|  |  |
| --- | --- |
| 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 №7**

**По курсу**

**«Анализ временных рядов»**

**«Модели корреляции нестационарных временных рядов»**

Подготовил:

Студент группы

**ИУ5-34Б Журавлев Н.В**

14.12.2024

Проверил:

**Лабунец Л.В.**

*2024 г*.

1. **Цели работы**
   1. Изучение моделей и методов формирования корреляционной матрицы НВР.
   2. Приобретение навыков выборочного оценивания коэффициента корреляции НВР в пакете «STATISTICA».
2. **Задачи работы**
   1. Формирование диаграммы рассеяния и временного ряда корреляционного произведения для различных значений корреляционного лага в пакете «STATISTICA».
   2. Формирование непараметрических моделей коэффициента корреляции НВР в пакете «STATISTICA».
   3. Формирование параметрических моделей корреляционной матрицы НВР в пакете «STATISTICA».
   4. Приобретение навыков визуализации 3D- графиков в пакете «STATISTICA».
3. **Теоретическая часть**

В рамках понятия автокорреляции, исследуют наличие вероятностных зависимостей текущих значений НВР от его прошлых значений. Если объектами анализа являются два временных ряда, один из которых рассматривают как фактор влияния, а второй – как цель прогнозирования, то говорят о взаимной корреляции, или кросс - корреляции [1 - 5].

Полезным инструментом качественного анализа наличия корреляции является двумерная диаграмма рассеяния (ДР). Для НВР выбирают два временных сечения. Интервал *m* между этими сечениями принято называть корреляционным лагом или окном. Значения ряда, измеренные в эти моменты времени, рассматривают, как декартовы координаты точки на плоскости. Значение процесса в предыдущий момент времени (фактор) удобно ассоциировать с координатой по оси абсцисс, а в текущий момент (прогноз) – с координатой по оси ординат. Перемещая корреляционное окно фиксированной ширины *m* по оси времени на один шаг получают множество точек на плоскости.

Форма сформированного облака точек характеризует наличие или отсутствие корреляционной зависимости в данных. Необходимо отметить, что для НВР важным параметром формирования ДР является не только корреляционный лаг *m*, но и количество шагов времени (количество точек) *N*, в соответствии с которым анализируют форму ДР. Если ДР демонстрирует наличие явно выраженного генерального направления, то этот факт свидетельствует о наличии значимой корреляции. В противном случае приходят к выводу об отсутствии корреляции.

При исследовании корреляции НВР важно также помнить следующее правило. Наибольшее значение корреляционного лага *m*, при котором дисперсия выборочной оценки коэффициента корреляции приемлемо мала, не превышает одной десятой от объема выборки, т.е.  [12 - 14].

Цифровую модель автокорреляции нормированного остатка  НВР представим как результат рекуррентной процедуры сглаживания, например, с помощью ЦФ [11]

,

,

,

, *i* = 1, 2, …

Объектом сглаживания, в этом случае, являются временные ряды  в виде корреляционных произведений. Здесь  - начальная оценка коэффициента корреляции НВР на фиксированном лаге *m* = 1, 2, …,  в дискретный момент времени *m* + ; *N* – объем выборки;  - импульсные характеристики фильтров, определяющие автоковариацию исходного ряда на лаге *m*.

Альтернативные начальные оценки коэффициента корреляции  НВР в рекуррентной процедуре сглаживания основаны на применении, так называемых, методов суммирования или вычитания [2]

, ,

, .

Критерии завершения итераций имеют вид

, *i* = 1, 2, …

или

, *i* = 2, 3, … ,

где  и  - выбранные пользователем достаточно малые, положительные уровни значимости критериев. Финальные оценки коэффициента корреляции НВР рассчитывают по аддитивным формулам



или

.

Ясно, что в этом случае процедуры сглаживания с помощью ЦФ или локально взвешенной полиномиальной регрессии должны обеспечивать выполнение ограничения . Это требование хорошо согласуется с моделями робастного взвешивания [11].

**4 Практическая часть**

Построим график нормированного временного ряда.

