

СРАВНИТЕЛЬНАЯ ОЦЕНКА МЕТОДОВ СРЕДНЕСРОЧНОГО ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ СЕЛЬСКОХОЗЯЙСТВЕННЫХ КУЛЬТУР

*М.С. ТЕЛЮК, ассистент кафедры теории и социально-экономической статистики,
РЭУ им. Г.В. Плеханова
email: myfranker@gmail.com*

Аннотация

Статья посвящена сравнительной оценке двух наиболее распространенных методов прогнозирования показателей урожайности сельскохозяйственных культур: авторегрессионных моделей – скользящего среднего и множественных линейных регрессий. Автор сравнивает данные методы исходя из качества прогноза на среднесрочную перспективу урожайности сельскохозяйственных культур.

Ключевые слова: урожайность, авторегрессионные модели, множественная линейная регрессия, прогнозирование, скользящее среднее.

Урожай и урожайность – важнейшие результативные показатели растениеводства и сельскохозяйственного производства в целом. Уровень урожайности отражает воздействие экономических и природных условий, в которых осуществляется сельскохозяйственное производство. Установление точных размеров урожайности различных сельскохозяйственных культур возможно только после завершения уборки и оприходования всей полученной продукции. Однако в большинстве случаев сведения об урожае необходимы намного раньше полного созревания сельскохозяйственных культур и начала их уборки. Например, для определения валовых сборов сельскохозяйственных культур в перспективном периоде в Центральном федеральном округе (ЦФО), на основе чего рассчитывается самообеспеченность федерального округа по основным культурам растениеводства на среднесрочный период. Поэтому определение ожидаемых (прогнозных) размеров уро-

жая и урожайности необходимо для принятия управленческих решений [1, 11].

В научной литературе имеется множество методов прогнозирования урожайности сельскохозяйственных культур:

- методы экспоненциального сглаживания и экстраполяции;
- модели прогнозирования – авторегрессионные с интегрированным скользящим средним (ARIMA);
- модели сингулярно-спектрального анализа (SSA);
- модели на основе дистанционных и метеорологических наблюдений и др. [2, 3, 5–10].

Однако в большинстве исследований прогнозирование проводилось без учёта материально-технического оснащения сельскохозяйственного производства, несмотря на то что ориентировочные данные о материально-техническом оснащении при прогнозных значениях урожайности упрощают принятие управленческих решений. В данной работе автор стремится восполнить этот пробел.

Далее в работе будут рассмотрены два метода прогнозирования урожайности:

- 1) множественная линейная регрессия с учётом материально-технического оснащения;
- 2) модель авторегрессии – скользящего среднего.

Целью исследования является выбор наилучшего, т.е. дающего наиболее достоверные результаты, метода прогнозирования урожайности на среднесрочную перспективу по ЦФО.

Исходные данные. Построение моделей производилось на основе данных ЦФО за 1996–2015 гг. Модели строились на основе 1996–2011 гг. (обучающая выборка), а 2012–2015 гг. (тестовая выборка) использовались

Сравнительная оценка методов среднесрочного прогнозирования...

для оценки качества построенных моделей. Значения урожайности рассматриваются по хозяйствам всех категорий. Данные были взяты из системы ЕМИСС¹ за весь доступный период с 1996–2015 гг. по пяти основным сельскохозяйственным культурам, выращиваемым на территории ЦФО (рис. 1).

Методы среднесрочного прогнозирования урожайности. В работе автор рассматривает два наиболее известных метода прогнозирования урожайности для среднесрочного периода (обзор данных методов):

– авторегрессионные модели – скользящего среднего;

– множественная линейная регрессия.

При построении авторегрессионных моделей – скользящего среднего использовались только временные ряды урожайности сельскохозяйственных культур. Методом перебора модель выбиралась из следующего списка [4]:

– «AR» – авторегрессионная модель;

– «MA» – модель скользящего среднего;

– «ARMA» – модель авторегрессии – скользящего среднего;

– «ARIMA» – интегрированная модель авторегрессии – скользящего среднего;

– «SARIMA» – сезонная модель авторегрессии – скользящего среднего;

– «SARIMA» – сезонная интегрированная модель авторегрессии – скользящего среднего;

– «ARCH» – авторегрессионная условная гетероскедастичная модель;

– «GARCH» – обобщённая авторегрессионная условная гетероскедастичная модель.

