22. Временные ряды. АКФ и ее использование при построении тренд-сезонной модели временного ряда. Определение сезонных составляющих.

Модели временных рядов

Если временной ряд представляется в виде суммы соответствующих компонент (см. 4.1), то полученная модель носит название **аддитивной** (от англ. *to add* – добавлять). Представление временного ряда в виде произведения компонент соответствует **мультипликативной** модели (4.2); также можно выделить еще один вид модели - **смешанного типа** (4.3):

$$y_t = u_t + s_t + v_t + \varepsilon_t; \qquad (4.1)$$

$$y_t = u_t \cdot s_t \cdot v_t \cdot \varepsilon_t; \qquad (4.2)$$

$$y_t = u_t \cdot s_t \cdot v_t + \varepsilon_t, \qquad (4.3)$$

где y_t – уровни временного ряда;

u_t – трендовая составляющая;

s_t – сезонная компонента;

v_t – циклическая компонента;

ε_t – случайная компонента.

Для выявления структуры временного ряда вычисляют коэффициенты автокорреляции разных порядков.

Коэффициентом автокорреляции r(τ) порядка τ называется коэффициент корреляции, вычисленный для исходного ряда данных, и того же ряда, но сдвинутого на τ интервалов времени:

$$\rho(\tau) = \frac{\mathrm{cov}(y_t, y_{t+\tau})}{D(y_t)}. \qquad (4.5)$$

Величину сдвига τ называют **лагом**.

Примечание: Автокорреляция измеряет, насколько сегодняшнее значение временного ряда "помнит" свои прошлые значения.

Простая аналогия:

Представьте, что вы измеряете температуру каждый день в 12:00:

- Автокорреляция с лагом 1 (τ=1): Насколько температура сегодня связана с температурой вчера?
- Автокорреляция с лагом 7 (τ=7): Насколько температура сегодня связана с температурой неделю назад?
- Автокорреляция с лагом 30 (τ=30): Насколько температура сегодня связана с температурой месяц назад?

Коэффициент корреляции называется коэффициентом **автокорреляции**,
так как характеризует статистическую связь между уровнями одного и того же временного ряда, отстоящими на τ временных тактов.

Последовательность коэффициентов автокорреляции уровней ряда с увеличением лага называется **автокорреляционной функцией**, *АКФ* (*autocorrelation function*, *ACF*)

Значения АКФ характеризуют тесноту (степень) статистической связи между уровнями временного ряда, разделенными τ временными тактами. Они безразмерны, не зависят от масштаба измерения уровней исследуемого временного ряда и подчиняются ограничению:

$$|p(\tau)| \leq 1.$$

(Очевидно, что $p(0) = 1$).

На практике значения АКФ статистически оцениваются по имеющимся уровням временного ряда. Выборочная оценка коэффициента автокорреляции $r(\tau)$ может быть определена следующим образом:

$$r(\tau) = \frac{\frac{1}{n-\tau}\sum_{t=1}^{n-\tau}(y_t - \overline{y})(y_{t+\tau} - \overline{y})}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \overline{y})^2} \qquad (4.6)$$

где n — длина временного ряда $y_1, y_2, \ldots, y_t, \ldots, y_n$;

τ — временной сдвиг ($\tau = 1, 2, \ldots, n-1$);

$\overline{y}$ — оценка среднего значения, найденная по формуле $\overline{y} = \frac{\sum_{t=1}^{n} y_t}{n}$.

Отметим, что числитель выражения (4.6) представляет выборочную оценку коэффициента автоковариации. График АКФ отражающий изменение величины $r(\tau)$ в зависимости от значений сдвига τ, называют коррелограммой (correlogram).

Очевидно, что с увеличением значения лага τ число пар наблюдений ($n - \tau$), используемых для расчета в (4.6), уменьшается.

Поэтому в практических руководствах рекомендуется поддерживать соотношение $\tau \leq n/4$. Для стационарного временного ряда с увеличением τ автокорреляционная функция должна демонстрировать свойство монотонного убывания по абсолютной величине, так как взаимосвязь между уровнями ряда с ростом τ ослабевает. Однако это условие может нарушаться для выборочных АКФ.

