УДК 336.01

ПРИЛОЖЕНИЯ МОДЕЛИ АВТОРЕГРЕССИИ И ПРОИНТЕГРИРОВАННОГО СКОЛЬЗЯЩЕГО СРЕДНЕГО (АРПСС) В ЭКОНОМИЧЕСКИХ ПРОЦЕССАХ

[1]Половников Д.С., [1,2]Колпаков И.Ю.

[1]ФГБОУ ВО «Пермский национальный исследовательский политехнический университет», Пермь, e-mail: polovnikov.161@mail.ru;

[2]ФГБОУ ВО «Пермский государственный национальный исследовательский университет», Пермь, e-mail: kolpakov.ilia@mail.ru

Исследование экономических процессов предполагает создание и анализ моделей динамических систем, для которых оказываются неэффективными методы, применяемые в стационарных эконометрических моделях. В рамках теории Бокса и Дженкинса были предложены модели анализа временных рядов, удовлетворяющих условиям стационарности. Для таких рядов со сдвигом времени не меняется их функция плотности вероятности, математическое ожидание и дисперсия. Также было показано, что текущее значение ряда можно представить линейной комбинацией его прошлых возмущений. Для описания стохастических стационарных процессов применяется общая линейная модель, частными случаями с конечным числом параметров которой являются модели скользящего среднего и авторегрессии, более применимые на практике. Дальнейшее обобщение моделей для временных рядов, описывающих реальные процессы и не являющихся стационарными – модели авторегрессии и проинтегрированного скользящего среднего – имеет широкую область применения. Ввиду сложности алгоритма и необходимости обработки большого количества статистических данных модель реализуют в математических и статистических программных пакетах. В статье представлена реализация модели АРПСС в ППП Statistica и ее применение в прогнозировании данных средней заработной платы граждан Российской Федерации.

Ключевые слова: временные ряды, стационарность, модель авторегрессии, модель скользящего среднего, АРПСС, прогнозирование

APPLICATIONS OF AUTOREGRESSIVE AND INTEGRATED MOVING AVERAGE (ARIMA) MODEL IN ECONOMIC PROCESSES

[1]Polovnikov D.S., [1,2]Kolpakov I.Yu.

[1]Perm National Research Polytechnic University, Perm, e-mail: polovnikov.161@mail.ru;

[2]Perm State University, Perm, e-mail: kolpakov.ilia@mail.ru

The study of economic processes involves the creation and analysis of models of dynamic systems for which the methods used in stationary econometric models are ineffective. In the framework of the theory of Box and Jenkins, models of the analysis of time series that satisfy the stationarity conditions were proposed. For such series with a time shift, their probability density function, expectation, and variance do not change. It was also shown that the current value of the series can be represented by a linear combination of its past disturbances. To describe stochastic stationary processes, a general linear model is used, special cases with a finite number of parameters of which are moving average and autoregressive models that are more applicable in practice. A further generalization of models for time series that describe real processes and are not stationary — autoregressive and integrated moving average models — have a wide field of application. Due to the complexity of the algorithm and the need to process a large amount of statistical data, the model is implemented in mathematical and statistical software packages. The article presents the implementation of the ARIMA model in Statistica program and its application in forecasting the average salary of citizens of the Russian Federation.

Keywords: time series, stationarity, autoregressive model, moving average model, ARIMA, forecasting

В стационарных эконометрических моделях значение результирующей переменной зависит от одновременных значений ее факторных признаков, то есть от текущего состояния экономической системы. В динамических системах значения переменной зависят от ее значений в предыдущие моменты времени, то есть ее значений со сдвигом по времени на один шаг назад. Это означает, что поведение системы зависит не только от ее текущего состояния, но и от траектории изменения системы. Эконометрическая модель такого типа представляет собой не функцию объясняющих переменных, а функционал от траектории экономических переменных.

Начало использования динамических эконометрических моделей можно отнести к диссертации Луи Башелье 1900 г., в которой была описана динамика поведения французских государственных облигаций, схожая с броуновским движением.

