Министерство образования и науки Российской Федерации

Федеральное государственное автономное образовательное учреждение   
высшего образования

«Южно-Уральский государственный университет

(Национальный исследовательский университет)»

Институт естественных и точных наук

Кафедра «Математическое и компьютерное моделирование»

Направление 02.03.01 «Математика и компьютерные науки»

КУРСОВАЯ РАБОТА

по дисциплине «Методы прогнозирования и анализ рынка»

Тема: «Прогнозирование изменения цены акций ЗАО “Яндекс”»

Руководитель, доцент МиКМ

\_\_\_\_\_\_\_\_\_\_\_\_\_\_Акимова А. А.

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2018 г.

Автор работы

Студент группы ЕТ-411

\_\_\_\_\_\_\_\_\_\_\_\_\_Смирнов Н. А.

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2018 г.

Работа защищена

с оценкой (прописью, цифрой)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2018 г.

Челябинск, 2018

ОГЛАВЛЕНИЕ

[ВВЕДЕНИЕ 3](#_Toc532536548)

[1. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ВРЕМЕННОГО РЯДА 4](#_Toc532536549)

[1.1. Выявление и устранение аномальных наблюдений 4](#_Toc532536550)

[1.2. Сглаживание временного ряда 5](#_Toc532536551)

[1.3. Предварительные расчеты 8](#_Toc532536552)

[2. АДДИТИВНАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА 10](#_Toc532536553)

[2.1. Выявление наличия неслучайной составляющей 10](#_Toc532536554)

[2.2. Построение моделей трендовой составляющей 13](#_Toc532536555)

[2.3. Выбор адекватной модели трендовой составляющей 16](#_Toc532536556)

[2.4. Выделение сезонной составляющей 21](#_Toc532536557)

[2.5. Построение аддитивной модели 23](#_Toc532536558)

[3. ЛИНЕЙНАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА 25](#_Toc532536559)

[3.1. Выявление наличия свойства стационарности временного ряда 25](#_Toc532536560)

[3.1.1 Математическое описание 25](#_Toc532536561)

[3.1.2. Реализация в пакете EViews 27](#_Toc532536562)

[3.2. Построение модели авторегрессии и проинтегрированного скользящего среднего ARIMA(p,k,q) 30](#_Toc532536563)

[3.1.1. Построение авторегрессионной модели AR(p) 31](#_Toc532536564)

[3.1.2. Построение модели скользящего среднего MA(q) 32](#_Toc532536565)

[3.1.3. Построение смешанной модели ARMA(p,q) 33](#_Toc532536566)

[3.1.4. Сравнительный анализ построенных моделей 34](#_Toc532536567)

[4. КОМБИНИРОВАННАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА 38](#_Toc532536568)

[5. ТОЧЕЧНЫЙ ПРОГНОЗ ПО ПОСТРОЕННЫМ МОДЕЛЯМ 41](#_Toc532536569)

[5.1. Точечный прогноз по аддитивной модели 41](#_Toc532536570)

[5.2. Точечный прогноз по линейной модели 41](#_Toc532536571)

[5.3. Точечный прогноз по комбинированной модели 42](#_Toc532536572)

[5.4. Сравнительный анализ построенных прогнозов 43](#_Toc532536573)

[ЗАКЛЮЧЕНИЕ 44](#_Toc532536574)

[ПРИЛОЖЕНИЕ А. 45](#_Toc532536575)

[БИБЛИОГРАФИЧЕСКИЙ СПИСОК 46](#_Toc532536576)

# ВВЕДЕНИЕ

Основной целью данной курсовой работы является анализ имеющихся данных об объекте и о выбранном секторе рынка, прогнозирование на некоторый будущий период времени средствами технического анализа.

Технический анализ - это прогнозирование изменений значений экономических показателей в будущем на основе анализа изменений этих значений в прошлом. В основе его лежит анализ временных рядов значений и их графиков. Наиболее часто методы технического анализа используются для анализа цен, изменяющихся свободно, например, на биржах.

Случайный процесс (СП) – параметризированное семейство случайных величин {Y(τ)}, где параметр τ – время.

СП называется СП с непрерывным временем, если система, в которой он протекает, меняет свои состояния в любой момент времени.

Временной ряд – набор упорядоченных во времени случайных величин:

{Y()} = Y(), Y(), …, Y(),

т.е. наблюдений над случайным процессом с непрерывным временем Y(τ) в моменты времени , где – моменты времени и i=1, 2, …, n; – наблюдение над случайной величиной Y(т.е. конкретная реализация не является случайной величиной. Далее наблюдения будем называть фактическими наблюдениями.

Все расчеты приведены на основе цен акций ЗАО «Яндекс» в период с 19.12.2016 по 28.02.2018 (количество наблюдений n=300). Исходные данные представлены в приложении 1.

# ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ВРЕМЕННОГО РЯДА

На этапе предварительной обработки производится построение и изучение графика временного ряда (диаграммы рассеяния). Графические методы анализа позволяют сделать предварительные выводы о характере процесса, которые затем могут быть проверены и уточнены с помощью расчета конкретных характеристик ряда.

В пункте 1.1 временной ряд будет проверен на наличие аномальных наблюдений с помощью критерия Ирвина. В пункте 1.2 мы проведем предварительные расчеты: найдем числовые характеристики временного ряда, построим коррелограмму.

## Выявление и устранение аномальных наблюдений

Под аномальным наблюдением понимается значение уровня ряда, нехарактерное для динамики изучаемого процесса.

Критерий Ирвина позволяет установить наличие аномальных уровней во временном ряду, то есть уровней, которые не отвечают возможностям исследуемой экономической системы и существенно влияют на основные характеристики ряда.

Расчетные значения статистики Ирвина λt вычисляются с помощью следующих формул:

Согласно критерию Ирвина, с вероятностью 5% во временном ряду присутствуют аномальные наблюдения, так как существуют такие λt, что λt>λкрит, где λкрит =0,851.

В случае если во временном ряду присутствуют аномальные наблюдения, необходимо их заменить, используя следующую формулу:

В данном временном ряду аномальных наблюдений не выявлено.

## Предварительные расчеты

Рассмотрим числовые характеристики временного ряда в случае аддитивной модели.

