

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное автономное образовательное  
учреждение высшего образования  
«Южно-Уральский государственный университет  
(национальный исследовательский университет)»  
Институт естественных и точных наук  
Факультет математики, механики и компьютерных технологий  
Кафедра прикладной математики и программирования  
Направление подготовки: 01.03.04 Прикладная математика

РАБОТА ПРОВЕРЕНА

Рецензент,

\_\_\_\_\_ Ю.В. Гасников

« \_\_\_\_ » \_\_\_\_\_ 2019 г.

ДОПУСТИТЬ К ЗАЩИТЕ

Заведующий кафедрой, д.ф.-м.н.,  
доцент

\_\_\_\_\_ А.А. Замышляева  
« \_\_\_\_ » \_\_\_\_\_ 2019 г.

Прогнозирование движения денежных средств в банкоматах с помощью  
авторегрессионной модели ARIMA

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА  
К ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЕ  
ЮУрГУ–01.03.04.2019.71.ПЗ ВКР

Руководитель работы,  
Заведующий кафедрой,  
д.ф.-м.н., доцент

\_\_\_\_\_ /А.А. Замышляева  
« \_\_\_\_ » \_\_\_\_\_ 2019 г.

Автор работы,

Студент группы ЕТ-413

\_\_\_\_\_ /М.Е. Сухоруков  
« \_\_\_\_ » \_\_\_\_\_ 2019 г.

Нормоконтролер, доцент, к.т.н.

\_\_\_\_\_ /Т.Ю. Оленчикова  
« \_\_\_\_ » \_\_\_\_\_ 2019 г.

Челябинск  
2019  
АННОТАЦИЯ

Сухоруков М.Е. Прогнозирование движения денежных средств с помощью авторегрессионной модели ARIMA.– Челябинск: ЮУрГУ, ET-413, 50 с., 12 ил., 1 табл., библиогр. список – 30 наим., 1 прил.

В работе рассмотрена авторегрессионная модель прогнозирования ARIMA. Также в данной выпускной квалификационной работе рассмотрены методы автоматического подбора входных параметров модели.

В первой главе рассмотрены основные теоретические вопросы, связанные с прогнозированием движения денежных средств в банкоматах. Рассмотрены самые популярные используемые в настоящее время модели.

Во второй главе рассмотрены основные математические методы, используемые в работе, построен алгоритм, а также рассмотрены программная реализация.

В третьей главе рассмотрены полученные результаты, а также проведено сравнение полученных результатов с данными, предоставленными компанией «Compass Plus». Рассмотрены возможные методы улучшения модели.

## ОГЛАВЛЕНИЕ

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ .....	7
ВВЕДЕНИЕ.....	8
1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ДВИЖЕНИЯ ДЕНЕЖНЫХ СРЕДСТВ.....	10
1.1 Анализ временного ряда .....	10
1.2 Виды и классификации методов прогнозирования.....	12
1.3 Анализ модели ARIMA .....	19
1.4 Выводы по разделу .....	20
2 МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ARIMA .....	22
2.1 Математическая модель ARIMA.....	22
2.2 Описание алгоритма .....	25
2.3 Условия применения разработанного программного обеспечения и оценка требуемых ресурсов .....	33
2.4 Выводы по разделу .....	35
3 ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ.....	36
3.1 Разработка методологии тестирования программы .....	36
3.2 Сравнительный анализ результатов прогноза .....	37
3.3 Методы повышения точности модели.....	41
3.4 Выводы по разделу .....	43
ЗАКЛЮЧЕНИЕ .....	44
БИБЛИОГРАФИЧЕСКИЙ СПИСОК .....	46
ПРИЛОЖЕНИЕ 1 Текст программы.....	50

## ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ARIMA – Авторегрессионное интегрированное скользящее среднее (autoregressive integrated moving average).

PACF – частичная автокорреляция.

ACF – автокорреляция.

AR – авторегрессионная часть.

MA – скользящее среднее.

I – часть, включающая первую производную временного ряда.

## ВВЕДЕНИЕ

Актуальность. Денежные средства представляют собой важный элемент оборотного капитала любой организации, тем более банков, так как являются одним из основных средств расчета. Для банкоматов, главное – непрерывное движение денежных средств, которое может обеспечить постоянный доступ к обналичиванию денежных средств, находящихся на банковских картах.

В стабильный период работы банковской системы экономический эффект в рассматриваемой области достигается за счет оптимального выбора сумм загрузки и периодов инкассации с учетом стоимости как загружаемых денежных средств, так и себестоимости инкассации. Потому основной задачей является построение верного прогноза движения денежных средств в банкоматах. В данной работе эта задача решается с помощью авторегрессионной модели ARIMA.

Прогнозирующая функция должна быть построена таким образом, чтобы обеспечивать максимальное правдоподобие между прогнозируемыми и фактическими данными. Это позволит составить максимально эффективное расписание загрузки денежных средств в банкоматы, которое позволит избежать таких ситуаций, когда инкассаторская служба загружает деньги в банкомат, который не нуждается в этом, либо напротив, не позволит банкомату простаивать без денег.

Целью настоящей работы является прогнозирование движения денежных средств. Для достижения поставленной цели необходимо решить следующие задачи:

- изучить теоретические основы эконометрики;
- рассмотреть особенности прогнозирования движения денежных средств в банкоматах;
- разработать алгоритм подбора входных параметров модели ARIMA и прогнозирования данной моделью;

– определить адекватность и применимость данной модели к прогнозированию денежных средств в банкоматах.

Объектом исследования в дипломной работе являются банкоматы, предоставленные компанией «Compass Plus». Среди них есть как банкоматы, которые находятся на заводах, предназначенные в первую очередь для выдачи зарплат, так и банкоматы, которые находятся на проходных местах, в центре города, торгово-развлекательных центрах.

Предметом исследования является расписание загрузки денежных средств в банкоматы.

В работе использовались следующие методы исследования:

- анализ теоретических источников по исследуемой проблеме;
- анализ документации языка python и подключаемых библиотек;
- систематизация теоретического материала и документаций;
- обобщение полученного материала.

Выпускная квалификационная работа состоит из введения, трех глав, заключения, списка используемой литературы и приложений. Первая глава раскрывает теоретические аспекты прогнозирования движения денежных средств, анализа временного ряда. Во второй главе разобран алгоритм прогнозирования, рассмотрены используемые методы. В третьей главе проведен анализ полученных результатов, представлены возможные доработки алгоритма.

Информационной базой исследования является документация языка python, документация сторонних библиотек, публикации экономическо-математических журналов, интернет ресурсы.

# 1 ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ ДВИЖЕНИЯ ДЕНЕЖНЫХ СРЕДСТВ

## 1.1 Анализ временного ряда

Прогноз – обоснованное суждение о возможном состоянии объекта в будущем или альтернативных путях и сроках достижения этих состояний.

Прогнозирование – это процесс разработки прогноза.

Этап прогнозирования – часть процесса разработки прогнозов, характеризующаяся своими задачами, методами и результатами. Деление на этапы связано со спецификой построения систематизированного описания объекта прогнозирования, сбора данных, с построением модели, верификацией прогноза.

Приём прогнозирования – одна или несколько математических или логических операций, направленных на получение конкретного результата в процессе разработки прогноза. В качестве приёма могут выступать сглаживание динамического ряда, определение компетентности эксперта, вычисление средневзвешенного значения оценок экспертов и так далее.

Модель прогнозирования – модель объекта прогнозирования, исследование которой позволяет получить информацию о возможных состояниях объекта прогнозирования в будущем и (или) путях и сроках их осуществления.

Метод прогнозирования – способ исследования объекта прогнозирования, направленный на разработку прогноза. Методы прогнозирования являются основанием для методик прогнозирования.

Методика прогнозирования – совокупность специальных правил и приёмов (одного или нескольких методов) разработки прогнозов.

Прогнозирующая система – система методов и средств их реализации, функционирующая в соответствии с основными принципами прогнозирования. Средствами реализации являются экспертная группа,

совокупность программ и так далее. Прогнозирующие системы могут быть автоматизированными и неавтоматизированными [1].

Временный ряд называется упорядоченная во времени совокупность измерений одной из характеристик исследуемого объекта.

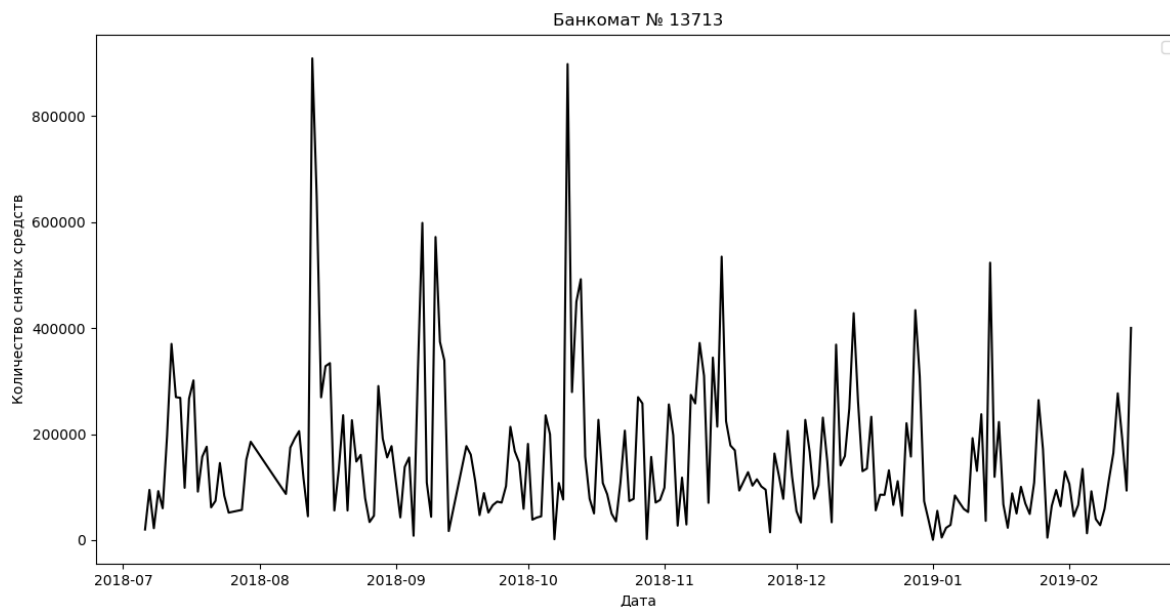


Рисунок 1.1 – Графическое представление ряда

Временные ряды дают возможность прогнозировать будущие значения. Основываясь на предыдущих значениях, временные ряды могут использоваться для прогнозирования тенденций в экономике, погоде и планировании мощностей, и многих других областях.

До построения математической модели проводится анализ имеющегося временного ряда.

Предполагается, что значения ряда состоят из двух компонент:

- систематическая составляющая (регулярная);
- случайный шум.

Аддитивная модель временного ряда:

$$Y_t = T_t + S_t + C_t + e, \quad (1)$$

где тренд ( $T$ ) – плавно изменяющаяся компонента, описывающая чистое влияние долговременных факторов;

сезонная компонента ( $S$ ) – состоит из последовательности почти повторяющихся циклов;



циклическая компонента ( $C$ ) – плавно изменяющаяся компонента, описывающая длительные периоды относительного подъема и спада, состоит из циклов,  $T_t$  относительного подъема и спада, состоит из циклов, меняющихся по амплитуде и порядку роста;

случайная компонента ( $e$ ) – остается после полного вычленения закономерных компонент [2].

