МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

В. А. Алексеева

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Учебное пособие
для студентов, обучающихся по направлению 01.03.04

Ульяновск
УлГТУ
2020

УДК 330.43(75)
ББК 65в6я73
А 47

Рецензенты:

кафедра «Телекоммуникационные технологии и сети» Ульяновского государственного университета (зав. кафедрой, д.т.н., профессор А.А. Смагин);

зав. кафедрой информационной безопасности и теории управления Ульяновского государственного университета, д.ф.-м.н., профессор А.С. Андреев

Утверждено редакционно-издательским советом университета в качестве учебного пособия

Алексеева, Венера Арифзяновна

А 47 Анализ временных рядов : учебное пособие / В. А. Алексеева. – Ульяновск : УлГТУ, 2020. – 147 с.

ISBN 978-5-9795-2091-9

Содержание учебного пособия включает основные методы анализа, моделирования и прогнозирования временных рядов. Представлены примеры решения задач, которые могут быть использованы для подготовки и выполнения контрольных и практических заданий, лабораторных работ по курсу «Анализ временных рядов». Приведены упражнения для самостоятельного выполнения.

Пособие предназначено для студентов технических вузов дневной формы обучения.

УДК 330.43(75)
ББК 65в6я73

© Алексеева В. А., 2020

ISBN 978-5-9795-2091-9 © Оформление. УлГТУ, 2020

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ 6

РАЗДЕЛ 1. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ 7

1. Понятие временного ряда 7

2. Составляющие временного ряда 8

3. Предварительный анализ временных рядов 9

3.1. Тест на стационарность ВР 10

3.2. Автокорреляционная функция....15

3.3. Проверка наличия тренда....17

3.4. Расчет показателей, характеризующих тенденцию динамики....20

3.5. Сглаживание временных рядов 22

3.6. Выделение периодических составляющих временного ряда....26

3.7. Фрактальный анализ временных рядов 30

Контрольные вопросы 32

РАЗДЕЛ 2. МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ....34

1. Построение моделей тренда 34

1.1. Простая линейная регрессия....35

1.2. Квадратичный тренд....36

1.3. Экспоненциальный тренд 36

1.4. Степенной тренд 37

1.5. Гиперболический тренд....38

2. Моделирование периодических колебаний временного ряда 38

3. Моделирование случайной составляющей временного ряда....40

3.1. Модели авторегрессии....40

3.2. Модели скользящего среднего 41

3.3. Модели авторегрессии-скользящего среднего....41

3.4. Авторегрессионная модель с условной гетероскедастичностью....41

Контрольные вопросы 44
РАЗДЕЛ 3. АНАЛИЗ КАЧЕСТВА МОДЕЛЕЙ И ДИАГНОСТИКА ОСТАТКОВ.....46
1. Анализ качества моделей 46
2. Диагностика остатков.....49
Контрольные вопросы 51
РАЗДЕЛ 4. АНАЛИЗ МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ.....52
1. Взаимные характеристики двух временных рядов 52
2. Взаимный спектральный анализ временных рядов 54
3. Кросс-корреляция.....57
Контрольные вопросы 58
ПРАКТИЧЕСКАЯ ЧАСТЬ 59
1. Примеры решения задач по анализу временных рядов 59
2. Задачи для самостоятельного выполнения.....81
МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ.....91
Лабораторная работа № 1. Анализ временных рядов в MS Excel. 91
Лабораторная работа № 2. Предварительный анализ временных рядов в системе Statistica..... 107
Лабораторная работа № 3. Модели авторегрессии, скользящего среднего и ARIMA..... 112
Лабораторная работа № 4. Спектральный (Фурье) анализ и кросс-спектральный анализ..... 117
Лабораторная работа № 5. Программный комплекс АС ДРМ. Предварительный анализ временных рядов 125
Лабораторная работа № 6. Программный комплекс АС ДРМ. Моделирование временных рядов..... 130

Лабораторные работы № 7, 8, 9. Программная реализация методов анализа, моделирования, прогнозирования временных рядов, критериев качества моделей и диагностики остатков 139
Приложение 1 141
Приложение 2 142
БИБЛИОГРАФИЧЕСКИЙ СПИСОК 145

ВВЕДЕНИЕ

Цель создания данного учебного пособия – помочь студентам, обучающимся по направлению 01.03.04 «Прикладная математика», получить более широкое представление о методах анализа, математического моделирования и прогнозирования явлений и процессов, изменяющихся во времени. В учебном пособии подробно излагаются методы распознавания типа тренда, алгоритмы расчета линейных и нелинейных трендов. Описываются различные модели сезонных колебаний, гармонические модели, модели авторегрессии, скользящего среднего, авторегрессии-скользящего среднего. Рассматриваются различные меры качества построенных моделей и методы диагностики остатков.

Учебное пособие состоит из 4 разделов: предварительный анализ временных рядов, моделирование временных рядов, анализ качества моделей временных рядов и диагностика остатков, анализ многомерных временных рядов.

В пособии содержатся необходимые теоретические сведения для освоения дисциплины, приводятся примеры решения практических задач и задания для самостоятельного выполнения. Также пособие содержит задания и методические рекомендации для выполнения лабораторных работ по дисциплине «Анализ временных рядов».

Учебное пособие по дисциплине «Анализ временных рядов» способствует формированию у студентов практических навыков по построению и анализу стохастических моделей различных явлений и процессов, изменяющихся во времени.

РАЗДЕЛ 1. ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

1. Понятие временного ряда

Ряд наблюдений $y(t_1)$, $y(t_2)$,…$y(t_N)$ анализируемой случайной величины $\xi(t)$, произведенных в последовательные моменты времени t_1, t_2,…, t_N, называется *временным рядом* (ВР), то есть ВР есть упорядоченная во времени последовательность наблюдений, характеризующих уровни развития изучаемого явления в последовательные моменты или периоды времени. В литературе также встречаются термины *динамический ряд*, *ряд динамики*, в англоязычной литературе принят термин *time series*.

Величины $y(t_i)$ называются *уровнями ряда*, а t_i – временными метками (моменты или интервалы наблюдения). Целью исследования ВР является выявление закономерностей в изменении уровней ряда и построении его модели в целях прогнозирования и исследования взаимосвязей между явлениями.

Классификация ВР:

1) по количеству анализируемых характеристик процесса или явления ВР подразделяются на *одномерные* (фиксированная числовая характеристика) и *многомерные* (рассматриваются несколько характеристик исследуемого процесса или явления);

2) по времени ВР подразделяются на *моментные* и *интервальные*. Интервальный ряд – последовательность, в которой уровень явления относят к результату, накопленному или вновь произведенному за определенный интервал времени. Моментный ряд – последовательность, в которой уровень ряда характеризует изучаемое явление в конкретный момент времени;

3) по расстоянию промежутку времени между последовательными наблюдениями ряды бывают *равномерные* (наблюдения через равные промежутки времени) и *неравномерные* (промежутки времени между наблюдениями изменяются);

4) по постоянству во времени ВР бывают *стационарными* и *нестационарными*. ВР называется строго стационарным (или стационарным в узком смысле), если совместное распределение вероятностей m наблюдений $Y(t)=\{y(t_1), y(t_2),\ldots, y(t_m)\}$ такое же, как и для m наблюдений $y(t_1+\tau), y(t_2+\tau),\ldots, y(t_m+\tau)$ при любых τ, t_1, $t_2,\ldots, t_N$ и m, т.е. если его свойства не меняются при изменении начала отсчета. Ряд $y(t)$ называется стационарным в широком смысле, если математическое ожидание, дисперсия ВР не зависят от t.

2. Составляющие временного ряда

Пусть (Ω, F, P) – вероятностное пространство, на котором задан стационарный процесс $Y(t)$, наблюдаемый в равноотстоящие моменты времени $t_1, t_2, \ldots, t_N$:

$$Y(t) = f(t) + \varphi(t) + \psi(t) + \varepsilon(t), \quad t = \overline{t_1, t_N}, \qquad (1.1)$$

где $Y(t_1)$, $Y(t_2)$, …, $Y(t_N)$ – ряд наблюдений случайной функции $\xi(t)$; $f(t)$ – неслучайная (долговременная) функция тренда; $\varphi(t)$ – неслучайная (сезонная) периодическая функция; $\psi(t)$ – неслучайная (долговременная, циклическая) функция; $\varepsilon(t)$ – нерегулярная компонента (случайная величина, ошибка).

Анализ ВР сводится к выделению регулярных компонент $f(t)$, $\varphi(t)$, $\psi(t)$ и описанию нерегулярной ее части $\varepsilon(t)$.

Как правило, наличие той или иной составляющей можно определить с помощью визуального анализа графика временного ряда.

Тренд, или тенденция $f(t)$, представляет собой устойчивую закономерность, наблюдаемую в течение длительного периода времени, например, возрастание или убывание исследуемой характеристики. Сезонная компонента $\varphi(t)$ связана с наличием факторов, действующих с заранее известной периодичностью менее 1 года. Циклическая компонента $\psi(t)$ – неслучайная функция, описывающая длительные периоды (более 1 года). Случайная компонента $\varepsilon(t)$ – это составная часть временного ряда, оставшаяся после выделения систематических компонент. Эта компонента отражает воздействие многочисленных факторов случайного характера и может иметь разнообразную структуру, начиная от простейшей в виде «белого шума» до весьма сложных, описываемых моделями авторегрессии — скользящего среднего.

Модель ВР, называемая мультипликативной моделью, представляется как произведение перечисленных компонент:

$$Y(t) = f(t) \cdot \varphi(t) \cdot \psi(t) + \varepsilon(t), \; t = \overline{t_1, t_N}. \qquad (1.2)$$

Перед построением модели исходные данные проверяются на сопоставимость (уровни ВР должны иметь одинаковые единицы измерения, шаг наблюдений, интервал времени, методику расчета, элементы, относящиеся к неизменной совокупности), однородность (отсутствие случайных выбросов, аномальных наблюдений, которые искажают результаты моделирования), устойчивость (наличие закономерности в изменении уровней ряда) и достаточность (число наблюдений должно в 5–15 превосходить число параметров модели).

3. Предварительный анализ временных рядов

В ходе предварительного анализа для получения общего представления о динамике исследуемого показателя целесообразно построить его график. К процедурам предварительного анализа

относятся тест на стационарность, проверка наличия тренда, сглаживание временных рядов, построение автокорреляционной функции, вычисление показателей динамики ВР, методы выявления периодических компонент ВР, в частности, спектральный и вейвлет-анализы.

3.1. Тест на стационарность ВР

При анализе случайных процессов задача проверки стационарности изучаемого процесса занимает важное место. Суждение о стационарности позволяет исследователю уточнить модель процесса или выявить причины нарушения стабильности хода эксперимента.

Тест на стационарность осуществляется путем проверки гипотез о постоянстве среднего, определяемого непараметрическим критерием сдвига, и о постоянстве дисперсии процесса или его остатков, определяемого критерием рассеяния.

В случае опровержения гипотезы о постоянстве среднего процесса проводится процедура удаления тренда. Если же процесс имеет ненулевое постоянное среднее, то он сглаживается процедурой удаления среднего. При любом исходе процесс центрируется относительно нуля.

Если при проверке постоянства дисперсии процесса или его остатков после удаления непериодического тренда выявлено нарушение постоянства среднего квадрата, тогда процесс относится к классу нестационарных процессов, причем первичным является предположение о том, что его нестационарность обусловлена только изменением дисперсии с течением времени.

В случае принятия гипотезы о постоянстве дисперсии рекомендуется проверить процесс на стационарность по критерию

Пирсона или любому другому критерию, позволяющему выявить нестационарность.

Проверка постоянства среднего значения. Существует несколько алгоритмов проверки постоянства среднего значения процесса. К ним относятся непараметрический критерий сдвига и критерий инверсий.

Непараметрический критерий сдвига. Во всех непараметрических методах не предполагается, что исследователь знает распределение исследуемого процесса. Нулевая гипотеза обычно состоит в предположении, что две сравниваемые выборки характеризуются одним и тем же распределением, причем безразлично каким. Пусть имеется по N независимых отсчетов с первого и второго участков записи, или одна реализация разбита на две одинаковые группы по N измерений: $X=(x_1, x_2,\ldots, x_N)$ и $Y = (y_1, y_2,\ldots, y_N)$. В последовательности Y ищется наибольшее значение y_{max} и подсчитывается число s отсчетов x_i, превышающих y_{max}. Если это число s превышает критическое значение $Z_{критич}$, то для выбранной доверительной вероятности $P_{дов}$ делается вывод о том, что в реализации присутствует сдвиг среднего вверх. Если этот вид нестационарности не обнаружен, следует проверить число значений y_i, превышающих x_{max}, что характеризует сдвиг среднего вниз. Критические значения $Z_{критич}$ равны:

для $N < 15$ $Z_{критич} = 4$ при $P_{дов} = 0.95$; $Z_{критич} = 6$ при $P_{дов} = 0.99$;

для $N > 15$ $Z_{критич} = 5$ при $P_{дов} = 0.95$; $Z_{критич} = 7$ при $P_{дов} = 0.99$.

Критерий инверсий. Критерий строится на понятии инверсии, которая заключается в том, что в последовательности чисел за большим следует меньшее. Если отсчеты отражают стационарный процесс, то число инверсий I независимо от распределения наблюдаемого процесса обладает известным распределением. Так, величина γ, определяемая формулой

$$\gamma = 1 - \frac{4I}{N(N-1)}, \tag{1.3}$$

имеет среднее значение, равное нулю, и дисперсию, равную:

$$\mathrm{D}_\gamma = \frac{2(2N-5)}{9N(N-1)}. \tag{1.4}$$

Величина γ, уже начиная с $N > 10$, имеет распределение, близкое к нормальному. Поэтому для проверки нулевой гипотезы о том, что дрейфа среднего значения нет, можно использовать неравенство

$$\left| \frac{\gamma}{\sqrt{\mathrm{D}_\gamma}} \right| < \mathrm{Z}_\Phi \left(\frac{1 + \mathrm{P}_\Д}{2} \right). \tag{1.5}$$

Критерий инверсий является одним из наиболее эффективных непараметрических критериев, применим как к последовательности отсчетов процесса, так и к последовательностям его характеристик.

Проверка постоянства дисперсии. К алгоритмам проверки постоянства дисперсии процесса относятся критерий Кохрена и критерий рассеяния.

Критерий Кохрена. Данный критерий применяется для выяснения вопроса о том, являются ли несколько оценок дисперсий $\widetilde{\sigma}_i^2$ принадлежащими одному и тому же нормальному распределению. Если по нескольким участкам реализации вычислены оценки дисперсий $\widetilde{\sigma}_i^2$ и вычисленная по формуле

$$\mathrm{G} = \frac{\widetilde{\sigma}_{\mathrm{i}}^2 \, max}{\sum\limits_{\mathrm{i}=1}^{\mathrm{K}} \widetilde{\sigma}_{\mathrm{i}}^2}, \tag{1.6}$$

статистика G превышает критическое значение $Z_G(\nu, K, P_{дов})$, где $\nu = N – 1$, N – объем каждой выборки, K – число участков (выборок), то предположение о стационарности следует отвергнуть.

Следует отметить, что данный критерий справедлив при предположении о нормальности и очень чувствителен к нормальности исходного распределения, и его применение может приводить к серьезным ошибкам, если предположение о нормальности не справедливо.

Критерий рассеяния. Используются отсчеты двух выборок по N измерений: $X = (x_1, x_2,\ldots, x_N)$ и $Y = (y_1, y_2,\ldots, y_N)$. По второй последовательности Y находится наибольший отсчет y_{max} и наименьший y_{min}. Затем подсчитывается число s отсчетов x_i, превышающих y_{max} или меньших, чем y_{min}. Нулевая гипотеза, состоящая в предположении равенства дисперсий для первого и второго участков реализации, принимается, если число s меньше критического значения $Z_{критич}$ с выбранной доверительной вероятностью $P_{дов}$. Критические значения $Z_{критич}$ приведены в табл.1. Объемы каждой выборки N целесообразно выбирать равными для обоих выборок.

Таблица 1

Критические значения критерия рассеяния

$P_{дов} = 0.95$		$P_{дов} = 0.99$	
Диапазон N	$Z_{критич}$	Диапазон N	$Z_{критич}$
Не более 6	5	Не более 10	7
От 7 до 25	6	От 11 до 20	8
От 26 до 50	7	От 21 до 50	9
От 51	8	От 51	10

Кроме того, для анализа постоянства дисперсий может применяться критерий инверсий. Для этого определяется стационарность ряда оценок $\tilde{\sigma}_i^2$ или $\tilde{\sigma}_i$, по результатам вычислений на нескольких интервалах.

Проверка стационарности. Для проверки стационарности распределений может применяться следующая логическая схема: из реализации извлекаются две или более выборки, состоящие из отсчетов x_i, y_i, …, снятых на нескольких участках записи. По полученным отсчетам рассчитывается статистика критерия, которая сравнивается с критическим значением при выбранной доверительной вероятности.

Для проверки согласия распределений разных выборок применимы критерии согласия эмпирических распределений с теоретическими, поэтому проверку стационарности распределений основывают на методах проверки согласия распределений.

Критерий χ^2. Пусть имеется n_x отсчетов x_i с первого участка записи,

n_y отсчетов y_i со второго участка и т.д. Общее количество отсчетов равно N. Выдвигается нулевая гипотеза о том, что эти выборки принадлежат одной и той же генеральной совокупности, а отклонения в распределениях для каждого участка объясняются лишь случайным характером процесса.

Использование χ^2 при анализе временных рядов возможно с применением следующего алгоритма: выборка объема N разбивается на L разрядов (интервалов) и параллельно на две группы с равным числом отсчетов $N/2$. Границы разрядов выбираются по каждой группе отдельно и по всей реализации в целом. В идеале предполагается, что если во всей реализации есть s_i элементов из i-го разряда, то на каждом из участков будет по s_i отсчетов этого же интервала. После сортировки каждой группы и всей реализации по интервалам количества отсчетов, попавшие в разряды, можно описать в форме таблицы:

Таблица 2

Разбиение выборки

n_{x1}	n_{x2}	…..	n_{xL}	N_x
n_{y1}	n_{y2}	…..	n_{yL}	N_y
n_1	n_2	…..	n_L	N

По таблице оцениваются вероятности принадлежности отсчетов каждому из участков X и Y, и вычисляется статистика

$$\chi^2 = 2\sum_{j=1}^{L}\frac{1}{n_j}\left[\left(n_{x_j} - \frac{n_j}{2}\right)^2 + \left(n_{y_j} - \frac{n_j}{2}\right)^2\right], \quad (1.7)$$

которая сравнивается с критическим значением $Z_{критич}(v, P_{дов})$, где $v=L-1$. Нулевая гипотеза отвергается при $\chi^2 > Z_{критич}(v, P_{дов})$.

3.2. Автокорреляционная функция

Автокорреляционная функция (АКФ) – последовательность коэффициентов автокорреляции уровней первого и последующих порядков, характеризующих тесноту линейной связи между последующими и предыдущими членами временного ряда, т.е. корреляцию между рядами Y_1, Y_2,…, Y_N и Y_{L+1}, Y_{L+2},…, Y_{L+N}, где L показывает длину временного смещения в ряду (величину лага). Коэффициенты автокорреляции определяются по формуле:

$$r_\tau = \frac{\sum_{t=\tau+1}^{n}(y_t - \bar{y}_{1\tau})(y_{t-\tau} - \bar{y}_{2\tau})}{\sqrt{\sum_{t=\tau+1}^{n}(y_t - \bar{y}_{1\tau})^2 \sum_{t=\tau+1}^{n}(y_{t-\tau} - \bar{y}_{2\tau})^2}}, \quad (1.8)$$

где τ – величина лага,

$\bar{y}_{1\tau} = \frac{\sum_{t=\tau+1}^{n} y_t}{n-\tau}$, $\bar{y}_{2\tau} = \frac{\sum_{t=\tau+1}^{n} y_{t-\tau}}{n-\tau}$.

Функцию $r(\tau) = r_\tau$ называют автокорреляционной функцией временного ряда, а ее график – коррелограммой.

Анализ автокорреляционной функции и коррелограммы позволяет выявить структуру ряда, т. е. определить присутствие в ряде той или иной компоненты. Так, если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка m, то ряд содержит циклические колебания с периодичностью в m моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, то ряд не содержит тенденции и циклических колебаний.

Линейные коэффициенты автокорреляции характеризуют тесноту только линейной связи текущего и предыдущих уровней ряда. Поэтому по коэффициентам автокорреляции можно судить только о наличии или отсутствии линейной (или близкой к линейной) зависимости. Если ряд имеет сильную нелинейную тенденцию, коэффициент автокорреляции может приближаться к нулю. Для проверки ряда на наличие нелинейной тенденции рекомендуется вычислить линейные коэффициенты автокорреляции для временного ряда, состоящего из логарифмов исходных уровней. Отличные от нуля значения коэффициентов автокорреляции будут свидетельствовать о наличии нелинейной тенденции.

Для проверки наличия автокорреляции используется критерий Дарбина – Уотсона (DW):

$$d = \frac{\sum_{t=1}^{N-1}\left(Y_{t+1} - Y_t\right)^2}{\sum_{t=1}^{N} Y_t^2}. \qquad (1.9)$$

Возможные значения критерия находятся в интервале [0;4]. Если ряд не содержит автокорреляцию, то значения критерия колеблются вокруг 2. При значении, близком к 4, говорят о сильной положительной

автокорреляции; если значение стремится к 0, то говорят об отрицательной автокорреляции.

3.3. Проверка наличия тренда

Для проверки наличия тенденции наиболее широко применяются следующие методы.

Метод сравнения средних. Применяется для выявления монотонно возрастающей или монотонно убывающей тенденции. ВР разбивается на две примерно одинаковые части $y_1, y_2, \dots, y_{n_1}$ и $y_{n_1+1}, y_{n_1+2}, \dots, y_{n=n_1+n_2}$ и вычисляются средние ($\bar{y}_1$ и $\bar{y}_2$) и выборочные дисперсии (s_1^2 и s_2^2) для обеих частей соответственно.

Далее рассчитывается значение критерия Стьюдента по формуле:

$$t_{\text{в}} = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{s_1^2}{n_1}} + \sqrt{\frac{s_2^2}{n_2}}}, \tag{1.10}$$

если предполагается, что значения дисперсий на этих участках не равны между собой, и по формуле:

$$t_{\text{в}} = \frac{|\bar{y}_1 - \bar{y}_2|}{s^2} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \tag{1.11}$$

где s^2 – общая выборочная дисперсия ряда, если предполагается, что дисперсии одинаковы.

Нулевая гипотеза о равенстве средних (об отсутствии тенденции) отвергается, если выполняется условие $t_в > t_{кр}(1-\alpha, m)$, где $t_{кр}(1-\alpha, m)$ – табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $m = n_1 + n_2 - 2$.

Метод Фостера-Стюарта. Является более универсальным и дает более надежные результаты. Каждому уровню ряда y_i, начиная с i=2, ставятся в соответствие два значения p_i и q_i:

$$p_i = \begin{cases} 1, & \text{если } y_i \text{ меньше всех предыдущих уровней} \\ 0, & \text{в противном случае,} \end{cases}$$

$$q_i = \begin{cases} 1, & \text{если } y_i \text{ больше всех предыдущих уровней} \\ 0, & \text{в противном случае.} \end{cases}$$

Вычисляется статистика критерия Стьюдента по формуле:

$$t_{\text{в}} = \frac{\sum_{i=2}^{n}(p_i - q_i)}{2\sum_{i=2}^{n}\frac{1}{i}}. \qquad (1.12)$$

Гипотеза об отсутствии тенденции отвергается, если выполняется условие $t_в > t_{кр}(1-\alpha, n-1)$, где $t_{кр}(1-\alpha, n-1)$ – табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $n - 1$.

Критерий серий, основанный на медиане. Представим уровни исследуемого ВР в виде вариационного ряда, т.е. расположим значения ВР в порядке возрастания. Найдем медиану y_m полученного ряда: середину ряда в случае нечетного n, среднее арифметическое двух значений в середине вариационного ряда в случае четного n.

По исходному ВР построим последовательность из плюсов и минусов следующим образом: вместо y_t ставится «+», если $y_t > y_m$, и «-», если $y_t < y_m$ (члены временного ряда, равные y_m, в полученной таким образом последовательности плюсов и минусов не учитываются). Образованная последовательность плюсов и минусов характеризуется общим числом серий $\nu(n)$ и протяженностью самой длинной серии $l(n)$. При этом под «серией» понимается последовательность подряд идущих плюсов и подряд идущих минусов. Если исследуемый ряд состоит из статистически независимых наблюдений, случайно варьирующих около некоторого постоянного уровня, то чередование «+» и «-» в построенной последовательности должно быть случайным. В этом случае справедлива гипотеза о неизменности среднего значения ВР. При этом построенная последовательность не должна содержать слишком длинных серий подряд идущих «+» или «-», и, соответственно, общее число серий не должно быть слишком малым. Так что в данном

критерии целесообразно рассматривать одновременно пару критических статистик (ν(n); $l(n)$). Если нарушается хотя бы одно из неравенств

$$\begin{cases} v(n) > \frac{1}{2}\left(n + 2 - 1.96\sqrt{n-1}\right), \\ l(n) < 1.43ln(n+1). \end{cases} \quad (1.13)$$

то гипотеза о неизменности среднего значения временного ряда отвергается с вероятностью ошибки α, такой, что 0,05 < α < 0,0975 и, тем самым, подтверждается наличие зависящей от времени неслучайной составляющей в разложении ВР.

Критерий «восходящих» и «нисходящих» серий

Этот критерий позволяет выявить смещение среднего значения в исследуемом распределении не только монотонного, но и более общего, например, периодического характера. Так же, как и в предыдущем критерии, исследуется последовательность знаков – плюсов и минусов, однако правило образования этой последовательности в данном критерии иное: на i-ом месте вспомогательной последовательности ставится «+», если $y_{i+1} > y_i$, и «-», $y_{i+1} < y_i$ (если два или несколько следующих друг за другом наблюдений равны между собой, то принимается во внимание только одно из них). Последовательность подряд идущих «+» (восходящая серия) будет соответствовать возрастанию результатов наблюдения, а последовательность «-» (нисходящая серия) – их убыванию. Критерий основан на том же соображении, что и предыдущий: если выборка случайна, то в образованной последовательности знаков общее число серий не может быть слишком малым, а их протяженность — слишком большой. При уровне значимости 0,05 < α < 0,0975 критерий имеет вид:

$$\begin{cases} v(n) > \frac{2n-1}{3} - 1.96\sqrt{\frac{16n-29}{90}}, \\ l(n) < l_0(n), \end{cases} \quad (1.14)$$

где $l_0(n) = \begin{cases} 5 \text{ при } n \leq 26, \\ 6 \text{ при } 26 < n \leq 153, \\ 7 \text{ при } 153 < n \leq 1170. \end{cases}$

Если хотя бы одно из неравенств (1.14) окажется нарушенным, то гипотезу о неизменности среднего значения временного ряда следует отвергнуть, что говорит о наличии тенденции в исследуемом ВР.

3.4. Расчет показателей, характеризующих тенденцию динамики

Абсолютное изменение уровней, или абсолютный прирост, – это разность между сравниваемым уровнем и уровнем более раннего периода, принятым за базу сравнения. Если эта база – непосредственно предыдущий уровень, показатель называют цепным:

$$\Delta_n = y_n - y_{n-1}. \quad (1.15)$$

Если за базу взят начальный уровень, показатель называют базисным:

$$\Delta_0 = y_n - y_0. \quad (1.16)$$

Если абсолютное изменение отрицательно, его следует называть абсолютным сокращением. Абсолютный прирост со временем возрастает, т.е. уровни ряда изменяются с ускорением. *Ускорение* - это разность между абсолютным изменением за данный период и абсолютным изменением за предыдущий период равной длительности:

$$\Delta_i = \Delta_i - \Delta_{i-1}. \quad (1.17)$$

Показатель абсолютного ускорения применяется только в цепном варианте, но не в базисном.

Темп роста – это отношение сравниваемого уровня (более позднего) к уровню, принятому за базу сравнения (более раннему). Темп роста исчисляется в цепном варианте – к уровню предыдущего года, а в базисном – к одному и тому же, обычно начальному уровню. Он свидетельствует о том, сколько процентов составляет сравниваемый

уровень по отношению к уровню, принятому за базу, или во сколько раз сравниваемый уровень больше уровня, принятого за базу.

Цепной темп роста определяется по формуле:

$$k_n = \frac{y_n}{y_{n-1}}. \quad (1.18)$$

Базисный темп роста определяется по формуле:

$$k_0 = \frac{y_n}{y_0}. \quad (1.19)$$

Если сравниваемый уровень выразить через уровень базисного периода и абсолютное изменение, получим:

$$k_n = \frac{y_{n-1}+\Delta}{y_{n-1}} = 1 + \frac{\Delta}{y_{n-1}}.$$

(1.20)

Величина $\frac{\Delta}{y_{n-1}}$, т.е. отношение абсолютного изменения к уровню предыдущего (или базисного) года, называется относительным приростом или темпом прироста. Он равен темпу изменения минус единица (минус 100 %).

Средние показатели динамики – средний уровень ряда, средние абсолютные изменения и ускорения, средние темпы роста – характеризуют тенденцию. Они необходимы при обобщении характеристик тенденции за длительный период, по различным периодам; при сравнении развития за неодинаковые по длительности отрезки времени и при выборе аналитического выражения тренда.

Средний уровень интервального ряда динамики определяется как простая арифметическая средняя из уровней за равные промежутки времени:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}, \quad (1.21)$$

или как взвешенная арифметическая средняя из уровней за неравные промежутки времени, длительность которых и является весами.

Средний абсолютный прирост (абсолютное изменение) определяется как простая арифметическая средняя из абсолютных изменений за равные промежутки времени или как частное от деления базисного абсолютного изменения на число усредняемых отрезков времени от базисного до сравниваемого периода:

$$\bar{\Delta} = \frac{\Delta_i}{n} = \frac{y_n - y_0}{n}. \quad (1.22)$$

Средний темп изменения определяется наиболее точно при аналитическом выравнивании динамического ряда по экспоненте. Средний темп определяют как геометрическую среднюю из цепных темпов роста за *n* лет или из общего (базисного) темпа роста за *n* лет:

$$\bar{k} = \sqrt[n]{\prod_{i=1}^{n} k_i} = \sqrt[n]{\frac{y_n}{y_0}}. \quad (1.23)$$

3.5. Сглаживание временных рядов

Сглаживание временного ряда, т.е. замена фактических уровней расчетными значениями, имеющими меньший разброс, чем исходные данные, является простым методом выявления тенденции развития. В ряде случаев при графическом изображении временного ряда тренд прослеживается недостаточно отчетливо. Поэтому ряд сглаживают, на график наносят сглаженные значения, и, как правило, тенденция проявляется более четко. Некоторые методы анализа и прогнозирования требуют в качестве предварительного условия сглаживание временного ряда. Сглаживание временных рядов используется при устранении аномальных наблюдений.

Аналитические методы. Сглаживание с использованием кривой, проведенной относительно фактических значений ряда так, чтобы эта кривая отображала тенденцию, присущую ряду и одновременно освобождала его от мелких незначительных колебаний. Такие кривые называют еще кривыми роста.

Методы механического сглаживания. При использовании этих методов производится сглаживание каждого отдельного уровня ряда с использованием фактических значений соседних с ним уровней. Для сглаживания временных рядов часто используются методы простой и взвешенной скользящей средней, экспоненциального сглаживания. В основе этих методов лежит простая идея: если «индивидуальный» разброс значений члена временного ряда $x(t)$ около своего среднего (сглаженного) значения μ характеризуется дисперсией σ^2, то разброс среднего из N членов временного ряда около того же значения будет характеризоваться гораздо меньшей величиной дисперсии, равной σ^2/N.

Метод простой скользящей средней. Согласно этому методу определяется количество наблюдений, входящих в интервал сглаживания. Сглаживание временного ряда по методу скользящей средней заключается в замене исходных уровней ряда y_t сглаженными значениями y'_t, которые получаются как среднее значение определенного числа уровней исходного ряда, симметрично окружающих значение y_t. В результате получается временной ряд y'_t, меньше подверженный колебаниям. Возможны два варианта расчета:

1) Если интервал сглаживания нечетный, т.е. $g=2p+1$, то

$$y'_t = \frac{y_{t-p}+y_{t-p+1}+\cdots+y_{t+p-1}+y_{t+p}}{2p+1} = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1}. \qquad (1.24)$$

2) Если интервал сглаживания четный, т.е. $g=2p$, то

$$y'_t = \frac{\frac{1}{2}y_{t-p}+y_{t-p+1}+\cdots+y_{t+p-1}+\frac{1}{2}y_{t+p}}{2p} = \frac{\frac{1}{2}y_{t-p}+\sum_{i=t-p+1}^{t+p-1} y_i+\frac{1}{2}y_{t+p}}{2p}. \qquad (1.25)$$

При использовании скользящей средней с длиной активного участка $g = 2p+1$ первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек является существенным недостатком, т. к. для исследователя последние

«свежие» данные обладают наибольшей информационной ценностью. Для восстановления потерянных значений временного ряда можно использовать следующий прием:

а) Вычисляется средний прирост Δ_y на последнем активном участке ($y_{n-g},\ldots,y_n$):

$$\Delta_y = \frac{y_n - y_{n-g}}{g-1} \qquad (1.26)$$

где g – длина активного участка.

б) Определяются значения последних $p=(g-1)/2$ уровней сглаженного временного ряда с помощью последовательного прибавления среднего абсолютного прироста Δ_y к последнему сглаженному значению.

Аналогичная процедура применяется для восстановления первых p уровней временного ряда. Важным свойством процедуры сглаживания является полное устранение периодических колебаний из временного ряда, если длина интервала сглаживания берется равной или кратной периоду колебаний. Если развитие процесса носит нелинейный характер, то применение метода простой скользящей средней может привести к значительным искажениям исследуемого процесса.

Метод взвешенной скользящей средней. Этот метод отличается от предыдущего тем, что сглаживание внутри интервала производится не по прямой, а по кривой более высокого порядка. Чаще всего используются полиномы 2-го и 3-го порядков. Так как при простой скользящей средней выравнивание на каждом активном участке производится по прямой (полиному первого порядка), то метод простой скользящей средней может рассматриваться как частный случай метода взвешенной скользящей средней. Простая скользящая средняя учитывает все уровни ряда, входящие в активный участок сглаживания, с равными весами, а взвешенная средняя приписывает каждому уровню

вес, зависящий от удаления уровня до уровня, стоящего в середине активного участка. Выравнивание с помощью взвешенной скользящей средней осуществляется следующим образом. Для каждого активного участка подбирается полином вида

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + \cdots,$$

параметры которого оцениваются по методу наименьших квадратов. При этом начало отсчета переносится в середину активного участка. Нет необходимости каждый раз заново вычислять весовые коэффициенты при уровнях ряда, входящих в активный участок сглаживания, так как они будут одинаковыми для каждого активного участка. При этом веса обладают свойствами:

• симметричность относительно центрального уровня,

• сумма весов с учетом общего множителя равна 1,

• наличие как положительных, так и отрицательных весов позволяет сглаженной кривой сохранять различные изгибы кривой тренда.

Существуют приемы, позволяющие с помощью дополнительных вычислений получить сглаженные значения для p начальных и конечных уровней ряда при длине интервала сглаживания $g = 2p + 1$.

