УДК 338

А.Ю. Трусова, А.И. Ильина*

МОДЕЛИРОВАНИЕ И АНАЛИЗ ДИНАМИЧЕСКИХ ДАННЫХ

В статье рассмотрены некоторые особенности авторегрессионых моделей, использовалась авторегрессионная интегрированная модель скользящего среднего Бокса – Дженкинса. Выполнен ретропрогноз на основе использованных моделей, рассчитаны важнейшие показатели моделей, проведена оценка адекватности и выбрана оптимальная модель авторегрессии.

Ключевые слова: динамические данные, авторегрессионые модели, прогнозирование, ретропрогноз.

Авторегрессионные модели широко используются для описания стационарных случайных процессов в экономических исследованиях.

В настоящее время достаточно хорошо изучены адаптивные модели прогнозирования уровней рядов динамики, такие как: модель скользящего среднего, модели авторегрессии – скользящего среднего, обобщенная линейная модель прогноза временных рядов, сезонные модели, модель Бокса – Дженкинса, нестационарные модели, модель со множеством состояний и др.

Построение авторегрессионных моделей в экономике основано на таком важном свойстве рядов экономических явлений и процессов, как взаимозависимость уровней одного и того же ряда друг от друга. Условие нормальности распределения ряда при построении их эконометрических моделей не является обязательным [1].

Авторегрессионные модели широко используются для описания стационарных случайных процессов. Характерной особенностью стационарных временных рядов является то, что их развитие происходит без выраженной тенденции в неизменных стабильных условиях, поэтому вероятностные свойства рядов не изменяются во времени. Функции распределения стационарных динамических рядов не меняются при сдвиге времени [2].

Рассмотрим модели, лежащие в основе процедуры прогнозирования и методы построения этих моделей, а затем проанализируем особенности их применения и конкретные результаты, полученные при проведении экспериментов.

1. Авторегрессионная модель.

В этой модели текущее значение процесса выражается через конечную линейную совокупность предыдущих значений процесса и возмущения ε_t:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \cdots + a_m y_{t-m} + \varepsilon_t, t = 2,3,\dots,n, \quad (1)$$

где α_j - параметры (коэффициенты) авторегрессионной модели $(j = 1,2,\dots m)$;

ε_t – случайная ошибка с нулевым математическим ожиданием, конечной дисперсией и единичной автокорреляционной матрицей, подтверждающей отсутствие автокорреляции между уровнями ряда ошибок (отклонений).

Трусова Алла Юрьевна, Ильина Алла Ивановна (a_yu_ssu@mail.ru), кафедра математики и бизнес-информатики Самарского государственного университета, 443011, Российская Федерация, г. Самара, ул. Акад. Павлова, 1.

2. Модель скользящего среднего.

Модель описывает конкретные уровни выборочного ряда (случайный процесс) как линейную функцию зависимости моделируемой величины y_t от разности (отклонений) между прошлыми фактическими (y_{t-i}) и прошлыми расчетными (смоделированными) $(\widehat{y_{t-1}})$ наблюдениями $(\varepsilon_{t-i} = y_{t-i} - \widehat{y_{t-i}},\ i = 1,2,\ldots,q; t = 2,3,\ldots,n)$.

Таким образом, данная модель задается уравнением вида

$$\hat{y}_t = \beta_0 + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \cdots + \beta_i\varepsilon_{t-i} + \beta_q\varepsilon_{t-q}, \qquad (2)$$

где β_0, β_i – параметры модели скользящего среднего.

3. Нестационарные модели.

В этих моделях используются идеи о возможности трансформировать нестационарные ряды в стационарные путем перехода от исходного ряда к его разностям соответствующего порядка d. Тогда преобразованный, стационарный ряд можно описать одной из рассмотренных выше моделей [3; 4].

Модель авторегрессии – скользящего среднего

Описывает временной ряд y_t выборочных данных, задается путем объединения АР-модели и СС-модели в следующем виде:

$$\widehat{y_t} = \alpha_0 + \sum_{j=1}^{p}\alpha_j y_{t-j} + \beta_0 + \sum_{i=1}^{q}\beta_i\varepsilon_{t-i}, \quad t = 2,3,\ldots,n, \qquad (3)$$

где α_0, α_j – коэффициенты авторегрессии $(j = 1,2,\ldots,p)$;

p — порядок уравнения авторегрессии;

q – порядок уравнения скользящего среднего;

β_0, β_i – коэффициенты (параметры) уравнения скользящего среднего $i = 1,2,\ldots,q)$;

ε_t – отклонения фактического y_t и расчетного $\widehat{y_t}$ уровней ряда.

