

XIV Международная научно-практическая конференция студентов аспирантов и молодых учёных  
«Молодёжь и современные информационные технологии»

## ОПРЕДЕЛЕНИЕ КАЧЕСТВА МОДЕЛИ ВРЕМЕННОГО РЯДА ЗАВИСИМОСТИ ОТ КОЛИЧЕСТВА ВХОДНЫХ ДАННЫХ

Бозняков Р.В.

Научный руководитель: Семенов М.Е.

Томский политехнический университет

romario1131@bk.ru

### Введение

В экономике часто встречаются явления, которые интересно и важно изучать, так как они изменяются во времени. С течением времени изменяются цены, экономические условия, режим протекания того или иного производственного процесса. Если уметь анализировать и моделировать данные процессы, то можно уменьшить вероятность возникновения неблагоприятных событий, то есть сократить риски. Риск является неотъемлемой частью деятельности людей и компаний, занимающихся выпуском или инвестированием в ценные бумаги, а именно в акции. Из всех ценных бумаг акции обладают наибольшей волатильностью, то есть являются наиболее изменчивыми. Поэтому для сокращения рисков и, соответственно, денежных потерь необходимо уметь анализировать и исследовать процессы изменения цен акций.

Объектом исследования в данной работе были выбраны цены акций ОАО «Газпром». ОАО «Газпром» - глобальная энергетическая компания. Основные направления деятельности - геологоразведка, добыча, транспортировка, хранение, переработка и реализация газа, газового конденсата и нефти, реализация газа в качестве моторного топлива, а также производство и сбыт тепло- и электроэнергии.

Цель данной работы - проведение анализа объекта исследования и построение модели, описывающий исследуемый процесс течение различных интервалов времени для определения влияния количества входных данных на качество построенных моделей. В качестве исследуемых интервалов были взяты периоды времени, равные 4 месяцам (с 01.02.2016 по 01.06.2016), 6 месяцам (с 01.12.2015 по 01.06.2016), 1 году (с 01.06.2015 по 01.06.2016).

Для исследуемых периодов времени были сформированы три временных ряда, содержащих 84, 124, 252 наблюдаемых значений соответственно.

### Проведение анализа

Полученные временные ряды были проверены на нормальность распределения и стационарность. Для этого были построены гистограммы для ряда доходностей акций в течение заданных периодов времени, а также применен критерий Колмогорова - Смирнова. Доходности акций рассчитывались по формуле:

$$X_t = \frac{X_{t+1} - X_t}{X_t}$$

Было получено, что все исследуемые временные ряды подчиняются нормальному закону распределения.

Для проверки на стационарность были исследованы автокорреляционные функции временных рядов. Исследование показало, что наблюдается значительная зависимость (корреляция) между смежными наблюдениями. Также можно отметить очень высокую величину коэффициента корреляции первого порядка для временных рядов и постепенное снижение при изменении лага. Такая сильная зависимость между наблюдениями указывает на наличие тренда в характере поведения временных рядов и дает основание полагать, что временной ряд при всех исследуемых периодах является нестационарным.

### Построение математической модели

Так как временные ряды являются нестационарными, то из класса линейных моделей подходящей моделью будет модель ARIMA (p, d, q), где параметр d=1, так как ряд является интегрируемым первого порядка.

В ходе исследования были определены оптимальные модели для исследуемых рядов. Для временного ряда с периодом 4 месяца оптимальной моделью является модель ARIMA (1, 1, 0). Данная модель имеет следующий вид:

$$X_t = 0,124 + 0,142X_{t-1} + \varepsilon_t$$

Для временного ряда с периодом 6 месяца оптимальной моделью является модель ARIMA (2, 1, 0), имеющая следующий вид:

$$X_t = 0,0597 + 0,96X_{t-1} - 0,87X_{t-2} + \varepsilon_t$$

Для ряда с периодом год была построена модель ARIMA (2, 1, 0):

$$X_t = 0,021 + 0,055X_{t-1} - 0,27X_{t-2} + \varepsilon_t$$

Построенные модели достаточно корректно описывают исследуемый процесс, что подтверждают АКФ и ЧАКФ ряда остатков, а именно малые значения коэффициентов автокорреляции в пределах доверительного интервала.

