МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

МОДЕЛИ ВРЕМЕННЫХ РЯДОВ В АНАЛИЗЕ РЫНКА АКЦИЙ

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

Студентки 2 курса 248 группы
направления 09.04.03 — Прикладная информатика

механико-математического факультета
Дранищевой Анны Лериевны

Научный руководитель
к. ф.-м. н. ______________ Н. Ю. Агафонова

Заведующий кафедрой
д. ф.-м. н., доцент ______________ С. П. Сидоров

Саратов 2019

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Актуальность темы. В условиях постоянно нарастающих объемов информации и повышения сложности технических, социальных, экономических и прочих систем, человеку, задействованному в управлении этими системами, становится все труднее оценить всю поступающую в органы управления и непосредственно к лицу, принимающему решения, информацию, а также оценить влияние множества различных факторов на поведение системы. Это мешает ему оперативно и безошибочно выбирать оптимальное решение.

Для достижения наилучших результатов, как в качественном, так и в количественном отношении и при этом минимизации рисков серьезных убытков и аварий, в современной науке активно развивается направление систем поддержки принятия решений. Перед данными системами ставится задача помочь людям, принимающим решение, в условиях большого количества данных и сложных реакций управляемой системы на проявление внешних факторов. Детально и качественно анализируя предметную область, системы поддержки принятия решений должны оперативно предоставлять пользователю полную и объективную информацию о наблюдаемой или управляемой системе, спрогнозированные показатели, варианты оптимальных или необходимых решений для разных условий.

Целью магистерской работы является исследование вопросов анализ рынка ценных бумаг, в частности, цен на ГМК "Норильский никель выделение закономерностей изменения, построение адекватных моделей с целью прогнозирования.

Предметом исследования магистерской работы являются временные ряды, как математическая модель экономических процессов. В качестве статистического материала для обработки данных методами анализа временных рядов был выбран курс акций горно-металлургической компании "Норильский никель"(ГМК "Норильский никель").

Практическая значимость заключается в том, что содержащиеся в работе положения и выводы могут быть использованы при дальнейшем более глубоком исследовании поведения цен. Теоретические и методологические положения, представленные в работе, значительно повышают возможности и качество анализа и прогнозирования временных рядов, уточняют особенности применения методов, обеспечивают глубокое понимание сущности происходящих процессов.

Структура и содержание магистерской работы. Работа состоит из введения, 3 разделов, заключения, списка использованных источников и приложения. Общий объем работы составляет 60 страниц.

Работа прошла апробацию на ежегодной студенческой конференции "Актуальные проблемы математики и механики которую проводил механико-математический факультет СГУ в апреле 2019 года, в секции "Анализ данных в VII Международной молодежной научно-практической конференции «Математическое и компьютерное моделирование в экономике, страховании и управлении рисками», ноябрь 2018 года.

1 Основное содержание работы

Регрессионная модель объединяет широкий класс универсальных функций, которые описывают некоторую закономерность. При этом для построения модели в основном используются измеряемые данные, а не знание свойств исследуемой закономерности. Такая модель часто неинтерпретируема, но более точна. Это объясняется либо большим числом моделей-претендентов, которые используются для построения оптимальной модели, либо большой сложностью модели. Нахождение параметров регрессионной модели называется обучением модели.

На сегодняшний день существует много универсальных программ обработки и анализа данных, представленных временными рядами: Stata, SPSS, EViews, Statistica. При выполнении работы будет использоваться пакет прикладных программ (ППП) Eviews (EconometricViews 5), так как в нем реализованы его возможности решения задач прогнозирования количественных показателей, представляющих собой именно временной ряд. В пакете Eviews имеется достаточно полный набор методов по обнаружению и борьбе с типичными для анализа временных рядов проблемами:

— гетероскедастичность,

— автокорреляция,

— нестационарность

— наличие коинтеграции,

Кроме того, EViews позволяет строить прогноз сразу же после построения модели. В теории временных рядов под авторегрессионной моделью порядка p обычно понимается модель вида:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + K + \alpha_p y_{t-p} + \varepsilon_t \tag{1}$$

где $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ - "белый шум последовательность независимых одинаково распределенных случайных величин со средним 0 и дисперсией σ_ε^2.