В случае когда коэффициенты автокорреляции существенно отличаются от нуля при больших лагах, ряд динамики можно описать моделью авторегрессии – скользящего среднего.

Модель АРСС (p,q) применительно к стационарному случайному процессу в общем виде представляется следующим уравнением с характерным «белым шумом» e_t:

$$y_t = \sum_{j=1}^{p} a_j y_{t-j} + e_t - \sum_{i=1}^{q} b_i e_{t-i}, \quad t = 2,3,\ldots,n, \qquad (4)$$

где a_j, b_i – истинные коэффициенты авторегрессии и скользящего среднего динамического процесса y_t; $-1 < a_j < 1$ и $-1 < b_i < 1$.

Модель Бокса – Дженкинса

Переход от авторегрессионной модели скользящего среднего к модели Бокса – Дженкинса происходит с помощью представления АРСС-модели без свободного члена для центрированных элементов ряда, задающихся в виде уравнений:

$$\hat{y}_t = \bar{y} + \alpha_1(y_{t-1} - \bar{y}) + \alpha_2(y_{t-2} - \bar{y}) + \cdots + \alpha_p(y_{t-p} - \bar{y}) - \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}$$

$$\Delta \hat{y}_t = \Delta \bar{y} + \alpha_1(\Delta y_{t-1} - \Delta \bar{y}) + \alpha_2(\Delta y_{t-2} - \Delta \bar{y}) + \cdots + \alpha_p(\Delta y_{t-p} - \Delta \bar{y}) - \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}, \quad (5)$$

где $\Delta \hat{y}_t, \hat{y}_t$ - расчетные значения исходного и разностного рядов в момент времени t;

$\Delta \bar{y}, \bar{y}$ – средние значения исходного и разностного рядов, относительно которых предполагается наличие статистического равновесия колеблемости уровней ряда;

ε_t – отклонение фактического и расчетного уровней ряда в момент времени t.

Авторегрессионная модель скользящего среднего такого вида нестационарного исходного ряда идентифицируется тремя параметрами:

1) порядком авторегрессионной модели,
2) порядком разностного ряда,
3) порядком уравнения скользящего среднего.

Определенная таким образом модель называется авторегрессионной интегрированной моделью скользящего среднего Бокса – Дженкинса. Данная модель используется для идентификации временного ряда (определение порядков конечной разности, авторегрессии и скользящего среднего), оценивания параметров и проверки адекватности модели [5].

Обощенная линейная модель прогноза временных рядов

Обобщенная линейная модель прогноза временного ряда (ОЛИМП) представляет собой, по сути, модель авторегрессии – скользящего среднего, применяемую для моделирования нестационарных временных рядов.

В модели ОЛИМП не используется параметр d – порядок разностного ряда. Таким образом, эта модель идентифицируется только двумя параметрами p и q и сокращенно записывается как ОЛИМП (p,q). Уравнение модели ОЛИМП имеет следующий вид:

$$\hat{y}_t = \bar{y} + \sum_{j=1}^{p} \alpha_j(y_{t-j} - \bar{y}) - \sum_{i=1}^{q} \beta_i \varepsilon_{t-i}, \qquad t = 2,3,\ldots,n, \quad (6)$$

где $y_t, \hat{y}_t$ – эмпирические (фактические) значения уровня ряда в момент времени t;

$\bar{y}$ – среднее значение исходного ряда;

α_j – коэффициент уравнения авторегрессии j-го порядка;

β_i – коэффициент уравнения скользящего среднего i-го порядка;

ε_t – отклонение фактического (y_t) и расчетного $(\hat{y}_t)$ уровней исходного ряда в момент времени t.

Построенные на практике модели показали, что применение специальных процедур адаптации моделей прогноза временных рядов, основанных на методах Брауна, Хольта, авторегрессии, авторегрессии – скользящего среднего, обобщенной линейной модели, дает достаточно точные результаты краткосрочного и среднесрочного ретропрогноза.

В алгоритмах процедур адаптивных моделей заложены схемы постоянного пошагового сопоставления оценок ретропрогноза, полученных на основе модели, с фактическими уровнями ряда и корректировки параметров модели в соответствии с имеющимися расхождениями.

Модель строилась на основе данных о пассажирских перевозках в разные промежутки времени, всего 144 наблюдения. По исходным данным построена модель проинтегрированного скользящего среднего, выявлены тенденции, проведена оценка параметров модели, проверена стационарность ряда и адекватность модели, построен прогноз на последующие промежутки времени. На рис. 2 представлен график исходных данных по 144 наблюдениям.