Математическое ожидание оценивается как выборочное среднее:

Дисперсия оценивается как выборочная дисперсия:

Степень статистической связи между последовательностями и , сдвинутыми относительно друг друга на *l* моментов времени, т.е. с лагом (порядком коэффициента автокорреляции) *l*, может быть определена с помощью коэффициента автокорреляции

Коэффициент автокорреляции *p(l)* оценивается как выборочный коэффициент автокорреляции**:**

Нормированная автокорреляционная функция временного ряда – последовательность коэффициентов автокорреляции *r(0), r(1), r(2),….*

Для анализа значений нормированной автокорреляционной функции удобно использовать график, который называется коррелограммой. Она изображает зависимость значений коэффициентов автокорреляцииот величины лага.

Частный коэффициент автокорреляции измеряет связь между текущим значением ряда и его предыдущими значениями , ,когда влияние всех промежуточных лагов устранено.

Математическое ожидание и дисперсию найдем с помощью средств MS Excel, а все остальные числовые характеристики временного ряда – в пакете EViews.

Расчетные значения для характеристик представлены в таблице.

Таблица1.2–расчетные значения

|  |  |
| --- | --- |
| Характеристика | Значение |
| Математическое ожидание | 29,10486667 |
| Дисперсия | 28,90084714 |

На рисунке 1.1 слева показана коррелограмма, справа–частная коррелограмма. (ПОМЕНЯТЬ)

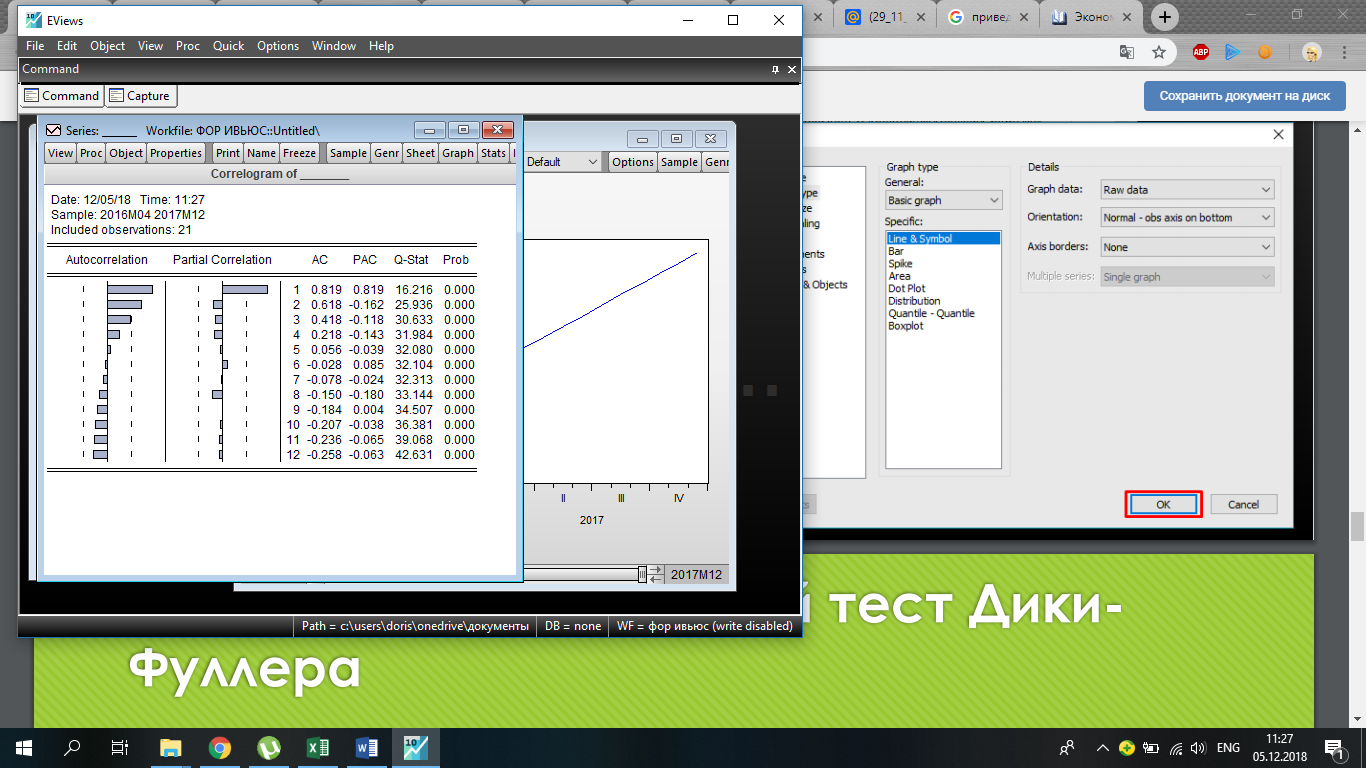


Рисунок 1.1 ­–Коррелограмма

Значения, полученные в ходе предварительной обработки, используются в дальнейших расчетах.

# АДДИТИВНАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА

Для анализа структуры временных рядов используют аддитивную модель, которая имеет следующий вид:

Y(τi) = q(τi) + *ε*(τi),

где *ε*(τi) - случайная составляющая, которая отражает случайные колебания или шумы процесса,q(τi) - неслучайная (систематическая) составляющая, которая может включать одну или несколько из следующих компонент:

- трендовая составляющая, наблюдаемая в течение длительного периода времени, которая отражает влияние долговременных факторов;

- сезонная компонента, которая описывает регулярные периодические колебания;

- периодическая (циклическая) компонента, которая описывает длительные периоды (более одного года) относительного подъема и спада.

Мы будем строить аддитивную модель, включающую только трендовую и сезонную составляющие.

Прежде всего, в разделе 2.1 мы покажем, что временной ряд имеет неслучайную составляющую. Далее, в разделе 2.1, построим 3 модели трендовой составляющей, после чего в разделе 2.3 сделаем выбор лучшей модели трендовой составляющей. В последнем разделе 2.4, на основании полученных результатов, построим аддитивную модель.

## Выявление наличия неслучайной составляющей

В аддитивной модели отсутствует неслучайная составляющая тогда и только тогда, когда (т. е., временной ряд состоит из статистически независимых наблюдений, случайно варьирующихся около некоторого постоянного значения).

Таким образом, вместо проверки наличия неслучайной составляющей будем проверять постоянство математического ожидания. Сформулируем две статистические гипотезы:

Рассмотрим два критерия для проверки этих гипотез.

Критерий 1 (метод доверительных разностейсредних уровней) имеет следующий алгоритм.

1. Разбить вариационный ряд на две части, и , где

;

*;*

*.*

2.Предположить, что и , и имеют нормальное распределение.

3. Найти для обеих частей выборочные средние и , выборочныедисперсии ; .

