



МГИМО МИД РФ



ЦЕНТР
ВОЕННО-ПОЛИТИЧЕСКИХ
ИССЛЕДОВАНИЙ

Сервис анализа данных для оценки динамики развития стран мира

Профессор МГИМО, д.и.н. Алексей Подберезкин
Зав. каф. МЭИТ МГИМО, к.ф.-м.н. Никита Артамонов
Зав. лаб. ЭВТ МЭИТ МГИМО Александр Немченко
Инж. лаб. ЭВТ МЭИТ МГИМО Михаил Звягинцев
Студентка 4 курса МЭО МГИМО Анна Белякова

Москва, август 2020 г.

Преамбула

Цель данного текста: презентация проекта, суть которого ясна из названия.

Здесь нет сложных формул, не приведены доказательства теорем Математической статистики, которые лежат в основе примененных в проекте методов. Мы ограничимся кратким пояснением терминов, встречающихся в интерфейсе пользователя, покажем на примерах возможности системы, наметим пути её дальнейшего развития.

Пример 1.

1. Аналитик имеет три таблицы (рис. 1-3) с данными по некоторому количеству стран и лет:

ВВП ППС										
Ед.изм.: млрд \$. Источник: http://svspb.net/danmark/vvp-stran.php . Скачать ZIP-файл источника										
#	Страны \ Годы	2018	2017	2016	2015	2014	2013	2012	2011	2010
1	Китай	25270.1	23189.9	21316.9	19756	18285.2	16724.8	15247.1	13864.9	12402.9
2	США	20494.1	19485.4	18707.2	18219.3	17521.8	16784.8	16197.1	15542.6	14992.1
3	Индия	10505.3	9596.8	8787.9	8036.3	7362.6	6727.3	6214.5	5782	5311.2
4	Япония	5594.5	5427.1	5225.1	5137.3	5021.5	4909.9	4730.6	4573.2	4484.8
5	Германия	4356.4	4199.3	4022.2	3894.7	3797.1	3647.2	3562.9	3471.8	3278.8
6	Россия	4213.4	4026.5	3887.4	3833.9	3890.6	3791.8	3660.5	3463.5	3229
7	Бразилия	3365.3	3254.9	3160.8	3233.7	3317.2	3239.2	3090.4	2974.9	2802.3
8	Индонезия	3494.7	3249.6	3035.2	2858.5	2696.8	2520.5	2346.6	2171.5	2003.5
9	Великобритания	3037.8	2929.9	2823.7	2744.1	2652.7	2528.9	2435.5	2355.6	2270
10	Франция	2962.8	2854.1	2741.6	2680.5	2623	2549.9	2491.6	2437	2336
11	Мексика	2569.8	2464	2369	2276.9	2181.1	2082.2	2018.9	1911.3	1806.1
12	Италия	2397.4	2324.1	2244.8	2196	2152.9	2110.5	2110.6	2130.9	2075.4
13	Турция	2292.5	2185.9	1996.5	1914	1785.1	1665.9	1509	1412.9	1245.6
14	Корея	2136.3	2034.9	1937.6	1862.1	1792.4	1702.2	1625.8	1559.4	1473.3
15	Испания	1864.4	1778.3	1694.6	1624.8	1551	1501.5	1501.2	1517.4	1501.3
16	Саудовская Аравия	1857.5	1777.2	1757.1	1709.5	1624.7	1538	1472.1	1370.2	1220.2

Рис.1

Военные расходы, млн \$ 2017 г., SIPRI										
Ед.изм.: млн \$ 2017 г.. Источник: https://www.sipri.org/databases/milex . Скачать ZIP-файл источника										
#	Страны \ Годы	2018	2017	2016	2015	2014	2013	2012	2011	2010
1	США	633565	605803	612889	616483	631513	673102	731086	775156	784835
2	Китай	239223	227829	215718	204202	191627	176476	161441	148656	137890
3	Саудовская Аравия	65843	70400	63141	88233	82732	70194	61251	54122	53396
4	Россия	64193	66527	82576	77023	71467	66682	63584	54877	51420
5	Индия	66578	64559	60311	54729	54214	51691	52075	52261	51759
6	Франция	59542	60417	58795	56672	54473	53679	53814	54568	55932
7	Великобритания	46883	46433	46903	46834	47921	49199	51237	53388	55407
8	Япония	45362	45387	45351	45627	44836	44363	44552	45095	44496
9	Германия	46192	45382	42918	41177	39951	40133	42591	41599	42824
10	Корея	41157	39171	38463	37523	36120	34753	33731	32893	32449
11	Бразилия	30769	29283	27491	28961	29648	29067	28910	28364	29283
12	Австралия	26836	27691	27741	25333	22982	21175	21360	22136	22447
13	Италия	26082	26448	25709	22695	23701	25674	26696	28880	29491
14	Канада	21352	21343	18436	18194	15836	15582	16857	17727	17148
15	Турция	22088	17824	16630	14355	13917	13838	13504	13214	13087
16	Испания	17039	16044	14497	15634	14708	14730	16884	16694	18097

Рис. 2

Публикации, Scopus										
Ед.изм.: количество публикаций. Источник: https://www.scimagojr.com/countryrank.php . Скачать ZIP-файл источника										
#	Страны \ Годы	2018	2017	2016	2015	2014	2013	2012	2011	2010
1	США	683003	683590	669204	669588	664528	665898	659053	621884	600345
2	Китай	599386	534879	496397	461547	491686	456895	415776	395431	344420
3	Великобритания	211710	209593	204237	200087	195048	198445	191251	177393	172923
4	Германия	180608	180712	179539	174519	173431	169607	166424	156045	150645
5	Индия	171356	154619	152760	140499	132805	117127	108751	98238	80802
6	Япония	131198	130823	130441	127862	132356	136822	135130	132152	129700
7	Франция	120908	123530	122519	121176	121185	122003	116324	111108	107610
8	Италия	119405	118116	114396	112195	108628	105467	98959	89990	86301
9	Канада	111561	108252	106434	104622	104481	102627	100535	93702	90926
10	Австралия	106228	101697	98802	95345	92969	89244	81844	75038	69794
11	Испания	96517	94657	92217	89936	90769	87819	85783	79478	73359
12	Россия	99099	89492	81689	67834	58244	49703	44724	43604	40390
13	Бразилия	81742	78299	74624	70267	68328	64711	61422	55487	50974
14	Нидерланды	63193	61463	59984	58875	58783	58161	57122	51964	50475
15	Швейцария	48317	48452	46424	45472	44316	43300	41340	38044	35830
16	Польша	49488	48239	46903	43946	41986	39497	36647	33203	31474
17	Турция	45582	44584	47138	44529	41400	40302	36766	34926	33325

Рис. 3

Аналитик хочет проанализировать соотношение сил на мировой шахматной доске для 10-12 стран, включая Россию, Китай и США.

Вначале ему придется освоить информацию, содержащуюся в 300 ячейках этих таблиц. Затем оценить относительный вес каждого из параметров, просчитать их соотношение, сделать несколько выводов по каждому из параметров и затем попытаться эти выводы объединить в один простой рейтинг.

Потратив некоторое количество сил и времени он сможет сделать вывод, что Россия уступает США:

- в экономической мощи в 5 раз;
- в военной мощи в 10 раз;
- в интеллектуальной мощи в 7 раз.

Заметим, что таблиц и ячеек в них может быть на несколько порядков больше, что сделает «ручной» анализ невозможным.

Использование электронных таблиц может помочь, но это потребует неоправданно много времени, да и масштабируются они только до определенных пределов.

Итак: Цель проекта — облегчить анализ данных.

2. Если открыть на сайте <http://eurasian-defence.ru/richdb> вот эту табличку:

Главные компоненты по 2018 году									
#	Страны \ Группы	Военно-экономическая сила		Человеческий капитал		Институты развития ЧК		ГК по всей базе	
		PCA1	PCA2	PCA1	PCA2	PCA1	PCA2	PCA1	PCA2
1	Китай	8.9194	-1.0846	11.6707	4.0098	-0.1346	1.0552	10.3254	1.5781
2	США	9.4416	0.3524	9.211	-3.5691	-1.6414	0.357	9.6385	-0.7721
3	Индия	2.3114	-0.0518	1.2339	-0.7529	1.5732	0.8582	1.4783	2.7024
4	Япония	1.0614	-0.8754	1.8481	0.7711	-1.8721	0.6558	1.2937	-1.6684
5	Германия	0.6623	-0.6176	2.0552	-1.0855	-1.9298	0.3868	1.187	-1.922
6	Великобритания	0.3168	-0.1719	2.4074	-1.5889	-0.9508	-0.3335	1.0218	-0.9239
7	Франция	0.5586	0.1401	1.1073	-0.7524	-1.0486	0.1589	0.5895	-0.6263
8	Россия	1.1708	1.1004	0.4672	-0.2209	0.1578	-0.1356	0.5243	1.3
9	Италия	0.0278	-0.4896	1.1366	-0.8226	-0.4975	-0.2833	0.1822	-0.5
10	Канада	0.1246	-0.511	0.9721	-0.7521	-0.8666	-0.4451	0.1767	-0.9172
11	Испания	-0.123	-0.4817	0.7386	-0.6264	-0.465	-0.4519	-0.1174	-0.4889
12	Израиль	-0.3291	1.5071	-0.2166	-0.0318	-2.8578	1.764	-0.1766	-1.5489
13	Нидерланды	-0.3981	-0.5161	0.4319	-0.445	-1.3188	-0.1562	-0.3183	-1.3998
14	Швеция	-0.4779	-0.6423	0.1105	-0.2387	-2.0682	0.5525	-0.3863	-2.0956

Рис. 4

то по колонке PCA1 можно сразу сказать: по сумме всех показателей Россия уступает США в 18, а Китаю почти в 20 раз. Там же можно найти несколько графиков очень наглядно это иллюстрирующих (см. раздел **Графики** ниже).

Конец примера 1.

Этот пример показывает, как можно получить более точный и наглядный результат, потратив на порядок меньше энергии/времени.

*Примечание: Термины **Главные компоненты** и **РСА**, используемые в таблице на рис. 4 раскрываются ниже на стр. 15-17.*

О проекте

Концепцию групп данных, их влияние на состояние обществ и государств, их соотношение, определяющее рост или падение влияния страны/коалиции на мировой арене впервые разработал профессор МГИМО, директор Центра военно-политических исследований Алексей Подберезкин.

Все параметры, с той или иной точностью, всесторонне характеризующие страны нашей планеты и/или их союзы, можно объединить в группы с условными названиями: мощь страны, качество человеческого материала, факторы формирующие это качество.

Мы назвали эти группы:

1. Военно-экономическая сила
2. Человеческий капитал
3. Институты развития ЧК

Конечно этот список не исчерпывает все параметры страны, но, на наш взгляд, именно они составляют остов, скелет на котором держится все остальное. Именно они обеспечивают суверенитет страны и благосостояние граждан.

Данная модель имеет ограничения и другого рода — она не учитывает отрицательные факторы действующие внутри государств и внешние воздействия на них, а также явления катастрофического характера непреодолимой силы длиннопериодического, редкого и редчайшего характера.

Ограничившись перечисленными параметрами мы сформулировали гипотезу, согласно которой определяющей группой факторов является группа № 3. Именно от этих факторов, как мы полагаем, зависит какого качества человеческим материалом будет располагать страна и, главное, какова будет динамика изменений этого качества. Соответственно, экономическая и военная мощь являются производными от фактора №2.

Цель нашего исследования: доказать или опровергнуть эту гипотезу.

Одним из инструментов этого исследования является данный сайт.

Сбор и анализ данных в процессе научно-исследовательской работы в гуманитарных областях всегда было трудоемким делом. Универсальные программные средства, такие как электронные таблицы, требуют серьезных усилий, чтобы добиться результатов, которые часто не удовлетворяют исследователей (универсальность имеет обратную сторону: ограниченность и отсутствие узко-специальных функций).

Мы решили попытаться создать такой сервис, который позволил бы решить часть этих проблем. Название «Сервис Ричардсона» выбрано в честь английского математика Льюиса Ф. Ричардсона, предложившего первую математическую модель гонки вооружений.

Сервис имеет следующую технологическую архитектуру:

- Бэкенд (back-end):
 - Сервер с операционной системой Linux/Debian.
 - База данных на основе СУБД Postgresql.
 - Язык программирования Python.
- Фронтенд (front-end):
 - Web-сервер Apache.
 - Язык программирования PHP.

Сервис доступен в интернете по адресу: <http://eurasian-defence.ru/richdb> и включает следующие функции:

- Просмотр базы данных (далее везде БД) с возможностью фильтрации по источникам, странам и годам, а также по группам источников и коалициям стран.
- Нормализованные данные (см. ниже).
- Данные по всем источникам и годам по группам источников.
- Главные компоненты (см. ниже подробное описание метода ГК).
- Сортировка и транспонирование таблиц.
- Более 1300 иллюстраций: диаграммы, графики, 3d-модели, в т.ч. интерактивные и анимированные.
- Дамп базы данных имеет размер 50 Мб, БД содержит более 100 тысяч основных единиц хранения (в таблице value) и вдвое больше дополнительных.

Описание БД

БД содержит таблицы, имеющие следующую структуру:

- **Справочники:**
 - **source** — Источники (13)
 - sid — идентификатор
 - sname — наименование
 - slink — ссылка на источник
 - unit — единица измерений
 - file — ссылка на файл-источник на сайте
 - status — статус
 - sgid — id группы
 - **source_group** — Группы источников
 - cggid — идентификатор
 - sgroupname — наименование
 - **country** — Страны (196)
 - cid — идентификатор
 - cname — наименование
 - cname_en — наименование на английском языке
 - cname_en2 — наименование на английском языке 2-й вариант
 - cgid — id группы
 - **country_group** — Группы стран (коалиции)
 - cggid — идентификатор
 - cgroupname — наименование
 - **year** — Годы (39)
 - yid — идентификатор
 - yname — наименование
- **Данные**
 - **value** - Сырые данные (32459)
 - vid — идентификатор данной таблицы
 - vsid — идентификатор источника
 - vyid — идентификатор года
 - vcid — идентификатор страны
 - val — значение
 - interpolat — признак интерполированного значения

- **norm** - Нормализованные данные (31466)
 - **nid** — идентификатор данной таблицы
 - **nsid** — идентификатор источника
 - **ncid** — идентификатор страны
 - **nyid** — идентификатор года
 - **nval** — значение
- **pca** — Главные компоненты (3634)
 - **pid** — идентификатор данной таблицы
 - **rpc** — номер компоненты (1-я ГК или 2-ГК)
 - **pcid** — идентификатор страны
 - **pyid** — идентификатор года
 - **pval** — значение
 - **pgsid** — id группы

Также в БД есть вспомогательные и временные таблицы, используемые скриптами конвертации данных из файлов источников.

Подробнее см. Журнал инженерных работ: <http://eurasian-defence.ru/richdb/src/readmy.html>

Функционал

Большинство ссылок и надписей, смысл которых неочевиден, снабжены всплывающими, при наведении на них курсора мыши, подсказками.

Все данные, представленные на сайте проекта <http://eurasian-defence.ru/richdb/>, имеют вид таблиц (двухмерных матриц) или графиков, диаграмм, 3d-моделей.

Таблицы имеют два измерения:

- по горизонтали это, как правило, годы убывающие слева направо (с таким расчетом, чтобы самые свежие данные были в пределах первого экрана и для их просмотра не требовался горизонтальный скроллинг);
- по вертикали это, чаще всего, названия стран расположенных в алфавитном порядке или отсортированные по первой числовой колонке (иногда по колонке базового 2017 года).

В шапке таблиц, содержащих дифференцированные по группам источников данные, соответствующие разделы выделены цветами, внизу таблиц есть расшифровка этих цветов.

Коалиции стран тоже имеют цветовую индексацию.

Страны не входящие в коалиции и ячейки шапки, не относящиеся к группам источников, имеют общий фон сайта. (На рис. 5 приведен пример раскраса таблицы. Состав параметров меняется — база постоянно пополняется и модернизируется.)

Пример раскраса шапки и части строк таблицы

Все источники по 2017 году												
#	Страны / Источники	Военно-экономическая сила				Человеческий капитал				Институты развития ЧК		
		ВВП ППС (млрд \$)	Военные расходы, млн \$ 2017 г., SIPRI (млн \$ 2017 г.)	Военные расходы, % от ВВП, SIPRI (% от ВВП)	Выработка электроэнергии (млн. кв.час)	ИРЧП, Human Development (ИРЧП)	Патенты (количество патентов)	Публикации, Scopus (количество публикаций)	Цитирование публикаций, Scopus (количество цитирований)	Число "мозговых центров", University of Pennsylvania TTCSP (шт.)	Число супер компьютеров в TOP500 (шт.)	Число университетов в Top500 ARWU (Шанхай) (шт.)
1	Китай	23189.9	227829	1.9	6604.5	0.753	1245709	534879	1642037	512	202	57
2	США	19485.4	605803	3.1	4302.5	0.919	293904	683590	2326280	1872	143	135
3	Индия	9596.8	64559	2.5	1470.3	0.643	14961	154619	310564	444	4	1
4	Япония	5427.1	45387	0.9	1050.1	0.913	260292	130823	344505	116	35	17
5	Германия	4199.3	45382	1.2	653.7	0.938	47785	180712	659344	214	21	37
6	Россия	4026.5	66527	4.2	1089.6	0.822	22777	89492	148470	103	3	3
7	Бразилия	3254.9	29283	1.4	588	0.76	5480	78299	176387	93		6
8	Индонезия	3249.6	8178	0.8	254.6	0.704	2271	20459	27746	26		
9	Великобритания	2929.9	46433	1.8	338.6	0.919	13301	209593	784331	311	15	38
10	Франция	2854.1	60417	2.3	554.1	0.89	14415	123530	428072	197	18	20
11	Мексика	2464	5781	0.5	329.1	0.765	1334	24318	57619	74	1	1
12	Италия	2324.1	26448	1.4	295.8	0.881	8643	118116	435700	103	6	16
13	Турция	2185.9	17824	2.1	297.3	0.805	8175	44584	88983	46		1
14	Корея	2034.9	39171	2.6	576.4	0.904				53	5	12
15	Испания	1778.3	16044	1.2	275.6	0.891	2167	94657	325432	63	1	11
16	Саудовская Аравия	1777.2	70400	10.3	382.1	0.856	909	21650	98575	8	4	4
17	Канада	1764	21343	1.3	662.7	0.921	4053	108252	404617	100	5	19

Рис. 5

Легенда цветowych индексов

Цветовые индексы стран	Россия и союзники	Китай и союзники	США и союзники	Исламский мир
Цветовые индексы групп источников	Военно-экономическая сила	Человеческий капитал	Институты развития ЧК	
Числа выделенные этим цветом , рассчитаны методом линейной интерполяции .				

Рис. 6

Все графики, построенные на основе базы данных, сведены в два раздела сайта:

- [Графики по странам](#)
- [3D-модели и диаграммы](#)

Графики по странам

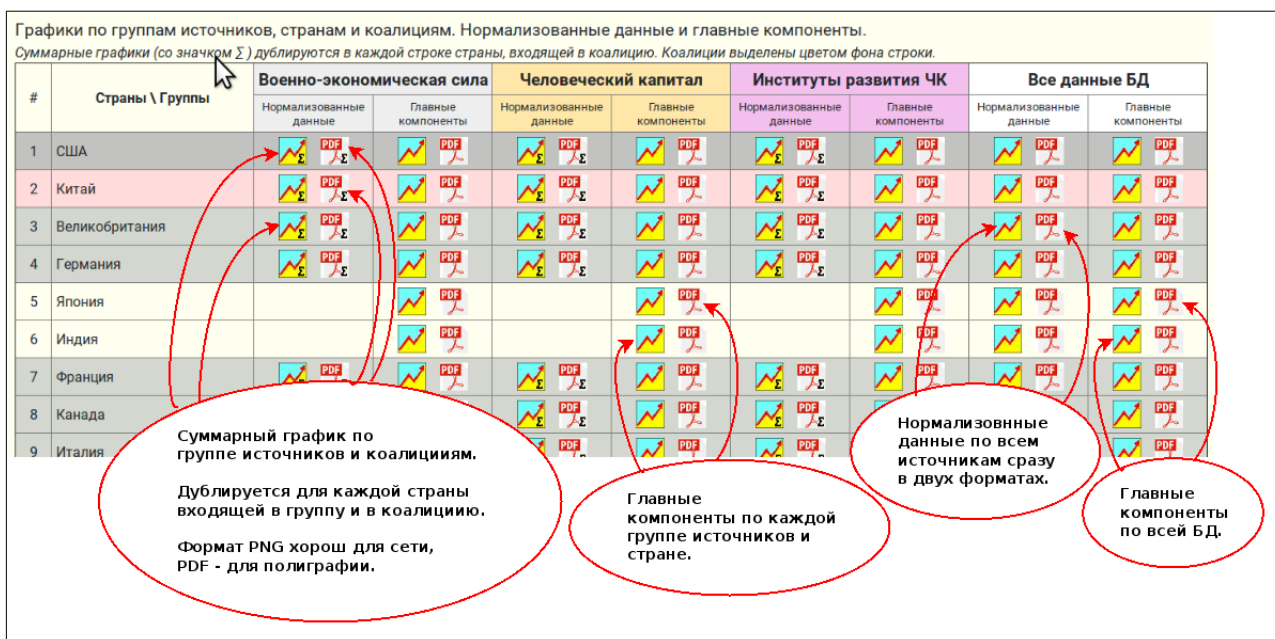


Рис. 7

3D-модели и диаграммы



Рис. 8

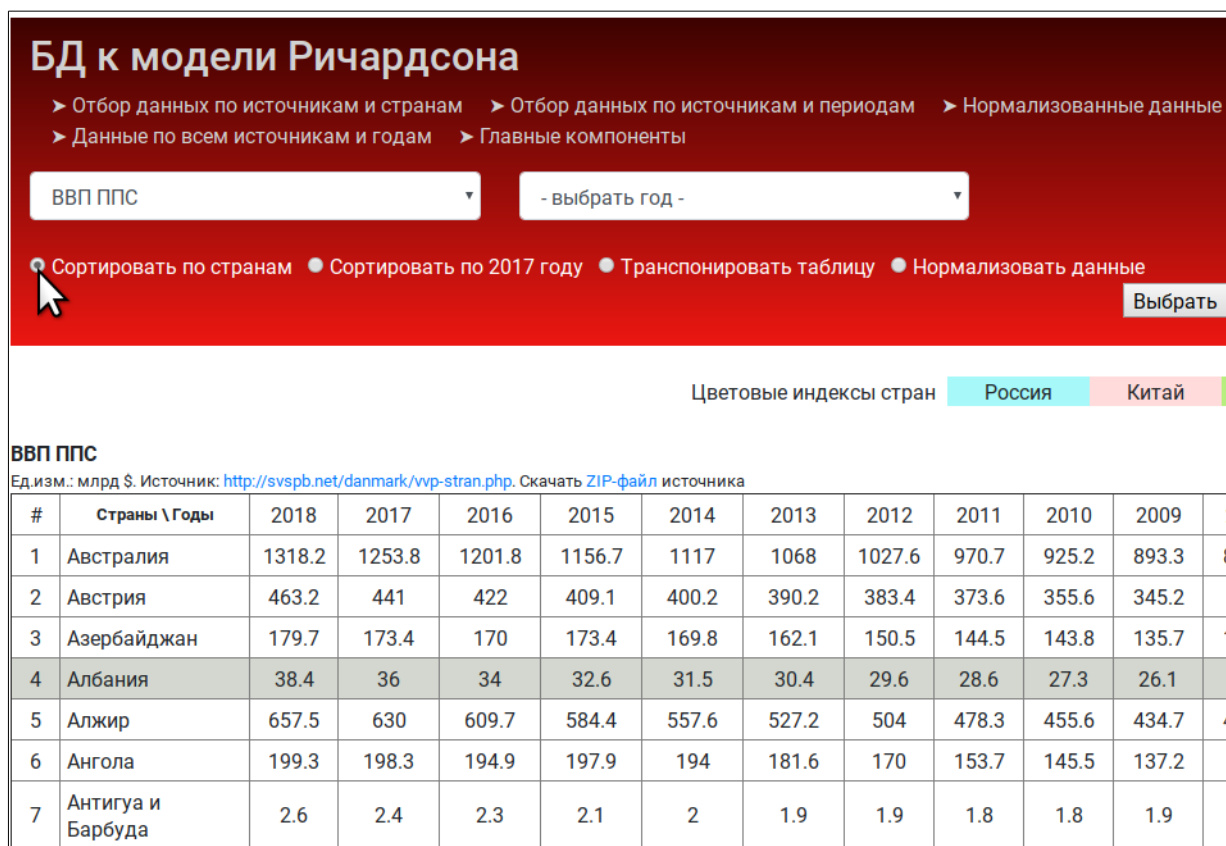
Основные понятия и их реализация в интерфейсе «Сервис Ричардсона»

Описания функций и терминов приводятся по принципу «от простого к сложному».

