

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.  
ЧЕРНЫШЕВСКОГО»

Кафедра дифференциальных уравнений и математической экономики

**ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ  
КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЦЕН НА РОССИЙСКОМ  
ОПТОВОМ РЫНКЕ ЭЛЕКТРОЭНЕРГИИ**

**АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ**

Студента 2 курса 247 группы

направления - **09.04.03 Прикладная информатика**

**механико-математического факультета**

**Горшуковой Регины Васильевны**

Научный руководитель  
Профессор, д.э.н., профессор

В.А. Балаш

Заведующий кафедрой:  
Зав.кафедрой, д.ф.-м.н., профессор

С.И. Дудов

Саратов 2023

**Введение.** В настоящее время в России действует оптовый рынок электроэнергии и мощности. Данный рынок необходим в современном обществе, потому что электроэнергия является ресурсом, который потребляется практически всеми жителями нашей планеты. Цена данного ресурса оказывает большое влияние на экономику. Все участники оптового рынка электроэнергии и мощности функционируют в условиях конкуренции, в этих условиях им необходимо вырабатывать рыночную стратегию и планировать будущие финансовые потоки. В данных условиях им необходимо прогнозировать цены на электроэнергию на так называемом «рынке на сутки вперед».

**Актуальность данной магистерской работы** связана с тем, что в условиях высокой неопределённости на российском рынке, нужно прогнозировать модели, независимые от прогноза экзогенных переменных, поскольку возможная ошибка предсказания данных параметров приведет к нежелательной ошибке прогноза искомого показателя - цены на рынке электроэнергии.

**Целью магистерской работы** является нахождение метода, который окажется более точным на обучающей и тестовой выборке.

Для достижения поставленной цели в работе необходимо решить следующие **задачи**: в теоретической части исследовать оптовый рынок электроэнергии и мощности, основные принципы его функционирования, основные методы краткосрочного прогнозирования цен в электроэнергетике; в практической части производим построение моделей: ETS, ARIMA, TBATS, NNAR, Prophet, Bootstrapping, ARFIMA.

Выявим более точный метод с помощью реализации их программой R. Построение моделей основывается на данных с сайта АО «АТС» с 28.02.2021 по 28.02.2023.

**Основное содержание работы.** Работа состоит из введения, четырёх разделов, заключения, списка использованных источников, содержащего 26 наименований. Общий объем работы составляет 58 страниц, включая 3 таблицы.

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

В **первом** разделе приводятся основные понятия оптового рынка электроэнергии и мощности и его основные принципы функционирования.

ОРЭиМ действует в так называемых ценовых и неценовых зонах. Ценовая зона делится ещё на две отдельные зоны, на первую и вторую.

В первую входят Приволжский округ и Уральский округ, Северо – Западный, Южный, Центральный и Северо – Кавказский округ. Во вторую входит Сибирский федеральный округ.

В неценовую входят те регионы, в которых по техническим причинам невозможна организация рыночных отношений. Регионы, входящие в данную зону: регионы Дальнего Востока, Калининградская область, Архангельская область, Республика Коми.

ОРЭиМ функционирует по правилам, которые определило Правительство Российской Федерации - это Постановление Правительства Российской Федерации №1172 от 27 декабря 2010 года.

Во **втором** разделе приводятся основные методы краткосрочного прогнозирования цен в электроэнергетике.

Исследователи изучили потребности данного рынка и предложили ряд подходов, по которым можно создавать модели прогнозирования. Данный ряд делится на три группы: 1. модели, которые строятся с помощью теории игр; 2. модели, которые называются фундаментальные или имитационные; 3. модели, которые строятся с помощью анализа временных рядов, они подразделяются на два направления: традиционные статистические модели и модели искусственного интеллекта.

В **третьем** разделе производится построение моделей прогноза цен на электроэнергию на рынке «на сутки вперед» и анализ данных.

ETS метод иначе называют «Метод экспоненциального сглаживания». Каждый метод помечается парой букв (T,S), которые определяют тип "трендовых" и "сезонных" компонентов, комбинации представлены в таблице 1.

