ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Государственное образовательное учреждение

высшего профессионального образования

«ЮЖНЫЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Л.П.Рунова

**Модель авторегрессии и скользящего среднего (ARMA)**

**Методические указания для студентов**

**3 курса бакалавриата направления**

**«Бизнес-информатика» (профиль «Бизнес-аналитика»)**

**Ростов-на-Дону**

**2013**

**УДК 004.45/338**

**ББК 32.973.26-018.2/65.23**

**М 57**

Рунова Л.П.

Модель авторегрессии и скользящего среднего (ARMA). – Ростов-на-Дону: Изд-во ЮФУ, 2013. – 59 с.

ISBN 978-5-699-22086-1

Предмет дисциплины “Методы и модели бизнес-прогнозирования” – самые разные вопросы профессиональной деятельности менеджеров, топ-менеджеров, управленческого персонала крупных, средних и малых предприятий, которые должны предвидеть ход развития бизнес-процесса, циклические колебания экономической конъюнктуры, уметь вырабатывать эффективную стратегию. Без этого принятые решения могут оказываться ошибочными, вложенные инвестиции не окупятся, последствия реализации бизнес-проектов будут противоположны ожидаемым.

Методические указания предназначены для студентов бакалавриата направления ”Бизнес-информатика” (профиль ”Бизнес-аналитика”) при изучении дисциплины ”Методы и модели бизнес-прогнозирования”. Эти методические указания могут быть полезны для студентов и других направлений бакалавриата, а также для магистрантов и аспирантов, решающих задачи социально-экономического прогнозирования.

Рецензенты: д-р экон. наук,

проф. С.В.Крюков,

д-р экон. наук,

проф. Е.И.Лазарева

**ISBN** **978-5-699-22-086-1**

Южный федеральный университет, 2013

**Оглавление**

1 Модель авторегрессии (AR)………………………………………………………...4

2 Модель авторегрессии первого порядка AR(1)…………………………………....8

3 Модель авторегрессии второго порядка AR(2)…………………………………..10

4 Модель скользящего среднего (МА)……………………………………………...16

5 Модель скользящего среднего первого порядка М А(1)…………………..… 18

6 Модель скользящего среднего второго порядка MA(2)…………………………20

7 Модель авторегрессии-скользящего среднего (ARMA)………………………....20

8 Модель авторегрессии и интегрированного скользящего среднего……………24

9 Идентификация модели ARMA..……………………………………………..…...28

Литература…………………………………………………………………..…….….59

**Модель авторегрессии и скользящего среднего (ARMA)**

Во многих случаях в прогнозировании в бизнесе и экономике используются эконометрические модели, построенные на основе временных рядов. Поскольку в данных, собранных на протяжении некоторого промежутка времени, обычно проявляется влияние тренда, сезонных изменений и другие подобные эффекты, наблюдения для разных периодов времени оказываются связанными между собой или, говоря иначе, **автокоррелируют**. Таким образом, для данных временных рядов выборка, составленная из серии имеющихся наблюдений, не может рассматриваться как обычная случайная выборка. Поэтому если стандартные методы регрессии применить к наблюдениям, следующим друг за другом на протяжении некоторого времени, при интерпретации результатов могут возникнуть определенные проблемы. Построение регрессионных моделей для данных временных рядов должно проводиться с особой тщательностью.

**1 Модель авторегрессии (AR)**

Общий вид модели авторегрессии *р*-ого порядка — AR(р) может быть выражен следующим уравнением:

, (1)

где *β1, β2 ,…,βp –* некоторые константы; *εt –* случайные ошибки, образующие «белый шум»:

M(εt) = 0 (2)

. (3)

Она (AR(p)-модель) описывает изучаемый процесс в момент *t* в зависимости от его значений в предыдущие моменты *t-1, t-2,…,t-p*.

Построение модели AR(p) вида (1), адекватной реальному временному ряду *yt*, предполагает решение двух взаимосвязанных задач: определения рационального порядка модели (величины p) и оценки значений ее коэффициентов.

Рассмотрим сначала общие подходы к оценке параметров модели AR(p).

Без ограничения общности будем предполагать, что математическое ожидание ряда *yt*, равно нулю, т.е. *M(yt)=0*. В противном случае вместо переменной *yt*, в выражении (1) можно рассмотреть центрированную переменную , где , но тогда , что доказывает наше предположение.

Из уравнения (1) следует, что параметры модели *β1, β2,…, βp* могут быть выражены через коэффициенты автокорреляции *ρ(τ)*. Для этого умножим уравнение (1) на *yt-τ* почленно и найдём математическое ожидание каждого получившегося слагаемого:

 (4)

Однако, зная, что , в предположении, что , так как *εt* – случайная величина со свойствами «белого шума», не имеющая корреляционной связи с предшествующими моменту *t* значениями рассматриваемого процесса *yt*, разделим левую и правую части выражения на дисперсию процесса **. Тогда всё выражение можно переписать представить следующим образом:

. (5)

Подставив в получившееся уравнение вместо истинных значений коэффициентов автокорреляции *ρτ* их выборочные оценки *rτ*, получим следующую систему линейных уравнений:

 (6)

в которой известными являются оценки коэффициентов автокорреляции *r1*, *r2*,…, *rp* , а неизвестными - оценки коэффициентов *β1, β2,…, βp* модели AR(p): *b1, b2,…, bp.*

Систему линейных уравнений (6) называют уравнениями Юла-Уокера, а полученные на ее основе значения *b1, b2,…, bp* - оценками коэффициентов модели авторегрессии AR(p) Юла-Уокера. Эти оценки могут быть получены с использованием определителей, либо на основе векторно-матричной формы записи системы (6).

На основе определителей оценки Юла-Уокера получают в следующем виде:

, (7)

где Δ – определитель системы (6).

 (8)

Δ τ - определитель, получаемый из определителя Δ путем замены его τ-го столбца на столбец, состоящий из коэффициентов автокорреляции, образующих левую часть системы (1.6) - *r1*, *r2*,…, *rp*.

В векторно-матричной форме записи систему (6) можно переписать в следующем виде:

 (9)

где *r -* вектор-столбец известных оценок коэффициентов автокорреляции с первого по *p-*й включительно, *r=(r1, r2,…, rp)'*; *a* - вектор-столбец неизвестных оценок параметров модели, *а=(а1, а2,..., аp)'*; *R* - матрица, составленная из оценок коэффициентов автокорреляции, определитель которой выражен формулой (8).

Непосредственно из выражения (9) вытекает, что неизвестные оценки коэффициентов модели авторегрессии определяются как

. (10)

Теоретически оценки Юла-Уокера должны обладать свойствами несмещённости и эффективности. Однако, на практике, в моделях авторегрессии большого порядка, эти свойства могут не подтверждаться. Особенно это относится к свойству несмещённости. Как и в моделях с лаговыми зависимыми переменными, смещённость в оценках коэффициентов моделей авторегрессии может быть обусловлена существующей зависимостью между сдвинутыми рядами рассматриваемой переменной *yt-1*, *yt-2* и ошибкой *εt*. Этой возможной зависимостью при построении системы уравнений Юла-Уокера обычно пренебрегают, полагая, что ошибки *εt* образуют «белым шум».

Неэффективность оценок может быть вызвана плохой обусловленностью матрицы *R*, что, как правило, является свидетельством зависимости уже между рядами *yt-1*, *yt-2,…, yt-τ.*

Вместе с тем при небольших порядках модели *(p =1,2,3)* оценки Юла-Уокера обычно являются достаточно «хорошими». В крайнем случае, их можно рассматривать как первое приближение к «оптимальным» оценкам, которые могут быть получены путем уточнения оценок Юла-Уокера на основе использования более мощных методов оценивания, например, нелинейных.

Качество оценок Юла-Уокера может быть проверено путем исследования свойств ряда ошибки *εt*. Если ее свойства близки к характеристикам «белого шума», то оценки Юла-Уокера можно считать «достаточно хорошими». Об этом, в частности, может свидетельствовать критерий Дарбина-Уотсона, значение которого должно лежать примерно в интервале от 1 до 3.

**2 Модель авторегрессии первого порядка AR(1)**

Модель авторегрессии первого порядка AR(1) записывается в следующем виде:

. (11)

Легко видеть, что система Юла-Уокера в этом случае сводится к одному уравнению, непосредственно определяющему оценку *b1* коэффициента *β1:*

*b1 = r1.* (12)

Учитывая, что *εt*и *yt* независимы

. (13)

Так как процесс *yt* *–* стационарный, то . Таким образом, полагая, что , имеем

 (14)

или

, (15)

откуда вытекает, что

. (16)

Из полученного равенства, учитывая, что дисперсия – величина положительная, получаем условие стационарности *-* *|r1|<1*.

Следовательно, при *|r1|>1* ряд оказывается нестационарным.

Найдем автокорреляционную функцию процесса *yt*. Умножая (11) на *yt-1* и вновь учитывая независимость *εt* и *yt*, найдем

 (17)

откуда коэффициент корреляции

 (18)

т. е. коэффициент авторегрессии *r1* представляет собой коэффициент корреляции между соседними возмущениями *yt* и *yt-1*, или ко­эффициент автокорреляции  *ρ1*.

**3 Модель авторегрессии второго порядка AR(2)**

Модель авторегрессии второго порядка AR(2) можно представить в виде следующего уравнения:

 (19)

Система уравнений Юла-Уокера в этом случае состоит из двух уравнений:

 (20)

Выразив *a1* и *а2* через коэффициенты автокорреляции *r1* и *r2*, получим

 (21)

Однако, система уравнений (20) может быть решена относительно *r1* и *r2*

 (22)

Если умножить обе части уравнения (19) на *yt,* взять математическое ожидание от каждого слагаемого, с учётом того, что в слагаемом *M*(*yt , εt*) мы *yt* заменяем на саму модель, и, зная, что *M(yt, yt-τ)=Cov(yt, yt-τ)=rτD(yt-τ)=rτσy2*, получаем:

. (23)

Из этого выражения легко можно получить соотношение между дисперсиями исходного процесса *yt* и ошибкой модели *εt*:

. (24)

Из полученного равенства, учитывая, что дисперсия – величина положительная, получаем необходимые и достаточные условие стационарности:

 (25)

или это можно переписать как

 (26)

**Пример 1**. Использование авторегрессионной модели порядка p **AR (p).**

Компания **X** специализируется на обслуживании портфеля ценных бумаг. Рассмотрим задачу разработки более четкой методики прогнозирования индекса Доу-Джонса (индекса перевозок), с использованием методологии Бокса-Дженкинса. В таблице 1 представлены последние 65 ежедневных средних заключительных значений индекса перевозок для летних месяцев.

Таблица 1 Ежедневные заключительные средние значения индекса перевозок

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Период | Индекс | Период | Индекс | Период | Индекс |
| 1 | 222,34 | 23 | 233,05 | 45 | 253,41 |
| 2 | 222,24 | 24 | 235,00 | 46 | 252,04 |
| 3 | 221,17 | 25 | 236,17 | 47 | 248,78 |
| 4 | 218,88 | 26 | 238,31 | 48 | 247,76 |
| 5 | 220,05 | 27 | 241,14 | 49 | 249,27 |
| 6 | 219,61 | 28 | 241,48 | 50 | 247,95 |
| 7 | 216,40 | 29 | 246,74 | 51 | 251,41 |
| 8 | 217,33 | 30 | 248,73 | 52 | 254,67 |
| 9 | 219,69 | 31 | 248,83 | 53 | 258,62 |
| 10 | 219,32 | 32 | 248,78 | 54 | 259,25 |
| 11 | 218,25 | 33 | 249,61 | 55 | 261,49 |
| 12 | 220,30 | 34 | 249,90 | 56 | 264,95 |
| 13 | 222,54 | 35 | 246,45 | 57 | 268,21 |
| 14 | 223,56 | 36 | 247,57 | 58 | 272,16 |
| 15 | 223,07 | 37 | 247,76 | 59 | 272,79 |
| 16 | 225,36 | 38 | 247,81 | 60 | 275,03 |
| 17 | 227,60 | 39 | 250,68 | 61 | 278,49 |
| 18 | 226,82 | 40 | 251,80 | 62 | 281,75 |
| 19 | 229,69 | 41 | 251,07 | 63 | 285,70 |
| 20 | 229,30 | 42 | 248,05 | 64 | 286,33 |
| 21 | 228,96 | 43 | 249,76 | 65 | 288,57 |
| 22 | 229,99 | 44 | 251,66 |  |  |

Начнем анализ с рассмотрения графика исходных данных, представленного на рис. 1. В ряду явно присутствует возрастающий тренд. Следующим шагом в определении пробной модели будет рассмотрение выборочной функции автокорреляции данных, показанной рис. 2. Следует отметить, что первые несколько коэффициентов автокорреляции постоянно имеют большое значение и стремятся к нулю весьма медленно. Следовательно, первоначальные выводы о наличии тренда были верными, и что исходный временной ряд яв­ляется нестационарным, т.е. его значения нельзя считать изменяющимися относительно не­которого фиксированного уровня.