4. Применить Критерий Фишера (проверяет равенство дисперсий): есливыполнено неравенство

, где ,

вExcel ,

то дисперсии одинаковы с уровнем значимости , поэтому можно применять критерий Стьюдента.

Если неравенство не выполнено, дисперсии неодинаковы, нужно либо применить другой критерий (не Критерий 1), либо принять гипотезу .

По результатам расчета получаем: неравенство выполняется, следовательно, дисперсии равны и можно применить критерий Стьюдента.

1. Применить критерий Стьюдента (проверить постоянство математических ожиданий). Если выполнено:

где

в Excel ,

то гипотеза принимается с уровнем значимости .

Идея критерия: если выборочные средние , то тренда во временном ряде нет, если значительно отличаются – тренд есть.

После расчетов , а t = 2,59243, значит гипотеза на данном этапе не принимается.

Критерий 2 (критерий серий) необходим для проверки постоянства математических ожиданий. Таким образом, он позволяет проверить, является ли случайным порядок появления двух значений переменной. Серия – это последовательность похожих наблюдений. Если в выборке либо слишком много серий, либо слишком мало, то эта выборка не является случайной. Для проверки выборки на случайность необходимо сделать следующее.

1. Упорядочить члены временного ряда по возрастанию:

.

2. Найти выборочную медиану:

3. По элементам , расположенным согласно исходному временному ряду и отличным от , образовать последовательность серий по правилу:

.

Полученная последовательность плюсов и минусов характеризуется общим числом серий  и протяженностью самой длинной серии . Под «серией» понимается последовательность подряд идущих плюсов или подряд идущих минусов (в частном случае серия может состоять только из одного плюса или только из одного минуса и тогда ее протяженность равна единице). Очевидно, что если наблюдения стохастически независимы (выборка случайна), то чередование плюсов и минусов в последовательности должно быть более или менее «случайным», т. е. эта последовательность не должна содержать слишком длинных серий подряд идущих плюсов или подряд идущих минусов, и соответственно общее число серий  не должно быть слишком малым.

Если хотя бы одно из неравенств:

протяжённость самой длинной серии;=; = взятие целой части от не выполняется, то гипотеза отвергается с вероятностью ошибки , где . Здесь – взятие целой части числа z.

При n = 300, расчётные значения: = 29,57; , . , а . Следовательно, не выполняются оба неравенства. Значит, отвергается гипотеза , принимается гипотеза , т.е. математическое ожидание не постоянно, временной ряд не случайный, т.е. содержит не случайную составляющую.

## Построение моделей трендовой составляющей

Тренд - это аналитическая функция, которая описывает тенденцию изменения явления и связывает единым законом развития все последующие уровни ряда динамики.

Трендовая составляющая отражает устойчивую и долговременную тенденцию временного ряда.

Таким образом, возникает задача выделения тренда, т.е. построения оценки для функции *t*(τ) (или оценок () для значений t ()) по заданной временной выборке {, yi}. При этом предполагается, что остальные составляющие *p*(τ), *s*(τ) временного ряда отсутствуют.

Полагая время τ независимой переменной, оценим функцию , используя метод парной регрессии, который основан на аддитивной модели временного ряда:

Y() = t() + ε(τi),

где случайные величины ε(τi) удовлетворяют условиям Гаусса-Маркова:

1. дисперсии δi – конечные, одинаковые, независимые от измерений и таковы, что

M(ε(τi)) = 0,

M(ε(τi), ε(τj)) = ;

2. случайные величины ε(τi) имеют нормальное распределение N (0, δ2).

Для того, чтобы определить характер тренда, необходимо выбрать вид функции . Предварительно мы анализируем графическое изображение ряда, т.е. строим диаграмму рассеивания по точкам *{yi}*. С помощью диаграммы рассеяния можно сделать выбор вида функции из представленных ниже.

1. Линейная функция

используется для представления процессов с постоянной скоростью изменения.

2. Экспоненциальная функция

используется для представления процессов, которые характеризуются увеличением значения трендовой составляющей в одинаковое число раз за равные промежутки времени.

3. Полиномиальная функция

используется для представления процессов, в которых с течением времени тенденция меняется с возрастающей на убывающую (или наоборот).

4. Логарифмическая функция

используется для представления процессов, для которых характерно насыщение рынка.

После выбора вида функции построим уравнение регрессии , зависящее от коэффициентов которые являются оценками коэффициентов функции тренда

Вычислим коэффициенты по методу наименьших квадратов, минимизируя отклонение расчетных значений от фактических:

Для этого нужно найти частные производные 1 порядка по каждому из коэффициентов приравнять к нулю и решить полученную систему из k уравнений.

Получив коэффициенты , примем составленное уравнение регрессии (τ) в качестве оценки для функции тренда которая может быть использована для дальнейшего анализа временного ряда или его прогнозирования.

В результате визуального анализа диаграммы рассеяния, представленной на рисунке 2.1, для определения характера тренда выбраны функции: линейная и полиномиальная пятой степени.

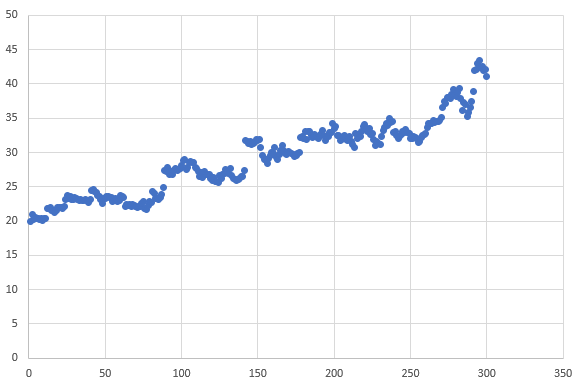


Рисунок 2.1 – Диаграмма рассеяния

С помощью встроенных функций MSExсel найдем коэффициенты уравнений регрессии. Полученные уравнения трендовой составляющей представлены в таблице 2.1.

Таблица 2.1 – Уравнения трендовой составляющей

|  |  |
| --- | --- |
| Модель |  |
| Линейная |  |
| Полиномиальная 5-ой степени |  |

## Выбор адекватной модели трендовой составляющей

Независимо от вида и способа построения трендовой составляющей вопрос о ее выборе и возможности применения в экономико-математической модели может быть решен только после установления адекватности и точности линии тренда. Построенная модель является адекватной если она соответствует данным наблюдения.

Проверка адекватности модели основывается на анализе ряда остатков

где расчетное значение неслучайной составляющей временного ряда.

