МИНОБРНАУКИ РОССИИ

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

Кафедра теории функций и стохастического анализа

ИДЕНТИФИКАЦИЯ ВРЕМЕННЫХ РЯДОВ И ВЫБОР ОПТИМАЛЬНОЙ МОДЕЛИ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

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

механико-математического факультета
Барышникова Максима Сергеевича

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

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

Саратов 2024

ВВЕДЕНИЕ

В современном мире, где данные становятся все более ценным ресурсом, анализ временных рядов играет ключевую роль в прогнозировании будущих событий и принятии обоснованных решений. Временные ряды, представляющие собой последовательности измерений, собранных через определенный интервал времени, используются во многих областях, от экономики до медицины, для понимания тенденций и прогнозирования будущих изменений. Однако, выбор подходящей модели для анализа этих данных является не менее важным аспектом, поскольку он напрямую влияет на точность и надежность получаемых прогнозов.

Актуальность темы. Актуальность темы выпускной квалификационной работы «Идентификация временных рядов и выбор оптимальной модели» заключается в значимости анализа временных рядов для предсказания будущих событий на основе исторических данных. Временные ряды являются основой для прогнозирования в различных областях, включая экономику, финансы, климатологию и медицину. Определение оптимальной модели для анализа временных рядов позволяет улучшить точность прогнозов и принятие решений.

Также актуальность темы обусловлена постоянным развитием технологий и появлением новых методов анализа данных, которые могут значительно улучшить качество прогнозов и анализа временных рядов. Поэтому, исследование в этой области не только способствует развитию науки, но и имеет практическое применение в реальных условиях.

В целом, данная работа актуальна и важна для развития науки и технологий, а также для улучшения качества принятия решений в различных сферах деятельности.

Целью бакалаврской работы является исследование методов идентификации временных рядов и выбора оптимальной модели для прогнозирования и анализа динамики сложных реальных систем и процессов.

Объект исследования – временные ряды в компьютерных системах, которые могут быть представлены различными типами данных, такими как сетевой трафик, данные информационной системы и другие.

Предмет исследования – анализ и выбор методов идентификации временных рядов для прогнозирования и анализа динамики сложных реальных систем и процессов.

Для достижения поставленных целей в работе необходимо решить следующие **задачи**:

– определить основные понятия, необходимые для описания временных рядов;

– разработать методологию исследования временных рядов и выбора оптимальной модели;

– проанализировать существующие методы и модели прогнозирования временных рядов;

– выбрать наиболее подходящие методы и модели для решения поставленной задачи;

– провести экспериментальное исследование на реальных данных с использованием выбранных методов и моделей;

– оценить качество полученных результатов и сравнить их с результатами, полученными другими методами;

– разработать рекомендации по выбору оптимальной модели прогнозирования для временных рядов.

Практическая значимость проводимого исследования состоит в том, что полученные результаты и разработанная методология могут быть использованы для выбора оптимальной модели прогнозирования временных рядов в различных областях, таких как финансы, экономика, прогнозирование погоды и другие. Это позволит повысить точность и надежность прогнозирования, а также снизить затраты на сбор и анализ данных.

Структура и содержание бакалаврской работы. Работа состоит из введения, 3 разделов, заключения, списка использованных источников, содержащего 20 наименований и двух приложений. Общий объем работы составляет 59 страницы.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы работы, формулируется цель работы и решаемые задачи, отмечается практическая значимость полученных результатов.

В **первом** разделе приводятся основные понятия теории временных рядов.

Второй раздел посвящен стационарным стохастическим моделям.

Под стационарной случайной последовательностью понимается последовательность $\{X_t\}_{t\geq1}$, в которой члены представляют собой случайные величины:

1. С постоянным математическим ожиданием:

$$E\{X_t\} = m = const;$$

2. С ковариацией, зависящей только от разности их номеров:

$$Cov(X_t, X_{t+k}) = C_{XX}(k), \quad \forall t,$$

где параметр t трактуется как дискретное время, а последовательность $\{X_t\}_{t\geq1}$ называется временным рядом.

Стационарные временные ряды используются при построении моделей, которые применяются для описания поведения случайных возмущений (остатков), т. е. того, что остается после исключения из временного ряда неслучайной составляющей. В качестве основы для построения большинства моделей стационарных временных рядов используются общие линейные модели (ОЛМ). В ОЛМ предполагается, что белый шум ν_t можно трансформировать в динамический ряд X_t при помощи линейного фильтра (некоторого линейного преобразования).

Выражение (1) представляет собой первую (прямую) форму ОЛМ:

$$\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \nu_{t-j} = \psi(B)\nu_t, \tag{1}$$

которая позволяет выразить значения динамического ряда ε_t через взвешенную сумму настоящего и прошлых значений белого шума ν.

