Компонентный анализ

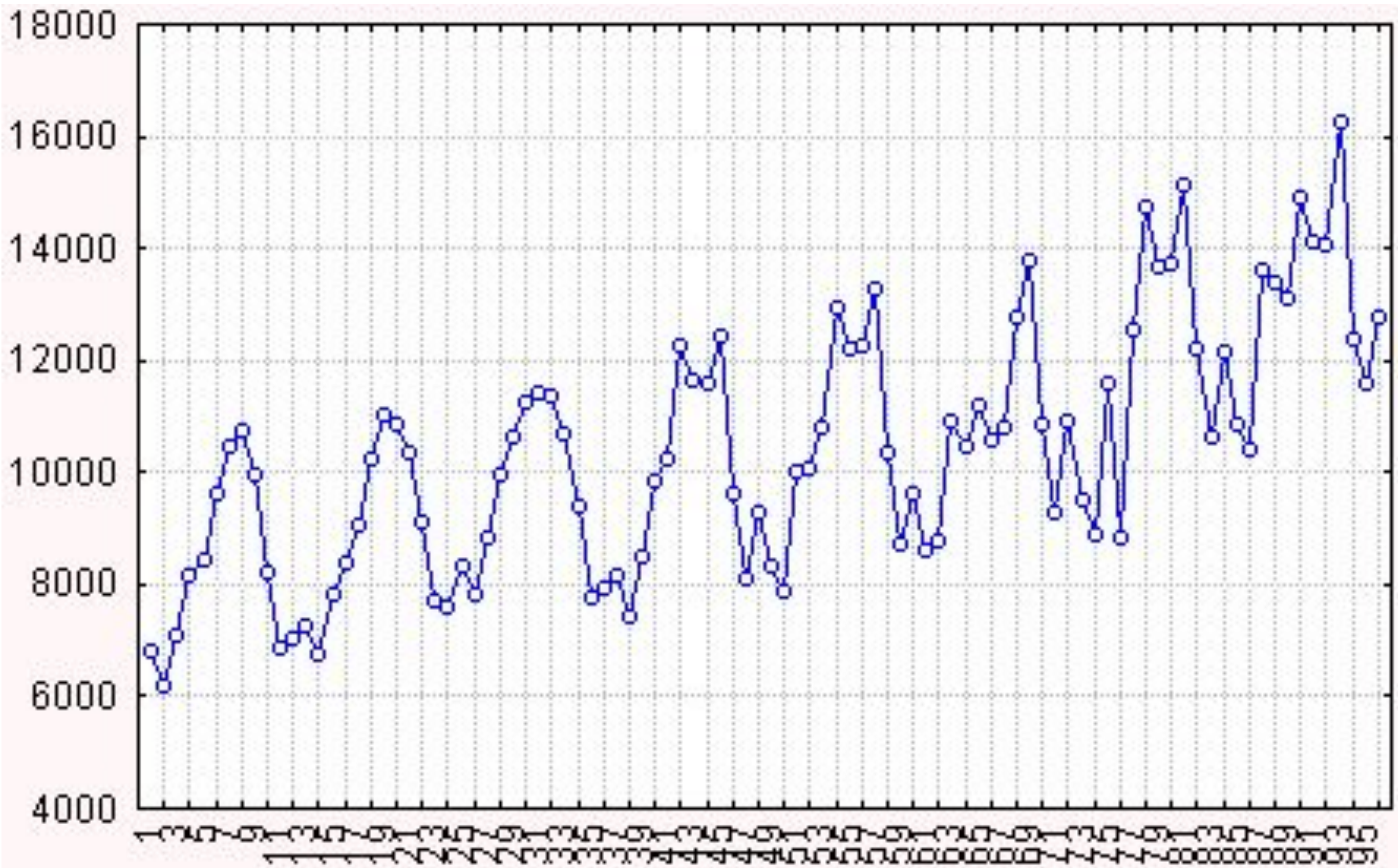


График среднемесячных объемов строительства жилых домов за 8 лет

Можно увидеть ряд компонент:

Тренд, сезонность, цикличность, стационарный случайный процесс, абсолютно случайная составляющая – белый шум.

