

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение

высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра « Прикладная математика и информатика»
(наименование кафедры)

01.03.02 Прикладная математика и информатика
(код и наименование направления подготовки, специальности)

Системное программирование и компьютерные технологии
(направленность (профиль)/специализация)

БАКАЛАВРСКАЯ РАБОТА

на тему Анализ и прогнозирование показателей потребления ресурсов (на примере ТСЖ)

| | | |
|--------------|-----------------------------------|------------------------|
| Студент | М.А. Хлебникова (И.О. Фамилия) | _____ (личная подпись) |
| Руководитель | С.В. Мкртычев (И.О. Фамилия) | _____ (личная подпись) |
| Консультанты | М.А. Четаева (И.О. Фамилия) | _____ (личная подпись) |

Допустить к защите

Заведующий кафедрой к.т.н., доцент, А.В. Очеповский _____
(ученая степень, звание, И.О. Фамилия) (личная подпись)

« _____ » _____ 20 _____ г.

Тольятти 2018

АННОТАЦИЯ

Название дипломной работы: «Анализ и прогнозирование показателей потребления ресурсов (на примере ТСЖ)».

Актуальность данной работы заключается в необходимости выбора метода анализа и прогнозирования и дальнейшего его применения для данных ТСЖ, что позволит вычислить прогноз для дальнейших периодов.

Объектом дипломной работы бакалавра являются показатели потребления ресурсов ТСЖ.

Предметом бакалаврской диссертации является прогнозирование показателей потребления ресурсов ТСЖ.

Целью данной работы является анализ данных ТСЖ с помощью метода прогнозирования, который является наиболее подходящим для такого типа данных, как данные ТСЖ, вычисление прогноза, опираясь на исходные данные и полученный результат прогнозирования.

Практическая значимость дипломной работы бакалавра заключается в анализе данных ТСЖ, вычислении прогноза.

Дипломная работа бакалавра состоит из пояснительной записки на 46 страниц, введения на 2 страницы, включая 18 рисунков, 4 таблицы, списка 21 источников, в том числе 5 источников на иностранном языке.

Первая глава посвящена изучению характеристик данных ТСЖ «Фрунзенское-1», изучению методов прогнозирования, выбору наиболее подходящего метода прогнозирования для данных ТСЖ.

Вторая глава посвящена выбору математической модели для данных ТСЖ, построению данной модели.

Третья глава посвящена проверке на адекватность выбранной математической модели, построению прогноза.

В заключении подводятся основные итоги и выводы по работе.

Результатом работы является анализ данных ТСЖ, вычисленный прогноз.

ABSTRACT

The title of the bachelor's thesis is Analysing and forecasting resource consumption indicators (by the example of homeowners association).

The relevance of the bachelor's thesis lies in the need to select the method of analysis and forecasting, and its further application to the homeowners association data, which will allow calculating the forecast for future periods.

The object of the study was the indicators of resource consumption of the homeowners association.

The subject of the study was the forecasting indicators of resource consumption of the homeowners association.

The aim of the work was to analyze the data of homeowners association using the forecasting method, which is the most suitable for such type of homeowners association data, to calculate the forecast on the basis of the source data and to get the forecast result.

The practical importance of the bachelor's thesis lies in the analysis of the homeowners association data, calculation of the forecast.

The bachelor's thesis consists of an explanatory note on 46 pages, introduction on 2 pages, including 18 figures, 4 tables, the list of 21 references including 5 foreign references.

The first chapter is devoted to the study of characteristics of the data of homeowners association Frunzenskoe-1, to the study of forecasting methods, the choice of the most appropriate forecasting method for homeowners association data.

The second chapter is devoted to the selection of a mathematical model for homeowners association data and to building this model.

The third chapter is devoted to checking the proposed model for adequacy and to calculating the forecast.

In conclusion, the main results and conclusions of the work were presented.

The result of the work is the analysis of the homeowners association data, calculated forecast.

СОДЕРЖАНИЕ

| | |
|--|----|
| ВВЕДЕНИЕ | 5 |
| 1 Анализ данных ТСЖ и выбор метода их прогнозирования..... | 7 |
| 1.1 Характеристика исходных данных ТСЖ..... | 7 |
| 1.2 Методы прогнозирования..... | 8 |
| 1.3 Основные эконометрические методы прогнозирования | 11 |
| 1.4 Выбор эконометрического метода прогнозирования | 13 |
| 2 Построение временных рядов по данным электроэнергетики и воды ТСЖ и расчет их числовых характеристик | 20 |
| 2.1 Проверка статистических гипотез о свойствах временного ряда | 20 |
| 2.1.1 Проверка гипотезы о наличии аномальных наблюдений во временных рядах значений по данным электроэнергетики и воды ТСЖ | 20 |
| 2.1.2 Проверка гипотез о наличии неслучайной составляющей временных рядов значений по данным электроэнергетики и воды ТСЖ | 22 |
| 2.2 Выделение неслучайной составляющей временных рядов по данным электроэнергетики и воды ТСЖ..... | 26 |
| 2.3 Выделение тригонометрической составляющей временных рядов по данным электроэнергетики и воды ТСЖ..... | 32 |
| 3 Проверка адекватности и качества построенных моделей временных рядов по данным электроэнергетики и воды ТСЖ и прогнозирование их трендовой составляющей..... | 36 |
| 3.1 Проверка математического ожидания ряда остатков временных рядов по данным электроэнергетики ТСЖ..... | 36 |
| 3.2 Проверка случайности ряда остатков по данным электроэнергетики ТСЖ | 37 |
| 3.3 Проверка независимости значений ряда остатков по данным электроэнергетики ТСЖ..... | 39 |
| 3.4 Тест на наличие автокорреляции временного ряда по данным воды | 40 |
| 3.5 Прогнозирование трендовой составляющей временных рядов по данным электроэнергетики и воды ТСЖ..... | 42 |
| ЗАКЛЮЧЕНИЕ | 43 |
| СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ..... | 44 |

ВВЕДЕНИЕ

Одним из видов хозяйственной деятельности ТСЖ, является предоставление клиентам таких ресурсов, как вода и электроэнергия. Данному предприятию важно знать, какой расход ресурсов, и какая стоимость за них будет значиться в последующих периодах, потому что каждое предприятие должно планировать свое будущее. Именно по этой причине анализ данных и дальнейшее прогнозирование является важной составляющей деятельности ТСЖ.

Актуальность данной работы заключается в необходимости анализа данных и прогнозирования показателей потребления ресурсов ТСЖ для повышения эффективности его деятельности.

Объект исследования: показатели потребления ресурсов ТСЖ.

Предмет исследования: прогнозирование показателей потребления ресурсов ТСЖ.

Целью данной работы является анализ данных и прогнозирование показателей потребления ресурсов ТСЖ.

Для достижения поставленной цели необходимо решить следующие **задачи:**

- Проанализировать необходимую научную и учебно-методическую литературу.
- Проанализировать деятельность ТСЖ, а также исходные данные этой организации.
- Изучить методы прогнозирования.
- Исходя из характеристики данных ТСЖ, выбрать наиболее подходящий для них метод прогнозирования.
- Провести анализ данных ТСЖ выбранным методом и вычислить прогноз.

Методы исследования: статистический анализ, методы прогнозирования.

Практическая значимость выпускной квалификационной работы заключается в анализе данных ТСЖ, вычислении прогноза.

Соответствие содержания бакалаврской работы профессиональным компетенциям по видам профессиональной деятельности выпускника:

- научно-исследовательская деятельность:

- способность собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям (ПК-1);

- способностью понимать, совершенствовать и применять современный математический аппарат (ПК-2).

В первой главе совершается исследование предметной области, описываются существующие методы прогнозирования, приводится обоснование и описание выбранного метода для анализа и прогнозирования данных ТСЖ.

Во второй главе совершается выбор вида модели временного ряда, строится выбранная модель по данным ТСЖ, вычисляется прогноз по данной модели.

В третьей главе совершается проверка на адекватность, подбор метода прогнозирования, который наилучшим образом прогнозирует данные в случае, если модель не прошла проверку на адекватность для какого-либо вида ресурсов (электроэнергия, вода).

В заключении подводятся итоги исследования, формируются окончательные выводы по изучаемой тематике.

1 Анализ данных ТСЖ и выбор метода их прогнозирования

1.1 Характеристика исходных данных ТСЖ

Компания ТСЖ «Фрунзенское-1» зарегистрирована 29 февраля 2008 года, регистратор — Межрайонная ИФНС России №2 по Самарской области.

Полное наименование — товарищество собственников жилья «Фрунзенское-1». Компания находится по адресу: 445028, Самарская область, г. Тольятти, бульвар Приморский, д. 40, кв. 44.

Таблица с исходными данными демонстрирует фактические цены на электроэнергию и воду за 2016 год. Источник данной информации – ТСЖ «Фрунзенское-1».

Стоимость электроэнергии включает: электроэнергию в местах общего пользования (дневной и ночной тарифы), электроэнергию личного использования (дневной и ночной тарифы), электроэнергию нежилых помещений.

Стоимость воды включает: теплоэнергию, химочищенную воду, хозяйственные бытовые сточные воды, хозпитьевую воду.

Ниже приведены данные цен (руб.) на расходы за электроэнергию и воду за 2016 год, распределенные по месяцам (таблица 1.1).