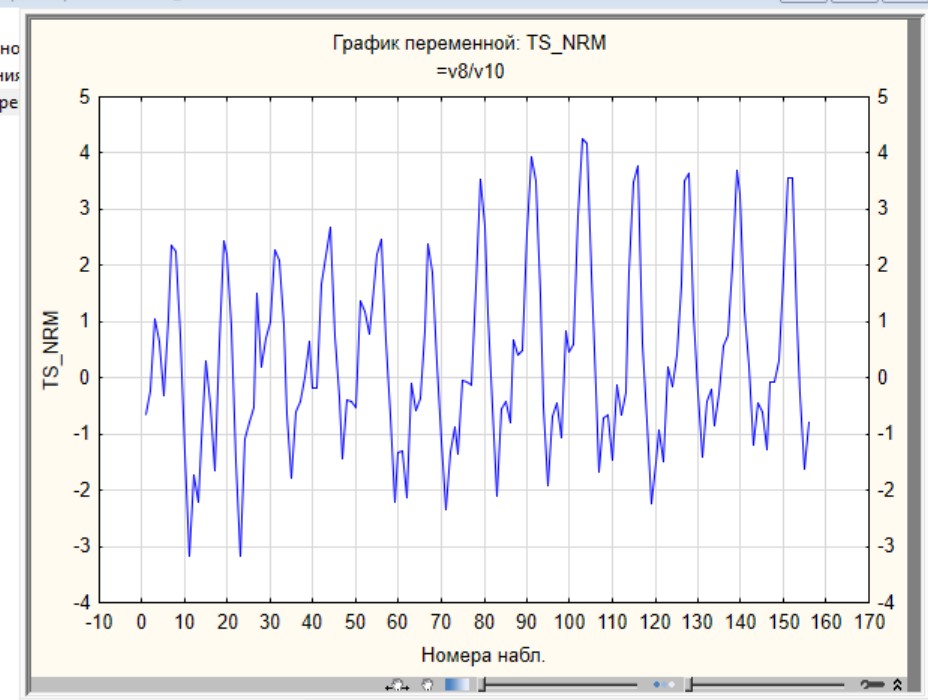


Рисунок – График нормированного временного ряда

Построим модель тренда. Во вкладке подгонка выберем модель Lowess на всём диапазоне с размером окна аппроксимации 0,3 и размером сетки среднее.



Рисунок – График с моделью Lowess

Необходимо сделать оцифровку. Для этого создадим табличку, в которую перенесём все месяца и продублируем их в соседний столбик, так же дополним их ещё 400 значениями. Затем вернувшись к графику нажмём правую кнопку мыши и выберем кнопку оцифровка графика, затем кнопку оцифровка подгонки. Полученные значения копируем в дополнительные 400 значений, вместе со значениями. Затем выполняем сортировку по второму столбцу и первое значение в третьем столбце заполняем его соседним значением снизу. Далее в анализе выбираем вкладку пропущенные значения. Затем нажимаем кнопку сохранить значения и вставляем их в 3 столбец, после чего сортируем по 1 столбцу.

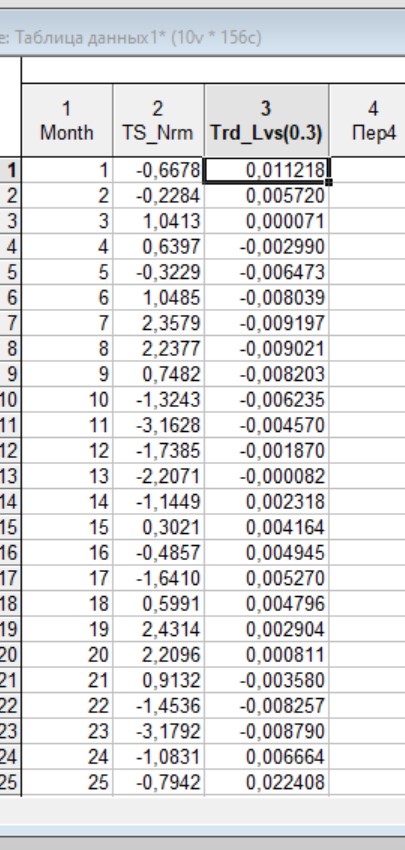


Рисунок – Оцифровка значений

Получившуюся оцифровку (первые 156 значений) заносим в основную таблицу. И считаем остаток, написав формулу v2-v3.

