МИНОБРНАУКИ РОССИИ

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

Кафедра теории функций и стохастического анализа

МОДЕЛИРОВАНИЕ ТЕНДЕНЦИЙ И ПРОГНОЗИРОВАНИЕ КУРСОВ ЦЕННЫХ БУМАГ С УЧЕТОМ АВТОКОРРЕЛЯЦИИ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студентки 4 курса 412 группы
направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета
Бодровой Татьяны Анатольевны

Научный руководитель
ст. препод. ______________ Н. В. Сергеева

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

Саратов 2019

ВВЕДЕНИЕ

Актуальность темы. В современных условиях становления российского фондового рынка особую значимость приобретают исследования по моделированию прогноза котировок ценных бумаг. Колебания биржевых индексов, кризис ипотечного кредитования в США и другие потрясения рынка ценных бумаг показывают, что необходимость в данных исследованиях назрела и актуальна. Как в России, так и в ведущих государствах колебания этого рынка все менее зависят от политического влияния и влияния других нерыночных факторов, что подтверждает необходимость проведения объективных исследований в этой области. Научно-методические разработки по данной тематике могут быть полезны как юридическим лицам, так и конкретным гражданам.

Целью бакалаврской работы является изучение теории временных рядов, исследование и программная реализация методов, используемых для идентификация временных рядов и построение моделей временных рядов при прогнозировании курсов акций.

Объектом исследования является рынок ценных бумаг, как один из главных финансовых элементов международной экономической системы.

Предметом исследования являются модели прогнозирования значений котировок ценных бумаг.

Для достижения указанной цели в работе необходимо решить следующие **задачи**:

— определить сущность ценных бумаг, их основные виды и характеристики;
— определить структуры и принципы функциональности рынка ценных бумаг, дать характеристики его составным частям;
— изучение классификаций экономических моделей рынка ценных бумаг;
— рассмотрение моделей пространственных данных и финансовых временных рядов;
— изучение регерессионных моделей доходностей ценных бумаг;
— построение «рыночной модели» на основе метода парной регрессии;
— прогнозирование курса акций с помощью метода авторегрессии.

Структура и содержание бакалаврской работы. Работа состоит из введения, четырех разделов, заключения, списка использованных источников, содержащего 23 наименования, и четырех приложений. Общий объем работы составляет 48 страниц.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы работы, отмечается практическая значимость полученных результатов.

В **первом** разделе определены сущность структуры и принципы функциональности рынка ценных бумаг, их основные виды и характеристики.

Во **втором** разделе рассмотрены модели и методы анализа временных рядов.

Модели временных рядов. Временным рядом обычно называют ряд значений $x_1, x_2, \ldots, x_T$ анализируемой случайной величины, полученных в последовательные моменты времени $t = 1, 2, \ldots, T$.

Самой общей моделью временного ряда $x_1, x_2, \ldots, x_T$ является совместная функция распределения $F(\cdot)$ случайных величин $x_1, x_2, \ldots, x_T$, определяемая соотношением:

$$F(x_1, x_2, \ldots, x_T) = F_1(x_1)F_2(x_2|x_1)F_3(x_3|x_1, x_2) \cdots F_T(x_T|x_1, x_2, \ldots, x_T), \quad (1)$$

где $F_1(x_1)$ - частная (маргинальная) функция распределения случайной величины x_1; $F_t(x_t|x_1, x_2, \ldots, x_{t-1})$ - условная функция распределения случайной величины x_1 при условии, что значения $x_1, x_2, \ldots, x_{t-1}$ являются фиксированными $(t = 1, 2, \ldots, T)$.