Таблица 1.1 – Данные ТСЖ

| Период | Электроэнергия | Вода |
|---------|----------------|-----------|
| Январь | 97426,66 | 242018,05 |
| Февраль | 95537,21 | 311829,49 |
| Март | 97424,18 | 244582,97 |
| Апрель | 93480,86 | 222754,32 |
| Май | 89851,9 | 105789,31 |
| Июнь | 83230,29 | 101497,29 |
| Июль | 89721,39 | 87391,14 |

Продолжение таблицы 1.1

| | | |
|----------|-----------|-----------|
| Август | 99873,72 | 96439,49 |
| Сентябрь | 106099,36 | 113373,62 |
| Октябрь | 100289,27 | 141178,60 |
| Ноябрь | 96951,42 | 251224,50 |
| Декабрь | 92748,95 | 302337,68 |

На основании представленных данных можно сделать следующие выводы:

- заметно определенное повышение и понижение цен на электроэнергию, а также воду в зависимости от периода;
- более ярко выражена разница по воде, так как в стоимость входит теплоэнергия;
- цены в холодные периоды, такие как январь, февраль, март, апрель, ноябрь и декабрь более высокие, чем в остальные периоды. А значит можно предположить, что в предыдущих и последующих годах будет наблюдаться аналогичная динамика цен.

Однако для более точных выводов необходимо провести анализ, а также использовать специальные методы прогнозирования.

1.2 Методы прогнозирования

Прогнозирование – это процесс научного выявления состояния и предполагаемых путей развития процессов и явлений, основой которого является система установленных причинных и следственных связей, а также закономерностей [8].

Ниже приведены виды методов и моделей прогнозирования (рисунок 1.1).

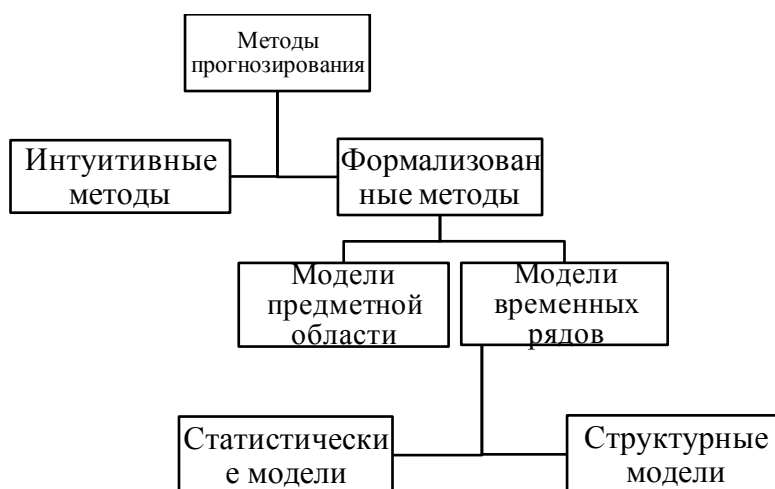


Рисунок 1.1 – Методы прогнозирования

Чтобы дать необходимую теоретическую основу для исследования, дана краткая характеристика каждого из представленных видов прогнозов.

Суть интуитивного метода прогнозирования заключается в проведении экспертами интуитивного и логического анализа проблемы, что позволяет получить правильное ее решение.

В основе формализованные методов прогнозирования лежит построении прогнозов формальными средствами математической теории, что гораздо повышает достоверность и точность прогнозов, при этом сократив сроки их выполнения.

Модели предметной области представляют собой математические модели прогнозирования, для построения которых используются законы предметной области.

Модели временных рядов представляют собой математические модели прогнозирования, которые устанавливают, есть ли зависимость между историческими значениями внутри самого процесса, чтобы в дальнейшем вычислить прогноз будущих значений. Данные модели универсальны для

сфер разного типа, так как их общий вид не зависит от природы временного ряда.

В статистических моделях определение зависимости между будущими и прошлыми значениями представляется в виде уравнения. Статистические модели включают в себя:

- 1) модели регрессии (линейной и нелинейной);
- 2) авторегрессионной модели;
- 3) модели с экспоненциальным сглаживанием;
- 4) модель, строящаяся по выборке максимального подобия.

В структурных моделях определение зависимости будущих значений от прошлых значений дается в виде структуры. Чаще всего под этим подразумевается наличие некоторых правил переходов. Данный тип модели состоит из:

- 1) модели нейронных сетей;
- 2) модели, которые основаны на использовании цепи Маркова;
- 3) модели, которые используют классификационные и регрессионные деревья.

Основной задачей является получение информации о возможных состояниях данных ТСЖ в будущем и дальнейших путях достижения этих целей. В таком случае наилучшим методом прогнозирования будет являться регрессионный анализ (статистический метод)[13].

Данный метод используется с целью определения зависимости некоторой величины от другой величины, а может и нескольких других величин, что дает возможность установить изменения в среде учета влияний этих изменений на исследуемый показатель.

Так как данные ТСЖ являются экономическими показателями, значит, прогноз по ним более полным, если использовать экономические исследования[1]. Данные исследования лучше всего проводить, используя методы такой науки, как эконометрика.

Эконометрика представляет собой науку, в которой, опираясь на фактические данные, строятся и анализируются математические модели различных экономических процессов.

Для того, что выбрать наиболее подходящий эконометрический метод анализа и прогнозирования, необходимо рассмотреть все методы эконометрического исследования.

1.3 Основные эконометрические методы прогнозирования

Как показывает практика, значимостью эконометрики заключается в применении ее методов. Такое применение позволяет обнаружить существенные связи между явлениями, а также дать прогноз дальнейшего развития данного явления, который имеет обоснование в некоторых заданных условиях, численно оценить и проверить экономические последствия, возникшие при принятии каких-либо важных решениях управления.

Чтобы выбрать наиболее подходящий метод эконометрического исследования, необходимо рассмотреть и описать основные эконометрические методы[9].

1. Первым таким методом является парный регрессионный анализ. Между экономическими переменными существуют лишь статистические и корреляционные зависимости.

Выражая переменные через X , Y , получим зависимость такого вида

$$M Y \setminus X = f x$$

Данная зависимость будет иметь название функции регрессии Y на X . X имеет название независимой переменной, которая также является объясняющей, Y – зависимая переменная или же объясняемая. В таком случае, рассматривая две случайные величины, используют парную регрессию.

2. Нелинейная регрессия. Не все экономические зависимости являются линейными, поэтому их моделирование линейными уравнениями регрессии не является целесообразным. Поэтому вследствие этого будет

использоваться такой вид регрессионного анализа, как нелинейная регрессия, в котором экспериментальные данные моделируются с помощью функции, которая представляет собой нелинейную комбинацию параметров модели и зависит от одной и более независимых переменных. Данные аппроксимируются методом последовательных приближений[12].

3. Множественная регрессия. При таких задачах, как решение проблем спроса, изучение функций издержек производства, доходность акций, в макроэкономических расчетах применяют множественную регрессию.

Уравнение множественной представляется в виде

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_m + e,$$

где y – зависимая переменная или объясняемая переменная, x_1, \dots, x_m – независимые факторы или объясняющие переменные; e – случайное возмещение, или стохастическая переменная, которая включает влияние факторов, которые не учтены в модели.

4. Фиктивные переменные. Если при построении модели требуется ввести качественные переменные, то их необходимо перевести в количественные. Переменные, имеющие такую конструкцию, носят название фиктивных переменных, количественным образом описывающих качественные признаки.

5. Системы экономических уравнений. Использование системы уравнений возникает по причине того, что некоторые процессы моделируются сразу несколькими уравнениями, которые содержат и собственные и повторяющиеся переменные одновременно. [15].

6. Временные ряды. Данный метод представляется в виде совокупности значений какого-либо показателя за некоторые моменты времени, следующие друг за другом. Каждый уровень y в t формируется тогда, когда на него влияют кратковременные, длительные и случайные факторы. Наиболее значимое влияние на явление оказывают те факторы,

которые длительны и постоянны. Данные факторы формируют основную тенденцию ряда – тренд $T t$. А такие факторы, как кратковременные и периодичные, образуют формирование сезонных колебаний ряда $S t$. Случайные факторы – это отражение случайных изменений уровней ряда.

1.4 Выбор эконометрического метода прогнозирования

Данные показателей потребления ресурсов, таких как электроэнергия и вода являются набором числовых данных в течение последовательных периодов времени. Для данных такого типа больше всего подходит метод анализа временных рядов, с помощью которого можно предсказать значение числовой переменной, опираясь на ее прошлые и настоящие значения.

Основу анализа временных рядов можно описать следующим предположением: факторы, которые влияют на объект исследования в настоящем и прошлом, будут оказывать на него влияние и в будущем. Значит, основные задачи анализа временных рядов состоят в выделении и идентификации факторов, которые имеют значение для прогнозирования. Чтобы решить эту задачу, были разработаны многие математические модели, которые предназначены для исследования колебаний компонентов, которые входят в модель временного ряда [6].

Вспомнив предыдущий раздел, еще раз уточним, что каждый уровень $y t$ временного ряда формируется под влиянием кратковременных, длительных и случайных факторов.

Основные определения, которые формируются факторами, описанными выше:

- $T t$ – основная тенденция ряда;
- $S t$ – сезонные колебания ряда;
- $C t$ – циклическая компонента;
- $E t$ – случайные факторы.

Уровни временного ряда можно выразить в виде суммы систематической, которая также может быть регулярной или детерминированной и случайной, которая от времени не зависима, составляющих. Получить регулярную составляющую можно сложив тренд, циклическую и сезонную компоненты. Однако, стоит отметить, что бывают случаи, когда данная составляющая может и не включать все три компоненты одновременно.

Существует два вида факторов, действие которых формирует случайную (нерегулярную) компоненту:

- 1) факторы резкого внезапного действия;
- 2) факторы, являющиеся текущими.

Факторы, которые носят характер резкого внезапного действия, такие как, например, стихийные бедствия, отклоняются более значительно, чем случайные колебания, которые также называются катастрофическими.

