

**Использование  
экспоненциальных методов для  
анализа временных рядов**

Графеева Н.Г.

2016

# Какие задачи могут решаться экспоненциальными методами?

- сглаживание временных рядов (smoothing);
- разметка временных рядов (labeling);
- краткосрочное и долгосрочное прогнозирование (forecasting);
- И др.

# Популярные экспоненциальные методы

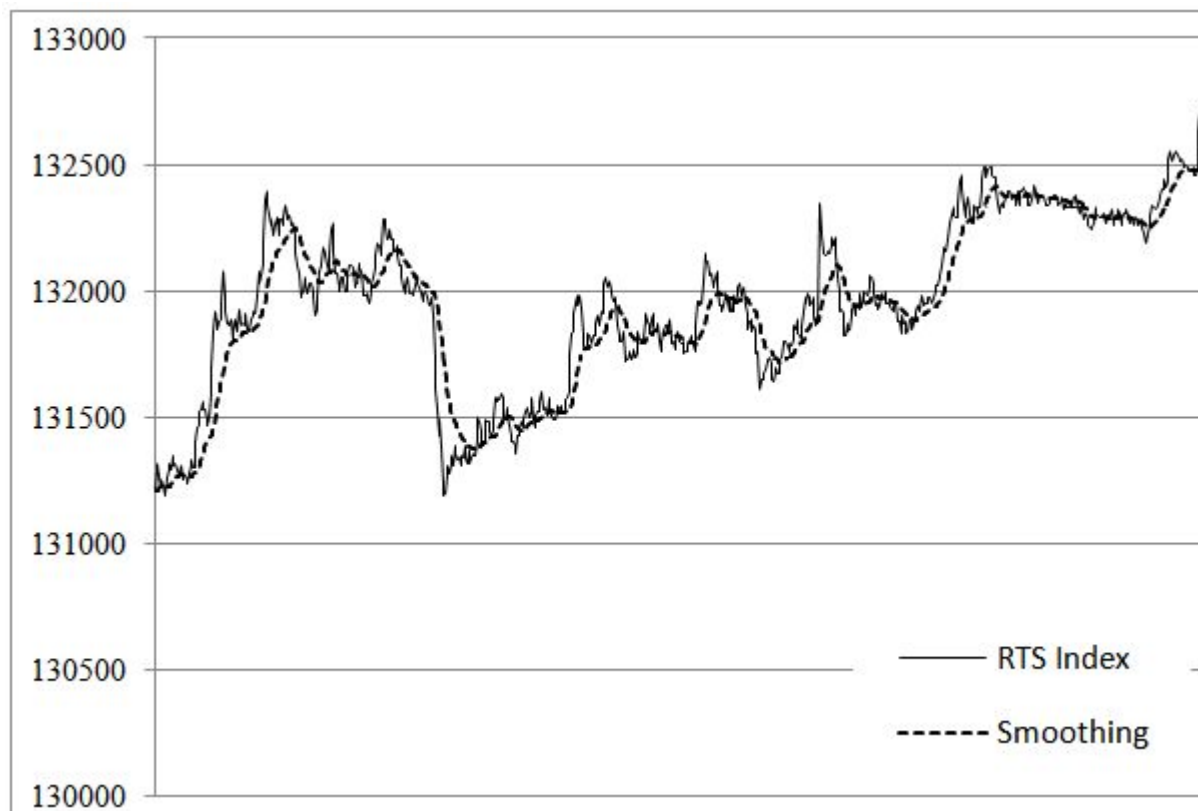
- Одинарный (учитывает предыдущие значения с коэффициентами);
- Двойной (добавляется учет трендов);
- Тройной (добавляется учет циклов).

