

УДК 330.4:338.43:637.1

ПРОГНОЗИРОВАНИЕ ОБЪЕМОВ ПРОИЗВОДСТВА МОЛОКА НА ОСНОВЕ СЕЗОННОЙ ARIMA-МОДЕЛИ

Заяц О.А.*ФГБОУ ВО «Волгоградский государственный аграрный университет», Волгоград,
e-mail: OlgaAZ15@Gmail.com*

В статье для получения прогноза выполнено эконометрическое моделирование временного ряда квартального производства молока в РФ за период с 2004 по 2018 г. Анализ графика исходного ряда и диаграммы автокорреляционной функции показал, что ряд нестационарен и содержит сезонную компоненту. В качестве основной методологии анализа и прогнозирования была использована модель авторегрессии интегрированного скользящего среднего (модель Бокса – Дженкинса, ARIMA-модель), которая позволила привести нестационарный временной ряд к стационарному виду путем взятия сезонной разности между значениями ряда. При исследовании нескольких различных моделей с целью получения значимых коэффициентов и остатков со свойствами «белого шума» для временного ряда производства молока были получены следующие модели: $ARIMA(1,0,1)(1,1,0)_4$, $ARIMA(2,0,0)(1,1,0)_4$, $ARIMA(2,0,1)(1,1,0)_4$. По показателям качества и точности все построенные ARIMA-модели одинаково хорошо соответствуют данным на ретроспективном участке. Прогноз объемов производства молока на 2019–2020 гг. был рассчитан путем объединения прогнозов, полученных по всем трем моделям. Для статистической обработки данных, построения ARIMA-моделей и анализа остатков были использованы следующие инструментальные средства: Statistica 10.0 и Econometric Views 3.1.

Ключевые слова: прогноз, моделирование, временной ряд, ARIMA-модели, модель авторегрессии интегрированного скользящего среднего, динамика производства молока, сезонные колебания

FORECASTING OF MILK PRODUCTION ON THE BASIS OF THE SEASONAL ARIMA-MODEL

Zayats O.A.*Volgograd State Agrarian University, Volgograd, e-mail: OlgaAZ15@Gmail.com*

In order to obtain a forecast it was designed an econometric model of the time series of quarterly milk production in the Russian Federation for the period from 2004 to 2018. Analysis of the initial series graph and the autocorrelation function diagram showed that the series was non-stationary and contained a seasonal component. As the main methodology of analysis and forecasting, the AutoRegressive Integrated Moving Average model (Box-Jenkins model, ARIMA-model) was used, which allowed the non-stationary time series to be brought into a stationary form by taking the seasonal difference between the series values. In the study of several different models in order to obtain significant coefficients and residues with the properties of «white noise» for the time series of milk production, the following models were obtained: $ARIMA(1,0,1)(1,1,0)_4$, $ARIMA(2,0,0)(1,1,0)_4$, $ARIMA(2,0,1)(1,1,0)_4$. Properties of ARIMA-built models were equally well in line with the data on the retrospective plot. The forecast of milk production volumes for 2019-2020 was calculated by combining the forecasts obtained for all three models. The following tools were used for statistical data processing, plotting ARIMA models and analyzing residues: Statistica 10.0 and Econometric Views 3.1.

Keywords: forecast, modeling, time series, ARIMA-models, model of autoregression integrated moving average, dynamics of milk production, seasonal fluctuations

Развитие молочной отрасли является одним из приоритетных направлений государственной политики в сфере АПК. Производство молока в России на протяжении ряда лет относительно стабильно, однако наблюдается существенное сокращение поголовья коров, которое компенсируется ростом продуктивности молочного стада. В 2018 г. производство молока в хозяйствах всех категорий составило 30,6 млн т, что на 1,5% больше чем в 2017 г. За последние пять лет производство молока увеличилось на 2,6%, а по отношению к 2004 г. – снизилось на 3,8%. В коммерческом секторе (сельскохозяйственные организации и КФХ) производство молока имеет устойчивую тенденцию к росту (в среднем на 215 тыс. т ежегодно),

а в хозяйствах населения, наоборот, сокращается (в среднем на 365 тыс. т в год). По состоянию на 2018 г. производство в хозяйствах населения составило 11,9 млн т или 38,7% от общего объема.