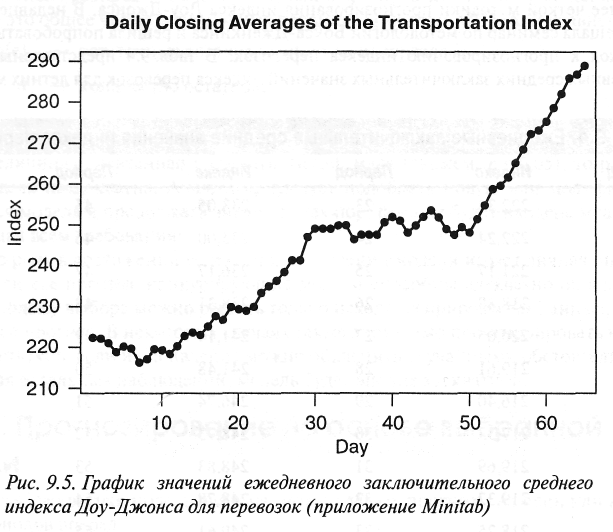


Рисунок 1 - График значений ежедневного заключительного среднего индекса Доу-Джонса для перевозок

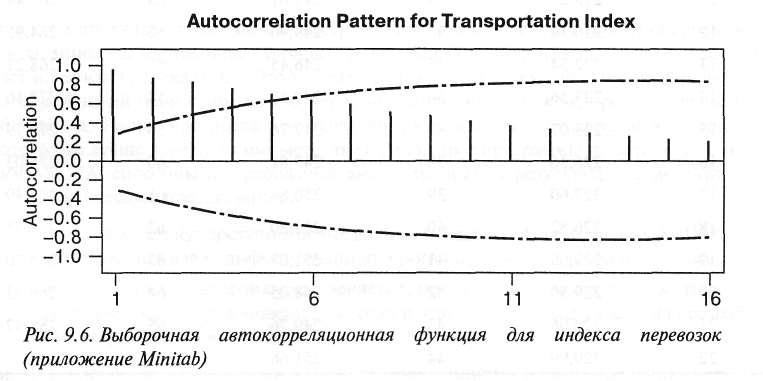


Рисунок 2 - Выборочная автокорреляционная функция для индекса перевозок

Вычислим разности данных, с целью проверить, позволит ли это устранить тренд и получить стационарный ряд. Все из­менения разностных данных происходят в окрестности определенного фиксированного уровня. Оказалось, что выборочным средним для разностей является значение 1,035. Выборочные автокорреляции для разностей показаны на рис. 3, а выборочные частные автокорреляции - на рис. 4.

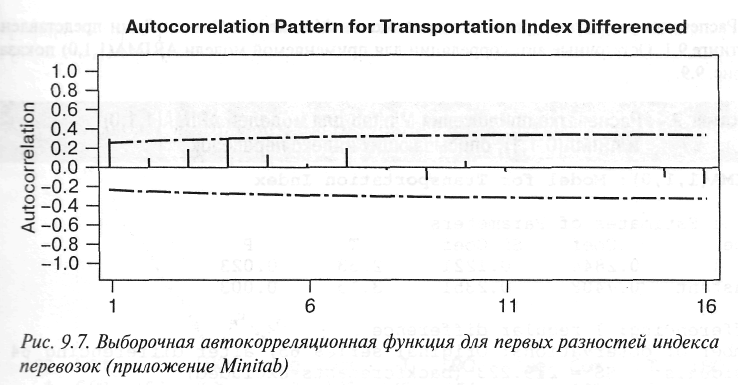


Рисунок 3 - Выборочная автокорреляционная функция для первых разностей индекса перевозок

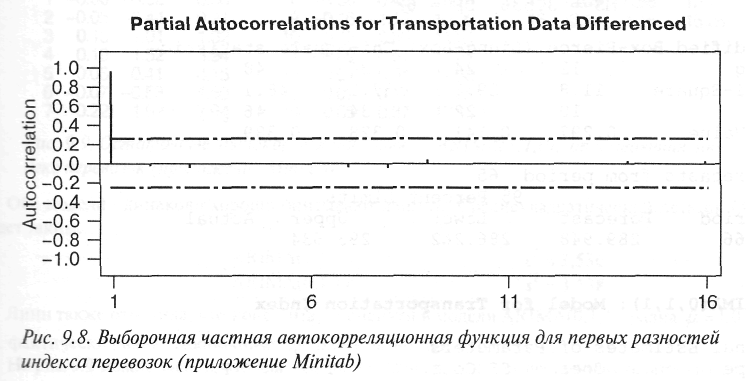


Рисунок 4 - Выборочная частная автокорреляционная функция для первых разностей индекса перевозок

Получаем весьма противоречивые результаты. Сравнение коэффициентов автокорреляции с их предельной ошибкой показало, что существенной была только автокорреляция на первом временном ин­тервале. Аналогично для коэффициентов частной автокорреляции существенным был только интервал 1. Коэффициенты автокорреляции отсекались после первого интервала, указывая на поведение, характерное для модели МА(1). И в то же время коэффициенты частной авто­корреляции также отсекались после этого же интервала, указывая на поведение, характерное ужедля модели AR(1).

Обе выборки не проявляют плавного убывания значений коэффициентов. Применим к индексу перевозок обе модели — ARIMA(1,1,0) и ARIMA(0,1,1). Кроме того, включим в каждую модель постоянное слагаемое, чтобы учесть тот факт, что изменения в ряду разностей проявляются в окрестности уровня, находящегося вы­ше нуля. Если индекс перевозок обозначить, как yt, то разностный ряд будет Δyt = yt - yt-1 и построенная модель, будет иметь следующий вид:

ARIMA(1,1,0): Δyt = φ0 + φ1 Δyt-1 + εt

ARIMA(0,1,1): Δyt = *ϻ +* εt – ω1 εt-1.

Обе модели одинаково хорошо описывают данные. Среднеквадратический остаток *(MS)* будет таким.

АRМА (1,1,0): s2 = 3,536,

ARIMA (0,1,1): s2 = 3,538.

Следует также отметить, что константа, оцененная в модели ARIMA(0,1,1), равна *ϻ=1,038*, т.е. фактически равна выборочному среднему разностей.

На рис. 6 можно видеть, что для модели ARIMA(1,1,0) нет существенных остаточных коэф­фициентов автокорреляции. Хотя остаточная автокорреляционная функция для модели ARIMA(0,1,1) здесь не показана, результат для нее такой же.

Qm - статистика Льюинга-Бокса, рас­считанная для групп интервалов *т =* 12, 24, 36 и 48, не существенна, на что указывает большая ве­личина *р* для обеих моделей. Поэтому можно сделать вывод о том, что обе модели адекватны. Кроме того, про­гнозы на один период вперед, сделанные с помощью этих двух моделей, почти совпадают.

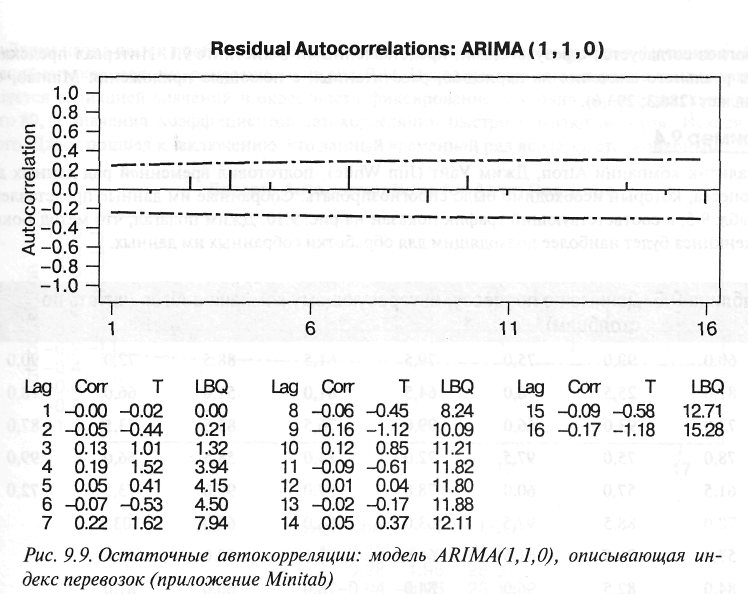


Рисунок 6 - Остаточные автокорреляции; модель ARIMA(1,1,0), описывающая индекс перевозок

Разрешая возникшую дилемму, отдаем предпочтение модели ARIMA(1,1,0), осно­вываясь на ее незначительном преимуществе в точности. Результаты проверки этой модели для периода 66 будут таковы:

yt - yt-1 = φ0 + φ1 (yt-1 - yt-2) + εt

или

yt = yt-1 + φ0 + φ1 (yt-1 - yt-2) + εt

так что при φ0 = 0,741 и φ1 = 0,284 уравнение прогноза примет следующий вид:

ŷ66 =y65 + 0741 + 0,284(y65 - y64) = 288, 57 + 0,741 + 0,284(288,57-286,33)=289,947

Рассчитанный интервал предсказа­ния реального значения на период 66, со­ставляет (286,3; 293,6).

**4 Модель скользящего среднего (МА)**

В моделях скользящего среднего текущее значение стационарного случайного процесса второго порядка *yt*, представляется в виде линейной комбинации текущего и прошедших значений ошибки *εt, εt-1,.., εt-q*, по своим свойствам соответствующей «белому шуму». Такое представление может быть выражено следующим уравнением (модель скользящего среднего порядка *q* – MA(q)):

 (27)

где *γ1, γ2,…, γq -* параметры модели.

В соответствии с определением «белого шума» ошибка *εt*, характеризуется следующими свойствами:

*M(εt)=0*  (28)

. (29)

Вследствие этого и автокорреляционная функция «белого шума» имеет очень простую форму:

 . (30)

С учетом свойств ошибки *εt*, несложно построить автокорреляционную функцию модели MA(q). Ее коэффициент ковариации *q-*ого порядка определяется следующим образом:

 (31)

При *τ=0* выражение (31) представляет собой дисперсию процесса *yt*, которая в силу свойств (28) и (29) выражается через коэффициенты модели MA(q): *γ1, γ2,…, γq*; и дисперсию ошибки  следующим образом:

. (32)

Для произвольного *τ* из (32) получим, что коэффициент ковариации определяется выражением

 (33)

Автокорреляционная функция модели MA(q) получается непосредственно из (7):

 (34)

Система из *q* уравнений (8), может служить основой для получения оценок *g1, g2,…, gq* неизвестных параметров модели MA(q) - *γ1, γ2,…, γq*. Для этого необходимо подставить в каждое ее уравнение вместо значений коэффициентов автокорреляции *ρτ* рассматриваемого процесса *yt* их рассчитанные оценки *rτ*.

Однако в отличие от уравнений Юла-Уокера эта система нелинейная и её решение требует использования специальных итеративных процедур расчетов за исключением наиболее простой модели MA(1).

**5 Модель скользящего среднего первого порядка (МА)**

Она представляется следующим выражением:

*yt= εt - γ1 εt-1.* (35)

Из (34) следует, что дисперсии процесса  и ошибки этой модели  связаны следующим соотношением:

. (36)

Её единственный отличный от нуля первый коэффициент автокорреляции выражается через коэффициент модели как

. (37)

Из соотношения (37) несложно получить квадратическое уравнение относительно оценки *g1* неизвестного параметра *γ1*

, (38)

где *r1 -* оценка коэффициента автокорреляции первого порядка, т.е. *ρ1*.

В свою очередь, из (38) следует, что существуют два решения этого уравнения, связанные между собой следующим соотношением:

. (39)

Условию стационарности процесса удовлетворяет только решение *g1*, по абсолютной величине меньшее единицы:

 (40)

при условии, что

. (41)

Из (41) следует, что модели скользящего среднего первого порядка могут применяться только для описания процессов с автокорреляционной функцией, обрывающейся после первой задержки и коэффициентом автокорреляции, по абсолютной величине не превышающем 0,5.

В заключение приведем основные результаты для модели скользящего среднего второго порядка - MA(2).

**6 Модель скользящего среднего второго порядка MA(2)**

Модель скользящего среднего второго порядка MA(2) в общем виде записывается следующим образом:

*yt= εt - γ1 εt-1- γ2 εt-2.* (42)

Из (39) непосредственно вытекает, что дисперсии процесса и ошибки  связаны следующим соотношением:

. (43)

Её автокорреляционная функция определяют значения коэффициентов автокорреляции, связанные с параметрами модели следующими соотношениями

 (44)

Из этих соотношений могут быть найдены оценки коэффициентов модели *g1* и *g2* при известных оценках коэффициентов автокорреляции *r1* и *r2*.