В 1987 г. Нельсон и Плоссер показали, что коэффициенты регрессии почти всех исторических макроэкономических рядов США являются статистически значимыми, то есть стандартные тесты регрессионного анализа не диагностируют нарушений предпосылок классической модели, но при этом никакой зависимости между экономическими показателями нет. Эти факты заставили пересмотреть все до тех пор полученные эконометрические результаты в области анализа экономических моделей [1; 2].

Целью работы является исследование статистических данных по средней заработной плате на территории Российской Федерации – изучаемый временной ряд не является стационарным. На данном примере покажем необходимые преобразования для использования модели АРПСС и ее применимость для прогнозирования будущих значений рядов.

Материалы и методы исследования

Для описания стационарных случайных процессов используется общая линейная модель, имеющая скорее теоретическое значение, так как ее трудно обрабатывать. На практике применяют частные модели, описываемые небольшим числом параметров.

В 1938 г. Вольд доказал, что стационарный случайный процесс может быть представлен в виде линейной комбинации прошлых возмущений, или белого шума с нулевым математическим ожиданием:

$$X_t = \varepsilon_t + \Psi_1\varepsilon_{t-1} + \Psi_2\varepsilon_{t-2} + \ldots = \sum_{\tau=0}^{\infty}\Psi_\tau\varepsilon_{t-\tau} = \sum_{\tau=-\infty}^{t}\Psi_{\tau-\tau}\varepsilon_t. \quad (1)$$

В операторной форме для обратимого процесса (1) имеем:

$$X_t = \left(\Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots\right) + \varepsilon_t. \quad (2)$$

Из (2) следует, что для обратимой общей линейной модели текущее значение процесса является линейной комбинацией всех его прошлых значений и случайного возмущения, не коррелирующего с этими значениями, т.е. можно построить прогнозное значение $X^*_{t+k}(k \geq 1)$ по его прошлым значениям [3].

Случайный процесс называется процессом скользящего среднего (moving average) порядка q, сокращенно MA(q), если в его разложении Вольда присутствует q слагаемых. Пусть $\Psi_k = 0$ при $k > q$, тогда разложение (1) примет вид:

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q} = (1 - \theta_1 B - \ldots - \theta_q B^q)\varepsilon_t \equiv \theta(B)\varepsilon_t.$$

Если известны все значения $X_\tau(\tau \leq t)$, то прогнозное значение $X^*_{t+k}(k \geq 1)$ с учетом равенств $\Psi_j = -\theta_j, j = 1,\ldots,q, \Psi_j = 0$ при $j > q$ примет вид:

$$X^*_{t+k} = -\left(\theta_k\varepsilon_t + \ldots + \theta_q\varepsilon_{t+k-q}\right).$$

Таким образом, в модели скользящего среднего порядка q можно построить прогноз максимум на q шагов вперед: при этом среднеквадратическая ошибка X_{t+k} равна $\left(1 + \theta_1^2 + \ldots\theta_{k-1}^2\right)\sigma_\varepsilon^2$.

Другой класс моделей с конечным числом параметров можно получить в предположении, что обращенная форма общей линейной модели (1) содержит лишь конечное число членов. Пусть $\Phi_k = 0$ при $k > p$, тогда (1) можно записать в следующем виде:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \ldots + \Phi_p X_{t-p} + \varepsilon_t. \quad (3)$$

Конечное число членов в разложении (3) определяется порядком p модели авторегрессии, или AR(p).

За счет подбора порядков p и q в описанных моделях можно удовлетворительно описывать многие реальные процессы. На практике для получения большей гибкости в подгонке модели к исследуемым рядам бывает целесообразно объединить авторегрессию и скользящее среднее таким образом, чтобы построить наиболее простую модель, дающую хорошую аппроксимацию при малом количестве параметров. Такая модель носит название ARMA (p, q) и описывается уравнением:

$$X_t = \Phi_1 X_{t-1} + \ldots + \Phi_p X_{t-p} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q}, \quad (4)$$

Пусть требуется построить прогноз $X^*_{t+k}, k \geq 1,\ k \leq q$ по имеющимся наблюдениям $X_\tau, \tau \leq t$. Тогда по (4)

$$X^*_{t+k} = \Phi_1 X_{t+k-1} + \ldots + \Phi_p X_{t+k-p} + \varepsilon_{t+k} - \theta_1 \varepsilon_{t+k-1} - \ldots - \theta_q \varepsilon_{t+k-q}.$$