Трендовая составляющая является адекватной, если остатки

1. являются случайными;

2. имеют равное нулю математическое ожидание;

3. независимы между собой.

1. Случайность ряда остатков проверяется по критерию серий. Предварительно выбирается медиана упорядоченного ряда остатков. Каждому элементу ряда остатков ставится в соответствие знак «+», если , и знак «–», если . Непрерывно идущую последовательность одинаковых знаков принято называть серией. Определяется максимальная длина серии *l* и число серий *v*. Остатки считаются случайными на уровне значимости 0,05, если одновременно выполняются два условия

Для полиномиальной модели 5 степени т.е. ряд остатков не случайный. Для линейной модели т.е. ряд остатков случайный.

2. Проверка равенства нулю математического ожидания проводится с помощью критерия Стьюдента. Сформулируем две гипотезы.

Если тренд оценен достаточно точно и математическое ожидание остатков равно нулю, то гипотеза считается справедливой. В противном случае неслучайная составляющая содержит систематическую ошибку.

Возьмем

где .

Если попадает в критическую область

где , то отвергается с уровнем значимости . определятся через встроенные функции MSExcel. .

Расчетные значения по выбранным трендовым моделям представлены в таблице 2.2. выберем равным 0,05, тогда интервал будет иметь вид

У полиномиальной и линейной модели математическое ожидание ряда остатков не равно нулю, следовательно, тренд по этим моделям оценен недостаточно точно, т.е. неслучайная составляющая содержит систематическую ошибку.

Таблица 2.2— Расчетные значения для трендовых моделей

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Модель |  |  |  | Вывод |
| Линейная | -47,8866 | 65,79792991 | -12,6055674 | Не попадает в интервал |
| Полиномиальная 5-ой степени | 0,004466667 | 1,690808252 | 0,045756186 | Попадает в интервал |

Под независимостью ряда остатков понимается отсутствие в нем автокорреляции, т. е. отсутствует зависимость каждого значения ряда от предыдущих значений. Для проверки ряда остатков на отсутствие автокорреляции уровней остатков используется критерий Дарбина-Уотсона. Этот критерий основан на идее, что корреляция между сохраняется на в невязках Для проверки критерия необходимо рассчитать величину:

представляющую собой отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии. Между критерием Дарбина–Уотсона *d* и коэффициентом автокорреляции остатков первого порядка имеет место следующее соотношение:

Коэффициент автокорреляции первого порядка определяется по формуле:

Отсюда следует, что .

При положительной автокорреляции:.

При отрицательной автокорреляции:.

При отсутствии автокорреляции: .

Соотношения между критерием Дарбина-Уотсона и коэффициентом автокорреляции свидетельствует о том, что с некоторой вероятностью отсутствует автокорреляция.

Для более точного определения, какое значение *d* свидетельствует об отсутствии автокорреляции, а какое о ее наличии, построена таблица критических точек распределения Дарбина-Уотсона. По этой таблице для заданного уровня значимости , числа наблюдений и количества объясняющих переменных определяются значения нижней и верхней границы.

Вывод о наличии или отсутствии автокорреляции остатков принимается в зависимости от того в какой промежуток попадает значение критерия *d.*

– остатки имеют положительную автокорреляцию.

– зона неопределенности.

– автокорреляция отсутствует.

– зона неопределенности.

– остатки имеют положительную автокорреляцию.

Данные по трем моделям представлены в таблице 2.3.

Таблица 2.3 — Автокорреляция остатков

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Модель |  | *d* |  |  | Промежуток |
| Линейная | 0,307 | 0,836 | 1,22 | 1,42 |  |
| Полиномиальная | 0,956 | 0,088 | 0,83 | 1,96 |  |
| Логарифмическая | 0,36 | 1,28 | 1,22 | 1,42 |  |

Таким образом у линейной и полиномиальной модели присутствует автокорреляция остатков, у логарифмической трендовой модели ее нет.

Оценка точности модели тренда заключается в оценке близости модельных значений тренда к фактическим уровням ряда и осуществляется с помощью вычисления индекса детерминации.

Введем суммы:

; ,

где .

Индекс детерминации показывает, какая доля изменения временного ряда обусловлена изменением переменной , характеризует близость уравнения тренда к исходным данным, которые содержат «нежелательную» случайную составляющую .

В случае линейного тренда справедливо тождество , где – коэффициент детерминации линейной регрессии:

Если , то справедлива гипотеза : нелинейную регрессию можно заменить линейной.

Индекс детерминации можно использовать только тогда, когда значения чисел коэффициентов регрессии одинаково. При различных значениях необходимо использовать приведенный индекс детерминации.

При выборе уравнения регрессии нужно учитывать не только величину , но сложность уравнения, определяемое количеством коэффициентов уравнения. Такой учет реализован в приведенный индексе детерминации

где m – число коэффициентов в уравнении регрессии.

Необходимо посчитать все и выбрать из них наибольшее, которое и соответствует лучшему уравнению регрессии, эти данные представлены в таблице 2.4.

Таблица 2.4 – Расчетные значения индекса детерминации

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Модель |  |  |  |  |
| Линейная | 8641,353295 | 854,796916 | -9,109247159 | -9,281173131 |
| Полиномиальная  5-ой степени | 8641,353295 | 1982418,844 | 0,995641005 | 0,995566873 |

По проведенным расчетом видно, что линейная модель не подходит для точного описания данных временного ряда. Наиболее точной оказалось полиномиальная модель пятой степени.

Таким образом наиболее точной и адекватной моделью трендовой составляющей будет полиномиальное уравнение тренда пятой степени вида:

## Построение аддитивной модели

Таким образом после проведения всех расчетов временной ряд содержит неслучайную составляющую, таким образом общий вид аддитивной модели:

Y() = + + *ε*(),

где*ε*() – случайная составляющая, которая отражает случайные колебания или шумы процесса;

– трендовая составляющая, наблюдаемая в течение длительного периода времени, которая отражает влияние долговременных факторов;

– сезонная компонента, которая описывает регулярные периодические колебания.

Исключим влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. Получим величины

Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Аддитивная модель будет иметь вид:

# ЛИНЕЙНАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА

Временной ряд называется стационарным в широком смысле (далее – стационарным), если числовые характеристики случайных величин Y() не зависят от момента времени .

