

# Сравнение алгоритмов прогнозирования рядов на базе M4

Факультет Экономических Наук НИУ ВШЭ

Волкова Анастасия

В 2018 году проходили соревнования по прогнозированию временных рядов **M4 Competition**. Второе место занял алгоритм **M4metalearning**.

Данный метод был разработан исследователем Пабло Монтеро-Мансо из университета Испании Ла-Корунья и его австралийскими коллегами из университета Монаша, включая известного статистика **Роба Хиндмана**. Новизна алгоритма состоит в линейном комбинировании прогнозов 9 стандартных методов, веса для которых вычисляются с помощью градиентного бустинга решающих деревьев.

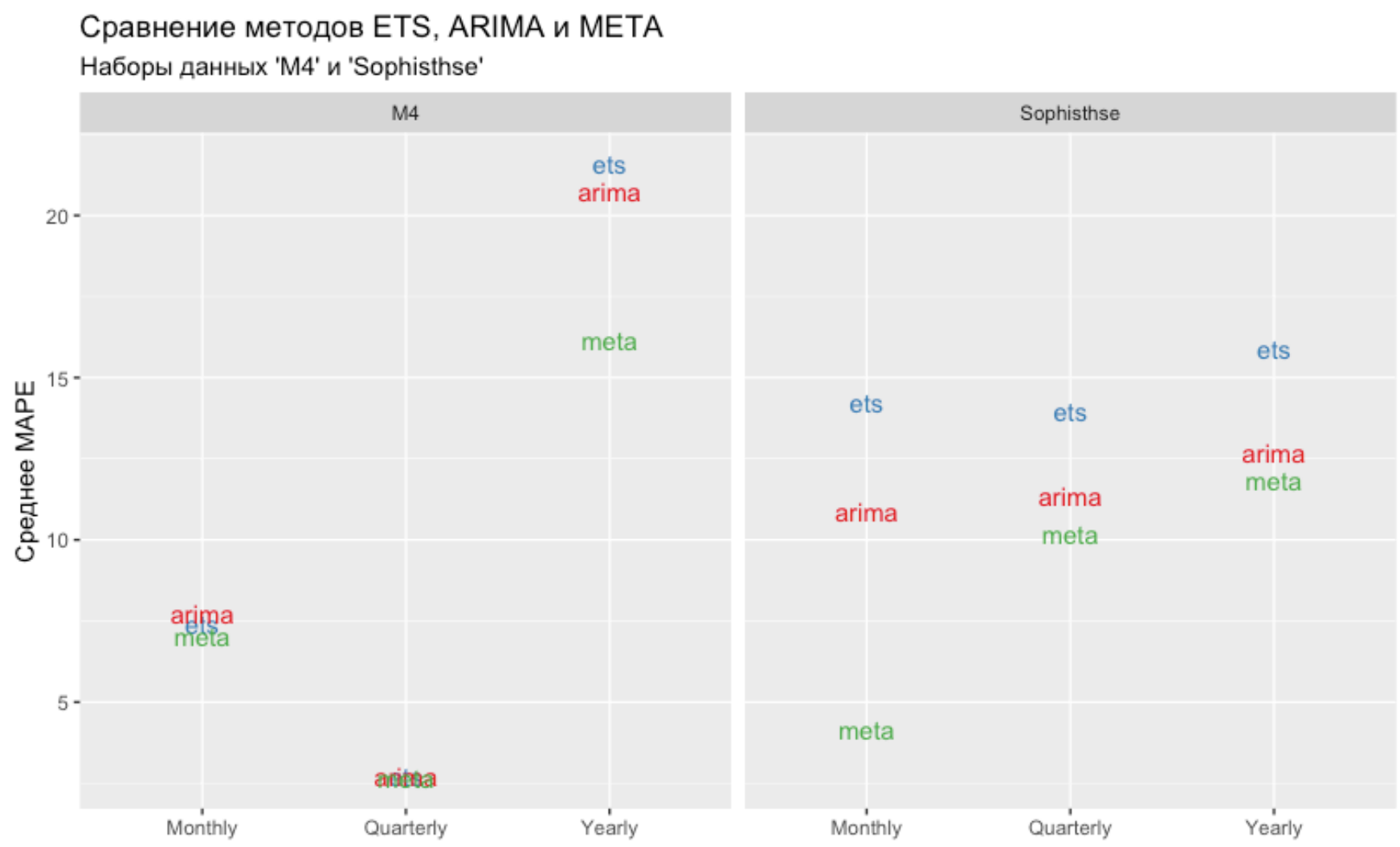
### Комбинация составляется из 9 базовых методов прогнозирования:

