

Министерство образования и науки Российской Федерации
(МИНОБРНАУКИ РОССИИ)
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (ТГУ)
Факультет прикладной математики и кибернетики
Кафедра прикладной математики

ДОПУСТИТЬ К ЗАЩИТЕ В ГЭК
Руководитель ООП
д-р техн.наук, профессор


 К.И. Лившиц
« 06 » июня 2016 г.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА
АНАЛИЗ ФИНАНСОВЫХ РЯДОВ МЕТОДОМ «ГУСЕНИЦА – SSA»

по основной образовательной программе подготовки бакалавров
по направлению подготовки
38.03.01 «Экономика»,
профилю подготовки «Математические методы в экономике»

Шкварская Анжелика Игоревна

Руководитель ВКР
доцент, канд.техн.наук

 С.А. Цветницкая
подпись
« 2 » июня 2016 г.

Автор работы
студент группы № 1123


 А.И. Шкварская
подпись

Министерство образования и науки Российской Федерации
(МИНОБРНАУКИ РОССИИ)
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (ТГУ)
Факультет прикладной математики и кибернетики
Кафедра прикладной математики

УТВЕРЖДАЮ

Руководитель ООП

Д.т.н., профессор

 Лившиц К.И.
« 16 » июня 2016 г.

ЗАДАНИЕ

по подготовке ВКР бакалавра

студентке Шкварской Анжелики Игоревны группы № 1123

1. Тема ВКР «Анализ финансовых рядов методом «Гусеница- SSA»\
2. Срок сдачи студентом выполненной ВКР:
 - а) на кафедре « 2 » июня 2016 г.
 - б) в ГЭК « 9 » июня 2016 г.
3. Цель дипломной работы состоит в изучении и применении метода сингулярного спектрального анализа для разложения временного ряда на аддитивные составляющие и прогнозирования временного ряда.
4. План работы включает:
 - а. Изучение литературы:
 - i. Голяндина Н.Э. Метод «Гусеница- SSA»: анализ временных рядов: учеб.пособие. СПб.,2004.
 - 4.1.2 Главные компоненты временных рядов: «Гусеница- SSA» / под ред. Д. Л.Данилова, А.А. Жиглявского. СПб: Пресском, 1997.
 - 4.1.3. Александров, Ф.И. автоматизация выделения трендовых и периодических составляющих временного ряда в рамках метода «Гусеница- SSA» / Ф.И. Александров, Н.Э. Голяндина.

4.1.4. Голяндина Н.Э. Метод «Гусеница- SSA»: прогноз временных рядов: Учеб. Пособие. –СПб., 2004.

б. Разработка алгоритмов анализа и предсказания временных рядов.

с. Реализация разработанных алгоритмов в пакете MathCad.

Дата выдачи задания « 19 » апрель 20 16 г.

Руководитель ВКР

канд. тех. наук, доцент

С.А. Цветницкая

Задание принял к исполнению

19.04.2016

С.А. Цветницкая

РЕФЕРАТ

ВКР содержит 47 страниц, 12 рисунков, 4 таблицы, 4 приложения, 6 источников литературы.

Цель работы: Изучение метода «Гусеница– SSA», особенностей его реализации, способов выделения трендовой составляющей ряда и способов прогнозирования значений ряда. Применение алгоритма «Гусеница– SSA» для сглаживания финансового ряда и суммы неслучайного временного ряда и шума. Применение алгоритма «Гусеница– SSA» для прогнозирования исходной временной последовательности.

Результат: Изучен метод «Гусеница– SSA». Получены результаты сглаживания по методу «Гусеница– SSA» и по методу скользящего среднего. Получены результаты прогнозирования на реальных активах и на модельных примерах. Найдены ошибки прогнозирования.

Расчеты производились в пакете MathCad 14.0.

ОГЛАВЛЕНИЕ

| | |
|---|----|
| Введение | 6 |
| 1 Метод анализа временных рядов «Гусеница-SSA» | 7 |
| 1.1 Алгоритм метода «Гусеница-SSA» | 8 |
| 2 Применение метода «Гусеница-SSA» для сглаживания | 11 |
| 2.1 Сглаживание финансового ряда | 12 |
| 2.2 Сглаживание на модельных примерах | 16 |
| 3 Прогнозирование исходной временной последовательности | 21 |
| 3.1 Построение прогноза на модельных примерах | 23 |
| 3.2 Построение прогноза для финансовых рядов | 26 |
| Заключение | 30 |
| Литература | 31 |
| Приложение А. Алгоритм метода «Гусеница- SSA» | 32 |
| Приложение Б. Результат сглаживания с использованием метода гусеница и метода скользящего среднего | 35 |
| Приложение В. Прогнозирование временной последовательности | 42 |
| Приложение Г. Построение прогноза на модельных данных | 45 |

ВВЕДЕНИЕ

В 80-х годах прошлого века возник метод анализа временных рядов, который в России получил название «Гусеница», а за рубежом – Singular Spectrum Analysis (SSA). Этот метод является новой и сильной техникой, применимой ко многим практическим проблемам, таким, как исследования в области многомерной статистики, многомерной геометрии, динамических систем и обработки сигналов.

Прикладные области SSA разнообразны: от математики и физики до экономики и финансовой математики. SSA возник, как метод анализа и первоначально применялся к одномерным рядам. Отличительной чертой метода является то, что он не требует предварительного задания модели ряда.

Метод SSA позволяет:

- 1) Различать составляющие временного ряда, полученные из последовательности какой-либо величины, взятой через равные промежутки времени: медленные тренды, сезонные и другие периодические или колебательные составляющие, а также шумовые компоненты;
- 2) Находить заранее известные периодичности рядов ;
- 3) Сглаживать исходные данные на основе отобранных составляющих;
- 4) Наилучшим образом выделять компоненту с заранее известным периодом;
- 5) Предсказывать дальнейшее поведение наблюдаемой зависимости.

1 Метод анализа временных рядов «Гусеница – SSA»

Для анализа временного ряда выбирается целый параметр L – длина окна. Этот параметр выбирается произвольно. При достаточно большой длине ряда и достаточно большом L результаты не будут зависеть от длины окна. Далее на основе ряда строится траекторная матрица, столбцами которой являются скользящие отрезки длины L . Следующий шаг – это сингулярное разложение траекторной матрицы в сумму элементарных матриц. Каждая элементарная матрица задается набором из собственного числа и собственных векторов.

Пусть исходный ряд является суммой нескольких рядов. Теоретические результаты позволяют определить по виду собственных чисел и собственных векторов. Суммируя элементарные матрицы внутри каждого набора и затем, переходя от результирующих матриц к ряду, мы получаем разложение ряда на аддитивные слагаемые, такие как сумма тренда, периодики и шум.

Возможность разбить совокупность элементарных матриц на группы тесно связана с понятием разделимости рядов.

Таким образом, целью метода является разложение временного ряда на интерпретируемые аддитивные составляющие. При этом метод не требует стационарности ряда, знания модели тренда. Это является достоинством метода.

Недостатками метода является неавтоматическая группировка компонент сингулярного разложения траекторной матрицы ряда для получения составляющих исходного ряда и отсутствие модели не позволяет проверять гипотезы о наличии в ряде той или иной составляющей.

2 1.1 Алгоритм метода «Гусеница- SSA»

Рассмотрим временной ряд $(f_i)_{i=1}^N$, где f_i – i – ая компонента ряда, N – длина ряда. Основным параметром алгоритма является длина окна L , $1 < L < N$. Результатом работы алгоритма является разбиение ряда на аддитивные составляющие.

Алгоритм метода «гусеница- SSA» можно разбить на четыре последовательно выполняемых шагов:

1 шаг. Построение траекторной матрицы

2 шаг. Сингулярное разложение

3 шаг. Группировка матриц

4 шаг. Диагональное усреднение

Шаг 1. Построение траекторной матрицы

Первый шаг заключается в построении траекторной матрицы X размерности $L \times K$, ($K = N - L + 1$).

Из временного ряда $(f_i)_{i=1}^N$ выбираются отрезки длины L , причем данные отрезки строятся последовательно, то есть первый отрезок будет включать в себя точки ряда с первой по L – ю, второй отрезок со второй по $L+1$ – ю точку и так далее.

И в качестве столбцов траекторной матрицы X выбираются построенные отрезки, которые принято называть векторами вложения.

Траекторная матрица принимает вид:

$$X = \begin{pmatrix} f_1 & f_2 & \cdots & f_{N-L+1} \\ f_2 & f_3 & \cdots & f_{N-L+2} \\ \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & \cdots & f_N \end{pmatrix} \quad (1)$$

Построенная матрица X является ганкелевой. Ганкелева матрица представляет собой матрицу, у которой на всех диагоналях стоят равные элементы.

Шаг 2. Сингулярное разложение

Результатом этого этапа является разложение траекторной матрицы X на элементарные части, т.е. представление матрицы X в виде суммы элементарных матриц.

Пусть $S = XX^T$. Обозначим через $\lambda_1, \lambda_2, \dots, \lambda_L$ собственные числа матрицы, взятые в убывающем порядке, а через U_1, U_2, \dots, U_L систему собственных векторов матрицы S , соответствующих собственным числам. Тогда сингулярное разложение траекторной матрицы X может быть записано следующим образом:

$$X_i = U_i U_i^T X \quad (2)$$

где $X = X_1 + X_2 + \dots + X_L$, обозначим $\tilde{X} = \sum_{i=1}^r X_i$, где r – количество медленно меняющихся собственных векторов.

Шаг 3. Группировка матриц

В результате сингулярного разложения матрицы X имеем набор собственных векторов и собственных значений. Матрица X может быть представлена в виде суммы матриц $X = X_1 + X_2 + \dots + X_L$,

$$\text{где } X_i = U_i U_i^T X$$

В данной работе матрицу X будем представлять в виде суммы двух матриц

$$X = \tilde{X} + X_{\text{шум}}, \text{ где} \quad (3)$$

\tilde{X} – сглаженная составляющая матрицы X , $X_{\text{шум}}$ – шумовая составляющая.

Для того чтобы выделить матрицы X_i , относящиеся к \tilde{X} , анализируются собственные векторы и собственные значения.

По виду собственных векторов можно определить, какие составляющие относятся к сглаженной группе. К сглаженной группе относят векторы U_i компоненты которых медленно меняются.

Второй способ выделения сглаженных составляющих – анализ собственных значений. Вклад составляющей матрицы X_i в матрицу X определяется собственным значением λ_i . Так как компоненты упорядочены в порядке убывания, следовательно, компонента с меньшим номером вносит больший вклад в вид ряда, а незначительные компоненты имеют большие номера.

Множество индексов L разбиваем на два подмножества I_1 и I_2 . В I_1 компоненты, относящиеся к сглаженной группе \tilde{X} , а в I_2 все остальные.

Каждой матрице X_i соответствует временной ряд f_i . Так как матрицы X_i , в общем случае, не являются ганкелевыми, то на четвертом шаге алгоритма производим операцию диагонального усреднения.

Шаг 4. Диагональное усреднение

Теперь исходный временной ряд, полученный из матрицы \tilde{X} , определится как:

$$f_k^i = \left\{ \begin{array}{l} \frac{1}{k} \sum_{i=1}^k \tilde{X}_{k,k-i+1} \text{ для } 1 \leq k \leq L \\ \frac{1}{L} \sum_{i=1}^L \tilde{X}_{k,k-i+1} \text{ для } L \leq k \leq K \\ \frac{1}{N-k+1} \sum_{i=1}^{N-k+1} \tilde{X}_{i+k-K, K-i+1} \text{ для } K \leq k \leq N \end{array} \right\} \quad (4)$$

Таким образом, результатом работы алгоритма является разложение временного ряда на интерпретируемые аддитивные составляющие, «гусеница- SSA» может решать различные задачи такие, как выделение тренда, сглаживание ряда, определение периодик.

2 Применение «Гусеница- SSA» для сглаживания

В результате сингулярного разложения матрицы X имеем набор собственных векторов и собственных значений. Матрица X представлена в виде суммы матриц

$$X = X_1 + X_2 + \dots + X_L \quad (5)$$

Вклад составляющей матрицы X_i в матрицу X определяется собственным значением λ_i .

В данной работе сглаженный ряд, результаты которого получены методом «гусеница- SSA», сравнивался с результатами, полученными методом скользящего среднего.

Скользящее среднее определяется по формуле

$$\widetilde{S}_n = \sum_{i=1}^p \frac{S_i}{p}, \text{ где} \quad (6)$$

S_i – цены(наиболее часто берут цены закрытия),

n – количество дней, за которое рассчитывается скользящее среднее, это основной параметр при построении, его еще называют длина сглаживания,

p – длина ряда.

2.1 Сглаживание финансового ряда

Пусть исходный временной ряд представляет собой ежедневные цены акции Роснефть с 16.03.2015 по 16.03.2016 г. Ряд изображен на рисунке 1 и состоит из 250 наблюдений

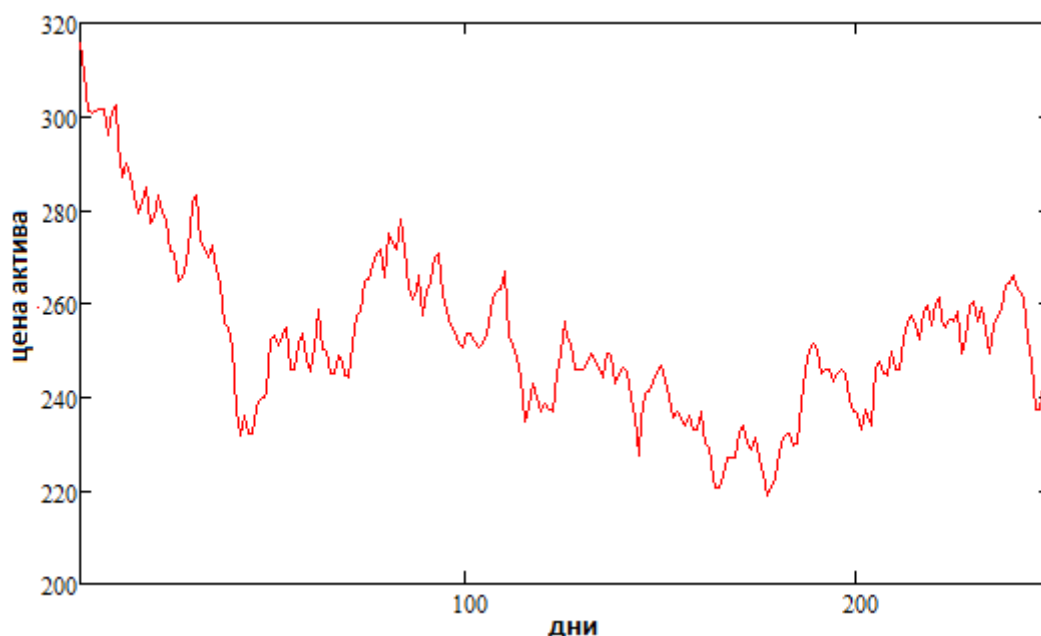


Рисунок 1 – Динамика цены акции Роснефть

Для выделения сглаженных составляющих можно учитывать информацию, которую несут собственные векторы и собственные значения. К сглаженной группе относят те составляющие, собственные векторы которых медленно меняются. Выделение таких составляющих можно производить визуально, анализируя графики собственных векторов.