Тренд

Пусть временной ряд можно представить в виде: $x_t = \xi_t + \varepsilon_t$

Величина ε_t - абсолютно случайная составляющая, называемая белым шумом и удовлетворяющая требованиям:

1) $M(\varepsilon_t)=0$;

2) $D(\varepsilon_t)=\sigma^2$;

3) $cov(\varepsilon_t,\varepsilon_{t+k})=0$ при любых t и $k\neq 0$.

Величина ξ_t может быть генерирована, либо детерминированной функцией, либо случайным процессом, либо их комбинацией.

Опр. Под *детерминированным трендом* понимают долговременную тенденцию развития, имеющую вид непериодической неслучайной, медленно изменяющейся функции времени. $\xi_t=f(t)=a_0+a_1t+a_2t^2$

Случайный (стохастический тренд) $\xi_t=\xi_{t-1}$

Комбинация трендов $\xi_t=\xi_{t-1}+a_0+a_1t+a_2t^2$

Детерминированный тренд может быть:

Линейный

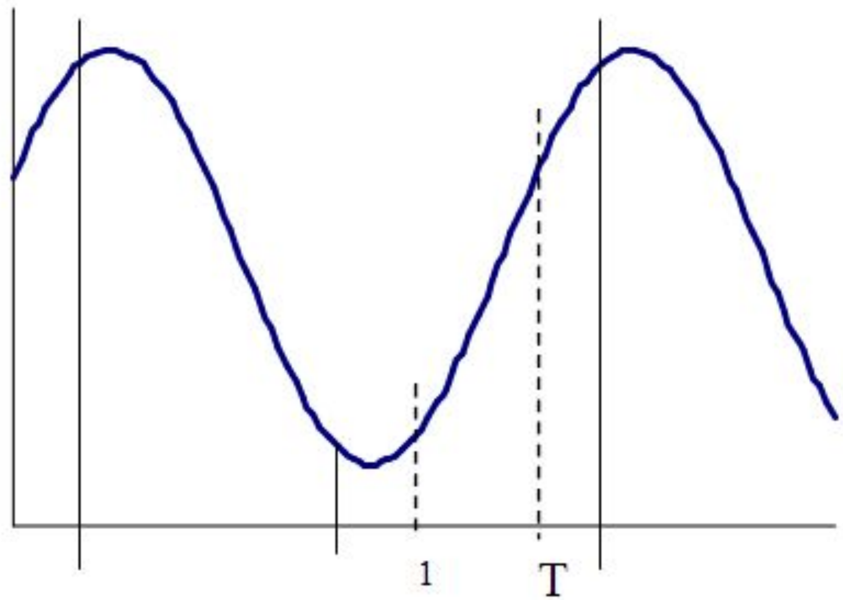
Кусочно-линейный

Нелинейный

Примеры трендов

Исследовании величины осадков:

в течение ста лет медленное увеличение в течение всего периода может быть воспринято как тренд. С другой стороны, если мы обладаем данными за 2000 лет, то рост количества осадков может оказаться частью некоторого медленного колебательного процесса. Поэтому, исходя из тренда в течение столетия, сделать вывод, что климат становится все более влажным, было бы ошибкой