Проблема прогнозирования социально-экономических процессов, отражающих перспективы развития сельскохозяйственных отраслей страны и ее регионов, является очень важной. В значительной степени это обусловлено тем, что возрастает число предприятий и организаций, эффективность деятельности которых непосредственно зависит от способности предвидеть развитие событий. Большинство методов, используемых для построения прогнозных моделей, исходят из следующих предполо-

жений: основные тенденции и зависимости, наблюдавшиеся в прошлом, сохраняются или можно предсказать направление их изменения в будущем; процессы имеют вероятностный характер и развитие исследуемого объекта определяется суммарным влиянием закономерности и случайности.

В практике прогнозирования значительного количества процессов различной природы применяются модели временных рядов. Структура экономических временных рядов может быть настолько сложной, что в результате моделирования трендовой и сезонной компоненты традиционными методами в остаточном ряду остаются статистические зависимости, которые можно моделировать. Для прогнозирования таких процессов используются модели авторегрессии скользящего среднего (ARMA-модели), а в случае нестационарности временного ряда – модели авторегрессии проинтегрированного скользящего среднего, разработанные Дж. Боксом и Г. Дженкинсом (ARIMA-модели) [1].

Материалы и методы исследования

Исследование базируется на использовании статистических данных поквартальных объемов производства молока в Российской Федерации за период с 2004 по 2018 г. Целью работы является построение нескольких адекватных ARIMA-моделей и выбор наиболее подходящей по показателям качества и точности модели для прогнозирования будущих значений исследуемого временного ряда.

В общем виде модель ARIMA(p, d, q) выражается формулой

$$\Delta^k y_t = \alpha_1 \Delta^k y_{t-1} + \dots + \alpha_p \Delta^k y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

или

$$\Phi_p(L)(1-L)^d y_t = \Theta_q(L)\varepsilon_t,$$

где $\Delta^k y_t$ – k -я последовательная разность уровней y_t ;

ε_t – «белый шум»;

p и q – порядок авторегрессии (AR) и порядок скользящей средней (MA);

d – порядок интегрируемости;

L – оператор сдвига, т.е. преобразование ряда, смещающее его на один временной такт;

$\Phi_p(L)$ и $\Theta_q(L)$ – функции операторов лага соответствующих AR(p) и MA(q) процессов.

Модель Бокса – Дженкинса может быть интерпретирована как модель множественной линейной регрессии, в которой в качестве факторных переменных выступают предшествующие значения зависимой

переменной, а в качестве регрессионного остатка – скользящие средние из элементов «белого шума».

Использование ARMA-моделей предполагает стационарность временных рядов, а многие экономические временные ряды нестационарны. В большинстве случаев для приведения ряда к стационарному виду достаточно исключить детерминированные компоненты (тренд, периодичность) или применить операцию взятия разности. Способ приведения нестационарного временного ряда к стационарному зависит от того, с каким типом нестационарности мы имеем дело (например, содержит ли ряд детерминистический или стохастический тренд). Во многих случаях получить стационарные временные ряды позволяет взятие d -й последовательной разности: $\Delta^d y_t = (1-L)^d y_t$. Если ряд становится стационарным после d -кратного взятия разностей, процесс называется интегрируемым d -го порядка. Для сезонных временных рядов с длиной периода s применяется взятие конечных разностей с лагом s :

$$\Delta_s y_t = y_t - y_{t-s} = (1-L^s) y_t,$$

где L^s – оператор сдвига на s периодов. Наиболее распространенным и эффективным из простых тестов на стационарность и порядок интегрируемости является обобщенный тест Дики – Фуллера [2].

Интегрированная модель авторегрессии скользящего среднего (ARIMA-модель) может достаточно хорошо описывать поведение нестационарных временных рядов, в том числе содержащих сезонную компоненту. Модели ARIMA в значительной степени подходят для прогнозирования временных рядов, характеризующих сельскохозяйственное производство, так как эти ряды в силу специфики влияющих факторов являются стохастическими. К тому же временные ряды параметров сельскохозяйственного производства часто являются нестационарными [3, 4].