Рассмотрим интегрируемый порядка d нестационарный процесс X_t. Если при этом стационарный процесс $Y_t = \Delta^d X_t$, составленный из разностей d-порядка, является ARMA(p,q), тогда X_t называется процессом ARMA(p,q).

Большинство эмпирических временных рядов является реализацией процесса ARIMA (p,d,q).

В конце моделирования произведем разынтеграцию.

$$y_t - 2 * y_{t-1} + y_{t-2} = \alpha_0 + \alpha_1 * (y_{t-1} - 2 * y_{t-2} + y_{t-3}) + \varepsilon_t \tag{2}$$

$$y_t = \alpha_0 + \alpha_1 * (y_{t-1} - 2 * y_{t-2} + y_{t-3}) + \varepsilon_t + 2 * y_{t-1} - y_{t-2} \tag{3}$$

$$y_t = \alpha_0 + (2 + \alpha_1) * y_{t-1} - (2 * \alpha_1 + 1) * y_{t-2} + \alpha_1 * y_{t-3} + \varepsilon_t \tag{4}$$

Для прогнозирования ARIMA-процессов применяются два подхода.

Получение прогнозных значений $\hat{Y_T}(h)$ ARIMA-процесса $Y_t = \Delta^d X_t$, по методике прогнозирования авторегрессионных моделей скользящего среднего с последующим вычислением прогнозных значений (до получения необходимого $\hat{X_T}(h)$). Например, для процесса ARMA(1,1) формулы прогнозирования имеют вид:

$$\hat{Y}_{T(+1)} = \alpha_0 + \alpha_1 Y_T - \beta_1 \varepsilon_{t-1} \tag{5}$$

$$\hat{Y}_{T(+h)} = \alpha_0 + \alpha_1 \hat{Y}_{T(+h-1)}, \text{при}\, h \geq 2 \tag{6}$$

При прогнозировании на практике реальные параметры ARMA(p,q) α_i и β_j заменяют их оценками $\hat{\alpha}_i$, $\hat{e}$, $\hat{\beta}_j$, а случайные воздействия ε_t заменяют на остатки $\hat{\varepsilon}_t$, полученные при оценивании модели.

Построение прогнозной формулы с помощью модификации уравнения (4) путем подстановки разностей, вместо Y_t и последующего разрешения полученного уравнения относительно X_t. В результате будет получена ARIMA (p,d,q) нестационарного процесса.

В модели случайного блуждания ARIMA(0,1,0) формула экстраполяции имеет вид:

$$X_{T+h} = X_{T+h-1} + \varepsilon_t \tag{7}$$

В то время, как формула для прогноза имеет вид:

$$\hat{X}_T\,(h) = \; X_T\,, \,\text{для} h \geq 1 \varepsilon_t \tag{8}$$

Дисперсия ошибки прогноза $ar\,(e_T\,(h)) = h\sigma_\varepsilon^2$ увеличивается с ростом h. Ширина доверительного интервала прогноза возрастает пропорционально $\sqrt{h}$.

Конечной целью моделирования является оценка и прогнозирование цены акции. Чтобы учесть влияние случайных факторов, помимо точечного прогноза в эконометрике строится также интервальный прогноз. В нем отклонение от закономерности в результате случайных событий определяется границами доверительных интервалов, которые находятся по следующей формуле:

$$y_{t+i} = \hat{y}_{t+i} \pm \sigma_{\varepsilon} * 1,96, \varepsilon_t \tag{9}$$

где i-номер нашего прогнозного значения.

