**Министерство науки и высшего образования Российской Федерации**

федеральное государственное автономное образовательное учреждение   
высшего образования

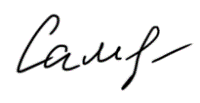
«Национальный исследовательский Томский политехнический университет»

|  |  |
| --- | --- |
| Школа / филиал | ИЯТШ ТПУ |
| Обеспечивающее подразделение | ОЭФ |
| Направление подготовки / специальность | 01.03.02  Прикладная математика и информатика |
| Образовательная программа (направленность (профиль)) | Математические и программные средства исследования операций в экономике;  Математические средства эконофизики |

**КУРСОВАЯ РАБОТА**

**«ИССЛЕДОВАНИЕ ДИНАМИКИ КУРСА GBR/RUB НА МБ МЕТОДАМИ РЕГРЕССИОННОГО АНАЛИЗА»**

|  |  |
| --- | --- |
| Выполнил обучающийся | Саматов Денис Сергеевич |
| Группа | 0В01 |



(подпись обучающегося)

**Согласовано:**

Дата проверки    26.05.22

Итоговая оценка\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(традиционная оценка, балл)

         Доцент ОЭФ ТПУ       \_\_\_\_\_\_\_\_\_\_     /Шинкеев М.Л./

Томск – 2022 г.

**Оглавление**

[Задание 3](#_Toc104452849)

[Теоретическая часть 4](#_Toc104452850)

[Регрессионный анализ 4](#_Toc104452851)

[Виды регрессионного анализа 5](#_Toc104452852)

[Оценка коэффициентов уравнения регрессии 6](#_Toc104452853)

[Анализ регрессионных остатков 8](#_Toc104452854)

[Практическая часть 10](#_Toc104452855)

[Сбор данных 10](#_Toc104452856)

[Анализ полученной регрессии 10](#_Toc104452857)

[Исследование остатков на независимость при помощи критерия серий 12](#_Toc104452858)

[Исследование остатков на нормальность 12](#_Toc104452859)

[Исследование остатков на гомоскедастичность 12](#_Toc104452860)

[Прогноз для средней цены закрытия акций на 4 недели вперед 13](#_Toc104452861)

[Вывод 15](#_Toc104452862)

[Список литературы 16](#_Toc104452863)

# **Задание**

**Вариант 18.**

**Цель работы:** Исследование динамики курса GBR/RUB на МБ методами регрессионного анализа.

**Задание:**

1. Используя открытые источники (сайт finam.ru), провести сбор исходных статистических данных для регрессионного анализа: получить данные о курсе GBR/RUB\_TOD на МБ на момент закрытия (close) за период с 01.01.21 по 30.11.21 с периодичностью одна неделя (на сайте finam.ru на странице «Котировки» выбираем нужную позицию, переходим в «Экспорт котировок» и заказываем нужные данные).
2. Построить подходящую регрессионную модель вида , характеризующую зависимость средней курсовой стоимости y от времени t (, ). Коэффициенты , нелинейно входящие в модель, определить используя, либо спектральный анализ, либо один из численных методов, минимизируя сумму квадратов остатков.
3. Оценить коэффициент детерминации, значимость модели по критерию Фишера, остаточную дисперсию, значимость коэффициентов модели и (в предположении, что остатки независимые нормальные случайные величины с одинаковой дисперсией, а коэффициенты известны и равны найденным оценкам).
4. Исследовать свойства остатков модели (проверить гипотезы о независимости, гомоскедастичности, нормальности), сделать выводы о адекватности модели.
5. Получить прогноз для среднего значения курса GBR/RUB\_TOD на 4 недели вперед, а также построить доверительные границы для прогноза (курса GBR/RUB\_TOD) и сравнить результат с фактическими значениями курса за данный период.

