and Economics, Prague, September 6-8, 2018

The traditional principal component approach

Suppose that there are *n* original variables that characterize economic performance of a *m* countries:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{pmatrix}$$

Denote by Σ the corresponding covariance matrix;
by $l_k = \{l_{ki}\}_{i=1}^{n}$ the normalized vector of *k*-th component loadings,
and by $z_{kj} = \sum_{i=1}^{n} l_{ki} x_{ij}$ the *k*-th component score for the *j*-th country.

The traditional principal component approach

Principal component loadings are determined as eigenvectors of the covariance matrix Σ.

The eigenvalues of the matrix Σ $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$ are equal to the variance of the corresponding principal component scores $Z_1, Z_2, \ldots, Z_n$ where $Z_k = \left(z_{k1}, \ldots, z_{km}\right)$ is the *k*-th principal component vector.

Total variance of principal component scores equals to the total variance of primary data, thus $\rho_k = \lambda_k \Big/ \sum_{k=1}^{n} \lambda_k$ is the share of data variance explained by the *k*-th principal component.

The traditional principal component approach

The first principal component score

$$z_{1j} = \sum_{i=1}^{n} l_{1i} x_{ij}$$

is known to be used as an aggregate economic indicator for the *j*-th country.

A modified principal component approach*

takes $$y_{1j} = \sum_{i=1}^{n} l_{1i}^2 x_{ij}$$

instead of $$z_{1j} = \sum_{i=1}^{n} l_{1i} x_{ij}$$

as an aggregate indicator of economic activity.

The approach is based on normalized property of component loadings: $$l_1 l_1^T = \sum_{i=1}^{n} l_{1i}^2 = 1$$

Unfortunately this approach fails to explain total variation of the data.

*Ayvazyan, S.A., Stepanov, V.S., & Kozlova, M. I. (2006). Measurement of Synthetic Categories of Quality of Life of the Regional Population and Identification of the Key Directions to Improve Regional Policy (in Russian). *Applied Econometrics, 2* (2), 18-84.

The generalized modified principal component (GMPC) approach

An aggregate measure of national economic activity – a weighted sum of all principal component scores:

$$I_j = \sum_{k=1}^{n} \rho_k y_{kj} = \sum_{k=1}^{n} \rho_k \sum_{i=1}^{n} l_{ki}^2 x_{ij} = \sum_{k=1}^{n} \left(\lambda_k \sum_{i=1}^{n} l_{ki}^2 x_{ij} \right) \Big/ \sum_{k=1}^{n} \lambda_k$$