Представление (1) учитывает взаимную зависимость и неодинаковую распределенность случайных величин $x_1, x_2, \ldots, x_T$. Однако модель временного ряда $x_1, x_2, \ldots, x_T$, определяемая функцией совместного распределения $F(x_1, x_2, \ldots, x_T)$, как правило, не пригодна для практического применения поскольку функция $F(x_1, x_2, \ldots, x_T)$ обычно неизвестна. Поэтому при описании временных рядов обычно принимаются во внимание лишь первые начальные и вторые центральные моменты совместного распределения $F(x_1, x_2, \ldots, x_T)$, т.е. математические ожидания $E(x_t)$, а также дисперсии $D(x_t)$ и ковариации $cov(x_t, x_\tau)$ $(t \neq \tau, t, \tau = 1, 2, \ldots, T)$.

Ковариация $cov(x_t, x_{t-k})$ между значениями x_t и $x_{t-k}(x_{t+k})$ зависит только от «сдвига по времени», т.е. лага k, и не зависит от времени t:

$$cov(x_t, x_{t-k}) = E((x_t - \mu_x)(x_{t-k} - \mu_x)) \equiv \gamma_k, \quad t = 1, 2, \ldots, T. \tag{2}$$

Автокорреляция с задержкой k равна

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \tag{3}$$

откуда вытекает, что $\rho_0 = 1$.

Статистическая оценка АКФ по реализации $x_1, x_2, \ldots, x_T$ временного ряда называется выборочной автокорреляционной функцией и вычисляется по формуле:

$$\tilde{\rho}_k = \frac{\tilde{\gamma}_k}{\tilde{\gamma}_0}, \quad k = 1, 2, \ldots, T-1, \tag{4}$$

где $\tilde{\gamma}_k \quad (k = 1, 2, \ldots, T-1)$, $\tilde{\gamma}_0$ - статистические оценки ковариационной функции и дисперсии временного ряда, определяемые по формулам:

$$\tilde{\gamma}_k = \sum_{t=k+1}^{T} \frac{(x_t - \overline{x})(x_{t-k} - \overline{x}_{t-k})}{T-k}, \quad \overline{x}_{t-k} = \sum_{t=k+1}^{T} \frac{x_{t-k}}{T-k}, \tag{5}$$

$$\tilde{\gamma}_0 = \tilde{\sigma}_x^2 = \sum_{t=1}^{T} \frac{(x_t - \overline{x})^2}{T}, \quad \overline{x} = \tilde{\mu}_x = \sum_{t=1}^{T} \frac{x_t}{T}. \tag{6}$$

Процесс авторегрессии p-го порядка имеет вид:

$$\widetilde{x}_t = \frac{1}{\varphi(B)}\varepsilon_t = \varphi^{-1}(B)\varepsilon_t. \tag{7}$$

где

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p, \tag{8}$$

ε_t - белый шум.

Наряду с АКФ при установлении свойств модели временного ряда используется так называемая частная автокорреляционная функция (ЧАКФ) и ее статистическая оценка. В отличие от автокорреляционной функции ЧАКФ применяется для измерения степени тесноты статистической связи между значениями x_t и x_{t-k} временного ряда $\{x_t\}$ $(t = 1, 2, \ldots, T)$, отстоящими друг от друга на лаг k, в предположении, что соответствующие промежуточные значения ряда $x_{t-k+1}, \ldots, x_{t-1}$ $(t = 1, 2, \ldots, T-k)$ являются фиксированными. Таким образом ЧАКФ характеризует степень статистической взаимосвязи между x_t и x_{t-k} при устранении опосредованного влияния на эту взаимосвязь со стороны всех промежуточных значений ряда. Временной ряд $\{x_t\}$ $(t = 1, 2, \ldots, T)$, обладающий свойствами (2), (3) называется слабостационарным или стационарным в широком смысле.

Модель авторегрессии является исключительно полезной стохастической моделью для описания некоторых встречающихся на практике рядов. В этой модели текущее значение процесса выражается как конечная линейная совокупность предыдущих значений процесса и импульса ε_t.