Для каждого метода существует две модели: с аддитивными ошибками(A) и с мультипликативными ошибками(M). Для отличия этих моделей, а также для отличия моделей от методов, добавляем третью букву ETS(\*,\*,\*).

Таблица 1

Тренд	Сезонная составляющая		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N(None)	(N,N)	(N,A)	(N,M)
A(Additive)	(A,N)	(A,A)	(A,M)
$A_d$ (Additive damped)	( $A_d$ ,N)	( $A_d$ ,A)	( $A_d$ ,M)

ARIMA метод расшифровывается как авторегрессионная интегрированная скользящая средняя. Многие из моделей являются частными случаями модели ARIMA, как показано в таблице 2.

Таблица 2

Белый шум	ARIMA (0,0,0)
Случайное блуждание	ARIMA(0,1,0)
Случайное блуждание с дрейфом	ARIMA(0,1,0)
Авторегрессия	ARIMA(p,0,0)
Скользящее среднее значение	ARIMA(0,0,q)

Модель выгладит так:  $ARIMA(p,d,q)$ , где  $p$  - порядок следования авторегрессионной части;  $d$  - степень первой вовлеченной дифференцировки;  $q$  - порядок расположения части скользящей средней.

TBATS метод основывается на комбинациях из рядов Фурье с экспоненциальной сглаживающей моделью и преобразованиях Бокса – Кокса.

NNAR метод иначе называется авторегрессией нейронной сети, обозначается  $NNAR(p,k)$ , где  $p$  - запаздывающие входы и  $k$  - число скрытых узлов. Этот метод основывается на данных временных рядов с запаздывающими значениями. Они используются в качестве входных данных для нейросети.

Prophet метод состоит в подгонке аддитивных регрессионных моделей со следующими четырьмя основными компонентами: тренд, годовая сезонность, недельная сезонность, праздники.

Метод Bootstrapping заключается в том, что он генерирует выборку размера  $B$ , которая называется bootstrap-выборка, из исходного набора данных

размера  $N$ . Осуществляется данный процесс с помощью случайного отбора с заменой  $B$  наблюдений.

ARFIMA модель иначе называют «Модель дробно-интегрированной авторегрессии скользящего среднего», она обеспечивает еще один подход к прогнозированию временных рядов считается, что параметр  $d$  может принимать дробные значения.

Произведем анализ моделей. Метод экспоненциального сглаживания оценивает временной ряд и автоматически выбирает модель с помощью реализации функции  $ets(y, h)$ . Результаты представлены на рисунке 1 для Северо-Западного округа (СЗО) и рисунке 2 для Сибирского федерального округа (СФО).

```
R Console
> xttest12=subset(x12, start = length(x12)-21)
> pe1=ets(x11)
> summary(pe1)
ETS(A,N,A)

Call:
ets(y = x11)

Smoothing parameters:
  alpha = 0.3549
  gamma = 1e-04

Initial states:
  l = 1386.4236
  s = -1.8461 11.6939 26.7066 8.7861 -63.7437 16.0495
    2.3536

sigma: 82.2876

AIC  AICc  BIC
10918.41 10918.72 10964.05

Training set error measures:
      ME  RMSE  MAE  MPE  MAPE  MASE
Training set -0.2044982 81.76362 61.46694 -0.3697983 5.0728 0.6100713
      ACF1
Training set 0.07732303
> autoplot(forecast(pe1,h=21))+ autolayer(xttest11)
```