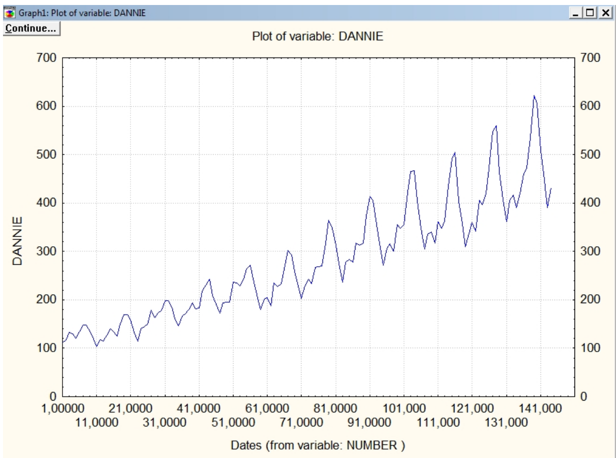

Рис. 1. Графическая иллюстрация исходного ряда

На рис. 1 видно, что ряд удовлетворяет всем требованиям, при этом нет резких скачков, просматриваются тренд ряда, который выражается в плавном увеличении объема перевозок, и некоторая сезонность. Для нахождения периодичностей используется спектральный анализ.

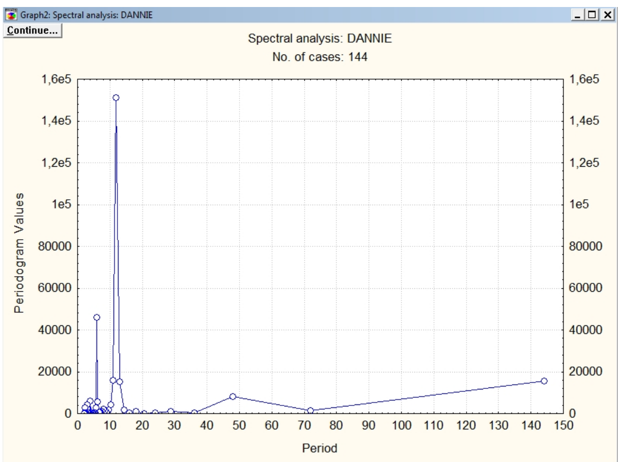

Рис. 2. Спектральный анализ, периодограмма

Узкий высокий пик на рис. 2 свидетельствует о наличии регулярных циклов, а широкие пики соответствуют нерегулярным циклам.

Следующим шагом являлось оценивание параметров АРИСС модели максимизацией функции правдоподобия. С помощью программы «Статистика» получился результат оценивания со стандартными ошибками, асимптотическими значениями t-статистик и т. д. (рис. 4).

В таблице видно, что оценки обоих параметров высоко значимы (т.к. p значительно меньше 0,05).

Далее строятся прогнозы и их доверительные интервалы для наблюдений, начиная с 145-го (рис. 5, 6).

Для анализа адекватности модели исследуются остатки (рис. 7).

На графике видно, что выборочная плотность распределения остатков успешно аппроксимируется нормальным законом распределения, что является признаком адекватности построенной модели.

Descriptive Statistics (данные.sta)

Continue...	Mean	Std.Dv.	Minimum	Maximum	First Case	Last Case	N
DANNIE	280,299	119,97	104,0000	622,0	1,000000	144,0000	144,0000
DANNIE : Periodogram vals;	4126,955	18508,66	,0000	151148,3	1,000000	73,0000	73,0000
DANNIE : Spectral density;	4178,534	11630,41	41,8273	75196,7	1,000000	73,0000	73,0000
DANNIE : ln(x)	5,542	,44	4,6444	6,4	1,000000	144,0000	144,0000

Рис. 3. Параметры ряда

Input: DANNIE (данные.sta)

Continue...

Transformations: ln(x),D(1),D(12)

Model:(0,1,1)(0,1,1) Seasonal lag: 12 MS Residual=,00141

Paramet.	Param.	Asympt. Std.Err.	Asympt. t(129)	p	Lower 95% Conf	Upper 95% Conf
q(1)	,377162	,089318	4,222698	,000045	,200445	,553880
Qs(1)	,572379	,071189	8,040232	,000000	,431529	,713229

Рис. 4. Оценка параметров

Forecasts; Model:(0,1,1)(0,1,1) Seasonal lag: 12 (данные.sta)

Continue...