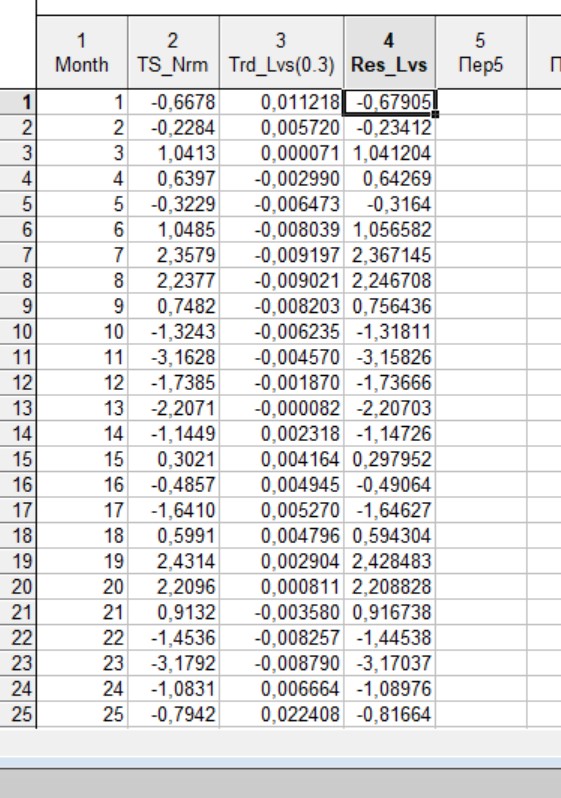


Рисунок – Значение остатка

Заходим в меню базовые статистики и таблицы, выбираем описательные статистики. Выбираем остаток и ставим галочку на полях медиана, минимум максимум и процентили со значениями 25 и 75.

В заголовок таблицы запишем значение медианы, правее неё среднее. Далее вычисляем центр размаха (среднее от минимум и максимума) и заносим его правее среднее. Аналогичным образом для интерквантильного диапазона и заносим результат правее медианы.

Сортируем остаток и находим крайние значения интерквантильного диапазона и копируем в соседний столбик его значения. Вычисляем выборочное среднее для интерквантильного диапазона аналогично предыдущему пункту. Затем получившееся значение помещаем правее медианы. Удаляем интерквантильный диапазон и возражаем сортировку по 1 столбику. Возведём 4 столбец в квадрат. Заходим в анализ, затем в панели 2 уровня выбираем вкладку x=f(x) и выбираем элемент возведение в степень и пишем значение 2.

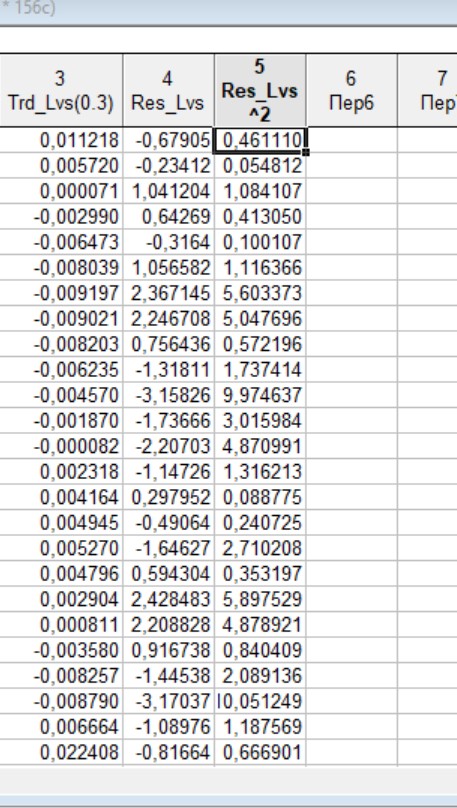


Рисунок – Квадрат остатка

Построим аналогично первому разу модель lowess, но уже для квадрата остатка.

