Моделирование одномерных временных рядов и прогнозирование

Составляющие временного ряда

Временным рядом (рядом динамики, динамическим рядом) называется упорядоченная во времени последовательность численных показателей $\{(y_i,t_i), i=1,2,...,n\}$, характеризующих уровни развития изучаемого явления в последовательные моменты или периоды времени

Таблица 1

Динамика ВВП Российской Федерации

	2000 г.	2001 г.	2002 г.	2003 г.	2004 г.
ВВП, млрд руб.	7305,6	8943,6	10834,2	13285,2	17048,1

Величины y_i называются уровнями ряда, а t_i – временными метками (моменты или интервалы наблюдения). Обычно рассматриваются временные ряды с равными интервалами между наблюдениями, в качестве значений t_i берутся порядковые номера наблюдений и временной ряд представляется в виде последовательности $y_1, y_2, ..., y_n$, где n – количество наблюдений.

- Целью исследования временного ряда является выявление закономерностей в изменении уровней ряда и построении его модели в целях прогнозирования и исследования взаимосвязей между явлениями.

- При исследовании экономического временного ряда его обычно представляют в виде совокупности трех составляющих:

– долговременной тенденции (T), т. е. устойчивого увеличения или уменьшения значений уровней ряда (тренда);

– периодических колебаний (S);

– случайных колебаний (E).

- Различным образом объединяя эти компоненты, можно получить различные модели временного ряда (Y):

