
Всему свое время:
время обнимать и
время объятий
сторониться
(книга Экклезиаста, 3)

Ф.Т. Алескеров
НИУ ВШЭ, ИПУ РАН

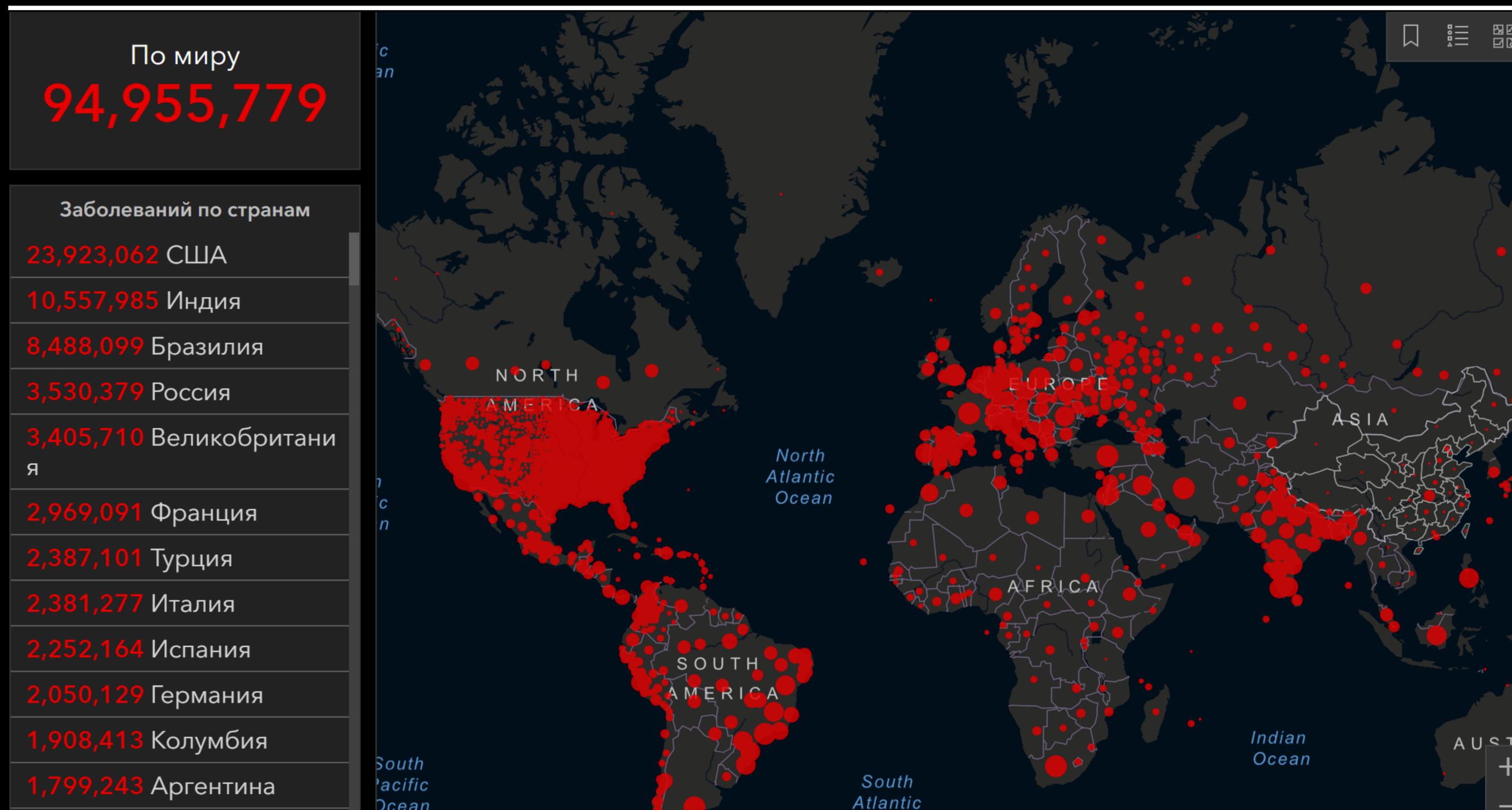
Выступление на УС ИПУ РАН, 25.01.2021

Соавторы

С. Демин, А. Мячин, В. Якуба

(НИУ ВШЭ, ИПУ РАН)

Мировая статистика по COVID-19



COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE)
at Johns Hopkins University (JHU): <https://coronavirus.jhu.edu/map.html>

Дата обращения: 18.01.2021

Используемые данные

Источник: The Oxford COVID-19 Government Response Tracker

Ежедневные данные с 1 января 2020 года

Данные по более чем 180 странам

<https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>



UNIVERSITY OF
OXFORD

- *Степень введения отдельных карантинных мер*
- *Количество смертей*
- *Количество зараженных*
- *Законопослушность*

Используемые данные

Степень введения отдельных карантинных мер

C1. Закрытие школ

C2. Закрытие предприятий

C3. Отмена публичных мероприятий

C4. Ограничение собраний

C5. Отмена общественного транспорта

C6. Требование не покидать жилище

C7. Ограничение на поездки внутри страны

C8. Ограничение на поездки за рубеж

Используемые данные

Законопослушность

*До какой степени граждане выполняют указания правительства
(данные основаны на экспертных оценках)*

Модель агрегирования уровня карантинных мер по коронавирусу

Критерий	Число градаций
Закрытие школ (с1)	4
Закрытие предприятий (с2)	4
Отмена публичных мероприятий (с3)	3
Ограничение собраний (с4)	5
Отмена общественного транспорта (с5)	3
Требование не покидать жилище (с6)	4
Ограничение на поездки внутри страны (с7)	3
Ограничение на поездки за рубеж (с8)	5

Модель агрегирования уровня карантинных мер по коронавирусу

Пороговая модель: Пример для 3-градационных оценок

	c1	c2	c3
Страна А	2	2	2
Страна В	3	3	1

A>B

Порог: альтернатива со всеми средними оценками лучше, чем альтернатива с хотя бы одной плохой оценкой, даже если все остальные оценки наилучшие.

Модель агрегирования уровня карантинных мер по коронавирусу

Агрегирование карантинных мер по нескольким критериям производится с помощью порогового правила, которое представляет собой лексикографическое упорядочение количества значений критериев. Страны сравниваются попарно, начиная с критериев, имеющих наименьшее значение, равное 0.

Страна, в которой число критериев с нулевыми значениями меньше, доминирует в парном сравнении с другой страной, в которой число критериев с нулевыми значениями больше.

Если для пары стран число критериев с нулевыми значениями равно, то сравнивается число критериев, имеющих значение 1, и т.д., до максимальных значений критериев.

Агрегированные карантинные меры содержат 468 градаций и нормированы на $[0;1]$

Модель агрегирования уровня карантинных мер по коронавирусу

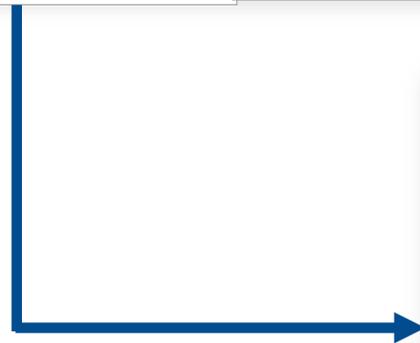
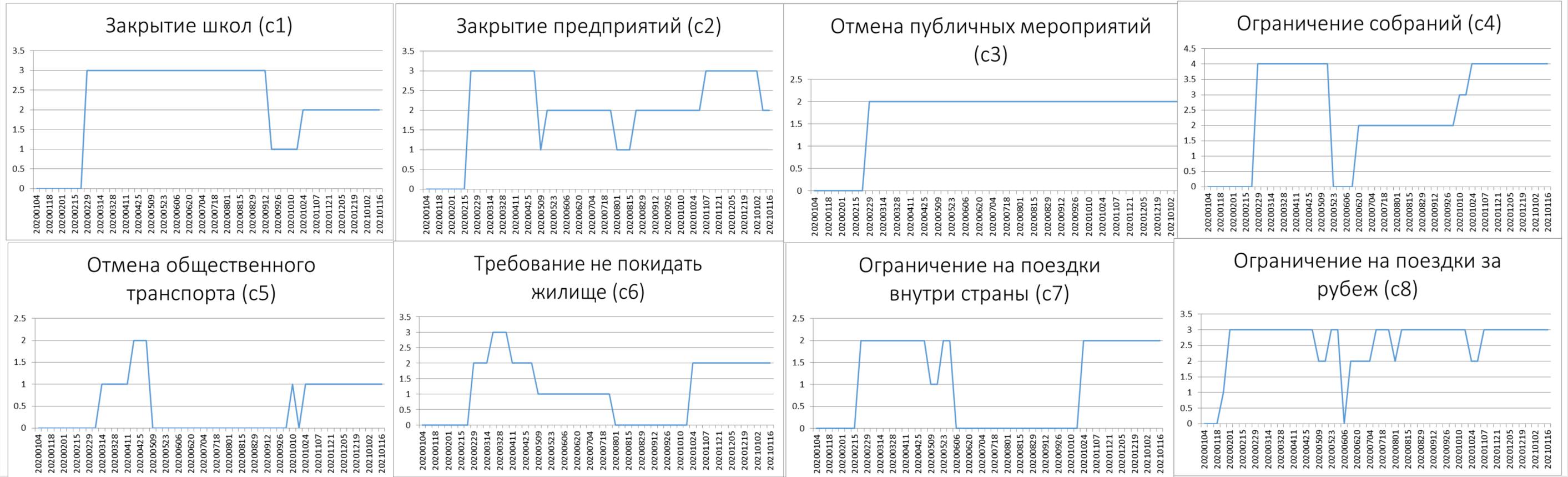
Запреты публичных мероприятий и собраний

- C2: Закрытие предприятий*
- C3: Отмена публичных мероприятий*
- C4: Ограничение собраний*
- C6: Требование не покидать жилище*

Государственные запреты

- C1: Закрытие школ*
- C5: Отмена общественного транспорта*
- C7: Ограничения на поездки внутри страны*
- C8: Ограничения на поездки за рубеж*

Карантинные меры (Италия)



Карантинные меры (Италия)



Карантинные меры: собрания



Агрегированные меры: собрания



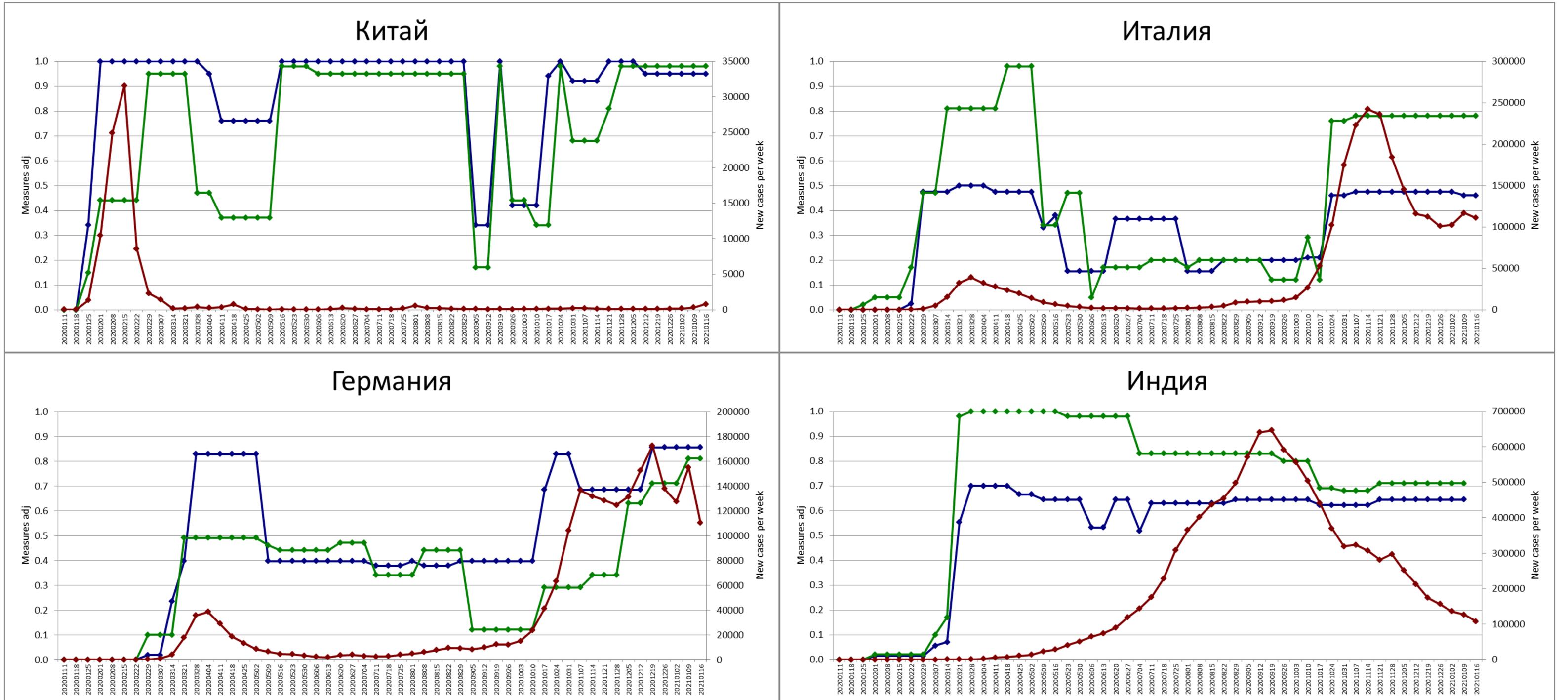
Карантинные меры: гос. запреты



Агрегированные меры: гос. запреты



Агрегированные меры и новые случаи заболеваний (понеделно)



Число новых заболеваний



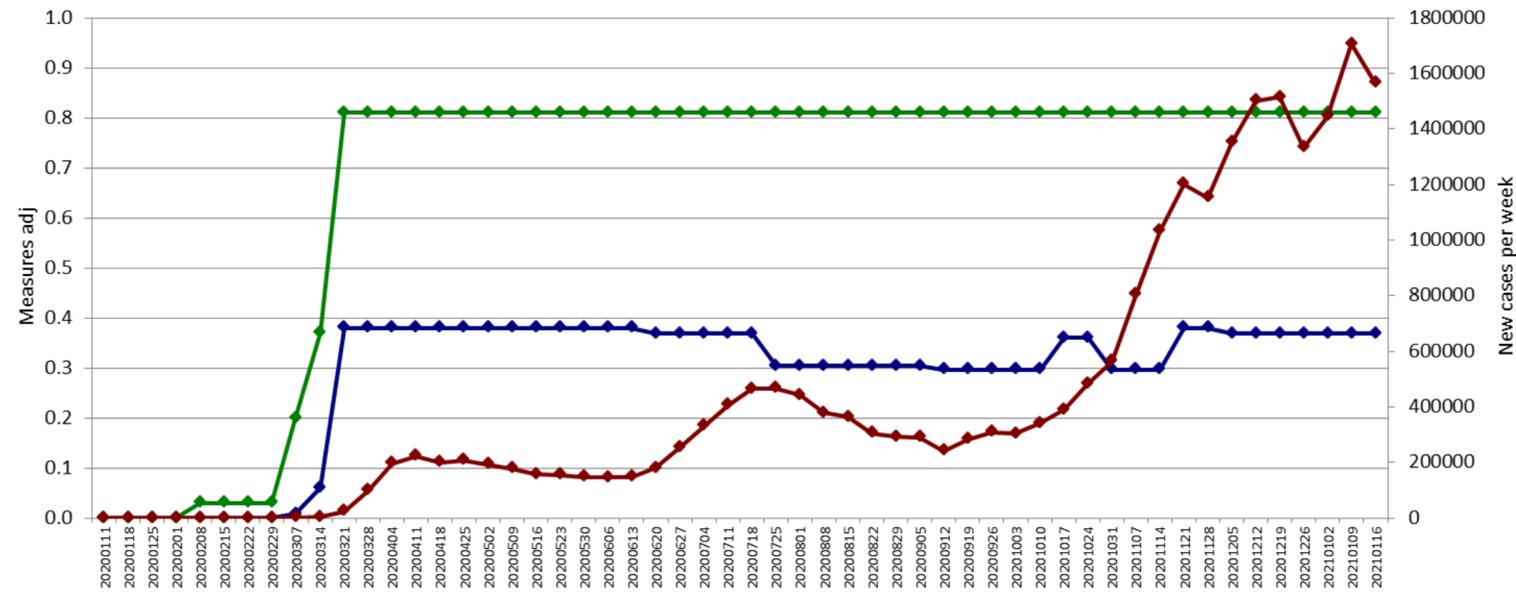
Меры: собрания



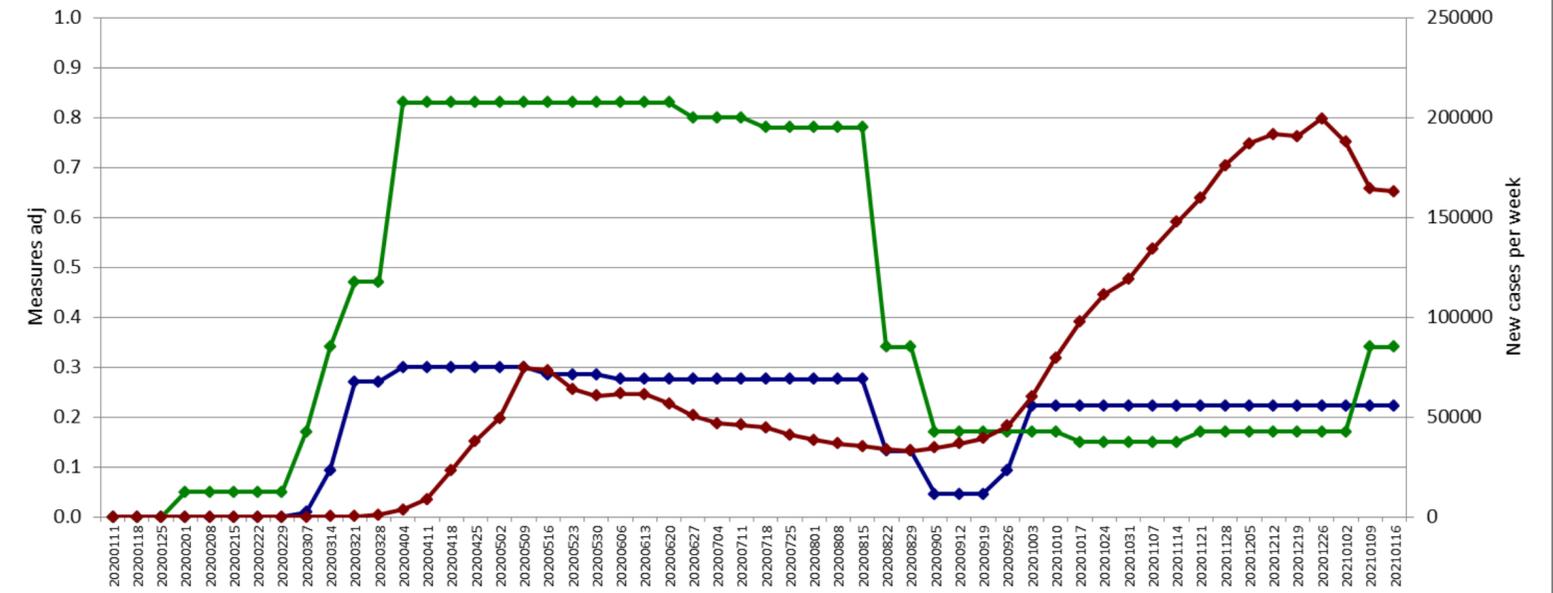
Меры: гос. запреты

Агрегированные меры и новые случаи заболеваний (понеделно)

