

# Прогнозирование макроэкономических рядов с помощью алгоритма SSA

Волкова Анастасия, НИУ ВШЭ, Москва, 2018

## Что за алгоритм?

Одним из сильных и перспективных методов анализа и прогнозирования временных рядов является метод SSA (Singular Spectrum Analysis), в России известный как «Гусеница». Одна из отличительных черт алгоритма – возможность проводить исследование структуры ряда (выявление трендовой, гармонических и шумовых составляющих) без предположения о его модели. Также, алгоритм неплохо работает на зашумленных рядах и рядах с выбросами. Зашумленные ряды метод гусеницы сглаживает и получается достаточно хороший прогноз. Кроме того, SSA достаточно быстро работает.

## Как это работает?

Алгоритм состоит из 4 основных этапов:

- Построение траекторной матрицы
- Сингулярное разложение
- Группировка матриц
- Диагональное усреднение

## Обозначения:

$(f_i)_{i=1}^N$  – исходный временной ряд  
 $N$  – длина ряда  
 $f_i$  –  $i$ -ая компонента ряда

## Траекторная матрица

Нужно уместить наш исходный ряд в матрицу  $X$  размерности  $L \times K$ , где  $K = N - L + 1$

$L$  – длина гусеницы (экзогенный параметр)

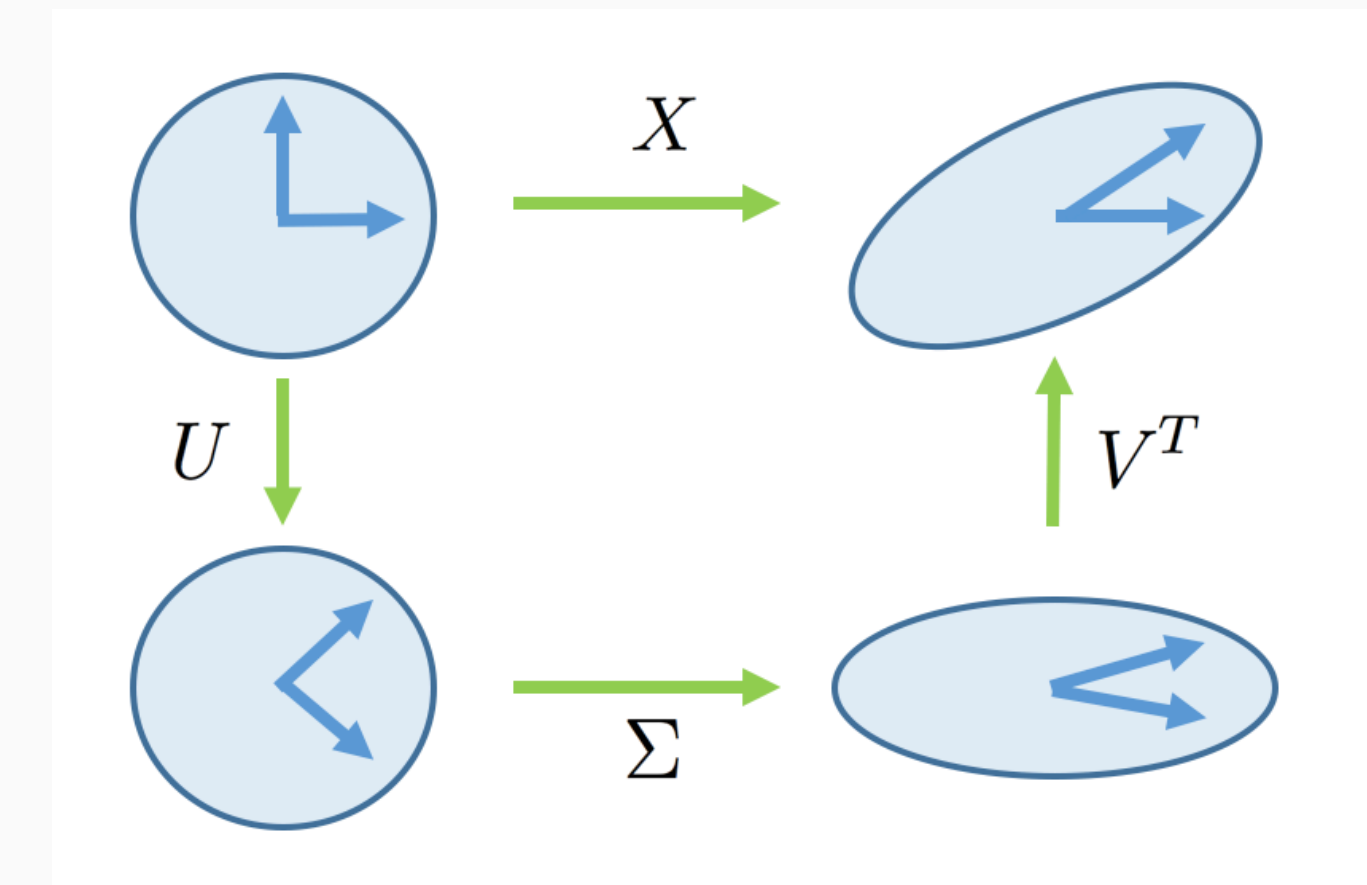
Можно провести аналогию с очередью в магазине: каждый первый человек из очереди уходит, а вновь пришедший встает в конец, и так  $L$  раз.

