Лекция 8

Временные ряды

1. *Понятие временного ряда и его составляющих*
2. *Стационарные временные ряды*
3. *Выравнивание временных рядов*
4. *Моделирование ряда при наличии циклических колебаний*

1. Понятие временного ряда и его составляющих.

Основная идея анализа ранее рассмотренных моделей заключается в том, что изменение результирующей переменной объясняется за счёт изменения одной или нескольких *другух* переменных.

В реальности результирующая переменная складывается под влиянием большого числа факторов, многие из которых не поддаются непосредственному наблюдению и измерению.

Поэтому наилучшим источником информации служат значения *самой исследуемой переменной* в прошлые моменты времени.

В этом случае мы имеем дело с другим видом статистических данных – *временными рядами* в отличие от *пространственной выборки*, как это было ранее.

Под *временным рядом* в экономике подразумевается совокупность наблюдений некоторого показателя Y, характеризующего один и тот же объект за несколько последовательных моментов или периодов времени.

Отдельные наблюдения этого показателя называются *уровнями* ряда и обозначаются символами y_t, $t = 1,2,...,n$, где n – число уровней ряда (число наблюдений).

Каждый уровень временного ряда y_t формируется под воздействием большого числа факторов, которые условно можно разделить на три группы:

- факторы, формирующие основную тенденцию ряда (трендовая компонента);
- факторы, определяющие циклические колебания ряда (циклическая компонента);
- случайные факторы (случайная компонента).

В большинстве случаев фактический уровень временного ряда можно представить как *сумму* или как *произведение* трендовой, циклической и случайных компонент.

Соответственно говорят об *аддитивной* или *мультипликативной* модели временного ряда.

Математическая запись этих моделей имеет вид:

- аддитивная модель $y_t = u_t + v_t + \varepsilon_t$;
- мультипликативная модель $y_t = u_t \cdot v_t \cdot \varepsilon_t$.

В этих уравнениях:

- u_t – *тренд*, описывающий влияние долговременных факторов, т.е. длительную, "вековую" тенденцию изменения признака , которая может быть либо возрастающей (рис. 1), либо убывающей (рис. 2);