Результаты исследования и их обсуждение

Анализируемый временной ряд объемов производства молока, как и многие экономические процессы в сельском хозяйстве, обладает сезонной периодичностью. Традиционные подходы моделирования временных рядов с периодическими колебаниями (расчет значений периодической компоненты и построение аддитивной или мультипликативной модели; использование рядов Фурье; применение моделей с переменной структурой [1]), как правило, не являются экономичными в том смысле, что

модель может содержать слишком много параметров. Уменьшить число параметров модели временного ряда с периодичностью возможно путем учета при построении прогнозной модели взаимосвязей между уровнями ряда, разделенными периодом колебаний, т.е. построения сезонной модели $ARIMA(p, d, q)(P, D, Q)_s$:

$$\Phi_p(L)\Phi_q^*(L^s)(1-L^s)^D(1-L)^d y_t = \Theta_q(L)\Theta_q^*(L^s)\epsilon_t,$$

где P и Q – порядок сезонной авторегрессии и порядок сезонной скользящей средней; D – порядок сезонной разности.

Определение порядков сезонной авторегрессии $AR(P)$ и сезонного скользящего среднего $MA(Q)$ основано на исследовании диаграмм автокорреляционной (АКФ) и частной автокорреляционной (ЧАКФ) функций, при этом все типичные проявления связаны с сезонными лагами [2, 5].

На этапе идентификации $ARIMA$ -модели необходимо, чтобы ряд первоначально нестационарный стал стационарным. Анализ графика исходного временного ряда, а также диаграмм АКФ и ЧАКФ (рис. 1) дает основания предположить, что ряд нестационарен. Значение автокорреляционной функции практически не убывает по мере роста лагов.

Для подтверждения наших предположений относительно нестационарности анализируемого ряда используем обобщенный тест Дики – Фуллера. При проведении теста необходимо решить проблему включения в тестовое уравнение константы, тренда и числа добавок (лаговых значений зависимой переменной). Так как в ряду производства молока за анализируемый период тренд

отсутствует, то нет смысла включать его в тестовое уравнение. Исходя из специфики данных (квартальные), в модель включается четыре лага. Критическое значение t -статистики для модели с константой составляет $t_{0,05} = -2,91$. Так как полученное значение t -критерия Стьюдента ($-1,89$) превышает критическое, с вероятностью 0,95 принимается гипотеза о нестационарности ряда.

Анализируемый временной ряд производства молока содержит ярко выраженную сезонную компоненту с периодом, равным четырем кварталам. Для приведения временного ряда к стационарному виду ряд был преобразован с помощью процедуры взятия сезонных разностей с лагом 4. Тест Дики – Фуллера для сезонной разности ряда показал, что значение t -статистики на 5%-ном уровне значимости равно $-3,92$ меньше критического ($t_{0,05} = -1,95$), что позволяет сделать вывод о том, что ряд сезонных разностей с вероятностью 95% стационарен, т.е. исходный ряд имеет первый порядок сезонной интегрируемости: $D = 1$.

Для выявления наиболее подходящих для описания исследуемого временного ряда моделей были проанализированы различные комбинации параметров p, q, P, Q . Выбор значений порядков авторегрессии и скользящего среднего осуществляли на основе анализа автокорреляционной и частной автокорреляционной функций (рис. 2) [2, 6].

Среди адекватных моделей были выбраны три наиболее подходящие сезонные модели $ARIMA(p, d, q)(P, D, Q)_s$, результаты оценивания которых представлены в табл. 1. Параметры $ARIMA$ -моделей оценивались методом максимального правдоподобия.