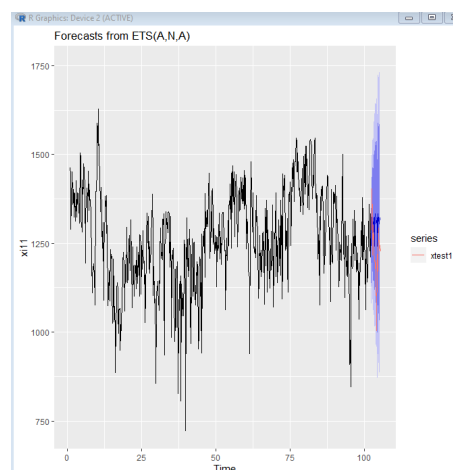


Рисунок 1 — Северо-Западный округ

```
R Console
> autoplot(forecast(pe1,h=21))+ autolayer(xttest11)
> pe2=ets(x12)
> summary(pe2)
ETS(A,N,A)

Call:
ets(y = x12)

Smoothing parameters:
  alpha = 0.3103
  gamma = 0.0324

Initial states:
  l = 1368.2081
  s = -24.5001 -11.3523 13.7548 49.436 -1.2787 5.353
    -31.4127

sigma: 88.2715

AIC  AICc  BIC
11017.95 11018.26 11063.59

Training set error measures:
      ME  RMSE  MAE  MPE  MAPE  MASE
Training set -1.54626 87.70941 58.04339 -2.090617 7.579609 0.6966246
      ACF1
Training set 0.1327244
> autoplot(forecast(pe2,h=21))+ autolayer(xttest12)
```

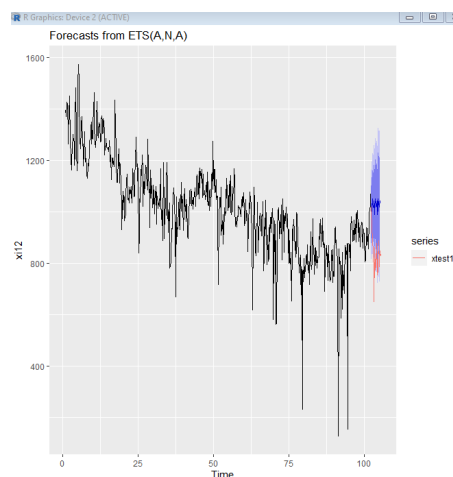


Рисунок 2 — Сибирский федеральный округ

В результате видим, что в данном случае для СЗО и СФО выбрана модель ETS( A, N, A): модель с отсутствующим трендом и аддитивной сезонностью с аддитивной ошибкой.

Выбор подходящих значений для параметров  $p$ ,  $d$  и  $q$  модель авторегрессии и скользящего среднего реализуется с помощью функции `auto.arima()`, которая автоматически выбирает параметры. Результаты для рядов модели ARIMA представлены на рисунке 3 для СЗО и рисунке 4 для СФО.

```
R Console
> pa1=auto.arima(xil1)
> summary(pa1)
Series: xil1
ARIMA(1,1,2) (0,0,2) [7]

Coefficients:
          ar1          ma1          ma2          sma1          sma2
          0.6801         -1.1383          0.1782          0.1500          0.1138
s.e.      0.0691          0.0861          0.0751          0.0391          0.0370

sigma^2 = 7213: log likelihood = -4147.46
AIC=8306.91  AICc=8307.03  BIC=8334.29

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -1.503605 84.57082 63.61688 -0.5823451 5.265388 0.6314099
              ACF1
Training set 0.004642925
> autoplot(forecast(pa1,h=21))+ autolayer(xtest11)
```

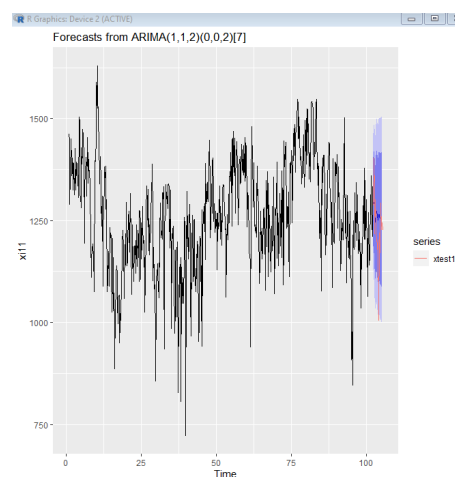


Рисунок 3 — Северо-Западный округ

```
> pa2=auto.arima(xil2)
> summary(pa2)
Series: xil2
ARIMA(1,1,2) (0,0,1) [7]

Coefficients:
          ar1          ma1          ma2          sma1
          0.5526         -1.1329          0.1877          0.1377
s.e.      0.0954          0.1055          0.0909          0.0373

sigma^2 = 7410: log likelihood = -4157.54
AIC=8325.07  AICc=8325.16  BIC=8347.89

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set -4.232099 85.77946 56.40793 -2.356305 7.380862 0.6769962
              ACF1
Training set -0.006194013
> autoplot(forecast(pa2,h=21))+ autolayer(xtest12)
```