Input: DANNIE

Start of origin: 1 End of origin: 144

CaseNo.	Forecast	Lower 90,0000%	Upper 90,0000%
145	450,1171	422,9655	479,0117
146	425,6620	395,5777	458,0341
147	479,5240	441,3696	520,9766
148	492,0412	449,0088	539,1979
149	508,5479	460,4357	561,6874
150	583,0166	524,0264	648,6473
151	669,1520	597,3584	749,5742
152	666,4152	591,1003	751,3264
153	557,9980	491,9233	632,9478
154	496,7552	435,3899	566,7696
155	429,6965	374,5207	493,0009
156	477,1535	413,6613	550,3910

Рис. 5. Прогноз на 12 наблюдений

В ходе работы были рассмотрены особенности авторегрессионых моделей, такие как неизменность динамических временных рядов, при сдвиге времени; понятие об информационной ценности наблюдений; колебания уровней ряда в условиях модификации моделей, понятие о взаимозависимости уровней ряда, оценка адекватности моделей и другие.

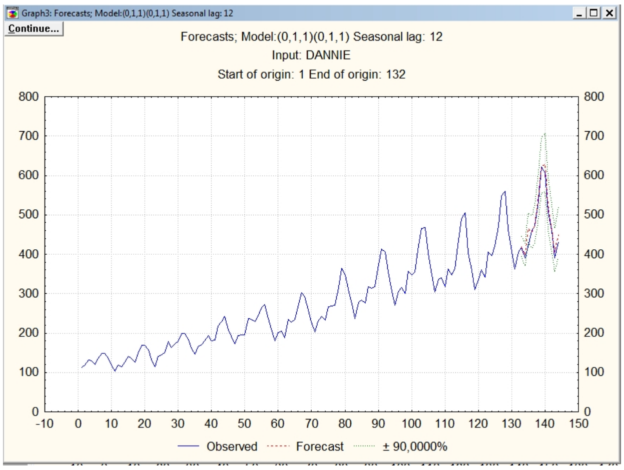

Рис. 6. Проверка прогноза

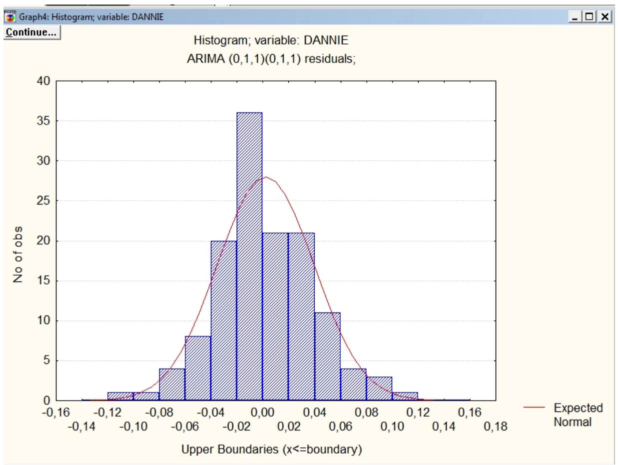

Рис. 7. Гистограмма остатков

В работе построены модель Бокса – Дженкинса. Выявлены основные показатели моделей, проверена адекватность и соответствие модели реальному положению.

Библиографический список

1. Хеннан Э. Многомерные временные ряды. М.: Мир, 1974.
2. Лукашин Ю. Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Мир, 2003.

3. Попов А.С. Резервы ресурсосбережения на машиностроительных предприятиях. М.: Финансы и статистика, 2006.

4. Сараев А.Л., Сараев Л.А. К теории структурной модернизации производственных предприятий // Вестник Самарского государственного университета. 2012. № 10(101). С. 160–169.

5. Эренберг А. Анализ и интерпретация статистических данных. М.: Финансы и статистика, 1999.

A.Yu. Trusova, A.I. Ilyina*

MODELING AND ANALYSIS OF DYNAMIC DATA

In this work some characteristics of autoregressive models are viewed, autoregressive integrated model of moving average of Box-Jenkins was used. Retro forecast was made based on this model, most important indicators were calculated, adequacy of model was tested and optimal autoregressive model was chosen.

Key words: dynamic data, autoregressive model, forecast, retro forecast.

* *Trusova Alla Yurievna, Ilyina Alla Ivanovna* (a_yu_ssu@mail.ru), the Dept. of Mathematics and Business Informatics, Samara State University, Samara, 443011, Russian Federation.