2.6. МОДЕЛИ АВТОРЕГРЕССИИ И МОДЕЛИ БОКСА-ДЖЕНКИНСА

Достаточно часто на практике встречаются стационарные процессы, каждое настоящее значение Y_t которых определяется предыдущими, накопленными ранее значениями Y_{t-1}, Y_{t-2} и т.д. То есть, имеет место взаимосвязь или корреляция между этими значениями. А поскольку коррелируют друг с другом значения одного и того же ряда, такое явление называют *автокорреляция*.

Для того чтобы определить насколько процесс является автокоррелированным, осуществляют расчет коэффициентов парной корреляции между значениями этого ряда и ими же, сдвинутыми на некоторый шаг назад. Такие коэффициенты будут называться *автокорреляционными*. Для их вычисления в формулу расчета коэффициента парной корреляции последовательно подставляют попарно сравниваемые значения показателя Y в момент t и показатели этого же процесса Y, но сдвинутые во времени на некоторый шаг τ, то есть $Y_{t-\tau}$:

$$r_\tau = \frac{\sum_{t=\tau+1}^{T}(Y_t - \overline{Y}_\tau)(Y_{t-\tau} - \overline{Y}^\tau)}{\sqrt{\sum_{t=\tau+1}^{T}(Y_t - \overline{Y}_\tau)^2 \sum_{t=\tau+1}^{T}(Y_{t-\tau} - \overline{Y}^\tau)^2}}, \qquad (2.6.1)$$

где

$$\overline{Y}_\tau = \frac{1}{T-\tau}\sum_{t=\tau+1}^{T} Y_t, \quad \overline{Y}^\tau = \frac{1}{T-\tau}\sum_{t=1}^{T-\tau} Y_t$$

Таким образом, в качестве двух случайных переменных, между которыми выявляется корреляция, выступают исходный ряд значений Y_t и ряд $Y_{t-\tau}$. Сам шаг τ изменяется от единицы до некоторого значения τ_M. Поэтому в распоряжении прогнозиста находится некоторая зависимость коэффициента парной корреляции r от шага τ:

$r = f(\tau)$.

Эту зависимость называют автокорреляционной функцией.

Наиболее наглядно свойства автокорреляции исходного ряда выявляются из графического анализа автокорреляционной функции. График зависимости значений коэффициента автокорреляции r_τ от шага τ называют коррелограммой. Анализ этого графика дает прогнозисту очень много ценной информации для выявления особенностей изучаемого процесса - периодичности некоторых явлений, их цикличности и сезонности. структура этой цикличности и т.п. Очевидно, что максимальные значения автокорреляционной функции могут изменяться в пределах от минус единицы до плюс единицы, а максимальное число сдвигов τ_M не должно быть близким к числу наблюдений показателей $\tau_M < T$.

Типичный график автокорреляционной функции приведён на рис. 2.6.1.

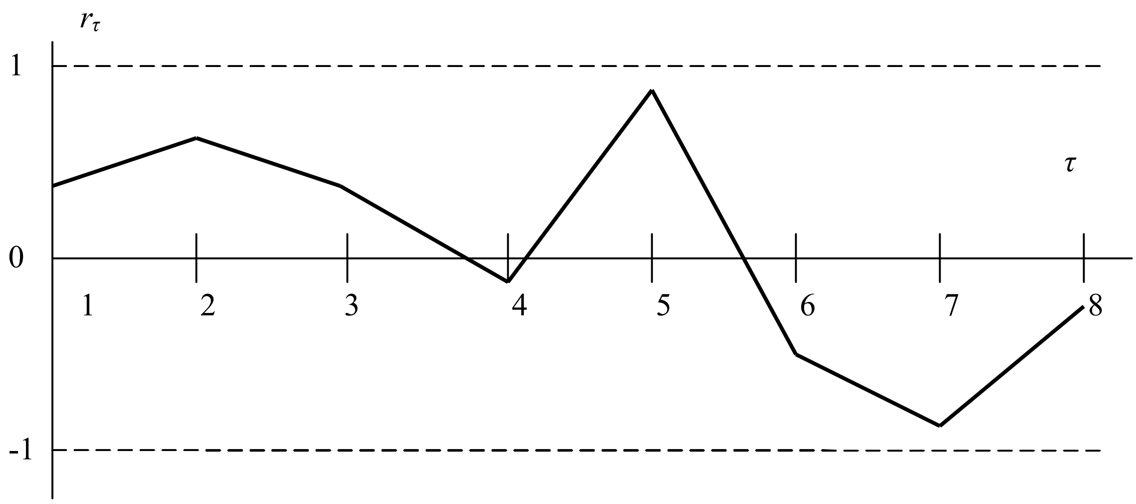

Рис. 2.6.1. Типичная коррелограмма при наличии автокорреляции

Для большей наглядности на график коррелограммы наносят не только значения коэффициентов автокорреляции при соответствующих сдвигах τ, но ещё и соединяют близлежащие точки отрезками прямых линий. Именно это и показано на рис. 2.6.1. В результате получается некоторая ломаная линия, максимумы и минимумы которой и являются предметом особого изучения, ведь они характеризуют приближение зависимости между значениями ряда Y_t и предыдущими значениями $Y_{t-\tau}$ к линейной, причём, чем ближе величина коэффициента автокорреляции при каком-то шаге τ к единице, тем ближе к линейной зависимость между указанными значениями.

Если при некотором сдвиге τ коэффициент автокорреляции по модулю окажется не менее чем *0,8*, то говорят о наличии этой зависимости, а сдвиг во времени τ, соответствующий этому высокому значению коэффициента, называют *лагом*.