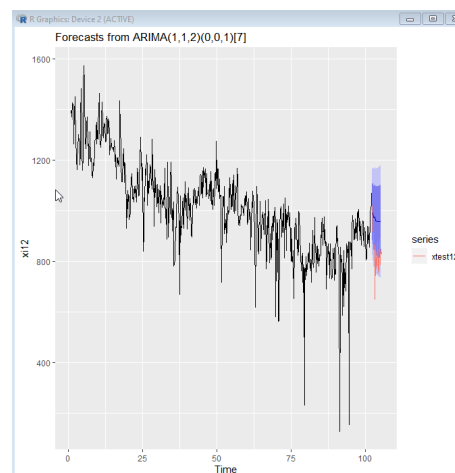


Рисунок 4 — Сибирский федеральный округ

По полученным результатам можно сделать вывод, что ряд по СЗО соответствует модели  $ARIMA(1,1,2)$ :  $y_t = c + 0.6801y_{t-1} - 1.1383\varepsilon_{t-1} + 0.1782\varepsilon_{t-2} + \varepsilon_t$ . Стандартное отклонение  $\sqrt{7213} = 84.93$ . Ряд по СФО соответствует моде-

ли  $ARIMA(1,1,2): y_t = c + 0.5526y_{t-1} - 1.1329\varepsilon_{t-1} + 0.1877\varepsilon_{t-2} + \varepsilon_t$ . Стандартное отклонение  $\sqrt{7410} = 86.08$ .

Функция  $tbats(y)$  реализуется для модель TBATS. Значение показателей функции представлены в таблице 3. Результаты для рядов модели TBARS представлены на рисунке 5 для СЗО и рисунке 6 для СФО.

Таблица 3

Компонент	Значение
1	Параметр преобразования Бокса-Кокса
{0,0}	Ошибка ARMA
-	Параметр демпфирования
{<0,0>}	Сезонный период, слагаемые Фурье

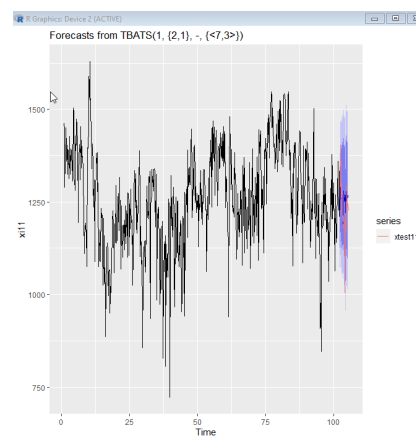
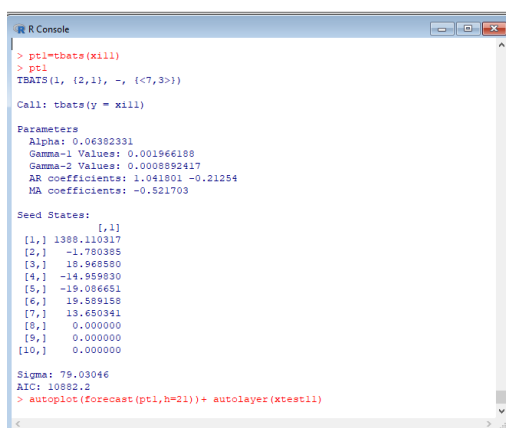