Методы определения наличия свойства стационарности данного ВР приведены в пункте 3.2. К таким методам относят следующие методы:

1. DF (DickeyFuller);

2. ADF (AdvancedDickeyFuller);

3. KPSS (KwiatkowkiPhillipsSchmidtShin).

Проверка наличия стационарности временного ряда позволяет определить класс моделей, которые могут быть построены для данного временного ряда. А именно, в пункте 3.3. строятся модели ARIMA(p,k,q) нестационарных временных рядов, а в подпунктах 3.3.1 – 3.3.3 – модели AR(p), MA(q), ARMA(p,q) стационарных временных рядов.

## Выявление наличия свойства стационарности временного ряда

### 3.1.1 Математическое описание

Будем говорить, что временной ряд имеет порядок интегрирования k т.е. , если сам временной ряд и все его ряды разностей порядка до k-1 включительно не стационарны, а ряд разностей порядка k - стационарен.

Здесь разность порядка k определяется как

,

где разность первого порядка .

Стационарный временной ряд имеет порядок интеграции ноль.

Последовательно проводится проверка первоначального временного ряда, ряда первых разностей, ряда вторых разностей и т.д., что позволяет определить порядок интегрирования временного ряда.

Для того, чтобы проверить наш временной ряд на стационарность, принято использовать несколько тестов. Основная идея тестов заключается в том, чтобы по данным временного ряда подобрать параметры одной из известных моделей, после чего, по значениям параметров модели, сделать вывод о наличии стационарности.

Проверку на стационарность будем проводить в пакете EViews.

Рассмотрим несколько тестов:

1. Тест Дикки-Фуллера(DF);

2. Расширенный Тест Дики-Фуллера(ADF);

3. Тест Квятковского-Филлипса-Шмидта-Шина(KPSS).

1. DF-тест основан на оценке параметра приведенногоуравнения

,

где - фактические наблюдения временного ряда;- ошибка.

Выдвигаются гипотезы:

1. - наш временной ряд не стационарен.

2. - наш временной ряд стационарен.

Если значение t-статистики Стьюдента для параметра меньше табличного(критического) значения DF-статистики, то принимается гипотеза , иначе принимается гипотеза .

Таблицы Дики-Фуллера (значения DF-статистики) рассчитаны для уровней значимости в 1, 5, 10 %.

DF-тест применим только если ряд случайных составляющих не автокоррелирован. Если данное условие нарушено, то необходимо включить разность с необходимым количеством лагов. Именно в этом случае применяется расширенный тест Дики-Фуллера.

2. Расширенный тест Дики-Фуллера (ADF-тест) имеет такие же критические значения.

Модель расширенного теста Дики-Фуллера имеет вид:

.

Например, модель расширенного теста Дики-Фуллера, использующая AR (2), имеет вид:

.

Данную модель можно представить в виде:

.

Для определения количества включаемых лагов необходимо при выборке объемом менее 81 наблюдения - 2 лага, при выборка объемом от 81 до 256 наблюдений - 3 лага, и т.д.

Вывод о наличии стационарности по ADF-тесту производится так же, как и в случае DF-теста.

3. Тест стационарности KPSS рассматривает ряд вида

где - стационарный процесс и - случайное блуждание, определяемое как ,

- нормально распределенная случайная величина с нулевым математическим ожиданием, и - коэфициент.

Выдвигаются две гипотезы:

1.: временной ряд стационарен ( или );

: временной ряд не стационарен (.

Вывод о наличии стационарности по тесту KPSS производится так же, как и в случае DF-теста.

### 3.1.2. Реализация в пакете EViews

Для того, чтобы проверить наличие стационарности временного ряда, воспользуемся статистическим пакетом EViews. Данный пакет позволяет быстро и наглядно получить искомые характеристики временного ряда.

Для выявления стационарности необходимо взять временной ряд с привязкой к дате.Встроенными функциями пакета EViews представим временной ряд графически, получим диаграмму рассеяния (рисунок 3.1).

По полученной диаграмме рассеяния можно сделать предположение, что временной ряд не стационарен, так как присутствует тренд. Для достоверности результатов при анализе временного ряда на принадлежность их к классу стационарных принято использовать несколько тестов. Основная идея тестов: по данным временного ряда подобрать параметры одной из известных моделей, по значениям параметров сделать вывод о наличии стационарности.

