

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САРАТОВСКИЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра математической экономики

**ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ
С ИСПОЛЬЗОВАНИЕМ ЯЗЫКА R**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

Студента 4 курса 451 группы
направления подготовки 38.03.05 – Бизнес-информатика
механико-математического факультета

Чуйкова Арсения Алексеевича

Научный руководитель:

доцент, к.ф.-м.н., доцент _____ В.В. Новиков

Заведующий кафедрой:

д.ф.-м.н., профессор _____ С.И. Дудов

Саратов 2018

Введение. При торговле товарами посредством сети Интернет постоянно возникает задача прогнозирования объёма продаж. Этот процесс нельзя осуществить без использования специальных знаний и приложений, которые позволяют уменьшить роль ошибок при принятии того или иного решения. Для повышения производительности труда, улучшения качества обслуживания и оптимизации процесса управления необходимо обратить внимание на планирование сбыта продукции.

Цель работы: спрогнозировать объём продаж интернет-магазина с использованием языка R.

Задачами работы являются: анализ теоретических основ прогнозирования динамических рядов, применение полученных знаний на практике для реализации вычислительного эксперимента.

Представленная работа состоит из трёх разделов.

Первый раздел посвящён моделированию временных рядов. В нём рассматриваются основные факты и методы, относящиеся к анализу и прогнозированию временного ряда. В частности, обсуждаются такие понятия как автокорреляция, тренд, сезонность и особенности аддитивной модели.

Второй раздел посвящён использованию среды R для анализа временных рядов. Среди прочего рассмотрены средства языка R, с помощью которых происходит чтение данных временного ряда, его графическое представление, разложение, прогнозирование и экспоненциальное сглаживание. Анализируется работа интернет-магазина и производится прогнозирование объёма продаж на конкретном примере в среде R.

Основное содержание работы. Эконометрическую модель можно построить, используя два типа исходных данных:

- данные, характеризующие совокупность различных объектов в определённый момент (период) времени;
- данные, характеризующие один объект за ряд последовательных моментов(периодов) времени.

Модели, построенные по данным первого типа, называются пространственными моделями. Модели, построенные по данным второго типа, называются моделями временных рядов.

Временной ряд — это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени. В отечественной литературе для этого термина используются синонимы «динамический ряд» и «ряд динамики». Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие циклические колебания ряда;
- случайные факторы.

При различных сочетаниях этих факторов зависимость уровней ряда от времени может принимать разные формы.

Во-первых, большинство временных рядов экономических показателей имеют тенденцию, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. По всей видимости, эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию. На рис. 1 показаны компоненты гипотетического временного ряда. Синим цветом изображена возрастающая тенденция.