0. Описывать пункты меню и селекторы выбора источников, стран и лет нет смысла — это проще попробовать, чем читать инструкцию.

1. **Сортировка по странам**, т. е. по алфавиту. Это сортировка по умолчанию.

На рис. 9 белой стрелкой показана выбранная радио-кнопка.



БД к модели Ричардсона

▶ Отбор данных по источникам и странам ▶ Отбор данных по источникам и периодам ▶ Нормализованные данные
▶ Данные по всем источникам и годам ▶ Главные компоненты

ВВП ППС - выбрать год -

Сортировать по странам Сортировать по 2017 году Транспонировать таблицу Нормализовать данные

Выбрать

Цветовые индексы стран Россия Китай

ВВП ППС
Ед.изм.: млрд \$. Источник: <http://svspb.net/danmark/vvp-stran.php>. Скачать ZIP-файл источника

#	Страны \ Годы	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008
1	Австралия	1318.2	1253.8	1201.8	1156.7	1117	1068	1027.6	970.7	925.2	893.3	868.2
2	Австрия	463.2	441	422	409.1	400.2	390.2	383.4	373.6	355.6	345.2	335.7
3	Азербайджан	179.7	173.4	170	173.4	169.8	162.1	150.5	144.5	143.8	135.7	127.6
4	Албания	38.4	36	34	32.6	31.5	30.4	29.6	28.6	27.3	26.1	25.1
5	Алжир	657.5	630	609.7	584.4	557.6	527.2	504	478.3	455.6	434.7	413.8
6	Ангола	199.3	198.3	194.9	197.9	194	181.6	170	153.7	145.5	137.2	128.9
7	Антигуа и Барбуда	2.6	2.4	2.3	2.1	2	1.9	1.9	1.8	1.8	1.9	1.9

Рис. 9

2. **Сортировка по 2017 году**.

2017 год взят за основу, т. к. он является базовым в некоторых источниках, например, SIPRI приводит военные расходы к ценам этого года.

(Для этого варианта иллюстрации нет — все очевидно.)

3. **Транспонирование таблицы** — трансформация таблицы таким образом, что имена колонок становятся названиями строк, а имена строк — заголовками колонок.

На рис. 10 белой стрелкой показана радио-кнопка для выбора этого действия.

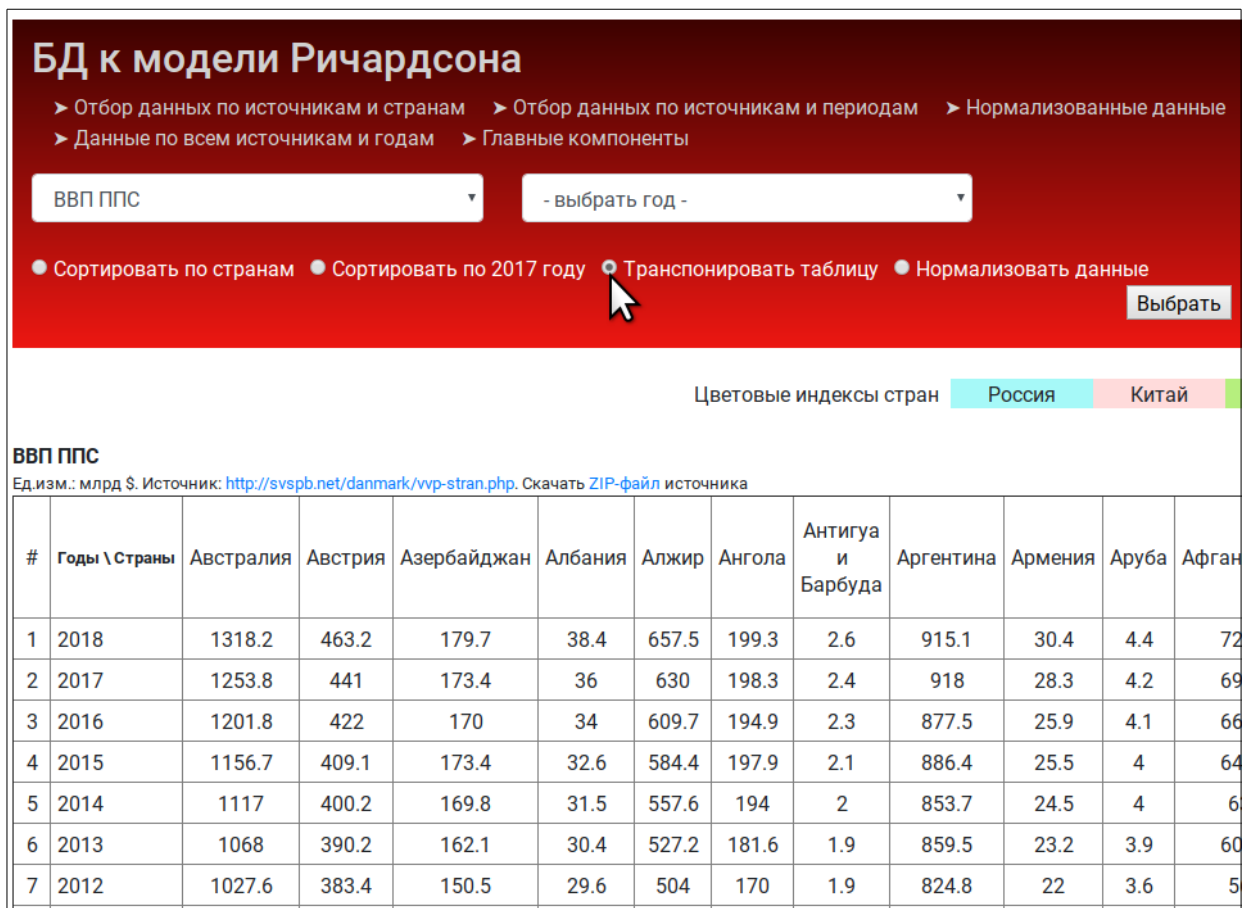


Рис. 10

4. Нормализация данных: довольно сложное преобразование (формула приведена на рис. 11)

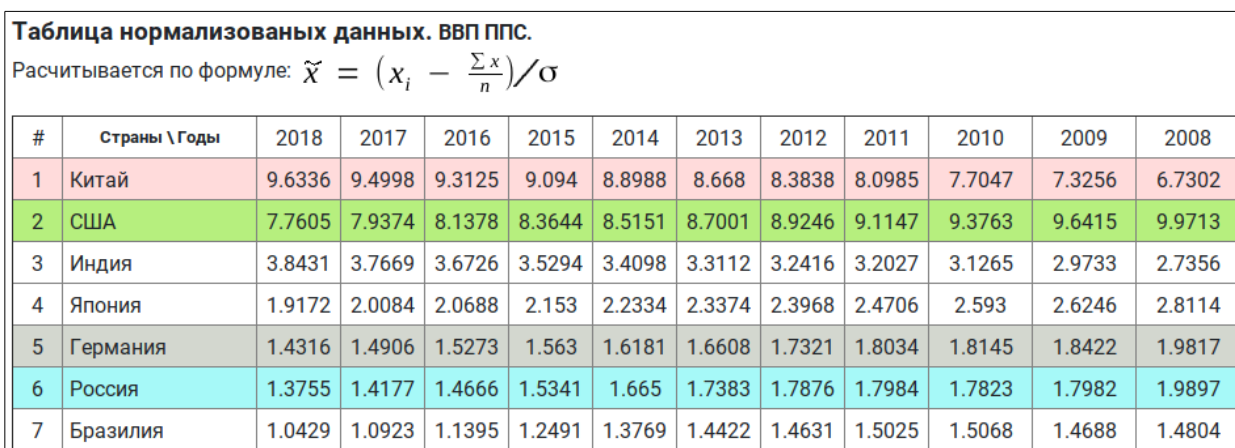


Рис. 11

Здесь данные лишены единиц измерений, таких как доллары или проценты, и приведены к единой шкале, причем в данной матрице значения принимают, как положительные, так и отрицательные величины.

При наличии качественных исходных данных матрица нормализованных данных после суммировании по строкам и колонкам в угол дает точный ноль. В реальности этого достичь сложно — всегда есть искажения, пропуски, шум. Здесь (см. фрагмент левого нижнего угла матрицы на рис. 12) это почти удалось. Только нули со знаками минус показывают, что где-то в 10-м знаке после запятой число отличается от чистого нуля.

																	-2.215
-0.2911	-0.293	-0.2954	-0.2979	-0.2972	-0.294	-0.2945	-0.2951	-0.2934	-0.2948	-0.2953	-0.2974	-0.3019	-0.306	-0.303	-0.3054	-11.2497	
3.9756	3.9384	4.0312	4.1183	4.0993	3.9861	3.836	3.795	3.7095	3.6546	3.6605	3.606	3.6647	3.6733	3.5133	3.4403	127.4326	
0	0	-0	-0	0	0	0	0	0	0	-0	0	-0	0	-0	0	0	

Рис. 12

Если таблица отсортирована по возрастанию числовых данных первой колонки, то вверху будут мировые лидеры с положительными индексами, в нижней части таблицы — аутсайдеры с отрицательными значениями (максимальными по абсолютной величине), а по середине близкие к нулю (см. фрагмент центра матрицы на рис. 13).

31	Колумбия	0.0151	0.0187	0.0259	0.0298	0.0313	0.0278	0.0244	0.0228	0.0111	0.0123	0.0074	0.0063	0.0021
32	Объединенные Арабские Эмираты	0.0069	0.013	0.0217	0.0228	0.0179	0.0154	0.0106	0.0072	-0.0029	0.0063	0.0224	0.0214	0.0281
33	Бангладеш	0.0218	0.0111	0.0006	-0.0099	-0.0192	-0.0268	-0.0338	-0.0415	-0.0495	-0.0529	-0.0672	-0.0728	-0.0749
34	Ирак	-0.0129	-0.0044	0.0108	-0.016	-0.0141	-0.0066	-0.0181	-0.0445	-0.0549	-0.0591	-0.0688	-0.0802	-0.072
35	Вьетнам	0.0016	-0.0069	-0.015	-0.0227	-0.0313	-0.0379	-0.0433	-0.0485	-0.0557	-0.06	-0.0742	-0.0799	-0.0831
36	Алжир	-0.0191	-0.0148	-0.0088	-0.0088	-0.0099	-0.0109	-0.0094	-0.0097	-0.0081	-0.005	-0.0111	-0.0094	-0.0036

Рис. 13

С точки зрения математики никакой разницы между положительными, отрицательными числами и нулем нет. Сильно упрощая можно сказать, что они просто показывают отклонение от некоторой средней величины (это конечно не среднее арифметическое — точная формула приведена в верхней части рис. 11.)

На основе таких матриц уже можно проводить сравнительный анализ между странами, коалициями, как по отдельным параметрам (источникам, группам источников), так и по всем данным базы сразу.

Для этого очень полезно построить графики нормализованных данных (см. раздел Графики по странам).

5. Метод главных компонент (далее МГК или по английски PCA — principal component analysis) один из самым популярных методов снижения размерности матриц.

Это значит, что вместо 10, 100 или 1000 параметров характеризующих событие, объект или страну, можно получить 1, 2 или в крайнем случае 3 показателя, которые включают в себя не менее 80% информации, содержащейся во всех начальных, сырых данных.

Часто один из эти, оставшихся после применения к матрице МГК, параметров (их называют 1-я, 2-я, 3-я главные компоненты) сильно больше остальных по абсолютной величине.

Это провоцирует исследователя отбросить остальные, но делать это нельзя, ибо **все** главные компоненты ортогональны друг-другу, т.е информация, содержащаяся в одном, отсутствует в остальных. Не учитывать какую-то из компонент, значит потерять важную часть информации характеризующую страну.

Надо понимать, что Главные компоненты напрямую не связаны с физическим смыслом тех параметров, на базе которых они вычислены (при дальнейшем анализе это иногда удастся определить). По сути они являются синтетическим индексом стран, который показывает, математически строго вычисленное их соотношение. Не менее 80% информации, которая содержалась в первичных параметрах, по группам или сразу во всех, сохраняется в Главных компонентах.

Рассмотрим схематичное графическое изображение матрицы абстрактного объекта на рис. 14.

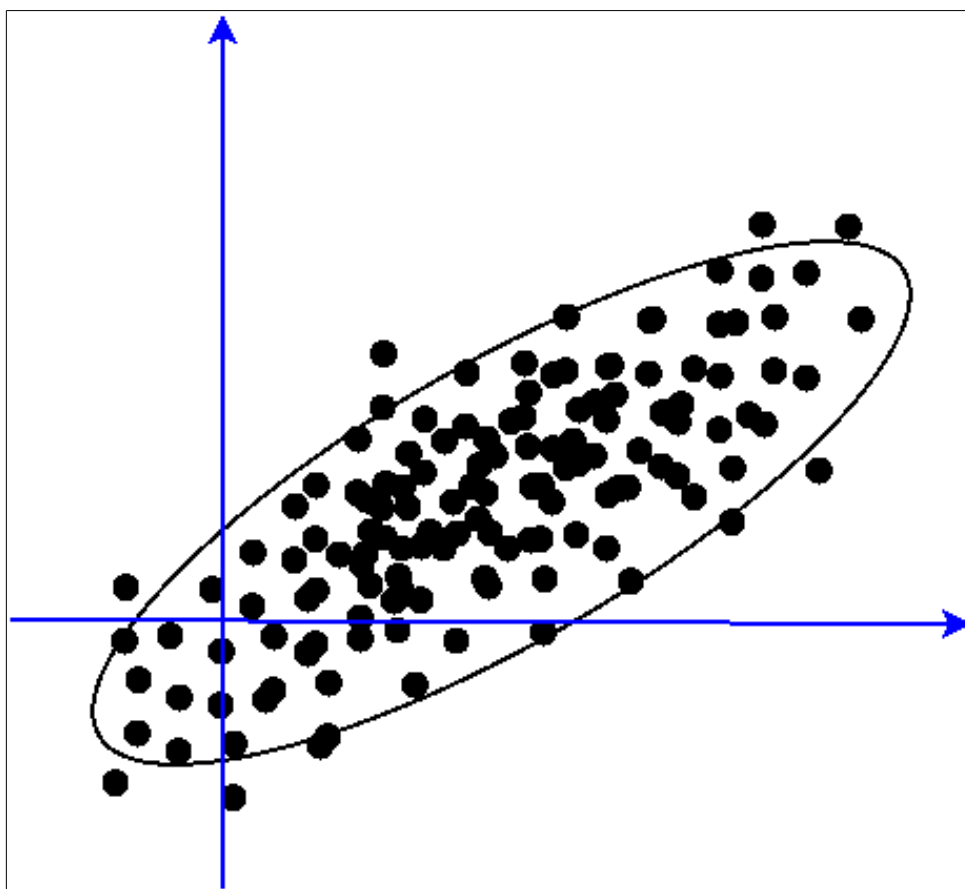


Рис. 14

Здесь:

- синие стрелки — оси координат,
- в которых черные точки — суть значения параметров нормализованной матрицы,
- эллипс — область, внутри которой поместились не менее 80% точек.

Если центр эллипса, вместе с точками внутри него, перенести в начало координат, игнорируя точки оставшиеся за пределами эллипса, координаты точек, конечно изменятся, но количество информации, которые они содержат сохранится в полном объеме, т. е. эта операция легитимна

Если теперь нанести на график большой и малый радиусы эллипса, мы получим два вектора: первая (большая) и вторая (меньшая) главные компоненты всей совокупности данных (см. рис. 15).

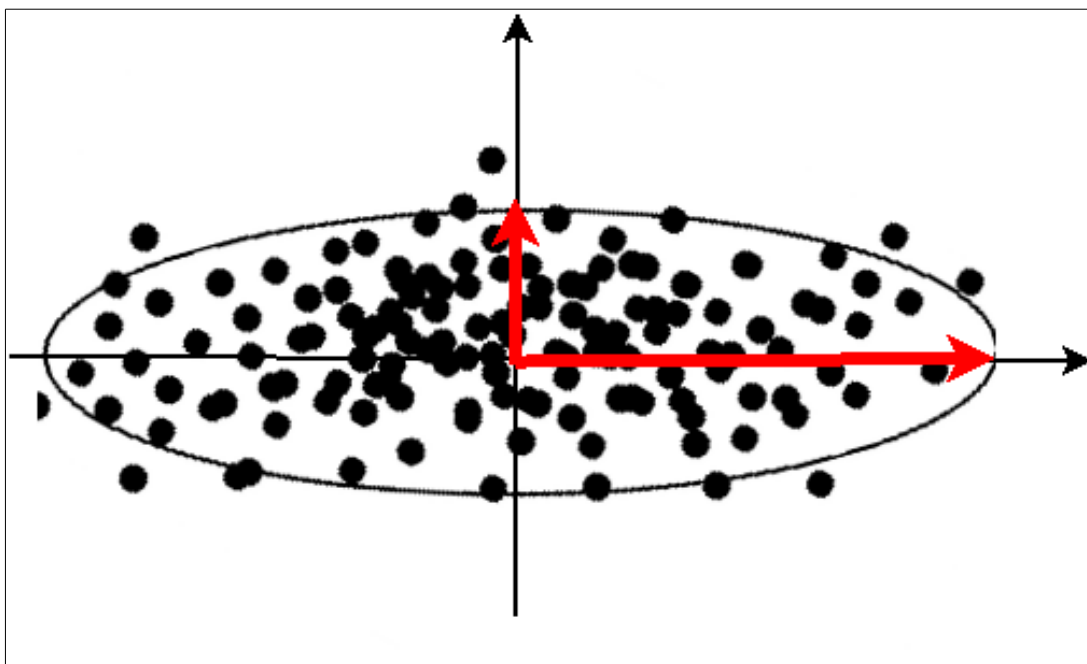


Рис. 15

Если посмотреть на этот рисунок, можно заметить, что только часть точек получит положительные координаты, остальные будут иметь отрицательные значения.

Подобный же эффект мы наблюдали при нормализации данных — это нормально.

Надо сказать, что все сложные матричные и векторные преобразования, которые необходимо произвести над сырыми данными, чтобы добиться конечного результата: нормализованных данных и главных компонент, доступны в виде готовых функций во многих библиотеках с открытым кодом на разных языках программирования. Мы использовали библиотеки языков Python и R.

Графики

Сейчас основная собранная база данных содержит более 90 тыс. чисел характеризующих 196 стран по 23 показателям за 39 лет, дампы базы имеет размер 49 Мб.

Человеческий мозг не может объективно проанализировать такой объем информации. Даже используя сведенные в таблицы и снабженные удобным интерфейсом для просмотра данных в разных разрезах, оценить такой объем чисел не легко.

Графическое представление полученных данных имеет два преимущества:

- плотность информации на единицу площади во много раз больше чем при табличной форме,
- сравнимость показателей на картинках мгновенна и очевидна.

Это обусловлено физиологией. Зрительные образы более привычны для мозга — мы мгновенно узнаем по фотографии лицо человека и легко отличаем одно от другого, а чтобы сделать то же самое по словесному описанию, необходима длительная тренировка, серьезное усилие, затрата энергии и времени.

Пример 1, приведенный в преамбуле, теперь можно продолжить с использованием инструментов визуализации данных, предоставляемых сервисом <http://eurasian-defence.ru/richdb/>.

Пример 2.

Откроем графики США, Китая и России для нормализованных данных.

Графики по группам источников, странам и коалициям. Нормализованные данные и главные компоненты.

#	Страны \ Группы	Военно-экономическая сила		Человеческий капитал		Институты развития ЧК		Все данные БД	
		Нормализованные данные	Главные компоненты	Нормализованные данные	Главные компоненты	Нормализованные данные	Главные компоненты	Нормализованные данные	Главные компоненты
1	США								
2	Китай								
3	Великобритания								
4	Германия								
5	Япония								
6	Индия								
7	Франция								
8	Канада								
9	Италия								
10	Россия								
11	Австралия								

Red circles with numbers 1, 2, and 3 are placed over the 'Институты развития ЧК' column for China, India, and Russia respectively. Red arrows point from these circles to the 'Все данные БД' column for the same countries.