- *ARIMA* модель, параметры  $p$ ,  $d$  и  $q$  настраиваются автоматически.
- *ETS* модель, параметры настраиваются автоматически.
- *NNAR* - нейронная сеть прямого распространения с одним скрытым слоем, обучающаяся на лагах. Количество лагов настраивается автоматически.
- *TBATS* модель включает в себя тригонометрические методы для распознавания сезонности, которая может меняться со временем, трансформацию Бокса-Кокса от гетероскедастичности, *ARMA* ошибки, тренд и сезонность подобно ETS. Параметры настраиваются автоматически.
- *STLM-AR* метод сезонной трансформации с помощью *LOESS* с *AR* моделированием очищенных от сезонности рядов.
- *THETA* метод выиграл M3 Forecasting Competition в 2000 году.
- *NAIVE* - случайное блуждание. Наивный прогноз с помощью последнего наблюдения.
- *RW-DRIFT* - модель случайного блуждания со сдвигом. Частный случай *ARIMA(0,1,0)* с константой.
- *SEASONAL NAIVE* – наивный прогноз с учетом сезонности.

Авторы предлагают использовать модель, предобученную ими на 100 000 временных рядах датасета M4, но ее можно обучить и самостоятельно.

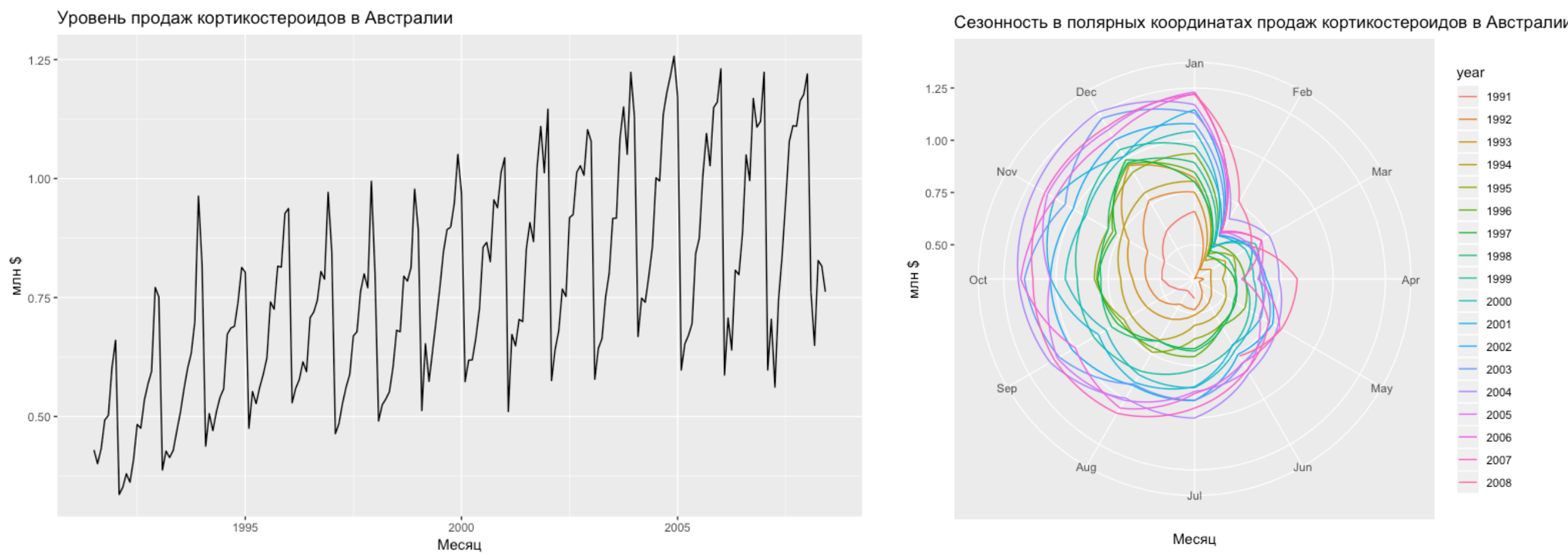
Протестируем качество алгоритма на некотором количестве рядов разной частоты из M4 и на не знакомых ей ранее рядах российской базы данных Sophisthse.

Каждый ряд был спрогнозирован тремя методами: ETS, ARIMA и META. В качестве метрики использована MAPE.



Проведем очень краткий обзор методов прогнозирования, используемых в алгоритме **M4metalearning**. Для наглядности будем применять их к конкретному временному ряду. Например, возьмем данные по ежемесячным продажам кортикостероидных препаратов H02 в Австралии с 1992 по 2010 год.

Заметим, что ряд нестационарный, присутствует возрастающий тренд, а также небольшой рост дисперсии по мере увеличения уровня ряда. Данную проблему можно исправить с помощью логарифмирования. Также присутствует сезонность, посмотрим на нее в полярных координатах.



Худшее качество у моделей случайного блуждания, которые не учитывают сезонность. Остальные методы прогнозирования показали относительно схожее качество на данном примере.