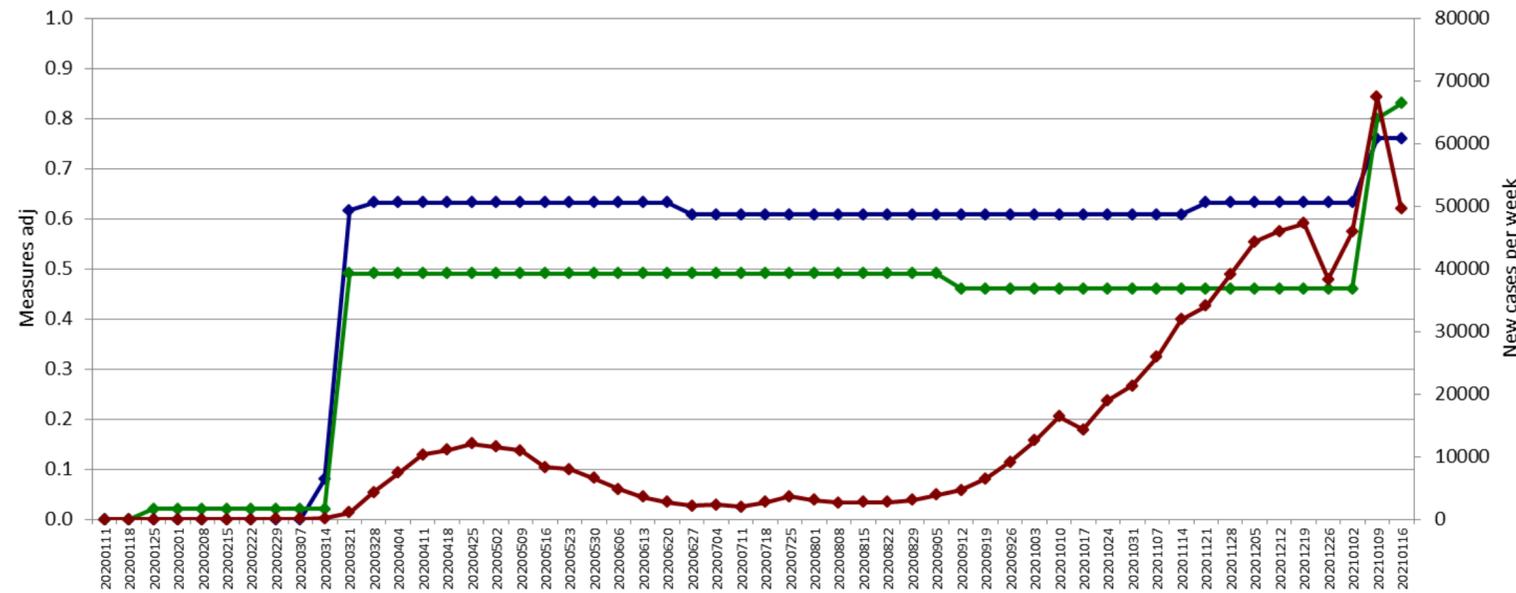
США



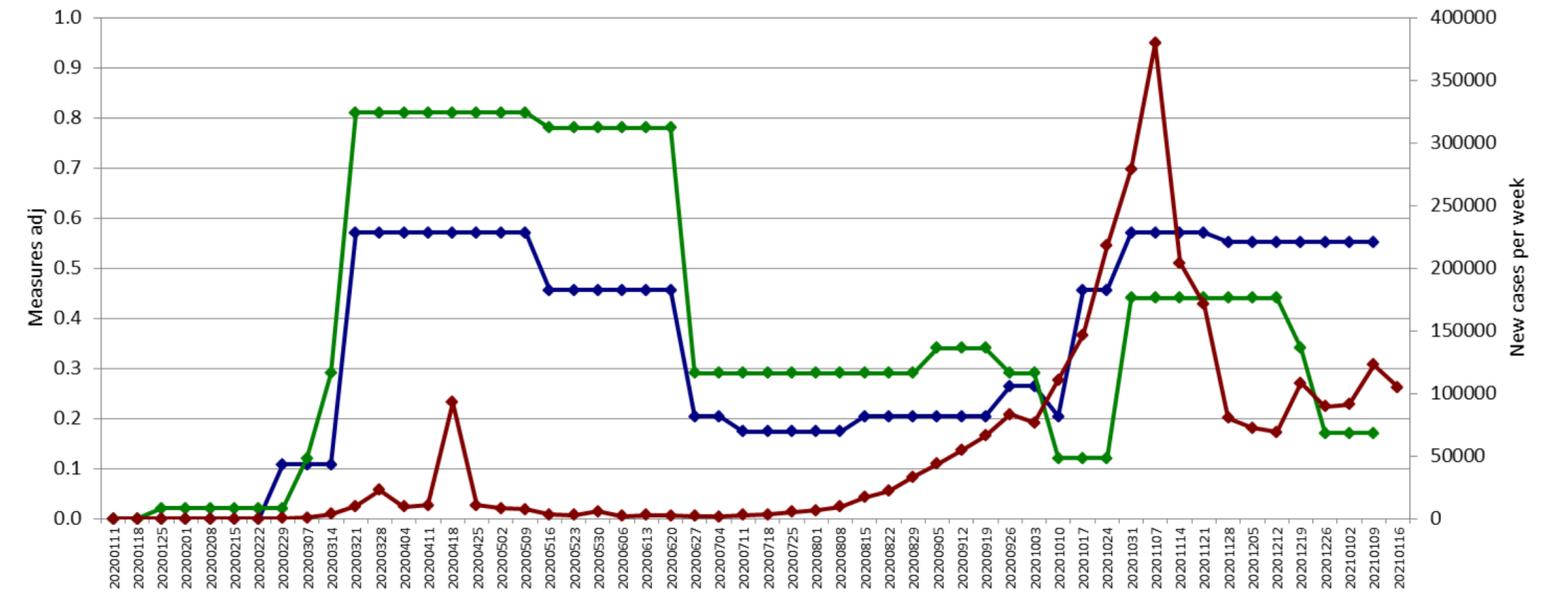
Россия



Канада



Франция

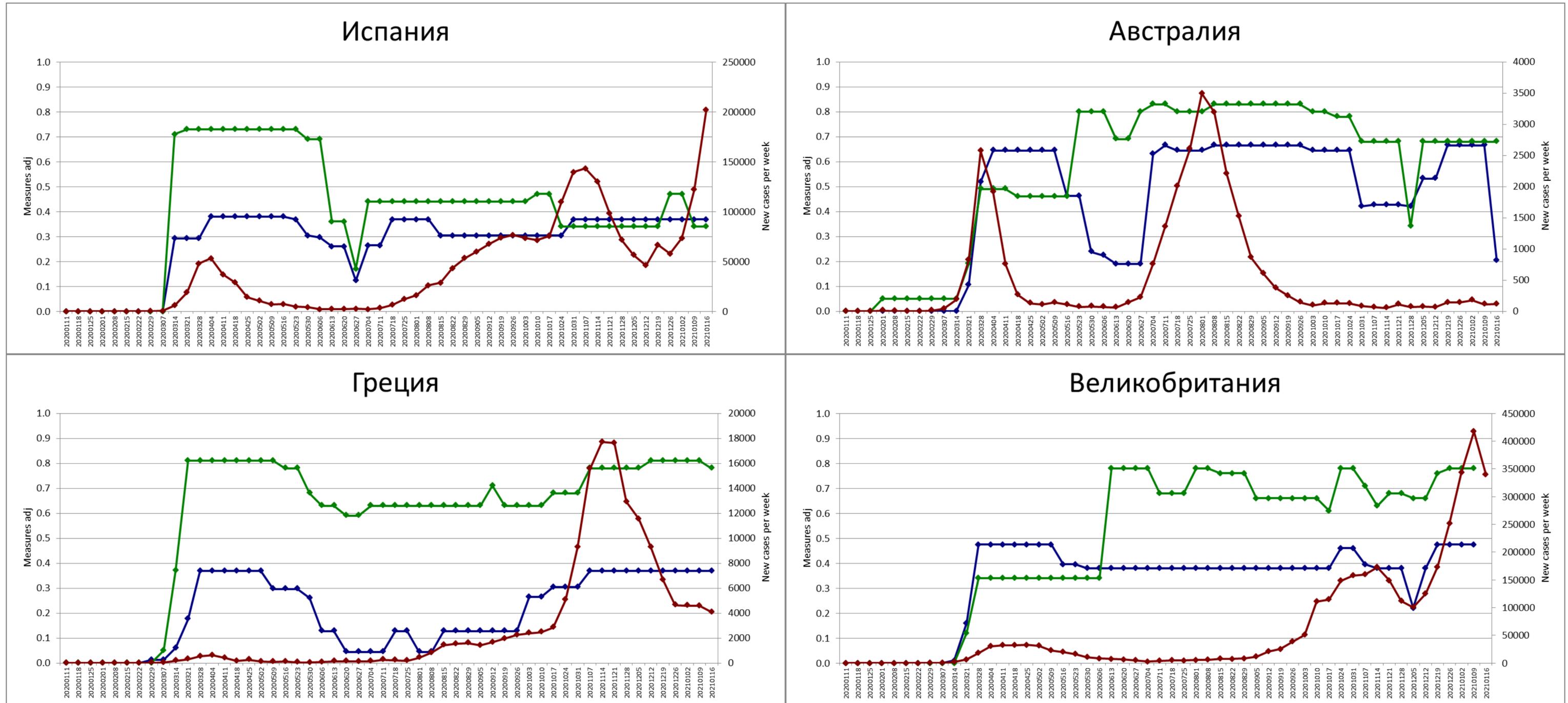


Число новых заболеваний

Меры: собрания

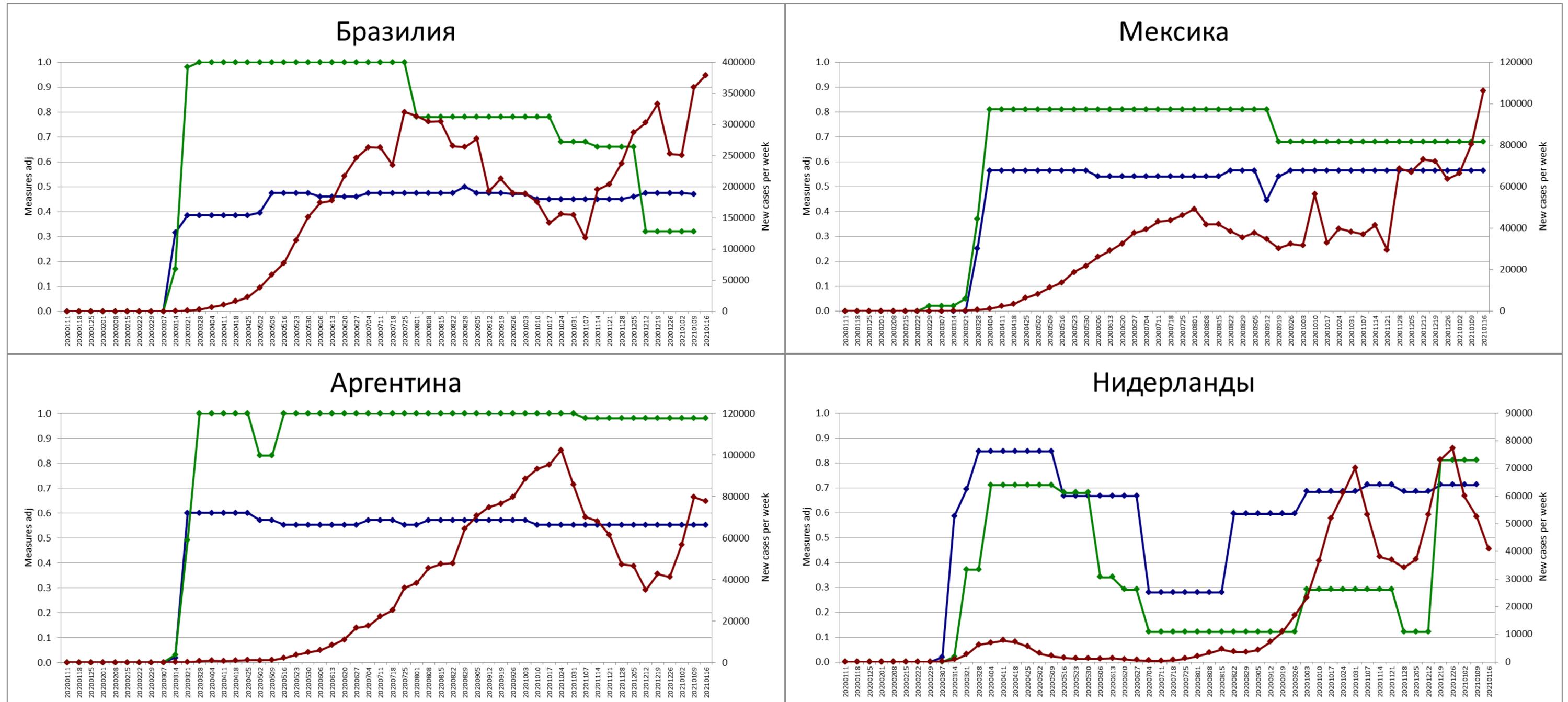
Меры: гос. запреты

Агрегированные меры и новые случаи заболеваний (понеделно)



■ Число новых заболеваний ■ Меры: собрания ■ Меры: гос. запреты

Агрегированные меры и новые случаи заболеваний (понеделно)



Число новых заболеваний



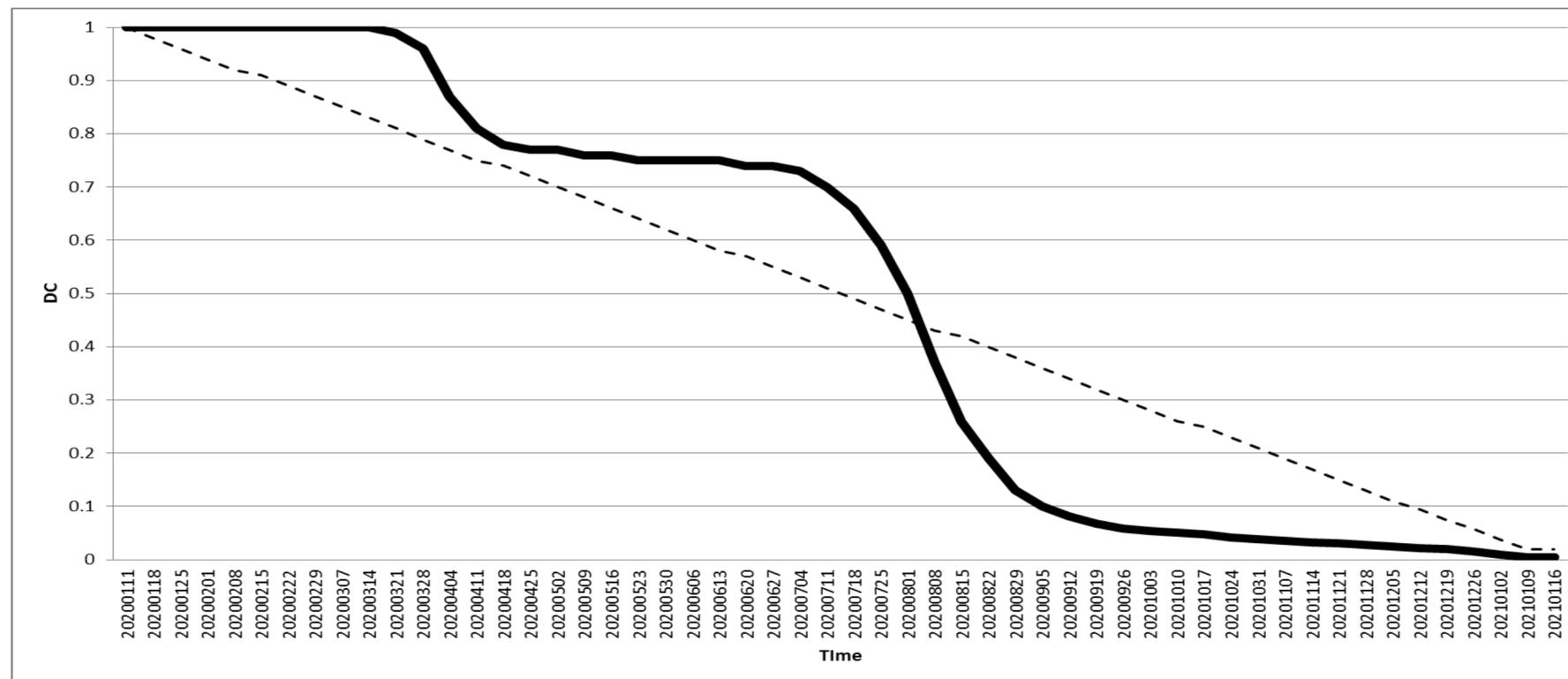
Меры: собрания



Меры: гос. запреты

Модель анализа временных рядов числа случаев заболевания коронавирусом

Для выбранного периода вычисляется сумма значений временного ряда за весь выбранный период (длины n), и для текущего k -ого момента времени вычисляется доля значений, начиная с этого момента до конца выбранного периода.



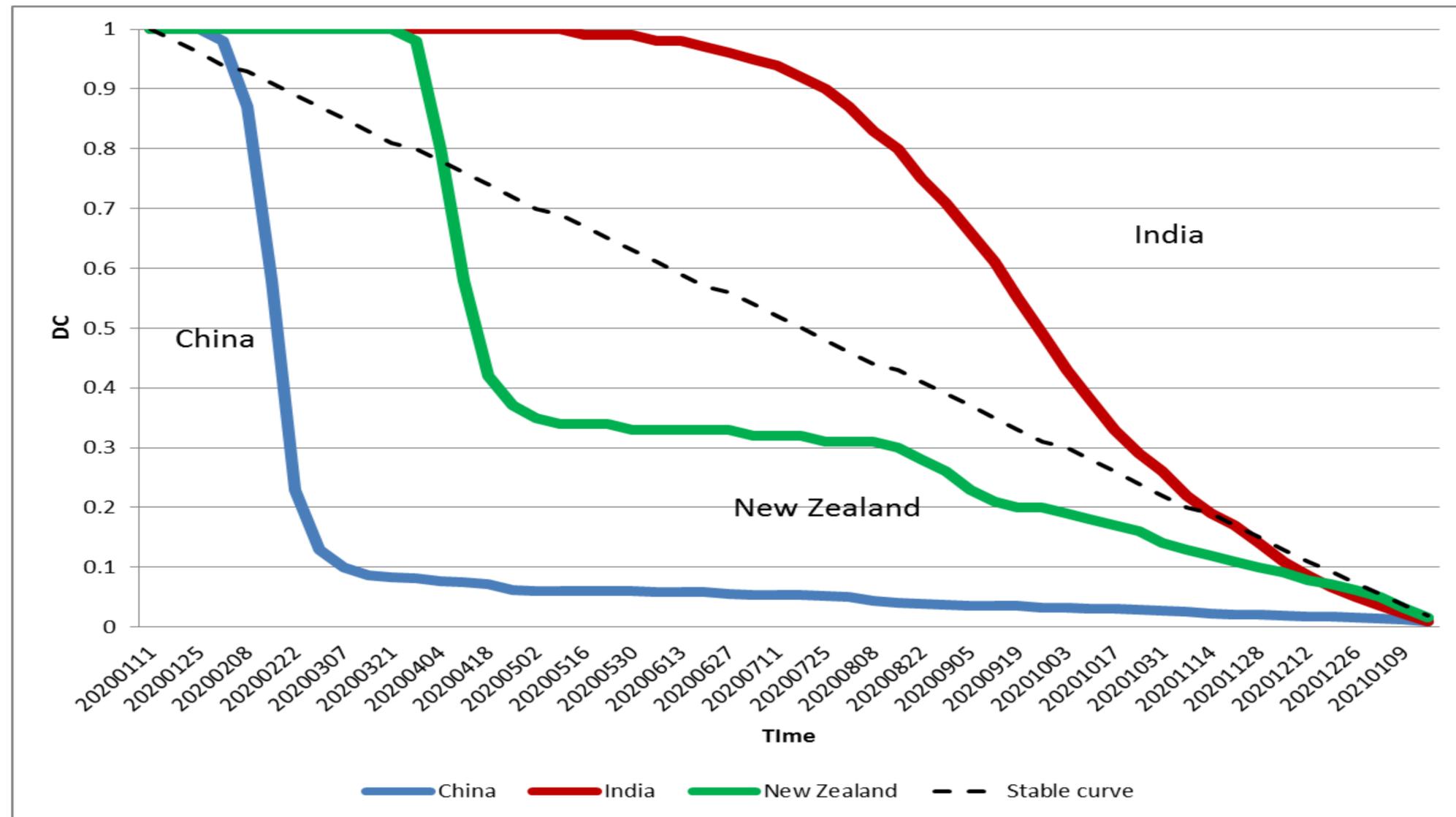
Модель анализа временных рядов числа случаев заболевания коронавирусом

$$LC(k) = \frac{\sum_{t=k}^n v_t}{\sum_{t=1}^n v_t},$$

где v_t - значение в момент времени t .

