# Использованные методы

## from sklearn.linear\_model import **LinearRegression**

Линейная регрессия. В качестве предикторов используется 5 независимых переменных, выгруженных с сайта [FRED](https://fred.stlouisfed.org/tags/series?t=eu%3Beurope%3Bmei%3Bmonthly&rt=europe&ob=pv&od=desc). Сдвиг предикторов -1 **⇒** регрессия вычисляет значения на 1 месяц вперёд. Сумма предикторов Exports + Imports = Net Trade ⇒ для избавления от избыточностинужно выбрать один. Мощность массива “train” 18 значений, мощность массива “test” 12 значений. Реализовано два метода вычисления коэффициентов:

* Append – использованный прогноз после появления «настоящих» данных добавляется в пул “train”, увеличивая его мощность .
* Override – использованный прогноз перезаписывает самое старое значение в пуле “train”, мощность сохраняется неизменной .

Append имеет выше , ниже среднее и медианное MAPE **⇒** Append работает лучше. Append имеет опасность «переобучения». Минусы:

* Преимущественно отрицательный нулевая гипотеза верна, данные лучше описываются прямой линией.
* Отсутствие предикторов для предсказания с горизонтом > 1 месяца.
* Предикторы появляются поздно. Сайт FRED имеет отставание в год, на сайте [OECD](https://www.oecd.org/sdd/oecdmaineconomicindicatorsmei.htm) данные за ноябрь на момент начала декабря отсутствуют.





## FRED Data

Информациясо страницы README в xlsx при выгрузке с сайта:

* CPHPTT01EUM661N, Consumer Price Index: Harmonized Prices: Total All Items for the European Union, Monthly, Not Seasonally Adjusted
* LRHUTTTTEUM156N, Harmonized Unemployment Rate: Total: All Persons for the European Union, Monthly, Not Seasonally Adjusted
* PITGCG01EUM661N, Producer Prices Index: Total Consumer Goods for the European Union, Monthly, Not Seasonally Adjusted
* XTEXVA01EUM667S, Exports: Value Goods for the European Union, Monthly, Seasonally Adjusted
* XTIMVA01EUM667S, Imports: Value Goods for the European Union, Monthly, Seasonally Adjusted
* XTNTVA01EUM667S, Net Trade: Value Goods for the European Union, Monthly, Seasonally Adjusted

Европейский Союз как крупнейшее «глобальное» объединения стран на мировом рынке.

## Таблица LinearReg Append



## Таблица LinearReg Override



## from statsmodels.tsa.api import **ExponentialSmoothing**

Экспоненциальное сглаживание/Экспоненциальное сглаживание «с трендом». Средний эксп. сглаживания – 0,29, эксп. сглаживания «с трендом» - 0,16 ⇒ данные имеют тренд. Использовались 2 метода вычисления сглаженной экспоненты: append и override, разницей можно пренебречь. Мощность массива “train” была установлена в 24 значения, мощность массива “test” в 12 значений.





## Таблица ExpSmoothing Append

## Таблица ExpSmoothing Override

## statsmodels.tsa.ar\_model import **AutoReg**

Авторегрессионная модель, в качестве предикторов используется 2 последних значения временного ряда. Таблицу значений предсказаний по commodities на 1 месяц вперёд можно посмотреть далее, количество коэффициентов равно параметру lag + 1. Мощность массива “train” 24 значения, мощность массива “test” 12 значений.

После получения данных за первый «предсказанный» месяц был выбран оптимальный (лучшее количество выбиралось путём минимизации суммы среднего значения MAPE и медианного значения MAPE. При равном минимуме считается, что меньшее количество коэффициентов оптимальнее) lag для каждого commodity, чтобы уточнить дальнейшие предсказания на 3, 6, … месяцев.



## Таблица AutoReg



## ARIMA

В общем виде модель ARMA(p,q), где p – порядок авторегрессии, q – порядок скользящего среднего. Поскольку наши временные ряды являются нестационарными, для приведения его к стационарному виду необходимо несколько разностей **⇒** модель становится ARIMA(p,d,q), где d – порядок разности. Параметры p, q, d = 1, 1, 1. Мощность массива “train” 24 значения, мощность массива “test” 12 значений.





## Таблица ARIMA Append



## Таблица ARIMA Override



# Выводы

При минимизации MAPE и максимизации можно выделить два варианта: авторегрессия и ARIMA моделирование. Авторегрессия имеет меньший разброс между минимальным и максимальным MAPE при чуть больших средних значениях на 1 месяц вперёд, нежели ARIMA. Точность предсказания с помощью авторегрессии высокая на месяц вперёд и сильно снижается на 3 и 6 месяцев.



Из-за скользящего среднего ARIMA гораздо лучше показывает себя при предсказании на 3 и 6 месяцев.



Точность ARIMA можно улучшить точным подбором p, q, d либо с помощью grid search либо посредством статистического анализа каждого commodity.



В сравнении с наивным предсказанием Danone от 2020 значительно меньше медианное и минимальное значение MAPE на 3 и 6 месяцев, лучше среднее на 3 месяца.