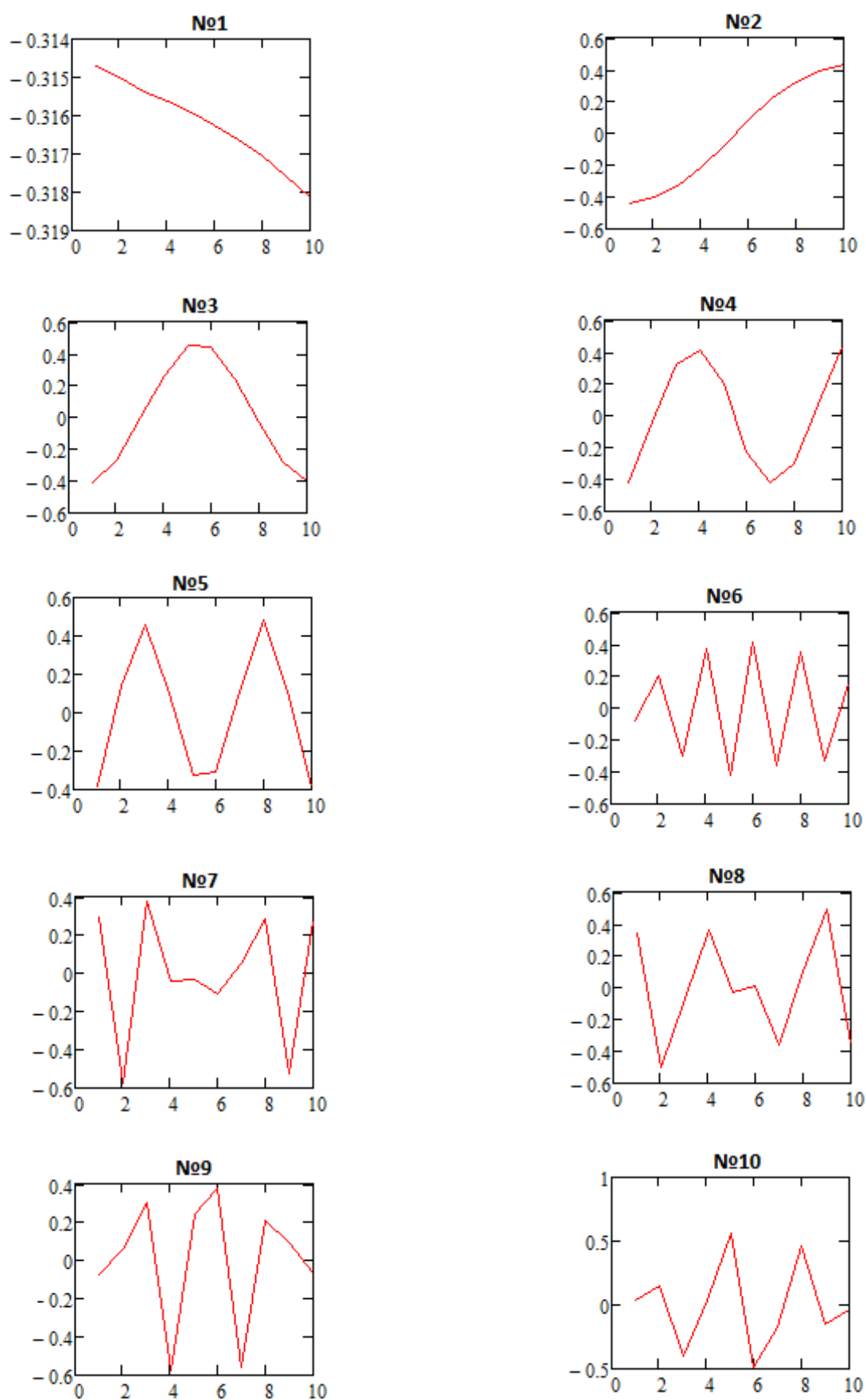


Рисунок 2 – Графики собственных векторов сингулярного разложения

На рисунке 2 видно, что первый и второй векторы изменяются медленно, и поэтому они относятся к сглаженной группе. Кроме собственных векторов, при исследовании ряда можно учитывать информацию, которую содержат в себе соответствующие собственные числа. Для идентификации компонент можно принять во внимание тот факт, что компоненты упорядочены в порядке убывания собственных чисел. Следовательно, компонента с меньшим номером вносит больший вклад в вид ряда, а незначительные компоненты имеют большие номера. Обозначим через I – множество выбранных компонент разложения относящихся к сглаженным составляющим. Тогда отношение

$$u = \frac{\sum_{j \in I} \lambda_j}{\sum_{i=1}^L \lambda_i} \quad (7)$$

отражает вклад сглаженных составляющих в исходный ряд f . Введем следующие обозначения $\alpha_i = \frac{\lambda_i}{\sum \lambda}$. В таблице 1 представлены собственные числа сингулярного разложения и отношение $\alpha_i = \frac{\lambda_i}{\sum \lambda}$ характеризует вклад составляющей f_i в разложение временного ряда f . Из таблицы 1 видно, что первой и второй компонентам разложения, выбранным ранее, соответствуют наибольший вклад в разложение и наибольшие значения собственных чисел, это еще раз подтверждает правильность их выбора в качестве сглаженных компонент. Для данного временного ряда наблюдалась достаточно хорошая разделимость компонент, так как собственные числа существенно различались между собой.

Таблица 1 – Собственные числа λ_i и отношение α_i

| | | | | | |
|--------------------|-----------|-----------|----------|---------|-----|
| Номер компоненты | 1 | 2 | 3 | 4 | ... |
| Вклад в разложение | 99,9% | 0,03% | 0,01% | 0,005% | ... |
| Собственное число | 4740720,8 | 1719,3928 | 507,3445 | 251,881 | ... |

Методами автоматической идентификации сглаженной составляющей являются: метод Кендала, метод низких частот, метод нулей и др..

Рассмотрим метод нулей. Основная идея этого метода состоит в построении графиков собственного вектора и подсчете числа нулей – точек, в которых график пересекает ось абсцисс. Чем больше число нулей, тем больше вероятность того, что рассматриваемый собственный вектор соответствует периодической составляющей или шуму, но не тренду.

Результат сглаживания исходного ряда по двум компонентам приведен на рисунке 3.

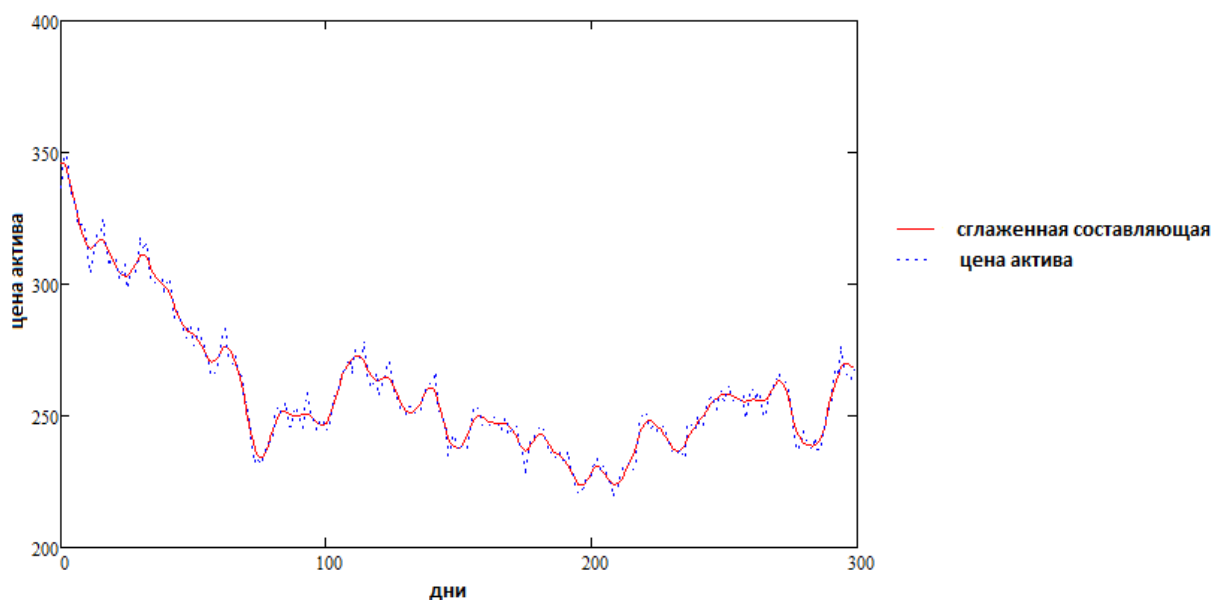


Рисунок 3 – Графики курса акции Роснефть и сглаженная составляющая

2.2 Сглаживание на модельных примерах

Рассмотрим выделение тренда на модельных примерах. В качестве примера рассмотрим три модели. Результаты оценивания тренда первой и 2 модели приведены в приложении Б.

1) $f(t) = \sin(t \cdot 0.1)$

2) $f(t) = \exp(t \cdot 0.01)$

3) $f(t) = \sin(t \cdot 0.1) \cdot \exp(t \cdot 0.01)$

$$f1(t) = f(t) + z(t), \text{ где } z(t) - \text{шум.} \quad (8)$$

Результаты оценивания тренда модели $f(t) = \sin(t \cdot 0.1) \cdot \exp(t \cdot 0.01)$ для двух значений σ приведены на рисунках 4 – 9.

В качестве шума z использовали нормально распределенные случайные величины с нулевым средним и среднеквадратическим отклонением σ . Выделение тренда осуществляли методом гусеница и методом скользящего среднего.

На рисунке 4-6 приведены результаты оценивания для $\sigma = 0.1$.