Рисунок 5 — Северо-Западный округ

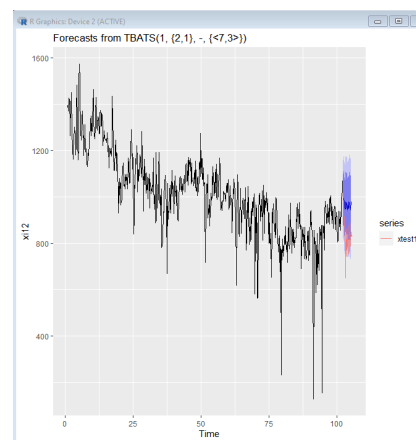
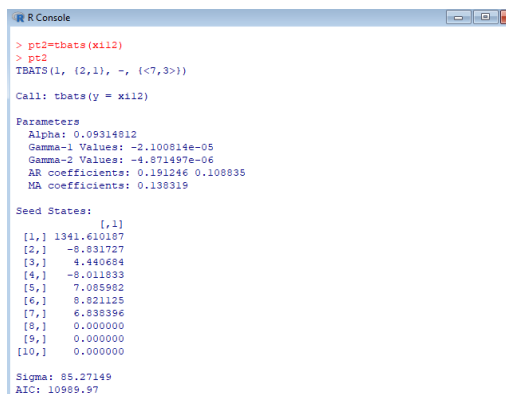


Рисунок 6 — Сибирский федеральный округ

Ряд по СЗО соответствует модели  $TBATS(1, \{2,1\}, -, \langle 7,3 \rangle)$ , где 1 означает, что преобразования Бокса-Кокса не было,  $\{2,1\}$  - ARMA(2,1), - влияние внутри колебательной системы отсутствует,  $\langle 7,3 \rangle$  - семидневная сезонность, три члена ряда Фурье. По СФО мы наблюдаем, что результат  $TBATS(1, \{2,1\}, -, \langle 7,3 \rangle)$ , где 1 означает, что преобразования Бокса-Кокса не было,  $\{2,1\}$  - ARMA(2,1), - влияние внутри колебательной системы отсутствует,  $\langle 7,3 \rangle$  - семидневная сезонность, три члена ряда Фурье.

Для построения нейросетевой модели ряда реализуется функции  $nnetar(y, lambda)$ , где трансформацию Бокса-Кокса  $lambda=0$ , чтобы гарантировать, что прогнозы будут положительными. Результаты представлены на рисунке 7 для СЗО и рисунке 8 для СФ.

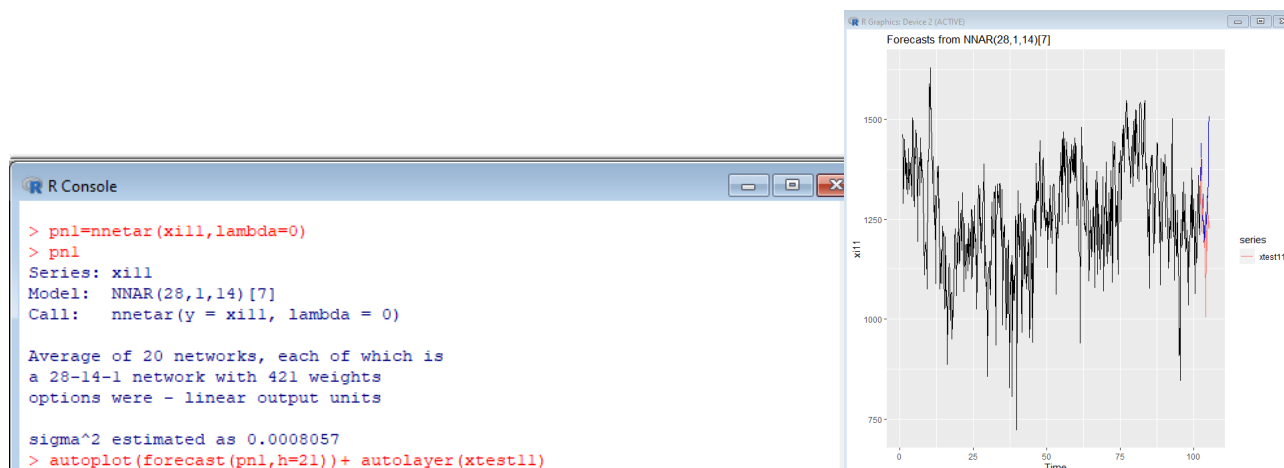


Рисунок 7 — Северо-Западный округ



Рисунок 8 — Сибирский федеральный округ



Ряд по СЗО соответствует модели NNAR  $(28, 1, 14)_7$ , она имеет  $y_{t-1}, y_{t-2}, y_{t-3}$  и  $y_{t-7}$  и четырнадцать нейронных скрытых узлов. По СФО мы наблюдаем, что модель выдала результат NNAR  $(28, 1, 14)_7$ , она имеет  $y_{t-1}, y_{t-2}, y_{t-3}$  и  $y_{t-7}$  и четырнадцать нейронных скрытых узлов.