$$X = \begin{pmatrix} f_1 & f_2 & \dots & f_{N-L+1} \\ f_2 & f_3 & \dots & f_{N-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & \dots & f_N \end{pmatrix}$$

## SVD - разложение

Представим матрицу  $X$  в виде произведения матриц.

$$X = U \times \Sigma \times V^T$$



## Группировка матриц

Делим все матрицы на группы: тренд, сезонность, шум.

## Диагональное усреднение

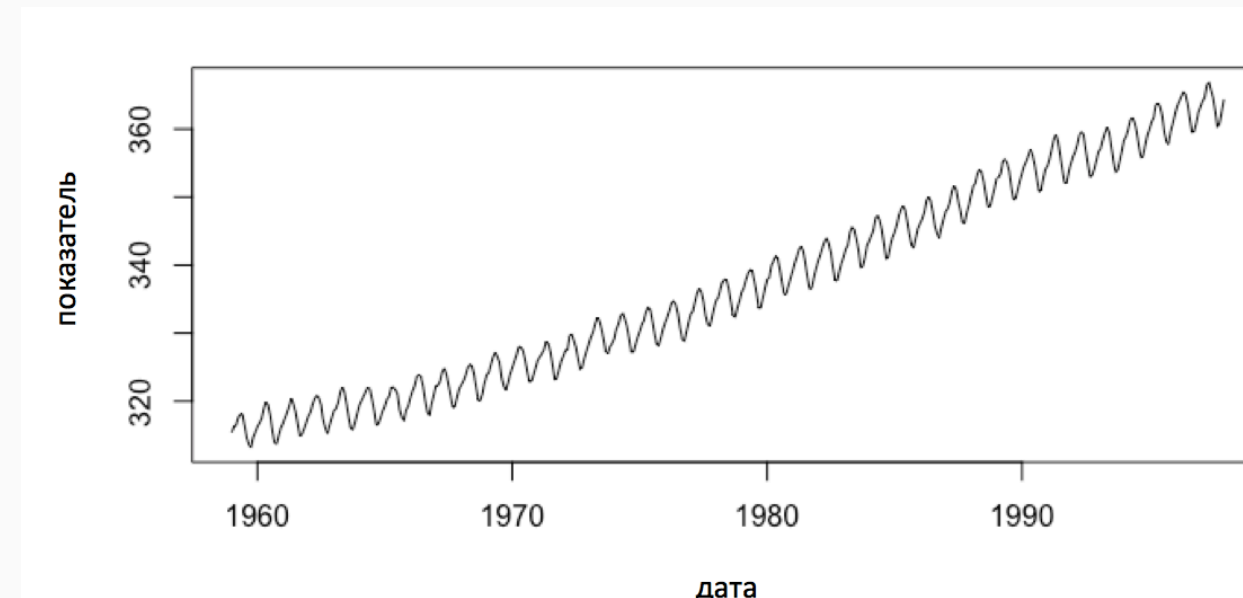
$$Y_i = \begin{pmatrix} f_1 & f_2 & \dots & f_{L^*} & \dots & f_{K^*} & \dots & f_{K^*+1} \\ f_2 & f_3 & \dots & f_{L^*} & \dots & f_{K^*} & \dots & f_{K^*+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & f_{L^*} & \dots & f_{K^*} & \dots & \vdots & \vdots & \vdots \\ f_{L^*} & \dots & f_{K^*} & \dots & \dots & \dots & \dots & f_N \end{pmatrix}$$