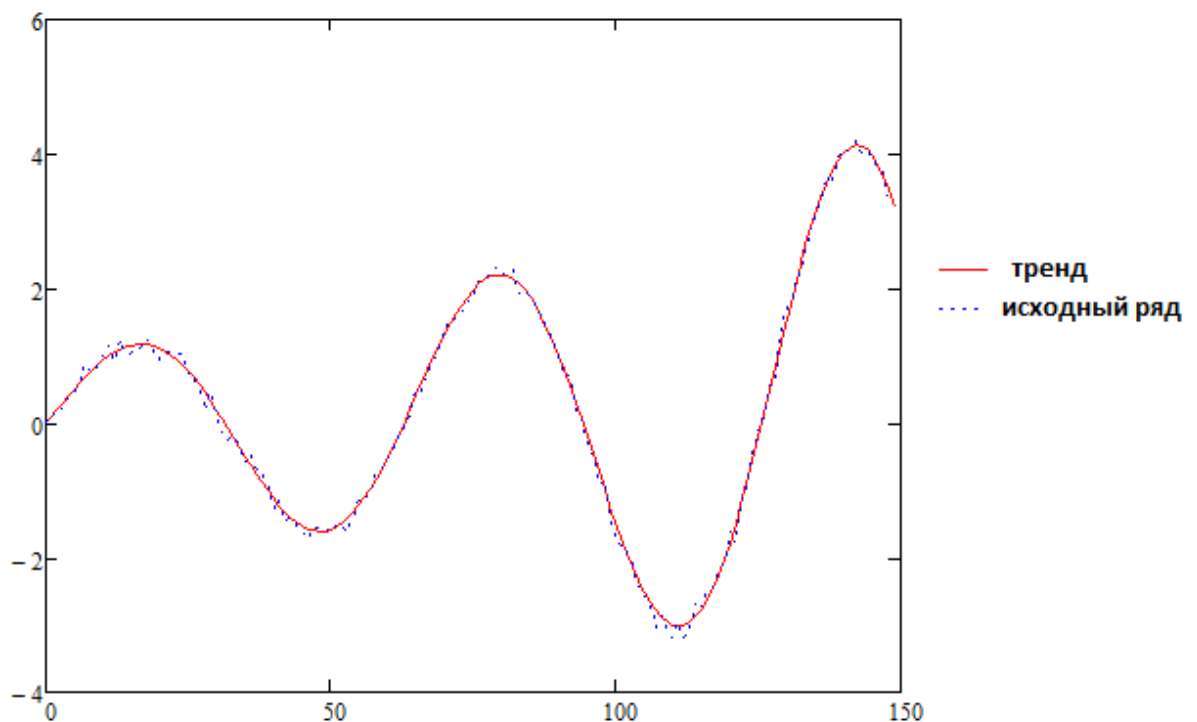


Рисунок 4 – Тренд и зашумленный временной ряд

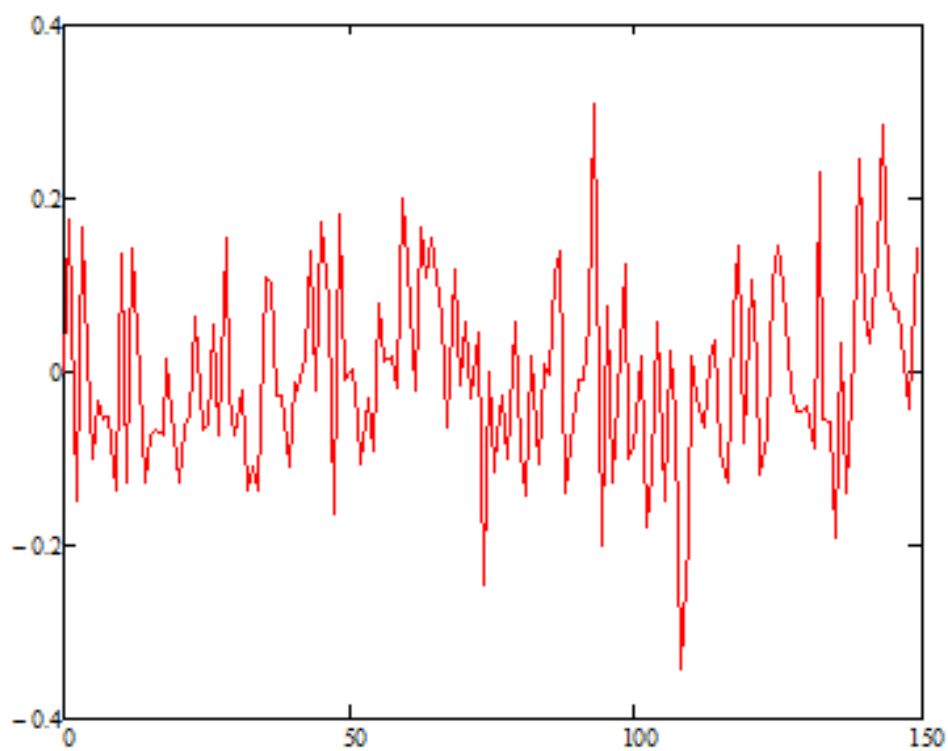


Рисунок 5 – Динамика шума

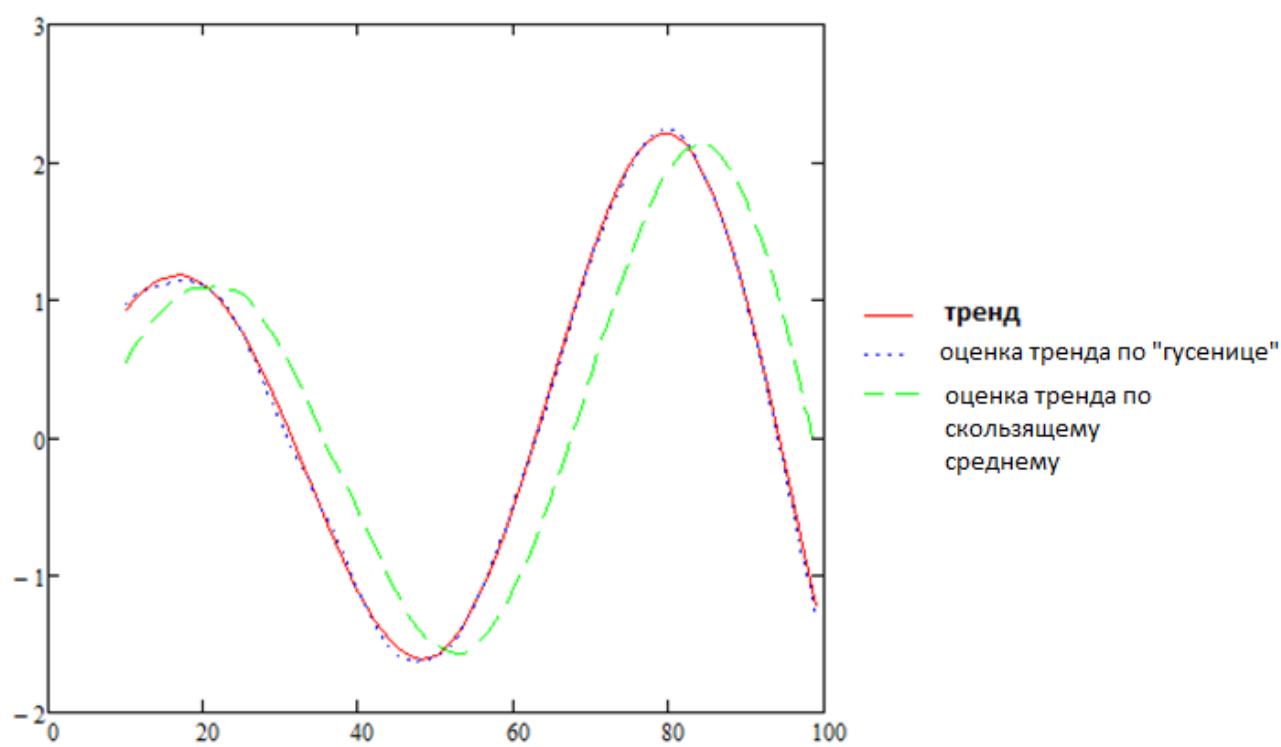


Рисунок 6 – Тренд, оценки тренда

Аналогичные результаты получены для $\delta = 0.3$

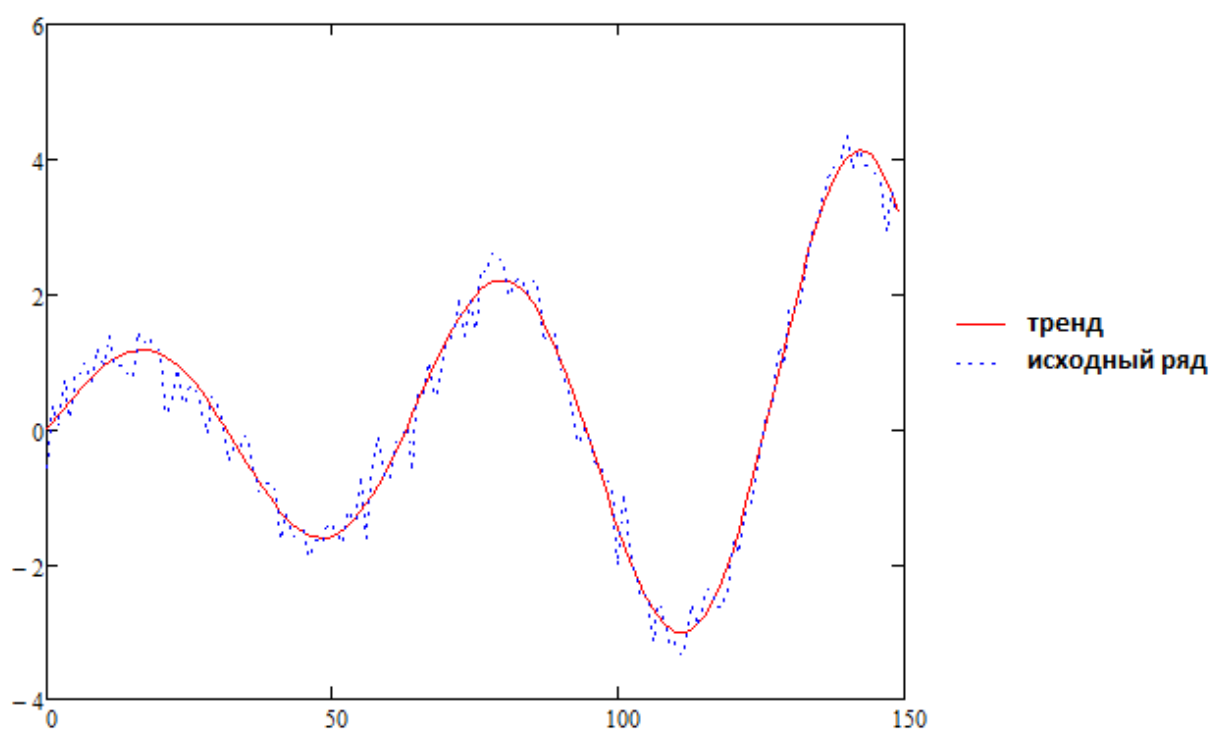


Рисунок 7 – Тренд и зашумленный временной ряд

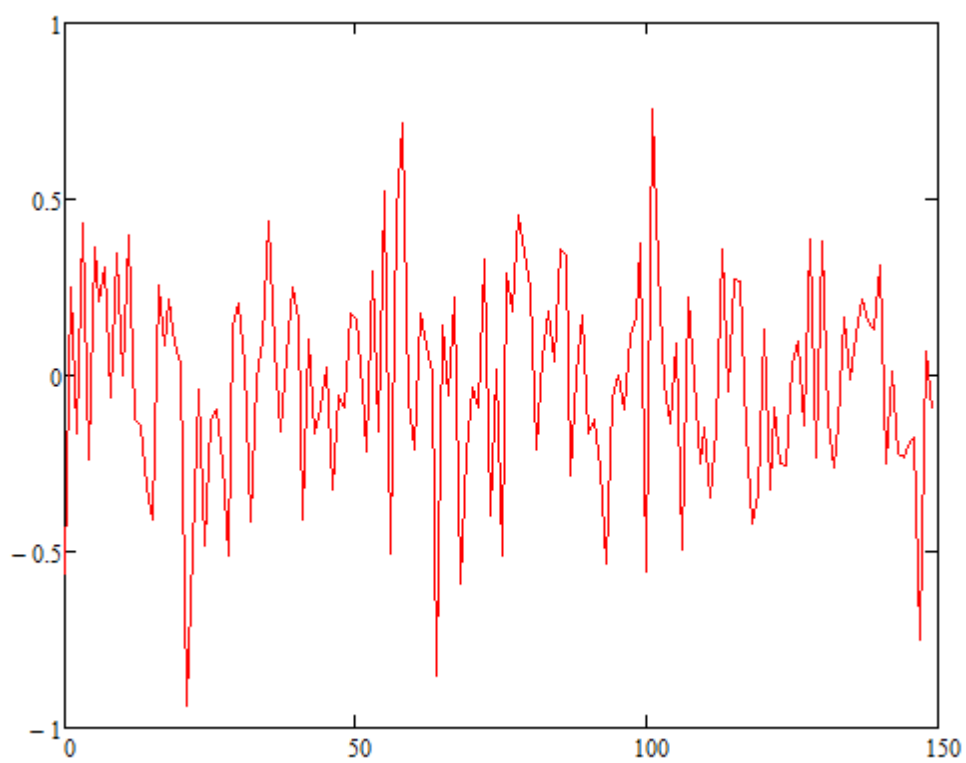


Рисунок 8 – Динамика шума

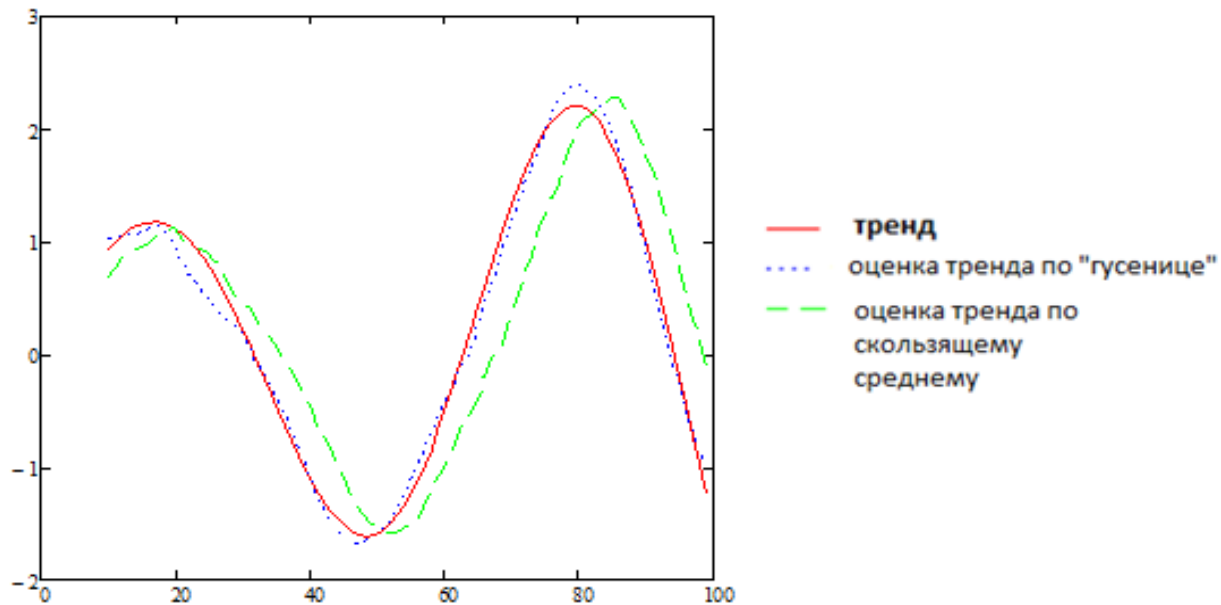


Рисунок 9 – Тренд, оценки тренда

Был проведен анализ ошибки сглаживания. Для вычисления оценки ошибки сглаживания была применена средняя абсолютная ошибка в процентах (*Mean Absolute Percentage Error*):

$$\text{MAPE} = \frac{1}{p} \sum_{i=1}^p \frac{|f_i - f1_i|}{f_i} \cdot 100\% \quad (9)$$

где f_i – тренд, $f1_i$ – оценка тренда, p – длина ряда.

Показатель MAPE, как правило, характеризует относительную точность прогноза. Для прогноза высокой точности $\text{MAPE} < 10\%$, хорошей – $10\% < \text{MAPE} < 20\%$, удовлетворительной – $\text{MAPE} > 50\%$.

В таблице 2 приведены результаты средней абсолютной ошибки MAPE.

Таблица 2 – Результаты MAPE

| Среднеквадратическое отклонение | Метод скользящего среднего | Метод гусеница |
|---------------------------------|----------------------------|----------------|
| $\delta = 0.1$ | 127% | 10,7% |
| $\delta = 0.3$ | 138,3% | 19,1% |

3 Прогнозирование исходной временной последовательности

Алгоритм метода «Гусеница» может быть использован для предсказания.

Пусть известен числовой ряд $f_{i=1}^N$. На основе его можно построить ганкелевую матрицу X . В качестве базиса поверхности, содержащей эту выборку, могут быть использованы векторы $[U^1, U^2, \dots, U^L]$ матрицы S .

Для целей предсказания мы будем использовать базис, составленный из векторов $U^j, j \in I_1, j = 1, 2, \dots, r$.

Разложение X^i столбца по базису $[U^1, U^2, \dots, U^r]$ примет вид

$$[U^1 \ U^2 \ \dots \ U^r] \begin{pmatrix} p_1^i \\ p_2^i \\ p_3^i \end{pmatrix} = X^i \quad (10)$$

Так как число уравнений в (10) больше числа неизвестных ($L > r$), то мы можем найти вектор параметров $(p_1^i, p_2^i, \dots, p_r^i)$ из системы $L - 1$ уравнений

$$V_* = \begin{bmatrix} U_1^1 & U_1^2 & \dots & U_1^r \\ \vdots & \vdots & \dots & \vdots \\ U_{L-1}^1 & U_{L-1}^2 & \dots & U_{L-1}^r \end{bmatrix} \begin{pmatrix} p_1^i \\ \dots \\ p_r^i \end{pmatrix} = \begin{pmatrix} X_1^i \\ \dots \\ X_{L-1}^i \end{pmatrix} \quad (11)$$

Систему (11) назовем усеченной системой.

Для предсказания значения f_{N+1} рассмотрим вектор столбец

$$X^{N-L+2} = (f_{N-L+1}, \dots, f_{N+1})^T. \quad (12)$$

Разложение вектора X^{N-L+2} примет вид

$$[U^1 U^2 \ \dots \ U^r] \begin{pmatrix} p_1^{N-L+2} \\ \dots \\ p_r^{N-L+2} \end{pmatrix} = \begin{pmatrix} f_{N-L+2} \\ \dots \\ f_{N+1} \end{pmatrix} \quad (13)$$

Параметры $(p_1^{N-L+2} \ \dots \ p_r^{N-L+2})$ найдем из усеченной системы

$$V_* \begin{pmatrix} p_1^{N-L+2} \\ \dots \\ p_r^{N-L+2} \end{pmatrix} = \begin{pmatrix} f_{N-L+2} \\ \dots \\ f_N \end{pmatrix} \quad (14)$$

Тогда предсказанное значение найдем из соотношения

$$f_{N+1} = [U^1 U^2 \dots U^r] \begin{pmatrix} p_1^{N-L+2} \\ \dots \\ p_r^{N-L+2} \end{pmatrix} \quad (15)$$

3.1 Построение прогноза на модельных примерах

В качестве примера рассмотрим модель :

$$f(t) = \sin(t \cdot 0.1) \cdot \exp(t \cdot 0.01) \quad (16)$$

$$f1(t) = f(t) + z(t), \text{ где } z(t) - \text{шум} \quad (17)$$

В качестве шума z использовали нормально распределенные случайные величины с нулевым средним и среднеквадратическим отклонением σ .

