

УДК 51.76

**ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И  
АЛГОРИТМИЧЕСКОГО ЯЗЫКА R В ПРОГНОЗИРОВАНИИ  
УРОЖАЙНОСТИ ЗЕРНОВЫХ КУЛЬТУР В ТАМБОВСКОЙ ОБЛАСТИ**

**Борис Игнатьевич Смагин**

доктор экономических наук, профессор

bismagin2023@mail.ru

**Николай Алексеевич Барков**

магистрант

barkov-92@bk.ru

Мичуринский государственный аграрный университет

г. Мичуринск, Россия

**Аннотация.** В статье рассмотрена методология Бокса-Дженкинса для прогнозирования урожайности зерновых культур региона на основе использования алгоритмического языка R. Проведение предварительного анализа при построении автокорреляционной и частной автокорреляционных функций позволяет выявить особенности временного ряда, а применение критерия Бокса-Пирса – диагностировать его на стационарность. В дальнейшем применение пакетов языка R позволило построить высококачественную прогнозирующую функцию.

**Ключевые слова:** алгоритмический язык R, зерновые культуры, урожайность, прогнозирование, стационарность.

Прогнозирование того или иного процесса или явления, как правило, основывается на анализе временных рядов, из которых извлекается сводная и статистическая информация точек данных, заданных в хронологическом порядке. В этом процессе появляется возможность не только изучить прошлые состояния, но и спрогнозировать состояние изучаемого явления в будущем.

Ранее нами было показано, что сельскохозяйственное производство обладает стохастическим принципом действия, в силу чего объективный анализ аграрного сектора экономики возможен лишь в рамках вероятностных категорий [3]. Поэтому прогноз любой сельскохозяйственной отрасли рассматривается как вероятностное суждение о будущем состоянии исследуемого объекта.

Для осуществления прогноза мы остановили свой выбор на алгоритмическом языке R, который в первую очередь ориентирован на вероятностно-статистическое моделирование изучаемых процессов и явлений.

При прогнозировании урожайности зерновых культур была использована методология Бокса-Дженкинса [1].

Мы осуществили сбор данных по зерновому производству в Тамбовской области за 1990 – 2023 гг., всего 34 наблюдения (идентификатор соответствующего файла – `tamb_obl`).

В качестве показателей были взяты: время в годах (YEAR), валовое производство зерна, тыс. т. (VAL), урожайность, ц/га (UROV), площадь посева зерновых культур, га (S).

Проверка структуры данного файла (`str(tamb_obl)`) показала, что файл действительно содержит 34 наблюдений и 4 переменные, обозначенные как YEAR, VAL, UROV, S.

```
str(tamb_obl)
```

```
tibble [34 × 4] (S3: tbl_df/tbl/data.frame)
 $ YEAR: num [1:34] 1990 1991 1992 1993 1994 ...
 $ VAL : num [1:34] 2449 1806 2249 2170 1904 ...
 $ UROV: num [1:34] 24.1 17.1 20.3 19.5 18.6 9.8 15.2 20.1 17.1 14 ...
 $ S   : num [1:34] 1165 1100 1126 1134 1040 ...
```

Для предварительного анализа и составления прогноза необходимо

загрузить соответствующие пакеты языка R:

```
library(stats)
> library(fable)
```

Загрузка требуемого пакета: fabletools

Registered S3 method overwritten by 'tsibble':

```
method          from
as_tibble.grouped_df dplyr> library(fabletools)
> library(tsibble)
```

Присоединяю пакет: 'tsibble'

Следующие объекты скрыты от 'package:base':

```
intersect, setdiff, union> library(ggplot2)
> library(forecast)
```

Registered S3 method overwritten by 'quantmod':

```
method          from
as.zoo.data.frame zoo
> library(feasts)
```

Построим коррелограмму автокорреляционной функции анализируемого временного ряда. Для стационарных процессов автокорреляционная функция зависит только от расстояния между значениями  $t_1, t_2$ , т. е. от величины лага  $\tau$ , а не от места их расположения.

```
y<-ts(tamb_obl$UROV)
> y|>as_tsibble()|>feasts::ACF(lag_max = 17)|>autoplot()+labs(title="Коррелограмма автокорреляционной функции",x="лаг",y="ACF")
Response variable not specified, automatically selected `var = value`
```