**7 Модель авторегрессии-скользящего среднего (ARMA)**

Общий вид модели авторегрессии-скользящего среднего – ARMA(p,q) определяется следующим уравнением:

, (45)

где *β1 , β2 ,…, βp ,γ1, γ2,…, γq* - коэффициенты модели; *р -*  порядок авторегрессии; *q -* порядок скользящего среднего.

Заметим, что модель (45) может быть преобразована либо в модель авторегрессии AR(p)

, (46)

где ошибка *ξt*, удовлетворяет свойствам процесса скользящего среднего порядка *q*, либо в модель скользящего среднего – MA(q):  путем выражения переменных *yt-τ*, через линейные комбинации ошибок

 (47)

и дальнейшего приведения подобных членов после раскрытия скобок.

Для этих модификаций модели (45) рассмотрим свойства ее автокорреляционной функции и возможные подходы к оценке ее параметров. Заметим, что при сдвигах, превышающих по своей величине порядок скользящего среднего *q*, т.е. при *τ>q*, коэффициенты автоковариации модели ARMA(p,q), определяемой выражением (45), не зависят от ошибок модели. В самом деле,

 (48)

Если *τ>q*, то в силу свойств «белого шума» все математические ожидания произведений ошибок *εt* и *εt-τ-j*, *j< q* оказываются равными нулю, т.е.

*M(εt-j , εt-τ-j)=0, τ=q+1,q+2,…; j=1,…, q.*

В этом случае, т.е. при *τ>q*, значения коэффициентов автоковариации модели ARMA(p,q) удовлетворяют свойствам этих коэффициентов, характерным для модели авторегрессии *р-*ого порядка AR(p):

 (49)

Из выражения (49) непосредственно вытекает, что неизвестные значения коэффициентов *β1 , β2,…, βp* в этом случае могут быть оценены из модификации системы уравнений Юла-Уокера, имеющей в данном случае следующий вид:

 (50)

С использованием найденных из системы (50) значений оценок коэффициентов *β1, β2,…, βp* на основании выражения (46) сформируем процесс скользящего среднего *q-*ого порядка – MA(q):

, (51)

где *ut* - фактическая ошибка, являющаяся оценкой ошибки *ξt*. Значения ошибки *ut*, получают путем подстановки в выражение (46) вместо неизвестных параметров *β1 , β2 ,…, βp* их оценок *b1, b2,…, bp* определенных из системы (50). *et -* фактическая ошибка, значение которой используется вместо истинной ошибки *εt*, при оценке коэффициентов скользящего среднего. Для определения оценок *g1,g2,…,gq* коэффициентов скользящего среднего применяются нелинейные методы оценивания, предполагающие решение системы нелинейных уравнений типа (48).

Рассмотрим наиболее «популярную» модификацию моделей авторегрессии-скользящего среднего ARMA(1,1). Эта широко используемая в практике эконометрических исследований модель может быть выражена следующей формулой:

*yt=β1yt-1+ εt-γ1 εt-1 .* (52)

Для определения дисперсии этой модели умножим под знаком математического ожидания левую и правую части выражения (52) на *yt*. В результате получим

 (53)

При выводе выражения (53) учтено, что *M(yt, εt)=M(β1yt-1+ εt -γ1 εt-1)=σ02* в силу свойств процесса «белого шума» *εt*.

Далее, умножив под знаком математического ожидания левую и правую части выражения (52) на *εt-1* , получим:

, (54)

поскольку .

Аналогично, получим первый коэффициент автоковариации процесса *yt* , умножив под знаком математического ожидания левую и правую части уравнения (52) на *yt-1*. С учетом того, что *yt-1=β1yt-2+ εt-1 - γ1 εt-2*  и в силу свойств «белого шума» *εt* получим, что

. (55)

Из выражений (53)-(55) непосредственно вытекает, что дисперсия  процесса *yt*, описываемого моделью ARMA(1,1), его первый коэффициент автоковариаций и дисперсия ошибки *εt*, оказываются связанными следующими соотношениями:

 (56)

а коэффициенты автоковариаций более высоких порядков (как следует из выражений (45) и (46)) - соотношениями вида:

. (57)

Из соотношения (54) несложно получить выражение, определяющее значение первого коэффициента автокорреляции процесса ARMA(1,1):

. (58)

Значения коэффициентов автокорреляции более высоких порядков связаны соотношением аналогичным (13) *ρτ= β1 ρτ-1, τ≥2.*

Таким образом, значения коэффициентов автокорреляции модели ARMA(1,1) подчиняется экспоненциальному закону

, (59)

где .

**8 Модель авторегрессии и интегрированного скользящего среднего**

В реальности исследуемые процессы свойством стационарности могут и не обладать, тогда с помощью достаточно несложных преобразований можно привести наблюдаемый ряд к стационарному процессу.

Одним из таких способов преобразования является взятие конечных разностей

, (60)

где  - первая разность. Это преобразование целесообразно использовать, когда закон изменения у, близок к линейному.

, (61)

где - вторая разность. Преобразование применяется, когда закон изменения *yt*, близок к квадратической зависимости.

К приведённому ряду можно применить модель авторегрессии и скользящего среднего, но на этом процесс построения модели нельзя считать завершенным. Для его окончания необходимо продолжить процесс построения модели изначального процесса, выполнив обратные преобразования, перейдя от преобразованных значений *xt*, к исходным значениям *yt*.

Пусть процесс *xt*, соответствует модели авторегрессии-скользящего среднего. Запишем преобразование с помощью оператора сдвига *В* для ряда *xt*.

Получим соответственно

, (62)

, (63)

, (64)

. (65)

где многочлены степеней *p* и *q* соответственно от оператора сдвига, используемые для получения эквивалентной записи модели ARMA(p,q).

Подставляя (62) в (63), получим уравнение для модели динамики исходного временного ряда *yt, t=1,2,…,T* в следующем виде:

. (66)

Отметим, что преобразование (62) не затрагивает ошибку *εt*. Рассмотрим описанную процедуру на примере модели ARMA(1,1).

Пусть

, (67)

что эквивалентно записи

. (68)

Объединяя эти два уравнения в одно, получим модель относительно исходного временного ряда *yt* в следующем виде:

 (69)

Заметим, что преобразование (61) с помощью оператора В записывается в следующем виде:

. (70)

В этом случае для произвольной модели ARMA(p, q) получим

. (71)

В частности, для модели ARMA(1,1), построенной для ряда *zt*, выражение (71) для исходного процесса *yt* приобретает следующий вид:

. (72)

В случае приведения исходного ряда *yt, t=1,2,…,T* к стационарному с использованием d-ой разности его результирующая модель определяется выражением:

. 73)

В практических исследованиях при проведении обратных преобразователей вместо параметров *β* и *γ*, в соответствующие выражения для моделей исходного временного ряда *yt*, необходимо подставить значения их оценок *a* и *b*, полученные для моделей преобразованного стационарного процесса *yt*.

Таким образом, из выражений (67) и (70) вытекает, что использование для преобразования исходного временного ряда *yt* в стационарный процесс *xt, t=1,2,…,T*, оператора разности не ведет к изменению вида модели, описывающей процесс *yt*. Она, как и модель ARMA(p,q), описывающая стационарный процесс *xt,* является линейной по форме.

Обратим также внимание на необходимость анализа свойств и оценки основных характеристик ошибки исходной, т.е. восстановленной модели. Это должно быть сделано, в том числе и для обоснования оценки качества самой модели. Для некоторых преобразований их значения дисперсии фактической ошибки можно определить, исходя из соответствующих значений дисперсии среднеквадратической ошибки преобразованной модели, используя свойства дисперсий линейных, логарифмических и других зависимостей, соответствующих сделанному преобразованию. В этой связи заметим, что ряд значений фактической ошибки модели  определяется в этом случае после формирования основного уравнения модели и расчета на его основе значений. Далее свойства фактической ошибки могут быть определены с использованием специальных тестов.

**9 Идентификация модели ARMA**

Из рассмотренного материала вытекает, что произвольный реальный стационарный процесс второго порядка может быть выражен разными вариантами моделей временных рядов. Чтобы показать это, запишем, например, модель AR(1) в более компактном виде с использованием оператора сдвига назад В. Его воздействие на любую переменную, зависящую от времени, определяется следующими соотношениями:

. (74)

С учетом (1) модель AR(1) можно представить в следующей форме записи:

. (75)

Поскольку *|β1|<1*, то  является суммой бесконечно убывающей геометрической прогрессии

 (76)

С учетом (2) модель (3) запишем в следующем виде:

, (77)

где в данном случае .

Из выражения (74) следует, что модель авторегрессии первого порядка оказывается эквивалентной модели скользящего среднего бесконечного порядка. Аналогичным образом можно показать и обратное соотношение между порядками этих моделей. Так, для модели MA(1) имеем

. (78)

Поскольку *|γ1|<1*(из условия стационарности процесса *yt*), то из выражения (75) получим

 (79)

В данном случае  - коэффициенты модели авторегрессии бесконечного порядка.

В общем случае модель авторегрессии p-ого порядка оказывается эквивалентной модели скользящего среднего q-огo порядка

, (80)

где многочлен q-ой степени  - результат деления единицы на многочлен q-ой степени.

Из рассмотренных соотношений вытекает важный вывод: на практике можно подобрать модель с минимальным числом параметров, которая описывает временной ряд *yt*, являющийся стационарным процессом, не «хуже», чем другие варианты моделей с большим числом параметров. Обычно понятие «не хуже» связывается с минимальной дисперсией модели и отсутствием автокорреляции в ряду ее ошибки.

Практическая ценность этого вывода состоит в следующем. При построении моделей временных рядов нужно стремиться к минимизации числа их параметров, а, следовательно, и порядка самой модели. Дело в том, что параметры таких моделей оцениваются на основе коэффициентов автокорреляции исходного процесса *yt*. С увеличением порядка модели для определения значений ее параметров необходимо использовать в качестве исходных данных и большее число выборочных коэффициентов автокорреляции (с большими номерами). Точность их оценки с ростом сдвига падает, да и их абсолютные значения либо стремятся к нулю, либо попадают в область повышенной неопределенности. Из-за этого снижается надежность оценок коэффициентов моделей временных рядов высоких порядков, как и качество самих моделей. Все это и заставляет искать для описания реальных процессов модели временных рядов с минимальным числом параметров.

Процесс выбора модели, в наилучшей степени соответствующей рассматриваемому реальному процессу, называется идентификацией модели. В нашем случае идентификация состоит в определении общего вида модели из класса моделей ARMA(p,q), характеризующейся наименьшим числом параметров по сравнению с другими возможными вариантами, без потерь в точности описания исходного процесса.

Вообще говоря, идентификация - это достаточно грубая процедура (последовательность процедур), целью которой является определение некоторой области приемлемых значений характеристик порядка p и q модели ARMA(p,q), которая в ходе дальнейших исследований должна быть сведена к конкретным их величинам.

Обычно в этой части идентификация сопровождается процедурами оценки параметров альтернативных вариантов моделей и выбора наилучшего из них на основе использования критериев качества.

Таким образом, в общем случае формирование модели, в наилучшей степени подходящей для описания реального процесса, как бы состоит трех пересекающихся и дополняющих друг друга этапов - идентификации, оценивания и диагностики (согласования модели с исходными данными с целью выявления ее недостатков и последующего улучшения)

Общая идея идентификации модели ARMA(p,q) состоит в том, что свойства реального процесса и свойства наилучшей модели должны быть близки друг к другу.

Эта близость, как это было показано ранее, практически целиком определяется на основе сопоставления поведения их автокорреляционных функций: теоретической - для модели и эмпирической - для реального процесса, выборочные коэффициенты автокорреляции которого оценены на основе наблюдаемых данных. Поскольку выборочные коэффициенты автокорреляции могут характеризоваться достаточно большими ошибками и, кроме того, сильными корреляционными взаимосвязями между собой, то на практике точного сходства между «теоретической» и «эмпирической» автокорреляционными функциями ожидать не следует, особенно при больших сдвигах. Например, вследствие статистической взаимосвязи между коэффициентами автокорреляции процесса относительно значимые уровни выборочных коэффициентов автокорреляции (всплески) могут иметь место и в областях сдвигов, где их теоретические аналоги близки к нулю. Поэтому при сопоставлении теоретических и выборочных автокорреляционных функций обычно учитывают лишь их главные свойства. Именно их совпадение позволяет значительно сузить круг приемлемых для описания реального процесса вариантов модели. Окончательный выбор в пользу одной из них обычно делается по результатам этапов оценивания и диагностики моделей.

Отметим наиболее характерные свойства автокорреляционных функций типовых моделей ARMA(p,q).