Значение v_t в момент времени t определено, как число новых случаев, $v_t = a_t - a_{t-1}$, где a_t - суммарное число случаев на момент времени t от 1 до n (предполагается, что данные по числу случаев a_t заданы для t от 0 до n).

Модель нестационарных временных рядов для случаев заражения людей за неделю

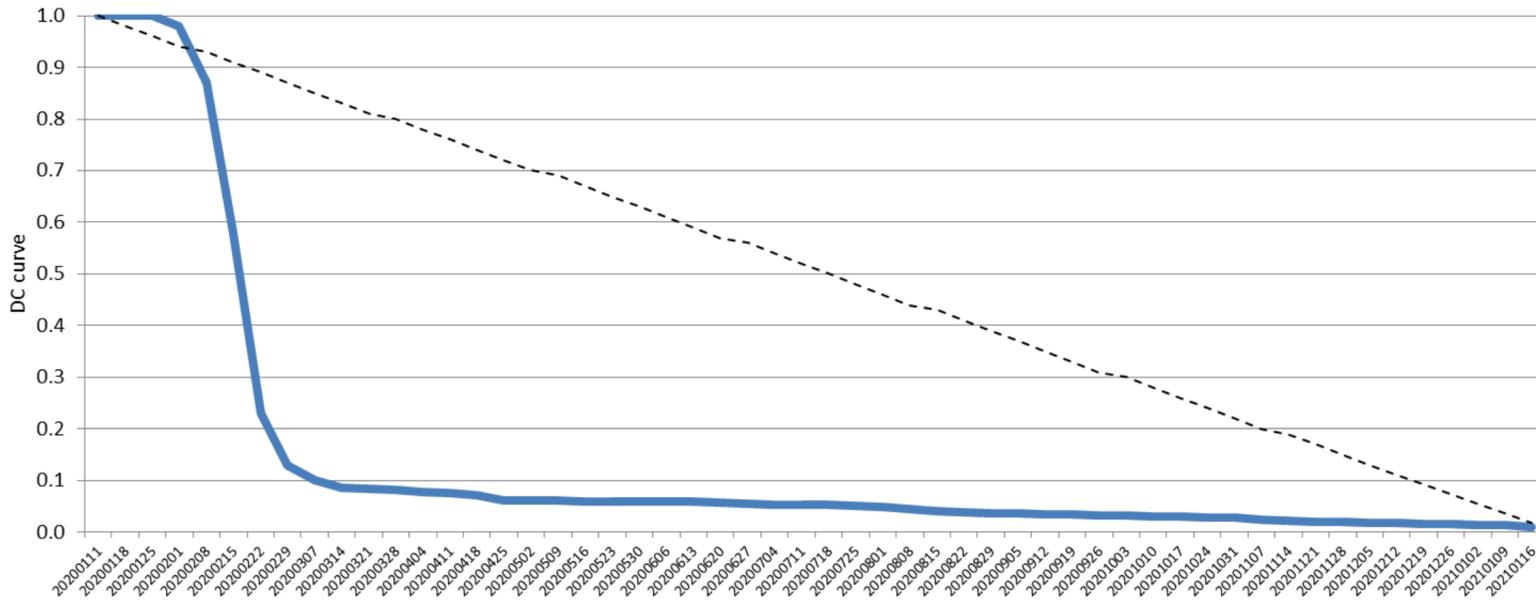


Пик в январе - График под диагональю (Китай)

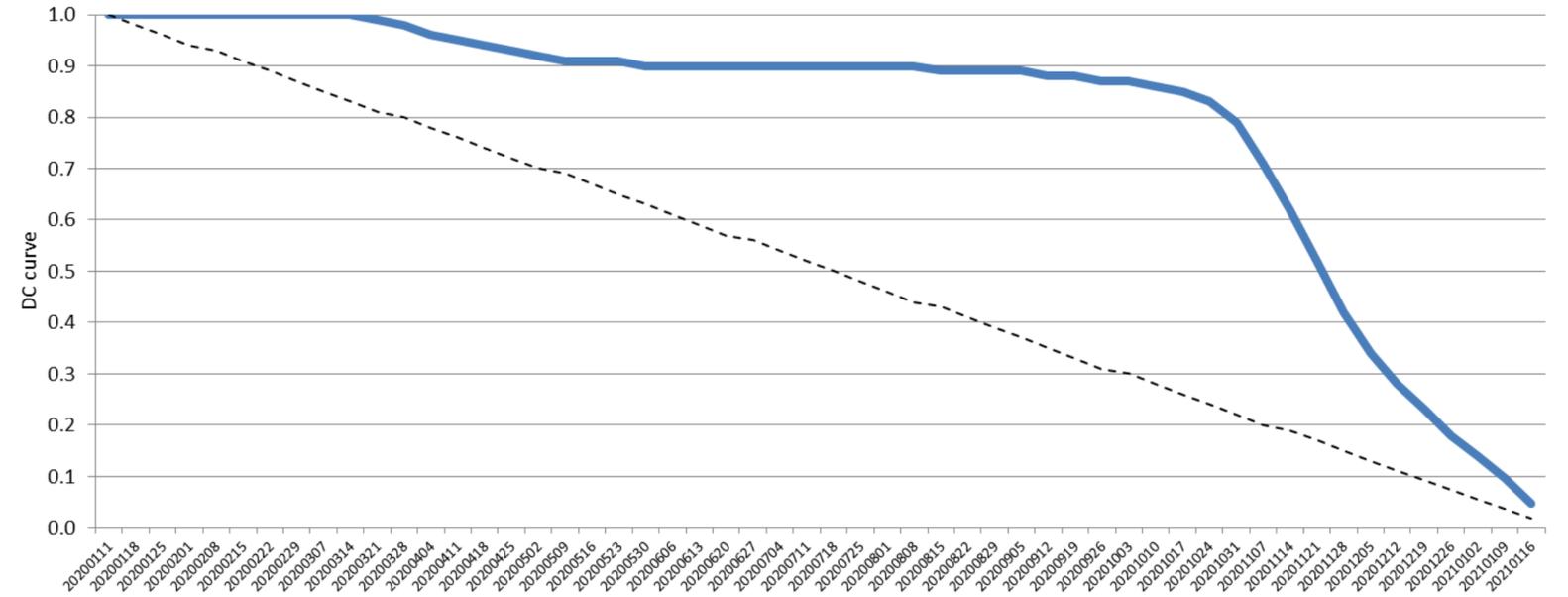
Пик в сентябре - График над диагональю (Индия)