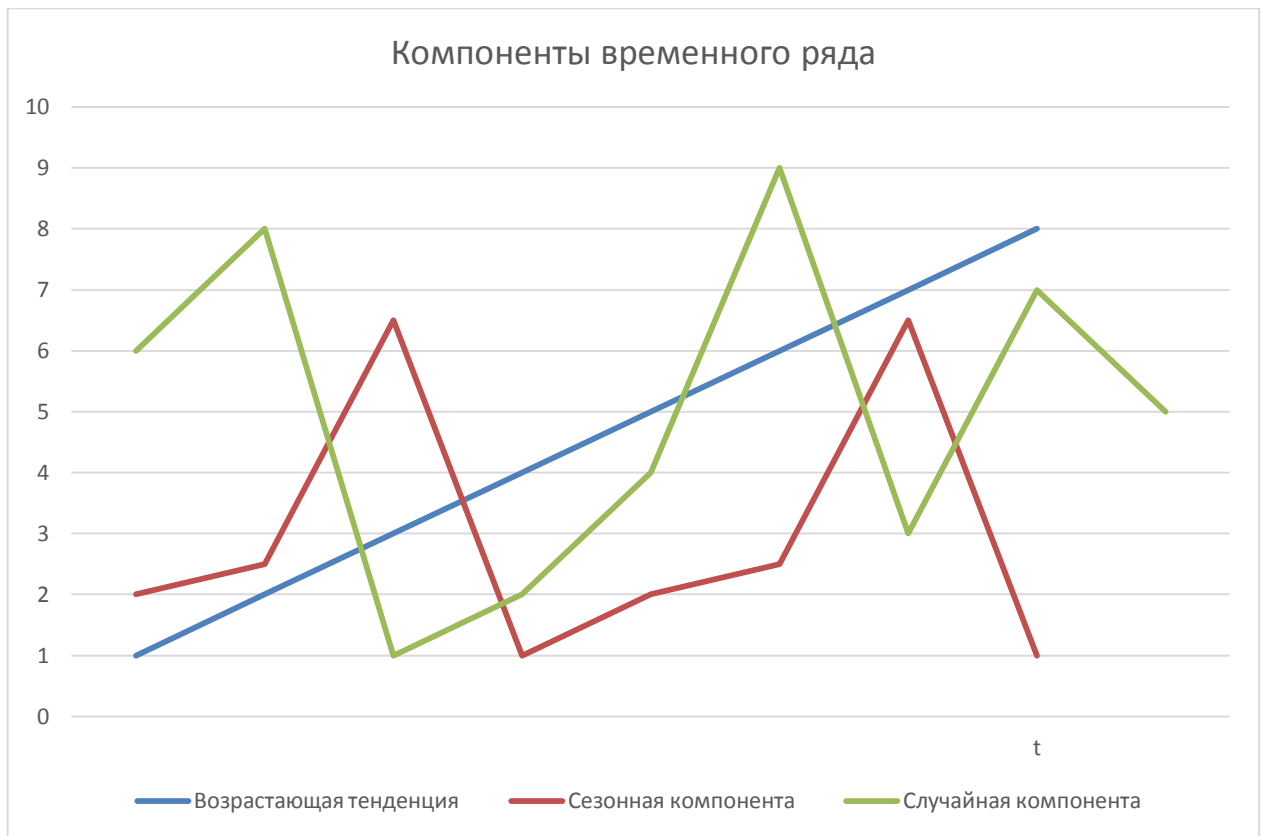


Рисунок 1 - Основные компоненты временного ряда

Во-вторых, изучаемый показатель может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей зависит от времени года. Например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним периодом. При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка, а также с фазой бизнес-цикла, в которой находится экономика страны. На рис. 1. представлен гипотетический временной ряд, содержащий только сезонную компоненту, что изображено красным цветом.

Некоторые временные ряды не содержат тенденции и циклическую компоненту, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты. Пример ряда, содержащего только

случайную компоненту, приведен на рис. 1 зелёным цветом. Очевидно, что реальные данные не соответствуют полностью ни одной из описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется аддитивной моделью временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется мультипликативной моделью временного ряда. Основная задача эконометрического исследования отдельного временного ряда - выявление и придание количественного выражения каждой из перечисленных выше компонент, с тем чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

Известно несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания. Моделирование циклических колебаний в целом осуществляется аналогично моделированию сезонных колебаний, поэтому мы рассмотрим только методы моделирования последних. Простейший подход - расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда. Общий вид аддитивной модели следующий:

$$Y = T + S + E \quad (1)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой T , сезонной S и случайной E компонент. Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E \quad (2)$$

Данная модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой T , сезонной S и случайной E

компонент. Выбор одной из двух моделей проводится на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Для временного ряда, который можно описать с использованием аддитивной модели с увеличением или уменьшением тренда и сезонности, используем экспоненциальное сглаживание Холта-Винтерса для составления краткосрочных прогнозов. Потенциальные сглаживающие оценки Холта-Винтерса учитывают тренд и сезонную составляющую в текущий момент времени. Сглаживание контролируется тремя параметрами: альфа, бета и гамма, для оценки уровня, наклона тренда и сезонной компоненты соответственно. Эти параметры принимают значения от 0 до 1. Значения, близкие к 0, означают, что относительно небольшой вес размещается на последних наблюдениях при прогнозировании будущих значений.

R — язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. R поддерживает широкий спектр статистических и численных методов и обладает хорошей расширяемостью с помощью пакетов. Пакеты представляют собой библиотеки для работы специфических функций или специальных областей применения.

Примером временного ряда, который может быть описан с использованием аддитивной модели с тенденцией и сезонностью, является временный ряд объёма продаж интернет-магазина. Amazon — это самый крупный в мире онлайн-каталог, в котором пользователи из разных стран могут купить практически любые товары. На официальном сайте www.amazon.com опубликованы годовые отчёты организации. Используем данные по объёму продаж за период 2010-2017г. и составим таблицу.