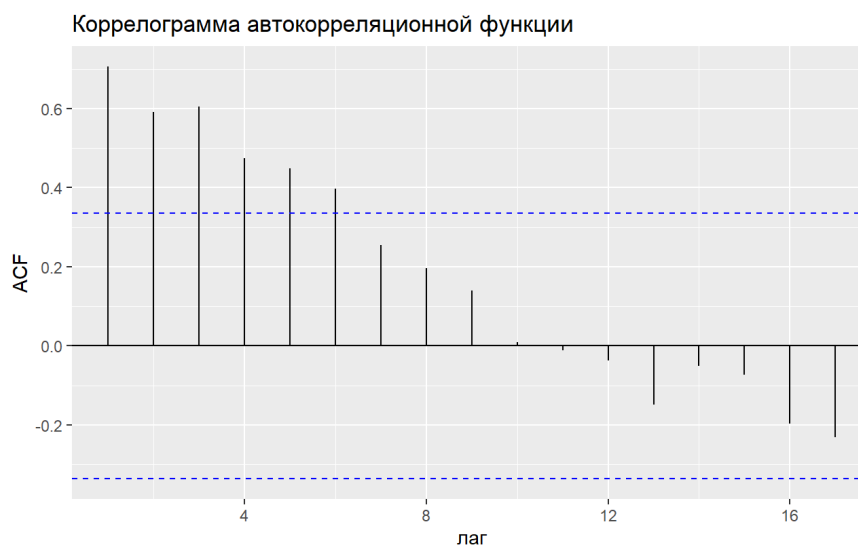


Рисунок 1 – Коррелограмма автокорреляционной функции.

Наряду с выборочным коэффициентом автокорреляции часто используются выборочные частные коэффициенты корреляции между значениями временного ряда  $y_j$  и  $y_{j+\tau}$ , в которых устранено влияние промежуточных (располагающихся между этими членами временного ряда). На основе библиотеки `ggplot2`, применим функцию этой библиотеки `gg_tsdisplay`. Данная функция предназначена для построения матричной диаграммы, состоящей из трех панелей, содержащих отдельные диаграммы. В верхней панели отображается линейная диаграмма, показывающая зависимость уровня временного ряда от времени. В левой нижней панели находится коррелограмма автокорреляционной функции. В правой нижней панели помещена диаграмма, определенная заданным режимом отображения. Мы будем использовать режим построения коррелограммы частной автокорреляционной функции (`partial`).

```
y|> as_tsibble()>gg_tsdisplay(plot_type = "partial")
```

Plot variable not specified, automatically selected `y = value`

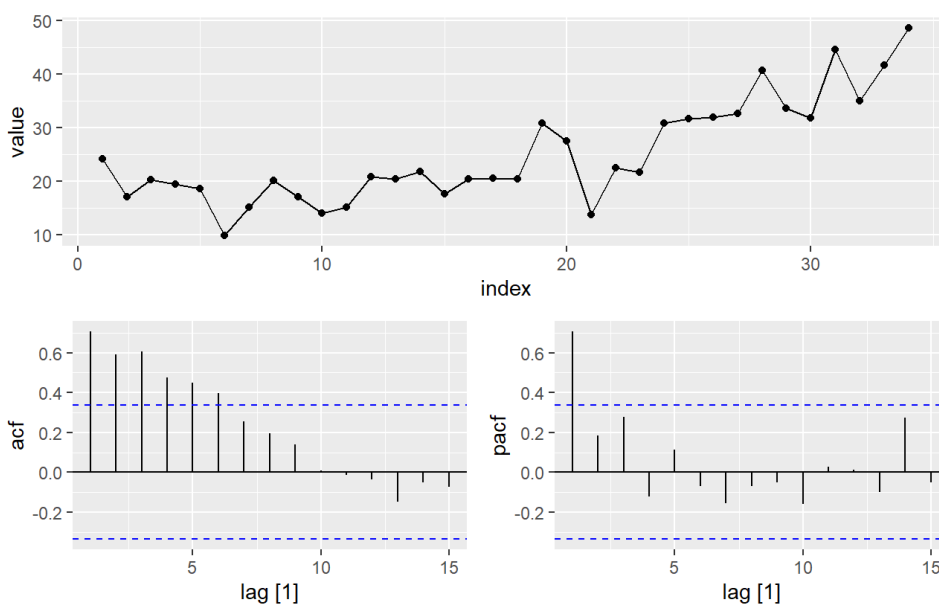


Рисунок 2 – Диаграмма временного ряда.