# **Теоретическая часть**

## **Регрессионный анализ**

Регрессионный анализ – это статистический метод исследования функциональной связи математического ожидания величины от переменных , , рассматриваемых в регрессионном анализе как неслучайные величины, независимо от истинного распределения этих величин. Регрессионный анализ включает в себя: определение (подбор) класса функций, связывающих результативный показатель (среднее значение величины и переменные ); оценку неизвестных параметров уравнения связи; анализ значимости и точности полученного уравнения.

Регрессия – зависимость математического ожидания случайной величины от одной или нескольких других случайных величин (свободных переменных), то есть регрессионным анализом называется поиск такой функции, которая описывает эту зависимость. Регрессия может быть представлена в виде суммы неслучайной и случайной составляющих:

|  |  |
| --- | --- |
| , | (1) |

где – функция регрессионной зависимости, а  – случайная величина с нулевым математическим ожиданием.

Обычно в регрессионном анализе исходят их следующей модели, устанавливающей связь между входными переменными и выходной переменной :

|  |  |
| --- | --- |
|  | (2) |

где – функция регрессии величины на величины , – случайная величина с математическим ожиданием равным нулю и дисперсией , в общем случае зависящей от набора значений величин , уравнение называют уравнением регрессии. Величину называют остаточной дисперсией или дисперсией предсказания.

В большинстве случаев полагают, что функция относится к некоторому параметрическому семейству функций , либо аппроксимируется функциями данного семейства. Выбор семейства функций делают либо, исходя из экспериментальных данных, либо из теоретических соображений. Как правило, аппроксимирующие функции выбирают в виде линейной комбинации некоторого набора линейно независимых базисных функций :

|  |  |
| --- | --- |
| , | (3) |

где функции предполагаются заданными. Поскольку данное уравнение является линейным относительно неизвестного вектора параметров , то в этом случае говорят о модели регрессии, линейной по параметрам. Задача регрессионного анализа состоит в том, чтобы на основе выборочных данных найти вектор оценок коэффициентов уравнения регрессии, исходя из заданного критерия качества аппроксимации. Заметим, что уравнение (3) путем введения новых переменных сводится к, так называемому, уравнению множественной линейно регрессии:

|  |  |
| --- | --- |
|  | (4) |

В матричной форме уравнение регрессии записывается в виде:

|  |  |
| --- | --- |
| , | (5) |

где  – вектор ошибок, – вектор значений переменной , – вектор параметров модели, а – это матрица значений переменных размерами . Первый столбец содержит единицы (значения фиксированной переменной ), остальные столбцы – значения переменных .

## **Виды регрессионного анализа**

Существует несколько видов регрессий, среди которых различают линейные и нелинейные регрессии. К линейной относится регрессия вида:

|  |  |
| --- | --- |
|  | (6) |

В свою очередь нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, нелинейные по объясняющим переменным: полиномы разных степеней, например:

|  |  |
| --- | --- |
|  | (7) |

равносторонняя гипербола:

|  |  |
| --- | --- |
|  | (8) |

Регрессии, нелинейные по оцениваемым параметрам: степенная:

|  |  |
| --- | --- |
|  | (9) |

показательная:

|  |  |
| --- | --- |
| , | (10) |

где во всех видах регрессии – возмущение, или стохастическая переменная, включающая влияние неучтенных факторов в модели.

## **Оценка коэффициентов уравнения регрессии**

Рассмотрим линейную по коэффициентам модель регрессии величины на :

|  |  |
| --- | --- |
| , | (11) |

где: – функция регрессии на (функции предполагаются заданными), – случайная величина с математическим ожиданием равным нулю и дисперсией .