Факторы текущего периода являются провокаторами случайных колебаний, которые являются результатом действия большого числа причин побочного характера. Ощутимо будет лишь суммарное воздействие этих факторов, влияние каждого по отдельности будет незначительным.

Разложение ряда на три составляющие, такие как: тренд, сезонную компоненту и нерегулярную составляющую является сутью корректировки временного ряда[11].

Если период колебаний длится не более одного года, то они носят название сезонных колебаний. Основной причиной их возникновения считаются природно-климатические условия. На примере ТСЖ это можно объяснить колебанием цен на электроэнергию и воду. Год за годом наблюдается понижение цен в теплые времена года и последующее повышение цен в холодные времена года, связанное с тем, что в холодные периоды люди проводят большую часть времени дома, а значит идет большее потребление электроэнергии и горячей воды, а также учитывается

отопление. Таким образом, в колебаниях показаний потребления ресурсов прослеживается устойчивая годовая периодичность.

Существует два основных способа взаимодействия компонентов временных рядов – аддитивно и мультипликативно:

- Аддитивная модель: $Y = T + S + E$;
- Мультипликативная модель: $Y = T \times C \times S \times E$;
- Модель смешанного типа: $Y_t = T \times C \times S + E$.

В зависимости от анализа структуры сезонных колебаний решается, какая из этих трех моделей будет использоваться. Если в аддитивной модели значения сезонной компоненты для различных циклов постоянны, то данная модель строится только тогда, когда амплитуда колебаний практически неизменна. Мультипликативная модель временного ряда, при которой уровни ряда имеют зависимость от значений сезонной компоненты, строится тогда, когда амплитуда сезонных колебаний имеет убывающий или возрастающий характер.

Рассматривая на примере ТСЖ потребление ресурсов, таких как электроэнергия и вода, можно сразу определить, что будет строиться аддитивная модель, так как амплитуда колебаний каждого года приблизительно постоянна.

Построение аддитивной и мультипликативной моделей проводится при расчете значений T , S и E для каждого уровня ряда

Оценив и удалив тренд, оценив и удалив сезонную компоненту, моделируя циклическую компоненту, конструируя прогнозную модель и выполняя прогноз можно выполнить прогнозирование с помощью компонентного анализа.

После того, как вычислен прогноз, необходимо проверить полученную модель на адекватность, или другими словами, проверить соответствует ли модель исследуемому объекту или процессу. Модель временного ряда можно

считать адекватной, если она верно отражает систематические компоненты временного ряда.

Важным примечанием является то, что анализ тренда существенно упрощается, если тренд является монотонным, то есть либо устойчиво возрастает, либо устойчиво убывает. Если во временных рядах содержится значительная ошибка, то в данном случае выделение тренда осуществляется сглаживанием.

Сглаживание всегда включает некоторый способ локального усреднения данных, результатом которого является то, что несистематические компоненты взаимно погашают друг друга. Наиболее общим методом сглаживания можно считать скользящее среднее, в котором происходит замена каждого члена ряда на простое или взвешенное среднее n соседних членов.

В случаях, когда ошибка измерения очень велика, прибегают к методу сглаживания методом наименьших квадратов, взвешенных относительно расстояния или методу отрицательного экспоненциально взвешенного сглаживания[8]. Данные методы устраняют шум и преобразуют данные в относительно гладкую кривую. Ряды, в которых содержится незначительное количество наблюдений и систематическое расположение точек, сглаживаются с помощью бикубических сплайнов.

В большинстве случаев монотонные временные ряды приближаются, линейные функции. Однако, если же имеет место быть явная монотонная нелинейная компонента, то в первую очередь необходимо преобразовать данные, чтобы устранить нелинейность. Наиболее часто применимы экспоненциальное, логарифмическое или полиномиальное преобразования данных.

Если взять разность соответствующего порядка, то периодическая составляющая для данного лага k может быть удалена. Отсюда следует, что из каждого i -го элемента ряда вычитается $(i-k)$ -й элемент. Значит, можно выявить скрытые периодические составляющие ряда. Удаление некоторых

автокорреляций может повлечь за собой изменение других автокорреляций, а также делает некоторые другие сезонные составляющие более заметными, так как автокорреляции на последовательных лагах зависимы.

Формализованные методы прогнозирования опираются на математическую теорию, которая делает прогнозы более точными и достоверными, а также существенно сокращает сроки их выполнения, позволяет обрабатывать информацию и оценивать результат.

Суть метода прогнозной экстраполяции заключается в приложении определенной тенденции развития экономического процесса для некоторого базисного периода к прогнозируемому периоду. При использовании данного метода необходимо иметь сведения об устойчивости тенденций развития объекта, срок которого, примерно в 2-3 раза превышает срок прогнозирования. Долгосрочная тенденция изменения экономических показателей представляет собой тренд.

Операцию экстраполяции в общей форме можно представить следующим уравнением:

$$Y_{i+L} = F \cdot Y_i \times L ,$$

где Y_{i+L} – экстраполируемое значение уровня;

L – период упреждения;

Y_i – уровень, который принимается за базу экстраполяции.

Самая простая экстраполяция проводится на основании средних характеристик ряда, которые представляют собой средний уровень, средний абсолютный прирост и средний темпа роста.

Метод скользящих средних является одним из наиболее простых из всех методов. С помощью данного метода происходит осуществление механического выравнивания временного ряда. Метод основывается на том, что фактические уровни ряда заменяются расчетными средними, в которых колебания погашаются [10].

Экстраполяция тренда имеет место, если уровни ряда зависимы от фактора времени t , в данном случае зависимость представляется в виде:

$$y_t = f \ t$$

Модель стационарного процесса, в которой значение показателя представляется как линейная комбинация конечного числа, предшествующих этому показателю и аддитивной случайной составляющей значений, является моделью авторегрессии.

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t ,$$

где α – константа;

β – параметр уравнения;

ε_t – случайная компонента.

Для выполнения краткосрочного прогноза может быть использован метод экспоненциального сглаживания.

Простое экспоненциальное сглаживание. Модель временного ряда в данном случае имеет следующий вид:

$$X_t = b + \varepsilon_t,$$

где b – константа;

ε – случайная ошибка.

Экспоненциальное сглаживание устроено так, что выделение b нужно для того, чтобы использовать сглаживание скользящим средним, в котором последним наблюдениям присваиваются большие веса, чем предпоследним, предпоследним большие веса, чем пред-предпоследним и т.д.

Формула простого экспоненциального сглаживания имеет следующий вид:

$$S_t = \alpha \times y_t + (1 - \alpha) \times S_{t-1},$$

Где S_t – экспоненциальная средняя (сглаженное значение уровня ряда) на момент t ;

α – вес текущего наблюдения при расчете экспоненциальной средней;

y_t – фактический уровень динамического ряда в момент времени t ;

S_{t-1} – экспоненциальная средняя предыдущего периода.

Обычно на практике α принимает значение в пределах от 0,1 до 0,3. Редкое исключение, когда значение превышает 0,5.

Данный метод применяют, прежде всего, при постоянном объеме потребления $\alpha = (0,1 - 0,3)$. При изменении структуры потребления, например, с учетом сезонных колебаний, $\alpha = (0,3 - 0,5)$.

Кроме выше описанного существует и другой подход к определению параметра сглаживания, который предложен Брауном:

$$\alpha = 2 / n + 1 ,$$

где n – длина исходного ряда динамики.

Для прогнозирования показателей потребления электроэнергии и воды ТСЖ, используя метод анализа временных рядов, необходимо вычислить трендовую составляющую, которая по всем статистическим характеристикам хорошо подходит для прогнозирования данных ТСЖ.

Выводы по первой главе:

Была изучена характеристика данных ТСЖ, которые являются экономическими показателями. В результате изучения этих данных был сделан вывод, что наиболее полное и точное прогнозирование будет вычислено с помощью такой науки, как эконометрика, проводя экономические исследования.

Были рассмотрены эконометрические методы прогнозирования, что позволило выбрать наиболее подходящий для данных ТСЖ.

2 Построение временных рядов по данным электроэнергии и воды ТСЖ и расчет их числовых характеристик

2.1 Проверка статистических гипотез о свойствах временного ряда

2.1.1 Проверка гипотезы о наличии аномальных наблюдений во временных рядах значений по данным электроэнергии и воды ТСЖ

На стадии предварительного анализа временного ряда необходимо провести проверку гипотезы о наличии аномальных наблюдений. Чтобы обнаружить данные наблюдения используется метод Ирвина.

Для наблюдений временного ряда формируются следующие гипотезы:

H_0 : i -е наблюдение не является аномальным;

H_1 : i -е наблюдение аномальное.

Чтобы проверить данные гипотезы нужно вычислить значение критерия по следующей формуле:

$$I_i = \frac{|y_i - y_{i-1}|}{S_y},$$

где

$$S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

S_y представляет собой выборочное среднеквадратическое отклонение, которое вычислено по выборке n .

Если величина I_i попадет в критическую точку (превысит предельное значение), то гипотеза H_0 отвергается с вероятностью α ошибки первого рода и принимается гипотеза H_1 (y_i признается аномальным наблюдением).

Так как предельное значение $I_{\text{пр}}$ зависит от количества наблюдений, то приведем лишь те значения n , с помощью которых будет вычисляться $I_{\text{пр}}$.

Чтобы вычислить предельное значение для $n=36$ необходимо использовать табличные предельные значения для $n=30$ и $n=50$.