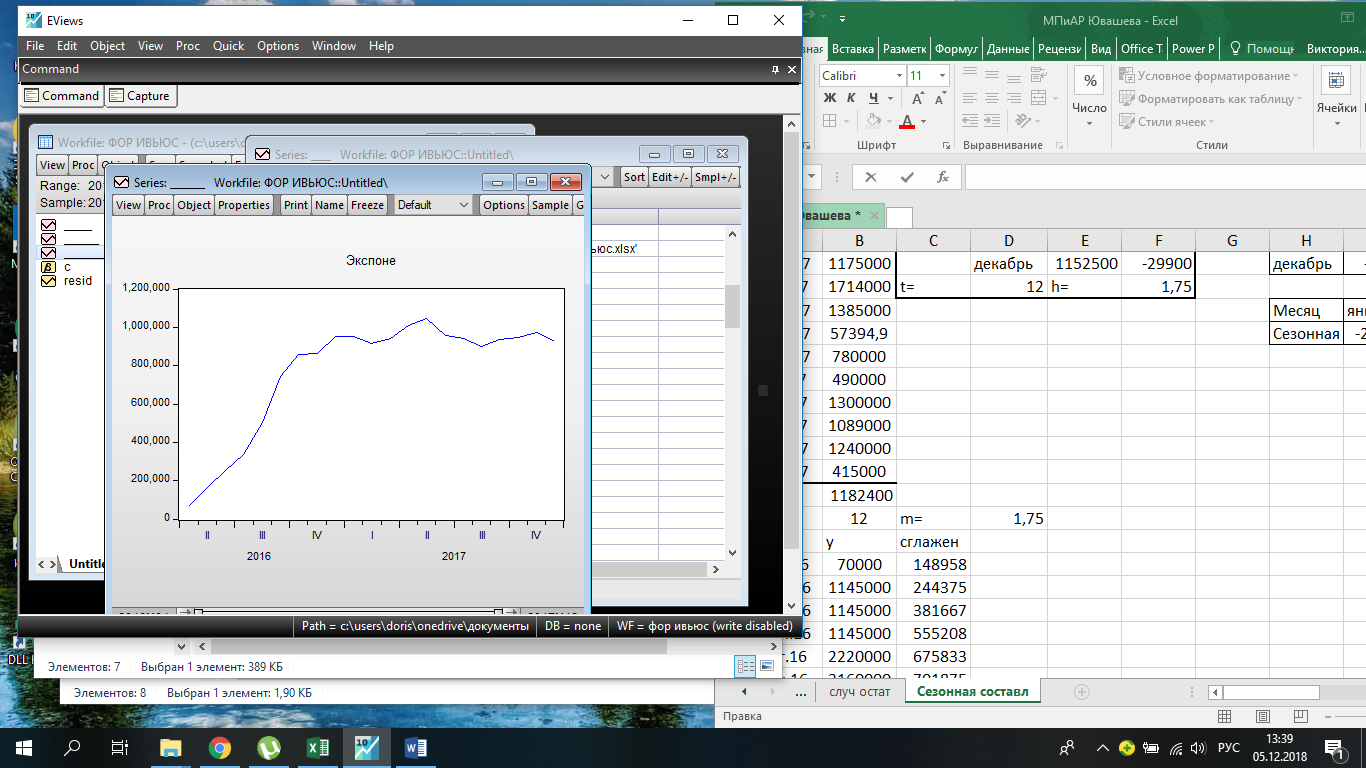


Рис 3.1 – Диаграмма рассеяния временного ряда

Пакет EViews предлагает несколько тестов на наличие стационарности. Рассмотрим следующие:

1. Тест Дикки-Фуллера.Результат теста представлен на рисунке 3.2. Таблицы Дики –Фуллера рассчитаны для уровней значимости в 1, 5, 10 %.



Рисунок 3.2 – Результаты теста Дикки-Фуллера

В таблице приведены расчетное значение статистики Дикки-Фуллераравно -2,8144и Prob. 0.077, а также критические значения этой статистики на 1, 5 и 10%-м уровнях значимости.

Результаты расширенного теста Дикки-Фуллера представлены на рисунке 3.3. Расчетное значение статистики Дикки-Фуллераравно -3,703 и Prob. 0.0125.



Рисунок 3.3 – Результаты расширенного теста Дикки-Фуллера

Поскольку расчетное значение получилось больше критического, на уровне значимости 1%, можно сделать предположение что ряд не стационарен. Но в силу того, что Prob. принимаетзначение близкое к 0, нельзя дать точное заключение о стационарности временного ряда.

Проведем расширенный тест Дикки-Фуллера для ряда первых разностей. Результаты представлены на рисунке 3.4.



Рисунок 3.4 – Результаты расширенного теста Дикки-Фуллера по ряду первых разностей

По результатам теста Prob. близко к 0, следовательно ряд первых разностей стационарен.

1. Тест KPSS (Кватковского- Филлипса-Шмидта-Шина). Результаты теста представлены на рисунках 3.5 и 3.6.



Рисунок 3.5 – Результаты KPSS-теста

На рисунке 3.5 представлены результаты расчетного значения статистики KPSS, равный 0,7978. А также критические значения этой статистики на 1, 5 и 10%-м уровнях значимости 0.739, 0.463, 0.347 соответственно.

Поскольку расчетное значение статистики KPSS получилось больше критического, причем на всех уровнях значимости, можно сделать вывод о том, что наш ряд не стационарен.



Рисунок 3.6 – Результаты KPSS-теста для ряда первых разностей

Однако для ряда первых разностей расчетное значение статистики KPSS получилось меньше критического на всех уровнях значимости, то можно смело говорить о том, что наш временной ряд стационарен.

1. Анализ коррелограммы. Также, чтобы проверить временной ряд на стационарность, можно построить коррелограмму, см. рисунок 1.1.

Значения в первом лаге, т.е ACF(1) близки к единице, а затем коррелограмма медленно убывает по угасающей экспоненте(синусоиде). Также значения в первом лаге ACF(1)=PACF(1) близко к единице, однако остальные значения коэффициентов корреляции статистически незначbмы, т.е значения функций выходят за пределы доверительного интервала

Таким образом можно сделать вывод, что временной ряд не стационарен, а ряд первых разностей стационарен.

## 3.2. Построение модели авторегрессии и проинтегрированного скользящего среднего ARIMA(p,k,q)

Модель авторегрессии и проинтегрированного скользящего среднего ARIMA(p,k,q) предназначена для описания ВР, имеющих порядок интеграции k, в виде ARMA(p,q) модели ряда разностей k-го порядка первоначального ВР:

*,*

где C, *µj* и – коэффициенты модели,*p* и *q* – порядки модели, – случайные величины, образующие «белый шум».

Рассмотренные далее в подпунктах 3.2.1. – 3.2.3. линейные модели можно понимать как частные случаи ARIMA(p,q) при k=0.

Поскольку временной ряд, исследуемый в данной курсовой является нестационарным, т.е. имеет порядок интеграции k=1>0 (как установлено в пункте 3.1), то в каждом из подпунктов 3.2.1. – 3.2.3. построим соответствующие модели для ряда разностей k-го порядка, являющегося стационарным.

### Построение авторегрессионной модели AR(p)

При построении аддитивной модели объясняющими переменными являются функции момента времени τ. В нашем случае (для линейной модели) объясняющие переменные - лаговые переменные *yi-1, yi-2, …, yi-p*.

Модель AR(p): (где коэффициенты βj (j = 0…p) оцениваются с помощью пакета EViews, p – порядок модели)

Возмущения εiудовлетворяют условиям Гаусса-Маркова:

P1.

P2.

P3. .

При построении модели нужно решить следующие задачи:

1. определение порядка *p* авторегрессионной модели временного ряда;

2. оценивание коэффициентов βi.

Рассмотрим решение первой задачи.

*Ρчаст(m)* – частный коэффициент автокорреляции, соответствующий лагу *m*. Величина *Ρчаст(m)* определяется как коэффициент корреляции между двумя случайными величинами Y(),…,Y() случайных величин.

Оценкой для *Ρчаст(m)* является выборочный частный коэффициент автокорреляции *rчаст(m)*.

Предположим, что все вычисленные частные коэффициенты автокорреляции значимы до порядка *p* включительно. Тогда можно принять порядок автокорреляции модели равным *p*.

В нашем случае однозначно определить порядок p по коррелограмме не удается, поэтому будем определять порядок p подбором.

Рассмотрим решение второй задачи.

Наиболее часто используются следующие модели AR(p):

1. авторегрессионная модель первого порядка (или модель AR(1)):

2. авторегрессионная модель второго порядка (или модель AR(2)):

Коэффициентыβj (j = 0…p) будем оценивать с помощью пакета E’Views. Для этого необходимо в среде EViews в контексном меню «Quick» выбрать пункт «Equation…». Далее в появившемся окне в поле «Equationspecification» вписать «Y c ar(1)» где Y *-* имя той таблицы, в которой хранятся значения массива временного ряда. Результат этих действий приведен на рисунке 3.7.