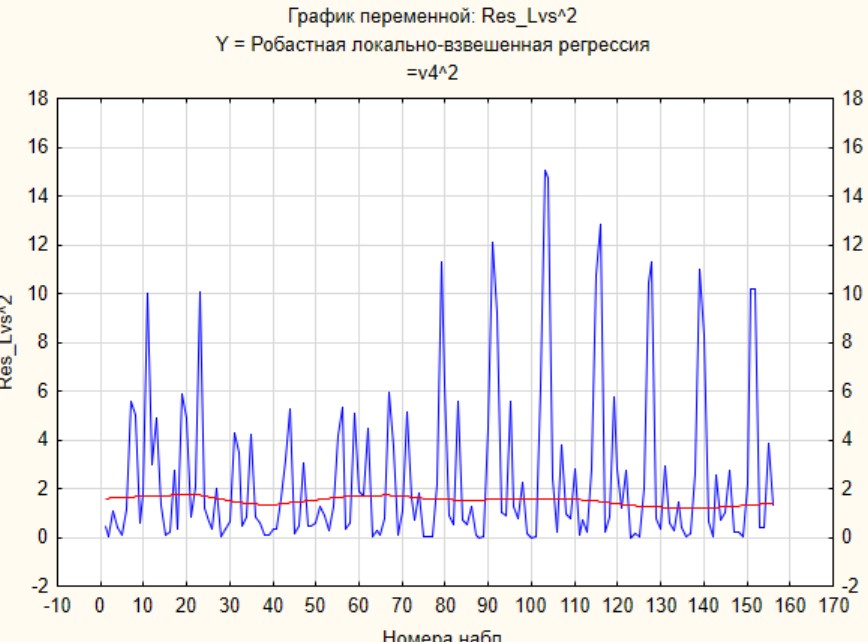


Рисунок – Модель Lowess для квадрата остатка

Оцифровываем данные аналогично предыдущему разу и заносим в 5 столбец.

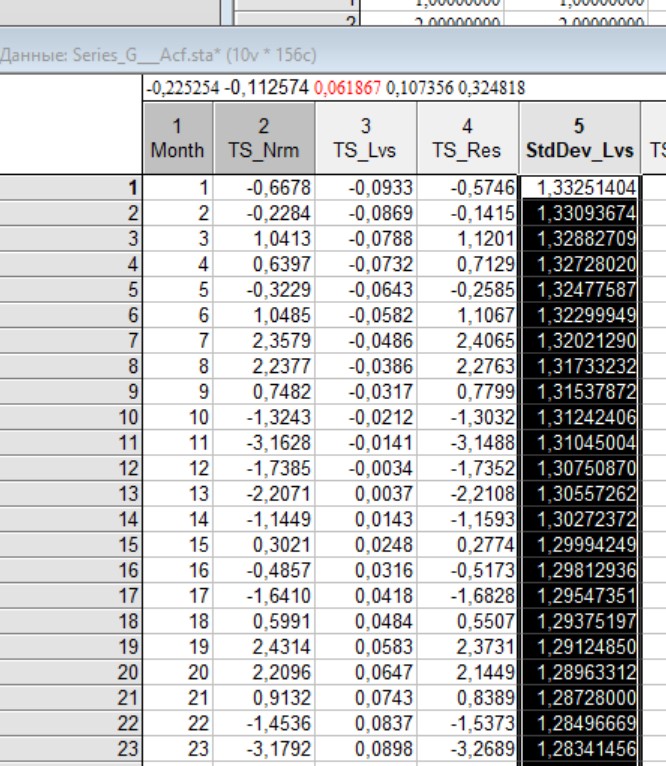


Рисунок – Оцифровка модели Lowess для квадрата остатка

Масштабируем остаток введя формулу v4/v5 в соседний столбик.

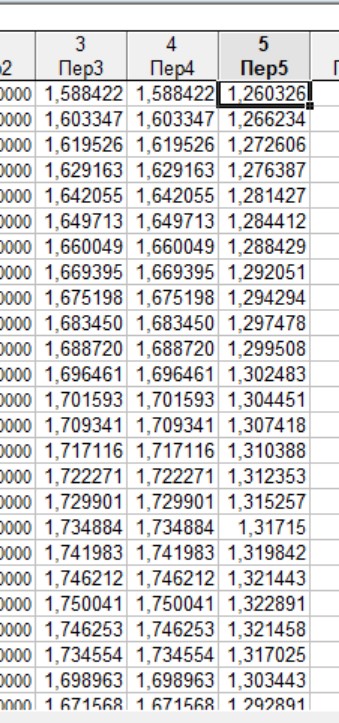


Рисунок – Масштабирование остатка

Проверим на сколько он является нормированным для этого запустим инструмент описательные статистики. Поделим ряд на значение среднего отклонения.

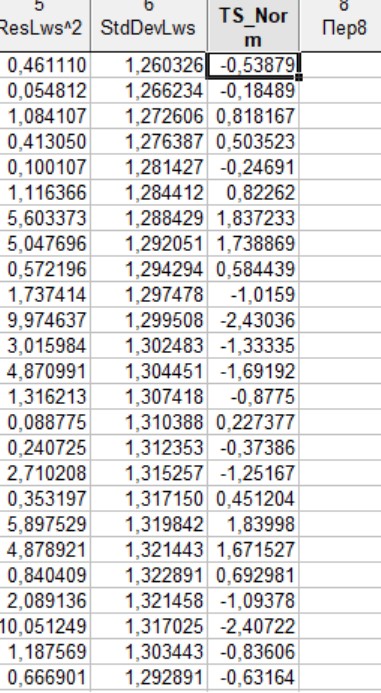


Рисунок – Новые значения масштабирования остатка

Проверяем аналогичным предыдущему разу значения.