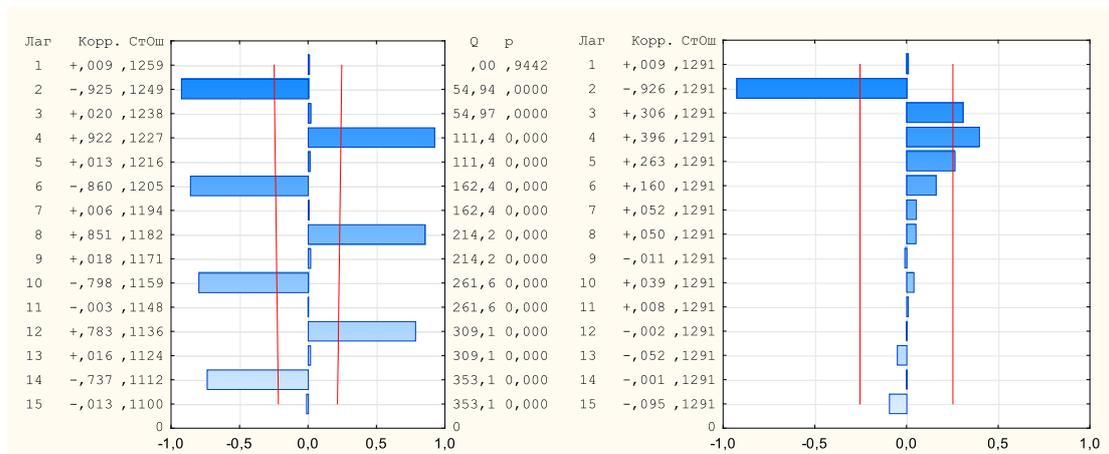


Рис. 1. Автокорреляционная и частная автокорреляционная функции ряда

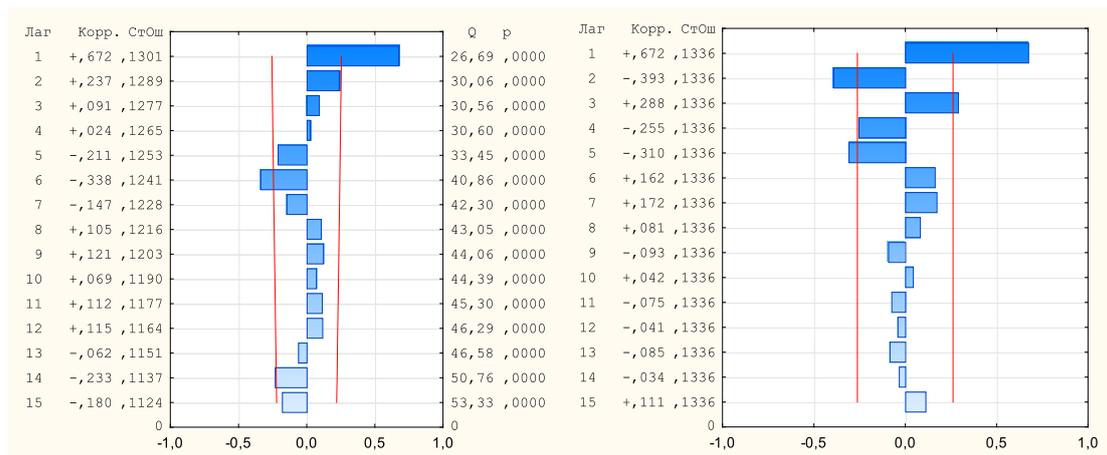


Рис. 2. Автокорреляционная и частная автокорреляционная функции сезонных разностей

Таблица 1

Результаты оценивания ARIMA-моделей

Переменная	Коэффициент	Стандартная ошибка	t-статистика	Значимость t-статистики
ARIMA(1,0,1)(1,1,0) ₄				
AR(1)	0,5185	0,1345	3,86	0,0003
MA(1)	-0,6288	0,1118	-5,62	0,0000
SAR(1)	0,3106	0,1423	2,18	0,0335
ARIMA(2,0,0)(1,1,0) ₄				
AR(1)	1,1503	0,1224	9,40	0,0000
AR(2)	-0,5652	0,1205	-4,69	0,0000
SAR(1)	0,4739	0,1328	3,57	0,0008
ARIMA(2,0,1)(1,1,0) ₄				
AR(1)	1,4657	0,1313	11,16	0,0000
AR(2)	-0,7965	0,1054	-7,56	0,0000
MA(1)	0,4824	0,2052	2,35	0,0225
SAR(1)	0,5836	0,1223	4,77	0,0000

По t -критерию Стьюдента все параметры моделей статистически значимы при 5%-ном уровне значимости. Для подтверждения того, что построенные модели адекватно описывают исходный временной ряд, проанализируем их остатки. Проверим некоррелированность остатков с помощью теста Бокса – Льюнга [1] (табл. 2).

Так как для разного числа лагов значимость Q -статистики больше 0,05, с вероятностью 0,95 автокорреляция в остатках всех трех ARIMA-моделей отсутствует.

Для проверки нормальности остатков используем тест Харке-Бера (табл. 3). Этот тест вычисляет для остатков модели выборочные значения коэффициентов асимметрии A и эксцесса E . При условии

нормальности распределения, статистика Харке-Бера

$$JB = \frac{(n - p - q - 1)}{6} \left[A^2 + \frac{1}{4}(E - 3)^2 \right]$$

имеет χ^2 распределение с двумя степенями свободы.