Переобучение модели является одной из основных проблем, с которыми приходится сталкиваться при её построении на основе коэффициента детерминации. Вследствие этого построение модели основывалось на критерии AIC (информационный критерий Акаике). Данный критерий был выбран потому, что он, в отличие от других (BIC, SBC), содержит функцию штрафа, линейно зависящую от числа параметров. Штраф за каждый последующий добавленный параметр в модель критерий AIC оценивает в равной мере. Поэтому данный критерий был использован для оценки качества модели, чтобы штрафы всех параметров были соразмерны. Качество модели определяется меньшим значением критерия AIC [5].

Все расчёты в данной работе произведены в среде Wolfram Research, Inc., Mathematica, Version 10.2, Champaign, IL (2015).

Построение моделей. Результаты полученных моделей представлены в табл. 1.

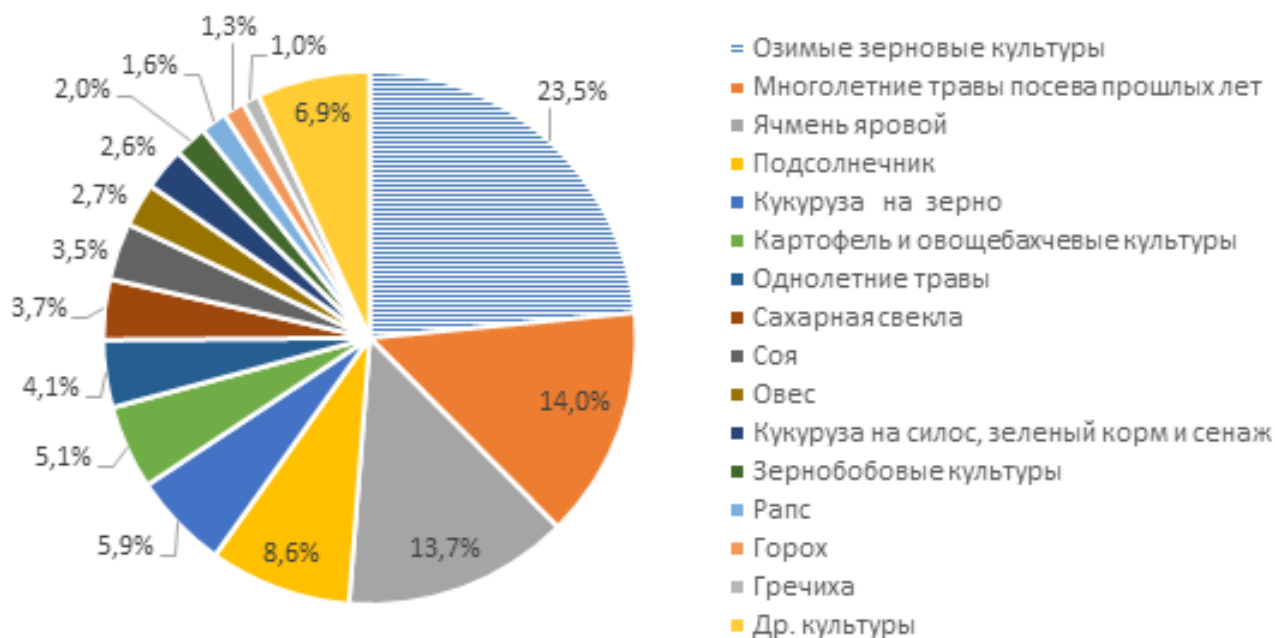


Рис. 1. Структура посевной площади сельскохозяйственной культур в ЦФО (источник: Федеральная служба государственной статистики. URL: <http://www.gks.ru>)

¹ Единая межведомственная информационно-статистическая система. URL: <http://www.fedstat.ru>.

Таблица 1

Результаты построенных авторегрессионных моделей

№ с/х культуры	Модель	AIC
1	ARIMA [2, 1, 1]	54,3297
2	MA [0]	5,15336
3	ARIMA [2, 1, 2]	56,3508
4	ARIMA [1, 1, 1]	38,8514
5	MA [0]	84,7006

Стоит отметить, что для некоторых сельскохозяйственных культур модель упростилась до скользящей средней нулевого порядка. Для удостоверения, что это не ошибка, рассмотрим список из 10 лучших кандидатов для многолетних трав посева прошлых лет (табл. 2).