– аддитивную

$$Y_t = T_t + S_t + E_t;$$

– мультипликативную

$$Y_t = T_t \cdot S_t \cdot E_t;$$

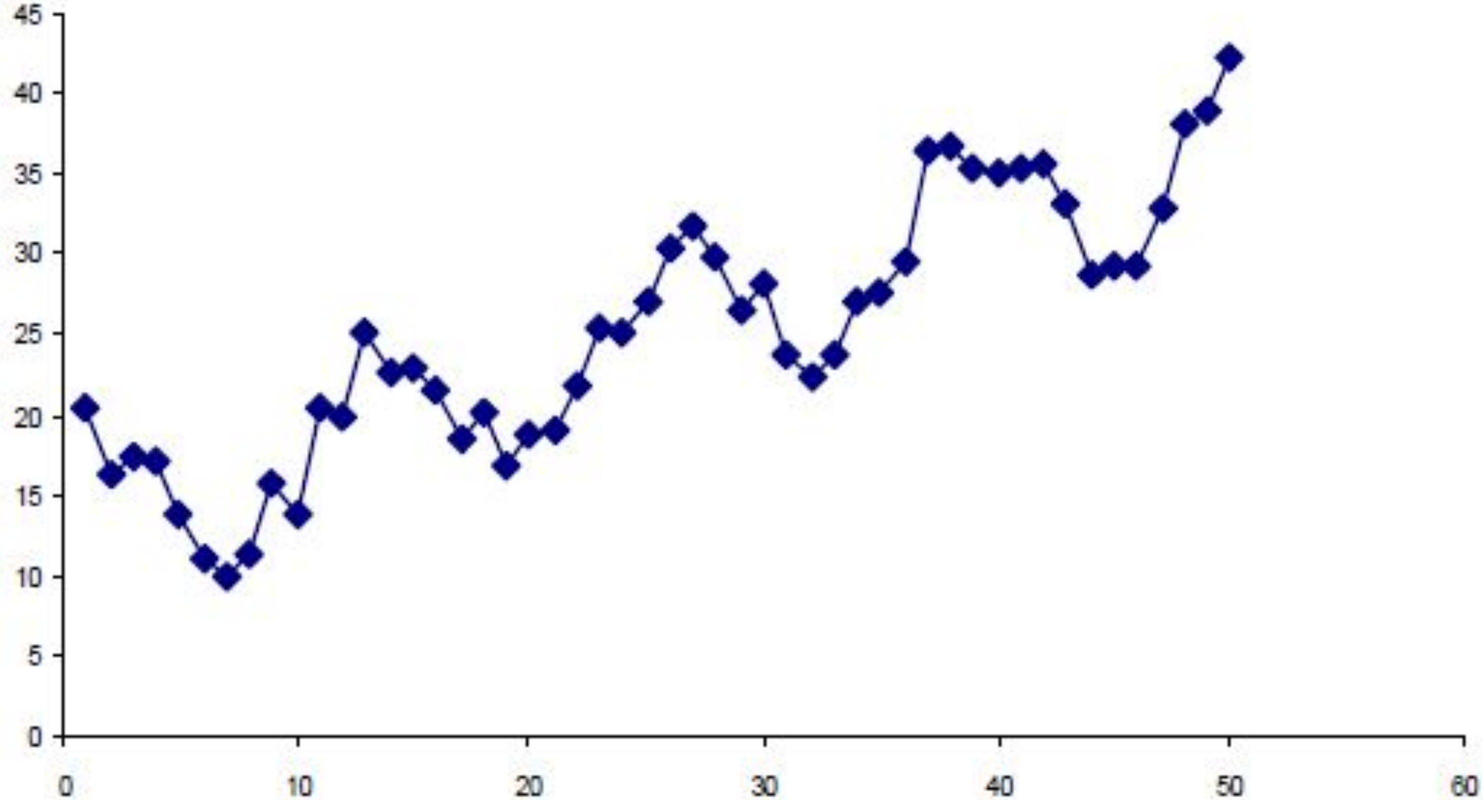
– смешанную

$$Y_t = T_t \cdot S_t + E_t.$$

- В экономике периодические колебания принято подразделять на *сезонные*, у которых период колебаний не превышает одного года (цены на сельскохозяйственную продукцию), вызванные климатическими или социально-экономическими причинами, и *циклические* с периодом колебаний несколько лет, связанные с циклами деловой активности.

- Основная задача эконометрического исследования временного ряда заключается в выявлении и придании количественного выражения составляющим его отдельным компонентам.

- Наличие той или иной составляющей можно определить с помощью визуального анализа графика временного ряда.


45
40
35
30
25
20
15
10
5
0
0
10
20
30
40
50
60

Перед построением модели исходные данные проверяются на:

- *сопоставимость* (применение одинаковой методики получения или расчета данных),
- *однородность* (отсутствие случайных выбросов), устойчивость (наличие закономерности в изменении уровней ряда),
- *достаточность* (число наблюдений должно в 7–10 превосходить число параметров модели).

Автокорреляция уровней временного ряда

- Важной особенностью временных рядов по сравнению с данными наблюдений, относящихся к одному периоду времени, является, как правило, наличие связи между последовательными уровнями ряда, вызванное действием каких-либо долговременных причин, что приводит к наличию таких составляющих ряда, как долговременная тенденция и периодическая составляющая.

- Корреляционная зависимость между последовательными уровнями временного ряда называется *автокорреляцией уровней* временного ряда.

- Степень тесноты автокорреляционной связи между уровнями ряда может быть определена с помощью *коэффициентов автокорреляции*, т. е. коэффициентов линейной корреляции между уровнями исходного временного ряда и уровнями ряда, сдвинутыми на несколько шагов назад во времени.

$$r_\tau = \frac{\sum_{t=\tau+1}^{n}(y_t - \overline{y}_{1\tau})\cdot(y_{t-\tau} - \overline{y}_{2\tau})}{\sqrt{\sum_{t=\tau+1}^{n}(y_t - \overline{y}_{1\tau})^2 \cdot \sum_{t=\tau+1}^{n}(y_{t-\tau} - \overline{y}_{2\tau})^2}},$$

где τ – величина сдвига, называемая *лагом*, определяет порядок коэффициента автокорреляции,

$$\overline{y}_{1\tau} = \frac{\sum_{t=\tau+1}^{n} y_t}{n-\tau}; \quad \overline{y}_{2\tau} = \frac{\sum_{t=\tau+1}^{n} y_{t-\tau}}{n-\tau}.$$

Функцию $r(\tau) = r_\tau$ называют *автокорреляционной функцией* временного ряда, а ее график – *коррелограммой.*

- Анализ автокорреляционной функции и коррелограммы позволяет выявить структуру ряда, т. е. определить присутствие в ряде той или иной компоненты.

