

УДК 519.22:338.43:631.55:633.1

## НЕКОТОРЫЕ ПОДХОДЫ К ПРОГНОЗИРОВАНИЮ УРОЖАЙНОСТИ ЗЕРНОВЫХ КУЛЬТУР

**Борис Игнатьевич Смагин**

доктор экономических наук, профессор

bismagin2023@mail.ru

Мичуринский государственный аграрный университет

г. Мичуринск, Россия

**Аннотация.** В статье показано, что сельскохозяйственное производство является системой, обладающей стохастическим принципом действия. В силу чего, объективный анализ аграрного сектора экономики возможен в рамках вероятностных категорий; при этом наиболее сложным является процесс прогнозирования. На примере сельскохозяйственных организаций Липецкой области была предпринята попытка составить прогноз урожайности зерновых культур с использованием методологии Бокса-Дженкинса и алгоритмического языка R. Построенная прогнозирующая функция обладает высокой степенью надежности, что позволяет использовать ее в целях краткосрочного прогнозирования.

**Ключевые слова:** сельскохозяйственное производство, урожайность, прогнозирование, методы Бокса-Дженкинса, алгоритмический язык R.

Ранее нами было показано, что сельскохозяйственное производство является системой, обладающей стохастическим принципом действия. Следовательно, объективный анализ аграрного сектора экономики возможен в рамках вероятностных категорий [4]. При этом наиболее сложным является процесс прогнозирования функционирования как всего АПК, так и его отдельных отраслей.

Прогнозирование в области сельскохозяйственного производства играет особую роль, т.к. степень риска крайне высока, что в значительной степени обусловлено природно-климатическими условиями и функционированием объектов биологической природы. Следовательно, использование методов прогнозирования, базирующихся на современных информационных технологиях с применением языков программирования, математических и статистических пакетов, является актуальной задачей управления АПК [2].

Основной этап связан непосредственно с получением и анализом прогностических утверждений, которые, как правило основаны на предположении о сохранении в будущем предшествующих тенденций развития. Предполагается, что наблюдения доступны в дискретные, равноотстоящие моменты времени.

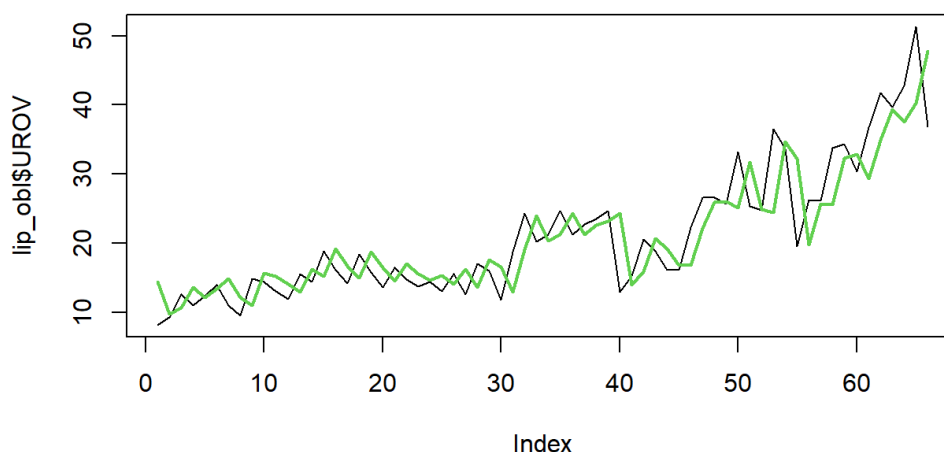
Одна из наиболее важных проблем связана с выбором модели прогнозирования. Мы считаем, что прогноз многих процессов может быть осуществлен на основе использования методологии Бокса-Дженкинса [1]. Если не учитывать сезонные явления, то в общем случае имеется модель авторегрессии проинтегрированного скользящего среднего (ARIMA). ARIMA (p, d, q) называется включает в себя процесс авторегрессии (AR(p)), процесс скользящего среднего (MA(q)), а d – количество различий, сделанных, когда временной ряд становится стационарным временным рядом.

