|  |  |
| --- | --- |
| lu135925on3bu_tmp_3360867a00ce4d37 | **Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования** **«Московский государственный технический университет имени Н.Э. Баумана** **(национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ Информатика и системы управления и искусственный интеллект

КАФЕДРА                  Системы обработки информации и управления

**Лабораторная работа №1**

**По курсу**

**«Анализ временных рядов»**

**«Модели трендов нестационарных временных рядов»**

Подготовил:

Студент группы

**ИУ5-34Б Журавлев Н.В**

14.09.2024

Проверил:

**Лабунец Л.В.**

*2024 г*.

**1 Цели работы**

1.1 Ознакомление и приобретение навыков работы с пакетом статистического анализа данных «Statistica» версии 6.0;

1.2 Изучение моделей и методов формирования трендов НВР.

**2 Задачи работы**

2.1 Приобретение навыков работы с электронной таблицей и рабочей книгой пакета «Statistica»;

2.2 Изучение параметрических моделей трендов НВР;

2.3 Изучение непараметрических и комбинированных моделей трендов НВР.

**3 Теоретическая часть**

Одним из наиболее популярных непараметрических методов сглаживания ВР является их цифровая фильтрация. Процедуру цифровой фильтрации данных удобно представить с помощью уравнения дискретной свертки

, (1)

где  – текущий отсчет ряда;  - импульсная характеристика цифрового фильтра (ЦФ), определяющая метод сглаживания;  - начальная модель тренда. Здесь и в дальнейшем полагают, что импульсные характеристики фильтров удовлетворяют условию нормировки .

В зависимости от типа ЦФ его импульсная характеристика может быть конечной или бесконечной. Например, в случае простой скользящей средней (simple moving average - SMA) импульсная характеристика фильтра имеет вид  для всех  и ноль в противном случае. Здесь *K* - конечный временн*о*й интервал усреднения данных. Выбранные пределы суммирования по индексу истории *k* реализуют схему прогнозирования «вперед». В этой модели оценку тренда  формируют на основе значений ряда , …,  в текущий *n*- й и предыдущие моменты времени. Это стандартная выборочная оценка математического ожидания, принятая в классической статистике в рамках гипотезы стационарности и эргодичности данных на текущем временн*о*м интервале [*n* – *K* + 1, *n*].

Опыт цифровой обработки данных показывает, что однократного сглаживания, как правило, не достаточно для полного выделения тренда из НВР. Иными словами, начальный остаток  может содержать трендовую компоненту

.

Эффективное элиминирование тренда из данных требует применения рекуррентной процедуры сглаживания остаточных ВР

, , *i* = 1, 2, …

(2)

Типичные критерии завершения итераций имеют вид

, *i* = 1, 2, …

или

, *i* = 2, 3, …,

где *N* – объем выборки,  и  - выбранные пользователем достаточно малые, положительные уровни значимости критериев. Финальные модели остаточного ряда и тренда в этом случае рассчитывают по формулам

, . (3)

Эффективной альтернативой цифровой фильтрации данных являются параметрические модели трендов НВР, основанные на полиномиальной, экспоненциальной или логарифмической аппроксимации ВР. В классе интеллектуальных моделей трендов следует отметить методы сглаживания ВР с помощью локально взвешенной полиномиальной регрессии. Для рекуррентного статистического оценивания НВР в соответствии с выражением (1.3) удобными являются модели:

* наименьших квадратов, взвешенных «расстоянием» (Distance Weighted Least Squares - DWLS) [11];
* локальной взвешенной робастной регрессии (Lowess) [12].

**4 Практическая часть**

Для выполнения задач лабораторной работы загрузим файл Series\_G.sta из учебной базы данных в каталоге Examples в папке программы.

Заголовок электронной таблицы содержит:

1. Количество столбцов (variables). Каждый столбец таблицы – уникальный временной ряд
2. Количество строк (cases). Каждая строка – отсчет времени.

Свойства электронной таблицы:

* Подзаголовок. Редактируется текстовым редактором, аналогичным Word. Для редактирования нужно дважды кликнуть по заголовку.

В меню имеются стандартные команды *Add, Move, Copy, Delete*… Помимо этого есть раздел *Statistics*, содержащий широкий спектр методов статистического анализа, моделирования и прогнозирования данных

Раздел *Advanced Linear/Nonlinear Models* позволяет проводить исследования ВР.

Подраздел *Time Series* в разделе *Advanced Linear/Nonlinear Models* позволяет графически отобразить временной ряд в удобном для пользователя виде.

Кнопка *Variables* предназначена для выбора необходимого ВР. Наличие записи в окне означает, что для ускорения работы системы медленный формат данных электронной таблицы преобразован в быстрый двоичный формат данных. Данный формат позволяет использовать методы изменения атрибутов и форматов объектов.

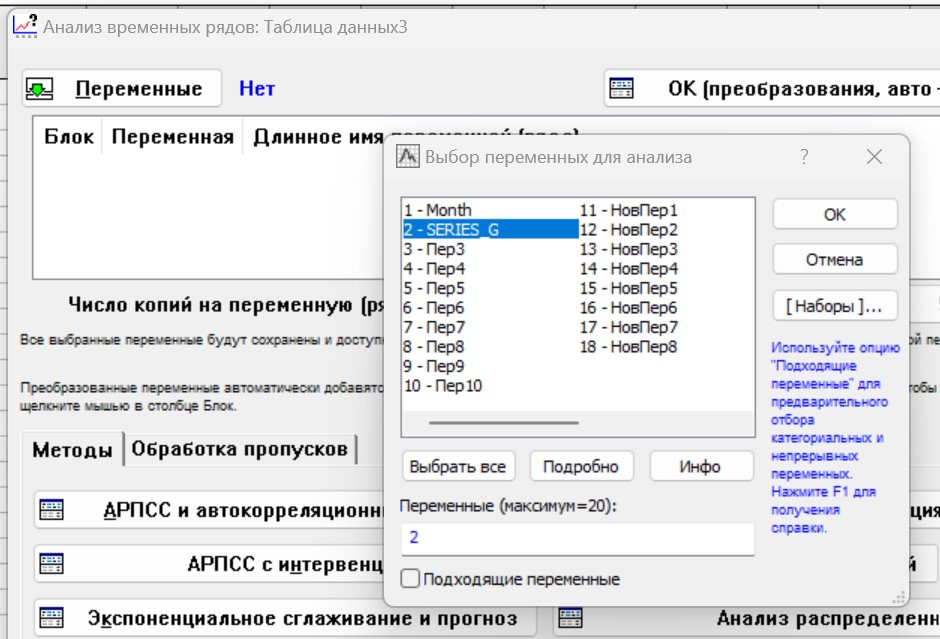


Рисунок 1 – Выбор ряда для анализа

Для серьезного анализа количество записей должно быть не менее 10. В нижней части таблицы есть 2 закладки: вкладка *Quick* обеспечивает доступ к методам моделирования, вкладка *Missing Data* позвляет восстанавливать потерянные данные. Кнопка *OK* вызывает очередное рабочее окно, в нижней части которого имеются вкладки для отображения предварительной обработки.