Автокорреляционная функция модели авторегрессии первого порядка - AR(1) спадает строго по экспоненте (точнее, этот вывод справедлив для абсолютных значений коэффициентов автокорреляции). Плавный характер уменьшения коэффициентов автокорреляции характерен и для моделей авторегрессии более высоких порядков. В одном случае спад происходит либо чуть быстрее, чем строго по экспоненте, либо чуть медленнее, а в другом - по закономерности, соответствующей затухающей синусоиде.

Чрезвычайно важная информация о порядке модели авторегрессии содержится в так называемой частной автокорреляционной функции.

Для процесса, описываемого моделью AP(p), ее значениями являются последние значения коэффициентов моделей авторегрессии порядков, не превосходящих p, т.е. моделей с порядками *τ=1,2,…, p*. Обозначим значения частной автокорреляционной функции модели AR(p) через *πp1, πp2,…, πpp.* Тогда для модели AP(p) значение *πp1* равно *ρ1* и на практике определяется как оценка коэффициента *β1* модели AR(1) по формуле:

 (81)

где значение  (см. выражение (21)) - как коэффициент модели AR(2). На практике значение *πp2*, таким образом, определяется по формуле:

. (82)

и оценка любого коэффициента  определяется как оценка коэффициента *βτ*, модели AR(τ) по формуле:

 (83)

Можно показать, что для модели AP(p) значения частной автокорреляционной функции являются значимыми (отличными от нуля) до задержки к включительно, т.е. *πp1>0, i≤p* и равными нулю при сдвигах, превышающих порядок модели, т.е. *πp1=0, i>p*. На практике этот результат следует понимать в «статистическом смысле», поскольку оценки значений коэффициентов частной автокорреляционной функции определяются на основании выборочных значений коэффициентов автокорреляции и поэтому сами являются случайными величинами, характеризующимися определенной ошибкой. Для оценок коэффициентов частной автокорреляционной функции, порядок которых превышает порядок модели, дисперсия ошибки приблизительно может быть оценена по следующей формуле:

, (84)

где *i>p; T* - объем динамического ряда показателя *yt.*

Таким образом, поведение частной автокорреляционной функции моделей авторегрессии аналогично поведению автокорреляционных функций моделей скользящего среднего. Для модели AR(p) ее частная автокорреляционная функция «обрывается» после задержки p, как это имело бы место у автокорреляционной функции модели MA(q). Это свойство частной автокорреляционной функции удобно использовать при идентификации моделей авторегрессии. Если значения такой функции, рассчитанной для реального процесса, обрываются (становятся нулевыми), начиная со сдвига p+1, это указывает на то, что модель авторегрессии p-ого порядка соответствует свойствам рассматриваемого процесса.

Как вытекает из выражения (34) теоретическая автокорреляционная функция модели MA(q) обрывается после задержки q. Поэтому, если автокорреляционная функция реального процесса обладает аналогичными свойствами, это указывает на то, что для его описания целесообразно использовать модель скользящего среднего соответствующего порядка. Иными словами, если у процесса *yt*, оказался значимым только первый коэффициент автокорреляции *r1* и при этом, в соответствии с выражением (41) *r1*<0,5, то данный факт указывает на целесообразность выбора для его описания модели MA(1). Если «обрыв» имеет место после второго сдвига - то модель MA(2) и т.д.

Точно так же как и для моделей авторегрессии частные автокорреляционные функции могут быть построены и для моделей скользящего среднего любых порядков. Для оценки их коэффициентов используются выражения (81)-(83). При этом с учетом того, что для модели MA(1) первый коэффициент автокорреляции *ρ1* и параметр модели *γ1* связаны соотношением , то для τ=2, 3,... с учетом того, что *ρ2= ρ3=…=0,* можно показать, что значения частных коэффициентов автокорреляции этой модели определяются по следующей формуле:

. (85)

Из (11) непосредственно вытекает, что

. (86)

откуда следует, что частная автокорреляционная функция модели MA(1) (т.е. абсолютные значения ее частных коэффициентов автокорреляции) затухает по закону, близкому к экспоненциальному. Иными словами, её поведение похоже на автокорреляционную функцию модели AR(1).

Можно показать, что аналогичное соответствие свойств характер для частной автокорреляционной функции модели MA(2) и автокорреляционной функции модели AR(2). Они представляют собой либо плавно спадающие с ростом сдвига зависимости экспоненциального типа, либо затухающие синусоиды. Такое соответствие автокорреляционных и частных автокорреляционных функций характерно и для моделей авторегрессии, и скользящего среднего более высоких порядков.

Для моделей ARMA(p,q) поведение автокорреляционной функции после задержки q похоже на поведение автокорреляционной функции модели AR(p). Однако на практике обычно используется модель ARMA(1,1), т.е. только первого порядка. Как было показано выше (см. выражения (75)—(80)), это связано с тем, что составляющая модели, относящаяся к авторегрессии первого порядка поглощает все процессы скользящего среднего более высоких порядков, и, наоборот, составляющая скользящего среднего первого порядка поглощает процессы авторегрессии высоких порядков. Вследствие этого и поведение автокорреляционной и частной автокорреляционной функций модели ARMA(1,1) характеризуется как бы комбинацией свойств этих функций, имевших место для моделей AR(1) и MA(1).

Иными словами, составляющая AR(1) способствует тому, что автокорреляционная функция модели ARMA(1,1) (абсолютные значения коэффициентов автокорреляции) затухает экспоненциально, но после первой задержки (первого сдвига). Это непосредственно вытекает из выражений (36) и (38). В свою очередь, составляющая MA(1) определяет закономерности поведения частной автокорреляционной функции модели ARMA(1,1), которая также затухает примерно экспоненциально в соответствии с выражением (85) и (86).

Рассмотренные подходы к идентификации основаны на сопоставлении свойств выборочных автокорреляционных и частных автокорреляционных функций реального стационарного процесса и предполагаемой для его описания модели. На практике идеальное совпадение свойств этих функций встречается не часто, поскольку и реальные процессы обычно не слишком точно соответствуют своим теоретическим аналогам-моделям, и оценки их коэффициентов автокорреляции характеризуются наличием ошибок. Вследствие этого процедура идентификации служит для обоснования выбора некоторой пробной модели из общей группы моделей типа ARMA(p,q), которая является. Как бы начальной точкой на пути построения «оптимального» теоретического аналога (модели) рассматриваемого процесса на основе пользования более точных процедур диагностики и методов оценки параметров модели.

Обычно с помощью процедур диагностики исследуют свойства фактической ошибки модели *et,* которую часто называют остаточной ошибкой. При этом целесообразно руководствоваться следующей логикой анализа временного ряда *et*, значения которого определяется как разность между фактическими и расчетными значениями процесса в момент t, т.е. , где - значения процесса, рассчитываемые по соответствующей модели.

Для «удачной» модели можно ожидать, что ряд ошибки *et, t=1,2,…,T* по своим свойствам будет достаточно близок к «белому шуму» — случайному процессу, характеризующемуся полным отсутствием каких-либо закономерностей в своих значениях, за исключением известного закона их распределения, обычно предполагаемого нормальным. Для нашего случая это означает, что математическое ожидание фактической ошибки должно быть равно нулю (*M(et)=0*), а дисперсия постоянна на любом участке ее измерения () и между рядами *et, et-1, et-2,..*. отсутствует автокорреляционная зависимость, т.е. первый и последующие выборочные коэффициенты автокорреляции ряда *et, t=1,2,…,T* близки к нулю.

Иными словами, фактическая ошибка модели *et,* должна быть «настолько случайна», что ее невозможно было бы уточнить никакой другой моделью.

Кроме того, как это было показано выше, желательно, чтобы дисперсия ошибки  была существенно меньше дисперсии процесса . В этом случае модель, описывающая процесс *yt* как бы снимает значительную часть неопределенности в его изменчивости, что позволяет с большей обоснованностью предсказывать его значения.

Наличие каких-либо закономерностей в ряду ошибки *et,* указывает, что построенная модель неадекватна рассматриваемому процессу *yt.* Причинами неадекватности могут быть ошибки в оценках параметров либо так называемая, неопределенность модели. Примерами такой неопределенности является использование модели AR(1) вместо адекватной процессу модели ARMA(1,1). В этом случае ошибка модели AR(1)  характеризуется свойствами модели MA(1). На это укажет отличный от нуля её первый коэффициент автокорреляции.

Отметим, что и неверно определенные значения параметров приводят к появлению «неслучайности» в ряду фактической ошибки.

Вследствие этого на практике однозначно указать какой-либо путь уточнения модели на основе анализа свойств ошибки *et,* отличных от свойств «белого шума», обычно не представляется возможным. В такой ситуации можно сначала рекомендовать уточнить значения параметров модели путем использования более эффективных процедур их оценки, а затем, если это окажется необходимым - доопределить модель.

Для этой цели могут быть использованы и другие более точные методы оценивания (например, нелинейные), в которых найденные оценки используются как начальные приближения к «оптимальным» значениям параметров модели ARMA(p,q).

Из приведенных выше рассуждений вытекает, что диагностика модели сводится к исследованию свойств ее ошибки с целью выявления степени соответствия её свойств свойствам «белого шума». Такие исследования в случае модели стационарных процессов обычно сводятся к проверке значимости коэффициентов автокорреляции фактической ошибки *et.*

Для проверки гипотезы о соответствии свойств ошибки модели свойствам белого шума могут использоваться процедуры проверки гипотез о постоянстве и равенстве нулю ее математического ожидания, постоянстве дисперсии, равенстве нулю ее коэффициентов автокорреляции.

Подход Бокса-Дженкинса к анализу временных рядов является весьма мощным инстру­ментом для построения точных прогнозов с малой дальностью прогнозирования. Модели ARIMA достаточно гибкие и могут описывать широкий спектр характеристик временных рядов, встречающихся на практике. Формальная процедура проверки модели на адекват­ность проста и доступна. Кроме того, прогнозы и интервалы предсказания следуют непо­средственно из подобранной модели.

Однако использование моделей ARIMA имеет и несколько недостатков.

1. Необходимо относительно большое количество исходных данных. Следует пони­мать, что если данные периодичны со, скажем, сезонным периодом 5=12, то на­блюдения за один полный год будут составлять фактически одно сезонное значение данных (один взгляд на сезонную структуру), а не двенадцать значений. Во­обще говоря, при использовании модели ARIMA для несезонных данных необхо­димо около 40 или более наблюдений. При построении модели ARIMA для сезонных данных нужны наблюдения приблизительно за 6-10 лет, в зависимости от величины периода сезонности.
2. Не существует простого способа корректировки параметров моделей ARIMA, та­кого как в некоторых сглаживающих методах, когда задействуются новые данные. Модель приходится периодически полностью перестраивать, а иногда требуется выбрать совершенно новую модель.
3. Построение удовлетворительной модели ARIMA зачастую требует больших затрат времени и ресурсов. Для моделей ARIMA расходы на построение модели, время выполнения вычислений и объемы необходимых баз данных могут оказаться существенно выше, чем для более традиционных методов прогнозирования, таких как сглаживание.

Согласно Бернштейну (Bernstein, 1996), прогнозирование является одной из важнейших составляющих менеджмента, которая оказывает значительную помощь в процессе при­нятия решений. Фактически любое важное управленческое решение в определенной сте­пени зависит от прогнозов. Накопление запасов связано *с* прогнозами ожидаемого спроса; производственный отдел должен планировать потребности в рабочей силе и сырье на следующий месяц или два; финансовый отдел должен производить краткосрочное фи­нансирование на следующий квартал; отдел кадров должен предвидеть необходимость приема или увольнения служащих. Список разнообразных применений прогнозирования может быть очень длинным.

Управленцы прекрасно осведомлены о необходимости прогнозирования. Несомнен­но, много времени уделяется изучению существующих тенденций в экономике и полити­ке, а также тому, как грядущие события могут повлиять на востребованность предлагае­мой продукции и/или обслуживания. Старшие должностные лица заинтересованы в ко­личественном прогнозе для сравнения его со своим собственным мнением. Интерес к прогнозированию особо обостряется в тех случаях, когда происходят события, способ­ные оказать серьезное влияние на уровень спроса. Недостатком методов количествен­ного прогноза является их зависимость от данных прошлых наблюдений. По этой при­чине они, естественно, менее эффективны в предсказании неожиданных перемен, при­водящих к резкому повышению или падению спроса.

Зачастую менеджерам необходимо сделать краткосрочный прогноз для большого чис­ла наименований продукции. Типичным примером является ситуация, когда перед ме­неджером стоит задача наладить производство на основе прогнозирования спроса на не­сколько сотен наименований продуктов, образующих одну линию. В данном случае наи­более оправданно использование методов сглаживания.