- Если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию.
- Если наиболее высоким оказался коэффициент автокорреляции порядка m, то ряд содержит циклические колебания с периодичностью в m моментов времени.
- Если ни один из коэффициентов автокорреляции не является значимым, то ряд не содержит тенденции и циклических колебаний

- Линейные коэффициенты автокорреляции характеризуют тесноту только линейной связи текущего и предыдущих уровней ряда.
- По коэффициентам автокорреляции можно судить только о наличии или отсутствии линейной (или близкой к линейной) зависимости.
- Для проверки ряда на наличие нелинейной тенденции рекомендуется вычислить линейные коэффициенты автокорреляции для временного ряда, состоящего из логарифмов исходных уровней. Отличные от нуля значения коэффициентов автокорреляции будут свидетельствовать о наличии нелинейной тенденции.

Моделирование тенденции временного ряда

- Моделирование тенденции временного ряда является важнейшей классической задачей анализа экономических временных рядов.
- Решение этой задачи начинается с проверки наличия тенденции и формулирования предложений о характере долговременной тенденции, после чего уже строится модель тенденции как функции времени.

Методы определения наличия тенденции

- Для диагностирования наличия тенденции наиболее широко применяются *метод сравнения средних* и *метод Фостера-Стюарта*.

Метод сравнения средних

- Метод сравнения средних применим для выявления монотонной тенденции.

Временной ряд разбивается на две примерно равные части $y_1, y_2, \ldots, y_{n_1}$ и $y_{n_1+1}, y_{n_1+2}, \ldots, y_{n=n_1+n_2}$ с количеством уровней n_1 и n_2 и для каждой части вычисляются средние $(\bar{y}_1, \bar{y}_2)$ и выборочные дисперсии (s_1^2, s_2^2) соответственно.

Далее рассчитывается значение критерия Стьюдента по формуле

$$\tau = \frac{\left|\bar{y}_1 - \bar{y}_2\right|}{\sqrt{\frac{s_1^2}{n_1}} + \sqrt{\frac{s_2^2}{n_2}}},$$

если предполагается, что значения дисперсий на этих участках не равны между собой, т. е. $\sigma^2{}_1 \neq \sigma^2{}_2$, и по формуле

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{s^2}\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}},$$

где s^2 – общая выборочная дисперсия ряда, если предполагается, что дисперсии одинаковы $\sigma^2{}_1 = \sigma^2{}_2 = \sigma^2$.

Нулевая гипотеза о равенстве средних (об отсутствии тенденции) отвергается, если выполняется условие

$$\tau > t_{1-\alpha,\, m},$$

где $t_{1-\alpha,\, m}$ – табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $m = n_1 + n_2 - 2$.

Метод Фостера-Стюарта

- Является более универсальным и дает более надежные результаты.

Каждому уровню ряда y_i, начиная со второго, ставится в соответствие два значения p_i q_i по следующим правилам:

$p_i = 1$, если уровень y_i меньше всех предыдущих уровней, т. е.

$$y_i < y_1, y_2, \ldots, y_{i-1},$$

и $p_i = 0$ в противном случае;

$q_i = 1$, если уровень y_i больше всех предыдущих уровней, т. е.

$$y_i > y_1, y_2, \ldots, y_{i-1},$$

и $q_i = 0$ в противном случае.

Вычисляется статистика

$$t_p = \frac{\sum_{i=2}^{n}(p_i - q_i)}{2\sum_{i=2}^{n}\frac{1}{i}} .$$

Гипотеза об отсутствии тенденции отвергается, если выполняется условие

$$t_{\text{p}} > t_{1-\alpha,\, n-1},$$

где $t_{1-\alpha,\, n-1}$ – табличное значение t-критерия Стьюдента при уровне значимости

α и числе степеней свободы $n - 1$.

Сглаживание временного ряда по методу скользящей средней

- Цель сглаживания временного ряда заключается в получении ряда с меньшим разбросом уровней, что в ряде случаев позволяет на основе визуального анализа сделать вывод о наличии тенденции и ее характерных особенностях.

Сглаживание временного ряда по методу скользящей средней заключается в замене исходных уровней ряда y_t сглаженными значениями y'_t, которые получаются как среднее значение определенного числа уровней исходного ряда, симметрично окружающих значение y_t.

В результате получается временной ряд

y'_t, меньше подверженный колебаниям. Если индивидуальный разброс значений временного ряда около своего среднего значения a характеризуется дисперсией σ^2, то средняя из m членов ряда $(y_1 + y_2 + \ldots + y_m)/m$ будет иметь в m раз меньшую дисперсию (σ^2/m).