## Guide для новичков

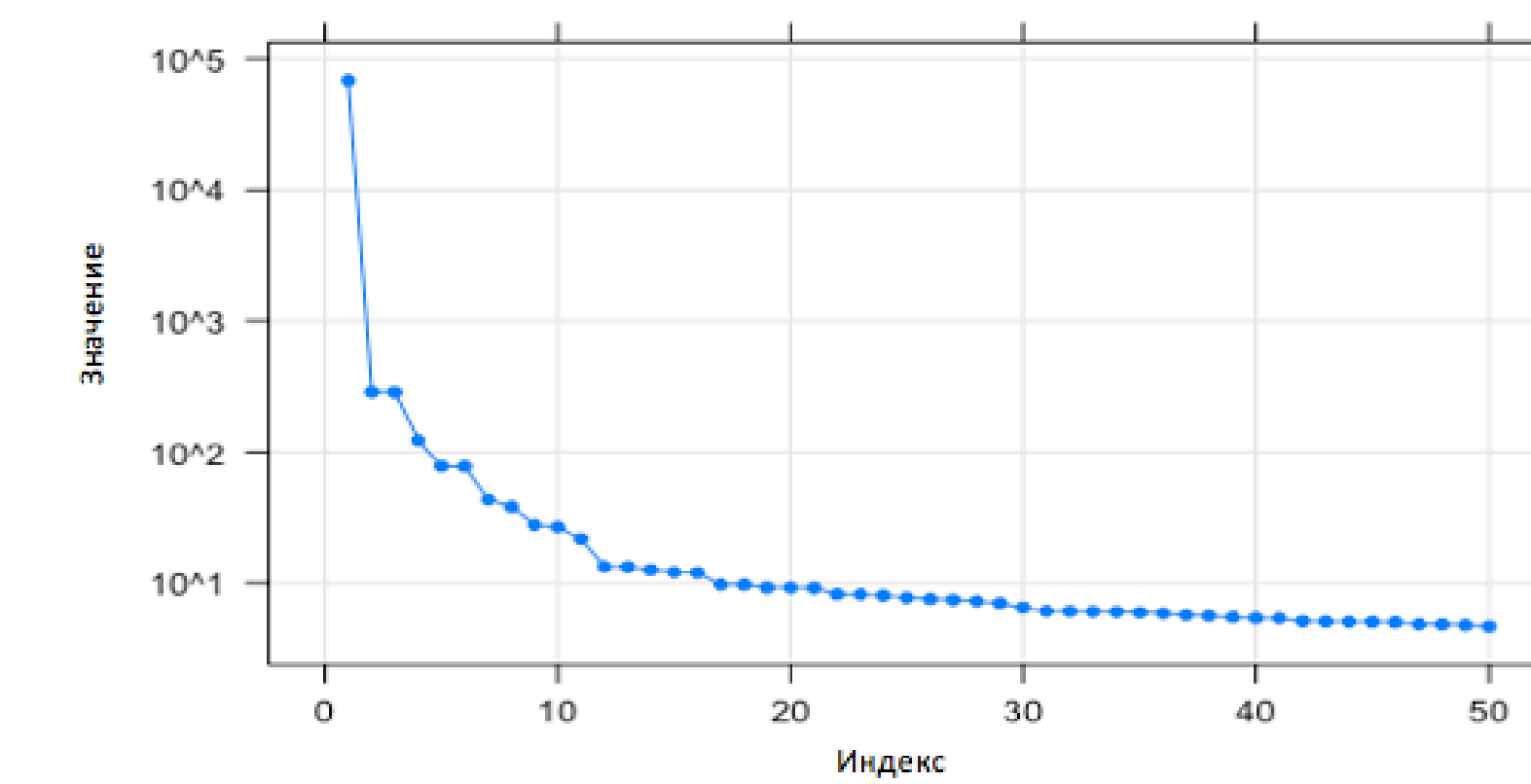
- Установите пакет Rssa.
- Загрузите данные, которые хотите исследовать.
- Для декомпозиции ряда используйте команду `ssa()`, указав длину гусеницы.
- Рекомендуется выбрать  $L$ , кратное периодике ряда (посчитать углы у фигур пар векторов).
- Для реконструкции ряда используйте команду `reconstruct()`, указав группы векторов.
- Пары сгруппированных векторов должны образовывать четкие геометрические фигуры.
- Построить прогноз можно с помощью команды `vforecast()`
- Проверить качество прогноза можно с помощью пакета `accuracy`