**

Рисунок 3.7 – Нахождение коэффициентов модели AR(1)

Таким образом, построенная модель AR(1) имеет следующий вид:

### Построение модели скользящего среднего MA(q)

Модель MA(q) имеет вид:

где C, – коэффициенты модели, q – порядок модели, – случайные величины, образующие «белый шум» (порядок найдем перебором).

Коэффициенты будем считать в пакете EViews.

Для этого необходимо в среде EViews в контексном меню «Quick» выбратьпункт «Estimate Equation…». Далее в появившемся окне в поле «Equationspecification» вписать «Y c ma(1)», где Y *-* имя той таблицы, в которой хранятся значения массива временного ряда. Далее нажать «ОК». Результат этих действий приведен на рисунке 3.8.



Рисунок 3.8 – Нахождение коэффициентов MA(1) модели

Таким образом, построенная модель MA(1) имеет следующий вид:

–754243,3 +

### Построение смешанной модели ARMA(p,q)

Для достижения большей гибкости при построении модели ВР полезно включать в нее и авторегрессионные члены, и члены скользящего среднего.

,

где C, *µj* и – коэффициенты модели,*p* и *q* – порядки модели, – случайные величины, образующие «белый шум».

Коэффициенты будем считать в пакете EViews.

Для этого необходимо в среде EViews в контексном меню «Quick» выбратьпункт «Estimate Equation…». Далее в появившемся окне в поле «Equationspecification» вписать «Y c ar(1) ma(1)» где Y *-* имя той таблицы, в которой хранятся значения массива временного ряда. Далее нажать «ОК». Результат этих действий приведен на рисунке 3.9.



Рисунок 3.9 - Построение смешанной модели ARMA(1,1)

Таким образом, построенная модель ARMA(1,1) имеет следующий вид:

### Сравнительныйанализ построенных моделей

Для выбора наиболее подходящей ARIMA(p,k,q) модели мы посчитаем в пакете EViews показатели адекватности для каждой модели, приведённые далее. Во всех формулах этого пункта приняты следующие обозначения:

*n* – число фактических значений;

*m*– число коэффициентов в модели;

*yi* – фактическое наблюдение в i-тый момент;

 - расчётное наблюдение в i-тый момент.

1. Приведённый индекс детерминации (AdjustedR-squared) рассчитывается следующим образом:

где - индекс детерминации, введённый в пункте 2.3.

Данный коэффициент позволяет сравнить модели с разным числом коэффициентов. Он учитывает число коэффициентов модели, вводя штраф за дополнительные регрессоры, которые не способствуют увеличению объясняющей силы регрессии. При включении в регрессию дополнительных переменных коэффициент может уменьшаться и может быть отрицательным, если модель плохо специфицирована.

2. Стандартная ошибка регрессии (S.E. ofregression) рассчитывается следующим образом:



Чем меньше значение стандартной ошибки, тем адекватнее модель.

3. F-статистика (F-statistic) рассчитывается следующим образом:

******

Уравнение нелинейной регрессии значимо с уровнем значимости α, если

******

При помощи F-статистики в предположении, что остатки модели распределены нормально, проверяется гипотеза о значимости регрессии в целом, т.е. о том, что коэффициенты при всех объясняющих переменных, включенных в модель, значимо отличаются от нуля.

4. P-значение для F-статистики (Prob(F-Statistic)).

Если значение меньше*,* чемвыбранный уровень значимости, то гипотезу о том, что все коэффициенты модели равны нулю, можно отвергнуть на этом уровне значимости.

Заметим, что регрессия может быть значимой, даже если каждый коэффициент в отдельности не значим.

5. Информационный критерий АкаикеAIC (Akaikeinfocriterion) рассчитывается следующим образом:



Используется для выбора лучшей модели из некоторого набора альтернативных моделей: чем меньше значение критерия, тем лучше модель.

6. Информационный критерий Шварца BIC или SC (Schwarzcriterion) рассчитывается следующим образом:



Используется для выбора лучшей модели из некоторого набора альтернативных моделей – чем меньше значение критерия, тем лучше модель. Всегда выбирает лучшую модель с числом параметров, не превышающим число параметров в модели, которая была выбрана по критерию Акаике.

7. Hanna-QuinnCriterion



Используется для выбора лучшей модели из некоторого набора альтернативных моделей – чем меньше значение критерия, тем лучше модель.

В таблице 3.1 приведены расчётные значения показателей адекватности моделей AR(2), MA(3), ARMA(2,1), найденные с помощью пакета EViews.

Таблица 3.1 – Показатели адекватности моделей AR(2), MA(3), ARMA(2,1)

|  |  |  |  |
| --- | --- | --- | --- |
| Критерий | Модель | | |
| AR(1) | MA(1) | ARMA(1,1) |
| Приведённый индекс детерминации | 0,9137 | 0,6986 | 0,94399 |
| Стандартная ошибка регрессии | 94231,83 | 176067,9 | 78095,48 |
| F-статистика | 95,2368 | 20,8577 | 95,5083 |
| P-значение для F-статистики | 0 | 0,000021 | 0 |
| Информационный критерий Акаике | 26,0242 | 27,2739 | 25,7239 |
| Критерий Шварца | 26,1734 | 27,4231 | 25,9229 |
| Hanna-Quinn | 26,0565 | 27,3063 | 25,7671 |

По приведенному индексу детерминации лучше всего данные описывает модель ARMA(1,1), она описывает 94% всех данных, когда модель MA(1) описывает только 69%. Стандартная ошибка регрессии меньше всего также у модели ARMA(1,1), а модель MA(1) является худшей по этому критерию, имею ошибку регрессии 176067,9 что почти в два раза больше, чем у модели ARMA(1,1). Все значения F-статистик больше критических которые для моделей AR(1), MA(1), ARMA(1,1) равны соответственно 3,55, 3,1196, 3,197, т.е. все уравнения регрессии значимы с уровнем значимости 0,05. P-значение для F-статистики меньше, чем уровень значимости следовательно гипотезу о том, что все коэффициенты модели равны нулю, можно отвергнуть на этом уровне значимости. По критерию Акаике, критерию Шварца и Hanna-Quinn самой подходящей является модель ARMA(1,1).

Таким образом наиболее подходящей моделью для описания данного нестационарного временного ряда является модель ARMA(1,1).

# КОМБИНИРОВАННАЯ МОДЕЛЬ ВРЕМЕННОГО РЯДА

Комбинированную модель можно рассматривать как аддитивную модель временного ряда

где сезонная и периодическая составляющие отсутствуют, уравнение тренда известно, а случайная составляющая представлена моделью AR(p) вида .