Главным преимуществом методов экспоненциального сглаживания является их низ­кая стоимость и простота. Они не дают такой точности, как сложные методы, например ARIMA. Но при построении прогнозов для тысяч наименований продуктов, методы сглаживания зачастую являются единственным разумным подходом.

Прогнозы перспектив, основанные на временных рядах, опираются на предположе­ние о том, что развитие будущих событий будет подобно прошлому, а структура прошлых событий поддается адекватному описанию. Методика временных рядов является одной из наиболее часто применяемых для прогнозирования переменных, с постоянной и ста­бильной структурой изменений.

Методология Бокса-Дженкинса является очень мощным инструментом точного крат­косрочного прогнозирования. Менеджеры должны учитывать, что создание удовлетвори­тельной модели ARIMA по методике Бокса-Дженкинса требует довольно большого ко­личества исторических данных и значительных затрат времени аналитика.

Практических применений методики Бокса-Дженкинса очень много. Модели ARIMA реально применялись для следующих целей:

1. оценка изменений в структуре цен в телефонной индустрии США;
2. изучение взаимосвязи между концентрацией аммония, скоростью течения и тем­пературой воды в реках;
3. прогнозирование годовых объемов запасов;
4. прогнозирование количества действующих нефтяных скважин;
5. анализ количества построенных частных жилищных единиц;
6. анализ ежедневных наблюдений процентного роста количества единиц продавае­мого товара;
7. анализ конкуренции между авиа- и железнодорожными перевозками;
8. прогнозирование уровня занятости;
9. анализ большого числа временных рядов энергопотребления для коммунальных  
   предприятий;
10. анализ эффектов стимулирования продаж потребительских продуктов;
11. прогнозирование различных категорий гарантий качества продукции.

**Пример 2**. Аналитиком компании **Y**, был подготовлен временной ряд данных для производственного процесса, который необходимо спрогнозировать. Собранные им данные представлены в таблице 2, а соответствующий график показан на рис.7. Представляется, что метод Бокса-Дженкинса будет наиболее подходящим для обработки собранных данных.

Таблица 2 Значения выпусков продукции компании Atron

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Период | Выпуск | Период | Выпуск | Период | Выпуск | Период | Выпуск | Период | Выпуск |
| 1 | 60,0 | 16 | 88,5 | 31 | 79,5 | 46 | 84,0 | 61 | 72,0 |
| 2 | 81,0 | 17 | 76,5 | 32 | 64,5 | 47 | 73,5 | 62 | 66,0 |
| 3 | 72,0 | 18 | 82,5 | 33 | 99,0 | 48 | 78,0 | 63 | 73,5 |
| 4 | 78,0 | 19 | 72,0 | 34 | 72,0 | 49 | 49,5 | 64 | 66,0 |
| 5 | 61,5 | 20 | 76,5 | 35 | 78,0 | 50 | 78,0 | 65 | 73,5 |
| 6 | 78,0 | 21 | 75,0 | 36 | 63,0 | 51 | 88,5 | 66 | 103,5 |
| 7 | 57,0 | 22 | 78,0 | 37 | 66,0 | 52 | 51,0 | 67 | 60,0 |
| 8 | 84,0 | 23 | 66,0 | 38 | 84,0 | 53 | 85,5 | 68 | 81,0 |
| 9 | 72,0 | 24 | 97,5 | 39 | 66,0 | 54 | 58,5 | 69 | 87,0 |
| 10 | 67,8 | 25 | 60,0 | 40 | 87,0 | 55 | 90,0 | 70 | 73,5 |
| 11 | 99,0 | 26 | 97,5 | 41 | 61,5 | 56 | 60,0 | 71 | 90,0 |
| 12 | 25,5 | 27 | 61,5 | 42 | 81,0 | 57 | 78,0 | 72 | 78,0 |
| 13 | 93,0 | 28 | 96,0 | 43 | 76,5 | 58 | 66,0 | 73 | 87,0 |
| 14 | 75,0 | 29 | 79,5 | 44 | 84,0 | 59 | 97,5 | 74 | 99,0 |
| 15 | 57,0 | 30 | 72,0 | 45 | 57,0 | 60 | 64,5 | 75 | 72,0 |

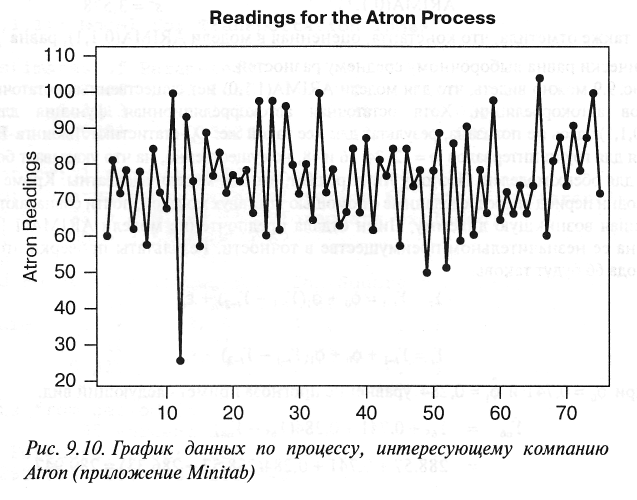


Рисунок 7- График данных по производственному процессу, интересующему компанию Atron

Начнем поиск пробной модели с анализа графика данных и графика функции выборочной автокорреляции, показанного на рис. 8. Исходный временной ряд данных характеризуется вариацией значений в окрестности фиксированного уровня, приблизительно равного 80, а значения коэффициентов автокорреляции быстро убывают до нуля. Исходя из этого, можно сделать вывод, что данный временной ряд является стационарным.

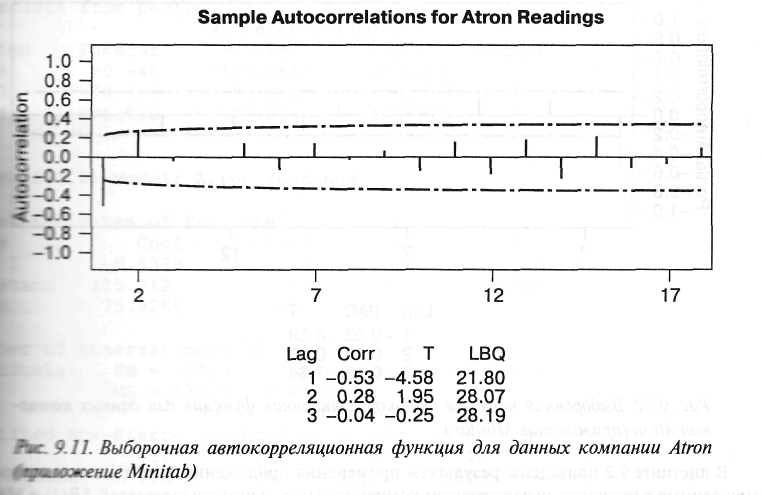


Рисунок 8 - Выборочная автокорреляционная функция для данных компании Atron

Первый выборочный коэффициент автокорреляции (-0,53) существенно отличается нуля для уровня 5%, поскольку находится вне диапазона

.

Автокорреляция для запаздывания в 2 периода ближе к пороговому значению для уровня 5% и противоположна по знаку автокорреляции r1на интервале 1. Остальные автокорреляции малы инаходятся в рамках установленных предельных ошибок. Можно предположить, что подобная структура коэффициентов автокорреляции соответствует либо модели AR(1), либо, что также допустимо, модели МА(2), если считать, что автокорреляции отсекаются (неотличимы от нуля) уже после второго интервала. В результате принимаем решение проанализировать график функции выборочной частной автокорреляции, показанный на рис. 9.

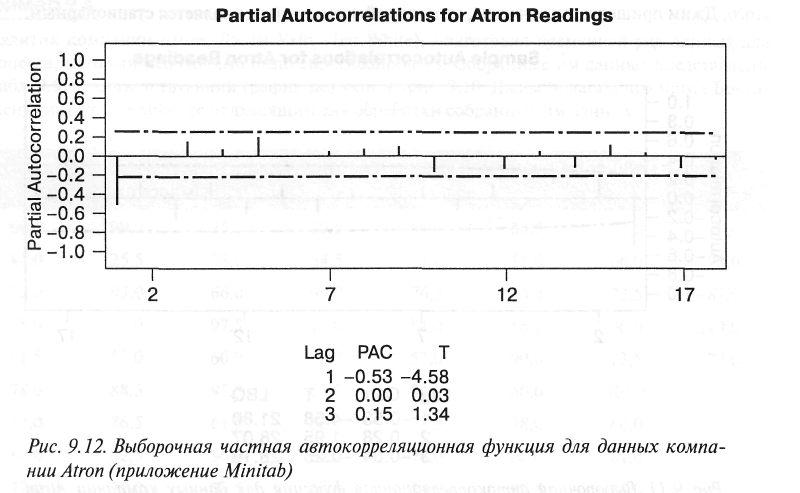


Рисунок 9 - Выборочная частная автокорреляционная функция для данных компании

Заметим, что первый коэффициент частной автокорреляции (-0,53) значительно отличается от нуля, но ни один из остальных коэффициентов частной автокорреляции не приближается к уровню значащего значения.. В результате приходим к заключению, что поведение функций выборочной автокорреляции и выборочной частной автокорреляции соответствует модели AR(1) (или, что то же самое, ARIMA(1,0,0)), однако чтобы полностью исключить риск, смоделируем данные также с помощью модели МА(2) (или АRIМА(0,0,2)). Если обе модели окажутся адекватными, возможно выбрать лучшую модель, исходя из принципа экономии (*Принцип экономии* состоит в предпочтении простой модели более сложной).

Постоянное слагаемое включено в обе модели, чтобы учесть тот факт, что данные изменяют­ся в окрестности уровня, отличного от нуля (если бы данные выражались как отклонение от выборочного среднего, то в обеих моделях постоянное слагаемое было бы ненужным).

Обемодели показали хорошее соответствие данным. Оцененные коэффициенты значительно отличаются от нуля. Среднеквадратические ошибки сходны.

МА(2): s2=135,1

AR(1): s2=137,9.

Прогнозы на один и два периода вперед для двух этих моделей отличаются в некоторых деталях, однако прогнозы на три периода вперед (период 78) весьма близки. При фиксированном источнике для предсказаний, прогнозы для стационарных процессов становятся, в конечном счете, равны предполагаемому среднему уровню. В рассматриваемом случае предполагаемый средний уровень приблизительно равен *ϻ* = 75 для обеих моделей.

Qm-статистика Льюинга-Бокса (модифицированная статистика Бокса-Пирса) незначи­тельна для коэффициентов корреляции на интервалах *т =* 12, 24, 36 и 48 для обеих моделей. Отдельные остаточные коэффициенты автокорреляции малы и находятся в рамках их предельных оши­бок. Остаточная автокорреляционная функция для модели МА(2) аналогична. Не вызывает сомнений тот факт, что ошибки случайны в обеих этих моделях.

Поскольку модель AR(1) имеет два параметра (включая постоянное слагаемое), а модель МА(2) — три (включая постоянное слагаемое), то, в соответствии с принципом экономии, для прогноза будущих значений данных решил воспользоваться более простой моделью AR(1).

Уравнение прогноза AR(1) будет иметь вид

ŷt *=* 115,842 + (-0,538) yt-1 = 115,842 - 0,538 yt-1,так что для периода 76

ŷt = 115,842 - 0,538 y75 = 115,842 - 0,538(72) = 77,11.

Помимо этого, прогноз на два периода вперед будет следующим.

ŷ77 = 115,842-0,538 y76 = 115,842-0,538(77,11) = 74,3.

**Пример 3**. Аналитик компании Atron, решил воспользоваться методом Бокса-Дженкинса для прогнозирования ошибок (отклонения от намеченных объемов производства компании), обнаруживаемых при контроле каче­ства производственного процесса, находящегося под его управлением. Соответствующие данные приведены в таблице 3, а график этого временного ряда ошибок показан на рис. 10.