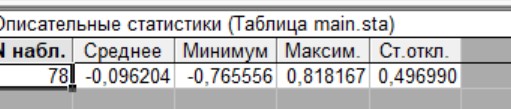


Рисунок – Результат инструмента описательные статистики

Сдвинем значения вперёд на 1. Для этого в анализе выберем вкладку shift и в поле укажем значение 1.

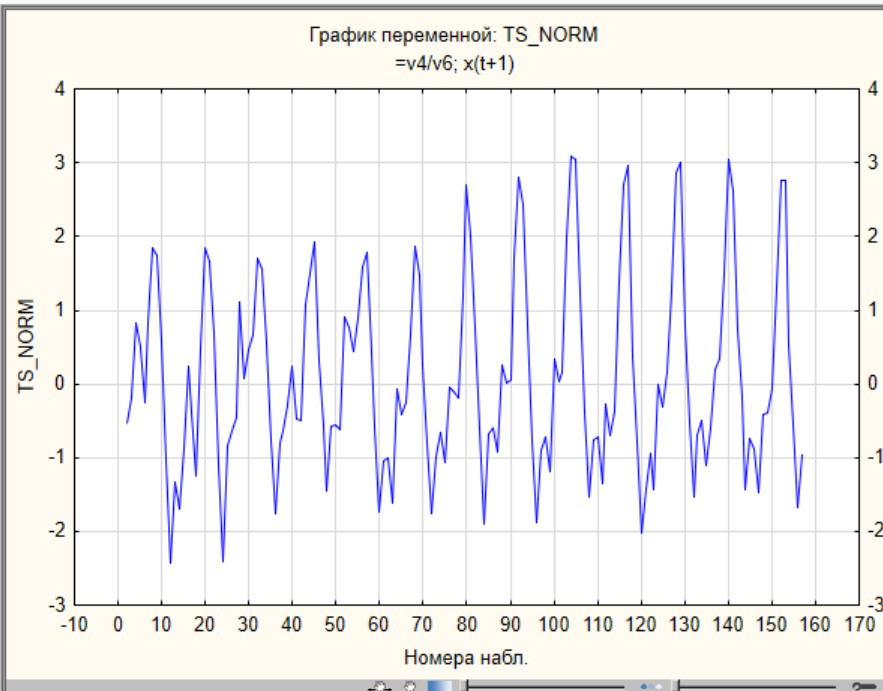


Рисунок – Сдвиг вперёд с лагом 1

Сохраним оцифровку со сдвигом. Получившиеся значения скопируем в 7 столбик. Затем в 8 столбике запишем временной ряд корреляционного произведения, для этого напишем формулу v6\*v7.

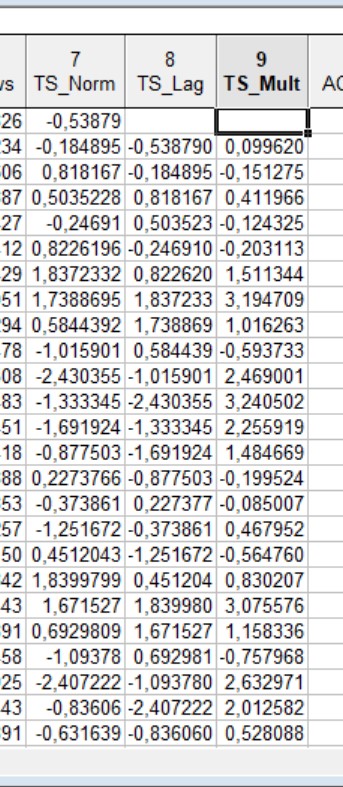


Рисунок – Временной ряд корреляционного произведения

Далее строим модель Lowess для получившегося графика.