Если оператор $\psi(B)$ обратим, то ОЛМ можно представить в виде обращенной формы – взвешенной суммы настоящего и прошлых значений процесса $\{\varepsilon_t\}_{t\geq0}$. Обозначим обратный оператор через $\pi(B)$:

$$\pi(B) = \psi^{-1}(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots = 1 - \sum_{j=1}^{\infty} \pi_j B^j,$$

если он существует, то процесс $\{\varepsilon_t\}_{t\geq 1}$ называется *обратимым.* Для обратимого процесса из представления (1) следует:

$$\nu_t = \psi^{-1}(B)\varepsilon_t = \pi(B)\varepsilon_t \left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right)\varepsilon_t. \tag{2}$$

Спецификация (2) – вторая (обращенная) форма ОЛМ, в которой текущее значение белого шума выражено в виде линейной комбинации настоящего и прошлых значений временного ряда $\{\varepsilon_t\}_{t\geq 0}$.

Модель авторегрессии AR(p). Пусть в спецификации (2), представляющей процесс $\{\varepsilon_t\}_{t\geq 1}$ в форме регрессии текущего значения на прошлые значения процесса и значения импульса ν_t:

$$\nu_t = \pi(B)\varepsilon_t = \left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right)\varepsilon_t = \varepsilon_t - \pi_1\varepsilon_{t-1} - \pi_2\varepsilon_{t-2} - \ldots,$$

нужно учесть, что только первые p параметров отличны от нуля.

Обозначим конечный вектор параметров через $\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_p)^T$, тогда спецификация модели примет вид:

$$\varepsilon_t = \Phi_1\varepsilon_{t-1} + \Phi_2\varepsilon_{t-2} + \ldots + \Phi_3\varepsilon_{t-p} + \nu_t.$$

Так как в этом уравнении текущее значение временного ряда определяется как линейная комбинация его предыдущих значений с максимальным лагом p (параметр модели) и текущей возмущающей переменной, то есть представляет собой регрессию на прошлые значения, то данная модель называется *авторегрессионной.* Операторная форма этой модели записывается следующим образом:

$$\Phi(B)\varepsilon_t = \nu_t \tag{3}$$

где оператор авторегрессии

$$\Phi(B) = 1 - \Phi_1 B - \Phi_2 B^2 - \ldots - \Phi_p B^p.$$

Чтобы процесс авторегрессии AR(P) порядка p был *стационарным*, необходимо наложить определённые ограничения на параметры модели. Согласно теории ОЛМ, корни характеристического уравнения $\Phi(B) = 0$ должны быть по модулю больше 1, то есть лежать вне единичного круга.

Прогнозирование в модели AR(p). Оптимальный предиктор определяется как условное математическое ожидание. Если доступны $n \geq p$ предыдущих значений процесса, формула для оптимального предиктора на один шаг вперёд будет иметь следующий вид:

$$\varepsilon^*_{t-1}(1) = E\{\varepsilon_t | \varepsilon_\tau, \tau \leq t-1\} = \Phi_1 \varepsilon_{t-1} + \Phi_2 \varepsilon_{t-2} + \ldots + \Phi_p \varepsilon_{t-p}.$$

При вычислении прогноза на два шага вперед используется прогноз $\varepsilon^*_{t-1}(1)$:

$$\varepsilon_{t+1} = \Phi_1 \varepsilon_1 + \Phi_2 \varepsilon_{t-1} + \ldots + \Phi_p \varepsilon_{t+1-p} + \nu_{t+1},$$

$$\varepsilon^*_{t-1}(2) = E\{\varepsilon_{t+1} | \varepsilon_\tau, \tau \leq t-1\} = \Phi_1 \varepsilon_1 + \Phi_2 \varepsilon_{t-1} + \ldots + \Phi_p \varepsilon_{t+1-p} =$$

$$= \Phi_1 \varepsilon^*_{t-1}(1) + \Phi_2 \varepsilon_{t-1} + \ldots + \Phi_p \varepsilon_{t+1-p}, \quad \text{и т.д.}$$

Модель скользящего среднего MA(q). Автокорреляционная функция процесса скользящего среднего порядка q обращается в ноль после задержки q. Это особое свойство отличает данный процесс от других. Если значение автокорреляций ρ_k для $k = 1, \ldots, q$ известны, то уравнения:

$$\rho_k = \begin{cases} \dfrac{\theta_k + \theta_1 \theta_{k+1} + \ldots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2}, & k = 1, 2, \ldots, q, \\ 0, \quad k > q \end{cases}$$

можно решить относительно весовых коэффициентов модели $\theta = (\theta_1, \theta_2, \ldots, \theta_q)^T$.