Модель нестационарных временных рядов

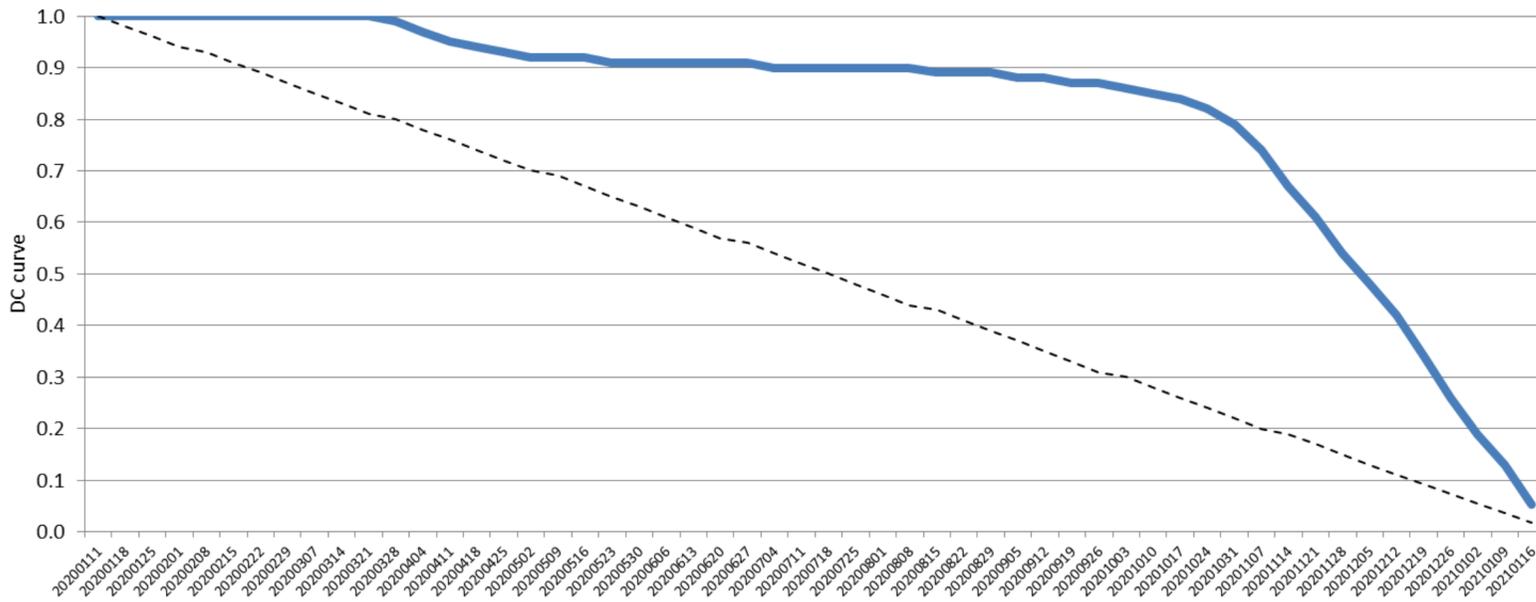
Китай



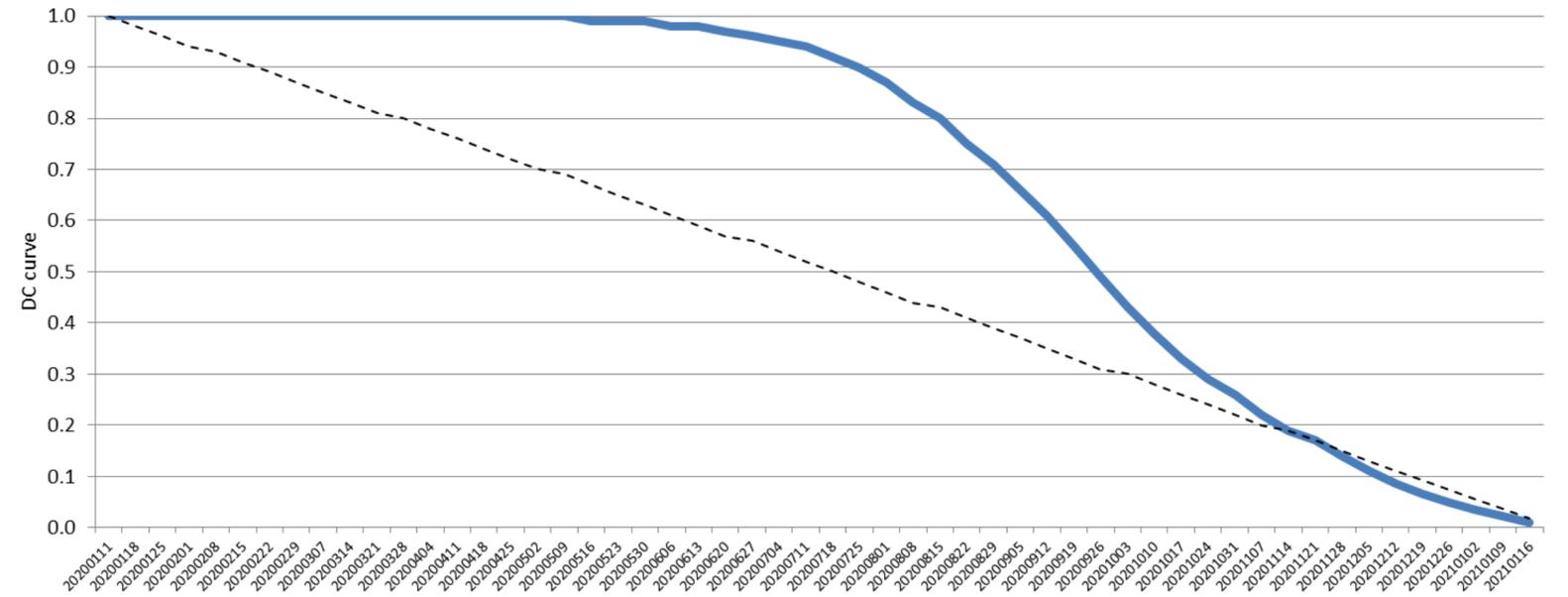
Италия



Германия

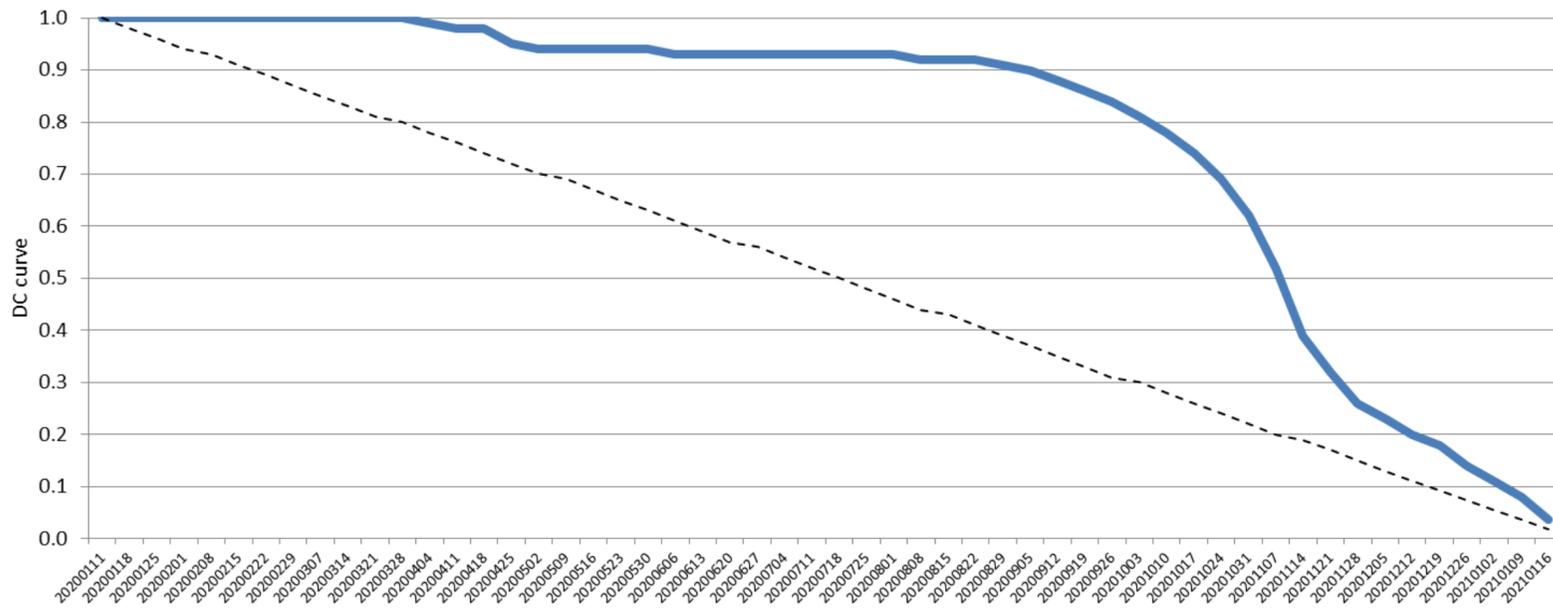


Индия

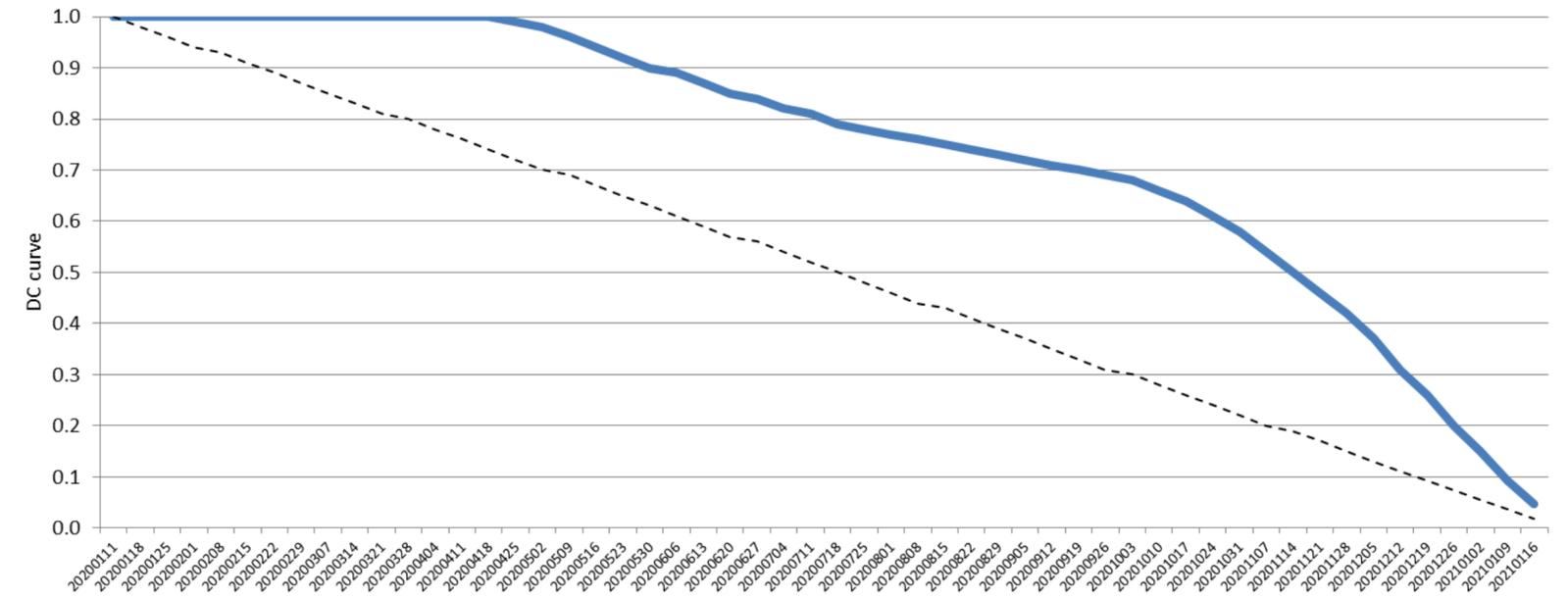


Модель нестационарных временных рядов

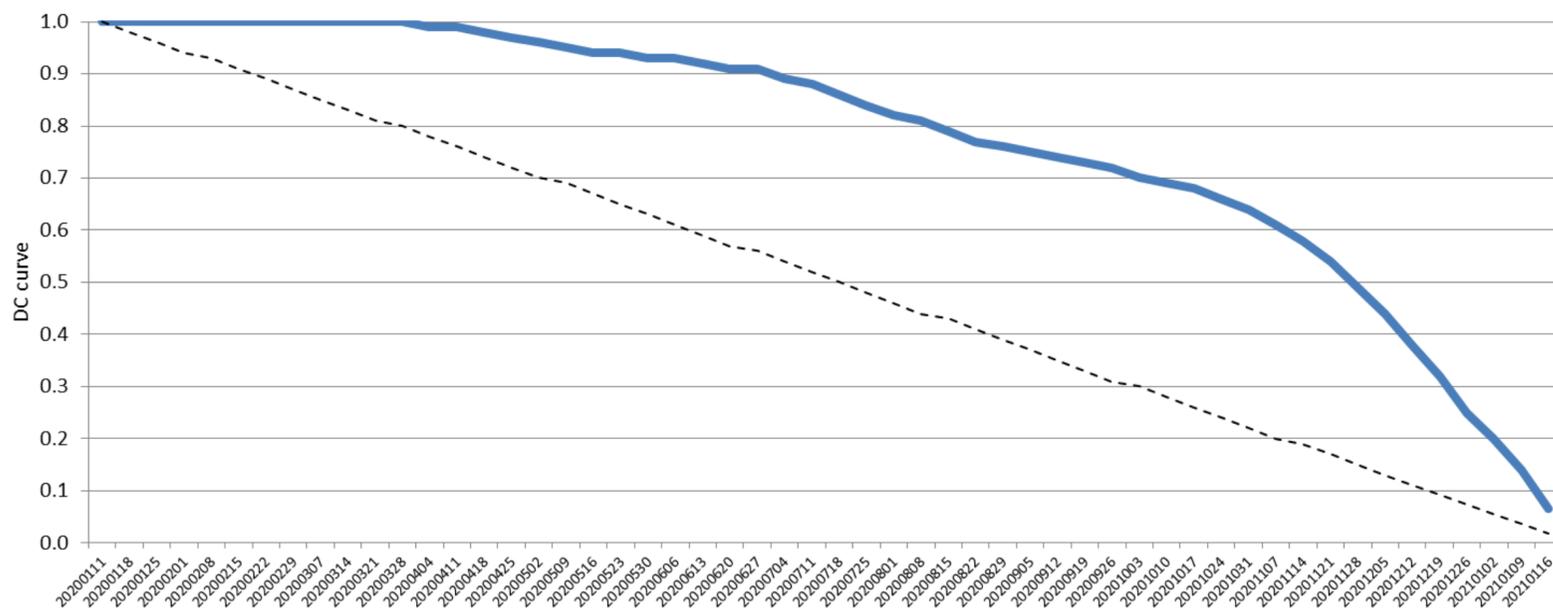
Франция



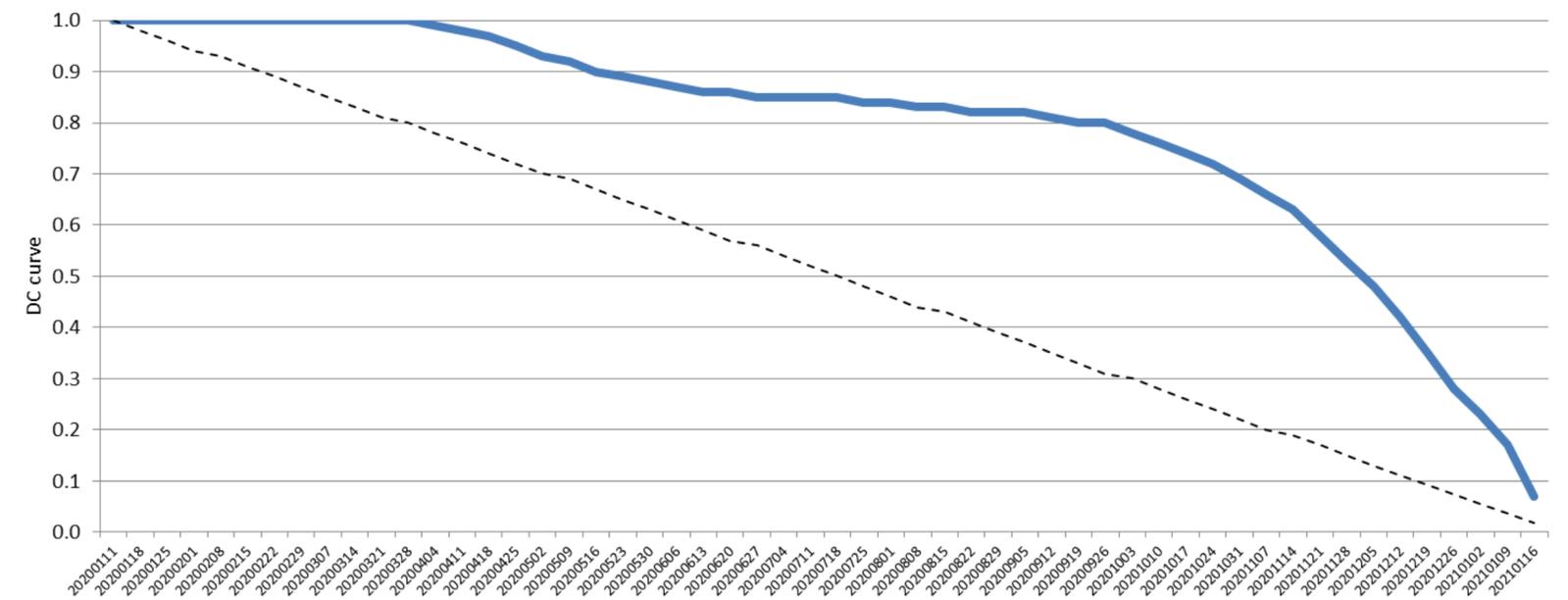
Россия



США

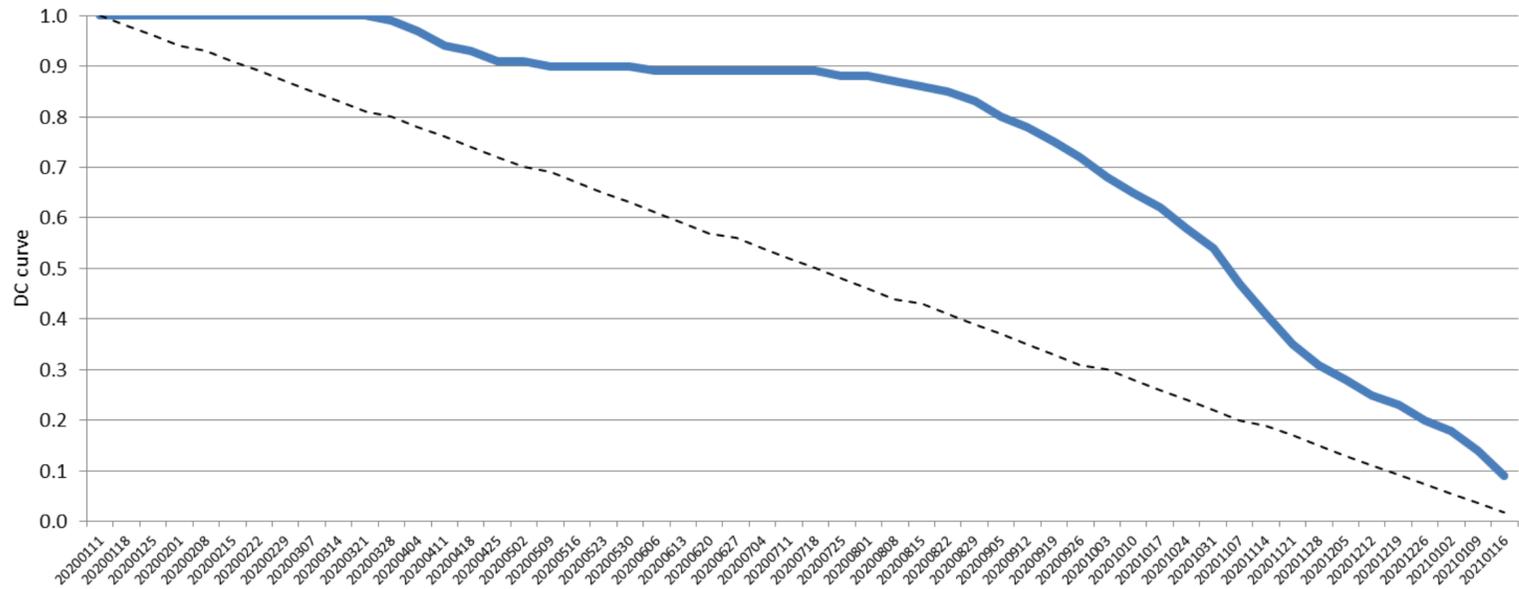


Канада

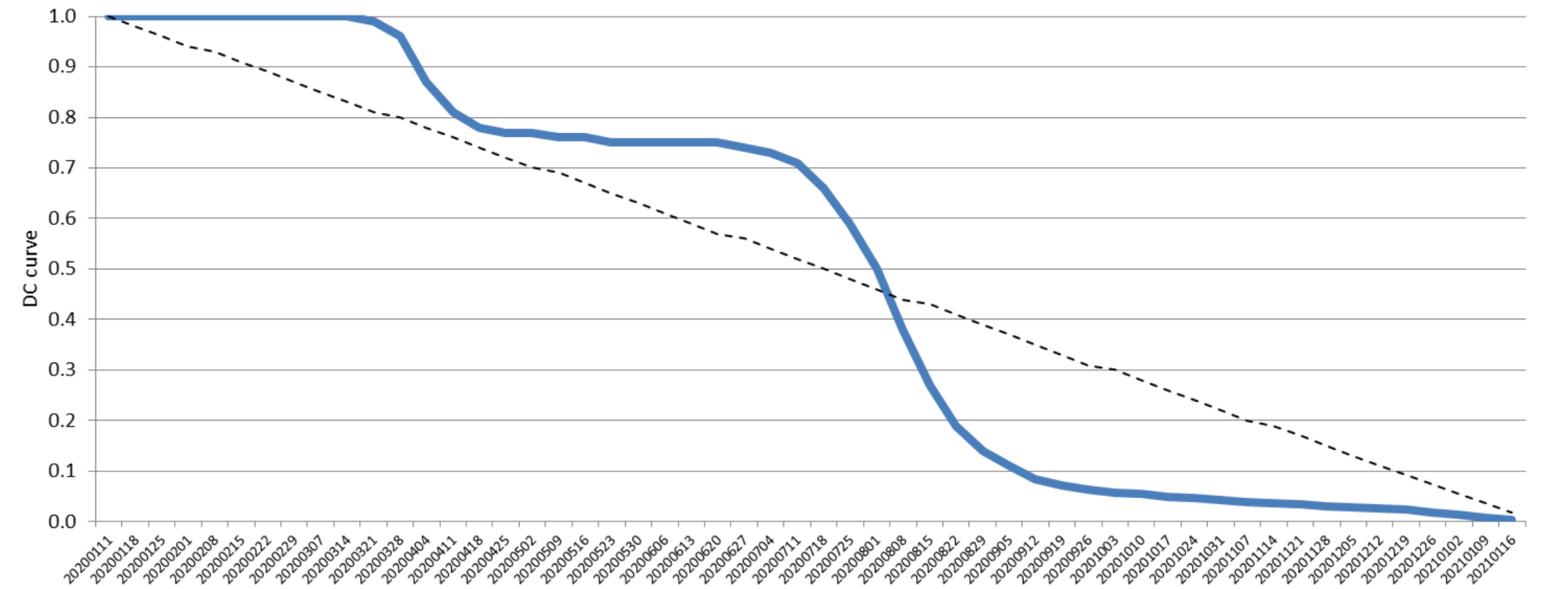


Модель нестационарных временных рядов

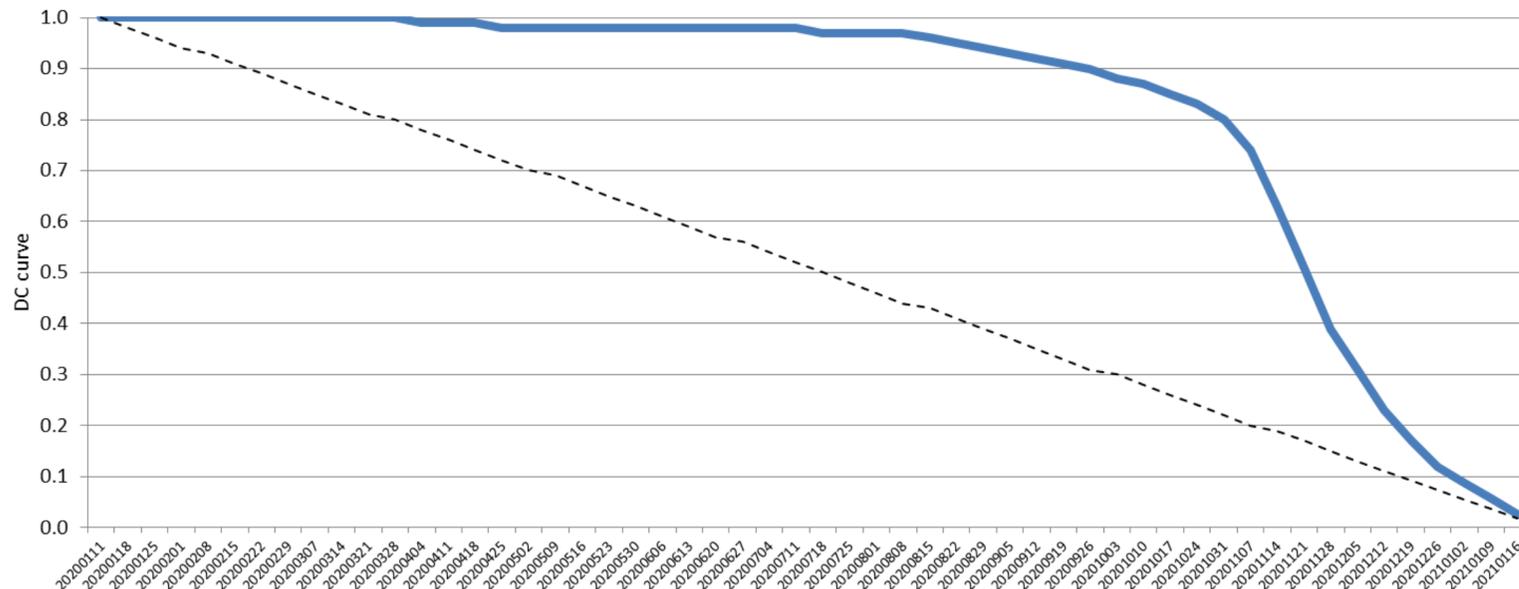
Испания



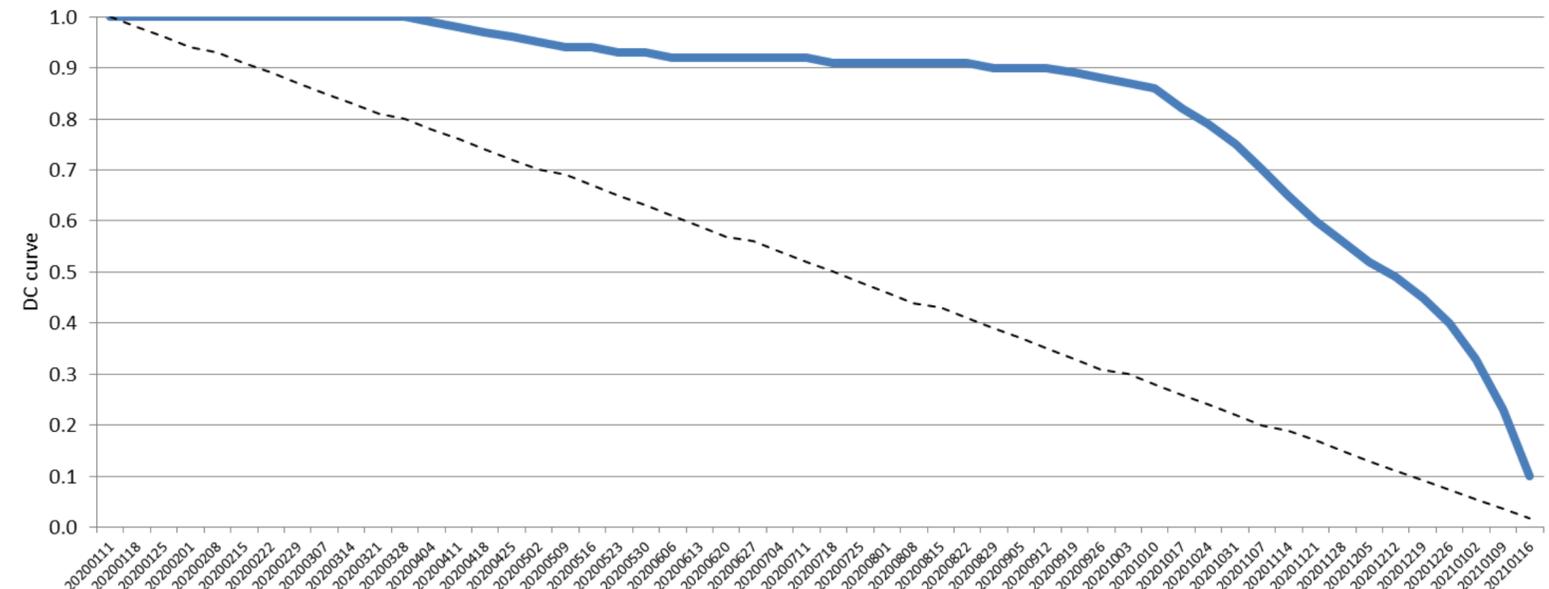
Австралия



Греция

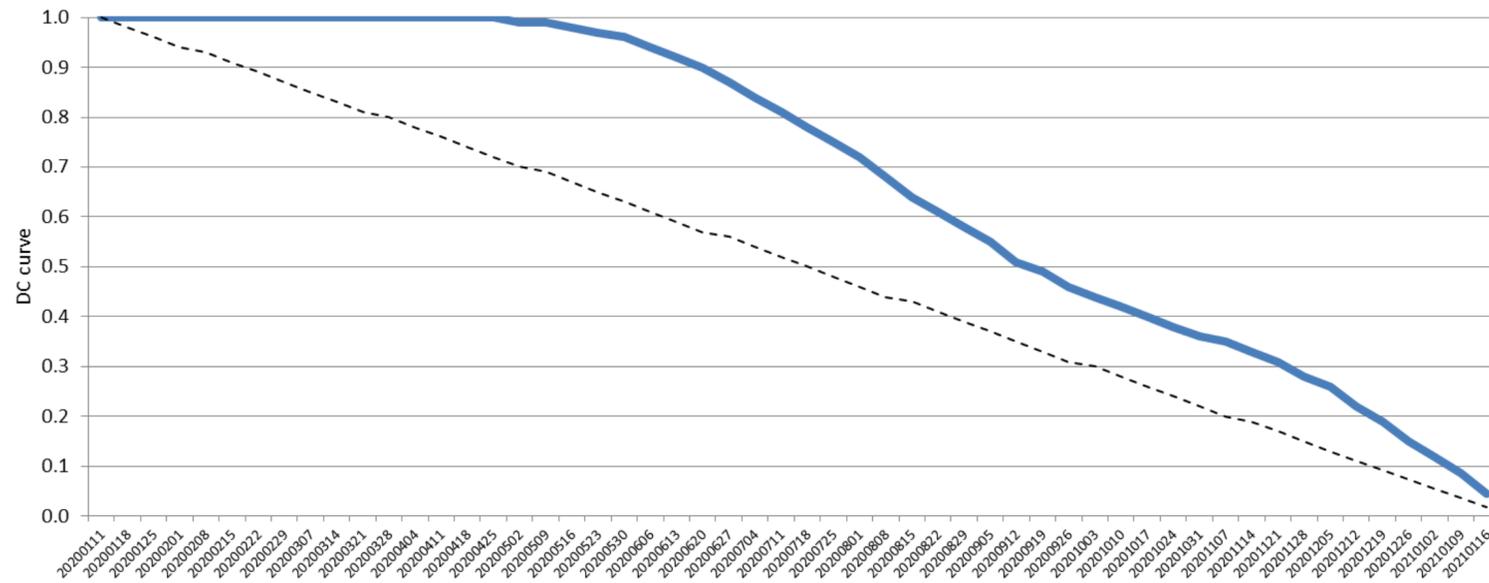


Великобритания

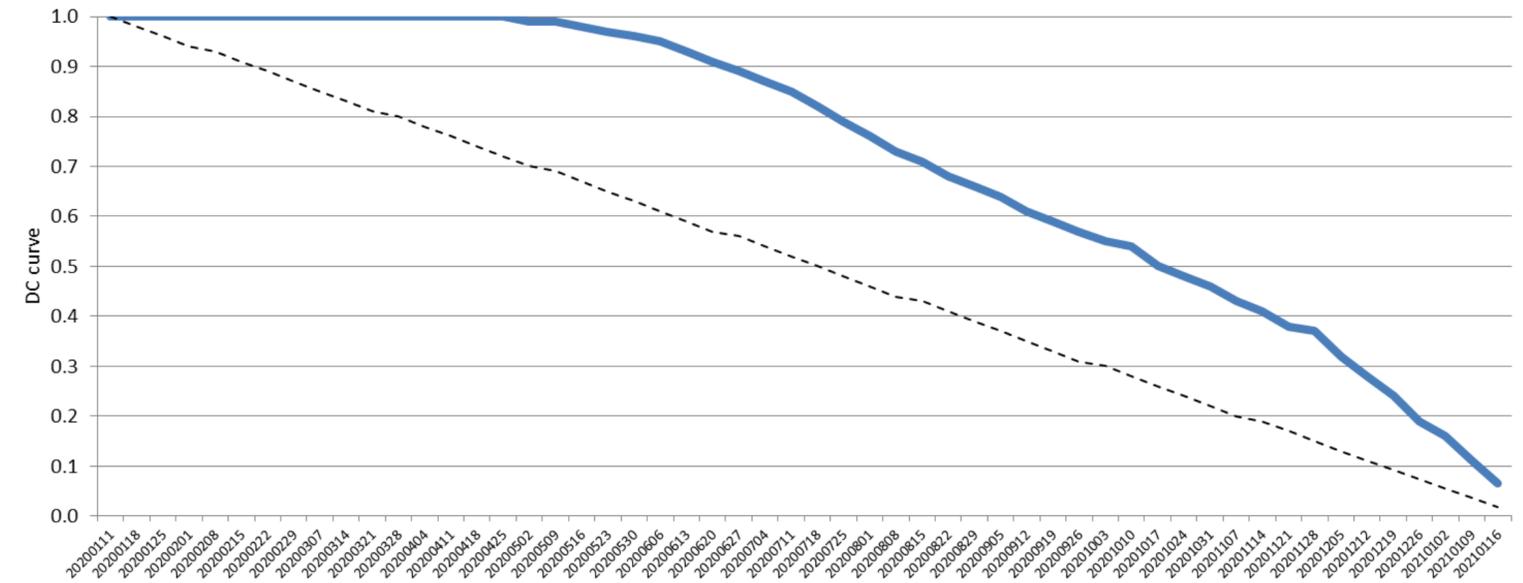


Модель нестационарных временных рядов

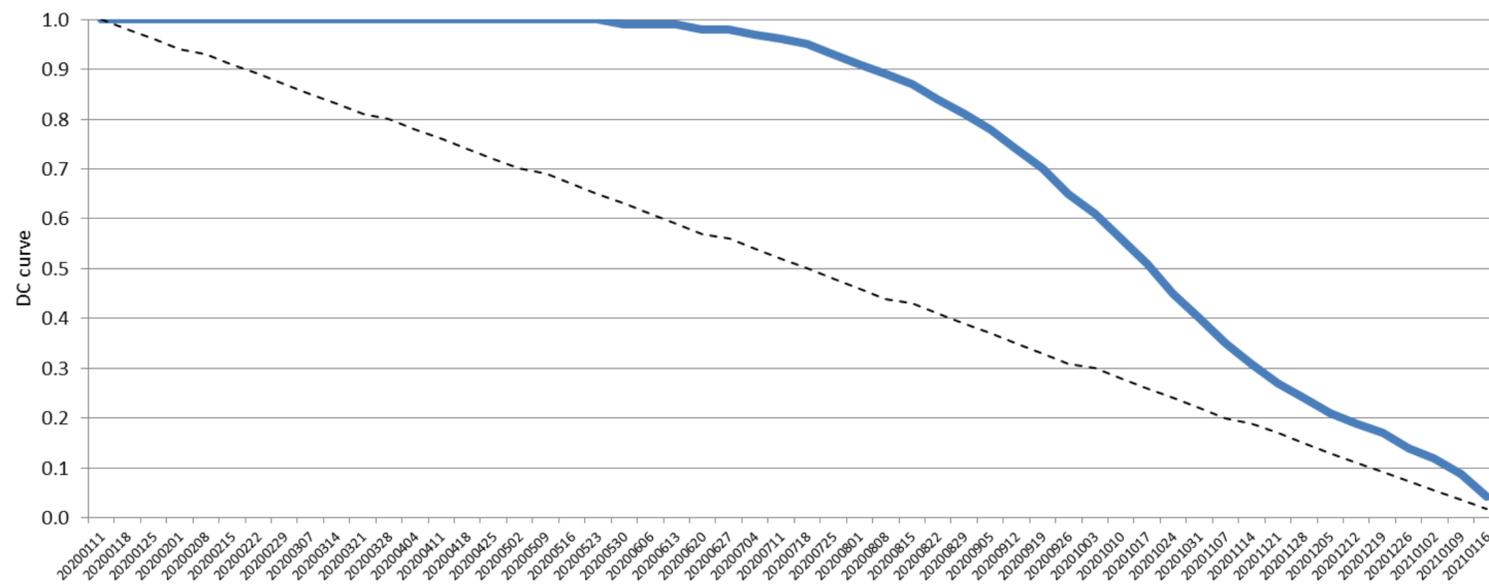
Бразилия



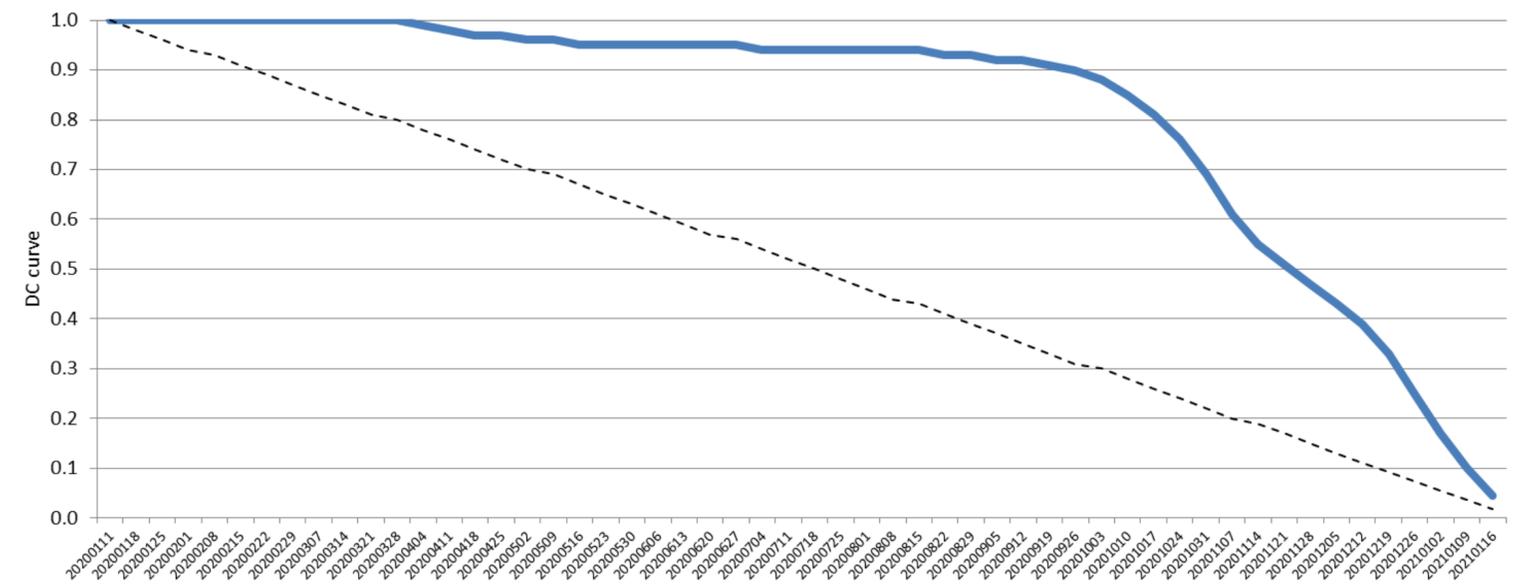
Мексика



Аргентина



Нидерланды



Выявление групп стран со схожей структурой по уровню образования, законопослушности и заболеваемости

Исходные данные:

- *66 стран*
- *Базовая система показателей, характеризующая общий уровень образования в стране, законопослушность граждан и заболеваемость COVID-19*

Рассматриваемый период:

01 февраля 2020 – 09 января 2021

Методология исследования:

классические методы кластерного анализа и анализа паттернов

Уравнения регрессии

$$N_t(\text{C1}) = 0,068 - 0,682N_{t-1} - 0,149N_{t-2} - 0,278N_{t-3} - 0,24N_{t-4} + 1,103N_{t-5}$$

$$\begin{aligned} N_t(\text{C2}) \\ = 0,078 + 0,234N_{t-1} - 0,29N_{t-2} - 0,258N_{t-3} + 0,184N_{t-4} - 0,334N_{t-5} + 0,787N_{t-6} \end{aligned}$$

$$N_t(\text{C3}) = -0,037 + 0,015N_{t-1} + 0,761N_{t-2} + 0,238N_{t-3} - 0,945N_{t-4} + 0,634N_{t-5}$$

$$N_t(\text{C4}) = 0,16 + 0,305N_{t-1} - 0,444N_{t-2} - 1,576N_{t-3} + 1,401N_{t-4} + 0,664N_{t-5}$$

$$N_t(\text{C5}) = 0,333 - 0,544N_{t-1} + 1,412N_{t-2} - 2,838N_{t-3} + 0,976N_{t-4} - 0,54N_{t-5}$$

Результаты регрессии

Множественный R	0,50167
R-квадрат	0,251673
Нормированный R-квадрат	0,187162
Стандартная ошибка	0,230075
Наблюдения	64

Все страны

Множественный R	0,878409