Таблица 3 Ошибки, обнаруженные при контроле качества в компании Atron (произв. 1)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Период  (П) | Ошибка | (П) | Ошибка | (П) | Ошибка | (П) | Ошибка | (П) | Ошибка |
| 1 | -0,23 | 19 | -0,20 | 37 | -1,93 | 55 | -0,97 | 73 | 0,10 |
| 2 | 0,63 | 20 | 0,21 | 38 | 1,87 | 56 | 0,83 | 74 | -0,62 |
| 3 | 0,48 | 21 | 0,91 | 39 | -0,97 | 57 | -0,33 | 75 | 2,27 |
| 4 | -0,83 | 22 | -0,36 | 40 | 0,46 | 58 | 0,91 | 76 | -0,62 |
| 5 | -0,03 | 23 | 0,48 | 41 | 2,12 | 59 | -1,13 | 77 | 0,74 |
| 6 | 1,31 | 24 | 0,61 | 42 | -2,11 | 60 | 2,22 | 78 | -0,16 |
| 7 | 0,86 | 25 | -1,38 | 43 | 0,70 | 61 | 0,80 | 79 | 1,34 |
| 8 | -1,28 | 26 | -0,04 | 44 | 0,69 | 62 | -1,95 | 80 | -1,83 |
| 9 | 0 | 27 | 0,90 | 45 | -0,24 | 63 | 2,61 | 81 | 0,31 |
| 10 | -0,63 | 28 | 1,79 | 46 | 0,34 | 64 | 0,59 | 82 | 1,13 |
| 11 | 0,08 | 29 | -0,37 | 47 | 0,60 | 65 | 0,71 | 83 | -0,87 |
| 12 | -1,30 | 30 | 0,40 | 48 | 0,15 | 66 | -0,84 | 84 | 1,45 |
| 13 | 1,48 | 31 | -1,19 | 49 | -0,02 | 67 | -0,11 | 85 | -1,95 |
| 14 | -0,28 | 32 | 0,98 | 50 | 0,46 | 68 | 1,27 | 86 | -0,51 |
| 15 | -0,79 | 33 | -1,51 | 51 | -0,54 | 69 | -0,80 | 87 | -0,41 |
| 16 | 1,86 | 34 | 0,90 | 52 | 0,89 | 70 | -0,76 | 88 | 0,49 |
| 17 | 0,07 | 35 | -1,56 | 53 | 1,07 | 71 | 1,58 | 89 | 1,54 |
| 18 | 0,09 | 36 | 2,18 | 54 | 0,20 | 72 | -0,38 | 90 | -0,96 |

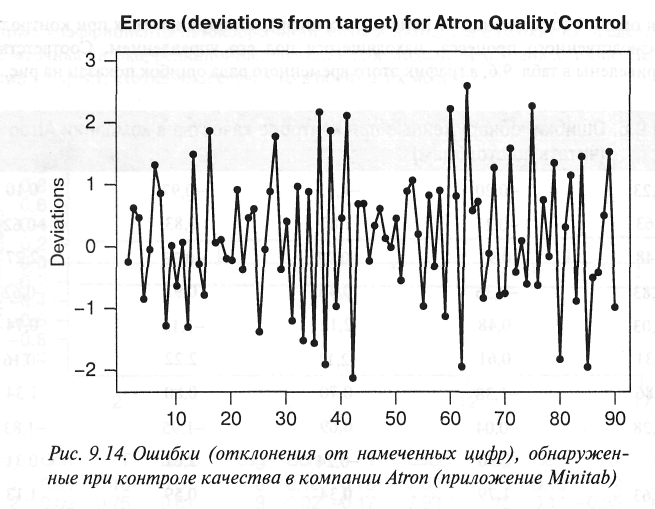


Рисунок 10 - Ошибки (отклонения от намеченных значений объемов), обнаруженные при контроле качества в компании Atron

Начнем процесс определения модели с изучения графика временного ряда ошибок, а также проверки функций автокорреляции и частной автокорреляции, показанных на рис. 11 и 12.

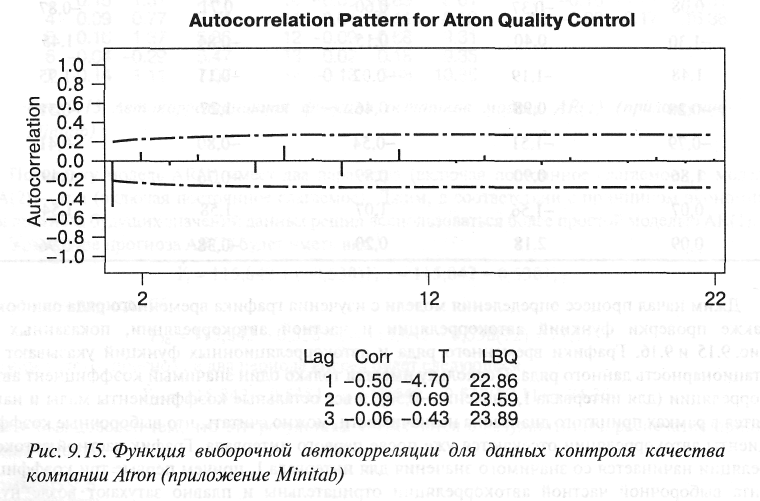


Рисунок 11 - Функция выборочной автокорреляции для данных контроля качества компании Atron

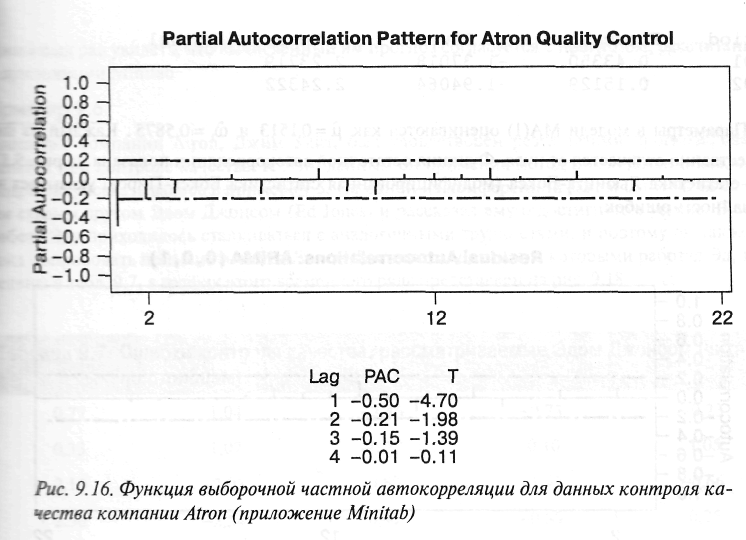


Рисунок 12 - Функция выборочной автокорреляции для данных контроля качества

Графики временного ряда и автокорреляционных функций указывают на стационарность данного ряда. Поскольку имеется только один значимый коэффициент авто­корреляции (для интервала 1, значение-0,50), а все остальные коэффициенты малы и нахо­дятся в рамках принятого диапазона незначимости, можно считать, что выборочные коэффи­циенты автокорреляции отсекаются уже после первого интервала.

График частной автокор­реляции начинается со значимого значения для интервала 1, причем первые три коэффици­ента выборочной частной автокорреляции отрицательны и плавно затухают возле нуля. Можно сделать вывод, что поведение выборочных коэффициентов автокорреляции и частной автокорреляции весьма сходно с теоретическими показателями для процесса МА(1) (или ARIMA(0,0,1))*.* Приходим к заключению, что исследуе­мый временной ряд можно описать с помощью модели МА(1).

Параметры в модели МА(1) оцениваются как *ϻ* = 0,1513 и ω1=0,5875. Каждый из них существенно отличается от нуля. Функция остаточной автокорреляции показана на рис. 13, а ᵡ2-статистика Льюинга-Бокса (модифицированная статистика Бокса-Пирса) указывает на случайность ошибок.

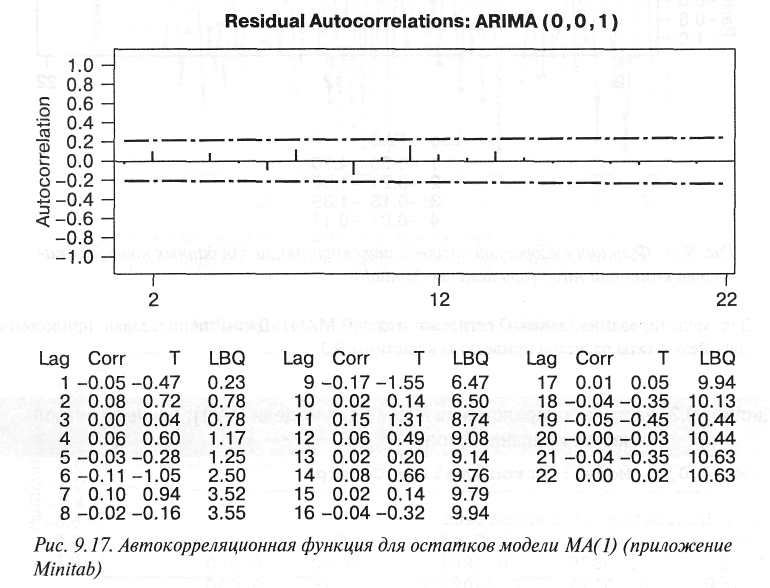


Рисунок 13 - Автокорреляционная функция для остатков модели МА(1)

Уравнение прогноза по модели МА(1) будет следующим:

ŷt = *ϻ - ω1εt-1,*

где εt-1 оценивается с помощью соответствующего остатка et-1. Для прогноза ошибки (отклонения от намеченных цифр) на период 91 нужен остаток для периода 90, e90*=* -0,4804. Вычисляем следующее:

ŷ91= 0,1513 - 0,5875(-0,4804) = 0,4335.

Прогноз относительно ошибки контроля качества в период 92 является просто пред­полагаемым средним ряда, так как, в начале прогноза *t =* 90, наилучшей оценкой порядка ошибки в период 91, ε91, является нуль. Таким образом,

ŷ92= 0,1513 -0,5875(0) = 0,1513.

**Пример 4**. Интерес представляет прогнозирование ошибок при контроле качества объема выпусков другого производства компании Atron. Попробуем применить метод Бокса-Дженкинса, с этим данным (таблица 4), а график этого временного ряда представлен на рис. 14.

Таблица 4. Ошибки, обнаруженные при контроле качества в компании Y (произв. 2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Период | Ошибка | Период | Ошибка | Период | Ошибка | Период | Ошибка |
| 1 | 0,77 | 21 | -0,23 | 41 | 2,16 | 61 | -0,75 |
| 2 | 0,33 | 22 | 0,49 | 42 | 0,04 | 62 | -1,24 |
| 3 | 2,15 | 23 | -0,87 | 43 | 1,91 | 63 | -0,62 |
| 4 | 2,50 | 24 | 0,61 | 44 | 0,43 | 64 | -0,54 |
| 5 | 1,36 | 25 | 0,20 | 45 | -0,32 | 65 | -0,23 |
| 6 | 0,48 | 26 | 0,98 | 46 | -0,48 | 66 | 1,05 |
| 7 | 2,05 | 27 | 0,78 | 47 | -0,13 | 67 | -0,66 |
| 8 | -1,46 | 28 | 0,80 | 48 | -2,26 | 68 | 0,25 |
| 9 | -1,13 | 29 | 0,86 | 49 | -0,73 | 69 | -0,63 |
| 10 | -2,85 | 30 | 1,72 | 50 | 0,10 | 70 | 0,91 |
| 11 | -2,67 | 31 | 0,15 | 51 | -1,47 | 71 | -0,21 |
| 12 | -2,71 | 32 | -1,15 | 52 | -0,89 | 72 | 0,24 |
| 13 | -1,30 | 33 | -2,46 | 53 | -0,53 | 73 | 0,05 |
| 14 | -0,88 | 34 | -0,37 | 54 | -0,20 | 74 | 0,85 |
| 15 | -0,07 | 35 | 0,80 | 55 | -0,70 | 75 | 1,55 |
| 16 | -1,47 | 36 | 0,49 | 56 | -0,27 | 76 | 0,40 |
| 17 | 1,04 | 37 | 0,50 | 57 | 0,39 | 77 | 1,82 |
| 18 | 1,02 | 38 | 0,07 | 58 | -0,07 | 78 | 0,81 |
| 19 | -2,03 | 39 | 1,92 | 59 | 0,89 | 79 | 0,28 |
| 20 | -2,54 | 40 | 1,00 | 60 | 0,37 | 80 | 1,06 |

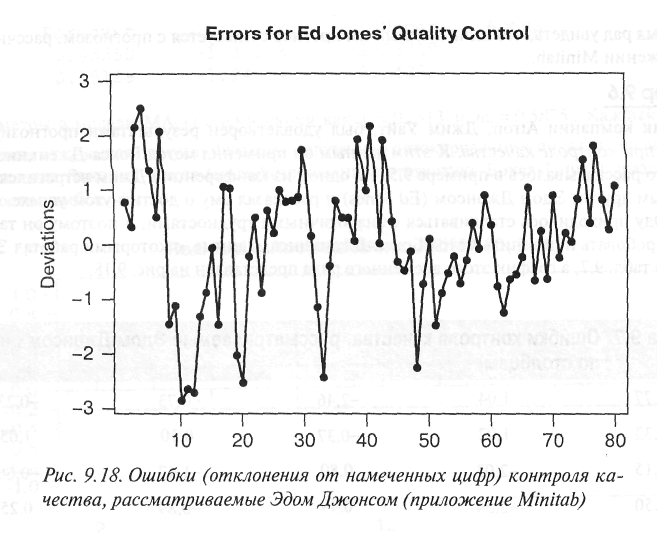


Рисунок 14 - Ошибки (отклонения от намеченных цифр) контроля качества

Общий вид графиков исходного временного ряда и функции выборочной автокорреляции (рис. 15) наводит на мысль о том, что исходный ряд ошибок контроля качества является стационарным. Значения ошибок колеблются около фиксированного уровня — нуля, а авто­корреляции быстро и плавно затухают.