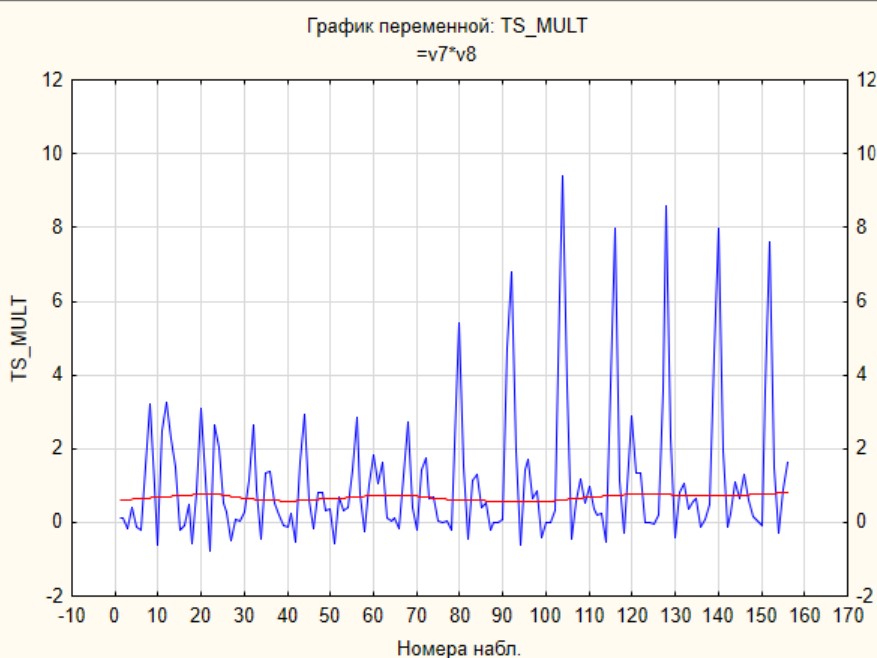


Рисунок – Модель Lowess для временного ряда корреляционного произведения

Производим аналогичную ранее сделанной оцифровку модели Lowess.

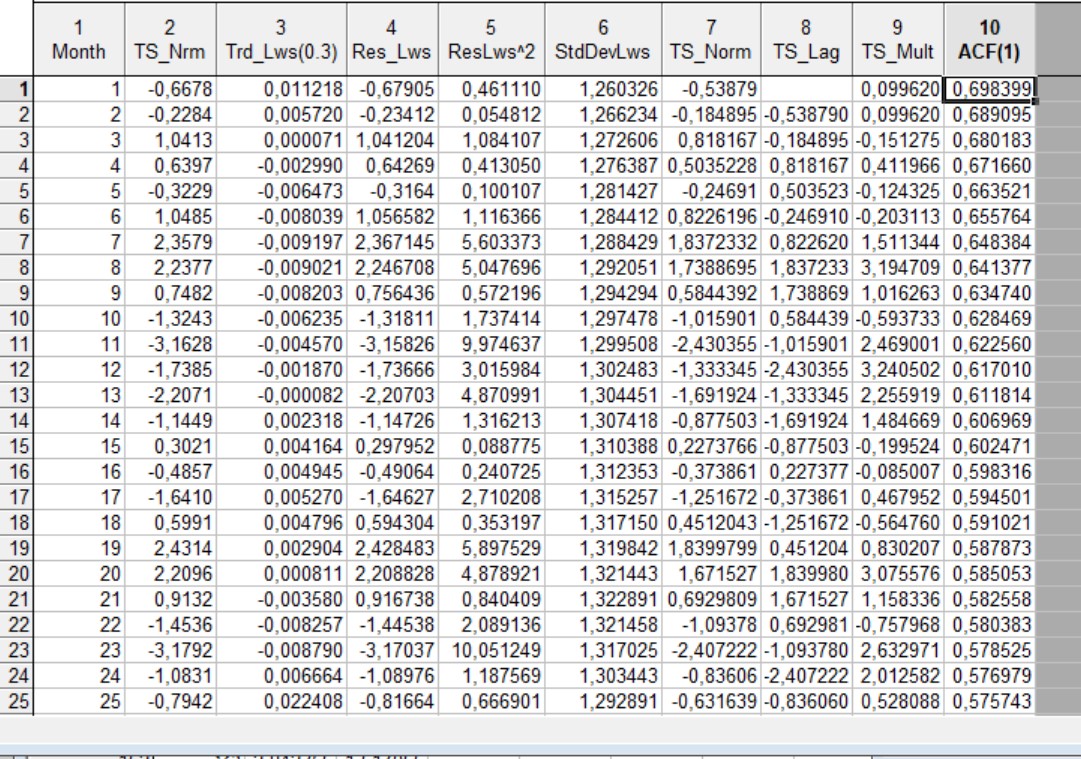


Рисунок – Оцифровка Lowess

Сравним с робастной оценки в виде медианного фильтра. На вкладки сглаживания поставим выбор медианный фильтр с окном 25. И строим на одном графике.