The modified principal component score $y_{kj} = \sum_{i=1}^{n} l_{ki}^2 x_{ij}$

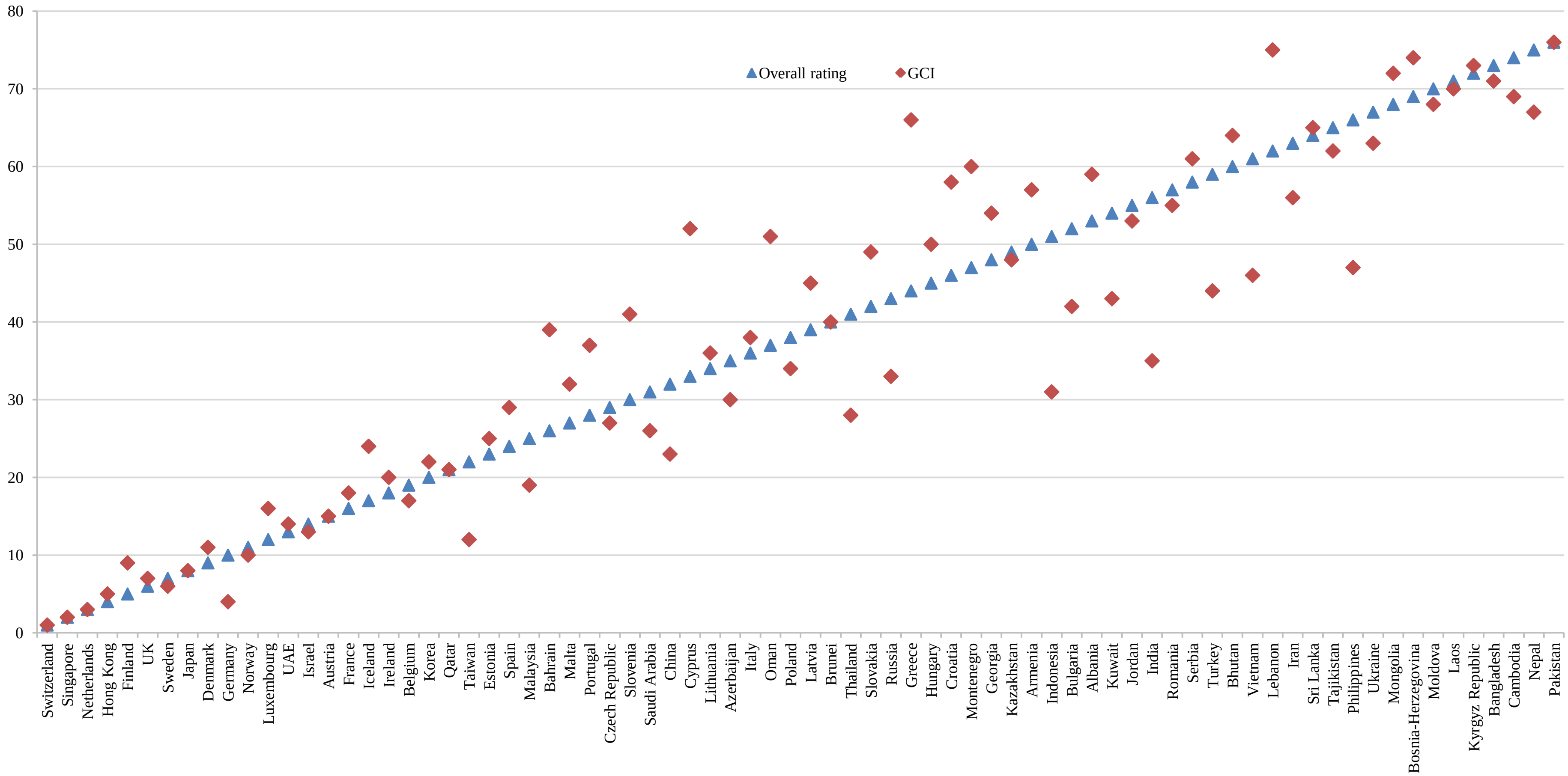
is an element of the aggregate index.

These elements are weighted by the corresponding shares of explained variance ρ_k so as to retain units of measure of the original variables.