Анализ временных рядов - одна из основных частей прикладной статистики. Под временным рядом понимается последовательность значений некоторого протекающего во времени процесса. Примерами временных рядов могут быть финансовые индексы, ежедневные курсы валют, котировки акций, годовые объемы продаж, квартальные объемы производства, деловая активность и т.д., то есть переменные, значения которых изменяются со временем. Обычно элементы временного ряда (члены, уровни ряда) нумеруют в соответствии с номером момента времени, к которому они относятся, и порядок уровня ряда играет важную роль в дальнейшем исследовании. Уровни временного ряда получают, как правило, или в результате непосредственного изменения ординат исследуемого процесса через определенные промежутки времени, или в процессе осреднения за определенный период времени (например, средняя цена продаж за день, средняя прибыль за год)

Если измерения производятся непрерывно, то говорят о временных рядах с непрерывным временем - случайных процессах. Если текущая переменная имеет временной характер, то говорят о случайных функциях (например, давление газа в газопроводе в зависимости от расстояния до точки, в которой оно измеряется). Случайные функции нескольки переменных - случайные поля. При одновременной регистрации нескольких характеристик исследуемого процесса рассматриваются многомерные временные ряды.

Основные цели анализа временного ряда можно сформулировать следующим образом:

— описание характерных особенностей ряда
— выяснение механизма, порождающего временной ряд
— подбор статистической модели, описывающей временной ряд
— прогноз будущих значений ряда на основе прошлых наблюдений

Практическая реализация перечисленных целей выполняется в результате следующих этапов:

1) графическое представление временного ряда;

2) выделение и удаление: тренда, сезонных и циклических составляющих (выравнивание, сглаживание);

3) выделение и удаление низко- или высокочастотных составляющих процесса (фильтрация);

4) исследование случайной составляющей временного ряда, оставшейся после сглаживания и фильтрации;

5) построение математической модели для описания случайной составляющей и проверка ее адекватности;

6) прогнозирование будущего развития процесса, описанного временным рядом;

7) исследование взаимосвязей между различными временными рядами.

Для решения перечисленных выше задач используются следующие методы и модели:

•методы корреляционного анализа, позволяющие выявить периодические зависимости и их лаги внутри одного процесса (автокорреляция) или между несколькими процессами (кросскорреляция);

•методы спектрального анализа, нацеленные на выявлениепериодических и квазипериодических составляющих временного ряда;

•методы преобразования временных рядов (сглаживание и фильтрация) с целью удаления из них высокочастотных или сезонныхсоставляющих;

•модели авторегрессии и скользящего среднего (для описанияпрогнозирования случайной составляющей временного ряда).

Сглаживание (выравнивание) уровней временного ряда выполняетсяпри помощи специально подобранных функций (тренда), описывающихзакономерности развития во времени исследуемых экономическихявлений. Выбор той или иной функции в качестве тренда является наиболееважным этапом анализа временного ряда, так как ошибка на данном этапе приводит к очень серьезным последствиям, особенно при прогнозировании уровней ряда. Основным инструментом, позволяющим отдать предпочтение той или иной модели, являются приросты уровней ряда, а также некоторые их преобразования.

В качестве простейших моделей тренда в анализе временных рядовиспользуются полиномы, экспоненты и логистические кривые.Приведем пример некоторых из них.

1. Полиномиальный тренд

Полиномиальный тренд имеет вид:

$$Y_t = b_0 + b_1 \times t + b_2 \times t^2 + \ldots + b_n \times t^n, \tag{10}$$

где t - независимая переменная (время), $b_i,\ i = 1, \ldots, n$ - параметры модели, которым, при небольших значениях i, можно дать конкретную интерпретацию. Например, b_0 - уровень ряда в начальный момент времени $(i = 0)$, b_1 - скорость роста, b_2 - ускорение роста, b_3 - изменение ускорения.