В математике и статистике стационарный процесс – процесс, у которого функция среднего постоянна, а ковариационная функция зависит лишь от разности аргументов [3]. Таким образом, стационарный процесс – это процесс без тренда.

Основная сложность, возникающая при прогнозировании данных с месячным периодом, заключается в том, что количество дней в месяце постоянно меняется, потому невозможно выбрать один период, по которому будет строиться функция. Помимо этого – есть сдвиги, основанные на сдвиге дней недели. Не все банкоматы выдают одинаковое количество денег в будни и выходные, из этого следует, что даже при верном выборе месячного периода, данные могут оказаться со сдвигом.

При анализе временного ряда необходимо определить два основных параметра: определение природы ряда и прогнозирование будущих значений. Обе эти цели требуют, чтобы модель ряда была математически описана. Как только модель определена, можно описывать рассматриваемые данные и строить предварительные прогнозы.

## 1.2 Виды и классификации методов прогнозирования

Для прогнозирования движения денежных средств были выбраны:

1) язык python – высокоуровневый язык программирования общего назначения. Python – очень мощный язык программирования, используемый для множества различных задач. Большое сообщество языка с открытым исходным кодом создало немало инструментов для эффективной работы с

Python. В последние годы ряд инструментов был создан специально для анализа данных;

2) среда Visual Studio Code. VS Code – IDE, интегрированная среда разработки (Integrated Development Environment), работающая в любой операционной системе с различными интерпретаторами Python. VS Code имеет инструменты для обработки ошибок, отладки и модульного тестирования, а также позволяет легко переключаться между средами Python, включая виртуальные;

3) библиотека Pandas. Это Python библиотека для анализа и обработки данных. Обладает высокой скоростью обработки данных;

4) библиотека Statsmodels. Данный модуль Python предоставляет классы и функции для оценки множества различных статистических моделей, а также необходим для проведения статистических тестов и исследования статистических данных. Результаты проверяются на соответствие существующим статистическим пакетам, чтобы убедиться в их правильности;

5) библиотека Matplotlib. Matplotlib – это библиотека Python для построения графиков, которая генерирует показатели в различных печатных форматах и интерактивных средах на разных платформах.

На данный момент существует множество методов прогнозирования движения денежных средств и соответственно несколько основных показателей эффективности генерируемых прогнозов денежных потоков. Каждый метод в своей основе имеет один и тот же принцип: в результате реализации проекта необходимо получить верную модель движения денежных средств (необходимо точно спрогнозировать количество снятых денег в банкомате).

Базовой информацией для прогнозирования денежных потоков являются данные, полученные при оценке остатка средств в банкомате. Именно они обуславливают выбор модели прогнозирования. Наиболее

пригодными для составления прогноза величины будущих денежных потоков, являются следующие модели:

- регрессионные модели (линейная регрессия, нелинейная регрессия);
- авторегрессионные модели (ARIMAX, GARCH);
- модель экспоненциального сглаживания;
- модель по выборке максимального подобия;
- нейросетевые модели;
- модели на базе цепей Маркова [4];

Модели авторегрессии наиболее широко применяемые, они отличаются от остальных универсальностью, высокой точностью прогнозов, простотой реализации и невысокими требованиями к вычислительной мощности.

К очевидным преимуществам можно отнести то, что эти модели имеют очень четкое математико-статистическое обоснование, что делает их одними из наиболее научно обоснованных моделей из всего множества моделей прогнозирования тенденций во временных рядах.

Одним из преимуществ модели ARIMA является формализованная и наиболее подробно разработанная методика, следуя которой можно подобрать модель, наиболее подходящую к каждому конкретному временному ряду. Формальная процедура проверки модели на адекватность достаточно проста, а также существуют разработанные методики по автоматическому подбору наилучшей ARIMA [5].

В сравнении с регрессионными моделями, модель ARIMA позволяет дать прогноз более чем на один шаг. Прогноз на один шаг (период) вперед может быть получен исключительно по значениям исходного временного ряда. Помимо этого, регрессионные модели, как и экспоненциальное сглаживание, хорошо вычисляют общий тренд, но они практически полностью игнорируют выбросы данных, что непозволительно, при прогнозировании движения денежных средств в банкоматах.

Регрессионная линейная зависимость имеет вид:

$$Y(t) = C + b_1 Y(t-1) + b_2 Y(t-2) + \dots + b_n Y(t-n) + \varepsilon(t) = \sum_{i=1}^n b_i Y_{t-i} + \varepsilon_i, \quad (2)$$

где  $C$  – константа, которую для простоты часто полагают равной 0;

$n$  – число ретроспективных значений ряда, учитываемых в модели (порядок модели);

$b_i$  – коэффициенты (параметры) модели, которые требуется оценить при её построении;

$\varepsilon(t)$  – случайная составляющая, отражающая вероятностный характер модели [6].

Авторегрессионная модель GARCH [7] же использует авторегрессивный процесс для дисперсии, то есть использует прошлые значения дисперсии для вычисления будущих значений дисперсии этой временной серии. Она объясняет гетероскедастичность модели влиянием предыдущих значений серии и белым шумом. GARCH позволяет кластеризовать данные временного ряда. Зачастую, при прогнозировании экономических данных, используют объединенную модель GARCH-ARIMA [8].

Кроме этого существуют различные модификации GARCH-моделей, такие, как A-GARCH, E-GARCH и другие, применяемые в различных специфических условиях. Например, A-GARCH или асимметричная GARCH-модель используется, когда для периодов подъема и спада на финансовых рынках дисперсии различны.

Проверку на предположении о подобии нужно проверять для каждой новой задачи. Всякая модель временного ряда есть инструмент обращения с этим рядом, и, как всякий инструмент, модель нужно использовать должным образом.

Модель по выборке максимального правдоподобия может показать хорошие результаты в тех случаях, когда в данных нет слишком сильных выбросов.

Модель по выборке максимального подобия не подходит по той причине, что движение денежных средств слишком хаотично, что не всегда позволяет применять данную модель.

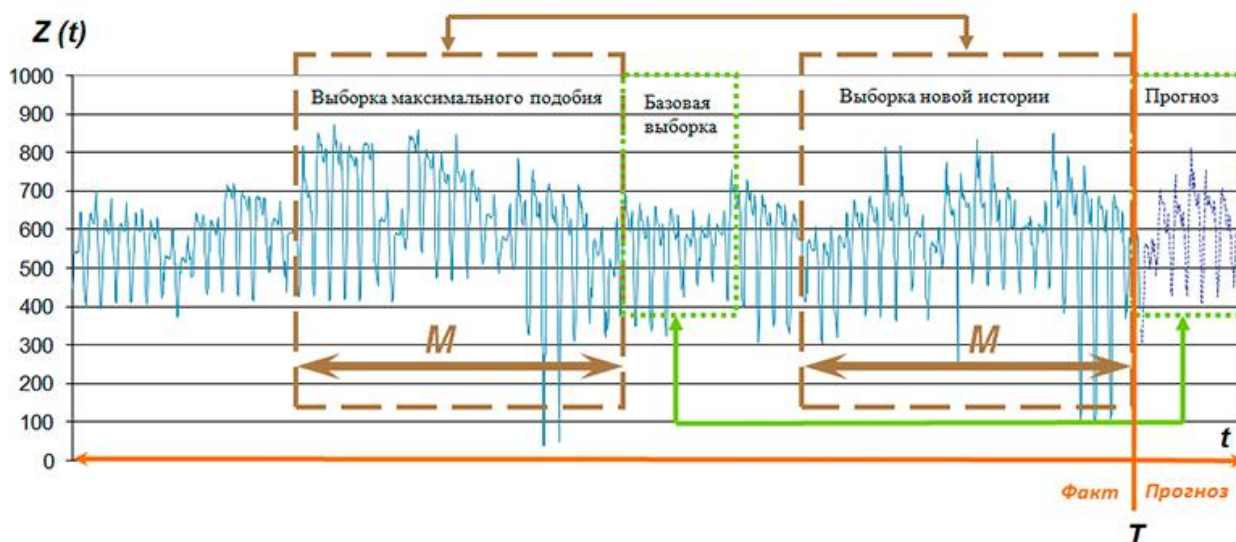


Рисунок 1.2 – Принцип работы модели по выборке максимального правдоподобия

Нейросетевые модели можно рассматривать в виде черного ящика с некоторыми входами и выходами. Ключевой особенностью, привлекающей исследователей, является то, что она умеет обучаться и абстрагировать входные данные. Это одновременно и достоинство, и недостаток. Достоинство заключается в том, что нейросетевую модель можно настроить так, что она будет достаточно точно работать и вычислять сложнопрогнозируемые участки с высокой степенью точности. Недостаток же в том, что невозможно отследить то, по каким параметрам и критериям вычисляются значения, что может привести к серьезным последствиям. С помощью нейронных сетей строится множество прогнозов [9, 10, 11, 12, 13, 14].

Другими словами, нейросеть может научиться отображать любую функцию входных сигналов. Это может быть как простейшая зависимость

(линейная, парабола), так и то, что с помощью математической функции выразить невозможно.

Нейронная сеть может научиться понимать любой принцип и использовать его на данных, которых не было в обучающей выборке.

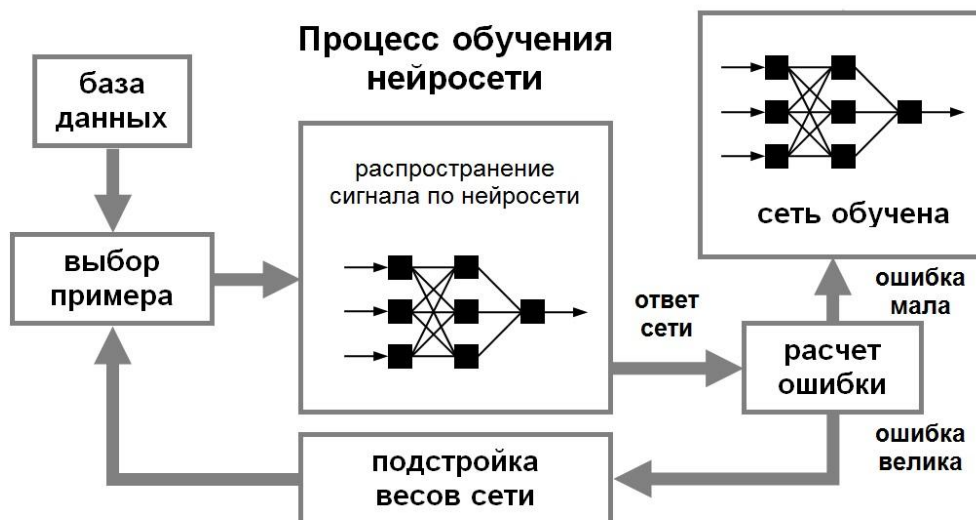


Рисунок 1.3 – Принцип работы нейросети

В моделях прогнозирования на основе цепей Маркова на состояние процесса влияет лишь текущего состояния, и не имеет никакого влияние предыдущее состояние. Данные модели подходят для решения задач только к процессам с короткой памятью [15].