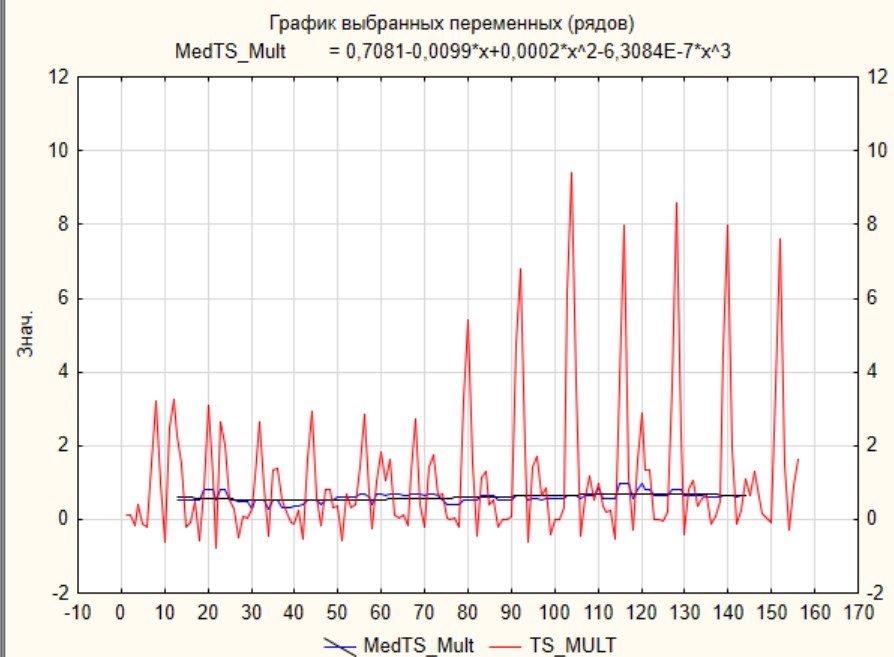


Рисунок – Сравнение с медианным фильтром

Оцифруем медиану.

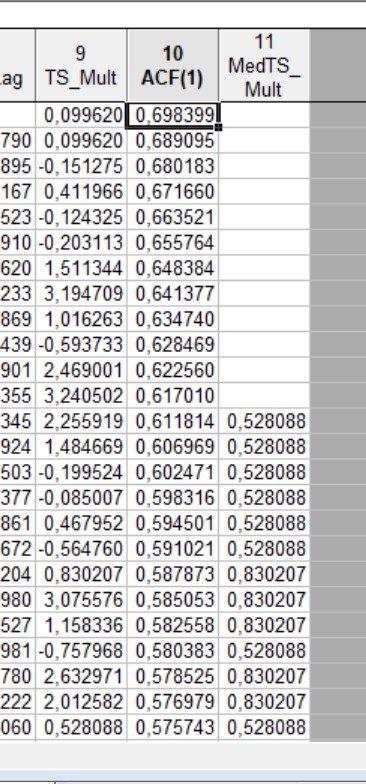


Рисунок – Оцифровка полиномиальной модели

Построим на одном графике временной ряд корреляционного произведения и медиану.