Примеры трендов

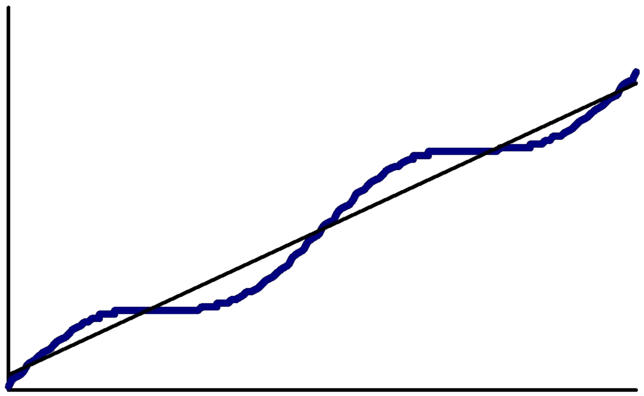
- Кусочно-линейный тренд

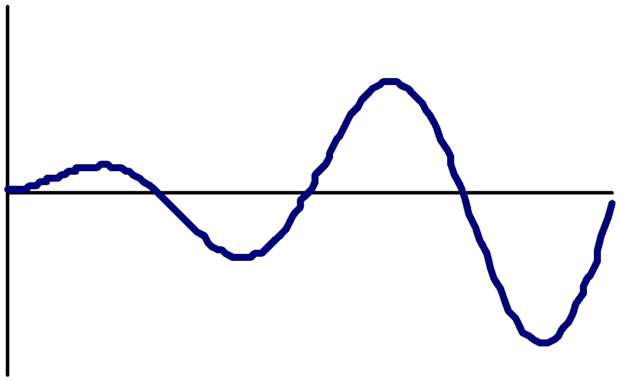
Сезонность -цикличность

<u>Опр.</u> *Сезонная компонента* s_t (эффект сезонности) носит характер периодической неслучайной функции.

Термин «сезонность» относят к эффектам, которые происходят с периодом в один год, но те же идеи могут быть использованы и при изучении любых других явлений, обусловленных строго периодическими процессами (циклическая компонента).

Аддитивная модель $x_t = m_t + s_t + \varepsilon_t$ **Мультипликативная модель** $x_t = m_t \cdot s_t \cdot \varepsilon_t$





Переход от мультипликативной к аддитивной модели:

логарифмируют ряд: $\ln x_t = \ln(m_t \cdot s_t \cdot \varepsilon_t) = \ln m_t + \ln s_t + \ln \varepsilon_t$

и переходим к рассмотрению:

$$x'_t = m'_t + s'_t + \varepsilon'_t$$

Определение наличия тренда и сезонной компоненты

Определения наличия тренда на основе:

- Визуального анализа графика исходных данных;
- Теоретических предпосылках о наличии тренда;
- Спектрального анализа;
- Анализа коррелограмм АКФ и ЧАКФ;
- Теста Форестера-Стюарта.

Определения наличия периодической компоненты на основе:

- Визуального анализа графика исходных данных;
- Теоретических предпосылках о наличии сезонности (цикличности);
- Спектрального анализа;
- Анализа коррелограмм АКФ и ЧАКФ;

Что такое спектральный анализ?

Ряд Фурье выглядит следующим образом:

$$F(x)=\sum_{j=1}^{\infty}a_j\cos(j\cdot x)+\frac{1}{2}b_0+\sum_{j=1}^{\infty}b_j\sin(j\cdot x)$$

Длина периода членов ряда, следующая: первый синусоидальный член имеет период 2π, второй имеет период 2π/2=π, третий - 2π/3 и т.д. Аналогично и для косинусоидальных членов. Вся сумма имеет период 2π.

Любую конечную последовательность $x_1, x_2, \ldots x_N$, не обязательно периодическую, можно представить в виде суммы:

$$x_t=b_0+\sum_{k=1}^{[N/2]}a_k\cos\frac{2\pi k}{N}t+\sum_{k=1}^{[N/2]}b_k\sin\frac{2\pi k}{N}t.$$

Коэффициенты $b_0, b_1, \ldots\ b_{[N/2]}$ и $a_1, a_2, \ldots\ a_{[N/2]}$ можно оценить через значение x_t и тригонометрические функции.