Результаты предсказания двух других моделей приведены в приложении Г.

Результаты предсказания этой модели для двух значений сигма приведены на рисунках.

На рисунке 10 приведен результат оценивания для $\sigma = 0.1$.

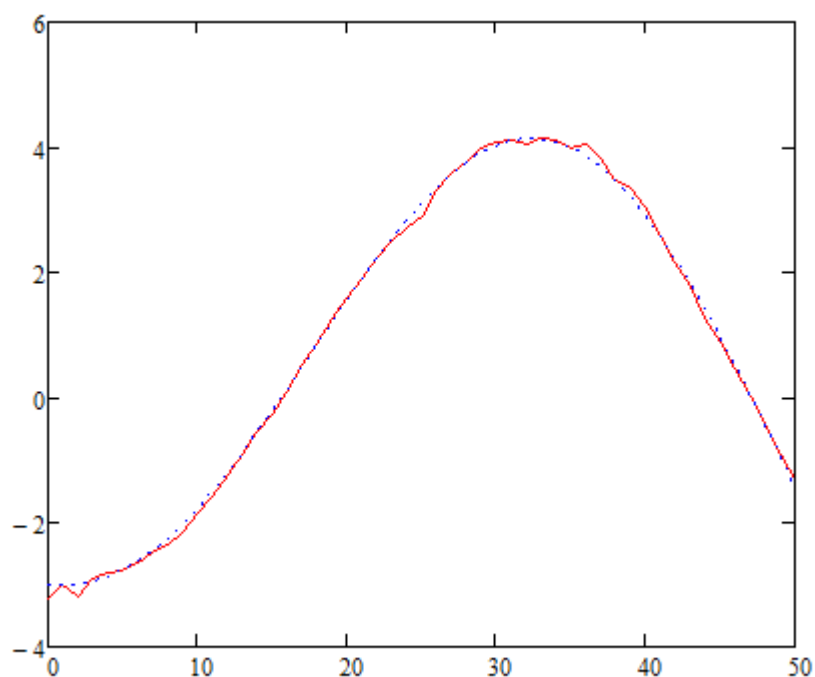


Рисунок 10 – Зашумленный временной ряд и предсказанное значение ряда

Аналогичные результаты получены для $\delta = 0.3$

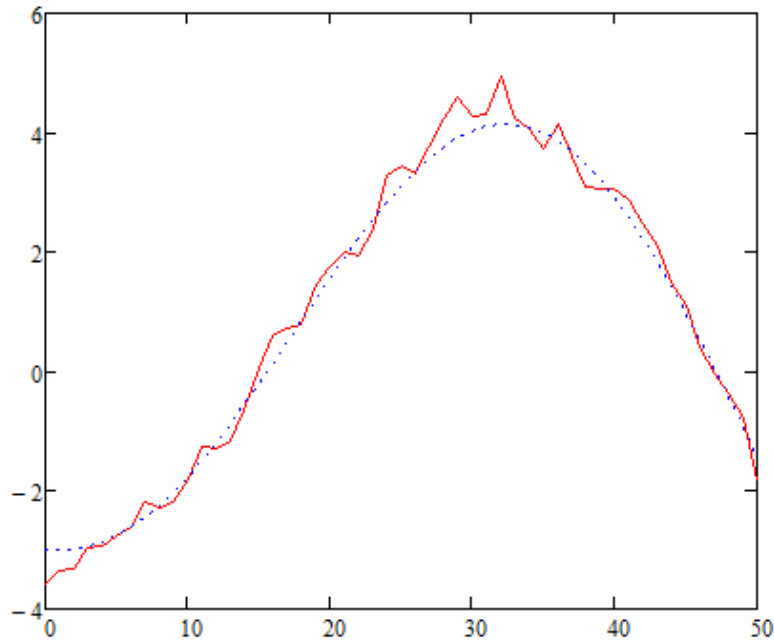


Рисунок 11 – Зашумленный временной ряд и предсказанное значение ряда

Был проведен анализ ошибки прогноза. Для вычисления оценки ошибки прогноза была применена средняя абсолютная ошибка в процентах (*Mean Absolute Percentage Error*):

$$\text{MAPE} = \frac{1}{p} \sum_{i=1}^p \frac{|Z_i - \hat{Z}_i|}{Z_i} \cdot 100\% \quad (18)$$

Где p – период, на который был сделан прогноз, \hat{Z}_i – спрогнозированное значение ряда в точке i , Z_i – фактическое значение ряда в точке i .

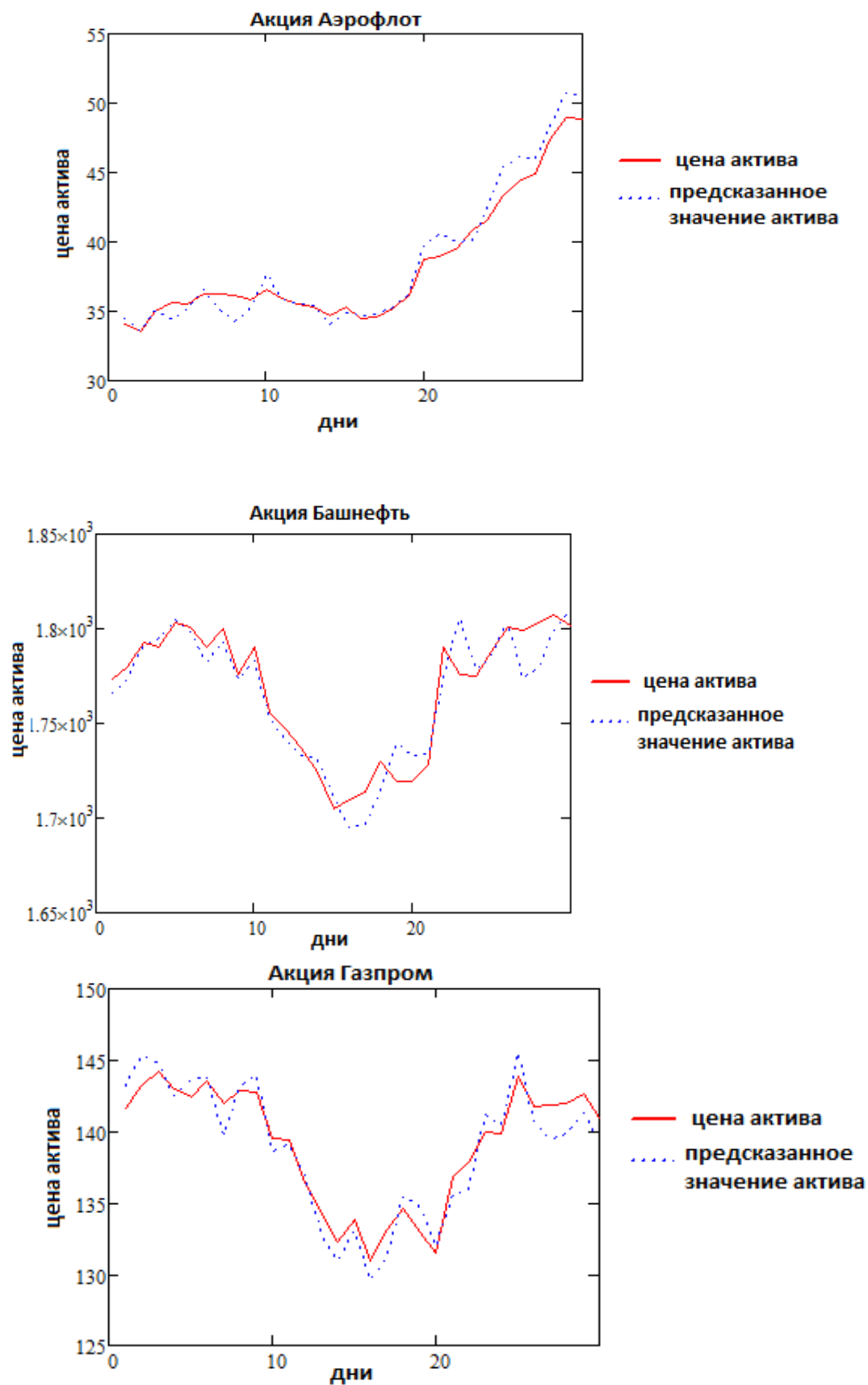
Показатель MAPE, как правило, характеризует относительную точность прогноза. Для прогноза высокой точности $\text{MAPE} < 10\%$, хорошей – $10\% < \text{MAPE} < 20\%$, удовлетворительной – $\text{MAPE} > 50\%$.

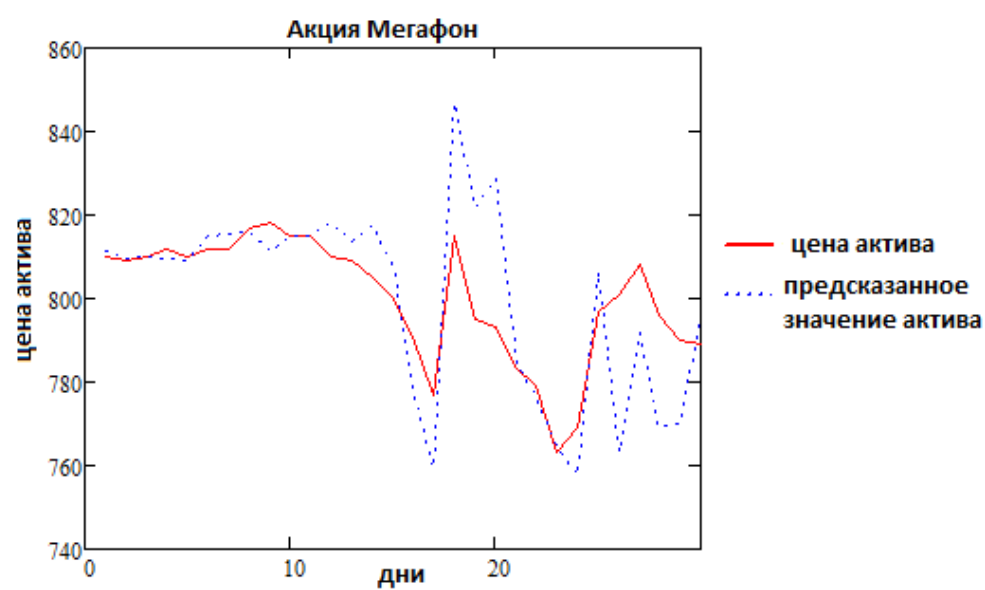
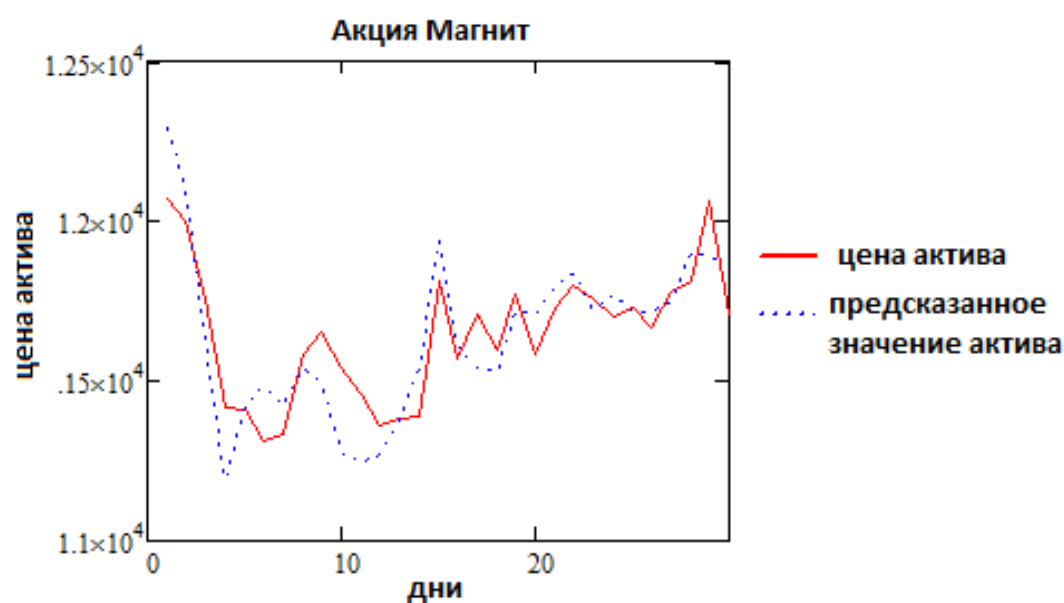
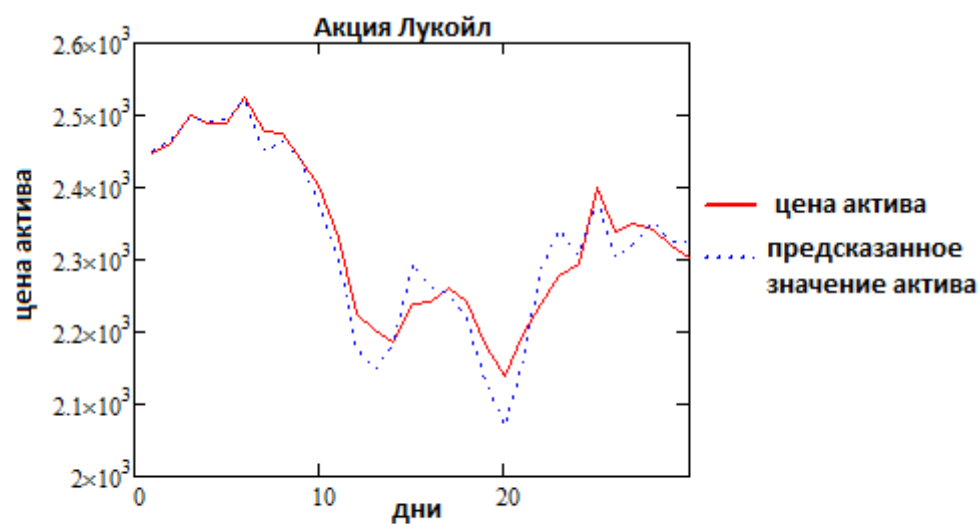
Таблица 3 – Результат MAPE

| Среднеквадратическое отклонение | Ошибка прогноза(в %) |
|---------------------------------|----------------------|
| $\delta = 0.1$ | 7,44 |
| $\delta = 0.3$ | 21,01 |

3.2 Построение прогноза для финансовых рядов

На рисунке 12 показаны графики прогноза цен акции за 30 дней.