Рис. 16

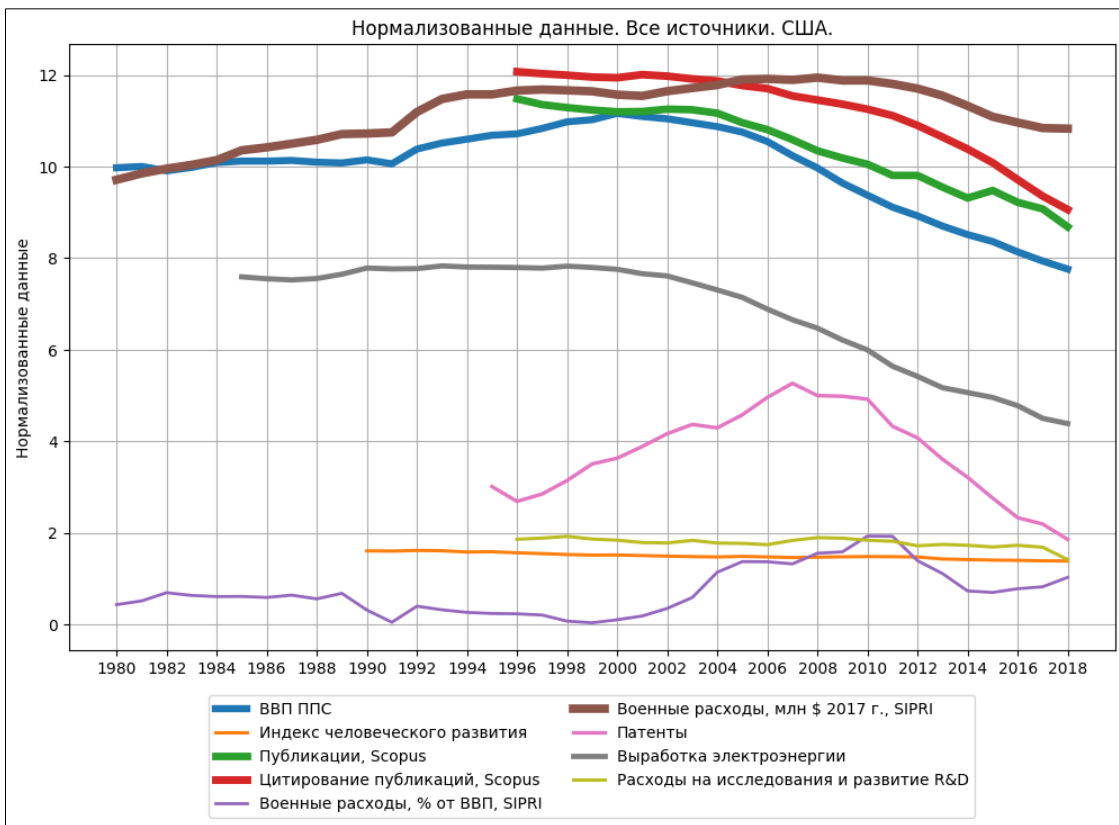


Рис. 17

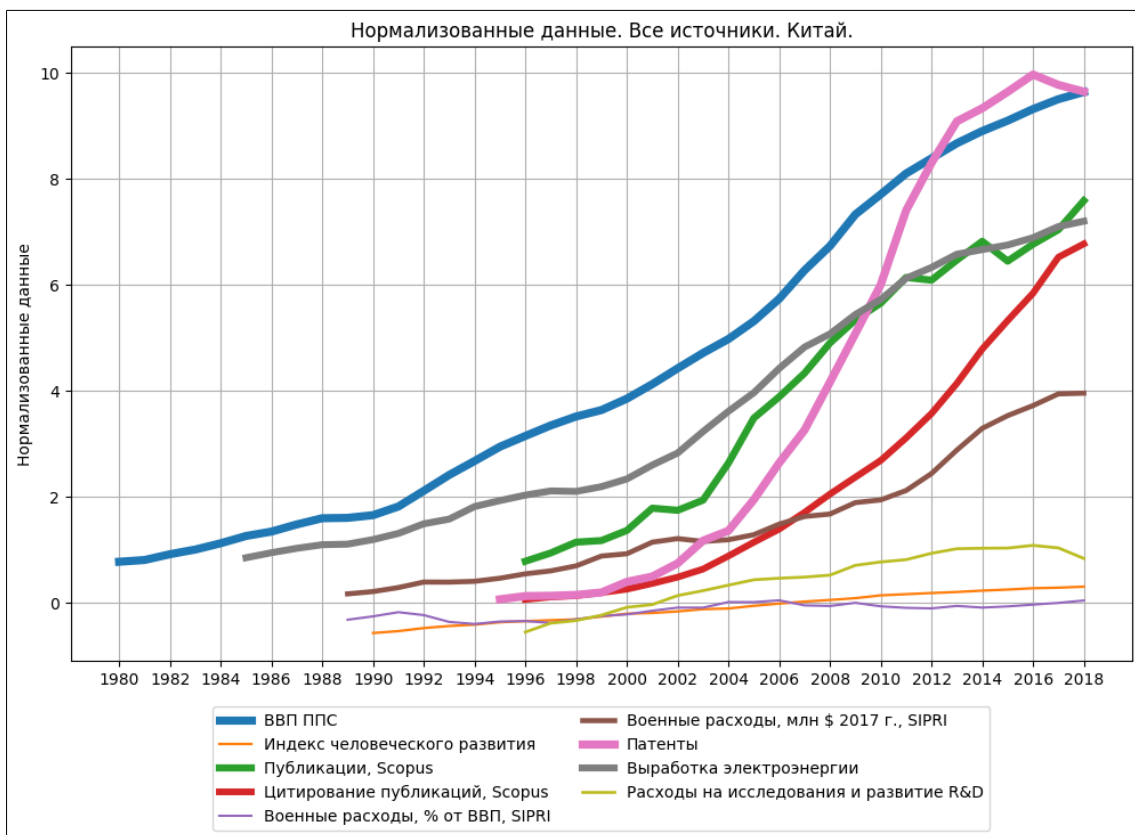


Рис. 18

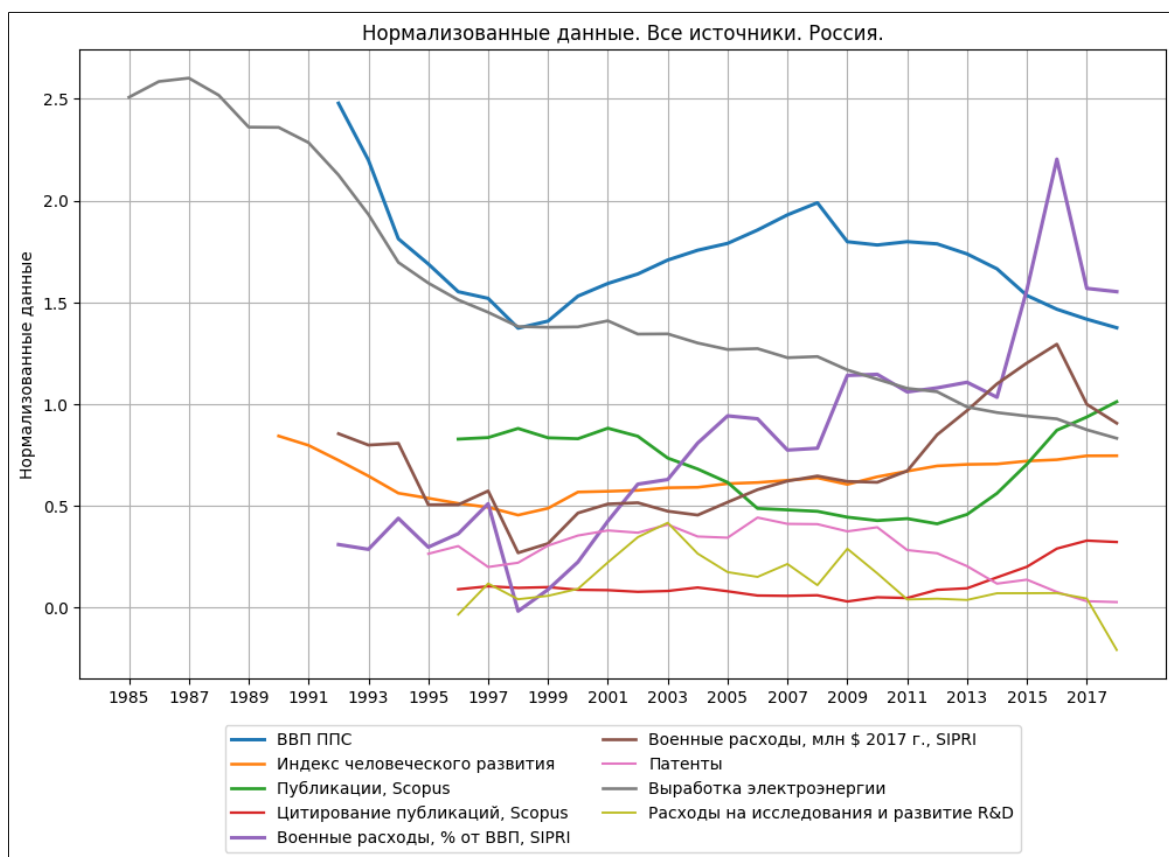


Рис. 19

Все числа из всех источников приведены к единой абстрактной шкале, очищенной от физического смысла разнообразных единиц измерения. При этом ни единого бита информации, содержащейся в источниках не потеряно, но теперь **соотношения** между источниками и странами стало возможным корректно сравнивать.

При внимательном изучении этих графиков можно увидеть много интересного.

Первое что бросается в глаза на графике всех источников России (рис. 19), это неизменное падение ВВП начиная с 2008, при этом выработка электроэнергии падала опережающими темпами, что свидетельствует об относительном отставании реального (производственного) сектора экономики.

*Заметим, что если аналитик обратится к абсолютным числам статистики по выработке электроэнергии, он не увидит падения, а скажет, что был даже небольшой рост. Это объясняется тем, что нормализованные данные будучи очищенными от физических единиц измерения, сохраняют **СООТНОШЕНИЯ** параметров между странами. Опережающий рост производства электроэнергии в Китае, определяет **ОТНОСИТЕЛЬНОЕ** падение этого параметра для всех стран, у которых **ТЕМПЫ** роста ниже, чем в Китае.*

У России виден резкий, но нестабильный рост военных расходов с 2010 (причины мы здесь анализировать не будем, хотя они довольно очевидны).

ИРЧП — это интегральный показатель здоровья и образования населения. Радует, что он, пусть не очень быстро, но стабильно растет с 1998 года. Этот показатель очень консервативен, его резкие колебания невозможны по определению, методология его расчета этого не позволяет. Рост этого показателя дает надежду на будущий рост всех остальных наблюдаемых параметров.

После предпринятых МинОбром усилий по стимулированию научно-педагогических кадров к публикационной активности, число публикаций в Scopus стало быстро нарастать. Правда число цитирований, хотя тоже росло, но с заметно меньшей скоростью, а в последний год так и вовсе переломилось вниз.

Число патентов стагнировало с 1996, а с 2006 уже 12 лет падает все быстрее.

Последний параметр «Расходы на исследования и развитие (R&D)» учитывается UNESCO в % от ВВП стран с 1996 года, однако он содержит много лагун (часть из них удалось заполнить методом линейной интерполяции). Мы отнесли его к группе №3.

Таким образом, первичный анализ наблюдаемых данных для России, показывает очевидный перекося в сторону военных расходов, на фоне падения ВВП, и серьезное отставание показателей интеллектуального развития креативной части общества.

Техническая подробность: толщина линий на графиках отражает логарифмическую зависимость весов показателей относительно друг-друга.

На графиках США и Китая это сильно заметнее, чем на графике России, так как масштаб у России заметно меньший: всего от 0 до 2,5 единиц. У США и Китая он в пять раз больше.

Посмотрим на показатели Китая — они гармоничны и даже красивы:

Мощный рост ВВП, военные расходы в процентах от ВВП мизерные, зато цитирование не уступает числу публикаций, а количество патентов росло почти экспоненциально и уже обогнало по весу ВВП.

ИРЧП из-за огромного бедного населения растет медленно, но растет.

Очень, даже просто визуально, из этой картинки понятно, что рост экономики опирается на рост Человеческого капитала. Военный потенциал, является производной от экономики, а не паразитом, удушающим ее.

Правильно отстроенная пирамида: Институты развития ЧК → Человеческий капитал → Рост экономики и военной силы — результат правильной долговременной стратегии руководства страны.

График на рис. 18 доказывает это зримо и неоспоримо.

А вот картина бывшего мирового лидера (рис. 17):

С 2000 года падает ВВП.

Следом покатились и показатели Человеческого капитала.

Только военные расходы растут относительно других и уже стали самым тяжелым (толстая линия) из всех наблюдаемых факторов, характеризующих государство.

Рассмотренные графики весьма информативны и подробны, однако еще более зримую картину может дать интеграция данных.

Рассмотрим графики нормализованные данных по группам источников и коалициям.

Графики по группам источников, странам и коалициям. Нормализованные данные и главные компоненты.

#	Страны \ Группы	Военно-экономическая сила		Человеческий капитал		Институты развития ЧК		Все данные БД	
		Нормализованные данные	Главные компоненты	Нормализованные данные	Главные компоненты	Нормализованные данные	Главные компоненты	Нормализованные данные	Главные компоненты
1	США								
2	Китай								
3	Великобритания								
4	Германия								
5	Япония								
6	Индия								
7	Франция								
8	Канада								
9	Италия								
10	Россия								
11	Австралия								

Рис. 19

Здесь не имеет значения в какой строке выбрать иконку — в пределах колонки они ведут на одну и ту же картинку. Всего их три, по каждой группе источников.

На каждом рисунке (20, 21, 22) по четыре линейных графика по числу коалиций (число и состав коалиций, выбраны условно, но не случайно; при необходимости их легко скорректировать).

Графики построены на основании суммы значений нормализованных данных стран входящих в коалицию ("Россия и союзники", "США и союзники", "Китай и союзники", "Исламский мир") и по сумме показателей входящих в группу источников ("Военно-экономическая сила", "Человеческий капитал", "Институты развития ЧК").

Можно сказать, что графики, довольно точно показывают соотношение сил коалиций в трех рассматриваемых областях.

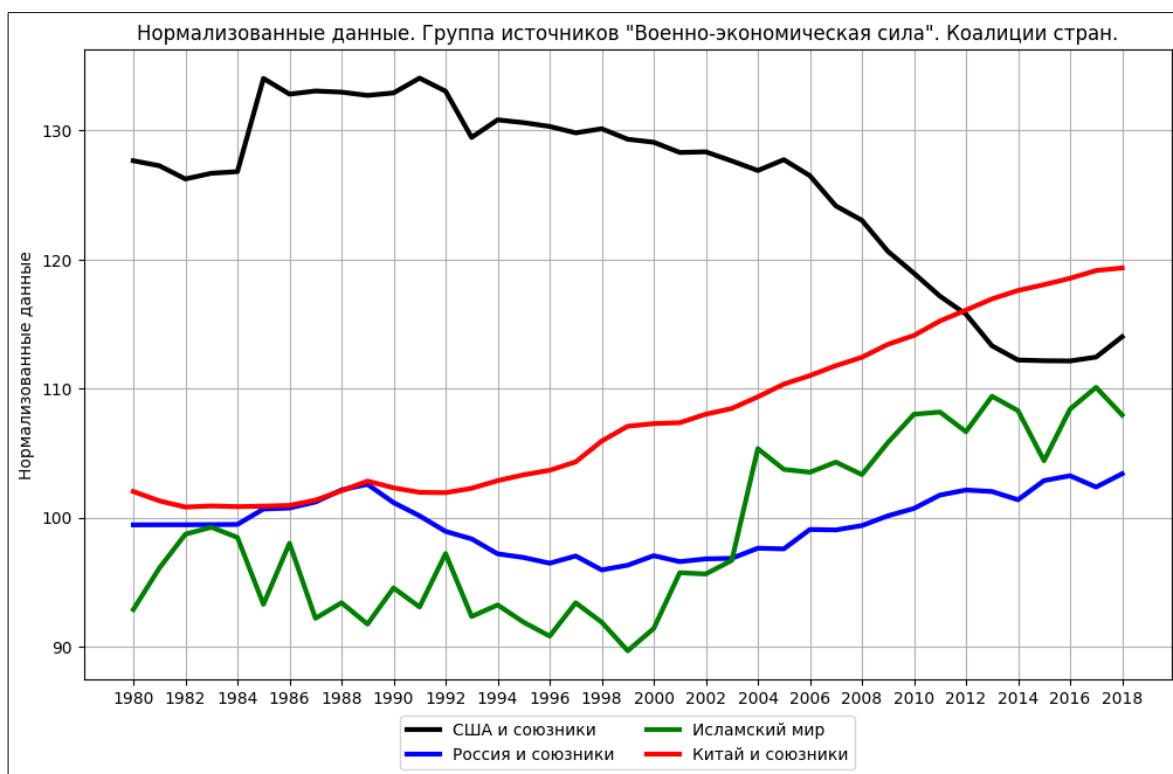


Рис. 20

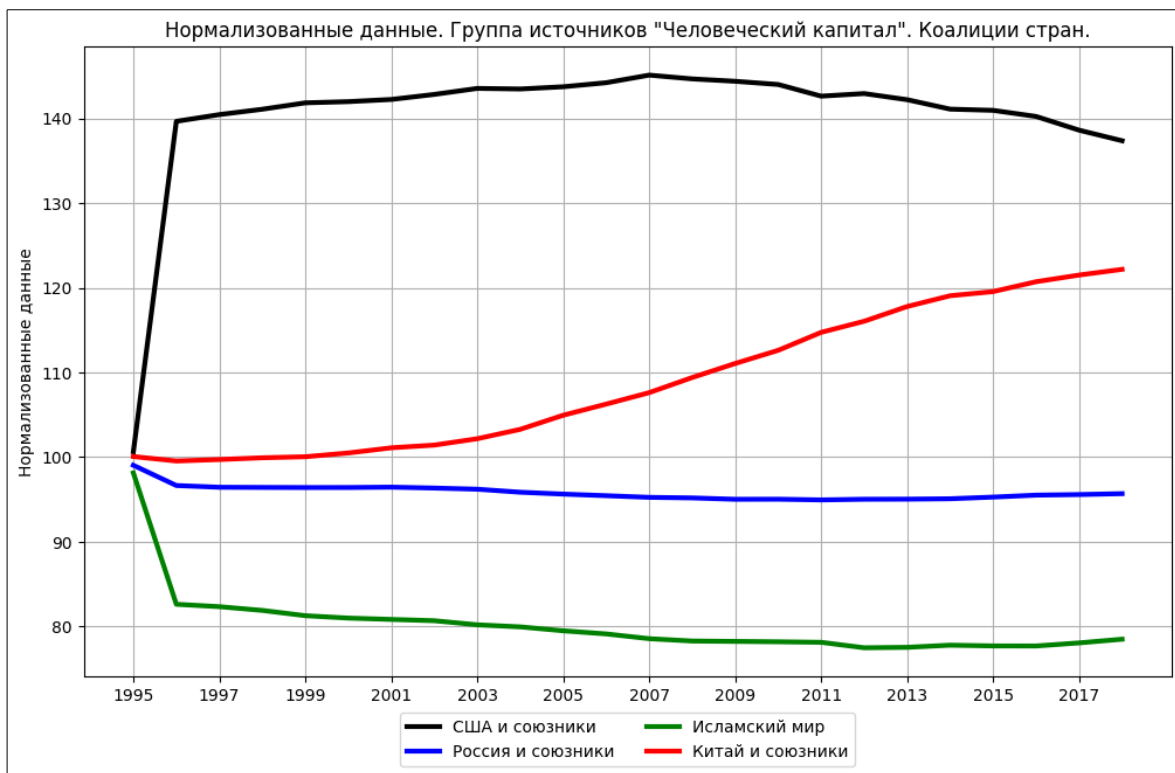


Рис. 21

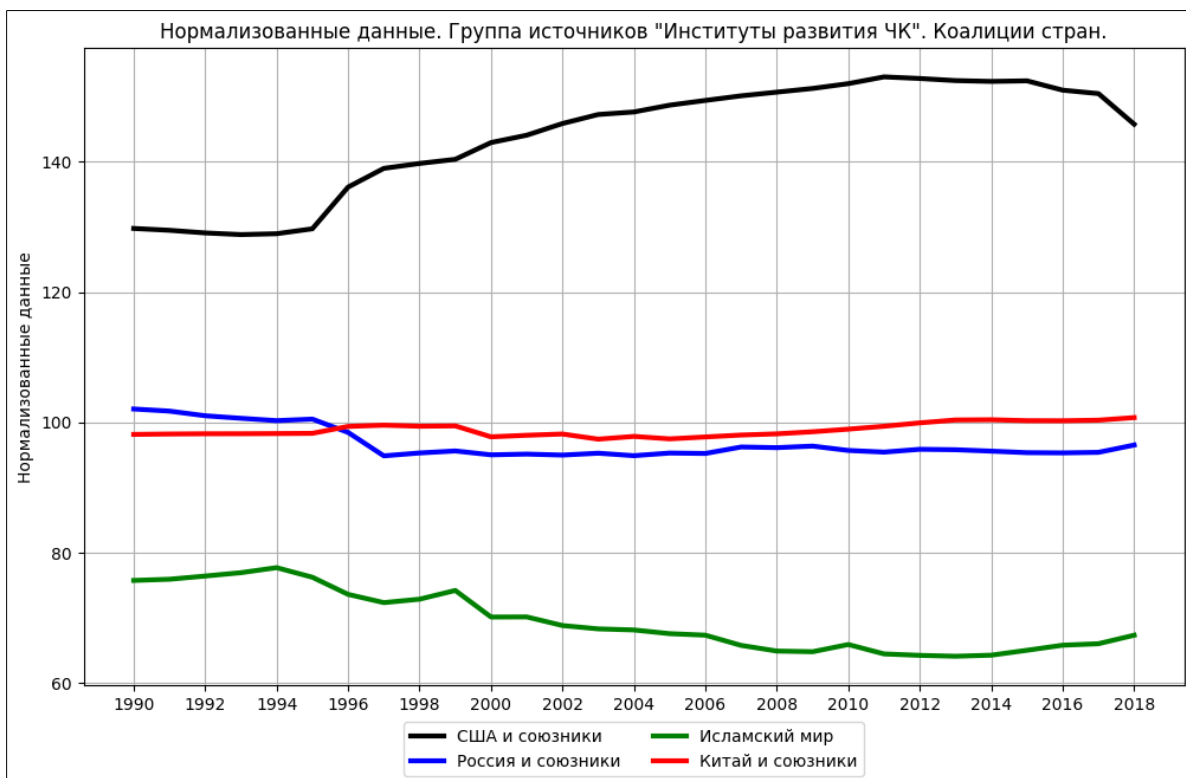


Рис. 22

На основании этих трех графиков (рис. 20-22) можно сделать следующие выводы:

1. Западная цивилизация больше не лидер в военно-экономической сфере — их обошел Китай с союзниками.
2. Однако человеческий потенциал и институты его развития в западном мире существенно превосходят всех и хотя, тренды и здесь у них отрицательные, однако темпы падения пока невелики, что дает им шанс выровнять ситуацию и в экономике.
3. Российские с союзниками интегральные показатели подтверждают: застойные явления ведут к стагнации.
4. Исламский мир, видимо, за счет числа и бедности населения и богатства недр, хотя и судорожными рывками, но растет в экономической сфере и даже в 2003 году обошел Российскую условную коалицию, но невнимание к человеческому фактору развития, еще долго не позволит им выйти на стратегический уровень Китая и Запада.
5. У России, в целом, шансов больше, но если не произойдет перелома в областях "Человеческий капитал" и "Институты развития ЧК", то и "Военно-экономическая сила" неизбежно будет деградировать. Последствия мы здесь описывать не будем.

Конец примера 2.

Приведенный выше пример из области Нормализованных данных — это один из первых этапов подготовки данных к исследованию их при помощи математической модели, но и он уже дал возможность наглядно сравнить как динамику развития стран, так и их отдельные показатели.

Сравнение Примера 1 и Примера 2, заставляет признать, что применение графических средств, не только создает некоторый комфорт для аналитика, но и позволяет намного глубже и исследовать предмет изучения.

Следующим этапом станет вычисление Главных компонент, т.е. снижение размерности матриц содержащих данные стран с 10 показателей до двух. (В базе накоплены данные по 20 показателям, но из-за неполноты по годам и/или по странам часть из них была исключена из расчетов.)

Это позволит сильно упростить построение математической модели, которую мы хотим создать в итоге.

На сайте <http://eurasian-defence.ru/richdb> можно детально рассмотреть содержание собранной базы данных, а также увидеть более 1300 диаграмм и графиков построенных на их основе.

Главные компоненты на графиках

Несколько примеров взятых с сайта (там их намного больше):

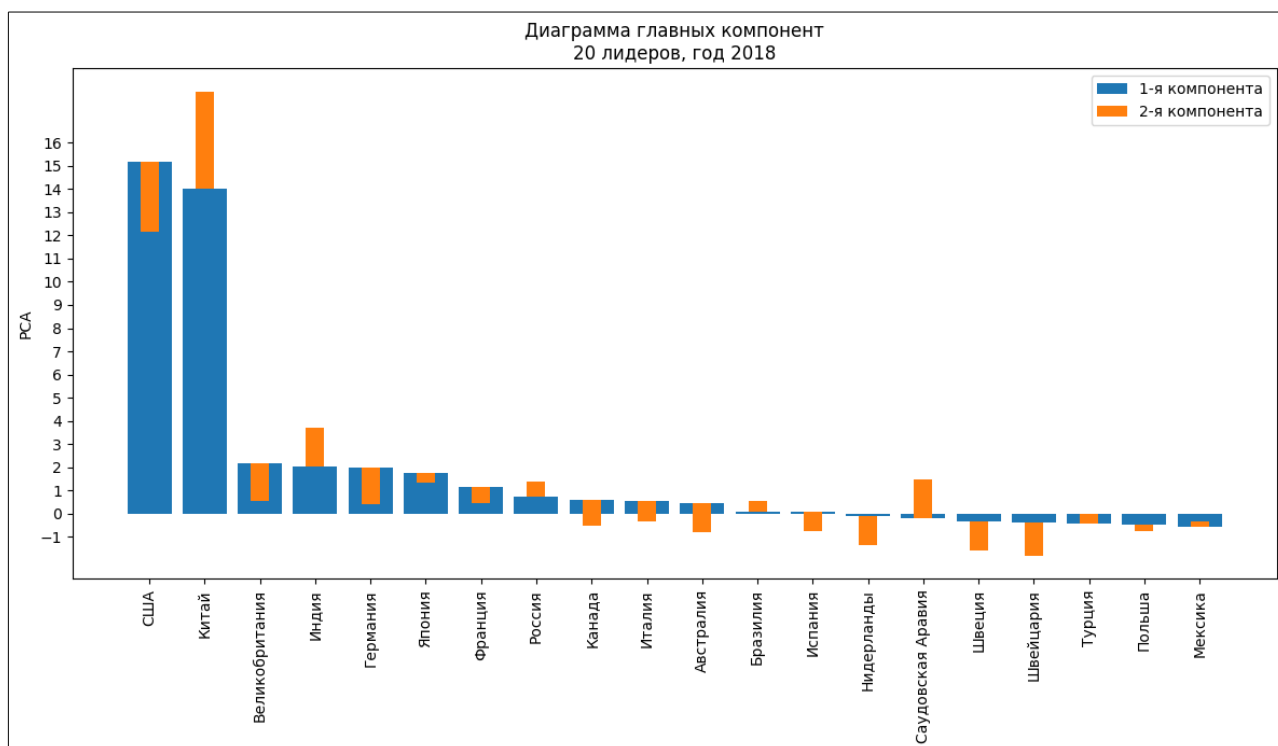


Рис. 23

Из этой диаграммы сразу видно, что в 2018 году:

- неравномерность развития, только для первых 20 стран составляет более 15 раз;
- США и Китай обгоняют следующих за ними 4 страны более чем в 7 раз;
- отличие развивающихся стран — заметная вторая компонента;
- причем у западных (старых) и у развивающихся (условно новых) экономик — они разнонаправлены: у старых оранжевый язык направлен вниз, а у новых вверх.

Это просто наблюдения, пока никаких выводов.