Таблица 2

Кандидаты моделей для многолетних трав посева прошлых лет

№	Кандидаты	AIC
1	MA[0]	5,1534
2	MA[1]	6,9918
3	AR[1]	7,0346
4	ARMA[1,1]	9,1146
5	ARIMA[0,1,1]	11,6092
6	ARIMA[0,1,2]	12,6210
7	ARIMA[1,1,1]	13,0257
8	ARIMA[1,1,0]	13,2571
9	ARIMA[0,1,0]	14,4788
10	ARIMA[1,1,2]	14,7771

Источник: авторские расчёты

Далее приведены результативные данные при использовании других критериев к построению авторегрессионных моделей – скользящего среднего (см. ниже).

Выбор факторных признаков при построении многомерных линейных регрессий урожайности основывался на следующих показателях:

x1 – число Вольфа;

x2 – внесение сельскохозяйственными организациями калийных удобрений в пересчё-

те на 100% питательных веществ на 1 га посева, кг;

x3 – внесение сельскохозяйственными организациями азотных удобрений в пересчёте на 100% питательных веществ на 1 га посева, кг;

x4 – внесение сельскохозяйственными организациями органических удобрений на 1 га посева, т;

x5 – внесение сельскохозяйственными организациями фосфорных удобрений в пересчёте на 100% питательных веществ на 1 га посева, кг;

x6 – бороны, шт. на га;

x7 – грабли тракторные, шт. на га;

x8 – жатки валковые, шт. на га;

x9 – косилки, шт. на га;

x10 – культиваторы, шт. на га;

x11 – опрыскиватели и опыливатели тракторные, шт. на га;

x12 – плуги, шт. на га;

x13 – пресс-подборщики, шт. на га;

x14 – протравливатели семян, шт. на га;

x15 – разбрасыватели твёрдых минеральных удобрений, шт. на га;

x16 – сеялки, шт. на га;

x17 – тракторные прицепы, шт. на га;

x18 – тракторы, шт. на га;

x19 – комбайны зерноуборочные, шт. на га;

x20 – комбайны картофелеуборочные, шт. на га;

x21 – комбайны кормоуборочные, шт. на га;

x22 – комбайны кукурузоуборочные, шт. на га;

x23 – комбайны льноуборочные, шт. на га;

x24 – свеклоуборочные машины (без ботвоуборочных), шт. на га;

x25 – машины для внесения в почву органических удобрений, шт. на га.

В качестве основного фактора, характеризующего природно-климатические условия возделывания сельскохозяйственной культуры, представлена солнечная активность, которая измеряется числом Вольфа. При прогнозировании урожайности с помощью многомерной линейной регрессии предполагается, что значения факторных признаков заранее известны, так как это данные об имеющемся материально-техническом оснащении сельскохозяйственного производства. Исключение составляет значение числа Вольфа, прогнозное значение которого берётся из прогнозных данных обсерватории¹.

¹ Центр анализа данных по влиянию Солнца (Бельгия). URL: <http://sidc.oma.be>.

Количество параметров (факторов) при построении многомерных регрессий превышает количество значений наблюдений рядов динамики. Поэтому целесообразнее использовать модифицированный (скорректированный) критерий Акаике (AICс) [12] для отбора наилучшей линейной регрессии. Продолжительность ряда динамики, на основе которого производится построение моделей, составляет 16 лет, поэтому более двух факторов в модель не будут включаться для получения корректного результата.

Многомерные линейные регрессии строились методом перебора всех возможных одно- и двухфакторных линейных регрессий. Наилучшей считалась, модель, имеющая наименьшее значение критерия AICс.

Основные показатели, характеризующие отобранные регрессии по пяти основным сельскохозяйственным культурам, выращиваемым в ЦФО, представлены в табл. 3.

Полученное уравнение регрессии можно считать в целом статистически значимым с вероятностью 95%, так как наблюдаемое значение F-критерия Фишера выше критического. Факторы, не имеющие содержательного смысла в характеристике урожайности сельскохозяйственной культуры, убирались автоматически как наименее влияющие на результативный признак. В качестве факторных признаков в большинстве своём выступают данные о технической вооружённости сельхозпроизводства. По некоторым культурам видно малое значение коэффициента детерминации. К примеру, урожайность «кукурузы на зерно» объясняется зависимостью от фактора «х1 – жатки валковые, шт. на га» всего на 45%.