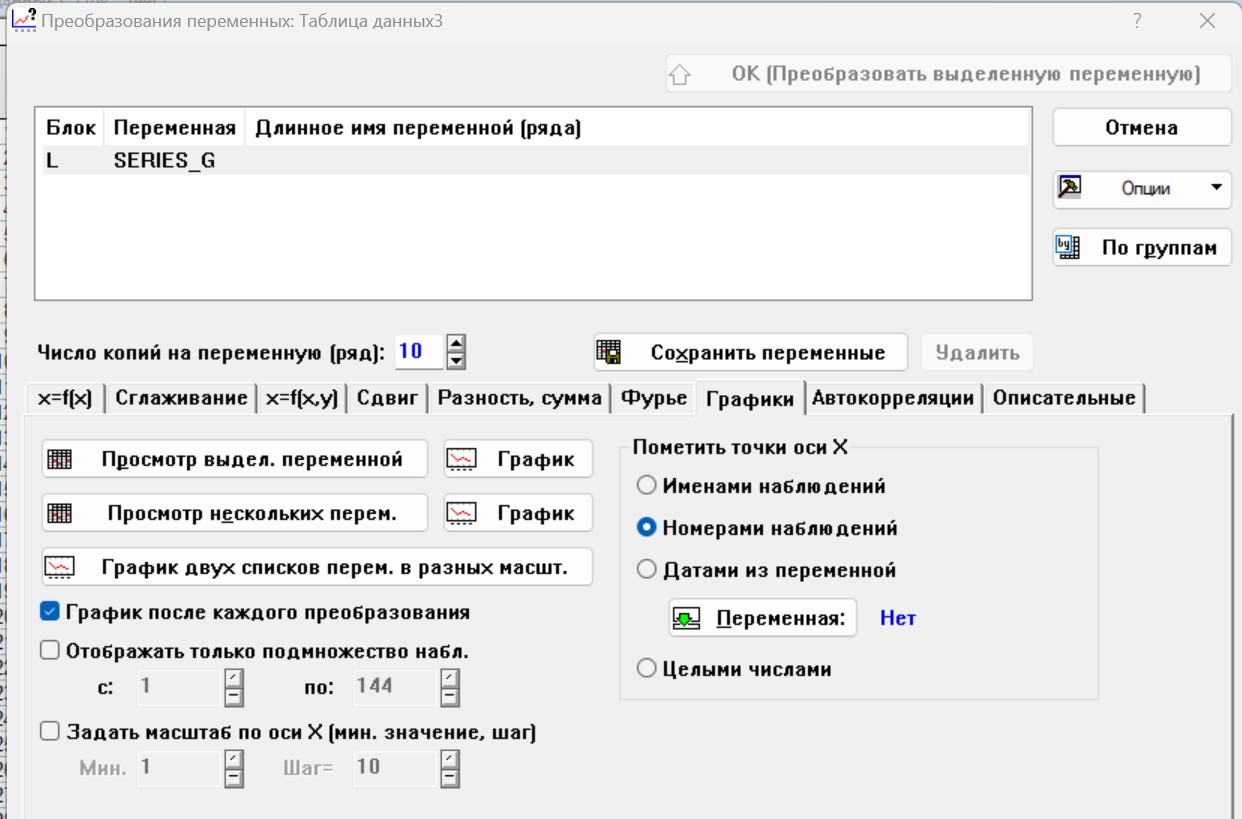


Рисунок 2 - Окно предварительной обработки данных

Вкладка *Preview Plots* предлагает 3 способа отображения графиков:

1. Одна линия на одном графике (верхняя кнопка);
2. Много линий на одном графике в одном масштабе;
3. Изображение нескольких графиков в масштабах левой и правой осей.

При нажатии на *Plot* появляется новый объект *WorkBook*. Раздел состоит из 2 частей. Пиктограмму можно перетащить в другое место, взяв ее мышкой, два раза щелкнув по названию можно редактировать его.

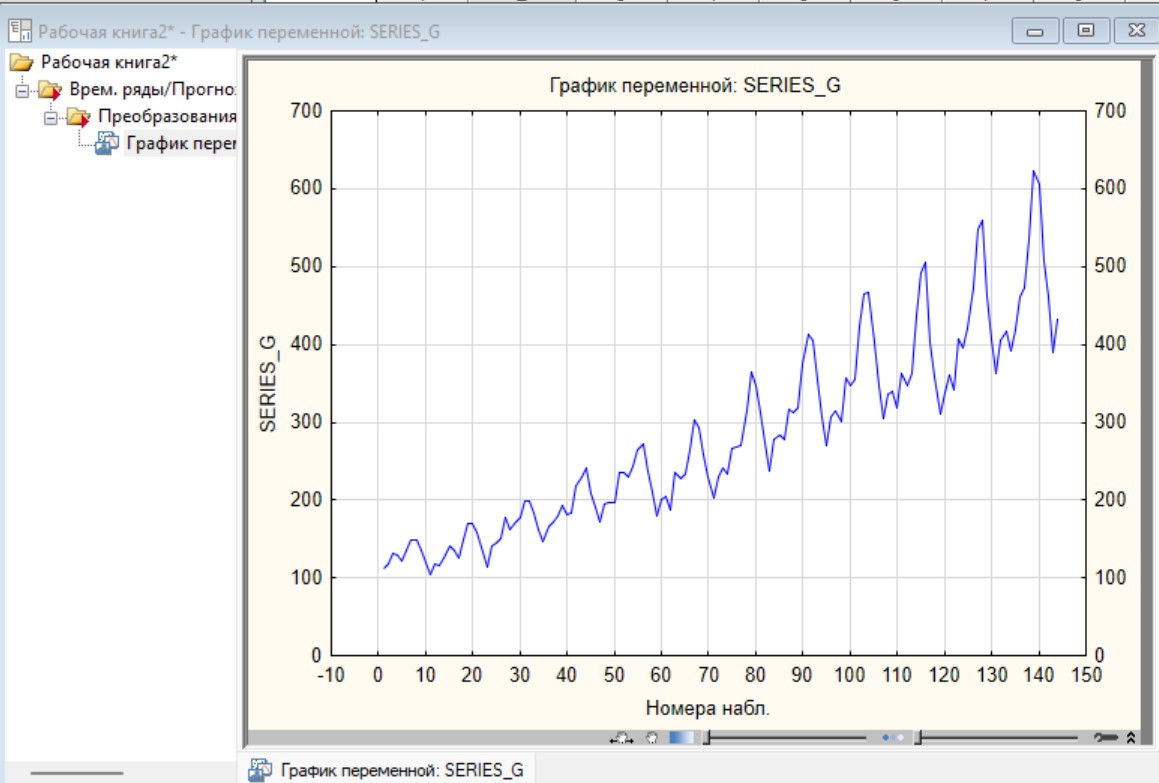


Рисунок 3. График ВР

Визуализация исходных данных и результатов в виде графиков необходима для формирования гипотез относительно механизма поведения ВР. Все элементы графика – это множество объектов, отдельно взятый элемент - объект. Например, название графика, заголовки и подзаголовки, ряды чисел, линии, сетка и т.д. Для редактирования объекта нужно его выделить и кликнуть 2 раза. Появится окно редактирования.