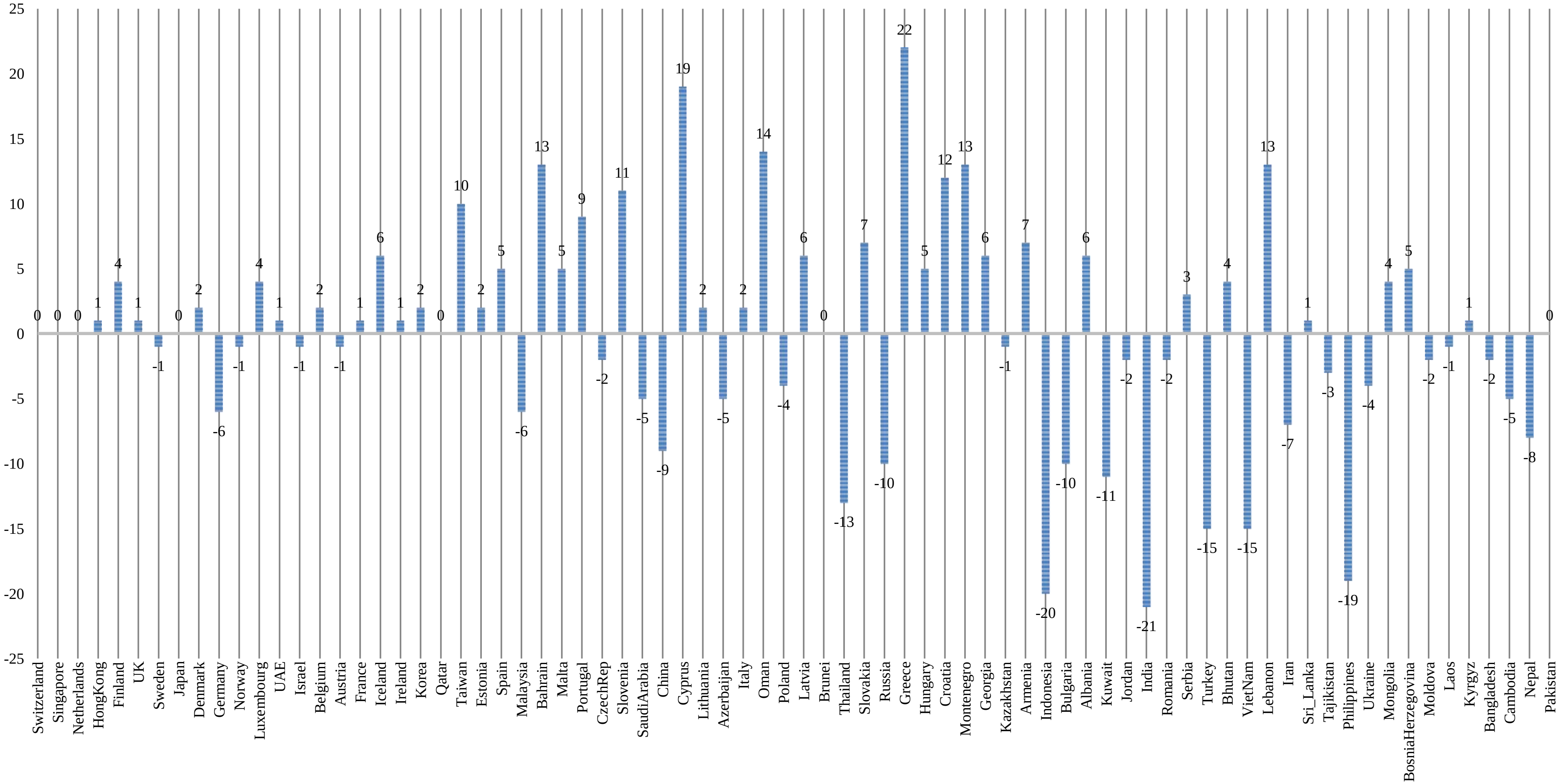
Global Competitiveness Index (GCI) as compared to GMPC ranking

Place	Country	+/- GCI	Place	Country	+/- GCI	Place	Country	+/- GCI
1	Switzerland	0	26	Bahrain	13	51	Indonesia	-20
2	Singapore	0	27	Malta	5	52	Bulgaria	-10
3	Netherlands	0	28	Portugal	9	53	Albania	6
4	HongKong	1	29	CzechRep	-2	54	Kuwait	-11
5	Finland	4	30	Slovenia	11	55	Jordan	-2
6	UK	1	31	SaudiArabia	-5	56	India	-21
7	Sweden	-1	32	China	-9	57	Romania	-2
8	Japan	0	33	Cyprus	19	58	Serbia	3
9	Denmark	2	34	Lithuania	2	59	Turkey	-15
10	Germany	-6	35	Azerbaijan	-5	60	Bhutan	4
11	Norway	-1	36	Italy	2	61	VietNam	-15
12	Luxembourg	4	37	Oman	14	62	Lebanon	13
13	UAE	1	38	Poland	-4	63	Iran	-7
14	Israel	-1	39	Latvia	6	64	Sri_Lanka	1
15	Belgium	2	40	Brunei	0	65	Tajikistan	-3
16	Austria	-1	41	Thailand	-13	66	Philippines	-19
17	France	1	42	Slovakia	7	67	Ukraine	-4
18	Iceland	6	43	Russia	-10	68	Mongolia	4
19	Ireland	1	44	Greece	22	69	Bosnia and Herzegovina	5
20	Korea	2	45	Hungary	5	70	Moldova	-2
21	Qatar	0	46	Croatia	12	71	Laos	-1
22	Taiwan	10	47	Montenegro	13	72	Kyrgyz	1
23	Estonia	+2	48	Georgia	6	73	Bangladesh	-2
24	Spain	5	49	Kazakhstan	-1	74	Cambodia	-5
25	Malaysia	-6	50	Armenia	7	75	Nepal	-8
						76	Pakistan	0