Сравнение прогнозных значений. Для получения более достоверных данных по ка-

честву прогноза указанные методы были применены ко всем сельскохозяйственным культурам, по которым имеются данные с 1996–2015 гг. в ЕМИСС. Результаты прогнозирования значений урожайности двумя методами по всем 27 сельскохозяйственным культурам¹ в сравнении с тестовой выборкой представлены на рис. 2.

Прогнозные значения модели за 2012–2014 гг. изображены на оси абсцисс, а на оси ординат – фактические значения за аналогичный период. Если прогнозное значение модели совпадает с фактическим, точечное значение должно находиться на биссектрисе $x = y$, т.е. на линии идеальных значений. Чем ближе значение к прямой $x = y$, тем более качественно модель даёт прогнозные результаты. Из рис. 2 видно, что наиболее точные результаты дают авторегрессионные модели – скользящего среднего по сравнению с множественной линейной регрессией. Для более наглядных результатов были рассчитаны среднеквадратические значения ошибок моделей на тестовой выборке, на основе которых было рассчитано среднее значение коэффициента вариации по всем сельскохозяйственным культурам:

$$\bar{v} = \frac{\sum_{i=1}^n \frac{\sigma}{\bar{y}_i}}{n},$$

где σ – среднеквадратическое отклонение ошибки по тестовой выборке; \bar{y}_i – среднее значение урожайности сельскохозяйственной культуры за весь рассматриваемый период; n – число сельскохозяйственных культур (в данном случае 27).

В табл. 4 представлены значения среднеквадратических ошибок при использовании следующих критериев:

Таблица 3

Результаты построения множественных линейных регрессий

№ С/Х культуры	Факторы, включённые в модель	R ²	AICс	F
1	19,9582 – 175871 x7 + 122847 x9	0,718733	88,7713	16,7814
2	20,7711 – 139809 x7 + 94678,7 x9	0,510561	101,544	4,48654
3	9,67058 + 3513,55 x17 – 19224,2 x7	0,496347	47,2219	12,7194
4	17,82 – 1787,31 x17	0,579314	77,2452	16,865
5	43,9504 – 25536,3 x8	0,450054	124,54	11,4571

1 По всем культурам, представленным в ЕМИСС с 1996 г. URL: <https://www.fedstat.ru>.

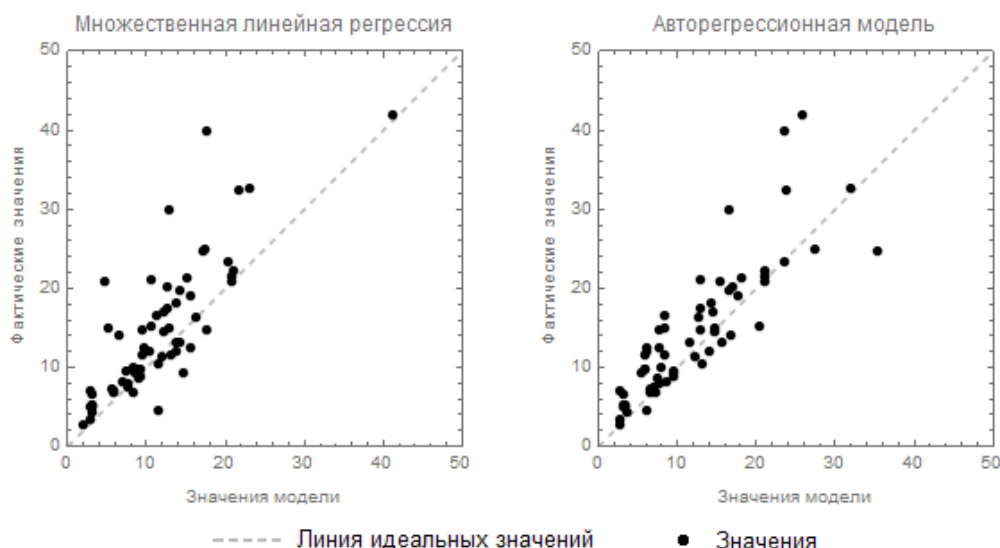


Рис. 2. Сравнение результатов прогноза с тестовой выборкой (2 метода)

Таблица 4

Среднее значение коэффициента вариации ошибки модели при разных критериях

Модель	AIC	AICc	BIC	SBC
Авторегрессионные модели – скользящего среднего	0,421628	0,50822	$2,27784 \times 10^{22}$	0,476775
Множественная линейная регрессия	0,467078	0,445283	0,45454	0,612057

1) AIC – информационный критерий Акаике;

2) AICc – скорректированный критерий Акаике;

3) BIC – Байесовский информационный критерий;

4) SBC – Байесовский критерий Шварца.