Если ввести вектор базисных функций: и вектор коэффициентов уравнения регрессии: , то уравнение (11) можно переписать в виде:

|  |  |
| --- | --- |
|  | (12) |

Пусть в результате эксперимента получена выборка из пар значений величин и . Требуется на основе данных значений найти оценки неизвестных параметров уравнения регрессии (11). Для нахождения оценок параметров используем метод наименьших квадратов, согласно которому в качестве оценок берутся такие, которые минимизируют сумму квадратов отклонений наблюдаемых значений от значений, определяемых функции регрессии, то есть, из условия минимума функции:

|  |  |
| --- | --- |
|  | (13) |

Оценка вектора коэффициентов уравнения регрессии, найденная из условия минимума функции (13) имеет вид:

|  |  |
| --- | --- |
| , | (14) |

где:

|  |  |
| --- | --- |
| , | (15) |

Матрицы и можно найти также по формулам:

|  |  |
| --- | --- |
| , , | (16) |

где:

|  |  |
| --- | --- |
| , . | (17) |

Остаточная дисперсия модели регрессии (несмещенная оценка дисперсии ) находится по формуле:

|  |  |
| --- | --- |
| , | (18) |

где – оценка функции регрессии.

Качество регрессионной модели можно оценить по значению коэффициента детерминации:

|  |  |
| --- | --- |
|  | (19) |

Чем ближе значения к 1, тем большую долю дисперсии величины объясняет функция регрессии. Однако, коэффициент детерминации просто характеризует качество аппроксимации и не учитывает точность полученных оценок дисперсий регрессионной модели. Поэтому более корректно использование исправленного коэффициента детерминации, рассчитываемого по формуле:

|  |  |
| --- | --- |
|  | (20) |

Для проверки значимости коэффициента детерминации используется статистика Фишера:

|  |  |
| --- | --- |
|  | (21) |

При условии истинности , в предположении, что остатки независимые нормальные случайные величины с одинаковой дисперсией, данная статистика имеет распределение Фишера с и числом степеней свободы. Если наблюденное значение статистики , где – критическая точка распределения Фишера уровня (квантиль распределения Фишера уровня ), то нулевая гипотеза отклоняется и считается, что значимо отличается от нуля и использование модели (11) статистически обосновано. В противном случае коэффициент детерминации и, соответственно, уравнение признаются незначимыми.

Для проверки значимости коэффициентов уравнения регрессии используется статистика: , где - соответствующий диагональный элемент матрицы . Статистика при истинности гипотезы , в предположении, что остатки независимые нормальные случайные величины с одинаковой дисперсией, имеет распределение Стьюдента с степенями свободы. Следовательно, если , где – квантиль распределения Стьюдента уровня с числом степеней свободы (двухсторонняя критическая точка уровня ), то нулевая гипотеза отвергается и коэффициент на уровне значимости признается значимым. В противном случае следует признать, что коэффициент незначимо отличается от нуля, а значит и соответствующее слагаемое из уравнения регрессии следует исключить. Можно не задавать изначально уровень значимости и, соответственно, не вычислять , а вычислить наблюдаемый уровень значимости , где – случайная величина, имеющая распределение Стьюдента с степенями свободы, и, на основе полученного значения, принять или отклонить гипотезу. Как правило, при анализе значимости коэффициентов регрессии, значимость коэффициента обычно не проверяют.

Доверительный интервал для значений в точке уровня в предположении, что остатки независимые нормальные случайные величины с одинаковой дисперсией имеет вид:

|  |  |
| --- | --- |
| , , | (22) |

где - квантиль распределения Стьюдента уровня с числом степеней свободы .