Таблица 1 - Объем продаж интернет-магазина Amazon, млн. долл.

| Год | Квартал | Объем продаж, млн. долл. |
|------|-------------|--------------------------|
| 2010 | I квартал | 7,131 |
| | II квартал | 6,566 |
| | III квартал | 7,560 |
| | IV квартал | 12,948 |
| 2011 | I квартал | 9,857 |
| | II квартал | 9,913 |
| | III квартал | 10,876 |
| | IV квартал | 17,431 |
| 2012 | I квартал | 13,185 |
| | II квартал | 12,834 |
| | III квартал | 13,806 |
| | IV квартал | 21,268 |
| 2013 | I квартал | 16,070 |
| | II квартал | 15,704 |
| | III квартал | 17,092 |
| | IV квартал | 25,587 |
| 2014 | I квартал | 19,741 |
| | II квартал | 19,340 |
| | III квартал | 20,579 |
| | IV квартал | 29,328 |
| 2015 | I квартал | 22,717 |
| | II квартал | 23,185 |
| | III квартал | 25,358 |
| | IV квартал | 35,747 |
| 2016 | I квартал | 29,128 |
| | II квартал | 30,404 |

| | | |
|------|-------------|--------|
| | III квартал | 32,714 |
| | IV квартал | 43,741 |
| 2017 | I квартал | 35,714 |
| | II квартал | 37,955 |
| | III квартал | 43,744 |
| | IV квартал | 60,453 |

Требуется, используя данные за 2010-2017 год (1-32 кварталы), составить прогноз объема продаж Amazon на 2018 год (33-36 кварталы).

Будем предполагать что исходные данные находятся в файле "Amazon.txt" в текущем каталоге.

Чтение данных read.table в набор d:

```
> d<- read.table("Amazon.txt",h=F)
```

Формирование временного ряда amazontimeseries из d:

```
> amazontimeseries <- ts(d, frequency=4, start=c(2010,1))
```

Вывод временного ряда:

```
> amazontimeseries
```

```
      Qtr1  Qtr2  Qtr3  Qtr4
2010  7.131  6.566  7.560 12.948
2011  9.857  9.913 10.876 17.431
2012 13.185 12.834 13.806 21.268
2013 16.070 15.704 17.092 25.587
2014 19.741 19.340 20.579 29.328
2015 22.717 23.185 25.358 35.747
2016 29.128 30.404 32.714 43.741
2017 35.714 37.955 43.744 60.453
```