Метод экспоненциального сглаживания. Рассмотренные методы простой и взвешенной скользящей средней не дают возможности сгладить первые и последние p наблюдений временного ряда. Отсутствие сглаженных первых наблюдений не так важно по сравнению с последними наблюдениями, особенно если целью исследования является прогнозирование развития процесса. Есть методы, позволяющие получить сглаженные значения последних уровней так же, как и всех остальных. К их числу относится метод экспоненциального сглаживания. Особенность этого метода заключена в том, что в процедуре выравнивания каждого наблюдения

используются только значения предыдущих уровней, взятых с определенным весом. Вес каждого наблюдения уменьшается по мере его удаления от момента, для которого определяется сглаживаемое значение. Сглаженное значение наблюдения ряда St на момент времени t определяется по формуле:

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1}, \quad (1.27)$$

где α – сглаживающий параметр, характеризующий вес выравниваемого наблюдения, причем $0 < \alpha < 1$. От численного значения параметра α зависит, насколько быстро будет уменьшаться вес предшествующих наблюдений и в соответствии с этим степень их влияния на сглаживаемый уровень. Чем больше значение параметра α, тем меньше сказывается влияние предшествующих уровней и соответственно меньшим оказывается сглаживающее воздействие экспоненциальной средней. Задачу выбора параметра, определяющего начальные условия, предлагается решать следующим образом: если есть данные о развитии процесса в прошлом, то их среднее значение можно принять в качестве y_0, если таких сведений нет, то в качестве y_0 используют исходное (первое) значение наблюдения временного ряда.

3.6. Выделение периодических составляющих временного ряда

Спектральный анализ. С помощью преобразования Фурье любой ряд динамики можно представить в виде суммы конечного числа гармоник. Но часто исследователю нужны не все гармоники, а лишь основные, порождающие основную часть дисперсии процесса.

Рассмотрим реализацию $x(t)$ длиной T, принадлежащую стационарному случайному процессу и имеющую нулевое среднее значение. При дискретном временном параметре такая реализация представлена N (т.к. $T=N\cdot\Delta t$) значениями временного ряда $\{x(t)\}$.

Дискретное преобразование Фурье дает значения спектральной плотности на частотах:

$$f_k = \frac{k}{T} = \frac{k}{\Delta t \cdot N}, \quad k = \overline{0, N-1}. \tag{1.28}$$

При этом коэффициенты Фурье определяются в виде:

$$X(f_k) = \Delta t \cdot \sum_{i=0}^{N-1} x(i) \cdot e^{-\frac{ij2\pi k}{N}}. \tag{1.29}$$

Оценку спектральной плотности:

$$\widetilde{S}(f_k) = \frac{1}{\Delta t \cdot N} |X_i(f_k)|^2 \tag{1.30}$$

можно рассматривать как характеристику степени тесноты связи между $x(i)$ и гармонической компонентой с периодом $2\pi/w$.

Вейвлет-анализ. Классический анализ Фурье основан на возможности исследования функций во временной ($| t | < \infty$) и частотной ($| v | < \infty$) областях с помощью прямого и обратного преобразования Фурье. Однако преобразование Фурье не позволяет достаточно полно исследовать иррегулярные функции, т.е. функции, характеристики которых эволюционируют во времени. Например, преобразование Фурье не отличает сигнал, представляющий собой сумму двух синусоид, от сигнала, состоящего из тех же синусоид, но включенных последовательно. Для устранения этого недостатка нужно локализовать преобразование Фурье на промежутках конечной длины.

Интегральным вейвлет–преобразованием функции $f(t) \in L^2(R)$ называется выражение

$$W(a,b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} f(t) \psi^*\left(\frac{t-b}{a}\right) dt, \tag{1.31}$$

где $a, b \in R$, $a \neq 0$.

Входящая в выражение (1.31) функция $\psi(t)$ называется вейвлетом. Вейвлет – это математическая функция, позволяющая анализировать различные частотные компоненты данных. График функции выглядит как волнообразные колебания с амплитудой, уменьшающейся до нуля вдали от начала координат. Существует много разных видов вейвлетов, например, вейвлет Морле, вейвлет Добеши, вейвлет Гаусса, вейвлет Хаара, MHat- вейвлет и другие.

Параметр a определяет размер вейвлета и называется масштабом. Его аналогом в Фурье-анализе является период (частота) гармонического колебания. Понятие масштаба – более широкое (хотя и менее наглядное), чем понятие периода. Связано это с тем, что в Фурье-преобразовании функциональный вид ядра преобразования зафиксирован, в то время как вейвлет преобразование одной и той же функции можно получить с помощью различных базисных вейвлетов (т.е. в разных системах масштабов).

Параметр b задает временную локализацию вейвлета и называется сдвигом. Этот параметр не имеет аналога в Фурье-преобразовании.

Обратное интегральное вейвлет-преобразование задается выражением

$$f(t) = C_{\psi}^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(a,b) \psi\left(\frac{t-b}{a}\right) \frac{1}{a^{1/2}} \frac{dadb}{a^2}, \qquad (1.32)$$

где C_{ψ} – нормирующий коэффициент:

$$C_{\psi} = \int_{-\infty}^{\infty} |\widehat{\psi}|^2 |\omega|^{-1} dw < \infty. \qquad (1.33)$$

Аналитическое представление вейвлета Морле и его преобразования Фурье задается следующими выражениями:

$$\psi(t) = e^{-t^2/\alpha^2} \left[e^{ik_0 t} - e^{-ik_0^2 \alpha^2 / 4} \right], \qquad (1.34)$$

$$\hat{\psi}(\omega) = \alpha\sqrt{\pi}\left[e^{-\alpha^2(k_0-\omega)^2/4} - e^{-\alpha^2(k_0+\omega)^2/4}\right] \qquad (1.35)$$

Вейвлет Морле – это плоская волна, модулированная гауссианой. Параметр α задает ширину гауссианы, параметр k_0 – частоту плоской волны. Обычно выбирают $\alpha^2 = 2$ и $k_0 = 2\pi$. При этих значениях с достаточной точностью можно принять:

$$\psi(t) = e^{-t^2/\alpha^2} e^{i2\pi t}, \qquad (1.36)$$

$$\hat{\psi}(\omega) = \alpha\sqrt{\pi} e^{-\alpha^2(2\pi-\omega)^2/4}. \qquad (1.37)$$

Практические аспекты вейвлет-анализа. Будем считать, что временной ряд задан значениями функции, следующими друг за другом с постоянным шагом Δt:

$f_k = f(t_k)$, $t_k = \Delta t\, k$, $k = 0,1,\ldots N\text{-}1$.

Для оценки вейвлет-преобразования этой последовательности воспользуемся следующим выражением:

$$W(a,b) = \frac{1}{n(a,b)} \sum_{k=0}^{N-1} f_k \psi^*\left(\frac{t_k - b}{a}\right), \qquad (1.38)$$

где

$$n(a,b) = \sum_{k=0}^{N-1} e^{-\frac{1}{B}\left(\frac{t_k-b}{a}\right)^2}. \qquad (1.39)$$

Для вейвлета Морле $b = \alpha^2$.

Введем оценку локального спектра энергии

$$S(a_i, b_j) = \left|W_a(a_i, b_j)\right|^2 \qquad (1.40)$$

Эту функцию обычно называют скалограммой, подчеркивая тем самым ее способность описывать распределение энергии по масштабам. Поскольку это распределение локализовано во времени с помощью параметра сдвига b, уместно называть локальной скалограммой.

3.7. Фрактальный анализ временных рядов

Фрактальными можно назвать эволюционные процессы, ВР показателей которых обладают долговременной памятью. К их числу относятся чаще всего природные ВР либо ВР основных показателей эволюционных процессов.

Фрактальный анализ позволяет выявить устойчивость тенденции ряда. Одной из составляющих фрактального анализа является R/S-анализ. История создания методологии R/S – анализа восходит к середине XX-го века, когда гидролог Херст решил проверить предположение о том, что ряды объемов стока рек подчиняются нормальному закону, он в результате ввел новую статистику – показатель Херста *(H)*. Как оказалось, этот показатель имеет широкое применение в анализе ВР благодаря своей устойчивости. Он содержит минимальные предположения об изучаемой системе и может классифицировать временные ряды. Он может отличить случайный ряд от неслучайного, даже если случайный ряд не гауссовский. Херст обнаружил, что большинство природных систем не следуют случайному блужданию – гауссовскому, т.е. поведение ВР показателей этих систем не подчиняется нормальному закону. Херст измерял колебания воды в резервуаре относительно среднего с течением времени и ввел безразмерное отношение посредствам деления размаха R на стандартное отклонение наблюдений S. Этот способ анализа стал называться методом нормированного размаха (*R/S*-анализа).

В представленном ВР *Z* последовательно выделяем его начальные отрезки

$$Z_\tau = z_1, z_2, \ldots z_\tau, \ \tau = 3,4\ldots,n, \qquad (1.41)$$

для каждого из которых вычисляем текущее среднее

$$\overline{Z}_\tau = \frac{1}{\tau}\sum_{i=1}^{\tau} Z_i \ . \qquad (1.42)$$

Далее для каждого фиксированного τ вычисляем накопленное отклонение для отрезков длины *t*:

$$X_{\tau,t} = \sum_{i=1}^{t} (z_i - z_\tau),\ t = \overline{1,\tau}. \qquad (1.43)$$

После чего вычисляем разность между максимальным и минимальным накопленными отклонениями (размах)

$$R = R(\tau) = max(X_{\tau,t}) - min(X_{\tau,t}). \qquad (1.44)$$

Этот размах нормируется, т.е. представляется в виде дроби *R**S*, где $S=S(\tau)$ – стандартное отклонение для отрезка ВР Z_τ, $3 \le \tau < n$.

Херст ввел следующее соотношение:

$$R/S=(a\cdot N)^H, \qquad (1.45)$$

где *R*/*S* – нормированный размах, *N* – число наблюдений, *a* – константа, *H* – показатель Херста.

Имеются три различных классификации для показателя Херста:

1) $H = 0.5$. Указывает на *случайный* ряд. События случайны и некоррелированны. Настоящее не влияет на будущее. Функция плотности вероятности может быть нормальной кривой, однако это не обязательное условие.

2) $0 \le H < 0.5$. Данный диапазон соответствует *антиперсистентным,* или *эргодическим*, рядам. Такой тип системы часто называют «возврат к среднему». Если система демонстрирует «рост» в предыдущий период, то, скорее всего, в следующем периоде начнется спад. И наоборот, если шло снижение, то вероятен близкий подъем. Устойчивость такого антиперсистентного поведения зависит от того, насколько *H* близко к нулю.

3) $0.5 < H < 1.0$. Имеем *персистентные,* или *трендоустойчивые,* ряды. Если ряд возрастает (убывает) в предыдущий период, то, вероятно, что он будет сохранять эту тенденцию какое-то время в будущем. Чем ближе H к 0.5, тем более зашумлен ряд и тем менее выражен его тренд.

Существует еще и четвертая характеристика показателя Херста, когда $H>1$. В этом случае говорят о статистике Леви и о процессе (или временном ряде) с фрактальным временем, о временных точках разрыва производной. Это означает, что происходят независимые скачки амплитуды, распределенные по Леви за время, определенное величиной скачка, и растущее вместе с ним.

Контрольные вопросы

1. Что такое временной ряд?
2. Какие бывают временные ряды?
3. Что такое стационарный и нестационарный временной ряд?
4. Назовите составляющие временного ряда.
5. Что собой представляет тренд?
6. Что такое периодическая и сезонная составляющие временного ряда?
7. Назовите методы предварительного анализа временных рядов.
8. Опишите критерии проверки временного ряда на стационарность.
9. Что собой представляет фрактальный анализ?
10. Для чего используется фрактальный анализ?
11. Что такое показатель Херста? О чем говорит показатель Херста, близкий к 0?

12. Назовите характеристики, описывающие динамику временного ряда.
13. Для чего используется спектральный анализ?
14. Для чего используется вейвлет-анализ?
15. Какие виды вейвлетов вы знаете?
16. Какие тесты на наличие тенденции вы знаете?
17. Опишите критерий Фостера-Стюарта.
18. Для чего используется критерий Дарбина – Уотсона?
19. Как определить коэффициент автокорреляции?
20. Что собой представляет автокорреляционная функция?

РАЗДЕЛ 2. МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

1. Построение моделей тренда

В предыдущем разделе мы рассмотрели методы, позволяющие установить наличие тенденции во временном ряде.

Встречающиеся во многих исследованиях временные ряды наряду с флуктуациями и нерегулярностями имеют общую тенденцию изменения. После обнаружения наличия тенденции перед исследователем возникает задача построения уравнения, описывающего данную тенденцию. Во многих случаях желательно делать выводы о тренде на основании самого наблюдаемого ряда, хотя при этом может быть известно, что тренд возникает в силу действия каких-то иных факторов. В ряде случаев, когда теория не может указать явный вид тренда как функции времени, тем не менее бывает возможно приблизить тренд полиномом достаточно низкой степени.

Аналитическим выравниванием ВР называется нахождение аналитической функции $\hat{y} = f(t)$, характеризующей основную тенденцию изменения уровней ряда с течением времени. Сама функция $f(t)$ носит название кривой роста. При аналитическом выравнивании (нахождении аналитической функции $\hat{y} = f(t)$) исходят из предположения, что аддитивная модель временного ряда может быть представлена как сумма двух компонент $y(t) = f(t) + \varepsilon(t)$, где $\varepsilon(t)$ – случайная компонента с нулевой средней и постоянной дисперсией выражает ошибку модели из-за действия случайных факторов. Чаще всего в качестве кривой роста применяются следующие функции:

– линейная $\hat{y}_t = a + bt$,

– парабола второго и более высоких порядков $\hat{y}_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k$,

– гиперболическая $\hat{y}_t = a + \frac{b}{t}$,

– экспоненциальная $\hat{y}_t = e^{a+bt}$,

– потенциальная $\hat{y}_t = ab^t$,

– степенная $\hat{y}_t = at^b$

и другие.

Выбор кривой чаще всего обусловлен графическим представлением значений исходного ВР. Рассмотрим некоторые виды моделей тренда.

1.1. Простая линейная регрессия

Самым простым типом линии тренда является прямая линия, описываемая линейным уравнением тренда:

$$\hat{y}_t = a + bt, \tag{2.1}$$

где $\hat{y}_t$ – предсказанные значения ВР для некоторого момента времени *t*.

Этот тип тренда подходит для отображения тенденции примерно равномерных изменений уровней: равных в среднем абсолютных приростов или абсолютных сокращений уровней за равные промежутки времени.

Для оценки параметров линейной регрессии используется метод наименьших квадратов, суть которого заключается в минимизации суммы квадратов отклонений исходных значений ВР от предсказанных:

$$S = \sum_{t=1}^{n} e_t^2 = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 \to min. \tag{2.2}$$

Оценки параметров линейной регрессии, полученные с помощью МНК, определяются по формулам:

$$b = \frac{n\sum_{i=1}^{n} t_i y_i - \sum_{i=1}^{n} t_i \cdot \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} t_i^2 - \left(\sum_{i=1}^{n} t_i\right)^2} = \frac{\overline{ty} - \bar{t}\cdot\bar{y}}{\overline{t^2} - (\bar{t})^2}; \tag{2.3}$$

$$a = \frac{1}{n}\sum_{i=1}^{n} y_i - b\frac{1}{n}\sum_{i=1}^{n} t_i = \bar{y} - b\bar{t}, \tag{2.4}$$

где $\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n}$ – среднее значение переменной *t*, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$ – среднее значение переменной *y*, $\overline{ty} = \frac{\sum_{i=1}^{n} (t_i \cdot y_i)}{n}$, $\overline{t^2} = \frac{\sum_{i=1}^{n} t_i^2}{n}$.

1.2. Квадратичный тренд

Квадратичным называется тренд, выраженный полиномом второй степени с уравнением

$$\hat{y}_t = a_0 + a_1 t + a_2 t^2. \quad (2.5)$$

Полиномы третьей степени и более высоких порядков редко применимы для выражения тенденции динамики и слишком сложны для получения надежных оценок параметров при ограниченной длине ВР.

Тренд в форме полинома второй степени применяется для отображения таких тенденций динамики, которым свойственно примерно постоянное ускорение абсолютных изменений уровней.

Парабола 2 порядка, как уравнение тренда, применяется к различным процессам, которые на некотором, как правило непродолжительном, этапе развития имеют примерно постоянное ускорение абсолютного прироста уровней.

Оценивание параметров квадратичного тренда также осуществляется с помощью МНК. Минимизация суммы квадратов отклонений приводит к построению системы нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum t_i + a_2 \sum t_i^2 = \sum y_i; \\ a_0 \sum t_i + a_1 \sum t_i^2 + a_2 \sum t_i^3 = \sum t_i y_i; \\ a_0 \sum t_i^2 + a_1 \sum t_i^3 + a_2 \sum t_i^4 = \sum t_i^2 y_i. \end{cases} \quad (2.6)$$

Решая эту систему уравнений, получаем оценки параметров квадратичной регрессии.

1.3. Экспоненциальный тренд

Применить обычный МНК для оценивания параметров большинства нелинейных моделей невозможно. Одним из способов

решения этой проблемы является преобразование нелинейной модели к линейному виду. Для этого можно использовать следующие процедуры:

1) логарифмирование;
2) замена переменных;
3) потенцирование.

Рассмотрим экспоненциальный тренд вида:

$$\hat{y}_t = e^{a+bt}. \quad (2.7)$$

Прологарифмируем обе части уравнения:

$ln\hat{y}_t = \ln(e^{a+bt})$.

Используя свойства логарифмов, получаем:

$ln\hat{y}_t = (a + bt)\ln e$.

$ln\hat{y}_t = a + bt$.

Сделаем замену переменной $y'_t = ln\hat{y}_t$. В результате получим линейное уравнение:

$y'_t = a + bt$.

Так как в модели используются не исходные уровни ВР, а их логарифмы, нужно преобразовать исходную таблицу данных, рассчитав натуральные логарифмы значений ВР. Применяем к полученной модели МНК и находим параметры модели по формулам (2.3) и (2.4).

1.4. Степенной тренд

Рассмотрим степенную модель тренда:

$$\hat{y}_t = at^b. \quad (2.8)$$

Здесь, как и в случае экспоненциального тренда, модель содержит степень. В подобных случаях лучше всего использовать процедуру логарифмирования. Проведем те же преобразования, что и в п. 1.3:

$ln\hat{y}_t = \ln(at^b)$,

$ln\hat{y}_t = lna + \ln(t^b)$;

$ln\hat{y}_t = lna + bln\,t$.

Проведем замену переменных:

$y'_t = ln\hat{y}_t; t' = lnt; a' = lna.$

В результате получим линейное уравнение:

$y'_t = a' + bt'.$

Преобразуем исходную таблицу данных: вместо исходных значений ВР и переменной *t* находим их натуральные логарифмы. Применяем к полученной модели МНК и находим параметры b и a' по формулам (2.3) и (2.4). Параметр a степенной модели определяется как $a = e^{a'}$.

1.5. Гиперболический тренд

Рассмотрим гиперболическую модель тренда:

$$\hat{y}_t = a + \frac{b}{t}. \quad (2.9)$$

Чтобы привести ее к линейному виду, достаточно произвести замену переменной z = $\frac{1}{t}$ и преобразовать исходную таблицу данных. Получаем линейный тренд:

$\hat{y}_t = a + bz,$

параметры которого оцениваются обычным МНК.

2. Моделирование периодических колебаний временного ряда

В ВР нередко содержатся периодические колебания вокруг общей тенденции. В таких случаях становится целесообразным применение гармонического анализа, целью которого является выявление и измерение периодических колебаний в рядах динамики. Задача гармонического анализа состоит в выявлении гармоник, несущих основную информацию о наблюдаемом процессе. С помощью преобразования Фурье любой ВР можно представить в виде суммы конечного числа гармоник.

После проведения аппроксимации ряда полиномом низкой степени и удаления тренда, в зависимости от того, обнаружены по спектральному анализу и вейвлет-анализу значимые и устойчивые гармоники или нет, проводится гармонический анализ.

С помощью преобразования Фурье любой ВР можно представить в виде суммы конечного числа гармоник.

Для каждой точки этого ряда справедливо выражение

$$y_t = f(t) + \sum \left(a_k \cos\left(kt \frac{2\pi}{n} \right) + b_k \sin\left(kt \frac{2\pi}{n} \right) \right). \ (t = 1,2, \dots, n) \qquad (2.10)$$

Здесь y_t – фактический уровень ряда в момент (интервал) времени t; $f(t)$ – выровненный уровень ряда в тот же момент времени, a_k, b_k – параметры колебательного процесса (гармоники) с номером k, в совокупности оценивающие размах отклонения от общей тенденции и сдвиг колебаний относительно начальной точки. Общее число колебательных процессов, которые можно выделить для ряда, состоящего из n уровней, равно $n/2$. Обычно ограничиваются меньшим числом наиболее важных гармоник, выделяемых с помощью спектрального и вейвлет-анализа. Параметры гармоники с номером k определяются по формулам:

$$\varepsilon_t = y_t - f(t); \qquad (2.11)$$

$$a_k = \frac{2}{n} \sum_{t=1}^{n} \varepsilon_t \cdot \cos\left(kt \frac{2\pi}{n} \right); \quad k = 1,2, \dots, \frac{n}{k} - 1 \qquad (2.12)$$

$$b_k = \frac{2}{n} \sum_{t=1}^{n} \varepsilon_t \cdot \sin\left(kt \frac{2\pi}{n} \right); \quad k = 1,2, \dots, \frac{n}{k} - 1 \qquad (2.13)$$

$$a_{n/2} = \frac{1}{n} \sum_{t=1}^{n} \varepsilon_t \cdot \cos(\pi t); \quad b_{n/2} = 0. \qquad (2.14)$$

Этот метод хорошо подходит для аналитического выражения периодических колебаний, имеющих синусоидальную форму.

3. Моделирование случайной составляющей временного ряда

3.1. Модели авторегрессии

Модель авторегрессии 1-го порядка – AR(1) может быть определена выражением:

$$y(t) = \alpha * y(t-1) + \varepsilon(t), \qquad (2.15)$$

где α – некоторый числовой коэффициент, не превосходящий по абсолютной величине единицу ($|\alpha| \leq 1$), а $\varepsilon(t)$ – последовательность случайных величин, образующих белый шум.

В качестве предварительного порядка модели AR(p) можно рассматривать такое число p, начиная с которого все последующие оценки выборочной частной автокорреляционной функции отклоняются от нуля не более чем на $\pm 2/\sqrt{n}$. Окончательный подбор порядка модели AR(p) процесса связан со статистической значимостью полученных коэффициентов модели и детальным изучением поведения остатков, получаемых вычитанием из исходного ряда значений подобранной модели. Если полученные остатки ведут себя как белый шум, то процесс подбора модели можно считать завершенным. В противном случае следует изменить порядок модели или перейти к более сложным комбинированным моделям авторегрессии-скользящего среднего.

В общем виде авторегрессионный процесс p-го порядка – AR(p) может быть записан так:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \varepsilon_t . \qquad (2.16)$$

Параметры β_j (j=1,…,p) модели (2.16) обычно оцениваются МНК при предположениях, что случайная величина ε_t распределена нормально и независимо от t и коэффициенты модели по абсолютной величине значения меньше единицы.

3.2. Модели скользящего среднего

В частном случае общего линейного процесса, когда только первые q из весовых коэффициентов β_j – ненулевые, процесс имеет вид:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}, \qquad (2.17)$$

где символы $-\theta_1, -\theta_2, \ldots, -\theta_q$ используются для обозначения конечного набора параметров β, участвующих в линейном процессе.

Процесс называется моделью скользящего среднего порядка q и сокращенно обозначается MA(q).

3.3. Модели авторегрессии-скользящего среднего

На практике для получения экономичной параметризации анализируемого процесса иногда бывает необходимо включить в модель как члены, описывающие авторегрессию, так и члены, моделирующие остаток в виде скользящего среднего. Такой линейный процесс имеет вид:

$$y_t = \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q} \qquad (2.18)$$

и называется процессом авторегрессии-скользящего среднего порядка (p,q). Сокращенное обозначение для него ARMA (p,q).

При очень общих условиях стационарный ARMA-процесс может быть представлен как бесконечный AR-процесс или как бесконечный MA-процесс. Использование ARMA-процессов позволяет строить более компактные модели реальных временных рядов по сравнению со схожими по поведению AR- или MA-процессами.

3.4. Авторегрессионная модель с условной гетероскедастичностью

Если задано распределение вероятностей y_t, условное по предопределенным к моменту t величинам $y_{t-1}, y_{t-2}, \cdots$, составляющим информационное множество Ω_t, то одношаговым прогнозом значения y_t на основе этой информации является условное математическое

ожидание $M(y_t|\Omega_t)$, ошибкой прогноза – условная дисперсия $D(y_t|\Omega_t)$. Оба этих выражения в принципе допускают зависимость от Ω_t. Если же условная дисперсия в действительности постоянна и не зависит от истории процесса, то такой процесс обладает свойством *условной гомоскедастичности*.

Рассмотрим в качестве примера авторегрессию

$$y_t = \gamma \cdot y_{t-1} + \varepsilon_t, \tag{2.19}$$

где ε – белый шум с дисперсией σ^2. Условное среднее равно $\gamma \cdot y_{t-1}$, тогда как условная дисперсия равна σ^2 и не зависит от истории процесса, который является поэтому условно гомоскедастичным. Являются таковыми и все процессы ARMA.

Ниже изучаются процессы, обладающие свойством условной гетероскедастичности, т.е. такие, условная дисперсия которых зависит от истории процесса и более точно характеризует степень присущей прогнозам неопределенности.

Модель ARCH, т.е. модель с авторегрессионной условной гетероскедастичностью (autoregressive conditional heteroskedasticity), предложена Р. Энглом в 1982 г. для моделирования кластеризации волатильности. Процесс ARCH *p*-го порядка, $\{\varepsilon_t\}_{t=-\infty}^{t=+\infty}$, задается следующим соотношением для условной дисперсии ε_t:

$$\varepsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2), \tag{2.20}$$

$$\sigma_t^2 = \mu + \gamma_1 \varepsilon_{t-1}^2 + \ldots + \gamma_p \varepsilon_{t-p}^2. \tag{2.21}$$

Здесь $\Omega_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ – предыстория процесса ε_t, а σ_t^2 – условная по предыстории дисперсия ε_t, т.е. $\sigma_t^2 = D(\varepsilon_t|\Omega_{t-1}) = M(\varepsilon_t^2|\Omega_{t-1})$. Условную дисперсию часто называют волатильностью процесса. Для

того чтобы условная дисперсия оставалась положительной, требуется выполнение соотношений $\mu > 0$ и $\gamma_1,\ldots,\gamma_p \geq 0$.

Смысл модели ARCH состоит в том, что если абсолютная величина ε_t оказывается большой, то это приводит к повышению условной дисперсии в последующие периоды. В свою очередь при высокой условной дисперсии более вероятно появление больших (по абсолютной величине) значений ε_t. Наоборот, если значения ε_t в течение нескольких периодов близки к 0, то это приводит к понижению условной дисперсии в последующие периоды практически до уровня μ. В свою очередь, при низкой условной дисперсии более вероятно появление малых (по абсолютной величине) значений ε_t. Таким образом, ARCH-процесс характеризуется инертностью условной дисперсии (кластеризацией волатильности).

Модель GARCH (generalized ARCH – обобщенная модель ARCH), предложенная Т. Боллерслевом, является альтернативной модификацией модели ARCH, позволяющей получить более длинные кластеры при малом числе параметров. Модель APCC зачастую позволяет получить более «сжатое» описание временных зависимостей для условного математического ожидания, чем модель AP. Подобным же образом модель GARCH дает возможность обойтись меньшим количеством параметров по сравнению с моделью ARCH, если речь идет об условной дисперсии.

Модель GARCH(p,q) описывается формулой:

$$\sigma_t^2 = \mu + \gamma_1\varepsilon_{t-1}^2 + \ldots + \gamma_p\varepsilon_{t-p}^2 + \delta_1\sigma_{t-1}^2 + \ldots + \delta_q\sigma_{t-q}^2 = \mu + \sum_{j=1}^{p}\gamma_j\varepsilon_{t-j}^2 + \sum_{j=1}^{q}\delta_j\sigma_{t-j}^2$$

(2.22)

При этом предполагается, что $\mu > 0$, $\gamma_1,\ldots,\gamma_p \geq 0$ и $\delta_1,\ldots,\delta_q \geq 0$. На практике, как правило, достаточно взять *p*=1 и *q*=1. Изредка используют GARCH(1,2) или GARCH(2,1).

Как и в модели ARCH, σ_t^2 служит условной дисперсией процесса:

$$\varepsilon_t | \Omega_t \sim N\left(0, \sigma_t^2\right). \tag{2.23}$$

Процесс GARCH можно записать в эквивалентной форме, если обозначить $\eta_t = \varepsilon_t^2 - \sigma_t^2$:

$$\varepsilon_t^2 = \mu + \sum_{j=1}^{m}\left(\delta_j + \gamma_j\right)\varepsilon_{t-j}^2 + \eta_t - \sum_{j=1}^{q}\delta_j \eta_{t-j}\ , \tag{2.24}$$

где *m=max(p,q)*. (В этой записи подразумевается $\gamma_j = 0$ при $j > p$ и $\delta_j = 0$ при $j > q$).

Стандартным методом оценивания для моделей GARCH является метод максимального правдоподобия. Оценки максимального правдоподобия являются состоятельными и асимптотически эффективными.

Важнейший вывод, который следует из анализа модели ARCH, состоит в том, что наблюдаемые изменения в дисперсии (волатильность) ВР могут иметь эндогенный характер, т.е. порождаться определенной нелинейной моделью, а не какими-то внешними структурными сдвигами.

Контрольные вопросы

1. Назовите известные вам модели, описывающие долговременную тенденцию временного ряда.
2. Опишите суть метода наименьших квадратов.
3. Как строится линейная модель временного ряда?

4. Как строится квадратичная модель временного ряда?
5. Опишите основные подходы к построению нелинейных моделей временных рядов.
6. Что собой представляет гармонический анализ?
7. Какие гармоники включаются в гармоническую модель?
8. Что собой представляет модель авторегрессии?
9. Какой метод используется для оценки параметров авторегрессии?
10. Что собой представляет модель скользящего среднего?
11. Что такое ARCH – модель?
12. Какие модели используются для описания случайной составляющей временного ряда?

РАЗДЕЛ 3. АНАЛИЗ КАЧЕСТВА МОДЕЛЕЙ И ДИАГНОСТИКА ОСТАТКОВ

1. Анализ качества моделей

Применение критериев качества позволяет выявить степень адекватности модели наблюдениям и ее пригодности для аппроксимации в данном выборочном пространстве по известным внутренним, а также смешанным и внешним статистическим мерам соответствия.

Внутренние меры соответствия. К внутренним мерам соответствия относятся меры качества, полученные по построенной по исходным данным модели, такие как:

1) остаточная сумма квадратов

$$SS_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \quad (3.1)$$

2) стандартная ошибка наблюдений

$$\sigma = \sqrt{\frac{SS_e}{v_e}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-p}}, \quad (3.2)$$

где p – число параметров модели.

3) коэффициент множественной корреляции

$$R = \frac{\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2}}. \quad (3.3)$$

4) коэффициент множественной детерминации R^2.

5) скорректированный коэффициент множественной корреляции

$$R_C = \sqrt{1 - \frac{n(1-R^2)}{n-p}}. \quad (3.4)$$

6) скорректированный коэффициент множественной детерминации R_C^2.

7) наблюдаемое значение F-статистики

$$F = \frac{R^2(n-p)}{(1-R^2)(p-1)}. \quad (3.5)$$

Смешанные меры соответствия. В регрессионном анализе различают несколько видов ошибок прогноза. Для модели, в которой отсутствует смещение, можно предложить вычислять дисперсию отдельного прогноза – дисперсию оценки регрессии, или дисперсию оценки отдельного отклика *Y*. Для серии наблюдений необходимо ввести обобщенную (среднюю) дисперсию.

Одну из смешанных мер, основанных на формуле для отдельного прогноза, предложил Меллоус. Используя наблюдения, по которым строилась регрессионная модель, он вводит шкалированную сумму дисперсий, подсчитанную для *n* наблюдений.

Общим свойством таких мер является прямое или косвенное включение составляющей, характеризующей систематическую ошибку модели. К ограничениям смешанных мер согласия следует отнести использование гипотезы о несмещенности полной модели. Кроме того для смешанных мер используются те же наблюдения, что и для внутренних критериев.

Из рассмотренных мер наиболее удобной представляется мера C_p. С ее помощью можно оценить, какая из конкурирующих моделей обеспечивает более точный прогноз с учетом в целом случайных и систематических ошибок.

К смешанным мерам относятся:

1) мера Мэлоуса

$$\mathrm{C}_p = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sigma^2} + 2p - n. \quad (3.6)$$

2) мера $$J_p = \frac{n+p}{n-p} SS_e. \quad (3.7)$$

3) усредненная дисперсия для MY

$$D\hat{Y} = \frac{\sigma^2 C_p}{n}. \quad (3.8)$$

4) усредненная дисперсия для Y

$$DY = \frac{\sigma^2 C_p}{n} + \sigma^2. \quad (3.9)$$

Внешние меры соответствия. Основаны либо на анализе устойчивости коэффициентов модели в выборочном пространстве, либо на анализе расхождений между прогнозом и известным наблюдаемым значением для объектов, не участвовавших в получении модели.

Выборка наблюдений делится на две части – обучающую и контрольную. Обучающая выборка наблюдений используется для построения модели, тогда как контрольная последовательность (помимо возможности проверки устойчивости коэффициентов модели) дает возможность оценить несмещенность модели, т.е. качество прогноза по мерам, основанным на разностях между наблюденным значением отклика и его прогнозом или на их функциях.

Для временных рядов контрольной выборкой является часть наблюдений в конце ряда, не использованная при построении комплексной модели.

К внешним мерам соответствия относятся:

1) сумма квадратов отклонений

$$SS_{ed} = \sum_{i=1}^{k} (y_i - \hat{y}_i)^2, \quad (3.10)$$

где k – объем контрольной выборки.

2) случайная ошибка (внешнее среднеквадратическое отклонение)

$$\sigma_\Delta = \sqrt{\frac{SS_{ed}}{v_{ed}}} = \sqrt{\frac{\sum_{i=1}^{k} {\Delta_i}^2}{k-p}}, \quad (3.11)$$

где k – объем контрольной выборки, p- число регрессоров в модели, $\Delta_i = y_i - \hat{y}_i$.

3) модуль систематической ошибки

$$|\bar{\Delta}| = \sum_{i=1}^{k} |\Delta_i| / k. \quad (3.12)$$

4) «внешний» коэффициент множественной корреляции

$$R = \frac{\sum_{i=1}^{k} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{k} (y_i - \bar{y})^2 \sum_{i=1}^{k} (\hat{y}_i - \bar{\hat{y}})^2}}. \quad (3.13)$$

2. Диагностика остатков

После выполнения расчетов с применением МНК уточняется степень соблюдения основных предположений регрессионного анализа и теоремы Гаусса-Маркова.

Алгоритм проверки соблюдения предположений предназначен для получения оптимальной модели. Существует возможность диагностики соблюдения предположений регрессионного анализа на основе следующих предположений: определенность модели, независимость регрессоров, нормальность распределения ошибок, нулевое значение математического ожидания ошибок, постоянство дисперсии, независимость ошибок. Для диагностики регрессионной модели существует набор специальных процедур, таких как *t*-критерий, χ^2-критерий, *F*-критерий, критерии проверки на нормальность, критерий Бартлетта, критерий Дарбина-Уотсона и др.

Модель, признанная пригодной к использованию, может быть либо недоопределенной, либо избыточной (переопределенной). Признаком избыточности является наличие регрессоров, для которых значения частных *t*-статистик оказываются меньше критического $t_T(\alpha/2; n - p)$.

Недоопределенность модели исследуется лишь визуально по соответствующим графикам остатков. Автоматически подсчитывается заполнение тренда. Результат выдается в процентном соотношении.

Для оценки степени независимости регрессоров (отсутствие мультиколлинеарности) проверяется гипотеза о независимости

переменных H_0: $Det|R|=1$, где $Det|R|$ – определитель матрицы парных коэффициентов корреляции между факторами. Величина $[m–1–1/6\cdot(2\cdot m+5)\cdot lg(Det|R|)]$ имеет приближенное распределение χ^2 с $(1/2 \cdot n\cdot(n-1))$ степенями свободы.