Следовательно, данная модель не применима к прогнозированию в банкоматах, так как тут важны данные, которые были получены задолго до прогнозируемого значения.

Таблица 1.1 – Сравнение моделей прогнозирования

Модель прогнозирования	Достоинства	Недостатки
Регрессионные модели	Низкая сложность алгоритмов; Малые требования к вычислительным мощностям; Доступность промежуточных вычислений; Разнородность решаемых задач.	Возможность моделирования лишь линейных процессов; Сложность определения вида функциональной зависимости;

		Невозможность прогнозирования всплесков данных.
--	--	---

Продолжение таблицы 1.1 – Сравнение моделей прогнозирования

Модель прогнозирования	Достоинства	Недостатки
Авторегрессионные модели	Быстрота получения результата; Доступность промежуточных вычислений; Простота моделей; Разнородность решаемых задач.	Трудоемкость определения параметров.
Модели максимального правдоподобия	Простота моделей; Быстрота получения результата.	Высокая вероятность получения неверных данных.
Модели экспоненциального сглаживания	Простота моделей; Быстрота получения результата; Решение задач долгосрочного прогнозирования.	Отсутствие гибкости.
Нейросетевые модели	Возможность моделирования нелинейных процессов; Адаптивность; Масштабируемость; Разнородность решаемых задач.	Сложность программной реализации; Отсутствие промежуточных вычислений; Высокие требования к непротиворечивости обучающей выборки.
Модели на базе цепей Маркова	Простота моделей.	Узкая применимость моделей; Невозможность решение задач прогнозирования с

		длинной памятью.
--	--	------------------

### 1.3 Анализ модели ARIMA

Модели ARIMA (авторегрессионное интегрированное скользящее среднее) создают мощный класс моделей, которые можно применять ко многим рядам реального времени. Модели ARIMA основаны на трех частях: (AR) авторегрессионная часть, (MA) скользящее среднее и (I) часть, включающая первую производную временного ряда [16].

В статистике и эконометрике, в частности в анализе временных рядов, модель авторегрессионного интегрированного скользящего среднего (ARIMA) является обобщением модели авторегрессионного скользящего среднего (ARMA). Обе эти модели адаптированы к данным временного ряда либо для лучшего понимания поведения данных, либо для прогнозирования будущих точек в ряду (прогнозирование). Модели ARIMA применяются в некоторых случаях, когда данные свидетельствуют о нестационарности, когда начальный шаг дифференцирования (соответствующий «интегрированной» части модели) можно применять один или несколько раз для устранения нестационарности [17].

Методы, которые используются в данной выпускной квалификационной работе, будут принимать только данные из разных временных рядов. Это означает, что рассматриваются только соотношение между значением оси Y и временными точками оси X. Внешние факторы не учитываются, хотя и могут влиять на временной ряд.

Для адекватной работы модели, необходимо правильно определить входные параметры для модели ARIMA, это  $p, d, q$ , где  $p$  – число авторегрессионных лагов,  $d$  – порядок разности временного ряда,  $q$  – количество ошибок в уравнении [18].



Значение  $p$  выбирается с помощью функции частичной автокорреляции PACF. Для этого по тестируемым значениям строится коррелограмма, на которой определяется максимальный номер коэффициента сильно отличный от нуля, возьмем значение равное 0.3, это и будет искомым значением  $p$ .

Значение  $d$  выбирается с помощью обобщенного теста Дикки-Фуллера. Если модель стационарна, ставим  $d = 0$ , модель ARIMA становится ARMA. Если наша модель не стационарна, то дифференцируем, и снова проверяем стационарность. Если модель стационарна, присваиваем  $d = 1$ . Для финансово-экономических процессов значение  $d > 1$  не свойственно, так как в этом случае процесс является «взрывным». Возникновение таких процессов маловероятно, так как финансово-экономическая среда достаточно инерционная, что не позволяет принимать бесконечно большие значения за малые промежутки времени [19].

Значение  $q$  – порядок скользящего среднего. На основе полученных параметров, с помощью модели ARIMA строится прогнозирующая функция, полученные данные сравниваются с реальными, находятся накопительные ошибки, среднеквадратичное значение ошибок и дисперсия ошибок. На основе этих данных по каждому банкомату делается предварительный вывод о применимости данной модели.

#### 1.4 Выводы по разделу

На основании рассмотренного теоретического материала по прогнозированию движения денежных средств было установлено следующее.

Данные типа временных рядов широко распространены в самых разных областях человеческой деятельности. В экономике это ежедневные цены на акции, курсы валют, еженедельные и месячные объемы продаж, годовые объемы производства и т.п. Прогнозирование – тема важная и широко применяемая в нашей жизни, следовательно, исследование этой темы актуально.

Прогнозирование движения денежных средств в банкоматах тема очень широкая и важная. При точном прогнозе, можно составить оптимальный план движения инкассации, что позволит минимизировать издержки, а также не допустит простоя банкомата.

Успешная деятельность коммерческого банка зависит не только от его финансовой и инвестиционной деятельности, значительную долю дохода банка обеспечивают операции с наличными средствами клиентов. С развитием техники банки могут размещать своих представителей буквально на каждом углу, эти представители – банкоматы. Пользователи банкоматов – зачастую люди, которые живут или работают рядом, и поэтому не вызовет трудностей определение случайных величин, которые влияют на работу банкомата. Однако в случае поломки банкомата или нехватки в нем наличных денег банк начинает терять своих клиентов, поскольку, если клиент решил снять определенную сумму со счета – он ее все равно снимет, только воспользуется другим банкоматом. А это означает, что банк может недополучить определенную часть прибыли, т. е. понесет неявные потери. Необходимо найти такие оптимальные значения параметров функционирования банкомата, чтобы коммерческий банк не был в убытке от его эксплуатации.

Важными достоинствами класса моделей ARIMA являются их простота и прозрачность моделирования. Еще одним достоинством является единообразие анализа и проектирования, заложенное в работе. На сегодняшний день, данный класс моделей является одним из наиболее популярных [20], а потому в открытом доступе легко найти примеры применения авторегрессионных моделей для решения задач прогнозирования временных рядов различных предметных областей.

В главе был проведен обзор текущих моделей, описаны их преимущества и недостатки, а также проведен краткий анализ по каждой модели.

## 2 МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ARIMA

### 2.1 Математическая модель ARIMA

Общая формула модели ARIMA выглядит следующим образом:

$$\Delta^d X_t = c + \varepsilon_t + \sum_{i=1}^p \alpha_i \Delta^d X_{t-i} + \sum_{i=1}^q \beta_i \varepsilon_{t-i}, \quad (3)$$

где  $\Delta^d$  – оператор разности временного ряда порядка  $d$ ;

$X_t$  – прогнозируемое значение;

$c$  – некоторая постоянная;

$\alpha_i$  – коэффициенты авторегрессии;

$\beta_i$  – параметры модели скользящего среднего;

$\varepsilon_t$  – белый шум;

$p$  – число авторегрессионных лагов.

График ACF – это гистограмма коэффициентов корреляции между временным рядом и лагами своих же значений.

Коэффициент автокорреляции считается по формуле [21]:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (4)$$

где

$$\bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}, \quad (5)$$

$$\bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1}, \quad (6)$$

Корреллограмма представлена на рисунке 2.1.

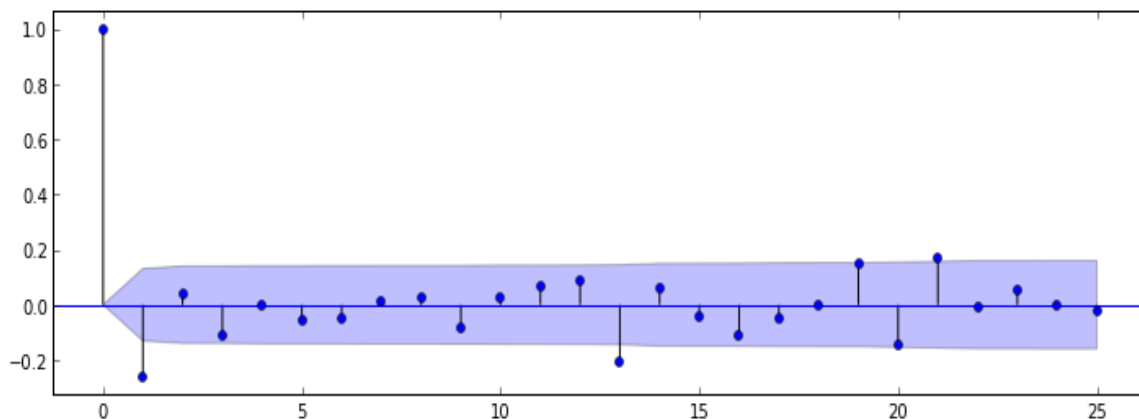


Рисунок 2.1 – График ACF

По корреллограмме ACF выбирается лаг, который обладает самой высокой корреляцией. Значение этого лага это - период данных. Это значение является значением  $p$ .

$p$  – число авторегрессионных членов (часть AR). Это позволяет включить влияние прошлых значений в модель ARIMA [22].

График PACF – это график коэффициентов частичной корреляции между сериями и их лагами. PACF поможет нам определить  $q$ , т. к. по ее корреллограмме можно определить количество автокорреляционных коэффициентов сильно отличных от 0 в модели MA. Для сглаживания подобных пиков и применяется модель скользящего среднего, которая, по сути дела, представляет собою простой фильтр низких частот [23]. Другими словами – MA необходима для сглаживания резких выбросов, которые, в свою очередь, могут сильно увеличить ошибку.

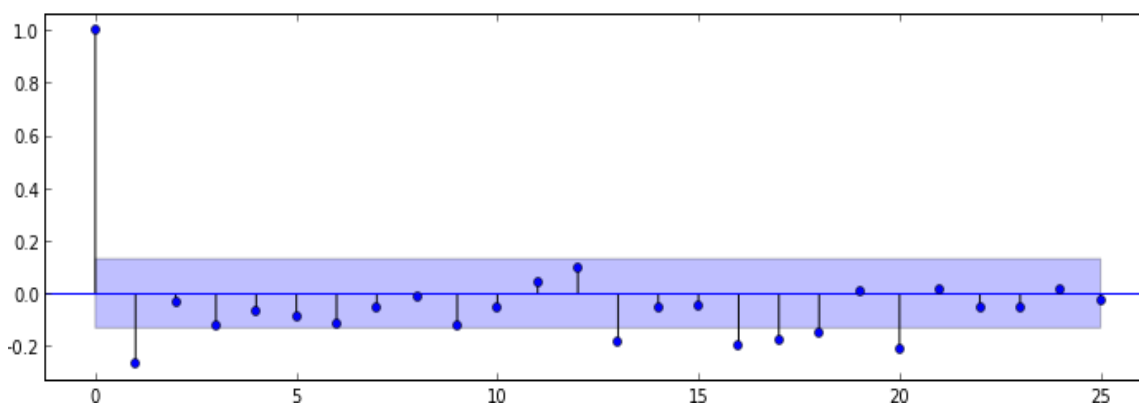


Рисунок 2.2 – График PACF

Для получения значения  $d$ , применяется обобщенный тест Дики–Фуллера:

$$\Delta y_t = b_0 + b_1 \cdot t + b \cdot y_{t-1} + \varepsilon_t, \quad (7)$$

где  $\Delta$  – оператор разности первого порядка;

$y_t$  – значение тестовой регрессии;

$b_0$  – константа;

$b_1 \cdot t$  – линейный тренд;

$\varepsilon_t$  – ошибка.