Распределение остатков для всех моделей соответствует нормальному, так как значимость статистики Харке-Бера больше 0,05. Таким образом, анализ остаточной последовательности свидетельствует об адекватности построенных ARIMA-моделей.

Общими показателями качества модели ARIMA являются критерий Акайка и байесовский критерий Шварца, основанные на

принципе снижения остаточной суммы квадратов при добавлении значимого фактора. В табл. 4 для отбора наиболее подходящей для прогнозирования модели представлены критерии Акайка и Шварца, а также показатели точности: средняя абсолютная процентная ошибка и средняя квадратическая ошибка [5].

В результате анализа данных табл. 4 установлено, что построенные ARIMA-модели демонстрируют высокий уровень аппроксимации на фактических данных. Сравнительная оценка моделей на основе

информационных критериев и показателей точности приводит к выводу о целесообразности использования для прогнозирования всех трех моделей с последующим усреднением индивидуальных прогнозов. Такой способ объединения прогнозов, полученных от прогнозирования по различным моделям, способствует улучшению результатов, так как учитывается больше информации. В табл. 5 приводится результат расчета прогнозных значений поквартальных объемов производства молока в России на 2019–2020 гг.

Таблица 2

Проверка остатков на наличие автокорреляции в остатках

Лар	ARIMA(1,0,1)(1,1,0) ₄			ARIMA(2,0,0)(1,1,0) ₄			ARIMA(2,0,1)(1,1,0) ₄		
	АКФ	Q-стат.	p	АКФ	Q-стат.	p	АКФ	Q-стат.	p
1	0,052	0,160	0,6890	-0,038	0,086	0,7698	0,043	0,110	0,7405
2	0,017	0,177	0,9152	-0,130	1,098	0,5775	-0,145	1,369	0,5042
3	-0,095	0,727	0,8669	0,148	2,442	0,4859	0,140	2,573	0,4623
4	-0,080	1,126	0,8901	-0,096	3,019	0,5547	-0,064	2,828	0,5871
5	-0,207	3,855	0,5705	-0,047	3,160	0,6753	0,060	3,055	0,6915
6	-0,247	7,810	0,2524	-0,186	5,414	0,4919	-0,051	3,225	0,7801
7	-0,133	8,974	0,2546	-0,119	6,359	0,4985	-0,079	3,641	0,8200
8	0,213	12,032	0,1498	0,131	7,518	0,4819	0,061	3,896	0,8664
9	0,074	12,415	0,1910	0,000	7,518	0,5833	-0,028	3,950	0,9147
10	0,105	13,189	0,2134	0,067	7,835	0,6450	0,043	4,081	0,9436
11	0,051	13,379	0,2693	0,011	7,844	0,7272	-0,042	4,209	0,9634
12	0,102	14,149	0,2913	0,074	8,245	0,7657	-0,027	4,263	0,9782
13	0,023	14,189	0,3608	0,073	8,642	0,7994	-0,004	4,265	0,9880
14	-0,240	18,638	0,1793	-0,260	13,883	0,4585	-0,228	8,295	0,8734
15	-0,017	18,661	0,2296	0,051	14,088	0,5189	0,103	9,143	0,8699

Примечание. Q-стат. – статистика Бокса – Льюнга; p – оценка значимости Q-статистики.

Таблица 3

Результаты теста Харке-Бера для проверки нормальности распределения

Показатель	ARIMA(1,0,1)(1,1,0) ₄	ARIMA(2,0,0)(1,1,0) ₄	ARIMA(2,0,1)(1,1,0) ₄
Коэффициент асимметрии	0,1271	0,1145	0,0684
Коэффициент эксцесса	3,0309	2,7297	2,8116
Статистика Харке-Бера	0,1530	0,2929	0,1265
Значимость статистики Харке-Бера	0,9264	0,8638	0,9387

Таблица 4

Сравнительная оценка точности и адекватности ARIMA-моделей

Показатель	ARIMA(1,0,1)(1,1,0) ₄	ARIMA(2,0,0)(1,1,0) ₄	ARIMA(2,0,1)(1,1,0) ₄
Критерий Акайка	12,46	12,42	12,42
Критерий Шварца	12,57	12,53	12,56
Средняя абсолютная процентная ошибка, %	1,6	1,6	1,6
Средняя квадратическая ошибка	119,78	117,62	116,22