Тренд, описываемый полиномом первой степени, имеет вид:

$$Y_t = b_0 + b_1 \times t \tag{11}$$

Характерной особенностью данного тренда является то, что приросты ординат (разности первого порядка) остаются постоянными:

$$u_1 = Y_1 - Y_{t-1} = b_0 + b_1 \times t - b_0 - b_1 (t-1) = b_1 = const \tag{12}$$

поэтому данная модель может быть использована для описания тенденции временного ряда, в котором уровни со временем или равномерно возрастают, или равномерно убывают.

2. Простейшая модель экспоненциального тренда имеет вид:

$$Y_t = a \times b^1, \tag{13}$$

где $a,\ b$ - параметры модели. Модель применяется для описания уровней ряда с постоянными темпами роста и прироста("лавинообразные"процессы), на которые не оказывают значительного воздействия различного рода ограничения. Характерной особенностью модели является то, что приросты уровней ряда зависят от величины самой функции.

4.Кривая Гомперца

Уравнение кривой имеет вид:

$$Y_t = k \times a^{b^1} \tag{14}$$

Математическое ожидание, оно же среднее значение выборки, рассчитывается, т.е. используется формула:

$$\bar{y} = \frac{1}{n} \sum\nolimits_{t=1}^{n} y_t \tag{15}$$

Среднеквадратическое отклонение (СКО) характеризует рассеивание значений независимой случайной величины относительно математического ожидания. Для расчетов используется формула (16).

$$\sigma = \sqrt{\frac{1}{n} \sum\nolimits_{t=1}^{n} (y_t - \bar{y})^2} \tag{16}$$

Коэффициент вариации характеризует относительную меру отклонения значений от математического ожидания. Чем больше значение коэффициента, тем больше разброс значений, а следовательно и меньше однородность данных. Если коэффициент вариации превышает 33%, то дальнейшие расчеты будут не точными. Чтобы избавиться от неоднородности данных принято выкидывать минимальное и максимальное значения ряда. Оптимальным признается интервал от 10% до 20%. Тогда изменчивость ряда признается незначительной. Рассчитывается коэффициент вариации по формуле (17).

$$V = \frac{\sigma}{\bar{y}} \cdot 100 \tag{17}$$

Коэффициент асимметрии - характеризует отклонение распределения от симметричного распределения, характерного для нормальной кривой. Асимметрия положительна, если мода находится слева от математического ожидания, и отрицательна, если справа. При симметричном расположении коэффициент асимметрии равен нулю.

Сглаживание временных рядов - это распространенный прием выявления тенденций развития рассматриваемого ряда. Суть различных приемов сводится к замене фактических значений расчетными, но с меньшим уровнем колебаний. В данной магистерской работе используются метод скользящей средней и метод экспоненциального сглаживания.

Метод скользящей средней основан на взаимном погашении случайных отклонений в средних величинах. При сглаживании временного ряда таким способом участвуют все уровни ряда. Чем шире берется интервал сглаживания, тем более плавным получается выделенный тренд. Однако следует учитывать, что при больших интервалах сглаживания сокращается и количество наблюдений, что создает неточности в расчетах.

Для начала, чтобы построить ряд скользящих средних следует выбрать интервал сглаживания m, как правило, берутся нечетные значения (3,5,7 и т.д). Для первых $m < n$ уровней рассчитывается среднее арифметическое. Это и является сглаженным значением для выбранного уровня. Эта процедура повторяется до $(n-m+1)$ -го уровня. В итоге получаем ряд скользящих средних порядка m.