Для тестовой регрессии существуют свои критические значения  $DF$ -статистики, которые берутся из специальной таблицы Дики-Фуллера. Если значение статистики лежит левее критического значения (критические значения – отрицательные) при данном уровне значимости, то нулевая гипотеза о единичном корне отклоняется и процесс признается стационарным (в смысле данного теста). В противном случае гипотеза не отвергается и процесс может содержать единичные корни, то есть быть нестационарным (интегрированным) временным рядом.

Очевидно, что построив модель

описание ряда. Это в итоге сказывается на оценке коэффициента  $a$ , а значит, и на финальном результате проверки гипотезы. Поэтому для того, чтобы учесть все элементы ряда, был предложен расширенный тест Дики-Фуллера, отличающийся от первоначального теста только тем, что в добавляются еще  $p$  значений предыдущих  $y$  в разностях (что в итоговой модели фактически соответствует авторегрессии  $p + 1$  порядка).

Предложенный авторами класс тестов, получивших название “расширенные тесты Дики-Фуллера”, по сей день является самой распространенной процедурой тестирования гипотезы о наличии единичных корней в данных.

Тест Дики-Фуллера, как и многие другие тесты, проверяют наличие лишь одного единичного корня. Однако, процесс может иметь теоретически

несколько единичных корней. В этом случае тест может быть некорректным. Поскольку обычно предполагается, что больше трёх единичных корней вряд ли могут встречаться в реальных экономических временных рядах, то теоретически обоснованным является тестирование в первую очередь вторых разностей ряда. Если гипотеза единичного корня для этого ряда отвергается, то тогда тестируется единичный корень в первых разностях. Если на этом этапе гипотеза не отвергается, то исходный ряд имеет два единичных корня. Если отвергается, то проверяется единичный корень в самом временном ряде, как описано выше. На практике часто все делают в обратной последовательности, что не совсем корректно. Для корректных выводов необходимы результаты тестов для вторых и первых разностей наряду с самим временным рядом [24].

С помощью теста Дики-Фуллера определяется значение коэффициента  $d$ . При отсутствии единичных корней  $d = 0$  и процесс является стационарным, при  $d > 0$  в процессе присутствует тренд.

## 2.2 Описание алгоритма

Ниже представлен алгоритм в виде блок-схем, построенный для вычисления входных параметров и работы модели с этими параметрами, алгоритм представлен на рисунках 2.3-2.5.

Далее этот алгоритм будет написан на языке Python, с использованием сторонних библиотек.

Алгоритм создает все необходимые переменные, заполняет пропуски в данных, считает полученную ошибку, сравнивает два результата и сохраняет их в виде csv-файла и картинки png. На вход алгоритм получает csv-файл с тестовыми данными.

