# Постановка задачи ИАД

Представлен временной ряд объема производства промышленного предприятия (в млн. руб.) по месяцам с 2004 по 2019 гг. Необходимо построить прогнозную модель на основе имеющихся данных и осуществить прогнозирование объема производства на январь-март 2020.

# Графический разведочный анализ данных

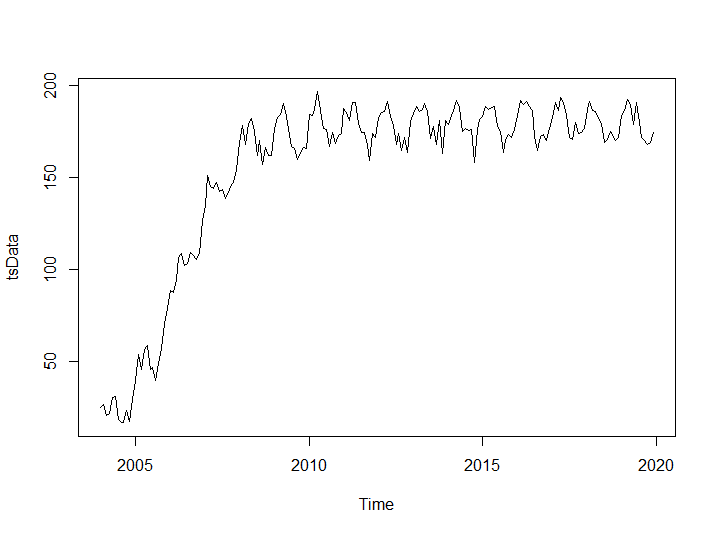


Рис. 1 График временного ряда

На рис.1 видно, что тренд данного временного ряда нелинейный. Скорее всего тренд является логистическим и устойчиво возрастает. Присутствует сезонная компонента. Также ряд не стационарен.

## Декомпозиция временного ряда на составляющие

Для того, чтобы приступить к декомпозиции временного ряда на составляющие, нужно определить тип сезонного компонента. Сезонная компонента постоянна и со временем не меняет свое значение, для декомпозиции используем тип аддитивный.

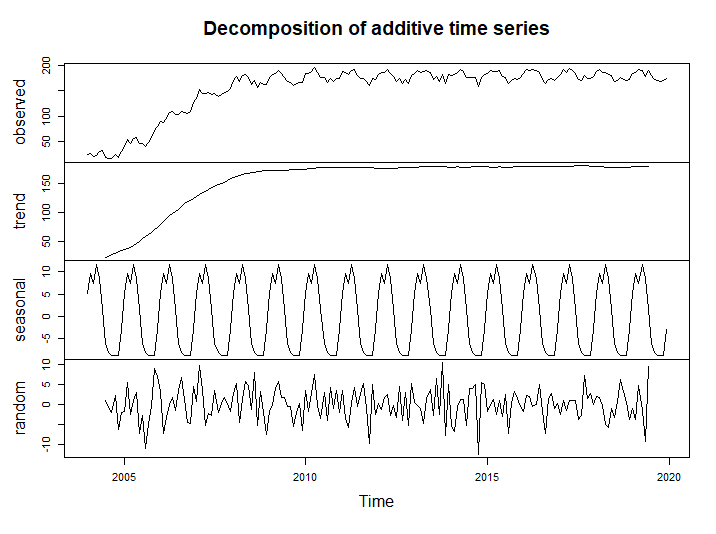


Рис. 2 Декомпозиция временного ряда

На рис.2 также видно, что тренд данного временного ряда нелинейный. Больше всего тренд похож логистический или логарифмический и устойчиво возрастает. Присутствует сезонная компонента, не меняющаяся со временем. Также ряд не стационарен.

# Построение модели на основе метода последовательной идентификации

## Идентификация тренда

По результатам визуального анализа графика и декомпозиции временного ряда на составляющие можно предположить, что у временного ряда есть логарифмический тренд. Однако для более точной идентификации вида тренда воспользуемся функцией nls(), которая оценивает параметры нелинейной модели на основе нелинейного метода наименьших квадратов.

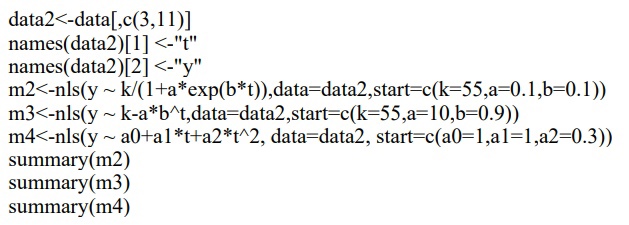
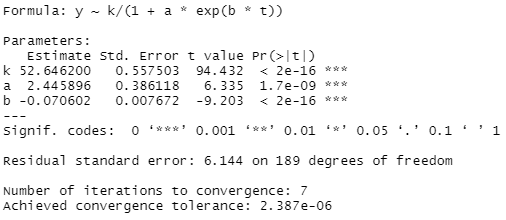
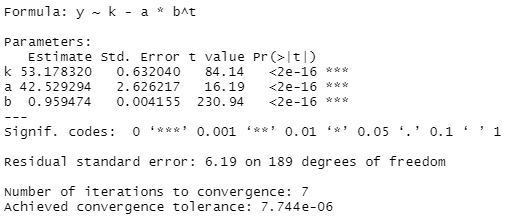


Рис. 4 Код для работы с функцией.