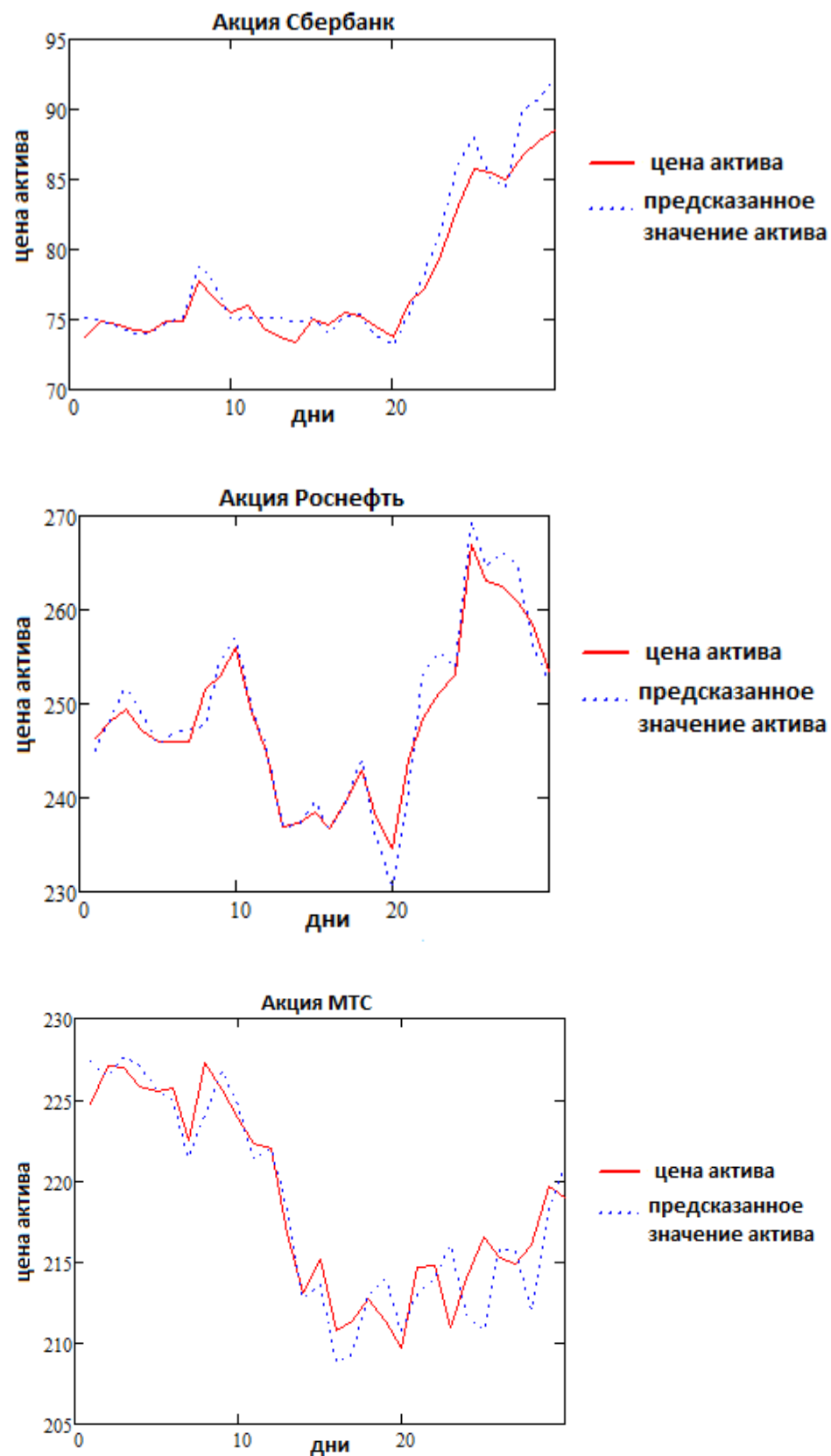


Рисунок 12 – Прогноз тренда и реальные цены активов

Был проведен анализ ошибки прогноза

$$\text{MAPE} = \frac{1}{p} \sum_{i=1}^p \frac{|z_i - \hat{z}_i|}{z_i} \cdot 100\% \quad (19)$$

Показатель MAPE, как правило, характеризует относительную точность прогноза. Для прогноза высокой точности $\text{MAPE} < 10\%$, хорошей – $10\% < \text{MAPE} < 20\%$, удовлетворительной – $\text{MAPE} > 50\%$.

Таблица 4 – Результат MAPE

| Название акции | Ошибка прогноза на 30 дней(в %) |
|----------------|---------------------------------|
| Сбербанк | 0,428 |
| Газпром | 0,32 |
| Роснефть | 0,665 |
| Магнит | 2,136 |
| Лукойл | 0,105 |
| Аэрофлот | 3,711 |
| Башнефть | 0,231 |
| Мегафон | 0,318 |
| МТС | 0,552 |

ЗАКЛЮЧЕНИЕ

В данной работе был рассмотрен метод анализа и прогнозирования временных рядов «Гусеница- SSA». Изучена структура и особенности применения данного метода. Получены результаты прогнозирования для модельных примеров и реальных активов. Получены результаты ошибки прогнозирования. Сделан вывод, что «Гусеница- SSA» дает хорошие результаты прогнозирования.

Применение этого метода для выделения сглаженной составляющей финансового ряда, представляющего ежедневные цена акции, и суммы неслучайного временного ряда и шума. Результаты по Методу «Гусеница- SSA» сравнивались с результатами, полученными по методу скользящего среднего. В результате метод «Гусеница- SSA» дает более точные результаты.

ЛИТЕРАТУРА

1. Голяндина Н.Э. Метод «Гусеница- SSA»: анализ временных рядов: учеб.пособие. СПб.,2004
2. Главные компоненты временных рядов: «Гусеница- SSA» / под ред. Д. Л.Данилова, А.А. Жиглявского. СПб: Пресском, 1997.
3. Александров, Ф.И. автоматизация выделения трендовых и периодических составляющих временного ряда в рамках метода «Гусеница- SSA» / Ф.И. Александров, Н.Э. Голяндина.
4. Голяндина Н.Э. Метод «Гусеница- SSA»: прогноз временных рядов: Учеб. Пособие. –СПб., 2004.
5. Цыплаков А. Введение в прогнозирование в классических моделях временных рядов //Квантиль. №1 (2006).
6. Н. Э. Голяндина, Е. В. Осипов. Метод «Гусеница- SSA»: для анализа временных рядов с про- пусками, С.-Петербургский государственный университет.

ПРИЛОЖЕНИЕ А. Алгоритм метода «Гусеница- SSA»

Построение траекторной матрицы

$$X(f, N, \tau) := \begin{cases} \text{for } i \in 0..N - \tau \\ \quad \text{for } j \in 0..\tau - 1 \\ \quad \quad X_{i,j} \leftarrow f_{i+j} \end{cases}$$

$X =$

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----|--------|--------|--------|--------|--------|--------|--------|--------|
| 0 | 8.005 | 9.773 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 |
| 1 | 9.773 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 |
| 2 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 |
| 3 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 |
| 4 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 |
| 5 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 |
| 6 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 |
| 7 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 |
| 8 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 |
| 9 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 |
| 10 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 |
| 11 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 |
| 12 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 |
| 13 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 |
| 14 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 | 32.824 |
| 15 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 | 32.824 | ... |

Сингулярное разложение

$$S := X \cdot X^T$$

Собственные числа матрицы S

eigenvecs (S) =

| | 0 | 1 | 2 | 3 | 4 |
|----|-------|------------------------|--------|------------------------|--------|
| 0 | 0.125 | 0.356 | -0.24 | -0.158 | -0.194 |
| 1 | 0.132 | 0.332 | 0.097 | -0.234 | 0.108 |
| 2 | 0.137 | 0.303 | 0.106 | -0.076 | -0.379 |
| 3 | 0.142 | 0.278 | 0.306 | 0.232 | -0.062 |
| 4 | 0.148 | 0.259 | -0.217 | 0.321 | -0.182 |
| 5 | 0.153 | 0.244 | -0.093 | 0.073 | 0.233 |
| 6 | 0.158 | 0.218 | -0.163 | -0.113 | -0.147 |
| 7 | 0.163 | 0.19 | 0.359 | $6.032 \cdot 10^{-3}$ | 0.126 |
| 8 | 0.168 | 0.157 | 0.043 | 0.293 | -0.208 |
| 9 | 0.173 | 0.123 | -0.021 | 0.303 | 0.457 |
| 10 | 0.179 | 0.097 | -0.383 | $-1.767 \cdot 10^{-3}$ | 0.05 |
| 11 | 0.184 | 0.09 | 0.082 | -0.23 | 0.291 |
| 12 | 0.19 | 0.058 | 0.093 | -0.112 | 0.042 |
| 13 | 0.195 | 0.017 | 0.186 | 0.086 | 0.2 |
| 14 | 0.201 | $-2.703 \cdot 10^{-3}$ | -0.175 | 0.085 | 0.067 |
| 15 | 0.206 | -0.021 | -0.131 | -0.156 | ... |

ПРИЛОЖЕНИЕ А

$$\frac{\lambda_j}{\sum \lambda} =$$

$\lambda := \text{eigenvals}(S) =$

| | 0 |
|----|--------------------|
| 0 | $9.13 \cdot 10^5$ |
| 1 | $2.327 \cdot 10^3$ |
| 2 | 89.411 |
| 3 | 85.299 |
| 4 | 78.24 |
| 5 | 74.08 |
| 6 | 63.155 |
| 7 | 54.024 |
| 8 | 53.42 |
| 9 | 42.559 |
| 10 | 45.705 |
| 11 | 44.766 |
| 12 | 27.363 |
| 13 | 23.572 |
| 14 | 18.514 |
| 15 | ... |

| |
|-----------------------|
| 0.997 |
| $2.541 \cdot 10^{-3}$ |
| $9.76 \cdot 10^{-5}$ |
| $9.311 \cdot 10^{-5}$ |
| $8.541 \cdot 10^{-5}$ |
| $8.087 \cdot 10^{-5}$ |
| $6.894 \cdot 10^{-5}$ |
| $5.897 \cdot 10^{-5}$ |
| $5.831 \cdot 10^{-5}$ |
| $4.646 \cdot 10^{-5}$ |
| $4.989 \cdot 10^{-5}$ |
| $4.887 \cdot 10^{-5}$ |
| $2.987 \cdot 10^{-5}$ |
| $2.573 \cdot 10^{-5}$ |
| $2.021 \cdot 10^{-5}$ |
| ... |

$$X_{\text{vost}} := \sum_{i=0}^{\tau-1} XX1_i =$$

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----|--------|--------|--------|--------|--------|--------|--------|--------|
| 0 | 8.005 | 9.773 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 |
| 1 | 9.773 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 |
| 2 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 |
| 3 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 |
| 4 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 |
| 5 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 |
| 6 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 |
| 7 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 |
| 8 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 |
| 9 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 |
| 10 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 |
| 11 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 |
| 12 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 |
| 13 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 |
| 14 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 | 32.824 |
| 15 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 | 32.824 | ... |

ПРИЛОЖЕНИЕ А

```

diag(X) := for s ∈ 1..N
|
|   
$$x_{s-1} \leftarrow \frac{1}{s} \cdot \sum_{i=1}^s Xl_{i-1, s-i} \quad \text{if } 1 \leq s \leq \tau$$

|
|   
$$x_{s-1} \leftarrow \frac{1}{\tau} \cdot \sum_{i=1}^{\tau} Xl_{i-1, s-i} \quad \text{if } \tau \leq s \leq n$$

|
|   
$$x_{s-1} \leftarrow \frac{1}{N-s+1} \cdot \sum_{i=1}^{N-s+1} Xl_{i+s-n-1, n-i} \quad \text{if } n \leq s \leq N$$

|
| for i ∈ 0..N - τ
|   for j ∈ 0..τ - 1
|     vosti,j ← xi+j
|
| vost

```

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----|--------|--------|--------|--------|--------|--------|--------|--------|
| 0 | 8.005 | 9.773 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 |
| 1 | 9.773 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 |
| 2 | 12.34 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 |
| 3 | 12.401 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 |
| 4 | 15.291 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 |
| 5 | 13.697 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 |
| 6 | 16.842 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 |
| 7 | 16.216 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 |
| 8 | 16.366 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 |
| 9 | 17.589 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 |
| 10 | 21.954 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 |
| 11 | 19.476 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 |
| 12 | 20.036 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 |
| 13 | 23.127 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 |
| 14 | 24.407 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 | 32.824 |
| 15 | 23.665 | 25.803 | 25.228 | 29.133 | 29.08 | 31.504 | 32.824 | ... |

trend1 := diag(trend) =

ПРИЛОЖЕНИЕ Б. Результат сглаживания с использованием метода гусеница и метода скользящего среднего

Результаты оценивания тренда модели $f(t) = \sin(t \cdot 0.1)$ приведены на рисунках оценивания сигма $\delta = 0.1$