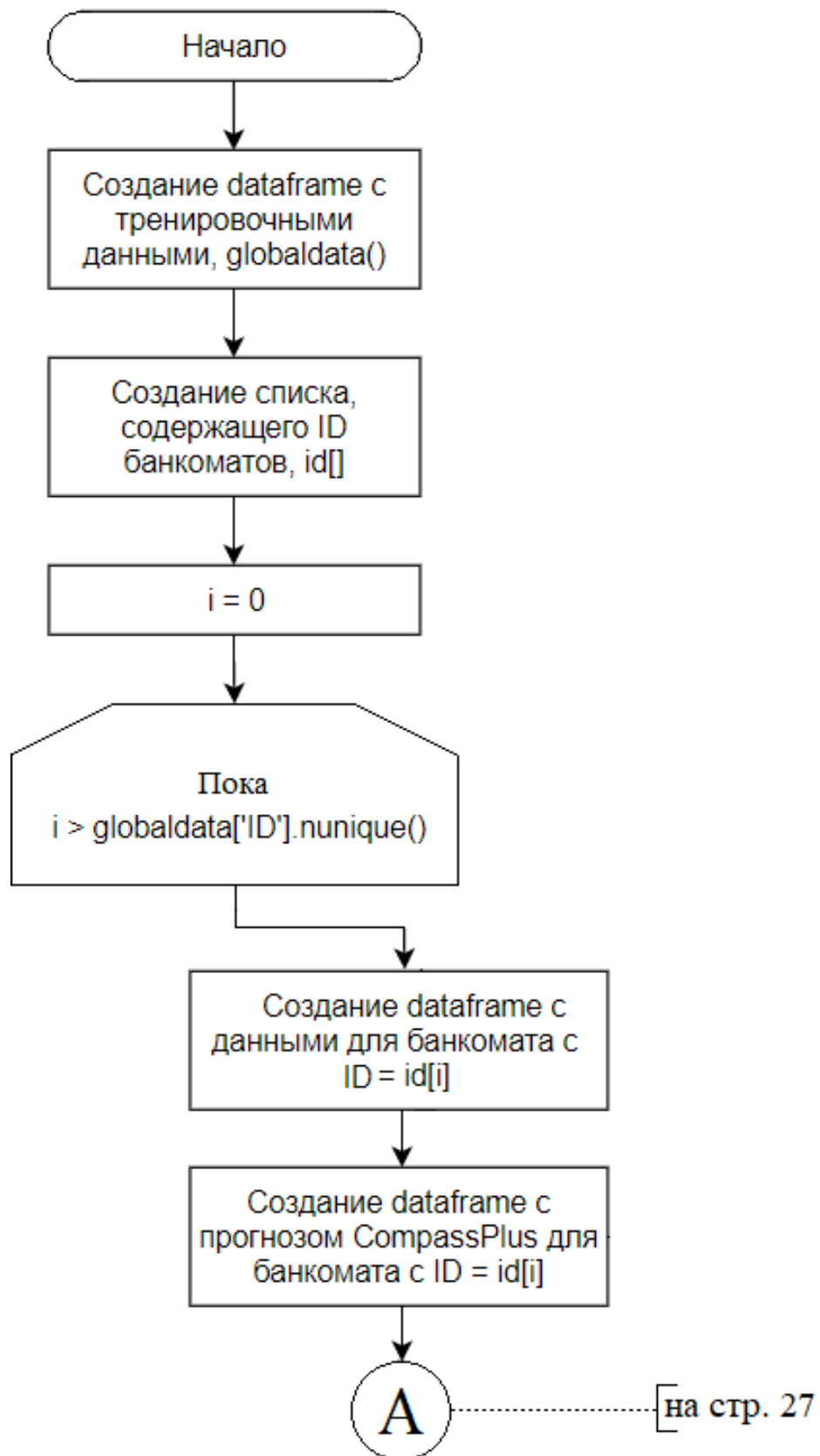


Рисунок 2.3 – Схема прогнозирующего алгоритма

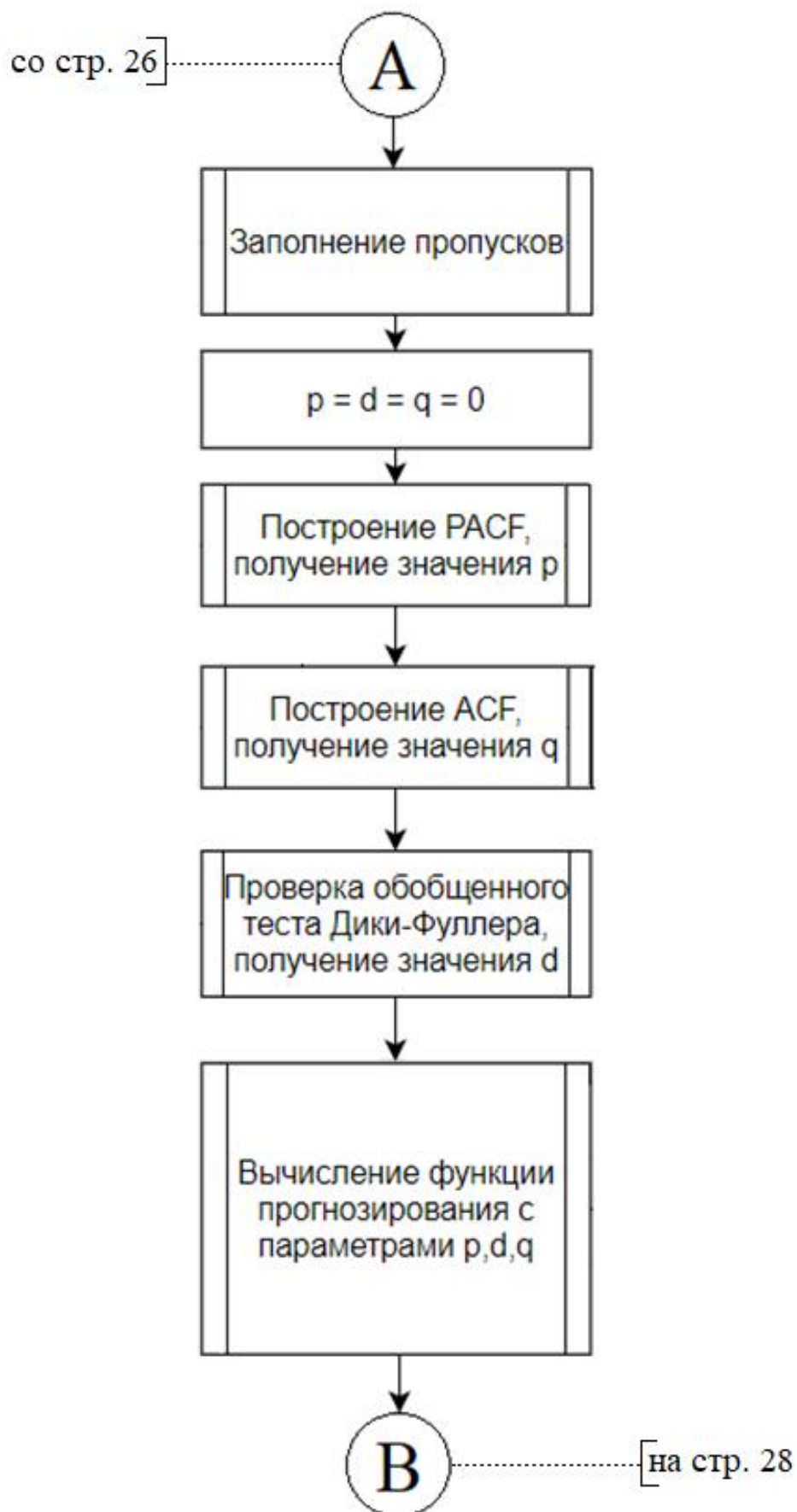


Рисунок 2.4 – Продолжение



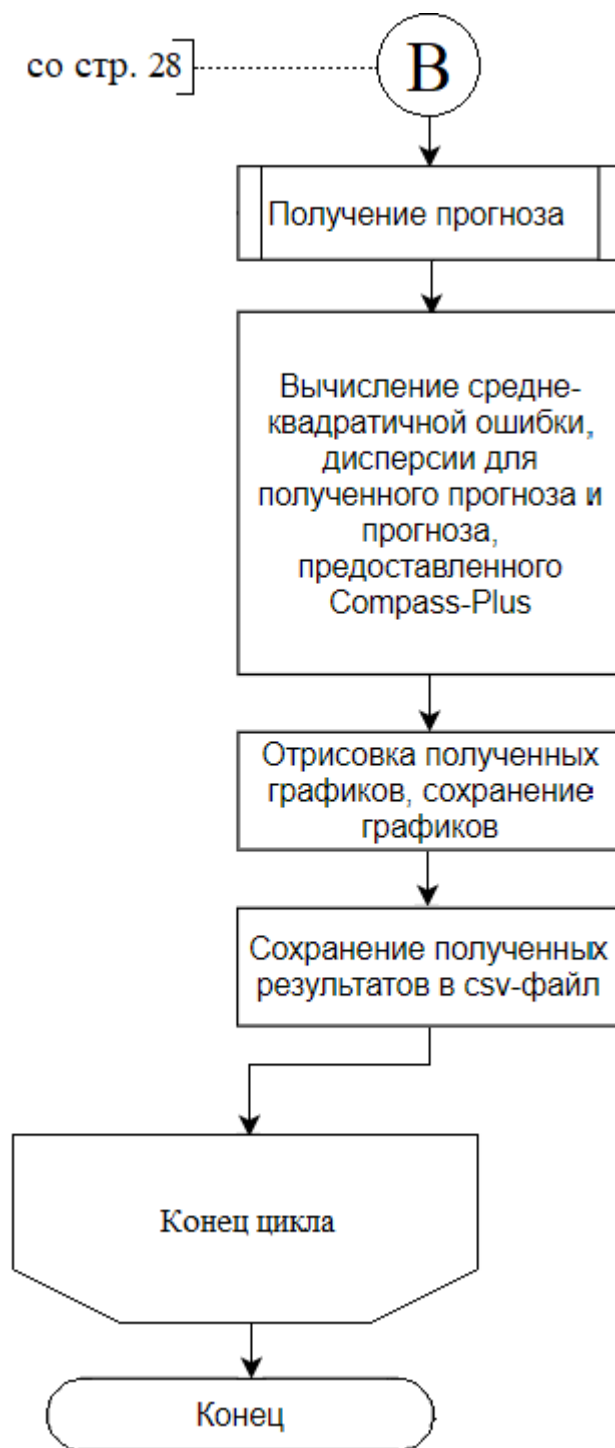


Рисунок 2.5 – Окончание

Ниже представлено описание алгоритма и программная реализация некоторых функций.

1) Из csv-файла считываются данные, которые записываются в DataFrame. Эти данные будут являться тренировочными данными, на основе которых будет считаться функция прогнозирования. При этом используются следующие функции:

Создание датафрейма:

```
globaldata = pd.DataFrame
```

Заполнение датафрейма из файла с данными:

```
globaldata = pd.read_csv('sравнение.csv', sep=',',  
index_col='DATE', parse_dates = True)
```

Сортировка датафрейма по индексам:

```
globaldata = globaldata.sort_index()
```

2) Из csv-файла также формируется список ID банкоматов, который будет необходим для перебора. Формирование списка происходит с помощью функции:

```
id = globaldata['ID'].drop_duplicates().tolist()
```

3) Цикл работает  $i$  раз, где  $i$  принимает значения от 1 до `globaldata['ID'].nunique()` – количества банкоматов.

4) Создание DataFrame с фактическими данными о банкомате с `ID = id[i]`. DataFrame содержит в себе данные о количестве снимаемых средств в предыдущие дни. Удаляются столбцы, которые содержат в себе информацию об ID – он находится в списке всех ID, валюте – она не важна и столбец, отвечающий за информацию о том, является данная сумма фактической или прогнозом, так как вся информация находится в одном файле.

Создание датафрейма с тренировочными данными из общего датафрейма со всеми данными:

```
datasumm = globaldata.loc[(globaldata['ID'] ==  
id[i]) & (globaldata['SIGN'] == "Fact")]
```

Удаление столбца CURRENCY:

```
datasumm.drop('CURRENCY', axis=1, inplace=True)
```

Удаление столбца ID:

```
datasumm.drop('ID', axis=1, inplace=True)
```

Удаление столбца DIRECT:

```
datasumm.drop('DIRECT', axis=1, inplace=True)
```

Удаление столбца SIGN:

```
datasumm.drop('SIGN', axis=1, inplace=True)
```

5) Создание DataFrame с прогнозируемыми Compass Plus данными о банкомате с ID = id[i]. DataFrame содержит в себе прогнозируемые суммы в конкретный период. Удаляются столбцы, которые содержат в себе информацию об ID – он находится в списке всех ID, валюте – она не важна и столбец, отвечающий за информацию о том, является данная сумма фактической или прогнозом, так как вся информация находится в одном файле.

Создание датафрейма с прогнозом компании «Compass Plus»:

```
predcompass = globaldata.loc[(globaldata['ID'] == id[i]) & (globaldata['SIGN'] == "Forecast")]
```

Удаление столбца CURRENCY:

```
predcompass.drop('CURRENCY', axis=1, inplace=True)
```

Удаление столбца ID:

```
predcompass.drop('ID', axis=1, inplace=True)
```

Удаление столбца DIRECT:

```
predcompass.drop('DIRECT', axis=1, inplace=True)
```

Удаление столбца SIGN:

```
predcompass.drop('SIGN', axis=1, inplace=True)
```

б) Заполнение пропусков в фактических данных, это необходимо для работы функции прогнозирования.

Получение даты первого тренировочного значения:

```
start = datasumm.index[0]
```

Получение даты последнего тренировочного значения:

```
finish = datasumm.index[-1]
```

Получение списка промежуточных дат:

```
idx = pd.date_range(start, finish)
```

Заполнение отсутствующих дат в датафрейме:

```
datasumm = datasumm.reindex(idx)
```

Заполнение отсутствующих значений в датафрейме:

```
datasumm['SUMM'].fillna(method = 'bfill',  
inplace=True)
```

7) Вызов функции, вычисляющей параметр  $p$  с помощью автокорреляции. На вход функция получает датафрейм с тренировочными данными, количество лагов, для которых вычисляется значение авторегрессии и метод подсчета.

Функция вычисляющая PACF:

```
pp = sm.tsa.stattools.pacf(uchebnie, nlags=40,  
method='ols', alpha=None)
```

Внутри функции осуществляется цикл, находящий номер лага, на котором достигается максимальное значение автокорреляции. Возвращает функция максимальное значение автокорреляции и номер лага, на котором достигается максимальная автокорреляция.

8) Вызов функции, вычисляющей параметр  $d$  с помощью обобщенного теста Дики–Фуллера. На вход функция получает временной ряд. Возвращает функция значение  $d$ .

Получение значения теста Дики-Фуллера:

```
test = sm.tsa.adfuller(datasumm['SUMM'])
```

Создание датафрейма из тренировочных сумм, не содержащего последнего значения:

```
buff = datasumm['SUMM'].diff(periods = 1).dropna()
```

Проверка значений теста и подбор значения  $d$ :

9) Вызов функции, вычисляющей параметр  $q$  с помощью частичной автокорреляции. На вход функция получает датафрейм с тренировочными данными, количество лагов, для которых вычисляется значение авторегрессии и метод подсчета. Возвращает функция максимальное значение автокорреляции и номер лага, на котором достигается максимальная автокорреляция.

10) Функция вычисляющая ACF:

```
qq = statsmodels.tsa.stattools.acf(uchebnie,  
unbiased=False, nlags=40, qstat=False, fft=None,  
alpha=None)
```

Цикл, находящий номер лага, на котором достигается максимальное значение автокорреляции:

11) Вызов функции, вычисляющей прогнозирующую функцию для временного ряда. На вход функция получает датафрейм с данными по одному банкомату  $b$  параметры  $p, d, q$ . Возвращает функция прогнозирующую функцию.

```
mod = sm.tsa.statespace.SARIMAX(uchebnie,  
order = (p, d, q),  
measurement_error = True,  
enforce_stationarity = False,  
simple_differencing = True,  
enforce_invertibility = False)  
results = mod.fit(dispatch = False)
```

12) Получение прогноза.

```
preds = results.predict(start = finish -  
datetime.timedelta(days=14), end = finish)
```

13) Получение значений среднеквадратичных отклонений полученного прогноза и прогноза компании «Compass Plus».

Создание списка ошибок для исследуемой модели ARIMA:

```
errormy = []
```

Создание списка ошибок прогноза компании «Compass Plus»:

```
errorcompass = []
```

14) Графическое изображение и сохранение полученных результатов.

Создание фигуры для отрисовки:

```
plt.figure(figsize=(15, 10))
```

Отрисовка фактических данных:

```
plt.plot(datasumm[finishdatetime.timedelta(days=14)
: finish],color='red',label = u'Фактические')
```

Отрисовка полученного прогноза моделью ARIMA:

```
plt.plot(preds,color='black',label = u'Прогноз
ARIMA')
```

Отрисовка прогноза компании «Compass Plus»:

```
plt.plot(predcompass,color='blue',label = u'Прогноз
Compass')
```

Подпись легенды:

```
plt.title(f'Банкомат № {id[i]}')
```

```
plt.legend()
```

Отображение графика:

```
plt.show()
```

Сохранение графика:

```
plt.savefig(f'Results\{i}.png')
```

15) Заполнение результатами в csv-файла:

```
errors = errors.append({'ID':id[i],
'my_std':np.std(errormy),'compass_std':np.std(
errorcompass)}, ignore_index=True)
```

Сохранение полученных данных в csv-файл:

```
errors.to_csv('factpredfinal.csv', sep=',')
```

### 2.3 Условия применения разработанного программного обеспечения и оценка требуемых ресурсов

Программное обеспечение (ПО) работает с любыми данными, которые находятся в csv-файле и представляют собой структуру, в которой находятся 6 столбцов данных:

1) ID – идентификатор банкомата, для которого строится прогнозирующая функция. Необходим для перебора всех банкоматов и выборке данных по конкретному банкомату. Обязателен;

2) DATE – даты и время, для которых известны суммы снятых средств. Выступает идентификатором при построении прогноза. Обязателен;

3) CURRENCY – код валюты по стандарту Международной организации по стандартизации (ISO 4217). Необходим только при наличии данных с несколькими видами валют. Необязателен;

4) DIRECT – столбец, показывающий, фактическая сумма или искусственно заполненная (1 для фактических данных, 0 для интерполированных данных). Необходим только если заполнение промежутков происходит в отдельном модуле. Необязателен;

5) SIGN – столбец, показывающий, фактическая сумма или спрогнозированная («Fact» - для фактических данных, «Forecast» - для полученного прогноза). Необходим только для сравнения прогноза, либо хранения данных для последующего использования. Необязателен;

6) SUMM – количество снятых средств в день, указанный в столбце Date (сумма снятых средств в течении суток). Обязателен.