Метод экспоненциального сглаживания заключается в сглаживании временных рядов, при котором обрабатываются все предыдущие наблюдения. При этом информация учитывается по мере удаления от прогнозного периода, т.е. чем старше наблюдение, тем меньше оно влияет на прогноз. Рабочая формула данного метода выглядит следующим образом:

$$S_t = a \cdot y_t + (1-a) \cdot S_{t-1}, \tag{18}$$

где $a \in (0;1)$

Здесь y_t - значение исходного временного ряда в момент времени t, S_t - значение в сглаженном ряду в момент t, a - коэффициент сглаживания. От величины a зависит скорость снижения влияния предыдущих наблюдений. Точного метода выбора коэффициента нет, поэтому в работе рассматриваются несколько значений от 0 до 1. При значениях близких к 0 влияние прошлых значений достаточно велико и наоборот, при значениях близких к 1, на прогноз влияют лишь последние наблюдения.

Данный метод хорош тем, что дает возможность получить оценку параметров тренда, характеризующих не усредненный процесс, а тенденцию, сформировавшуюся к моменту последних наблюдений.

Для проверки постоянства дисперсии ряда применяется тест Сиджела-Тьюки. Он заключается в выдвижении гипотезы о постоянстве дисперсии. Все значения индексируются, а затем делятся на две части. Далее значения ряда сортируются по возрастанию, и каждому из них присваивается ранг. Причем минимальному значения присваивается 1, а максимальному - 2, второму по минимальности - 3, а по максимальности - 4 и т.д. Далее ряд приводят в изначальный вид, сохраняя за каждым значением его ранг. Затем подсчитывается сумма рангов первой совокупности R_1 и расчетное значение z по формуле (19).

$$z = \frac{R_1 - \frac{1}{n} n_1 (n_1 + n_2 + 1)}{\sqrt{\frac{1}{12} * n_1 * n_2 * (n_1 + n_2)}} \tag{19}$$

Если расчетное значение $z > |1,96|$, тогда предположение о постоянстве дисперсии изучаемого временного ряда отвергается.

Еще одним тестом на постоянство дисперсии является тест Вальда-Вальфовица.Рассчитывается медиана[1] ряда. Если значение $y_t > Me$, тогда ему присваивается ранг 1, если же меньше, то 0. При условии, что y_t численно равно медиане, значение не учитывается в дальнейших расчетах. После ранжирования всего ряда находятся число последовательностей одного и того же ранга (серий) и длина максимальной серии, υ и δ соответственно. Критические значения рассчитываются по формулам.

$$\nu_{\text{крит}} = \frac{1}{2} * \left(n + 2 - 1 * \sqrt{n - 1}\right) \tag{20}$$

$$\delta_{\text{крит}} = 1,43 * \ln (n + 1)) \tag{21}$$

Расширенный тест Дики-Фуллера применяется для проверки гипотезы о наличии авторегрессии более чем первого порядка во временном ряду. Он основан на оценке параметра $\delta = \alpha_1 - 1$, уравнения $\Delta y_t = \alpha_1 y_{t-1} + \varepsilon_t$ Выдвигаются следующие гипотезы:

[1]статистика, делящая ранжированную совокупность на две равные части: 50 % "нижних"членов ряда данных будут иметь значение не больше медианного, а "верхние"50 % —не меньше, чем медиана.(источник: http://ru.wikipedia.org.,дата обращения 27.11.18)

$$H_0 : \delta = 0$$

$$H_1 : \delta \neq 0$$

Отклонение нулевой гипотезы в пользу альтернативной означает, что $|\alpha_1| < 1$ и временной ряд стационарный или же интегрируемый нулевого порядка.

Проведем анализ, моделирование и прогнозирование на примере курса акций ГМК "Норильский Никель"в краткосрочной, среднесрочной и долгосрочной перспективе.Источником послужил интернет-сайт investfunds.ru[2].

В рамках краткосрочного прогноза проанализируем данные, собранные за 2 года, проведем исследование рядов на стационарность, с последующим моделированием типа ARIMA.