На практике часто проверяется матрица парных коэффициентов корреляции. Если в матрице имеются статистически значимые коэффициенты корреляции, то мультиколлинеарность считается обнаруженной.

Проверка нормальности распределения может быть осуществлена как графическими методами, так и с помощью статистических критериев.

Для графической проверки может быть использована нормальная вероятностная бумага. Пусть $d_1<d_2<...<d_n$ – нормированные остатки, расположенные в порядке возрастания. Для умеренно больших значений n построенный на вероятностной бумаге график зависимости d_i от $(i–1)/n$ будет выявлять любые заметные отклонения от нормальности, поскольку при наличии нормальности точки графика должны располагаться приблизительно вдоль прямой $y = x$.

Для оценки соблюдения предположения о нормальности можно воспользоваться также графиком $(d, \hat{Y})$: равномерно заполненный на 95 % тренд шириной $\pm 2\sigma$ помимо адекватности модели свидетельствует о нормальности ε.

Из статистических критериев используется критерий χ^2, при котором проверяется гипотеза H0: $d \sim N(0, 1)$. Для проверки нормальности могут быть использованы критерий Шапиро-Уилка, критерий Колмогорова-Смирнова и другие. Приблизительная нормальность для небольшой по объему выборки может оцениваться по методу Айвазяна.

Для анализа предположения о равенстве нулю математического ожидания применяется критерий проверки значения среднего. Нарушение этого предположения сигнализирует об ошибках в расчетах.

Нарушение условия однородности наблюдений или постоянства дисперсии обычно проверяется по графикам остатков, а также статистическими критериями: критерий равенства (неравенства) двух дисперсий позволяет проверить гипотезу об однородности дисперсий в двух нормальных совокупностях с параметрами (μ_1, σ_1^2) и (μ_2, σ_2^2) объемом n_1, n_2; критерий Бартлетта позволяет проверить гипотезу об однородности дисперсий в нескольких (более двух) нормальных совокупностях.

Для проверки нарушения условия независимости ошибок привлекаются графики остатков (d, T), где T – время наблюдения.

Наряду с графическим представлением используется и аналитический критерий Дарбина-Уотсона.

Контрольные вопросы

1. Назовите внутренние меры качества модели.
2. Назовите смешанные меры качества модели.
3. Назовите внешние меры качества модели.
4. Что такое обучающая и контрольная выборки?
5. Какие предположения регрессионного анализа проверяются после построения модели?
6. Какие способы проверки на нормальность распределения вы знаете?
7. Для чего используется критерий Бартлетта?
8. Что такое недоопределенность и избыточность модели?
9. Как проверить наличие автокорреляции в остатках?
10. Для чего нужна диагностика остатков?

РАЗДЕЛ 4. АНАЛИЗ МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

1. Взаимные характеристики двух временных рядов

При изучении общих характеристик двух временных рядов используются понятия *взаимного спектра мощности* и *взаимной корреляционной функции*. В области оценок этим понятиям соответствуют *взаимная периодограмма* и *взаимная коррелограмма*.

Пусть имеются две дискретные центрированные последовательности действительных чисел

$$x_k = x(t_k), t_k = \Delta t k, k = 0,1,..., N-1, \quad (4.1)$$

$$y_k = y(t_k), t_k = \Delta t k, k = 0,1,..., N-1. \quad (4.2)$$

Обозначая символом * операцию комплексного сопряжения, определим взаимную периодограмму рядов x_k и y_k с помощью соотношения

$$D_{xy}(\upsilon) = \frac{1}{N^2} X^*(\upsilon) Y(\upsilon), \quad 4.3)$$

где

$$X(\upsilon) = \sum_{k=0}^{N-1} x_k e^{-i 2\pi \upsilon t_k}, \quad (4.4)$$

$$Y(\upsilon) = \sum_{k=0}^{N-1} y_k e^{-i 2\pi \upsilon t_k}. \quad (4.5)$$

Взаимная периодограмма используется для обнаружения частот общих гармоник.

Взаимная коррелограмма ВР определяется следующим образом:

$$\bar{c}_{xy}[m] = \frac{1}{N} \sum_{k=0}^{N-m-1} x_k y_{k+m}, \quad (4.6)$$

m=0,1, …, N-1.

При этом имеет место свойство

$$\bar{c}_{xy}[-m] = \bar{c}_{yx}[m]. \tag{4.7}$$

Взаимные периодограмма и коррелограмма связаны между собой соотношением

$$D_{xy}(\upsilon) = \frac{1}{N}\left[\sum_{m=0}^{N-1}\bar{c}_{xy}[m]e^{-i2\pi\upsilon m\Delta t} + \sum_{m=0}^{N-1}\bar{c}_{yx}[m]e^{-i2\pi\upsilon m\Delta t} - \bar{c}_0\right]. \tag{4.8}$$

В случае, когда два ряда x_k и y_k тождественны, взаимные периодограмма и коррелограмма соответствуют обычным понятиям периодограммы и коррелограммы. При этом выражение (4.8) переходит в

$$D(\upsilon) = \frac{1}{N}\left[2\,\mathrm{Re}\sum_{m=0}^{N-1}\bar{c}_m e^{-i2\pi\upsilon m\Delta t} - \bar{c}_0\right]. \tag{4.9}$$

В отличие от обычной периодограммы взаимная периодограмма представляет собой комплекснозначную функцию

$$D_{xy}(\upsilon) = P_{xy}(\upsilon) - iQ_{xy}(\upsilon), \tag{4.10}$$

где

$$P_{xy}(\upsilon) = \mathrm{Re}\,X(\upsilon)\,\mathrm{Re}\,Y(\upsilon) + \mathrm{Im}\,X(\upsilon)\,\mathrm{Im}\,Y(\upsilon), \tag{4.11}$$

$$Q_{xy}(\upsilon) = \mathrm{Re}\,X(\upsilon)\,\mathrm{Im}\,Y(\upsilon) - \mathrm{Re}\,Y(\upsilon)\,\mathrm{Im}\,X(\upsilon). \tag{4.12}$$

Функцию $P_{xy}(\upsilon)$ называют *коспектром*, а функцию $Q_{xy}(\upsilon)$ – *квадратурным спектром*. Для получения конкретной информации об общих свойствах исследуемых рядов вводятся в рассмотрение *модуль взаимной периодограммы, или кросс-амплитуда*

$$A_{xy}(\upsilon) = \sqrt{P_{xy}^2(\upsilon) + Q_{xy}^2(\upsilon)} \tag{4.13}$$

и *фазовый спектр*

$$\Phi_{xy}(\upsilon) = -arctg\frac{Q_{xy}(\upsilon)}{P_{xy}(\upsilon)}. \tag{4.14}$$

Кросс-амплитуда может интерпретироваться как мера ковариации между соответствующими частотными компонентами двух рядов. Фазовый спектр показывает разность фаз общих гармоник на данной частоте. Оценки фазового спектра измеряют, насколько каждая частотная компонента одного ряда опережает частотные компоненты другого.

Можно нормировать значения кросс-амплитуды, возведя их в квадрат и разделив на произведение оценок спектральной плотности каждого ряда. Результат называется *квадратом когерентности*, и он может быть интерпретирован как квадрат коэффициента корреляции между циклическими компонентами двух рядов соответствующей частоты.

Значение усиления в анализе вычисляется делением значения кросс-амплитуды на оценки спектральной плотности одного или двух рядов. Следовательно, может быть вычислено два значения усиления, которые могут интерпретироваться как стандартные коэффициенты регрессии соответствующей частоты, полученные методом наименьших квадратов.

2. Взаимный спектральный анализ временных рядов

Ниже приводится алгоритм взаимного спектрально-корреляционного анализа двух ВР, который является обобщением соответствующего алгоритма анализа одного временного ряда.

Исходные данные:

- Δt - временной шаг задания рядов;
- равномерные временные ряды:

$$x_k = x(t_k), t_k = \Delta tk, k = 0,1,..., N-1, \quad (4.15)$$

$$y_k = y(t_k), t_k = \Delta tk, k = 0,1,..., N-1; \quad (4.16)$$

Ниже мы опишем основные шаги алгоритма:

$$x_k = A_1 \cos(2\pi \upsilon_1 t_k + \phi_1) + \sigma_x \xi_k, \tag{4.17}$$

$$y_k = \sum_{i=1}^{2} B_i \cos(2\pi \upsilon_i t_k + \phi_i) + \sigma_y \eta_k, \tag{4.18}$$

$$t_k = \Delta t k, k = 0,1,..., N-1,$$

где Δt – постоянный шаг выборки; A_1, υ_1, ϕ_1 – амплитуда, частота и фаза ряда x_k; $B_i, \upsilon_i, \phi_i, i = 1,2,$ – амплитуды, частоты и фазы ряда y_k; $\xi_k, \eta_k, k = 0,1,2,..., N-1$ – случайные величины с нулевым математическим ожиданием и единичной дисперсией, распределенные по нормальному закону (шумовые компоненты рядов x_k и y_k); σ_x, σ_y – среднеквадратические отклонения шумовых компонентов рядов x_k и y_k:

$$\sigma_x = \sqrt{\frac{A_1^2}{2\gamma_x}}, \ \sigma_y = \sqrt{\frac{B_1^2 + B_2^2}{2\gamma_y}}, \tag{4.19}$$

где γ_x, γ_y – отношения «сигнал к шуму» соответственно для рядов x_k и y_k.

Алгоритм кросс-спектрального анализа состоит из следующих шагов.

1. Графическое представление исходных рядов во временной области.

2. Центрирование рядов:

$$\begin{aligned} x_k^0 &= x_k - m_x, k = 0,1,..., N-1, \\ y_k^0 &= x_k - m_y, k = 0,1,..., N-1, \end{aligned} \tag{4.20}$$

где $m_x = \frac{1}{N} \sum_{k=0}^{N-1} x_k$; $m_y = \frac{1}{N} \sum_{k=0}^{N-1} y_k$.

3. Графическое представление центрированных рядов.

4. Дополнение рядов нулями.

5. Вычисление взаимной периодограммы.

Сначала вычисляются дискретные преобразования Фурье по формулам

$$X_j = FFT_j\left\{x_k^0\right\}_{k=0}^{N_2-1} = \sum_{k=0}^{N_2-1} x_k^0 e^{-i\frac{2\pi}{N_2}kj},\ j = 0,1,\ldots,N_2-1, \qquad (4.21)$$

$$Y_j = FFT_j\left\{y_k^0\right\}_{k=0}^{N_2-1} = \sum_{k=0}^{N_2-1} y_k^0 e^{-i\frac{2\pi}{N_2}kj},\ j = 0,1,\ldots,N_2-1, \qquad (4.22)$$

после чего взаимная периодограмма определяется по формуле

$$D_{xy}[j] = \frac{1}{N^2} X_j^* Y_j,\ j = 0,1,\ldots,N_2/2. \qquad (4.23)$$

Отсчеты взаимной периодограммы соответствуют частотам

$\upsilon_j = \Delta\upsilon j,\ j = 0,1,\ldots,N_2/2$,

где

$$\Delta\upsilon = \frac{1}{N_2\Delta t}. \qquad (4.24)$$

6. Вычисление коспектра и квадратурного спектра:

$$P_{xy}[j] = \mathrm{Re}\, D_{xy}[j];\quad Q_{xy}[j] = -\mathrm{Im}\, D_{xy}[j]. \qquad (4.25)$$

7. Вычисление квадрата амплитуды взаимной периодограммы:

$$A_{xy}^2[j] = P_{xy}^2[j] + Q_{xy}^2[j]. \qquad (4.26)$$

8. Вычисление фазового спектра:

$$\Phi_{xy}[j] = -arctg\frac{Q_{xy}[j]}{P_{xy}[j]}. \qquad (4.27)$$

9. Оценивание дисперсий рядов:

$$\sigma_x^2 = \frac{1}{N-1}\sum_{k=0}^{N-1}(x_k^0)^2,\quad \sigma_y^2 = \frac{1}{N-1}\sum_{k=0}^{N-1}(y_k^0)^2. \qquad (4.28)$$

10. Графическое представление квадрата амплитуды взаимной периодограммы и порога обнаружения сигнала.

Этот график позволяет отождествить значимые спектральные линии, общие для двух рядов. Все пики квадрата амплитуды взаимной

периодограммы, превышающие критический уровень $\sigma_x^2\sigma_y^2 Z_1 / N^2$, считаются значимыми, т.е. принадлежащими общему детерминированному компоненту рядов.

11. Графическое представление фазового спектра.

Можно напечатать значения фазового спектра для всех значений j, однако целесообразно получить фазовый спектр только для тех частот, на которых обнаружены общие значимые гармоники. Другими словами, на печать выводятся только те значения фазового спектра, для которых выполняются соотношения

$$A_{xy}^2[j] \geq \frac{\sigma_x^2\sigma_y^2}{N^2} Z_1. \tag{4.29}$$

12. Вычисление взаимных коррелограмм.

$$\overline{c}_{xy}[m] = \frac{1}{N} \operatorname{Re} FFT_m^{-1}\left\{X_j^* Y_j\right\}_{j=0}^{N_2-1}, \tag{4.30}$$

$$\overline{c}_{yx}[m] = \frac{1}{N} \operatorname{Re} FFT_m^{-1}\left\{X_j Y_j^*\right\}_{j=0}^{N_2-1}, \tag{4.31}$$

$m = 0,1,2,...,N^* - 1.$

13. *Графическое представление коррелограмм.*

3. Кросс-корреляция

При коррелировании ВР часто имеет место взаимное влияние процессов при условии сдвига временных серий друг относительно друга на некоторый временной промежуток, называемый лагом, т.е. влияние одного явления на другое проявляется с некоторым запаздыванием или опережением. Для определения наличия временного лага необходимо проанализировать ряды динамики и выявить их характерные черты. В большинстве случаев нельзя судить о наличии временного лага и тем более о конкретной его длине.

Взаимная корреляционная функция определяется для двух стационарных временных рядов как коэффициент корреляции между x_t и y_{t+k} в зависимости от лага k:

$$r_k = \frac{\sum_{t=1}^{n-k} X_t Y_{t+k} - \sum_{t=1}^{n-k} Y_t \sum_{t=1}^{n-k} \frac{X_{t+k}}{n-k}}{\sqrt{\left[\sum_{t=k}^{n-k} Y_t^2 - \sum_{t=k}^{n-k} \frac{Y_t^2}{n-k}\right]\left[\sum_{t=k+1}^{n} X_t^2 - \sum_{t=k+1}^{n} \frac{X_t^2}{n-k}\right]}}. \quad (4.32)$$

Ряд $r_k = r(k)$ представляет собой таблично заданную корреляционную функцию, «затухающую» достаточно быстро. Наличие пиков в функции указывает на наличие временного лага. Если пики в функции $r(k)$ повторяются через определенное время, то взаимное влияние рядов носит периодический характер.

Контрольные вопросы

1. Что такое многомерный временной ряд?
2. Опишите кросс-спектральный анализ временных рядов.
3. Что собой представляют компоненты кросс-спектрального анализа: кросс-периодограмма, коспектр, квадратурный спектр, кросс-амплитуда, фазовый сдвиг, усиление и квадрат когерентности?
4. Что такое кросс-корреляции?
5. Что такое лаг?

ПРАКТИЧЕСКАЯ ЧАСТЬ

1. Примеры решения задач по анализу временных рядов

Пример 1. Имеются данные об уровне безработицы (в %) за 10 месяцев:

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	8,9	8,6	8,8	8,4	8,2	8,3	7,9	7,6	7,2	7,4

1) Определить коэффициенты автокорреляции уровней ряда первого и второго порядка.
2) Выбрать уравнение тренда и определить его параметры с помощью МНК.

Решение.

1) Составим таблицу для расчета коэффициента автокорреляции первого порядка.

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	8,9	-	-	-	-	-	-
2	8,6	8,9	0,556	0,689	0,3827	0,3086	0,4746
3	8,8	8,6	0,756	0,389	0,2938	0,5709	0,1512
4	8,4	8,8	0,356	0,589	0,2094	0,1264	0,3468
5	8,2	8,4	0,156	0,189	0,0294	0,0242	0,0357
6	8,3	8,2	0,256	-0,011	-0,0028	0,0653	0,0001
7	7,9	8,3	-0,144	0,089	-0,0128	0,0209	0,0079
8	7,6	7,9	-0,444	-0,311	0,1383	0,1975	0,0968
9	7,2	7,6	-0,844	-0,611	0,516	0,7131	0,3735
10	7,4	7,2	-0,644	-1,011	0,6516	0,4153	1,0223
Σ	81,3	73,9	2,665E-15	-7,994E-15	2,2056	2,4422	2,5089

Средние значения уровней ряда равны:

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1} = \frac{8{,}6+8{,}8+8{,}4+8{,}2+8{,}3+7{,}9+7{,}6+7{,}2+7{,}4}{10-1} = \frac{72{,}4}{9} = 8{,}044,$$

$$\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{8{,}9+8{,}6+8{,}8+8{,}4+8{,}2+8{,}3+7{,}9+7{,}6+7{,}2}{10-1} = \frac{73{,}9}{9} = 8{,}211.$$

Определим коэффициент корреляции между уровнями y_t и y_{t-1} по формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t-\bar{y}_1)(y_{t-1}-\bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t-\bar{y}_1)^2 \sum_{t=2}^{n}(y_{t-1}-\bar{y}_2)^2}}.$$

Получим коэффициент автокорреляции первого порядка

$$r_1 = \frac{2{,}2056}{\sqrt{2{,}4422 \cdot 2{,}5089}} = 0{,}891.$$

Полученное значение свидетельствует об очень тесной зависимости между уровнем безработицы текущего и предыдущего месяцев и, следовательно, о наличии во временном ряду сильной линейной тенденции.

Определим коэффициент автокорреляции второго порядка. Построим таблицу:

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	8,9	-	-	-	-	-	-
2	8,6	-	-	-	-	-	-
3	8,8	8,9	0,825	0,563	0,4641	0,6806	0,3164
4	8,4	8,6	0,425	0,262	0,1116	0,1806	0,0689
5	8,2	8,8	0,225	0,463	0,1041	0,0506	0,2139
6	8,3	8,4	0,325	0,063	0,0203	0,1056	0,0039
7	7,9	8,2	-0,075	-0,138	0,0103	0,0056	0,0189
8	7,6	8,3	-0,375	-0,037	0,0141	0,1406	0,0014
9	7,2	7,9	-0,775	-0,438	0,3391	0,6006	0,1914
10	7,4	7,6	-0,575	-0,738	0,4241	0,3306	0,5439
Σ	81,3	66,7	-2,665E-15	-1,776E-15	1,4875	2,0950	1,3588

Средние значения уровней ряда равны:

$$\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2} = \frac{8{,}8+8{,}4+8{,}2+8{,}3+7{,}9+7{,}6+7{,}2+7{,}4}{10-2} = \frac{63{,}8}{8} = 7{,}975,$$

$$\bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2} = \frac{8{,}9+8{,}6+8{,}8+8{,}4+8{,}2+8{,}3+7{,}9+7{,}6}{10-2} = \frac{66{,}7}{8} = 8{,}3375.$$

Определим коэффициент автокорреляции второго порядка:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t-\bar{y}_3)(y_{t-2}-\bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t-\bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2}-\bar{y}_4)^2}} = \frac{1{,}4875}{\sqrt{2{,}0950 \cdot 1{,}3588}} = 0{,}882.$$

Полученные результаты еще раз подтверждают, что ряд уровня безработицы содержит линейную тенденцию.

2) Построим линейный тренд: $\hat{y}_t = a + bt$.

Методом наименьших квадратов найдем оценки параметров тренда по формулам:

$$b = \frac{\overline{ty} - \bar{t} \cdot \bar{y}}{\overline{t^2} - (\bar{t})^2};$$

$$a = \bar{y} - b\bar{t}.$$

Проведем следующие расчеты:

$$\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n} = \frac{1+2+3+4+5+6+7+8+9+10}{10} = \frac{55}{10} = 5{,}5;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{8{,}9+8{,}6+8{,}8+8{,}4+8{,}2+8{,}3+7{,}9+7{,}6+7{,}2+7{,}4}{10} = \frac{81{,}3}{10} = 8{,}13;$$

$$\overline{ty} = \frac{\sum_{i=1}^{n} (t_i \cdot y_i)}{n} = \frac{1 \cdot 8{,}9+2 \cdot 8{,}6+3 \cdot 8{,}8+4 \cdot 8{,}4+5 \cdot 8{,}2+6 \cdot 8{,}3+7 \cdot 7{,}9+8 \cdot 7{,}6+9 \cdot 7{,}2+10 \cdot 7{,}4}{10} =$$

$$\frac{431{,}8}{10} = 43{,}18\ ,$$

$$\overline{t^2} = \frac{\sum_{i=1}^{n} t_i^2}{n} = \frac{1^2+2^2+3^2+4^2+5^2+6^2+7^2+8^2+9^2+10^2}{10} = \frac{385}{10} = 38{,}5.$$

Тогда оценки параметров тренда будут равны:

$$b = \frac{43{,}18 - 5{,}5 \cdot 8{,}13}{38{,}5 - 5{,}5^2} = -0.186;$$

$$a = \bar{y} - b\bar{t} = 8.13 + 0.186 \cdot 5.5 = 9.153.$$

Следовательно, уравнение линейного тренда имеет вид:

$$\hat{y}_t = 9{,}153 - 0{,}186t.$$

Пример 2. Имеются данные об объеме продаж за 10 месяцев:

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	12	13	14	11	12	16	20	20	21	23

1) Провести сглаживание временного ряда простой скользящей средней с длиной интервала сглаживания: а) 3, б) 4.
2) Построить линейный тренд временного ряда.
3) Спрогнозировать ожидаемый объем продаж на 2 месяца вперед.

Решение.

1) а) Сглаживание простой скользящей средней проведем по формуле (1.24), так как интервал сглаживания является нечетным числом:

$$y'_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}.$$

При этом для первого и последнего периодов вычислить сглаженные значения невозможно. Для остальных месяцев получим:

$$y'_2 = \frac{12 + 13 + 14}{3} = 13,$$

$$y'_3 = \frac{13 + 14 + 11}{3} = 12{,}7,$$

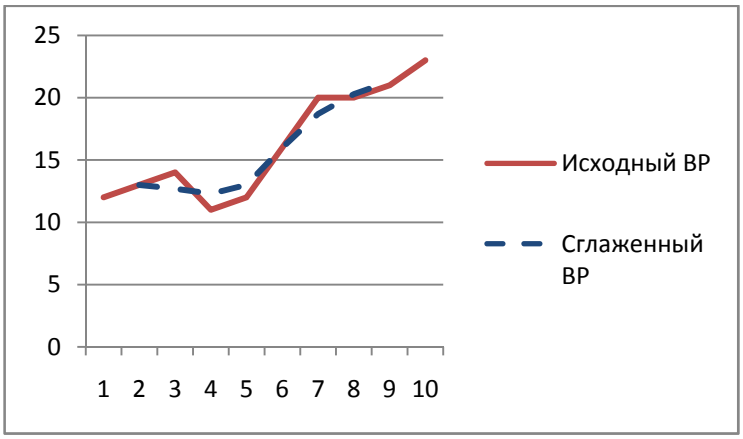
…,

$$y'_9 = \frac{20 + 21 + 23}{3} = 21{,}3.$$

Запишем полученные результаты в таблицу:

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	12	13	14	11	12	16	20	20	21	23
y'_t	-	13	12,7	12,3	13	16	18,7	20,3	21,3	-

Нанесем полученный сглаженный ряд на график исходного ВР.



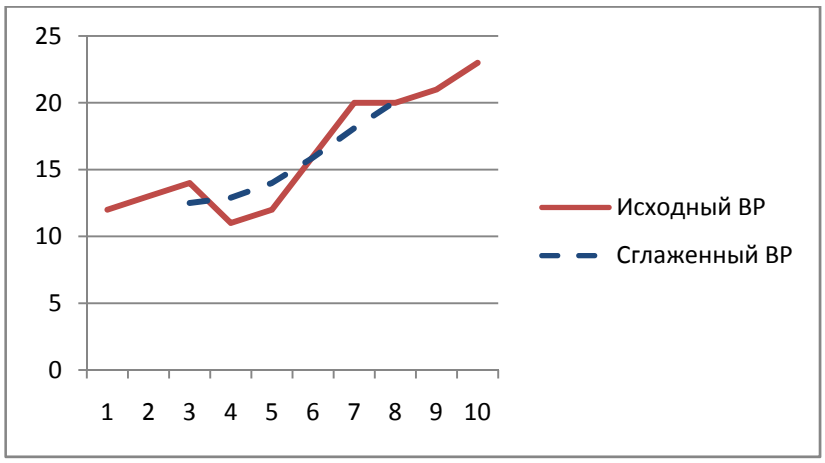
1) б) Сглаживание простой скользящей средней проведем по формуле (1.25), так как интервал сглаживания является четным числом:

$$y'_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

Запишем полученные результаты в таблицу:

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	12	13	14	11	12	16	20	20	21	23
y'_t	-	-	12,5	12,9	14	15,9	18,1	20,1	-	-

График для исходного и сглаженного ВР имеет вид:



2) Построим линейный тренд: $\hat{y}_t = a + bt$.

Методом наименьших квадратов найдем оценки параметров тренда по формулам:

$$b = \frac{\overline{ty} - \bar{t} \cdot \bar{y}}{\overline{t^2} - (\bar{t})^2};$$

$$a = \bar{y} - b\bar{t}.$$

Проведем следующие расчеты:

$$\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n} = \frac{1+2+3+4+5+6+7+8+9+10}{10} = \frac{55}{10} = 5{,}5;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{12+13+\cdots+21+23}{10} = \frac{162}{10} = 16{,}2;$$

$$\overline{ty} = \frac{\sum_{i=1}^{n} (t_i \cdot y_i)}{n} = \frac{1 \cdot 12 + 2 \cdot 13 + \cdots + 9 \cdot 21 + 10 \cdot 23}{10} = \frac{999}{10} = 99{,}9\ ,$$

$$\overline{t^2} = \frac{\sum_{i=1}^{n} t_i^2}{n} = \frac{1^2+2^2+3^2+4^2+5^2+6^2+7^2+8^2+9^2+10^2}{10} = \frac{385}{10} = 38{,}5.$$

Тогда оценки параметров тренда будут равны:

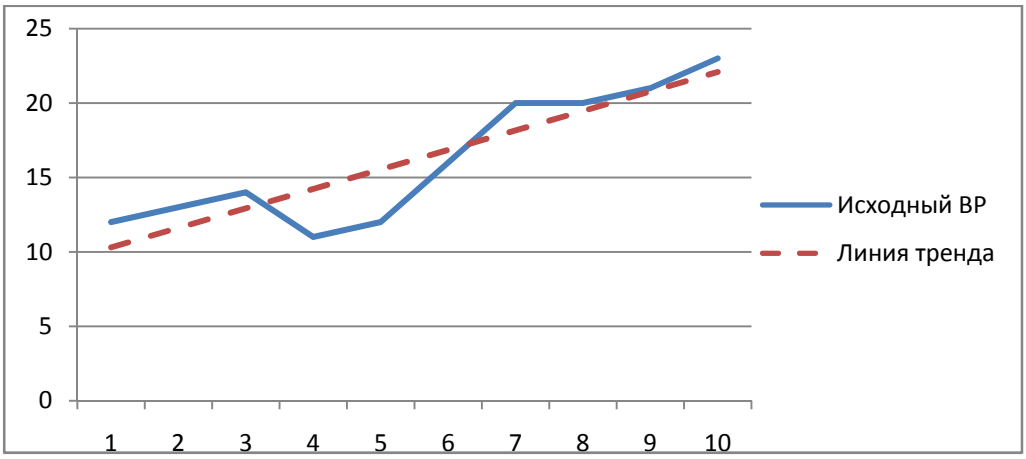
$$b = \frac{99{,}9 - 5{,}5 \cdot 16{,}2}{38{,}5 - 5{,}5^2} = 1{,}309;$$

$$a = \bar{y} - b\bar{t} = 16{,}2 - 1{,}309 \cdot 5{,}5 = 9.$$

Следовательно, уравнение линейного тренда имеет вид:

$$\hat{y}_t = 9 + 1{,}309t.$$

Нанесем полученный линейный тренд на график исходного ВР:



3) Спрогнозируем ожидаемый объем продаж на 2 месяца вперед, подставив в полученное уравнение тренда значения t =11 и t =12.

$\hat{y}_{11} = 23{,}4.$

$\hat{y}_{12} = 24{,}7.$

Пример 3. Имеются данные курса акций за 12 месяцев с января по декабрь.

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
y_t	8,9	8,6	8,8	8,4	8,2	8,3	7,9	7,6	7,2	7,3	7,2	7,4

1) Построить модель авторегрессии первого порядка.

2) Найти прогноз курса акций на первый месяц следующего года.

Решение.

1) Построим модель авторегрессии первого порядка:

$y_t = \beta_1 y_{t-1} + \varepsilon_t$.

Построение уравнения авторегрессии первого порядка аналогично построению уравнения линейного тренда, где в качестве регрессора выступают предыдущие уровни ряда y_{t-1}, а в качестве отклика - y_t, при этом свободный коэффициент $a = 0$.

Найдем оценку параметра модели авторегрессии методом наименьших квадратов по формуле:

$$\beta_1 = \frac{\sum_{t=2}^{n} y_{t-1} \cdot y_t}{\sum_{t=2}^{n} y_{t-1}^2} = \frac{701{,}81}{714{,}44} = 0{,}982.$$

Следовательно, модель авторегрессии первого порядка будет иметь вид:

$\hat{y}_t = 0{,}982 y_{t-1}$.

2) Найдем прогноз курса акций на первый месяц следующего года по модели авторегрессии первого порядка.

Учитывая, что $y_{12} = 7{,}4$, прогнозируемое значение курса акций за 13-ый месяц составит:

$\hat{y}_{13} = 0{,}982 \cdot 7{,}4 = 7{,}27$.

Пример 4. Имеются данные ВВП США за 10 лет (с 1961-го по 1970-ый год).

Год	1	2	3	4	5	6	7	8	9	10
y_t	563	605	639	686	744	815	862	943	1020	1076

1) Построить уравнение парной линейной регрессии.
2) Найти остаточную дисперсию, оценки дисперсий параметров модели.
3) Вычислить среднюю ошибку аппроксимации.
4) Вычислить прогноз ВВП на 1 год.
5) Найти доверительный интервал для прогноза с вероятностью 0,95.

Решение.

1) Построим уравнение парной линейной регрессии.

Методом наименьших квадратов найдем оценки параметров тренда по формулам:

$$b = \frac{\overline{ty} - \bar{t} \cdot \bar{y}}{\overline{t^2} - (\bar{t})^2};$$

$$a = \bar{y} - b\bar{t}.$$

Проведем следующие расчеты:

$$\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n} = \frac{1+2+3+4+5+6+7+8+9+10}{10} = \frac{55}{10} = 5{,}5;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{563+605+\cdots+1020+1076}{10} = \frac{7953}{10} = 795{,}3;$$

$$\overline{ty} = \frac{\sum_{i=1}^{n} (t_i \cdot y_i)}{n} = \frac{1\cdot 563+2\cdot 605+\cdots+9\cdot 1020+10\cdot 1076}{10} = \frac{48562}{10} = 4856{,}2\ ,$$

$$\overline{t^2} = \frac{\sum_{i=1}^{n} t_i^2}{n} = \frac{1^2+2^2+3^2+4^2+5^2+6^2+7^2+8^2+9^2+10^2}{10} = \frac{385}{10} = 38{,}5.$$

Тогда оценки параметров тренда будут равны:

$$b = \frac{4856{,}2 - 5{,}5 \cdot 795{,}3}{38{,}5 - 5{,}5^2} = 58{,}43;$$

$$a = \bar{y} - b\bar{t} = 795{,}3 - 58{,}43 \cdot 5{,}5 = 473{,}93.$$

Следовательно, уравнение парной линейной регрессии имеет вид:

$$\hat{y}_t = 473{,}93 + 58{,}43t.$$

2) Найдем остаточную дисперсию, оценки дисперсий параметров модели.

Для определения остаточной дисперсии найдем значения остатков e_t как разность между наблюдаемыми значениями исходного ВР и прогнозируемыми по построенному уравнению линейной регрессии значениями, т. е. $e_t = y_t - \hat{y}_t$. Результаты запишем в таблицу:

Год	1	2	3	4	5	6	7	8	9	10
y_t	563	605	639	686	744	815	862	943	1020	1076
$\hat{y}_t$	532,36	590,79	649,22	707,65	766,08	824,52	882,95	941,38	999,81	1058,24
e_t	30,64	14,21	-10,22	-21,65	-22,08	-9,52	-20,95	1,62	20,19	17,76

Используя значения таблицы, получим значение остаточной дисперсии:

$$S^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{30{,}64^2 + 14{,}21^2 + \cdots + 17{,}76^2}{10-2} = \frac{3456{,}82}{8} = 432{,}1.$$

Для определения оценок дисперсий параметров модели проведем расчеты:

$$\sum_{i=1}^{n} t_i^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2 = 385,$$

$$\sum_{i=1}^{n}(t_i - \bar{t})^2 = (1-5{,}5)^2 + (2-5{,}5)^2 + \cdots + (10-5{,}5)^2 = 82{,}5.$$

Оценки дисперсий параметров модели определим по формулам:

$$S_a^2 = S^2 \frac{\sum_{i=1}^{n} t_i^2}{n\sum_{i=1}^{n}(t_i - \bar{t})^2} = 432{,}1\frac{385}{10 \cdot 82{,}5} = 201{,}65,$$

$$S_b^2 = S^2 \frac{1}{\sum_{i=1}^{n}(t_i - \bar{t})^2} = 432{,}1\frac{1}{82{,}5} = 5{,}24$$

3) Вычислим среднюю ошибку аппроксимации по формуле:

$$\bar{A} = \frac{1}{n} \cdot \left(\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) \cdot 100\%$$

$$= \frac{1}{10} \cdot \left(\left| \frac{30{,}64}{563} \right| + \left| \frac{14{,}21}{605} \right| + \cdots + \left| \frac{17{,}76}{1076} \right| \right) \cdot 100\% =$$

$$= \frac{1}{10} \cdot 0{,}229 \cdot 100\% = 2{,}29\%.$$

Ошибка аппроксимации говорит о хорошем качестве подбора уравнения регрессии, так как принимает значение ниже 10 %.

4) Вычислим прогноз ВВП на 1 год:

$$\hat{y}_{11} = 473{,}93 + 58{,}43 \cdot 11 = 1116{,}67.$$

5) Найдем доверительный интервал для прогноза с вероятностью 0,95.

Вычислим стандартную ошибку индивидуального значения прогноза, найденного в предыдущем пункте:

$$S_{\hat{y}_k} = S \cdot \sqrt{1 + \frac{1}{n} + \frac{(t_k - \bar{t})^2}{\sum_{i=1}^{n}(t_i - \bar{t})^2}} = \sqrt{432{,}1} \cdot \sqrt{1 + \frac{1}{10} + \frac{(11 - 5{,}5)^2}{82{,}5}}$$

$$= 25{,}17.$$

Из таблицы квантилей распределения Стьюдента (Приложение 2) найдем табличное значение *t*-статистики:

$t_{табл}(\alpha,n\text{-}2)= t_{табл}(0.05,10\text{-}2)= t_{табл}(0.05,8)=1{,}86$.

Нижняя и верхняя границы доверительного интервала будут соответственно равны:

$$\hat{y}_k - t_{табл}(\alpha, n-2) \cdot S_{\hat{y}_k} = 1116{,}67 - 1{,}86 \cdot 25{,}17 = 1069{,}85,$$

$$\hat{y}_k + t_{табл}(\alpha, n-2) \cdot S_{\hat{y}_k} = 1116{,}67 + 1{,}86 \cdot 25{,}17 = 1163{,}49.$$

Доверительный интервал для прогноза ВВП с доверительной вероятностью 0,95 имеет вид:

$I_{0.95}$=(1069,85;1163,49).