Удаление, изменение или добавление входных данных происходит в коде программы. Выходные данные предоставляют собой csv-файл, содержащий результат работы прогнозирующей функции и значения среднеквадратичного отклонения и дисперсии, а также графики, сохраненные в .png формате для визуального представления.

Программный продукт разрабатывался на электронно-вычислительной машине, имеющей:

1) процессор Intel Core i5-4210M, 3 МБ кэш-памяти, тактовая частота до 3,20 ГГц;

2) оперативная память DDR3, объемом 6Gb;

3) операционная система – Windows 10.

Среднее время работы программы составляет 20 минут. За это время была вычислена прогнозирующая функция, среднеквадратичная ошибка и дисперсия для полученного в результате работы программы прогноза и

прогноза компании «Compass Plus» для 171 банкомата. Среднее время выполнения программы для одного банкомата составило 7 секунд.

## 2.4 Выводы по разделу

В данной главе была рассмотрена математическая модель и программная реализация алгоритма. Данное изучение позволило сделать ряд выводов.

Авторегрессия – модель временных рядов, в которой значение временного ряда в данный момент времени может быть выражено в виде линейной комбинации предыдущих значений этого же ряда и случайной ошибки, обладающей свойством «белого шума».

Процесс скользящего среднего – текущее значение случайного процесса представляется в виде линейной комбинации текущего и прошлых значений ошибки, по своим свойствам соответствующей «белому шуму».

Интегрированная составляющая отвечает за порядок дифференциации ряда при необходимости привести ряд к стационарному. Обычно рассчитывается разница до 2-го порядка. В эконометрике обычно до 1-го порядка.

Главной задачей является определение значений входных параметров модели  $p, d, q$ . Для применения модели ARIMA необходимо иметь данные в разрезе времени любого уровня динамики (годы, месяцы, дни и т.д.). В данном случае предоставлены данные о снятых средствах по дням.

Был построен и программно реализован алгоритм, позволяющий автоматически вычислять входные параметры  $p, d, q$ ; а также вычислять прогнозирующую функцию и с помощью данной функции получать значения в следующий момент времени.

Таким образом, можно сделать вывод о том, что математическая модель была успешно исследована, выбраны верные математические методы. Алгоритм успешно реализован на языке Python в виде программного модуля.



### 3 ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

#### 3.1 Разработка методологии тестирования программы

Среднеквадратичное отклонение измеряется в единицах измерения самой случайной величины. Оно используется как обобщающая характеристика набора отклонений.

В качестве оценки погрешности прогнозирования выбранной модели, будем использовать среднеквадратичное отклонение ошибки, причем ошибкой будем считать функцию разности между прогнозируемым значением и реальным.

Подсчет ошибки выглядит следующим образом:

$$\sigma = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}, \quad (8)$$

где  $e_t$  – функция ошибки;

$\hat{y}_t$  – прогнозируемые значения временного ряда;

$y_t$  – реальные значения временного ряда.

Среднее значение выборки рассчитывается по формуле:

$$\bar{y} = \frac{1}{n} \sum_{t=1}^n y_t \quad (9)$$

где  $\bar{y}$  – среднее значение выборки;

$y_t$  – значение выборки;

$n$  – количество элементов в выборки.

Выборочная дисперсия, необходимая для среднего математического отклонения, рассчитывается по формуле:

$$s^2 = \frac{1}{n} \sum_{t=1}^n (y_t - \bar{y})^2 \quad (10)$$

где  $s^2$  – выборочная дисперсия;

$\bar{y}$  – среднее значение выборки;

$y_t$  – значение выборки;

$n$  – количество элементов в выборки.

Среднее математическое отклонения выборки рассчитывается по формуле:

$$\text{---} \quad (11)$$

где  $\sigma$  – среднее математическое отклонения выборки;

$s^2$  – выборочная дисперсия.

### 3.2 Сравнительный анализ результатов прогноза

Компанией «Compass Plus» помимо тренировочных данных, был так же предоставлен прогноз, который используется в текущий момент компанией. Сама модель, используемая компанией неизвестна, потому вывод будет строиться только на основании среднеквадратичного отклонения, не беря во внимание сложности алгоритмов.

Сравнительные результаты работы алгоритмов прогнозирования представлены на рисунках 3.1-3.4. Помимо полученного моделью ARIMA прогноза и прогноза, полученного компанией «Compass Plus» представлен график фактического снятия денежных средств с банкоматов, для визуальной оценки работы алгоритмов.

Как видно на рисунке 3.1, графики полученного прогноза и прогноза компании «Compass Plus» различны, они достигают максимальной точности в различных точках. Для сравнения, среднеквадратичное отклонение для полученного прогноза составляет 104 187 рублей, а для прогноза компании «Compass Plus» – 103 442 рубля. Разница в прогнозах составляет менее одного процента в пользу «Compass Plus».

Преимуществом модели ARIMA является то, что в первую неделю среднеквадратичная ошибка меньше, чем у прогноза «Compass Plus», следовательно, модель смотрится более выигрышно.

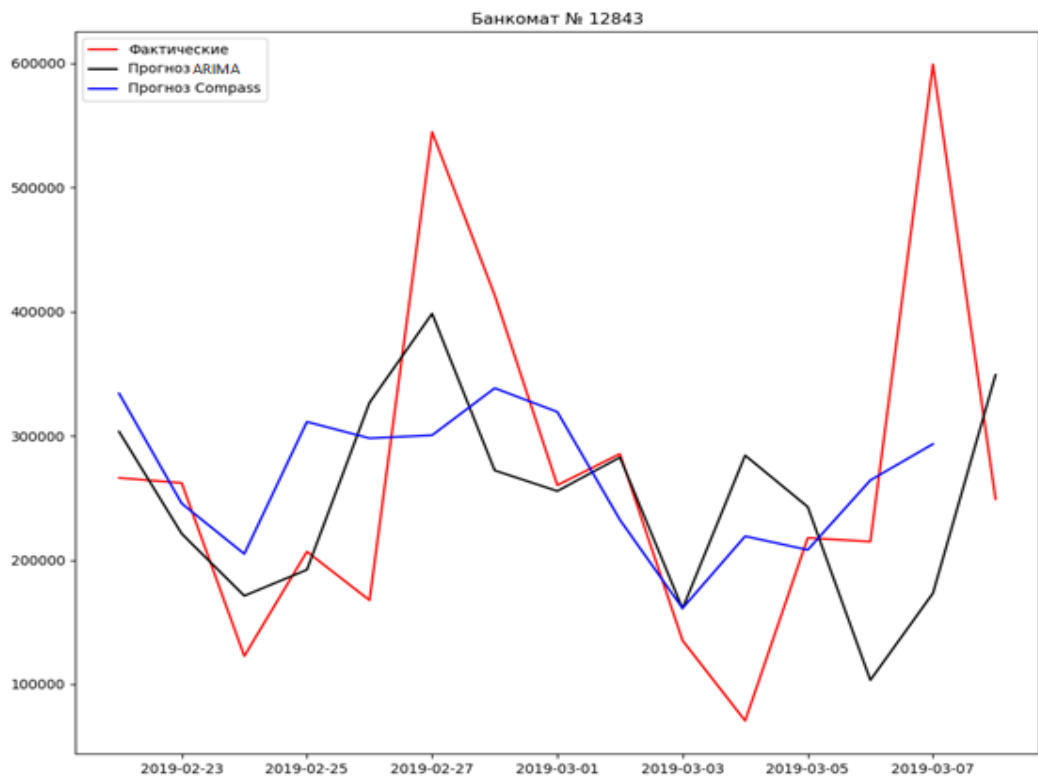


Рисунок 3.1 – Результат работы программы, для банкомата под номером 12843

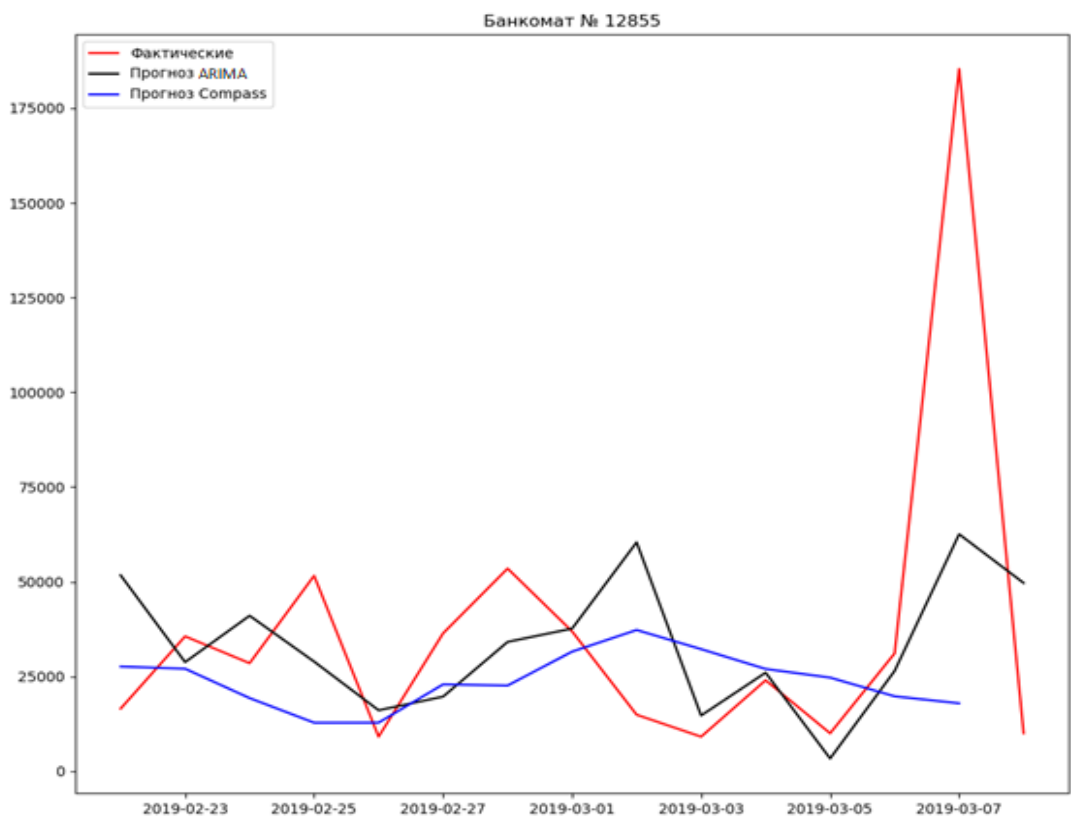


Рисунок 3.2 – Результат работы программы, для банкомата под номером 12855

Как видно на рисунке 3.2, в целом графики ведут себя похоже, но за счет того, что полученный прогноз на 07-03-2019 был более точен, хоть и допустил значительное отклонение, оказался более точным. Среднеквадратичное отклонение полученного прогноза составляет 17 306 рублей, компании «Compass Plus» – 18 294 рубля. Разница прогнозов составила 5,5% в пользу полученных исследуемой моделью данных.