На той же странице [3D-модели и диаграммы](#), где мы открыли предыдущую диаграмму можно открыть анимированную серию диаграмм по всем годам. Ролики зациклены, т. е. можно бесконечно медитировать, глядя на то, как от года к году страны-лидеры меняются местами.

Динамику изменений главных компонент во времени хорошо иллюстрируют следующие графики (на сайте есть для всех стран, здесь приведены для России, Китая и США).

Для каждой страны мы приведем по 4 рисунка. По одному на каждую группу источников и суммарный построенный по всей БД.

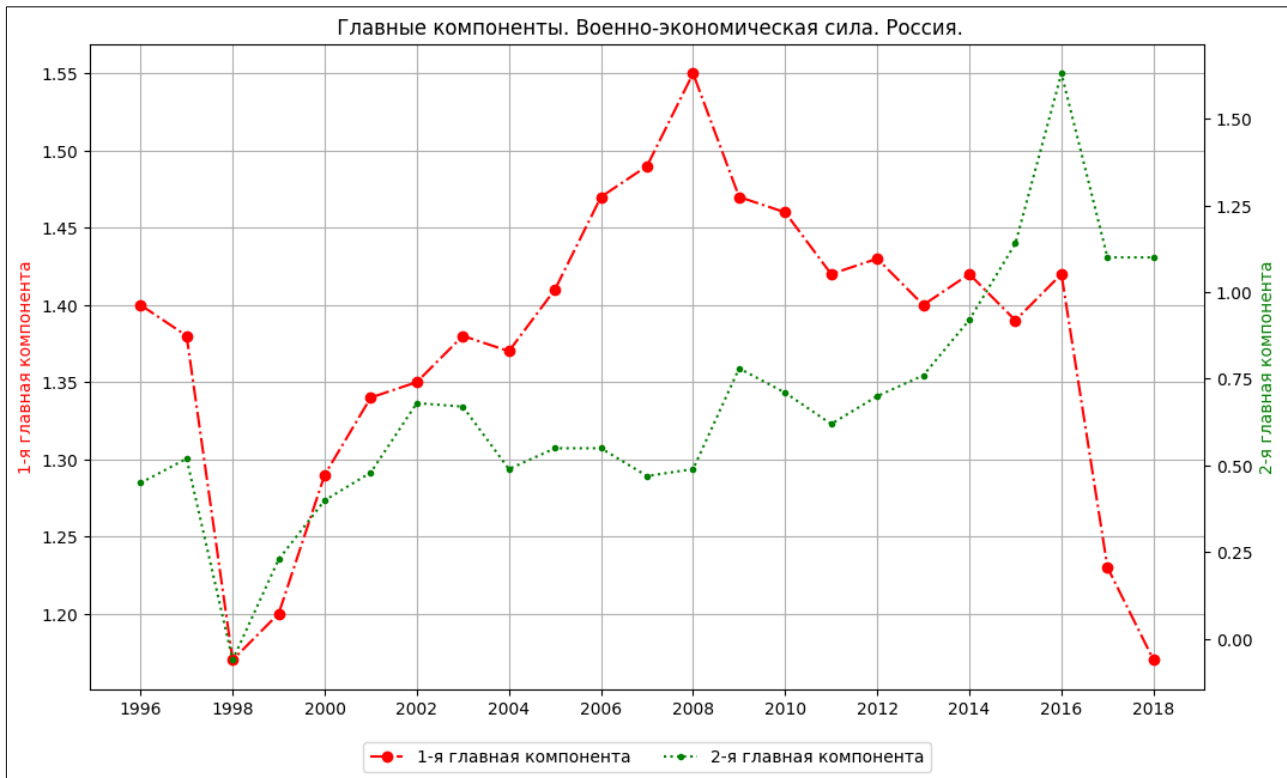


Рис. 24

Здесь отчетливо видны провалы в развитии России, обусловленные известными причинами: 1998 — дефолт, 2008 — мировой кризис, 2014 — санкции (угол падения говорит о степени их влияния на экономику).

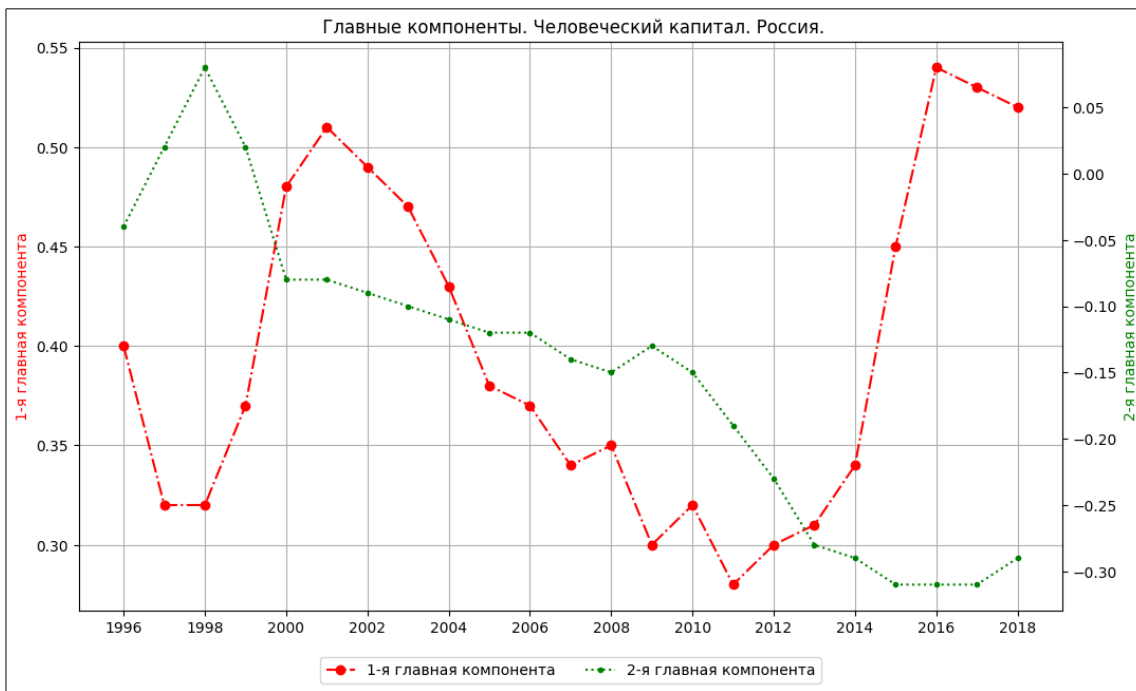


Рис. 25

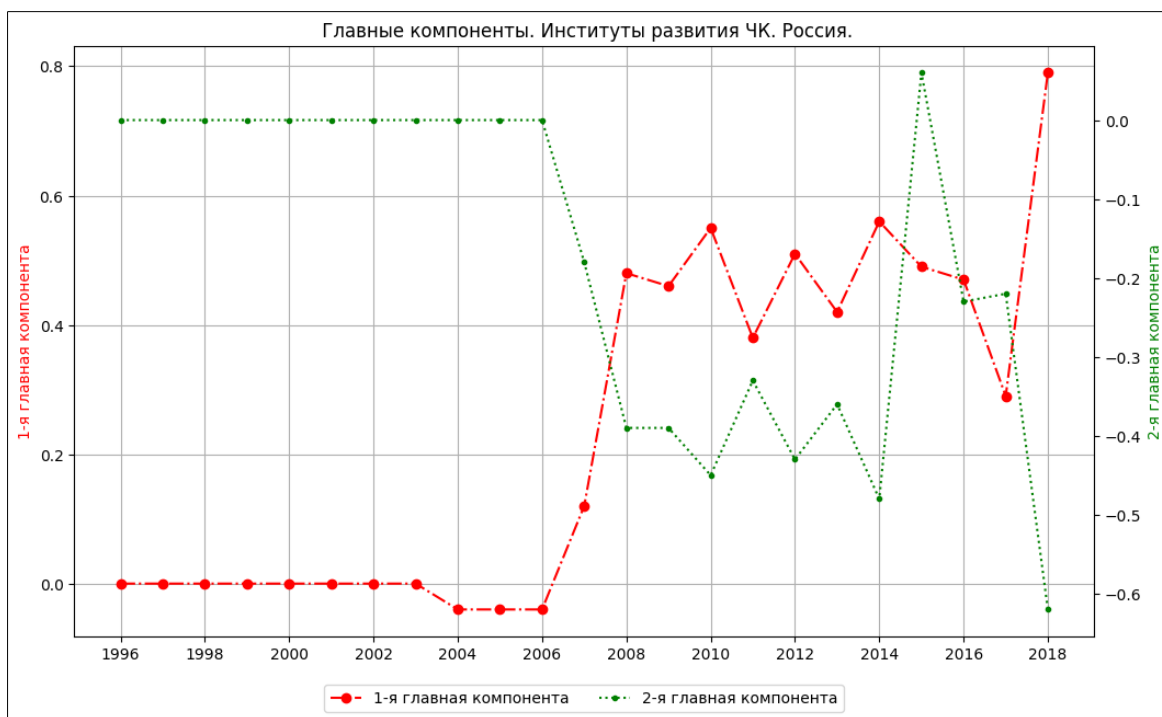


Рис. 26

Рассмотрение рис. 25 и 26 дает понимание того, что является первичным, а что вторичным. В 2006 году впервые стали заметны позитивные изменения в институтах развития человеческого капитала и только спустя пять лет пошел вверх график собственно ЧК. До начала его падения, после изменения первичного тренда прошло всего два года.

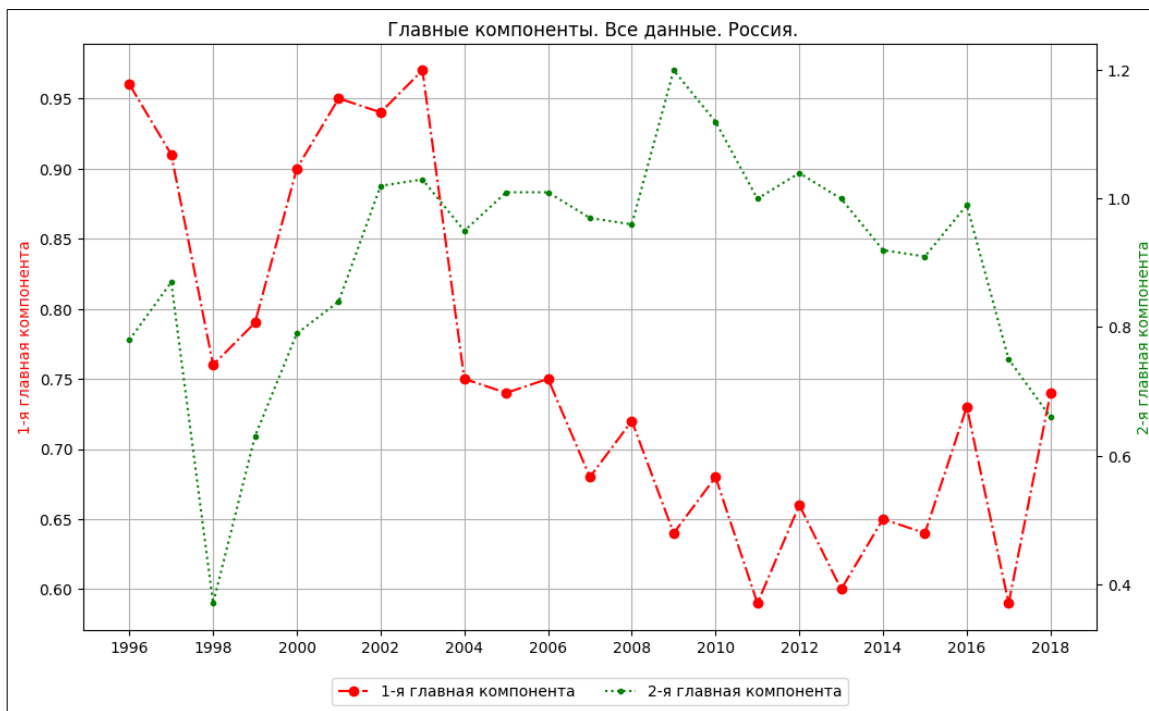


Рис. 27

Обобщенный график подтверждает ранее сделанный вывод о стагнации России. В таком масштабе это видно особенно отчетливо.

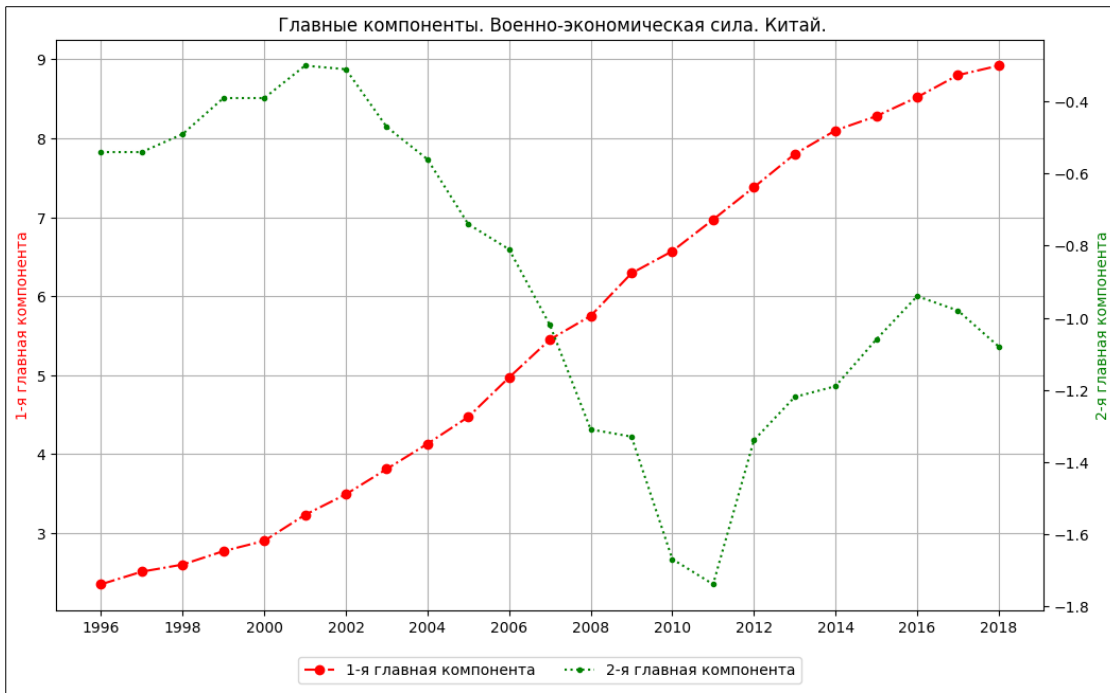


Рис. 28

Четкое поступательное движение вверх Первой главной компоненты (1-й ГК) Китая — свидетельство правильности выбранной стратегии и отличной управляемости, позволяющей вовремя реагировать на тактические колебания мировых экономических и политических процессов.

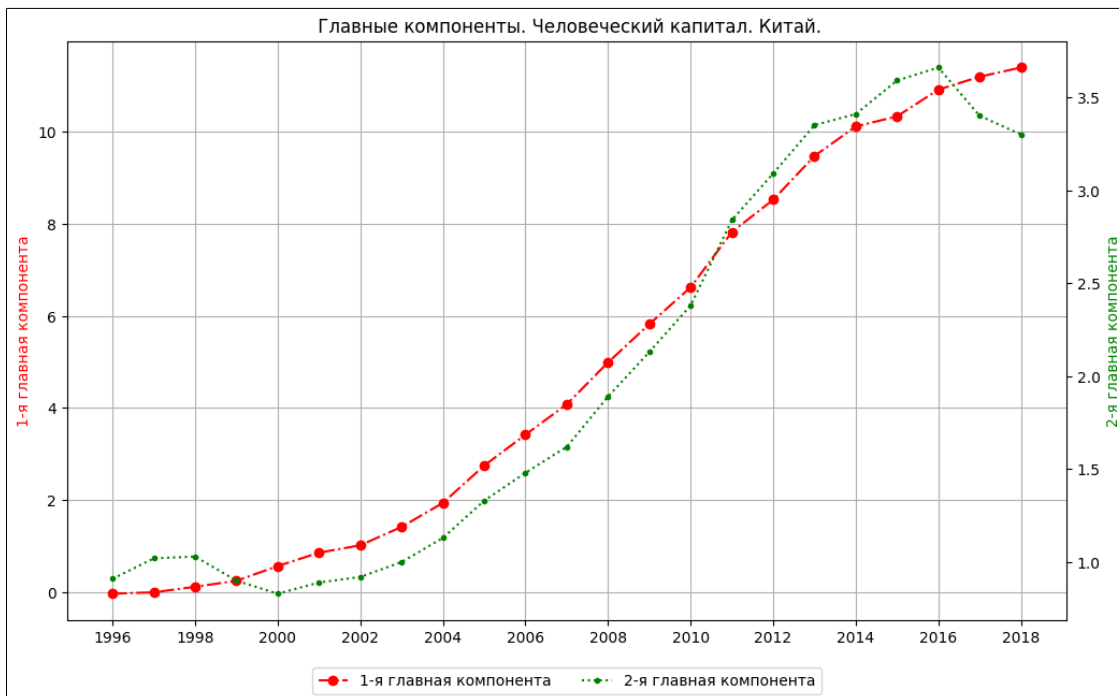


Рис. 29

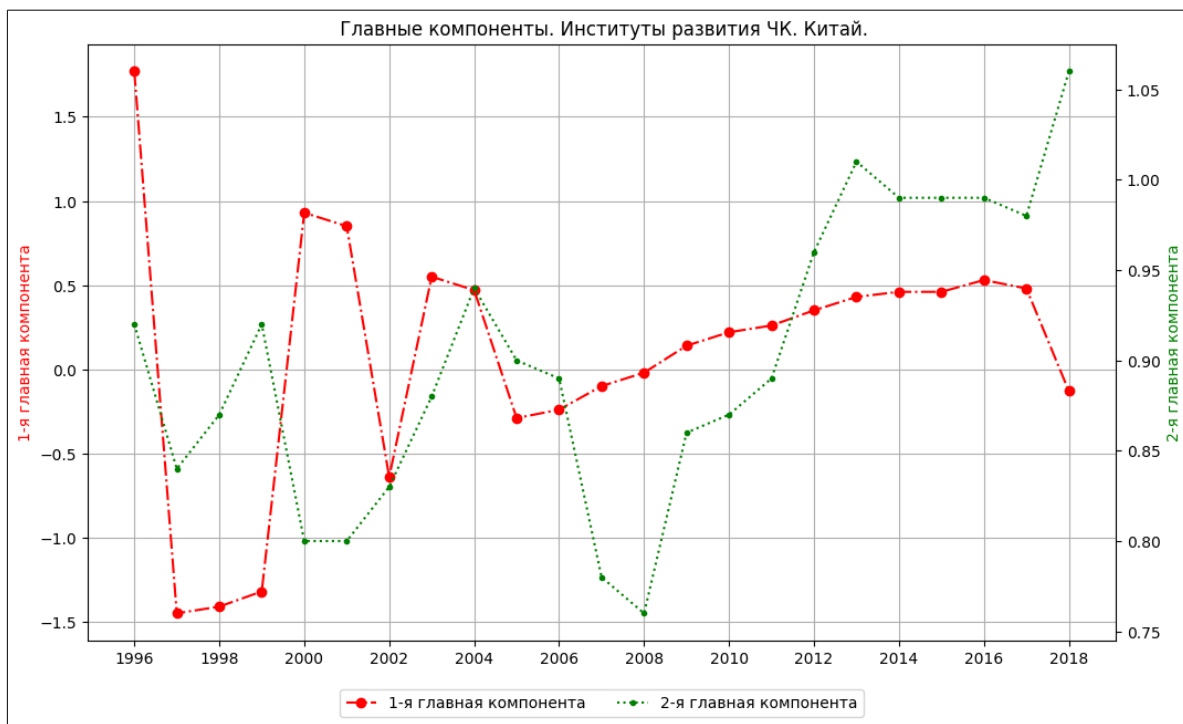


Рис. 30

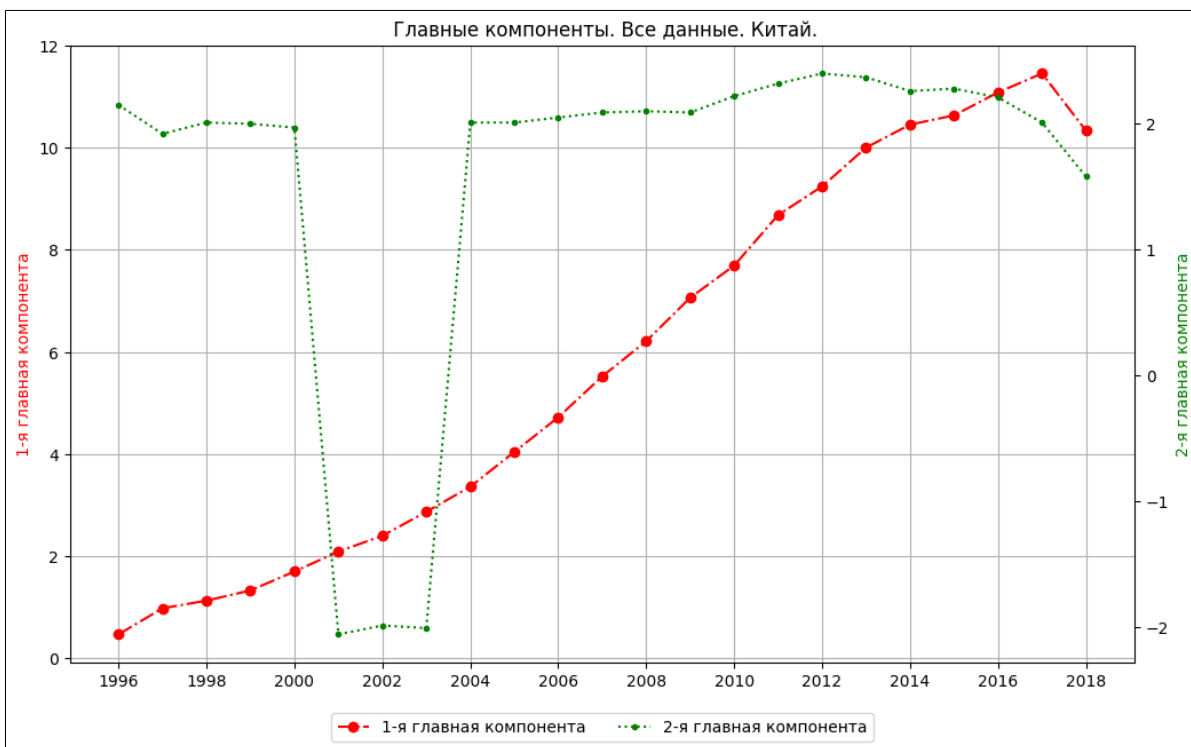


Рис. 31

Все графики показывают синхронный рост — Китай все делает правильно.