В качестве первоначальных данных используются ежедневные цены на момент закрытия биржи в период с 01.04.2013 по 31.03.2015. Для последующей работы с данными, значения были восстановлены по схеме:

1. Если требовалось восстановить цены для выходных, то субботнему дню присваивалось значение пятницы, а воскресному – значение понедельника;
2. Если было пропущено от трех до пяти ней, то значения восстанавливались с помощью методами интерполяции и среднего арифметического;
3. При отсутствии данных на более длительные сроки восстановление происходило с помощью графика, построенного по двум крайним точкам с учетом СКО за предыдущие дни.

Расчеты показывают, что рассматриваемый временной ряд имеет математическое ожидание равное 6574,35 и среднеквадратическое отклонение 1986,04. Коэффициент вариации меньше 33%, что говорит об относительной однородности данных. Стоит заметить, что коэффициенты асимметрии и эксцесса равны 11,1598 и 0,737144 соответственно. Они не попадают в интервал [-2;2], необходимый для нормального распределения. Из этого следует, что закон распределения для рассматриваемого временного ряда отличен от нор-

[2]Котировки акций "ГМК Норильский Никель"(Тикер:GMKN).URL: http://stocks.investfunds.ru/stocks/251(дата обращения 27.11.2018).

мального. Подтверждением этого служит визуальный анализ ряда.

График плотности распределения смещен влево и имеет плоскую вершину, что подтверждается расчетами. Основным подтверждением того, что закон ряда отличен от нормального, является гистограмма распределения (рисунок **??**). Отчетливо видны две вершины, что говорит о бимодальности распределения. Это значит, что в рассматриваемом ряду существуют два максимума, смесь двух нормальных распределения, что в дальнейшем может осложнить процедуру приведения выборки к стационарному виду. Чтобы избавится от одного из законов распределения, отбросим первые 199 значений и рассмотрим описанные выше характеристики для полученного ряда.

Коэффициент детерминации при использовании метода экспоненциального сглаживания с различными α оказался больше, чем при использовании метода скользящей средней. Наибольший R^2 имеет метод экспоненциального сглаживания при $\alpha = 0,1$, очень близок к 90%. При использовании метода скользящей средней на больших выборка, поэтому мы отказываемся от построения тренда данным способом.

Итак, оптимальным вариантом является экспоненциальное сглаживание с коэффициентом сглаживания равным 0,1. Тогда уравнение линейного тренда выглядит следующим образом:

$$y = 11,404x + 4145,498 \tag{22}$$

Для дальнейшей работы с временным рядом необходимо выполнение условия стационарности. Это означает, что его математическое ожидание и дисперсия должны быть постоянными.

Применим расширенный тест Дики-Фуллера для проверки гипотезы о наличии авторегрессии более чем первого порядка во временном ряду.

Наиболее подходящим для дальнейшего моделирования признан ряд вторых разностей. Исходя, из ЧАКФ и АКФ, делаем вывод о том, что наилучшей моделью будет являться AR(1). Проанализируем оценки коэффициентов модели, полученные в статистическом пакете "StatgraphicsPlus 5.1".

Первый коэффициент является значимым, а также его значение удовлетворяет требуемым ограничениям $|\alpha_1| = 0,0341594 < 1$). Тогда модель имеет следующий вид:

$$y_t = -1,08717 - 0,0341594 y_{t-1} \tag{23}$$

2 Основные результаты

В магистерской работе были применены навыки анализа и обработки большого объема статистической информации, проведено исследование исходных данных на стационарность и наличие трендовой составляющей Также были составлены прогнозы для временных перспектив получены результаты:

— курс акций ГМК "Норильский никель"является нестационарным, как в каждодневных значениях, так и при агрегации данных;
— доверительные интервалы в долгосрочном прогнозе имеют большую длину.
— выполнен анализ курса акций
— проведен процесс сглаживания различными методами (сглаживание по методу скользящей средней,экспоненциальное сглаживание)
— построены модели AR
— выполнено построение прогноза
— качество модели оценено тестами Манна-Уитни, Сиджела-Тьюки, Вальда-Вольфовица