## Проанализируем ряд

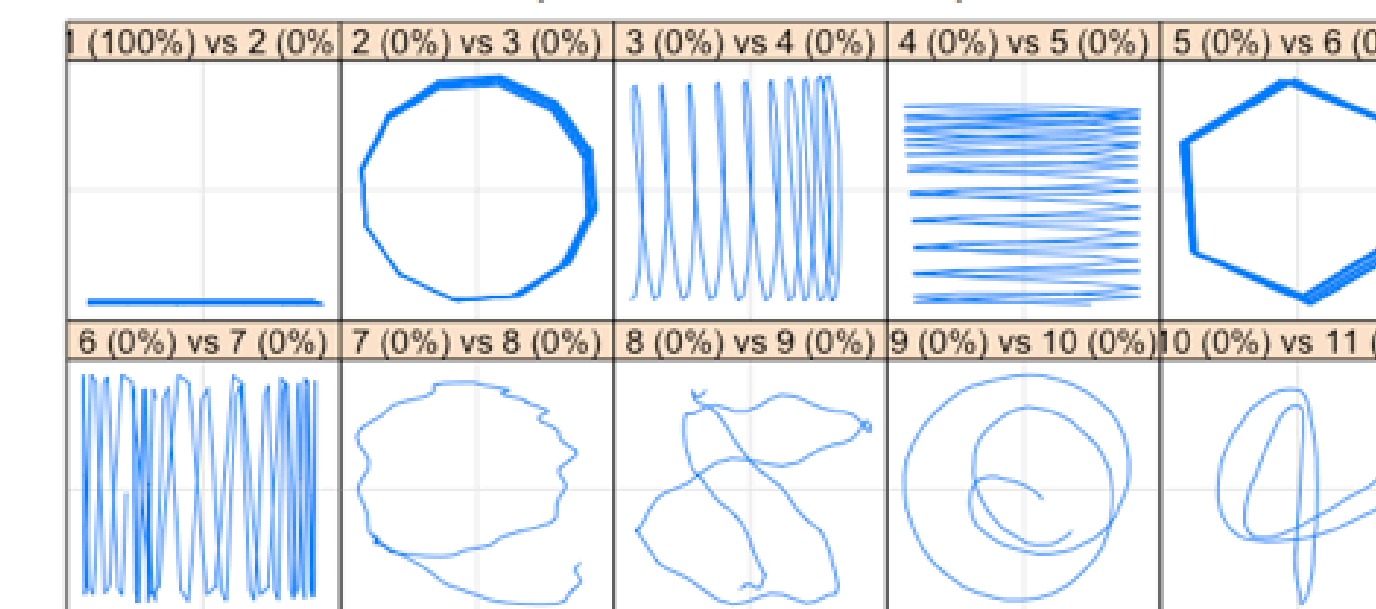
Ряд `so2`. 468 ежемесячных наблюдений уровня CO2 близ вулкана Мауна Лоа с 1959 по 1997 годы.