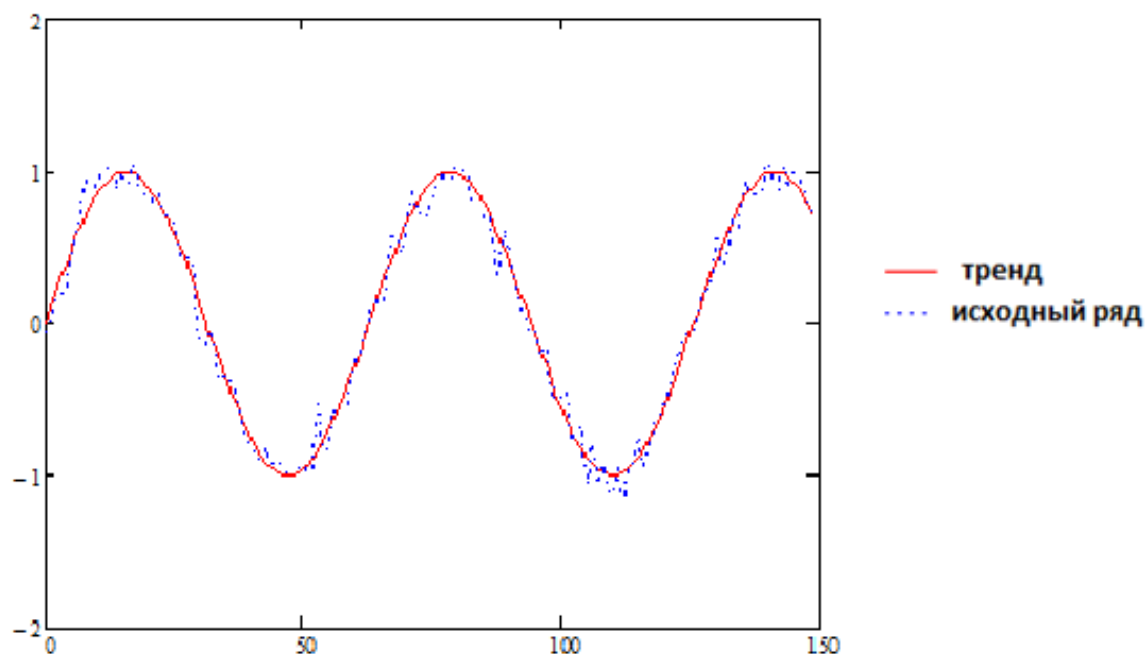


Рисунок 1— Тренд и зашумленный временной ряд

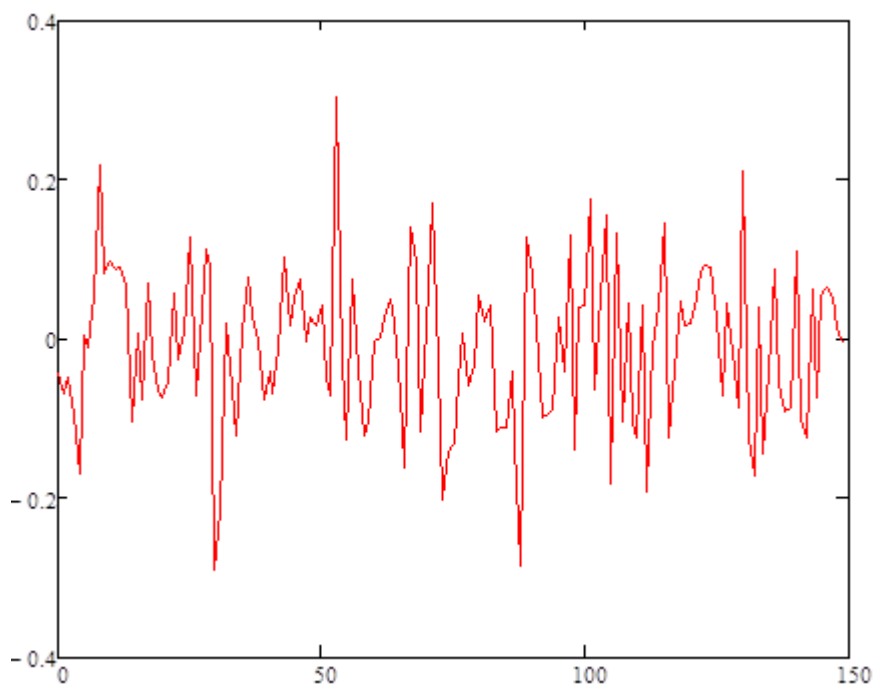


Рисунок 2 – динамика шума

ПРИЛОЖЕНИЕ Б

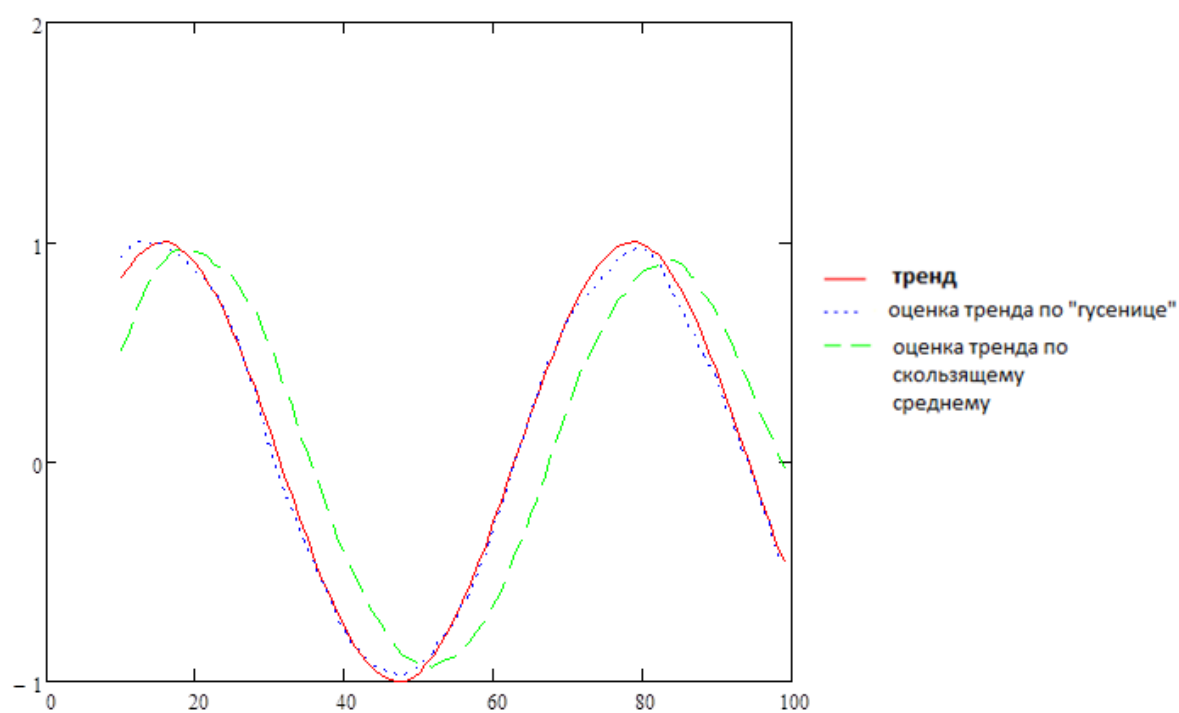


Рисунок 3 – Тренд, оценки тренда

Аналогичные результаты получены для $\delta = 0.3$

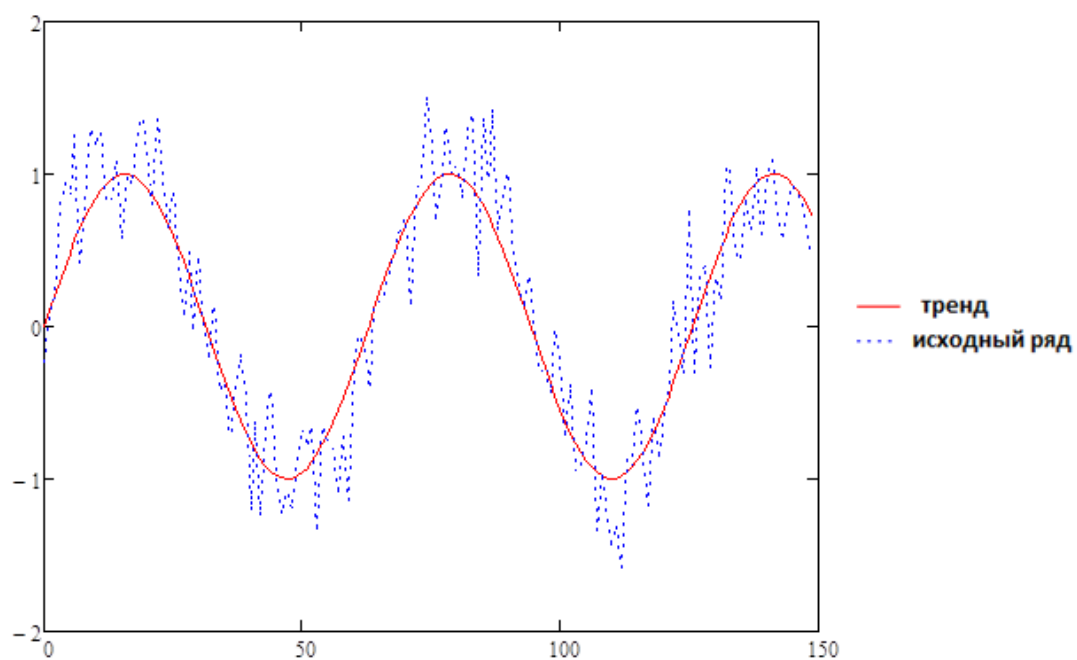


Рисунок 4– Тренд и зашумленный временной ряд

ПРИЛОЖЕНИЕ Б

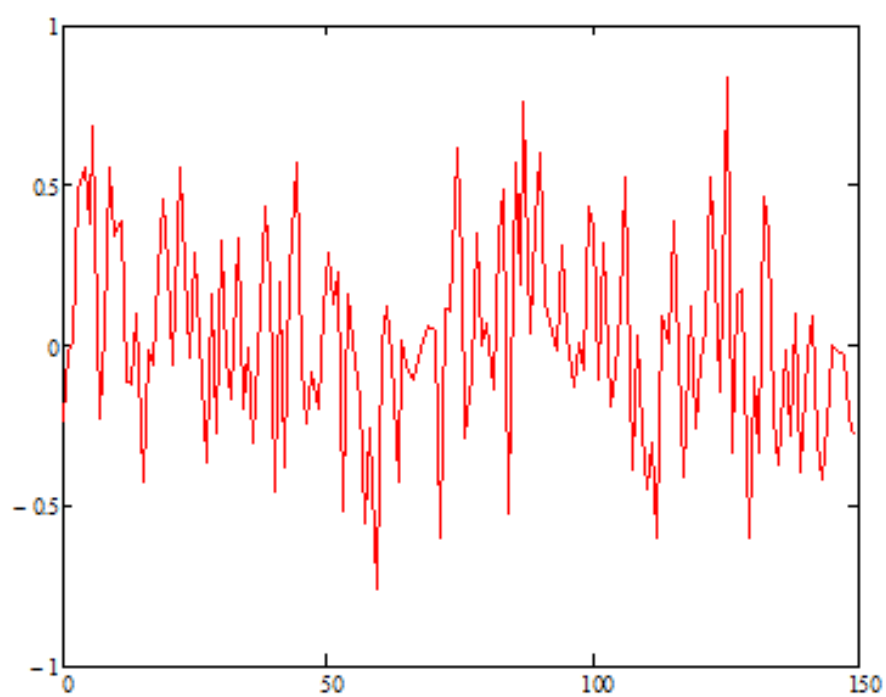


Рисунок 5 – динамика шума

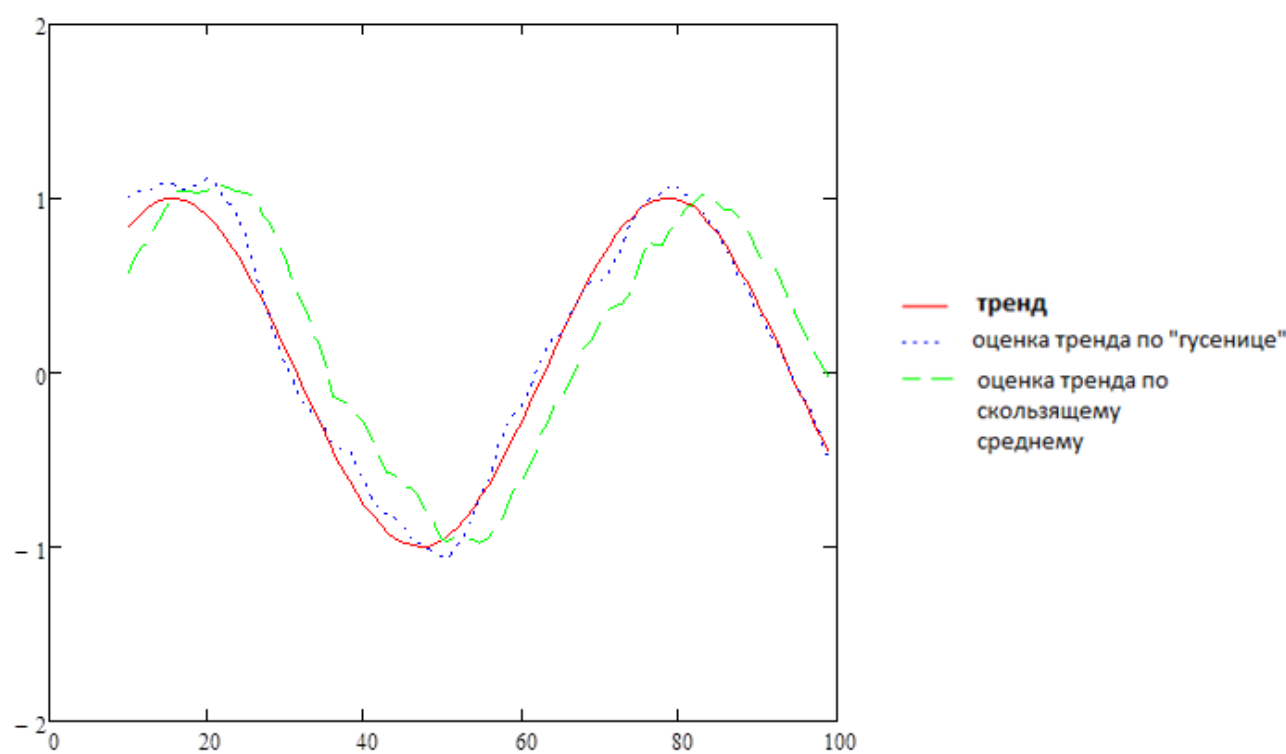


Рисунок 6 – Тренд, оценки тренда

ПРИЛОЖЕНИЕ Б

В таблице 1 приведены результаты MAPE

| Среднеквадратическое отклонение | Метод скользящего среднего | Метод гусеница |
|---------------------------------|----------------------------|----------------|
| $\delta = 0.1$ | 133,3% | 10,9% |
| $\delta = 0.3$ | 135,9% | 27,07% |