R-квадрат	0,771602
Нормированный R-квадрат	0,58127
Стандартная ошибка	0,159486
Наблюдения	12

C1: Албания, Франция,
Казахстан, Россия, Сербия,
Словения, Испания

Множественный R	0,687243
R-квадрат	0,472303
Нормированный R-квадрат	0,184468
Стандартная ошибка	0,188185
Наблюдения	18

C2: Бельгия, Болгария, Канада,
Финляндия, Япония, Новая
Зеландия, Швейцария

Множественный R	0,945596
R-квадрат	0,894152
Нормированный R-квадрат	0,788304
Стандартная ошибка	0,11532
Наблюдения	11

C3: Бразилия, Египет, Индия,
Мексика, Непал, Сирия,
Турция

Множественный R	0,829122
R-квадрат	0,687443
Нормированный R-квадрат	0,464188
Стандартная ошибка	0,234508
Наблюдения	13

C4: Аргентина, Австралия,
Чили, Италия, США,
Великобритания, Израиль

Множественный R	0,863159
R-квадрат	0,745043
Нормированный R-квадрат	-0,52974
Стандартная ошибка	0,372768
Наблюдения	7

C5: Австрия, Азербайджан,
Китай, Германия, Люксембург,
Нидерланды, Гонконг

Полный состав кластеров

C1

Albania
Bosnia and Herzegovina
France
Kazakhstan
Kyrgyz Republic
Moldova
Romania
Russia
Serbia
Slovak Republic
Slovenia
Spain

C2

Belgium
Belarus
Bulgaria
Canada
Czech Republic
Denmark
Estonia
Finland
Hungary
Iceland
Japan
Latvia
New Zealand
Norway
Saudi Arabia
Singapore
Switzerland
United Arab Emirates

C3

Brazil
Egypt
India
Mexico
Nepal
Nicaragua
Nigeria
South Africa
Syria
Turkey
Vietnam

C4

Argentina
Australia
Chile
Georgia
Greece
Ireland
Israel
Italy
Lithuania
Poland
Ukraine
United Kingdom
United States

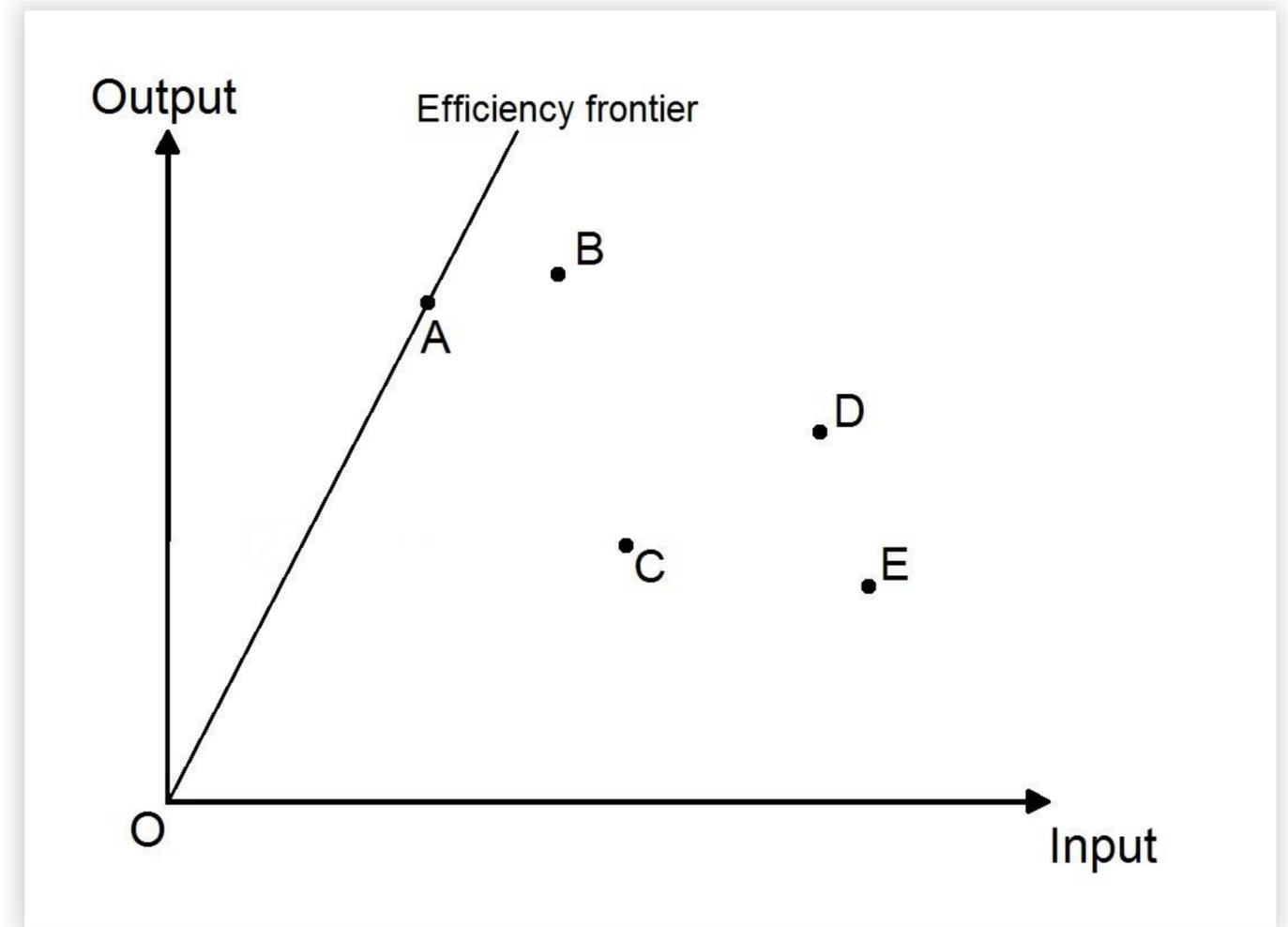
C5

Austria
Azerbaijan
China
Germany
Hong Kong
Luxembourg
Netherlands



Оболочечный Анализ Данных (DEA) Эффективность мер

Базовая модель DEA



Базовая модель DEA

$$e_k = \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}}$$

e_k – эффективность k -го объекта

u_i, v_j – весовые коэффициенты, показывающие вклад каждого из параметров

x_{ik} – входные параметры, показывающие затраты

y_{jk} – выходные параметры, иллюстрирующие полученные результаты

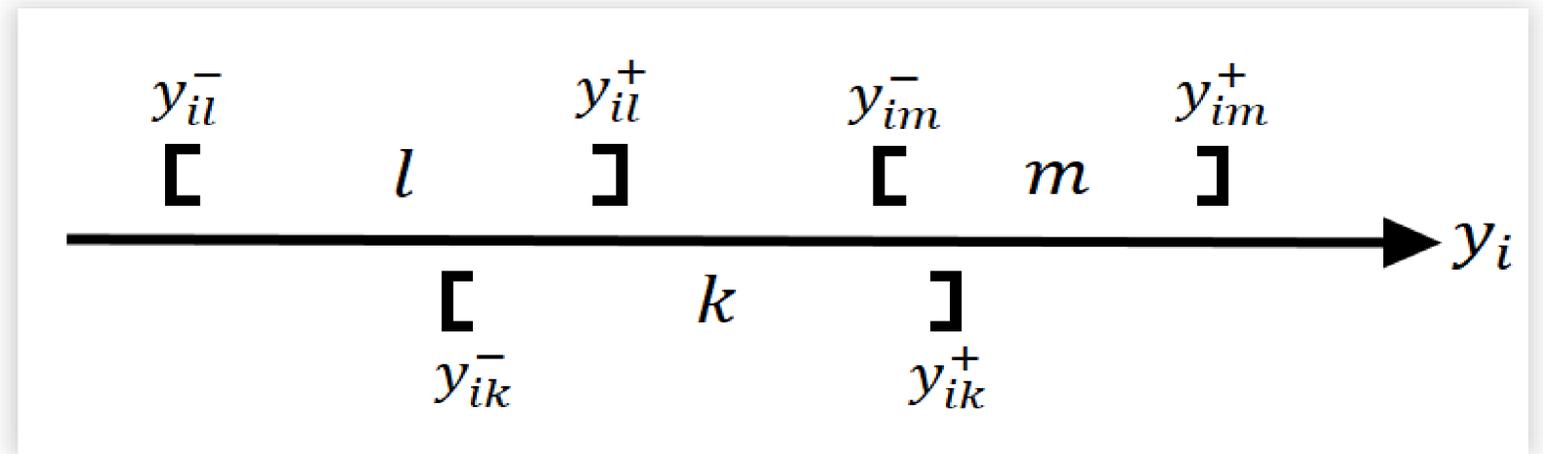
M – количество выходных параметров; N – количество входных параметров