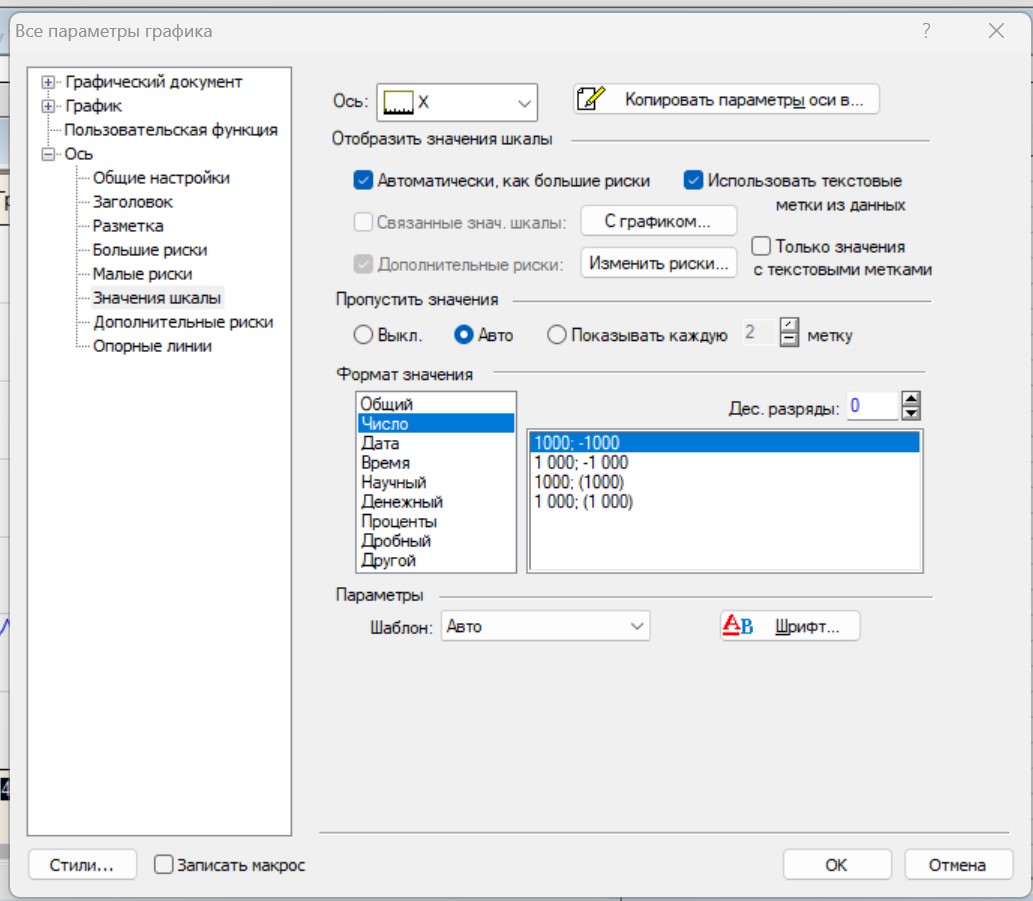


Рисунок 4 - Редактирование горизонтальной оси графика

Имеется BackUp на панели быстрого управления. Редактирование горизонтальной оси графика с шагом 12 подтверждает гипотезу наличия сезонной компоненты с периодом 12 месяцев:

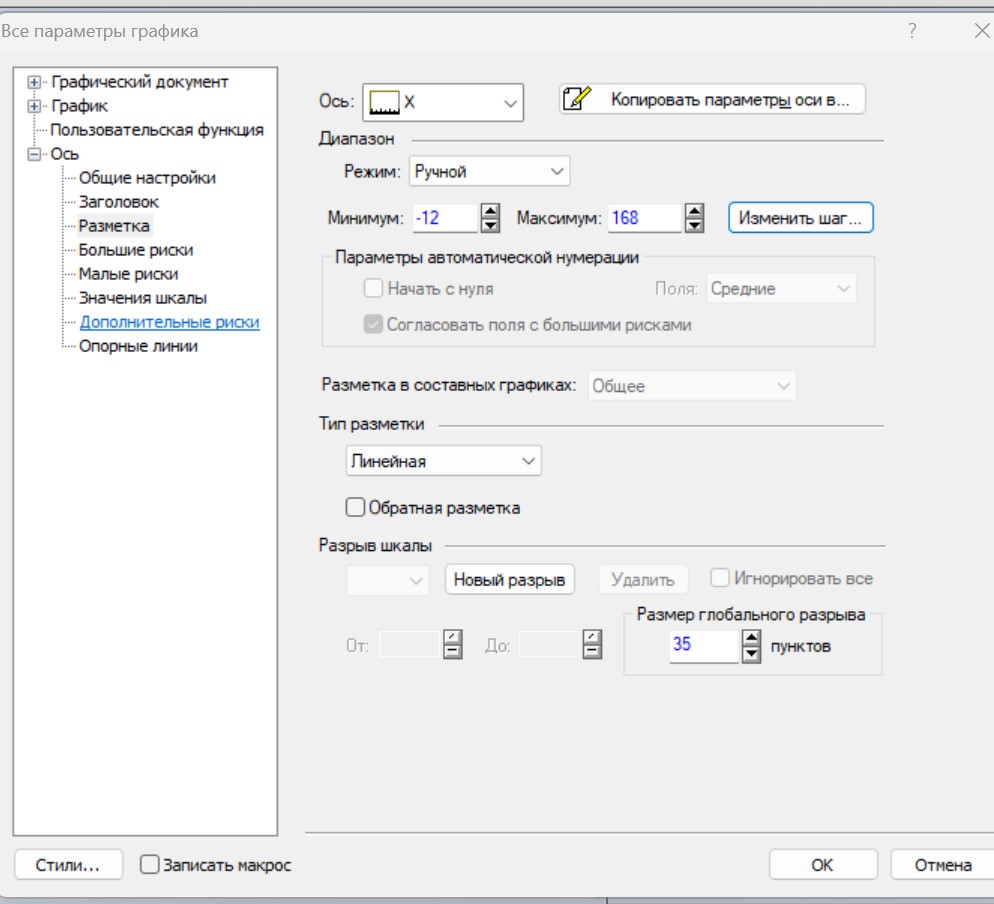


Рисунок 5 - Редактирование горизонтальной оси графика

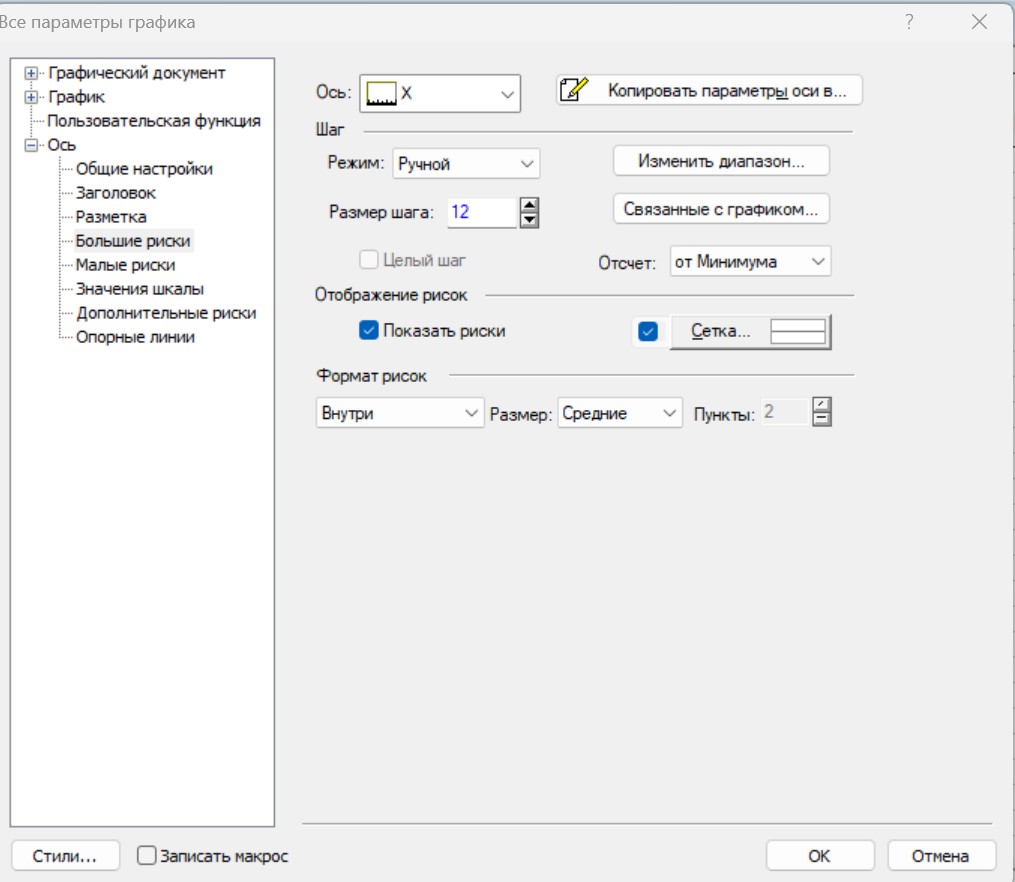


Рисунок 6 - Редактирование горизонтальной оси графика

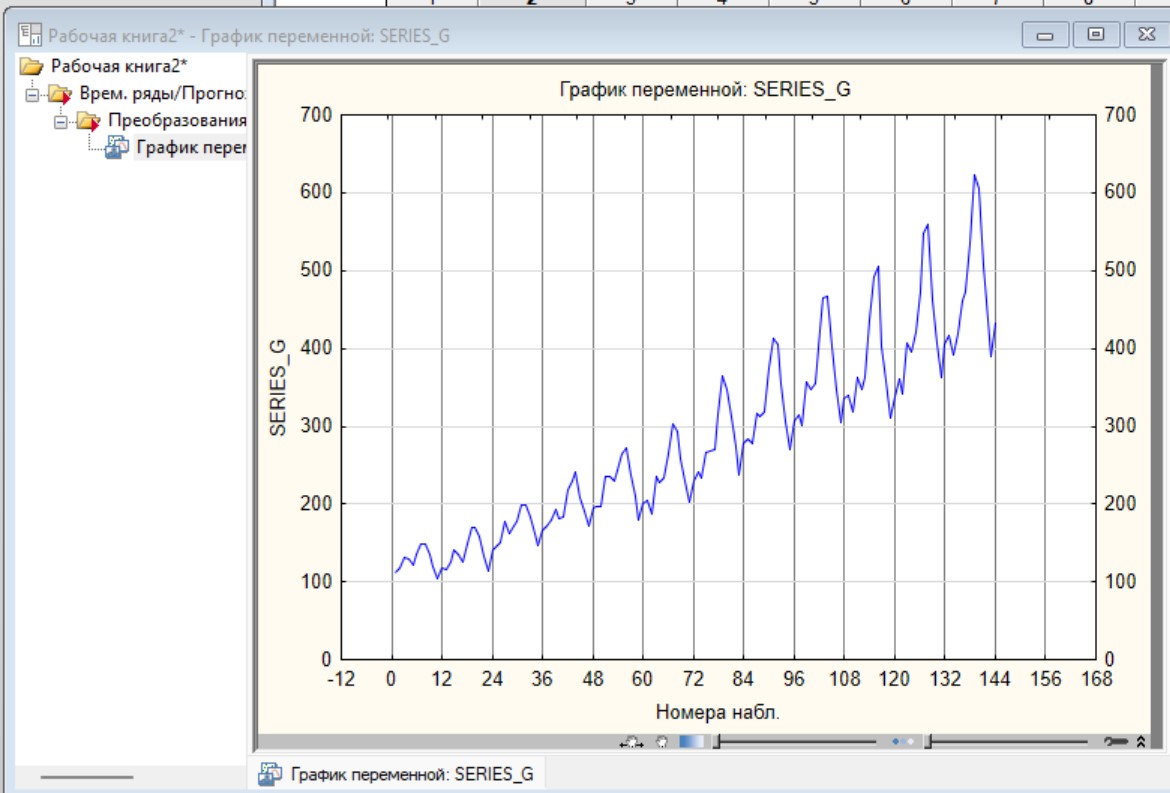


Рисунок 7 – Новая ось X

Данные в среднем растут. Построим модель тренда. Анализируем модели тренда 2 типов: непараметрические и параметрические.

Непараметрическую модель тренда формируют методами сглаживания (цифровой фильтрации). Для этого реанимируем рабочую панель. Во вкладке *Smoothing* имеются несколько моделей фильтров. 1 – SMA. N – интервал сглаживания. Для сглаживания сезонной составляющей возьмем N=12 и более. Флажок *Prior* ассоциирует SMA с правой границей интервала сглаживания. Без флажка – центрированный тренд. Ставя флажок *Weighted*, можно выбрать веса для КИХ- фильтра, рассчитанного пользователем.

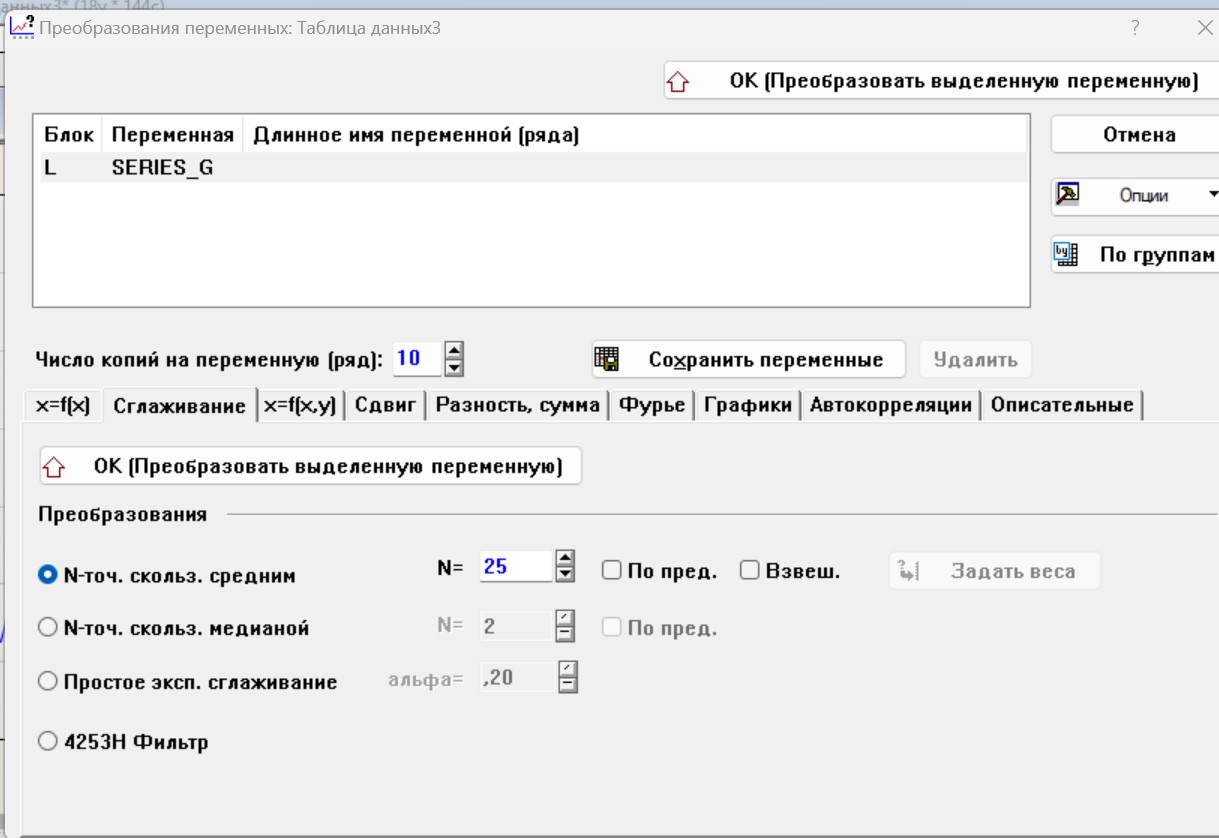


Рисунок 8 – Формирование модели тренда

Итак, для формирования непараметрической модели тренда необходимо выбрать: метод сглаживания, период сглаживания, способ ассоциации значения скользящей средней с моментом времени в пределах интервала сглаживания.

После нажатия кнопки *OK (Transform selected series)* выполняется сглаживание. В каталог рабочей книги добавляется график SMA.