Вывод графика с фактическими значениями:

```
> plot.ts(amazontimeseries)
```

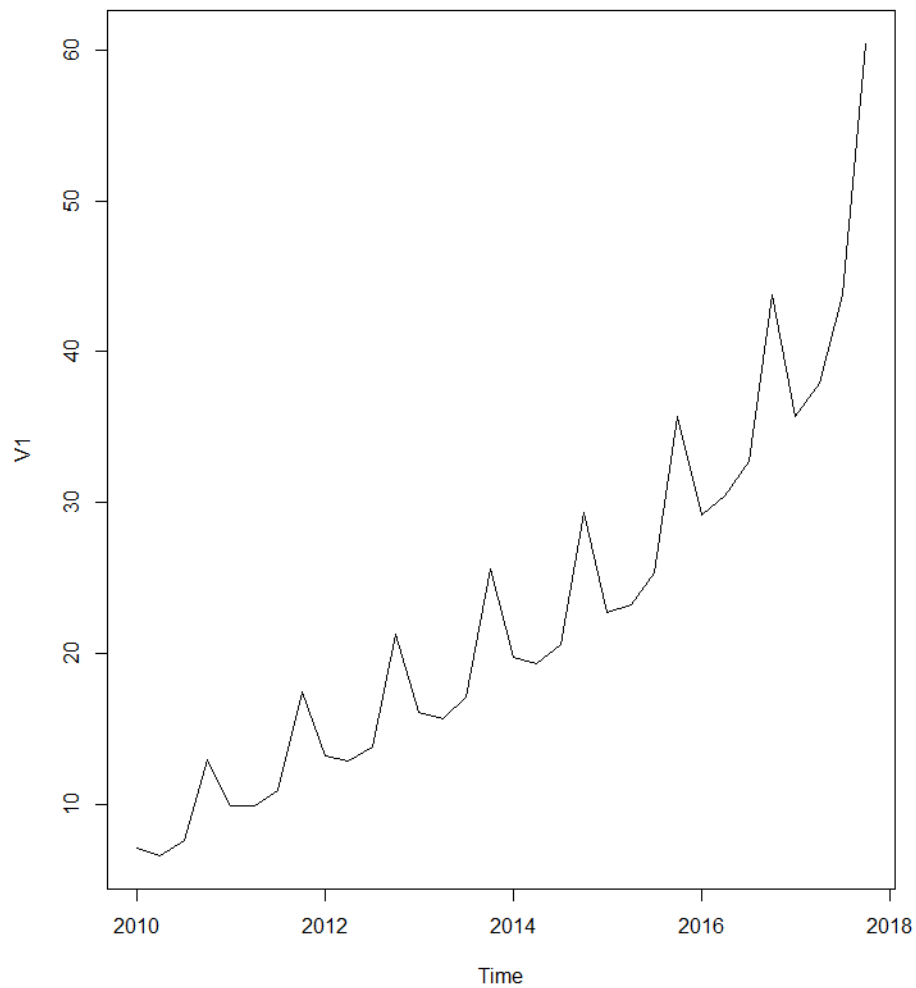



Рисунок 3 – Фактические значения объёма продаж Amazon

Функция `decompose()` возвращает список объектов в качестве результата, где содержатся оценки периодической составляющей, тренда и нерегулярной компоненты, хранящиеся в именованных элементах этого списка объектов, называемых «seasonal», «trend» и «random» соответственно:

```
> amazontimeseriescomponents <- decompose(amazontimeseries)
```