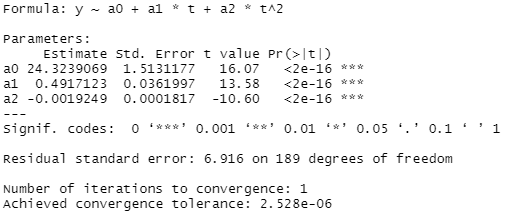
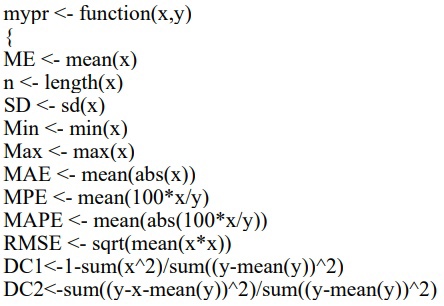


Рис. 5 Данные трех моделей.

Нахождение временных рядов остатков модели реализуется с помощью функции residuals().

Для вычисления статистических характеристик модельных остатков удобно написать собственную функцию. В функции mypr() реализовано вычисление характеристик точности модели. Аргументы функции x – временной ряд остатков; y – исходный временной ряд:



Результаты применения функции mypr() к ВР остатков построенных моделей тренда приведены в табл. 1:

|  |  |  |  |
| --- | --- | --- | --- |
| Статистическая характеристика | Модель 2 | Модель 3 | Модель 4 |
| n | 192 | 192 | 192 |
| SD | 6.11 | 6,16 | 6.88 |
| Min | -14.3 | -13.7 | -15 |
| Max | 15.7 | 15.3 | 15.4 |
| ME | -0.00242 | 2.71e-08 | -1.09e-05 |
| MAE | 5.02 | 5.06 | 5.59 |
| MPE | -1.89 | -1.87 | -3.2 |
| MAPE | 11.3 | 11.5 | 13.3 |
| RMSE | 6.1 | 6.14 | 6.86 |
| DC1 | 0.689 | 0.684 | 0.606 |
| DC2 | 0.687 | 0.684 | 0.606 |

Анализ данных табл. 1 позволяет сделать вывод о хорошей точности построенных моделей тренда: средняя абсолютная процентная ошибка (MAPE) составляет 11.3…13.3 %, коэффициент детерминации (DC1, DC2) равен 0.606…0.689. Для всех моделей тренда получены близкие по значениям статистические характеристики остатков. Формально по совокупности характеристик лучшая модель тренда – логистическая.

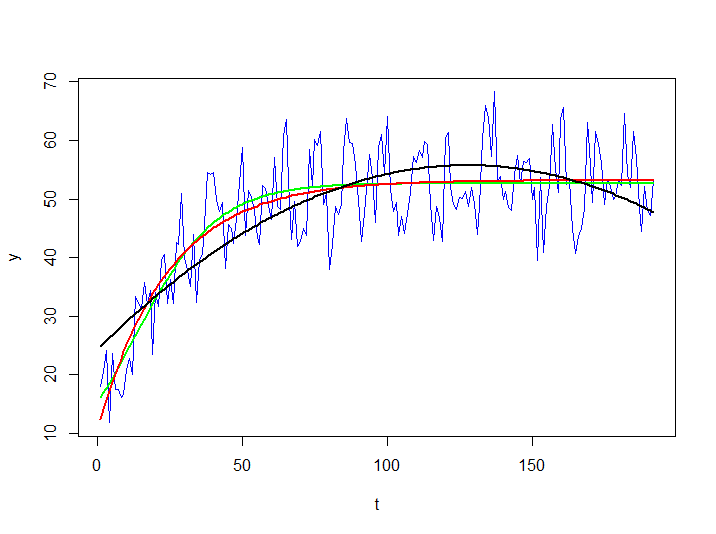


Рис. 7. Временной ряд с наложением моделей тренда

На графике (рис. 7) модель модифицированной экспоненты и логистическая модель расположены очень близко друг к другу и визуально практически совпадают. Параболическая модель проходит точку экстремума (максимума) на наблюдаемом интервале изменения временного ряда и в дальнейшем прогнозирует уменьшение значений ВР, что визуально плохо согласуется с имеющимися наблюдениями. Поэтому в качестве модели тренда можно рекомендовать либо логистическую кривую, либо модель модифицированной экспоненты.