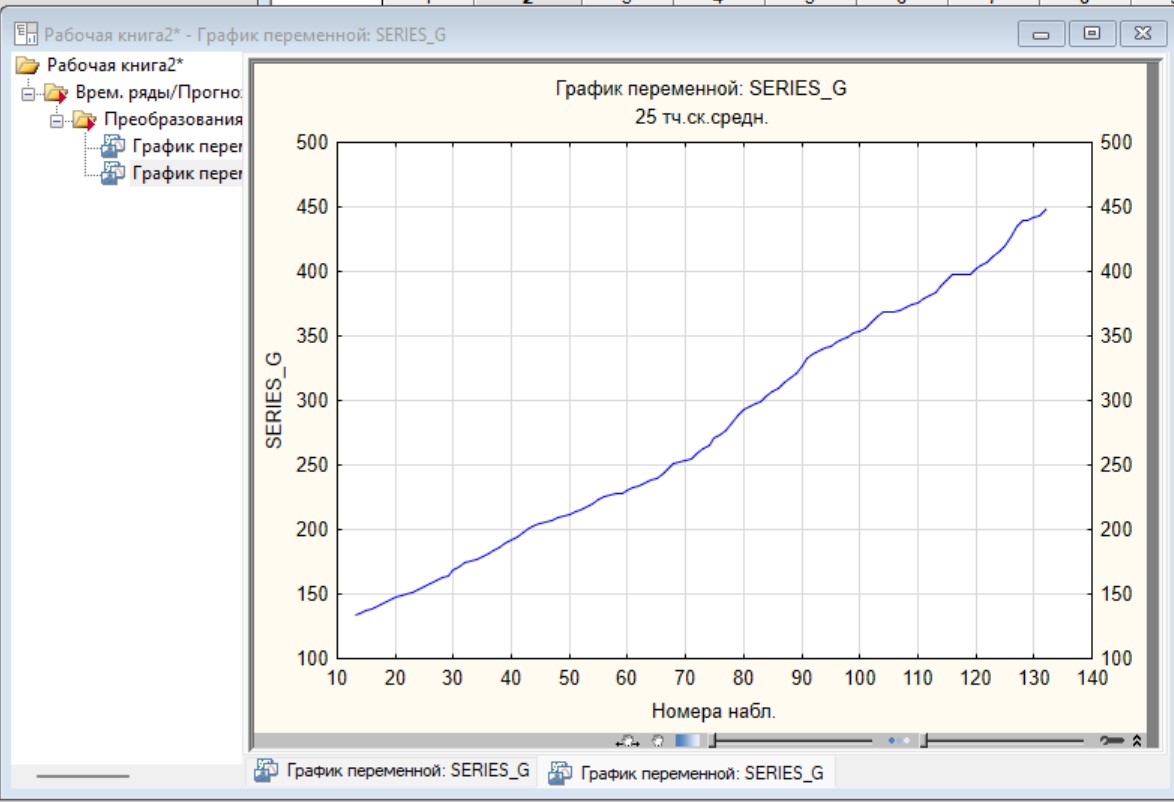


Рисунок 9 - график SMA

Для наглядности нужно поместить ВР и SMA на один график. На рисунке видно, что центрированная модель тренда обладает некоторыми недостатками, а именно, модель не охватывает левый и правый (что более важно) сегменты данных.

Простейший способ устранения этого недостатка состоит в повторном сглаживании тренда с помощью подходящей параметрической модели.

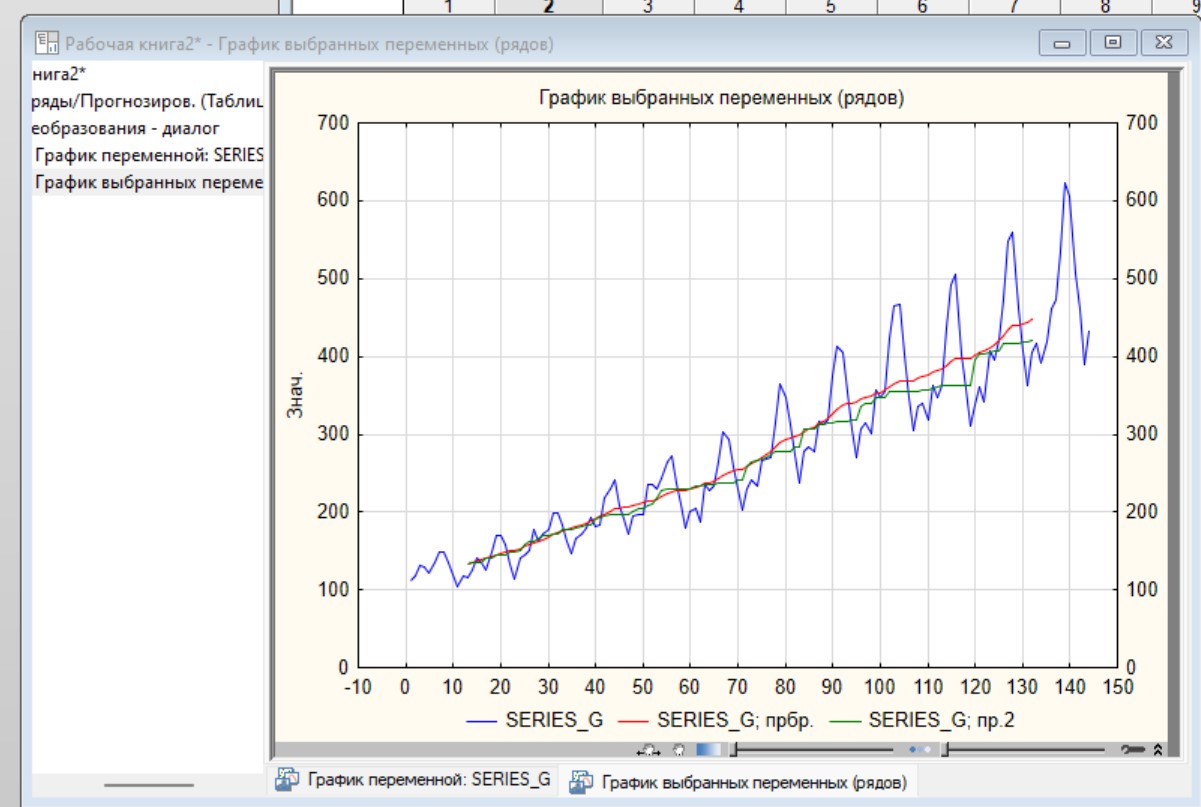


Рисунок 10 - Модель простой скользящей средней (SMA)

Доступ к библиотеке параметрических моделей тренда обеспечивает интегральная панель атрибутов рабочей книги.

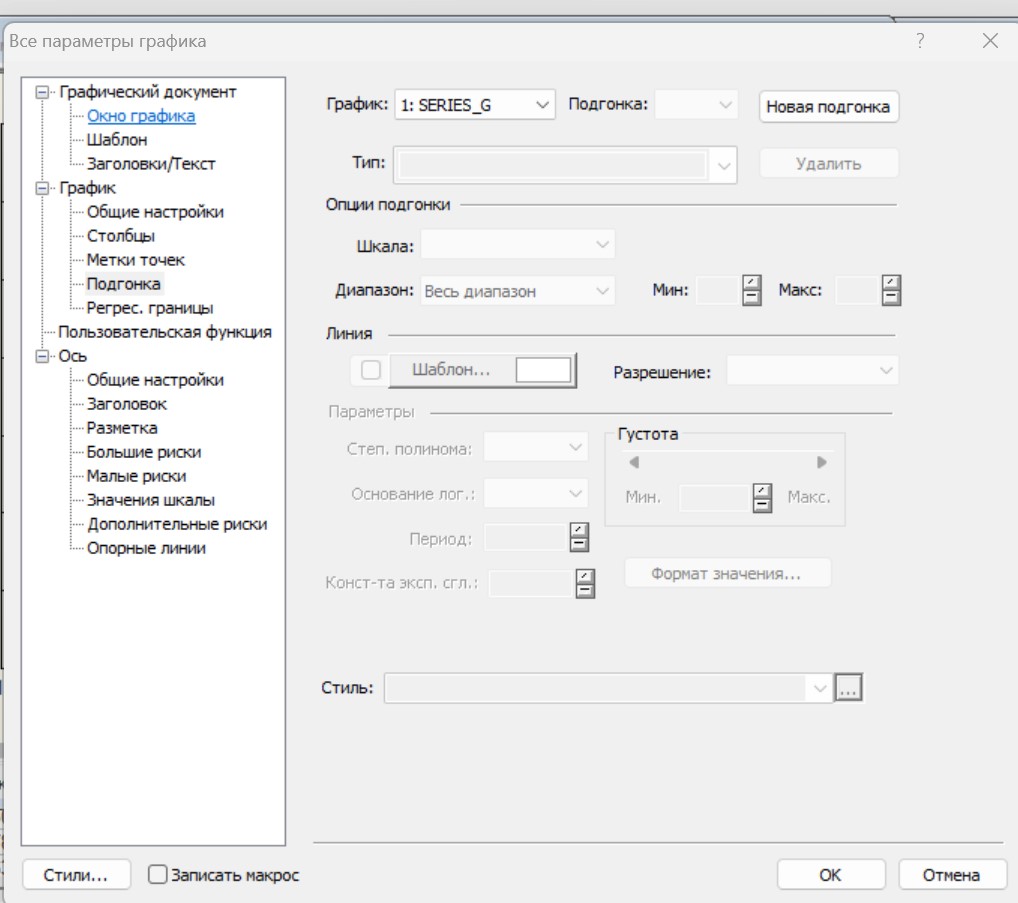


Рисунок 11 - Параметрические модели аппроксимации ВР

Для ее инициализации необходимо навести мышь на поле графика и дважды кликнуть по нему. В появившемся окне выберем закладку *Plot: Fitting*. Имеются модели: линейные, полиноминальные и т. д. (Рисунок 13). Для создания новой модели необходимо нажать *Add new fit* вверху рабочего окна. Выберем экспоненциальную модель. Нажмем ОК. На графике ВР и его SMA появляется экспоненциальная модель сглаживания.