Базовая модель DEA

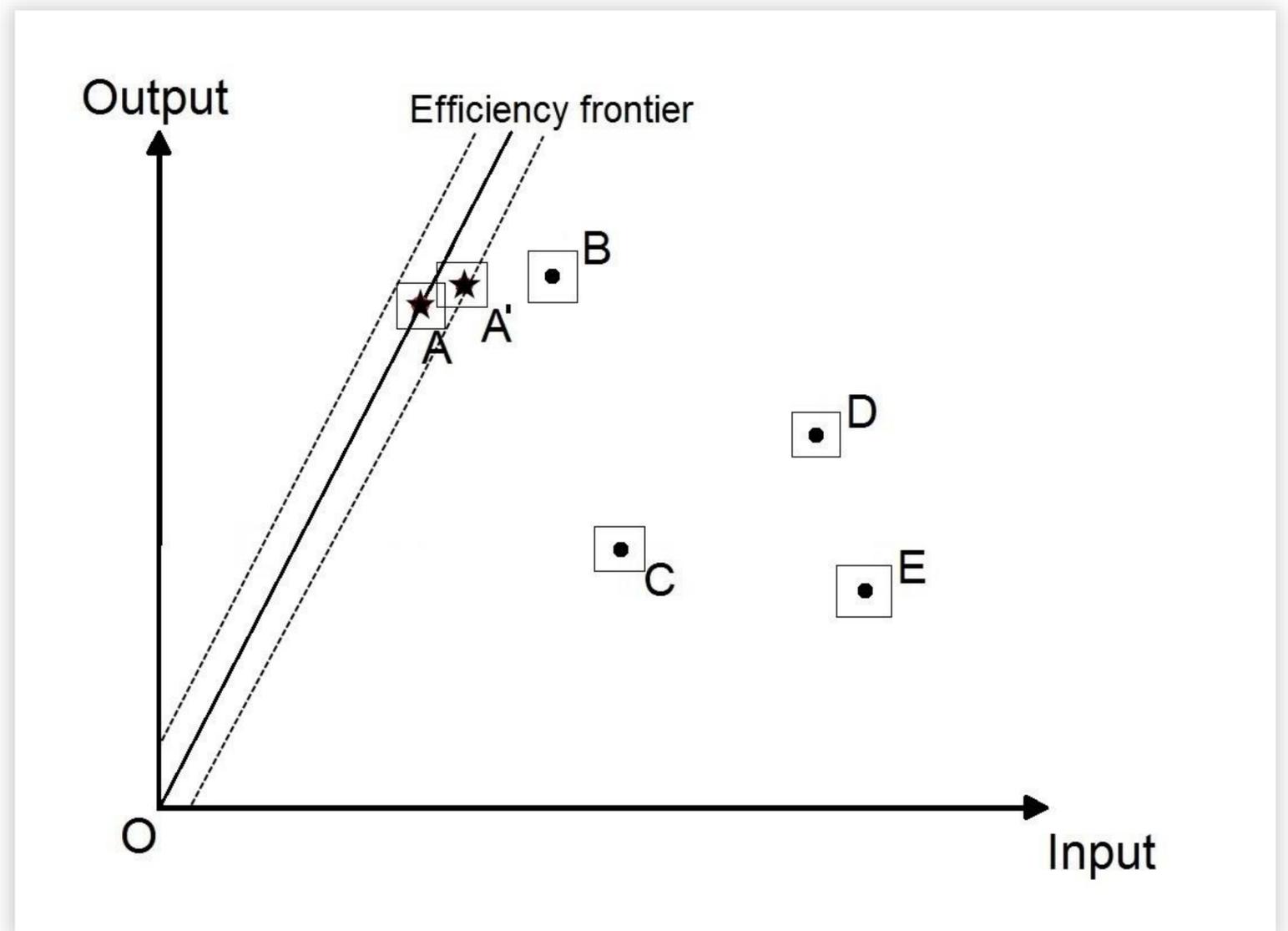
$$\max_{u_i, v_j} \left(e_k = \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}} \right)$$

$$\left\{ \begin{array}{l} \frac{\sum_{i=1}^M u_i y_{ik}}{\sum_{j=1}^N v_j x_{jk}} \leq 1, \quad k = 1, \dots, R \\ u_i > 0 \quad \forall i \\ v_j > 0 \quad \forall j \end{array} \right.$$

Интервальная оценка значений



Интервальный DEA



Параметры модели

Входные данные:

Степень введения отдельных карантинных мер

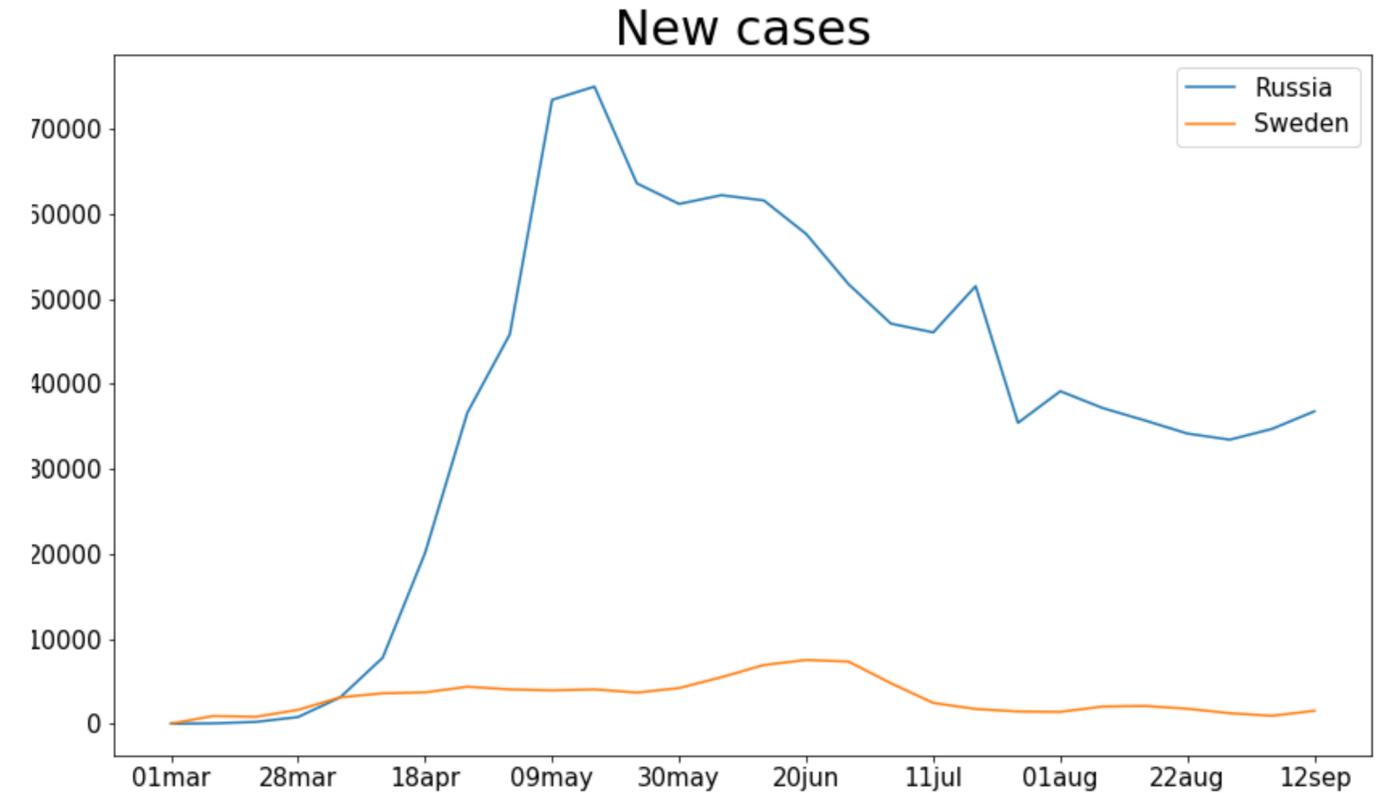
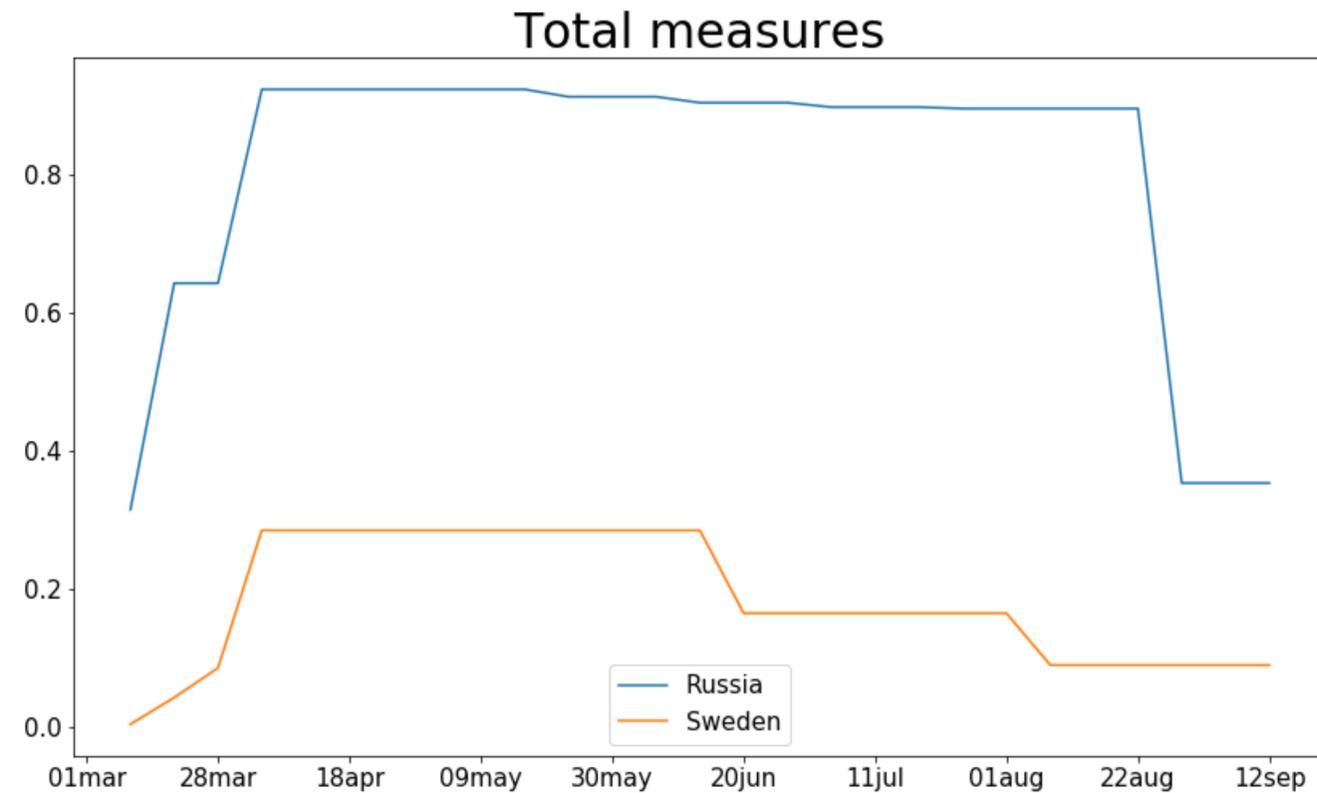
Законопослушность

Выходные параметры:

Прирост количества заболевших

Страна	Запрет загранич. путешествий	Запрет внутр. путешествий	Закрывание школ	Закрывание рабочих мест	Отмена публичных мероприятий	Запрет на массовые собрания	Остановка обществ. транспорта	Домашний режим	Законопослушность
Россия	3-5	1-3	2-4	1-3	1-3	3-5	0-2	1-3	2-4
Швеция	2-4	0-2	0-2	0-2	0-2	2-4	0-1	0-1	8-10

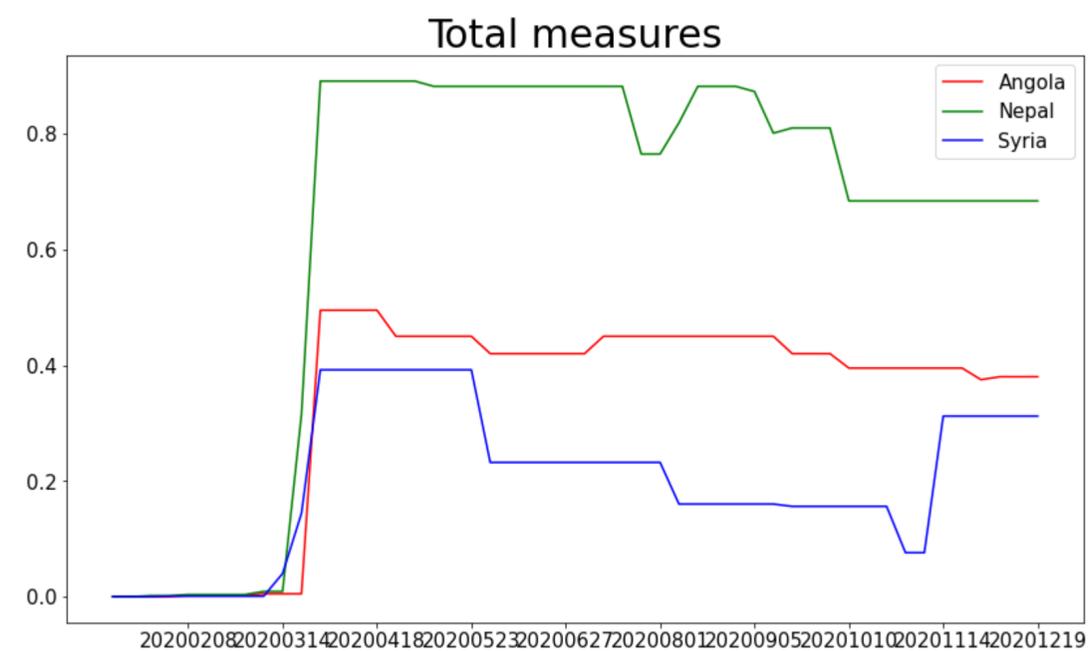
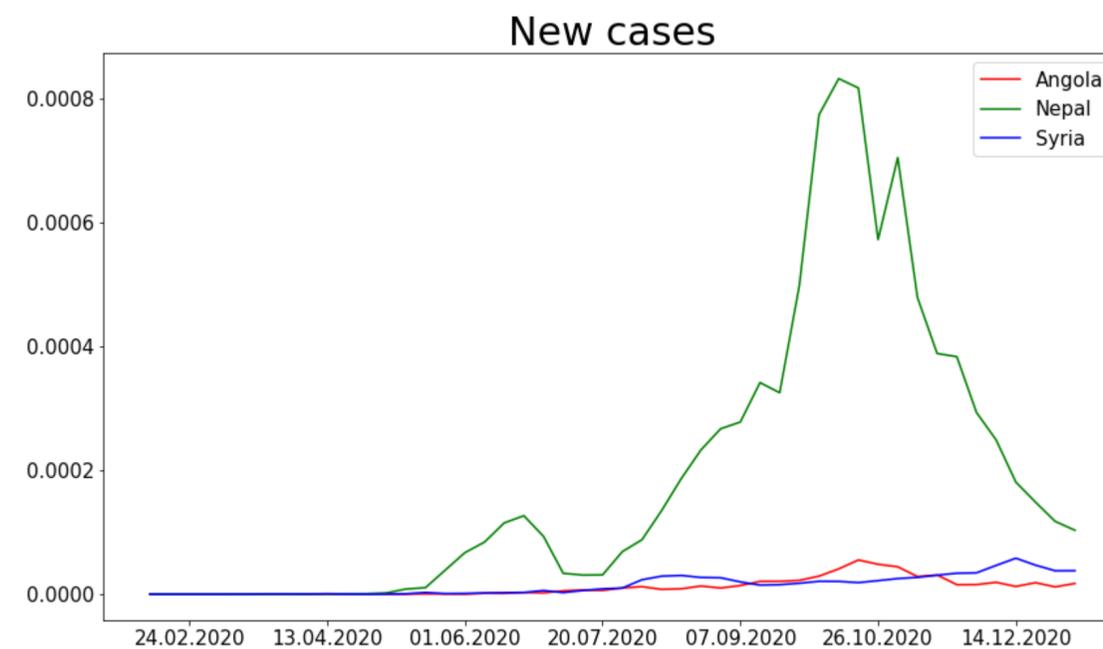
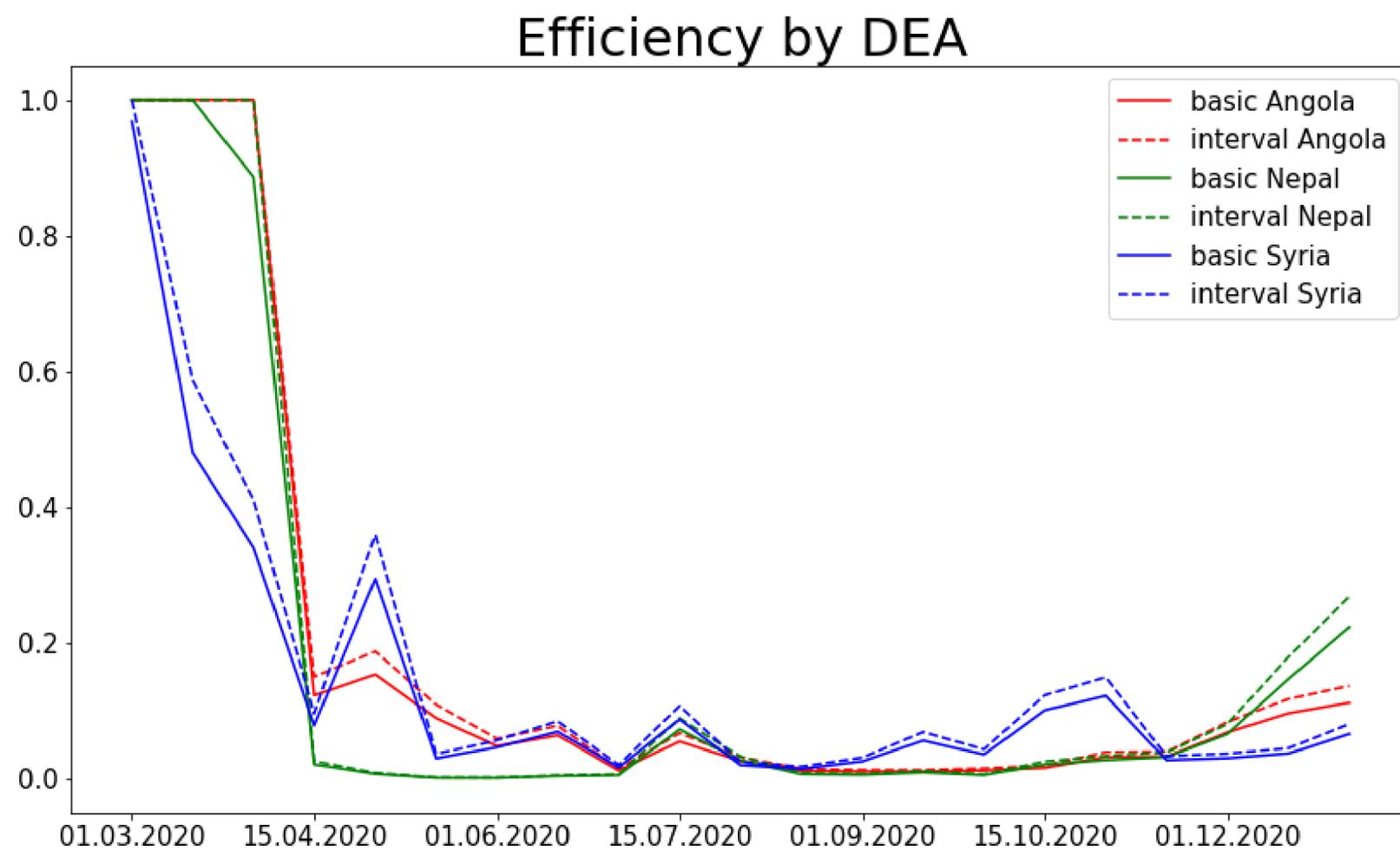
Параметры модели