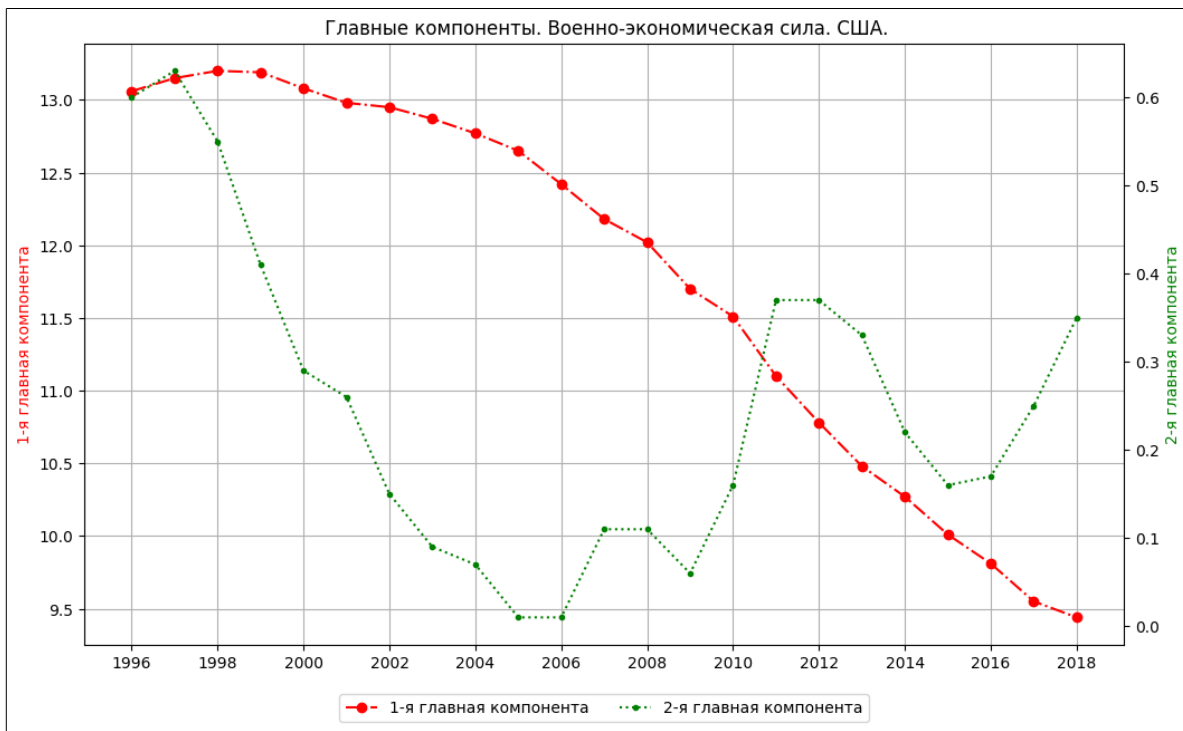


Рис. 32

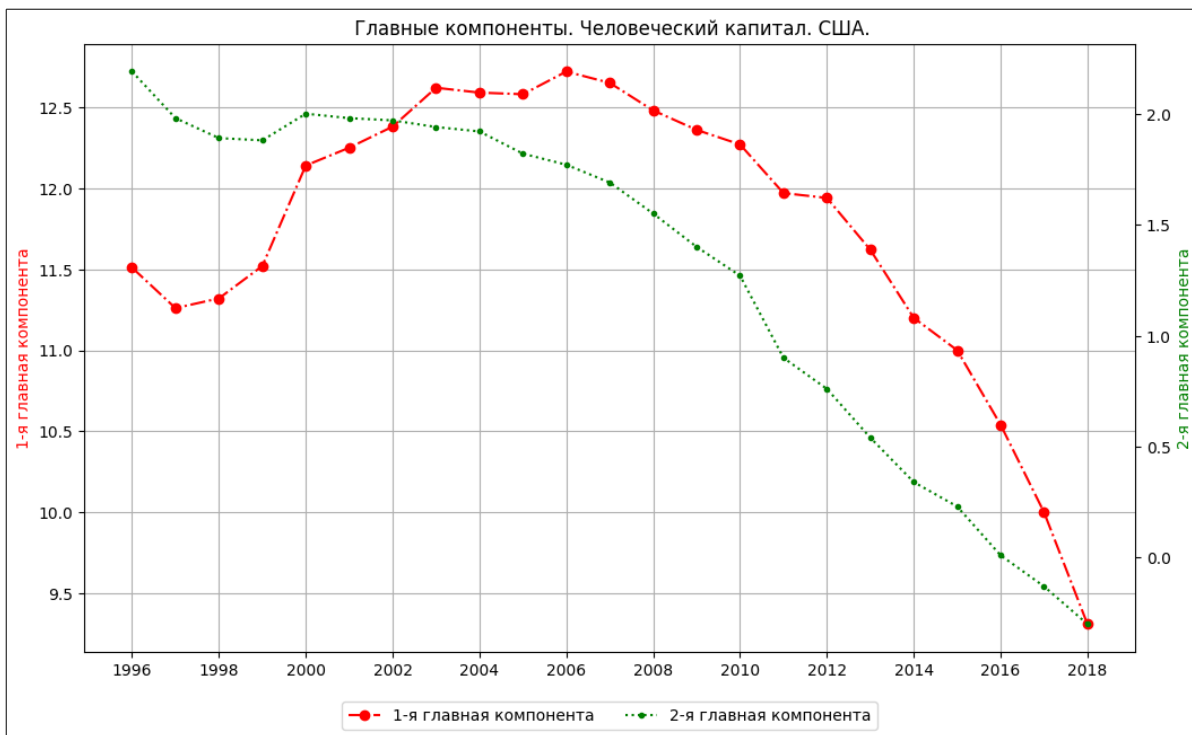


Рис. 33

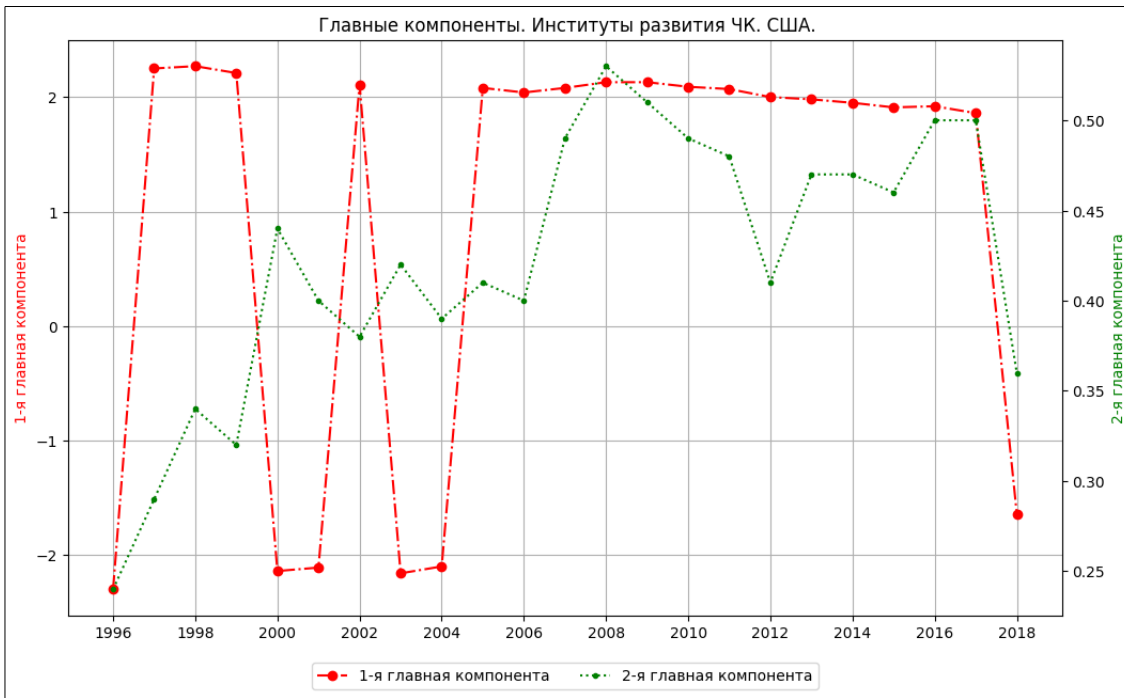


Рис. 34

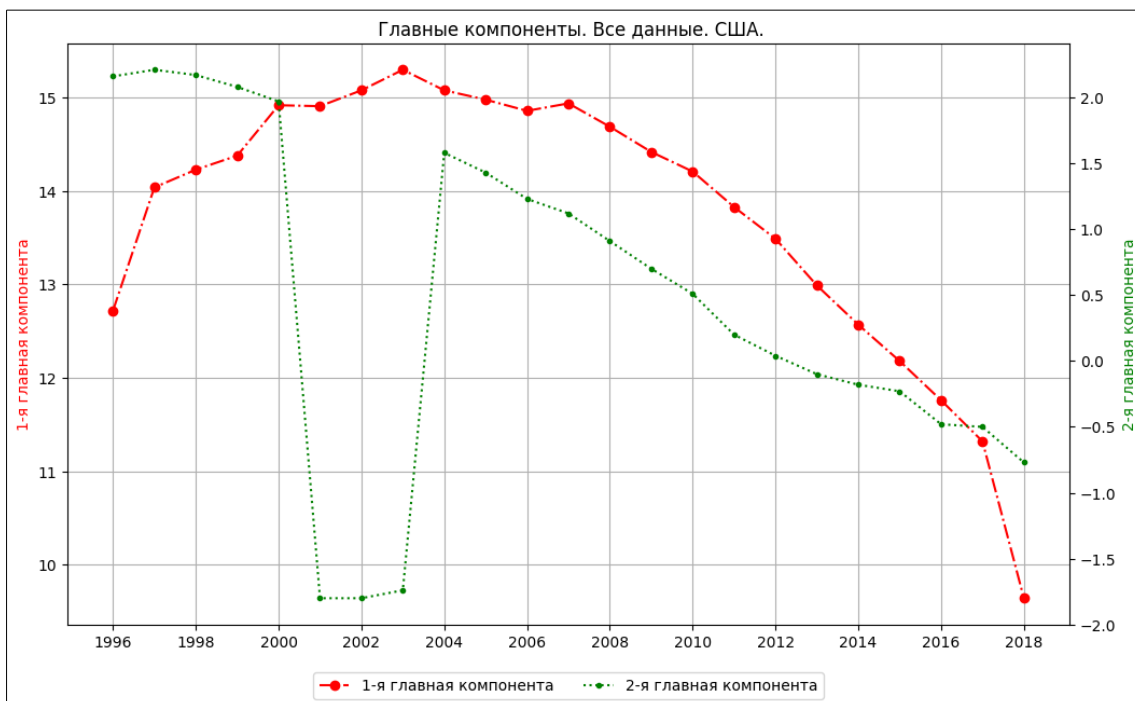


Рис. 35

Похоже здесь комментарии излишни.

Глиссада падения Америки вынуждает Трампа тянуть штурвал на себя. Есть ли у него шанс вывести ее из пике — очень большой вопрос, однако утянуть в воронку тонущего Титаника всю сложившуюся после второй мировой войны планетарную систему, этот процесс точно сможет.

Ниже диаграмм Главных компонент стран лидеров, на той же странице есть таблица озаглавленная «Точечные диаграммы распределений главных компонент»

Точечные диаграммы распределений главных компонент														
Группы\Годы	Анимация	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006
Военно-экономическая сила														
Человеческий капитал														
Институты развития ЧК														
Все данные														

Рис. 36

Здесь более 160 точечных графиков, на которых очень хорошо видно, как неравномерно распределяются страны на плоскости в координатах Главных компонент.

На следующих трех рисунках, взятых для примера, если смотреть их последовательно видно как меняются места точек России, Индии и особенно Китая, относительно более стабильных мест Японии, Франции и США.

Видно как Россия медленно, Индия быстрее, а Китай очень быстро отдаляется от плотной группы аутсайдеров, как Китай догоняет Америку.

Все это совсем хорошо наблюдать на анимированных картинках, ведь они по сути те же самые рисунки, просто сшитые в один закольцованный ролик.

В этой таблице можно по отдельности рассмотреть распределения главных компонент, рассчитанных по каждому году и по группам источников:

- Военно-экономическая сила
- Человеческий капитал
- Институты развития ЧК
- По всей базе

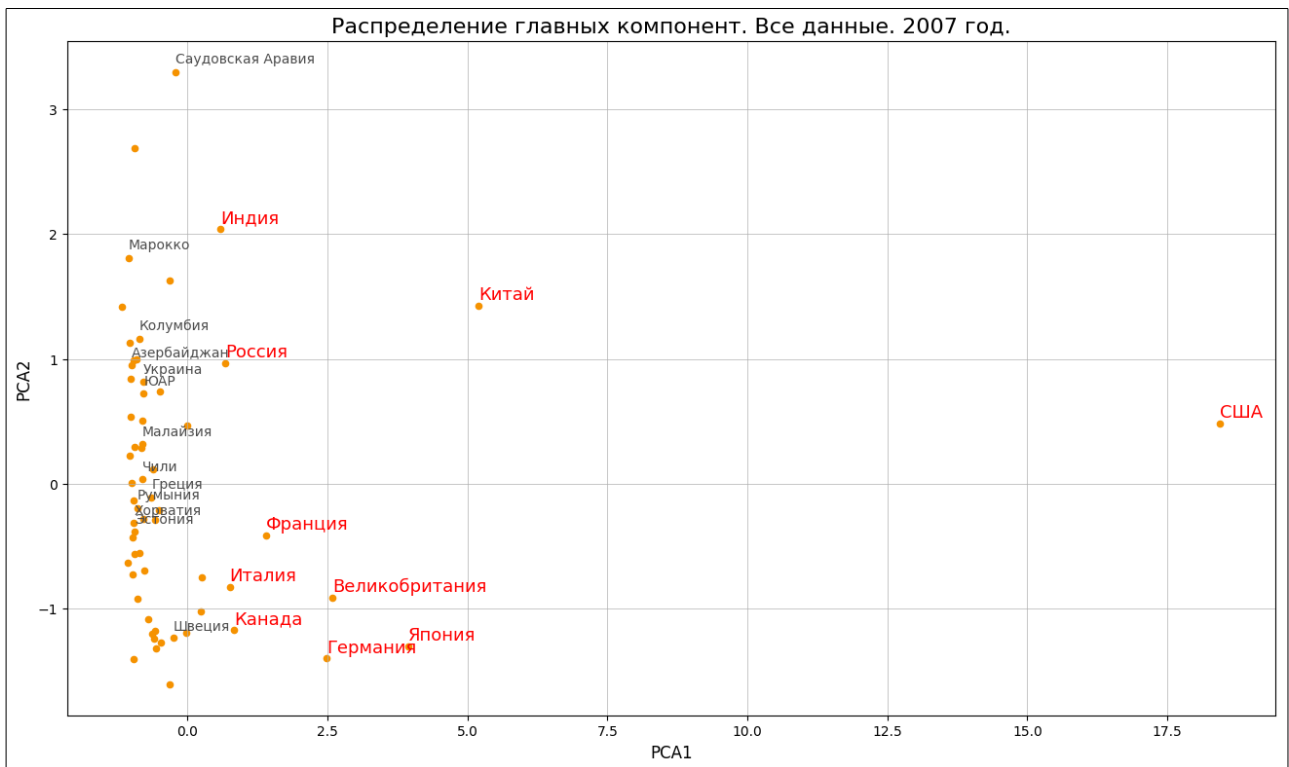


Рис. 37

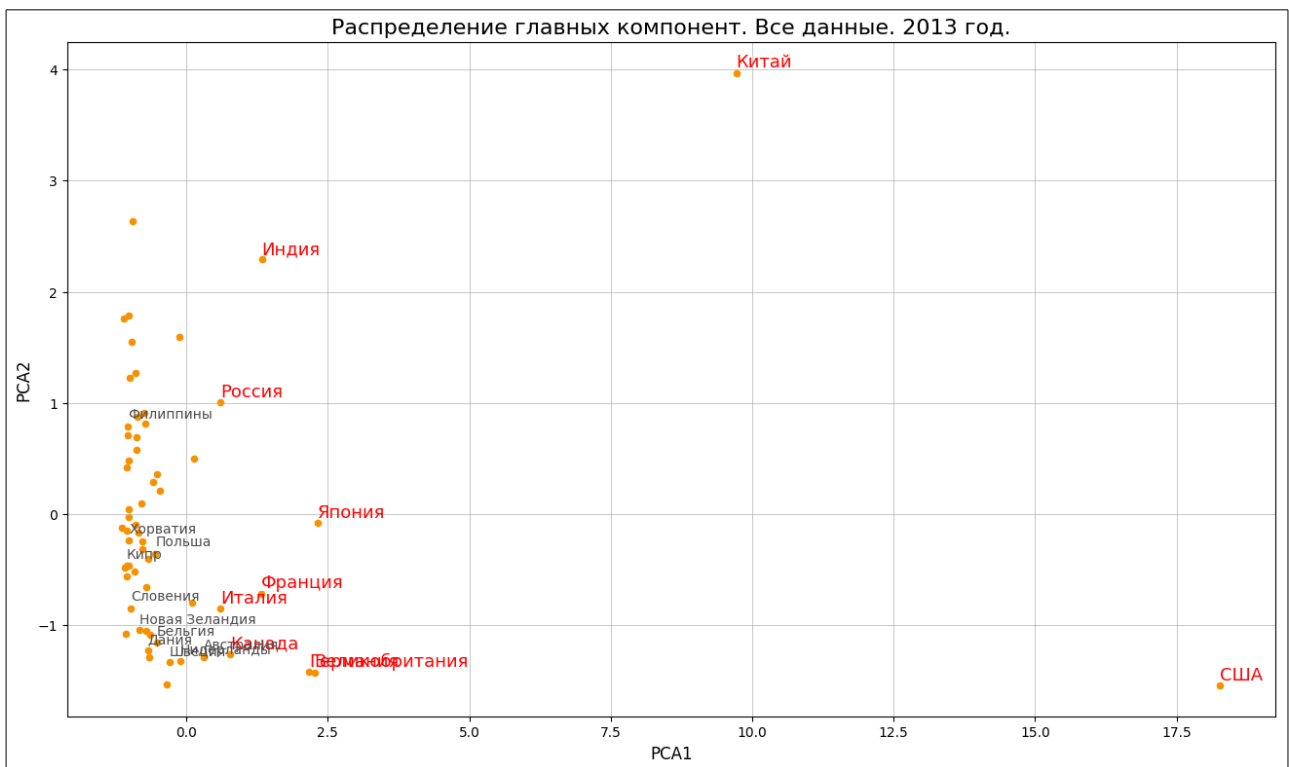


Рис. 38

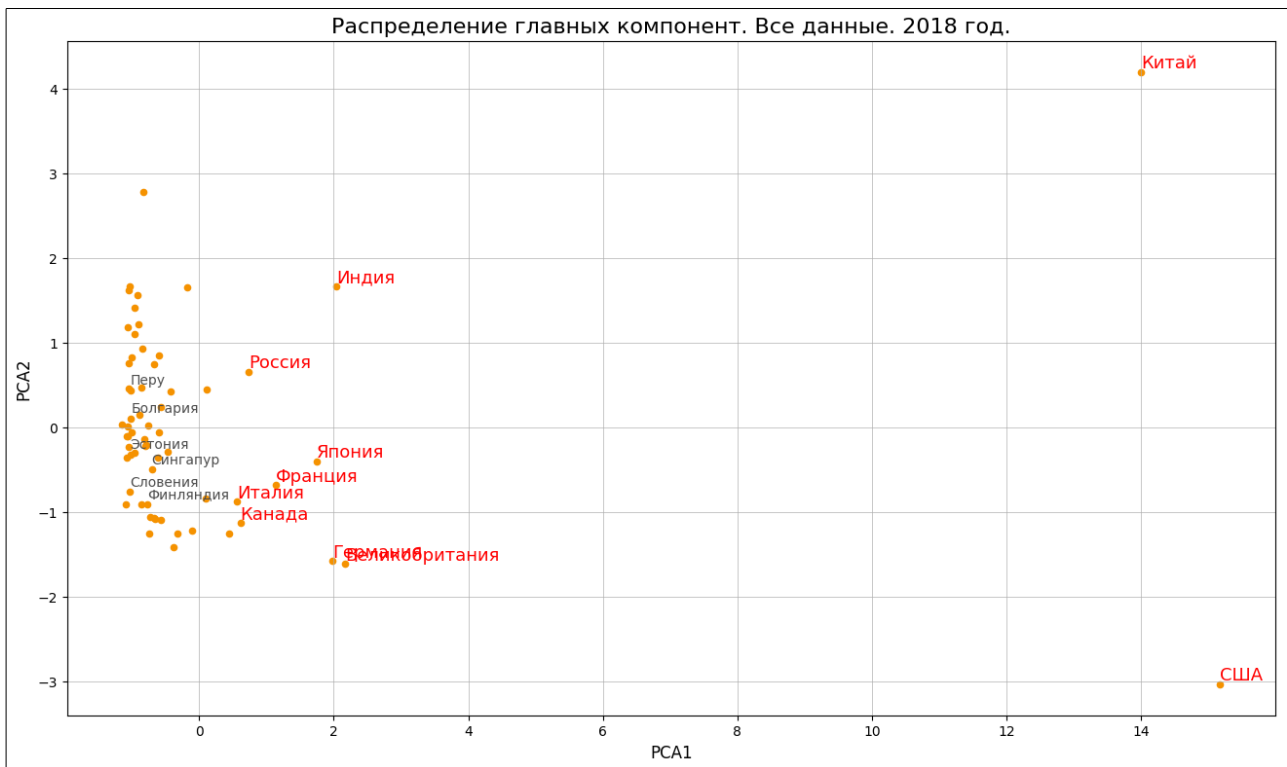


Рис. 39

Несовершенство алгоритма вывода названий стран, на этих рисунках иногда приводит к накладкам, поэтому здесь они выводятся не все, а случайным образом выбранные (кроме первой десятки).

На следующем этапе анализа, группы стран будут выделены обученным алгоритмом в анклавов (кластеры), хотя уже и сейчас видно, что кроме двух явных лидеров, есть красные — твердые середняки и оранжевые — аутсайдеры. Впрочем глубину дифференциации можно увеличить.

Следующая небольшая таблица «Объединение точек в кластеры на диаграммах распределений ГК» является продолжением предыдущей, однако содержит в себе гораздо больше интеллекта. Обученный на реальных числах (но отнюдь не имеющий понятия, ни о политике, ни об экономике, ни о человеческом капитале) алгоритм сформировал для каждого года по два кластера точек, основываясь лишь на их числовых характеристиках: первую (РСА1) и вторую (РСА2) главные компоненты, вычисленные по довольно сложной схеме из совокупности всех данных имеющихся в базе.

На рис. 40, 41, 42 (на сайте их 23 шт.) выделены два кластера точек:

1. Правый кластер - черные точки на голубом фоне;
2. Левый кластер - синие точки на желтом фоне.

Только в 1997 алгоритм поменял цвета местами (предоставляю аналитикам выдвигать гипотезы по поводу особенности этого года).

Кроме того часть точек не включены не в один из кластеров, несмотря даже на то, что некоторые из них, находятся внутри выделенного контура. По какой-то причине, алгоритм не посчитал их пригодными для включения в кластер — такие точки имеют бирюзовый цвет.

Двенадцать стран имеющие наибольшие первые ГК, всегда подписаны красным цветом. Остальные подписаны черным шрифтом помельче. Причем, чтобы предотвратить наложение множества надписей друг на друга, подписаны только некоторые, случайным образом выбранные, точки/страны.

Не всегда все 12 лидеров попадают в Правый кластер, например, в 2007 и в 2009 годах Бразилия вылетала в Левый кластер, но позже всё же закрепилась в Правом. Это обусловлено, как неравномерным развитием самой Бразилии, так и небольшим расстоянием между кластерами.

Однако в 2018 году разрыв (зазор) между кластерами заметно увеличился, причем не столько за счет роста сильных, сколько за счет дальнейшего ослабления слабых (см. нижнюю шкалу РСА1 на слайдах 2017 и 2018 гг).

В качестве примера использования подобных графиков в исследованиях, проследим положение на них Украины. Оно обозначено на рисунках малиновой стрелкой.