Если же ряд $x_1, x_2, \ldots x_N$ является периодическим, с периодом T, причем N кратно T, значения x_t достаточно представить на множестве $1, 2, \ldots\ T$, поскольку $x_{t+T}=x_t$ (в силу периодичности последовательности). Разложение Фурье для такой последовательности будет иметь вид:

$$x_t=b_0+\sum_{k=1}^{[T/2]}a_k\cos\frac{2\pi k}{T}t+\sum_{k=1}^{[T/2]}b_k\sin\frac{2\pi k}{T}t,$$

$$a_k=\frac{2}{T}\sum_{t=1}^{T}x_t\cdot\cos\frac{2\pi k}{T}\cdot t,$$

$$b_k=\frac{2}{T}\sum_{t=1}^{T}x_t\cdot\sin\frac{2\pi k}{T}\cdot t,$$

$$b_0=\frac{\sum_{t=1}^{T}x_t}{T}.$$

Периодограмма и спектрограмма

Опр. График, на котором по оси ординат отложено r_k^2, а по оси абсцисс – период , называется *периодограммой*.(здесь $r_k^2 = a_k^2 + b_k^2$)

Опр. График, на котором по оси ординат отложено r_k^2, а по оси абсцисс – частота , называется *спектрограммой* (спектром).

Спектр является необходимым инструментом при определении периода колебаний периодической компоненты.

Рассмотрим одну тригонометрическую пару в разложении Фурье.

Коэффициенты a_k и b_k определяются из условия:

$$\sum_{t=1}^{T}\left[x_t - \left(a_k \cdot \cos\frac{2\pi k}{T}t + b_k \cdot \sin\frac{2\pi k}{T}t\right)\right]^2 \Rightarrow \min.$$

Преобразовав выражение, получим:

$$\sum[x_t^2 - 2x_t \cdot a_k \cdot \cos\frac{2\pi k}{T}t - 2x_t \cdot b_k \cdot \sin\frac{2\pi k}{T}t + +a_k^2 \cdot \cos^2\frac{2\pi k}{T}t + b_k^2 \cdot \sin^2\frac{2\pi k}{T}t +$$

$$2a_k \cdot b_k \cdot \cos\frac{2\pi k}{T}t \cdot \sin\frac{2\pi k}{T}t] = \sum x_t^2 - a_k^2 \cdot \frac{T}{2} - b_k^2 \cdot \frac{T}{2} = \sum x_t^2 - \left(a_k^2 + b_k^2\right)\cdot\frac{T}{2}$$

Определение наличия периодической составляющей

Если среднее значение временного ряда равно нулю (т.е. тренда нет), то выражение $\sum x_t^2=\sum(x_t-0)^2=D(x_t)$ – есть дисперсия. И в этом случае r_k^2 показывает вклад тригонометрической пары с периодом $\frac{T}{2}$ в общую дисперсию ряда. Для тригонометрической пары с периодом $\frac{T}{k}$ r_k^2 показывает степень синхронных колебаний исходного ряда и тригонометрической функции с периодом $\frac{T}{k}$.

СЕЗОННОСТЬ

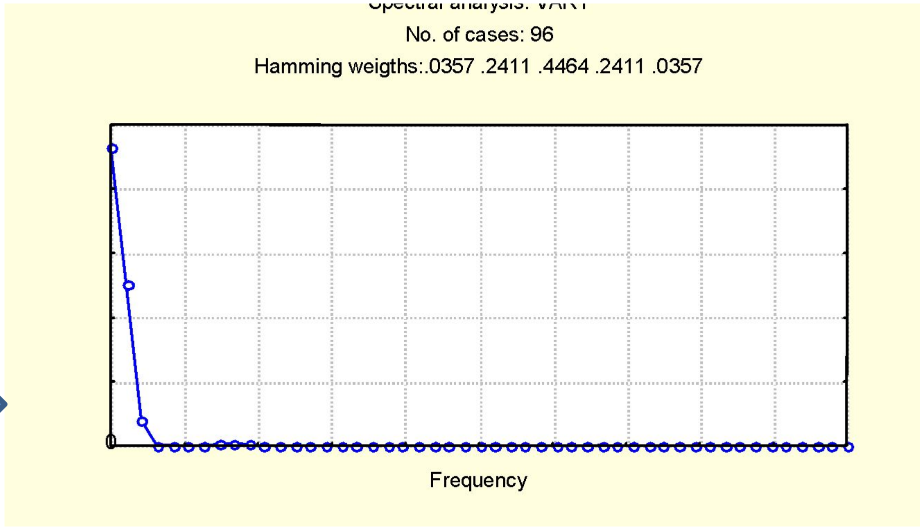
/цикличность

Периодограмма ряда, содержащего сезонную компоненту (с периодом 12)

Спектрограмма ряда, содержащего тренд.