Ниже приведены предельные значения $I_{\text{пр}}$ для $n=30$ и $n=50$ (таблица 2.1):

Таблица 2.1 – Предельные значения

| n | $\alpha = 0,05$ | $\alpha = 0,01$ |
|----|-----------------|-----------------|
| 30 | 1,2 | 1,7 |
| 50 | 1,1 | 1,6 |

Все наблюдения, которые являются аномальными необходимо исключить из временного ряда, путем замены их расчетными значениями средней двух соседних значений аномального наблюдения.

Чтобы вычислить $I_{\text{пр}}$ для временных рядов по данным электроэнергии и воды, каждый из которых состоит из 36 наблюдений, выполним линейную интерполяцию первой ($n=30$) и второй ($n=50$) строк таблицы 2.1 для $\alpha = 0,05$. Получим следующее значение:

$$I_{\text{пр}} = \frac{1.1 - 1.2}{20} * 36 - 30 + 1.2 = 1.17$$

Ниже приведены данные потребления электроэнергии с вычисленными величинами I_i (рисунок 2.1).

| i | Y_i | I_i |
|-------|----------|-------------|
| 1 | 89563,95 | |
| 2 | 72991,81 | 0,984906774 |
| 3 | 79064,49 | 0,580771487 |
| 4 | 76236,26 | 0,270482924 |
| 5 | 71337,78 | 0,468474732 |
| 6 | 70969,79 | 0,035193075 |
| 7 | 67272,29 | 0,35361778 |
| 8 | 73111,18 | 0,558412873 |
| 9 | 72525,05 | 0,056054972 |
| 10 | 72337,91 | 0,017898112 |
| 11 | 72148,73 | 0,018091936 |
| 36 | 92748,95 | 0,401910549 |
| S_y | 10456,22 | |

Рисунок 2.1 – Определение аномальных наблюдений по данным электроэнергии

В данном случае нет таких значений I_i , которые бы превысили $I_{пр}$.
Значит, данные по электроэнергии остаются без изменения.

Ниже приведены данные потребления воды с вычисленными величинами I_i (рисунок 2.2).

| i | Y_i | I |
|-------|----------|-----------|
| 1 | 271864,4 | |
| 2 | 326039,3 | 0,6162253 |
| 3 | 214817,5 | 1,2651183 |
| 4 | 169215,6 | 0,5187098 |
| 15 | 234894 | 0,9879578 |
| 16 | 213438,7 | 0,2440494 |
| 17 | 104161 | 1,2430048 |
| 18 | 93697,71 | 0,1190166 |
| 28 | 222754,3 | 0,2482867 |
| 29 | 105789,3 | 1,3304459 |
| 30 | 101497,3 | 0,0488206 |
| 34 | 141178,6 | 0,3162743 |
| 35 | 251224,5 | 1,2517429 |
| 36 | 302337,7 | 0,5813988 |
| S_y | 87914,14 | |

Рисунок 2.2 – Определение аномальных наблюдений по данным воды

В данном случае аномальные наблюдения, где $I_i > I_{пр}$ присутствуют в 3,17,29 и 35 периодах. Для дальнейшего анализа необходимо заменить аномальные наблюдения вычисленными средними двух соседних периодов данных наблюдений .

Таким образом, 3, 17, 29 и 35 периоды будут иметь следующие значения соответственно: 247627.4, 153568.2, 162125.8, 221758.1.

2.1.2 Проверка гипотез о наличии неслучайной составляющей временных рядов значений по данным электроэнергии и воды ТСЖ

Наличие или отсутствие неслучайной составляющей (тренда) можно выявить, проверив гипотезы о постоянстве среднего значения временного ряда.

Выдвинем формулировки двух статистических гипотез:

$$H_0: M Y_i = const ;$$

$$H_1: M Y_i \neq const .$$

Чтобы проверить данные гипотезы, разобьем временной ряд на две части, которые по числу значений примерно равны. Каждая из этих частей имеет нормальное распределение и является самостоятельной выборочной совокупностью.

Временной ряд будет иметь тенденцию к тренду в том случае, если средние, вычисленные для каждой совокупности значительно отличаются друг от друга. При несущественном расхождении средних значений совокупностей можно утверждать, что временной ряд не имеет тенденции.

Проверяя данные гипотезы о равенстве средних двух нормально распределенных совокупностей, проверим, имеют ли временные тренд

Первую часть временного ряда будет обозначаться как $Y^{(1)}$, которая содержит n_1 наблюдений, $i = 1, 2, \dots, n_1$ наблюдений, тогда вторая часть – $Y^{(2)}$ будет содержать n_2 наблюдений $Y_i, i = n_1 + 1, \dots, n_1 + n_2$.

Чтобы проверить гипотезу о равенстве дисперсий, с помощью опции «Анализ данных» в программе Excel, вычислим двухвыборочный F-тест для дисперсий.

Ниже представлен результат выполнения двухвыборочного теста для данных по **электроэнергии** (рисунок 2.3).

| Двухвыборочный F-тест для дисперсии | | |
|-------------------------------------|--------------|--------------|
| | Переменная 1 | Переменная 2 |
| Среднее | 75296,53926 | 92516,02944 |
| Дисперсия | 22561209,98 | 45559026,38 |
| Наблюдения | 18 | 18 |
| df | 17 | 17 |
| F | 0,495208343 | |
| P(F<=f) одностороннее | 0,078740681 | |
| F критическое одностороннее | 0,440161596 | |

Рисунок 2.3 – Результаты выполнения F-теста для данных по электроэнергии

Критическая область представляет собой интервал $(X_{\text{кр},\alpha}, \infty)$, где точка $X_{\text{кр},\alpha}$ будет определяться из условия:

$$P(F > X_{\text{кр},\alpha}) = \alpha,$$

где α – вероятность ошибки первого рода.

Исходя из полученных значений в таблице со значениями по данным **электроэнергии** (рисунок 2.3), находим: $F = 0.4952$, $X_{\text{кр},\alpha} = 0.4402$. Можно заметить, что значение $F = 0.4952$ попадает в критическую область и нулевая гипотеза H_0 о равенстве дисперсий отклоняется.

Ниже представлен результат выполнения двухвыборочного теста для данных по **воде** (рисунок 2.4).

| Двухвыборочный F-тест для дисперсии | | |
|-------------------------------------|--------------|--------------|
| | Переменная 1 | Переменная 2 |
| Среднее | 202594,625 | 181097,9625 |
| Дисперсия | 13248560109 | 7106675868 |
| Наблюдения | 18 | 18 |
| df | 17 | 17 |
| F | 1,864241504 | |
| P(F<=f) одностороннее | 0,104656909 | |
| F критическое одностороннее | 2,271892889 | |

Рисунок 2.4 – Результаты выполнения F-теста для данных по воде

Исходя из полученных значений в таблице со значениями по данным **воды** (рисунок 2.4), находим: $F = 1.8643$, $X_{\text{кр},\alpha} = 2.2719$. В данном случае значение $F = 1.8643$ не попадает в критическую область и нулевая гипотеза H_0 о равенстве дисперсий принимается.

Чтобы проверить гипотезу о равенстве математических ожиданий, с помощью команды «Анализ данных» в программе Excel, вычислим двухвыборочный t-тест с одинаковыми дисперсиями. В данном тесте критическая область представляет собой объединение двух интервалов и имеет следующий вид:

$$-\infty, -t_{\text{кр},\text{дв}} \cup t_{\text{кр},\text{дв}}, \infty.$$

В таблице представлены результаты двухвыборочного t-теста для данных по **электроэнергии** (рисунок 2.5).

| Двухвыборочный t-тест с одинаковыми дисперсиями | | |
|---|--------------|--------------|
| | Переменная 1 | Переменная 2 |
| Среднее | 75296,53926 | 92516,02944 |
| Дисперсия | 22561209,98 | 45559026,38 |
| Наблюдения | 18 | 18 |
| Объединенная дисперсия | 34060118,18 | |
| Гипотетическая разность средних | 0 | |
| df | 34 | |
| t-статистика | -8,851532508 | |
| P(T<=t) одностороннее | 1,20386E-10 | |
| t критическое одностороннее | 1,690924255 | |
| P(T<=t) двухстороннее | 2,40771E-10 | |
| t критическое двухстороннее | 2,032244509 | |

Рисунок 2.5 – Результаты выполнения t-теста для данных по электроэнергии

Для данных по электроэнергии критическая область имеет вид:

$$-\infty, -2.0323 \cup 2.0323, \infty.$$

Наблюдаемый критерий в данном случае принимает значение -8.8515. Данное значение попадает в критическую область. Отсюда следует, что гипотеза H_0 о равенстве математических ожиданий будет отклонена. Так как данные гипотезы (о равенстве математических ожиданий и равенстве дисперсий) отклонены, то это позволяет сделать вывод о том, что временной ряд по данным электроэнергии содержит неслучайную составляющую (тренд).

Ниже в таблице представлены результаты двухвыборочного t-теста для данных по **воде** (рисунок 2.6).

| Двухвыборочный t-тест с одинаковыми дисперсиями | | |
|---|--------------|--------------|
| | Переменная 1 | Переменная 2 |
| Среднее | 202594,625 | 181097,9625 |
| Дисперсия | 13248560109 | 7106675868 |
| Наблюдения | 18 | 18 |
| Объединенная дисперсия | 10177617989 | |
| Гипотетическая разность средних | 0 | |
| df | 34 | |
| t-статистика | 0,639247767 | |
| P(T<=t) одностороннее | 0,263472732 | |
| t критическое одностороннее | 1,690924198 | |
| P(T<=t) двухстороннее | 0,526945464 | |
| t критическое двухстороннее | 2,032244498 | |

Рисунок 2.6 – Результаты выполнения t-теста для данных по воде

Для данных по воде критическая область имеет вид:

$$-\infty, -2.0323 \cup 2.0323, \infty.$$

Наблюдаемый критерий в данном случае принимает значение 0.6392.