Global Competitiveness Index (GCI) as compared to GMPC ranking



Global Competitiveness Index (GCI) as compared to GMPC ranking



Ranking based on the first principal component (PC1) as compared to GMPC ranking

Country	Place	+/- Place PC1	Country	Place	+/- Place PC1
Switzerland	1	0	Estonia	23	0
Singapore	2	0	Spain	24	-3
Netherlands	3	0	Malaysia	25	+1
HongKong	4	0	Bahrain	26	+1
Finland	5	+1	Malta	27	+1
UK	6	-2	Portugal	28	0
Sweden	7	+2	CzechRep	29	-1
Japan	8	+2	Slovenia	30	-2
Denmark	9	-3	SaudiArabia	31	+2
Germany	10	-1	China	32	+1
Norway	11	+1	Cyprus	33	-4
Luxembourg	12	+3	Lithuania	34	-1
UAE	13	0	Azerbaijan	35	+2
Israel	14	-3	Italy	36	0
Belgium	15	+1	Oman	37	+3
Austria	16	+1	Poland	38	0
France	17	+1	Latvia	39	-1
Iceland	18	0	Brunei	40	-4
Ireland	19	-1	Thailand	41	+2
Korea	20	-2	Slovakia	42	0
Qatar	21	+2	Russia	43	-3
Taiwan	22	+1	…	…	…

The generalized modified principal component (GMPC) approach

Usually original data can be grouped into a number (θ) of subsets or pillars that reflect definite attributes of social and economic pattern.

M21 | *fx* =РАНГ.РВ(L21;$L$3:$L$78;0)