Экспоненциальная модель тренда хорошо аппроксимирует тенденцию поведения НВР в среднем. Для более детального анализа поведения данных рационально рассматривать более сложные модели, например, модель ядерной робастной регрессии *Lowess*.

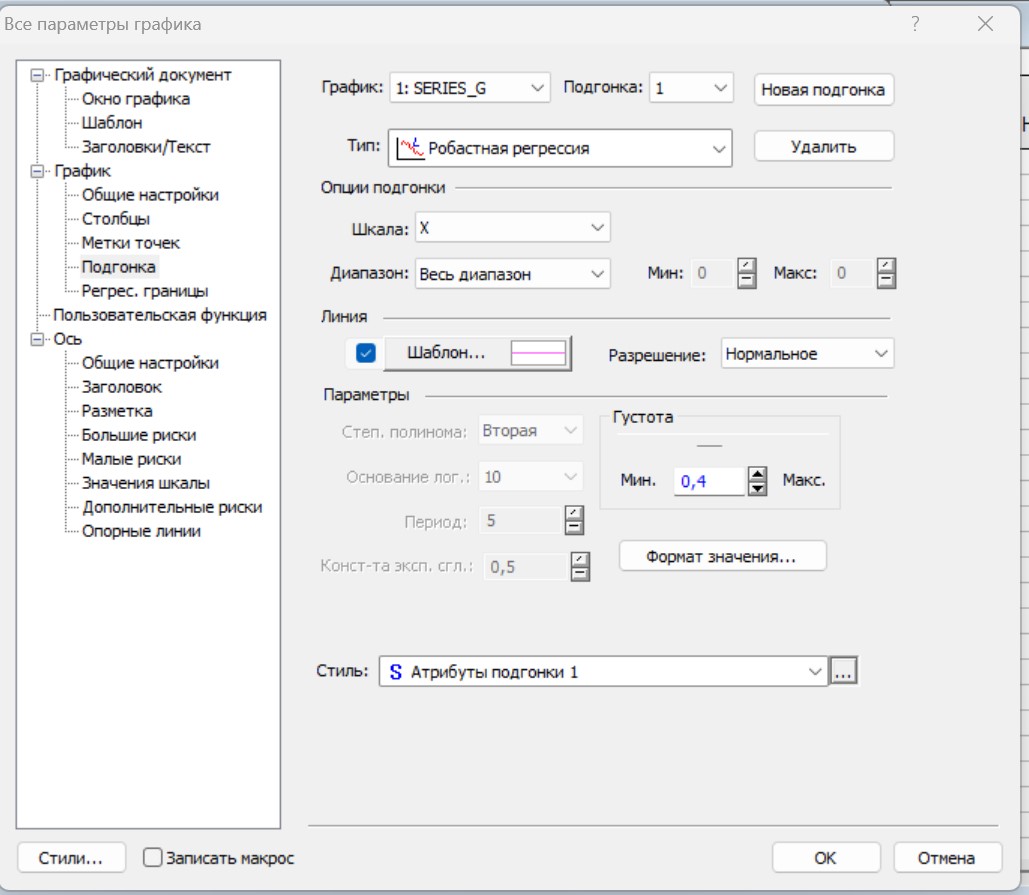


Рисунок 12- Параметры модели Lowess сглаживания НВР и его SMA

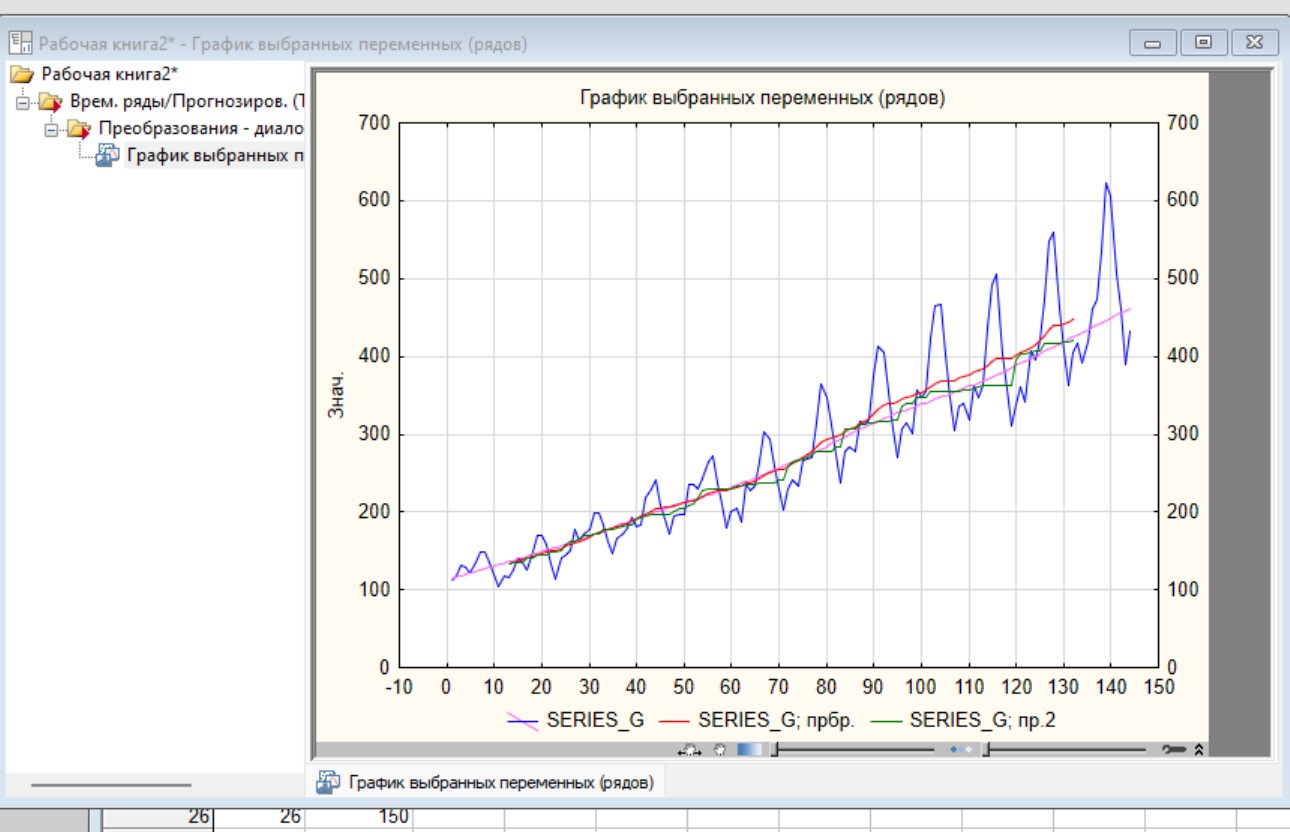


Рисунок 13 - Модель Lowess сглаживания НВР и его SMA

Мы можем добавить столбцы в нашу электронную таблицу. Для этого на панели быстрого доступа зайдем в меню Vars -> Add. Добавим 3 столбца. Первый столбец назовем Months. Для его заполнения, аналогично заполнению таблицы Excel, запишем в 2-х первых строках числа 1 и 2, выделим их и протащим до 156 строки (это больше количества отсчетов исходного ВР). Столбец заполнился числами от 1 до 156. Второй столбец заполним отсчетами SMA. Для этого дважды щелкаем по названию столбца. Редактируем название и в поле в нижней части рабочего окна записываем выражение для экспоненциальной оценки тренда.