Прогнозирование в модели MA(q). Оптимальный предиктор модели MA(q) имеет вид:

$$\varepsilon_t^*(k) = \begin{cases} -\sum_{j=k}^{q} \theta_j \nu_{t+k-j} = -(\theta_k \nu_t + \ldots + \theta_q \nu_{t+k-q}) \text{ при } k \leq q, \\ 0 \quad \text{при } k > q. \end{cases}$$

Прогноз возможен только на ограниченное количество q шагов вперед, с определенным уровнем среднеквадратической ошибки предиктора

$$Var(e_t(k)) = \sigma_\nu^2 \sum_{j=0}^{k-1} \theta_j^2 = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_{k-1}^2)\sigma_\nu^2, \quad \theta_0 = 1.$$

В **третьем** разделе диплома проведен анализ временного ряда с использованием различных статистических тестов. В частности, были проведены исследования математического ожидания, дисперсии, коэффициентов автокорреляции, а также применены тесты Манна-Уитни и Сиджела-Тьюки. Эти методы позволили оценить статистические характеристики временного ряда, включая его стационарность, а также выявить возможные аномалии и тенденции в данных.

За основу исследований был взят искусственно смоделированный временной ряд. Это позволило точно контролировать характеристики данных и проводить анализ в условиях, максимально приближенных к реальным, но при этом обеспечивая возможность проверки гипотез и методов анализа на искусственных данных.

В ходе исследования все тесты были подтверждены, что означает, что статистические характеристики временного ряда, включая его математическое ожидание, дисперсию, коэффициенты автокорреляции, а также стационарность, были обнаружены и подтверждены с использованием выбранных статистических методов. Это подтверждение важно, так как оно указывает на то, что используемые методы анализа были применимы и эффективны для анализа данных. Однако, стоит помнить, что статистическая значимость не всегда означает практическую значимость результатов исследования. Важно также оценивать практическую значимость результатов, чтобы понять, насколько они могут быть применены в реальных ситуациях или для принятия решений.

В следующем этапе исследования были проведены аналитические работы с данными акций компаний Яндекс, Газпром, Татнефть и Сбербанк, взятые за 2014 год и преобразованные в временные ряды. Этот процесс включал в себя несколько ключевых этапов, каждый из которых был критически важным для понимания динамики акций и их будущего развития.

Первым шагом было проведение проверки стационарности и сезонности данных. Стационарность данных является важным условием для многих статистических моделей временных рядов, включая ARIMA и SARIMA, которые используются для прогнозирования будущих значений временного ряда. Сезонность, в свою очередь, отражает периодические колебания в данных, которые могут быть вызваны различными факторами, такими как экономические циклы или сезонные события. Проверка стационарности и сезонности позволила определить, какие модели временных рядов могут быть наиболее подходящими для анализа данных акций.

Следующим этапом был разработка моделей авторегрессии для последующего анализа. Авторегрессионные модели используются для анализа временных рядов, где текущее значение переменной зависит от ее предыдущих значений. Разработка этих моделей позволила оценить влияние прошлых цен акций на их будущие значения, что является ключевым для прогнозирования будущих тенденций в ценах акций.

В завершение работы были выполнены прогнозы на основе собранных данных. Этот этап включал в себя использование разработанных моделей для генерации прогнозов на будущее развитие акций выбранных компаний. Прогнозы были основаны на анализе статистических характеристик данных и использовании моделей авторегрессии для предсказания будущих значений акций.

Такой подход позволил оценить динамику акций выбранных компаний и предсказать их будущее развитие. Результаты исследования могут быть использованы для принятия инвестиционных решений, анализа рыночных тенденций и разработки стратегий инвестирования. Важно отметить, что точность прогнозов зависит от множества факторов, включая качество данных, выбранную модель временного ряда и актуальность анализа.

В **заключении** приведены результаты бакалаврской работы.

Основные результаты

1. Определены основные понятия, необходимые для описания временных рядов. Изучены спецификации моделей временного ряда, моделей тренда и выделения сезонной составляющей.
2. Определены основные понятия стационарных стохастических моделей. Изучены модели авторегрессии и скользящего среднего.
3. Изучено прогнозировние в моделях авторегрессии и скользящего среднего.
4. В ходе исследования была разработана и смоделирована модель авторегрессии временного ряда. Созданная программа выполняет тестирование математического ожидания, дисперсии и автокорреляционных коэффициентов, а также проводит анализ с использованием тестов Манна-Уитни и Сиджела-Тьюки.
5. Были получены и проанализированы данные по акциям нескольких компаний. На основе обработанной информации были созданы модели авторегрессии и выполнены прогнозы по дальнейшему развитию акций выбранных компаний.