$$f(t) = \exp(t \cdot 0.01)$$

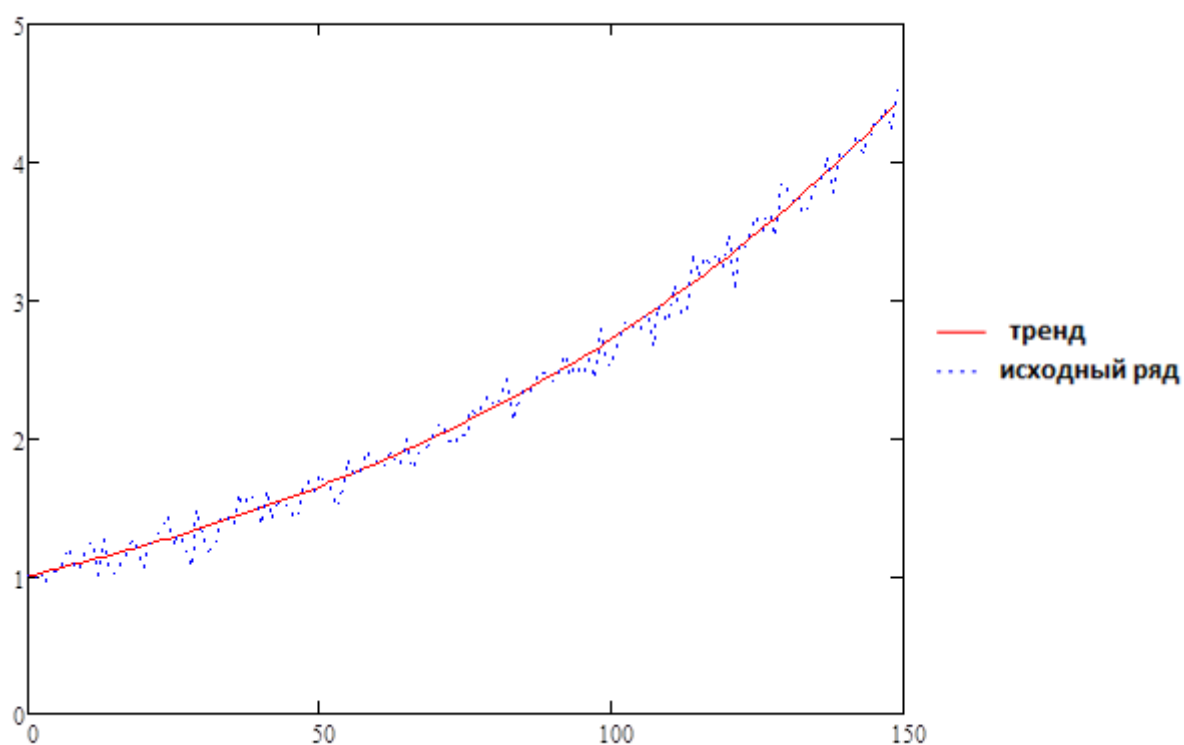


Рисунок 7– Тренд и зашумленный временной ряд

ПРИЛОЖЕНИЕ Б

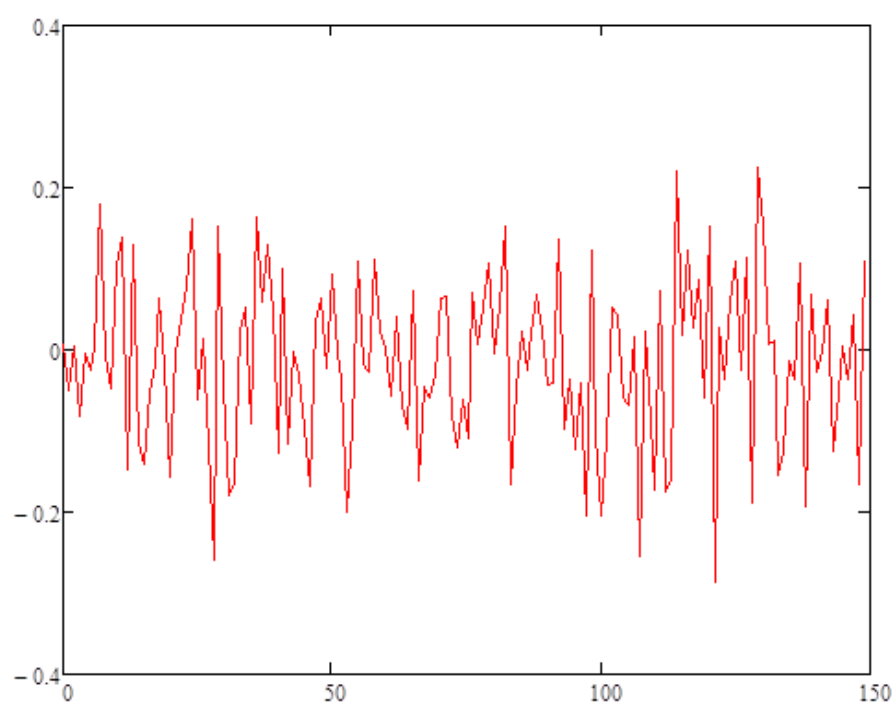


Рисунок 8 – Динамика шума

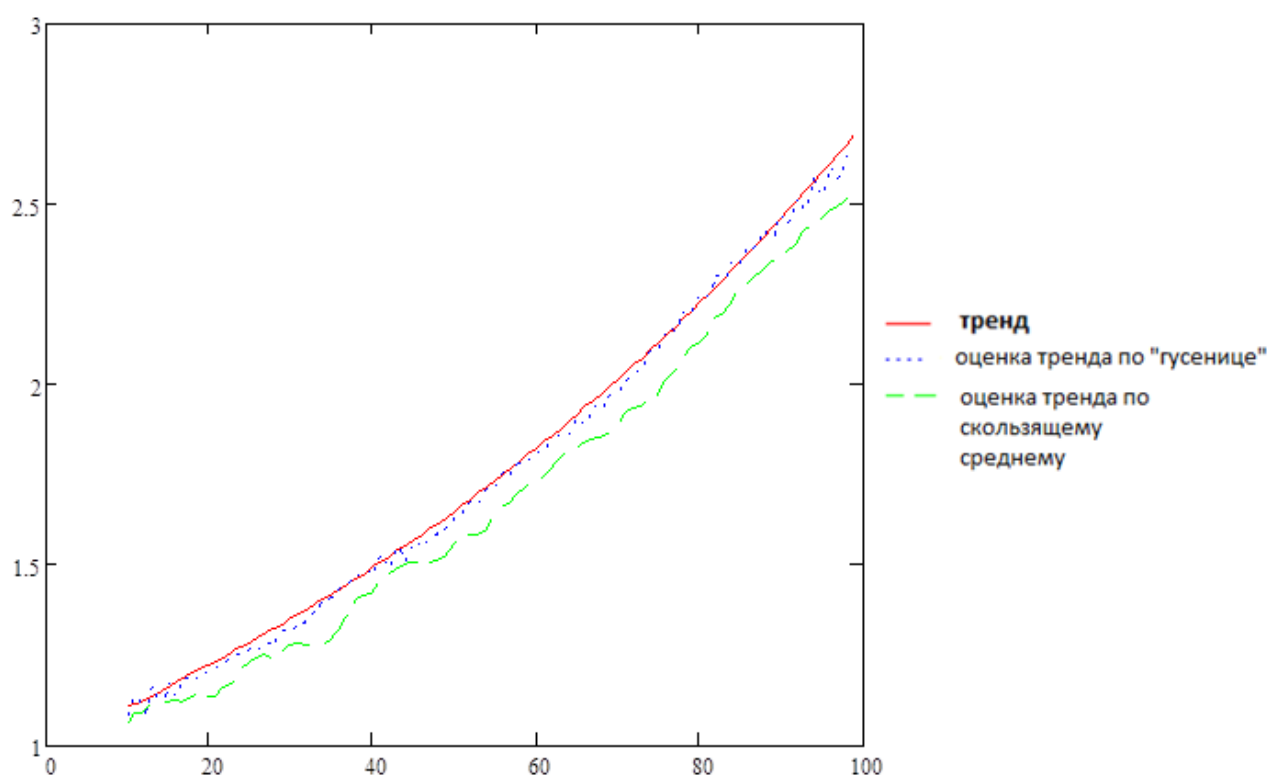


Рисунок 9 – Тренд, оценки тренда

ПРИЛОЖЕНИЕ Б

Аналогичные результаты получены для $\delta = 0.3$

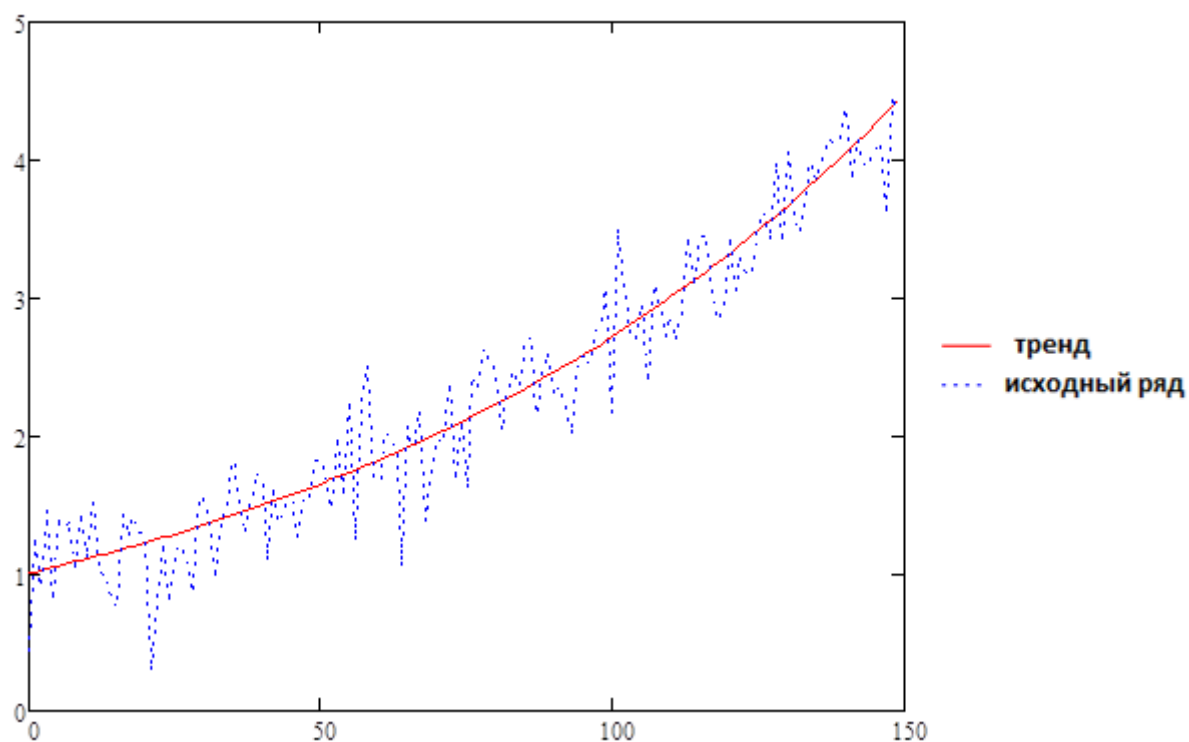


Рисунок 10– Тренд и зашумленный временной ряд

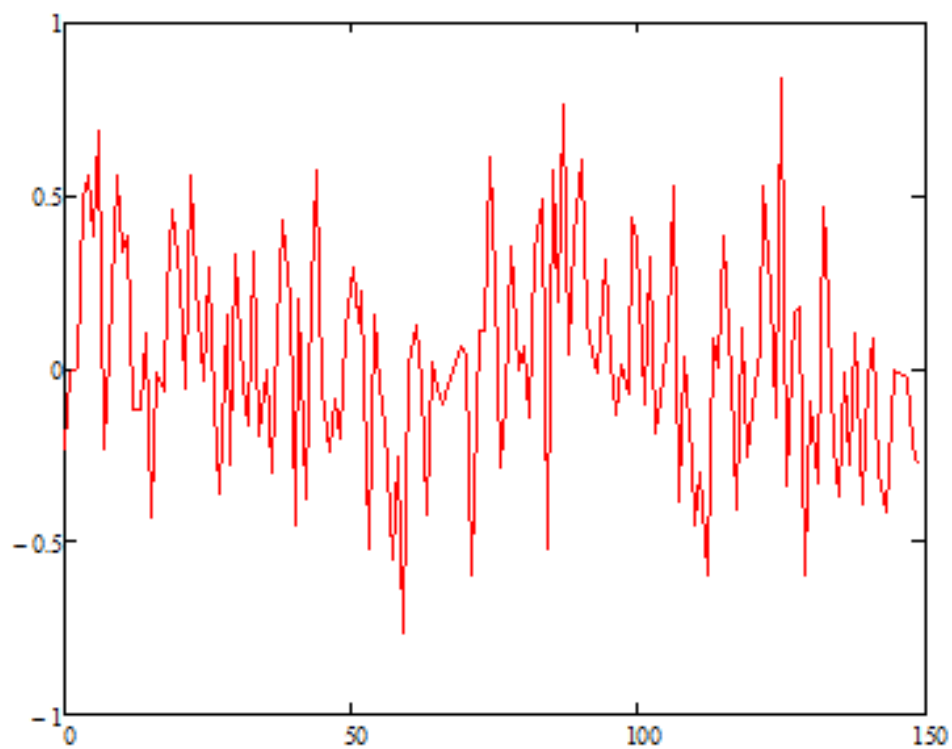


Рисунок 11 – Динамика шума

ПРИЛОЖЕНИЕ Б

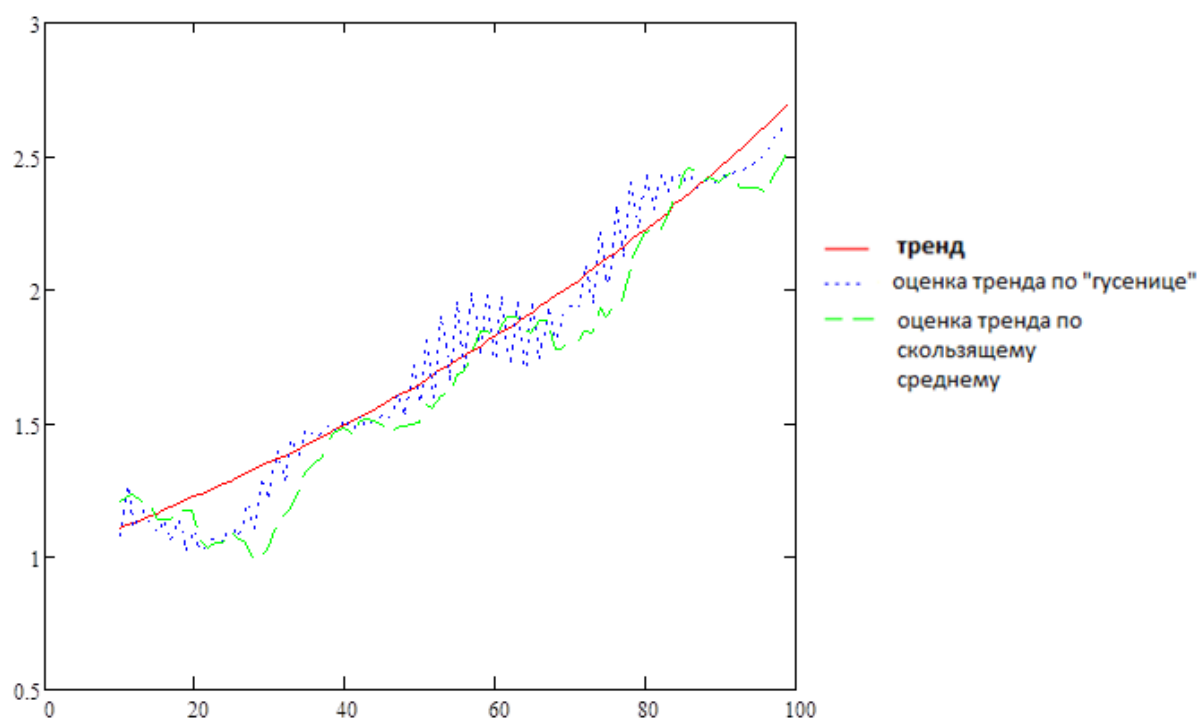


Рисунок 12 – Тренд и зашумленный временной ряд

В таблице 2 приведены результаты MAPE

| Среднеквадратическое отклонение | Метод скользящего среднего | Метод гусеница |
|---------------------------------|----------------------------|----------------|
| $\delta = 0.1$ | 5% | 2,1% |
| $\delta = 0.3$ | 6,89% | 5,75% |