Здесь – коэффициенты модели, p – порядок модели, – случайные величины.

Общий вид комбинированной модели временного ряда, имеющего порядок интеграции 1:

где – расчетное значение в момент времени ,– трендовая составляющая в момент времени ,p – порядок авторегрессионой модели случайной составляющей , – коэффициенты авторегрессионой модели случайной составляющей ,– фактические наблюдения.

Алгоритм построения комбинированной модели.

1. Провести визуальный анализ диаграммы рассеивания и убедиться в наличии тренда.

2. Построить уравнение тренда t(τ).

3. Вычислить остатки

4. Проверить гипотезу о равенстве нулю математического ожидания ряда остатков c помощью критерия

*,*

где *; .*

Если вычисленное значение непопадает в критическую область

*,*

где

то с уровнем значимости α можно считать, что математическое ожидание ряда остатков нулевое.

Для расчета можно использовать встроенную функцию Excel РАСПСТЬЮДОБР.

5. Вычислить коэффициенты авторегрессионной модели AR(p) ряда остатков с помощью пакета Eviews.

6. С учетом центрирования , построить комбинированную модель:

Рассмотрим применение алгоритма для построения комбинированной модели анализируемого временного ряда.

1. В результате визуального анализа диаграммы рассеяния, представленной на рисунке 2.1, можно сделать вывод о наличии трендовой составляющей. Кроме того, в пункте 2.1 было установлено наличие не случайной составляющей.

* 1. В пункте 2.3 было была построена логарифмическая модель трендовой составляющей

найдены значения ряда остатков и с помощью критерия показано, что ряд остатков имеет нулевое математическое ожидание.

5. Коэффициенты авторегрессионной модели AR(1) ряда остатков были найдены в пункте 3.2.1.

; С.

6. С учетом найденных коэффициентов, применим центрирование и построим комбинированную модель:

# 5.ТОЧЕЧНЫЙ ПРОГНОЗ ПО ПОСТРОЕННЫМ МОДЕЛЯМ

Точечный прогноз заключается в получении прогнозного значения на основе построенной модели временного ряда.

## 5.1. Точечный прогноз по аддитивной модели

Точечный прогноз по аддитивной модели заключается в получении прогнозного значения путем подстановки в уравнение модели момента времени .

В разделе 2.5 была построена аддитивная модель вида:

где значения сезонной составляющей приведены в таблице 2.5.

Построим точечный прогноз, для следующего момента времени, используя аддитивную модель. Поскольку последнее фактическое значение временного ряда, полученное для 21-го момента времени, найдено в декабре 2017 года, то построим прогноз для 22-го элемента, января 2018:

Таким образом согласно аддитивной модели в январе 2018 года, общая сумма выплат пострадавшим от несчастного случая на Южно-Уральской железной дороге составит 590016,3 рублей.

## Точечный прогноз по линейной модели

Точечный прогноз по линейной модели заключается в получении прогнозного значения путем подстановки в уравнение модели необходимых предыдущих фактических значений временного ряда.

В разделе 1.3. были построены несколько линейных моделей, среди которых лучшей была признана модель ARMA (1,1). Построим прогнозное значение для момента времени , следующего за моментом времени последнего фактического наблюдения .

Общая формула для нахождения имеет вид:

,

где ­­и - коэффициенты модели; и - фактические значения временного ряда; -расчетное значение для момента времени τn, которое находится по формуле:

.

Модель ARMA(1,1) имеет вид:

Таким образом согласно аддитивной модели в январе 2018 года, общая сумма выплат пострадавшим от несчастного случая на Южно-Уральской железной дороге составит 810427,8 рублей.

## Точечный прогноз по комбинированной модели

Точечный прогноз по аддитивной модели заключается в получении прогнозного значения путем подстановки в уравнение модели момента времени , а также необходимых предыдущих фактических значений временного ряда и моментов времени.

В главе 4 разработана комбинированная модель вида:

Построим точечный прогноз, для следующего момента времени, используя комбинированную модель. Поскольку последнее фактическое значение временного ряда получено для 24-го момента времени, то построим прогноз для 25-го момента времени, полагая что

получим:

Таким образом согласно аддитивной модели в январе 2018 года, общая сумма выплат пострадавшим от несчастного случая на Южно-Уральской железной дороге составит 944149,3.

## Сравнительный анализ построенных прогнозов

Исходя из полученных значений приведённого индекса детерминации и F-статистики (см. таблицу 5.1), можно сделать вывод о том, что наиболее адекватной моделью является аддитивная. Таким образом, наиболее вероятное прогнозное значение 590016,3.

Таблица 5.1 – Сравнительный анализ построенных прогнозов

|  |  |  |  |
| --- | --- | --- | --- |
| Характеристики | Модель | | |
| Аддитивная | Линейная | Комбинированная |
| Приведённый индекс детерминации R2adj | 0,9994 | 0,94399 | 0,72307 |
| F-статистика | 9994 | 95,5083 | 24,80428754 |
| Прогноз для момента времени, следующего за моментом последнего фактического наблюдения, yi+1 | 590016,3 | 810427,8 | 944149,25 |

# ЗАКЛЮЧЕНИЕ

# ПРИЛОЖЕНИЕ А.

Исходный ряд данных

|  |  |  |
| --- | --- | --- |
|  | Дата | Y |
|
| 1 | апр.16 | 70000 |
| 2 | май.16 | 4354000 |
| 3 | июн.16 | 120000 |
| 4 | июл.16 | 1760000 |
| 5 | авг.16 | 2220000 |
| 6 | сен.16 | 3160000 |
| 7 | окт.16 | 1275000 |
| 8 | ноя.16 | 890000 |
| 9 | дек.16 | 1890000 |
| 10 | янв.17 | 940000 |
| 11 | фев.17 | 555000 |
| 12 | мар.17 | 1175000 |
| 13 | апр.17 | 1714000 |
| 14 | май.17 | 1385000 |
| 15 | июн.17 | 57394,9 |
| 16 | июл.17 | 780000 |
| 17 | авг.17 | 490000 |
| 18 | сен.17 | 1300000 |
| 19 | окт.17 | 1089000 |
| 20 | ноя.17 | 1240000 |
| 21 | дек.17 | 415000 |

# БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. 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.
2. Э.Е. Тихонов. Методы прогнозирования в условиях рынка: учебное пособие. - Невинномысск, 2006. - 221 с.