Далее рассмотрим модель

$$\varphi(B) X_t = \theta(B) \varepsilon_t, \quad (5)$$

где $\varphi(B)$ – нестационарный оператор авторегрессии порядка $p + d$, такой, что d корней уравнения $\varphi(B) = 0$ равны единице, а остальные p корней лежат вне единичного круга; $\theta(B)$ – оператор скользящего среднего. Тогда (5) можно записать в следующем виде:

$$\varphi(B) = \Phi(B)(1 - B)^d, \quad (6)$$

где $\Phi(B)$ – стационарный оператор авторегрессии порядка p. Если ввести разностный оператор $\nabla = 1 - B : \nabla X_t = X_t - X_{t-1}$, то $\varphi(B)$ запишется как $\varphi(B) = \Phi(B)\nabla^d$, и (6) можно записать в виде

$$\Phi(B) \nabla^d X_t = \theta(B) \varepsilon_t, \quad (7)$$

где $\nabla^d X_t = \left(\sum_{j=0}^{d} (-1)^j C_d^j B^j \right) X_t = \sum_{j=0}^{d} (-1)^j C_d^j X_{t-j}$, и, следовательно, (7) уже является стационарным обратимым процессом ARMA(p,q). Если ввести обратный к ∇ оператор $S = \nabla^{-1} = (1 - B)^{-1} = \sum_{j=0}^{\infty} B^j$, называемый оператором суммирования, то из (10) следует, что

$$X_t = S^d \Phi^{-1}(B) \theta(B) \varepsilon_t.$$

Таким образом, процесс $\{X_t\}$ можно получить d-кратным интегрированием стационарного процесса (7), т.е. процесс $\{X_t\}$ получается из белого шума ε_t с помощью трех операций фильтрации: скользящего среднего $\theta(B)$, стационарной авторегрессии $\Phi^{-1}(B)$ и суммирования S^d [3].

Для прогнозирования временных рядов с помощью моделей авторегрессии и проинтегрированного скользящего среднего воспользуемся прикладным программным пакетом Statistica 12. Для работы потребуется модуль Times Series/Forecasting, раздел ARIMA & autocorrelation functions [4].

Рассмотрим данные о среднемесячной заработной плате по РФ за период 2013–2019 гг. [5] Для анализа используем данные за 2013–2018 гг. (табл. 1), чтобы сравнить прогноз с фактическими данными за 2019 год для корректировки значений.

Таблица 1

Средние месячные заработные платы по РФ в 2013–2019 гг.

Месяц / Год	01	02	03	04	05	06	07	08	09	10	11	12
2013	26840	26620	28693	30026	29723	30986	30229	29226	29346	30069	30290	39648
2014	29535	29255	31486	32947	32272	33726	32515	30763	31929	32439	32546	42136
2015	30929	31325	32642	34377	34380	35395	33901	32176	32911	33357	33347	43408
2016	32660	33873	35501	36497	37270	38447	35888	35405	35843	35749	36195	47554
2017	34422	35497	37899	39225	39679	41454	38073	37099	38047	38333	38848	51197
2018	39017	40443	42364	43381	44076	45848	42413	41364	41774	42332	42595	55569
2019	42263	43062	46324	48030	47926	49348	46509	44961	45541	46549	46285	62239