Если автокорреляционная функция имеет несколько лагов, то говорят о том, что у этого ряда имеются *распределённые лаги*. Впрочем, иногда о распределённых лагах говорят, если показатель y_t находят в зависимости от другого фактора x_t.

Поскольку лаг означает наличие зависимости значений самого ряда от его же значений, но сдвинутых на величину лага, то эту зависимость можно описать математически. В общем случае модель авторегрессии может описываться следующей формулой:

$$\hat{Y}_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + a_2 Y_{t-3} + a_2 Y_{t-4} + \ldots \qquad (2.6.2)$$

Применительно к графику автокорреляционной функции рис. 2.6.1, на котором выделяются два лага, равные 5 и 7 соответственно, можно говорить о том, что модель авторегрессии будет содержать две переменные - Y_{t-5} и Y_{t-7}. Поскольку при лаге, равном пяти, коэффициент автокорреляции имеет положительный знак, то коэффициент при переменной Y_{t-5} будет положительным, а так как коэффициент автокорреляции при лаге, равном семи, имеет отрицательный знак, что свидетельствует об обратной линейной зависимости, то и коэффициент при переменной Y_{t-7} будет отрицательным:

$$\hat{Y}_t = a_0 + a_1 Y_{t-5} - a_2 Y_{t-7}$$

Для нахождения коэффициентов модели авторегрессии используются соответствующие разделы математической статистики, в большинстве случаев для этого используется МНК.

Интересным развитием моделей авторегрессии является раздел, который в нашей стране принято называть моделями Бокса-Дженкинса. Дж.Бокс и Г.Дженкинс широко пользуются различного рода операторами.

Первый из них – это оператор сдвига назад B, определяемый как[1]

$$Bz_t = z_{t-1}. \qquad (2.6.3)$$

Для сдвига назад на m используется соответствующий оператор:

$$B^m z_t = z_{t-m}. \qquad (2.6.4)$$

Обратная операция выполнятся оператором сдвига вперёд $F=B^{-1}$, задаваемым как

$$Fz_t = z_{t+1}. \qquad (2.6.5)$$

Для сдвига вперёд на m используется оператор:

$$F^m z_t = z_{t+m}. \qquad (2.6.6)$$

Кроме этих операторов предлагается использовать разностный оператор со сдвигом назад:

$$\nabla z_t = z_t - z_{t-1} = (1-B)z_t. \qquad (2.6.7)$$

В свою очередь оператор, обратный разностному оператору, - это оператор суммирования, выражаемый так:

$$Sz_t = \nabla^{-1} z_t = \sum_{\tau=0}^{\infty} z_{t-\tau} = z_t + z_{t-1} + z_{t-2} + ... = (1 + B + B^2 + ...)z_t = (1-B)^{-1} z_t. \qquad (2.6.8)$$

Авторегрессионная модель вида:

$$z_t = a_1 z_{t-1} + a_2 z_{t-2} + ... + a_p z_{t-p} + \varepsilon_t, \qquad (2.6.9)$$

где ε_t – случайная нормально распределённая величина с нулевым математическим ожиданием и конечной дисперсией (называемой «белым шумом»), представляет собой авторегрессию p-го порядка, которую с учётом введённых операторов можно записать в такой эквивалентной форме:

$$(1 - a_1 B - a_2 B^2 - ... - a_p B^p) z_t = \varepsilon_t \text{ или } a(B) z_t = \varepsilon_t. \qquad (2.6.10)$$

Откуда

$$z_t = a^{-1}(B)\varepsilon_t. \qquad (2.6.11)$$

В этой форме записи появляется новая трактовка взаимосвязи – белый шум ε_t как бы «фильтруется» через соответствующий оператор и превращается в переменную z_t. Поэтому иногда говорят, что рассматривается модель линейной фильтрации. Линейной она является потому, что используется линейный фильтр. Запись (2.6.11), как легко заметить, более компактная, нежели (2.6.9) и на основе этого принципа Дж.Бокс и Г.Дженкинс предложили

[1] Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. Вып. 1, – М.: Мир, 1974. - С. 23

многообразные модификации и расширения моделей авторегрессии для различных задач моделирования и прогнозирования стационарных процессов.

Модели фильтрации представляют собой весьма многообразный раздел прогнозирования временных рядов[1].

Отдельно рассматривается задача построения нелинейных авторегрессионных моделей. Легко заметить, что например нелинейная авторегрессионная модель вида:

$$\hat{Y}_t = a_0 + a_1 Y_{t-1}^2$$

будет неустойчивой, поскольку значения ряда на каждом последующем шаге возрастают по сравнению с предыдущим во второй степени. Поэтому, если уровни ряда превышают единицу – ряд возрастает, а если они меньше единицы – они стремятся к нулю с ростом числа шагов. Зная эти особенности авторегрессионных моделей, на практике используют «пороговые модели», когда при достижении некоторого порогового значения уровня моделируемого ряда происходит изменение модели авторегрессии; билинейные модели, когда предыдущие значения ряда раскладываются на две составляющие и некоторые другие.

[1] Моделі і методи соціально-економиного прогнозування: Підручник / Геэць В.М., Клебанова Т.С., Черняк О.І. и др. – Х.: ВД «ІНЖЕК», 2008. – 396 с.; Кобринский Н.Е. Информационные фильтры в экономике (Анализ одномерных временных рядов). – М.: Статистика, 1978. – 287 с.