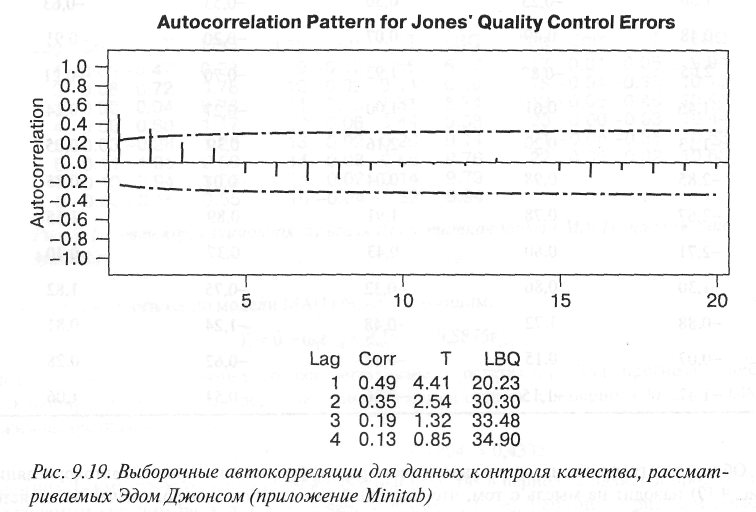


Рисунок 15 - Выборочные автокорреляции для данных контроля качества

Отметим, что два первых коэффициента автокорреляции существенно отличны от нуля и, что, при прочих равных условиях, более важно, коэффициенты автокорреляции для первых нескольких ин­тервалов затухают подобно тому, как это определено в теоретическом описании процессов типа AR(1)*.* Проанализируем также график функции выборочной частной автокорреляции, представленный на рис. 16. Все коэффициенты частной автокорреляции, кроме первого, практически незначимы. В совокупности структура функций выборочной автокорреляции и выборочной частотной автокорреляции точно соответствовала процессам типа AR(p). Поэтому представляется, что данные ряда ошибок (отклонений от намеченных значений выпусков) можно адекватно смоделировать как процесс AR(1) или же ARIMA(1,0,0).

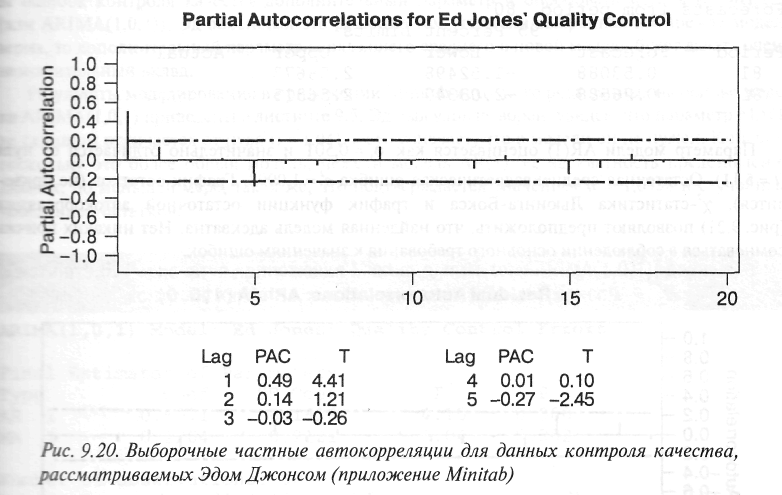


Рисунок 16 - Выборочные частные автокорреляции для данных контроля качества

Поскольку выборочное среднее ряд ошибокчрезвычайно мало (порядка нуля) по сравнению со стандартным отклонением, вмодель не будет включен постоянный член.

Параметр модели AR(1) оценивается как φ1=0,501 и значительно отличается от нуля (t = 5,11). Остаточная среднеквадратическая ошибка *s2 =* 1,0998. ᵡ2-статистика Льюинга-Бокса и график функции остаточной автокорреляции (рис. 17) позволяют предположить, что найденная модель адекватна. Нет никаких причин сомневаться в соблюдении основного требования к значениям ошибок.

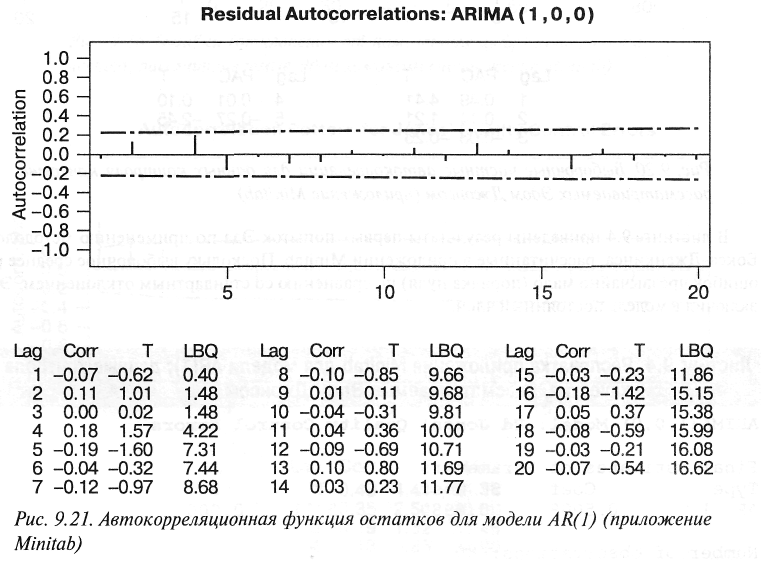


Рисунок 17 - Автокорреляционная функция для модели AR(1)

Уравнение прогноза имеет следующий вид:

ŷt = 0,501yt-1.

Таким образом, прогнозы на периоды 81 и 82 будут следующими:

ŷ81 = 0,501y80 = 0,501(1,06) = 0,531

ŷ82 = 0,501y81 = 0,501(0,531) = 0,266.

Опробуем чуть более сложную модель, чтобы получить результаты, ко­торые подтверждали бы выбор в пользу модели AR(1). Используем для анали­за ошибок контроля качества дополнительный параметр и опробуем модель ARMA(1,1) (или ARIMA(1,0,1)). Последнее можно обосновать тем, что если выбранная прежде модель верна, то дополнительный параметр скользящего среднего в новой модели будет давать очень незначительный вклад.

Результаты моделирования исходного ряда данных на основе моде­ли ARIMA(1,0,1) показали, что параметр МА(1) неслишком отличается от нуля *(t =* 1,04), а это означает, что в модели он не нужен. Конечно, поскольку это более общая модель, нежели модель AR(1), ее адекватность представления данных, по меньшей мере, такая же, что подтверждается значением *s2* = 1,0958 и случайным поведением остатков.

Пример 4. Рассмотрим прогно­зирование объемов продаж компании Keytron. Имеются данные об объемах продаж за 115 месяцев. Эти данные, охватывающие период с января 1987 г. по август 1996 г., представлены в таблице 5

Таблица 5 Ежемесячные объемы продаж компании Keytron

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Месяц | Объем | Месяц | Объем | Месяц | Объем |
| 1 | 1736,8 | 41 | 1796,6 | 82 | 2441,4 |
| 2 | 1297,4 | 42 | 1822,6 | 83 | 2113,8 |
| 3 | 559,0 | 43 | 1835,6 | 84 | 2035,8 |
| 4 | 1455,6 | 45 | 1944,8 | 85 | 2152,8 |
| 5 | 1526,2 | 46 | 2009,8 | 86 | 1708,2 |
| 6 | 1419,6 | 47 | 2116,4 | 87 | 806,0 |
| 7 | 1484,6 | 48 | 1994,2 | 88 | 2028,0 |
| 8 | 1651,0 | 49 | 1895,4 | 89 | 2236,0 |
| 9 | 1661,4 | 50 | 1947,4 | 90 | 2028,0 |
| 10 | 1851,2 | 51 | 1770,6 | 91 | 2100,8 |
| 11 | 1617,2 | 52 | 626,6 | 92 | 2327,0 |
| 12 | 1614,6 | 53 | 1768,0 | 93 | 2225,6 |
| 13 | 1757,6 | 54 | 1840,8 | 94 | 2321,8 |
| 14 | 1302,6 | 55 | 1804,4 | 95 | 2275,0 |
| 15 | 572,0 | 56 | 2007,2 | 96 | 2171,0 |
| 16 | 1458,6 | 57 | 2067,0 | 97 | 2431,0 |
| 17 | 1567,8 | 58 | 2048,8 | 98 | 2165,8 |
| 18 | 1627,6 | 59 | 2314,0 | 99 | 780,0 |
| 19 | 1575,6 | 60 | 2072,6 | 100 | 2056,6 |
| 20 | 16,82,2 | 61 | 2134,6 | 101 | 2340,0 |
| 21 | 1710,0 | 62 | 1799,2 | 102 | 2033,2 |
| 22 | 1853,8 | 63 | 756,6 | 103 | 2288,0 |
| 23 | 1788,8 | 64 | 1890,2 | 104 | 2275,0 |
| 24 | 1822,4 | 65 | 2256,8 | 105 | 2581,8 |
| 25 | 1838,2 | 66 | 2111,2 | 106 | 2540,2 |
| 26 | 1635,4 | 67 | 2080,0 | 107 | 2519,4 |
| 27 | 618,8 | 68 | 2191,8 | 108 | 2267,2 |
| 28 | 1593,8 | 69 | 2202,2 | 109 | 2615,6 |
| 29 | 1898,0 | 70 | 2449,2 | 110 | 2163,2 |
| 30 | 1911,0 | 71 | 2090,4 | 111 | 899,6 |
| 31 | 1695,0 | 72 | 2184,0 | 112 | 2210,0 |
| 32 | 1757,6 | 73 | 2267,2 | 113 | 2376,4 |
| 33 | 1944,8 | 74 | 1705,6 | 114 | 2259,4 |
| 34 | 2108,6 | 75 | 962,0 | 115 | 2584,4 |
| 35 | 1895,4 | 76 | 1929,2 |  |  |
| 36 | 1822,6 | 77 | 2202,2 |  |  |
| 37 | 2054,0 | 78 | 1903,2 |  |  |
| 38 | 1544,4 | 79 | 2337,4 |  |  |
| 39 | 600,6 | 80 | 2022,8 |  |  |
| 40 | 1604,2 | 81 | 2225,6 |  |  |

Внимательно изучив временной ряд, график которого показан на рис. 18, можно об­наружить в нем, наряду с возрастающим трендом, отчетливо проявляющуюся сезонную структуру. Приходим к заключению, что данный ряд является нестационарным, и поэтому следует применить к нему сезонную модель ARIMA

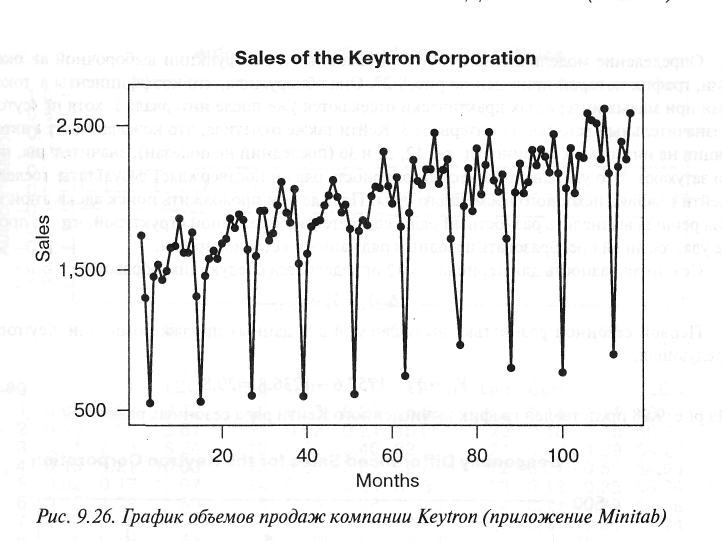


Рисунок 18 - График объемов продаж компании

Определение модели данных начнем с изучения функции выборочной автокорреля­ции, график которой приведен на рис. 19. Коэффициенты автокорреля­ции при малых интервалах практически отсекаются уже после интервала 1, хотя присутствует и незначительный всплеск на интервале 3. Следует также отметить, что коэффициенты автокорре­ляции на интервалах сезонности, т.е. 12, 24 и 36 (последний не показан), значительны, но быст­ро затухают. Это указывает на нестационарность ряда и подтверждает результаты исследования графика исходного временного ряда.