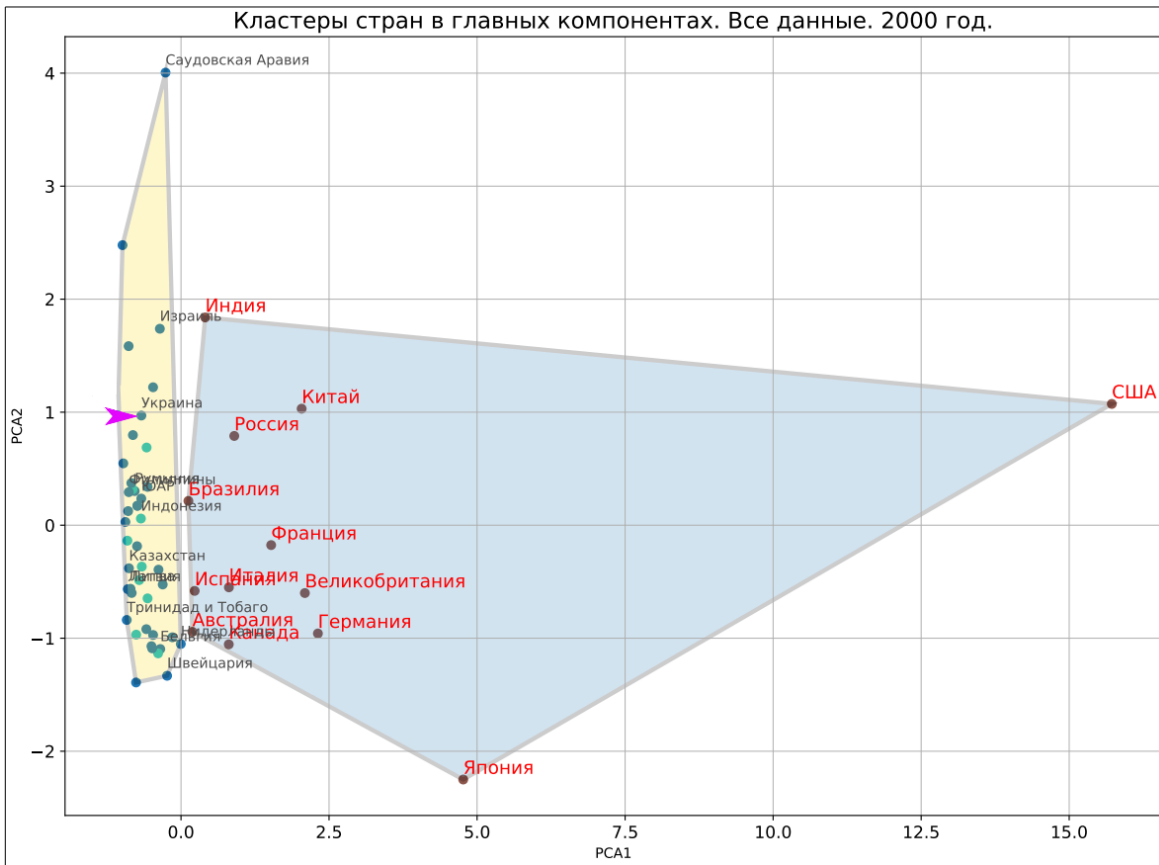


Рис. 40

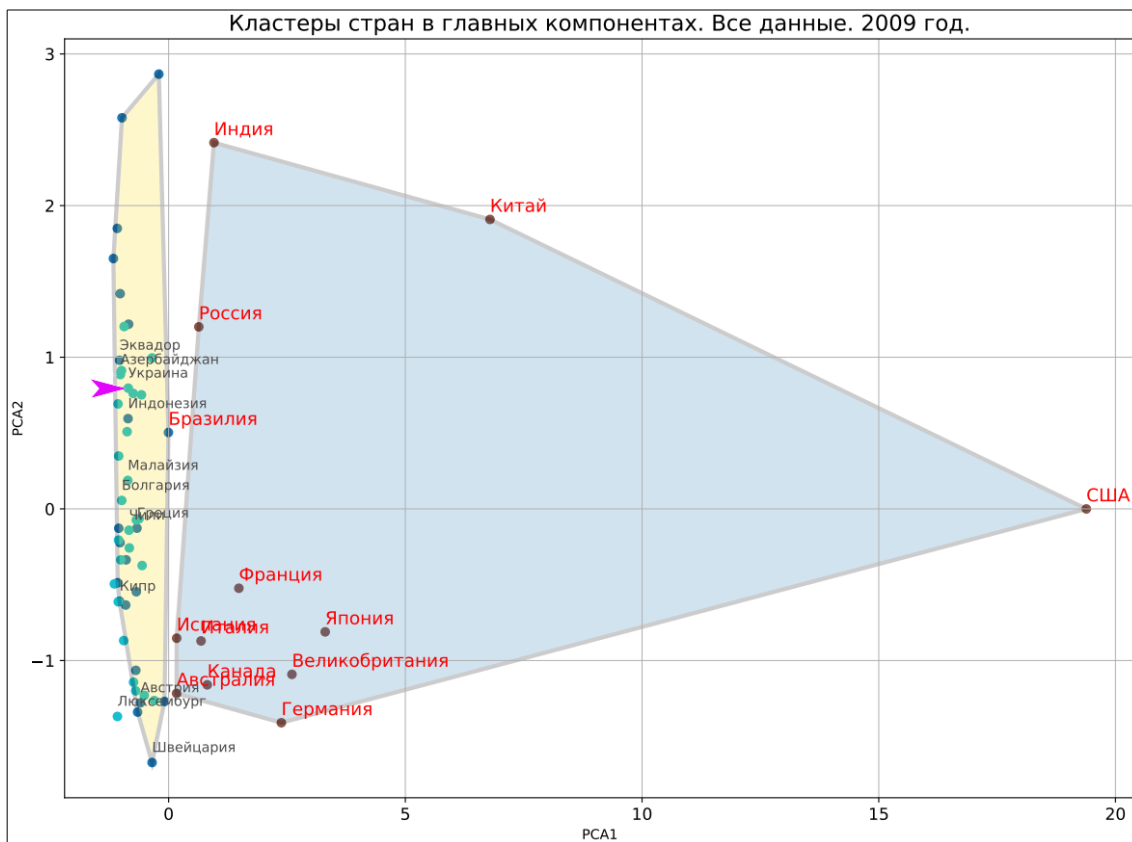


Рис. 41

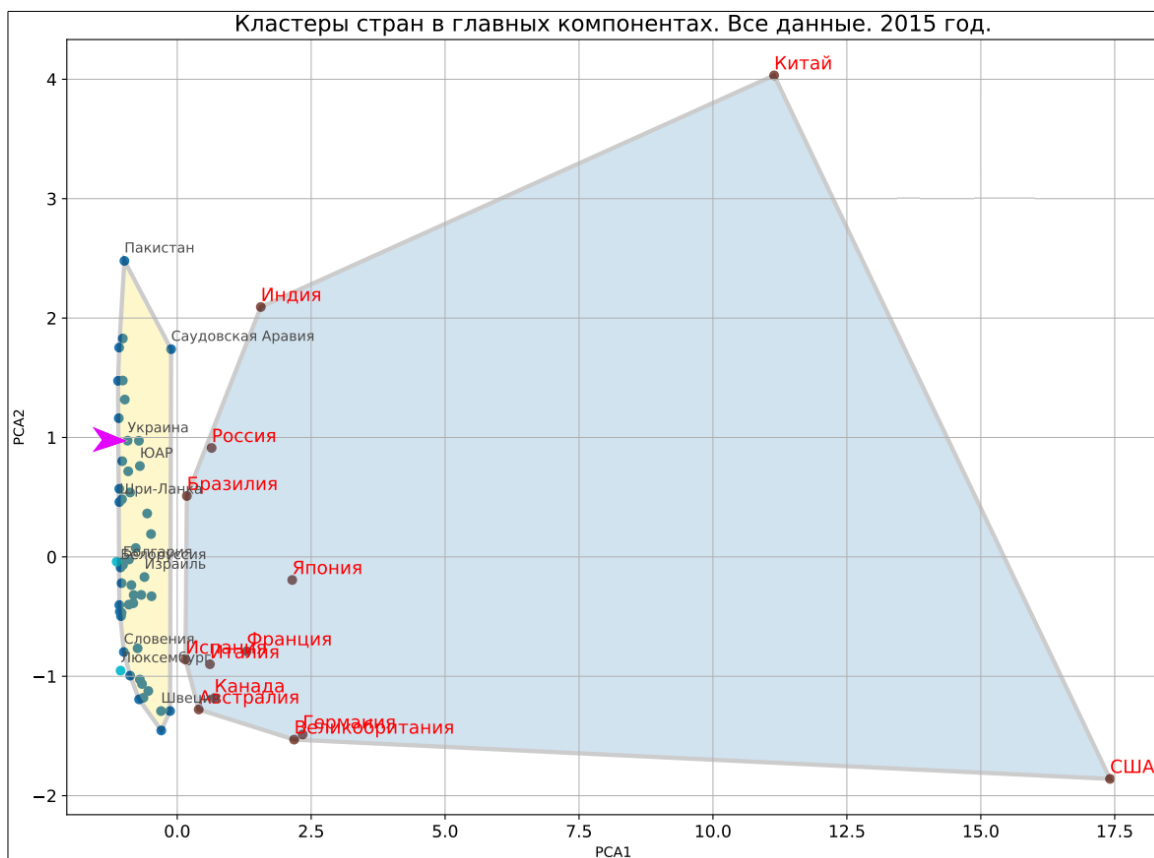


Рис. 42

Какие факты мы можем извлечь из этих, всего трёх, рисунков:

1. Украина в 2000 и 2015 гг. была включена в Левый кластер.
2. Украина в 2009 г. не вошла ни в один из кластеров.
3. Нижний контур обоих кластеров, соответствующий минимальным значениям PCA2, замыкают «старые демократии»: Великобритания, Германия, Япония, США в Правом кластере и Швеция, Швейцария в Левом.
4. Верхний контур, соответствующий максимальным значениям PCA2, замыкают динамично развивающиеся страны: Индия, Китай, Россия в Правом и Пакистан, Саудовская Аравия в Левом.

Оставляю аналитикам возможность интерпретации этих фактов.

Трёхмерные модели

3D-модели строятся на тех же данных, что и все представленные выше иллюстрации. Для их построения и обеспечения интерактивности использованы библиотеки Python: `plotly`, `matplotlib`, `pandas`, `pumpru` и другие.

В печатном (фиксированном) тексте невозможно продемонстрировать всю прелесть трёхмерной модели, которую на сайте можно вертеть, масштабировать, выбрать ракурс с которого наиболее отчетливо видны те или иные тенденции, зависимости, связи.

Наиболее удачные варианты можно зафиксировать в простую плоскую картинку. Этим и воспользуемся. Ниже приведено несколько снимков с трёхмерных моделей, которые на сайте можно найти в разделе [3D-модели и диаграммы](#).

Здесь ось X — первая, Z — вторая компонента, Y — время в годах.

На этом трёхмерном графике: синие точки — Россия, красные — Китай, черные — США.

PCA1 и PCA2 — английские аббревиатуры 1-й и 2-й главных компонент.

Нулевой год здесь 1996, последний год 2018.

При вращении этой коробочки координаты могут меняться местами.

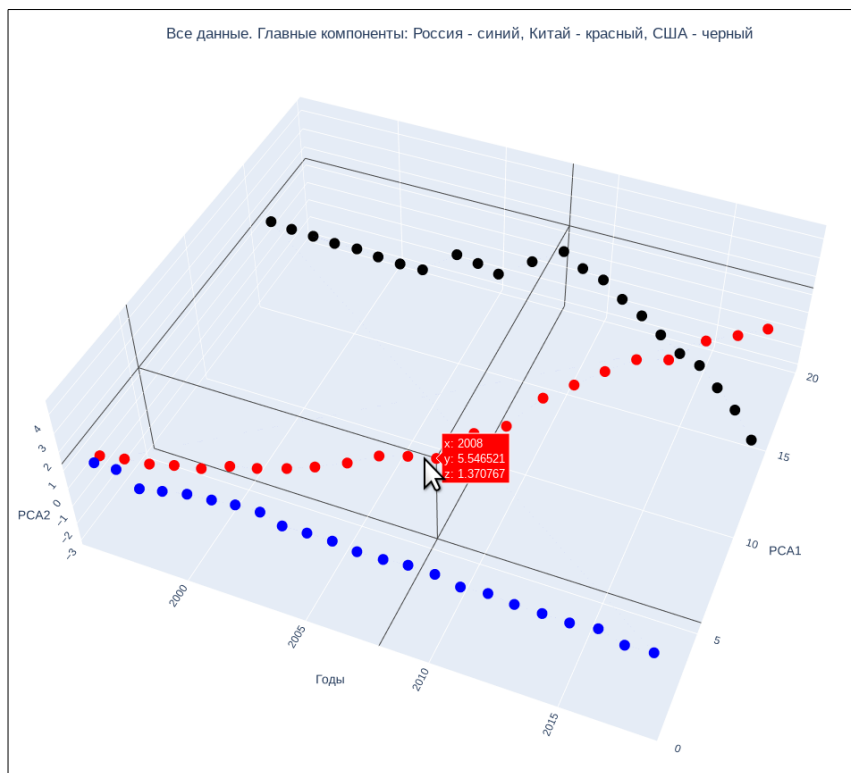


Рис. 43

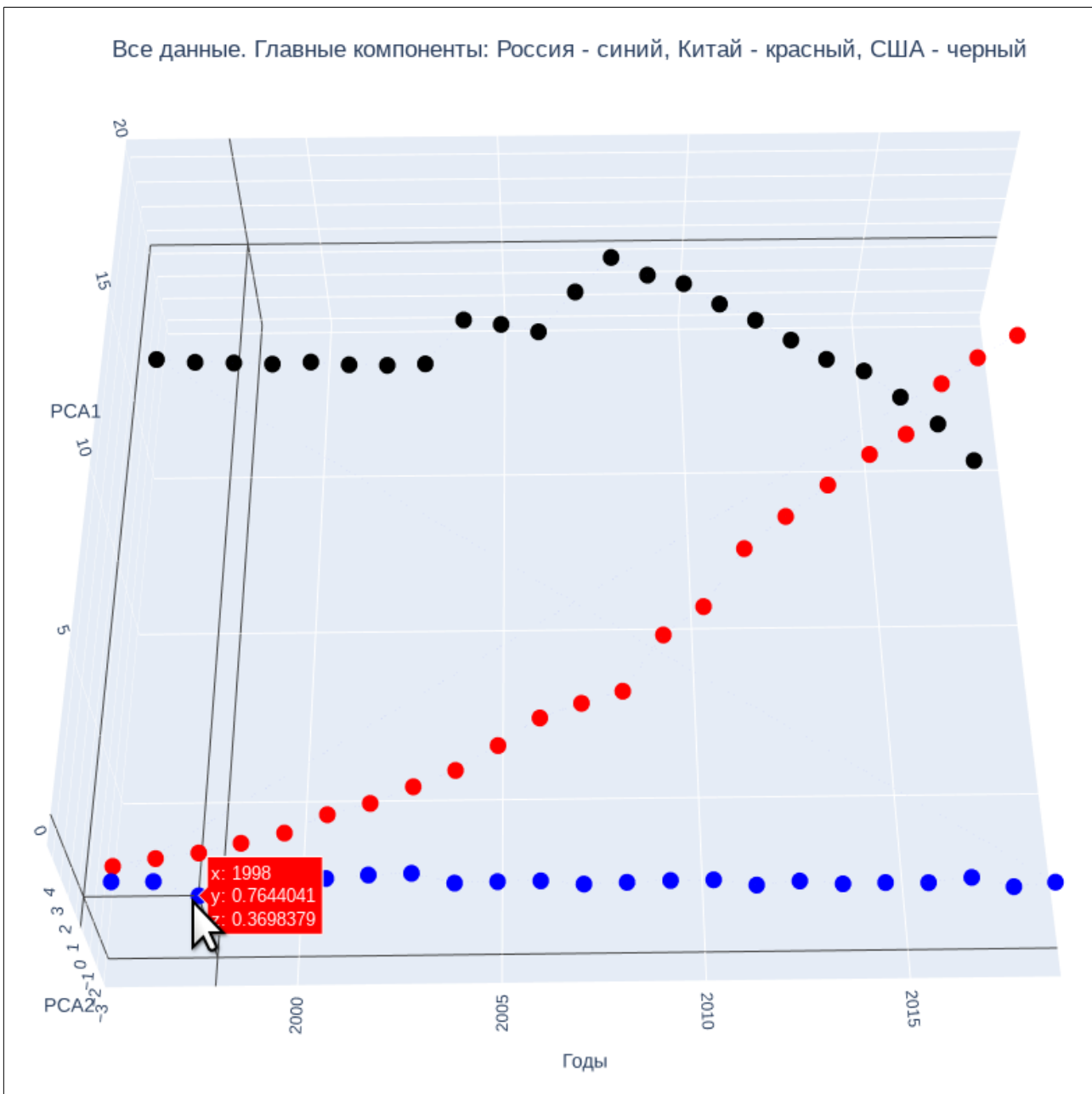


Рис. 44

На этом слайде отчетливо видны тенденции: Китай обгоняет США, если опираться на собранные в БД данные.

Но все же лучше Модель смотреть на сайте.