## **Анализ регрессионных остатков**

Для получения информации об адекватности построенной модели [регрессии](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BC%D0%B5%D1%80%D0%BD%D0%B0%D1%8F_%D0%BB%D0%B8%D0%BD%D0%B5%D0%B9%D0%BD%D0%B0%D1%8F_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F) исследуют [регрессионные остатки](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BC%D0%B5%D1%80%D0%BD%D0%B0%D1%8F_%D0%BB%D0%B8%D0%BD%D0%B5%D0%B9%D0%BD%D0%B0%D1%8F_%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F). Если выбранная регрессионная модель хорошо описывает истинную зависимость, то остатки должны быть [независимыми](http://www.machinelearning.ru/wiki/index.php?title=%D0%92%D1%8B%D0%B1%D0%BE%D1%80%D0%BA%D0%B0) [нормально распределенными](http://www.machinelearning.ru/wiki/index.php?title=%D0%9D%D0%BE%D1%80%D0%BC%D0%B0%D0%BB%D1%8C%D0%BD%D0%BE%D0%B5_%D1%80%D0%B0%D1%81%D0%BF%D1%80%D0%B5%D0%B4%D0%B5%D0%BB%D0%B5%D0%BD%D0%B8%D0%B5) [случайными величинами](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BC%D0%B5%D1%80%D0%BD%D0%B0%D1%8F_%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0) с нулевым [средним](http://www.machinelearning.ru/wiki/index.php?title=%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BC%D0%B5%D1%80%D0%BD%D0%B0%D1%8F_%D1%81%D0%BB%D1%83%D1%87%D0%B0%D0%B9%D0%BD%D0%B0%D1%8F_%D0%B2%D0%B5%D0%BB%D0%B8%D1%87%D0%B8%D0%BD%D0%B0). Анализ регрессионных остатков – это процесс проверки выполнения этих условий.

Для проверки остатков на независимость будет использоваться критерий серий, основанный на медиане.

Идея этого метода заключается в том, что при его использовании для проверки утверждения о присутствии во временном ряду трендовой компоненты, временной ряд объёмом N ранжируется, т. е. все наблюдения упорядочиваются по возрастанию, и рассчитывается медиана ранжированного ряда. Медианой при этом называется наблюдение, которое делит ранжированный временной ряд на две равные части.

Если временной ряд содержит нечётное количество наблюдений, то в качестве медианы принимается значение, стоящее в середине данного ряда.

Если временной ряд содержит чётное количество наблюдений, то в качестве медианы берётся среднее арифметическое значение двух наблюдений, находящихся посередине временного ряда.

После этого уровни исходного временного ряда сравниваются с медианой по следующему принципу:

1. если уровень временного ряда больше медианы, то ему приписывается знак «+»;
2. если уровень временного ряда меньше медианы, то ему приписывается знак «-».

Основная гипотеза формулируется как утверждение об отсутствии трендовой компоненты во временном ряду.

Если хотя бы одно из следующих неравенств не выполняется, то основная гипотеза об отсутствии тренда в изучаемом временем ряду отклоняется:

|  |  |
| --- | --- |
| , | (23) |

где – квантиль нормального распределения уровня , γ – общее количество серий данного временного ряда, φ – длинна самой большой серии из плюсов или минусов [5].

При этом числа и необходимо округлить вниз до ближайшего целого значения.

Для проверки нормальности распределения остатков воспользуемся критерием Жарка-Бера.

Идея метода состоит в том, что для совокупности остатков оценивается «скошенность» (асимметрия) и «вытянутость» (эксцесс) фактического распределения ряда остатков и сравнивается с нормальным. При этом за оценку «скошенности» распределения отвечает коэффициент асимметрии S , а за оценку «вытянутость» распределения – коэффициент эксцесса K.

Алгоритм критерия Жарка-Бера следующий:

1. Выдвигается гипотеза  о нормальном распределении остатков выборки.
2. Вычисляется наблюдаемое значение критерия по формуле:

|  |  |
| --- | --- |
| , | (24) |

где N – количество наблюдений.

1. По таблице критических точек распределения Пирсона (число степеней свободы 2) определяется критическое значение (на уровне значимости 0,05 оно равняется 5,991).
2. Если  , то гипотеза  о нормальном распределении остатков отклоняется, т.е. распределение остатков не является нормальным. Если  , то гипотеза  о нормальном распределении остатков принимается [6].

Для проверки остатков на гомоскедастичность воспользуемся критерием Фишера.