Пусть $\tilde{x}_t, \tilde{x}_{t-1}, \tilde{x}_{t-2}, \ldots$ будут отклонениями от μ, например $\tilde{x}_t = x_t - \mu$. Тогда

$$\tilde{x}_t = \varphi_1 \tilde{x}_{t-1} + \varphi_2 \tilde{x}_{t-2} + \ldots + \varphi_p \tilde{x}_{t-p} + \varepsilon_t \tag{9}$$

называется процессом авторегрессии AR порядка p. Такое название объясняется тем, что линейная модель

$$\tilde{x} = \varphi_1 \tilde{x}_1 + \varphi_2 \tilde{x}_2 + \ldots + \varphi_p \tilde{x}_p + \varepsilon,$$

связывающая «зависимое» переменное x с множеством «независимых» переменных $x_1, x_2, \ldots, x_p$, плюс член ε, описывающий ошибку, часто называются моделью регрессии, то есть x «регрессирует» на $x_1, x_2, \ldots, x_p$. В формуле (3) переменная x регрессирует на своих предшествующих значениях, поэтому модель авторегрессирующая. Если определить оператор авторегрессии порядка как

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_p B^p,$$

то модель авторегрессии сжато записывается как

$$\varphi(B)\tilde{x}_t = \varepsilon_t.$$

В **третьем** разделе описана процессы авторегрессии.

Автокорреляционная функция авторегрессии.

При переходе в выражении (9) к математическим ожиданиям получается разностное уравнение

$$\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \ldots + \varphi_p \gamma_{k-p}, \quad k > 0. \tag{10}$$

В этом выражении последнее слагаемое - математическое ожидание $E[\widetilde{x}_{t-k}\varepsilon_t]$ равно нулю при $k > 0$, так как $\widetilde{x}_{t-k}$ может включать лишь импульсы ε_j до момента $t-k$, не коррелированные с ε_t. Деление всех членов формулы (10) на γ_0 приводит к тому, что автокорреляционная функция удовлетворяет аналогичному разностному уравнению

$$\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \ldots + \varphi_p \rho_{k-p}, \quad k > 0. \tag{11}$$

Параметры авторегрессии, выраженные через автокорреляции: уравнения Юла — Уокера

Если подставить в (10) значения $k = 1, 2, \ldots, p$, то получается система линейных уравнений для $\varphi_1, \varphi_2, \ldots, \varphi_p$, со свободными членами $\rho_1, \rho_2, \ldots, \rho_p$:

$$\begin{cases} \rho_1 = \varphi_1 + \varphi_2 \rho_1 + \ldots + \varphi_p \rho_{p-1}, \\ \rho_2 = \varphi_1 \rho_1 + \varphi_2 + \ldots + \varphi_p \rho_{p-2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \rho_p = \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \ldots + \varphi_p. \end{cases} \tag{12}$$

Эти уравнения называются уравнениями Юла — Уокера.

Оценки Юла — Уокера для параметров процесса получаются при замене теоретических значений автокорреляции ρ_k выборочными автокорреляциями r_k.

В **четвертом** разделе построена рыночная модель цен акций ОАО «Нефтяная компания Лукойл» и сделан прогноз курса акций с помощью метода авторегрессии. С сайта www.finam.ru были взяты еженедельные котировки акций указанной компании за период с 01.01.2017 по 01.04.2019.

Были вычислены оценки АКФ и ЧАКФ, которые представлены в таблице 1.

Таблица 1 – Статистические оценки АКФ и ЧАКФ для еженедельных котировок

n	1	2	3	4	5	6	7
ρ_n	0.9917	0.9834	0.9818	0.98152	0.97965	0.97535	0.9696
$\widehat{\varphi}_{nn}$	0.9917	-0.0080	0.4047	0.0877	0.0868	-0.0984	-0.1111
n	8	9	10	11	12	13	14
ρ_n	0.9672	0.9663	0.96043	0.9517	0.9462	0.9419	0.9377
$\widehat{\varphi}_{nn}$	0.1079	0.0074	-0.2431	-0.1417	-0.0077	-0.0862	0.0455