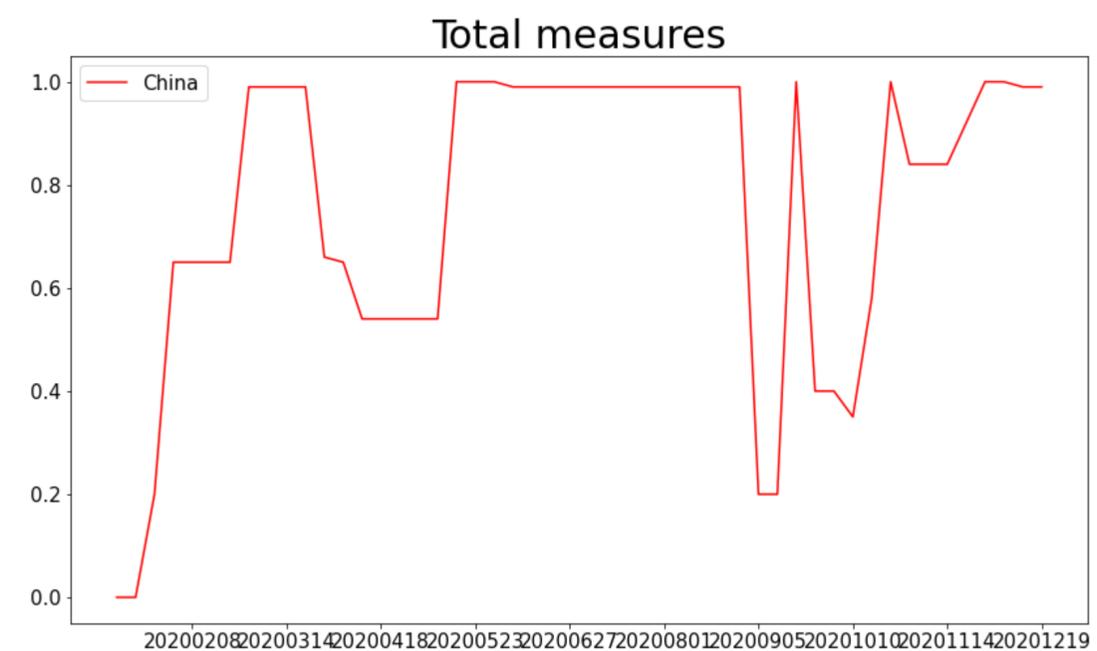
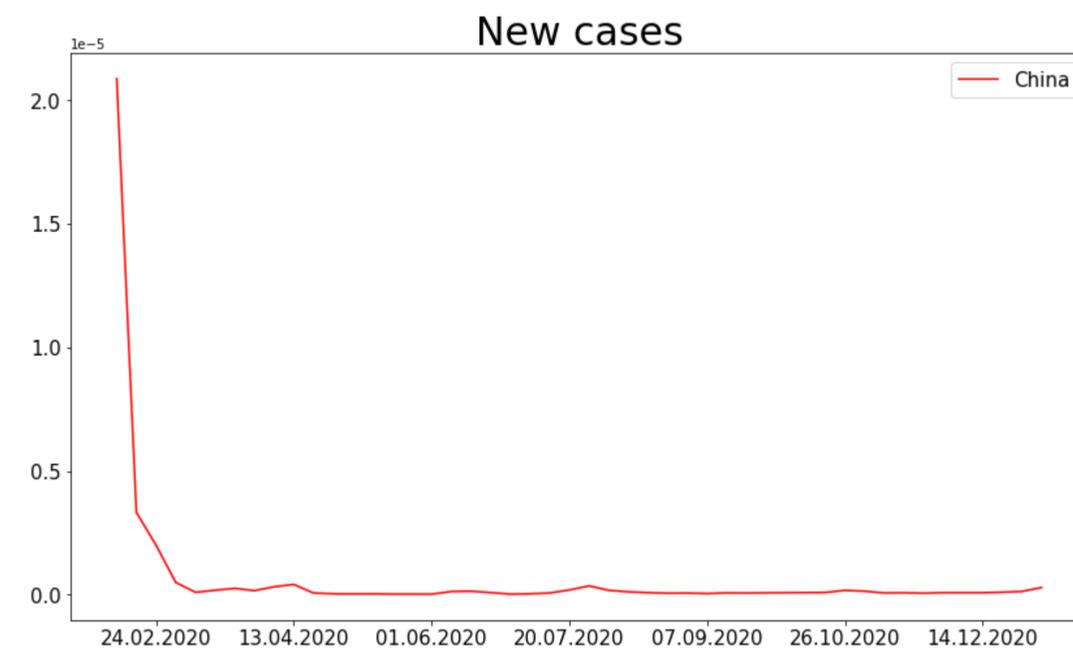
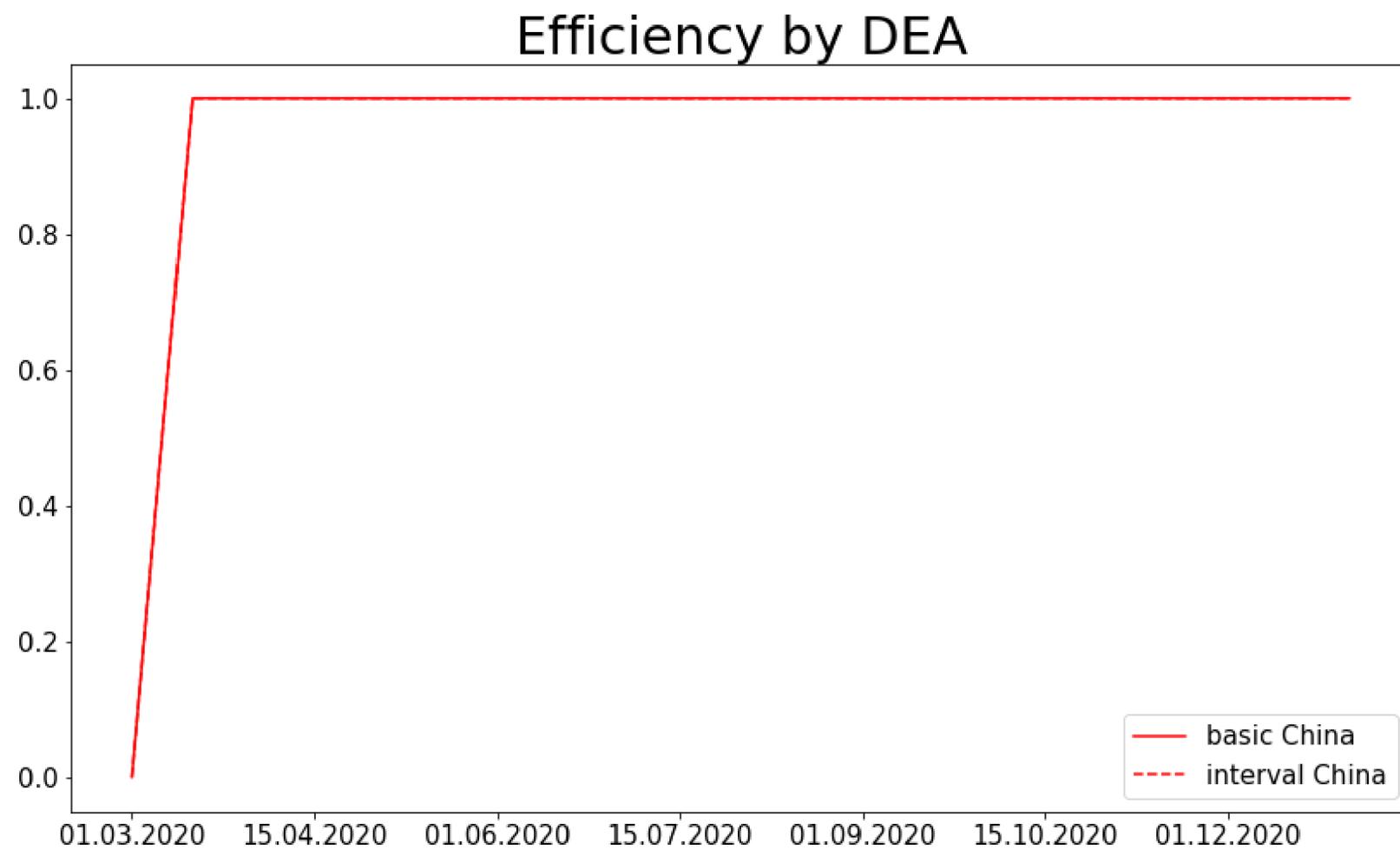
Страна	Запрет загранич. путешеств.	Запрет внутр. путешеств.	Закрывание школ	Закрывание рабочих мест	Отмена публичных мероприятий	Запрет на массовые собрания	Остановка обществ. транспорта	Домашний режим	Законопослушность
Россия	3-5	1-3	2-4	1-3	1-3	3-5	0-2	1-3	2-4
Швеция	2-4	0-2	0-2	0-2	0-2	2-4	0-1	0-1	8-10

Эффективность карантинных мер

Группа 1 (Поздний приход вируса)

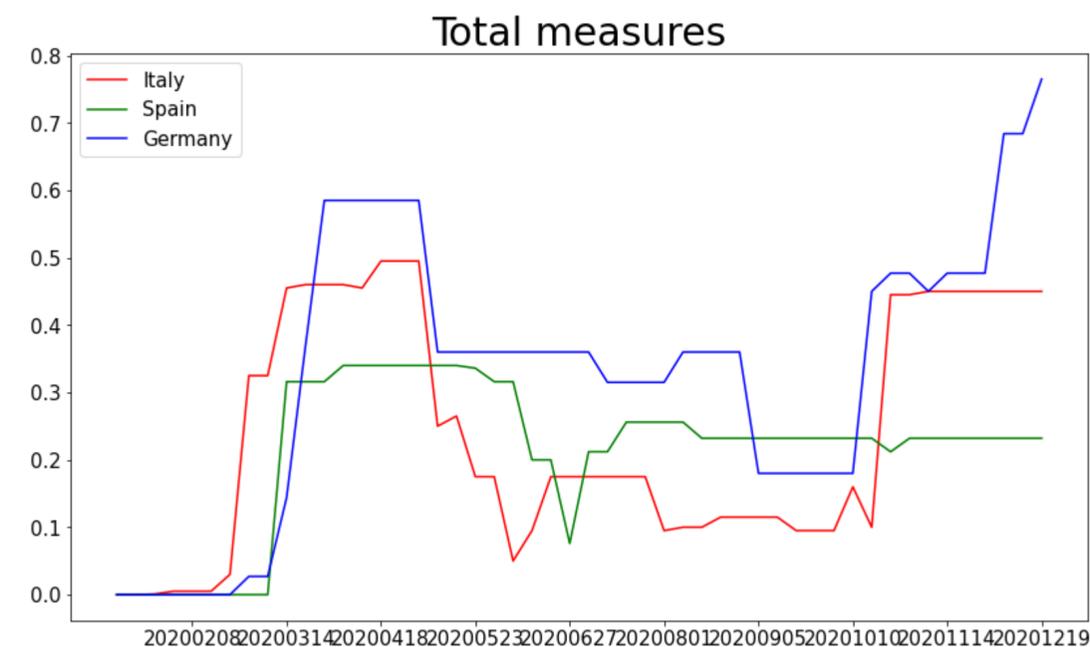
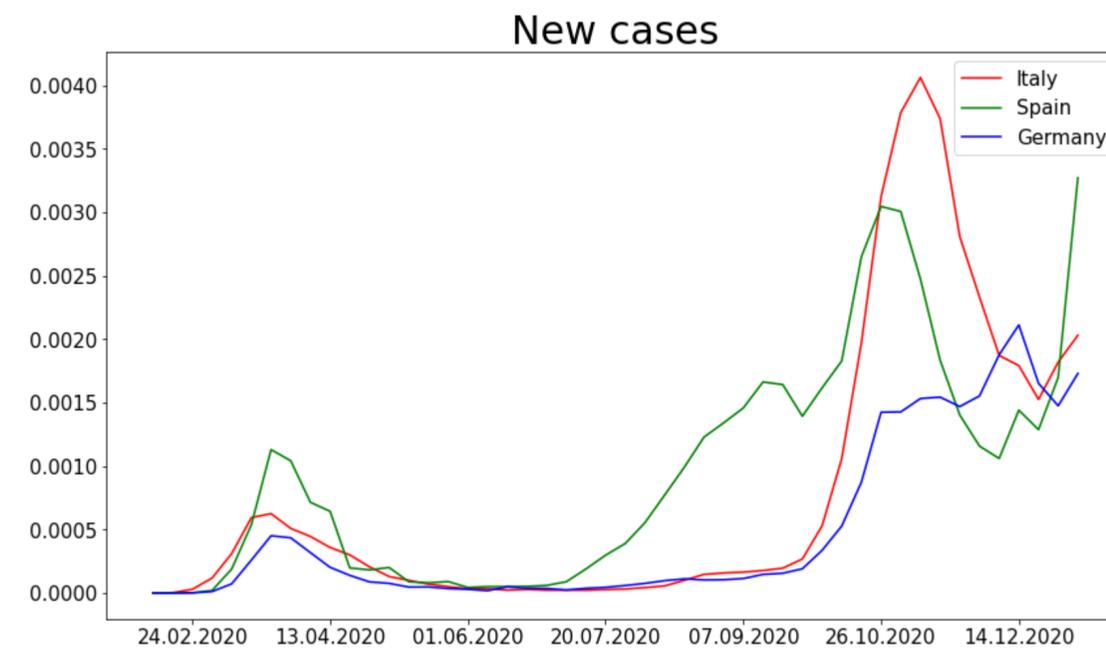
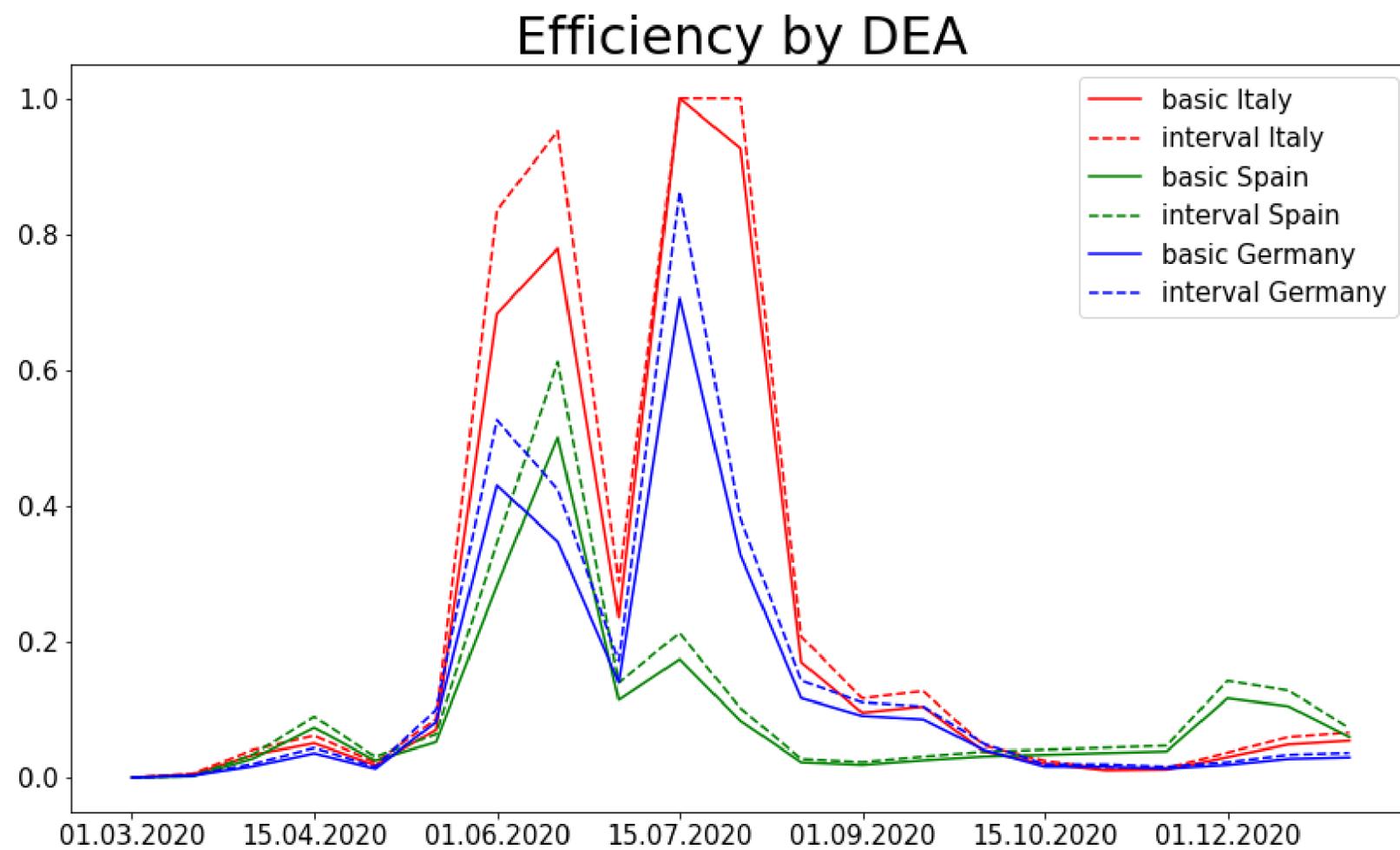