Рассмотрим один частный вопрос, касающийся прогнозирования урожайности зерновых культур. Исходным временным рядом являются данные об урожайности зерновых культур Липецкой области за 1956 – 2021гг. (65

наблюдений). Используя алгоритмический язык R мы обозначили данную совокупность как `lip_obl`.

В качестве первоначального приближения рассмотрим случай прогнозирования с помощью известной (оценочной) AR-модели. Построим прогноз, используя функцию `fitted()` пакета `forecast`. Ниже приведен полный код решения.

```
> library(stats)
> library(forecast)
> lip_obl
plot(lip_obl$UROV,type = 'l')
est<-arima(x=lip_obl$UROV,order = c(1,0,0))
> lines(fitted(est),col=3,lwd=2)
```



```
est
Call:
arima(x = lip_obl$UROV, order = c(1, 0, 0))
Coefficients:
    ar1 intercept
    0.8815  21.3660
s.e. 0.0599  4.4031
sigma^2 estimated as 22.01: log likelihood = -196.42, aic = 398.84
Тем самым получена зависимость вида
```

$$y_t = 21,366 + 0,8815y_{t-1},$$

где  $y_t$  – урожайность в момент времени  $t$ .

Модель ARIMA характеризуется тремя терминами:  $p, d, q$  где,  $p$  – это порядок следования члена AR;  $q$  – порядок следования термина MA;  $d$  –

количество разностей, необходимых для того, чтобы сделать временной ряд стационарным.

```
> fitted(est, h = 3)
library(forecast)
> est<-arima(x=lip_obl$UROV,order = c(3,0,0)); est
Call:
arima(x = lip_obl$UROV, order = c(3, 0, 0))
Coefficients:
      ar1      ar2      ar3 intercept
      0.6069 -0.1264  0.4939  24.0797
s.e.  0.1146  0.1457  0.1173  11.0463
sigma^2 estimated as 16.51: log likelihood = -187.68, aic = 385.36
```

Воспользуемся функцией `fitted()` с дополнительным параметром `h` для установки горизонта прогнозирования. По мере роста временного горизонта прогноз будущих значений приближается к математическому ожиданию ряда, и следовательно, дисперсия как ошибки, так и прогнозных значений будет стремиться к нулю, поскольку прогнозные значения стремятся к безусловному математическому ожиданию.

```
> var(fitted(est,h=3),na.rm = TRUE)
[1] 51.26358
> var(fitted(est,h=5),na.rm = TRUE)
[1] 36.51305
> var(fitted(est,h=10),na.rm = TRUE)
[1] 24.58342
> var(fitted(est,h=20),na.rm = TRUE)
[1] 7.879318
> var(fitted(est,h=30),na.rm = TRUE)
[1] 4.864404
```

Из этого можно сделать важный вывод о том, что модели AR (а также MA, ARMA и ARIMA) лучше всего подходят для краткосрочного прогнозирования. Они теряют предсказательную силу для отдаленных горизонтов в будущем.

#### Подбор параметров процесса $MA(q)$

Шаблоны автокорреляционной функции (ACF) и частной автокорреляционной функции (PACF) различаются в процессах скользящего среднего (MA) и авторегрессии (AR). В отличие от авторегрессионного процесса,

который характеризуется медленно убывающей ACF, процессу MA свойственно резкое обнуление ACF при любом значении порядка, превышающем  $q$ .

Это связано с тем, что авторегрессионный процесс зависит от предыдущих членов, а они учитывают предыдущие импульсы в системе, в то время как модель MA, учитывающая импульсы непосредственно по их значениям, обладает механизмом предотвращения бесконечного влияния.

```
> ma.est=arima(x=lip_obl$UROV,order=c(0,0,9))
> ma.est
Call:
arima(x = lip_obl$UROV, order = c(0, 0, 9))
Coefficients:
ma1  ma2  ma3  ma4  ma5  ma6  ma7  ma8  ma9 intercept
0.5671 0.1181 0.7602 1.1349 0.9440 0.5872 0.3670 0.4102 0.6121
21.3226
s.e. 0.1302 0.1394 0.1223 0.1348 0.1673 0.1419 0.1346 0.1257 0.1356
2.5523
```