Проверим значимость коэффициентов частной автокорреляции по критерию стандартной ошибки. Коэффициенты частной автокорреляции рассчитывались на основе выборки из n наблюдений. Тогда стандартная ошибка

$$\text{С. О.}\left[\widehat{\varphi}_{nn}\right] = \frac{1}{\sqrt{n}}.$$

Пусть гипотеза $H_0 = \{\widehat{\varphi}_{nn} = 0\}$. В качестве альтернативной гипотезы возьмем гипотезу $H_1 = \{\widehat{\varphi}_{nn} \neq 0\}$. Для проверки гипотезы в общем случае необходимо рассчитать значение t-статистики,но поскольку выборка большая можно использовать Z-статистику (нормальное распределение):

$$Z_n = \frac{\widehat{\varphi}_{nn} - 0}{\text{С. О.}\left[\widehat{\varphi}_{nn}\right]}.$$

Для проверки гипотезы при уровне значимости $\alpha = 0.05$ необходимо сравнить вычисленные значения Z_n с критическим табличным значением статистики.

Если $|Z_n| < |Z_{\text{кр}}|$, принимается основная гипотеза H_0. Иначе она отвергается.

$$\left|\frac{\widehat{\varphi}_{nn} - 0}{\text{С. О.}\left[\widehat{\varphi}_{nn}\right]}\right| < |Z_{\text{кр}}|,$$

где $Z_{\text{кр}}$ – аргумент функции Лапласа при $\Phi\left(Z\right) = \dfrac{1-\alpha}{2}$. В итоге получаем неравенство

$$-1.96 \leq \frac{\widehat{\varphi}_{nn}0}{\text{С. О.}\left[\widehat{\varphi}_{nn}\right]} \leq 1.96.$$

Для выборки $n = 118$, тогда $-0.18 \leq \widehat{\varphi}_{nn} \leq 0.18$. В этом случае первый и третий коэффициент являются значимыми. Поэтому построим модели $AR(1)$ и $AR(3)$ и сравним полученные результаты. Модель $AR(1)$ для еженедельных котировок будет иметь вид

$$x_n^w = 270.05 + 0.9474x_{n-1}^w. \tag{13}$$

Для того, чтобы записать модель $AR(3)$ необходимо найти φ_1, φ_2, φ_3. Решим систему уравнений Юла-Уокера третьего порядка. Получим $\varphi_1 = 1.0029$, $\varphi_2 = -0.4125$, $\varphi_3 = 0.4047$. Проверено выполнение условие стационарности, тогда модель $AR(3)$ примет вид

$$x_n^w = 63.7016 + 1.0029x_{n-1}^w - 0.4125x_{n-2}^w + 0.4047x_{n-3}^w. \tag{14}$$

Спрогнозируем цену акций на несколько периодов вперед, используя полученные модели. Объем выборки $n = 118$. Для прогноза используем модель (13). В результате

$$x_{119}^w = 270.05 + 0.9474x_{118}^w = 270.05 + 0.9474 \cdot 5822 = 5844.18,$$

$$x_{120}^w = 270.05 + 0.9474x_{119}^w = 270.05 + 0.9474 \cdot 5844.18 = 5866.18,$$

$$x_{121}^w = 270.05 + 0.9474x_{120}^w = 270.05 + 0.9474 \cdot 5866.18 = 5887.99.$$

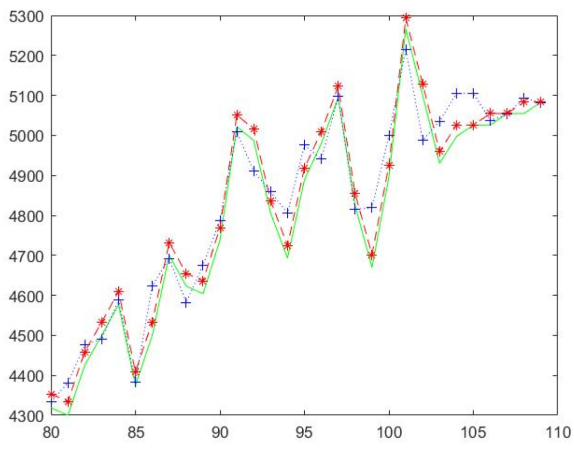

Рисунок 1 – Прогноз на пять периодов вперед по моделям $AR(1)$ и $AR(3)$ для еженедельных котировок

Из графиков рисунка 1 видно, что модели модели $AR(1)$ и $AR(3)$ хорошо описывают временной ряд, поэтому можно использовать модель $AR(1)$ поскольку она более простая.