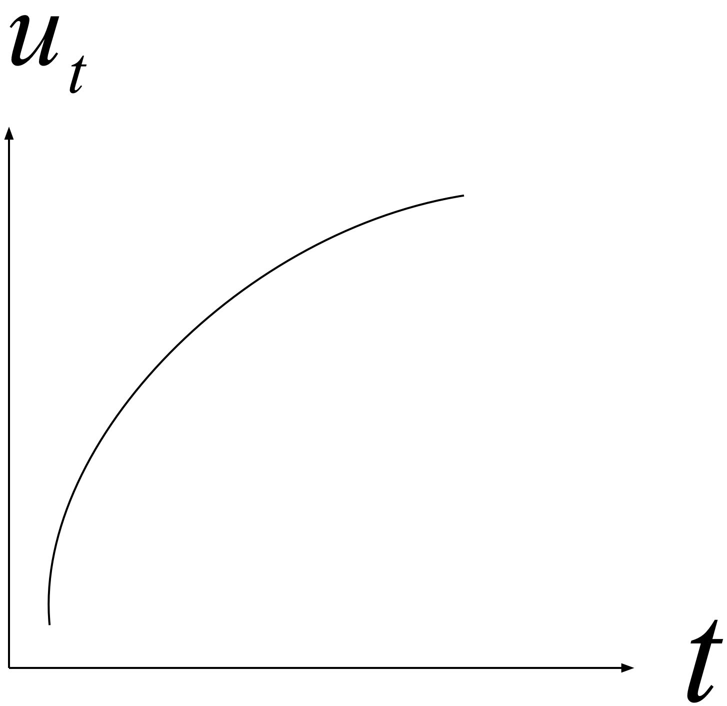

Рис. 1

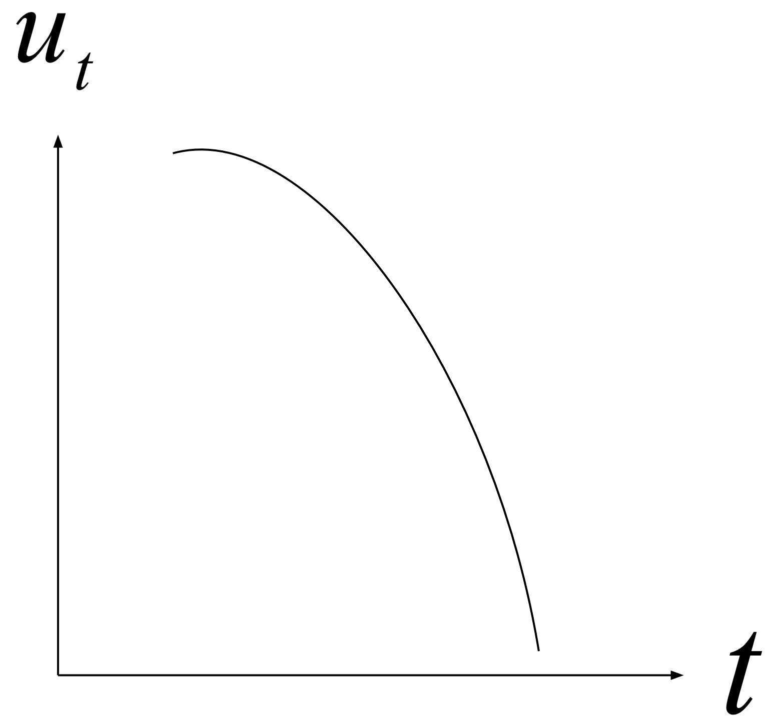

Рис. 2

- v_t – *циклическая* компонента, отражающая повторяемость экономических процессов. Циклические колебания могут носить *сезонный* характер, и связаны они с внутригодовыми колебаниями временного ряда. При наличии данных за более длительные промежутки времени могут выявляться *конъюнктурные* циклические колебания, формирующиеся под влиянием долговременных циклов экономической, демографической и прочей природы (рис. 3);