Прогнозирование сезонных колебаний

4-членная скользящая средняя:

$$\boxed{\hat{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4};} \qquad (4.11)$$

12-членная скользящая средняя:

$$\boxed{\hat{y}_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}} \qquad (4.12)$$

Рассмотрим алгоритм расчета для случая *мультипликативной*
сезонности.

1. Для описания тенденции воспользуемся процедурой скользящей средней при четной длине интервала сглаживания $\ell = 2p$. Тогда для временных рядов месячной динамики скользящая средняя при $\ell = 12$ на
   каждом активном участке будет определяться выражением (4.12), для рядов квартальной динамики можно использовать аналогичное выражение (4.11)
   при $\ell = 4$. Получим сглаженный ряд y'_t.
2. Рассчитаем отношение фактических значений y_t к уровням
   сглаженного ряда y'_t, полученным на предыдущем шаге:

$$x_t = \frac{y_t}{y'_t} \qquad (4.26)$$

Уровни вновь полученного ряда x_t отражают эффект сезонности и
случайности.

3. Для элиминирования (исключения из рассмотрения) влияния случайных факторов определим предварительные значения (оценки) сезонной составляющей как средние значения из уровней x_t для одноименных месяцев (кварталов). Например, для временных рядов месячной динамики процедура усреднения может быть описана выражением:

$$x_i = \begin{cases} \frac{1}{k}\sum_{j=1}^{k} x_{12j+i} & \text{при } i = 1, 2, \ldots, 6 \\ \frac{1}{k}\sum_{j=0}^{k-1} x_{12j+i} & \text{при } i = 7, 8, \ldots, 12 \end{cases}$$

где k — число целых периодов (циклов) во временном ряду, полученном на втором шаге.

Разные пределы суммирования объясняются тем, что при использовании скользящей средней с четным значением длины интервала сглаживания ($\ell = 2p$) p первых и p последних уровней ряда будут потеряны.
(В нашем случае потери составят по 6 уровней в начале и в конце ряда).

4. Проведем корректировку первоначальных значений сезонной составляющей, вызванную тем, что суммарное воздействие сезонности на динамику предполагается нейтральным. Это свойство для сезонных колебаний в мультипликативной форме выражается в том, что средняя арифметическая из значений коэффициентов сезонности для полного сезонного цикла должна быть равна 1. Поэтому окончательные оценки сезонного цикла должна быть равна 1. Поэтому окончательные оценки коэффициентов сезонности получим с помощью следующего выражения:

$$s_i = \bar{x}_i k \quad (i = 1, 2, \ldots, m), \qquad (4.27)$$

где $k = \frac{m}{\sum_{i=1}^{m} \overline{x_i}}$,

m — число фаз в полном сезонном цикле (как правило, $m = 12$ для рядов месячной динамики и $m = 4$ для квартальных данных).

Для **аддитивной формы** сезонности меняется содержание 2 и 4 этапов алгоритма. Очевидно, что на втором шаге знак деления необходимо заменить вычитанием:

$$x_t = y_t - y_t'. \qquad (4.28)$$

После получения предварительных оценок сезонности усреднением на шаге 3, следует выполнить процедуру их корректировки.

Взаимопогашаемость сезонных колебаний выражается в том, что для аддитивного случая сумма значений сезонной составляющей для полного сезонного цикла должна быть равна нулю. Поэтому окончательные, скорректированные оценки сезонной компоненты определяются с помощью следующего выражения:

$$S_i = \bar{x}_i - \bar{x} \quad (i = 1, 2, \ldots, m). \qquad (4.29)$$

где

$$\bar{x} = \frac{1}{m} \sum_{i=1}^{m} \overline{x_i},$$

m — число фаз в полном сезонном цикле.

Построение тренд-сезонных составляющих

Процедуру построения тренд-сезонных моделей можно описать в виде следующей последовательности шагов:

1. Оценивание сезонной составляющей с учетом характера сезонности (аддитивной или мультипликативной)
2. Десезонализация (сезонная корректировка) исходных данных
3. Расчет параметров тренда на основе временного ряда, полученного на втором шаге
4. Моделирование динамики исходного ряда с учетом трендовой и сезонной составляющих
5. Использование построенной модели для прогнозирования, в случае если исследователя удовлетворили полученные характеристики точности и адекватности модели.