Таблица 5

Результат прогнозирования объемов производства молока в РФ, тыс. т

Год Квартал	2019 г.				2020 г.			
	I	II	III	IV	I	II	III	IV
ARIMA(1,0,1)(1,1,0) ₄								
Прогнозное значение	6538	8757	8803	6854	6586	8787	8817	6871
Нижняя граница, 95%	6297	8392	8410	6455	6053	8131	8132	6178
Верхняя граница, 95%	6778	9123	9195	7254	7120	9442	9502	7563
ARIMA(2,0,0)(1,1,0) ₄								
Прогноз. значение	6498	8742	8786	6854	6537	8770	8797	6880
Нижняя граница, 95%	6262	8383	8385	6449	6028	8168	8162	6244
Верхняя граница, 95%	6734	9102	9188	7259	7045	9373	9431	7515
ARIMA(2,0,1)(1,1,0)								
Прогноз. значение	6463	8696	8744	6842	6490	8723	8760	6893
Нижняя граница, 95%	6230	8369	8384	6480	6016	8193	8220	6351
Верхняя граница, 95%	6696	9023	9104	7204	6963	9253	9301	7434
«Усредненный» прогноз								
Прогноз. значение	6500	8732	8778	6850	6538	8760	8791	6881
Нижняя граница, 95%	6263	8381	8393	6461	6033	8164	8171	6258
Верхняя граница, 95%	6736	9083	9162	7239	7042	9356	9411	7504

Согласно долгосрочному прогнозу Минэкономразвития [7] производство молока в РФ в ближайшие годы будет развиваться умеренными темпами – к 2024 г. прирост составит 3,8% по отношению к 2018 г. При этом основной прирост обеспечит корпоративный сектор, где в последние годы отмечается рост инвестиционных вложений и расширение использования усовершенствованных технологий производства молока. Прогнозные расчеты, проведенные с использованием разработанных ARIMA-моделей (табл. 5), показывают, что производство молока в России в ближайшей перспективе будет увеличиваться, но меньшими темпами. Так, согласно «усредненному» прогнозу в 2019 г. ожидается увеличение производства молока на 0,7%, а в 2020 г. – на 1,1% по отношению к уровню 2018 г.

Выводы

Применение стохастических сезонных моделей ARIMA является достаточно эффективным методом краткосрочного и среднесрочного прогнозирования временных рядов с периодичностью, характеризующей изменение показателей сельскохозяйственного производства. Модели авторегрессии проинтегрированного скользящего среднего с сезонностью позволяют достаточно точно аппроксимировать широкий класс случайных процессов, нестационарность которых обусловлена наличием во временных рядах периодической составляющей. Основным

недостатком моделей класса ARMA является то, что при добавлении к исходному ряду новой информации, необходимо корректировать модель. При этом значения порядков AR и MA процессов могут значительно измениться, что приведет к построению совершенно другой модели.

Список литературы

1. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование. М.: Финансы и статистика, 2010. 320 с.
2. Елисева И.И., Курьшева С.В., Костеева Т.В., Пантина И.В., Михайлов Б.А., Нерадовская Ю.В., Штрое Г.Г., Бартелс К., Рыбкина Л.Р. Эконометрика. М.: Финансы и статистика, 2006. 576 с.
3. Реймер В.В., Улезько А.В., Тютюников А.А. Прогноз развития сельского хозяйства Амурской области // Научный журнал КубГАУ – Scientific Journal of KubSAU. 2015. № 114. [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/prognoz-razvitiya-selskogo-hozyaystva-amurskoy-oblasti> (дата обращения: 03.06.2019).
4. Рудой Е.В., Артюх К.А., Федяев П.М. Прогноз производства молока в Кемеровской области на основе моделирования // Достижения науки и техники АПК. 2017. № 2. С. 77–80.
5. Дуброва Т.А. Прогнозирование социально-экономических процессов. М.: Маркет ДС, 2010. 192 с.
6. Рогачев А.Ф. Системный анализ и прогнозирование временных рядов урожайности на основе автокорреляционных функций и нейросетевых технологий // Известия Нижневолжского агроуниверситетского комплекса: Наука и высшее профессиональное образование. 2018. № 3 (51). С. 277–285.
7. Прогноз социально-экономического развития Российской Федерации на период до 2036 года [Электронный ресурс]. URL: <http://economy.gov.ru/minec/about/structure/depmacro/201828113> (дата обращения: 03.06.2019).