Вывод графика с элементами ряда:

```
> plot(amazontimeseriescomponents)
```

Decomposition of additive time series

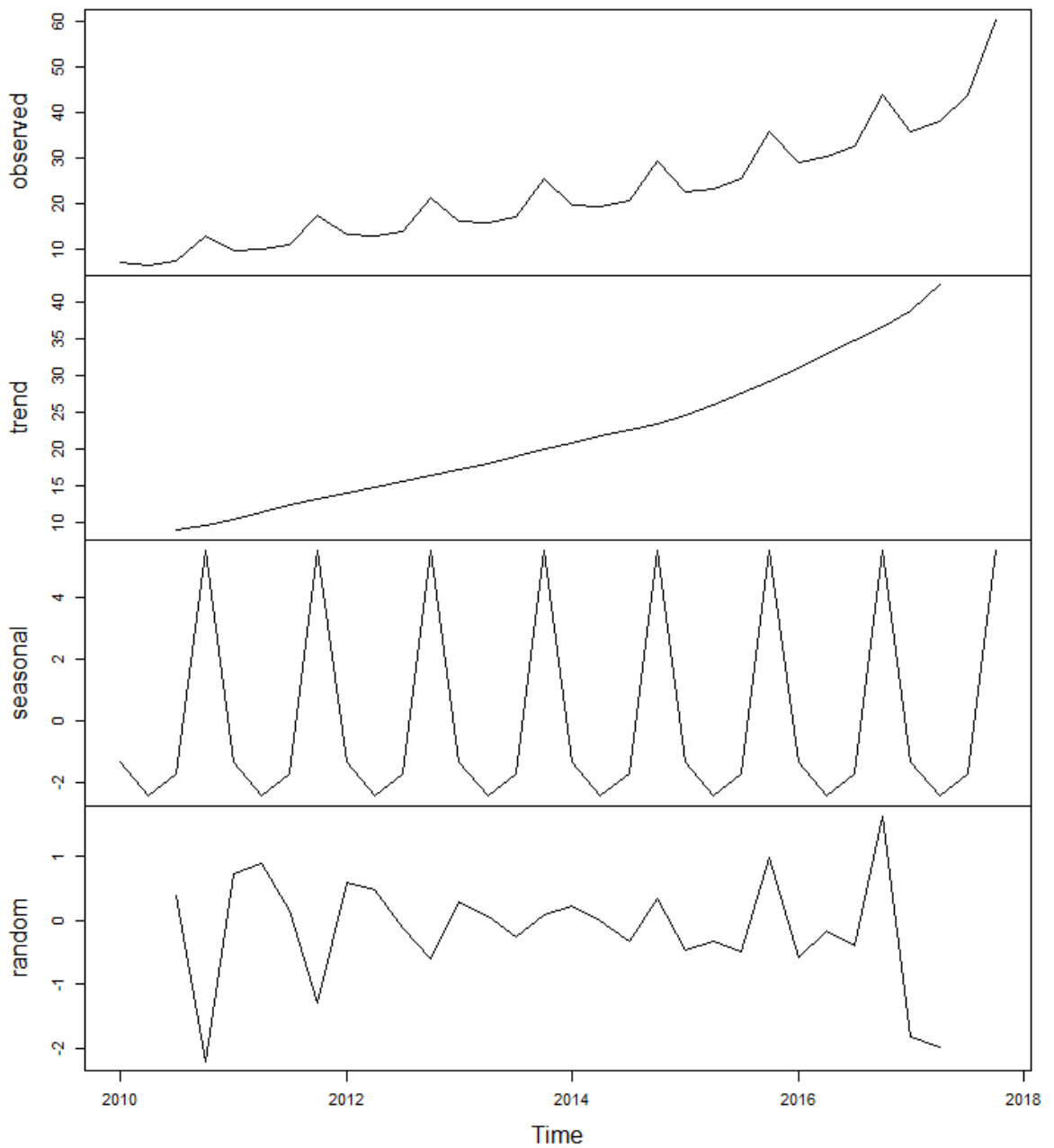


Рисунок 10 – Элементы ряда Amazon

Построение модели HoltWinters. Результат находится в объекте mdHoltWinters:

```
> mdHoltWinters<- HoltWinters(amazontimeseries, alpha = 0.1, beta = 0.2, gamma = 0.9, seasonal = "additive")
```

Вывод списка свойств объекта mdHoltWinters:

```
> summary(mdHoltWinters)
```

| | Length | Class | Mode |
|--------------|--------|--------|-----------|
| fitted | 112 | mts | numeric |
| x | 32 | ts | numeric |
| alpha | 1 | -none- | numeric |
| beta | 1 | -none- | numeric |
| gamma | 1 | -none- | numeric |
| coefficients | 6 | -none- | numeric |
| seasonal | 1 | -none- | character |
| SSE | 1 | -none- | numeric |
| call | 3 | -none- | call |

Вывод значений свойств объекта mdHoltWinters. SSE - сумма квадратов отклонений:

```
> mdHoltWinters$SSE
[1] 79.88424
```

Построение прогнозных значений (точечный и интервальный #прогнозы) на 4 периода вперед с помощью модели mdHoltWinters. Результат находится в объекте dw:

```
> dw <- predict(mdHoltWinters, 4, prediction.interval = TRUE)
```

Вывод результата прогнозирования из объекта dw, где значения точечного прогноза: fit ; интервальный прогноз: lwr - нижняя граница, upr - верхняя граница:

```
> dw
      fit      upr      lwr
2018 Q1 56.43162 59.72876 53.13447
2018 Q2 63.44145 68.75169 58.13122
2018 Q3 72.06540 80.26206 63.86874
2018 Q4 88.77440 100.45751 77.09129
```

Вывод графиков с фактическими, теоретическими и прогнозными значениями:

```
> plot(mdHoltWinters,dw,main="Объём продаж интернет-магазина Amazon",ylab="Факт / Прогноз",xlab="Годы")
```