Данное значение не попадает в критическую область. Отсюда следует, что гипотеза H_0 о равенстве математических ожиданий принимается. Опираясь на то, что гипотезы о равенстве математических ожиданий и равенстве дисперсий принимаются, можно сделать вывод о принятии гипотезы об отсутствии трендовой составляющей во временном ряду по данным воды. Также стоит отметить, что такие характеристики, как постоянство математического ожидания и дисперсии позволяют принять гипотезу о стационарности временного ряда по данным воды.

2.2 Выделение неслучайной составляющей временных рядов по данным электроэнергии и воды ТСЖ

Чтобы осуществлять долговременное прогнозирование, необходимо вычислить трендовую составляющую. Данная составляющая $T(t)$ представляет собой отражение влияния долговременных факторов и соответствует устойчивой и долговременной тенденции изменения временного ряда. Построение оценки $T(t)$ для функции $T(t)$ является основной задачей выделения тренда. При этом выдвигается предположение о том, что составляющие $P(t)$, $S(t)$ временного ряда отсутствуют [10].

Для выделения трендовой составляющей временных рядов по данным электроэнергии и воду будем использовать регрессионные методы.

Основой регрессионного метода является следующая модель временного ряда:

$$Y_{t_i} = T(t_i) + \varepsilon(t_i),$$

где случайные величины удовлетворяют условиям:

$$\begin{aligned} M \varepsilon_{t_i} &= 0; \\ M \varepsilon_{t_i} \varepsilon_{t_j} &= \begin{cases} \sigma^2, i = j; \\ 0, i \neq j. \end{cases} \end{aligned}$$

Если рассматривать время t как некоторую независимую переменную, то оценку функции $t(i)$ можно получить с помощью регрессионного метода (метода парной регрессии).

Наряду с линейной функцией $T t = \beta_0 + \beta_1 t$ на практике часто используют нелинейные функции, такие как:

- полиномиальная

$$T t = \beta_0 + \beta_1 t + \dots + \beta_p t^p,$$

где p – степень полинома;

- экспоненциальная

$$T t = \beta_0 e^{\beta_1 t};$$

- логистическая

$$T t = \frac{\beta_0}{1 + \beta_1 e^{-\beta_2 t}}.$$

После выбора вида функции $T(t)$ строится уравнение регрессии $T t$, которое зависит от коэффициентов $\beta_0, \beta_1, \dots, \beta_k$ функции тренда

$$Y' t' = \beta'_0 + \beta'_1 t' + \varepsilon',$$

которой соответствует уравнение тренда

$$T' t' = b'_0 + b'_1 t'.$$

Вычисление коэффициентов b'_0, b'_1 происходит на основе метода наименьших квадратов, который имеет следующий вид:

$$a_0 n + a_1 t = y$$

$$a_0 t + a_1 t^2 = y \cdot t$$

Выполняя обратное преобразование $b_0 = e^{b'_0}$, получаем искомые оценки b_0, b_1 для коэффициентов регрессии, которая является нелинейной.

Чтобы выделить трендовую составляющую необходимо построить графики временных рядов по данным электроэнергии и воды, также построить линию тренда, подобрав уравнение регрессии с наибольшим коэффициентом детерминации.

По данным электроэнергии построим уравнения регрессии, и выберем уравнение с наибольшим коэффициентов детерминации R_t^2 .

Ниже приведены вычисленные уравнения регрессии по данным электроэнергии с соответствующими им коэффициентами детерминации (таблица 2.1).

Таблица 2.1 – Уравнения регрессии по данным электроэнергии

| № | Уравнение | R_t^2 |
|---|--|------------------|
| 1 | $T = 806,21t + 68991$ (линейное) | $R_t^2 = 0,6599$ |
| 2 | $T = 69805e^{0,0095t}$ (экспоненциальное) | $R_t^2 = 0,6621$ |
| 3 | $T = 7225,4\ln(t) + 64695$ (логарифмическое) | $R_t^2 = 0,3576$ |
| 4 | $T = 66318t^{0,0857}$ (степенное) | $R_t^2 = 0,36$ |
| 5 | $T = 23,149t^2 - 50,307t + 74416$ (полиномиальное, m=2) | $R_t^2 = 0,7068$ |
| 6 | $T = -3,9441t^3 + 242,05t^2 - 3334,6t + 85229$ (полиномиальное, m=3) | $R_t^2 = 0,8793$ |

Сравнив величины коэффициента детерминации R_t^2 для этих уравнений, выбираем наилучшее уравнение – полиномиальное третьей степени.

Ниже приведен график построения тренда временного ряда по данным электроэнергии (рисунок 2.7).

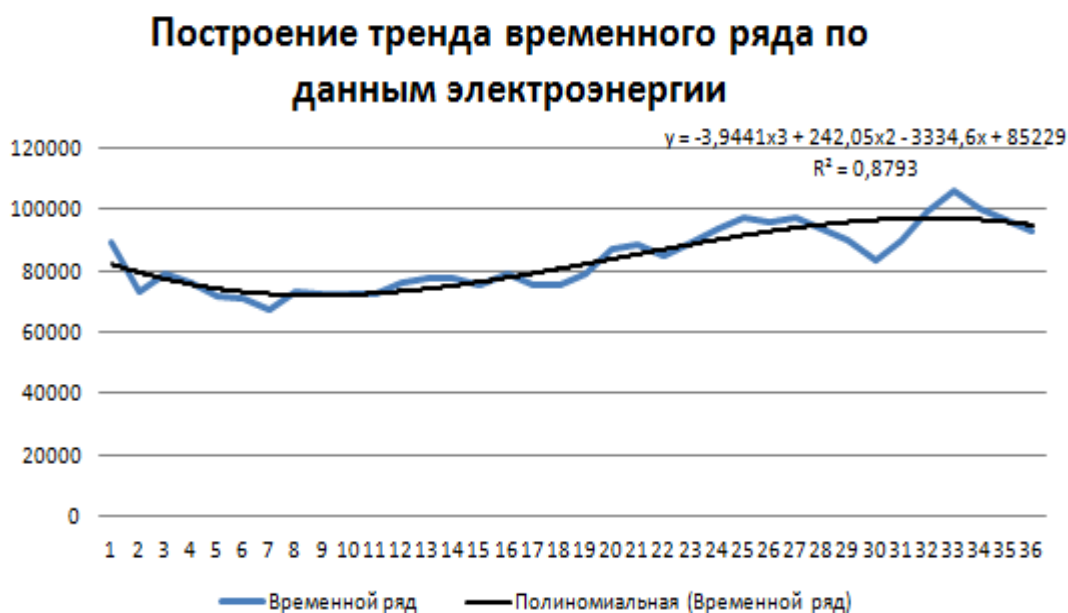


Рисунок 2.7 – График тренда временного ряда по данным электроэнергии

Вычислим уравнения регрессии для временного ряда по данным воды.

Ниже приведены вычисленные уравнения регрессии по данным воды с соответствующими им коэффициентами детерминации (таблица 2.2).

Таблица 2.2 – Уравнения регрессии по данным воды

| № | Уравнение | R_t^2 |
|---|--|------------------|
| 1 | $T = 172,79t + 188164$ (линейное) | $R_t^2 = 0,0004$ |
| 2 | $T = 162925e^{0,0007t}$ (экспоненциальное) | $R_t^2 = 0,0002$ |
| 3 | $T = -10372\ln(t) + 212544$ (логарифмическое) | $R_t^2 = 0,0109$ |
| 4 | $T = 183590t^{-0,04}$ (степенное) | $R_t^2 = 0,0049$ |
| 5 | $T = 47,995t^2 - 1948,6t + 199411$ (полиномиальное, m=2) | $R_t^2 = 0,0034$ |

Продолжение таблицы 2.2

| | | |
|---|--|-----------------|
| 6 | $T = 0,0612t^6 - 6,8068t^5 +$ $293,24t^4 - 6155,8t^3 +$ $64695t^2 - 306639t +$ $+631604 + 85229$ (полиномиальное, m=6) | $R_t^2 = 0,398$ |
|---|--|-----------------|

Исходя из данных, приведенных в таблице 2.2, можно сделать вывод, что ни одно из данных уравнений не подходит для прогнозирования данных по воде, так как коэффициент детерминации R_t^2 во всех приведенных уравнениях очень мал. Значит, данные уравнения нельзя использовать в качестве регрессии.

Для временного ряда по данным воде, при выделении тренда, в качестве объясняющих переменных будем использовать лаговые переменные, оказывающие влияние с некоторым замедлением. Наиболее часто используемыми в качестве такой модели являются авторегрессионные модели. Общий вид авторегрессионной модели имеет следующий вид:

$$y_i = \beta_0 + \beta_1 y_{i-1} + \beta_2 y_{i-2} + \dots + \beta_p y_{i-p} + \varepsilon_i, i = 1, 2, \dots, n,$$

Где $\beta_0, \beta_1, \dots, \beta_p$ – коэффициенты модели, y_{i-1} – лаговые переменные, которые определяют зависимость между значением y_i временного ряда в момент t и значениями в предыдущие моменты времени. Лаговые переменные и будут являться объясняющими переменными.

Так как временной ряд является стационарным, то коэффициент авторегрессионной модели β_0 следует приравнять нулю. Вычислим оценки b_1, b_2 модели авторегрессии второго порядка для коэффициентов β_1, β_2 .