# Одинарное экспоненциальное сглаживание (smoothing)

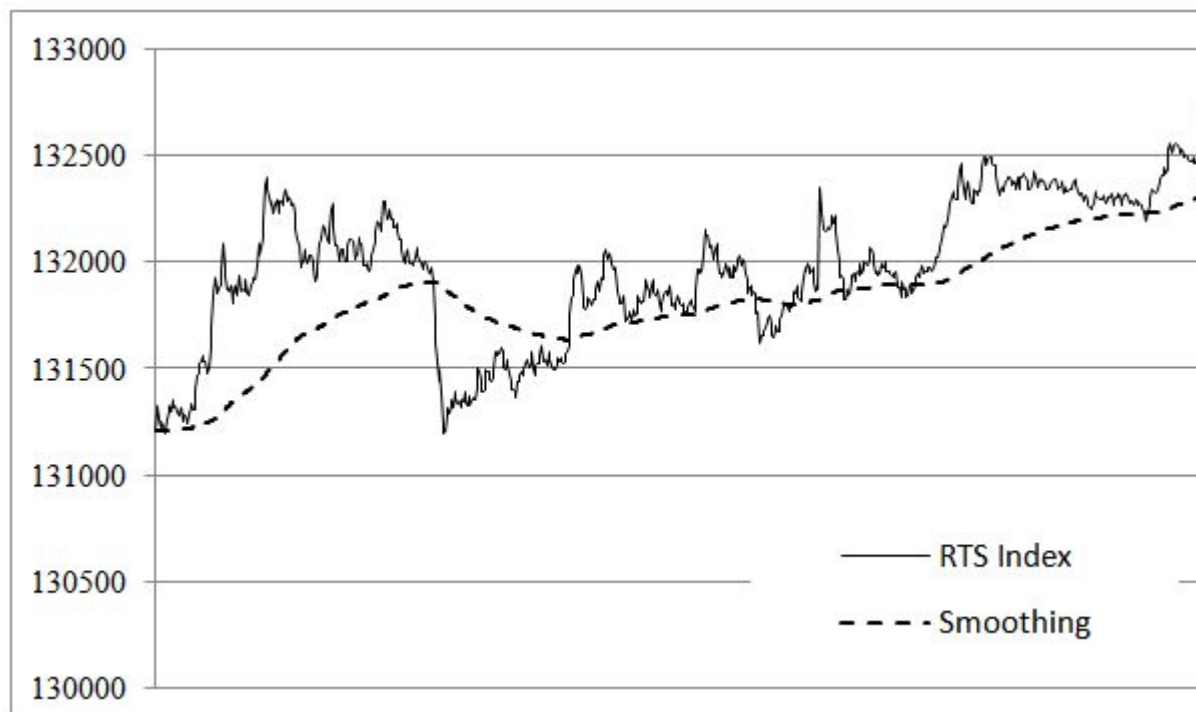
$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad t > 1, \quad 0 \leq \alpha \leq 1$$

$$S_1 = y_1, \quad t = 1$$

# Пример. Одинарное экспоненциальное сглаживание ( $\alpha$ =0.1)



# Пример. Одинарное экспоненциальное сглаживание ( $\alpha$ =0.01)



# Разметка трендов с помощью одинарного экспоненциального сглаживания

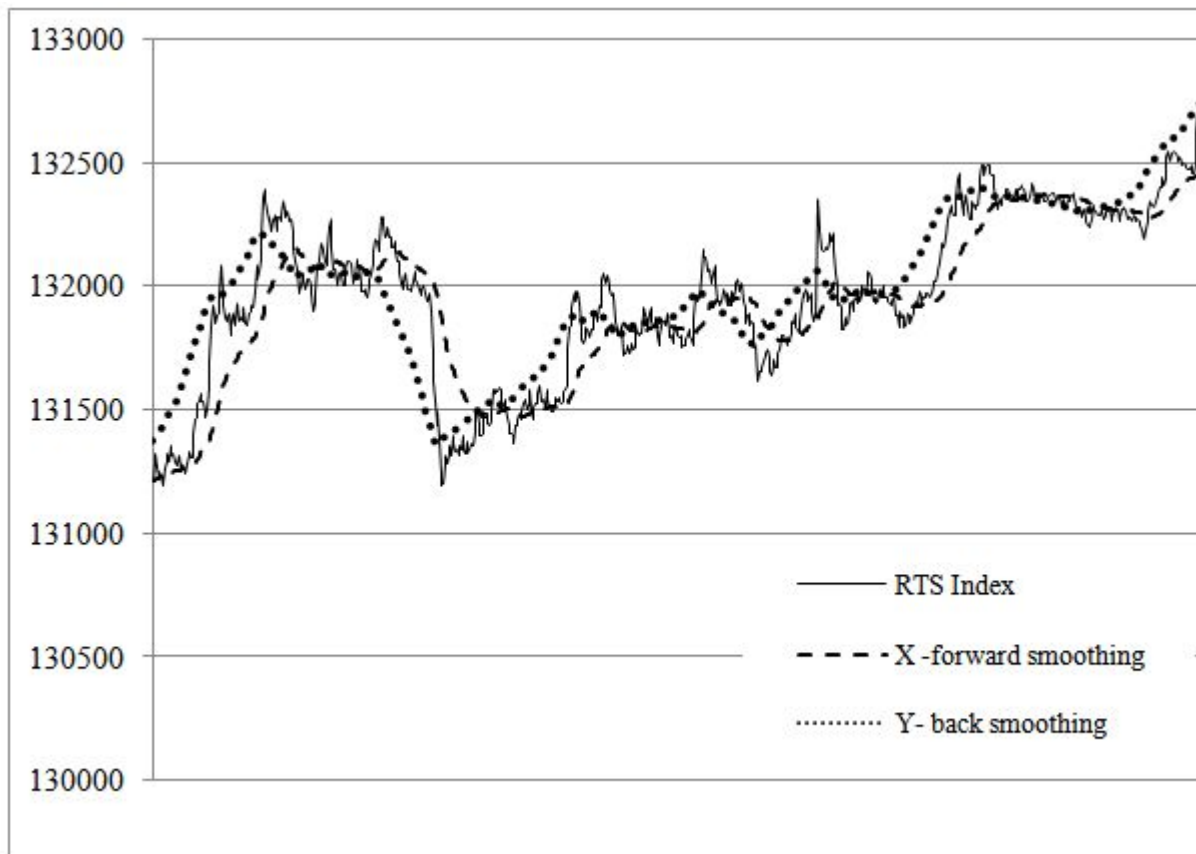
$$X_t = \alpha y_t + (1 - \alpha)X_{t-1}, \quad t > 1 \text{ (forward smoothing)}$$