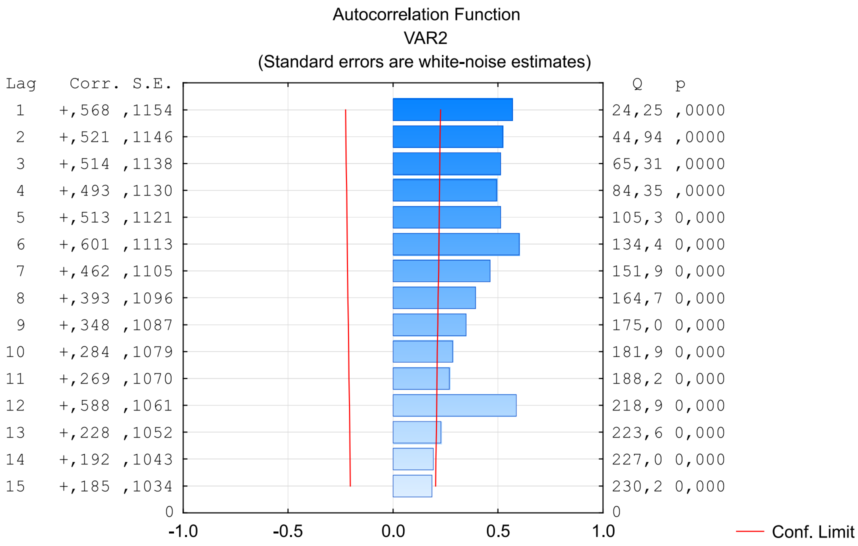

Рис. 1. Функция автокорреляции остатков

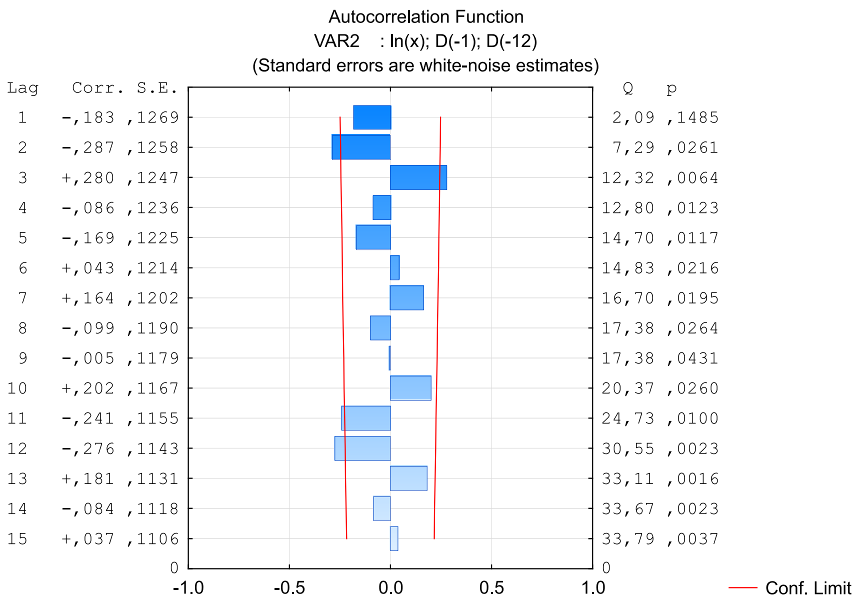

Рис. 2. Функция автокорреляции остатков

Формально мы можем прогнозировать только стационарные ряды. Для того чтобы работать с представленным выше рядом, необходимо провести преобразования, суть которых заключается в том, чтобы избавиться от коррелированности остатков (стандартных ошибок). Для исходного ряда функция автокорреляции выглядит следующим образом (рис. 1).

Отметим, что остатки коррелированы, и необходимо применить разностный оператор. Кроме того, сильно коррелированы остатки с лагом 12, поэтому в модели нужно использовать сезонную компоненту с данным лагом.

Преобразуем исходный ряд: прологарифмируем и применим разностный оператор второго порядка с лагами 1 и 12. Для нового ряда функция автокорреляции примет вид (рис. 2).

Количество статистически значимых коэффициентов определяет порядок модели ARMA(p, q). Можно отметить, что по функции убывания коэффициентов автокорреляции со временем (лагом) можно судить о том, какие знаки будут принимать коэффициенты модели [6].

С учетом автокорреляции остатков для анализа ряда необходимо применять авторегрессию 2-го порядка с сезонными компонентами $Ps = 1$, $Qs = 1$. Результат выполнения алгоритма и оценки параметров представлен ниже (рис. 3).

Программный пакет Statistica автоматически проверяет статистические гипотезы для коэффициентов, поэтому выделенные красным цветом коэффициенты модели являются статистически значимыми (при $\alpha = 0{,}05$). В пакете есть возможность построить прогноз на N шагов вперед с доверительным интервалом, а также визуализировать данные (рис. 4).

Табличные значения прогноза и фактические показатели (табл. 2).