Для вычисления сглаженных значений y'_t по методу простой скользящей средней используются следующие формулы:

1) Нечетный интервал сглаживания

$g = 2p+1$ (интервал сглаживания – количество исходных уровней ряда (y_t), используемых для сглаживания):

$$y'_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p-1} + y_{t+p}}{2p+1},$$

где y_t – фактическое значение уровня исходного ряда в момент t; y'_t – значение скользящей средней в момент t; $2p+1$- длина интервала сглаживания.

Формула при интервалах сглаживания $g = 3$ и $g = 5$ принимает вид

$$y'_t = \frac{y_{t-1} + y_t + y_{t+1}}{3};$$

$$y'_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}.$$

- 2) *Четный интервал сглаживания*

g = 2*p*:

$$y'_t = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \ldots + y_{t-1} + y_t + y_{t+1} + \ldots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} =$$

$$= \frac{\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}.$$

Формула при интервалах сглаживания $g = 2$ и $g = 4$ принимает вид

$$y'_t = \frac{\frac{1}{2} y_{t-1} + y_t + \frac{1}{2} y_{t+1}}{2};$$

$$y'_t = \frac{\frac{1}{2} y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2} y_{t+2}}{4}.$$

- При использовании скользящей средней с длиной активного участка

$g = 2p+1$ первые и последние p уровней ряда сгладить нельзя, их значения теряются. Очевидно, что потеря значений последних точек является существенным недостатком, т. к. для исследователя последние «свежие» данные обладают наибольшей информационной ценностью.

- Для восстановления потерянных значений временного ряда можно использовать следующий прием:

а) Вычисляется средний прирост Δ_y на последнем активном участке $(y_{n-g}, \ldots, y_n)$

$$\Delta_y = \frac{y_n - y_{n-g}}{g-1},$$ где g – длина активного участка.

б) Определяются значения последних $p = (g-1)/2$ уровней сглаженного временного ряда с помощью последовательного прибавления среднего абсолютного прироста Δ_y к последнему сглаженному значению y'_{n-p}

$$y'_{n-p+1} = y'_{n-p} + \Delta_y,\ y'_{n-p+2} = y'_{n-p+1} + \Delta_y,\ \ldots,\ y'_n = y'_{n-1} + \Delta_y.$$

- Аналогичная процедура применяется для восстановления первых *p* уровней временного ряда.
- Важным свойством процедуры сглаживания является *полное устранение* периодических колебаний из временного ряда, если длина интервала сглаживания берется равной или кратной периоду колебаний.

Метод аналитического выравнивания

Аналитическим выравниванием временного ряда называют нахождение аналитической функции $\hat{y} = f(t)$, характеризующей основную тенденцию изменения уровней ряда с течением времени. Сама функция $f(t)$ носит название *кривой роста*.

При аналитическом выравнивании (нахождении аналитической функции $\hat{y} = f(t)$) исходят из предположения, что аддитивная модель временного ряда может быть представлена как сумма двух компонент

$$y(t) = f(t) + \varepsilon_t,$$

где ε_t – случайная компонента с нулевой средней и постоянной дисперсией выражает ошибку модели из-за действия случайных факторов.

Чаще всего в качестве кривой роста применяются следующие функции

– линейная $y_t = a_0 + a_1 t$;

– парабола второго и более высоких порядков

$$y_t = a_0 + a_1 t^1 + a_2 t^2 + \ldots + a_k t^k;$$

– гиперболическая $y_t = a_0 + a_1 / t$;

– экспонента $y_t = e^{a_0 + a_1 t}$;

– потенциальная $y_t = a_0 \times a_1{}^t$;

– модифицированная экспонента $y_t = K + a_0 \times a_1{}^t$;

– степенная $y_t = a_0 \times t^{a_1}$;

– логистическая кривая $y_t = \dfrac{K}{1 + a_0 e^{-a_1 t}}$;

– кривая Гомперца $y_t = K \times a_0{}^{a_1^t}$.