Для проверки стационарности временных рядов разработан целый ряд специальных статистических критериев, среди которых наиболее популярным является критерий Бокса – Пирса:

$$Q(m) = T \sum_{\tau=1}^m r^2(\tau),$$

где  $m$  – максимальный размер лага;  $\tau$  – величина лага [2].

```
Box.test(y)
```

```
Box-Pierce test
```

```
data: y  
X-squared = 16.965, df = 1, p-value = 3.807e-05
```

Тест показывает, что наблюдаемое значение критерия  $\chi^2 = 16,965$ . Число степеней свободы  $df$  для  $m = 1$  равно 1. Для данных значений критерия и параметра закона распределения уровень значимости  $p$ -value очень мал ( $p$ -value < 3,807e-05. Поэтому нулевая гипотеза должна быть отклонена. Ряд является нестационарным, так как исходный и лагированный на единицу временные ряды зависимы.

Для получения стационарности временного ряда используют так называемую технологию дифференцирования. Однако мы можем отказаться от пошаговой подгонки модели вручную в пользу автоматического способа ее подбора. Выбор целевой модели осуществляется на основе самых разных информационных критериев с помощью функции `auto.arima()` пакета `forecast`:

```
est<-auto.arima(tamb_obl$UROV,stepwise = FALSE,max.p = 5,max.q = 9)
```

```
> est
```

```
Series: tamb_obl$UROV  
ARIMA(0,1,1) with drift
```

```
Coefficients:
```

```
      ma1  drift  
-0.6269  0.7423  
s.e.    0.1438  0.3574
```

```
sigma^2 = 28.92: log likelihood = -101.56
```

```
AIC=209.12  AICc=209.94  BIC=213.61
```

Важным моментом является вычисление коэффициента корреляции между исходным временным рядом и прогнозируемой моделью, построенной с помощью функции `auto.arima()`.

```
cor(tamb_obl$UROV,est$fitted)
```

```
[1] 0.8478889
```

Полученное столь высокое значение коэффициента корреляции (0,8479) свидетельствует о высоком качестве построенной модели. Рассмотрим

графическое сравнение между исходным временным рядом и прогнозируемой моделью

```
plot(tamb_obl$UROV, type="l")  
> lines(fitted(est), col=3, lwd=2)
```



Рисунок 3 – Результаты сглаживания.

Приведенные графики подтверждают адекватность построенной прогнозирующей модели.

### Список литературы:

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Выпуск 1. М.: Мир, 1974. 408с.
2. Наумов В.Н. Методы прогнозирования временных рядов: учебное пособие для вузов. Санкт-Петербург: Лань, 2024. 196 с.
3. Смагин Б.И. Стохастичность функционирования как атрибут аграрной сферы производства // Вестник Мичуринского государственного аграрного университета, 2020. №4. С. 196 – 203.

**UDC 51.76**

**THE USE OF INFORMATION TECHNOLOGIES AND THE  
ALGORITHMIC LANGUAGE R IN FORECASTING THE YIELD OF  
GRAIN CROPS IN THE TAMBOV REGION**

**Boris Ig. Smagin**

doctor of economics, professor

bismagin2023@mail.ru

**Nikolay Al. Barkov**

master's student

barkov-92@bk.ru

Michurinsk State Agrarian University

Michurinsk, Russia

**Abstract.** The article considers the Box-Jenkins methodology for predicting the yield of grain crops in the region based on the use of the algorithmic language R. Conducting a preliminary analysis during the construction of autocorrelation and partial autocorrelation functions makes it possible to identify the features of the time series, and the application of the Box-Pierce criterion makes it possible to diagnose its stationarity. Later, the use of R language packages allowed us to build a high-quality predictive function.

**Keywords:** algorithmic language R, crops, productivity, forecasting, stationarity.

Статья поступила в редакцию 25.02.2026; одобрена после рецензирования 20.03.2026; принята к публикации 31.03.2026.

The article was submitted 25.02.2026; approved after reviewing 20.03.2026; accepted for publication 31.03.2026.