Спектр будет выражаться кривой
типа $y=1/x^2$ с высоким пиком в начале
отсчета.

ТРЕНД



Что такое коррелограммы

С помощью коэффициентов автокорреляции, можно измерить связь между текущими и прошлыми значениями исходного временного ряда (сдвинутую на заданное количество лагов назад переменную называют *лаговой*) Совокупность коэффициентов автокорреляции образует автокорреляционную функцию (АКФ, (ACF)).

Коэффициент автокорреляции измеряет связь между текущими и прошлыми наблюдениями временного ряда и рассчитывается следующим образом:

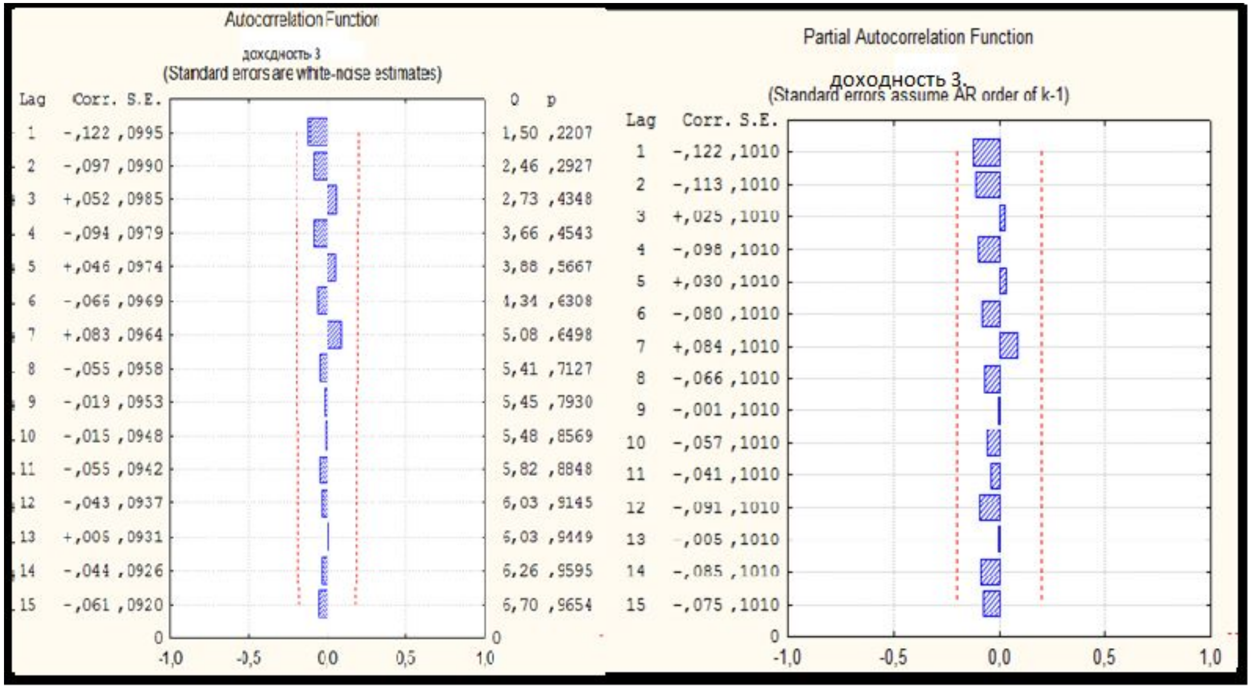
$$p_k = \frac{\sum_{t=1}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}$$

где k – количество лагов запаздывания, Y - среднее значение ряда.