- Построение таких функций ничем не отличается от построения уравнений парной регрессии (линейной или нелинейной) с учетом того, что в качестве зависимой переменной используются фактические уровни временного ряда y_t, а в качестве независимой переменной моменты времени $t = 1,2, ..., n$.
- Для построения кривой роста необходимо выбрать вид аналитической зависимости и затем оценить значения ее параметров.

- Для определения вида тенденции (аналитической зависимости) применяются такие методы, как

– качественный анализ изучаемого процесса;

– построение и визуальный анализ графика зависимости уровней ряда от времени;

– расчет и анализ показателей динамики временного ряда (абсолютные приросты, темпы роста и др.);

– анализ автокорреляционной функции исходного и преобразованного временного ряда;

– метод перебора, при котором строятся кривые роста различного вида со следующим выбором наилучшей на основании значения скорректированного коэффициента детерминации $\overline{R}^2$

Выбор вида тенденции

- **Выбор вида тенденции на основе качественного анализа**
- Социально-экономические процессы в зависимости от характера их протекания можно разделить на три класса:

- 1) *Процессы с монотонным характером развития и отсутствием пределов роста*
- Эти условия справедливы для поведения многих экономических показателей, например, для большинства натуральных показателей промышленного производства. В этом случае для моделирования тенденции могут использоваться: *линейная, параболическая, экспоненциальная, степенная* функции.

2) *Процессы, которые имеют предел роста (падения) в исследуемом периоде, так называемые процессы с «насыщением».*

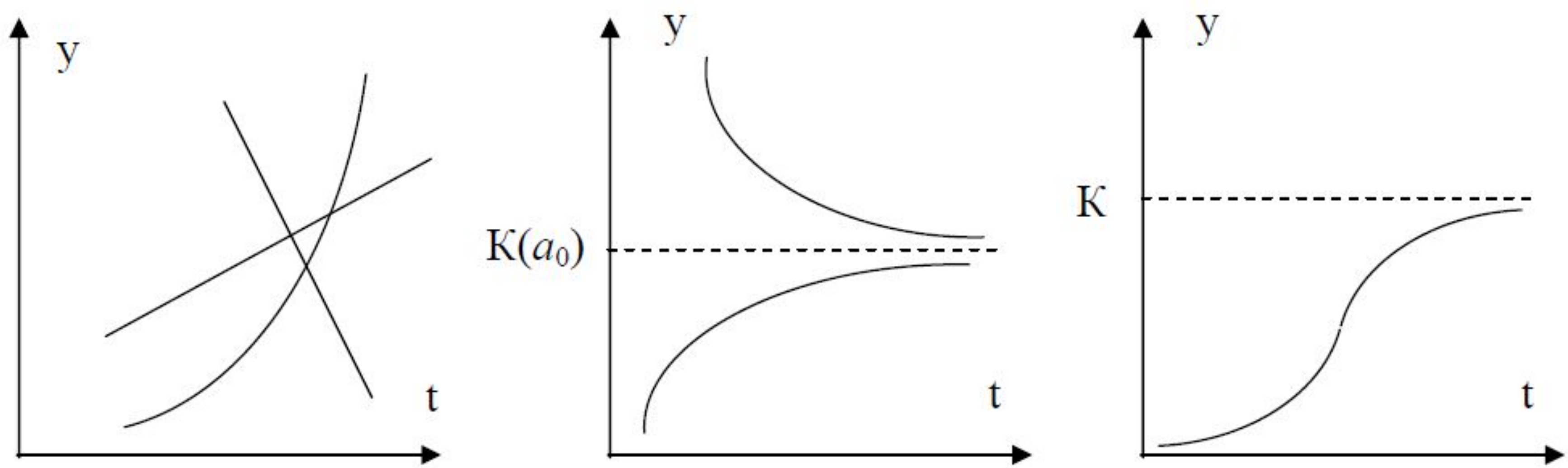
Развитие процесса происходит под влиянием некоторых ограничивающих факторов, величина воздействия которых растет вместе с ростом достигнутого уровня. С такими процессами часто сталкиваются в демографии, при изучении потребностей в товарах и услугах (в расчете на душу населения), при исследовании эффективности использования ресурсов и т. д.

Примерами показателей, для которых могут быть указаны пределы роста, являются среднедушевое потребление определенных продуктов питания, расход удобрений на единицу площади и т. п.

В этом случае для моделирования тенденции используются гиперболическая функция.