Таблица 2

Прогноз модели и фактические данные

CaseNo.	Forecasts; Model:(2,1,0)(1,0,1) Seasonal lag: 12 Start of origin: 1 End of origin: 72			
	Forecast	Lower	Upper	2019
73	42008,3	40583,1	43483,5	42263,2
74	43406,0	41665,3	45219,4	43062,4
75	45623,8	43686,0	47647,5	46324,2
76	46930,3	44710,1	49260,8	48029,8
77	47599,2	45138,9	50193,7	47926,2
78	49481,7	46767,8	52353,1	49347,9
79	45868,4	43203,8	48697,4	46509,4
80	44719,5	41977,2	47640,9	44961,3
81	45371,1	42457,1	48485,1	45540,9
82	45843,7	42772,6	49135,3	46549
83	46219,3	42998,7	49681,2	46284,5
84	60320,8	55962,9	65018,1	62239,2

```
      Variable: VAR2
Transformations: ln(x),D(1)
          Model: (2,1,0)(1,0,1)  Seasonal lag:  12
No. of obs.:71  Initial SS=,95127  Final SS=,02869 (3,016%)  MS=,00043
 Parameters (p/Ps-Autoregressive, q/Qs-Moving aver.); highlight: p<.05
             p(1)   p(2)  Ps(1)  Qs(1)
  Estimate: -,3626 -,3506 ,99654 ,39406
  Std.Err.: ,11781 ,11532 ,00218 ,13306
```

Рис. 3. Оценка модели ARMA(2,0)

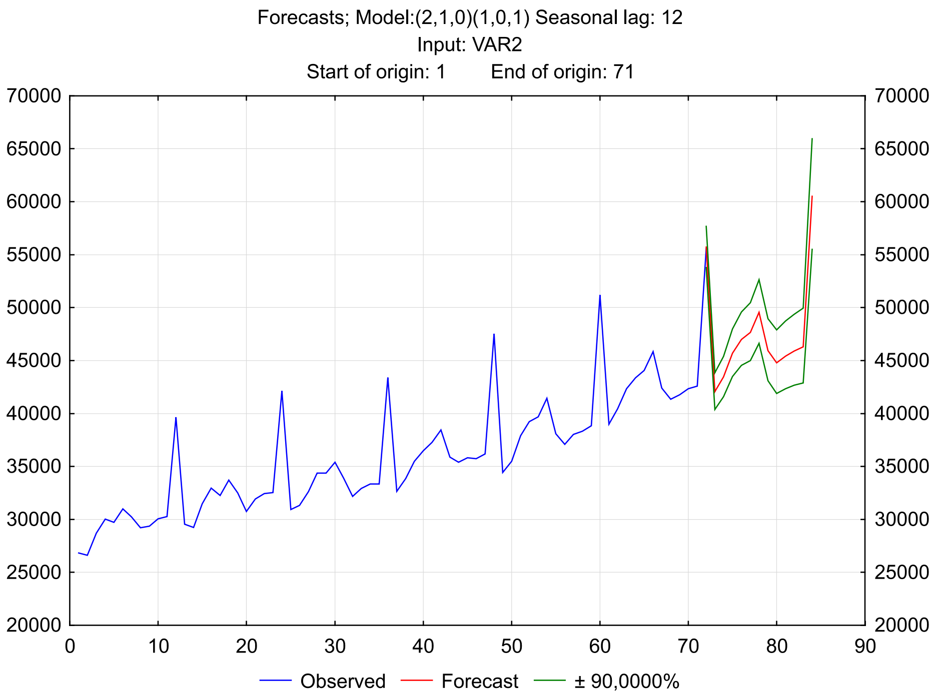

Рис. 4. График прогнозных значений ряда на 12 шагов вперед

Модель ARIMA позволяет строить точечный прогноз для будущих значений временного ряда. В силу того что исследуемый процесс является стохастическим, прогнозное значение также является случайной величиной. Для непрерывных случайных величин необходимо строить интервал, размер которого определяется вероятностью попадания в него реализации случайной величины. С помощью программного пакета Statistica строим доверительный интервал при заданном уровне значимости.

Далее используем данные 2019 г., чтобы сделать модель более точной, и построим прогноз на первое полугодие 2020 г. Сохраним те же настройки модели ARIMA. Полученный результат представлен в табл. 3.

Таблица 3

Прогноз модели на первое полугодие 2020 г.