Пример 5. Имеются данные о рождаемости в России (в млн человек) за 10 лет.

Год	1	2	3	4	5	6	7	8	9	10
y_t	1,214	1,266	1,311	1,396	1,477	1,502	1,457	1,479	1,610	1,713

1) Построить уравнение парной линейной регрессии.
2) Построить таблицу дисперсионного анализа.
3) Оценить качество модели по критерию Фишера.
4) Найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Решение.

1) Построим уравнение парной линейной регрессии.

Проведем следующие расчеты:

$$\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n} = \frac{1+2+3+4+5+6+7+8+9+10}{10} = \frac{55}{10} = 5{,}5;$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{1{,}214+1{,}266+\cdots+1{,}713}{10} = \frac{14{,}425}{10} = 1{,}4425;$$

$$\overline{ty} = \frac{\sum_{i=1}^{n} (t_i \cdot y_i)}{n} = \frac{1\cdot 1{,}214+2\cdot 1{,}266+\cdots+10\cdot 1{,}713}{10} = \frac{83{,}311}{10} = 8{,}3311\ ,$$

$$\overline{t^2} = \frac{\sum_{i=1}^{n} t_i^2}{n} = \frac{1^2+2^2+3^2+4^2+5^2+6^2+7^2+8^2+9^2+10^2}{10} = \frac{385}{10} = 38{,}5.$$

Тогда оценки параметров тренда будут равны:

$$b = \frac{8{,}3311 - 5{,}5 \cdot 1{,}4425}{38{,}5 - 5{,}5^2} = 0{,}048;$$

$$a = \bar{y} - b\bar{t} = 1{,}4425 - 0{,}048 \cdot 5{,}5 = 1{,}1776.$$

Следовательно, уравнение парной линейной регрессии имеет вид:

$$\hat{y}_t = 1{,}1776 + 0{,}048t.$$

2) Построим таблицу дисперсионного анализа.

В уравнении линейного тренда присутствует только один фактор (время), следовательно, имеется одна степень свободы. Для остаточной дисперсии число степеней свободы равно n-2=10-2=8. Для полной дисперсии: n-1=10-1=9.

Определим суммы квадратов по формулам:

- остаточная сумма квадратов $SS_e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = 0{,}019$;

- полная сумма квадратов $SS = \sum_{i=1}^{n} (y_i - \bar{y})^2 = 0{,}21$;

- факторная сумма квадратов $SS_R = SS - SS_e = 0{,}21 - 0{,}019 = 0{,}191$.

Средние квадраты (дисперсии) рассчитываются как отношение суммы квадратов к соответствующему числу степеней свободы.

$$MS_R = \frac{0{,}191}{1} = 0{,}191, \quad S^2 = \frac{0{,}019}{8} = 0{,}0024.$$

Значение статистики Фишера определяется как отношение средних квадратов:

$F = \frac{MS_R}{S^2} = 79{,}58.$

Все полученные данные сведем в таблицу дисперсионного анализа:

Вариация *y*	Число степеней свободы	Сумма квадратов	Средний квадрат	F
Факторная	1	0,191	0,191	79,58
Остаточная	8	0,019	0,0024	
Полная	9	0,21		

3) Оценим качество модели по критерию Фишера.

Проверим гипотезу H_0: b=0, т. е. гипотезу о том, что фактор времени не влияет существенно на рождаемость.

Значение статистики Фишера возьмем из таблицы дисперсионного анализа:

$F_{факт}$=79,58.

Из таблицы квантилей распределения Фишера (Приложение 2) найдем табличное значение статистики Фишера

$F_{табл}(\alpha,k_1,k_2)= F_{табл}(0{,}05;1;8)=5{,}32$.

$F_{факт}$=79,58 > $F_{табл}$=5,32, следовательно, гипотеза H_0 отвергается, и уравнение линейного тренда признается значимым в целом.

Для оценки адекватности модели проведем сравнение:

$F_{факт}$=79,58 > 4·$F_{табл}$=21,28, следовательно, модель признается значимой и пригодной для прогноза.

4) Найдем коэффициент множественной корреляции и коэффициент множественной детерминации.

Пользуясь таблицей дисперсионного анализа, вычислим значение коэффициента корреляции по формуле:

$$R = (\text{знак } b)\sqrt{\frac{SS_R}{SS}} = +\sqrt{\frac{0{,}191}{0{,}21}} = 0{,}954.$$

Полученное значение коэффициента множественной корреляции близко к 1, что свидетельствует о сильной степени линейной зависимости рождаемости от времени.

Коэффициент множественной детерминации $R^2 = 0{,}954^2 = 0{,}91$. Коэффициент детерминации показывает, что 91 % дисперсии переменной *y* объясняется влиянием фактора времени.

Пример 6. Имеются данные по объему экспорта России за 12 кварталов.

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	82,6	83,8	84,4	102,3	101,5	108,8	110,4	122,4	102,6	101,5	102,9	110,8

Построить модели зависимости ВР от времени:

а) квадратичную;

б) степенную.

Решение.

а) Квадратичным называется тренд, выраженный полиномом второй степени с уравнением

$$\boldsymbol{\hat{y}_t = a_0 + a_1 t + a_2 t^2}.$$

Оценивание параметров квадратичного тренда также осуществляется с помощью МНК. Минимизация суммы квадратов отклонений приводит к построению системы нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum t_i + a_2 \sum t_i^2 = \sum y_i; \\ a_0 \sum t_i + a_1 \sum t_i^2 + a_2 \sum t_i^3 = \sum t_i y_i; \\ a_0 \sum t_i^2 + a_1 \sum t_i^3 + a_2 \sum t_i^4 = \sum t_i^2 y_i. \end{cases}$$

Найдем соответствующие суммы, получаем следующую систему нормальных уравнений:

$$\begin{cases} 12a_0 + 78a_1 + 650a_2 = 1214; \\ 78a_0 + 650a_1 + 6084a_2 = 8224{,}8; \\ 650a_0 + 6084a_1 + 60710a_2 = 69378{,}4. \end{cases}$$

Решая эту систему уравнений, получаем оценки параметров квадратичной регрессии. Полученное уравнение квадратичного тренда имеет вид:

$$\hat{y} = 69{,}645 + 9{,}341t - 0{,}539t^2.$$

б) Рассмотрим степенную модель тренда:

$\hat{y}_t = at^b$.

Проведем процедуру логарифмирования:

$ln\hat{y}_t = \ln(at^b)$,

$ln\hat{y}_t = lna + \ln(t^b)$;

$ln\hat{y}_t = lna + bln\, t$.

Проведем замену переменных:

$y'_t = ln\hat{y}_t; t' = lnt; a' = lna$.

В результате получим линейное уравнение:

$y'_t = a' + bt'$.

Преобразуем исходную таблицу данных: вместо исходных значений ВР и переменной *t* находим их натуральные логарифмы:

t'	0	0,693	1,099	1,386	1,609	1,792	1,946	2,079	2,197	2,303	2,398	2,485
y'_t	4,414	4,428	4,436	4,628	4,620	4,690	4,704	4,807	4,631	4,620	4,634	4,708

Применяем к полученной модели МНК и находим параметры по формулам (2.3) и (2.4):

b=0,132, a'=4,39.

Параметр a степенной модели определяется как $a = e^{a'} = 80{,}64$. Получаем уравнение степенного тренда:

$\hat{y}_t = 80{,}64 \cdot t^{0{,}132}$.

Пример 7. Имеются данные цены 1 барреля нефти за 12 месяцев:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	49,52	49,56	47,69	48,79	44,44	37,22	34,24	36,81	38,67	45,83	49,72	50,35

Построить модели зависимости ВР от времени:

а) экспоненциальную;

б) потенциальную.

Решение.

а) Рассмотрим экспоненциальный тренд вида:

$\hat{y}_t = e^{a+bt}$.

Прологарифмируем обе части уравнения:

$ln\hat{y}_t = \ln(e^{a+bt})$.

Используя свойства логарифмов, получаем:

$ln\hat{y}_t = (a + bt)\ln e$.

$ln\hat{y}_t = a + bt$.

Сделаем замену переменной $y'_t = ln\hat{y}_t$. В результате получим линейное уравнение:

$y'_t = a + bt$.

Так как в модели используются не исходные уровни ВР, а их логарифмы, нужно преобразовать исходную таблицу данных, рассчитав натуральные логарифмы значений ВР. Получаем таблицу:

t	1	2	3	4	5	6	7	8	9	10	11	12
lny_t	3,902	3,903	3,865	3,888	3,794	3,617	3,533	3,606	3,655	3,825	3,906	3,919

Применяем к полученным данным МНК и находим параметры модели по формулам (2.3) и (2.4).

b=-0,0066, a=3,83.

Получаем модель экспоненциального тренда:

$\hat{y}_t = e^{3,83-0,0066t}$.

б) Рассмотрим потенциальную модель вида:

$\hat{y}_t = ab^t$.

Проведем процедуру логарифмирования:

$ln\hat{y}_t = \ln(ab^t)$,

$ln\hat{y}_t = lna + \ln(b^t)$;

$ln\hat{y}_t = lna + tln\, b$.

Проведем замену переменных:

$y'_t = ln\hat{y}_t; b' = lnb; a' = lna$.

В результате получим линейное уравнение:

$y'_t = a' + b't$.

Преобразуем исходную таблицу данных: вместо исходных значений ВР находим их натуральные логарифмы.

Получаем таблицу:

t	1	2	3	4	5	6	7	8	9	10	11	12
lny_t	3,902	3,903	3,865	3,888	3,794	3,617	3,533	3,606	3,655	3,825	3,906	3,919

Применяем к полученным данным МНК и находим параметры модели по формулам (2.3) и (2.4).

b'=-0,0066, a'=3,83.

Определим коэффициенты потенциальной модели:

$b = e^{b'} = 0,99, \qquad a = e^{a'} = 46,06$.

Получаем модель тренда:

$\hat{y}_t = 46,06 \cdot 0,99^t$.

Пример 8. Имеются данные по курсу доллара США за 12 месяцев:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	60,35	66,72	65,74	63,8	66,74	72,93	76,33	75,9	67,86	66,17	66	64,18

Построить модели зависимости ВР от времени:

а) логарифмическую;

б) гиперболическую.

а) Рассмотрим логарифмическую модель тренда:

$\hat{y}_t = a + b \cdot lnt.$

Чтобы привести ее к линейному виду, достаточно произвести замену переменной z = lnt и преобразовать исходную таблицу данных. Получаем линейный тренд:

$\hat{y}_t = a + bz,$

параметры которого оцениваются обычным МНК.

Преобразованная таблица имеет вид:

z	0	0,693	1,099	1,386	1,609	1,792	1,946	2,079	2,197	2,303	2,398	2,485
y_t	60,35	66,72	65,74	63,8	66,74	72,93	76,33	75,9	67,86	66,17	66	64,18

Применяем к полученным данным МНК и находим параметры модели по формулам (2.3) и (2.4).

b=2,69, a=63,25.

Получаем модель тренда:

$\hat{y}_t = 63{,}25 + 2{,}69 \cdot lnt.$

б) Рассмотрим гиперболическую модель тренда:

$\hat{y}_t = a + \frac{b}{t}.$

Чтобы привести ее к линейному виду, достаточно произвести замену переменной z = $\frac{1}{t}$ и преобразовать исходную таблицу данных. Получаем линейный тренд:

$$\hat{y}_t = a + bz,$$

параметры которого оцениваются обычным МНК.

Преобразованная таблица имеет вид:

z	1	0,5	0,333	0,25	0,2	0,167	0,143	0,125	0,111	0,1	0,091	0,083
y_t	60,35	66,72	65,74	63,8	66,74	72,93	76,33	75,9	67,86	66,17	66	64,18

Применяем к полученным данным МНК и находим параметры модели по формулам (2.3) и (2.4).

b=-9,44, a=70,17.

Получаем модель тренда:

$$\hat{y}_t = 70{,}17 - \frac{9{,}44}{t}.$$

Пример 9. Имеются данные о величине доходов и расходов на душу населения за 10 лет.

Год	1	2	3	4	5	6	7	8	9	10
Доходы, x_t	15	10	12	16	18	20	22	22	23	25
Расходы, y_t	12	12	13	14	16	15	18	19	21	22

1) Определить ежегодные абсолютные приросты доходов и расходов и сделать выводы о тенденции развития каждого ряда.
2) Построить линейную модель расходов, используя первые разности уровней исходных динамических рядов.
3) Включить в линейную модель зависимости фактор времени.

Решение.

1) Определим ежегодные абсолютные приросты доходов и расходов по формулам:

$$\Delta y_t = y_t - y_{t-1},\ \ \Delta x_t = x_t - x_{t-1}.$$

Расчеты представим в виде таблицы:

x_t	Δx_t	y_t	Δy_t
15	-	12	
10	-5	12	0
12	2	13	1
16	4	14	1
18	2	16	2
20	2	15	-1
22	2	18	3
22	0	19	1
23	1	21	2
25	2	22	1

Абсолютные приросты не имеют четко выраженной направленности, они варьируют вокруг среднего уровня, следовательно, оба ряда характеризуются линейной тенденцией.

2) Построим линейную модель расходов, используя первые разности уровней исходных динамических рядов.

Построим линейную регрессионную модель, используя первые разности, т. е. модель вида:

$\Delta y = f(\Delta x) = a + b \cdot \Delta x.$

С помощью метода наименьших квадратов определяем коэффициенты модели. Конечная модель имеет вид:

$\Delta y = 0{,}96 + 0{,}135 \cdot \Delta x.$

3) Включим в линейную модель зависимости фактор времени. Модель имеет вид:

$$\hat{y} = a + b \cdot x + c \cdot t.$$

Применяя МНК, получим систему нормальных уравнений:

$$na + b\sum x_i + c\sum t_i = \sum y_i;$$
$$a\sum x_i + b\sum x_i^2 + c\sum x_i \cdot t_i = \sum x_i y_i;$$
$$a\sum t_i + b\sum x_i \cdot t_i + c\sum t_i^2 = \sum t_i\ y_i.$$

Подставив исходные значения, получаем систему уравнений:

$$10a + 183b + 55c = 162;$$
$$183a + 3571b + 1132c = 3115;$$
$$55a + 1132b + 385c = 988.$$

Уравнение регрессии имеет вид:

$$\hat{y} = 8{,}83 + 0{,}09 \cdot x + 1{,}04 \cdot t.$$

Пример 10. Имеются данные по смещению Северного полюса Земли за 10 дней.

День	1	2	3	4	5	6	7	8	9	10
y_t	0,046	0,042	0,04	0,037	0,036	0,033	0,032	0,03	0,029	0,028

1) Построить уравнение авторегрессии первого порядка.
2) Проверить остатки на наличие автокорреляции с помощью критерия Дарбина-Уотсона.

Решение.

1) Построим модель авторегрессии первого порядка:

$y_t = \beta_1 y_{t-1} + \varepsilon_t$.

Найдем оценку параметра модели авторегрессии методом наименьших квадратов по формуле

$$\beta_1 = \frac{\sum_{t=2}^{n} y_{t-1} \cdot y_t}{\sum_{t=2}^{n} y_{t-1}^2} = \frac{0{,}0113}{0{,}012} = 0{,}942.$$

Следовательно, модель авторегрессии первого порядка будет иметь вид:

$\hat{y}_t = 0{,}942 y_{t-1}$.

2) Проверим остатки на наличие автокорреляции с помощью критерия Дарбина-Уотсона. Для этого запишем в таблицу предсказанные значения ВР и остатки:

День	1	2	3	4	5	6	7	8	9	10
y_t	0,046	0,042	0,04	0,037	0,036	0,033	0,032	0,03	0,029	0,028
$\hat{y}$	-	0,0433	0,0396	0,0377	0,0349	0,0339	0,0311	0,0301	0,0283	0,0273
e_t	-	-0,0013	0,0004	-0,0007	0,0011	-0,0009	0,0009	-0,0001	0,0007	0,0007

Критерий Дарбина-Уотсона рассчитывается по формуле:

$$d = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} = \frac{0{,}0000172}{0{,}0000047} = 3{,}66.$$

Так как фактическое значение критерия близко к 4, можно считать, что автокорреляция в остатках характеризуется отрицательной величиной.

Пример 11. На основании данных об изменении индекса потребительских цен на товары и услуги за 20 месяцев, приведенных в таблице, проверить наличие аномальных наблюдений.

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	100,35	100,5	100,84	101,01	100,44	100,32	100,29	100,34	100,04	100,2
Месяц	11	12	13	14	15	16	17	18	19	20
y_t	99,76	99,84	100,13	100,28	100,36	100,4	100,33	100,55	100,83	100,27

Решение.

Используем метод Ирвина, основанный на определении λ_t-статистик по формуле:

$$\lambda_t = \left|\frac{y_t - y_{t-1}}{S_y}\right|,$$

где $S_y = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}$, $\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i$.

Проведем расчеты:

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \frac{1}{20} \cdot 2007{,}08 = 100{,}354;$$

$$S_y = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2} = \sqrt{\frac{1}{19} \cdot 1{,}77} = 0{,}305.$$

Расчетные значения λ_t запишем в таблицу:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8	λ_9	λ_{10}
-	0,492	1,115	0,557	1,869	0,393	0,098	0,164	0,984	0,525
λ_{11}	λ_{12}	λ_{13}	λ_{14}	λ_{15}	λ_{16}	λ_{17}	λ_{18}	λ_{19}	λ_{20}
1,443	0,262	0,951	0,492	0,262	0,131	0,230	0,721	0,918	1,836

Расчетные значения λ_2, λ_3, ... сравниваются с табличными значениями критерия Ирвина λ_α. Если оказывается, что расчетное значение λ_t больше табличного λ_α, то соответствующее значение y_t уровня ряда считается аномальным.

Сравнивая найденные значения коэффициента λ_t с табличным значением λ_α, равным 1,3 для уровня значимости $\alpha = 0{,}05$ и при $n = 20$ (число уровней временного ряда), получаем, что отдельные значения уровней ряда превосходят значение λ_α, следовательно, делаем вывод о том, что в данной модели присутствуют аномальные колебания, вызванные ошибками второго рода, которые устранению не подлежат.

2. Задачи для самостоятельного выполнения

Задача 1. Имеются данные об уровне инфляции (в %) за 10 лет:

Год	1	2	3	4	5	6	7	8	9	10
y_t	8,9	8,6	9,1	8,2	7,4	5,6	6,2	7,6	12	8,4

1) Определить коэффициенты автокорреляции уровней ряда первого и второго порядка.
2) Выбрать уравнение тренда и определить его параметры с помощью МНК.

Задача 2. Имеется ряд динамики курса швейцарского франка к российскому рублю за 14 дней:

День	1	2	3	4	5	6	7
y_t	65,43	65,23	64,46	65,08	65,08	65,08	65,24
День	8	9	10	11	12	13	14
y_t	65,36	64,61	65,09	64,70	64,70	64,70	64,49

1) Определить коэффициенты автокорреляции уровней ряда первого и второго порядка.
2) Выбрать уравнение тренда и определить его параметры с помощью МНК.

Задача 3. Имеются данные депозитов физических лиц в течение года:

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
y_t	12	10	14	13	15	16	12	20	19	21	25	23

1) Провести сглаживание временного ряда простой скользящей средней с длиной интервала сглаживания: а) 3, б) 4.
2) Построить линейный тренд временного ряда.
3) Спрогнозировать значения ВР на 2 месяца вперед.

Задача 4. Имеются ряды динамики курса английских фунтов стерлингов к российскому рублю за 14 дней:

День	1	2	3	4	5	6	7
y_t	82,26	82,42	81,55	82,07	82,07	82,07	82,11
День	8	9	10	11	12	13	14
y_t	82,51	81,76	82,54	82,78	82,78	82,78	82,57

1) Провести сглаживание временного ряда простой скользящей средней с длиной интервала сглаживания: а) 4, б) 5.
2) Построить линейный тренд временного ряда.
3) Спрогнозировать ожидаемый курс валюты на 2 дня вперед.

Задача 5. Имеются данные об уровне безработицы в регионе за 14 месяцев.

Месяц	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y_t	5,3	5,8	5,6	5,4	5,2	4,8	4,8	5,1	4,7	4,6	4,8	4,5	4,1	4,1

1) Построить модель авторегрессии первого порядка.
2) Найти прогноз уровня безработицы в 15-м месяце по модели авторегрессии.

Задача 6. Имеется ряд динамики средних расходов на душу населения в сфере медицинских и оздоровительных процедур и услуг (Y), в тыс. руб., за 9 месяцев.

Месяц	1	2	3	4	5	6	7	8	9
y_t	1,263	1,979	1,718	2,237	2,211	1,762	3,068	2,359	3,109

1) Построить модель авторегрессии первого порядка.
2) Найти прогноз среднедушевых расходов в сфере медицинских и оздоровительных процедур и услуг на 10-й месяц по модели авторегрессии.

Задача 7. Имеются данные уровня безработицы в США за 10 лет (с 1961-го по 1970-ый год).

Год	1	2	3	4	5	6	7	8	9	10
y_t	6,7	5,5	5,7	5,2	4,5	3,8	3,8	3,6	3,5	4,9

1) Построить уравнение парной линейной регрессии.
2) Найти остаточную дисперсию, оценки дисперсий параметров модели.
3) Вычислить среднюю ошибку аппроксимации.
4) Вычислить прогноз уровня безработицы на 1 год.
5) Найти доверительный интервал для прогноза с вероятностью 0,95.

Задача 8. Имеются данные заявленной потребности в работниках за 12 месяцев.

Год	1	2	3	4	5	6	7	8	9	10	11	12
y_t	1461	1512	1537	1572	1733	1764	1755	1744	1722	1651	1579	1488

1) Построить уравнение парной линейной регрессии.
2) Найти остаточную дисперсию, оценки дисперсий параметров модели.
3) Вычислить среднюю ошибку аппроксимации.
4) Вычислить прогноз на 1 месяц вперед.
5) Найти доверительный интервал для прогноза с вероятностью 0,95.

Задача 9. Имеются данные о рождаемости в Беларуси (в млн человек) за 10 лет.

Год	1	2	3	4	5	6	7	8	9	10
y_t	0,168	0,165	0,171	0,162	0,163	0,153	0,142	0,132	0,127	0,117

1) Построить уравнение парной линейной регрессии.
2) Построить таблицу дисперсионного анализа.
3) Оценить качество модели по критерию Фишера.

4) Найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 10. Имеются данные по объему экспорта России за 12 кварталов.

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	87,1	107,4	10,4	126	134,5	97,4	57,8	63,6	81,6	94,2	90,1	95,4

1) Построить уравнение парной линейной регрессии.
2) Построить таблицу дисперсионного анализа.
3) Оценить качество модели по критерию Фишера.
4) Найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 11. Имеются данные по средней заработной плате за 12 дней.

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	133	150	145	156	167	181	157	152	164	159	173	180

1) Построить модели зависимости ВР от времени:
   а) квадратичную;
   б) степенную;
2) Для каждой модели:
   а) построить таблицу дисперсионного анализа;
   б) проверить значимость и адекватность модели;
   в) найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 12. Имеются данные по солнечной активности (числа Вольфа) за 12 месяцев 2000 года.

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	90	113	139	125	122	125	170	131	110	99	107	104

1) Построить модели зависимости ВР от времени:

   а) квадратичную;

   б) степенную;

2) Для каждой модели:

   а) построить таблицу дисперсионного анализа;

   б) проверить значимость и адекватность модели;

   в) найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 13. Имеются данные геомагнитной активности Земли за 12 месяцев (1990 г.):

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	536	415	664	439	565	433	447	703	436	569	619	672

1) Построить модели зависимости ВР от времени:

   а) экспоненциальную;

   б) потенциальную;

2) Для каждой модели:

   а) построить таблицу дисперсионного анализа;

   б) проверить значимость и адекватность модели;

   в) найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 14. Имеются данные по солнечной активности (числа Вольфа) за 12 месяцев 2010 года:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	13,2	18,8	15,4	8	8,7	13,6	16,1	19,6	25,2	23,5	22	14

1) Построить модели зависимости ВР от времени:

   а) экспоненциальную;

   б) потенциальную;

2) Для каждой модели:

а) построить таблицу дисперсионного анализа;

б) проверить значимость и адекватность модели;

в) найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 15. Имеются данные по солнечной активности (поток радиоизлучения на длине волны 10,7) за 12 месяцев 200 года:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	647	625	650	636	629	612	610	612	611	610	604	603

1) Построить модели зависимости ВР от времени:

   а) логарифмическую;

   б) гиперболическую;

2) Для каждой модели:

   а) построить таблицу дисперсионного анализа;

   б) проверить значимость и адекватность модели;

   в) найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 16. Имеются данные по ценам на серебро за 12 дней:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	12,64	12,22	12,46	12,84	12,64	12,82	12,96	12,79	12,54	12,38	12,48	12,83

1) Построить модели зависимости ВР от времени:

   а) логарифмическую;

   б) гиперболическую;

2) Для каждой модели:

   а) построить таблицу дисперсионного анализа;

   б) проверить значимость и адекватность модели;

   в) найти коэффициент множественной корреляции и коэффициент множественной детерминации.

Задача 17. Имеются данные о величине доходов и расходов на душу населения за 10 лет

Год	1	2	3	4	5	6	7	8	9	10
Доходы	10	16	18	20	18	23	22	24	28	30
Расходы	12	18	15	20	22	27	19	24	23	26

1) Определить ежегодные абсолютные приросты доходов и расходов и сделать выводы о тенденции развития каждого ряда.
2) Построить линейную модель расходов, используя первые разности уровней исходных динамических рядов.
3) Включить в линейную модель зависимости фактор времени.

Задача 18. Имеются два временных ряда объемом 9 наблюдений:

Год	1	2	3	4	5	6	7	8	9
x_t	12	11	5	9	6	5	3	2	4
y_t	18	19	13	10	8	5	6	4	4

1) Определить ежегодные абсолютные приросты рядов и сделать выводы о тенденции развития каждого ряда.
2) Построить линейную зависимость y от x, используя первые разности уровней исходных динамических рядов.
3) Включить в линейную модель зависимости фактор времени.

Задача 19. Имеются данные по смещению Северного полюса Земли за 10 дней:

День	1	2	3	4	5	6	7	8	9	10
y_t	0,203	0,2037	0,2043	0,2051	0,2056	0,206	0,206	0,2065	0,2074	0,2085

1) Построить уравнение авторегрессии первого порядка.
2) Проверить остатки на наличие автокорреляции с помощью критерия Дарбина-Уотсона.

Задача 20. Имеются ежедневные данные по солнечной активности (числа Вольфа):

День	1	2	3	4	5	6	7	8	9	10
y_t	32	32	24	14	11	16	11	10	14	19

1) Построить уравнение авторегрессии первого порядка.
2) Проверить остатки на наличие автокорреляции с помощью критерия Дарбина-Уотсона.

Задача 21. На основании данных об изменении индекса потребительских цен на непродовольственные товары за 20 месяцев, приведенных в таблице, проверить наличие аномальных наблюдений.

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	100,45	100,39	100,24	100,64	100,26	100,25	100,19	100,21	100,17	100,17
Месяц	11	12	13	14	15	16	17	18	19	20
y_t	100,17	100,19	100,31	100,23	100,14	100,23	100,04	100,48	100,44	100,25

Задача 22. На основании данных об изменении индекса потребительских цен на продовольственные товары за 20 месяцев, приведенных в таблице, проверить наличие аномальных наблюдений.

Месяц	1	2	3	4	5	6	7	8	9	10
y_t	100,55	100,99	101,70	101,27	100,79	100,53	100,43	100,41	99,52	99,70
Месяц	11	12	13	14	15	16	17	18	19	20
y_t	99,09	99,56	100,18	100,45	100,65	100,69	100,57	100,96	100,72	100,16

Задача 23. Рассчитать показатели динамики временного ряда для числа опасных гидрологических явлений на территории России с 2010 по 2019 гг.

Год	1	2	3	4	5	6	7	8	9	10
y_t	8	17	21	16	7	4	15	13	12	17

Задача 24. Рассчитать показатели динамики временного ряда для количества браков на 1000 человек населения на территории России с 2010 по 2019 гг.

Год	1	2	3	4	5	6	7	8	9	10
y_t	8,5	9,2	8,5	8,8	8,4	7,9	6,7	7,1	6,1	6,5

Задача 25. Имеются данные по численности населения России (млн человек) за 11 лет:

Год	1	2	3	4	5	6	7	8	9	10	11
y_t	142,9	142,9	143	143,3	143,7	146,3	146,5	146,8	146,9	146,8	146,7

Проверить наличие тренда в ВР, используя:

а) критерий сравнения средних;

б) критерий Фостера-Стюарта;

в) критерий серий, основанный на медиане;

г) критерий восходящих и нисходящих серий.

Задача 26. Имеются данные по приему в аспирантуру в российские вузы за 10 лет (тыс. человек):

Год	1	2	3	4	5	6	7	8	9	10
y_t	139,9	139,5	131,2	117,8	107	97,8	87,2	82,6	79,6	72,5

Проверить наличие тренда в ВР, используя:

а) критерий сравнения средних;

б) критерий Фостера-Стюарта;

в) критерий серий, основанный на медиане;

г) критерий восходящих и нисходящих серий.

Задача 27. Имеются данные по коэффициенту рождаемости в России за 10 лет:

Год	1	2	3	4	5	6	7	8	9	10
y_t	2,111	2,194	2,130	2,007	1,887	1,732	1,552	1,385	1,288	1,198

1) Постройте прогноз ВР на 3 года вперед, используя методы: скользящей средней, экспоненциального сглаживания, наименьших квадратов.
2) Рассчитайте ошибки полученных прогнозов при использовании каждого метода.
3) Сравните полученные результаты, сделайте выводы.

Задача 28. Имеются данные по средней продолжительности жизни мужчин за 10 лет:

Год	1	2	3	4	5	6	7	8	9	10
y_t	61,3	62,5	63,5	64,5	64,3	64	63,2	63,2	62	59

1) Постройте прогноз ВР на 3 года вперед, используя методы: скользящей средней, экспоненциального сглаживания, наименьших квадратов.
2) Рассчитайте ошибки полученных прогнозов при использовании каждого метода.
3) Сравните полученные результаты, сделайте выводы.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

Для выполнения лабораторных работ студенты должны самостоятельно найти 2 взаимосвязанных временных ряда по 50 наблюдений. Временные ряды могут быть с любыми интервалами между наблюдениями (ежедневные, ежемесячные, квартальные и т.д.). Примеры возможных временных рядов: курсы валют, объемы продаж, урожайность сельхозкультур, интернет-трафик, координаты полюса Земли, данные солнечной активности, ежегодный прием студентов на первый курс и т.д.

Отчет по лабораторным работам должен содержать:

1. Титульный лист (образец представлен в приложении 1).
2. Описание исходных данных.
3. Краткие теоретические сведения.
4. Все расчеты, графики, формулы, требуемые для выполнения задания лабораторной работы.
5. Основные результаты и выводы.

Лабораторная работа № 1. Анализ временных рядов в MS Excel.

Задание к лабораторной работе № 1.
Для двух временных рядов:

1. Найти коэффициенты автокорреляции 1, 2 и 3 порядков.

2. Построить линейный тренд (зависимость от времени), используя пакет анализа данных в Excel. Проверить качество тренда и построить прогноз на 5 значений вперед. Построить диаграмму рассеяния.

3. Построить экспоненциальный, степенной и квадратичный тренды. Проверить качество построенных моделей.

4. Провести сглаживание, используя методы скользящего среднего и экспоненциального сглаживания.

5. Построить модели авторегрессии первого и второго порядка, используя пакет анализа данных.

Цель работы: в ходе выполнения лабораторной работы студент приобретает навыки анализа временных рядов в MS Excel с помощью автокорреляционного, дисперсионного, регрессионного анализов, построения различных графиков, применения Пакета анализа.

Пример выполнения.

Исходные данные:

Известны динамика среднедушевого дохода населения (X) в Ульяновской области за 2015–2019 гг. по месяцам в тыс. руб., а также средние расходы на душу населения в сфере медицинских и оздоровительных процедур и услуг (Y) в тыс. руб.

t	x	y
1	12.578	1.263
2	17.378	1.979
3	14.537	1.718
4	19.491	2.237
5	19.307	2.211
6	17.778	1.762
7	22.090	3.068
8	19.312	2.359
9	21.355	3.109
10	19.349	1.677
11	22.663	2.843
12	20.502	2.441
13	20.066	2.568
14	19.925	2.307
15	22.667	2.976
16	22.575	3.220
17	23.413	3.394
18	21.541	2.951
19	19.201	2.521
20	22.707	2.859
21	22.669	3.358
22	21.526	2.903
23	24.147	3.546
24	20.217	2.403
25	22.076	2.876

t	x	y
26	21.836	2.908
27	23.652	3.871
28	20.807	2.637
29	20.626	2.409
30	21.789	2.883
31	23.442	3.638
32	21.605	2.799
33	31.660	4.368
34	22.779	2.913
35	14.499	1.398
36	21.685	2.417
37	22.960	2.780
38	19.715	2.356
39	23.188	3.290
40	20.834	2.619
41	22.594	2.761
42	20.960	2.513
43	21.463	2.697
44	24.108	3.732
45	20.945	2.608
46	21.364	2.725
47	22.675	2.857
48	22.100	2.446
49	25.263	3.907
50	22.485	2.849

1. Рассчитаем коэффициенты автокорреляции 1, 2 и 3 порядков. Рассмотрим подробнее на примере ряда X. Добавим к столбцу X еще 3 столбца, в которых запишем исходный ряд со смещением на 1, 2, 3 лага.

t	xt	x(t-1)	x(t-2)	x(t-3)
1	12.578	-	-	-
2	17.378	12.578	-	-
3	14.537	17.378	12.578	-
4	19.491	14.537	17.378	12.578
5	19.307	19.491	14.537	17.378
6	17.778	19.307	19.491	14.537
7	22.090	17.778	19.307	19.491
8	19.312	22.090	17.778	19.307
9	21.355	19.312	22.090	17.778
10	19.349	21.355	19.312	22.090
11	22.663	19.349	21.355	19.312
12	20.502	22.663	19.349	21.355
13	20.066	20.502	22.663	19.349
14	19.925	20.066	20.502	22.663
15	22.667	19.925	20.066	20.502
16	22.575	22.667	19.925	20.066
17	23.413	22.575	22.667	19.925
18	21.541	23.413	22.575	22.667
19	19.201	21.541	23.413	22.575
20	22.707	19.201	21.541	23.413
21	22.669	22.707	19.201	21.541
22	21.526	22.669	22.707	19.201
23	24.147	21.526	22.669	22.707
24	20.217	24.147	21.526	22.669
25	22.076	20.217	24.147	21.526
26	21.836	22.076	20.217	24.147
27	23.652	21.836	22.076	20.217
28	20.807	23.652	21.836	22.076
29	20.626	20.807	23.652	21.836

30	21.789	20.626	20.807	23.652
31	23.442	21.789	20.626	20.807
32	21.605	23.442	21.789	20.626
33	31.660	21.605	23.442	21.789
34	22.779	31.660	21.605	23.442
35	14.499	22.779	31.660	21.605
36	21.685	14.499	22.779	31.660
37	22.960	21.685	14.499	22.779
38	19.715	22.960	21.685	14.499
39	23.188	19.715	22.960	21.685
40	20.834	23.188	19.715	22.960
41	22.594	20.834	23.188	19.715
42	20.960	22.594	20.834	23.188
43	21.463	20.960	22.594	20.834
44	24.108	21.463	20.960	22.594
45	20.945	24.108	21.463	20.960
46	21.364	20.945	24.108	21.463
47	22.675	21.364	20.945	24.108
48	22.100	22.675	21.364	20.945
49	25.263	22.100	22.675	21.364
50	22.485	25.263	22.100	22.675

Для использования электронных таблиц Excel при работе со статистическими методами могут применяться как обычные средства, такие как вставка функций (в первую очередь статистических), мастер диаграмм, так и специальные, в частности, надстройка «Пакет анализа».