Следует отметить, что рассмотренные временные ряды трудно анализировать, поскольку они является трендовыми. Об этом говорит очень медленно затухающие автокорреляционные функции. Ряды без тенденции, как правило, не представляют интереса для экономистов. AR-модели вообще не предназначены для описания процессов с тенденцией, однако они хорошо описывают колебания, что весьма важно для отображения развития неустойчивых показателей. Чтобы применить AR-модели к экономическим процессам с тенденцией, на первом этапе формируют стационарный ряд, исключая тенденцию, путем перехода от исходного ряда к ряду разностей соседних значений членов ряда. Первоначальный (исходный) ряд является интегрированным рядом первого порядка, когда его первые разности образуют стационарный ряд динамики. Если для формирования стационарного временного ряда требуется получить ряд вторых разностей, то исходный ряд называется интегрированным рядом второго порядка и т.д.

Таким образом, AR-модели разностных временных рядов характеризуются двумя параметрами: p (порядок авторегрессии) и d (порядок конечных разностей), поэтому записываются как $AR(p, d)$.

Простейшим способом определения наиболее подходящего разностного ряда является вычисление для каждого ряда ($d = 0, 1, 2, ...$) его дисперсии. Для дальнейшей обработки выбирается ряд, у которого величина этого показателя минимальна.

В работе рассмотрен ряд первых разностей $\nabla x_n = x_n - x_{n-1}$ для еженедельных котировок акций компании «ЛУКОЙЛ». Первый порядок разностей был выбран потому что у него минимальная дисперсия.

Для определения параметра p модели были получены статистические оценки автокорреляционной и частной автокорреляционной функций. Для первых разностей была выбрана и построена модель авторегреесии $AR(2)$. На основе этой модели был осуществлен прогноз на 5 периодов вперед.

Для проверки адекватности моделей использовался критерий Дарбина-Уотсона для автокорреляция в остатках.

В моделях $AR(1)$ и $AR(3)$ присутствует положительная автокорреляция. В модели $AR(2)$ корреляция отсутствует, что говорит об адекватности модели. Следовательно, для прогноза нужно выбрать модель $AR(2)$.

Рассмотренные модели временных рядов, характеризуются наличием долговременной зависимости (долгой памяти). Для построения более качественных моделей нужно использовать комбинированные модели авторегрессии $AR(p)$ и скользящего среднего $MA(q)$. Так же предложено использовать модели класса $ARIMA(p, d, q)$ с дробным показателем d.

В работе так же были построены AR - модели для ежедневных котировок той же компании, с помощью которых осуществлялся прогноз.

Построение моделей авторегрессии и получение прогнозных оценок исходного ряда возможно только с использованием программных средств. Для расчета АКФ, ЧАКФ, коэффициентов AR моделей, построения графиков в системе Matlab была написана программа.

Основные результаты

1. Определены сущность ценных бумаг, их основные виды и характеристики, структуры и принципы функциональности рынка ценных бумаг, даны характеристики его составным частям.

2. Изучена классификаций экономических моделей рынка ценных бумаг.

3. Рассмотрены модели авторегрессии и скользящего среднего для моделирования временных рядов.

4. Найдены статистические оценки автокорреляционной функции.

5. Спрогнозирована цена акций на несколько периодов вперед, используя модели $AR(1)$ и $AR(2)$.

6. Найдена AR - модель для временного ряда первых разностей.

7. Для проверки адекватности моделей использован критерий Дарбина - Уотсона для автокорреляции в остатках.