О наличии или отсутствии гетероскедастичности остатков судят по величине F-критерия Фишера для регрессии остатков. Если фактическое значение F-критерия выше табличного, то, следовательно, существует четкая корреляционная связь дисперсии ошибок от значений факторов, включенных в регрессию, и имеет место гетероскедастичность остатков. В противном случае ( <) делается вывод об отсутствии гeтероскедастичности остатков регрессии. Фактическое значение F-критерия находится по формуле:

|  |  |
| --- | --- |
| , | (25) |

где – это несмещенная выборочная дисперсия первой выборки, – это несмещенная выборочная дисперсия второй выборки

Практическая часть

Сбор данных

Используя открытые источники (сайт finam.ru), проведен сбор исходных статистических данных для регрессионного анализа: получены данные о курсе GBR/RUB\_TOD на МБ за период с 01.01.21 по 30.11.21 с периодичностью одна неделя. Выборка размером . Выборка представляет собой массив целых чисел от 1 до 47 включительно, то есть, номер каждой недели в году. Графически экспериментальные данные выглядят следующим образом.

Рисунок 1. Исходные статистические данные.

Анализ полученной регрессии

Для построения и анализа была выбрана модель: , где и . Частоты для регрессионной модели подбирались итерационным методом. Исходя из соотношения , были найдены частоты:

С помощью полученных данных были произведены оценки коэффициентов для модели регрессии.

Таблица 1. Коэффициенты модели регрессии.

|  |  |
| --- | --- |
| Коэффициенты | Значения |
|  | 455,9212 |
|  | 302,9652 |
|  | -6,2359 |
|  | 240,8966 |
|  | -4550,8164 |
|  | -590,4288 |
|  | -59,2667 |
|  | -9,6945 |
|  | 69,9110 |

Полученное уравнение имеет вид:

Ниже представлен график смоделированной кривой на фоне экспериментальных данных. Как видно, модель достаточно хорошо согласуется с изначальными данными.

Рисунок 2. Регрессия.

Рассмотрим результаты регрессионной статистики. Коэффициент детерминации, . Полученная модель в 89 % случаев объясняет разброс исходных данных. Другими словами, модель регрессии дает хороший результат. Этот результат подтверждает ее значимость по критерию Фишера: . Это число значительно меньше, чем . Следовательно, модель является высокозначимой. Остаточная дисперсия: .

Значения коэффициентов регрессии были рассчитаны по данным выборки. Необходимо убедиться, что рассчитанные коэффициенты будут статистически значимы (т.е. отличны от нуля для значительной части выборок из рассматриваемой генеральной совокупности) и войдут в модель. Для оценки статистической значимости коэффициента регрессии выдвигается нулевая гипотеза о равенстве коэффициентов регрессии нулю. Ниже приведена таблица значимости коэффициентов модели:

Таблица 2. Значимость коэффициентов модели регрессии.

|  |  |  |
| --- | --- | --- |
| Коэффициенты | P-значение | Значимость коэф. |
|  |  | Значим |
|  |  | Значим |
|  |  | Значим |
|  |  | Значим |
|  |  | Значим |
|  |  | Значим |
|  |  | Значим |
|  |  | Значим |

Исследование остатков на независимость при помощи критерия серий

Проведя необходимые вычисления для системы неравенств, получим следующий результат:

Округляя правые значения вниз до ближайшего целого значения, заметим, что гипотеза об отсутствии тренда принимается.

Исследование остатков на нормальность

Используем критерий Жарка-Бера, вычислим по формуле (24) значение . Получим:

|  |  |
| --- | --- |
|  |  |
|  |  |

Таким образом, , следовательно гипотеза о нормальном распределении остатков отвергается, т.е. распределение не является нормальной.

Исследование остатков на гомоскедастичность

Для исследования однородности остатков (гомоскедастичности) построим диаграмму рассеяния остатков от значений переменной :

Рисунок 3. Диаграмма рассеяния остатков.

Из приведенной диаграммы можно сделать вывод о достаточной однородности разброса остатков относительно значений переменной , то есть можно считать, что остатки гомоскедастичны.

Для более точной оценки остатков на скедастичность воспользуемся критерием Фишера.

Для этого разделим остатки на две одинаковые группы, и вычислим выборочную дисперсию каждого из них, и с помощью этих значений найдем значение статистики критерия Фишера. Ниже представлены все полученные данные.