ПРИЛОЖЕНИЕ В. Прогнозирование временной последовательности

| | | | | | | |
|-------|-------|--------------------|--------|--------------------|--------|--------|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 | 77.85 | $3.005 \cdot 10^3$ | 154.8 | $2.71 \cdot 10^3$ | 101.6 | 329.75 |
| 1 | 78.1 | $2.96 \cdot 10^3$ | 155.9 | $2.696 \cdot 10^3$ | 109.91 | 330 |
| 2 | 78.69 | $3.07 \cdot 10^3$ | 155.5 | $2.707 \cdot 10^3$ | 106.95 | 329.5 |
| 3 | 80.56 | $2.932 \cdot 10^3$ | 156.8 | $2.671 \cdot 10^3$ | 106.9 | 322.1 |
| 4 | 81.8 | $2.844 \cdot 10^3$ | 154.6 | $2.662 \cdot 10^3$ | 105.98 | 322.6 |
| 5 | 80.51 | $2.816 \cdot 10^3$ | 153.85 | $2.7 \cdot 10^3$ | 104.45 | 322.65 |
| 6 | 81 | $2.832 \cdot 10^3$ | 153.8 | $2.674 \cdot 10^3$ | 104.4 | 319 |
| S = 7 | 78.9 | $2.873 \cdot 10^3$ | 155.6 | $2.643 \cdot 10^3$ | 105 | 324.6 |
| 8 | 79.87 | $2.936 \cdot 10^3$ | 154.4 | $2.586 \cdot 10^3$ | 105.99 | 323.5 |
| 9 | 78.5 | $2.894 \cdot 10^3$ | 153.95 | $2.629 \cdot 10^3$ | 105.99 | 329 |
| 10 | 78.4 | $2.92 \cdot 10^3$ | 153.1 | $2.63 \cdot 10^3$ | 105.78 | 336.5 |
| 11 | 76.89 | $2.9 \cdot 10^3$ | 152.9 | $2.699 \cdot 10^3$ | 104.99 | 346 |
| 12 | 77.48 | $2.878 \cdot 10^3$ | 152.4 | $2.752 \cdot 10^3$ | 105 | 351 |
| 13 | 76.04 | $2.854 \cdot 10^3$ | 153.2 | $2.8 \cdot 10^3$ | 103.99 | 341.75 |
| 14 | 75 | $2.857 \cdot 10^3$ | 152.9 | $2.733 \cdot 10^3$ | 103.54 | 334.95 |
| 15 | 75.16 | $2.872 \cdot 10^3$ | 154.15 | $2.796 \cdot 10^3$ | 101.64 | ... |

$$r_{i,j} := \frac{S_{i,j} - S_{i-1,j}}{S_{i-1,j}}$$

$$\text{gank}(f, mm, N) := \begin{array}{l} \text{for } j \in 0..N - mm \\ \quad \text{for } i \in 0..mm - 1 \\ \quad \quad X_{i,j} \leftarrow f_{i+j} \\ \quad X \end{array}$$

ПРИЛОЖЕНИЕ В

```

predscazanieq (V,mm,N,kkk,q) :=
    vz ← submatrix(V,0,mm-2,0,kkk-1)
    v ← submatrix(V,mm-1,mm-1,0,kkk-1)
    rab ← vzT · vz
    rab1 ← rab-1
    rab2 ← v · rab1
    rav3 ← vzT · q
    oo ← rab2 · rav3
    oo

```

```

diagonal(a,N,L) :=
    n ← N - L
    for i1 ∈ 0..L-1
        fi1 ←  $\frac{1}{i1+1} \cdot \sum_{r=0}^{i1} a_{r,i1-r}$ 
    for i2 ∈ L..n
        fi2 ←  $\frac{1}{L} \cdot \sum_{r=0}^{L-1} a_{r,i2-r}$ 
    for i3 ∈ n+1..N-1
        fi3 ←  $\frac{1}{N-i3} \cdot \sum_{r=0}^{N-i3-1} a_{r+i3-n,n-r}$ 
    f

```

```

trendex(f,N,L,k) :=
    nn ← N - L
    X ← gank(f,L,N)
    SS ← X · XT
    U ← eigenvcs(SS)
    lam ← eigenvals(SS)
    for i ∈ 0..L-1
        XSi ← U⟨i⟩ · U⟨i⟩T · X
    tr ← sX(k,XS)
    trfd ← diagonal(tr,N,L)
    (trfd U lam XS)

```

ПРИЛОЖЕНИЕ В

```

predscazanieq (V,mm,N,kkk,q) :=
| vz ← submatrix(V,0,mm-2,0,kkk-1)
| v ← submatrix(V,mm-1,mm-1,0,kkk-1)
| rab ← vzT · vz
| rab1 ← rab-1
| rab2 ← v · rab1
| rav3 ← vzT · q
| oo ← rab2 · rav3
| oo

```

```

genpredsk (f,N,L,k) :=
| (trfd U lam XS) ← trendex(f,N,L,k)
| X ← gank(f,L,N)
| q ← submatrix(X,1,L-1,N-L,N-L)
| rez ← predscazanieq (U,L,N,k,q)

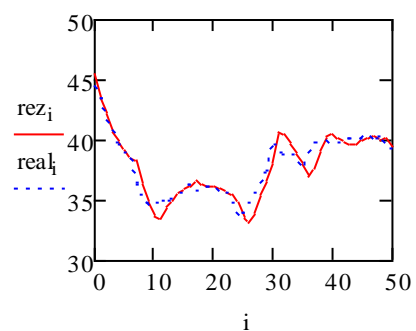
```

```

ngenpredsk (f,N,L,k,np,nn) :=
| for i ∈ 0..np
|   | fl ← submatrix(f,nn+i,nn+N-1+i,0,0)
|   | rezi ← genpredsk (fl,N,L,k)
| rez

```

$$\left[\sum_{j=0}^{np} \left(\frac{|rez_j - real_j|}{real_j} \right) \right] \cdot \frac{100}{np} = 1.876$$



ПРИЛОЖЕНИЕ Г. Построение прогноза на модельных данных

$$f(t) = \exp(t \cdot 0.01)$$

$$f1(t) = f(t) + z(t), \text{ где } z(t) - \text{шум.}$$

Результаты предсказания этой модели для двух значений сигма приведены на рисунках.

Приведен результат оценивания для $\sigma = 0.1$

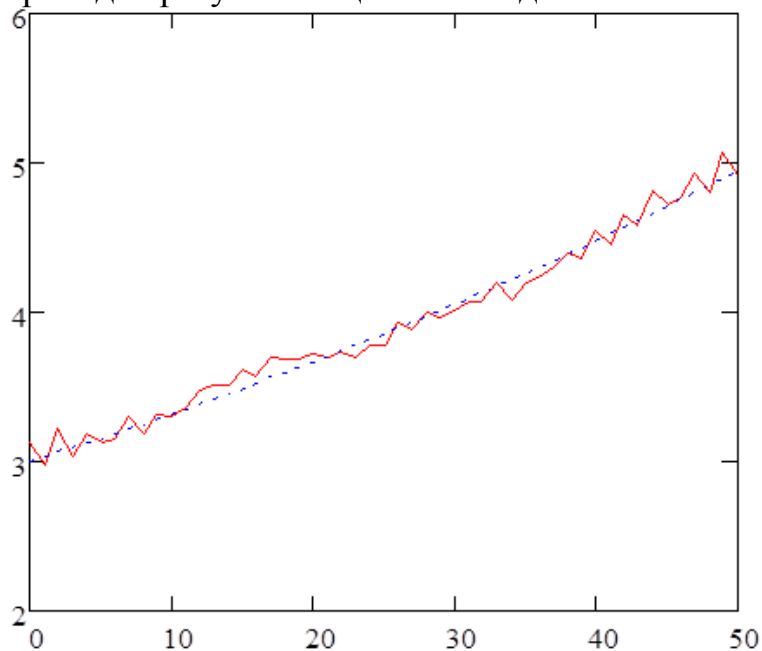


Рисунок 13 – Зашумленный временной ряд и предсказанное значение ряда

Аналогичные результаты получены для $\delta = 0.3$

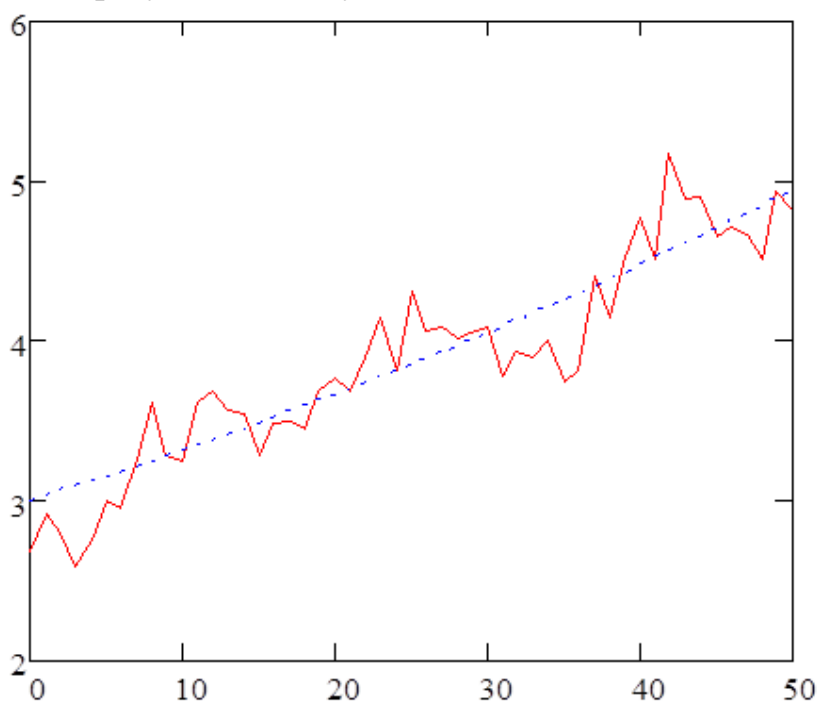


Рисунок 14 – Зашумленный временной ряд и предсказанное значение ряда

ПРИЛОЖЕНИЕ Г

Результаты предсказания модели $f(t) = \sin(t \cdot 0.1)$ для двух значений сигма приведены на рисунках.

Приведен результат оценивания для $\sigma = 0.1$

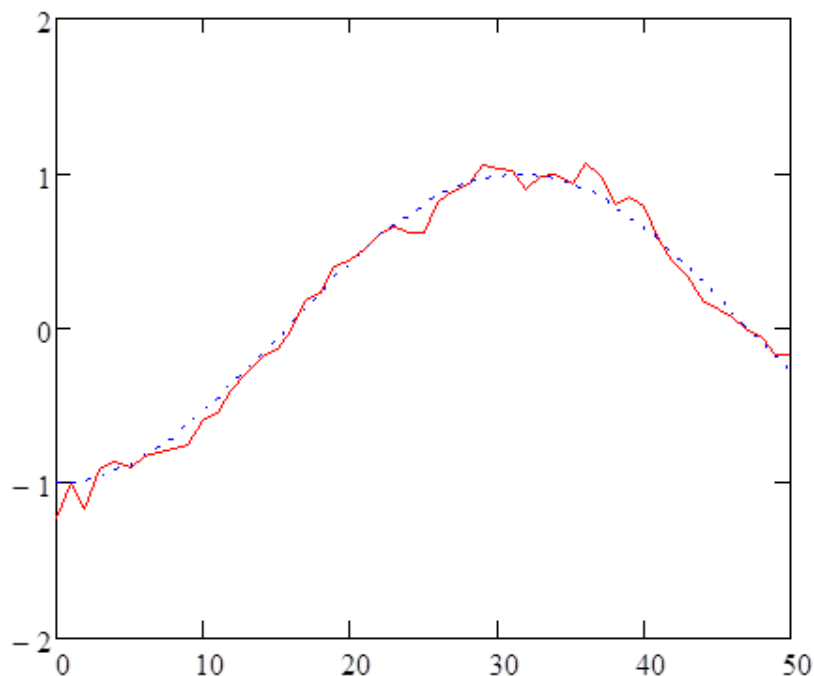


Рисунок 15 – Зашумленный временной ряд и предсказанное значение ряда

Аналогичные результаты получены для $\delta = 0.3$

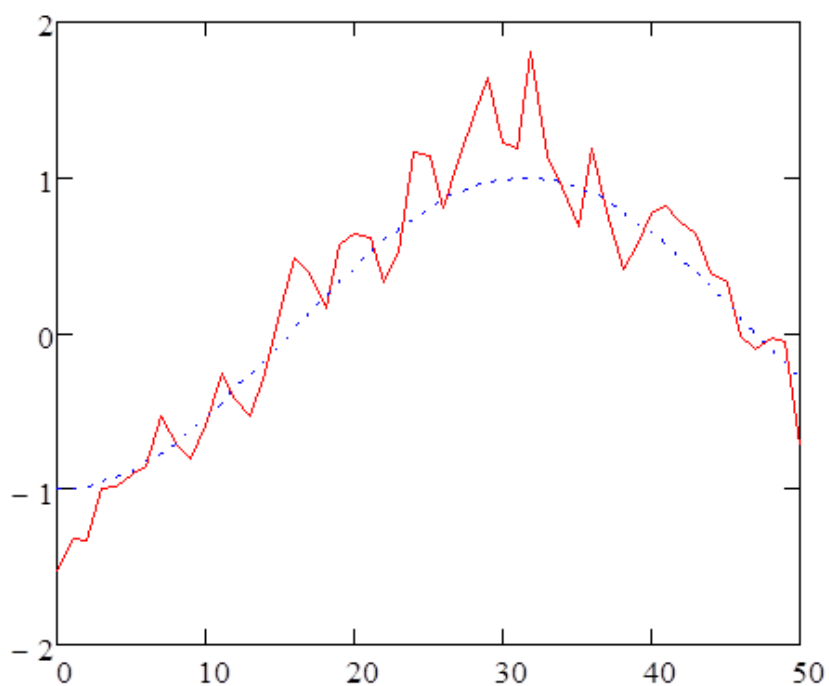


Рисунок 16 – Зашумленный временной ряд и предсказанное значение ряда

Уважаемый пользователь! Обращаем ваше внимание, что система «Антиплагиат» отвечает на вопрос, является ли тот или иной фрагмент текста заимствованным или нет. Ответ на вопрос, является ли заимствованный фрагмент именно плагиатом, а не законной цитатой, система оставляет на ваше усмотрение.

Отчет о проверке № 1

дата загрузки: 06.06.2016 10:06:58
пользователь: mari.zavgor@mail.ru / ID: 3355970
отчет предоставлен сервисом «Антиплагиат»
на сайте <http://www.antiplagiat.ru>

Информация о документе

№ документа: 5
Имя исходного файла: вкр.docx
Размер текста: 543 кБ
Тип документа: Прочее
Символов в тексте: 21313
Слов в тексте: 2714
Число предложений: 130

Информация об отчете

Дата: Отчет от 06.06.2016 10:06:59 - Последний готовый отчет
Комментарии: не указано
Оценка оригинальности: 81.75%
Заимствования: 18.25%
Цитирование: 0%



Оригинальность: 81.75%
Заимствования: 18.25%
Цитирование: 0%

Источники

| Доля в тексте | Источник | Ссылка | Дата | Найдено в |
|---------------|--|---|------------------|------------------------|
| 8.03% | [1] Текст (2/2) | http://se.math.spbu.ru | 16.04.2016 | Модуль поиска Интернет |
| 6.84% | [2] Загрузить 1,20 МБ | http://mirkin.ru | раньше 2011 года | Модуль поиска Интернет |
| 6.66% | [3] ИСПЫТАНИЕ АЛГОРИТМА МЕТОДА «ГУСЕНИЦА-SSA» ДЛЯ ВОССТАНОВЛЕНИЯ ВРЕМЕННОГО РЯДА | http://cyberleninka.ru | 08.10.2015 | Модуль поиска Интернет |