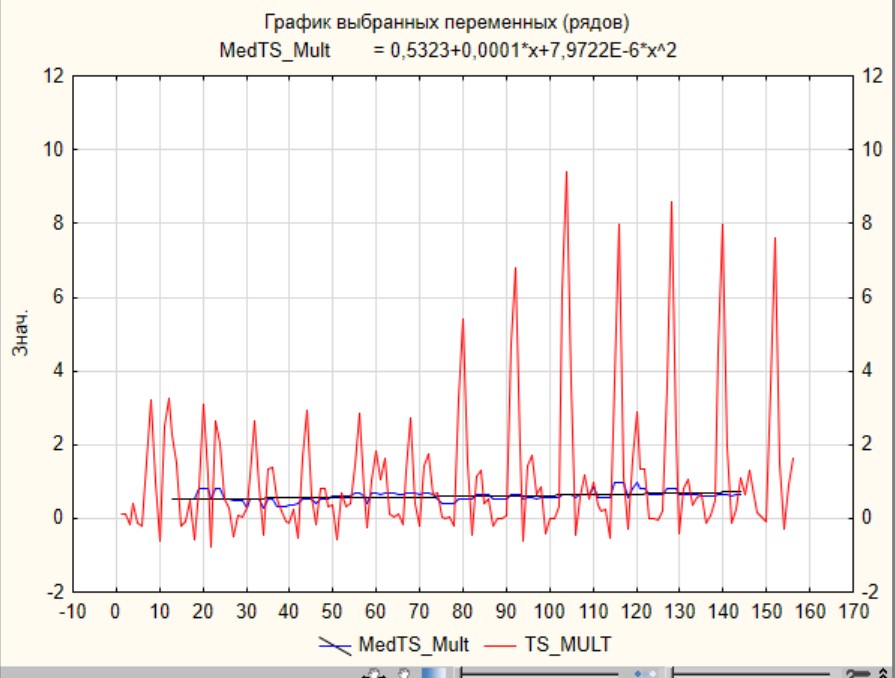


Рисунок – График временного ряда корреляционного произведения и медианы

Аппроксимируем медиану квадратичной моделью. Возьмём формулу и занесём в соседний столбец.

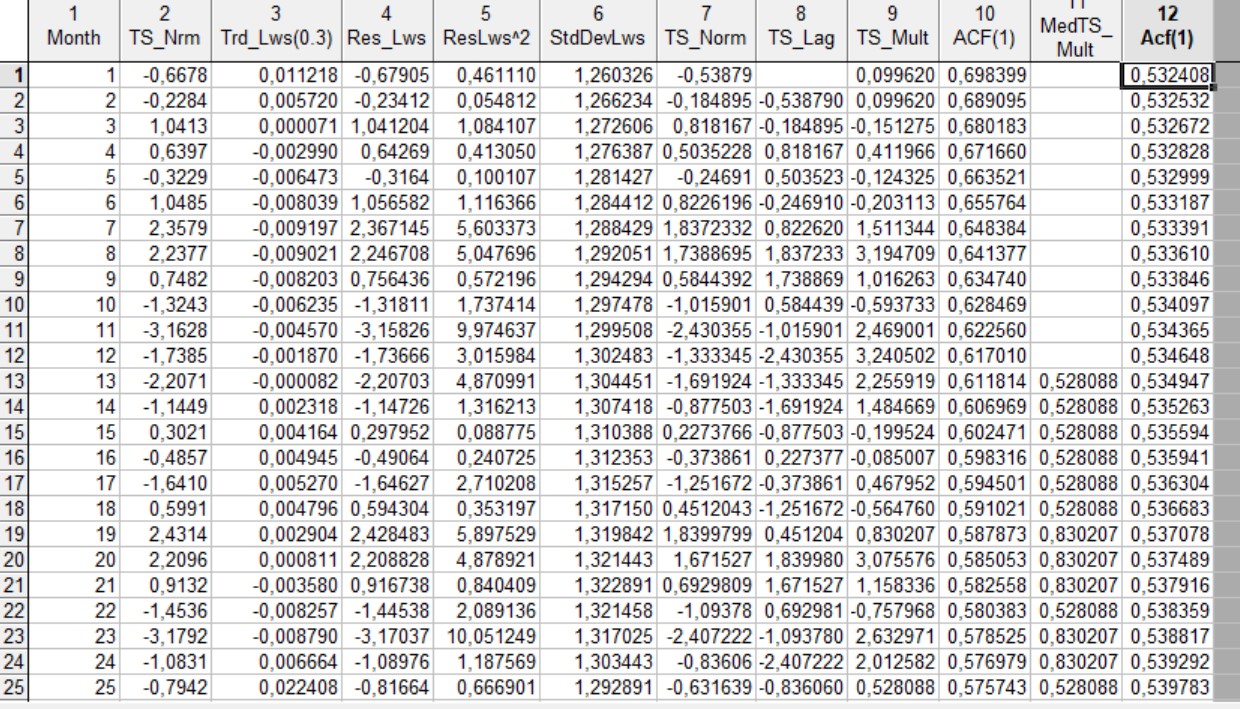


Рисунок – Значения аппроксимации медианы квадратичной моделью

**5 Выводы**

Устойчивость оценки АКФ сильно зависит от размера максимального значения корреляционного лага. Самая грубая оценка может быть получена с помощью диаграммы рассеяния.

Используя средства визуализации пакета STASTICA можно получить наглядные результаты моделирования, позволяющие проводить анализ и прогнозирование.

**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.