$$X_1 = y_1, \quad t = 1$$

$$Z_t = \alpha y_t + (1 - \alpha)Z_{t+1}, \quad t < n \text{ (back smoothing)}$$

$$Z_n = y_n, \quad t = n$$

# Пример. Разметка трендов ( $\alpha = 0.05$ )





# Как формально определить тренды?

$$X_t > Z_t, |X_t - Z_t| > \sigma$$

$$X_t < Z_t, |X_t - Z_t| > \sigma$$

$$|X_t - Z_t| < \sigma$$

descending trend

ascending trend

sideways trend

# Прогнозирование на один шаг вперед с помощью одинарного сглаживания (Single Smoothing Forecast)

$$S_t = \alpha y_t + (1 - \alpha)S_{t-1}, \quad t > 1, \quad 0 \leq \alpha \leq 1$$

$$S_1 = y_1, \quad t = 1$$

$$F_{t+1} = S_t \text{ (forecasting)}$$

# Прогнозирование на несколько шагов вперед (Bootstrap Forecast)

$$S_t = \alpha y_{origin} + (1 - \alpha)S_{t-1} \text{ (smoothing)}$$

$$F_{t+1} = S_t \text{ (forecasting)}$$

# Пример. Сглаживание на несколько шагов вперед

Пусть последнее значение (в точке  $t$ ) последовательности равно 70, а его сглаженное значение 71.5. При этом  $\alpha = 0.1$ . Тогда значение в точке  $(t+1)$  вычисляется так:

$$\begin{aligned} S_{t+1} &= \alpha y_{origin} + (1 - \alpha) S_t \\ &= 0.1(70) + 0.9(71.5) \\ &= 71.5. \end{aligned}$$

А в точке  $(t+2)$  так:

$$S_{t+2} = 0.1(70) + 0.9(71.5) = 71.35$$

# Сравнение прогнозирования на один шаг и на несколько шагов

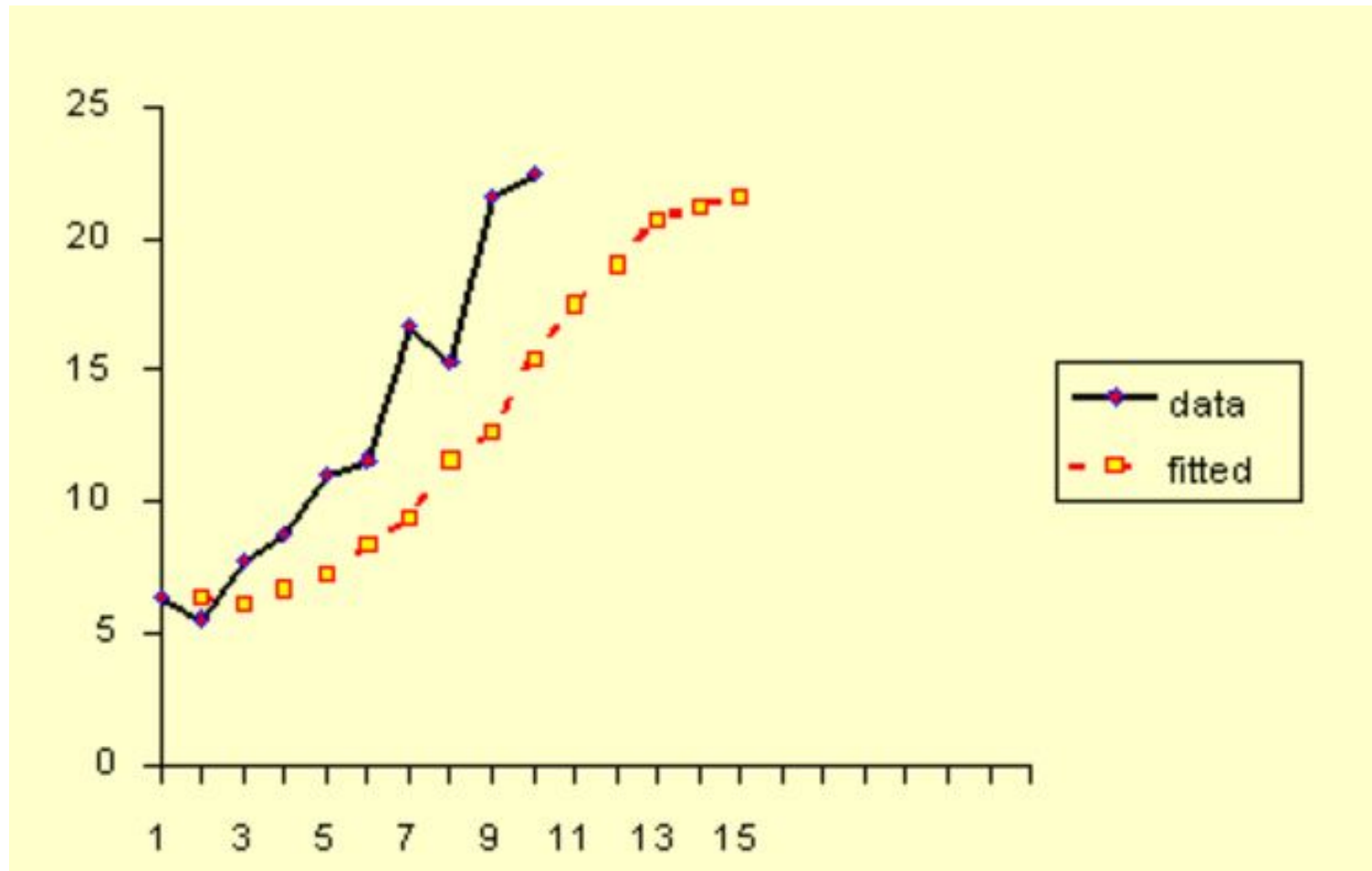
<b>Bootstrap forecast</b>	<b>Data</b>	<b>Single Smoothing Forecast</b>
71.50	75	71.5
71.35	75	71.9
71.21	74	72.2
71.09	78	72.4
70.98	86	73.0

# Одинарное экспоненциальное сглаживание и тренды

Одинарное экспоненциальное сглаживание не слишком хорошо сглаживает тренды. Приведем пример данных, имеющих склонность к наличию тренда, и соответствующих вычисленных сглаженных значений при  $\alpha = 0.3$ .

y	6.4	5.6	7.8	8.8	11.0	11.6	16.7	15.3	21.6	22.4
S		6.4	6.2	6.7	7.3	8.4	9.4	11.6	12.7	15.4

# Пример одинарного экспоненциального сглаживания



# Двойное экспоненциальное сглаживание (Double Exponential Smoothing)

Гораздо лучше трендовые тенденции в исходных данных сглаживаются с помощью технологии двойного экспоненциального сглаживания. Соответствующие формулы, учитывающие трендовые тенденции, принимают вид:

$$S_t = \alpha y_t + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad 0 \leq \gamma \leq 1$$



# Начальные значения для трендовой компоненты

Есть несколько вариантов, как можно задать начальные значения для рекуррентных формул. Значение для  $S_1$  устанавливается как  $y_1$ . А вот для  $b_1$  возможны варианты. Например:

$$b_1 = y_2 - y_1$$

$$b_1 = \frac{1}{3}[(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]$$

$$b_1 = \frac{y_n - y_1}{n - 1}$$

# Как подобрать подходящие параметры?

- Оптимальные значения для  $\alpha$  и  $\gamma$  могут быть получены с помощью нелинейной оптимизационной технологии известной под названием Marquardt Algorithm (или **Алгоритм Левенберга — Марквардта**) либо с использованием самого примитивного перебора с равномерным шагом по сетке в диапазоне  $[0-1, 0-1]$ .

# Прогнозирование с двойным экспоненциальным сглаживанием

Прогнозирование на один шаг:

$$F_{t+1} = S_t + b_t$$

Прогнозирование на  $m$  шагов:

$$F_{t+m} = S_t + mb_t$$

# Пример

Подготовим пример, позволяющий сравнить два подхода к сглаживанию и прогнозированию. Исходные данные – ряд:

**6.4, 5.6, 7.8, 8.8, 11, 11.6, 16.7, 15.3, 21.6, 22.4**

При двойном сглаживании лучшие результаты были получены при  $\alpha = 0.3623$  и  $\gamma = 0.1$ . Для одинарного сглаживания лучшим параметром оказался  $\alpha = 0.977$ . Начальное приближение для  $b_1$  было выбрано так:

$$b_1 = ((y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3))/3 = 0.8$$

# Результаты сглаживания (и прогнозирования на один шаг)

<b>Data</b>	<b>Double</b>	<b>Single</b>
6.4	6.4	
5.6	6.6 (Forecast = 7.2)	6.4
7.8	7.2 (Forecast = 6.8)	5.6
8.8	8.1 (Forecast = 7.8)	7.8
11.0	9.8 (Forecast = 9.1)	8.8
11.6	11.5 (Forecast = 11.4)	10.9
16.7	14.5 (Forecast = 13.2)	11.6
15.3	16.7 (Forecast = 17.4)	16.6
21.6	19.9 (Forecast = 18.9)	15.3
22.4	22.8 (Forecast = 23.1)	21.5

# Сравнение результатов прогнозирования двойным и одинарным экспоненциальными методами

<b>Period</b>	<b>Single</b>	<b>Double</b>
---------------	---------------	---------------

<b>11</b>	22.4	25.8
-----------	------	------

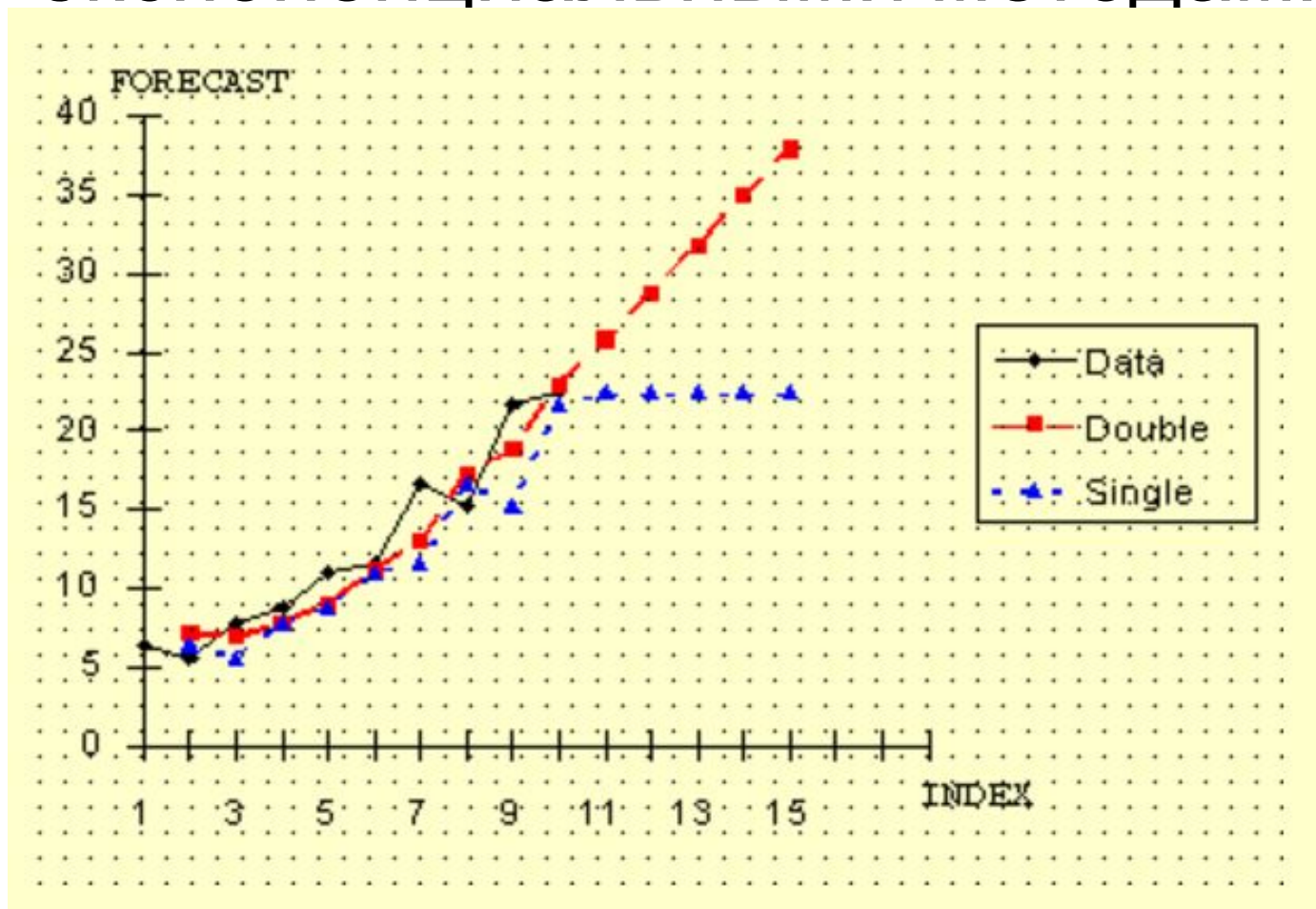
<b>12</b>	22.4	28.7
-----------	------	------