Результаты для модели prophet первых нескольких предсказанных значений отклика, их (принятые по умолчанию) 80%-ные доверительные границы и график представлены на рисунке 9 для СЗО и рисунке 10 для СФО.



Рисунок 9 — Северо-Западный округ

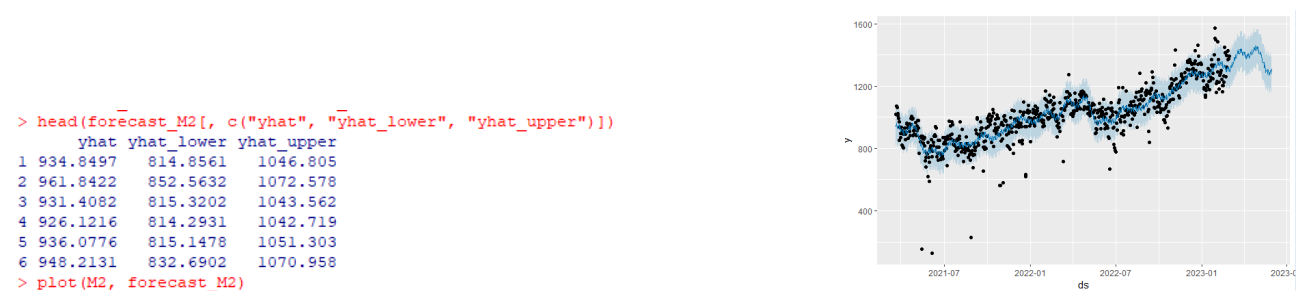


Рисунок 10 — Сибирский федеральный округ

Черные точки на рисунке 9 и 10 соответствуют значениям отклика из обучающей выборки. Сплошная голубая линия - это предсказанные моделью значения отклика, а огибающая эту линию светло-голубая "лента" это 80% доверительные интервалы соответствующих предсказанных значений. Прогнозные значения  $y$  на следующие 90 дней видны в правой части графика.

С помощью функции prophet\_plot\_components() можно изобразить отдельные компоненты модели на рисунке 11 для СЗО и СФО.

На рисунке 11 в верхней части изображен тренд, посередине изображена недельная сезонность, а внизу можно увидеть годовую сезонность.

В СЗО мы наблюдаем, что тренд сложный, а также что в этом временном ряду максимальная цена по понедельникам и минимальная цена в воскре-

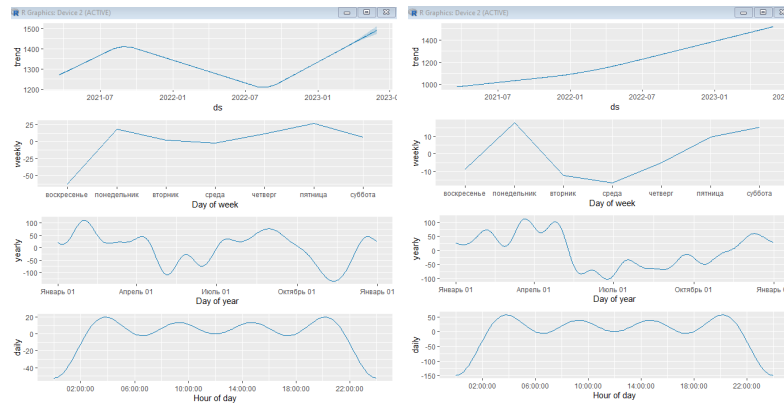


Рисунок 11 — Северо-Западный округ и Сибирский федеральный округ

сенью. Если рассматривать внутри-годовые колебания, можно заметить закономерность, что высокая цена обычно в зимние месяцы, а минимальная в осенний период. В СФО мы наблюдаем, что тренд восходящий, а также что в этом временном ряду максимальная цена по понедельникам, а минимальная цена в среду. Если рассматривать внутри-годовые колебания, можно заметить закономерность, что высокая цена обычно в весенние месяцы, а минимальная в летний.