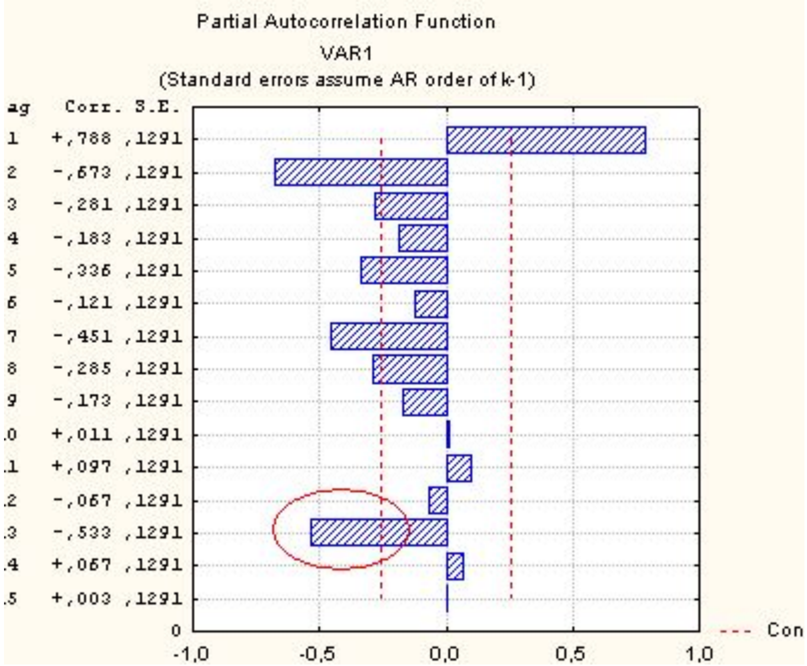
Частный коэффициент автокорреляции измеряет связь между текущим значением переменной X_t и предыдущими значениями этой переменной X_{t-1}, X_{t-2}, …, X_{t-k}, когда влияние всех промежуточных временных лагов устранено. Частный коэффициент автокорреляции лежит в основе частной автокорреляционной функции (ЧАКФ).

Что такое коррелограммы

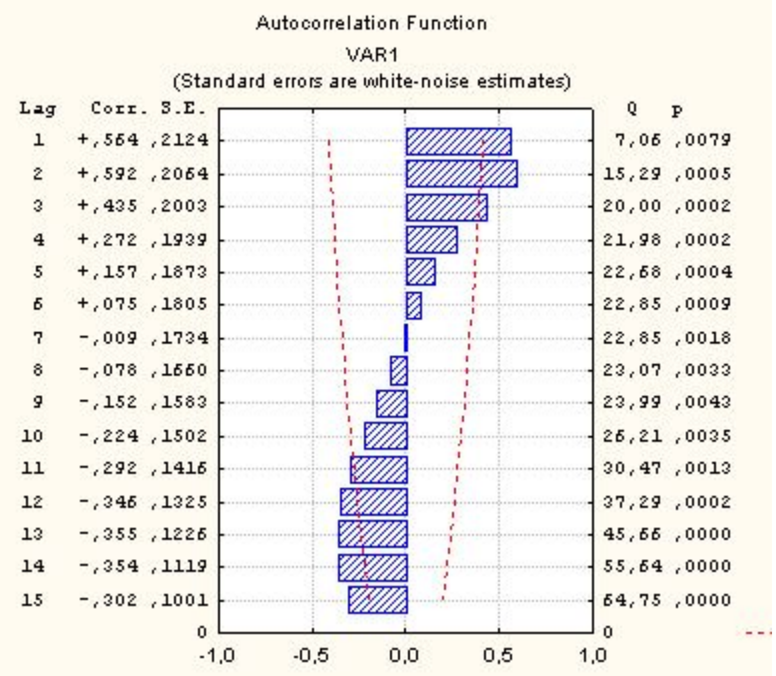
<u>Опр.</u> График автокорреляционной функции, где по оси абсцисс откладывается *k* – количество лагов запаздывания, а по оси ординат соответствующие значения коэффициентов автокорреляции, называется *коррелограммой АКФ*. Аналогичный график частной автокорреляционной функции называется *частной коррелограммой.*



Определение тренда и сезонности на основе коррелограмм



Сезонность (период 12)



Наличие тренда