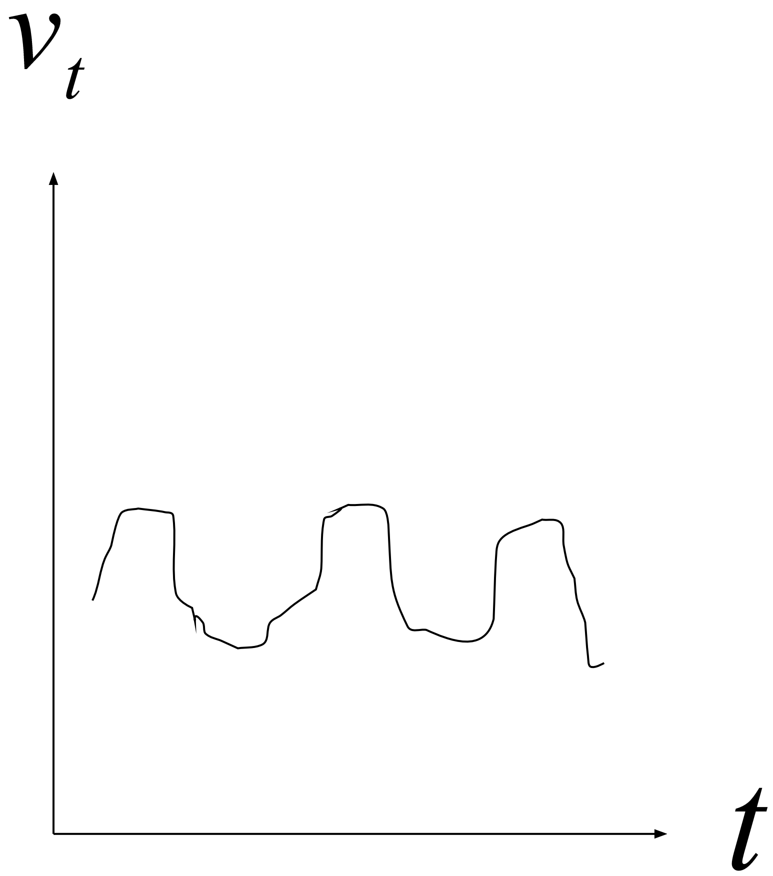

Рис. 3

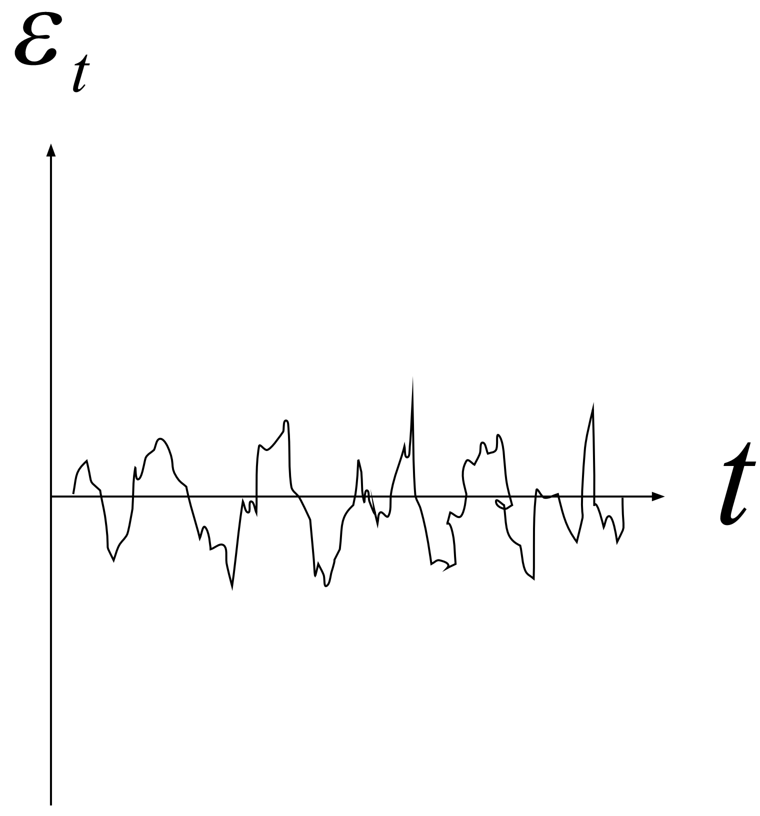

Рис. 4

- ε_t – *случайная* компонента, отражающая влияние не поддающихся регистрации случайных факторов (рис. 4).

2. Стационарные временные ряды.

Для того чтобы задача анализа временных рядов была практически реализуемой, необходимо определенным образом ограничить класс рассматриваемых моделей с точки зрения структуры ряда и его вероятностных характеристик.

Поиск адекватной модели ряда обычно начинают в рамках класса *стационарных временных рядов.*

Ряд y_t называется *строго стационар-ным*, если совместное распределение вероят-ностей m наблюдений

$$y_{t_1}, y_{t_2}, \ldots, y_{t_m}$$

такое же, как и для m наблюдений

$$y_{t_1+\tau}, y_{t_2+\tau}, \ldots, y_{t_m+\tau}$$

для любых $m, t_1, t_2, \ldots, t_m, \tau$.

Ряд называется *слабо стационарным* (стационарным в широком смысле), если для него выполняются следующие соотношения:

1. $M(y_t) = a = const$

2. $D(y_t) = \sigma^2 = const$.

3. $Cov(y_t, y_{t+\tau}) = \gamma(\tau)$.

Другими словами ряд слабо стационарен, если математическое ожидание, дисперсия и автоковариация ряда y_t не зависят от времени t .