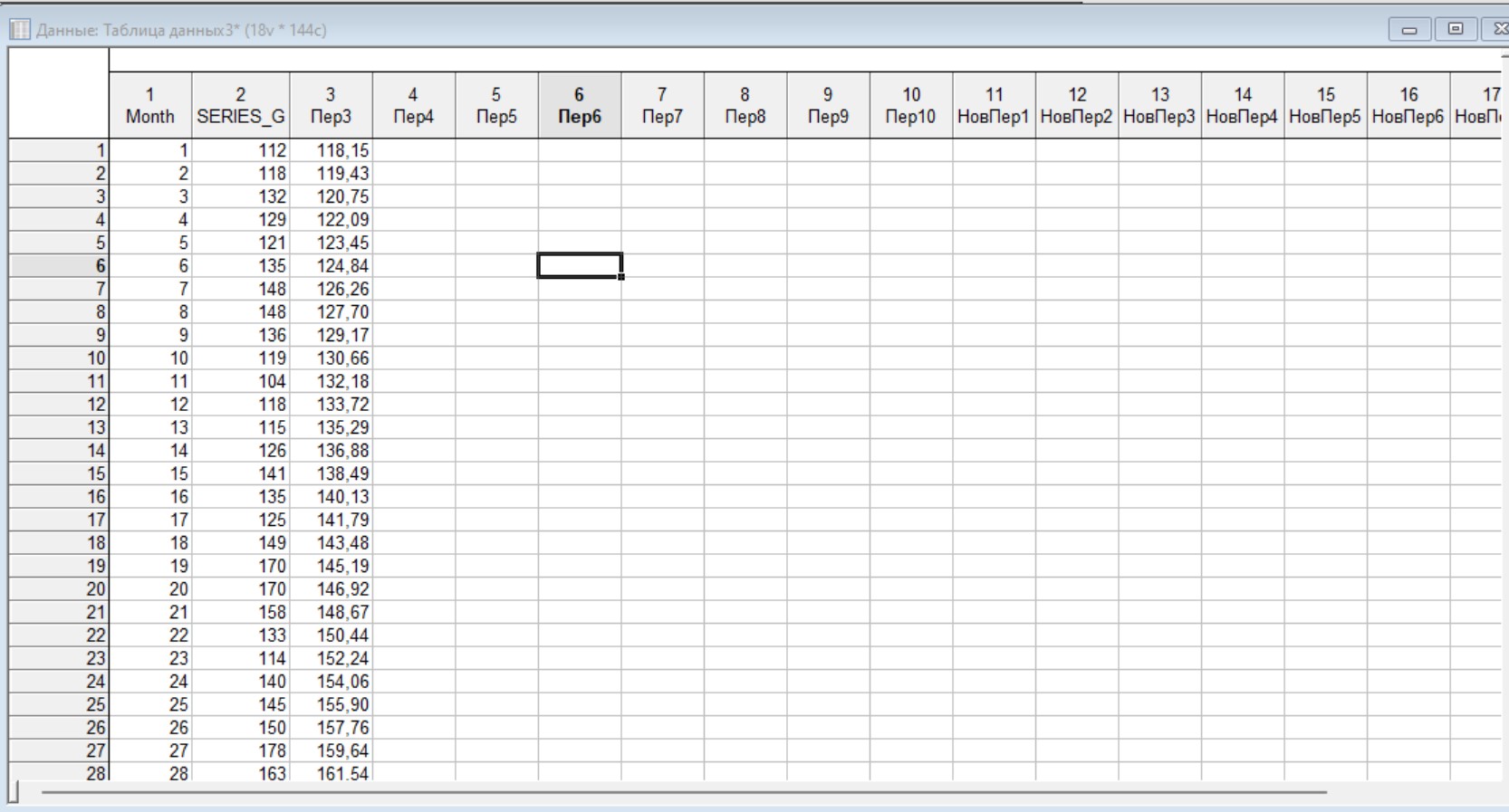


Рисунок 14 – Экспоненциальная аппроксимация SMA

После нажатия кнопки ОК столбец заполнился отсчетами данных. Затем сформирует остаточный ВР. Для этого в нижнее окно редактирования запишем формулу для вычисления разности между исходным ВР и экспоненциальной моделью тренда.

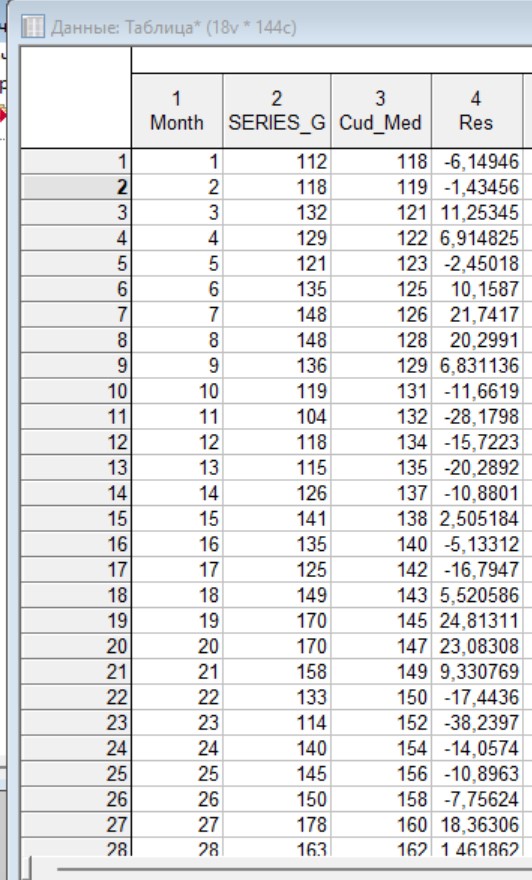


Рисунок 15 - ВР ошибки сглаживания

Для наглядности отобразим исходный ряд, модель тренда и ошибку аппроксимации НВР на одном графике в двух масштабах

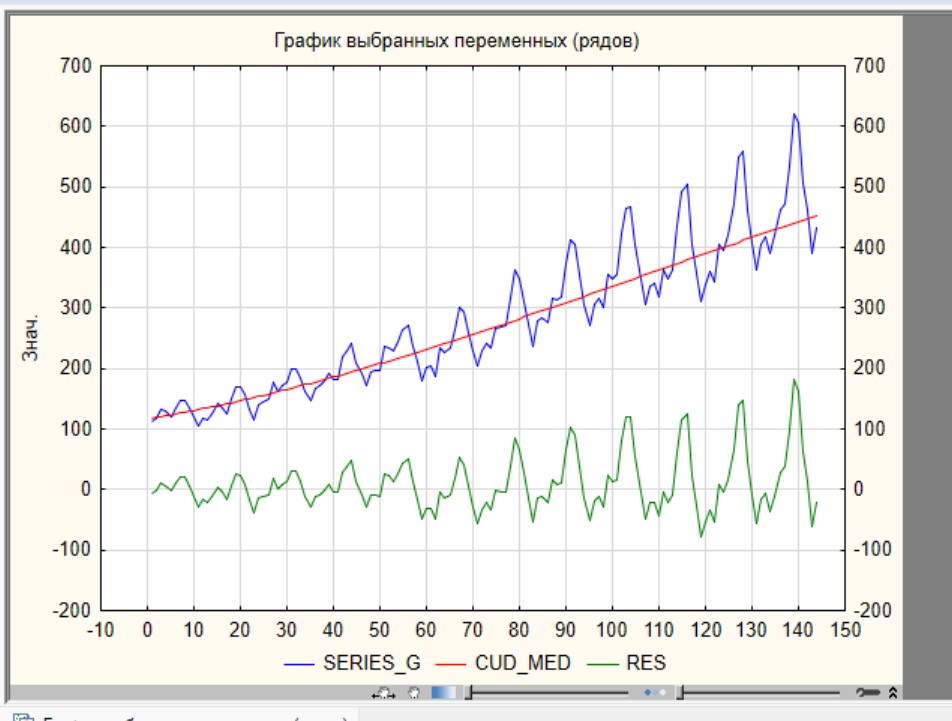


Рисунок 16 - Графики исходного ряда, экспоненциальной модели тренда

и остаточного НВР

В результате получили модель тренда и остаточный НВР как разность между исходным рядом и трендом.