Функция `bld.mbb.bootstrap()` генерирует новые временные ряды, из исходного ряда, используя выборку `bootstrap`. Результаты для рядов метода Bootstrapping представлены для СЗО и для СФО на рисунке 12.

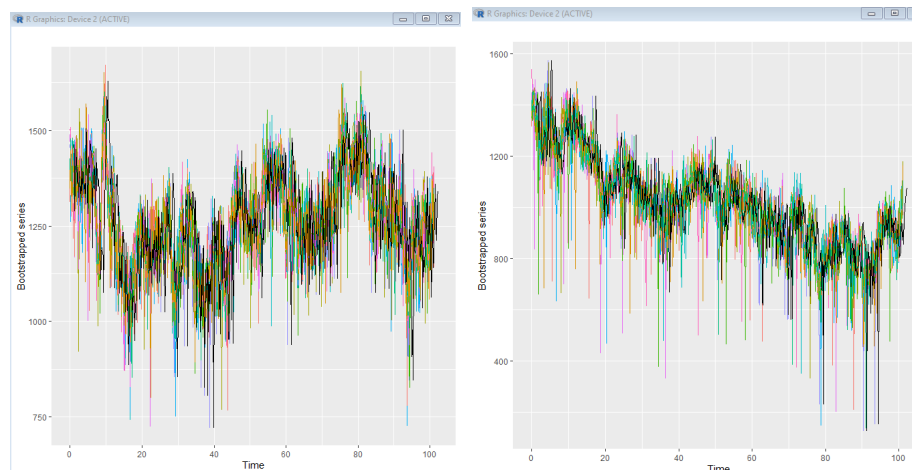


Рисунок 12 — Северо-Западный округ и Сибирский федеральный округ

На основе представленного графика для СЗО округа и для СФО видно, что bootstrapping наглядно показывает, что прогноз неточен, на графике представлено 10 выборок и они практически не пересекаются.

Функция `arfima(y)` реализуется для модель ARFIMA, функция объединяет `fracdiff` и `auto.arima` для автоматического выбора и оценки модели ARFIMA. Результаты для рядов модели ARFIMA представлены на рисунке 13 для СЗО и рисунке 14 для СФО.

```
R Console
> arf1=arfima(xil1)
> summary(arf1)

Call:
arfima(y = xil1)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
d          0.44198    0.02633  16.787 <2e-16 ***
ar.ar1     0.08237    0.04688   1.757  0.0789 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
sigma[eps] = 86.10648
[d.tol = 0.0001221, M = 100, h = 4.391e-05]
Log likelihood: -4166 ==> AIC = 8338.17 [3 deg.freedom]
> autoplot(forecast(arf1,h=21))+ autolayer(xtest11)
-----
```

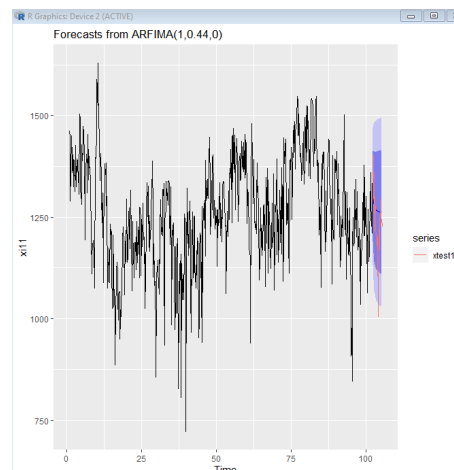


Рисунок 13 — Северо-Западный округ

```
R Console
> arf2=arfima(xil2)
> summary(arf2)

Call:
arfima(y = xil2)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
d          0.476076    0.015546  30.623 <2e-16 ***
ar.ar1    -0.015738    0.040809  -0.386  0.70
ar.ar2     0.008834    0.038749   0.228  0.82
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
sigma[eps] = 86.64409
[d.tol = 0.0001221, M = 100, h = 4.396e-05]
Log likelihood: -4171 ==> AIC = 8349.959 [4 deg.freedom]
> autoplot(forecast(arf2,h=21))+ autolayer(xtest12)
-----
```