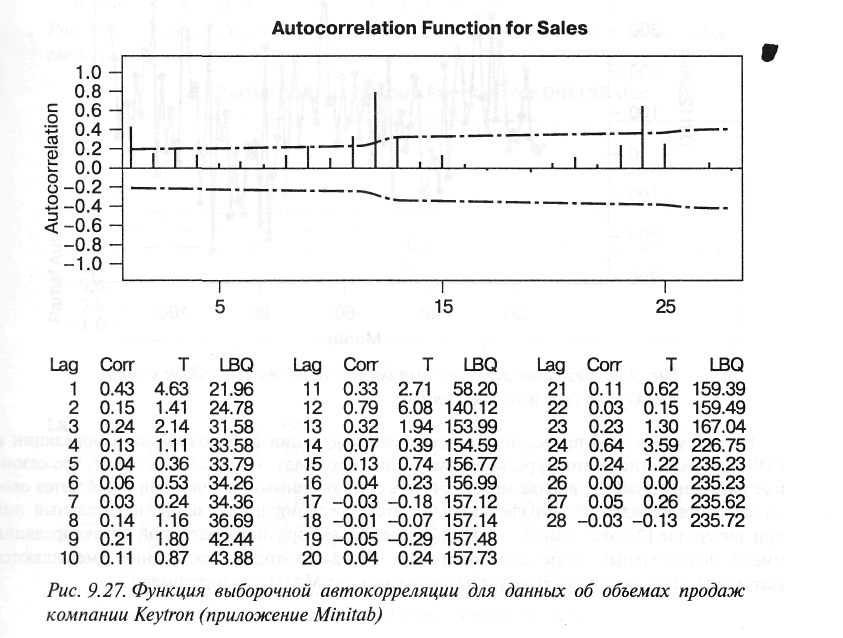


Рисунок 19 - Функция выборочной автокорреляции для данных об объемах компании

Прежде чем продолжить поиск адекватной модели, вычислим разностный ряд в соответствии с сезонной структурой, чтобы проверить, не удастся ли преобразовать исходный ряд данных в стационарный.

Сезонная разность для периода *S=*12 определяется следующим образом:

Δ12yt = yt - yt-12.

Первой сезонной разностью, вычисляемой для данных продаж компании Keytron, будет следующая:

y13 – y1 = 1757,6 - 1736,8 = 20,8.

На рис. 20 представлен график вычисленного ряда сезонных разностей.

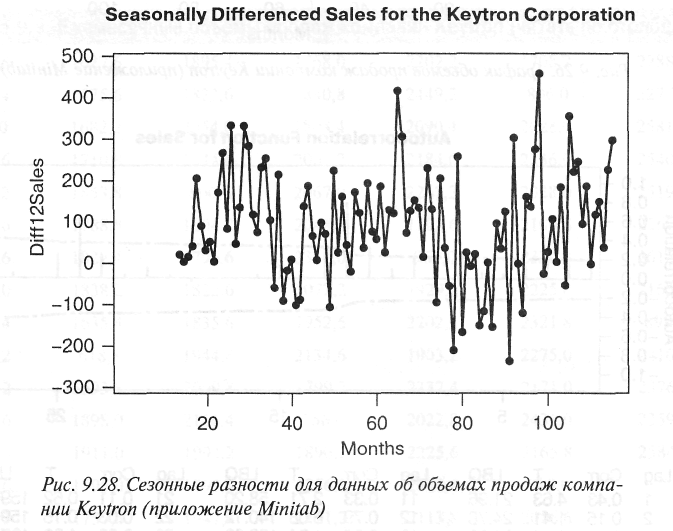


Рисунок 20 - Сезонные разности для данных об объемах продаж компании

На рис. 21 и 22 приведены соответственно функции выборочной автокорреляции и выборочной частной автокорреляции для разностного ряда. Из рис. 19 следует, что сезон­ные разностные данные вполне можно считать стационарными, причем они колеблются око­ло значения порядка 100. Коэффициенты автокорреляции имеют один значительный пик при интервале 12 (отсеченный), а коэффициенты выборочной частичной автокорреляции имеют значительные пики при интервалах 12 и 24, которые постепенно уменьшают» (затухают). Подобное поведение указывает на элемент МА(1) при интервале 12.

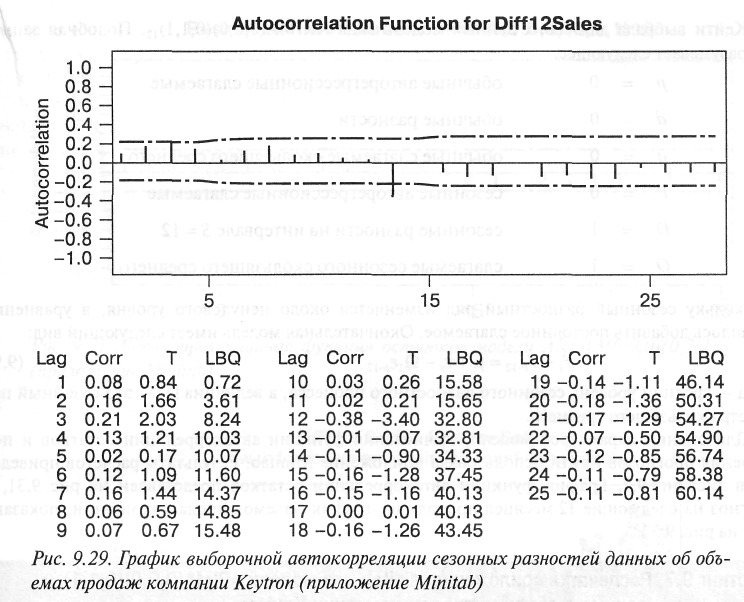


Рисунок 21 - График выборочной автокорреляции сезонных разностей   
данных об объемах продаж компании

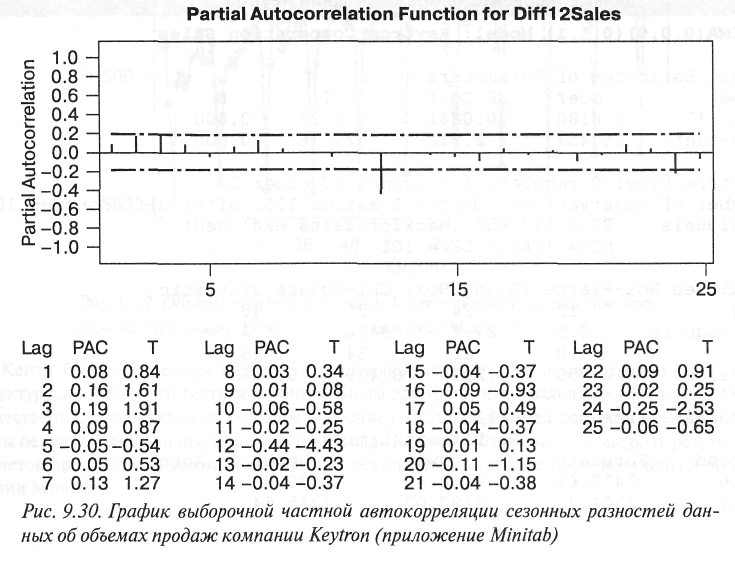


Рисунок 22 - График выборочной частной корреляции сезонных разностей данных об объемах продаж компании

Выберем для данных модель вида ARIMA(0,0,0)(0,1,1). Подобная запись подразумевает следующее.

*р* = 0 - обычные авторегрессионные слагаемые

*d* = 0 - обычные разности

*q* = 0 - обычные слагаемые скользящего среднего

*Р =* 0 - сезонные авторегрессионные слагаемые

*D =* 1 - сезонные разности на интервале 5-12

*Q =* 1 - слагаемые сезонного скользящего среднего.

Поскольку сезонный разностный ряд изменяется около ненулевого уровня, в уравнение пришлось добавить постоянное слагаемое. Окончательная модель имеет следующий вид

yt - yt-12 =*ϻ* + εt + ψ1εt-12, (87)

где *ϻ* — средний уровень сезонного разностного процесса, а величина ψ *—* это сезонный па­раметр скользящего среднего.

График функции автокорреляции остатков представлен на рис. 23, а прогноз на следующие 12 месяцев продолжает график объемов продаж компании (рис. 24).

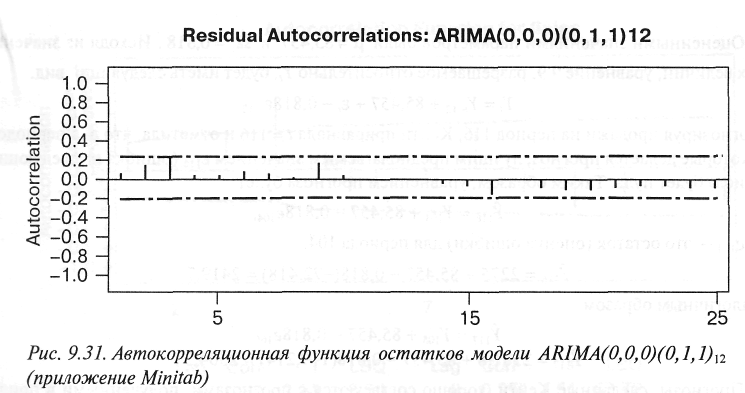


Рисунок 23 - Авторегрессионная функция остатков модели ARIMA(0,0,0)(0,1,1)

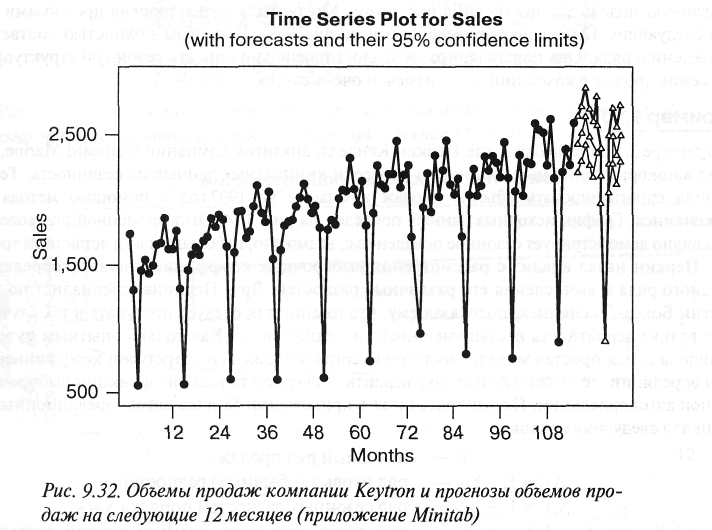


Рисунок 24 - Объемы продаж компании и прогнозы объемов продаж

Получаем, что первоначальная модель хорошо описывает структуру данных. ᵡ2- статистика Льюинга-Бокса для групп интервалов *т =* 12, 24, 36 и 48 не существенна, что показывает большое значение р. Автокорреляции остатков все одинаково малы без какой-либо видимой структуры.

Оцененными значениями параметров были  *ϻ* = 85,457 и ψ = 0,818. Исходя из значений этих величин, уравнение (87), разрешаемое относительно yt*,* будет иметь следующий вид:

yt = yt-12 + 85,457 + 0,818εt-12.

Прогнозируя продажи на период 116, приравниваем *t =* 116 и видим, что для периодов, на которые делается прогноз, лучшим предполагаемым значением ε116(ошибка на следующий период) будет нуль. Таким образом, уравнением прогноза будет

ŷ116 = y114 + 85,147 – 0,818e104,

где e104— это остаток (оценка ошибки) для периода 104.

ŷ116= 2275 + 85,457 - 0,818(-72,418) = 2419,7. Аналогичным образом

ŷ117 = y105+ 85,457-0,8186,05e105

ŷ117= 2581,8 + 85,457 – 0,818(119,214) = 2504,3.

Прогнозы полностью соответствуют поведению ряда. Можно предположить правильность описания сезонной структуры и в скором времени в компании будет отмечен рост объемов продаж.

**ЛИТЕРАТУРА**

1. Кендэл М. Временные ряды. - М.: "Финансы и статистика", 1981.

2. Рунова Л.П., Рунов И.Л. Анализ временных рядов и прогнозирование. Учебно-методические материалы по дисциплине   
“Методы социально-экономического прогнозирования” для студентов

са специальности “Математические методы в экономике”. Ростов-на-Дону, РГУ, 2006.

1. Скучалина Л. Н., Крутова Т. А. Организация и ведение базы данных временных рядов. Система показателей, методы определения, оценки прогнозирования информационных процессов. ГКС РФ. М., 1995.
2. Статистическое моделирование и прогнозирование. Учебное пособие / Под ред. А. Г. Гранберга. - М.:"Финансы и статистика", 1990.
3. Четыркин Е.Н. Статистические методы прогнозирования. -М.: ”Статистика”, 1975.