Данные оценки вычисляются по следующим формулам:

$$b_1 = \frac{r(1)(1 - r^2)}{1 - r^2(1)};$$

$$b_2 = \frac{r(2) - r^2(1)}{1 - r^2(1)},$$

а уравнение регрессии имеет вид:

$$y = b_1 y_{i-1} + b_2 y_{i-2}.$$

Ниже представлены вычисленные значения b_1 и b_2 (рисунок 2.8).

| | | | | | | | |
|----|--------------------------|----------|------------------|------------------|----------|----------------|----------|
| 2 | Выявление тренда по воде | | | | | | |
| 3 | | | | | | | |
| 4 | t | Y | Y _{i-1} | Y _{i-2} | T | b ₁ | 1,151159 |
| 5 | 1 | 271864,4 | | | | b ₂ | -0,48249 |
| 6 | 2 | 326039,3 | 271864,4 | | | | |
| 7 | 3 | 247627,4 | 326039,3 | 271864,4 | 244150,1 | | |
| 8 | 4 | 169215,6 | 247627,4 | 326039,3 | 127746,5 | | |
| 9 | 5 | 81097,17 | 169215,6 | 247627,4 | 75315,26 | | |
| 10 | 6 | 83784,24 | 81097,17 | 169215,6 | 11710,24 | | |
| 40 | 36 | 302337,7 | 221758,1 | 141178,6 | 187161,1 | | |
| 41 | | | | | | | |
| 42 | Порядок | 1 | 2 | | | | |
| 43 | p(L) | 0,776502 | 0,411383 | | | | |
| 44 | abs(p(L)) | 0,776502 | 0,411383 | | | | |

Рисунок 2.8 – Вычисленные значения b_1 и b_2

Таким образом, получаем следующее уравнение регрессии:

$$y = 1.151159 y_{i-1} - 0.48249 y_{i-2}.$$

Вычислим коэффициент детерминации построенного уравнения регрессии с помощью программы Excel.

Ниже представлены результаты вычислений (рисунок 2.9).

| Регрессионная статистика | |
|--------------------------|-------------|
| Множественный R | 0,831779455 |
| R-квадрат | 0,791857063 |
| Нормированный R-квадрат | 0,781916968 |
| Стандартная ошибка | 38062,18253 |
| Наблюдения | 33 |

Рисунок 2.9 – Результат вычисленной дисперсии по данным воды

2.3 Выделение тригонометрической составляющей временных рядов по данным электроэнергии и воды ТСЖ

К тригонометрической составляющей $C(t)$ временного ряда относят сезонную составляющую $S(t)$ и периодическую компоненту $P(t)$. Сезонная компонента представляет собой повторяемость экономических процессов в течение некоторого не слишком длительного периода, например, месяца или года, а периодическая компонента представляет собой повторяемость экономических процессов в течение длительных периодов.

Для вычисления тригонометрической составляющей воспользуемся методами гармонического анализа периодических функций.

Гармонический анализ выражает периодическую функцию в виде линейной комбинации косинусов и синусов.

Если предположить, что функция $\varphi(t)$ есть непрерывная функция с периодом T , то данную функцию $\varphi(t)$ можно выразить рядом Фурье. Ряд Фурье имеет следующий вид:

$$\varphi(t) = a_0 + \sum_{k=1}^{\infty} \left[a_k \cos \frac{2\pi k}{T} t + b_k \sin \frac{2\pi k}{T} t \right], 0 \leq t \leq T,$$

где k – номер гармоники. Можно заметить, что уменьшение периода функций $\cos \frac{2\pi k}{T} t$, $\sin \frac{2\pi k}{T} t$ происходит при увеличении номера. Коэффициенты разложения выражаются следующими формулами:

$$a_0 = \frac{1}{T} \int_0^T \varphi(t) dt;$$

$$a_k = \frac{2}{T} \int_0^T \varphi(t) \cos \frac{2\pi k}{T} t dt;$$

$$b_k = \frac{2}{T} \int_0^T \varphi(t) \sin \frac{2\pi k}{T} t dt;$$

$$\varphi_k = \frac{2\pi}{T}k.$$

Ниже представлены результаты вычисления тригонометрической составляющей с помощью ряда Фурье по данным электроэнергии (рисунок 2.10).

Рисунок 2.10 – Вычисление тригонометрической составляющей по
данным электроэнергии

Ниже приведен результат расчета сезонной компоненты по данным электроэнергии путем вычета трендовой составляющей из тригонометрической (рисунок 2.11).

| Сезонная компонента | | |
|---------------------|----------|----------|
| 2014 | 2015 | 2016 |
| 5314,403 | 13326,45 | -4071,78 |
| 6291,122 | 10623,78 | -7046,13 |
| 6326,982 | 7264,229 | -10393,4 |
| 6025,336 | 3851,147 | -13510,2 |
| 5969,28 | 967,6308 | -15813,5 |
| 6571,754 | -973,38 | -16890,3 |
| 7971,236 | -1833,41 | -16602,1 |
| 10000,98 | -1779,2 | -15115,7 |
| 12240,03 | -1231,71 | -12852,1 |
| 14132,35 | -746,968 | -10367,2 |
| 15142,2 | -860,734 | -8196,91 |
| 14903,94 | -1938,62 | -6706,73 |

Рисунок 2.11 – Сезонная компонента по данным электроэнергии

Ниже представлены результаты вычисления тригонометрической составляющей с помощью ряда Фурье по данным воды (рисунок 2.12).

| | | | | | | | | |
|-------------|--------------|----------|--------------|----------|----|---------------|----------|--------------|
| 273989,98 | 237282,2831 | 136995 | 136994,99 | 237282,3 | i | γ_{cp} | P_i | e_i |
| 319872,7533 | 159936,3767 | 277017,9 | -159936,3767 | 277017,9 | 1 | 273990 | 298494,6 | -24504,5761 |
| 242367,8933 | 1,48468E-11 | 242367,9 | -242367,8933 | 2,97E-11 | 2 | 319872,8 | 292162,9 | 27709,88723 |
| 201802,84 | -100901,42 | 174766,4 | -100901,42 | -174766 | 3 | 242367,9 | 257108,3 | -14740,43922 |
| 132263,7183 | -114543,7401 | 66131,86 | 66131,85917 | -114544 | 4 | 201802,8 | 202723,8 | -920,9493918 |
| 92993,08 | -92993,08 | 1,14E-11 | 92993,08 | -2,3E-11 | 5 | 132263,7 | 143581,5 | -11317,81271 |
| 89049,27 | -77118,93001 | -44524,6 | 44524,635 | 77118,93 | 6 | 92993,08 | 95528,68 | -2535,597868 |
| 91865,97667 | -45932,98833 | -79558,3 | -45932,98833 | 79558,27 | 7 | 89049,27 | 71440,95 | 17608,31693 |
| 102017,55 | -1,8748E-11 | -102018 | -102017,55 | 3,75E-11 | 8 | 91865,98 | 77772,64 | 14093,33361 |
| 144140,4867 | 72070,24333 | -124829 | -72070,24333 | -124829 | 9 | 102017,6 | 112827,2 | -10809,62661 |
| 227011,6267 | 196597,8356 | -113506 | 113505,8133 | -196598 | 10 | 144140,5 | 167211,7 | -23071,23311 |
| 302237,88 | 302237,88 | -7,4E-11 | 302237,88 | -1,5E-10 | 11 | 227011,6 | 226354 | 657,648542 |
| 184967,7546 | 89439,07672 | 72140,58 | 5526,964306 | 10040,02 | 12 | 302237,9 | 274406,8 | 27831,0487 |
| a0 | a1 | b1 | a2 | b2 | | | | -1,94026E-11 |
| 34213070236 | 13203611434 | | 131349363,1 | | | | | |
| S0 | S1 | | S2 | | | | | |

Рисунок 2.12 – Вычисление тригонометрической составляющей по данным воды

Ниже приведен результат расчета сезонной компоненты по данным воды (рисунок 2.13).

| Сезонная компонента | | |
|---------------------|-----------|----------|
| 2014 | 2015 | 2016 |
| | 95019,144 | 27835,37 |
| | 67960,844 | 174713,7 |
| 12958,19 | 35373,877 | 14915,24 |
| 74977,26 | 87565,611 | 71626,62 |
| 68266,27 | 11214,67 | 5165,346 |
| 83818,44 | 21730,158 | 16373,73 |
| 14120,87 | 37675,723 | 32826,17 |
| 22866,84 | 11383,909 | 26143,38 |
| 55133,86 | 51012,156 | 43975,69 |
| 95250,62 | 103811,48 | 83232,13 |
| 118769 | 91638,579 | 118537 |
| 82202,36 | 78433,637 | 87245,76 |

Рисунок 2.13 – Сезонная компонента по данным воды

В обоих временных рядах (электроэнергия, вода) видно достаточно хорошее совпадение y_i и p_i . Это же доказывает почти нулевое среднее значение остатков e_i в обоих случаях.

Выводы по второй главе:

В данной главе была проведена проверка статистических гипотез временных рядов по данным электроэнергии и воды ТСЖ, вычислена трендовая составляющая каждого временного ряда, которая по всем статистическим характеристикам хорошо подходит для прогнозирования данных ТСЖ, а также получена тригонометрическая составляющая временных рядов этих данных.

3 Проверка адекватности и качества построенных моделей временных рядов по данным электроэнергетики и воды ТСЖ и прогнозирование их трендовой составляющей

Для данной проверки исходными данными будет обозначаться ряд остатков:

$$e_i = y_i - q t_i, \quad i = 1, 2, \dots, n$$

где $q t_i$ – значение, которое вычислено по полученной модели при $t = t_i$.

Стоит отметить, что $q t_i$ может содержать, как только лишь тренд $t r$, так и тригонометрическую составляющую вместе с трендом.