В меню «Формулы» выбираем пункт «Вставить функцию». В появившемся окне (рис. 1) выбираем категорию функций – статистические.

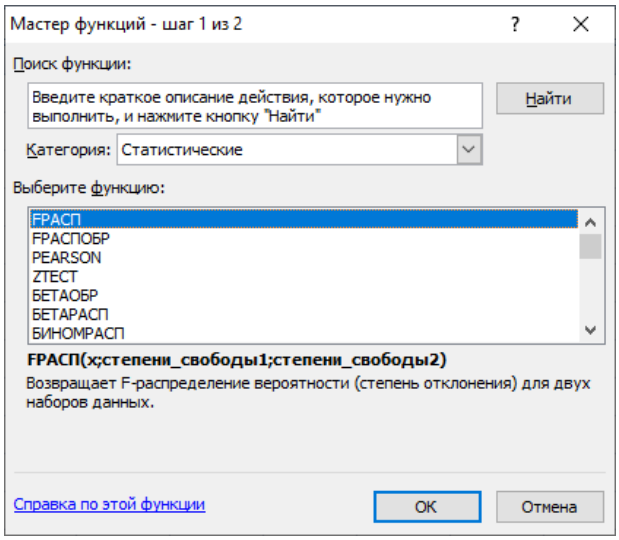


Рис. 1. Вставка статистических функций в MS Excel

Выбираем функцию КОРРЕЛ (рис. 2), вставляем диапазоны для двух характеристик, для которых ищется коэффициент корреляции, и нажимаем кнопку «ОК». Длины массивов должны совпадать.

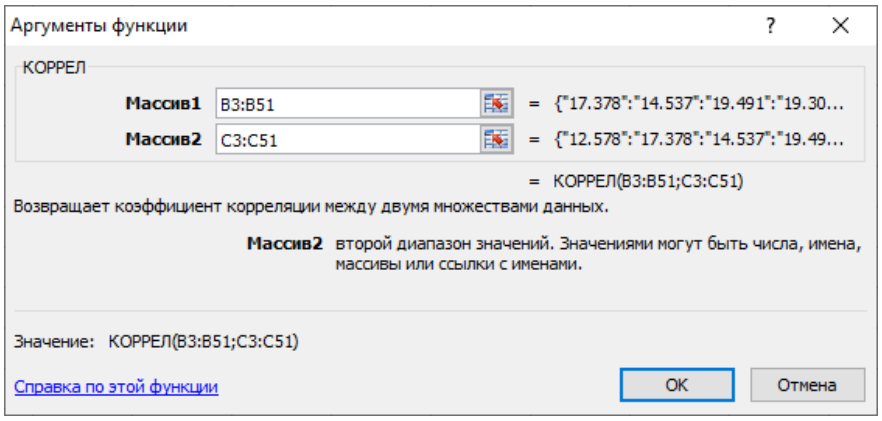


Рис. 2. Расчет коэффициента корреляции в MS Excel

Для ряда X получаем следующие значения автокорреляции:

Коэффициент автокорреляции 1 порядка	0,238804
Коэффициент автокорреляции 2 порядка	0,137095
Коэффициент автокорреляции 3 порядка	0,235299

Значения автокорреляции близки к 0, что говорит о слабой зависимости между текущими и предыдущими значениями ряда.

Аналогичные расчеты проводятся для ряда Y.

Коэффициент корреляции в Excel может быть также найден с помощью надстройки «Пакет анализа».

Подключение пакета «Анализ данных» в различных версиях MS Excel производится по-разному из-за различий графического интерфейса программы.

Для подключения пакета в MS Excel 2003 нужно выбрать в меню вкладку «Сервис»/«Надстройки». Появится окно «Надстройки», в котором надо установить флажок «Пакет анализа» и нажать кнопку «ОК» (рис. 3).

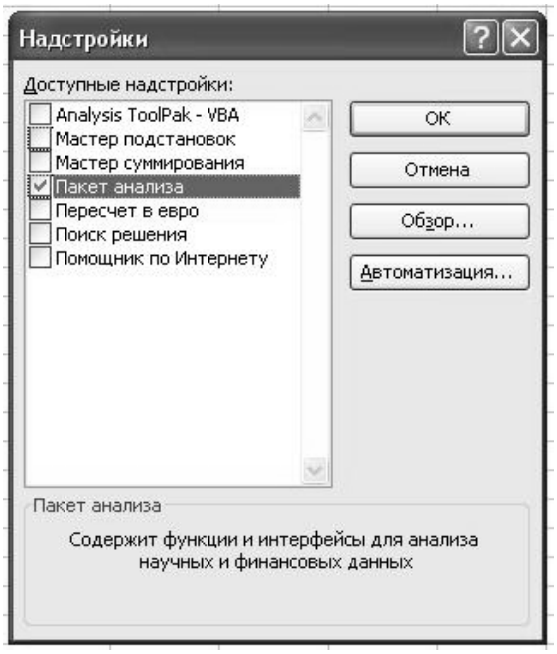


Рис. 3. Выбор надстройки «Пакет анализа»

Затем во вкладке главного меню «Сервис» появится позиция «Анализ данных», при выборе которой открывается новое окно (рис. 4).

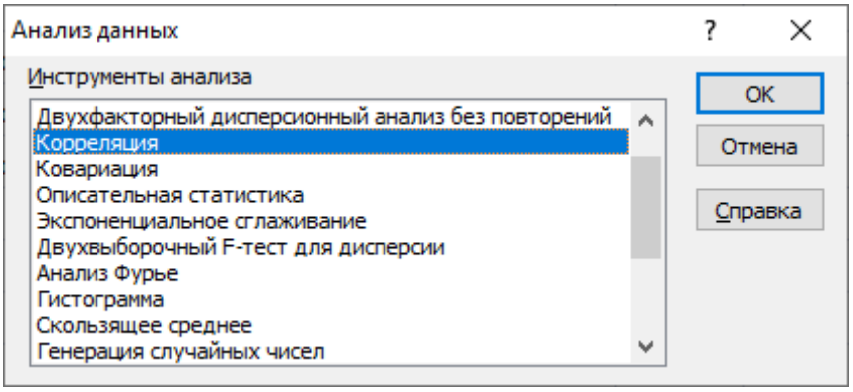


Рис. 4. Выбор инструментов пакета «Анализ данных»

Если надстройка «Пакет анализа» отсутствует в списке доступных надстроек (рис. 3), то для ее поиска следует нажать на кнопку «Обзор…».

Для подключения пакета в MS Excel 2007 требуется нажать на кнопку «Microsoft Office» в левом верхнем углу главного окна, а затем в выпавшем меню нажать на кнопку «Параметры Excel». В появившемся окне в списке команд выбрать пункт «Надстройки», в списке «Управление» выбрать позицию «Надстройки Excel» и нажать кнопку «Перейти …» (рис. 5). В появившемся списке нужно выделить пункт «Пакет анализа» и нажать кнопку «ОК».

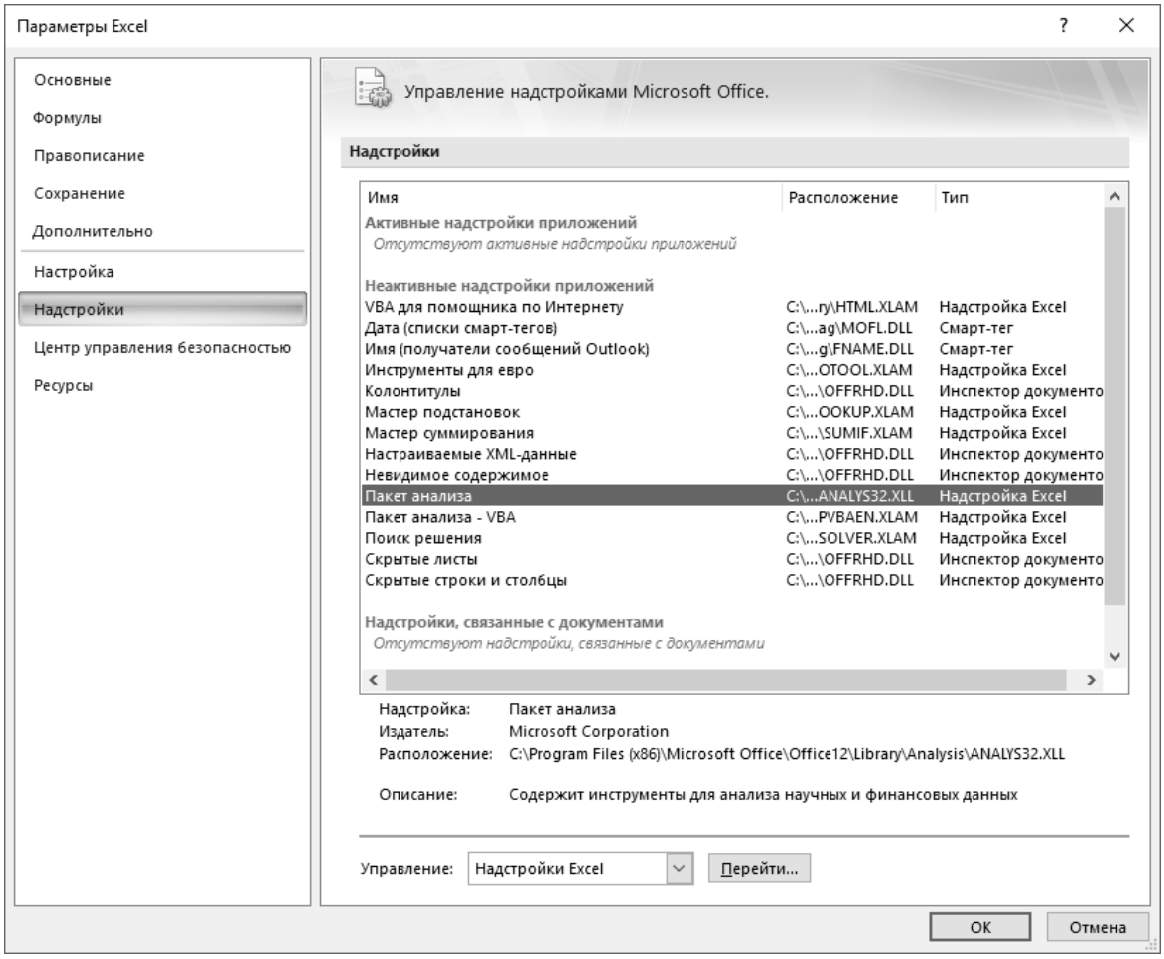


Рис. 5. Выбор меню «Надстройки»

В результате подключения надстройки на вкладке «Данные» в группе «Анализ» станет доступна команда «Анализ данных». При выборе этой команды будет открываться окно «Анализ данных» (4).

Для подключения пакета в MS Excel 2010 и старше нужно выбрать на вкладке «Файл» команду «Параметры» для открытия окна «Параметры Excel». Дальнейшие действия аналогичны действиям, описанным для MS Excel 2007.

Для определения значений коэффициентов корреляции между исходными и смещенными уровнями ряда используем инструмент Корреляция пакета анализа. На рис. 6 показано заполнение соответствующего диалогового окна.

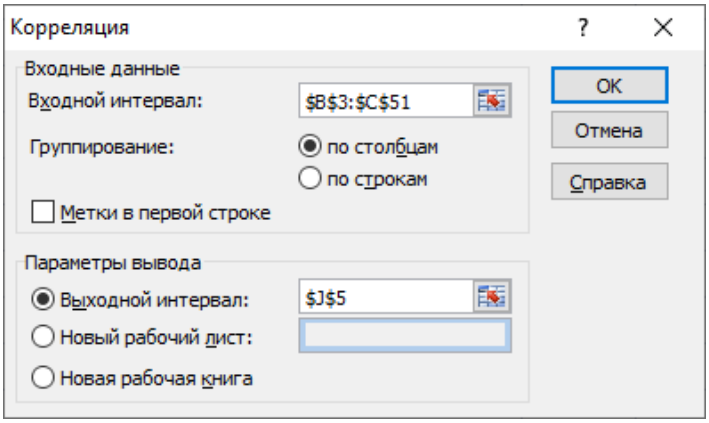


Рис. 6. Заполнение диалогового окна «Корреляция»

В результате получаем корреляционную матрицу вида

	Столбец 1	*Столбец 2*
Столбец 1	1	
Столбец 2	0,238803841	1

2. Существует несколько способов построения линейного тренда в MS Excel:

1) с помощью вставки статистической функции ЛИНЕЙН (рис.7).

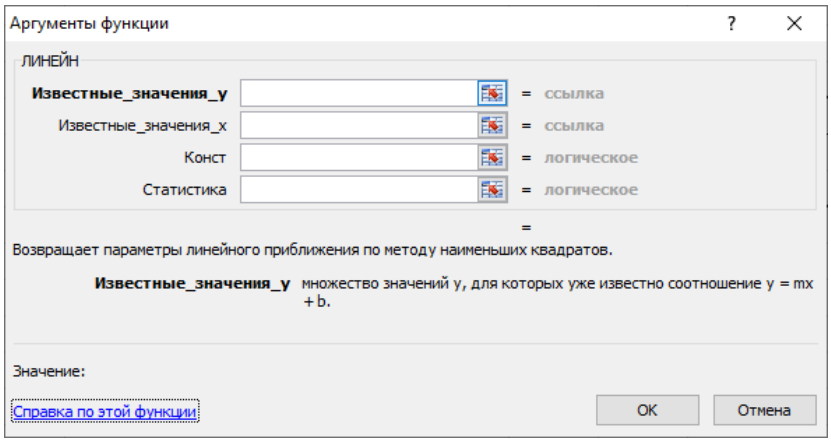


Рис. 7. Диалоговое окно функции «ЛИНЕЙН»

Для того чтобы вычислить оценки параметров *a* и *b*, нужно:

- выделить 2 ячейки в 1 строке;
- вставить функцию ЛИНЕЙН: в качестве *y* берутся значения ВР, в качестве x – моменты времени (значения от 1 до 50) (рис. 8)
- нажать одновременно клавиши CTRL+SHIFT+Enter.

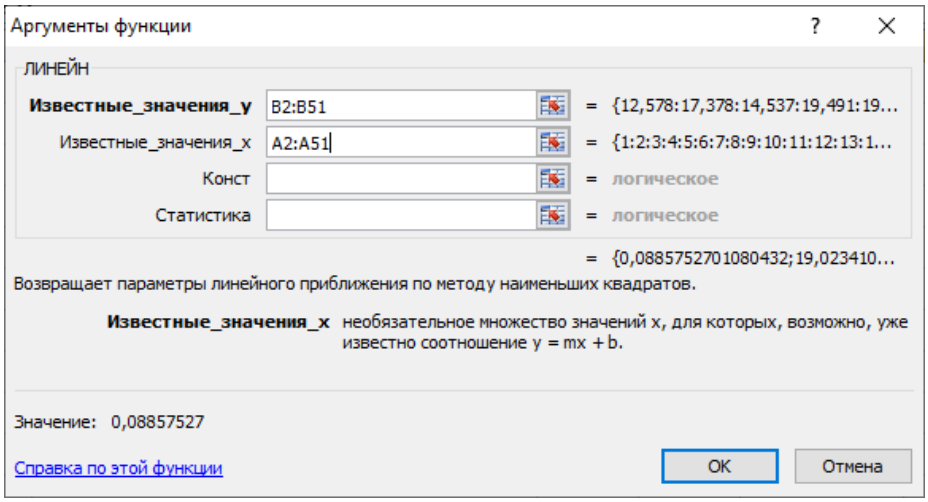


Рис. 8. Заполнение диалогового окна функции «ЛИНЕЙН»

В результате получим значения оценок параметров:

b	a
0,08857527	19,02341061

2) построение диаграммы и добавление линии тренда.

Выделяем столбец времени t и столбец с ВР X и выбираем вкладку «Вставка», нажимаем «Гистограмма» и среди различных типов гистограмм выбираем точечную. Полученный график представлен на рис. 9.

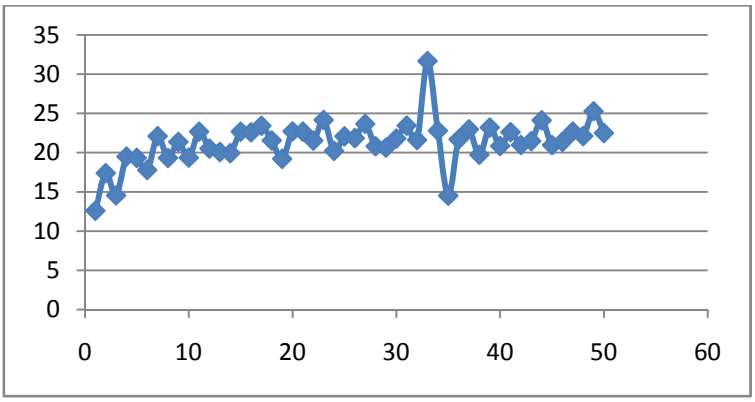


Рис. 9. График временного ряда

Нажимаем на любую точку на графике, вызываем контекстное меню правой кнопкой мыши и выбираем пункт «Добавить линию тренда».

В появившемся диалоговом окне выбираем вид линии тренда Линейная, проставляем галочки для вывода уравнения регрессии и коэффициента детерминации (рис. 10).

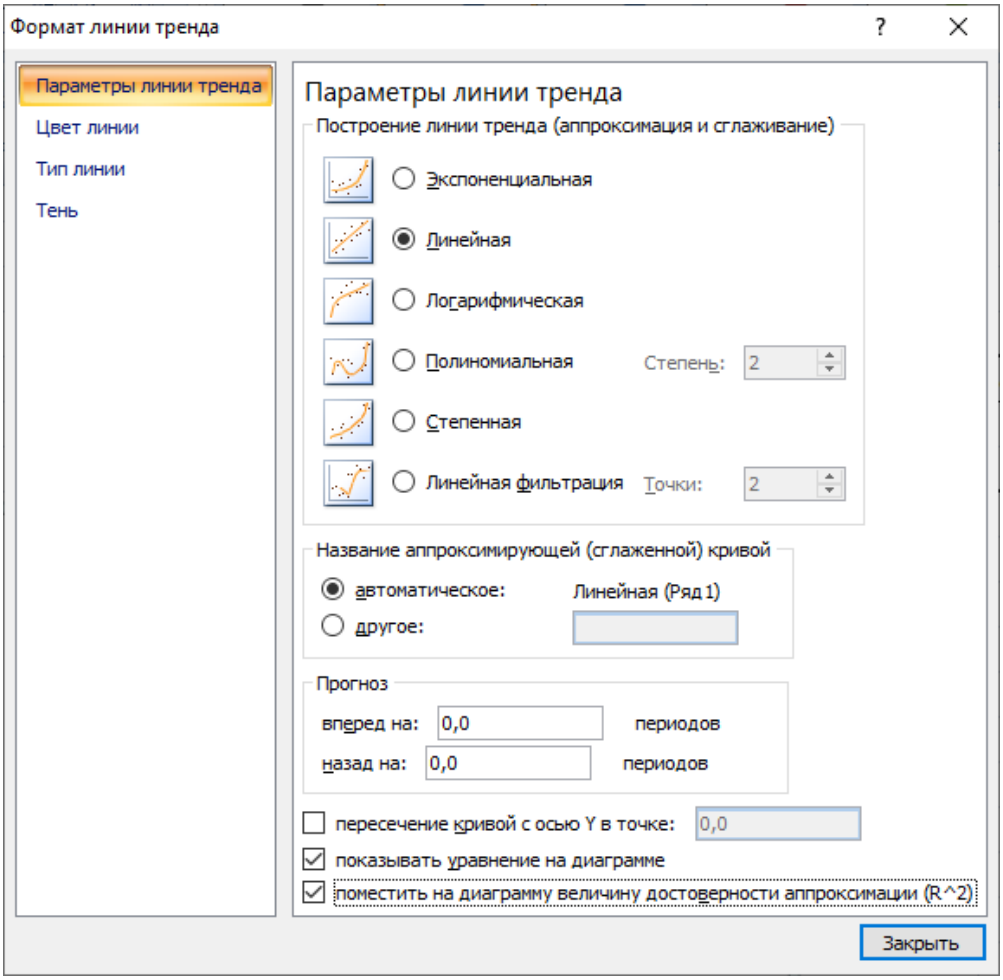


Рис. 10. Заполнение диалогового окна «Добавить линию тренда»

Результат построения линии тренда представлен на рис. 11.

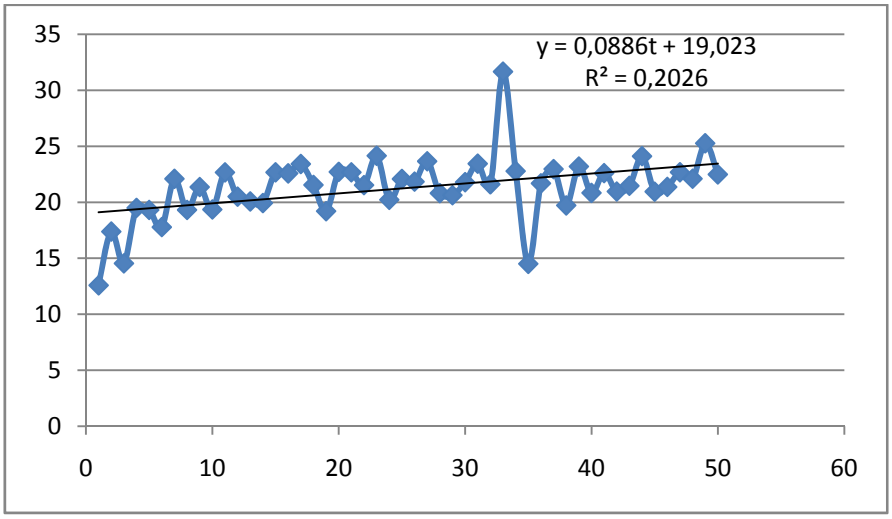


Рис. 11. Вывод линейной регрессии

Коэффициент детерминации R^2 показывает, что всего лишь 20 % разброса значений ВР обусловлены влиянием показателя времени. Это говорит о плохом качестве подбора модели.

3) Применение пакета анализа.

Используя надстройку «Пакет анализа», можно построить линейную регрессию. Пакет анализа дает возможность без дополнительных манипуляций сразу найти оценки параметров регрессии, построить таблицу дисперсионного анализа, рассчитать различные меры проверки значимости и адекватности модели и ее параметров. В диалоговом окне «Анализ данных» выбирается пункт «Регрессия» (рис. 12).

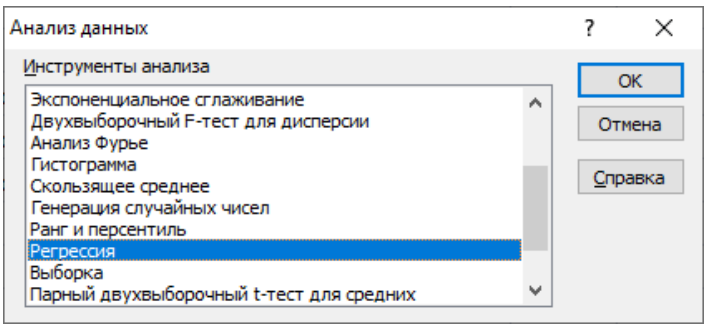


Рис. 12. Построение регрессии с помощью пакета анализа

Результат построения линейной регрессии представлен на рис. 13.

Регрессионная статистика	
Множественны	0,450118414
R-квадрат	0,202606586
Нормированны	0,185994224
Стандартная ош	2,588087268
Наблюдения	50

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ачимость F*
Регрессия	1	81,69208587	81,69209	12,19613	0,001039
Остаток	48	321,5133938	6,698196		
Итого	49	403,2054797			

	Коэффициенты	Стандартная ош	t-статисти	P-Значени	Нижние 9	Верхние 9	Нижние 9	Верхние 95,0%
Y-пересечение	19,02341061	0,743141575	25,59864	1,24E-29	17,52922	20,5176	17,52922	20,5176
Переменная X 1	0,08857527	0,025363046	3,492296	0,001039	0,037579	0,139571	0,037579	0,139571

Рис. 13. Результат построения линейной регрессии с помощью пакета анализа

Помимо коэффициента детерминации, здесь используются и другие меры качества модели, такие как:

- коэффициент множественной корреляции (0,45) говорит о слабой зависимости между исходными и предсказанными значениями ВР;

- остаточная дисперсия (6,7);

- критерий Фишера (F-критерий): показатель значимости F (меньше заданного уровня значимости, например, 0,05) говорит, что модель регрессии значима;

- критерий Стьюдена (t-критерий) используется для оценки статистической значимости параметров модели. При заданном уровне значимости α=0,05 параметры модели являются значимыми, так как *p*-значения меньше α.

Прогноз для ВР X на 5 значений вперед рассчитывается по уравнению линейного тренда: вместо *t* подставляем в уравнение следующие 5 значений времени (51,52,53,54,55) и получаем таблицу спрогнозированных значений ВР X.

51	23,54075
52	23,62932
53	23,7179
54	23,80648
55	23,89505

Диаграмма рассеяния для ВР X (построение точечной гистограммы без соединяющих точки линий):

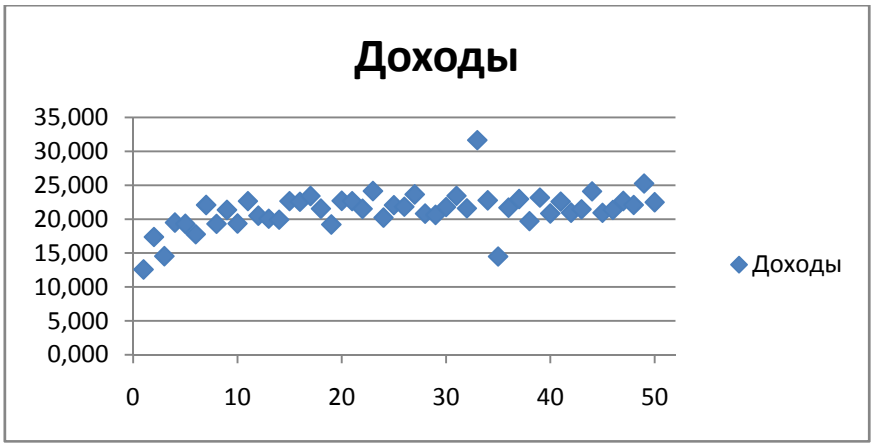
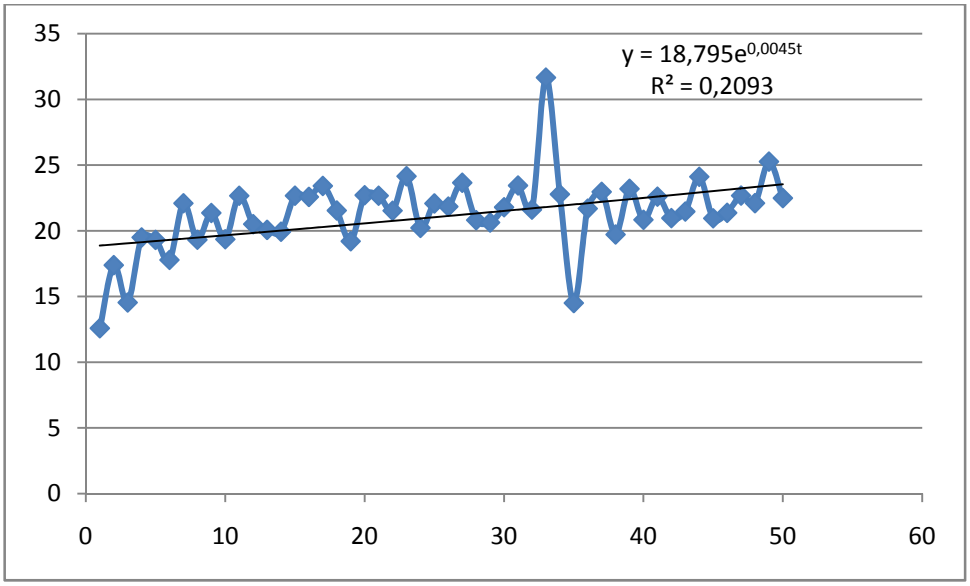


Рис. 14. Диаграмма рассеяния для ВР X

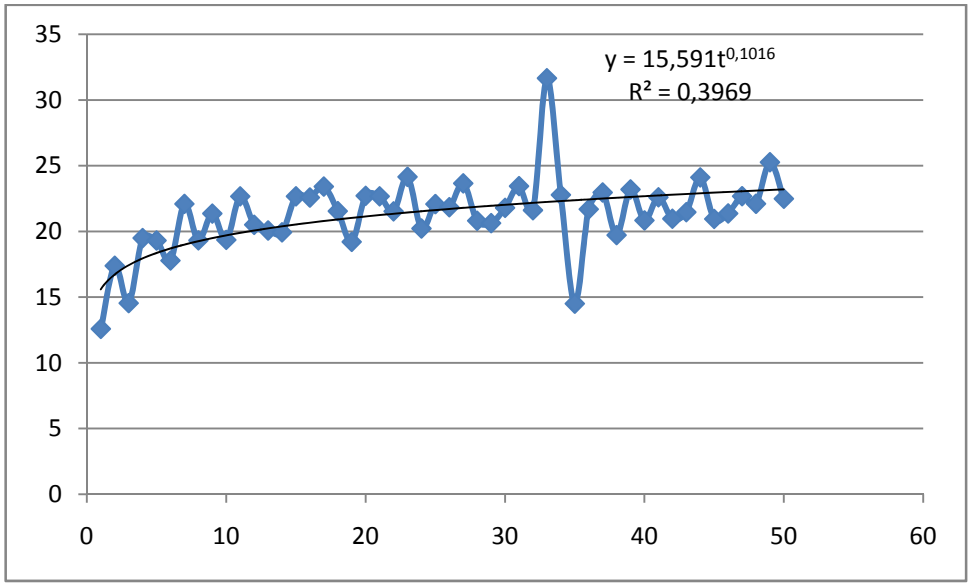
Ряд имеет незначительную возрастающую тенденцию и слабую положительную корреляцию.

3. Построим нелинейные тренды для ВР X с помощью построения точечных гистограмм.

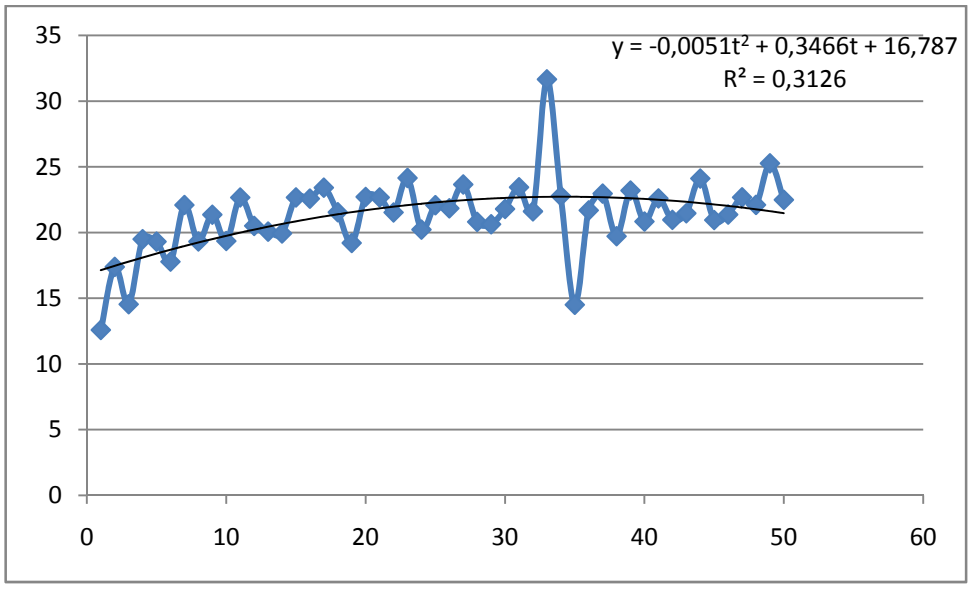
Экспоненциальная линия тренда:



Степенная линия тренда:



Квадратичная линия тренда:



По построенным графикам и значениям коэффициентов детерминации можно сделать вывод, что исходные данные лучше всего аппроксимируются степенной линией тренда.

Для построения нелинейных трендов также можно использовать Пакет анализа, но при необходимости исходные данные должны быть преобразованы.

4. Проведем сглаживание ВР X двумя способами:

1) Сглаживание методом скользящего среднего.

Вызываем диалоговое окно «Анализ данных», выбираем пункт «Скользящее среднее». В результате получаем таблицу скользящих средних при интервале сглаживания 3 и строим по этой таблице гистограмму (рис. 15).

Рис. 15. Сглаживание методом скользящего среднего

2) Экспоненциальное сглаживание.

Вызываем диалоговое окно «Анализ данных», выбираем пункт «Экспоненциальное сглаживание». В результате получаем таблицу сглаженных значений ВР и строим по этой таблице гистограмму (рис. 16).

Экспоненциальное сглаживание:

Рис. 16. Экспоненциальное сглаживание

Метод скользящего среднего сглаживает ряд лучше, чем метод экспоненциального сглаживания, и является наиболее достоверным, поскольку средняя относительная ошибка данного метода наименьшая и составляет не более 10 %.

5. Построим модели авторегрессии 1 и 2 порядка для ВР X. Для этого используем построение линейной регрессии с помощью Пакета анализа. В качестве зависимой переменной будем принимать текущие значения ВР, в качестве регрессоров – сдвинутые на 1 и 2 лага значения ВР.