sigma<sup>2</sup> estimated as 11.41: log likelihood = -179.99, aic = 381.97

Перейдем теперь к рассмотрению модели авторегрессии проинтегрированного скользящего среднего. Модели ARIMA по-прежнему обеспечивают точность, близкую к современному уровню, особенно на небольших наборах данных, на которых более сложные модели машинного обучения или глубокого обучения выглядят не наилучшим образом. Однако даже модели ARIMA подвержены опасности чрезмерной подгонки, несмотря на их относительную простоту.

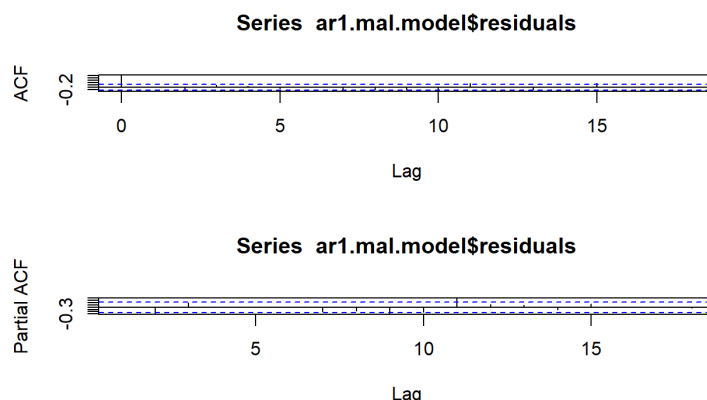
Совмещение моделей AR и MA приводит к получению модели авторегрессионного скользящего среднего (ARMA – Autoregressive Moving Average), которая применяется в случаях, когда ни AR, ни MA не позволяет описать наблюдаемую динамику временного ряда с достаточной степенью точности.

Теорема Вольда утверждает, что каждый ковариационно-стационарный временной ряд можно представить как сумму двух временных рядов – детерминированного и стохастического [3].

Остановимся на вопросе выбора параметров. Модель ARIMA описывается тремя параметрами ( $p, d, q$ ). Значения  $p, d$  и  $q$  выбираются согласно имеющемуся набору данных.

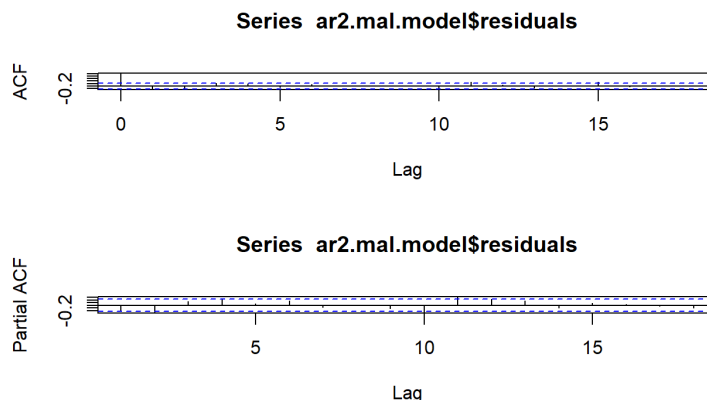
Порядок модели выбирается, исходя из знаний о предметной области, значений оцениваемых критериев (таких, как AIC) и сведений о поведении функций PACF и ACF для исследуемого процесса. Давайте попробуем настроить модель ARIMA как вручную, используя сведения о функциях PACF и ACF, так и автоматически – с помощью программных инструментов выбора параметров модели, – в частности функции *auto.arima()* пакета *forecast*.

```
require(forecast)
> set.seed(1017)
> y=lip_obl$UROV
> ar1.mal.model=Arima(y,order=c(1,0,1))
> par(mfrow=c(2,1))
> acf(ar1.mal.model$residuals)
> pacf(ar1.mal.model$residuals)
```