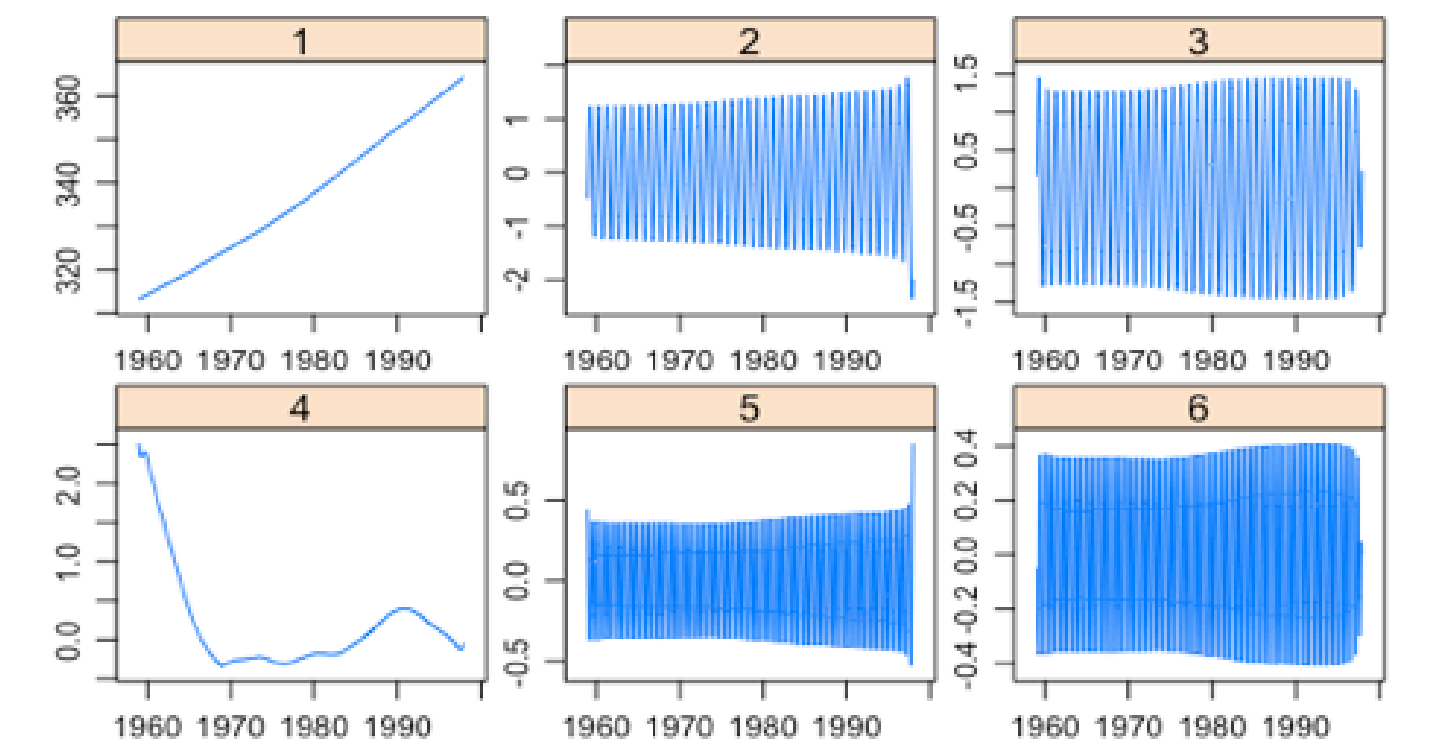
Собственные значения



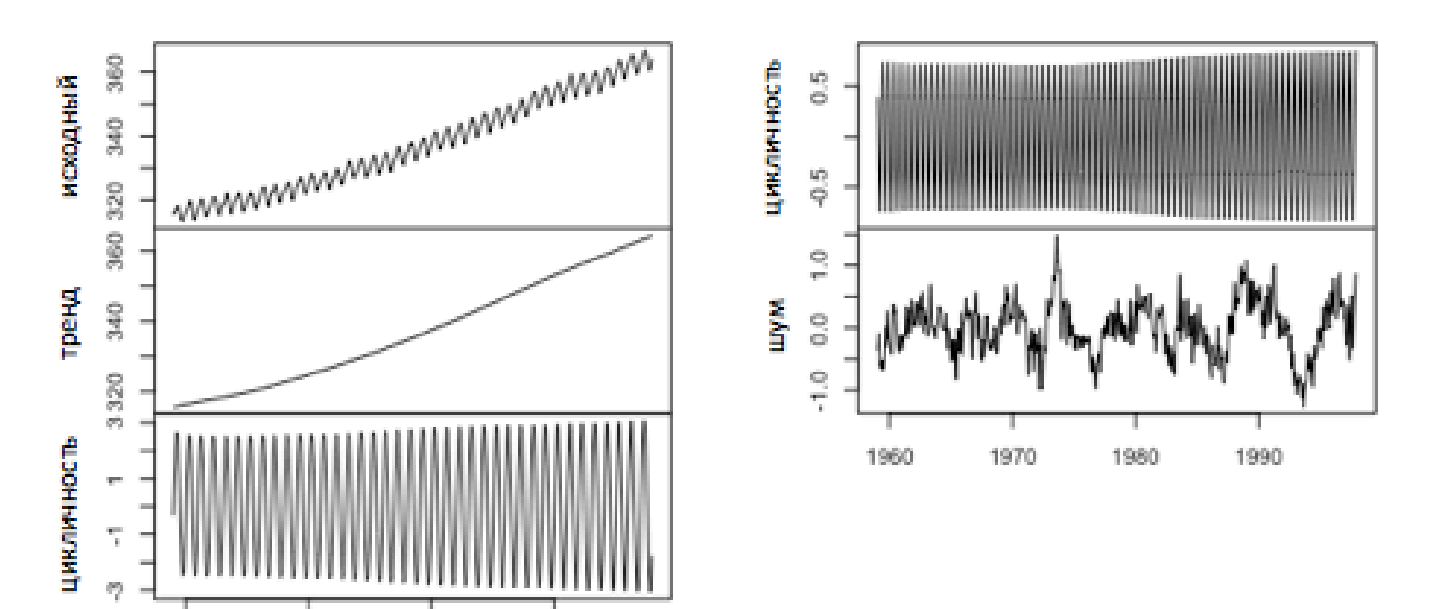
Пары собственных векторов



Реконструкция



Реконструкция по компонентам



## Подбор универсальной длины

Чтобы проверить есть ли некая зависимость точности прогнозирования от длины окна и количества выбранных компонентов в общем случае, проанализируем изменения качества алгоритма при разных параметрах для нескольких рядов. Для исследования возьмем первые 10 рядов из дата-сета `M4comp`. Исходя из графиков, нельзя сделать однозначные выводы о том, какое значение лучше. Что хорошо подходит одним рядам, совершенно не подходит другим. Хотя можно заметить, что  $L = 120$  в среднем дает неплохой прогноз для всех рядов (качество не самое лучшее, но для всех 10 рядов достаточно однородно).

В качестве тренда возьмем 1 и 4 вектора, в качестве сезонности и шума 2 и 3, 5 и 6, так как видно, что они имеют похожую структуру и четкие фигуры. Произведем реконструкцию в соответствии с новыми группами.

## Выводы

Подводя итог, можно сделать вывод, что для алгоритма SSA лучше настраивать параметры вручную. Данный алгоритм дает неплохой прогноз, но для этого необходим тщательный подбор параметров модели, исходя из особенностей исследуемого ряда. Прежде всего необходимо учитывать периодичность ряда и фигуры пар векторов. Кроме того, сложные ряды рекомендуется сначала сглаживать, выбрав относительно небольшое значение длины гусеницы, а уже затем проводить анализ реконструированного ряда.