<b>13</b>	22.4	31.7
-----------	------	------

<b>14</b>	22.4	34.6
-----------	------	------

<b>15</b>	22.4	37.6
-----------	------	------

# Сравнение результатов прогнозирования двойным и одинарным экспоненциальными методами



# Тройное экспоненциальное сглаживание и прогнозирование

Что происходит, если исходный временной ряд содержит не только тренды, но и периодичность (например, сезонные колебания)? Двойное экспоненциальное предсказание не работает! В этом случае имеет смысл обратить внимание на тройное экспоненциальное сглаживание и предсказание (с учетом и трендов, и периодичности). Эта технология носит название Хольта-Винтера или HW(Holt-Winters).

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1 - \alpha)(S_{t-1} + b_{t-1}) \quad \text{OVERALL SMOOTHING}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1} \quad \text{TREND SMOOTHING}$$

$$I_t = \beta \frac{y_t}{S_t} + (1 - \beta)I_{t-L} \quad \text{SEASONAL SMOOTHING}$$

$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \quad \text{FORECAST}$$



# Периодичность

- $L$  – длина периода (должна быть определена заранее). Исходные данные должны содержать как минимум – два периода.

# Параметры

- Все параметры  $(\alpha, \beta, \gamma)$  – это значения в интервале  $(0,1)$ . Подбор параметров можно осуществлять все тем же методом **Левенберга – Марквардта** либо перебором по сетке куба  $[0-1,0-1,0-1]$ .

# Начальное значение для трендового компонента

$$b = \frac{1}{L} \left( \frac{y_{L+1} - y_1}{L} + \frac{y_{L+2} - y_2}{L} + \dots + \frac{y_{L+L} - y_L}{L} \right)$$

# Начальные значения для индексов сезонности

- Самая существенная деталь – среднее значение всех сезонных индексов должно быть равно 1. Этот параметр отражает влияние наблюдений внутри периода. Простейший способ расчета начальных значений для сезонных индексов:

$$I_i = L \frac{y_i}{\sum_{1}^L y_i}, \quad 1 \leq i \leq L$$

# Начальные значения для индексов сезонности

Самая существенная деталь – среднее значение всех сезонных индексов должно быть равно 1. Этот параметр отражает влияние наблюдений внутри периода. Простейший способ расчета начальных значений для сезонных индексов:

$$I_i = L \frac{y_i}{\sum_1^L y_i}, \quad 1 \leq i \leq L$$

# Пример. Возможный вариант расчета индексов сезонности

Технологию расчета индексов сезонности (индексов периодичности) поясним на конкретном примере. Пусть имеются наблюдения за 6 периодов (соответствуют 6 годам). Каждый период состоит из четырех наблюдений (соответствует данным о продажах за квартал). Тогда следующая последовательность шагов продемонстрирует технологию расчета индексов сезонности.

Шаг 1.

Вычислим средние продажи для каждого года:

$$A_p = \frac{\sum_{i=1}^4 y_i}{4}, \quad p = 1, 2, \dots, 6.$$

Шаг 2.

Определим индексы влияния для каждого наблюдения:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>
$y_1/A_1$	$y_5/A_2$	$y_9/A_3$	$y_{13}/A_4$	$y_{17}/A_5$	$y_{21}/A_6$
$y_2/A_1$	$y_6/A_2$	$y_{10}/A_3$	$y_{14}/A_4$	$y_{18}/A_5$	$y_{22}/A_6$
$y_3/A_1$	$y_7/A_2$	$y_{11}/A_3$	$y_{15}/A_4$	$y_{19}/A_5$	$y_{23}/A_6$
$y_4/A_1$	$y_8/A_2$	$y_{12}/A_3$	$y_{16}/A_4$	$y_{20}/A_5$	$y_{24}/A_6$

Шаг 3.

Определим начальные значения для индексов сезонности:

$$I_1 = (y_1/A_1 + y_5/A_2 + y_9/A_3 + y_{13}/A_4 + y_{17}/A_5 + y_{21}/A_6) / 6$$

$$I_2 = (y_2/A_1 + y_6/A_2 + y_{10}/A_3 + y_{14}/A_4 + y_{18}/A_5 + y_{22}/A_6) / 6$$

$$I_3 = (y_3/A_1 + y_6/A_2 + y_{11}/A_3 + y_{15}/A_4 + y_{19}/A_5 + y_{23}/A_6) / 6$$