В следующем коде тестируется модель ARIMA(2, 0, 1) и строятся графики ACF и PACF для невязок улучшенной, более сложной модели.

```
ar2.mal.model=Arima(y,order=c(2,0,1))
> plot(y,type = 'l'); lines(ar2.mal.model$fitted,col=2)
> plot(y,ar2.mal.model$fitted); par(mfrow=c(2,1))
> acf(ar2.mal.model$residuals); pacf(ar2.mal.model$residuals)
```



Ниже приведен еще один, быстрый, способ сравнения моделей, который заключается в изучении корреляции спрогнозированных значений и фактических данных временного ряда.

```
> cor(y,ar1.mal.model$fitted)
[1] 0.8878196
> ar2.mal.model=Arima(y,order=c(2,0,1))
> cor(y,ar2.mal.model$fitted)
[1] 0.8757787
```

Таким образом, наблюдается очень тесная связь между моделями.

Используя преимущества алгоритмического языка R, осуществим автоматическую подгонку модели. В настоящее время мы можем отказаться от пошаговой подгонки модели вручную в пользу автоматического способа ее подбора. Выбор целевой модели осуществляется на основе самых разных информационных критериев (например, AIC) с помощью функции *auto.arima* () пакета *forecast*:

```
> est=auto.arima(lip_obl$UROV,stepwise = FALSE,max.p = 3,max.q = 9)
> est
Series: lip_obl$UROV
ARIMA(2,1,0) with drift

Coefficients:
    ar1    ar2  drift
-0.4273 -0.5392  0.5036
s.e.  0.1133  0.1133  0.2524
sigma^2 = 16.37: log likelihood = -181.92
AIC=371.85 AICc=372.51 BIC=380.54
```

В этом примере функции передаются входные данные, выбор которых основан на результатах предыдущих исследований. В частности, мы указали

наиболее подходящие с нашей точки зрения максимальные значения порядков процессов AR и MA, но программой была выбрана более простая модель, чем предполагалось, – полностью лишенная MA-составляющей.

После построения прогнозирующей функции следует вычислить корреляцию между прогнозируемыми и фактическими значениями.

```
> cor(lip_obl$UROV,est$fitted)
[1] 0.9076161
```

Столь высокий уровень корреляции между прогнозируемыми и фактическими значениями при значительном числе наблюдений (65), позволяет сделать вывод о возможности использования данной модели в целях краткосрочного прогнозирования.

#### Список литературы:

1. Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Выпуск 1. М.: Мир. 1974. 408с.
2. Буховец А.Г., Семин Е.А., Бирючинская Т.Я. Современные подходы и методы в прогнозировании урожайности отдельных видов зерновых культур: монография. Воронеж: ФГБОУ ВО Воронежский ГАУ. 2016. 214с.
3. Нильсен Э. Практический анализ временных рядов. Прогнозирование со статистикой и машинное обучение. СПб.: ООО «Диалектика». 2021. 544с.
4. Смагин Б.И. Стохастичность функционирования как атрибут аграрной сферы производства// Вестник Мичуринского государственного аграрного университета. 2020. №4. С. 196 203.



UDC 519.22:338.43:631.55:633.1

## SOME APPROACHES TO FORECASTING THE YIELD OF GRAIN CROPS

**Boris Ig. Smagin**

doctor of economics, professor

bismagin2023@mail.ru

Michurinsk State Agrarian University

Michurinsk, Russia

**Abstract.** The article shows that agricultural production is a system with a stochastic principle of action. Therefore, an objective analysis of the agricultural sector of the economy is possible within the framework of probabilistic categories; at the same time, the forecasting process is the most difficult. On the example of agricultural organizations of the Lipetsk region, an attempt was made to make a forecast of grain yield using the Box-Jenkins methodology and the algorithmic language R. The constructed predictive function has a high degree of reliability, which allows it to be used for short-term forecasting purposes.

**Keywords:** agricultural production, yield, forecasting, Box-Jenkins methods, algorithmic language.

Статья поступила в редакцию 03.05.2024; одобрена после рецензирования 13.06.2024; принята к публикации 27.06.2024.

The article was submitted 03.05.2024; approved after reviewing 13.06.2024; accepted for publication 27.06.2024.