Эффективность карантинных мер Группа 2 (Китай)



Эффективность карантинных мер

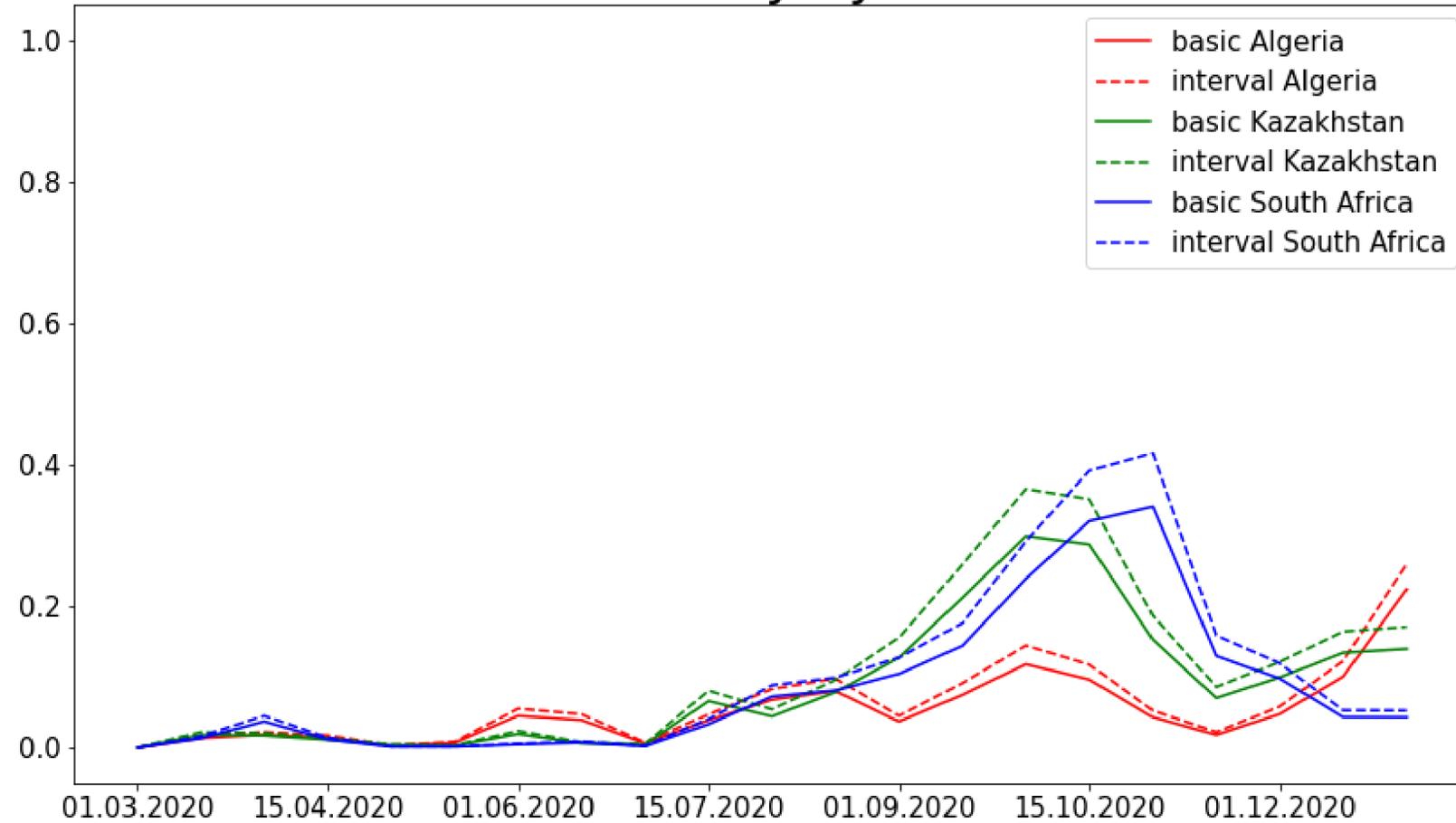
Группа 3 (Страны с ранней второй волной)



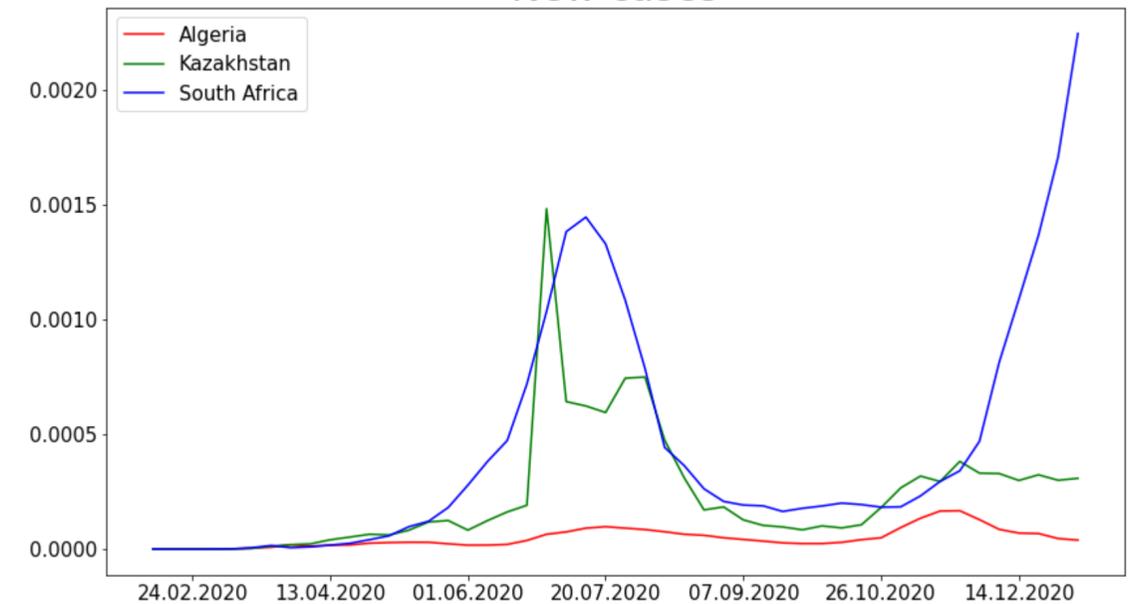
Эффективность карантинных мер

Группа 4 (Страны с поздним завершением первой волны)

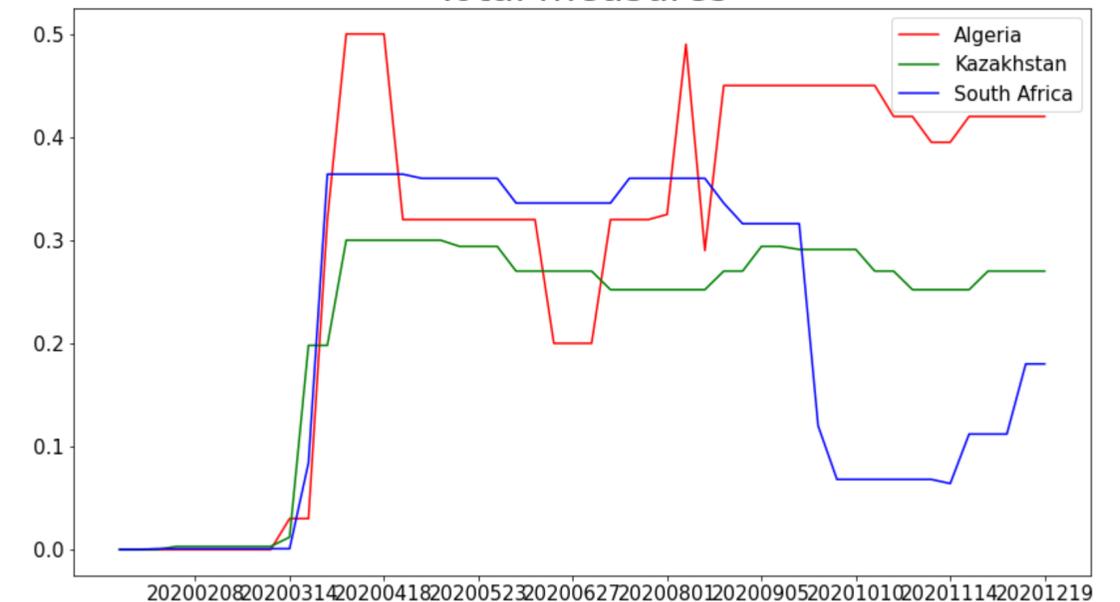
Efficiency by DEA



New cases



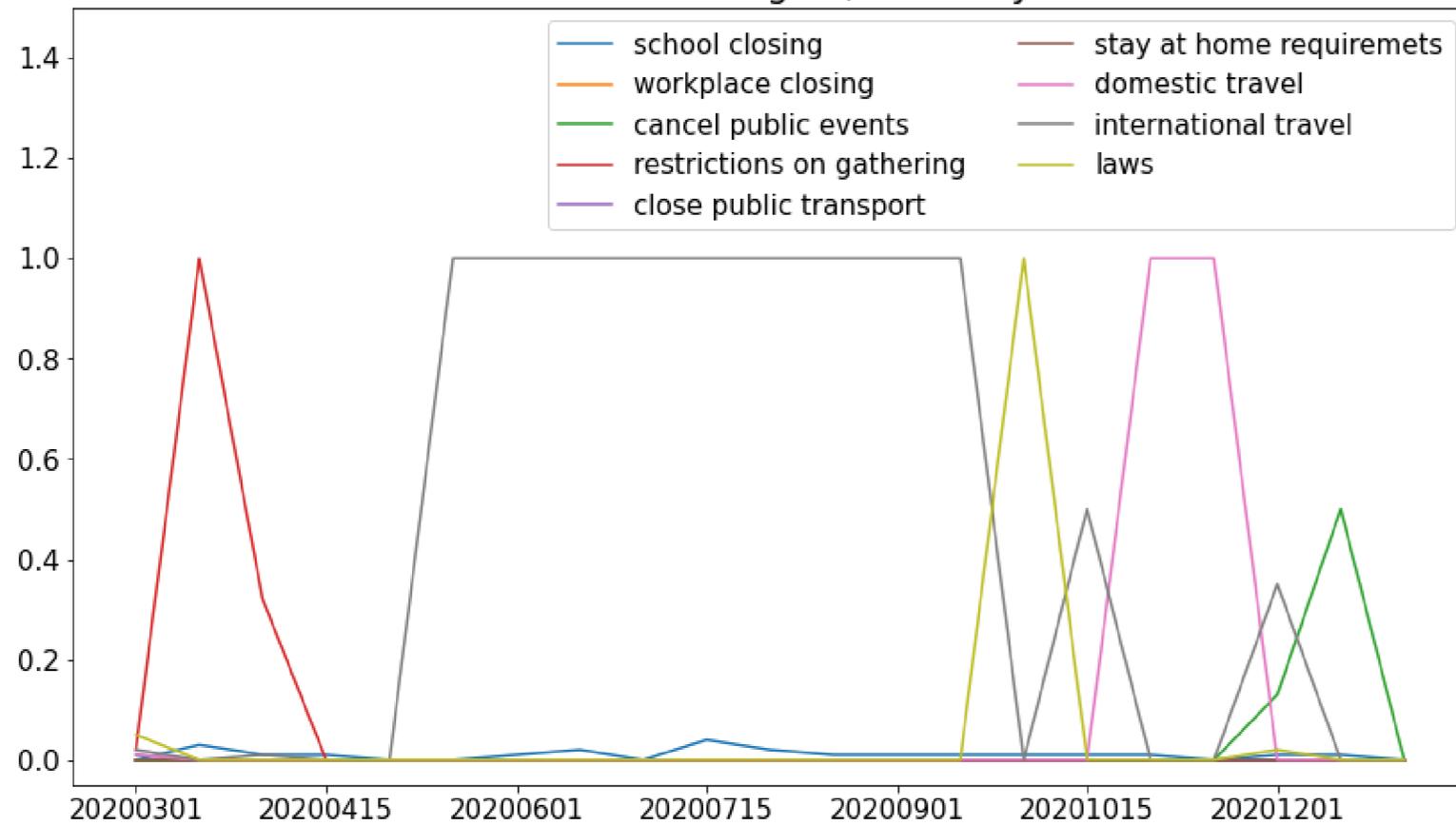
Total measures



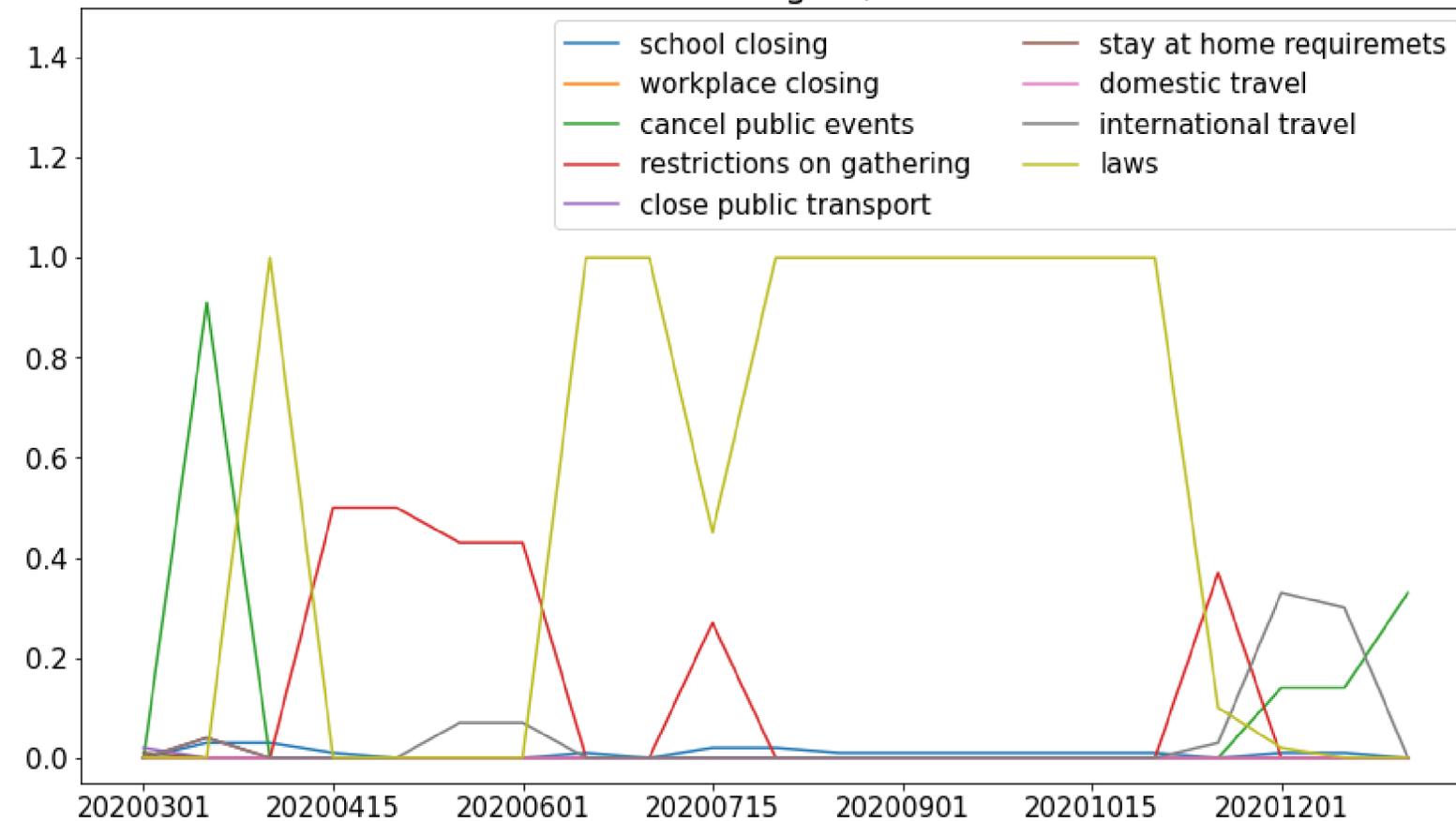
Эффективность карантинных мер

Анализ весов параметров

Parameter weights, Germany



Parameter weights, Sweden



Заключение

Представлена модель агрегирования уровня карантинных мер по 8 критериям, таких, как отмена публичных мероприятий, ограничение на поездки и др. Построена модель и проведен анализ временных рядов числа случаев заболевания коронавирусом.

На основе классических методов кластерного анализа (k-means, иерархическая кластеризация и DBSCAN) и методов анализа паттернов выявлены группы стран со схожей внутренней структурой по совокупности трех показателей: уровень образования, принятые карантинные меры и законопослушность граждан (экспертные оценки). В полученных группах построены регрессионные модели, позволяющие в краткосрочной перспективе рассчитывать прирост числа заболевших COVID-19.

С помощью известных методов DEA и новых методов интервального оболочечного анализа данных (Interval DEA) оценена эффективность применения карантинных мер в крупнейших странах мира.

| Литература

- M.J. Farrell, The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society*, 120(3) (1957) pp. 253–281
- A. Charnes, W.W. Cooper, E. Rhodes, Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2 (1978) pp. 429–444
- F. Aleskerov, S. Demin, An Assessment of the Impact of Natural and Technological Disasters Using a DEA Approach, *Dynamics of Disasters – Key Concepts, Models, Algorithms, and Insights* (2016) pp. 1–14
- F. Aleskerov, V. Petrushenko, DEA by sequential exclusion of alternatives in heterogeneous samples, *International Journal of Information Technology & Decision Making*, 15(01) (2016) pp. 5–22
- F.T. Aleskerov, V. Belousova, S. Petrushchenko, Models of Data Envelopment Analysis and Stochastic Frontier Analysis in the Efficiency Assessment of Universities, *Automation and Remote Control*. 78(5) (2017) pp. 902–923

Литература

- Mirkin B., Myachin A. Ordinal Equivalence Classes for Parallel Coordinates // Intelligent Data Engineering and Automated Learning - IDEAL2019. – Springer. – 2019.
- Myachin A. Pattern Analysis in Parallel Coordinates Based on Pairwise Comparison of Parameters // Automation and Remote Control. – V.80. – №1. – 2019. – P. 112-123.
- Mirkin B. Core Data Analysis: Summarization, Correlation, and Visualization. Cham, Switzerland: Springer, 2019
- Aleskerov F., Chistyakov V., Kalyagin V., Social threshold aggregations // Social Choice and Welfare. 2010. Vol. 35, No. 4, pp. 627-646.
- Алескеров Ф.Т., Якуба В.И. Метод порогового агрегирования трехградационных ранжировок // ДАН. 2007. Т. 413, №2. С. 181-183.
- Ф.Т. Алескеров, М.А. Голубенко Об оценке симметричности политических взглядов и поляризованности общества Препринт WP7/2003/04
- Алескеров Ф.Т., Вольский В.И., Якуба В.И. Применение индексов симметричности и поляризованности для анализа динамики потребления / 4-я Международная конференция по проблемам управления (МКПУ) 2009.

Спасибо за
внимание!