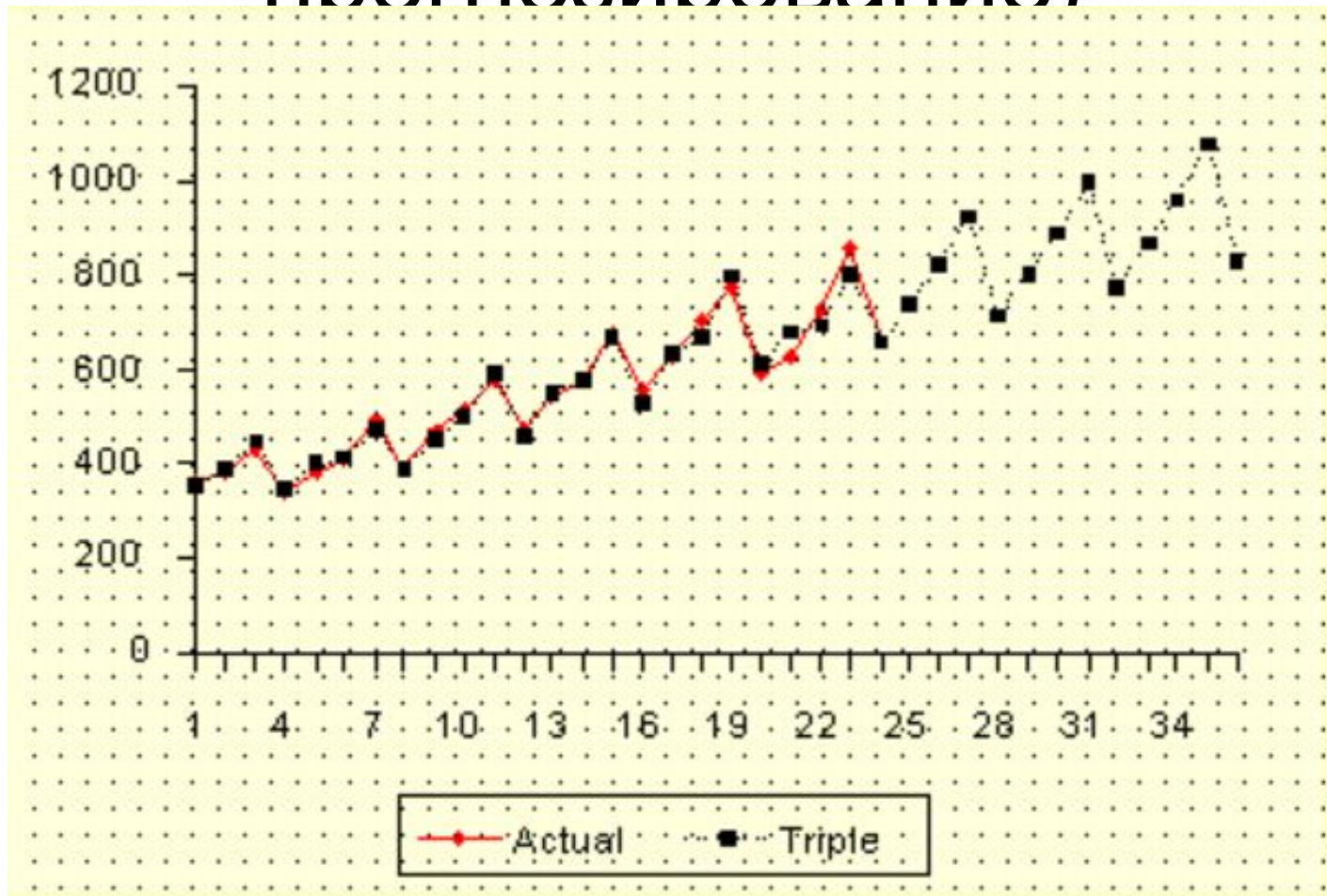
$$I_4 = (y_4/A_1 + y_6/A_2 + y_{12}/A_3 + y_{16}/A_4 + y_{20}/A_5 + y_{24}/A_6) / 6$$