Автоковариация $Cov(y_t, y_{t+\tau})$ характе-ризует ковариационную зависимость между различными уровнями одного временного ряда y_t , т.к. при наличии тренда и цикличе-ской компоненты значения последующих уровней ряда зависят от предыдущих значений.

Автоковариация имеет те же недостатки, что и ковариация: с трудом поддаётся непосредственной интерпретации и зависит от единиц измерения .

Отсюда более удобным для практики является *коэффициент автокорреляции:*

$$\rho(\tau) = \frac{M\left((y_t - a)(y_{t+\tau} - a)\right)}{\sigma^2} = \frac{Cov(y_t, y_{t+\tau})}{\sigma^2}.$$

Число периодов , по которым рассчитывается коэффициент автокорреляции, называется *лагом*. Если $\tau = 1$, то имеем коэффициент автокорреляции 1-го порядка, при $\tau = 2$ - коэффициент автокорреляции 2-го порядка и т.д.

С увеличением τ число пар значений, по которым рассчитывается $\rho(\tau)$, уменьшается и для обеспечения статистической достоверности лаг τ не должен превышать четверть объёма выборки ($\tau < n/4$).

Отметим две особенности $\rho(\tau)$.

Во-первых, он изменяется в пределах

$$-1 \le \rho \le 1$$

и характеризует тесноту только *линейной связи* текущего и предыдущих уровней ряда.

Для некоторых временных рядов, имеющих сильную нелинейную тенденцию, коэффициент $\rho \approx 0$.

Во-вторых, по знаку $\rho(\tau)$ нельзя делать вывод о возрастающей или убывающей тенденции уровней ряда.

Бывает так, что $\rho > 0$, но ряд y_t при этом имеет убывающую тенденцию.

Зависимость ρ от величины τ называют *автокорреляционной функцией ряда*, а её график – *коррелограммой* (рис. 5-8).