- 3) *Так называемые S-образные процессы, представляющие как бы два последовательных лавинообразных процесса* (когда прирост зависит от уже достигнутого уровня): один с ускорением развития, а другой – с замедлением. С такими процессами часто сталкиваются в демографических исследованиях, в страховых расчетах, при решении задач прогнозирования научно-технического прогресса, при определении спроса на новый вид продукции.

- К *S*-образным процессам можно отнести процесс развитие новой отрасли (нового производства). Вначале производство развивается очень медленно вследствие того, что технические методы производства еще недостаточно разработаны, издержки производства высоки и спрос на рынке на данный товар еще очень мал, поэтому производство развивается медленно.
- В дальнейшем, благодаря усовершенствованию технических методов изготовления, переходу к массовому производству и увеличению емкости рынка для данного товара производство растет быстрее.
- Затем наступает период насыщения рынка, рост производства все более замедляется, и, наконец, почти прекращается. Наступает стабилизация производства на определенном уровне.



а) I класс *б*) II класс *в*) III класс

Рис. Схемы протекания процессов

Оценка адекватности и точности модели тенденции

- После построения модели тенденции осуществляется проверка ее качества по характеристикам адекватности (соответствия данным наблюдения) и точности.

- **Проверка адекватности** модели основывается на анализе ряда остатков

$$e_t = y_t - \hat{y}_t. \qquad (i = 1, 2, \ldots, n)$$

Модель считается адекватной, если остатки:

- являются случайными;
- распределены по нормальному закону;
- имеют равное нулю среднее значение $e = 0$;
- независимы между собой.

Моделирование периодических колебаний

- **Выделение периодической компоненты по методу скользящей средней**
- Простейшим приемом выделения периодической компоненты основан на использовании сглаживания временного ряда по методу простой скользящей средней.
- Предварительно следует определиться с видом модели временного ряда – аддитивной или мультипликативной.
- Это можно сделать на основе анализа графика временного ряда.

- Если амплитуда периодических колебаний примерно постоянна, то следует выбрать аддитивную модель:

$$Y = T + S + E$$

- Здесь амплитуда колебаний периодической компоненты предполагается постоянной, не зависящей от времени.
- Если амплитуда периодических колебаний возрастает с ростом уровней ряда, то следует выбрать мультипликативную модель временного ряда:

$$Y = T \cdot S \cdot E$$

- Выделение периодической компоненты основывается на том, что если исходный временной ряд содержит периодическую компоненту с периодом g, то сглаженный по методу простой скользящей средней временной с интервалом сглаживания g такой компоненты уже не содержит.

- В случае аддитивной модели периодическая компонента выделяется путем нахождения разности между соответствующими уровнями исходного и сглаженного ряда.
- В случае мультипликативной модели периодическая компонента выделяется путем нахождения отношения между соответствующими уровнями исходного и сглаженного ряда.
- Затем вычисляются средние значения, соответствующие наблюдениям внутри одного периода колебаний.

- **Моделирование сезонных колебаний с помощью фиктивных переменных**
- Рассмотрим метод моделирования временного ряда, содержащего сезонные колебания, основанный на включении в модель фиктивных переменных.
- Количество фиктивных переменных принимается равным числу наблюдений в пределах одного цикла колебаний без единицы.

- Например, при моделировании поквартальных данных необходимо ввести три дополнительные переменные:

$$z_1 = \begin{cases} 1, & \textit{весна}, \\ 0, & \textit{не весна}, \end{cases} \quad z_2 = \begin{cases} 1, & \textit{лето}, \\ 0, & \textit{не лето}, \end{cases} \quad z_3 = \begin{cases} 1, & \textit{осень}, \\ 0, & \textit{не осень}. \end{cases}$$

- Зиме в этом случае соответствуют нулевые значения всех фиктивных переменных.

- Уравнение регрессии с учетом фиктивных переменных принимает вид:

$$y = a + b \cdot t + c_1 \cdot z_1 + c_2 \cdot z_2 + c_3 \cdot z_3 + \varepsilon \, .$$

- Коэффициенты c_i характеризуют отклонение уровней первых трех сезонов по отношению к последнему.
- Модель с фиктивными переменными может рассматриваться как частный случай аддитивной модели временного ряда.