В результате получаем следующие таблицы и графики моделей авторгерессии:

АР I-го порядка

ВЫВОД ИТОГОВ

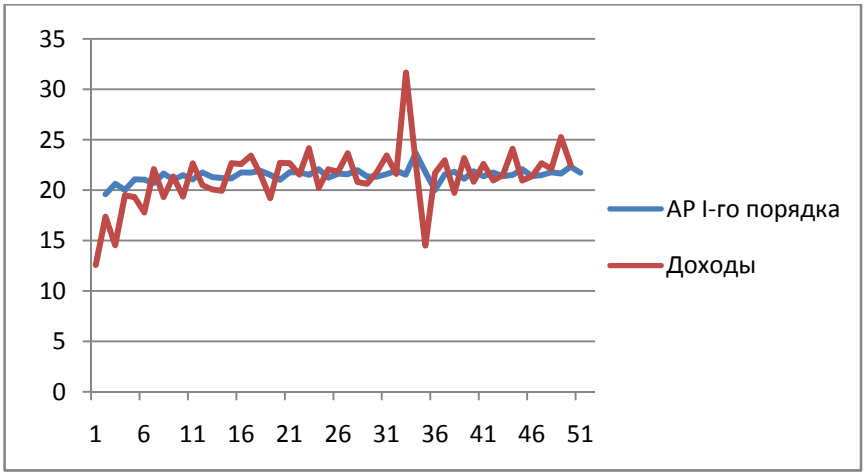
Регрессионная статистика	
Множественный R	0,238803841
R-квадрат	0,057027274
Нормированный R-квадрат	0,036964025
Стандартная ошибка	2,557064618
Наблюдения	49

Дисперсионный анализ

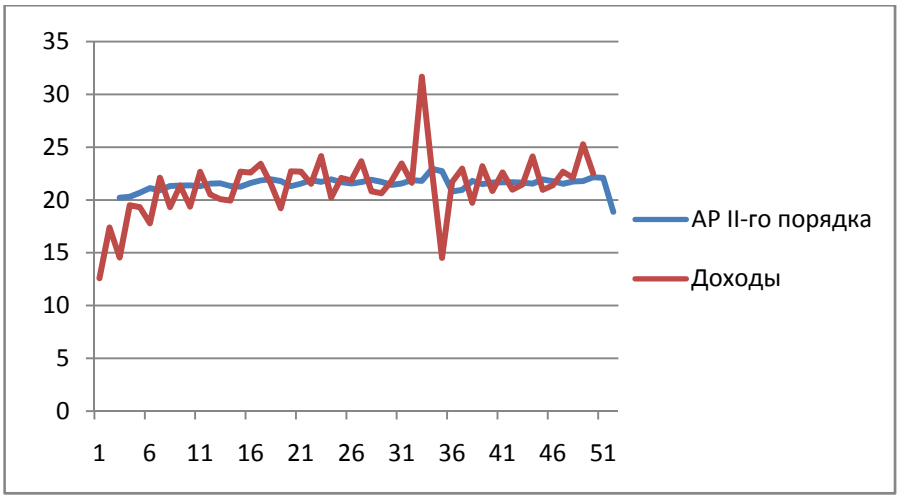
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	18,58509339	18,58509339	2,842374786	0,098437076
Остаток	47	307,3132346	6,53857946		
Итого	48	325,898328			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	16,88747836	2,736480585	6,171239968	1,48128E-07	11,3823895	22,39256722	11,3823895	22,39256722
Переменная X 1	0,21508782	0,127577811	1,685934395	0,098437076	-0,041565632	0,471741272	-0,041565632	0,471741272

$y = 16{,}887 + 0{,}215 * y_{t-1}$



АР II-го порядка									
ВЫВОД ИТОГОВ									
Регрессионная статистика									
Множественный R	0,189152566								
R-квадрат	0,035778693								
Нормированный R-квадрат	-0,007075587								
Стандартная ошибка	2,572673206								
Наблюдения	48								
Дисперсионный анализ									
	df	*SS*	*MS*	*F*	*Значимость F*				
Регрессия	2	11,05171079	5,525855397	0,834891941	0,440530813				
Остаток	45	297,8391342	6,618647427						
Итого	47	308,890845							
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*	
Y-пересечение	16,73487958	3,742383381	4,471717053	5,21628E-05	9,197332531	24,27242663	9,197332531	24,27242663	
Переменная X 1	0,130661616	0,146762181	0,890294865	0,37804282	-0,16493259	0,426255823	-0,16493259	0,426255823	
Переменная X 2	0,094865992	0,134744435	0,704043857	0,485032971	-0,176523231	0,366255214	-0,176523231	0,366255214	
y = 16,735+0,131*yt-1+0,095*yt-2									



По результатам анализа моделей АР I-ого и II-ого порядка получено, что модели не являются значимыми, т.к. *значимость* $F > 0.05$. Коэффициенты детерминации АР I-ого и II-ого порядка R_1^2=0.06 и R_2^2=0.04 также говорят о плохом качестве подбора моделей.

Аналогичные вычисления производятся для второго ВР.

Лабораторная работа № 2. Предварительный анализ временных рядов в системе Statistica

Задание к лабораторной работе № 2.

1. Построить графики временных рядов. Исходя из графиков, сделать выводы о возможном наличии тренда.
2. Провести экспоненциальное сглаживание рядов.
3. Построить автокорреляционную функцию.

Цель работы: в ходе выполнения лабораторной работы студент приобретает навыки предварительного анализа временных рядов в системе Statistica.

Пример выполнения.

В качестве исходных данных возьмем ВР X.

Универсальный пакет Statistica занимает одно из первых мест в мире среди программ статистической обработки данных. В пакете Statistica анализ и моделирование ВР осуществляется в модуле Временные ряды и прогнозирование (Time Series/Forecasting).

Последовательность выполнения работы.

1. Загрузите пакет Statistica. Установите удобный Вам режим сохранения результатов работы с помощью Диспетчера вывода Файл - Диспетчера вывода (File - Output Manager).

2. Создайте новую Рабочую книгу с таблицей для исходных данных Файл - Создать - Создать новый документ (File - New - Create New Document) и скопируйте в нее временной ряд. Созданную рабочую книгу лучше сразу переименовать и сохранить в своей папке. Если данные уже сохранены в таблице Excel, можно открыть уже готовую таблицу.

3. Каждому ряду дайте уникальное имя вместо стандартных Var1, Var2…, а в поле Длинная метка или формула (Long Name) вы можете занести любую полезную информацию о ряде. Переименование осуществляется в окне спецификаций переменной, которое можно вызвать двойным щелчком правой кнопки мыши на имени столбца.

Рекомендуется использовать вышеперечисленную последовательность действий как стандартное начало любого сеанса работы в пакете. Вызовите модуль для работы с временными рядами Анализ - Углубленные методы анализа - Временные ряды и

прогнозирование (Statistics - Advanced Linear/Nonlinear Models - Time Series/Foregasting) (рис.17).

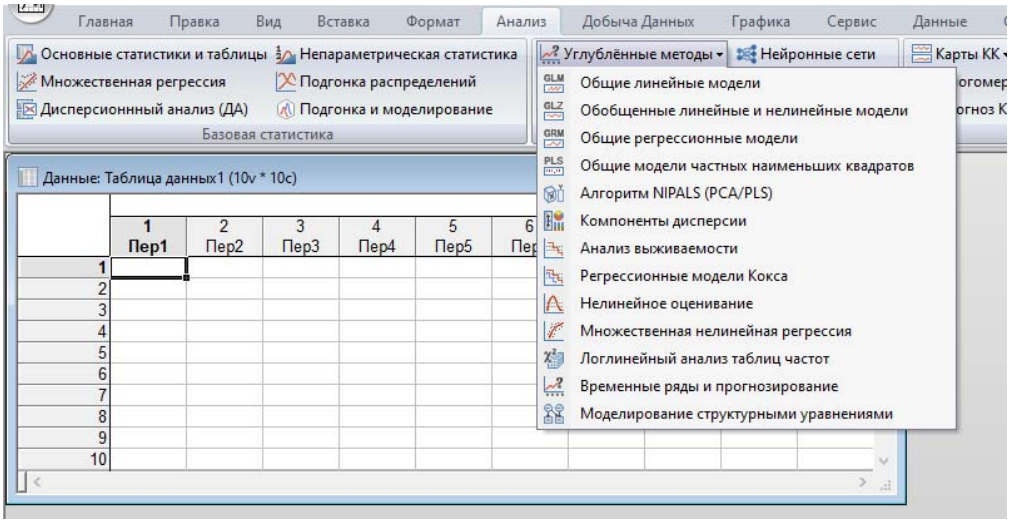


Рис. 17. Выбор модуля Временные ряды и прогнозирование в пакете Statistica

Стартовое окно модуля Временные ряды и прогнозирование представлено на рис. 18.

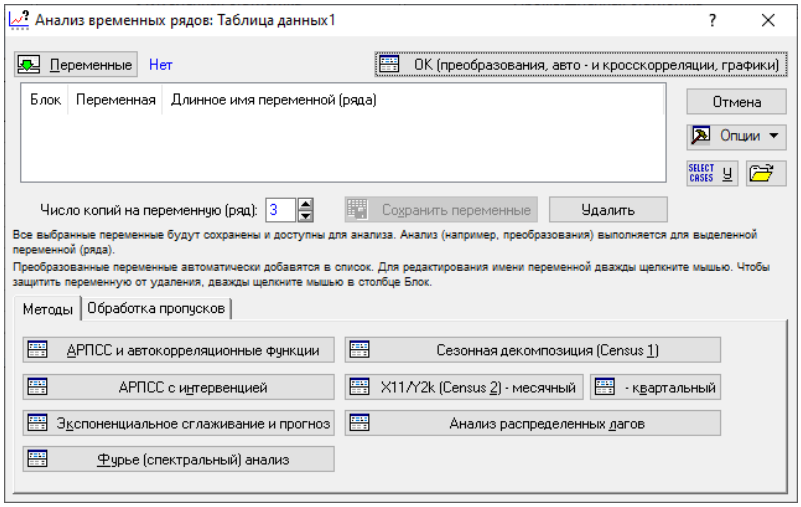


Рис. 18. Стартовое окно модуля Временные ряды и прогнозирование

1. Построим график временного ряда. Для этого выберем в пункте меню «Графика» линейные графики.

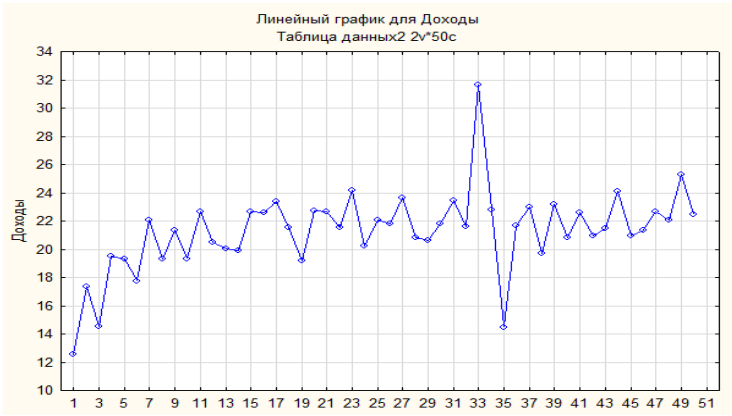


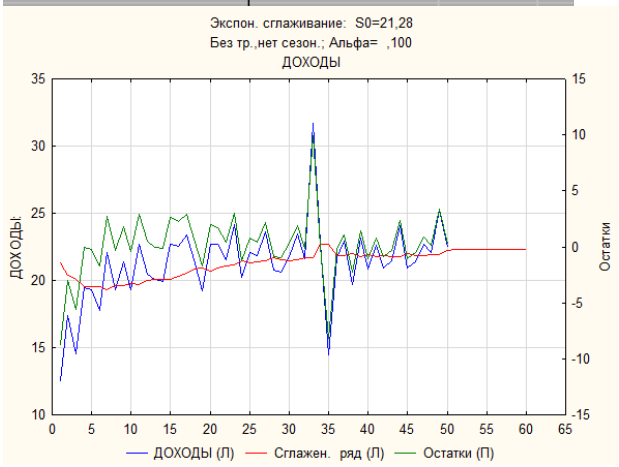
Рис. 19. Результат построения графика ВР в пакете Statistica

По виду графика можно сделать вывод, что рассматриваемый ВР содержит незначительно возрастающий тренд и периодические колебания.

2. Проведем экспоненциальное сглаживание ряда. В модуле Временные ряды полностью реализованы все 12 классических моделей экспоненциального сглаживания. Задание модели может включать аддитивную или мультипликативную сезонную составляющую и/или линейный, экспоненциальный или демпфированый тренд; в частности, доступны популярные модели с линейным трендом Холта-Винтера. Пользователь может задавать начальное значение параметров сглаживания, начальное значение тренда и (если требуется) сезонные факторы. Для тренда и сезонной составляющей могут быть заданы независимые параметры сглаживания. Для определения лучшей комбинации параметров используется метод поиска на сетке; в таблицах результатов для всех комбинаций значений параметров сглаживания вычисляется средняя ошибка, средняя абсолютная ошибка, сумма квадратов ошибок, среднеквадратическая ошибка, средняя относительная ошибка и средняя абсолютная относительная ошибка. Наименьшие значения этих ошибок выделяются цветом. Имеется возможность автоматического поиска лучшего набора параметров в смысле среднеквадратической, средней абсолютной или средней абсолютной относительной ошибки (для этого используется общая процедура минимизации). На стартовом окне модуля выбираем кнопку «Экспоненциальное сглаживание и прогноз». Построим для примера 3 варианта экспоненциального сглаживания:

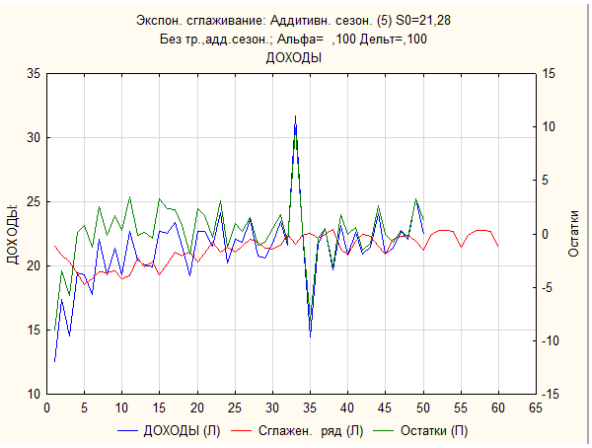
1) Нет сезонности

Итоговая ошибка	Экспон. сглаживание: S0=21,28 (Без тр.,нет сезон.; Альфа= ,100 ДОХОДЫ
	Ошибка
Средн. ош.	0,200514297665
Ср. абсол. ошибка	1,781375135555
Суммы квадратов	387,774407119757
Ср. квадрат	7,755488142395
Средн. относ. ошибка	-0,971106547193
Ср. абс. отн. ошибка	9,107986237426



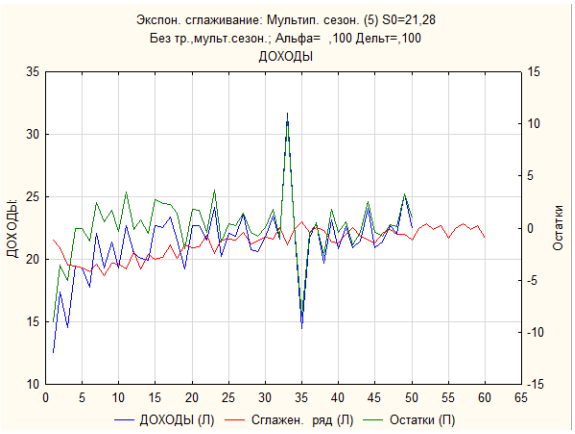
2) С аддитивной сезонностью

Итоговая ошибка	Экспон. сглаживание: Аддитивн. с Без тр.,адд.сезон.; Альфа= ,100 Д ДОХОДЫ
	Ошибка
Средн. ош.	0,195619553464
Ср. абсол. ошибка	1,843158085873
Суммы квадратов	405,777311788765
Ср. квадрат	8,115546235775
Средн. относ. ошибка	-0,999770802258
Ср. абс. отн. ошибка	9,422187720255



3) С мультипликативной сезонностью

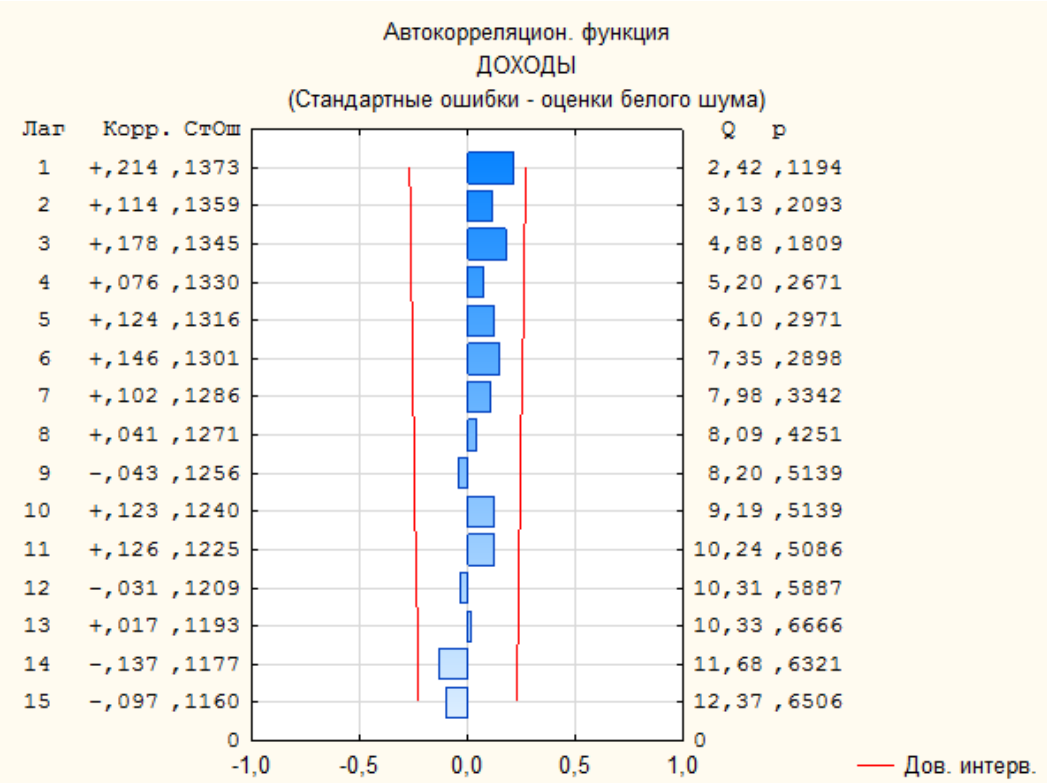
Итоговая ошибка	Экспон. сглаживание: Мультип. сез Без тр.,мульт.сезон.; Альфа= ,100 / ДОХОДЫ
	Ошибка
Средн. ош.	0,182813490585
Ср. абсол. ошибка	1,811869076701
Суммы квадратов	409,952810371304
Ср. квадрат	8,199056207426
Средн. относ. ошибка	-1,060471723027
Ср. абс. отн. ошибка	9,254652096856



Все построенные модели сглаживания не содержат тренда. Лучшей считается модель с наименьшими абсолютными значениями ошибки: среди трех построенных вариантов сглаживания лучшей является модель, не содержащая сезонности.

3. Построим автокорреляционную функцию.

На стартовом окне модуля выбираем кнопку «АРСС и автокорреляционные функции».



Автокорреляция 1-го порядка имеет наибольшее значение. Все коэффициенты автокорреляции незначимы, следовательно, предыдущие значения ВР оказывают слабое влияние на текущие значения.

Лабораторная работа № 3. Модели авторегрессии, скользящего среднего и ARIMA

Задание к лабораторной работе № 3.

- построить модели авторегрессии 1-го и 2-го порядка с константой и без нее;

- построить модели скользящего среднего 1-го и 2-го порядка с константой и без нее;

- построить прогнозы для всех моделей;

- вычислить приближенные и точные суммы квадратов из условия максимума правдоподобия;

- вычислить оценки параметров, стандартные ошибки и корреляции;

- исследовать остатки модели ARIMA.

Цель работы: в ходе выполнения лабораторной работы студент приобретает навыки построения ARIMA моделей в системе Statistica.

Пример выполнения.

В качестве исходных данных возьмем ВР X. На стартовом окне модуля выбираем кнопку «АРСС и автокорреляционные функции». В появившемся диалоговом окне задайте параметры модели. Рассмотрим в качестве примера только модели первого и второго порядка с константой. Модели авторегрессии 1-го и 2-го порядка для доходов:

Исход.:ДОХОДЫ (Таблица данных2)
Преобразования: Нет
Модель(1,0,0) MS Остаток= 8,0130

Параметр	Парам.	Асимпт. Ст.ошиб.	Асимпт. t(48)	p	Нижняя 95% дов.	Верхняя 95% дов.
Конст.	21,22825	0,515764	41,15882	0,000000	20,19124	22,26526
p(1)	0,21589	0,143558	1,50387	0,139167	-0,07275	0,50453

```
      Переменная: ДОХОДЫ
Преобразования:
          Модель: (1,0,0)
Число набл.:50  Начальная SS=23050,  Итоговая SS=384,62 (1,669%)  MS=8,0130
 Параметры (p/Ps-авторегрессии, q/Qs-скольз. средн.); выделение: p<.05
             Конст.  p(1)
   Оценка: 21,228 ,21589
 Ст.ошиб.: ,51576 ,14356
```

Т.е. модель авторегрессии первого порядка имеет вид: $Y=21,228+0,22y_{t-1}$.

Исход.:ДОХОДЫ (Таблица данных2)
Преобразования: Нет
Модель(2,0,0) MS Остаток= 8,1351

Параметр	Парам.	Асимпт. Ст.ошиб.	Асимпт. t(47)	p	Нижняя 95% дов.	Верхняя 95% дов.
Конст.	21,20166	0,567579	37,35458	0,000000	20,05984	22,34348
p(1)	0,20068	0,147864	1,35722	0,181193	-0,09678	0,49815
p(2)	0,07890	0,151020	0,52247	0,603795	-0,22491	0,38272

```
      Переменная: ДОХОДЫ
Преобразования:
         Модель: (2,0,0)
Число набл.:50  Начальная SS=23050,  Итоговая SS=382,35 (1,659%)  MS=8,1351
 Параметры (p/Ps-авторегрессии, q/Qs-скольз. средн.); выделение: p<.05
            Конст.   p(1)    p(2)
    Оценка: 21,202 ,20068 ,07890
  Ст.ошиб.: ,56758 ,14786 ,15102
```

Модель авторегрессии второго порядка имеет вид $Y=21,202+0,2y_{t-1}+0,08y_{t-2}$

Коэффициенты во всех моделях устойчивы, т.к. они, по меньшей мере, в два раза превышают свои стандартные ошибки.

Модели скользящего среднего 1-го и 2-го порядка для доходов:

Исход.:ДОХОДЫ (Таблица данных2)
Преобразования: Нет
Модель(0,0,1) MS Остаток= 8,0570

Параметр	Парам.	Асимпт. Ст.ошиб.	Асимпт. t(48)	p	Нижняя 95% дов.	Верхняя 95% дов.
Конст.	21,24402	0,484913	43,80994	0,000000	20,26904	22,21901
q(1)	-0,19665	0,140366	-1,40098	0,167651	-0,47887	0,08557

```
      Переменная: ДОХОДЫ
Преобразования:
         Модель: (0,0,1)
Число набл.:50  Начальная SS=23050,  Итоговая SS=386,73 (1,678%)  MS=8,0570
 Параметры (p/Ps-авторегрессии, q/Qs-скольз. средн.); выделение: p<.05
            Конст.   q(1)
    Оценка: 21,244 -,1967
  Ст.ошиб.: ,48491 ,14037
```

Модель скользящего среднего первого порядка имеет вид:
$Y=\varepsilon_t+0,20\varepsilon_{t-1}+21,244$

Исход.:ДОХОДЫ (Таблица данных2)
Преобразования: Нет
Модель(0,0,2) MS Остаток= 8,2044

Параметр	Парам.	Асимпт. Ст.ошиб.	Асимпт. t(47)	p	Нижняя 95% дов.	Верхняя 95% дов.
Конст.	21,23752	0,508620	41,75523	0,000000	20,21431	22,26073
q(1)	-0,19014	0,151020	-1,25904	0,214232	-0,49395	0,11367
q(2)	-0,05490	0,150233	-0,36543	0,716431	-0,35713	0,24733

```
      Переменная: ДОХОДЫ
Преобразования:
         Модель: (0,0,2)
Число набл.:50  Начальная SS=23050,  Итоговая SS=385,61 (1,673%)  MS=8,2044
 Параметры (p/Ps-авторегрессии, q/Qs-скольз. средн.); выделение: p<.05
            Конст.   q(1)    q(2)
    Оценка: 21,238 -,1901 -,0549
  Ст.ошиб.: ,50862 ,15102 ,15023
```

Модель скользящего среднего второго порядка имеет вид:

$Y=\varepsilon_t+0,19\varepsilon_{t-1}+0,05\varepsilon_{t-2}+21,238$

Коэффициенты практически во всех моделях устойчивы, т.к. они, по меньшей мере, в два раза превышают свои стандартные ошибки (по модулю).

Прогнозы для модели авторегрессии (1-го и 2-го порядка) для доходов:

Прогнозы; Модель:(1,0,0) Сезонный лаг: 1
Исход.:ДОХОДЫ
Начало исходных: 1 Конец исходн.: 50

Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
51	21,49957	16,75181	26,24733	2,830725
52	21,28683	16,42968	26,14397	2,895942
53	21,24090	16,37871	26,10308	2,898946
54	21,23098	16,36856	26,09340	2,899086
55	21,22884	16,36641	26,09127	2,899093
56	21,22838	16,36595	26,09081	2,899093
57	21,22828	16,36585	26,09071	2,899093
58	21,22826	16,36583	26,09069	2,899093
59	21,22825	16,36582	26,09068	2,899093
60	21,22825	16,36582	26,09068	2,899093

Прогнозы; Модель:(2,0,0) Сезонный лаг: 1
Исход.:ДОХОДЫ
Начало исходных: 1 Конец исходн.: 50

Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
51	21,77966	16,99386	26,56546	2,852212
52	21,41892	16,53769	26,30014	2,909080
53	21,29087	16,37643	26,20530	2,928872
54	21,23670	16,31859	26,15482	2,931066
55	21,21573	16,29691	26,13455	2,931485
56	21,20725	16,28833	26,12617	2,931546
57	21,20389	16,28496	26,12283	2,931556
58	21,20255	16,28361	26,12149	2,931558
59	21,20201	16,28308	26,12095	2,931558
60	21,20180	16,28286	26,12074	2,931558

Прогнозы для модели скользящего среднего (1-го и 2-го порядка) для доходов:

С константой:

Прогнозы; Модель:(0,0,1) Сезонный лаг: 1
Исход.:ДОХОДЫ
Начало исходных: 1 Конец исходн.: 50

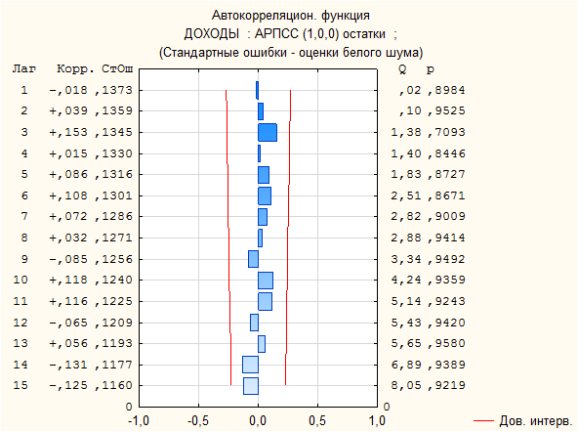
Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
51	21,33710	16,57633	26,09786	2,838479
52	21,24402	16,39208	26,09597	2,892842
53	21,24402	16,39208	26,09597	2,892842
54	21,24402	16,39208	26,09597	2,892842
55	21,24402	16,39208	26,09597	2,892842
56	21,24402	16,39208	26,09597	2,892842
57	21,24402	16,39208	26,09597	2,892842
58	21,24402	16,39208	26,09597	2,892842
59	21,24402	16,39208	26,09597	2,892842
60	21,24402	16,39208	26,09597	2,892842

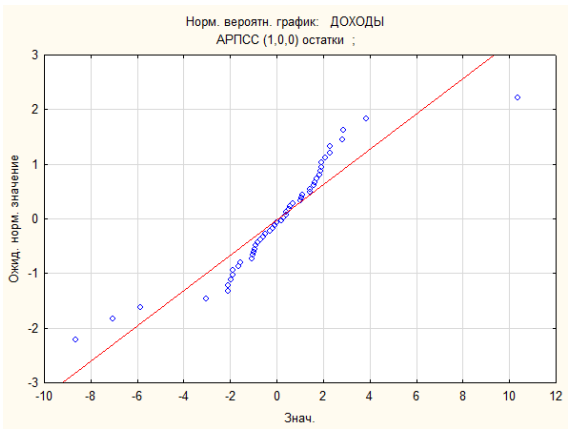
Прогнозы; Модель:(0,0,2) Сезонный лаг: 1
Исход.:ДОХОДЫ
Начало исходных: 1 Конец исходн.: 50

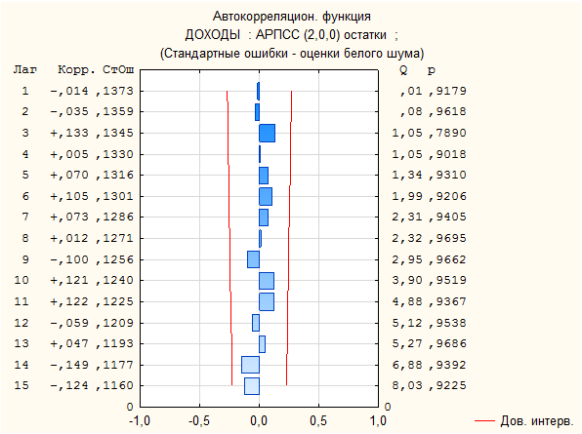
Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
51	21,54059	16,73445	26,34673	2,864332
52	21,26423	16,37198	26,15648	2,915651
53	21,23752	16,33817	26,13688	2,919888
54	21,23752	16,33817	26,13688	2,919888
55	21,23752	16,33817	26,13688	2,919888
56	21,23752	16,33817	26,13688	2,919888
57	21,23752	16,33817	26,13688	2,919888
58	21,23752	16,33817	26,13688	2,919888
59	21,23752	16,33817	26,13688	2,919888
60	21,23752	16,33817	26,13688	2,919888

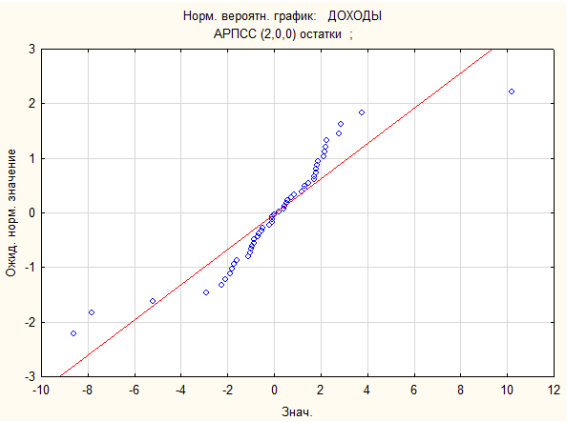
Проведя дальнейшие вычисления и сравнив модели, получаем, что лучшая модель АР – это модель АР 2-го порядка с константой с точным методом оценивания, а лучшая модель СС – это модель СС 2-го порядка с константой с приближенным методом оценивания.

Корреляция и остатки модели для доходов:
Модели авторегрессии:

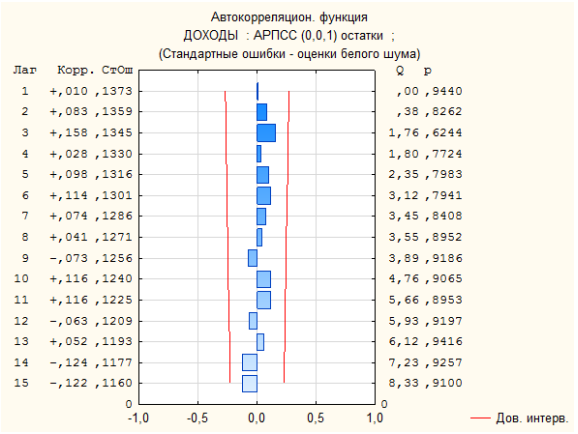


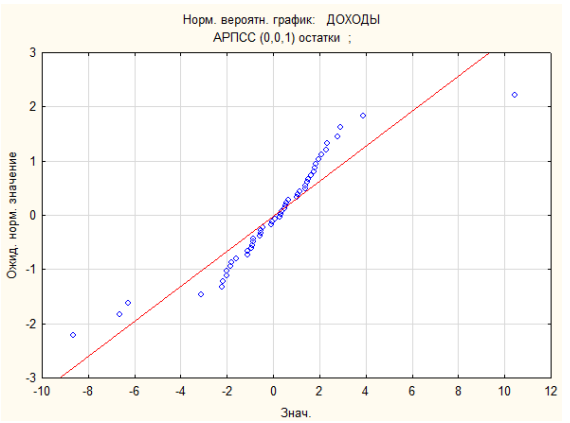


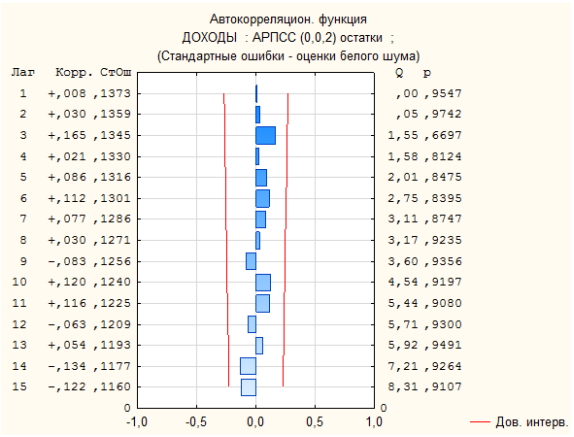


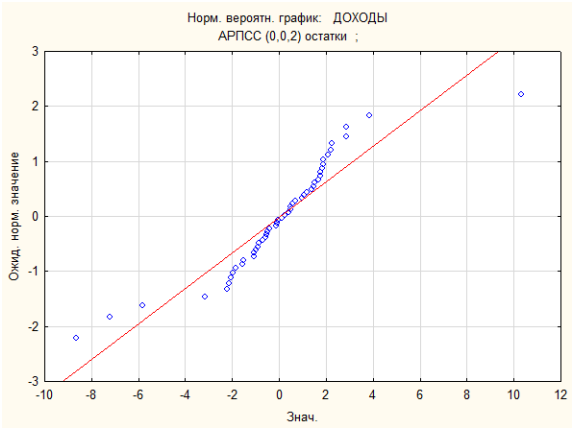


Модели скользящего среднего:









По данным рисункам можно сделать вывод, что остатки после построения моделей не имеют значимой автокорреляции и подчиняются распределению, близкому к нормальному.

Лабораторная работа № 4. Спектральный (Фурье) анализ и кросс-спектральный анализ.

Задание к лабораторной работе № 4.

- провести спектральный анализ ВР;
- провести кросс-спектральный анализ ВР.

Цель работы: в ходе выполнения лабораторной работы студент приобретает навыки проведения спектрального анализа временных рядов в пакете Statistica.

Пример выполнения.

Для проведения спектрального анализа в качестве исходных данных возьмем ВР X, для проведения кросс-спектрального анализа нам потребуются оба ряда X и Y. На стартовом окне модуля выбираем кнопку «Фурье (спектральный) анализ».

Модуль Временные ряды включает полную реализацию методов спектрального, или Фурье, анализа одного ряда и кросс-спектральный анализ двух рядов. Стандартные методы предварительной обработки ряда включают косинус-сглаживание, вычитание среднего и удаление тренда. Результаты обычного спектрального анализа содержат коэффициенты частоты, периода, коэффициенты при синусах и

косинусах, периодограммы и оценки спектральной плотности. Оценка плотности может быть вычислена с помощью весов Даниеля, Хэмминга, Бартлетта, Тьюки, Парзена или с весами и шириной, заданными пользователем. Очень полезно, особенно при работе с длинными рядами, иметь возможность выводить в убывающем порядке заранее заданное число точек периодограммы или спектральной плотности; таким образом можно легко обнаружить резкие пики периодограммы и спектральной плотности для длинных рядов. Имеется возможность вычислить d-критерий Колмогорова-Смирнова для значений периодограммы, чтобы проверить, является ряд белым шумом или нет. Для представления результатов анализа имеются различные типы графиков; можно отобразить коэффициенты при синусах и косинусах, периодограмму, лог-периодограмму, спектральную и лог-спектральную плотность по отношению к частотам, периодам и лог-периодам. В случае длинного исходного ряда имеется возможность выбрать конкретный сегмент (период), для которого будут изображаться соответствующие периодограмма и график спектральной плотности, тем самым будет улучшено их «разрешение». При кросс-спектральном анализе, в дополнение к результатам обычного спектрального анализа каждого отдельного ряда, вычисляется кросс-периодограмма (вещественная и мнимая часть), ко-спектральная плотность, квадратурный спектр, кросс-амплитуда, значения когерентности, усиления и фазовый спектр. Все эти величины могут быть выведены на график, где по горизонтальной оси будет откладываться частота, период или лог-период либо для всего интервала периодов (соответственно, частот), либо для выбранного пользователем диапазона. Указанное пользователем количество наибольших значений кросс-периодограммы (вещественных или мнимых) может быть выведено в убывающем порядке в виде таблицы результатов, что позволяет легко выявлять на ней резкие пики для длинных исходных рядов. Как и во всех других

процедурах модуля Временные ряды, все полученные ряды могут быть добавлены в активную рабочую область и затем подвергнуты дальнейшему исследованию с помощью других методов анализа временных рядов или средствами других модулей системы STATISTICA.