3.1 Проверка математического ожидания ряда остатков временных рядов по данным электроэнергетики ТСЖ

Если трендовая составляющая $t r$ имеет достаточно точную оценку, а также $M \varepsilon t_i = 0$, то математическое ожидание остатков $e t_i$ должно быть нулевым. следующие статистические гипотезы:

$$H_0: M \varepsilon t_i = 0;$$

$$H_1: M \varepsilon t_i \neq 0, \quad i = 1, 2, \dots, n.$$

Для проверки этих гипотез, необходимо обратиться к критерию, который применим в математической статистике. Введем величину

$$T_e = \frac{e}{s_e} \cdot \bar{n},$$

где

$$e = \frac{1}{n} \sum_{i=1}^n e_i; S_e = \sqrt{\frac{\sum_{i=1}^n e_i^2 - e^2}{n-1}}.$$

Гипотеза H_0 будет принята в случае, когда критерий T_e подчиняется распределению Стьюдента с $(n - 1)$ степенями свободы.

Критическая область в данном случае имеет следующий вид:

$$-\infty, \chi_{\text{лев.}, \frac{\alpha}{2}} \cup \chi_{\text{пр.}, \frac{\alpha}{2}}, \infty,$$

где

$$\chi_{\text{лев.}, \frac{\alpha}{2}} = -\chi_{\text{пр.}, \frac{\alpha}{2}}, \chi_{\text{пр.}, \frac{\alpha}{2}} = t_{1-\alpha, n-1}.$$

При попадании вычисленного значения критерия в критическую область, нулевая гипотеза H_0 будет отклонена с уровнем значимости α и примется гипотеза, которая утверждает об отличии математического ожидания остатков от нуля. В данном случае неслучайная составляющая q t_i будет содержать ненулевую систематическую ошибку.

Ниже представлены вычисленные значения критерия и критической области для проверки равенства математического ожидания нулю по данным электроэнергии (рисунок 3.1).

| Оценка качества | | Проверка математического ожидания ряда остатков | | | | | |
|-----------------|-------------|---|--------------------|-------------------|--|--|--|
| e_{cp} | S_e | T_e | $X_{лев.\alpha/2}$ | $X_{пр.\alpha/2}$ | | | |
| -0,201214815 | 1,190402898 | -1,014185106 | -2,03011 | 2,030108 | | | |
| | | | | | | | |
| | | | | | | | |

Рисунок 3.1 – Вычисленные значения критерия и критической области по данным электроэнергии

3.2 Проверка случайности ряда остатков по данным электроэнергии ТСЖ

Если модель неслучайной составляющей временного ряда подобрана правильно, тогда ряд остатков $e_i, i = 1, 2, \dots, n$ является набором случайных величин. Чтобы проверить свойство случайности значений e_i будем использовать критерий под названием «поворотные точки».

Обозначим значение e_i как поворотную точку, в случае если выполняется одно из следующих систем неравенств:

$$1) e_i < e_{i-1}; e_i < e_{i+1},$$

$$2) e_i > e_{i-1}; e_i > e_{i+1}.$$

Через n_p обозначим количество поворотных точек, обнаруженных в ряде остатков. Если выполняется следующее неравенство

$$n_p > \text{ent} \frac{2}{3} n - 2 - 1.96 \sqrt{\frac{16n - 29}{90}},$$

Где $\text{ent} [.]$ – целая часть. В данном случае ряд остатков считается случайным. Это свидетельствует о том, что построенная модель не содержит систематическую ошибку, а значит, является адекватной данному временному ряду.

Ниже представлены лишь те наблюдения, которые содержат поворотные точки, а также вычисленное неравенство (рисунок 3.2).

| Проверка случайности ряда остатков | | | | |
|------------------------------------|-------------------------|----------|----------|----------------|
| n_p | Вычисленное неравенство | | | |
| 17 | 16,43464937 | | | |
| t | Y_i | T | e | Точки поворота |
| 2 | 72991,81 | 79496,45 | -6504,64 | *** |
| 3 | 79064,49 | 77297,16 | 1767,327 | *** |
| 7 | 67272,29 | 72394,42 | -5122,14 | *** |
| 8 | 73111,18 | 72024,02 | 1087,156 | *** |
| 11 | 72148,73 | 72586,85 | -438,12 | *** |
| 13 | 77605,19 | 74120,46 | 3484,731 | *** |
| 15 | 75679,68 | 76359,91 | -680,232 | *** |
| 16 | 79333,06 | 77685,17 | 1647,89 | *** |
| 18 | 75527,71 | 80628,41 | -5100,7 | *** |
| 21 | 88685,1 | 85420,14 | 3264,957 | *** |
| 22 | 85098,76 | 87023,22 | -1924,46 | *** |
| 25 | 97426,66 | 91518,69 | 5907,969 | *** |
| 26 | 95537,21 | 92833,7 | 2703,515 | *** |
| 27 | 97424,18 | 94017,53 | 3406,654 | *** |
| 30 | 83230,29 | 96545,3 | -13315 | *** |
| 33 | 106099,4 | 97040,53 | 9058,828 | *** |

Рисунок 3.2 – Критерий «поворотных точек»

Исходя из данных, приведенных на рисунке 3.2, можно сделать вывод, что данная модель является адекватной по временному ряду электроэнергии.

3.3 Проверка независимости значений ряда остатков по данным электроэнергии ТСЖ

О неполноте построенной модели может свидетельствовать наличие автокорреляции между значениями ряда остатков (зависимость между этими значениями).

Чтобы определить наличие автокорреляции между значениями временного ряда по данным электроэнергии будет использоваться тест Дарбина-Уотсона.

Основная мысль данного теста выражается в том, что корреляция между ε_i и ε_{i+1} , не равная нулю, свидетельствует о том, что она присутствует и в остатках $e_i = y_i - \hat{y}_i$ регрессионной модели, где $\hat{y}_i = q(t_i)$ – оценка объясненной части временного ряда, которая строится обычным методом наименьших квадратов.

Определение статистики:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2}.$$

Если наблюдается значение:

- 1) $d_B < d < 4 - d_B$ – гипотеза H_0 принимается;
- 2) $d_H < d < d_B$ или $4 - d_B < d < 4 - d_H$ – вопрос о принятии или отвержении гипотезы H_0 остается открытым;
- 3) $0 < d < d_H$ – гипотеза H_0 отвергается и принимается альтернативная гипотеза о положительной автокорреляции;
- 4) $4 - d_H < d < 4$ – гипотеза H_0 отвергается и принимается альтернативная гипотеза о наличии отрицательной автокорреляции.

Пороговые значения d_B, d_H зависят от числа наблюдений, числа объясняющих переменных в функции $q(t)$ и уровня значимости.

Ниже приведено вычисление статистики Дарбина-Уотсона для данных по электроэнергии (рисунок 3.3).

| Проверка автокорреляции остатков | | | | | | |
|----------------------------------|-------------|-------------|----------|----------|-------------|-------------|
| dn | db | | | | | |
| 1,346 | 1,588 | | | | | |
| t | Yi | T | e | ei-ei-1 | e^2 | (ei-ei-1)^2 |
| 1 | 89563,95 | 82132,5059 | 7431,444 | | 55226361,41 | |
| 2 | 72991,81 | 79496,4472 | -6504,64 | -13936,1 | 42310305,1 | 194214362 |
| 3 | 79064,48667 | 77297,1593 | 1767,327 | 8271,965 | 3123446,021 | 68425398 |
| 4 | 76236,25667 | 75510,9776 | 725,2791 | -1042,05 | 526029,7245 | 1085864,7 |
| 5 | 71337,78 | 74114,2375 | -2776,46 | -3501,74 | 7708716,249 | 12262159 |
| 6 | 70969,79333 | 73083,2744 | -2113,48 | 662,9764 | 4466802,219 | 439537,75 |
| 7 | 67272,28667 | 72394,4237 | -5122,14 | -3008,66 | 26236287,79 | 9052010,7 |
| 8 | 73111,17667 | 72024,0208 | 1087,156 | 6209,293 | 1181907,878 | 38555318 |
| 9 | 72525,05333 | 71948,4011 | 576,6522 | -510,504 | 332527,7982 | 260613,96 |
| 36 | 92748,95333 | 94864,2704 | -2115,32 | -3140,77 | 4474566,293 | 9864464,9 |
| | | | | | 691461181 | 806510414 |
| | d | 1,596385671 | | | | |

Рисунок 3.3 – Статистика Дарбина-Уотсона

Исходя из данных, приведенных на рисунке 3.3 можно заметить, что вычисленное значение d находится в пределах от $d_{\text{в}} = 1.588$ до $4 - d_{\text{в}} = 2.412$. Значит, гипотеза об отсутствии автокорреляции принимается.

3.4 Тест на наличие автокорреляции временного ряда по данным воды

Для того, чтобы проверить стационарный временной ряд по данным воды на наличие в нем автокорреляции будем использовать тест Льюинга-Бокса. Выдвинем для данной проверки следующие статистические гипотезы:

H_0 : автокорреляция отсутствует;

H_1 : автокорреляция присутствует.

Критерий теста Льюинга-Бокса определяется следующим выражением:

$$Q_p = n \sum_{i=1}^p \frac{r^2}{n-l'}$$

где n – число наблюдений временного ряда, p – предполагаемый порядок авторегрессионной модели.