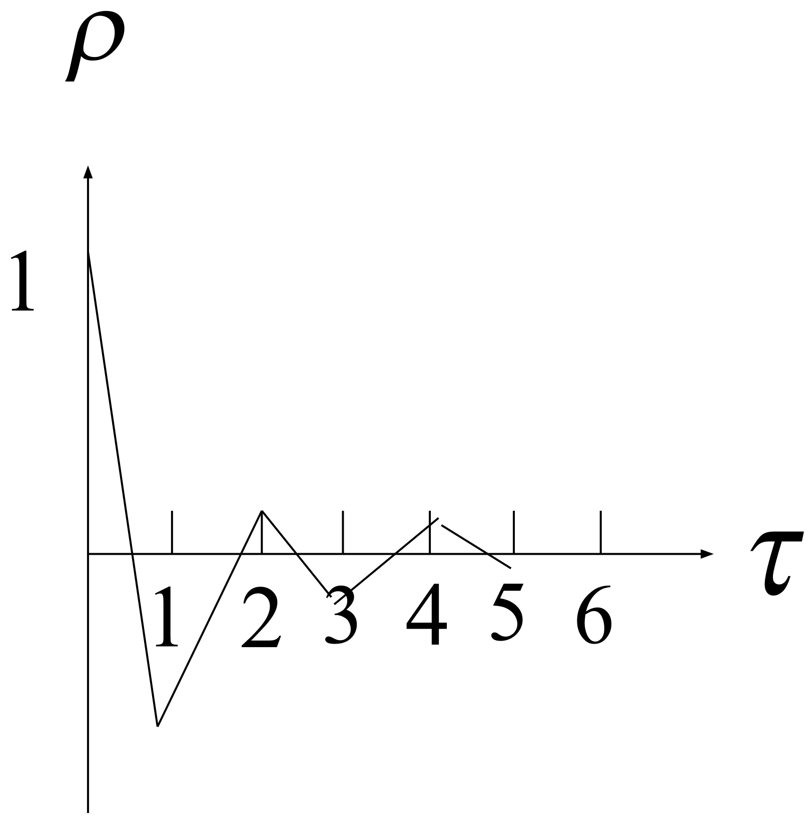

Рис. 5

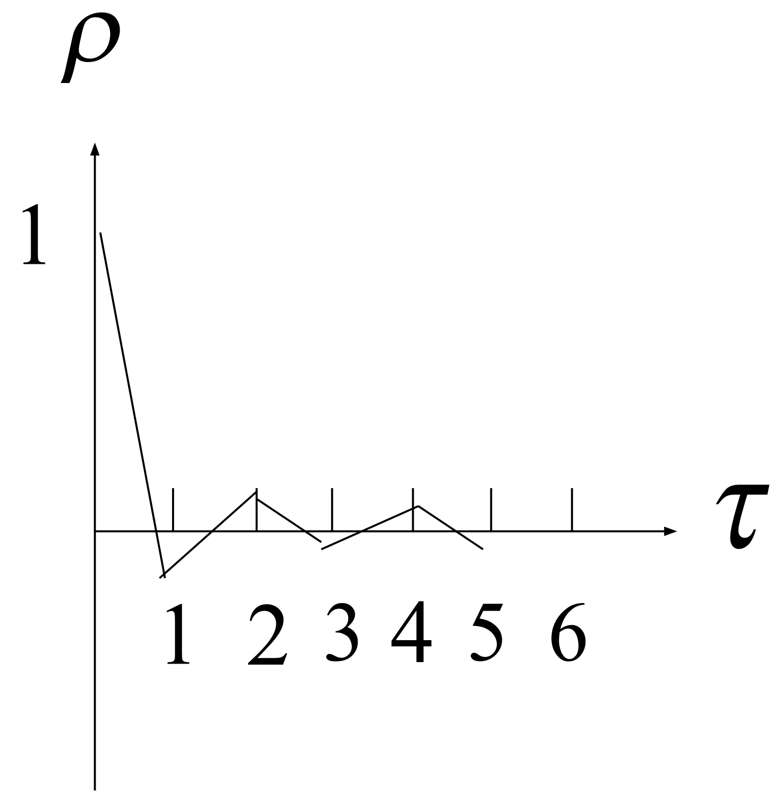

Рис. 6

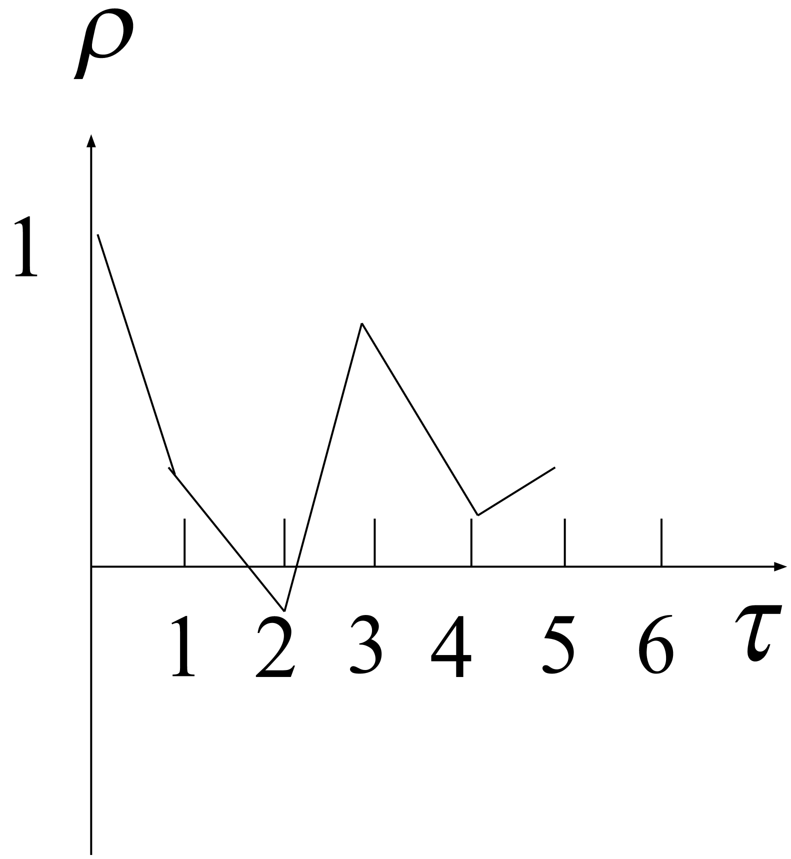

Рис. 7

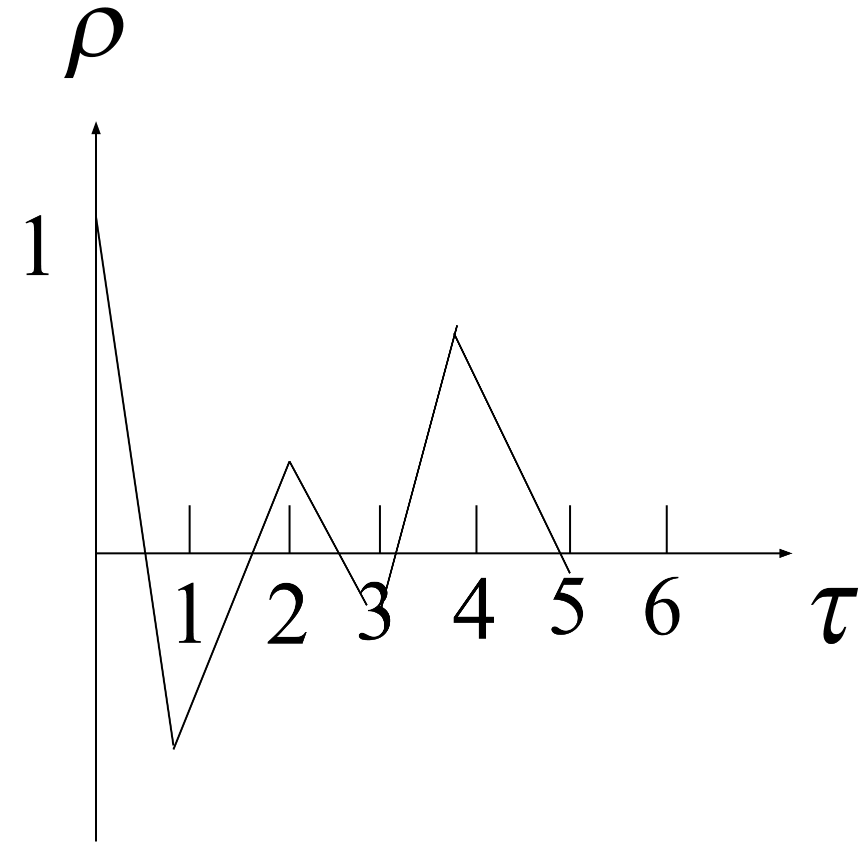

Рис. 8

Анализ автокорреляционной функции и её графика помогает выявить структуру ряда.

Если чередуются затухающие положительные и отрицательные значения

$$\rho(\tau), \tau = 1,2,3...,$$

то это характерно для стационарного ряда.

Если наиболее большим по модулю оказалось значение $\rho(1)$, то исследуемый ряд содержит только *тенденцию* (рис. 5).

Если наиболее высоким оказался коэффициент автокорреляции k – го порядка, то ряд содержит циклические колебания с периодичностью в k моментов времени (рис. 7, $k = 3$, рис. 8, $k = 4$).

Если ни одно из значений не является доминирующим (рис. 6), то либо ряд не содержит тренда и циклической составляющей и имеет только случайную компоненту, либо ряд имеет сильную нелинейную тенденцию.