Спектральный (Фурье) анализ временного ряда X.

```
Перемен.: ДОХОДЫ

  Число наблюдений:  50      Первое: 1  Последнее: 50
N набл. после доб. конст: 50
          Преобразования: Сред=21,282    вычтено; Уд. тренд

Пять наиб. пиков периодогр. (значение:частота):
  (1)48,86:,0200 (2)45,30:,3800 (3)44,66:,1800 (4)24,24:,1200 (5)23,24:,3200
```

Из данных видно пять наибольших пиков периодограммы (по частоте). Эта информация часто полезна при анализе очень больших рядов, которые непросто показать на одном графике. Гармоники с наибольшими значениями периодограммы также будут полезны при построении гармонической модели.

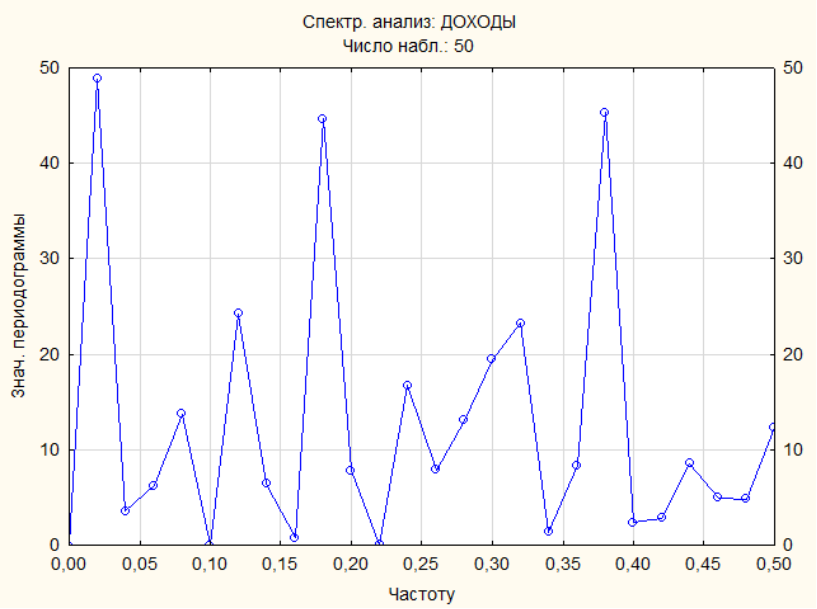
Результатом проведения спектрального анализа будет таблица вида (на примере применения весов Хемминга):

	Частоту	Период	Косинус коэфф.	Синус коэфф.	Периодограмма	Плотн.	Хемминг веса
0	0,000000		0,00000	0,00000	0,00000	23,81182	0,035714
1	0,020000	50,00000	-1,18159	0,74702	48,85503	24,64439	0,241071
2	0,040000	25,00000	-0,37192	0,07371	3,59395	15,37651	0,446429
3	0,060000	16,66667	-0,21422	-0,45137	6,24073	8,70533	0,241071
4	0,080000	12,50000	-0,72401	-0,15708	13,72152	8,62541	0,035714
5	0,100000	10,00000	0,01255	-0,00807	0,00557	9,60423	
6	0,120000	8,33333	-0,15560	-0,97223	24,23601	12,88023	
7	0,140000	7,14286	-0,47490	0,17397	6,39493	10,47855	
8	0,160000	6,25000	0,17293	0,03088	0,77148	13,79612	
9	0,180000	5,55556	-0,56328	-1,21201	44,65663	22,23480	
10	0,200000	5,00000	-0,27398	0,48733	7,81388	14,88355	
11	0,220000	4,54545	-0,00440	0,02827	0,02047	7,80107	
12	0,240000	4,16667	-0,63659	-0,51354	16,72408	10,11843	
13	0,260000	3,84615	-0,42734	0,36445	7,88613	11,40172	
14	0,280000	3,57143	0,71806	0,08754	13,08172	13,86003	
15	0,300000	3,33333	-0,27196	-0,83935	19,46190	17,77318	
16	0,320000	3,12500	-0,33483	0,90406	23,23607	16,15521	
17	0,340000	2,94118	0,23175	-0,00728	1,34404	10,53291	
18	0,360000	2,77778	-0,48569	-0,31470	8,37324	15,89711	
19	0,380000	2,63158	-0,53588	1,23478	45,29625	22,96792	
20	0,400000	2,50000	0,08745	0,29694	2,39554	13,28321	
21	0,420000	2,38095	-0,18956	0,28049	2,86518	5,70671	
22	0,440000	2,27273	-0,56829	0,13399	8,52275	5,95464	
23	0,460000	2,17391	-0,25462	-0,36648	4,97851	5,99150	
24	0,480000	2,08333	-0,43971	-0,02925	4,85498	6,82635	
25	0,500000	2,00000	0,70330	0,00000	12,36561	8,21676	

	Частоту	Период	Косинус коэфф.	Синус коэфф.	Периодограмма	Плотн.
1	0,020000	50,00000	-1,18159	0,74702	48,85503	24,64439
0	0,000000		0,00000	0,00000	0,00000	23,81182
19	0,380000	2,63158	-0,53588	1,23478	45,29625	22,96792
9	0,180000	5,55556	-0,56328	-1,21201	44,65663	22,23480
15	0,300000	3,33333	-0,27196	-0,83935	19,46190	17,77318
16	0,320000	3,12500	-0,33483	0,90406	23,23607	16,15521
18	0,360000	2,77778	-0,48569	-0,31470	8,37324	15,89711
2	0,040000	25,00000	-0,37192	0,07371	3,59395	15,37651
10	0,200000	5,00000	-0,27398	0,48733	7,81388	14,88355
14	0,280000	3,57143	0,71806	0,08754	13,08172	13,86003
8	0,160000	6,25000	0,17293	0,03088	0,77148	13,79612
20	0,400000	2,50000	0,08745	0,29694	2,39554	13,28321
6	0,120000	8,33333	-0,15560	-0,97223	24,23601	12,88023
13	0,260000	3,84615	-0,42734	0,36445	7,88613	11,40172
17	0,340000	2,94118	0,23175	-0,00728	1,34404	10,53291
7	0,140000	7,14286	-0,47490	0,17397	6,39493	10,47855
12	0,240000	4,16667	-0,63659	-0,51354	16,72408	10,11843
5	0,100000	10,00000	0,01255	-0,00807	0,00557	9,60423
3	0,060000	16,66667	-0,21422	-0,45137	6,24073	8,70533
4	0,080000	12,50000	-0,72401	-0,15708	13,72152	8,62541
25	0,500000	2,00000	0,70330	0,00000	12,36561	8,21676
11	0,220000	4,54545	-0,00440	0,02827	0,02047	7,80107
24	0,480000	2,08333	-0,43971	-0,02925	4,85498	6,82635
23	0,460000	2,17391	-0,25462	-0,36648	4,97851	5,99150
22	0,440000	2,27273	-0,56829	0,13399	8,52275	5,95464
21	0,420000	2,38095	-0,18956	0,28049	2,86518	5,70671

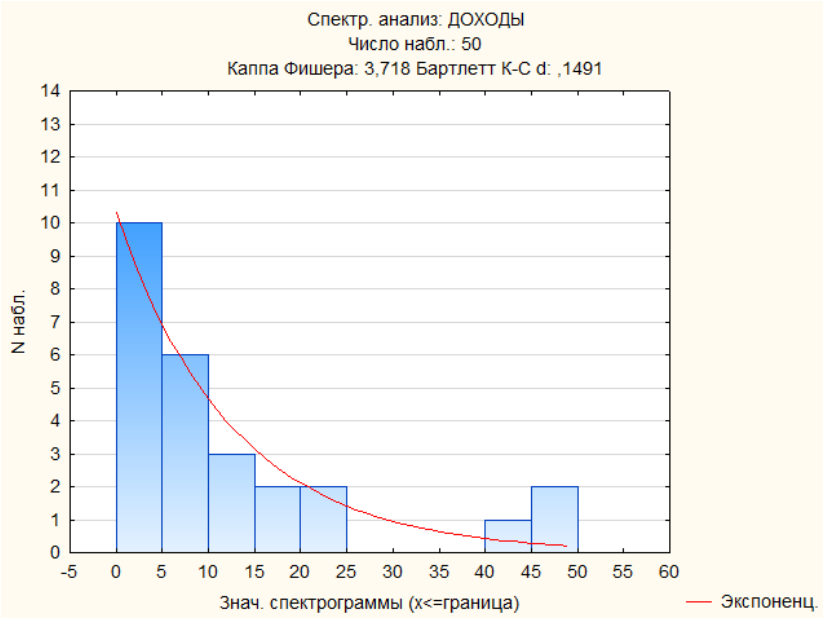
и графики, например:

1) Периодограмма:



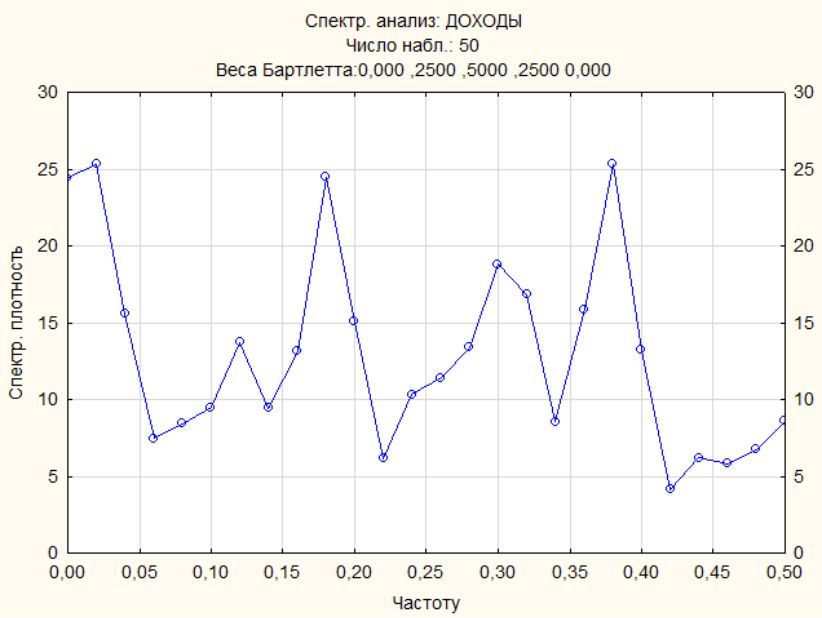
На графике периодограммы видно 8 чётких пиков, максимальный на частоте 0,04.

2) Гистограмма периодограммы:



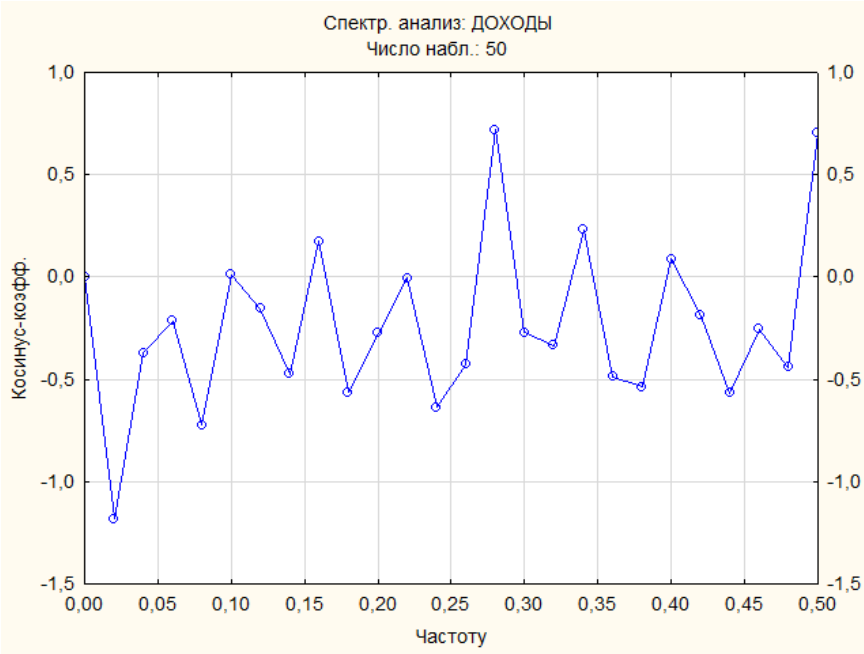
Из графика видно, что вычисленный d-критерий Колмогорова-Смирнова для значений периодограммы не подчиняется экспоненциальному распределению.

3) Спектральная плоскость:



На графике спектральной плотности видно 5 четких пиков, максимальный на частоте 0,38.

4) Косинус-коэффициент:



На графике косинус-коэфф. видно 8 четких пиков, максимальный на частоте 0,28.

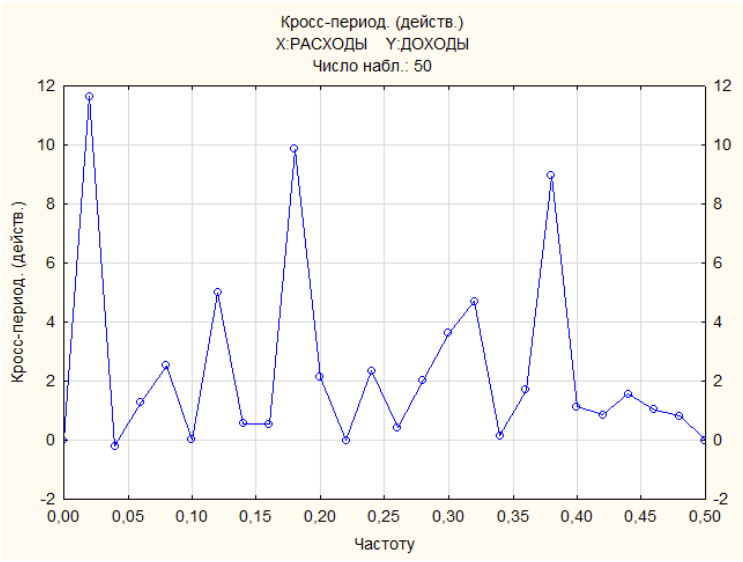
Все остальные расчеты и графики студент рассматривает самостоятельно.

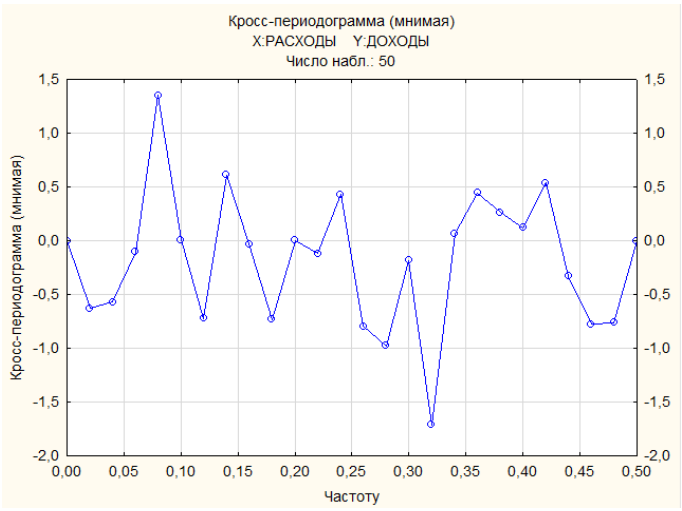
Кросс-спектральный анализ

Аналогично проводим процедуру кросс-спектрального анализа для 2 рядов. В результате получим графики составляющих кросс-спектрального анализа и таблицу. Интерпретация составляющих кросс-

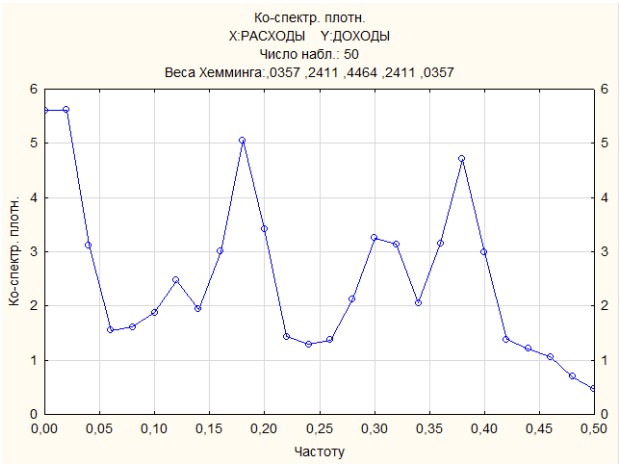
спектрального анализа представлена в разделе 4 данного учебного пособия.

Кросс-период (действительных и мнимых значений):

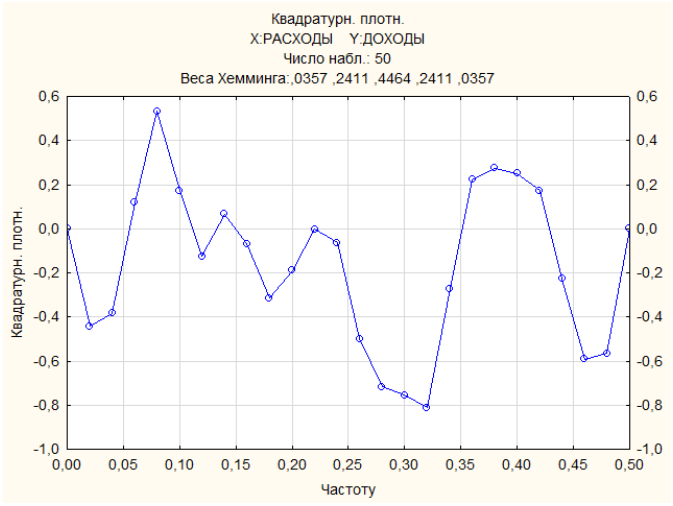




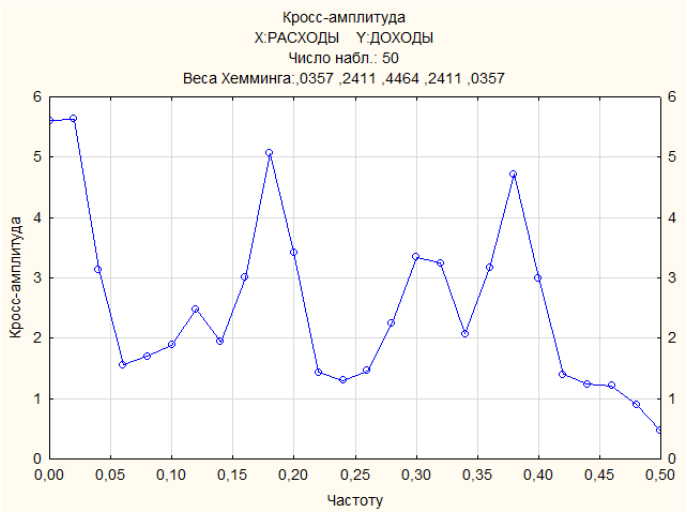
Ко-спектральная плотность:



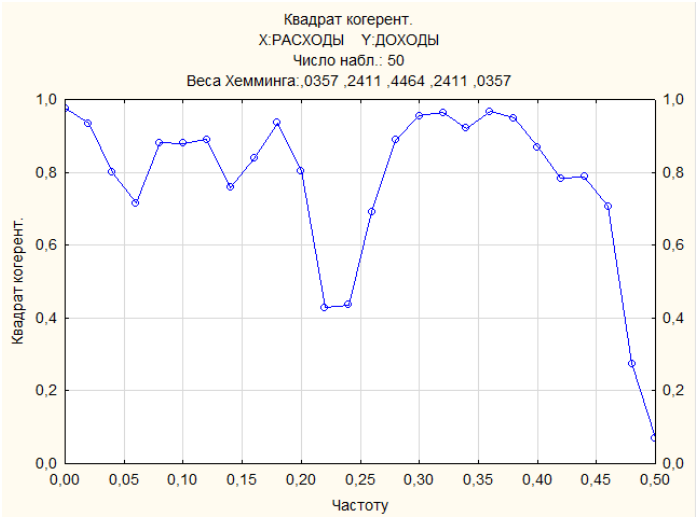
Квадратурная плотность:



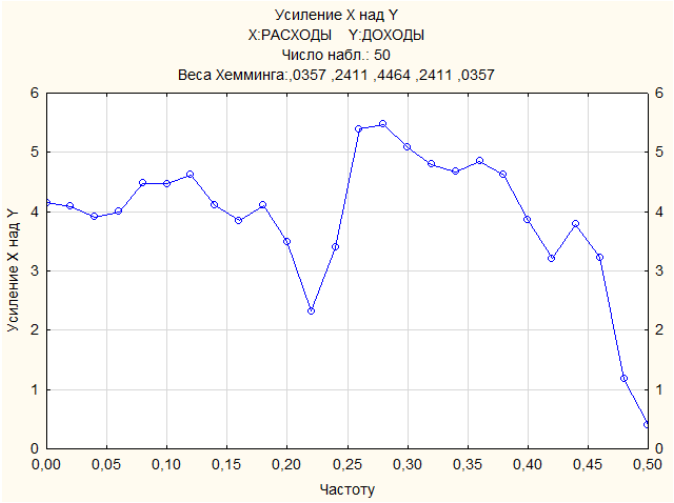
Кросс-амплитуда:



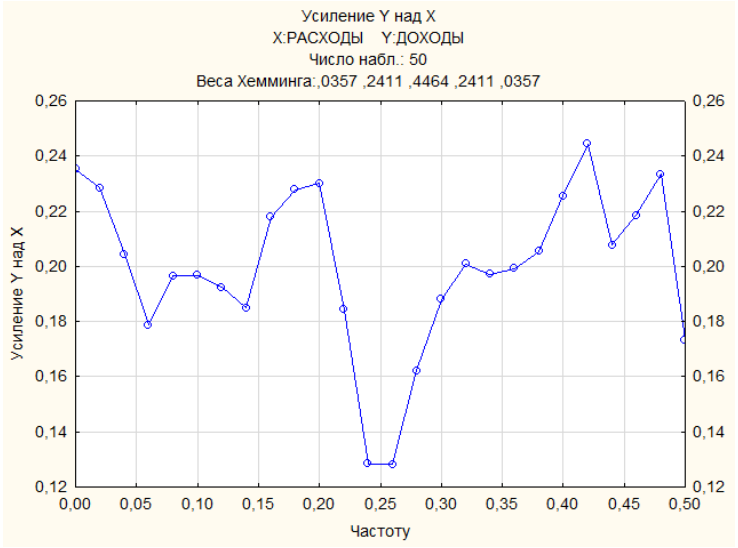
Квадрат когерентности:



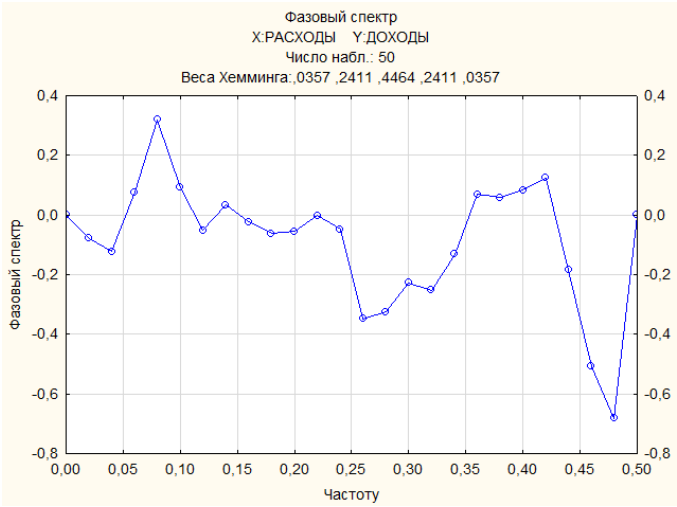
Усиление X над Y:



Усиление Y над X:



Фазовый спектр:



Сделайте выводы самостоятельно.

Лабораторная работа № 5. Программный комплекс АС ДРМ. Предварительный анализ временных рядов.

Задание к лабораторной работе № 5.

Для двух ВР осуществить предварительный анализ с помощью пакета программ АС ДРМ:

1) вывести описательную статистику;

2) проверить на стационарность, сделать выводы о стационарности ряда;

3) построить график, сделать вывод о наличии тренда (линейного, квадратичного и т. д.), циклических изменений, о наличии выбросов и аномалий;

4) построить автокорреляционную функцию, сделать выводы о стационарности ВР и о параметрах авторегрессионной модели;

5) провести спектральный и вейвлет-анализ, сделать выводы о наиболее значимых гармониках в ВР (по максимальным значениям спектральной плотности);

6) провести фрактальный анализ ВР, сделать выводы о наличии тренда, привести все возможные графики и таблицы.

Цель работы: знакомство с программным комплексом АС ДРМ, приобретение навыков проведения предварительного анализа временных рядов в АС ДРМ.

Пример выполнения.

Программный комплекс АС ДРМ (Автоматизированная система динамического регрессионного моделирования) разработан на кафедре «Прикладная математика и информатика» Ульяновского государственного технического университета.

Программный комплекс АС ДРМ состоит из интерфейсной, управляющей части и модулей, реализующих заданные алгоритмы.

Интерфейс АС ДРМ представлен на рис. 20.

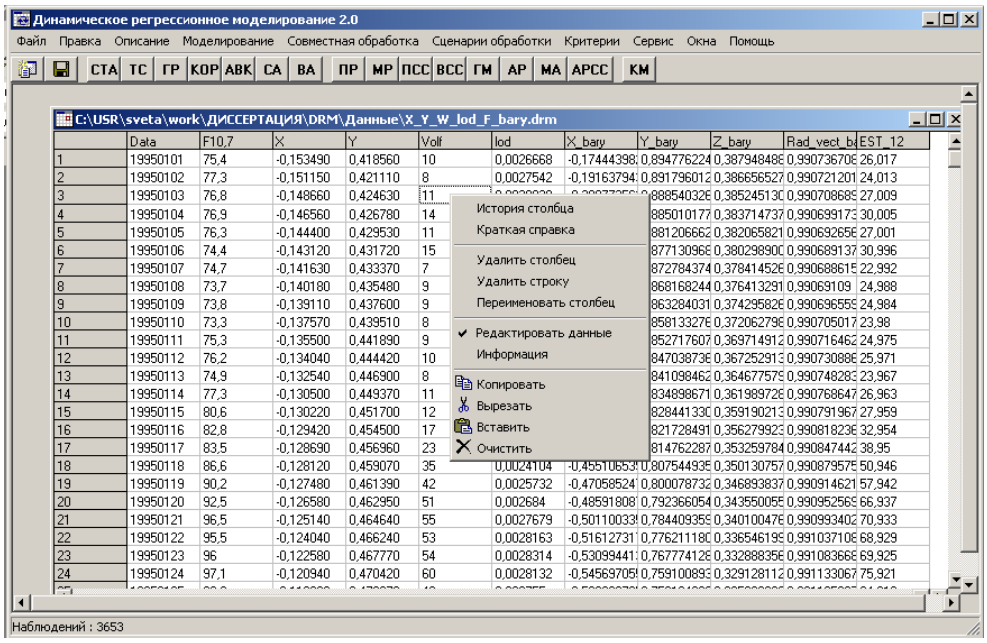


Рис. 20. Внешний вид основного окна пакета АС ДРМ

При запуске программы на экране появляется окно с меню, перечисляющее предоставляемые АС ДРМ возможности: Файл, Правка, Описание, Моделирование, Совместная обработка, Критерии, Сервис, Окна, Помощь и кнопки быстрого доступа к некоторым процедурам.

Пункт меню **Файл** позволяет осуществить стандартные действия по работе с файлами (открыть, создать, сохранить), а также корректно завершить работу программы.

Правка позволяет осуществить работу с данными. Реализованы процедуры работы со столбцами и строками: добавление/удаление столбцов/строк, добавление случайно сформированного столбца, столбцов времени, констант, процедура умножения столбца на число, функции округления и проверки данных на наличие ошибок.

Сервис предоставляет быстрый доступ к блокноту и калькулятору.

Пункт меню **Окна** позволяет управлять расположением окон на экране.

Справка выводит информацию о программе; содержит справочник для пользователей АС ДРМ.

Описание позволяет проводить предварительный анализ данных: вывод описательной статистики, исследование рядов на стационарность, построение графиков ВР, вычисление корреляционной и автокорреляционной функций, спектральный и вейвлет-анализы, фрактальный анализ, центрирование и усреднение данных.

Пункт **Моделирование** включает формирование моделей трендовой составляющей, множественной регрессии, гармонической модели, авторегрессии, моделей АРСС (ARMA), мартингальной аппроксимации, ARCH-моделей, построение комплексной модели и прогноза на будущее.

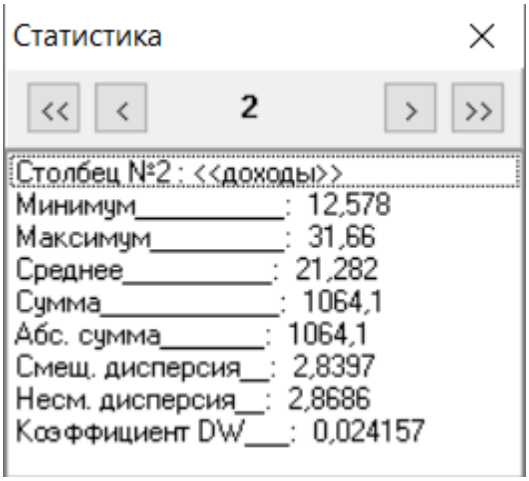
Пункт **Критерии** предназначен для исследования качества моделирования на каждом этапе внешними, смешанными и внутренними мерами соответствия, а также для диагностики остатков.

Совместная обработка позволяет строить совместные графики нескольких временных рядов, проводить кросс-корреляционный анализ и кросс-спектральный анализ ВР.

В качестве исходных данных возьмем ВР X. В АС ДРМ можно открывать файлы Excel.

В соответствии с заданиями проведем предварительный анализ ВР.

Описательная статистика ВР (основные числовые характеристики ряда):

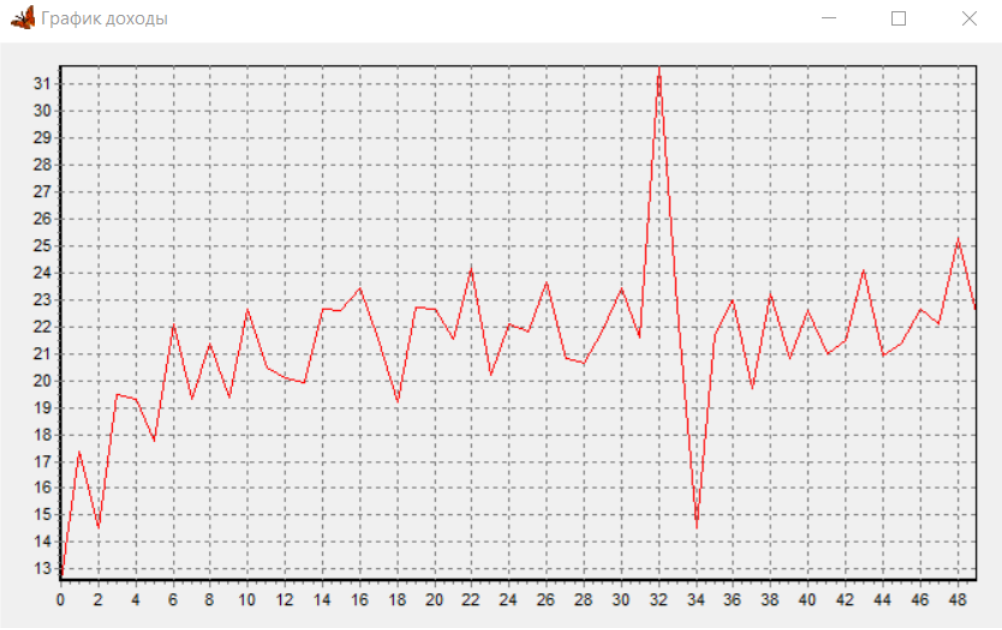


Проверка на стационарность:

```
Проверка на стационарность : доходы

_____
 Непараметрический критерий сдвига:
_____
Данные разбиты на две одинаковые группы объемом 25.
Доверительная вероятность - 0,95
Критическое значение - 5

Рассматривается первая группа.
Количество значений, превышающих максимум второй группы - 0
С доверительной вероятностью 0,95 отвергается гипотеза о том, что в реализации присутствует сдвиг среднего вверх.
Рассматривается вторая группа.
Количество значений, превышающих максимум первой группы - 24
С доверительной вероятностью 0,95 принимается гипотеза о том, что в реализации присутствует сдвиг среднего вниз.
_____

_____
 Критерий рассеяния.
_____
Данные разбиты на две одинаковые группы объемом 25.
Доверительная вероятность - 0,95
Критическое значение - 5

Число измерений, удовлетворяющих условию:
 x>max или x<min, где
      x принадлежит первой группе,
      max - максимум второй группы,
      min - минимум второй группы,
  равно 0
 Гипотеза о равенстве дисперсий первой и второй группы принимается.
```

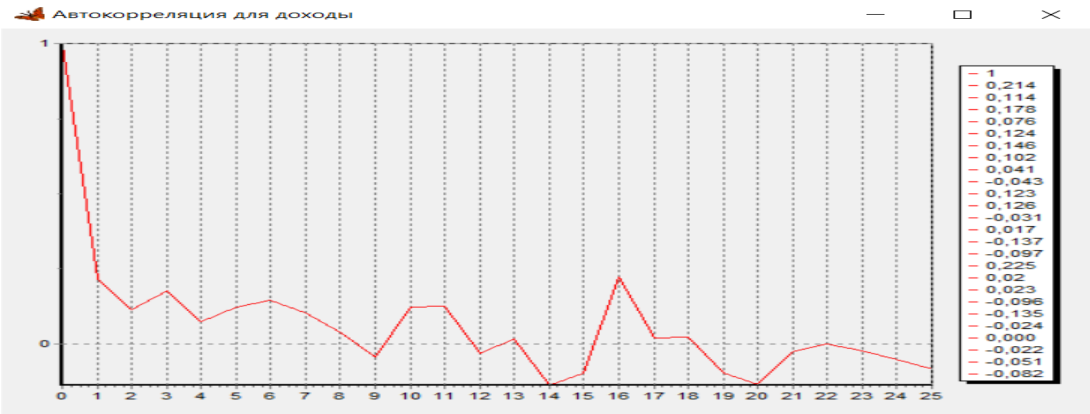
При проверке на стационарность получили, что исследуемый ряд является стационарным по дисперсии и нестационарным по среднему.

График ВР:



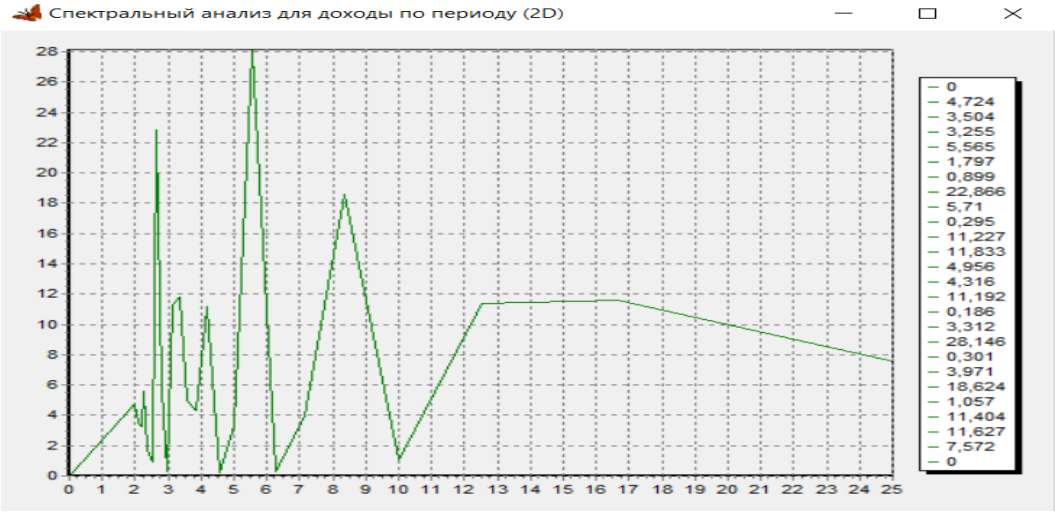
По графику можно выявить наличие возрастающего тренда и присутствие циклических изменений.