### Сравнение полученных моделей

Для сравнения качества полученных моделей были выбраны следующие характеристики: САПО – средняя абсолютная процентная ошибка; САО – средняя абсолютная ошибка; СКО – среднее квадратическое отклонение;

ВИС- информационный критерий Байерса;

$R^2$  – коэффициент детерминации.

Используя данные характеристики, было произведено сравнение построенных моделей для каждого из исследуемых временных рядов. Полученные результаты представлены в таблице 1.

Таблица 1 – Характеристики временных рядов

	4 месяца (84 набл. значений)		6 месяцев (124 набл. значений)		1 год (252набл. значений)	
САПО, %	1,30	1,29	-0,01	1,21	-0,08	
САО	1,89	1,86	-0,03	1,73	-0,13	
СКО	6,17	5,96	-0,21	5,11	-0,85	
ВИС	1,87	1,86	-0,01	1,67	-0,19	
$R^2$ , %	97,72	98,4	+1,7	99,63	+1,21	

Как видно из таблицы 1, при увеличении количества входных данных, значение рассчитанных характеристик уменьшается. Это означает, что с увеличением количества входных данных величина отклонения наблюдаемых значений от расчетных (модельных) убывает, следовательно, модель более качественно описывает исследуемый процесс. В данном случае наилучшей является модель, построенная для временного ряда с периодом, равным 1 году.

#### Определение оптимального количества входных данных временного ряда с периодом 1 год

Ранее было определено, что количество входных данных влияет на качество модели. В данном случае временной ряд содержит 252 наблюдаемых значений. Но действительно ли временной ряд должен содержать определенное количество, или возможно сократить количество данных, при том не значительно потеряв в качестве. Для того, чтобы это проверить, было принято решение построить модель ARIMA(2, 1, 0) для данного временного ряда, постепенно убирая из него наблюдения. Исключаться будут те наблюдения, которые вошли во временной ряд в конце периода исследования. Прodelывать это необходимо до тех пор, пока расхождения не станут значительными. Данные полученных расчетов представлены в таблице 2.

Таблица 2 – Результаты расчетов

Количество входных данных	Количество исключенных данных	СКО	САО
252	0	5,17	1,73
251	1	5,19	1,73
250	2	5,14	1,72
245	7	5,20	1,74
242	10	5,17	1,73
239	13	5,12	1,71
236	16	5,16	1,73
229	23	4,96	1,69
226	26	4,99	1,7
216	36	4,45	1,62
206	46	4,5	1,63

Сравнение моделей производилось только по двум характеристикам: САО и СКО. Как видно из таблицы, при исключении данных в количестве от 0 до 16, СКО лежит в интервале от 5,1 до 5,2, изменения происходит в сотых долях. Увеличивая количество исключенных данных до 23, СКО уже становится равным 4,96, САО=1,69. Наилучшие показатели имеет модель, содержащая 216 наблюдаемых значений (-36). Можно сделать вывод, оптимальную модель для временного промежутка можно построить и с меньшим количеством данных, при этом не теряя в качестве модели.

#### Заключение

В данной работе было изучено влияние количества входных данных на качество построенной модели. Для этого были сформированы три временных ряда в течение трех различных временных промежутков: 4 месяца, 6 месяцев, 1 год. Было получено, что количество входных данных действительно влияет на качество модели, а именно на величину отклонения наблюдаемых данных от смоделированных. Также было количественно определено влияние данного фактора. Для временного ряда с периодом, равным 1 году было получено, что можно откинуть часть исходных данных, не теряя при этом в качестве.

#### Список использованных источников

1. Гребенников А.В., Крюков Ю.А., Чернягин Д.В. Моделирование сетевого трафика и прогнозирование с помощью модели ARIMA // Системный анализ в науке и образовании, 2011. – Вып. 1. – [www.sanse.ru/download/79](http://www.sanse.ru/download/79)
2. Бокс Дж., Дженкинс Г.М. Анализ временных рядов, прогноз и управление. М.: Мир, 1974. 406 с.
3. Draper N., Smith H. Applied regression analysis. New York: Wiley, In press, 1981. 693 p.

4. Крейнина М.Н. Финансовый менеджмент:  
учеб. пособие. - М.: "Дело и Сервис", 2001. - 400  
с.;