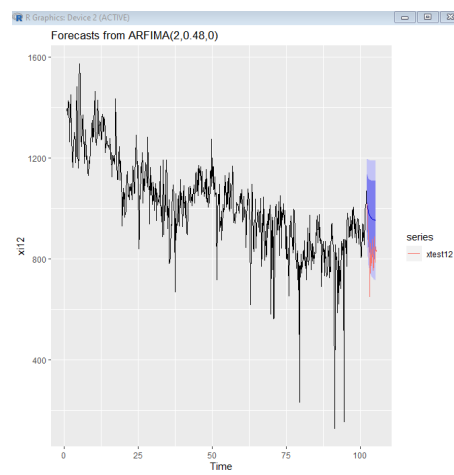


Рисунок 14 — Сибирский федеральный округ

По полученным результатам можно сделать вывод, что ряд для СЗО соответствует модели ARFIMA(1,0.44,0), где параметр  $d = 0.44$ . По СФО модель выдала результат, что ряд соответствует модели ARFIMA(2,0.48,0), где параметр  $d = 0.48$ .

В четвёртом разделе рассмотрим оценку точности результатов прогнозирования и выберем итоговую модель.

В качестве характеристики точности моделей могут быть использованы следующие показатели: Рассмотрим для моделей следующие показатели

точности: средняя ошибка (ME), среднеквадратичная ошибка корня (RMSE), средняя абсолютная ошибка (MAE), средняя процентная ошибка прогнозирования (MPE), средняя абсолютная процентная ошибка (MAPE).

Все показатели можно найти с помощью функции `accuracy()`. Если точность прогноза равна приближена к 0, то выбранная модель описывает фактические значения на 100%, т. е. очень точно, но надо сразу обратить внимание, что такого точного показателя никогда не будет. На рисунке 15 представлены ошибки для моделей.

	RMSE	MAE	MPE	MAPE
Training ETS	81.76362	61.46694	-0.36979833	5.0728002
Test ETS	91.18519	78.54962	-4.79753356	6.4193049
Training ARIMA	84.57082	63.61688	-0.58234512	5.2653878
Test ARIMA	60.73933	54.10715	-1.94956730	4.3472799
Training TBATS	79.03046	59.24492	-0.49215775	4.8986079
Test TBATS	63.74517	55.60859	-1.21918884	4.4573753
Training NNAR	17.18718	11.37673	-0.02727521	0.9267085
Test NNAR	54.25187	46.33561	-2.25695860	3.7098242
Training Prophet	198.49720	165.99825	0.13960584	0.1382594
Test Prophet	2753.36530	2173.34352	0.78594556	1.6829462
Training Bootstrapping	90.98802	62.68281	-1.49012454	4.7040993
Test Bootstrapping	96.92393	-53.11616	-2.52318888	4.9496000
Training ARFIMA	86.40069	65.57205	-0.52169050	5.4245060
Test ARFIMA	80.85395	62.11661	-3.84059280	5.2781050

Рисунок 15 — Северо-Западный округ и Сибирский федеральный округ

Из показателей можно сделать вывод, что в СЗО на обучающей и тестовой выборке наиболее точной моделью является NNAR, в СФО на обучающей и тестовой выборке наиболее точной моделью является TBATS.

В **заключении** приведены результаты магистерской работы.

### Основные результаты

1. Определены основные понятия оптового рынка электроэнергии и мощности и его принципы функционирования;
2. Изучены основные методы краткосрочного прогнозирования цен в электроэнергетике;
3. Построены модели и графики на так называемом «рынке на сутки вперед»;
4. Проанализированы результаты Северо-Западного округа и Сибирского федерального округа;
5. Выявлены точные модели для Северо-Западного округа и Сибирского федерального округа.