Автокорреляционная функция:



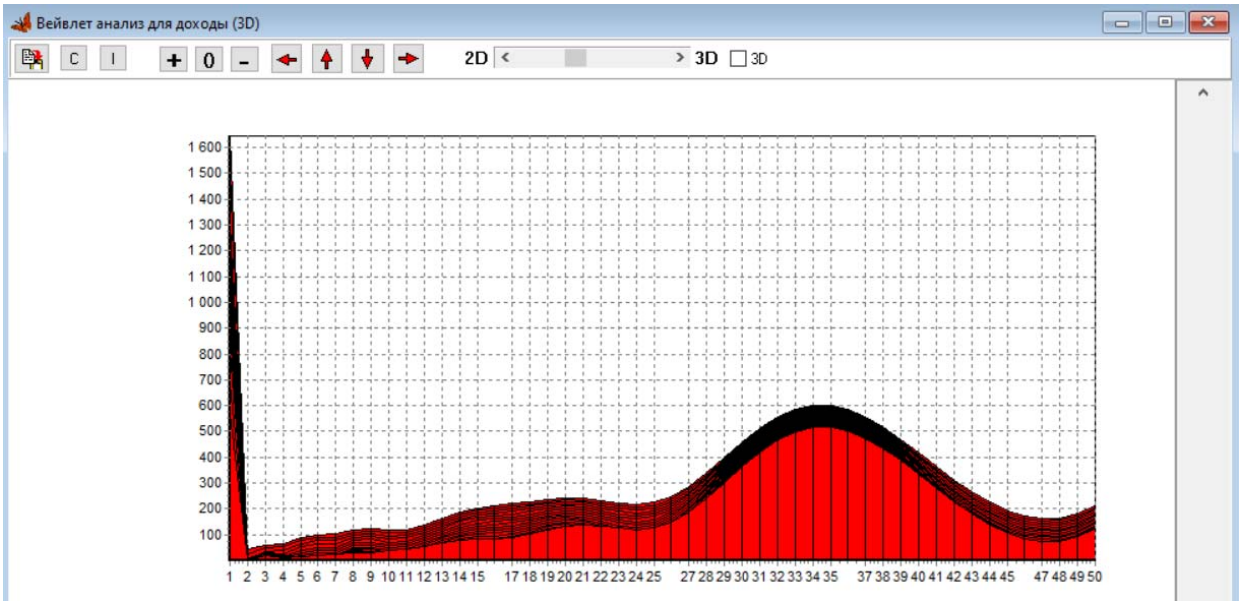
Автокорреляционная функция (АКФ) для доходов затухает значительно быстро, значения автокорреляции близки к 0, что говорит об отсутствии влияния предыдущих значений ВР на текущие.

Спектральный анализ:



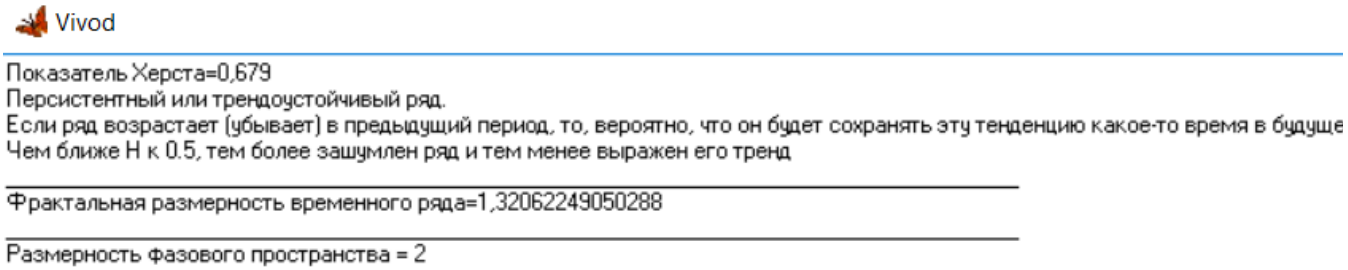
В спектре присутствует высокочастотная составляющая с большой амплитудой, следовательно, для временного ряда это указывает на небольшую случайную составляющую, небольшую неустойчивость описываемых им процессов, наличие шумов в данных.

Вейвлет-анализ:



Вейвлет анализ позволяет выявить не только значимые, но и устойчивые во времени гармоники.

Фрактальный анализ:



Показатель Херста H больше 0,5. Делаем вывод, что в данном ВР присутствует тенденция, и она достаточно устойчивая.

Лабораторная работа № 6. Программный комплекс АС ДРМ. Моделирование временных рядов.

Задание к лабораторной работе № 6.

Для двух ВР построить комбинированную модель:

1) построить простую регрессию (выбрать из имеющихся одну модель с наименьшим среднеквадратическим отклонением (СКО)), вывести график, СКО;

2) построить по остаткам от предыдущего шага простое и взвешенное скользящее среднее, вывести график, СКО;

3) построить по остаткам от предыдущего шага гармоническую модель (значимые гармоники выбирать по спектральному анализу, меняя значение допустимого уровня спектральной плотности), вывести график, СКО;

4) построить по остаткам от предыдущего шага модель авторегрессии (порядок выбрать с помощью программы), вывести график, СКО;

5) построить по остаткам от предыдущего шага мартингальную аппроксимацию, вывести график, СКО;

6) построить по остаткам от предыдущего шага модель АРСС, вывести график, СКО;

7) построить по остаткам от предыдущего шага ARCH-модель (все виды этих моделей), вывести график, СКО;

8) к остаткам от предыдущего шага применить фильтр Калмана, вывести график, СКО.

После каждого шага нужно:

1) делать вывод об улучшении модели (уменьшилось или увеличилось СКО и «внешнее» СКО);

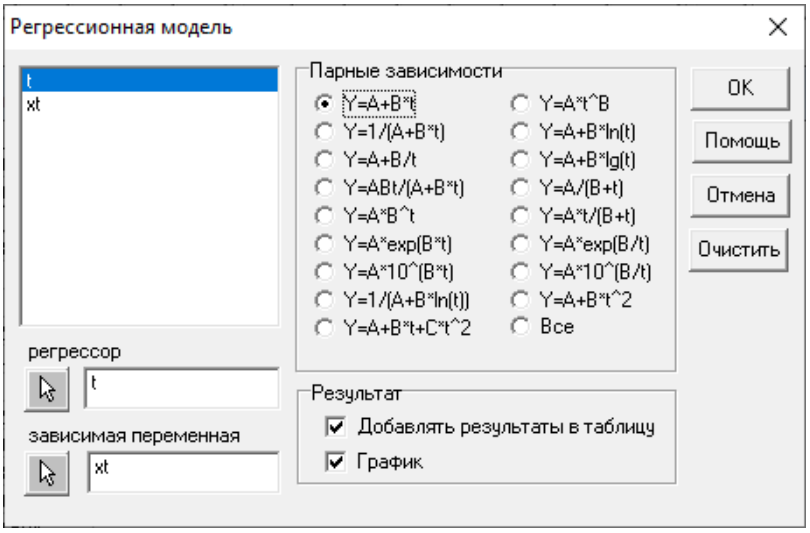
2) с помощью пункта меню «Критерии» провести диагностику построенных моделей (остатков): вывести значения внутренних, внешних (контрольная выборка) и смешанных мер соответствия, скользящего экзамена, сделать выводы об определенности модели, о нормальности ошибок, о равенстве нулю математического ожидания, о гетероскедастичности и наличии авторегрессии в остатках.

Цель работы: в ходе выполнения лабораторной работы студент приобретает навыки моедлирования ВР.

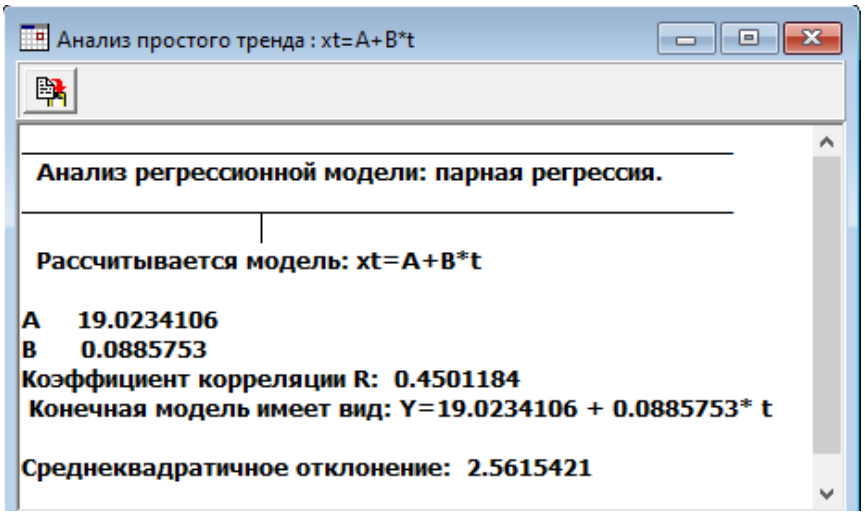
Пример выполнения.

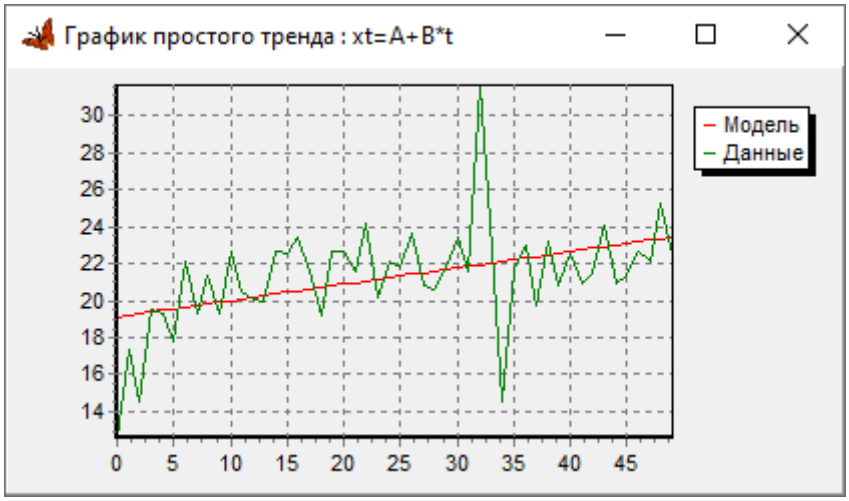
В качестве исходных данных возьмем ВР X. Построим ряд независимой переменной: в данном случае это столбец времени *t*. Его можно добавить вручную или через пункт меню Правка. Выполним первые два шага. Остальные шаги студенты выполняют самостоятельно.

1. Построим простую регрессию (для этого нужно выбрать один из имеющихся в АС ДРМ 17 видов тренда с наименьшим СКО). Чтобы построить линейную регрессию, нужно выбрать в меню: Моделирование – Простая регрессия или нажать на кнопку ПР. В появившемя диалоговом окне выбираем в качестве регрессора столбец времени, в качестве независимой переменной – временной ряд, выбираем вид регрессии и нажимаем ОК.



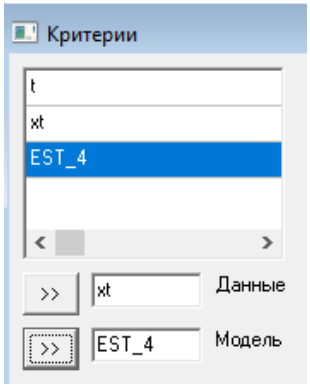
Результатом вышеописанной процедуры будут два окна, в одном из которых отображаются модели и некоторые меры качества, во втором – график. Обратите внимание, что после построения модели в таблице формируется новый столбец, в котором высчитываются остатки после удаления тренда из ВР.





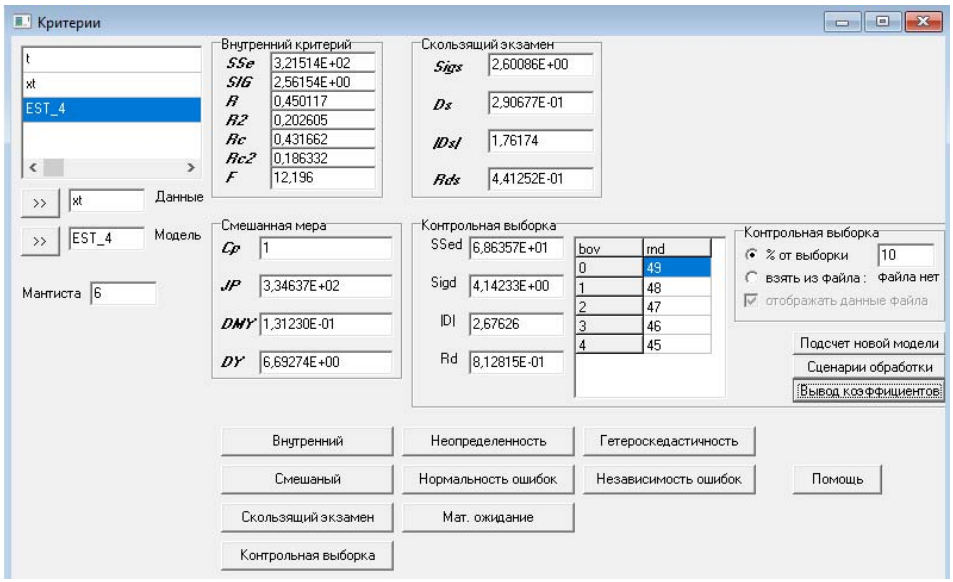
Далее проводим анализ качества построенной модели и диагностику остатков.

В меню выбираем Критерии. В появившемся диалоговом окне в качестве данных выбирается исходный временной ряд, в качестве модели – остатки, полученные после построения линейной регрессии.



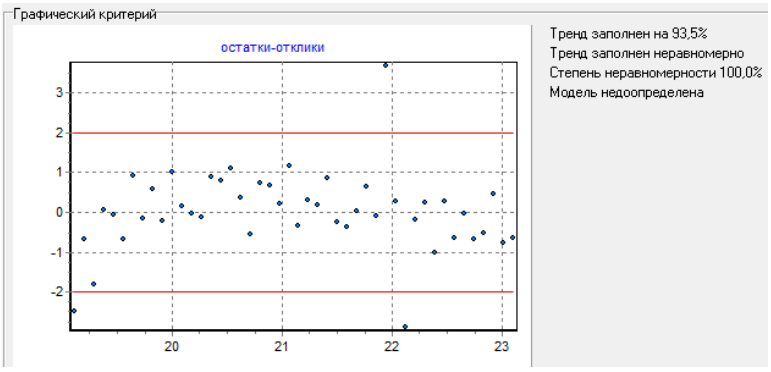
По очереди выбираем вкладки в окне диагностики остатков:

1) Вкладки «Внутренний», «Смешанный» и «Контрольная выборка» выводят значения внутренних, смешанных и внешних мер соответственно.



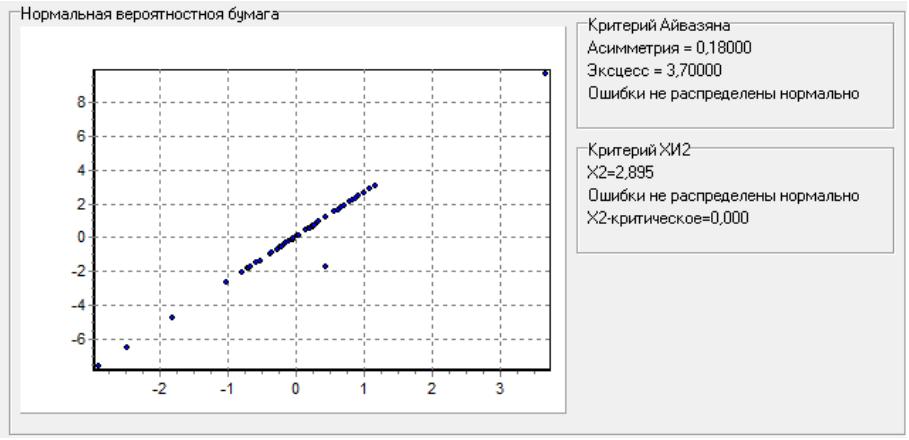
По данным мерам качества делаем вывод, что модель значима по F-критерию, но недостаточно хорошо описывает исходный ВР, так как значения стандартных отклонений очень большие, а значения коэффициентов корреляции и детерминации близки к нулю. При этом внешние меры говорят, что данная модель хорошо описывает тенденцию ряда, и по построенной модели можно предсказать поведение ряда в будущем.

2) Вкладка «Неопределенность» показывает, что модель недоопределена. Это означает, что в модель включены не все факторы, существенно влияющие на исследуемый процесс, и нужно добавить в модель новые регрессоры.

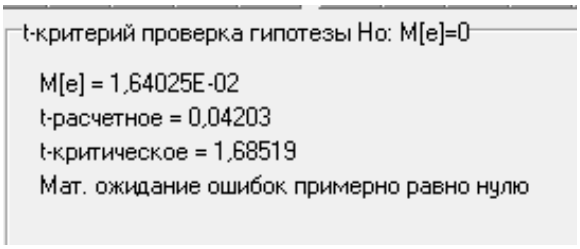


Избыточность модели, наоборот, говорит о том, что в модель включены факторы, которые не оказывают существенного влияния на исследуемый процесс.

3) Вкладка «Нормальность» проверяет условие теоремы Гаусса–Маркова о нормальности распределения остатков. В АС ДРМ реализованы два критерия проверки на нормальность: критерий Айвазяна и критерий хи-квадрат. Оба критерия показывают, что остатки после построения парной линейной регрессии не распределены нормально, что говорит о нарушении предположения регрессионного анализа.



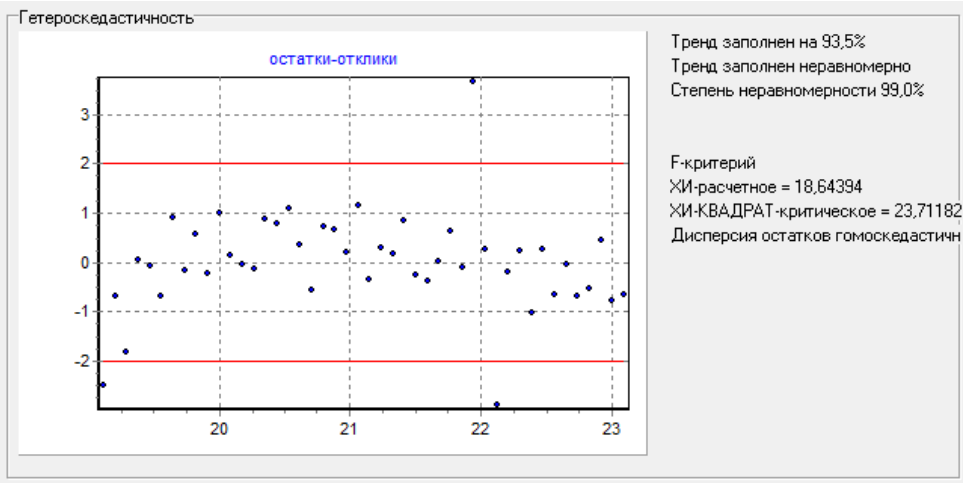
4) Вкладка «Мат.ожидание» проверяет условие теоремы Гаусса –Маркова о равенстве нулю математического ожидания остатков. Для проверки используется t-критерий. Это условие для нашего примера выполняется:



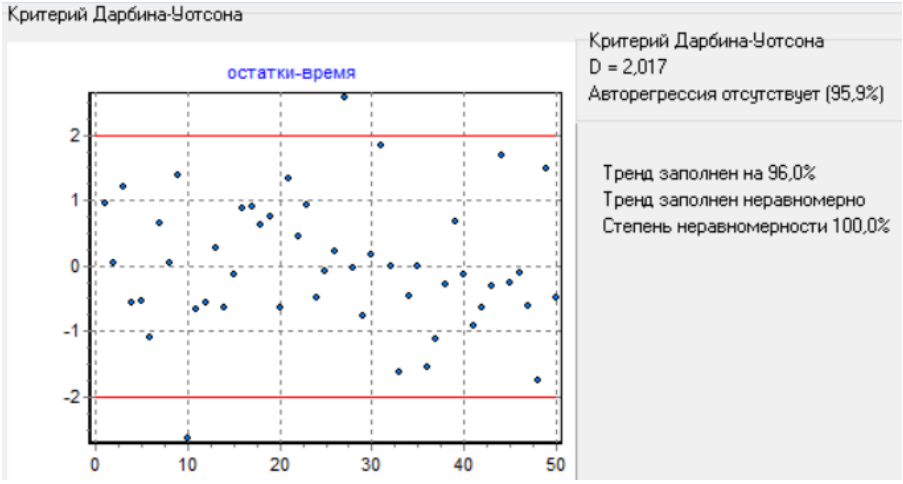
5) Вкладка «Гетероскедастичность» проверяет постоянство дисперсии в остатках, т.е. являются ли остатки гомоскедастичными или гетероскедастичными. Гетероскедастичность — понятие, используемое в прикладной статистике, означающее неоднородность наблюдений, выражающуюся в непостоянной дисперсии случайной ошибки регрессионной модели. Гетероскедастичность противоположна гомоскедастичности, означающей однородность наблюдений, то есть постоянство дисперсии случайных ошибок модели.

Наличие гетероскедастичности случайных ошибок приводит к неэффективности оценок, полученных с помощью МНК. Статистические выводы о качестве полученных оценок могут быть неадекватными. В связи с этим тестирование моделей на гетероскедастичность является одной из необходимых процедур при построении регрессионных моделей.

Для исследуемого в работе ВР получаем, что остатки гомоскедастичны, т.е. условие теоремы Гаусса-Маркова о постоянстве дисперсии остатков соблюдается.

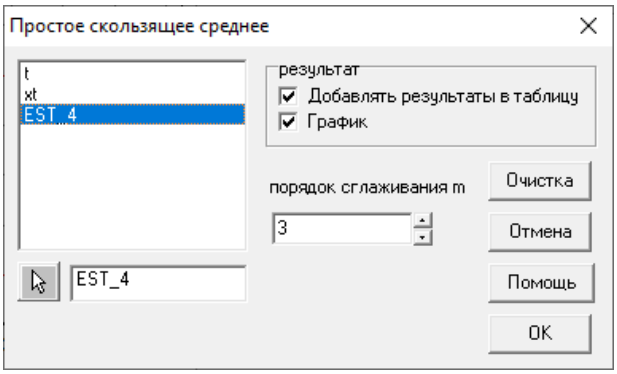


6) Вкладка «Независимость остатков» проверяет на наличие или отсутствие авторегрессии в остатках. Проверка осуществляется с помощью критерия Дарбина–Уотсона. Если значение критерия близко к 2, как в нашем случае (DW=2,017), то делается вывод об отсутствии авторегрессии. Если значение ближе к 0 или к 4, говорится, что авторегрессия присутствует в остатках.



Вывод: модель недоопределена, ошибки не распределены нормально, математическое ожидание ошибок примерно равно нулю, дисперсия остатков гомоскедантична, авторегрессия отсутствует.

2. Построим по остаткам от предыдущего шага простое и взвешенное скользящее среднее, выберем из них наилучшее по минимуму среднеквадратического отклонения. Для этого откроем окно построения простого скользящего среднего нажатием кнопки ПСС или через меню Моделирование – Простое скользящее среднее. В качестве данных для моделирования выбираются остатки после предыдущего шага. Длину интервала сглаживания примем равной 3.



В результате получаем модель вида:

Сренеквадратическое отклонение модели 2,1659.

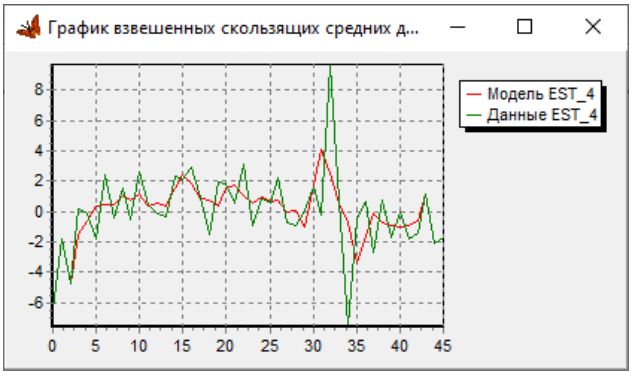
Аналогично проведем построение модели с помощью взвешенного скользящего среднего.

Анализ взвешенных скользящих средних:

Анализ переменной EST_4

Порядок сглаживания: 3
Веса:
w[-3] = -0,095238
w[-2] = 0,14286
w[-1] = 0,28571
w[0] = 0,33333
w[1] = 0,28571
w[2] = 0,14286
w[3] = -0,095238

Среднеквадратичное отклонение: 2,0124



Вывод: СКО для скользящих средних уменьшилось по сравнению с предыдущим шагом (построение простой регрессии), следовательно, модель улучшилась и в том, и в другом случае. Дальше вычисления проводятся для остатков после ВСС, так она дала меньшее СКО.

Проводим диагностику остатков по аналогии с предыдущим шагом, получаем следующие результаты: модель недоопределена, ошибки не распределены нормально, математическое ожидание ошибок

примерно равно нулю, дисперсия остатков гомоскедантична, авторегрессия отсутствует.

Лабораторные работы № 7, 8, 9. Программная реализация методов анализа, моделирования, прогнозирования временных рядов, критериев качества моделей и диагностики остатков

Задание к лабораторным работам № 7,8,9. Написать программу, реализующую один из методов анализа, моделирования, прогнозирования временных рядов, критериев качества моделей и диагностики остатков.

Цель работы: изучить и уметь применять на практике методы анализа, моделирования, прогнозирования временных рядов, критериев качества моделей и диагностики остатков, сформировать навыки программной реализации этих методов.

Рекомендации. Задания к лабораторным работам № 7, 8, 9 выдаются преподавателем каждому студенту индивидуально. Язык программирования студенты выбирают самостоятельно. Результаты работы программы требуется продемонстрировать на 3-4 примерах временных рядов. Примерные задания к лабораторным работам:

1. Программная реализация фрактального анализа: расчет показателя Херста, построение R/S и H траекторий.
2. Программная реализация спектрального анализа.
3. Программная реализация кросс-спектрального анализа.
4. Программная реализация построения автокорреляционной функции и частной автокорреляционной функции.
5. Программная реализация поиска коэффициентов кросс-корреляции.
6. Программная реализация построения модели авторегрессии порядка p.

7. Программная реализация алгоритмов поиска порядка авторегрессии.

8. Программная реализация построения модели скользящего среднего порядка q.

9. Программная реализация построения модели ARMA(p, q).

10. Программная реализация построения модели линейного и квадратичного тренда с проверкой качества моделей по критерию Фишера.

11. Программная реализация построения модели степенного и экспоненциального тренда с проверкой качества моделей по средней ошибке аппроксимации.

12. Программная реализация критерия Дарбина-Уотсона.

13. Программная реализация критерия Шапиро-Уилка.

14. Программная реализация критерия Фостера-Стюарта.

15. Программная реализация критерия Аббе.

16. Программная реализация критерия восходящих и нисходящих серий.

17. Программная реализация MHat-вейвлета.

18. Программная реализация вычислений показателей динамики временного ряда.

19. Программная реализация экспоненциального сглаживания ВР.

20. Программная реализация теста Энгеля-Грэйнджера.

21. Программная реализация критерия рассеяния.

22. Программная реализация теста Манна-Уитни.

23. Программная реализация теста Сиджела-Тьюки.

Приложение 1

Образец титульного листа отчета по лабораторной работе

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Факультет информационных систем и технологий
Кафедра «Прикладная математика и информатика»
Дисциплина «Анализ временных рядов»

Отчёт по лабораторной работе № 1
«Анализ временных рядов средствами Excel»

Выполнила:
студентка группы ПМбд-31
Савельева Д.А.
Проверила:
к.т.н., доцент кафедры ПМИ
Алексеева В.А.

Ульяновск
<год>

Приложение 2

Квантили нормального распределения

p	**0,90**	**0,95**	**0,975**	**0,99**	**0,995**	**0,999**	**0,9995**
u	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Квантили распределения хи-квадрат

$k \setminus p$	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**	**0,999**
2	4,605	5,991	7,378	9,210	10,597	13,815
3	6,251	7,815	9,348	11,345	12,838	16,266
4	7,779	9,488	11,143	13,277	14,860	18,466
5	9,236	11,070	12,832	15,086	16,750	20,515
6	10,645	12,592	14,449	16,812	18,548	22,457
7	12,017	14,067	16,013	18,475	20,278	24,321
8	13,362	15,507	17,535	20,090	21,955	26,124
9	14,684	16,919	19,023	21,666	23,589	27,877
10	15,987	18,307	20,483	23,209	25,188	29,588
11	17,275	19,675	21,920	24,725	26,757	31,264
12	18,549	21,026	23,337	26,217	28,300	32,909
13	19,812	22,362	24,736	27,688	29,819	34,527
14	21,064	23,685	26,119	29,141	31,319	36,124
15	22,307	24,996	27,488	30,578	32,801	37,698
16	23,542	26,296	28,845	32,000	34,267	39,252
17	24,769	27,587	30,191	33,409	35,718	40,791
18	25,989	28,869	31,526	34,805	37,156	42,312
19	27,204	30,144	32,852	36,191	38,582	43,819
20	28,412	31,410	34,170	37,566	39,997	45,314
21	29,615	32,671	35,479	38,932	41,401	46,796
22	30,813	33,924	36,781	40,289	42,796	48,268
23	32,007	35,172	38,076	41,638	44,181	49,728
24	33,196	36,415	39,364	42,980	45,558	51,179
25	34,382	37,652	40,646	44,314	46,928	52,619
26	35,563	38,885	41,923	45,642	48,290	54,051
27	36,741	40,113	43,195	46,963	49,645	55,475
28	37,916	41,337	44,461	48,278	50,994	56,892
29	39,087	42,557	45,722	49,588	52,335	58,301
30	40,256	43,773	46,979	50,892	53,672	59,702
35	46,059	49,802	53,203	57,342	60,275	66,619
40	51,805	55,758	59,342	63,691	66,766	73,403
45	57,505	61,656	65,410	69,957	73,166	80,078
50	63,167	67,505	71,420	76,154	79,490	86,660
60	74,397	79,082	83,298	88,379	91,952	99,608
70	85,527	90,531	95,023	100,425	104,215	112,317

Квантили распределения Стьюдента

$k \setminus p$	**0,900**	**0,950**	**0,975**	**0,990**	**0,995**	**0,999**
2	1,886	2,920	4,303	6,965	9,925	22,3
3	1,638	2,353	3,182	4,541	5,841	10,2
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,106	4,025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	17,34	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,458
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,398
30	1,310	1,697	2,042	2,457	2,750	3,385
40	1,303	1,684	2,021	2,423	2,704	3,307
60	1,296	1,671	2,000	2,390	2,660	3,232
120	1,289	1,658	1,98	2,358	2,617	3,160

Квантили распределения Фишера при $p = 0{,}95$

$k_2 \mid k_1$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**20**	**30**	**40**	**60**	**120**
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	248,0	250,1	251,1	252,2	253,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,46	19,47	19,48	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,62	8,59	8,57	8,55
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,75	5,72	5,69	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,50	4,46	4,43	4,40
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,81	3,77	3,74	3,70
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,38	3,34	3,30	3,27
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,08	3,04	3,01	2,97
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,86	2,83	2,79	2,75
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,77	2,70	2,66	2,62	2,58
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,54	2,47	2,43	2,38	2,34
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	2,31	2,27	2,22	2,18
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,19	2,15	2,11	2,06
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,11	2,06	2,02	1,97
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,04	1,99	1,95	1,90
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,01	1,92	1,87	1,82	1,77
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,84	1,79	1,74	1,68
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,84	1,74	1,69	1,64	1,58
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,75	1,65	1,59	1,53	1,47
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,66	1,55	1,50	1,43	1,35

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Прогнозирование временных рядов в пакете Statistica : методические указания / сост. Л.И. Дубровская. – Томск : Томский государственный университет, 2012. – 36 с.

2. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование: учебник. — М.: Финансы и статистика, 2001. — 228 с.

3. Агапова, Е. Г. Анализ временных рядов : учебное пособие / Е. Г. Агапова ; [науч. ред. Т. М. Попова]. - Хабаровск : Изд-во Тихоокеан. гос. ун-та, 2015. – 55 с.

4. Шанченко, Н. И. Лекции по эконометрике : учебное пособие для студентов высших учебных заведений, обучающихся по специальности «Прикладная информатика (в экономике)» / Н. И. Шанченко. – Ульяновск : УлГТУ, 2008. – 139 с.

5. Елисеева И.И., Курышева С.В., Гордеенко Н.М. Эконометрика. – М.: Финансы и статистика, 2003.

6. Бокс Дж., Дженкинс Т. Анализ временных рядов. Прогноз и управление. – М.: Мир, 1974. – 242 с.

7. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. – 758 с.

8. Канторович Г.Г. Анализ временных рядов // Экономический журнал ВШЭ. – 2002. – № 4. – С. 498–523.

9. Богачев В.В. Моделирование нестационарных процессов авторегрессионными моделями // Моделирование экономических процессов. – М.: МЭСИ, 1989.

10. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. – М.: ЮНИТИ, 1998. – 1024 с.

11. Валеев С.Г. Регрессионное моделирование при обработке наблюдений. – М.: Наука, 1991. – 272 с. (второе издание, дополненное и

переработанное: Валеев С.Г. Регрессионное моделирование при обработке данных. Казань: ФЭН, 2001. 296 с.)

12. Витязев В.В. Спектрально-корреляционный анализ равномерных временных рядов: учеб. пособие. – СПб.: Изд-во С.-Петерб. ун-та, 2001. – 48с.

13. Цветков В.П., Цветков И.В., Гуляева О.С. Фрактальный анализ валютных временных рядов // Финансы и кредит. – 2007. – № 9. – С. 30–35.

14. Астафьева Н.М. Вейвлет-анализ: основы теории и применения. // Успехи физических наук. – 1998. – Т.166. – № 11. – С. 1145–1170.

15. Витязев В.В. Вейвлет-анализ временных рядов: учебное пособие. – СПб.: Изд. С.-Петерб. унив-та, 2001. – 58 с.

16. Валеев С.Г., Куркина С.В. Программная реализация ДРМ-подхода для обработки и анализа временных рядов // Изв. вузов. Геодезия и аэрофотосъемка. – 2006. – № 5. – С. 10-21.

17. Валеев С.Г., Куркина С.В., Кувайскова Ю.Е. Модели сглаживания временных рядов // Труды междун. конф. «Континуальные алгебраические логики, исчисления и нейроинформатика в науке и технике»: Информатика, системы искусственного интеллекта и моделирование технических систем. – Ульяновск: УлГТУ, 2006. – С. 83-85.

18. Валеев С.Г., Фасхутдинова В.А. Модификация программного комплекса АС ДРМ применительно к обработке гео- и гелиофизических данных // Вопросы современной науки и практики. Университет им. В.И. Вернадского. – Тамбов: ТГТУ, 2008. – № 2 (12). – Том 2. – С. 64–68.

19. Валеев С.Г., Кувайскова Ю.Е., Куркина С.В., Фасхутдинова В.А. Программный комплекс для обработки временных рядов //

Вопросы современной науки и практики. Университет им. В.И. Вернадского. – Тамбов: ТГТУ, 2008. – № 4 (14). Том 2. – С. 102–107.

Учебное издание

АЛЕКСЕЕВА Венера Арифзяновна

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

Учебное пособие

Редактор А. В. Ганина

ЛР № 020640 от 22.10.97.
Подписано в печать 22.12.2020. Формат 60×84/16.
Усл. печ. л. 8,6. Тираж 50 экз. Заказ 61. ЭИ № 1647.

Ульяновский государственный технический университет,
432027, г. Ульяновск, ул. Сев. Венец, д. 32.

ИПК «Венец» УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.