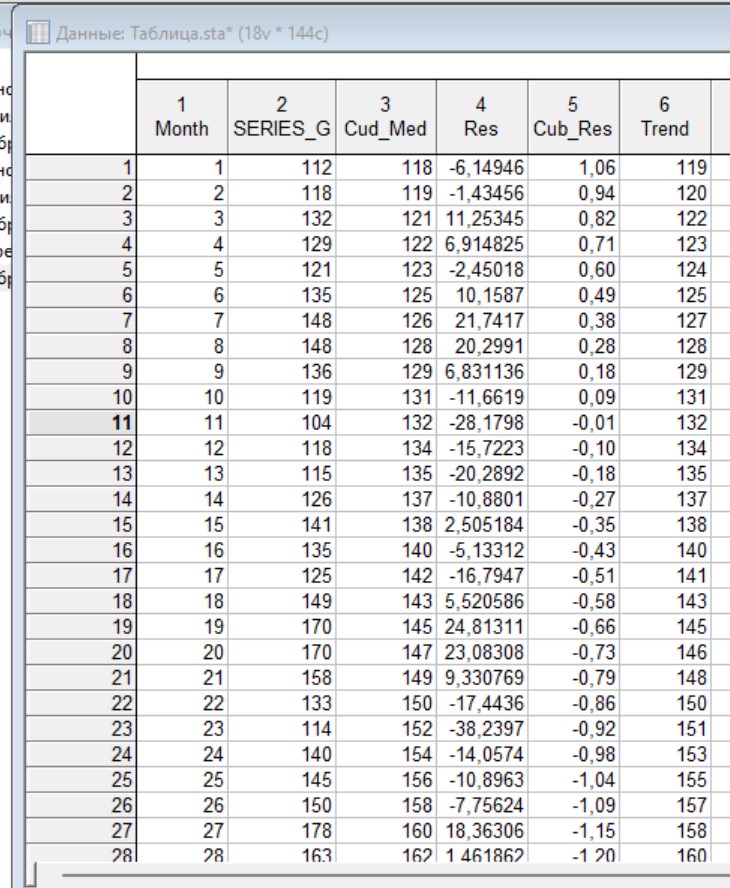


Рисунок 18 – Модель тренда

Посчитаем остаток ещё раз, для этого из исходного временного ряда вычтем модель тренда.

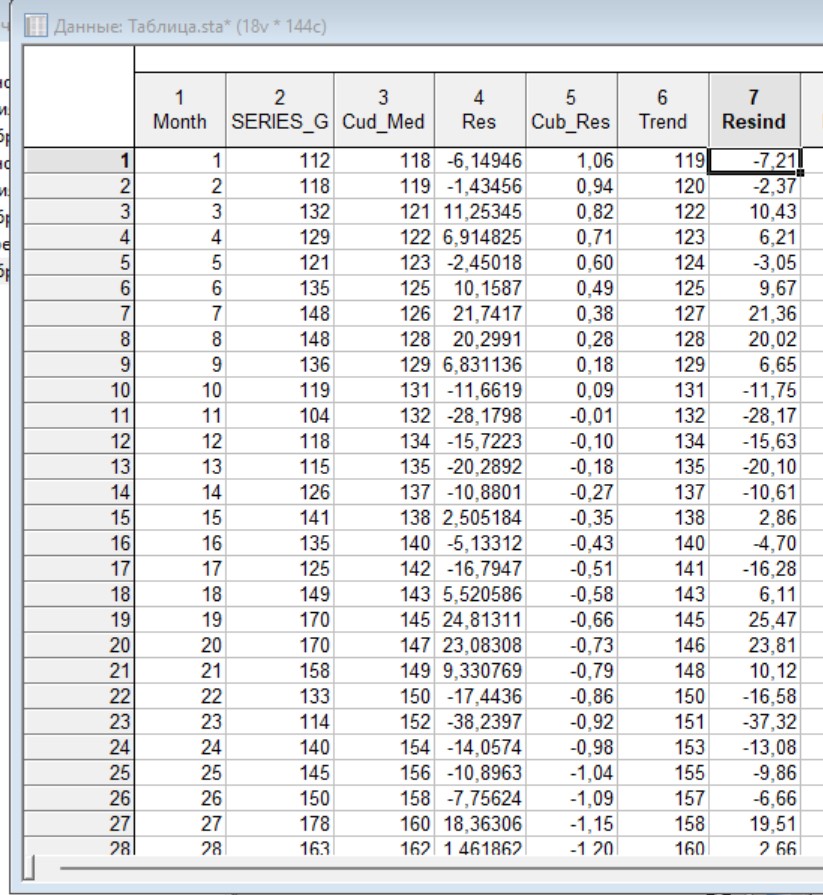


Рисунок 19 – Подсчёт остатка

**5 Выводы**

Для прогнозирования данных необходимо знать их статистические характеристики. Одной из статистических характеристик является математическое ожидание или тренд. Для формирования модели тренда по одной реализации в пакете «STATISTICA» имеются параметрические и непараметрические методы. К непараметрическим моделям относят скользящие средние. Они имеют известный недостаток, а именно, центрированная модель, наиболее точно описывающая процесс, не охватывает граничные сегменты данных. Для устранения этого недостатка применяют параметрические модели аппроксимации НВР. В пакете «STATISTICA» имеется богатый выбор параметрических моделей: линейная, экспоненциальная, полиномиальная, а также несколько более интеллектуальных моделей, например, модель Lowess.

**Список литературы**

1. Цветков Э. И. Нестационарные случайные процессы и их анализ. – М.: Энергия, 1973. – 128 с.

2. Мирский Г. Я. Аппаратурное определение характеристик случайных процессов. – М.: Энергия, 1972. – 456 с.

3. Мирский Г. Я. Характеристики стохастической взаимосвязи и их измерения. – М.: Энергоиздат, 1982. – 320 с.

4. Ольшевский В. В. Основы теории статистический измерений. – Таганрог: ТРТИ, 1976. – 107 с.

5. Бендат Дж., Пирсол А. Прикладной анализ случайных данных: Пер. с англ. – М.: Мир, 1989. – 540 с.

6. Пугачев В. С. Теория случайных функций и ее применение к задачам автоматического управления. – М.: Физматгиз, 1962. – 883 с.

7. Романенко А. Ф., Сергеев Г. А. Аппроксимативные методы анализа случайных процессов. – М.: Энергия, 1974. – 176 с.

8. Прикладной анализ случайных процессов. Под ред. Прохорова С. А. / Самара: СНЦ РАН, 2007. – 582 с.

9. Голяндина Н. Э. Метод «Гусеница» - SSA: анализ временных рядов: Учеб. пособие. – СПб.: С. – Петерб. гос. ун-т, 2004. – 76 с.

10. Голяндина Н. Э. Метод «Гусеница» - SSA: прогноз временных рядов: Учеб. пособие. – СПб.: С. – Петерб. гос. ун-т, 2004. – 76 с.

11 Лабунец Л. В., Лебедева Н.Л. . Чижов М. Ю. Рекуррентные статистики нестационарных временных рядов // Радиотехника и электроника, 2011, т. 56, № 12, с. 1468 – 1489.

12. Боровиков В. П. Популярное введение в программу STATISTICA. - 2000. – 269с.

13. Боровиков В. П. STATISTICA. Искусство анализа данных на компьютере: для профессионалов. 2-е изд. - СПб.: Питер, 2003. – 688с.

14. Тихонов Э. Е. Методы прогнозирования. – Невинномысск, 2006. - 206с.

15. Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов: Учебное пособие. – М.: Финансы и статистика, 2003. – 416 с.