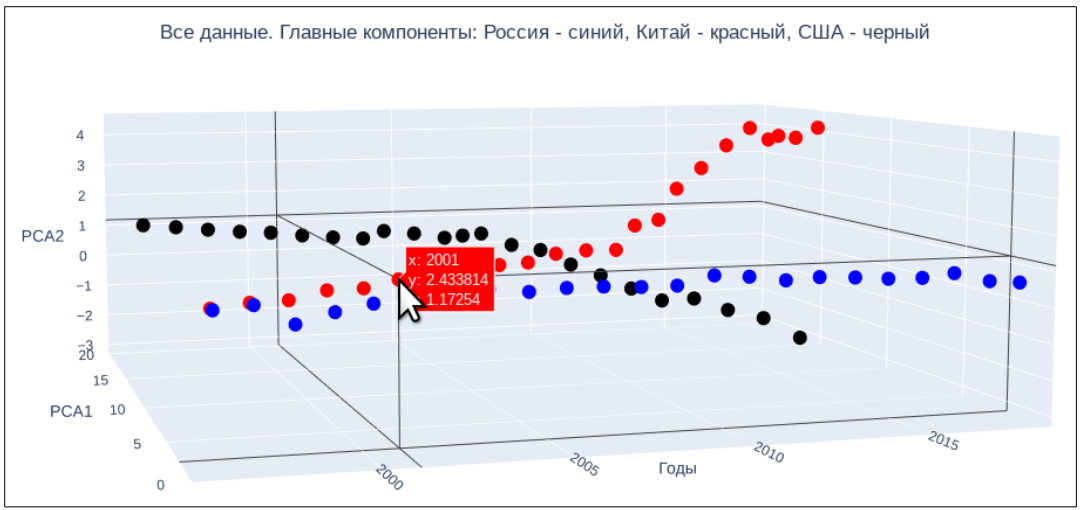


Рис. 45

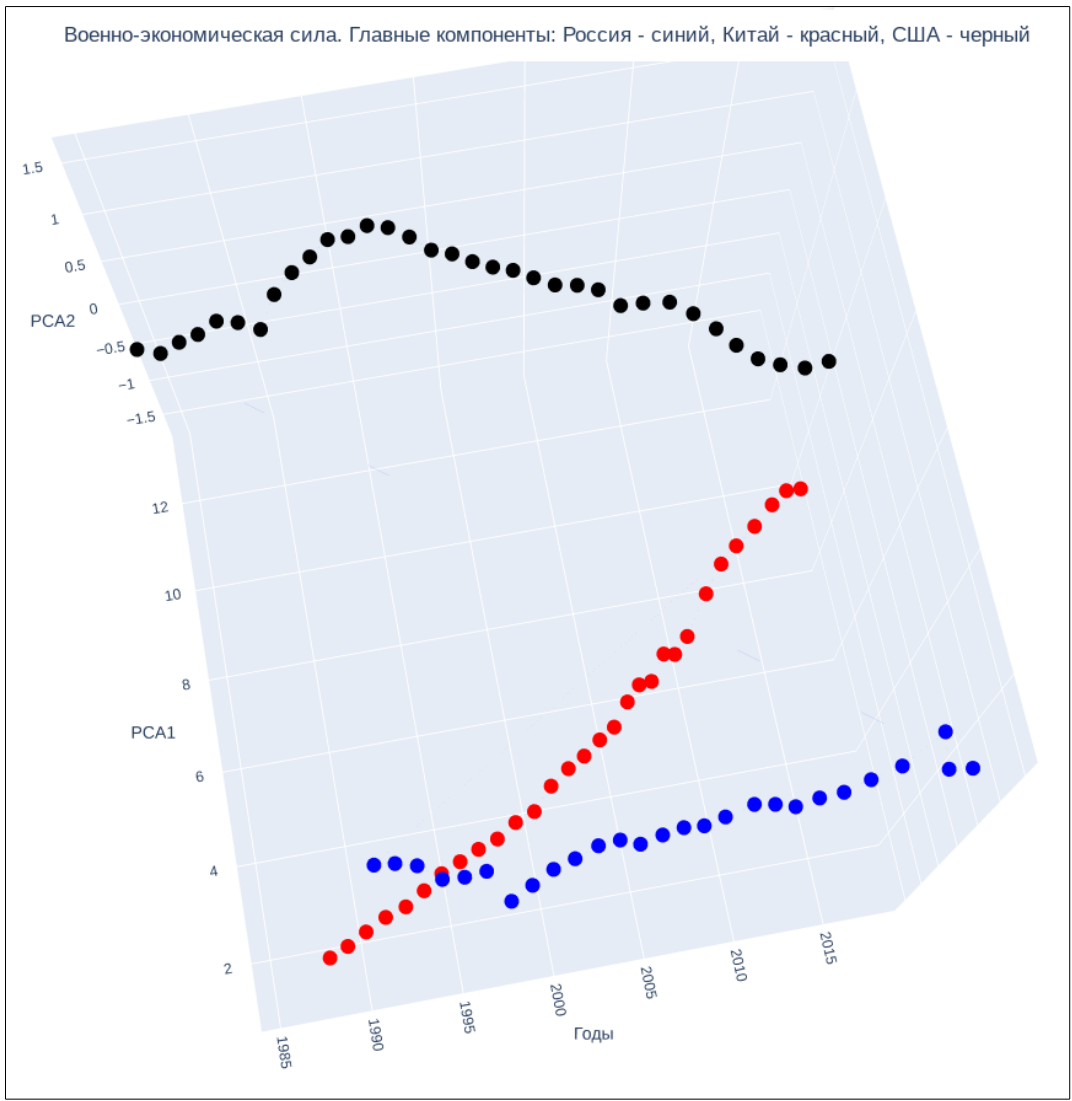


Рис. 46

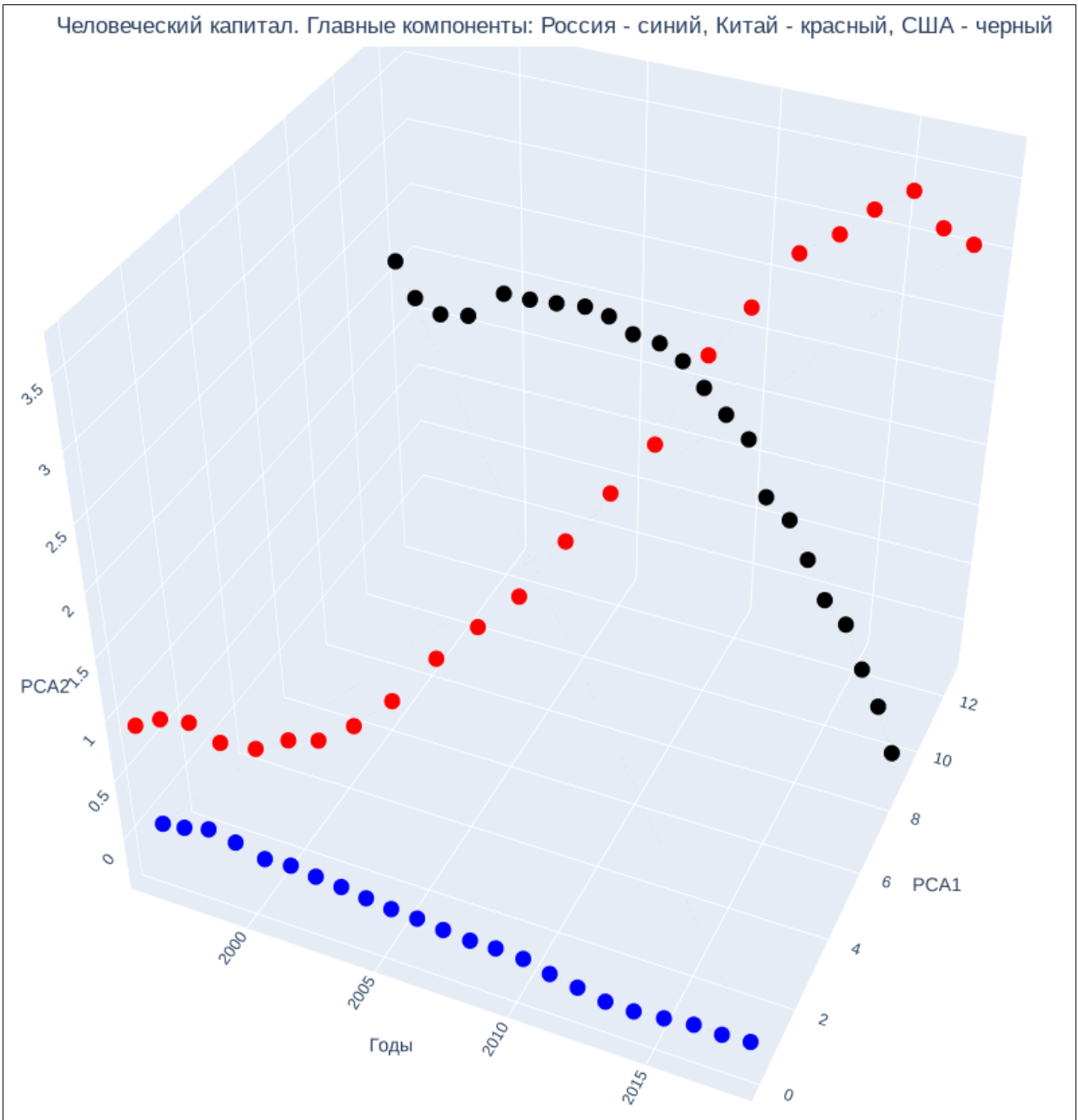


Рис. 47

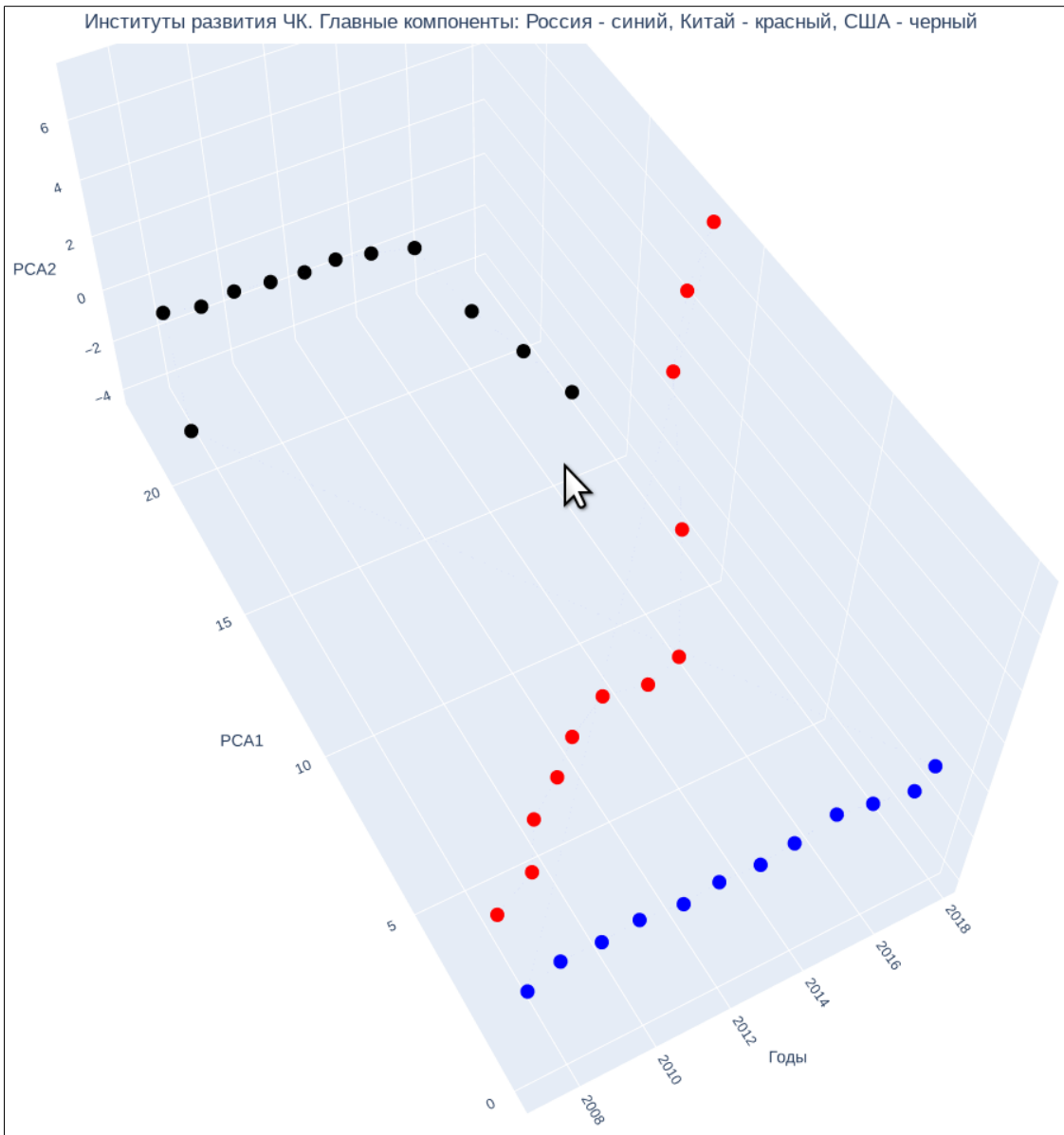
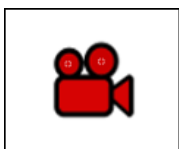


Рис. 48

Для самых ленивых на сайте есть ролик, который снят во время ручного исследования одной из моделей. Всего за одну минуту зритель получит представление о возможностях модели и, если преодолеет тяготение, может попробовать сам проиграть похожий сценарий.



Так выглядит ссылка на ролик.

Математическая модель

Введение

Данные организованы в три группы переменных (факторов): Военно-экономическая сила, Человеческий капитал и Институты развития ЧК. Каждая группа содержит несколько переменных. Для агрегирования переменных внутри каждой группы используется стандартный метод выделения главных компонент.

Пусть x, y, z – переменные из 1й, 2й и 3й группы соответственно (например, выделенные первые главные компоненты). Наш набор данных имеет структуру панельных данных (пространственно-временную структуру), каждую переменную индексируем двумя индексами: $x_{i,t}, y_{i,t}, z_{i,t}$, где i – номер страны (пространственный индекс), t – года (временной индекс, год).

Для исследования взаимного влияния переменных их трёх групп будем использовать **динамическую модель панельных данных**. Основные преимущества таких моделей состоит в возможности явно учитывать “индивидуальные особенности” стран

Модель состоит из трёх (одновременных) уравнений. Её спецификация имеет вид

$$\begin{aligned}x_{i,t} &= \gamma_1 x_{i,t-1} + \beta_{11} y_{i,t-1} + \beta_{12} z_{i,t-1} + \eta_i^{(1)} + u_{i,t} \\y_{i,t} &= \gamma_2 y_{i,t-1} + \beta_{21} x_{i,t-1} + \beta_{22} z_{i,t-1} + \eta_i^{(2)} + v_{i,t} \\z_{i,t} &= \gamma_3 z_{i,t-1} + \beta_{31} x_{i,t-1} + \beta_{32} y_{i,t-1} + \eta_i^{(3)} + w_{i,t}\end{aligned}$$

Важно обратить внимание, что в правой части переменные берутся с лагом по времени. В этой модели γ, β – коэффициенты, которые оцениваются по наблюдениям. $\eta_i^{(1)}, \eta_i^{(2)}, \eta_i^{(3)}$ – т.н. индивидуальные эффекты, возможно ненаблюдаемые. Они учитывают индивидуальные особенности стран. u, v, w – ненаблюдаемые шоки. Они агрегируют влияние остальных (неучтённых в модели) факторов.

Данная модель позволяет исследовать причинность по Гренджеру и прогнозировать с учётом кросс-зависимости между переменными. Метод оценивания уравнений модели: обобщённый метод моментов (GMM)

Эмпирический анализ

Вариант №1: рассмотрим базовый вариант, когда $x = pval11, y = pval12, z = pval13$ – первые главные компоненты в каждой группе.

Результаты приведены в таблице ниже. В скобках указаны стандартные ошибки коэффициентов, lag означает временной лаг.

	Зависимая переменная		
	pval1_1 (1)	pval1_2 (2)	pval1_3 (3)
lag(pval1_1, 1)	1.056*** (0.013)	0.144*** (0.040)	0.286 (0.345)
lag(pval1_2, 1)	-0.055*** (0.014)	0.937*** (0.020)	0.580* (0.339)
lag(pval1_3, 1)	-0.001 (0.004)	0.003 (0.002)	0.824*** (0.027)
Observations	56	56	56

Note: *p<0.1; **p<0.05; ***p<0.01

Выводы:

1. На первую группу значимо влияет только вторая, а третья группа незначимо
2. На вторую группу значимо влияет только первая группа, а третья влияет незначимо.
3. На третью группу вторая влияет значимо.
4. Вторая являются причинностью по Гренджеру для третьей.

Гренджер: Сэр Клайв Уильям Джон Грейнджер — английский экономист, лауреат Нобелевской премии по экономике 2003 «За разработку методов анализа экономических временных рядов с общими трендами».

Вариант №2: рассмотрим теперь модель для исходных данных. Предварительные тесты указывают на единичный корень для переменных. Следовательно, будем строить модель для первых разностей переменных (дифференцируем их). Результаты оценивания представлены в таблице ниже (*diff* означает дифференцирование, *lag* означает временной лаг, в скобках указаны стандартные ошибки коэффициентов).

	<i>Dependent variable:</i>											
	diff(log(val1))	diff(log(val5))	diff(log(val6))	diff(log(val8))	diff(log(val2))	diff(log(val4))	diff(log(val7))	diff(log(val22))	diff(log(val15))	diff(log(val16))	diff(log(val19))	diff(log(val12))
lag(diff(log(val1)), 1)	0.387*** (0.062)	0.906** (0.381)	-0.457 (0.361)	-0.112 (0.177)	0.001 (0.010)	-0.343 (0.498)	-0.030 (0.385)	0.0004 (0.002)	-0.408 (0.361)	-0.156 (0.205)	0.279** (0.139)	0.008 (0.006)
lag(diff(log(val5)), 1)	0.029 (0.026)	0.350* (0.204)	-0.915*** (0.204)	-0.044 (0.061)	0.001 (0.004)	0.467** (0.196)	-0.008 (0.155)	-0.001 (0.001)	0.050 (0.187)	0.037 (0.107)	0.054 (0.167)	-0.001 (0.001)
lag(diff(log(val6)), 1)	-0.020 (0.029)	-0.546*** (0.192)	0.891*** (0.272)	0.053 (0.079)	-0.002 (0.004)	-0.562** (0.246)	0.038 (0.165)	0.001 (0.001)	-0.295 (0.189)	-0.123 (0.134)	-0.0003 (0.143)	0.001 (0.001)
lag(diff(log(val8)), 1)	-0.016 (0.015)	-0.032 (0.051)	-0.037 (0.059)	-0.078 (0.051)	-0.002 (0.002)	-0.286 (0.257)	-0.084 (0.212)	-0.001 (0.001)	0.077 (0.084)	0.055 (0.052)	0.024 (0.083)	0.0001 (0.002)
lag(diff(log(val2)), 1)	-0.651** (0.308)	-0.039 (2.479)	-1.530 (1.682)	-0.435 (0.873)	0.015 (0.103)	4.043 (4.801)	1.624 (2.596)	-0.012 (0.011)	-0.051 (2.220)	1.779 (1.926)	-0.715 (1.510)	0.0002 (0.011)
lag(diff(log(val4)), 1)	-0.017*** (0.005)	0.018 (0.025)	-0.045* (0.024)	-0.017 (0.012)	0.003*** (0.001)	0.562*** (0.189)	0.068 (0.069)	-0.0005*** (0.0002)	-0.050 (0.043)	0.025 (0.039)	0.050 (0.031)	0.001 (0.0004)
lag(diff(log(val7)), 1)	-0.014*** (0.005)	0.019 (0.021)	-0.006 (0.016)	-0.010 (0.009)	-0.001 (0.001)	0.062 (0.062)	-0.187*** (0.065)	-0.0002 (0.0002)	-0.022 (0.038)	-0.034* (0.020)	-0.002 (0.021)	-0.0001 (0.0002)
lag(diff(log(val22)), 1)	3.603*** (0.980)	-14.958 (11.587)	2.124 (7.382)	5.151 (4.115)	0.347 (0.246)	32.885*** (9.801)	-15.335 (20.461)	0.928*** (0.044)	7.141 (5.706)	6.663** (3.009)	-3.088 (6.891)	0.098 (0.095)
lag(diff(log(val15)), 1)	-0.002 (0.007)	-0.003 (0.021)	-0.032 (0.034)	0.022 (0.015)	-0.0001 (0.001)	-0.108*** (0.036)	-0.028 (0.038)	0.0003 (0.0003)	0.251*** (0.087)	0.056 (0.045)	-0.006 (0.020)	0.001 (0.001)
lag(diff(log(val16)), 1)	-0.004 (0.011)	0.001 (0.030)	-0.019 (0.035)	0.031 (0.031)	0.004** (0.002)	0.072 (0.057)	0.013 (0.051)	0.001** (0.0005)	0.376*** (0.068)	0.618*** (0.051)	0.023 (0.037)	-0.0003 (0.001)
lag(diff(log(val19)), 1)	0.022 (0.014)	-0.084** (0.033)	0.018 (0.050)	-0.001 (0.027)	0.003 (0.002)	0.123 (0.091)	0.043 (0.084)	0.001 (0.001)	0.021 (0.045)	0.0001 (0.032)	0.138*** (0.044)	0.0001 (0.001)
lag(diff(log(val21)), 1)	-1.545* (0.832)	3.196 (5.869)	-1.762 (5.458)	-0.020 (2.102)	-0.112 (0.083)	2.268 (7.739)	-2.256 (10.501)	-0.012 (0.019)	1.282 (5.378)	-1.420 (2.971)	-1.722 (4.137)	0.560*** (0.066)
Observations	63	63	63	63	63	63	63	63	63	63	63	63

Note:

*p<0.1; **p<0.05; ***p<0.01

Выводы:

Анализ имеющихся в БД панельных данных показывает:

1. val21 (Общее население) причина по Гренджеру для val1 (ВВП ППС).
2. val4 (Цитирование публикаций) причина по Гренджеру для val6 (Военные расходы, млн \$).
3. val7 (Патенты) причина по Гренджеру для val16 (Абоненты мобильной сотовой связи).
4. Между остальными причинности не обнаружены.

Пояснения к данным вариантов 1 и 2

Подготовленные данные можно получить с сайта <http://eurasian-defence.ru/richdb09/>

[Первичные данные \(csv.zip\)](#)

[Нормализованные данные \(csv.zip\)](#)

[Главные компоненты \(csv.zip\)](#)

[Скрипт расчета мат.моделей \(R.zip\)](#)

Нужно иметь в виду, что нумерация полей в CSV-файлах следует не прямой последовательности "как на сайте", а берётся из справочной таблицы:

№ группы sgid	Имя группы sgname	№ источника sid	Имя источника sname
1	Военно-экономическая сила	1	ВВП по паритету покупательной способности
1	Военно-экономическая сила	5	Военные расходы, % от ВВП, SIPRI
1	Военно-экономическая сила	6	Военные расходы, млн \$ 2017 г., SIPRI
1	Военно-экономическая сила	8	Выработка электроэнергии
2	Человеческий капитал	2	Индекс человеческого развития, ПРООН
2	Человеческий капитал	4	Цитирование публикаций, Scopus
2	Человеческий капитал	7	Патенты
2	Человеческий капитал	22	Доля населения трудоспособного возраста 15-64 лет
3	Институты развития ЧК	15	Пользователи Интернета, ITU
3	Институты развития ЧК	16	Абоненты мобильной сотовой связи, ITU
3	Институты развития ЧК	19	Расходы на исследования и развитие (R&D), UNESCO
3	Институты развития ЧК	21	Общее население независимо от гражданства/статуса

Например:

в таблице **richard_value.csv** поля расшифровываются так:

val1 - ВВП ППС

val2 - ИРЧП, Human Development (а вовсе не Военные расходы!)

val3 - Публикации, Scopus

val4 - Цитирование публикаций

val5 - Военные расходы, % от ВВП, SIPRI

и т.д.

в таблице **richard_pca.csv**

rval1_0 - 1-я ГК для всех данных БД

rval2_0 - 2-я ГК для всех данных БД

rval1_1 - 1-я ГК для группы № 1 "Военно-экономическая сила"

rval2_1 - 2-я ГК для группы № 1 "Военно-экономическая сила"

rval1_2 - 1-я ГК для группы № 2 "Человеческий капитал"

rval2_2 - 2-я ГК для группы № 2 "Человеческий капитал"

и т.д.

в таблице **richard_norm.csv**

nval1 - ВВП ППС

nval2 - ИРЧП, Human Development (а не Военные расходы!)

nval3 - Публикации, Scopus

и т. д.

Вариант 3: Нелинейная модель

Модель состоит из трёх (одновременных) уравнений. Её спецификация модели имеет вид:

$$x_{i,t} = \gamma_{11} x_{i,t-1} + \gamma_{12} x_{i,t-1}^2 + \beta_{11} y_{i,t-1} + \beta_{12} z_{i,t-1} + \delta_{11} y_{i,t-1}^2 + \delta_{12} z_{i,t-1}^2 + \eta^{(1)}_i + u_{i,t}$$

$$y_{i,t} = \gamma_{21} y_{i,t-1} + \gamma_{22} y_{i,t-1}^2 + \beta_{21} x_{i,t-1} + \beta_{22} z_{i,t-1} + \delta_{21} x_{i,t-1}^2 + \delta_{22} z_{i,t-1}^2 + \eta^{(2)}_i + v_{i,t}$$

$$z_{i,t} = \gamma_{31} z_{i,t-1} + \gamma_{32} z_{i,t-1}^2 + \beta_{31} x_{i,t-1} + \beta_{32} y_{i,t-1} + \delta_{31} x_{i,t-1}^2 + \delta_{32} y_{i,t-1}^2 + \eta^{(3)}_i + w_{i,t}$$

Важно обратить внимание, что в правой части переменные берутся с лагом по времени. В этой модели γ, β, δ – коэффициенты, которые оцениваются по наблюдениям. $\eta^{(1)}, \eta^{(2)}, \eta^{(3)}$ – т.н. индивидуальные эффекты, возможно ненаблюдаемые. Они учитывают индивидуальные особенности стран. u, v, w – ненаблюдаемые шоки. Они агрегируют влияние остальных (неучтённых в модели) фактором.

В Таблице ниже приведены результаты расчётов (в скобках указаны стандартные ошибки коэффициентов):

	Зависимая переменная		
	pval1_1 (1)	pval1_2 (2)	pval1_3 (3)
lag(pval1_1)	0.834*** (0.051)	0.146 (0.155)	-0.119 (0.159)
lag(pval1_1^2)	0.012*** (0.002)	0.005 (0.006)	0.005 (0.008)
lag(pval1_2)	0.123*** (0.025)	1.053*** (0.074)	-0.014 (0.205)
lag(pval1_3)	0.015 (0.009)	0.033** (0.015)	0.776*** (0.044)
lag(pval1_2^2)	-0.010*** (0.001)	-0.013** (0.007)	0.003 (0.013)
lag(pval1_3^2)	-0.001 (0.002)	0.001 (0.007)	0.001 (0.016)
Observations	56	56	56
Note:	* p<0.1; ** p<0.05; *** p<0.01		

Выводы:

1. Н первую группу значимо влияет вторая группа.
2. На вторую группу значимо влияет третья группа.
3. На третью группу значимо не влияет ни одна из групп.
4. Вторая группа является причиной по Гренджеру для первой.
5. Третья группа является причиной по Гренджеру для второй.

Пояснения к данным варианта 3

Подготовленные данные можно получить с сайта <http://eurasian-defence.ru/richdb/>

[Первичные данные \(csv.zip\)](#)

[Нормализованные данные \(csv.zip\)](#)

[Главные компоненты \(csv.zip\)](#)

[Скрипт расчета модели \(R.zip\)](#)

Справочная таблица

№ группы sgid	Имя группы sgname	№ источника sid	Имя источника sname
1	Военно-экономическая сила	1	ВВП по паритету покупательной способности
1	Военно-экономическая сила	5	Военные расходы, % от ВВП, SIPRI
1	Военно-экономическая сила	6	Военные расходы, млн \$ 2017 г., SIPRI
1	Военно-экономическая сила	8	Выработка электроэнергии
2	Человеческий капитал	2	Индекс человеческого развития, ПРООН
2	Человеческий капитал	3	Публикации, Scopus
2	Человеческий капитал	4	Цитирование публикаций, Scopus
2	Человеческий капитал	7	Патенты
3	Институты развития ЧК	19	Расходы на исследования и развитие (R&D), UNESCO
3	Институты развития ЧК	24	Ожидаемая продолжительность образования на момент поступления в школу
3	Институты развития ЧК	25	Средняя продолжительность образования населения старше 25 лет
3	Институты развития ЧК	26	Ожидаемая продолжительность жизни на момент рождения

Выводы по итогам данного этапа исследования

Гипотеза, согласно которой определяющей группой факторов является группа «Институты развития человеческого капитала» (группа № 3), от этой группы факторов зависит «Качество человеческого капитала» (группа № 2), а «Экономическая и военная мощь» (группа № 1) является производной от группы № 2, статистическим анализом подтверждается с использованием нелинейной модели (вариант расчета №3).

Проявляется также зависимость между группами факторов причинности по Гренджеру.

Возможные причины неполного соответствия результатов анализа в вариантах 2 и 3 гипотезе о зависимостях групп факторов: неточно подобранные для исследования факторы и недостаточно полные данные (особенно это касается группы факторов № 3 «Институты развития ЧК»).

Анализ графиков динамики изменений Главных компонент для трех групп факторов в варианте №3 позволяет утверждать, что гипотеза также подтверждается: на рис. 49-50

отчетливо видно, что росту Человеческого капитала предшествует рост Институтов развития ЧК.

Зависимость Военно-экономической силы от ЧК менее очевидна для США. Видимо кроме исследованных были и другие факторы повлиявшие на их падение начавшееся в 1997 году. Либо рост факторов групп 2 и 3 оказался недостаточным для столь огромной и значит инерционной экономики, как США.