Критерий Q_p имеет $\chi^2_{p,1-\alpha}$, если принимается гипотеза о равенстве нулю всех коэффициентов p $l, l = 1, 2, \dots, p$. Если выполняется неравенство $Q_p > \chi^2_{p,1-\alpha}$, то гипотеза H_0 отвергается и принимается предположение о том, что исследуемый временной ряд имеет автокорреляцию. Ниже приведены значения коэффициентов автокорреляции $r(l)$, вычисленные по временному ряду данных по воде (рисунок 3.4).

| Тест на наличие автокорреляции | | | | |
|--------------------------------|----------|-----|----------|-------------------|
| l | $r(l)$ | p | Q | $\chi^2_{p;0,95}$ |
| 1 | 0,756207 | 1 | 22,35114 | 3,841459 |
| 2 | 0,400408 | 2 | 28,6176 | 5,991465 |
| 3 | 0,054559 | 3 | 28,73395 | 7,814728 |

Рисунок 3.4 – Коэффициенты автокорреляции временного ряда по данным воды

Видно, что для всех p выполняется неравенство $Q_p > \chi^2_{p,1-\alpha}$. Отсюда следует, что гипотезы H_0 для $p = 1, 2, 3$ отклоняются с вероятностью ошибки первого рода, которая равна 0.05.

Проведем анализ автокорреляционной функции.

Данный анализ позволяет выявить наличие во временном ряду таких составляющих, как: $T(t), P(t), S(t)$.

Ниже представлено вычисление коэффициентов автокорреляции временного ряда по данным воды (рисунок 3.5).

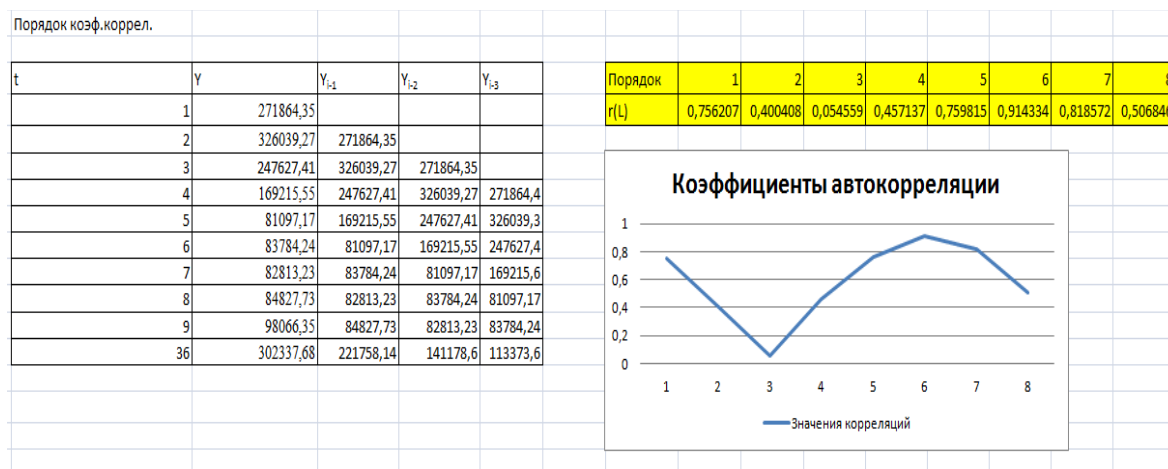


Рисунок 3.5 – Коэффициенты автокорреляции

Результаты анализа значений $r(l)$ и коррелограммы позволяют сделать вывод о том, что временной ряд содержит сезонные колебания периодичностью в 6 месяцев, так как r_6 принимает наибольшее значение, равное 0.9143, а также нелинейный тренд.

3.5 Прогнозирование трендовой составляющей временных рядов по данным электроэнергетики и воды ТСЖ

Основываясь на построенные модели по данным электроэнергетики и воды, рассчитаем точечные прогнозы этих данных. Точечный прогноз получим путем подстановки в уравнение тренда (модель) соответствующего значения фактора времени.

Модель тренда по данным электроэнергетики имеет следующий вид:

$$T(t) = -3,9441t^3 + 242,05t^2 - 3334,6t + 85229.$$

Вычислим точечный прогноз по данным электроэнергетики:

$$T_{37} = -3,9441 \cdot 37^3 + 242,05 \cdot 37^2 - 3334,6 \cdot 37 + 85229 \approx 93434.75$$

$$T_{38} = -3,9441 \cdot 38^3 + 242,05 \cdot 38^2 - 3334,6 \cdot 38 + 8522 \approx 91613.74$$

$$T_{39} = -3,9441 \cdot 39^3 + 242,05 \cdot 39^2 - 3334,6 \cdot 39 + 8522 \approx 89377.58$$

Модель тренда по данным воды имеет следующий вид:

$$T(y_i) = b_1 y_{i-1} + b_2 y_{i-2}.$$

Вычислим точечный прогноз по данным воды:

$$T_{y_{37}} = b_1 y_{37-1} + b_2 y_{37-2} \approx 241041.8$$

Выводы по третьей главе:

Была проведена проверка на адекватность полученных трендовых составляющих временных рядов по данным электроэнергетики и воды, а также вычислен точечный прогноз по данным моделям трендовым составляющим.

ЗАКЛЮЧЕНИЕ

В процессе выполнения выпускной квалификационной работы была изучена характеристика данных ТСЖ, которые являются экономическими показателями. В результате изучения этих данных был сделан вывод, что наиболее полное и точное прогнозирование будет вычислено с помощью такой науки, как эконометрика, проводя экономические исследования.

Для достижения поставленной в работе цели решены следующие задачи:

1) Рассмотрены эконометрические методы прогнозирования с целью выбора наиболее подходящего для вычисления прогноза показателей потребления ресурсов ТСЖ.

2) Используя полученное уравнение регрессии для каждого временного ряда, были вычислены трендовые составляющие временных рядов по данным электроэнергии и воды, по результатам которых были вычислены сезонная и случайная составляющая каждого временного ряда.

3) Проведена проверка на адекватность полученных моделей, а также вычислен точечный прогноз показателей потребления воды и электроэнергии на последующие периоды.

Результаты бакалаврской работы могут быть рекомендованы для анализа и прогнозирования показателей потребления ресурсов в управляющих компаниях и ТСЖ.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

Научная и методическая литература

1. Алексеева, М.М. Планирование деятельности фирмы [Текст]: учеб.-метод. пособие / М.М. Алексеева. - М.: Финансы и статистика, 1998. – 248 с.
2. Афанасьев, В. Н. Анализ временных рядов и прогнозирование / В.Н. Афанасьев, М.М. Юзбашев. - М.: Финансы и статистика, Инфра-М, 2015. - 320 с.
3. Бокс, Дж. Анализ временных рядов прогноз и управление (часть 2) / Дж. Бокс, Г. Дженкинс. - М.: [не указано], 2016. - 185 с.
4. Бокс, Дж. Анализ временных рядов прогноз и управление. Выпуск 1 / Дж. Бокс, Г. Дженкинс. - М.: Мир, 2016. - 408 с.
5. Бриллинджер, Д. Временные ряды. Обработка данных и теория / Д. Бриллинджер. - М.: [не указано], 2015. - 694 с.
6. Дуброва Т.А. Статистические методы прогнозирования: учеб.пособие/ Т. А. Дуброва, М.Ю. Архипова. Московский государственный университет экономики, статистики и информатики. — М., 2004. — 136 с.
7. Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов / Ю.П. Лукашин. - М.: Финансы и статистика, 2015. - 416 с.
8. Романенко, И.В. Социальное и экономическое прогнозирование [Текст]: конспект лекций / И.В. Романенко.- СПб.: Изд-во Михайлова В.А., 2000. – 64 с.
9. Тиндова, М.Г. Экономика и управление: учеб. Пособие / М.Г. Тиндова, О.С. Кузнецова. – Саратов: ССЭИ РЭУ им. Г.В. Плеханова, 2015. – 108 с.
10. Эконометрика в Excel : учеб. пособие. Ч. 2. Анализ временных рядов / Ю. Е. Воскобойников ; Новосиб. гос. архитектур.-строит. ун-т. – Новосибирск : НГАСУ (Сибстрин), 2008. – 39 с.

Электронные ресурсы

11. Временные ряды и их характеристики [Электронный ресурс]. Режим доступа: https://studopedia.ru/7_120330_metod-skolzyashchego-srednego.html.
12. Голик Е.С. Теория и методы статистического прогнозирования [Электронный ресурс]: учеб.пособие / Е.С. Голик, О.В. Афанасьева. – СПб.: Изд-во СЗТУ, 2008. – 72 с.
13. Методы корреляционного и регрессионного анализа [Электронный ресурс]. Режим доступа: http://edu.alnam.ru/book_mkor.php?id=35.
14. Методы регрессионного анализа [Электронный ресурс]. Режим доступа: <https://studfiles.net/preview/5579950/page:8/>.
15. Статистический анализ и прогнозирование доходов бюджета [Электронный ресурс]. Режим доступа: <http://works.doklad.ru/view/m1dAPzbNBmg/4.html>.
16. Моделирование тенденции временного ряда при наличии структурных изменений [Электронный ресурс]. Режим доступа: http://studbooks.net/2244368/matematika_himiya_fizika/modelirovanie_tendentsii_vremennogo_ryada_nalichii_strukturnyh_izmeneniy.

Литература на иностранном языке

17. A Complete Tutorial on Time Series Modeling [Электронный ресурс]. Режим доступа: <https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>.
18. How To Identify Patterns in Time Series Data: Time Series Analysis [Электронный ресурс]. Режим доступа: <https://datamarket.com/data/list/?q=provider%3Atdsl>.
19. Introduction to Time Series Analysis [Электронный ресурс]. Режим доступа: <https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm>.

20. Seizure prediction: the long and winding road / Mormann, Florian; Andrzejak, Ralph G.; Elger, Christian E.; Lehnertz, Klaus// – 2007. – 130 (2). – 314-333.

21. Time Series [Электронный ресурс]. Режим доступа: <https://www.investopedia.com/terms/t/timeseries.asp>.