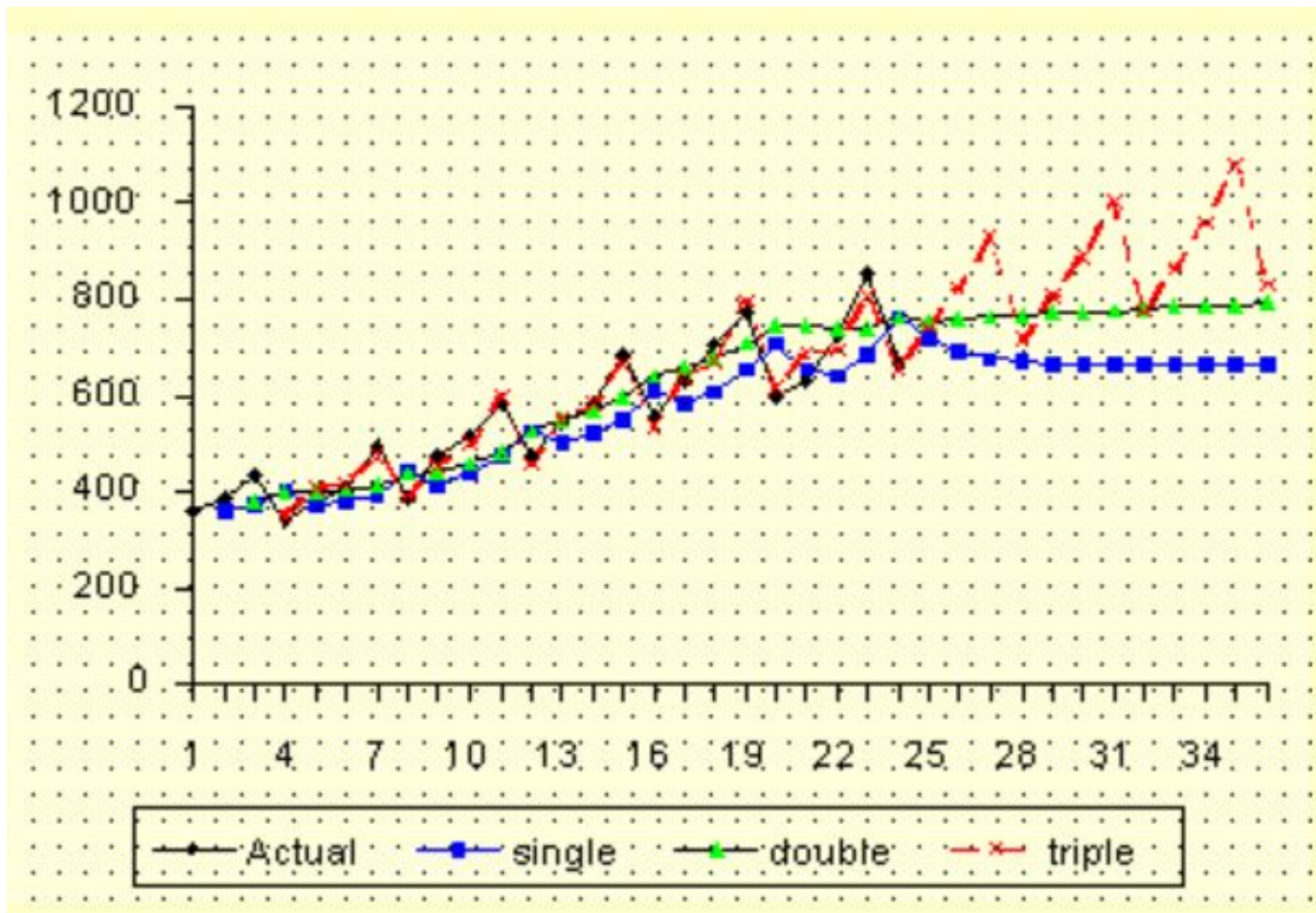
# Пример (исходные данные)

Quarter Period Sales				Quarter Period Sales			
<b>90</b>	1	1	362	<b>93</b>	1	13	544
	2	2	385		2	14	582
	3	3	432		3	15	681
	4	4	341		4	16	557
<b>91</b>	1	5	382	<b>94</b>	1	17	628
	2	6	409		2	18	707
	3	7	498		3	19	773
	4	8	387		4	20	592
<b>92</b>	1	9	473	<b>95</b>	1	21	627
	2	10	513		2	22	725
	3	11	582		3	23	854
	4	12	474		4	24	661

# Пример (тройное экспоненциальное прогнозирование)



# Пример (три вида экспоненциального прогнозирования)



- Методы экспоненциального сглаживания и прогнозирования доказали на протяжении многих лет, что являются очень полезным во многих приложениях, связанных с прогнозированием. Метод впервые был предложен С.Холта в 1957 году и использовался для не сезонных и не трендовых рядов. Позднее (в 1958) С.Холт предложил модификацию с учетом трендов. А затем Winter (1965) обобщил идею с учетом сезонности. Так и появилось название метода Хольта-Винтера...

# А что же на эту тему есть в аналитических пакетах СУБД?

- Как ни странно, в аналитических пакетах на сегодняшний день экспоненциальные методы отсутствуют. Они есть только в статистических или data mining – библиотеках.

# Задание 2

Создать приложение, позволяющее:

- Делать ежедневные прогнозы потребления творога на основе метода Holt-Winter в разрезе всей сети ресторанов;
  - Подобрать оптимальные параметры для метода;
  - для выбранного интервала определять точность прогноза;
  - Выводить результаты прогнозирования в виде графика и таблицы.
- 
- Ссылку на приложение, логин и пароль для входа отправлять по адресу: [N.Grafeeva@spbu.ru](mailto:N.Grafeeva@spbu.ru)
  - Тема - Data\_Mining\_2016\_job2