Статистическими оценками числовых характеристик слабо стационарного временного ряда являются:

- выборочное среднее $\bar{y}_t = \frac{1}{n}\sum_{t=1}^{n} y_t;$

- выборочная дисперсия

$$D_{y_t} = \frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y}_t)^2;$$

● выборочный коэффициент автокорреляции ($\tau = 1,2,3,\ldots$)

$$r(\tau) = \frac{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum\limits_{t=1}^{n-\tau} y_t \cdot \sum\limits_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_t^2 - \left(\sum\limits_{t=1}^{n-\tau} y_t\right)^2} \cdot \sqrt{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_{t+\tau}^2 - \left(\sum\limits_{t=1}^{n-\tau} y_{t+\tau}\right)^2}}.$$

Функцию $r(\tau)$ переменной τ называют *выборочной автокорреляционной функцией*.

Наряду с $r(\tau)$ рассматривают *частные коэффициенты автокорреляции* $r_{част}(\tau)$, которые характеризуют тесноту линейной связи уровней ряда y_t и $y_{t+\tau}$ при устранении влияния уровней $y_{t+1}, y_{t+2}, \ldots, y_{t+\tau-1}$, находящихся между ними. Например, частный коэффициент второго порядка $r_{част}(2)$ оценивает тесноту связи y_t и y_{t+2} при элиминировании уровня y_{t+1}.

Далее находится *выборочная частная автокорреляционная* функция:

$$r_{\text{част}} = r_{част}(\tau), \tau = 2,3,\ldots$$

3. Выравнивание временных рядов.

Если при анализе структуры временного ряда обнаружена только тенденция и отсутствуют циклические колебания, то можно приступить к моделированию тенденции ряда. Если же во временном ряде имеют место и циклические колебания, то, прежде всего, требуется исключить циклическую составляющую и лишь затем приступить к моделированию тенденции.

Для выявления основной тенденции в уровнях ряда, т.е. *выравнивания* ряда, ис-пользуются различные методы:

- механическое (алгоритмическое) выравнивание;
- аналитическое выравнивание.

Из методов первого типа рассмотрим *метод скользящих средних*. Он основан на переходе от исходных значений ряда к их средним значениям на некотором интервале времени, длина которого фиксирована и определена заранее.

Если интервал содержит нечётное число $m = 2p + 1$ уровней ряда, то среднее значение ряда находится по формуле

$$\bar{y}_t = \frac{1}{m} \sum_{i=t-p}^{t+p} y_i, \quad p = 1,2,\ldots \quad .$$

Чаще всего $m = 3,5,7$. В итоге получается сглаженный ряд средних значений $\bar{y}_t$, но число уровней у него будет меньше, чем в исходном ряде. Например, при $m = 3$ в полученном новом ряде теряется два уровня: $\bar{y}_1$ и $\bar{y}_n$. В общем случае число уровней сглаженного ряда уменьшается на $2p$ значений.

Если выбранный интервал содержит чётное число $m = 2p$ уровней ряда, то вначале находятся скользящие средние

$$\bar{y}_{t-0,5} = \frac{1}{m}\sum_{i=t-p}^{t+p-1} y_i, \quad p = 1,2,...$$

для промежуточных уровней ряда, а затем выполняется центрирование полученных скользящих средних

$$\bar{y}_t = \frac{1}{2}(\bar{y}_{t-0,5} + \bar{y}_{t+0,5}), \quad t = p+1, p+2, ..., n-p$$

с целью приведения их к фактическим временным периодам исходного ряда.

Существуют и другие методы механического выравнивания ряда: метод взвешенных скользящих средних, метод экспоненциального сглаживания (метод Брауна), метод последовательных разностей.

Однако в эконометрике основное внимание уделяется **аналитическому** выравниванию ряда. Данный метод заключается в построении аналитической функции, характеризующей зависимость уровней ряда от времени, т.е. в построении парной регрессии

$$\widetilde{y}_t = f(t).$$

Для этого можно использовать различные виды функций:

- линейный тренд $\tilde{y}_t = b_0 + b_1 t$;
- гиперболический тренд $\tilde{y}_t = b_0 + b_1 \dfrac{1}{t}$;
- степенной тренд $\tilde{y}_t = b_0 t^{b_1}$

и т.д.