Графики для США

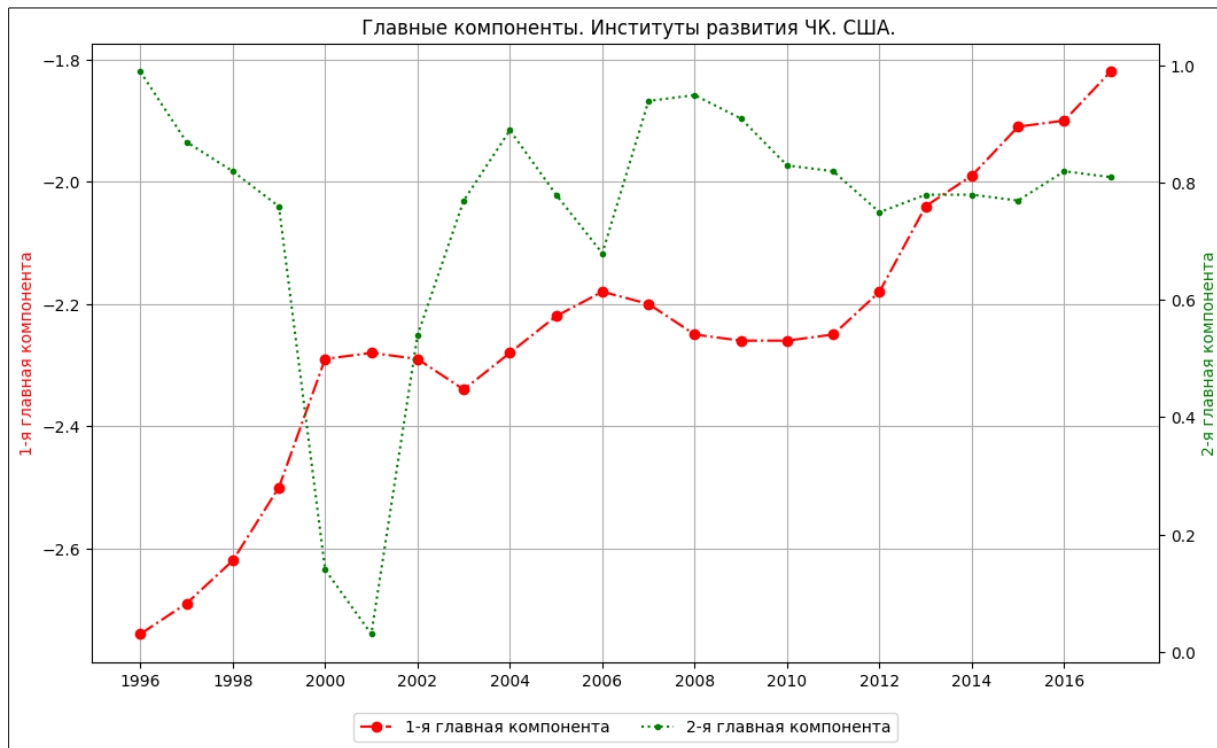


Рис. 49

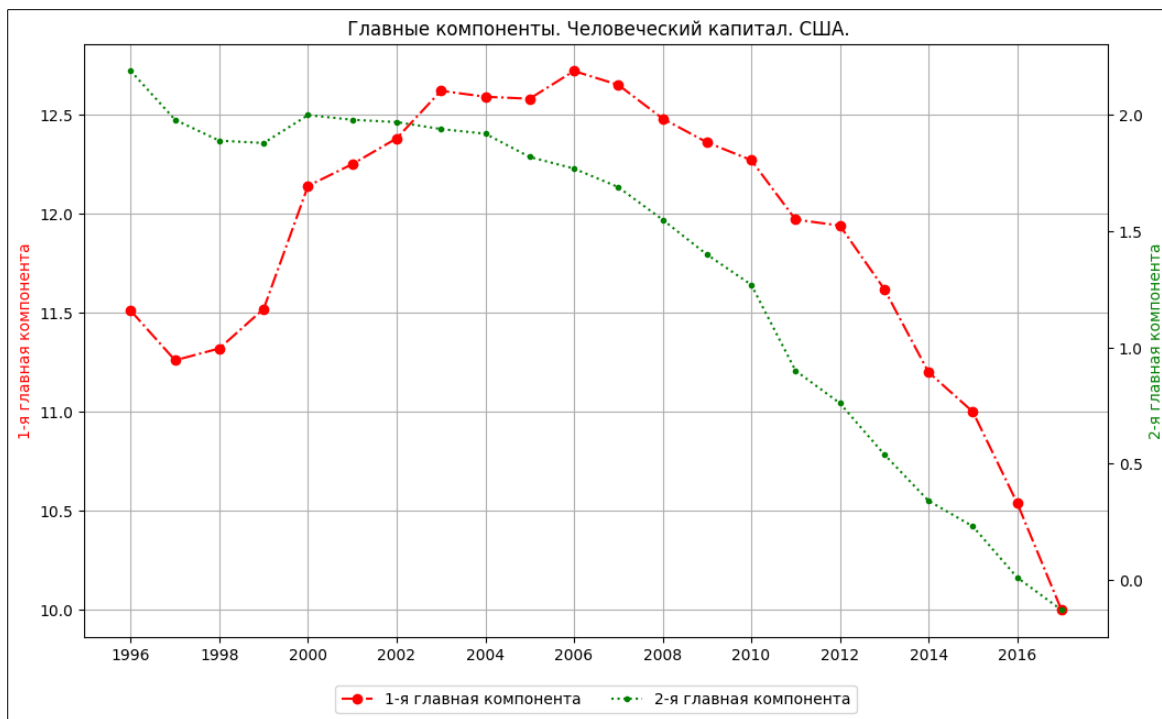


Рис.50

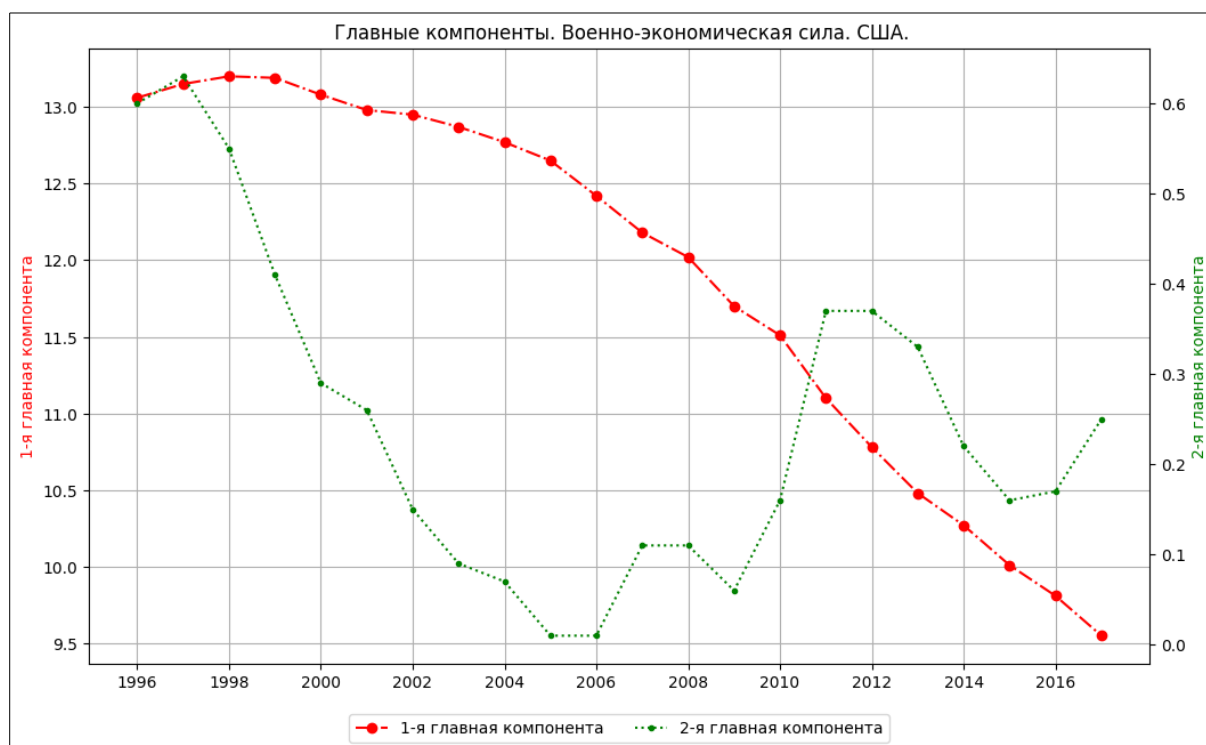


Рис. 51

Для неприсоединившихся стран таких как Швеция и Финляндия, у которых милитаристская составляющая истеблишмента не является столь влиятельной, как с США, зависимость групп факторов на графиках более выражена (рис.52-54).

Графики для Финляндии

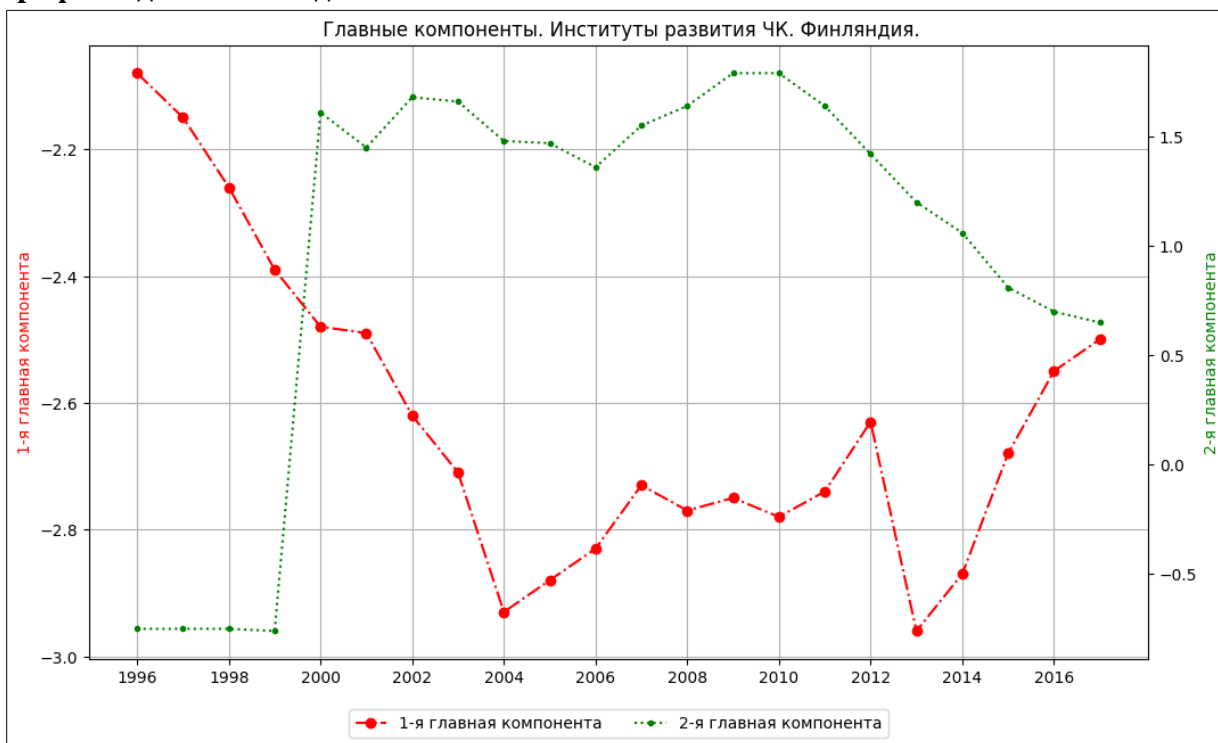


Рис. 52

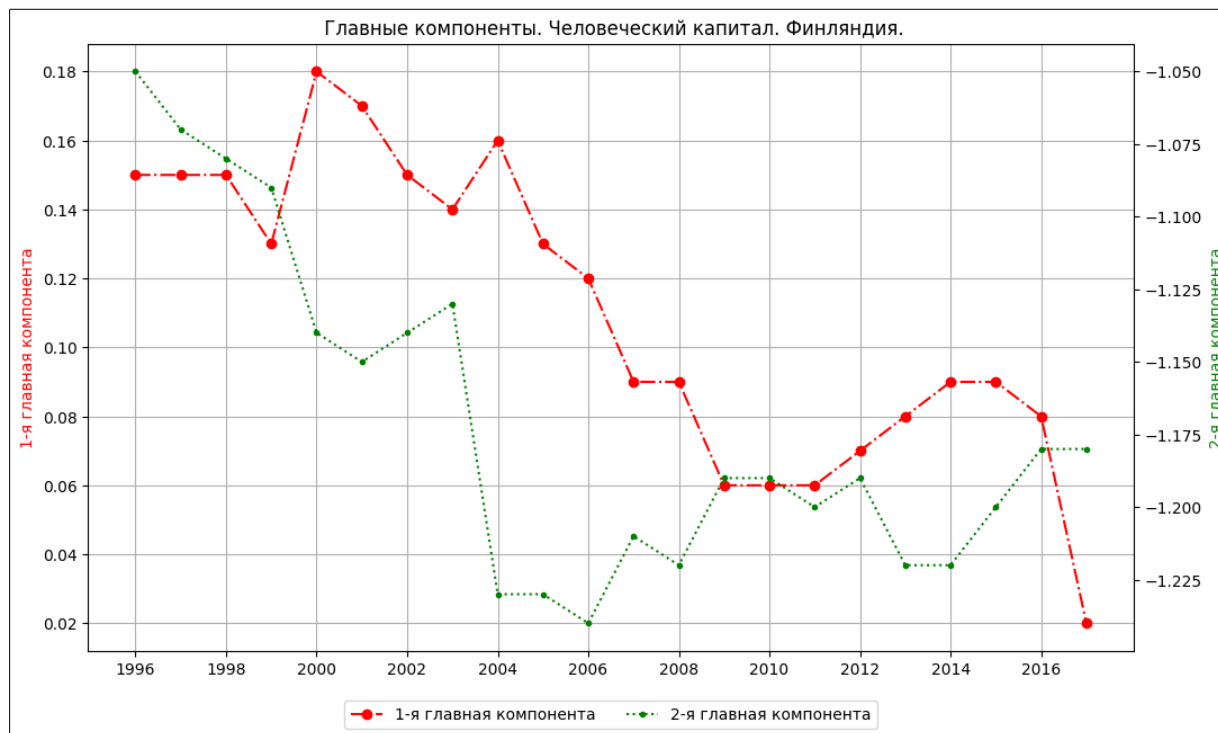


Рис. 53

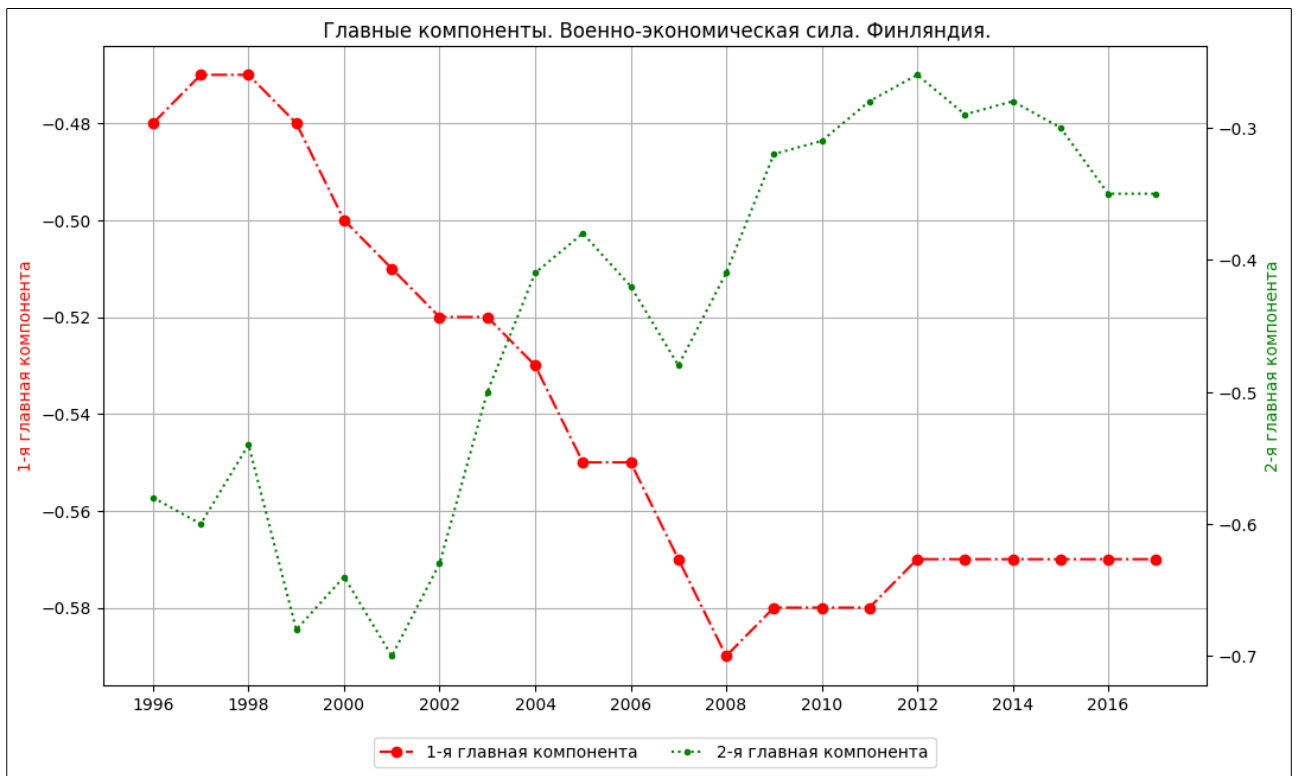


Рис. 54

Для стран из нижней части таблицы откуда взяты все эти графики выражены тенденции медленной, последовательной деградации всех групп факторов (рис.55-57).

Графики для Болгарии

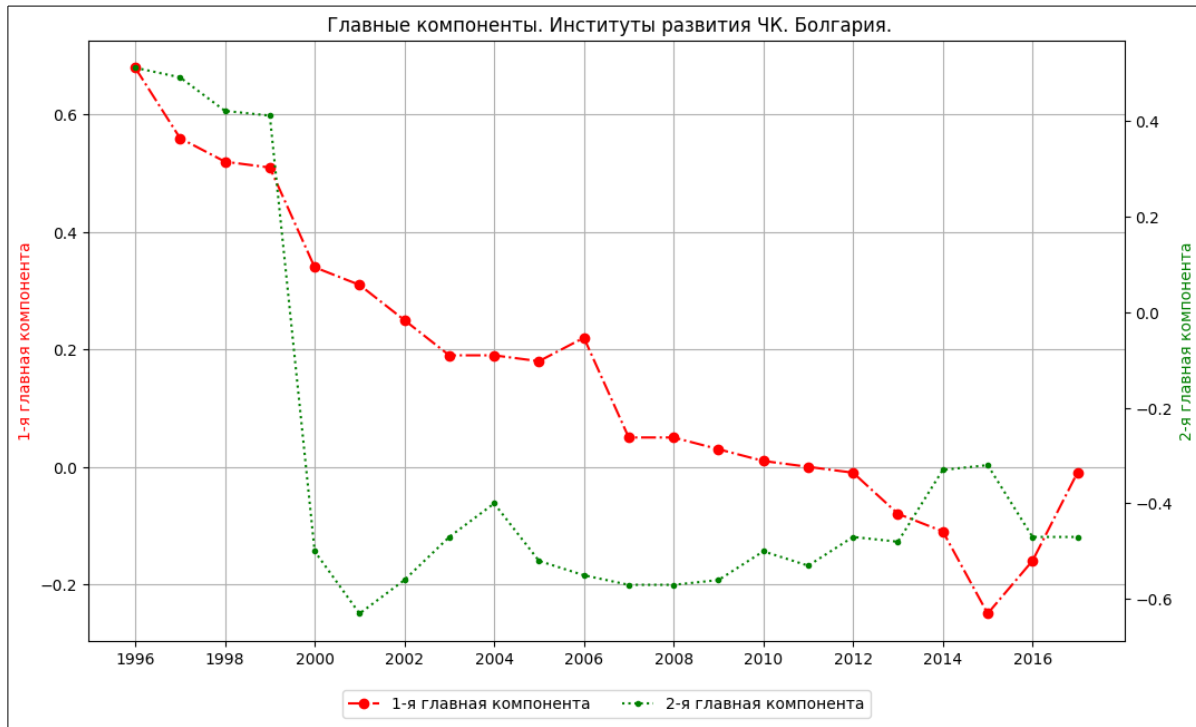


Рис. 55

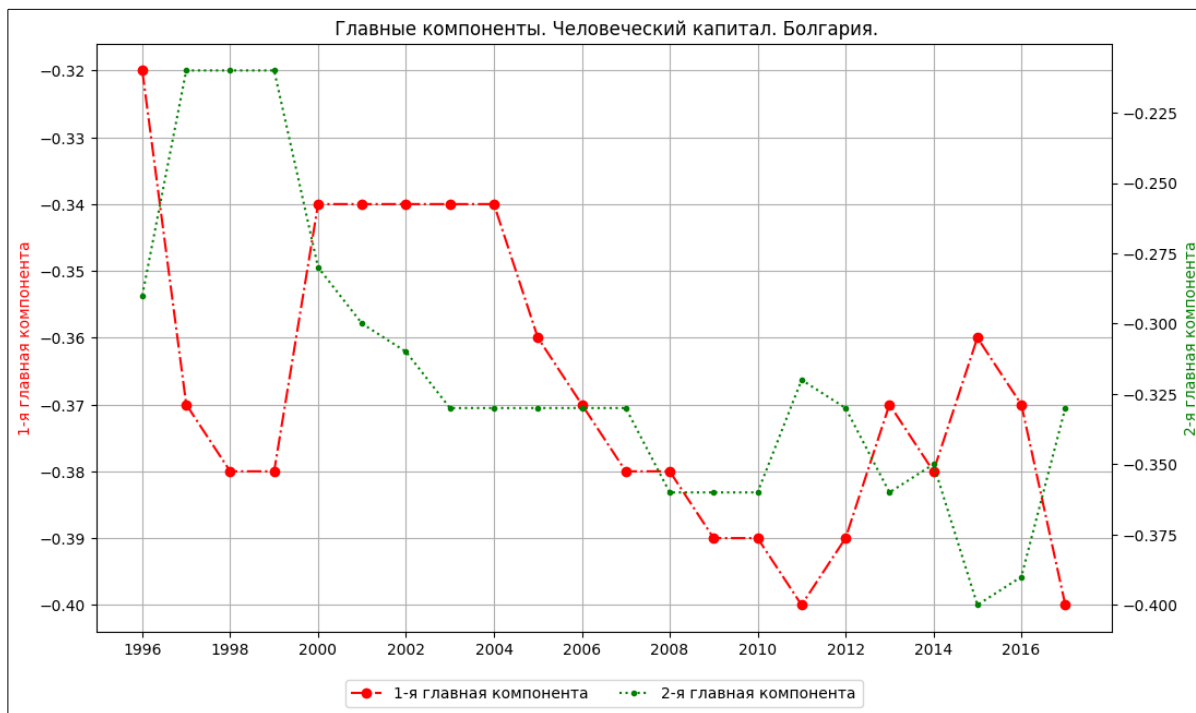


Рис. 56

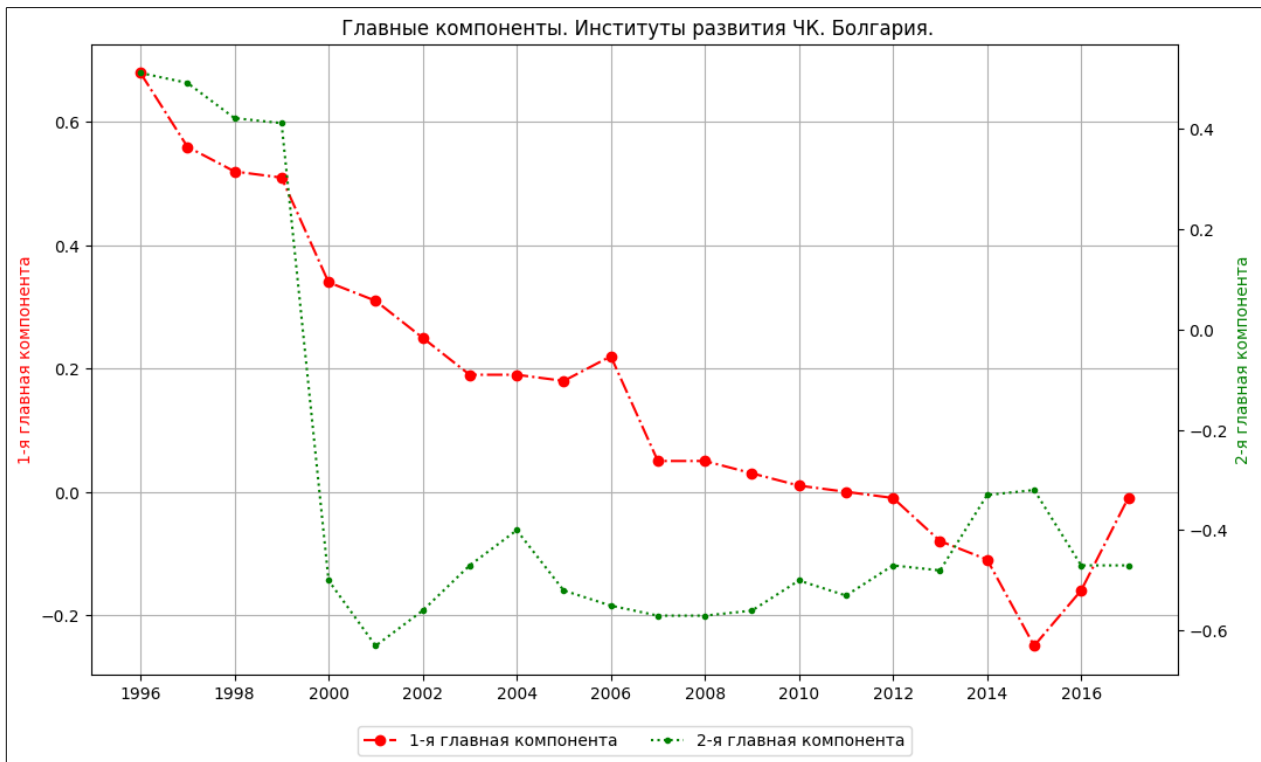


Рис. 57

Здесь видны кратковременные всплески, обусловленные, видимо, размерами помощи от Евросоюза, однако общую тенденцию это может лишь замедлить.

Прогнозирование

Прогнозы на основе математических моделей, а тем более рекомендации политическому руководству страны, должны быть многократно подтверждены альтернативными исследованиями, способами и источниками.

Очень осторожный прогноз-гипотеза на основании данного математического и программного аппарата может быть таким: при увеличении (усиленном финансировании) факторов относящихся к третьей группе (Институты развития ЧК), даже в ущерб первой группе (Военно-экономическая сила), будет обеспечен рост показателей второй группы (Человеческий капитал), который является фундаментальной основой для роста Военно-экономических показателей.

Рекомендации по продолжению исследований

Необходимы дальнейшие исследования, прежде всего полевые по сбору данных, после чего необходимо провести повторный эмпирический анализ для линейной и нелинейной моделей.