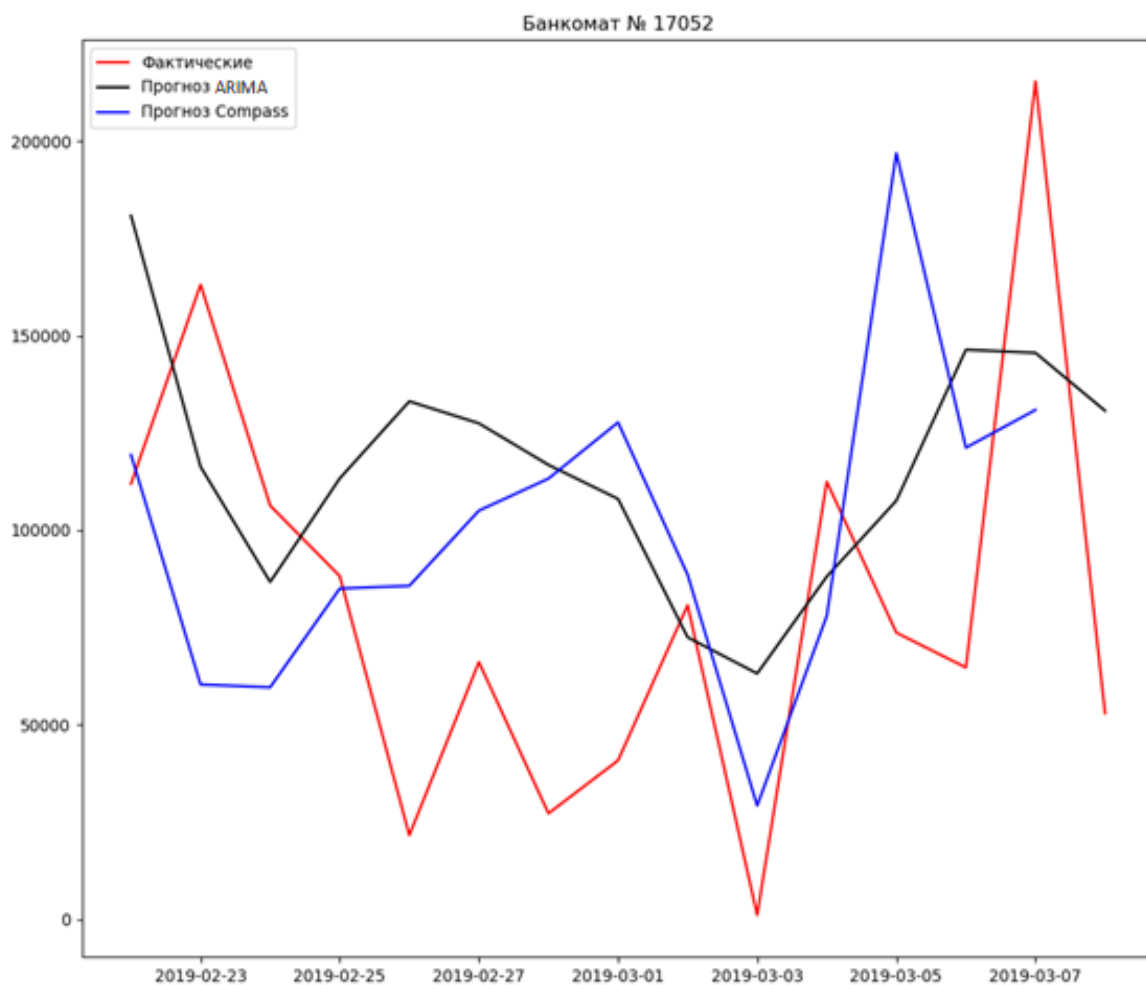


Рисунок 3.3 – Результат работы программы, для банкомата под номером 17052

В данном случае, из-за того, что банкоматы мало используются, их прогнозирование достаточно трудоёмкий процесс, так как при низких значениях, показатели банкомата более похожи на случайные блуждания, но ими не являются. В результате полученных результатов, среднеквадратичное отклонение прогноза компании «Compass Plus» составил 61933 рубля,

среднеквадратичное отклонение для полученного прогноза составляет 48846 рублей. Разница прогнозов составляет более 21%, что существенно.



Рисунок 3.4 – Результат работы программы, для банкомата под номером 23523

Модель ARIMA более успешно справилась с прогнозированием резких спадов и всплесков. В данном случае, среднеквадратичное отклонение прогноза компании «Compass Plus» составил 160293 рубля, среднеквадратичное отклонение для полученного прогноза составляет 136920 рублей. Разница прогнозов составляет почти 15%, к тому же, данные на ближайшую неделю – самые ценные, а модель ARIMA спрогнозировала поведение движения средств очень точно именно в первой неделе.

Полученные в результате прогнозирования суммы либо примерно соответствуют точности модели, применяемой в компании «Compass Plus», либо превосходят ее. Все сравнения производились в соответствии с результатами расчетов среднеквадратичной ошибки.

Таким образом, программа отвечает предъявленным требованиям.

### 3.3 Методы повышения точности модели

Полученные при прогнозировании результаты обладают достаточно высокой точностью, но их все же можно улучшить. Предлагается несколько способов, которые теоретически могут дать заметный прирост в точности результата.

Первый способ – добавление в модель экзогенных данных.

Экзогенная переменная – переменная характеристика объекта, отражающая условия и свойства внешней по отношению к данному объекту среды. Используется при решении задач, связанных с экономико-математическим моделированием. Экзогенные переменные чаще всего учитываются для ограничения экономико-математической модели. Экзогенные переменные – это известные данные, которые подаются на вход модели, они обуславливают внешние факторы, которые учитывает модель.

Модель ARIMA, с добавлением экзогенных данных, называется моделью ARIMAX (autoregressive integrated moving average eXtended). Расширенное авторегрессионное интегрированное скользящее среднее. Есть некоторые работы, использующие данный метод [25, 26].

Для улучшения точности модели, можно добавить такие экзогенные факторы, как:

1) дни недели, так как у банкоматов, находящихся на заводах, основной поток будет происходить в будние дни, в то время, как основной поток денежных средств в ТРК происходит в выходные дни;

2) праздничные дни, основные всплески, которые очень важно контролировать, происходят перед крупными праздниками (новый год, пасха, Рождество и т.д.);

3) месторасположение банкоматов, можно использовать как дополнительные экзогенные данные к пункту 1.

Второй способ улучшения точности модели – использование дополнительно к модели ARIMA модели GARCH. Данный способ описан в пункте 1.2. GARCH широко используется в финансовой индустрии, так как многие цены на активы условно гетероскедастичны. Использование смешанной модели [27, 28].

Гетероскедастичность, неоднородность – понятие математической статистики и эконометрии; означает случай, когда дисперсия ошибки в уравнении регрессии изменяется от наблюдения к наблюдению.

Очевидно, что данные о снятии средств в банкоматах – гетероскедастичны, следовательно, есть высокая вероятность того, что модель ARIMA+GARCH даст более точный прогноз.

Кроме того, для улучшения точности модели, можно воспользоваться другим интерполяционным методом для заполнения пропусков данных, ведь часто ошибки возникают в результате отсутствия данных, в следствии чего достаточно тяжело опираться на такие данные. Заполнение таких пропусков позволит снизить общую ошибку прогнозирования.

Еще одним способом, который, возможно, даст более точный прогноз, является добавление в модель ARIMA сезонных параметров. Расширение ARIMA, которое поддерживает прямое моделирование сезонного компонента ряда, называется SARIMA [29, 30].

SARIMA добавляет три новых гиперпараметра для указания авторегрессии (AR), разности (I) и скользящего среднего (MA) для сезонной составляющей ряда, а также дополнительный параметр для периода сезонности.

Сезонная модель ARIMA формируется путем включения дополнительных сезонных терминов в ARIMA. Сезонная часть модели состоит из параметров, которые очень похожи на несезонные компоненты модели, но включают обратные сдвиги сезонного периода.

### 3.4 Выводы по разделу

Полученные результаты показывают, что модель ARIMA не менее эффективна для прогнозирования движения денежных средств в некоторых банкоматах, чем модель, используемая в текущий момент компанией «Compass Plus». Изменение используемой модели для прогнозирования позволит снизить издержки, связанные с работой инкассаторской службы. Необходимо произвести исследование, выявляющее, на каких банкоматах модель ARIMA более эффективна. Так как это позволит увеличить точность прогнозирования некоторых банкоматов до 25%, что благополучно скажется на финансовой составляющей компании.

В разделе также были рассмотрены варианты изменения модели, которые в теории могут снизить среднеквадратичную ошибку вычисления прогноза.



## ЗАКЛЮЧЕНИЕ

В результате проведенного исследования была достигнута поставленная цель – было спрогнозировано движение денежных средств с помощью модели ARIMA.

Основное предназначение банкоматов – непрерывное движение денежных средств, которое может обеспечить постоянный доступ к обналичиванию денежных средств, находящихся на банковских картах. Так как наличные средства до сих пор остаются основным средством расчета, то прогнозирование снимаемых средств с банкоматов – тема особенно актуальная.

Целью работы является прогнозирование движения денежных средств в банкоматах. В следствии чего в ходе выполнения работы были поставлены и достигнуты следующие задачи:

- изучить теоретические основы эконометрики;
- рассмотреть особенности прогнозирования движения денежных средств в банкоматах;
- разработать алгоритм подбора входных параметров модели ARIMA и прогнозирования данной моделью;
- определить адекватность и применимость данной модели к прогнозированию денежных средств в банкоматах.

Были рассмотрены основные модели прогнозирования временных рядов. Важными достоинствами класса моделей ARIMA являются их простота и прозрачность моделирования, единообразие.

Главной задачей является определение значений входных параметров модели  $p, d, q$ , где  $p$  – порядок авторегрессии,  $d$  – порядок разности,  $q$  – порядок скользящего среднего.

Был построен и программно реализован алгоритм, автоматически вычисляющий входные параметры  $p, d, q$ ; а также вычисляющий прогно-

зирующую функцию и с помощью данной функции получающий значения в следующий момент времени.

Полученные функции показывают похожее поведение с фактическими данными – это значит они принадлежат одному случайному процессу. В сравнении с прогнозом, полученным компанией «Compass Plus», полученные результаты моделью ARIMA показывают результаты не хуже, из чего следует ее применимость.

Также были предложены следующие пути развития, связанные с моделью ARIMA:

- ARIMA + GARCH. Модель GARCH позволяет учитывать гетероскедатичность данных;

- SARIMA. Добавление дополнительных сезонных параметров в ARIMA;

- ARIMAX. Добавление экзогенных параметров в модель ARIMA.