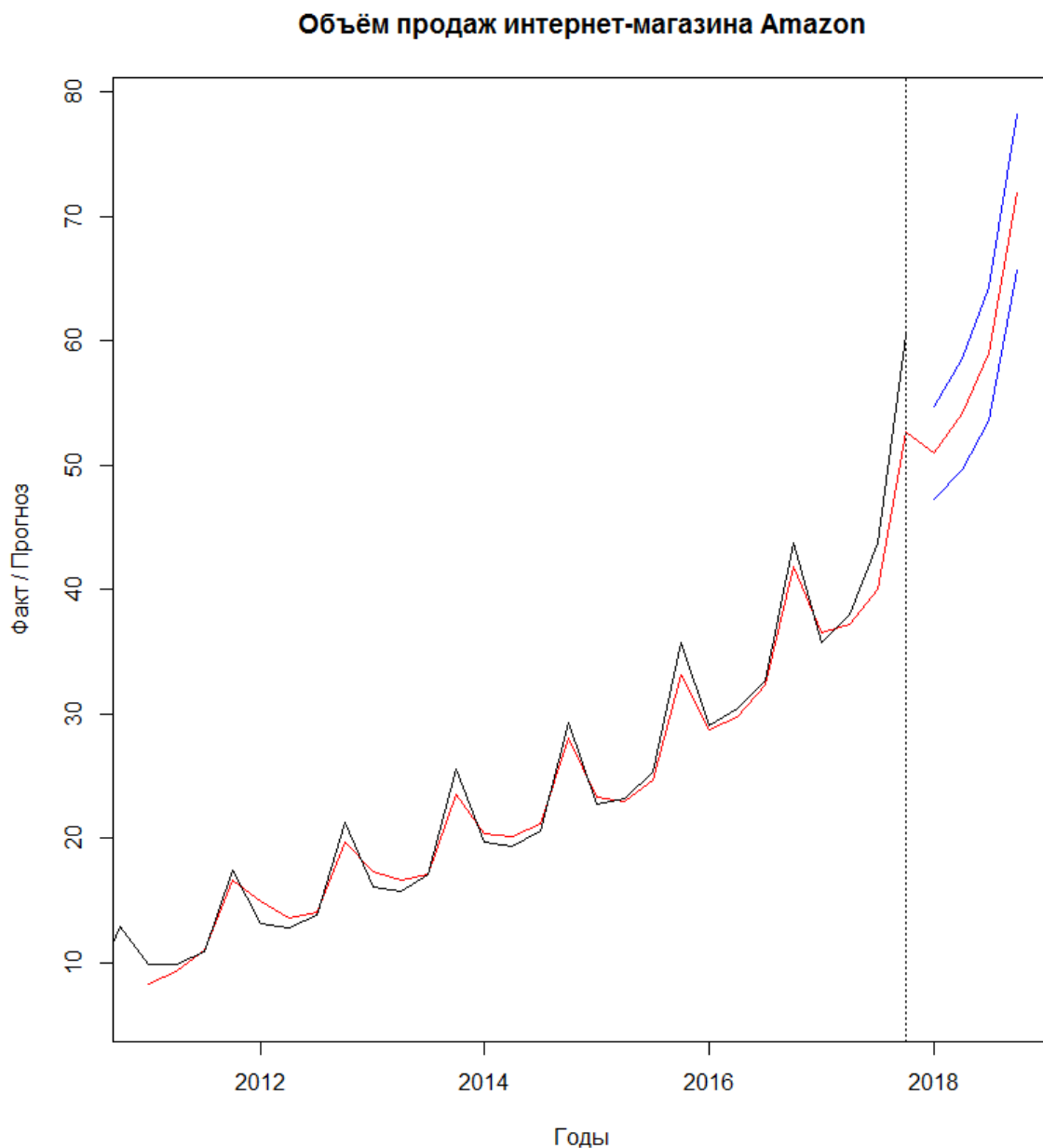


Рисунок 11 – Фактические, теоретические и прогнозные значения ряда Amazon

Заключение. В данной работе был сделан обзор статистических методов анализа временных рядов, а также рассмотрены средства программной реализации этих методов в среде программирования R. В качестве

практического приложения изученной теории было выполнено прогнозирование объёма продаж интернет-магазина средствами языка R.

В заключении можно отметить, что практически все экономические процессы подвержены такому явлению, как сезонность. Она способна оказывать значительное влияние на различные хозяйственные отрасли страны. Однако при помощи существующих универсальных моделей, а также прикладных компьютерных программ возможно осуществить комплексное исследование тренд-сезонных временных рядов, провести их всесторонний анализ и предугадать поведение того или иного экономического процесса. Благодаря этому можно скорректировать деятельность интернет-магазина, предприятия и определенной хозяйственной сферы, чтобы сезонность как можно меньше в будущем повлияла на их нормальное функционирование.