Выбранные критерии в построенных моделях дают наименьшую ошибку в тестовой выборке, что свидетельствует о правильности выбранных критериев.

В итоге стоит отметить, что авторегрессионные модели – скользящего среднего дают наименьшее среднее значение коэффициента вариации ошибки на тестовой выборке по всем рассмотренным сельскохозяйственным культурам.

Итак, в работе были рассмотрены два метода прогнозирования урожайности на среднесрочную перспективу. Был выбран наилучший метод, дающий наименьшую относительную ошибку при прогнозировании на среднесрочную перспективу на уровне ЦФО.

Авторегрессионные модели – скользящего среднего, построенные на основе данных материально-технического оснащения, де-

монстрируют более качественные результаты по сравнению с множественной линейной регрессией. Стоит отметить, что это утверждение верно только при прогнозировании на среднесрочную перспективу и на макроуровне (т.е. с использованием данных на уровне федерального округа). Но это не свидетельствует о малоприменимости множественной регрессии при прогнозировании урожайности на макроуровне. Такие результаты были получены, по мнению автора, из-за короткого известного временного ряда урожайности, имеющегося в базе данных ЕМИСС, вследствие чего не удалось построить более качественные множественные регрессии и достаточно достоверно оценить влияние материально-технического оснащения на урожайность сельскохозяйственных культур.

Библиографический список

1. Афанасьев В.Н., Маркова А.И. Статистика сельского хозяйства: учеб. пособие. М., 2003.
2. Горбачев И.В., Панова Т.В., Панов М.В. Прогнозирование урожайности зерновых и зернобобовых культур в хозяйствах ЦФО РФ

до 2020 года // Вестник Брянской государственной сельскохозяйственной академии. 2014. № 2. С. 43–46.

3. *Иванов П.В., Ткаченко И.В.* Экономико-математическое моделирование в АПК. Ростов н/Д, 2013.

4. *Канторович Г.Г.* Лекции по курсу «Анализ временных рядов» // Экономический журнал ВШЭ. 2002. №1–4, 2003. №1.

5. *Низомов С.С.* Применение методов корреляционно-регрессионного и кластерного анализа при прогнозировании урожайности зерновых культур // Гуманитарные и социальные науки. 2014. № 2. С. 768–772.

6. *Низомов С.С.* Применение методов статистического моделирования при прогнозировании урожайности зерновых культур // Вестник БГАУ. 2014. №2. С. 132–135.

7. *Петрова И.Ф., Свешникова И.В., Малых Н.И.* Прогнозирование развития и размещения зернового хозяйства в стране // Экономика сельского хозяйства России. 2015. № 5. С. 68–73.

8. *Семин Е.А.* Совершенствование способов прогнозирования урожайности зерновых

культур: автореф. дис. ... канд. экон. наук. Воронеж, 2015.

9. *Симоненко Е.И.* Методология среднесрочного и долгосрочного прогнозирования урожайности и валового сбора зерновых культур // Актуальные проблемы гуманитарных и естественных наук. 2014. № 9. С. 182–184.

10. *Тихонов В.Е., Неверов А.А.* Долгосрочное прогнозирование урожайности полевых культур на основе планетно-солнечно-земных связей в степном Предуралье // Бюллетень Оренбургского научного центра УрО РАН. 2014. № 4. С. 1–14.

11. *Ткаченко И.В., Ткаченко Н.И.* Экономико-стохастическое моделирование как аппарат системного подхода к управлению сельскохозяйственными предприятиями // Вестник Волжского университета имени В.Н. Татищева. 2014. № 2(31). С. 109–116.

12. *Burnham K.P., Anderson D.R.* Multimodel inference: understanding AIC and BIC in Model Selection // Sociological Methods & Research. 2004. Vol. 33, pp. 261–304.