Таким образом, основные задачи работы выполнены, цель достигнута.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- 1 Основные понятия прогностики. URL: <https://finlit.online/metodyi-modelirovanie-matematicheskie/osnovnyie-ponyatiya-prognostiki-32215.html> (дата обращения: 12.01.2019).
- 2 Анализ временных рядов и прогнозирование. URL: [https://pokrovka11.files.wordpress.com/2011/12/emetrix\\_time\\_series.pdf](https://pokrovka11.files.wordpress.com/2011/12/emetrix_time_series.pdf) (дата обращения 15.01.2019).
- 3 Жуковский, М.Е. Основы теории случайных процессов: учебное пособие / М.Е. Жуковский, И.В. Родионов, Д.А. Шабанов. – М.: Изд-во МФТИ, 2016. – 122 с.
- 4 Прогнозирование денежных потоков. URL: <https://studfiles.net/preview/2400639/> (дата обращения 29.01.2019).
- 5 Преимущества и недостатки моделей ARIMA. URL: [https://studme.org/41020/ekonomika/preimuschestva\\_nedostatki\\_modeley\\_arima](https://studme.org/41020/ekonomika/preimuschestva_nedostatki_modeley_arima) (дата обращения 15.01.2019).
- 6 Эконометрика: учебное пособие / В.А. Балаш, О.С. Балаш, А.И. Землянухин и др. – Саратов, 2005. – 80 с.
- 7 Федорова, Е.А. Прогнозирование фондового рынка Российской Федерации с помощью GARCH-моделирования / Е.А. Федорова, Д.А. Бузлов. URL: <https://cyberleninka.ru/article/v/prognozirovanie-fondovogo-rynka-rossiy-skojfederatsii-s-pomoschyu-garch-modelirovaniya> (дата обращения 03.03.2019).
- 8 Hosseinipoor, Saeid. (2016). Application of ARIMA and GARCH Models in Forecasting the Natural Gas Prices. / Hosseinipoor, Saeid. – 2016.
- 9 Time series and forecasting with Neural Networks. URL: <https://www.neuraldesigner.com/blog/time-series-and-forecasting-using-neural-networks> (дата обращения 02.04.2019).

10 Gheyas, I.A., Smith L.S. A Neural Network Approach to Time Series Forecasting / I.A. Gheyas, L.S. Smith // A Proceedings of the World Congress on Engineering, London, 2009, Vol 2 [электронный ресурс]. P. 1292 – 1296. URL: [www.iaeng.org/publication/WCE2009/WCE2009\\_pp1292-1296.pdf](http://www.iaeng.org/publication/WCE2009/WCE2009_pp1292-1296.pdf) (дата обращения 20.04.2019).

11 Neural Networks for Forecasting Financial and Economic Time Series. URL: <https://medium.com/microsoftazure/neural-networks-for-forecasting-financial-and-economic-time-series-6aca370ff412> (дата обращения 22.04.2019).

12 Gheyas, I.A. A Neural Network Approach to Time Series Forecasting // Gheyas I.A., Smith L.S. // Proceedings of the World Congress on Engineering, London, 2009, Vol 2 [электронный ресурс]. P. 1292 – 1296. URL: [www.iaeng.org/publication/WCE2009/WCE2009\\_pp1292-1296.pdf](http://www.iaeng.org/publication/WCE2009/WCE2009_pp1292-1296.pdf) (дата обращения 29.04.2019).

13 Morariu, N. Neural network model for time series forecasting // N. Morariu, E. Iancu, S. Vlad // Romanian Journal of Economic Forecasting. 2009, No. 4. P. 213 – 223.

14 Соломкин, А.В. Краткосрочное прогнозирование потребления электроэнергии с помощью нейросетевых методов // А.В. Соломкин, Электроника и информационные технологии [электронный ресурс]. 2011, №1. 5 с. URL: [http://fetmag.mrsu.ru/2009-3/pdf/Forecasting\\_electricity\\_consumption.pdf](http://fetmag.mrsu.ru/2009-3/pdf/Forecasting_electricity_consumption.pdf) (дата обращения 29.04.2019).

15 Селиверстова, А.В. Сравнительный анализ моделей и методов прогнозирования // Селиверстова А.В., Современные научные исследования и инновации. 2016. № 11 [Электронный ресурс]. URL: <http://web.snauka.ru/issues/2016/11/74271> (дата обращения: 25.03.2019).

16 Кизбикенов, К.О. Прогнозирование и временные ряды: учебное пособие / К. О. Кизбикенов. – Барнаул : АлтГПУ, 2017. – 115 с.

17 Балонишников, А.М. ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ МЕТОДОМ ФАРМЕРА-СИДОРОВИЧА И БОКСА-ДЖЕНКИНСА. // А.М. Балонишников, В.А. Балонишникова, А.В. Копыльцов, URL:

<https://cyberleninka.ru/article/v/prognozirovanie-vremennyh-ryadov-metodami-farmera-sidorovicha-i-boksa-dzhenkinsa>.

18 Time series Forecasting ARIMA models. URL: <https://towardsdatascience.com/time-series-forecasting-arma-models-7f221e9eee06> (дата обращения 14.04.2019).

19 Тест Дики-Фуллера. URL: [https://ru.wikipedia.org/wiki/%D0%A2%-D0%B5%D1%81%D1%82\\_%D0%94%D0%B8%D0%BA%D0%B8\\_%E2%80%94%D0%A4%D1%83%D0%BB%D0%BB%D0%B5%D1%80%D0%B0](https://ru.wikipedia.org/wiki/%D0%A2%-D0%B5%D1%81%D1%82_%D0%94%D0%B8%D0%BA%D0%B8_%E2%80%94%D0%A4%D1%83%D0%BB%D0%BB%D0%B5%D1%80%D0%B0) (дата обращения 11.04.2019).

20 Чучуева, И.А., МОДЕЛЬ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ ПО ВЫБОРКЕ МАКСИМАЛЬНОГО ПОДОБИЯ // Чучуева, И.А., МГТУ им. Н.Э. Баумана, Москва, 2012.

21 Автокорреляция, коэффициент автокорреляции. URL: <http://univer-nn.ru/ekonometrika/avtokorrelyaciya-koefficient-avtokorrelyacii/> (дата обращения 20.04.2019).

22 Шанченко, Н.И. Эконометрика: лабораторный практикум : учебное пособие / Н. И. Шанченко. – Ульяновск : УлГТУ, 2011. – 117 с.

23 Модель прогнозирования ARIMAX: скользящее среднее. URL: <http://www.mbureau.ru/blog/model-prognozirovaniya-arimax-skolzyashchee-srednee> (дата обращения 30.03.2019).

24 Тест Дики-Фуллера. URL: <http://poivs.tsput.ru/ru/Math/ProbabilityAndStatistics/MathStatistics/TimeSeriesAnalysis/DickeyFullerTest> (дата обращения 05.04.2019).

25 KOFI AGYARKO ABABIO COMPARATIVE STUDY OF STOCK PRICE FORECASTING USING ARIMA AND ARIMAX MODELS, 2012, URL: <http://ir.knust.edu.gh/bitstream/123456789/3904/1/Final%20Thesis%202012%2C%20KOFI%20AGYARKO%20ABABIO.pdf> (дата обращения 14.05.2019).

26 ARIMAX forecasting : Applied to the Baltic Supramax Index, Erlandsen, Karoline Håseth, 2008.

27 Применение модели ARIMA-GARCH для прогнозирования курса рубля на R. URL: <https://smart-lab.ru/blog/327500.php> (дата обращения 11.05.2019).

28 ARIMA+GARCH Trading Strategy on the S&P500 Stock Market Index Using R. URL: <https://www.quantstart.com/articles/ARIMA-GARCH-Trading-Strategy-on-the-SP500-Stock-Market-Index-Using-R> (дата обращения 11.05.2019).

29 Fredrik Nikolaisen Sävås Forecast Comparison of Models Based on SARIMA and the Kalman Filter for Inflation/Fredrik Nikolaisen Sävås/– Independent Thesis Advanced Level Department of Statistics Uppsala University, 2013.

30 Oduro-Gyimah, F. K. Sarima Time Series Model Application to Microwave Transmission of Yeji-Salaga (Ghana) Line-Of-Sight Link // F. K. Oduro-Gyimah, E. Harris, K. F. Darkwah, Department of Mathematics Kwame Nkrumah University of Science and Technology Kumasi-Ghana / 2012.

## ПРИЛОЖЕНИЕ 1

### Текст программы

```
# Загрузка библиотек
import pandas as pd
import datetime
import matplotlib.pyplot as plt
import statsmodels.api as sm
import statsmodels.formula.api as smf

# Создание, заполнение общего датафрейма с тренировочными
данными, сортировка
globaldata = pd.DataFrame
globaldata = pd.read_csv('otsort.csv', sep=',', index_col='DATE',
parse_dates = True)
globaldata = globaldata.sort_index()
globaldata["real"] = 1

errors = pd.DataFrame
errors = pd.read_csv('error.csv', sep=',', index_col=None, parse_dates =
True)
errors = errors.sort_index()

# Создание списка ID банкоматов для перебора
id = globaldata['ID'].drop_duplicates().tolist()
id.sort()
globaldata.resample('D')

# Проход по всем банкоматам
for i in range(0,globaldata['ID'].nunique()): #globaldata['ID'].nunique()
```

```
# Создание датафрейма с тренировочными данными для конкретного  
банкомата с данными DATE/SUMM
```

```
datasumm = globaldata.loc[globaldata['ID'] == id[i]]
```

```
datasumm.drop('ID', axis=1, inplace=True)
```

```
# Заполнение неизвестных дней в банкоматах
```

```
start = datasumm.index[0]
```

```
finish = datasumm.index[-1]
```

```
idx = pd.date_range(start, finish)
```

```
datasumm = datasumm.reindex(idx)
```

```
datasumm['SUMM'].fillna(method = 'bfill', inplace=True)
```

```
datasumm.fillna(0)
```

```
print('----- ID = ',id[i],'-----')
```

```
for j in range (1,14) :
```

```
    start = datetime.date(2018,11,17) + datetime.timedelta(days=7*(j-1))
```

```
    finish = start + datetime.timedelta(days=6)
```

```
    vivod = datasumm[start:finish]
```

```
    uchebnie = datasumm['SUMM'][:start - datetime.timedelta(days=1)]
```

```
    max = p = d = 0
```

```
    pp = sm.tsa.stattools.pacf(ucheбnie, nlags=40, method='ols',  
alpha=None)
```

```
    for k in range(5,34) :
```

```
        if (pp[k] > max) :
```

```
            max = pp[k]
```

```
            p = k
```



```

test = sm.tsa.adfuller(datasumm['SUMM'])
buff = datasumm['SUMM'].diff( periods = 1).dropna()
if test[0]> test[4]['5%']:
    test1 = sm.tsa.adfuller(buff)
    if test1[0]> test1[4]['5%']:
        d = 1
    else :
        d = 2

q = 1
mod = sm.tsa.statespace.SARIMAX(uchebnie,
    order = (p, d, q),
    #seasonal_order = (7, 0, 1, 7),
    measurement_error = True,          # Для адекватной обработки
пропущенных данных
    enforce_stationarity = False,      # Для работы
    simple_differencing = True,
    enforce_invertibility = False)    # получения прогнозирования
results = mod.fit(dispatch = False)
preds = results.predict(start = start, end = finish)

for k in range(0,len(preds)) :
    if (preds[k] < 0) :
        preds[k] = 0

for k in range(0,7) :
    ddate = start + datetime.timedelta(days=k)
    errors      =      errors.append({'ID':id[i],      'DATE':ddate,
'FACT':vivot['SUMM'][k],  'PRED':preds[k],  'real':datasumm['real'][k]},
ignore_index=True) #=[id[i], ddate, vivot['SUMM'][k], preds[k]]

```

```
    print(f'STADIA {j}/14')
print(f'BANKOMAT # {i}')

errors.to_csv('factpred1.csv', sep=',')
# plt.plot(vivod)
# plt.plot(preds)
# plt.show()
```