CaseNo.	Forecasts; Model:(2,1,1)(1,0,1) Seasonal lag: 12 Start of origin: 1 End of origin: 83		
	Forecast	Lower (0,9)	Upper (0,9)
85 (01.2020)	46089.88	44384.74	47860.54
86 (02.2020)	47035.64	45181.37	48966.01
87 (03.2020)	50137.72	47861.43	52522.27
88 (04.2020)	51882.66	49325.57	54572.31
89 (05.2020)	52058.62	49337.83	54929.44
90 (06.2020)	53827.39	50816.59	57016.57

Результаты исследования и их обсуждение

По имеющимся данным за 2019 г. можно сравнить прогноз ARIMA и фактические значения. Как видно из рис. 1 и табл. 2, модель достаточно точно описывает исследуемый нестационарный процесс, отражая тенденцию ряда и циклические колебания (периоды меньших и больших выплат заработной платы, объясняемые летними сезонами отпусков и декабрьскими премиями). Все фактические значения находятся в пределах построенного доверительного интервала.

Полученные результаты и оценки параметров модели, прошедшие проверку на статистическую значимость, позволяют считать модель адекватной и использовать для прогнозирования исследуемого процесса – изменения среднемесячных заработных плат резидентов РФ. Прогнозные значения на первое полугодие 2020 г. приведены в табл. 3.

Дополнительно можно подбирать разные наборы параметров модели ARIMA, дающие сходные удовлетворительные аппроксимации исследуемого процесса, и использовать усредненный прогноз [7]. При этом не всегда удается найти параметры модели, при которых оценки коэффициентов были бы статистически значимыми. В частных случаях построенная модель может не иметь статистически значимых оценок, но давать адекватный прогноз на прошедшие периоды, что, однако, не позволяет считать модель надежной и использовать для прогнозирования будущих периодов.

Также существуют обобщения и их модификации и аналоги модели ARIMA(p, d, q): ARFIMA(p, d, q), ARFIMA-GARCH(p, d, q) – учитывающие фрактальную размерность исследуемого ряда. Каждый из них имеет свою специфику применения и преимущества [8].

Заключение

При исследовании процессов, описываемых временным рядом, имеющим скрытые закономерности, определяющие изменения числовых показателей процесса, необходимо учитывать их фрактальные свойства. В частных случаях для анализа применяется модель Бокса-Дженкинса, направленная на исследование нестационарных временных рядов. При преобразованиях временного ряда и подборе параметров модели можно получить удовлетворительный прогноз для изучаемого процесса на k шагов с некоторым доверительным интервалом.

Использование модели авторегрессии – проинтегрированного скользящего среднего особенно полезно при исследовании экономических процессов, которые имеют определенную тенденцию и цикличность, поскольку построение точного прогноза будущих экономических показателей необходимо при выборе экономической стратегии государства, предприятия, составления планов производства, хеджирования рисков и управления экономической активностью компаний.

Список литературы

1. Яковлев В.П. Эконометрика. М.: Дашков и К°, 2017. 383 с.

2. Елисеева И.И. Эконометрика: учебник для бакалавриата и магистратуры. М.: Юрайт, 2017. 449 с.

3. Афанасьев В.Н. Анализ временных рядов и прогнозирование: учебник для вузов. М.: Финансы и статистика, ИНФРА-М, 2012. 318 с.

4. Яковлев В.Б. Эконометрика в Excel и Statistica. Часть 2. Анализ временных рядов. М.: Эдитус, 2018. 160 с.

5. Федеральная служба государственной статистики. 2020. [Электронный ресурс]. URL: https://www.gks.ru (дата обращения: 19.06.2020).

6. Баллод Б.А. Методы и алгоритмы принятия решений в экономике. СПб.: Лань, 2018. 272 с.

7. Заяц О.А. Прогнозирование объемов производства молока на основе сезонной ARIMA-модели // Фундаментальные исследования. 2019. № 6. С. 61–66.

8. Симонов П.М., Гарафутдинов Р.В. Моделирование и прогнозирование динамики курсов финансовых инструментов с применением эконометрических моделей и фрактального анализа // Вестник Пермского университета. Серия. Экономика. 2019. Т. 14. № 2. С. 268–288.