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Так как , следовательно, принимается гипотеза о гомоскедастичности остатков.

Прогноз для средней цены закрытия акций на 4 недели вперед

Сравним значения полученного прогноза средних цен значения курса GBR/RUB\_TOD на МБ на 4 недели вперед с фактическими значениями курса:

Таблица 3. Сравнение данных.

|  |  |
| --- | --- |
| Фактическое значение | Предсказанное значение |
| 96,8175 | 93,0535 |
| 98 | 86,4191 |
| 98,5 | 75,6972 |
| 100,66 | 59,5195 |

Найдем границы доверительных интервалов уровня 0,95 для линии регрессионной модели(в предположении, что остатки независимые нормальные случайные величины с одинаковой дисперсией). Для этого вычисляем оценку дисперсии оценки значения функции регрессии в произвольной точке :

|  |  |
| --- | --- |
|  | (26) |

Определяем величину доверительного отклонения для вероятности в точке : , где – квантиль распределения Стьюдента уровня с числом степеней свободы . Соответственно, получим границы доверительного интервала в точке :

|  |  |
| --- | --- |
|  | (27) |

Результаты приведены в таблице ниже:

Таблица 4. Полученное значение для построения доверительных интервалов.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Неделя | 48 | 49 | 50 | 51 |
|  | 6,2566 | 11,0747 | 18,1691 | 28,1691 |

Графически полученные результаты выглядят так:

Рисунок 4. Графики фактических и предсказанных значений.

Вывод

В ходе выполнения работы, исследована динамика курса GBR/RUB\_TOD на МБ методами регрессионного анализа:

* Построена подходящая регрессионная модель вида, характеризующая зависимость средней курсовой стоимости y от времени t.
* Произведена оценка коэффициента детерминации, значимости модели по критерию Фишера, остаточная дисперсия, значимости коэффициентов модели.
* Подробно исследованы свойства остатков модели (проверены гипотезы о независимости, гомоскедастичности, нормальности).

Построенная модель хорошо объясняет примерно 89 процентов экспериментальных данных и является высокозначимой. При этом было обнаружено, что данная модель не точно делает прогноз. Полученные предсказанные значения, для средних значений курса GBR/RUB\_TOD на 4 недели вперёд, отличаются от фактических значений курса за данный период. Можно сделать вывод о нерациональности использования данной модели для предсказания курса валют.

Список литературы

1. Айвазян, Сергей Артемьевич. Прикладная статистика Основы эконометрики: учебник: в 2 т.. — 2-е изд., испр.. — М.: ЮНИТИ-ДАНА, 2001-Т. 1: Теория вероятностей и прикладная статистика. — 2001. — 656 с.: ил.. — Библиогр.: с. 642-643. — Алф.-предм. указ.: с. 644-656.. — ISBN 5-238-00304-8.
2. Дубров А.М. Многомерные статистические методы : учебник для экономистов и менеджеров / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. — Москва: Финансы и статистика, 1998. — 350 с.. — ISBN 5279019453.
3. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. — Москва: Физматлит, 2006. — 816 с. — ISBN 5-9221-0707-0.
4. Многомерный статистический анализ в экономике : учебное пособие / Л. А. Сошникова [и др.]; под ред. В. Н. Тамашевича. — Москва: ЮНИТИ, 1999. — 598 с.. — ISBN 5238000995.
5. [Прикладная статистика: Основы моделирования и первичная обработка данных](https://scask.ru/q_book_stat1.php) [Электронный ресурс] : инф. ресурс // [сайт]. – Ред. 2022 г. – URL: https://scask.ru/q\_book\_stat1.php?id=152, свободный. Загл. с экрана (дата обращения: 10.05.2022).
6. Статистика Жака-Бера [Электронный ресурс] : инф. ресурс // [сайт]. – Ред. 2022 г. – URL: https://help.fsight.ru/ru/mergedProjects/lib/05\_statistics/uimodelling\_jarqueberastat.htm, свободный. Загл. с экрана (дата обращения: 11.05.2022).