	A	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	
1		IER EDUCATION A		GOODS MARKET EFFI		LABOR MARKET EFFI		FINANCIAL MARKET		TECHNOLOGICAL RI		MARKET SIZE		BUSINESS SOPHISTI		R&
2			Rating		Rating		Rating		Rating		Rating		Rating		Rating	
3	Albania	49760504	37	0,54863774	39	0,34246657	63	0,1712124	55	0,2877611	58	0,0634549	66	0,1758732	58	0,
4	Armenia	22252277	59	0,6019609	17	0,3873353	41	0,1702452	57	0,2849419	59	0,0591675	70	0,1899019	44	0,
5	Austria	98561258	16	0,59572097	19	0,4220214	22	0,2209608	24	0,4338985	22	0,0957737	31	0,2800982	3	0,
6	Azerbaijan	22586567	58	0,58253674	24	0,44845082	12	0,1930635	36	0,3565171	41	0,0830512	49	0,2120853	32	0,
7	Bahrain	25516619	35	0,6402625	4	0,40258839	32	0,220988	23	0,4666087	16	0,0706477	58	0,2117892	33	0,
8	Bangladesh	55901214	74	0,50887145	61	0,32277777	69	0,1677138	60	0,1515961	73	0,0959421	29	0,1654131	68	0,
9	Belgium	38535599	4	0,55373403	35	0,41298623	24	0,2326999	17	0,4559754	18	0,101035	25	0,2672234	8	0,
10	Bhutan	92340848	70	0,48166495	67	0,42242992	21	0,1943564	35	0,2176441	66	0,0420615	76	0,1808667	53	0,
11	BosniaHerzegovina	99076095	68	0,4634615	74	0,31153073	71	0,1507968	69	0,3094134	54	0,0667309	62	0,1584508	71	0,
12	Brunei	24671628	56	0,59115615	23	0,40657662	30	0,170344	56	0,3484801	43	0,0623422	68	0,1703543	62	0,
13	Bulgaria	46679936	40	0,5470267	40	0,37888989	46	0,1796609	49	0,3572454	40	0,0830853	48	0,1780935	56	0,
14	Cambodia	50366666	76	0,4774427	71	0,40168502	33	0,1635609	62	0,1831303	69	0,0724958	57	0,1656722	67	0,
15	China	41045737	44	0,48146602	68	0,41264315	26	0,2080395	30	0,3259559	51	0,1412471	1	0,2179512	27	0,
16	Croatia	39210227	45	0,54095569	45	0,3409829	64	0,1677822	59	0,3681968	37	0,0767336	52	0,1685891	65	0,
17	Cyprus	51991147	36	0,60403029	15	0,41212059	27	0,147551	72	0,4261284	24	0,0627139	67	0,2003825	38	0,
18	CzechRep	73762204	27	0,53100589	50	0,39776761	35	0,2205743	25	0,4028985	29	0,0952002	32	0,2125276	30	0,
19	Denmark	32327046	5	0,62717052	8	0,4689847	6	0,2225915	21	0,5338816	1	0,089786	41	0,2656319	10	0,
20	Estonia	99915222	14	0,58183208	25	0,4352997	18	0,2263618	18	0,4715038	15	0,0674338	61	0,19946	40	0,
21	Finland	35666123	1	0,62236116	10	0,44033613	15	0,2638422	2	0,5013165	7	0,086538	45	0,2603312	15	0,
22	France	37711266	21	0,53971624	47	0,40131577	34	0,22465	19	0,483843	13	0,1178779	6	0,261085	13	0,
23	Georgia	07768373	64	0,61921082	11	0,39140664	38	0,169541	58	0,2811203	60	0,0652248	63	0,1740074	60	0,
24	Germany	99598836	15	0,57206392	29	0,45409913	11	0,2443089	12	0,4728564	14	0,1232573	3	0,277074	5	0,
25	Greece	95272352	18	0,51299066	59	0,33567125	66	0,1143463	76	0,3749134	35	0,0884105	43	0,1887681	46	0,
26	HongKong	00161711	13	0,63001133	6	0,49280303	4	0,2613737	3	0,4943622	11	0,1022121	22	0,2659522	9	0,

The generalized modified principal component (GMPC) approach

The aggregate index can be imagined as a sum of linear transforms of the pillars that constitute original data:

$$I_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{n}\lambda_k l_{ki}^2 x_{ij}}{\sum_{k=1}^{n}\lambda_k} = \frac{\sum_{i=1}^{n}\sum_{k=1}^{n}\lambda_k l_{ki}^2 x_{ij}}{\sum_{k=1}^{n}\lambda_k} = \frac{\sum_{\alpha=1}^{\theta}\sum_{i=n_{\alpha-1}+1}^{n_\alpha}\sum_{k=1}^{n}\lambda_k l_{ki}^2 x_{ij}}{\sum_{k=1}^{n}\lambda_k}$$

$$0 = n_0 < n_1 < \cdots < n_\theta = n$$

Thus, the resulting aggregate indicator can be decomposed into a sum of partial indicators that reflect the effect of definite pillars on general economic performance:

$$I_j = \sum_{\alpha=1}^{\theta} I_{j\alpha}, \text{ where } I_{j\alpha} = \sum_{i=n_{\alpha-1}+1}^{n_\alpha}\sum_{k=1}^{n}\lambda_k l_{ki}^2 x_{ij} \Big/ \sum_{k=1}^{n}\lambda_k$$

An application of principal component analysis to international comparison of economic activities

Alexei Verenikin verenikin@mail.ru
Anna Verenikina a.verenikina@yandex.ru

Thank You for attention!