Параметры каждого из перечисленных трендов можно определять обычным МНК, используя в качестве независимой переменной время t, а в качестве зависимой – уровни ряда y_t.

Особенность заключается в том, что независимая переменная принимает целочисленные значения ($t = 1,2,3,\ldots$), что даже облегчает вычисления.

Для нелинейных трендов предварите-льно проводят стандартную процедуру их линеаризации.

4. Моделирование ряда при наличии циклических колебаний.

Существует несколько подходов при моделировании рядов с циклическими колебаниями. Для определенности пусть они представляют *сезонные* изменения.

Наиболее простым методом является расчёт значений сезонной компоненты и построение аддитивной или мультипликативной модели ряда.

Если амплитуда сезонных колебаний со временем не меняется, то применяют аддитивную модель $y_t = u_t + v_t + \varepsilon_t$. В противном случае используют мультипликативную модель $y_t = u_t \cdot v_t \cdot \varepsilon_t$. Построение обеих моделей сводится к расчёту значений u_t, v_t, ε_t для каждого уровня.

Сезонные компоненты при этом должны удовлетворять следующим требованиям:

● в случае аддитивной модели сумма всех сезонных компонент за год должна быть равна *нулю*;

● для мультипликативной модели произведение всех сезонных компонент должна равняться *единице*.

Процесс построение аддитивной модели включает в себя следующие шаги.

1. Выравнивание временного ряда методом скользящей средней.

В итоге получается выровненный ряд , который не содержит сезонной

2. *Расчет значений сезонной компоненты*.

Оценки сезонной компоненты находятся как разность между фактическими уровнями ряда и скользящими средними

$$v_t = y_t - \bar{y}_t.$$

Далее вычисляются средние значения за каждый сезон оценки сезонной компоненты по всем годам, по которым имеются данные:

$$\bar{v}_{ti} = \frac{1}{m}\sum v_t.$$

В аддитивной модели сумма значений сезонной компоненты по всем сезонам должна быть равна нулю. Если это не выполняется, т.е.

$$\sum_{i=1}^{p} \overline{v}_{ti} = d \neq 0,$$

где p –число сезонов в году, то вычисляется корректирующий коэффициент:

$$k = \frac{d}{p}.$$

Затем рассчитываются скорректированные значения сезонной компоненты как разность между ее средней оценкой и корректирующим коэффициентом k :

$$v_i = \bar{v}_{ti} - k, \quad i = \overline{1, p}.$$

При этом должно выполняться равенство:

$$\sum_{i=1}^{p} v_i = 0.$$

3. Устранение сезонной компоненты из исходных уровней ряда.

Из каждого уровня исходного ряда y_t вычитается скорректированное значение сезонной компоненты v_i, в результате получается ряд, содержащий только тенденцию и случайную компоненту:

$$u_t + \varepsilon_t = y_t - v_i.$$

4. Аналитическое выравнивание уровней $u_t + \varepsilon_t$.

Поскольку эти данные не содержат цик-лической компоненты можно выполнить мо-делирование тенденции ряда. Форму тренда выявляют либо визуально по полю корреля-ции, либо другими известными

5. Расчет суммы значений трендовой и сезонной компонент $\tilde{u}_t + v_i$.

К значениям выровненных уровней ряда $\tilde{u}_t$ прибавляются значения скорректированной сезонной компоненты v_i для соответствующих сезонов.

6. Расчет ошибок.

Расчет абсолютной ошибки производится по формуле:

$$e_t = y_t - (\tilde{u}_t + v_i).$$

Сумму квадратов полученных абсолютных ошибок $\sum e_t^2$ относят к общей сумме квадратов отклонений уровней ряда и по значению

$$R^2 = 1 - \frac{\sum e_t^2}{\sum (y_t - \bar{y}_t)^2} = 1 - \frac{\sum e_t^2}{n \cdot \sigma_{y_t}^2}$$

делают вывод о качестве модели.